RNAlign2D – a rapid tool for combined RNA structure and sequence-based alignment using pseudo-amino acid substitution matrix

Tomasz Woźniak[1], Małgorzata Sajek[2], Jadwiga Jaruzelska[1], Marcin Sajek[1]

INSTITUTE OF HUMAN GENETICS
POLISH ACADEMY OF SCIENCES

1. Institute of Human Genetics, Polish Academy of Sciences, Poznań, Poland
2. Department of Human Molecular Genetics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

INTRODUCTION

The functions of RNA molecules are mainly determined by their secondary structures. These functions can also be predicted using bioinformatic tools that enable the alignment of multiple RNAs to determine functional domains and/or classify RNA molecules into RNA families. Here, we introduce an extremely fast Python-based tool called RNAlign2D. This tool is dedicated to multiple alignment of RNA molecules with known secondary structures. It converts RNA sequences to pseudo-amino acid sequences that incorporate structural information and uses a customizable scoring matrix to align these RNA molecules using the multiple protein sequence alignment tool MUSCLE. This approach can be customized for virtually all protein aligners. RNAlign2D is freely available from https://github.com/tomaszwozniakihg/rnalign2d .

MATERIALS AND METHODS

RNAlign2D is a command line tool written as a Python script that works in UNIX-based operation systems. To compare RNAlign2D with other tools that can use fixed 2D structure for multiple RNA alignment LocARNA and CARNA, we used 2 available benchmark datasets: BraliBase 2.1 and RNAStralign.In the next step, the sum-of-pairs-scores (SPSs) and positive predictive value (PPVs) were calculated for each alignment. Alignment time was also measured for subset of datasets from RNAStralign benchmark.

RESULTS

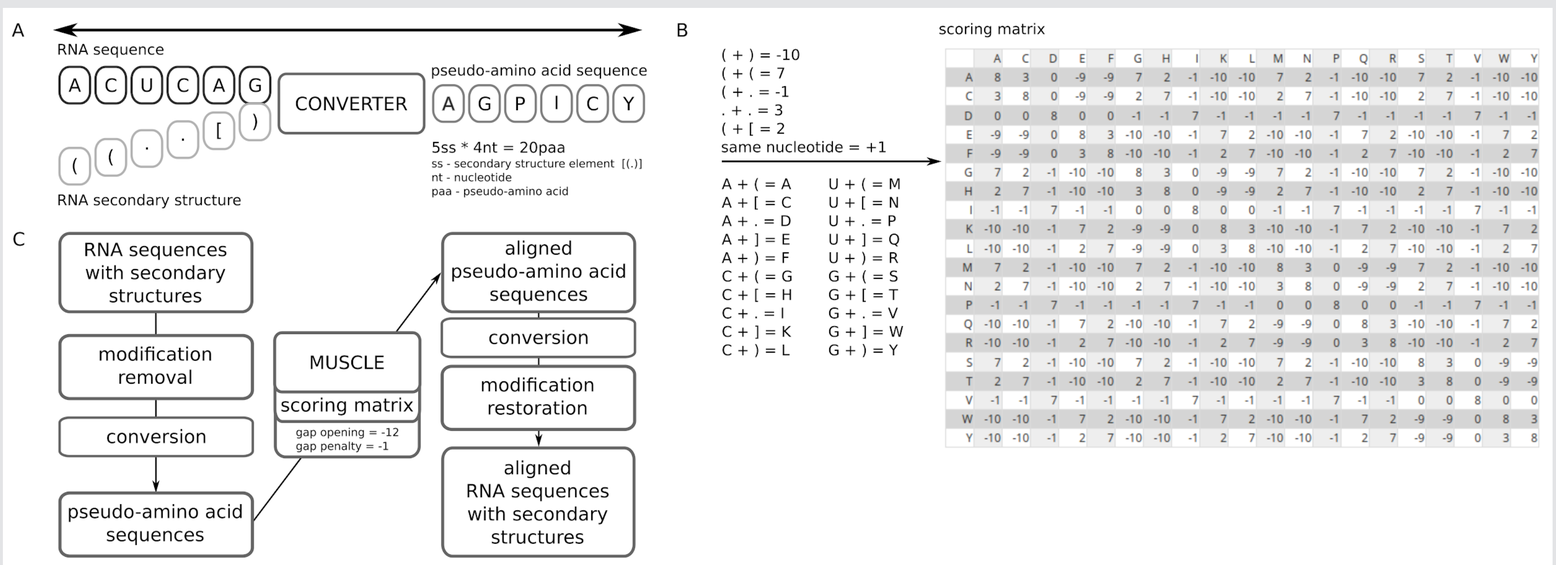


Figure 1. Schematic representation of the RNAlign2D workflow. **A**. Basic concept of RNA sequence-structure conversion to a pseudo-amino acid sequence. **B**. Conversion of 20 RNA sequence-structure elements to pseudo-amino acids and their scores (left) and the default scoring matrix (right). **C**. Block diagram of the RNAlign2D workflow.

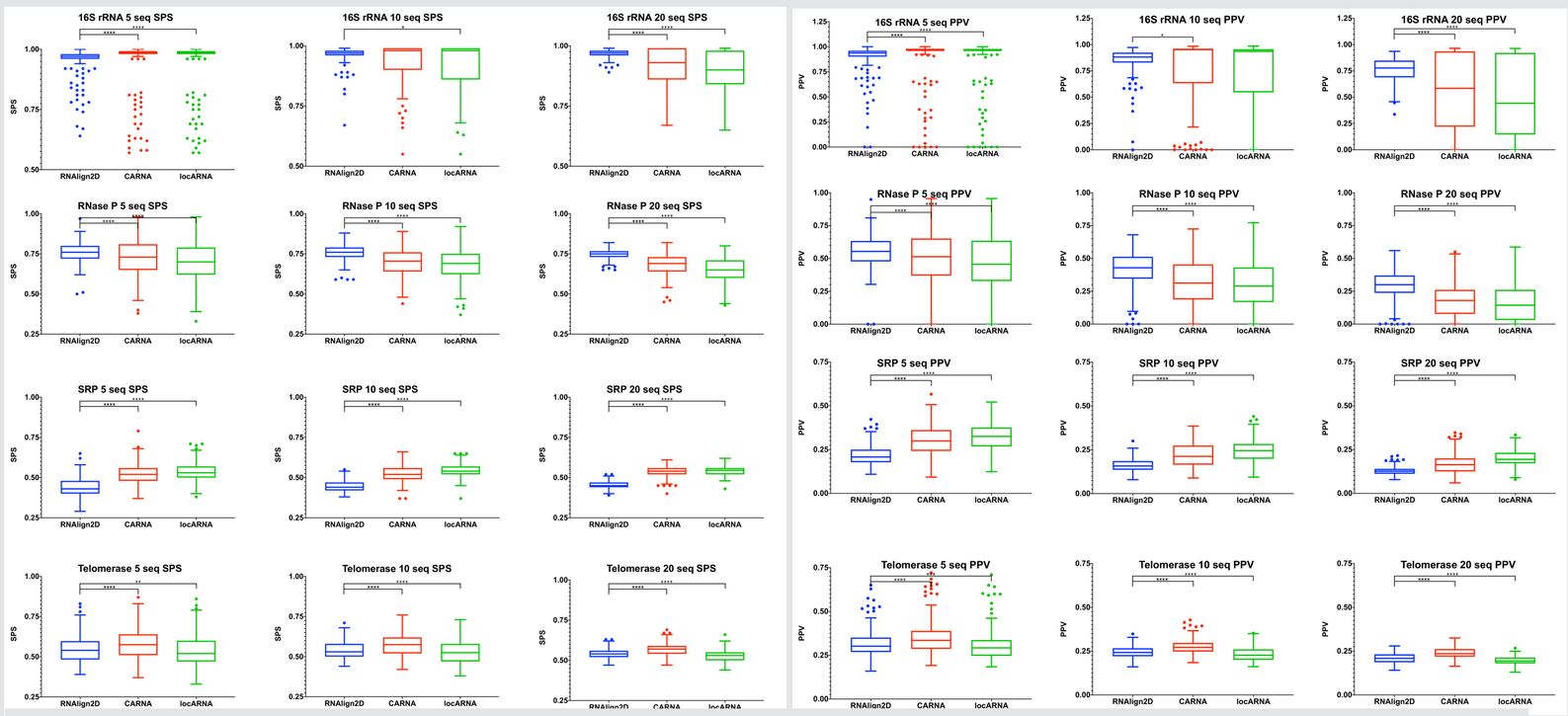


Figure 2. Box and whisker plots comparing SPSs (left) and PPV (right) for the alignment of 200 groups of 5, 10 and 20 homologous sequences from the entire RNAStralign benchmark dataset with RNAlign2D, CARNA and LocARNA. **** p-value < 0.0001; *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05.

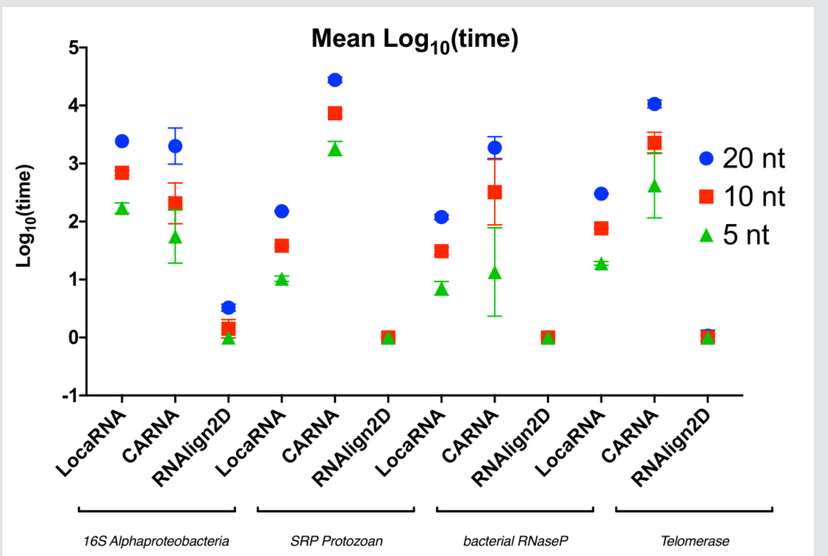


Figure 3. Comparison of alignment performance times for RNAlign2D, CARNA and LocARNA presented as a graph with standard errors indicated.

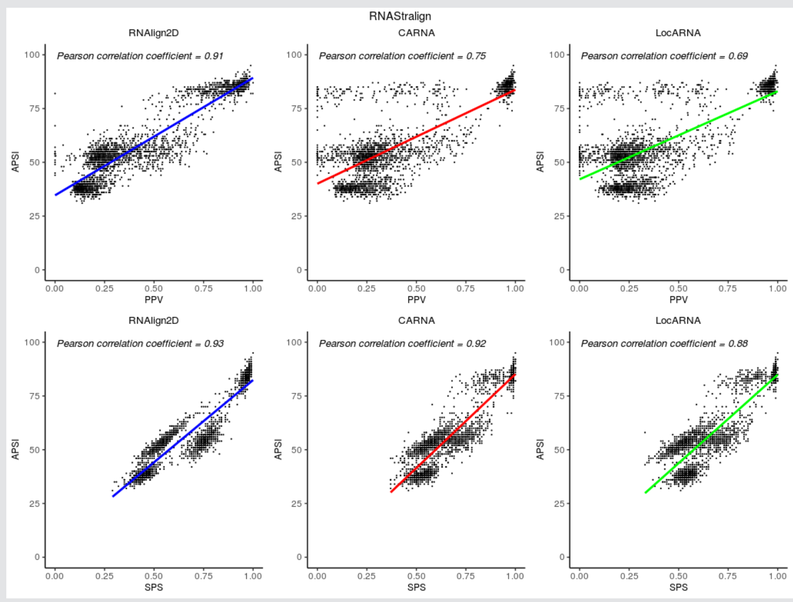
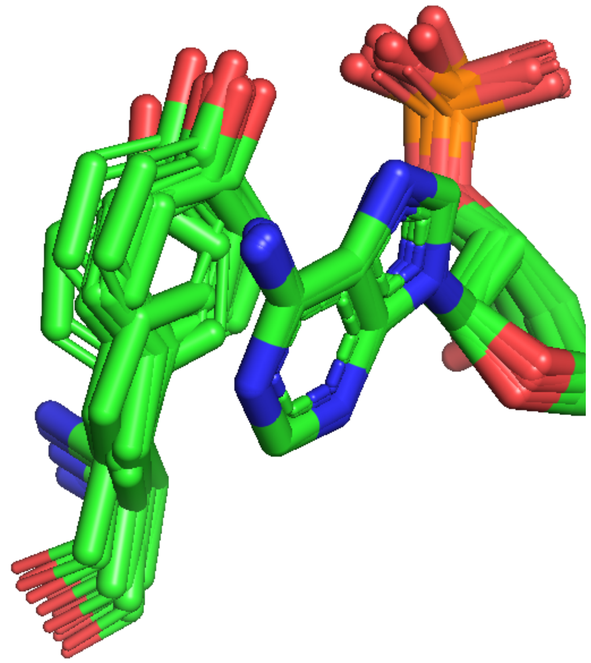


Figure 4. Pearson correlation plots of APSI vs. PPV (upper panel) and APSI vs. SPS (lower panel) for alignment of the RNAStralign benchmark dataset with RNAlign2D (left panel), CARNA (middle panel) and LocARNA (right panel). The correlation coefficients are shown at the top of each plot.

Mining biomacromolecular interactions with the BioShell package

Justyna Krys, Monika Pikuzinska, Dominik Gront
Faculty of Chemistry,University of Warsaw



INTRODUCTION

Molecular modelling is a technique commonly used in deciphering life on a molecular level. Outputs from these simulations usually comprise 3D structures and their energies, evaluated in a given force field. These tools however rarely explain how the biomolecules interact.

Thirteen close homologs from CYP family which bind NAP molecule were analysed as an example. Only interactions between ligand and a protein were taken into account.

STACKING INTERACTIONS

In the table there are all stacking interactions between NAP and a protein. It turns out that in all cases it is the same amino acid TYR604 (in 3QFC it is TYR 607). Geometry is presented on a figure. Thirteen proteins was superimposed to have NAP in the same place so the differences between homologs can be easily seen.

VAN DER WAALS INTERACTIONS

Multiple sequence alignment was created using MAFT program and coloured by Van der Waals interactions. Color scale was used to show the distances between interacting residues.

HYDROGEN BOND INTERACTIONS

Hydrogens was added to the structures with reduce[1] program. Residues that creates hydrogen bonds with NAP was colored orange on multiple sequence alignment.

PDB code	1st residue	2nd residue	r	angle	xy	z
1JA0	NAP852 B	Y604 B	3.751	166.402	1.519	3.430
1JA0	NAP752 A	Y604 A	3.698	167.154	0.997	3.561
1JA1	NAP1852 B	Y604 B	3.884	13.503	1.600	3.539
1JA1	NAP1752 A	Y604 A	3.715	7.488	1.235	3.503
3ES9	NAP753 A	Y604 A	4.269	55.074	1.139	4.114
3ES9	NAP753 B	Y604 B	4.196	32.883	0.896	4.100
3OJX	NAP753 A	Y604 A	3.930	12.828	1.702	3.542
1J9Z	NAP752 A	Y604 A	3.644	7.551	0.828	3.549
1J9Z	NAP852 B	Y604 B	3.711	6.020	1.296	3.477
6NJR	NAP703 A	Y604 A	3.751	162.636	0.884	3.645
6NJR	NAP703 B	Y604 B	3.961	164.295	1.076	3.812
3QFC	NAP753 A	Y607 A	3.865	10.405	1.604	3.516
3QFC	NAP753 B	Y607 B	3.725	12.987	1.099	3.559

Funding: This research was funded by the National Science Centre (Poland) Grant No. 2018/29/B/ST6/01989.

1. Word, et al.(1999) "Asparagine and glutamine: using hydrogen atom contacts in the choice of sidechain amide orientation" J. Mol. Biol. 285, 1735-1747.

Macnar, J.M.; Szulc, N.A.; Kryś, J.D.; Badaczewska-Dawid, A.E.; Gront, D. BioShell 3.0: Library for Processing Structural Biology Data. Biomolecules 2020, 10, 461.

grahics prepared with VisuaLife Package https://visualife.readthedocs.io

fingeRNAt - a novel tool for high-throughput analysis of nucleic acid - ligand interactions

Natalia A. Szulc[1], Zuzanna Mackiewicz[1], Janusz M. Bujnicki[1,2,*], and Filip Stefaniak[1,*]

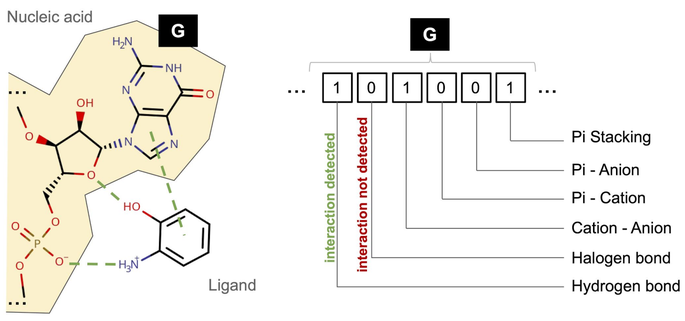
1. International Institute of Molecular and Cell Biology in Warsaw, Poland; 2. Adam Mickiewicz University, Poznań, Poland;
* - corresponding author's email: {iamb, fstefaniak}@genesilico.pl (J.M.B. and F.S.)

Abstract

Nucleic acids are becoming increasingly attractive targets for potential drugs. Since most targets of small molecule drugs are proteins, the portfolio of nucleic acids-oriented bioinformatics tools is limited. Here we present **fingeRNAt** - a novel and open-source software for **calculation of Structural Interactions Fingerprints (SIFs) for nucleic acid - ligand complexes**. SIFs translate information about 3D interactions in a target-ligand complex into a string, where the respective bit in the fingerprint is e.g. set to 1 in case of detecting particular interaction, and to 0 otherwise. By using SIFs, the interactions are represented in a unified fashion, thus allowing for easy analysis and comparison, as they provide a full picture of all interactions within the complex.

Background

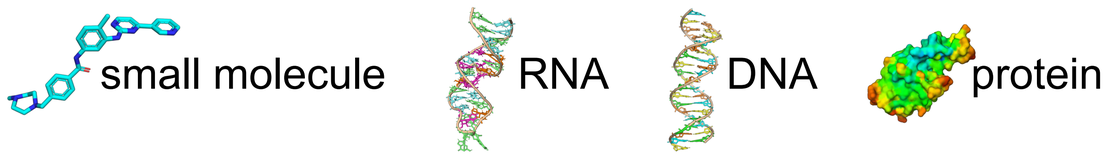
- Many nucleic acids are disease-associated with ability to adapt a tertiary structure hence constituting promising targets for drugs.
- Structural Interactions Fingerprints (SIFs) represent interactions within a complex in a form of a binary or a hologram string, a convenient input to computational analyses.



- No freely available tool to calculate SIFs for nucleic acid - ligand complexes.

Overview

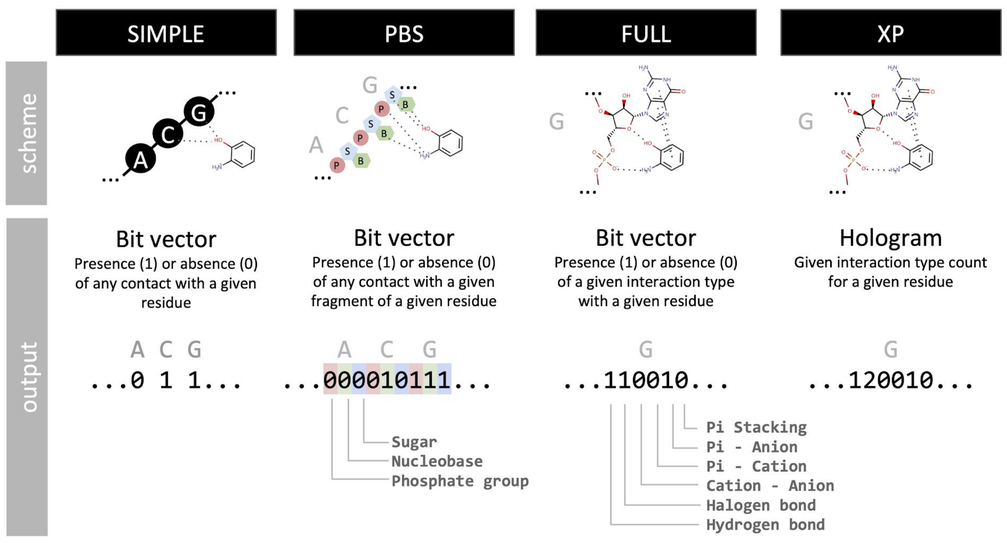
- fingeRNAt is a Python 3.x program which calculates SIFs in complexes of RNA/DNA and:



- **Input/Output**

  Requires (i) RNA/DNA structure in pdb/mol2 format and (ii) ligands' structures in sdf format.
  The output is a SIF calculated for each complex saved to separate row of a tab-separated file.

- **SIFs types**



fingeRNAt applications

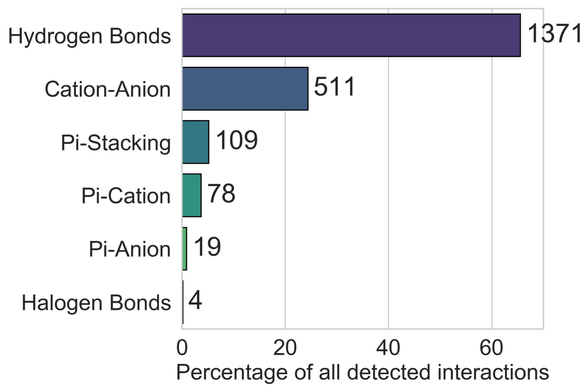
What are the non-covalent interactions statistics in RNA - ligands complexes?

Dataset

Non-redundant complexes of RNA with small molecule ligands.

Calculation of interactions

Non-covalent interactions in all the complexes from the dataset were detected and converted to **SIFs using fingeRNAt**. SIFs were used to calculate interactions statistics.



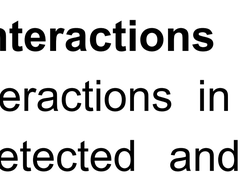
Hydrogen bonds are most frequent (over 65%), but ionic interactions play second most important role, constituting almost one quarter of all interactions.

fingeRNAt applications

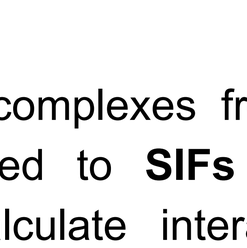
Can interaction patterns be used to discriminate between active and inactive compounds?

Dataset

Target: HIV TAR



Ligands:



Calculation of interactions

Docking was performed using rDock. Non-covalent interactions in all the complexes were detected and converted to **SIFs using fingeRNAt**. SIFs were used to calculate average number of contacts for each interaction.

residue	nucleotide	interaction type	active	inactive	difference	p-value
21	G	Pi-Cation	0.000	0.011	-0.011	0.00
21	G	Pi-Anion	0.000	0.004	-0.004	0.01
22	A	Halogen Bonds	0.000	0.003	-0.003	0.03
22	A	Pi-Anion	0.000	0.011	-0.011	0.00
23	U	Hydrogen Bonds	1.000	0.963	0.037	0.00
23	U	Halogen Bonds	0.000	0.009	-0.009	0.00
23	U	Cation-Anion	0.000	0.007	-0.007	0.00
23	U	Pi-Anion	0.000	0.003	-0.003	0.03
26	G	Hydrogen Bonds	1.000	0.952	0.048	0.00
26	G	Pi-Anion	0.000	0.012	-0.012	0.00
27	A	Halogen Bonds	0.000	0.009	-0.009	0.00
27	A	Pi-Anion	0.000	0.024	-0.024	0.00
39	C	Halogen Bonds	0.000	0.012	-0.012	0.00
39	C	Pi-Anion	0.000	0.004	-0.004	0.01
40	U	Pi-Anion	0.000	0.007	-0.007	0.00

Active and inactive ligands have different binding patterns and this variance may be utilized in rational drug design.

Code availability
github.com/n-szulc/fingeRNAt

References
Deng, Chuaqui, Singh, *J. Med. Chem.* **2004**, 47(2), 337−344.
Salentin et al., *Nucleic Acids Res.* **2015**, 43(W1), W443-W447.
O'Boyle et al., *J Cheminform.* **2011**, 3(1), 33.
Philips et al., *RNA.* **2013**, 19(12),1605-1616.
Ruiz-Carmona et al., *PLoS Comput Biol.* **2014**, 10(4), e1003571.

Acknowledgments
The „Modeling of dynamic interactions between RNA and small molecules and its practical applications" project is carried out within the TEAM programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund (Grant TEAM/2016-3/18 to J.M.B.).

European Funds Smart Growth
FNP Foundation for Polish Science
European Union European Regional Development Fund