Insight into the LGP2 Helicase Domain – An *in Silico* Approach

Subhomoi Borkotoky, Chetan Kumar Meena, Aswathy Soman, Ayaluru Murali*

Bioinformatics Centre, School of Life Sciences, Pondicherry University, Puducherry, India

Copyright © 2015 Horizon Research Publishing All rights reserved.

Abstract LGP2, a member of retinoic acid inducible gene-I like receptors (RLR), encoded by the gene DHX58 in human induces antiviral response against many RNA viruses. The LGP2 shares a considerable similarity to the amino acid sequence of Hef Helicase Domain, the helicase domain of RIG-I-like protein helicase-associated endonuclease (Hef), which is involved in RNA binding. Earlier studies suggest that RLR contains a conserved C-terminal domain (CTD), which is responsible for the binding specificity to the viral RNAs and C-terminal domain of LGP2 also takes part in RNA binding. The present study is aimed to explore the interactions of LGP2 and RNA, thereby finding the crucial residues for the interaction, with the help of *in-silico* tools. The predicted crucial residues were validated by docking and molecular dynamics simulation studies as well as by fitting the model on a LGP2 density map. Our results in this study suggest that the residues in the helicase domain of LGP2 are crucial for RNA binding and it is positioned around the groove region of LGP2. Though earlier experimental studies identified the RNA binding residues, but our *in silico* binding studies with the full length LGP2 predicted some additional residues that can be valuable

Keywords LGP2, Homology Modeling, Molecular Docking, Molecular Dynamics Simulation, Flexible Fitting

1. Introduction

The first line of defense, innate immune response is dependent on pattern recognition receptors; one class of pattern recognition receptors is toll like receptors, found on the cell surface or endosomes in effector cells, that recognize pathogen associated molecular pattern typically located on the surface of pathogen cells [1], whereas RIG-I like receptors (RLRs) are in the cytoplasm. The RLR family, within superfamily 2 (SF2) RNA helicases, include RIG-I (retinoic acid-inducible gene-I), MDA5 (melanoma-differentiation-associated gene 5), and LGP2 (Laboratory of Genetics and Physiology 2) and are closely related to DEAD-box helicases that can sense RNAs to initiate antiviral responses [2, 3]. In human, LGP2 is encoded by DHX58 gene [4].

Though LGP2 is quite similar to other RLR family members RIG-I and MDA-5, it lacks the signature N-terminal caspase activation and recruitment domain (CARD) (Fig. 1) that is required for signaling [5]. Different studies defined LGP2 as a positive as well as negative regulator of RIG-I and MDA5-mediated viral recognition [6, 7]. Further, it was demonstrated that LGP2 protein produced in insect cells can bind both single- and double-stranded RNA (ss- and ds-RNA), with higher affinity for dsRNA and significant change in the conformation of the LGP2 monomer to a stable dimer takes place when it was complexed to a dsRNA [8]. The RLR contains a conserved C-terminal domain (CTD), which is responsible for the binding specificity to the viral RNAs. The solution structure of LGP2 (PDB ID: 2RQA) C-terminal domain (CTD) (546 – 678) has shown that it binds to dsRNA and 5'ppp-ssRNA with higher affinity than CTD of RIG-I [9]. A similar view was also reported in the 2.0-Å resolution crystal structure of human LGP2 CTD and 8-bp dsRNA complex (PDB ID: 3EQT), which was also supported by gel filtration chromatography and analytical ultracentrifugation studies [10].

The helicase domain of RIG-I-like protein, helicase-associated endonuclease (Hef), consists of three structural sub-domains that share common conserved sequences like other SF2 helicase members, e.g. RIG-I and DEAD-box. The Hef helicase domain (Hhd, residues 1–494) contains seven helicase motifs that were spread in sub-domains 1 and 3 [11, 12]. The crystal structure of Hhd resembles a partially opened clamp with a concave groove associated with its lower portion [11]. Murali *et al* [8], using electron microscopy and single particle analysis reported the low resolution structure of LGP2. By superimposing the crystal structure of Hef helicase domain into the electron density map of LGP2, it was observed that LGP2 shares a considerable structural similarity with the three domains of Hhd [7]. Further LGP2 also shares a considerable sequence similarity with Hhd (Figs. 1 and 3.i). It was also reported that the RNA binding motifs that were observed in Hhd [11] were also conserved in LGP2 (Fig. 3.i) and that they aligned in the groove part of the LGP2 structure when the crystal structure of Hhd is superimposed on the LGP2 density map [8].

Figure 1. Graphical representation of domains present in MDA5, RIG I, LGP2 and Hhd. Blue color represent CARD domain and green color shows Helicase–ATP binding domain. While orange color shows helicase C-terminal domain. MDA5, RIG I contains all three domains while LGP2 and Hhd are missing the CARD domains.

Though the structure of LGP2 CTD complexed with RNA was extensively studied by X-Ray diffraction and NMR techniques [9, 10], the interaction of RNA with the full length LGP2 (C- and N-termini) will be more biologically relevant. In this study, we made an attempt to understand the RNA interactions with full length LGP2 with the help of several *in silico* tools.

2. Materials and Methods

2.1. Protein Structure Prediction

In the absence of a full length, experimentally determined high resolution structure of LGP2, homology modeling approach was used to predict the full length structure using MODELLER 9.10 [13]. Amino acid sequence of LGP2 was retrieved from UniProtKB [14] database (Q96C10) for three-dimensional structure prediction. PDB database was searched for homology by using BLAST tool. From homology search, Helicase associated endonuclease (HEF) helicase domain was retrieved as best suitable template for homology modeling with 42% sequence identity.

2.2. Cryo-EM Flexible Fitting

The LGP2 density map was already solved by electron microscopy and was used in this study [8]. Since this density map has a low resolution to determine the secondary structural elements, we have used modeling based on flexible fitting approach to predict its structure. The structure of LGP2 that resulted in homology approach was fitted to the density map, with the Flex-EM module [15] of MODELLER.

2.3. Active Site Prediction

In the present work, CASTp (Computed Atlas of Surface Topography of Proteins) [16] is used for identification of active site regions. The *in silico* model of LGP2 was submitted to CASTp server for active site identification. It

provides identification and measurements of surface accessible pockets as well as interior accessible cavities.

2.4. Docking Studies

As we have already found helicase domains of LGP2 as active site region through sequence alignment, docking was performed for the confirmation of those active site residues by using RNA strand. Li *et al* [10] reported a crystal structure of C-terminal domain of LGP2 complexed with RNA. This RNA was used as a reference ligand for the identification of active site residues. Docking was performed with the help of FFT-based protein docking server HEX [17].

2.5. Molecular Dynamics Simulations (MDS)

The stability of the three-dimensional structure of LGP2 along with the interaction strength was checked by using Gromacs 4.5.3 simulation package [18] with the help of Amber force field [19]. A maximum of 50,000 steps of energy minimization was carried out for the system with a tolerance of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ using conjugant gradient algorithm followed by steepest descent minimization with the same number of minimization steps. PME method was used to apply a twin-range cutoff (1.0 nm for Vander Waals and electrostatic interactions) for long interactions [20]. The minimized and solvated systems upon satisfying the geometry and solvent orientations were further used in the simulation steps. The LINCS [21] algorithm was used to constrain all bond angles, while the SETTLE [22] algorithm used to constrain the geometry of water molecule. Equilibration of the entire system was carried out for 50 ps each for position restraints of both NVT (constant number of particles, volume and temperature) and NPT (constant number of particles, pressure and temperature). Equilibration of the entire ensemble confirmed the uniformity of temperature, pressure, density and total energy of the system. The pre-equilibrated system was subsequently used in the production MDS of 15000 ps (15 ns) with a time-step of 2 fs. For every two ps structural coordinates were saved and analyzed using the analytical tools available in the Gromacs

package. Conformation with the lowest potential energy was selected from the 15 ns MDS trajectory, which was further refined by energy minimization. The refined models were validated using the structural analysis and verification server (SAVES) (<http://nihserver.mbi.ucla.edu/SAVES/>), which uses various tools, including PROCHECK [23], ERRAT [24], and VERIFY_3D [25].

3. Results and Discussion

3.1. Sequence Alignment

For structure prediction, Hef helicase domain (PDB ID: 1WP9) of *Pyrococcus furiosus* was used as template because it shares a considerable structural similarity with LGP2 (Fig. 1). To gain an insight into the residual arrangement of protein, we performed sequence alignment by ClustalW [26] (shown in Fig. 3.i). From Fig. 3.i, it can be seen that both the sequences are sharing several conserved regions, including the helicase motifs reported for Hhd [11] (marked in black boxes in Fig. 3.i).

Figure 2. Secondary structural component of LGP2's three-dimensional structure. LGP2 contains 6 sheets, 4 beta alpha beta units, 4 beta hairpins, 2 beta bulges, 20 strands, 30 helices, 27 helix-helix interactions, 50 beta turns, 5 gamma turns, and 1 disulphide bond for a proper folding.

3.2. Three Dimensional Structure of LGP2

For the better understanding of secondary structural components for active site formation in three-dimensional space, three dimensional structure predictions are necessary. Murali *et al* [8] reported the low-resolution LGP2 structure in its monomeric and dimeric forms. However, in order to explore the active site, it is required to have a better resolution. The three-dimensional model of LGP2 was modeled using homology modeling with the help of CryoEM module of modeller 9.10. The model having the lowest DOPE score (-68165) and molpdf value (4987.37) was chosen for further analysis. Analysis of the secondary structure elements of model was done with the help of PDBsum [27]. The PDBsum results suggested that LGP2 three-dimensional structure contains 6 β sheets, 4 $\beta\alpha\beta$ units, 4 β hairpins, 2 β bulges, 20 β strands, 30 α helices, 26 helix-helix interactions, 50 β turns, and 5 γ turns (Fig. 2). Information of the helix-helix interactions are given in the table 1. There are 26 helix-helix interactions present in the three-dimensional structure of LGP2 protein. Out of them, interaction between 2-19, 3-18, 8-21, 9-21, 13-17, 14-17,

16-17, 16-20 and 19-20 are important for the proper folding of this protein. There are total 50 β turns present in the three-dimensional structure of LGP2 and out of them 6 β turns (present between helix 9-10, helix 19-20, helix 10-11, helix 15-16, helix 17-18, and helix 11-12) are playing important role in the proper folding of this protein along with them one γ turn that is present between helix 15-16 is also important. β hairpin turn present between residues 570-575, 610-617 and 620-630 are also important for the formation of C-terminal domain, which is the RNA binding site of this protein. Total 3 β sheets are present in this protein and these sheets contain 20 β strands. These secondary structure elements of the protein are necessary for the better understanding of spatial arrangement of LGP2 structure as interactions between these elements defines the folds, which eventually forecasts the biological relevance.

The 3D model was then subjected to validation. The Ramachandran plot analysis shows that the model is of good quality as 92.5% of the residues fall into the most favoured regions. The model was also submitted to ProSA Server [28], which shows Z-Score: -9.72 for overall model quality.

Table 1. Helix-helix interaction present in the helices of LGP2 protein. C: beyond the C terminus of the helix, N: beyond the N terminus of the helix, I: internal to the helix.

No.	Helices		Distance (Å)	Angle (ω°)	Interaction type		No. interacting residues		
							Total	Helix 1	Helix 2
1.	1	2	9.4	-43.0	I	I	11	6	6
2.	1	10	8.1	-131.3	N	I	3	2	2
3.	2	4	9.8	-40.3	I	I	7	4	3
4.	2	19	13.9	-98.4	N	C	1	1	1
5.	3	4	3.5	-13.8	C	N	7	4	3
6.	3	18	9.3	-58.5	N	I	4	2	3
7.	5	6	10.3	-107.8	N	I	2	1	2
8.	6	8	7.7	44.3	N	N	10	5	5
9.	8	9	7.9	66.4	I	C	4	3	2
10.	8	21	10.0	-110.2	I	I	6	5	4
11.	9	20	8.3	-58.5	I	I	7	5	4
12.	9	21	8.3	-74.3	I	I	14	6	5
13.	11	12	15.4	26.7	I	I	2	2	1
14.	11	13	10.7	-171.0	I	I	12	6	6
15.	11	15	10.6	171.1	N	C	11	6	6
16.	12	13	7.9	-159.9	I	N	16	8	8
17.	12	20	10.8	84.8	N	N	1	1	1
18.	13	14	6.8	155.9	C	N	12	7	5
19.	13	15	10.6	-11.7	N	N	7	4	4
20.	13	16	5.8	-27.2	C	N	3	2	2
21.	13	17	9.6	-35.8	C	I	4	2	3
22.	14	15	9.9	-154.8	I	I	11	6	8
23.	14	17	10.0	135.1	N	I	2	1	2
24.	16	17	13.1	-61.6	I	I	5	4	3
25.	16	20	15.2	-65.7	N	N	1	1	1
26.	19	20	8.4	-95.4	N	I	8	4	5

3.3. Fitting of Electron Density Map

The flexible fitting was done by optimizing the model simultaneously with its position and orientation in the density map of LGP2 [8] using Flex-EM, which integrates rigid and flexible fitting. First, rigid fitting of the model structure into the density was performed using Mod-EM guided by a cross-correlation coefficient (CCF) between the atomic structure and the density map. Next, the best fit (with CCF ~ 0.8296) was refined by using Flex-EM. The rigid residues were found with help of RIBFIND server [29]. All residues, except the rigid bodies found from RIBFIND server, were allowed to move flexibly by using a molecular dynamics optimization. The final model fitted into the electron density map using the fit-in-map utility of UCSF Chimera is shown in Fig. 3.ii. The flexible fitting of the model structure into the LGP2 density map and the position of the helicase domain in the density confirms the earlier observations [8] with a marginal difference. The earlier study have shown that all Hhd helicase motifs fall into the groove part of the LGP2 density map, while in our study we

found that the conserved helicase domains of LGP2 are spread along the curvature and groove of the density map (Fig. 3.ii). Some of residues of the conserved helicase motifs of LGP2 moved away from the groove. However in docking studies we have seen that these residues are making interaction with the RNA (see docking results in next section).

3.4. Binding Site Prediction

After confirming the position of helicase motifs in LGP2 density map Fig. 3.iii.a, we used CASTp to identify the active site residues. In CASTp analysis, we have chosen the pocket having the largest area and the volume. The residues, which were falling in the active site pocket, are given in table 2 and we can see the position of those residues in Fig. 3.iii.b in three-dimensional space. The common residues (132-135, 374-377, 439-444) from both the conserved helicase motifs and CASTp calculations were selected as active site residues for further work and they are shown in Fig. 3.iii.c.

Table 2. Residues found in sequence alignment, CASTp and common residues found in both study.

Conserved residues found in sequence alignment	Active site residues calculated by CASTp	Common residues in both sequence alignment and CASTp calculation
25, 26, 27, 28, 29, 30, 31, 129, 130, 131, 132, 133, 134, 165, 166, 167, 168, 169, 170, 171, 371, 372, 373, 374, 375, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 467, 468, 469, 470, 471	4, 6, 23, 24, 25, 26, 27, 28, 29, 30, 31, 34, 52, 58, 59, 62, 63, 65, 66, 67, 129, 132, 134, 135, 166, 168, 169, 170, 197, 199, 200, 201, 207, 208, 210, 211, 215, 217, 351, 354, 357, 358, 360, 361, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 383, 384, 385, 386, 387, 393, 398, 400, 403, 423, 424, 427, 428, 429, 430, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 500, 503, 504, 507, 508, 511, 512	25-31, 129, 132, 134, 166, 168-170, 371-375, 438-447, 467-471.

Figure 3. i) Alignment of amino acid sequences of Hhd (PDB ID: 1WPG) and LGP2. Helicase motifs of both proteins are boxed in black. ii) Two different views of the predicted three-dimensional model of LGP2 are superimposed in the electron density map of LGP2. Here the green color residues show the position of helicase motifs of LGP2. iii) Predicted three-dimensional structure of LGP2 protein. a) The position of helicase domain found through sequence alignment between Hhd and LGP2 (shown in green). b) The active site residues of LGP2 calculated by using CASTp server (shown in blue). c) The common residues seen in both (a) and (b) (shown in yellow).

3.5. Docking

Table 3. LGP2-RNA docking results from Hex server.

Receptor	Ligand	cluster	Docked energy (kcal/mol)
LGP2	RNA	1	-7.59
		2	-6.42
		3	-6.16
		4	-6.03
		5	-5.92
		6	-5.84
		7	-5.77
		8	-5.72
		9	-5.66
		10	-5.61
		11	-5.56
		12	-5.52
		13	-5.48
		14	-5.44
		15	-5.41

Docking studies were performed for finding the hotspot residues of LGP2 to confirm whether the binding residues correlate with our active site residues identified by CASTp and sequence alignment. Here protein-RNA docking was performed with help of HEX protein docking server without providing the active site residues. HEX generated 15 clusters of docked protein and RNA. Out of those 15-clusters, best cluster with the maximum dock score was selected for further analysis (Table 3). This complex was analyzed with the help of DIMPLOT module of LIGPLOT [30] software for visualization of interactions (with a cutoff value of 3.5 Å chosen for hydrogen bonds). The atoms, which are forming Hydrogen bonds and hydrophobic interactions, are given in table 3 along with their residue numbers. Interestingly, those residues, which are common in both CASTp calculations and sequence alignment are also making bond with RNA (as identified by DIMPLOT) (Fig. 4.i).

The residues forming H-bonds in our *in silico* studies are also common in the other experimental studies such as NMR titration experiments (T 575, H 576, K 634, R 636, K 651 and W 652) as well as X-ray crystallography (H 576, K 651 and W 652). Apart from these, we have seen some additional residues (377 and 635) making different interactions with LGP2 (table 4). These results suggest that the RNA binding cavity is not confined to the C-terminal.

Figure 4. i) The docked complex of LGP2 and dsRNA. The helicase motifs of LGP2 are shown in blue and black color. Here the blue color shows the helix which makes H-bond with the RNA molecule (see the text). The close-up view of all the residues which are interacting with RNA is shown in the inset. There are 8 hydrogen bonds present between LGP2 and RNA (see text), but only the H-bond between helicase domain and RNA is shown. ii) The profile of hydrogen bonds throughout the simulation. Most of the time the number of H-bonds remained between 15 and 20. iii) The map of hydrogen bond existence is plotted between hydrogen bond index and simulation time. The amino acids involved in the H-bonding are shown on the right side. The H-bonds that stay consistently throughout the simulation period are seen as a continuous horizontal line. About 13 residues involved in H-bonding were stable for 80 % of the simulation period or greater.

Table 4. Interacting residues of LGP2 and RNA

S. No.	Hydrogen Bonds			Hydrophobic Interactions	Special Interactions	
	H-Bond Donor	H-Bond Acceptor	Distance (Å)		Pi-Pi Interactions	Pi-Cation Interactions
LGP2-RNA	Arg 377 (NH2)*	Gua 11 (O5)	3.21	Thr 575- Cyt 10	His414-Gua1	Gua 7-Arg 36 Cyt 10-Arg 377
	Thr 575 (OG1) [@]	Cyt 16 (OP2)	2.33	Thr 376- Cyt 12		
	Arg 636 (NE) [@]	Cyt 12(O2)	3.20	Gln 378- Gua 5		
	Arg 636 (NH1) [@]	Cyt 6 (O2')	2.67	Arg 286- Cyt 8		
	Lys 634 (2HZ) [@]	Gua 5 (N3)	2.24	Leu 282-Gua 13		
	Lys 651 (NZ) ^{@,§}	Cyt 14 (OP2)	2.53	Val 635- Gua 7		
	Trp 652 (N) ^{@,§}	Cyt 14 (OP1)	2.95	His 381- Gua 15		
	His 576 (NE2) [§]	Gua 15 (O2')	3.24	Lys 650- Gua 1		
	Val 635 (N)	Gua 13 (O2')	3.17	Gly 407- Cyt 14		
				Val 632- Cyt 14		

* Residues identified in both conserved helicase motifs and CASTp.

[@] Residues reported in NMR Titration experiment [8].

[§] Residues reported by X-ray Crystallography of CTD of LGP2 [9].

Figure 5. RMSD and RMSF profiles of LGP2 throughout the simulation period. i) Shows the RMSD profile. LGP2 got local minima at points I and II 9ns but at point III (after 10 ns) stability was achieved and remained same for rest of the simulation period. ii) RMSF profile of LGP2. The C-terminal part of LGP2 shows more flexibility compared to rest of the structure. This is the region where RNA binds. The flexibility of this region may help in the complex association and dissociation.

3.6. Molecular Dynamics Simulation

A 15-nano second molecular dynamics simulation study was carried out to check the stability of RNA-protein interactions.

The stability of the LGP2-RNA complex was checked through RMSD profile (shown in Fig. 5.i). This graph was generated between the conformation of protein LGP2 before and after 15 ns simulation. The RMSD profile of any structure is generated based on the scalar distance between same atoms of two different conformations. Here the RMSD profile was generated for only the backbone residues. As we can see in the Fig. 5.i, the LGP2 was showing fluctuation up to 3 ns, after that protein got minimum deviation for some extent. Again, at 7 ns, it returned to almost same minimum deviation as seen at 3ns. After a simulation for some more time it got stability at 9 ns and it was stabilized at that conformation for rest of simulation period. Protein was deviating up to 3 Å for getting stability in structure from modeled structure. The fluctuation range is acceptable for modeled structure. After getting the RMSD profile and

reaching stable conformation, identification of most fluctuating region of LGP2 was done by generating its RMSF profile (shown in Fig. 5.ii). The most fluctuating part of LGP2 consists in C-terminal region (between residue numbers 500-555); this is the region, which is adjacent to RNA binding residues. The RMSF profile of LGP2 is generated only for α carbon of backbone.

After identification of these fluctuating moieties, the interaction strength of protein and RNA was analyzed by generating graph of hydrogen bonds (Fig.4.ii). As we can see, the numbers of H-bonds were ranging between 15-25 (occasionally going to a minimum of 12) and it was remaining at ~20 for most of the simulation period. This shows that interaction between protein and RNA is strong enough for its functioning in immune system.

An H-bond existing graph was generated with the cutoff range of 3.5Å for H-bonds. Here the H-bonds were calculated from each MDS trajectory in 2fs interval. Multiple occurrences of H-bonds between same base-residue pair were counted as one. The total numbers of H-bonds made between LGP2 and RNA were around 186 throughout

simulation. Out of 186 H-bonds, around 13 hydrogen bonds were showing consistency (Fig. 4.iii). Out of these 13 consistent H-bonds, 4 bonds were matching with LigPlot analysis. The details of interactive residues along with distances are given in table 3.

The results were further analyzed by molecular dynamics simulations. The full length model with RNA, LGP2 C terminal-RNA model and NMR structure of LGP2 C-terminal (PDB ID: 2RQA) were simulated for 15 ns using Gromacs to compare our modeled structure with the experimental structure. The NMR structure 2RQA was chosen for the comparison as it has more similarity with our modeled structure. From Fig. 6, we can see that both the full-length LGP2-RNA as well as LGP2 C-terminal-RNA model are showing a similar pattern of fluctuation with the experimentally determined structure of C-terminal domain (2RQA). This similar pattern clearly indicates that our *in-silico* observations are supported by NMR observations.

4. Conclusions

In summary, our study provides a more detailed insight into the LGP2 RNA binding scenario. Though the crystal structure of CTD bound with RNA and the NMR structure without RNA were solved, there is no full-length structure available for LGP2, and information regarding the active site of LGP2 is also limited. In this study we have modeled the full-length structure of LGP2 through homology modelling approach. We have also predicted the active site residues for LGP2 with *in-silico* tool, which was supported by the sequence analysis of Hef helicase domain.

The flexible fitting of the model structure into the LGP2 density map and the position of the helicase domain in the density map is correlating with earlier studies [8]. The successful non biased docking and molecular dynamics simulation studies of RNA-LGP2 complex also supports the position of active site residues. We have also seen that the hydrogen bonding was stable throughout the simulation confirming the tighter bonding between LGP2 and RNA. This study suggests that, in the case of full length LGP2 structure, the critical RNA binding residues spreads out of its niche compared to the earlier reports focusing only on the CTD. Thus our structural information provides the basis for a molecular understanding of the way LGP2 binds, interactions required for binding and conformational change. Further *in vitro* and *in vivo* studies on LGP2 will be helpful for developing new methods against viral attacks.

Figure 6. RMSD profile of full length LGP2 in Green color, predicted C-terminal domain of LGP2 in Red color and NMR structure of C-terminal domain (2RQA) in Black color.

Acknowledgements

AM thanks UGC, New Delhi for providing Major Research Project. CKM acknowledges UGC-RGNF, for the Junior Research Fellowship.

Authors' Contribution

Subhomoi Borkotoky and Chetan Kumar Meena Contributed equally and hence can be considered as co-first authors.

REFERENCES

- [1] Knight, J.C. Genomic modulators of the immune response. *Trends in genetics* : TIG. 2013, 29, 74-83.
- [2] Kolakofsky, D., Kowalinski, E., Cusack, S. A structure-based model of RIG-I activation. *RNA*. 2012, 18, 2118-27.
- [3] Fairman-Williams, M.E., Guenther, U.P., Jankowsky, E. SF1 and SF2 helicases: family matters. *Current opinion in structural biology*. 2010, 20, 313-24.
- [4] Zhu, Z.X., Zhang, X.L., Wang, G.Q., Zheng, H.X. The Laboratory of Genetics and Physiology 2: Emerging Insights into the Controversial Functions of This RIG-I-Like Receptor. *Biomed Res Int*. 2014.
- [5] Rothenfusser, S., Goutagny, N., DiPerna, G., Gong, M., Monks, B.G., Schoenemeyer, A., et al. The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. *J Immunol*. 2005, 175, 5260-8.
- [6] Saito, T., Hirai, R., Loo, Y.M., Owen, D., Johnson, C.L., Sinha, S.C., et al. Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. *P Natl Acad Sci USA*. 2007, 104, 582-7.
- [7] Satoh, T., Kato, H., Kumagai, Y., Yoneyama, M., Sato, S., Matsushita, K., et al. LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *P Natl Acad Sci USA*. 2010, 107, 1512-7.
- [8] Murali, A., Li, X.J., Ranjith-Kumar, C.T., Bhardwaj, K., Holzenburg, A., Li, P.W., et al. Structure and function of LGP2, a DEX(D/H) helicase that regulates the innate immunity response. *J Biol Chem*. 2008, 283, 15825-33.
- [9] Takahashi, K., Kumeta, H., Tsuduki, N., Narita, R., Shigemoto, T., Hirai, R., et al. Solution structures of cytosolic RNA sensor MDA5 and LGP2 C-terminal domains: identification of the RNA recognition loop in RIG-I-like receptors. *J Biol Chem*. 2009, 284, 17465-74.
- [10] Li, X., Ranjith-Kumar, C.T., Brooks, M.T., Dharmiah, S., Herr, A.B., Kao, C., et al. The RIG-I-like receptor LGP2 recognizes the termini of double-stranded RNA. *J Biol Chem*. 2009, 284, 13881-91.
- [11] Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y., Morikawa, K. Crystal structure and functional implications of *Pyrococcus furiosus* hef helicase domain involved in branched DNA processing. *Structure*. 2005, 13, 143-53.
- [12] Cordin, O., Banroques, J., Tanner, N.K., Linder, P. The DEAD-box protein family of RNA helicases. *Gene*. 2006, 367, 17-37.
- [13] Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., et al. Comparative protein structure modeling using MODELLER. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]*. 2007, 50, 2.9.1-2.9.31.
- [14] Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bairoch, A. UniProtKB/Swiss-Prot. *Methods Mol Biol*. 2007, 406, 89-112.
- [15] Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., Sali, A. Protein structure fitting and refinement guided by cryo-EM density. *Structure*. 2008, 16, 295-307.
- [16] Binkowski, T.A., Naghibzadeh, S., Liang, J. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic acids research*. 2003, 31, 3352-5.
- [17] Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M.D., Ritchie, D.W. HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic acids research*. 2010, 38, W445-9.
- [18] Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 2008, 4, 435-47.
- [19] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*. 1995, 117, 5179-97.
- [20] Kawata, M., Nagashima, U. Particle mesh Ewald method for three-dimensional systems with two-dimensional periodicity. *Chem Phys Lett*. 2001, 340, 165-72.
- [21] Hess, B., Bekker, H., Berendsen, H.J., Fraaije, J.G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry*. 1997, 18, 1463-72.
- [22] Miyamoto, S., Kollman, P.A. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry*. 1992, 13, 952-62.
- [23] Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*. 1993, 26, 283-91.
- [24] Colovos, C., Yeates, T.O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science : a publication of the Protein Society*. 1993, 2, 1511-9.
- [25] Eisenberg, D., Luthy, R., Bowie, J.U. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods in enzymology*. 1997, 277, 396-404.
- [26] Thompson, J.D., Higgins, D.G., Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*. 1994, 22, 4673-80.
- [27] Laskowski, R.A. PDBsum: summaries and analyses of PDB structures. *Nucleic acids research*. 2001, 29, 221-2.
- [28] Wiederstein, M., Sippl, M.J. ProSA-web: interactive web

service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*. 2007, 35, W407-W10.

- [29] Pandurangan, A.P., Topf, M. RIBFIND: a web server for identifying rigid bodies in protein structures and to aid

flexible fitting into cryo EM maps. *Bioinformatics*. 2012, 28, 2391-3.

- [30] Wallace, A.C., Laskowski, R.A., Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering*. 1995, 8, 127-34.