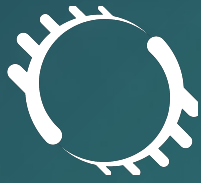




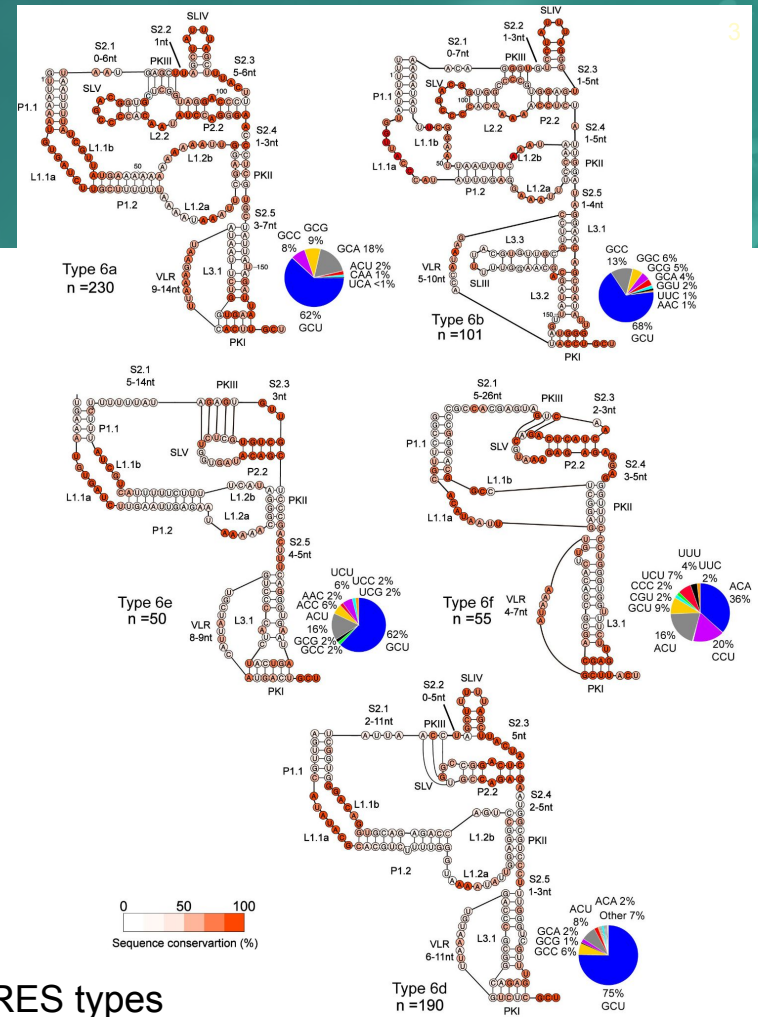
Zooming out to See Clearer

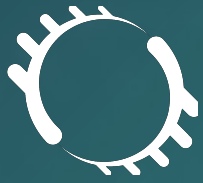
Hosna Jabbari
University of Alberta



Challenge

How to define structural similarity in highly variable structures?

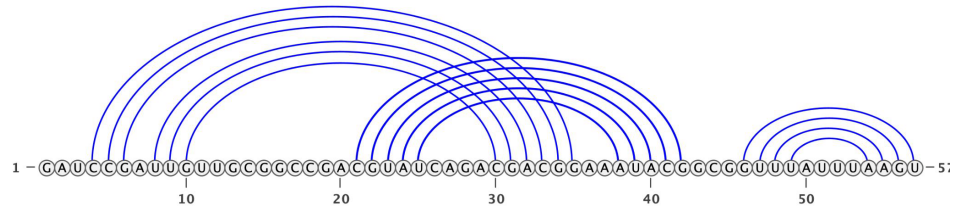
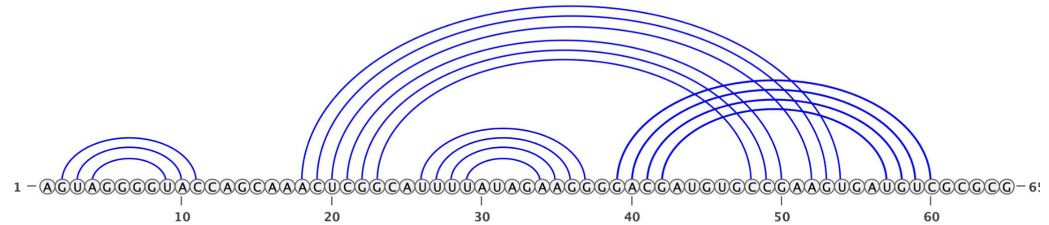


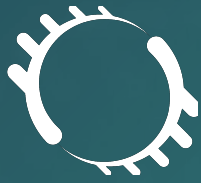


Structure Alignment

4

Input: Two possibly pseudoknotted secondary structures





Existing methods

5

Nucleotide-to-Nucleotide mapping:

- LocARNA [Will et al. 2007]
- LaRA and LaRA 2 [Markus et al. 2007, Winkler et al. 2022]

Descriptor-based search:

- RNARobo [Rampášek et al. 2016]

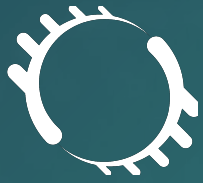


RNA Abstract Shapes

$(((.((..((...)))..((...))..)))..)$

(Level 1) $(_((_))(_)_)_$

(Level 5) $((())())$



Shadow Abstraction

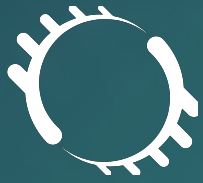
7

$((\dots((\dots))\dots[[\dots))\dots])]$

(Level 6)

$([])$

[Reidys et al., Bioinformatics, 2011]



Our Abstraction

8

$((\dots((\dots))\dots[[\dots))\dots])]$



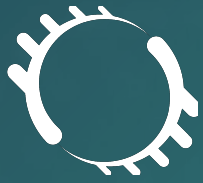
(Level 5)

$((\) [\])$

(Level 6)

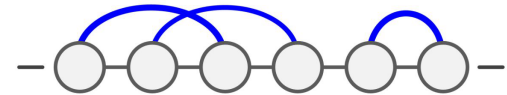
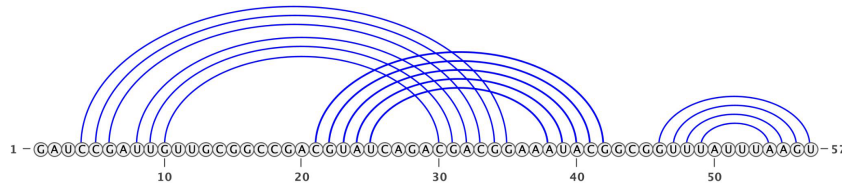
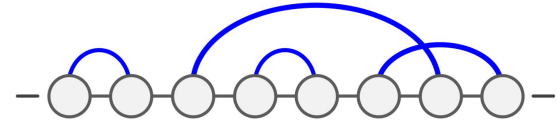
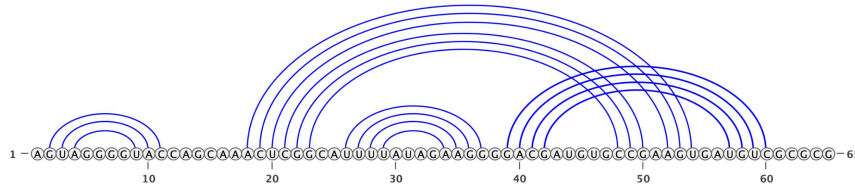
$([\])$

[Reidys et al., Bioinformatics, 2011]



Reduced Topology

9





CoSTAR: Coarse Stem-Topology Alignment of Pseudoknotted RNA Structures by Relation-Constrained Search

Alignment of secondary structures.

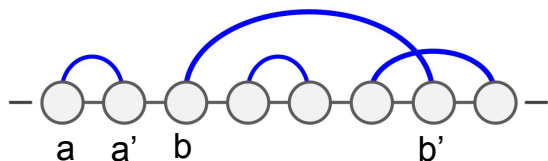
Works on stems rather than bases.

Tree search bounded by
directionality and topology.





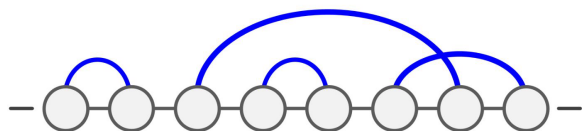
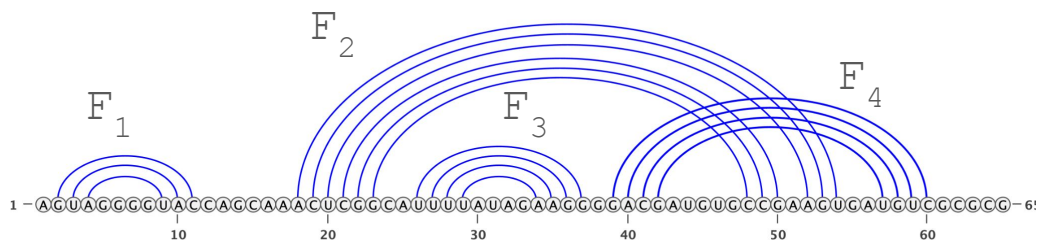
How does each stem “relate” to each other?



ℓ	condition	relation of B with respect to A	code alias
1	$b' \prec a$	5'-disjoint	5'SL
2	$b \prec a \prec b' \prec a'$	5'-crossing	5'PK
3	$b \prec a \prec a' \prec b'$	contains	outer
4	$a \prec b \prec a' \prec b'$	3'-crossing	3'PK
5	$a' \prec b$	3'-disjoint	3'SL
6	$a \prec b \prec b' \prec a'$	inside	inner.



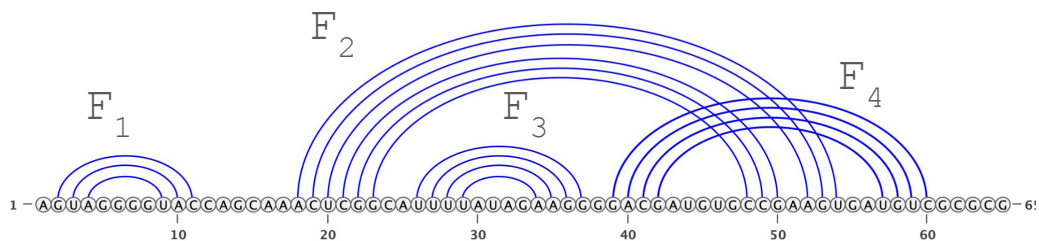
Stem Topology



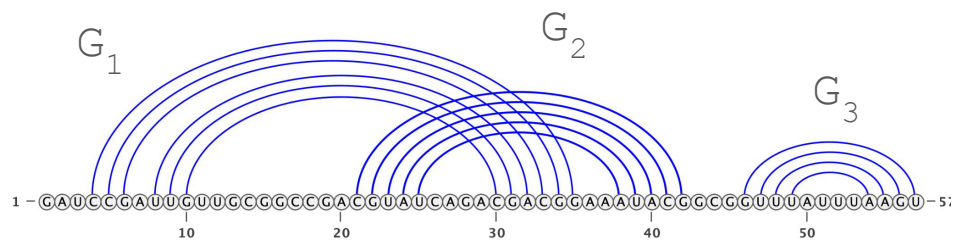
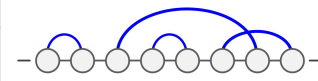
	F1	F2	F3	F4
F1		3' SL	3' SL	3' SL
F2	5' SL		Inner	3' PK
F3	5' SL	Outer		3' SL
F4	5' SL	5' PK	5' SL	



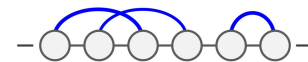
Stem Topology

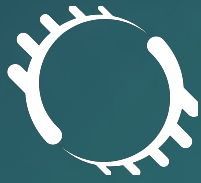


F1	F2	F3	F4
	3' SL	3' SL	3' SL
5' SL		Inner	3' PK
5' SL	Outer		3' SL
5' SL	5' PK	5' SL	

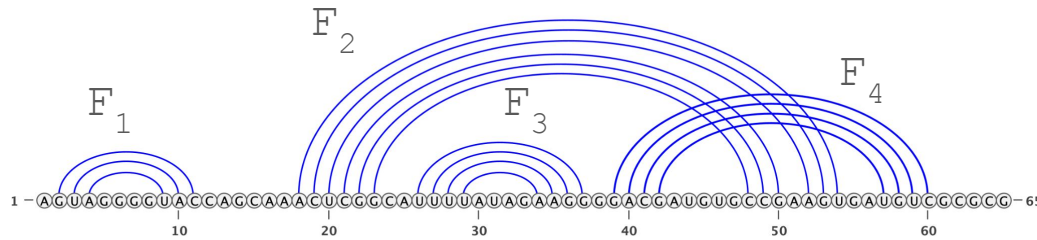


	3' PK	3' SL
5' PK		3' SL
5' PK	5' PK	





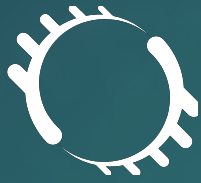
Stem Features



	F1	F2	F3	F4
H1	3	6	4	4
H2	4	25	4	14
H3	2.4	7.2	2.13	4.6

Stems are characterized by:

- **H1**: Number of basepairs
- **H2**: Distance between arms
- **H3**: Negative Stacking Energy

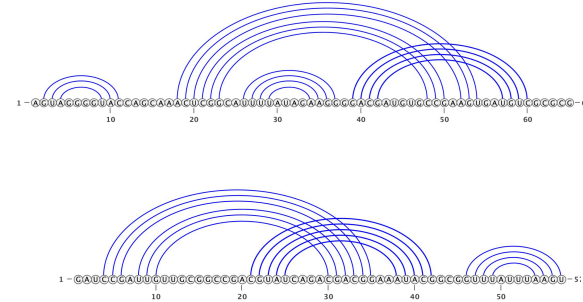


Input: Two Sequence-Structure pairs, (S^F, R^F) and (S^G, R^G)

- For each pair precompute
 - Ordered stem list
 - Reduced topology graph
 - Relation array
 - Feature vector

Output: Minimum-cost partial one-to-one mapping between stems of F and G:

- injective, direction-preserving and topologically consistent.





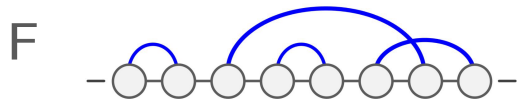
Alignment Cost

Cost of an alignment is:

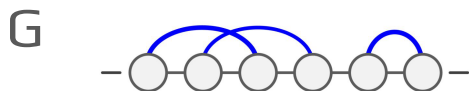
- Sum of features for skipped F stems
- Sum of features for skipped G stems
- For each mapped stem:
 - Absolute sum of difference in features between F_i and G_j



Search Alignment



Initialize Search Tree 5' → 3' with partial alignments starting from each pair ($|F| \times |G|$).

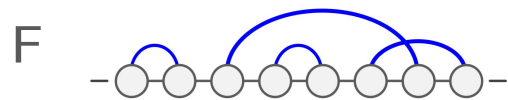


Partial alignments are represented as lists:

- Index is the F position
- Entry is the G position
- "-1" indicates stem is skipped
- "-2" indicates stem is not yet assigned



Search Alignment

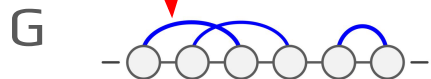


Initial alignment examples:

→ $F_{\text{align}} = [1, -2, -2, -2], G_{\text{align}} = [1, -2, -2]$

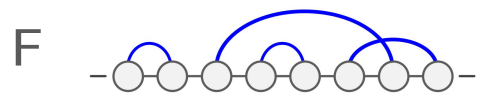
$F_{\text{align}} = [-1, 1, -2, -2], G_{\text{align}} = [2, -2, -2]$

$F_{\text{align}} = [-1, -1, \underline{2}, -2], G_{\text{align}} = [-1, \underline{3}, -2]$





Search Alignment

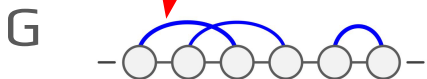


Initial alignment examples:

$$F_{\text{align}} = [1, -2, -2, -2], G_{\text{align}} = [1, -2, -2]$$

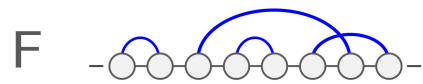
$$F_{\text{align}} = [-1, 1, -2, -2], G_{\text{align}} = [2, -2, -2]$$

$$F_{\text{align}} = [-1, -1, \underline{2}, -2], G_{\text{align}} = [-1, \underline{3}, -2]$$





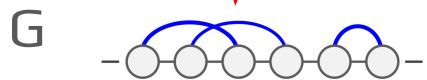
Search Alignment



Initial alignment examples:

$$F_{\text{align}} = [0, -2, -2, -2], G_{\text{align}} = [0, -2, -2]$$

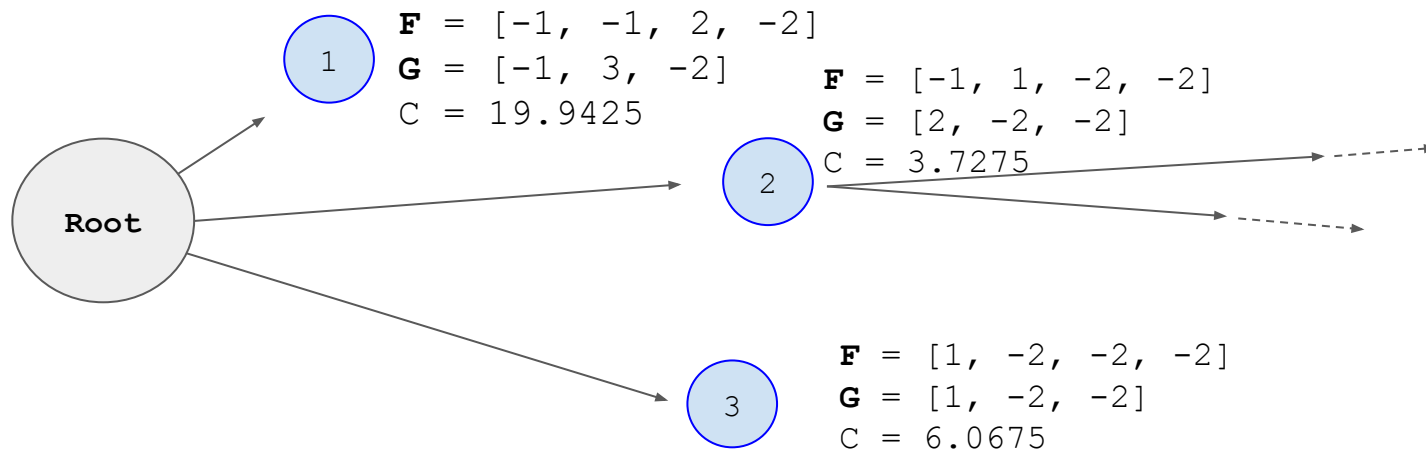
$$F_{\text{align}} = [-1, 0, -2, -2], G_{\text{align}} = [1, -2, -2]$$



$$F_{\text{align}} = [-1, -1, \underline{1}, -2], G_{\text{align}} = [-1, \underline{2}, -2]$$



Search Alignment





Search Constraint

Extending alignment requires:

- Next stem must be 3' of previously aligned stems
- **Directionality**
- Must have the same “relation type” in for all prior stems in F as in G
- **Topology**
- Expanding the node “be worth it”

Mapped and skipped decisions are final.



Mapping to Nucleotides

CoSTAR finds a mapping. To get back to nucleotide resolution:

- Topology format inspired by RNArobo
 - S0 H1 S1 H2 S2 H1' S3 H2' S4
- Build the nucleotide sequence (and structure) from the 5' side using that topology.



Benchmark

Selected the first 10 from each family.

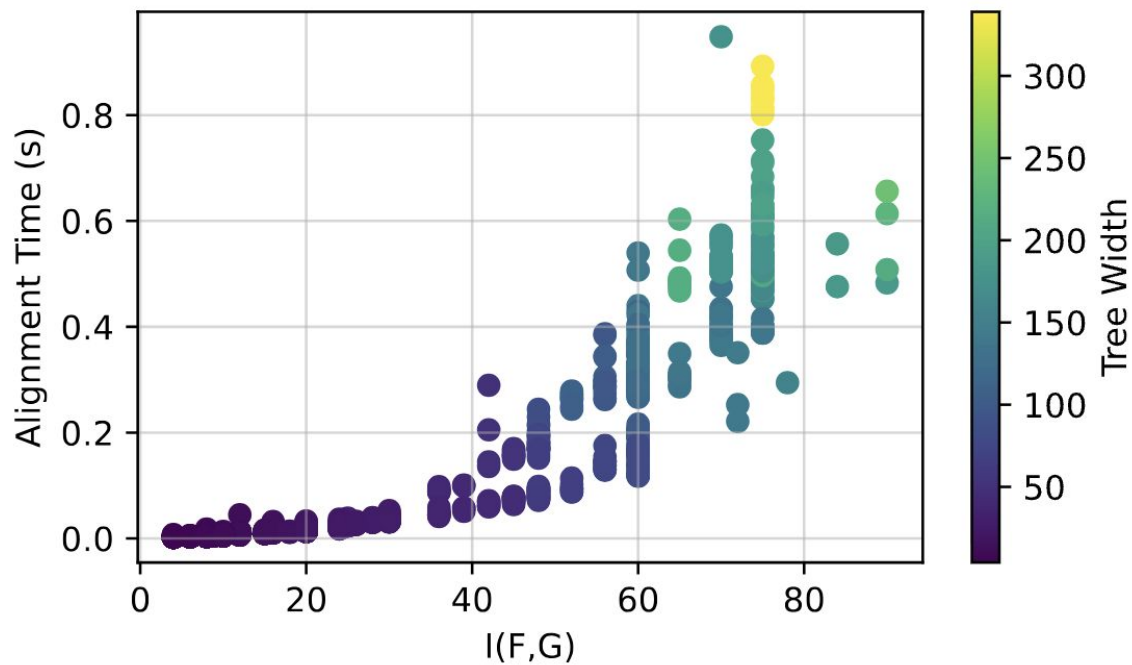
Used consensus structure.

Compared within families and between families

Rfam family	N. Seq.	Seq. Id.	PK	Min. NT	Max. NT	Min. stems	Max. stems
RF00001	712	0.75	No	116	118	5	6
RF00003	100	0.68	No	162	164	5	5
RF00008	85	0.71	No	40	82	3	4
RF00165	52	0.60	Yes	61	63	2	2
RF00174	434	0.73	Yes	180	224	12	15
RF00381	13	0.82	Yes	56	59	2	2
RF00507	51	0.63	Yes	78	81	3	4

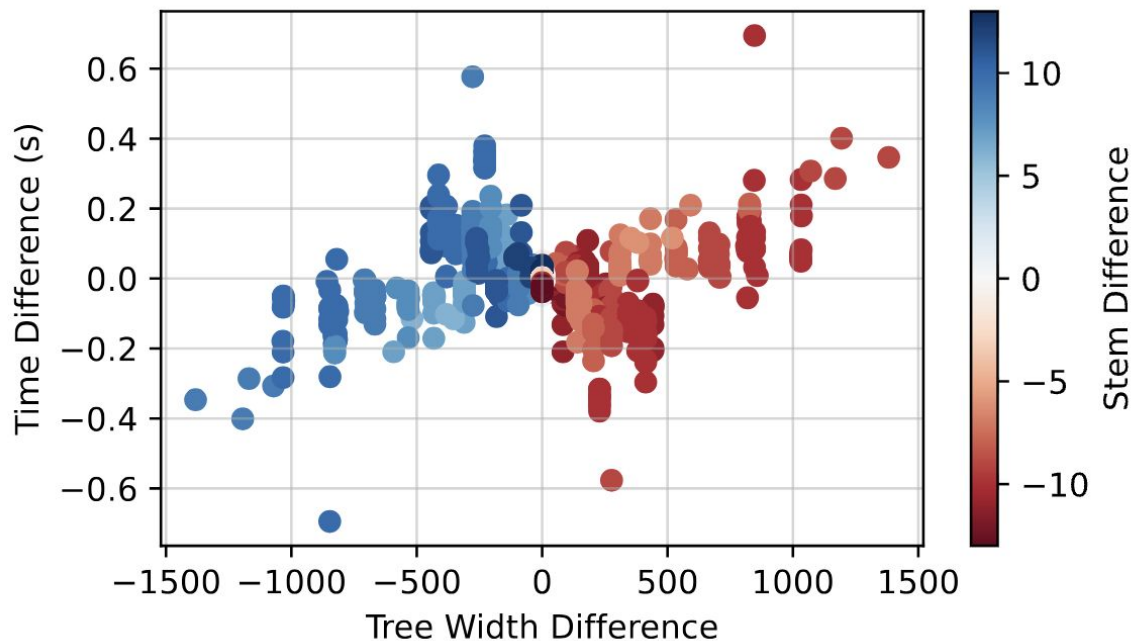


Time Growth by F and G





Order Matters





Comparison with LaRA

CoSTAR F

```
|AGUAGGGGUACCAGCAA|ACU_CGG|CAUUUUUAUAGAAGGG|GACG_|AUGUG|CCGAAGU|GA|UGUC_|GCGCG_____|
|..(((.....))).....|(((_(((|..(((.....)))..|[[[[_|.....|)))..))|..|]]]]_|....._____|
```

CoSTAR G

```
|GAU_____|CCGAUUG|UUGCGGCCGA_____|CGUAU|CAGA_|CGA_CGG|AA|AUACG|GCGGUUUUAUUUAAGU|
|..._____|(((.(((|....._____|[[[[[|....._|)))_)|..|]]]]]|...(((.....)))|
```

LaRA F

```
AGUAGGGGUACCAGCAAACUCGGCAUUUUUAUAGAAGGGGACGAUGUGCCGAAGUGAUGUCGCGCG
..(((.....))).....((((.....(((.....)))..[[[.....]))..]]]).....
```

LaRA G

```
GAUCCGAUUGUU-GCGGCCGACGUAUCAGAC-GACGGAAUACGCGGUUUUAUUUA-----AGU
...(((.(((.._.....[[[[[....._)]))..]]]]]...(((.....)_____)))
```



Resulting Nucleotide Alignment

CoSTAR F

```
| GGGUUCGGGGUACUAGUGUGAA | UGCCC | GGCUAGUACCCUGU | GCUA | GUG | GU | UUAUCU | AC | UGA | UGUU | CAAUUA | GGGCA | UUUG |
| .....((((((((.....|[[[[[...)))))...|(.((|...|((|.....|))|...|))|.)|.....|]]]]|....|
```

CoSTAR G

```
| _____ | UGCCC | GGCUAGUACCCUGU | GCUA | GUG | GU | UUAUCU | AC | UGA | UGUU | CAAUUA | GGGCA | _____ |
| _____ | (((((| .....|(.((|...|((|.....|))|...|))|.)|.....|))))| _____ |
```

Lara F

```
GGGUUCGGGGUACUAGUGUGAAUGCCCGGCUAGUACCCUGUGCUAGUGGUUUAUCUACUGAUGUUCAAUUAAGGGCAUUUG
.....((((((((.....[[[[[...)))))...|.((...((.....))...))|.)|.....|]]]]|....
```

Lara G

```
UG-----C--CCGGCUAGUACCCUGUGCUAGUGGUUUAUCUACUGAUGUUCAAUUAAGGGCA----
(( _____ ( ____ ((.....|.((...((.....))...))|.)|.....|)))) _____
```



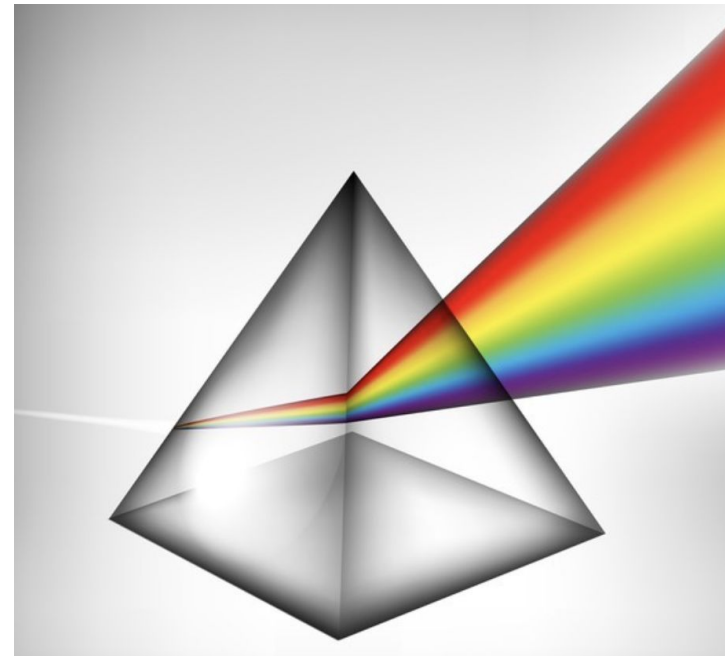


CoSTAR

- Coarse alignment algorithm for RNA secondary structures that may contain pseudoknots
 - dominant input size is number of stems, not number of nucleotides
 - directionality and relation-array constraints reduce the space
- Pruning can be enhanced
- Anchoring step for nucleotide-level alignment

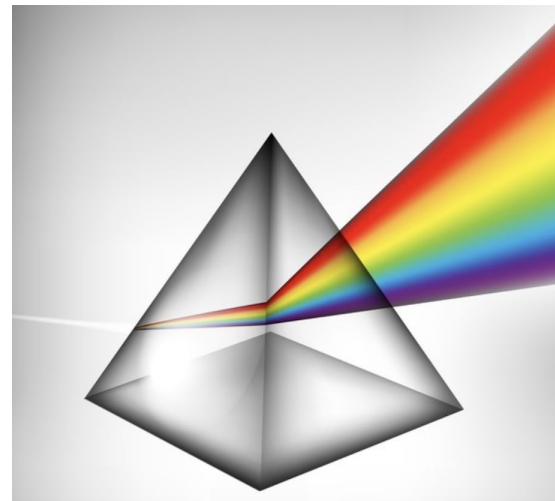


PRISM: Partition-Function Decomposition into Structural Classes for Hierarchically Constrained RNA Pseudoknot Ensembles





- CParty [Gray et al. 2025]
 - Partition function
 - Ensemble free energy
- PRISM
 - posterior probability analysis:
 - *Estimated Base Pair Probabilities*
 - **Ensemble Representations**
 - *Centroid*
 - *Maximum Expected accuracy structure*
 - *Can be guided by SHAPE reactivity*





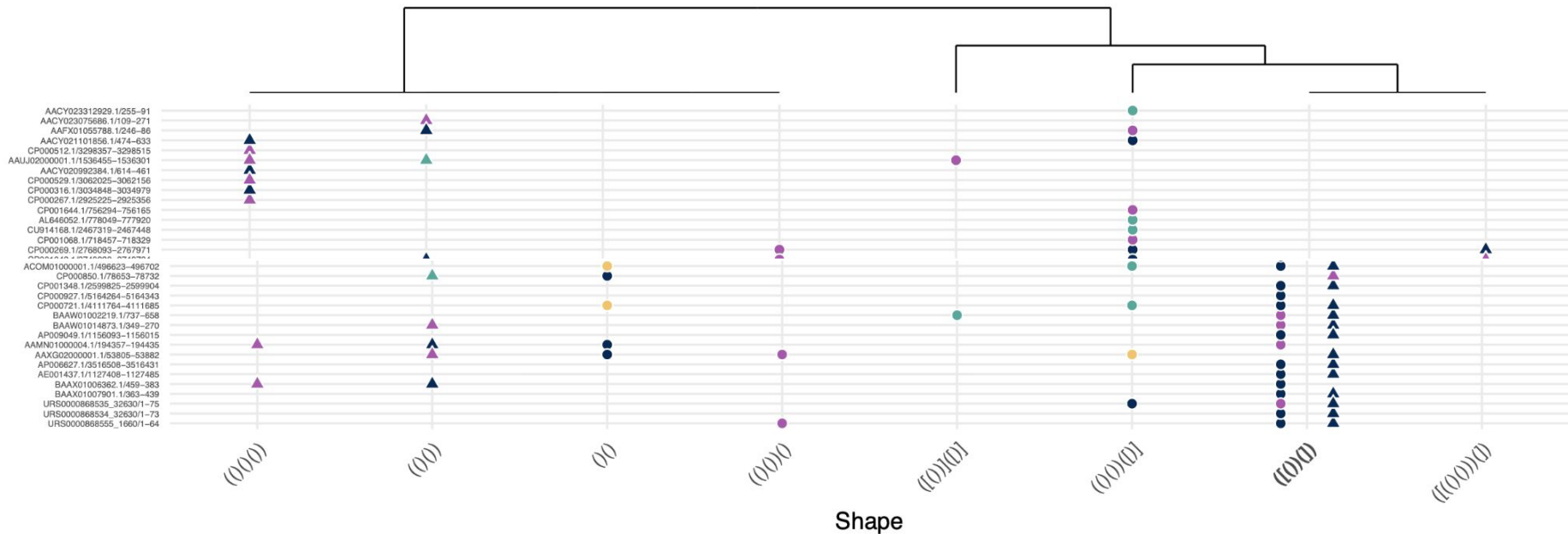
Recover pseudoknotted motifs missed by centroid

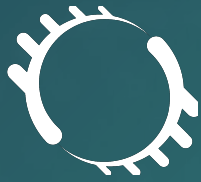
- | | | |
|-----|--|---------------------|
| (1) | GGGUGGUGUGUACCACCCCUGAUGAGUCCGAAAGGACGAA | sequence, |
| (2) | ((((((((.[[[])))))...]]).(((....))))...) | MFE structure, |
| (3) | ((((((((....))))))).....(((....))))...) | centroid structure, |
| (4) | [[]] () (0.806), () () (0.175) | top two shapes. |



RNA-shape distributions for family-level analysis

- Rfam ZMP-ZTP riboswitch





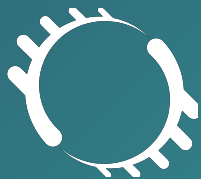
Summary

Zooming out helps to see similarity in structures (individually or ensembles)

- CoSTAR
- PRISM



github.com/TheCOBRALab



Thank you!



Sebastian Will



UNIVERSITY OF ALBERTA



Our Team



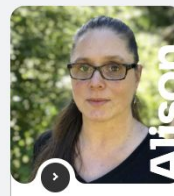
Stephani Da Costa

/ M.Sc. student, in Biomedical Engineering, University of Alberta /



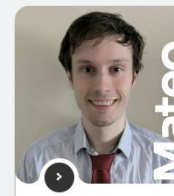
Lacey Haddon

/ Research Associate, Biomedical Engineering & Chemistry /



Alison Ziesel

/ PhD student, Biomedical Engineering /



Mateo Gray

/ PhD Student, Biomedical Engineering /



Finn Archinuk

/ PhD Student, Biomedical Engineering /



Amirali Loghmani

/ M.Sc. Student in Biomedical Engineering, University of Alberta /



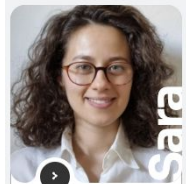
Davy Ouedraogo

/ Postdoctoral Fellow, Bioinformatics /



Leyana Haile

/ M.Sc. Student, Biomedical Engineering, University of Alberta /



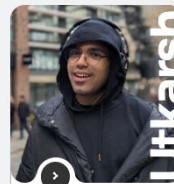
Sara Joubbi

/ Postdoctoral Research Scholar, Department of Biomedical Engineering, University of Alberta /



Sophie Sigfstead

/ B.Sc. Mathematics and Computational Science, University of Alberta /



Utkarsh Kalra

/ M.Sc. Student, Biomedical Engineering, University of Alberta /



Yahya Al-Shamali

/ Master's Student, Bioinformatics /