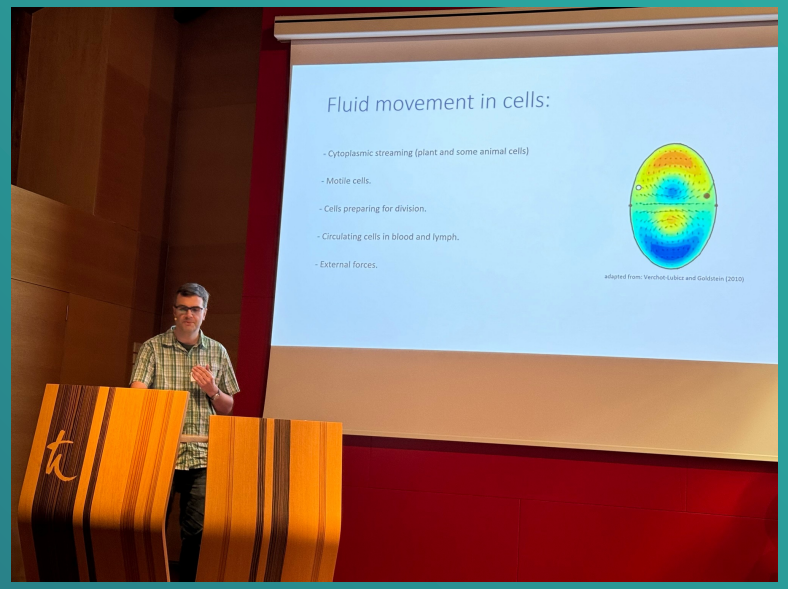
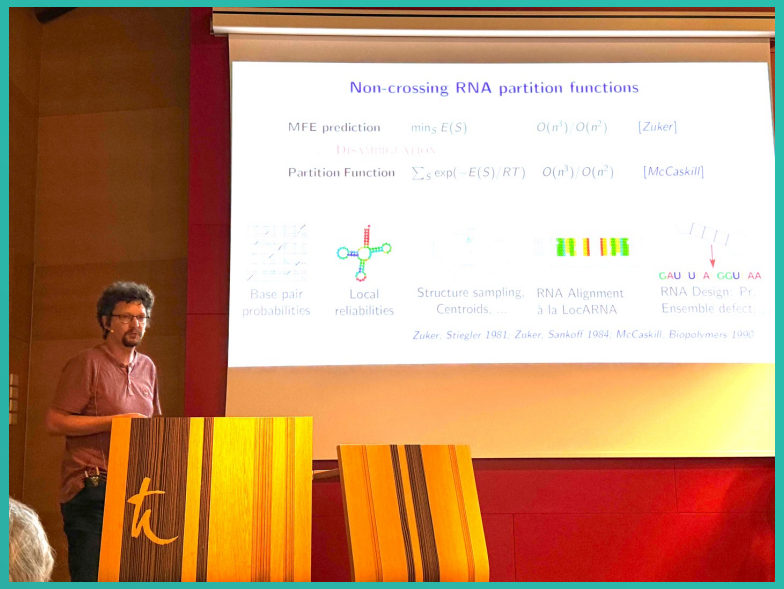
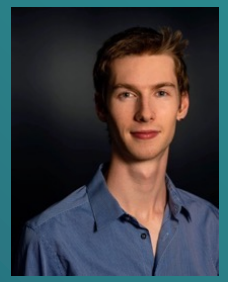
COBRA
COMPUTATIONAL BIOLOGY
RESEARCH AND ANALYTICS

Unravelling RNA function through structure prediction

Hosna Jabbari
Jabbari@ualberta.ca



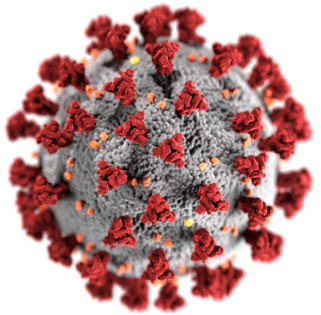




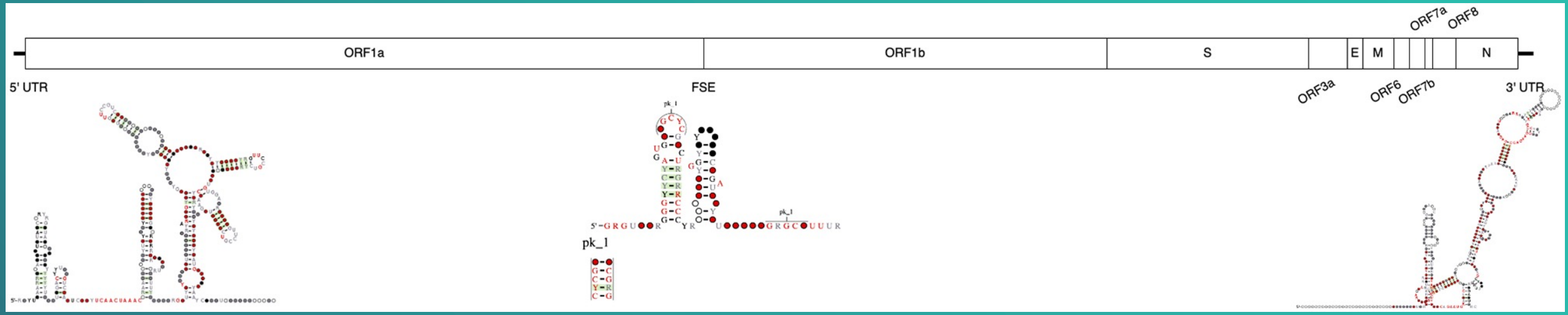
COBRA COMPUTATIONAL BIOLOGY RESEARCH AND ANALYTICS

THECOBRALAB.COM

Known Example!



- SARS-CoV-2
  - A positive sense single stranded RNA
  - Length: 30k base



Structure information from experimental data

- Incarnato laboratory [Manfredonia et al. 2020]
  - (in vivo and in vitro) SHAPE and DMS
  - 87 regions (not 3'UTR or FSE)
- Pyle laboratory [Huston et al. 2021]
  - (in vivo) SHAPE
  - 3' UTR and FSE, and dispersed throughout ORF1ab
- Cliff Zhang's laboratory [Sun et al. 2021]
  - (in vivo click) SHAPE
  - 37 structural elements
  - 5' UTR but not pseudoknot 3' UTR, and no mention of FSE.

Structure information from computational work

- Moss laboratory [Andrews et al. 2020]
  - ScanFold-based computational pipeline
  - 8 regions
- Pyle laboratory [de Cesaris et al. 2021]
  - SuperFold
  - 61% of the genome as being base paired
- Das laboratory [Rangan et al. 2020]
  - combination of CONTRAfold and RNAz
  - 44 regions (components of 5' UTR, FSE, and 3' UTR)
- Mathews laboratory [Li et al. 2021]
  - LinearTurboFold
  - 50 elements, 26 of which were not in previous studies

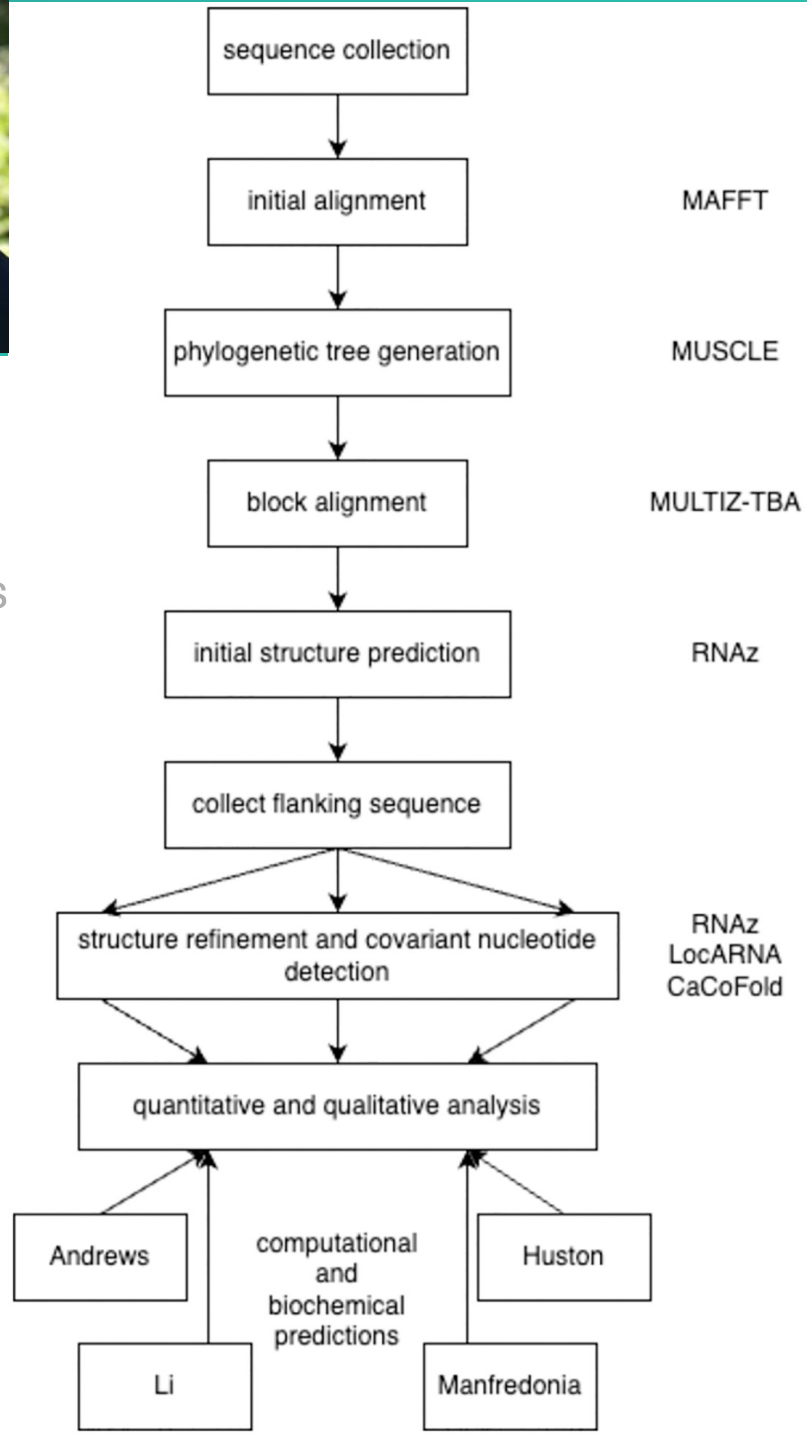
- RNA Puzzles [Gumna et al. 2022]
  - 74 3D models of UTRs, including some predicting pseudoknots in 3' UTR, and some extended 5' UTR structures beyond the currently accepted 5 stem loops

- Little is known about structures in coding regions

Structural Pipeline

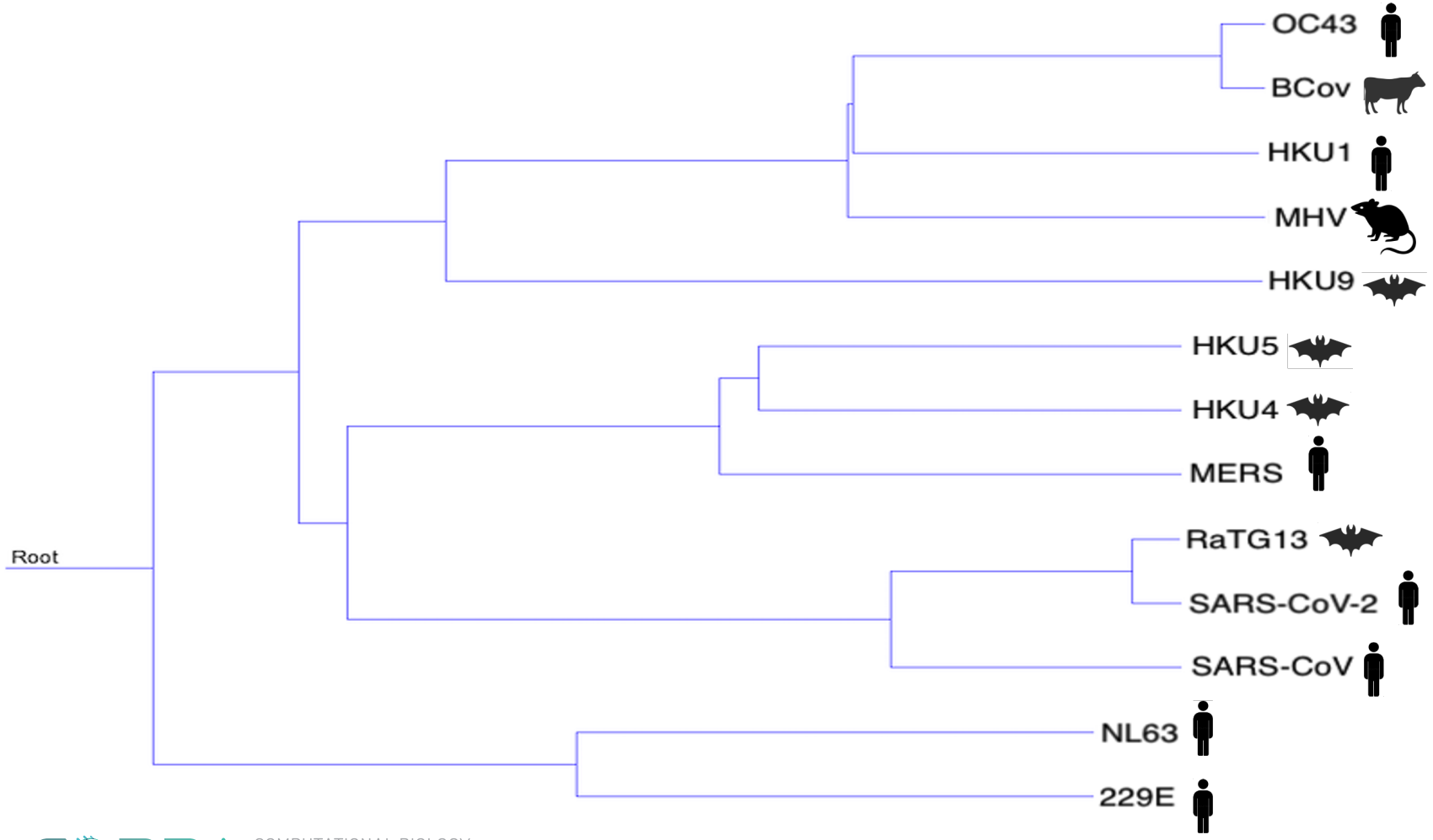
- Started with a known pipeline
  - Conserved structures in vertebrates [Thiel et al. Genes 2018]
  - Conserved structures in non-coding RNA [Will et al. Genome Res 2013]

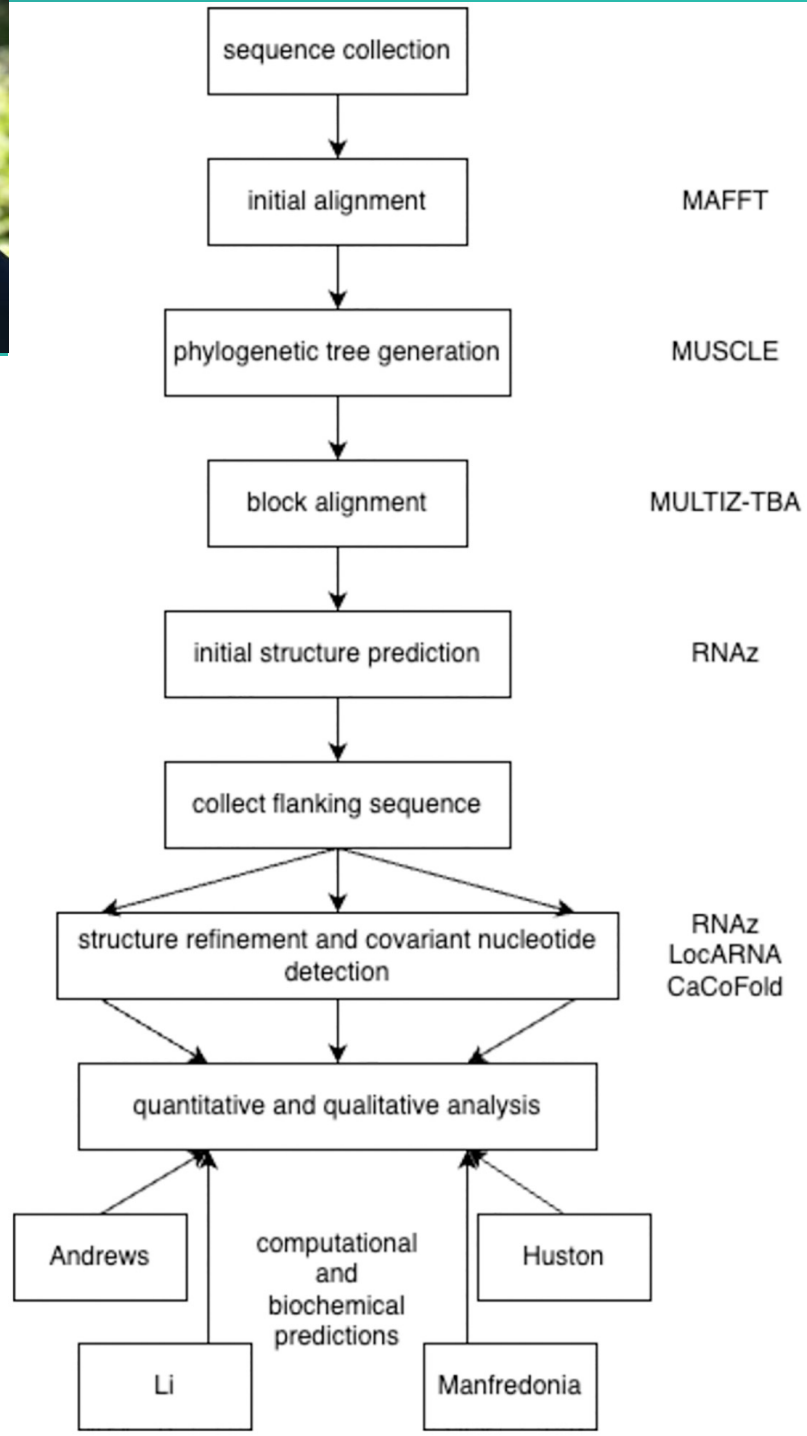
- Modified it for shorter sequences
  - More stringent criteria
  - Added Cacofold [Rivas PLoS Comp Bio. 2020]
  - Cross referenced findings with others



Ziesel and Jabbari 2024

Structural Pipeline

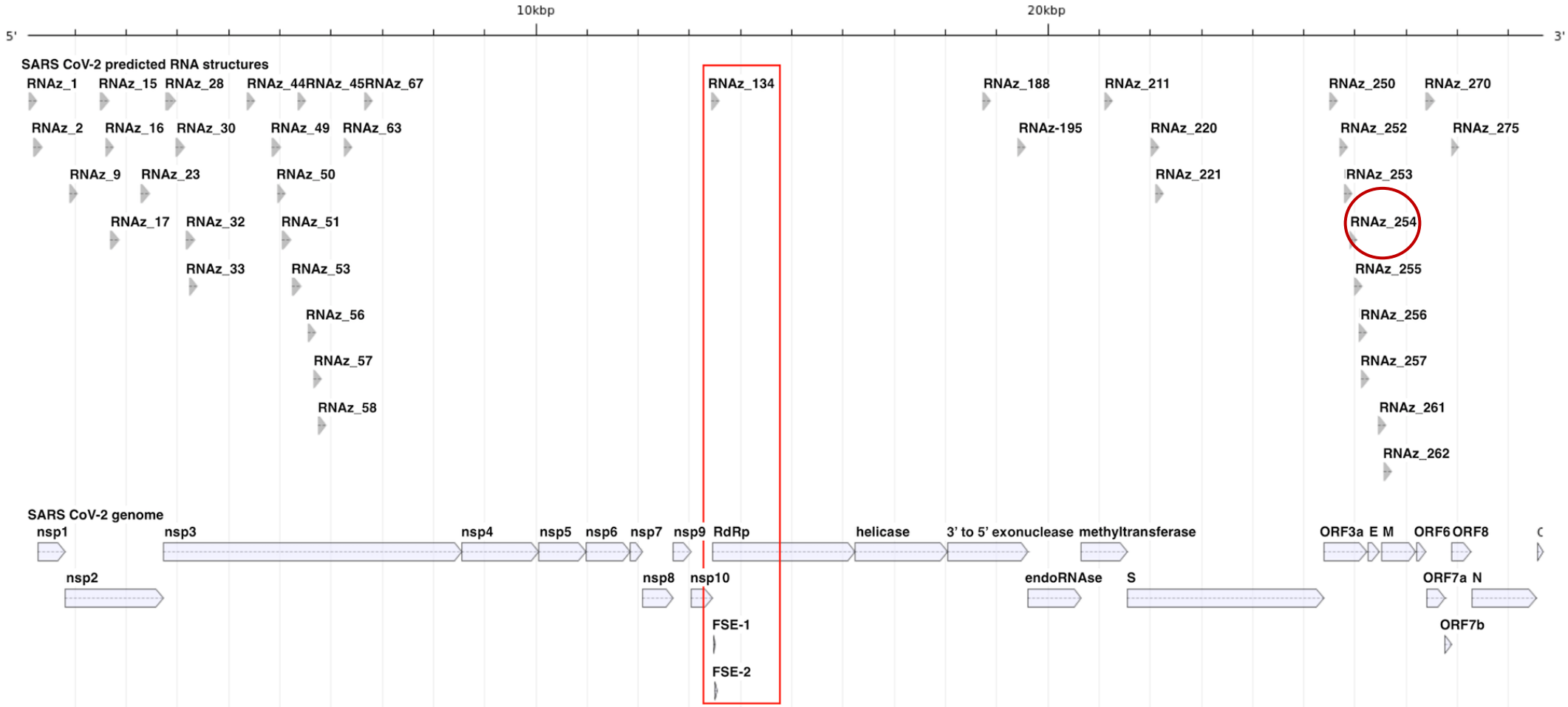




COBRA COMPUTATIONAL BIOLOGY RESEARCH AND ANALYTICS

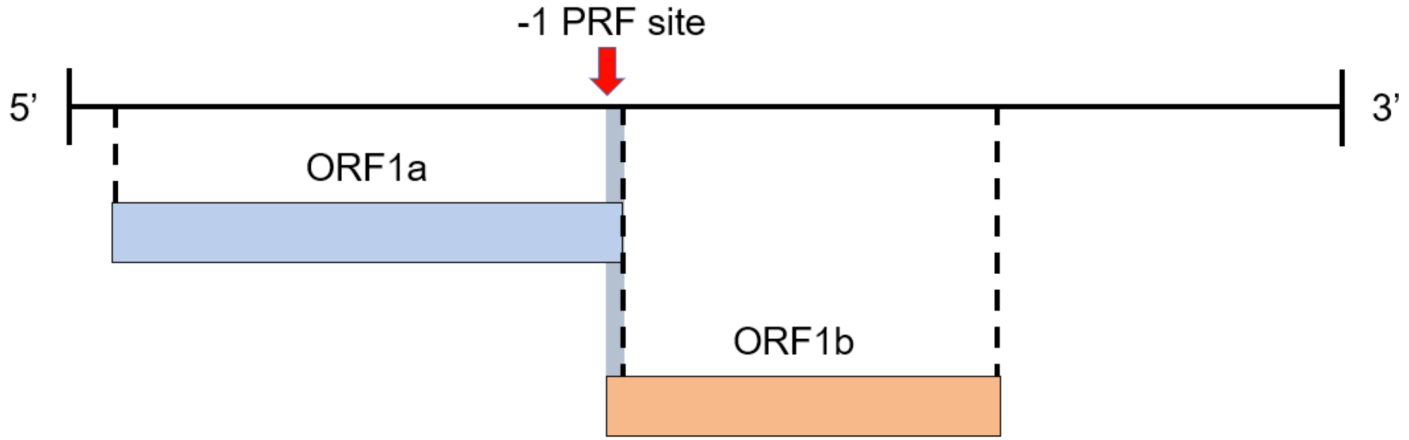
Ziesel and Jabbari 2024

Structure Comparison



Frameshifting Pseudoknot

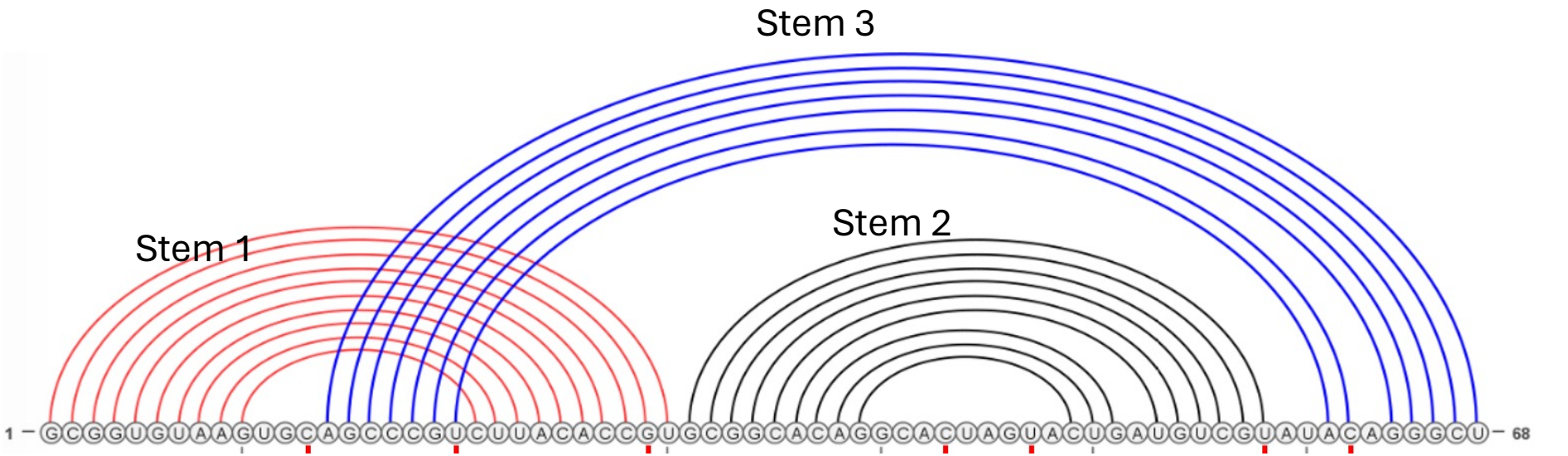
- Key in replication of coronaviruses: *ribosomal frameshifting*
- SARS-CoV-2 has 2 ORF
  - Normally only one translated
  - Pseudoknot facilitates frameshifting and translation of both!



What we know

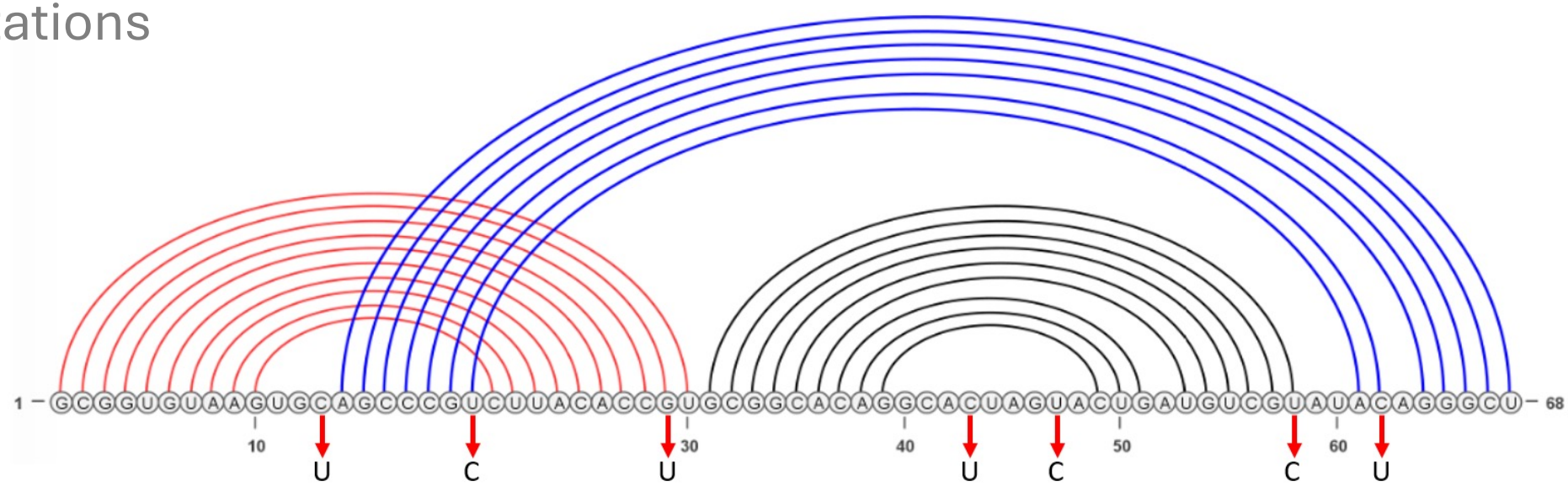
- SARS-CoV −1 PRF conserved in SARS-CoV-2
  - only a single base difference

- Pseudoknot plasticity



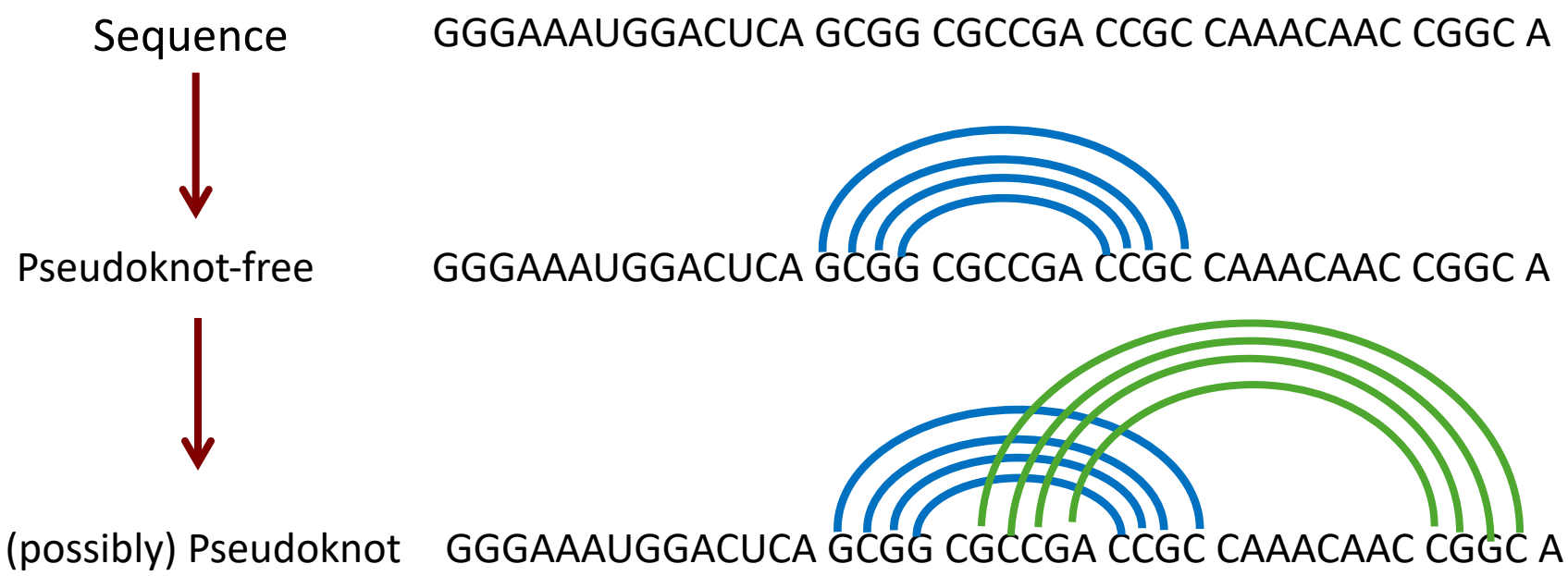
What we don’t know

- Characterization of the structure and its pathways
  - a sampling of land- scape of previously unidentified–non-native structures
  - Following a hierarchical folding view
  - Incorporating experimental data (SHAPE)
- Effects of mutations



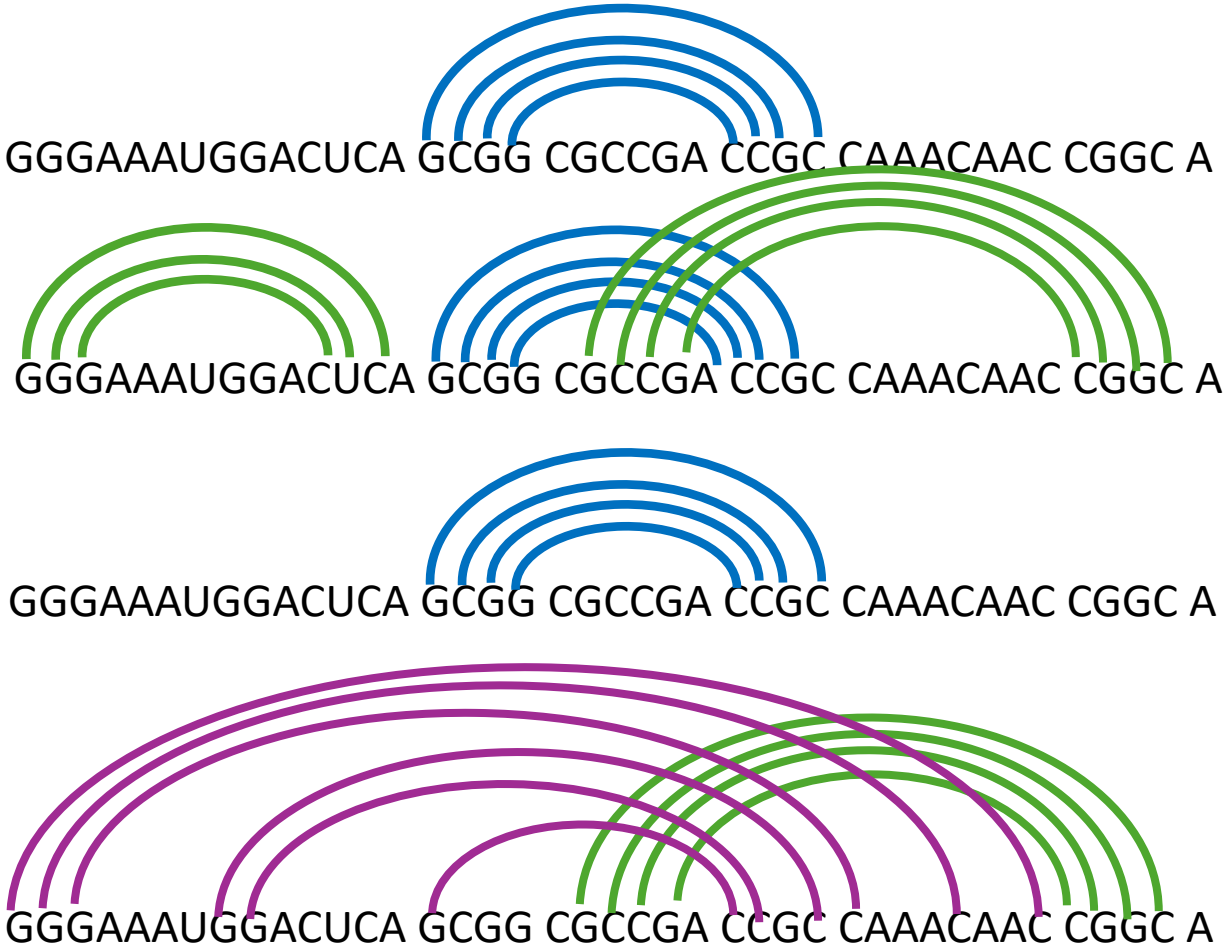
Hierarchical Folding

- Evidence that RNA structure formation is hierarchical [Tinoco & Bustamante 1999, Cho et al. 2009, Bailor et al. 2010]



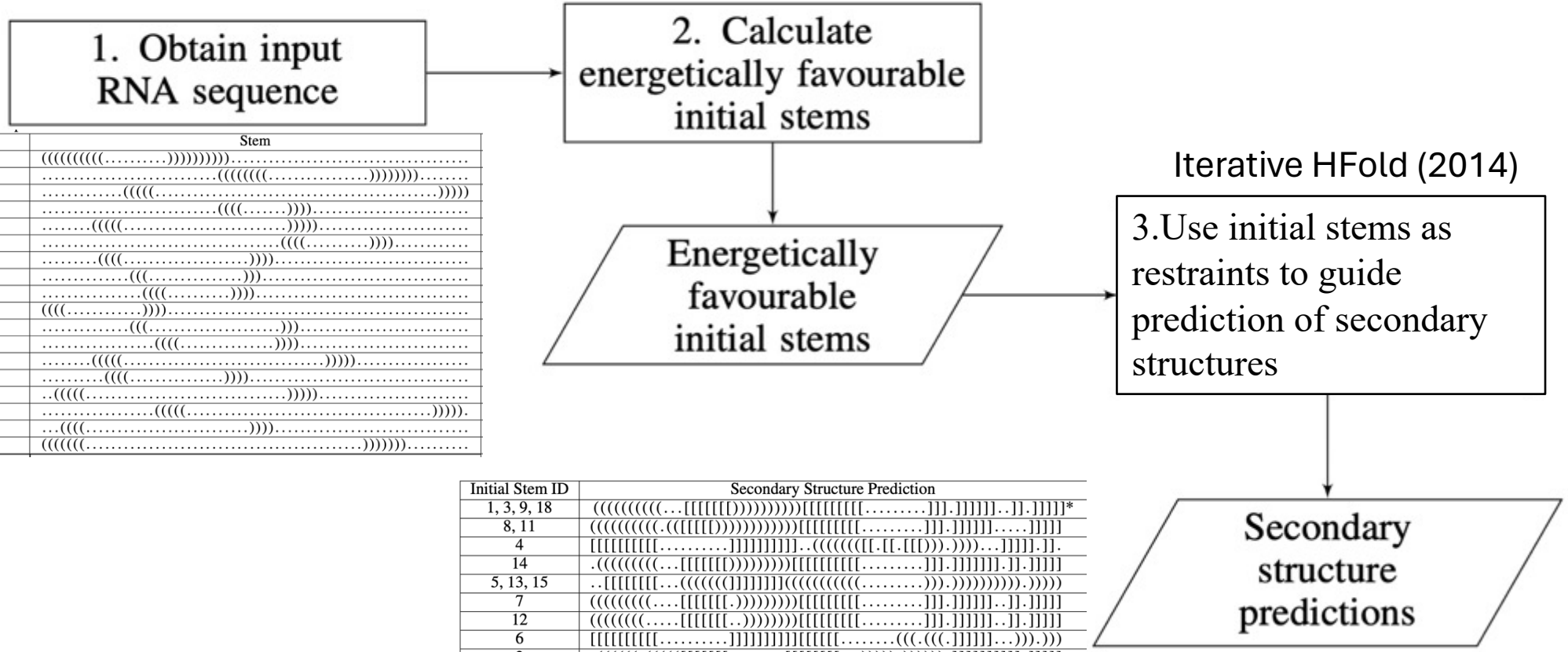
Hierarchical folding-based methods

- HFold [Jabbari et al. 2008]
  - Input: RNA Seq + a pk-free structure
  - Input structure as a **constraint**
  - Output: MFE given input
- Iterative HFold [Jabbari et al. 2014]
  - Input: RNA Seq + a pk-free structure
  - Input structure as a **restraint**
  - Output: MFE guided by the input



SARS-CoV-2 PRF Structure Prediction

GCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGCU

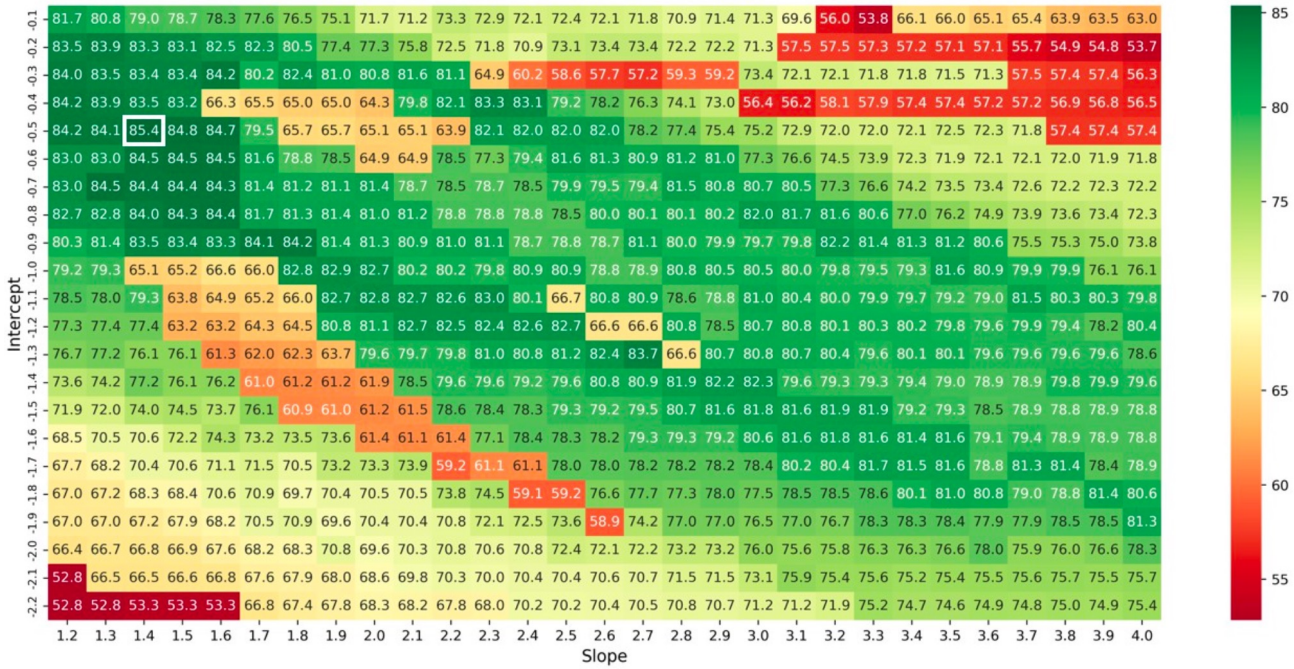


ID	Stem
1	(((((((((.....))))))))).....
2	.....(((((((.....))))))).....
3	.....(((((.....)))))
4	.....((((.....)))).....
5	.....(((((.....))))).....
6	.....((((.....)))).....
7	.....((((.....)))).....
8	.....(((.....))).....
9	.....((((.....)))).....
10	((((.....)))).....
11	.....(((.....))).....
12	.....((((.....)))).....
13	.....(((((.....))))).....
14	.....((((.....)))).....
15	..(((((.....))))).....
16	.....(((((.....))))).
17	...((((.....)))).....
18	(((((((.....)))))).....

Initial Stem ID	Secondary Structure Prediction
1, 3, 9, 18	((((((((((...[[[[[[[)))))))))))[[[[[[[[[.....]]].]]]]]]..]].]]]]]*
8, 11	((((((((((.(([[[[[)))))))))))))[[[[[[[[[.....]]].]]]]]].....]]]]]
4	[[[[[[[[[[.....]]]]]]]]]]..(((((((([.[[.[[[))).))))...]]]]].]].
14	.((((((((...[[[[[[[)))))))))[[[[[[[[[[.....]]].]]]]]]].]].]]]]]
5, 13, 15	..[[[[[[[[...(((((((]]]]]]]]((((((((((.....))).))))))))).)))))
7	((((((((....[[[[[[[.)))))))))[[[[[[[[[.....]]].]]]]]]..]].]]]]]
12	((((((((.....[[[[[[[..)))))))))[[[[[[[[[.....]]].]]]]]]..]].]]]]]
6	[[[[[[[[[[.....]]]]]]]]]]]][[[[[[.....(((.(((.]]]]]]...))).)))
2	.(((((((.((((([[[[[[[.....[[[[[[[[...)))))).))))))).]]]]]]]]]].]]]]]
10	.....(((((([[[[[[[.....[[[[[[[[[[[)))))....]]].]]]]]]]]]].]]]]]
16	[[[[[[[.....((((()]]]]]]]][[[[[[[[[.....]]].]]]]]]....))))).
17	.....[[[[(((((....(((((.((((.]]]]..))))..)).....)))....)))))
Input sequence	GCGGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGCU

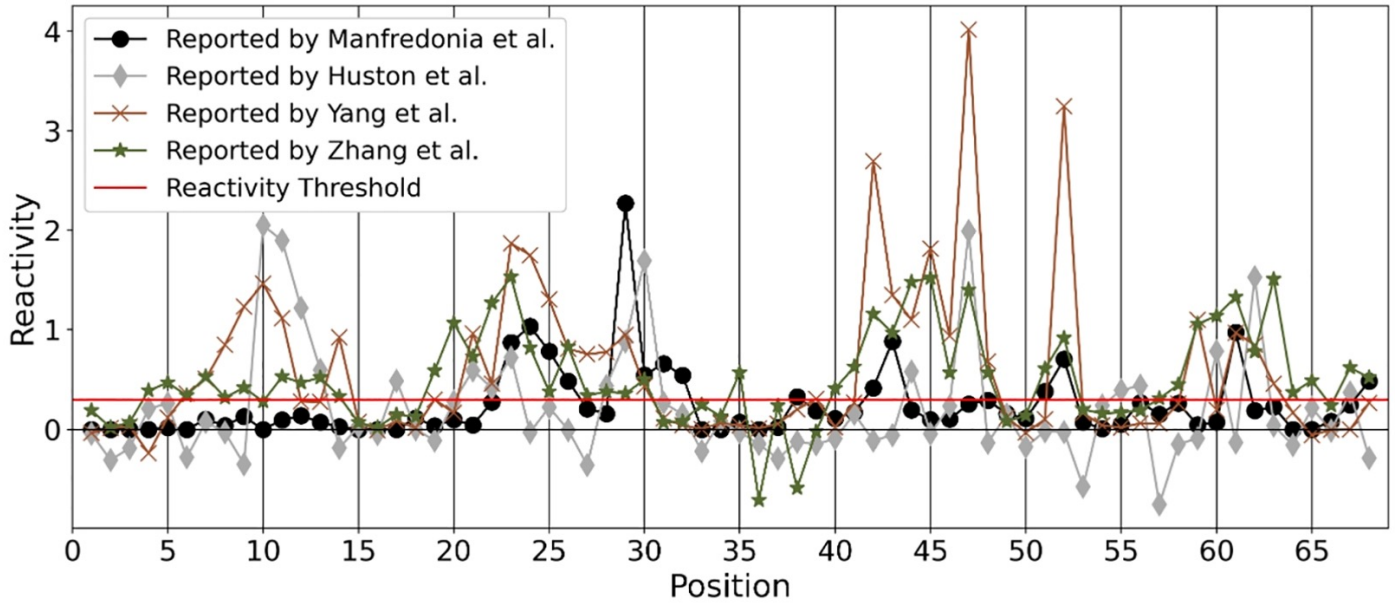
Incorporating SHAPE

- ShapeKnots data
  - 5 RNAs > 300
  - 5 riboswitches
  - 4 hard to predict
  - 3 RNAs in interactions
- Extended
  - 6 from RNA Mapping Database
  - 3 riboregulators
  - SARS-CoV2 3’ and 5’ UTR
  - ribonuclease domain of Bacillus subtilis

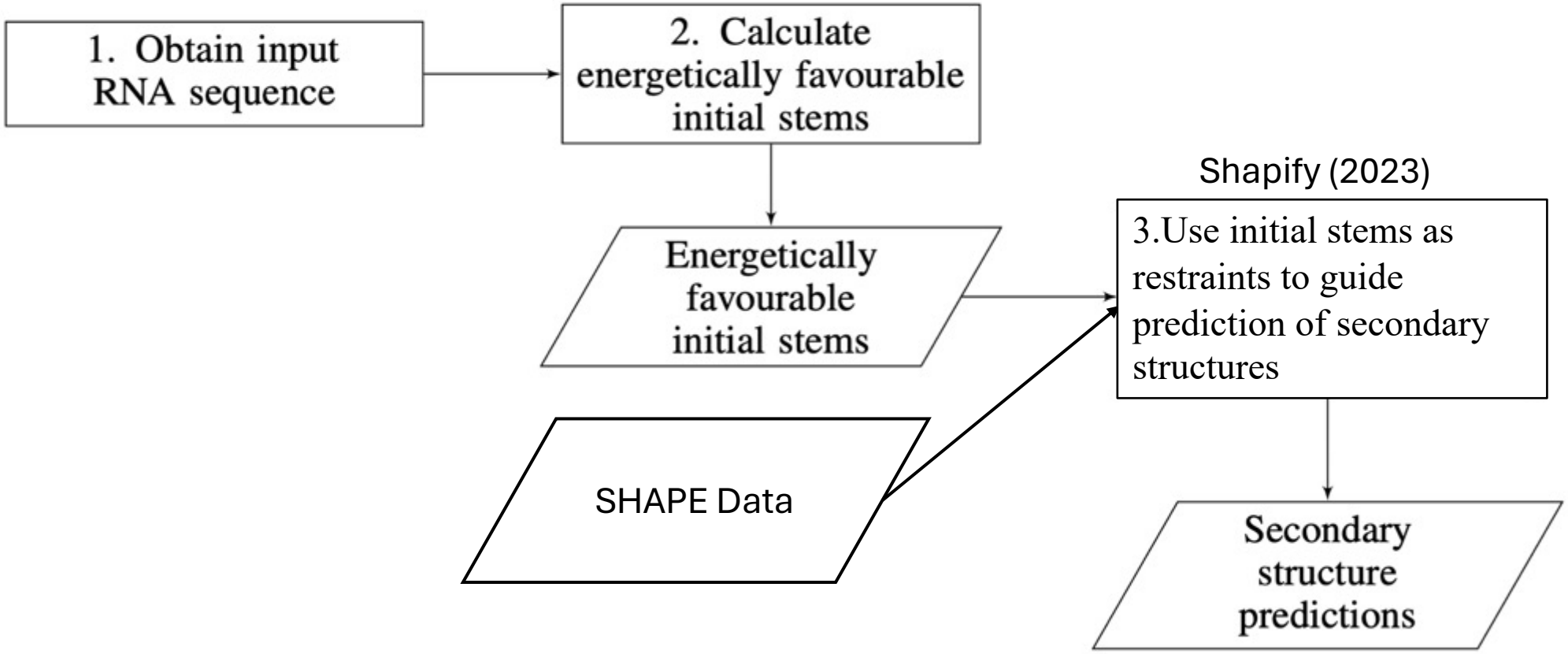


Shapify [Trinity et al. 2023]

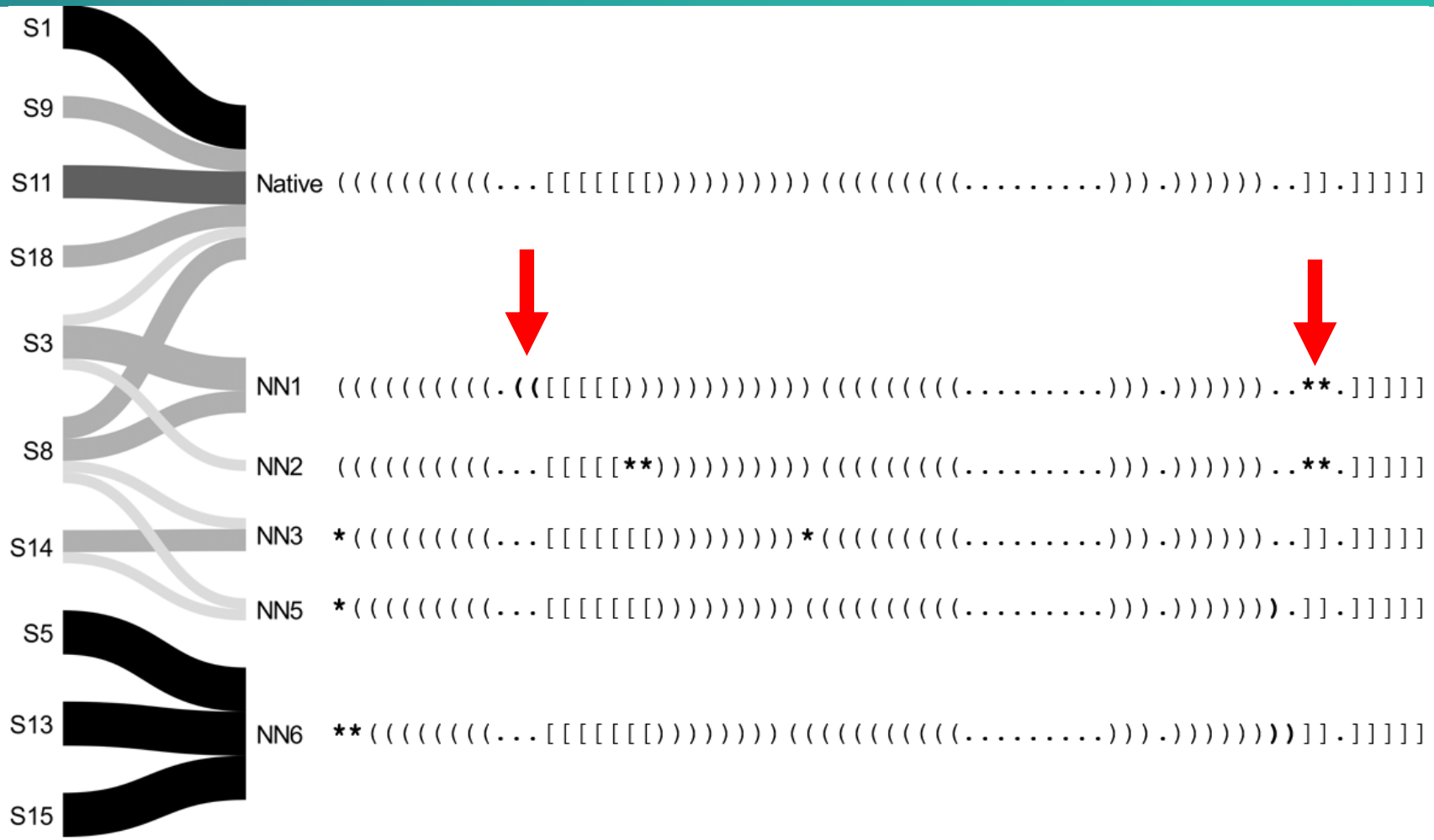
- Huston 2021 (in vivo, genome-wide)
- Manfredonia 2020 (in vivo, genome-wide)
- Yang 2021 (in vivo, genome-wide)
- Zhang 2021 (in vitro, PRF)



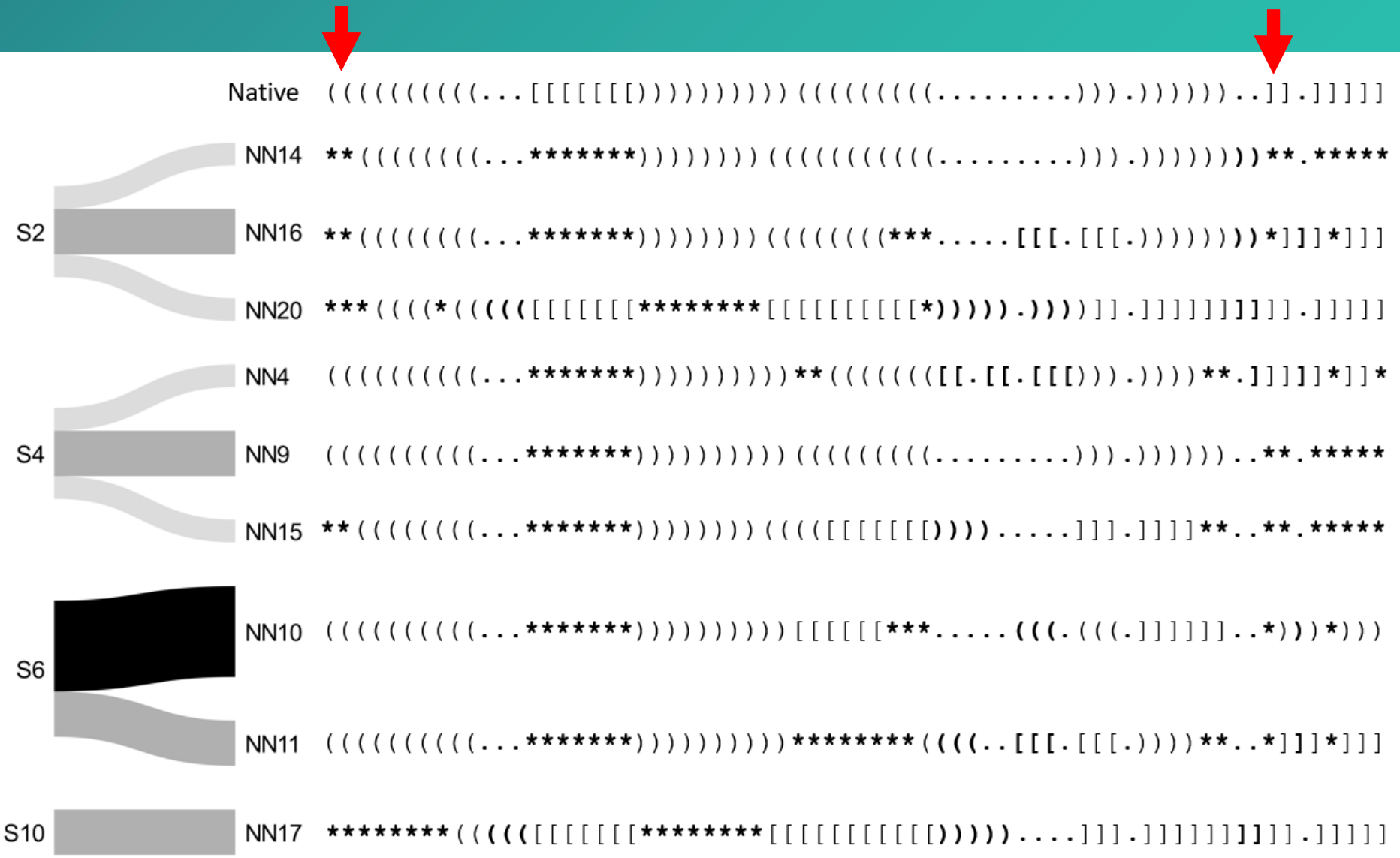
SARS-CoV-2 PRF Structure Prediction



Native-adjacent Structures

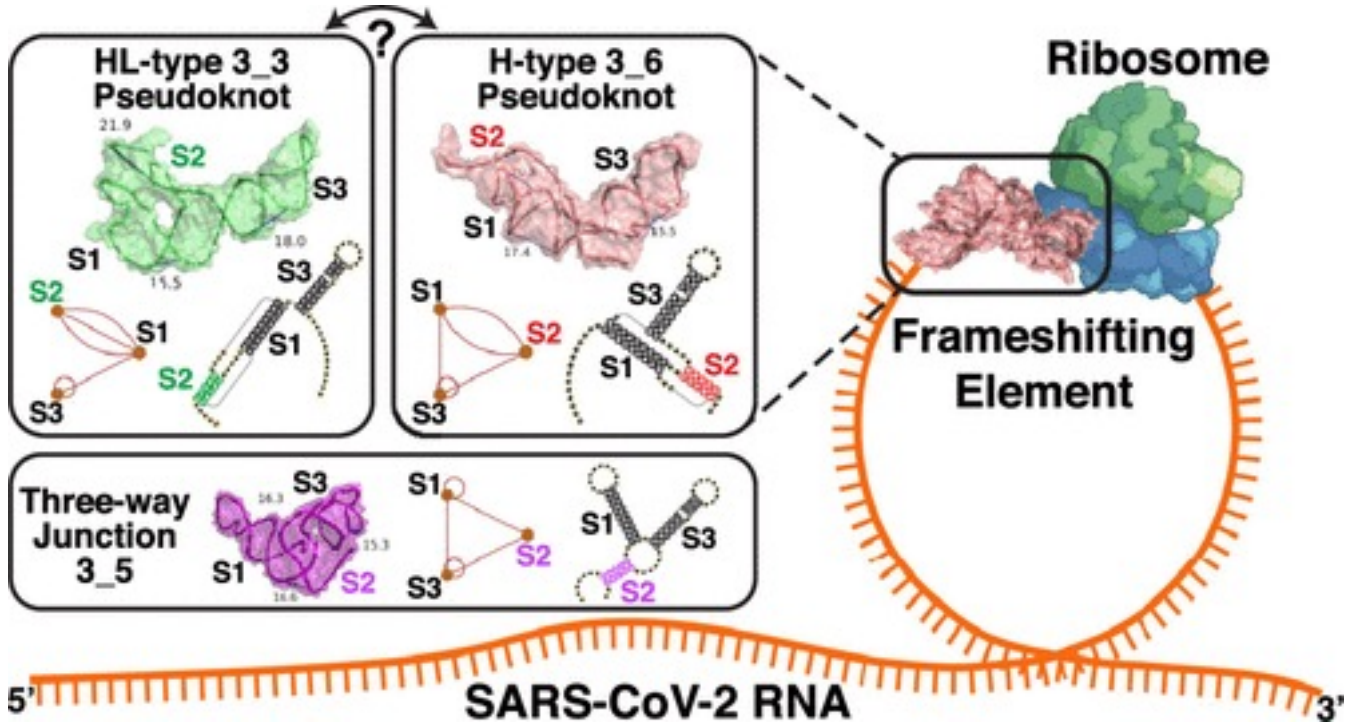


Native Non-adjacent Structures



What others found

- Schlick et al. identified 3 motifs [J Am Chem Soc. 2021]
  - SHAPE
  - ‘3_6’ motif corresponds with NN1, NN2, and NN4.
- Bhatt et al. [Science 2021]
  - cryo-EM
  - Structure that matched ‘3_6’ motif



Dual Graphs

2_1

((((.....))))((((.....))))

2_2

((((...((((...))))...))))

2_3

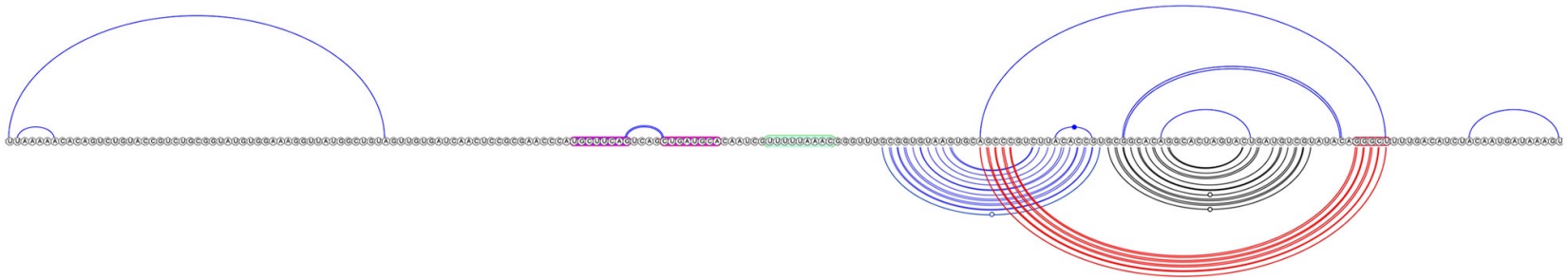
((((...[[[[...))))...]]]]

Tying the Knot [Trinity et al. 2024]

- Expanded FSE window size
  - 90 to 222 nt

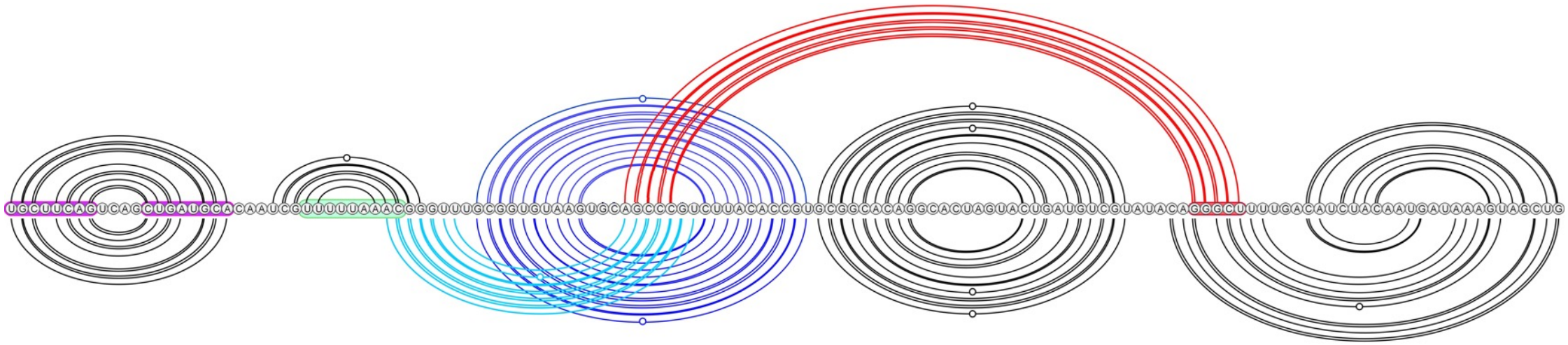
$$\text{Free energy per nt} = \frac{\text{Structure free energy}}{\text{Sequence length}}$$

- Considered covariation
  - KnotAli [Gray et al. 2020]
  - MSA obtained from Schlick et al. JACS 2021
- Considered refolding



144 nt window

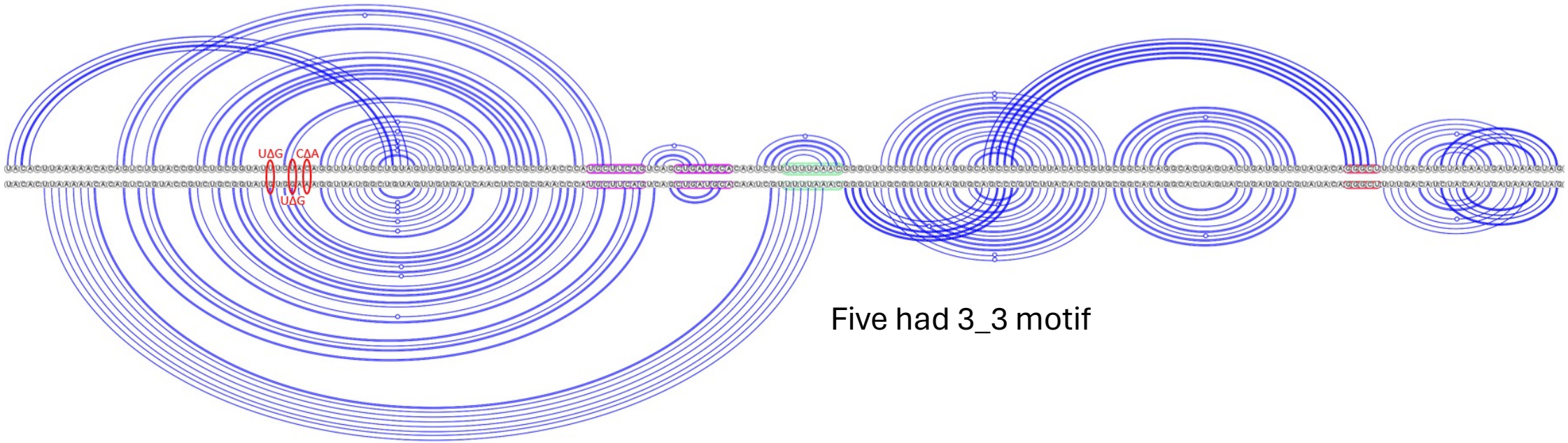
3_6 motif – Native structure



3_3 motif

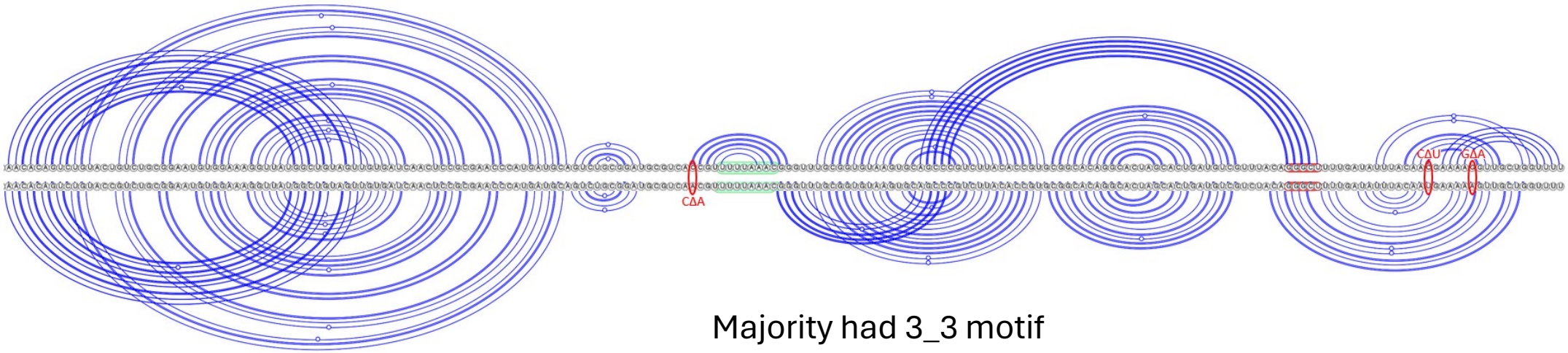
Trinity et al. 2024

Incorporating Covariation – 7 SARS-CoV2 sequences



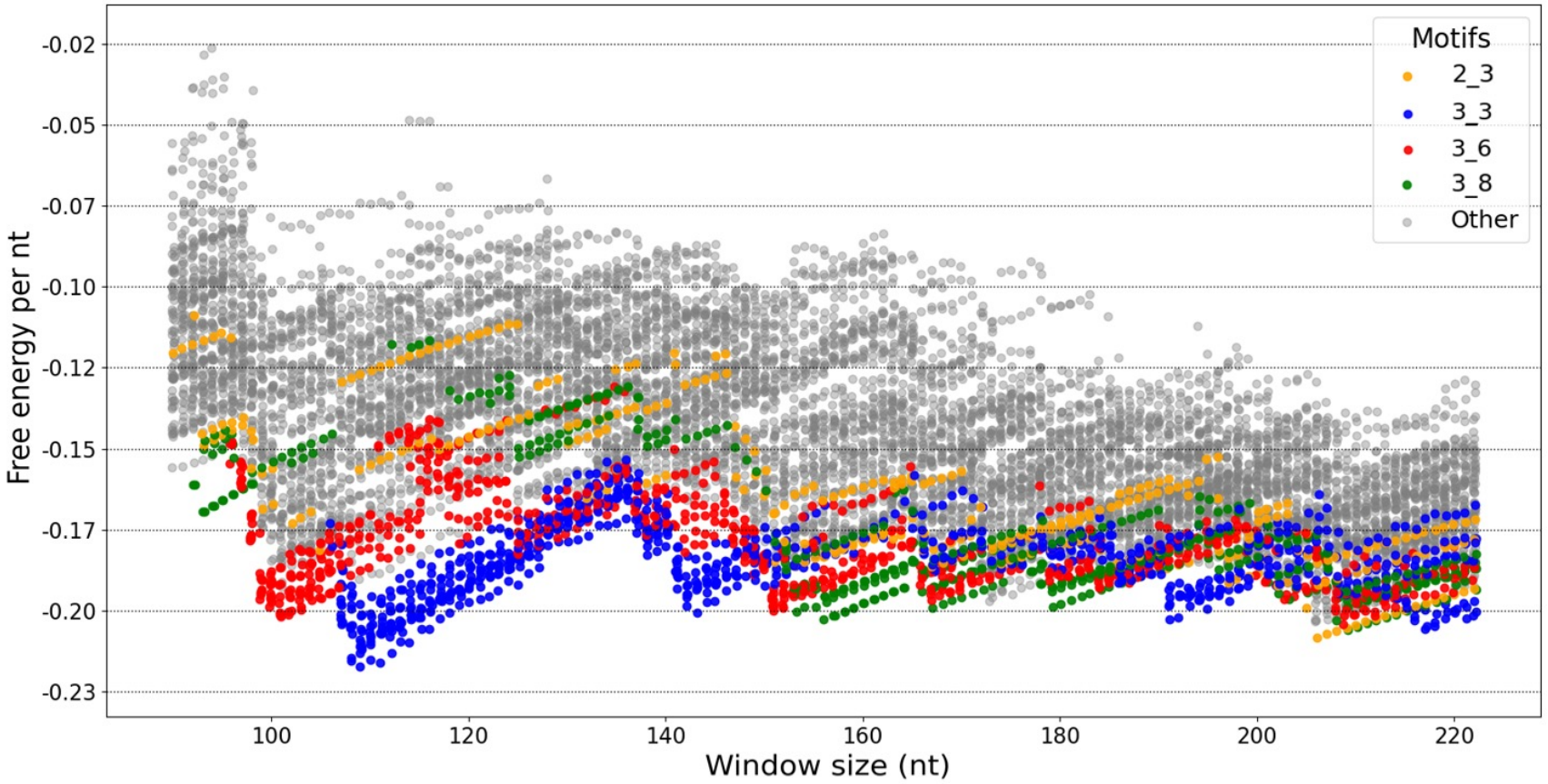
Five had 3_3 motif

Incorporating Covariation – sarbecoviruses



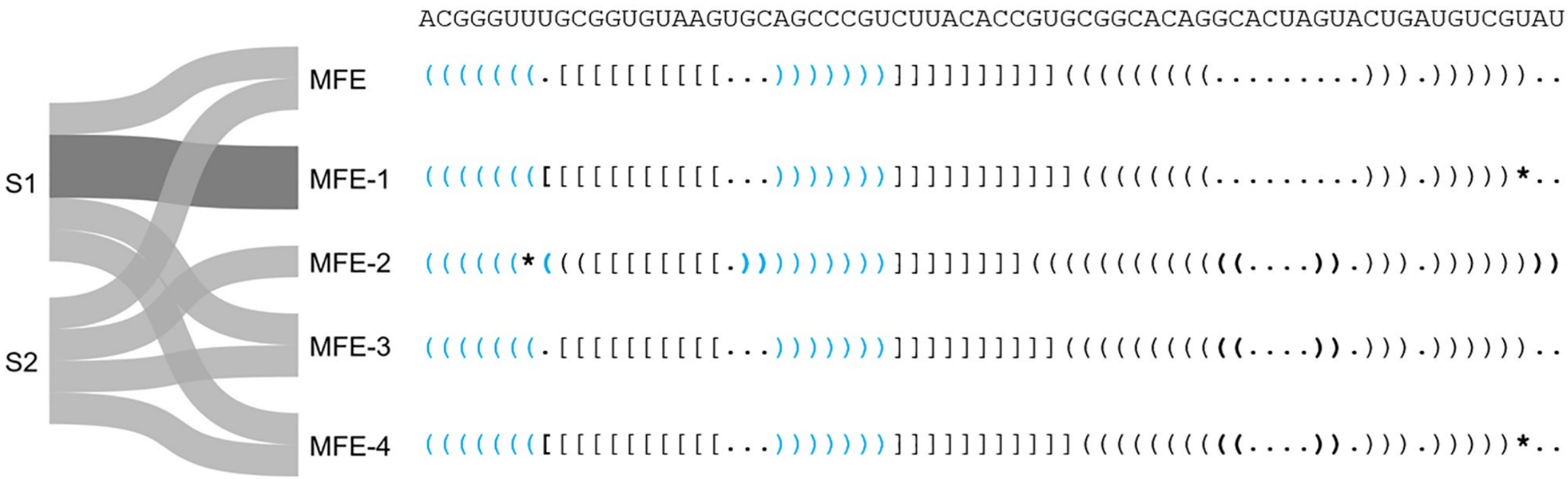
Majority had 3_3 motif

Multiple Structures – Varying Window Size and SHAPE

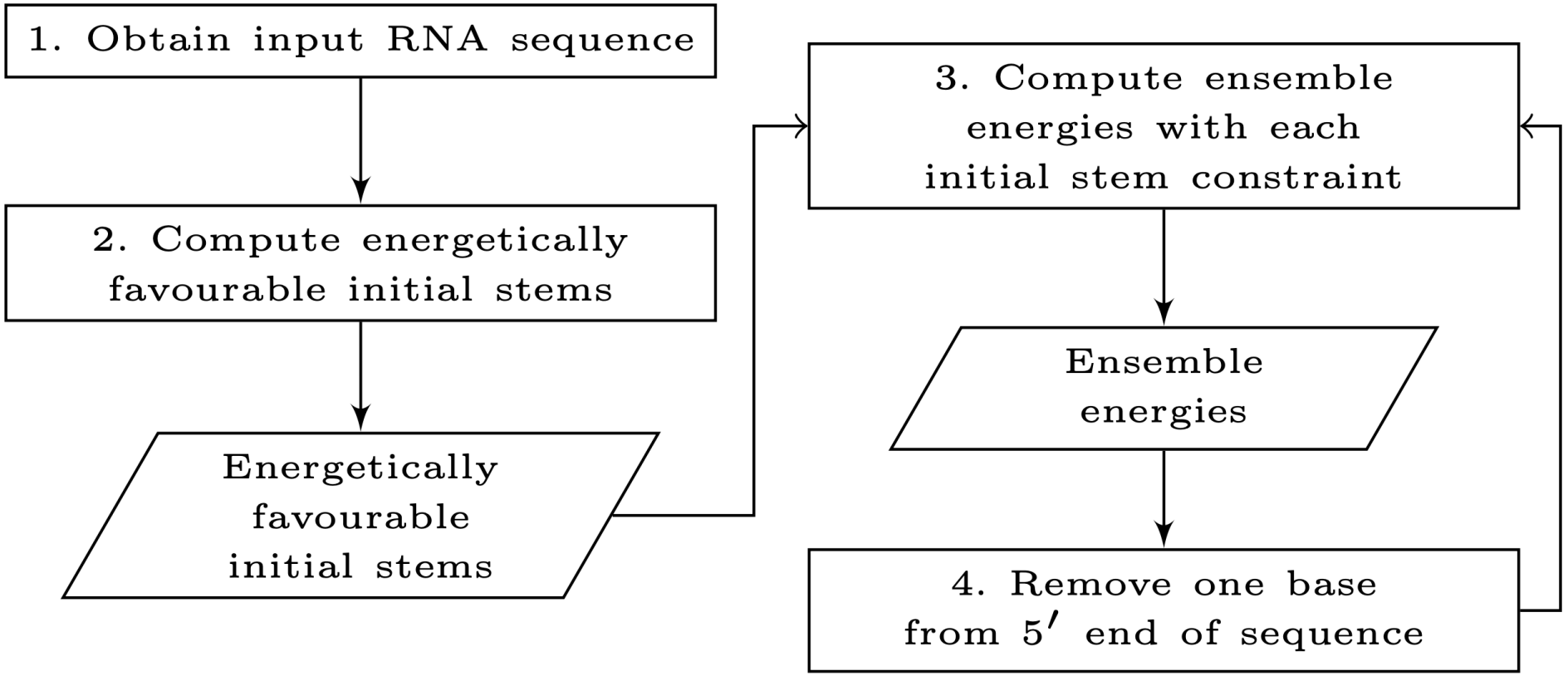


~11k sequences

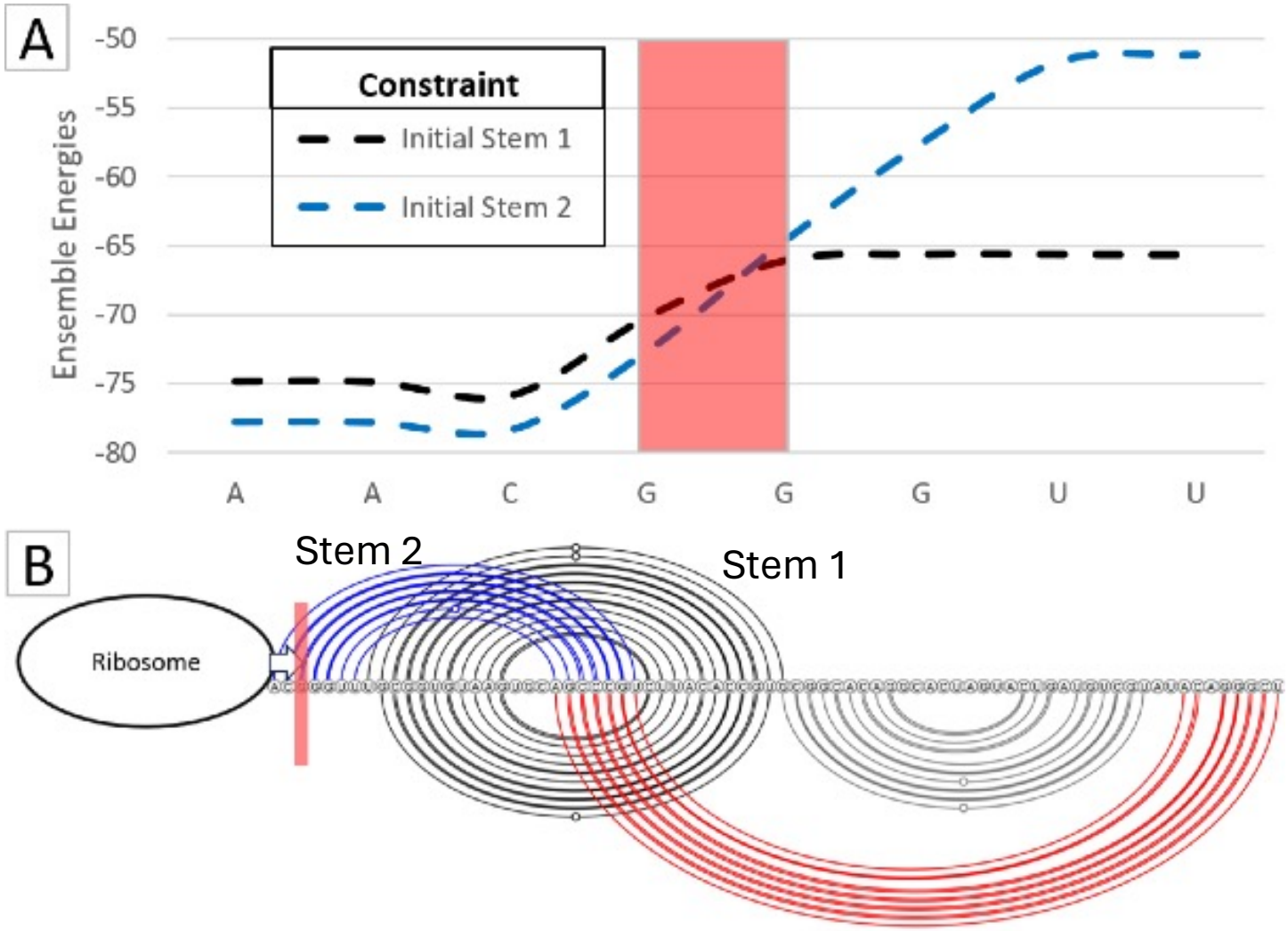
Path convergence for 3_3 motif



Simulating refolding– CParty [Trinity et al. 2024, submitted]



Decreasing sequence length from 77 to 70



CParty

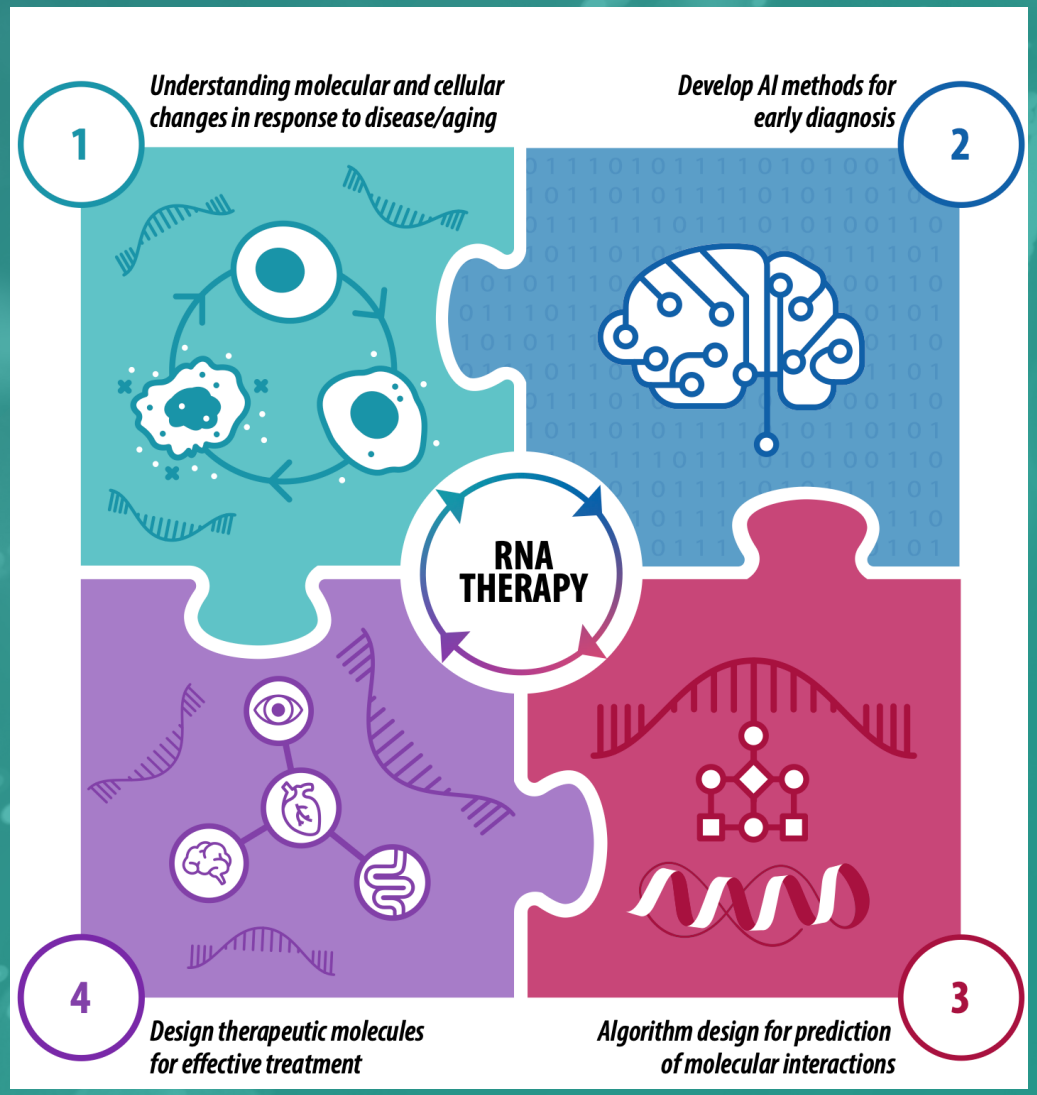
Iterative HFold

What we found

- Conserved Similarity
- Structural paths
- SHAPE incorporation
- Covariations and length dependence
- Motif transitions

We are hiring PhD students and Postdocs!

COBRA
COMPUTATIONAL BIOLOGY
RESEARCH AND ANALYTICS



THECOBRALAB.COM

Thank you!

Sebastian Will

Ulrike. Stege

Yan Ponty

All members of the COBRA Lab

UNIVERSITY OF ALBERTA