PTRNAmark: an all-atomic distance-dependent knowledge-based potential for 3D RNA structure evaluation

Yi Yang[1,2], Qi Gu[2], Ya-Zhou Shi[2], Zhi-Gang Shao[1] and Zhi-Jie Tan[2*]

[1]*Guangdong Provincial Key Laboratory of Quantum Engineering and Quantum Materials, SPTE, South China Normal University, Guangzhou 510006, China.*
[2]*Center for Theoretical Physics and Key Laboratory of Artificial Micro & Nano-structures of Ministry of Education, School of Physics and Technology, Wuhan University, Wuhan 430072, China*

ABSTRACT

RNA molecules play vital biological roles, and understanding their structures gives us crucial insights into their biological functions. Model evaluation is a necessary step for better prediction and design of 3D RNA structures. Knowledge-based statistical potential has been proved to be a powerful approach for evaluating models of protein tertiary structures. In present, several knowledge-based potentials have also been proposed to assess models of RNA 3D structures. However, further amelioration is required to rank near-native structures and pick out the native structure from near-native structures, which is crucial in the prediction of RNA tertiary structures. In this work, we built a novel RNA knowledge-based potential—PTRNAmark, which not only combines nucleotides' mutual and self energies but also fully considers the specificity of every RNA. The benchmarks on different testing data sets all show that PTRNAmark are more efficient than existing evaluation methods in recognizing native state from a pool of near-native states of RNAs as well as in ranking near-native states of RNA models.

Keywords: RNA structure; knowledge-based potential; scoring

---

* To whom correspondence should be addressed: zjtan@whu.edu.cn

INTRODUCTION

RNA molecules play essential roles in living systems, and its catalytic prowess, biological importance and ability to form complex folds have recently come to prominence [1-3]. Like proteins, in order to understand the functions of RNAs deeply, we need to study the 3D structure. In recent years, several computational methods have been proposed for structural modeling or RNA tertiary structures prediction, for the number of available RNA experimental structures is finite at the present time [4-13]. For all the methods, a large set of decoys are generated when models predict RNA tertiary structures, so it is so significant to evaluate these candidates.

One of approach for evaluating models of protein tertiary structures is knowledge-based statistical potential, which has been proved to be an efficient method [14-16]. At the present, some of the knowledge-based potentials have also been proposed to evaluate models of RNA tertiary structures [17-21]. For instance, firstly Capriotti et al. has developed the Ribonucleic Acids Statistical Potential (RASP) in 2011[17,19]. Secondly the coarse-grained and all-atom KB potentials have been proposed by Bernauer et al. in 2011 [18]. Thirdly in 2015 Yi Xiao et al. has developed a novel all-atom knowledge-based statistical potential, 3dRNAscore [24]. The KB potential is a distance-dependent statistical potential, which used a Dirichlet process mixture model to obtain the distance distributions instead of bin counting, and the fully differentiable feature also makes it possible for molecular dynamics simulations [18,23]. RASP is also a distance-dependent statistical potential which is a detailed full-atom potential that includes a representation of local and non-local interactions in RNA structures and is trained on a non-redundant training set (randstr) generated by MODELLER [17,22]. Unlike the aforementioned two knowledge-based potentials that utilize the distances between atoms, 3dRNAscore uses not only distance-dependent potential, but also a dihedral-dependent potential, involving seven RNA dihedral angles [24]. The benchmark tests showed that the RASP, KB and 3dRNAscore potentials could identify the native state structures effectively [17,18,24]. What's more, other statistical potentials which embedded in the RNA tertiary

structure prediction programs have been proposed for evaluating RNA tertiary structures. For example, a full-atom RNA potential (FARFAR, fragment assembly of RNA with full-atom refinement) available within the ROSETTA suite is efficient to be used for the de novo prediction and design of non-canonical RNA 3D structures [6,11]. And MC-fold, which can predict RNA 3D structures, also use an all-atom score function that embedded in the RNA tertiary structure prediction programs to rank RNA tertiary structures [9]. However, for protein, it is the problem that it is the distinction on universality and pertinence that makes the potentials perform diversely in different decoy sets [25]. There is a problem which is the same as protein in RNA, which means that there is a contradiction between universality and pertinence. And it is essential that further enhancements are needed to rank near-native structures and pick out the structure closest to the native state from near-native structures.

In this work, we build an all-heavy-atom knowledge-based statistical potential called Precision Training RNA Mark (PTRNAmark). Unlike the aforementioned knowledge-based potentials that only consider non-bond interactions which belongmutual nucleotides, a new energy contribution based on the inside of nucleotide is involved in PTRNAmark. Furthermore, to consider the specificity of physical interaction, PTRNAmark is trained twice by two sorts of training sets, which is unchanging and changing respectively. For a decoy, firstly PTRNAmark is trained by a constant training set, like the aforementioned knowledge-based potentials, then PTRNAMark is trained by another training set which is some structures, originating from decoys, that are the lowest energy ranked by first time scoring. We think that this method could fully consider the characteristic of every RNA model and the specificity of physical interaction. It turns out that PTRNAmark performs better than 3DRNAScore, RASP, KB potentials and ROSETTA in ranking a tremendous amount of near-native RNA tertiary structures as well as recognizing native state from a pool of near-native states of RNAs.

MATERIALS AND METHODS

The steps for generating PTRNAmark are as follows. First, we design the functional form of PTRNAmark from Boltzmann distribution, which merges nucleotides' mutual and self-energies. Second, in order to train the parameters of the scoring function, which have been used to score for the decoys firstly, we select a training set of non-redundant RNA tertiary structures in which the structure features are representative and the structures of high similarity and low quality are removed. Third, we use the two test sets, which are occurring now, to test the accomplishment of PTRNAmark. Here, for every decoy, PTRNAmark is trained by some structures of the lowest energy which are from decoys and are scored by the PTRNAmark that is building in the second step. And, we use different metrics to compare the performance of PTRNAmark with other scoring methods. The more point of each building step of PTRNAmark is described in the follows.

Generation of RNA potential

Our knowledge-based potential PTRNAMark is made of two terms. The first term based on the distance between any two non-bonded heavy atoms located at different nucleotides in the molecule, and the second term based on distance between any two non-bonded heavy atoms located at inside nucleotides in the molecule. According to three assumptions which were pointed out by Samudrala [27], the total energy score of a given RNA sequence S_q with conformation C_p is calculated by

$$\mathrm{Score}(S_q, C_p) = \sum_{m=1} \sum_{n=m+1} \bar{u}_{i_m, i_n}(r_{m,n}) \quad (1)$$

Where $r_{m,n}$ are the distance between mth and nth atoms, and i_m and i_n are the residue-specific atom types, respectively.

The energy term of mutual nucleotide

The knowledge-based potentials were derived based on the Boltzmann or Bayesian formulations. For the atomic distance-specific contact potentials, the potential can be written as:

$$\bar{u}_{i,j}(r) = -RTIn\left[\frac{f_{i,j}^{OBS}(r)}{f_{i,j}^{REF}(r)}\right] \quad (2)$$

where R and T are Boltzmann constant and Kelvin temperature, respectively. $f_{i,j}^{OBS}(r)$ is the observed probability of atomic pairs (i,j) within a distance bin r to r+dr in experimental RNA conformations. $f_{i,j}^{REF}(r)$ is the expected probability of atomic pairs (i,j) in the corresponding distance from random conformations without atomic interactions, which is so-called reference state. Here atomic pair (i,j) runs through all the atomic pairs in the RNA chain except for those pairs of the same nucleotides. Because of the reason that the equal size of datasets is used for calculating $f_{i,j}^{OBS}(r)$ and $f_{i,j}^{REF}(r)$ in RNA statistical potentials at present, the probabilities in Eq. (2) can be replaced by the frequency counts of atomic pairs:

$$\bar{u}_{i,j}(r) = -RTIn\left[\frac{\frac{N_{i,j}^{OBS}(r)}{N_{i,j}^{OBS}}}{\frac{N_{i.j}^{REF}(r)}{N_{i.j}^{REF}}}\right] = -RTIn\left[\frac{N_{i.j}^{OBS}(r)}{N_{i.j}^{REF}(r)}\right] \quad (3)$$

Here, $N_{i.j}^{OBS}(r)$ is the observed number of atom pairs (i,j) at the distance r in experimental RNA structures. $N_{i.j}^{REF}(r)$ is the expected number of atomic pairs (i,j) if there were no interactions between atoms. $N_{i,j}^{OBS} \equiv N_{i.j}^{REF} = \sum_{r}^{r_{cut}} N_{i,j}^{OBS}(r)$ is the total number of atomic pairs (i,j) in thestructure samples, where r_{cut} is the cutoff distance. Cut off distance means the maximum value of the distance d in the process of statistics, and we find that when cut off distance is 20 Å, the number of atom pair observed is maxed, so we take 20 Å to be cut off distance in PTRNAmark. The counts of $N_{i.j}^{REF}(r)$ could not be compiled from experimental structures directly. It relies on which reference state we choose. The RASP and 3DRNAScore potentials both used averaged (RAPDF) reference states [27] and the KB potentials used quasi-chemical (KBP) [29] approximation reference states [30]. Our statistical potential chooses the averaged reference state, which ignores the type of atom. In averaged reference state, $N_{i.j}^{REF}(r)$ can be calculated as follows [27]:

$$N_{i,j}^{REF}(r) = f^{REF}(r)N_{i,j}^{REF} = \frac{\sum_{i,j} N_{i,j}^{OBS}(r)}{\sum_{i,j}\sum_{r} N_{i,j}^{OBS}(r)} \sum_{r} N_{i,j}^{OBS}(r) = \frac{N^{OBS}(r)}{N_{total}^{OBS}} N_{i,j}^{OBS} \quad (4)$$

Here $N^{OBS}(r)$ is the counts of observed contacts between all pairs of atom types at a particular distance r. $N_{i,j}^{OBS}$ is the number of the occurrence of atom pairs of types i and j in whole distance region. N_{total}^{OBS} is the total number of contacts between all pairs of atom types summed over all distance r, it means the total counts. So the first term of functional form of PTRNAmark can be written as:

$$\bar{u}_{i,j}(r) = -RTIn\left[\frac{N_{i,j}^{OBS}(r)N_{total}^{OBS}}{N^{OBS}(r)N_{i,j}^{OBS}}\right] \quad (5)$$

The energy term of nucleotide's inside

We use not only mutual nucleotide potential, but also the inside potential of nucleotide, involving all atom pairs excluding these atom pairs which belong to bond stretching and angle bending in four RNA nucleotide inside. First, we calculated their statistical distribution over the training set. Once we get their statistical distributions, just like the mutual nucleotide potential, the potential can be written as:

$$\bar{u}_{q,w}(r) = -RTIn\left[\frac{f_{q,w}^{OBS}(r)}{f_{q,w}^{REF}(r)}\right] \quad (6)$$

where R and T are Boltzmann constant and Kelvin temperature, respectively. $f_{q,w}^{OBS}(r)$ is the observed probability of atomic pairs (q,w) within a distance bin r to r+dr in experimental RNA conformations. $f_{i,j}^{REF}(r)$ is the expected probability of atomic pairs (q,w) within a distance bin r to r+dr in a reference state. Here atomic pair (q,w) runs through all the atom-pairs precluding these atom pairs which belong to bond stretching and angle bending in four RNA nucleotide inside. Then we could also get the second term of functional form of PTRNAmark in the same way as the mutual nucleotide potential energy, it can be written :

$$\bar{u}_{q,w}(r) = -RTIn\left[\frac{N_{q,w}^{OBS}(r)N_{total}^{OBS}}{N^{OBS}(r)N_{q,w}^{OBS}}\right] \quad (7)$$

Combination of the two energy terms

In PTRNAMark the two energy terms are combined together to get the final total energy:

$$\bar{u}_{\text{total}} = \bar{u}_{\text{mutual}} + \varepsilon \bar{u}_{\text{inside}} \quad (8)$$

where $\bar{u}_{total}$ is the total energy, $\bar{u}_{\text{mutual}}$ is the energy of mutual nucleotide $\bar{u}_{\text{inside}}$ is the energy of nucleotide's inside, and ε is the weight. To get an appropriate value for ε, first we generate five decoy sets by using EFOLD program[31](PDBID:1MSY, 1ZIH, 1KKA, 1Q9A, 255D), then we calculated each decoy's RMSD and the energy score, after that we maximize the ES by trying using a series of ε value. In the end, we find the optimized ε value is 0.15(see supplementary data VIII).

Our all-heavy-atom distance-dependent potential utilizes all the 85 atom types in the four nucleotides: 22 atom types in adenine (A), 20 atom types in cytosine (C), 23 atom types in guanine (G) and 20 atom types in uracil (U). In general, the distance-dependent statistical potential just considers nonbonding interactions between the different nucleotide. For PTRNAmark, the first term considers the nonbonding interaction between different nucleotide, and the second term includes all interactions of nucleotide inside except for bond stretching and angle bending interaction. so the atom pair in which the two atoms belong to the homogeneous nucleotide would not be considered in the first term and the second term only considers the atom pair which belongs to nucleotide's inside except for bond stretching and angle bending.

For discrete statistics of data, the size of the bin has a great influence on the probability distribution. If the bin width is oversized, the result would be not truly precise. When the bin width is undersized, an unsuitable and artificial discontinuity of the result will occur, because of none or little samples located in certain bins .For protein potential, Sippl [26] used a bin width of 1 Å, Samudrala used a bin width of 1 Å and then carried out spline fitting [27]. For RNA potential, Capriotti (RASP) also used a bin width of 1 Å, Bernauer (KB) used a Dirichlet process mixture model, which leads to analytically differentiable potential functions, rather than fixed binning and spline fitting. Yi Xiao's group (3DRNAScore), used a bin width of 0.3 Å,

studying Scott's work [32] in 1979 and extracting the bin width for 3dRNAscore from experimental structure. Here, considering that there is an observed diversity of the number of atom pairs in different bin width, we use the varying bin width. When the distance of atom pair < 3 Å, the bin width is 1 Å. If the distance of atom pair >= 3 Å, the bin width is 0.3 Å.

As for the problem of sparse data, in 1990, Sippl developed a method to address this problem. He approximated the genuine frequency by the sum of the total densities and the statistical frequencies [26]. Yi Xiao's group (3DRNAScore) utilize a penalty to solve this problem, giving a penalty to the total energy score when the distance between two atoms is < 3 Å. In our PTRNAmark, we use a constant value to solve this problem.

Training sets

Since the performance of scoring function depends on training sets in some degree, it is considerable for a statistical potential to train the parameters of the scoring function. Here, in order to score a decoy, PTRNAmark needs to score a decoy two times. It means that PTRNAmark uses two sets of parameters, ones of which are trained by a certain training sets, the other of which are depended on some structures which are the lowest energy and originate from decoys, ranked by the first time score whose parameters is trained by a certain training sets. The detail of each scoring step of PTRNAmark is described in the following figure1.

In first scoring, PTRNAmark uses a certain training sets like the aforementioned knowledge-based potentials. To generate the training set, firstly we gathered 1369 structures by RNA 3D Hub non-redundant RNA set. RNA 3D Hub hosts the non-redundant set of RNA-containing 3D structures extracted from experimental RNA structures [33]. Then we discard all the RNAs with identity > 80% and coverage greater than 80% by using blastn[34]. Furthermore, we remove low-quality structures whose resolution >3.5 Å. After that some typical representative structures which

respectively contain internal loop, hairpin loop, double helix, triple helix, junction, bugle, non-Waston-Crick base pairing, mismatch base pairing are selected by using a visual conformation software PYMOL(http://www.pymol.org/). Finally, 380 structures selected are used to be a training set (PDB ID see supplementary data VI).

In second scoring, when you hope to rank the near-native structures for a decoy set, PTRNAmark uses N structures with energies in the lowest N of the energy range to be a training set, which are scored by the first time scoring. For most decoy sets, when the energy gap of adjacent structures with energies in the lowest 10 of the energy range is less than $1K_bT$, the N=10. However, if the energy gap of adjacent structures with energies in the lowest 10 of the energy range is large (>1 K_bT), in this moment N<10, and we select those structures, whose energy gap is small, with energies in the lowest N of the energy range to be a training set(see supplementary data VII). That means that parameters in PTRNAmark are changed for every decoy: if you score a new decoy, the parameters in PTRNAmark should have needed to be trained again.

Test Sets

We tested our knowledge-based potential using two different decoy sets available at present. Test set I is a randstr decoy set [17], which is the largest decoy dataset in the benchmark. This is generated by MODELLER [35] with a set of Gaussian restraints for dihedral angles and atom distances from 85 native structures. It can be downloaded from http://melolab.org/supmat/RNApot/Home.html.

Test set II is made up of two parts. The first part are the decoys established by Bernauer group [18], which is generated by position restrained dynamics, normal-mode perturbation approach and REMD simulation[36,37]. In the normal-mode perturbation approach, the structures possess stereochemically correct bond lengths and angles but without correct contacts [38]. In the REMD simulation whose temperature is roughly distributed RNA structures from 285 to 592 K for 50 different temperatures, each RNA structure was generated by 1ns REMD simulations. This decoys can be downloaded form http://csb.stanford.edu/rna/. Second part are the

FARNA decoys which consist of lots of near-native tertiary models [11]. It can be downloaded from https://daslab.stanford.edu/resources/.

Metrics of measuring RNA structures

It is significant for comparing any two RNA structures quantitatively to make use of some metrics to assess their tertiary structures. RMSD (root mean square deviation) is the most universal, which depicts the global geometry differences between two RNA 3D structures but it is usually difficult to describe the hydrogen bond networks of RNA molecules. Hence, accounting for hydrogen-bonding networks intra-molecular in RNA some metrics have been proposed to assess RNA structures. One of the commonly used metrics specifically devised for RNA is DI (deformation index) proposed by Parisien [39]. The DI is defined as

$$\mathrm{DI}(\mathrm{A},\mathrm{B}) = \frac{\mathrm{RMSD}(\mathrm{A},\mathrm{B})}{\mathrm{INF}(\mathrm{A},\mathrm{B})} \quad (6)$$

Here, for comparing hydrogen-bonding networks intramolecular between two RNA structures firstly let Sr be the set of interactions in a reference structure (usually an experimentally resolved structure) and Sm the set of interactions of a modeled structure. The interactions found in the intersection of both sets are true positives, $\mathrm{TP} = s_r \cap s_m$. The interactions in Sm that are not present in Sr are false positives, $\mathrm{FP} = s_m \backslash s_r$. The interactions absent in Sm but present in Sr are false negatives, $\mathrm{FN} = \mathrm{Sr} \backslash \mathrm{Sm}$. Then Parisien defines the interaction network fidelity (INF) between structures A and B as the MCC, and $\mathrm{INF}(\mathrm{A},\mathrm{B}) = \mathrm{MCC}(\mathrm{A},\mathrm{B})$, which is estimated by[40,41]:

$$\mathrm{MCC} = \sqrt{\mathrm{PPV} \times \mathrm{STY}}$$

$$\mathrm{PPV} = \frac{|\mathrm{TP}|}{|\mathrm{TP}| + |\mathrm{FP}|}$$

$$\mathrm{STY} = \frac{|\mathrm{TP}|}{|\mathrm{TP}|+|\mathrm{FN}|} \quad (7)$$

In this work, we use both RMSD and DI to measure how well an RNA model recapitulates the corresponding experimental structure in the benchmark of the performance of PTRNAmark and other scoring functions.

Metrics of measuring scoring function

To describe the ranking near-native RNA structures performance of a scoring function, the ES (enrichment score [18,44] is employed, which based on identifying the top 10% scoring ($\mathrm{E}_{\mathrm{top10\%}}$) and best 10% RMSD values ($\mathrm{R}_{\mathrm{top10\%}}$), then evaluating their degree of overlap (this choice percentage is somewhat arbitrary). The Enrichment Score is defined as

$$\mathrm{ES} = \frac{|\mathrm{E}_{\mathrm{top10\%}} \cap \mathrm{R}_{\mathrm{top10\%}}|}{0.1\times 0.1\times \mathrm{N}_{\mathrm{decoys}}} \quad (8)$$

where $\mathrm{E}_{\mathrm{top10\%}}$ is the number of structures with energies (scores given by scoring function) in the lowest 10% of the energy range. For RMSD-based ES, $\mathrm{R}_{\mathrm{top10\%}}$ is the number of structures with RMSD in the lowest 10%. For DI-based ES, $\mathrm{R}_{\mathrm{top10\%}}$ is the number of structures with DI in the lowest 10%. $|\mathrm{E}_{\mathrm{top10\%}} \cap \mathrm{R}_{\mathrm{top10\%}}|$ is the intersection of $\mathrm{E}_{\mathrm{top10\%}}$and $\mathrm{R}_{\mathrm{top10\%}}$. If the relationship between the scores and RMSD or DI is completely linear, then ES is equal to 10. If the relationship is random, ES is equal to 1, so

$$\mathrm{ES} = \begin{cases} 10 & \text{perfect scoring} \\ 1 & \text{perfectly random} \\ < 1 & \text{bad scoring} \end{cases}$$

In this work, we use ES, which based on DI and RMSD receptively, to measure scoring function in ranking near-native RNA structures performances.

RESULTS

The performances of selecting native structure from decoys

It is essential function for the statistical potential to identify the native-like tertiary structure of target RNA in a pool of structural decoys correctly. In order to illustrate the performances of selecting native structures from decoys of PTRNAmark, we compare three existing RNA knowledge-based potentials: RASP [17], KB [18] and Rosetta [6], using an all-heavy-atom representation for all the potentials and evaluating them over the same decoy sets for the purpose of equity.

We use two test sets which are test set I and test set II respectively for all potentials. (see Figure 2)When using PTRNAmark, 83 out of 85 native structures are identified in test set I, and 34 out of 39 in test set II. RASP could identify 77 out of 85 native structures in test set I, 34 out of 39 in test set II. For KB potential , the former is 80 out of 85 native structures and the latter is 33 out of 39 in test set II. For Rosetta 53 out of 85 native structure are identified in test set I, and 26 out of 39 in test set II. It is worth nothing that that RASP has a better performance than NAST [42] and the molecular force field energy function: AMBER pseudo-energies [43] has been proved by Capriotti's group. These results show that PTRNAmark has a better performance than other methods on identifying native RNA structures. A more detailed table than figure 2 and energy-RMSD plots are provided in the supplementary data I and II.

The performances of ranking near-native RNA structures

Ranking near-native structures within reason is another essential function of a scoring function. In the RNA structure prediction, how to confirm that a predicted structure is closer than others to the native state is a natural and significant problem which we have to face spontaneously. For RNA, to evaluate structure distinction, RMSD and DI which represent geometrical and topological unlikeness are common and useful metrics. In this work, we have benchmarked the performance of ranking near-native RNA structures of PTRNAmark, 3dRNAscore, RASP, KB and Rosetta on

test set II, using both RMSD and DI metrics. Figure3 shows that five different RNA decoys from test set II were used to make figures with RMSD and energy. In all five cases, PTRNAmark was very effective than others in identifying near-native decoys. what's more, table1 shows that when we are using ES of RMSD, both PTRNAmark (ES=5.1) outperforms other four scoring function on the overall average level:3DRNAscore (ES=4.5), KB (ES=3.7), RASP (ES=3.8) and Rosetta (ES=2.7). PTRNAmark also outperforms them on the REMD decoys, normal mode decoys and FARNA decoys of test set II. The result is the same for PTRNAmark when the test set II is divided into two parts: NMR and X-ray. Besides when we are using ES of DI, the average value of FARNA decoys is lower a little than 3DRNAscore—ES of DI of PTRNAmark is 2.6, ES of DI of 3DRNAscore is 2.8—and another result is similar as ES of RMSD. These results suggest that PTRNAmark is better than other methods when it's used to rank near-native structures. A more detailed table than table 1 and energy-RMSD (DI) plots are provided in the supplementary data III and V.

The contribution of nucleotides' self-energy and retraining process

For PTRNAmark, on the one hand, unlike most potential for RNA which only consider the non-bond interaction between different nucleotides, PTRNAmark combines the energy of mutual nucleotides and nucleotides' inside. On the other hand, most statistical potentials for RNA are based on a constant training set. PTRNAmark uses two training sets, one of which is constant, and the others are varying. To verify the contribution of the term of nucleotides' self-energy and retraining process, we respectively tested the performance of energy of mutual nucleotides, total energy, and total energy including retraining process, using ES of RMSD on test set II. The results are shown in Table 2. On average, on the REMD decoys, normal mode decoys, FARNA decoys and overall average level, total energy without retraining process all performs better than mutual nucleotides' energy. For retraining process, the result is similar and the promotion of ES is stronger. And on average, on the REMD decoys, FARNA decoys and overall average level, total energy without retraining process all

performs better than other scoring function (See table 1). However, in some case, the method of retraining process cut ES down. If ten structures which are selected to be retraining set is ideal, the method of retraining process could promote statistical potential's precise (see supplementary data XI). But when the selecting ten structures are unsatisfactory, the case is reverse. That means that if the general method of score function is a little bit precise, the method of retraining process could enhance it to be stronger. But if the common method of score function is terrible, the method make it get worse. These results indicate that for most decoys nucleotides' self-energy and retraining process can further improve the accuracy of the distance-dependent energy on average.

Captured structural features

Common RNA base interactions typically explicitly represented in RNA potentials or force-fields are base-pairing interactions and base-stacking interactions. Figure 4 depicts the distance distribution of the atom pair between N1 of adenine and N3 of uracil. Two apparent peaks appear on the distance distribution. The first peak is at the distance of 3 Å, and it results from the base pairing interaction between adenine and uracil. The second peak is at the distance of 5.1 Å, and it stems from the base-stacking interaction between adjacent residues along the nucleotide residues chain. Figure5(A) shows the secondary structure and 3D structure of 434d, and figure 5(B) shows the base-stacking and base-pairing energies analyzed by PTRNAscore in RNA 434d. Besides the base-pairing energies between the nucleotides 1-14, 2-13, 3-12, 4-11, 6-9 and 5-6 are the six lowest energies, which exactly represent the four Watson–Crick base-pairing in 434d. What' more, for base-stacking energies between adjacent two nucleotides, the lower energy, the stronger base-stacking.

DISCUSSION

According to the thermodynamic hypothesis proposed by Anfinsen, the native structure tends to have the lowest free energy [28]. So accurately to say, using free energy to evaluate a structure is necessary. For classical molecular force fields, it is

light to calculate the enthalpy of a structure but for its entropy, it needs to consume too much time. In contrast, to force fields, both enthalpy and entropy information are contained statistical potentials extracted from experimental data of known RNA structures. Although they are not directly equal to free energy, they in principle correlate with the latter. Furthermore, it is easy and fast to calculation the energy.so it is useful, effective and efficient for RNA and protein structure scoring comparing with molecular force fields.

One of the considerable differences between PTRNAmark and other RNA statistical potentials is the association of the traditional different nucleotides' energy with the energy of nucleotides' inside. So PTRNAmark gives thought to both different nucleotides' non-bond interaction and the interaction of nucleotides' inside. It is common knowledge that molecular force fields are the basis of molecular dynamics simulation to study conformations of biomolecules. Many force fields consist of four components: bond stretching, angle bending, the rotation of bonds and non-bonded interactions. Furthermore, non-bonded interaction is made up two components: the interaction of different nucleotides and nucleotides' inside. The potential of mutual nucleotides corresponds to non-bonded interactions originating from different nucleotides, besides the potential of nucleotides' inside corresponds to non-bonded interactions being from nucleotides' inside and the rotation of bonds. Initially, we plan to consider all interaction of nucleotides' inside, but we find atom-pairs corresponding to bond stretching and angle bending have no significant impression. Yi Xiao's group found similar result too when they built 3DRNAscore[24]. What's more, the reason that the weight of energy of nucleotides' inside is too small may be that for RNA non-bonded interaction being from different nucleotides is stronger, and the diversity of conformation for a decoy set depends on this interaction in a large part.

For an ideal scoring function, it is very important that physical interactions are implicitly contained in distance-dependent statistical potential in RNA. The distance-dependent statistical potential corresponds to non-bonded interactions part of a force field. And for an atom-pairs, statistical potential energy implicitly contains not

only this atomic pair's non-bonded interactions but also other atoms' non-bond interactions being around this atom-pair. However, since the structural features of every RNA molecules are different, furthermore the frequency of structural features of every RNA molecules—like the frequency of the base pairing UN3 and AN1— is diverse for statistical process, so the non-bonded interactions which are implicitly contained in distance-dependent statistical potential should be of specificity(see supplementary data X). Figure6 shows that for base-pairing GO6-CN4 and base-stacking GC4-GC4, we have benchmarked the performance of phase matching between distance and energy of PTRNAmark, 3DRNAscore, KB, and RASP, comparing the native structure 1duq. Besides because of constant training sets for most statistical potential, the specificity of RNA structure features is not fully considered, Which means that there is a contradiction between universality and pertinence. And for a decoy set when you rank near-native structures, PTRNAmark uses a varying training set in second time scoring where structures belong to decoy set, furthermore we test the performance of PTRNAmark when we select the different counts of decoys to be a training set using a decoy set generated by EFOLD, we find that when we choose ten structures which are the lowest energy ranking by the first time scoring, PTRNAmark outperforms the other situation(see supplementary data). PTRNAmark fully considers the specificity of physical interactions in RNA by using this method, so the performance of phase matching of PTRNAmark is better than other statistical potential (see Figure6). And the specificity of RNA may be considered by using another method, and we still study the problem.

As for statistical potential, on the one hand, although knowledge-based statistical potential has been proved to be an efficient method, there exists an inherent limitation for the knowledge-based scoring function because it involves calculation of a reference state. An ideal reference states are not achievable, and the current methods to construct reference states are normally based on randomizing disconnected atoms and implicitly ignore excluded volume, sequences, and connectivity [45]. Therefore, the extracted potentials by these methods are not equal to the true potentials. Thomas and Dill built an iterative method to circumvent the calculation of the reference state

[46] and Xiao-Qin Zou et al. proposed a new, efficient iterative method to extract effective interaction potentials from a database of protein–ligand complex structures, which circumvents the calculation of the reference state [47，48]. And for RNA model this iterative method which circumvents the calculation of the reference state may be a more accurate method to assess RNA structures. On the other hand, since RNA backbone is highly (negatively) charged, the formation of the compact tertiary structure requires the polyanionic chain to overcome the massive charge repulsion from the backbone.so precisely to say, it is sensitive to electrostatic interactions for RNAs. And when RNAs are short of positive ions, it perhaps can be unable to form the functional folds [49-51]. The knowledge-based potentials only implicitly consider this effect by counting the experimental structures. And we do not know whether we more consider focusing electrostatic interactions for RNAs, we will study these problems in future.

CONCLUSION

In this work, we build a novel RNA knowledge-based potential for identifying native RNA structures and ranking near-native structures. We not only combine nucleotides' mutual and self energies but also fully consider the specificity of every RNA by using retraining process to retrain the parameters. Although some limitations, the benchmark tests show that our method could outperform existing methods in selecting native RNA structures and ranking near-native structures.

REFERENCE

1 Gesteland, R.F., Cech, T.R. & Atkins, J.F. The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 2006).

2 Guttman,M. and Rinn,J.L. (2012) Modular regulatory principles of large non-coding RNAs. Nature, 482, 339–346.

3 Dethoff,E.A., Chugh,J., Mustoe,A.M. and Al-Hashimi,H.M. (2012)Functional complexity and regulation through RNA dynamics.Nature, 482, 322–330.

4 Laing,C. and Schlick,T. (2011) Computational approaches to RNA structure prediction, analysis, and design. Curr. Opin. Struct. Biol.,21, 306–318.

5. Cao,S. and Chen,S.J. (2011) Physics-based de novo prediction of RNA 3D structures. J. Phys. Chem. B, 115, 4216–4226.

6. Das,R., Karanicolas,J. and Baker,D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. Nat.Methods, 7, 291–294.

7. Frellsen,J., Moltke,I., Thiim,M., Mardia,K.V., Ferkinghoff-Borg,J. and Hamelryck,T. (2009) A probabilistic model of RNA conformational space. PLoS Comput. Biol., 5, e1000406.

8. Sharma,S., Ding,F. and Dokholyan,N.V. (2008) iFoldRNA: three-dimensional RNA structure prediction and folding. Bioinformatics, 24, 1951–1952.

9. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature, 452, 51–55.

10. Martinez,H.M. Jr, Maizel,J.V. and Shapiro,B.A. (2008) RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. J. Biomol. Struct. Dyn., 25, 669–683.

11. Das,R. and Baker,D. (2007) Automated de novo prediction of native-like RNA tertiary structures. Proc. Natl. Acad. Sci. U.S.A.,104, 14664–14669.

12. Zhao,Y., Huang,Y., Gong,Z., Wang,Y., Man,J. and Xiao,Y. (2012) Automated and fast building of three-dimensional RNA structures. Sci. Rep., 2, 734.

13. Zhao,Y., Gong,Z. and Xiao,Y. (2011) Improvements of the hierarchical approach for predicting RNA tertiary structure. J. Biomol. Struct. Dyn., 28, 815–826.

14. Zhang,Y. (2008) Progress and challenges in protein structure prediction. Curr. Opin. Struct. Biol., 18, 342–348.

15. Zhou,H. and Skolnick,J. (2011) GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. Biophys. J., 101, 2043–2052.

16. Adhikari,A.N., Freed,K.F. and Sosnick,T.R. (2013) Simplified protein models: Predicting folding pathways and structure using amino acid sequences. Phys. Rev. Lett., 111, 28103.

17. Capriotti,E., Norambuena,T.,Marti-Renom,M.A. and Melo,F. (2011) All-atom knowledge-based potential for RNA structure prediction and assessment. Bioinformatics, 27, 1086–1093.

18. Bernauer,J., Huang,X., Sim,A.Y.L. and Levitt,M. (2011) Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. RNA, 17, 1066–1075.

19. Norambuena,T., Cares,J.F., Capriotti,E. and Melo,F. (2013) WebRASP: a server for computing energy scores to assess the accuracy and stability of RNA 3D structures. Bioinformatics, 29, 2649–2650.

20. Sim,A.Y.L., Schwander,O., Levitt,M and Bernauer,J. (2012) evaluating mixture models for building rna knowledge-based potentials. J. Bioinform. Comput. Biol., 10, 1241010.

21. Olechnovic,K. and Venclovas,C. (2014) The use of interatomic contact areas to quantify discrepancies between RNA 3D models and reference structures. Nucleic Acids Res., 42, 5407–5415.
22. Melo,F., Nchez,R. and Sali,A. (2002) Statistical potentials for fold assessment. Protein Sci., 11, 430–448.
23. Neal,R.M. (2000) Markov chain sampling methods for Dirichlet process mixture models. J. Comp. Graph. Stat., 9, 249–265.
24. J. Wang, Y. Zhao, C. Zhu, Y. Xiao, 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures, Nucleic acids research, 43 (2015) e63.
25. H. Deng, Y. Jia, Y. Wei, Y. Zhang, What is the best reference state for designing statistical atomic potentials in protein structure prediction, Proteins, 80 (2012) 2311-2322.
26. Sippl,M.J. (1990) Calculation of conformational ensembles from potentials of mena force: an approach to the knowledge-based prediction of local structures in globular proteins. J. Mol. Biol., 213, 859–883.
27. Samudrala,R. and Moult,J. (1998) An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J. Mol. Biol., 275, 895–916.
28. Anfinsen,C.B. (1973) Principles that govern the folding of protein chains. Science, 181, 223–230.
29. Lu,H. and Skolnick,J. (2001) A distance-dependent atomic knowledge-based potential for improved protein structure selection. Proteins: Struct. Funct. Bioinform., 44, 223–232.
30. Rykunov,D. and Fiser,A.A.S. (2007) Effects of amino acid composition, finite size of proteins, and sparse statistics on distance-dependent statistical pair potentials. Proteins: Struct. Funct. Bioinform., 67, 559–568.
31. Shi, Y.Z., Wang, F.H., Wu, Y.Y. and Tan, Z.J. (2014) A coarse-grained model with implicit salt for RNAs: predicting 3D structure, stability and salt effect. The Journal of chemical physics, 141, 105102.
32. Scott,D.W. (1979) On optimal and data-based histograms. Biometrika, 66, 605–610.
33. Leontis,N.B.Z.C. (2012) . Nonredundant 3D Structure Datasets for RNA Knowledge Extraction and Benchmarking Springer, Berlin, Heidelberg27, 281–298.
34. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. J. Mol. Biol., 215, 403–410.
35. Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol., 234, 779–815.
36. Huang,E.S., Subbiah,S., Tsai,J. and Levitt,M. (1996) Using a hydrophobic contact potential to evaluate native and near-native folds generated by molecular dynamics simulations. J. Mol. Biol., 257, 716–725.
37. Hansmann,U.H. and Okamoto,Y. (1999) New Monte Carlo algorithms for protein folding. Curr. Opin. Struct. Biol., 9, 177–183.
38. Summa,C.M. and Levitt,M. (2007) Near-native structure refinement using in vacuo energy minimization. Proc. Natl. Acad. Sci. U.S.A., 104, 3177–3182.
39. Parisien,M., Cruz,J.A., Westhof,E. and Major,F. (2009) New metrics for comparing and assessing discrepancies between RNA 3D structures and models. RNA, 15, 1875–1885.
40. Gorodkin,J., Stricklin,S.L. and Stormo,G.D. (2001) Discovering common stem–loop motifs in unaligned RNA sequences. Nucleic Acids Res., 29, 2135–2144.
41. Matthews,B.W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. BBA-Protein Struct. Mol., 405, 442–451.

42. Jonikas,M.A., Radmer,R.J., Laederach,A., Das,R., Pearlman,S., Herschlag,D. and Altman,R.B. (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. RNA, 15, 189–199.
43. Wang,J., Cieplak,P. and Kollman,P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem., 21, 1049–1074.
44. Tsai,J., Bonneau,R., Morozov,A.V., Kuhlman,B., Rohl,C.A. and Baker,D. (2003) An improved protein decoy set for testing energy functions for protein structure prediction. Proteins, 53, 76–87.
45.Thomas, P. D.; Dill, K. A. J Mol Biol 1996, 257, 457.
46. Thomas, P. D.; Dill, K. A. Proc Natl Acad Sci USA 1996, 93, 11628.
47. Huang, S.Y. and Zou, X. (2006) An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. Journal of computational chemistry, 27, 1866-1875.
48. Huang, S.Y. and Zou, X. (2006) An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. Journal of computational chemistry, 27, 1876-1882.
49. Chen,S. (2008) RNA folding: conformational statistics, folding kinetics, and ion electrostatics. Annu. Rev. Biophys., 37, 197.
50. Chin,K., Sharp,K.A., Honig,B. and Pyle,A.M. (1999) Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. Nat. Struct. Mol. Biol., 6, 1055–1061.
51. Draper,D.E. (2013) Folding of RNA tertiary structure: linkages between backbone phosphates, ions, and water. Biopolymers, 99, 1105–1113.

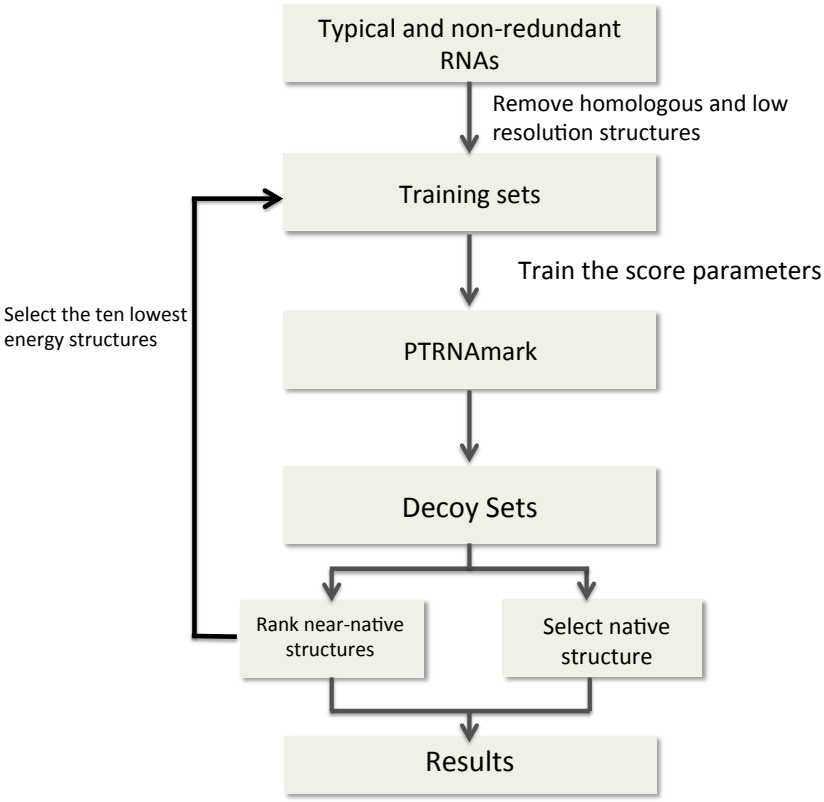

Figure1. Flow chart of training steps of PTRNAmark

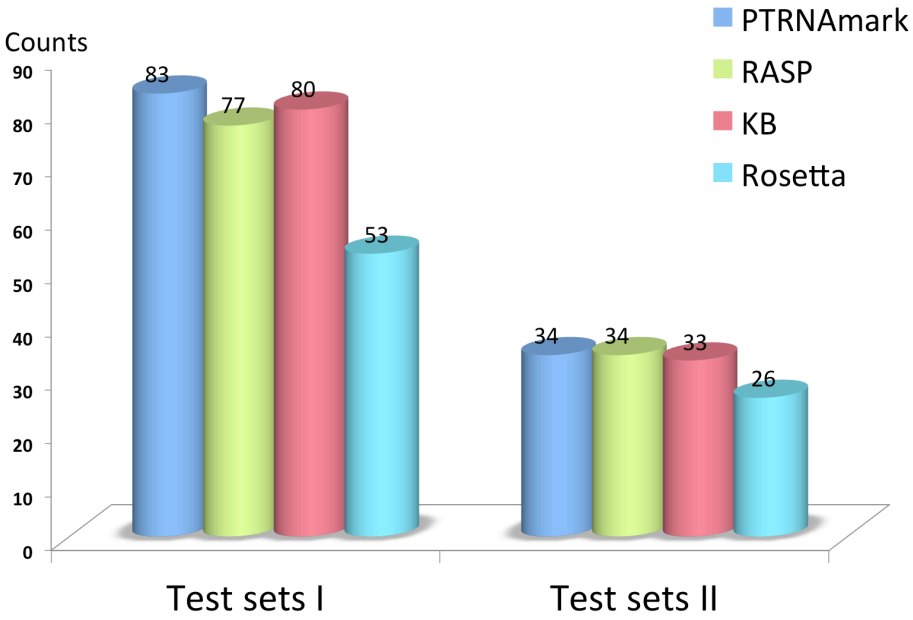

Figure 2. Counts of native states identified correctly from test set I and test set II by PTRNAmark, RASP, KB and Rosetta, respectively.

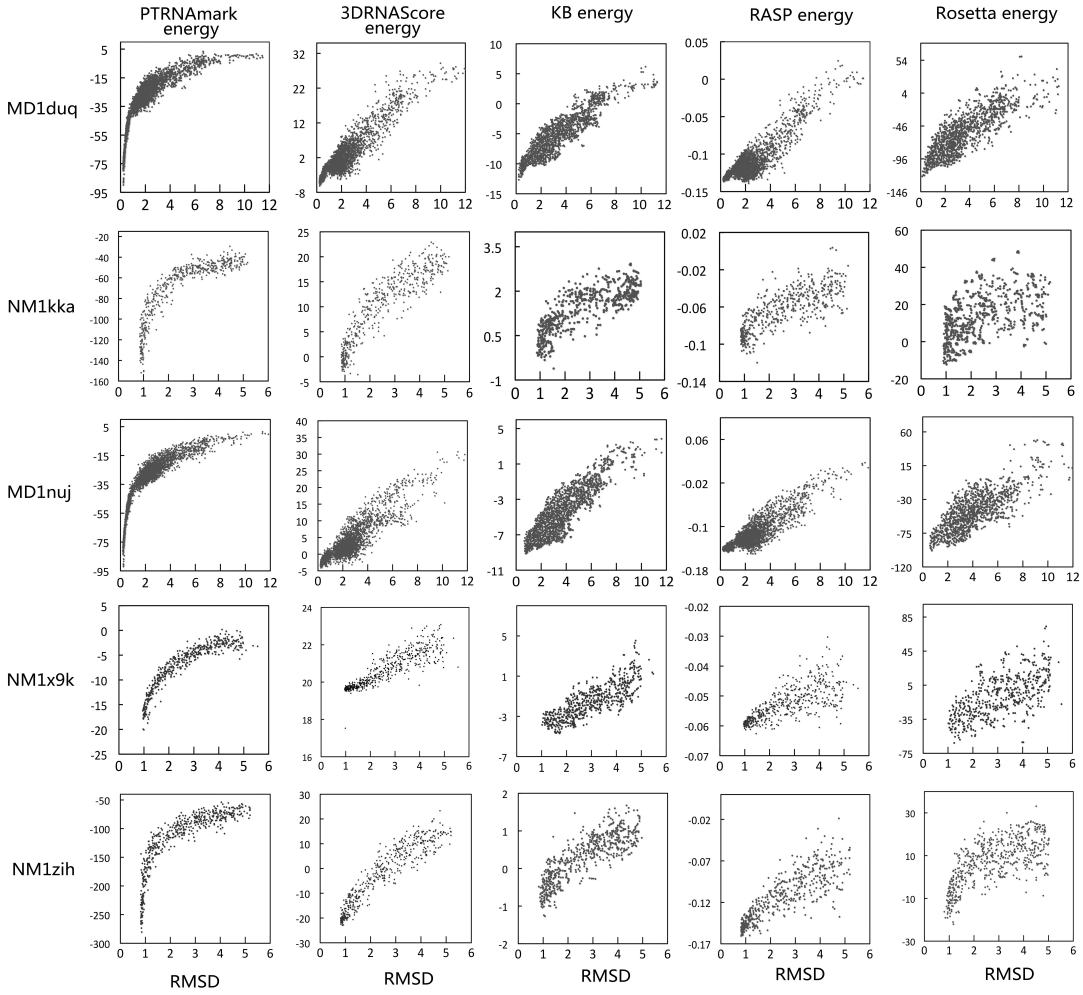

Figure3.Energy as a function of RMSD for decoys generated using MD decoys and NM decoys for five different systems. PTRNAscore, 3DRNAscore, KB, RASP and Rosetta are shown in turns respectively

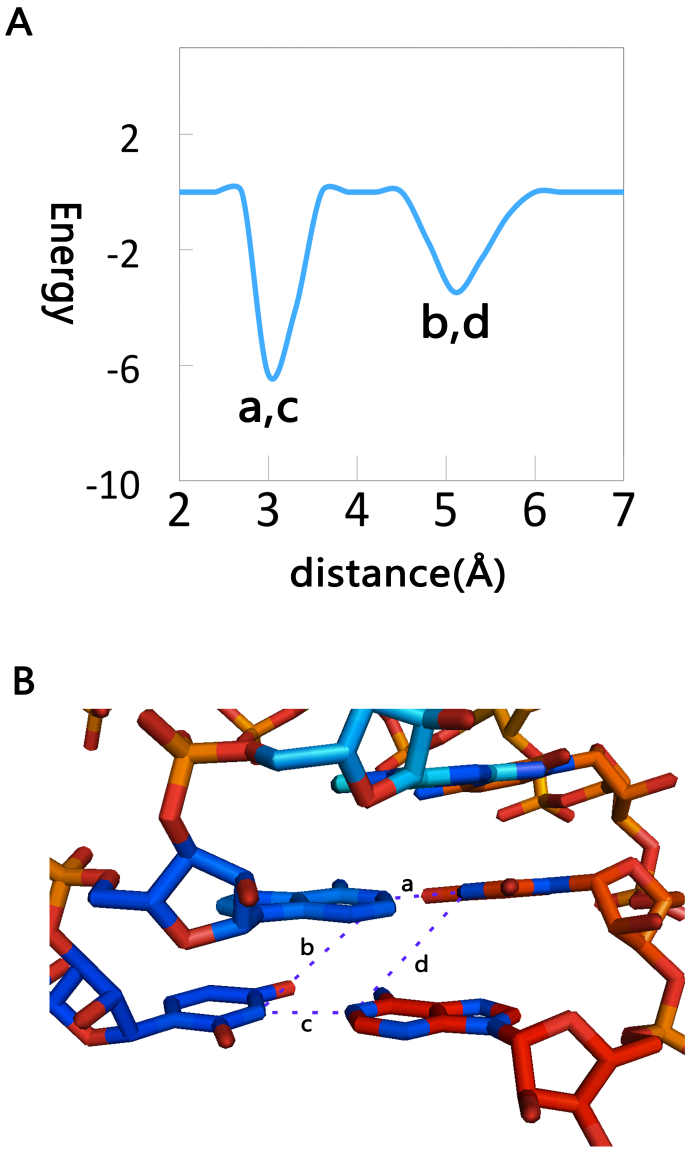

Figure4. (A) Energy distribution of the distance between N1 of adenine and N3 of uracil. (B) Diagram of the four representative distance between N1 of adenine and N3 of uracil.

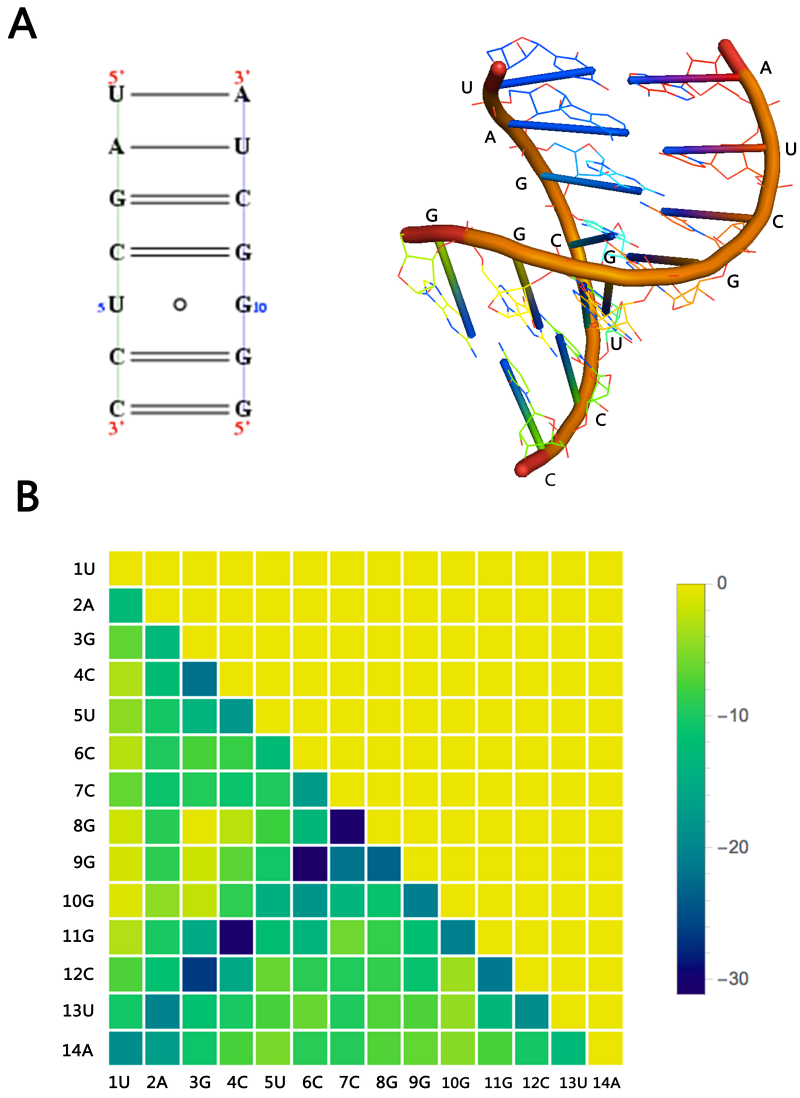

Figure 5. (A) Secondary structure and 3D structure of 434D.(B) Base-stacking and base-pairing energies between each possible base-pair in 434D calculated by PTRNAscore. The lower the energy, the stronger interaction.

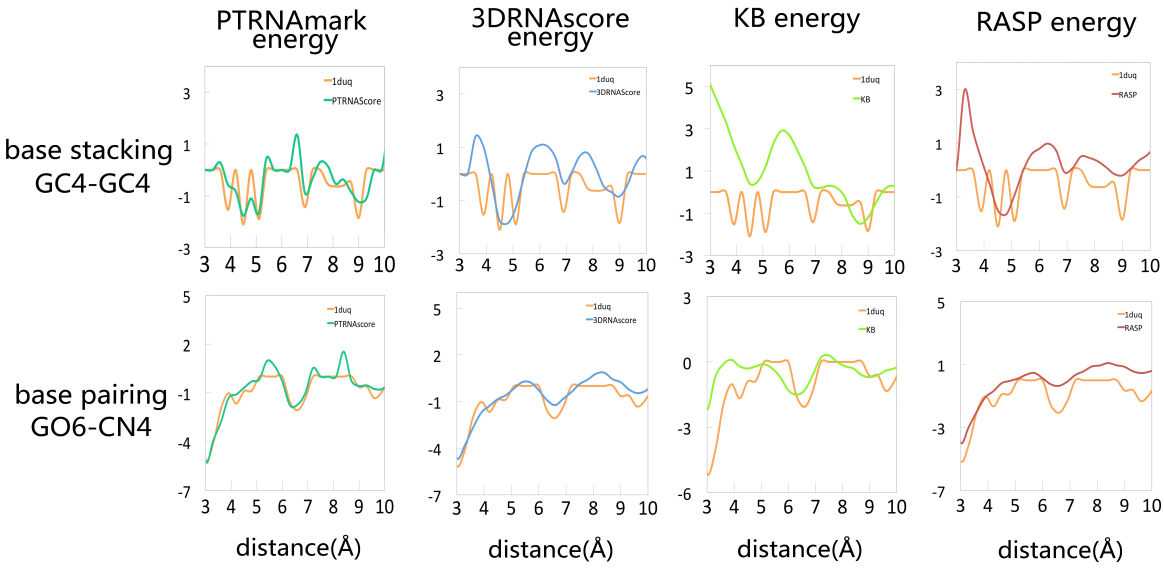

Figure6. Energy as a function of distance for base stacking GC4-GC4 and base pairing GO6-CN4, compaing the native structure in RNA 1DUQ. PTRNAmark, 3DRNAscore,KB, and RASP are shown in turns respectively.

Table1. Comparison of the performance of ranking near native structures of PTRNAmark, 3DRNAscore, KB potential and Rosetta methods in test set II

Decoy	RNA	Length	Method	Enrichment score (RMSD)					Enrichment score (DI)				
				PTRNAScore	KB	RASP	3DRNAScore	Rosetta	PTRNAScore	KB	RASP	3DRNAScore	Rosetta
Position restrained dynamics and REMD(A)	1duq	26	X-ray	9.5	7.6	7.6	8.5	7.1	9.6	7.5	7.6	8.3	7.0
	1f27	30	X-ray	9.6	7.9	6.6	8.3	6.2	9.6	7.8	6.6	8.1	6.2
	1msy	27	X-ray	9.3	5.7	5.7	7.5	3.6	9.2	6.0	5.6	7.6	3.5
	1nuj	24	X-ray	9.5	7.3	5.2	7.7	6.9	9.5	7.2	5.2	7.4	6.7
	434d	14	X-ray	9.4	7.7	7.0	8.0	6.8	9.4	7.7	6.9	7.6	6.8
Normal modes(B)	1duq	26	X-ray	8.4	7.0	5.7	7.5	3.8	8.2	7.0	5.7	6.9	3.5
	1esy	19	NMR	7.9	5.4	4.5	4.9	5.6	7.6	5.5	4.7	6.5	5.7
	1f27	30	X-ray	8.6	5.8	3.7	5.9	2.6	8.8	5.8	3.7	6.7	2.5
	1i9v	76	X-ray	8.2	2.6	5.3	6.1	3.0	8.2	2.7	5.1	5.5	3.0
	1kka	17	NMR	6.9	4.6	4.1	5.7	4.6	7.4	4.3	4.1	6.9	4.8
	1msy	27	X-ray	9.0	5.6	2.2	5.9	4.6	9.2	5.4	2.7	6.7	4.5
	1nuj	24	X-ray	8.6	7.4	5.9	7.1	2.4	8.6	7.0	5.9	4.5	2.2
	1qwa	21	NMR	5.1	3.2	2.0	3.5	3.8	4.5	3.3	2.2	3.7	3.9
	1x9k	62	X-ray	8.5	1.6	5.2	6.7	3.0	8.6	1.9	5.2	3.1	2.8
	1xjr	46	X-ray	8.7	5.4	7.9	7.7	2.2	8.5	5.0	7.9	6.2	2.5
	1ykq	19	X-ray	8.3	3.4	3.5	4.6	2.8	8.1	3.3	3.8	5.0	2.5
	1zih	12	NMR	8.4	5.4	5.7	7.7	6.6	8.0	5.3	5.7	6.9	6.5
	28sp	28	NMR	8.6	4.0	6.5	5.7	1.8	6.4	4.5	6.7	5.7	2.9
	2f88	34	NMR	8.4	5.4	4.9	6.8	4.4	7.9	5.4	4.7	4.1	4.2
	434d	14	X-ray	7.3	7.4	7.4	7.7	5.2	7.6	7.4	7.6	6.9	5.2
FARNA(C)	1a4d	41	NMR	5.2	3.8	2.0	2.4	0.8	6.8	4.0	2.0	2.1	0.7
	1csl	28	X-ray	2.0	1.5	1.6	2.0	1.3	3.4	1.4	1.6	2.4	1.4
	1dqf	19	X-ray	4.8	1.8	2.8	4.2	1.0	1.4	2.0	3.2	3.6	1.0
	1esy	19	NMR	4.4	3.7	4.8	3.2	1.2	2.8	3.5	4.2	3.8	1.1
	1i9x	26	X-ray	2.0	1.3	2.2	3.4	1.5	1.4	1.5	3.6	4.0	1.6
	1j6s	24	X-ray	0.6	1.4	0.2	0.2	0.6	1.8	1.7	0.8	1.6	0.5
	1kd5	22	X-ray	1.0	0.3	0.8	3.2	0.2	3.2	0.5	1.6	2.4	0.3
	1kka	17	NMR	0.6	1.2	0.6	1.4	0.6	0.6	0.8	0.8	2.0	0.6
	1l2x	27	X-ray	0.4	3.2	1.0	0.6	1.8	2.6	3.2	1.2	1.5	1.8
	1mhk	32	X-ray	0.6	1.2	1.6	1.6	1.0	0.2	1.2	1.4	1.4	1.2
	1q9a	27	X-ray	2.6	0.5	0.8	2.6	0.8	0.4	0.5	1.2	3.2	0.8
	1qwa	21	NMR	1.4	1.2	0.4	2.2	1.0	3.2	1.3	0.6	1.4	1.0
	1xjr	46	X-ray	3.6	2.0	2.4	2.4	1.2	3.4	1.9	3.4	3.2	1.0
	1zih	12	NMR	7.4	5.0	4.8	4.8	2.0	7.6	5.5	7.2	6.8	1.9
	255d	24	X-ray	1.6	0.7	0.6	2.4	1.3	2.2	0.7	0.6	2.0	1.1
	283d	24	X-ray	3.0	0.8	0.8	1.4	0.7	1.8	0.8	1.0	1.8	0.7
	28sp	28	NMR	3.8	1.5	3.0	2.8	1.7	1.8	1.2	4.2	3.8	1.8
	2a43	36	X-ray	3.2	2.0	1.4	2.2	0.6	2.6	1.8	2.0	3.2	0.6
	2f88	34	NMR	3.6	1.3	2.2	3.2	1.3	2.6	1.3	1.6	3.6	1.2
Average value		(A)		9.5	7.2	6.4	8.0	6.1	9.5	7.2	6.4	7.8	6.0
		(B)		8.1	4.9	5.0	6.2	3.8	7.8	4.9	5.0	5.7	3.8
		(C)		2.7	1.8	1.8	2.3	1.1	2.6	1.8	2.2	2.8	1.1
		X-ray		5.9	3.8	3.7	4.8	2.8	5.7	3.8	3.9	4.7	2.7
		NMR		5.5	3.5	3.5	4.2	2.7	5.2	3.5	3.7	4.4	2.8
		All		5.6	3.7	3.6	4.5	2.7	5.5	3.7	3.8	4.6	2.7

Table2. Comparison of performance of mutual nucleotides' energy, total energy, and total energy with retraining process of PTRNAmark in test set II.

Decoy	RNA	Length	Method	Enrichment score(RMSD)		
				mutual nucleotides	Total	Retraining
Position restrained dynamics and REMD(A)	1duq	26.0	X-ray	8.2	8.3	9.5
	1f27	30.0	X-ray	8.6	8.7	9.6
	1msy	27.0	X-ray	7.3	7.5	9.3
	1nuj	24.0	X-ray	7.8	7.9	9.5
	434d	14.0	X-ray	8.2	8.4	9.4
Normal modes(B)	1duq	26.0	X-ray	7.5	7.5	8.4
	1esy	19.0	NMR	4.4	4.7	8.0
	1f27	30.0	X-ray	6.7	6.7	8.6
	1i9v	76.0	X-ray	6.5	6.5	8.2
	1kka	17.0	NMR	4.6	4.7	6.9
	1msy	27.0	X-ray	5.7	5.9	9.0
	1nuj	24.0	X-ray	7.1	7.1	8.6
	1qwa	21.0	NMR	3.2	3.3	5.3
	1x9k	62.0	X-ray	6.2	6.2	8.5
	1xjr	46.0	X-ray	7.6	7.7	8.7
	1ykq	19.0	X-ray	3.9	3.9	8.3
	1zih	12.0	NMR	7.7	7.7	8.4
	28sp	28.0	NMR	5.7	5.9	8.6
	2f88	34.0	NMR	7.1	7.2	8.4
	434d	14.0	X-ray	7.2	7.7	7.3
FARNA(C)	1a4d	41.0	NMR	4.4	4.4	5.2
	1csl	28.0	X-ray	1.9	2.0	2.0
	1dqf	19.0	X-ray	5.3	5.4	4.8
	1esy	19.0	NMR	3.1	3.2	4.4
	1i9x	26.0	X-ray	2.5	2.4	2.0
	1j6s	24.0	X-ray	1.5	2.2	0.6
	1kd5	22.0	X-ray	1.1	1.2	1.0
	1kka	17.0	NMR	1.5	1.6	0.6
	1l2x	27.0	X-ray	0.3	0.4	0.4
	1mhk	32.0	X-ray	0.9	1.2	0.6
	1q9a	27.0	X-ray	1.3	1.4	2.6
	1qwa	21.0	NMR	2.9	3.2	1.4
	1xjr	46.0	X-ray	3.5	3.6	3.6
	1zih	12.0	NMR	6.2	6.2	7.4
	255d	24.0	X-ray	2.3	2.4	1.6
	283d	24.0	X-ray	1.1	1.2	3.0
	28sp	28.0	NMR	1.3	1.4	3.8
	2a43	36.0	X-ray	1.7	1.8	3.2
	2f88	34.0	NMR	1.1	1.2	3.6
Average value		(A)		8.0	8.2	9.5
		(B)		6.1	6.2	8.1
		(C)		2.3	2.4	2.7
		X-ray		4.7	4.8	5.9
		NMR		4.0	4.2	5.5
		All		4.5	4.6	5.6