PCCP

The key role of electrostatic interactions in the induced folding in RNA recognition by DCL1-A

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-11-2017-007889.R1
Article Type:	Paper
Date Submitted by the Author:	06-Feb-2018

Complete List of Authors:	Zhao, Lingci; Chang Chun Institute of Applied Chemistry Chinese Academy of Sciences, State Key Laboratory of Electroanalytical Chemistry; jilin university , department of physics Rasia, Rodolfo; The Institute of Molecular and Cellular Biology of Rosario , Biophysics; School of Biochemistry and Pharmacy, University of Rosario, Chemical Biology Suarez, Irina; The Institute of Molecular and Cellular Biology of Rosario , Biophysics Gauto, Diego; The Institute of Molecular and Cellular Biology of Rosario , Biophysics Wang, Jin; State University of New York at Stony Brook, Department of Chemistry and Department of Physics

PCCP

ROYAL SOCIETY OF CHEMISTRY

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

The key role of electrostatic interactions in the induced folding in RNA recognition by DCL1-A

Lingci Zhao,[ab] Rodolfo M. Rasia,[c] Irina P. Suarez,[c] Diego F. Gauto,[c] and Jin Wang*[abd]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

The intrinsically disordered protein domain DCL1-A is the first report of a complete double stranded RNA binding domain folding upon binding. DCL1-A recognizes the dsRNA by acquiring a well-folded structure after engagement with its interaction partner. Despite the structural characterization of the interaction complex underlying the recognition of dsRNA has been established, the dynamics of disorder-to-order transitions in the binding process remains elusive. Here we have developed a coarse-grained structure-based model with consideration of electrostatic interactions to explore the mechanism of the coupled folding and binding. Our approach led to remarkable agreements with both experimental and theoretical results. We quantified the global binding-folding landscape, which indicates a synergistic binding induced folding mechanism. We further investigated the effect of electrostatic interactions in this coupled folding and binding process. It reveals that non-native electrostatic interactions dominate the initial stage of the recognition. Our results help improve our understanding of the induced folding of the IDP DCL1-A upon binding to dsRNA. Such methods developed here can be applied for further explorations of the dynamics of coupled folding and binding systems.

Introduction

Intrinsically disordered proteins (IDP) are a class of naturally abundant proteins and protein regions that lack well-defined three-dimensional structures in their free forms under physiological conditions. Such IDPs play essential roles in various biological functions, frequently related to signaling, molecular recognition, cell regulation, and closely implicated in some human diseases[1]. Early bioinformatics studies have shown that more than 47% of eukaryotic proteins have consecutive disordered residues, and the ratio further raises up to 66% for signaling proteins[2], and only about 32% of the crystal structures of the Protein Data Bank show completely ordered structures (containing no missing election density)[3]. The advantages conferred by the dynamic nature of unbound IDPs have been suggested to derive from the weak contacts between the intrinsically disordered protein and the target that is subsequently "reeled in" to the folded complex. This gives the IDP a greater effective capture radius and rapid stepwise association that increases the chances of binding[4]. The intrinsic flexibility leads to moderate affinity coupled with high specificity[5], thus enabling IDPs to adapt their structures so as to perfectly match the binding sites of their native partners[6]. This smart One-to-Many binding mode has become a hallmark of IDPs that allow the structural plasticity for binding multiple targets[1]. A number of experimental data suggested that the unbound IDPs often transiently acquire the secondary structures present in the well-ordered complexes, that is, they are not totally unfolded random coils. Since IDPs can fold into ordered structures after engagement with their interaction partners, the binding process from isolated unfolded states towards the well-structured native bound state involves significant disorder-to-order transitions. However, the mechanisms of the coupled binding and folding are not yet well understood. Addressing this issue will contribute to answer fundamental questions about how the intrinsic flexibility as well as transiently preformed structural elements influences the functioning of the IDPs.

Traditional experimental methods, such as X-ray crystallography and NMR, can only provide the static snapshot or the averaged ensemble description of biomolecules, making them improper to probe the determining dynamics of the IDPs. Although we have seen impressive progress in biophysical techniques in recent years, including single-molecule spectroscopy such as florescence energy transfer[7], and small-angle X-ray scattering (SAXS),

* Corresponding authors
[a] College of Physics, Jilin University, Changchun, Jilin, 130012, People's Republic of China
[b] State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, 130022, People's Republic of China
[c] Instituto de BiologÃ■a Molecular y Celular de Rosario. 27 de Febrero 210 bis, predio CCT, 2000 Rosario, Argentina
[d] Department of Chemistry and Physics, State University of New York, Stony Brook, NY 11794-3400 **E-mail**: jin.wang.1@stonybrook.edu

it is still challenging to obtain a detailed description of the intrinsic flexibility of the disordered structures and of the mechanism of the coupled folding and binding, due to the limited spatial and temporal resolution of these measurements.

On the other hand, molecular dynamics simulation, as a prevalent computational tool, can be used to trace the dynamics of many kinds of biomolecules at the atomistic scale. It has been very successful in identifying key residues for function and providing insights of explorations of various biological systems. For instance, Tan et al. reported an interesting study on the specific recognition between zinc-finger domains and RNA involving binding induced conformational change[29]. However, if one wishes to exhaustively explore the energy landscape of a moderately sized protein, the traditional simulations of protein structures and dynamics with atomistic detail remain difficult for many protein systems, owing to the limitation of accessible timescale compared to the large size of protein conformational space. Despite considerable developments in hardware, the use of lower resolution models that degrades the level of protein representation from all-atom to coarse-grained is still the only way to efficiently explore the fundamental features of biomolecular systems in longer timescales. The conformational space can be restricted via replacing residue fragments or even the whole residue with united atoms. Therefore, coarse-grained models, with the reduction of degrees of freedom and a consequent longer accessible timescale, can provide an insightful first order approximation to the global conformational landscape. Since natural proteins are specific copolymers encoded by evolution, the molecular mechanics of the well-defined three-dimensional structures involve a complex interplay of van de Waals interactions, hydrogen bonds, electrostatic interactions, etc. The reduced degrees of freedom at the coarse-grained scale make it difficult to produce an accurate description of the structural properties and the specific processes of conformational transitions of biomolecules. One widely used strategy to overcome this difficulty is to employ a specific force field, in which only the information of the known native structures are taken into account. This structure-based model (SBM), also called G$\bar{o}$-like model, ensures that the native fold of a protein is the global minimum of the free energy landscape. From the energy landscape perspective, proteins have evolved a funnel-like thermodynamic landscape biased towards the native state, allowing them to fold to the unique stable conformation sufficiently quickly. The funnel is believed to originate from the evolutionary selection, so unstable and nonfunctional conformations would be evolutionarily unfavorable[8].

In the present work, we set up a coarse-grained structure-based model to study the folding of IDP DCL1-A upon binding to a double stranded RNA. DICER-LIKE1 (DCL1) is the ribonuclease that carries out microRNA biogenesis within the microRNA processing complex in plants, by excising the long microRNA precursors aided by dsRNA-binding protein HYL1 and the zinc-finger protein SERRATE[9–11]. The enzyme DCL1 has two tandem double stranded RNA binding domains (dsRBD) in its C-terminus, in which the first one is referred to as DCL1-A. Notably, this domain is intrinsically disordered in its free form, but acquires a folded conformation in the presence of its substrate RNA, being the first report of a complete dsRBD folding upon binding[12]. The IDP DCL1-A has been previously investigated by experiments, including fluorescence anisotropy assays, NMR, and circular dichroism[12], and by traditional physics-based simulations[13]. Rasia et al. proposed a two-step equilibrium binding model for the interaction of DCL1-A with substrate RNA, which predicted a dominant sequential pathway for the coupled folding and binding system that is initialed by loosely binding to the partner via a partially preformed folded structure, and followed by rearranging the interactions to fit to the binding site and simultaneously accomplish the final fold. They subsequently resorted to molecular dynamics simulations to obtain atomistic detail information on the behavior of DCL1-A bound to dsRNA and characterize the key residues involved in binding. Despite the global aspects of pri-miRNA processing and the structural features of the interaction complex underlying the recognition of substrate dsRNA have been established, there is little information on the conformational dynamics of the disorder-to-order transitions along the binding pathway. Here, from a global perspective, we developed a coarse grained structure-based model to explore the conformational transitions of DCL1-A during its binding to substrate RNA and quantify the underlying functional conformation energy landscape. Owing to the simplicity and computational efficiency of our coarse grained structure based model at the residue level compared to the detailed atomistic force field model, we provide a quantitative description of the binding process in terms of thermodynamics and kinetics from enough samplings of binding-folding transition events during the simulations. In particular, we implemented Debye-Hückel model to describe the electrostatic interactions due to the highly charged characteristic of DCL1-A and dsRNA. The predicted recognition mechanism is in remarkable agreement with the experimental observation that the folding event appears to happen on the surface of dsRNA being preceded by the formation of a loosely bound complex[12]. We also investigated the structural features of DCL1-A at different binding stages. Our results are in agreement with the experimental data showing that the third binding region which has tendency to transiently populate folded conformations does lead the IDP to its partner in the initial recognition. We further predict the distributions of active residues in the encounter states during the initial binding process.

Methods

The structure-based models (SBM) have been proven successful at combining the energy landscape theory of protein folding/binding with an efficient molecular dynamics simulation. The plain SBM only considers the interactions existing in the native structure resulting in a smooth energy landscape to guarantee the feasibility of simulation. In order to study the effect of electrostatic interactions on our system, we developed a modified coarse-grained structure-based model with consideration of electrostatic interactions described by the Debye-Hückel model. Each amino acid in DCL1-A is represented by a Cα atom and each nucleotide in the double stranded RNA is represented by three beads, located at the centroid of phosphate, sugar, and base groups without heterogeneity. In our model, Lys and Arg in DCL1-

A have one positive point charge, and Asp and Glu in DCL1-A, as well as the phosphate beads in the dsRNA carry one negative point charge. This model contains the IDP DCL1-A with 71 residues and a double-stranded RNA hairpin corresponding to the lower stem region of pri-miR172a from A. thaliana. The contact map gives all possible interactions between pairs of residues in a given structure. In this study, 143 intramolecular folding contacts within DCL1-A were identified by Contacts of Structural Units (CSU) software[28]. The residue-residue contact map is based on a full list of atomistic contacts derived from a detailed calculation of the solvent-accessible surface of every atom. And 73 intermolecular binding contacts were constructed by cutoff algorithm: if the distance between any heavy atoms in DCL1-A and RNA is smaller than 5 Å, there is a native intermolecular contact between DCL1-A and RNA. We introduce separate parameters ε_f and ε_b to rescale the nonbonded potentials of intramolecular and intermolecular parts to match the experimental data (see the details of parameter calibration in the Supporting Information). Notably, the dsRNA is kept rigid and frozen in space while the DCL1-A is set to be free during all the simulations. The rigidity of the RNA segment reduces the flexibility of the DCL1-A:dsRNA interface, and will promote the efficiency of the binding. However, the binding pattern is supposed not to be qualitatively changed. Therefore, the functional form of the potential energy used in our model can be expressed by:

$$\begin{aligned} V = & \sum_{bonds} \varepsilon_r (r-r_0)^2 + \sum_{angles} \varepsilon_\theta (\theta-\theta_0)^2 \\ & + \sum_{dihedrals} K_\Phi^{(n)} (1-\cos(n\times(\Phi-\Phi_0))) \\ & + \sum_{contacts} \varepsilon_{ij} \left(5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} \right) \\ & + \sum_{non-native} \varepsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12} + V_{Debye-Hückel} \end{aligned} \quad (1)$$

In our work, the parameters are derived from the original folding/binding studies[23,24], namely, ε_r=100ε, ε_θ=20ε, $K_\Phi{}^{(1)}$=ε, $K_\Phi{}^{(3)}$=0.5ε, ε_{NC}=ε and σ_{NC}=4$\mathring{A}$. The interaction strength of Lennard-Jones potential is proportional to the residue-type-based statistical potential reported by Miyazawa and Jernigan to modulate the energetic heterogeneity[25]. The parameter of the nonbonded contacts, ε_{ij}, is set as follows:

$$\varepsilon_{ij} = \left(\gamma \left(\frac{\varepsilon_{ij}^{MJ}}{\bar{\varepsilon}^{MJ}} - 1 \right) + 1 \right) \quad (2)$$

In which $\varepsilon_{ij}{}^{MJ}$ is the original MJ potential, $\bar{\varepsilon}^{MJ}$ is the mean value of the entire set of LJ weights in the whole system, γ has been set to 1.0 corresponding to the "flavored model"[26].

The electrostatic interaction is described by the Debye-Hückel model, which can tune the strength of charge-charge interactions via varying the salt concentrations:

$$V_{Debye-Hückel} = \Gamma_{DH} \times K_{Coulomb} \times B(\kappa) \sum_{i,j} \frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon r_{ij}} \quad (3)$$

$K_{coulomb}$=$4\pi\varepsilon_0$= 138.94 kJ·mol^{-1}·nm·e^{-2} is the electric conversion factor; B(κ) is the salt-dependent coefficient; κ^{-1} is the Debye screening length, which is modulated by the salt concentration C_{Salt} ($\kappa \approx 3.2\sqrt{C_{Salt}}$). ε is the dielectric constant and was set to 80 throughout the simulations. Γ_{DH} is the energy scaled parameter which was introduced to balance the total energy. In the present simulation, considering that κ is 1.24 nm^{-1} under the physiological salt concentration of 150mM, we set Γ_{DH} = 0.535, so that the DH potential for two opposite charges located at a distance of 0.5 nm could match their native contact energy. More detailed meaning of Debye-Hückel model can be found here[27].

All pairs of residues that do not involved in bond, angle, dihedral and native contact terms are considered as non-native contacts and assumed to have only volume-exclusion repulsions (σ_{NC}). It is supported by many studies that the folding and binding mechanisms are encoded in the topology of proteins in their native states[15–17]. Therefore, we have the significant simplification that only native interaction patterns are taken into account so as to remove the energetic roughness. Energy landscape theory proposes that proteins have evolved a smooth and funnel-like thermodynamic landscape biased toward the native state, allowing them to fold to a unique stable structure on biological timescales[30]. A random sequence polypeptide often does not have sufficient time to explore all possible conformations in search of the native state; therefore native interactions must play a key role in guiding the conformational search in the direction of the native structure as a result of the evolution selection, while the non-native contacts are much less preferred from natural selection. For polymer physics, if the interactions between the non-native residues are very weak (due to the far away distance in the native structure) compared to the interactions between the native residues, then the only remaining significant interactions will be excluded volume repulsions. It is important to notice that non-native contact pairs can have electrostatic interactions if the associated residues are polar or charged.

Our simulations were performed at a series of dilute solution in the range of 50mM $\sim$ 300mM to control the strength of electrostatic interactions. To explore the thermodynamics, we performed a group of constant temperature simulations at T_s = 160.90 K (see details in the SI Appendix), which were initialized from 40 different configurations either dissociative or dimeric attempting to observe more transitions. And the total simulations accumulated 416 binding/unbinding transitions within 40 μs, ensuring the statistical reliability for thermodynamic properties analysis.

For kinetics, we simulated 200 individual trajectories started from varying dissociative configurations with different initial velocities at a lower temperature T = 0.95T_s. Each simulation accounts for a single binding event, and was ended as soon as reaching the native bound state. The unbound state with no intermolecular contact (Q_i = 0) is the initial state. And the loosely bound encounter complex is formed by capture event. Then the

encounter complex proceeds either to evolve to the bound state, or escape to the unbound state. The first passage time (FPT_{on}) of binding-on events (from unbound state to bound state), the mean passage time (MPT_{cap}) of capture events (from unbound state to encounter complex) and the first passage time (FPT_{evo}) of evolution events (from encounter complex to bound state) were collected to compare the rate of binding at different salt concentrations by averaging the corresponding 200 trajectories.

Results and discussion

Binding induced folding in dsRNA recognition by DCL1-A

To illustrate the process of coupled folding and binding of DCL1-A with dsRNA, we performed computer simulations by developing a coarse-grained structure-based model with consideration of electrostatic interactions. Firstly, the free energy landscape of DCL1-A and dsRNA complex was plotted along Q_i and Q_f to obtain a global thermodynamic view of the mechanism of the coupled folding and binding process. Q_i, which is the fraction of native contacts between DCL1-A and dsRNA, is the binding reaction coordinate, and monitors the degree of binding process. Q_f, which is the fraction of native contacts for folding of DCL1-A, is the folding reaction coordinate, and shows the ratio of folded content. As shown in Fig. 1, our IDP DCL1-A prefers the route from the conformational region with high Q_i value and low Q_f value (bound-unfolded state) to the region with high Q_i value and high Q_f value (bound-folded state). Thus the 2D free energy surface corresponds to the induced folding route rather than the conformational selection route (a path along the lower right corner, which indicates folding before binding). Therefore, our SBM model gives a strong support to an induced folding mechanism since the folding of DCL1-A takes place after binding. This result is highly consistent with the view from the previously published experimental study[12]. In the unbound state, the isolated DCL1-A comprises a large number of disordered or partially disordered conformations with Q_f about 0.2 (but less than 0.5), according with its description as an IDP. There is no apparent binding transition barrier (means a low activation energy) in the free energy profile between the unbound state (Q_i=0.0) and the encounter complex (An encounter complex was defined when the system evolved from Q_i=0.0 to Q_i>0, and usually had only one intermolecular native contact)[14]. Thus the initial recognition of dsRNA by the IDP DCL1-A occurs very quickly and early. The snapshots of the DCL1-A binding to the dsRNA with various Q_i are shown around the free energy landscape in Fig. 1 to illustrate the binding process. In the encounter complex ensemble, some local regions of DCL1-A always make native contacts with dsRNA while the majority of DCL1-A remain largely disordered. And the initial captures are inclined to involve one of the three regions (R1, R2, and R3) of contacts being formed first. The representative frames of the three binding regions forming the intermolecular contacts respectively with dsRNA as the initial recognition are highlighted in the blue box in Fig. 1. To recognize the double-stranded RNA segments, DCL1-A adopts a topology α-β-β-β-α, where the two alpha helices are packed on the same side of the beta sheets (see the top of Fig. 1). The three binding regions can be identified in the structure: the first is the α-helix 1 (R1), the second is located on the loop between β1 and β2 (R2), and the third is related to the N-terminus of α-helix 2 in addition to the loop between β3 and α-helix 2 (R3).

The Characterization of the IDP DCL1-A in its unbound state

We can see that the free form of DCL1-A is overall flexible with little folded content (Q_f about 0.2). However, considering that its canonical fold in the presence of the dsRNA has different secondary structure elements, we examined the fraction of native contacts for folding of α-helix 1, β-sheet, and α-helix 2, respectively. Fig. 2(a) shows that α-helix 2 contributes most of the folded content to the IDP DCL1-A in its unbound state with a comparatively large Q_f (more than 0.5), while the other two secondary structure regions contain more disordered conformations with Q_f less than 0.22. Although the structure based model will slightly overestimate the populations of helical content in the free form, this result is quite consistent with our experiments that the free protein is not completely unstructured (3% helical content at 298K), and further, different from the N-terminal half of the IDP, the C-terminal half shows some tendency to populate helical structures[12].

On the other hand, it has been shown that intrinsically disordered regions with charged residues are widely found in DNA/RNA-binding proteins, and play an important role in many molecular events, especially the initial recognition[15,16]. In the present work, we extended the structure-based model (SBM) to consider the long-range electrostatic interactions described by Debye-Hückel model due to the abundant electrostatic component in the formation of the DCL1-A and dsRNA complex. We investigated the effect of different strengths of electrostatic interactions upon the structure of isolated DCL1-A via varying the salt concentrations, because in Debye-Hückel model, the screening effect of implicit ion decreases with lowering the salt concentration and hence increases the strength of electrostatic interactions. We calculated the Q_f of the three secondary structure regions with increasing the salt concentrations (gradually decreasing the electrostatic interactions) from 50 to 300 mM. Fig. 2(a) shows that the unbound DCL1-A was almost independent of the salt concentrations, implying that the electrostatic interactions do have hardly impact on the unbound state of DCL1-A. We also examined the inter-residue distances within the unbound IDP DCL1-A. The distance map in Fig. 2(b) reveals that DCL1-A lacks an overall well-defined structural features, and behaves just as a typical intrinsically disordered protein. Notably, the C-terminal half shows some tendency to preform some helical structures, consistent with the abovementioned analysis of Q_f and the experimental results. The inter-residue distances of the unbound state seem to be determined simply by the separation distance within the amino acid sequence[17]. On the contrary, the distance map of the native bound DCL1-A in Fig. 2(c) displays a well-structured protein, in which the secondary structure elements (α-β-β-β-α) precisely correspond to that shown in Fig. 1.

Non-native electrostatic interactions dominate the encounter complex

In the initial DCL1-A:dsRNA recognition the encounter complex is quickly formed by 3D diffusion. To reach the encounter complex ensemble, some stabilizing interactions between DCL1-A and dsRNA have to compensate the translation entropy lost. As shown in the blue box in Fig. 1, our system does not always involve the same contacts being formed first. All three binding regions are possible candidates to guide the system to the encounter complex. However, there are limited native intramolecular protein interactions formed in the encounter complex as shown in the free energy landscape with Q_i<0.2, indicating that the stabilizing interactions in the encounter complex are mostly non-native (transient and vanish in the native structure). By investigating the protein-RNA interaction energy during different binding stages (Table 1), we can see clearly that there are most non-native electrostatic interactions (-8.67) in the encounter complex, while the energy of native interactions (-1.65) are relatively low. It is the non-native electrostatic interactions that dominate the formation of the encounter complex. To better characterize the interactions in the encounter complex and identify the key residues for the recognition of dsRNA by DCL1-A, we introduced a cut-off algorithm which can take the non-native contacts into account to calculate the contact map and average contacts of each residue or nucleotide (for a detailed definition of cut-off algorithm,see SI). In the encounter complex, the C-terminal half of DCL1-A, precisely speaking, the third binding region (R3) forms wide contacts with the dsRNA, while the N-terminal half has little intermolecular interactions (Fig. 3(a)). It is noteworthy that most of the contacts are actually non-native contacts. The averaged contact number shown in Fig. 3(b) makes clear that the C-terminal half corresponding to the third binding region of DCL1-A forms most of the contacts within the encounter complex. A detailed analysis of the contacts for the phosphate, sugar, and base within the RNA segment indicates that the interactions involving phosphates have a strong influence on the encounter complex (Fig. 3(c)). For this reason we thought that electrostatic interactions play a determinant role in the initial recognition. The contact map in the encounter complex with the salt concentration 50mM (Fig. S1) reveals more and more widespread non-native contacts formed in the C-terminal half of DCL1-A as we increase the electrostatic interactions. We also plotted the average contacts between DCL1-A and dsRNA in the native bound state in Fig. 7. Compared to the other two types of contacts within the RNA segment in Fig. 7(c), we found that the interactions involving the phosphates do not make significant contributions to the native bound state, in contrast with the encounter complex. In conclusion, the non-native electrostatic interactions can act as "steering forces" in the initial recognition to facilitate the early binding.

The intermediate state emerges as lowering the salt concentration

The results of titration of DCL1-A with substrate RNA followed by NMR shows the existence of an intermediate state, which is unfolded but bound to the dsRNA[12]. In our simulations at the physiological salt concentration, we do not see an apparent intermediate state along the induced folding route, which could be in agreement with the unfolded intermediate detected by the experimental measurements. Considering that the electrostatic interactions play an essential role in facilitating the initial recognition as the "steering forces" in the encounter complex, we again tuned the strengths of electrostatic interactions via varying the salt concentrations to investigate the unfolded but bound intermediate state. In Fig. 4(a), we can see the conformational distribution of DCL1-A in complex with the dsRNA. The conformational distribution is measured by RMSD, which quantifies the structural similarity to the protein DCL1-A in the bound complex. The sharp probability distribution located at the region with RMSD less than 0.2nm corresponds to the native folded state; and the wide range of conformations with RMSD from 1nm to 2.6nm relate to disordered and partially disordered states that are also sampled. It shows clearly that the most native bound states can be achieved at the physiological salt concentration 150mM. In contrast, it is harder to get native bound states at 200mM, and these even disappear at 300mM. Most notably, an intermediate state emerges at around RMSD=0.8nm at salt concentration of 100mM when we decrease the salt concentration, or in other words increase the electrostatic interactions. This becomes obvious as we keep on lowering the salt concentration to 50mM. Another intermediate state can be found next to the native structure only at salt concentration at 50mM. The strong electrostatic interactions result in a quick and tight binding before the folding step on the surface of dsRNA. The presence of considerable protein-RNA interactions in the unfolded-bound state that have to be disrupted for the protein to acquire its folded conformation, traps DCL1-A in this intermediate state, very close to the native structure with Q_i approximately at 0.8. In Fig. 4(b), the typical time trajectory as a function of Q_i from the simulation at C_{Salt}=50mM is shown. It illustrates the binding/unbinding transitions. It is worth noticing that an intermediate state (Q_i at about 0.7) emerges in addition to the unbound and native bound state. We would like to point out that this main intermediate state corresponds to that in Fig. 4(a) at around RMSD=0.8nm after investigating the structure features. In Fig. 4(c-h), the free energy surfaces are shown in the two-dimensional space of Q_i and Q_f for corresponding salt concentrations. The free energy minimum with respect to the intermediate state gradually appears as the salt concentrations decrease. Considering the nature of our structure-based model, that the native state is emphasized in the force field, it is worth noting that the theoretical analysis here was based on a range of salt concentrations higher than that used in the experimental conditions. Nevertheless, our model has succeeded in predicting the monotonic reduction of the affinity with increasing salt concentration.

The effects of electrostatic interactions on kinetics

To further investigate the role of electrostatic interactions in the binding process of DCL1-A with dsRNA complex, we performed 200 constant temperature simulations starting from different dissociative DCL1-A and dsRNA with different initial velocities in

the salt concentration range from 50mM to 300mM. Each simulation contains one single binding event, started from the unbound states and expected to reach the native bound state. Then we calculated the first passage time (FPT) for the three important binding regions (R1, R2, and R3). We defined the binding of certain region as complete when the corresponding fraction of native binding contacts exceeds 0.8 at the first time. As shown in Fig. 5(a), region 3 needs the shorter time to complete its binding, while the region 1 has the largest FPT. Interestingly, the third binding region of DCL1-A is also the region, which as mentioned above, has preformed helical structures in the C-terminal half of the protein in its free form. On the other hand, it is worth noting that the binding event completes faster as the salt concentrations decrease, that is, shielding charges from each other at higher salt concentrations results in slowing down the process. This is more obvious in R3, probably due to its abundant charged component. The phenomenon can be observed in Fig. 5(b) where we show the binding rate. To dissect the basis of this phenomenon we considered the association process as two steps: an encounter (or capture) step and a further evolution step from the encounter states to the bound complex. Thus we calculated three binding rates for describing the kinetics of the binding process by using the reciprocals of FPT_{on}, MPT_{cap}, and FPT_{evo}. All three binding rates vary in a salt concentration dependent manner. In the same way as the overall binding rate shown in Fig. 5(b), the capture rate shown in Fig. 5(c) decreases monotonously with decreasing the electrostatic interactions (as increasing the salt concentrations). We can conclude that the electrostatic interactions facilitate the initial recognition as the steering forces at the first step. The increase of salt concentration roughly accelerates the rate of evolution, probably due to the trapping in the intermediate states as a result of the stabilizing effect of the non-native electrostatic interactions at lower salt concentrations. This agrees with the thermodynamic results, which show more intermediate states with lower salt concentrations (Fig. 4(c-h)).

Next, we explored the previous frames just before the native bound state is achieved to see how the three binding regions contribute to the last part of the binding process (Fig. 6). We calculated the probability of which of the three binding regions has the largest degree of completion of binding. Surprisingly R3, which is the fastest bound region in the initial recognition, turns out to be the least complete. In other words, the intermolecular binding contacts between R3 as well as R1 (containing intensive folding contacts in their potential helical structures) of DCL1-A and the dsRNA appear to backtrack in some extent. We speculate that DCL1-A needs an adjustment upon the surface of dsRNA before optimizing the interactions, especially the intramolecular folding ones, and the orientation to the final receptor site.

Conclusions

In general, a whole protein usually consists of many segments with diverse biophysical properties, including well-ordered core, flexible loops and linkers, and intrinsic disordered binding domain. These segments together fulfill a rich spectrum of biological functions. The typical enzyme DCL1 has two tandem double stranded RNA binding domains in its C-terminus. The first intrinsically disordered domain DCL1-A is not only an independent binding unit, but also cooperates with DCL1-B, whose function is closely correlated to the intracellular localization of DCL1 as an anchor, to bind to substrate RNA. Considering the context, we propose that the IDP DCL1-A, acting as a flexible linker between DCL1-B and the rest of protein, has the ability to facilitate the recruitment of alternative partners within the nucleus, and promote the dsRNA recognition through its folding step after forming the encounter complex. The functional diversity of proteins is believed to derive from the evolutionary selection of different peptide segments for assembly.

There are many force fields available for molecular dynamics simulations of biological systems, each suited to account for different aspects of biological features. Directed by the expected range of applicability and different philosophy of thinking about the molecular picture, we divided the force fields into three categories: physics-based, structure-based, and knowledge-based. The last one depends on the statistical analysis of structural regularities derived from a growing databases of experimental structures, but it lacks transferability. The physics-based force fields, which are empirically parameterized, are capable of providing detailed description of energetics and predicting the heterogeneous ensembles of biomolecules, but are always limited by huge computational cost. So the vast conformational space searching is intractable for them. In contrast, the structure-based models with the smoother topology of energy landscape enable sufficient sampling via the tremendous advantage of speed, especially combined with the coarse-grained methods.

Recent experiments have shown that the dsRBDs have the ability to bind to any element in dsRNA molecules in a non-specific-sequence fashion, since the complexes form few contacts with bases in the minor groove[18]. This feature is very suitable for the recognition of microRNA precursors because of the essential lack of sequence conservation[13]. In the present work, DCL1-A interacts mostly with the phosphate backbone and the ribose moieties of substrate RNA, thus owing the binding specificity to the recognition of three-dimensional conformation of the RNA molecule. In this case, the structure-based model seems to be an appropriate method for our simulation.

What is the benefit of intrinsic flexibility for IDPs? It has been argued that flexibility (disorder) is kinetically favored, because it renders IDPs the adaptability of configurations before tight binding to its native partners. For a folded globular protein, binding to a well-ordered target takes place via translational and rotational diffusions to reach an intermediate encounter complex, and nearly at once forming almost all of its interactions with the partner upon their interface. As for the IDP, such a scenario corresponds to conformational selection, however, it seems unreasonable considering the following aspects: on one hand, the capability of forming most of the binding interactions at once calls for the near-native conformations of IDPs, while without the preceding aid of native partners, this is inconsistent with the basic features of unbound IDPs; on the other hand, when the extended conformations adjust to fit to all the close contacts on the surface of partners at the same time, the severe orientational restraints and the consequent large entropy lost will definitely result in an exces-

sively low rate constant[19]. Instead the binding of IDPs involves a multistep process containing intermediate encounter complexes, in which only part of the protein binds to the partner while the remaining parts still undergo conformational searching. Such sequential pathways involving binding induced folding, which differ from the typical single step all-or-none mode in the docking of canonical folded proteins, have been invoked in a number of experimental and theoretical studies of IDP-containing complexes, such as IA3-YPrA[20], Chz1-Histone[16], and pKID-KIX[21,22]. We have proposed in our previous paper that IDPs achieve the multistep kinetics via a divide-and-conquer mechanism[6], in which the forming and breaking of interactions between the IDP and its partner take place in a gradual way. IDPs with intrinsic flexibility sacrifice affinity to retrieve the adaptability with rapid association and disassociation, which is conductive to signal transduction and gene regulation. The adaptability enables IDPs promiscuous binding with fuzzy identification, which allows rearrangement of recognition motifs to mediate interactions with alternative partners. The stepwise binding process confers them high specificity, that is, IDPs can explore many different positions within the long heterogeneous precursor RNA during fast binding and unbinding, while not trapped in the nonnative binding sites thanks to the low affinity.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

L.Z. and J.W. thank support from National Science Foundation of China Grant 91430217 and Ministry of Science and Technology, China, Grants 2016YFA0203200 and 2013YQ170585. J.W. thanks the support in part from National Science Foundation Grant NSF-PHY-76066.

References

1 Ganguly, D., and J. H. Chen. 2011. Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions. Proteins-Structure Function and Bioinformatics 79:1251-1266.

2 Iakoucheva, L. M., C. J. Brown, J. D. Lawson, Z. Obradovic, and A. K. Dunker. 2002. Intrinsic disorder in cell-signaling and cancer-associated proteins. Journal of Molecular Biology 323:573-584.

3 Habchi, J., P. Tompa, S. Longhi, and V. N. Uversky. 2014. Introducing Protein Intrinsic Disorder. Chem Rev 114:6561-6588.

4 Iesmantavicius, V., J. Dogan, P. Jemth, K. Teilum, and M. Kjaergaard. 2014. Helical Propensity in an Intrinsically Disordered Protein Accelerates Ligand Binding. Angew Chem Int Edit 53:1548-1551.

5 Chu, X. K., and J. Wang. 2014. Specificity and Affinity Quantification of Flexible Recognition from Underlying Energy Landscape Topography. Plos Computational Biology 10.

6 Wang, Y., X. K. Chu, S. Longhi, P. Roche, W. Han, E. K. Wang, and J. Wang. 2013. Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. P Natl Acad Sci USA 110:E3743-E3752.

7 He, Y., and P. Lu. 2012. Manipulating Protein Conformations. ACS NANO.

8 Wang, J., Q. Lu, and H. P. Lu. 2006. Single-molecule dynamics reveals cooperative binding-folding in protein recognition. Plos Computational Biology 2:842-852.

9 Dong, Z., M. H. Han, and N. Fedoroff. 2008. The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. P Natl Acad Sci USA 105:9970-9975.

10 Xie, Z. X., K. Khanna, and S. L. Ruan. 2010. Expression of microRNAs and its regulation in plants. Semin Cell Dev Biol 21:790-797.

11 Bologna, N. G., J. L. Mateos, E. G. Bresso, and J. F. Palatnik. 2009. A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. Embo J 28:3646-3656.

12 Suarez, I. P., P. Burdisso, M. P. M. H. Benoit, J. Boisbouvier, and R. M. Rasia. 2015. Induced folding in RNA recognition by Arabidopsis thaliana DCL1. Nucleic Acids Research 43:6607-6619.

13 Drusin, S. I., I. P. Suarez, D. F. Gauto, R. M. Rasia, and D. M. Moreno. 2016. dsRNA-protein interactions studied by molecular dynamics techniques. Unravelling dsRNA recognition by DCL1. Arch Biochem Biophys 596:118-125.

14 Huang, Y. Q., and Z. R. Liu. 2009. Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding-Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism. Journal of Molecular Biology 393:1143-1159.

15 Chu, X., F. Liu, B. A. Maxwell, Y. Wang, Z. Suo, H. Wang, W. Han, and J. Wang. 2014. Dynamic conformational change regulates the protein-DNA recognition: an investigation on binding of a Y-family polymerase to its target DNA. PLoS Comput Biol 10:e1003804.

16 Chu, X. K., Y. Wang, L. F. Gan, Y. W. Bai, W. Han, E. K. Wang, and J. Wang. 2012. Importance of Electrostatic Interactions in the Association of Intrinsically Disordered Histone Chaperone Chz1 and Histone H2A.Z-H2B. Plos Computational Biology 8.

17 Chu, W. T., J. Clarke, S. L. Shammas, and J. Wang. 2017. Role of non-native electrostatic interactions in the coupled folding and binding of PUMA with Mcl-1. Plos Computational Biology 13.

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

19 Zhou, H. X., X. D. Pang, and C. Lu. 2012. Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. Physical Chemistry Chemical Physics 14:10466-10476.

20 Wang, J., Y. Wang, and X. Chu. 2011. Multi-Scaled Explorations of Binding-Induced Folding. PLoS Computational Biology.

21 Turjanski, A. G., J. S. Gutkind, R. B. Best, and G. Hummer. 2008. Binding-induced folding of a natively unstructured tran-

scription factor. Plos Computational Biology 4.

22 Sugase, K., H. J. Dyson, and P. E. Wright. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447:1021-U1011.

23 Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. Journal of Molecular Biology 298:937-953.

24 Whitford, P. C., J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu, and J. N. Onuchic. 2009. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. Proteins 75:430-441.

25 Miyazawa, S., and R. L. Jernigan. 1996. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. Journal of Molecular Biology 256:623-644.

26 Cho, S. S., Y. Levy, and P. G. Wolynes. 2009. Quantitative criteria for native energetic heterogeneity influences in the prediction of protein folding kinetics. Proc Natl Acad Sci U S A 106:434-439.

27 Azia, A., and Y. Levy. 2009. Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. J Mol Biol 393:527-542.

28 Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. (1999). Automated analysis of interatomic contacts in proteins. Bioinformatics, 15(4), 327-332.

29 Tan, C., Li, W., and Wang, W. (2013). Localized frustration and binding-induced conformational change in recognition of 5s rna by tfiiia zinc finger. Journal of Physical Chemistry B, 117(50), 15917-25.

30 Jr, Ronald D. Hills,, and C. L. B. Iii. "Insights from Coarse-Grained G$\bar{o}$ Models for Protein Folding and Dynamics." International Journal of Molecular Sciences 10.3(2009):889.

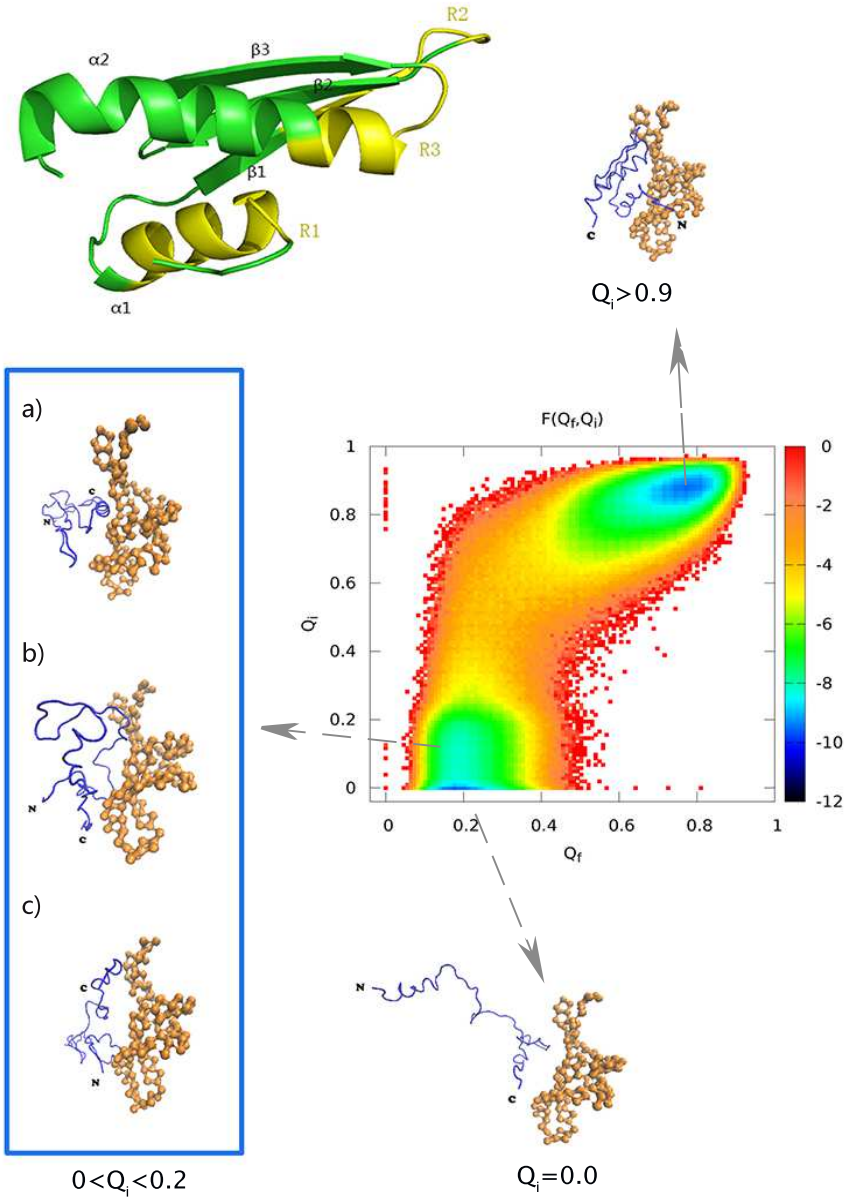
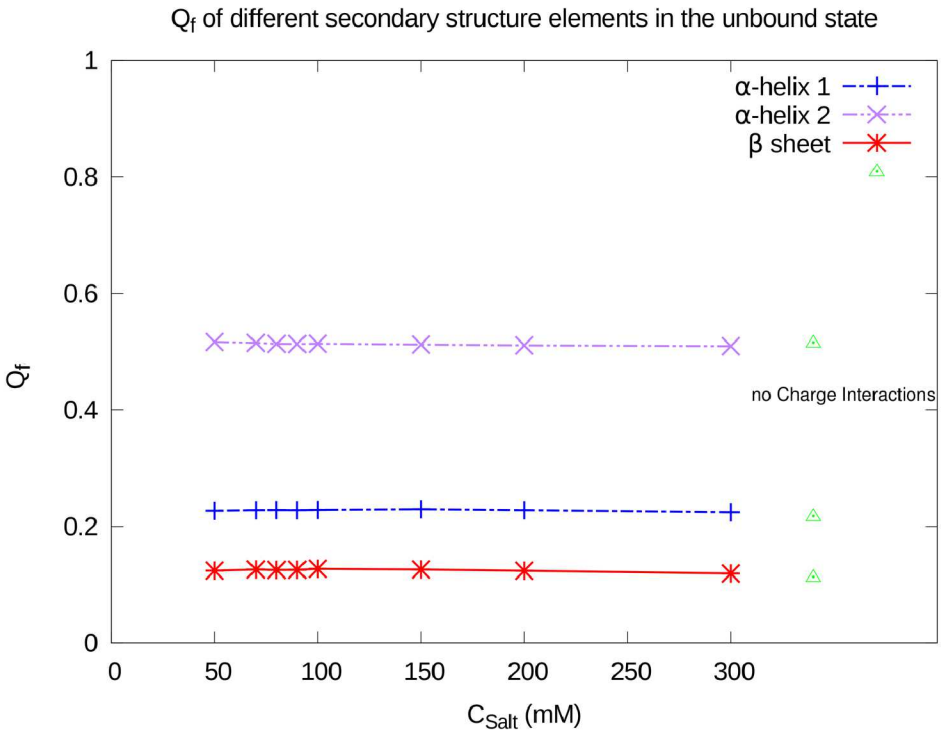
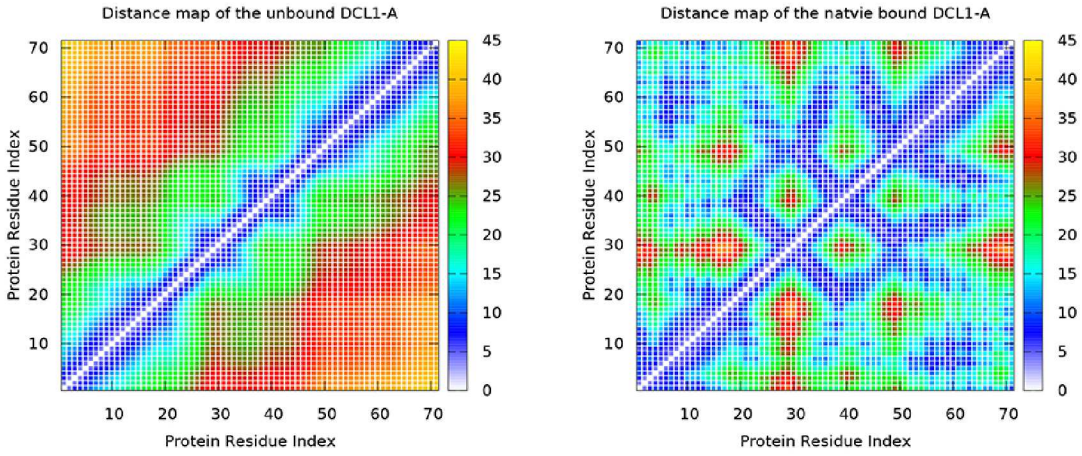


Fig. 1 Structural representation of DCL1-A and free energy landscape. On the *Top* is shown the structure of DCL1-A domain. It folds with a topology α-β-β-β-α upon binding to the dsRNA (not shown here). The three binding regions are colored in yellow. The picture on the *Right* corresponds to the free energy surface as a function of Q_i (binding reaction coordinate) and Q_f (folding reaction coordinate), which supports a coupled folding and binding process via an induced folding mechanism. The system favors the route along the upper left corner (binding precedes folding), which passes through the conformational region with high Q_i value and low Q_f value (bound-unfolded state) to the region with both high values of Q_i and Q_f (bound-folded state, or the native bound state). The structural representations of the DCL1-A and dsRNA complex throughout the binding process are shown around the free energy landscape. Specifically, the typical encounter complexes are highlighted in the blue box, in which the binding region R3 [R2,R1] makes the first contact with the dsRNA in a) [b), c)].



(a)



(b) (c)

Fig. 2 The detailed folded content and distance maps. (a) Q_f of different secondary structure elements in the unbound state with various salt concentrations. Q_f of α-helix 1, α-helix 2, and β sheet are plotted with points colored by blue, purple, and red, respectively. In particular, the simulation in the absence of charge-charge interactions is also performed and the data are plotted as a benchmark. (b) The inter-residue distance map of the unbound,and (c) the native bound DCL1-A. The value of distance (C_α-C_α atoms) increases from 0 Å (white) to 45 Å (yellow). The distance map of unbound state reveals a lack of overall well-defined structural features, while that of the bound state displays a well-structured protein with the topology α-β-β-β-α.

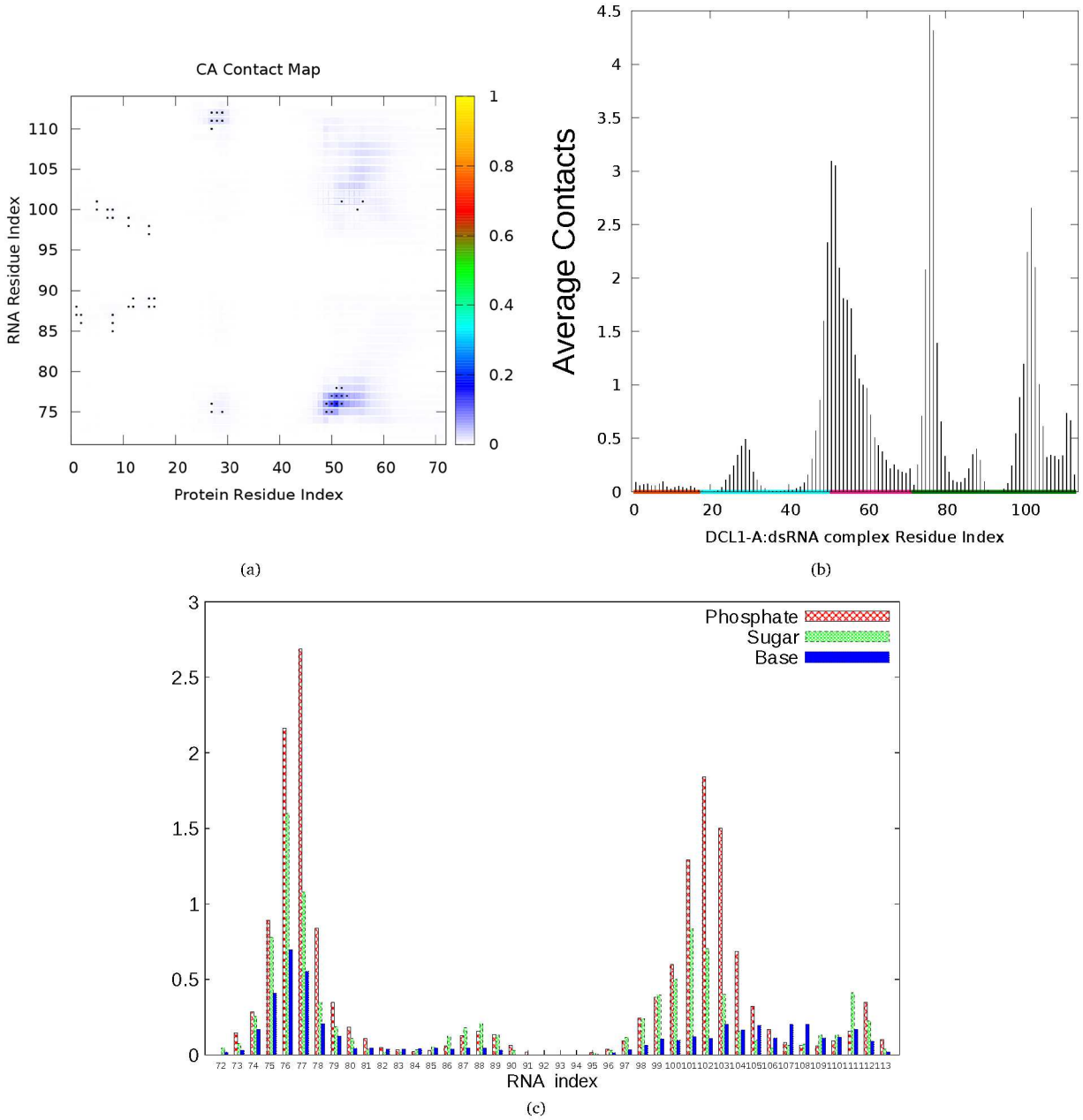


Fig. 3 Contact information in the encounter complex. (a) The C_α-C_α contact map between DCL1-A and dsRNA. The colored squares with gradational changes indicate the probability of the contact formed in the encounter state. The black points represent the contacts existed in the native structure. (b) The average contact number of each residue or nucleotide formed in the encounter complex. The colors on the x-axis represent different secondary structure regions: red, the α-helix 1; cyan, the β sheets; magenta, the α-helix 2, and dark-green, the RNA segment. (c) The exact contacts for phosphate, sugar, and base in the dsRNA. The calculation of contacts is based on the cut-off algorithm.

Table 1 Protein-RNA interaction energy in different binding stages.In the structure-based model, the non-native LJ potential is represented by the excluded volume repulsive term in the energy function. Energy is in the unit of ε.See more details in "Methods".

Stage	**Unbound**		**Encounter**		**Bound**	
Energy part	Native	Non-native	Native	Non-native	Native	Non-native
E_{LJ}	-0.01±0.28	0.52±0.81	-3.54±3.03	0.62±0.87	-42.59±4.21	0.13±0.24
E_{elec}	-0.08±0.07	-7.24±3.74	-1.65±1.11	-8.67±2.78	-4.23±0.46	-6.08±0.63

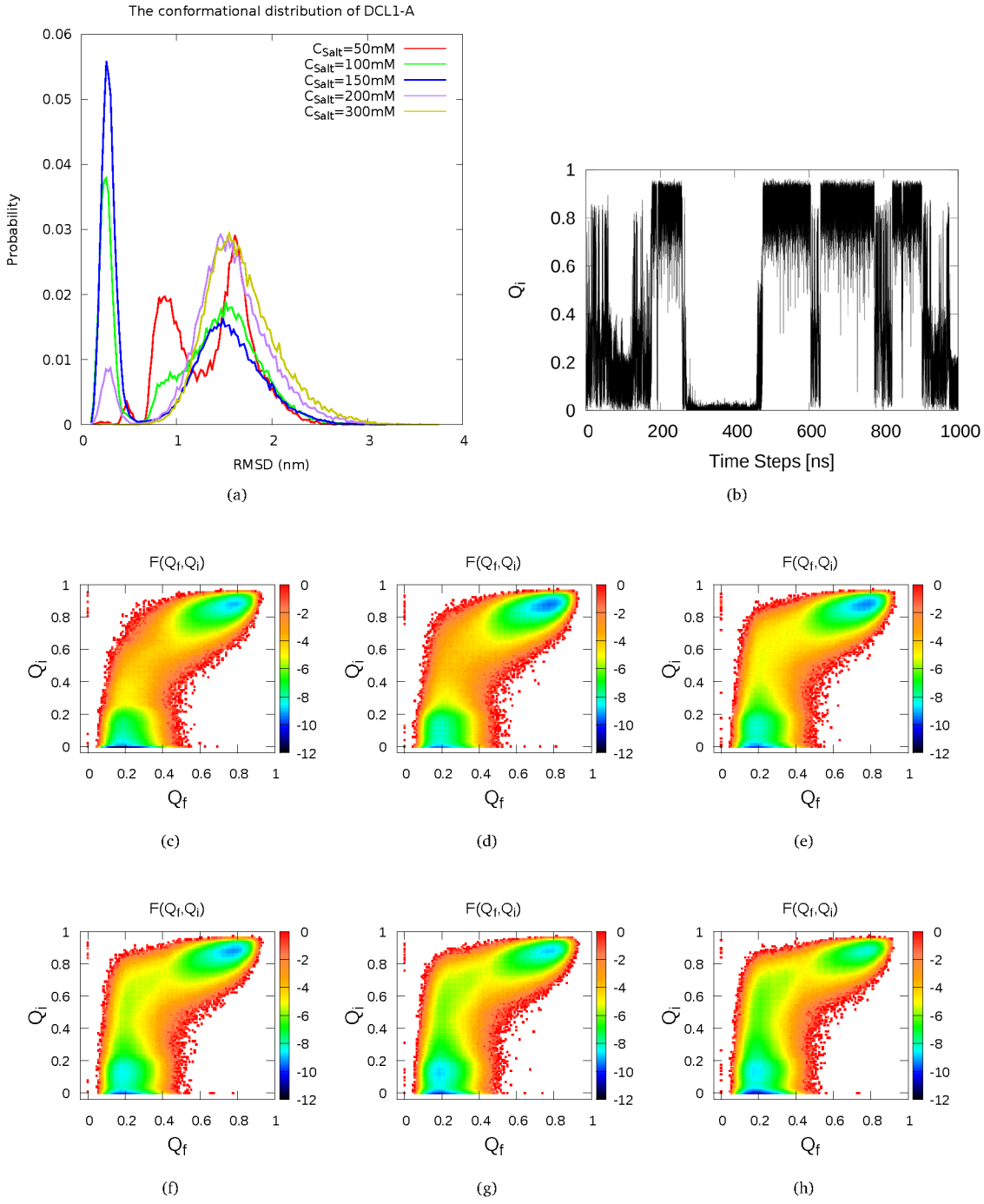


Fig. 4 The shift of conformational distributions at different salt concentrations. (a) The conformational distribution of DCL1-A when binding to dsRNA at the simulation temperature Ts=160.90 K. The conformational distribution is measured by RMSD, which quantifies the structural similarity to the protein DCL1-A in the bound complex. The sharp probability distribution located at the region with RMSD less than 0.2nm corresponds to the native folded state; and a wide range of conformations with RMSD from 1nm to 2.6nm relating to disordered and partially disordered stats are also sampled. It shows clearly that the most native bound states can be achieved at the physiological salt concentration 150mM. As a contrast, it turns to be harder to get the native bound states at 200mM, and even disappeared at 300mM. Most notably, an intermediate state occurs when lowering the salt concentration, that is, to increase the electrostatic interactions. (b) Typical time trajectory of binding/unbinding transitions at C_{Salt}=50mM. Markedly, the intermediate state (Q_i about 0.7) emerges in addition to the unbound and native bound state. (c-h) the free energy surfaces are shown in the two-dimensional space of Q_i and Q_f in the salt concentration range of 200mM, 150mM, 90mM, 80mM, 70mM, and 50mM. The free energy minimum with respect to the intermediate state gradually appears as the salt concentrations decrease.

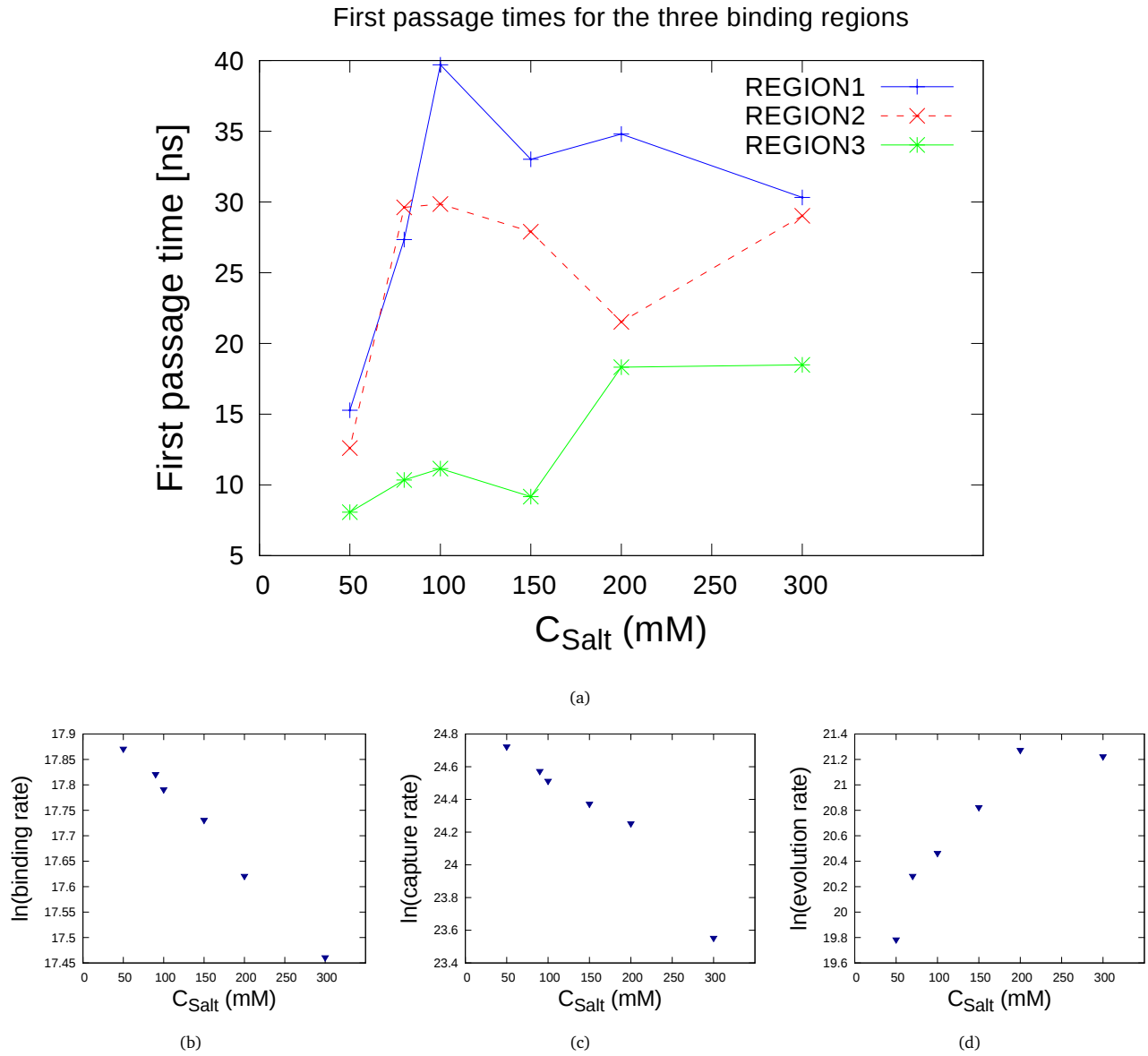


Fig. 5 Role of electrostatic interactions in the binding process. (a) First passage time (FPT) for the three binding regions at different salt concentrations. REGION3 needs the fewest time to complete binding, REGION2 slower, and REGION1 binds the last. It is worth noting that the binding event completes faster as the salt concentrations decrease, that is to say, shielding more charges from each other at higher salt concentrations will result in slowing down. And this is more evident in REGION3. (b) Binding rate, (c) capture rate, and (d) evolution rate for DCL1-A interacting with dsRNA in the salt concentration range from 50mM to 300mM. They were calculated by using the reciprocals of FPT_{on}, MPT_{cap}, and FPT_{evo}. All the units of rates are s^{-1}. Broadly speaking they are all modulated by the salt concentration.The overall binding rate as well as the capture rate, which monitors the rate of formation of encounter complex, decrease monotonously with increasing the salt concentrations. And the increasing of salt concentrations roughly accelerates the rate of evolution from the encounter states to the native bound complex.

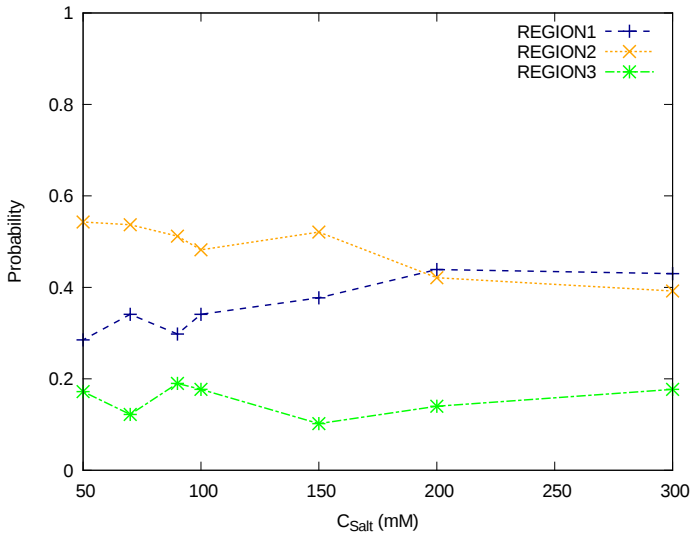


Fig. 6 The probability of the pre-bound regions at different salt concentrations. Probability of which in the three binding regions has the largest degree of binding completion before the final binding is shown. Surprisingly, REGION3, which is the fastest bound region in the initial recognition, turns out to be least. However, the most flexible REGION2 acquires the largest probability in the previous frame just before the native bound state.

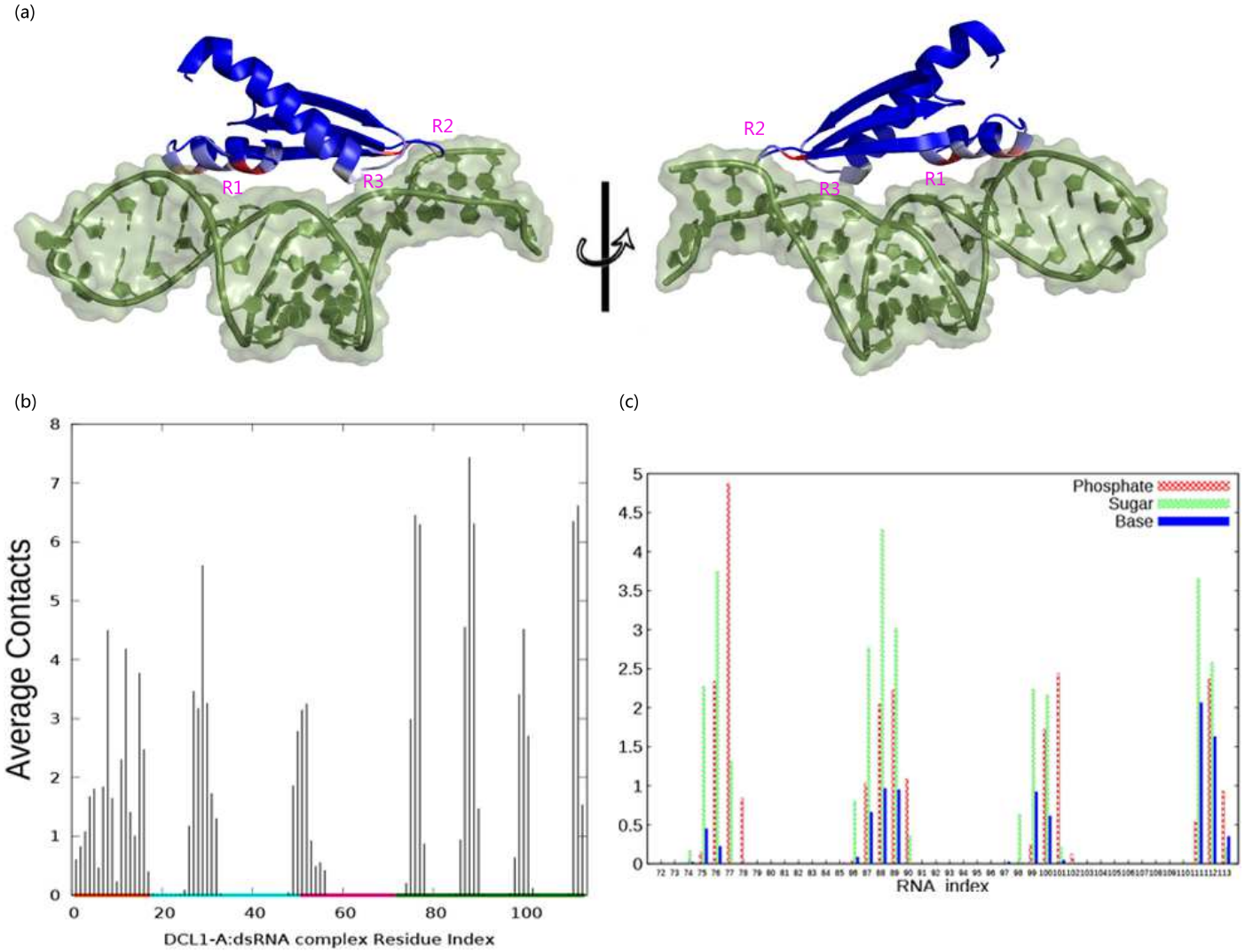


Fig. 7 The Average contacts between DCL1-A and dsRNA in the native bound state. (a) The native structure of DCL1-A:dsRNA complex. The three important binding regions are highlighted from blue to red, corresponding to the intermolecular contact number from 0 to 8, while the dsRNA is colored greyish-green. (b) The average of contacts of each residue or nucleotide in the DCL1-A:dsRNA complex are shown. The colors on the x-axis represent different secondary structure regions: red, the α-helix 1; cyan, the β sheets; magenta, the α-helix 2, and dark-green, the RNA segment. (c) The exact contacts for phosphate, sugar, and base in the dsRNA. The calculation of contacts is based on the cut-off algorithm.