

Structural and Functional Characterization of RNA-binding Site of Rev and Rev-Response-Element RNA Complex via All Atom Molecular Dynamics Simulations

Zaheer UL-Haq^{*,1}, Nizakat Ali¹, Mohamed H. Al-Agamy^{2,3} and Assem Barakat⁴

¹Dr. Panjwani Center for Molecular Medicine and Drug Research, H.E.J. Research Institute of Chemistry, International Center for Chemical and Biological Sciences, University of Karachi, Karachi-75270, Pakistan.

²Division of Microbiology, Pharmaceutics Department, College of Pharmacy, King Saud University, P.O. Box 2457, Riyadh 11451, Saudi Arabia.

³Microbiology and Immunology Department, Faculty of Pharmacy, Al-Azhar University, Cairo, Egypt.

⁴Department of Chemistry, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

* Corresponding author
Email: zaheer.qasmi@iccs.edu

Abstract: Nuclear export of viral mRNAs, is an essential step in the HIV replication cycle. This role is played by a small regulatory protein of HIV-1 called Rev. The N-terminal region of Rev contains an arginine-rich sequence. The arginine-rich motif (ARM) is located between amino acids 38-50 and forms an alpha-helical secondary structure. Expression of the structural proteins of human immunodeficiency virus type 1 requires the direct interaction of multiple copies of the viral Rev protein with its highly structured RNA target sequence, the Rev Response element (RRE). The major viral proteins are not produced if this transport of RNA is stopped. Therefore, knowledge of Rev structure is essential for understanding of its cooperative binding to the RRE, for understanding the mechanism of HIV infection and for the development of antiviral drugs that interfere with Rev's essential functions and for acknowledgment of good candidate drugs for treatment of AIDS. To understand how REV interact with RRE element of HIV-RNA and its formation of oligomeric complex it is better to characterize the domain wise structure of REV with regard in function of each domain. Due to lack of structural data on Rev no single compound is reported as inhibitor of REV expect antiviral drugs. Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev Protein is much more important. The ARM is a highly specific sequence which allows for the multimerization of Rev and also binding of REV with RNA. Here we are first time exploring the structural characteristics of REV protein both in free form and in complex with RNA at domain function level especially explore the role of ARM motif in REV HIV-1 protein as RNA binding sites by molecular dynamics (MD) simulation and homology modeling studies. Results indicate that the arginine-rich motif (ARM) is crucial in stability of this complex. The residues ARG38, 39, 41, 43, 44, 48, 50, and ASN40 are most interacting with nucleobases of RRE in Crystal structure of Rev and Rev-response-element RNA complex. Our study plays a major role in elaboration of binding of RNA with REV and pave the way for further investigation for therapeutically agent for HIV.

Key words: molecular dynamics simulation; RNA-binding; rev protein; rev response element (RRE); HIV-virus.

1. Introduction

HIV virus is a retrovirus belongs to family of virus called Retroviridae contain ribonucleic acid (RNA) as a genetic material [1-3]. Total size of genome of HIV virus is 9 kb and encode for 15 proteins [4-5]. Before translation of this RNA into protein it undergoes various strategies such as overlapping reading frames and alternative splicing [6-8]. After splicing three types of mRNA transcripts are produced one is unspliced, singly-spliced and other is full-spliced RNA [9-10]. HIV-1 regulatory proteins (including Rev) are translated from completely spliced mRNA transcripts, while structural proteins are translated from incompletely spliced transcripts [11]. Both the single-spliced viral mRNAs and unspliced RNA must be exported from nucleus to cytoplasm where it is translated into all essential structural protein that make complete new virions [12]. This nuclear export is promoted by the 13 kDa HIV-1 Rev protein [13-14]. Before exporting of RNA, REV protein migrates into nucleus where it undergoes oligomerization and then make complex with RNA [15]. REV protein recognizes a highly conserved 350 nucleotide structured intronic RNA element, called the Rev Response Element (RRE) [16]. Then the export of RNA is carried out by Crm1 nuclear export pathway [17]. First Rev oligomerizes onto the RRE to form a ribonucleoprotein (RNP) complex of 200–300 kDa containing 8–10 Rev molecules [18-19], that binds Crm-1 (exportin-1), GTP-bound Ran, and other host cell proteins via the Rev nuclear export sequence (NES), to promote nuclear export [20]. Structure of REV contain 116 a.a in a single chain [21]. It contains multiple conserved domain which are known essential for the function of REV. At N-terminal it contains arginine-rich RNA-binding motif (ARM) that specifically binds to a purine-rich bulge within stem loop IIb of the RRE [14] [22]. The RRE contains several stem loops such as Stem iA, iiA, iii/ivA, vA, Stem iiB and Stem iiC. The most well characterized being is the iiB for binding sites for REV ARM domain [23] [24]. HIV RNA and REV make a ribo-nucleus RNA protein (RNP) complex that have to export from nucleus to cytoplasm. For this export REV protein has a conserved sequence of amino acids At C-terminal (71-82 residues) called NES (Nuclear export sequence) [1]. It is also called REV activating domain and rich with leucine residues [17]. As befits their biological importance, the interactions whereby Rev recognizes RRE-containing RNAs have been intensively studied, as has Rev's propensity to oligomerize, both in the absence of RNA and in the context of the RRE [23-26].

Rev is a good target for antiviral therapies, since Rev is absolutely necessary for HIV-1 replication [27, 28]. It has been shown that various organic compounds have the ability to target the Rev/RRE interaction [29]. Neomycin B, diphenylfuran cation, and proflavine are

small molecules that can prevent Rev from binding to the RRE sequence [27, 29, 30]. If Rev is incapable of binding to the RRE on the pre-mRNA, the RNA will not be exported to the cytoplasm, also resulting in lack of necessary structural proteins.

In our study we elaborated the large-scale molecular dynamics simulation behavior of a Crystal structure of Rev and RRE complex REV, and free REV protein (homology model), in order to see the structural part of REV protein that help in assembly of REV units and then binding of these units with RNA. Our results suggest that in structural domain (ARM) some residues such as ARG and ASN in position (38-50) are essential for binding with RNA and some residues of oligo-domain (9-26) and (51-65) such as LEU18,22, ILE59,55 and PHE21 provide a central hydrophobic core and are necessary in oligomerization. In this paper we also identified the nucleotide sequence of HIV RNA that are most reactive with REV protein. Our result suggested that some nucleotide of HIV RNA such as U45, 66, 72/G46, 47, 67, 70, and A44 belong to stem iiB of RRE elements are in contact with REV protein residues. The ARM motif of REV is well elaborated structurally and its importance in interaction of viral Rev protein with its highly structured RNA target sequence, the Rev Response element (RRE) is identified. We also elaborated Nucleotides interaction of RRE with REV protein.

2. Material and Methods

2.1 Molecular Dynamics Simulation

Crystal structure of Rev and Rev-response-element RNA complex (PDB: ID 4PMI) was downloaded. In crystal structure two molecules of REV protein (1-72) residues bound with RRE element (stem iiB). Crystal structure were partially charged, Protonated and energy minimized by MOE software. MD simulations of 4PMI and homology model was carried out by version 2016.4 of GROMACS. The AMBER96 protein, nucleic AMBER94 force field was used throughout MD simulations. We used a simple cubic box as the unit cell, the modeled protein was centered in this box and placed at least 1.0 nm from the box edge (-d 1.0). We used spc216.gro, which is a generic equilibrated 3-point solvent model Water to solvate the box. The ions were added to neutralize the system by using the genion algorithm of Gromacs. The structure was relaxed through a process called energy minimization (EM) with steepest descent minimization. The system was equilibrate both at temperature around 300 K and pressure at 1 bar, by using Berendsen thermostat and Parrinello-Rahman barostat respectively. To constrain all bonds Lincs algorithm was used and for integration the leapfrog algorithm was used.

Particle-mesh Ewald (PME) summation was used to account the Electrostatic forces with cutoff was 10 Å for both electrostatics and van der Waals interactions. A 20 ns of md was run with these conditions for both homology model and crystal structure. After MD simulation the results were analyzed by GROMACS tools and VMD.

2.2 Homology modeling

The HIV-REV sequence (residues 1–116; NCBI Reference Sequence: NP_057854.1, www.ncbi.nlm.nih.gov) was defined as the target sequence. 3D structure of Rev protein was modeled using I-TASSER [31].

I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical approach used both for prediction of protein structure and function of that protein based on its structure [32]. The full-length sequence of REV protein from NCBI server with ID mentioned before was submitted in FASTA format to I-TASSER. The first step in homology modeling is identification of threading templates. For this purpose, I-TASSER used LOMETS (meta-server threading approach). This server search template from the PDB library. From targeted template I-TASSER form a set structural conformation called as decoys. For the construction of 3-D models, I-TASSER uses the SPICKER program. This program cluster all the decoys based on the pair-wise structure similarity and reports up to five models. Biological annotation of 5 selected model was the carried out by COFACTOR and COACH [21].

3. Results and Discussion

3.1 Homology modeling

As there is no homologous protein structures deposited in the Protein Data Bank (PDB) of the HIV-1 protein Rev, so we used Protein threading, also known as fold recognition method to model 3-D structure of REV protein. In Protein threading we searched template that have same fold as query and of known structures. It differs from the homology modeling method of structure prediction as it (protein threading) is used for proteins which do not have their homologous protein structures deposited in the Protein Data Bank (PDB). Threading works by using statistical knowledge of the relationship between the structures deposited in the PDB and the sequence of the protein which one wishes to model.

3.2 Top 10 threading templates used by I-TASSER

I-TASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading

programs, where each threading program can generate tens of thousands of template alignments [33]. I-TASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the 10 best templates selected from the LOMETS threading programs. Table no 1 shows top 10 templates used in modeling and their alignments against query sequence is shown in Figure 1.

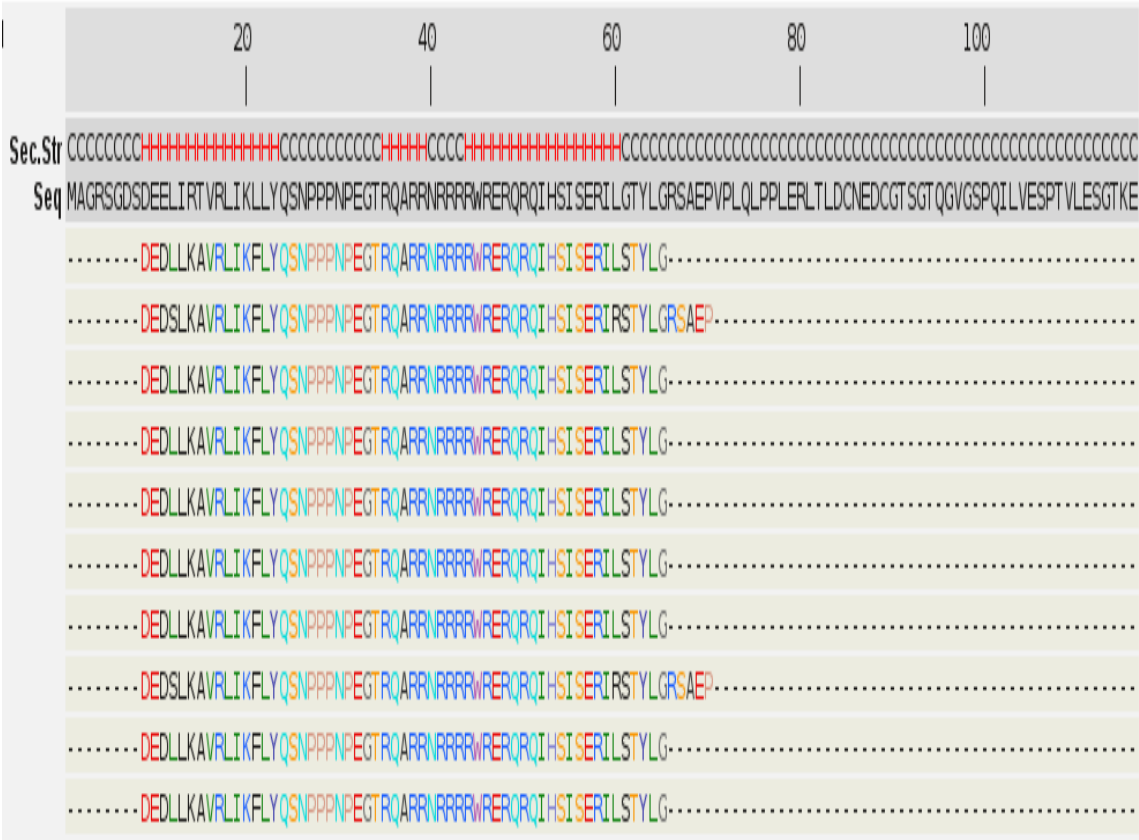


Figure 1: The top 10 template-query alignments generated by LOMETS. All the residues are coloured in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in colour. Colouring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are coloured in dark shade.

Rank	Iden1	Iden2	Cov	PDB:hits	Z-score
1	0.89	0.44	0.49	2X7LM	2.75
2	0.87	0.47	0.53	3LPHA	2.01
3	0.89	0.44	0.49	2X7LM	3.30
4	0.89	0.44	0.49	2X7LM	9.07
5	0.89	0.44	0.49	2X7LM	6.33
6	0.89	0.44	0.49	2X7LM	2.98
7	0.89	0.44	0.49	2X7LM	9.45
8	0.87	0.47	0.53	3LPHA	2.27
9	0.89	0.44	0.49	2X7LM	3.11
10	0.89	0.44	0.49	2X7LM	1.19

Table 1: Top 10 threading templates used by I-TASSER. **Rank** of templates represents the top ten threading templates used by I-TASSER. **Ident1** is the percentage sequence identity of the templates in the threading aligned region with the query sequence. **Ident2** is the percentage sequence identity of the whole template chains with query sequence. **Cov** represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein. Normal **Z-score** is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.

3.3 Top 5 final models predicted by I-TASSER

For each target, I-TASSER simulations generate a large ensemble of structural conformations, called decoys. To select the final models, I-TASSER uses the SPICKER program [34] to cluster all the decoys based on the pair-wise structure similarity and reports up to five models which corresponds to the five largest structure clusters (Figure 2). The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa. Since the top 5 models are ranked by the cluster size, it is possible that the lower-rank models have a higher

C-score in rare cases. Although the first model has a better quality in most cases, it is also possible that the lower-rank models have a better quality than the higher-rank models. If the I-TASSER simulations converge, it is possible to have less than 5 clusters generated; this is usually an indication that the models have a good quality because of the converged simulations.

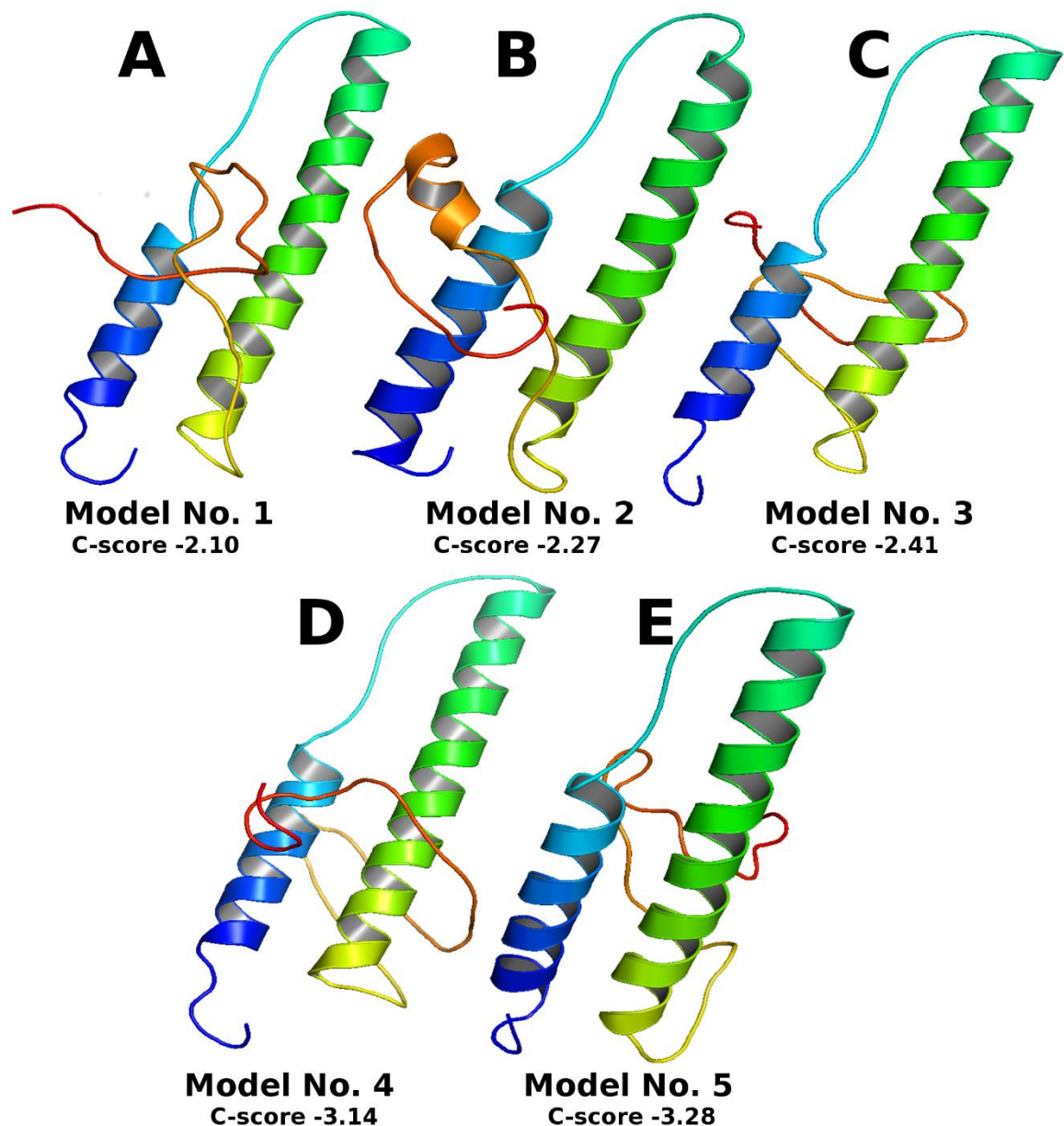


Figure 2: Top 5 final models predicted by I-TASSER

3.4 Predicted function using COFACTOR and COACH

In the NMR structure of REV protein its α -helical ARM motif bound to the IIB RNA hairpin [8], and a second binding site is in stem IA [35]. To find out how the individual

subunits of Rev recognize the RRE, we predicted the binding sites of REV model by COFACTOR and COACH based on the I-TASSER structure prediction. While COFACTOR [36] deduces protein functions (ligand-binding sites, EC and GO) using structure comparison and protein-protein networks, COACH [37] is a meta-server approach that combines multiple function annotation results (on ligand-binding sites) from the COFACTOR, TM-SITE and S-SITE program. We only predicted ligand binding sites of Rev model. This result showed that in REV protein model the most interacting domain from residues 36-60 is ARM domain. Table 2 and Figure 3 shows ligand binding sites.

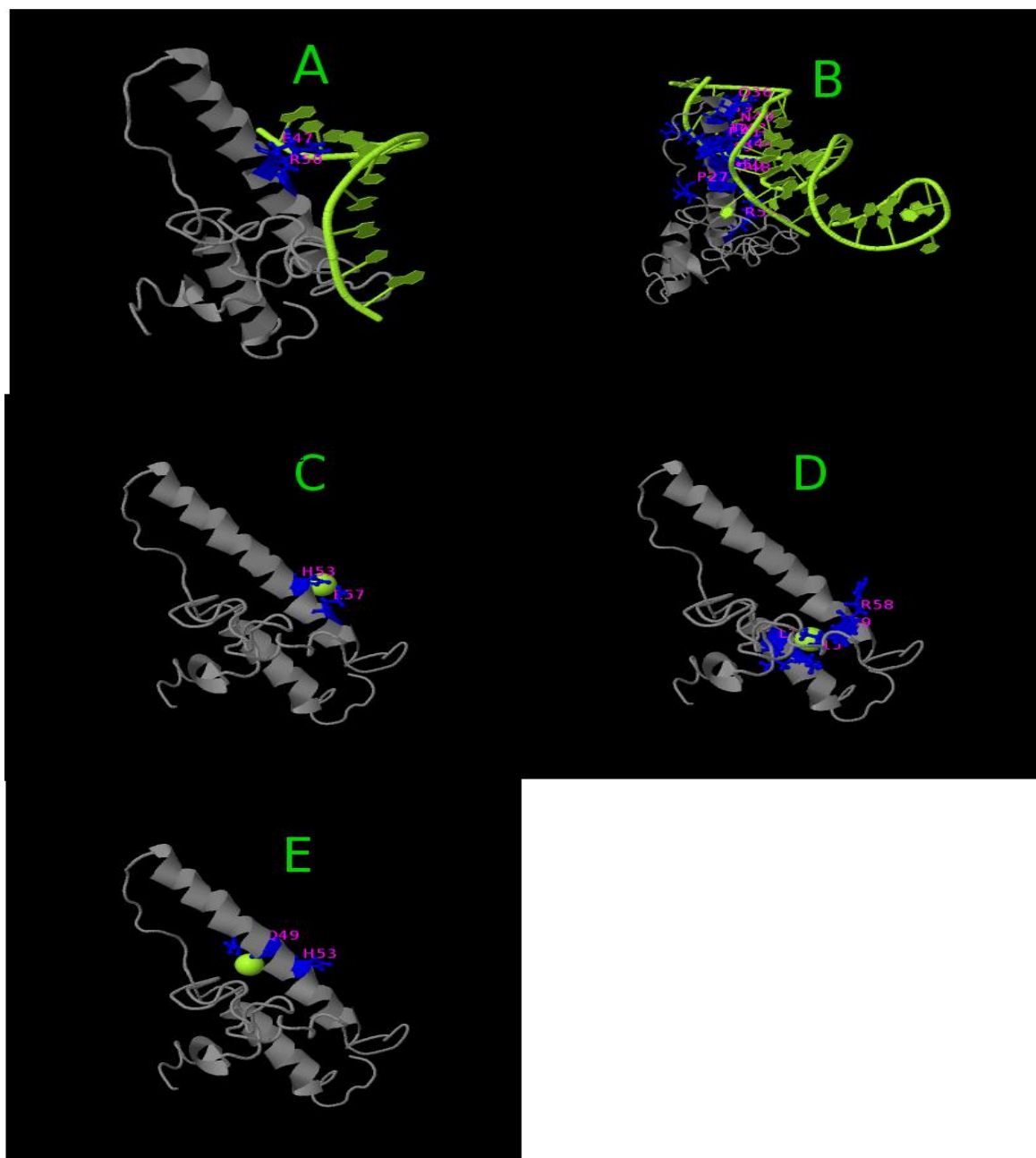


Figure 3: Predicted ligand binding sites. Binding residues are shown in blue ball & stick.

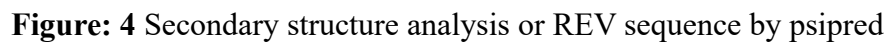
Predicted binding ligand are shown in green sphere. (A) complex with nucleic acid. (B) complex with Zn. (C) complex with Cu. (D) complex with XE.

Table 2: C-score is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction. **Cluster size** is the total number of templates in a cluster.

Rank	C-Score	Cluster size	Ligand name	Ligand binding sites residues
1	0.10	5	Nucleic acid	47, 50
2	0.06	3	Nucleic acid	27,36,39,40,41,42,43,44,46,49,50,58
3	0.04	2	Zn	53, 57
4	0.04	2	XE	15,18,58,59
5	0.04	2	CU	49, 53

3.5 Secondary analysis of REV sequence by psipred on line webserver

The HIV-REV sequence (residues 1–116; NCBI Reference Sequence: NP_057854.1, www.ncbi.nlm.nih.gov) was used to predict secondary structure of REV by PSIPRED. The structure of REV protein consists of highly conserved motif called helix-loop-helix motif [41]. Secondary analysis of REV sequence also showed these conserved motifs. The region from 9-26 (oligomerization domain) that contain conserved hydrophobic residues such as LEU, PRO and ILU and serve as a hydrophobic interface providing surface area for the contact of different units of REV to be organized into highly oligomeric form during binding with RRE elements form alpha-helix structure (Figure 4). The region from 34-60 (Arginine-rich motif ARM) also form alpha helix. These two α helix are connected with by short coiled loop. Coiled loops are present between these two helix and at N-Terminal (1-10) and C-terminal Domain (60-116) and form same conserved helix-loop-helix motif (Figure 4). The conformation of these helices changes and come close to each other during the binding with RNA. However, no β strand are found throughout the sequence.



Inspection of secondary structure of Crystal structure of Rev and Rev-response-element RNA complex and homology model indicate that there is tremendous conformational flexibility in REV protein in free form. This result support the idea that REV protein has intrinsic property of aggregation and fold incorrectly when it is not bound with RNA, and when it binds to RNA its conformational flexibility vanished away and it becomes more rigid and become stable in helix-loop-helix conformation [22] [38]. As in Figure 5 A, two helices are formed one in region 9-26 (oligomerization domain) and of Arginine-rich motif ARM (from 34-60). REV protein in free form (homology model) also form same secondary elements but there are some conformational changes during MD simulation as a function of time during 20 ns (Figure 6 B). In free form some residues of oligo domain (9-26) undergoes conformational change. LEU21, 22, TYR23, GLN24, SER25 and ASN26 undergoes huge conformational change from α -helix to turn and β strand. These residues provide hydrophobic surface area interface between different units of REV protein during RNA binding. Their fluctuation in free form but not in complex support the idea that during binding

with RNA these residues help in oligomerization of REV molecules around RNA. The ARM motif (38-50) also showed some structural changes. During the 20ns simulation they change their alpha helix structure into turn and coil after a time of about 10 ns. This also support that during the RNA binding these motifs form fully well organized into alpha-helix structure. The NES domain in case of free REV (homology model) but not in complex also have structural flexibility. The residue LEU81 form beta strand in this domain. Other residues at C-terminal GLU87, ASP88, CYS89 also form beta strand, this is not in case of well model 3-D structure of REV. Finally, MD simulation analysis clearly showed that REV proteins secondary elements become more rigid with binding to RNA. It may be possible that RRE element make it possible for REV to change its shape and its rigidity [22].

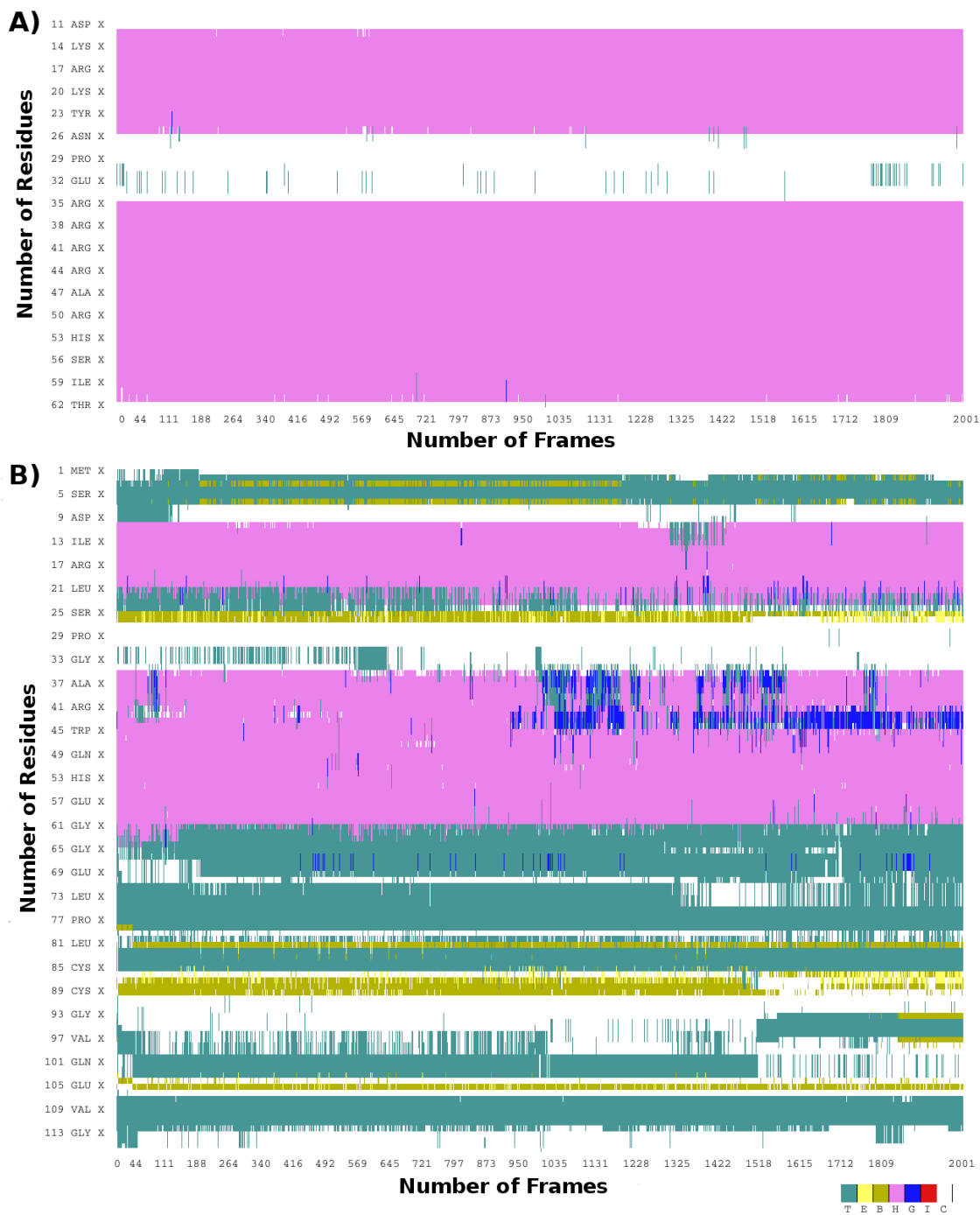


Figure: 5 Secondary structure analysis after MD simulation. (A) Crystal structure of Rev and Rev-response-element RNA complex REV (B) REV protein homology model. Amino acids sequence is shown on Y-axis and number of frame is given on X-axis. The colour plate shows the secondary elements T = turn C = coil E= extended conformation B= isolated bridge H = alpha-helix G = 3-10 helix and I = pi-helix.

3.7 The Molecular Dynamics of Crystal Structure of Rev and Rev-Response-Element RNA Complex and Homology Model

For looking structural stability and conformational changes RMSD of the backbone atoms of the whole trajectories of 20 ns of the Crystal structure of Rev and Rev-response-element RNA complex and homology model of REV was calculated. Figure 6 depicts the RMSD values as calculated for the whole Crystal structure of Rev and Rev-response-element RNA complex (black), and homology model of rev protein (red). There is a large conformational changes between complex and homology model with an increased RMSD of 0.2 nm (in case of complex) and 0.5 nm (in case of homology model). The RMSD of the REV protein in crystal structure in complex with RNA exhibits very little structural fluctuations. It increases until ~ 0.35 nm, stays around this value for ~ 4 ns, decreases for a short while, and then stabilizes at ~ 0.18 nm. The RMSD track of the REV protein (homology model), exhibits almost same pattern as of crystal structure in regard with conformational stability. From starting until the end of the simulation, the RMSD of the REV model is mostly stable.

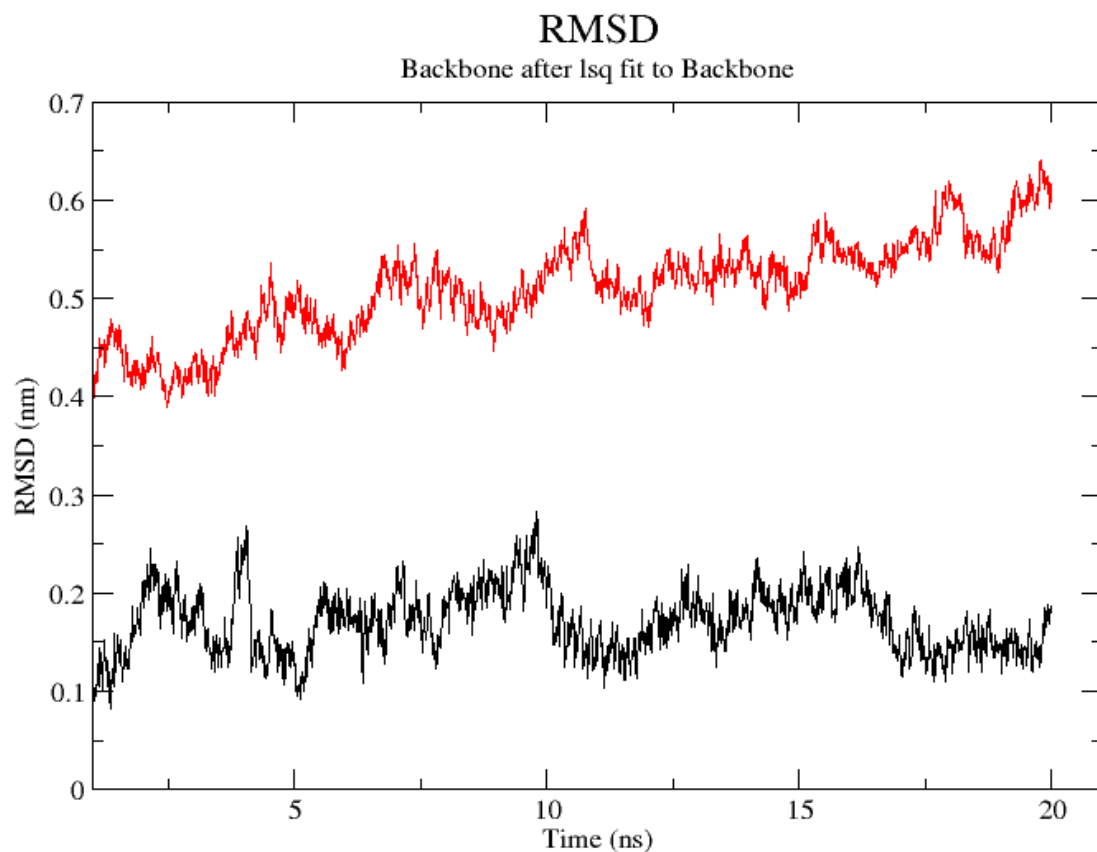


Figure 6: RMSD of Crystal structure of Rev and Rev-response-element RNA complex REV (black colored) and REV protein homology model (red colored).

3.8 The Root Mean-Square Fluctuation (RMSF) Analysis

To further analyze the trajectories of the Crystal structure of Rev and Rev-response-element RNA complex REV and REV protein homology model, we computed the standard deviation from the RMSD for each of its residues, i.e., their root mean-square fluctuations (RMSF). Figure 7 presents the RMSF of the REV protein complex with RNA (black Lines) and REV protein homology model (Red Line).

It should be noted that PDB 4pmi that is Crystal structure of Rev and Rev-response-element RNA complex REV contain only N-terminal domain of REV protein from 1-72 amino acids. And we are also focused our study to this region which is RNA binding. In general, the crystal structure of REV protein complex with RNA is more rigid and confined and tend to be less flexible than its protein structure without RNA [38]. In free form (homology model) the residues LEU21, 22, TYR23, GLN24, SER25 and ASN26 shows higher fluctuation and an average value of RMSF 0.4nm. This can be correlate with that these residues also showed structural flexibility in secondary structure analysis. And in case of complex these residues become rigid and showing very less flexibility as their RMSF values are less than 0.1 nm in complex. The RMSF value of ARM motif (38-50) they showed very little flexibility (0.15nm) and as their secondary structure analysis showed they have little flexibility. In case of complex this motif (ARM) become stable and showing no flexibility (0.05nm). The other two domain of REV protein in free form from residues 65-70 and 90-100 at C-terminal shows higher values of RMSF around 0.5nm.

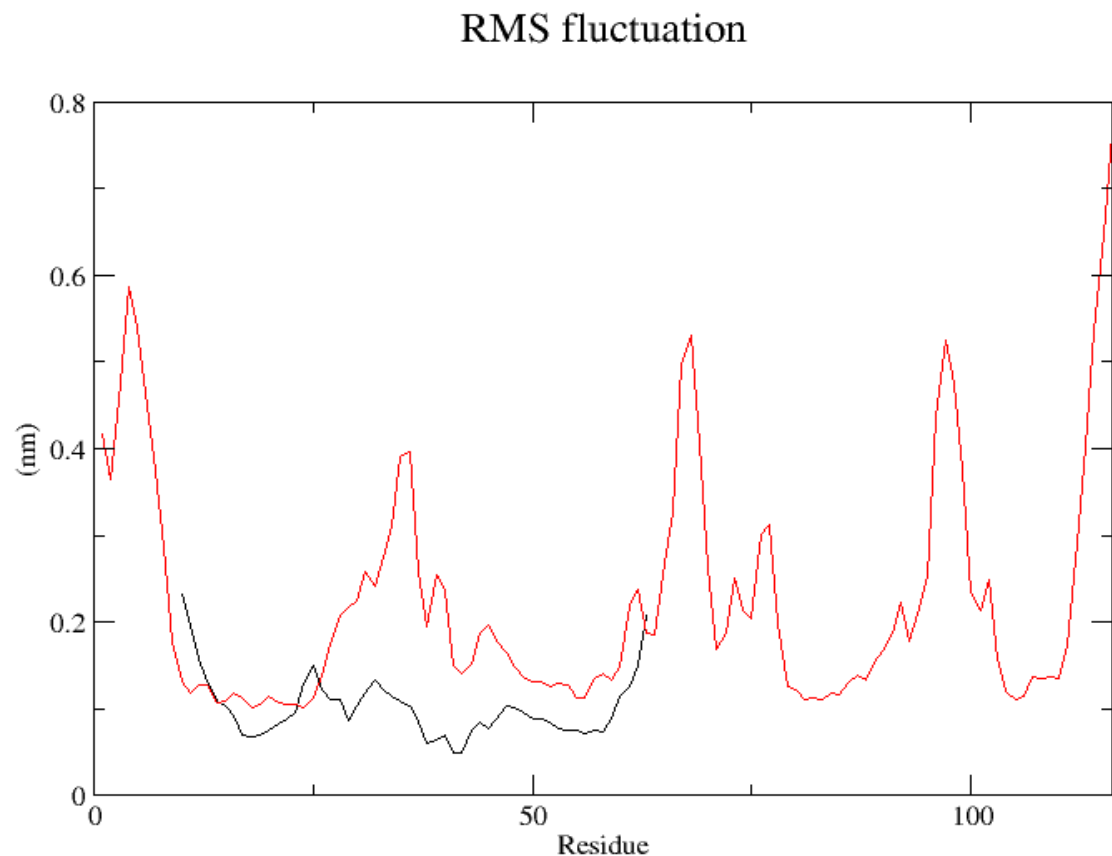


Figure 7: RMSF of Crystal structure of REV and Rev-response-element RNA complex (black colored) and free REV protein homology model (red colored).

3.9 Hydrogen-bond analysis of of Crystal Structure of Rev and Rev-Response-Element RNA Complex After MD

In the Crystal structure of Rev and Rev-response-element RNA complex REV [39] the R38 make hydrogen bond with U66 and G67 of RNA, R39 with G70, N40 with G47, R41 with G46, R43 with U72, R44 with U45, R48 with A44 and R50 with U72 as in Figure 8. The residues at position 42, 45, 46, and 47 in ARM motif make no interaction.

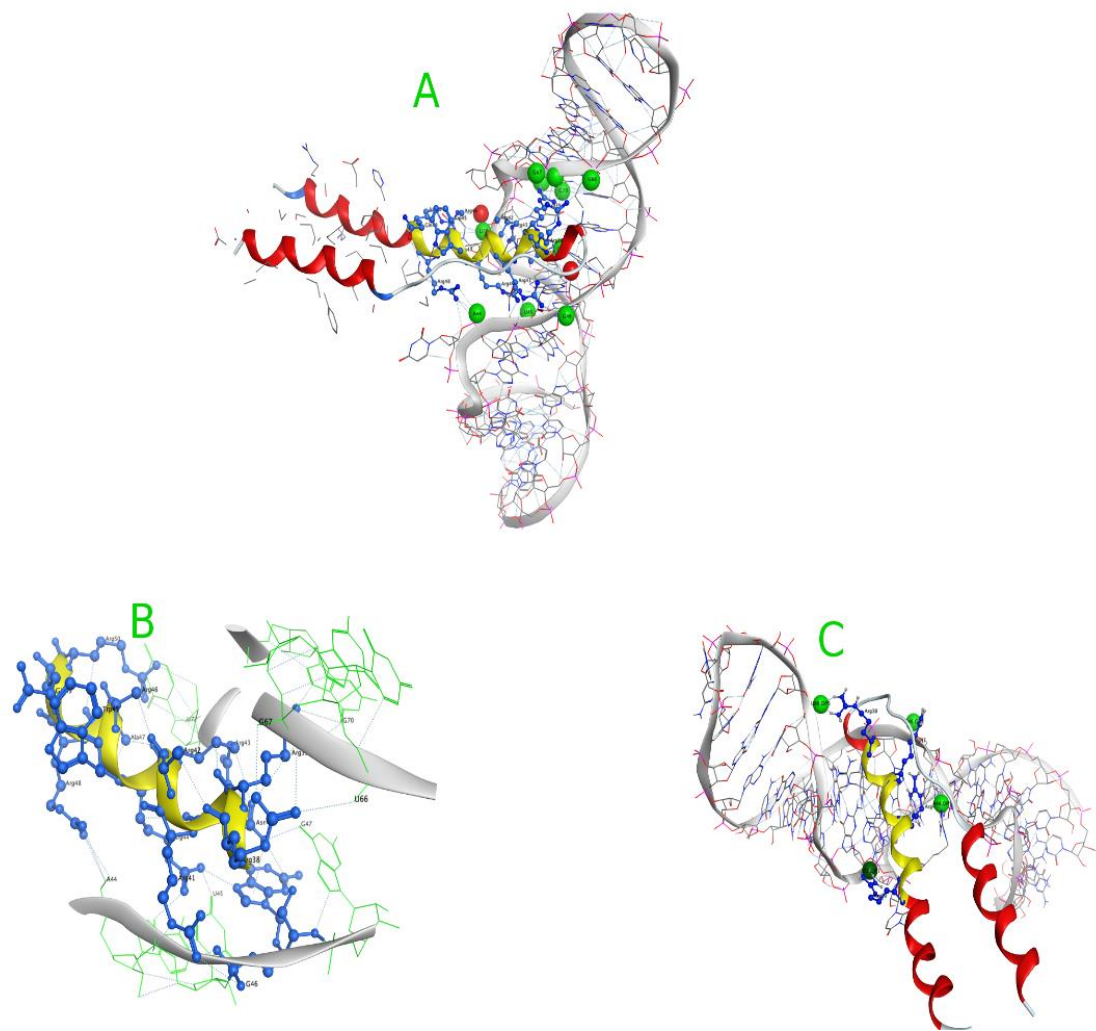


Figure:8 (A) Crystal structure of Rev and Rev-response-element RNA complex REV (PDB: ID 4PMI). (B) ARM(arginine rich motif) of REV protein (C) crystal structure of REV complex after MD simulation of 20ns. ARM motif is shown in yellow ribbon, interacting residues are shown in blue ball&stick and interacting nucleobases are shown in green sphere.

After MD simulation hydrogen bond analysis was carried out. During MD simulation of 20ns the average hydrogen bond between R38, U66, G67, R39, G70, R41, G46, R48, A44 and R50, U72, U72, are 1.4, 1.9 and 1.7, 1.2 and 1.3, respectively. All the trajectories show that these hydrogen bond are constant throughout the MD simulation as Figure 9. But the hydrogen bond between N40G47, R43U72, R44U45 are not constant as in Figure 9.

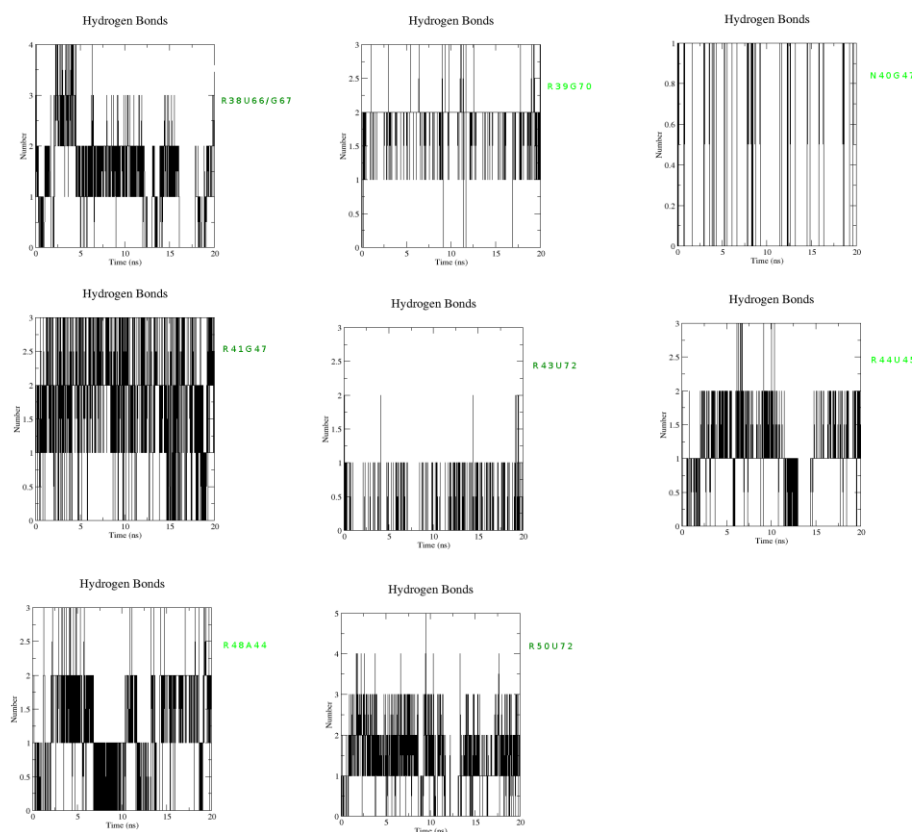


Figure:9 Hydrogen bond analysis.

3.10 Solvent Accessible Surface Area (SASA)

Prior to binding of REV with HIV-RNA, REV protein forms a oligomeric complex. Oligomerization process of REV protein is cooperative. One molecule of REV helps other molecule for binding to RNA. The domain (9-26) called oligo domain and ARM-domain (RNA binding domain) in REV protein help in oligomerization [14] [35] [39]. Some residues such as LEU and ILE provide a hydrophobic interface surface in the center of REV structure for uniting different molecule of REV protein and make ARM domain in such position ready for interacting with RRE elements. SASA calculation help in recognition of these hydrophobic and hydrophilic part of REV protein. So, we calculate the SASA value of ARM domain, oligo-domain and NES-domain (NES domain is only present in REV homology model but not in crystal complex structure). Total solvent accessible surface area as a function of time was is higher in free form (homology model) of REV protein (is about 90 nm^2) compared with crystal structure (50 nm^2). As in Figure 10 A and B. Average area per residues calculated that shows much similarity in homology model and crystal structure, as in Figure 11 A and B. In homology model ARG residues have the higher value of SASA due to hydrophilic nature throughout the sequence. Otherwise ARM domain has higher values about

330 Å. the NESs domain on the other side that recruit Crm1 [40] [41] has lower values 274 Å [28]. Due to hydrophobic residues (LEU 75, 78, 81, 83, PRO 76, 77 and THR 82. in crystal structure the average SASA of ARM motif is 328 Å that is not difference much as compared with homology model. Crystal structures of Rev and homology model some of residues (LEU12, ILU13, VAL16, LEU18, ILU19, LEU21, LEU22, ILU52, ILU55, ILU59) have lowest SASA value so they form a central hydrophobic core that mediates interactions between Rev subunits and positions the ARMs on one side of the Rev oligomer to bind RNA [14]. Lower SASA value of residue ILE55 (283 Å in free form and 278 Å in complex) indicate that this residue is essential to maintain the dimeric state of REV. As Edgcomb et al describe that the Rev subunit bound to the junction site forms an extensive surface with the RNA that causes the dimer interface to pivot around a single residue, Ile55, that is known to be critical for dimer integrity and function (Jain and Belasco, 2001; Edgcomb et al., 2008). Estimated solvation free energy as a function of time of homology model of REV was 0 as compare with crystal structure in complex with RNA was about 5 as in Figure 10 C and D. also support the idea that free REV has ability for interaction.

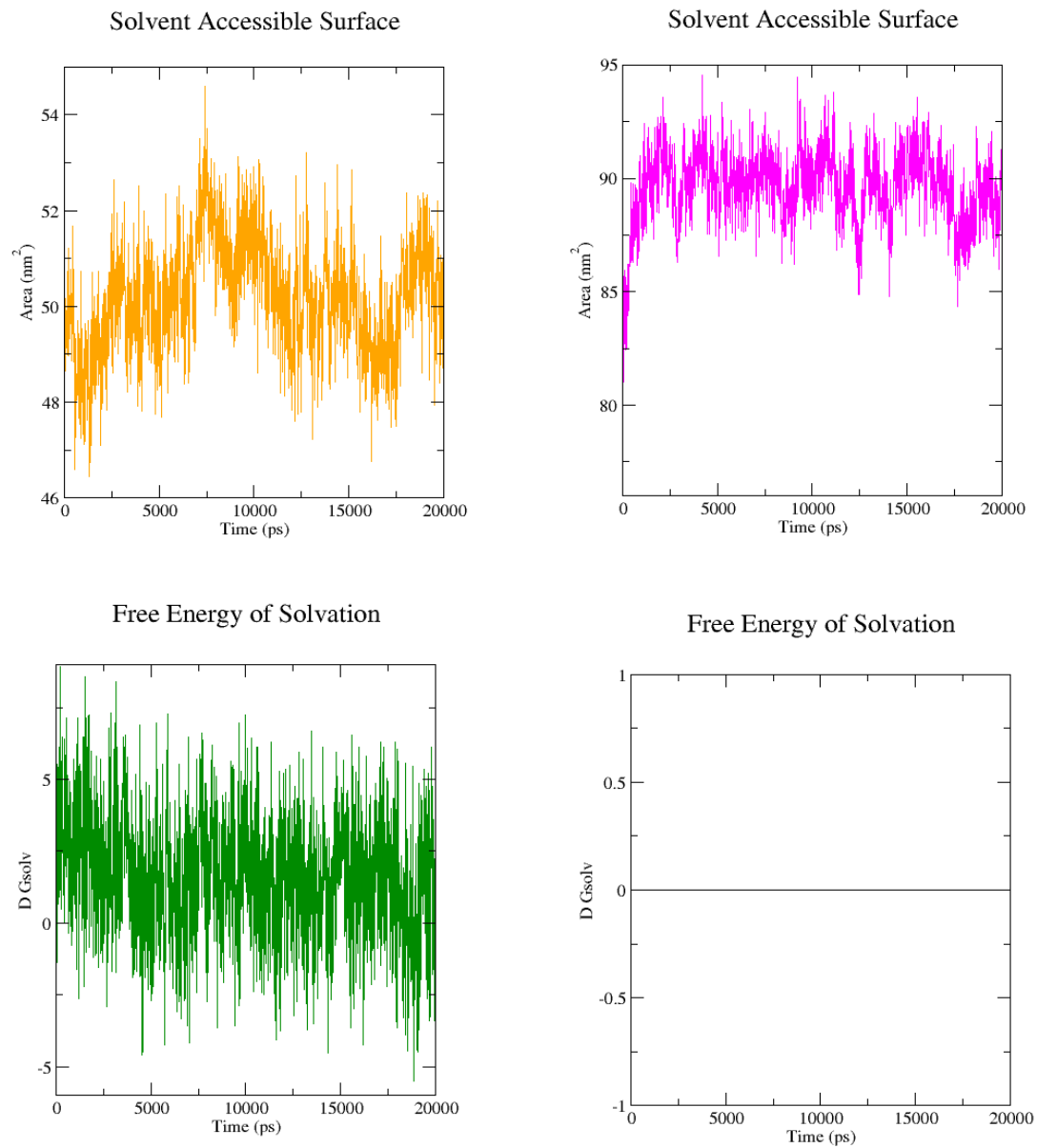


Figure 10: Solvent accessible surface area (SASA) analysis of crystal structure of REV and homology model. (A,B) Total area as a function of time (C,D) Estimated solvation free energy as a function of time.

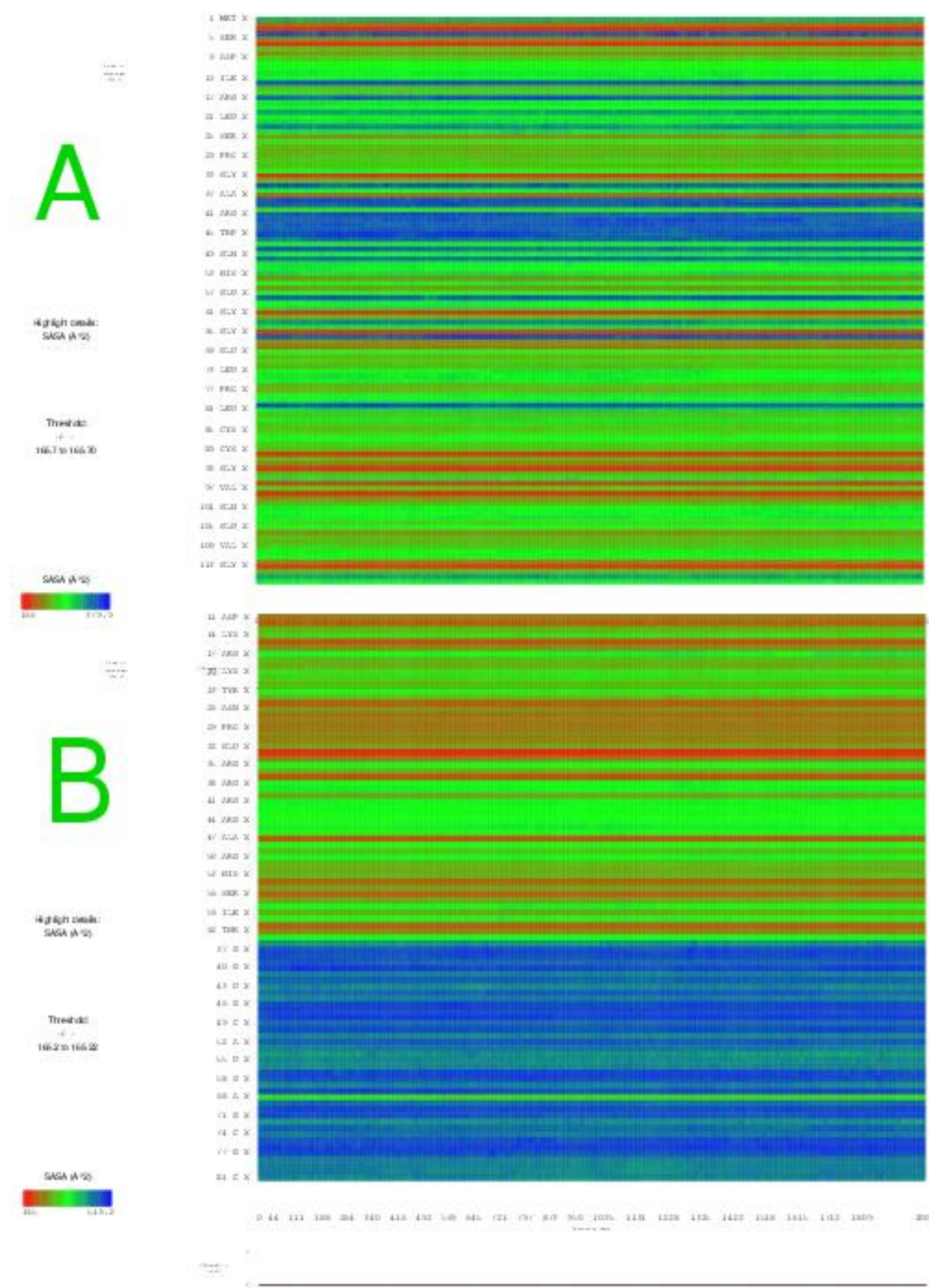


Figure 11: Solvent accessible surface area (SASA) of individual residue.(A) homology mode (B) complex crystal structure. Analysis were performed by VMD.

4. Conclusion

We have developed homology model of REV protein and MD simulation of crystal structure of Rev and Rev-response-element RNA complex, aiming to characterize the ligand-

binding sites of REV protein. The ligand-binding sites of homology model was deduced by COFACTOR indicating that REV has ability to bind with nucleic acids such as RNA. The most interacting residues with nucleic acids belong to domain called ARM (arginine rich domain). In crystal structure of REV we found that R38, 39, 41, 48, 50 and N40 are in contact with nucleotides of RRE. And these interactions are constant throughout MD simulation. In COFACTOR results we identified that domain ARM is interacting domain with nucleic acids. Further secondary analysis of REV model indicates that there are conformational changes occurs between complex and free REV protein. Further, these conformational changes are conformed with RMSD and RMSF calculation of REV in free form and with complex. SASA calculation also divide the protein into its hydrophobic and hydrophilic part and help in identification of role of different domain inside the structure of REV protein. The RNA binding sites that is ARM motif in REV protein is well characterized structurally and functionally.

References

1. Malim, MH; McCarn, DF; Tiley, LS; Cullen, BR. Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. *J. Virology*. **1991**, 65 (8): 4248–4254.
2. Meyer, BE; Meinkoth, JL; Malim, MH. Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals. *J. Virology*. **1996**, 70 (4): 2350–2359.
3. Feinberg, Mark B. et al. HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell*. **1986**, 46 (6): 807–817.
4. Gallo, RF; Wong-Staal, F; Montagnier, L; Haseltine, WA; Yoshida, M. HIV/HTLV gene nomenclature. *Nature*. **1988**, 333 (6173): 504.
5. Feinberg, MB; Jarrett, RF; Aldovini, A; Gallo, RC; Wong-Staal, F. "HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral. *Cell* **1986**, 46:807-817. 10.
6. Cochrane, A; Kramer, R; Ruben, S; Levine, J; Rosen, CA. The human immunodeficiency virus rev protein is a nuclear phosphoprotein. *Virology*. **1989**, 171 (1): 264–266.
7. Sodroski, JW; Goh, WC; Rosen, C; Dayton, A; Terwilliger, E; Haseltine, W. A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature*. **1986**, 321 (6068): 412–417.
8. Surendran, R; Herman, P; Cheng, Z; Daly, TJ; Ching Lee, J. HIV Rev self-assembly is linked to a molten-globule to compact structural transition. *Biophysical Chemistry*. **2004**, 108 (1–3): 101–119.
9. Kjems, J; Calnan, BJ; Frankel, AD; Sharp, PA. Specific binding of a basic peptide from HIV-1 Rev. *European Molecular Biology Organization Journal*. **1992**, 11 (3): 1119–1129.
10. Hope, TJ. The ins and outs of HIV Rev". *Archives of Biochemistry and Biophysics*. **1999**, 365 (2): 186–191.
11. Jayaraman, B., Crosby, D. C., Homer, C., Ribeiro, I., Mavor, D., & Frankel, A. D. (2014). RNA-directed remodeling of the HIV-1 protein Rev orchestrates assembly of the Rev–Rev response element complex. *Elife*. **2014**, 8; 3:e04120.
12. Zapp, M; Hope, T; Parslow, T; Green, M. Oligomerization and RNA binding domains

- of the type 1 human immunodeficiency virus Rev protein: a dual function for an arginine-rich binding motif. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, 88 (17): 7734–7738.
13. Tiley LS, Malim MH, Tewary HK, Stockley PG, Cullen BR. Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 758–762.
 14. Cook KS, et al. Characterization of HIV-1 Rev protein: Binding stoichiometry and minimal RNA substrate. *Nucleic Acids Res.* **1991**, 19:1577–1583.
 15. Pollard VW, Malim MH. The HIV-1 Rev protein. *Annu Rev Microbiol.* **1998**, 52:491–532.
 16. Daly TJ, Cook KS, Gray GS, Maione TE, Rusche JR. Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. *Nature.* **1989**, 342:816–819.
 17. Malim MH, Cullen BR. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: Implications for HIV-1 latency. *Cell.* **1991**, 65:241–248.
 18. Stahl SJ, et al. Generation and characterization of a chimeric rabbit/human Fab for co-crystallization of HIV-1 Rev. *J Mol Biol.* **2010** in press.
 19. Mann DA, et al. A molecular rheostat. Co-operative Rev binding to stem I of the Rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J Mol Biol.* **1994**, 241:193–207.
 20. Daly TJ, et al. Biochemical characterization of binding of multiple HIV-1 Rev monomeric proteins to the Rev-responsive element. *Biochemistry.* **1993**, 32:10497–10505.
 21. Suhasini M, Reddy TR. Cellular proteins and HIV-1 Rev function. *Curr HIV Res.* **2009**, 7:91–100.
 22. Jianyi Yang, Ambrish Roy, and Yang Zhang. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics.* **2013**, 29:2588-2595.
 23. Ambrish Roy, Jianyi Yang, and Yang Zhang. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research.* **2012**, 40:W471-W477.
 24. Bogerd H, Greene WC. Dominant negative mutants of human T-cell leukemia virus type I Rex and human immunodeficiency virus type 1 Rev fail to multimerize in vivo. *J Virol.* **1993**, 67:2496–2502.

25. Malim MH, Bohnlein S, Hauber J, Cullen BR. Functional dissection of the HIV-1 Rev trans-activator—derivation of a trans-dominant repressor of Rev function. *Cell*. **1989**, 58:205–214.
26. Izaurralde, E; Adam, S (April 1998). Transport of macromolecules between the nucleus and the cytoplasm. *RNA*. **1998**, 4 (4): 351–364.
27. Wingfield PT, et al. HIV-1 Rev expressed in recombinant *Escherichia coli*: Purification, polymerization, and conformational properties. *Biochemistry*. **1991**, 30:7527–7534.
28. Xiao, G; Kumar, A; Li, K; Rigl, CT; Bajic, M; Davis, TM; Boykin, DW; Wilson, WD. Inhibition of the HIV-1 rev-RRE complex formation by unfused aromatic cations. *Bioorganic and Medicinal Chemistry*. **2001**, 9 (5): 1097–1113.
29. Xiao, G; Kumar, A; Li, K; Rigl, CT; Bajic, M; Davis, TM; Boykin, DW; Wilson, WD. Inhibition of the HIV-1 rev-RRE complex formation by unfused aromatic cations. *Bioorganic and Medicinal Chemistry*. **2001**, 9 (5): 1097–1113.
30. Baba, M. Inhibitors of HIV-1 gene expression and transcription. *Current Topics in Medicinal Chemistry*. **2004**, 4 (9): 871–882.
31. DeJong, ES; Chang, CE; Gilson, MK; Marino, JP. Proflavine acts as a Rev inhibitor by targeting the high-affinity Rev binding site of the Rev responsive element of HIV-1. *Biochemistry*. **2003**, 42 (26): 8035–8046.
32. Yang, J., & Zhang, Y. (2015). Protein Structure and Function Prediction Using I-TASSER. *Current Protocols in Bioinformatics* / Editorial Board, Andreas D. Baxevanis ... [et Al.]. **2015**, 52, 5.8.1–5.815.
33. Y Zhang. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. **2008**, 9: 40.
34. S. Wu, Y. Zhang. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research*. **2007**, 35: 3375-3382.
35. Y. Zhang, J. Skolnick, SPICKER: A Clustering Approach to Identify Near-Native Protein Folds. *Journal of Computational Chemistry*. **2004**, 25: 865-871.
36. Daugherty MD, D’Orso I, Frankel AD. A solution to limited genomic capacity: Using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol Cell*. **2008**, 31:824–834.
37. Battiste JL, et al. Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science*. **1996**, 273:1547–1551.
38. Pond SJ, Ridgeway WK, Robertson R, Wang J, Millar DP. HIV-1 Rev protein

- assembles on viral RNA one molecule at a time. *Proc Natl Acad Sci USA*. **2009**, 106:1404–1408.
39. Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature*. **1989**, 338:254–257.
 40. Watts NR, et al. Three-dimensional structure of HIV-1 Rev protein filaments. *J Struct Biol*. **1998**, 121:41–52.
 41. Surendran, R; Herman, P; Cheng, Z; Daly, TJ; Ching Lee, J. HIV Rev self-assembly is linked to a molten-globule to compact structural transition. *Biophysical Chemistry*. **2004**, 108 (1–3): 101–119.
 42. Fornerod, M; Ohno, M; Yoshida, M; Mattaj, IW. CRM1 is an export receptor for leucine-rich nuclear export signals". *Cell*. **1997**, 90 (6): 1051–1060.
 43. DiMattia, M. A., Watts, N. R., Stahl, S. J., Rader, C., Wingfield, P. T., Stuart, D. I., ... Grimes, J. M. Implications of the HIV-1 Rev dimer structure at 3.2 Å resolution for multimeric binding to the Rev response element. *Proceedings of the National Academy of Sciences of the United States of America*. **2010**, 107(13), 5810–5814.