Archives of Biochemistry and Biophysics 596 (2016) 118–125

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

dsRNA-protein interactions studied by molecular dynamics techniques. Unravelling dsRNA recognition by DCL1

Salvador I. Drusin [a, b], Irina P. Suarez [a], Diego F. Gauto [a], Rodolfo M. Rasia [a, c], Diego M. Moreno [d, e, *]

[a] Instituto de Biología Molecular y Celular de Rosario (CONICET-UNR), Ocampo y Esmeralda, Predio CCT, 2000 Rosario, Argentina
[b] Área Física, Departamento de Químico-Física, Facultad de Ciencias, Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, S2002LRK Rosario, Santa Fe, Argentina
[c] Área Biofísica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, S2002LRK Rosario, Santa Fe, Argentina
[d] Instituto de Química de Rosario (CONICET-UNR), Suipacha 570, S2002LRK Rosario, Santa Fe, Argentina
[e] Área Inorgánica, Departamento de Químico-Física, Facultad de Ciencias, Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, S2002LRK Rosario, Santa Fe, Argentina

ARTICLE INFO

Article history:
Received 10 December 2015
Received in revised form
2 March 2016
Accepted 9 March 2016
Available online 14 March 2016

Keywords:
Molecular dynamics
dsRNA recognition
dsRBD
miRNA processing
Binding free energy
Mismatch base pair

ABSTRACT

Double stranded RNA (dsRNA) participates in several biological processes, where RNA molecules acquire secondary structure inside the cell through base complementarity. The double stranded RNA binding domain (dsRBD) is one of the main protein folds that is able to recognize and bind to dsRNA regions. The N-terminal dsRBD of DCL1 in *Arabidopsis thaliana* (DCL1-1), in contrast to other studied dsRBDs, lacks a stable structure, behaving as an intrinsically disordered protein. DCL1-1 does however recognize dsRNA by acquiring a canonical fold in the presence of its substrate. Here we present a detailed modeling and molecular dynamics study of dsRNA recognition by DCL1-1.

We found that DCL1-1 forms stable complexes with different RNAs and we characterized the residues involved in binding. Although the domain shows a binding loop substantially shorter than other homologs, it can still interact with the dsRNA and results in bending of the dsRNA A-type helix.

Furthermore, we found that R8, a non-conserved residue located in the first dsRNA binding region, recognizes preferentially mismatched base pairs. We discuss our findings in the context of the function of DCL1-1 within the microRNA processing complex.

1. Introduction

MicroRNAs (miRNAs) are a large class of small RNAs that act as post-transcriptional regulators by recognizing target mRNAs through base pair complementarity. In plants, they have major roles in the regulation of development and physiology of the organism [1]. MicroRNAs are peculiar among small RNA molecules because they originate in an endogenous transcript, primary-microRNA (pri-miRNA), that contains the miRNA itself in a hairpin structure. The transcript is then processed to yield the 22 nucleotide product, the mature miRNA. The biogenesis of miRNAs in plants is carried out in the nucleus by a complex formed by the proteins Dicer-like 1 (DCL1), HYL1 and SERRATE. DCL1 is the active protein, as it harbors the tandem RNAseIII domains that perform the staggered cuts leading to the separation of the miRNA from the pri-miRNAs [2]. DCL1 shares the same domain architecture as other Dicer enzymes, with two C-terminal double stranded RNA binding domains (dsRBDs) [3]. Disruption of the tandem dsRBDs is lethal to the organism, showing that these domains are essential for the function of the enzyme [4].

The dsRBD domains recognize double-stranded RNA segments. These domains adopt a common fold, α-β-β-β-α, where the two α-helices are supported on the same face of the β-sheet (Fig. 1) [5]. The interaction of dsRBDs with double-stranded RNA has been characterized through the resolution of structures of dsRBD:dsRNA complexes [6–9]. Three binding regions can be identified in the structures. The first is the α-helix 1 (Region 1), the second corresponds to loop β1- β2 (Region 2) and the third one is located in the loop region between β3 and α2 (Region 3). Regions 1 and 2 interact

* Corresponding author. Instituto de Química de Rosario (CONICET-UNR), Suipacha 570, S2002LRK Rosario, Santa Fe, Argentina.
E-mail address: moreno@iquir-conicet.gov.ar (D.M. Moreno).

http://dx.doi.org/10.1016/j.abb.2016.03.013
0003-9861/

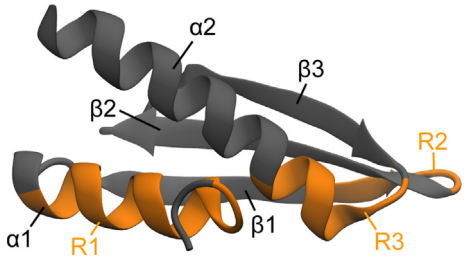

Fig. 1. Structure of the DCL1-1 domain. The three binding regions within the dsRBD are highlighted in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mostly with the ribose moieties of RNA and region 3 with the phosphate backbone. This binding mode gives dsRBDs specificity towards dsRNA binding against the dsDNA or RNA-DNA hybrids.

dsRBDs are widely distributed among eukaryote, bacterial and viral proteins, where they are usually present in multiple copies, alone or combined with other functional domains. The dsRBDs are in principle capable of binding to any element in a double-stranded RNA molecule, since complexes show few contacts with bases in the minor groove making sequence readout not possible. It has been argued that dsRBDs bind dsRNA in a non-sequence-specific fashion, and owe their specificity to the recognition of the 3D conformation of the RNA molecule [5]. However, some specificity can be achieved through the presence of imperfections in dsRNA or through interaction with other proteins or domains [6,7]. This feature is especially suited for recognition of miRNA precursors because of the essential lack of sequence conservation.

Recent findings show that some dsRBDs are capable of not only recognizing the imperfect structure of dsRNA, but show in some cases base-specific readout of the substrate. In the case of the dsRBD from yeast Rnt1p ribonuclease, an extension of the C-terminal helix allows the recognition of the AGNN loop present in its substrates [8,10]. More recently both dsRBDs from adenosine deaminase 2 (ADAR2) were shown to bind to particular locations on its substrate RNA guided through base-specific contacts in the minor groove of dsRNA. Notably, for each domain, the specific contacts occur in a mismatched base pair, G-G in one case and A-C in the other, that allow access to the bases into the minor groove of the A-form double strand [7]. In this work, the authors suggested that the distance between the ends of loop 2 and helix 1, which is finely modulated by the overall structure of the dsRBD, together with imperfections in the dsRNA, allow for the specificity of these tandem dsRBDs towards its substrate dsRNA. This evidence suggests that the dsRBDs within DCL1 could recognize the digestion sites by binding to specific regions of the precursor RNAs, either because of their structure or their sequence.

RNAse III enzymes from the Dicer family usually act together with helper proteins to achieve maximum specificity and activity. These helper proteins have no catalytic activity and contain only dsRBD domains. The role of the dsRBDs from these proteins and those from the Dicer enzyme itself in the reaction has not yet been elucidated. It is possible that the dsRBDs help to regularize regions of irregular secondary structure. Alternatively, binding of the dsRBDs may induce some distortion in the regular A-form structure of dsRNA that could be necessary for optimum activity of the RNAse III domains. In this sense, previous reports show that dsRBDs can induce bends in the elongated dsRNA structure [11]. Therefore, a potential function for some of the dsRBDs involved in the processing of miRNAs in plants could be the stabilization of the secondary structure elements in the precursor or the distortion of the regular A-form helix to allow the correct positioning of the active sites of DCL1 in the processing complex.

Whereas Dicer enzymes from animals have a single dsRBD at their C-terminus, some plant Dicer-like enzymes have two similar domains. In particular in DCL1 the dsRBDs differ in sequence and, presumably in function. Both dsRBDs bind dsRNAs [12,13] but dsRBD2 was shown to be essential for the correct localization of DCL1 in the “dicing bodies” of the nucleus [14] and it was also suggested to function as a non-canonical nuclear localization signal. The region of the substrate (pri-miRNA) to which they bind has not been determined so far, but based on the size of the precursors (>40 bp) and the length of dsRNA necessary to accommodate a single dsRBD, both can bind independently to any pri-miRNA.

Molecular Dynamics simulations (MD) are extremely useful to obtain detailed atomistic level information [15–19]. To the present, only two MD simulations of dsRBD:dsRNA complexes have been reported. A 2 ns simulation of the Drosophila Staufen dsRBD3 free bound to dsRNA showed that a high degree of conformational flexibility is retained upon complex formation, particularly in loop 2 [20]. More recently, a study on TRBP-dsRBD2 allowed describing the molecular basis for the discrimination between dsRNA- from DNA-containing duplexes by these domains [21]. This work also showed that the conformation of the DNA–RNA duplex can be altered by dsRBDs.

In previous works, we have characterized both tandem dsRBDs of DCL1 from *Arabidopsis thaliana*. The second domain shows a canonical fold and binds dsRNA forming heterogeneous complexes [12]. In contrast, the first domain (DCL1-1) is intrinsically disordered, but acquires the dsRBD fold when bound to substrate RNA [13]. We have shown that the complex between DCL1-1 and its substrate presents two species in slow exchange, a major one in which the protein is folded and a minor species in which the domain remains mostly unfolded but still bound to the RNA. This situation makes it extremely difficult to obtain direct experimental data on the structure of the complex, either by crystallization or by NMR.

In the present work we have resorted to molecular modeling and MD in order to obtain atomistic detail information on the behavior of DCL1-1 bound to dsRNA. We could characterize the binding mode of DCL1-1, and found that it induces a distortion in the A-form dsRNA helix. Furthermore a non-conserved residue in region 1 shows a preferential binding towards mismatched base pairs. Our findings reveal a potential function for this domain within the plant miRNA processing complex.

2. Methods

In order to build a structure of a DCL1-1:dsRNA complex, we resorted to the structure of another dsRBD:RNA complex formed by the second dsRBD of ADAR2 (ADAR2-2) and its substrate, GluR-2 LSL (dsRNA-A).

DCL1-1:dsRNA-A was built by generating a structural alignment of DCL1-1 with the structure of the ADAR2-2:dsRNA-A complex (Fig. 2A). The structure of dsRNA-D was calculated *in silico* (as explained below) and aligned with the previous structures in order to obtain the ADAR2-2:dsRNA-D and DCL1-1:dsRNA-D complexes. This procedure is illustrated in Fig. 2B.

2.1. Initial structures

The NMR structure of ADAR2-2 bound to dsRNA-A was obtained from the Protein Data Bank (entry 2L2K [7]). The structure of the folded form of DCL1-1 was calculated using the chemical shifts of the complex as restraints for the software CS-Rosetta [13,22]. DCL1-1:dsRNA-A was built *in silico* by structurally aligning the NMR

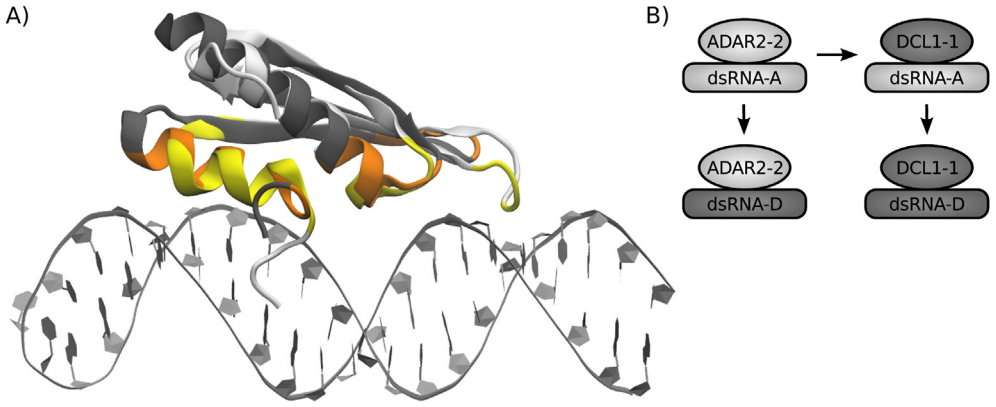

Fig. 2. **(A)** Aligned structures of ADAR2-2:dsRNA-A (white, with binding regions in yellow) and DCL1-1 (gray with binding regions in orange), showing the same structural features and binding regions. **(B)** Scheme of the strategy used to build the protein-dsRNA models *in silico*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

structure of DCL1-1 over the ADAR2-2:dsRNA-A complex using Multiseq plug-in of VMD [23,24]. The starting structure of the pri-miR172a lower stem region (dsRNA-D) was created from scratch with the MC-Fold | MC-Sym pipeline server service [25] and then subjected to a simulation protocol of 40 ns before using it for the construction of ADAR2-2:dsRNA-D and DCL1-1:dsRNA-D. The corresponding RNA sequences were: 5′-GCUGCUGUGGCAUCAUCAA-GAUUCA-3′ and 5′-AGAAUCUUGAUGAUG CUGCAUCGGC-3′ corresponding to the 5′ and 3′ ends of dsRNA-D, respectively. These two complexes were created by positioning the protein structure over the generated dsRNA manually using VMD. In both cases, several different positions were tested, and the ones with the lowest binding energy were selected for the simulations.

2.2. Classical molecular dynamic simulations

MD simulations were performed starting from NMR structures of protein-dsRNA complexes or from the models built *in silico* as described below. Each complex was immersed in a truncated octahedral periodic box with a minimum solute-wall distance of 8 Å, filled with explicit TIP3P water molecules [26], and then neutralized by addition of Na^+ cations using the AMBER leap module.

Molecular dynamic simulations were performed with the AMBER14 package [27,28], using the ff14SB [29] force field to describe the protein and OL3 [30,31] modifications to describe the dsRNAs. Particle-mesh Ewald (PME) was implemented for long range interactions with a cutoff distance of 12 Å [32]. Temperature and pressure were regulated with the Berendsen thermostat and barostat, as implemented in the AMBER package, using a time constant of 2 ps [33]. All bonds involving hydrogen were fixed using the SHAKE algorithm [34]. Each initial system was minimized using a multistep protocol, then heated from 0 to 300 K, and finally a short simulation at constant temperature of 300 K, under constant pressure of 1 bar, was performed to allow the systems to reach proper density. These equilibrated structures were the starting point for 200 ns of MD simulations at 300 K in the NVT ensemble.

2.3. Binding free energy calculations

The binding free energy of each complex and its per residue decomposition were performed with Molecular mechanics with generalized Born and surface-area solvation (MM-GBSA) [35] in AmberTools. Binding free energy is described through the following equations:

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand}$$

$$\Delta G_{bind} = \Delta E_{vdw} + \Delta E_{ele} + \Delta E_{GB} + \Delta E_{SA} - T\Delta S$$

where $G_{complex}$, $G_{receptor}$ and G_{ligand} are respectively complex, receptor and ligand free energy; ΔE_{vdw} and ΔE_{ele} are the contributions of van der Waals and electrostatic energy; ΔE_{GB} and ΔE_{SA} are the polar and non-polar desolvation contributions; and $-T\Delta S$ is the entropic contribution of the conformation at temperature T. Entropy change was calculated through the normal mode method. A total of 500 frames over the last 20 ns of each trajectory were used for the energy calculations, while only 50 of these frames were used for the normal mode analysis.

2.4. Calculation of structural parameters extracted from the MD simulation

Hydrogen bonds were defined with a maximum donor-acceptor distance of 3.5 Å and angle of 30°. dsRNA axial bending was calculated using Curves+ [36]. RMSD, RMSF and average structure calculations were made using the ptraj module from AMBER [37]. Sequence alignments were performed using Jalview [38].

2.5. RNA binding monitored by CD spectroscopy

An experimental assessment of the accuracy of the MD results was obtained through the determination of RNA binding by a DCL1-1 mutant by means of Circular Dichroism spectroscopy in order to confirm the importance of K51 and K52 in the binding energy. Titrations were conducted in 10 mM phosphate buffer (pH 7.0), 1 mM β-mercaptoethanol. Aliquots of a 90 μM DCL1-1-K51A/K52A solution were added to a 1 μM dsRNA-D. At each step, a Circular Dichroism spectrum was acquired between 190 and 300 nm, including the 260 nm RNA bases CD band. Spectra were acquired at 298 K on a Jasco-J810 CD spectrometer.

3. Results and discussion

3.1. Binding and comparison of dsRBD:dsRNA structures

The main goal of this work is understand the structural basis of the pri-miRNA recognition by DCL1-1. To this end, we performed molecular dynamics simulations in two different DCL1-1:dsRNA systems and we analyzed their structural features and thermodynamic properties.

DCL1-1 is intrinsically disordered in its free form and acquires its dsRBD fold when bound to substrate dsRNA. We had previously calculated a structure of dsRNA bound form DCL1-1 [13] using NMR techniques, but the size and exchange properties of the complex precluded the calculation of the RNA structure and of the binding interface. Moreover, the large change in the structure of the protein when going from the free to the bound forms results in a substantial displacement of all signals in the NMR spectra. For this reason, chemical shift perturbation, a method normally used to detect binding interfaces from NMR data, could not be applied in this case.

In order to build up an initial model structure, we resorted to using the structure of ADAR2-2 bound to dsRNA-A previously elucidated by NMR as a template. We chose this structure as our basis model as it was the only dsRBD:dsRNA complex available at the beginning of the present work where the dsRNA partner is continuous and with a natural sequence. ADAR2-2 presents 27% sequence identity with DCL1-1 (Fig. S1) and has a similar dsRBD structure (Fig. 2A), so we were able to build a model for DCL1-1:dsRNA-A by replacing ADAR2-2 with DCL1-1 in the complex structure. Furthermore, by replacing dsRNA-A with a fragment of the lower-stem region of pri-miR172a (dsRNA-D), which is a natural substrate for DCL1, we constructed models for ADAR2-2:dsRNA-D and DCL1-1:dsRNA-D.

These four models were used to study the interaction of DCL1-1 with dsRNA, and to compare the two proteins. A scheme of this protocol is shown in Fig. 2B.

3.2. Classical MD simulations of protein-dsRNA complexes

We performed 200 ns classical MD simulation of each complex in the NVT ensemble. RMSD and RMSF values are shown in Fig. S2. In each case, the protein RMSD value rises at the beginning and quickly converges to a value that varies less than 1 Å in the last 100 ns and all the final values are in the range of 1.5–3.5 Å. This clearly shows the stability of the system during the MD simulation. The RMSF, calculated through the last 100 ns of the MD, has peaks corresponding with the protein loops, as it would be expected. The only exception to this is the β3-α2 loop, which is very short and is part of substrate binding region 3.

Analysis of the MD shows that both ADAR2-2 complexes did not suffer any pronounced conformational changes, neither in the dsRBD nor the dsRNA. On the contrary, in the DCL1-1 complexes, a change in the dsRNA's conformation can be observed early in the simulation. In these two systems, the dsRNA adopts a significant curvature within the first 60 ns, and is conserved through the rest of the simulation. This unexpected change in the axis gets the dsRNA close to region 3 of DCL1-1. An average structure of the four complexes is shown in Fig. 3A. This figure shows how DCL1-1 induces an axial bend in its substrates. In contrast, ADAR2-2 shows no significant bending in any of its dsRNAs. This confirms that this effect is specific of DCL1-1.

Using the software curves+ [36], we measured the total axial bending of both dsRNAs in three different states: free, bound to ADAR2-2, and bound to DCL1-1. Axial bend was found to be greater in DCL1-1 complexes than in ADAR2-2 complexes or in its free state, as can be seen in Fig. 3B. This effect is more prominent in DCL1-1:dsRNA-A, where the bend is almost 30° larger than in ADAR2-2:dsRNA-A. In DCL1-1:dsRNA-D, while less outstanding, it is still significant, with a difference of slightly more than 10° with respect to ADAR2-2:dsRNA-D.

Induction of axial bend on dsRNA by dsRBDs was previously suggested based on RNA mobility shifts [11]. Our results on DCL1-1 describe a mechanism for this effect. DsRNA bending could be related to the domain's role within the miRNA processing complex in plants.

3.3. Binding free energies calculations

Binding free energy calculations of all four complexes were made using the MM-GBSA method. The binding free energy calculated throughout the simulation is plotted in Fig. S3. We used the free energy values calculated over the last 20 ns of simulation to perform a detailed analysis of energy components and per residue free energy, due to the fact that all models reached to an equilibrated conformation and steady energy values.

The MM-GBSA method for calculating free energies is useful to identify the individual energy contributions in the system. Additionally, the binding energy values can be compared when similar systems are studied, although it does not always provide accurate absolute free energies.

Table 1 shows the different contributions to the total binding energy (van der Waals, electrostatic, polar desolvation, and non-polar desolvation energies) for the four studied complexes. Analysis from the different components reveals that the main contribution comes from the electrostatic interaction that originates between the polar or charged residues of the protein and the phosphates, 2′ hydroxyl groups and bases of the dsRNA. However, these large values are in part neutralized by a very large polar desolvation. Van der Waals interactions are also favorable but contribute less to the total interaction energy.

In order to study the dsRBD:dsRNA interaction in greater detail, we decomposed the binding energy per residue to identify the residues that contribute to the interaction. Per-residue free energy decomposition throughout each protein is shown in Fig. 4. The three dsRNA binding regions of both proteins can be clearly distinguished. Residues outside the binding regions show interaction energy values close to zero, or even positive in the case of negatively charged residues.

As can be seen in the right panels of Fig. 4, of the three binding regions, region 3 is the one that contributes the most to binding with values around −30 kcal/mol, followed by region 1 around −25 kcal/mol, and finally region 2 around −10 kcal/mol. This data goes in line with experimental evidence obtained both in vitro [39–42] and in vivo [42] for several dsRBDs. It can be seen that in spite of DCL1-1 having a shorter β1-β2 loop, region 2 does bind to the dsRNA just as well as in ADAR2-2. This can be explained by the induced curvature of the axis of the dsRNA, which is driven into a position that allows the interaction with this shorter region.

The per-residue decomposition of the binding free energies highlights those residues that are essential for the interaction. In DCL1-1 the interaction is mostly driven by three residues one on each binding region: R8 in region 1, R27 in region 2 and K52 in region 3. In ADAR2-2 most residues in region 3 show favorable interaction energies, between −6 and −10 kcal/mol.

In order to validate these results we produced a mutant form of DCL1-1 where we replaced the two consecutive lysine residues in region 3 with alanine residues. We then followed the RNA binding/folding reaction by CD spectroscopy (Fig. S4). The mutant protein failed to bind dsRNA-D, thus giving support to the binding free energy results obtained from the MD simulation.

3.4. Protein-dsRNA interactions

Hydrogen bond and hydrophobic interaction analyses were made throughout the last 20 ns of the dynamics. The most relevant interactions are illustrated in Fig. 5 and Fig. S5. In our MD simulations of ADAR2-2:dsRNA-A we observed the same interactions reported by Stefl et al. on the structure solved by NMR [7] (V237, M238, N241, S258, N280, K281, K282 and K285). The MD simulation allowed us to detect other residues that participate in interactions with the RNA that are not evident on the analysis of the static

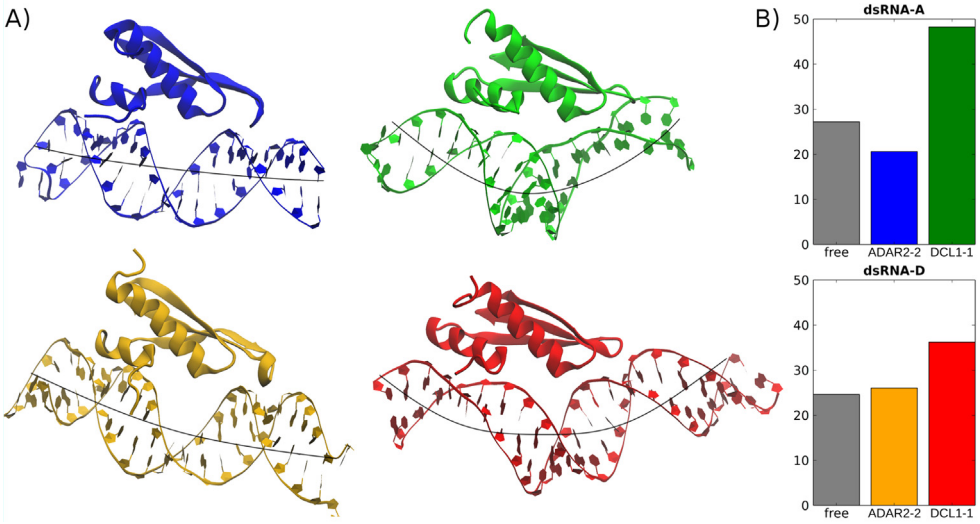

Fig. 3. (A) Average structure of the last 20 ns of the MD simulations from the four model complexes: ADAR2-2:dsRNA-A (blue), ADAR2-2:dsRNA-D (yellow), DCL1-1:dsRNA-A (green) and DCL1-1:dsRNA-D (red). The axial bend of the dsRNAs is illustrated as a black line. **(B)** Total axis bend of the dsRNA in its free state, bound to ADAR2-2 and bound to DCL1-1, respectively, for both dsRNAs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Total free binding energies of the four model complexes and their different components.

	ΔE_{vdw}	ΔE_{ELE}	ΔE_{GB}	ΔE_{GBsurf}	$\Delta E_{GBtotal}$	$-T\Delta S$	ΔG_{total}
ADAR2-2:dsRNA-A	−98.81	−3454.12	3476.20	−13.83	−90.56	48.58	−41.98
ADAR2-2:dsRNA-D	−105.97	−3552.02	3586.64	−14.85	−86.21	48.31	−37.89
DCL1-1:dsRNA-A	−103.97	−3014.34	3035.48	−14.16	−96.99	43.02	−53.97
DCL1-1:dsRNA-D	−105.42	−3137.95	3155.09	−14.49	−102.75	44.06	−58.69

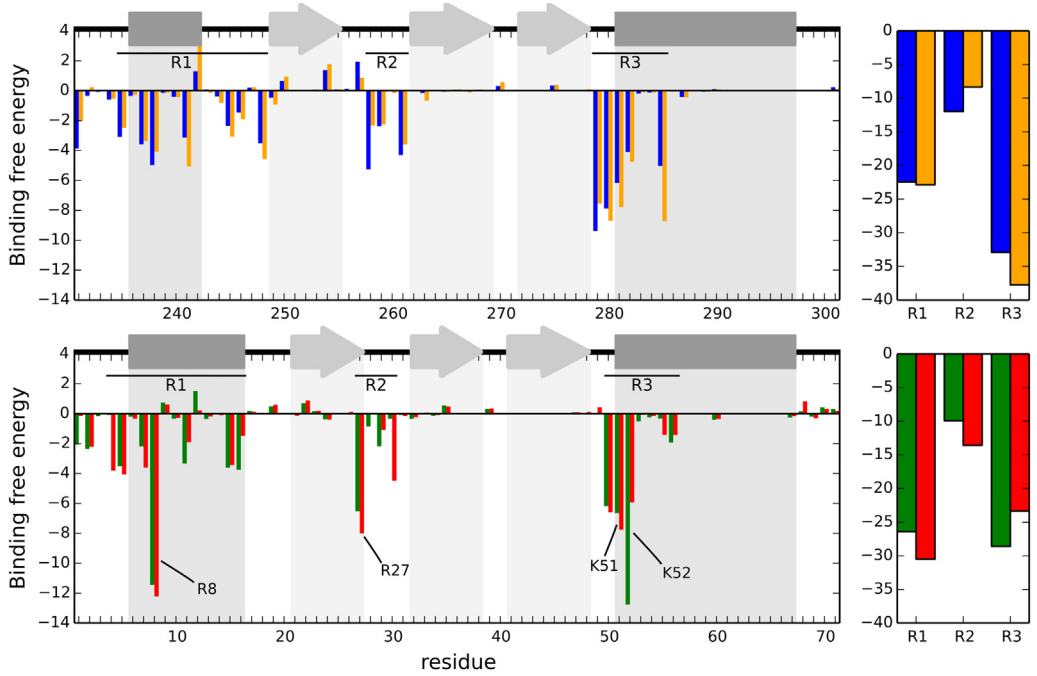

Fig. 4. Per-residue binding free energy of the four dsRBD:dsRNA complexes. ADAR2-2:dsRNA-A (blue) and ADAR2-2:dsRNA-D (yellow) are shown in the top plot, while DCL1-1:dsRNA-A (green) and DCL1-1:dsRNA-D (red) are shown in the bottom plot. The secondary structure of each domain is shown on top of each plot, with the three binding regions indicated as R1, R2 and R3. The total energy of each binding region is shown on the right of each plot. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

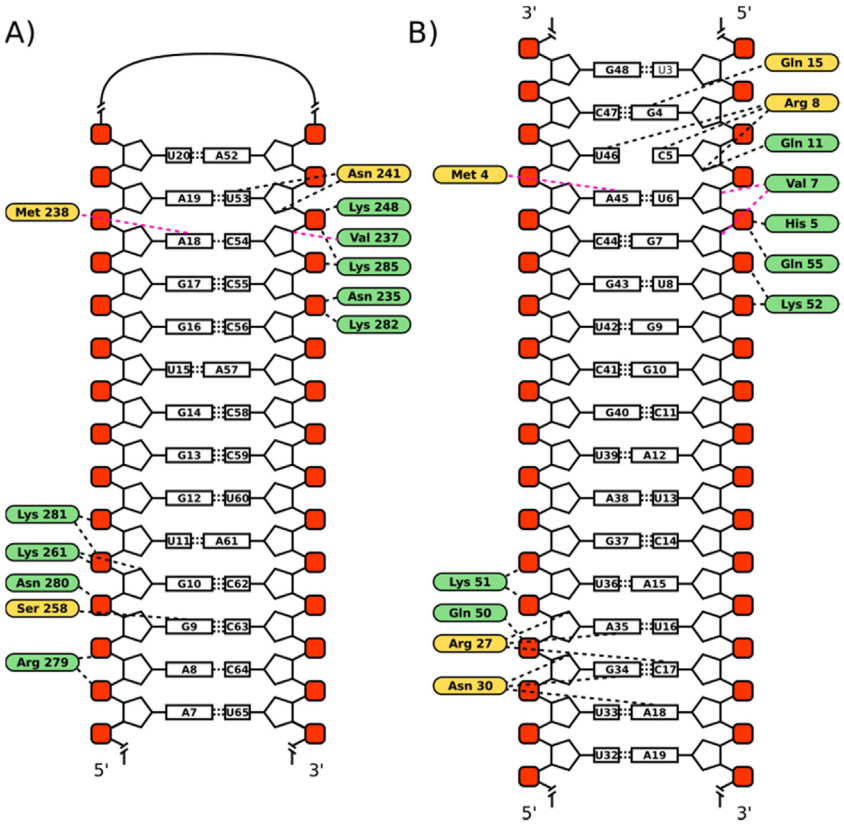

Fig. 5. Diagram showing the main interaction between the dsRBDs and the dsRNA in ADAR2-2:dsRNA-A **(A)** and DCL1-1:dsRNA-D **(B)**. In each dsRNA, the phosphate group is represented by a red square, the ribose sugar is represented by a pentagon, and the base is represented by a rectangle. Each basepair shows the corresponding number of hydrogen bonds in dotted lines. Contact residues from the dsRBD that interact with the backbone are colored green, while those that interact with the bases are highlighted in yellow. Hydrogen bonds are represented as black dotted lines, while hydrophobic interactions are represented as magenta dotted lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

structure (N235, K248, K261 and R279). The complex of ADAR2-2 with dsRNA-D shows roughly the same interactions as the one with dsRNA-A, and both DCL1-1 complexes behave very similarly, therefore contributing to the validation of the MD methodology.

Hydrogen bonds occur mainly between the nitrogen atoms of sidechain groups from arginine, asparagine, histidine, glutamine or lysine residues; and phosphate or hydroxyl groups from the RNA backbone. We found some residues, however, that form hydrogen bonds not with the backbone of the dsRNA, but with the nitrogenous bases. These residues are N241 and S258 in ADAR2-2, and R8, Q15, R27, G29 and N30 in DCL1-1. These bonds between DCL1-1 residues and the nitrogenous bases could mean that the domain could recognize key aspects from the dsRNA sequence in order to bind the precursor RNA in the correct position. All the hydrogen bonds from the four complexes are shown in detail in Fig. S6.

The hydrophobic contacts, on the other hand, are provided mainly by valine residues that contact the ribose in the RNA backbone as well as from methionine residues that in some cases interact with the nitrogenous bases of the dsRNA. An average structure from the simulation of DCL1-1:dsRNA-D is shown in Fig. 6 highlighting the main interactions.

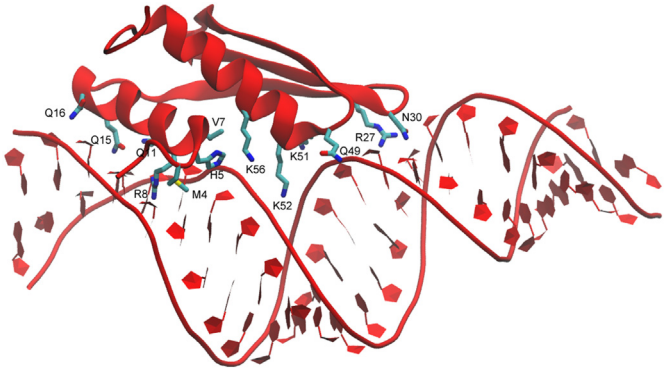

Fig. 6. Structure of the DCL1-1:dsRNA-D complex, showing the main residues interacting with the dsRNA.

3.5. Mismatch recognition by R8 in DCL1-1

Interactions between non-canonical bases are weaker than its canonical counterparts, which results in a disruption of the RNA homogeneity. This effect may act as a marker for certain types of aminoacids and be the structural key for dsRNA recognition by the miRNA processing proteins. Our results show that in DCL1-1, R8 strongly binds to mismatched base pairs. Fig. 7 shows the interaction of this residue with an A-C base of dsRNA-A and with a U-C base in dsRNA-D. Additionally, we have found that the presence of a G base in the base pair hinders the formation of the bond, as a G base's amine group repels the arginine's amine group and obstructs the H bond. In a similar way results obtained on the dsRBDs of ADAR2 show that modification of the identity of the base contacting the residue in an equivalent position causes a reduction in the binding affinity, further supporting the importance of this sequence specific contact in the formation of the complex [7].

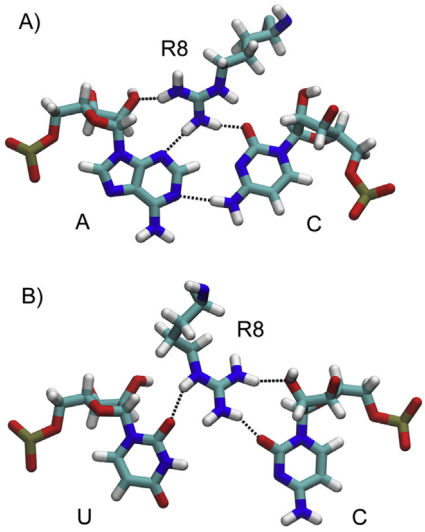

Fig. 7. Interaction of R8 from DCL1-1 with mismatched base pairs from dsRNA-A **(A)** and dsRNA-D **(B).**

These results suggest the key role of R8 in the recognition of dsRNA sequence which is further supported by the NMR experiments described below.

3.6. R8 participates in the DCL1-1 dsRNA-D encounter complex

In a previous work we had thoroughly characterized experimentally the interaction between DCL1-1 and dsRNA-D using NMR [13]. We had found that the free protein, which is intrinsically disordered, binds the substrate and acquires a folded conformation via an intermediate unfolded bound species. The disordered nature of the free protein gives rise to a $^{1}H^{15}N$-HSQC spectrum with little chemical shift dispersion and results in a global change when the folded bound species is formed. This in turn hinders a chemical shift perturbation analysis that would directly pinpoint those residues most affected upon formation of the RNA:protein complex. However, the analysis of the chemical shift perturbation between the unfolded free and unfolded bound species gives insight on those residues that participate in the initial recognition of the RNA, before the protein acquires its folded conformation.

The information gathered in the present work is further supported by a new analysis of the NMR data obtained previously, bringing interesting conclusions to light. We compared the difference in chemical shifts between the free unfolded form and the encounter complex obtained on the titration of DCL1-1 with dsRNA-D [13]. This analysis shows that the residues that are more affected in the encounter complex are distributed between helix 1, helix 2 and sheet 3. Perturbations of residues in helix 2 correlate with the increase in helical propensity as calculated from the ^{13}C secondary shifts. Therefore these perturbations are most probably due to the stabilization of the secondary structure in this region. In contrast the HN-N resonances of residues R8, E9, I38, G40, Q42 and G44 show a significant perturbation without a corresponding change in the secondary structure propensity. This indicates that the perturbation is brought about by the interaction with dsRNA and suggests that residues R8, E9, and Q42, i.e. those with longer hydrophilic sidechains, may be contacting the dsRNA in the initial encounter complex. These results are shown in Fig. S7.

The results of our MD simulation reveal that R8 is the residue that gives the largest contribution to the binding energy of the complex formed with dsRNA, and that the sidechain binds preferentially to mismatched base pairs within the helical framework. The previous analysis of NMR chemical shift perturbations suggests that R8 has a significant role in the initial recognition of the substrate.

The position occupied by R8 in helix 1 is involved in substrate recognition in many dsRBDs. While there is no sequence conservation at this position and the physicochemical character of the interaction varies among the different dsRBDs [5], there is conservation among different plant DCLs and animal Dicer proteins. Multiple sequence alignments carried on this region are shown in Fig. S8. The methionine residues located in an equivalent position in ADAR2-1 and ADAR2-2 were shown to be important for the protein activity, and the structure of the complexes show that they participate in base identity recognition via a hydrophobic contact between the methionine methyl group and the adenine H2 atom. Our results put forward a role for R8 in the recognition of mismatched base pairs in DCL1-1 and add further evidence for the function of equivalent residues in other dsRBDs in the discrimination of different dsRNA substrates and the role of the sequence readout mechanism in these domains.

4. Conclusions

In the present work we studied a total of four dsRBD:dsRNA complexes, which were built based on the structure of the ADAR2-2:dsRNA-A complex. These systems were subject to MD simulations in order to understand several properties of the complexes like stability, binding energy and axial bending.

We were able to determine the relative importance of each binding region of the dsRBD, and to show that although the β2-β3 loop of region 2 in DCL1-1 is shorter than in other similar dsRBDs, it can bind to the dsRNA just as well. The interaction with this binding region induces the bending of the dsRNA axis. This bending may be important for the protein's function in miARN precursors processing by DCL1.

Through the per-residue decomposition of the binding energy, we were able to identify the aminoacids that define this interaction. Among these residues, we found that in DCL1-1 R8 interacts preferentially with non-canonical base-pairs, as these mismatched bases open the minor groove of the dsRNA making the bases more accessible for the aminoacid sidechain. The presence of a guanine base in the mismatch, on the other hand, hinders the interaction because of its amine group, which repels the arginine's guanidine group.

The fact that R8 is one of the first residues of DCL1-1 to interact with the dsRNA, inducing the folding of the domain, highlights the importance of this residue in the dsRNA recognition, and suggests that mismatched base pairs within the whole precursor could provide an initial recognition and anchoring point for this residue.

Overall, in the present work we employed molecular dynamics simulations as an alternative to the use of spectroscopic techniques for the study of biological complexes. In this way, we could build a structure and characterize the binding features of this system, which would have been extremely difficult to accomplish through other structural biology techniques due to its dynamic nature. Moreover, these computational techniques gave us an atomistic picture of the dsRBD-RNA interaction, which could be applied in other biological systems.

Acknowledgements

This work was supported by grants from CONICET and ANPCyT (PICT-2012-1702 and PICT 2013–3281) to RMR and (PICT-2011-2332) to DMM. RMR and DMM are members of CONICET. SID, IPS and DFG thank CONICET for PhD and postdoctoral fellowships.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.abb.2016.03.013.

References

[1] W. Filipowicz, S.N. Bhattacharyya, N. Sonenberg, Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9 (2008) 102–114.

[2] N.G. Bologna, O. Voinnet, The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*, Annu. Rev. Plant Biol. 65 (2014) 473–503.

[3] Q. Liu, Y. Feng, Z. Zhu, Dicer-like (DCL) proteins in plants, Funct. Integr. Genomics 9 (2009) 277–286.

[4] S.E. Schauer, S.E. Jacobsen, D.W. Meinke, A. Ray, DICER-LIKE1: blind men and elephants in Arabidopsis development, Trends Plant Sci. 7 (2002) 487–491.

[5] G.G. Masliah, P. Barraud, H.T. Allain Fée, F.H.-T. Allain, RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence, Cell Mol. Life Sci. 70 (2013) 1875–1895.

[6] J. Gan, G. Shaw, J.E. Tropea, D.S. Waugh, D.L. Court, X. Ji, et al., A stepwise model for double-stranded RNA processing by ribonuclease III, Mol. Microbiol. 67 (2008) 143–154.

[7] R. Stefl, F.C. Oberstrass, J.L. Hood, M. Jourdan, M. Zimmermann, L. Skrisovska, et al., The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove, Cell 143 (2010) 225–237.

[8] Z. Wang, E. Hartman, K. Roy, G. Chanfreau, J. Feigon, Structure of a yeast RNase III dsRBD complex with a noncanonical RNA substrate provides new insights into binding specificity of dsRBDs, Structure 19 (2011) 999–1010.

[9] J.M. Ryter, S.C. Schultz, Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA, EMBO J. 17 (1998) 7505–7513.

[10] H. Wu, A. Henras, G. Chanfreau, J. Feigon, Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 8307–8312.

[11] E. Hitti, A. Neunteufl, M.F. Jantsch, The double-stranded RNA-binding protein Xlrbpa promotes RNA strand annealing, Nucleic Acids Res. 26 (1998) 4382–4388.

[12] P. Burdisso, I.P. Suarez, N.G. Bologna, J.F. Palatnik, B. Bersch, R.M. Rasia, Second double-stranded RNA binding domain of dicer-like ribonuclease 1: structural and biochemical characterization, Biochemistry 51 (2012) 10159–10166.

[13] I.P. Suarez, P. Burdisso, M.P.M.H. Benoit, J. Boisbouvier, R.M. Rasia, Induced folding in RNA recognition by *Arabidopsis thaliana* DCL1, Nucleic Acids Res. 43 (2015) 6607–6619.

[14] Y. Fang, D.L. Spector, Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants, Curr. Biol. 17 (2007) 818–823.

[15] D.M. Moreno, M. a Martí, P.M. De Biase, D.A. Estrin, V. Demicheli, R. Radi, et al., Exploring the molecular basis of human manganese superoxide dismutase inactivation mediated by tyrosine 34 nitration, Arch. Biochem. Biophys. 507 (2011) 304–309.

[16] P. Arroyo-Mañez, D.E. Bikiel, L. Boechi, L. Capece, S. Di Lella, D.A. Estrin, et al., Protein dynamics and ligand migration interplay as studied by computer simulation, Biochim. Biophys. Acta Proteins Proteom. 1814 (2011) 1054–1064.

[17] L.D. Alvarez, D.A. Estrin, Exploring the molecular basis of neurosteroid binding to the β3 homopentameric GABAA receptor, J. Steroid Biochem. Mol. Biol. 154 (2015) 159–167.

[18] N.V. Di Russo, M.A. Martí, A.E. Roitberg, Underlying thermodynamics of pH-dependent allostery, J. Phys. Chem. B 118 (2014) 12818–12826.

[19] J.P. Bustamante, A. Bonamore, A.D. Nadra, N. Sciamanna, A. Boffi, D.A. Estrin, et al., Molecular basis of thermal stability in truncated (2/2) hemoglobins, Biochim. Biophys. Acta Gen. Subj. 1840 (2014) 2281–2288.

[20] T. Castrignanò, G. Chillemi, G. Varani, A. Desideri, Molecular dynamics simulation of the RNA complex of a double-stranded RNA-binding domain reveals dynamic features of the intermolecular interface and its hydration, Biophys. J. 83 (2002) 3542–3552.

[21] L. Vukovic, H.R. Koh, S. Myong, K. Schulten, Substrate recognition and specificity of double-stranded RNA binding proteins, Biochemistry 53 (2014) 3457–3466.

[22] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J.M. Aramini, G. Liu, et al., Consistent blind protein structure generation from NMR chemical shift data, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 4685–4690.

[23] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph 14 (1996) 27–28, 33–38.

[24] E. Roberts, J. Eargle, D. Wright, Z. Luthey-Schulten, MultiSeq: unifying sequence and structure data for evolutionary analysis, BMC Bioinforma. 7 (2006) 382.

[25] M. Parisien, F. Major, The MC-fold and MC-Sym pipeline infers RNA structure from sequence data, Nature 452 (2008) 51–55.

[26] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926.

[27] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. DeBolt, et al., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, Comp. Phys. Commun. 91 (1995) 1–41.

[28] D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, et al., AMBER 14, University of California, San Francisco, 2014.

[29] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, J. Chem. Theory Comput. (2015), 150723121218006.

[30] M. Zgarbová, M. Otyepka, J. Sponer, A. Mládek, P. Banáš, T.E. Cheatham, et al., Refinement of the Cornell, et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles, J. Chem. Theory Comput. 7 (2011) 2886–2902.

[31] P. Banáš, D. Hollas, M. Zgarbová, P. Jurečka, M. Orozco, T.E. Cheatham, et al., Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins, J. Chem. Theory Comput. 6 (2010) 3836–3849.

[32] B.A. Luty, I.G. Tironi, W.F. van Gunsteren, W.F. van Gunsteren, Lattice-sum methods for calculating electrostatic interactions in molecular simulations, J. Chem. Phys. 103 (1995) 3014.

[33] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.

[34] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341.

[35] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, et al., Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, Acc. Chem. Res. 33 (2000) 889–897.

[36] R. Lavery, M. Moakher, J.H. Maddocks, D. Petkeviciute, K. Zakrzewska, Conformational analysis of nucleic acids revisited: curves+, Nucleic Acids Res. 37 (2009) 5917–5929.

[37] D.R. Roe, T.E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular synamics trajectory data, J. Chem. Theory Com. 9 (2013) 3084–3095.

[38] A.M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, G.J. Barton, Jalview version 2–a multiple sequence alignment editor and analysis workbench, Bioinformatics 25 (2009) 1189–1191.

[39] A. Ramos, S. Grünert, J. Adams, D.R. Micklem, M.R. Proctor, S. Freund, et al., RNA recognition by a Staufen double-stranded RNA-binding domain, EMBO J. 19 (2000) 997–1009.

[40] 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.

[41] B.C. Krovat, M.F. Jantsch, Comparative mutational analysis of the double-stranded RNA binding domains of Xenopus laevis RNA-binding protein A, J. Biol. Chem. 271 (1996) 28112–28119.

[42] P. Burdisso, F. Milia, A.L. Schapire, 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.