Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography

J.R. Stagno[1], Y. Liu[1], Y.R. Bhandari[1], C.E. Conrad[2,3], S. Panja[4], M. Swain[1], L. Fan[5], G. Nelson[6], C. Li[6], D.R. Wendel[1], T.A. White[7], J.D. Coe[2,3], M.O. Wiedorn[7,8], J. Knoska[7,8], D. Oberthuer[7], R.A. Tuckey[1], P. Yu[1], M. Dyba[1], S.G. Tarasov[1], U. Weierstall[3,6], T.D. Grant[9], C.D. Schwieters[10], J. Zhang[11], A.R. Ferré-D'Amaré[12], P. Fromme[2,3], D.E. Draper[13], M. Liang[14], M.S. Hunter[14], S. Boutet[14], K. Tan[15], X. Zuo[16], X. Ji[17], A. Barty[7], N.A. Zatsepin[3,6], H. N. Chapman[7,8], J.C.H. Spence[3,6], S. A. Woodson[4] & Y.-X. Wang[1]*

[1]Structural Biophysics Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA; [2]Department of Biochemistry, Arizona State University, Tempe, AZ 85287, USA; [3]Center for Applied Structural Discovery, The Biodesign Institute, Arizona State University, Tempe, AZ, 85287, USA; [4]Department of Biophysics, Johns Hopkins University, Baltimore, MD 21218; [5]Small Angle X-ray Scattering Core Facility, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA; [6]Department of Physics, Arizona State University, Tempe, AZ 85287, USA; [7]Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany; [8]Department of Physics, University of Hamburg, Luruper Chaussee 149, 22607 Hamburg, Germany; [9]Hauptmann-Woodward Medical Research Institute, Buffalo, NY 14203, USA; [10]Center for Information Technology, National Institutes of Health, Bethesda, MD 20892-5624, USA; [11]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, [12]Laboratory of RNA Biophysics and Cellular Physiology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA. [13]Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, USA; [14]Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, California, USA; [15]Structural Biology Center, Biosciences Division, [16]X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA; [17]Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA. *Corresponding author:

Yun-Xing Wang. Email: wangyunx@mail.nih.gov; phone: 301-846-5985

Riboswitches are RNA structural elements generally located in the 5′ untranslated region (5′ UTR) of mRNA. In the genetic regulation, ligand binding to the aptamer domain of a riboswitch triggers a signal to the downstream expression platform[1-3]. A complete understanding of the structural basis for this mechanism requires the ability to study structural changes over time[4]. Here we apply femtosecond X-ray free electron laser (XFEL) pulses [5,6] to obtain structural measurements from crystals so small that diffusion of a ligand can be timed to initiate a reaction prior to diffraction. We demonstrate this approach by determining four structures of the adenine riboswitch aptamer domain during the course of a reaction involving two apo, one ligand-bound intermediate, and the final bound states. These structures support a reaction mechanism model with at least four states and illustrate the structural basis for signal transmission. The two apo conformers differ significantly in the three-way junction and the P1 switch helix relative to the ligand-bound conformation. Our time-resolved crystallographic measurements with a 10-second delay captured the structure of an intermediate with changes in the binding pocket that accommodate the ligand. With a >10-minute delay, the RNA molecules were fully converted to the bound state, in which the substantial conformational changes resulted in conversion of the space group. Such drastic changes *in crystallo* highlight the important opportunities that micro/nanocrystals may offer in these and similar time-resolved diffraction studies. These results all together demonstrate the potential of "mix-and-inject" time-resolved serial crystallography to study biochemically important interactions between biomacromolecules and ligands, including those involving large conformational changes.

The *add* adenine riboswitch is a translational control riboswitch from *V. vulnificus*[7]. Its aptamer domain is a 71-nucleotide RNA (rA71). The term "riboswitch" refers to the distinct conformational switching that takes place in a ligand-sensing aptamer domain upon ligand binding, resulting in an altered pattern of gene expression[2,8]. At the core of the mechanism is a "switching sequence"[9-11], which forms the 3' half of the first helix (P1) of the aptamer domain. We first determined the structures of apo-rA71 using the serial femtosecond crystallography (SFX), revealing two molecules with distinct conformations (apo1 and apo2) in the asymmetric unit (Fig. 1). The two structures likely represent snapshots of the conformational ensemble present in solution (Extended Data Fig. 2), of which only one (apo2) is ligand-binding competent. Both apo conformers exhibit an overall purine riboswitch architecture[7,12], where helices P2 and P3 are fully formed and the L2-L3 kissing loop interaction is present. However, they differ significantly from the ligand-bound form (RCSB:4TZX[13]) in the three-way junction (Figs. 1, 2a and 2b) and in P1 (Fig. 1), with the heavy-atom RMSDs are > 6 Å for the three-way junctions, and 6.5 Å and 10.3 Å for P1 of apo2 and apo1, respectively (Extended Data Table 2).

The binding pocket was suggested to be locally disordered in the absence of ligand[14]. Our structures show that the empty binding pocket is structured. Both apo conformers reveal new sets of hydrogen-bonding and stacking interactions involving key residues (Fig. 2a and 2b). Consistent with solution studies (Supplementary Discussion)[14,15], U48, which is solvent-exposed in the ligand-bound form, flips in to form a wobble base pair either with the ligand-recognition residue U75 in apo1, or with U74 in apo2 (Fig. 2a and 2b). In particular, the U48•U75 (apo1) and U48•U74 (apo2) interactions are reminiscent of a mutant adenine riboswitch[16]. In addition, the role of U48 is similar to that of A14 in the preQ1 riboswitch, both of which rotate into the binding pocket in the absence of ligand[17]. In the J2/3 latch region, G46, U47 and U49 form

Watson-Crick base pairs with C53, A52 and A73, respectively, and C50 is stacked onto the U47-A52 base pair in both conformers. The J1/2 hinge region exhibits marked differences among all three structures, with RMSDs of ~7.8 Å (apo1 vs. bound), ~5.9 Å (apo2 vs. bound), and ~ 7.2 Å (apo1 vs. apo2) (Extended Data Table 2). These differences arise from altered base interactions of U22 and A23 (Fig. 2a and 2b). In apo2, both U22 and A23 flip out. In apo1, only U22 flips out, forming a hydrogen bond with the backbone of U25 in P2, whereas A23 forms a Hoogsteen base pair with U74, and is wedged between the bases of U48 and U49 (Fig. 2a and 2b). Both apo structures lack the three base-triples, A73•(A52-U22), C50•(U75-A21) and U51•(adenine-U74) that lock P1 and J2/3 together in the bound structure (Fig. 2a)[7].

In the absence of ligand-directed coaxial stacking between P1 and P3, the J1/2 hinge serves as a point of rotation for P1. The angular differences in the helical axis of P1 in the apo structures relative to the ligand-bound form are ~22° (apo2) and ~15° (apo1) (Fig. 1). The rotation and "kinking" at the hinge region in apo1 results in a ~8 Å translation of P1 relative to the bound structure (Fig. 1b), and a compression of the major groove formed by helices P1 and P3 in both apo conformers, which is ~10 Å wide compared to >16 Å for the ligand-bound conformer (Extended Data Fig. 3b). These changes are the direct results of 1) the loss of the ligand-facilitated coaxial stacking between P1 and P3, whereby junction J1/2 acts as a flexible hinge about which P1 pivots, and 2) the loss of key interactions that lock P1 and junction J2/3 together.

Our stopped-flow spectroscopy data (Fig. 2c; Extended Data Fig. 4) suggests at least a four-state kinetics model that not only supports two apo conformations, but also predicts an intermediate-bound (IB) state (Fig. 2c; Extended Data Fig. 4; Supplementary Discussion):

$$\text{apo1} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{apo2} + \text{ade} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{IB*•ade} \underset{k_{r}}{\overset{k_{f}}{\rightleftharpoons}} \text{B•ade}$$

Visualizing reaction intermediates using trapping or time-resolved crystallography has been reported.[4,18-21] Here, we use ligand "mix-and-inject" SFX[22,23] with RNA microcrystals to determine the structures of four reaction states of the adenine riboswitch in real time (Fig. 3a). Micro and nano crystals enable rapid diffusion of the substrate[23] and can accommodate large conformational changes of the molecules (leading to large changes in crystal volume) without breaking the crystal, while the serial measurement approach of measuring a crystal only once is applicable to the study of both reversible or irreversible reactions. The feasibility of mix-and-inject SFX hinges on at least three factors. First, the time of ligand diffusion throughout any given crystal must be negligible compared to that required for conformational changes associated with a reaction in crystals. The linkage time between dynamics and activities of most enzymes and RNA is on the ms timescale[24,25], whereas diffusion into crystals smaller than about 10 μm^3 volume is on the sub-ms timescale[23]. Second, diffusion must be sufficiently uniform so as to minimize crystal strain without destroying crystals during the process of transition. This is achieved because of a much larger surface to volume ratio and a much smaller total volume of a micro/nanocrystal compared to a macroscopic crystal. Third, mixing and delivery of a sample to an XFEL beam must be accomplished sufficiently faster than the rate of conversion to capture conformational intermediates, which can be achieved with a suitable nozzle[22].

We first performed mix-and-inject SFX with a long mixing delay of >10 min (10min-mix) to demonstrate the feasibility of converting the RNA i*n crystallo* to the final ligand-bound reaction state via relatively large conformation changes (Fig. 1). Both apo1 and apo2 were fully converted to the ligand-bound state, resulting in a polymorphic phase transition to a different space group and unit cell matching that of a previously reported structure[13] (Fig. 3b and 3c; Extended Data Fig. 5b; Methods, Extended Data Table 1). Such a transition may suggest that, in

cases where Bragg diffraction is still observable after conversion, rA71 molecules have undergone "sufficiently ordered" structural changes *in crystallo* that enabled the formation of a new lattice. This might be otherwise impossible in macroscopic crystals. It is noteworthy to point out that reactions involving proteins that have been studied thus far using time-resolved crystallography are those whose structural changes are not large enough to cause a lattice transition[4,21,26].

In order to capture the structure of IB using mix-and-inject SFX, the lifetime of the IB state in crystals must be longer than the mixing and diffusion times. In the case of rA71, the concentration of IB is expected to peak within milliseconds-seconds according to the time-course derived from solution kinetics (Fig. 2d). However, the peak time for IB depends on the actual ligand concentration in the crystals for this non-first-order reaction as well as the crystal lattice constraint. Given that the motions of both RNA and ligands are restricted in the microfluidic crystal environment, we postulated that the ligand-induced transition might not strictly follow the time-course predicted by solution kinetics, and that the conversion rate to B might be on the second time-scale, sufficiently slow to accumulate a large enough population of IB to be detectable by diffraction using the mixing device currently available (Fig. 3a; Extended Data Fig. 5a; Methods).

We performed mix-and-inject SFX with a 10-sec mixing delay (10sec-mix) to determine the structure of the intermediate state. The crystal lattice was unchanged (Extended Data Figs. 5b and 5c), suggesting that no major conformational changes take place within this time interval, characteristic of the IB based on our kinetic data (Supplementary Discussion). However, the difference electron density revealed key local conformational changes around the binding pocket (Extended Data Fig. 6). Structure analysis and refinement revealed the adenine in the ligand-

binding pocket, accompanied by altered conformations for U48 and A21 in apo2 (Figs. 4a and 4b). These changes were not observed in the binding-incompetent apo1 (Extended Data Fig. 7c).

In the IB structure, the ligand displaces U48 and is stacked between U48 and U49 (U49-Ade-U48) (Fig. 4b), similar to the U49-A21-U48 stacking in apo2 prior to binding (Fig. 4a). This is consistent with the low fluorescence of IB and little change in the fluorescence from the apo2 to IB states in our kinetic model (Supplementary Discussion). The movement of U48 in turn displaces A21 (Fig. 4b). The adenine ligand in the binding pocket of IB is stabilized by a network of H-bonds and base-stacking similar but not identical to those in the bound state (compare Figs. 4f and 4h). Specifically, the ligand H-bonds with O4 and N3 of U74 as well as the 2′-OH and a phosphate oxygen of U48 (Fig. 4f). In the absence of ligand, the ligand-recognizing residue U74 is base-paired with A23 or U48 in apo1 and apo2, respectively (Figs. 4e and 4g). In IB, the stacking amongst U49, ligand, U48, and U75 are reminiscent of apo1, except that the role of the ligand in IB is fulfilled by A23 in apo1 (Fig. 4c). In apo1, this base stacking together with a unique U75-A21-U48 base-triple (Fig. 2a) stabilizes apo1 in a "pseudo-intermediate" structural state, which may explain why apo1 is not the first to convert to IB within the crystal lattice. The rest of the IB structure, including P1, is similar to that of apo2. This is consistent with our kinetics model that predicts the formation of IB prior to the major conformational change (Supplementary Discussion). Furthermore, like apo1 and apo2, the IB structure still lacks the three base-triple interactions that stabilize P1 in the bound state[7], resulting in the partial disorder of P1 relative to the rest of the molecule, as indicated by the incomplete electron density for these residues, especially the 3′ strand containing the switching sequence (Fig. 4i), and high average B-factors (Extended Data Table 3). These structural heterogeneities,

arising from the instability of P1, are also evident from time-averaged molecular dynamics ensemble refinement[27].

Our results provide both a kinetics mechanistic model and the structural basis for how the signal of ligand binding is transmitted through P1. The low stability of the P1 helix in the context of a riboswitch is not unique to rA71[9,28,29]. Thus, modulation from a marginally stable helix in the ligand-free state to a stable helix in the ligand-bound state may be a common strategy to sensitize RNA structural switching to the presence or absence of cognate ligands. Our structures indicate that this is accomplished by ligand-induced rewiring of base interactions in the binding pocket. Equally important, the use of mix-and-inject SFX to determine these structures highlights the potential impact of this method for structural studies of non-photoactivated biological reactions in real time. Since most of these reactions are on the μs to ms timescale, the ensemble-averaging nature of SFX measurements and the properties of micro/nanocrystals may offer a new route for studying biological reactions and dynamics. In principle, many enzymatic reactions (>5000)[30] can be studied this way. These include many other types of interactions involving proteins or nucleic acids with ligands or substrates, such as where a very few conformation species in a reaction are dominant and the signals are discernible. As long as micro/nanocrystals of those biomacromolecules can be obtained, further improvements of both XFELs and mixing devices will lead to real time crystallographic studies of fundamental importance.

Acknowledgements

Portions of this research were carried out at the Linac Coherent Light Source, a National User Facility operated by Stanford University on behalf of the US Department of Energy, Office of

Basic Energy Sciences. The CXI instrument was funded by the LCLS Ultrafast Science Instruments (LUSI) project funded by the US Department of Energy, Office of Basic Energy Sciences. Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. We thank Drs. J. Strathern and M. Dunne for their support and S. Wakatsuki for insightful discussions. This work is supported in part by the NSF-STC "BioXFEL" (NSF-1231306), the NIH Intramural Research Programs of NCI, CIT, NHLBI, and the US Department of Energy, Office of Biological and Environmental Research under Contract DE-AC02-06CH11357, the European Research Council, "Frontiers in Attosecond X-ray Science: Imaging and Spectroscopy (AXSIS)", ERC-2013-SyG 609920, and the BMBF through project 05K16GU1.

Author Contributions

JRS and YXW designed experiments; YL and PY provided the RNA samples; JRS and YL crystallized the RNA; JRS, YL, YRB, DRW, CEC, JDC, GN, CL. NAZ, MW, DO, JK, TDG, MSH, SB, ML and YXW collected the SFX data; MW, DO, JK and HNC designed the mixing setup; JRS, YRB, DRW, TAW, AB, RAT, NAZ, XJ and TDG processed and analyzed the SFX data; JRS, YRB, MS, XJ and YXW interpreted the SFX data; NAZ, AB, MSH, SB, ML, UW, PF HNC and JCHS contributed the XFEL expertise and support; YRB, LF, XZ, CDS and YXW collected, analyzed and interpreted SAXS data; JRS, CEC, KT and YXW characterized samples; MD and YL performed mass spectroscopy; SGT and YL performed ITC and fluorescence titration; SP and YL collected and analyzed binding data; XZ and SAW modeled ligand binding kinetics; XZ, SP, SAW and YXW interpreted kinetic data; DED, ADF and JZ contributed

insightful discussions; JRS and YXW drafted the manuscript and all authors contributed to the revision.

Competing financial interests

The authors declare no competing financial interests.

Corresponding author

Correspondence to Yun-Xing Wang and Sarah Woodson

Reference

1 Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* **113**, 577-586 (2003).

2 Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem Biol* **9**, 1043 (2002).

3 Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* **4** (2012).

4 Hajdu, J. *et al.* Analyzing protein functions in four dimensions. *Nat Struct Biol* **7**, 1006-1012 (2000).

5 Boutet, S. *et al.* High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362-364 (2012).

6 Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73-77 (2011).

7 Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* **11**, 1729-1741 (2004).

8 Winkler, W. C. & Breaker, R. R. Genetic control by metabolite-binding riboswitches. *Chembiochem* **4**, 1024-1032 (2003).

9 Batey, R. T. Structure and mechanism of purine-binding riboswitches. *Q Rev Biophys* **45**, 345-381 (2012).

10 Di Palma, F., Colizzi, F. & Bussi, G. Ligand-induced stabilization of the aptamer terminal helix in the add adenine riboswitch. *RNA* **19**, 1517-1524 (2013).

11 Lemay, J. F. & Lafontaine, D. A. [The adenine riboswitch: a new gene regulation mechanism]. *Med Sci (Paris)* **22**, 1053-1059 (2006).

12 Batey, R. T., Gilbert, S. D. & Montange, R. K. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**, 411-415 (2004).

13 Zhang, J. & Ferre-D'Amare, A. R. Dramatic improvement of crystals of large RNAs by cation replacement and dehydration. *Structure* **22**, 1363-1371 (2014).

14 Rieder, R., Lang, K., Graber, D. & Micura, R. Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *Chembiochem* **8**, 896-902 (2007).

15 Gilbert, S. D., Stoddard, C. D., Wise, S. J. & Batey, R. T. Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J Mol Biol* **359**, 754-768 (2006).

16 Delfosse, V. *et al.* Riboswitch structure: an internal residue mimicking the purine ligand. *Nucleic Acids Res* **38**, 2057-2068 (2010).

17 Jenkins, J. L., Krucinska, J., McCarty, R. M., Bandarian, V. & Wedekind, J. E. Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J Biol Chem* **286**, 24626-24637 (2011).

18 Pande, K. *et al.* Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725-729 (2016).

19 Tenboer, J. *et al.* Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242-1246 (2014).

20 Barends, T. R. *et al.* Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350**, 445-450 (2015).

21 Hajdu, J. *et al.* Millisecond X-ray diffraction and the first electron density map from Laue photographs of a protein crystal. *Nature* **329**, 178-181 (1987).
22 Wang, D., Weierstall, U., Pollack, L. & Spence, J. Double-focusing mixing jet for XFEL study of chemical kinetics. *J Synchrotron Radiat* **21**, 1364-1366 (2014).
23 Schmidt, M. Mix and Inject: Reaction Initiation by Diffusion for Time-Resolved Macromolecular Crystallography. *Advances in Condensed Matter Physics* (2013).
24 Wolf-Watz, M. *et al.* Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol* **11**, 945-949 (2004).
25 Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr Opin Struct Biol* **18**, 321-329 (2008).
26 Campbell, J. W. *et al.* Calcium binding sites in tomato bushy stunt virus visualized by Laue crystallography. *J Mol Biol* **214**, 627-632 (1990).
27 Burnley, B. T., Afonine, P. V., Adams, P. D. & Gros, P. Modelling dynamics in protein crystal structures by ensemble refinement. *Elife* **1**, e00311 (2012).
28 Huang, L., Ishibe-Murakami, S., Patel, D. J. & Serganov, A. Long-range pseudoknot interactions dictate the regulatory response in the tetrahydrofolate riboswitch. *Proc Natl Acad Sci U S A* **108**, 14801-14806 (2011).
29 Haller, A., Altman, R. B., Souliere, M. F., Blanchard, S. C. & Micura, R. Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc Natl Acad Sci U S A* **110**, 4188-4193 (2013).
30 Spence, J. & Lattman, E. Imaging enzyme kinetics at atomic resolution. *IUCrJ* **3**, 228-229 (2016).

Figure Captions

Figure 1. Structure comparison of the ligand-bound and apo conformers. a, Secondary structure of rA71 with key regions highlighted in yellow (P1), red (hinge), and green (latch). The three consecutive GC pairs in P1 are illustrated in bold. **b-d**, Cartoon representations in three different views, with key regions color-coded as in (a), comparing the structures of: **b**, apo1 (blue) and ligand-bound (gray, RCSB:4TZX); **c**, apo2 (cyan) and ligand-bound (gray); and **d**, apo1 (blue) and apo2 (cyan). The structures were aligned against the kissing-loops (L2 and L3), residues 29-41 and 58-68.

Figure 2. Structures of the three-way junction in the absence of ligand. a, The three-way junction, as observed in apo1 (blue), apo2 (cyan), and ligand-bound (magenta, RCSB:4TZX) structures, with base triples (yellow, gray, and green) and adenine ligand (black) indicated. **b**, Secondary structures and detailed interactions (shown as ball-and-stick) of key residues (red) within the junctions of each of the apo conformers. In the secondary structure map, base-pairs are depicted as solid black lines, and the movements of residues that flip outward are shown as black arrows. **c,** Trajectories are shown for 5 concentrations of adenine and are globally fit to the model in eq. 1 (Supplementary Discussion). **d,** Evolution of species concentrations over time (Supplementary Discussion).

Figure 3. Setup of mix-and-inject SFX and conversion of the structure and crystal lattice. **a.** Cartoon of the SFX ligand-mixing experiment. For details, see the Methods. **b**, The unit cells of the crystals of apo ($P2_1$) and ligand-bound ($P2_12_12$), whose structure was converted *in crystallo* from the apo structures after >10 min mixing with adenine ligand (Extended Data Fig.

5b&c). **c**, Superimposition (left) of the ligand-bound structures of the 10min-mix (green) and RCSB:4TZX (magenta), and 2Fo-Fc electron density maps contoured at 1σ for the whole structure of 10min-mix (middle) and the binding pocket (right) showing the ligand (black).

Figure 4. Visualizing the ligand-bound intermediate state.

a, Fo-Fc map (3σ, green) of the binding-pocket of apo2 (cyan), calculated after refining the apo-rA71 structure against the 10sec-mix data, with adenine in the binding pocket and U48 removed, revealing altered conformations for U48 (large peak) and A21 (smaller peak). **b**, Superimposition of apo2 (cyan) and IB (yellow) showing the different conformations for U48 and A21. Adenine (red) binding to apo2 (cyan) results in displacement of U48, and consequently, A21 to form IB (yellow). **c**, Superimposition of apo1 (blue) and IB (yellow), revealing very similar base-stacking interactions, in which A23 of apo1 (blue) takes the place of the IB ligand (red). **d**, Three-way junction with adenine ligand (black) of ligand-bound state (RCSB:4TZX), shown for comparison. **e-h**, Ball-and-stick models, with H-bond interactions shown, of key residues in the ligand-binding pockets of apo2 (e), IB (f), apo1 (g), and ligand-bound (h). **i**, $2F_o$-F_c electron density maps contoured at 1σ for apo2 (cyan), IB (yellow), apo1 (blue), and ligand-bound (magenta). **j**, Structure ensembles of apo2 (cyan), IB (yellow), apo1 (blue), and ligand-bound (magenta) from time-averaged molecular dynamics ensemble refinement, demonstrating the flexibility/stability of P1 in each of the four conformations.

Methods

Apo-rA71 Sample Preparation and Characterization

Apo-rA71 was generated by *in vitro* transcription using a solid-state-immobilized DNA template [31] and purified by 12% polyacrylamide gel electrophoresis under denaturing conditions. The RNA was eluted from the gel in buffer containing 50 mM sodium acetate, pH 5.3, 2 mM EDTA, and refolded by heating at 80 °C for 2 min and then incubation at 4 °C overnight. The RNA sample was characterized using high-resolution mass spectrometry. Purified apo-rA71 was buffer-exchanged into RNase-free water, followed by buffer containing 10 mM HEPES, pH 7.5, 100 mM KCl, 0.5 mM EDTA (RNA buffer), and concentrated to 40 g/L. Apo-rA71 was crystallized using the batch method by mixing in 1.5-mL microcentrifuge tubes equal volumes (25 μL each) of RNA solution and crystallization buffer containing 40 mM sodium cacodylate, pH 6.5, 80 mM KCl, 100 mM $MgCl_2$, 12 mM spermine tetrahydrochloride, and 65% v/v (+/-)-2-methyl-2,4-pentanediol (crystallization buffer), followed by incubation at 19-22 °C. Rectangular plate-like microcrystals formed after 1 day at high crystal density and amongst moderate to heavy precipitate. Crystal length typically ranged from 1-10 μm, the majority of which were ~5 μm, with some as large as 20 μm. Microcrystals were characterized visually by bright-field stereomicroscopy with cross-polarization, as well as by UV-TPEF microscopy and second-order nonlinear imaging of chiral crystals (SONICC). Relative diffraction quality of microcrystals, the resolution of which extended to ~6-8 Å, was assessed by powder diffraction using a synchrotron source (Extended Data Fig. 1).

To minimize sample consumption, a viscous stream using an agarose medium was chosen over the conventional liquid-jet sample delivery for SFX [32]. In preparing the sample for SFX data collection, 0.5 mL of microcrystal suspension was centrifuged briefly to form a

concentrated slurry of microcrystals and precipitate, and the pellet was drawn up into a Hamilton syringe. The agarose medium for viscous-injection was prepared by dissolving at 95 °C ultra-low gelling agarose (Sigma) at 9% w/v in buffer containing 40 mM sodium cacodylate, pH 6.5, 80 mM KCl, 100 mM $MgCl_2$, 12 mM spermine tetrahydrochloride, and 37.5% v/v (+/-)-2-methyl-2,4-pentanediol. The agarose medium was drawn up into a separate Hamilton syringe and allowed to cool to room temperature, forming a gel. The volumes of RNA crystals and agarose medium were adjusted to a ratio of 1:1.7. The syringes were attached to a syringe coupler[33] and the microcrystals were embedded in the agarose medium by mixing to homogeneity. All syringes, lines, and devices were purged with RNase-free water and dried prior to contact with sample.

High-resolution Mass Spectrometry

Mass spectrometry data were acquired on an Agilent 6520 Accurate-Mass Q-TOF LC/MS System, (Agilent Technologies, Inc., Santa Clara, CA) equipped with a dual electro-spray source, operated in the negative-ion mode. Separation was performed on Clarity 2.6u Oligo-MS 100A column (2.1 mm x 100 mm). Mobile phase buffer A consisted of 45 mM hexafluoreisopropanol (HFIP), 3 mM triethylamine (TEA) and 10 μM ethylenediaminetetraacetic acid (EDTA) in water at pH 7.3. Mobile phase buffer B consisted of 45 mM HFIP, 3 mM TEA and 10 μM EDTA in 90% methanol/water. The analytes were eluted at a flow rate of 0.2 ml/min with a 10 to 50% organic gradient exchange over 15 minutes. The instrument was used in full-scan TOF mode. MS source parameters were set with a capillary voltage of 4 kV, the fragmentor voltage of 200 V and skimmer 65 V. The gas temperature was 350 °C, drying gas flow 10 l/min and nebulizer pressure 25 psig. Data were acquired at high

resolution (3,200 *m/z*), 4 GHz. TOF-MS mass spectra were recorded across the range 200–3,200 *m*/z. Data acquisition and analysis were performed using MassHunter Workstation Software (version B.07.00). To maintain mass accuracy during the run time, an internal mass calibration sample was infused continuously during the LC/MS runs.

Characterization of Micro- and nanocrystals by Powder Diffraction

The relative quality of riboA71 microcrystals was assessed by room-temperature powder diffraction at the Advanced Photon Source 19ID beamline (Structural Biology Center, Argonne National Laboratory). An aliquot of microcrystal slurry was pelleted in a MicroRT (MiTeGen) capillary, and mounted on a goniometer base. The impact of capillary material was not characterized and may have contributed to background. The sample was exposed to X-rays with a beam size of 0.1 x 0.1 mm^2, a flux of 3x10^14 photons/sec/mm^2 (before an attenuation factor of 8), sample rotation of 0.1 degree, and exposure time of 1 second. Diffraction rings could be seen to a resolution of ~ 8 Å, with individual spots extending to ~ 6 Å (Extended Data Fig.1f).

Apo-rA71 SFX Data Collection and Analysis

Data were collected using the Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source at the SLAC National Accelerator Laboratory. A continuous stream of sample was extruded from a 50 μm capillary into the X-ray interaction region using a viscous injector [34] at a flow-rate of ~300 nL/min. The crystals were exposed to ~9.2 keV XFEL pulses, each with a duration of ~50 fs, at a repetition rate of 120 Hz. The X-rays were focused to 2-3 μm using beryllium lenses, and the transmission was varied from 22 to 100 % to minimize saturation, based on live tracking of Bragg peak intensities. Diffraction snapshots from RNA

microcrystals were recorded on a Cornell-SLAC Pixel Array Detector (CSPAD). A total of 823,198 frames were recorded, amounting to ~1.9 hr of total acquisition time and 3.8 TB of raw data.

Cheetah [35] was used to extract diffraction patterns ("hits"), defined as those images containing a minimum of 15 peaks. Key peak-finding parameters included the 'peak-finder 8' algorithm, a minimum pixel intensity of 60 analog-to-digital units (ADU) and a minimum signal-to-noise of 4.2, which yielded an average hit-rate of 22.4% (184,772 hits). The selected diffraction patterns were indexed by CrystFEL 0.6.0[36-38] using a combination of XDS, MOSFLM, and DirAx. The detector geometry was refined to sub-pixel accuracy using CrystFEL's *geoptimiser*, and the peaks used for indexing were those output from *Cheetah*. Preliminary indexing results strongly suggested a primitive monoclinic lattice, whose unit cell parameters were then enforced, with a tolerance of 8% for cell axes and 2% for cell angles, to a subsequent round of indexing. The final indexing yield was 13.2% (24,411 indexed patterns). The intensities were integrated in CrystFEL using peak centering, optimized integration radii, and a fixed reciprocal-space profile radius of 0.02 nm^{-1}, and were merged using the average-based Monte Carlo method. For this particular data set, we found that scaling or post refinement helped only marginally overall, but worsened the data in the higher resolution shells, as judged by figures of merit, R_{split} and CC*[39]. The data were truncated [40] to 2.29 Å resolution based on the CC* (0.689) of the highest resolution shell (Extended Data Table 1).

Apo-rA71 Sample Preparation for mix-and-inject SFX using XFEL

Apo-rA71 RNA for ligand-mixing SFX was generated by in vitro transcription from PCR template, and purified by phenol/chloroform extraction and ethanol precipitation. The RNA

pellets were resuspended in DEPC-treated H_2O and buffer exchanged into 10 mM HEPES, pH 7.5, 100 mM KCl, 0.5 mM EDTA, and concentrated to 20-40 g/L using a 20-ml Amicon Ultra-4 Centrifugal Filter device (Merck Milipore, CO). The RNA was crystallized by mixing equal volumes of 5 g/L RNA and crystallization buffer and 3% v/v of a concentrated crystal seed stock, followed by constant mixing/rotation for two hours at 19-22 °C. The plate-like microcrystals (1-10 μm in length; ~5 μm on average) were harvested by centrifugation at 1000 g and the supernatant was removed. The crystals were then washed three times with stabilization buffer (containing equal volumes of RNA buffer and crystallization buffer) to remove any RNA in solution, and the final crystal pellet was resuspended in a minimal volume of stabilization buffer.

Mix-and-inject SFX Data Collection and Analysis

The diffusion-triggered experiments were first proposed in 2000 by Hajdu et al., in proposal "First Scientific Experiments for LCLS" (http://www.slac.stanford.edu/pubs/slacreports/reports03/slac-r-611.pdf). The scheme of the mixing device is illustrated in Fig. 3a. The setup included two HPLCs with switchboxes. HPLC1 controlled the delivery of sample (filtered through a 20-micron inline filter) or buffer, and HPLC2 controlled the delivery of adenine-ligand or buffer. The HPLC lines fed into a T-junction (mixer), which was connected to a gas-directed virtual nozzle (1.5 m in length with 75 μm inner diameter capillary), to provide mixing delay times of several seconds, depending on flow rates. For the 10sec-mix data, the data were collected at a combined flow rate of 30 μL/min: 15 μL/min each of sample and ligand (10 mM adenine in stabilization buffer).

To determine the structure of the bound-state rA71 after ligand diffusion and in-crystal conversion, the sample was premixed with an equal volume of 10 mM adenine in the crystallization solution, and loaded into the sample reservoir. Data acquisition commenced at ca. 10 min after mixing, and continued for ca. 30 min with a flow rate of 29 μL/min mixed sample and 1 μL/min ligand solution.

The crystals were exposed to ~9.5 keV XFEL pulses, each with a duration of ~45 fs and a pulse energy of about 10^{11} photons, at a repetition rate of 120 Hz. The X-rays were focused to a spot diameter of about 1 μm using KB mirrors, and the beam was attenuated to minimize saturation, based on live tracking of Bragg peak intensities. For the 10sec-mix data, *Cheetah* hit-finding parameters included a minimum number of 20 peaks, a minimum pixel intensity of 300 ADU and a min SNR of 6. Hit-finding parameters for 10min-mix data were similar, and included 15 peaks, 300 ADU, and min SNR of 7. CrystFEL (version 0.6.2+90ff333) was used for indexing, integration and scaling, with very similar parameters to those used for the apo-rA71 data. Whereas the 10sec-mix crystals were isomorphous with the apo-rA71 crystals, the 10min mixing of ligand into the crystals converted the molecules to the bound state, which was clearly indicated by a change in the space group and unit cell dimensions to that of a previously reported ligand-bound structure. The hit-rate and indexing rates for the 10sec-mix data were 17.0% and 15.8%, respectively, with 14,038 indexed patterns. The hit-rate and indexing rates for the 10min-mix data were 4.1% and 3.8%, respectively. The ligand-bound crystals suffered dramatically from structure conversion, as indicated by diffraction quality and a very low indexing rate. This, and limited available beam time of six hours for the entire set of measurements, resulted in a final data set of only 378 indexed patterns (see Extended Table 1).

Structure Solution and Refinement

After numerous attempts using various search models derived from the ligand-bound form of the adenine riboswitch aptamer domain (RCSB code 4TZX), the structure of apo-rA71 was solved by PHASER [41] molecular replacement using a truncated version of 4TZX in which residues 13-18, 43-53, and 78-83 were removed. The 10sec-mix structure was solved by refining the apo-rA71 model against the 10sec-mix data, while conserving the test set of reflections from the apo-rA71 data. Fo-Fc and Fo-Fo difference electron density maps calculated using the apo-rA71 model revealed variation between the two data sets in the form of isolated peaks, most likely corresponding to backbone phosphates, predominantly in the apo2 molecule (Extended Data Figs. 6a and 7a). This indicated a mixture of at least two conformers. The occupancies of U48 and A21 of apo2 were set to 0.5 and refined in the same manner. The Fo-Fc map (Extended Data Fig. 6b) clearly indicated the alternate (IB) conformation of U48 with a blob of density in the original U48 position corresponding to the adenine ligand. The alternate conformation of A21 was much less pronounced and is partially disordered along with the adjacent hinge residues (U22 and A23). Keeping the occupancy of A21 at 0.5, the structure was refined with U48 omitted. The Fo-Fc map (Extended Data Fig. 6c) again strongly supported the alternate configuration of U48, A21, and density for the adenine ligand. The positions of the ligand, U48, and A21 are reminiscent of the apo1 molecule, in which A23 is in place of the ligand, forming similar stacking and H-bond interactions. The final model for the 10sec-mix data contained adenine ligand and two alternate conformations for both U48 and A21, all with occupancies of 0.5 (Extended Data Fig. 6d).

The 10min-mix converted bound structure was solved by PHASER[42] MR using RCSB:4TZX as a search model, and was refined using Refmac5[43]. Despite such a small SFX

dataset (378 indexed patterns), the structure was solved to 3.0 Å resolution by PHASER MR with surprisingly good electron density for the entire structure, including the adenine ligand, and the flipped-out U48, a signature of the bound state (Fig. 3b). The structure refinement, however, reflects higher-than-normal R-factors due to such low multiplicity in the SFX data. All crystal data and refinement statistics are summarized in Extended Data Table 1. Final models were obtained using a combination of ERRASER [44] for automated model building, manual placement of residues in COOT [45], and iterative cycles of restrained refinement in PHENIX [42] and Refmac5[43] with TLS (apo and IB only). The structures were deposited in the RCSB PDB under accession codes 5E54 (apo-rA71), 5SWD (IB), and 5SWE (B). All figures displaying structural coordinates were generated using PyMOL[46].

Ensemble Refinement Calculations

Time-averaged molecular dynamics ensemble refinement was performed using PHENIX[27,42] (version dev-2447) with default parameters, including 0.8 for the fraction of atoms included in TLS fitting, and a temperature of 300 K with 10 steps in the molecular dynamics simulation. The results for each were an ensemble of 20 models. Ensemble refinement was not used for the final model calculations, but rather as a qualitative assessment to demonstrate the varying structural flexibility of the P1 helix in each of the rA71 structures.

Simulation of the Time-course of IB Accumulation *in crystallo*

Assuming only apo2 is converted to IB in the crystal, the binding reaction in crystals can be approximately described by a three-state model. Changing from IB to final binding state

B•ade in the crystal is expected to be slow, with k_f assumed to be 0.01 s^{-1} and k_r= 0 s^{-1} (no dissociation); k_{on} and k_{off} to be 0.1 $\mu M^{-1}S^{-1}$ and 37 s^{-1} respectively (Supplementary Discussion).

$$\text{apo + ade} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{IB*•ade} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \text{B•ade}$$

Due to the slow k_f, a significant population of IB*•ade may accumulate within a certain time window. Given the two time points where 50 and 100% of RNA are converted at t=10 sec and 10 min., respectively, and the initial bulk ade concentration ~5 mM and initial in-crystal concentration of apo2 ~15 mM, the time course of the IB*•ade buildup is simulated to rationalize the SFX results (Extended Data Fig. 5a).

SAXS Data Collection and Analysis

The procedure for SAXS data collection, processing and analysis are similar to that previously described[47] using an in-house program package NCI-SAXS or a program package by Svergun and co-workers (http://www.embl-hamburg.de/biosaxs/). These procedures and protocols for data collection and processing are in full conformity with the recently published recommended standards[48]. X-ray scattering measurements were carried out at room temperature at the beamlines 12-ID-B of the Advanced Photon Source, the Argonne National Laboratory. The setups were adjusted to achieve scattering q values of $0.006 < q < 0.72$ $Å^{-1}$, where $q=(4\pi/\lambda)\sin\theta$, and 2θ is the scattering angle. Twenty two-dimensional (2D) images were recorded for each buffer or sample solution using a flow cell, with the exposure time of 0.5-1 seconds to minimize radiation damage and obtain good signal-to-noise ratio. No radiation damage was observed as confirmed by the absence of systematic signal changes in sequentially collected X-ray scattering images and also confirmed later by gel electrophoresis. The 2D images were reduced to one-dimensional scattering profiles using the MATLAB software package at the beamlines. Scattering profiles of the RNAs were calculated by subtracting the background

buffer contribution from the sample-buffer profile. The experimental radius of gyration (R_g) was calculated from the data at low q values in the range of $qR_g < 1.3$, using the Guinier approximation of $\ln I(q) \approx \ln(I(0)) - R_g^2 q^2/3$.

The solution X-ray scattering curves (Extended Data Fig. 2) were back-calculated using program SolX[49,50], and atomic volume parameters for nucleic acids[51-53]. The same volume parameters were also used in the ensemble calculations[54,55]. The difference between the back-calculated curve and the experimental data after q=0.2 $Å^{-1}$ may arise from the flexibility of the molecule in solution, and imperfect simulations on the solvation layer and ion cloud around the molecule.

SAXS Ensemble Calculations

The Kratky plots of the experimental SAXS data of the aptamer domain of the adenine riboswitch RNA in both the bound and apo states show asymmetric Bell-shape curves (Extended Data Fig. 2c), characteristic of a folded structure with partial flexibility. As discussed in the main text, the holistic view of the apo-rA71 is that flexibility is most likely associated with the hinge and latch regions, and the P1 helix, whereas the kissing loops together with the top parts of the P2 and P3 are relatively rigid. The detailed procedure for the ensemble calculations was carried out using Xplor-NIH as described previously[55], with exception that the ensemble weight optimization facility[56] was used in this work. During the calculations, the P2 and P3 helices were treated as rigid bodies, allowing them to rotate and translate freely during simulated annealing. For the P1 helix, the upper part of the helix, A17-U79, G18-C78, A19-U77, U20-A76 and A21-U73, was set as rigid, whereas the residues G13, G14 and G16 and A17 on the 5' end of P1 were set semi-free, and residues C83, C82, C81 and U80 at the 3' end in P1 were allowed to rotate and translate freely, restrained only by covalent linkage. Such restraints were to reflect the "uncertainty" and account for the difference in occupancy and B-factors of the residues in P1. The force constant for SAXS was set at 100 kcal/mol, which was ramped so that the calculation could sample conformation space freely without being trapped in a local minimum. The annealing temperature was set

at 1,000 *K*, which was high enough to enable segments of duplexes to freely translate and rotate, but not so high as to disrupt basepairs in duplex regions.

Stopped-flow Fluorescence

Riboswitch RNA containing 2-aminopurine (2AP) at position 48 in the ligand binding pocket (Dharmacon) was gel-purified and deprotected following the manufacturer's protocol. The final material had the expected molecular mass (23,000) by mass spectrometry. The kinetics of structural changes due to adenine binding in SF buffer (10 mM HEPES, pH 7.5, 100 mM KCl, 10 mM $MgCl_2$) at 25 °C was measured using an Applied Photophysics SX 18MV stopped-flow spectrometer. The samples were excited at 318 nm and 2AP fluorescence emission measured using a 335 nm longpass filter. Unless otherwise stated, 0.5 μM 2AP-RNA and 0.5 – 1600 μM adenine were mixed with a dead time of 1.8 ms. Six to eight trajectories per sample were averaged and the background intensity was subtracted from each progress curve. The adenine binding kinetics in 5-800 μM adenine was globally fit to the four-state model in eq. 1 using MATLAB (MathWorks, MA). See Supplementary Discussion for further details.

Data availability statement

The coordinates for apo, IB, and B, are deposited in the Protein Data Bank under accession codes 5E54, 5SWD, and 5SWE, respectively.

Reference

31. Liu, Y. *et al.* Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature* **522**, 368-372 (2015).
32 Conrad, C. E. *et al.* A novel inert crystal delivery medium for serial femtosecond crystallography. *IUCrJ* **2**, 421-430 (2015).
33 Cheng, A., Hummel, B., Qiu, H. & Caffrey, M. A simple mechanical mixer for small viscous lipid-containing samples. *Chem Phys Lipids* **95**, 11-21 (1998).
34 Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat Commun* **5**, 3309 (2014).
35 Barty, A. *et al. Cheetah*: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J Appl Crystallogr* **47**, 1118-1131 (2014).
36 White, T. A., Kirian, R.A., Martin, A.V., Aquila, A., Nass, K., Barty, A., Chapman, H.N. CrystFEL: a software suite for snapshot serial crystallography. *J Appl Crystallogr* **45** (2012).
37 Kirian, R. A. *et al.* Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals. *Acta Crystallogr A* **67**, 131-140 (2011).
38 White, T. A. *et al.* Crystallographic data processing for free-electron laser sources. *Acta Crystallogr D Biol Crystallogr* **69**, 1231-1240 (2013).
39 Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).
40 French, G. S. & Wilson, K. S. On the treatment of negative intensity observations. *Acta Cryst.* **A34**, 517-525 (1978).
41 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
42 Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-1954 (2002).
43 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367 (2011).
44 Chou, F. C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat Methods* **10**, 74-76 (2013).
45 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
46 The PyMOL Molecular Graphics System, Version 1.7.6.4 Schrödinger, LLC.
47 Wang, J. *et al.* A method for helical RNA global structure determination in solution using small-angle x-ray scattering and NMR measurements. *J Mol Biol* **393**, 717-734 (2009).
48 Jacques, D. A., Guss, J. M. & Trewhella, J. Reliable structural interpretation of small-angle scattering data from bio-molecules in solution--the importance of quality control and a standard reporting framework. *BMC Struct Biol* **12**, 9 (2012).
49 Zuo, X. *et al.* X-ray diffraction "fingerprinting" of DNA structure in solution for quantitative evaluation of molecular dynamics simulation. *Proc Natl Acad Sci U S A* **103**, 3534-3539 (2006).

50 Tiede, D. M., Mardis, K. L. & Zuo, X. X-ray scattering combined with coordinate-based analyses for applications in natural and artificial photosynthesis. *Photosynth Res* **102**, 267-279 (2009).
51 Zuo, X. *et al.* Global molecular structure and interfaces: refining an RNA:RNA complex structure using solution X-ray scattering data. *J Am Chem Soc* **130**, 3292-3293 (2008).
52 Nadassy, K., Tomas-Oliveira, I., Alberts, I., Janin, J. & Wodak, S. J. Standard atomic volumes in double-stranded DNA and packing in protein--DNA interfaces. *Nucleic Acids Res* **29**, 3362-3376 (2001).
53 Voss, N. R. & Gerstein, M. Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J Mol Biol* **346**, 477-492 (2005).
54 Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**, 65-73 (2003).
55 Fang, X. *et al.* An unusual topological structure of the HIV-1 Rev response element. *Cell* **155**, 594-605 (2013).
56 Deshmukh, L. *et al.* Structure and dynamics of full-length HIV-1 capsid protein in solution. *J Am Chem Soc* **135**, 16133-16147 (2013).

Extended Data Figure Captions

Extended Data Figure 1: Characterization of RNA crystals.

Microcrystals of apo-rA71 were grown using batch crystallization as described in Methods. A SONICC Imager (Formulatrix) was used to image each sample (0.5-1 μL) of crystals using three different methods: **a**, visible light; **b**, UV-TPEF; and (**c**), second-order nonlinear imaging of chiral crystals (SONICC). **d**, Crystal samples were observed using a stereomicroscope (Zeiss) under cross-polarized light. **e**, Without cross-polarization, crystals were barely observable; **f**, The relative quality of the crystalline samples was measured by powder X-ray diffraction (APS beamline 19-ID), with a maximum observable resolution of ~6 Å. The resolution ring (red) corresponds to 6.8 Å.

Extended Data Figure 2. Characterization of rA71 in solution by small angle X-ray scattering.

Comparison of the **a**, Kratky plots of the solution X-ray scattering curves of rA71 in the apo (black) and bound (red) states. **b**, Plot of back-calculated SAXS profiles of apo1 and apo2 conformers along with solution X-ray scattering curves of rA71 in the apo (red) states. **c,** Experimental SAXS curve shown in red with red error bars and superimposed with the SAXS curves that were back-calculated from 128 structures using the two-member ensemble calculation. The ratio of the two conformers (apo2: apo1) is approximately 0.9:0.1 to give the best fit to the experiment data with χ^2 values ranging from ~0.2 +/- 0.1.

Extended Data Figure 3. Three-way junction undergoes dramatic structural rearrangements to accommodate ligand, compressing the major groove.

a, The three-way junction, depicted in three orientations, as observed in apo1 (blue), apo2 (cyan,), and ligand-bound (magenta, RCSB:4TZX) structures. Virtually all residues in the three-way junction undergo significant conformational changes upon ligand binding. Most notable are the "swinging" residues in the hinge (U22, A23) and latch (U48, U49, U51) regions, whose atomic positions differ by as much as 17 Å in the apo conformers relative to the ligand-bound conformer. **b**, In the absence of ligand, concerted movement of the hinge (depicted as white surface and stick model) and latch regions results in significant narrowing of the major groove formed between helices P1 and P3, which measures 9.4 Å, 10.0 Å, , and 16.6 Å for apo1, apo2, and ligand-bound conformers, respectively. Major groove distances were measured between the phosphorous atoms of U71 and A19 (or U20 in the case of apo1 due to a difference in register).

Extended Data Figure 4. **Multistage ligand binding kinetics to the rA71 riboswitch.**

Ligand binding to the rA71 riboswitch was monitored by replacing U48 with the fluorescent base analog 2AP. The 2AP fluorescence emission intensity increases upon ligand-induced reorganization of the binding pocket. The 2AP substitution does not change the secondary structure of the apo rA71 riboswitch, as judged by partial ribonuclease digestion. **a,** Equilibrium titration of rA71-U482AP with adenine yields K_d = 5 μM (black symbols). A similar K_d was obtained from the endpoints of the kinetics progress curves (red symbols) and from in-line probing experiments. This value is about 10-fold higher than adenine binding to the unmodified riboswitch, possibly because 2AP forms more stable base stacking interactions in the apo structure. Error bars show the standard deviation from the average of two or more independent

trials. **b.** The ligand binding kinetics is consistent with a four-state mechanism. Binding of 0.5-1600 μM adenine to 0.5 μM rA71-U482AP was measured by stopped-flow fluorescence (1.8 ms deadtime) as described in SI Methods. The apparent rate constants for adenine association, λ_{fast} and λ_{slow}, were obtained from fits of individual trajectories to a biphasic rate equation, $\Delta F = A_{fast} (1 - \exp(-\lambda_{fast}t)) + A_{slow} (1 - \exp(-\lambda_{slow}t))$. Error bars as in (**a**). The non-linear increase in λ_{fast} (filled symbols) with adenine concentration over the full range of ligand concentrations indicates the presence of one or intermediates in the binding mechanism. The ligand-independent phase, λ_{slow} (open symbols), results in biphasic trajectories above 50 μM adenine and is explained by slow exchange between binding competent and binding incompetent forms of the riboswitch. **c**. The apparent bimolecular rate constant for adenine association is slower than diffusion and was obtained from λ_{fast} versus [adenine], under pseudo-first order conditions (0.5 – 25 μM adenine) in which ligand binding to the competent (open) riboswitch is rate-limiting. In 10 mM $MgCl_2$ (filled symbols), $k_{on} = 1.9 \bullet 10^5$ $M^{-1}s^{-1}$ and $k_{off} = 1.7$ s^{-1}. In 1.25 mM $MgCl_2$ (open symbols), $k_{on} = 5.2 \bullet 10^4$ $M^{-1}s^{-1}$ and $k_{off} = 2.1$ s^{-1}. **d,e**, The same set of reduced experimental data in Fig. 2c was globally fit to simpler three-state kinetic mechanisms as described in the Supplementary Discussion. These models were not able to describe the solution binding kinetics over the full-range of ligand concentrations tested. Therefore, the four-state model in eq. 1 is the simplest mechanism capable of describing the data. We do not exclude the possibility that the riboswitch samples additional apo states and intermediate complexes that may contribute to the robustness of the switch mechanism. **d.** Three-state mechanism with only one apo state. The parameters obtained from the fitting are: $k_{on} = 0.28$ $\mu M^{-1}s^{-1}$, $k_{off} = 37$ s^{-1}, $k_f = 103$ s^{-1}, $k_r = 5.1$ s^{-1}, $sc = 2.03$. Err(k,sc) = 0.053. **e.** Three-state mechanism with two apo states and no binding intermediate.

Parameters: k_{op} = 2.5 s^{-1}, k_{cl} = 0.58 s^{-1}, k_{on} = 0.16 $\mu M^{-1}s^{-1}$, k_{off} = 0.79 s^{-1}, *sc* = 2.44. Err(k,sc) = 0.056.

Extended Data Figure 5. Time-course simulation of the IB concentration in the crystal and comparison of the unit cell dimensions of the apo, IB, and B structures.

a, Simulated time-courses of the IB buildup and changes in concentrations of ligand, apo2, and bound states in the crystal. See also Methods. **b-c**, Space group and unit-cell dimensions of the crystals of apo, IB, and B. The structure was converted in the crystal from apo to B after >10 min mixing with adenine ligand. The crystal lattice remains unchanged after 10 seconds of mixing with ligand.

Extended Data Figure 6. Elucidating the structure of IB.

a, To first verify whether there were changes in IB relative to apo2, the structure of apo2 was refined against the 10sec-mix data; 2Fo-Fc (1σ, blue) and Fo-Fc (3σ, green) electron density maps are shown. Both maps indicated alternate positions for both U48 and A21. **b**, The occupancies of U48 and A21 of apo2 were set to 0.5 and refined in the same manner. The Fo-Fc map (3σ, green) clearly indicated the alternate (IB) conformation of U48 with a blob of density in the original U48 position corresponding to the adenine ligand. The alternate conformation of A21 was much less pronounced and is partially disordered along with the adjacent hinge residues (U22 and A23). **c**, Keeping the occupancy of A21 at 0.5, the structure was refined with U48 omitted. The Fo-Fc map (3σ, green) again supports the alternate configuration of U48, A21, and density for the adenine ligand. **d**, The final refined structure of IB with the adenine ligand (red,

and alternate conformations for both U48 and A21 modeled at 0.5 occupancy, shown with 2Fo-Fc electron density map (1σ, blue).

Extended Data Figure 7. Fo-Fc and 2Fo-Fc electron density maps, and ensemble refinement model for the 10sec-mix data.

a, Weighted Fo-Fc difference electron density maps in green (3σ) and red (-3σ), computed using the apo structure (apo2, cyan; apo1, blue) and the structure factors from the 10sec-mix data. The isolated peaks, most likely corresponding to backbone phosphates, indicate a mixture of conformational states, and are predominantly in and around the three-way junction of apo2. **b**, 2Fo-Fc electron density map (1σ, left) and time-averaged molecular dynamics ensemble refinement model (right) for the "apo1-like" molecule of the 10sec-mix structure. **c**, Superimposition of apo1 (blue) and the apo1-like molecule of the 10sec-mix structure (orange), indicating no structural changes to apo1 after 10 seconds of mixing with ligand.

Extended Data Table 1: Crystal Data and Refinement Statistics

Crystal Data	**Apo (5E54)** [a]	**IB (5SWD)**	**B (5SWE)**
Space group	$P2_1$	$P2_1$	$P2_12_12$
Unit cell parameters			
a, *b*, *c* (Å)	47.9, 46.7, 92.3	47.9, 46.7, 92.3	49.9, 154.9, 25.2
b (°)	94.1	94.1	90
Matthews coefficient (Å^3/Da)	2.33	2.33	2.21
Resolution (Å)	25.02–2.30 (2.38–2.30)[b]	31.06-2.50 (2.59-2.50)	26.32-3.00 (3.31-3.00)
Data acquisition time (min)	114	72	33
Hit rate / Indexing rate (%)	22.4 / 13.2	17.0 / 15.8	4.1 / 3.8
No. indexed patterns	24,411	14,038	378
No. unique reflections	18,368 (1,823)	14,357 (1,431)	4115 (945)
Completeness	1.00 (1.00)	1.00 (1.00)	0.95 (0.87)
R_{split} (%) [c]	10.92 (81.51)	14.10 (86.84)	66.89 (101.57)
SNR	4.37 (0.27)	3.49 (0.31)	1.37 (0.30)
Multiplicity	461.3 (314.6)	346.8 (188.3)	20.3 (13.0)
CC^* [d]	0.998 (0.689)	0.996 (0.462)	0.833 (0.210)
Structure Refinement			
Resolution range (Å)	25.02–2.30 (2.36–2.30)	30.92-2.50 (2.57-2.50)	26.32-3.00 (3.08-3.00)
No. reflections used	16,995 (972)	13,212 (735)	3,919 (230)
Completeness for range	0.98 (0.79)	0.98 (0.74)	0.95 (0.76)
R_{work} [e]	0.213 (0.640)	0.195 (0.548)	0.349 (0.421)
R_{free} [f]	0.256 (0.660)	0.229 (0.735)	0.379 (0.407)
Wilson B (Å^2)	70.9	80.9	66.4
Average B (Å^2)	77.2	90.6	58.1
No. atoms	2,844	2896	1512
RNA	2,826	2868	1512
Magnesium	4	3	0
Water	14	15	0
Ligand	-	10	10
R.M.S. deviations			
Bond lengths (Å)	0.007	0.007	0.003
Bond angles (°)	1.62	1.52	0.856
Coordinate error (DPI) (Å)	0.24	0.26	0.70

[a] PDB accession code

[b] Values in parentheses are for the highest-resolution shell.

[c] $R_{\text{split}} = \frac{1}{2^{1/2}} \frac{\Sigma \left| I_{even} - I_{odd} \right|}{\frac{1}{2} \Sigma \left(I_{even} + I_{odd} \right)}$

[d] $CC^* = [2CC_{1/2} / (1 + CC_{1/2})]^{1/2}$

[e] $R_{\text{work}} = \Sigma_{\text{hkl}} \left| \, |F_o| - |F_c| \, \right| / \Sigma_{\text{hkl}} |F_o|$, calculated from the working data set

[f] R_{free} is calculated from ~5 % of data randomly chosen and not included in refinement.

Extended Data Table 2:

a, Root-mean-square deviations (Å) of all atoms among apo and ligand-bound conformers with structures aligned with the kissing loops

	3-way jct.	P1 [a]	U48	U51	U74	Hinge	Latch	Kissing loops
apo2 vs. 4TZX	6.4	6.5	13.3	10.1	2.8	5.9	7.5	0.8
apo1 vs. 4TZX	6.2	10.3	11.5	10.0	2.8	7.8	6.8	0.8
apo1 vs. Apo2	3.4	6.4	2.8	4.7	1.7	7.2	2.0	0.5
apo1 vs. IB	3.5	5.0	2.3	4.8	1.6	6.4	2.0	0.8
apo2 vs. IB	1.3	1.4	2.4	0.4	0.6	2.5	0.8	0.3
IB vs. 4TZX	5.6	6.6	12.2	9.9	3.2	5.4	6.5	0.89

b, Root-mean-square deviations (Å) of all atoms among apo and ligand-bound conformers with whole structure alignments

	3-way jct.	P1 [a]	U48	U51	U74	Hinge	Latch	Kissing loops
apo2 vs. 4TZX	5.6	2.5	10.6	10.0	2.3	1.6	4.9	0.9
apo1 vs. 4TZX	5.6	3.4	7.9	10.6	2.3	1.7	4.8	0.8
apo1 vs. Apo2	3.6	1.5	2.6	2.0	1.6	2.4	1.4	0.3
apo1 vs. IB	3.3	3.8	2.3	3.9	1.8	6.2	1.8	1.1
apo2 vs. IB	1.3	1.3	2.4	0.3	0.4	0.3	0.8	0.3
IB vs. 4TZX	5.5	4.1	10.7	10.9	2.4	5.1	6.4	1.4

[a] The RMSDs of P1 when compared alone are among the three structures:
1.3 Å (apo2 vs. bound)
0.9 Å (apo1 vs. bound)
1.0 Å (apo2 vs. apo1)
1.4 Å (IB vs. bound)
1.3 Å (IB vs. apo1)
1.3 Å (IB vs. apo2)

Extended Data Table 3: Average B-factors (Å^2)

	Overall	**P1** [b]
Apo1 (13-77) [a]	76.7	122.9
Apo2 (13-80)	77.1	131.0
IB (13-77)	92.7	158.9
IB-Apo1 (13-80)	88.1	137.1
Bound (13-83)	57.9	69.1

[a] Numbers in parentheses indicate the range of residues modelled.
[b] Residues 13-21 and 75-83.

FIGURE 1

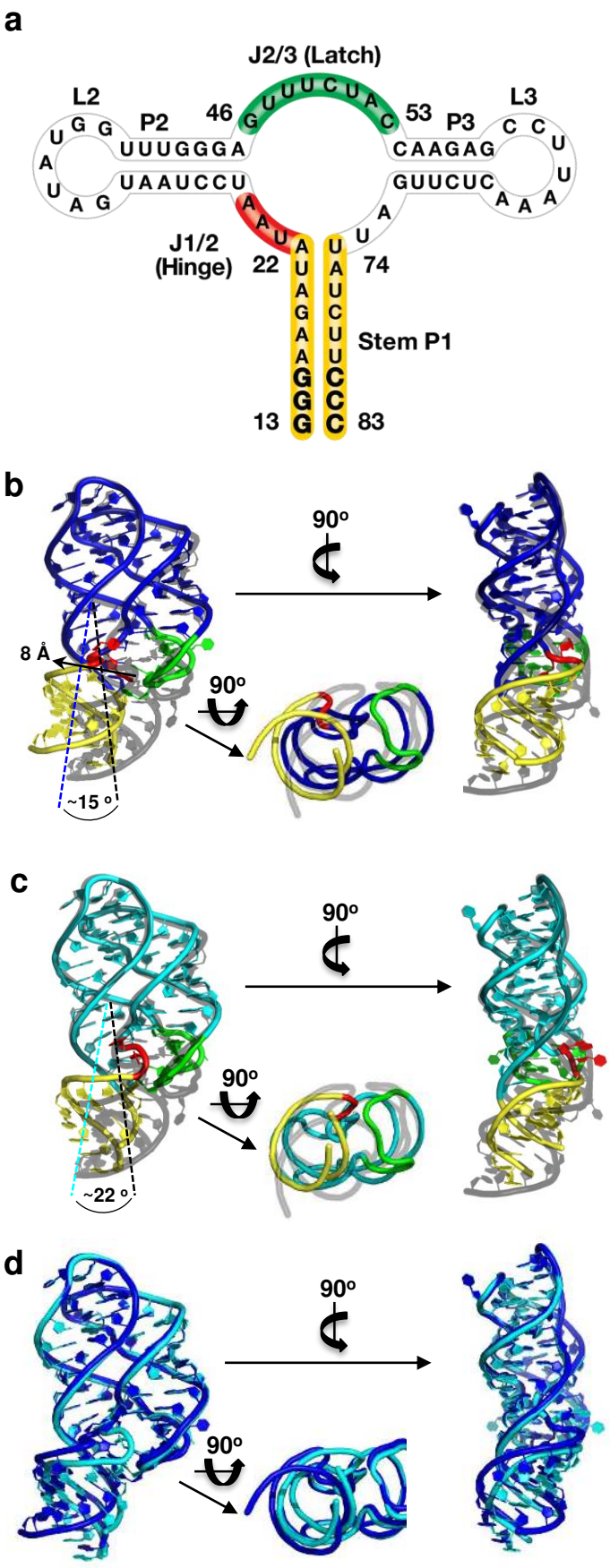

a
J2/3 (Latch)
L2
P2
46
53
P3
L3
J1/2
(Hinge)
22
74
Stem P1
13
83
b
90°
8 Å
90°
~15 °
c
90°
90°
~22 °
d
90°
90°

FIGURE 2

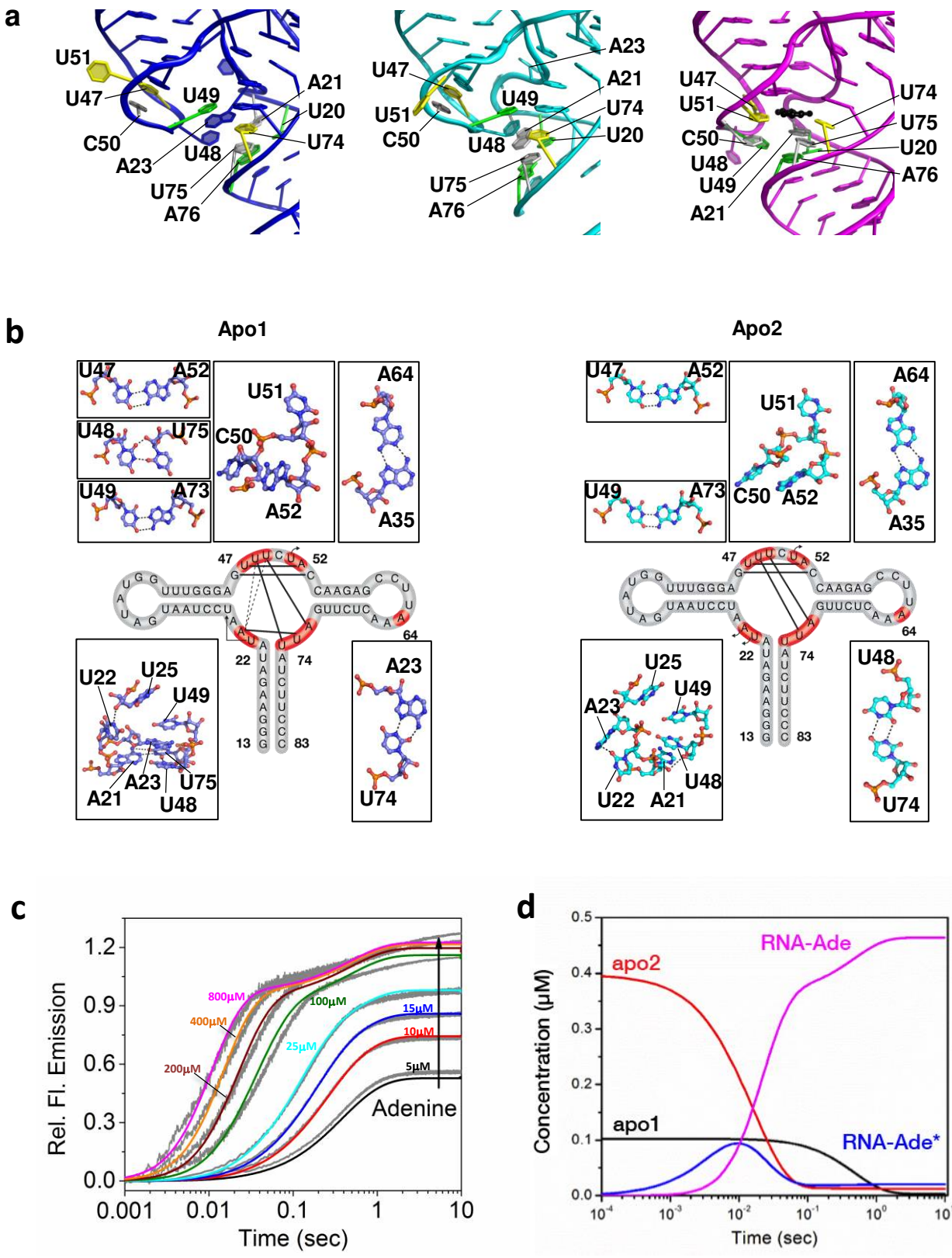
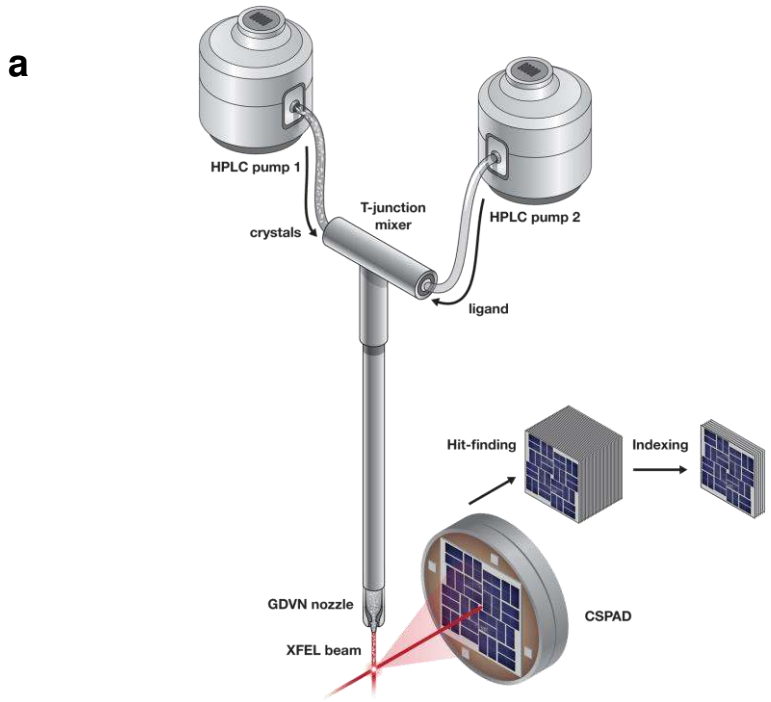
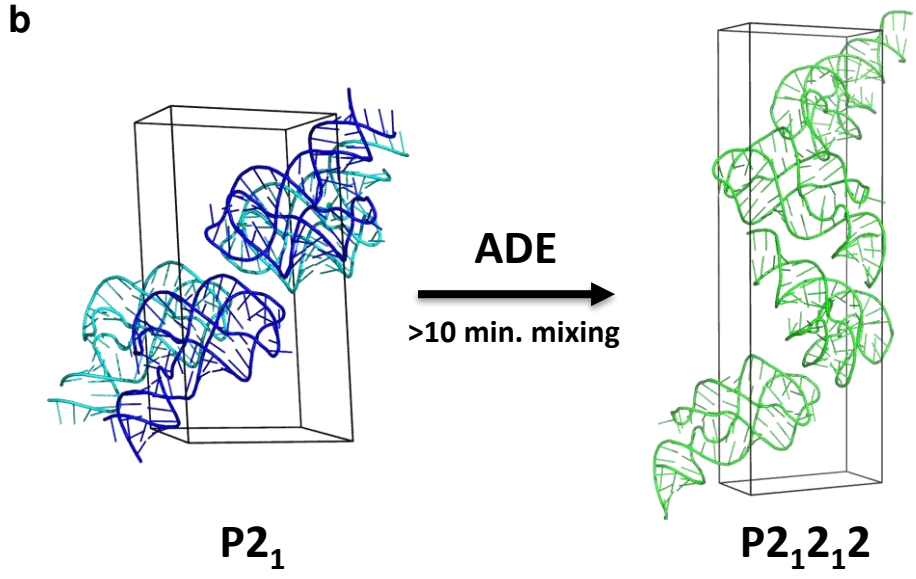

a
U51
U47
U49
A21
U20
C50
U48
U74
A23
U75
A76
U47
A23
A21
U49
U51
U74
C50
U48
U20
U75
A76
U47
U74
U51
U75
C50
U20
U48
A76
U49
A21
b
Apo1
Apo2
U47
A52
U48
U75
U49
A73
U51
C50
A52
A64
A35
U22
U25
U49
A21
A23
U75
U48
A23
U74
47
52
64
22
74
13
83
UUUGGGA
UAAUCCU
CAAGAG
GUUCUC
AUAGAAGGG
AUCUUUCCC
c
Rel. Fl. Emission
Time (sec)
800μM
400μM
200μM
100μM
25μM
15μM
10μM
5μM
Adenine
d
Concentration (μM)
Time (sec)
apo2
RNA-Ade
apo1
RNA-Ade*

FIGURE 3


a
HPLC pump 1
crystals
T-junction mixer
HPLC pump 2
ligand
Hit-finding
Indexing
GDVN nozzle
XFEL beam
CSPAD


b
ADE
>10 min. mixing
P2$_1$
P2$_1$2$_1$2

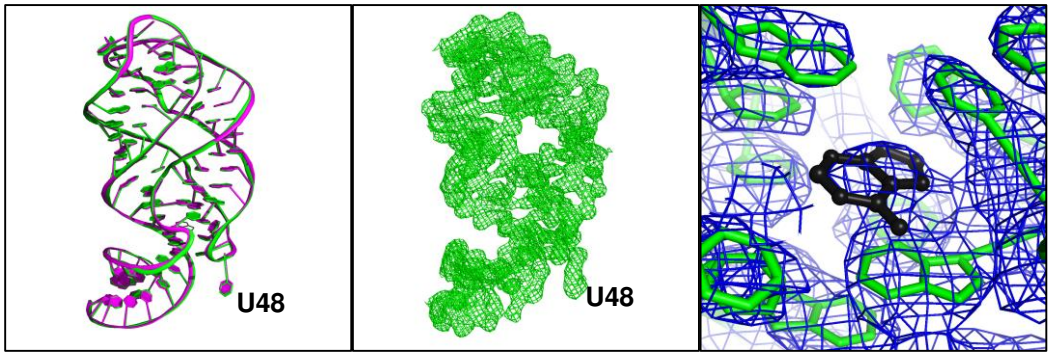
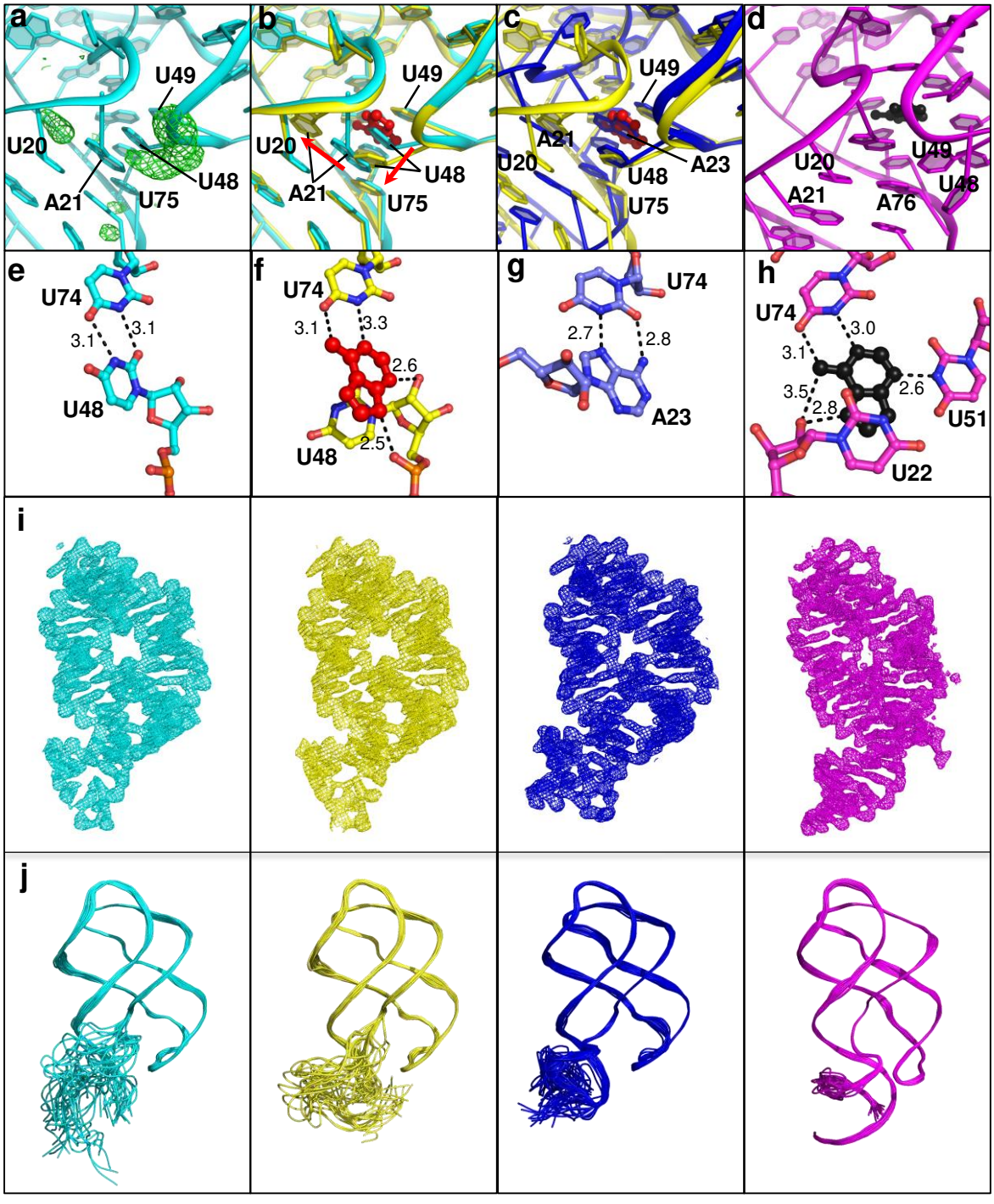
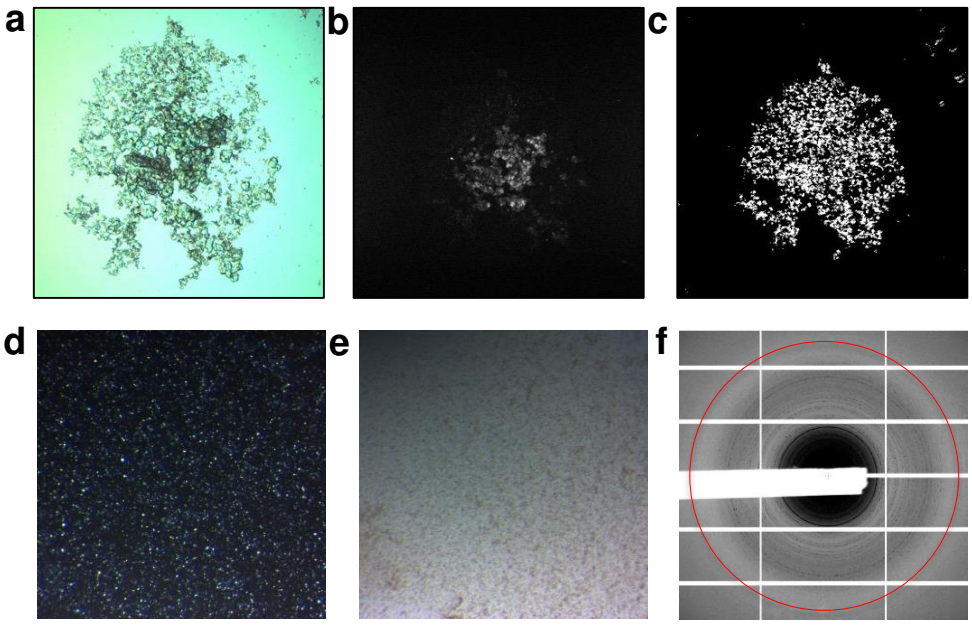

c
U48
U48

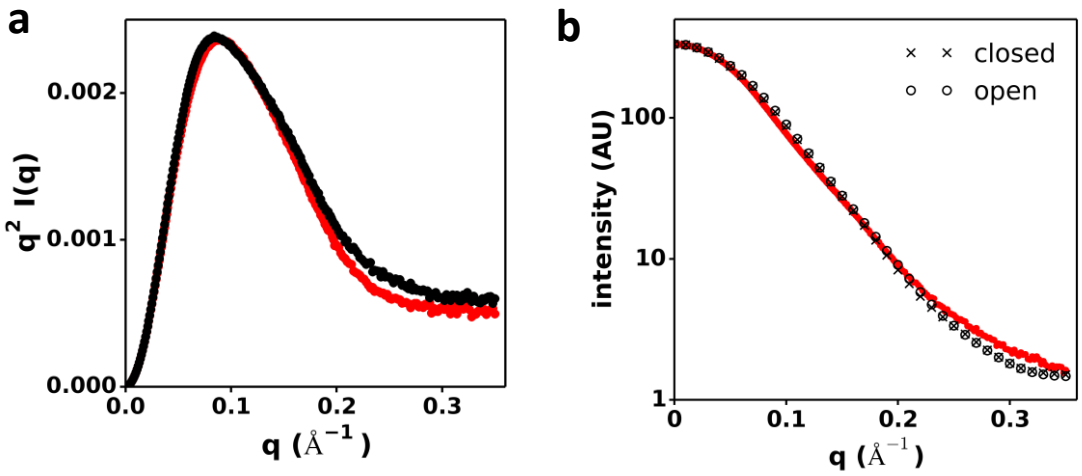
FIGURE 4

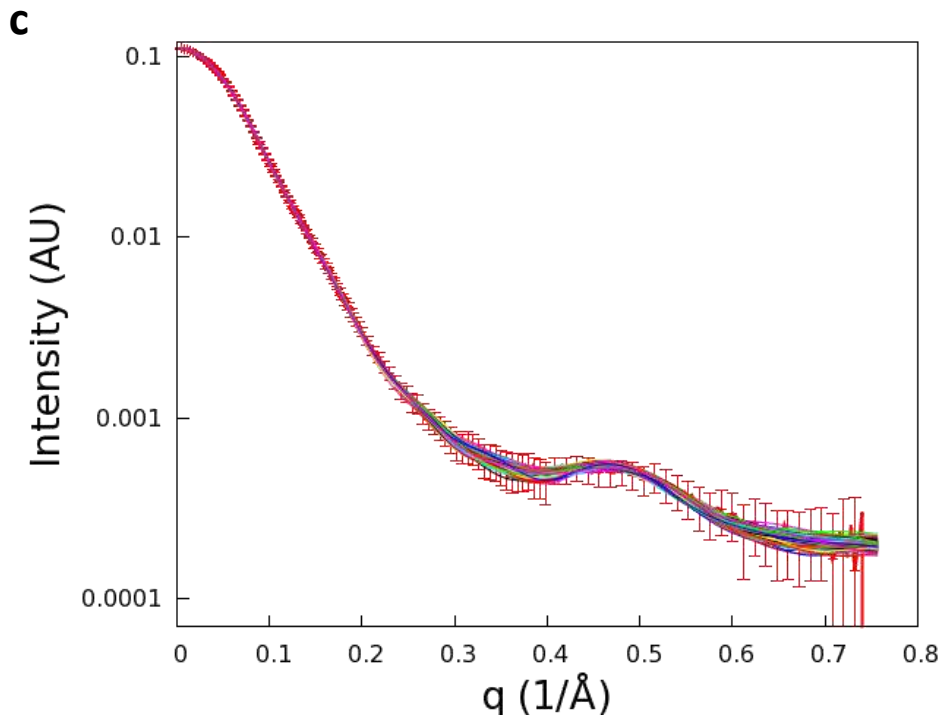

a
U49
U20
U48
A21
U75
b
U49
U20
U48
A21
U75
c
U49
A21
U20
A23
U48
U75
d
U20
U49
A21
A76
U48
e
U74
3.1
3.1
U48
f
U74
3.1
3.3
2.6
U48
2.5
g
U74
2.7
2.8
A23
h
U74
3.0
3.1
3.5
2.8
2.6
U51
U22
i
j

Extended Data Figure 1

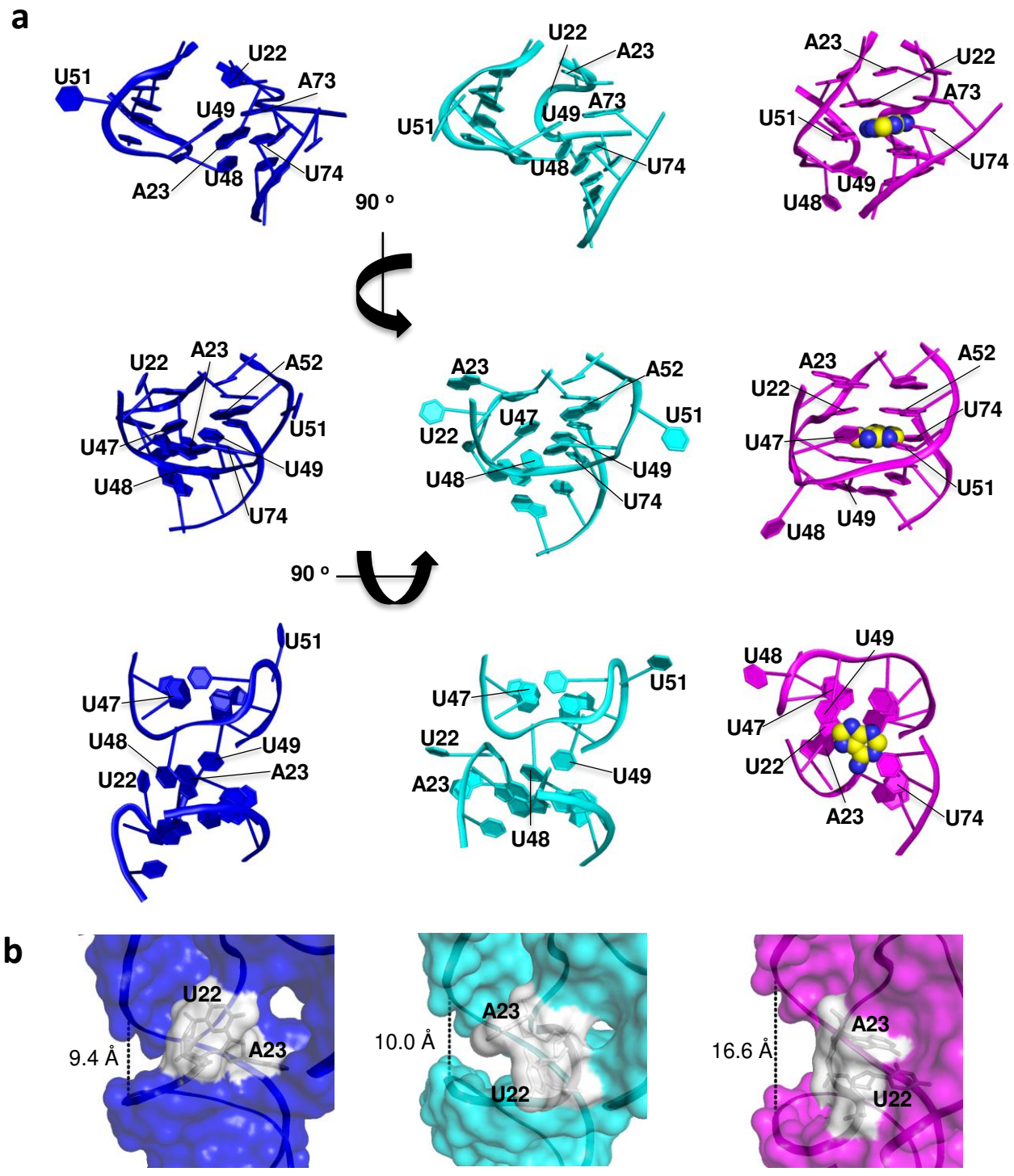


Extended Data Figure 2

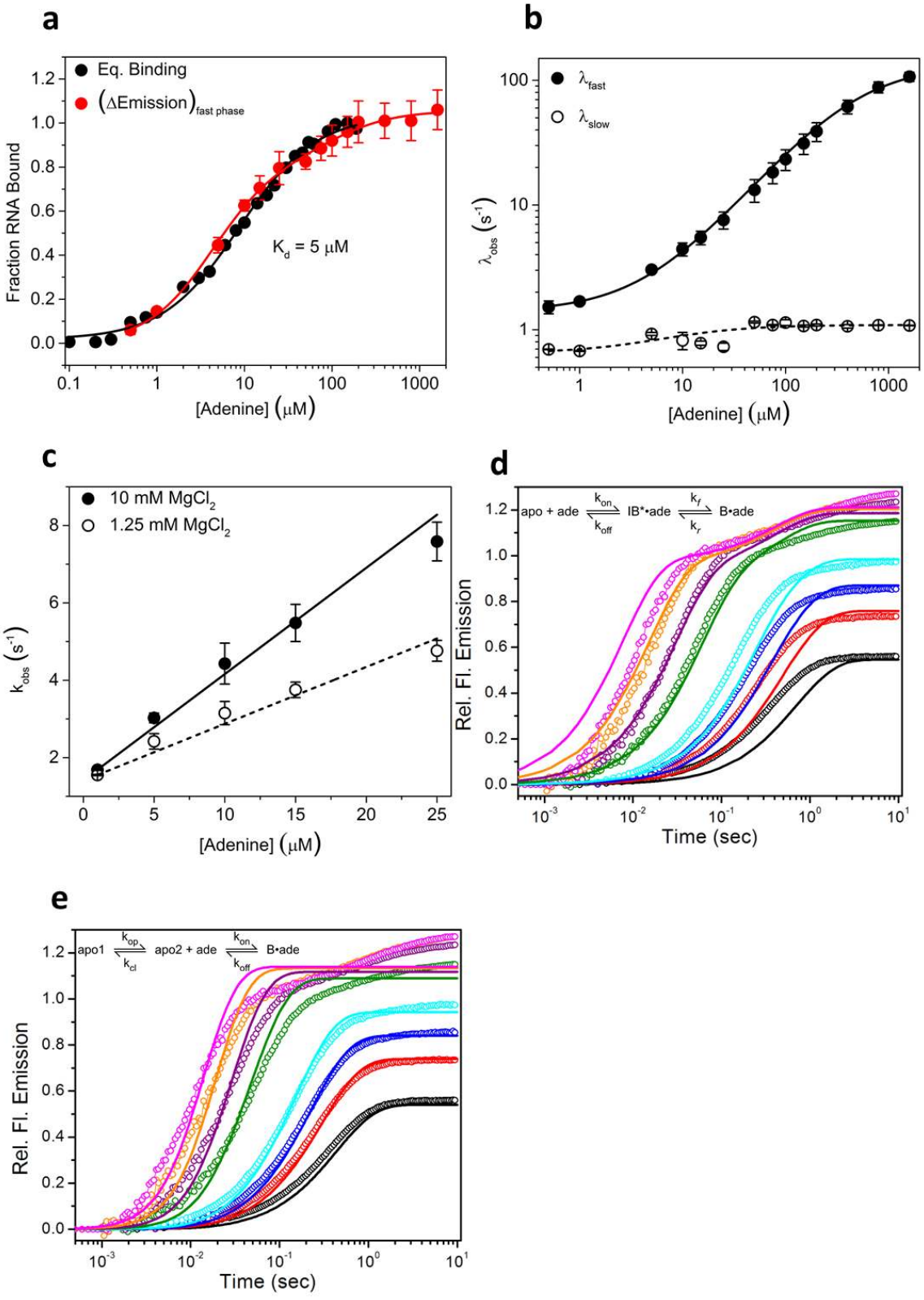




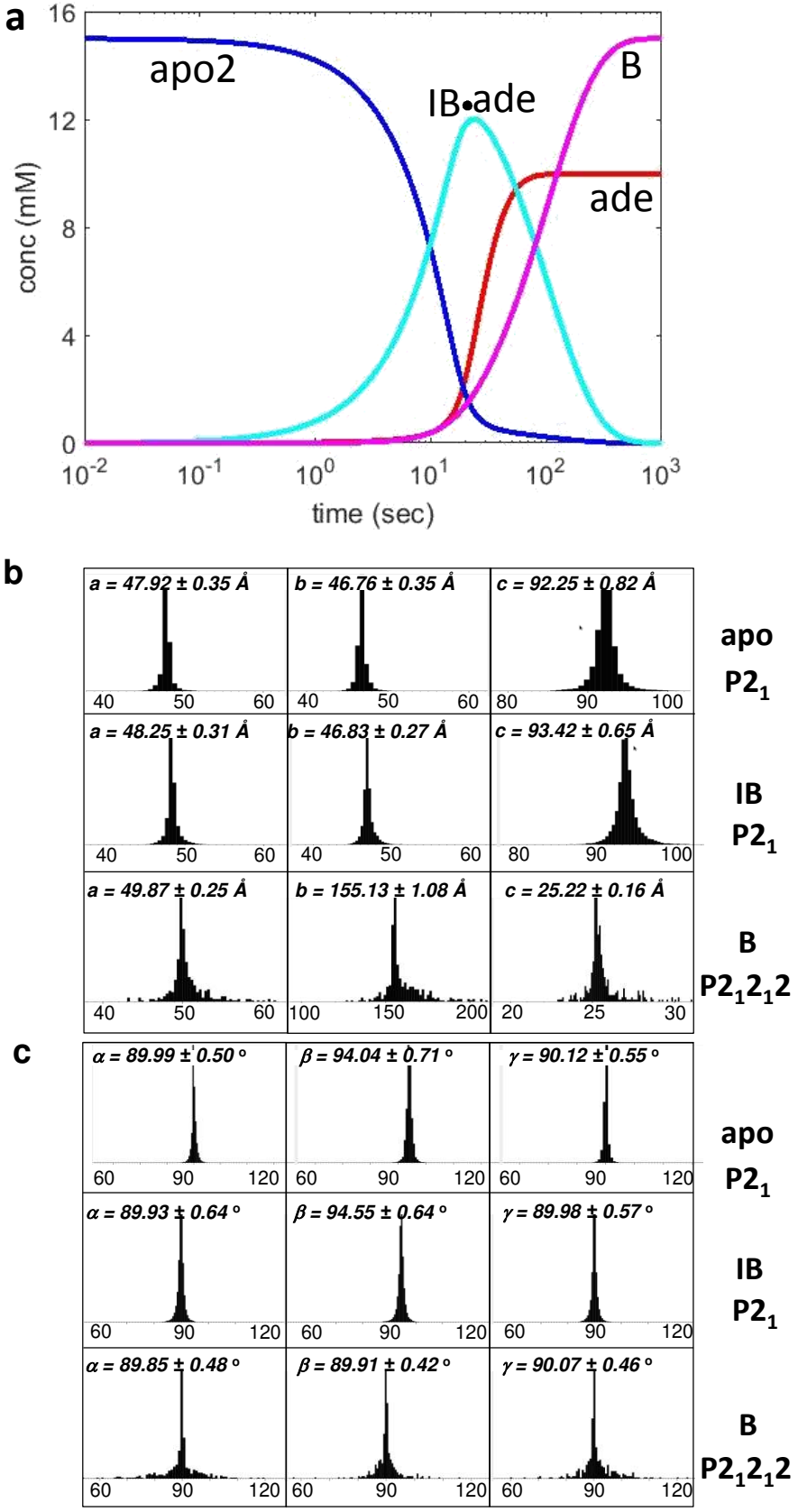
Extended Data Figure 3



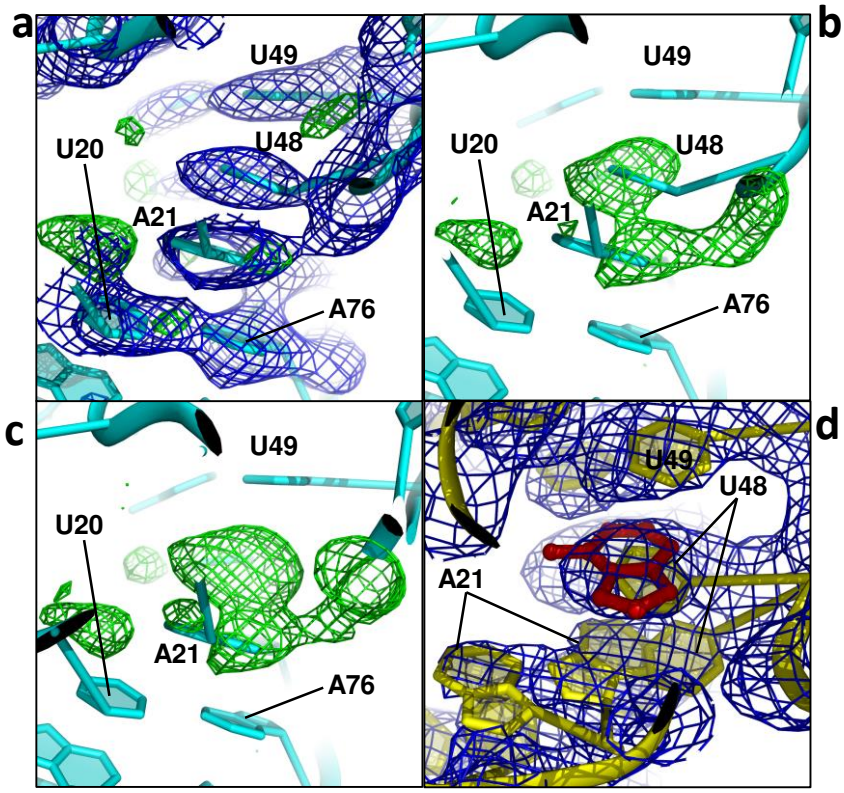
Extended Data Figure 4



Extended Data Figure 5



Extended Data Figure 6



Extended Data Figure 7