Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

STRUCTURE-FUNCTION DYNAMICS HYBRID MODELING: RNA DEGRADATION

Hua Zheng
Wei Xie
Chunsheng Fang
Wandi Xu

Mechanical and Industrial Engineering
Northeastern University
360 Huntington Ave
Boston, MA 02115, USA

Paul C. Whitford
Ailun Wang

Department of Physics and
Center for Theoretical Biological Physics
Northeastern University
360 Huntington Ave
Boston, MA 02115, USA

ABSTRACT

RNA structure and functional dynamics play fundamental roles in controlling biological systems. Molecular dynamics simulation, which can characterize interactions at an atomistic level, can advance the understanding on new drug discovery, manufacturing, and delivery mechanisms. However, it is computationally unattainable to support the development of a digital twin for enzymatic reaction network mechanism learning, and end-to-end bioprocess design and control. Thus, we create a hybrid ("mechanistic + machine learning") model characterizing the interdependence of RNA structure and functional dynamics from atomistic to macroscopic levels. To assess the proposed modeling strategy, we consider RNA degradation which is a critical process in cellular biology that affects gene expression. The empirical study of RNA lifetime prediction demonstrates the promising performance of the proposed multi-scale bioprocess hybrid modeling strategy.

1 INTRODUCTION

Understanding RNA structure and functional dynamics directly influences drug (e.g., mRNA vaccines) discovery, manufacturing, and delivery. RNA structure affects: a) its functions and interactions with other molecules, such as DNA, proteins, and ions; and b) the regulation of enzymatic reaction network. For example, RNA structure directly affects stability, translation, and delivery efficacy of RNA vaccines. The rate of RNA degradation depends on various factors, including pH, temperature, and ionic concentrations.

RNA structure is described in terms of three levels: primary, secondary, and tertiary. The primary structure is the nucleotide sequence of the RNA molecule, represented by four base letters (i.e., A, U, C, G). The secondary structure refers to the pattern of hydrogen bonding (base pairing) along the chain (i.e., helices), while the tertiary structure denotes the final 3D shape of the RNA molecule, determined by both secondary structure hydrogen bonding and additional nucleotide interactions. In recent years, there is a surge in deep learning-based algorithms for biomolecular 3D structure prediction; for example the success of AlphaFold2 (Jumper et al. 2021) has garnered the most attention. However, the field of RNA structure prediction and structure-function dynamics modeling remains largely unexplored within the OR community, presenting potential opportunities for new insights and advancements (Xie and Pedrielli 2022).

The functions of RNA molecules are closely linked to structure and dynamics. Computer simulations, in particular molecular dynamics (MD) methods, allow structural dynamics of biomolecular systems to be investigated with unprecedented temporal and spatial resolution; see Šponer et al. (2018). However, MD simulations are complex and time-consuming. Typically, all-atom MD simulations can only probe structure conformation change at very short time scales, i.e., 10^{-6} second.

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

To model conformational dynamics and ensure scientific interpretability, we propose a hybrid ("mechanistic + machine learning") model characterizing the interdependencies of RNA structure-function dynamics from atomistic to macroscopic levels. This approach is based on physics-based dimensional reduction to describe some key properties: (1) interatomic interactions and potential energies quantifying global connectivity and atomic interdependencies; (2) solvation shells that approximate the aggregated effects of diffuse ions; and (3) free energy barriers and lifetimes for RNA conformational changes. *The proposed hybrid modeling strategy provides insight into energetics and dynamics of enzymatic reactions and biomolecular conformational change, which supports regulation mechanism learning and reaction rate prediction.*

To assess its performance, we consider RNA degradation, as measured based on unfolding times, and introduce an RNA lifetime hybrid network model (RNA-LifeTime) to quantify the structural changes of RNA during the degradation processes. By analyzing the lifetime of native contacts (defined based on interactions found in the folded molecule), RNA-LifeTime provides valuable insights into structural stability of RNA molecules under varying environmental conditions, such as temperature and ionic concentrations, and serves as a versatile probe for exploring the RNA degradation processes. We employ the effective energy potential associated with atom-atom electrostatic interactions and solvent-mediated ionic interactions as the driving force. The proposed RNA-LifeTime can improve the prediction of RNA degradation rate.

In summary, we made the following contributions: 1) We developed the RNA-LifeTime hybrid model that effectively incorporates 3D structural information and conformational dynamics of biomolecules in a diffuse ionic environment; 2) To the best of our knowledge, RNA-LifeTime is the first 3D molecular hybrid model capable of predicting the kinetics of RNA degradation and unfolding; 3) We develop a potential energy aggregate model employing an effective 3D spatial modeling technique called "multi-headed Gaussians;" 4) The empirical study demonstrates the efficacy of our approach and shows that RNA-LifeTime can achieve high accuracy on RNA degradation rate estimation; and 5) By offering accurate probabilistic predictions for RNA lifetimes, our work provides valuable insights into the factors governing RNA folding processes, which establishes the groundwork for advancements in RNA-based therapeutics and diagnostics.

The organization of the paper is as follows. In Section 2, we present the physical foundation for the proposed hybrid model on RNA structure-function dynamics. Then, we introduce the RNA-LifeTime model and utilize MD simulations of RNA degradation processes to evaluate the performance of our proposed approach and compare it with baseline models in Sections 3 and 4. We conclude this paper in Section 5.

2 RNA STRUCTURE-FUNCTION AND MOLECULAR DYNAMICS

The proposed RNA structure-function dynamic hybrid model can facilitate the learning of enzymatic reaction network regulation mechanisms through MD simulations. We consider RNA structural dynamics (e.g., degradation and unfolding rate) as a function of environment conditions, including ion concentrations and temperature. At any time t, RNA structure evolution can be modeled with a state transition function, where the state $\boldsymbol{s}_t = (\mathbf{X}_t, \boldsymbol{z}_t)$ includes the RNA structure, denoted by $\mathbf{X}_t$ (or key features characterizing RNA structure-function), and the environment conditions denoted by $\boldsymbol{z}_t$,

$$\boldsymbol{s}_{t+1} = f(\boldsymbol{s}_t; \boldsymbol{\beta}_t(\boldsymbol{s}_t)) + \boldsymbol{e}_t, \tag{1}$$

where $\boldsymbol{\beta}_t(\boldsymbol{s}_t)$ represents the kinetic parameters (such as degradation rate) characterizing the regulation mechanism. The residual $\boldsymbol{e}_t$ represents the impact from other factors and model error of $f(\cdot)$.

The proposed model is built on the scientific understanding: (1) interatomic interactions and potential energy in Section 2.1.1, accounting for RNA system global connectivity and interdependencies; (2) Gaussian mixture distribution approximating the aggregate effect from hydration shells in Section 2.1.2; and (3) free energy barrier crossing required for RNA conformation change in Section 2.1.3; and (4) RNA structure dynamics balancing the driving forces induced by energy potential and thermodynamics in Section 2.2.

2.1 RNA Structure and Environmental Impact

As a cyber-physical system, an RNA system is composed of atoms with charge; see Figure 1. The interactions between atoms from the same RNA molecule include: (1) short-range bonded interactions such as bond stretching, angle bending, and torsion, which can be related to the physical network; and (2) long-range non-bonded interactions, such as van der Waals and electrostatic forces, which can be related to the cyber network. These interactions give rise to the potential energy that governs folding and conformational changes of the RNA structure. At the same time, atoms have random vibrations due to thermal energy. *Therefore, the balance of driving forces introduced by the potential energy and thermal energy influences critical pathways and reaction rates of regulatory reaction networks, e.g., RNA stability.*

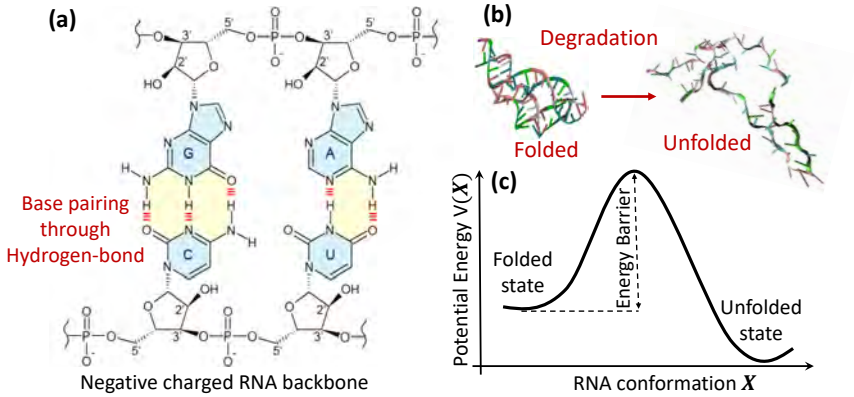

Figure 1: RNA structural dynamics is influenced by atom interactions and thermodynamics: **(a)** Base pairing through hydrogen bond; **(b)** RNA conformational change; **(c)** Free-energy barrier crossing for (b).

2.1.1 RNA System Structure-Function and Potential Energy

RNA function depends on its structure, which is typically specified in terms of primary, secondary, and the tertiary atomic structure. Primary structure defines an RNA sequence in terms of its constituent nucleotides. The secondary structure of RNA molecule, including the 2D structure, is determined by the formation of hydrogen bonds between complementary base pairs (i.e., A=U and C=G base pairs). The 2D structure can impact the accessibility of the RNA molecule to other molecules, such as proteins and ions, and can therefore affect its function. Further building on the basic architecture defined by the 2D structure, the 3D structure is determined by the interactions between the atoms in the RNA molecule, and it is also important for biological functions, e.g., stability, activity, and binding specificity. For example, RNA molecules with a stable 3D structure may be less susceptible to degradation, while RNA molecules with a flexible 3D structure may be more dynamic and adaptable to different environments.

Atom-atom bonded and non-bonded interactions determine the potential energy, i.e., $V = V_{bonded} + V_{non-bonded}$, which induce the constraints on RNA structural dynamics and are associated with an energy barrier to undergo a conformational change (see Figure 1). The bonded interactions play a crucial role of defining the precise stereochemistry of the molecule. These short-range interactions are typically modeled using harmonic, or periodic potentials, as implemented using semi-empirical potentials, such as the AMBER force field, which take into account the interactions between bonded and angles and dihedrals potentials.

The non-bonded interactions, occurring between atoms that are distant in the RNA sequence, are typically modeled using distance-dependent potentials, such as the Lennard-Jones potential and the Coulomb potential. The Lennard-Jones potential represents the van der Waals (VDW) interactions between atoms, while the Coulomb (C) potential represents the electrostatic interactions between charged atoms: $V_{non-bonded} =$

$V_C + V_{VDW}$. The Coulomb potential is induced by any pair of atoms i and j with charge q_i and q_j,

$$V_C = \sum_{ij \in \mathscr{R}} \frac{q_i q_j}{4\pi\varepsilon\varepsilon_0 r_{ij}}, \tag{2}$$

where r_{ij} is the interatomic distance, $\mathscr{R}$ represents the set of atoms in the RNA system, ε is the dielectric constant for water, and ε_0 is the permittivity of free space.

2.1.2 RNA-Ion Interactions and Effect on Energy Potential

RNA structure can be impacted by the ion concentration environment through ion-mediated electrostatic interactions. Here we consider ion-RNA interactions and study the impact on the energy potential (Wang et al. 2022). Positively charged ions (e.g., Mg^{2+}) can interact with negatively charged RNA backbone through Coulomb electrostatics (see Figure 2). Due to the creation of *inner- and outer- hydration shells*, the solvent environment around RNA is generally described as containing chelated and diffuse ions. In the inner-shell, chelated ions are partially dehydrated, which allows them to form strong direct contacts with RNA. As a result, chelated ions can remain bound on RNA for millisecond time scale. In the outer-shell, diffuse ions remain fully hydrated (e.g., $Mg(H_2O)_6^{2+}$) and associate less strongly with RNA. Despite the transient and weak influence of individual diffuse ions, the behavior of a diffuse ion is primarily determined by longer-range electrostatic interactions, and their collective effect on RNA structure can be significant. Therefore, we consider the *aggregated effects* of ion-RNA interactions on energy potential due to the existence of inner- and outer-shells during the RNA structural dynamics hybrid modeling development.

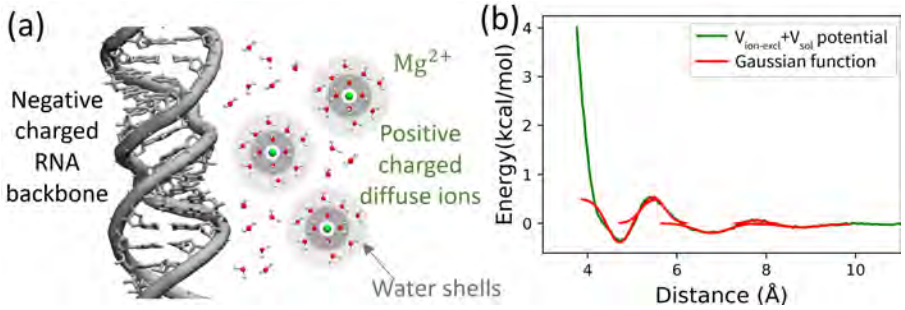

Figure 2: **(a)** An illustration of RNA and diffuse ion interactions. **(b)** Gaussian mixture (red line) approximates the energy potential of ions excluding Coulomb potential, $V_{ion-excl} + V_{sol}$ in Equation (3) (green line).

The electrostatic potential energy (V_E) includes Coulomb interactions (V_C'), and effective potentials that describe ionic solvation effects (V_{sol}) and excluded volume of ions ($V_{ion-excl}$) (Wang et al. 2022),

$$V_E = V_C' + [V_{sol} + V_{ion-excl}] = \sum_{ij \in \mathscr{R}} \frac{q_i' q_j'}{4\pi\varepsilon\varepsilon_0 r_{ij}} + \left[\sum_{ij \in \mathscr{R}} \left(\sum_{k=1}^{C_g} B^{(k)} e^{-C^{(k)}\left[r_{ij} - R^{(k)}\right]^2} \right) \right]. \tag{3}$$

The potential energy V_C' represents the direct Coulomb interactions with updated charges q_i' and q_j' accounting for the impact from chelated ions with strong contacts with RNA molecules. For simplification, the mean-field approach is used here. The potential energies related to solvent-mediated ionic interactions and excluded volume of ions, represented by $V_G \equiv V_{sol} + V_{ion-excl}$, are modeled by a sum of C_g Gaussians, characterizing the aggregated effect of diffuse ions in the hydration shells. It accounts for up to three outer hydration shells, as shown in Figure 2 (adapted from Wang et al. (2022)).

For RNA stability analysis, the energy function used to study the difference of energy between folded and unfolded state can be simplified. First, we ignore the change in bond potential because the bonds are too strong to break unless the temperature is extremely high. Second, the harmonic angle potential is similar and it is less related to the native contact definition used to measure RNA folding. Third, the Coulomb potential is a key driving force of inter-atomic interactions that contribute to native contacts. *Thus, the energy barrier used in the paper for RNA degradation rate estimation is related to V_C' and $V_{sol} + V_{ion-excl}$.*

2.1.3 Free-Energy Barrier Crossing for Conformational Change

The potential energy V imposes constraints on the RNA structure by favoring conformations that have lower local potential energy. When we have the conformational change from state $\mathbf{X}_1$ to $\mathbf{X}_2$, the *energy barrier*, defined as $\Delta G = \max G(\mathbf{X}) - G(\mathbf{X}_1)$, represents the difference in the free energy between the starting state $\mathbf{X}_1$ and the maximum, denoted by $\max G(\mathbf{X})$, occurring during the transition process; as shown in Figure 1. Differing with the potential energy, free energy includes the entropic contribution. Then, the rate of accommodation denoted by k_a (similar to the rate acceleration in enzymatic reactions) and the mean-first passage time (e.g., the lifetime of contacts transitioning from the folded to the unfolded state) denoted by T have the following relationship with ΔG,

$$k_a = \frac{1}{T} \propto \frac{1}{C_a} \exp\left(-\frac{\Delta G}{k_B \mathbb{T}}\right), \tag{4}$$

where k_B is the Boltzmann constant, $\mathbb{T}$ is the temperature, and C_a is the barrier-crossing attempt frequency. Methods for estimating these prefactors or coefficients are described in Whitford et al. (2010). For the RNA degradation process with monotonic change in free energy, we have $\Delta G = G(\mathbf{X}_1) - G(\mathbf{X}_2)$. *The energy barrier, referring to the amount of energy that must be supplied to a biomolecular system in order for it to undergo a particular transformation, is influenced by factors, such as temperature and ion concentration.*

The hybrid model we propose can extend to estimate reaction rates, uncover regulatory mechanisms, and facilitate optimal learning and control in enzymatic reaction networks, where the catalytic power of enzymes depends on factors like temperature, pH, and ion concentration.

2.2 RNA Structure-Function Dynamics and Thermodynamics

Langevin dynamics is used to determine RNA structure evolution accounting for: a) regulation from the potential energy gradient; and b) thermal energy modelled by Brownian motion due to *temperature effects*. For any RNA system composed of N' atoms with masses $\mathbf{m}$ and coordinates $\mathbf{X}_t$, Langevin equation states,

$$\mathbf{m}\ddot{\mathbf{X}}_t = -\nabla V(\mathbf{X}_t) - \gamma \dot{\mathbf{X}}_t + \sqrt{2\gamma k_B \mathbb{T}}\mathscr{G}_t \tag{5}$$

where the gradient $-\nabla V(\mathbf{X}_t)$ over the energy potential gives the driving force calculated from the atoms interaction potentials, $\dot{\mathbf{X}}_t$ is the velocity, $\ddot{\mathbf{X}}_t$ is the acceleration, γ is friction coefficient, $\mathbb{T}$ is the temperature, k_B is Boltzmann's constant, and $\mathscr{G}_t$ is a delta-correlated stationary Gaussian process with zero mean representing thermal fluctuations. RNA structural dynamics can be studied using a variety of computational techniques, including MD simulations, Monte Carlo simulations, and coarse-grained models.

3 RNA DEGRADATION DYNAMICS HYBRID MODELING

While the RNA backbone often remains stable during the degradation process, the disruption of interatomic interactions can lead to changes in the secondary and tertiary structures of RNA. *To quantify these structural changes, native contacts of residues or atoms are used in this study.* In RNA folding and degradation processes, they refer to the natural interactions, such as base-pairing and ionic interactions, between the residues or atoms of RNA molecules in their folded 3D structure. Native contacts are defined by using the Shadow Contact Map algorithm (Noel et al. 2012) with cutoff parameters from Wang et al. (2022). In addition, the fraction of native contacts is used to measure the deviation from the native folded state of RNA structure through MD simulations (Wang et al. 2019). By identifying changes in native contacts, our model sheds light on how the structural integrity of RNA molecules changes over the degradation process.

Notation: The numbers of simulation trajectories, types of native contact, and environmental features are denoted by N_s, C_t, and C_z respectively. Denote the number of residues in the input primary sequence by N. Sequences of varying lengths were padded with zeros at the end to ensure they were of equal length. Use $\odot$ for the element-wise multiplication, $\otimes$ for the outer product, and $\oslash$ for element-wise division. We use $[x]$

to denote the sequence of positive integers from 1 to x, where x is an integer and $\lceil x \rceil$ to denote the smallest integer greater than or equal to x. We denote the standard Dropout with the operator $\texttt{Dropout}_x$, where x is the dropout rate, i.e., the probability setting an entry to zero. Use `BatchNorm` for the batch normalization, such that the mean and standard deviation are calculated per dimension over the mini-batches. We use $\texttt{L}(\cdot)$ for a linear transformation with a weight vector or matrix, denoted by $\boldsymbol{w}$ and W, and a bias vector $\mathbf{b}$. The feedforward network is two-layer fully connected with ReLu activation function having the hidden layer dimension multiplier h, i.e., a factor by which the dimensions of the hidden layers are scaled

$$\texttt{FeedForward}(\mathbf{X};\boldsymbol{b}_3,\boldsymbol{b}_4,\mathbf{W}_3,\mathbf{W}_4) = \boldsymbol{b}_4 + \mathbf{W}_4^\top \texttt{ReLu}(\boldsymbol{b}_3 + \mathbf{W}_3^\top \mathbf{X}) \text{ with ReLu(x)=max}(0,x).$$

3.1 Native Contact and Its Lifetime

The lifetime of a native contact has been introduced by Best et al. (2013) to quantify the importance of an individual inter-residue contact in the protein folding mechanism. We adopt this concept and apply it to RNA degradation process (Figure 3) by studying the trajectory of fraction of native contacts, i.e., the ratio of the number of contacts present in the current structure to the number of contacts in the native structure.

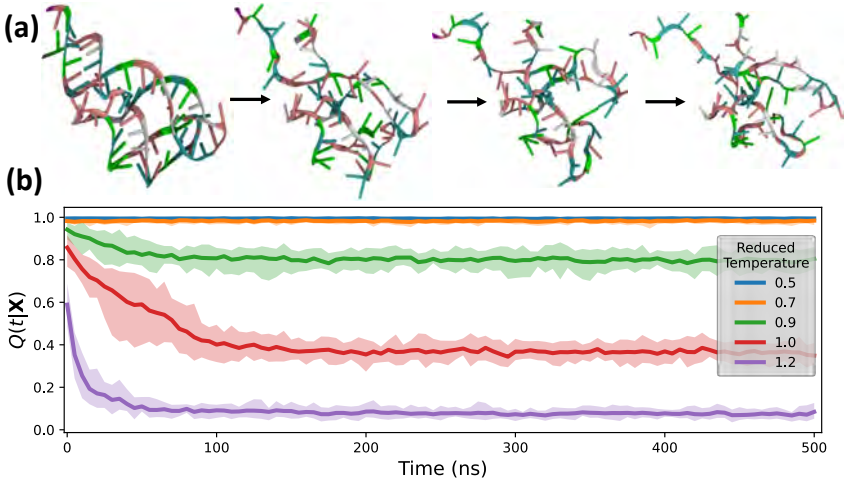

Figure 3: **(a)** The degradation of RNA (PDB 1Y26) in the first 20 ns of MD simulation. **(b)** The fraction of native contacts $Q(t|\mathbf{X}_0,\boldsymbol{z})$ is changing with time at varying reduced temperatures. The reduced temperature is a unit used in MD simulation where 0.5 corresponds to room temperature, and 1.0 corresponds to 350K.

As the RNA molecule degrades, the fraction of native contacts reflects the percentage of the sequence that remains integrated. Let $\mathbf{X}_0$ represent the initial native conformation of an RNA sequence of length N, with each element encoding location, base, and environmental condition of the residue. For simplification, suppose the environmental conditions (i.e., temperature, ionic concentration) are fixed, denoted by $\boldsymbol{z}$. Let T_{ij} represent the lifetime of native contact pair (i,j) and r_{ij} the distance between residue i and j for any $i,j \in [N]$. The probability of the native contact pair (i,j) remaining integrated until time t is modelled by

$$\Pr(T_{ij} > t|\mathbf{X}_0,\boldsymbol{z}) \approx \frac{1}{1+e^{\beta(r_{ij}(t|\mathbf{X}_0,\boldsymbol{z})-\lambda r_{ij}(0|\mathbf{X}_0,\boldsymbol{z}))}} \quad \textbf{(Lifetime Probability)},$$

where $\beta = 50$ Å^{-1} is a smoothing parameter and the factor $\lambda = 1.2$ accounts for thermodynamic fluctuation when contact is formed (Best et al. 2013). Let $\mathbb{C}(\mathbf{X}_0)$ represent the set of native contact pairs for initial conformation $\mathbf{X}_0$. Then, following Best et al. (2013), we define the integrity function of the RNA molecule,

$$Q(t|\mathbf{X}_0,\boldsymbol{z}) \equiv \frac{1}{|\mathbb{C}(\mathbf{X}_0)|} \sum_{(i,j)\in\mathbb{C}(\mathbf{X}_0)} \Pr(T_{ij} > t|\mathbf{X}_0,\boldsymbol{z}) \quad \textbf{(Integrity Function)}.$$

Thus, we interpret $Q(t|\mathbf{X}_0,\boldsymbol{z})$ as the averaged probability of native contacts remaining by time t. As the key characteristics of RNA integrity, the probability (6) will be used as the **output** of our predictive model.

3.2 RNA Lifetime Modeling

In this study, the native conformation of the RNA molecule is represented by the inter-residue distances and their contact types, i.e., $\mathbf{X}_0 = (\mathbf{D}, \mathbf{C})$, where the (i, j)-entry of the matrix $\mathbf{D}$ represents the pair distance between residues i and j in the folded native state; the (i, j, k)-entry of the tensor $\mathbf{C}$ represents the k-th type of contact between residues i and j in the folded native state. We consider *input* as the initial conformation of a RNA molecule and the environmental conditions, that is $\mathbf{S}_0 = (\mathbf{X}_0, \boldsymbol{z})$. The input $\mathbf{S}_0$ corresponds to the initial state of the RNA system in Equation (1), i.e., $\boldsymbol{s}_0 = \text{vec}(\mathbf{S}_0) \equiv (\text{vec}(\mathbf{D}), \text{vec}(\mathbf{C}), \text{vec}(\boldsymbol{z}))^\top$. For simplification, suppose the environmental conditions $\boldsymbol{z}_t$ are fixed during the RNA degradation process; thus drop t subscript. At any time t, the *output* of interest is the lifetime probability matrix,

$$\mathbf{Y}(t) = [y_{ij}(t)]_{N\times N} \quad \text{where } y_{ij}(t) = \begin{cases} \Pr(T_{ij} > t|\mathbf{S}_0) & \text{if } (i,j) \in \mathbb{C}(\mathbf{X}_0); \\ 0 & \text{otherwise.} \end{cases}$$

Then, we can calculate the observed fraction of native contacts by $Q(t|\mathbf{S}_0) = \frac{1}{|\mathbb{C}(\mathbf{X}_0)|}\sum_{(i,j)\in\mathbb{C}(\mathbf{X}_0)} y_{ij}(t)$.

Through connecting the lifetime of a native contact with the potential energy difference in Equation (4), we can express the logarithm of the lifetime T_{ij} of a native contact (i, j) as a function of potential energy barrier for the RNA degradation, represented by ΔG_{ij}, i.e., $\mathtt{L}(\log T_{ij}) = \frac{\Delta G_{ij}(\mathbf{S}_0)}{\mathtt{L}_\mathbb{T}(\mathbb{T})} + \sigma_{ij}(\boldsymbol{z})W$, where $\sigma(\boldsymbol{z}) > 0$ is a scale parameter function, $\mathtt{L}(\cdot)$ and $\mathtt{L}_\mathbb{T}(\cdot)$ are linear functions of the logarithm of time $\log T_{ij}$ and temperature $\mathbb{T}$, and the noise W follows a standard normal distribution. The term $\sigma_{ij}(\boldsymbol{z})W$ is introduced to account for the random motion induced by thermodynamics; see Equation (5).

Then the lifetime probability of the native contact $(i, j) \in \mathbb{C}(\mathbf{X}_0)$ is modeled by

$$\begin{aligned} H_{ij}(t|\mathbf{S}_0) &\equiv \Pr(T_{ij} \geq t|\mathbf{S}_0) = \Pr\left(\mathtt{L}(\log T_{ij}) \geq \mathtt{L}(\log t)|\mathbf{S}_0\right) = \Pr\left(W \geq \frac{\mathtt{L}(\log t) - \Delta G_{ij}(\mathbf{S}_0)/\mathtt{L}_\mathbb{T}(\mathbb{T})}{\sigma_{ij}(\boldsymbol{z})}\right) \\ &= 1 - \Phi\left(\frac{\mathtt{L}(\log t) - \Delta G_{ij}(\mathbf{S}_0)/\mathtt{L}_\mathbb{T}(\mathbb{T})}{\sigma_{ij}(\boldsymbol{z})}\right), \text{ if } W \sim \mathscr{N}(0,1). \end{aligned}$$

Assume $H_{ij}(t|\mathbf{S}_0) = 0$ if $(i, j) \notin \mathbb{C}(\mathbf{X}_0)$. Then we can have the lifetime probability model in a matrix form

$$\mathbf{H}(t|\mathbf{S}_0) = \left[1 - \Phi\left(\frac{\mathtt{L}(\log t) - \Delta\mathbf{G}(\mathbf{S}_0)/\mathtt{L}_\mathbb{T}(\mathbb{T})}{\sigma(\boldsymbol{z})}\right)\right] \odot \mathbf{M} \tag{6}$$

where $\Delta\mathbf{G}(\mathbf{S}_0)$ has (i, j)-entry $\Delta G_{ij}(\mathbf{S}_0)$ and $\mathbf{M}$ is the matrix where (i, j)-entry is one if $(i, j) \in \mathbb{C}(\mathbf{X}_0)$ and 0 otherwise. In sum, $\mathbf{M}$ is a binary matrix that masks out the elements of $\mathbb{C}(\mathbf{X}_0)$ from the original matrix.

We conclude this section by connecting the lifetime probability model with the famous accelerated failure time (AFT) model in survival analysis (Lee and Wang 2003). While both models assume that the covariate affects the log-time scale in an additive form, our proposed model offers a more flexible functional expression by incorporating the effects of environmental conditions and RNA structures.

3.3 RNA Lifetime Hybrid Network Model

Algorithm 1 presents the architecture of RNA-LifeTime. The mean function of potential difference, denoted as $\Delta\mathbf{G}(\mathbf{S}_0)$, incorporates the potential energy associated with electrostatic interactions. The functional expressions for both Coulomb potential energy and the potential energy related to solvent-mediated ionic interactions and excluded volume of ions are derived from Wang et al. (2022) with the trainable weights. The potential energy related to solvent-mediated ionic interactions and excluded volume of ions is represented by a sum of Gaussians (see Line 3-7). For each type of interactions considered, a Gaussian mixture with C_g components is used to approximate the aggregated effect from hydration shells, with distinct Gaussian parameters designated for each specific interaction type. *The newly introduced mechanism, referred to as*

"multi-headed Gaussians", is an effective architecture for modeling the molecular 3D structural information. For a coarse-grained model, the Coulomb interaction of two residues (i, j) is dominated by the negatively charged phosphate group located along RNA backbone, i.e., $q_i = q_j = -1$. Therefore, the Coulomb interactions in Equation (2) between any two RNA residues can be approximated as $V_{C,ij} = \frac{W_{2,ij}}{r_{ij}}$, where the trainable parameter $W_{2,ij}$, representing the term $\frac{q_i q_j}{4\pi\varepsilon\varepsilon_0}$, is a pairwise potential parameter that captures the strength of the electrostatic interaction between residues i and j (see Line 10).

Algorithm 1: RNA Lifetime Hybrid Network Model (RNA-LifeTime)

Input: The pair distance matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$, native contact tensor $\mathbf{C} \in \mathbb{R}^{N\times N\times C_t}$, mask matrix $\mathbf{M} \in \{0,1\}^{N\times N}$, environmental feature $\boldsymbol{z} \in \mathbb{R}^{C_z}$, number of contact types $C_t = 10$, number of Gaussians C_g, and hidden layer dimension multiplier $h = 4$. **Output**: $\mathbf{H}(t|\mathbf{S}_0;\boldsymbol{\theta}) \in [0,1]^{N\times N}$

Function LifetimeModel(t, $\mathbf{D}$, $\mathbf{C}$, $\mathbf{M}$, $\boldsymbol{z}$, $C_t = 10$, $C_z = 2$, $C_g = 5$, $h = 4$): Trainable Parameters

1 # Solvent-mediated ionic interactions as a sum of Gaussians

2 $\mathbf{J} = \mathbf{C}\mathbf{W}_1$ $\quad \mathbf{W}_1 \in \mathbb{R}^{C_t\times C_g\times 3}$

3 **for all** $i \in \{1,2,\ldots,C_g\}$ **do**

4 $\quad \mathbf{B}_i, \mathbf{\Lambda}_i, \mathbf{R}_i = \mathbf{J}_{\cdot\cdot i1}, \mathbf{J}_{\cdot\cdot i2}, \mathbf{J}_{\cdot\cdot i3} \quad \left(\mathbf{B}_i, \mathbf{\Lambda}_i, \mathbf{R}_i \in \mathbb{R}^{N\times N}\right)$

5 $\quad \mathbf{V}_G^{(i)} = \mathbf{B}_i \odot \exp\left[-\mathbf{\Lambda}_i \odot (\mathbf{D} - \mathbf{R}_i) \odot (\mathbf{D} - \mathbf{R}_i)\right]$

6 **end for**

7 $\mathbf{V}_G = \sum_{i=1}^{C_g} \mathbf{V}_G^{(i)}$

8 $\bar{\mathbf{V}}_G = \texttt{Sigmoid}\left(\texttt{L}^G(\boldsymbol{z}; b_1, \boldsymbol{w}_1)\right) \cdot \mathbf{V}_G$ $\quad b_1 \in \mathbb{R}^1, \boldsymbol{w}_1 \in \mathbb{R}^{C_z}$

9 # Coulomb electrostatic potential as a reciprocal of distance

10 $\mathbf{V}_C = \mathbf{W}_2 \oslash \mathbf{D}$ $\quad \mathbf{W}_2 \in \mathbb{R}^{N\times N}$

11 $\bar{\mathbf{V}}_C = \texttt{Sigmoid}\left(\texttt{L}^C(\boldsymbol{z}; b_2, \boldsymbol{w}_2)\right) \cdot \mathbf{V}_C$ $\quad b_2 \in \mathbb{R}^1, \boldsymbol{w}_2 \in \mathbb{R}^{C_z}$

12 # Total effective potential energy by following Equation (3)

13 $\mathbf{V}_1 = \texttt{Dropout}_{0.2}\left(\bar{\mathbf{V}}_G + \bar{\mathbf{V}}_C\right)$

14 $\mathbf{V}_2 = \mathbf{V}_1 + \texttt{FeedForward}(\mathbf{V}_1; \boldsymbol{b}_3, \boldsymbol{b}_4, \mathbf{W}_3, \mathbf{W}_4)$ $\quad \boldsymbol{b}_3 \in \mathbb{R}^{hN}, \boldsymbol{b}_4 \in \mathbb{R}^N$

15 $\Delta\mathbf{G}(\mathbf{S}_0) = \texttt{BatchNorm}(\mathbf{V}_2)$ $\quad \mathbf{W}_3 \in \mathbb{R}^{N\times(hN)}, \mathbf{W}_4 \in \mathbb{R}^{(hN)\times N}$

16 # Scale parameter model

17 $\sigma(\boldsymbol{z}) = \texttt{Softplus}\left(\texttt{L}^\sigma(\boldsymbol{z}; b_5, \boldsymbol{w}_5)\right)$ $\quad b_5 \in \mathbb{R}^1, \boldsymbol{w}_5 \in \mathbb{R}^{C_z}$

18 # Lifetime probability model by following Equation (6)

19 $y = \texttt{L}(\log t; b_6, w_6)$ $\quad b_6 \in \mathbb{R}^1, w_6 \in \mathbb{R}^1$

20 $\mathbf{H}(t|\mathbf{S}_0;\boldsymbol{\theta}) = \left[1 - \Phi\left(\frac{y - \Delta\mathbf{G}(\mathbf{S}_0)/\texttt{L}_\mathbb{T}(\mathbb{T}; b_6, w_6)}{\sigma(\boldsymbol{z})}\right)\right] \odot \mathbf{M}$ $\quad b_7 \in \mathbb{R}^1, w_7 \in \mathbb{R}^1$

Return $\mathbf{H}(t|\mathbf{S}_0;\boldsymbol{\theta})$ $\quad \boldsymbol{\theta}$: collection of model parameters

The free energy barrier for an RNA system refers to the energy difference between the folded native and unfolded states of the sequence (Bryngelson and Wolynes 1989; Morgan and Higgs 1996). This barrier is crucial in determining the stability and folding kinetics of the RNA. The magnitude of the free energy barrier depends on various factors, such as the RNA sequence, temperature, and ionic concentration. To represent the energy barrier crossing, a sigmoid function is employed to modulate the impact of temperature and ionic concentration on effective potential energies $\bar{\mathbf{V}}_G$ (see Line 8).

In RNA-LifeTime, we use a mean-field approach to account for the effect of ionic concentrations. The positively charged ions (Mg^{2+}) are considered to uniformly neutralize the negativel charge of all residues, ensuring that each residue carries an averaged negative charge between 0 and 1 (i.e., $0 \le q_1 = q_2 = \ldots = q_N < 1$). In the future research, we will consider the unevenly distributed ions. Consequently, the Coulomb potential described in Equation (2) decreases (but remains greater than or equal to zero) as the concentration of positively charged ions increases, due to the fact that $0 \le q_i q_j < 1$. Thus, a sigmoid function is employed to represent the role of ionic concentration in reducing the Coulomb potential (see Line 11). Then, based on Equation (3), the total effective potential energy of each residue pair is calculated as the sum of Coulomb

electrostatic potential and solvent-mediated ionic interactions (see Line 13). It follows by a two-layer fully-connected feedforward network characterizing the effect between native contacts (see Line 14).

The system exhibits increased thermodynamic variability in response to elevated temperatures. Therefore, the scale parameter is modeled by a linear function followed by a softplus activation function, i.e., $\sigma(\boldsymbol{z}) = \texttt{Softplus}\left(\texttt{L}^{\sigma}(\boldsymbol{z}; b_5, \boldsymbol{w}_5)\right)$, as shown in Algorithm 1 Line 17. The softplus function $\texttt{Softplus} = \log(1+\exp(x))$ ensures the output of the linear function is positive and has a smooth, continuous gradient.

3.4 Loss Function and Training Procedure

We express the lifetime probability model $\mathbf{H}(t|\mathbf{S}_0;\boldsymbol{\theta})$ specified by the parameters $\boldsymbol{\theta} = (b_1, b_2, \ldots, b_7, \mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{W}_4, \boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_5, w_6, w_7)$. Given the trajectory observations, denoted by $\mathscr{D} = \{(\mathbf{S}_{0,n}, \mathbf{Y}_n(t_\ell)) \,|\, \ell \in [L], n \in [N_s]\}$, where the output trajectory $\mathbf{Y}(t)$ is measured at discrete time t_ℓ for $\ell \in [L]$. Here N_s is the total number of simulation trajectories, and L is the number of timestep records in each trajectory. The model is trained by minimizing mean absolute error (MAE, $p = 1$)

$$\text{Loss}_p\left(\mathbf{H}(t_\ell|\mathbf{S}_0;\boldsymbol{\theta}), \mathbf{Y}(t_\ell)\right) = \frac{1}{|\mathbb{C}(\mathbf{X}_0)|} \sum_{(i,j)\in\mathbb{C}(\mathbf{X}_0)} \|H_{ij}(t_\ell|\mathbf{S}_0;\boldsymbol{\theta}) - \mathbf{Y}_{ij}(t_\ell)\|_p. \tag{7}$$

We found that degradation displayed an all-or-none behavior. Once triggered, the degradation process quickly finishes: the fraction of native contacts stops decreasing within the first 25 timesteps before fluctuating around a constant. This observation implies that samples from earlier times contain more valuable information about the degradation process (referred to as "positive samples"). Consequently, we implemented a technique to upsample data collected in the early stages of the process (i.e., when $\ell \le U$), where U denotes the upsampling threshold. We define the sampling probability as,

$$P(\ell) = \begin{cases} (U-\ell+2)/\left(\frac{U(U+1)}{2} + L\right) & \text{if } \ell \le U, \\ 1/\left(\frac{U(U+1)}{2} + L\right) & \text{if } \ell > U; \end{cases} \tag{8}$$

which assigns a higher weight to samples closer to the starting time. It is also easy to verify that $\sum_{\ell=1}^{L} P(\ell) = 1$.

Algorithm 2: Model Training Procedure

Input: Dataset $\mathscr{D} = \{(\mathbf{S}_{0,n}, \mathbf{Y}_n(t_\ell)) \,|\, \ell \in [L], n \in [N_s]\}$ where the n-th trajectory input $\mathbf{S}_{0,n}$ includes the pair distance matrix $\mathbf{D}_n \in \mathbb{R}^{N\times N}$, native contact tensor $\mathbf{C}_n \in \mathbb{R}^{N\times N\times C_g}$, $\mathbf{M}_n \in \{0,1\}^{N\times N}$, environmental conditions $\boldsymbol{z}_n \in \mathbb{R}^{C_z}$, upsample threshold U, the number of epochs K, and mini-batch size B.

1 **for all** $k \in \{1, 2, \ldots, K\}$ **do**
2 **for all** $h \in \{1, 2, \ldots, \lceil LN_s/B \rceil\}$ **do**
3 Randomly sample a mini-batch $\mathscr{B} = \{(n,\ell)_b\}_{b=1}^{B}$ from the dataset $\mathscr{D}$ by applying Equation (8)
4 **for all** $(n,\ell) \in \{(n,\ell)_b\}_{b=1}^{B}$ **do**
5 $\mathbf{H}_n(t_\ell|\mathbf{S}_0;\boldsymbol{\theta})$ =LifetimeModel(t_ℓ, $\mathbf{D}_n$, $\mathbf{C}_n$, $\mathbf{M}_n$, $\boldsymbol{z}_n$)
6 **end for**
7 $\mathscr{L}(\boldsymbol{\theta}) = \frac{1}{|\mathscr{B}|}\sum_{(n,\ell)\in\mathscr{B}} \text{Loss}_1\left(\mathbf{H}_n(t_\ell|\mathbf{S}_0;\boldsymbol{\theta}), \mathbf{Y}_n(t_\ell)\right)$ by applying Equation (7)
8 $\boldsymbol{\theta}_k \leftarrow \text{Adam}(\mathscr{L}(\boldsymbol{\theta}_{k-1}))$
Return $\boldsymbol{\theta}_K$

A min-batch gradient descent approach is used to train the RNA-LifeTime hybrid model; see Algorithm 2. This algorithm iteratively trains the model over multiple epochs, where each epoch generates mini-batches of the dataset by applying the upsampling probability in Equation (8). Then, a sample average approximate of the expected loss function is computed to compare the predictions to the actual lifetime probabilities. In the training, we use Adam optimizer with a base learning rate 0.003, $\beta_1 = 0.9$, and $\beta_2 = 0.999$ to search for the best fit parameters $\boldsymbol{\theta}$ for the RNA-LifeTime model. By default, the weights of the Linear layers are initialized using the LeCun (fan-in) initialization strategy (Orr and Müller 1998).

4 EMPIRICAL STUDY

In this section, we study the performance of RNA-LifeTime of multiple single RNA molecules in their degradation processes. We use the classic AFT model with the linear mean function (AFT-Linear) as the baseline, which uses $\boldsymbol{s}_0$ as input, and its scale parameter model $\sigma(\boldsymbol{z})$ is the same as RNA-LifeTime.

4.1 Data Sources and Training Procedure

The RNA molecules (14 in total) are manually selected from PDB database (wwPDB consortium. 2018), and preprocessed using SMOG 2 (Noel et al. 2016). The preprocessing includes the steps: (1) remove water molecules and hydrogen atoms (only consider heavy atoms); (2) add Mg^{2+} and K^+ ions and calculate the total charge of the molecular simulation system; and (3) neutralize the system with Cl^- ions.

The MD simulations were generated by utilizing OpenSMOG (de Oliveira Jr et al. 2022) with the force field calculated by "AA_ions_Wang22.v1" (Wang et al. 2022). Simulation experiments were conducted under eight reduced temperatures (r.t.) ranging from 0.5 to 1.2 and three Mg^{2+} concentrations (0.1mM, 1mM, and 10mM). Here, 0.5 r.t. corresponds to room temperature, and 1.0 r.t. corresponds to 350K. For each combination of RNA conformation, temperature, and Mg^{2+} concentration, we conducted 10 simulation replications, each taking 0.5 hours per GPU. The MD simulations were conducted using 6-8 GPUs and required approximately one week to complete, i.e., $8 \times 3 \times 14 \times 10 \times 0.5 = 1680$ hours/GPU, where 8 is the number of temperature levels, 3 is the number of Mg^{2+} concentration levels, and 14 is the number of RNA molecules. After MD simulations, the simulated trajectories were post-processed to generate pair distance matrices, contact type, environmental features, and the fractions of native contacts. Then they were utilized to train and evaluate the performance of proposed RNA-LifeTime hybrid model.

MD simulation trajectories were padded to 256 residues. For each trajectory, we used the initial state $\mathbf{S}_0$ as the input and selected the lifetime probability $\Pr(T_{ij} > t|\mathbf{S}_0)$ for each $(i,j) \in \mathbb{C}(\mathbf{X}_0)$ at every other timestep within the first 1000 timesteps as outputs, thus utilizing 500 timesteps in total (i.e., $L = 500$). Then we selected five sequences at random and designated their trajectories as the test set. The remaining nine sequences were assigned to the training and validation sets. We trained the model on a cluster of 64 CPUs and 128 GB memory until convergence (i.e., the validation loss stops decreasing after 5 epochs). The mini-batch size, epochs, and upsample threshold were selected to be $B = 512$, $K = 20$, and $U = 25$.

4.2 Model Performance

The expected lifetime $T_{ij} > 0$ can be calculated as follows: $\mathbb{E}[T_{ij}|\mathbf{S}_0] = \int_0^\infty \Pr(T_{ij} > t|\mathbf{S}_0)\mathrm{d}t$. The expected RNA lifetime from MD simulations can be approximated by $\mathbb{E}^{MD}[T_{ij}|\mathbf{S}_0] = \int_0^\infty Q(t|\mathbf{S}_0)\mathrm{d}t \approx \sum_\ell^L Q(t_\ell|\mathbf{S}_0)\Delta t$ with $\Delta t = t_\ell - t_{\ell-1}$. Similarly, the expected RNA lifetime predicted by RNA-LifeTime becomes $\mathbb{E}^{RNA-LT}[T_{ij}|\mathbf{S}_0] = \int_0^\infty H_{ij}(t|\mathbf{S}_0)\mathrm{d}t \approx \sum_\ell^L H_{ij}(t_\ell|\mathbf{S}_0)\Delta t$. Then we define the mean absolute error of expected lifetime (MAE-LT) between MD simulation and RNA-LifeTime prediction as

$$\text{MAE-LT} = \frac{1}{N_s}\sum_{n=1}^{N_s}\frac{1}{|\mathbb{C}(\mathbf{X}_{0,n})|}\sum_{(i,j)\in\mathbb{C}(\mathbf{X}_{0,n})}\left|\mathbb{E}^{MD}[T_{ij}|\mathbf{S}_0] - \mathbb{E}^{RNA-LT}[T_{ij}|\mathbf{S}_0]\right|,$$

where N_s is the total number of simulation trajectories. Overall, the performances are evaluated by three metrics: MAE-LT, MAE (Equation (7) with $p = 1$), mean squared error (MSE) (Equation (7) with $p = 2$).

RNA-LifeTime accurately predicts native contact fractions for RNA molecules; see the representative prediction results in Figure 4. Since it was trained on a small set of MD simulation data containing only nine RNA molecules, whereas AlphaFold 2 was trained on a large dataset of 170,000 protein structures, we anticipate that the accuracy could be further improved by including more simulations of additional RNA molecules in the training set.

The results in Table 1 demonstrate that the RNA-LifeTime hybrid model can achieve high accuracy. Here the MD simulation is supposed as the ground truth model. RNA-LifeTime performs better than

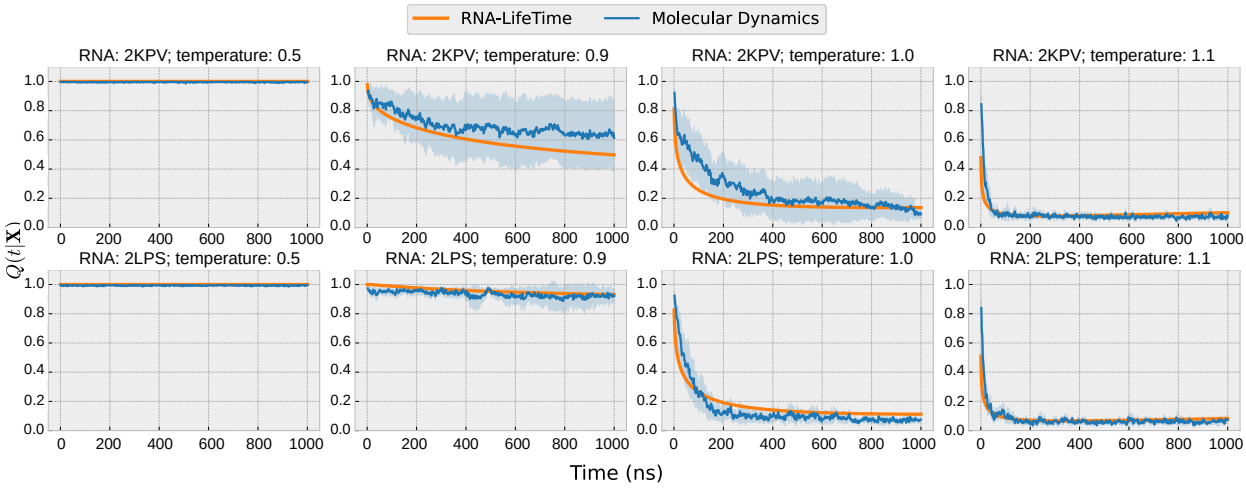

Figure 4: Predicted (orange) vs. MD simulated (blue) trajectories for a Let-7 miRNA (PDB 1ATO, top row) and Yeast ai5(gamma) Group II Intron (PDB 2LPS, bottom row). The results were obtained at a 10 mM concentration of Mg^{2+} ions. Error bars indicate the confidence intervals for the fractions of native contacts, based on 10 replicate simulations.

Table 1: RNA-LifeTime prediction performance on RNA degradation rate. Test errors are summarized as the mean (standard deviation) across 3 macro-replications.

	# Parameters	MAE-LT	MAE	MSE	Training Time
AFT-Linear	295,627,971	262.7 (10.2)	0.3744 (0.013)	0.300 (0.012)	$\approx$ 1 day
RNA-LifeTime ($C_g = 1$)	42,246	67.84 (2.8)	0.1531 (0.003)	0.106 (0.003)	$\approx$ 2 Hours
RNA-LifeTime ($C_g = 3$)	42,306	70.24 (2.2)	0.1461 (0.002)	0.093 (0.002)	$\approx$ 3 Hours
RNA-LifeTime ($C_g = 5$)	42,366	67.44 (2.1)	0.1461 (0.002)	0.090 (0.002)	$\approx$ 4 Hours
MD-Average	-	52.5	0.1238	0.067	$\approx$ 1 weeks (using 6-8 GPUs)

AFT-Linear in terms of MAE-LT, MAE, and MSE, while having much fewer parameters and requiring a shorter training time. This suggests that RNA-LifeTime is more efficient and accurate for predicting RNA degradation rate. Among the RNA-LifeTime models tested, the one with $C_g = 5$ achieves the best performance. Furthermore, we compute the MAE-LT, MAE, and MSE for the mean trajectories of the ground truth MD simulation (MD-average) to represent the intrinsic randomness in the system (refer to the last row of Table 1). These results approximate the lowest achievable error. In comparison, the errors of RNA-LifeTime are close to those of the MD-average, suggesting the effectivness of the RNA-LifeTime.

5 CONCLUSION

The proposed hybrid model for RNA structure and functional dynamics builds on the scientific understanding of biomolecular interactions and it allows us to efficiently predict molecular conformational changes while providing insights into the energetics and dynamics of an enzymatic reaction network. Our RNA-LifeTime approach is capable of predicting RNA lifetime with high accuracy in short time, unlike traditional methods that may require a significantly longer time of molecular dynamics simulations. This modeling strategy can be extended to enzymatic molecular reaction networks and facilitate the development of multi-scale bioprocess digital twin. Further coupling with process analytical technologies and optimal design of experiments, we can accelerate new drug discovery and manufacturing process development.

ACKNOWLEDGMENTS

Dr. Wei Xei was supported by the National Institute of Standards and Technology (Grant no. 70NANB17H002 and 70NANB21H086). Dr. Paul Whitford was supported by NSF grant MCB-1915843 and his Work at the Center for Theoretical Biological Physics was also supported by the NSF (Grant PHY-2019745).

REFERENCES

Best, R. B., G. Hummer, and W. A. Eaton. 2013. "Native Contacts Determine Protein Folding Mechanisms in Atomistic Simulations". *Proceedings of the National Academy of Sciences* 110(44):17874–17879.

Bryngelson, J. D., and P. G. Wolynes. 1989. "Intermediates and Barrier Crossing in a Random Energy Model (with Applications to Protein Folding)". *The Journal of Physical Chemistry* 93(19):6902–6915.

de Oliveira Jr, A. B., V. G. Contessoto, A. Hassan, S. Byju, A. Wang, Y. Wang, E. Dodero-Rojas, U. Mohanty, J. K. Noel, J. N. Onuchic, and P. C. Whitford. 2022. "SMOG 2 and OpenSMOG: Extending the Limits of Structure-Based Models". *Protein Science* 31(1):158–172.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko et al. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold". *Nature* 596(7873):583–589.

Lee, E. T., and J. Wang. 2003. *Statistical Methods for Survival Data Analysis*, Volume 476. John Wiley & Sons.

Morgan, S. R., and P. G. Higgs. 1996. "Evidence for Kinetic Effects in the Folding of Large RNA Molecules". *The Journal of Chemical Physics* 105(16):7152–7157.

Noel, J. K., M. Levi, M. Raghunathan, H. Lammert, R. L. Hayes, J. N. Onuchic, and P. C. Whitford. 2016, 03. "SMOG 2: A Versatile Software Package for Generating Structure-Based Models". *PLOS Computational Biology* 12(3):1–14.

Noel, J. K., P. C. Whitford, and J. N. Onuchic. 2012. "The Shadow Map: a General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function". *The Journal of Physical Chemistry B* 116(29):8692–8702.

Orr, G. B., and K.-R. Müller. 1998. *Neural Networks: Tricks of the Trade*. Springer.

Šponer, J., G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka et al. 2018. "RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview". *Chemical Reviews* 118(8):4177–4338.

Wang, A., M. Levi, U. Mohanty, and P. C. Whitford. 2022. "Diffuse Ions Coordinate Dynamics in a Ribonucleoprotein Assembly". *Journal of the American Chemical Society* 144(21):9510–9522.

Wang, J., B. Williams, V. R. Chirasani, A. Krokhotin, R. Das, and N. V. Dokholyan. 2019. "Limits in Accuracy and a Strategy of RNA Structure Prediction Using Experimental Information". *Nucleic Acids Research* 47(11):5563–5572.

Whitford, P. C., J. N. Onuchic, and K. Y. Sanbonmatsu. 2010. "Connecting Energy Landscapes with Experimental Rates for Aminoacyl-tRNA Accommodation in the Ribosome". *Journal of the American Chemical Society* 132(38):13170–13171.

wwPDB consortium. 2018. "Protein Data Bank: the Single Global Archive for 3D Macromolecular Structure Data". *Nucleic Acids Research* 47(D1):D520–D528.

Xie, W., and G. Pedrielli. 2022. "From Discovery to Production: Challenges and Novel Methodologies for Next Generation Biomanufacturing". In *Proceedings of the 2022 Winter Simulation Conference 2022*, edited by B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. Corlu, L. Lee, and P. Lendermann, 238–252. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

HUA ZHENG is Ph.D. candidate of Mechanical and Industrial Engineering (MIE) at Northeastern University (NU) with expertise in machine learning and computer simulation. Email: zheng.hua1@northeastern.edu.

WEI XIE is Assistant Professor of MIE at NU with expertise in computer simulation and ML/AI. Email: w.xie@northeastern.edu.

PAUL WHITFORD is Associate Professor of Physics at NU with expertise in biomolecular simulations. Email: p.whitford@neu.edu.

AILUN WANG is Postdoctoral Research Associate of the Center for Theoretical Biological Physics at NU with expertise in biomolecular simulations. Email: a.wang@northeastern.edu.

CHUNSHENG FANG is Ph.D. candidate of MIE at NU with expertise in machine learning. Email: fang.chun@northeastern.edu.

WANDI XU is Ph.D. candidate of MIE at NU with expertise in operations research. Email: xu.wand@northeastern.edu.