

information needed for such recognition is not yet available for most viruses, and crystal structures of drugs bound to RNA structures are rare (and not necessarily representative); new molecular-level understanding of such binding is a critical need. Structural studies on RNA are further complicated by the inherent flexibility of RNA molecules, which requires an understanding of their dynamics not just their ground state conformation. Consequently, simple molecular docking will not suffice; by contrast molecular dynamics potentially allows the energy landscape and structural flexibility to be probed. Herein we employ molecular dynamics to explore in detail, for the first time, a nano-scale drug inserting into a bulge in a UTR viral RNA, replicating experimental observations and gaining fundamental new insight into the dynamics of the RNA and of the drug entry process; crucial intelligence to inform design of new UTR-structure-targeting drugs. The nano-scale drugs studied are supramolecular cylinders, which not only have unprecedented RNA bulge-binding ability but are the first in class of metallo-supramolecular architectures to show potent anti-viral activity in cellular assays.³⁵ There is a growing interest in the application of metallo-supramolecular architectures in biology.^{36–41}

Results and discussion

As a suitable UTR structure for our studies we chose HIV-1 TAR RNA which is both experimentally well described and representative of wider viral UTR structural motifs. As a drug we chose a nanoscale metallo-supramolecular cylinder because it is unique as a nano-drug that has previously been

crystallographically characterised when bound within an RNA cavity (a perfect three-way junction (3WJ)) (Fig. 1).^{42,43} It is also unique in threading through an RNA cavity, interacting with all of the internal structure. These cylinders also bind bulge structures in RNA, prevent TAT protein from recognizing the binding site in the TAR sequence of HIV^{35,44} and arrest HIV replication in mammalian cells.³⁵ The strong evidence of binding and in-cell efficacy, makes this an ideal test-bed to investigate whether molecular dynamics simulations can identify the processes that underpin the kinetics of targeting highly flexible RNA strands. At the same time, it provides a suitable challenging size of drug, and one with large, nanoscale, 3-dimensional molecular surfaces whose match and strong binding to the 3D shape of RNA structural motifs should collapse the RNA's conformational landscape to a non-functional (impotent) state. The cylinder exists in two enantiomeric forms, both of which bind RNA bulges. Experimental X-ray crystal structures are also available for unbound cylinders;^{45,46} the calculated DFT structures herein are almost identical.

Simulations of RNAs (uncomplexed)

For multi-microsecond simulations, classical MD forcefields describing the dynamics of both RNA and DNA have until very recently^{47–52} been found to be unsatisfactory – over such time-scales they induced structures not seen experimentally. With longer simulations being available, the conformational space sampled can deviate further from the absolute minimum energy point and explore the importance of non-covalent interaction dynamics as pi-stacking and hydrogen bonding.^{53–57} However new forcefields^{50–52} have become available and we show now that the Rochester-Mathews force-field^{51,53} can be used to simulate RNA over long timescales, reproducibly, not only for free RNA but for drug-bound complexes. The Rochester-Mathews forcefield is publicly available giving it the potential to be accessed and implemented by all. It uses the same underpinning level of DFT theory as that applied to metal-containing cylinder coordination compounds creating an overall consistency. Moreover, there are ways to accurately model NMR ensembles of RNA structures without the need of extensive MD simulations.⁵⁸ Collectively we accumulated over 200 μ s of simulated time; such long and data-rich simulations on a flexible RNA system, brought new challenges in analysis. We address these by applying Markov state modeling^{59,60} to the problem and show that this enables us to identify stable and metastable conformations among the millions of frames.

Overall we have performed 123 simulations of at least 1 μ s and up to 10 μ s, overall \sim 200 μ s including several shorter runs with varying initial conditions. To analyse the vast volume of data, over 200 000 000 coordinate frames, we employ the PyEmma workflow⁶⁰ and Markov State Modelling (MSM). This involves reducing the dimensionality by choosing appropriate features of the simulation and identifying macrostates of each simulation using MSM and extracting those metastable structures with Perron-cluster cluster analysis (PCCA). Those

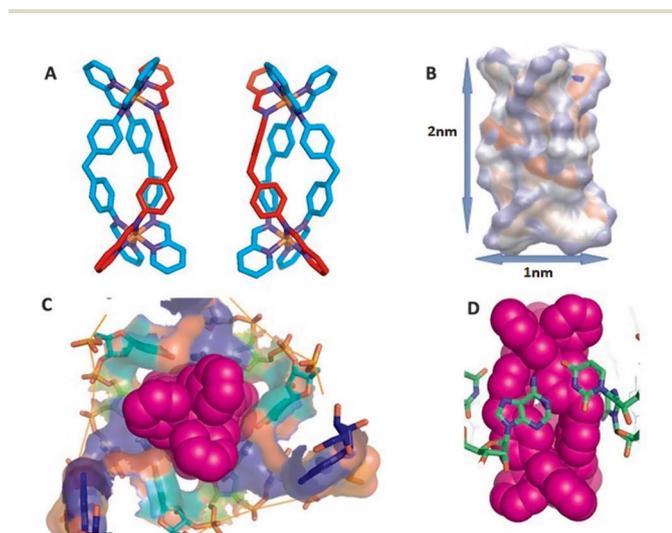


Fig. 1 (A) *P* and *M* enantiomers of the iron cylinder $[\text{Fe}_2\text{L}_3]^{4+}$ optimized by DFT. (B) Distribution of partial charges for *P* enantiomer as calculated by *ssb-d-D3/Def2-SVP* level of DFT theory, visualized by VMD, and also showing approximate cylinder size. (C) Surface of the RNA 3-way junction cavity stabilized by the *M* enantiomer of the cylinder from the crystal structure *pdb 4JIY*.⁴³ (D) Stacking of RNA bases to the cylinder in the centre of the 3-way junction in *pdb 4JIY*.⁴³ Analogous stacking is also seen with cylinders located at the terminal base pairs of the strands (see ESI part B†). Hydrogen are omitted for clarity in A, C and D.



extracted structures and the whole simulation are also presented in the Leontis–Westhof⁶¹ nomenclature using Barnaba.⁶² A detailed explanation of this workflow is included in ESI.†

To confirm the ability of the forcefield^{50–52} to conserve structural features of viral stem-loop RNAs (as observed, dynamically, in NMR), and to establish the effectiveness of our approach to analysis, we first explored the dynamics of poliovirus stem loop (pdb: 2GRW)⁶³ coxsackievirus stem loop (pdb: 1RFR)⁶⁴ and HIV2-TAR (pdb: 1AJU)⁶⁵ RNA with no bound drugs. The simulations reliably reproduced NMR observations for the stem loops (including regions of non-Watson–Crick pairing) and the predicted effect of a small bound ligand on the HIV2 TAR. Indeed for poliovirus stem-loop, the MD simulations reveal and explain features that are observed in the NMR structural data, but have not previously been satisfactorily captured in the deposited conformations, and for HIV2-TAR shows how the ligand-free RNA structure deviates from the conformation of the bound state, demonstrating the effect a binding molecule can have on an RNA structure: a detailed analysis of these free RNA simulations is included in ESI.†

HIV1-TAR. We now turned to a more in depth study of the dynamics of our test UTR stem-loop, the HIV-1 TAR RNA. While in the coxsackievirus, poliovirus and HIV-2 simulations we had focused on the proposed ground state of the RNA as the starting point for the simulations, now we expanded our attention beyond the ground state to look also at other conformations within the experimentally suggested (NMR; pdb 1ANR) structures. In an effort to avoid introducing biases and acceleration methods to the simulation we chose to explore the conformation landscape by starting simulations from different local minima as described in the original HIV-1 TAR NMR solution structure.⁶⁶ There are 20 NMR solutions proposed and we started from five such minima (first, third, fourth, seventh and twelfth). For each of these higher energy solutions a 2 μ s simulation retained the characteristics consistent with the NMR structure and did not deviate into unnatural (loosely bound) conformations. From each starting point similar features can be observed as the simulation proceeds which indicates that the forcefield can reproduce transitions within the landscape of a few μ s per solution. These unbiased MD simulations capture the conformational changes of the RNA across the energy landscape for the first time, and clearly reveal the variation possible in the RNA structure and the range of conformations sampled (and which a drug could encounter and sample). Importantly, time-lagged independent component analysis (TICA) of the trajectories (Fig. 2) revealed a broad energy minimum in the ground state which shows that small perturbations in the conformation have minimal effect on the energy. Moreover, a single 10 μ s long simulation (as well as an independent 6 μ s long simulation) of the ground state reveals the conformational richness near to the minimum. These observations highlight the limitation of a simple docking approach for flexible RNAs.

Across the simulations, the helical regions remain relatively stable with strong WC base pairing. The only stem base pair not retaining the WC pairing is A22:U40, which often drifts apart as

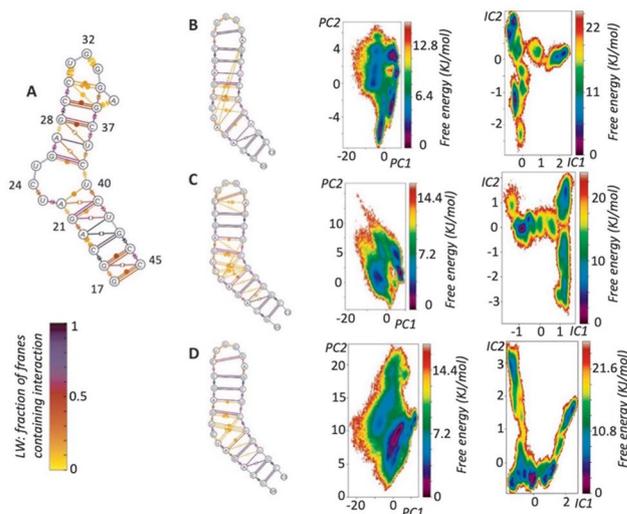


Fig. 2 (A) Summary of the 1ANR 20 NMR solutions presented in Leontis Westhof (LW) nomenclature. (B) LW nomenclature of 10 μ s simulation and PCA and TICA free energy surfaces, demonstrating: how the simulation reproduces 1ANR NMR structure but also reveals transient pairings (LW yellow) not well defined by (but nevertheless noted in) NMR; the greater richness of information in TICA analysis over PCA; the many conformations (TICA minima) that are accessible in the simulation at this temperature (310 K). (C) LW nomenclature of 6 μ s simulation and PCA and TICA free energy surfaces. (D) Combined results of seven 2 μ s simulations (see ESI Methods†) starting with different NMR solutions.

the U40 retains strong stacking with C39. It is often the case that U40 seems to be in the 2nd rather than the first stem.

While a variety of transient base-pairings of all types were observed in the bulge region, as expected from the experimental NMR observations, no new stable base pairings were observed apart from that between C30 and G33/4 which is not observed by NMR but is observed in gel electrophoresis. On the unbulged strand stacking is strong and continuous, but this is a lot less evident on the bulge strand. Of the three bulged nucleotides, U23 and C24 are more likely to stack whereas U25 is the most likely to be fully outside the helix and can even create long range interactions with the loop nucleotides (G33 and A35) creating a transient folding up of the second stem. Such a folding was not observed in the HIV-2 TAR simulation.

The loop region is characterised by limited stacking between bases and common WC pairing between C30 and G34. Transient non-WC pairing can include C30 *cis* or *trans* WC/Hoogsteen to A35.

Examining the runs starting from the different local energy minima; the first simulation starting from 1anr1 identified 3 distinct states, that can be recognised even by the PCA analysis. All are energetically and conformationally close together as seen by the RMSD and ERMSD. PCCA analysis shows one to be in much higher occupancy, clearly the ground state. A second simulation also starting from 1anr1 sampled a wider conformational space. Base pairing of stems was retained although stacking between C19:G43 and A20:U21 was not, although it is observed in the NMR. After that, stacking does continue all the



way to the loop. At the loop a few different conformations were sampled that mostly gave rise to the different MSM states identified. C30 base pairs with either G33 or G34. In the bulge region U40 is stacked strongly with C39 but not always to C41 and transient, short lived pairing takes place between all bulge residues and either of C39 and U40, with pairing types including both sugar and Hoogsteen edges as well as in the trans position. MSM analysis gave 6 different states.

The simulation starting from the third NMR solution, 1anr3, yielded 5 local minima in the TiCA projections and CK test allowed for 5 states in MSM analysis. Overall, stacking and pairing throughout the stems is conserved and transient pairing within the bulge region is similar to that of the previous run. Most importantly the second state is very reminiscent of the ground state.

As we planned to apply significant external forces to the RNA structure, by introducing the cylinder into the system, we also tested the behaviour of the forcefield with higher NMR energy solutions. An experimental analysis of higher energy RNA conformations (when in the presence of a bound ligand) has been discussed by Orlovsky *et al.*⁶⁷ In that work, 3 nucleotide bulges are observed to adopt multiple conformations; we replicate these multiple conformations in our simulations (Fig. 2B–D, ESI†) providing further experimental validation of our model. Going up the energy ladder from the starting conformation one might expect to encounter more structures that deviate significantly from the ground state. Nevertheless starting from the fourth solution, 1anr4, most of the important structural features were retained. Pairing and stacking remains consistent with the exception of the U23 to C24 stacking. PCA revealed 3 stationary points which become 5 with TiCA. Also notable is that from this state up, examining the first 4 TiCA vectors instead of just two showed much higher diversity. In the loop, pairing C30:G34 is seen again, as well as the usual transient non-traditional pairing, but now interactions between U23 and U38 and trans Hoogsteen to sugar between U23 and C39 are observed. Stacking of U40 to C39 remains strong but stacking of U23 to C24 was less prevalent.

The seventh, 1anr7, and twelfth, 1anr12, structures are quite different from the ground state and this brings challenges for the simulation: specifically, the loss of A helix structures which is characterised by the overall elongation of G17 to G33 distance can be testing to any forcefield. Nevertheless, starting from 1anr7, the stacking and pairing remains consistent. PCA identified 2 states whereas TiCA suggested 6 states and the CK test is also passed with 6 states. The first 4 states are reminiscent of the ground state with different loop configurations, namely sugar to Hoogsteen between C30 and A35, or less often trans WC to Hoogsteen. In the other two states, U25, which generally points outside the bulge can create temporary long-range interactions with loop residue G33.

Starting from 1anr12, which is also very elongated with a sharp backbone kink in the bulge area, also retrieved most of the properties of the ground state. Pairing and stacking remain consistent for the stems. In the loop the common C30 to G34 pairing is stable along with a transient Hoogsteen to sugar between A35 and C30. In the bulge region stacking between C39

and U40 is strong and most of the transient non traditional base pairings are also seen. PCA revealed 2 states whereas TiCA revealed 5.

The results demonstrate that the forcefield can satisfactorily retain characteristics of the structure as described by the NMR experimental constraints.

In addition to the unbound 1anr structure, there are some TAR RNA structures with various different bound drugs, and so for comparison we also explored as a starting point one such structure (the only solution of pdb; 1UUI)⁶⁸ from which we had removed the drug. The structure, after removing the ligand, has some differences with the 1anr structure: pairing on the stems is the same, but stacking is disturbed before the bulge, probably since U23 is WC paired with A27.

When using this as the starting point for a 2 μ s simulation, the loop folded back onto the bulge (from which the ligand had been removed) forming interactions from U23 and C24 to A35, and the stem remained folded for much of the simulation. The bulge stacking did not return to the transient pairings seen in the earlier simulations. PCA analysis of the simulation revealed 3 states and TiCA 6, which was also passed the CK test on with MSM with the sixth state being ground state of this run. The simulation demonstrates how ligand binding can modify the structure and dynamics of the TAR RNA and again highlights that docking, while a useful guide, may miss key features and opportunities. The Rochester forcefield⁵² behaved well for every case of RNA molecular dynamics, even in cases outside the ground state of the structure in question.

Cylinders binding to HIV1-TAR

Docking studies. Disney has recently used docking to screen libraries of small molecules binding to RNAs including TAR.^{3,52} We initially undertook simple docking calculations as described in methods using all 20 structures from pdb; 1anr TAR RNA NMR study. The results are dominated by different forms of bulge region binding. While the two enantiomers do show slightly different binding energies, the Autodock Vina⁶⁹ as other docking software (used as it is one of few that allow incorporation of first row d-block metal centres) as other docking software tends to underestimate the electrostatic contribution when a charged molecule is involved. Nevertheless the docking scores are high compared to other small molecule drugs assessed by this method reflecting the larger available surface of the cylinder.

It is interesting to compare the results of docking with overall results of subsequent MD simulations. In particular in the MD simulations, capping of the open terminal bases is a transient, but relatively stable (more than 2 μ s) location seen with both enantiomers. Although only a local minimum in the interaction of cylinders with TAR it highlights the limitations of docking in targeting nucleic acids because, across all 20 NMR solutions of TAR RNA, the terminal bases are coplanar only in one (the ninth). Consequently only in this structure solution does the docking reveal the end capping as a potential binding site. So docking outcomes are constrained by the rigid RNA structure(s) used in the docking, whereas in reality – as we shall



see – RNAs are highly fluxional and dynamic molecules that access much structural space. Thus while such simple docking studies are valuable for high throughput screening they might be more suited to small molecules where the molecule is less likely to have a major effect on RNA conformation. For the larger cylinders the size of the binding surface means that induced conformational change is more likely and so more sophisticated MD can offer greater insight into the interaction. Crucially, while the docking showed bulge region binding, bulge insertion by the cylinder was not observed.

Molecular dynamics simulations. To examine the interaction between TAR and the cylinders, simulations (112) started with the cylinder (DFT optimized – Fig. 1A and B) in random places 1 nm away from the RNA as well as from sites identified by docking studies with initial TAR conformations derived from multiple experimental 1ANR solutions examined earlier.

The size of the cylinder restricts how rapidly it will move between sites (local minima) in the simulations' timescale. Consequently a single simulation would fail to explore all binding sites and conformations. Instead we take the quite different approach of using multiple simulations (1–10 μ s) from different starting points which allows the cylinder to explore a much greater range of RNA conformations and to encounter multiple potential binding sites. By combining this with Markov state modelling analysis we are now able to explore effectively the dynamic conformational landscape of the TAR RNA – cylinder complex.

The simulations show the cylinder moving up, down and around the DNA exploring different sites and positions, and moving between them, until it ultimately inserts into the 3-base bulge. Such a dynamic exploration of different positions is what is anticipated for such a polycation with a sophisticated RNA polyanion in these timescales. There are a number of different, kinetically-accessible, positions that the cylinder explores and occupies transiently *en route*, of which some represent local minima with longer residence times (though still transient) and are identified from the MSM analysis (Fig. S8 and S9[†]). We and we will describe these briefly before turning to the 3-base bulge that is the ultimate binding site.

Transient end-stacking interactions. Often the cylinder (both enantiomers) found a local minimum, which it occupied for at least 1 μ s at a time, and in which it capped the terminal G17:C45 bases (Fig. 3). Some RNA forcefields have been suggested to over-emphasise base-stacking.^{70,71} However, in this RNA system this binding position is among the most accessible kinetically and, since such cylinder binding has also been observed in X-ray crystal structures,^{42,43} it demonstrates that the simulation is replicating an experimentally validated binding location. To assess how well the forcefield and the parameterisation (now including the cylinder) reproduces this binding as captured by the crystal structure we extracted the G17:C45 bases and the cylinder from a frame of the longest lived position and we then optimised that structure at the *ssb-d-D3/LANL2DZ* (DFT and semi-empirical (PM7)) and superimposed it on the binding mode extracted from a crystal structure. The overlap (Fig. 3) is extremely good, implying that the forcefield is working as desired, and that the crystallographically observed binding is

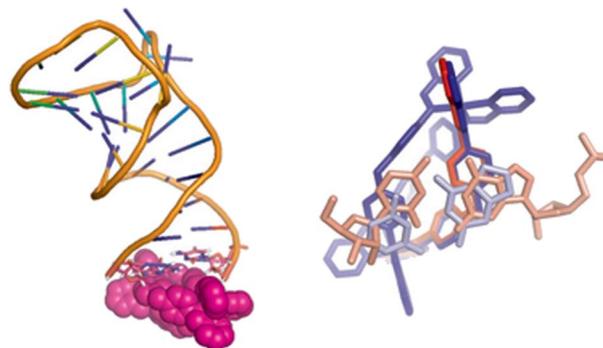


Fig. 3 Left: end-capping of the cylinder observed in an MD simulation. Right: the end-stacking experimentally observed in crystal structure 4JIY⁴³ (red), overlain with that observed in an MD simulation followed by DFT optimisation (blue).

reproduced. This end capping is to some extent a feature of using a shortened oligonucleotide both in these simulations and in X-ray crystal structures: it certainly does demonstrate the affinity of the cylinder for extended planar pi-surfaces, but such end capping sites are not so common in biologically active RNAs.

Transient groove interactions. The cylinder is commonly observed exploring the RNA grooves, primarily the groove of the first stem. The residence time for the *M* enantiomer on average is longer than for the *P* implying that the *M* enantiomer may have a higher affinity for the grooves although the kinetics were not adequately sampled to quantify difference.

Transient loop interactions. The cylinder can take advantage of unpaired open bases of the loop and interact transiently there (also seen in simulations with the coxsackievirus stem), but this is less commonly observed in the simulation compared to other locations. Loops are a common feature in RNA structures (and indeed in non-canonical DNA structures such as G-quadruplexes and i-motifs) but seem not to be a particular target for the cylinder, consistent with our experimental observations.

Transient interactions in the bulge area. The cylinder is most frequently found around or on the bulge (Fig. 4) in the simulation (and as confirmed by experimental data^{35,44}), with *M* and *P*

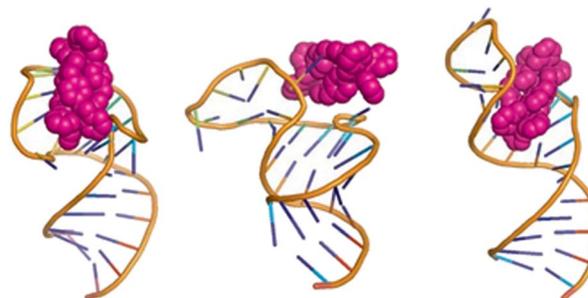


Fig. 4 Exemplar bulge-binding interactions observed in the simulations, *en route* to bulge insertion, including bridging from bulge to loop. The right hand figure has the cylinder in the position where a cyclic peptide has been observed to bind to TAR.



being very similar in their preference for this location. RNA conformations that involve the loop bridging to the bulge (U25) can be stabilised for longer (compared to free TAR), with the cylinder sitting on top of the bridge or mediating stacking. In the absence of the bridge, the cylinder can also sit between the bulge and the opposite RNA strand, in a position in which it opens up the base pairing protecting the TAT binding site. In the case that the cylinder sits on the bulge nucleotides, it stabilises the transient base pairing and dislocates the counterions that would normally reside there which leads to an overall elongated structure of the RNA with minimal helicity.

In this context it is noteworthy that Keene and Collins have explored the binding of a dinuclear ruthenium polypyridyl agent (but of quite different shape to the cylinders) to a TAR-like RNA and proposed that it might bind around the groove near the bulge.^{72,73} Given that the bulge-area is the most frequent location for the cylinder prior to bulge-insertion, it seems likely that this region could also be a preferred area of binding for other dinuclear complexes that cannot insert inside the bulge; for example differently shaped metallo-helices have been reported to not remain bound to TAR in electrophoresis,⁷⁴ in contrast to the bulge-inserting cylinders herein,^{35,44} and might be more loosely associated outside the bulge.

Bulge insertion. For both *M* and *P* enantiomers, insertion into the bulge is observed; once in the bulge the cylinder is strongly bound and remains there. In this unique binding mode, the cylinder sits in a V-shaped cleft (Fig. 5) that resembles the 3WJ structure (Fig. 1C). The effect of the binding is to restrict/collapse the conformational flexibility of the RNA, prevent the transient loop-bulge interactions and lessen the helicity of the stems. It is striking that, although this is the most stable binding mode in simulations, it fails to be identified in docking studies from any of the 20 1ANR solutions, because docking does not account for RNA flexibility. The bulge insertion and its effects are consistent with and explain both experimental RNase A footprinting results⁴⁴ and the ability of this cylinder to remain bound in electrophoresis when other metallo-helices do not.⁷⁴

The MD simulations also provide intriguing molecular-level insight into how an insertion is possible:

Entry mechanism for *M* enantiomer (Movies S1, S2; ‡ Fig. 6A–E). The cylinder first associates with the RNA outside

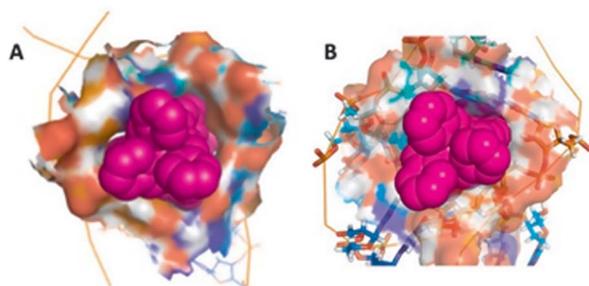


Fig. 5 The bulge insertion mode: the surface of the RNA cavity shows the extremely high contact surface for (A) the *M* enantiomer and (B) the *P* enantiomer, and the similarity to each other and to the 3WJ-binding (compare Fig. 1C).

the bulge (Fig. 6A and B). It interacts with the two base pairs at the bulge; A22–U40 and G26–C39. The G26–C39 base pair stacks onto a pair of phenyls (drawn from different strands of the cylinder; Fig. 6C). The A22–U40 pair is transient and we see it both paired and unpaired and interacting (stacking) with the cylinder with the U40 having a particular tendency to stack on a phenyl even when not paired (Fig. 6C and D). From here the mechanism of entry proceeds by two very similar processes, differing primarily in whether the A22–U40 is paired during entry or not. The entry process seems to be quicker when A22–U40 is paired, but entry can take place without this pairing (Fig. 6E). The stacking of the paired bases A22–U40, along with the stacking of paired G26–C39 to the cylinder is effectively a V-shaped cleft about the cylinder and is reminiscent of the stacking observed in the 3WJ structure. The bulge itself is initially folded (rather than open) (Fig. 6D) and neutralised by sodium cations, implicating the kinetic contribution of the ionic environment.

As the simulation proceeds, the sodium cations leave and the bulge opens. U25 and C24 are flipped out and stack with each other. The cylinder remains stacked in the V-shaped cleft afforded by U40 (or U40–A22) and C39–G26. The cylinder starts to slide around placing its pyridyls into the bulge; these pyridyls initially encounter the sugar of U23. U25 and C24 swing back and forth with U25 also encountering the pyridyls and transiently stacking with pyridyls as does A22. The crucial point of insertion involves the cylinder stacked with G26–C39, twisting around and inserting through the centre of the bulge (Fig. 6E). It does so facilitated by transient stacking interactions with U25, A22 and C41 which help to guide it into the cavity. With the cylinder now in the cavity, U40–A22 stack onto a pair of the cylinder phenyls, and so (re-)form the V-shaped cleft (now U40–A22; C39–G26) that is similar to two sides of the 3WJ structure. This process has been replicated in 5 independent simulations.

The MSM analysis of this entry process shows just two principal states; once the cylinder has moved from its location just outside and starts to open and enter the bulge, the energy landscape drops rapidly down into the final position where the cylinder is fully inserted and where it remains (Fig. 6M, S147 and ESI Table 5‡).

Entry mechanism for *P* enantiomer (Movies S3, S4; ‡ Fig. 6G–L). In the case of the *P* enantiomer from the same starting position (Fig. 6G), the entry mechanism is different but has similar features. The cylinder splits the U25 G26 bulge nucleotides and still stacks the G26–C39 base pair while stabilizing it (Fig. 6H and J). On the other side, the cylinder pyridyls press upon the A22:U40 base pair (Fig. 6K). Within 3 ns the base pair opens, the cylinder stacking aligns to G26–C39 and the A22–U40 base pair re-forms, now enclosing the cylinder in the bulge pocket (Fig. 6L). For the rest of the simulation the cylinder resides in the familiar triangle only this time it is splitting nucleotides U25 and G26 as opposed to C39 and U40 with the *M* enantiomer. U40 now plays a supportive role in stacking the cylinder phenyls and its base pairing with A22 becomes transient. This mechanism has been replicated in 4 independent simulations.

It is instructive that both cylinder enantiomers slide into the cleft down the RNA bases and locate in the V-shaped cleft of the



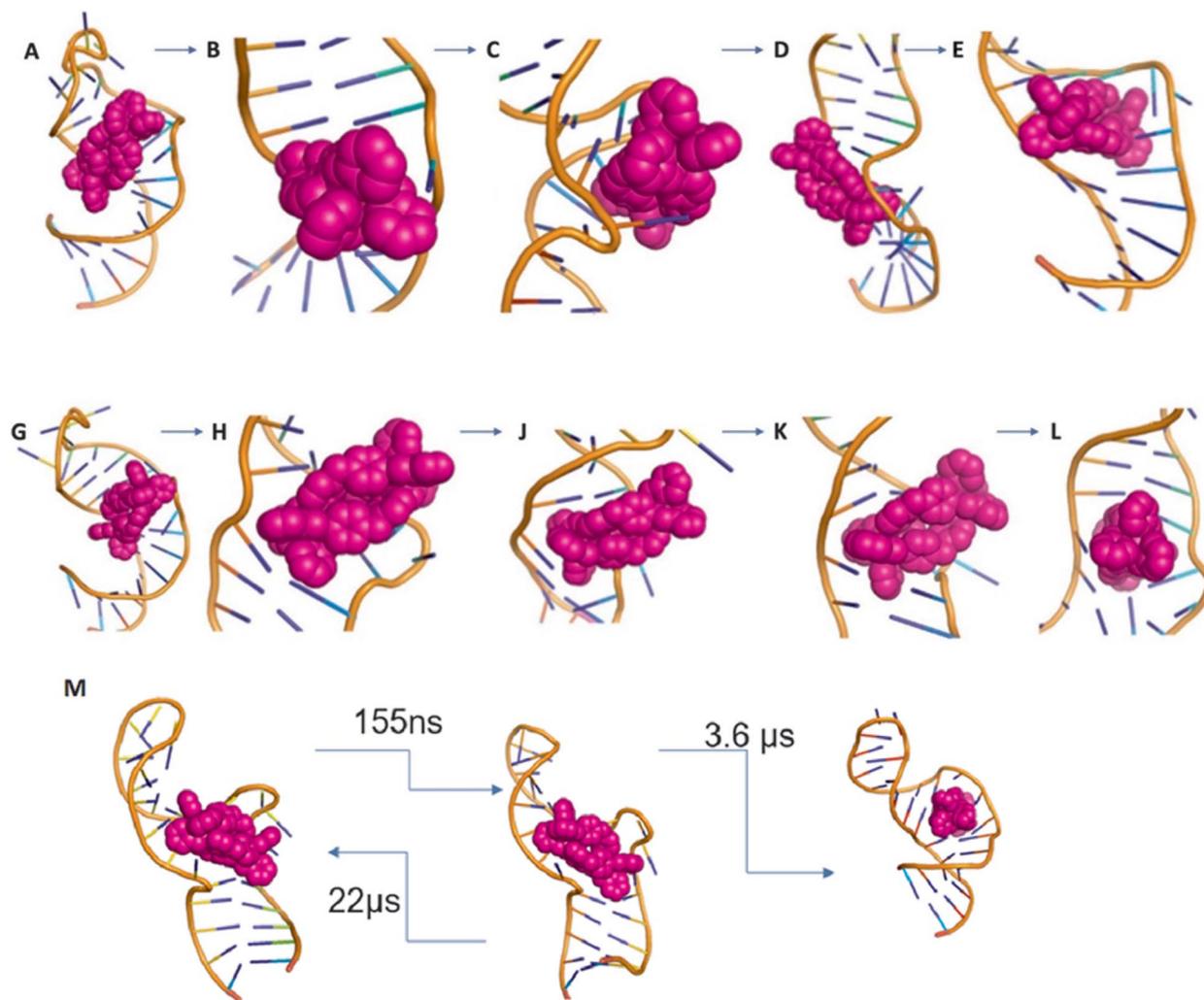


Fig. 6 (A–E) Entry of *M* enantiomer: (A) starting position of *M* cylinder on 1ANR1. (B) Cylinder rotates to split the U25 G26 and (C) aligns in parallel to the G26:C39 base pair (order of microseconds). (D) After relaxation of the backbone (order of microseconds), (E) the cylinder is inserted into the cavity (order of nanoseconds). In contrast to the *P* cylinder the *M* cylinder splits the C39 U40 and makes contact transiently stacking the 3 nucleotides of the bulge. (G–L) Entry of *P* enantiomer: (G) starting position of *P* cylinder on 1anr1. (H) Cylinder splits the CU nucleotides at the non-bulged strand and (J) pushes the AU base pair (order of microseconds). (K) The bp opens and the cylinder aligns parallel to the GC base pair (order of nanoseconds) and (L) after the AU closes the *P* cylinder is in the centre of the bulge. (M) Transition timescales for the *M* cylinder between states.

bulge which is similar to that in the 3WJ (Fig. 5). The longer range effect of the insertion is that the helicity of the second stem is disturbed which is consistent with the experimentally observed increased cutting of the C30:U31 by RNAase A.⁴⁴

This bulge insertion is a fascinating illustration of how a three dimensional nano-size agent might target the interior of an RNA structural feature, not by hydrogen bonding to the bases but rather by using its external pi-surfaces to recognize the surfaces inside the structure. To that extent the structure resembles a three-dimensional version of intercalation, and in that context it is notable that Barton has shown that the ‘light-switch’ intercalator $[\text{Ru}(\text{bpy})_2(\text{dppz})]^{2+}$, which doesn’t intercalate into duplex RNA, can bind at RNA mismatch sites,⁷⁵ where it is proposed to do so by insertion, with extrusion of the mispaired bases. The organic intercalator ethidium has been

proposed to bind one-base bulges in RNA,⁷⁶ and metal complexes bearing a ‘phi’ intercalator suggested to bind near the TAR bulge from cleavage experiments, though that is not yet well understood at a structural level.^{77–80} This insertion of a three-dimensional structure represents a unique and exciting approach to target RNA structures.

Considerations regarding free energy landscape of RNA-cylinder complex

The simulations suggest that the binding interaction between the cylinder and the TAR-RNA should be characterised as an “induced fit” interaction, meaning that the cylinder does not recognise the bulge cavity in the traditional lock-key manner but rather it induces the precise conformation of the RNA. This complicates the free energy landscape estimation. Although we



do get an idea of the landscape using TiCA and PCA we do not believe that the space is sufficiently sampled and therefore MSM probabilities only reflect the sampled space. Mmpbsa techniques cannot be used as removing the cylinder from the final complex exposes a large hydrophobic cavity and an RNA structure that is not in a minimum. Therefore in this paper we have focused on the kinetics and mechanics of the binding process and not on the free energy estimation of the binding. However, in other systems, metadynamics and transition path sampling (TPS) have previously been applied to study the interaction of metal complexes with nucleic acids and proteins.^{81,82}

Methods

DFT of cylinders

Density functional theory optimisation of the two cylinders were performed in Nwchem 6.8.1 (ref. 83) with SSB-D,⁸⁴ becke97-d,⁸⁵ and TPSSH⁸⁶ with D3 dispersion correction⁸⁷ for the first two and D3BJ for the last with of Def2-SVP basis set. The optimisation was performed under tight driver criteria and increased grid to xfine settings for convergence. Partial charge distribution on atomic positions was calculated with the ESP module under overall restrain of charge. Visualisation of the charge distribution at the surface was done in VMD 1.9.2 (ref. 88) on surface after converting the nwchem output .molden and .esp files to mol2.

Docking

Autodock vina⁶⁹ was used to create pdbqt files for all solutions of pdb 1ANR as well as the first solutions of coxackievirus stem loop and HIV-2. The cylinder structure after DFT optimisation was entered as a ligand – the searching box was big enough to contain the entire molecule and the cylinder (at least 20 Å away from the biomolecule). Exhaustiveness was set to 1000. Additional docking to just the terminal bases, specifying the docking box to the first 3 base pairs showed that only 2 out of the 20 solutions allowed for capping-mode docking.

Molecular dynamics simulations

Parametrisation of supramolecular cylinder: already DFT optimised geometries of the cylinders were split into 5 residues (3 ligands and 2 metal ions) that were fed to MCPB.py⁸⁹ that generated parameters for the metal centres at the wb97XD9/6-31G*⁸⁶ level of theory using Gaussian09 (ref. 90) as well as partial charges using RESP. The coordinate and parameter files were converted to gromacs using ParmEd (<http://parmed.github.io/ParmEd/html/index.html>).

Preparation of parameters with AMBER99SB⁶⁶ was achieved with pdb2gmx program of GROMACS 2019.2 (ref. 91) whereas for the ROC forcefield⁵¹ it was achieved using tleap program of Amber18 (ref. 92) and the files provided in ref. 51. The parameters and coordinates were then converted to gromacs using ParmEd.

In all systems, unless otherwise stated, the RNA was put in a dodecahedral box with edges at least 1.5 nm from the solute

filled with TIP3P water. Initial minimisation was carried to at least 500 kJ mol⁻¹ nm⁻¹ or 50 000 steps followed by heating and NVT equilibration for 1000 ps using V-rescale modified Berendsen thermostat, coupling the cylinder with the RNA at 310 K. All simulations use 2 fs time step and Parrinello–Rahman pressure coupling and PME electrostatics at 1.0 nm cut-off. Attempts to run the simulation with a 4 fs time step led quickly to blow up of the system, although 3 fs time step was more stable.

After completion the compressed trajectories (.xtc) were analysed to remove periodic boundary conditions and rotations using gromacs' trjconv program. After removing the water the trajectories were analysed with pyemma2.5.6 and pyemma 2.5.7,⁶⁰ barnaba.⁶² Free energy calculations used g_mmpbsa.⁹³

We also explored simulations for the ruthenium cylinder (total 17.3 μs) in place of the iron cylinder. The ruthenium cylinder behaved analogously in its binding, though its movement was slower due to the increased molecular mass.

Simulation analysis

To analyse the simulations and identify different micro-states on the energy landscape of each run, we followed the Pyemma workflow.⁶⁰ The workflow involves principal component analysis, time dependent component analysis, and Markov state modeling and Perron cluster cluster analysis.

To identify the best features to apply the workflow to, we explored a variety of potential different features to see which best captured the kinetic variance that occurred during the simulations:

1. Position of centre of mass (COM) of each residue is a low dimensional and relatively efficient way to capture different states, including simulations that involve one or more cylinders.
2. Taking advantage of the fact that each residue has an atom named N3, which is away from the backbone, we created a matrix of distances between these N3 atoms, which although high in dimensionality captures nearly all the kinetic variance. For the cylinder simulations, we also added the distances of the metal ions (Fe or Ru) and the resulting matrix can capture adequately the kinetics of the system during the simulation.
3. The distances between the phosphorus atoms in the backbone.

Of these approaches 2 proved the most useful and was applied to all the simulations.

For each simulation, Principal Component Analysis (PCA) was carried and the projections between the first 4 PCs are plotted, followed by time-lagged independent component analysis (TICA) for lag times 1 to 5000 steps. The lag time for which the fewer number of TICA dimensions were necessary to capture 95% of the kinetic variance was chosen for further analysis. The number of clusters was chosen by examining the convergence with regards to VAMP2 as described the original paper and <http://www.emma-project.org/latest/index.html>. Lag times for MSM model were chosen from the convergence at timescales of identified processes. Only models that used all of the states and could pass the Chapman–Kolmogorov test



were continued to Perron-cluster cluster analysis (PCCA) which led to extraction of states with certain probability and structure in pdb format. Not all simulations were long enough to produce an appropriate Markov state model, and it should be noted that the Markov state models as used here are meant to describe or sum up the particular simulations and not the whole system.

The extracted state and the full length of the simulation were analysed with Barnaba:⁶² all long production molecular dynamics runs, as well as states identified by PCCA, were analysed using barnaba resulting in 2D Leontis/Westhof classification⁶¹ of base interactions as well as E-RMSD as defined by barnaba software, RMSD and J-couplings.

Conclusions

This study provides an unprecedented platform to inform design of agents that target different important RNA structural motifs found in nucleic acid nanoscience and biology, such as this bulge cavity present in the UTR of many different viruses. We show that MD simulations, in conjunction with Markov state modeling, allow the dynamic conformational landscape of RNA to be probed and thus different and more relevant binding modes and capabilities of a potential drug to be identified; by contrast, docking to rigid RNA structures is not sufficient to guide such drug designs. The simulations provide crucial new information, not readily accessible by experiment: they show insertion of the cylinders into the cavity of the RNA bulge in a similar binding to that seen for RNA 3-way junctions; they not only provide insight into the ultimate bound structure but also its wider effect on RNA conformation reducing the RNA conformational flexibility once the cavity is bound; and, for the first time, they provide insight about the molecular mechanism through which a drug might enter a cavity in the RNA UTR, involving stacking on and sliding down bases and base pairs. Together these new molecular insights and the combined modelling and analysis approaches that have enabled them and can be more widely applied, will transform understanding of how to create supramolecular drugs that insert effectively into RNA cavities and can guide new designs against a spectrum of critical RNA viruses that threaten human well-being.

Author contributions

DFT calculations, MD simulations and MSM analyses were designed and undertaken by LM. MJH conceived the project, which was supervised by MJH and IS. All authors analysed the data and discussed the results, and MJH and LM drafted the paper. All authors commented on the manuscript and contributed to the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Drs Dwaipayan Chakrabarti, Pawel Grzechnik, Robert Neely and Prof Zoe Pikramenou (all U. Birmingham) for helpful advice. This work was funded by the EPSRC Physical Sciences for Health Centre (EP/L016346/1) and the University of Birmingham. Simulations used the BlueBEAR HPC and CaStLes resources⁹⁴ (U. Birmingham).

Notes and references

- 1 S. P. Velagapudi, M. D. Cameron, C. L. Haga, L. H. Rosenberg, M. Lafitte, D. R. Duckett, D. G. Phinney and M. D. Disney, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **113**, 5898–5903.
- 2 N. F. Rizvi and G. F. Smith, *Bioorg. Med. Chem. Lett.*, 2017, **27**, 5083–5088.
- 3 M. D. Disney and A. J. Angelbello, *Acc. Chem. Res.*, 2016, **49**, 2698–2704.
- 4 M. D. Disney, B. G. Dwyer and J. L. Childs-Disney, *Cold Spring Harbor Perspect. Biol.*, 2018, **10**, a034769.
- 5 G. J. R. Zaman, P. J. A. Michiels and C. A. A. Van Boeckel, *Drug Discov. Today*, 2003, **8**, 297–306.
- 6 C. H. Li and Y. Chen, *Int. J. Biochem. Cell Biol.*, 2013, **45**, 1895–1910.
- 7 V. Smirnova, I. M. Terenin, A. Khutornenko, D. E. Andreev, S. E. Dmitriev and I. N. Shatsky, *Biochimie*, 2016, **121**, 228–237.
- 8 M. De Nova-Ocampo, M. C. Soliman and W. Espinosa-Hernández, *Mol. Biol. Rep.*, 2019, **46**, 1413–1424.
- 9 J. Gilmore, K. Deguchi and K. Takeyasu, Nanoimaging of RNA Molecules with Atomic Force Microscopy. in *Microscopy and imaging Science: Practical approaches to Applied Research and education*, ed. A. Méndez-Vilas, Formatex Research Center, 2017, pp. 300–306.
- 10 R. Comandur, E. D. Olson and K. Musier-Forsyth, *RNA*, 2017, **23**, 1850–1859.
- 11 S. M. Villordo, C. V. Filomatori, I. Sánchez-Vargas, C. D. Blair and A. V. Gamarnik, *PLoS Pathog.*, 2015, **11**, 1–22.
- 12 D. E. Alvarez, A. L. S. De Lella Ezcurra Fucito and A. V. Gamarnik, *Virology*, 2005, **339**, 200–212.
- 13 E. J. Kelly, E. M. Hadac, S. Greiner and S. J. Russell, *Nat. Med.*, 2008, **14**, 1278–1283.
- 14 C. K. Damgaard, E. S. Andersen, B. Knudsen, J. Gorodkin and J. Kjems, *J. Mol. Biol.*, 2004, **336**, 369–379.
- 15 I. Boeras, B. Seufzer, S. Brady, A. Rendahl, X. Heng and K. Boris-Lawrie, *Sci. Rep.*, 2017, **7**, 6902.
- 16 D. Yang and J. L. Leibowitz, *Virus Res.*, 2015, **206**, 120–133.
- 17 B. Hsue and P. S. Masters, *J. Virol.*, 1997, **71**, 7567–7578.
- 18 L. Li, H. Kang, P. Liu, N. Makkinje, S. T. Williamson, J. L. Leibowitz and D. P. Giedroc, *J. Mol. Biol.*, 2008, **377**, 790–803.
- 19 K. C. Liao, V. Chuo, W. C. Ng, S. P. Neo, J. Pompon, J. Gunaratne, E. E. Ooi and M. A. Garcia-Blanco, *RNA*, 2018, **24**, 803–814.
- 20 L. de Borja, S. M. Villordo, F. L. Marsico, J. M. Carballeda, C. V. Filomatori, L. G. Gebhard, H. M. Pallarés,



- S. Lequime, L. Lambrechts, I. Sánchez Vargas, C. D. Blair and A. V. Gamarnik, *mBio*, 2019, **10**, e02506–e02518.
- 21 A. M. Fleming, Y. Ding, A. Alenko and C. J. Burrows, *ACS Infect. Dis.*, 2016, **2**, 674–681.
- 22 W. C. Ng, R. Soto-Acosta, S. S. Bradrick, M. A. Garcia-Blanco and E. E. Ooi, *Viruses*, 2017, **9**, 1–14.
- 23 L. Sethaphong, A. Singh, A. E. Marlowe and Y. G. Yingling, *J. Phys. Chem. C*, 2010, **114**, 5506–5512.
- 24 T. Kulinski, M. Olejniczak, H. Huthoff, L. Bielecki, K. Pachulska-Wieczorek, A. T. Das, B. Berkhout and R. W. Adamiak, *J. Biol. Chem.*, 2003, **278**, 38892–38901.
- 25 T. N. Do, E. Ippoliti, P. Carloni, G. Varani and M. Parrinello, *J. Chem. Theory Comput.*, 2012, **8**, 688–694.
- 26 L. Pascale, S. Azoulay, A. Di Giorgio, L. Zenacker, M. Gaysinski, P. Clayette and N. Patino, *Nucleic Acids Res.*, 2013, **41**, 5851–5863.
- 27 D. Maity, S. Kumar, F. Curreli, A. K. Debnath and A. D. Hamilton, *Chem.–Eur. J.*, 2019, **25**, 7265–7269.
- 28 M. J. Selby, E. S. Bain, P. A. Luciw and B. M. Peterlin, *Genes Dev.*, 1989, **3**, 547–558.
- 29 R. Nifosi, *Nucleic Acids Res.*, 2000, **28**, 4944–4955.
- 30 F. Musiani, G. Rossetti, L. Capece, T. M. Gerger, C. Micheletti, G. Varani and P. Carloni, *J. Am. Chem. Soc.*, 2014, **136**, 15631–15637.
- 31 D. M. Krüger, J. Bergs, S. Kazemi and H. Gohlke, *ACS Med. Chem. Lett.*, 2011, **2**, 489–493.
- 32 E. Ennifar, J. C. Paillart, A. Bodlenner, P. Walter, J. M. Weibel, A. M. Aubertin, P. Pale, P. Dumas and R. Marquet, *Nucleic Acids Res.*, 2006, **34**, 2328–2339.
- 33 H. Dong, D. Ray, S. Ren, B. Zhang, F. Puig-Basagoiti, Y. Takagi, C. K. Ho, H. Li and P. Y. Shi, *J. Virol.*, 2007, **1**, 4412–4421.
- 34 H. Ling, M. Fabbri and G. A. Calin, *Nat. Rev. Drug Discovery*, 2013, **12**, 847–865.
- 35 L. Cardo, I. Nawroth, P. J. Cail, J. A. McKeating and M. J. Hannon, *Sci. Rep.*, 2018, **8**, 13342.
- 36 H. Sepehrpour, W. Fu, Y. Sun and P. J. Stang, *J. Am. Chem. Soc.*, 2019, **141**, 14005–14020.
- 37 A. Casini, B. Woods and M. Wenzel, *Inorg. Chem.*, 2017, **56**, 14715–14729.
- 38 A. Pöthig and A. Casini, *Theranostics*, 2019, **9**, 3150–3169.
- 39 B. Woods, R. D. M. Silva, C. Schmidt, D. Wragg, M. Cavaco, V. Neves, V. F. C. Ferreira, L. Gano, T. S. Morais, F. Mendes, J. D. G. Correia and A. Casini, *Bioconjugate Chem.*, 2021, DOI: 10.1021/acs.bioconjchem.0c00659.
- 40 J. Han, A. F. B. Räder, F. Reichart, B. Aikman, M. N. Wenzel, B. Woods, M. Weinmüller, B. S. Ludwig, S. Stürup, G. M. M. Groothuis, H. P. Permentier, R. Bischoff, H. Kessler, P. Horvatovich and A. Casini, *Bioconjugate Chem.*, 2018, **29**, 3856–3865.
- 41 L. Cardo and M. J. Hannon, *Met. Ions Life Sci.*, 2018, **18**, 303–324.
- 42 A. Oleksi, A. G. Blanco, R. Boer, I. Usón, J. Aymamí, A. Rodger, M. J. Hannon and M. Coll, *Angew. Chem. Int. Ed.*, 2006, **45**, 1227–1231.
- 43 S. Phongtongpasuk, S. Paulus, J. Schnabl, R. K. Sigel, B. Spingler, M. J. Hannon and E. Freisinger, *Angew. Chem. Int. Ed.*, 2013, **52**, 11513–11516.
- 44 J. Malina, M. J. Hannon and V. Brabec, *Sci. Rep.*, 2016, **6**, 29674.
- 45 J. M. Kerckhoffs, J. C. Peberdy, I. Meistermann, L. J. Childs, C. J. Isaac, C. R. Pearmund, V. Reudegger, S. Khalid, N. W. Alcock, M. J. Hannon and A. Rodger, *Dalton Trans.*, 2007, **21**, 734–742.
- 46 G. I. Pascu, A. C. G. Hotze, C. Sanchez-Cano, B. M. Kariuki and M. J. Hannon, *Angew. Chem. Int. Ed.*, 2007, **46**, 4374–4378.
- 47 J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter and M. Otyepka, *Chem. Rev.*, 2018, **118**, 4177–4338.
- 48 A. Cesari, S. Bottaro, K. Lindorff-Larsen, P. Banáš, J. Šponer and G. Bussi, *J. Chem. Theory Comput.*, 2019, **15**, 3425–3431.
- 49 D. Tan, S. Piana, R. M. Dirks and D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.*, 2018, **115**, 1346–1355.
- 50 S. Vangaveti, S. V. Ranganathan and A. A. Chen, *Wiley Interdiscip. Rev.: RNA*, 2017, **8**, e1396.
- 51 A. H. Aytenfisu, A. Spasic, A. Grossfield, H. A. Stern and D. H. Mathews, *J. Chem. Theory Comput.*, 2017, **13**, 900–915.
- 52 A. J. Angelbello, R. I. Benhamou, S. G. Rzuczek, S. Choudhary, Z. Tang, J. L. Chen, M. Roy, K. W. Wang, I. Yildirim, A. S. Jun, C. A. Thornton and M. D. Disney, *Cell Chem. Biol.*, 2021, **28**, 34–45.
- 53 D. H. Mathews, *Methods*, 2019, **162–163**, 60–67.
- 54 P. D. Dans, D. Gallego, A. Balaceanu, L. Darré, H. Gómez and M. Orozco, *Chem*, 2019, **5**, 51–73.
- 55 N. Gresh, J. E. Sponer, M. Devereux, K. Gkionis, B. De Courcy, J. P. Piquemal and J. Sponer, *J. Phys. Chem. B*, 2015, **119**, 9477–9495.
- 56 M. Zgarbová, M. Otyepka, J. Šponer, F. Lankaš and P. Jurečka, *J. Chem. Theory Comput.*, 2014, **10**, 3177–3189.
- 57 G. Pinamonti, F. Paul, F. Noe, A. Rodriguez and G. Bussi, *J. Chem. Phys.*, 2019, **150**, 154123.
- 58 H. Shi, A. Rangadurai, H. Abou Assi, R. Roy, D. A. Case, D. Herschlag, J. D. Yesselman and H. M. Al-Hashimi, *Nat. Commun.*, 2020, **11**, 5531.
- 59 J. Copperman and D. Zuckerman, Accelerated estimation of long-timescale kinetics by combining weighted ensemble simulation with Markov model microstates using non-Markovian theory, 2019, ArXiv.1903.04673, DOI: 10.1016/j.bpj.2019.11.1099.
- 60 M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J. H. Prinz and F. Noé, *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.
- 61 N. B. Leontis, *Nucleic Acids Res.*, 2002, **30**, 3497–3531.
- 62 S. Bottaro, G. Bussi, G. Pinamonti, S. Reißer, W. Boomsma and K. Lindorff-Larsen, *RNA*, 2019, **25**, 219–231.
- 63 J. A. N. Zoll, M. Tessari, F. J. M. Van Kuppeveld, W. J. G. Melchers and H. A. Heus, *RNA*, 2007, **13**, 781–792.



- 64 O. Ohlenschläger, J. Wöhnert, E. Bucci, S. Seitz, S. Häfner, R. Ramachandran, R. Zell and M. Görlach, *Structure*, 2004, **12**, 237–248.
- 65 K. T. Dayie, A. S. Brodsky and J. R. Williamson, *J. Mol. Biol.*, 2002, **317**, 263–278.
- 66 F. Aboul-ela, J. Karn and G. Varani, *Nucleic Acids Res.*, 1996, **24**, 3974–3981.
- 67 N. I. Orlovsky, H. M. Al-Hashimi and T. G. Oas, *J. Am. Chem. Soc.*, 2020, **142**, 907–921.
- 68 B. Davis, M. Afshar, G. Varani, A. I. H. Murchie, J. Karn, G. Lentzen, M. J. Drysdale, J. Bower, A. J. Potter and F. Aboul-Ela, *J. Mol. Biol.*, 2004, **336**, 343.
- 69 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2019, **31**, 455–461.
- 70 M. A. Ditzler, M. Otyepka, J. Šponer and N. G. Walter, *Acc. Chem. Res.*, 2010, **43**, 40–47.
- 71 G. A. Bermejo, G. M. Clore and C. D. Schwieters, *Structure*, 2016, **24**, 806–815.
- 72 D. P. Buck, C. B. Spillane, J. G. Collins and F. R. Keene, *Mol. Biosyst.*, 2008, **4**, 851–854.
- 73 C. B. Spillane, J. A. Smith, D. P. Buck, J. G. Collins and F. R. Keene, *Dalton Trans.*, 2007, **7**, 5290–5296.
- 74 J. Malina, P. Scott and V. Brabec, *Chem.–Eur. J.*, 2020, **26**, 8435–8442.
- 75 A. J. McConnell, H. Song and J. K. Barton, *Inorg. Chem.*, 2013, **52**, 10131–10136.
- 76 L. S. Ratmeyer, R. Vinayak, G. Zon and W. D. Wilson, *J. Med. Chem.*, 1992, **35**, 966–968.
- 77 E. Alberti, M. Zampakou and D. Donghi, *J. Inorg. Biochem.*, 2016, **163**, 278–291.
- 78 H. R. Neenhold and T. M. Rana, *Biochemistry*, 1995, **34**, 6303–6309.
- 79 A. C. Lim and J. K. Barton, *Bioorg. Med. Chem.*, 1997, **5**, 1131–1136.
- 80 P. J. Carter, C. C. Cheng and H. H. Thorp, *J. Am. Chem. Soc.*, 1998, **120**, 632–642.
- 81 R. C. Bernardi, M. C. R. Melo and K. Schulten, *Biochim. Biophys. Acta, Gen. Subj.*, 2015, **1850**, 872–877.
- 82 D. Wragg, A. de Almeida, R. Bonsignore, F. E. Kühn, S. Leoni and A. Casini, *Angew. Chem. Int. Ed.*, 2018, **57**, 14524–14528.
- 83 E. Aprá, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning Jr, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero-de-la-Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, A. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao and R. J. Harrison, *J. Chem. Phys.*, 2020, **152**, 184102.
- 84 M. Swart, M. Solà and F. M. Bickelhaupt, *J. Chem. Phys.*, 2009, **131**, 094103.
- 85 H. P. Varbanov, M. A. Jakupec, A. Roller, F. Jensen, M. Galanski and B. K. Keppler, *J. Med. Chem.*, 2013, **56**, 330–344.
- 86 K. P. Kepp, *Inorg. Chem.*, 2016, **55**, 2717–2727.
- 87 S. Grimme, J. Antony, T. Schwabe and C. Mück-Lichtenfeld, *Org. Biomol. Chem.*, 2007, **5**, 741–758.
- 88 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.
- 89 P. Li and K. M. Merz, *J. Chem. Inf. Model.*, 2016, **56**, 599–604.
- 90 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 09*, Gaussian, Inc., Wallingford CT, 2016.
- 91 M. J. Abraham, T. Murtola, R. Schulz, S. Pall, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
- 92 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **3**, 198–210.
- 93 R. Kumari, R. Kumar and A. Lynn, *J. Chem. Inf. Model.*, 2014, **54**, 1951–1962.
- 94 S. J. Thompson, S. E. M. Thompson and J.-B. Cazier, *CaStLeS (Compute and Storage for the Life Sciences): a collection of compute and storage resources for supporting research at the University of Birmingham*, DOI: 10.5281/zenodo.3250616.

