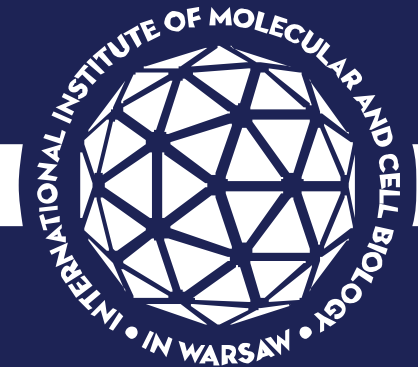


DesiRNA:

Application to EteRNA benchmark, bacterial and viral structures

Tomasz Wirecki

Laboratory of Bioinformatics and Protein Engineering
International Institute of Molecular and Cell Biology in Warsaw



Benasque 24.07.2024


Memories...



*People who love to eat
are always
the best people.*

Julia Child

Outline

- 
- 1. Introduction**
 - 2. The Past**
 - 3. The Now**
 - 4. Results**
 - 5. Conclusions**

Outline

1. Introduction

2. The Past

3. The Now

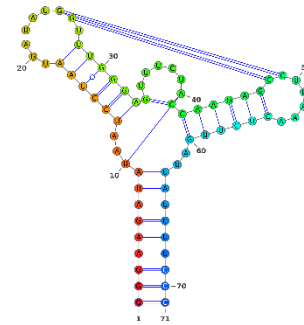
4. Results

5. Conclusions

RNA 2D structure prediction

Given a target sequence, predict its 2D structure

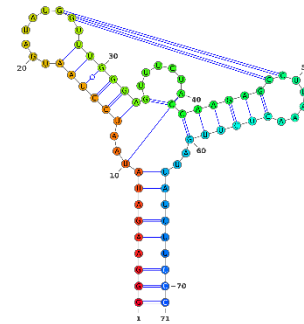
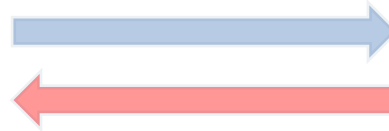
GGGAAGAUUAAUCCUAA
UGAUAUGGUUUGGGAGUU
UCUACCAAGAGCCUAAA
CUCUUGAUUAUCUCC



RNA 2D sequence design

Given target 2D structure, predict a sequence that folds to form that structure

GGGAAGAUUAAUCCUAA
UGAUAUGGUUUGGGAGUU
UCUACCAAGAGCCUAAA
CUCUUGAUUAUCUCC



RNA Design – brute force?



Brute force approach – $4^u \cdot 6^{p/2}$ for an RNA having u unpaired and p paired residues ...

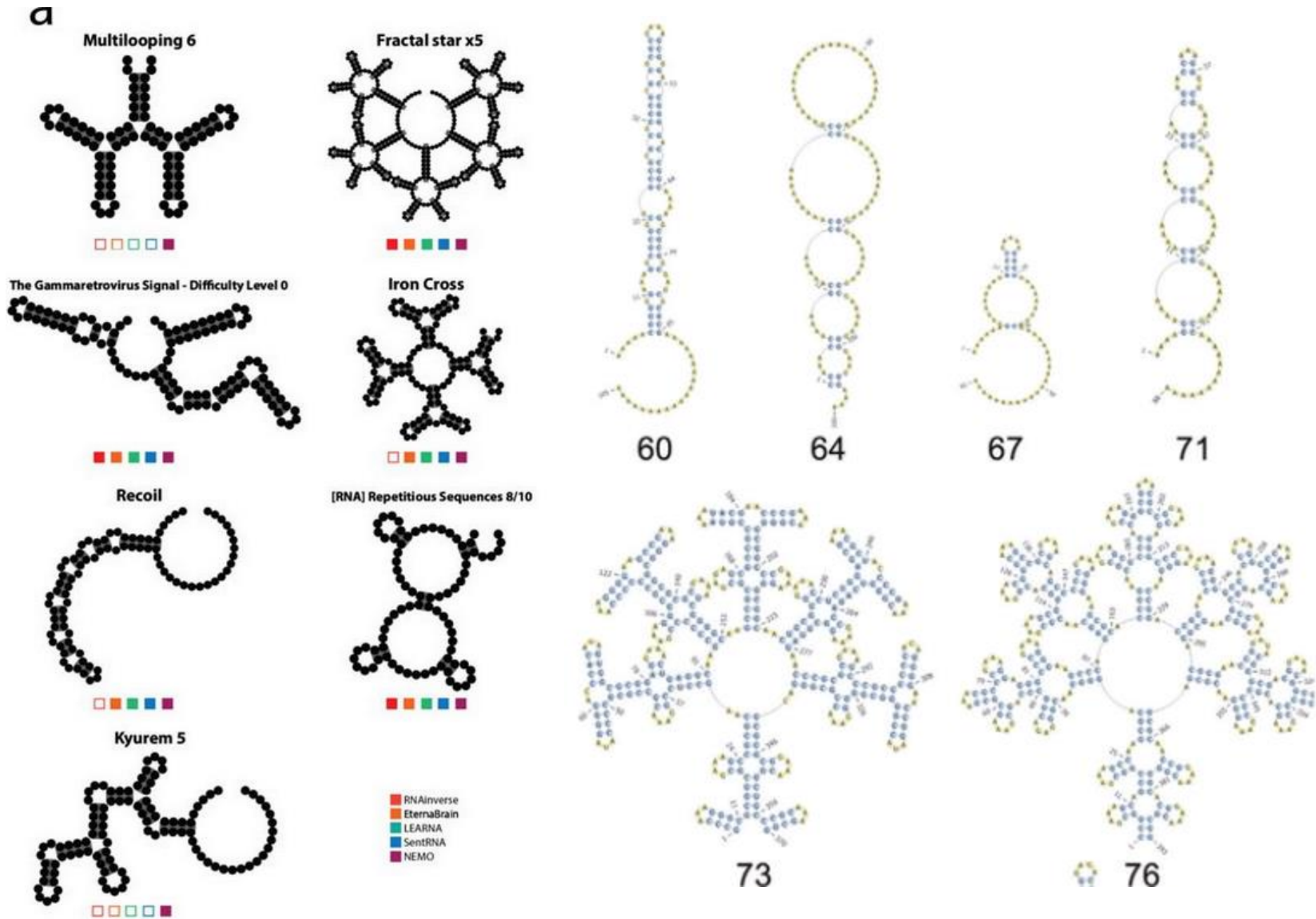
((.....)) $nt = 10, u = 6, p = 4$

No. of possible sequences = **147456**

(((((...((.....))...)))))) $nt = 22, u = 10, p = 12$

No. of possible sequences = **4.98^{10} !!!**

EteRNA benchmark



RNA Design – existing methods - limitations

- No pseudoknots
- No multimer design
- Not checking for oligomerization
- No temperature dependency
- No alternative structures
- No „natural” sequences

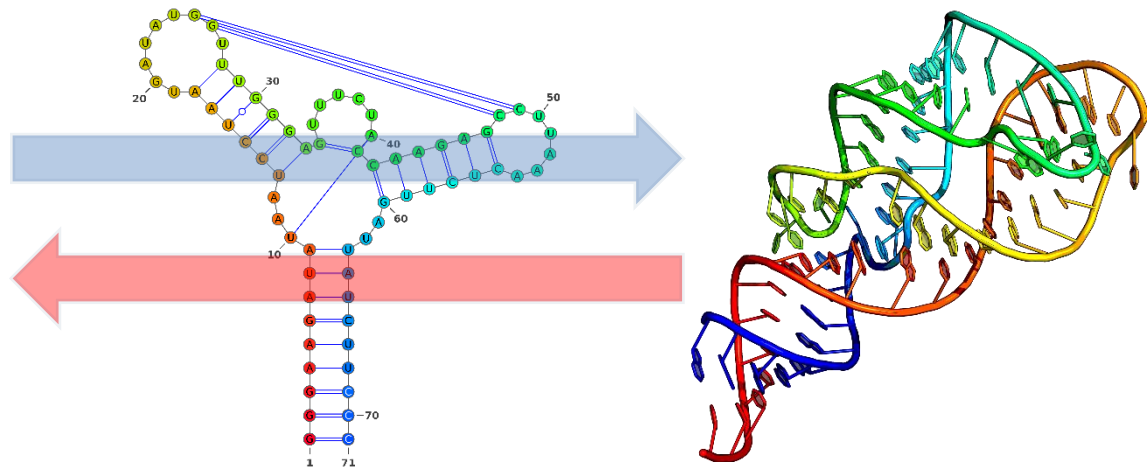


DesiRNA – 2D design

DesiRNA – algorithm for secondary structure design

- use both positive and negative design
- an algorithm which maximizes the appearance of the target secondary structure in the thermodynamic ensemble
- consider oligomerization (monomers vs homooligomers)
- designs alternative structures

GGGAAGUAU
UAAUCCUAA
UGAUAUGGU
UUGGGAGUU
UCUACCAAG
AGCCUUAAA
CUCUUGAAU
AUCUUCCC



Outline

1. Introduction

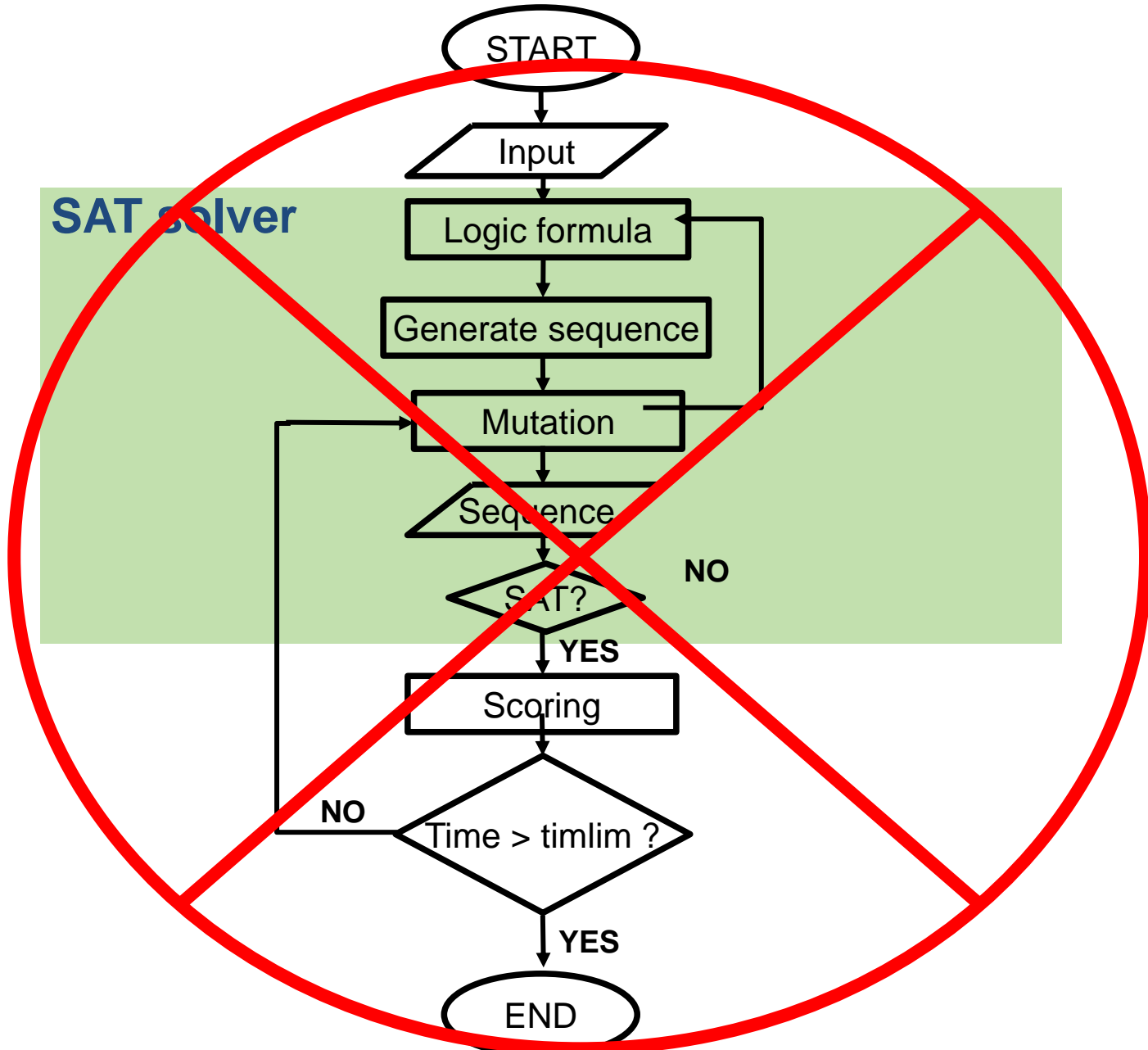
2. The Past

3. The Now

4. Results

5. Conclusions

DesiRNA - algorithm



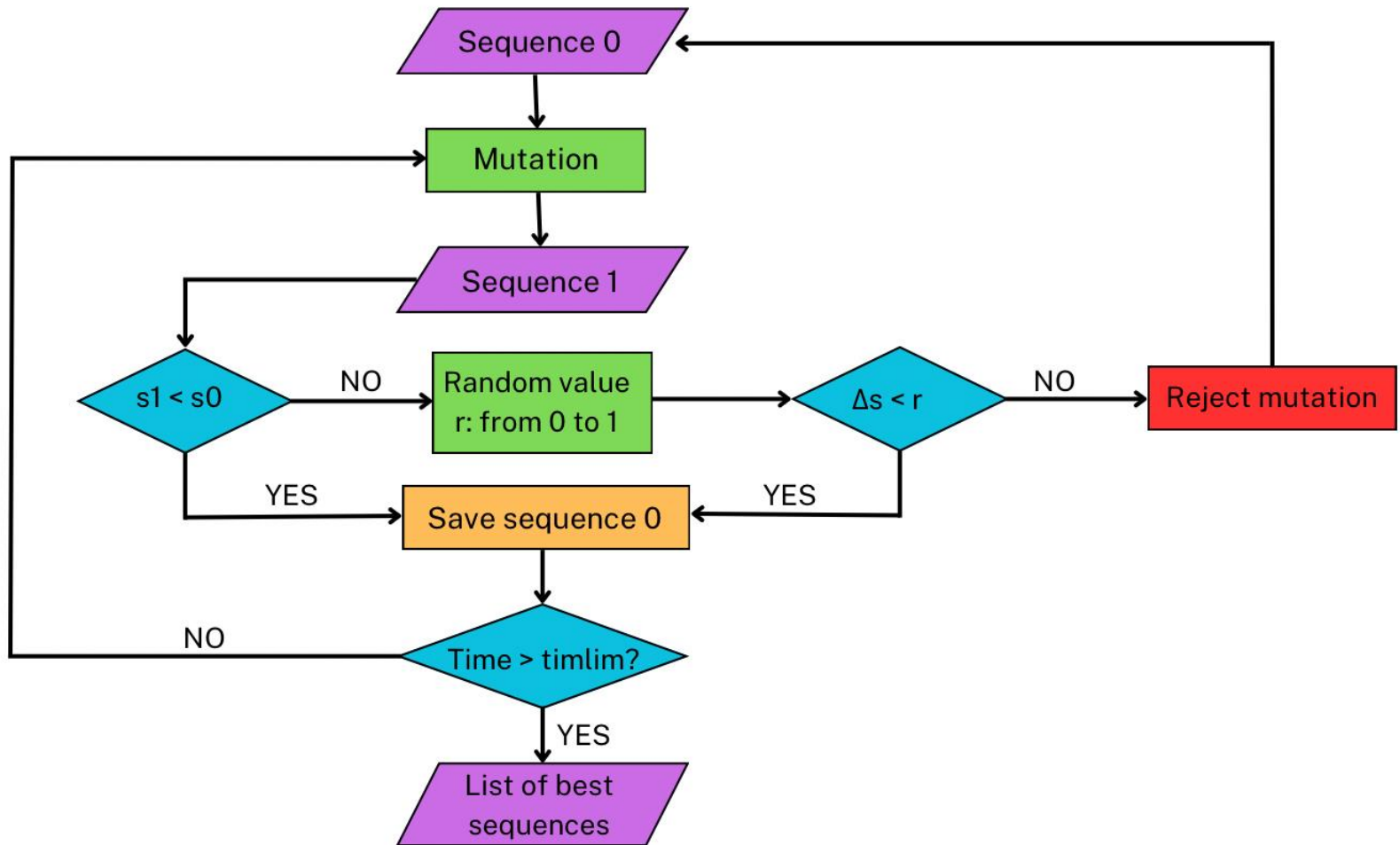
Outline

1. Introduction
2. The Past
- 3. The Now**
4. Results
5. Conclusions

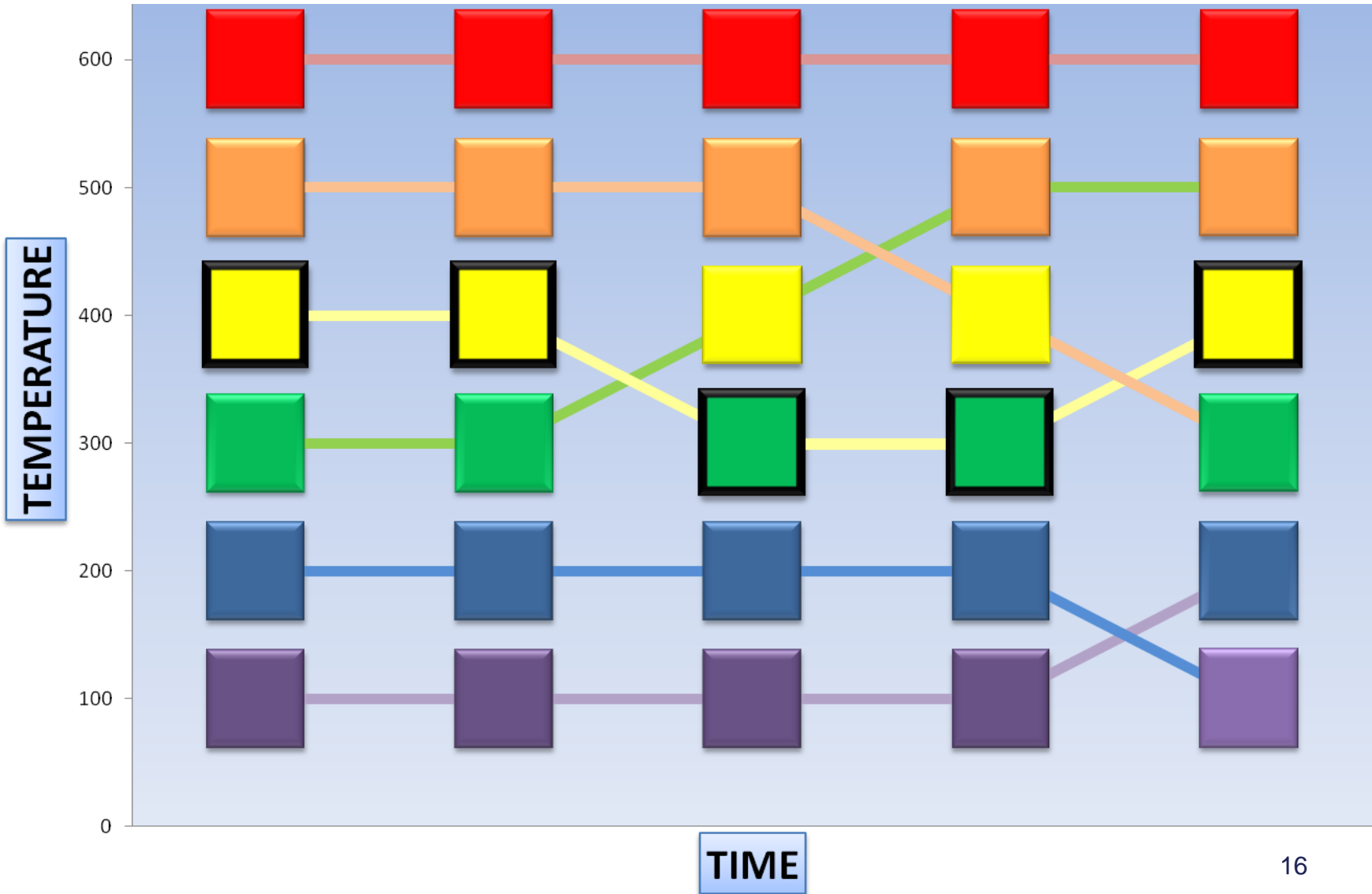


DEFIRNA

Monte Carlo



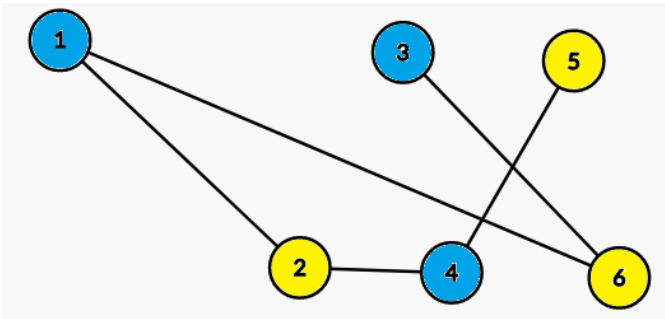
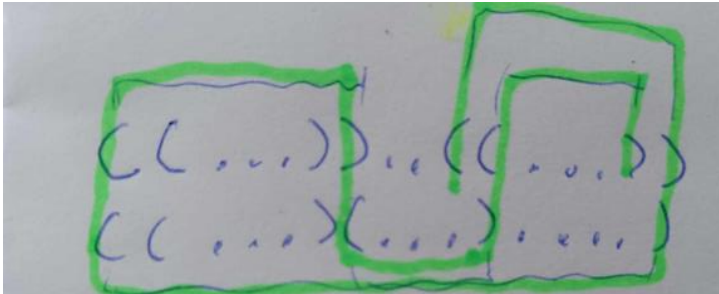
Replica Exchange



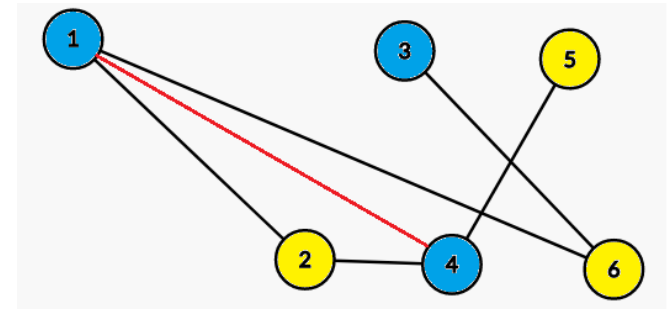
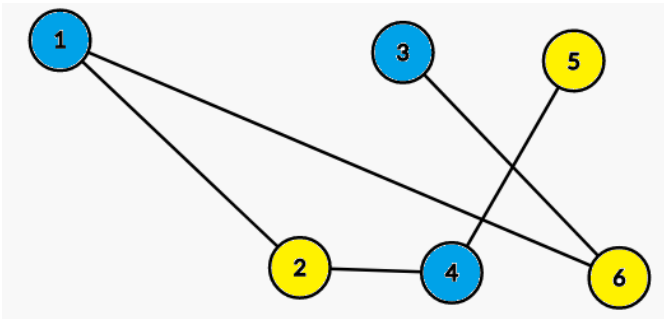
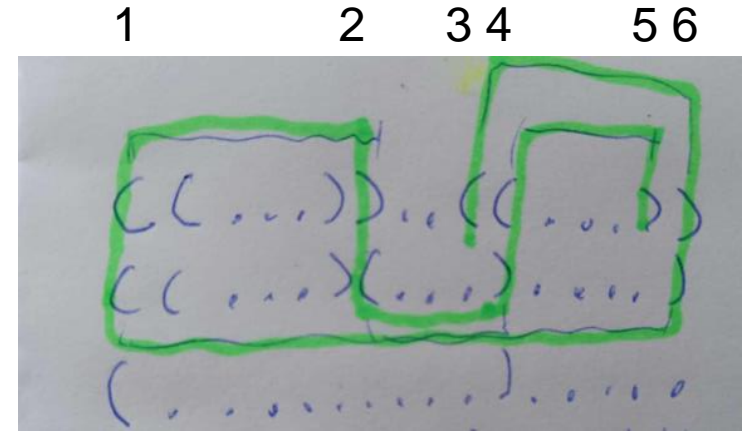
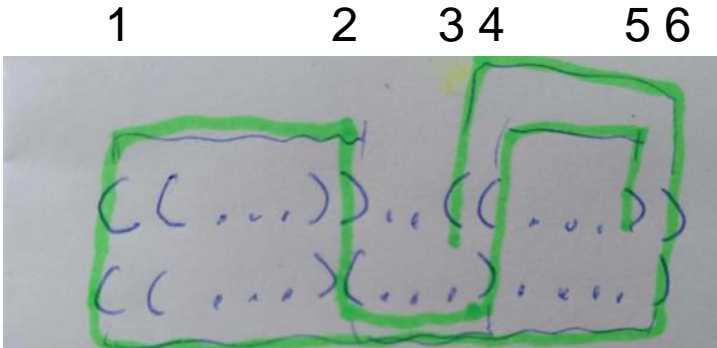
Alternative structures– graph coloring



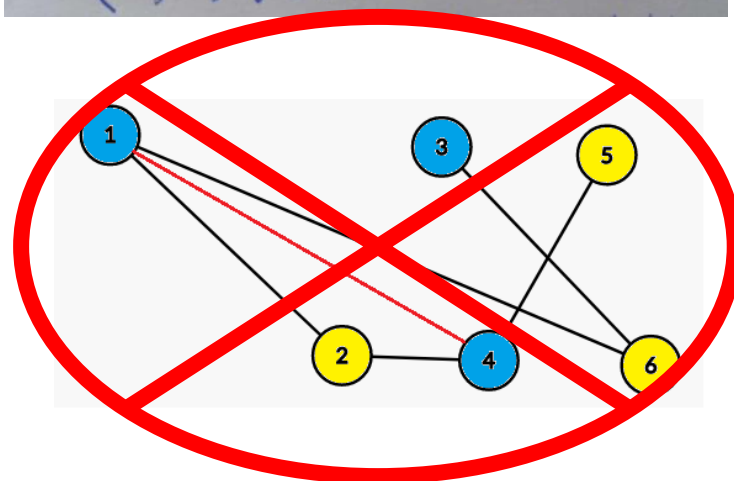
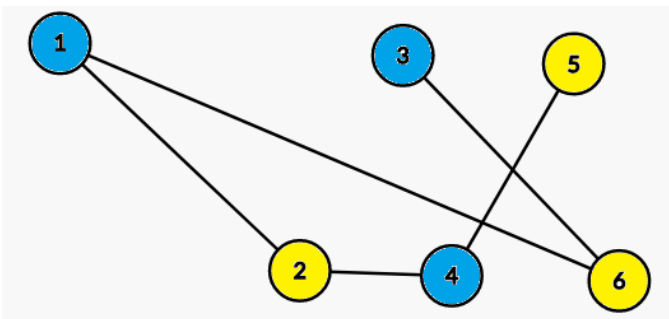
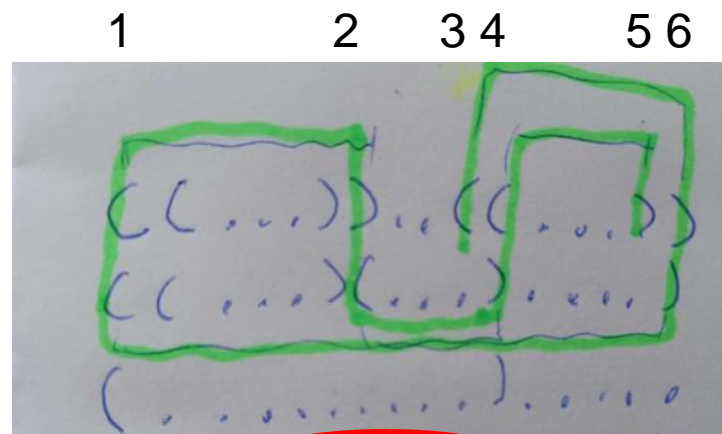
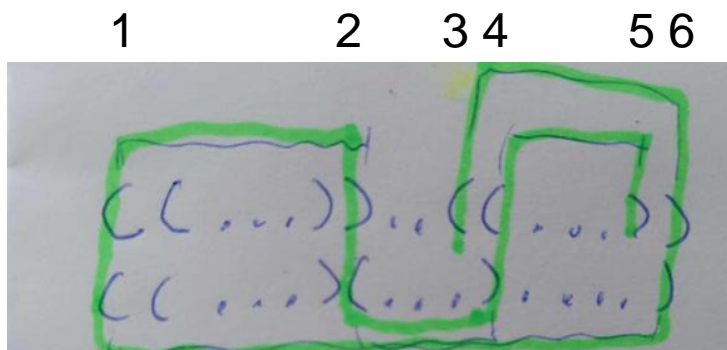
1 2 3 4 5 6



Alternative structures– graph coloring



Alternative structures– graph coloring



In alternative structures sequence design the graphs have to be „bipartite” (node of one color can be only connected to nodes of the second color)

DesiRNA – scoring functions



Ed-Epf: Energy of the desired structure (E_d) minus free energy of the thermodynamic ensemble (E_{pf})

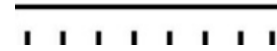
$$s = \exp\left[\frac{1}{kT} \times E_d - E_{pf}\right]$$

- Ed-MFE: Energy of the desired structure (E_d) minus Minimum Free Energy (MFE)
- 1-MCC: One minus Matthews Correlation Coefficient (MCC).
- sln_Epf : Sequence Length Normalized Epf.
- 1-precision: One minus precision ($TP/(TP+FP)$).
- 1-recall: One minus recall ($TP/(TP+FN)$).
- Edef: deviation of the RNA secondary structure ensemble from the target structure, normalized by sequence length

DesiRNA



- Design single chain RNA
- Design multimeric RNA
- Various types of constraints:



- Structure

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

- Sequence

GAUUANNNNAAGCAGUNNNNUCCANNNNN

- Alternative structures

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

- Pseudoknots

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

- Oligomerization



- Homo/ Heterodimers

- Prevented patterns

AAAA, GNRA, GGGC

- ACGU%

- **BONUS: It is extremely fast and lightweight!**

Outline

1. Introduction
2. The Past
3. The Now
- 4. Results**
5. Conclusions

EteRNA benchmark - results

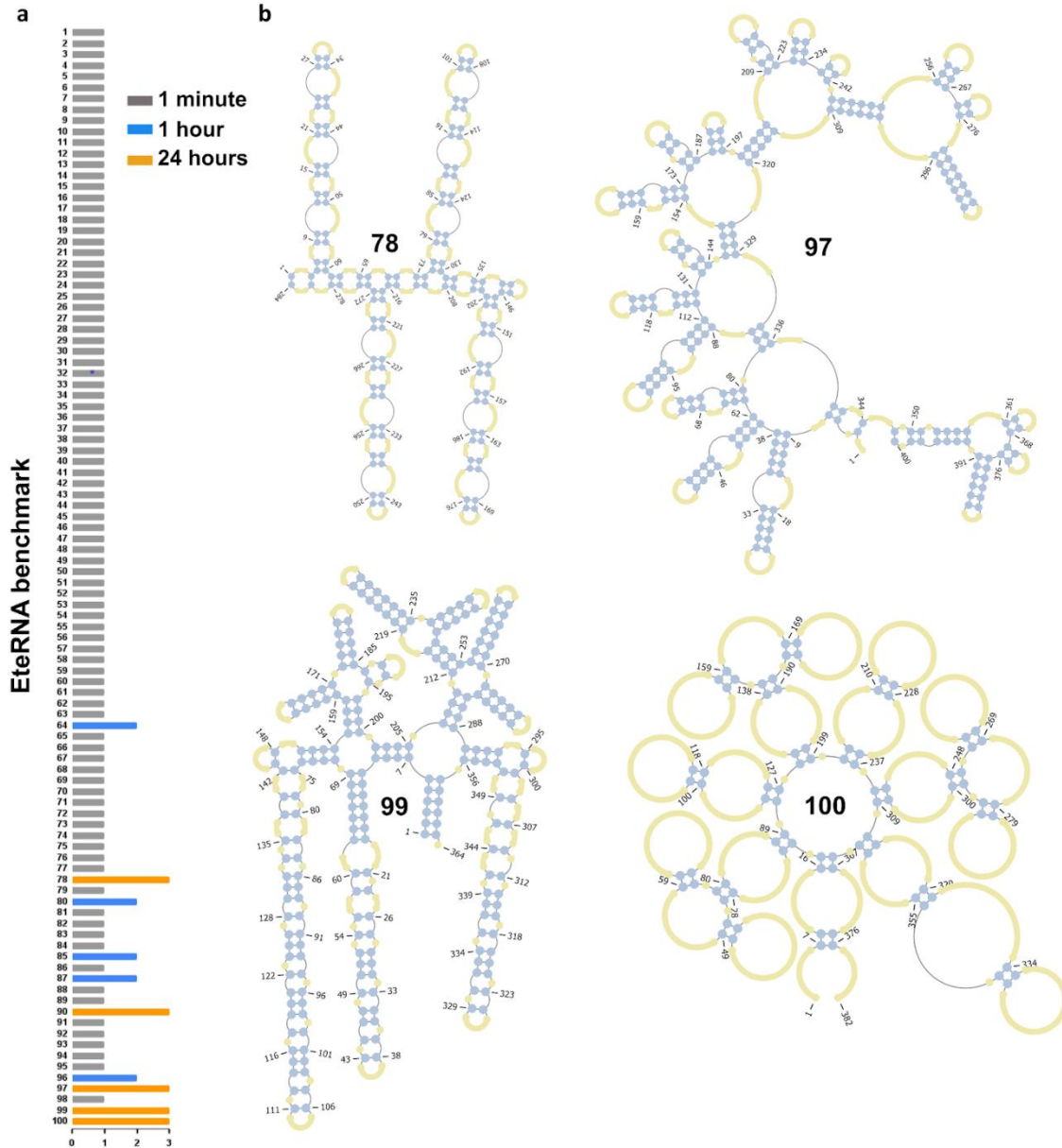


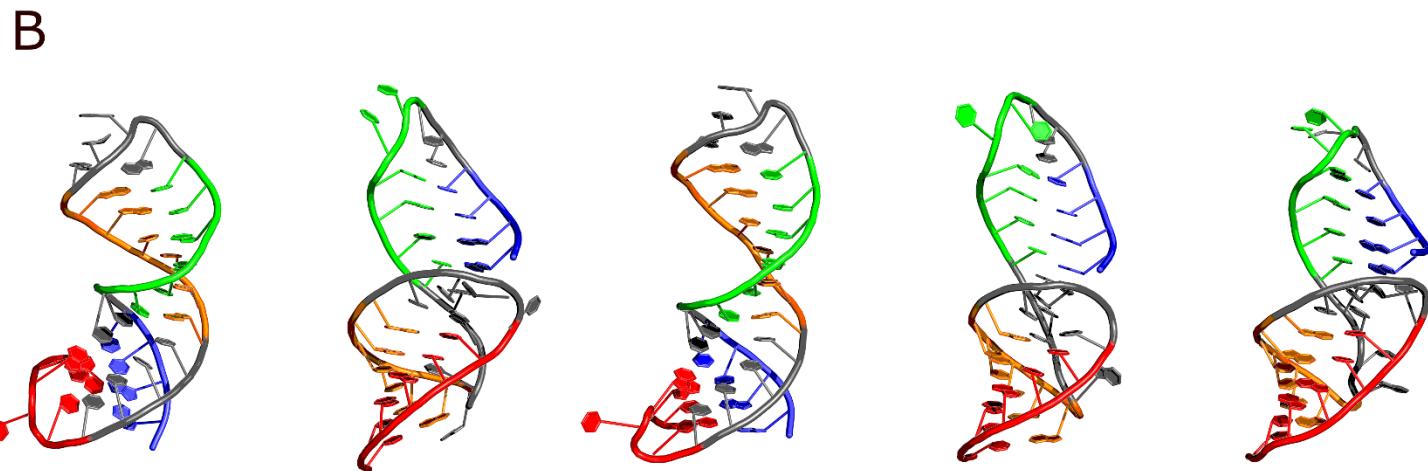
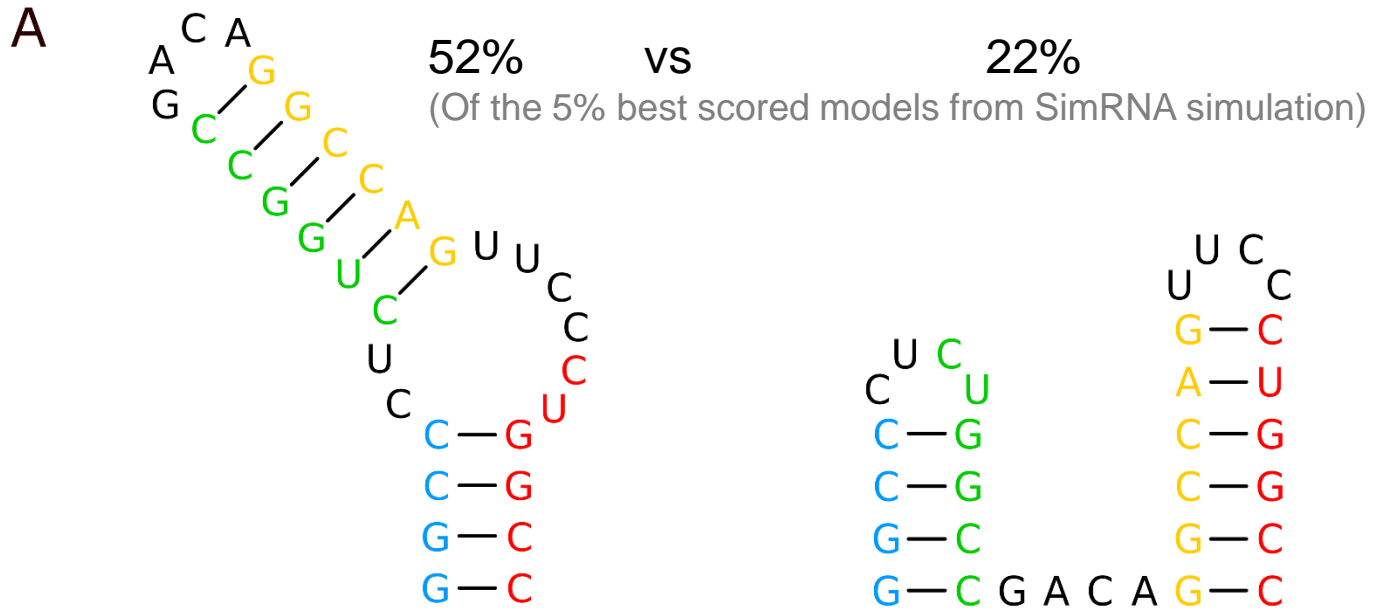
Image:
N. Badepally

The performance

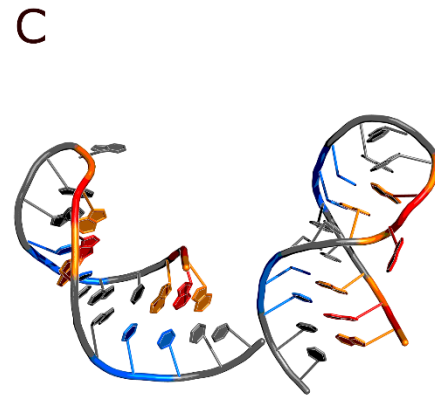
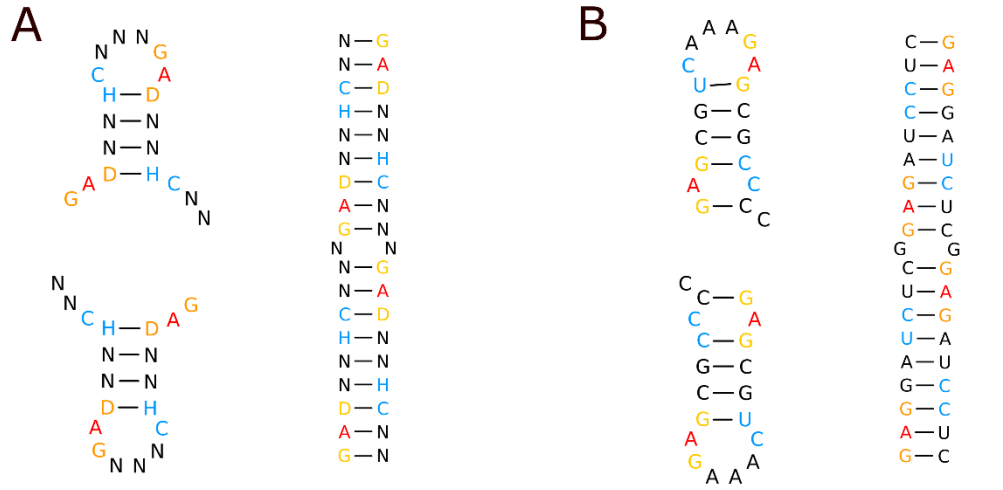


time	DesiRNA	
	E100 V1	E100 V2
60s	90	85
1h	95	95
24h	100	97

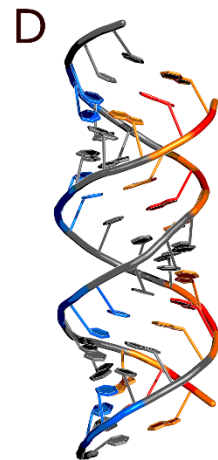
Alternative structures



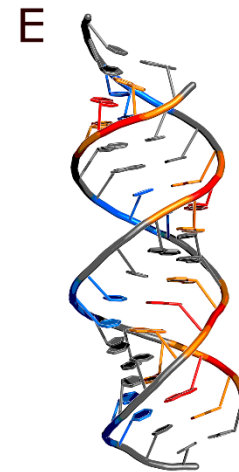
Homodimer vs homonomer



Mom19-I
98.675% of best
 scored models

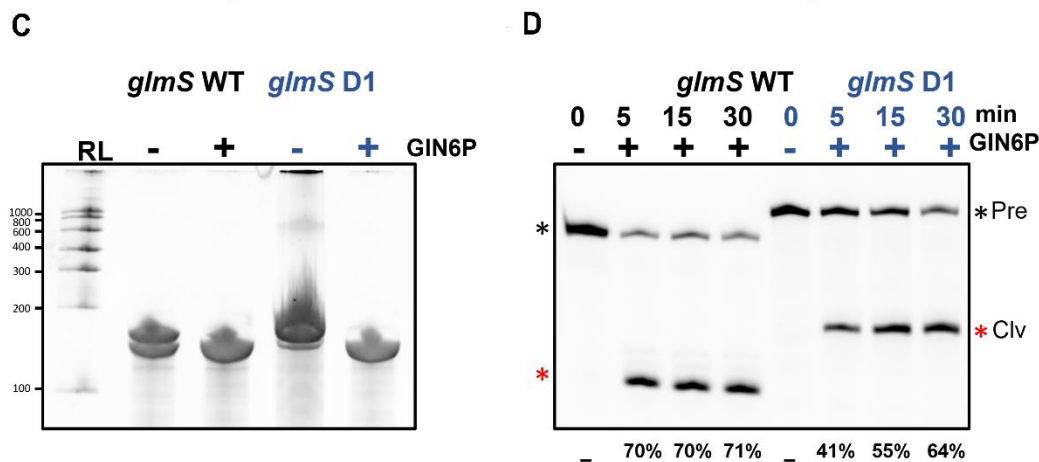
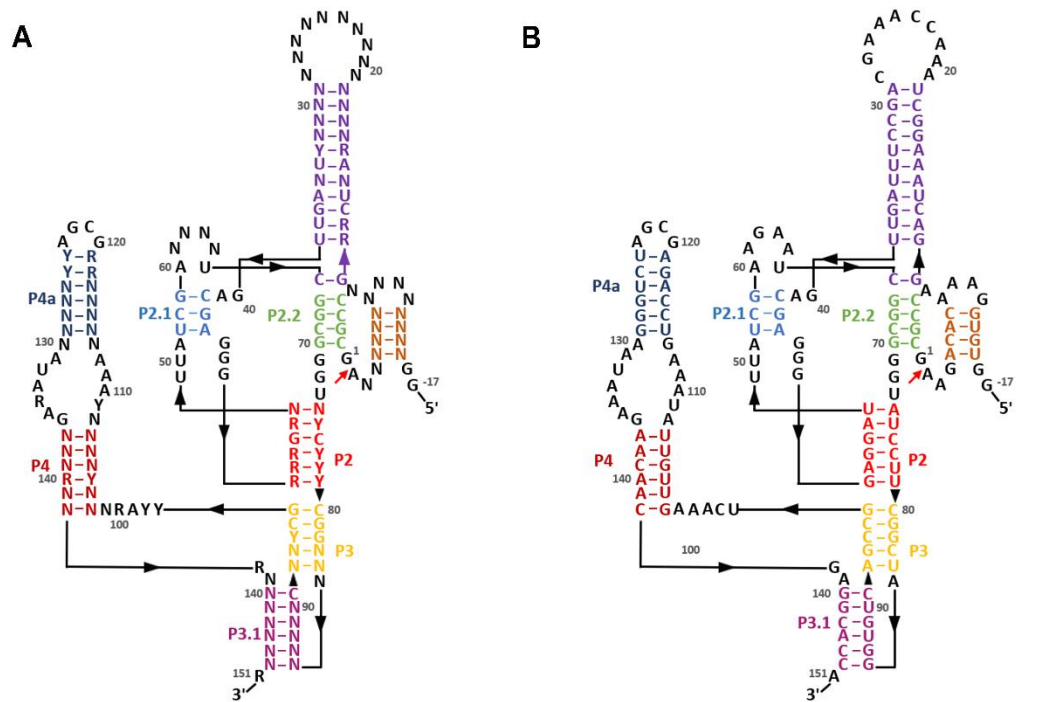


Mom19-II
100% of best
 scored models



RNA-II
 structure
 PDB ID:
 6IA2

Pseudoknotted structure - glmS



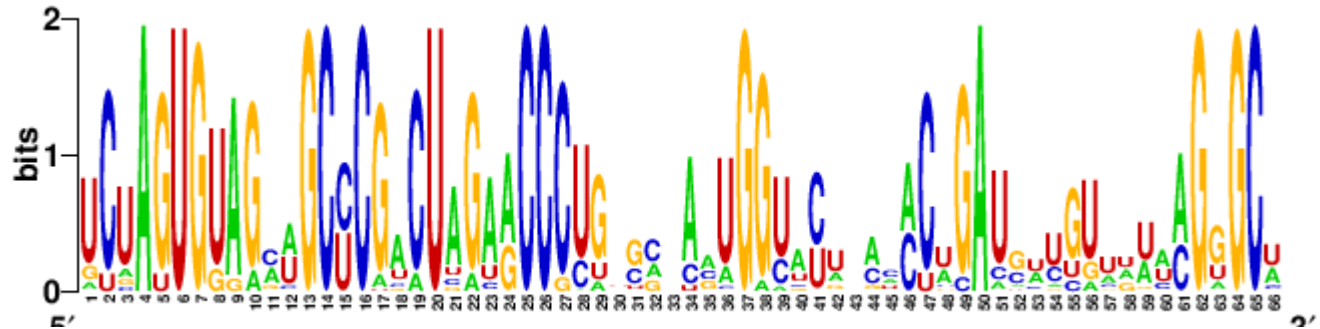
Pseudoknotted structure - FSE



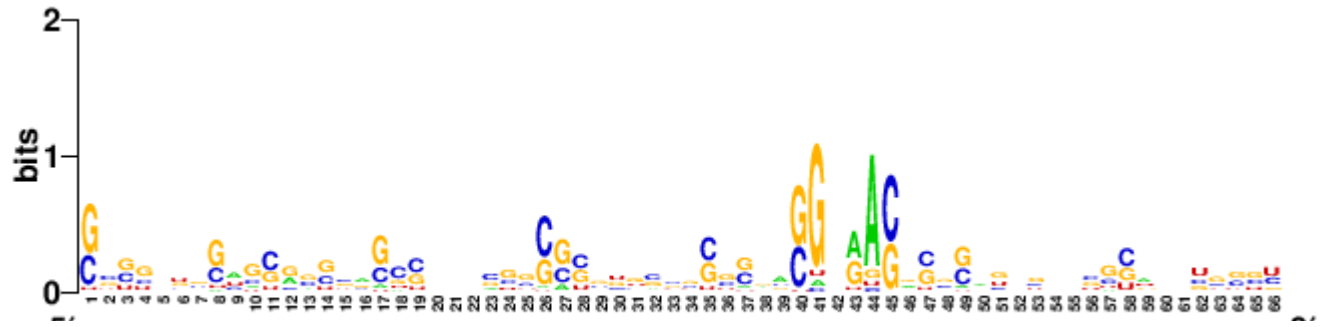
NNNAKUGKRRNNGCYCRNMUNRNNCCSYNNNNNNNNNGRNNNNNNNNNYNSANNNNNNNNNNMGNGCN

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

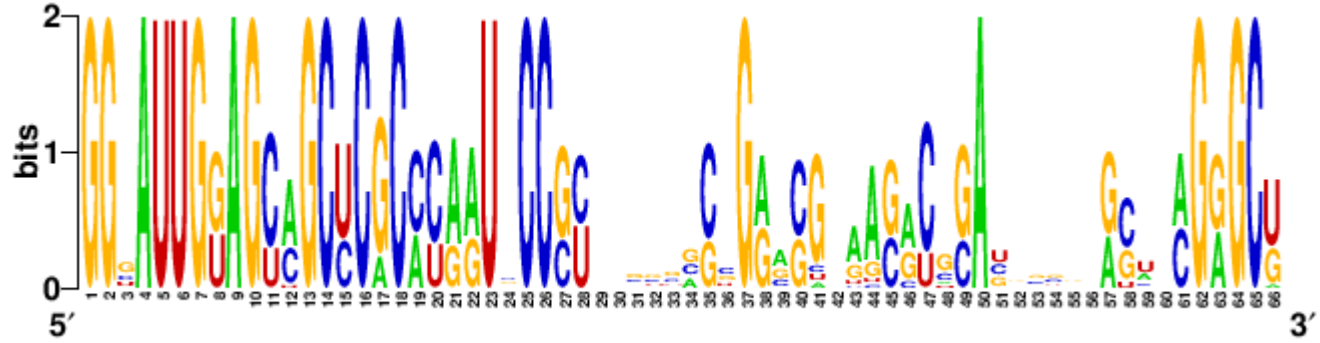
RF00507_seed



Design
no_seq_restr



Design
with_seq_restr

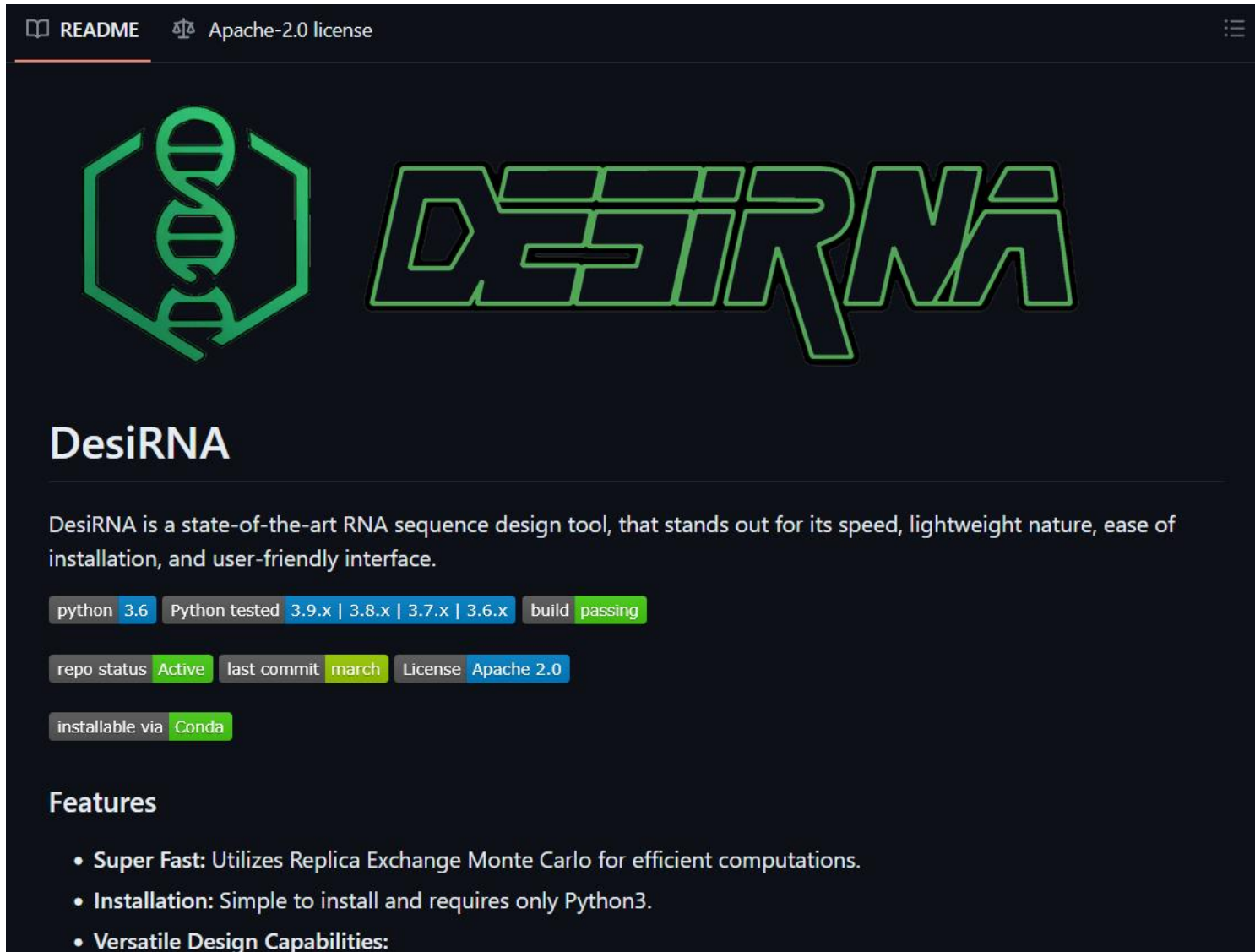


Outline


1. Introduction
2. The Past
3. The Now
4. Results
- 5. Conclusions**

GitHub

<https://github.com/fryzjergda/DesiRNA>



README Apache-2.0 license



DesiRNA

DesiRNA is a state-of-the-art RNA sequence design tool, that stands out for its speed, lightweight nature, ease of installation, and user-friendly interface.

python 3.6 Python tested 3.9.x | 3.8.x | 3.7.x | 3.6.x build passing

repo status Active last commit march License Apache 2.0

installable via Conda

Features

- **Super Fast:** Utilizes Replica Exchange Monte Carlo for efficient computations.
- **Installation:** Simple to install and requires only Python3.
- **Versatile Design Capabilities:**

Conclusions

- DesiRNA works like a charm
- The only method that solves the whole Eterna benchmark
- Extremely fast and lightweight
- User-friendly, easy to install
- Good documentation
- Alternative structures design, negative design, homodimer/monomer design – and more!

Thank you!



Acknowledgements



Janusz M. Bujnicki

Grzegorz

Farhang

Naeim

Nagendar

Eugene

Kalina

Filips



Funding:

National Science Centre (NCN) [UMO-2017/25/N22/01294 to J.M





Email: tomasz.wirecki@gmail.com

Linkedin: <https://www.linkedin.com/in/tomasz-wirecki-476b7637/>

GitHub: <https://github.com/fryzjergda>