Structural Modelling

ceRNA Network

Circular RNA (hsa_circ_0000064) Structural Modelling

We have used Vienna RNA package[1] for generating structural models of circular RNA hsa_circ_0000064. RNA sequence was retrieved from circ-Interactome database[2] to be used as an input for Vienna package. Vienna RNA package depends on extension of linear folding algorithms. Circular RNA molecules are modelled through post-processing of computed linear arrays. Using Vienna RNA Package, we could compare structural modifications between linear and circular structures in a memory-effective manner.[3] The energy contribution of Exterior loop should be scored in circular structures, on the other hand, exterior loops have no energy contribution in linear structures. RNAfold structure prediction tool was used to calculate the minimum free energy (MFE) and backtraces an optimal secondary structure. To compute centroid structure we used McCaskill's algorithm[4] through -p option. Mountain plot was produced with mountain.pl, Dot plot was also generated by RNAplfold.

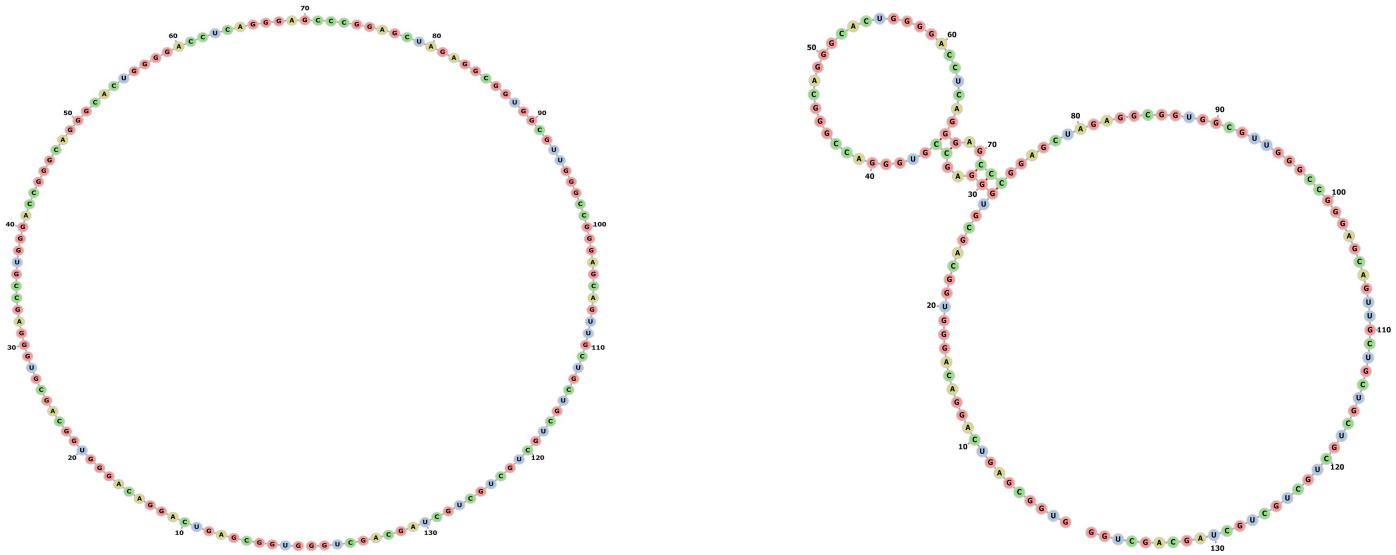

Figure-1 RNAFOLD simulation of circular RNA structure

Minimum free energy prediction using RNAFOLD generated an optimal secondary structure in dot-bracket notation from a centroid structure of 0.00 kcal/mol minimum free energy to 1.78 kcal/mol. Thermodynamic ensemble prediction using RNAFOLD computed a free energy of -51.72 kcal/mol, The frequency of the MFE structure in the ensemble is 0.07 % and the ensemble diversity is 65.27.

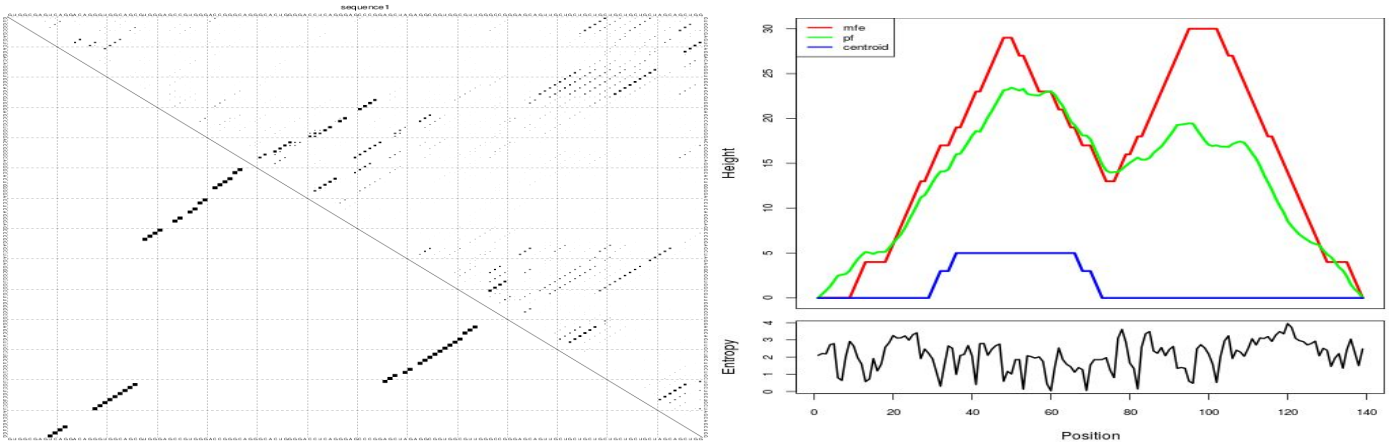

Figure-2 (left) energy dot-plot of circRNA model, (Right) Mountain plot of the same model

To the left is the The energy dot 2D plot which indicates all of the base pairs involved in optimal and suboptimal secondary structures, both axes of the graph represent the same RNA sequence. Each point drawn indicates a base pair between the ribonucleotides whose positions in the sequence are the coordinates of that point on the graph.
To the right is the Mountain plot plotting the number of base pairs enclosing a sequence position versus the position. The plot includes the MFE structure, the thermodynamic ensemble of RNA structure, and the centroid structure. Additionally we used it to estimate the positional entropy for each position.

miRNA mir-1825 Structural Modelling

We have used SimRNA [5] Tool for simulating circular structure of miRNA mir-1825 as SimRNA generates a circular starting conformation with the 5' and 3' ends close to each other as a starting structure for simulation. After specifying Secondary structure restraints using multiline dots-and-brackets format, The dots-and-brackets input is parsed and internally converted into the dedicated list of restraints. W used a default of 500 steps and 1% of the lowest energy frames taken to clustering.

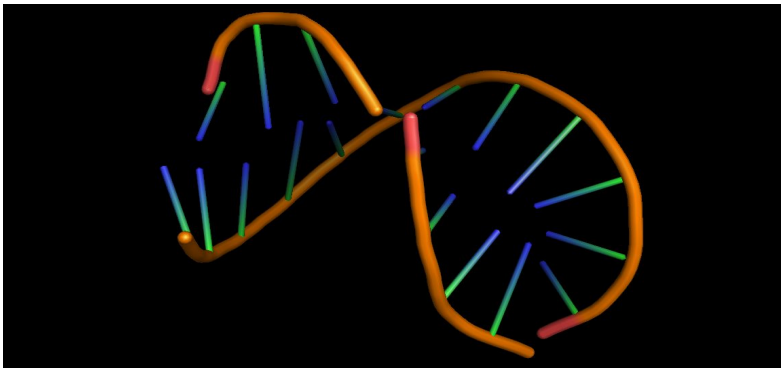

Figure-3 Best Secondary Structure Cluster Predicted by SimRNA for miRNA mir-1825

RNA Interaction Modeling

IntaRNA[6] is a program for fast and accurate prediction of interactions between two RNA molecules. It has been used to predict mRNA circRNA sites, to represent the interaction energy in the RNA Sponge.

Table-1 IntaRNA interaction energy prediction of RNA sponge

circRNA	Position	miRNA	Position	Energy
hsa_circ_0000064	44-58	hsa-mir-1825	1-15	-15.21740

Energy	-15.21740 kcal/mol
Hybridization Energy	-22.2
Unfolding Energy - circRNA	6.43246
Unfolding Energy - miRNA	0.53411
Position - circRNA RNA	44 -- 58
Position - miRNA RNA	1 -- 15
Position Seed - circRNA RNA	52 -- 58
Position Seed - miRNA RNA	1 -- 7

Figure-4 RNA sponge graphical representation of circRNA hsa_circ_000004 and miRNA hsa-mir-1825 in vienna format

Position-wise minimal energy profile

The following plots give us insights into the overall circRNA-miRNA interaction abundance. a minimal energy profile is provided for both sequences of both miRNA and circRNA, taking RNA-RNA interaction in consideration.

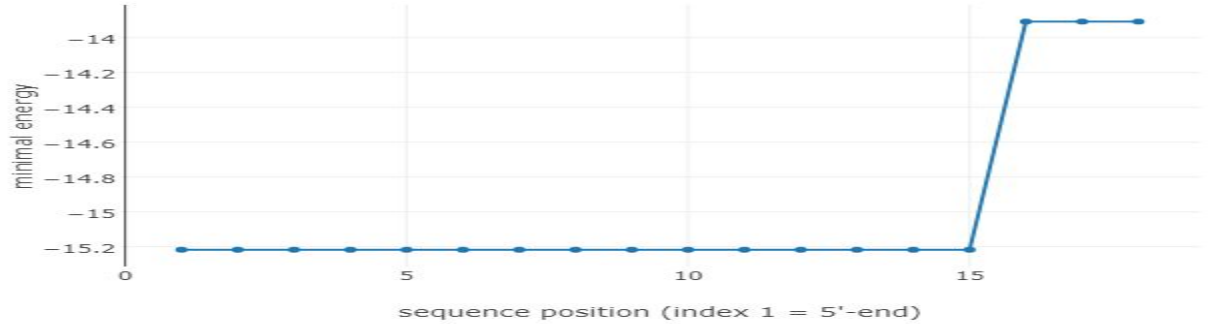

Figure-5 Minimal energy profile is provided for miRNA sequence

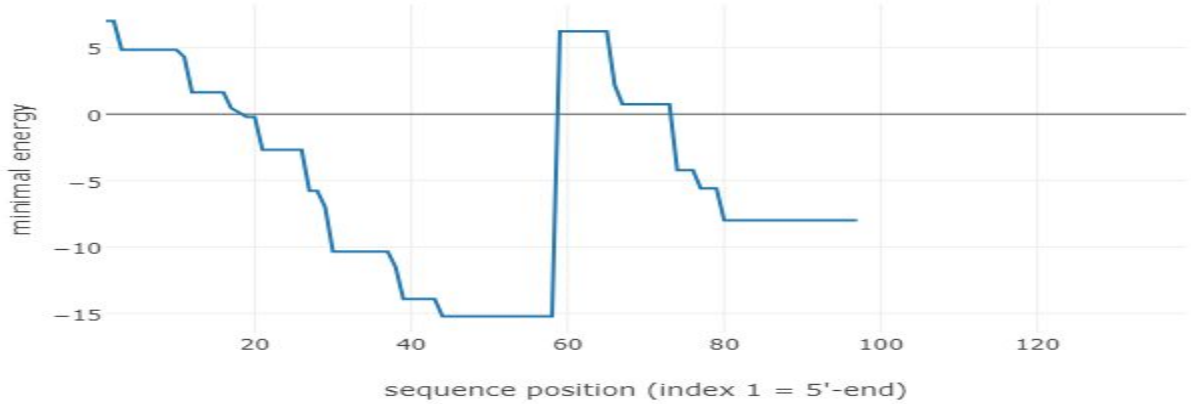

Figure-6 Minimal energy profile is provided for Circular RNA sequence representing sponge interaction.

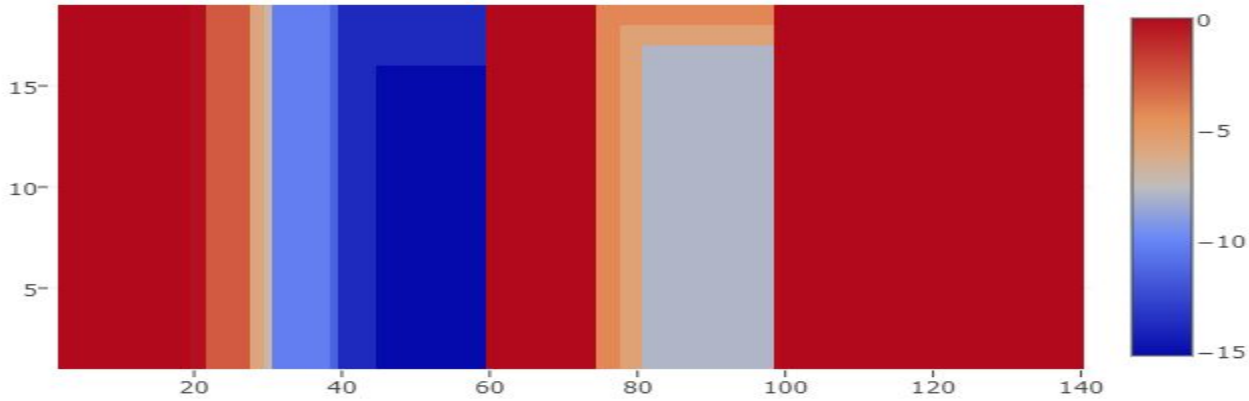

Figure-7 Heatmap visualization of the minimal energy for each non-coding RNA in the sponge

CRISPR Circuit Structural Modelling

Cas9 Modelling

Cas9 of S. pyogenes (BBa_K1218011) part was translated and modelled by SWISS Model server[7,8] Using SWISS-Model web server the modelling process was initiated by template recognition process where templates were selected according to the maximum sequence similarity 5FQ5 was of highest sequence identity (Sequence identity: 100.00), Finally, the geometry of the resulting model is regularized by using a force field. The global and per-residue model quality has been assessed using the QMEAN scoring function. For improved performance, weights of the individual QMEAN terms have been trained specifically for SWISS-MODEL.Models were selected based on their sequence identity as well as Swiss-MODEL quality assessment parameters GMQE and QMEAN4.

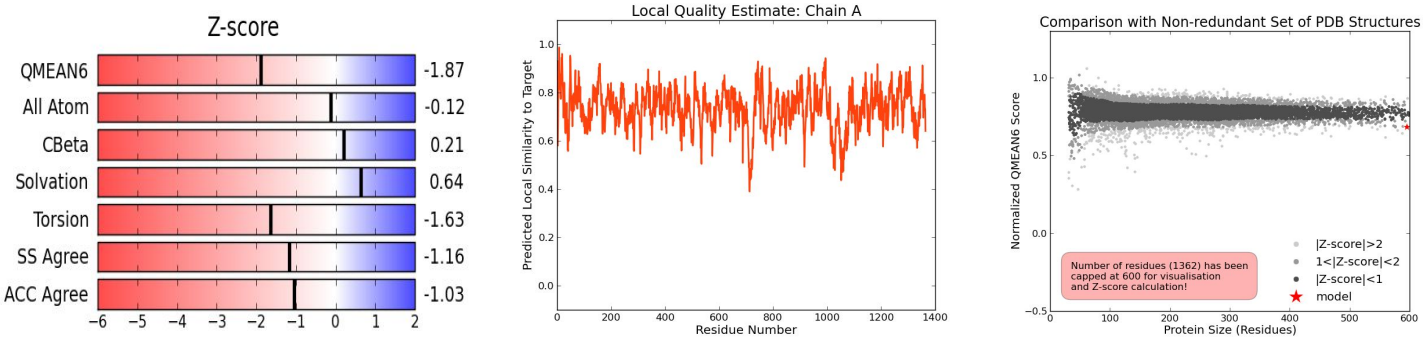

Figure-8 Quality estimation plots based on SWISS-MODEL parameters for Cas9

Nucleic acid Modelling

The most stable 2D structure of gRNA was generated using vfold[9]. The Rosetta package FARFAR[10,11] was used to build the 3D structure of gRNA, 3D-DART[12] was used to generate a 3D structure of the target DNA representing the cleavage site and PAM of cas9, while PAM flexibility was studied using Naflex[13].

Docking protocol

Following HADDOCK[14,15] docking protocol, consisting of randomized orientations and rigid body energy minimization, we have calculated 1,000 complex structures. The 200 complexes with the lowest intermolecular energies have been selected for semi-flexible simulated annealing in torsion angle space. The resulting structures have been then refined in explicit water. Finally, the solutions have been clustered using a threshold value of 1.5 A˚ for the pairwise backbone RMSD at the interface, and the

resulting clusters have been ranked according to their average interaction energy (defined as the sum of van der Waals, electrostatic and AIRs energy terms) as well as buried surface area. HADDOCK scoring is performed according to the weighted sum (HADDOCK score) of different energy terms which includes van der Waals energy, electrostatic energy, distance restraints energy, direct RDC restraint energy, intervector projection angle restraints energy, diffusion anisotropy energy, dihedral angle restraints energy, symmetry restraints energy, binding energy, desolvation energy and buried surface area. One lowest energy structure of the lowest intermolecular energy cluster was selected for analysis. This lowest energy structure displayed no AIR restraint violations within 0.3 A˚ threshold and was accepted as the final docked structure for the complex.

Table-2 Docking scores of both complexes using HADDOCK

Complex	Cas9.gRNA Cluster	Cas9.gRNA:Target DNA Cluster
HADDOCK score	-75.64 +/- 22.9	-340.7 +/- 23.9
clusters	13	17
Total Interaction energy(Kcal mol-1)	1.0 +/- 0.6	6.3 +/- 2.1
Van der Waals energy(Kcal mol-1)	-66.7 +/- 19.5	-97.3 +/- 14.1
Electrostatic Energy(Kcal mol-1)	-284.4 +/- 21.7	-342.0 +/- 73.6
Desolvation energy(Kcal mol-1)	20.6 +/- 4.0	-201.2 +/- 14.6
Restraints violation energy(Kcal mol-1)	273.9 +/- 32.63	262.1 +/- 60.51
Buried Surface area	1912.6 +/- 303.2	2785.3 +/- 168.1
z-score	-2	-2

For Cas9.gRNA docking, HADDOCK clustered 105 structures in 13 cluster(s), which represents 52.5 % of the water-refined models that were generated by HADDOCK. Note that currently the maximum number of models considered for clustering is 200. While for Cas9.gRNA:Target DNA docking, HADDOCK clustered 107 structures in 12 cluster(s), which represents 53.5 % of the water-refined models were generated by HADDOCK.

Results analysis

The results and graphics presented below are based on water-refined models generated by HADDOCK.

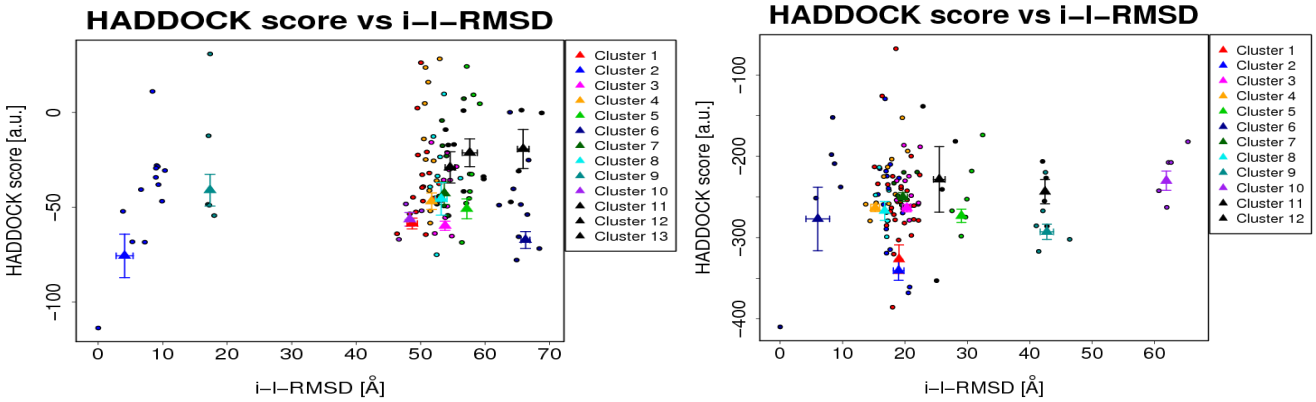

Figure-9 The clusters (indicated in color in the graphs) are calculated based on the interface-ligand RMSDs calculated by HADDOCK, with the interface defined automatically based on all observed contacts to the left is Cas9.gRNA Cluster and to the right is Cas9.gRNA:Target Cluster of lowest energy

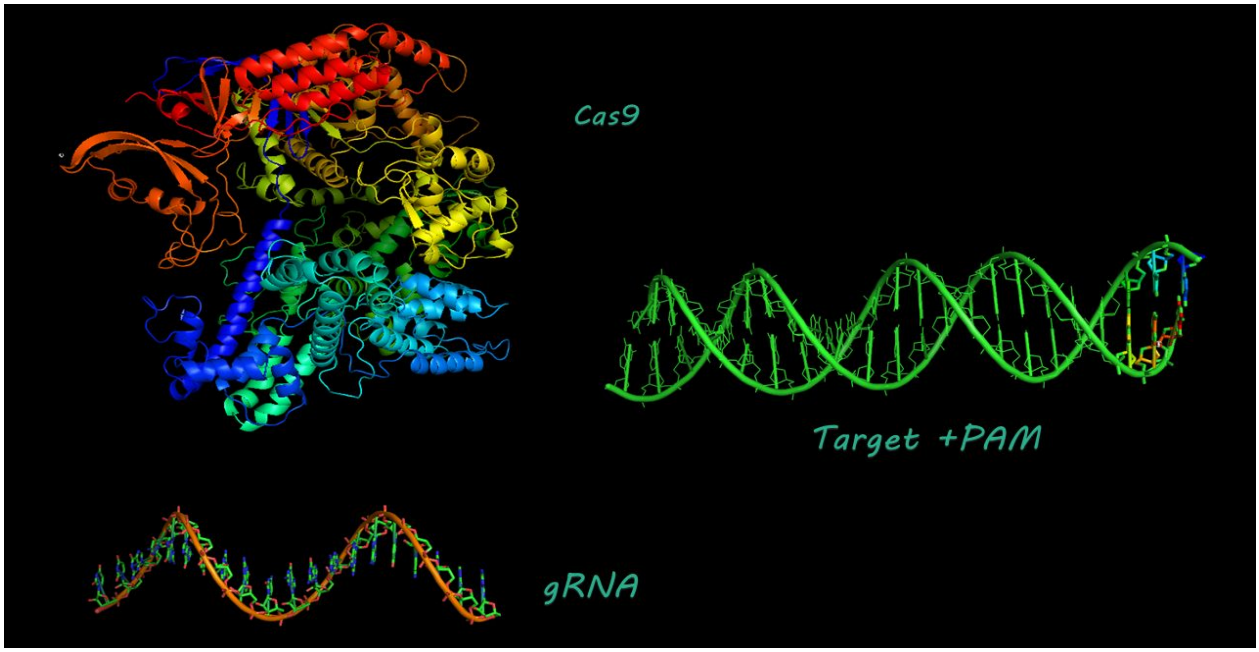

Figure-10 PyMol visualization of Modelled Structures of cas9, gRNA and Target DNA+PAM

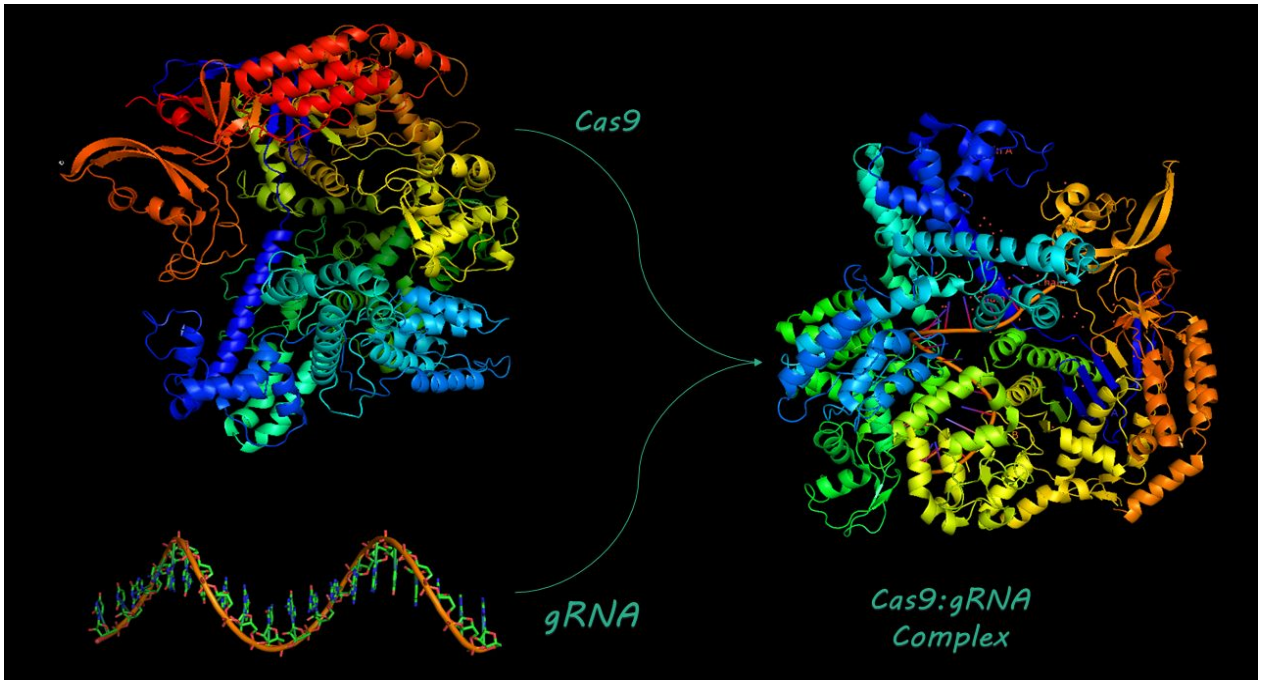

Figure-11 PyMol visualization of Modelled Structures of cas9, gRNA and cas9.gRNA docked complex

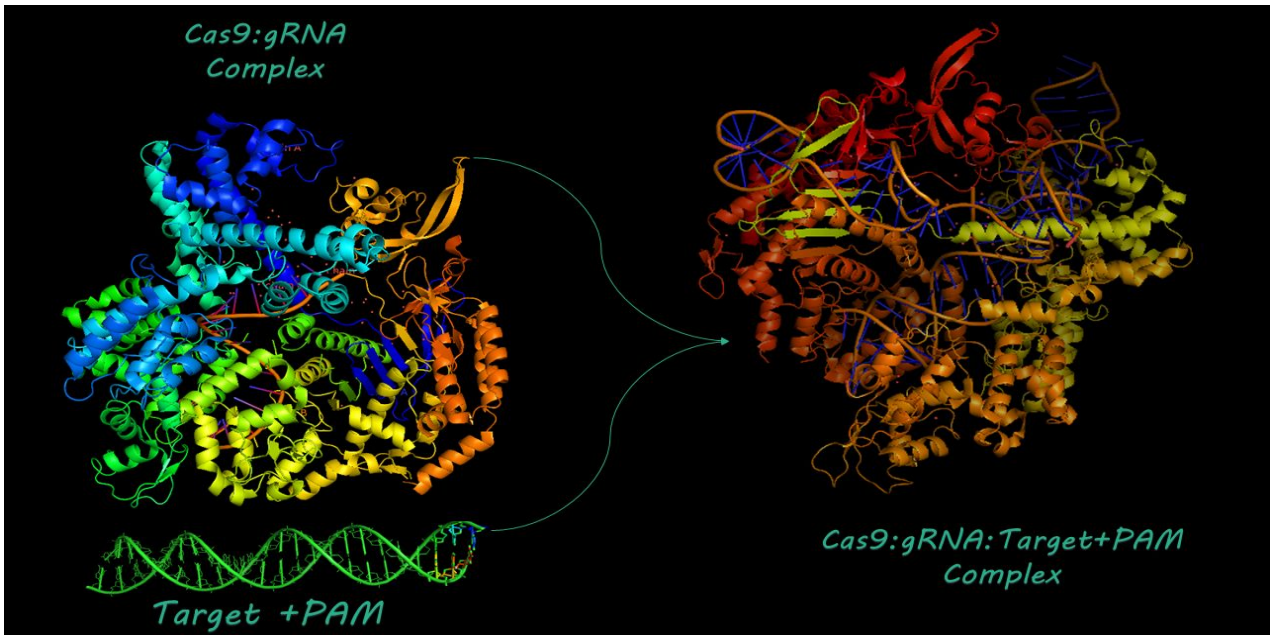

Figure-12 PyMol visualization for cas9:gRNA docked to Target DNA to identify cas9 cleavage of target DNA

References

[1] Lorenz, R. and Bernhart, S.H. and Höner zu Siederdissen, C. and Tafer, H. and Flamm, C. and Stadler, P.F. and Hofacker, I.L. "ViennaRNA Package 2.0", Algorithms for Molecular Biology, 6:1 page(s): 26, 2011

[2] Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. RNA Biology. 2016;13(1):34-42. doi:10.1080/15476286.2015.1128065.

[3] Hofacker I L, Stadler P F. Memory efficient folding algorithms for circular RNA secondary structures.Bioinformatics, 2006, 22: 1172-1176

[4] McCaskill J. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers. 1990;29:1105–1119.

[5] Boniecki MJ, Lach G, Dawson WK, et al. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Research*. 2016;44(7):e63. doi:10.1093/nar/gkv1479.

[6] Anke Busch, Andreas S. Richter, and Rolf Backofen, IntaRNA: efficient prediction of bacterial sRNA circRNAs incorporating circRNA site accessibility and seed regions, Bioinformatics, 2008, 24 no. 24 pp. 2849-56.

[7] Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics, 22, 195-201.

[8] Benkert, P., Kunzli, M. and Schwede, T. (2009) QMEAN server for protein model quality estimation. Nucleic Acids Res, 37, W510-514.

[9] Xu, XJ., Zhao, PN., Chen, SJ. (2014) Vfold: a web server for RNA structure and folding thermodynamics prediction.

[10] Das, R., Karanicolas, J., and Baker, D. (2010), "Atomic accuracy in predicting and designing noncanonical RNA structure". Nature Methods 7:291-294.

[11] Lyskov S, Chou FC, Conchúir SÓ, Der BS, Drew K, Kuroda D, et al. "Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE)". PLoS One. 2013 May 22;8(5):e63906

[12] Van Dijk M, Bonvin AMJJ. 3D-DART: a DNA structure modelling server. Nucleic Acids Research. 2009;37(Web Server issue):W235-W239.

[13] Hospital A, Faustino I, Collepardo-Guevara R, González C, Gelpí JL, Orozco M. NAFlex: a web server for the study of nucleic acid flexibility. Nucleic Acids Research. 2013;41(Web Server issue):W47-W55.

[14] Cyril Dominguez, Rolf Boelens and Alexandre M.J.J. Bonvin. HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. J. Am. Chem. Soc. 125, 1731-1737 (2003).

[15] G.C.P van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries and A.M.J.J. Bonvin. "The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes."J. Mol. Biol., 428, 720-725 (2016).