



Exploring the dynamics of RNA molecules with multiscale Gaussian network model

Shihao Wang, Weikang Gong, Xueqing Deng, Yang Liu, Chunhua Li*

College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China

ARTICLE INFO

Keywords:
 Multiscale Gaussian network model (mGNM)
 B-factor
 RNA molecules
 Divide domain
 Intramolecular interactions

ABSTRACT

RNA molecules play important roles in biological processes, their functions are intimately related to structural dynamics. Elastic network model (ENM) has achieved great success in predicting the large-amplitude collective behavior of proteins. However, for loosely-packed RNA structures, ENM models can not reproduce their dynamics as accurate as the densely-packed ones. In this work, the multiscale Gaussian network model (mGNM) is extended to predict dynamic properties of RNAs. All tests are performed on a non-redundant RNA structure database we constructed. In results, for B-factor reproduction, encouragingly mGNM achieves a significant improvement with the average value of Pearson correlation coefficient (PCC) between theoretical and experimental B-factors being 0.732, much higher than 0.494 and 0.321 obtained by conventional GNM and parameter-free GNM (pfGNM) models, respectively. Furtherly, mGNM attains a larger improvement in B-factor prediction for loosely-packed parts. Additionally, based on the analysis of functional movements, mGNM can properly make domain decompositions for tRNA^{Asp} and xrRNA. This work can strengthen the understanding of the intrinsic dynamics of RNAs, and mGNM is expected to have a bright prospect in dynamic analyses for loosely folded biomolecules, especially RNAs.

1. Introduction

Although composed of only four chemically similar nucleotides, RNAs play important roles in gene expression and regulation [1]. Specifically, RNA is capable of catalytic activity [2] and even classic RNAs such as ribosomal, transfer and messenger RNAs play surprisingly complex roles in protein synthesis [3]. The complex functions of RNAs are intimately related to their abilities to modulate structural dynamics [4]. Characterizing the functional dynamics of RNA molecules becomes increasingly important in RNA biology.

Molecular dynamics (MD) simulations can be used to study the dynamic behaviors of RNAs. However, the atomic-level simulation is often found too expensive and time consuming to study the large-scale conformational motion of RNAs. Meanwhile, the limited accuracy of the physics-based force field has been a bottleneck to describe the subtle atomic interactions within RNA molecules surrounded by solvent and ions because any atomic interaction error can dramatically impact the distribution of the MD trajectories [5,6].

The elastic network model (ENM) has been shown to be a particularly effective computational technique to investigate the function-relevant motions of proteins and even RNAs [7,8]. Tirion first represented the intramolecular interactions as elastic potentials of springs and

reproduced the molecular low frequency motional modes [9]. Yang et al. presented a web server, oGNM, to calculate the normal modes of motions for proteins and oligonucleotides and their complexes [10]. For nucleic acids, researchers developed the ENM models with different number of nodes representing a nucleotide to study their functional cooperative motions [2,8,11–14]. Our group proposed ENM-based methods to study the issues involved in the folding and allosteric processes of biomolecules [15–17]. In the conventional ENM, a biomolecule structure is modeled as an elastic network of certain atoms in which the node pairs within a given cutoff distance are considered to have interactions and are connected by a set of Hookean springs with a uniform force constant [18]. Generally, the low-frequency motion modes obtained by ENM represent the large-scale collective motions relevant to molecular functions. In order to improve the method, many modifications have been proposed in recent years, mainly aiming at the two points of the conventional ENM: the fixed cutoff distance and uniform force constant [19]. As we know, in molecules different interactions, such as covalent, Van der Waals, hydrophobic and electrostatic interactions [20], have different action ranges and do not suddenly decay to zero when the distances are beyond certain values. Thus, Riccardi et al. changed the identical spring constant in conventional ENM into a distance dependent function [21]. Later, Yang et al. further

* Corresponding author.

E-mail address: chunhuali@bjut.edu.cn (C. Li).

developed a parameter-free Gaussian network model (pfGNM) [22] where there is no cutoff distance and all residue pairs are considered to have interactions with the strength inversely proportional to the square of the inter-residue distance. Recently, considering the different action ranges of different interactions, Wei and coworkers proposed a multiscale Gaussian network model (mGNM) [23] where inter-residue interactions are considered as the sum of different interactions with different spring's strengthens represented by kernel functions. The mGNM has been successfully applied on dynamic analysis of proteins with the prediction of B-factor significantly improved by more than 13% compared with the conventional GNM [23].

Although ENM model has been successfully used to investigate the large-scale functional motions of many proteins including GroEL, GroEL-GroES complex and calcium ATPase [24–27], it meets difficulties in dynamic analysis for the loosely folded systems (not well-packed structures), especially for RNA molecules [28,29]. Sen et al. also found that with the conventional GNM, the flexible prediction of protein internal residues (packed compactly) is generally more accurate than that of surface residues (relatively packed loosely) [30]. The partial reason may be related to the identical connection strength in the conventional ENM, which has also been pointed out by Robert et al. [30]. Additionally, for RNA molecules which are highly negatively charged, the long-range electrostatic interactions within them are much stronger than those within proteins, which makes the conventional ENM with a cutoff distance not suitable for RNA dynamic analysis. Based on the analyses above, we think that the mGNM model with multiscale interactions considered should perform well on the loosely folded RNA structures.

In this work, we extend the mGNM method to the analyses of RNA dynamics, and compare the results obtained by mGNM with those by conventional GNM and pfGNM models. Additionally, we analyze the motional coupling with mGNM on RNA molecules and found this model can properly make domain decompositions.

2. Methods and materials

2.1. Database of RNA structures

We downloaded all the 1,374 structures containing only RNA molecules (October 2018) from Protein Data Bank (PDB) [31] (PDB, <http://www.rcsb.org/>). Considering the study aim, these structures were further filtered, and those meeting the following criteria were retained: [1] X-ray crystal structure, [2] structure resolution better than 3.0 Å, [3] number of nucleotides between 60 and 200, [4] sequence identity less than 90%. Here, we used CD-HIT [32–34] to cluster the RNAs. The sequences with more than 90% identity were put into one cluster, and the structure owning the most nucleotides in each cluster was selected as a representative. Finally, the database consists of 77 RNA structures (Table S1 in Supplementary material).

2.2. Gaussian network model

2.2.1. Conventional Gaussian network model

In the conventional Gaussian network model (GNM), a RNA structure is represented as a coarse-grained and elastic network where several nodes (here, P atom selected) are used to replace one nucleotide and the node pairs that are less than a cutoff distance (R_c) apart are connected by springs with a uniform force constant [35]. Thus, the total internal potential energy of the network of N nucleotides can be written as:

$$V = \frac{1}{2} \gamma \sum_{i,j} (\Delta\mathbf{R}_i - \Delta\mathbf{R}_j)^2 = \frac{1}{2} \gamma (\Delta\mathbf{R}^T \Gamma \Delta\mathbf{R}) \quad (1)$$

where $\Delta\mathbf{R}_i$ and $\Delta\mathbf{R}_j$ are the displacement for the i th and j th nodes respectively, the column vector $\Delta\mathbf{R}$ represents the fluctuation of the N

nucleotides, and γ is the force constant of the springs. The Kirchhoff matrix Γ is an $N \times N$ symmetric matrix where the elements can be described as:

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \text{ and } R_{ij} \leq R_c \\ 0, & \text{if } i \neq j \text{ and } R_{ij} > R_c \\ -\sum_{j,j \neq i}^N \Gamma_{ij}, & \text{if } i = j \end{cases} \quad (2)$$

where R_c is the cutoff distance and R_{ij} is the distance between the i th and j th nodes.

2.2.2. Parameter-free Gaussian network model

Different from the conventional cutoff-based GNM, the parameter-free Gaussian network model (pfGNM) adopts a distance-dependent spring constant set [22]. Here in the method, also P atoms are selected as nodes. All the node pairs are considered to have interactions and connected by the springs with a force constant being inversely proportional to the square of the distance between the two nodes. By this simplification, the total potential energy has the same formula as Eq. (1), but the Kirchhoff matrix Γ has the different form that can be described as:

$$\Gamma_{ij} = \begin{cases} -R_{ij}^{-2} & \text{if } i \neq j \\ -\sum_{j,j \neq i}^N \Gamma_{ij} & \text{if } i = j \end{cases} \quad (3)$$

where R_{ij} is the distance between the i th and j th nodes. The node pairs that are far apart have weaker interactions than those pairs that are close to each other.

2.2.3. Multiscale Gaussian network model

Based on the conventional GNM, the multiscale Gaussian network model (mGNM) [23] is improved in two aspects: [1] not one but multiscale interactions are considered in a superposition form of springs with different strengths; [2] not cutoff but different kernel function based spring force constants are adopted to represent interactions. Here, still P atoms are selected as nodes, and different exponential decay kernel functions [36] are used to represent interactions which can be written as:

$$\Phi(R_{ij}; \eta, \kappa) = e^{-(R_{ij}/\eta)^\kappa}, \quad \kappa > 0 \quad (4)$$

where R_{ij} is the distance between the i th and j th nodes, and the parameters η and κ control the decay extent of the kernel function. Intramolecular interactions can be represented as the sum of different kernel functions, and the corresponding Kirchhoff matrix of the n th kernel function can be described as:

$$[\Gamma_n]_{ij} = \begin{cases} -\Phi(R_{ij}; \eta_n, \kappa_n), & \text{if } i \neq j \\ -\sum_{j,j \neq i}^N [\Gamma_n]_{ij}, & \text{if } i = j \end{cases} \quad (5)$$

When n (in this work $n = 3$ in order to balance the accuracy and computational time) kernel functions are considered, the corresponding multiscale Kirchhoff matrix can be given by:

$$\Gamma = \sum_n a_n \cdot \Gamma_n + b \cdot \mathbf{M} \quad (6)$$

where Γ_n is the corresponding Kirchhoff matrix of the n th kernel function which can be obtained from Eq. (5), \mathbf{M} is a $N \times N$ matrix with all elements being 1, and the parameters a_n and b can be obtained by the least square method:

$$\text{Min}_{a_n, b} \left\{ \sum_i \left| \sum_n a_n [\Gamma_n]_{ii} + b - \frac{1}{B_i^{\exp}} \right|^2 \right\} \quad (7)$$

where B_i^{\exp} is the experimental B-factor of the i th node. The above formula is from the method of mFRI (multiscale flexibility-rigidity index) proposed by Wei and coworkers [23,36,37]. Theoretically, for

the Kirchhoff matrix (diagonally dominant matrix), the diagonal elements (connectivity strengths) reflect the rigidities of nucleotides, and are also taken as rigid indexes. Here, the kernel function weights are determined by minimizing the difference between the diagonal elements of Kirchhoff matrix and the inverses of experimental B-factors, which is consistent with mFRI theory proposed by Wei and coworkers [23].

2.3. Fluctuation and the cross-correlation

Gaussian network model method simplifies the complicated potential into a quadratic function in the vicinity of the equilibrium state, which allows for decomposing the motions into normal modes with different frequencies. These normal modes can be obtained by decomposing the inverse matrix of Kirchhoff matrix Γ^{-1} :

$$\Gamma^{-1} = \sum_{k=2}^N \lambda_k^{-1} \mathbf{u}_k \cdot \mathbf{u}_k^T \quad (8)$$

where λ_k and \mathbf{u}_k are the eigenvalue (motional frequency) and eigenvector of the k th motional mode, respectively. The mean square fluctuation of the i th nucleotide and the cross-correlation between the i th and j th nucleotides can be calculated by the following two equations as the sum of contributions from individual normal modes.

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle = \left(\frac{3k_B T}{\gamma} \right) \cdot \Gamma_{ii}^{-1} = \left(\frac{3k_B T}{\gamma} \right) \cdot \left[\sum_{k=2}^N \lambda_k^{-1} \mathbf{u}_k \cdot \mathbf{u}_k^T \right]_{ii} \quad (9)$$

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \left(\frac{3k_B T}{\gamma} \right) \cdot \Gamma_{ij}^{-1} = \left(\frac{3k_B T}{\gamma} \right) \cdot \left[\sum_{k=2}^N \lambda_k^{-1} \mathbf{u}_k \cdot \mathbf{u}_k^T \right]_{ij}, \quad (10)$$

where k_B is the Boltzmann constant, T is the absolute temperature, and the meaning of γ is the same as that in Eq. (1). B-factors, also called the crystallographic temperature factors, contain important information about individual atoms' thermal motions in a macromolecule, reflecting the local structural flexibility of the ground-state conformation. Theoretically, according to the Debye-Waller theory, the B-factor of the i th nucleotide can be calculated according to the Eq. (11), which is proportional to its mean square fluctuation (MSF).

$$B_i = \frac{8\pi^2}{3} \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle \quad (11)$$

The normalized cross-correlation between the two nucleotides can be written as:

$$C_{ij} = \frac{\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle}{\sqrt{\langle (\Delta \mathbf{R}_i)^2 \rangle} \cdot \sqrt{\langle (\Delta \mathbf{R}_j)^2 \rangle}} \quad (12)$$

This value varies from -1 to 1 . Positive values mean correlated motions happening along the same direction and the negative indicate correlated motions along the opposite direction. A higher absolute value represents a stronger correlation between the two nucleotides. The value $C_{ij} = 0$ means that the motions of nucleotides are completely uncorrelated.

For conventional GNM, pfGNM and mGNM models, the B-factors and cross-correlations are all calculated according to Eq. (11) and (12) respectively based their own Kirchhoff matrixes obtained from Eqs. (2), (3) and (6) respectively.

2.4. Pearson correlation coefficient

The Pearson correlation coefficient (PCC) defined as follows is used to optimize the parameters in conventional GNM and mGNM models.

$$PCC = \frac{\sum (B_i^{the} - \bar{B}^{the}) \cdot (B_i^{exp} - \bar{B}^{exp})}{\sqrt{\sum (B_i^{the} - \bar{B}^{the})^2} \cdot \sqrt{\sum (B_i^{exp} - \bar{B}^{exp})^2}} \quad (13)$$

where B_i^{the} and B_i^{exp} represent theoretical and experimental B-factors of the i th nucleotide, respectively, and \bar{B}^{the} and \bar{B}^{exp} the corresponding expected values. The PCC value ranges from -1 to 1 , and a larger PCC value means the theoretical result is closer to the experimental data.

It should be pointed out that for conventional GNM and mGNM models with parameters (except for pfGNM since there is no parameters), the optimized parameters are obtained to construct the corresponding GNM models through maximizing the PCC value between the theoretical and experimental B-factors for each RNA. During this process, for conventional GNM, the parameter cutoff distance value systematically varies in the range of $1\text{--}25 \text{ \AA}$ with a step size of 0.1 \AA , and for mGNM, the six parameters $\{\eta_1, \kappa_1, \eta_2, \kappa_2, \eta_3, \kappa_3\}$ vary in the ranges of $1\text{--}15, 1\text{--}10, 1\text{--}20, 1\text{--}10, 1\text{--}30$, and $1\text{--}10$, respectively with the same step size of 1.0 . The predefined ranges are based on the work by Wei and coworkers [23] where the parameters vary in the range of $\{\eta: 1\text{--}25, \kappa: 1\}$ for proteins. Here it is noted that in mGNM optimization process, for a given set of six parameters, there are still other parameters a_n and b in Eq. (6) to be determined, for which we use the least square method as shown in Eq. (7). For clarity, this process is shown in Fig. S2.

3. Results and discussion

3.1. Reproduction of B-factor on RNA dataset

The accurate prediction of B-factors provides an effective starting point to understand the dynamics of biomolecules. For the 77 RNAs, we constructed their mGNM models, calculated the theoretical B-factors and obtained PCC values under the optimal parameter set. In order to evaluate the performance of mGNM on RNA molecules, we compared the B-factor reproduction by mGNM with the corresponding results by conventional GNM and pfGNM models. The results are shown in Fig. 1 and the detailed data are given in Table S1.

From Fig. 1, the PCC values obtained by mGNM are significantly better than those by the other two methods for almost all the RNAs. Compared with the conventional GNM (see Fig. 1 (A)), mGNM achieves the improvements for 76 out of 77 cases with only one case (GlmS ribozyme, PDB ID: 2GCS) having a drop of 0.009 in PCC. Compared with pfGNM (see From Fig. 1 (B)), generally mGNM obtains relatively larger improvements. The average value corresponding to mGNM is 0.732 , much higher than 0.494 and 0.321 obtained by conventional GNM and pfGNM models respectively. The improvements reach to 48% and 128% respectively. In previous work, Xia et al. performed mGNM on 364 proteins and found the improvement is 13% compared with the conventional GNM [23]. Here it should be pointed out that in our work the optimized parameters for each case were used (the optimization process of parameters is described in detail at the end of Methods and Materials.) and they used a fixed set of parameters regardless of mGNM and conventional GNM models. Although this, the results still have comparability to some extent. The comparison suggests that more improvements can be achieved by mGNM method to be performed on RNAs than on proteins. In the following part, we will give the possible reasons (see section of B-factor reproduction for regions with different packing density). Furtherly from Fig. 1, evidently, the improvements are more remarkable for the cases where the conventional GNM and pfGNM models perform poorly, which is very important for a modified method. Additionally, comparing conventional GNM with pfGNM, it can be seen that the former performs slightly better than the latter on RNAs. On proteins, the similar result was also found by Mendonca et al. [38], while the contrary one was found by Zhang et al. [39], which is probably related to the different proteins and different fixed cutoff distances used in the two works.

For pfGNM, Yang et al. tried different inverse powers of the inter-residue distance to mimic spring constants, and found that the inverse 2nd power clearly outperforms others in B-factor predictions for

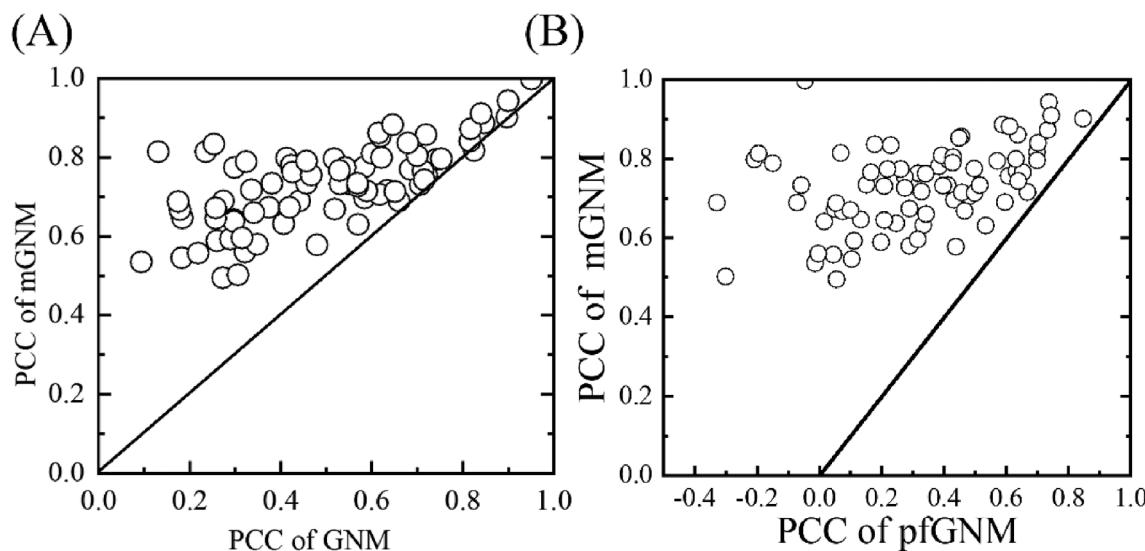


Fig. 1. Comparison of Pearson correlation coefficient between experimental and theoretical B-factors obtained by mGNM with those obtained by conventional GNM (A) and pfGNM (B) on the 77 RNAs.

proteins [22]. We also constructed the pfGNM models with different inverse powers 2–10 for the RNAs, and compared the average PCC values obtained from them, as shown in Table S2. The results show that the average PCC value corresponding to the inverse 2nd power is the highest, consistent with the previous result on proteins, which is why we adopted the inverse 2nd power here.

3.2. Case study

Table 1 lists the PCC values obtained by the three methods for the four cases where the conventional GNM performs the worst (2GDI: 0.095, 1D4R: 0.132, 4XW7: 0.175, 1Y27: 0.236) while mGNM achieves the most improvements more than 0.5 for three cases (1D4R: 0.814, 4XW7: 0.688, 1Y27: 0.813) and a medium one for one case (2GDI: 0.535). In the following, the detailed results and analyses will be given for the three cases 1D4R, 4XW7 and 1Y27. It should be pointed out that for the above three methods we mentioned, the corresponding force constants were taken into account to obtain the absolute B-factor values.

1D4R is a fragment of helix 6 of human signal recognition particle (SRP) RNA [40] which presents an irregular shape and makes the RNA structure (loosely folded) quite different from proteins (see Fig. S1 (A)). It can be deduced that the conventional GNM (cutoff 7.9 Å; the force constant 0.861 within the range recommended by the literature [41]) with a fixed cutoff distance and a uniform force constant can not give a good flexible prediction (PCC: 0.132), as shown in Fig. 2 (A) and (B) (the reproduced B-factors and RNA surface with them mapped onto). The pfGNM model also fails with PCC value being 0.069. When the multiscale interactions considered, the mGNM elevates the value to 0.814. The parameters of the three kernel functions are Φ_1 : $\eta_1 = 1$, $\kappa_1 = 2$; Φ_2 : $\eta_2 = 7$, $\kappa_2 = 4$; Φ_3 : $\eta_3 = 7$, $\kappa_3 = 5$, respectively. From the Fig. 2 (A), the fluctuations obtained by conventional GNM and pfGNM

models are far different from the experimental values (the corresponding force constant has been multiplied), while mGNM can accurately predict the highest peak of fluctuation although the other predicted highest peaks are not as high as the experimental ones. Additionally, for most of low valley regions in fluctuation, mGNM can relatively accurately give their locations in structure, while their predicted B-factors are more or less higher than the corresponding experimental ones.

4XW7 is a crystal structure of ZMP riboswitch [42] which includes two substructure domains (P1-P2 and P3 stems) long-range coupling with each other, and a junction J13 and pseudoknots (PK regions) that plays an important role in stabilizing the structure (see Fig. S1 (B)). Still, the conventional GNM (cutoff 16.1 Å, the force constant 0.854 within the range recommended by the literature [41]) and pfGNM models fail in predicting its flexibility (PCC values: 0.175 and 0.054 respectively), while mGNM can reproduce a relatively good result (see Fig. 3 (A) and (B)) with PCC being 0.688. The parameters of the three kernel functions are Φ_1 : $\eta_1 = 11$, $\kappa_1 = 7$; Φ_2 : $\eta_2 = 16$, $\kappa_2 = 5$; Φ_3 : $\eta_3 = 10$, $\kappa_3 = 7$, respectively. From the two figures, mGNM can well predict the two most flexible regions L2 and J13. Studies have shown that the hydrogen bonds in J13 are easy to be destroyed, indicating its relatively large flexibility [43]. For the three low valleys (two PK regions and partial P1 stem, some of which consist the ligand binding pocket), mGNM gives a relatively even prediction. Around these regions, we find some Mg^{2+} ions near PK regions, and iridium hexamine near the P1 stem in the crystal structure. Their existence can stabilize the local structures to some extent. Maybe this is the main reason why mGNM (with ions and iridium hexamine not considered) does not give a relatively rigid prediction for these regions.

1Y27 is a G-riboswitch-guanine complex [44] where the G-riboswitch forms a structure similar to a tuning fork with the ligand guanine binding to its pocket (see Fig. S1 (C)). In the network construction with mGNM, the parameters of the three kernel functions are Φ_1 : $\eta_1 = 4$, $\kappa_1 = 2$; Φ_2 : $\eta_2 = 4$, $\kappa_2 = 4$; Φ_3 : $\eta_3 = 7$, $\kappa_3 = 9$, respectively. The PCC value reaches to 0.813, much higher than the corresponding values 0.236 and −0.195 obtained by the conventional GNM (cutoff 12 Å, the force constant 0.259 within the range recommended by the literature [41]) and pfGNM models, respectively (see Fig. 4 (A)). From Fig. 4 (A), the evident fluctuation peaks (except for the last one corresponding to 3' end of P1 stem) can be well reproduced by mGNM while there are almost no changes for the fluctuations produced by the other two models. From Fig. 4 (B), the predicted highest peak is corresponding to 5' end of P1 stem which will unfold first when the ligand

Table 1

PCC values obtained by the three GNMs for the four cases where the conventional GNM performs the worst while mGNM achieves the most improvements for three cases and a medium one for one case.

PDB code	GNM	pfGNM	mGNM
1D4R	0.132	0.069	0.814
2GDI	0.095	−0.013	0.535
1Y27	0.236	−0.195	0.813
4XW7	0.175	0.054	0.688

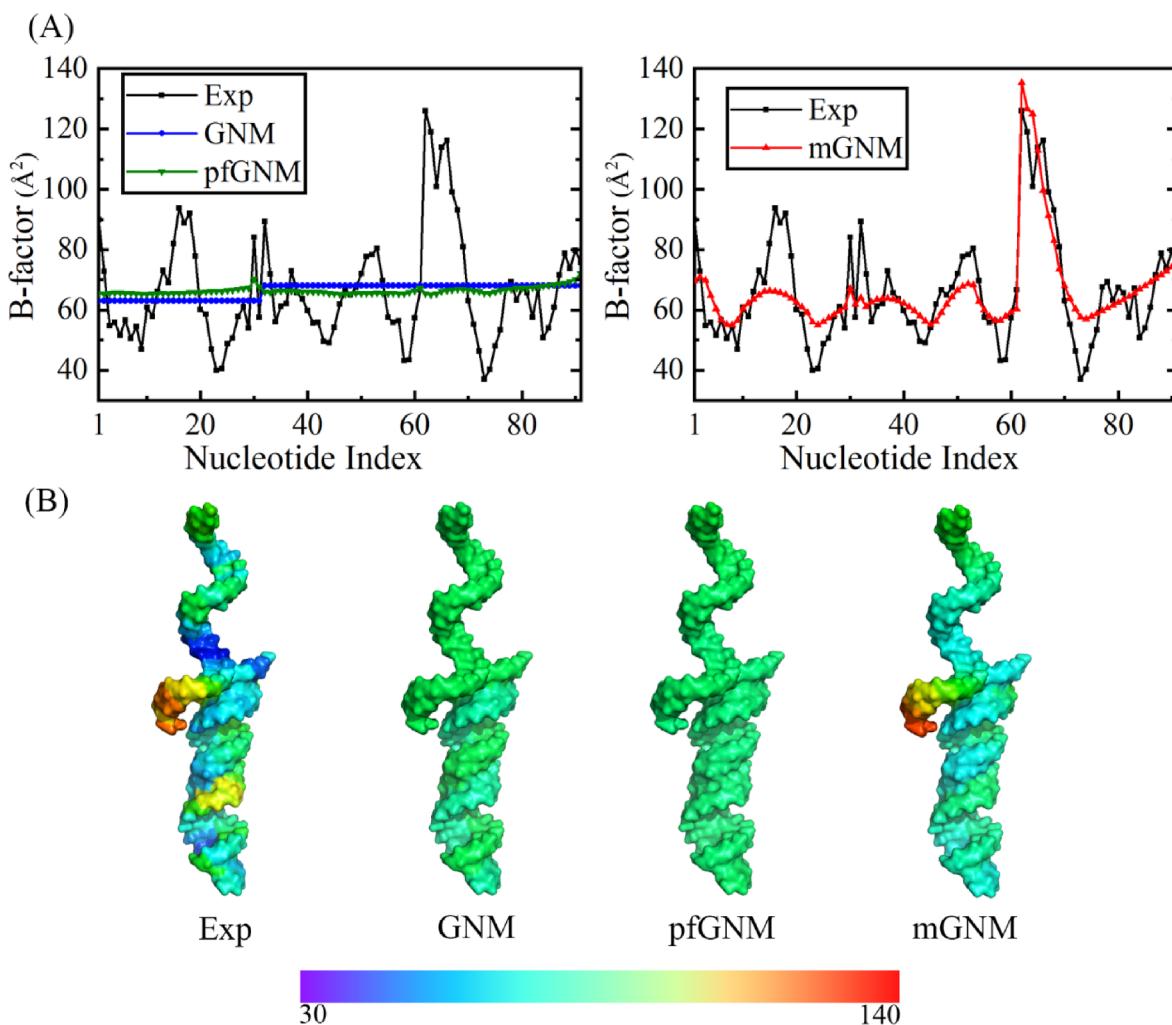


Fig. 2. (A) Theoretical B-factors of 1D4R obtained by mGNM, conventional GNM and pfGNM with the experimental values shown for comparison (the absolute B-factor values were obtained for the three methods). (B) RNA surfaces colored by experimental B-factors (Exp) and theoretical ones obtained with the three GNMs.

disassociates from the RNA [44]. The other four predicted highest peaks correspond to P2 and P3 stems which will unfold after P1 stem [16,44]. The junction region including J12, J23 and J31, constitutes the ligand binding pocket [44] and presents low fluctuation in our prediction, consistent with the experimental trend. As for the main shortcoming of mGNM in this case, we can see that some valleys like L2 and J23 segments in the experimental fluctuation can be predicted as valleys but the predicted values are still a little higher than the corresponding experimental ones, which may also be due to the solvation and ion effects not considered in mGNM method.

Based on the analyses above, mGNM can better reproduce the flexibility of RNA molecules than conventional GNM and pfGNM models. The main reason is that the multiple kernel functions integrated in mGNM can take into account not only the multiscale interactions to some extent but also these interactions in a decay form instead of in a cutoff-based form, which is more important in constructing the elastic networks for RNAs (with low packing density and rich negative charges) than for proteins. Additionally, it can be seen that mGNM has a good capability to predict peaks of fluctuation, while for some low valleys (rigid regions) mGNM performs not very well. For the former, the likely reason is that the soft connections considered in kernel functions can confer high flexibility to some regions, and for the latter the partial reason lies in that the solvent and ion effects are not considered in mGNM. Taking into account these effects can enhance the connection strength and therefore give a relatively rigid prediction.

We noted that the cutoff distance values used in our optimized

conventional GNM models for the three cases are different from the ones 10–20 Å [12] usually adopted in previous studies for the conventional one-node GNM model. Thus, we performed the conventional GNM with different cutoff values for the three cases, and the PCC values are 0.0889, -0.1589 and -0.0674 under 16 Å for cases 1D4R, 4XW7 and 1Y27, respectively, which are all lower than those from our optimized models respectively, and the case is the same for other cutoff values. In fact, by reviewing literatures, we found that evidently different cutoff values were ever used in constructing conventional one-node GNM models for different kinds RNA molecules, such as 15 Å for Ribozyme [12], 20 Å for thiM thiamine pyrophosphate riboswitch [8], which we think is partially due to the complexity of RNA structures.

3.3. B-factor reproduction for regions with different packing density

It has been a difficulty for GNM methods to perform on loosely folded structures or regions [29], we want to know whether the mGNM can give any improvement in dealing the issue. To detect it, we analyzed the performance of mGNM on regions with different packing densities. The packing density of where a nucleotide exists is defined as the number of its neighboring nucleotides less than 9 Å apart from it. The smaller the number is, the more loosely the nucleotide folds. For nucleotides with different packing densities, we calculated their absolute deviations of the theoretical from experimental B-factors to reflect the performances of mGNM on them for flexibility prediction, as shown in Fig. 5. We also give the corresponding results obtained by

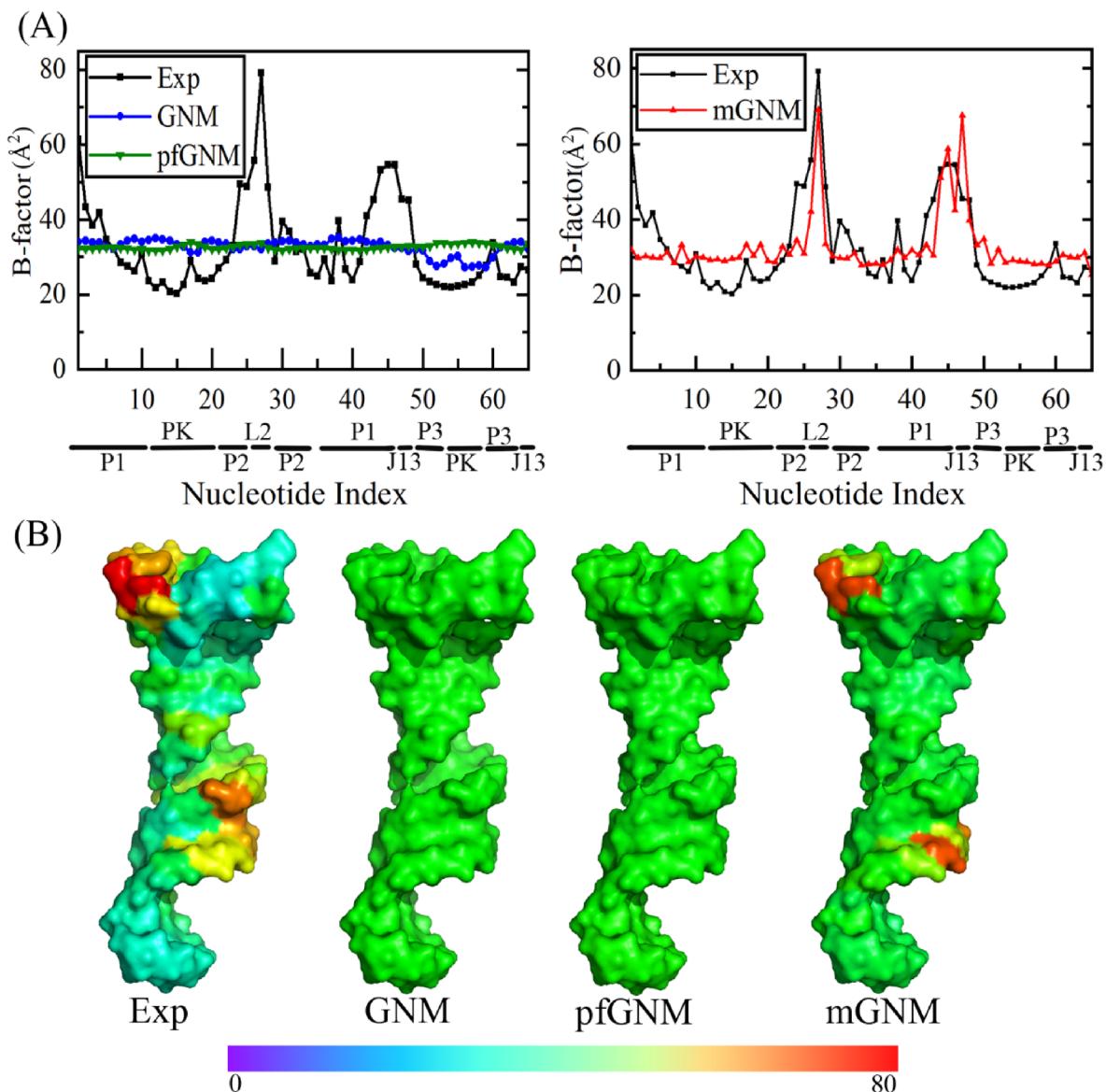


Fig. 3. (A) Theoretical B-factors of 4XW7 obtained by mGNM, conventional GNM and pfGNM with the experimental values shown for comparison (the absolute B-factor values were obtained for the three methods). (B) RNA surfaces colored by experimental B-factors (Exp) and theoretical ones obtained with the three GNMs.

conventional GNM for comparison (the results from pfGNM are not given due to its not good performance). From Fig. 5, it can be seen that with the increasing of packing density the absolute deviations obtained by both conventional GNM and mGNM have an evident decrease, which is consistent with previous studies that GNMs are more suitable to be performed on the densely folded structures [21]. Furthermore, we found that in terms of the average deviations, their values from mGNM are far less than those from conventional GNM regardless of how the packing density is, which verifies again the better potency of mGNM in reproducing molecular flexibility. Additionally, encouragingly, the improvements are much larger for the relatively more loosely folded regions, which can explain the reason why the better flexibility prediction can be obtained by mGNM on RNAs (loosely folded) than on proteins (densely folded). Thus, mGNM with multiscale interactions considered has a strong capacity for molecular flexibility prediction, especially for the loosely folded structures or regions.

3.4. Domain decomposition with mGNM

The mGNM has a good performance in RNA flexibility analysis.

Then we want to know whether the model can capture the functional coupling within RNAs. To detect it, we selected two biologically important RNAs, yeast tRNA^{Asp} (PDB ID: 2TRA) [45] and xrRNA (PDB ID: 5TPY) [46], as representatives to explore the issue. In order to obtain the motional coupling, we calculated the normalized cross-correlations (Eq. (12)). The cross-correlation between two nucleotides is between -1 and 1 , with positive values representing cooperative motions, and vice versa.

The crystal structure of tRNA^{Asp} is shown in Fig. 6 (A), which consists of 73 nucleotides folding in an L-shape conformation. In constructing the network by mGNM, the parameters of three kernel functions are Φ_1 : $\eta_1 = 9$, $\kappa_1 = 1$; Φ_2 : $\eta_2 = 20$, $\kappa_2 = 2$; Φ_3 : $\eta_3 = 25$, $\kappa_3 = 4$, respectively. The PCC between theoretical and experimental B-factors is 0.807. Fig. 6 (B) shows the cross-correlation map of tRNA^{Asp}. From Fig. 6 (B), the structure is roughly divided into five parts, of which parts 1 and 5 correspond to the acceptor arm, part 3 the anticodon region, and parts 2 and 4 the D-loop and T-loop in the Fig. 6 (A), respectively. In addition, evidently regions 1 and 5, 2 and 4 are positively correlated, indicating that they are coupled and cooperative in motion. Thus, the five parts can be grouped into three domains with parts 1 and 5 coupled

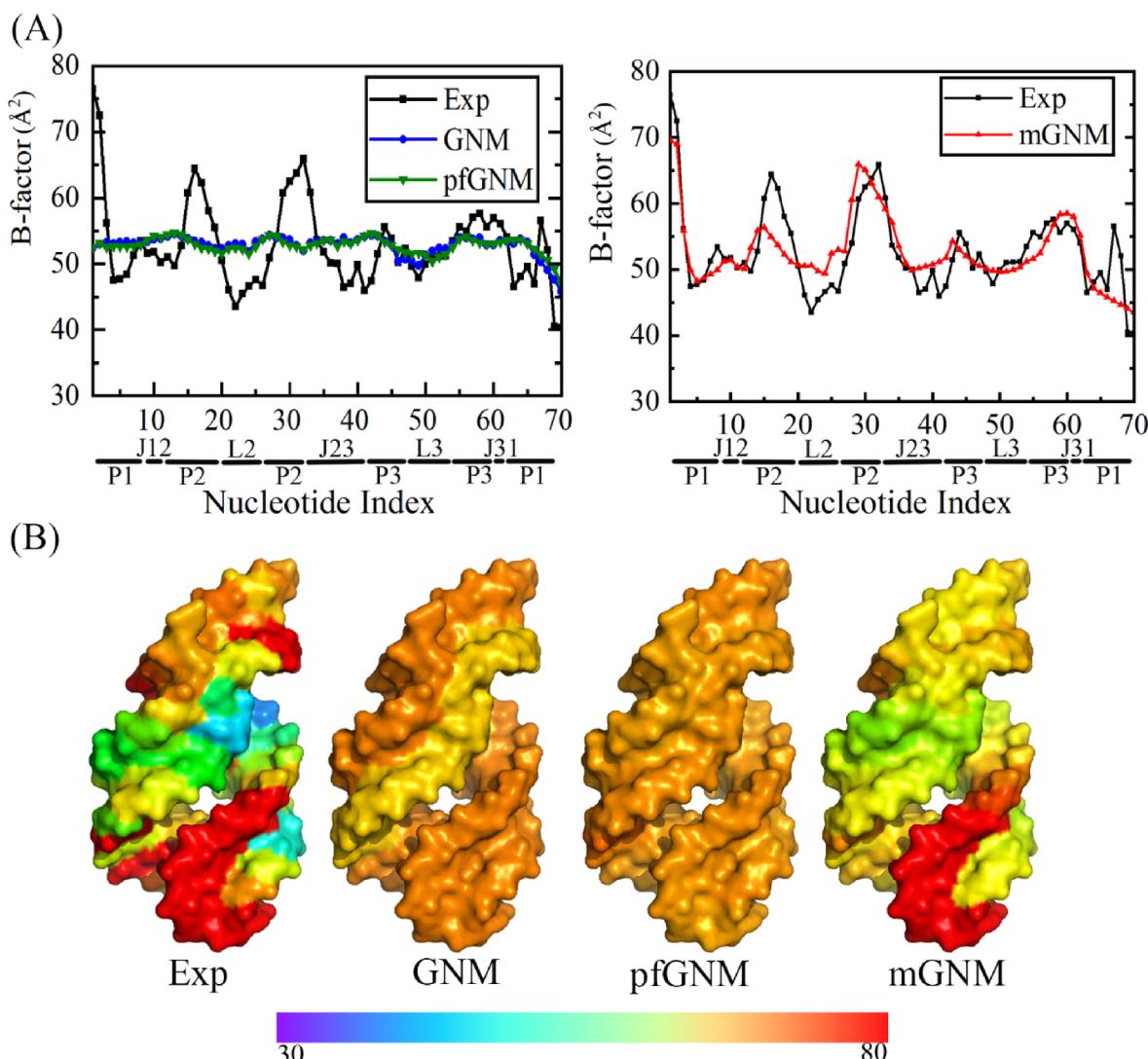


Fig. 4. (A) Theoretical B-factors of 1Y27 obtained by mGNM, conventional GNM and pfGNM with the experimental values shown for comparison (the absolute B-factor values were obtained for the three methods). (B) RNA surfaces colored by experimental B-factors (Exp) and theoretical ones obtained with the three GNMs.

to each other as an arm of the L-shaped conformation corresponding to acceptor stem, part 3 forming the other arm corresponding to anticodon loop and stem, and parts 2 and 4 constituting the elbow, which is consistent with the collective motions of tRNA^{Asp} low-frequency modes [47].

The xrRNA containing 71 nucleotides, used by the Zika virus to make sfRNA by co-opting a cellular exoribonuclease [46], is shown in Fig. 6 (C). In constructing the network by mGNM, the parameters of three kernel functions are Φ_1 : $\eta_1 = 15$, $\kappa_1 = 8$; Φ_2 : $\eta_2 = 15$, $\kappa_2 = 10$; Φ_3 : $\eta_3 = 5$, $\kappa_3 = 5$, respectively. The PCC between theoretical and experimental B-factors is 0.743. Fig. 6 (D) shows the cross-correlation map of xrRNA. From Fig. 6 (D), the structure can be roughly divided into four parts, of which part 1 and 2 correspond to the P1 and P2 stems, respectively, part 3 contains P3 and S4, and Part 4 consists of the P4-L4 in Fig. 6 (C). Evidently, parts 1 and 3 are positively correlated, indicating that they are coupled together in motion, which has been described in a previous study [46] that is part 1 (P1) and a part of part 3 (P3) form a ring and the 5' end passes through P3. For part 4 (P4-L4), previous study has pointed out that P4-L4 is highly conserved and the stacking stabilizes its structure, and additionally it can bind to exoribonuclease and change the enzyme's activity [46]. From above, the decomposed domains by mGNM based on the cross-correlations are in agreement with the experimental data to some extent.

3.5. Discussion on the difference in parameter sets of kernel functions for RNAs

It is noted that there exists a big difference in parameter sets of kernel functions for different RNAs, as shown in the section of Case study. We have analyzed the possible reasons as follows. Firstly, as we know, ion types and numbers embedded in RNA structures have a significant effect on RNA stabilities due to their strong negative electricity. In some experimental structures such as 1TOD and 3BNN there are no ions, while there are many ions of different types in others such as 2YIE (2 K⁺ and 14 Mg²⁺) and 4JF2 (15 Cs⁺ and 4 Mg²⁺). However, the current mGNM model does not explicitly take ions into account. Thus, the effects of the ignored ions will be considered into the optimized parameter sets in order to reach the best correlations between the theoretical and experimental B-factors for RNAs, which may induce a big difference in parameter sets for different RNAs. Secondly, the large difference in packing density may also be a factor. RNAs tend to fold into complex structures, some of which are densely packed, but others are loosely packed [48]. Different packing densities lead to different distribution patterns of nucleotides in space. The complexity of the interactions between nucleotides makes it difficult to use the same or similar functions to describe the interactions within RNAs of different packing densities. The last but not the least, due to the negative

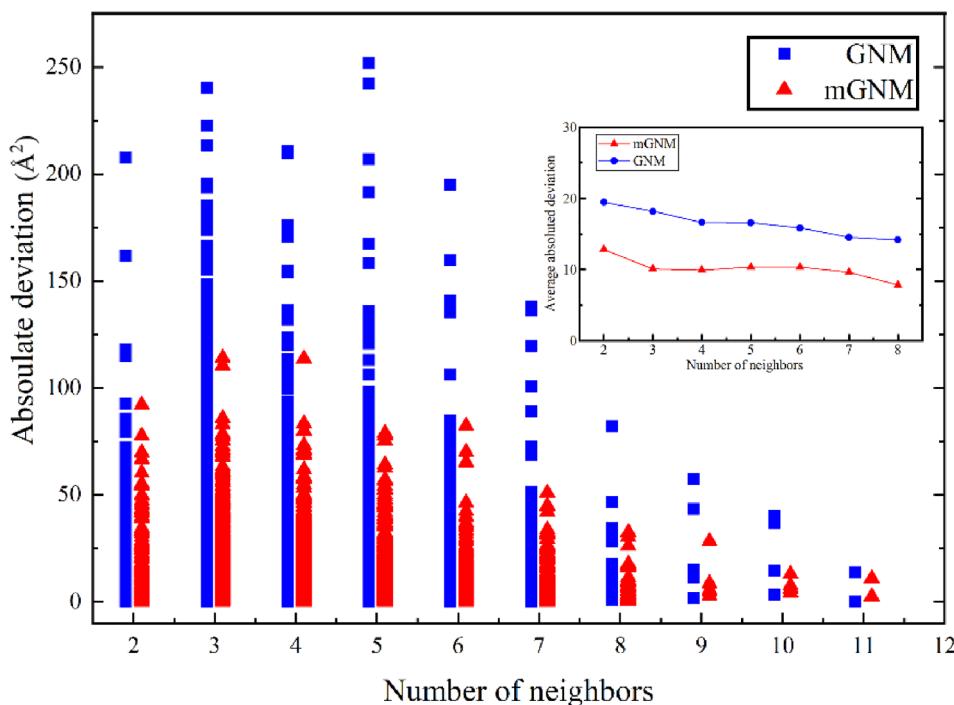


Fig. 5. Absolute deviations between experimental and theoretical B-factors obtained by conventional GNM and mGNM models for the nucleotides with different packing densities (denoted as number of neighbors).

electricity of RNAs, the different solution conditions (such as pH value, temperature, ion concentration and so on) for RNA folding may also bring about the big difference in parameter sets. In fact, in the traditional GNM model based on one node, the only parameter cutoff distance also has a larger variation range for RNAs (10–20 Å) [12] than for proteins (6–12 Å) [49].

Since the parameter set of kernel functions needs to be optimized, the mGNM can not be truly applied in the blind B-factor prediction for RNAs. It is worth noting that the recent work by Bramer et al. has achieved the blind B-factor prediction for proteins by integrating the machine learning and advanced graph theory, namely, multiscale weighted colored graphs (MWCGs) method [50]. In this work, the convolutional neural network (CNN) is trained with the fixed matrix composed of 30×8 kernel functions and other nine quantities as features. We will try the strategy in the future work and hope to achieve a blind B-factor prediction for RNAs. Additionally, it should be noted that the aim of our work is to build a Gaussian network model which is able to analyze the functional dynamic properties of RNAs such as motional correlations between different parts, and low-frequency collective motions encoded in RNA structures, which is different from the aim of the work by Bramer et al.

4. Limitations

It should be mentioned that there are some limitations in our current study. First, we only performed one-node (P atom) GNM models, and did not try two-node and three-node GNM models. Yang et al. pointed out that the conventional GNM with one node per nucleotide is better than the models with three nodes per nucleotide in reproducing B-factors for 64 oligonucleotides [10]. Studies show that the conventional GNM model with two nodes presenting a nucleotide can give a better result in reproducing B-factor for a DNA octahedron [51]. Here, the three kinds of GNM models all use one P atom to represent a nucleotide, and thus they are of comparability. Second, the domain decomposition based on the coupling analyses needs to be done on more RNA molecules in order to verify the effectiveness of mGNM. For the two points above, the further deep research works are now underway.

5. Conclusion

Accurately characterizing the dynamic properties of RNA molecules is important for us to understand their various biological functions. Currently, the elastic network model is an effective coarse-grained method for exploring the large-scale functional motions of bio-macromolecules. The conventional Gaussian network model (GNM) works well on proteins, but usually performs not ideally on the loosely folded RNAs. In this work, we extend the multiscale Gaussian network model (mGNM) to the analyses of RNA dynamics. We evaluate the performance of mGNM on a non-redundant database containing 77 RNAs constructed by us, and compare the results with those obtained by conventional GNM and pfGNM models. Under the optimal parameter set, mGNM achieves a significant improvement in reproducing B-factors with the average value of PCC between the theoretical and experimental B-factors being 0.732, much higher than 0.495 and 0.321 obtained by the conventional GNM and pfGNM models, respectively. The flexibility prediction for nucleotides with different packing densities indicates that GNMs are more suitable to be applied on the densely folded regions. And encouragingly, compared with conventional GNM, mGNM achieves a great improvement (regardless of how the packing density is), especially for the loosely folded regions, which can explain why much more improvement can be obtained by mGNM on RNAs than on proteins. Finally, the functional coupling is analyzed for tRNA^{Asp} and xrRNA systems and the results show that mGNM can properly make domain decompositions. In short, mGNM with the multiscale interactions considered is expected to have a bright prospect in RNA dynamic analysis.

6. Significance

Accurately obtaining the dynamical information is a pivotal step toward the understanding of RNA functions. Conventional ENM model, a particularly effective harmonic potential-based method for exploring the large-amplitude collective motions, performs not ideally for usually loosely folded RNAs. In this study, the mGNM with multiscale interactions considered is extended to the reproduction of dynamical

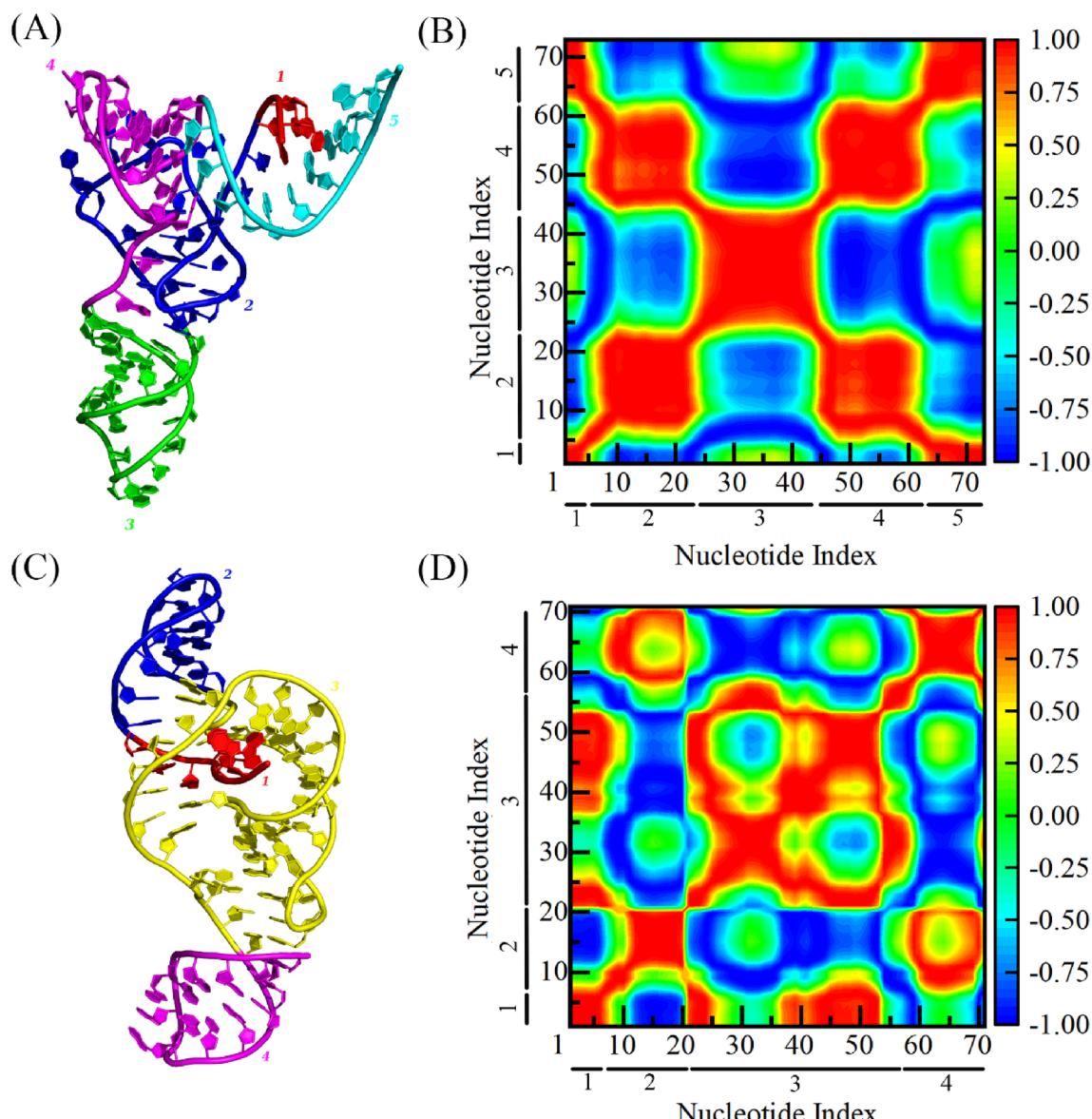


Fig. 6. (A) Crystal structure of tRNA^{Asp} (PDB ID: 2TRA). (B) Cross-correlations calculated by mGNM of tRNA^{Asp}. (C) Crystal structure of xrRNA (PDB ID: 5TPY). (D) Cross-correlations calculated by mGNM of xrRNA.

information for RNAs. The results show that mGNM achieves a significant improvement in flexibility prediction, especially for the loosely packed parts, and additionally, mGNM can properly make domain decompositions.

Author contributions

Shihao Wang, Chunhua Li and Weikang Gong designed the research. Shihao Wang performed mGNM calculations and Xueqing Deng and Yang Liu carried out the conventional GNM and pfGNM calculations. Shihao Wang and Chunhua Li performed data analyses and wrote the manuscript.

CRediT authorship contribution statement

Shihao Wang: Conceptualization, Methodology, Software, Writing - original draft. **Weikang Gong:** Conceptualization. **Xueqing Deng:** Methodology, Software. **Yang Liu:** Software. **Chunhua Li:** Conceptualization, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [31971180, 11474013], the Beijing Natural Science Foundation [4152011].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemphys.2020.110820>.

References

- [1] M.X. Cao, Y.L. Tang, W.L. Zhang, Y.J. Tang, X.H. Liang, Non-coding RNAs as

- Regulators of Lymphangiogenesis in Lymphatic Development, Inflammation, and Cancer Metastasis, *Front. Oncol.* 9 (2019) 916.
- [2] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme, *Cell* 35 (1983) 849–857.
- [3] R.M. Voorhees, V. Ramakrishnan, Structural basis of the translational elongation cycle, *Annu. Rev. Biochem.* 82 (2013) 203–236.
- [4] E.A. Dethoff, J. Chugh, A.M. Mustoe, H.M. Al-Hashimi, Functional complexity and regulation through RNA dynamics, *Nature* 482 (2012) 322–330.
- [5] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules, *Nat. Struct. Biol.* 9 (2002) 646–652.
- [6] R.O. Dror, R.M. Dirks, J.P. Grossman, H. Xu, D.E. Shaw, Biomolecular simulation: a computational microscope for molecular biology, *Annu. Rev. Biophys.* 41 (2012) 429–452.
- [7] S. Kmiecik, M. Kouza, A.E. Badaczewska-Dawid, A. Kloczkowski, A. Kolinski, Modeling of Protein Structural Flexibility and Large-Scale Dynamics: Coarse-Grained Simulations and Elastic Network Models, *Int. J. Mol. Sci.* 19 (2018).
- [8] G. Pinamonti, S. Bottaro, C. Micheletti, G. Bussi, Elastic network models for RNA: a comparative assessment with molecular dynamics and SHAPE experiments, *Nucleic Acids Res.* 43 (2015) 7260–7269.
- [9] M.M. Tirion, Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis, *Phys. Rev Lett* 77 (1996) 1905–1908.
- [10] L.W. Yang, A.J. Rader, X. Liu, C.J. Jursa, S.C. Chen, H.A. Karimi, I. Bahar, oGNM: online computation of structural dynamics using the Gaussian Network Model, *Nucleic Acids Res.* 34 (2006) W24–W31.
- [11] I. Bahar, R.L. Jernigan, Vibrational dynamics of transfer RNAs: comparison of the free and synthetase-bound forms, *J. Mol. Biol.* 281 (1998) 871–884.
- [12] C. Hyeon, D. Thirumalai, Ripping RNA by Force Using Gaussian Network Models, *J. Phys. Chem. B* 121 (2017) 3515–3522.
- [13] Y. Wang, A.J. Rader, I. Bahar, R.L. Jernigan, Global ribosome motions revealed with elastic network model, *J. Struct. Biol.* 147 (2004) 302–314.
- [14] M. Delarue, Y.H. Sanejouand, Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model, *J. Mol. Biol.* 320 (2002) 1011–1024.
- [15] J.G. Su, C.H. Li, R. Hao, W.Z. Chen, C.X. Wang, Protein unfolding behavior studied by elastic network model, *Biophys. J.* 94 (2008) 4586–4596.
- [16] C. Li, D. Lv, L. Zhang, F. Yang, C. Wang, J. Su, Y. Zhang, Approach to the unfolding and folding dynamics of adA-Riboswitch upon adenine dissociation using a coarse-grained elastic network model, *J. Chem. Phys.* 145 (2016) 14104.
- [17] Z. Han, Q. Shao, W. Gong, S. Wang, J. Su, C. Li, Y. Zhang, Interpreting the Dynamics of Binding Interactions of snRNA and U1A Using a Coarse-Grained Model, *Biophys. J.* 116 (2019) 1625–1636.
- [18] A.R. Attilan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, I. Bahar, Anisotropy of fluctuation dynamics of proteins with an elastic network model, *Biophys. J.* 80 (2001) 505–515.
- [19] C. Dieterich, P.F. Stadler, Computational biology of RNA interactions, Wiley Interdiscip Rev RNA 4 (2013) 107–120.
- [20] M. Hodges, M. Barahona, S.N. Yaliraki, Allostery and cooperativity in multimeric proteins: bond-to-bond propensities in ATCase, *Sci. Rep.* 8 (2018) 11079.
- [21] D. Riccardi, Q. Cui, G.J. Phillips, Application of elastic network models to proteins in the crystalline state, *Biophys. J.* 96 (2009) 464–475.
- [22] L. Yang, G. Song, R.L. Jernigan, Protein elastic network models and the ranges of cooperativity, *Proc. Natl. Acad. Sci. U.S.A* 106 (2009) 12347–12352.
- [23] K. Xia, K. Opron, G. Wei, Multiscale Gaussian network model (mGNM) and multi-scale anisotropic network model (mANM), *J. Chem. Phys.* 143 (2015).
- [24] C. Chennubhotla, I. Bahar, Markov methods for hierarchical coarse-graining of large protein dynamics, *J. Comput. Biol.* 14 (2007) 765–776.
- [25] H. Na, R.L. Jernigan, G. Song, Bridging between NMA and Elastic Network Models: Preserving All-Atom Accuracy in Coarse-Grained Models, *PLoS Comput. Biol.* 11 (2015) e1004542.
- [26] W. Zheng, H. Wen, A survey of coarse-grained methods for modeling protein conformational transitions, *Curr. Opin. Struct. Biol.* 42 (2017) 24–30.
- [27] W. Zheng, B.R. Brooks, D. Thirumalai, Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations, *Biophys. J.* 93 (2007) 2289–2299.
- [28] M.T. Zimmermann, R.L. Jernigan, Elastic network models capture the motions apparent within ensembles of RNA structures, *RNA* 20 (2014) 792–804.
- [29] A.W. Van Wynsberghe, Q. Cui, Comparison of mode analyses at different resolutions applied to nucleic acid systems, *Biophys. J.* 89 (2005) 2939–2949.
- [30] T.Z. Sen, Y. Feng, J.V. Garcia, A. Kloczkowski, R.L. Jernigan, The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models, *J. Chem. Theory Comput.* 2 (2006) 696–704.
- [31] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, *Nucleic Acids Res.* 28 (2000) 235–242.
- [32] W. Li, L. Jaroszewski, A. Godzik, Clustering of highly homologous sequences to reduce the size of large protein databases, *Bioinformatics* 17 (2001) 282–283.
- [33] W. Li, L. Jaroszewski, A. Godzik, Tolerating some redundancy significantly speeds up clustering of large protein databases, *Bioinformatics* 18 (2002) 77–82.
- [34] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, *Bioinformatics* 22 (2006) 1658–1659.
- [35] I. Bahar, A.R. Attilan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, *Fold Des.* 2 (1997) 173–181.
- [36] K. Xia, K. Opron, G.W. Wei, Multiscale multiphysics and multidomain models—flexibility and rigidity, *J. Chem. Phys.* 139 (2013) 194109.
- [37] K. Opron, K. Xia, G.W. Wei, Communication: Capturing protein multiscale thermal fluctuations, *J. Chem. Phys.* 142 (2015) 211101.
- [38] M.R. de Mendonca, L.G. Rizzi, V. Contessoto, V.B. Leite, N.A. Alves, Inferring a weighted elastic network from partial unfolding with coarse-grained simulations, *Proteins* 82 (2014) 119–129.
- [39] H. Zhang, T. Jiang, G. Shan, S. Xu, Y. Song, Gaussian network model can be enhanced by combining solvent accessibility in proteins, *Sci. Rep.* 7 (2017).
- [40] K. Wild, O. Weichenrieder, G.A. Leonard, S. Cusack, The 2 A structure of helix 6 of the human signal recognition particle RNA, *Structure* 7 (1999) 1345–1352.
- [41] D. Ming, M.E. Wall, Allostery in a coarse-grained model of protein dynamics, *Phys. Rev. Lett.* 95 (2005) 198103.
- [42] J.J. Trausch, J.G. Marcano-Velazquez, M.M. Matyjasik, R.T. Batey, Metal Ion-Mediated Nucleobase Recognition by the ZTP Riboswitch, *Chem. Biol.* 22 (2015) 829–837.
- [43] P.B. Kim, J.W. Nelson, R.R. Breaker, An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism, *Mol. Cell* 57 (2015) 317–328.
- [44] A. Serganov, Y.R. Yuan, O. Pikovskaya, A. Polonskaya, L. Malinina, A.T. Phan, C. Hobartner, R. Micura, R.R. Breaker, D.J. Patel, Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs, *Chem. Biol.* 11 (2004) 1729–1741.
- [45] E. Westhof, P. Dumas, D. Moras, Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals, *Acta Crystallogr. A* 44 (Pt 2) (1988) 112–123.
- [46] B.M. Akiyama, H.M. Laurence, A.R. Massey, D.A. Costantino, X. Xie, Y. Yang, P. Shi, J.C. Nix, J.D. Beckham, J.S. Kieft, Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease, *Science* 354 (2016) 1148–1152.
- [47] S. Nakamura, J. Doi, Dynamics of transfer RNAs analyzed by normal mode calculation, *Nucleic Acids Res.* 22 (1994) 514–521.
- [48] W. Huggins, S.K. Ghosh, P. Wollenzien, Hydrogen bonding and packing density are factors most strongly connected to limiting sites of high flexibility in the 16S rRNA in the 30S ribosome, *BMC Struct. Biol.* 9 (2009) 49.
- [49] D.A. Kondrashov, Q. Cui, G.J. Phillips, Optimization and evaluation of a coarse-grained model of protein motion using x-ray crystal data, *Biophys. J.* 91 (2006) 2760–2767.
- [50] D. Brämer, G.W. Wei, Blind prediction of protein B-factor and flexibility, *J. Chem. Phys.* 149 (2018) 134107.
- [51] G. Hu, L. He, F. Iacovelli, M. Falconi, Intrinsic dynamics analysis of a DNA octahedron by elastic network model, *Molecules* 22 (2017).