**Detecting Conformational Differences Between RNA 3D Structures**

Ryan R. Rahrig1 and Craig L. Zirbel2

1Department of Mathematics and Statistics, Ohio Northern University, Ada, OH 45810

2Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403

**Abstract**

A method is described for detecting local conformational changes between two 3D structures of the same RNA molecule. These could be different 3D structures of the same molecule from the same organism, or they may be 3D structures of homologous molecules from different, yet related, organisms. In this approach, we study the variability that exists among the relative translation and rotation that are needed to superimpose local neighborhoods after global superposition. Each translation is represented by a three-dimensional vector as is each rotation. Thus, we investigate the variability that exists among sets of multivariate data. We demonstrate that the method is able to identify both small- and large-scale conformational changes. Matlab/Octave programs to read RNA 3D structure files and compare structures have been developed and are freely accessible at <https://github.com/BGSU-RNA/ConformationalChange>.

**Key** **Words:** multivariate, outliers, RNA

1. Introduction

Over 3100 RNA 3D structures (Leontis, N.B.; Zirbel, C.L. 2012) have been deposited with the Protein Data Bank (PDB) (Berman, H.M.; Battistuz, T., et al. 2002) and its partner, the Nucleic Acid Data Bank (NDB) (Berman, H.M.; Olson, W.K., et al. 1992). Most of the 3D structures deposited have been determined to atomic precision or near-atomic precision by x-ray crystallography, but some have been determined by cryo-electron microscopy or nuclear magnetic resonance (NMR). In every case, one is attempting to determine the coordinates of hundreds or thousands of atoms in the molecule and its interaction partners such as other RNAs or proteins.

Not all of the crystal structures in the PDB represent distinct molecules. Of the 3100 RNA 3D structures deposited, roughly only 1400 represent distinct RNA molecules (Leontis, N.B.; Zirbel, C.L. 2012). The same RNA molecules have been crystallized by multiple research groups or have been crystallized multiple times by the same research group seeking improved versions of the crystal structure or examining the effects of binding drugs or other molecules. In some cases, small molecules like antibiotics can be diffused into existing crystals to bind to specific sites, and then the same crystal can be subjected to x-ray crystallography again (Auerbach, T.; Bashan, A., et al. 2002). As a concrete example, there are over 250 3D structures of the *Thermus thermophilus* (*T.th.*) small ribosomal subunit alone,

cf. <http://rna.bgsu.edu/rna3dhub/nrlist/view/NR_4.0_81883.33>

Molecules are dynamic entities which can change shape as they bind to different interaction partners. Riboswitches change shape in response to their ligand or metabolite binding. The ribosome changes shape in dramatic ways during different stages of translation as different protein factors and tRNAs are bound. Some 3D structures catch the two subunits of the ribosome in different stages of ratcheting, as described by Zhang et al. (2009). Also of interest is the rotation of the head domain of the small (30S) subunit of the ribosome, which is required for movement of mRNA and tRNAs (Mohan, Donahue, Noller, 2014).

Moreover, RNA 3D structures are subject to a variety of smaller influences. Even in crystallized form, each copy of the RNA molecule can be in slightly different geometric conformations. Some parts of the molecule might be loose enough to move about quite a bit, which will make it more difficult to determine the locations of the atoms in these parts. In x-ray crystallography, one diffracts x-rays from the crystal, measures the x-ray intensity and scattering as a function of scattering angle, and infers the density of electrons at spatial points throughout a single copy of the molecule by 3D Fourier transformation. This is inherently imprecise due to variability in detection of x-rays and the limited resolution of the detectors. Then a crystallographer attempts to fit atoms and RNA nucleotides to the electron density. To the extent that parts of the molecule are mobile, the electron density may be too diffuse to fit. Similar considerations apply to cryo-electron microscopy and NMR.

Thus, RNA 3D structures are subject to two different kinds of variability: variability due to differences in interaction partners and experimental conditions, and variability due to thermal fluctuations, experimental measurement, and fitting. We would like to determine what variability between RNA 3D structures is routine, and what variability is significant, so that we focus on meaningful changes and differences.

For example, research groups may want to analyze which regions of two crystal structures deviate from one another to determine any errors that may have been made during the modeling process. Or it may be beneficial to learn what conformational changes take place within the 3D structure of a molecule when an antibiotic or other ligand is bound.

Fortunately, we have many instances of duplicate structures, and so are able to characterize routine variability by comparing such 3D structures. We note that when we compare two different crystal structures of the same molecule from the same organism, we know exactly what nucleotides correspond between the two structures; thus, no alignment of the structures is required. When comparing homologous molecules from different organisms, we can make a structural alignment using R3D Align (Rahrig, 2010).

The classical tool for comparing two different models of a molecule is the average root-mean-square deviations (RMSD) after optimal superposition of the two structures. However, while RMSDs do capture differences in the general shapes of the molecules, they do not provide any information regarding the variability that exists between the structures at a more local level. In (Parisien 2009), two metrics suitable to RNA comparisons are described – the deformation index and deformation profile. The deformation index considers both base–base-stacking and base–base-pairing interactions while the deformation profile compares the structures at the nucleotide level for both intradomain and interdomain interactions. However, these do not indicate the nature of the deformation.

The idea in the present paper is to characterize the variability in local neighborhood locations and orientations in RNA 3D structures. Regions of variability will indicate conformational differences between the two structures. The method we describe in this paper finds local conformational changes with clear intuitive interpretation.

2. Method

2.1 Using Translations and Rotations to Detect Conformational Changes

The key idea of this paper is to analyze the variability that exists in the optimal superpositions of local neighborhoods, in comparison with the optimal superposition of the global structures. This will allow us to identify regions that are translated or rotated differently, relative to the rest of the structure.

An optimal superposition of nucleotides *a*1,..., *an* of structure *A* onto nucleotides *b*1,..., *bn* of structure *B* is computed as follows. Let α1,..., α*n* and *β*1,..., *βn* be the geometric centers of the bases of *a*1,..., *an* and *b*1,..., *bn*, respectively. The optimal superposition then consists of a translation vector *t* and a rotation matrix *R* that minimize the squared error,



First, the optimal global rotation matrix *R\** and translation vector *T\** are obtained using the technique described by Berthold et al. (1988). These are used to superimpose the two structures globally.

Then for each nucleotide *i* in the first structure, its five-nucleotide local neighborhood is found (according to Euclidean distance), and then the translation vector *ti* and rotation matrix *Ri* of the optimal superposition onto the corresponding five nucleotides in the other structure are found. Thus, *ti* and *Ri* are found relative to *T\** and *R\*.*

We note that while a rotation is typically represented as a 3×3 matrix, it can be expressed as a 3-dimensional vector using the axis of rotation and making its length equal to the angle of rotation in degrees. For our purposes, we will represent each *Ri* in this vector form.

2.2 Illustration of Using Translations to Detect Conformational Changes

As an example we compare the crystal structures found in the PDB files 2UUB and 2UUC. Overall, we expect them to be very similar since they are both crystallizations of the same organism with the same antibiotic bound. Also, the crystal structures were determined by the same research group and were deposited in the PDB only one week apart. Both files contain crystal structures of *T.th*. 16S RNA bound to an mRNA with a codon in the A-site and the antibiotic paromomycin. 2UUB is bound with a GUU-codon and 2UUC is bound with a GUA-codon (Weixlbaumer et al., 2007).

First, the translation vectors are analyzed. For each nucleotide *i* in the structure, the translation vector *ti* is calculated as described above. Figure 1 displays a 2D view of a scatterplot of the translation vectors. It can be seen that the three-dimensional data points are roughly centered at zero. What is of interest is the variability of the points about the mean as that provides a measure of the similarity (or dissimilarity) of the two structures. The more a point deviates from the mean, the greater the conformational difference among the neighborhood of nucleotides represented by that point.



**Figure 1:** View of 3D scatterplot of translation vectors for files 2UUB and 2UUC. Units are

Ångstroms (1Å=10−10 m). Points are centered about the origin and colored according to the corresponding nucleotide’s distance from the center of the structure. The colors range from blues to reds according to the color bar shown on the right. Extreme data points are labeled using the method outlined in Section 2.3, using a cut-off value of 25 for D*i*2.

For example, in the lower left of Figure 1, nucleotide 1257 certainly appears to be an extreme data point as it is set apart from the rest of the data points. Figure 2 shows a portion of the global superposition showing nucleotide 1257 and its local neighborhood. Nucleotides 1257 from the two structures do not superimpose well, which makes it clear why the translation *t*1257 is farther removed from the origin than most other translation vectors. Its norm is approximately 1.4 Ångstroms.



**Figure 2:** Portion of the global superposition showing the neighborhoods of nucleotide 1257 in 2UUB and 2UUC. The neighborhood includes nucleotide 1257 along with the nearest four nucleotides in 3D space, which are nucleotides 1256, 1258, 1259, and 1260.

The rotation vectors can be analyzed in a similar way as the translation vectors, although the interpretation is different. The rotation vectors provide additional information by indicating a different type of conformational change. For example, the 5-nucleotide local neighborhood around base U1257 in 2UUB, shown in Figure 2, requires roughly a 26.2 degree rotation relative to the overall rotation between the structures.

We are also able to detect larger regions of conformational difference by discovering clusters of data points that are scattered further from the mean. For example nucleotides 1026-1039 have relatively extreme values. A superposition of these nucleotides is provided in Figure S1 found in the supplementary figures provided in the corresponding Github repository.

Although we have just seen that extreme data points can be found by simply observing the scatterplot, a less subjective, more formal statistical process may be desired. In the next section, we provide a description of the statistical methods that can be employed in order to measure variability and detect extreme values within multivariate data of this type.

**2.3 Detecting Outliers in Multivariate Data**

We are essentially dealing with the problem of identifying outliers and other extreme data points within a sample of three-dimensional data. Detecting outliers is more diﬃcult when dealing with multivariate data than with univariate data. This is because multivariate data cannot be ordered in the same way that a univariate sample can. Thus, the data cannot simply be ordered in such a way that the extreme values show up on either end. With multivariate data, an observation vector may have a large error in one of its components or smaller errors in several components. However, if the multivariate data is two-dimensional or three-dimensional, we may still perceive an observation to be particularly extreme when the data is displayed as a scatterplot as in Figures 1 and 4.

Despite no obvious ordering of the data being present, it is necessary to adopt some notion of ordering in order to determine extremeness. The most common method of ordering multivariate data is to reduce each multivariate *x* to a scalar quantity *D*(*x*) (often some type of distance measure), thereby creating a univariate data set from which extreme values can be detected (Barnett and Lewis, 1994).

We will represent each observation vector *xi* (which may either represent *ti* or *Ri* ) by its standardized distance from the mean. We denote *D*(*xi*) as *Di* . Then we have,



where *S* is the sample variance-covariance matrix:



The higher the value of *Di* 2, the greater the distance of the point from the mean. Thus the task of finding conformational changes among structures amounts to finding nucleotides whose corresponding data points have larger values for *Di* 2.

When the data is normally distributed, a formal outlier test exists that is based on *Di* 2. This is because the Wilks’ statistic (Wilks, S.S., 1963) can be expressed in terms of *D*2(*n*), where



The Wilks’ statistic is



where *S*−*i* is the sample variance-covariance matrix of the data with the *ith* observation deleted. The Wilks’ statistic can also be written as



Thus a test for an outlier can be based on *D*2(*n*). Tables are available containing the upper 5% and 1% critical values of *D*2(*n*) (Rencher, 2002). However, in our case no normality assumption is made since we are not particularly interested in the formal declaration of outliers. We are more interested in which points are more extreme relative to the others, for which an ordering of the *D*2*i* values is sufficient. It is left to a future study to determine whether the data points *ti* and *Ri* can be modeled as normally distributed.

In addition to discovering local regions of conformational change, we are also interested in determining the overall similarity of two structures. We know that the more similar two structures are, the closer the data points should be centered about the mean. Thus, if we study the overall variability of the data points, we can learn more about the overall similarity of the two structures. While this information is contained within the sample variance-covariance matrix *S*, a single numerical value for the overall multivariate scatter is often desired. The total sample variance, which is the trace of *S*, can serve this purpose. Since we are working with three dimensional data, we have

Total sample variance = *s*11 + *s*22 + *s*33

3. Results

**3.1 Analyzing Total Variability**

First, we want to understand the variability in local conformations between two RNA 3D crystal structures of the same molecule from the same organism and produced by the same research group.

As our ﬁrst example, we use two crystal structures of the *T.th.* 16S rRNA that were produced by the same research group. We use the crystal structure found in PDB ﬁle 1FJG [Carter, 2000] and a more recent structure found in PDB ﬁle 1J5E. Each of these crystal structures was determined with no ligands or other complexes bound to the molecule.

As in section 2.2, we can produce scatterplots of the translation and rotation vectors and use *Di*2 values to ﬁnd regions of conformational change. The scatterplots of the translation and rotation vectors are included in the supplementary data as figures S2 and S3, respectively. Here we focus on the overall variability of the data points since that provides an indication of how similar the two structures are. The total sample variance for the translation vectors is found to be 0.2296, and the total sample variance for the rotation vectors is 10.6139.

While it is diﬃcult to analyze these numbers in and of themselves, these variances are useful for the sake of comparison. We next compare two crystal structures of the same molecule from the same organism, but where the crystal structures have been determined independently by diﬀerent research groups. We expect there to be greater variability in the data points than in this case since there is additional variability in the modeling processes. We use the 1J5E crystal structure again, but instead of using 1FJG which was determined by the same group that determined 1J5E, we use 2ZM6 which was solved more recently by an independent research lab (Kaminishi, T. et al., 2009).

As expected, the variances of the translation and rotation vectors are larger in the case when the structures were crystallized by diﬀerent groups than by the same group. The total sample variance for the translation and rotation data is 0.8938 and 37.5299, respectively. These values are much larger than in the example described in the previous section.

The scatterplots of the translation and rotation vectors are included in the supplementary data as figures S4 and S5, respectively.

**3.2 Head rotation in small ribosomal subunit**

A dramatic example of conformational change occurs between two *T.th.* small ribosomal subunit structures, 1FJG [Carter, 2000] and 4QS0 [Zhou, 2014]. In the second structure, the head of the small subunit has been caught in a rotated state, as described by [Mohan, 2014]. The fact that two large regions of the structure have different local rotation matrices compared to the global rotation is apparent in Figure 3.



**Figure 3.** Relative rotation vectors between chain A of 1FJG and chain A of 4QS0, colored by position in the sequence (from1 to 1507). Nucleotides in the head occur between positions 910 to 1380, and these are colored yellow to orange in the figure. Their relative rotation vectors clearly stand apart from the others, on either side of the vector (0,0,0) representing the global rotation.

Furthermore, the relative translation graph shows a spread of points, as the nucleotides in the head that are most distant from the hinges move the most, as is clearly shown in Figure 4.



**Figure 4.** Relative translation vectors between chain A of 1FJG and chain A of 4QS0, colored by position in the sequence (from1 to 1507). Nucleotides in the head occur between positions 910 to 1380, and these are colored yellow to orange in the figure. Their relative translation vectors clearly stand apart from the others, on either side of the vector (0,0,0) representing the global translation.

**Remove the following paragraph?**

Note that relatively few nucleotides are labeled in Figure 4, while many nucleotides are labeled in Figure 3, even though the cutoff value [explain better!] was set to 15 in both cases. This is apparently because the distribution of translation vectors is so much more spread out than the distribution of relative rotation vectors. [Really? Or could something calculated differently/wrongly in the programs?]

**3.3 Comparing Homologous Molecules from Different Organisms**

Our examples thus far have compared two different crystal structures of the same molecule form the same organism. In this section, we analyze two 16S rRNAs of homologous molecules from diﬀerent organisms. When we compare two molecules from diﬀerent organisms, we do not know exactly which nucleotides correspond, so an alignment of the two structures is ﬁrst required. Structural alignments are obtained using R3D Align (REF).

He we analyze the conformational differences between the *T.th*. 1J5E and *E.Coli* 2AVY ﬁles [REFs]. A structural alignment was obtained using R3D Align [REF]. Because 1j5e and 2avy are crystal structures of diﬀerent molecules, we expect there to be greater variability in the translation and rotation vectors than in the previous cases where the crystal structures are of the same molecule. The scatterplots for the translation and rotation vectors are shown in Figures \_\_ and \_\_\_, respectively, in the supplementary data. There are clear clusters of local neighborhoods that vary all in the same way. Most of these are at some distance from the center of the molecule, where more ﬂexibility is allowed.

It is evident from the scatterplots that there is indeed more variability in this case. This is conﬁrmed by analyzing the total sample variances which are computed to be 9.8974 and 925.0557. These values are much higher than in the other cases where the structures were from the same molecule such as in Section 3.1.

**4. Software and availability**

The software consists of three new programs that extend the FR3D and R3D Align packages. We describe them briefly here; additional documentation within the programs explains how they work and how they can be easily modified.

**CompareStructures.m** This program loads the RNA3D structures, calculates the global superposition, superimposes local neighborhoods, and calculates the relative rotation and translation vectors. It optionally displays scatterplots as shown in the paper, returns a data structure containing all data, and writes the data to an Excel spreadsheet. Points can optionally be colored by distance from the center of the structure, position in the sequence, or local geometric discrepancy between local neighborhoods (to bring out neighborhoods that are locally deformed).

**Interact.m** This program creates the two-dimensional scatterplot shown in Figure F3. The user may click a point in the graph to visualize the local neighborhood, and then enlarge or shrink the local neighborhood. Nearby amino acids, if any, are also shown, in case they cause a conformational change.

**Example.m** This script contains the commands to evaluate structures 1FJG and 4QS0, display the relative rotation and translation vectors, and interact with Figure F3.

The software is available through Github at this URL: <https://github.com/BGSU-RNA/ConformationalChange> As explained in the Readme file, one also needs to install the FR3D program suite and, if desired, the R3D Align programs. Support for reading mmCIF files is included in the current release of FR3D. The software has been extensively tested in the Matlab and Octave environments. The Readme file gives a number of pointers for working with Octave.

****

**Figure F3.** Scatterplot of rotation angle versus norm of translation vector, for the example of 1FJG and 4QS0 discussed above.

5. Conclusion

The procedure described in this paper, implemented in the Matlab programming language, is useful for quickly determining local conformational differences between two 3D structures of the same RNA. These structures may be crystallizations from the same molecule from the same organism, or they may be crystallizations of molecules from different, yet related, organisms. Comparing molecules from different organisms first requires a 3D alignment, which can be determined by programs such as R3D Align (Rahrig, R. et al., 2010). As a whole, this suite of programs provides valuable tools for researchers to quickly gain further insight into the function and structure of RNA 3D molecules.

References

[1] T. Auerbach, A. Bashan, et al., Antiobiotics Targeting Ribosomes: Crystallographic Studies. Curr. Drug Targets - Infect. Dis. 2 (2002), 169-186.

[2] Vic Barnett and Toby Lewis, Outliers in Statistical Data. Wiley & Sons. 3rd edition (1994).

[3] H.M. Berman, T. Battistuz, et al., The protein data bank. Acta crystallographica Section D, Biological crystallography. 58 (Pt 6 No 1) (2002):899–907.

[4] H.M. Berman, W.K. Olson, et al., The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. Biophysical journal. 63(3)(1992):751–759.

[5] Berthold K.P. Horn, H.M. Hilden, and Shariar Negahdaripour, Closed-form solution of absolute orientation using orthonormal matrices. Journal of the Optical Society of America. 5(7) (1988):1127–1135.

[] A.P. Carter, et al., Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature. 407 (2000): 340-348.

[6] T. Kaminishi, et al., Crystal structure of the Thermus thermophilus 30S ribosomal subunit. Submitted for publication.

[7] Srividya Mohan, John Paul Donohue, and Harry F. Noller, Molecular mechanics of 30S subunit head rotation. Proceeding of the National Academy of Sciences. 111(37) (2014): 13325-13330.

[8] N.B. Leontis and C.L. Zirbel, Nonredundant 3D Structure Datasets for RNA Knowledge Extraction and Benchmarking. In N. Leontis & E. Westhof (Eds.). 27 (2012):281–298.

[9] J.M. Ogle, et al., Selection of tRNA by the ribosome requires a transition from an open to a closed form. Cell. 111(5) (2002):721–732.

[10] M. Parisien, J.A. Cruz, E. Westhof, and F. Major,New metrics for comparind and assessing discrepancies between RNA 3D structures and models. RNA. 15 (2009):1875-1885.

[11] R.R. Rahrig, C.L. Zirbel, and N.B. Leontis, R3D Align: Global pairwise alignment of RNA 3D structures using local superpositions. Bioinformatics. 26 (2010): 2689-2697.

[12] Alvin C. Rencher, Methods of Multivariate Analysis. Wiley & Sons. 2nd edition (2002).

[13] B.S. Schuwirth, et al., Structures of the Bacterial Ribosome at 3.5 A Resolution. Science. 310 (2005): 827-834.

[14] B.S. Schuwirth, et al., Structures of the Bacterial Ribosome at 3.5 A Resolution. Science. 310 (2005): 827-834.

[15] A. Weixlbaumer, et al., Mechanism for expanding the decoding capacity of transfer RNAs by modiﬁcation of uridines. Nature structural & molecular biology. 14(6) (2007):498–502.

[16] S.S. Wilks, Multivariate statistical outliers. Sankhy: The Indian Journal of Statistics. Series A, 25(4) (1963):407–426.

[17] B.T. Wimberly, et al., Structure of the 30S ribosomal subunit. Nature. 407 (2000): 327-339.

[18] W. Zhang and J.A. Dunkle, et al., Structures of the ribosome in intermediate states of ratcheting. Science. 325(5943) (2009):1014–1017.

[] J. Zhou, et al., How the ribosome hands the A-site tRNA to the P site during EF-G-catalyzed translocation. Science. 345 (2014): 1188-1191.