

Ripping RNA by Force using Gaussian Network Models

Changbong Hyeon

*School of Computational Sciences, Korea Institute for
Advanced Study, Seoul 130-722, Republic of Korea*

D. Thirumalai

Department of Chemistry, University of Texas at Austin, TX 78712, USA

(Dated: September 20, 2016)

Abstract

Using force as a probe to map the folding landscapes of RNA molecules has become a reality thanks to major advances in single molecule pulling experiments. Although the unfolding pathways under tension are complicated to predict studies in the context of proteins have shown that topology plays is the major determinant of the unfolding landscapes. By building on this finding we study the responses of RNA molecules to force by adapting Gaussian network model (GNM) that represents RNAs using a bead-spring network with isotropic interactions. Cross-correlation matrices of residue fluctuations, which are analytically calculated using GNM even upon application of mechanical force, show distinct allosteric communication as RNAs rupture. The model is used to calculate the force-extension curves at full thermodynamic equilibrium, and the corresponding unfolding pathways of four RNA molecules subject to a quasi-statically increased force. Our study finds that the analysis using GNM captures qualitatively the unfolding pathway of *T. ribozyme* elucidated by the optical tweezers measurement. However, the simple model is not sufficient to capture subtle features, such as bifurcation in the unfolding pathways or the ion effects, in the forced-unfolding of RNAs.

I. INTRODUCTION

While rigid body dynamics under mechanical stress is well studied in classical mechanics and engineering [1], similar considerations in the context of biopolymers began in earnest only recently with the advent of single molecule (SM) techniques. These methods are now routinely used to manipulate the biopolymers [2–8]. The renewed interest in the semiflexible chain [9–11] is partly due to the stretching experiments on dsDNA and multidomain proteins whose force-extension curves are well characterized by the worm-like chain model [12]. SM manipulation techniques using the atomic force microscopy (AFM) and laser optical tweezers (LOT) have yielded great insights into the elasticity of nucleic acids as well as quantitative glimpses into the folding landscapes of proteins and RNA and their complexes. In particular, the use of mechanical force provides a viable means to measure and decipher the underlying characteristics of the free energy landscape, such as the position of transition state, ruggedness, and the height of free energy barrier [4, 6, 13–17], and the extent of dynamic/static disorder in the pulled molecules [18, 19]. Quantitative understandings into structure and energetics of DNA [6, 20], RNA [4, 5, 7], and proteins [21–27], which are difficult to decipher using conventional ensemble experiments, have become available by decoding the origin of molecular responses under tension.

Thanks to major advances in SM force measurement, developments in theories, computations, and experiments have cross-fertilized to produce comprehensive understanding of biomolecules on nanometer scales. In particular, steered molecular dynamics simulations pioneered by Schulten and coworkers made a major contribution to the field at the early stage of the development [28, 29]. These studies, which have become the standard tool for experimentalists, have been used to visualize the sequence of unfolding events that might occur during the unraveling of proteins under tension at atomic resolution. However, the time scale gap between the computational approach using atomically detailed model and experiment has not been resolved despite progress in computer technologies. In a practically limited computation time, large forces and loading rates (or pulling speeds) are typically used to accelerate the unbinding or unfolding processes. It is likely that the use of artificially large forces inevitably alters the unfolding pathways from those elucidated from force experiments [30]. The currently accessible time scale accessible in atomistic detailed simulations is $\lesssim \mathcal{O}(1)$ ms [31]. This restrains the minimum loading rate that can be reached by computer simulation to at best 10^4 pN/s if one were to study the forced

unfolding of a simple RNA hairpin, which reversibly unfolds and folds at $f = 10 - 20$ pN at $r_f \approx 1$ pN/s [4]. This is four orders of magnitude faster than the experimental loading rate ~ 1 pN/s used in LOT. The force ramped at rates faster than the relaxation process of the molecules confines their conformations to a narrow phase space. Therefore, it is necessary to explore alternate computational approaches to probe forced-unfolding.

A decade ago, in an effort to study forced-unfolding dynamics of RNA and proteins at time scales compatible to those in force measurements, we proposed a new class of minimal coarse-grained models, called the self-organized (SOP) polymer model [30, 32, 33]. The premise underlying the creation of the SOP model is that unfolding pathways of biomolecule is mainly determined by the overall network of inter-residue interactions in the native state, which are disrupted stochastically under a given perturbation, and the chain connectivity along the polymer backbone. The SOP model allowed us to simulate small-to-medium sized biomolecules at loading rates comparable to the real experiments [30, 32]. In the process, we demonstrated that the unfolding pathways of biomolecule are explicitly dependent on the pulling speed [30]. Furthermore, we revealed that the unfolding pathways of GFP bifurcate into two alternative pathways (opening of GFP barrel from either N- or C- terminus), which was difficult to capture by analyzing force-extension curves from AFM pulling experiments alone [22]. These predictions were not only unanticipated in experiments but are difficult to predict using standard molecular dynamics simulations.

The Gaussian Network Model (GNM) [34] (or elastic network model (ENM) [35, 36]), used to describe conformational fluctuations and intra-molecular allosteric communications of biomolecules in the native states [37–41], provides an alternative way to predict the force-induced rupture of RNA and proteins. Conformational fluctuations of proteins in their native state is faithfully described by low-frequency normal modes [39], which can be captured by using the topological information in the native contact maps. The low frequency modes remain insensitive to the details of interaction, such as strength and nature of interactions (vdW, H-bond, electrostatic). This observation justifies the use of coarse-grained representations of molecular systems for simulating at large length and time scales. Indeed, GNM is excellent in reproducing the relative amplitude of residue fluctuation, which is often compared with Debye-Waller factor.

Here, adapting the study of Srivastava and Graneck, who considered thermal and force-induced unfolding of proteins by extending the formalism of GNM [42], we explore force-induced

rupture of RNA molecules. In the framework of GNM, the dynamic cross-correlation between residue fluctuations at finite temperature is related to the inverse Kirchhoff matrix, containing information on network connectivity. The response of the system under force is related to squared inverse Kirchhoff matrix. Mechanical force can, in principle, be applied to any site in the structure, revealing the allosteric communication spread throughout the structure. In addition, the force-induced unfolding experiment is mimicked by applying forces on the two ends (or two distinct sites). Increasing forces induces a large fluctuation in the non-covalently bonded residue pairs ($|i - j| \geq 2$) and disrupts them, which can be implemented by modifying the connectivity matrix at each value of the applied force. We show that GNM using topological information alone can satisfactorily unveil the forced-unfolding pathway of RNA molecule at equilibrium.

II. THEORY

Gaussian network model under force: One of the aims of using GNM [34] is to elucidate the natural conformational dynamics arising solely from the native state contact map. The reasonable agreement between theoretically computed mean square fluctuation (MSF) and the crystallographic B-factor suggests that residue-specific interactions are of secondary importance in characterizing the dynamics of a biomolecule in the native state [34]. This is further substantiated by a number of studies that show fluctuations and correlated motions of the native conformation are satisfactorily described solely using the information on native contacts.

In GNM, vectors linking residue pairs in contact ($\mathbf{R}_{ij} \equiv \mathbf{R}_i - \mathbf{R}_j$) relative to those in the native state $\mathbf{R}_{ij}^o \equiv \mathbf{R}_i^o - \mathbf{R}_j^o$ are harmonically restrained to have the following form,

$$\mathcal{H}_0 = \frac{\gamma}{2} \sum_{(i,j)} (\mathbf{R}_{ij} - \mathbf{R}_{ij}^o)^2 = \frac{\gamma}{2} \sum_{(i,j)} (\delta\mathbf{R}_i - \delta\mathbf{R}_j)^2. \quad (1)$$

where (i, j) denotes the list of residue pairs in contact satisfying the geometrical condition $|\mathbf{R}_{ij}^o| < R_c$ in the native state. The native state molecular coordinates are obtained from the structures in the Protein Data Bank (PDB). Let $\delta\mathbf{R}_i \equiv \mathbf{R}_i - \mathbf{R}_i^o$ be the deviation of the i -th bead from its coordinate in the native state. This Hamiltonian given above can be written as $\mathcal{H}_0 = \frac{\gamma}{2} \delta\mathbf{R}^T \cdot \mathbf{\Gamma} \cdot \delta\mathbf{R}$ where $\delta\mathbf{R} = (\delta\mathbf{R}_1, \delta\mathbf{R}_2, \dots, \delta\mathbf{R}_N)^T$. Although one could consider an

improvement of the model, GNM [34] merely uses a uniform bond strength γ regardless of the type of the bond. This is sufficient to illustrate the essential ideas. The Laplacian (Kirchhoff) matrix $\mathbf{\Gamma}$ where Γ_{ij} results from the second derivative of harmonic potential with respect to $\delta\mathbf{R}_i$ and $\delta\mathbf{R}_j$. Note that the residue motion is “isotropic” in GNM, which reduces $3N - 6$ degrees of freedom to $N - 1$. The cross-correlation between the i -th and j -th residues is obtained from the partition function $Z_N = \int \mathcal{D}[\delta\mathbf{R}] e^{-\beta\mathcal{H}_0} = (2\pi k_B T / \gamma)^{3(N-1)/2}$ as,

$$\langle \delta\mathbf{R}_i \cdot \delta\mathbf{R}_j \rangle = -\frac{2k_B T}{\gamma} \frac{\partial \log Z_N}{\partial \Gamma_{ij}} = \frac{3k_B T}{\gamma} (\mathbf{\Gamma}^{-1})_{ij}. \quad (2)$$

The inverse Kirchhoff matrix can be computed by decomposing it into eigenmodes using $\mathbf{\Gamma}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T = \sum_{k=2}^N \lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T$ with λ_k and \mathbf{u}_k being the k -th eigenvalue and eigenvector, respectively. Note that the lowest mode ($k = 1$), corresponding to the overall translation, is removed from the sum. In the absence of force ($f = 0$), the self-correlation (or mean square fluctuation, MSF) $\langle \delta\mathbf{R}_i^2 \rangle$ is compared with the crystallographic (Debye-Waller) temperature factor, $B_i = \frac{8\pi^2}{3} \langle (\delta\mathbf{R}_i)^2 \rangle$, to determine the value of the parameter γ . The cut-off distance R_c is also determined by appealing to experimental B-factors. An optimum choice of cut-off distance R_c , which minimizes the difference between the temperature factor and self-correlation is expected. For RNA, when the center of mass in each nucleotide is set to be the position of each nucleotide, $R_c \sim (10 - 20) \text{ \AA}$ reproduces reasonable results for B-factor with $\gamma \sim 10 \text{ pN}\cdot\text{nm}/\text{\AA}^2$. For all the RNAs tested here, we chose $R_c = 15 \text{ \AA}$. This choice of R_c also provides better sensitivity to the system in the bond rupture process. When $R_c = 15 \text{ \AA}$ is used, we find that the difference between B_i and $8\pi^2/3 \times \langle (\delta\mathbf{R}_i)^2 \rangle$ for P4P6 domain (PDB code: 1gid) is minimized with $\gamma \approx 7.4 \text{ pN}\cdot\text{nm}/\text{\AA}^2$.

Allosteric communication, inter-residue cross-correlation under tension. When the k -th residue of the molecule is perturbed by mechanical force, the Hamiltonian becomes

$$\mathcal{H} = \frac{\gamma}{2} \delta\mathbf{R}^T \cdot \mathbf{\Gamma} \cdot \delta\mathbf{R} - \mathbf{f}_k^T \cdot \delta\mathbf{R}. \quad (3)$$

\mathbf{f}_k is the general force vector acting on the k -th residue; thus $\mathbf{f}_k = f \mathbf{e}_k$. The cross-correlation between the i -th and j -th residues under mechanical perturbation at k -th residue is obtained from the partition function,

$$Z_N(\mathbf{f}_k) = \int \mathcal{D}[\delta\mathbf{R}] e^{-\beta\mathcal{H}} = \left(\frac{2\pi k_B T}{\gamma} \right)^{3(N-1)/2} [\det \mathbf{\Gamma}]^{-3/2} \exp \left(\frac{\mathbf{f}_k^T \cdot \mathbf{\Gamma}^{-1} \cdot \mathbf{f}_k}{2\gamma k_B T} \right) \quad (4)$$

as follows:

$$\begin{aligned}
\langle \delta \mathbf{R}_i \cdot \delta \mathbf{R}_j \rangle &= -\frac{2k_B T}{\gamma} \frac{\partial \log Z_N(\mathbf{f}_k)}{\partial \Gamma_{ij}} \\
&= \mathbf{T}_{ij} + \mathbf{F}_{ij}^k \\
&= \frac{3k_B T}{\gamma} (\mathbf{\Gamma}^{-1})_{ij} + \frac{f^2}{\gamma^2} (\mathbf{\Gamma}^{-1})_{ki} (\mathbf{\Gamma}^{-1})_{jk}
\end{aligned} \tag{5}$$

Because of the linear coupling between force and displacement in Eq. (3) the fluctuation-fluctuation cross-correlation ($\langle \delta \mathbf{R}_i \cdot \delta \mathbf{R}_j \rangle$) is fully decomposed into temperature (\mathbf{T}_{ij}) and force (\mathbf{F}_{ij}^k) contributions. The f -dependent term in the Eq.5 results from the following matrix calculus: $\partial(\mathbf{\Gamma}\mathbf{\Gamma}^{-1})/\partial\Gamma_{ij} = 0$ and $\partial\mathbf{\Gamma}/\partial\Gamma_{ij} = \mathbf{e}_i\mathbf{e}_j^T$, thus, $\partial\mathbf{\Gamma}^{-1}/\partial\Gamma_{ij} = -\mathbf{\Gamma}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{\Gamma}^{-1}$. It is noteworthy that the thermal fluctuations (\mathbf{T}_{ij}) are decoupled from the effect of the applied perturbation. The allosteric cross-correlation between the two distinct parts of the structure due to mechanical perturbation on k -th residue is gleaned by inspecting \mathbf{F}_{ij}^k .

Force-extension curves: When tension is applied to the two ends of a molecule at $i = 1$ and N , the energy Hamiltonian is written as

$$\mathcal{H} = \frac{\gamma}{2} \delta \mathbf{R}^T \cdot \mathbf{\Gamma}_f \cdot \delta \mathbf{R} - (\mathbf{f}_N^T - \mathbf{f}_1^T) \cdot \delta \mathbf{R}. \tag{6}$$

and the corresponding partition function and dynamical cross-correlation are:

$$Z_N = \int \mathcal{D}[\delta \mathbf{R}] e^{-\beta \mathcal{H}} = \left(\frac{2\pi k_B T}{\gamma} \right)^{3(N-1)/2} [\det \mathbf{\Gamma}]^{-3/2} \exp \left(\frac{(\mathbf{f}_N^T - \mathbf{f}_1^T) \cdot \mathbf{\Gamma}_f^{-1} \cdot (\mathbf{f}_N - \mathbf{f}_1)}{2\gamma k_B T} \right) \tag{7}$$

and

$$\begin{aligned}
\langle \delta \mathbf{R}_i \cdot \delta \mathbf{R}_j \rangle &= -\frac{2k_B T}{\gamma} \frac{\partial \log Z_N}{\partial (\Gamma_f)_{ij}} \\
&= \mathbf{T}_{ij} + \mathbf{F}_{ij}^{N,1} \\
&= \frac{3k_B T}{\gamma} (\mathbf{\Gamma}_f^{-1})_{ij} + \frac{f^2}{\gamma^2} [(\mathbf{\Gamma}_f^{-1})_{Ni} - (\mathbf{\Gamma}_f^{-1})_{1i}] [(\mathbf{\Gamma}_f^{-1})_{jN} - (\mathbf{\Gamma}_f^{-1})_{j1}].
\end{aligned} \tag{8}$$

Here, we make the f -dependence of Kirchhoff matrix explicit by putting a subscript, which becomes clear in the algorithm used to calculate force-extension curves (see below).

The equilibrium force-extension curve is computed using Eq.8. For the residue pairs making bond, say a pair between i -th and j -th bead, mean square fluctuation $\langle (\delta \mathbf{R}_{ij})^2 \rangle \equiv \langle (\delta \mathbf{R}_i -$

$\langle \delta \mathbf{R}_j \rangle^2 = \langle (\delta \mathbf{R}_i)^2 \rangle + \langle (\delta \mathbf{R}_j)^2 \rangle - 2\langle \delta \mathbf{R}_i \cdot \delta \mathbf{R}_j \rangle$. To be specific,

$$\langle (\delta \mathbf{R}_{ij})^2 \rangle = \langle (\delta \mathbf{R}_{ij})^2 \rangle_T + \langle (\delta \mathbf{R}_{ij})^2 \rangle_f \quad (9)$$

where

$$\begin{aligned} \langle (\delta \mathbf{R}_{ij})^2 \rangle_T &= \frac{3k_B T}{\gamma} [(\mathbf{\Gamma}_f^{-1})_{ii} + (\mathbf{\Gamma}_f^{-1})_{jj} - 2(\mathbf{\Gamma}_f^{-1})_{ij}] \\ \langle (\delta \mathbf{R}_{ij})^2 \rangle_f &= \frac{f^2}{\gamma^2} [(\mathbf{\Gamma}_f^{-1})_{Ni} - (\mathbf{\Gamma}_f^{-1})_{1i}]^2 \\ &\quad + \frac{f^2}{\gamma^2} [(\mathbf{\Gamma}_f^{-1})_{Nj} - (\mathbf{\Gamma}_f^{-1})_{1j}]^2 \\ &\quad - 2\frac{f^2}{\gamma^2} [(\mathbf{\Gamma}_f^{-1})_{Ni} - (\mathbf{\Gamma}_f^{-1})_{1i}] [(\mathbf{\Gamma}_f^{-1})_{jN} - (\mathbf{\Gamma}_f^{-1})_{j1}] \end{aligned}$$

For a given f , we assume that a bound residue pair between i and j is disrupted when

$$\phi \equiv \frac{\langle (\delta \mathbf{R}_{ij})^2 \rangle_f}{\langle (\delta \mathbf{R}_{ij})^2 \rangle_T} > 1. \quad (10)$$

Breakage of the non-covalent bond between i and j pair with $(\mathbf{\Gamma}_f)_{ij} = -1$ and $|i - j| \geq 2$ is implemented in the computer algorithm by modifying the elements of the Kirchhoff matrix as

$$\begin{aligned} (\mathbf{\Gamma}_f)_{ij} = -1 &\longrightarrow 0 \\ (\mathbf{\Gamma}_f)_{ji} = -1 &\longrightarrow 0 \\ (\mathbf{\Gamma}_f)_{ii} &\longrightarrow (\mathbf{\Gamma}_f)_{ii} - 1 \\ (\mathbf{\Gamma}_f)_{jj} &\longrightarrow (\mathbf{\Gamma}_f)_{jj} - 1. \end{aligned} \quad (11)$$

The original Kirchhoff matrix is modified at each value of an increasing f whenever the condition $\phi > 1$ is identified for any pair of residue until all the non-covalent residue pairs are ripped. The extension is computed as the size of free ends using the Kirchhoff matrix, $\mathbf{\Gamma}_f$. Thus, reconstruction of Kirchhoff matrix with increasing force produces force-extension curve. We selected $\phi = 0.002$, so that the transition force for P4P6 domain is at $f_c \sim (10 - 15)$ pN.

III. RESULTS

We applied the GNM formalism under mechanical perturbation to four RNA molecules whose 3D structures are available: (i) 55-nt domain IIa of the Hepatitis C Viral (HCV) genome

(PDB code: 1p5m) [43]; (ii) prohead RNA (PDB code: 1foq) [44]; (iii) P4P6 domain of *T.ribozyme* (PDB code: 1gid) [45]; (iv) *T. ribozyme* (the atomic coordinates of a theoretically modeled *T. thermophila* ribozyme, TtLSU.pdb, was obtained from the Group I and II sections in the website <http://www-ibmc.u-strasbg.fr/upr9002/westhof>) [46]. We first discuss the effect of mechanical perturbation on the dynamical cross-correlation using *T. ribozyme*, followed by the calculation of force-extension curves for all four RNAs.

Allosteric communication of *T. ribozyme* in the presence of mechanical force:

Using the Kirchhoff matrix $\mathbf{\Gamma}$ of *T. ribozyme* we calculated the temperature-induced dynamical cross-correlation with MSF, $\langle(\delta\mathbf{R})^2\rangle$ (Fig.2) and the force-induced cross-correlation matrix \mathbf{F}^k for $k = 1, \dots, N$ without the prefactor f^2/γ^2 (Fig.3 and Supporting Movie). Some of the results which exhibit dramatic effect of tension on cross-correlation are demonstrated along with the map of secondary structure of *T. ribozyme* (Fig.3), and full results of perturbations are provided in the Supporting Movie. The extent of response to the perturbation is larger for those residues with large MSF that are not constrained by the rest of the structure. A perturbation on P9.2, which is clearly isolated from other parts (see contact map and 3D structure in Fig.1), induces large cross-correlations over the entire structure. Strong correlation is found between two sites when they are in direct contact in the structure. For example, the perturbation on $k = 325$, a part of P9 helix, induces enhanced fluctuation on P5 helix which is in direct contact with P9 via tertiary interactions. There is a strong cross-correlation between P5b and P6b when $k = 151$ is perturbed, between P2 and P5c for $k = 45$, between P2.1 and P9.1a for $k = 74$, between the residues consisting of the internal multiloop formed among P9, P9.1, P9.2 for $k = 336$. While one may argue that these are trivial outcomes of the contact information shown in Fig.1C, quantitative prediction of the force-induced, site-dependent cross-correlations demonstrated in Fig.3 is not possible without calculating \mathbf{F}_{ij}^k . Furthermore, the *secondary* cross-correlation between P9.2 and P6b, P9.1a and P7 induced by perturbing P2.1 ($k = 74$) is also difficult to predict *a priori*.

In summary, the entire network responds to mechanical stress at a particular site. Although force exerted to the network is linear, the response of the network (*T. ribozyme*) is not trivial, but depends on the network structure. The presence of secondary cross-correlation is viewed as a realization of allosteric communication over the entire molecular architecture.

Equilibrium force-extension curve: We computed the equilibrium force-extension curves of four RNA molecules. All the force-extension curves exhibit a single sharp transition around 10–15 pN, whose value can be adjusted by changing the value of ϕ . Unlike the FECs observed in LOT measurements or simulations which use a finite pulling speed, the details of domain-by-domain unraveling are not clear enough to extract from FEC. This is because the FEC from GNM is by nature an outcome at full thermodynamic equilibrium, which corresponds to the situation when the molecule is pulled reversibly using a slow pulling speed [47]. Even if there is a possibility of unfolding bifurcation, as in our previous Brownian dynamics simulation study of the forced-unfolding of *T. ribozyme* using the SOP model [30] it cannot be extracted from a purely elastic network model. If time is long enough to allow the *T. ribozyme* to fully sample the distinct configurations along the bifurcating pathways, the molecular extension is effectively a value averaged over such stochastic variations of individual trajectories.

While FEC from GNM in itself is of little use to decode the order of unfolding events, the number of native contacts in each residue, which can be read out from the diagonal elements of the Kirchhoff matrix ($Q_i = (\mathbf{\Gamma}_f)_{ii}$) at each f , provides us with information on how each residue loses its contact shared with others as f increases.

Ila domain of HCV IRES RNA: Both force-extension curve and history of contact disruption (Fig.4A) indicate that this hairpin with several bulges unfolds in an all-or-none fashion. However, our previous study using Brownian dynamics simulation of SOP model on this RNA has revealed two intermediate states between native and fully unfolded states [32]. The absence of signature of intermediate states is due to the inability of using finite loading rate in the GNM in contrast to the BD simulation used to probe forced-unfolding of Ila domain [32].

Prohead RNA (Two way junction): This two way junctions unfolds via [P1]→[P3]→[P2] (Fig.4B), which is in full accord with the our previous BD simulation result using SOP model at finite loading rate [32].

P4P6 domain: [P6,P6a,P6b] unfolds after the tertiary contacts between P5b and P6b is disrupted, which is followed by the unfolding of P5 helix, and P5abc domain unfolds at the

final stage. This agrees well with LOT experiment [5].

T. ribozyme: Like other RNA, the unfolding is initiated from the two ends of the ribozyme which composing the peripheral domains [P1], [P2,P2.1], and [P9,P9.1,P9.2], and bond disruptions proceed into the core regions including [P4,P5,P6] and [P3,P7,P8]. The evolution profile of bond rupture also captures the early events of tertiary contact disruptions, such as the breakages of P2-P5c and P2.1-P9.1a kissing loop pairs and P3 helix before the disruptions of core helices occur. Of note, in our previous work [30] on the mechanical unfolding simulations of *T. ribozyme* using SOP model, we found two alternative pathways: (1) [N]→[P9.2]→[P9.1, P9, P9.1a]→[P2]→[P2.1]→[P3, P7, P8]→[P6]→[P4, P5]→[P5a, P5b, P5c]; (2) [N]→[P2]→[P2.1]→[P9.2]→[P9, P9.1, P9.1a]→[P3, P7, P8]→[P6]→[P4, P5]→[P5a, P5b, P5c]. The first pathway is compatible with the experimentally inferred pathway is [N]→[P9.2]→[P9.1]→[P9, P9.1a]→[P2, P2.1]→[P3, P7, P8]→[P6, P4]→[P5]→[P5a, P5b, P5c]. Given the simplicity of the method, the general agreement of the unfolding pathway predicted by GNM with that from the LOT experiment [5] is remarkable.

IV. CONCLUDING REMARKS

To a first approximation, the conformational dynamics of biomolecules in response to the influence of the ambient temperature and mechanical perturbation is encoded in the three dimensional structure of the molecule [48]. By using this finding we showed that the force response of RNA in the native state can be well described with GNM which incorporated the 3D topology of the molecule in terms of Kirchhoff matrix. Importantly, the pattern of cross-correlation over the molecular architecture under mechanical perturbation differs from the thermal cross-correlation (Fig.2) and also shows a large variation depending on where the force is applied (Fig.3, SI Movie).

The accuracy of our prediction on dynamical cross-correlation ($\langle \delta \mathbf{R}_i \cdot \delta \mathbf{R}_j \rangle$) which is used to capture the essence of conformational dynamics is expected to increase with the growing system size (N) because main contributions to the value of $\langle \delta \mathbf{R}_i \cdot \delta \mathbf{R}_j \rangle$ come from low frequency modes. This is evident in the mathematical structure, $\langle \delta \mathbf{R}_i \cdot \delta \mathbf{R}_j \rangle \sim \left(\sum_{k=2}^N \lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T \right)_{ij}$ with

$\lambda_2 < \lambda_3 < \dots < \lambda_N$ [34].

Finally, given the simplicity of the model, the extent of agreement of the predicted unfolding pathways from GNM with those identified by SM force measurement is surprising. We note that simulation of forced-unfolding process at the experimentally meaningful pulling speed using even the minimal SOP model is time-consuming for large RNA molecules. While the current model can be further improved to obtain better agreement with experimentally determined properties, we rediscover that a number of key features of interest (conformational fluctuation, allostery, unfolding/folding pathways) are already encoded in the 3D native topology [30, 48]. We should note, however, subtle features such as pathway bifurcation cannot be obtained by GNM which only uses the contact map of the native state. Nevertheless, because of the simplicity the proposed method for RNA unfolding and the earlier study on proteins [42] show that approximate features of tension-induced rupture can be gleaned from the present model.

Acknowledgement We are pleased to dedicate this work to Klaus Schulten who had the vision decades ago that computations could become an integral part of solving major problems in biology. His pioneering theoretical and computational studies is a testimony to that vision. We thank the KIAS Center for Advanced Computation for providing computing resources for this work. This work was supported in part by a grant (CHE 16-36424) from the National Science Foundation to DT.

-
- [1] L. D. Landau and E. Lifshitz, *Course of Theoretical Physics* **3**, 109 (1986).
 - [2] E. Evans and K. Ritchie, *Biophys. J.* **72**, 1541 (1997).
 - [3] T. Strick, J. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, *Science* **271**, 1835 (1996).
 - [4] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco Jr., and C. Bustamante, *Science* **292**, 733 (2001).
 - [5] B. Onoa, S. Dumont, J. Liphardt, S. B. Smith, I. Tinoco, and C. Bustamante, *Science* **299**, 1892 (2003).
 - [6] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block, *Science* **314**, 1001 (2006).
 - [7] W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside, and S. M. Block, *Science* **319**,

630 (2008).

- [8] A. Dittmore, D. B. McIntosh, S. Halliday, and O. A. Saleh, *Phys. Rev. Lett.* **107**, 148301 (2011).
- [9] J. Wilhelm and E. Frey, *Phys. Rev. Lett.* **77**, 2581 (1996).
- [10] B.-Y. Ha and D. Thirumalai, *J. Chem. Phys.* **103**, 9408 (1995).
- [11] B. Y. Ha and D. Thirumalai, *J. Chem. Phys.* **106**, 4243 (1997).
- [12] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, *Science* **265**, 1599 (1994).
- [13] I. Tinoco Jr., *Ann. Rev. Biophys. Biomol. Struct.* **33**, 363 (2004).
- [14] C. Hyeon and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10249 (2003).
- [15] C. Hyeon and D. Thirumalai, *J. Phys.: Condens. Matter* **19**, 113101 (2007).
- [16] C. Hyeon, G. Morrison, and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9604 (2008).
- [17] M. Hinczewski, J. C. M. Gebhardt, M. Rief, and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4500 (2013).
- [18] C. Hyeon, M. Hinczewski, and D. Thirumalai, *Phys. Rev. Lett.* **112**, 138101 (2014).
- [19] M. Hinczewski, C. Hyeon, and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3852 (2016).
- [20] C. Danilowicz, V. W. Coljee, C. Bouzigues, D. K. Lubensky, D. R. Nelson, and M. Prentiss, *Proceedings of the National Academy of Sciences* **100**, 1694 (2003).
- [21] M. Rief, H. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, *Science* **276**, 1109 (1997).
- [22] M. Mickler, R. I. Dima, H. Dietz, C. Hyeon, D. Thirumalai, and M. Rief, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20268 (2007).
- [23] P. E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez, *Nature* **402**, 100 (1999).
- [24] J. M. Fernandez and H. Li, *Science* **303**, 1674 (2004).
- [25] J. Stigler, F. Ziegler, A. Gieseke, J. Gebhardt, and M. Rief, *Science* **334**, 512 (2011).
- [26] P. I. Zhuravlev, M. Hinczewski, S. Chakrabarti, S. Marqusee, and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **113**, E715 (2016).
- [27] D. K. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6166 (1999).
- [28] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, *Biophys. J.* **75**, 662 (1998).
- [29] B. Isralewitz, M. Gao, and K. Schulten, *Curr. Opin. Struct. Biol* **11**, 224 (2001).
- [30] C. Hyeon, R. I. Dima, and D. Thirumalai, *Structure* **14**, 1633 (2006).
- [31] D. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. Dror, M. Eastwood, J. Bank, J. Jumper,

- J. Salmon, Y. Shan, et al., *Science* **330**, 341 (2010).
- [32] C. Hyeon and D. Thirumalai, *Biophys. J.* **92**, 731 (2007).
- [33] A. Zhmurov, A. E. Brown, R. I. Litvinov, R. I. Dima, J. W. Weisel, and V. Barsegov, *Structure* **19**, 1615 (2011).
- [34] T. Haliloglu, I. Bahar, and B. Erman, *Phys. Rev. Lett.* **79**, 3090 (1997).
- [35] M. M. Tirion, *Phys. Rev. Lett.* **77**, 1905 (1996).
- [36] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, *Biophys. J.* **80**, 505 (2001).
- [37] C. Chennubhotla and I. Bahar, *Mol. Syst. Biol.* **2**, msb4100075 (2006).
- [38] W. Zheng, B. R. Brooks, S. Doniach, and D. Thirumalai, *Structure* **13**, 565 (2005).
- [39] W. Zheng, B. R. Brooks, and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7664 (2006).
- [40] I. Bahar and A. J. Rader, *Curr. Opin. Struct. Biol.* **15**, 586 (2005).
- [41] F. Tama, M. Valle, J. Frank, and C. L. Brooks III, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9319 (2003).
- [42] A. Srivastava and R. Granek, *Phys. Rev. Lett.* **110**, 138101 (2013).
- [43] P. J. Lukavsky, I. Kim, G. A. Otto, and J. D. Puglisi, *Nature. Struct. Biol.* **10**, 1033 (2003).
- [44] A. A. Simpson, Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, et al., *Nature* **408**, 745 (2000).
- [45] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, et al., *Science* **273**, 1678 (1996).
- [46] V. Lehnert, L. Jaeger, F. Michel, and E. Westhof, *Chemistry & Biology* **3**, 993 (1996), URL <http://www-ibmc.u-strabg.fr/upr9002/westhof/Group/TtLSU.pdb>.
- [47] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, *Science* **296**, 1832 (2002).
- [48] D. K. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7254 (2000).

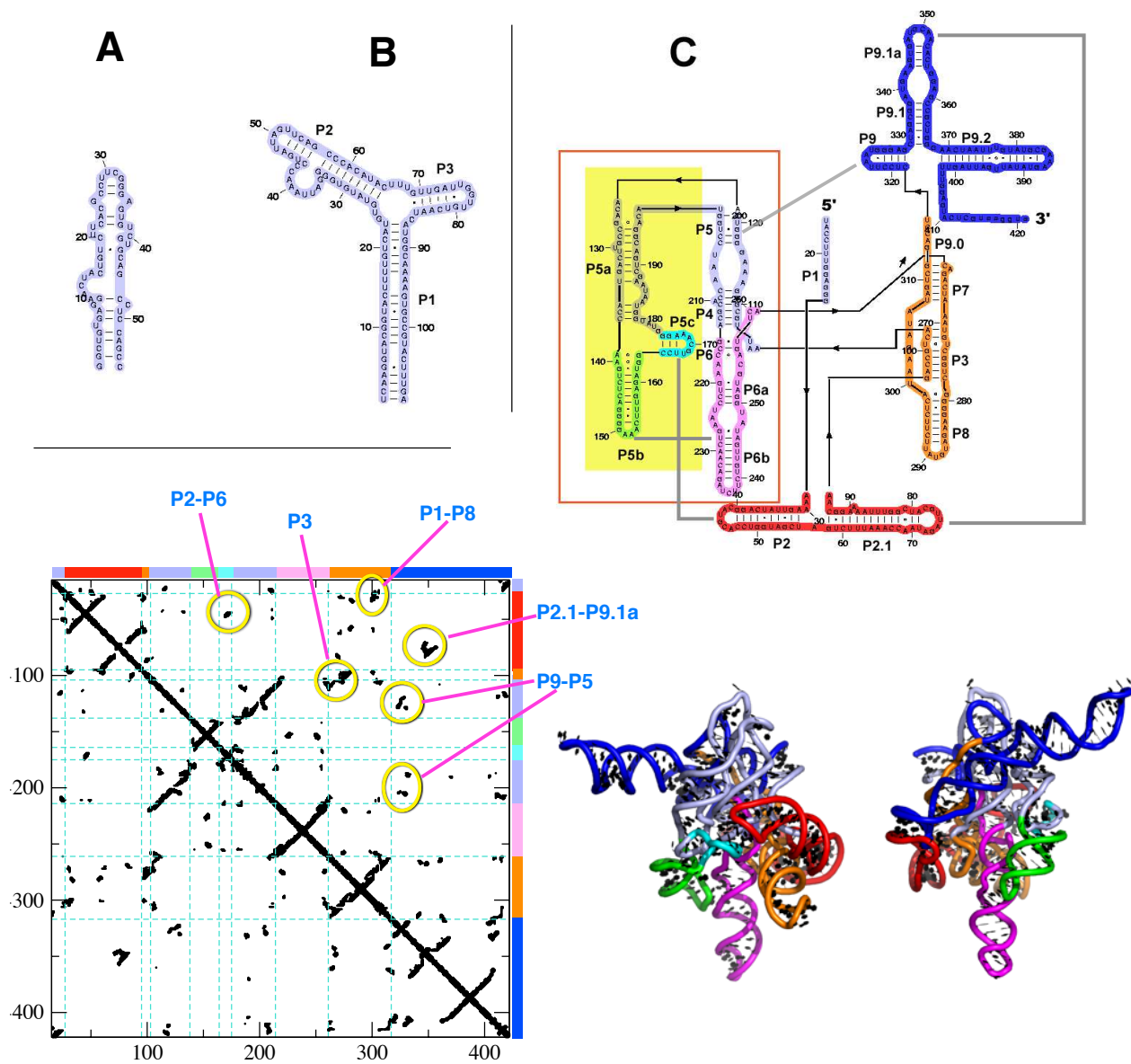


FIG. 1: Secondary structures of RNA. (A) IIa domain of HCV IRES RNA (PDB code: 1p5m). (B) prohead RNA (PDB code : 1foq). (C) *Tetrahymena* ribozyme. The region enclosed by a box corresponds to P4P6 domain (PDB code: 1gid). The contact map of *T. ribozyme* and 3D structures are shown at the bottom. Some of the key tertiary contacts are highlighted with yellow circles.

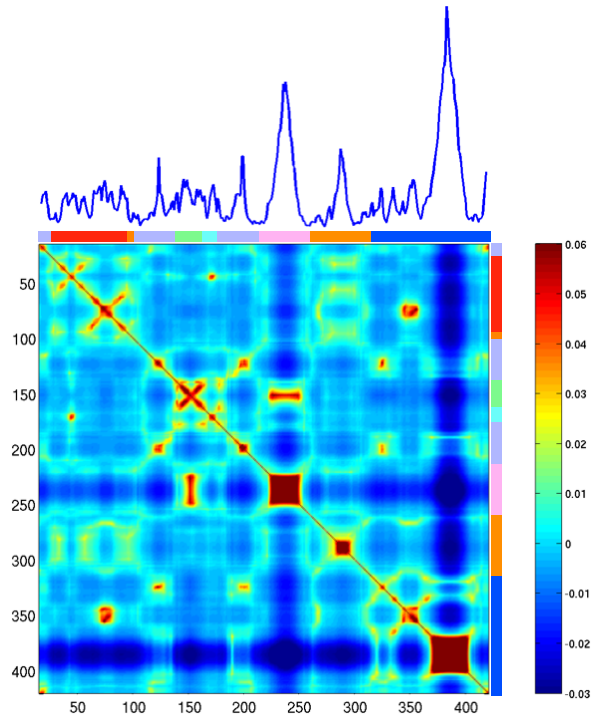


FIG. 2: Dynamic cross-correlation (\mathbf{T}_{ij}) of *T.* ribozyme at $f = 0$. Mean square fluctuation is shown on the top.

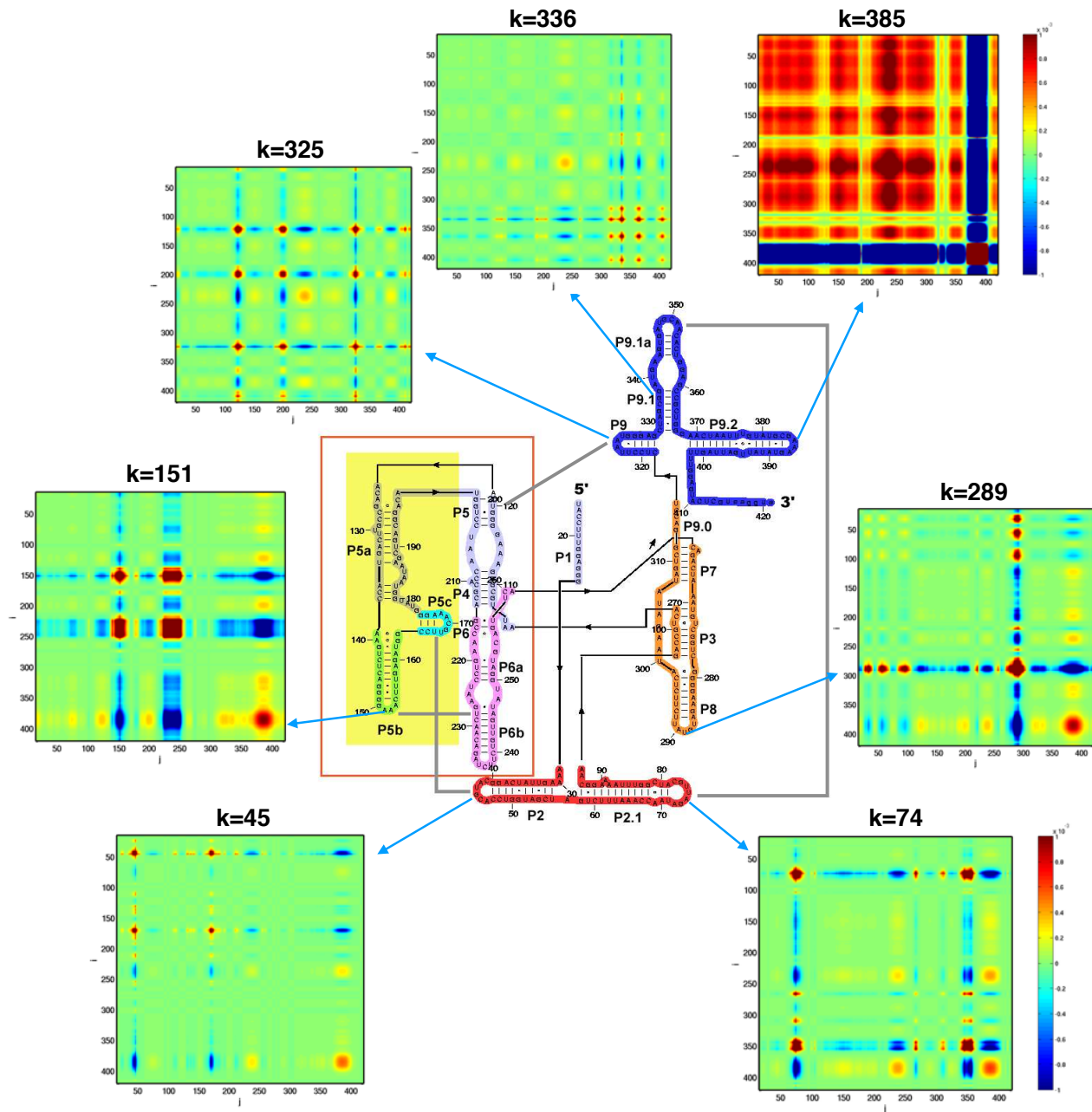


FIG. 3: Allosteric communication over the *T* ribozyme structure when k -th residue is perturbed by mechanical force. Cross-correlation matrices \mathbf{F}^k are shown for seven different k values. Location of each the perturbed site is marked with the arrow in the secondary structure. All the results for $k = 1, \dots, N$ are available in the Supporting Movie.

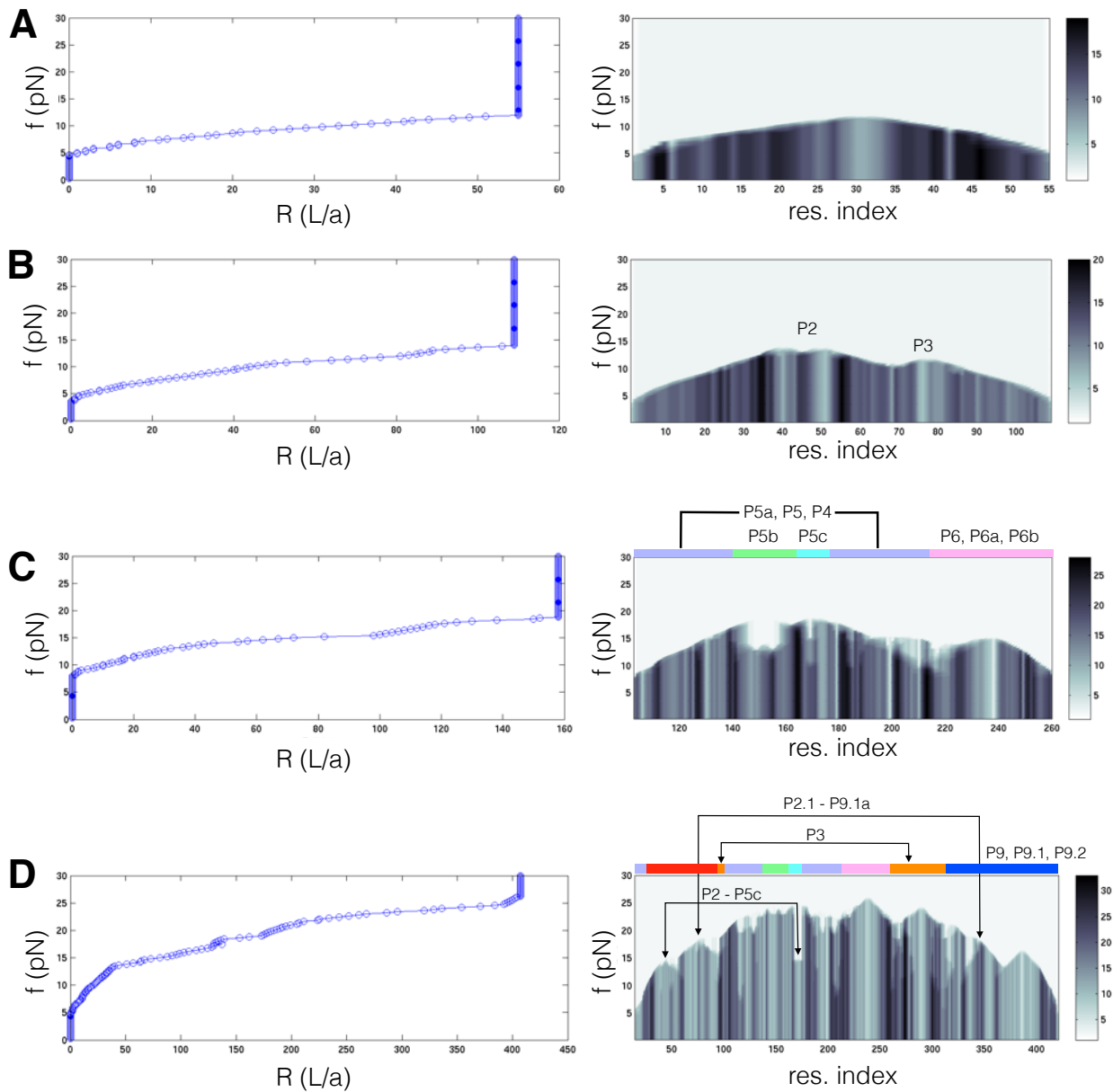


FIG. 4: Force-extension curves and number of native contacts ($Q_i = (\Gamma_f)_{ii}$, diagonal elements of Kirchhoff matrix) with increasing force calculated for (A) Ila domain of HCV IRES RNA (PDB code: 1p5m), (B) prohead RNA (PDB code : 1foq), (C) P4P6 domain, and (D) *Tetrahymena* ribozyme.