Computations Lab -3   Vicky (B21BB033)

Class Exercise 1

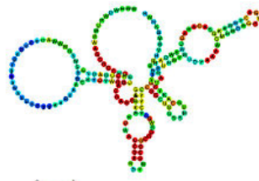Retrieving all information regarding **Riboswitch** from **RFAM**

Query: ribozyme AND rna_type:"riboswitch"

1. Name the family and RFAM ID

Ans: **Family: *glmS and RFAM id:* RF00234**

  Description: *glmS glucosamine-6-phosphate activated ribozyme aptamer.*

glmS glucosamine-6-phosphate activated ribozyme

Predicted secondary structure and sequence conservation of glmS

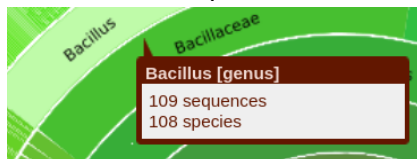2. How many species altogether are included in this family?

Ans: *1332* species

3. Where they are predominantly found, Eukaryotes or prokaryotes?

Ans: Mainly Procaryotes(Bacteria) But around 20% Eucaryotes

4. Are they present in Bacillus? How many?

Ans: Yes . 108 Species and 109 sequences.

Bacillus   Bacillaceae

Bacillus [genus]
109 sequences
108 species

5. In the evolutionary tree, how many clades do you see? Are they related? Which method was used to generate the tree?

Ans: 14 clades.A clade (also known as a monophyletic group) is a group of organisms that includes a single ancestor and all of its descendents.

This tree was built using the *fasttree* method.

This tree was built using the *fasttree* method.

Label the trees using ◉ species names or ◯ accessions.
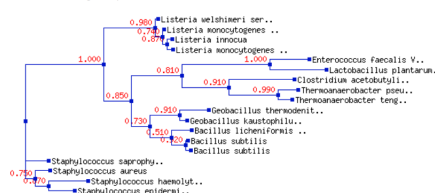
6. Change the tree-building method and see whether you get the same results or not.
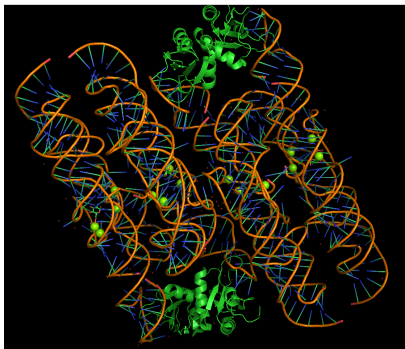Ans. Yes, I got the same results

7. How many structures are known to have this domain, belonging to the said family?
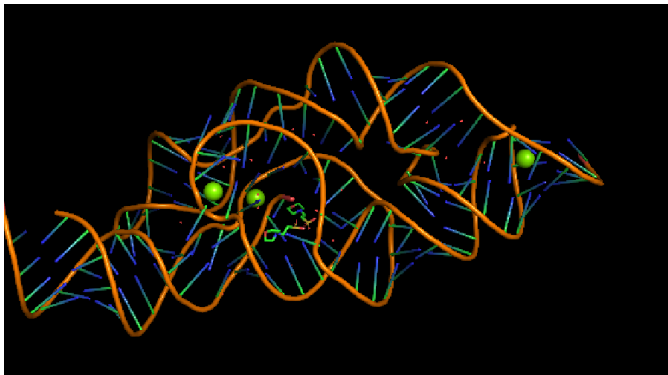Ans: 39 structures
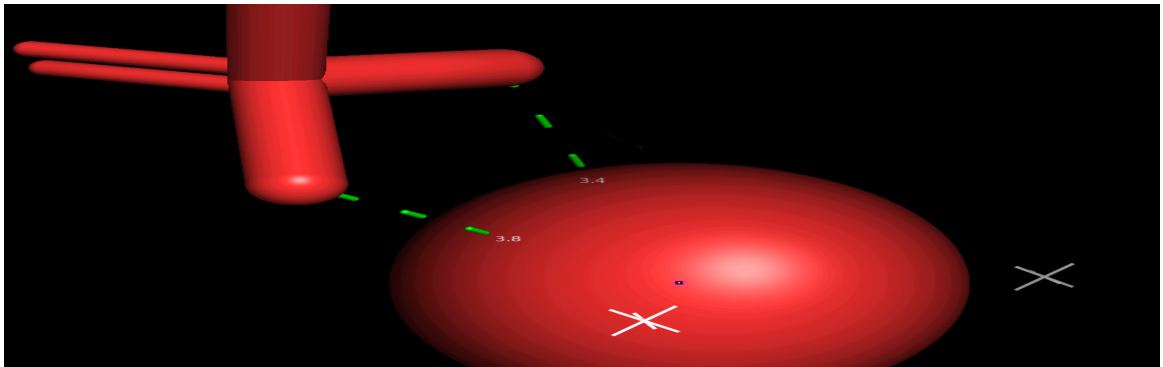8. Visualize in Pymol (any two structures), how many RNA molecules are there?
Ans: 2NZ4: one RNA molecule (U1 Small Nuclear Ribonucleoprotein A) and four chains A, J, K, L.

2HO6:It has two molecules with one chain each

9. Highlight the active site, is it ligand/ activator bound or unbound structure?
Ans: 2NZ4:

10. Are there any motifs which matches with this riboswitch? What are they?
Ans:

Motif Accession	Motif Description	Number of Hits	Fraction of Hits	Sum of Bits	Image
RM00004	CRC binding motif	2	0.111	21.0	🖼
RM00008	GNRA tetraloop	11	0.611	125.7	🖼

11. Is this motif seen in any other RNAs? Give example
Ans: YES, RM00008 has a structure 1G1X. STRUCTURE OF RIBOSOMAL PROTEINS S15, S6, S18, AND 16S RIBOSOMAL RNA.

12. What are the molecular functions known associated with this riboswitch? (refer to GO)
Ans:
Catalysis of a biochemical reaction at physiological temperatures. In biologically catalyzed reactions, the reactants are known as substrates, and the catalysts are naturally occurring macromolecular substances known as enzymes. Enzymes possess specific binding sites for substrates, and are usually composed wholly or largely of protein, but RNA that has catalytic activity (ribozyme) is often also regarded as enzymatic

13. How many sequences are used in the full alignment, for this particular riboswitch?
Ans: 4

14. Can you comment on the disease associations of this riboswitch (if any)?
Ans:

Class Exercise 2

Get this information on "**piwi domain**" from **PFAM**

Query: piwi domain

1. Name the PFAM ID and the clan
   Ans. PFAM ID: PF02171

   Clan: RNase_H

2. How many proteins have experimentally solved structures containing this domain?
   Ans. 109 domains

3. Name one such protein and the source organism, and PDB ID
   Ans. https://www.ebi.ac.uk/interpro/structure/PDB/1u04/

4. How many proteins have a domain architecture with only one piwi domain?
   Ans. 427 proteins

727 domain architectures

There are 10339 proteins with this architecture (represented by P34681):
PF16486 · PF08699 · PF02170 · PF16488 · PF16487 · PF02171

ArgoN
ArgoL1
PAZ
ArgoL2
ArgoMid
Piwi

There are 6304 proteins with this architecture (represented by O28951):
PF02171

Piwi

There are 3414 proteins with this architecture (represented by Q9SDG8):
PF16486 · PF08699 · PF02170 · PF16488 · PF02171

ArgoN
ArgoL1
PAZ
ArgoL2
Piwi

There are 2207 proteins with this architecture (represented by Q21691):
PF02170 · PF02171

PAZ

5. Predominantly, in which organisms are these proteins mostly found?
   Ans.

6. How big is the piwi domain (length)?
   Ans.
   427 amino acids

7. How many proteins have a domain architecture with one piwi and one "PAZ" domain?

Ans.

8. Name one such protein, and name the source organism.
   Ans. Nuclear RNAi defective-3 protein - Caenorhabditis elegans
9. What is the primary function of this protein?
   Ans. Transports small interfering RNAs (siRNAs) from the cytoplasm to the nucleus. Required for RNA interference (RNAi) in nuclei (PubMed:18653886, PubMed:34365510). Required for exogenous RNAi-induced H3K27 methylation
10. Are there any structures available? name them
    Ans.

Source organism	Caenorhabditis elegans go to search ⏷
UniProt	Q21691 go to UniProt ⏷
Experimental structures	None available in the PDB
Biological function	Transports small interfering RNAs (siRNAs) from the cytoplasm to the nucleus. Required for RNA interference (RNAi) in nuclei (PubMed:18653886, PubMed:34365510). Required for exogenous RNAi-induced H3K27 methylation (PubMed:26365259). go to UniProt ⏷

Structure viewer

Model Confidence ⓘ

■ Very high (pLDDT > 90) ■ High (90 > pLDDT > 70) ■ Low (70 > pLDDT > 50) ■ Very low (pLDDT < 50)

AlphaFold produces a per-residue model confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Predicted aligned error (PAE)

Click and drag a box on the PAE viewer to select regions of the structure and highlight them on the 3D viewer.

PAE data is useful for assessing inter-domain accuracy – go to Help section below for more information.