VIRTUAL RNA BIOLOGY: FROM SEQUENCE TO HIGH QUALITY 3D STRUCTURE MODEL

[1]MARTA SZACHNIUK, [2]JACEK BLAZEWICZ

[1]Institute of Computing Science, Poznan University of Technology, Poland,
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poland
E-mail: mszachniuk@cs.put.poznan.pl, jblazewicz@cs.put.poznan.pl

Abstract— Contemporary biomedicine aims to use specialized molecules in medical diagnostics and treatment of diseases. This goal can be achieved if computer algorithms, being the heart of virtual biology, are able to model high quality structures with predefined functions. Existing computer methods can quite well predict simple three-dimensional molecular structures, but many of the predicted models do not yet meet the high requirements set for them by the scientists. Here, we focus on RNA molecules and we show how to get from RNA sequence to a high quality tertiary structure using existing bioinformatics tools. In particular, we present the data processing scheme that makes use of RNAComposer and RNAfitme - our own tools for RNA 3D structure prediction and refinement.

Index Terms- 3D structure modeling, computational prediction, model refinement, RNA bioinformatics

I. INTRODUCTION

The last two decades have brought intensive development of life sciences. New technologies and computer algorithms dedicated to issues that originate in biology, chemistry, and biophysics have largely contributed to this phenomenon. The latter allowed for the creation of virtual biology, in which biological research takes place exclusively in computer space. Virtual biology is primarily based on the use of bioinformatics methods to conduct scientific experiments (e.g. simulation of intracellular processes) and to analyze the results of these experiments [7]. One of the most important topics in this field of science is structure prediction of biomolecules: proteins, nucleic acids, and macromolecular complexes.

For many years, proteins have been the main focus of interest of structural biologists and bioinformaticians. Many research groups developed methods of protein structure prediction, the quality and effectiveness of which was then evaluated in the CASP (Critical Assessment of protein Structure Prediction) competition [6]. In recent years, RNA (ribonucleic acid) molecules have also been included in the interest of scientists. Several methods for automated prediction of tertiary RNA structure have been proposed. As in the case of methods dedicated to proteins, they were developed in one of the following approaches: de novo modelling, homologous modelling, or hybrid methods. At the same time, new algorithms started to appear to support the comparison of predicted 3D structures and their verification and quality assessment [5]. As a natural consequence of this research, the RNA-Puzzles competition, similar to CASP, was organized to allow for a blind assessment of automatic predictions of the RNA structure [4]. Subsequent challenges of RNA-Puzzles showed the strengths and weaknesses of modern RNA 3D structure modeling algorithms. It turned out that many RNA models generated automatically have poor quality and accuracy. Often a relatively small correction could significantly improve the model and increase its chance in the competition. However, few existing methods can effectively improve the quality and accuracy of the envisaged RNA three-dimensional structure.

In this paper we show how to get from RNA sequence to a high quality tertiary structure using automated bioinformatics tools. We focus on the use of own webserver tools. We recall how RNAComposer, our RNA 3D structure prediction system, generates the prototype 3D model of RNA [9]. The system itself is equipped with mechanisms that allow the user to correct the raw structure and improve its quality [1]. In addition, we show how to improve the quality of the model containing incomplete data, incorrect angles or clashed atoms. The latter tasks can be successfully accomplished with RNAfitme, our recently released bioinformatics tool [2].

II. PROTOTYPING AND REFINING OF RNA 3D STRUCTURE

In virtual biology, the way from RNA sequence to the 3D structure usually boils down to a one-time action of launching a program for structure prediction. However, with existing bioinformatics tools, the computer modelling of a high quality RNA structure should be a multi-stage process. The first stage, in which the user applies one of the commonly available tools to predict the 3D RNA structure, should be followed by a quality assessment and model refinement (Fig. 1). Thus, the first obtained 3D model should be treated as a prototype that can be improved in subsequent steps, often processed iteratively.

Prototyping of RNA 3D structure can be done by one of the existing tools dedicated to RNA 3D structure prediction [5]. Here, we propose an application of RNAComposer [1][9]. This system, by default running in the interactive mode, allows the user to model the RNA 3D structure immediately and does not require any information from the user except for the sequence (optionally also the secondary structure). The obtained prototype model can be next automatically validated using e.g. MolProbity [3]. This tool has been developed to check the quality of crystallographic structures, but it can be successfully applied to evaluate models derived from computer simulation as well. MolProbity reports, among other things, on the bad angles, bad bonds, and clashes found in the analyzed structure [2].

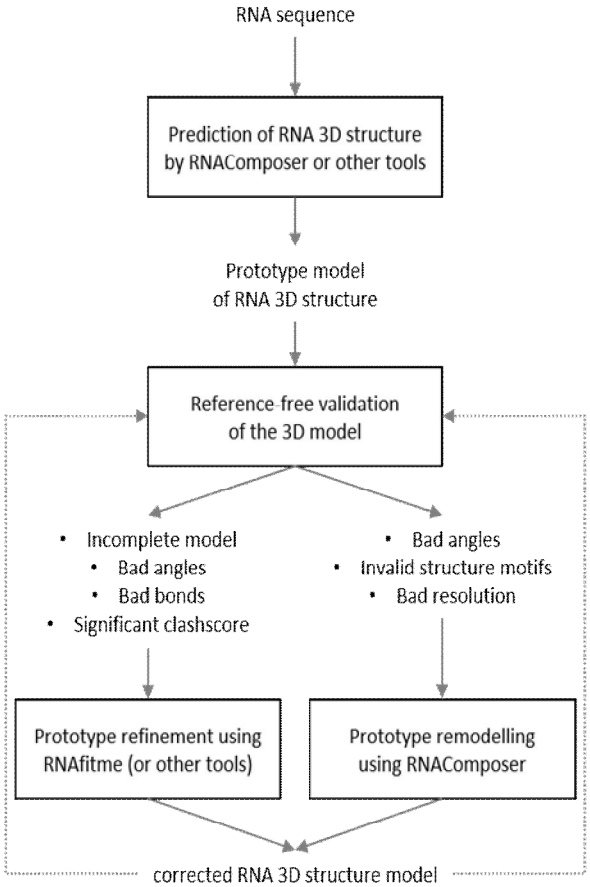


Figure 1. From sequence to high-quality RNA 3D structure.

A visual inspection of the predicted model can provide other types of problems. These can be, for example, missing residues or single atoms, invalid structure motifs, or poor resolution of the model.

The stage of prototype model refinement can be carried out using different tools, depending on the type of problems observed. The errors and irregularities reported by MolProbity [3] can be corrected using RNAfitme [2] or NAMD [8]. Among these two tools, RNAfitme can also complete the model if it contains missing nucleotide residues, bases (adenine, guanine, cytosine or uracyl) or atoms lying outside the RNA backbone. RNAfitme uses a library of conformers representing stereochemically correct structures of nucleotide residues and nucleobases [10]. The description of each conformer contains, among others, a list of atom coordinates, torsional angles and a distance metric that allows its optimum fitting and placement in the analyzed RNA structure. During RNA remodeling, RNAfitme optimizes its structure for packing in space and minimizes the number of clashes between atoms. Moreover, while adjusting the remodeled structure fragments, it takes into account their mutual position and tries to arrange them in such a way as to preserve the typical conformation of torsional angles. This leads to reduction of incorrect angle values and bad bonds in the prototype model.

Other problems, like incorrect motifs, or bad structure resolution can be corrected using RNAComposer run in the batch mode, which is available after logging to the system. Invalid motifs can be replaced in the prototype model if the user is able to provide the corresponding motif with the required structure characteristics. This can be done when an option of inserting own structure elements is chosen. Structure resolution can be improved automatically when the user filters the structure elements repository and requires to select only high-resolution candidates [1].

The two schemes of prototype model improvement can be processed independently. Although, if many different types of problems are identified in the initial model, it is strongly recommended to run RNAComposer-based remodeling first, and then proceed with RNAfitme if necessary. We believe, such iterative refinement significantly improves the predicted RNA 3D model in most cases and leads to high-quality RNA structures.

REFERENCES

[1] M. Antczak, M. Popenda, T. Zok, J. Sarzynska, T. Ratajczak, K. Tomczyk, et al., „New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure", Acta Biochimica Polonica, vol. 63, pp. 737-744, 2016.

[2] M. Antczak, T. Zok, M. Osowiecki, M. Popenda, R.W. Adamiak, M. Szachniuk, „RNAfitme: a webserver for modeling nucleobase and nucleoside residue conformation in fixed-backbone RNA structures", BMC Bioinformatics, vol. 19, pp. 304, 2018.

[3] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, "MolProbity: all-atom structure validation for macromolecular crystallography", Acta Crystallogr DBiolCrystallogr, vol. 66, pp. 12-21, 2010.

[4] Z. Miao, R.W. Adamiak, M. Antczak, R.T. Batey, A.J. Becka, M. Biesiada, et al., "RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme", RNA, vol. 23, pp. 655-672, 2017.

[5] Z. Miao, E. Westhof, "RNA Structure: Advances and Assessment of 3D Structure Prediction", Annual Review of Biophysics, vol. 46, pp. 483-503, 2017.

[6] J. Moult, K. Fidelis, B. Rost, T. Hubbard, A. Tramontano, "Critical assessment of methods of protein structure prediction (CASP) -round 6", Proteins, vol. 61, pp. 3-7, 2005.

[7] M. Muhamad, H. B. Zaman, A. Ahmad „Virtual Biology Laboratory (VLab-Bio): Scenario-Based Learning Approach", Procedia - Social and Behavioral Sciences, vol. 69, pp. 162-168, 2012.

[8] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, "Scalable molecular dynamics with NAMD", J Comp Chem. Vol. 26, pp. 1781-802, 2005.

[9] M. Popenda, M. Szachniuk, M. Antczak, K.J. Purzycka, P. Lukasiak, N. Bartol, J. Blazewicz, R.W. Adamiak, „Automated 3D structure composition for large RNAs", Nucleic Acids Research, vol. 40, pp. e112, 2012.

[10] T. Zok, M. Antczak, M. Riedel, D. Nebel, T. Villmann, P. Lukasiak, J. Blazewicz, M. Szachniuk, "Building the library of RNA 3D nucleotide conformations using clustering approach", International Journal of Applied Mathematics and Computer Science, vol. 25, pp. 689-700, 2015.

★★★