





ARTICLE

<https://doi.org/10.1038/s41467-019-13683-4>

OPEN

Metal ions and sugar puckering balance single-molecule kinetic heterogeneity in RNA and DNA tertiary contacts

Fabio D. Steffen ^{1,4}, Mokrane Khier^{1,4}, Danny Kowerko ^{1,2}, Richard A. Cunha¹, Richard Börner ^{1,3*} & Roland K.O. Sigel ^{1*}

The fidelity of group II intron self-splicing and retrohoming relies on long-range tertiary interactions between the intron and its flanking exons. By single-molecule FRET, we explore the binding kinetics of the most important, structurally conserved contact, the exon and intron binding site 1 (EBS1/IBS1). A comparison of RNA-RNA and RNA-DNA hybrid contacts identifies transient metal ion binding as a major source of kinetic heterogeneity which typically appears in the form of degenerate FRET states. Molecular dynamics simulations suggest a structural link between heterogeneity and the sugar conformation at the exon-intron binding interface. While Mg^{2+} ions lock the exon in place and give rise to long dwell times in the exon bound FRET state, sugar puckering alleviates this structural rigidity and likely promotes exon release. The interplay of sugar puckering and metal ion coordination may be an important mechanism to balance binding affinities of RNA and DNA interactions in general.

¹Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. ²Present address: Department of Informatics, Technical University Chemnitz, Straße der Nationen 62, 09111 Chemnitz, Germany. ³Present address: Laserinstitut Hochschule Mittweida, University of Applied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany. ⁴These authors contributed equally: Fabio D. Steffen, Mokrane Khier. *email: richard.boerner@hs-mittweida.de; roland.sigel@chem.uzh.ch

In the course of RNA maturation, self-splicing ribozymes catalyze two consecutive reactions, that is, the excision of the intron and the ligation of its flanking exons¹. To precisely locate the cleavage site, group I and II introns engage specific RNA recognition elements that base pair with complementary stretches on the exons (Fig. 1a)^{2–4}. In group II introns, these sequences are not conserved, instead the RNA identifies the structural transition between single- and double-stranded bases as the 5'-splice site⁵. The long-range tertiary interactions embed the 5'-exon in the active core of the ribozyme (Fig. 1b)^{6–8}, where a hydrogen-bond network and several coordinating metal ions convey stability to the tertiary contacts^{9,10}. Two Mg²⁺ directly participate in catalysis by activating the scissile bond for cleavage¹¹. Ideally, the metal ions fine-tune the interaction such that the 5'-exon is held in place until the second transesterification has occurred and the two exons are ligated. If the interaction between intron and exon is too weak, splicing either does not occur at all (intron retention) or stops after the first step, with the 5'-exon leaving the active site without being ligated. Conversely, if the association of intron and exon is too strong, the ribozyme no longer discriminates between correct and mismatched targets, which may lead to gene disruption and disease if reverse splicing occurs in tumor suppressor genes like p53^{12,13}. To minimize such errors, most group II introns use two independent exon recognition sites to keep hold of the 5'-exon: exon binding site 1 (EBS1) contributes the most to thermodynamic stability, while the surface-exposed EBS2 is responsible for target selectivity^{13,14}. Mono- and divalent ions play their part in strengthening the intron–exon interaction both through non-specific charge screening as well as site-directed coordination. A number of such

specific binding sites have been identified near active site elements^{9,10}. Like a padlock, the metal ion packs the strands together, thereby retaining the exon in the active site for a longer time than if no gatekeeping ion was present. As ion coordination is transient though, designated binding pockets are often only partially occupied and exon dissociation is thus kinetically heterogeneous¹⁵. Particularly, Mg²⁺ is known to induce such kinetic partitioning by interacting with RNA directly (inner-sphere coordination) or via a water molecule (outer-sphere coordination)^{16–18}.

Single-molecule techniques are well suited to detect subpopulations of molecules in different conformational states^{15,19–21}. This involves ion-induced collapse of RNA secondary structure elements as well as formation of more distant tertiary contacts¹⁷. Spectroscopic rulers such as Förster resonance energy transfer (FRET) capture these rearrangements based on a set of predefined distance coordinates. Yet, in some cases, be it due to the choice of the dye positions or the local geometry of the coordination environment, metal ion binding yields no detectable change in resonance energy transfer and results in a degenerate FRET state that comprises multiple kinetic states. Their existence is only inferred from the multiexponential decay rates from the one FRET state.

As metal ions direct folding and catalysis, heterogeneity is inherent to many RNAs^{16,22}. RNA folds hierarchically into a set of interconnected topological modules. The kinetics of their tertiary assembly is often rate determining in the folding process¹⁷. Pseudoknots are a recurrent motif in riboswitches, ribozymes, and the ribosome where they interface secondary and tertiary structure. They are often involved in the formation of the catalytic core or the aptamer domain^{23–26}. In its simplest form, an H-type

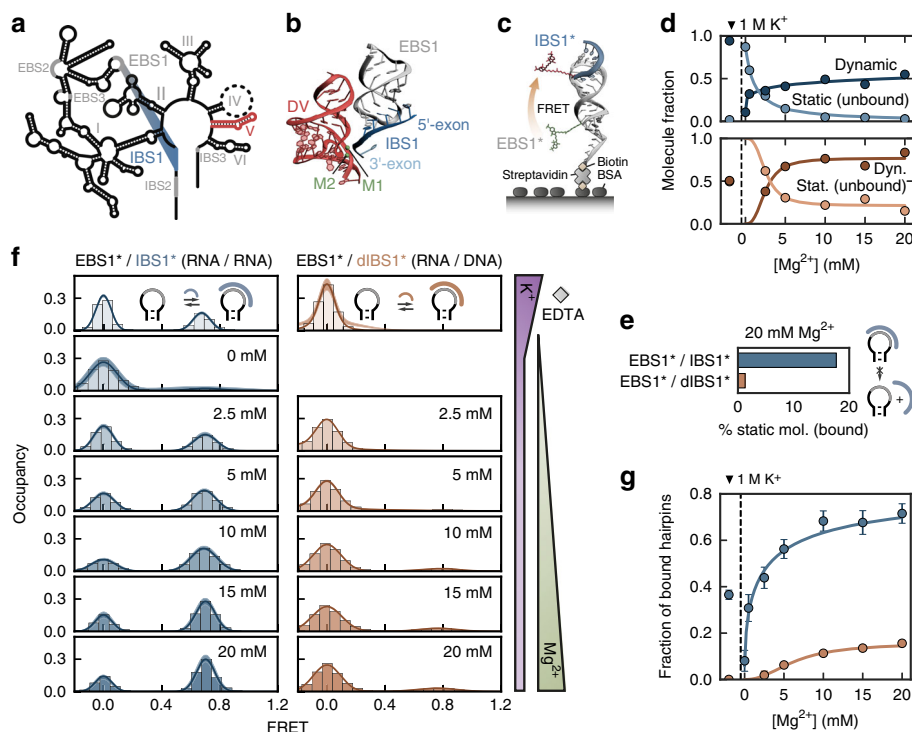


Fig. 1 Thermodynamics of Mg²⁺-induced tertiary contact formation from single-molecule FRET. **a** Secondary structure of the group IIB intron *Sc.ai5y* with the exon–intron binding site 1 (EBS1/IBS1, gray/blue) and the catalytic domain V (red). **b** Crystal structure of the active site of the P.ii. *LSU12* intron (PDB: 4ROD⁸) featuring domain V (DV, red), EBS1/IBS1 (gray/blue), and two catalytic metal ions (M1/M2, green). **c** Surface-immobilized FRET model system EBS1*/IBS1* labeled with Cy3 and Cy5 (BSA, streptavidin, and biotin are not drawn to scale). **d** Relative abundance of dynamic and statically unbound IBS1* (blue) or dIBS1* (orange) molecules as a function of Mg²⁺. **e** Percentage of long-lived, statically bound molecules at 20 mM Mg²⁺ and 100 mM K⁺. **f** FRET histograms of dynamic molecules showing a rising population of the high FRET state with increasing Mg²⁺ concentration. **g** Fraction of EBS1* hairpins bound to (d) IBS1* as a function of the Mg²⁺ concentration calculated from the integrals of the FRET states. Error bars correspond to the mean \pm 2 s.d. of 100 bootstrap samples. Source data are provided as a Source Data file.

pseudoknot consists of a hairpin that base pairs intramolecularly with a stretch of nucleotides outside the stem²⁷. In group II introns such a hairpin reaches out to the flanking exon and forms the recognition site for splicing and retrohoming^{2,28–30}. Here we use single-molecule FRET in combination with hidden Markov modeling to probe the pairing of EBS1 in the prototypical group II intron *Sc.ai5y* with its cognate intron binding site 1 (IBS1). To facilitate the interpretation of the binding free energy landscape, we have isolated the tertiary contact and tagged both interaction partner covalently with a fluorescent dye. In this way, we establish FRET as a reaction coordinate, which directly reports on the binding of exon and intron. Formation of this contact is a prerequisite for subsequent catalysis. While structures of the exon–intron complex before the first step of forward and reverse splicing have been solved recently^{7,31,32}, dynamic information on the association and dissociation reaction is still largely missing to date. By comparing how RNA and DNA exons bind to the isolated EBS1 hairpin, we can dissect the kinetics of the recognition step, which precedes forward and reverse splicing.

We find that RNA–DNA hybrids not only dissociate much faster than their pure RNA counterparts, but they do so in a more kinetically homogeneous manner. Our molecular dynamics simulations show that the origin lies in the conformational fit between the binding partners. Fast switching of the sugar puckers in the DNA exon compromises tertiary contact stability and likely impairs metal ion binding. This smooths out the rugged energy landscape and homogenizes the kinetics³³. Hence, sugar puckering is an effective way to counteract ion-induced heterogeneity in exon recognition and is a potentially widespread mechanism to modulate macromolecular interactions.

Results

DNA target recognition requires Mg^{2+} . We monitor the binding and unbinding of two seven nucleotides long exon strands, IBS1* (RNA) and dIBS1* (DNA), to surface-immobilized EBS1* hairpins by following the anti-correlated Cy3 and Cy5 emission over several minutes (Fig. 1c and Supplementary Fig. 1). The two intensity signals are converted into transfer efficiencies that fluctuate between a zero FRET state, corresponding to the unbound hairpin, and a high FRET state around 0.75 of the formed tertiary contact. We classified the molecules according to their interconversion frequency into (i) static zero (no binding event within observation time), (ii) static high (no unbinding event), (iii) one transition and (iv) dynamic (at least two transitions). The relative abundance of all four classes across different Mg^{2+} concentrations is summarized in Supplementary Fig. 2. Both RNA–RNA and RNA–DNA contacts are remarkably sensitive to Mg^{2+} , as indicated by the strong depletion of static zero FRET molecules in favor of dynamic molecules when Mg^{2+} is added (Fig. 1d). Static traces in the high FRET state only appear in the presence of IBS1*, but not dIBS1* (Fig. 1e). These persistent contacts (>400 s) are unique to the RNA–RNA contact and are strictly dependent on Mg^{2+} (Supplementary Fig. 2). This suggests that Mg^{2+} not only condenses around the RNA but binds site specifically to the tertiary contact. Further evidence for such a binding site comes from nuclear magnetic resonance (NMR) chemical shift mappings with Mg^{2+} and $[Co(NH_3)_6]^{3+}$, as well as Mn^{2+} -induced line broadening^{34,35}. The affected nucleotides coincide with a patch of negative surface potential near the 5′-splice site. Upon docking of IBS1* or dIBS1*, a cavity is formed, which readily accommodates a Mg^{2+} ion^{34,35}.

Divalent ions are therefore expected to promote the interaction of exon and intron by a combination of non-specific charge compensation and site-specific binding^{18,36}. The fraction of formed contacts are deduced from FRET efficiency histograms of all

dynamic molecules (Fig. 1f). The relative occupancy of the high FRET state increases as a function of the Mg^{2+} concentration (Fig. 1g). The normalized binding isotherm saturates at around 70% of bound EBS1*/IBS1* molecules. The fraction of bound RNA–DNA on the contrary does not exceed 20% even at 20 mM Mg^{2+} . Complete saturation of the hairpin would require exon concentrations in the high micromolar to millimolar range, which is inaccessible to single-molecule total internal reflection fluorescence (TIRF) imaging because of the high background signal (Supplementary Fig. 3). The dissociation constant calculated from the equilibrium population of the zero and high FRET states shows that EBS1*/dIBS1* is about an order of magnitude less stable than EBS1*/IBS1* (Supplementary Fig. 4a). This observation is consistent with previous reports of RNA–DNA hybrid duplexes being more labile than their canonical RNA–RNA counterparts^{37–39}. Notably, no EBS1*/dIBS1* formation at all is observed in the absence of Mg^{2+} , highlighting the importance of divalent ions in stabilizing tertiary structure motifs. The free energy contribution of Mg^{2+} binding to tertiary contact formation is given by the difference in Mg^{2+} binding to the hairpin alone and to the formed contact, involving both diffuse and site-specific components (Supplementary Information and Supplementary Fig. 4b). Diffuse binding is expected to be similar for RNA and DNA exons, thus the shift of the midpoint of the binding curve (RNA–RNA: $[Mg^{2+}]_{mid} = 3.7 \pm 0.9$ mM versus RNA–DNA: $[Mg^{2+}]_{mid} = 7.9 \pm 1.9$ mM) suggests a tighter inner- and/or outer-sphere coordination of Mg^{2+} at the RNA–RNA interface in line with previous binding affinities calculated from NMR chemical shifts^{15,40}.

Mg^{2+} induces heterogeneity by slowing down exon dissociation. Among the four molecular classes defined above, dynamic molecules are the most informative ones as they all represent functional RNA tertiary contacts that associate and dissociate within the observation window. To explore the kinetics of the interaction between the exon and intron binding sites, we discretized the FRET trajectories of the dynamic molecules into bound and unbound segments as shown in Fig. 2a. A visual inspection of the traces reveals both longer dwell times and more frequent transitions to the high FRET state, the more Mg^{2+} is present. By computing a mean dwell time (t_{zero}) and (t_{high}) for each trace, we can assess how uniformly the immobilized EBS1* hairpins react to IBS1* and dIBS1* binding (Fig. 2b). If there are kinetic subpopulations of EBS1* molecules, some of which bind IBS1* stronger than others, we would expect them to separate into individual clusters in the dwell time scatter plot as suggested by simulations (Supplementary Fig. 5a). On the other hand, if short and long binding events occur within the same trace, they will average out into a single stretched cluster. The width of the distribution is further influenced by the number of dwell times and thus by the limited observation time (Supplementary Fig. 5b, c). There is indeed no clear partitioning of the molecules into separate clusters (low molecule/subspecies heterogeneity, Fig. 2b). We observe, however, a positive correlation between the mean dwell time and the variability within a trace, measured by the difference between the shortest and the longest dwell time (Δt_{zero} or Δt_{high} , Supplementary Fig. 6). It follows that the most heterogeneous molecules (large Δt) also feature long mean dwell times.

To test whether these long dwell times are reinforced by divalent metal ions, we monitor the center of the mean dwell time ensemble across different Mg^{2+} concentrations (Fig. 2c and Supplementary Table 1). In the absence of Mg^{2+} , diffuse K^+ ions stimulate strand association mainly through charge compensation, but keep the excursions to the bound state relatively short. On average, IBS1* falls off the hairpin after about 20 s, and

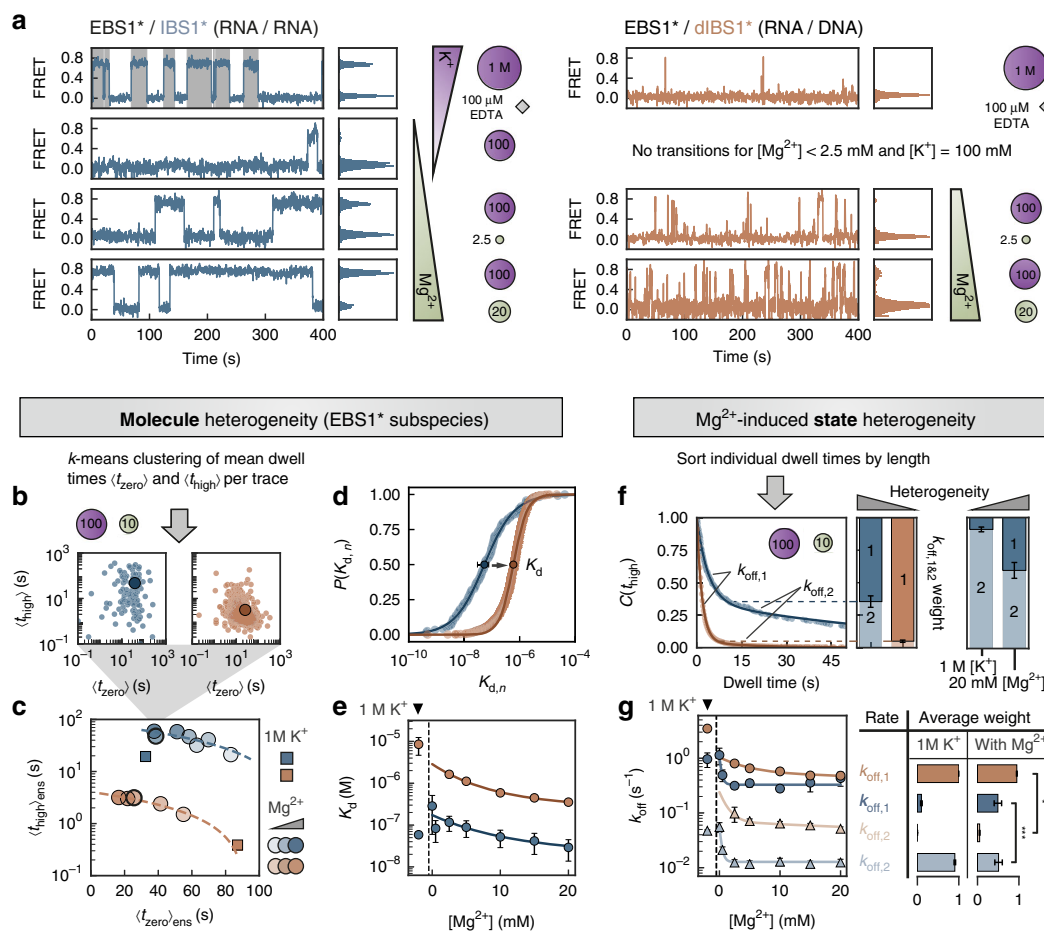


Fig. 2 Kinetic heterogeneity is tuned by Mg^{2+} and the nucleic acid strand type. **a** Representative single-molecule traces of EBS1*/IBS1* (blue) and EBS1*/dIBS1* (orange) displaying different residence times in the zero and high FRET state in response to gradual changes in the composition of the ionic environment. The traces are selected to reflect the overall histogram shown in Fig. 1f. The discretized state sequence is depicted by alternating backgrounds (white: zero FRET; gray: high FRET) for the top left trace and omitted elsewhere for clarity. Only dynamic molecules were used for the analysis. **b** The molecule-to-molecule variability is visualized by correlating the mean dwell times in the bound, (t_{high}), and unbound state, (t_{zero}), for every time trace. There are no distinct kinetic subspecies, but rather a single stretched cluster. Data is shown for 100 mM K^+ and 10 mM Mg^{2+} . **c** The cluster center is calculated by k -means and depicted as a circle in different shades of blue or orange for every Mg^{2+} concentration (10 mM Mg^{2+} is highlighted with a thick, black edge). Contact formation under high ionic strength (1 M K^+) in the absence of Mg^{2+} (100 μM EDTA) is indicated by a blue or orange square. **d** Cumulative $K_{d,n}$ histograms (10 mM Mg^{2+}) are fitted to a logistic function (Supplementary Eq. 13). **e** Stability constants of EBS1*/IBS1* and EBS1*/dIBS1* as a function of Mg^{2+} are quantified by the transition point, K_d , from fits in **c**. Error bars are drawn as mean ± 2 s.d. of 100 bootstrap samples. **f** Heterogeneity within the bound state is evaluated by computing cumulative dwell time histograms of the bound state. The presence of multiple decays is a characteristic of kinetic heterogeneity. The weighting factors (mean \pm s.d.) of the dissociation rate coefficients $k_{off,1}$ and $k_{off,2}$ are indicated alongside the decay. The ion dependence of the prefactor is displayed on the right for EBS1*/IBS1*. **g** Rate coefficients and average weighting factors (mean \pm s.d.) over the whole Mg^{2+} titration. The star (*) indicates the degree of heterogeneity. The latter is largest if the rate coefficients are far apart and equally populated (weighting factor ≈ 0.5). Source data are provided as a Source Data file.

dIBS1* unbinds almost immediately after docking (0.4 s). In contrast, addition of 20 mM Mg^{2+} ion triples the average residence time of the IBS1* exon and prolongates the dwell time of dIBS1* by a factor of 10. This goes along with an overall broadening of the distribution of mean dwell times (Supplementary Fig. 7). Hence, we propose that kinetic heterogeneity is induced predominantly by specific binding of Mg^{2+} to preformed tertiary contacts, which slows down the exon dissociation rate. Because Mg^{2+} coordination occurs on a similar timescale as exon unbinding, as shown by previous Mg^{2+} pulse experiments¹⁵, but is usually faster than the observation time, short- and long-lived dwell times coexist within the same trace.

Off-rates determine tertiary contact stability. A kinetic comparison of the relatively stronger RNA–RNA and the weaker

RNA–DNA interaction is useful in identifying the stability determining rate of tertiary contact formation. It can be viewed as a global ϕ -analysis, where instead of specific mutations, RNA is substituted with the same sequence of DNA to perturb the energetic levels of the bound native state, the transition state, or both^{41,42}. We found that the RNA–RNA and RNA–DNA differ mainly in the average time IBS1* or dIBS1* spends bound to the hairpin, that is, (t_{high}). In contrast, the average unbound dwell (t_{zero}) time is largely independent of the strand type. The ratio of unbound and bound mean dwell times translates into a dissociation constant, $K_{d,n}$ for each trace. The distribution of all $K_{d,n}$ is represented in a cumulative histogram and fitted to a logistic function (Supplementary Eq. 13, Fig. 2d, and Supplementary Fig. 8). The K_d at the inflection point is the expected value of the distribution and a measure for the stability of the tertiary contact. Over the Mg^{2+} titration (0–20 mM) the K_d of EBS1*/IBS1*

decreases by a factor of 10, resulting in an affinity of 29 ± 8 nM at 20 mM Mg^{2+} (Fig. 2e). The K_d of EBS1*/dIBS1* is about one order of magnitude higher under the same salt conditions (352 ± 8 nM at 20 mM Mg^{2+}), reflecting the lower stability of the RNA–DNA hybrid compared to the RNA–RNA, consistent with previous surface plasmon resonance (SPR) experiments (Supplementary Table 2)³⁵.

The similar on-rates of the two contacts suggest a reactant-like transition state where the exon and intron are in proximity, but base pairs are not yet fully formed. The energetic barrier originates mostly from the entropic cost of freezing backbone motions as well as uptake and localization of metal ions in the transition state^{20,43}. Because the energetic level of the transition state is invariant to the type of exon, the stability of the tertiary contact is dictated by the off-rates. These observations are in line with the emerging paradigm of early transition states being a hallmark of RNA folding^{20,24,39,41}. On a rugged energy landscape, it is the rate at which wrongly formed base pairs can be broken again, which limits the overall speed of folding¹⁷. Mg^{2+} ions provide both shortcuts along the folding route but also trap molecules in misfolded conformations. It is this dual role that leads to the stretched folding times characteristic for many ribozymes^{22,44}.

RNA–DNA is less kinetically heterogeneous. So far, we have looked at the average dwell time of (d)IBS1* in its bound and unbound state. This analysis gives us a robust estimate of the tertiary contact stability from the ensemble of single molecules. On the downside, variations in dwell time length within one trace are averaged out. Mg^{2+} titrations suggest that these variations are caused by ion coordination. Mg^{2+} binding is subtle such that it does not alter the inter-fluorophore distance and consequently ion bound and unbound populations display the very same FRET efficiency. In other words, a single FRET state degenerates into two interconverting kinetic states, each with its own off-rate. We refer to this scenario as state heterogeneity.

To determine the different rates, we computed complementary cumulative dwell time distributions of the bound and unbound state (Fig. 2f, Supplementary Fig. 9, and Supplementary Table 3). Binding kinetics are well described by a single exponential and a stretching factor β to account for small deviations at longer times. Unbinding, on the other hand, is not homogeneous, but partitions into two exponential terms in the presence of low millimolar amounts of Mg^{2+} . The relative weight of these two components differs considerably depending on the nature of the complementary strand. At 20 mM Mg^{2+} , the fast off-rate, $k_{off,1}$, accounts for $40 \pm 6\%$ of the RNA–RNA contact decay, whereas its amplitude rises to $96 \pm 1\%$ in the RNA–DNA contact, meaning the decay is practically homogeneous for the hybrid interaction, even though a small percentage of longer-lived contacts persists (Supplementary Fig. 9).

In either case, off-rates are most sensitive to low millimolar Mg^{2+} concentrations (0–2.5 mM), which coincides with the K_d of Mg^{2+} to the tertiary contact (Mg^{2+} binding in the loop: 1.78 ± 0.01 mM, binding at the stem-loop transition: 0.87 ± 0.01 , Fig. 2g and Supplementary Fig. 10)¹⁵. A comparison of the IBS1* and dIBS1* dissociation rates reveals that the higher stability of EBS1*/IBS1* over EBS1*/dIBS1* is conveyed by the slower and more prominent $k_{off,2}$ of the RNA–RNA contact. The fast $k_{off,1}$ is likely to originate from an RNA conformation where no Mg^{2+} is bound, while the slower $k_{off,2}$ links to a contact, which is stabilized by one or multiple site-bound Mg^{2+} ions. Since Mg^{2+} binding is transient, these states are exchanging on the timescale of imaging.

Degenerate FRET state is kinetically resolvable. To identify all interconversion rates between bound and unbound species, we

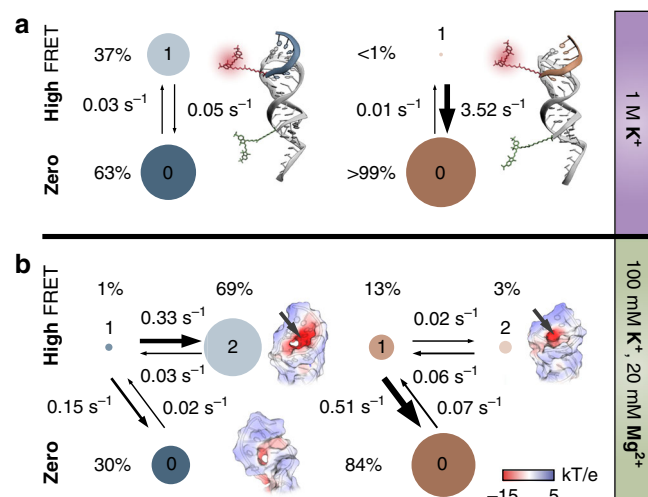


Fig. 3 Kinetic model of metal ion-induced IBS1* and dIBS1* recognition.

a Two-state binding of IBS1* (blue) and dIBS1* (orange) to the EBS1* hairpin under high monovalent ionic strength (0: exon unbound; 1: exon bound). The area of the circle is proportional to the state population (specified in percentage) and the arrow width reflects the relative rate of the transition. Binding rates, $k'_{on,01}$, are indicated as pseudo-first-order rates coefficients calculated from the trained HMM. **b** Three-state model with two degenerate high FRET states reflecting the coordination of Mg^{2+} to the formed tertiary contact (0: exon unbound, Mg^{2+} unbound; 1: exon bound, Mg^{2+} unbound; 2: exon bound, Mg^{2+} bound). Insets show surface representations of the RNA–RNA and RNA–DNA interaction, color coded by the electrostatic potential (red: -15 kT/e⁻¹; blue 5 kT/e⁻¹). Arrows indicate coordination sites for Mg^{2+} . Source data are provided as a Source Data file.

used a maximum-likelihood approach that trains a global hidden Markov model (HMM) on all dynamic FRET traces⁴⁵. We tested various kinetic networks and found a sequential three-state model with one zero and a twofold degenerate, high FRET state describing the data best whenever Mg^{2+} is present (Fig. 3b, Table 1, and Supplementary Fig. 11 and Supplementary Table 4, see also Supplementary Discussion). The most appropriate model has been selected based on the Bayesian information criterion (BIC, Supplementary Fig. 12) and prior biochemical knowledge about the system as outlined below. Furthermore, it is critically evaluated by re-simulating the obtained rate system and by reproducing experimental dwell time distributions (Supplementary Figs. 13 and 14 and Supplementary Table 5). Biochemically, the kinetic model couples IBS1* or dIBS1* binding to Mg^{2+} coordination and can be interpreted in the following way: docking of the complement to the hairpin involves a detectable change in FRET efficiency, followed by a state transition, which is insensitive to FRET, thus degenerate, but kinetically resolvable. We attribute this second transition to one or multiple Mg^{2+} ions that specifically coordinate to the exon bound hairpin. Two key observations support this hypothesis: (i) in the absence of Mg^{2+} , the mechanism simplifies to a two-state system with one on- and one off-rate (Fig. 3a and Table 1); (ii) as seen from the NMR structure, a patch of negative electrostatic surface potential is exposed once IBS1* or dIBS1* binds to EBS1*, thereby forming a cavity where Mg^{2+} can coordinate. This pocket is particularly pronounced in the RNA–RNA interaction and absent when no exon is bound. There are two alternative metal ion-binding modes at the exon–intron interface between EBS1*/dIBS1* (Fig. 4a): an outer-sphere bound ion has been proposed based on nuclear Overhauser enhancement (NOE) restraints from the nucleic acids to the NMR-active $[Co(NH_3)_6]^{3+}$ (green sphere)³⁵.

Table 1 Kinetic parameters determined with a hidden Markov model on dynamic single molecules.

Exon	Metal ions	$k'_{on,01}$ ($10^{-2} s^{-1}$) ^a	$k_{on,01}$ ($10^5 M^{-1} s^{-1}$) ^b	$k_{off,10}$ (s^{-1})	k_{12} ($10^{-2} s^{-1}$)	k_{21} ($10^{-2} s^{-1}$)	K_d (nM) ^c	ΔG_{bind} (kJ mol ⁻¹)	ΔG^\ddagger (kJ mol ⁻¹)
IBS1*	1 M K ⁺ ^d	2.96 ± 0.04	8.45 ± 0.11	0.051 ± 0.001	–	–	59.0	–41.2	39.2 ± 0.1
IBS1*	20 mM Mg ²⁺	2.14 ± 0.07	6.11 ± 0.20	0.146 ± 0.006	33.1 ± 1.4	2.56 ± 0.10	15.2	–44.6	40.0 ± 0.1
dIBS1*	1 M K ⁺ ^d	0.86 ± 0.03	1.72 ± 0.06	3.52 ± 0.11	–	–	17,500	–27.2	43.1 ± 0.1
dIBS1*	20 mM Mg ²⁺	7.07 ± 0.03	14.1 ± 0.1	0.511 ± 0.002	2.35 ± 0.07	6.33 ± 0.16	255	–37.6	37.9 ± 0.1

Errors are standard deviations computed from likelihood ratio tests⁷⁶

^aPseudo-first-order association rate coefficients, $k'_{on,01}$, are extracted from the HMM

^bSecond-order rate coefficients, $k_{on,01}$, are calculated from the pseudo-first-order association rate coefficient and the total IBS1* concentration in solution (35 nM IBS1* or 50 nM dIBS1*). For IBS1* this is $k_{on,01} = k'_{on,01} \times c_{IBS1*}^{-1}$

^cDissociation constants are calculated from the relative state population S_{bound} and $S_{unbound}$ and the ligand concentration. For IBS1* the K_d is thus given by $K_d = S_{unbound} \times c_{IBS1*} \times S_{bound}^{-1}$

^dEDTA (100 μM) are added to chelate any traces of divalent metal ions

Alternatively, a Mg²⁺ may exchange one or more of its coordinated water molecules with RNA atoms and bind deeper within the pocket (dark green sphere). It seems likely that inner-sphere binding of Mg²⁺ in that tunnel locks the tertiary contact in a rigid and stable conformation, which prevents fast dissociation of IBS1*. In the HMM model, the corresponding exchange rate $k_{ex,12}$, which describes the Mg²⁺-induced stabilization of IBS1*, increases at the expense of the dissociation rate $k_{off,10}$ over the course of the Mg²⁺ titration (Supplementary Fig. 11). In the RNA–DNA hybrid, on the other hand, the off-rate outweighs the exchange between different EBS1*/dIBS1* conformations, thereby attenuating the degree of heterogeneity in dIBS1* unbinding.

Sugar puckering increases dynamics at the RNA–DNA interface.

What is the molecular origin of the lower stability of the RNA–DNA contact and the reduction in kinetic heterogeneity? Structurally RNA and DNA are set apart by a methyl group on thymine and the lack of a 2'-OH on the DNA sugar. Differences in duplex stability have been attributed to both features⁴⁶. The 2'-hydroxyl group contributes to the structural diversity of RNA by forming hydrogen bonds that make up tertiary structure elements such as the ribose zipper or the A-minor motif^{47,48}. In helical regions, the puckering mode of the sugar is critically influenced by the presence or absence of the 2'-OH group. The most common helical topologies in nature are the A and B forms. While DNA mostly adopts the B-form (C2'-endo), RNA prefers the A-form (C3'-endo), in which the 2'-OH is in a sterically and electronically more favorable axial position (Fig. 4b)⁴⁹. The type of sugar pucker can be inferred from TOCSY experiments, where ³J coupling between H1' and H2' protons is related to the dihedral angle via the Karplus equation^{49,50}. Coupling between the two nuclei is strong if the sugar is in a C2'-endo conformation with the protons in axial alignment. A switch to C3'-endo rotates the protons into an equatorial position where they are oriented perpendicularly to each other and coupling is weak^{34,35}. Puckering dynamics have been observed on different timescales from picoseconds up to milliseconds^{51–53}.

The seven nucleotides of IBS1* and their base-paired complements on the EBS1* hairpin show no crosspeaks between H1' and H2' as opposed to the unpaired loop residues A10, U11, and U12, which display a sharp signal (Supplementary Fig. 15). In previous NMR structure calculations, the torsion angles have therefore been restrained to either a C3'-endo (IBS1* and paired nucleotides of EBS1*) or a C2'-endo (A10, U11, and U12) conformation^{34,35}. In dIBS1*, correlations between H1' and H2'/H2'' are observable for all seven residues, yet the linewidths of these peaks are broader compared to U11 and U12. Quantification of the peak volumes yields intermediary values, suggesting a fast exchange between puckering modes on the NMR timescale. Consequently, no restraints have been applied to these residues

and the resulting lowest energy NMR structures show pucker from C2'-exo to C2'-endo (Fig. 4c).

To assess whether the intermediary ring pucker (C4'-exo, O4'-endo, C1'-exo) are a result of interconverting conformations averaged out in the NMR measurement, we run molecular dynamics simulations for both tertiary contacts in their docked form. We first simulated the two contacts in the presence of K⁺ only, and in a second set of runs, we added Mg²⁺ at the positions expected from NMR chemical shifts, NOE restraints, and electrostatic calculations (Supplementary Fig. 16). In both cases, the ribose of the IBS1* residues is predominantly in the canonical C3'-endo conformation as expected from the absence of IBS1* crosspeaks in the TOCSY spectrum (Fig. 4c and Supplementary Figs. 17–20). Excursions to a C2'-endo occur mostly at the flexible 3'-terminus (C65). Overall, the canonical A-form helix of EBS1*/IBS1* is relatively rigid^{49,52}. In contrast, residues in dIBS1* undergo frequent repuckerings with correlation times in the mid picosecond range both in the presence and absence of Mg²⁺ (Supplementary Figs. 17 and 19). Their pucker profiles are very broad and cover most of the Eastern half of the pseudorotation cycle. The simulations show no significant energy barrier separating north and south pucker and thus a variety of intermediary deoxyribose conformations are explored (Supplementary Figs. 18 and 20). Interestingly, the unpaired loop residues A10–U12 of the EBS1* hairpin also switch between C2'-endo and C3'-endo conformations, but on a timescale of several nanoseconds (Supplementary Figs. 21 and 22). Unlike the residues in dIBS1*, intermediary pucker in A10–U12 are unfavored due to the presence of a distinct energy barrier.

Because the dihedral angles of the sugar ring are all interdependent, the pucker dynamics propagate onto the backbone where they alter the distance between neighboring phosphates and the shape of the helix⁴⁹. As a result, the RNA/DNA hybrid adopts neither a pure A-form nor a B-form geometry^{34,54,55}. The conformational exchange at the RNA–DNA interface probably weakens the binding, thus contributing to the lower stability of hybrid contacts in general and in our case to faster dissociation rates of dIBS1*^{37,38}.

Discussion

Group II introns have developed a highly specific recognition mechanism to cut and paste themselves from one part of the genome to another. Each of the three classes, IIA, IIB, and IIC, relies on a slightly different set of tertiary interactions to position the target strand in the active site for cleavage^{1,2}. However, common to all is the EBS1, which forms a duplex directly adjacent to the 5'-splice site. Here, we have investigated the thermodynamics and kinetics governing the stability of this tertiary contact using single-molecule FRET in combination with molecular dynamic simulations.

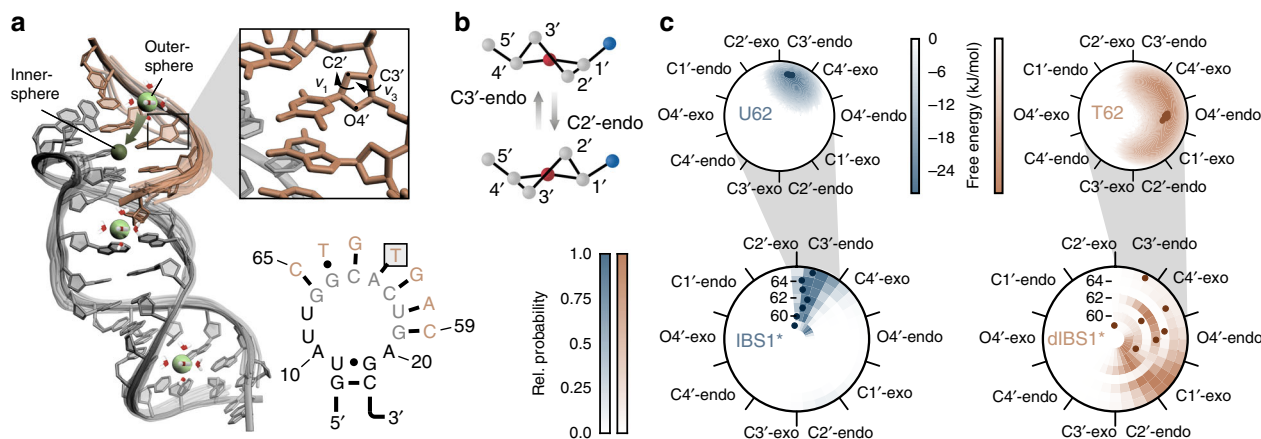


Fig. 4 Sugar pucker distributions from all-atom MD simulations. **a** NMR structure of EBS1*/dIBS1* (PDB: 2m1v³⁴) with overlaid backbones of the 18 lowest energy structures. Outer-sphere binding sites based on NOE restraints from dIBS1* protons (i.e., H3 and CH₃ of T62/64 and H5/6 of C65) to [Co(NH₃)₆]³⁺ (large green spheres; H₂O molecules are added for clarity) and a potential inner-sphere coordination mode where Mg²⁺ binds further inside the tunnel (smaller dark green sphere)³⁵. **b** Switch between energetically favored sugar pucker conformations C3'-endo and C2'-endo. **c** Circular histogram of pseudorotation angles along the seven nucleotides of IBS1* or dIBS1* (C59–C65) from MD simulations with 100 mM K⁺ and 20 mM Mg²⁺. Insets show distribution of pucker phases and amplitudes of U62 and T62. Puckers of the lowest energy NMR structures and their medians are indicated as dots for each residue. Source data are provided as a Source Data file.

Our FRET trajectories capture binding and unbinding events across various K⁺ and Mg²⁺ concentrations. Consistent with previous experiments, IBS1 (un)docking and Mg²⁺ coordination are coupled, resulting in Mg²⁺ exchange rates on a timescale of milliseconds to seconds, thus accessible to camera-based detection^{15,16}. We find that Mg²⁺ association and dissociation from the EBS1*/IBS1* contact are linked to very subtle conformational changes within the duplex, indiscernible by FRET. As a spectroscopic ruler, the FRET efficiency fails to resolve these kinetic states. In the time dimension, however, the states can be distinguished with regard to their dwell times (Fig. 5). Hence, exon unbinding is characterized by multiexponential decays from one degenerate FRET state, conceptually known as kinetic heterogeneity. Heterogeneity is common to many nucleic acid interactions and increases the complexity of the kinetic analysis¹⁶. Here, we have shown how to dissect such a state degeneracy and elucidate its molecular origin. For this purpose, we have drawn a state network of EBS1*/(d)IBS1* using a combination of dwell time analysis and hidden Markov modeling on an ensemble of FRET trajectories.

Based on our model system, we propose that the 5'-exon dissociates at different rates depending on whether the Mg²⁺ binding sites are occupied or not. While outer-sphere Mg²⁺ ions are typically thought to exchange on a millisecond timescale⁵⁶, our previous single-molecule Mg²⁺ pulse experiments suggest that some ions remain associated with the RNA for seconds or even minutes¹⁵. These ions are probably chelated by the RNA¹⁸ and buried in a binding pocket formed upon docking of the 5'-exon³⁴. The said cavity is shallower in the hybrid EBS1*/dIBS1* structure and Mg²⁺ binding to RNA–DNA contacts is generally weaker³⁵. Yet, Mg²⁺ is required to stabilize the intrinsically labile RNA–DNA contact. Loss of a coordinated Mg²⁺ ion is therefore likely to be followed by immediate strand dissociation. Such a concerted mechanism of a Mg²⁺ and exon dissociation is consistent with the observed reduction in kinetic heterogeneity from IBS1* to dIBS1*.

How does ribozyme catalysis benefit from heterogeneity? Group II introns have evolved to find just the right balance of keeping a correct substrate in the active site long enough for the two transesterifications to occur, while releasing mismatched sequences before the first chemical step can take place¹³. To

achieve such high selectivity, ribozymes make use of kinetic partitioning. Mg²⁺ prolongates the residence time of the exon given that there is a suitable binding site. Selective stabilization of the structurally and catalytically important exon–intron tertiary interaction forms the basis of its heterogeneous kinetics. A depletion of Mg²⁺ homogenizes the rates. Furthermore, splicing is downregulated, because K⁺ alone at physiological concentrations is neither able to stabilize the exon–intron contact sufficiently nor activate the phosphodiester bond for cleavage^{11,57}.

In the cell, where Mg²⁺ is present in very low millimolar concentrations, Ca²⁺ is tightly regulated, and other divalent metal ions are generally scarce, intron-encoded or host supplied proteins are recruited to stabilize the weak interactions and promote catalysis. The group IIB intron *Sc.ai5y* from which the EBS1*/IBS1* model interaction is derived, is a splicing-only intron, and resides in the housekeeping gene *COX1* of yeast mitochondria, where it is assisted by splicing factors such as the ATP-dependent helicase *Mss116*^{58–60}. Mobile group II introns, in contrast, still have their own open reading frame encoding a maturase that stabilizes the exon–intron contact^{61,62}. Only recently, cryo-EM structures of a maturase–intron complex from *Lactococcus lactis* and *Thermosynechococcus elongatus* have revealed how the exon is sandwiched between the protein's thumb domain and EBS1/2^{31,32}. This ternary complex between intron, exon, and protein might be of particular importance to convey stability to the labile RNA–DNA interaction during reverse splicing. Notably, the intron-encoded protein (IEP) has been found to bind more tightly to an intron lariat with the 5'-exon added *in trans* than to a spliced intron lacking the EBS/IBS interactions⁶¹. The exon–intron recognition complex and the IEP thus seem to cooperate to resolve instabilities in DNA target binding.

Our molecular simulations provide an atomic view on the exon–intron binding interaction and suggest that fast sugar pucker dynamics may be responsible for the lower affinity of the intron towards DNA exons. The simulations show that sugar puckers of dIBS1* are dispersed across the Eastern half of the pseudorotation cycle. In this energy basin the sugars are free to adopt different conformations. Does this switching have any regulatory function? On a microsecond to millisecond timescale, exchange between 3'-endo and 2'-endo sugars was found to be associated with helical to non-helical transitions, leading to

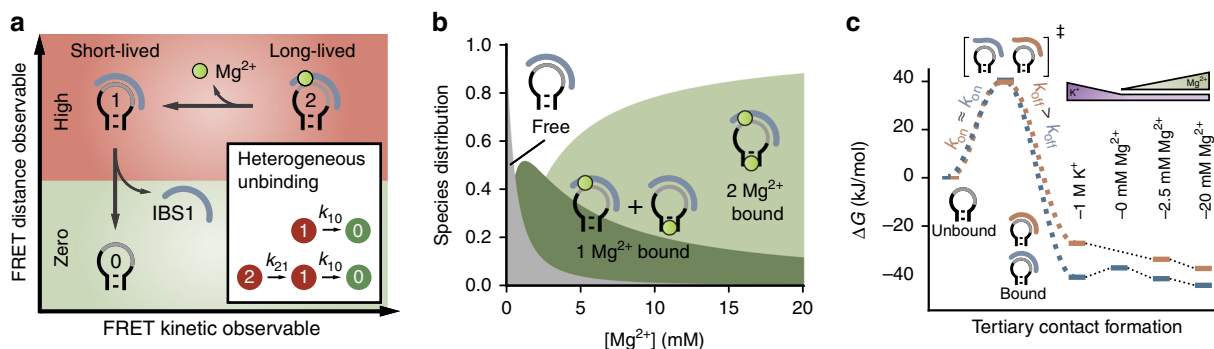


Fig. 5 Tertiary contact formation monitored by FRET in space and time. **a** Exon release is described by two FRET observables: dye distance and state dwell times. Intron–exon dissociation can occur from state 1 (Mg^{2+} free) or state 2 (Mg^{2+} bound) and is thus kinetically heterogeneous. **b** Occupation of Mg^{2+} binding sites as a function of the ion concentration using Mg^{2+} binding constants from NMR chemical shift perturbations (Eq. 5)¹⁵. Cartoons depict possible species with no (gray), one (dark green), or two Mg^{2+} bound (light green). The Mg^{2+} free species corresponds to the high FRET state 1 in **a**, while the other species together make up state 2. Diffuse Mg^{2+} ions are omitted for clarity. **c** Free energy diagram of tertiary contact formation featuring a reactant-like, unstructured transition state. Off-rates determine the stability of the interactions. Dotted black lines connect levels of the bound state at different metal ion concentrations. Source data are provided as a Source Data file.

alternative secondary structures, a feature common in riboswitches^{51,63}. These rather slow puckering modes are further reminiscent of the *cis/trans* isomerization in peptide bonds containing proline residues, which were recently discovered to be a source of heterogeneity in intrinsically disordered protein interactions⁶⁴. Unlike proline isomerization and secondary structure rearrangements, sugar puckering in dIBS1* happens on a timescale of picoseconds and is thus much faster than exon release. Consequently, the intron is prone to let go of its DNA substrate unless the contact is properly stabilized by Mg^{2+} and/or a protein. We see this facile reversibility of EBS1/dIBS1 binding as a sort of quality control mechanism that ensures correct integration site selection as the ribonucleoprotein (RNP) scans along the exon⁶⁵.

To test this hypothesis, the EBS1/IBS1 interaction needs to be put back into the context of the full intron. Such large and architecturally complex RNAs and RNPs have long eluded a detailed single-molecule kinetic analysis mainly because the site-specific introduction of suitable reporters, like fluorophores, has proven to be challenging⁶⁶. For instance, to probe the same reaction coordinate as in our EBS1*/(d)IBS1* model construct, a dye needs to be placed adjacent to EBS1, which is located around nucleotide position 330 in *Sc.ai5y*. Making such a FRET construct would require multiple fragments to be ligated enzymatically. Recent advances in co- and post-transcriptional nucleic acid labeling now provide more flexibility in choosing dye positions, which are close to functionally important elements, and thus provide a basis for studying the kinetics of large RNA–protein assemblies on a single-molecule level^{67,68}.

Here, we have used instead a minimal construct, which is chemically synthesized and covalently labeled, giving us a more straightforward access to studying its binding kinetics. We have shown that ion binding and sugar puckering are two interacting processes that shape kinetic heterogeneity. While site specifically coordinated Mg^{2+} ions delay strand dissociation, sugar switching effectively promotes it. Off-rates are therefore the determining factor for contact stability, in line with previous studies on RNA kissing loops and various kinds of RNA, DNA, and hybrid duplexes^{39,69}. A question that remains is whether sugar puckering attenuates heterogeneity by directly interfering with metal ion coordination. To explore the relative binding affinities of Mg^{2+} to pure RNA and RNA–DNA hybrid contacts, MD simulations with enhanced sampling may prove useful^{43,70}. Such simulation might also detect small structural changes induced upon Mg^{2+} coordination that FRET struggles to resolve. Altering the degree of

residual dynamics at binding interfaces is a common type regulation in biomolecules. By balancing flexibility and rigidity, sugar puckering and metal ions control RNA dynamics and form the molecular basis of kinetic heterogeneity.

Methods

Construct design. Oligos were purchased from IBA Lifesciences (Göttingen, DE) or Microsynth (Balgach, Switzerland) with cyanine labels at the 5' end (Cy3-EBS1* and Cy5-IBS1* or Cy5-dIBS1*) and a biotin at the 3' end. Photophysical parameters (fluorescence lifetime, time-resolved anisotropy) of the labeled oligos have been previously characterized in detail⁷¹. Sequences are derived from the group IIB intron *Sc.ai5y* and feature a slightly elongated stem as well as two A-to-C transversions in the loop. These were previously introduced at residues 15 and 17 to stabilize the exon–intron interaction (see Supplementary Methods). For surface immobilization on the microscopy slides, a single-stranded overhang was added that contains four uracils and a biotin at the 3' terminus¹⁵. We refer to these mutated and labeled constructs as EBS1*/IBS1* (RNA–RNA) and EBS1*/dIBS1* (RNA–DNA) to differentiate them from the original sequence in the group II intron.

Single-molecule FRET experiments and trace processing. RNA hairpins and their RNA or DNA complements were buffered in 50 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS) at pH 6.9. The imaging solution further contained 100 mM KCl, 1% glucose (w/v), 1 mM Trolox, an oxygen-scavenging system (2170 U/mL catalase, 165 U/mL glucose oxidase), and varying amounts of MgCl_2 (0–20 mM) or alternatively 1 M KCl with 100 μM EDTA to remove any traces of divalent ions. The biotinylated EBS1* RNA was immobilized on a quartz surface via a streptavidin linkage⁷². For this purpose, the microfluidic channel was first coated with biotinylated bovine serum albumin (5 min), treated with streptavidin (10 min), and finally incubated with biotinylated EBS1* RNA (10 pM, 5 min). IBS1* or dIBS1* strands were then flushed into the chamber at concentrations of 35 or 50 nM, respectively. Molecules were imaged on a custom-built TIRF microscope using a water-immersion objective (UPlanSApo 60 \times /1.2-W, Olympus) at a camera frame rate of 5 or 10 Hz (Andor iXon DU-897). Single molecules with anti-correlated donor/acceptor emission were selected and corrected for background and donor bleedthrough into the acceptor channel using the freely available software package MASH-FRET (<https://github.com/RNA-FRETtools/MASH-FRET.git>)^{16,73}. The FRET efficiency was then calculated from the corrected donor and acceptor intensities after donor excitation according to

$$\text{FRET} = \frac{I_A}{I_A + I_D} \quad (1)$$

FRET histograms were fitted to a Gaussian mixture model and uncertainties were estimated by bootstrapping⁷⁴.

Single-molecule kinetic analysis. Individual single-molecule traces were discretized based on a threshold criterion to generate sets of dwell times in the well-separated zero and high FRET state. A mean dwell time, $\langle t_{\text{zero},n} \rangle$ and $\langle t_{\text{high},n} \rangle$ was calculated for each trace n . On the other hand, cumulative dwell time histograms were computed by sorting dwell times from all dynamic molecules. Only molecules showing at least two transitions between the two states were considered as dynamic. Characteristic decay constants, $\tau_{\text{zero},m}$ and $\tau_{\text{high},m}$ were obtained by fitting a biexponential (Supplementary Eq. 2) or a stretched exponential (Supplementary Eq. 3) to the distribution. For the binding reaction, second-order rate coefficients, k_{on} , were calculated by dividing the

pseudo-first-order on-rate, k'_{on} , by the total IBS1* or dIBS1* concentration in solution: $k'_{\text{on}} = k_{\text{on}} \times c_{(\text{d})\text{IBS1}^*}^{-1}$. Dissociation constants, K_{d} , were computed from the on- and off-rates for each decay component m or alternatively for each individual molecule n . In case of the latter, cumulative distributions of $K_{\text{d},n}$ were built and fitted to a normalized logistic function

$$P(K_{\text{d},n}) = \frac{1}{1 + (K_{\text{d}}/K_{\text{d},n})^p}, \quad (2)$$

Error estimates on all variables were computed from a set of 100 bootstrap samples. For a more detailed mathematical description of the kinetic analysis, we refer the reader to the Supplementary Information. The Gibbs free energy of the bound state with respect to the free hairpin was determined using

$$\Delta G = RT \ln(K_{\text{d}}), \quad (3)$$

with the gas constant R and temperature T . The energy barrier separating the bound and unbound state was computed according to transition state theory as

$$\Delta G^\ddagger = -RT \ln\left(\frac{k_{\text{on}}h}{\kappa k_{\text{B}}T}\right), \quad (4)$$

with the binding rate k_{on} , Planck's constant h , and Boltzmann's constant k_{B} . Our calculations of ΔG^\ddagger assume a transmission coefficient $\kappa = 1$, thus providing a lower limit for the difference in free energy between the reactants and the transition state^{20,75}.

The fraction of Mg^{2+} bound RNA Θ is computed using association constants, K_1 and K_2 , derived from NMR chemical shift perturbations¹⁵

$$\Theta = \frac{K_1[\text{Mg}^{2+}] + K_2[\text{Mg}^{2+}] + K_1K_2[\text{Mg}^{2+}]^2}{1 + K_1[\text{Mg}^{2+}] + K_2[\text{Mg}^{2+}] + K_1K_2[\text{Mg}^{2+}]^2}. \quad (5)$$

Global HMM and FRET trace simulation. For each salt condition, a HMM was first trained on a trace-by-trace basis as implemented in the software package SMACKS^{45,76}. In a second step, the predetermined emission probabilities were fixed and the start and transition probabilities of a global HMM were optimized collectively on the entire ensemble of dynamic FRET traces. Kinetic models with different connectivities and degeneracies were tested and evaluated based on a BIC and biochemical significance (see Supplementary Discussion). To validate the obtained rate system, degenerate FRET traces were re-simulated in a subroutine of MASH-FRET⁷⁷ from the underlying rate matrix

$$\mathbf{K} = \begin{pmatrix} -k_{01} & k_{01} & 0 \\ k_{10} & -(k_{10} + k_{12}) & k_{12} \\ 0 & k_{12} & -k_{12} \end{pmatrix}. \quad (6)$$

Dwell time histograms were built analogously to the experimental traces.

Poisson-Boltzmann continuum electrostatics. The electrostatic surface potential of EBS1* alone and EBS1* in complex with either IBS1* or dIBS1* were computed with the APBS Electrostatics plugin⁷⁸ for PyMOL. Partial charges and VdW radii were assigned with PDB2PQR⁷⁹ using parameters from AMBER-ff99. Calculations were run at 300 K with 100 mM monovalent and 20 mM divalent ions.

Molecular dynamics simulation. NMR structures of the tertiary contact formed between EBS1* and IBS1* or dIBS1* (PDB: 2M23 and 2M1V) were used as starting points for molecular dynamics simulations with GROMACS 5.1⁸⁰. The simulations use the AMBER-ff99 force field⁸¹ with parmbsc0⁸² and $\chi\text{OL3}^{83,84}$ corrections (corresponds to AMBER-ff14). Monovalent ions were modeled with parameters from Joung and Cheatham⁸⁵, while those for Mg^{2+} were taken from Li et al.⁸⁶. The molecules were solvated in a dodecahedral box filled with TIP3P water and randomly interspersed with K^+ and Cl^- ions to neutralize the net charge and reach a concentration of about 100 mM. In a second set of simulations, Mg^{2+} ions were placed at specific locations suggested by NMR chemical shifts and NOE restraints to $[\text{Co}(\text{NH}_3)_6]^{3+}$. The nucleic acids were then charge neutralized and the K^+ and Mg^{2+} concentrations set to 100 mM and 20 mM, respectively (concentrations are estimates because of the limited box size). In an equilibration phase (2×5 ns, NVT and the NPT ensembles), the temperature was adjusted to 300 K using the velocity rescale thermostat and the pressure was kept constant at 1 bar by the Parrinello-Rahman barostat. Bonds were constrained by the LINCS algorithm with an integration time step of 2 fs. Nonbonded interactions use the Verlet scheme with a cut-off of 1.4 nm, while long-range electrostatics were treated with the particle mesh Ewald algorithm. Simulations of EBS1*/IBS1* and EBS1*/dIBS1* (both with and without Mg^{2+}) were run for 2 μs each. Convergence was assessed by computing the root mean square deviation over the nucleic acids and time binned histograms of the sugar puckers, showing that the repuckering time is much shorter than the overall simulation length. The pseudorotation angles of the riboses were calculated with PLUMED 2.5.2⁸⁷ and assigned to one of ten puckering modes (phase angle increments of 36°)⁴⁹ according to the definition by Huang et al.⁸⁸ (Supplementary Eqs. 15–18), which are related to the one introduced originally by Altona and Sundaralingam⁸⁹. The spatial distribution of Mg^{2+} ions, starting from a set of experimentally observed positions, is visualized using a Gaussian kernel density estimate. The point cloud is normalized by the ion

concentration, color coded by the density, and mapped onto the NMR structure (Supplementary Fig. 16).

NMR linewidth analysis. Linewidths and peak volumes in ^1H - ^1H]TOCSY spectra were calculated using NMRFAM-Sparky⁹⁰ by fitting a set of two-dimensional Gaussians to the assigned peaks.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request. The source data underlying Figs. 1d–g, 2a–g, 3a, b, 4c, 5b, c and Supplementary Figs. 1, 2, 7, 9, 15, 17–22 are provided as a Source Data file.

Code availability

A Matlab-based software package for video processing, evaluation of single-molecule trajectories, histogram, and dwell time distribution analysis is freely available at <https://github.com/RNA-FRETtools/MASH-FRET.git>. Documentation of the entire workflow of MASH-FRET can be found at <https://rna-frettools.github.io/MASH-FRET/>.

Received: 21 May 2019; Accepted: 15 November 2019;

Published online: 08 January 2020

References

- Pyle, A. M. Group II intron self-splicing. *Annu. Rev. Biophys.* **45**, 183–205 (2016).
- Jacquier, A. & Michel, F. Multiple exon-binding sites in class II self-splicing introns. *Cell* **50**, 17–29 (1987).
- Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45–50 (2004).
- Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature* **300**, 719–724 (1982).
- Su, L. J., Qin, P. Z., Michels, W. J. & Pyle, A. M. Guiding ribozyme cleavage through motif recognition: the mechanism of cleavage site selection by a group II intron ribozyme. *J. Mol. Biol.* **306**, 655–668 (2001).
- Toor, N., Rajashankar, K., Keating, K. S. & Pyle, A. M. Structural basis for exon recognition by a group II intron. *Nat. Struct. Mol. Biol.* **15**, 1221–1222 (2008).
- Chan, R. T., Robart, A. R., Rajashankar, K. R., Pyle, A. M. & Toor, N. Crystal structure of a group II intron in the pre-catalytic state. *Nat. Struct. Mol. Biol.* **19**, 555–557 (2012).
- Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514**, 193–197 (2014).
- Marcia, M. & Pyle, A. M. Principles of ion recognition in RNA: insights from the group II intron structures. *RNA* **20**, 516–527 (2014).
- Sigel, R. K. O., Vaidya, A. & Pyle, A. M. Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.* **7**, 1111–1116 (2000).
- Gordon, P. M. & Piccirilli, J. A. Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nat. Struct. Biol.* **8**, 893–898 (2001).
- Deininger, P. L. & Batzer, M. A. Mammalian retroelements. *Genome Res.* **12**, 1455–1465 (2002).
- Xiang, Q., Qin, P. Z., Michels, W. J., Freeland, K. & Pyle, A. M. Sequence specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry* **37**, 3839–3849 (1998).
- Somarowthu, S., Legiewicz, M., Keating, K. S. & Pyle, A. M. Visualizing the a15y group IIB intron. *Nucleic Acids Res.* **42**, 1947–1958 (2014).
- Kowerko, D. et al. Cation-induced kinetic heterogeneity of the intron–exon recognition in single group II introns. *Proc. Natl Acad. Sci. USA* **112**, 3403–3408 (2015).
- Börner, R., Kowerko, D., Miserachs, H. G., Schaffer, M. F. & Sigel, R. K. O. Metal ion induced heterogeneity in RNA folding studied by smFRET. *Coord. Chem. Rev.* **327–328**, 123–142 (2016).
- Woodson, S. A. Compact intermediates in RNA folding. *Annu. Rev. Biophys.* **39**, 61–77 (2010).
- Bowman, J. C., Lenz, T. K., Hud, N. V. & Williams, L. D. Cations in charge: magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.* **22**, 262–272 (2012).

19. Hyeon, C., Lee, J., Yoon, J., Hohng, S. & Thirumalai, D. Hidden complexity in the isomerization dynamics of Holliday junctions. *Nat. Chem.* **4**, 907–914 (2012).
20. Fiore, J. L., Holmstrom, E. D. & Nesbitt, D. J. Entropic origin of Mg²⁺-facilitated RNA folding. *Proc. Natl Acad. Sci. USA* **109**, 2902–2907 (2012).
21. Suddala, K. C., Wang, J., Hou, Q. & Walter, N. G. Mg(2+) shifts ligand-mediated folding of a riboswitch from induced-fit to conformational selection. *J. Am. Chem. Soc.* **137**, 14075–14083 (2015).
22. Bokinsky, G. & Zhuang, X. Single-molecule RNA folding. *ACS Chem. Biol.* **38**, 566–573 (2005).
23. Montange, R. K. & Batey, R. T. Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.* **37**, 117–133 (2008).
24. Panja, S., Hua, B., Zegarra, D., Ha, T. & Woodson, S. A. Metals induce transient folding and activation of the twister ribozyme. *Nat. Chem. Biol.* **13**, 1109–1114 (2017).
25. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
26. Rieder, U., Kreutz, C. & Micura, R. Folding of a transcriptionally acting preQ1 riboswitch. *Proc. Natl Acad. Sci. USA* **107**, 10804–10809 (2010).
27. Staple, D. W. & Butcher, S. E. Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* **3**, e213 (2005).
28. Lambowitz, A. M. & Zimmerly, S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* **3**, a003616 (2011).
29. Eskes, R., Yang, J., Lambowitz, A. M. & Perlman, P. S. Mobility of yeast mitochondrial group II introns. Engineering a new site specificity and retrohoming via full reverse splicing. *Cell* **88**, 865–874 (1997).
30. Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381**, 332–335 (1996).
31. Haack, D. B. et al. Cryo-EM structures of a group II intron reverse splicing into DNA. *Cell* **178**, 612–623 (2019).
32. Qu, G. et al. Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.* **23**, 549–557 (2016).
33. Solomatin, S. V., Greenfeld, M., Chu, S. & Herschlag, D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463**, 681–684 (2010).
34. Kruschel, D., Skilandat, M. & Sigel, R. K. O. NMR structure of the 5' splice site in the group IIB intron Sc.ai5γ—conformational requirements for exon–intron recognition. *RNA* **20**, 295–307 (2014).
35. Skilandat, M. & Sigel, R. K. O. The role of Mg(II) in DNA cleavage site recognition in group II intron ribozymes: solution structure and metal ion binding sites of the RNA–DNA complex. *J. Biol. Chem.* **289**, 20650–20663 (2014).
36. Draper, D. E. A guide to ions and RNA structure. *RNA* **10**, 335–343 (2004).
37. Lesnik, E. A. & Freier, S. M. Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry* **34**, 10807–10815 (1995).
38. Hall, K. B. & McLaughlin, L. W. Thermodynamic and structural properties of pentamer DNA–DNA, RNA–RNA and DNA–RNA duplexes of identical sequence. *Biochemistry* **30**, 10606–10613 (1991).
39. Rauzan, B. et al. Kinetics and thermodynamics of DNA, RNA, and hybrid duplex formation. *Biochemistry* **52**, 765–772 (2013).
40. Misra, V. K. & Draper, D. E. The linkage between magnesium binding and RNA folding. *J. Mol. Biol.* **317**, 507–521 (2002).
41. Bokinsky, G. et al. Single-molecule transition-state analysis of RNA folding. *Proc. Natl Acad. Sci. USA* **100**, 9302–9307 (2003).
42. Bartley, L. E., Zhuang, X., Das, R., Chu, S. & Herschlag, D. Exploration of the transition state for tertiary structure formation between an RNA helix and a large structured RNA. *J. Mol. Biol.* **328**, 1011–1026 (2003).
43. Cunha, R. A. & Bussi, G. Unraveling Mg²⁺-RNA binding with atomistic molecular dynamics. *RNA* **23**, 628–638 (2017).
44. Pan, J., Thirumalai, D. & Woodson, S. A. Folding of RNA involves parallel pathways. *J. Mol. Biol.* **273**, 7–13 (1997).
45. Schmid, S., Götz, M. & Hugel, T. Single-molecule analysis beyond dwell times: demonstration and assessment in and out of equilibrium. *Biophys. J.* **111**, 1375–1384 (2016).
46. Wang, S. & Kool, E. T. Origins of the large differences in stability of DNA and RNA helices: c-5 methyl and 2'-hydroxyl effects. *Biochemistry* **34**, 4125–4132 (1995).
47. Batey, R. T., Rambo, R. P. & Doudna, J. A. Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed.* **38**, 2326–2343 (1999).
48. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit. The A-minor motif. *Proc. Natl Acad. Sci. USA* **98**, 4899–4903 (2001).
49. Saenger, W. *Principles of Nucleic Acid Structure* 1st edn (Springer, New York, 1984).
50. Karplus, M. Contact electron–spin coupling of nuclear magnetic moments. *J. Chem. Phys.* **30**, 11–15 (1959).
51. Clay, M. C., Ganser, L. R., Merriman, D. K. & Al-Hashimi, H. M. Resolving sugar puckers in RNA excited states exposes slow modes of repuckering dynamics. *Nucleic Acids Res.* **45**, e134 (2017).
52. Olson, W. K. & Sussman, J. L. How flexible is the furanose ring? 1. A comparison of experimental and theoretical studies. *J. Am. Chem. Soc.* **104**, 270–278 (1982).
53. Duchardt, E. & Schwalbe, H. Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by MNMR 13C relaxation. *J. Biomol. NMR* **32**, 295–308 (2005).
54. Gyi, J. I., Lane, A. N., Conn, G. L. & Brown, T. Solution structures of DNA–RNA hybrids with purine-rich and pyrimidine-rich strands: comparison with the homologous DNA and RNA duplexes. *Biochemistry* **37**, 73–80 (1998).
55. Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S. & Reid, B. R. The DNA strand in DNA–RNA hybrid duplexes is neither B-form nor A-form in solution. *Biochemistry* **32**, 4207–4215 (1993).
56. Cowan, J. A. Coordination chemistry of magnesium ions and 5S rRNA (*Escherichia coli*): binding parameters, ligand symmetry, and implications for activity. *J. Am. Chem. Soc.* **113**, 675–676 (1991).
57. Peebles, C. L. et al. A self-splicing RNA excises an intron lariat. *Cell* **44**, 213–223 (1986).
58. Huang, H.-R. et al. The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc. Natl Acad. Sci. USA* **102**, 163–168 (2005).
59. Séraphin, B., Simon, M., Boulet, A. & Faye, G. Mitochondrial splicing requires a protein from a novel helicase family. *Nature* **337**, 84–87 (1989).
60. Karunatilaka, K. S., Solem, A., Pyle, A. M. & Rueda, D. Single-molecule analysis of Mss116-mediated group II intron folding. *Nature* **467**, 935–939 (2010).
61. Wank, H., SanFilippo, J., Singh, R. N., Matsuura, M. & Lambowitz, A. M. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol. Cell* **4**, 239–250 (1999).
62. Zimmerly, S. & Semper, C. Evolution of group II introns. *Mob. DNA* **6**, 7 (2015).
63. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell. Biol.* **5**, 451–463 (2004).
64. Zosel, F., Mercadante, D., Nettels, D. & Schuler, B. A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nat. Commun.* **9**, 3332 (2018).
65. Aizawa, Y., Xiang, Q., Lambowitz, A. M. & Pyle, A. M. The pathway for DNA recognition and RNA integration by a group II intron retrotransposon. *Mol. Cell* **11**, 795–805 (2003).
66. Steffen, F. D., Börner, R., Freisinger, E. & Sigel, R. K. O. Stick, flick, click: DNA-guided fluorescent labeling of long RNA for single-molecule FRET. *CHIMIA* **73**, 257–261 (2019).
67. Zhao, M. et al. Site-specific dual-color labeling of long RNAs for single-molecule spectroscopy. *Nucleic Acids Res.* **46**, e13 (2018).
68. Liu, Y. et al. Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature* **522**, 368–372 (2015).
69. Salim, N. et al. Thermodynamic and kinetic analysis of an RNA kissing interaction and its resolution into an extended duplex. *Biophys. J.* **102**, 1097–1107 (2012).
70. Šponer, J. et al. RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* **118**, 4177–4338 (2018).
71. Steffen, F. D., Sigel, R. K. O. & Börner, R. An atomistic view on carbocyanine photophysics in the realm of RNA. *Phys. Chem. Chem. Phys.* **18**, 29045–29055 (2016).
72. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507–516 (2008).
73. Hadzic, M. C. A. S., Börner, R., König, S. L. B., Kowanko, D. & Sigel, R. K. O. Reliable state identification and state transition detection in fluorescence intensity-based single-molecule Förster resonance energy-transfer data. *J. Phys. Chem. B* **122**, 6134–6147 (2018).
74. König, S. L. B. et al. BOBA FRET: bootstrap-based analysis of single-molecule FRET data. *PLoS ONE* **8**, e84157 (2013).
75. Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107–115 (1935).
76. Schmid, S. & Hugel, T. Efficient use of single molecule time traces to resolve kinetic rates, models and uncertainties. *J. Chem. Phys.* **148**, 123312 (2018).
77. Börner, R. et al. Simulations of camera-based single-molecule fluorescence experiments. *PLoS ONE* **13**, e0195277 (2018).
78. Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).

79. Dolinsky, T. J. et al. PDB2PQR. Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).
80. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
81. Cornell, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
82. Pérez, A. et al. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **92**, 3817–3829 (2007).
83. Banáš, P. et al. Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **6**, 3836–3849 (2010).
84. Zgarbová, M. et al. Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 (2011).
85. Joung, I. S. & Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
86. Li, P., Roberts, B. P., Chakravorty, D. K. & Merz, K. M. Rational design of particle mesh Ewald compatible Lennard–Jones parameters for +2 metal cations in explicit solvent. *J. Chem. Theory Comput.* **9**, 2733–2748 (2013).
87. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
88. Huang, M., Giese, T. J., Lee, T.-S. & York, D. M. Improvement of DNA and RNA sugar pucker profiles from semiempirical quantum methods. *J. Chem. Theory Comput.* **10**, 1538–1545 (2014).
89. Altona, C. & Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94**, 8205–8212 (1972).
90. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

Acknowledgements

We thank Ben Schuler, Helmut Grubmüller and Mauro Schilling for discussions on single-molecule kinetics and molecular dynamics simulations; Silke Johannsen for suggestions on NMR data analysis; Sonja Schmid for inputs on the ensemble HMM; Christian G. Hübner for providing computational resources for data analysis. This work was supported by the European Research Council (ERC Starting Grant MIRNA No. 259092, to R.K.O.S.) and the Forschungskredit of the University of Zurich (FK-17-098 to

F.D.S., FK-14-096/15-96 to R.B.). We are thankful for further financial support by the Swiss National Science Foundation (to R.K.O.S.) and the Stiftung für Wissenschaftliche Forschung (to R.K.O.S. and R.B.) of the University of Zurich.

Author contributions

F.D.S., D.K., R.B., and R.K.O.S. designed the research. M.K. collected single-molecule FRET data. F.D.S., M.K., and R.B. analyzed and interpreted the data. F.D.S. and R.A.C. ran and analyzed MD simulations. R.B., D.K., and R.K.O.S. supervised the work. F.D.S. and R.B. wrote the paper with the help of all authors.

Competing interest

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-13683-4>.

Correspondence and requests for materials should be addressed to R.B. or R.K.O.S.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020