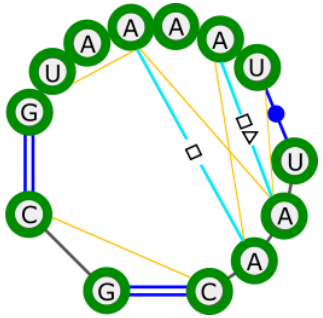
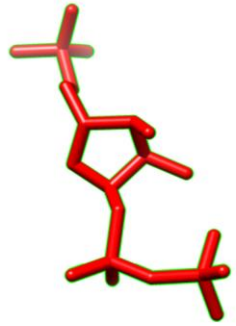
Relational Graph Convolutional Networks for RNA structure analysis and binding predictions

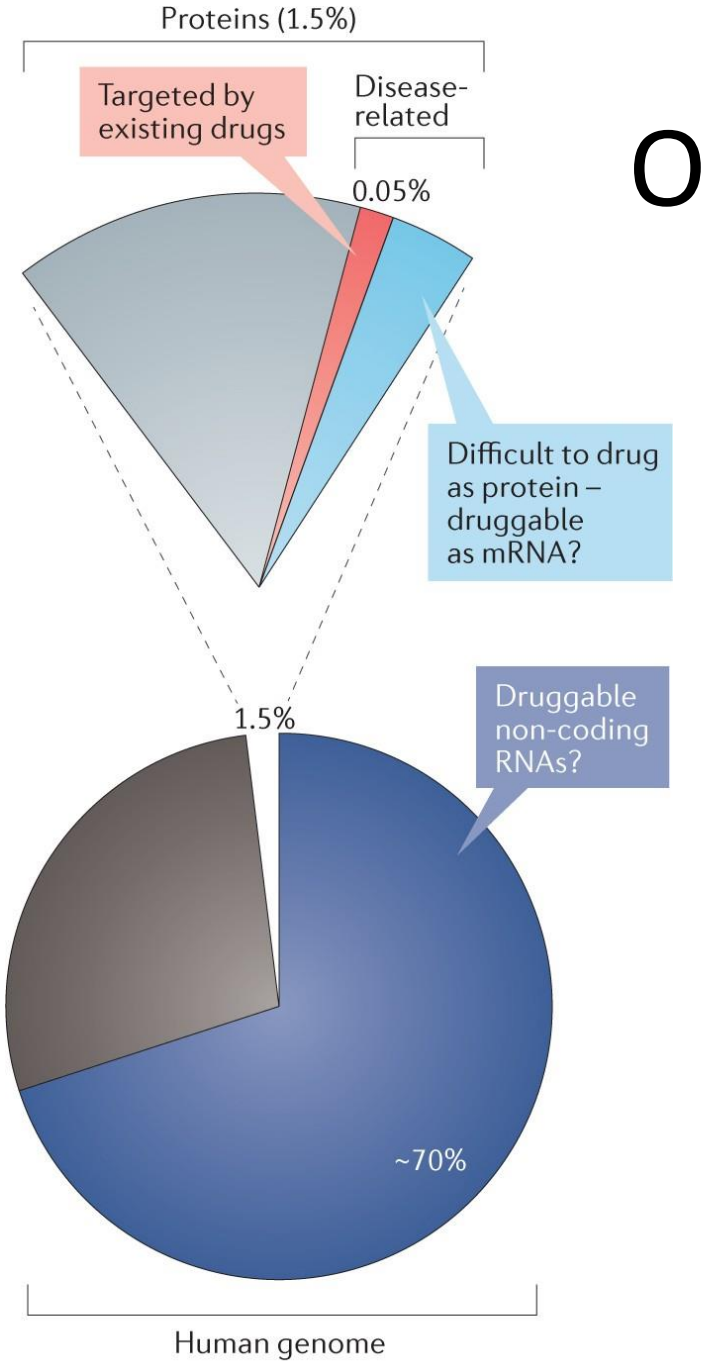


Jérôme Waldispühl
School of Computer Science
McGill University

✉ jerome.waldispuhl@mcgill.ca



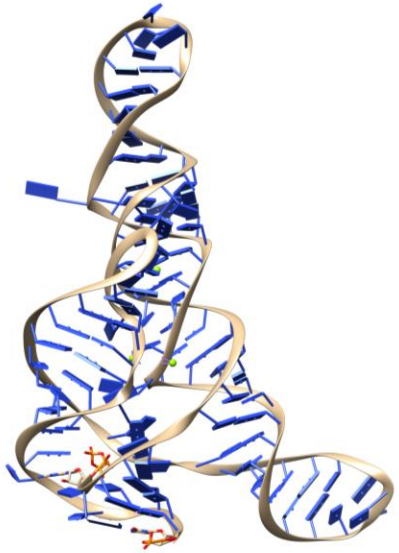
http://csb.cs.mcgill.ca



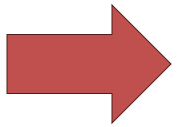
(Warner et al., 2018)

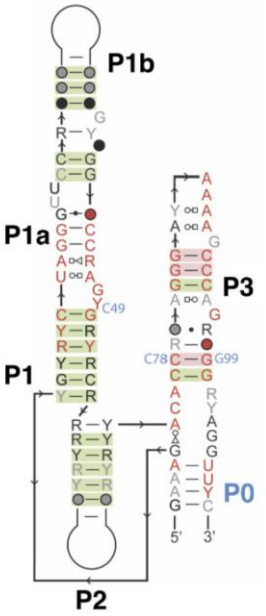
Objectives & Motivations

Challenge: How to screen the full data set?



PRPP riboswitch (Perselis & Serganov, 2018).





- 3D data is rare and 3D structures hard to predict
- Geometric DL is sensitive

- Lightweight
- Fast & robust algorithms
- Easier to acquire
- Retain structural signal

Relational Graph Convolutional Network (RGCN)

$$h_i^{(l+1)} = \phi\left(W_0^{(l)} h_i^{(l)} + \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_r(i)} \frac{1}{|\mathcal{N}_r(i)|} W_r^{(l)} h_j^{(l)}\right)$$

$h_i^{(l+1)}$: representation of node i at layer l
$\mathcal{R}$: relations (edge types)
$\mathcal{N}_r(i)$: neighbors of node i

(Schlichtkrull et al., 2018)

Semi-supervised learning and docking data augmentation for small molecule binding affinity prediction

RNAmigos

(Oliver *et al*., 2020)

(Carjaval-Patiño et al., 2025)

Carlos G. Oliver
Vanderbilt University

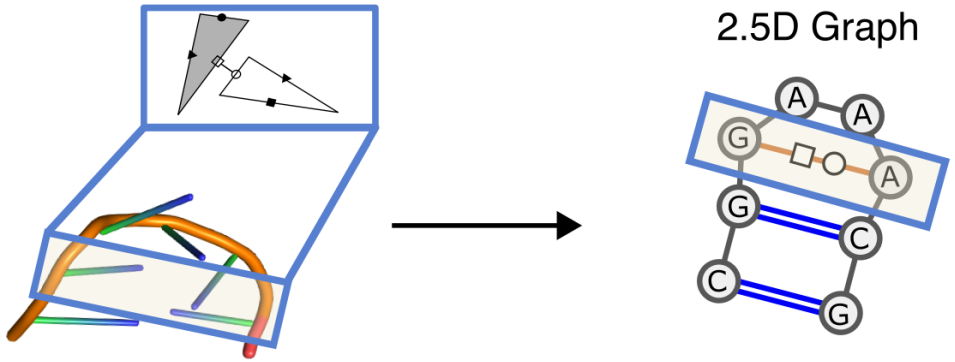
Vincent Mallet
Les Mines

Juan Carjaval
Univ. Nac. de Colombia

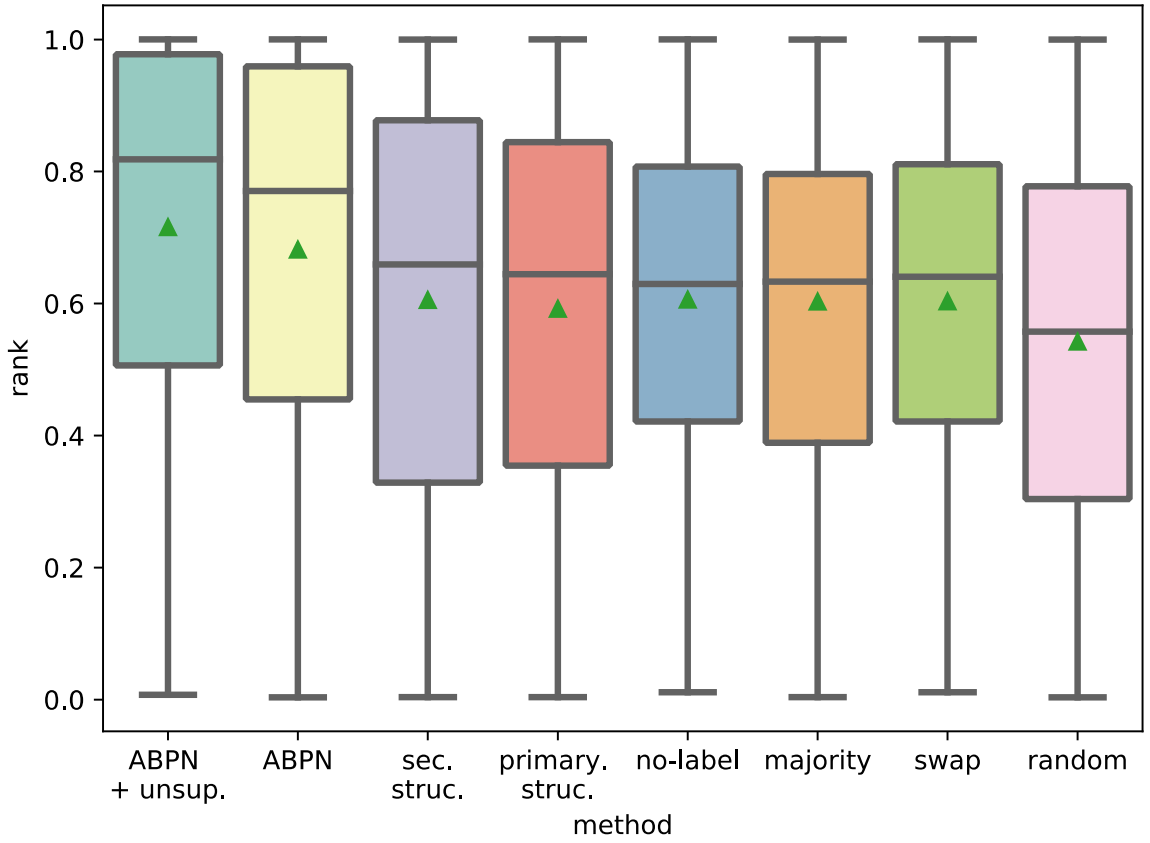
David Becerra
McGill University

RNAmigos1: Proof of concept

Encode binding pockets as graphs labelled with Leontis-Westhof base pair classification.



Given a motif **RNAmigos1** outputs the **fingerprint** of a ligand.



(Oliver *et al.*, 2020)

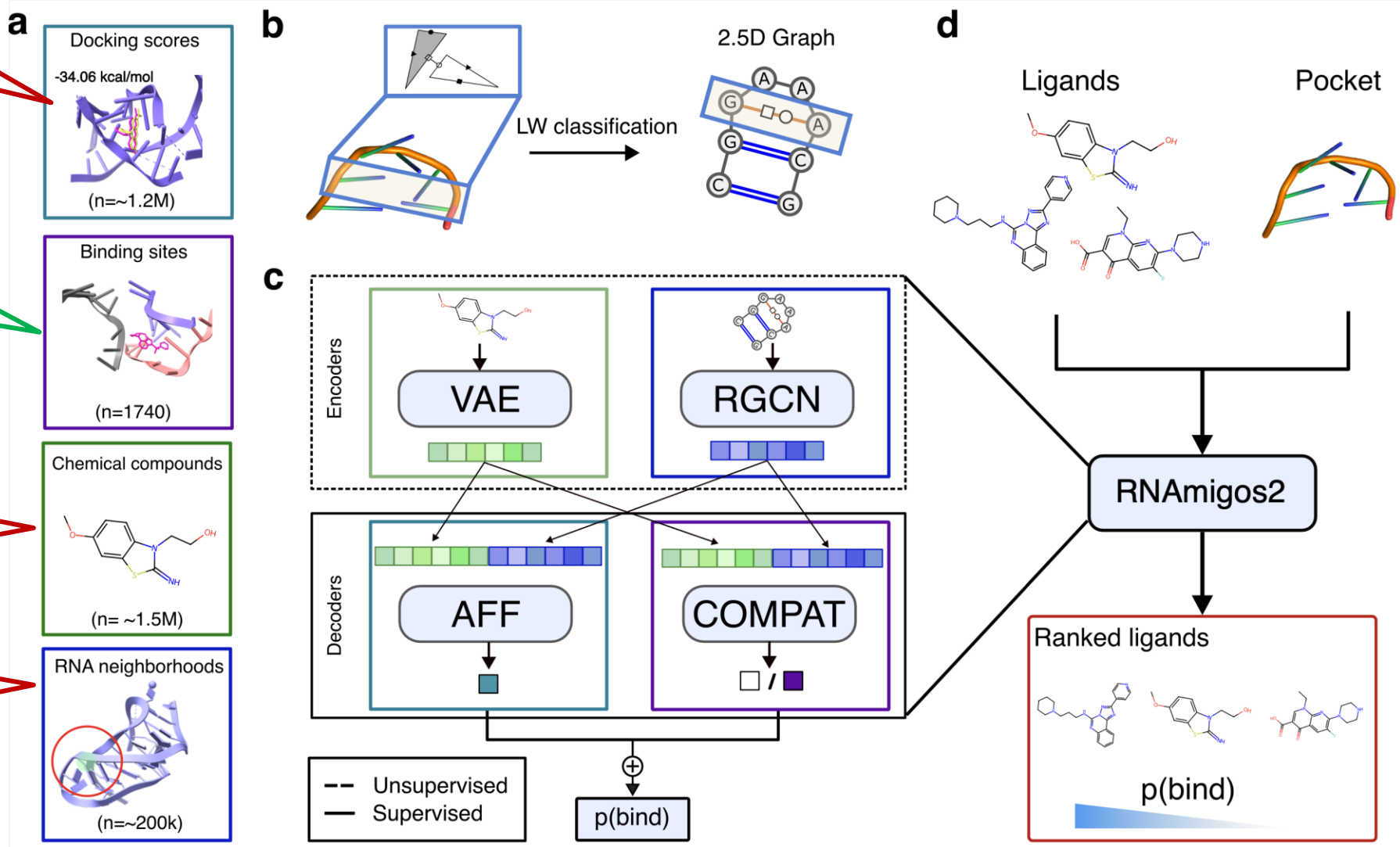
RNAmigos2 overview

large-scale docking experiment

Original PDB dataset

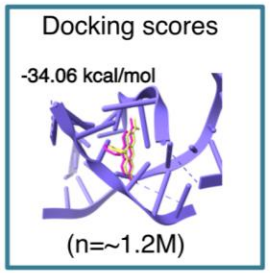
Unlabeled drug-like compounds

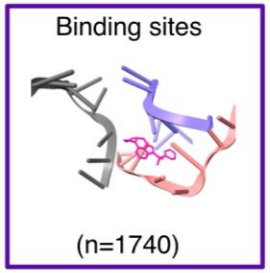
Unlabeled RNA target sites

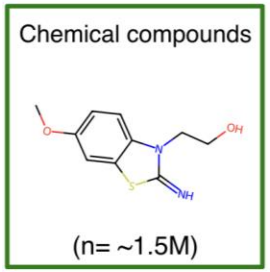


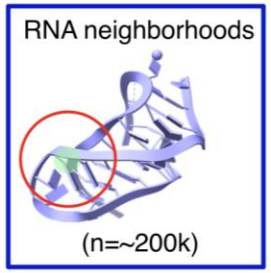
(Carjaval-Patiño *et al.*, 2025)

Data source





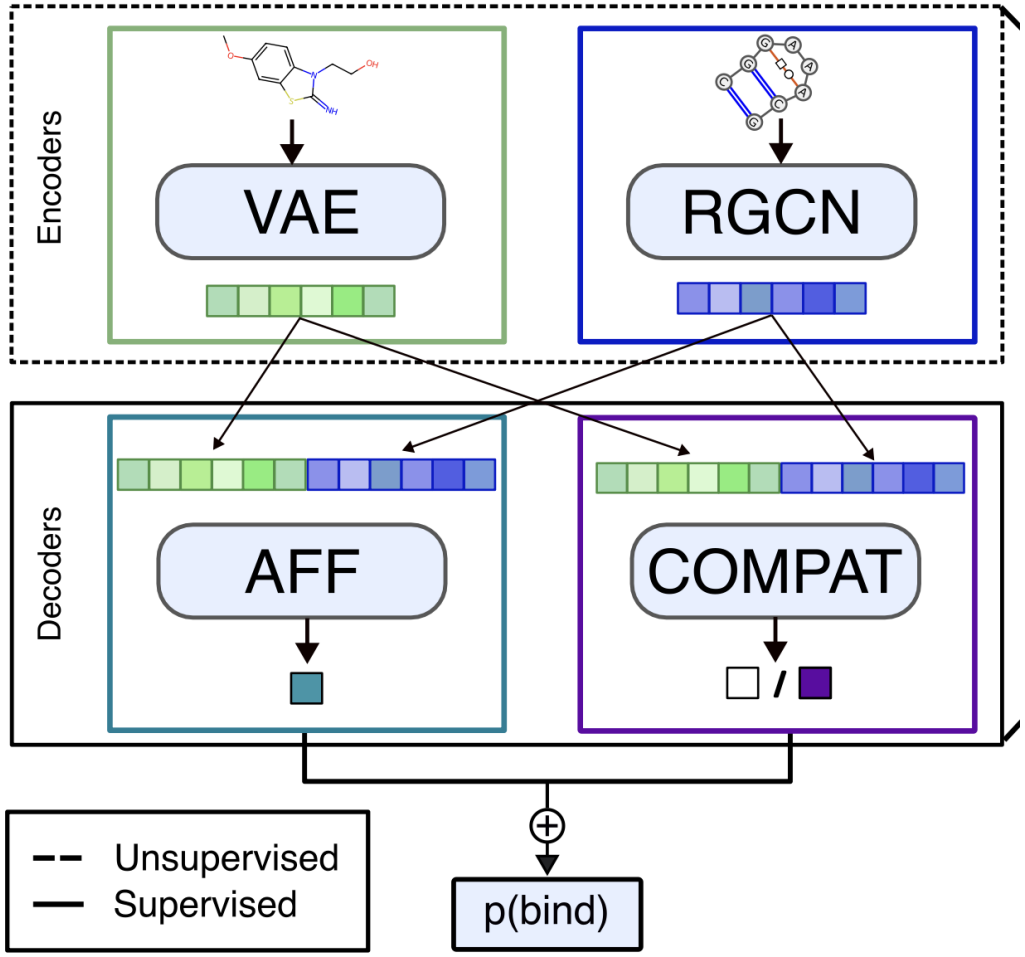




- **Docking scores** : ~1.2M RNA-pocket / ligand pairs scored by rDock.
- **Binding sites** : 1,740 native RNA–small-molecule complexes from the PDB.
- **Chemical compounds** : ~1.5M unlabeled drug-like molecules for VAE pre-training.
- **RNA neighborhoods** : ~200k binding-site-like subgraphs from the PDB for RGCN pre-training.

(Carjaval-Patiño *et al.*, 2025)

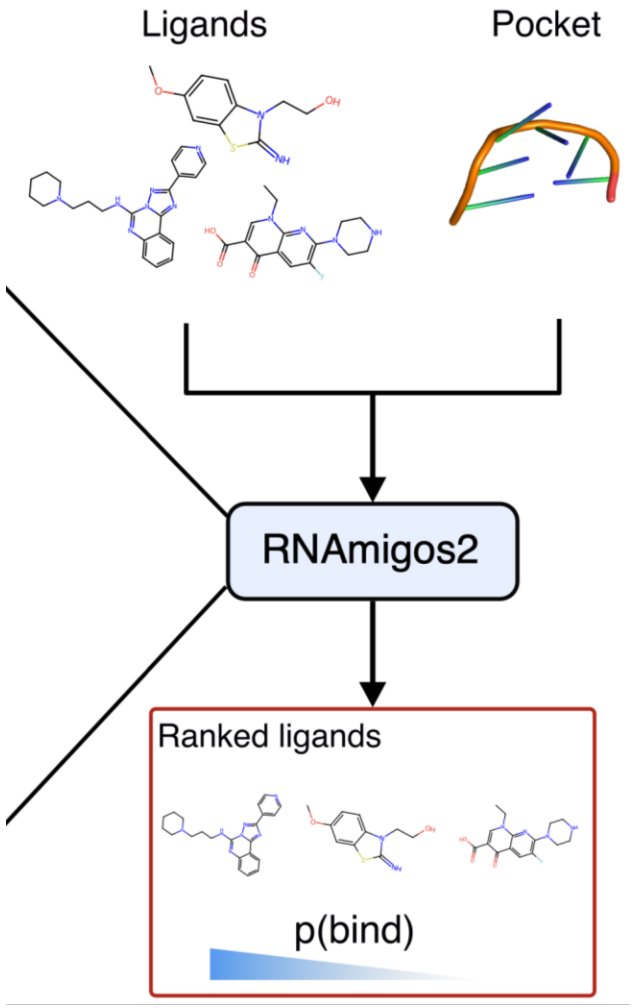
Encoder & Decoder



- **Encoders** (unsupervised): a VAE on small molecules and an RGCN on pocket 2.5D graphs.
- **Decoders** (supervised, joint): AFF predicts the rDock docking score; COMPAT predicts native vs. decoy.
- All four heads share the same encoders.

(Carjaval-Patiño *et al.*, 2025)

Scoring



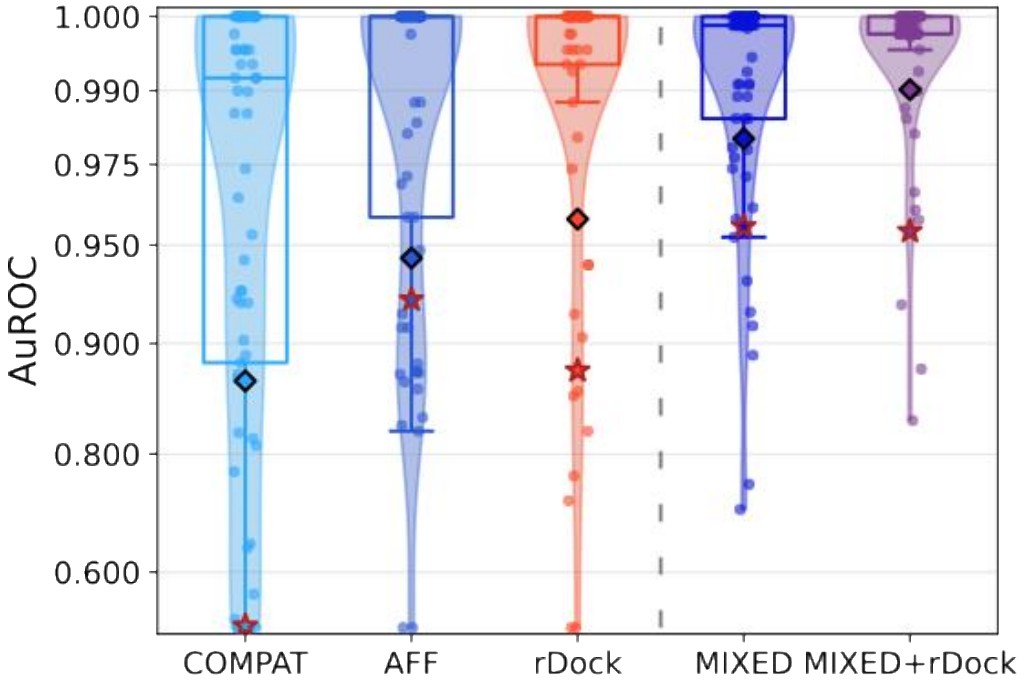
Input: a target pocket + a library of compounds:

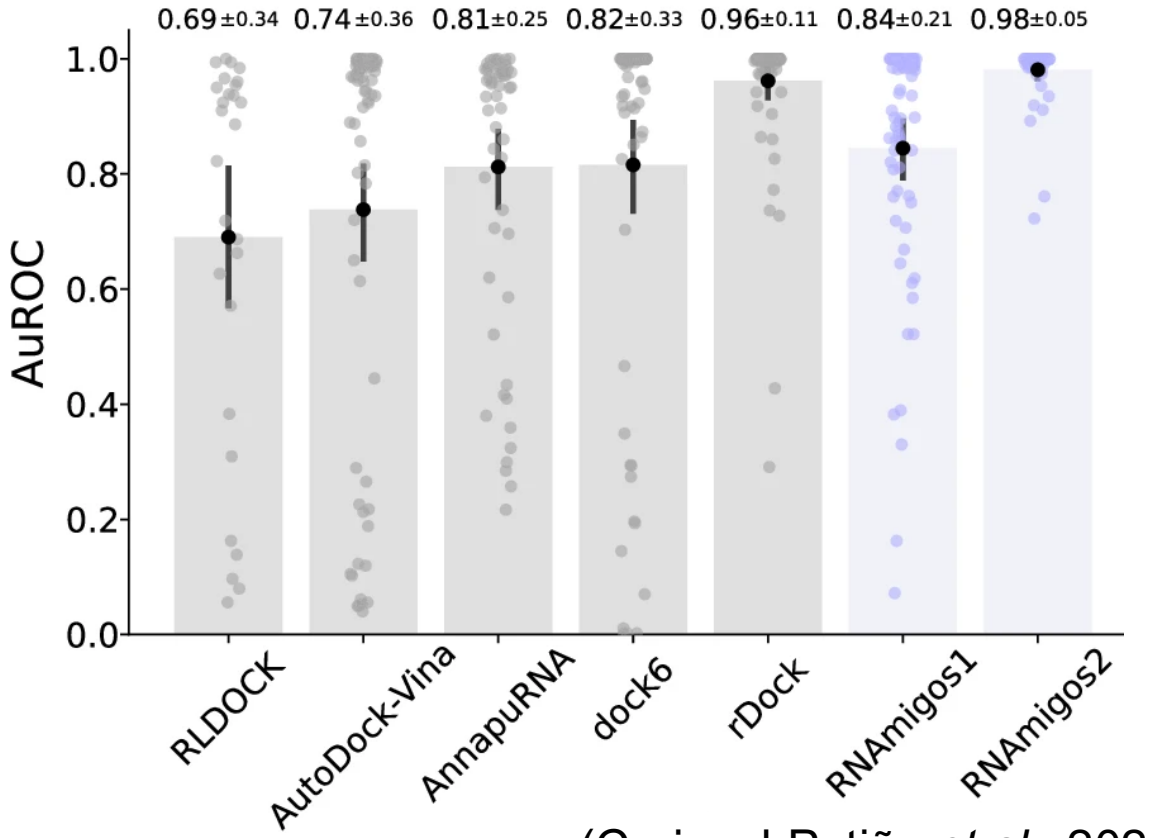
1. Encode pocket once with the trained RGCN.
2. Encode every compound with the VAE.
3. For each (pocket, compound), get scores from both the AFF and COMPAT heads.
4. Ensemble →final p(bind).

(Carjaval-Patiño *et al.*, 2025)

Virtual Screening (VS) Performance

- Candidate ligands: PDB (native) + filtered ChEMBL (decoy) compounds.
- For each PDB binding sites, score all compounds and sort them.

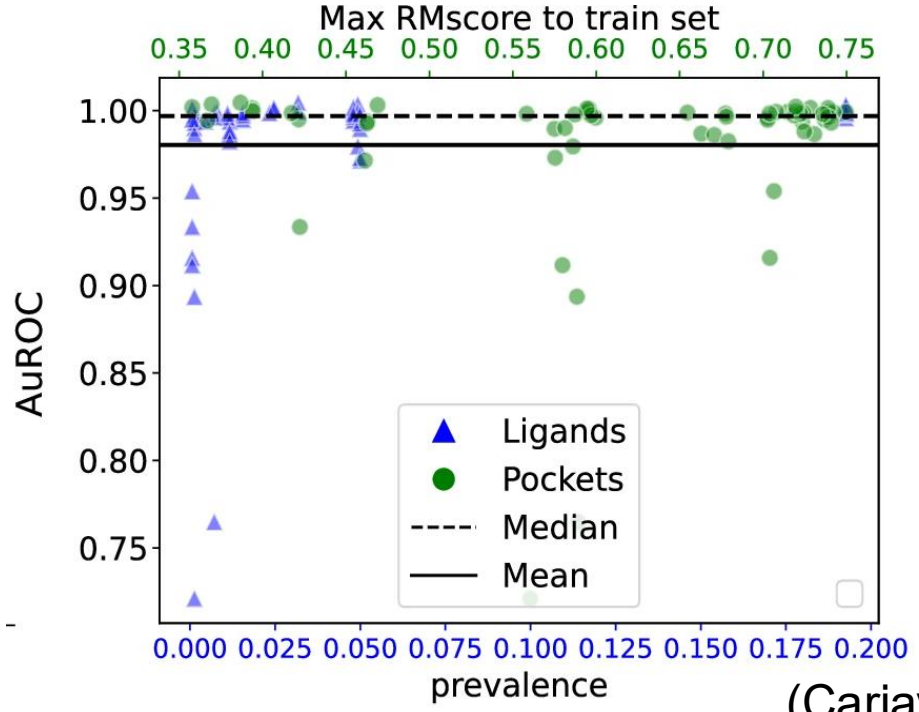


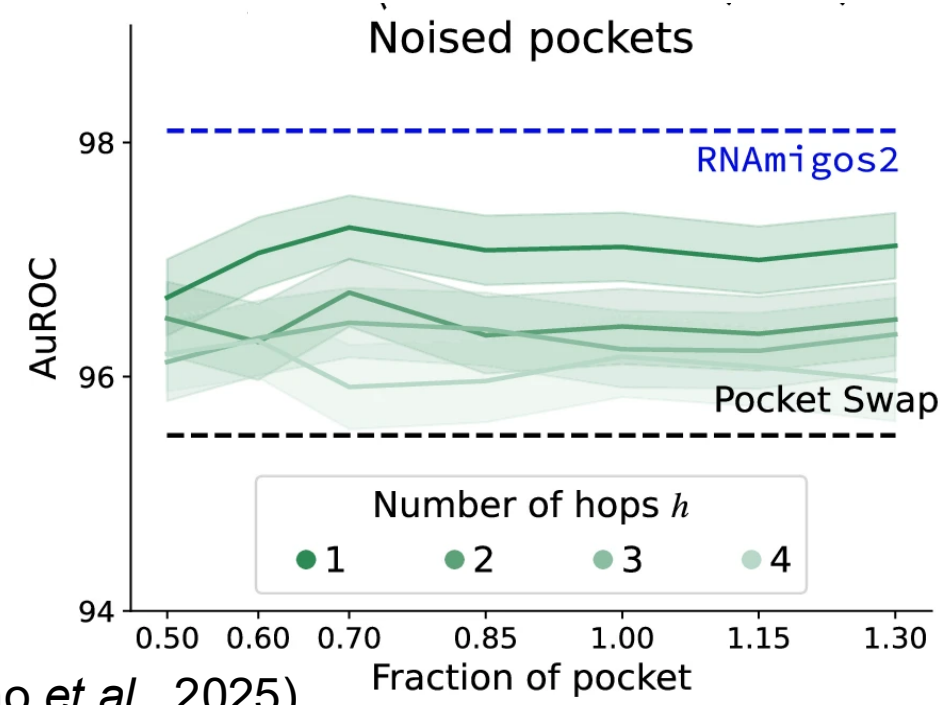


(Carjaval-Patiño *et al.*, 2025)

(some) Challenges of validation

- Leakage from pre-training
- Repeat experiments with multiple random seeds
- Class imbalance (e.g., PDB ligand dont display drug-like features)
- Generalization to unseen (distant) instances
- Validate the model leverage ligand **AND** structure features



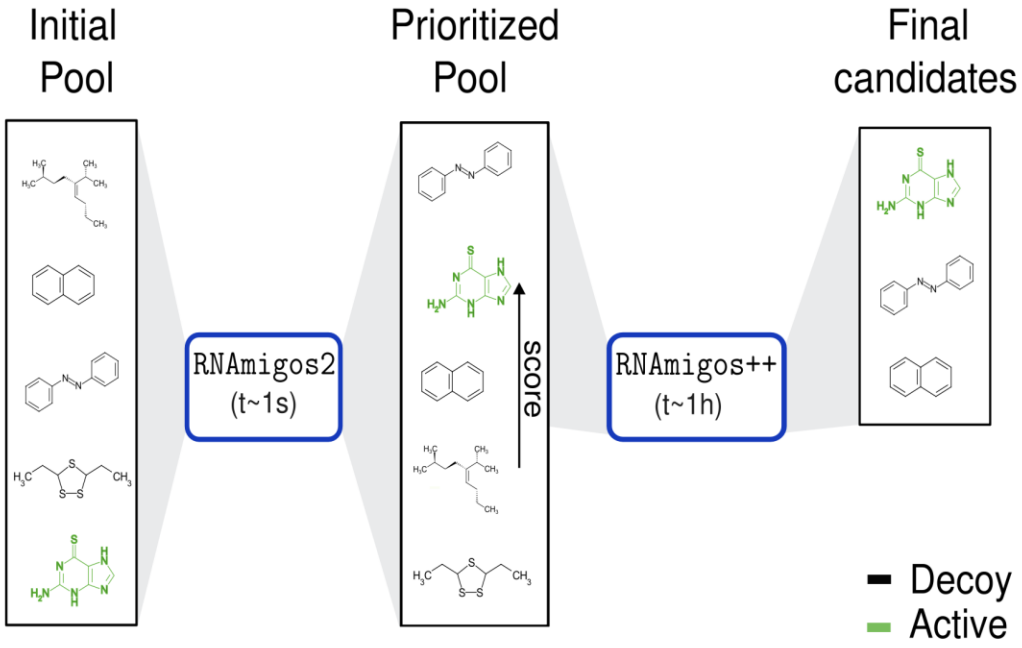


(Carjaval-Patiño *et al.*, 2025)

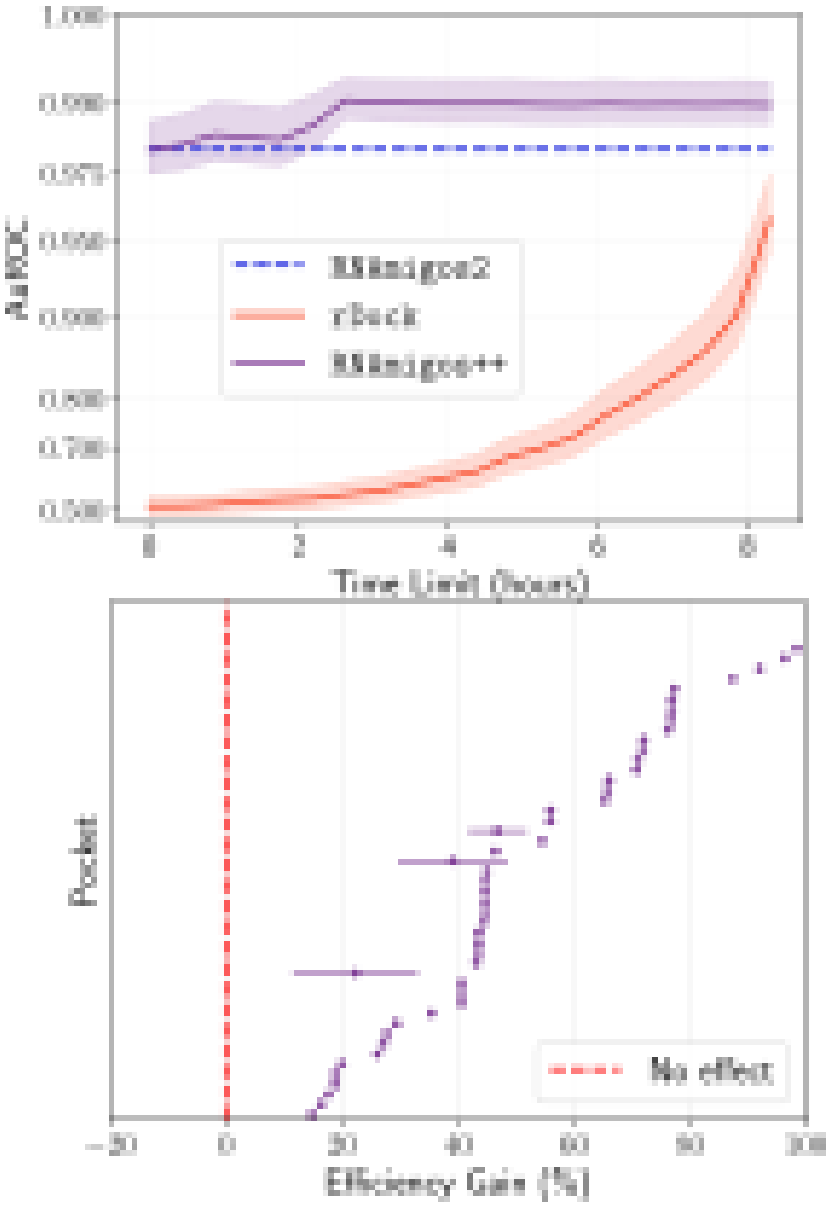
A hybrid docking strategy

Objective: Even if you do not fully trust RNAmigos, you can use it to accelerate classical docking

1. RNAmigos2 ranks candidates
2. Docking tool (rDock) scores the best candidates

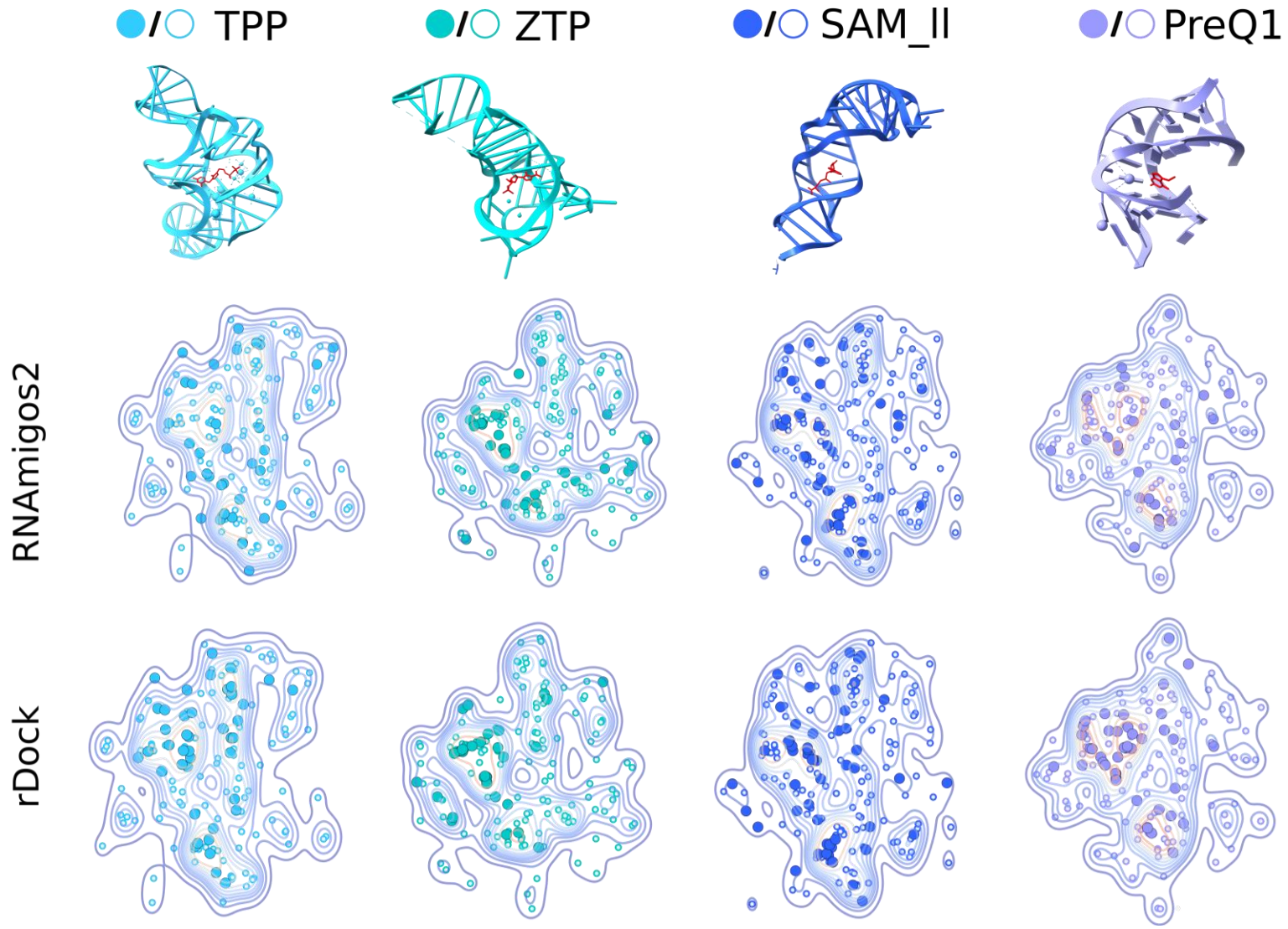


Performance:



(Blind) Test on ROBIN large-scale in-vitro assay

- 24,572 small molecules
- 36 Nucleic acid targets
- Mixed model (combined with rDock)
- Similar accuracy at fraction of running time



Try it yourself!

README MIT license

Using the tool with Collab

The easiest way to use the tool is to use Google Colab.

Open in Colab

You will need to provide a cif file, a binding site in the form of a list of binding pocket nodes and a list of ligand smiles.

Using the tool locally

A local use of the tool is also possible by following the next steps. NOTE: This has been tested on Linux Ubuntu 24 and Mac OS 13 and 14. No special hardware requirement, inference code runs on common desktops and laptops.

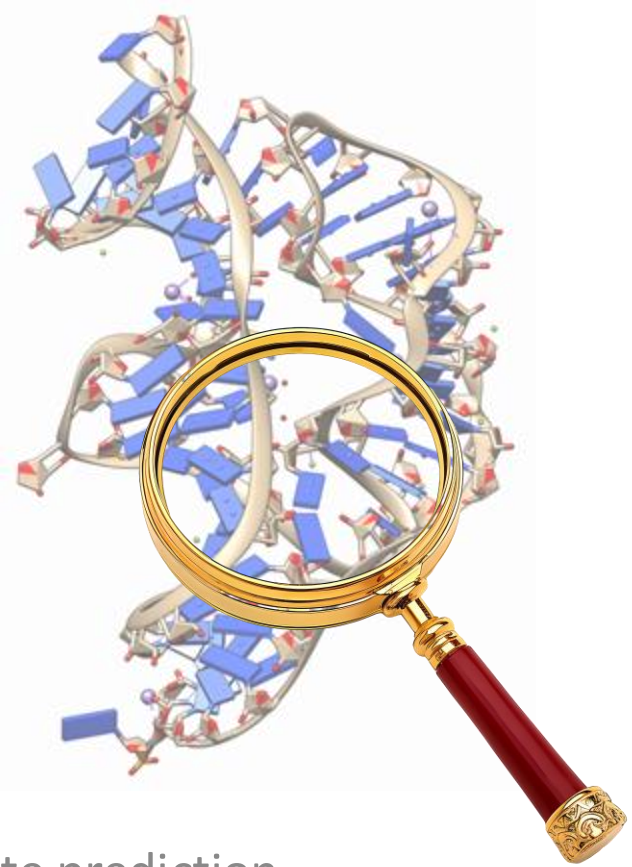
First, create a conda environment:

```
git clone https://github.com/cgoliver/rnamigos2.git
cd rnamigos2/
conda create -n rnamigos2
conda activate rnamigos2
pip install numpy==1.26
pip install torch==2.2.2+cpu torchaudio==2.2.2 torchdata==0.7.1 torchvision==0.17.2 --index-ur
pip install  dgl -f https://data.dgl.ai/wheels/torch-2.2/repo.html
pip install -r requirements.txt
```

To run RNAmigos2.0 on your own target and ligands, use the `rnamigos/inference.py` script.

https://github.com/cgoliver/rnamigos2

carlos.oliver@vanderbilt.edu



Binding site prediction

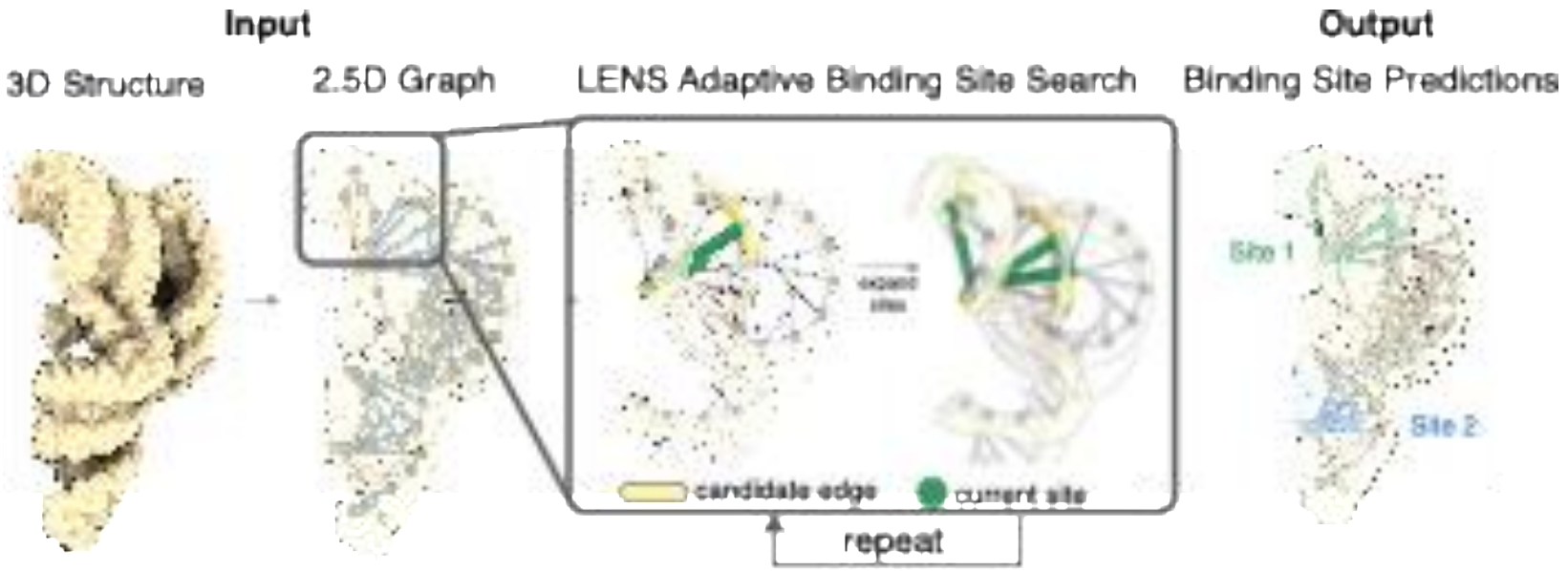
Ribo-LENS

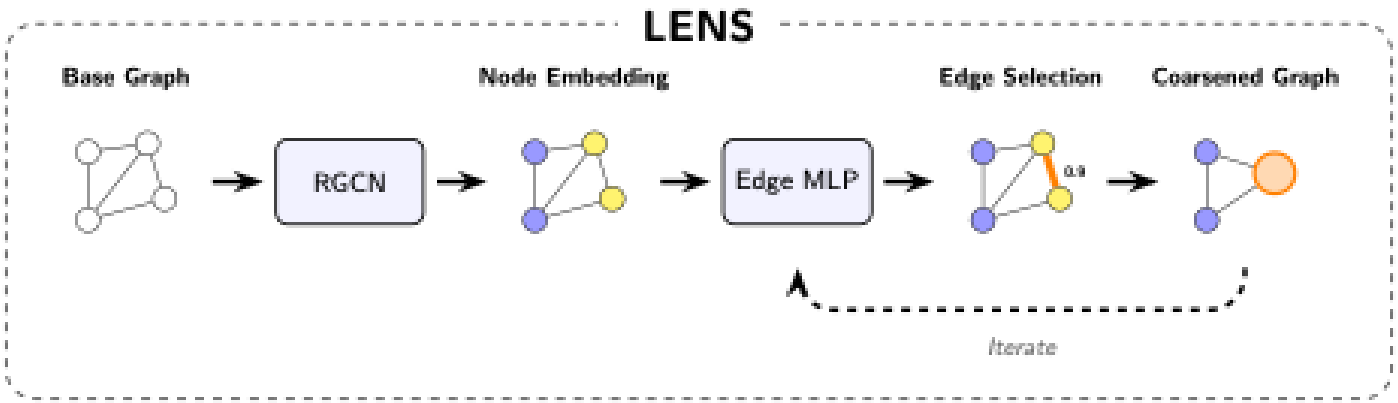
(Nitchi *et al*., submitted)

David Nitchi
McGill University

Carlos G. Oliver
Vanderbilt University

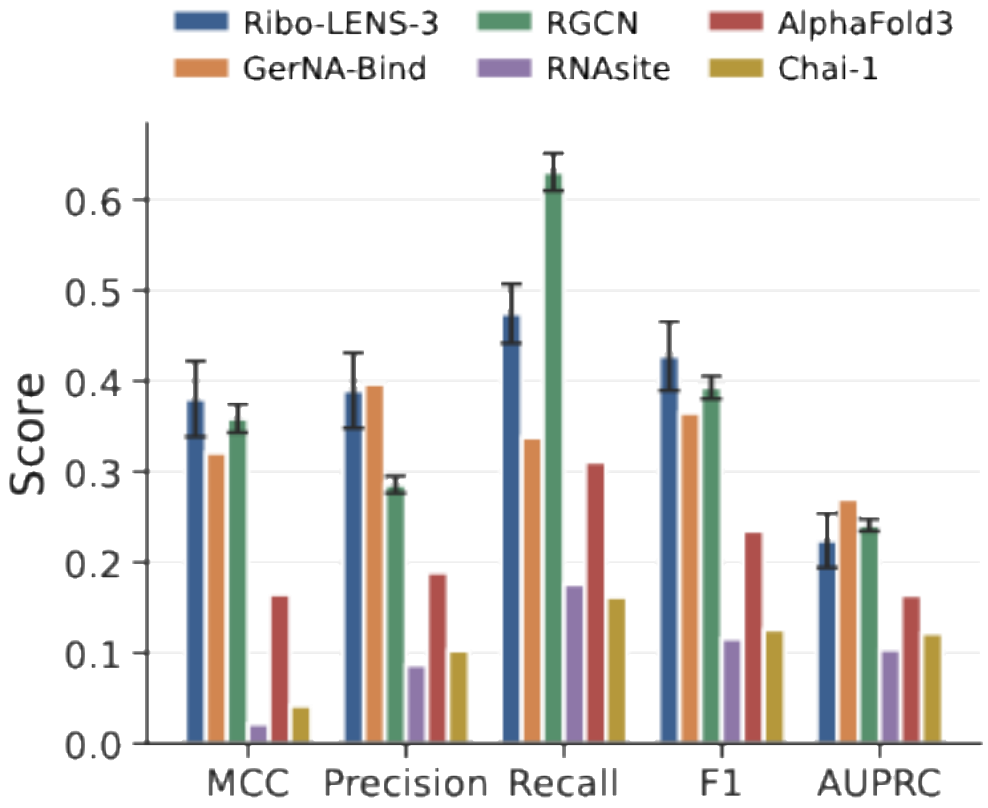
Ribo-LENS framework

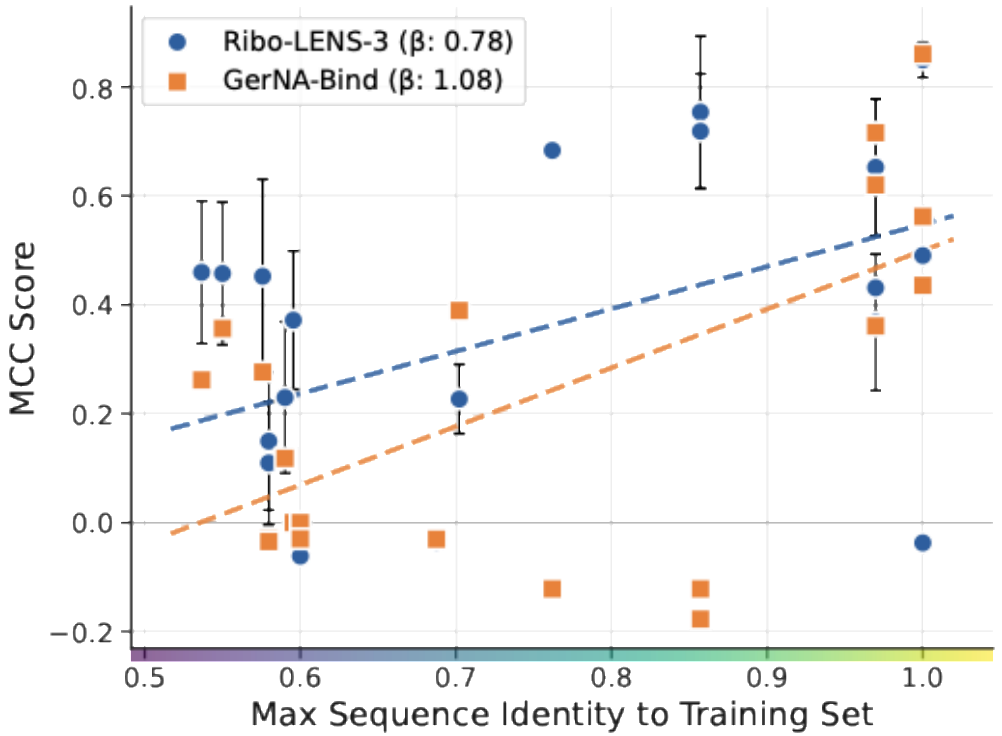




(Nitchi *et al.*, 2026)

Benchmark on GerNA-Bind test set

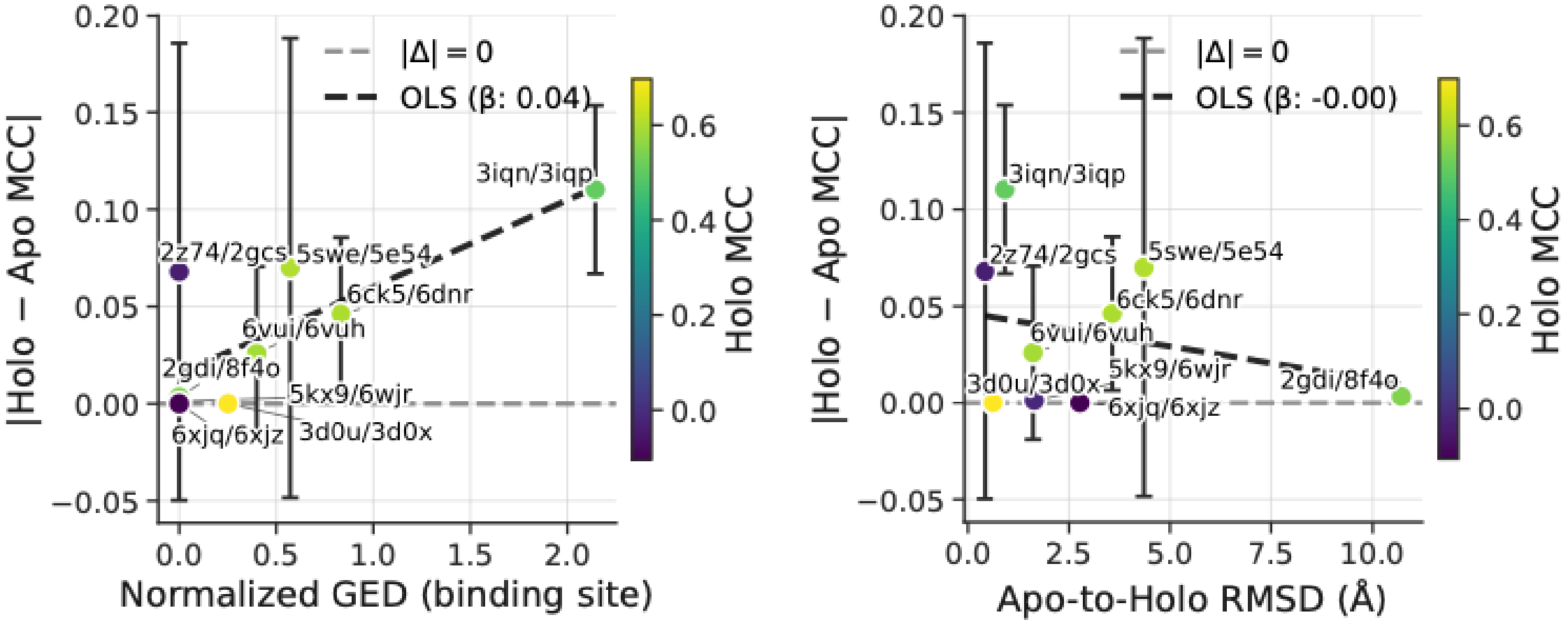




- GerNA-Bind test set (Xia *et al*., 2025) from HARIBOSS (Panei *et al*., 2022).
- Split: training set (structures <2020) & test sets (structures >2020).
- Positive: nucleotides $< 4\text{Å}$

(Nitchi *et al.*, 2026)

APO/HOLO conformational robustness



- Benchmark from fPocketR (Veenbass et al., 2025)
- GED: Graph Edit Distance
- RMDS: Backbone

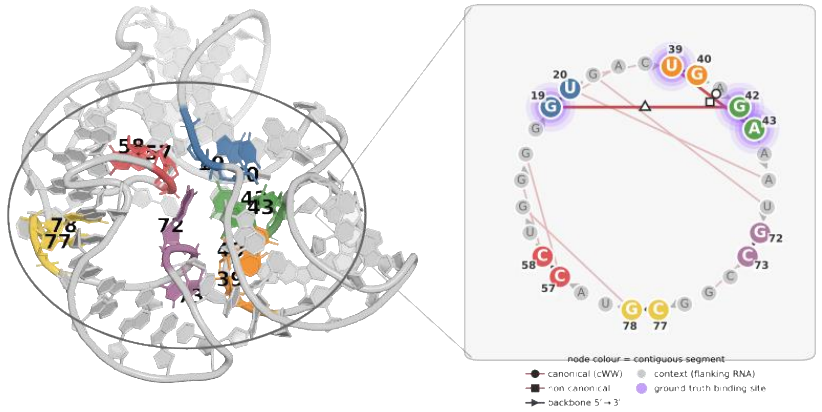
(Nitchi *et al.*, 2026)

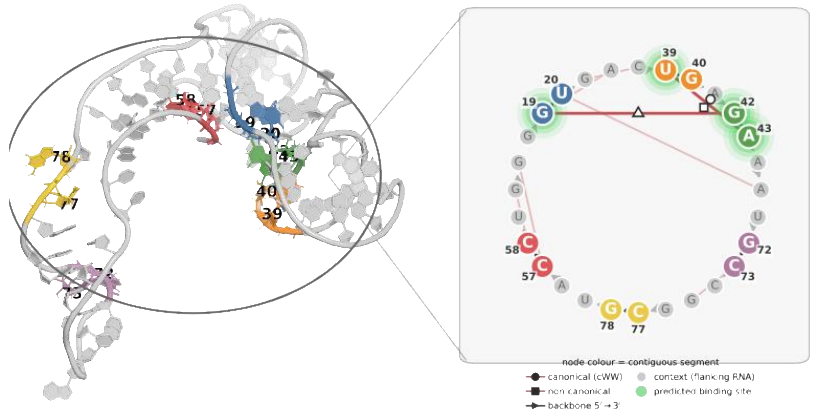
APO/HOLO case study

HOLO

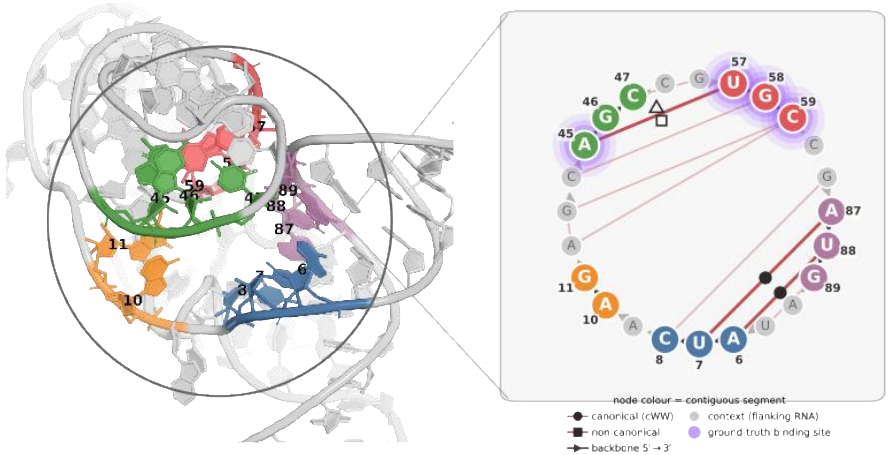
APO

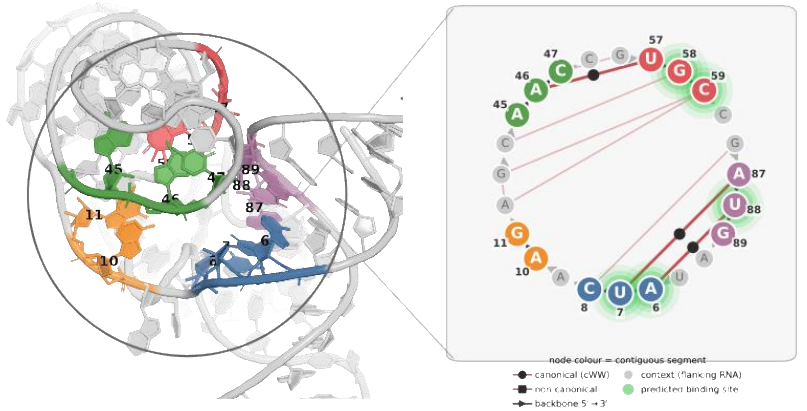
Conserved 2.5D
Distant 3D





Distant 2.5D
Conserved 3D

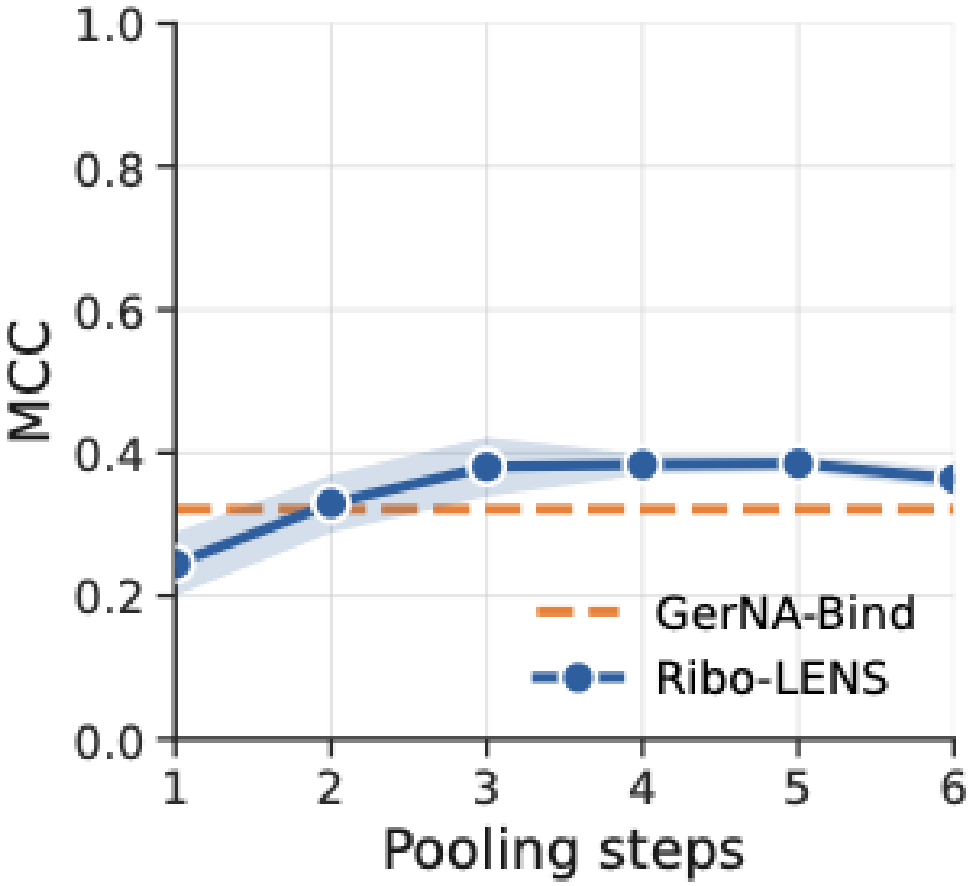


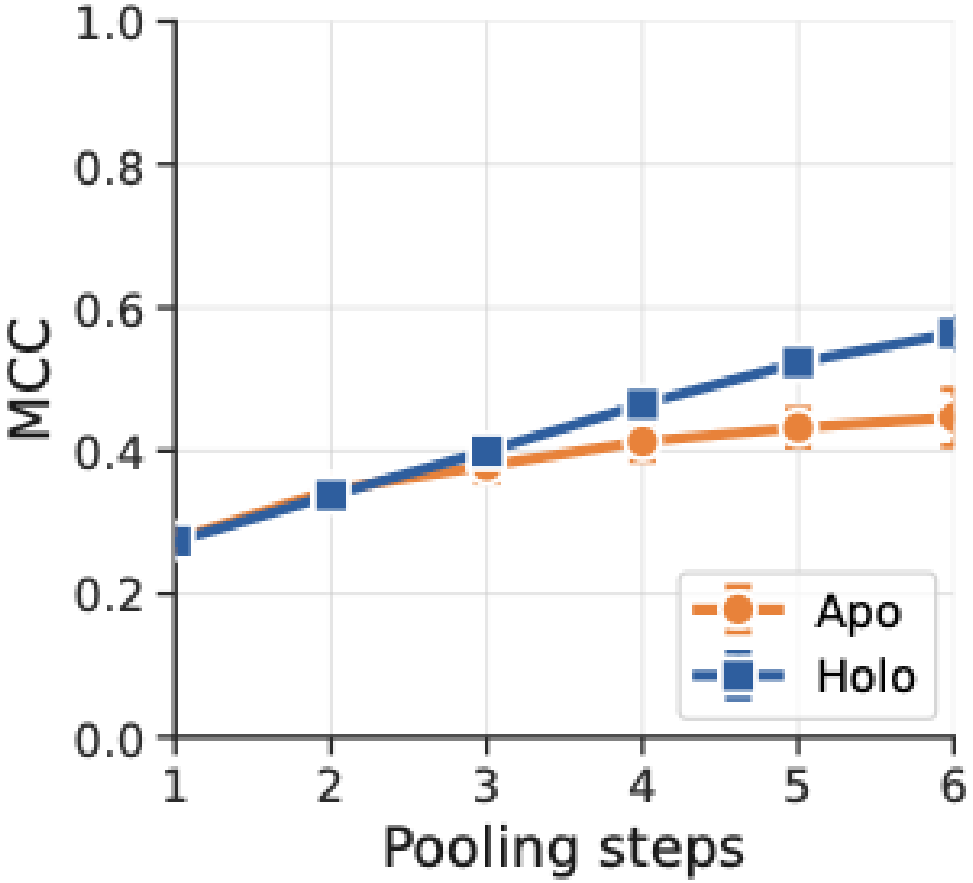


(Nitchi *et al.*, 2026)

Purple glow: True site; Green glow: Ribo-LENS prediction

What is the impact of pooling?

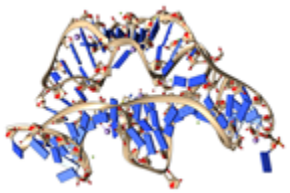




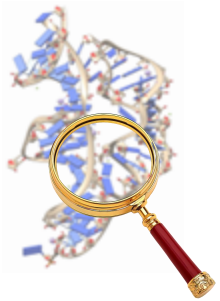
(Nitchi *et al.*, 2026)

Simulation of sequence-based virtual-screening

- ~25,000 small molecules binding against target 27 RNAs
- Retain only RNAs for which we can predict a structure (14)



- Predict 3D structure (RhoFold+)
- Extract 2.5D graph (for Ribo-LENS)

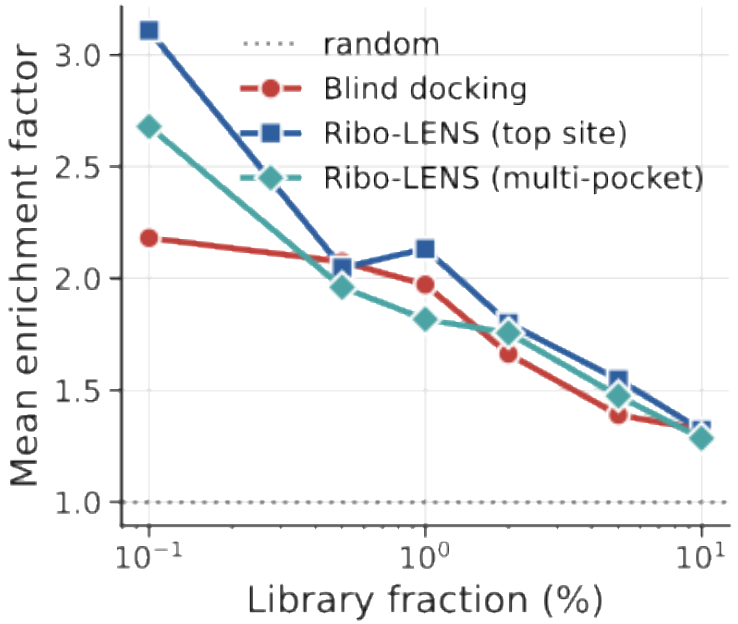


- Predict binding site from 2.5D with Ribo-LENS
- 3 pooling steps, no threshold

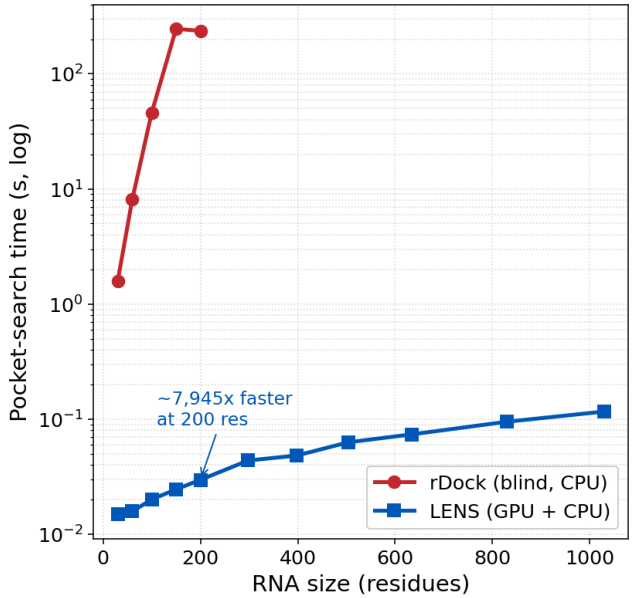
- Dock small molecule library on candidate sites with rDock
- Additional blind docking with rDock for a baseline
- Compute enrichment score

Ribo-LENS vs Blind Docking

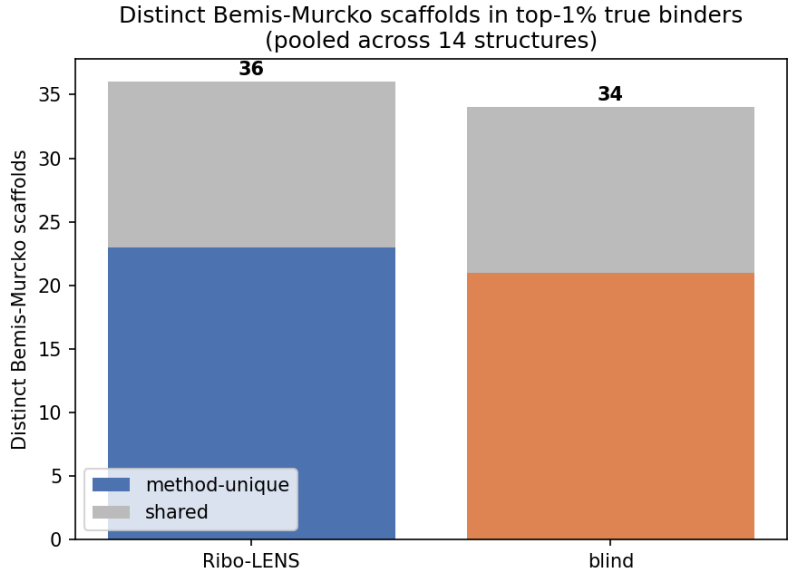
Accuracy



Time



Diversity



Enrichment factor:

$$EF_{x\%} = \frac{a_x / n_x}{A / N}$$

a_x : #compounds in top $x\%$ of ranking
n_x : #experimentally confirmed active among them

(Nitchi *et al.*, 2026)

Carlos G. Oliver
Vanderbilt University

Vincent Mallet
Les Mines

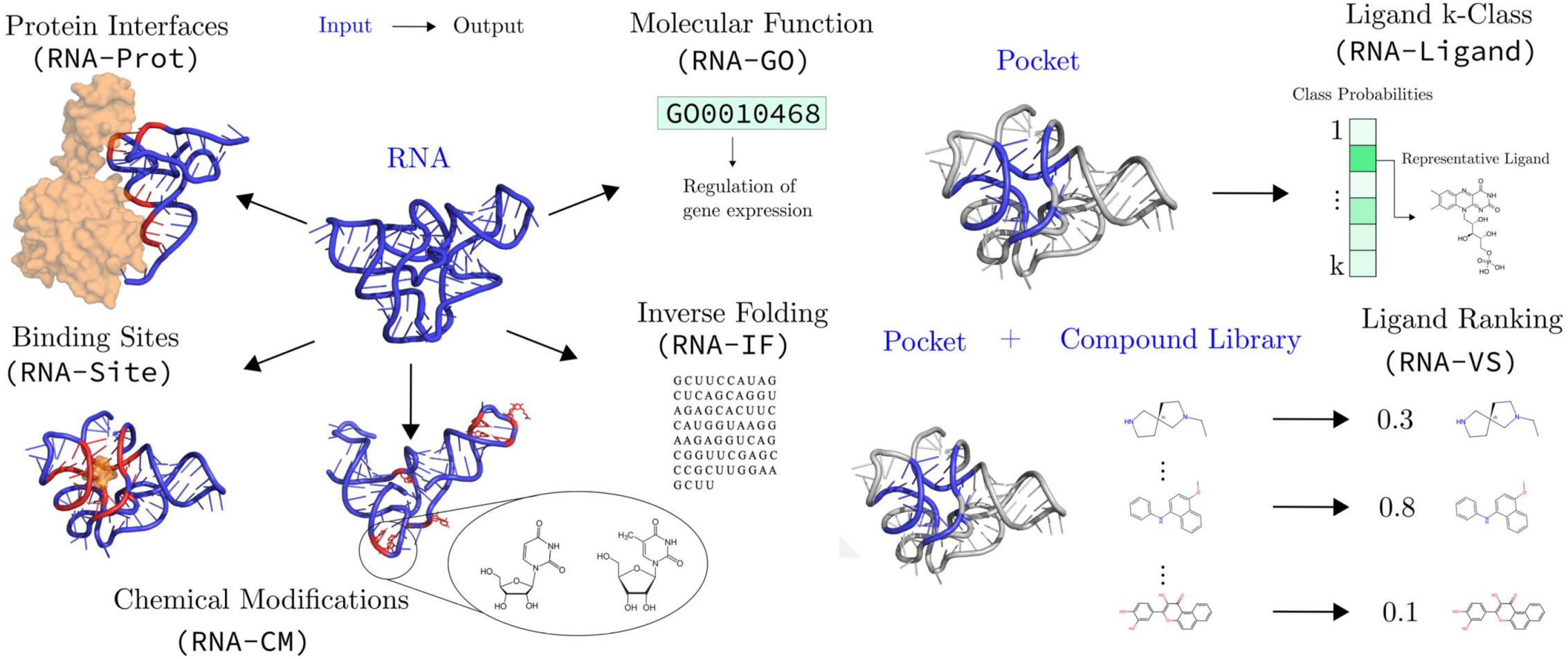
Prediction on small molecules binding RNA

RNAglib

(Mallet *et al*., 2022)

(Oliver et al., 2025)

rnaglib: A framework for RNA ML applications



(Mallet *et al.*, 2022; Wyss *et al.*, 2026)

Shoutout

Reasoning in RNA-targeting drug design

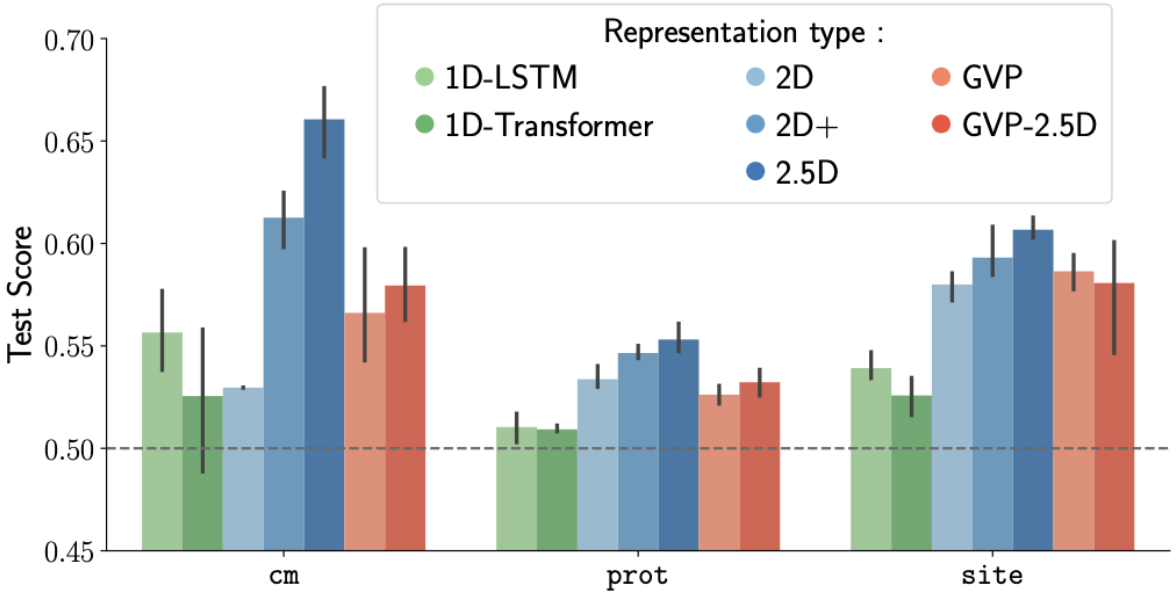
Review and benchmarks of ML methods

	Metric		Without swapping	Target swapping	Performance gap
Boltz 2	AuROC	global	0.535	0.533 ± 0.004	-0.4%
		target	0.530	0.529 ± 0.002	-0.2%
RNAmigos 2	AuROC	target	0.899	0.791 ± 0.007	-12.0%
RSAPred	AuROC	global	0.600	0.670 ± 0.057	+11.7%
		target	0.581	0.449 ± 0.047	-22.7%
DeepRSMA	PCC	global	0.784	0.424 ± 0.064	-46.2%
		target	0.430	0.351 ± 0.021	-22.5%
GerNA-Bind	AuROC	global	0.783	0.514 ± 0.005	-34.4%
		target	0.765	0.529 ± 0.005	-30.9%
RNAsmol (Mol perturbation)	AuROC	global	0.974	0.975 ± 0.000	+0.1%
		target	0.972	0.973 ± 0.002	+0.1%
RNAsmol (Net perturbation)	AuROC	global	0.717	0.546 ± 0.009	-23.8%
		target	0.651	0.569 ± 0.008	-12.6%

(Karroucha *et al.*, 2026)

Comprehensive benchmarking for RNA structure-function modeling

Exploring structure representations with rnaglib



(Wyss *et al.*, 2026)

Acknowledgments

- Jacques Boitreaud, Chai discovery
- David Becerra, McGill
- Juan Carjaval-Patiño, U. Nacional de Colombia
- Arnaud Chol, McGill
- Will Hamilton, MILA
- David Hiraki, McGill
- Vincent Mallet, Les Mines
- Nicolas Moitessier, McGill
- David Nitchi, McGill
- Carlos G. Oliver, Vanderbilt University
- Yann Ponty, CNRS/X
- Étienne Reboul, IBPC
- Vladimir Reinharz, UQàM
- Roman Sarrazin-Gendron, UQàM
- Antoine Soulé, MILA
- Antoine Taly, IBPC
- Éric Westhof, Université de Strasbourg
- Javona Whitebear, McGill
- Hua-Ting Yao, Univ. Vienna

References

Nitchi, D., Waldispühl, J., Oliver, C. Adaptive 2.5D base-pairing subgraph search detects RNA small-molecule binding sites. bioRxiv 2026.07.07.737024 (2026). https://doi.org/10.64898/2026.07.07.737024

Carvajal-Patiño, J.G., Mallet, V., Becerra, D. *et al.* RNAmigos2: accelerated structure-based RNA virtual screening with deep graph learning. *Nat Commun* **16**, 2799 (2025). https://doi.org/10.1038/s41467-025-57852-0

Mallet, V., Oliver, C., Broadbent, J. *et al.* RNAglib: a python package for RNA 2.5 D graphs, *Bioinformatics*, 38:5, 1458–1459 (2022). https://doi.org/10.1093/bioinformatics/btab844

Oliver, C., Mallet, V., Sarrazin Gendron, R. *et al.* Augmented base pairing networks encode RNA-small molecule binding preferences, *Nucleic Acids Research*, 48:14, 690–7699 (2020). https://doi.org/10.1093/nar/gkaa583