RNAsolo

in the RNA structure study

Computational Approaches to RNA Structure and Function, Benasque 2024

Maciej Antczak
Institute of Computing Science, PUT,
Institute of Bioorganic Chemistry, PAS

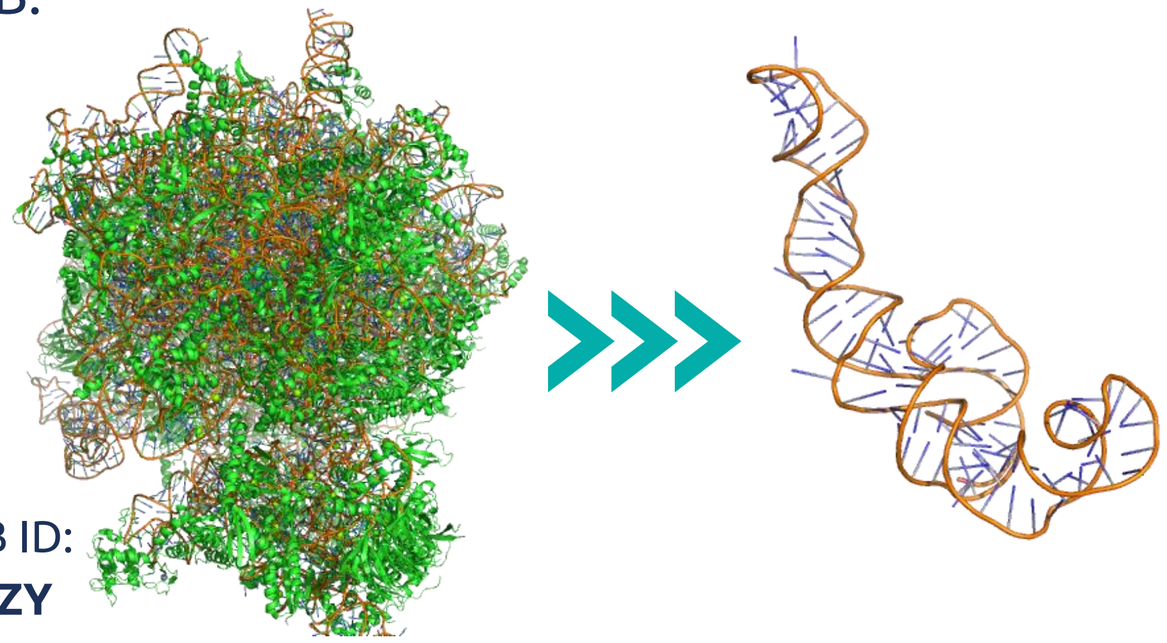
Motivation

Goal: You developed a new tool working on 3D RNA structures and you want to assess its quality and performance.

Question: What can you do?

1. Download all 3D RNA structures from the PDB.
2. Filter all non-RNA chains, ligands, and ions.
3. Process *somehow* modified residues.
4. Eliminate redundancy from sequential or family-oriented point of view.

THIS PROCESS MIGHT BE REALIZED OVER AND OVER AGAIN BY RESEARCHERS INTERESTED IN 3D RNA STRUCTURE ANALYSIS AND PREDICTION!



PDB ID: **7PZY**

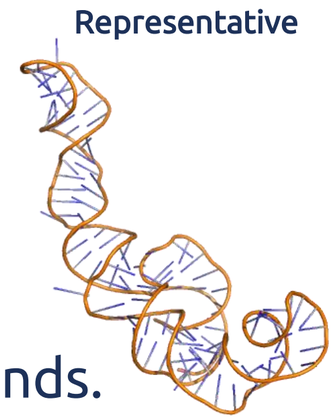
Motivation (2)

Goal: You need to develop a non-redundant and comprehensive training set for machine learning-based models.

Problem: You need multimodal data describing 3D RNA structures.

1. Sequence (FASTA).
2. Secondary structure (BPS, EXDBN).
3. Tertiary structure (PDB/mmCIF).
4. Torsion angles.
5. Interactions with other molecular compounds.

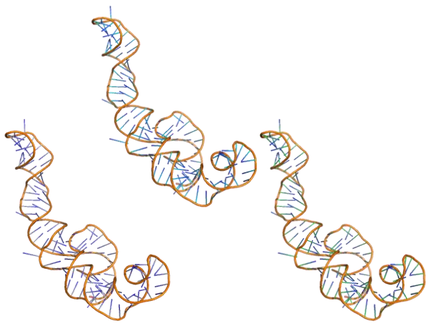
HOW TO EFFICIENTLY ELIMINATE REDUNDANCY?



Representative

Selected chain of **7PZY**

&



Class members

Structurally similar chains: **7PZY, 7Q0F, 7Q0P**

=

Equivalence class

06697.1

BGSU

What is the solution?
RNAsolo^2

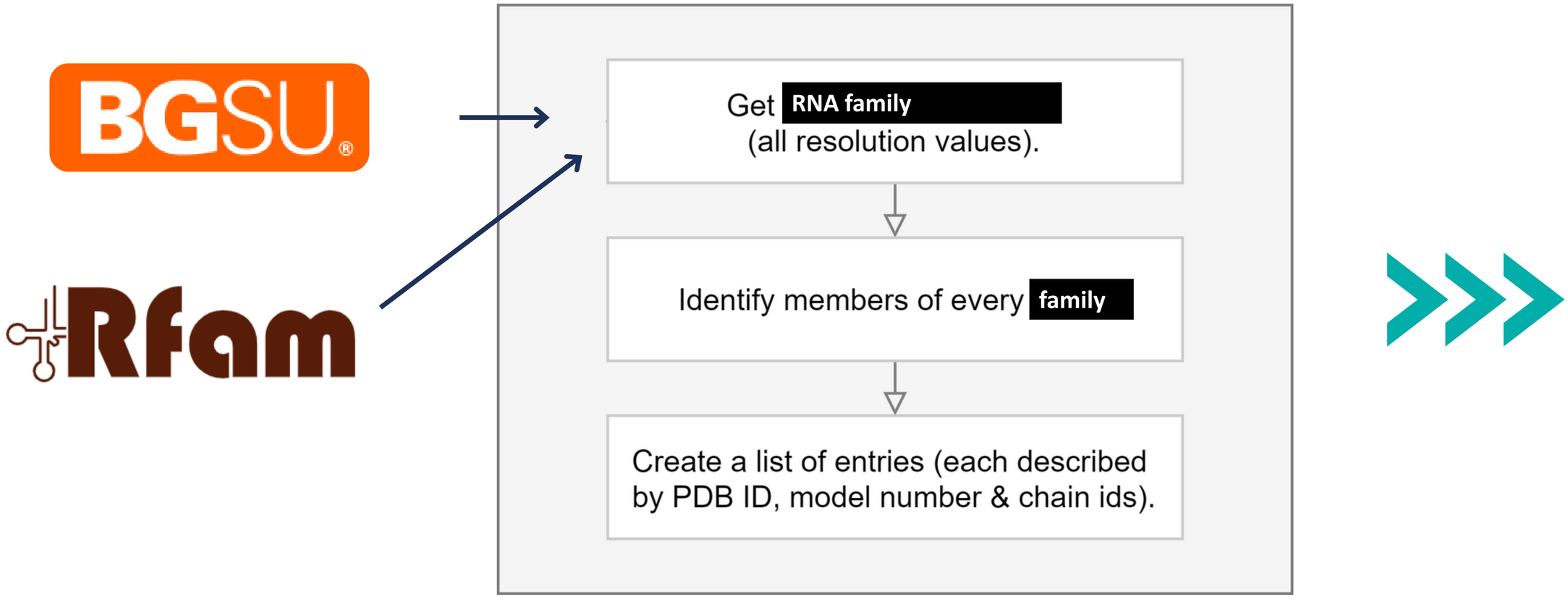
Desktop
Mobile

https://rnasolo.cs.put.poznan.pl/
https://mirna.cs.put.poznan.pl/

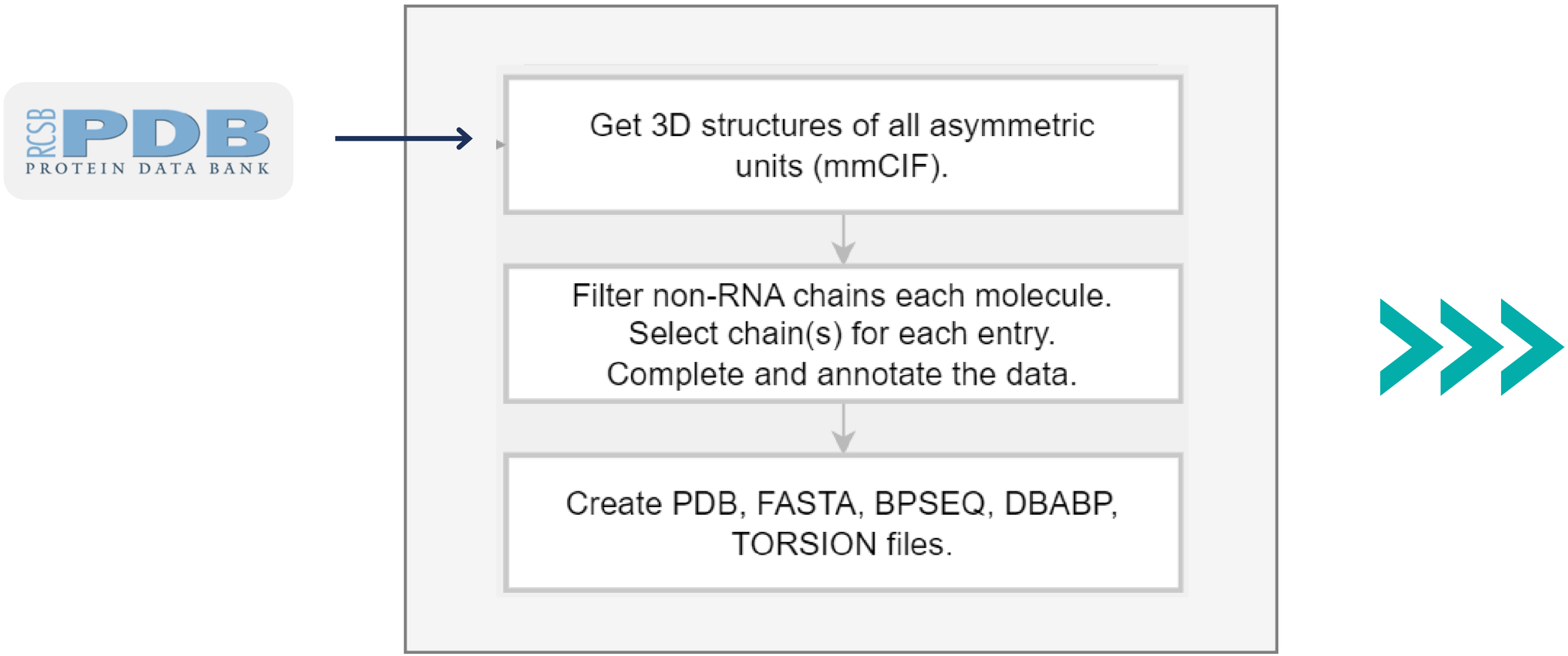
API endpoints

- /api/query/structure/package?pdbid=430D&format=CIF&models=1
- /api/query/structure?pdbid=6VFF&format=CIF&chains=D,C&model=1
- /api/query/class?identifier=02656&format=CIF
- /api/query/structure/contact?pdbid=8OIP&chains=AA

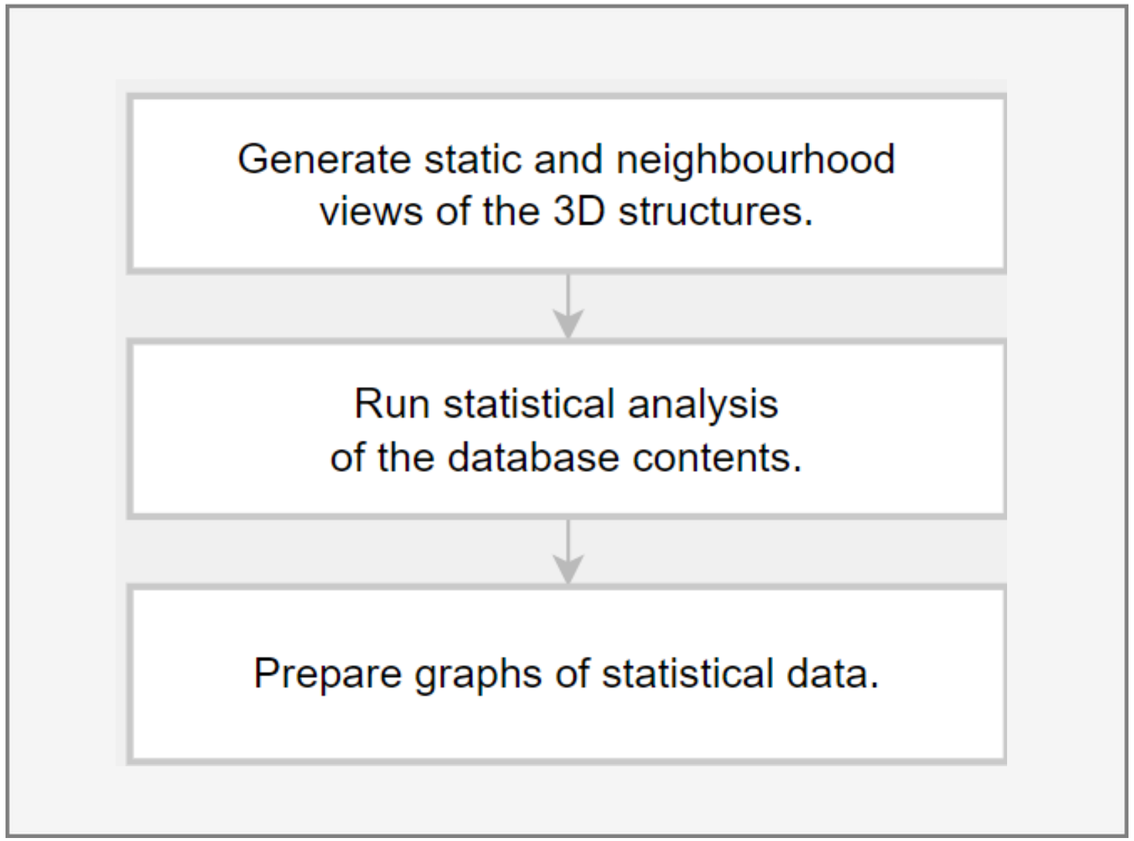
Retrieve a set of RNA entries



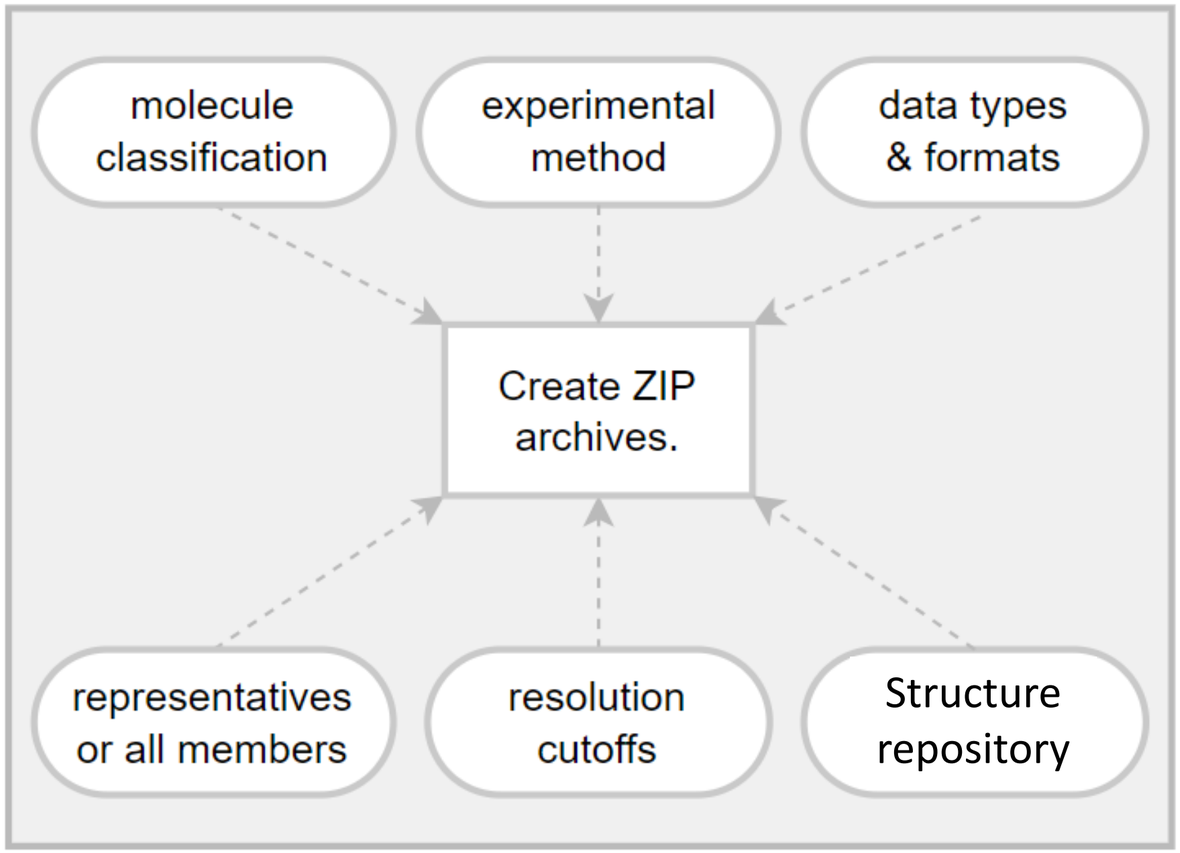
Prepare structural data



Analyze and visualize the data



Create benchmark sets ready to download



Home page (BGSU)

RNAsolo²

RNAsolo is a self-updating database that collects experimentally determined 3D RNA structures, cleans them from non-RNA data, and groups them into equivalence classes. After th
recent update (**06-06-2024**), *RNAsolo* contains **14,886** RNA structures clustered into **3,340** equivalence classes - each of the latter is exemplified by a cluster representative.

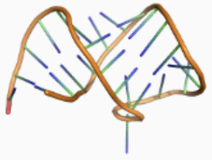
Browse RNAsolo by:

Representative sets | Rfam families | PDB id

Download CSV

Representative	Equivalence class

PDB ID: 8JHP
Model: 1
Chain(s): A
Molecule: Solo RNA
Method: NMR Spectroscopy
#nts: 26



Identifier: 60832.1
Cardinality: 1

Download

representative structure

Results in format(s):

Click here to specify file format

- PDB
- PDBx/mmCIF
- FASTA
- DBABP
- BPSEQ
- Torsion angles (CSV)

PDB ID: 8C0O

Home page (PDBs distribution)

RNAsolo²

RNAsolo is a self-updating database that collects experimentally determined 3D RNA structures, cleans them from non-RNA data, and groups them into equivalence classes. After the recent update (**06-06-2024**), *RNAsolo* contains **14,886** RNA structures clustered into **3,340** equivalence classes - each of the latter is exemplified by a cluster representative.

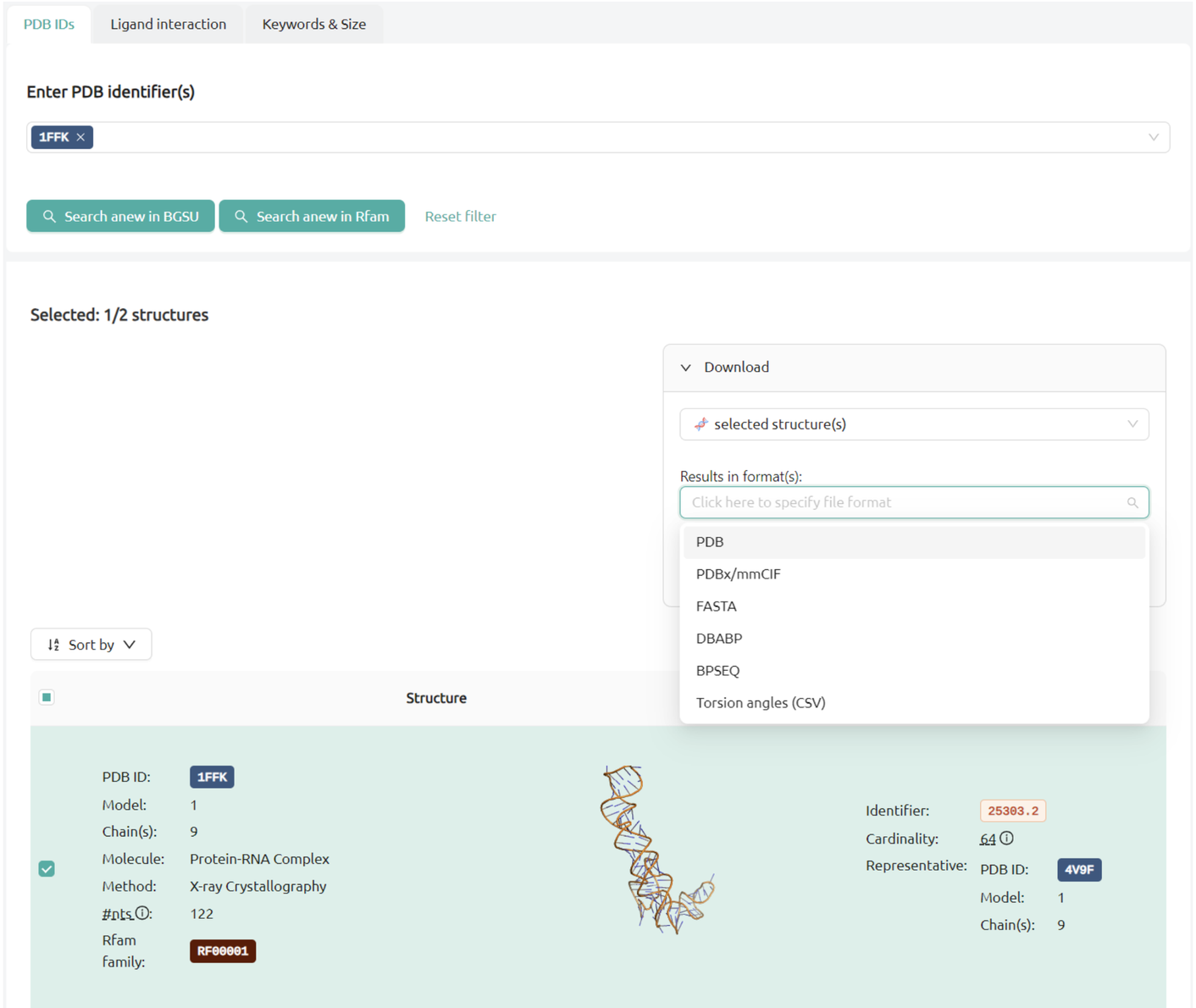
Browse RNAsolo by:

Representative sets | Rfam families | PDB id

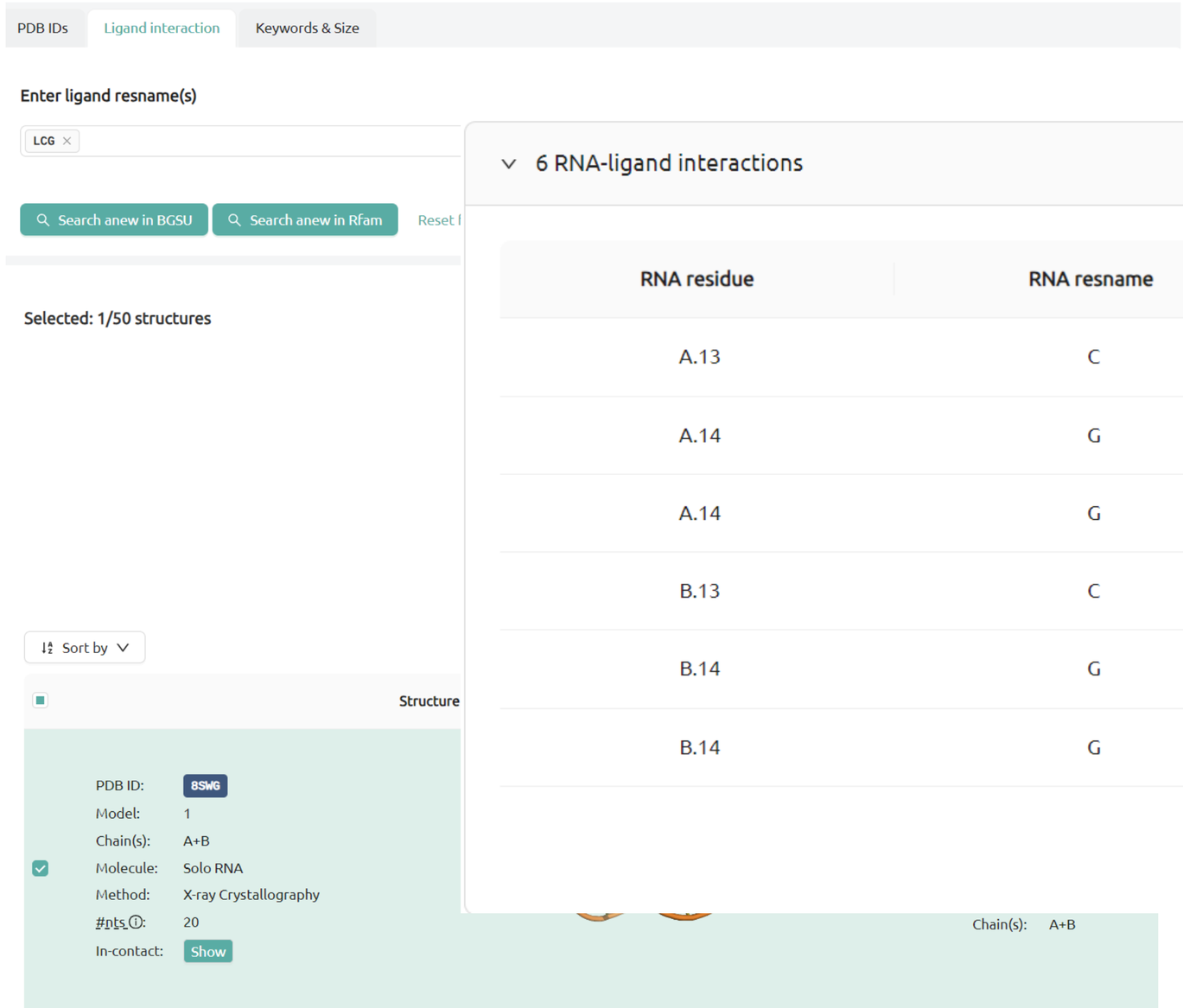
Download CSV

	PDB id	#members of Rfam family	#members of equivalence class
+	1EFW	2	2

Search the data via PDB ID(s)



Search the data via ligand symbol

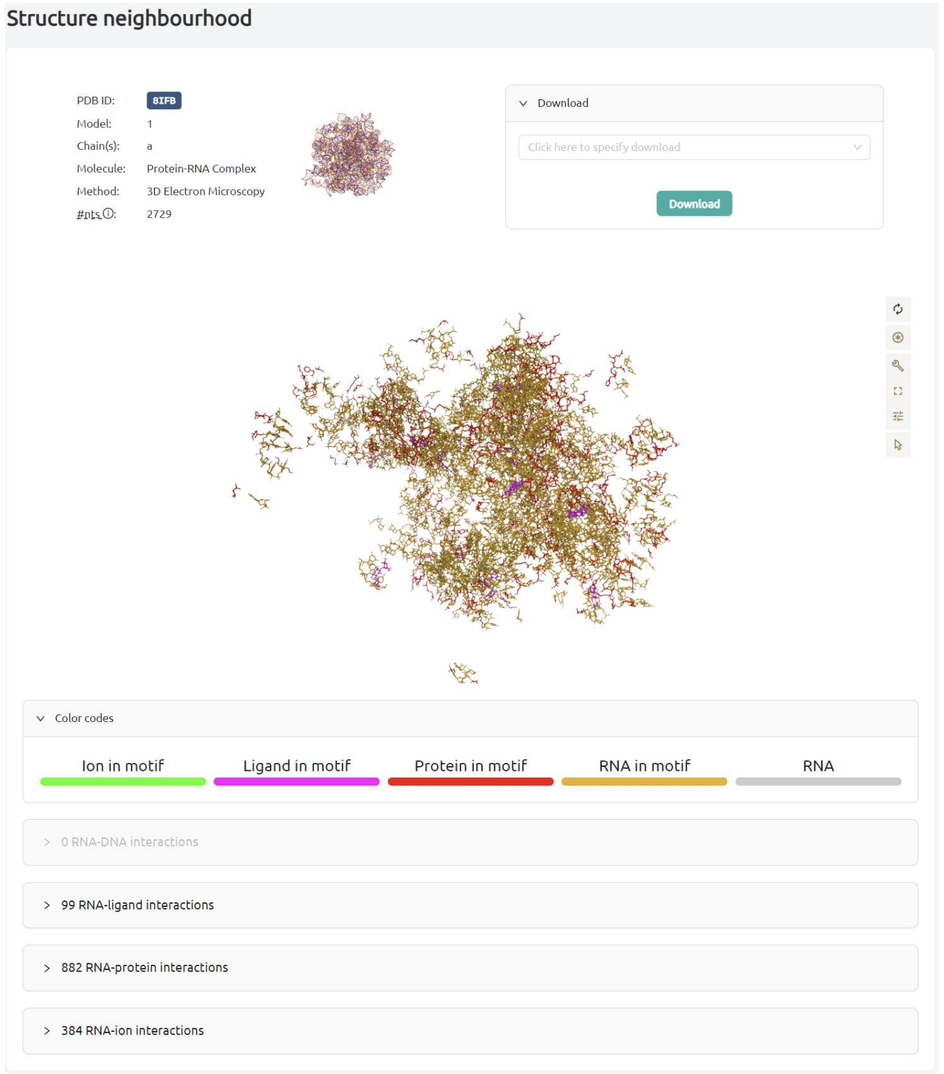


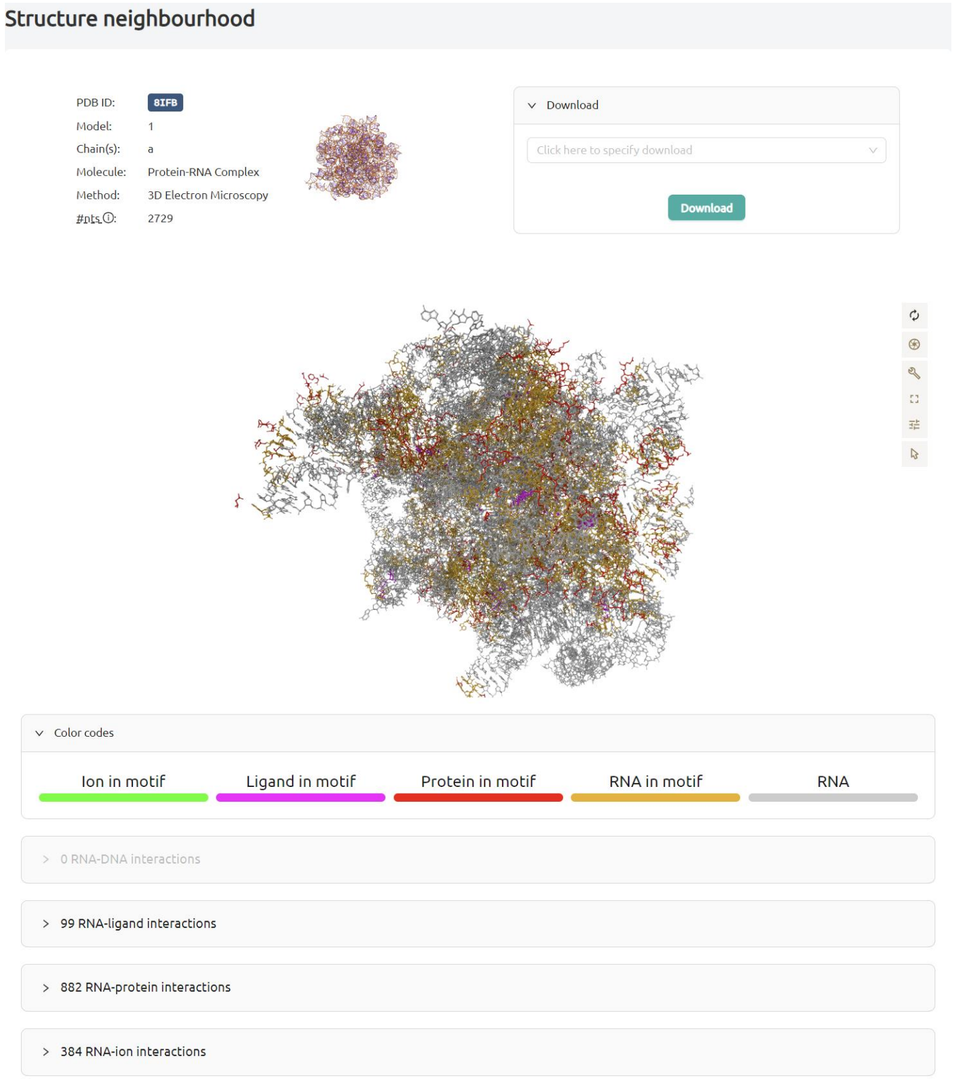
6 RNA-ligand interactions

RNA residue	RNA resname	Ligand id	Ligand name
A.13	C	B.4	LCG
A.14	G	B.101	G3A
A.14	G	B.3	LCC
B.13	C	A.4	LCG
B.14	G	A.101	G3A
B.14	G	A.3	LCC

1

Intermolecular interactions





https://mirna.cs.put.poznan.pl/neighbourhood/23e7090c-4bb0-42cb-b173-fc50310896bb

Intermolecular interactions (2)

882 RNA-protein interactions

RNA residue	RNA resname	Protein resid	Protein resname
a.6	A	i.132	HIS
a.6	A	i.135	GLN
a.14	A	z.18	SER
a.15	G	z.14	GLY
a.16	C	z.11	SER
a.18	U	p.26	GLY
a.18	U	p.30	ARG
a.19	A	p.23	GLY
a.23	G	r.77	ASP
a.24	G	r.78	GLU

< 1 2 3 4 5 ··· 89 > 10 / page

99 RNA-ligand interactions

RNA residue	RNA resname	Ligand id	Ligand name
a.28	A	a.3175	SPD
a.325	G	a.3173	SPD
a.326	G	a.3173	SPD
a.327	G	a.3173	SPD
a.333	G	a.3173	SPD
a.338	G	a.3173	SPD
a.387	U	a.3172	SPD
a.410	G	a.3171	SPD
a.411	G	a.3171	SPD
a.447	A	a.3175	SPD

< 1 2 3 4 5 ··· 10 > 10 / page

384 RNA-ion interactions

RNA residue	RNA resname	Ion id	Ion name
a.24	G	a.3120	MG
a.25	U	a.3120	MG
a.31	C	a.3210	MG
a.120	U	a.3002	MG
a.120	U	a.3001	MG
a.124	G	a.3002	MG
a.192	C	a.3089	MG
a.195	A	a.3092	MG
a.195	A	a.3122	MG
a.196	A	a.3122	MG

< 1 2 3 4 5 ··· 39 > 10 / page

https://mirna.cs.put.poznan.pl/neighbourhood/23e7090c-4bb0-42cb-b173-fc50310896bb

Q&A?

Bartosz Adamczyk Marta Szachniuk

Thank you for your attention!

JOURNAL ARTICLE

RNAsolo: a repository of cleaned PDB-derived RNA 3D structures

Bartosz Adamczyk, Maciej Antczak, Marta Szachniuk

Bioinformatics, Volume 38, Issue 14, July 2022, Pages 3668–3670,
https://doi.org/10.1093/bioinformatics/btac386
Published: 08 June 2022 **Article history**

Abstract

Motivation

The development of algorithms dedicated to RNA three-dimensional (3D) structures contributes to the demand for training, testing and benchmarking data. A reliable source of such data derived from computational prediction is the RNA-Puzzles repository. In contrast, the largest resource with experimentally determined structures is the Protein Data Bank. However, files in this archive often contain other molecular data in addition to the RNA structure itself, which —to be used by RNA processing algorithms—should be removed.

Results

RNAsolo is a self-updating database dedicated to RNA bioinformatics. It systematically collects experimentally determined RNA 3D structures stored in the PDB, cleans them from non-RNA chains, and groups them into equivalence classes. It allows users to download various subsets of data—clustered by resolution, source, data format, etc.—for further processing and analysis with a single click.

Availability and implementation

The repository is publicly available at https://rnasolo.cs.put.poznan.pl.

Issue Section: APPLICATIONS NOTES > DATABASES AND ONTOLOGIES

Associate Editor: Christina Kendziorski