COMP 564: Introduction to RNA 3D structures

Jérôme Waldispühl
School of Computer Science
McGill University
Motivations and challenges

- Secondary structure is a coarse-grained model.
- Function is achieved through the 3D structure.
- Experimental determination the RNA 3D structure is hard.
- Modeling the 3D structure is also hard!
- Before the prediction, a work has to be done on modeling and alignment of 3D structure.
Beyond the secondary structure

The type of interactions is not restricted to Watson-Crick base pairs:

SECIS element

Non-canonical interactions

NB: non-stacking interaction

5S ribosomal RNA
Classification of non Watson-Crick base pair interactions

What are we seeing when looking at the 3D structure?

“Loops” are not loops!

Sites for non Watson-Crick base pairs.
Classification of non Watson-Crick base pair interactions

Modeling the nucleotide side-chain with interacting edges
Classification of non Watson-Crick base pair interactions

**Consequence:** 3 edges available for base-pairing.
Classification of non Watson-Crick base pair interactions

Orientation of edge interaction is also important: The glycosidic bond orientation.

Cys (default):

Trans:
Classification of non Watson-Crick base pair interactions

12 edge-to-edge interacting motifs
Classification of interactions

But the picture is still far to be complete!

Base interacting with all 3 edges
Classification of interactions

The interacting motif is extended to model base triple.
More Features…

Base-Sugar conformation.

Anti (default):

Syn (Purines only):
More features...

Local strand orientation:

- Anti-parallel (default)
- Parallel

Locally parallel strands:
New symbols

- Indicates Base Stacking

- Indicates Change in Strand Orientation

- Indicates syn conformation for base
Example: 5S motif
Example: 5S motif

Loop A

Loop B

Loop C

Loop D

Loop E

E. coli
Bacteria

3' - 5'
G = C
104 A □ G 72
103 U □ A 73
102 G □ U 74
101 A □ G 75
100 G □ G 76
99 A □ U 77
98 G □ A 78
C = G
5' - 3'
More features (2)...

Superposition of tetra and penta GNRA loops:

Interaction of GNRA loops are also conserved:

23S H. marismortui  23S T. thermophilus
RNA 3D motifs

Non canonical base pairs are primarily found in loops

TPP riboswitch (Serganov et al., 2006). Image by Chimera.

Base pair networks provide a strong signature of 3D structures

IL_51265 from RNA 3D Motif Atlas. Images by VARNA & Chimera.

We search for recurrent graphs in RNA structures
Finding RNA motifs in 3D structures

Q: Given a description of a “known” motif, how to identify this motif in target structures?

Use graph theory, the problem of identifying a known pattern in a target graph reduces to the following:

1. Searching for isomorphic occurrences of the pattern (subgraph isomorphism).

2. Finding similar occurrences of the pattern (identifying a maximum common subgraph).

But it’s NP-complete…
3D Motif databases

- Rna3Dмотif (Djelloul & Denise, 2008)
- FRABASE (Popenda et al., 2010)
- RNA 3D Motif Atlas (Petrov et al., 2013)
- RNAbricks (Chojnowski et al., 2014)
- Rfam (Kalvari et al., 2017)
- CaRNAval (Reinharz et al., 2018)
- VeRNAL (Oliver et al., 2021)
FR3D: Find RNA 3D
(Sarver et al., 2008)
Leontis + Zirbel groups

Find small RNA motifs (two to 20 nucleotides) in PDB files
FR3D example: C-loop search

Output:

<table>
<thead>
<tr>
<th>Filename</th>
<th>Discrepancy from query</th>
<th>Motif Nucleotides</th>
<th>Pairwise Interactions</th>
<th>Structural Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2AW4</td>
<td>0.000</td>
<td>U 2680 C 2681 C 2683 U 2684 A 2725 A 2727</td>
<td>s35 cWW tWH s35 cWS cWW s35</td>
<td>UCA-CU ... AA-A</td>
</tr>
<tr>
<td>1s72</td>
<td>0.127</td>
<td>C 2717 C 2718 C 2720 U 2721 A 2761 G 2763</td>
<td>s35 cWW tWH s35 cWS cWW s35</td>
<td>CCA-CU ... AC-G</td>
</tr>
<tr>
<td>1kog</td>
<td>0.136</td>
<td>C 96 C 97 C 99 U 100 A 74 G 76</td>
<td>s35 cWW tWH s35 cWS cWW s35</td>
<td>CCA-CU ... AU-G</td>
</tr>
<tr>
<td>2p01</td>
<td>0.229</td>
<td>C 1319 C 1320 A 1322 U 1323 A 1331 C 1333</td>
<td>s35 cWW tWH s35 ncWS ncWW s35</td>
<td>GCA-AU ... AG-C</td>
</tr>
<tr>
<td>2AW4</td>
<td>0.232</td>
<td>C 1319 C 1320 A 1322 C 1323 G 1331 G 1333</td>
<td>s35 cWW tWH s35 ncWS ncWW s35</td>
<td>CCA-AC ... GG-G</td>
</tr>
<tr>
<td>2AW4</td>
<td>0.244</td>
<td>C 864 C 865 C 867 U 868 A 909 C 912</td>
<td>s35 cWW tWH s35 ncWS ncWW s35</td>
<td>GCA-CU ... AAAC</td>
</tr>
<tr>
<td>1s72</td>
<td>0.256</td>
<td>G 1425 C 1426 C 1428 U 1429 A 1437 C 1439</td>
<td>s35 cWW tWH s35 cWS cWW s35</td>
<td>GCA-CU ... AG-C</td>
</tr>
<tr>
<td>2p01</td>
<td>0.278</td>
<td>G 864 C 865 C 867 U 868 A 909 C 912</td>
<td>s35 cWW tWH s35 ncWS ncWW s35</td>
<td>GCA-CU ... AAAC</td>
</tr>
<tr>
<td>156c</td>
<td>0.380</td>
<td>G 371 C 372 A 374 U 375 A 389 C 390</td>
<td>s35 cWW tWH s35 cWS cWW s35</td>
<td>GCA-AU ... AA-C</td>
</tr>
<tr>
<td>1s72</td>
<td>0.402</td>
<td>G 958 C 959 C 962 C 963 A 1005 C 1008</td>
<td>s35 cWW tWH s35 cWS cWW s35</td>
<td>GCGACC ... AAAC</td>
</tr>
<tr>
<td>2AVY</td>
<td>0.415</td>
<td>A 371 C 372 A 374 U 375 A 389 U 390</td>
<td>s35 cWW ntWH s35 cWS cWW s35</td>
<td>ACA-AU ... A--U</td>
</tr>
</tbody>
</table>
What do we learn?

- Positions of insertions/deletions
- Base-pair co-variations
- Base conservations
- Problem: Limited number of examples

More at http://rna.bgsu.edu/rna3dhub/
RNA3Dmotif (Djelloul & Denise, 2008)

How to identify “unknown” motifs?

1. Identify secondary structure elements in 3D structure;

   Rationale: motifs are “often embedded within helical regions forming internal loops but may also comprise hairpin or junction loops.”

2. Calculate a similarity for each pair of structural elements;

   Rationale: Find the largest common sub-graphs

3. Cluster the graphs using the similarity measure.
Recurrent motifs found in ribosomal structures.

Known motifs

(C) C-loop
(K) K-turn
(S) Sarcin–ricin
(H) Hook–turn
(A) A–minor

(E) E–loop
(F) E–loop
(G) E–loop
(R) Reverse–Kturn
(T) Tandem sheared

Unknown motifs

(B)
(D)
(I)
(J)

Djelloul M, Denise A RNA 2008;14:2489-2497
Crystal structures of four putative new motifs superimposed.

Djelloul M, Denise A RNA 2008;14:2489-2497
Long-range interactions

3',3'-cGAMP riboswitch (Ren et al., 2015)

Images from https://en.wikipedia.org/wiki/Nucleic_acid_tertiary_structure
CaRNAval pipeline

1) Extract secondary structures

2) Extract all interaction graphs

3) Find maximal interaction networks

4) Cluster the recurrent interaction networks
Hierarchy of RNA Interaction Networks
VeRNAl: Mining for Fuzzy Base Pair Motifs

**Approach:** Clustering motifs in the embedding space.

(Oliver et al., 2021)
Results: Fuzzy motifs

Classes of motif automatically clustered with VeRNAI

(Oliver et al., 2021)