

# Identification of RNA sequence and structure motifs for protein interaction

Gerardo Pepe<sup>1,#</sup>, Andrea Guarracino<sup>1,#</sup>, Francesco Ballesio<sup>1</sup>, Marta Adinolfi<sup>1</sup>, Marco Pietrosanto<sup>1</sup>, Elisa Sangiovanni<sup>1</sup>, Ilio Vitale<sup>2,3</sup>, Gabriele Ausiello<sup>1</sup>, Manuela Helmer –Citterich<sup>1</sup>

<sup>1</sup> Department of Biology, University of Rome “Tor Vergata”, Rome, Italy.

<sup>2</sup> IIGM – Italian Institute for Genomic Medicine, c/o IRCSS Candiolo, Italy.

<sup>3</sup> Candiolo Cancer Institute, FPO – IRCCS, Candiolo, Italy.

# The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

## Introduction

RNA-binding proteins (RBPs) play a central role in RNA metabolism, in transcript regulation, splicing, stability, and translation, as well as in the regulation of non coding RNAs.

The BRIO web server (1) enables the identification of sequence and structure protein binding motifs in a set of one or more RNA molecules.

The BRIO motifs dataset encompasses 2296 sequence and 2508 structure motifs, associated with 186 RNA binding proteins and 69 protein domains by analyzing the results of several CLIP experiments (2).

BRIO takes advantage of the BEAR encoding to represent the structural motifs. This string encoding allows us to include also the structural context of each nucleotide in the secondary structure representation, without increasing the algorithm complexity (3).

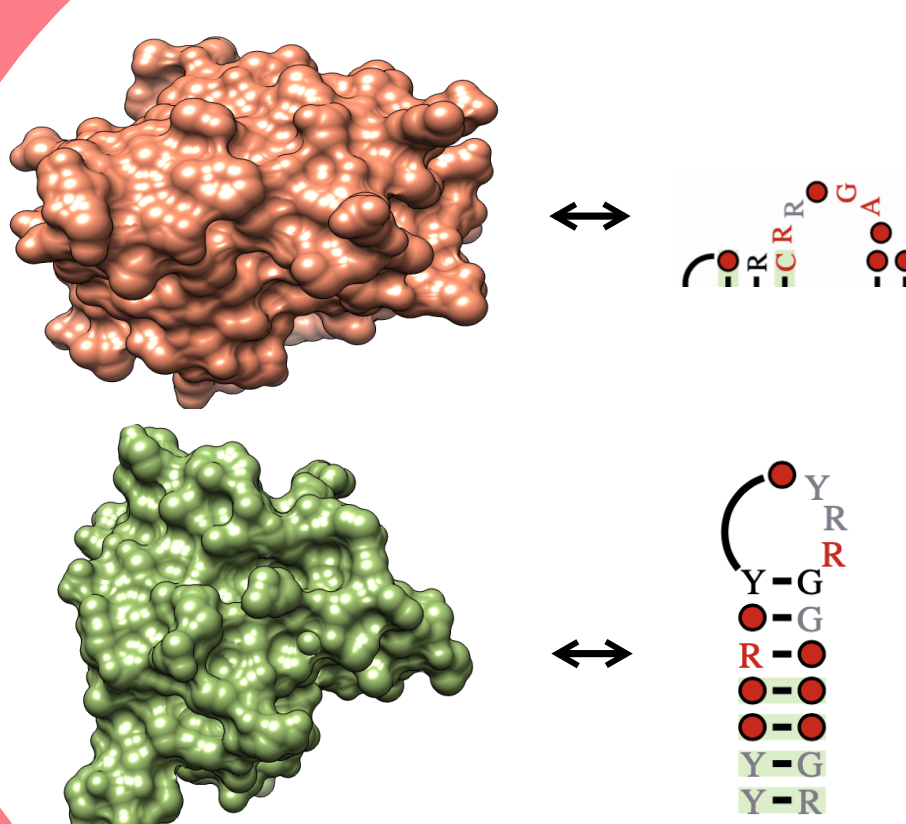
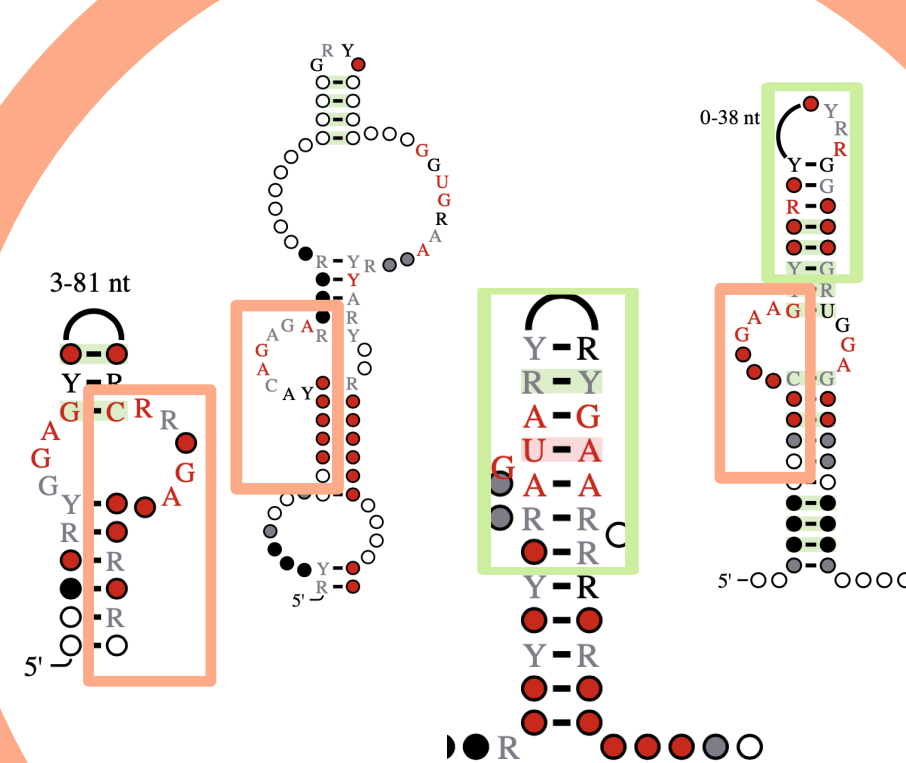
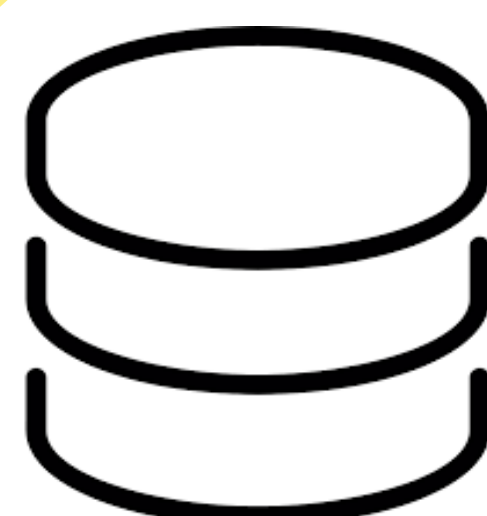
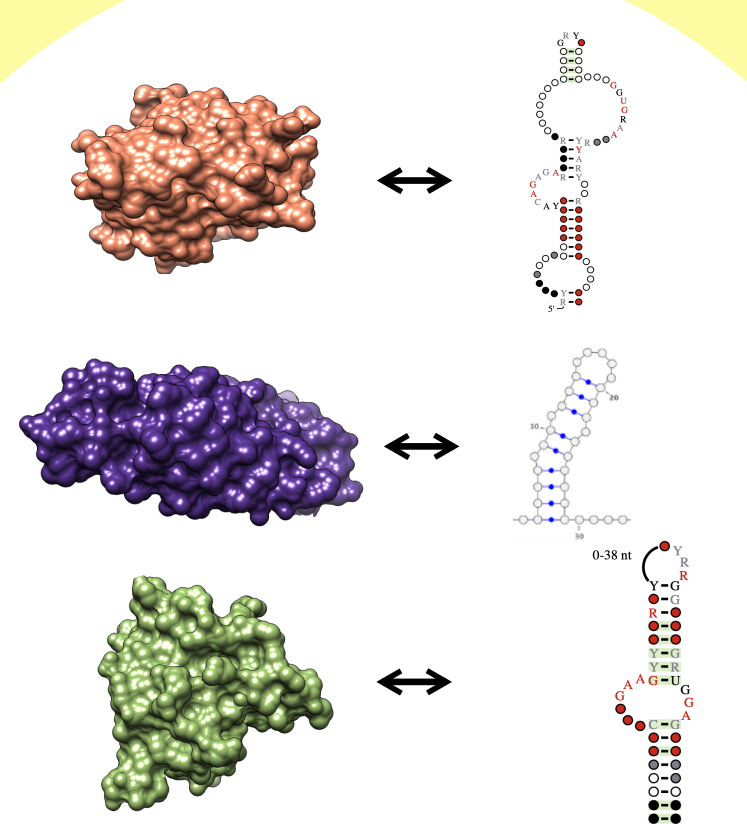
## Conclusion

Together with its friendly interface, BRIO can support scientists in their investigations on groups of RNA molecules of interest, their putative RBPs, and the roles these proteins play in RNA regulation.

To the best of our knowledge, no other web server exists offering the possibility to search for sequence and structure motifs associated with RNA binding proteins.



Beam RNA Interaction motif search tool



## Aim

The BRIO webserver allows the user to search on a database of sequence and structure motifs specifically associated with RNA binding proteins in a set of one or more input RNA molecules.

## Methods

The input RNA molecules are:  
– folded using RNAfold (4)  
– encoded using the BEAR alphabet (2)  
– scanned for sequence and structure motifs of our database (3).  
One-sided Fisher's Test is applied to evaluate if a motif is enriched in the input RNA molecules with respect to a set of background RNAs.

	qBEAR		qBEAR
STEM	z (1-5nt)	INTERNAL LOOP	c (2-6nt, left/right)
	a (6-9nt)		d (7-10nt, left/right)
	q (>10nt)		e (>10nt, left/right)
STEM BRANCH	v (1-5nt)	INTERNAL LOOP BRANCH	n (2-9nt)
	f (6-9nt)		h (10-18nt)
	r (>10nt)		y (19-22nt, >23nt)
LOOP	x (3-9nt)	BRANCH	t
	s (10-17nt)	BULGE	b (left/right)
	w (>18nt)	BULGE BRANCH	g (left/right)

## Results/Outputs

BRIO returns a collection of protein binding motifs identified in the input RNA molecules associated with their corresponding RBPs from CLIP experiments.

Logo	Type	Region	Coverage	p-value	Experiment	Protein	Domains	Cell line	Experiment info	Organism	Download
	Sequence	CDS	1	0.013	PARCLIP	WTAP			Link	Homo sapiens	Download
	Sequence	UTR	1	0.016	PARCLIP	ELAVL1/NMASE			Link	Homo sapiens	Download
	Structure	UTR	1	0.017	PARCLIP	FMR1			Link	Homo sapiens	Download
	Structure	CDS	1	0.02	PARCLIP	FMR1			Link	Homo sapiens	Download
	Structure	CDS	1	0.024	PARCLIP	HUR/Nase			Link	Homo sapiens	Download
	Sequence	CDS	1	0.031	PARCLIP	FMR1			Link	Homo sapiens	Download

Enriched Motifs		Sequences	
Name	# Sequence motifs	# Structure motifs	Length
+ chr1:149783661-149783992(-)	3	7	332

Start	End	Enriched Motif	Type	Protein	Experiment
90	96	UCUUGUC	Sequence	ELAVL1/NMASE	PARCLIP
96	127	b[***]d[d](ccmmmmccdd[d]333b)	Structure	FMR1	PARCLIP
99	124	b[***]d[d](ccmmmmccdd[d]333b)	Structure	CAPRIN1	PARCLIP
99	125	b[***]d[d](ccmmmmccdd[d]333b)	Structure	HuRMNase	PARCLIP
138	144	CCUGUC	Sequence	WTAP	PARCLIP
184	190	AGAAGGA	Sequence	FMR1	PARCLIP
280	295	eeeeeeccccccccc	Structure	QKI	PARCLIP
281	296	eeeeeeccccccccc	Structure	FMR1	PARCLIP
281	297	eeeeeeccccccccc	Structure	AGO2	PARCLIP
282	295	eeeeeeccccccccc	Structure	ALKBH5	PARCLIP
+ chr1:149784741-149784985(-)	3	7	245		

The web server is freely available at <http://brio.bio.uniroma2.it>.  
The source code is available at:  
<https://github.com/helmercitterich-lab/BRIO>.

- Guarracino A, Pepe G, Ballesio F, Adinolfi M, Pietrosanto M, Sangiovanni E, Vitale I, Ausiello G, Helmer –Citterich M. BRIO: a web server for RNA sequence and structure motif scan. Nucleic Acids Res (In Press).
- Mattei,E., Ausiello,G., Ferrè,F. and Helmer-Citterich,M. (2014) A novel approach to represent and compare RNA secondary structures. Nucleic Acids Res., 42, 6146–6157
- Adinolfi,M., Pietrosanto,M., Parca,L., Ausiello,G., Ferrè,F. and Helmer-Citterich,M. (2019) Discovering sequence and structure landscapes in RNA interaction motifs. Nucleic Acids Res., 47, 4958–4969.
- Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. Algorithms Mol. Biol., 6, 26.

