Bioinformatics, 2016, 1–3
doi: 10.1093/bioinformatics/btw502
Advance Access Publication Date: 6 August 2016
Applications note

OXFORD

Structural bioinformatics

WebSTAR3D: a web server for RNA 3D structural alignment

Erwin Holzhauser, Ping Ge and Shaojie Zhang*

Department of Computer Science, University of Central Florida, Orlando, FL 32816, USA

*To whom correspondence should be addressed.

Associate Editor: Ivo Hofacker

Received on January 25, 2016; revised on June 17, 2016; accepted on July 26, 2016

Abstract

Summary: The WebSTAR3D web server is a user-friendly online interface for the alignment of RNA 3D structures. The website takes as input two files, each of which can be in either PDB or mmCIF format, containing the desired structures to align, via a PDB code or user upload. In return, the user is presented with a visualization of the aligned structures in Jmol or JSmol, along with the corresponding sequence alignment, and the option to download the nucleotide mapping of the structures and a PDB file containing the aligned, superimposed structures.
Availability and Implementation: The WebSTAR3D is available at http://rna.ucf.edu/WebSTAR3D.
Contact: shzhang@cs.ucf.edu

1 Introduction

It is recently discovered that noncoding RNAs (ncRNAs) are involved in nearly every cellular process, including gene expression and other aspects of cellular metabolism (Hannon *et al.*, 2006). As ncRNAs are genetic end products, their functions are implicated primarily by their secondary and tertiary structures. This has motivated an increasing research interest in methods for the comparative analysis of RNA structures, as key evolutionary and functional relationships between RNA molecules can be inferred from their structural alignments.

With the rapidly growing archive of RNA 3D structures in the Protein Data Bank (PDB), numerous tools have been developed for the alignment of their tertiary structures, such as SARA (Capriotti and Marti-Renom, 2009), R3DAlign (Rahrig *et al.*, 2010) and LaJolla (Bauer *et al.*, 2009). Typically, these tools utilize the comparison of base pairing interactions between molecules (e.g. SARA, R3DAlign), or they encode key structural features of these molecules, such as RNA backbones and nucleotide torsion angles, as sequential information that can be further processed (e.g. LaJolla).

STAR3D implements an alternative approach that identifies the consensus of stacks between two RNA molecules, and uses this information to constrain the alignment of loop regions (Ge and Zhang, 2015). First, sub-stacks with similar 3D structures are identified and assembled into conserved stack pairs, from which a sparse compatible graph is generated based on the secondary structure relations and spatial distances. Then, a tree-like consensus structure of the two molecules is generated by efficiently finding the maximal clique in the compatible graph. Finally, loop regions between the two molecules are ordered based on the consensus structure such that a loop region from one RNA structure only needs to be compared to its matching partner from the other structure. The stack and loop alignments are then combined into the final result. This approach results in efficient RNA 3D structural alignment that matches the precision of other state-of-the-art tools. Here, we present the WebSTAR3D web server: a web interface for the alignment of RNA 3D structures that provides the core functionality of STAR3D, along with some additional features.

2 Methods

2.1 Implementation

The front-end of the webpage is implemented in HTML, Javascript and JQuery (https://jquery.com). The alignment is computed by the Java implementation of the STAR3D algorithm available at http://genome.ucf.edu/STAR3D, modified to additionally accept mmCIF files as input, accept user uploaded PDB/mmCIF files, and return the sequence alignment corresponding to the structural alignment. Communication between the program and the webpage (i.e. passing input from the front-end to the Java implementation and returning the resulting alignment back to the front-end) is handled with PHP.

The docking of aligned structures is visualized in Jmol or JSmol (http://www.jmol.org), based on user preference.

Before aligning structures, STAR3D preprocesses PDB files with base-pairing annotation using either MC-Annotate (Gendron *et al.*, 2001; Lemieux and Major, 2002) (for PDB inputs) or DSSR (Lu *et al.*, 2015) (for PDB and mmCIF inputs) and pseudo-knot removal using RemovePseudoknots (Smit *et al.*, 2008). Preprocessing is the most time-consuming portion of the entire STAR3D procedure. Also, the result of preprocessing for a particular PDB-chain pair can be reused for any alignment containing that pair. Thus, to improve runtime, the preprocessing for all parsable RNA PDB files (representative and class members) used in the v1.89 of non-redundant list of RNA structures under a 4.0 Å resolution cutoff (Leontis and Zirbel, 2012) has been stored on the hosting server. The cache of preprocessed PDB files will periodically be updated with new releases.

2.2 Input

Upon visiting the webpage, the user is presented with a form with which to input a pair of PDB files encoding the desired RNA 3D structures to align. For each structure, the user must provide a valid PDB identification code or upload their own properly formatted PDB or mmCIF file, along with a valid chain identifier for the molecule. The user must also specify the format (PDB or mmCIF) of the inputs, along with a desired base-pair annotation method to be used in the pipeline. When entering PDB codes into the form, the form offers suggestions for the codes corresponding to the preprocessed PDBs, along with suggestions for their available chain identifiers. Finally, there is a drop-down form to modify the parameters of the STAR3D algorithm. These include the dynamic programming parameters for the nucleotide alignment of loop regions: gap open cost, gap extension cost and match/mismatch scores, along with the minimum stack size to consider when searching for similar stacks between the two structures, and the root-mean-square distance cutoff used for both stack and loop region alignments. The default parameter values have been empirically chosen. The STAR3D algorithm, and how the aforementioned parameters are used within it, are described in more detail in Ge and Zhang (2015).

2.3 Output

The user is presented with the sequence alignment corresponding to the tertiary structure alignment, the secondary structure for each sequence in dot-bracket notation, a description of each structure and alignment metrics. As Jmol/JSmol rendering can be time-consuming, the user is presented with a visualization of the 3D docking of aligned regions in Jmol or JSmol only after clicking a 'Render Docking' button, giving the user faster access to the downloads. For the structure alignment and sequence alignment, the first and second structures are highlighted in red and blue, respectively, with aligned regions appearing in a darker hue than the unaligned regions. The user can download the alignment file generated by STAR3D, which includes the mapping between the nucleotides of the two structures, a PDB file containing the atomic coordinates of the superimposed structures, as well as the sequence alignment. Additionally, the user has the option to view just a portion of the aligned structures by specifying a continuous interval of interest for each structure, using slider widgets at bottom of the page. An example of the resulting page for the alignment of *E.coli* initiator tRNA (PDB: 3cw5, chain A) and mouse tRNA(Sec) (PDB: 3rg5, chain A) is shown in Figure 1.

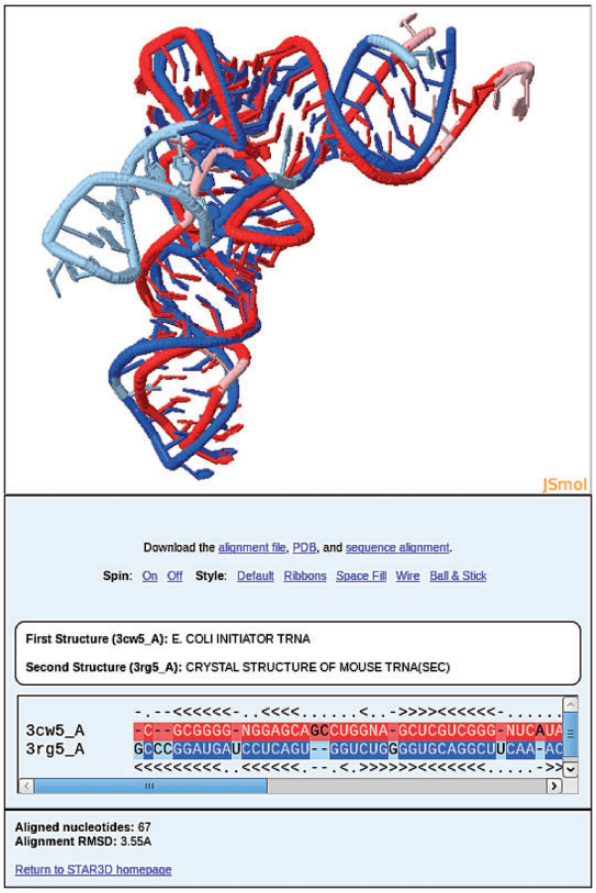

Fig. 1. WebSTAR3D alignment result of *E. coli* initiator tRNA (PDB: 3cw5, chain A) and mouse tRNA (PDB: 3rg5, chain A). The topmost container shows the Jmol docking of the aligned structures; the first and second structures are colored red and blue, respectively, with aligned regions colored in a darker hue. Directly below are the options to download the nucleotide mapping, the PDB of the docked structures and the sequence alignment. Then follows the secondary structure alignment corresponding to the 3D structural alignment

2.4 WebSTAR3D advantages

The primary advantages of WebSTAR3D stem from the algorithmic approach of STAR3D. Because STAR3D avoids computationally expensive secondary structure comparisons, WebSTAR3D can perform very fast 3D structural alignment, even for RNAs that are prohibitively large and/or non-homologous for other platforms, without sacrificing prediction accuracy. Where many structural alignment web servers redirect the user to an intermediary loading page that notifies the user of the status of their query, where the user may wait several minutes, STAR3D directly presents the user with the results page. On average, the user needs only wait several seconds—in fact, the non-homologous alignment of the sarcin-ricin motif (PDB: 483d, chain A) to the large 23S rRNA (PDB: 1qvg, chain 0) only takes 7 seconds to compute and then render with Jmol.

3 Conclusion

As ncRNAs are involved in numerous cellular processes, and their functions are implicated by their structures, there is a growing interest in tools for the comparative analysis of RNA structures. For this reason, we have presented a user-friendly web server tool for the tertiary structure alignment of RNA molecules. Because the underlying STAR3D algorithm takes a fundamentally different approach to

RNA 3D structural alignment than existing tools, WebSTAR3D performs faster on alignments of comparable, or better, quality compared to state-of-the-art tools.

Funding

This work has been supported by the US National Institutes of Health (R01GM102515).

Conflict of Interest: none declared.

References

Bauer,R.A. *et al*. (2009) Fast structural alignment of biomolecules using a hash table, n-grams and string descriptors. *Algorithms*, **2**, 692–709.

Capriotti,E. and Marti-Renom,M.A. (2009) SARA: a server for function annotation of RNA structures. *Nucleic Acids Res*., **37**, W260–W265.

Ge,P. and Zhang,S. (2015) STAR3D: a stack-based RNA 3D structural alignment tool. *Nucleic Acids Res*., **43**, e137.

Gendron,P. *et al*. (2001) Quantitative analysis of nucleic acid three-dimensional structures. *J. Mol. Biol*., **308**, 919–936.

Hannon,G.J. *et al*. (2006) The expanding universe of noncoding RNAs. *Cold Spring Harbor Symp. Quant. Biol*., **71**, 551–564.

Lemieux,S. and Major,F. (2002) RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res*., **30**, 4250–4263.

Leontis,N.B. and Zirbel,C.L. (2012) Nonredundant 3D structure datasets for RNA knowledge extraction and benchmarking. In: Neocles,L and Eric,W. (eds) *RNA 3D Structure Analysis and Prediction*, vol. **27**. Springer, Berlin, Heidelberg, pp. 281–298.

Lu,X.J. *et al*. (2015) DSSR: an integrated software tool for dissecting the spatial structure of RNA, *Nucleic Acids Res*., **43**, e142.

Rahrig,R.R. *et al*. (2010) R3D Align: global pairwise alignment of RNA 3D structures using local superpositions. *Bioinformatics*, **26**, 2689–2697.

Smit,S. *et al*. (2008) From knotted to nested RNA structures: a variety of computational methods for pseudoknot removal. *RNA*, **14**, 410–416.