SARA: a tool for RNA structural alignment

Emidio Capriotti and Marc A. Marti-Renom

Structural Genomics Unit, Centro de Investigación Príncipe Felipe, Valencia, Spain.

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

Aim

The pace of RNA structural determination has been accelerating for the last years. Currently (June, 2007), there are about 1,200 available structures containing one or more ribonucleotides. However, a complete characterization and classification of the RNA structural space have not been yet addressed. We aim to develop a new method for RNA pairwise structure alignment and apply it to an all-against-all comparison of RNA structures in the Protein Data Bank (PDB) (1). To do so, we have addressed the following objectives:

Characterize a set of RNA structural properties
Select an atom type for RNA structural representation
Derive an statistical framework of the alignment significance
Apply to an all-against-all comparison of RNA structures

Methods

Most available methods for protein structure alignment use the Cα atoms of residues to find the equivalences between two protein structures. In the case of RNA structural alignment, there is not a widely accepted standard atom type for representing its structure. For example, the ARTS program (2) uses the phosphate atom and the PRIMOS program (3) uses a reduced representation of the RNA conformation by calculating the η and θ torsion angles. To select a representative atom for the RNA structure, we have calculated the average and standard deviation of the distance between two consecutive nucleotides. We expect that the most significantly associated atom type to structural properties of the RNA would result in the less variable distribution of distances. The C3' atom resulted in an average inter-distance of 5.81Å with 0.44 Å standard deviation. From the C3' atom trace of the RNA backbone, our alignment method called **SARA** (Structure Alignment of Ribonucleic Acid) computes a unit-vector root mean square (URMS) distance between all pairs of nucleotide heptamers. Once an all-against-all matrix is computed, a Dynamic Programming algorithm (4) identifies the common similar regions between the two structures. Finally, a statistical significance of the alignment is calculated similarly to the MAMMOTH algorithm (5). To assess the statistical significance of our results, we built a set of 300 random RNA backbone structures of length spamming from 20 to 320 nucleotides. The random structures were generated by combining randomly selected backbone angles from a set of 42 rotamers (6). More than 44,000 pairwise structural alignments comparing each random structure were used as background distribution to calculate the fitting of the *Mu* and *Sigma* parameters for the extreme value distribution.

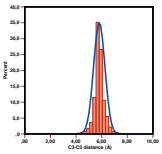
Results

Each of the 310 non-redundant (90% sequence identity) RNA chains in the PDB database (November 2006) with lengths between 20 and 500 nucleotides were structurally aligned to all other chains using the SARA program. More than 39,000 pairwise structural alignments were generated from which 5,421 pairwise alignments resulted in at least 20 C3 atoms superimposed within 4.0Å. The results of analyzing those alignments show that pairs whose sequence identity is higher than 50% superimpose on average ~36 C3 atoms with a local RMSD of ~3.1Å. While pairs whose sequence identity drops under 30% superimpose only on average ~24 of their C3 atoms with a local RMSD of ~3.3Å. Therefore, similarly to what was observed for proteins (7), there may be a detectable relationship between the sequence variation and the structure variation in homologous RNA molecules.

Availability

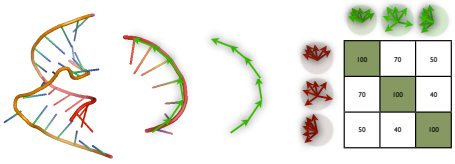
The results of our all-against-all alignments will be included as part of the DBAli server (http://bioinfo.cipf.es/sgu/services/DBAli/).

Atom selection



	Average	Standard deviation
c3-c3	**5.81**	**0.44**
c4-c4	6.10	0.52
c5-c5	6.25	0.55
o3-o3	6.92	0.83
o5-o5	5.81	0.61
P-P	6.07	0.65
cmn-cmr	6.92	0.83

Unit-vector RMS



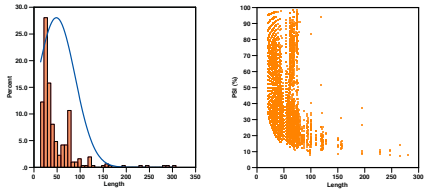
$$URMS^{R} = \sqrt{2.0 - \frac{2.84}{\sqrt{n}}}$$

$$S_{AB} = \frac{(URMS^{R} - URMS^{AB})}{URMS^{R}} \Delta(URMS^{R} \cdot URMS^{AB})$$

$$\begin{bmatrix} \Delta(URMS^{A} \cdot URMS^{AB}) = 10 \rightarrow URMS^{A} > URMS^{AB} \\ \Delta(URMS^{A} \cdot URMS^{AB}) = 0 \rightarrow URMS^{A} \approx URMS^{AB} \end{bmatrix}$$

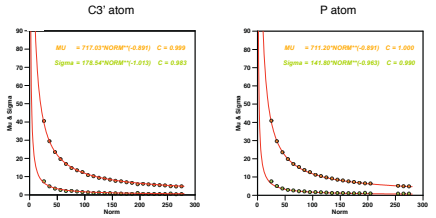
Ortiz AR, Strauss CE, Olmea O. *Protein Sci* 2002, **11**(11):2606-2621.

The problem of length

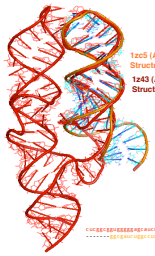


+ Ribosomal RNA with > 1,000 nucleotides

Statistical significance



Example



1zc5 (A)
Structure of the RNA signal essential for translational frame shifting in HIV-1.

1z43 (A)
Structural insights into SRP RNA: An induced fit mechanism for SRP assembly

PSI	78.0%
RMS	2.41 Angstroms
E-value	5.36×10^{-6}
P-value	12.13

References

1. H. M. Berman, et al., *Acta Crystallogr D Biol Crystallogr* 58, 899 (2002).
2. O. Dror, R. Nussinov and H. Wolfson, *Bioinformatics* 21 Suppl 2 (2005).
3. C. M. Duarte, L. M. Wadley and A. M. Pyle, *Nucleic Acids Res* 31, 4755 (2003).
4. T. F. Smith and M. S. Waterman, *J.Mol.Biol.* 147, 195 (1981).
5. A. R. Ortiz, C. E. Strauss and O. Olmea, *Protein Sci* 11, 2606 (2002).
6. L. J. Murray, et al.n, *Proc Natl Acad Sci U S A* 100, 13904 (2003).
7. C. Chothia and A. M. Lesk, *EMBO J* 5, 823 (1986).