Supporting Information : Multiple probes are required to explore and control the rugged energy landscape of RNA hairpins

Changbong Hyeon[1,2] and D. Thirumalai[3]

[1]*Center for Theoretical Biological Physics,*

University of California at San Diego, La Jolla, California 92093

[2]*Department of Chemistry, Chung-Ang University, Seoul 156-756, Republic of Korea*

[2]*Biophysics Program, Institute For Physical Science and Technology,*

University of Maryland, College Park, Maryland 20742

Definition of dihedral angle in TIS model: While six torsion angles are, in principle, required to specify the backbone conformation from one phosphate atom (P_n) relative to the next one (P_{n+1}) along the backbone, several methods for reduced representations for RNA have been proposed [1–5]. The TIS model [4, 5] belongs to such class that represents the nucleotide backbone by using a virtual bond. The high dimensional representation of entire nucleotide is substantially simplified so that exhaustive sampling of all conformations is possible. The dihedral angle for the TIS model is defined by the angle formed between four successive interaction sites along the sugar-phosphate backbone ($S_{i-1}P_iS_iP_{i+1}$ or $P_iS_iP_{i+1}S_{i+1}$), which is similar to the psuedo-torsion angles η ($C4'_{n-1}, P_n, C4'_n, P_{n+1}$) and θ ($P_n, C4'_n, P_{n+1}, C4'_{n+1}$) used by others [2, 3] (Note that P_i, S_i, and B_i in the TIS model are the positions of the centers of the phosphate, ribose, and base groups, respectively. In contrast, P_n and $C4'_n$ denote specific atom in the n^{th} nucleotide). To be specific, the i^{th} dihedral angle ϕ_i, which is the angle formed between the two planes formed from i^{th} to $(i+3)^{th}$ bead, is defined as $\cos\phi_i = (\vec{r}_{i+1,i} \times \vec{r}_{i+1,i+2}) \cdot (\vec{r}_{i+2,i+1} \times \vec{r}_{i+2,i+3})$ where $\vec{r}_{m,n} \equiv \vec{r}_m - \vec{r}_n$. To simulate as well as design the RNA molecule, we bias the ϕ_i to ϕ_i^o, where ϕ_i^o is the dihedral angle in the native state, using the dihedral angle potential with three minima, namely,

$$V_{DIH}(\phi_i) = [A_{1i}^{\eta} + B_{1i}^{\eta} + C_{1i}^{\eta} + A_{2i}^{\eta} cos(\phi_i - \phi_i^o + \phi_i^{\eta}) + B_{2i}^{\eta} cos3(\phi_i - \phi_i^o + \phi_i^{\eta}) + C_{2i}^{\eta} sin(\phi_i - \phi_i^o + \phi_i^{\eta})] \quad (1)$$

where the parameters (in $kcal/mol$) involving dihedral angle potential are determined depending on the value of ϕ_i^o as follows:

$A_{1i} = -A_{2i} = 1.0$, $B_{1i} = B_{2i} = 1.6$, $C_{1i} = 2.0$, $C_{2i} = -2.0$, $\phi_i^{g^+} = \pi/3$ for gauche(+) ($\eta = g^+$, $0 < \phi_i^o < 2\pi/3$),

$A_{1i} = -A_{2i} = 1.0$, $B_{1i} = B_{2i} = 1.6$, $C_{1i} = 2.0$, $C_{2i} = 2.0$, $\phi_i^{g^-} = 5\pi/3$ for gauche(-) ($\eta = g^-$, $4\pi/3 < \phi_i^o < 2\pi$),

$A_{1i} = A_{2i} = 1.2$, $B_{1i} = B_{2i} = 1.2$, $C_{1i} = C_{2i} = 0.0$, $\phi_i^t = \pi$ for trans ($\eta = t$, $2\pi/3 < \phi_i^o < 4\pi/3$).

Computation of free energy surfaces: Multidimensional free energy surface is computed by adapting the multiple histogram technique [6, 7]. For example, the two-dimensional free energy surface can be obtained at arbitrary external conditions specified by a temperature and a force. If the conformational states are well sampled over a range of temperatures and forces, the probability of finding the RNA conformation with order parameters (R,Φ) at a condition (T,f) is

$$P(R,\Phi|T,f) = \frac{\sum_E e^{-(E-fR)/T} \frac{\sum_{k=1}^K h_k(E,R,\Phi)}{\sum_{k=1}^K n_k e^{(F_k-(E-f_kR))/T_k}}}{\sum_E \sum_R \sum_\Phi e^{-(E-fR)/T} \frac{\sum_{k=1}^K h_k(E,R,\Phi)}{\sum_{k=1}^K n_k e^{(F_k-(E-f_kR))/T_k}}} \tag{2}$$

where E is the configurational energy, K is the number of histograms, $h_k(E,R,\Phi)$ is the number of states in (E,$E+\delta E$), (R,$R+\delta R$), (Φ,$\Phi+\delta\Phi$) of the k-th histogram, $n_k = \sum_{E,R,\Phi} h_k(E,R,\Phi)$, T_k and f_k are the pair of temperature and force in the simulations used to generate the k^{th} histogram, respectively. The free energy F_k (or partition function Z_k [$= \sum_R \sum_\Phi P(R,\Phi|T_k,f_k)$]), that is calculated self-consistently, satisfies

$$\begin{aligned} Z_r &= e^{-F_r/T_r} \\ &= \sum_E \sum_R \sum_\Phi e^{-(E-f_rR)/T_r} \frac{\sum_{k=1}^K h_k(E,R,\Phi)}{\sum_{k=1}^K n_k e^{(F_k-(E-f_kR))/T_k}} \\ &= \sum_E \sum_R \sum_\Phi e^{-(E-f_rR)/T_r} \frac{\sum_{k=1}^K h_k(E,R,\Phi)}{\sum_{k=1}^K [n_k e^{(E-f_kR)/T_k}]/Z_k}. \end{aligned} \tag{3}$$

With a choice of the vector $(Z_1,\ldots,Z_r,\ldots,Z_K) = (1,\ldots,1,\ldots,1)$ as an initial input, Eq.3 quickly converges to a stable solution. Using the low friction Langevin dynamics, we sampled the conformational states at the (T,f) in the range $\{0\ K < T < 500\ K, f = 0.0\ pN\}$ and $\{0.0\ pN < f < 20.0\ pN, T = 305\ K\}$. Exhaustive samplings around the transition regions at $\{305\ K \le T \le 356\ K,\ f = 0.0\ pN\}$ and $\{5.0\ pN \le f \le 7.0\ pN,\ T = 305\ K\}$ is required to

obtain reliable estimates of the thermodynamic quantities. The two-dimensional free energy surface, with R and Φ as order parameters, is calculated using

$$\begin{aligned} F(R,\Phi|T,f) &= F(T,f) - k_B T \log P(R,\Phi|T,f) \\ &= -k_B T \log \sum_E e^{-(E-fR)/T} \frac{\sum_{k=1}^{K} h_k(E,R,\Phi)}{\sum_{k=1}^{K} n_k e^{(F_k-(E-f_kR))/T_k}}, \end{aligned} \tag{4}$$

where $F(T,f) = -k_B T \log Z(T,f)$ with $Z(T,f) = \sum_{E,R,\Phi} e^{-(E-fR)/T} \frac{\sum_{k=1}^{K} h_k(E,R,\Phi)}{\sum_{k=1}^{K} n_k e^{(F_k-(E-f_kR))/T_k}}$.

Simulations: We assume that the dynamics of the RNA molecules can be described by the Langevin equation. The system of Langevin equations is integrated as described before [4, 8, 9].

Sampling conformations : Low friction Langevin dynamics, including inertia term, is performed to efficiently sample the conformational states of RNA, whose results are used to calculate the equilibrium properties of RNA (e.g., phase diagram and free energy surface).

$$m\frac{d^2\vec{r}_i}{dt^2} = -\zeta_L \frac{d\vec{r}_i}{dt} - \vec{\nabla}_i V(\{\vec{r}\}) + \vec{\Gamma}_i(t) \tag{5}$$

where $\vec{r}_i$ is the position vector of coarse-grained center, and $\vec{\Gamma}_i(t)$ is the random force satisfying the fluctuation-dissipation theorem, $\langle \vec{\Gamma}_i(t) \cdot \vec{\Gamma}_j(t') \rangle = 6k_B T \zeta_L \delta(t-t')\delta_{ij}$. In the TIS representation (base B, sugar ring S or phosphate P), the mass of a bead $m = 100 - 160$ g/mol ($= 1.7 \times 10^{-22} - 2.7 \times 10^{-22}$ g), the average distance between the adjacent beads $a = 4.6$ Å, and the energy scale $\epsilon_h = 1 \sim 2$ $kcal/mol$ ($= 7.0 \times 10^{-21} - 1.4 \times 10^{-20}$ J). The natural measure for time of Eq.5 is $\tau_L = (\frac{ma^2}{\epsilon_h})^{1/2} = 1.6 \sim 2.8$ ps. One can perform stable simulations with $\zeta_L = 0.05m/\tau_L$ and a simulation time step $\delta t = 0.0025\tau_L$. The equation of motion is solved using Verlet algorithm. Note, however, that the purpose of the low friction dynamics simulation, in which the inertia term dominate over the friction term, is only for the sampling purpose in the context of coarse-grained simulation of nano-scale system. Therefore, the dynamic trajectory in low friction environment is far from realistic, nor the natural time τ_L is a meaningful quantity without taking the explicit solvent environment into account.

Kinetic simulations: To perform kinetic simulation with a realistic friction coefficient, the over-damped simulations are performed using a Brownian dynamics algorithm

$$\zeta_H \frac{d\vec{r}_i}{dt} = -\vec{\nabla}_i V(\{\vec{r}\}) + \vec{\Gamma}_i(t). \tag{6}$$

From Eq.6 and $\tau_L^2 = ma^2/\epsilon_h$, the natural measure for time for over-damped condition at simulation temperature T_s is

$$\tau_H \approx \frac{\zeta_H a^2}{k_B T_s} = \frac{(\zeta_H \tau_L/m)\epsilon_h}{k_B T_s}\tau_L. \tag{7}$$

For the purpose of converting simulation time into real time, we choose the parameters as $\epsilon_h = 1$ kcal/mol, $m = 1.8 \times 10^{-22}$ g, $a = 4$ Å, which lead to $\tau_L = 2ps$, $\zeta_H = 6\pi\eta a \left(= 6\pi \times 0.01 g \cdot cm^{-1}s^{-1} \times 4\mathring{A}\right) = 87m/\tau_L$. We set $\zeta_H \approx 50m/\tau_L$ for convenience. The equations of motion using Eq.6 are stably integrated with a time step of $\delta t_L = 0.02\tau_L$. A single simulation time step is converted to the real time as $\delta t_L \rightarrow \delta t_H = 0.02\tau_H$, hence, for example, the 10^6 simulation time steps ($10^6 \times \delta t_H$) at $T_s = 290$ K corresponds to 3.5 μsec in real time. The folding kinetics simulations under force-quench condition are performed over the ensemble of equilibrated structures at $f = 22pN$. We switch the force $f = 22pN \rightarrow 0pN$ and measure the folding times for 100 different trajectories. The temperature-quench kinetics is similarly implemented from the ensemble of structure equilibrated at $T = 400$ K to the low temperature condition at $T = 290$ K. The time scale for f and T-quench (τ_{quench}) is instantaneous ($\sim$ 3.5 nsec) in the present work. Although this quench rate is hard to achieve in the real experiments, our prediction will not change as long as the time scale of folding (τ_{fold}) and time scale of quenching the external condition (τ_{quench}) are well separated, namely $\tau_{quench} \ll \tau_{fold}$.

[1] Murray, J. J. W., Arendall III, B., Richardson, D. C., & Richardson, J. S. (2003) *Proc. Natl. Acad. Sci.* **100**, 13904–13909.

[2] Wadley, L. M., Keating, K. S., Duarte, C. M., & Pyle, A. M. (2007) *J. Mol. Biol.* **372**, 942–957.

[3] Cao, S. & Chen, S. J. (2005) *RNA* **11**, 1884–1897.

[4] Hyeon, C. & Thirumalai, D. (2005) *Proc. Natl. Acad. Sci.* **102**, 6789–6794.

[5] Hyeon, C. & Thirumalai, D. (2006) *Biophys. J.* **90**, 3410–3427.

[6] Ferrenberg, A. M. & Swendsen, R. H. (1988) *Phys. Rev. Lett.* **61**(23), 2635–2638.

[7] Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., & Rosenberg, J. M. (1992) *J. Comp.*

Chem. **13**(8), 1011–1021.

[8] Veitshans, T., Klimov, D.K., & Thirumalai, D. (1997) *Folding Des.* **2**, 1–22.

[9] Klimov, D. K., Betancourt, M. R., & Thirumalai, D. (1998) *Folding Des.* **3**, 481–498.

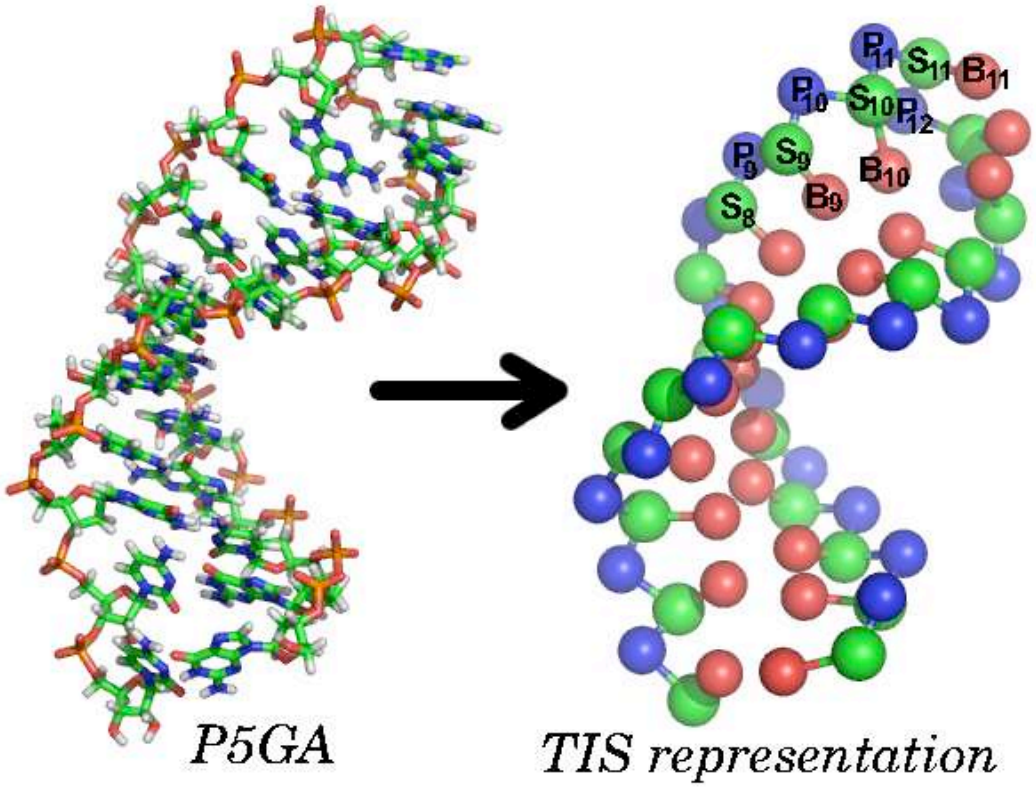

FIG. 3: Three Interaction Site (TIS) representation of P5GA hairpin.

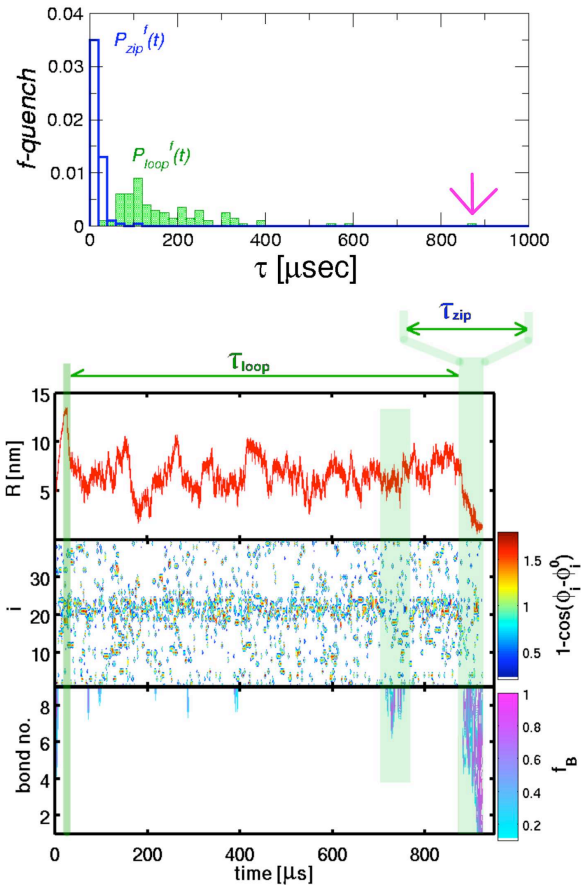

FIG. 4: A detailed analysis on the force-quench dynamics of RNA hairpin formation. Force-quench refolding of RNA hairpins undergoes two step dynamics; (i) loop formation and (ii) zipping. The time scales of these distinct processes are designated by τ_{loop} and τ_{zip}, respectively. For most of the RNA hairpins that are released from the rod-like state, the loop formation can be a hard-to-achieve and heterogeneous process because of the vast number of combinatorics formed by the loop dihedral angle degrees of freedom. The formation of correct dihedral angles in the loop region is correlated with the base pair formation that locks the loop. The distributions of τ_{loop} and τ_{zip} are computed for a hundred of force-quench refolding trajectories (top panel). For a very slow refolding trajectory (the event indicated by an arrow on the top panel, $\tau_{loop} \gtrsim 800\ \mu sec$), the time evolutions of end-to-end distance (R), deviations from the native dihedral angles ($1-\cos(\phi_i-\phi_i^o)$), and the degrees of base pair formation in the stem region (f_B) are shown on the bottom panel. Occasional unsuccessful attempts for the loop formation (see green shadow around $\tau \sim 750\ \mu sec$ for example) are observed before the successful loop formation and subsequent zipping event ($\tau \gtrsim 900\ \mu sec$).

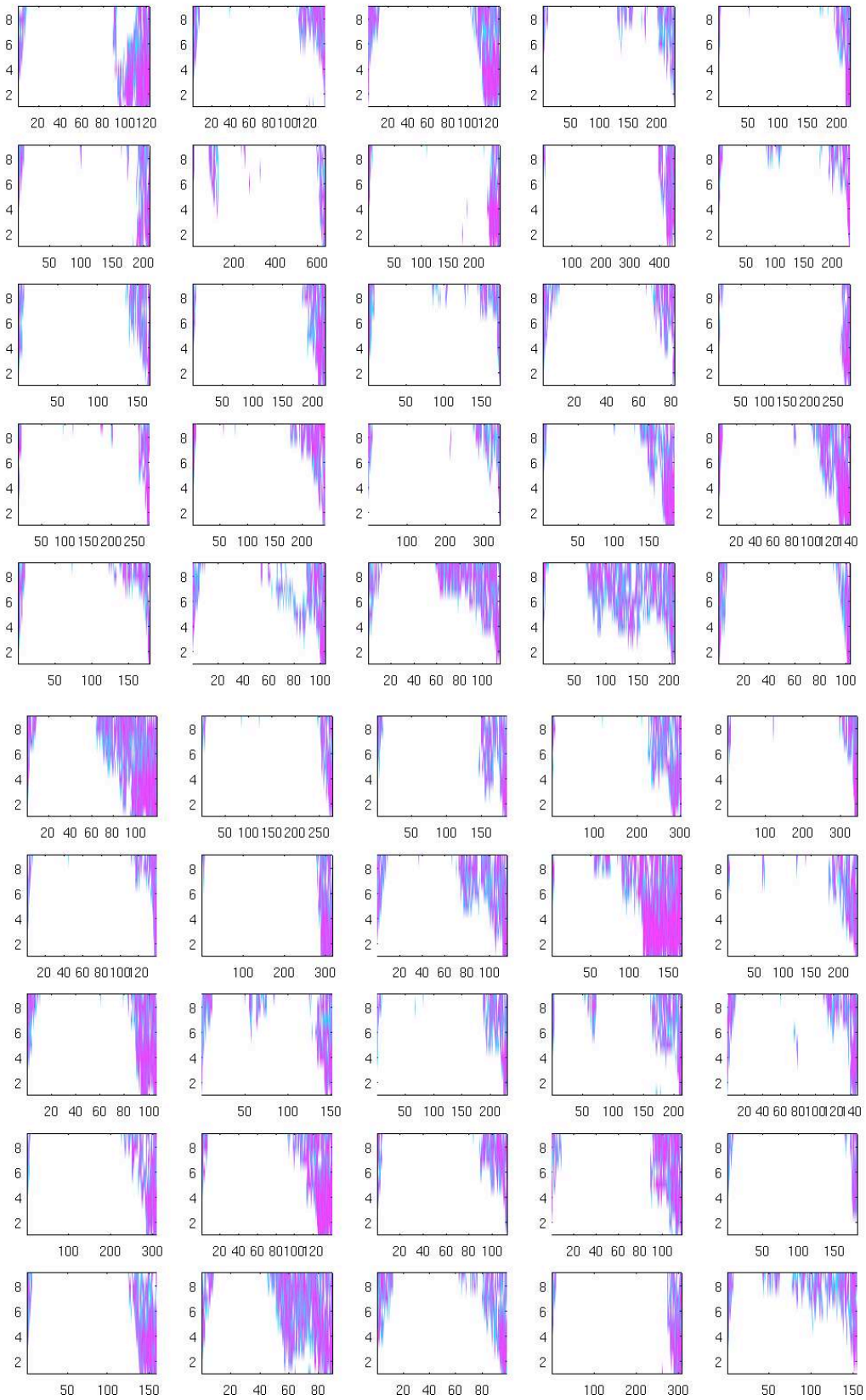

FIG. 5A: Time traces of f-quench dynamics probed by native base pair index. Time is in μsec.

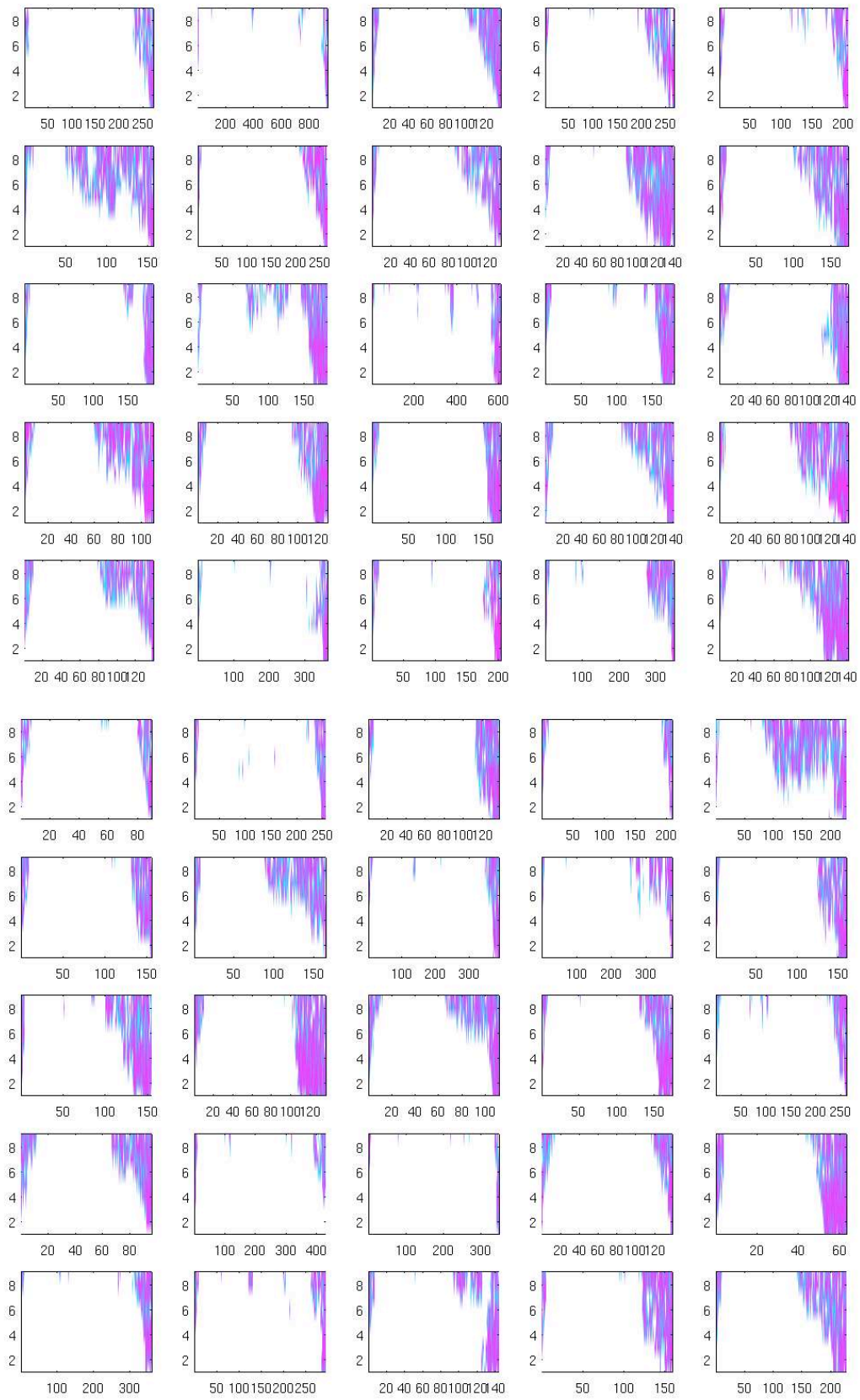

FIG. 5B: Time traces of f-quench dynamics probed by native base pair index (continued)

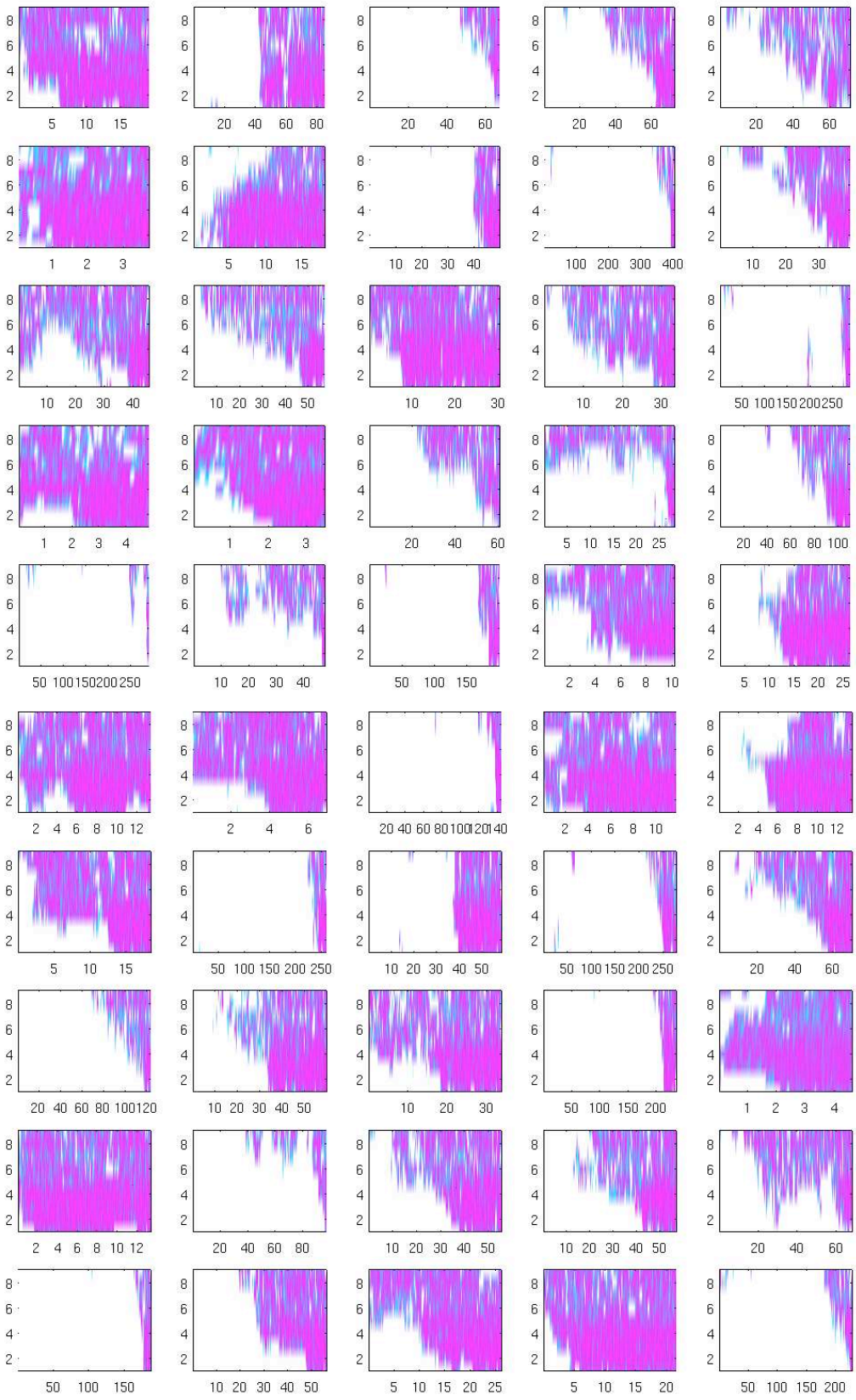

FIG. 6A: Time traces of T-quench dynamics probed by native base pair index. Time is in μsec.

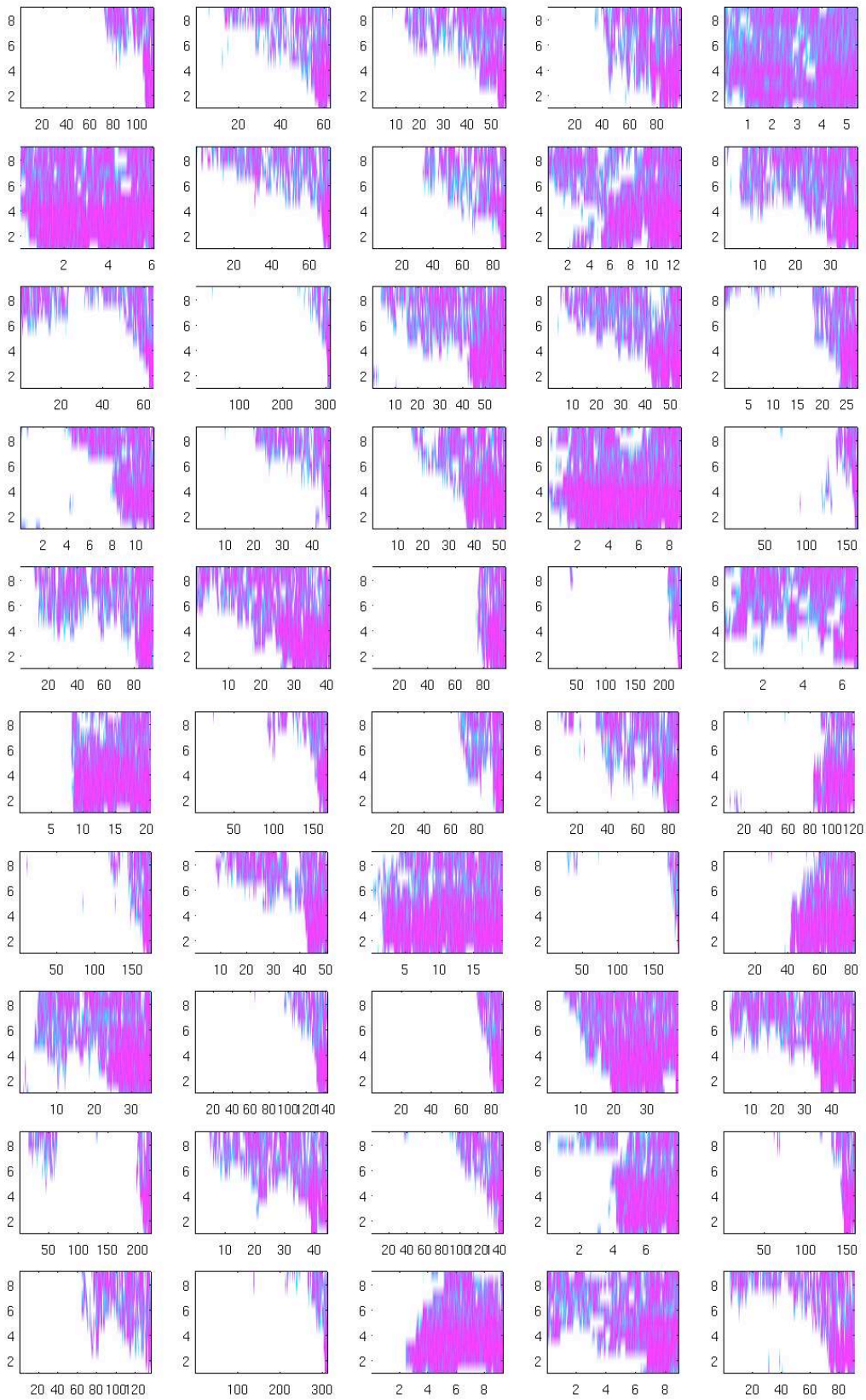

FIG. 6B: Time traces of T-quench dynamics probed by native base pair index (continued)