Published online 23 May 2008 Nucleic Acids Research, 2008, Vol. 36, Web Server issue W19–W24
doi:10.1093/nar/gkn327

SARSA: a web tool for structural alignment of RNA using a structural alphabet

Yen-Fu Chang[1], Yen-Lin Huang[2] and Chin Lung Lu[1,3,*]

[1]Institute of Bioinformatics, National Chioa Tung University, [2]Department of Computer Science, National Tsing Hua University and [3]Department of Biological Science and Technology, National Chioa Tung University, Hsinchu 300, Taiwan

Received February 19, 2008; Revised April 30, 2008; Accepted May 7, 2008

ABSTRACT

SARSA is a web tool that can be used to align two or more RNA tertiary structures. The basic idea behind SARSA is that we use the vector quantization approach to derive a structural alphabet (SA) of 23 nucleotide conformations, via which we transform RNA 3D structures into 1D sequences of SA letters and then utilize classical sequence alignment methods to compare these 1D SA-encoded sequences and determine their structural similarities. In SARSA, we provide two RNA structural alignment tools, PARTS for pairwise alignment of RNA tertiary structures and MARTS for multiple alignment of RNA tertiary structures. Particularly in PARTS, we have implemented four kinds of pairwise alignments for a variety of practical applications: (i) global alignment for comparing whole structural similarity, (ii) semiglobal alignment for detecting structural motifs, (iii) local alignment for finding locally similar substructures and (iv) normalized local alignment for eliminating the mosaic effect of local alignment. Both tools in SARSA take as input RNA 3D structures in the PDB format and in their outputs provide graphical display that allows the user to visually view, rotate and enlarge the superposition of aligned RNA molecules. SARSA is available online at http://bioalgorithm.life.nctu.edu.tw/SARSA/.

INTRODUCTION

Recently, it has become clear that RNA molecules have a variety of important biological functions in cells, including protein synthesis, RNA processing and modification, mRNA translation, gene regulation, chromosome replication and so on (1–3). Since structures are typically more evolutionarily conserved than sequences, detecting structural similarities among RNA 3D structures can bring more significant insights into their functional and even evolutionary relationships that would not be detected by sequence information alone. Recently, however, the number and size of RNA 3D structures deposited in the Protein Data Bank (4), as well as the Nucleic Acid Database (5), have been substantially and rapidly increasing, making it difficult and time-consuming to manually compare and analyze these RNA tertiary structures. Therefore, it has become more and more crucial to develop automatic tools that are able to efficiently and accurately perform RNA structural comparison.

Theoretically, detecting structural similarities in two RNA molecules at the tertiary structure level is a difficult problem, since it has been shown to be NP-hard to find a constant ratio approximation algorithm for computing a pair of maximal substructures from two protein/RNA 3D structures with exhibiting the highest degree of similarity, if the two proteins/RNAs being compared lie in a general 3D metric (not necessarily Euclidean) space (6). Therefore, currently available tools, such as ARTS (7,8) and DIAL (9), are all based on some heuristic approaches and particularly they are all dedicated to pairwise alignment/comparison of RNA tertiary structures. We refer the reader to (7,9) for briefly reviewing these tools and their approaches.

ARTS was implemented based on a cubic time algorithm that proceeds by a seed match followed by a greedy extension to approximately compute the largest common point set between phosphate atoms of two RNA molecules (7,8). While ARTS can serve as an excellent tool for detecting structural motifs, it is a little time-consuming job for ARTS to compare large RNA molecules (e.g. ribosomal RNAs) due to its cubic time complexity and sometimes the alignments produced by ARTS may be incorrect, as were demonstrated in ref. (9). To overcome these problems, DIAL was then developed based on a quadratic time dynamic programming algorithm by accounting for torsion/pseudo-torsion angle, nucleotide and/or base-pairing similarities (9). DIAL is a versatile tool of pairwise RNA structural alignment, because it can perform three types of alignments: (i) global alignment,

*To whom correspondence should be addressed. Tel: +886 3 5712121, ext. 56949; Fax: +886 3 5729288; Email: cllu@mail.nctu.edu.tw

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

(ii) local alignment and (iii) semiglobal alignment (i.e. a kind of global alignment without penalizing those end gaps that appear in the beginning and end of the alignment). However, we observed that the global, local and semiglobal alignments obtained by DIAL may still be incorrect for some pairs of RNA 3D structures (for details see the Experiments section).

In this study, we utilize the vector quantization (VQ) approach, a technique of high-dimensional clustering commonly used in classical signal processing (10), to derive an RNA structural alphabet (SA) of 23 letters that represent distinct and most common backbone conformations of residues in RNAs with known tertiary structures. Using this SA, we reduce RNA 3D structures to 1D sequences of SA letters and then use classical and efficient sequence alignment algorithms to compare these 1D SA-encoded sequences and determine their structural similarities. Based on such an SA-based approach, we have developed a novel web-based tool, called SARSA (http://bioalgorithm.life.nctu.edu.tw/SARSA/), which provides two RNA structural alignment tools, PARTS (http://bioalgorithm.life.nctu.edu.tw/PARTS/) for pairwise alignment of RNA tertiary structures and MARTS (http://bioalgorithm.life.nctu.edu.tw/MARTS/) for multiple alignment of RNA tertiary structures. For a variety of practical applications, four kinds of pairwise alignments were implemented in PARTS: (i) global alignment (11) for comparing whole structural similarity, (ii) semiglobal alignment (12) for detecting structural motifs, (iii) local alignment (13) for finding locally similar substructures and (iv) normalized local alignments (14) for eliminating the *mosaic effect* of local alignment (i.e. removing poor internal fragments in a local alignment), and a multiple global alignment (15) in MARTS. It is worth mentioning that in SARSA we provide a number of features that are not available in DIAL and other RNA structural alignment tools, such as the normalized local pairwise structural alignment in PARTS and the multiple structural alignment in MARTS. In addition, our experiments have shown that the pairwise alignments produced by our PARTS were comparable to those obtained by DIAL and the performance computation of PARTS was generally faster than that of DIAL. In some cases, our PARTS can actually produce more accurate global, semiglobal and local pairwise alignments when compared with DIAL (for details refer to the Experiments section).

METHODS

For protein backbones, two torsion (or dihedral) angles (ϕ and ψ) are sufficient to describe the conformation of each amino acid residue. In contrast, RNA molecules have much higher dimensionality, since for each nucleotide residue there are six backbone torsion angles ($\alpha, \beta, \gamma, \delta, \varepsilon$ and ζ) and a torsion angle of the bond between base and ribose ring (χ). Recently, Hershkovitz *et al.* (16) have performed a statistical analysis of RNA backbones to search for clusters in RNA conformational space using the so-called VQ, a technique of high-dimensional clustering commonly used in classical signal processing (10). The advantage of employing the VQ technique is that it allows all seven dimensions of RNA conformation to be analyzed simultaneously, so that the smaller number of clusters is needed to classify the RNA structure, as compared to the manual binning method used previously in ref. (17) by analyzing one torsion angle at a time. In fact, as was demonstrated in (16) by plotting torsion angle distributions using a dataset of 132 RNA crystal structures with at least 3.0 Å resolution or better, four torsion angles α, γ, δ and ζ are sufficient for specifying fundamental RNA conformations, since the others are either dependent on these four angles or have distributions with a single peak.

Here, we utilized the VQ approach, followed by a cluster merging, to classify all the residues in the dataset of 132 RNA crystal structures, as was used in ref. (16), only according to their four torsion angles of α, γ, δ and ζ. Consequently, we divided a total of 9,826 residues into a collection of 23 conformational clusters. For our purpose of transforming RNA 3D structures into 1D sequences, we further assigned a letter to each of 23 clusters. We used the set of these 23 letters as a SA and then encoded RNA 3D structures as 1D sequences of SA letters by assigning each residue in the RNA molecules with the letter of the cluster whose center is nearest to the residue being encoded. Like ordinary nucleotide sequences, these SA-encoded 1D sequences can then be aligned using classical methods of pairwise and multiple sequence alignments. For the accuracy of the resulting alignments, we derived two 23×23 scoring matrices, one based on the Hamming distance between each pair of SA-encoded letters and the other based on the statistical method that was used by Henikoff and Henikoff (18) to derive the BLOSUM family of substitution matrices (for details refer to the Supplementary Material).

Currently, four different types of pairwise alignments, global (11), semiglobal (12), local (13) and normalized local (14) alignments, as well as a multiple global alignment (15), were implemented in our web server for a variety of practical applications. Recall that the Smith–Waterman algorithm (13) was originally designed to discard nonsimilar initial and terminal fragments in the pairwise local alignment. However, it was not able to exclude nonsimilar internal fragments, which may lead to a so-called *mosaic effect* by including poor internal fragments in a local alignment (14). As will be illustrated in the Experiments section, such a mosaic effect can still be observed in the comparison of RNA tertiary structures. To overcome the mosaic effect in local alignment, Arslan *et al.* (14) introduced the *normalized local alignment problem* that aims to find the subsequences, say I and J, of two given sequences that maximizes $S(I,J)/(|I|+|J|)$ among all subsequences I and J with $|I|+|J| \geq T$, where $S(I,J)$ is the alignment score between I and J, and T is a threshold for the minimal overall length of I and J. The above length constraint of requiring $|I|+|J| \geq T$ is necessary, because length normalization favors short alignment, but the alignment should be sufficiently long to be biologically meaningful. Actually, the setting value of T can affect the optimal normalized local alignment. If T is small, then the optimal normalized local alignment

tends to be short; otherwise, it tends to be a long local alignment in which may contain nonsimilar internal fragments. In (14), Arslan *et al.* have also proposed a polynomial-time algorithm to solve the normalized local alignment problem. Here, we have implemented such an algorithm for the normalized local alignment of the SA-encoded sequences of two RNA 3D structures.

Based on the SA-based approach described in this study, we have developed the SARSA web server that provides two RNA structural alignment tools, PARTS for pairwise alignment of RNA tertiary structures and MARTS for multiple alignment of RNA tertiary structures. For details, the reader is referred to the Supplementary Material.

USAGE OF SARSA

Input

In PARTS and MARTS, we provide intuitive user interfaces for their usage. PARTS requires the user to input two RNA 3D structures by entering their PDB/NDB ids (4- and 6-character codes, respectively) or uploading them in the PDB format (4). For each input RNA 3D structure, it is optional for the user to enter its chain id as well as its starting and ending residue numbers in sequence. Particularly, note that the user has to specify a chain id if the given RNA molecule has multiple chains. In addition, the user is allowed to change the default settings of all the parameters. For instance, the user can select a pairwise alignment method that can be either global, semiglobal (default), local or normalized local alignment, modify the real values of gap open penalty and gap extension penalty, and specify the number of suboptimal alignments (at least one) if the selected alignment method is semiglobal, local or normalized local alignment.

To run MARTS, the user is required to input multiple (at least two) RNA 3D structures in the format of <pdb|ndb id>:<chain id>:[residue]-[residue], where id items are mandatory and residue items, representing starting and ending residues in the input RNA sequence, are optional. Moreover, the user is allowed to upload the structures in the PDB format and modify the default values of gap open penalty and gap extension penalty.

Output

In the output pages, PARTS and MARTS will first show the details of input RNA molecules, as well as the user-specified parameters. Next, they will show their alignment results, including alignment score, RMSD (root mean square deviation), and resulting alignment of SA-encoded sequences and its corresponding alignment of input RNA sequences. In addition, the user can click the 'Superposition display' link to visually view, rotate and enlarge the 3D structures of input RNA molecules and the superposition of their aligned 3D structures in a Jmol window.

EXPERIMENTS

To assess the accuracy of our PARTS, we calculated its receiver operating characteristic (ROC) curves, depicting the trade-off between true positive rate (i.e. sensitivity) and false positive rate (i.e. 1 minus specificity) and compared them with the best ROC curve of DIAL. By following the procedure that was used by Ferrè *et al.* (9) to compute DIAL's ROC curves, we obtained a filtered, nonredundant dataset that consisted of 51 families and altogether 186 nonfragmented motifs from the SCOR database (19,20), which currently organizes many RNA structural motifs in a hierarchical classification system similar to the SCOP database for protein domains (21). According to this dataset, we computed the ROC curves using the semiglobal alignment of PARTS with two different scoring matrices. As illustrated in Figure 1, the ROC curve using BLOSUM-like scoring matrix performed better than that using Hamming scoring matrix, where the AUC (area under ROC curve) of the former ROC curve is 0.75, while the AUC of the latter is just 0.66. In fact, the alignment results of our PARTS were still comparable to those obtained by DIAL, because as demonstrated in ref. (9) the AUCs of the ROC curves computed by DIAL using the semiglobal alignment method with different parameter settings are between 0.69 and 0.80 [refer to Figure 5 in ref. (9)]. In addition, the performance computation of our PARTS was faster than that of DIAL, even though DIAL was executed on a Linux cluster with 20 computational nodes, each with double CPUs of 1.3–3 GHz and 2 GB RAM, while our PARTS was run only on a single Linux PC with 2.8 GHz CPU and 3 GB RAM. Actually, in some cases as will be demonstrated below, our PARTS produced more accurate global, semiglobal and local pairwise alignments when compared with DIAL. Subsequently, unless otherwise specified, all the experiments were run using our PARTS and MARTS, as well as DIAL, with their default parameters.

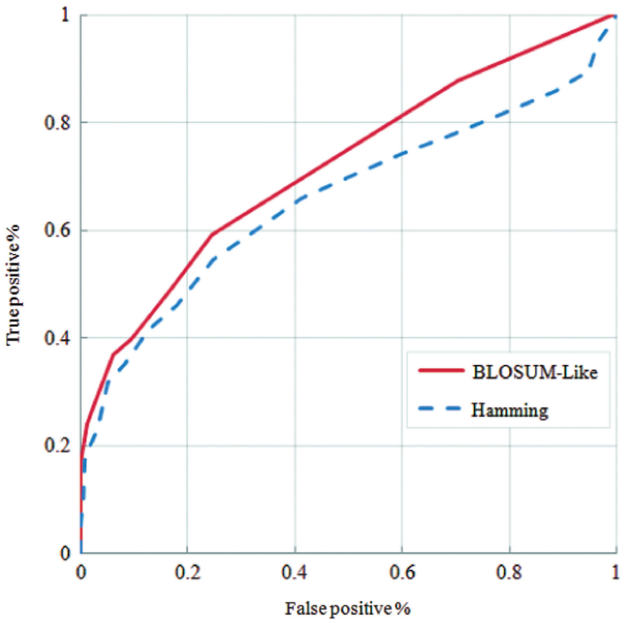

Figure 1. The ROC curves when using the pairwise semiglobal alignment of PARTS to align RNA structural motifs from the SCOR database with two different scoring matrices.

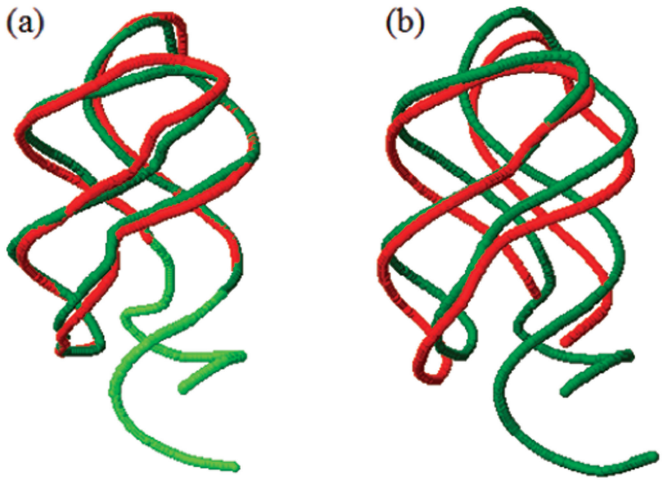

Figure 2. Superposition display of (**a**) PARTS global alignment and (**b**) DIAL global alignment between 1U8D:A (green) and 1Y26:X:25-72 (red).

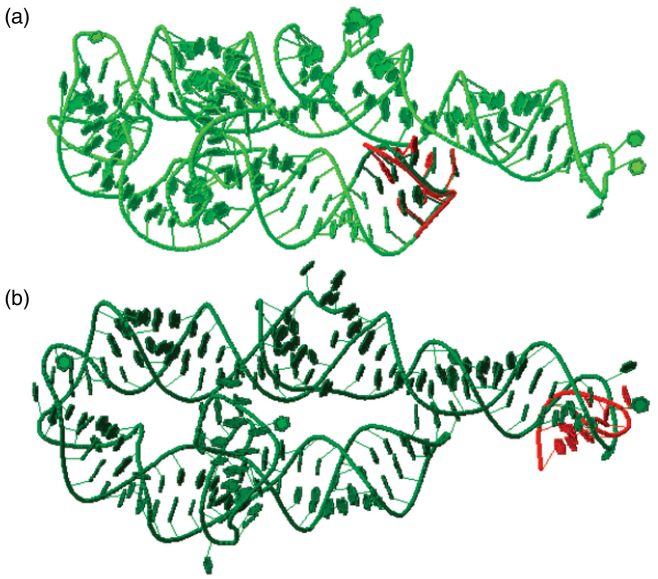

Figure 3. Superposition display of (**a**) PARTS semiglobal alignment and (**b**) DIAL semiglobal alignment using 1J5A:A:2530-2536 (red) and 1HR2:A:103-260 (green).

Pairwise global structural alignment

First of all, we used the tertiary structures of two riboswitches to test our PARTS and DIAL for their capabilities of globally aligning two RNA 3D structures. They are 1U8D (chain A) and 1Y26 (chain X), where 1U8D is the aptamer domain of the guanine-specific riboswitch from the *xpt-pbuX* operon of *Bacillus subtilis*, and 1Y26 is the aptamer domain of the adenine-specific riboswitch from the *Vibrio vulnificus add* gene. In fact, these two riboswitches have nearly identical tertiary structure, although they share only less than 60% sequence identity (22). Consequently, both PARTS and DIAL globally aligned their tertiary structures very well. To demonstrate the difference in global alignment accuracy of PARTS and DIAL, we tested them again by using complete 1U8D:A structure and incomplete 1Y26:X structure that ranges from 25 to 72 residues (1Y26:X:25-72) (i.e. two similar RNA tertiary structures with a little difference in length). As a result, our PARTS globally aligned 1U8D:A and 1Y26:X:25-72 with an RMSD of 1.70 Å (Figure 2a), whereas DIAL globally aligned them with an RMSD of 12.03 Å (Figure 2b). The reason for DIAL's result is that DIAL mis-aligned a fragment of 1Y26:X:68-72 with a fragment of 1U8D:A:76-80, but actually it should be aligned with 1U8D:X:68-72.

Pairwise semiglobal structural alignment

Note that Ferrè *et al.* (9) were the first to utilize the semiglobal alignment, a kind of global alignment without penalizing those end gaps appearing in the beginning and end of the alignment, for the detection of structural motifs in RNA 3D structures. In this experiment, we tested our PARTS, as well as DIAL, by using 1J5A:A with residues 2530–2536 as a query structural motif and 1HR2:A with residues 103–260 as a target RNA molecule. Consequently, our PARTS correctly detected the position of the query structural motif in the target RNA molecule and also returned a semiglobal alignment of IJ5A:A:2530-2536 and 1HR2:A:149-155 with an RMSD of 1.63 Å (Figure 3a). However, DIAL mis-aligned the query structural motif with a different portion of the target RNA molecule (i.e., 1HR2:A:234-240) and returned their semiglobal alignment with an RMSD of 2.43 Å (Figure 3b).

Pairwise local structural alignment

To illustrate the difference in local alignment accuracy of PARTS and DIAL, we applied them to a complete structure of riboswitch 1U8D:A and a partial structure of riboswitch 1Y26:X ranging from 39 to 45 residues (1Y26:X:39-45). Basically, as mentioned above, both of 1Y26:X and 1U8D:A have nearly identical tertiary structure and hence the partial structure 1Y26:X:39-45 should be aligned together with its corresponding substructure in 1U8D (i.e. 1U8D:A:39-45). Consequently, our PARTS aligned them by shifting a residue position, producing a alignment of 1Y26:X:39-45 and 1U8D:A:40-46 with an RMSD of 0.70 Å (Figure 4a). However, DIAL completely misaligned 1Y26:X:39-45 with 1U8D:A:74-80 in its local alignment that has an RMSD of 1.32 Å (Figure 4b).

Pairwise normalized local structural alignment

In fact, it can be observed that a long local alignment between two RNA 3D structures may contain some nonsimilar internal fragments (i.e. the so-called mosaic effect). For instance, Figure 5a displays the structural superposition for the optimal local alignment of two RNA pseudoknots 1L2X:A and 2A43:A that was obtained by our PARTS using Hamming scoring matrix with default parameters (RMSD = 2.40 Å). As shown in this figure, the similar substructures in the initial and terminal regions were very well fitted, but the nonsimilar substructures

in the internal region, ranging from residues 13 to 20, was not fitted well. This mosaic effect in the ordinary local alignment can actually be improved using the normalized local alignment method as implemented in PARTS. For instance, if we run PARTS normalized local alignment by using Hamming scoring matrix and specifying two for the number of suboptimal alignments and 16 for the value of T, then the substructures in the returned normalized local alignments, as shown in Figure 5b and 5c, are fitted very well and have RMSDs of 0.20 Å and 0.26 Å, respectively. As mentioned before, the setting value of T can affect the results of normalized local alignments. For example, if T is set to a large value (e.g. $T = 46$), then PARTS returns a long normalized local alignment, similar to the one depicted in Figure 5a, with nonsimilar internal substructures.

Multiple global structural alignment

We demonstrated the multiple alignment capability of our MARTS by applying it to six tRNA structures (i.e. 1H4S:T, 1ASZ:R:620-660, 1IL2:C, 2CSX:C, 1EVV:A and 1J2B:C) and five RNA pseudoknots (i.e. 1L2X:A, 2AP5:A, 1KPY:A, 2AP0:A and 1YG4:A). Consequently, our MARTS returned a global alignment of the six tRNA 3D structures with an RMSD of 10.73 Å (Figure 6a) and a global alignment of the five RNA pseudoknotted structures with an RMSD of 5.89 Å (Figure 6b). Note that the RMSD we used here for the multiple RNA structural alignment is defined to be the square root of the average sum of all squared pairwise distances.

SUMMARY

In this study, we have developed a web tool SARSA by providing two RNA structural alignment tools that are PARTS that can perform global, semiglobal, local and normalized local pairwise alignment of RNA 3D structures, and MARTS that can perform global multiple alignment of RNA 3D structures. It is worth mentioning again that the normalized local pairwise structural alignment in PARTS and the multiple structural alignment in MARTS are not available in other RNA structural alignment tools. In addition, according to our experiments, our PARTS indeed can quickly produce global, semiglobal and local pairwise structural alignments that are comparable to those obtained by DIAL.

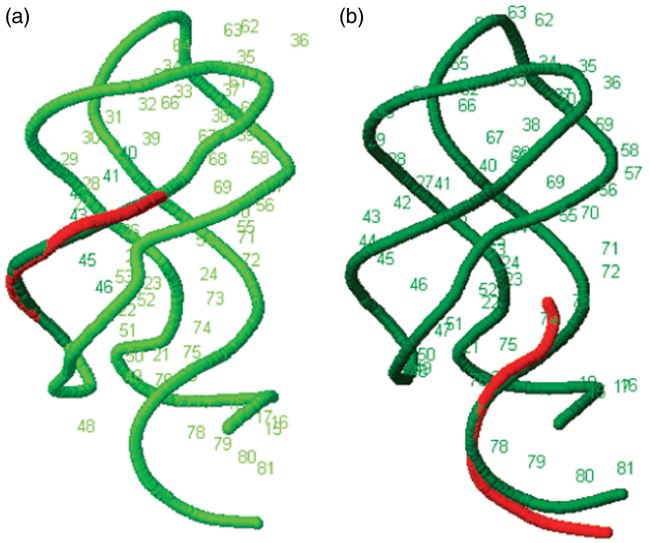

Figure 4. Superposition display of (**a**) PARTS local alignment and (**b**) DIAL local alignment between 1U8D:A (green) and 1Y26:X:39-45 (red).

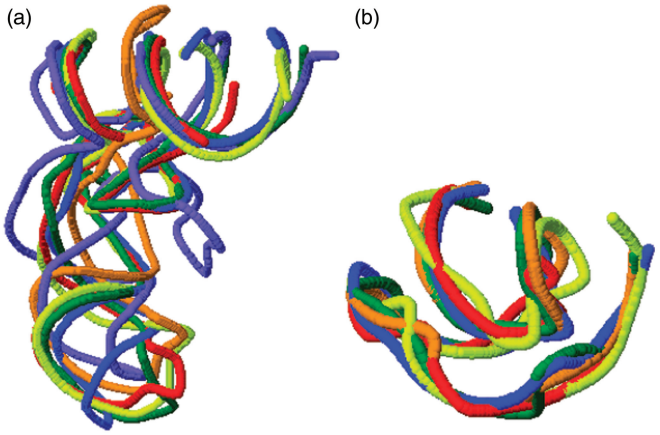

Figure 6. Multiple structural alignments obtained by MARTS for (**a**) six tRNA 3D structures and (**b**) five RNA pseudoknots.

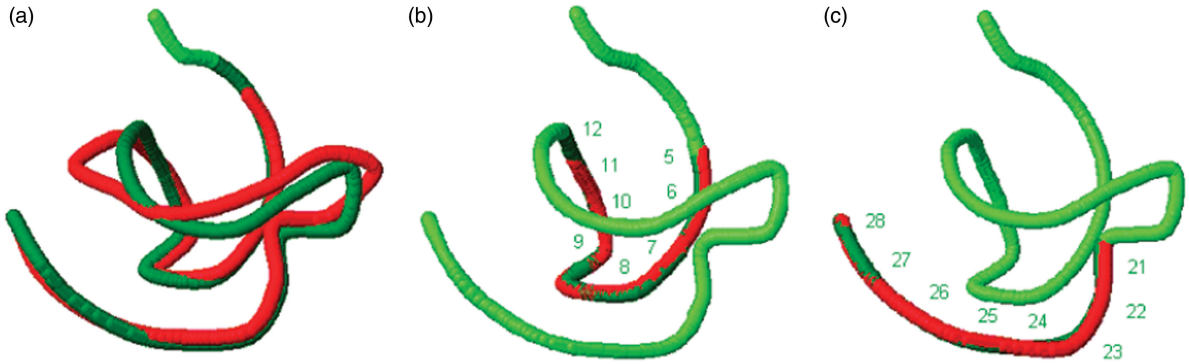

Figure 5. Superposition display of (**a**) PARTS local alignment between two RNA pseudoknots 1L2X:A and 2A43:A, and their (**b**) best and (**c**) second best PARTS normalized local alignments.

Therefore, we believe that our SARSA can serve as a useful tool in the study of structural biology.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported in part by National Science Council of Republic of China under grant NSC96-2221-E-009-126. Funding to pay the Open Access publication charges for this article was provided by the ATU plan of MOE.

Conflict of interest statement. None declared.

REFERENCES

1. Doudna,J.A. (2000) Structural genomics of RNA. *Nat. Struct. Biol.*, **7**, 954–956.
2. Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
3. Storz,G. (2002) An expanding universe of noncoding RNAs. *Science*, **296**, 1260–1263.
4. 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.
5. Berman,H.M., Olson,W.K., Beveridge,D.L., Westbrook,J., Gelbin,A., Demeny,T., Hsieh,S.H., Srinivasan,A.R. and Schneider,B. (1992) The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751–759.
6. Kolodny,R. and Linial,N. (2004) Approximate protein structural alignment in polynomial time. *Proc. Nat. Acad. Sci. USA*, **101**, 12201–12206.
7. Dror,O., Nussinov,R. and Wolfson,H. (2005) ARTS: alignment of RNA tertiary structures. *Bioinformatics*, **21 (Suppl 2)**, 47–53.
8. Dror,O., Nussinov,R. and Wolfson,H.J. (2006) The ARTS web server for aligning RNA tertiary structures. *Nucleic Acids Res.*, **34**, W412–W415.
9. 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.
10. Linde,Y., Buzo,A. and Gray,R.M. (1980) An algorithm for vector quantizer design. *IEEE Trans. Commun.*, **28**, 84–95.
11. Needleman,S. and Wunsch,C. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Evol.*, **48**, 443–453.
12. Setubal, J. and Meidanis, J. (1997) *Introduction to Computational Molecular Biology*, PWS Publishing Company, Boston.
13. Smith,T. and Waterman,M. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
14. Arslan,A.N., Eğecioğlu,O. and Pevzner,P.A. (2001) A new approach to sequence comparison: normalized sequence alignment. *Bioinformatics*, **17**, 327–337.
15. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
16. Hershkovitz,E., Sapiro,G., Tannenbaum,A. and Williams,L.D. (2006) Statistical analysis of RNA backbone. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **3**, 33–46.
17. Hershkovitz,E., Tannenbaum,E., Howerton,S.B., Sheth,A., Tannenbaum,A. and Williams,L.D. (2003) Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Res.*, **31**, 6249–6257.
18. Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Nat. Acad. Sci. USA*, **89**, 10915–10919.
19. Klosterman,P.S., Tamura,M., Holbrook,S.R. and Brenner,S.E. (2002) SCOR: a structural classification of RNA database. *Nucleic Acids Res.*, **30**, 392–394.
20. 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.
21. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
22. Serganov,A., Yuan,Y.R., Pikovskaya,O., Polonskaia,A., Malinina,L., Phan,A.T., Hobartner,C., Micura,R., Breaker,R.R. and Patel,D.J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.*, **11**, 1729–1741.