Yet another definition of a long-range RNA tertiary motif

Eugene Baulin
Laboratory of Bioinformatics and Protein Engineering (prof. Janusz M. Bujnicki)
International Institute of Molecular and Cell Biology in Warsaw

Benasque, 18 August 2022

Yet another definition of a long-range RNA tertiary ~~motif~~ module

---

Eugene Baulin
Laboratory of Bioinformatics and Protein Engineering (prof. Janusz M. Bujnicki)
International Institute of Molecular and Cell Biology in Warsaw

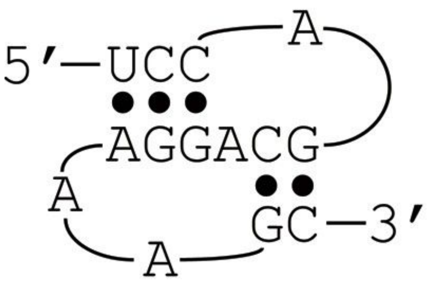
Benasque, 18 August 2022

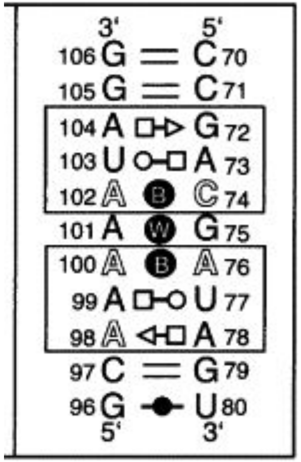
RNA tertiary structure

2D ≤ X ≤ 3D

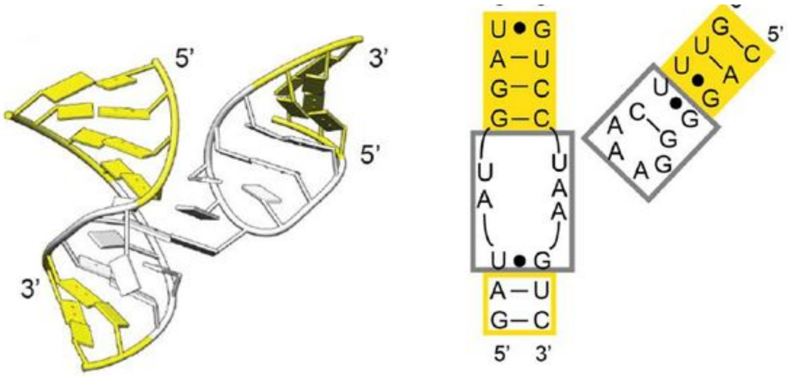
2D ≤ X ≤ **2.5D**

2.5D ≤ X ≤ 3D





- Pseudoknots
- “Nested” non-canonical interactions

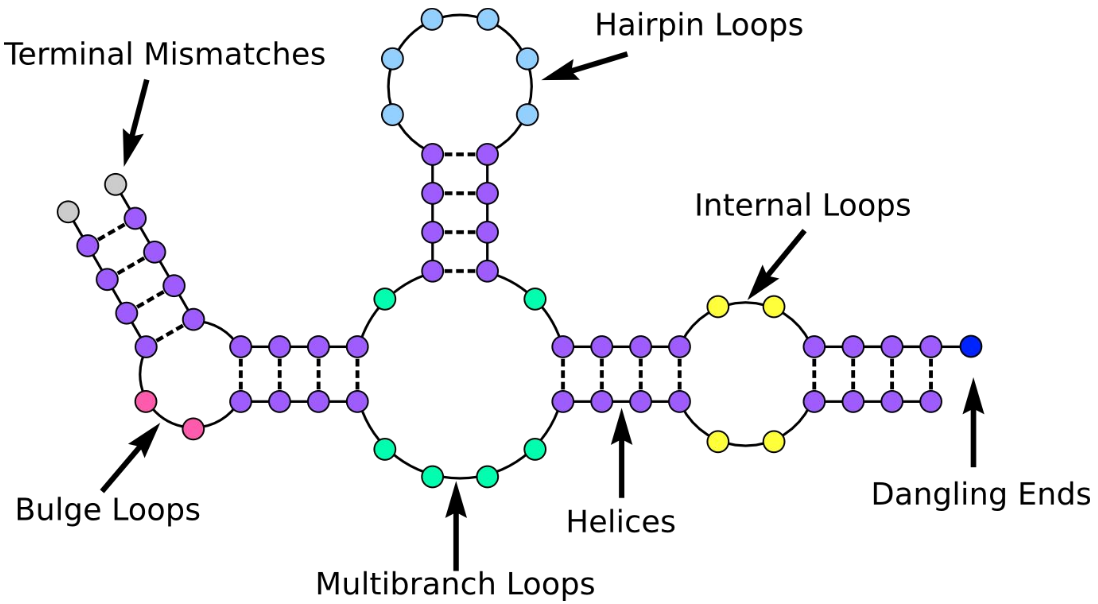


GAAA/11nt

Local tertiary motifs

2D ≤ X ≤ 3D

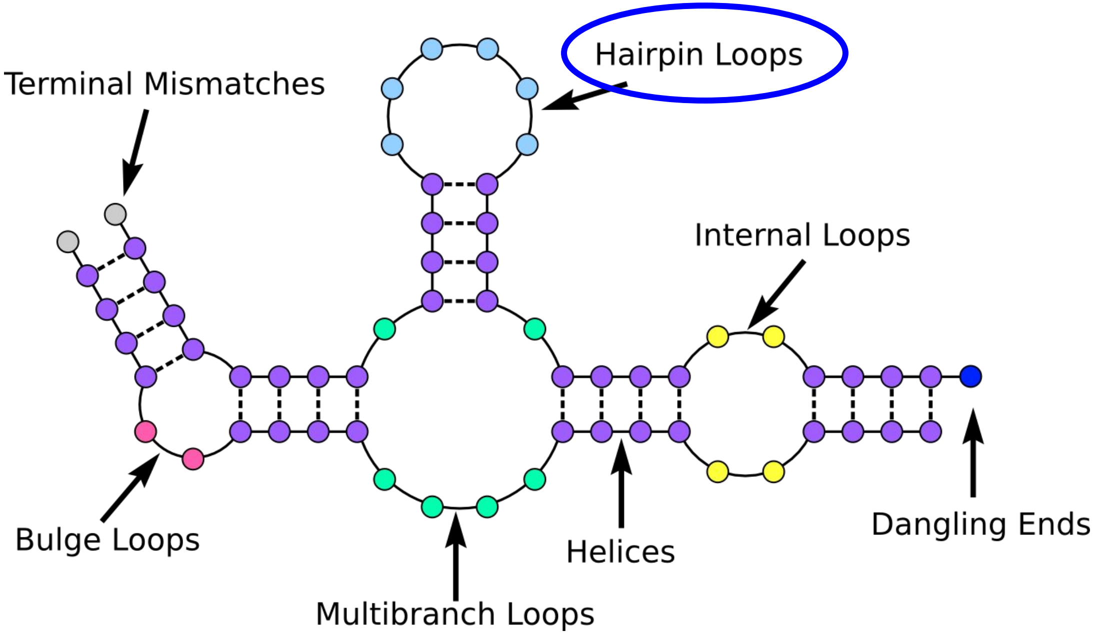
2D ≤ X ≤ **2.5D**

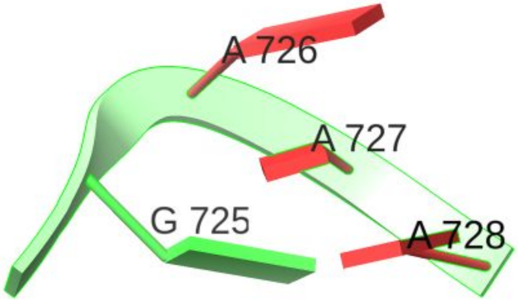


Local tertiary motifs

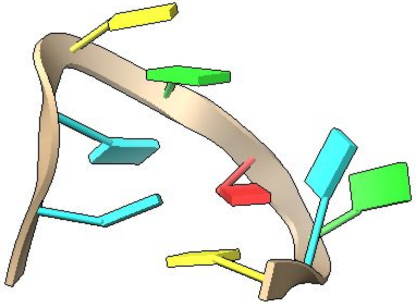
2D ≤ X ≤ 3D

2D ≤ X ≤ **2.5D**





GAAA tetraloop

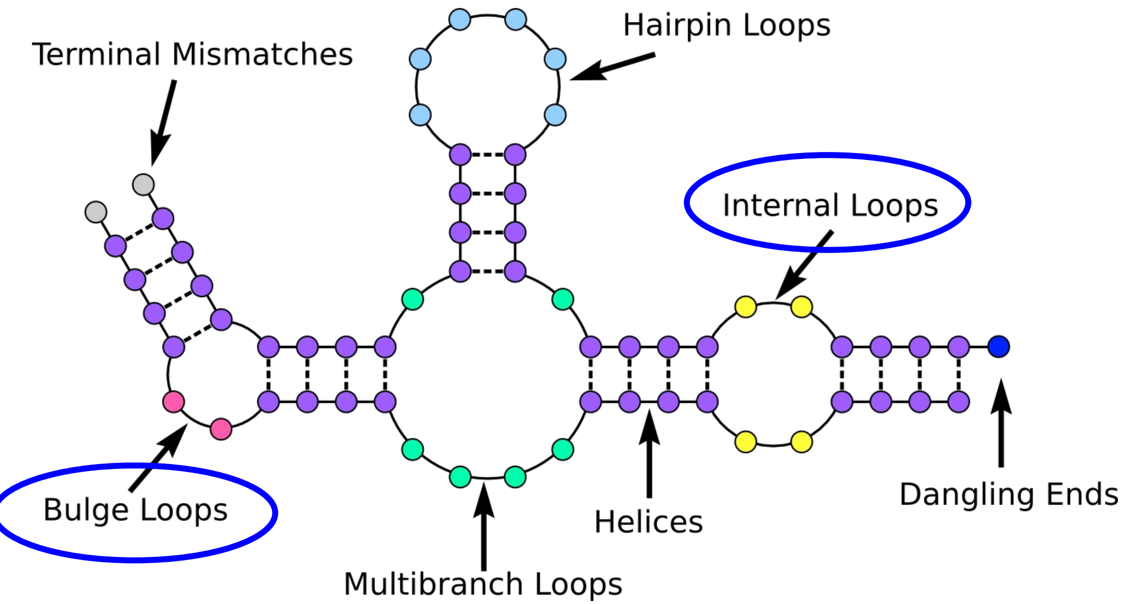


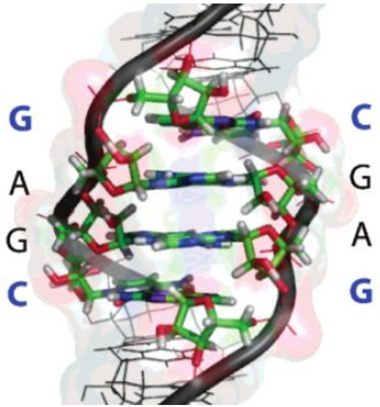
T-loop

Local tertiary motifs

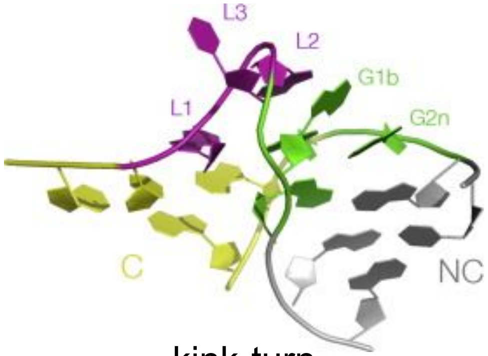
2D ≤ X ≤ 3D

2D ≤ X ≤ **2.5D**





tandem-sheared-GA

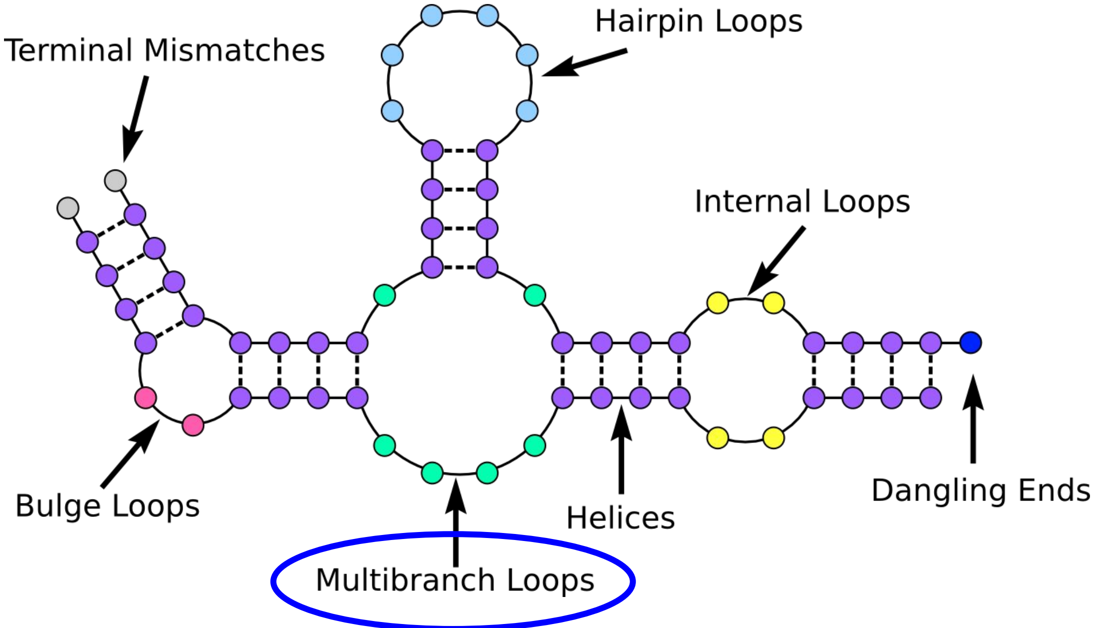


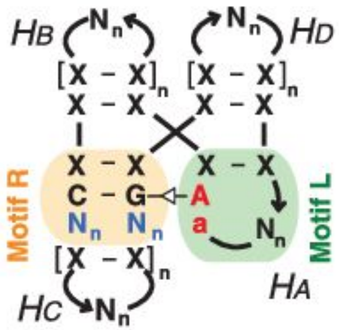
kink-turn

Local tertiary motifs

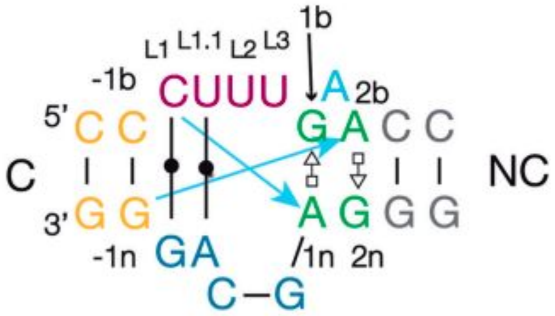
2D ≤ X ≤ 3D

2D ≤ X ≤ **2.5D**





A-minor junction



k-junction

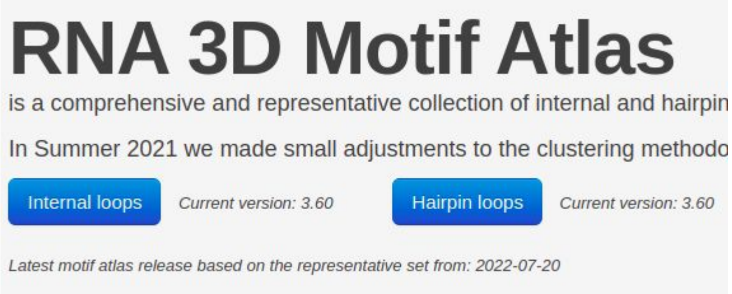
Local tertiary motifs

2D ≤ X ≤ 3D

2D ≤ X ≤ **2.5D**

RNA CoSSMos
Characterization of Secondary Structure Motifs

RNA Bricks



RNA 3D Motif Atlas

is a comprehensive and representative collection of internal and hairpin

In Summer 2021 we made small adjustments to the clustering methodo

Current version: 3.60

Current version: 3.60

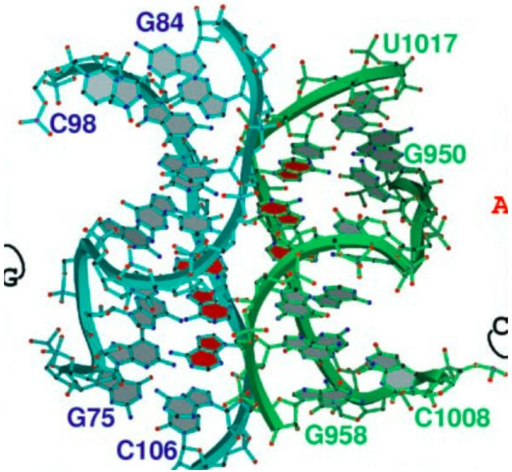
Latest motif atlas release based on the representative set from: 2022-07-20

RNAJunction
A database of RNA junction and kissing loop structures

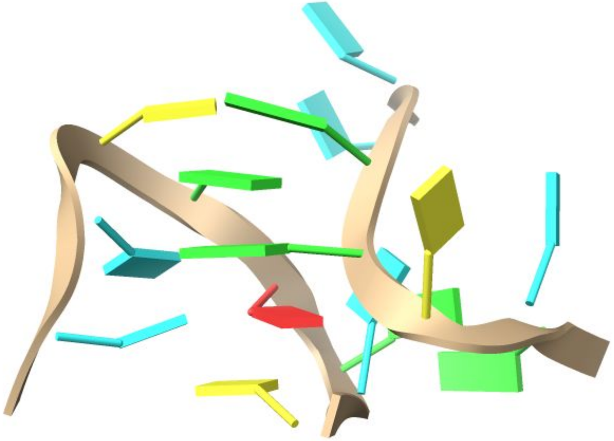
Long-range tertiary motifs

2D ≤ X ≤ 3D

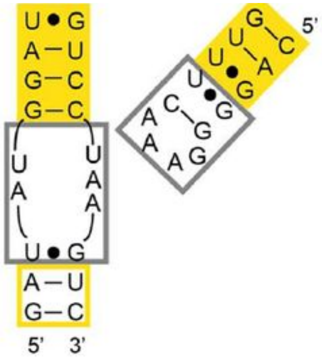
2.5D ≤ X ≤ 3D



A-patch



D-loop/T-loop



GAAA/11nt

Long-range tertiary motifs

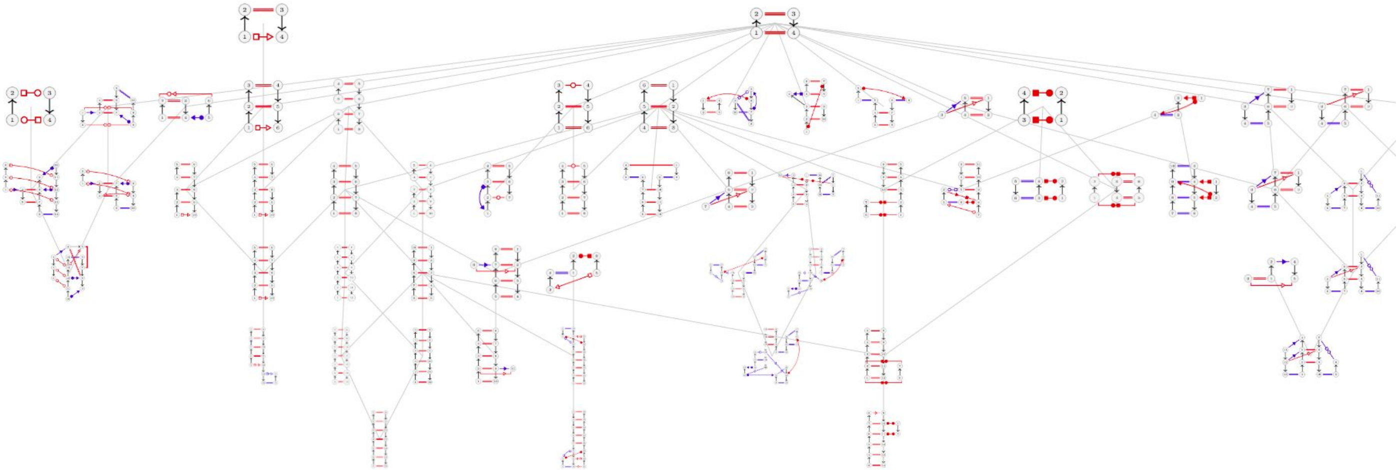
2D ≤ X ≤ 3D

2.5D ≤ X ≤ 3D

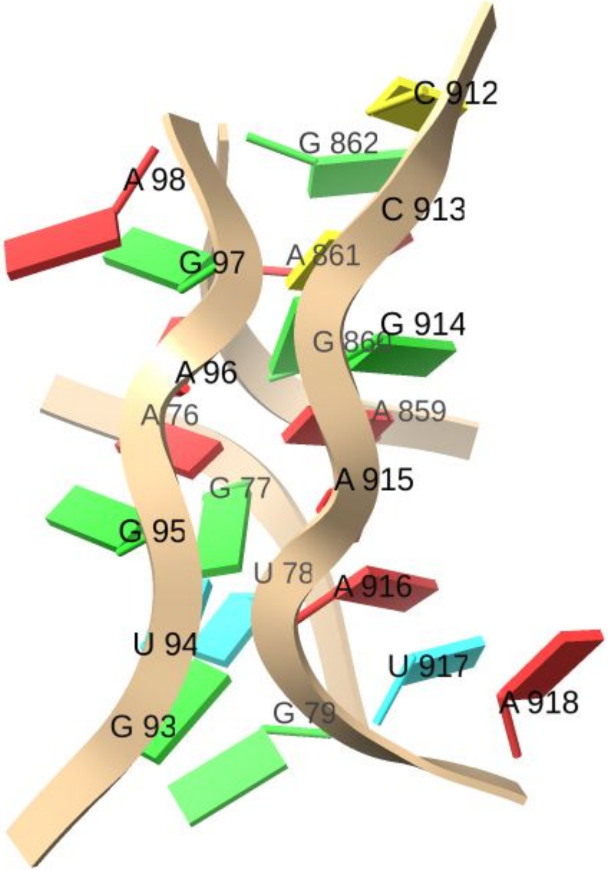
Distant SSEs RINs

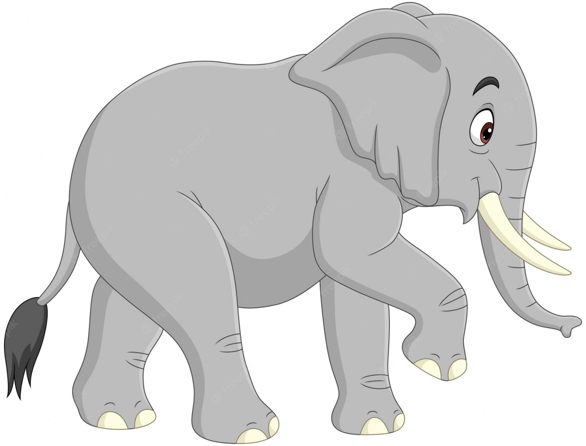
Pseudoknot mesh
A-minor mesh
A-minor type I
A-minor type I/II
trans-W-C/H mesh
Ribose zipper
GNRA
A-rich loop

CaRNAval
Homepage



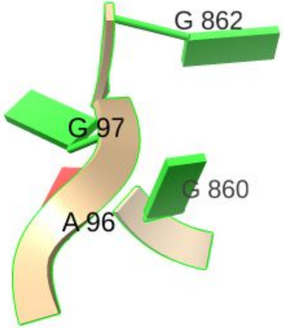
Double symmetrical A-patch (23S-5S rRNA)

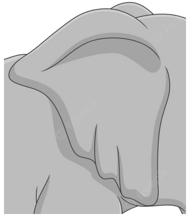




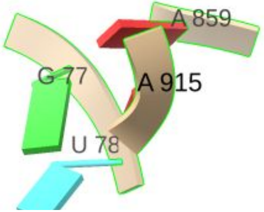
Double symmetrical A-patch (23S-5S rRNA)

```
7M4Y|1|B|A|96   ntSS  7M4Y|1|A|G|860
7M4Y|1|B|G|97    1BPh 7M4Y|1|A|G|862
```

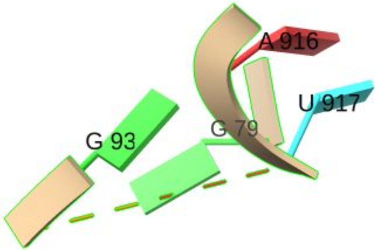


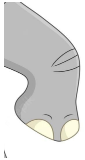


```
7M4Y|1|A|A|915   2BR  7M4Y|1|B|U|78
7M4Y|1|B|G|77   ntSS  7M4Y|1|A|A|915
7M4Y|1|B|G|77   ncSS  7M4Y|1|A|A|859
```



```
7M4Y|1|A|A|916   2BPh 7M4Y|1|B|G|79
7M4Y|1|B|G|93   n1BPh 7M4Y|1|A|U|917
7M4Y|1|B|G|93    3BR  7M4Y|1|A|A|916
```

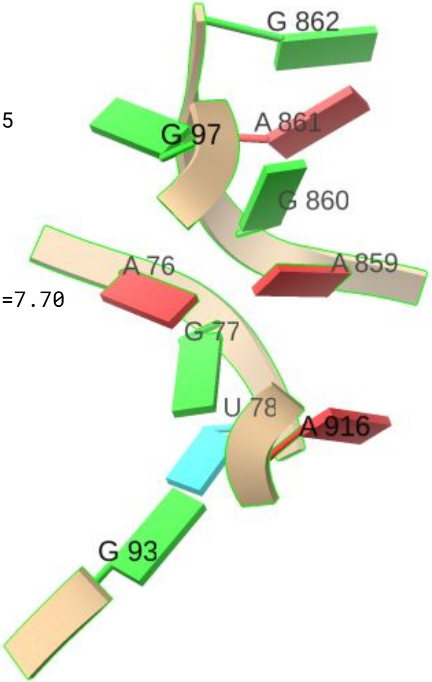




http://rna.bgsu.edu/rna3dhub/pdb/7M4Y/interactions/fr3d/all

Double symmetrical A-patch (23S-5S rRNA)

```
1..A.A.859.    1..B.G.77.     interBase-angle=66 H-bonds[2]: "O2'(hydroxyl)-O3'[3.06],N3-O2'(hydroxyl)[2.60]" min-baseDist=4.83
1..A.G.860.    1..B.A.76.     interBase-angle=65 H-bonds[2]: "O3'-O2'(hydroxyl)[2.97],O2'(hydroxyl)-N3[3.04]" min-baseDist=5.75
1..A.G.860.    1..B.G.97.     interBase-angle=71 H-bonds[1]: "N2(amino)-O4'[3.50]" min-baseDist=5.19
1..A.A.861.    1..B.G.97.     interBase-angle=74 H-bonds[2]: "O3'-N2(amino)[3.32],O2'(hydroxyl)-O2'(hydroxyl)[2.74]"
min-baseDist=6.52
1..A.G.862.    1..B.G.97.     interBase-angle=78 H-bonds[1]: "OP1-N2(amino)[3.43]" min-baseDist=7.69
1..A.A.916.    1..B.G.77.     interBase-angle=65 H-bonds[1]: "O4'-N2(amino)[3.18]" min-baseDist=4.60
1..A.A.916.    1..B.U.78.     interBase-angle=67 H-bonds[3]:
"O2'(hydroxyl)-O2'(hydroxyl)[3.03],O2'(hydroxyl)-O2(carbonyl)[3.48],N3-O2'(hydroxyl)[2.57]" min-baseDist=5.70
1..A.A.916.    1..B.G.93.     interBase-angle=69 H-bonds[1]: "O2'(hydroxyl)-N2(amino)[2.94]" min-baseDist=6.55
1..A.G.862.     OP1-hbonds[2]: "N2@1..B.G.97.[3.43]"
nts=4 AGAG 1..A.A.859.,1..A.G.860.,1..B.A.76.,1..B.G.77. (ribose zipper)
```

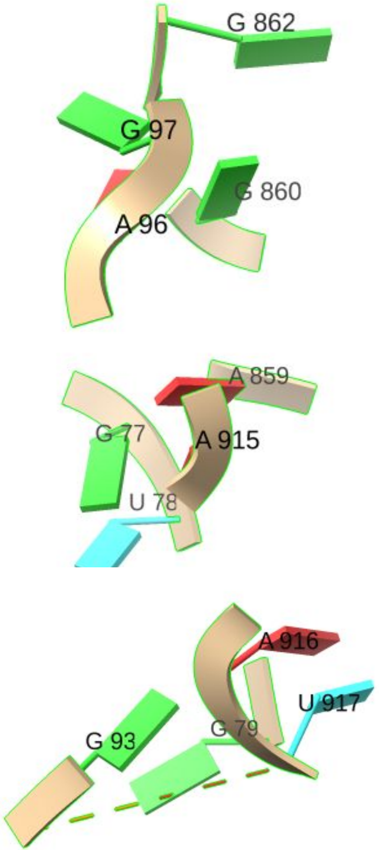


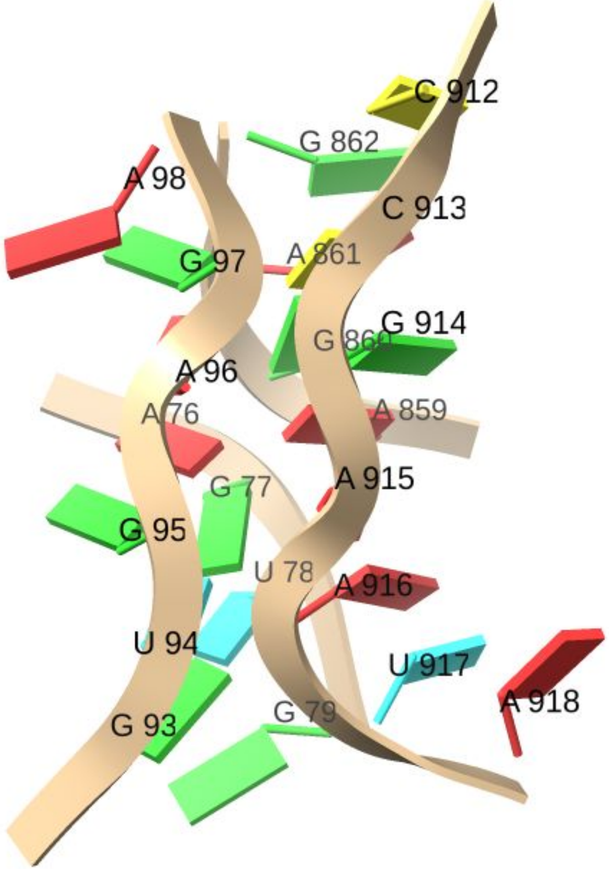
```
1..A.G.914.    1..B.A.96.     interBase-angle=60 H-bonds[1]: "O2'(hydroxyl)-O2'(hydroxyl)[3.28]" min-baseDist=7.70
```

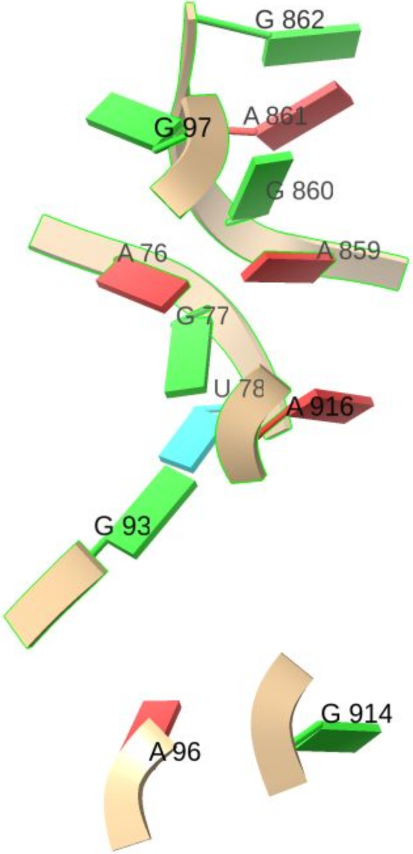
0 A-minors!!

http://forum.x3dna.org/rna-structures/ (DSSR, version 2.0)

And what about the “consensus”?



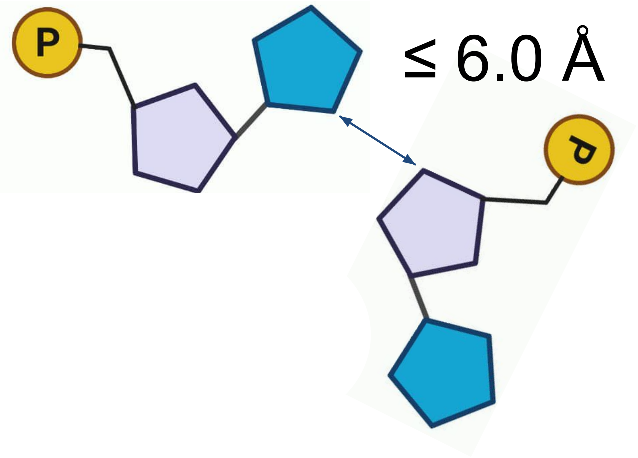


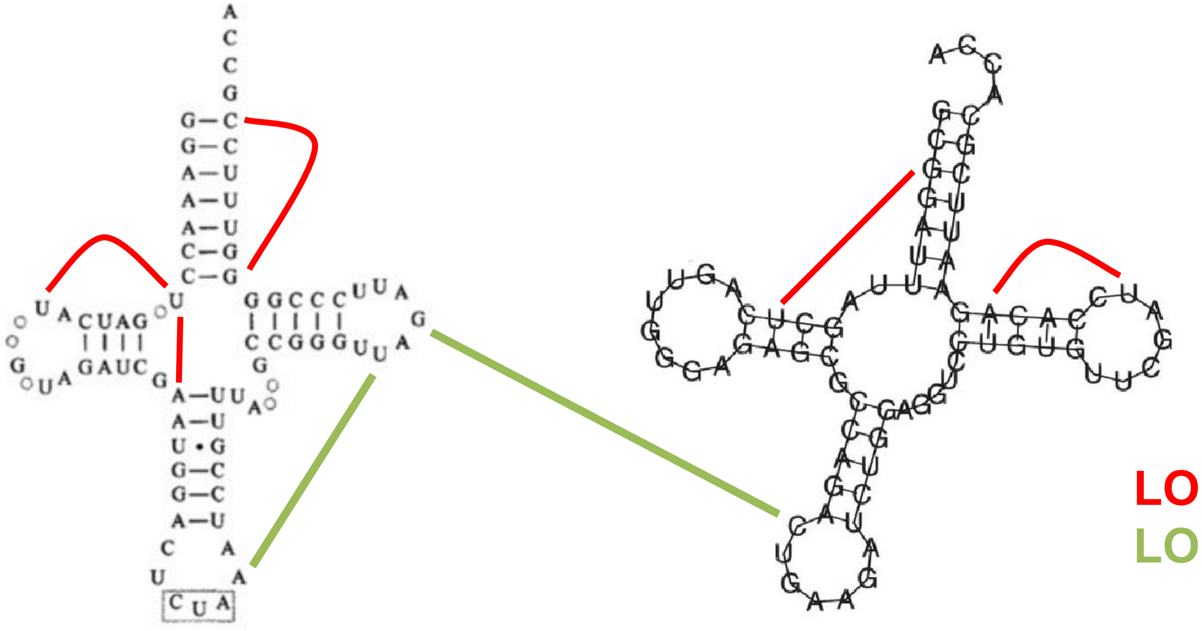


Long-range (LR) residue-residue pairs

BGSU NR-list, 4.0Å cutoff -> one chain per Rfam family

-> 97 RNA chains -> 70 chains with at least one LR pair





LOCAL

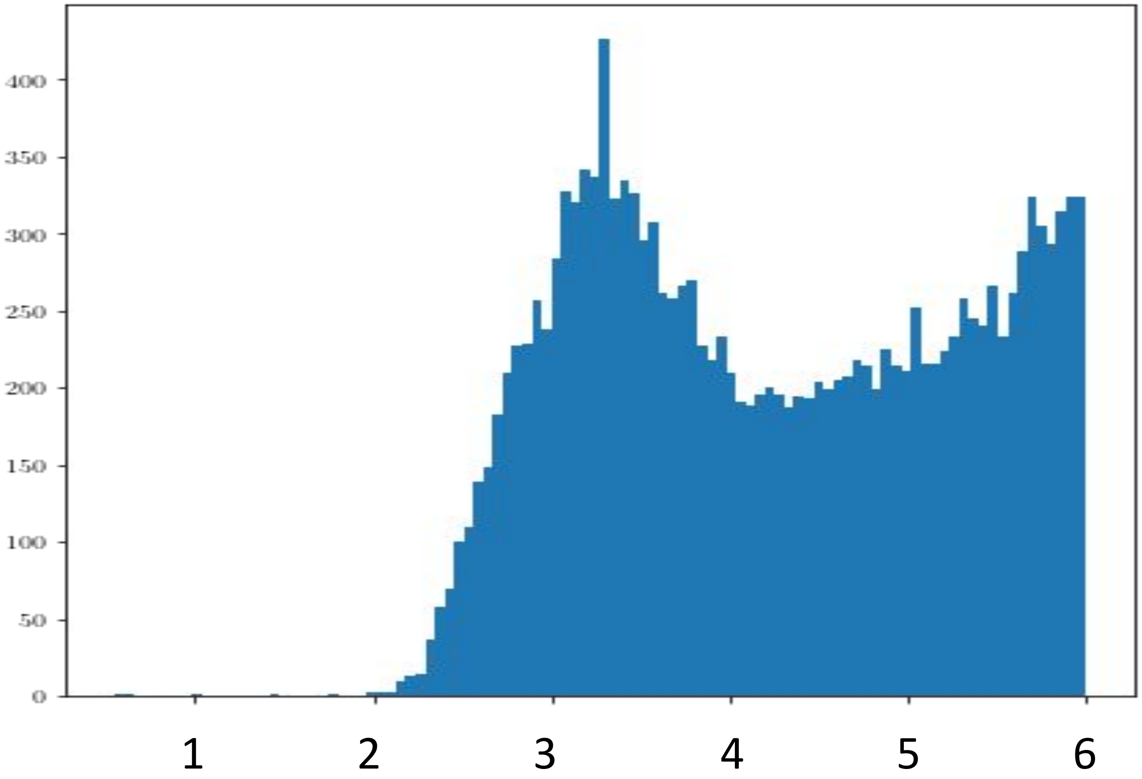
LONG-RANGE

Long-range (LR) residue-residue pairs

OVERALL	Stem	Loop	Total	Total %
A	550 (19.5%)	**2269 (80.5%)**	2819	26.30%
G	**2118 (61.2%)**	1340 (38.8%)	3458	32.30%
C	**1733 (69.3%)**	769 (30.7%)	2502	23.40%
U	840 (43.4%)	**1095 (56.6%)**	1935	18.00%
Total	5241 (48.9%)	5473 (51.1%)	10714	100%

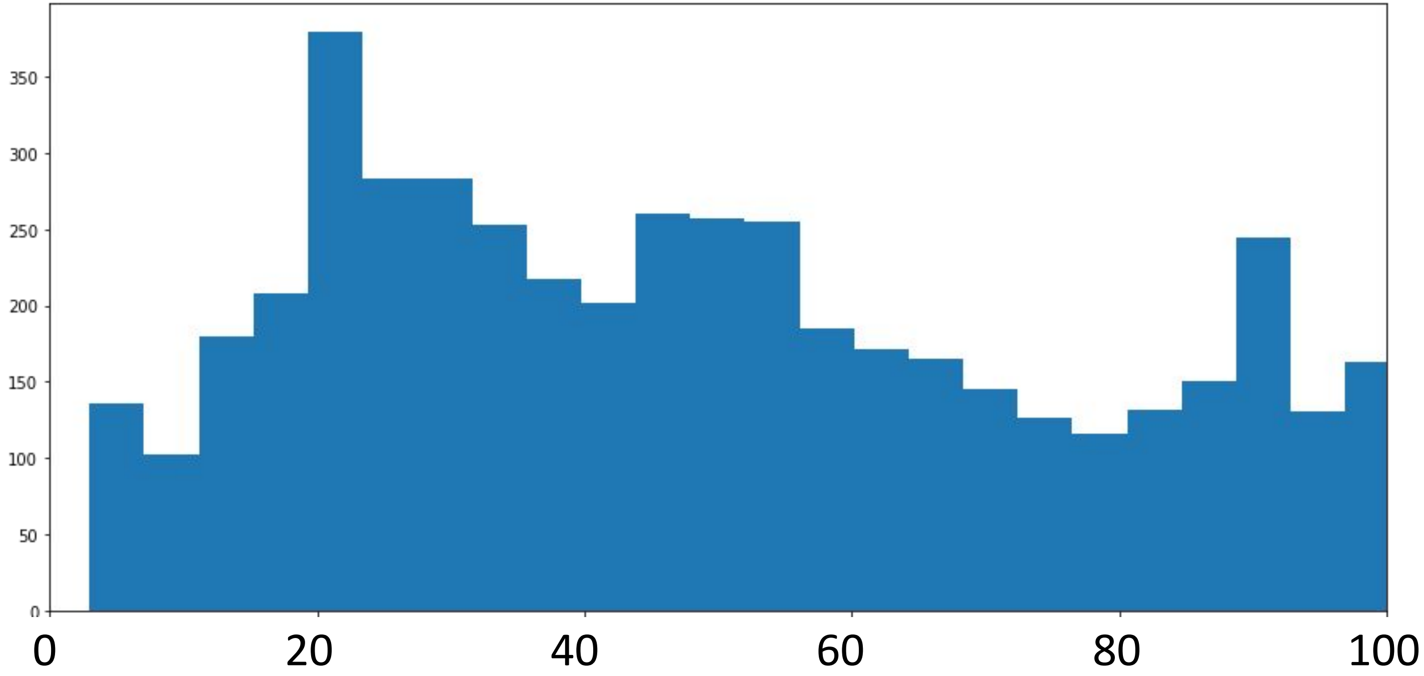
Long-range (LR) residue-residue pairs

Closest atom-atom pair distance (in angstroms)

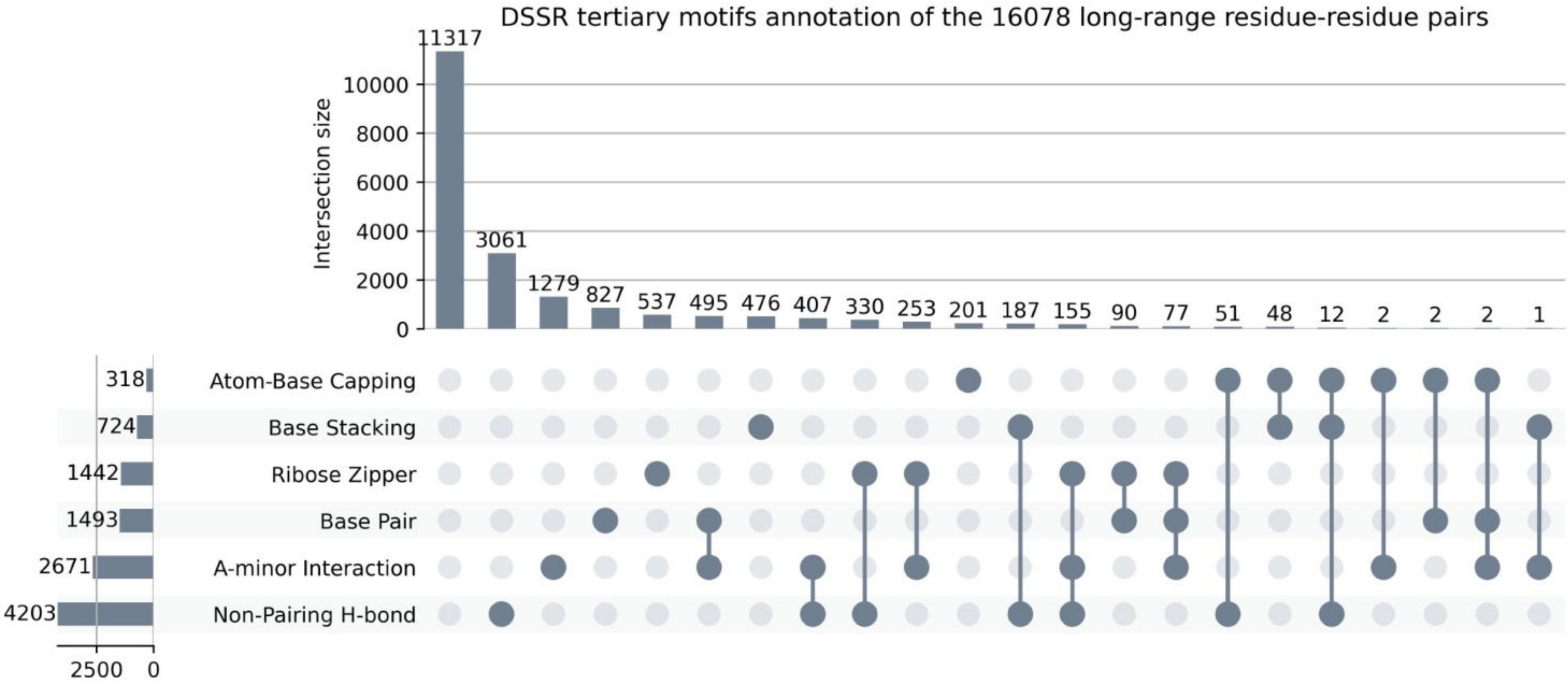


Long-range (LR) residue-residue pairs

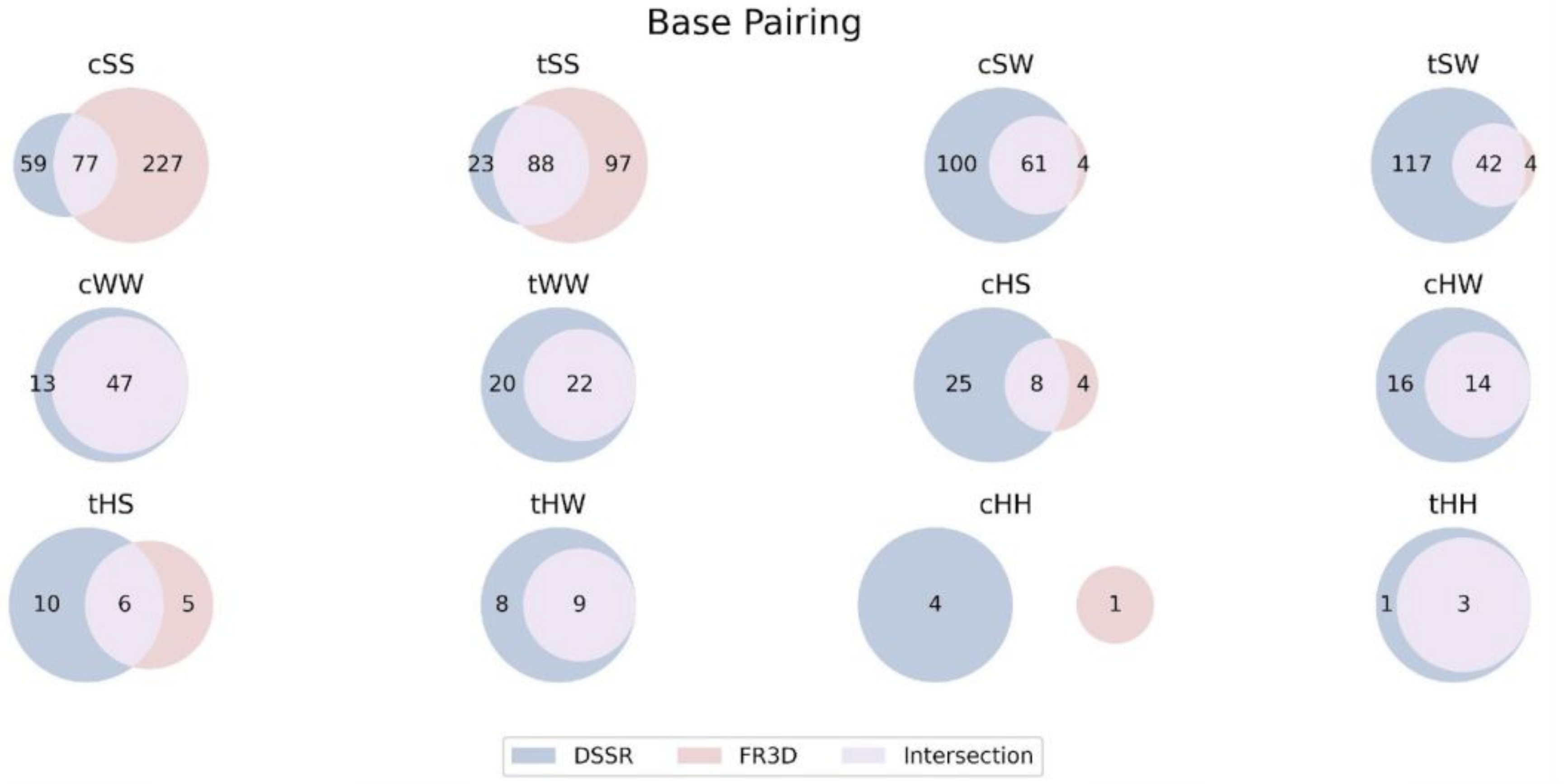
Sequence distance (in residues, cut at 100, median = 200, mean = 500)



Long-range (LR) residue-residue pairs

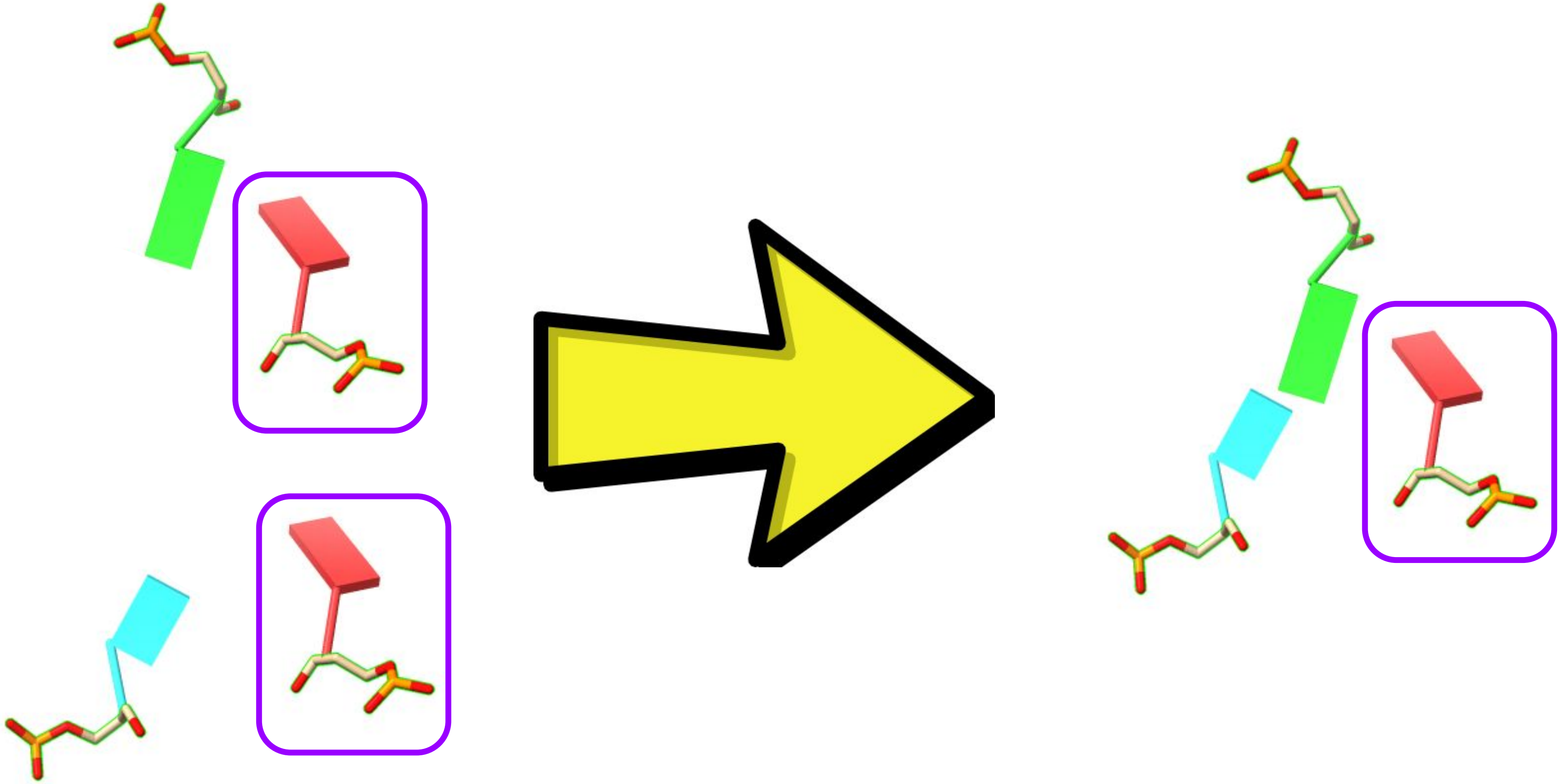

DSSR tertiary motifs annotation of the 16078 long-range residue-residue pairs

Long-range (LR) residue-residue pairs

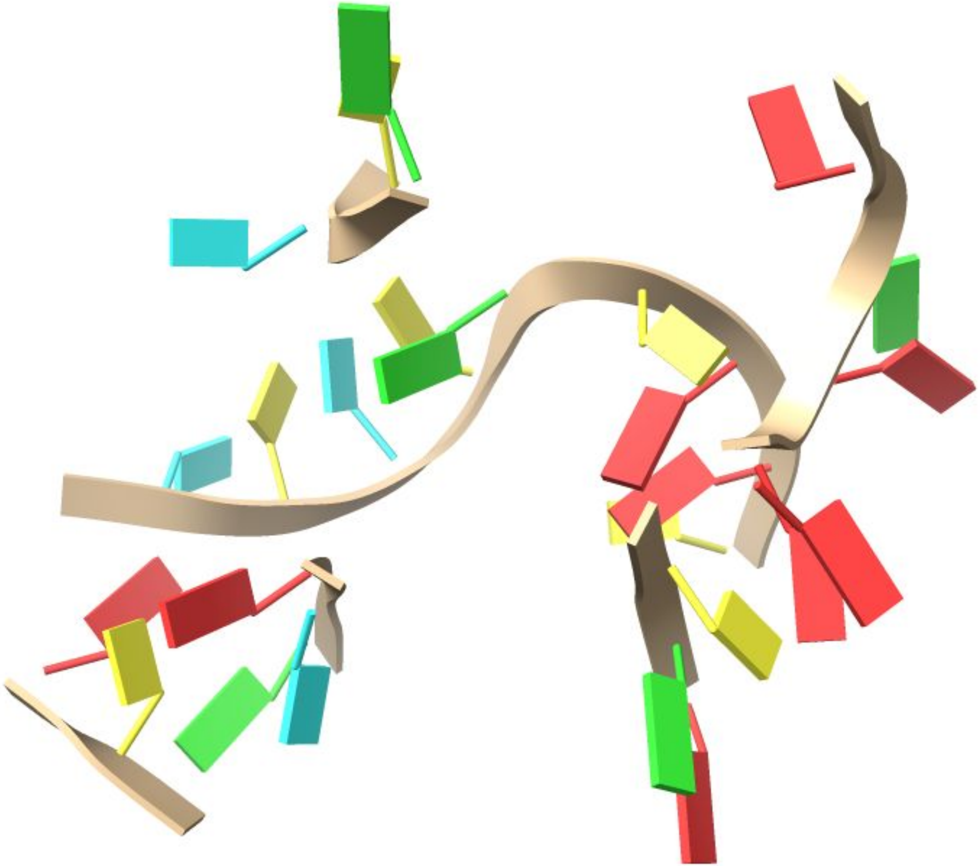


How to assemble LR modules?

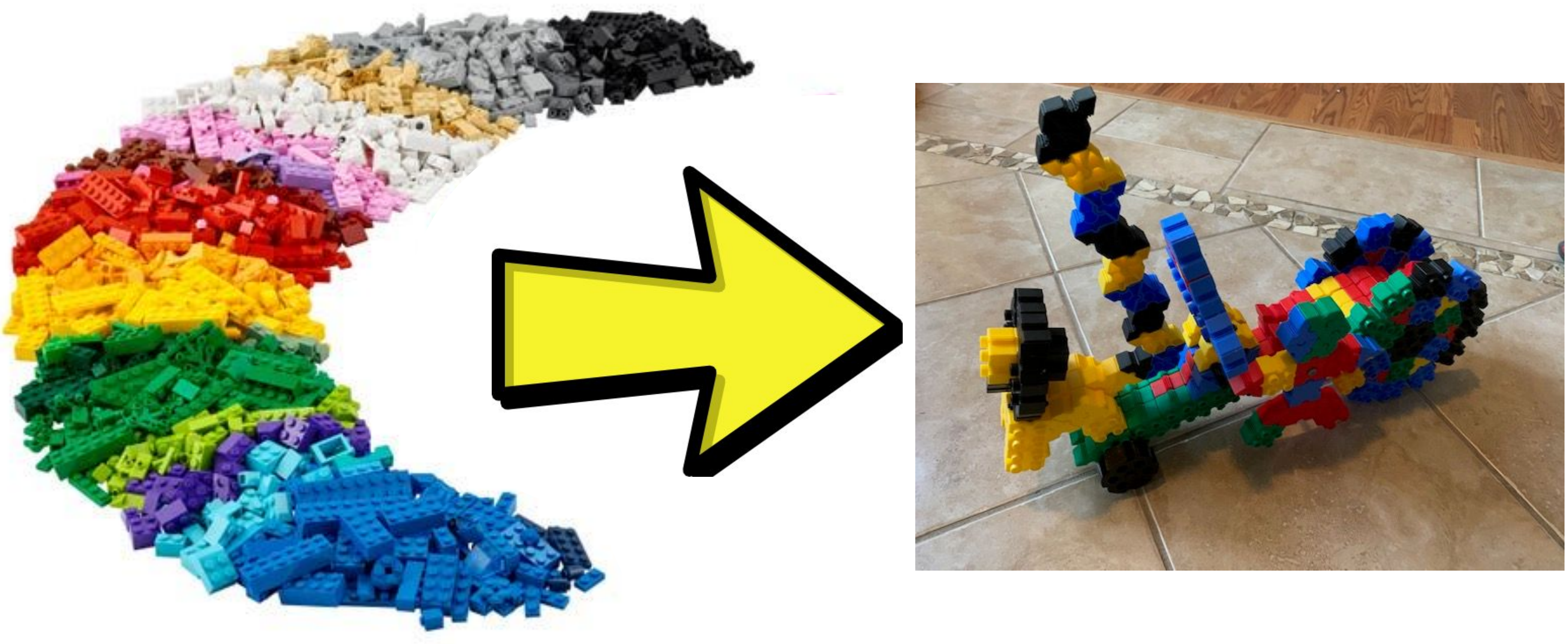
Glue pairs together?



Glue pairs together?

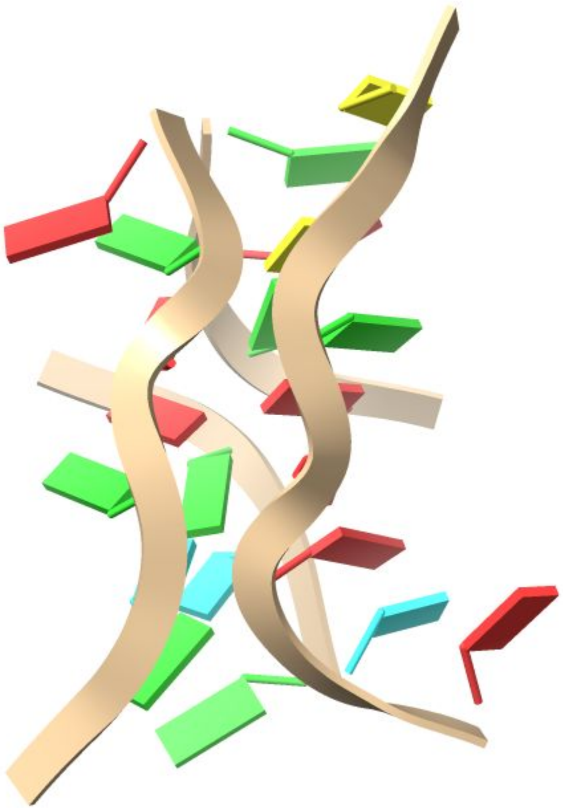


Glue pairs together?

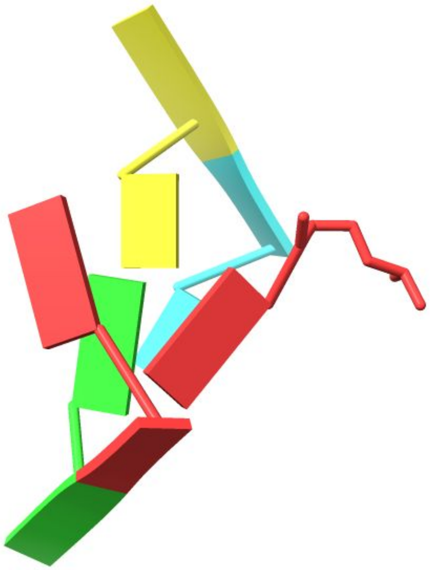


Another simple solution worked!

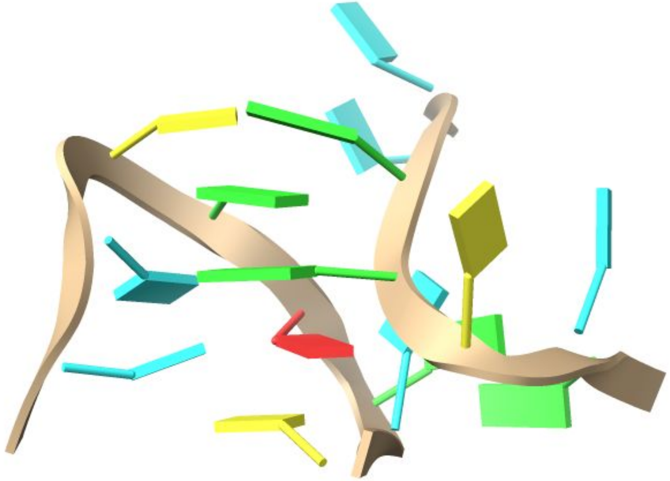
Double A-patch



Type-I A-minor

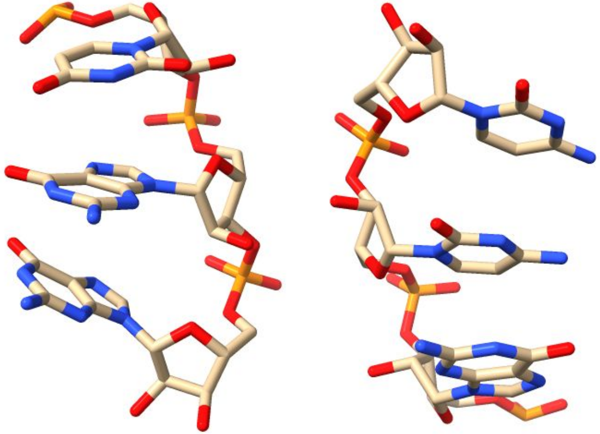


D-loop/T-loop interaction

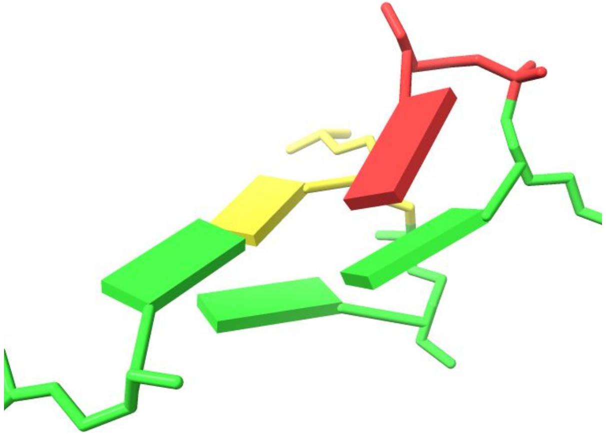


Another simple solution worked!

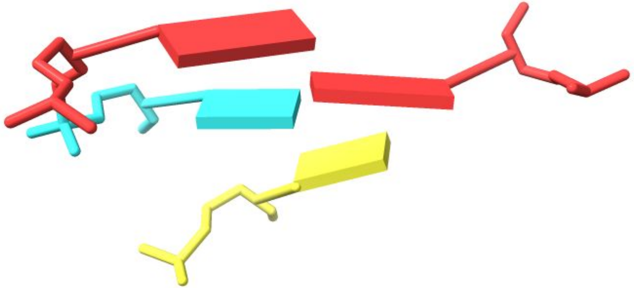
Ribose-zipper



Type-I/II
A-minor motif
(GA-variant)

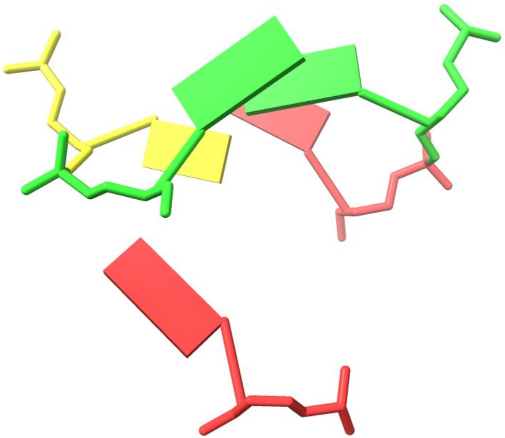


Cross-strand stack

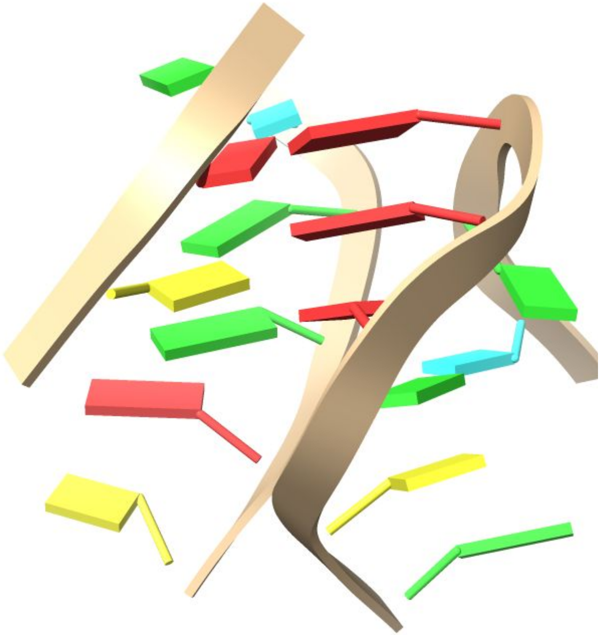


Another simple solution worked!

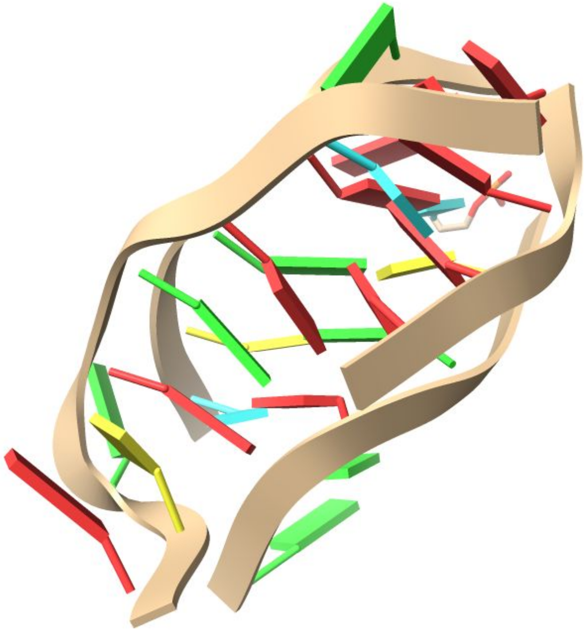
Non-canonical
A-minor



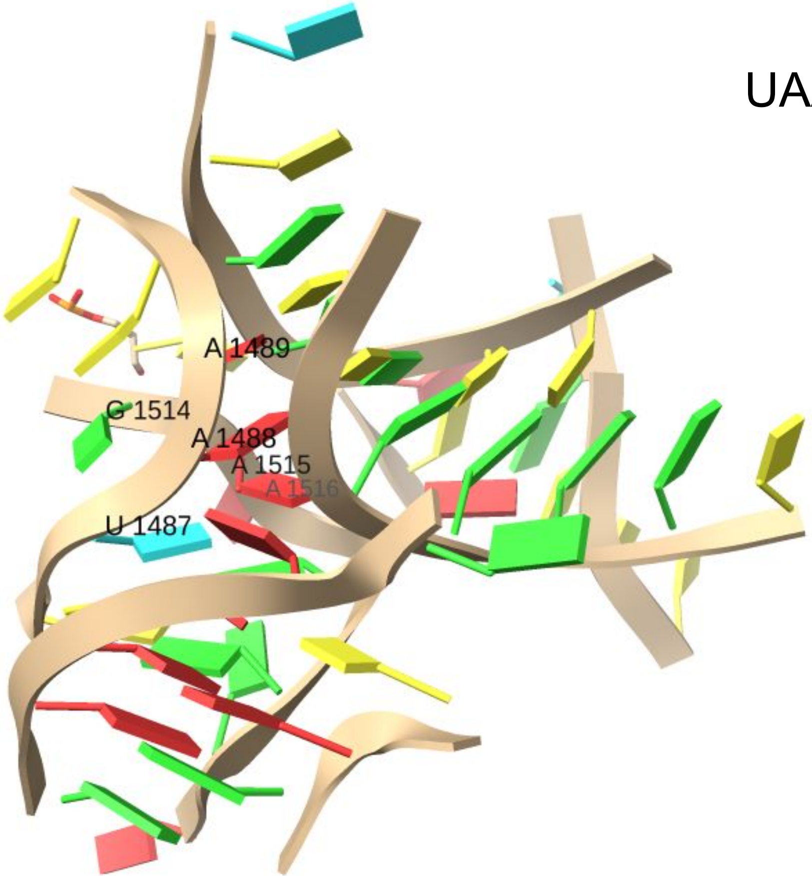
GAAA-tetraloop/receptor



A-patch of length 6



Context!



UAA/GAN internal loop

Better moment for contact annotations

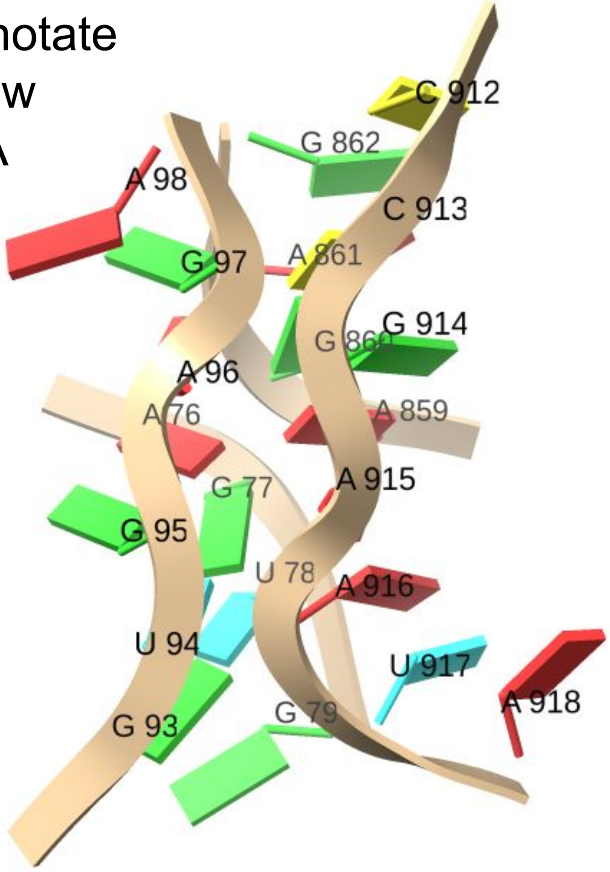
```
7M4Y|1|B|A|96   ntSS  7M4Y|1|A|G|860
7M4Y|1|B|G|97    1BPh 7M4Y|1|A|G|862

7M4Y|1|A|A|915   2BR  7M4Y|1|B|U|78
7M4Y|1|B|G|77   ntSS  7M4Y|1|A|A|915
7M4Y|1|B|G|77   ncSS  7M4Y|1|A|A|859

7M4Y|1|A|A|916   2BPh 7M4Y|1|B|G|79
7M4Y|1|B|G|93   n1BPh 7M4Y|1|A|U|917
7M4Y|1|B|G|93    3BR  7M4Y|1|A|A|916
```

FR3D

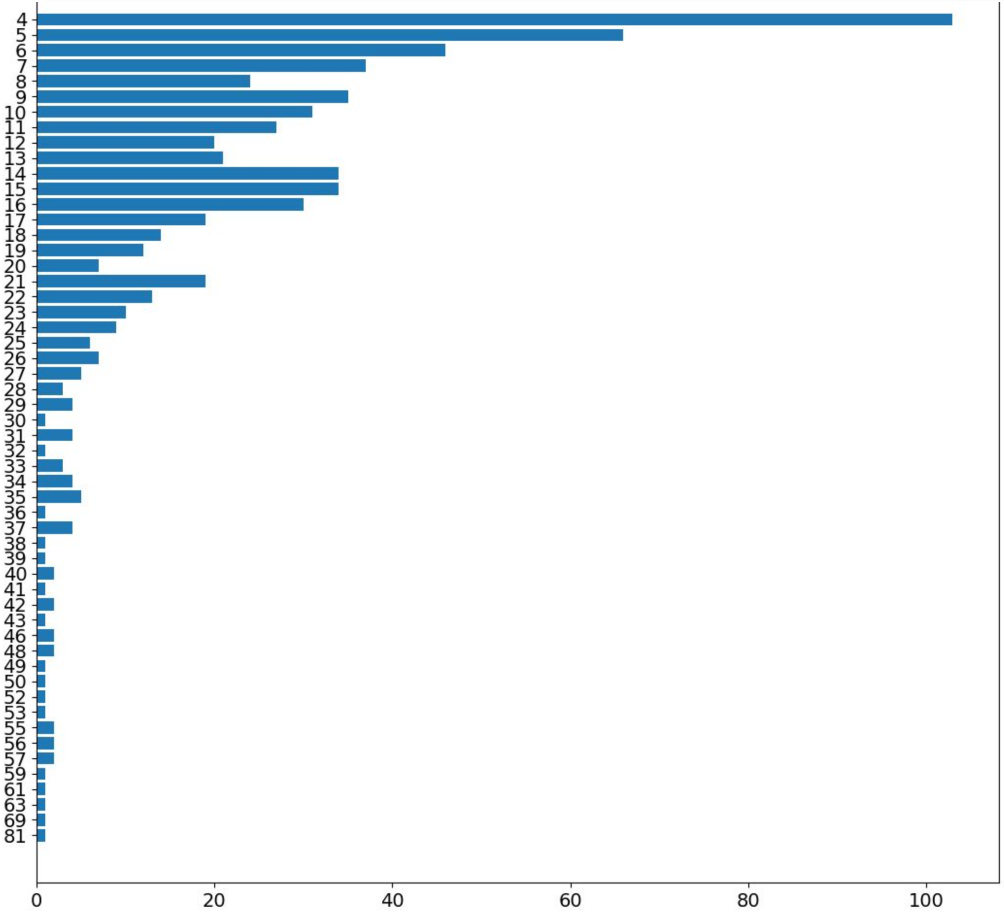
```
1..A.A.859.    1..B.G.77.     interBase-angle=66 H-bonds[2]:
"O2'(hydroxyl)-O3'[3.06],N3-O2'(hydroxyl)[2.60]" min-baseDist=4.83
1..A.G.860.    1..B.A.76.     interBase-angle=65 H-bonds[2]:
"O3'-O2'(hydroxyl)[2.97],O2'(hydroxyl)-N3[3.04]" min-baseDist=5.75
1..A.G.860.    1..B.G.97.     interBase-angle=71 H-bonds[1]:
"N2(amino)-O4'[3.50]" min-baseDist=5.19
1..A.A.861.    1..B.G.97.     interBase-angle=74 H-bonds[2]:
"O3'-N2(amino)[3.32],O2'(hydroxyl)-O2'(hydroxyl)[2.74]" min-baseDist=6.52
1..A.G.862.    1..B.G.97.     interBase-angle=78 H-bonds[1]:
"OP1-N2(amino)[3.43]" min-baseDist=7.69
1..A.A.916.    1..B.G.77.     interBase-angle=65 H-bonds[1]:
"O4'-N2(amino)[3.18]" min-baseDist=4.60
1..A.A.916.    1..B.U.78.     interBase-angle=67 H-bonds[3]:
"O2'(hydroxyl)-O2'(hydroxyl)[3.03],O2'(hydroxyl)-O2(carbonyl)[3.48],
N3-O2'(hydroxyl)[2.57]" min-baseDist=5.70
1..A.A.916.    1..B.G.93.     interBase-angle=69 H-bonds[1]:
"O2'(hydroxyl)-N2(amino)[2.94]" min-baseDist=6.55
1..A.G.862.    OP1-hbonds[2]: "N2@1..B.G.97.[3.43]"
nts=4 AGAG 1..A.A.859.,1..A.G.860.,1..B.A.76.,1..B.G.77. (ribose zipper)

1..A.G.914.    1..B.A.96.     interBase-angle=60 H-bonds[1]:
"O2'(hydroxyl)-O2'(hydroxyl)[3.28]" min-baseDist=7.70
```

DSSR

+ MC-Annotate
+ RNAView
+ ClaRNA



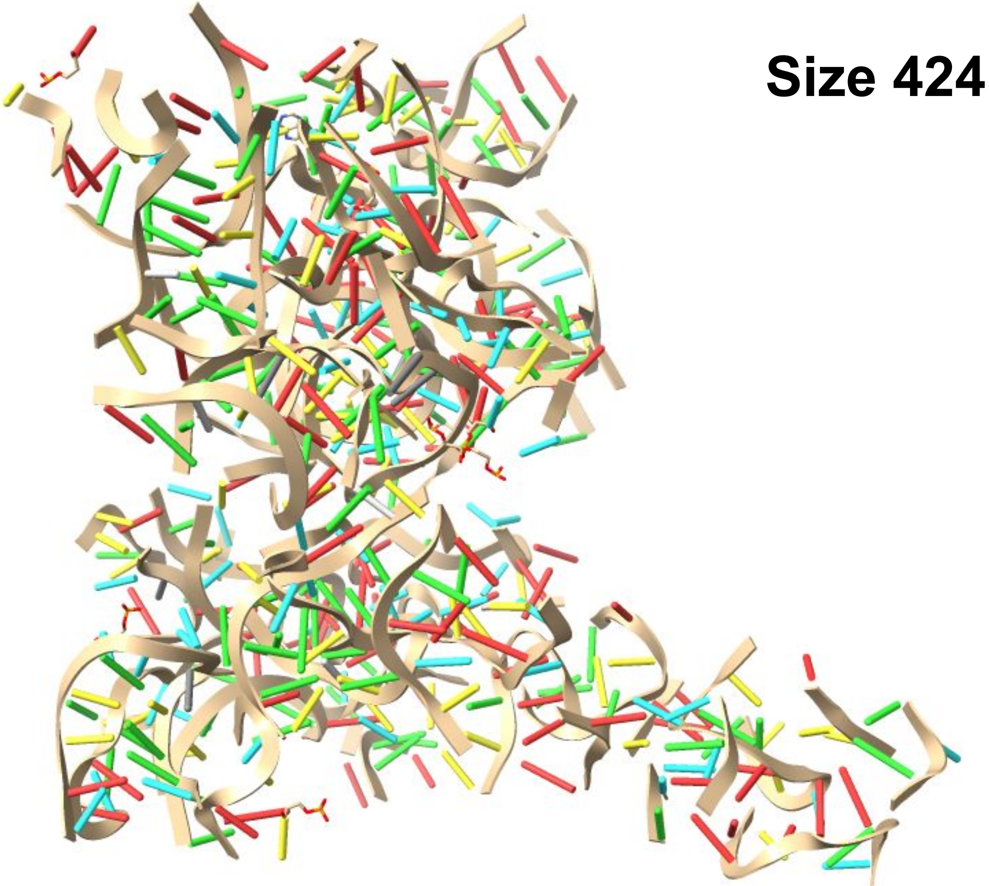
Well, to some extent…



Size 2 = ~400
Size 3 = ~200

Size ~100 = 3
Size ~500 = 3

Well, to some extent…



Size 424 (LSU rRNA)

Summary

- We suggest a new approach to long-range RNA tertiary module annotation

- The method does not rely on any interaction type

- How to go from modules to motifs?

Acknowledgements

Davyd Bohdan
MIPT, Moscow

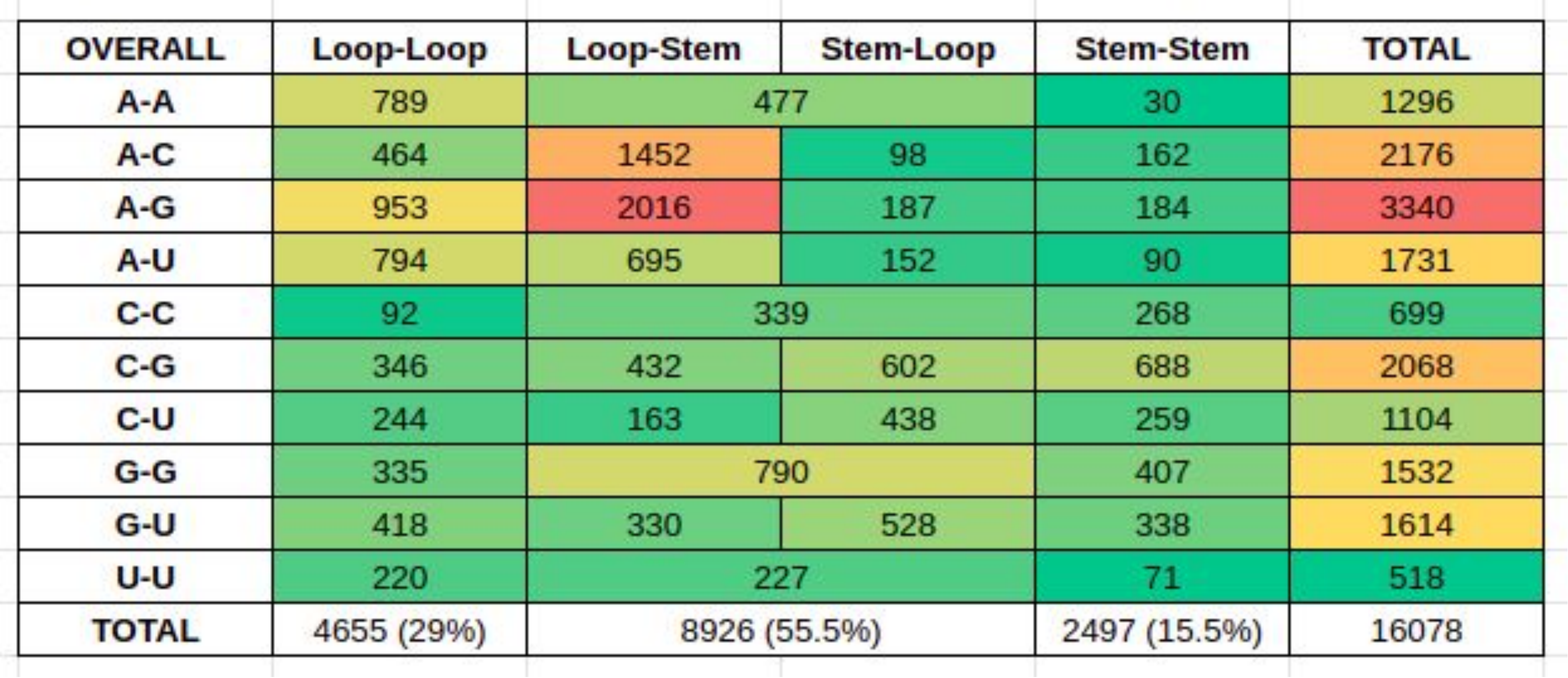
Valeria Voronina
UlSTU, Ulyanovsk

Janusz M Bujnicki
IIMCB in Warsaw

ALTF-525-2022 fellowship
to Evgenii Baulin

Thank you!

Long-range (LR) residue-residue pairs



OVERALL	Loop-Loop	Loop-Stem	Stem-Loop	Stem-Stem	TOTAL
A-A	789	477		30	1296
A-C	464	1452	98	162	2176
A-G	953	2016	187	184	3340
A-U	794	695	152	90	1731
C-C	92	339		268	699
C-G	346	432	602	688	2068
C-U	244	163	438	259	1104
G-G	335	790		407	1532
G-U	418	330	528	338	1614
U-U	220	227		71	518
TOTAL	4655 (29%)	8926 (55.5%)		2497 (15.5%)	16078