

Enhancements of the Gaussian network model in describing nucleotide residue fluctuations for RNA

Wen-Jing Wang(王文静), Ji-Guo Su(苏计国)

Citation: Chin. Phys. B, 2021, 30 (**5**): 058701. DOI: 10.1088/1674-1056/abe1a5
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Statistical potentials for 3D structure evaluation: From proteins to RNAs

Ya-Lan Tan(谭雅岚), Chen-Jie Feng(封晨洁), Xunxun Wang(王勋勋), Wenbing Zhang(张文炳), and Zhi-Jie Tan(谭志杰)

Chin. Phys. B, 2021, 30 (**2**): 028705. DOI: 10.1088/1674-1056/abc0d6

The theory of helix-based RNA folding kinetics and its application

Sha Gong(龚沙), Taigang Liu(刘太刚), Yanli Wang(王晏莉), and Wenbing Zhang(张文炳)†

Chin. Phys. B, 2020, 29 (**10**): 108703. DOI: 10.1088/1674-1056/abab84

Molecular dynamic simulation of the thermodynamic and kinetic properties of nucleotide base pair

Yu-Jie Wang(王宇杰), Zhen Wang(王珍), Yan-Li Wang(王晏莉), Wen-Bing Zhang(张文炳)

Chin. Phys. B, 2017, 26 (**12**): 128705. DOI: 10.1088/1674-1056/26/12/128705

Confinement of spherical colloid particles in a soft fluid membrane tube

Zheng Bin, Meng Qing-Tian

Chin. Phys. B, 2014, 23 (**3**): 038701. DOI: 10.1088/1674-1056/23/3/038701

Symmetrical adhesion of two cylindrical colloids to a tubular membrane

Niu Yu-Quan, Wei Wei, Zheng Bin, Zhang Cai-Xia, Meng Qing-Tian

Chin. Phys. B, 2013, 22 (**12**): 128701. DOI: 10.1088/1674-1056/22/12/128701

RAPID COMMUNICATION

Enhancements of the Gaussian network model in describing nucleotide residue fluctuations for RNA

Wen-Jing Wang(王文静) and Ji-Guo Su(苏计国)†

Key Laboratory for Microstructural Material Physics of Hebei Province, School of Science, Yanshan University, Qinhuangdao 066004, *China*

(Received 7 December 2020; revised manuscript received 10 January 2021; accepted manuscript online 1 February 2021)

Gaussian network model (GNM) is an efficient method to investigate the structural dynamics of biomolecules. However, the application of GNM on RNAs is not as good as that on proteins, and there is still room to improve the model. In this study, two novel approaches, named the weighted GNM (wGNM) and the force-constant-decayed GNM (fcdGNM), were proposed to enhance the performance of ENM in investigating the structural dynamics of RNAs. In wGNM, the force constant for each spring is weighted by the number of interacting heavy atom pairs between two nucleotides. In fcdGNM, all the pairwise nucleotides were connected by springs and the force constant decayed exponentially with the separate distance of the nucleotide pairs. The performance of these two proposed models was evaluated by using a non-redundant RNA structure database composed of 51 RNA molecules. The calculation results show that both the proposed models outperform the conventional GNM in reproducing the experimental *B*-factors of RNA structures. Compared with the conventional GNM, the Pearson correlation coefficient between the predicted and experimental *B*-factors was improved by 9.85% and 6.76% for wGNM and fcdGNM, respectively. Our studies provide two candidate methods for better revealing the dynamical properties encoded in RNA structures.

Keywords: Gaussian network model (GNM), *B*-factor, RNA molecules

PACS: 87.15.ad, 87.14.gn, 87.15.hp **DOI:** 10.1088/1674-1056/abe1a5

1. Introduction

RNA is one of the important biological macromolecules in the cell. For a long time, RNA is considered to be only served as a messenger that intermediates the transmission of the genetic information from DNA to protein sequence.[1] However, in recent decades more and more evidences have revealed that the biological functions carried out by RNA are much broader than we previously believed. It is increasingly realized that RNA can perform various functions including catalysis of biochemical reactions, regulation of biological processes, transduction of cellular signals, synthesis of proteins, post-translational modification of proteins, regulation of gene expressions, and so on.[1–4] Similar to proteins, RNA can fold into subtle tertiary structures, and the biological functions of RNA are essentially determined by its specific three-dimensional structure and the intrinsic dynamics encoded into the structure.[3–9] How to effectively extract the dynamical properties from the tertiary structure of RNA is critical for our understanding of RNA functions.

Elastic network model (ENM) describes a biomacromolecule as a group of nodes connected by springs, which is a simple and efficient method to investigate the intrinsic dynamics of biological macromolecules based on their native structures. ENM has been extensively used to analyze the conformational dynamics of proteins.[10–14] Many studies have proved that the residue fluctuations in proteins calculated by ENM are well consistent with the *B*-factors measured by x-ray crystallographic experiments as well as the results obtained by molecular dynamics (MD) simulations.[15] In addition, ENM has also been successfully applied in investigating the function-related collective motions, predicting folding cores and functional hotspots, identifying allosteric signaling pathways, enhancing conformational sampling in MD simulations, and realizing protein–protein soft docking.[16–21] In recently years, ENM has been extended to study the structure-encoded dynamic properties for RNA. Yang *et al*. constructed ENM for DNA/RNA-containing systems and the computed nucleotide fluctuations by ENM exhibit a quite good positive correlation with the experimental *B* factors.[22] Pinamonti *et al*. systematically assessed the performance of ENM by using different types of RNA structures, and they found that this simple model can properly reproduce the structural fluctuations obtained by all-atomistic MD simulations and SHAPE experiments.[23] The study of Zimmermann and Jernigan showed that the low frequency normal modes computed by ENM are capable of capturing the principal components of the structural flexibility determined by NMR experiments for dense-packed RNA molecules.[24] ENM was also utilized by Wang *et al*. to study the functional cooperative motions for large RNA-containing systems, and the results showed that ENM can provide valuable insights into RNA functional dynamics.[25] The study of

†Corresponding author. E-mail: jiguosu@ysu.edu.cn

© 2021 Chinese Physical Society and IOP Publishing Ltd

Afonin *et al*. indicated that ENM can characterize the experimentally observed dynamic behaviors of large RNA cubic nanoscaffolds.[26]

Although the usefulness of ENM in characterizing RNA dynamics has been validated, some studies also pointed out that the correlation between the ENM-computed and experimental residue fluctuations is lower for RNAs than that of proteins.[27] It is believed that the lower fluctuation correlation for RNA should be attributed to its loosely-packed tertiary structure compared with the well-packed protein structure.[27–29] In the conventional ENM, the interactions between residues are simplified as springs with a uniform force constant, and the existence of springs between residues is determined by their pairwise distance less than a given cutoff value.[15,18,22,29,30] Several improved strategies in constructing ENM have been proposed to enhance the performance of the model for RNA. For the nucleotide representation in ENM, the research work of Pinamonti *et al*. showed that the all-atomistic and three-nodes-per-nucleotide models are better than the single-node-per-nucleotide model in investigating RNA structural dynamics.[23] Diggins *et al*. further explored that the coarse-grained site selection in nucleotide representation also influences the performance of the ENM.[31] For the inter-nucleotide interactions in ENM, Mailhot *et al*. adopted a sequence-dependent potential function, which takes into account the chemical properties of the residues, to construct the elastic network contact model (ENCoM). It is found that ENCoM can better simulate the dynamics of RNA compared with the conventional ENM.[32] Moreover, the study of Wang *et al*. illustrated that a multiscale ENM, in which the multiscale harmonic interactions with variable force constants were adopted, distinctly improves the characterization of the dynamical properties of RNA. However, in the multiscale ENM, the force constants need to be optimized by using the experimental *B*-factors for each studied RNA molecule separately, and thus the optimized model for one RNA molecule is not generally applicable for other RNA systems.[33]

In the present work, two novel approaches are proposed to improve the performance of the ENM in describing the structural flexibility of RNA molecules. As mentioned above, in the conventional ENM the interaction strength is assumed to be identical for all the nucleotide pairs. In this study, the difference in the inter-nucleotide interaction strength is taken into account to build a weighted ENM (wENM), in which the force constant for each spring is weighted by the number of interacting heavy atom pairs between the two related nucleotides. Besides that, another simplification in the conventional ENM is that the inter-nucleotide interactions are neglected for the nucleotide pairs beyond a given cutoff distance. However, the nucleotides in RNA molecules are negatively charged, and the long-range electrostatic interactions are important for RNA structural dynamics. The construction of ENM based on a cutoff distance in the conventional method for proteins may be not suitable for RNA molecules. Aimed at this problem, another new ENM-building method, named force-constant-decayed ENM (fcdENM), is proposed. In our method, the distance cutoff is removed and all residue pairs are connected by springs with the force constant decayed exponentially with the separate distance. The performance of these two proposed models (wENM and fcdENM) is tested by using a large non-redundant RNA structure dataset. Our calculation results show that both the proposed models outperform the conventional ENM in prediction of the experimental *B*-factors.

2. Materials and methods

2.1. The conventional Gaussian network model

Gaussian network model (GNM) is one type of the ENM, which is built based on the coordinates of RNA structures downloaded from the Protein Data Bank (PDB). In GNM, each nucleotide in RNA structure is treated as a node located on its P atom. The inter-nucleotide interactions are represented by a linear harmonic potential, where each node in the model is connected by springs to other nodes within a given cutoff distance and the force constant for all the springs is identical.[34] The potential function of the GNM for a RNA molecule is the sum of the harmonic potential energies of all the springs in the model, which can be written as

$$\begin{aligned} V &= \frac{1}{2}\gamma\sum_{ij}(\Delta R_{ij})^2 = \frac{1}{2}\gamma\sum_{ij}(\Delta R_j - \Delta R_i)^2 \\ &= \frac{1}{2}\gamma\left(\Delta\boldsymbol{R}^{\mathrm{T}}\boldsymbol{\Gamma}\Delta\boldsymbol{R}\right), \end{aligned} \tag{1}$$

where $\Delta R_{ij} = R_{ij} - R_{ij}^0 = \Delta R_j - \Delta R_i$; R_{ij} and R_{ij}^0 are respectively the instantaneous and the equilibrium distances between the i^{th} and j^{th} nucleotides; ΔR_i and ΔR_j are the displacements from the equilibrium position for the i^{th} and j^{th} nucleotides, respectively; ΔR_{ij} is the change in the distance of the pairwise nucleotides; γ is the force constant of the springs in the model; $\Delta\boldsymbol{R} = \{\Delta R_i\}$, $i = 1,2,\ldots,N$, which represents the N-dimensional displacement vector and N is the nucleotide number in the RNA system; The superscript T denotes the transpose of the column vector; and $\boldsymbol{\Gamma}$ is a $N \times N$ symmetric matrix named as Kirchhoff matrix,[34] which can be built based on the inter-nucleotide contacts in the RNA structure as follows:

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j,\ R_{ij} \leq R_{\mathrm{c}}, \\ 0, & \text{if } i \neq j,\ R_{ij} > R_{\mathrm{c}}, \\ -\sum\limits_{k,k\neq i}^{N}\Gamma_{ik}, & \text{if } i = j, \end{cases} \tag{2}$$

here R_{c} is the cutoff distance, which is used to determine whether or not there exist inter-node springs.

According to the normal mode analysis theory, the dynamical properties of the RNA system are fully determined by the Kirchhoff matrix $\boldsymbol{\Gamma}$. The normal modes of the system can be calculated through diagonalization of the matrix $\boldsymbol{\Gamma}$, and the details of the calculation can be found in the references.[15,18,22,29,30,34] Based on the statistical thermodynamic theory, the mean-square fluctuation for each nucleotide can be computed as

$$\langle(\Delta R_i)^2\rangle = \frac{\int (\Delta R_i \Delta R_i) \exp\left(-\frac{V}{k_B T}\right) \mathrm{d}\{\Delta \boldsymbol{R}\}}{\int \exp\left(-\frac{V}{k_B T}\right) \mathrm{d}\{\Delta \boldsymbol{R}\}} = \frac{k_B T}{\gamma}\left[\boldsymbol{\Gamma}^{-1}\right]_{ii} = \frac{k_B T}{\gamma}\sum_{k=2}^{N} \lambda_k^{-1}[u_k]_i[u_k]_i, \quad (3)$$

where V is the potential function of the system given in Eq. (1); k_B is the Boltzmann constant; T is the temperature; u_k and λ_k are the k^{th} eigenvector as well as the corresponding eigenvalue of $\boldsymbol{\Gamma}$, which represent the k^{th} normal mode and the square of the frequency of the mode, respectively. Then, the B-factor for each nucleotide can be obtained by using the Debye–Waller formula

$$B_i = \frac{8\pi^2}{3}\langle(\Delta R_i)^2\rangle = \frac{8\pi^2 k_B T}{3\gamma}\sum_{k=2}^{N} \lambda_k^{-1}[u_k]_i[u_k]_i. \quad (4)$$

In the present work, the theoretical B-factors calculated by Eq. (4) are compared with the experimental data obtained by x-ray crystallographic method to assess the performance of the conventional GNM as well as our proposed new models. The Pearson correlation coefficient (PCC) is used to evaluate the accuracy of the predicted B-factors relative to the experimental data. The PPC between the predicted and experimental B-factors can be expressed as

$$\mathrm{PCC} = \frac{\sum_{i=1}^{N}\left(B_i^{\text{pred}} - \bar{B}^{\text{pred}}\right)\left(B_i^{\text{exp}} - \bar{B}^{\text{exp}}\right)}{\sqrt{\sum_{i=1}^{N}\left(B_i^{\text{pred}} - \bar{B}_i^{\text{pred}}\right)^2}\sqrt{\sum_{i=1}^{N}\left(B_i^{\text{exp}} - \bar{B}_i^{\text{exp}}\right)^2}}, \quad (5)$$

where B_i^{pred} and B_i^{exp} are respectively the predicted and experimental B-factors of the i^{th} nucleotide; $\bar{B}^{\text{pred}}$ and $\bar{B}^{\text{exp}}$ are the average values of the predicted and experimental B-factors of all the nucleotides in the RNA structure, respectively; and N is the total number of nucleotide in the RNA structure.

2.2. Weighted Gaussian network model

In the conventional GNM, the same force constant is adopted for all the springs in the model. However, RNA structure and dynamics are controlled by various inter-nucleotide interactions, in which the covalent interactions as well as the base–base paring and stacking are much stronger than other interactions.[35–38] The various interactions with different intensities play different roles in controlling RNA structural stability and dynamics.[35–38] For better characterizing RNA structural dynamics, the variation of interaction strength should be considered. In the present work, a weighted GNM (wGNM) was proposed, in which the differences in the strength of the inter-nucleotide interactions was taken into account by using different weights. In our model, the force constant for each spring was weighted simply by the number of atom–atom contacts between the corresponding pairwise nucleotides.

The proposed wGNM was built according to the following procedure.

(1) Based on the coordinates of all the heavy atoms in the RNA structure, the atom–atom contacts were determined by a given atomistic cutoff distance. An atom was considered to form a contact with another atom if their distance is less than the cutoff value. Otherwise, no contact exists between them.

(2) Similar to the conventional GNM, each nucleotide in the RNA structure was simplified as a node located on its P atom to construct the nucleotide-level coarse-grained model. But different from the conventional GNM, whether or not a spring exists between pairwise nucleotides was determined by the atomistic contacts between them, instead of the nucleotide-level distance cutoff. If a heavy atom in a nucleotide forms contact with any heavy atom of another nucleotide, these two nucleotides were considered to be interacted with each other and they were connected by a spring. Besides that, in wGNM, the force constant specific to each spring was weighted by the total number of atom-atom contacts between the corresponding nucleotides.

(3) Based on the above model-construction method, the Kirchhoff matrix $\boldsymbol{\Gamma}$ for wGNM can be computed as

$$\Gamma_{ij} = \begin{cases} -n_{ij}, & \text{if } i \neq j, \\ -\sum\limits_{k,k\neq i}^{N} \Gamma_{ik}, & \text{if } i = j, \end{cases} \quad (6)$$

where n_{ij} is the total number of heavy atom contacts between the i^{th} and j^{th} nucleotides. Then, the normal modes of the system can be calculated through diagonalizing the matrix $\boldsymbol{\Gamma}$, and the nucleotide B-factors of the RNA structure can be computed by using Eqs. (3) and (4). In this study, the computed B-factors were compared with the experimental data by using the PCC calculated with Eq. (5) to evaluate the accuracy of wGNM.

2.3. Force-constant-decayed Gaussian network model

As noted in our previous discussion, a distance cutoff was used to determine whether an interaction exists between two nucleotides, and the force constant for all the inter-nucleotide interactions was assumed to be identical in the conventional GNM. However, the nucleotides in RNAs are nega-

tively charged, and the conventional cutoff-based GNM construction method may lead to the neglect of the long-range electrostatic interactions that are weak but important for RNA structural dynamics. In the present study, another new GNM construction method was proposed, in which the distance cutoff was removed and all the pairwise nucleotides were connected by springs with the force constant decayed exponentially with their separate distance.

In our proposed force-constant-decayed GNM (fcdGNM), the Kirchhoff matrix $\boldsymbol{\Gamma}$ can be calculated as

$$\Gamma_{ij} = \begin{cases} -\exp\left(-\dfrac{R_{ij}^0}{\sigma}\right), & \text{if } i \neq j, \\ -\sum\limits_{k,k\neq i}^{N} \Gamma_{ik}, & \text{if } i = j, \end{cases} \tag{7}$$

where R_{ij}^0 is the equilibrium distance between the i^{th} and j^{th} nucleotides; σ is a variable parameter that controls the decay rate of the force constant with increasing distance. The diagonalization of $\boldsymbol{\Gamma}$ gets the motion modes and the associated frequencies of the studied system. Then, the nucleotide *B*-factors were calculated with Eqs. (3) and (4), and the computed results were compared with the x-ray crystallographic experimental data to assess the performance of our FcdGNM approach in *B*-factors prediction. The PPC between the computed and experimental *B*-factors was calculated by using Eq. (5).

2.4. Dataset of the non-redundant RNA tertiary structures

In this study, the proposed new models, i.e., wGNM and fcdGNM, were tested by using a non-redundant RNA structure dataset, and the calculation results were compared with those computed by the conventional GNM to evaluate the performance of these models. The non-redundant RNA tertiary structures were obtained from the representative sets of RNA 3D structures collected by Leontis and Zirbel.[39] The accuracy of the experimental *B*-factors largely depends on the resolution of RNA structures determined by x-ray crystallography. Higher resolution generally implies more precise of *B*-factor values. Therefore, the RNA crystal structures with 3.0 Å resolution or better were chosen in our study. In addition, the GNM is more applicable to the well-packed micromolecular structures. However, the RNA structures with small size are usually loosely-packed and also generally complexed with proteins or other RNAs. The binding partners may have significant influences on the dynamics of the small-size RNA. Thus, in the present work, only the large-size RNA structures were considered. Based on the above considerations, the following filter criteria were used to get the RNA structure dataset for our studies from the representative set 3.128 of RNA 3D structures released on May 27 in 2020: (1) structural determination by x-ray crystallographic experiments; (2) structural resolution better than 3.0 Å; (3) the number of nucleotides in RNA structure larger than 122; (4) the nucleotide *B*-factor data having been measured. According to the above criteria, a total of 51 RNA structures were obtained as the database for the present work. The PDB codes of these RNA structures are listed in Table S1 in the supplementary materials.

3. Results and discussion

3.1. The optimization of the variable parameters in the models

In the conventional GNM, the cutoff distance R_c is the only adjustable parameter. In order to investigate the effect of the cutoff value on the calculation results, the PCC between the predicted and experimental *B*-factors was computed for different discrete values of R_c. Figure 1(a) shows the change of the average PCC for all the RNA structures in the non-redundant dataset with the value of R_c increasing from 9 Å to 23 Å with an interval of 1 Å. The PCC values with different cutoff distances for each individual RNA molecule in the dataset are given in Table S2 in the supplementary materials. It is found that the cutoff value has obvious influence on the computation results, where the average PCC between the predicted and experimental *B*-factors is in the range of 0.463–0.518. The maximum average PCC value is 0.518, which corresponds to the cutoff distance of 12 Å. In the present study, the cutoff of 12 Å was used in our calculation for the conventional GNM.

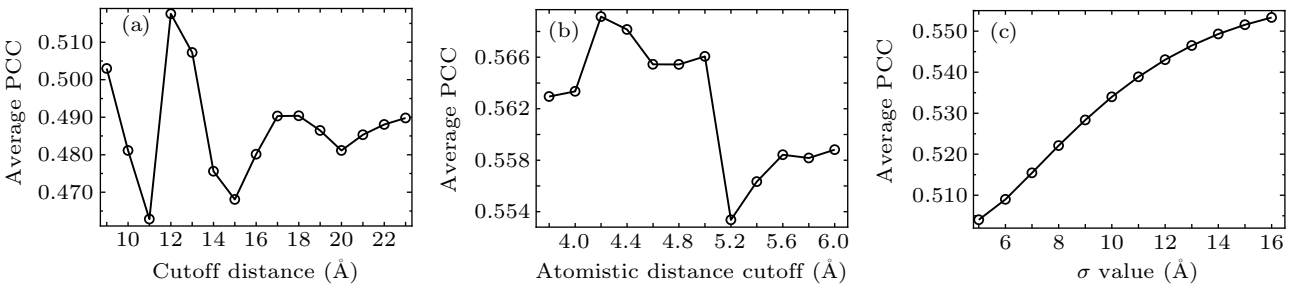


Fig. 1. The change of the average PCC between the predicted and experimental *B*-factors for the RNA structures in the non-redundant dataset calculated with different discrete values of the variable parameters in the models. (a) The change of the average PCC computed by the conventional GNM with different values of the cutoff distance R_c. (b) The change of the average PCC calculated by the wGNM with different values of the atomistic distance cutoff. (c) The change of the average PCC with the variation of the σ value computed by the fcdGNM.

For the proposed wGNM, the only variable parameter is the atomistic distance cutoff, which is used to determine the formation of contacts between heavy atoms. To explore the impact of the atomistic cutoff value on the performance of wGNM, the PCC between the predicted and experimental *B*-factors was calculated with twelve discrete cutoff values of 3.8 Å, 4.0 Å, 4.2 Å, 4.4 Å, 4.6 Å, 4.8 Å, 5.0 Å, 5.2 Å, 5.4 Å, 5.6 Å, 5.8 Å, 6.0 Å, respectively. The calculation results are shown in Fig. 1(b) and Table S3 in the supplementary materials. From Fig. 1(b), it is found that the average PCC of the RNAs in the database was varied from 0.553 to 0.569 when the above different cutoff values were adopted. For the cutoff value of 4.2 Å, the average PCC reached the highest value of 0.569. Thus, in this study, 4.2 Å was adopted as the atomistic distance cutoff value for the wGNM.

In the proposed fcdGNM, σ is the only adjustable parameter as discussed in Eq. (7). The influence of the σ value on the calculation results was studied, in which the value of σ increased from 5 Å to 16 Å with a step of 1 Å and the PCCs between the predicted and experimental *B*-factors were calculated for each σ value. Figure 1(c) displays the change of the average PCC with the variation of the σ value. The detailed PCC values for each RNA molecule in the non-redundant database can be found in Table S4 in the supplementary materials. From Fig. 1(c), the maximum value of the average PCC is 0.553 with the parameter σ to be 16 Å. Therefore, the parameter σ was set to be 16 Å for the fcdGNM.

3.2. The improved performance in reproduction of the experimental *B*-factors by our proposed models

For each RNA in the non-redundant database, the conventional GNM as well as the proposed wGNM and fcdGNM were respectively constructed based on the RNA crystal structure. The nucleotide *B*-factors were calculated by these three models, respectively, where the optimal value of the adjustable parameters in the models was used in the calculations as discussed in the previous section. Then, the PCCs between the predicted and the experimental *B*-factors were computed for wGNM, fcdGNM, and the conventional GNM, respectively, to evaluate their performance in reproducing the experimental-observed RNA structural dynamics. The PCC values obtained by the three models for all the 51 RNA structures in the non-redundant database are given in Table 1. The results show that the average PCC value is increased from 0.518 to 0.569 for the wGNM compared with the conventional GNM. The proposed new model, i.e., wGNM, obviously raises the *B*-factor prediction accuracy by 9.85%. For the other new model fcdGNM, the average value of PCCs is also improved from 0.518 to 0.553, with the increase rate of 6.76%. These results indicate that both our proposed models exhibit better performance in prediction nucleotide fluctuations than the conventional GNM. Our non-redundant dataset contains different types of RNA including ribosome, ribozyme, riboswitch, and other types. The performance of our proposed methods for different types of RNA was also evaluated separately. For ribosomal RNAs, the average PCC is improved from 0.579 to 0.657 and 0.645 for wGNM and fcdGNM, respectively. For ribozyme RNAs, the average PCC is increased from 0.507 to 0.547 and 0.524 for wGNM and fcdGNM, respectively. For riboswitch RNAs, the value is improved from 0.382 to 0.562 and 0.503 for wGNM and fcdGNM, respectively. These results indicate that our proposed methods outperform the conventional GNM for all these different RNA types.

From Table 1, it is found that the nucleotide *B*-factors of 8 RNA systems cannot be calculated by the conventional GNM due to more than one mode with zero frequency. The reason is that these RNA structures are not well packed and the number of springs connected to some nucleotides in these systems is too little. Several studies have also pointed out that the GNM is more applicable for tightly packed structures rather than the loosely packed structures.[40] Our fcdGNM method can effectively overcome this defect in the conventional GNM, where all the pairwise nucleotides were connected by springs in the fcdGNM and the force constant decayed exponentially with the pairwise separate distance. The results in Table 1 exhibit that the *B*-factors for these 8 RNA molecules can be effectively predicted by using our fcdGNM method. It should also be noted that the PPC values predicted by the conventional GNM for two RNA systems, i.e., chian B of 5T2A and chain E of 5T5H, are negative. The crystal structure of these two systems was checked manually, and it is found that the calculated RNA chain is packed tightly with other surrounding chains. In our calculation, only the studied RNA chain was considered separately, and the impacts of the surrounding chains were not taken into account, which may be responsible for the negative value of PCC. For the remaining 41 RNA structures, the PCC values predicted by the conventional GNM are positive and the average PCC is 0.557.

The 41 RNA structures with positive PCC values in the conventional GNM were also calculated with the new proposed wGNM and fcdGNM, as listed in Table 1. The average PCC is 0.631and 0.612 for wGNM and fcdGNM, respectively. Compared with the conventional GNM, the average PCC improved respectively by 13.3% and 9.87% for wGNM and fcdGNM, which indicates that the proposed models can obviously enhance the performance of GNM in describing RNA structural dynamics. From Table 1, it is also found that most of the RNAs with negative PCC values in the conventional GNM also exhibit negative PCC values in the wGNM and fcdGNM. This result implies that our proposed models have little effect on the RNA systems that are not suitable for the application of GNM.

Table 1. The PCC values for all the 51 RNA structures in the non-redundant database obtained by the conventional GNM, wGNM, and fcdGNM, respectively.

PDB code	Chain	Category	GNM	wGNM	fcdGNM
6QZP	L5+L8	ribosomal RNA	0.597	0.538	0.758
6SGC	54+84	ribosomal RNA	0.642	0.715	0.808
5TBW	1+4	ribosomal RNA	0.596	0.610	0.698
6S0Z	A	ribosomal RNA	0.702	0.916	0.721
5J7L	DA	ribosomal RNA	0.814	0.881	0.785
6SPB	A	ribosomal RNA	0.727	0.754	0.768
6V3A	AN1	ribosomal RNA	0.741	0.761	0.731
4Y4O	2A	ribosomal RNA	0.766	/[a)]	0.815
4V9F	0	ribosomal RNA	0.713	0.760	0.745
5DM6	X	ribosomal RNA	0.740	0.836	0.743
6ERI	AA	ribosomal RNA	0.588	0.841	0.744
4V88	A6	ribosomal RNA	0.716	0.740	0.747
6AZ1	1	ribosomal RNA	/	/	0.748
6QZP	S2	ribosomal RNA	0.747	0.836	0.773
6SGC	A1	ribosomal RNA	0.727	0.874	0.791
6AZ3	1+7	ribosomal RNA	0.571	0.667	0.630
6S0X	a	ribosomal RNA	0.262	0.413	0.414
5J7L	AA	ribosomal RNA	0.668	0.880	0.768
6SPC	a	ribosomal RNA	0.587	0.714	0.567
6V3A	sN1	ribosomal RNA	0.483	0.536	0.595
4Y4O	1a	ribosomal RNA	0.562	0.710	0.684
6TMF	A	ribosomal RNA	0.542	0.670	0.538
6ERI	BA	ribosomal RNA	0.603	0.777	0.770
5T5H	A+C	ribosomal RNA	/	/	0.461
5T2A	B	ribosomal RNA	–0.569	–0.502	–0.184
6AZ3	2	ribosomal RNA	/	0.675	0.757
5T5H	B	ribosomal RNA	/	/	0.381
4FAR	A	ribozyme RNA	0.624	0.677	0.633
4P95	A	ribozyme RNA	0.432	0.598	0.453
3JCS	3	ribosomal RNA	0.170	0.094	0.245
6AZ3	4	ribosomal RNA	0.788	0.797	0.764
6AZ3	3	ribosomal RNA	0.572	0.586	0.576
3DIL	A	riboswitch RNA	0.326	0.454	0.496
3D0U	A	riboswitch RNA	0.245	0.323	0.198
3PDR	X	riboswitch RNA	0.284	0.478	0.359
6UFG	A	riboswitch RNA	0.461	0.505	0.582
2R8S	R	ribozyme RNA	0.599	0.791	0.747
1U9S	A	ribozyme RNA	0.545	0.425	0.439
4ZDO	E	tRNA	/	–0.140	–0.081
5T5H	E	srRNA	–0.026	–0.247	0.016
3G9C	Q	ribozyme RNA	0.330	0.320	0.364
6ICZ	H+F	snRNA	/	/	0.454
3NDB	M	scRNA	0.659	0.651	0.487
3HHN	E	ribozyme RNA	0.293	0.311	0.399
6ID1	H+F	snRNA	/	/	0.337
6KC8	B	sgRNA	0.244	0.008	0.377
3V7E	C	riboswitch RNA	0.639	0.751	0.717
6N5P	A	riboswitch RNA	0.336	0.862	0.663
6QW6	4	snRNA	/	/	0.244
1L9A	B	scRNA	0.489	0.493	0.363
2Z75	B	ribozyme RNA	0.724	0.707	0.634
	average		0.518	0.569	0.553

[a)]The symbol "/" in the table means that the theoretical *B*-factor cannot be calculated because there is more than one normal mode with zero frequency.

It should be noted that the most time and memory consuming operation in the calculation of GNM is the diagonalization of the Kirchhoff matrix. For the proposed wGNM, the atomistic information was implicitly taken into account, and the size of the Kirchhoff matrix in wGNM is the same as that of the conventional GNM. For the fcdGNM, each nucleotide in the RNA structure is simplified as a node and the dimension of the Kirchhoff matrix is also same as that of the conventional GNM. Therefore, compared with the conventional GNM, these two new models have improved performance but do not increase the calculation burden.

3.3. Case study

In order to illustrate the improved performance of wGNM and fcdGNM relative to the conventional GNM in extracting the nucleotide fluctuations in RNA structures, the "AA" chain of the chloroplast ribosome (PDB code:6ERI) was investigated as a case study.[41] The nucleotide *B*-factors of the studied RNA structure were theoretically predicted by using the conventional GNM as well as the proposed wGNM and fcdGNM, respectively, and the predicted results were compared with the experimental data. The "AA" chain of the chloroplast ribosome is a large-size RNA structure composed of 2696 nucleotides, which forms a well-packed tertiary structure, as shown in Fig. 2(a). The PCC between the predicted and experimental *B*-factors calculated by the conventional GNM is 0.588. For the wGNM and fcdGNM, the PCC value is obviously improved to 0.841 and 0.744, respectively. These two new models are superior to the conventional GNM.

The exact calculation results of the *B*-factors by the wGNM and fcdGNM were compared with those by the conventional GNM in Figs. 2(b) and 2(c), respectively. It is found that the flexibility of a subdomain composed of 1079–1137 nucleotides, highlighted in red in Fig. 2(a), predicted by the conventional GNM is much lower than the experimental results. This may be caused by the large value of the distance cutoff adopting in the model. The large cutoff value may introduce non-existing interactions between this subdomain with the remain part of the RNA structure, which reduces the fluctuation of the subdomain. While for most of other flexible nucleotides, which correspond to the peaks in the *B*-factor curves, the fluctuations predicted by the conventional GNM are obviously higher than the experimental observations, as shown in Figs. 2(b) and 2(c). These flexible nucleotides are usually loops and protrude out from the RNA structure. This result indicates that the distance cutoff in the conventional GNM is insufficient to restrict the dynamics of these flexible nucleotides. However, for wGNM, the predicted results are significantly improved. The flexibility of the 1079–1137 subdomain calculated by wGNM agrees well with the experimental data, which is higher than the flexibility predicted by the conventional GNM, as shown in Fig. 2(b). In addition, for other flexible nucleotides corresponding to *B*-factor profile peaks, the fluctuations predicted by wGNM are also more con-

sistent with the experimental results, which are usually lower than the fluctuations predicted by the conventional GNM, as displayed in Fig. 2(b). These results imply that the wGNM can better capture the interaction distributions in RNA structures than the conventional GNM. For fcdGNM, the predicted results are also better than the conventional GNM. The fluctuations of the flexible nucleotides in protruding loops predicted by fcdGNM are lower than those calculated by the conventional GNM, and they better agree with the experimental results, as shown in Fig. 2(c). But for the 1079–1137 subdomain, the fluctuations predicted by fcdGNM are similar with those predicted by the conventional GNM.

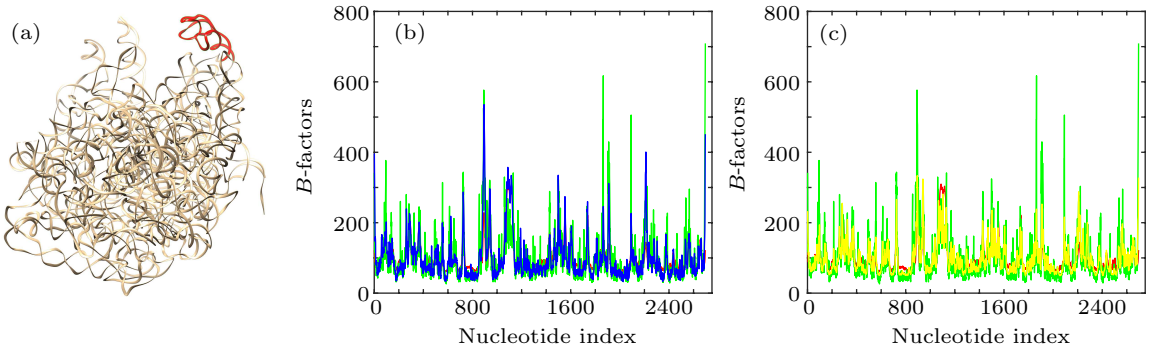


Fig. 2. The calculation results for the "AA" chain of the chloroplast ribosome (PDB code:6ERI). (a) The tertiary structure of the "AA" chain of the chloroplast ribosome, where the subdomain composed of 1079–1137 nucleotides is highlighted in red. (b) The nucleotide *B*-factors predicted by the wGNM (blue color) compared with the results obtained by the conventional GNM (green color) as well as the experimental data (red color). (c) The nucleotide *B*-factors predicted by the fcdGNM (yellow color) compared with the results obtained by the conventional GNM (green color) as well as the experimental data (red color).

4. Conclusion

More and more evidences have revealed that RNA molecules can fold into subtle three-dimensional structures as proteins and carry out various important functions in many biological processes. The exertion of the biological functions of RNAs depends on their structure-encoded conformational dynamics. How to effectively predict the dynamical properties from the tertiary structure of RNA molecules is important for the understanding of the molecular mechanism of RNA functions. The Gaussian network model (GNM), as one type of the elastic network model (ENM), is an efficient method to explore the intrinsic dynamics of biomacromolecules, which has been widely applied for proteins. In recent years, GNM has also been extended to study the structural dynamics of RNA. However, owing to the more loosely packed tertiary structure, the performance of the conventional GNM on RNAs is not as good as that on proteins. It is a valuable problem worth studying to improve the performance of GNM applied to RNAs.

In the conventional GNM, a cutoff distance was used to construct the model and a uniform force constant was applied for all the nucleotide pairs. Considering the structural characteristics and the various forces distribution properties specific to RNAs, the above simplifications in the conventional GNM may not work well for RNAs. In the present work, aiming at these simplifications, two approaches, i.e., wGNM and fcdGNM, were proposed to improve the application of GNM in investigating the structural dynamics of RNAs. In wGNM, the force constant is specific to the corresponding pairwise nucleotides, which is weighted by the number of atom–atom contacts between them. In fcdGNM, the force constant decays exponentially with the separate distance between the pairwise nucleotides. These two new models were tested by using a non-redundant RNA structure dataset composed of 51 RNAs, and the calculation results show that the proposed two models can better reproducing the experimental *B*-factors than the conventional GNM. Compared with the conventional GNM, the average PCC between the predicted and experimental *B*-factors improved by 9.85% and 6.76% for wGNM and fcdGNM, respectively. It is also found that GNM is more applicable for the RNA structures with large size, and 41 out of 51 RNA structures exhibit positive PCC values. For these 41 RNAs, the average PCC computed by wGNM and fcdGNM is improved respectively by 13.3% and 9.87% compared with the conventional GNM. The improved performance of wGNM and fcdGNM was illustrated by using the chain "AA" of the chloroplast ribosome as a case study.

The improved models outperform the conventional GNM in prediction nucleotide fluctuations. Besides that, the nucleotide-specific information was introduced implicitly into the model, which does not increase the calculation burden. Our studies provide two promising candidate models for better revealing the dynamical properties encoded in RNA structures.

References

[1] Breaker R R and Joyce G F 2014 *Chem. Biol.* **21** 1059
[2] Storz G 2002 *Science* **296** 1260
[3] Mustoe A M, Brooks C L and Al-Hashimi H M 2014 *Annu. Rev. Biochem.* **83** 441
[4] Dethoff E A, Chugh J, Mustoe A M and Al-Hashimi H M 2012 *Nature* **482** 322
[5] Manzano I, Vezeau G, Salis H and Zydney A L 2020 *Sep. Purif. Technol.* **237** 116372

[6] Hamdani H Y, Yahaya Z H and Firdaus-Raih M 2018 *Encyclopedia of Bioinformatics and Computational Biology*, 1st edn. (Amsterdam: Elsevier) pp. 546–553
[7] Huang B, Du Y, Zhang S, Li W, Wang J and Zhang J 2020 *Chin. Phys. B* **29** 108704
[8] Shi Y Z, Wu Y Y, Wang F H and Tan Z J 2014 *Chin. Phys. B* **23** 078701
[9] Zhang J, Zhang Y J and Wang W 2010 *Chin. Phys. Lett.* **27** 118702
[10] Kmiecik S, Kouza M, Badaczewska-Dawid A E, Kloczkowski A and Kolinski A 2018 *Int. J. Mol.Sci.* **19** 3496
[11] Haliloglu T, Bahar I and Erman B 1997 *Phys. Rev. Lett.* **79** 3090
[12] Bahar I, Atilgan A R, Demirel M C and Erman B 1998 *Phys. Rev. Lett.* **80** 2733
[13] Su J G, Jiao X, Sun T G, Li C H, Chen W Z and Wang C X 2007 *Biophys. J.* **92** 1326
[14] Zheng W and Wen H 2017 *Curr. Opin. Struct. Biol.* **42** 24
[15] Chennubhotla C, Rader A J, Yang L W and Bahar I 2005 *Phys. Biol.* **2** 173
[16] Rader A J and Bahar I 2004 *Polymer* **45** 659
[17] Su J G, Du H J, Hao R, Xu X J, Li C H, Chen W Z and Wang C X 2013 *J. Phys. Chem.* **117** 8689
[18] Zhang P F and Su J G 2019 *J. Chem. Phys.* **151** 045101
[19] Zhang Z, Shi Y and Liu H 2003 *Biophys. J.* **84** 3583
[20] Zacharias M 2010 *Curr. Opin. Struct. Biol.* **20** 180
[21] Liu Z C, Liu Q, Chen C Y, Zeng C, Ran P, Zhao Y J and Pan L 2020 *Chin. Phys. B* **29** 108709
[22] Yang L W, Rader A J, Liu X, Jursa C J, Chen S C, Karimi H A and Bahar I 2006 *Nucleic Acids Res.* **34** 24
[23] Pinamonti G, Bottaro S, Micheletti, C Bussi G and Giovanni B 2015 *Nucleic Acids Res.* **43** 7260
[24] Zimmermann M T and Jernigan R L 2014 *RNA* **20** 792
[25] Wang Y M, Rader A, Bahar I and Jernigan R L 2004 *J. Struct. Biol.* **147** 302
[26] Afonin K A, Kasprzak W, Bindewald E, Puppala P S, Diehl A R, Hall K T and Tae J K, Zimmermann M T, Jernigan R L, Jaeger L and Shapiro B A 2014 *Methods* **67** 256
[27] Sen T Z, Feng Y P, Garcia J V, Kloczkowski A and Jernigan R L 2006 *J. Chem. Theory Comput.* **2** 696
[28] Van Wynsberghe A W and Cui Q 2005 *Biophys. J.* **89** 2939
[29] Zimmermann M T and Jernigan R L 2014 *RNA* **20** 792
[30] Atilgan AR, Durell S R, Jernigan R L, Demirel M C, Keskin O and Bahar I 2001 *Biophys. J.* **80** 505
[31] Diggins P, Liu C J, Deserno M and Potestio R 2019 *J. Chem. Theory Comput.* **15** 648
[32] Mailhot O, Frappier V, Major F and Najmanovich R 2017 bioRxiv
[33] Wang S H, Gong W K, Deng X Q, Liu Y and Li C H 2020 *Chem. Phys.* **538** 110820
[34] Bahar I, Atilgan A R and Erman B 1997 *Fold Des.* **2** 173
[35] Chawla M, Chermak E, Zhang Q, Bujnicki J M, Oliva R and Cavallo L 2017 *Nucleic Acids Res.* **45** 11019
[36] Bottaro S, Palma F D and Bussi G 2014 *Nucleic Acids Res.* **42** 13306
[37] Butcher S E and Pyle A M 2011 *Acc. Chem. Res.* **44** 1302
[38] Sweeney B A, Roy P and Leontis N B 2015 *WIREs RNA* **6** 17
[39] Leontis N B and Zirbel C L 2012 *RNA 3D Structure Analysis and Prediction* (Berlin Heidelberg: Springer) p. 281
[40] Riccardi D, Cui Q and Phillips G N 2009 *Biophys. J.* **96** 464
[41] Boerema A P, Aibara S, Paul B, Tobiasson V, Kimanius D, Forsberg B O, Wallden K, Lindahl E and Amunts A 2018 *Nat. Plants* **4** 212