

RASS: a web server for RNA alignment in the joint sequence-structure space

Gewen He¹, Albert Steppi², Jose Laborde², Anuj Srivastava², Peixiang Zhao¹ and Jinfeng Zhang^{2,*}

¹Department of Computer Science, Florida State University, Tallahassee, FL 32306, USA and ²Department of Statistics, Florida State University, Tallahassee, FL 32306, USA

Received January 31, 2014; Revised April 24, 2014; Accepted May 3, 2014

ABSTRACT

Comparison of ribonucleic acid (RNA) molecules is important for revealing their evolutionary relationships, predicting their functions and predicting their structures. Many methods have been developed for comparing RNAs using either sequence or three-dimensional (3D) structure (backbone geometry) information. Sequences and 3D structures contain non-overlapping sets of information that both determine RNA functions. When comparing RNA 3D structures, both types of information need to be taken into account. However, few methods compare RNA structures using both sequence and 3D structure information. Recently, we have developed a new method based on elastic shape analysis (ESA) that compares RNA molecules by combining both sequence and 3D structure information. ESA treats RNA structures as 3D curves with sequence information encoded on additional coordinates so that the alignment can be performed in the joint sequence-structure space. The similarity between two RNA molecules is quantified by a formal distance, geodesic distance. In this study, we implement a web server for the method, called RASS, to make it publicly available to research community. The web server is located at <http://cloud.stat.fsu.edu/RASS/>.

INTRODUCTION

Comparison of ribonucleic acid (RNA) structures is an effective tool for studying the functions of RNA molecules and their evolutionary relationships. There have been a number of methods developed for RNA structure alignment/comparison (1–15). Databases for RNA structures have also been built to facilitate easy retrieval of such data (14,16,17). There are mainly three types of structure comparison methods. The first type of methods focuses

on detection of local structural motifs to identify functional domains, such as NASSAM (9), COMPADRES (8), RNAMotifScan (11) and FR3D (7). The second type of methods reduces the three-dimensional (3D) structures to one-dimensional (1D) sequences by representing nucleotide residues with some local structure features. Existing sequence alignment methods can then be applied to align the resulting 1D sequences. Among this type of methods, iPARTS discretized backbone torsion angles to form structural alphabet with 23 letters (states). A substitution matrix is then derived for the 23 states and used in the sequence alignment (2). SARA uses a set of unit vectors derived from consecutive nucleotides to represent each nucleotide, which can be compared with other nucleotide using unit-vector root mean square (URMS) as distance (1,18). LaJolla uses an n-gram model to analyze sequences derived from nucleotide torsion angles (4). Similarly, PRIMOS/AMIGOS (5) and DIAL (6) also represent nucleotides with torsion angles and align the sequences encoded by the torsion angle representation. These methods do not necessarily produce globally similar alignment between two RNA structures (i.e. with small RMSDs (root-mean-square-deviations)). To minimize RMSD for the aligned parts between two RNA structures, extra steps are required after the sequence alignment. The third type of methods starts from aligning similar local structures and then obtains larger scale alignment by extending the initial local alignment. For example, R3D Align employs a maximum clique algorithm on a specially defined graph, called a local alignment graph, to merge local alignments to form a global alignment (10); ARTS uses P (phosphor) atoms of two consecutive base pairs as seeds to first find structurally similar seed quadrants and then aligns overall structures based on the alignment of the seed quadrants (3); SETTER decomposes RNA structures to larger local structure units, called generalized secondary structure units (GSSUs), and uses a pairwise comparison method based on 3D similarity of the GSSUs (15). Several web servers for RNA structure alignment have been developed implementing some of the above methods, including ARTS

*To whom correspondence should be addressed. Tel: +1 850 644 3218; Fax: +1 850 644 5271; Email: Jinfeng@stat.fsu.edu
Present address: Jose Laborde, Department of Mathematics, Florida Atlantic University, Boca Raton, FL 33431, USA.

(3,19), SARA (1,18), SETTER (15,20), RADAR (21) and SARSA (22).

Sequence alignment and structure alignment use different sets of information: the former uses information of side chains, which are reduced to single letters, and the latter uses information of the backbone geometry. Both types of information play important roles in determining RNA functions. In principle, structure alignment methods should utilize sequence information since such information is almost always available. Recently, we have developed a RNA structure alignment method (23,24) that aligns RNAs in the joint sequence-structure space using a framework based on elastic shape analysis (ESA) originally designed for protein structure comparison (25,26). We have shown that the method performed better than using either sequence or structure information alone. The method also performed better than previous methods in RNA function prediction when tested on a benchmark dataset. In this study, we develop a web server for joint sequence-structure comparison of RNA structures using ESA, called RASS (RNA alignment in the joint sequence-structure space). In the next section, we first briefly describe the methodology of RASS, which is followed by a detailed description of the usage of the web server.

MATERIALS AND METHODS

We use ESA to compare two RNA structures or structural fragments. ESA treats RNA structures as 3D curves and utilizes a geometric framework that has been developed originally in image analysis and computer vision for shape analysis of parameterized curves and surfaces (27–30). The basic idea in this framework is to design an infinite-dimensional topological space (manifold) of curves, endow it with a metric structure, and compare any two objects by computing the distances between them on this manifold. Under this framework, we will be able to quantify the similarity between any two RNA structures by a formal distance, geodesic distance, computed on their respective shape manifolds. A geodesic path, the shortest path connecting the two curves (two points in the manifold), can also be generated. It can be seen as an optimal deformation from one structure into the other. The framework allows us to seamlessly incorporate sequence information into 3D structures in the computation of geodesic distances. The detailed description of the ESA method is given in (23) and (25).

The implementation is done using Matlab (with some embedded C functions). Due to Matlab initialization, RNA structure comparison at the web server takes more time than running the programs locally.

WEB INTERFACE AND USAGE

Input

The user interface of RASS is shown in Figure 1. The input of the server takes two RNA chains (structure A and B). Users can provide a Protein Data Bank (PDB) code or upload a PDB file for each RNA structure. A chain name should also be provided for each RNA structure.

Parameter selection

The parameters are preset at default values. Users have the flexibility to specify values for the following:

- Starting (for each structure A and B): the index of the starting residue from the RNA chains. The default value is 1.
- Ending (for each structure A and B): the index of the ending residue from the RNA chains. The default value is 10 000. That means for RNA chains with more than 10 000 residues, without specifying the starting and ending positions of the residues, the first 10 000 residues will be used in comparison. If the length of the RNA chain is less than 10 000, then the actual chain end position will be used. This value is large enough for the RNA structures currently in PDB. Being able to select starting and ending positions from the RNA chains to be compared is a feature most of the existing web servers do not provide. With this option users can select a partial structure within a larger RNA molecule for comparison, which can be interesting either structurally or functionally. Since our method is a global alignment method, this option can be especially useful if a user wants to compare a partial structure based on some prior knowledge/information.
- Lambda: weight for the nucleotide sequence (23). The value of lambda needs to be greater or equal to zero. When zero is specified, the resulting geodesic distance takes into account only backbone geometry information but not the nucleotide sequences. When a large number (e.g. 70) is specified, then the distance computation is dominated mainly by nucleotide sequences. Recommended values for lambda is between 0 and 10. The optimal value for lambda obtained by cross-validation on a benchmark data is 5, which is set as the default value.

Output

The output is displayed in a series of drop-down tabs (see Figures 2–4):

- Distance and *P*-value: geodesic distance calculated by ESA between two RNA chains is displayed under the Distance and *P*-value tab. *P*-value is obtained by comparing the geodesic distance with an empirical Gaussian distribution derived from a set of pairwise distances computed using a large number of non-homologous RNA structures taken from PDB (31). A discussion on the distribution of geodesic distances is given in Supplementary material. A small *P*-value indicates that the chains are related/similar statistically in the joint sequence-structure space.
- Geodesic path: this tab displays the optimal structural deformation from RNA molecule A to RNA molecule B viewed from three different angles (Figure 2). These views are slightly distorted by the sequence weight. To view an undistorted version of the geodesic path, one can set lambda = 0 (i.e. use structure only for the comparison).
- Sequence alignment: this tab shows the global sequence alignment between two RNA molecules (Figure 3). The

Home Publications Help People Downloads

RASS server: RNA alignment in the joint sequence-structure space

RNA molecule A

PDB code or upload PDB file No file chosen

Chain ID

RNA molecule B

PDB code or upload PDB file No file chosen

Chain ID


[Parameters](#) 

Figure 1. The user interface of RASS.

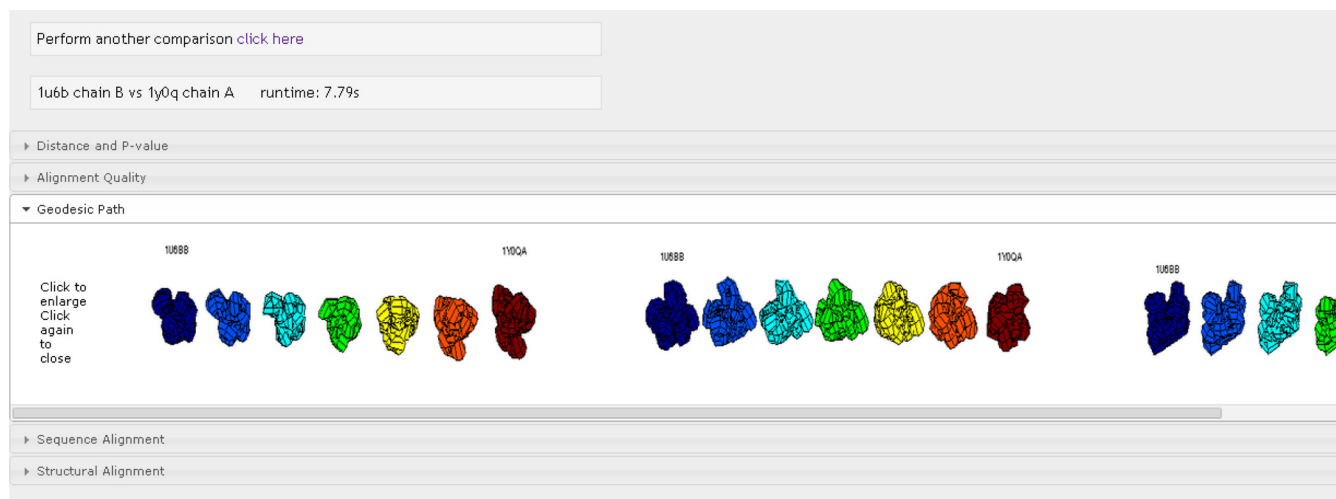


Figure 2. Geodesic path for chain B of RNA 1u6b and chain A of RNA 1y0q.

residue indices for each molecule are shown above or below the corresponding sequence for every 10 residues.

- **Structural Alignment:** the 3D optimal superposition is displayed with Jmol (Figure 4). Users can download the alignment files in PDB format through the link provided in this tab. Some users may not be able to see the graphical display for the first time using the web server. A link to Jmol tutorial page is also provided.

CONCLUSION

In this study, a web server is implemented to provide a tool for users to compare and align two RNA molecules in the joint sequence-structure space. A typical scenario for a user when using our server is as follows: The user will provide two RNA molecules as input where one may be a RNA molecule with known function and the other is a

ACKNOWLEDGMENT

We thank anonymous reviewers for their constructive comments and suggestions.

FUNDING

National Institutes of Health [1R21GM101552 to J.Z. and A.S.]. Funding for open access charge: National Institutes of Health [1R21GM101552].

Conflict of interest statement. None declared.

REFERENCES

1. Capriotti,E. and Marti-Renom,M.A. (2008) RNA structure alignment by a unit-vector approach. *Bioinformatics*, **24**, i112–i118.
2. Wang,C.W., Chen,K.T. and Lu,C.L. (2010) iPARTS: an improved tool of pairwise alignment of RNA tertiary structures. *Nucleic Acids Res.*, **38**, W340–W347.
3. Dror,O., Nussinov,R. and Wolfson,H. (2005) ARTS: alignment of RNA tertiary structures. *Bioinformatics*, **21**(Suppl. 2), ii47–ii53.
4. Bauer,R.A., Rother,K., Moor,P., Reinert,K., Steinke,T., Bujnicki,J.M. and Preissner,R. (2009) Fast structural alignment of biomolecules using a hash table, n-Grams and string descriptors. *Algorithms*, **2**, 692–709.
5. Duarte,C.M., Wadley,L.M. and Pyle,A.M. (2003) RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Res.*, **31**, 4755–4761.
6. Ferrè,F., Ponty,Y., Lorenz,W.A. and Clote,P. (2007) DIAL: a web server for the pairwise alignment of two RNA three-dimensional structures using nucleotide, dihedral angle and base-pairing similarities. *Nucleic Acids Res.*, **35**, W659–W668.
7. Sarver,M., Zirbel,C.L., Stombaugh,J., Mokdad,A. and Leontis,N.B. (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.*, **56**, 215–252.
8. Wadley,L.M. and Pyle,A.M. (2004) The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery. *Nucleic Acids Res.*, **32**, 6650–6659.
9. Harrison,A.M., South,D.R., Willett,P. and Artymiuk,P.J. (2003) Representation, searching and discovery of patterns of bases in complex RNA structures. *J. Comput. Aided Mol. Des.*, **17**, 537–549.
10. Rahrig,R.R., Leontis,N.B. and Zirbel,C.L. (2010) R3D Align: global pairwise alignment of RNA 3D structures using local superpositions. *Bioinformatics*, **26**, 2689–2697.
11. Zhong,C., Tang,H. and Zhang,S. (2010) RNAMotifScan: automatic identification of RNA structural motifs using secondary structural alignment. *Nucleic Acids Res.*, **38**, e176.
12. Laing,C. and Schlick,T. (2010) Computational approaches to 3D modeling of RNA. *J. Phys. Cond. Matter*, **22**, 283101.
13. Laing,C. and Schlick,T. (2011) Computational approaches to RNA structure prediction, analysis, and design. *Curr. Opin. Struct. Biol.*, **21**, 306–318.
14. Abraham,M., Dror,O., Nussinov,R. and Wolfson,H.J. (2008) Analysis and classification of RNA tertiary structures. *RNA*, **14**, 2274–2289.
15. Hoksza,D. and Svozil,D. (2012) Efficient RNA pairwise structure comparison by SETTER method. *Bioinformatics*, **28**, 1858–1864.
16. Murthy,V.L. and Rose,G.D. (2003) RNABase: an annotated database of RNA structures. *Nucleic Acids Res.*, **31**, 502–504.
17. Tamura,M., Hendrix,D.K., Klosterman,P.S., Schimmelman,N.R., Brenner,S.E. and Holbrook,S.R. (2004) SCOR: structural classification of RNA, version 2.0. *Nucleic Acids Res.*, **32**, D182–D184.
18. Capriotti,E. and Marti-Renom,M.A. (2009) SARA: a server for function annotation of RNA structures. *Nucleic Acids Res.*, **37**, W260–W265.
19. Dror,O., Nussinov,R. and Wolfson,H.J. (2006) The ARTS web server for aligning RNA tertiary structures. *Nucleic Acids Res.*, **34**, W412–W415.
20. Cech,P., Svozil,D. and Hoksza,D. (2012) SETTER: web server for RNA structure comparison. *Nucleic Acids Res.*, **40**, W42–W48.
21. Khaladkar,M., Bellofatto,V., Wang,J.T., Tian,B. and Shapiro,B.A. (2007) RADAR: a web server for RNA data analysis and research. *Nucleic Acids Res.*, **35**, W300–W304.
22. Chang,Y.F., Huang,Y.L. and Lu,C.L. (2008) SARSA: a web tool for structural alignment of RNA using a structural alphabet. *Nucleic Acids Res.*, **36**, W19–W24.
23. Laborde,J., Robinson,D., Srivastava,A., Klassen,E. and Zhang,J. (2013) RNA global alignment in the joint sequence-structure space using elastic shape analysis. *Nucleic Acids Res.*, **41**, e114.
24. Laborde,J., Srivastava,A. and Zhang,J. (2011) Structure-based RNA Function Prediction using Elastic Shape Analysis. *IEEE Int. Conf. Bioinformatics Biomed.*, 16–21.
25. Liu,W., Srivastava,A. and Zhang,J. (2011) A mathematical framework for protein structure comparison. *PLoS Comput. Biol.*, **7**, e1001075.
26. Liu,W., Srivastava,A. and Zhang,J. (2010) Protein structure alignment using elastic shape analysis. *Proc. ACM-BCB*, 62–70.
27. Srivastava,A., Klassen,E., Joshi,S. H. and Jermyn,I. H.. (2011) Shape analysis of elastic curves in euclidean spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, **33**, 1415–1428.
28. Joshi,S.H., Klassen,E., Srivastava,A. and Jermyn,I.H. (2007) A novel representation for Riemannian analysis of elastic curves in R^n . *IEEE Conf. Comput. Vis. Pattern Recogn.*, **7**, 1–7.
29. Mio,W., Srivastava,A. and Joshi,S. (2007) On shape of plane elastic curves. *Int. J. Comput. Vis.*, **73**, 307–324.
30. Klassen,E., Srivastava,A., Mio,W. and Joshi,S.H. (2004) Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, **26**, 372–383.
31. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.