

Assessing Variability of RNA Molecule Crystallizations

Ryan R. Rahrig¹

¹Ohio Northern University, Ada, OH 45810

Abstract

A method is described for detecting local conformational changes 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. In this approach, we study the variability that exists among the 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.

Key Words: multivariate, outliers, RNA

1. Introduction

Over 2600 RNA 3D structures have been deposited within 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 a process known as x-ray crystallography, but some have been determined by NMR (nuclear magnetic resonance) or electron microscopy. In every case, one is attempting to determine the atomic coordinates of hundreds or thousands of atoms in a particular type of molecule. But molecules themselves are dynamic entities; the positions of atoms fluctuate due to thermal noise about their equilibrium positions and because the molecule gets hit often by water molecules in the solution. Moreover, every crystal will contain many copies of the same molecule, and these copies will all be in slightly different geometric conformations.

Crystallography reduces the variability somewhat, because the molecules align to form a periodic crystal structure (with the unit cell as the repetitive element), and so are locked into place, but they can still jiggle about internally. 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.

Thus, 3D structures as determined by x-ray crystallography are subject to four kinds of variability. There is variability in the conformation of the molecule itself, as well as thermal fluctuations which make parts of the electron density difficult to fit, detection errors in determining the electron density, and also human error in fitting atoms to the electron density.

Each structure can result in a 3D crystal structure that is unique and differs from the determination of other crystal structures. RNA molecules represented in PDB files 1j5e, described in Wimberly et al. (2000), and 2avy, described in Schuwirth et al. (2005), are crystal structures of the 16S ribosomal RNA molecule from two different organisms, *Thermus thermophilus* (*T.th.*) and *E. Coli* (*E.c.*). The molecules themselves are different, so it is no surprise that the 3D structures differ, but they are homologous, so it is also no surprise that in many places, the 3D structures are quite similar. But not all of the crystal structures in the PDB represent distinct molecules. Of the 2600 RNA 3D structures deposited, roughly only 700 represent distinct RNA molecules.

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). Also, some molecules have been crystallized in complex with functional ligands such as mRNA or tRNA (in the case of ribosomes). Other times 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.

It is necessary to develop methods capable of detecting and assessing the variability among these structures since there are a variety of instances in which it is useful to detect the conformational changes between two structures. 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. Some crystallizations catch the two subunits of the ribosome in different stages of ratcheting, as described by Zhang et al. (2009). In all cases, we need to determine what variability between RNA 3D structures is typical, and what variability is unusual.

Fortunately, we have many instances of duplicate structures, and so are able to characterize the variability by comparing two (or more) 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.

The goal here is to characterize the variability in local neighborhood locations and orientations in RNA 3D crystal structures, and between RNA crystal structures of different organisms.

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 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. Finding local conformational changes can help improve the modeling process by determining local areas of deviation.

2. Method

2.1 Using Translations and Rotations to Detect Conformational Changes

We analyze the variability that exists in the deviations among the optimal superpositions of local neighborhoods, in comparison with the optimal superposition of the global structures.

An optimal superposition of nucleotides c_1, \dots, c_n onto l_1, \dots, l_n , is one that consists of a translation vector t and a rotation matrix R that minimize the squared error,

$$L^2 = \min_R \min_t \|l_i - R(c_i - t)\|^2$$

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 t_i and rotation matrix R_i of the optimal superposition onto the corresponding five nucleotides in the other structure are found. Thus, t_i and R_i are found relative to T^* and R^* . Careful attention must be paid to the calculation of each R_i and t_i so that each neighborhood's rotation and translation corresponds to the same reference orientation.

The calculation of t_i and R_i is now described in more detail. Suppose nucleotides $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, and α_5 are to be superimposed onto the neighborhood consisting of nucleotides $\beta_1, \beta_2, \beta_3, \beta_4$, and β_5 . We let $\bar{\alpha}$ be the center of mass for the neighborhood in the first structure and $\bar{\beta}$ be the center of mass for the neighborhood in the second structure. That is,

$$\bar{\alpha} = \frac{\sum_{j=1}^5 \alpha_j}{5} \quad \text{and} \quad \bar{\beta} = \frac{\sum_{j=1}^5 \beta_j}{5}.$$

Then we have $t_i = \bar{\beta} - \bar{\alpha}$ and R_i determined as in Berthold (1988).

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 R_i in this vector form.

It is important to point out once again that the translation and rotation vectors only indicate deviations from the global superposition. Therefore, the translation vectors t_i should be small and centered around 0. The rotation vectors should also be short (small angle of rotation).

2.2 Examples Using Translations and Rotations to Detect Conformational Changes

As an example we compare the crystal structures found in the PDB files 2uub and 2uuc.

We select these two structures because 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, A. et al., 2007).

First, the translation vectors are analyzed. For each nucleotide i in the structure, the translation vector t_i is calculated as described above. Figure 1 displays 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.

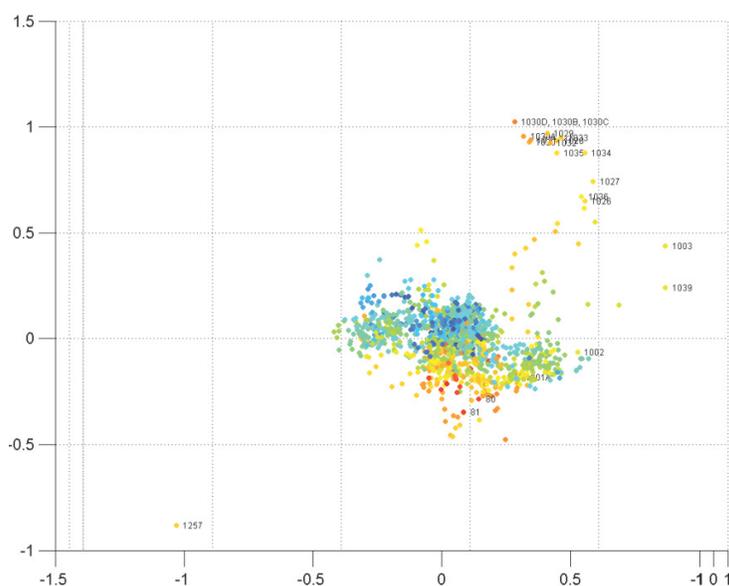


Figure 1: View of 3D scatterplot of translation vectors for files 2uub and 2uuc. Units are Angstroms ($1\text{\AA}=10^{-10}$ m). Points are centered about the origin and colored according to the corresponding nucleotide's distance from the center of the structure. The colors range from blues to reds according to the color bar shown in Figure 2. Extreme data points are labeled using the method outlined in Section 2.3, using a cut-off value of 25 for D_i^2 .

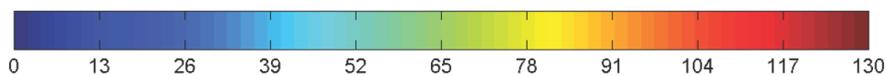


Figure 2: The data points in the 3D scatterplots of this chapter are colored according to the distance to the center of the molecule, as given by this color bar. The color bar ranges from 0 Angstroms to 130 Angstroms since nucleotides of 16S *T.th.* rRNA molecules typically are no more than 130 Angstroms from the center of the structure.

For example, in 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 3 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 t_{1257} is farther removed from the origin.

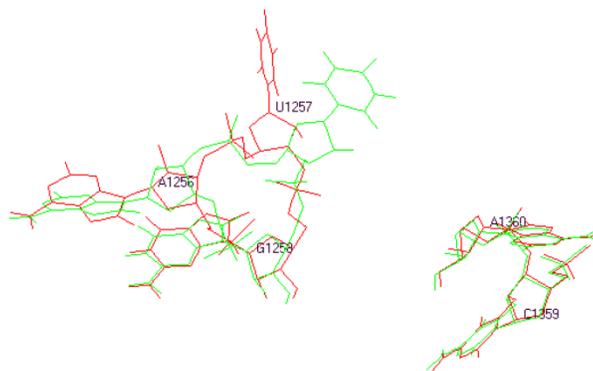


Figure 3: Portion of the global superposition showing the neighborhoods of nucleotide 1257 in 2uub and 2uuc. The neighborhood includes nucleotide 1257 along with the nearest four nucleotides in 3D space, which are nucleotides 1256, 1258, 1359, and 1360.

We are also able to detect larger regions of conformational difference by discovering clusters of data points that are scattered further from the mean. In Figure 1, there is a cluster of data points representing nucleotides 1026-1039 that have relatively extreme values. These nucleotides comprise a hairpin loop and basepairs in the helical region adjacent to the hairpin loop. Figure 4 displays the portion of the global superposition with these nucleotides along with nucleotides 1019-1025 and 1040-1045. It can be seen that nucleotides 1026-1039 do not superimpose as well in 3D space as the other nucleotides, which explains why the corresponding t_i values are extreme.

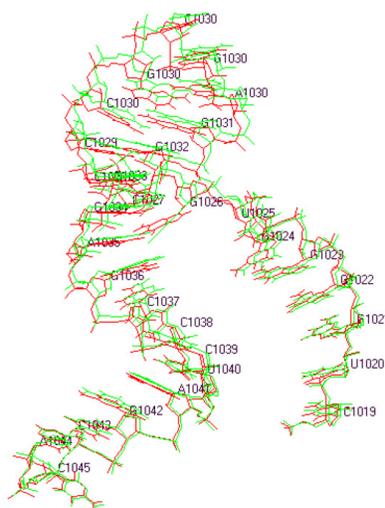


Figure 4: Superposition of nucleotides 1019-1045 of 2uub and 2uuc. Nucleotides in the hairpin and near the hairpin do not superimpose as well as the others which illustrates why the corresponding points on the scatterplot are relatively extreme.

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.

Figure 5 shows a view of the 3D scatterplot of the points representing the rotation deviations for the nucleotides of the crystal structures 1n33 and 1n34. Again, these structures were selected because they were produced by the same lab so we expect there to be many similarities. However, they have different complexes bound so we expect there to be some variability between the two structures. The 1n33 file contains the crystal structure of the *T.th.* 30S ribosomal subunit bound to codon and near-cognate transfer RNA anticodon stem-loop mismatched at the second codon position at the a site with paromomycin, and the 1n34 file contains the crystal structure of the *T.th.* 30S ribosomal subunit in the presence of codon and crystallographically disordered near-cognate transfer RNA anticodon stem-loop mismatched at the first codon position (Ogle, J.M. et al., 2002).

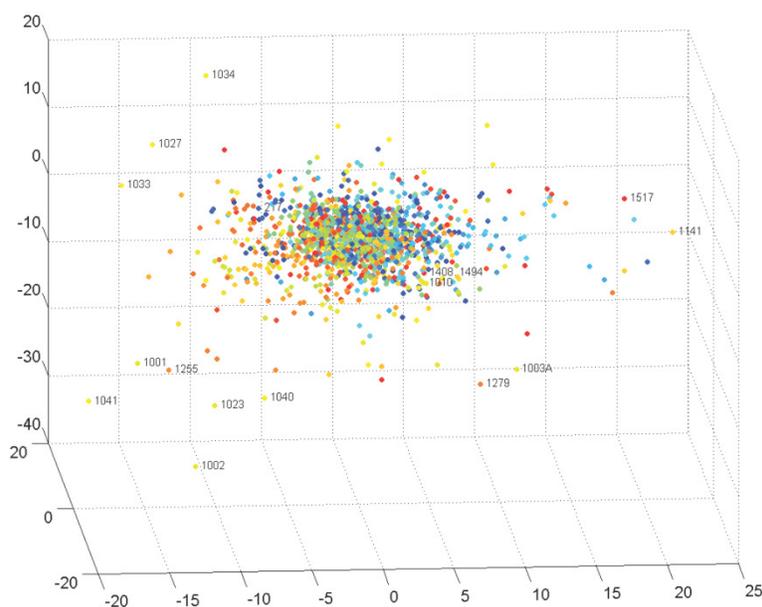


Figure 5: View of 3D scatterplot of rotation vectors for files 1n33 and 1n34. Units are degrees. Extreme data points are labeled using the method outlined in Section 2.3, using a cut-off value of 25 for D_i^2 .

The point labeled 1141 appears to be an extreme value. That is, there is a greater variability between the optimal rotation of the neighborhood of nucleotide 1141 and the global optimal rotation than the optimal rotations of other neighborhoods. Figure 6 shows the superposition of the neighborhood taken from the global superposition.

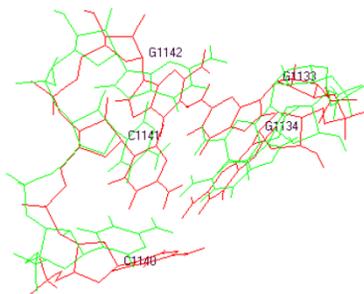


Figure 6: Superposition of local neighborhoods of nucleotide 1141 of 1n33 and 1n34. The two neighborhoods have different conformations resulting in a more extreme rotation vector.

Although we have just seen that extreme data points can be found by simply observing the scatterplot, a less subjective, more formal statistical process is needed. In the next section, we provide a description of the statistical methods that will 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 5. 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 x_i (which may either represent t_i or R_i) by its standardized distance from the mean. We denote $D(x_i)$ as D_i . Then we have,

$$D_i^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x}),$$

where S is the sample variance-covariance matrix:

$$S = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

The higher the value of D_i^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 D_i^2 .

When the data is normally distributed, a formal outlier test exists that is based on D_i^2 . This is because the Wilks' statistic (Wilks, S.S., 1963) can be expressed in terms of $D_{(n)}^2$, where

$$D_{(n)}^2 = \max_i D_i^2.$$

The Wilks' statistic is

$$w = \max_i \frac{|(n-2)S_{-i}|}{|(n-1)S|},$$

where S_{-i} is the sample variance-covariance matrix of the data with the i^{th} observation deleted. The Wilks' statistic can also be written as

$$w = 1 - \frac{nD_{(n)}^2}{(n-1)^2}$$

Thus a test for an outlier can be based on $D_{(n)}^2$. Tables are available containing the upper 5% and 1% critical values of $D_{(n)}^2$ (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_i^2 values is sufficient. It is left to a future study to determine whether the data points t_i and R_i 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 , it is desirable to have a single numerical value for the overall multivariate scatter. For that we use the total sample variance, which is simply the trace of S . Since we are working with three dimensional data, we have

$$\text{Total sample variance} = s_{11} + s_{22} + s_{33}$$

3. Results

3.1 Same Molecule, No Ligands Bound, Same Lab Analysis

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.thermophilus* 16S rRNA that were produced by the same research group. We use the crystal structure found in PDB file 1fjf and a more recent structure found in PDB file 1j5e. 1fjf was released in September of 2000 while 1j5e was released in April of 2002. Each of these crystal structures was determined with no ligands or other complexes bound to the molecule.

The scatterplots of the translation and rotation vectors are given in Figures 7 and 8, respectively. Points on the plot have been labeled with the corresponding nucleotide number if the value for D_i^2 is greater than 25. This number was used as a cut-off value since that is approximately the critical value for the outlier test described above at the .05 level for a sample of this size (with normal data). Figure 9(a) gives the largest 20 values of D_i^2 along with the corresponding nucleotide numbers.

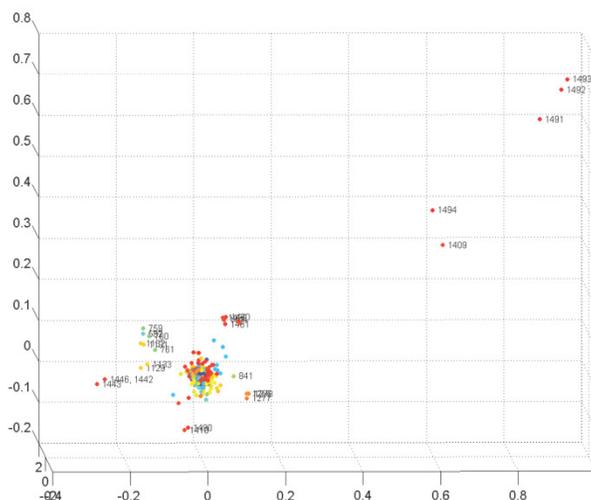
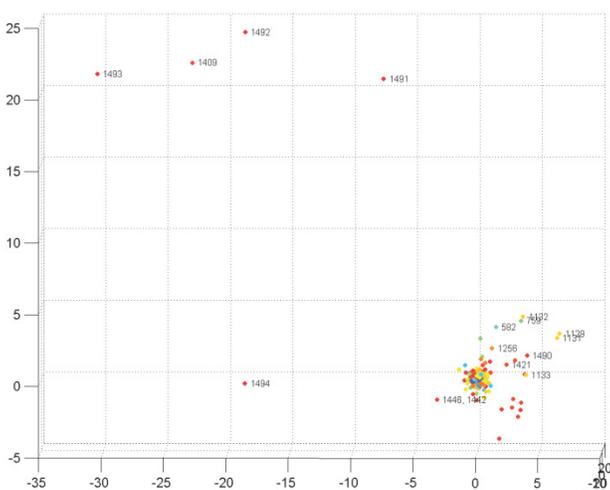


Figure 7: View of 3D scatterplot of translation vectors for files 1j5e and 1fjf. Units are Angstroms. Points are colored according to the distance to the center of the structure (See Figure 2). Points are labeled if the corresponding value for D_i^2 is greater than 25.



1j5e / 1jff		1j5e / 1jff	
Nucleotide	D_i^2	Nucleotide	D_i^2
A1493	401.90548	G1494	899.96391
A1492	381.09277	A1493	555.36737
G1491	321.21804	A1492	452.25642
A1446	238.64272	G1491	420.96963
G1442	238.64272	C1409	373.32662
G1443	208.41493	C1129	301.95947
A759	200.24573	C1132	164.08598
U582	189.03301	G1133	139.29572
G760	177.93137	A759	137.25964
C1409	161.26991	G1131	101.54401
G761	155.68751	U582	78.16921
G1494	155.22516	C1490	73.65867
G1131	87.02887	A1256	41.84380
C1132	86.51069	A1446	31.37154
G1133	50.13626	G1442	31.37154

(a) Translation Data

(b) Rotation Data

Figure 9: Largest values of D_i^2 , which indicates extreme data points.

As described above, we can use the scatterplots and D_i^2 values to find regions of conformational change. This would indicate which regions of the structure were modeled differently from one experiment to the next. However, we are also interested in the overall variability of the data points since that provides an indication of how similar the two structures are. The variance-covariance matrix for the translation data and rotation data (S^T and S^R , respectively) are

$$S^T = \begin{bmatrix} 0.0025 & -0.0026 & 0.0016 \\ -0.0026 & 0.0034 & -0.0020 \\ 0.0016 & -0.0020 & 0.0014 \end{bmatrix} \quad S^R = \begin{bmatrix} 1.6857 & 0.1350 & -1.1464 \\ 0.1350 & 0.4843 & 0.2426 \\ -1.1464 & 0.2426 & 1.4301 \end{bmatrix}$$

Thus the total sample variance for the translation vectors is 0.0073, and the total sample variance for the rotation vectors is 3.6001. While it is difficult to analyze these numbers in and of themselves, we can see that there is very little variability in the data points, implying the two structures are very similar. These variances will also be useful in the following sections where structures are compared that have been determined under other situations than here where the same lab crystallized the same molecule from the same organism under the same conditions. For example in the next section, we again compare two structures of the same molecule from the same organism; however, they have been crystallized by different research labs. We expect there to be greater variability in the data points than in that case since there is additional variability in the modeling processes.

3.2 Same Molecule, No Ligands Bound, Different Labs Analysis

We again consider two crystal structures of the same molecule of the same organism, but where the crystal structures have been determined independently by different research groups. We again consider two crystal structures of 16S *T.th.*, with no ligands or other complexes bound to the molecule, as before. We use the 1j5e crystal structure again, but instead of using 1jff which was determined by the same group that determined 1j5e, we use 2zm6 which was solved more recently (2009) by an independent research lab (Kaminishi, T. et al., 2009).

The two variance-covariance matrices are

$$S^T = \begin{bmatrix} 0.3265 & 0.0298 & -0.0945 \\ 0.0298 & 0.3412 & -0.0473 \\ -0.0945 & -0.0473 & 0.2320 \end{bmatrix} \quad S^R = \begin{bmatrix} 11.1536 & -3.0655 & -2.5901 \\ -3.0655 & 9.2062 & 3.4793 \\ -2.5901 & 3.4793 & 17.4081 \end{bmatrix}$$

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 0.8997 and 37.7679, respectively, compared to 0.0073 and 3.6001 before.

The scatterplots are given in Figures 10 and 11. They also show that the data is much more spread out in this case. The extreme values are listed in Figure 12.

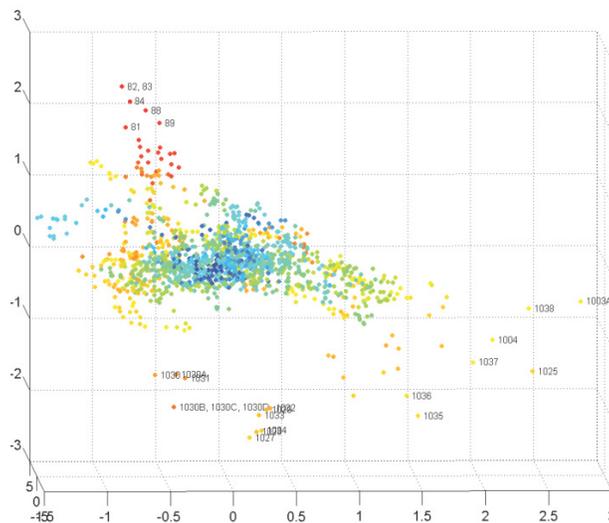


Figure 10: View of 3D scatterplot of translation vectors for files 1j5e and 2zm6. Units are Angstroms. Points are labeled if the corresponding value for D_i^2 is greater than 25.

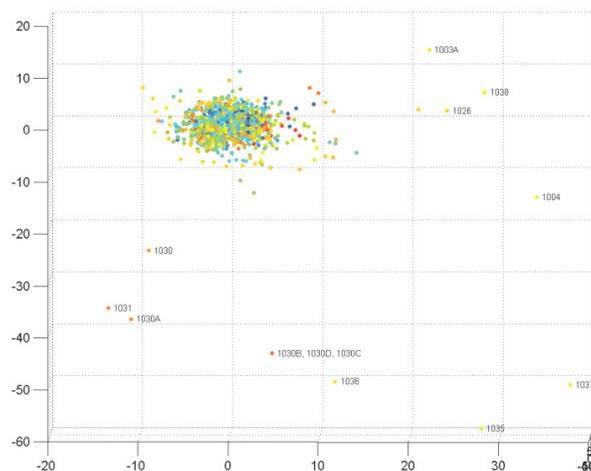


Figure 11: View of 3D scatterplot of rotation vectors for files 1j5e and 2zm6. Units are degrees. Points are labeled if the corresponding value for D_i^2 is greater than 25.

1j5e / 2zm6		1j5e / 2zm6	
Nucleotide	D_i^2	Nucleotide	D_i^2
C1030B	68.399	C1037	321.854
G1030C	68.399	A1035	238.314
A1030D	68.399	G1031	166.902
G1031	48.285	G1036	152.992
A1035	46.914	C1030	152.802
G1030A	44.151	A1004	139.285
G1032	41.235	C1038	128.619
C1030	38.476	G1030A	124.267
C1027	37.902	C1030B	123.947
G1034	36.988	A1030D	123.947
C1028	36.430	G1030C	123.947
G1003A	36.073	G1003A	78.237
U82	34.949	G1026	57.229
U83	34.949	U1025	50.408
C1029	33.926	C848	20.525
G1033	31.823	U84	20.191
G1036	31.740	U1086	19.956
A1004	30.889	C1006	19.869
U84	29.558	G1032	18.343
C1038	27.960	G1087	16.593

(a) Translation Data

(b) Rotation Data

Figure 12: Largest values of D_i^2 for 1j5e and 2zm6 data, indicating extreme data points.

4. All-Against-All Comparison of *T.th.* 16S Structures

As we've discussed, not all of the crystal structures in the PDB represent distinct organisms. In fact, there are nearly sixty PDB files that include a crystal structure of *T.th.* 16S rRNA. Most of these crystal structures include the 16S molecule in complex with mRNA, tRNA, one or more antibiotics, or some other type of ligand.

We have already shown that it is of interest to compare pairs of structures and analyze the similarities and differences. Also important is to do an all-against-all comparison of the structures and group them according to the overall similarities of their structures. As a measure of similarity between two structures, we can use the total sample variance of the translation or rotation vectors, as described above.

Here we focus on the translation vectors to analyze a subset of the *T.th.* 16S rRNA structures. The PDB files were selected that had identical base sequences to the 1j5e crystal structure in the first 1500 nucleotides, so that it was not necessary to determine an alignment. This procedure yielded 28 PDB files (including 1j5e). PDB file 2zm6 is also equivalent when disregarding an extra nucleotide at the beginning of the sequence, so that structure is considered as well, giving a set of 29 PDB files.

Figure 13 displays the mutual total sample variances among the 29 crystal structures. The variances are represented by colors as indicated by the color bars. The figure illustrates which structures and groups are more similar and which are not.

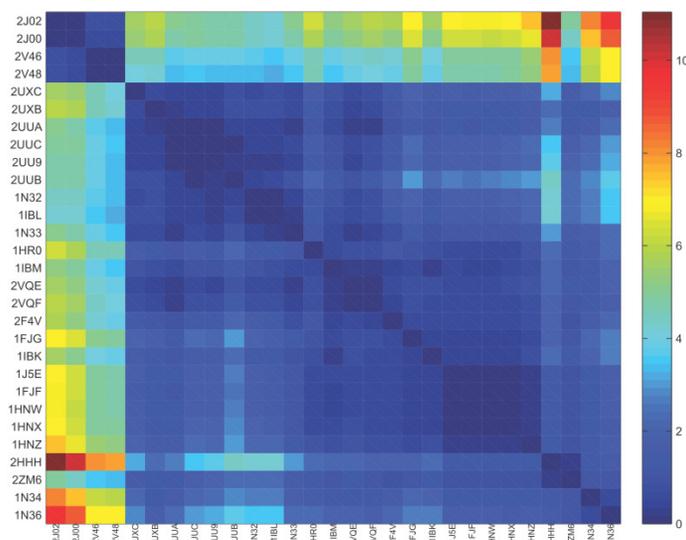


Figure 13: Mutual translation vector total sample variance matrix

A complete linkage cluster analysis was also performed on the structures. The dendrogram is provided in Figure 14. As expected, many structures crystallized by the same group are in the same cluster. For example, the crystal structures 2uua, 2uub, 2uuc, and 2uu9, which were all crystallized by the same group under the same conditions, are clustered together

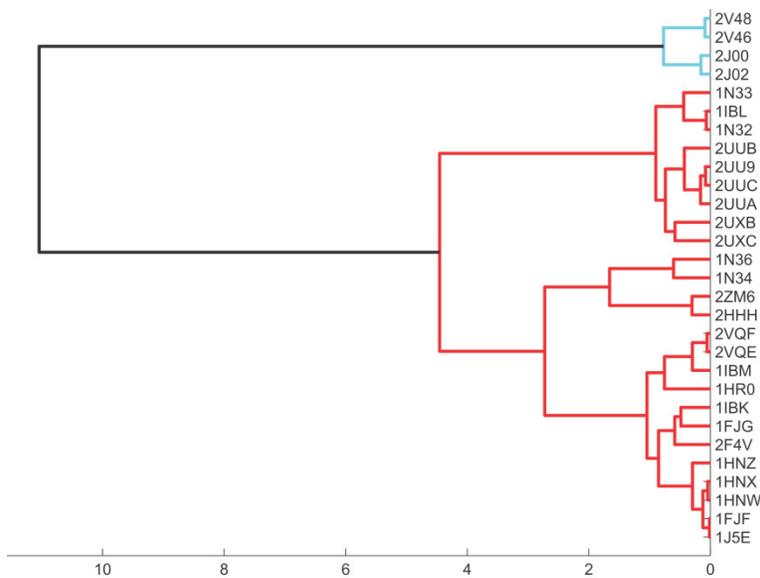


Figure 14: Dendrogram of 1j5e and the 28 structures with equivalent sequences

In some cases, common structures are related in that they are both in complex with the same type of molecule. For example, the dendrogram illustrates that the 1ibl and 1n32 crystal structures are closely related. While 1ibl and 1n32 were deposited in the PDB 18 months apart, they are both crystal structures of the 16S rRNA in complex with the antibiotic paromomycin.

4. 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) or SA Align (Rahrig, 2012). 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

- Barnett, Vic and Lewis, Toby (1994) *Outliers in Statistical Data*. Wiley & Sons, Chichester, New York, 3rd edition.
- Berman, H.M., Battistuz, T., et al. (2002) The protein data bank. *Acta crystallographica. Section D, Biological crystallography*, 58 (Pt 6 No 1):899–907.
- Berman, H.M., Olson, W.K., et al. (1992) The nucleic acid database. a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophysical journal*, 63(3):751–759, Sep 1992.
- Horn, Berthold K.P. and Hilden, H.M. and Shariar Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America*, 5(7):1127–1135, 1988.
- Kaminishi, T., et al. Crystal structure of the *Thermus thermophilus* 30S ribosomal subunit. Submitted for publication.
- Ogle, J.M., et al. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, 111(5):721–732, Nov 27 2002.
- Rahrig, R. (2012). Using Simulated Annealing for RNA 3D Structural Comparisons. *JSM Proceedings*.
- Rahrig, R., Zirbel, C.L., Leontis, N.B. (2010) R3D Align: Global pairwise alignment of RNA 3D structures using local superpositions. *Bioinformatics*, 26, 2689-2697.
- Rencher, Alvin C. (2002) *Methods of Multivariate Analysis*. Wiley & Sons, New York, 2nd edition.
- Schuwirth, B.S., et al. (2005) Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science*, 310, 827-834.
- Weixlbaumer, A., et al. (2007) Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nature structural & molecular biology*, 14(6):498–502, Jun 2007.
- Wilks, S.S. (1963) Multivariate statistical outliers. *Sankhy: The Indian Journal of Statistics, Series A*, 25(4):407–426.
- Wimberly, B.T., et al. (2000) Structure of the 30S ribosomal subunit. *Nature*, 407, 327-339.
- Zhang, W., Dunkle, J.A., et al. (2009) Structures of the ribosome in intermediate states of ratcheting. *Science*, 325(5943):1014–1017.