Sindh Univ. Res. Jour. (Sci. Ser.) Vol.44 (2) 303-312 (2012)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Quantifying the Variability in Graphs of RNA Structures Determined by Nuclear Magnetic Resonance

Z.U.A. KHUHRO, F. N. MEMON, A. P. HARRISON*

Institute of Mathematics and Computer Science, University of Sindh, Jamshoro
Corresponding author, Zain-ul-Abdin Khuhro, email: kzainkhuhro@yahoo.com Ph. No.

Received 16th May 2012 and Revised 30th June 2012

Abstract: This paper presents the comparison of an ensemble of RNA structures from the same molecule. Such structures are determined by Nuclear Magnetic Resonance (NMR) method. The means and standard deviations of length, dot product angle and dihedral angle values among bases of ensemble of RNA structures determined by NMR are calculated. Results show significant structural differences in overlaps within ensembles of NMR RNA structures.

Keywords: RNA, deviations, tolerance values, overlaps.

1. INTRODUCTION

The RNA structures are determined and deposited into the Protein Data Bank (PDB) (Berman, 2006) in the form of three dimensional coordinates. These three dimensional coordinates make it possible to compare many RNA structures and detect structural differences (Grindley,1993 Firdaus,2008). The geometric variability seen in RNA structures limit what information can be obtained in structural overlaps. The tolerance values are used with mean and standard deviations which limit whether two bases can be considered geometrically equivalent or not in the two structures. It is therefore important to establish secure choice of tolerance values. In this paper multiple models of same RNA structure determined by NMR method are studied so as to asses appropriate tolerances.

Ensemble of models

The determination of RNA structures by NMR method is fundamentally of a different nature to that using the X-ray method (Deigan, 2009). The major steps involved in NMR structures are the derivation and calculation protocols, e.g., nuclear overhauser effect (nOe) and residual dipolar couplings (RDC), (Guentert, 1998 Guantert, 2003). It is believed that X-ray crystallography perceive more accurate representation of structures than the NMR method because of a lack of clear quality assessment of NMR structures (Ddoreleijers, 1999). However, the NMR spectroscopy determines the structures in solution which is biologically more relevant than the X-ray crystallography which reveals structures in crystals (Egli, 2004).

This paper presents a method for detecting significant structural differences in overlaps seen between an ensemble of NMR RNA structures (**Fig.1),** which are based on the geometric relations.

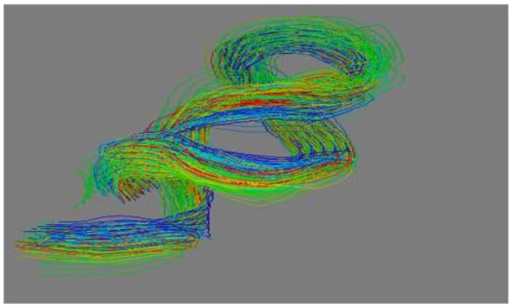


Fig. 1: Figure of an ensemble of RNA Structures of PDB code 1JO7 determined by NMR method. The RNA structure contains 32 models which are illustrated with different colours in the figure. The figure is created in RASMOL (Sayle and Bissell, 1992) software that can be used to view molecular structures in 3-D perspective.

2. METHOD

The approach described in this paper uses the conformational lengths and angles which are formed using cartesian coordinates of backbone atoms from an ensemble of RNA structures. Vectors between two atoms are selected from each base of every model. The geometric relations between vector pairs are measured. The mean and the standard deviation of the distances and angles between vectors of atoms (Khuhro, 2009) have been calculated along with the maximum and minimum distances and angles.

---

*Departments of Mathematical Sciences and Biological Sciences, University of Essex, UK

The following subsections describe the method and calculation of quantities.

Cartesian coordinates of NMR RNA structure

Step-I: Select a vector between pairs of atoms from each base of every model. Thus a RNA structure of different models which consists of different bases is converted into geometric parameters between pairs of vectors, i.e., length τ, dot product angle θ, and dihedral angle φ (Khuhro, 2009).

Step-II: Calculate the mean and the standard deviation of τ, θ, and φ values in each RNA structure.

Step-III: The maximum and minimum values of τ, θ, and φ are selected for particular bases in different models.

Measure of dispersion of NMR RNA data

A RNA structure determined by NMR method contains a number of bases and a number of different models. Each model is converted into a graph which describes the τ, θ and φ parameters between pairs of bases.

For a given pair of bases in all the models, the mean and standard deviation of τ, θ and φ are calculated. For example, if an ensemble of RNA structures consists of 3 models and a particular pair of base 1 and base 5 is given then the arithmetic mean and the standard deviation of τ, θ and φ are measured using length, dot product angle and dihedral angle values of the base 1 and base 5 in all the models as follows:

$$\bar{\tau} = \sum_{i=1}^{Models} \tau_i \qquad (1.1)$$

$$\sigma_\tau = \sqrt{\frac{\sum_{i=1}^{Models} (\tau_i - \bar{\tau})^2}{Models}} \qquad (1.2)$$

$$\bar{\theta} = \frac{\sum_{i=1}^{Models} \tan^{-1}\left(\frac{\sin(\theta_i)}{\cos(\theta_i)}\right)}{Models} \qquad (1.3)$$

$$\sigma_\theta = \sqrt{\frac{\sum_{i=1}^{Models} (\theta_i - \bar{\theta})^2}{Models}} \qquad (1.4)$$

$$\bar{\phi} = \frac{\sum_{i=1}^{Models} \tan^{-1}\left(\frac{\sin(\phi_i)}{\cos(\phi_i)}\right)}{Models} \qquad (1.5)$$

$$\sigma_\phi = \sqrt{\frac{\sum_{i=1}^{Models} (\phi_i - \bar{\phi})^2}{Models}} \qquad (1.6)$$

The selection of maximum and minimum values of τ, θ, and φ is described as follows:

1. Convert the NMR RNA structure into a graph by selecting O5´and C5´ toms.
2. Select the maximum and the minimum values of length, dot product angle and dihedral angle values for pair of bases.
3. Calculate the difference of maximum and minimum values.
4. Of all possible differences, select the highest difference of τ, θ and φ for a particular base pair as is shown in the **(Table 1)**.

2. RESULTS

In this section some of the RNA structures determined by NMR are examined. The selection of 10 particular RNA molecules in Table 1 belongs to different RNA structures which will help to understand the quantification of RNA structures variability. Though there are many other examples of RNA structures but in order to show an example just some of the structures are selected for the purpose to see the variability in different RNA structures. The selected RNA structures present in Table 1 have different number of nucleotides and different number of models. As the average number of models is 20, therefore, in Table 1, the models of size 10, 20 and 32 are selected. For example RNA structure PDB code 2JWV, PDB code 2K64 and PDB code 1JSE has 10, 20 and 32 models respectively.

The selected RNA structures are of different types which are indicated in Table 1. In addition, these RNAs are not ligand bound. Even if they could be ligand bound the method can compare the models of RNA structures determined by NMR because the method works on three dimensional coordinates of molecules.

In this section, vectors are constructed between the O5´ and C5´ atoms from the nucleotides of each RNA structures.

Difference of maximum and minimum values in bases

The aim is to find overlaps in the RNA structures by looking at the differences in lengths, dot product angle and dihedral angle values. As it is not possible to show all the base pairs in the table, therefore, only some of the base pairs are given in Table 1.

The maximum and minimum differences suggest how discrepant the geometry between bases in different models is.

For understanding the variability of RNA structures, determined by NMR method, the tolerance values which is an arbitrary selection are proposed for length τ, dot product angle θ and dihedral angles φ values such as 2 Å, 40 degrees, and 40 degrees respectively which is a random selection in order to find the optimal overlapping in RNA structures. It is observed that in some of the RNA structures in Table 1, the maximum difference in length, dot product angle and dihedral angle is within the proposed tolerance values. In Table 1, PDB code 2JXQ shows the maximum difference in length, dot product angle and dihedral angle which is 1.35 Å, 9.12 degrees, and 10.02 degrees respectively, which are all smaller than the proposed tolerance values. However, there are some RNA structures whose parameter values are more than the proposed threshold values. For example, in Table 1, for PDB code 1JO7 the maximum difference in length, dot product angle and dihedral angle is 14.83 Å, 75.10 degrees, and 59.07 degrees, respectively **(Fig. 2a).**

In order to understand further, all the pairs of bases are selected and their differences between maximum and minimum values are examined. The NMR RNA structure PDB code 1JO7 **(Fig. 3a)** is analysed where τ between base 1_A and base 2_G in model 1 is 6.62 Å and τ between base 1_A and base 2_G in model 19 is 5.89 Å. The maximum difference between the two τ is 0.73 Å smaller than the threshold of 2 Å. However, in the same RNA molecule PDB code 1JO7 the τ between base 1_A and base 18_G in model 9 is 45.464 Å and the τ between base 1_A and base 18_G in model 18 is 32.121 Å. Thus the maximum difference between the two τ is 13.34 Å greater than the threshold value of 2 Å. This τ difference is almost the same as the maximum difference between 7_A and 19_G in Table 1. **(Fig. 2a)** illustrates the two nucleotides 7_A and 19_G which are created in Rasmol (Sayle and Bissell, 1992).

Similarly, NMR RNA structure PDB code 2JTP **(Fig. 3b)** is analysed where τ between base 1_A and base 2_G in model 14 is 6.09 Å and τ between base 1_A and base 2_G in model 16 is 5.79 Å. The maximum difference between the two τ is 0.30 Å, smaller than the threshold value of 2 Å. However, in the same RNA molecule PDB code 2JTP the τ between base 1_A and base 17_C in model 16 is 49.05 Å and the τ between base 1_A and base 17_C in model 2 is 40.52 Å. Thus the maximum difference between the two τ is 8.53 Å, greater than the threshold value of 2 Å.

Another NMR RNA structure PDB code 2K64 **(Fig. 3c)** is analysed where τ between base 1_G and base 2_G in model 12 is 6.47 Å and τ between base 1_G and base 2_G in model 15 is 5.83 Å. The maximum difference between the two τ is 0.63 Å, smaller than the threshold value of 2 Å. In the same RNA molecule PDB code 2K64 the τ between base 1_G and base 19_G in model 8 is 32.85 Å and the τ between base 1_G and base 19_G in model 14 is 30.30 Å. Thus the maximum difference between the two τ is 2.54 Å, greater than the threshold value of 2 Å **(Fig. 2b)**.

Thus the overall τ difference in PDB code 1JO7 and PDB code 2JTP is considerably greater than the 2 Å, the proposed minimum length tolerance value. Whereas the maximum τ difference in PDB code 2K64 is only marginally greater than the 2 Å threshold value. In general, ensemble of RNA structures shows a difference typically < 2 Å among their bases.

Relationship of bases among models

Here the length τ is calculated for all the bases in an ensemble of RNA structures. To avoid complications, just 6 bases are selected for illustration. In Fig.4a, b and c,x-axis represents the base number whereas y-axis shows τ (in Å) of the selected base with other bases. A different colour scheme has been used for each model in the **(Fig. 4a, band c).**

It is important to note that there are some pairs of bases which show a significant difference in their geometry. For example, in PDB code 1JO7 the maximum difference between the bases amongst the models is about 14.83 Å, as given in Table 1 and which is also shown in **(Fig. 4a)** and (Fig. 2a).

Similarly for PDB code 2JTP **(Fig.4b)** the maximum difference between the bases amongst the models is about 11.71, as is given in Table 1.

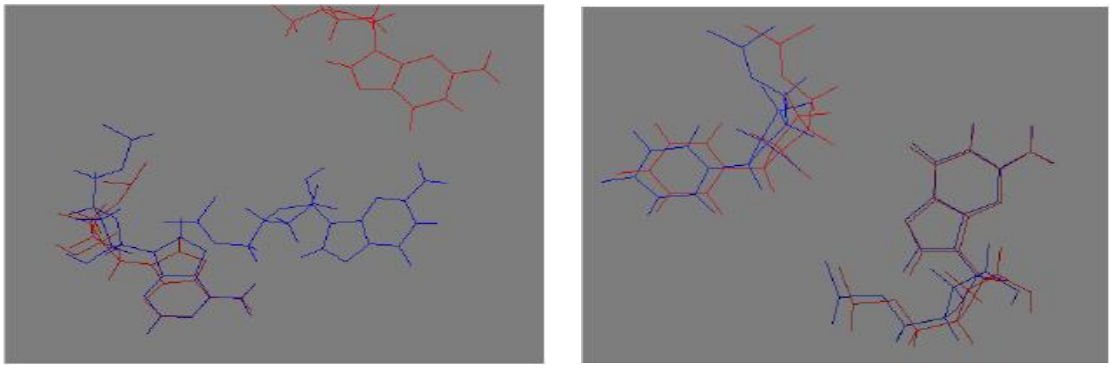
In the **(Fig. 4c)** for PDB code 2K64 the maximum difference between the bases amongst the models is about 2.41 Å, as suggested in **(Table 1)** and (Fig 2b).

Mean and standard deviation of τ in ensemble of RNA structures

Some statistical calculations are carried out for some RNA structures determined by NMR method and shown in **(Fig. 5a),b),c),d),e) and f).** In Fig. 5a, b, c, d, e and f) the mean and the standard deviation for each base in every model are shown, where the x-axis represents the standard deviation and y-axis represents the mean. Different colours and symbols are chosen in order to see each base clearly.

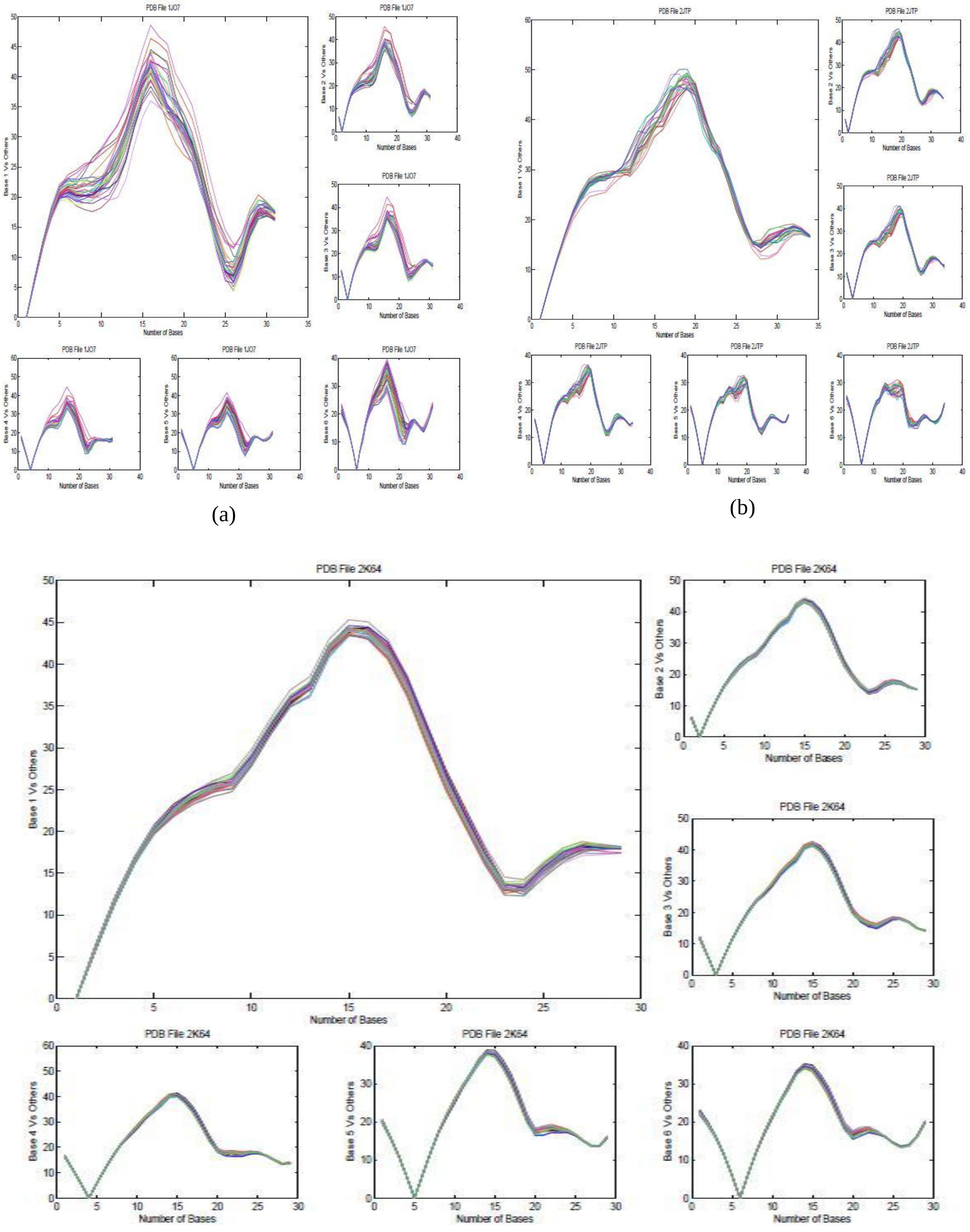
Table 1: Maximum and minimum values with their corresponding differences for length, dot product angle and dihedral angle based on O5′ and C5′ atoms among the bases have been calculated in different models where the first column represents PDB code, second column shows the RNA type, third and fourth columns show the bases with their location. Fifth and seventh columns show the maximum and minimum values for lengths, dot product angle and dihedral angle values for each structure. The sixth and eighth columns represent model I and model II respectively. Column nine shows overall difference between maximum and minimum values whereas last column shows the parameter for that particular base.

PDB ID:	Type of RNA	Base-1	Base-2	Maximum Value	Model-I	Minimum Value	Model-II	Difference	Parameter
1JO7	Viral RNA	7-A	19-G	28.869	13	14.034	28	14.835	τ
		7-A	19-G	142.868	12	67.763	28	75.105	θ
		7-A	19-G	141.420	19	82.341	20	59.079	φ
2JSE	rRNA	3-A	14-G	19.244	8	14.409	9	4.835	τ
		3-A	14-G	150.416	4	125.887	6	24.529	θ
		3-A	14-G	170.130	5	142.747	6	27.383	φ
2JTP	RNA stem-loop	17-C	34-C	50.723	2	39.011	16	11.712	τ
		17-C	34-C	163.635	15	80.257	17	83.378	θ
		17-C	34-C	162.000	17	106.274	19	55.726	φ
2JWV	RNA Aptamer	3-U	19-G	27.608	8	19.626	1	7.982	τ
		3-U	19-G	118.596	1	94.775	9	23.814	θ
		3-U	19-G	168.042	4	154.486	6	13.556	φ
2JXQ	RNA Duplex	1-G	9-C	25.001	1	23.650	9	1.351	τ
		1-G	9-C	79.474	7	70.346	8	9.128	θ
		1-G	9-C	72.663	8	62.640	9	10.023	φ
2K3Z	RNA Duplex	1-G	7-C	24.349	1	21.778	5	2.571	τ
		1-G	7-C	177.283	2	138.445	4	38.838	θ
		1-G	7-C	179.464	2	134.478	3	44.986	φ
2K64	Group-II introns	2-G	18-U	36.211	2	33.797	4	2.414	τ
		2-G	18-U	32.555	15	12.444	16	20.111	θ
		2-G	18-U	33.242	15	10.251	3	22.991	φ
2K95	Telomerase RNA	1-G	44-C	41.829	3	33.740	11	8.089	τ
		1-G	44-C	86.415	1	20.914	14	65.501	θ
		1-G	44-C	84.755	1	19.903	14	64.852	φ
2PN9	TAR RNA	1-G	10-G	18.653	5	15.257	9	3.396	τ
		1-G	10-G	111.411	7	87.499	10	23.912	θ
		1-G	10-G	147.456	5	90.678	8	56.778	φ
2RLU	mRNA	4-U	12-U	21.978	5	18.126	9	3.825	τ
		4-U	12-U	169.562	5	74.645	9	94.917	θ
		4-U	12-U	172.282	5	119.201	8	53.081	φ



(a) (b)

(Fig. 2a) NMR RNA structure PDB code 1JO7 where two nucleotides represent 7_A bases from models 13 and 28 respectively are on top of each other whereas the other two nucleotides represent 19_G bases from models 13 and 28 respectively are far from each other. The distance between 7_A and 19_G nucleotides is 14.83 Å which are also shown in Table 1. (b) NMR RNA structure PDB code 2K64 where two nucleotides represent 2_G bases from models 2 and 4 respectively are on top of each other whereas the other two nucleotides represent 18_U bases from models 2 and 4 respectively show little deviation between them. The distance between 2_G and 18_U nucleotides is 2.41 Å which are also shown in Table 1. Figures are created by Rasmol (Sayle and Bissell, 1992).

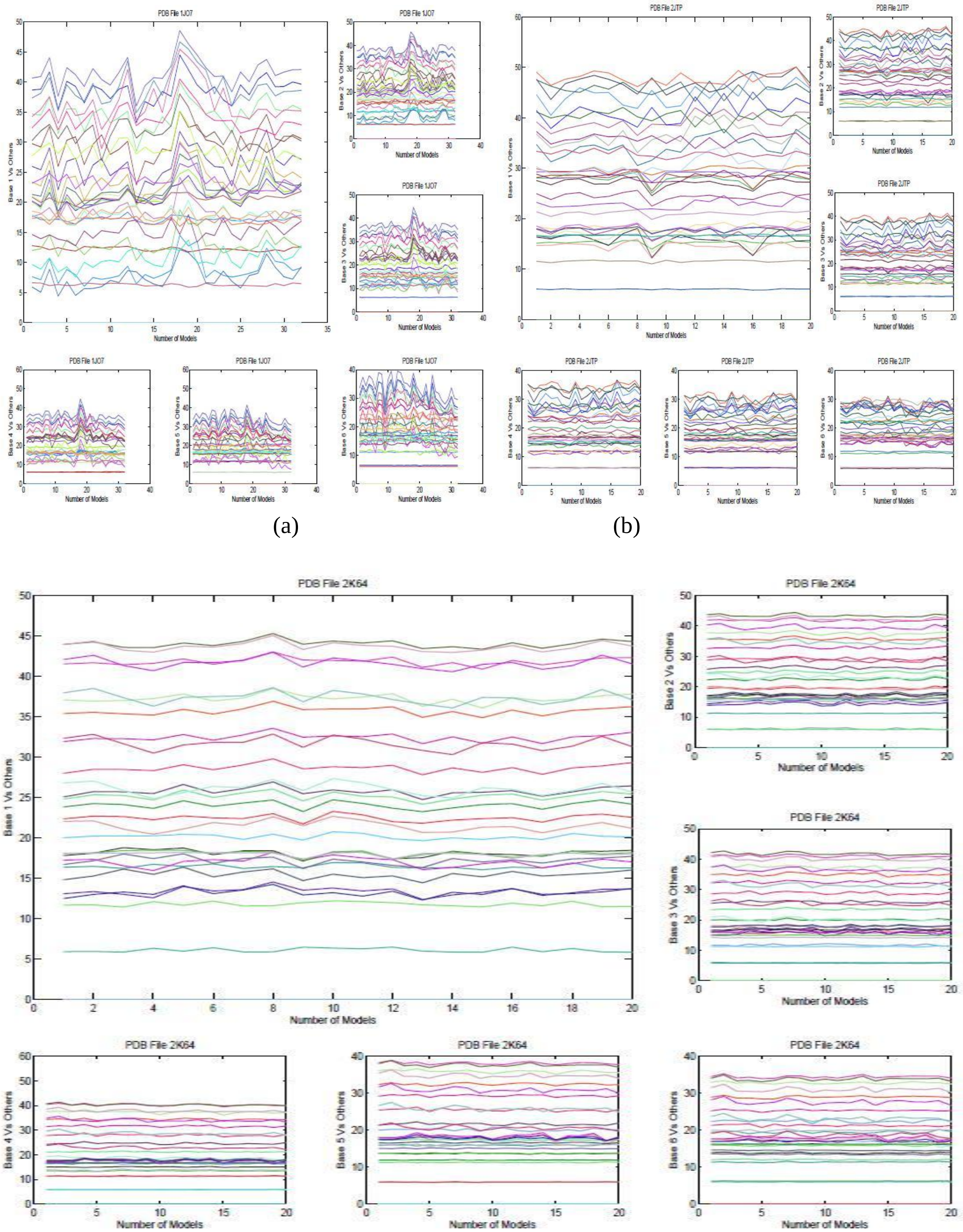


(a)

(b)

(c)

Fig. 3: X-axis represents number of bases and Y-axis represents Base(n) Vs all others bases. (a) PDB code 1JO7 contains 31 bases in 32 models. (b) PDB code 2JTP contains 34 bases in 20 models. (c) PDB code 2K64 contains 29 bases in 20 models. In each figure x-axis represents the model number in the RNA structure whereas y-axis shows τ (in $\mathring{A}$) among the selected base and the other bases. Different colours help to identify each base. One colour represents same base pair in all the figures presented. This colour scheme makes easy to identify the bases and their relationship with their corresponding models. In total, 6 particular bases are selected which are illustrated in six different subfigures and compared with all other bases in all models. The results show a significant difference in the geometry of base pairs.



(a) (b)

(c)

Fig. 4: X-axis represents number of models and Y-axis represents base (n) vs all other bases. (a) PDB file 1JO7 contains 31 bases in 32 models. (b) PDB file 2JTP contains 34 bases in 20 models. (c) PDB file 2K64 contains 29 bases in 20 models. In the graph x-axis represents the base number in the RNA structure whereas y-axis shows the difference in τ (in Å) among the selected bases and the other bases. Different colours are used in order to see the curves for each model. Colours help to identify the curves. One colour represents same model in all the figures presented. This makes easy to identify the bases and their relationship with the other models. In total, 6 particular bases are selected and presented in six different subfigures. In each subfigure different peaks are observed for each model horizontally and the differences of bases among the models can be seen in each Figure.

Six NMR RNA structures PDB code 2JTP, PDB code 1HWQ, PDB code 1IKD, PDB code 2YJM, PDB code 1JO7 and PDB code 2K63, are plotted, where every RNA structure has a different number of bases and a different number of models.

In order to examine the bases' relationship with each other in each NMR RNA structure, the measure of degree of spread among particular bases is presented with the tendency of individual base to vary from the mean values. This is helpful in comparing the bases, which may have the same mean but a different range. Some of the subfigures show that the standard deviation increases that means bases are far away from each other, for example the PDB code 1HWQ in Figure 5(a). Whereas some of the subfigures show that bases lie close to the mean as the standard deviation is small such as PDB code 2K64 in Figure 5(e). Figure 5((a), (b), (c), (d), (e) and (f)) show that for some of the structures, the NMR models closely resemble each other.

Difference of maximum and minimum values for different choice of atom pairs within bases

In Table 1, the O5′ and C5′ atoms are used for finding the τ. In this section, the approach has been extended to the O5′ and C2′ (Figure 6) atoms of RNA structures determined by the NMR method, as a way to estimate the accuracy of overlaps in structures. The aim is to find the overlaps in the RNA structures by looking at the differences in lengths, dot product angle and dihedral angle values when pair of different atoms and their vectors is made.

In Table 2, the RNA structures are examined for length, dot product angle and dihedral angle values with their corresponding differences. As it is not possible to show all the base pairs in the table, therefore, only some of the base pairs are given in Table 2.

The selection of maximum and minimum values of τ, θ, and φ is described as follows:

1. Convert the NMR RNA structure into a graph by selecting O5′ and C2′ atoms.
2. Select the maximum and the minimum values of length, dot product angle and dihedral angle values for all pairs of bases.

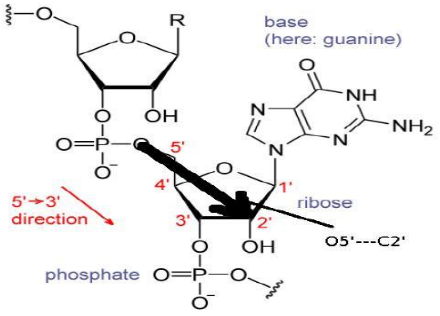


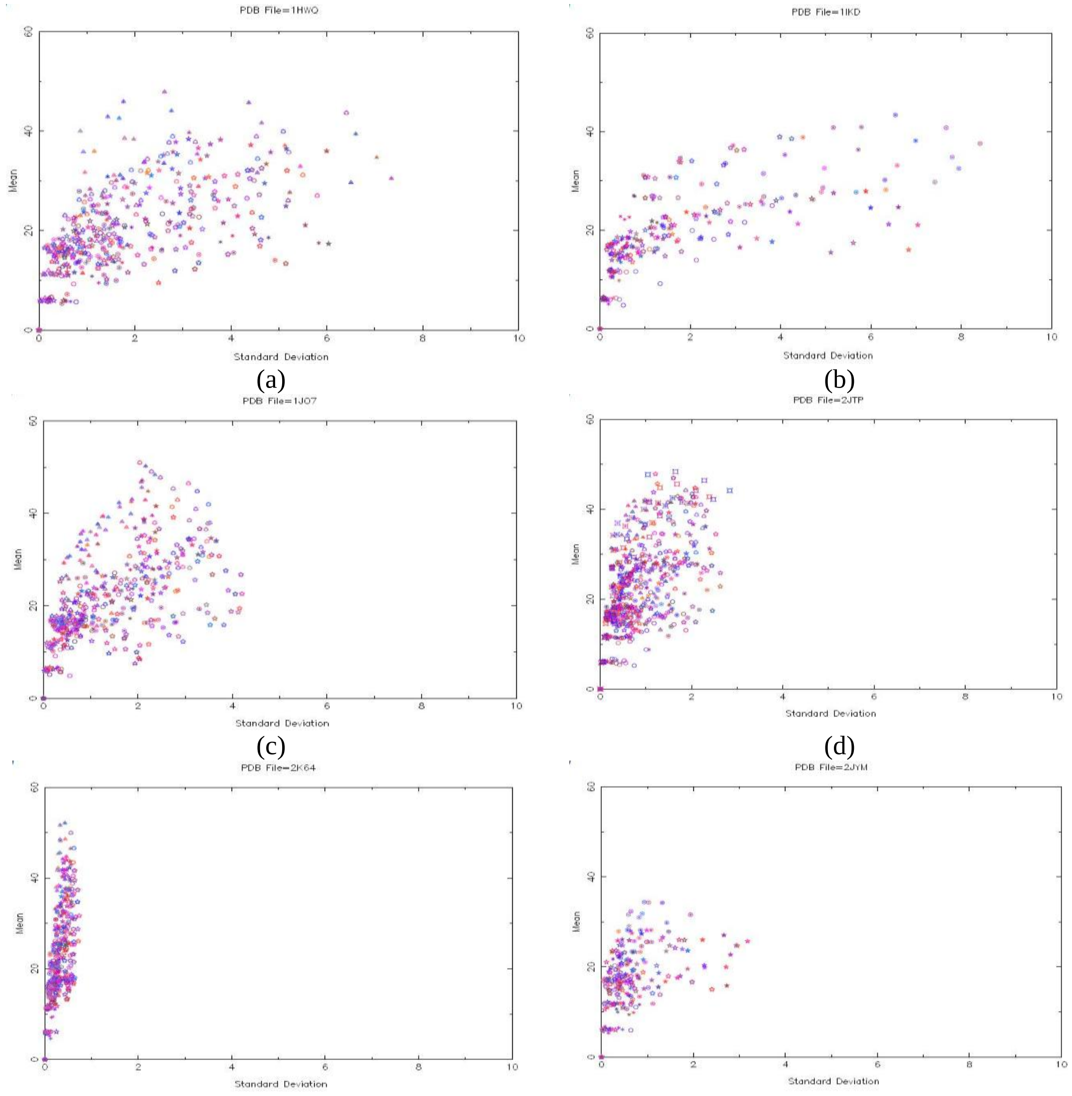
Fig. 6: Shows the chemical structure of RNA and describes the directionality of backbone with the positions of pentose sugar, phosphate groups and nitrogenous bases. It also shows the direction vector between the O5′ and C2′ atoms. (http://wiki.biomine.skelleftea.se/wiki/index.php/RNA, August, 2011).

3. Calculate the difference of maximum and minimum values.
4. Of all possible differences, select the highest value of τ, θ and φ for a particular base pairs as is shown in Table 2.

The whole procedure suggests the differences of bases in different models. It is also observed that in some of the RNA structures in Table 2, the difference in length, dot product angle and dihedral angles is within the proposed values which are suggested for τ, θ and φ (2 Å, 40 degrees, and 40 degrees). For example, in Table 2, PDB code 2JXQ the bases 1_G and 7_C show τ difference as 1.05 Å, dot product angle as 8.62 degrees and dihedral angle as 12.85 degree, which is smaller than the proposed tolerance value. However, there are some RNA structures whose parameter values are more than the proposed threshold values. These include PDB code 2RLU which has the values shown in Table 2 for the maximum τ, dot product angle, and dihedral angle of 2.68 Å, 34.81 degrees, and 87.48 degrees, respectively.

Comparisons of results of Tables 1 and 2

From the results of Table 1 and Table 2, it is observed that the choice of different atom set makes a little difference in the values of maximum differences in length, dot product angle and dihedral angle parameters. For some RNA structures, for example PDB code 2K64, the values in Table 1 for length is 2.41 Å, dot product



(a) (b) (c) (d) (e) (f)

Fig. 5: Y-axis shows the mean and x-axis shows the standard deviation of (a) PDB code 1HWQ (b) PDB code 1IKD (c) PDB code 1JO7 (d) PDB code 2JTP (e) PDB code 2K64 (f) PDB code 2JYM. These quantities have been calculated for the particular lengths between atoms O5′ and C5′ in all possible bases in all different models. Different symbols are adopted where every symbol represents different bases in all models.

angle is 20.11 degrees, and dihedral angle is 22.99 degrees where the O5′ and C5′ are used,. Whereas the value for the same RNA molecule in Table 2, where the O5′ and C2′ are used, for length is 2.29 Å, dot product angle is 13.05 degrees, and dihedral angle is 25.24 degrees.

In some of the RNA molecules, however, the difference in geometric parameters between bases is greater than the maximum values when O5′ and C2′ are used instead of O5′ and C5′ atoms. For example, in the NMR RNA structure PDB code 1JO7 (Figure 4(a)) the τ between base 1_A and base 2_G in model 1 is 6.59 and τ between base 1_A and base 2_G in model 28 is 5.45 Å. The maximum difference between the two τ is 1.14 Å whereas, as discussed in section 3, the maximum difference between 1_A and base 2_G is 0.73 Å. In the same RNA molecule PDB code 1JO7 the τ between base 1_A and base 18_G in model 18 is 44.59 Å and the τ between base 1_A and base 18_G in model 9 is 32.74 Å. Thus the maximum difference between the two τ is 11.84 Å when using O5′ and C2′ atoms whereas, as discussed in section 3, the maximum difference between 1_A and base 18_G is 13.345 Å when using O5′ and C5′ atoms.

Similarly, NMR RNA structure PDB code 2JTP (Figure 4(b)) is analysed where τ between base 1_G and base 2_G in model 18 is 5.76 Å and τ between base 1_G and base 2_G in model 16 is 5.20 Å. The maximum difference between the two τ is 0.55 Å whereas, as discussed in section 3, the maximum difference between 1_A and base 2_G is 0.30 Å. In the same RNA molecule PDB code 2JTP the τ between base 1_G and base 17_C in model 16 is 48.48 Å and the τ between base 1_G and base 17_C in model 2 is 41.00 Å. Thus the maximum difference between the two τ is 7.48 Å when using O5′ and C2′ whereas, as discussed in section 3, the maximum difference between 1_A and base 17_C is 8.54 Å when using O5′ and C5′.

Another NMR RNA structure PDB code 2K64 (Figure 4(c)) is analysed where τ between base 1_G and base 2_G in model 12 is 6.30 Å and τ between base 1_G and base 2_G in model 15 is 5.71 Å. The maximum difference between the two τ is 0.59 Å whereas, as discussed in section 3, the maximum difference between 1_G and base 2_G is 0.63 Å. In the same RNA molecule PDB code 2K64 the τ between base 1_G and base 19_G in model 2 is 31.00 Å and the τ between base 1_G and base 19_G in model 14 is 28.63 Å. Thus the maximum difference between the two τ is 2.36 Å when using O5′ and C2′ whereas, as discussed in section 3, the maximum difference between 1_G and base 19_G is 2.54 Å when using O5′ and C5′.

Thus the overall τ difference in PDB code 1JO7, PDB code 2JTP and PDB code 3K64, when O5′ and C2′ atoms are used instead of O5′ and C5′ atoms, does not make a big difference and the change is less than 2 Å threshold value, the proposed minimum length tolerance value.

4. DISCUSSION

Implicit in much of the preceding discussions is the assumption that if the bases would represent very little standard deviation in each model then the solution structures would closely resemble the representative model. However, this is not seen to be true in the presented results. It appears in Tables and Figures that some of the structures in an ensemble of RNA structures are not very close to their representative model. Thus, the parameter sets used to compute solution structure are very important. In addition, errors were made in the geometric interpretation of some of the NMR data that there might be some restraints which were used during computations (Rife, 1999). Considering the entire possibilities one can conclude that some solution structure's models require tolerance values more than initially proposed values for length, dot product angle and dihedral angle. Such differences for some of the RNA structures are also shown in Fig. 4a, b and c, Table 1 and **(Table 2)**.

The experiments were extended to the other backbone atoms, i.e., O5′ and C2′ of bases in every model. It is observed that in some RNA molecules the difference in distances between models is considerably greater than 2 Å threshold values whereas in some RNA structures the difference in distance between models is less than 2 Å threshold values. However, the use of O5′ and C2′ atoms instead of O5′ and C5′ atoms, to assign vectors to each nucleotide, do not make a significant difference in structural overlapping.

The proposed threshold values for geometric parameters are 2 Å, 40 degrees, and 40 degrees, for length, dot product angle and dihedral angles respectively. It appears that optimal geometric arrangement should be adopted where the geometric parameters should exhibit no/little differences in τ, θ and φ. This will help to find the overlapping differences in the compared RNA structures determined by NMR and will provide more conformational information. The whole process of comparison of an ensemble of RNA structure will help biologists to locate the structural differences for further analysis.

5. CONCLUSION

In this paper, a quantitative method has been developed which examines the ensemble of RNA structures. It is typically believed that the first model in each NMR RNA structure is a representative model, therefore, how to describe the optimal geometric arrangement which describes the relationship of representative model with the other models in RNA structures, is of great interest.

The mean and standard deviation of length, dot product angle and dihedral angle values is calculated for the bases in every model in order to see the closeness of each base in every model. In all the experiments it is observed that in some of the structures the bases difference is greater than 2 Å and in other structure the difference is less than 2 Å that is a proposed tolerance value for τ.

In future a strategy can be proposed to suggest the suitable tolerance values for finding optimal overlapping among the RNA structure of an ensemble.

Table 2: Maximum and minimum values with their corresponding differences for length, dot product angle and dihedral angle based on O5´ and C2´ atoms among the bases have been calculated in different models where the first column represents PDB code, second and third columns show the bases with their location. Fourth and sixth columns show the maximum and minimum values for lengths, dot product angle and dihedral angle values for each structure. The fifth and seventh columns represent model I and model II respectively. Column eight shows overall difference between maximum and minimum values whereas last column shows the parameter for that particular base.

PDB ID:	Type of RNA	Base-1	Base-2	Maximum Value	Model-I	Minimum Value	Model-II	Difference	Parameter
1JO7	Viral RNA	7-A	19-G	27.358	13	13.056	28	14.302	τ
		7-A	19-G	138.063	18	94.774	19	43.289	θ
		7-A	19-G	137.426	17	97.573	19	39.853	φ
2JSE	rRNA	3-A	14-G	20.271	8	15.926	9	4.345	τ
		3-A	14-G	91.019	5	75.505	7	15.514	θ
		3-A	14-G	70.275	5	51.768	6	18.507	φ
2JTP	RNA stem-loop	17-C	34-C	51.631	2	41.310	16	10.321	τ
		17-C	34-C	141.785	13	95.174	17	46.611	θ
		17-C	34-C	140.773	9	86.823	13	53.950	φ
2JWV	RNA Aptamer	3-U	19-G	27.755	1	20.442	8	7.313	τ
		3-U	19-G	143.869	4	131.849	9	12.020	θ
		3-U	19-G	159.789	4	142.262	6	17.527	φ
2JXQ	RNA Duplex	1-G	9-C	34.232	1	23.176	9	1.056	τ
		1-G	9-C	84.056	1	75.432	7	8.624	θ
		1-G	9-C	101.526	1	88.673	7	12.853	φ
2K3Z	RNA Duplex	1-G	7-C	22.692	1	21.114	5	1.578	τ
		1-G	7-C	173.709	2	75.891	3	43.754	θ
		1-G	7-C	178.195	2	146.625	10	31.570	φ
2K64	Group-II introns	2-G	18-U	34.468	2	32.176	14	2.292	τ
		2-G	18-U	108.218	5	95.163	13	13.055	θ
		2-G	18-U	131.017	4	105.771	15	25.246	φ
2K95	Telomerase RNA	1-G	44-C	40.449	11	34.454	19	5.995	τ
		1-G	44-C	61.605	17	21.331	18	40.274	θ
		1-G	44-C	53.429	16	1.688	17	51.741	φ
2PN9	TAR RNA	1-G	10-G	19.809	5	16.765	10	3.044	τ
		1-G	10-G	119.645	1	75.891	9	43.754	θ
		1-G	10-G	113.019	6	55.385	10	57.634	φ
2RLU	mRNA	4-U	12-U	22.157	4	19.474	9	2.683	τ
		4-U	12-U	89.451	6	54.637	9	34.814	θ
		4-U	12-U	119.034	6	31.545	9	87.489	φ

REFERENCES:

Berman, H., K. Henrick, H. Nakamura, and J. Markley, (2006) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*.

Deigan, K.E., T.W. Li, D.H. Mathews, and K.M. Weeks, (2009) Accurate SHAPE-directed RNA structure determination, *Proceedings of the National Academy of Sciences*, **(106):** 1, 97-102,

Doreleijers, J., M. Raves, T. Rullmann, and R. Kaptein, (1999) Completeness of NOEs in protein structures: A statistical analysis of NMR data. *Journal of Biomolecular NMR*, **(14):** 123–132.

Egli, M. (2004) Nucleic acid crystallography: current progress. *Current Opinion in Chemical Biology*, **(8):** 580-591.

Firdaus-Raih, M. (2008) Computational Approaches for Investigating Three-Dimensional Structures of RNA. *Malaysian Journal of Biochemistry and Molecular Biology*, **(16):** 1–9.

Grindley, H., P. Artymiuk, D. Rice and P. Willet, (1993) Identification of tertiary structure resemblance in proteins using a maximal common sub-graph isomorphism algorithm. *Journal of Molecular Biology*, **(229):** 707–721.

Guentert, P. (1998) Structure calculation of biological macromolecules from NMR data. *Quarterly Reviews of Biophysics*, **(31):** 145–237.

Guntert, P. (2003) Automated NMR protein structure calculation. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **(43):** 105–125.

Khuhro, Z.U.A, F. N. Memon, A. P. Harrison (2009). Ribonucleic acid tertiary structure comparison using graph theory. Conference proceedings, CIB, 2009, Hangzhou, China.

Klosterman, P., M. Tamura, S. Holbrook and S. Brenner, (2002) SCOR: a structural classification of RNA database. *Nucleic Acids Research*, **(30):** 392.

Rife, J., S. Stallings, C. Correll, A. Dallas, T. Steitz, and P. Moore, (1999) Comparison of the crystal and solution structures of two RNA oligonucleotides. *Biophysical Journal*, **(76):** 65–75.

Sayle, R. and A. Bissell, (1992) RasMol: A program for fast, realistic rendering of molecular structures with shadows. *In Proceedings of the 10th Eurographics UK*, vol. **(92):**.