

Internet **Electronic** Journal of **Molecular Design**

December 2006, Volume 5, Number 12, Pages 570–584

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Lemont B. Kier on the occasion of the 75th birthday

Structural Characteristics of the Nucleotides Pairing in RNA: Principal Component Analysis

Huai Cao,¹ Weixian Cheng,¹ Taohong Li,¹ Xulin Pan,¹ and Ciquan Liu¹

¹ Modern Biological Research Center, Yunnan University, Kunming 650091, The People's
Republic of China

Received: December 10, 2005; Revised: June 1, 2006; Accepted: June 23, 2006; Published: December 21, 2006

Citation of the article:

H. Cao, W. Cheng, T. Li, X. Pan, and C. Liu, Structural Characteristics of the Nucleotides Pairing in RNA: Principal Component Analysis, *Internet Electron. J. Mol. Des.* **2006**, 5, 570–584, <http://www.biochempress.com>.

Structural Characteristics of the Nucleotides Pairing in RNA: Principal Component Analysis[#]

Huai Cao,^{1,*} Weixian Cheng,¹ Taohong Li,¹ Xulin Pan,¹ and Ciquan Liu¹

¹ Modern Biological Research Center, Yunnan University, Kunming 650091, The People's Republic of China

Received: December 10, 2005; Revised: June 1, 2006; Accepted: June 23, 2006; Published: December 21, 2006

Internet Electron. J. Mol. Des. 2006, 5 (12), 570–584

Abstract

Motivation. RNA secondary structure motifs are important to the biological function of an RNA molecule. Close attention has been paid to the pairing state of a nucleotide presented in a variety of motifs in many reported studies of RNA structure and folding. More and more work was crowded on identifying, classifying, and discovering motifs because it is necessary that the interactions between secondary structure motifs facilitate the higher order structure to come into being. However, the paired–unpaired state for a nucleotide should be alterable while a given functional higher order structure is resulted from the secondary structure motifs. Studying the conformational factors of affecting the state change of a motif and mastering the factors to arrange the type of a motif is undoubtedly a key problem to understand, find out, and further develop the biological activity of RNA in the gene expression.

Method. In this article we collected the structural data of 1320 nucleotides of 45 RNA molecules in the various motifs from the databases, and analyzed the structural characteristics of their torsion angles by using principal component analysis. We adopted three grouping methods to analyze the 45 RNA molecules: in original motif, in molecular type, and in new motif. The variables representing conformation feature are six backbone torsion angles α , β , γ , δ , ε , ζ and χ between the sugar and the base, and the distance between the two phosphorus atoms P. A pre–treatment of the data, the equal weight and mean–centering for each variable, was used. The biplot representation, in which the scores on the first two PCs together with the loadings of the original variables are depicted in the same plot, was a proper illustration for the PCA results.

Results. In the PC1–PC2 biplots, the conformation feature of the nucleotides didn't display over–particular with the variety of motifs, or the difference of molecules. Only the three of the eight parameters, *i.e.*, α , γ , and ζ , had the most loadings and a negative correlation between α and γ . By comparison, the rest parameters represented a little of contribution to PC1 and PC2. Considering the scores on PC1, we grouped the nucleotides into the two clusters: one gathering in the angle between the vectors α and γ , and another dispersing out the angle. Significantly, the former were the paired nucleotides, and the latter unpaired.

Conclusions. All results revealed that the conformational factors, α and ζ , especially α , play an irreplaceable role in determining the pairing tendency of a nucleotide. Nucleotides in RNAs may fall into two states according to the values of their torsion angles, α and ζ , or only α , paired and unpaired. In comparison with the previous studies on the characteristics of RNA secondary structure motifs, this work shows that, by dominating a nucleotide having a range of α values, we may transform motifs to design desired RNA molecules.

Keywords. RNA secondary structure motifs; paired/unpaired nucleotides; torsion angle characteristics; Principal Component Analysis.

[#] Dedicated to Professor Lemont B. Kier on the occasion of the 75th birthday.

* Correspondence author; phone: 86–871–5033496; fax: 86–871–5036373; E–mail: caohuai@ynu.edu.cn.

Abbreviations and notations

HM, <i>Haloarcula marismortui</i>	PCA, principal component analysis
LSU, large subunit	SAM, S-adenosyl methionine
PC, principal component	TPP, thiamin pyrophosphate

1 INTRODUCTION

With the development of x-ray diffraction, NMR spectrum and chemical synthesis, some complex and intricate RNA molecules, such as 16S, 23S rRNAs, their structures have been resolved [1,2], which rapidly increased the available molecular structural information in RNA databases, such as PDB [3], NDB [4], RNABASE [5] and PseudoBase [6]. As RNA can both encode genetic information and catalyze chemical reactions [7], analyzing the structural data in the databases to recognize functional RNA and to discover structural rules in diverse RNAs have been carried out, which may help answer basic biological and biochemical questions including those related to the origins of life.

The relevant investigations are essentially developed in two approaches. One is the sequence-based. It reduces the backbone torsion angles to a few conformation parameters, or defines the RNA framework with the transformation and the structure parameters. For instance, based on the framework of RNA, Duarte and Pyle [8] defined two pseudobonds between three adjacent nucleotides. In nature, there are two pseudotorsion angles round the two bonds, η and θ , which characterize the conformation of a nucleotide. Then, a reduced representation of motif conformational space of RNA secondary structure was created. And a plot called “RNA worm”, which is a virtual roadmap for an RNA structure, was used for describing the conformational changes of the nucleotides in a specific RNA sequence. Using this method, it becomes very convenient and efficient to search, compare and discover the diversity of motif in RNA, and the major regions involving conformational changes in the 50S and 30S ribosome subunits have been successfully identified [9]. Hershkovitz *et al.* [10] developed an approach of torsion matching and binning for recognizing and cataloging conformational states of RNA. Their methods allow one to use the backbone (α , β , γ , δ , ϵ , ζ) and the glycosidic (χ) torsion angles and ribose pseudorotation phase angle (P), as generally defining for nucleic acids, to describe the conformation for a residue. After given the cutoff angles for each angle, they identified the 18 A-helical regions of length > 9 and 25 tetraloops in the HM 23S rRNA. In binning, they reduced the eight torsion angles to four and represented ingeniously the conformation of one nucleotide by a distinct ASCII symbol. The method requires three contiguous residues (matching) and five (binning) separately at least in the identification of RNA conformational motifs. Substituting for the traditional nucleic acid residues which defined them from phosphate to phosphate, Murray *et al.* [11] used a base-to-base (or sugar-to-sugar) division into “suites” to parse the RNA backbone repeats. A suite conformer contains δ , ϵ , ζ of the residue i and α , β , γ of the residue $i + 1$, that is, two heminucleotides. In addition to the six torsion angles, a named quality-filtering technique invested with resolution, crystallographic B

factor, and all-atom steric clashes are applied to define the RNA backbone conformer. Finally, they defined 42 “suites”–RNA backbone conformers from 132 RNA crystal structure files. The “suite” approach eliminates the influence of data inaccuracies. However, the 42 “suites” seems too complicated and would be not suited to non-crystal structure data.

Another approach is based on the topology. It only involves the adjacency between nucleotides instead of their conformations. Representing the RNA secondary structure as a tree, or a dual graph, the eigenvalue spectrum are obtained from the *Laplacian matrix* of its tree graph, depicting RNA motif characteristic. With RNA topologies, Barash [12] gave a method for the prediction of deleterious mutations in the secondary structure of RNAs by the spectrum analysis of RNA tree graphs. This method may solve the 2-fold problem in RNA mutation, to find the minimal number of nucleotide mutations required to disrupt a stable motif and to specify their locations within the sequence. As the illustrative examples, Barash introduced the predictions using in the L5b tetraloop GAAA of the P5abc subdomain in the group I intron ribozyme of *Tetrahymena thermophila*, the two riboswitches of prokaryotic transcription termination, TPP and SAM [12]. Schlick and co-workers [13,14] applied a graph theoretical approach to construct the RNA structure space encompassing the topologies of existing and hypothetical RNAs and cluster all RNA topologies into two groups, “RNA-like” and “non-RNA-like”, using topological descriptors and a standard clustering algorithm. Their works can help to direct the design of functional RNAs and identify the novel RNA folds in genomes through an efficient topology-directed search, which grows much more slowly in complexity of RNA size compared to the traditional sequence-based search.

It is certainly that the studies reported on the searching, comparing, analyzing and discovering RNA secondary structure motif connected with nucleic acid databases are very helpful for understanding the biological functions that could appear in RNA folding, especially for identification and design of novel functional RNAs. Considering the same RNA sequence may adopt different motifs, then, can anyone make a sequence take a functional type, or change one to another? That is, we must answer a question by using the structure information in database, is there any possibility from understanding a motif to transforming a motif.

In all motifs of RNA secondary structure, its constituent of nucleotides can only be in two states, “paired” or “unpaired”. Zorn *et al.* analyzed the frequency distribution of the paired and unpaired bases from the large 16S and 23S rRNAs of *E. coli*. They suggested that natural RNAs may maintain certain proportions of bases in various motifs to ensure structural integrity [15]. Hence, if we are able to “control” an RNA fragment to pass into the motif form expected or to change one or several nucleotides’ state in the motif, as a result, the motifs would implement the particular biological function by figuring out in which state the nucleotide is. Obviously, it will be useful to find out the conformational factor of being able to achieve this purpose. However, this would be ignored in most of analysis reported on the RNA structure by utilizing the databases. Here, the

different motifs of RNA secondary structure were collected from the databases. Based on the PCA of the experimental structural data, we found out the conformational factors affecting the paired or unpaired state of nucleotides. After the tests of different RNA molecules and a new RNA secondary structure motif, we also discussed the probability of changing the motif form of RNA secondary structure through controlling the state that a nucleotide takes.

2 MATERIALS AND METHODS

2.1 Structure Organization

The RNA molecules used for analysis were taken from PDB and NDB, and their structures were determined by X-ray or NMR. The model 1 was selected from the NMR structures containing more than one models, and for X-ray structures, the highest resolution was selected. The relevant geometric data and the pairs were obtained from Swiss-Pdb Viewer (<http://www.expasy.ch/spdbv/text/main.htm>) by processing the NMR structure. Forty-five molecules are included. Firstly, after checked in the databases, we found out that the one strand RNA of non-hairpin (at 2.8 Å resolution or better) was limited in amount. When the first and last nucleotides in every sequence were eliminated because of the shortage of the conformational data, we got the strand motifs with one hundred nucleotides. As for other motifs, we chose those RNA molecules possessing one hundred nucleotides after the two terminal nucleotides were eliminated.

Table 1. The Three Groups of RNA Molecules and Their PDB Codes

Grouping I	Non-hairpin strand		Hairpins					
			Standard		Multiple loop	Symmetric internal loop	Asymmetric internal loop	Bulge
	1DUH	1G2J	1AFX	1SCL	1EBR	1EOR	1JO7	1A9L
	1OSU	1KFO	1AJF	1KKA	1F1T	1L1W	1BGZ	1F6X
	373D	255D	1ESH	3PHP	1M5L	1HWQ	1FYO	1RHT
	283D	333D				28SP	480D	1SLO
Nucleotides	100		100	100	100	100	100	100
Grouping II	Hairpins		tRNAs			rRNA		
	1ATV	1A9L	1E4P	1EHZ	1GRZ			
	1AQO	1ATO	1ATW	1FIR				
	1BN0	1RHT	1UUU	1YFG				
	3PHP	1SCL	1FYO	TRNA05 ^a				
Nucleotides	244		246			247		
Grouping III	Hook-turns							
	1MHK	1194 ^b						
		1KC9 ^b						
		1JJ2 ^b						
Nucleotides	92							

^a NDB code, the modified bases in tRNAs are not included.

^b The nucleitides involved in h-turns

The six RNA secondary structure motifs are shown in the column on the top of Table 1. The structure feature analysis was performed within the six hundreds nucleotides. Then, the same analysis of comparison and contrast was carried out for every motif set contained one hundred

nucleotides. Secondly, an asymmetric unit of *Tetrahymena ribozyme* LSU rRNA group I intron [16] which included two hundred and forty–seven nucleotides, matched twelve hairpins, and four tRNAs (middle column in Table 1) were analyzed by the method in the first stage in order to improve the results above. Lastly, a new motif, four h–turns [17] in the column on the bottom of Table 1, was used to verify it.

2.2 Conformation Parameters

The conformation of a nucleotide in an RNA molecule was described by six backbone torsion angles α , β , γ , δ , ϵ , ζ and χ between the sugar and the base (Figure 1). Of which only δ is involved in the sugar ring. Our aim was not to compare sequences, and neither was to find motifs. Therefore, when we examined the state of the paired nucleotides one by one, the seven torsion angles were used and the distance between two phosphorus atoms was also used as the eighth parameter in addition.

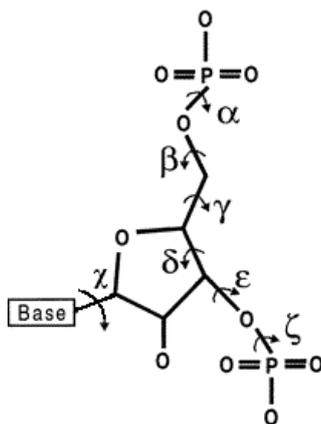


Figure 1. The torsion angles used as parameters in this paper.

2.3 Analysis Method: Principal Component Analysis

Principal component analysis (PCA) [18] is a popular multivariate technique. It is commonly used to elucidate the structure in a data matrix from an original experiment. It tries to represent the most important aspects of the original variables by using a smaller set of new variables (namely PCs). The original $m \times n$ data matrix, X , is decomposed into an $m \times n$ scores matrix, S , and an $n \times n$ loadings matrix, L . When the first PC has been defined, the second PC is chosen to be orthogonal to the first PC. The followed PCs are demanded to orthogonal with them. The total amount of variation, explained by a PC, is measured by the eigenvalue, λ , of the PCs rank according to the eigenvalues from large to little. The first PC explains the most variance and the last PCs usually explain very little variance. The scores of relevant PC1–PC2 (or/and PC1–PC3) in the S matrix are plotted into a two–dimension coordinate so that a score plot is formed. Correspondingly, a loading plot is formed by the loadings in the L matrix. Instead of interpreting score and loading plots

separately, so called biplot [19], in which the scores on the two PCs together with the loadings of the original variables are depicted in the same plot, can be commonly used. Correlations between variables are evaluated in terms of a PC model and it is proportional to its *Euclidean* distance from the origin. Similarities between objects on these PCs are quantified in terms of the distance between them. Before performing the PCA the original data are usually scaled. We applied a scaling technique that is an equal weight and mean-centering for each original variable.

As for the analysis of the structure data of RNA nucleotides, some authors discussed the influences of different structure representations on the outcome of multivariate technique specially, they showed much solicitude for the “the circularity” of the data [20–22]. We know that whether a nucleotide is similar to another is determined by the interposition of the corresponding atoms in the molecules, and this does not vary in terms of the circularity of the “angle” used to describe the interposition between atoms. It is necessary to regularize a kind of description of nucleotide molecule geometry as a popular standard reference frame since the different representations in describing conformation might bring the conflicting outcomes [23]. On the basis of the biomacromolecule conformational term of a torsion angle, there is a torsion angle between every four atoms, A–B–C–D. Viewed from B to C, an angle is named as positive when C–D is turned to B–A in an anticlockwise direction, and negative in a clockwise direction. Therefore, -120° and 240° indicate the angle represented the same interposition of the A, B, C, and D. In fact, it is identical without consideration of -180 to 180 and 0 to 360 . As far as two–dimension figure, the Figure 6 in Buydens’ paper [22], if we overlap 0° and 360° edges and then -180° and 180° edges, we can get the identical distributions of the dots on two cylinder planes. In our study, PCA was denoted by 0 – 360 conformation representation, and to corresponding to the phosphorus atom distances in number scale, we used “radian”, other than “degree” in the first stage. PCA was performed with Sirius for windows 6.5 for PIV computer by Pattern Recognition Systems AS.

3 RESULTS AND DISCUSSION

3.1 Analysis of Motifs

PCA of all 600 nucleotides included 6 types of motifs were shown in Figure 2a and 2b. Figure 2a is PC1–PC2 biplot, in which the loadings of α and γ on PC1 are bigger and have an opposite sign of values, ζ on PC2 is bigger and orthogonal with α and γ . But the loadings of other variables are very small, which indicates that the three variables are the primary factors for conformational characteristics. Because of the negative correlation between α and γ (the two vectors are almost placed along the same straight line), it can be figured that the torsion angles α and ζ determine the basic conformation of a nucleotide. Figure 2b shows the PC1–PC2 scores. And in the figure, the distribution of the 600 nucleotides is not relative to the different motifs, but displays approximately

in two states according to the sign of the scores on PC1. Combined the biplot 2a, it can be concluded that the nucleotides located between the two vectors α and ζ belong to one class, and the rest belong to another. After contrast with the corresponding 600 nucleotides, an interesting phenomenon appears in the classification of PC1–PC2 biplot: the nucleotide cluster surrounded by α and ζ in the circle of Figure 2b is basically paired, and those dispersed around the cluster are mostly unpaired. This phenomenon shows that it is possible to represent the conformation of a nucleotide to describe the states of paired and unpaired by choosing two from the 8 parameters.

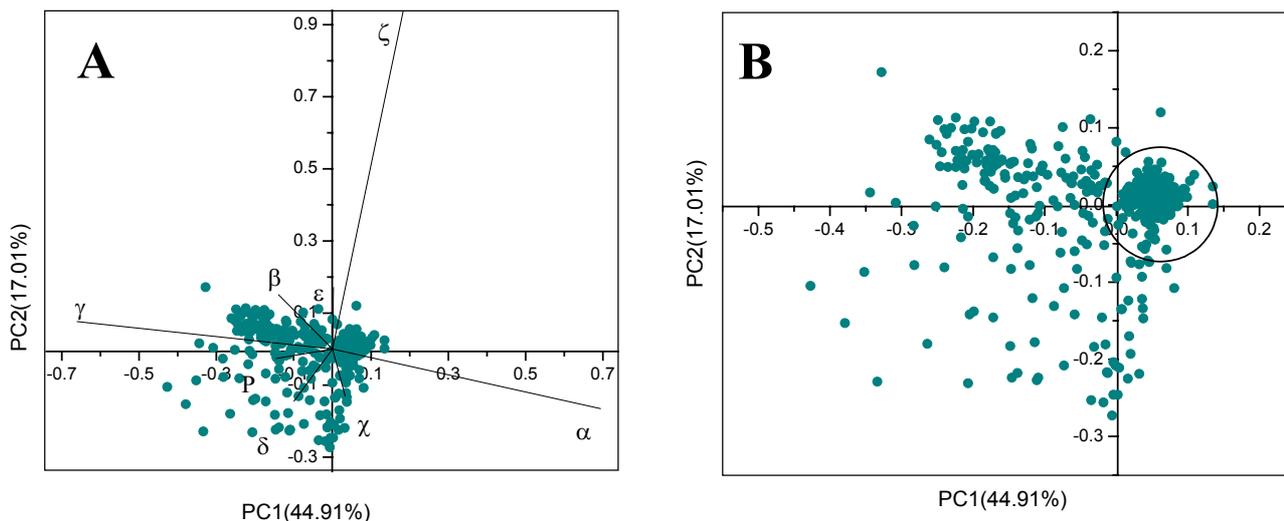


Figure 2. The results of PCA for the structural data of the 600 nucleotides: (a) biplot and (b) scores.

To substantiate the view, the same PCA was performed respectively on the 600 nucleotides according to the six different motifs, nonhairpin strand, standard hairpin, the hairpins contained bulge, internal loop, and multiple loop. The PC1–PC2 biplots are shown in Figure 3a–f. In Figure 3a the torsion angles α , γ and ζ obviously are more important than other 5 factors to examine structure characteristics. Two vectors which represent variables α and γ are just arranged in opposite directions but in a line (correlation coefficient = -1), ζ is orthogonal with them. So according to the scores on PC1 (total variance of 62.04 %), 100 nucleotides in the strand motifs, almost 90 percent is negative and other 10 percent positive. These nucleotides are divided into two groups on their positions in the angle between the two vectors α and ζ or outside of them. Notably, as the definition of the motifs, the elementary nucleotides of a pure strand should be unpaired, this definition makes it difficult to interpret the nucleotide distribution of the strand motifs which are similar to that of the 600 nucleotides above. The eight RNA molecules in the strand motifs are the synthetic matters. Though each asymmetric unit appears to be a strand, the present form in the assumed biological molecule (in crystal cell) is double helix. Those experimental results showed that the most nucleotides are in the paired, and the unpaired nucleotides, such as A39, C40, C41, A42, G54, A55 in 1DUH, just are arranged out the angle between α and ζ .

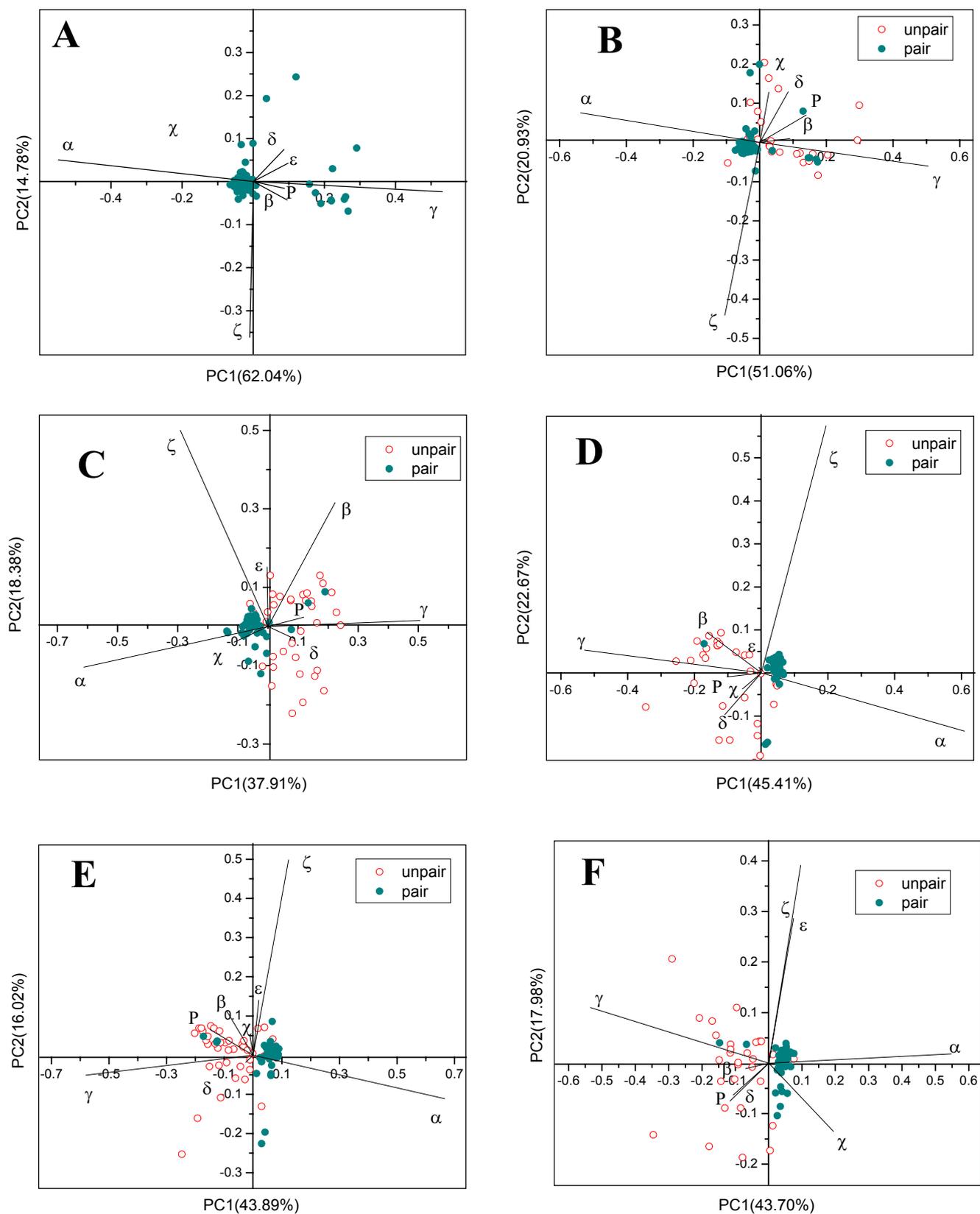


Figure 3. The results of PCA for the structural data of the six motifs. (a) The single strands. (b) The hairpins contained asymmetric internal loops. (c) The hairpins contained bulges. (d) The standard hairpins. (e) The hairpins contained symmetric internal loops. (f) The hairpins contained junctions.

Figure 3b showed the structural characteristics and distribution of nucleotides in the hairpin motifs including asymmetric internal loops. Similar to Figure 3a, in 3b, both the loading value and the interrelation of α , γ and ζ , as well as the region in which most paired or unpaired nucleotides were distributed, it demonstrated a structure characteristic with a close relationship to α and ζ .

The PC1–PC2 biplot of the hairpin motifs including bulge (Figure 3c) was slightly different from the above two figures, for example, the sign of variable loadings on PC2 was changed, and torsion angle β became more remarkable. But, a characteristic such as Figures 3a and 3b is still remarkable; comparing PC1 loading of β to those of α and γ , β to which assign a smaller value; comparing with PC2 loading of ζ , β to which assign a smaller value. α , γ and ζ are still the main factors to examine the conformational properties of the nucleotides, the interrelations between them still remain. Almost paired nucleotides distribute inside the angle between α and ζ .

The PC1–PC2 biplots from the nucleotides of the other three hairpin motifs appeared regular. Figures 3d, 3e, and 3f indicated the standard, the symmetric internal loop, and the complex, individually. Compared with the above three figures, the loading sign of α and γ on PC1 was exchanged, but the interrelationship was still negative, and the orthogonal to ζ was retentive (Figure 3f is light).

Demonstrated by the standard hairpins in Figure 3d, the positive or negative of loadings on PC1 can distinguish the paired or the unpaired nucleotides strictly, any nucleotide located at the region of the angle between α and ζ was paired. The symmetric internal loops remained the same feature, but the situation of the complex hairpins had slight alteration. The latter displayed a decrease of the angle between α and ζ , meanwhile, the loadings of ϵ and χ on PC1 were smaller than those of α , γ , and the loadings on PC2 are smaller than that of ζ , but larger than those of β , δ and P. The paired or unpaired location also corresponded to the positive or negative of scores on PC1.

These analyses of the 600 nucleotides in the whole and the six different motifs gave a common result. That is, the nucleotides in RNA secondary structure motifs can alter their paired or unpaired states and this alteration is subjected to adjustment of the factors α and ζ .

3.2 Analysis of Hairpin, tRNA and rRNA Molecules

As the amount of RNA molecules in the six secondary structure motifs analyzed above were small (less than 40 nucleotides), we chose larger number of RNA molecules *Tetrahymena ribozyme* LSU rRNA group I intron for further testing. In addition, 12 hairpins and 4 tRNA molecules had been used to contrast (Table 1 middle). Here we reserved seven torsion angle variables since the effect of the parameter P was not significant, and used 0–360 conformational representation. PCA results were demonstrated in Figure 4a, 4b and 4c.

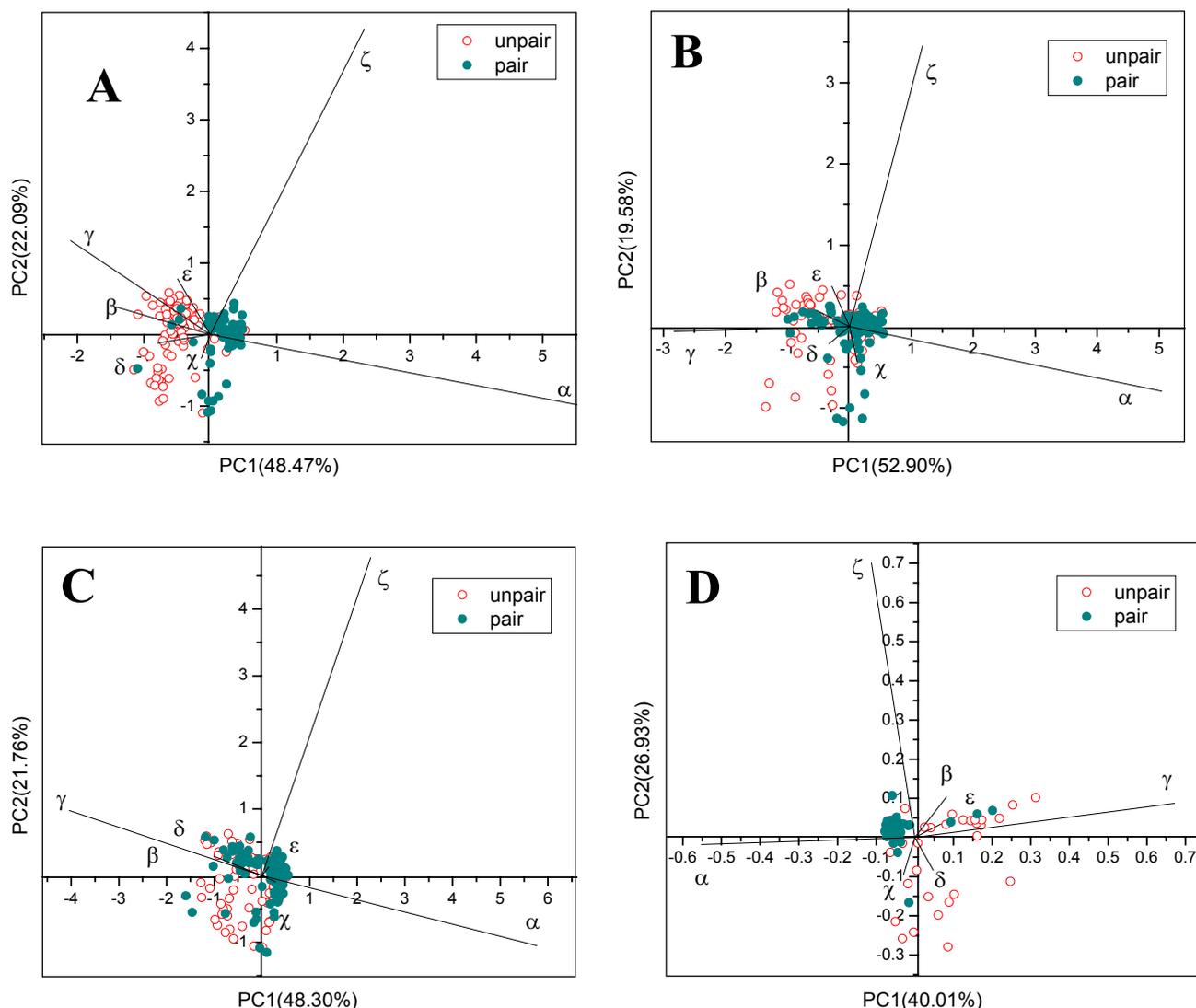


Figure 4. The results of PCA for the structural data of the 3 groups of RNA molecules: (a) the hairpins, (b) the tRNAs, (c) the rRNA, and (d) the hook–turn motifs.

Figure 4a showed the PC1–PC2 biplot of the 12–hairpin motifs. Compared to the standard hairpins in Figure 3d, the negative correlation of α to γ weakens, and the distribution of the paired or unpaired nucleotides was not strictly on their positive–negative scores of PC1. In addition, the loading of γ on PC1 decreases. Such change should result from that the 12 molecules contain nonstandard hairpin motifs, and 0–360 was used for representation. Surprisingly, the paired and the unpaired nucleotides held fast to their position inside and outside of the angle of the vectors α and ζ . The results from tRNAs and rRNA (Figure 4b, c) demonstrated the dominant position of α , γ and ζ , and their interrelationships as the same as those in figures 2 and 3. The nucleotides in the two RNA molecules inside of the positive and negative directions of PC1 were accord with the state of paired and unpaired. A difference was that a few of paired nucleosides inserted into the position of the unpaired, and so did on the contrary. This could be attributed to the interactions between the secondary structure motifs in tRNA and rRNA molecules. This tertiary interaction caused unpaired

nucleotides to be paired, while those new paired nucleotides flowed to the previous paired nucleotides. Actually, the nucleotides with error position mostly involved the inferior hydrogen-bond, the stem-loop join, and the mistake pair.

3.3 Analysis of Hook–Turn Motifs

Recently, Moore *et al.* [17] reported the crystal structure of a 26–nucleotide RNA. It consisted of an A–form helix that splits into two separate strands following a sheared A–G base pair. The backbone of the strand containing the G of the A–G pair made a turn of almost 180° in the space of two nucleotides. Similar structures, which they called “hook–turn motifs”, occurred in 16S and 23S rRNA [2, 24–26]. In order to test the results of PCA by which we analyzed the six RNA secondary structure motifs, the 12–hairpins, tRNA, and rRNA structures, PCA to the hook–turn motifs (Table 1 down) had been carried out (Figure 4d). The PC1–PC2 biplot which was similar to the hairpin motifs contained the bulge (Figure 2c) carried more obviously characteristics, that was, α and γ are negative correlation, and orthogonal with ζ . The sign of scores on PC1 determines the pair–state of nucleotides, and more than 90 percent nucleotides were concentrated inside the angle of the vectors α and ζ , meanwhile unpaired nucleotides were outside of it.

4 CONCLUSIONS

The result deduced from all PCA is encouraging. Among the seven torsion angles, which represented the nucleotide backbone conformation in RNA secondary structure motif, both α and ζ establish the state of a nucleotide–paired or unpaired? If considered the relationships between the variables and PCs, PC1 represented α (and/or γ), and PC2 represented ζ . α was even more significant because of the sign of scores on PC1. The simple statistics analysis (data not shown) of the torsion angles α and ζ for all nucleotides participated in the PCA disclosed that for α angle, the paired nucleotides located at the forth quadrant preferentially, meanwhile the unpaired nucleotides possessed the angle values from 45° to 135°, or the third quadrant. The ζ angle had not any special tendency. The preference of α and ζ was corresponding to the “binning” where α was mostly binned Bin 3, next were Bin1 and Bin2; ζ was widely distributed inside and outside of a bin [10]. Moreover, although Murray *et al.* used a two–heminucleotides “suite” and added a variety of level quality–filtering included resolution, crystallographic B factor, and all–atom steric clashes, they obtained such a result in the 3D of 8 636 nucleotides, of which α was fundamentally distributed in three grids and was not relevant to the sugar pucker, and ζ was a wide region (see Figure 4 in [11]). The works in these two groups did not connect the distributions of angles α and ζ with the paired–unpaired states of nucleotides. As our hypothesis, if there was a means which can be used to limit the value of α angle to a certain range, the pairing–tendency of nucleotides will be adjusted.

Are there any possibilities? α and ζ indicate the two torsion angles, O3'–P–O5'–C5' and C3'–

O3'–P–O5', which were closely connected to the backbone phosphorus atom. The P is bounded with four oxygen atoms. Change from bond angle O–P–O forces the two torsion angles to rearrange. Liebmann *et al.* [27] designed a model by which they studied the sensitivity of nucleic acid structure to the concentration and type of metal cation by using *ab initio* self-consistent field methods. They showed a specific change of the O–P–O angle. This change in the phosphate geometry could be propagated along the backbone of the strand if, as the one angle closes, another angle opens and can be translated to a change in backbone angle (torsion angle). The other conformation change in DNA, such as the transition between B and Z undergoes a process from a paired to an unpaired then a repaired. These changes are affected by solute ions and a chemical modification called methylation [28–30].

As for RNA, the transformation between paired and unpaired from a nucleotide is displayed in the interactions between motifs. Many functional RNA structures, for example, the catalytic center of the hammerhead ribozyme [31–33], the 16S ribosomal RNA junction [34,35], and the P5abc subdomain (a 56–nt RNA) of *Tetrahymena thermophila* group I intron [36,37], their forming proved the transformation between the paired and unpaired nucleotides and the requirement of metal ions. In addition, Bustamante *et al.* had achieved the unfolding and refolding to a single RNA molecule [38]. It is positive that the folding of RNA into a higher order structure with a biological significance undergoes a change between the two states, pairing and unpairing.

Compared with the reported conformational analyses on RNA, our results can answer perfectly the question what are the main conformational factors to bring a state change of a RNA secondary structure motif. The PCA for the backbone angles of 1320 nucleotides of 45 RNA molecules revealed a conformational feature, that was, whether a nucleotide was partial to pair or unpair. Among the seven torsion angles, both α and ζ , especially α , affects critically the conformational property of a RNA secondary structure motif because the angle distribution of α delimits basically that every nucleotide in the motif is located at stem (paired) or loop (unpaired). The theoretical and experimental studies on DNA mentioned above answered a question that the state changed between the paired and the unpaired nucleotides might be achieved by chemical means to alter the conformational factor. Furthermore, the transformation between paired and unpaired from a nucleotide displayed in the interactions between RNA secondary structure motifs required the help of some metal ions. This provided a possibility to achieve the state change via the external condition affecting the torsion angle α . Of course, applications of the graph theory will help to find the candidate sequences in RNA universe to synthesize the desired motifs. Moreover, the PCA results here pointed out a theoretical approach to transform motifs and for obtaining the functional RNA motifs were required. As a hope, we may use some physical methods, for example, radiation, mechanical force, or chemical methods, cation bonding, group replacing, and so on, which are single molecule means, to make a nucleotide to have a range of α value. If so, the aim of dominating the special type of RNA secondary structure motif should be realized.

Appendix 1

Table A1. The descriptions of the molecules in Table 1

PDB ID	# Res. ^c	Resolution	Source	Title
1DUH	45	2.7	Synthetic	The conserved domain IV of <i>E. coli</i> 4.5S RNA
1G2J	8	1.97	Synthetic	RNA octamer r(cccpgggg) containing phenyl ribonucleotide
1OSU	6	1.4	Synthetic	A 5'-UU-overhang exhibiting Hoogsteen-like trans U. U base pairs.
1KFO	19	1.6	Synthetic	An RNA helix recognized by a zinc-finger protein
413D	13	1.8	Synthetic	A'-form of r(ugagcuucggcuc)
255D	12	2	not available	An RNA double helix incorporating a track of non-Watson-Crick base pairs
283D	12	2.3	Synthetic	A curved RNA helix incorporating an internal loop with non-Watson-Crick base pairing
333D	8	2.52	Synthetic	An RNA oligomer incorporating tandem A-I mismatches
1AFX	12	13 struct.	<i>Eukaryotae</i>	UGAA Eukaryotic rRNA tetraloop
1SCL	29	NMR	<i>Rattus norvegicus</i>	The conformation of the sarcin/ricin loop from 28S rRNA.
1AJF	18	NMR	<i>Tetrahymena thermophila</i>	The P5b stem loop from A Group I intron complexed with Co (III) hexammine
1KKA	17	8 struct.	Synthetic	The unmodified anticodon stem-loop from <i>E. coli</i> tRNA(Phe)
1ESH	13	NMR	Synthetic	The stem loop C 5'aua3' triloop of brome mosaic virus (+) strand RNA
3PHP	23	10 struct.	<i>Turnip yellow mosaic virus</i>	The 3' hairpin of the tymv pseudoknot: preformation in RNA folding
1EBR	30	5 struct.	Synthetic	RNA (5'-ggugggcagcuucggcugcgguaccac-3')
1F1T	38	2.8	Synthetic	The malachite green aptamer complexed with tetramethyl-rosamine
1M5L	38	15 struct.	Synthetic	Wild-type and mutant internal loops from the SI-1 domain of the HIV-1 packaging signal
1EOR	22	NMR	Synthetic	A 22-nucleotide hairpin similar to the P5abc region of Group I ribozyme
1L1W	29	NMR	Synthetic	A Srp19 binding domain in human Srp RNA
1HWQ	30	20 struct.	Synthetic	The Vs ribozyme substrate stem-loop
28SP	28	7 struct.	Synthetic	NMR Structure Of The Most Conserved RNA Motif In Srp RNA
1JO7	31	32 struct.	Synthetic	Influenza A virus promoter
1BGZ	23	6 struct.	<i>E. coli</i>	S8 rRNA binding site from <i>E. coli</i>
1FYO	27	25 struct.	Synthetic	Eukaryotic decoding region A-site RNA
480D	27	1.5	Synthetic	The sarcin/ricin domain from <i>E. coli</i> 23 S rRNA
1A9L	38	12 struct.	<i>Haloferax volcanii</i>	A substrate for the archaeal pre-tRNA splicing endonucleases
1F6X	27	NMR	Synthetic	The RNase P RNA (M1 RNA) P4 stem oligoribonucleotide
1RHT	24	NMR	not available	24-Mer RNA hairpin coat protein binding site for bacteriophage R17
1SLO	19	NMR	<i>Caenorhabditis elegans</i>	First stem loop of the S11 RNA from <i>caenorhabditis elegans</i>
1ATV	17	4 struct.	not available	Hairpin with AGAA tetraloop
1E4P	24	20 struct.	Synthetic <i>Neurospora varkud satellite</i>	The ribozyme substrate hairpin of <i>neurospora</i> vs RNA
1AQO	29	15 struct.	not available	Iron responsive element RNA hairpin
1ATO	19	10 Structures	<i>Hepatitis d virus</i>	The isolated, central hairpin of the HDV antigenomic ribozyme
1ATW	15	3 struct.	not available	Hairpin with AGAU tetraloop
1BN0	20	11 struct.	Synthetic	SI3 hairpin from the packaging signal of HIV-1
1UUU	19	15 struct.	not available	An RNA hairpin loop with a 5'-cguuucg-3' loop motif
1EHZ	76	1.93	Synthetic	Yeast Phenylalanine tRNA
1FIR	76	3.3	<i>Bos taurus</i>	HIV-1 Reverse Transcription Primer tRNA(Lys3)
1YFG	75	3	<i>Saccharomyces cerevisiae</i>	Yeast Initiator tRNA
tRNA05 ^a	75	3	<i>Saccharomyces cerevisiae</i>	Yeast tRNA-Asp
1GRZ	247	5	<i>Tetrahymena thermophila</i>	A preorganized active site in the <i>Tetrahymena</i> ribozyme
1MHK	14(L) 12(S)	2.5	Synthetic construct	A 26-Mer RNA molecule, representing a new RNA motif, the Hook-turn
1I94 ^b	1514(A)	3.2	<i>Thermus thermophilus</i>	The small ribosomal subunit with tetracycline, Edeine and If3
1NKW ^b	2880(0) 124(9)	3.1	<i>Deinococcus radiodurans</i>	The large ribosomal subunit from <i>Deinococcus radiodurans</i>
1JJ2 ^b	2922(0) 122(9)	2.4	<i>Haloarcula marismortui</i>	The <i>Haloarcula marismortui</i> large ribosomal subunit

^a NDB code. ^b The nucleotides involved in h-turns are 25, 22, 23 separately. ^c Number of residues.

Acknowledgment

The authors acknowledge the financial of this research by National Natural Science Foundation of China (30160036, 90208018, 90303018). We thank Dr. Donghai Ye for his drawing.

5 REFERENCES

- [1] N. Ban, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*. **2000**, *289*, 905–920.
- [2] B. T. Wimberly, D. E. Brodersen, W. M. Jr Clemons, R. J. Morgan–Warren, A. P. Carter, C. Vornrhein, T. Hartsch and V. Ramakrishnan, Structure of the 30S ribosomal subunit. *Nature*. **2000**, *407*, 327–339.
- [3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, The Protein DataBank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- [4] H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S. H. Hsieh, A. R. Srinivasan and B. Schneider, The nucleic acid database: a comprehensive relational database of three dimensional structures of nucleic acids. *Biophys. J.* **1992**, *63*, 751–759.
- [5] V. L. Murthy, R. Srinivasan, D. E. Draper and G. D. Rose, A complete conformational map for RNA. *J. Mol. Biol.* **1999**, *291*, 313–327.
- [6] F. H. D. van Batenburg, A. P. Gulyaev, C. W. A. Pleij, J. Ng and J. Oliehoek, PseudoBase: a database with RNA pseudoknots. *Nucleic Acids Res.* **2000**, *28*, 201–204.
- [7] T. R. Cech, Ribozymes, the first 20 years. *Biochem. Soc. Trans.* **2001**, *30*, 1162–1166.
- [8] C. M. Duarte and A. M. Pyle, Stepping through an RNA structure: A novel approach to conformational analysis. *J. Mol. Biol.* **1998**, *284*, 1465–1478.
- [9] C. M. Duarte, L. M. Wadley and A. M. Pyle, RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Res.* **2003**, *31*, 4755–4761.
- [10] E. Hershkovitz, E. Tannenbaum, S. B. Howerton, A. Sheth, A. Tannenbaum and L. D. Williams, Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Res.* **2003**, *31*, 6249–6257.
- [11] L. J. W. Murray, W. B. Arendall III, D. C. Richardson and J. S. Richardson, RNA backbone is rotameric. *Proc. Natl. Acad. Sci. USA.* **2003**, *100*, 13904–13909.
- [12] D. Barash, Deleterious mutation prediction in the secondary structure of RNAs. *Nucleic Acids Res.* **2003**, *31*, 6578–6584.
- [13] H.H. Gan, S. Pasquali, and T. Schlick, Exploring the repertoire of RNA secondary motifs using graph theory: Implications for RNA design. *Nucleic Acids Res.* **2003**, *31*, 2926–2943.
- [14] N. Kim, N. Shiffeldrim, H.H. Gan and T. Schlick, Candidates for novel RNA Topologies. *J. Mol. Biol.* **2004**, *341*, 1129–1144.
- [15] J. Zorn, H. H. Gan, N. Shiffeldrim and T. Schlick, Structural motifs in ribosomal RNAs: implications for RNA design and genomics. *Biopolymers.* **2004**, *73*, 340–347.
- [16] B. L. Golden, A. R. Gooding, E. R. Podell and T. R. Cech, A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme, *Science*. **1998**, *282*, 259–264.
- [17] S. Szep, J. Wang and P. B. Moore, The crystal structure of a 26–nucleotide RNA containing a hook–turn. *RNA*, **2003**, *9*, 44–51.
- [18] S. World, K. Esbensen and P. Geladi, Principal component analysis. *Chemometrics Intell. Lab. Syst.* **1987**, *2*, 37–52.
- [19] K. R. Gabriel, The biplot graphical display of matrices with application to principal component analysis. *Biometrika.* **1971**, *58*, 453–467.
- [20] L. M. C. Buydens, T. H. Reijmers, M. L. M. Beckers and R. Wehrens, Molecular data–mining: a challenge for chemometrics. *Chemometrics Intell. Lab. Syst.* **1999**, *49*, 121–133.
- [21] T. H. Reijmers, R. Wehrens and L. M. C. Buydens, The influence of different structure representations on the clustering of an RNA nucleotides data set. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1388–1394.
- [22] T. H. Reijmers, R. Wehrens and L. M. C. Buydens, Circular effects in representations of an RNA nucleotides data set in relation with principal components analysis. *Chemometrics Intell. Lab. Syst.* **2001**, *56*, 61–71.

- [23] X. J. Lu and W. K. Olson, Resolving the discrepancies among nucleic acid conformation analyses. *J. Mol. Biol.* **1999**, *285*, 1563–1575.
- [24] D. J. Klein, T. M. Schmeing, P. B. Moore and T. A. Steitz, The kink–turn: A new RNA secondary structure motif. *EMBO J.* **2001**, *20*, 4214–4221.
- [25] F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon and F. Franceschi, Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell.* **2000**, *102*, 615–623.
- [26] J. Harms, F. Schluenzen, R. Zarivach, A. Bashan, S. Gat, I. Agmon, H. Bartels, F. Franceschi and A. Yonath, High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell.* **2001**, *107*, 679–688.
- [27] P. Liebmann, G. Loew, A. D. Mclean and G. R. Pack, *Ab initio* SCF studies of interactions of Li^+ , Na^+ , Be^{2+} and Mg^{2+} with H_2PO_4^- : model for cations binding to nucleic acids. *J. Am. Chem. Soc.* **1982**, *104*, 691–697.
- [28] M. Gueron, J.–Ph. Demaret and M. Filoche, A unified theory of the B–Z transition of DNA in high and low concentrations of multivalent ions. *Biophys. J.* **2000**, *78*, 1070–1083.
- [29] J. Klysik, S. M. Stirdivant, C. K. Singleton, W. Zacharias and R. D. Wells, The effects of 5–cytosine methylation on the B–Z transition in DNA restriction fragments and recombinant plasmids. *J. Mol. Biol.* **1983**, *168*, 51–71.
- [30] J. J. Van Lier, M. T. Smits and H. M. Buck, B–Z transition in methylated DNA. A quantum–chemical study. *European J. Biochem.* **1983**, *132*, 55–62.
- [31] Pley, K. M. Flaherty and D. B. McKay, Three–dimensional structure of a hammerhead ribozyme. *Nature.* **1994**, *372*, 68–74.
- [32] W. G. Scott, J. T. Finch and A. Klug, The crystal structure of an all–RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell.* **1995**, *81*, 991–1002.
- [33] J. B. Murray, D. P. Terwey, L. Maloney, A. Karpeisky, N. Usman, L. Beigelman and W. G. Scott, The structural basis of hammerhead ribozyme self–cleavage. *Cell.* **1998**, *92*, 665–673.
- [34] T. Ha, X. W. Zhuang, H. D. Kim, J. W. Orr, J. R. Williamson and S. Chu, Ligand–induced conformational changes observed in single RNA molecules. *Proc. Natl. Acad. Sci. USA.* **1999**, *96*, 9077–9082.
- [35] H. D. Kim, G. U. Nienhaus, T. Ha, J. W. Orr, J. R. Williamson and S. Chu, Mg^{2+} –dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules. *Proc. Natl. Acad. Sci. USA.* **2002**, *99*, 4284–4289.
- [36] R. T. Batey, R. P. Rambo and J. A. Doudna, Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed.* **1999**, *38*, 2326–2343.
- [37] M. Wu and I. Tinoco, Jr., RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. USA.* **1998**, *95*, 11555–11560.
- [38] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, Jr. and C. Bustamante, Reversible unfolding of single RNA molecules by mechanical force. *Science.* **2001**, *292*, 733–737.