Supporting Information

Revealing the distinct folding phases of an RNA three-helix junction

Alex Plumridge[1], Andrea M. Katz[1], George D. Calvey[1], Ron Elber[2], Serdal Kirmizialtin[3] and Lois Pollack[1,*]

[1]School of Applied and Engineering Physics, Cornell University, Ithaca, NY, USA

[2] Department of Chemistry and Institute for Computational Engineering and Sciences (ICES) University of Texas at Austin, Austin, TX, USA

[3] Chemistry Program, Science Division, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

* To whom correspondence should be addressed. Tel: 607 255-8695; Fax: 607 255-7658; Email: lp26@cornell.edu

The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors.

Sample Preparation

tP5abc synthesis 3

Static SAXS sample preparation 3

Time-resolved SAXS sample preparation 3

Static SAXS Experiments

Experimental setup 4

Sample cell 4

Data acquisition 4

Basic SAXS analysis 5

Time-resolved (TR) SAXS Experiments

Experimental setup..... 6

Mixer fabrication..... 7

Mixer principles 10

Data collection 12

Basic SAXS analysis 12

Simulations

Structures of extended and collapsed states 13

Reaction coordinate of the extended to collapsed transition 15

Ensemble Fitting

SAXS profile computation 17

Ensemble optimization 17

Static ensemble analysis 17

Time-resolved ensemble analysis 31

1D projections of static and time-resolved ensembles 44

Role of Mg ions in key intermediates..... 45

Critical assessment of the ensemble approach..... 46

Supplementary References 47

Sample Preparation

tP5abc synthesis

tP5abc was transcribed *in vitro* with T7 polymerase from PCR amplified DNA. The template strand (containing the RNA and T7 promoter sequence) and primers for the PCR were purchased from Integrated DNA Technologies (Coralville, IA, USA). The T7 reactions were carried out using T7 RiboMAX express large scale kits, purchased from Promega (Madison, WI, USA). The RNA product was purified by size exclusion chromatography using a Superdex HiLoad 16/600 200pg column on an ÄKTA FPLC purifier. Purified RNA was buffer exchanged against the appropriate salt concentration, before being stored overnight (see below).

Static SAXS sample preparation

RNA samples used in the static experiments were synthesized using the protocol described above. All samples were initially prepared and buffer exchanged into 10mM KMOPS + 20mM KCl + 20µM EDTA, pH7. The low [K] was chosen to select for extended RNA states. RNA used for the [K] salt series was buffer exchanged 5 times against a KCl containing solution of the final salt concentration + 10mM KMOPS + 20µM EDTA, pH 7 in Amicon Ultra-0.5 10KdA concentrators (EMD Millipore, Billerica, MA, USA) hours before experiments. RNA used for the [Mg] salt series was prepared first by buffer exchange (5 times) into the initial, low salt buffer. Immediately before placement in the beam, $MgCl_2$ was added to achieve the appropriate final salt concentration. All RNA was annealed at 65°C for 5 minutes, before slow cooling to room temperature.

Time-resolved SAXS sample preparation

RNA used in the time-resolved experiments was synthesized and buffer exchanged as described above. The initial buffer used for these experiments contained 10mM KMOPS + 20mM KCl + 20µM EDTA, pH7. For mixer experiments, the initial [RNA] was 80µM.

Static SAXS Experiments

Experimental setup

SAXS profiles were measured at the G1 station of the Cornell High Energy Synchrotron Source (CHESS). 11.325 keV x-rays were incident on a sample cell (described below). Images were acquired with a Pilatus 100K detector (Dectris). The direct beam was measured after passing through a semi-transparent beamstop (a 250 µm molybdenum foil) and used for normalization. Two sets of tungsten slits (ADC) were used to define the beam size and minimize parasitic scatter upstream of the sample cell. Samples were positioned in the sample cell using a peristaltic pump, and oscillated throughout the exposure with a syringe pump (Harvard Apparatus) to avoid radiation damage. The entire beam path from the beam defining slit to the beamstop was kept in vacuum to reduce parasitic scatter.

Sample cell

The static SAXS sample cell was comprised of a ~2" long section of 1.74 mm diameter polyimide tube (Microlumen) attached to stainless steel sample supply tubing on either end. A plastic support held the sample cell securely in place while providing ample room for the x-rays to pass through unimpeded. A picture of the static sample cell and support device are shown in Figure S1.

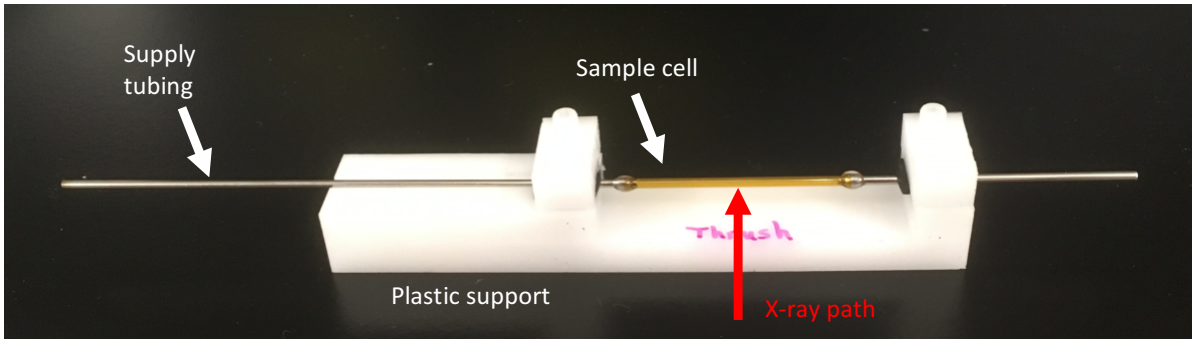

Fig S1. Static SAXS sample cell and support.

Data acquisition

At each salt concentration, measurements were performed at different [RNA] to control for interparticle interference effects (manifesting as changes in the low q SAXS profiles). RNA concentrations of 80, 60 or 40µM were used, and at least 2 RNA concentrations were sampled at each salt condition.

Basic SAXS analysis

All data were analyzed with MATLAB (MathWorks, Natick, Ma, USA) using in-house code. Buffer subtracted curves were matched in the range q>0.1 $Å^{-1}$ for accurate sample concentration normalization. Interparticle interaction, in this case repulsion, was only detected at the lowest monovalent salt condition, 20 mM KCl. For this data point, the SAXS profile at infinite dilution was created by linear extrapolation to the zero-concentration limit from profiles acquired at higher sample concentrations. For all other cases (with no interparticle interference), matched SAXS profiles from all sample concentrations were averaged to give the final SAXS curves.

Basic SAXS analysis of this data is shown in Figure S2 below. All Kratky plots have been scaled to match in the range q=0.01-0.06 $Å^{-1}$ and are unsmoothed to demonstrate the signal-to-noise achieved in the static experiments. The distance distribution functions (P(r)) were calculated using GNOM [1], and the curve areas normalized to unity.

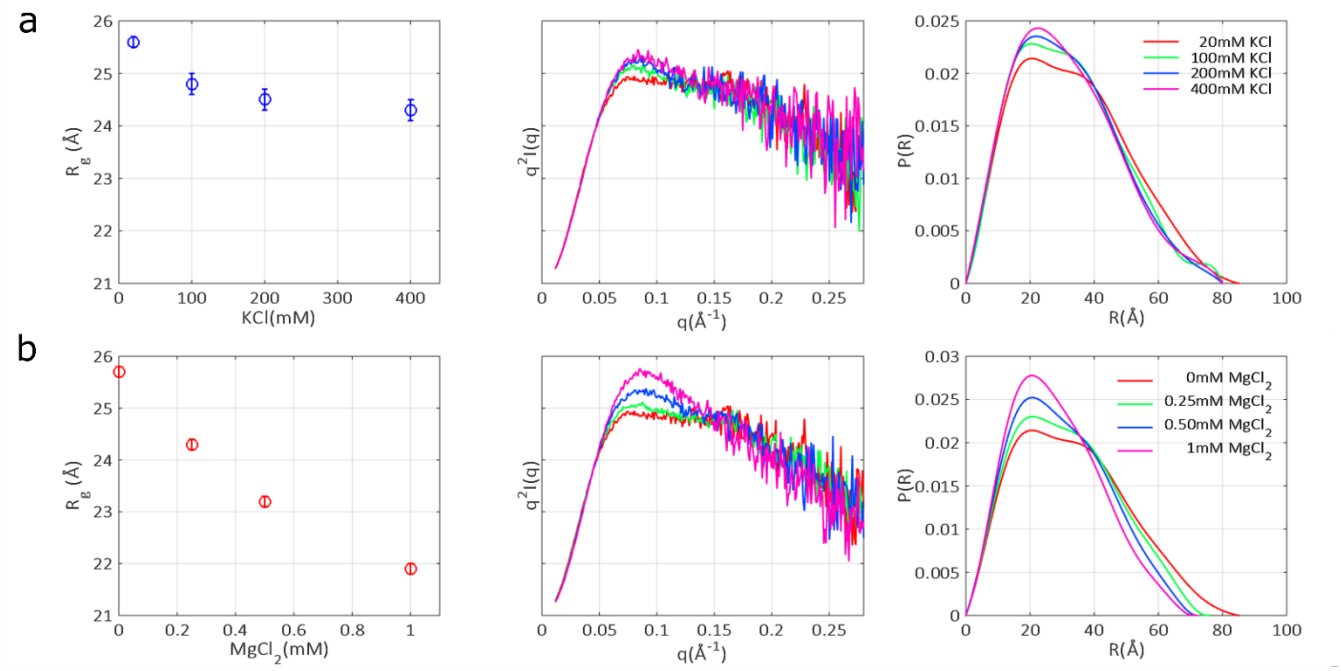

Fig S2. Basic SAXS analysis of the static salt titrations. The radii of gyration R_g (derived from a Guinier fit), Kratky plots and distance distribution functions (P(r)) derived from GNOM are shown for titrations in KCl **a)** and $MgCl_2$ **b)**. In some cases, the uncertainty on the R_g values are comparable to the marker size. Kratky plots are unsmoothed.

Time-resolved (TR) SAXS Experiments

Experimental setup

TR-SAXS data were acquired at the G1 Station of the Cornell High Energy Synchrotron Source (CHESS). An illustration of the experimental setup is shown in Figure S3. The mixer (described below) was housed in a 6" vacuum cube (Ideal Vacuum). 11.18 keV x-rays passed through a 50 µm SCATEX scatterless tantalum pinhole (Incoatec) about 25 mm upstream of the mixer. A 250 µm thick molybdenum foil was used as a semi-transparent beamstop, allowing images to be normalized by the transmitted photon counts. We collected scattering images with an Eiger 1M detector (Dectris). The entire x-ray flight path was under kept under vacuum to minimize parasitic scatter from windows. The flight tubes and sample chambers were connected by flexible vacuum bellows. Standard ¼-28 flat bottom fittings served as a vacuum feedthrough for the liquid supply lines. We used an OB1 pressure controller and MFS flow meters (Elveflow) to control both sample and buffer flows.

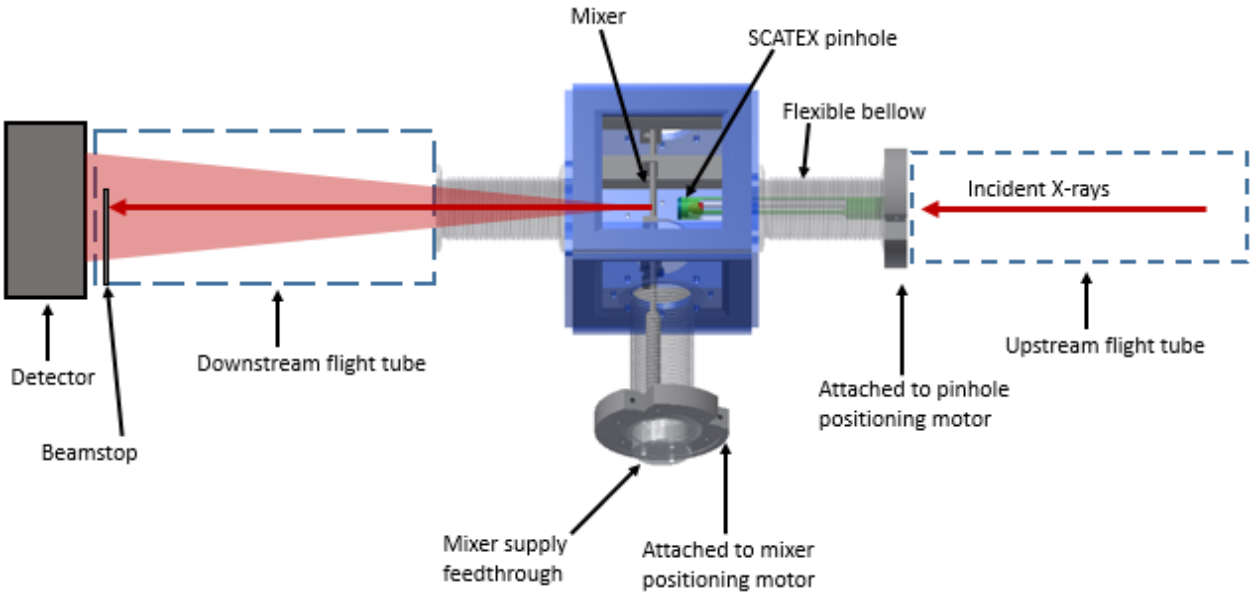

Fig S3. Illustration of the time-resolved SAXS setup used to establish the RNA folding pathway.

Mixer fabrication

Microfluidic capillary mixers based on the design by Calvey et al. were fabricated as shown in Figure S4. Most fabrication procedures are described in detail in [2]; however these devices differ in several respects from the ones described in that publication. Briefly, the fabrication followed four steps:

1) First, we scribed glass capillaries with standard polyimide coating (Polymicro Technologies and Postnova Analytics Inc.) to length with a diamond scribe, then flattened and beveled the tips on an Allied Multiprep polisher (Allied High Tech Products, Inc.), (Figure S4a).

2) Using a butane torch, we removed the polyimide coating from the end of the outer capillary. A femtosecond laser was used to cut the centering spacers from a two mil polyimide sheet, after which they were installed over the inner capillaries, (Figure S4b).

3) Next, we inserted the inner into the outer capillary, and used UV curable epoxy (UV15, Masterbond) to secure the downstream inner capillary and completely fill the gap between it and the outer capillary. We cut the bonded outer and inner capillaries to length with a diamond saw, then polished the ends flat using the Multiprep polisher, (Figure S4c).

4) To create the observation region, a piece of polyimide medical tubing with 1 mil wall thickness (MicroLumen) was fit over the flattened capillary assembly and sealed in place with Double Bubble Extra Fast Setting epoxy (Hardman Adhesives), as shown in Figure S4d. We secured the mixers to a metal frame, and connected the supply lines with standard 10-32 coned fittings (IDEX HS). Figure S5 shows a schematic of fluidic connections and a picture of completed mixers mounted on their frames.

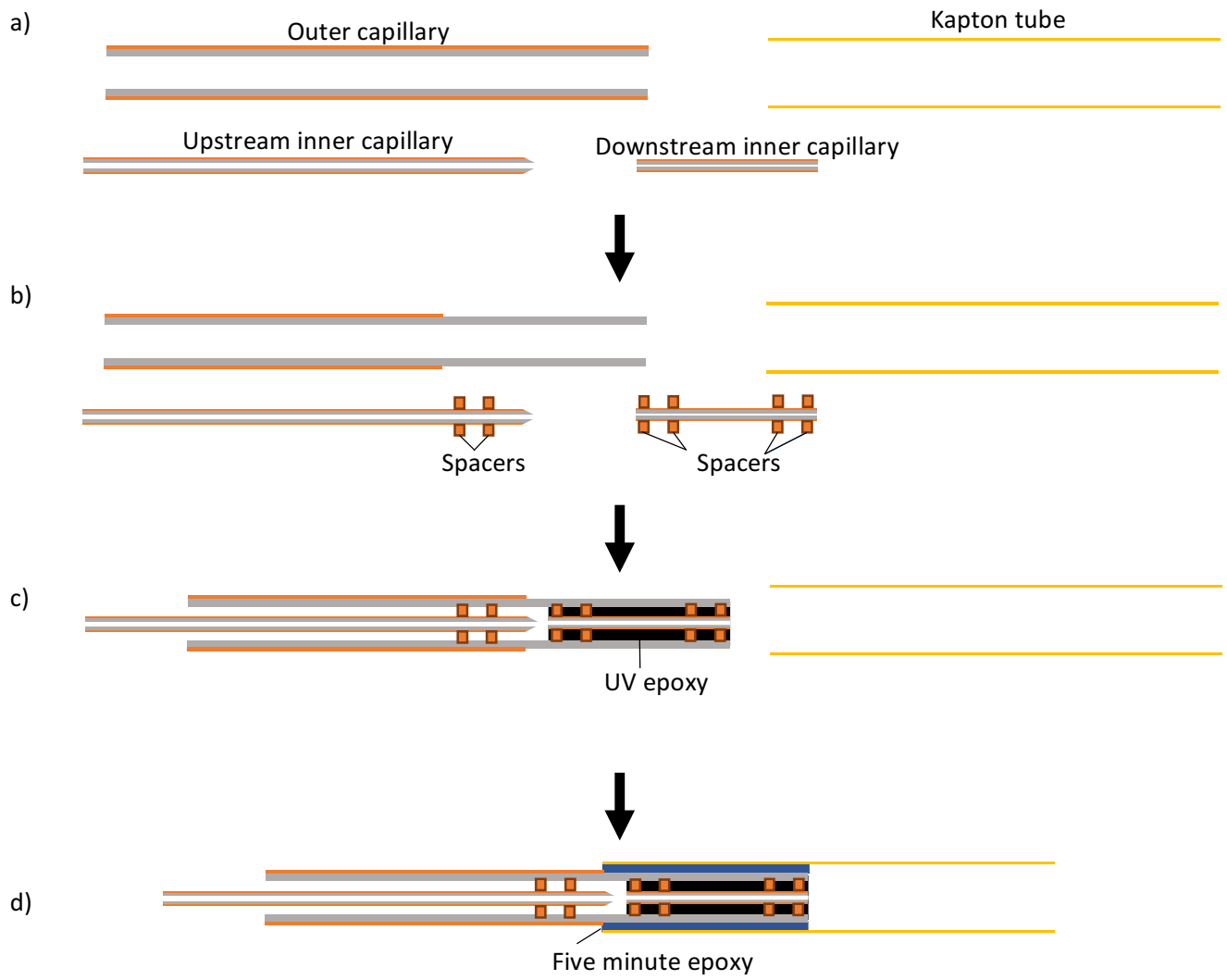

Fig S4. Schematic of key steps in the mixer fabrication process.

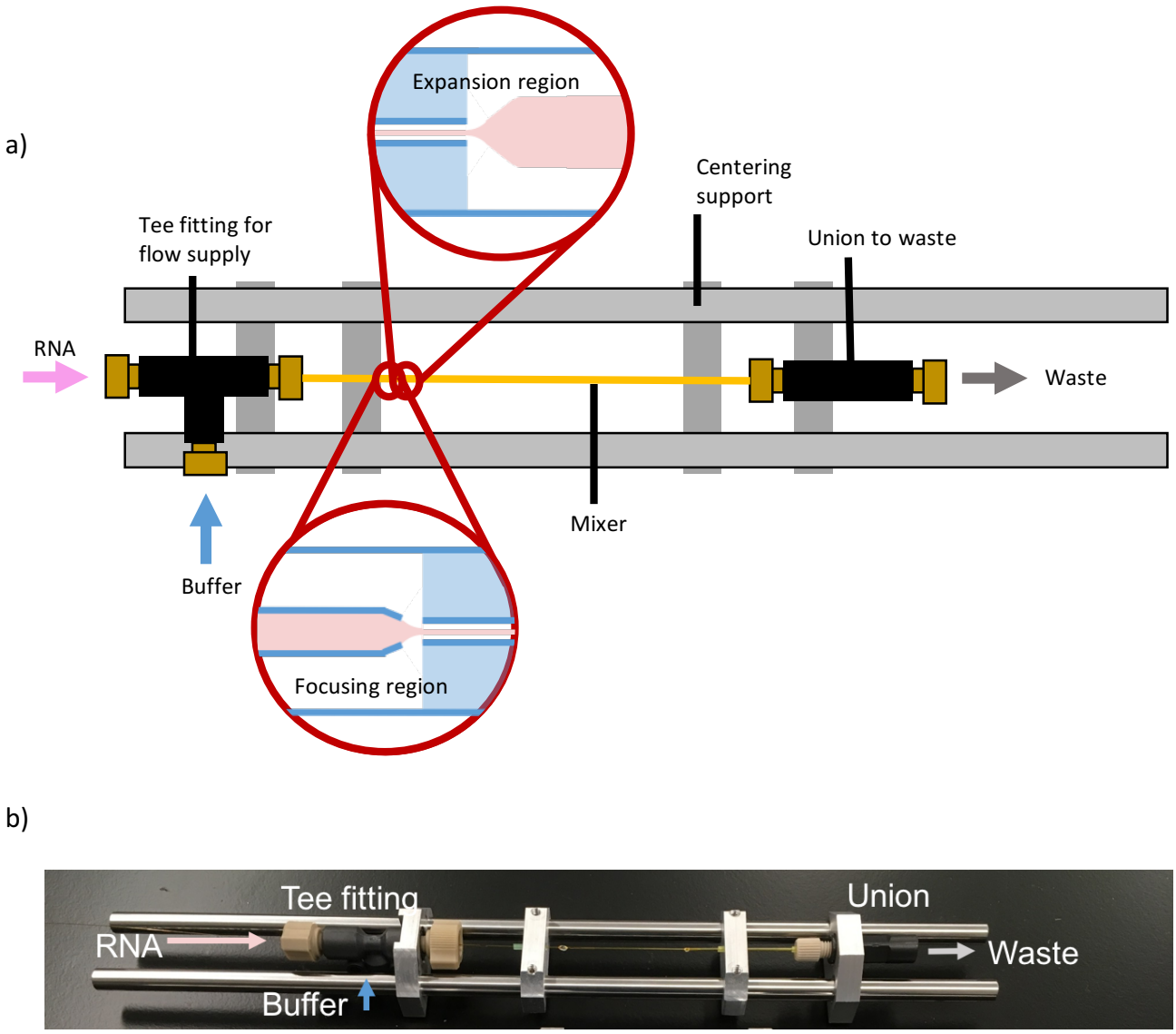

Fig S5. (a) Schematic of the fluidic attachments. **(b)** Completed mounted mixer, with fluidic fittings. Each frame rail is 8 inches long.

Mixer principles

The completed mixer can be described by considering three functional sections, labelled in Figure 2 of the main text. In the *focusing region*, RNA solution flowing out of the inner capillary is focused down into a thin (6-12 μm wide) jet by a coaxial flow of buffer containing magnesium. A constriction forces both the RNA solution and its coaxial sheath into a small diameter capillary, allowing the flow profile to develop quickly and minimizing pre-mixing as described in [2].

The constriction is the *mixing region* of the device. Here, magnesium diffuses rapidly across the thin RNA sample jet. The RNA, due to its larger size, does not diffuse appreciably out of the central jet. For each time point, the flowrates were chosen to ensure the sample traverses the length of the constriction in a time short compared to the time point of interest, as shown in Table S1.

The *delay and probe region* is comprised of the thin walled polyimide tubing. The sample jet slows and expands as it moves from the constriction to the delay and probe region, creating a larger x-ray path length. This larger sample volume improves the signal to noise ratio, and reduces the rate of diffusion so that the average magnesium concentration along the RNA jet remains roughly constant. The latter greatly simplifies interpretation and analysis of the data. The delay time is defined to be the time elapsed as the sample flows from the outlet of the constriction to its intersection with the x-ray beam. Different delay times were accessed by either moving the mixer or changing the flow rates.

Three different flow schemes were used to access the large range of time points required for this experiment (ms to seconds). To ensure consistency between conditions, the [Mg] in the outer buffer was adjusted to maintain an average concentration of 1 mM at the end of the constriction (thus throughout the observation region). Table S1 shows the Mg^{2+} concentration of the outer flow, buffer and sample flowrates, and timing dispersion for each time point. This dispersion arises from both the travel time of the sample through the finite size of the beam ("beam smearing") as well as from the parabolic flow profile within the channel ("flow dispersion"). These are added in quadrature to yield the total dispersion.

	Timepoint (ms)					
	10	30	100	300	1000	3000
Sample Flow Rate (μL/min)	5	5	5	1.7	1.7	1.7
Total flow rate (μL/min)	36.4	17	17	5.7	5.7	5.7
Time in constriction (ms)	3.5	7.5	7.5	22.5	22.5	22.5
Flow Dispersion (ms)	1.1	7.6	19.8	61.5	183.3	531.3
Beam Smearing (ms)	2.5	5.4	5.4	16	16	16
Total dispersion (ms)	±1.4	±4.7	±10	±30	±90	±270
Sample Path Length (μm)	80	148	148	148	148	148
Outer flow [Mg^{2+}] (mM)	2	2	2	1.5	1.5	1.5

Table S1. Flow parameters and dispersions for time points probed.

Magnesium diffusion was simulated using Comsol version 4.2. The diffusion coefficient was set to (1 x 10^{-9})m^2/s, intermediate in the range of reported literature values [3,4]. Figure S6a shows the simulated $MgCl_2$ concentration averaged over the intersection volume of the x-ray beam and the sample jet for each time point probed. These values account for both the shape of the intersection volume and the unequal weighting of different points across the radial profile of the sample jet due to its cylindrical shape. The concentration in a radial slice through the sample jet is shown in Figure S6b to show concentration extremes experienced by some molecules in the jet.

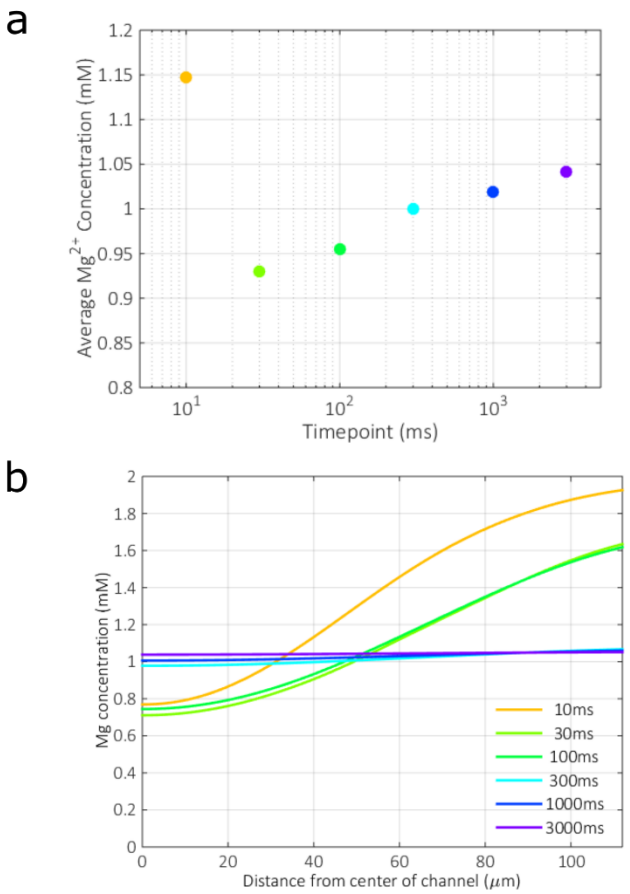

Fig S6. a) The average magnesium concentration observed at each time point in the experiment. **b)** The magnesium concentration found along a radial slice through the RNA jet at different time points.

Data collection

For each timepoint, we employed the following data acquisition procedure. We first positioned the mixer so the beam passed through the appropriate location in the probe region. With the sample flow off, we collected a buffer background signal. We then turned on the sample flow and acquired the sample signal. Finally, we turned off the sample flow and repeated the buffer background acquisition to ensure reliable background subtraction.

Basic SAXS analysis

As with the static data, all time-resolved data were analyzed with MATLAB (MathWorks, Natick, Ma, USA) using in-house code. SAXS derived parameters as calculated for the static data are shown in Figure S7 below. Note that slight differences in R_g compared to the manuscript reflect the different methods used to compute this parameter. Here, Guinier fitting is used to parallel the static data, whereas the main text uses GNOM derived values. The latter is preferable for TR experiments, as the whole curve is considered in the computation of R_g, rather than a limited q-range.

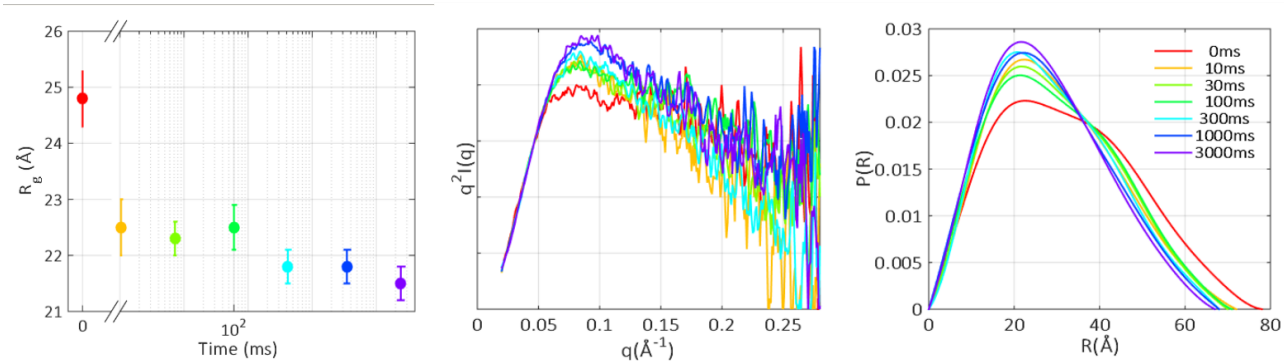

Fig S7. Basic SAXS analysis of the time-resolved SAXS data. The radii of gyration (derived from a Guinier fit), Kratky plots and distance distribution functions (P(r)) derived from GNOM are shown for each time point.

Simulations

Structures of the extended and collapsed states

We compute an ensemble of RNA conformations that represents the entire folding process of tP5abc. We guide the calculation by two end conformations. One conformation is the reactant (the unfolded structure) and the second conformation the product (folded conformation). The crystal structure 1HR2 [5] represents the product state after the removal of four base pairs in the P5b stem following [6]. The unfolded or extended structure of this RNA follows a model proposed by [6] that exploits NMR constraints.

From the two end configurations four structures are simulated: i) A collapsed (folded) state that was generated from the crystal structure with Mg^{2+} ions in the RNA core ii) A collapsed state with no Mg^{2+} ions, iii) An extended state with Mg^{2+} ions still at the RNA core and iv) extended state with no Mg^{2+} ions. The four configurations (instead of only two) allow us to examine the distributions of conformations as a function of tightly bound (or not) magnesium. The energies of the initial structures are minimized utilizing the GBSA implicit solvent model [7].

Subsequently, the RNA is solvated with explicit water molecules and ions in a triclinic cell that extends at least 10 Å from any RNA atom. The dimensions of the boxes are 63.8 x 56.4 x 71.4 $Å^3$ for the collapsed states and 93.9 x 64.0 x 73.3$Å^3$ for the extended states. To generate ensembles of equilibrium configurations in the neighborhood of each of the four structures we conducted MD simulations with the GROMACS 5.0.4 [8] program. The water model is SPC/E [9], and Mg^{2+}, K^+, and Cl^-, ions parameters taken from Refs [10,11]. The RNA force field parameters are from AMBER03 [12]. The equations of motion are integrated with a time step of 2 fs using the Leap Frog integrator [13] constraining the bond lengths at ideal values with the LINCS algorithm [14]. Particle Mesh Ewald (PME) summation [15] computes the long-range electrostatic interactions. The real space distance cutoff (for electrostatics and van der Waals energies) is set to 11Å. The grid for the Fourier space summation in PME is of 1.6Å, and fourth order splines interpolate the charge density on the grid. A dispersion correction is made for the van der Waals cutoff [16]. The bulk ion concentration of the aqueous solution is 55mM K^+ and 42mM Cl^- and it consists of 65 K^+ and 10 Cl^- ions. For a mixed aqueous solution of monovalent and magnesium ions, we have 31 K^+, 17 Mg^{2+} ions, and 10 Cl^- ions, equivalent to bulk concentrations of 32, 2.8 and 30 mM respectively. The extended states, on the other hand, use a bigger simulation box with the same number of ions that result in about 40% reduction of each ion concentration in comparison to their collapsed counterpart. Bulk concentrations are measured by spherically averaging the concentration far from the center of mass of RNA. A spherical shell at the radius R=26Å with dR=2Å is used to estimate the bulk concentrations.

Equilibrium simulations of the four states are conducted in three steps: i) Equilibration of volume size with constant pressure simulations, ii) Equilibration of the ion distributions, and iii) Equilibration of the RNA molecule. For the first stage, 10ns Isothermal Isobaric ensemble (NPT) simulations were conducted by fixing the pressure at 1 bar and temperature 300K. Temperature coupling was accomplished using velocity rescaling with a stochastic term [17] while for pressure

coupling we use the Parrinello-Rahman method [18] with a frequency of 1ps. The heavy atoms of RNA are restrained to their original positions using harmonic potentials with force constants of 1000kJ/nm^2. The volume of the last snapshot of the NPT simulation is used in the ion equilibration phase that follows. An 80ns constant volume and temperature simulation (NVT) is conducted to equilibrate the aqueous solution while keeping the RNA molecule restrained. The restraints are removed in the third phase of the simulations that samples RNA conformations in the neighborhood of the reactant and product states. Data are recorded for every 10ps for analysis and structural clustering. Clusters are constructed for the extended and collapsed sets based on Root Mean Square Deviation (RMSD) using gromos method [19].

The conformational ensembles of the folded and extended states show subtle differences in their structure and overall sizes. We monitor changes in the shape and size using the Radius of gyration (R_g) and the Root mean square distance (RMSD). The results are summarized in Table S2. The structural heterogeneity is measured by the average RMSD (avRMSD), and by the number of clusters in the conformational ensemble. Clusters with weights 10% or higher are reported. In the extended state, the RNA size is largest in the absence of Mg^{2+} ions with average R_g =25.2Å in comparison to R_g=21.5Å in Mg^{2+} solution. Three major clusters with avRMSD of 5.6Å suggest a marked structural heterogeneity in the extended state ensemble. Upon a collapse in monovalent salt solution, the average size of the RNA undergoes 25% reduction with lesser structural heterogeneity. Similar to findings in monovalent salt only, the conformational ensemble shows the existence of extended and collapsed states in the presence of Mg ions. The RNA in the presence of Mg^{2+} is more compact in both states than in the monovalent-only solution because of a more effective shielding of the electrostatic repulsion in Mg^{2+}/K^{+} ion mixture. The structural heterogeneity observed for monovalent ions is reduced dramatically for the ion mixture, making the collapsed state more stable.

In later phases, clusters define the reactant and product sets. The centers of most dominant clusters are used in reaction path calculations between folded and unfolded conformations. Here "unfolded" means more extended conformation with the majority of the secondary structure still intact. The dominant difference is in tertiary contacts, even though significant re-arrangement at the RNA core can also be highly significant in the presence of magnesium. The observation that the range of "unfolded" structures is quite small, especially in comparison to the related problem of protein folding, validates the use of the reaction path approach between two end structures.

	Extended-Mg^{2+}	Extended	Collapsed-Mg^{2+}	Collapsed
R_g (Å)	21.5	25.2	17.7	18.6
RMSD (Å)	4.6	5.3	2.3	3.4
$P_{cluster}$	0.36, 0.23, 0.12	0.15, 0.10	0.92	0.75, 0.16
Simulation Time	395ns	390ns	600ns	600ns

Table S2. Conformational dynamics and average dimensions of RNA in the extended and collapsed states in mono and divalent salt concentrations.

Reaction coordinate of the extended to collapsed transition

The folded and unfolded structures are the initial and final states for the Self Penalty Walk (SPW) algorithm to compute reaction coordinates. The SPW algorithm is available in the MOIL package [20]. In its present form, it requires distinguishable particles and therefore cannot use explicit solvation. Instead, the Generalized Born Surface Area [7] (GBSA), which is an implicit solvent model, is used in the path calculations. Given the two end states, Minimum Energy Paths (MEP) of 64 structures (including the end points) are computed with and without the Mg^{2+} ions at the core. Each of the MEP configurations was then solvated with SPC/E water [8] and explicit ions at 300K. The simulations were conducted with the MOIL package [17]. The isokinetic ensemble kept the temperature constant and the volume was 93.9x 64.0x 73.3Å^3. The free energy calculations uses SHAKE [21] to constrain bonds of the RNA molecule to their ideal value and MSHAKE [22] to fix the geometry of water molecules. Velocity Verlet algorithm is used to integrate the equations of motion. The non-bonded potential parameters of RNA are from OPLS-AA [23] while parameters for bonded terms are from the Amber99 forcefield [24]. This combination of the two force fields has been successfully applied to model different RNA systems in the past [25–27].

In preparation for the calculations of the free energy pathway, we conduct 80ns simulations of each of the aqueous solutions of the SPW path. The simulations fix the non-hydrogen atoms of the RNA letting the rest of the system move. After the equilibration of the aqueous solutions around the RNA, we used the String method [28,29] to compute minimum free energy paths. The String method is implemented in MOIL as an extension of LUP [30]. LUP is a similar algorithm to the zero temperature string [29] that was designed to compute Steepest Descent Paths (SDP). In calculations of minimum free energy paths (MFEP), a discrete representation of the reaction coordinate is quenched in the space of coarse variables. The quenching requires a thermal average over the non-coarse degrees of freedom (e.g. the water coordinates). Note that during the calculations the two end points (extended and collapsed configurations) are kept fixed.

The following differential equation provides quenched paths for the vector of coarse variables Q_i at path position i. We solve $dQ_i / dt = -\nabla_{Q_i} \langle U \rangle \left(\mathbf{I} - s_i s_i^t \right)$ for all intermediate structures i. τ is a fictitious time that is used for the quenching. In the present paper the coarse variables, Q, are the coordinates of the heavy atoms of the RNA molecule. The potential energy is $U\left(R_i\right)$ and s_i is the path slope in the space of coarse variables. It is approximated by $s_i \approx (Q_{i+1} - Q_{i-1}) / \left| Q_{i+1} - Q_{i-1} \right|$.

A time step, Δt, of 0.5 fs was used to compute the average force - $\left\langle \nabla_{Q_i} U \right\rangle$ in the NVT ensemble at a fixed value of the coarse variable (Q_i). Note that Δt is not the quenching time but the time step used to compute the average forces. The average forces are computed every 10 picoseconds and are used to quench the path as a function of the fictitious time τ. To ensure that the configurations are distributed uniformly along the path we re-parameterize the path after each iteration. We repeated this procedure for about 3ns of averaging time and used the last 1ns to generate the conformational ensembles for the extended-to-collapsed transition pathways.

Structure pool

About 350 configurations were selected from each path point providing a total of ~38,000 configurations in the pool. This procedure generates a large, varied and physically motivated pool from which ensemble selections could be made. The conformational landscape for the pool is shown in Figure S8 illustrating its breadth.

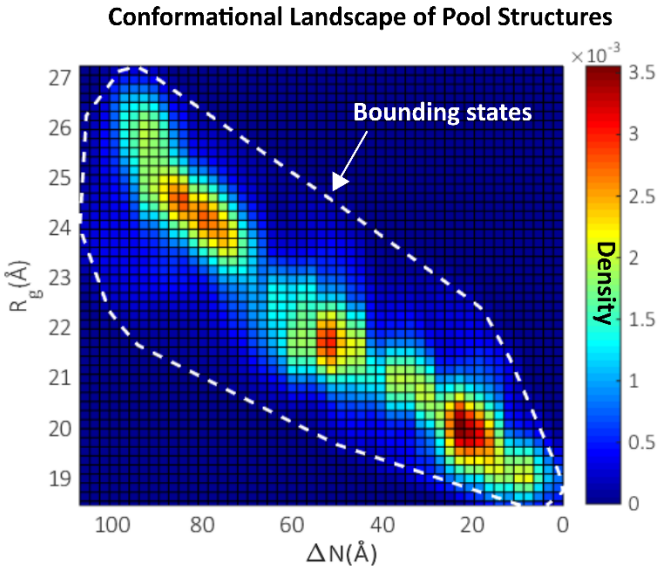

Fig S8. Conformational landscape of all models in the pool. The extreme (bounding) tP5 states are roughly shown by the white dotted line. This large pool provides a wide variety of possible conformations from which ensemble selections can be drawn.

The breadth and quality of models in the structure pool are two critical considerations for successful application of an ensemble selection scheme. For this system, we are fortunate that other experiments have provided high-resolution structures of unfolded and native RNA conformations to constrain and guide pool construction. Our process for constructing the pool is described above; essentially we computed the lowest free energy path between unfolded and folded structures, and computed orthogonal states for many structures along the path. We performed many tests to ensure that the final pool was as physically reasonable as possible. For example, inclusion of orthogonal states on the free energy path was essential to fit the data in multiple conditions, especially at the lower ionic strengths. This method allowed us to create a substantial pool of structures.

We did explore the effect of adding unphysical structures, for example using RNA conformations with non-native secondary structures, un-related to the known conformations of (either unfolded or folded) tp5abc. These conformations did not improve the fitting to the data, and often dramatically decreased the quality of the fitting when previously selected structures were removed. This experience justified the omission of these eccentric models from the structure set.

Ensemble Fitting

SAXS profile computation

The SAXS profiles for all MD generated structures were calculated using CRYSOL 2.8 [31], with maximum harmonic order 15, Fibonacci grid of order 18 and default hydration parameters. Due to concerns over hydration models, CRYSOL 3.0 (which implements a more realistic solvation protocol) was tested. The differences in the calculated SAXS profiles between the two hydration models are small, and much less than both the experimental uncertainty and noise on the data. Furthermore, CRYSOL is regularly implemented in the study of proteins (e.g. [32]) where hydration models are emphasized due to the lower scattering contrast of the molecular backbone [33]. Due to the vast array of salt conditions probed, and the uncertainty associated with current force-fields description of RNA, the set of structures had the ions removed before computation of the SAXS profiles. As with hydration models, the effects of the ion-atmosphere around RNA is small, and much less than the experimental uncertainty on the data.

Ensemble optimization

Ensemble optimization was run on each processed SAXS curve using GAJOE 2.0 for 2000 generations, 8 ensembles were selected with a self-optimizing ensemble size, and the scheme was iterated 100 times to accumulate statistics with constant subtraction allowed [34]. The latter was implemented to approximate the effect of the ion atmosphere around RNA, and ensures the shape of the SAXS profile is the more relevant fitting parameter than small variations in absolute intensity caused by hydration or ion atmosphere models. No change in selections or improvement in fits (as assessed by the reduced chi-square) occurred when the algorithm was run for 10,000 generations; the scheme had converged by 2000 generations.

Static ensemble analysis

The fits achieved using this ensemble optimization approach are shown in Figure S9 below. The goodness of fit (χ^2) for each individual ensemble was assessed by comparison of the ensemble (I_{ens}) and experimentally derived (I_{exp}) SAXS curves:

$$\chi^2 = \frac{1}{K-1}\sum_{i=1}^{K}\left(\frac{I_{exp}(q_i) - cI_{ens}(q_i)}{\sigma_{exp}(q_i)}\right)^2$$

where K is the total number of points in q-space, σ_{exp} is the experimental error at each q point and c is a scaling factor. The global fit for the full ensembles were assessed through the reduced chi-square for each experimental condition:

$$\chi_{Red}{}^2 = \frac{1}{N}\sum_{j=1}^{N}\chi_j{}^2$$

here, the summation is over all generated ensembles (N=800). Given the small size of the uncertainties on the static SAXS data, the fits achieved are in excellent agreement with the experimental data.

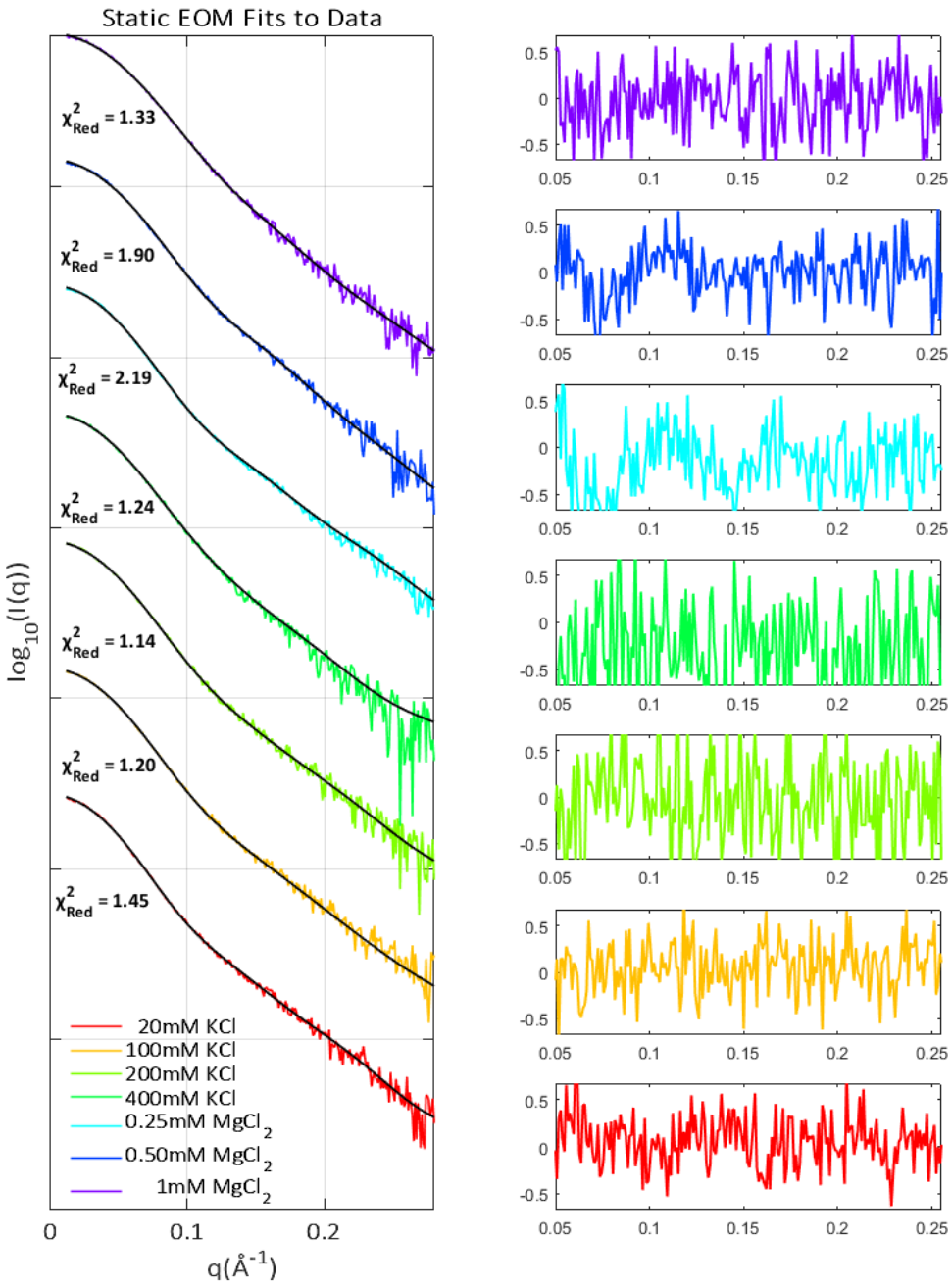

Fig S9: Fits achieved by ensemble fitting the structure pool to each static salt point (left) with the associated residuals (right).

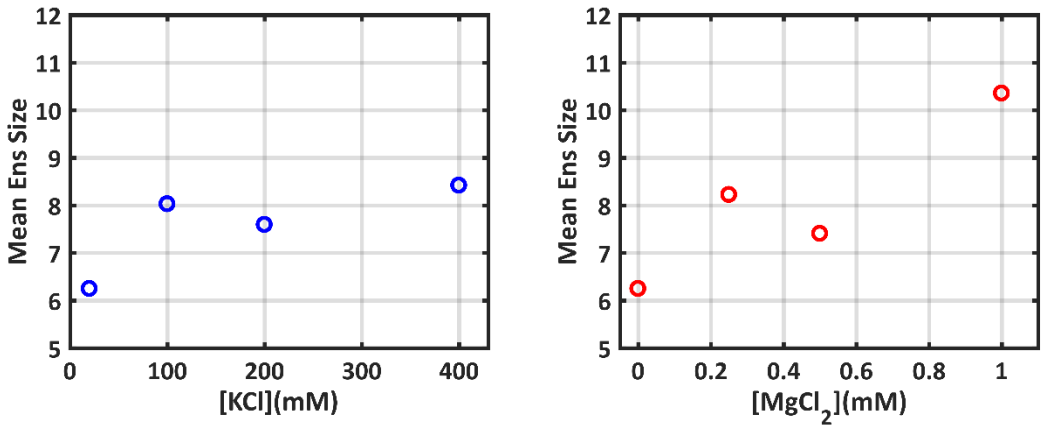

Fig S10: The mean ensemble size for each salt condition using the ensemble fitting approach discussed in the main text.

To parameterize the selected structures in a SAXS distinguishable manner, we calculated the R_g and Distance to Native state (ΔN) for each conformation. The former was calculated using the CRYSOL output (from a Guinier fit to the theoretical curve), while the latter was defined as the summed distance between pairs of tertiary contacts (as shown in Figure S11) relative to the native state. The C1 atoms in the appropriate residues were used when calculating the pairwise distances for ΔN.

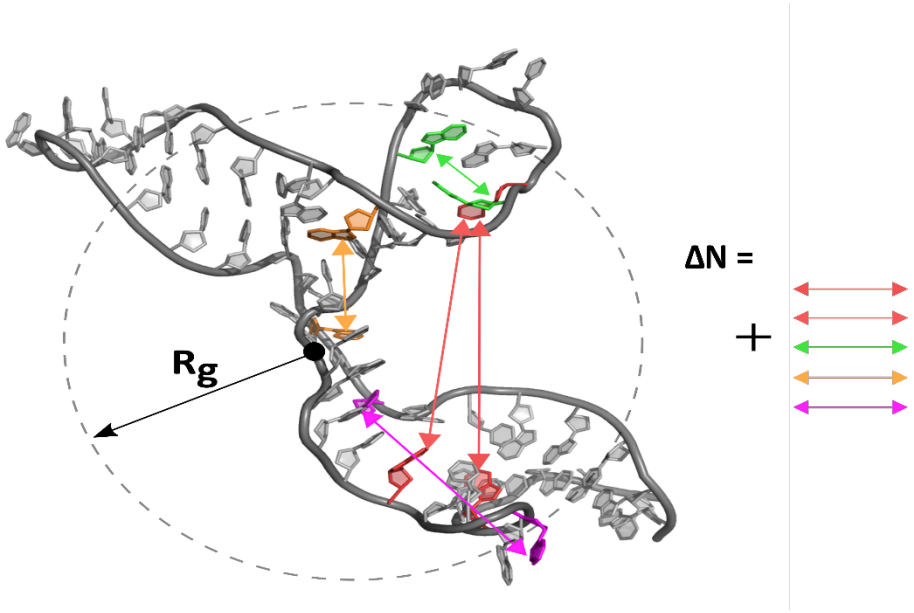

Fig S11: This panel shows an enlarged version of Figure 4d from the main text. The two metrics used in this analysis were the radius of gyration (R_g), which defines the rough global size of the molecule, and the Distance to Native state (ΔN) which tracks how close to the native state a given conformation is.

The conformational distributions defining the selected structures are shown as 1d projections onto R_g and ΔN in Figure S12 below. The distributions have been offset in each salt series for presentation purposes.

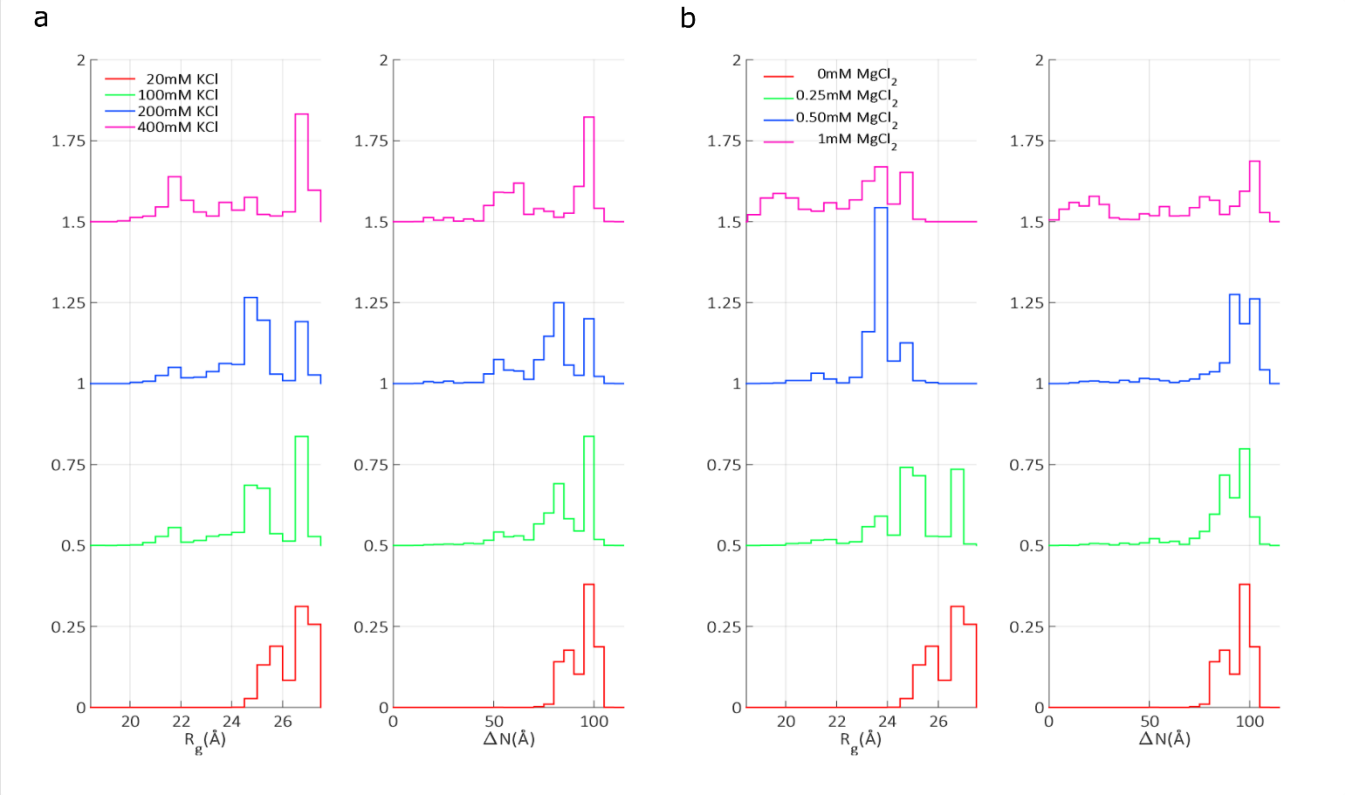

Fig S12. The R_g and ΔN distributions from the selected ensemble structures are shown for **(a)** KCl and **(b)** $MgCl_2$. The distributions areas are normalized to one, and offset in 0.5 increments to aid presentation.

As described in the main text, a convenient way to visualize the selected models is in terms of 2D landscapes defined by R_g and ΔN. In this representation, each selected structure contributes one point to the plot, located by its calculated value of R_g and ΔN. The conformational space has been divided into a 50x50 grid, and the heat on the map quantifies the density of structures within a 0.5Å circle centered about each point in the space. The densities have been smoothed by a running mean filter across 5 grid points to aid presentation. A common color map is used for landscapes within the main text Figures 4 and 6, but the colormap across these Figures differs for presentation purposes.

The two-dimensional representation of these histograms for each salt condition are presented below in Figures S13-S19. Key areas in the landscapes are circled, and 12 randomly drawn structures within these regions shown. Note that a combination of the presented structures is in general required to reconstitute the data. The sampling of these structures is large enough to capture the general theme of all structures contained within these regions, albeit without the intricate fluctuations the MD simulations provide.

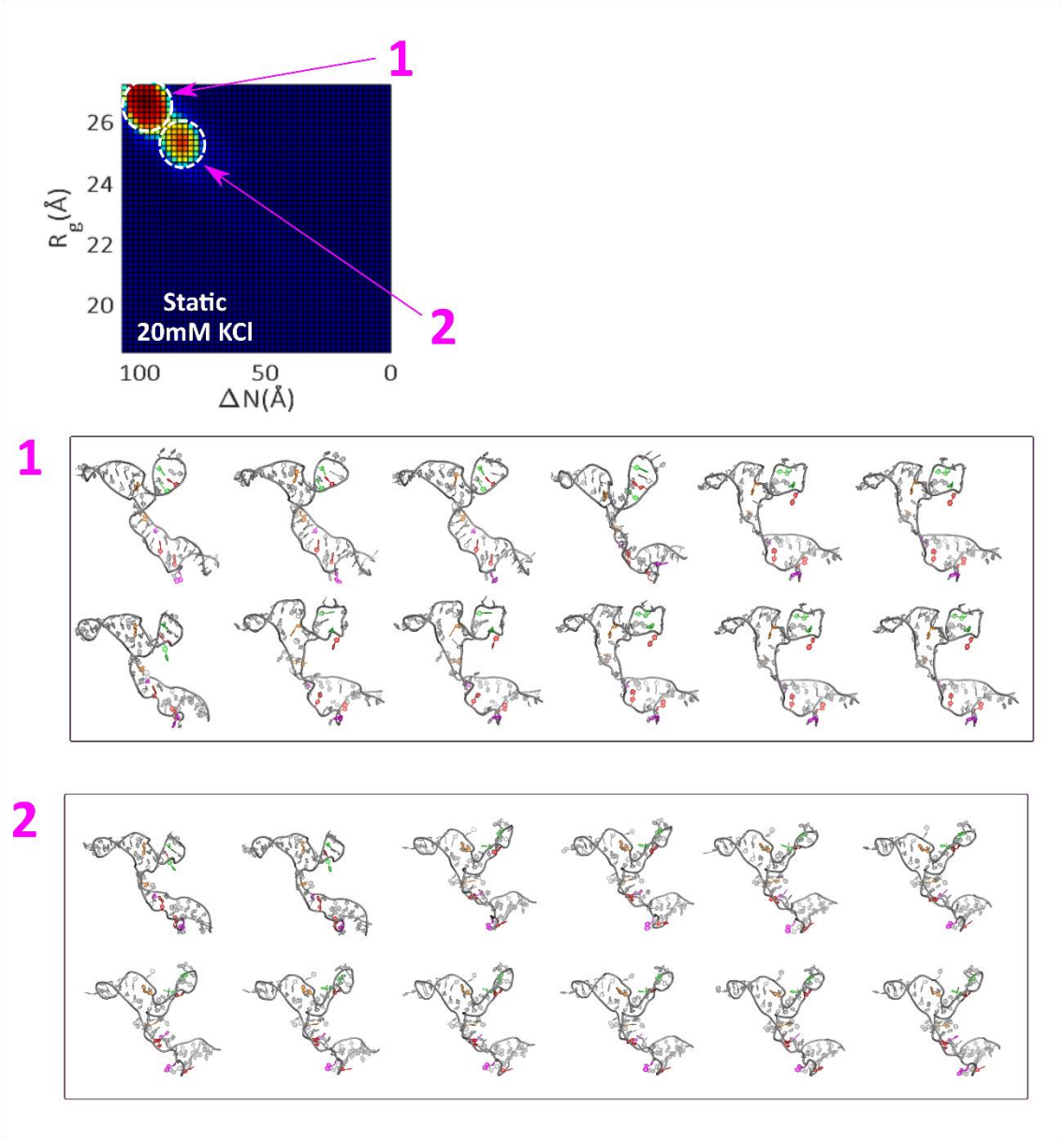

Fig S13. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

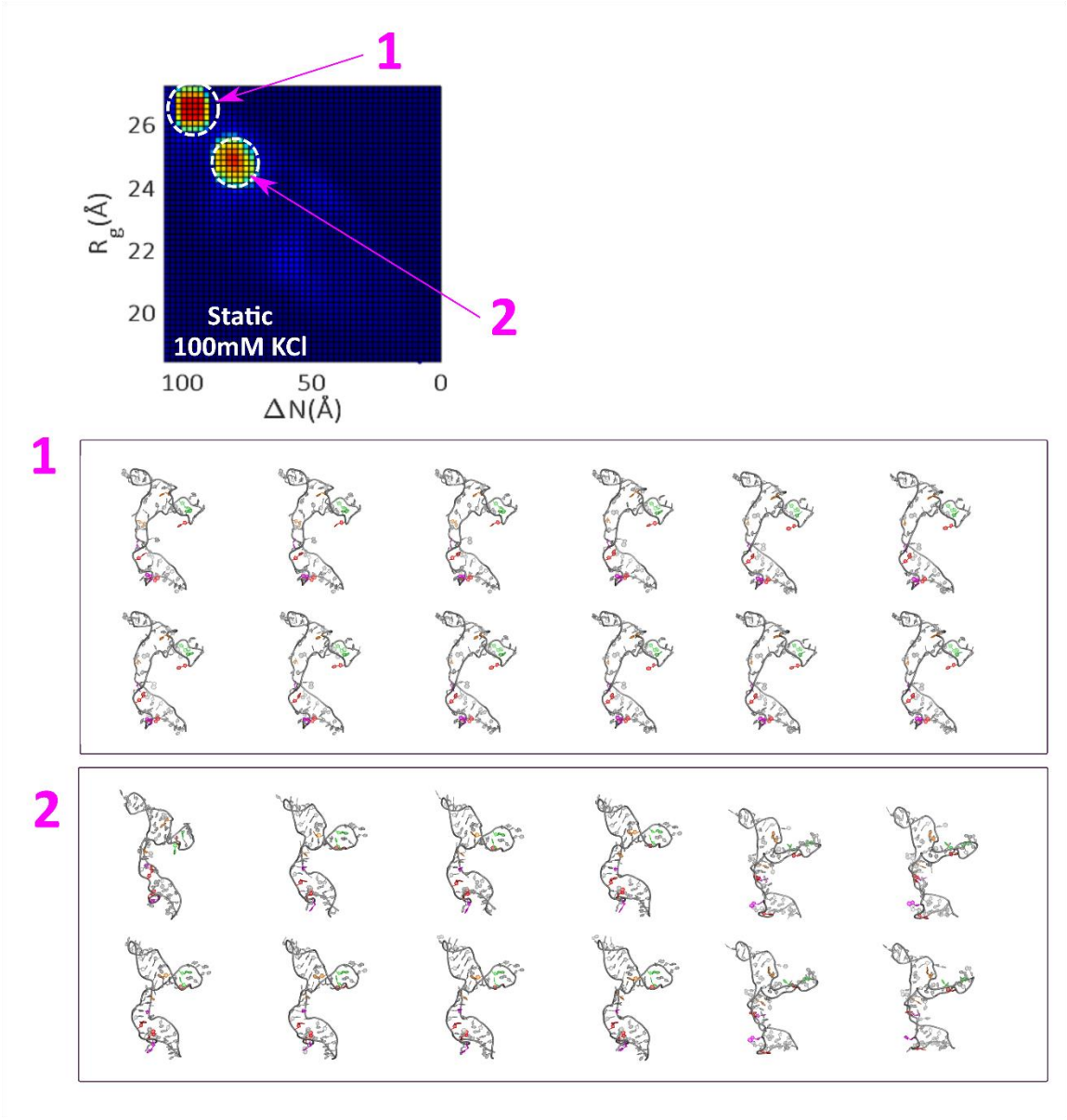

Fig S14. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

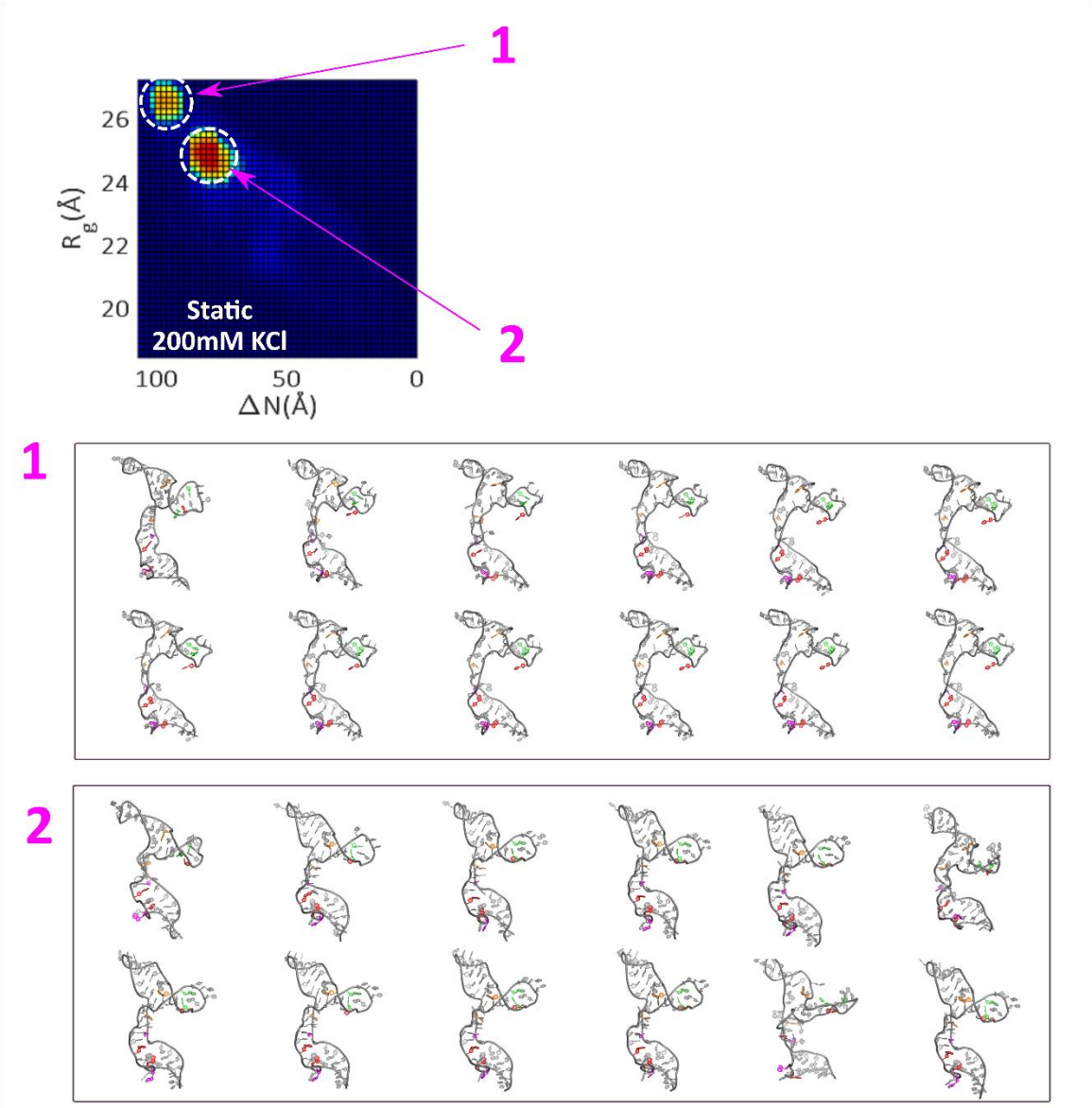

Fig S15. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

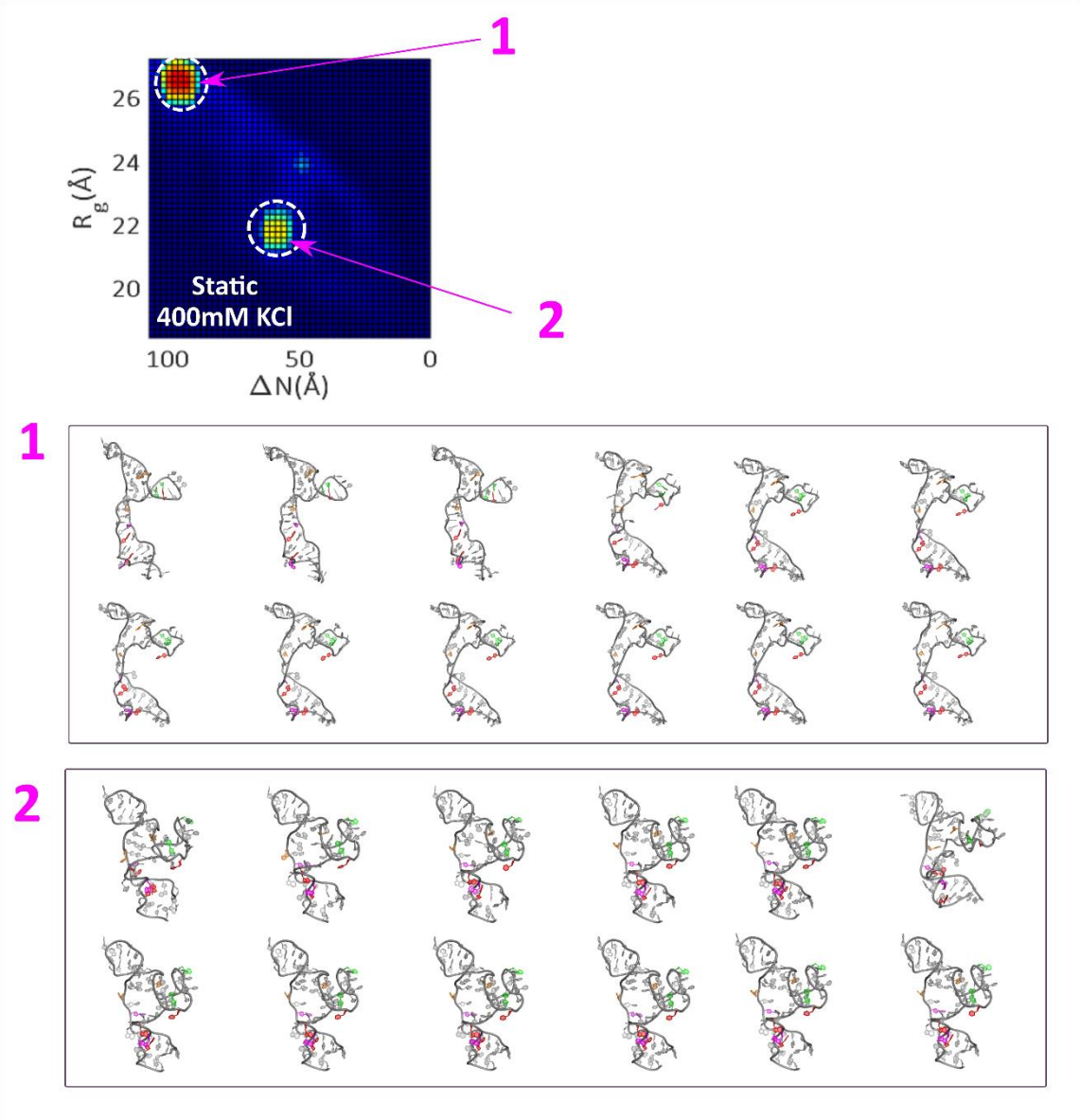

Fig S16. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

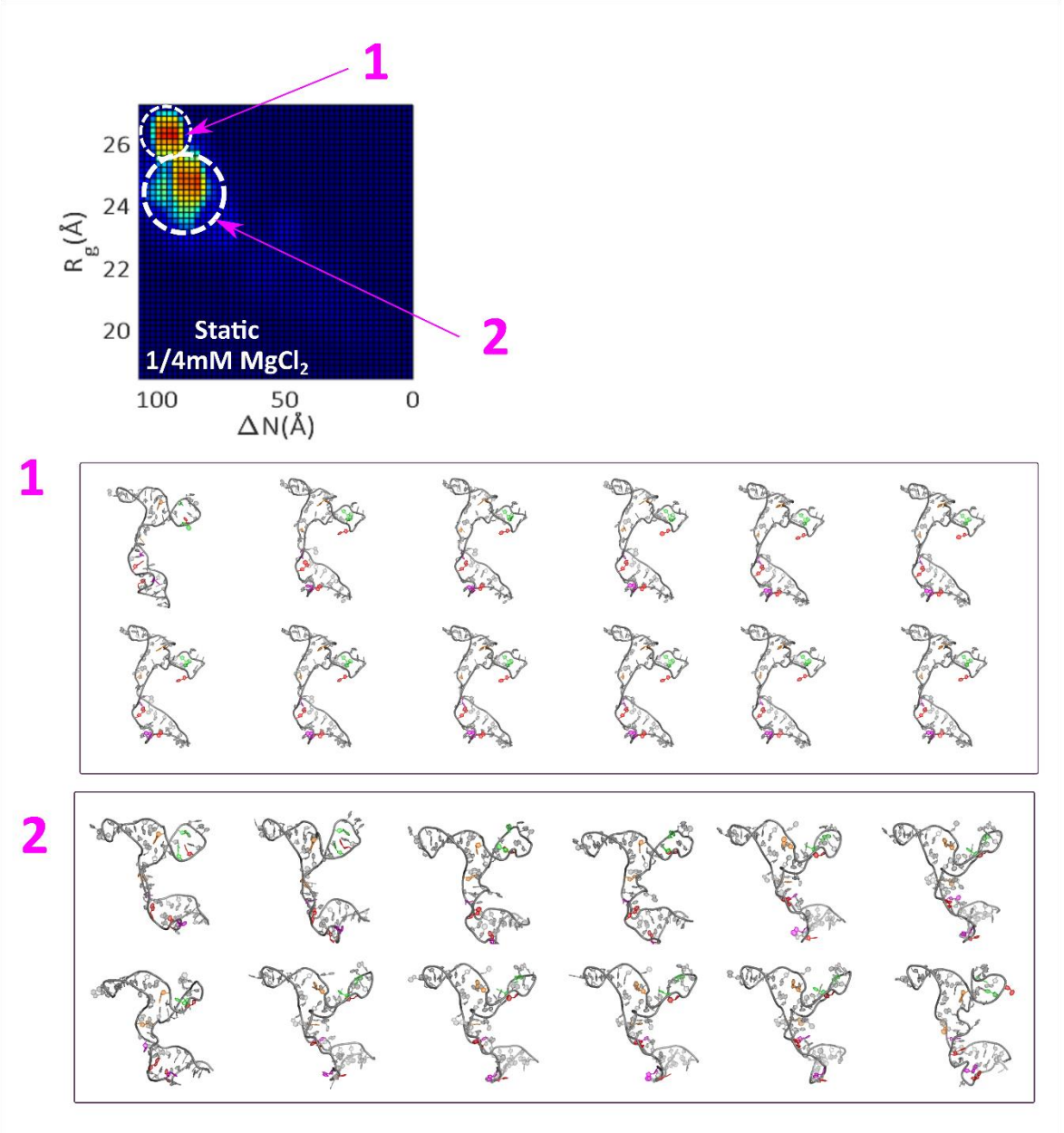

Fig S17. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

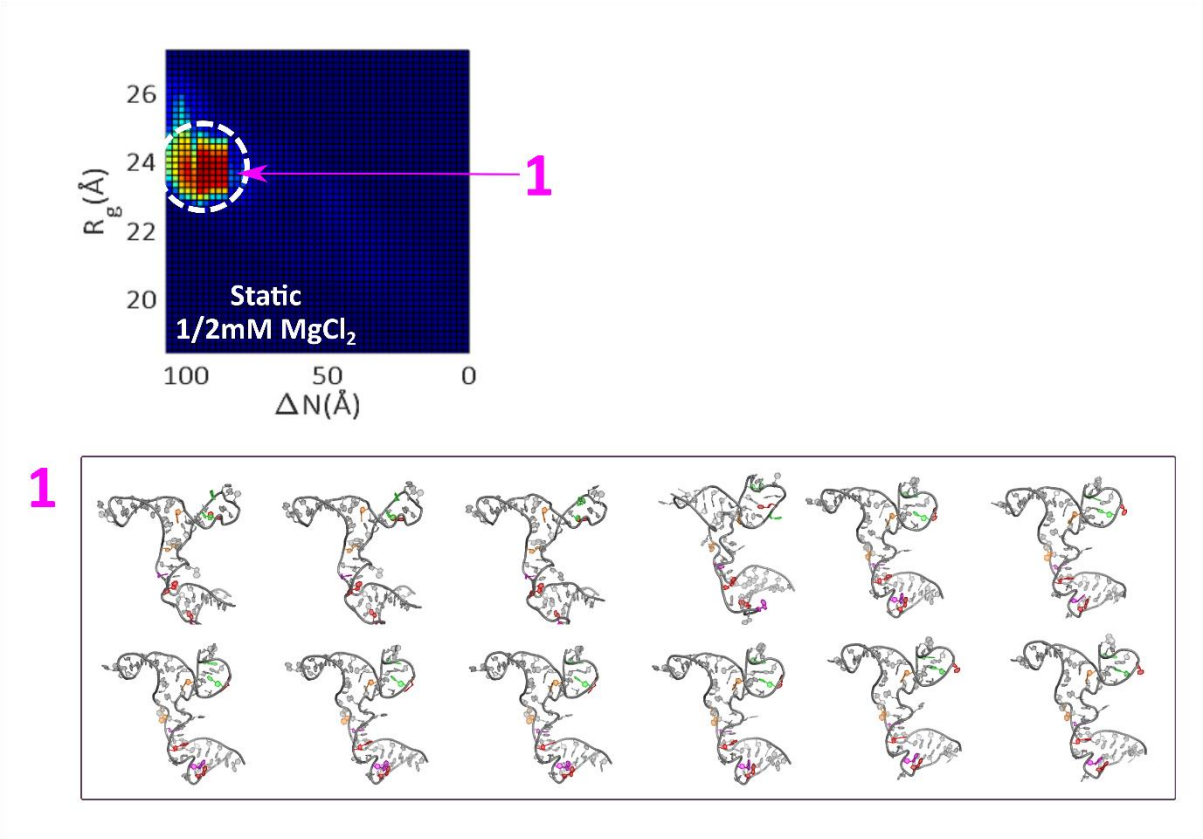

Fig S18. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

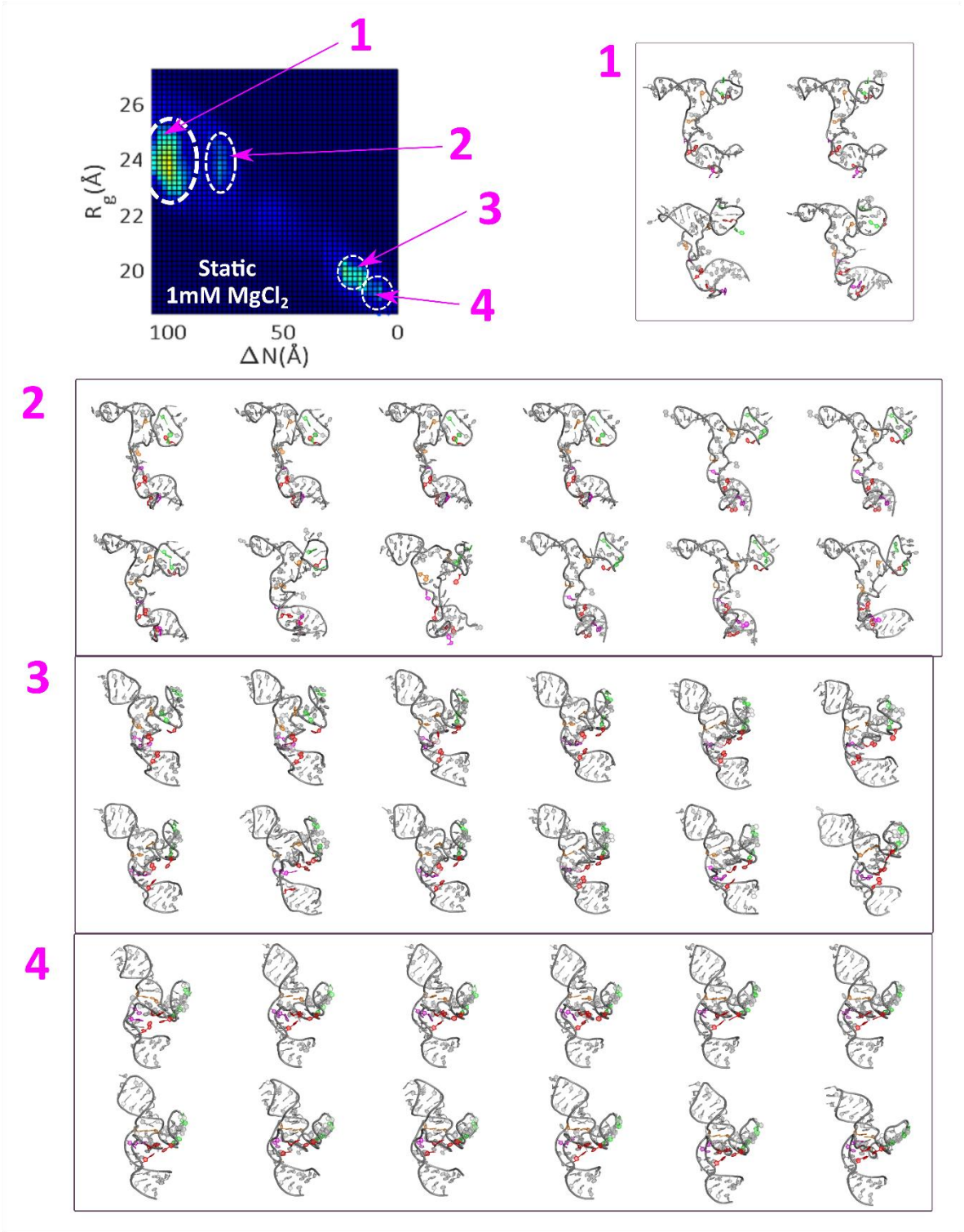

Fig S19. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Due to space constraints, only 4 open intermediate structures are shown. Pairs of tertiary contacts are colored as in Figure 1 of the main text.

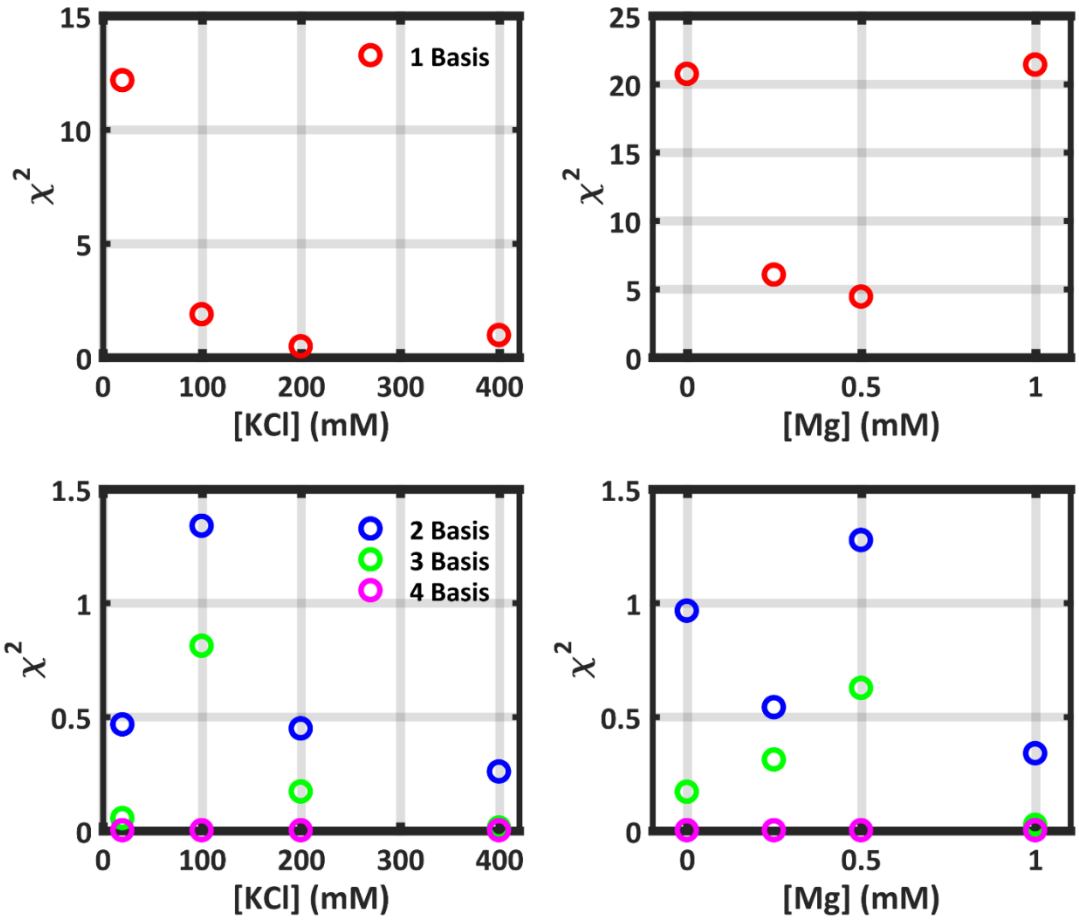

Fig S20. SVD identifies the minimum number of basis components required to fit the data for both KCl (left) and $MgCl_2$ (right) salt titrations. SVD was performed for each salt series, and the fits to the experimental data with increasing number of the identified basis vectors judged through calculation of χ2.

The fraction of tP5abc molecules in the native state as a function of salt are plotted in the Figure S21 below. At 1mM Mg, we measure ~11%, which is in good agreement with the 8% value reported in Reference [35]. The native state is never achieved at lower [Mg] or at any [K]. States which lack all native tertiary contacts are collectively refered to as 'non-native' in the Figure S21 (e.g. both helix-docked and open intermediate are considered 'non-native').

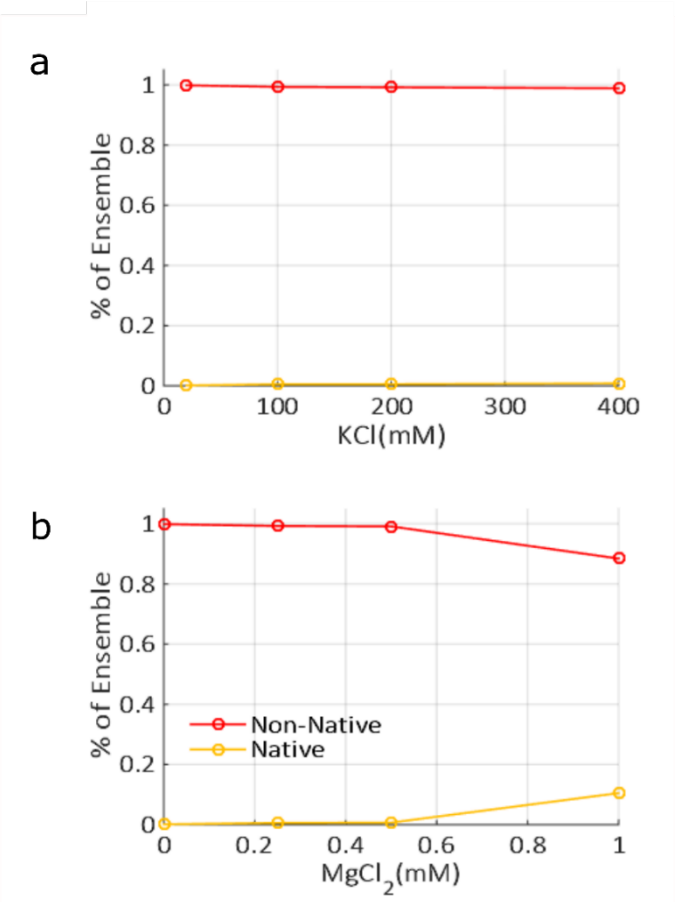

Fig S21. The fraction of tP5abc molecules occupying the native and non-native states for **(a)** KCl and **(b)** $MgCl_2$ in the static salt titration experiments.

Time-resolved ensemble analysis

The fits achieved following the application of the ensemble optimization technique to the TR data are shown in Figure S22 below, and were assessed as defined for the static data. Excellent agreement between selected structures and experimental data is observed; all computed values of chi-square are close to one.

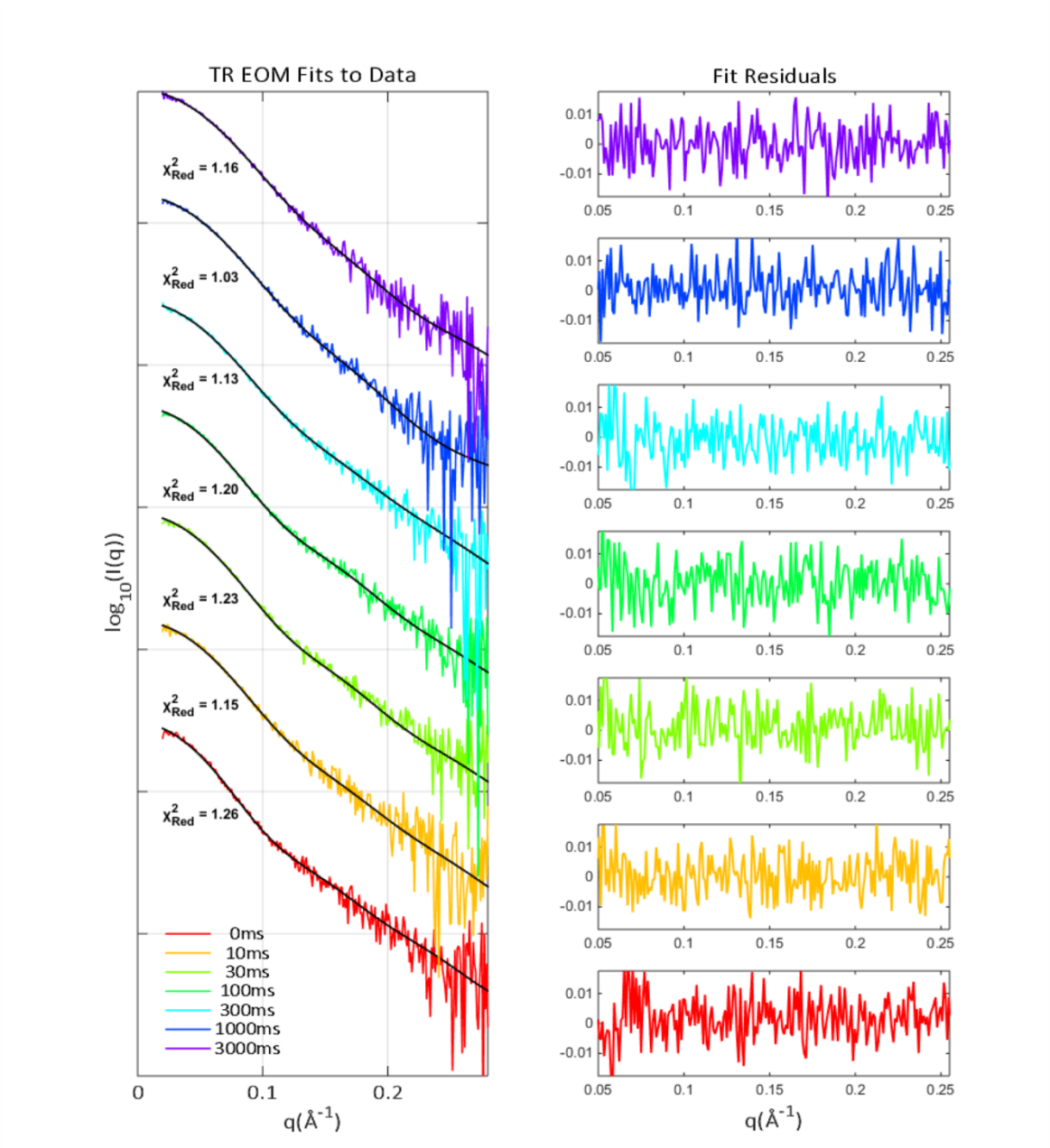

Fig S22. Fits achieved by ensemble fitting the structure pool to each time point (left), with the associated residuals (right).

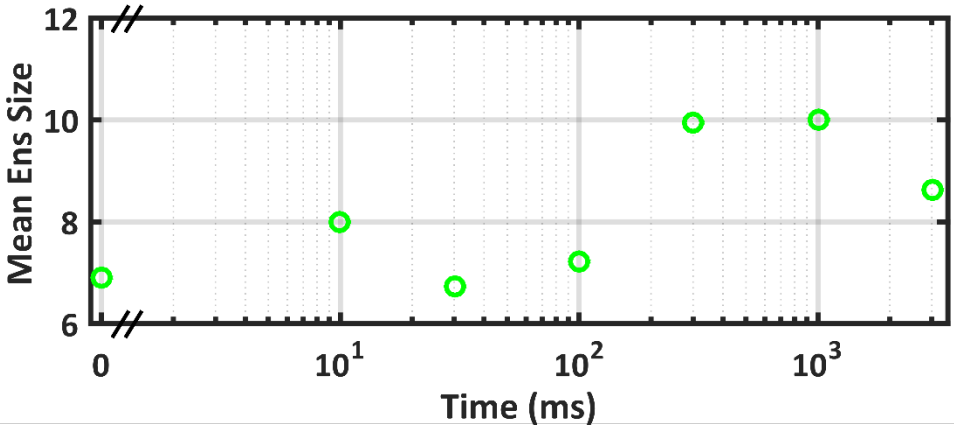

Fig S23: The mean ensemble size for each time-point using the ensemble fitting approach discussed in the main text.

The distributions of R_g and ΔN obtained through this ensemble fitting procedure for each experimental condition are shown as 1d histograms in Figure S24. These 1d projections allow the large swath of electrostatically relaxed states detected at 10ms to be seen more clearly (orange curve right panel), and furthermore emphasize the 'funneling' observed between 10-100ms after the initiation of folding.

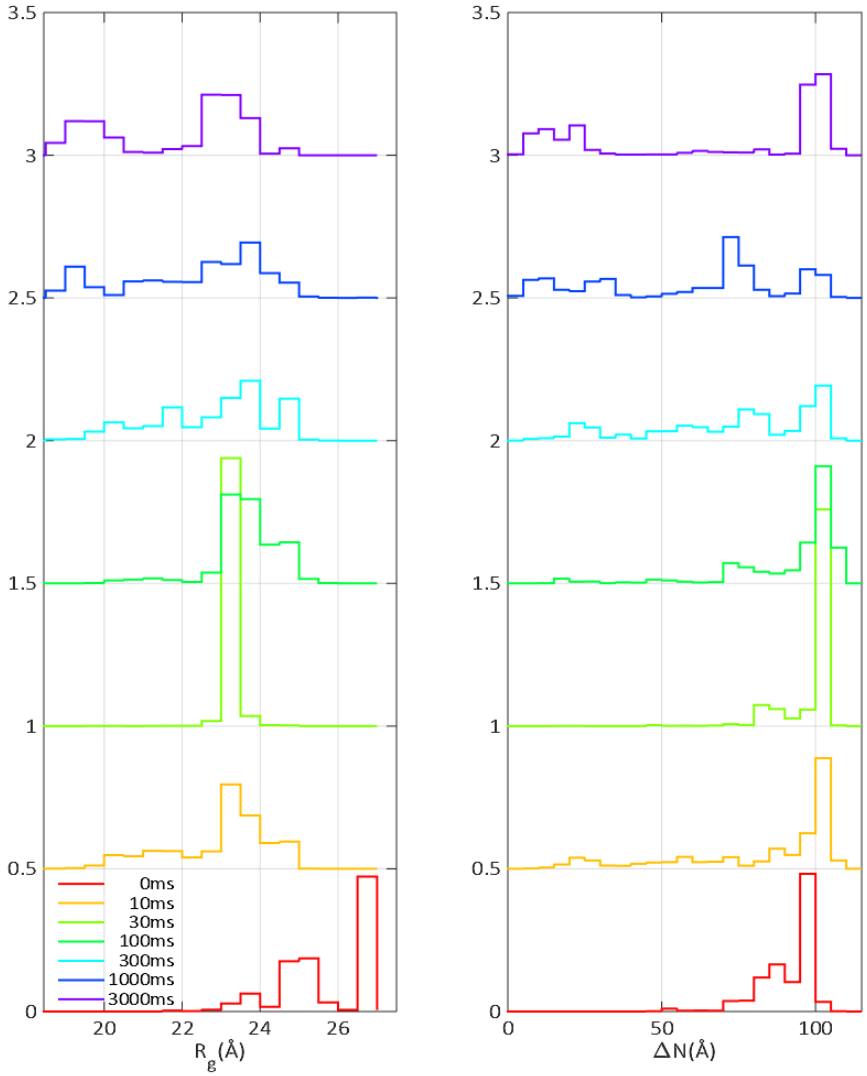

Fig S24. Distributions of R_g and ΔN for the underlying structures selected for each time point, normalized to one. Distributions have been offset in increments of 0.5 for presentation purposes.

Figures S25-S31 show the 2d landscapes defined by R_g and ΔN for each timepoint measured. Circles highlight important areas in each landscape. For each region, 12 randomly selected structures are displayed. The sampling of these structures is large enough to capture the general theme of all structures contained within these regions, albeit without the intricate fluctuations the MD simulations provide.

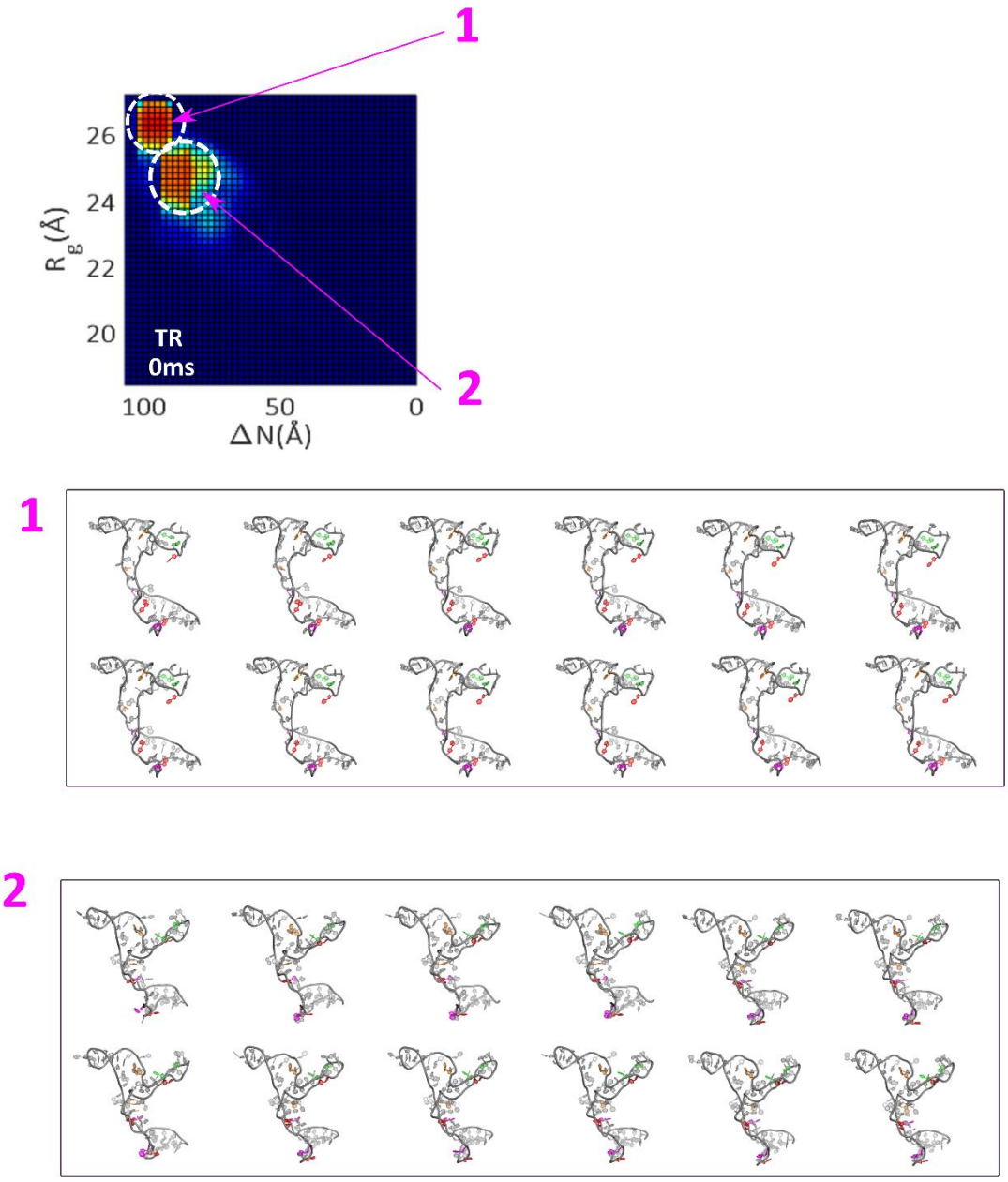

Fig S25. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

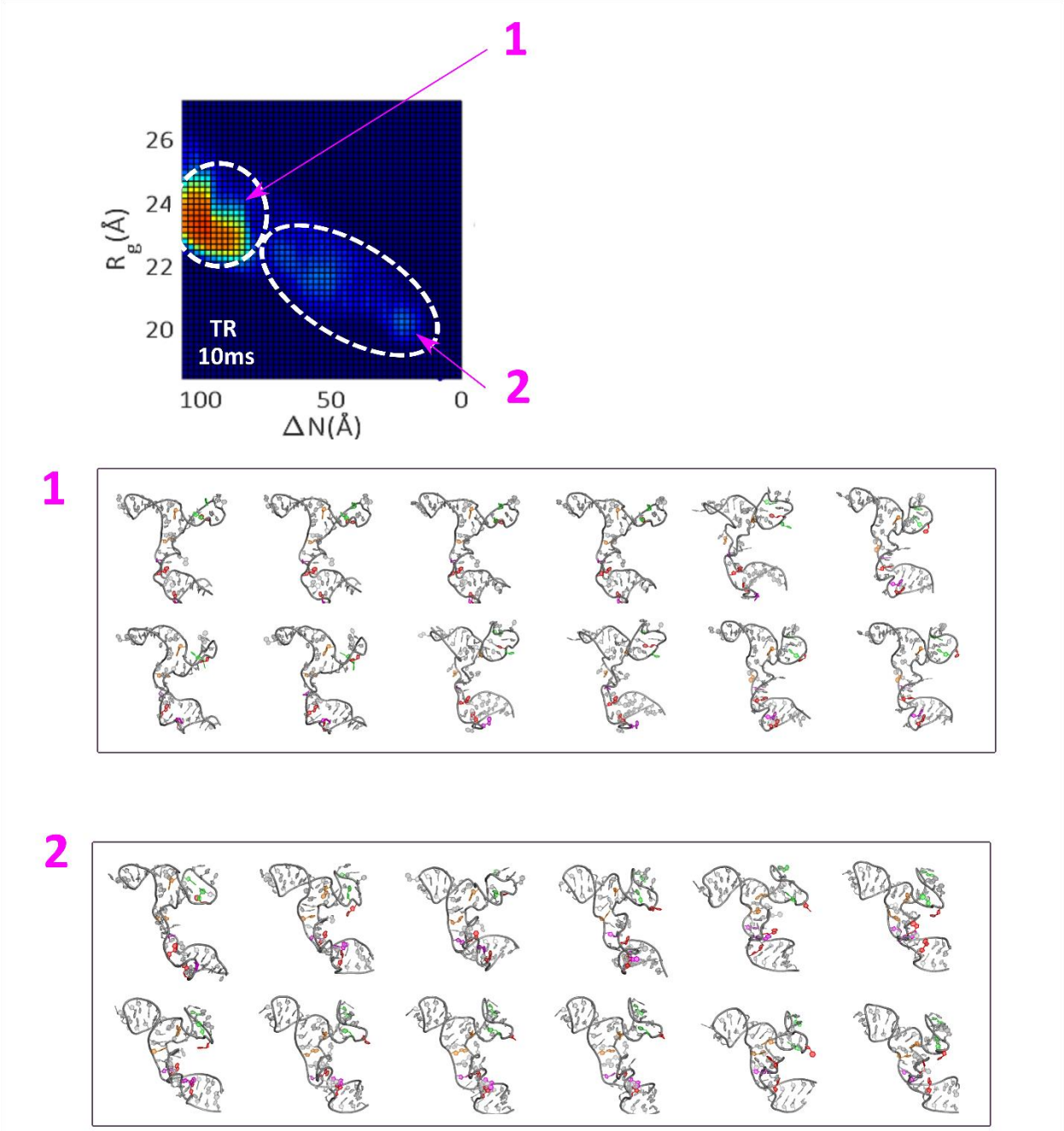

Fig S26. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

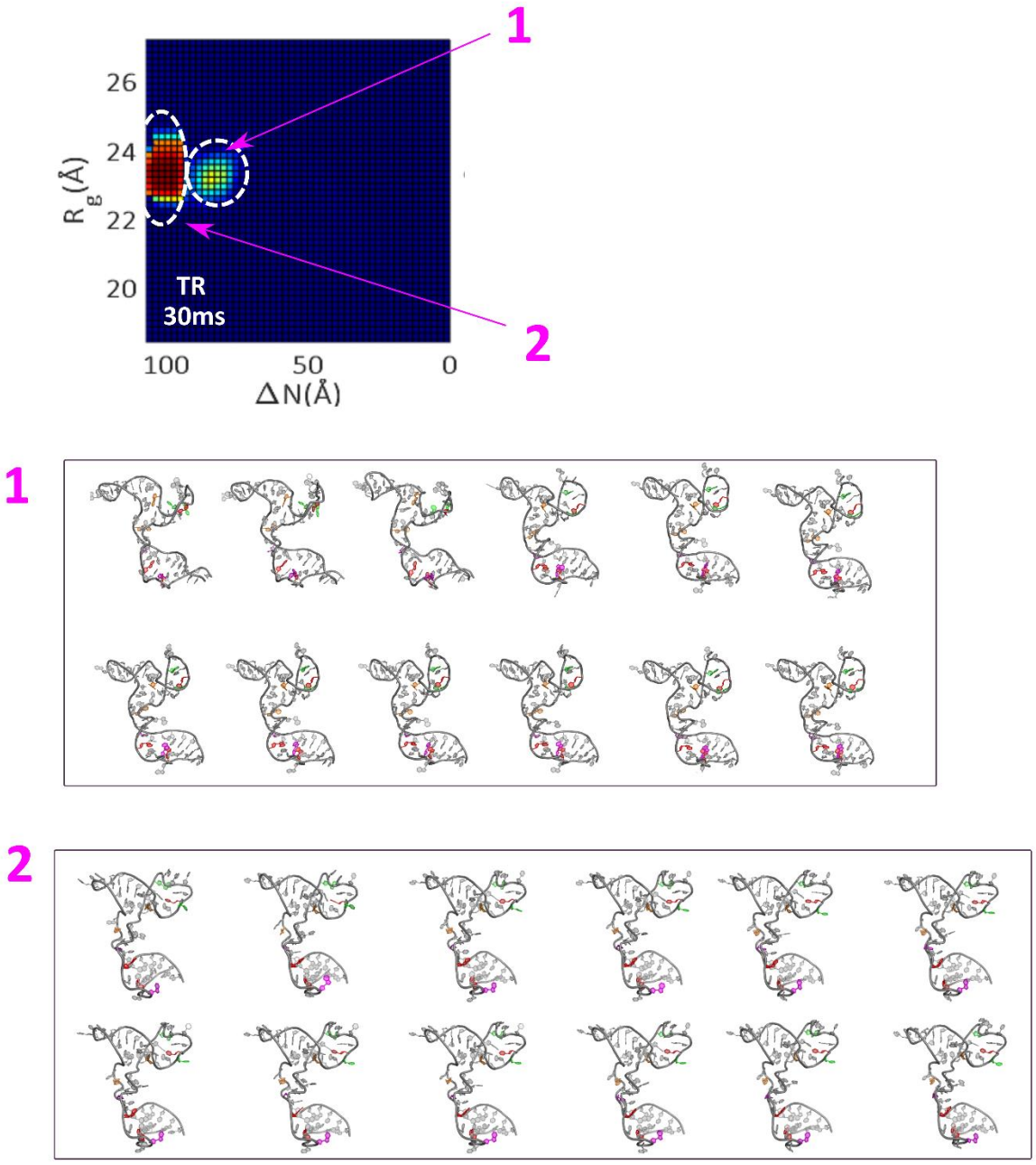

Fig S27. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

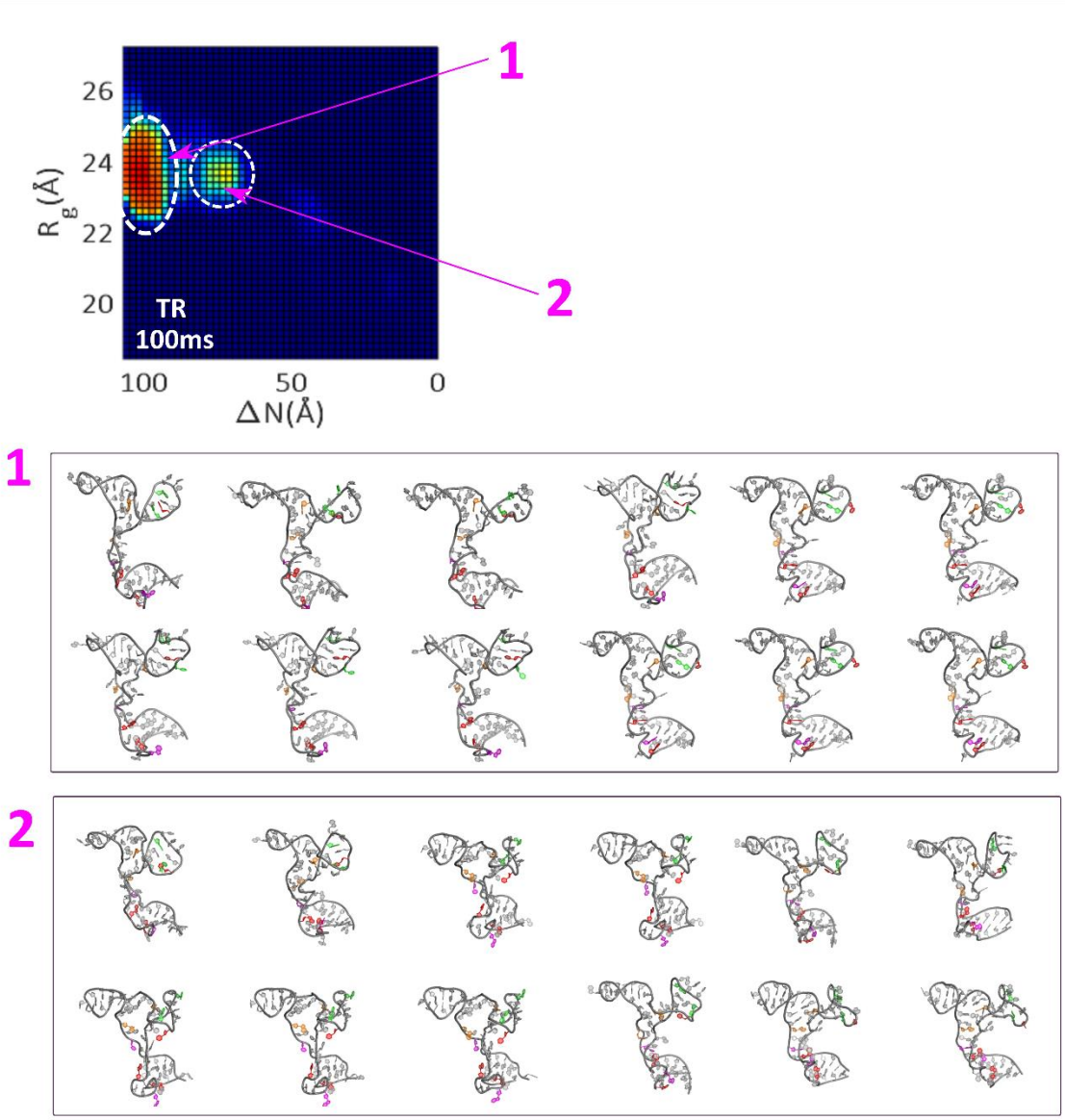

Fig S28. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

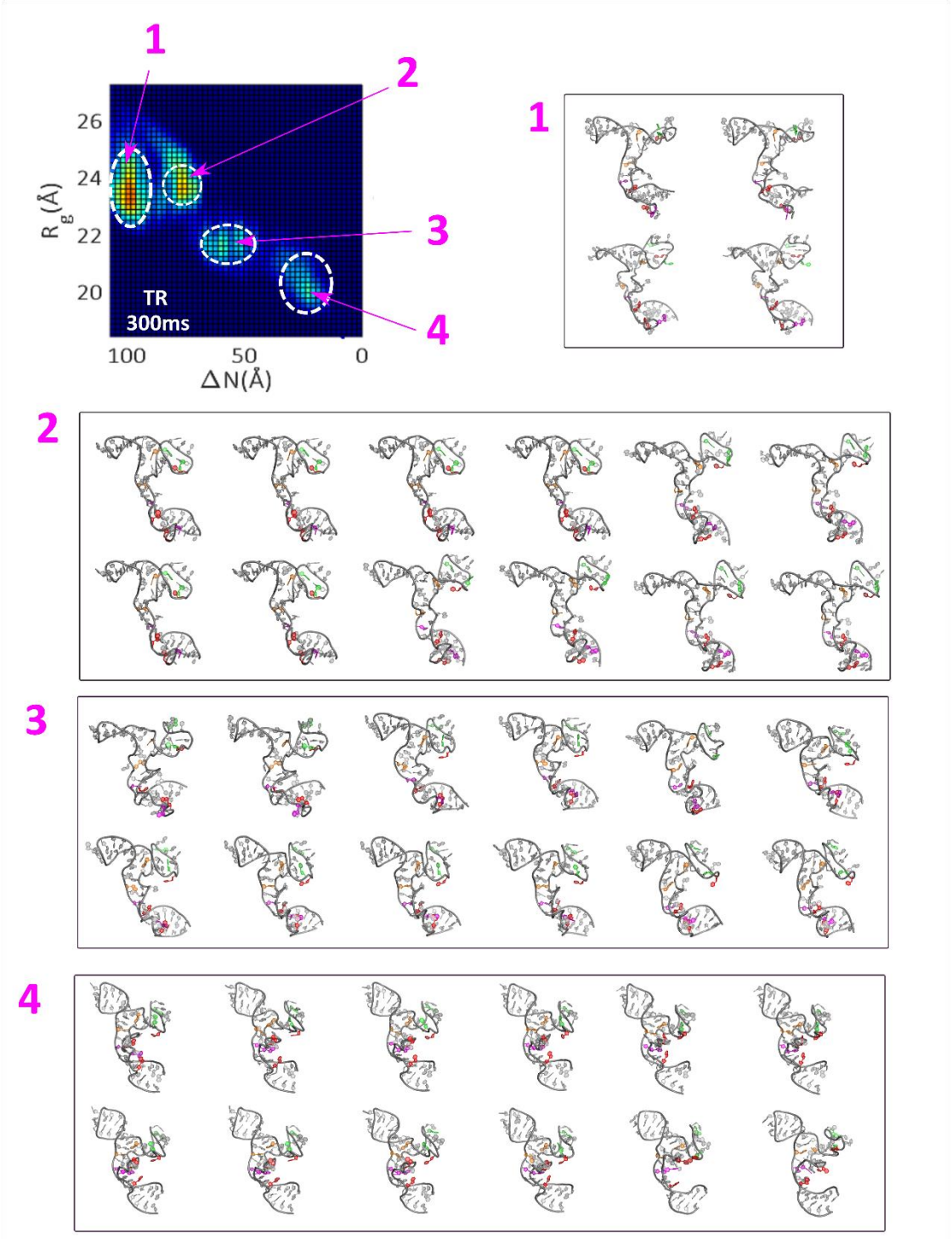

Fig S29. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Due to space constraints, only 4 open intermediate structures are shown. Pairs of tertiary contacts are colored as in Figure 1 of the main text.

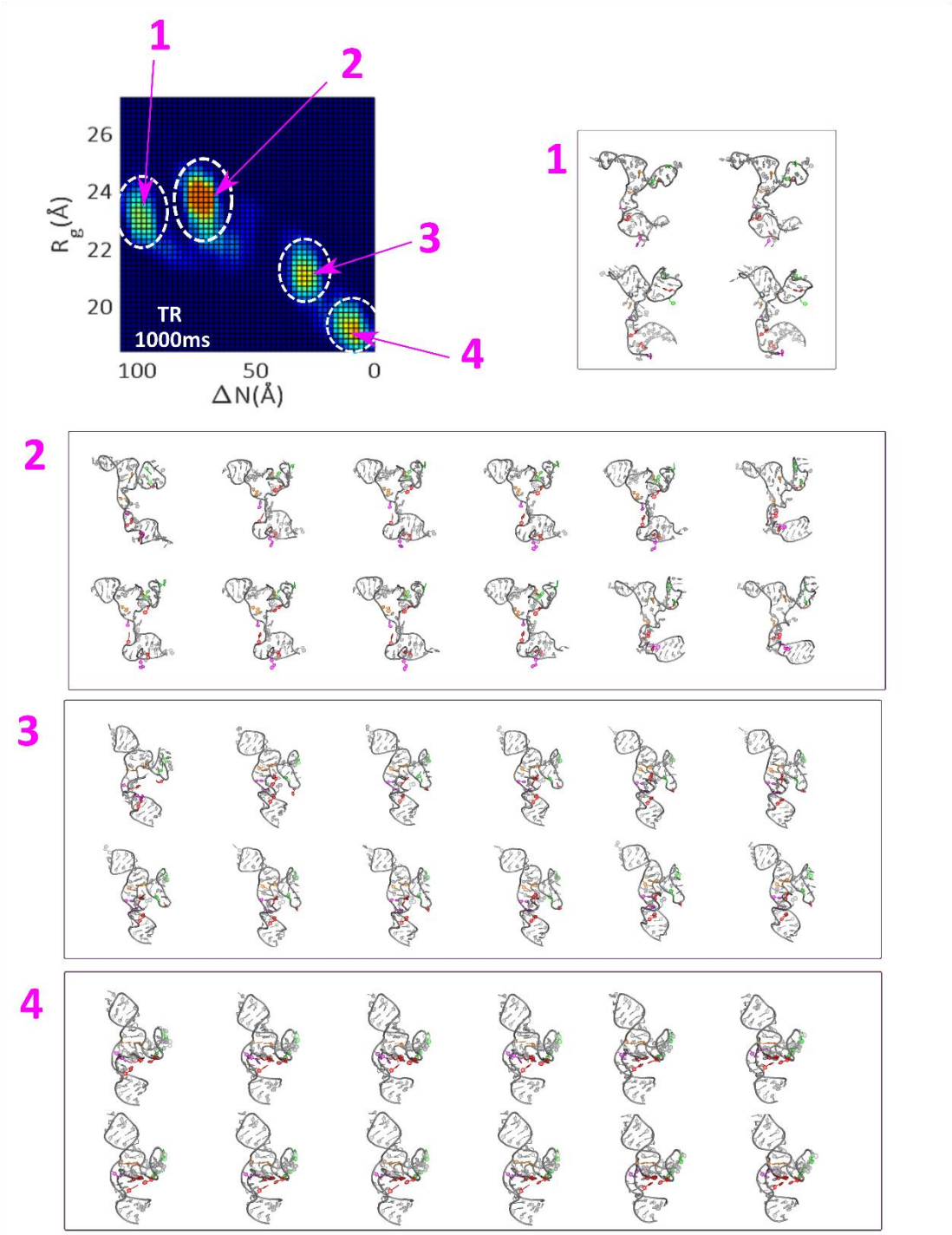

Fig S30. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Due to space constraints, only 4 open intermediate structures are shown. Pairs of tertiary contacts are colored as in Figure 1 of the main text.

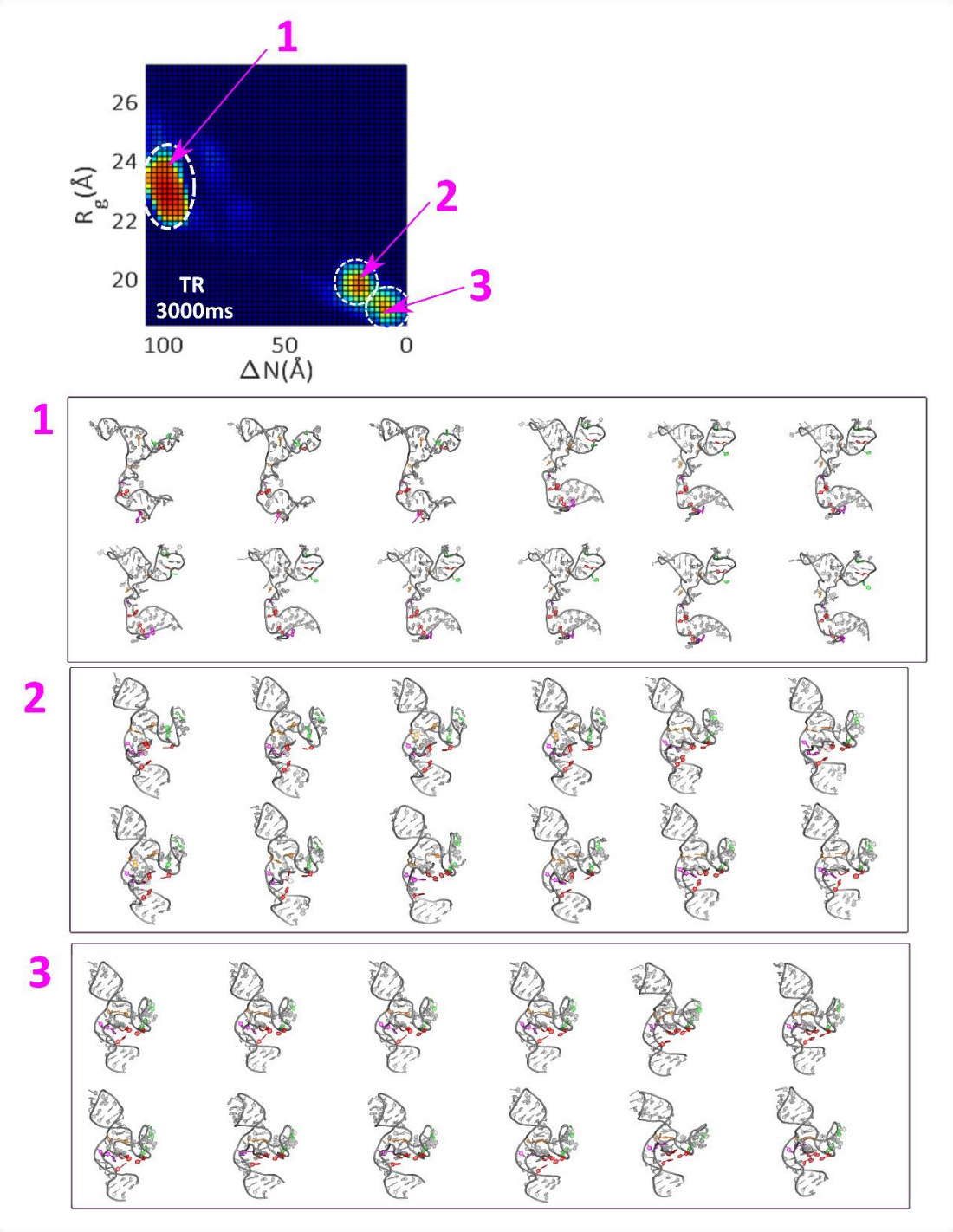

Fig S31. Conformational landscape of tP5abc in the condition listed (bold white text in heatmap). Regions of interest are circled, and 12 random structures are shown from each of these locales (boxed). Pairs of tertiary contacts are colored as in Figure 1 of the main text.

The fraction of tP5abc molcules in the native state is plotted as a function of time in Figure S32 below. In these experiments, ~15% of the population achieves this fold at the final time point, a slightly larger value than measured in the 1mM Mg static experiments. Given that the [Mg] in time resolved studies is slightly larger than the 1 mM of static experiments, this small variation is not surprising. Interestingly, the native state is never achieved before 300ms. As with the static analysis, conformations which lack all native contacts are collectively refered to as 'non-native'.

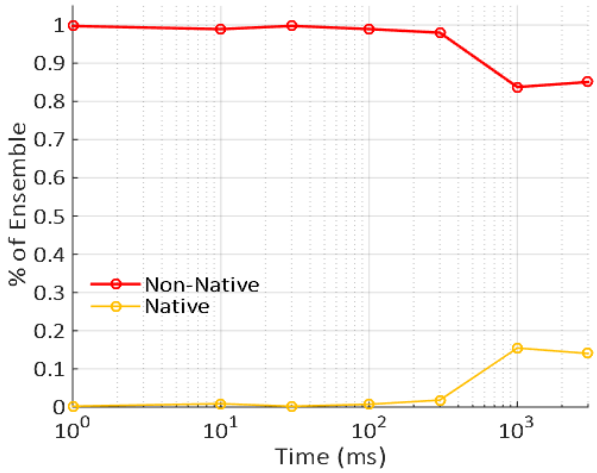

Fig S32. The fractions of tP5 molecules occupying the native and non-native states at each measured time point.

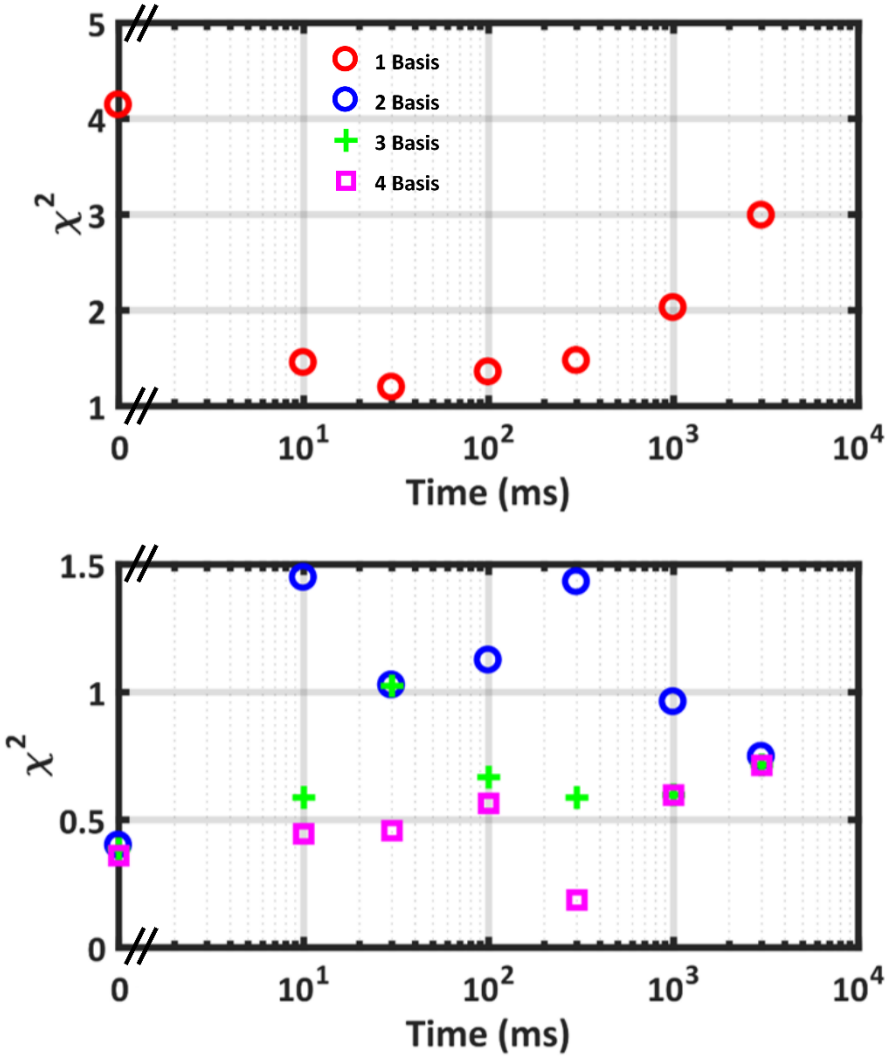

Fig S33: SVD identifies the minimum number of basis components required to fit the data for the time-resolved data. SVD was performed for the series, and the fits to the experimental data with increasing number of the identified basis vectors judged through calculation of χ2.

1D projections of static and time resolved ensembles

In Figure S33, we compare the distributions of R_g and ΔN for the relevant static salt points to the start/end points of the time-resolved experiments. Additionally, the 100ms time point is compared to the 0.50mM Mg static distribution to highlight the similarities between these two conditions. Good agreement is obtained between the start and end points; minor variations in the start point result from a slight structure factor present in the time-resolved data, the end point varies due to the slight variation in [Mg] of the time-resolved experiments.

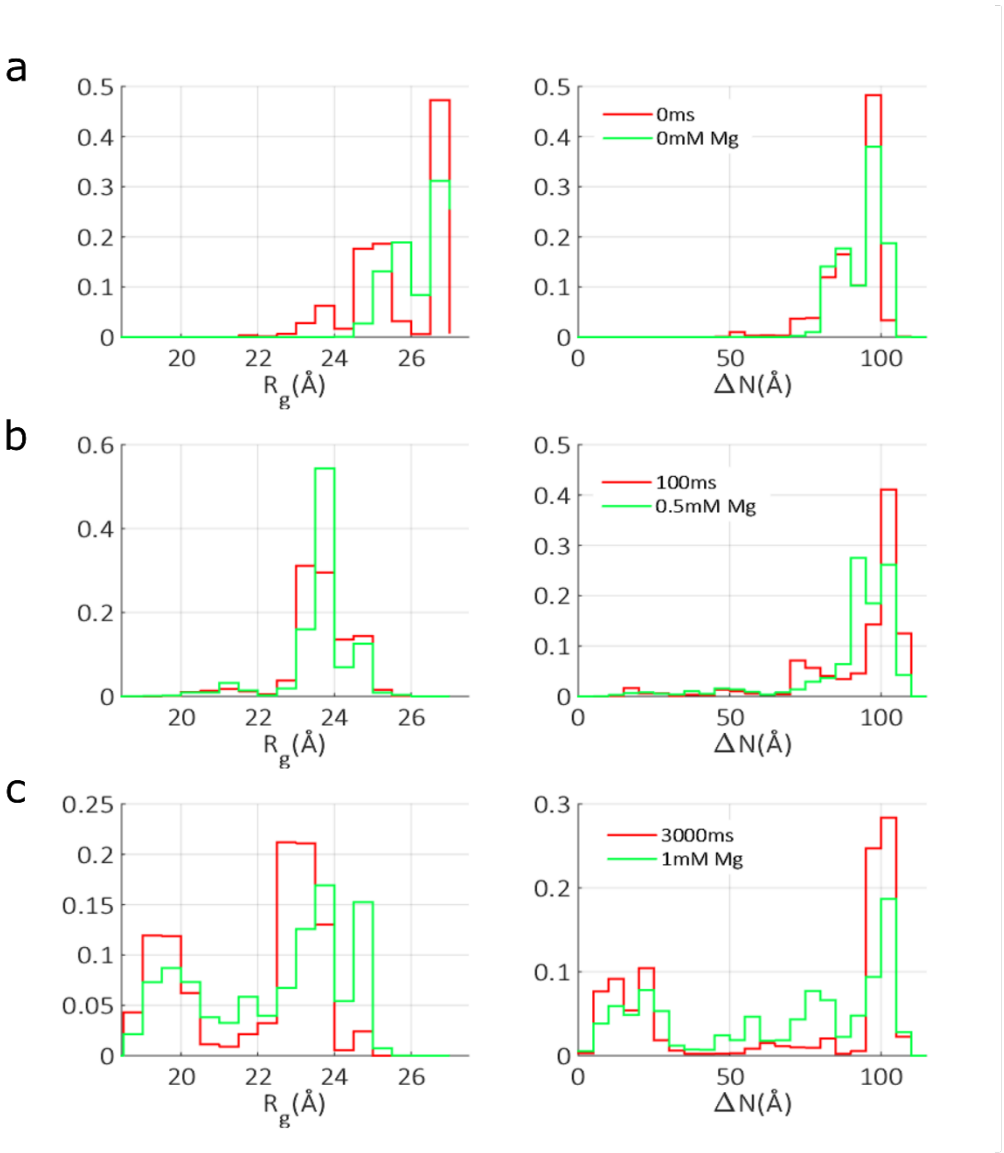

Fig S34. Comparison of R_g and ΔN distributions for the underlying structures selected for TR and static experimental conditions. **a)** The initial starting point for both experiments. **b)** Intermediate conditions for the time-resolved (100ms) and static (0.50mM Mg) experiments. **c)** End points for both experiments: 3000ms and 1mM Mg.

Role of Mg ions in key intermediates

Below, we show surface Mg ion density maps around three key intermediates in the folding pathway: the open, second and helix-docked intermediates. The figures illustrate the Mg ion atmosphere, as well as potential ion binding/localization regions that are prevalent in each model. For each of the states we select a representative conformation and conduct an equilibrium MD simulation to sample magnesium ion distributions for 200ns. To compute the magnesium ion localization we divide the simulation box into grids of 1Å in each direction x, y and z. All the simulation data where equilibrium is reached was used to compute the magnesium ion density. The average magnesium ion density at a grid k is simply computed as $< \rho_k > = < n_k > /dV$ where $< n_k > = \sum_t n_k(t)/L$ is the time average of the magnesium ion occupancy at that grid and *dV* is the grid volume. Results are normalized with the bulk density and shown as heat map in Fig S35.

Early in the folding pathway, the open intermediate shows a high Mg ion density around P5c, and ion localization around numerous other areas including the A-rich bulge, P5b and the pocket between the P5b/c stem and P5a.

In the second intermediate (which occurs at later times), the Mg ion density within P5c remains high, and a strongly localized ion density appears in the A-rich bulge. From the models, this appears concurrent with a 'rolling' up of P5a, placing this helix vertically under the molecular core. These calculations also reveal a growing ion density in the top stem of P5b, and in the pocket under P5c.

Finally, in the helix-docked state, an enhanced density of Mg ions appears in the pocket under P5c which accommodates the partially dehydrated Mg ion found in the native structure.

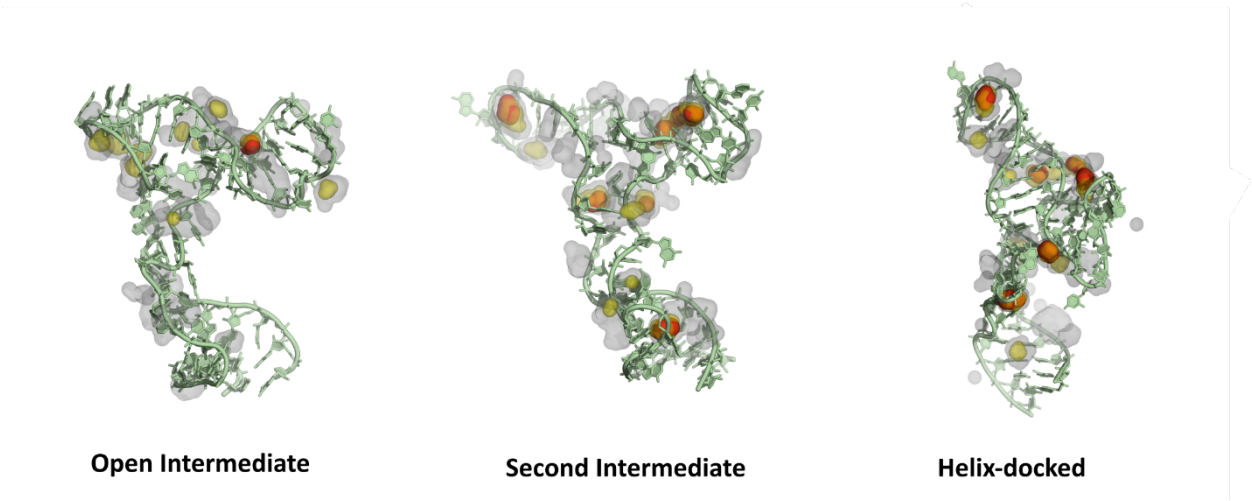

Fig S35. Mg ion density profiles along the folding reaction for three key intermediates. The local ion densities are computed by dividing the space into cubic grids. For clarity, only high-density regions are shown. 3d densities are shown with surface representation, where the gray represents regions with ion density $\rho > 5\rho_{bulk}$, yellow represents $\rho > 20\rho_{bulk}$ and red represents $\rho > 50\rho_{bulk}$.

Critical assessment of the ensemble approach

Finally, we note that ensemble fitting approaches must be used with caution. Valid concerns about overfitting the data can be remediated by testing the robustness of the results, and by benchmarking features of the ensembles against trends observed in the raw data. As an example, key features of the full folding pathway are revealed in the raw SAXS data. The small, but significant changes in the high angle scatter reflect subtle expansions or contractions and provide an outline that is not redrawn, but simply augmented by the models. An SVD analysis of the raw SAXS data also confirms EOM findings that a small number of distinct states are detected at each time point. Finally, the uniqueness of the solutions can be explored by removing frequently selected structures from the pool, or proposing unphysical structures as a solution. In these cases when the pool is deliberately altered, the quality of the fit is reduced in correspondence with the degree of alteration. Atomic scale agreement between computation and experiment can only be achieved by benchmarking against techniques that report data with Å resolution; however, with the appropriate caution, restrained interpretation, tests and careful validations, much can be learned from the methods described herein.

Supplementary References

1. Svergun,D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495–503.

2. Calvey,G.D., Katz,A.M., Schaffer,C.B. and Pollack,L. (2016) Mixing injector enables time-resolved crystallography with high hit rate at X-ray free electron lasers. *Struct. Dyn.*, **3**, 54301.

3. Miller,D.G., Rard,J.A., Eppstein,L.B. and Albright,J.G. (1984) Mutual diffusion coefficients and ionic transport coefficients lij of MgCl2-H2O at 25˚C. *J. Phys. Chem*, **88**, 5739–5748.

4. Li,Y.-H. and Gregory,S. (1974) Diffusion of ions in sea water and in deep sea sediments. *Geochim. Cosmochim. Acta*, **38**, 703–714.

5. Juneau,K., Podell,E., Harrington,D.J. and Cech,T.R. (2001) Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA-solvent interactions. *Structure*, **9**, 221–231.

6. Wu,M. and Tinoco Jr.,I. (1998) RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 11555–11560.

7. Onufriev,A., Bashford,D. and Case,D.A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins Struct. Funct. Genet.*, **55**, 383–394.

8. Abraham,M.J., Roland Schulz,T.M., Páll,S., Smith,J.C., Hess,B. and Lindahl,E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.

9. Berendsen,H.J.C., Grigera,J.R. and Straatsma,T.P. (1987) The missing term in effective pair potentials. *J. Phys. Chem.*, **91**, 6269–6271.

10. Allnér,O., Nilsson,L. and Villa,A. (2012) Magnesium ion water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.*, **8**, 1493–1502.

11. Smith,D.E. and Dang,L.X. (1994) Computer simulations of NaCl association in polarizable water. *J. Chem. Phys.*, **100**, 3757–3766.

12. Duan,Y., Wu,C., Chowdhury,S., Lee,M.C., Xiong,G., Zhang,W., Yang,R., Cieplak,P., Luo,R., Lee,T., *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, **24**, 1999–2012.

13. Van Gunsteren,W.F. and Berendsen,H.J.C. (1988) A leap-frog algorithm for stochastic dynamics. *Mol. Simul.*, **1**, 173–185.

14. Hess,B. (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.*, **4**, 116–122.

15. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: An N ·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.

16. Wennberg,C.L., Murtola,T., Hess,B. and Lindahl,E. (2013) Lennard-Jones lattice summation in

bilayer simulations has critical effects on surface tension and lipid properties. *J. Chem. Theory Comput.*, **9**, 3527–3537.

17. Bussi,G., Donadio,D. and Parrinello,M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**.
18. Parrinello,M. and Rahman,A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, **52**.
19. Daura,X., Gademann,K., Jaun,B., Seebach,D., van Gunsteren,W.F. and Mark,A.E. (1999) Peptide folding: When simulation meets experiment. *Angew. Chemie Int. Ed.*, **38**, 236–240.
20. Elber,R., Roitberg,A., Simmerling,C., Goldstein,R., Li,H., Verkhivker,G., Keasar,C., Zhang,J. and Ulitsky,A. (1995) MOIL: A program for simulations of macromolecules. *Comput. Phys. Commun.*, **91**, 159–189.
21. Ryckaert,J.P., Ciccotti,G. and Berendsen,H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, **23**, 327–341.
22. Weinbach,Y. and Elber,R. (2005) Revisiting and parallelizing SHAKE. *J. Comput. Phys.*, **209**, 193–206.
23. Pranata,J., Wierschke,S.G. and Jorgensen,W.L. (1991) OPLS potential functions for nucleotide bases. Relative association constants of hydrogen-bonded base pairs in chloroform. *J. Am. Chem. Soc.*, **113**, 2810–2819.
24. Kirmizialtin,S. and Elber,R. (2010) Computational exploration of thermodynamics and kinetics of mobile ions around RNA duplex. *J. Phys. Chem. B*, **114**, 8207–8220.
25. Kirmizialtin,S., Pabit,S.A., Meisburger,S.P., Pollack,L. and Elber,R. (2012) RNA and its ionic cloud: solution scattering experiments and atomically detailed simulations. *Biophys. J.*, **102**, 819–828.
26. Kirmizialtin,S., Silalahi,A.R.J., Elber,R. and Fenley,M.O. (2012) The ionic atmosphere around A-RNA: Poisson-Boltzmann and molecular dynamics simulations. *Biophys. J.*, **102**, 829–838.
27. Meisburger,S.P., Sutton,J.L., Chen,H., Pabit,S.A., Kirmizialtin,S., Elber,R. and Pollack,L. (2013) Polyelectrolyte properties of single stranded DNA measured using SAXS and single-molecule FRET: Beyond the wormlike chain model. *Biopolymers*, **99**, 1032–1045.
28. Maragliano,L., Fischer,A., Vanden-Eijnden,E. and Ciccotti,G. (2006) String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.*, **125**.
29. Ren,W., Vanden-Eijnden,E., Maragakis,P. and E,W. (2005) Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.*, **123**.
30. Ulitsky,A. and Elber,R. (1990) A new technique to calculate steepest descent paths in flexible polyatomic systems. *J. Chem. Phys.*, **92**, 1510.
31. Svergun,D., Barberato,C. and Koch,M.H. (1995) CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl.*

Crystallogr., **28**, 768–773.

32. Gibbs,E.B. and Showalter,S.A. (2016) Quantification of compactness and local order in the ensemble of the intrinsically disordered protein FCP1. *J. Phys. Chem. B*, **120**, 8960–8969.

33. Tokuda,J.M., Pabit,S.A. and Pollack,L. (2016) Protein–DNA and ion–DNA interactions revealed through contrast variation SAXS. *Biophys. Rev.*, **8**, 139–149.

34. Tria,G., Mertens,H.D.T., Kachala,M. and Svergun,D.I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*, 10.1107/S205225251500202X.

35. Xue,Y., Gracia,B., Herschlag,D., Russell,R. and Al-Hashimi,H.M. (2016) Visualizing the formation of an RNA folding intermediate through a fast highly modular secondary structure switch. *Nat. Commun.*, **7**, ncomms11768.