Computer-aided comprehensive explorations of RNA structural polymorphism through complementary simulation methods

Konstantin Röder[1], Guillaume Stinermann[2], Pietro Faccioli[3], and Samuela Pasquali[4,5]*

[1] Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK

[2] CNRS Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, PSL University, Université de Paris, 13 rue Pierre et Marie Curie, 75005, Paris, France

[3] Department of Physics, University of Trento and INFN-TIFPA, Trento, Italy

[4] Laboratoire CiTCoM, CNRS UMR 8038, Université Paris Cité, 4 avenue de l'Observatoire, 75006, Paris, France

[5] Laboratoire BFA, CNRS UMR 8251, Université Paris Cité, 35 rue Hélène Brion, 75013, Paris, France

*Corresponding author. Email: samuela.pasquali@u-paris.fr

Abstract

While RNA folding was originally seen as a simple problem to solve, it has been shown that the promiscuous interactions of the nucleobases result in structural polymorphism, with several competing structures generally observed for non-coding RNA. This inherent complexity limits our understanding of these molecules from experiments alone, and computational methods are commonly used to study RNA. Here, we discuss three advanced sampling schemes, namely Hamiltonian-replica exchange molecular dynamics, ratchet-and-pawl molecular dynamics and discrete pathsampling, as well as the HiRE-RNA coarse-graining scheme, and highlight how these approaches are complementary with reference to recent case studies. While all computational methods have their shortcomings, the plurality of simulation methods

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI.

10.1017/qrd.2022.19

leads to a better understanding of experimental findings and can inform and guide experimental work on RNA polymorphism.

The complexity of RNA folding

After the seminal experiments showing the hierarchical folding of RNA, RNA folding was thought to be an easier problem to solve than protein folding (Li, Vieregg, et al. 2008; Tinoco and Bustamante 1999). With an alphabet composed of only four letters, and with key interactions leading to the observed secondary structure dictated by canonical base pairing (G with C and A with T/U), what remained to be solved was "only" a combinatorial problem of finding the best pairing scheme for a given sequence.

About two decades later, we know that the problem is much more complex. Even searching for the optimal secondary structure remains a challenge as exhaustive sampling of all relevant conformations is unfeasible for most systems of biological interest, even though the advent of machine learning and the extensive use of chemical probing data are contributing to making the problem more tractable (Lorenz et al. 2016; Zhao et al. 2021). A common feature in complex RNA architectures are pseudoknots – non-nested arrangements of base pairs. Traditional secondary structure prediction algorithms do not treat these structures well and combining these approaches with machine learning has led to some progress (Sato and Kato 2022; Wang, Liu, et al. 2019). The situation is even more complex considering that canonical base pairing, even though dominant, is not the only form of base pairing. The multiple hydrogen bond donor and acceptor sites of the nucleobases allow for a multitude of base pairs, which have been reported experimentally. Around 150 non-canonical base pairs have been found and classified in terms of interaction "edges" (Watson-Crick, Hoogsteen and Sugar) (Leontis and Westhof 2001; Stombaugh et al. 2009). The full list can be found in the RNA Basepair Catalog of the Nucleic Acids Databank.

As it is the case in general for heteropolymers, a smaller alphabet results in an increase of frustration of the conformational space accessible to the molecule. In the case of RNA, the alphabet composed of only four different nucleobases, further complicated by the multitude of possible base pairs, results in a folding process possibly more complex to predict than for proteins (Ferreiro et al. 2014). The observation that proteins fold reliably and fast into their native confirmation has been explained by the principle of minimal frustration (Bryngelson and Wolynes 1987). Every sequence defines interactions between different parts of the molecule. The more of these are formed, the lower the frustration and the more stable the resulting structure. The native state exhibits a conformation that fulfils all packing requirements, i.e. the system shows minimal frustration. Minimal frustration is linked to the topography of the energy landscape, and in the case of globular proteins a single funnel anchored around the native fold is observed (Bryngelson, Onuchic,

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

et al. 1995; Leopold et al. 1992). [1] As a result, the number of native contacts observed is a good proxy for the progress of the highly cooperative folding of proteins.

In contrast, RNA is characterized by the existence of several stable structural ensembles with different secondary structures, and many of these systems are highly dynamic (Brillet et al. 2020). The number of alternative contacts in RNA lead to large frustration and disorder, as the sequence allows for multiple competing interactions. This higher frustration has been highlighted both by experiments (Burge et al. 2006; Garst et al. 2011; Kolesnikova and Curtis 2019; Lightfoot et al. 2019; Martinez-Zapien et al. 2017; Saldi et al. 2021; Yu et al. 2021) and by simulations (Cragnolini, Laurin, et al. 2015; Denesyuk and Thirumalai 2011; Rissone et al. 2022; Röder, Barker, et al. 2022; Röder, Stirnemann, et al. 2020; Schlick et al. 2021; Šponer et al. 2018; Yan et al. 2022), and its main manifestation is structural polymorphism. Within these distinct structures, there must not necessarily be a distinct global minimum, and therefore a native state does not necessarily exist, as has been noticed by others (Vicens and Kieft 2022).

Therefore, in our opinion, an ensemble approach should be chosen when talking about RNA. The relative population of these structural ensembles depends on experimental conditions, as observed for riboswithches and several other non-coding regulatory RNAs (Brillet et al. 2020; Fay et al. 2017; Halvorsen et al. 2010; Kolesnikova and Curtis 2019). Post-transcriptional modifications and single point mutations also can shift the equilibrium between the alternative structures (Liu et al. 2017; Martinez-Zapien et al. 2017; Röder, Barker, et al. 2022; Schlick et al. 2021). Finally, many RNAs interact with proteins and these interactions often lead to changes to the observed fold (Jaeger et al. 2009). Which structure is detected in experiments therefore depends on the details of the experiment itself, and at times more than one structure is detected in the same experiment (Martinez-Zapien et al. 2017).

Given this plurality of possible structures, simulations cannot be limited to the prediction of a single structure (which is what is achieved by most bioinformatic approaches), and focus must shift to a global view, which centres around the molecular energy landscape. All information about the structures, their energy, and interconversion pathways between them can be calculated from knowledge of the energy landscape (EL). Insight can also be obtained on the influence of external factors such as ionic conditions, pH, temperature, presence of ligands, and chemical changes in the sequence by considering the EL. Any experiment or simulation probes the energy landscape directly or indirectly. Various methods do so in different ways, and often the EL is not directly mapped.

The most common simulation method is molecular dynamics (MD) simulations. However, due to the broken ergodicity exhibited biomolecular energy landscapes (Wales and Salamon 2014) there are many practical difficulties. In brief, the structural ensembles are separated by high barriers, making transitions between them rare events. This kinetic

[1] This description extends to proteins that exhibit more than one structural ensemble, and which have a multifunnel energy landscape. Such landscapes are also governed by the principle of minimal frustration (Röder and Wales 2018).

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

partition between different regions will make observation of transitions in standard MD simulations very unlikely. As a result, so called enhanced sampling approaches have been developed, which for example include pathsampling methods.

Here, we present our perspective on how simulations can be used to gather information on RNA energy landscapes and structural polymorphism. There are two approaches commonly employed. The first option is the use of enhanced sampling methods (Mlýnský and Bussi 2018), and here we briefly present three of these, namely Hamiltonian Replica Exchange (H-REX, REST2) simulations (Wang, Friesner, et al. 2011), discrete pathsampling (DPS) (Wales 2002; Wales 2004) and a variationally optimized ratchet-and-pawl molecular dynamics (rMD) simulation scheme (Tiana and Camilloni 2012) called Bias Functional approach (A Beccara et al. 2015). The second option is to smooth the energy landscape through coarse-graining (Papoian 2018). A pictorial illustration on how each of these methods samples the energy landscape is given in Fig.1. By considering several examples, we show that these approaches are complementary, and that the best results are obtained when combining multiple simulation methods.

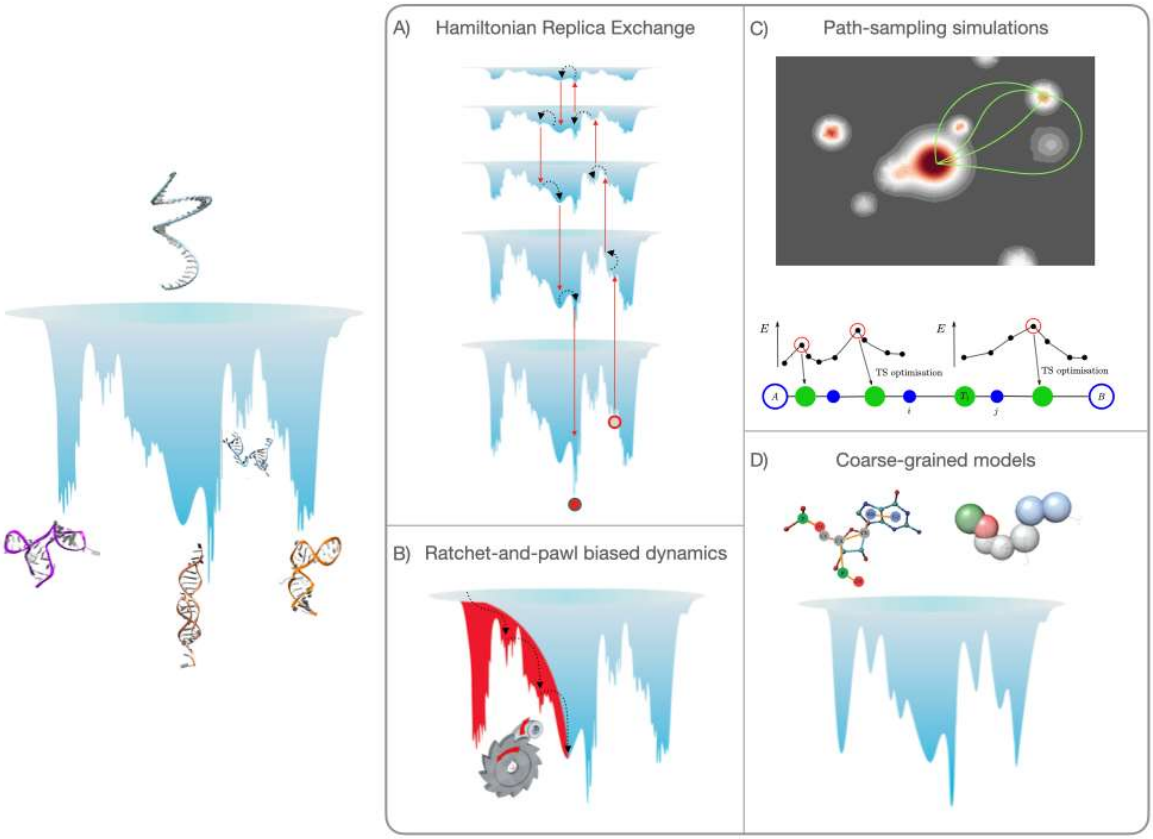


Figure 1: Left: Illustration of how the energy landscape of a polymorphic RNA might look like (the vertical axis represents the energy or free energy of the system). At the top of the landscape, we find high energy unfolded conformations, while we note several deep minima, separated by high barriers, all corresponding to substantially different structures of the molecule. One

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

of these minima might be observed experimentally and referred to as the "native structure". Right panel (A-D): illustration of how each of the method presented samples the landscape.

An overview of the simulation methods

Hamiltonian-Replica Exchange simulations

Despite the increased time scales that can be probed with unperturbed MD simulations – now routinely on the order of μs – the relevant conformational motions cannot be sampled as the associated time scales still exceed computational feasibility, prompting interest for enhanced sampling strategies that have been developed and widely applied to biomolecules, including RNA (see e.g. Mlýnský and Bussi 2018 for a recent review).

One way to improve sampling in an unguided way (i.e. without assuming or imposing predetermined collective variables along which the transitions will occur) is through the use of replica exchange MD simulations (Sugita and Okamoto 1999). Multiple copies of the system are simulated at different temperatures, increasing the accessible time scales. However, this approach is very sensitive to the overlap between the replicas, which depends on the number of degrees of freedom, and at the moment it is hardly applicable at an all-atom resolution for nucleic acids exceeding a handful of residues in explicit solvent.

This problem may be overcome by using Hamiltonian replica-exchange (H-REX) simulation schemes. In particular, we used the Replica-Exchange with Solute Tempering (REST2) strategy (Wang, Friesner, et al. 2011), where all replicas evolve at the same physical temperature, but they can exchange their Hamiltonian with a scaled potential energy for the biomolecule (Fig.1A), decreasing the number of degrees of freedom. As a result, fewer replicas are required, and sampling is enhanced. For example, for proteins containing 100-200 residues, one to two dozen replicas were shown to lead to satisfactory exchange probabilities (Maffucci, Laage, Sterpone, et al. 2020; Maffucci, Laage, Stirnemann, et al. 2020; Stirnemann and Sterpone 2017). However, this technique has mostly been applied to short oligonucleotides, and in particular to the sampling of tetraloops conformational space (Bottaro et al. 2020; Kührová et al. 2016; Mlýnský, Janecek, et al. 2022). While a recent work pointed to limitations in the ability of such an approach to actually fold even short RNAs (Mlýnský, Janeček, et al. 2022), REST2 remains a very attractive strategy to ease and to accelerate conformational sampling, which eventually enables to escape the kinetic traps in which brute-force simulations may be stuck for long times.

In this short perspective, we exclusively focus on REST2, which we have applied to an RNA much larger than these tetraloops (Röder, Stirnemann, et al. 2020), but other applications to reasonably large biomolecules are mostly limited to DNAs and proteins. For these applications, recent success of REST2 in identifying important conformations that were not revealed by long brute-force MD (Gillet et al. 2021; Maffucci, Laage, Sterpone, et al. 2020; Maffucci, Laage, Stirnemann, et al. 2020; Stirnemann and Sterpone

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

2017) offer promising perspectives for the RNA field. However, it should be noted that when employed with atomistic resolution models, the computational costs remain high. This shortcoming may be overcome by focusing on a specific region of the system under investigation, reducing the size of the perturbed region, and thus the number of required replicas.

Ratchet-and-pawl molecular dynamics (rMD) and the Bias Functional approach

Ratchet-and-pawl (rMD) simulations are based on introducing a soft history-dependent biasing force to enhance the generation of productive folding trajectories towards a given target structure (Paci and Karplus 1999). In practice, once a target structure is known experimentally, it is possible to extract some features characteristics of its configuration and define a collective variable that can be used to guide unfolded structures toward it in a biased molecular dynamic simulation. In the literature, many collective variables exist for biomolecules, ranging from a simple atomic distance or dihedral angle to the radius of gyration, the RMSD and many more depending on the specific feature relevant for the fold of the molecule (Fiorini et al. 2013). The system is free to explore the energy landscape, as long as it follows broadly this predetermined collective variable (CV), which is a proxy for the reaction coordinate. An external biasing force is switched on when the system backtracks with respect to the CV (see Fig.1B). In RNA and protein folding simulations, one choice for the predetermined CV is obtained from the overlap of the instantaneous and the target atomistic contact map (Camilloni et al. 2011). This approach produces folding trajectories efficiently but requires structural information about the target.

In the ideal case in which CV coincides with the reaction coordinate (the committor function (E and Vanden-Eijnden 2010)), rMD trajectories sample the correct region of configuration space (Bartolucci et al. 2018; Cameron and Vanden-Eijnden 2014). However, the choice of CV used in RNA folding simulations is only a proxy of the ideal reaction coordinate. Therefore, with rMD it is only possible to obtain an approximate reconstruction of the folding energy landscape. Systematic errors from the biasing force can be minimised by applying Bias Functional (BF) filtering procedure (A Beccara et al. 2015). In this approach, a variational principle derived from the path integral representation of Langevin dynamics (Onsager and Machlup 1951) is used to select the folding trajectories generated by rMD that have the highest probability of occurring in the absence of any biasing force.

Apart from the requirement to use structural information about the folded structure, another drawback of rMD simulations is that the generated trajectories only explore part of the energy landscape, namely the region most likely traversed by productive pathways towards the predetermined target structure. While this approach greatly enhances computational efficiency, it prevents the method from exploring other parts of the landscape that may be associated with kinetic trapping.

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

Discrete pathsampling for RNA

H-REX and rMD simulations compute trajectories of molecules moving on the energy landscape. Discrete pathsampling (DPS) (Wales 2002; Wales 2004) focuses on the topography of the energy landscape. The energy landscape is considered coarse-grained, where only the local minima and transition states that connect them are used as representation. Each transition state connects two local minima, and between any pair of minima, we can identify a discrete path consisting of a series of minima connected by transition states. This representation results in a kinetic transition network, which can then be analysed to obtain kinetic and thermodynamic characteristic, including the associated structures and transition mechanisms.

Through this approach, the topography of the energy landscape is obtained, and this information allows readily for interpretation of mutational data (Röder, Stirnemann, et al. 2020). As local minima and transition states are well-defined geometrically, they can be located by geometry optimisation, overcoming the dependence on long time scales other simulations suffer from. A shortcoming of the method is the use of implicit solvent representations, which introduces a source of error (Šponer et al. 2018). While it is theoretically possible to use explicit solvent, the increased computational cost currently prevents such setups. While free energies can be readily obtained, explorations of higher entropy configurations are difficult. As such, structural transitions between folded structures are generally well resolved, while unfolding events are not. More information and details on how the energy landscapes are explored with DPS can be found in various reviews (Joseph et al. 2017; Röder, Joseph, et al. 2019).

While DPS is most efficient when folded structures are known, the methodology can locate unknown folded structures and new funnels, as demonstrated in the exploration of mutational changes for example in 7SK RNA (Röder, Stirnemann, et al. 2020). However, currently there is no algorithm to guarantee the location of all structures. A useful way around this limitation is to create several possible alternative structures and connect them. Importantly, this approach does not require the structures to be optimised as long as key interactions, such as base pairs are formed.

Coarse-grained RNA representations

By grouping several atoms into larger particles (grains), the computational exploration of the energy landscape is aided in two ways. Firstly, the coarse-graining smooths the energy landscape (see Fig.1D), which removes kinetic traps for the exploration. Secondly, the number of degrees of freedom is reduced, making the computations more tractable. The choice of the mapping between atoms and grains depends on the level of details required and on the kind of interactions that are considered relevant (see (Li and Chen 2021) for a recent review on the different existing RNA coarse-grained models). For RNA structures, key elements are base pairing, stacking and electrostatic interactions. In the HiRE-RNA model (High-Resolution Energy model for RNA) (Cragnolini, Laurin, et al. 2015; Pasquali and Derreumaux 2010), we have chosen to preserve a relatively high resolution with each

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

nucleotide described by 6 or 7 beads. This level of detail, while significantly reducing the number of particles, allows the definition of planes for the nucleobases, reflecting the aromatic rings stacking, and distinguishes different edges of the bases to account for both canonical and non-canonical pairings. While using an implicit solvent, long-range electrostatic effects are accounted for by a Debye-Hückel potential energy term dependent on experimental ionic concentrations in solution. While the development of this coarse-grained model is still on-going, its usefulness for small systems (Cragnolini, Chakraborty, et al. 2017; Stadlbauer et al. 2016) and when coupled to experimental data (Mazzanti et al. 2021; Pasquali, Frezza, et al. 2019) has been demonstrated.

The obvious shortcoming of any coarse-graining (CG) methodology is the loss of detail, due to the lower model resolution. In addition, the implicit nature of solvent and ions will impact the observed features. These drawbacks mean the entropy is not faithfully produced within coarse grained simulations. However, the reduced complexity will allow the study of larger systems and larger scale rearrangements, providing otherwise inaccessible insights.

Despite the fewer degrees of freedom, our coarse-grained MD simulations can still be expensive, with several days of CPU needed to achieve folding of a small molecule of 20-30 nucleotides, although we will be able to achieve much greater speed once the force-field will be ported to parallel MD computing.

A small showcase

In this section, we discuss a few illustrative applications of these methods, emphasising their complementary nature.

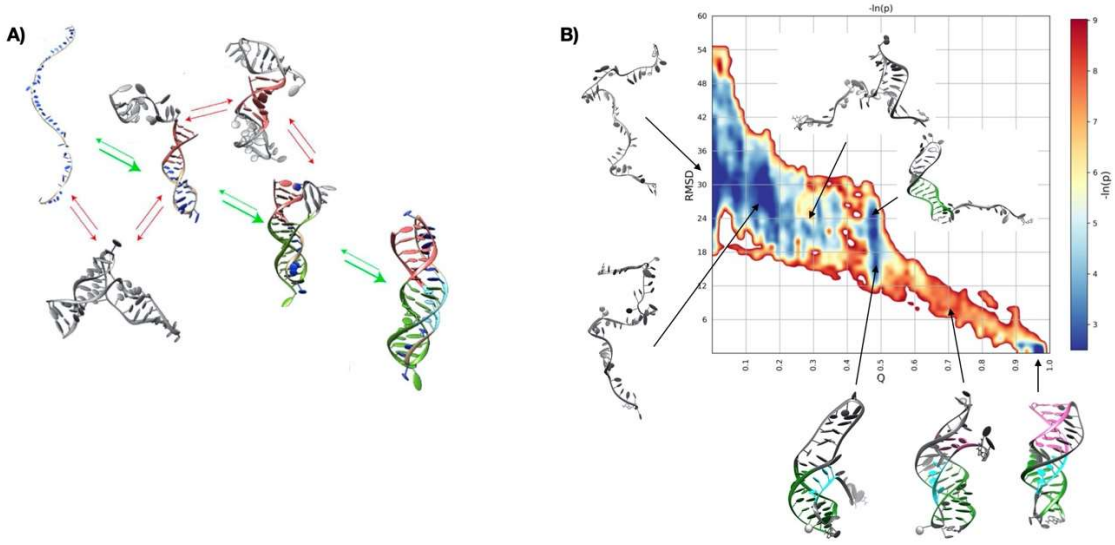


Figure 2: Folding of the human telomerase triple helix performed with unbiased coarse-grained simulations (REMD), allowing to widely explore alternative conformations (A) and with biased atomistic simulations allowing to explore the details of intermediate states (B).

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

Folding pathway of the human telomerase H-pseudoknot triple helix

This example is a 47-nucleotide RNA, exhibiting a H-pseudoknot (two-interlacing strands) further stabilized by a triple helix (PDB ID 2K96). The system has been studied extensively experimentally (Gavory et al. 2006; Kim et al. 2008; Theimer et al. 2005) and has become a benchmark for modelling (Biyun et al. 2011; Cho et al. 2009; Denesyuk and Thirumalai 2011), as it contains a pseudoknot, a challenging structural feature, and non-canonical interactions leading to triplet formation in the triple helix. This system was also used as test case for the HiRE-RNA model (Cragnolini, Laurin, et al. 2015), and, more recently, to validate the application of variationally optimized rMD to RNAs (Lazzeri et al. 2022). Folding simulations were performed in both instances starting from fully unfolded conformations.

The coarse-grained simulations consisted of a long run with the HiRE-RNA model and replica exchange MD simulations at 64 different temperatures. rMD folding simulations consisted of 100 short runs (each lasting a nominal time interval of 5 ns) with the AMBER ff99 with the Barcelona α/γ backbone modification (Perez et al. 2007) and the χ modification (Zgarbova et al. 2011). It should be emphasized that the simulation time does not directly correlate with the physical transition path time, as the history-dependent bias brakes microscopic reversibility and alters the kinetics. CG simulations required two weeks of computation on local cluster in 2015 to achieve the native structure for the first time. rMD simulations required roughly a week of simulation to generate all trajectories on a GPU cluster in 2022. The results of the two simulations are shown in Fig.2.

The HiRE-RNA simulations yielded the correct native state and identified a sensible folding pathway. Moreover, alternative states were observed, which constitute kinetic traps and are characterized by the formation of non-native secondary structures, leading to alternative folds. These results were in qualitative agreement with experimental evidence of the formation of metastable states and folding intermediates (Kim et al. 2008). Despite the coarse-graining, these simulations were computationally expensive and the sampling was not optimal. In particular, it was not possible to give a full assessment of the relative populations of the observed states. The rMD simulations produced more insight into the productive folding pathway to the experimentally observed target structure. It was possible to collect statistically relevant populations for the different conformations and generate a heatmap illustrating the folding in terms of formation of native contacts and RMSD with respect to target (see Fig.2B). The results highlighted the ruggedness of the folding landscape, characterized by a multitude of intermediate states. Moreover, we were able to infer the existence a pronounced bottleneck towards the final stage of folding, when the formation of the pseudoknot takes place. It was also possible to characterize the order of the events of folding in terms of formation of stems and loops and the main path found by our simulations corresponded well with the experimental evidence from thermodynamic studies (Kim et al. 2008) and by simulations by other groups (Cho et al. 2009).

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

Importantly, both methods lead to the correct folding path and folding intermediates, giving credibility to both methods. In addition, the unbiased CG simulations also identified alternative structures, which are not on the folding pathway. The rMD simulations can provide statistics of the explored states and detailed insight into the interactions along the folding pathway.

Energy landscape and folding pathway of a small H-pseudoknot

For the 22-nucleotide long tmRNA pseudoknot taken from Aquifex aeolicus (PDB ID 2G1W) (Nonin-Lecomte et al. 2006), we performed both a full exploration of the energy landscape with discrete pathsampling and an exploration of the folding pathway with rMD (see Fig.3). Both sets of simulations used the atomistic AMBER ff99 force field with the Barcelona α/γ backbone modification and the χ modification (Zgarbova et al. 2011). The energy landscape exploration used an implicit solvent model, while the rMD simulations were in explicit solvent.

As the system is small (in fact, it is the smallest known pseudoknot), it was possible to exhaustively explore the energy landscape (Ma et al. 2021). From these simulations, we can again appreciate the presence of a rugged folding funnel (see Fig.3A). The energy landscape is characterized by one main funnel anchored by the native, experimentally observed structure. Some smaller subfunnels exist on the energy landscape, but only small barriers separate them from the main funnel. When analysing the ensemble of structures corresponding to these subfunnels, we detect partially folded states, but no states with alternative secondary structure competing with the native fold. The rMD simulations provided insight into the folding mechanism. As in the previous case, the folding pathways cross a multitude of metastable intermediate states (see Fig.3B).

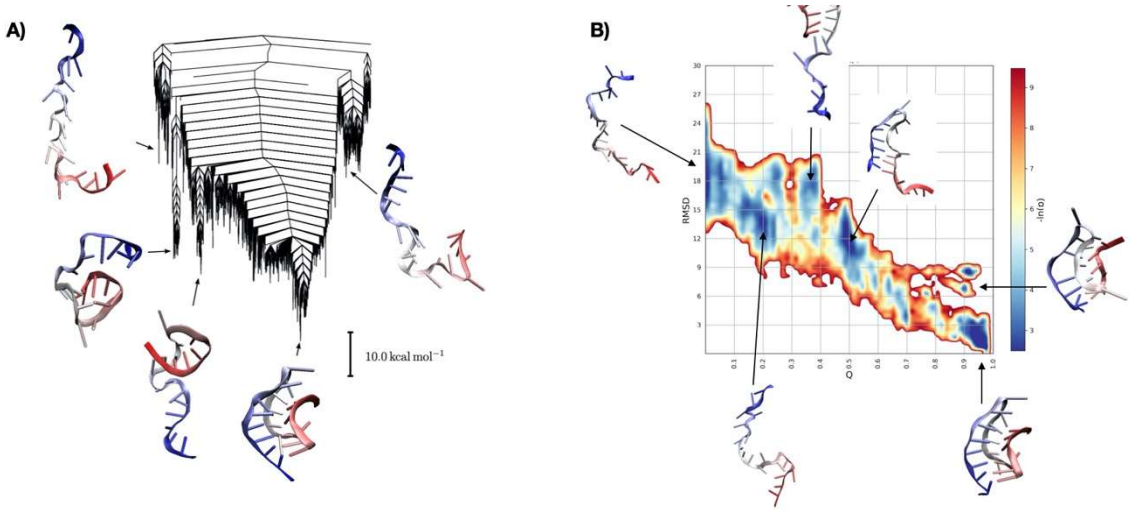


Figure 3: folding of the small H-pseudoknot PK1 studied with path sampling to obtain its energy landscape (A) and with biased folding simulations to study the native folding mechanism (B).

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

The two approaches support each other in this conclusion. While DPS provides a complete view of the energy landscape, the interaction with the solvent and the high entropy regions are not fully resolved, and rMD provided the missing details for the folding pathway.

Exploring polymorphism: 7SK RNA and KSHV's ORF50 transcript

7SK RNA is a non-coding RNA and part of a ribonucleoprotein complex, which is crucial to transcription regulation by RNA polymerase II (Wassarman and Steitz 1991). Its 5′ hairpin (HP1) was characterized experimentally by different methods including X-ray crystallography (Martinez-Zapien et al. 2017), NMR (Bourbigot et al. 2016), SAXS (Brillet et al. 2020), and chemical probing (Lebars et al. 2010; Olson et al. 2022). As a perfect example of RNA polymorphism, the high-resolution methods (NMR and X-ray) detected substantially different structures for this hairpin, including two distinct structures within the same crystal.

The three alternative structures are characterized by a reorganization of base pairing in the upper portion of the stem. The NMR structure is a hairpin with bulges and only canonical base pairing, while the X-ray structures exhibit non-canonical pairings and some triplets, organized differently in the two structures. The hairpin is the binding site for an affector protein (Egloff et al. 2018), which is crucial to the important biological function of RNA 7SK (Nguyen et al. 2001; Yang et al. 2001). Hence, understanding this structural polymorphism and its implication is of significant importance.

We used DPS and H-REX simulations to study the upper portion of HP1 (27 nucleotides), and further studied a set of mutations (Röder, Stirnemann, et al. 2020), reported to affect the binding affinity of HP1 (Martinez-Zapien et al. 2017). DPS was initiated from the experimentally observed

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

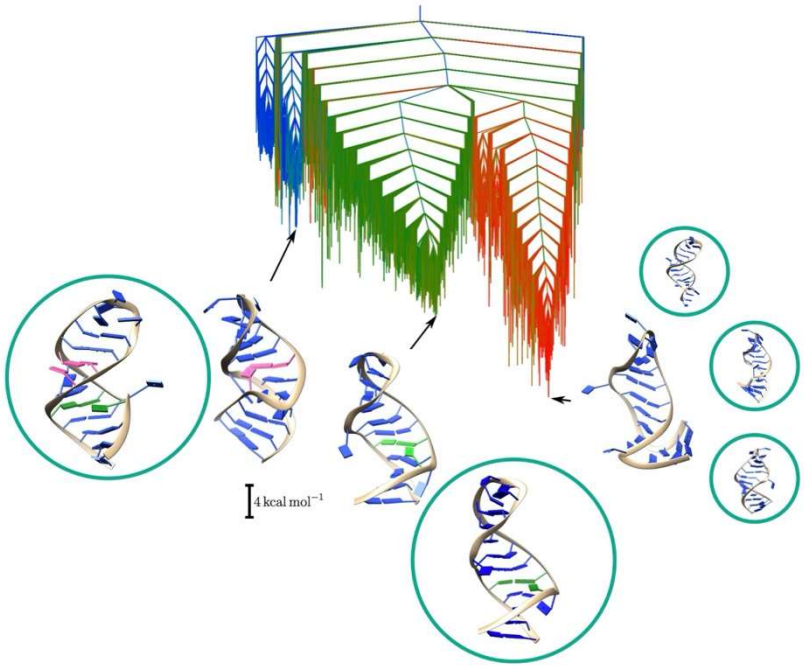


Figure 4: Energy landscape obtained from DPS for the 7SK RNA HP1 hairpin with key structures shown. The structural polymorphism is clearly observable, with three main funnels corresponding to more compact stem loops as observed in X-ray crystallography (blue and green), and more extended structures as observed by NMR experiments (red). Encircled structures correspond to the main clusters observed in H-REX simulations.

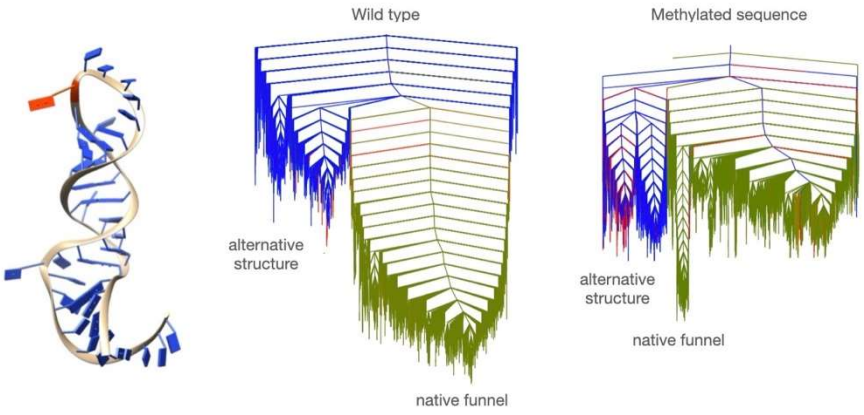
crystal structures and a multifunnel energy landscape was obtained from sampling, including descriptions of the relative stability and interconversion pathways (see Fig.4).[2]

The energy landscape revealed the polymorphic character of the HP1 hairpin, and based on the observed structures and their relative energies, we formulated a hypothesis relating the lowest energy X-ray hairpin structure to the binding of the affector protein. From our exploration of the energy landscapes for various mutants, we were able to draw a correlation between specific mutations and protein binding affinity, providing a mechanical explanation of the observed mutational effects.

Two sets of H-REX simulations complemented the energy landscape exploration with DPS, each starting from one of the observed crystal structures. Simulations were performed for 100ns using 20 replicas, with a Hamiltonian coupling λ ranging from 1 to 0. Due to the large size of the solvated system (roughly 30,000 atoms), convergence for such simulations is difficult to achieve. Nonetheless, the findings from the extended trajectories agree with the observations from pathsampling. The clusters of structures corresponded well to the structures found in the main funnels of the energy landscape. The agreement between the simulations enabled us to exclude a prominent role of structural water molecules or ions, which is not possible from the implicit solvent representation in used

[2] A detailed description of how this computational study was conducted is reported in (Röder and Pasquali 2021).

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press



in DPS. While DPS did not necessarily explore the high-energy portions of the landscape, which would require

Figure 5: EL of wild type and methylated sequence for ORF50. One of the lowest energies structures is shown with the site of methylation highlighted in red.

unfolding and potential refolding into different structures, the H-REX simulations did explore these regions. We would therefore expect H-REX to be able to depart more significantly from the initial structures and possibly find new structures with a full reorganization of the molecule. In our simulations, however, we only located states already explored by discrete pathsampling. The experimental evidence combined with the exploration of two different kind of simulations, give us some confidence that we have probably explored all the biologically relevant structures of the system.

Another area of growing interest is the study of post-translational modifications. In a recent study we investigated a post-translational methylation of an RNA hairpin (Röder, Barker, et al. 2022). As many RNAs are subject to methylation (Zaccara et al. 2019), it is important to understand how this modification impacts the adopted structures and how the equilibrium between possible alternative structures is altered. In the case of the RNA transcript of open reading frame 50 (ORF50) of Kaposi's sarcoma-associated herpes virus, which encodes the replication and transcription activator protein required for viral activation (Guito and Lukac 2012), methylation stabilises the RNA transcript, leading to effective viral replication (Baquero-Perez et al. 2019). Here, discrete pathsampling was used based on experimentally observed secondary structures (Baquero-Perez et al. 2019) only.

Our study revealed the existence of several structural basins, with the native structure occupying the lowest energy states. A set of higher energy structures, which allow interactions of the transcript with proteins (in this case a m6A reader), is found as well. In the unmethylated system, these structures are effectively inaccessible, while the methylation reduces the energy difference significantly, leading higher occupation of these states, which can recruit the m6A reader (see Fig.5).

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

These results highlight the importance of studies of mutations and chemical modifications, as the unmodified sequence might exhibit polymorphism, but it cannot be detected in experiment, while small alterations lead to significant changes in the energy landscape, which may lead to detectable polymorphism.

Conclusions and Perspectives

From these case studies, we can draw some general conclusions on the specificity of RNA folding and on what can constitute a profitable strategy to tackle larger and more complex systems. For all systems we have studied, even the simplest, we find a rugged energy landscape, i.e it is characterized by the presence of many locally stabilised structures, corresponding to the subsequent formation of local secondary structures. These configurations may be part of a single funnel, most likely for small systems, or, if they exhibit significant difference in their secondary structure, belong to competing funnels, a feature likely observed for larger RNA molecules.

A successful strategy to investigate the possible structures is the combination of several simulation methods. Secondary structure predictions based on bioinformatics may already reveal some of the complexity of the RNA folding problem. If a single dominant structure emerges, it is likely that the energy landscape is less complex and exhibits a single main funnel. In such cases, it is likely that canonical base pairs are adopted, and chemical probing may give further confidence in such predictions. Non-canonical interactions may be important to local structural details. Of course, in these scenarios, complexity may arise from additional effects, such as post-translational modifications.

If, on the other hand, multiple secondary structures are proposed, a multiscale approach can be useful. Using a coarse-grained model, initial scouting of the energy landscape can yield a survey of possible alternative structures, leading to an identification of the main folding funnels. Subsequent all atom simulations can then be used to investigate details on the energy landscape, seeded by the structures obtained from the coarse-grained simulations.

With this approach in mind, we recently started investigating the frameshifting pseudoknot of SARS-CoV-2, for which a structure is known experimentally but for which both experimental and simulation data suggest the existence of alternative structures. We first simulated the system at a relatively high temperature with the CG model to generate seeds for DPS in order to speed up the energy landscape exploration. Then, using discrete pathsampling, a search for the conversion path between the native structure and a proposed structure lacking the characteristic pseudoknot was initiated with these seeds. Given the size of the system, simulations are computationally very demanding and still ongoing, but preliminary results show the possible conversion path between the two states (Fig.6).

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

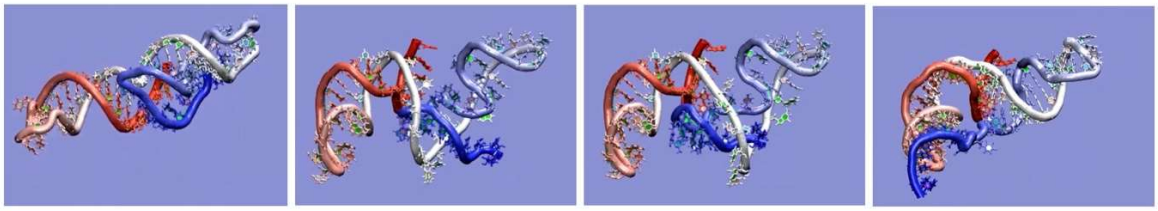


Figure 6: Sample structures on the conversion pathway between the native state of the frameshifting pseudoknot of SARS-CoV-2 (left) and an alternative structure with no pseudoknot (right).

Another research direction is based on the extensive conformational sampling and access to free-energy landscape explorations provided by H-REX, which could offer decisive insights into key phenomena related to the RNA World hypothesis and the origins of life, as currently studied by some of us. Previous studies on protein enzyme systems have shown the crucial importance of proper conformational sampling of the reactant and product states to understand the chemical reactivity of these biological objects (Maffucci, Laage, Sterpone, et al. 2020). We thus aim at understanding how a ribozyme's accessible conformations can affect its reactivity. Secondly, the end product of template-based RNA replication in abiotic conditions is an RNA dimer. However, these are known to be very stable constructs, with high denaturation temperatures. Inspired by the use of the REST2 strategy for the study of protein melting properties (Maffucci, Laage, Stirnemann, et al. 2020; Stirnemann and Sterpone 2015; Stirnemann and Sterpone 2017), we are currently trying to understand how RNA duplexes separate upon temperature increase, and how this depends on the strand sequence.

While many computational methods exist to study biomolecules, the challenges encountered by the complexity of RNA folding means that the best strategy for computational studies to yield useful biological insight, rests on the combination of multiple approaches to overcome individual shortcomings and access time and size scales otherwise inaccessible. While data-based structural predictions are important, they require the additional insight from physical modelling, especially to understanding the dynamic, polymorphic nature of RNA. Simulations of RNA have been used for decades, but the maturity of many methods and the growing understanding of RNA means that a new chapter of research has opened up, where computational approaches, if used properly, will routinely provide exciting insights into the nature of RNA.

Author Contributions

KR, GS, PF and SP all contributed to the developments of the various methods presented in the manuscript and all contributed to its writing. SP conceived the article and made the figures.

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

Financial Support

KR is funded by the Cambridge Philosophical Society. KR and SP thank Université Paris Cité for a visiting fellowship for KR. Some of the research mentioned here has received funding from the European Research Council under the European Union's Eighth Framework Program (H2020/20142020)/ERC Grant Agreement No. 757111 (G.S.). This work was also supported by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01 to GS).SP is thankful for the computing time allocated to the French national computing grant A0090710584. This work was partially supported by a STSM Grant from COST Action CA17139 (eutopia.unitn.eu) funded by COST (www.cost.eu).

Conflicts of Interest

KR, GS and SP do not have any conflict of interest. P.F. is co-founder and shareholder of Sibylla Biotech SPA, a company exploiting molecular simulations to perform early-stage drug discovery.

Data availability statement

Simulation data is available through links relative to the original articles presenting the material.

References

A Beccara, S., Fant, L., & Faccioli, P. (2015). Variational scheme to compute protein reaction pathways using atomistic force fields with explicit solvent. Physical Review Letters, 114(9), 098103.

Baquero-Perez, B., Antanaviciute, A., Yonchev, I. D., Carr, I. M., Wilson, S. A., & Whitehouse, A. (2019). The tudor snd1 protein is an m6a rna reader essential for replication of kaposi's sarcoma-associated herpesvirus. eLife, 8, e47261.

Bartolucci, G., Orioli, S., & Faccioli, P. (2018). Transition path theory from biased simulations. Journal of Chemical Physics, 149(7), 072336.

Biyun, S., Cho, S. S., & Thirumalai, D. (2011). Folding of human telomerase rna pseudoknot using ion-jump and temperature-quench simulations. Journal of the American Chemical Society, 133(50), 20634–20643.

Bottaro, S., Nichols, P. J., Vögeli, B., Parrinello, M., & Lindorff-Larsen, K. (2020). Integrating NMR and simulations reveals motions in the UUCG tetraloop. Nucleic Acids Research, 48(11), 5839–5848. https://doi.org/10.1093/nar/gkaa399

Bourbigot, S., Dock-Bregeon, A.-C., Eberling, P., Coutant, J., Kieffer, B., & Lebars, I. (2016). Solution structure of the 5'-terminal hairpin of the 7SK small nuclear RNA. RNA, 22(12), 1844–1858.

Brillet, K., Martinez-Zapien, D., Bec, G., Ennifar, E., Dock-Bregeon, A.-C., & Lebars, I. (2020). Different views of the dynamic landscape covered by the 5'-hairpin of the 7SK small nuclear RNA. RNA, 26(9), 1184–1197.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D., & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins, 21(3), 167–195.

Bryngelson, J. D., & Wolynes, P. G. (1987). Spin glasses and the statistical mechanics of protein folding. Proceedings of the National Academy of Sciences USA, 84, 7524–7528.

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex dna: Sequence, topology and structure. Nucleic Acids Research, 34(19), 5402–5415.

Cameron, M., & Vanden-Eijnden, E. (2014). Flows in complex networks: Theory, algorithms, and application to Lennard-Jones cluster rearrangement. Journal of Statistical Physics, 156(3), 427–454.

Camilloni, C., Broglia, R. A., & Tiana, G. (2011). Hierarchy of folding and unfolding events of protein g, ci2, and acbp from explicit-solvent simulations. Journal of Chemical Physics, 134(4), 045105.

Cho, S. S., Pincus, D. L., & Thirumalai, D. (2009). Assembly mechanisms of rna pseudoknots are determined by the stabilities of constituent secondary structures. Proceedings of the National Academy of Sciences 106(41), 17349–17354.

Cragnolini, T., Chakraborty, D., Šponer, J., Derreumaux, P., Pasquali, S., & Wales, D. J. (2017). Multifunctional energy landscape for a dna g-quadruplex: An evolved molecular switch. Journal of Chemical Physics, 147(15), 152715.

Cragnolini, T., Laurin, Y., Derreumaux, P., & Pasquali, S. (2015). Coarse-grained hire-rna model for ab initio rna folding beyond simple molecules, including noncanonical and multiple base pairings. Journal of Chemical Theory and Computation, 11(7), 3510–3522.

Denesyuk, N. A., & Thirumalai, D. (2011). Crowding promotes the switch from hairpin to pseudoknot conformation in human telomerase rna. Journal of the American Chemical Society, 133(31), 11858–11861.

E, W., & Vanden-Eijnden, E. (2010). Transition-path theory and path-finding algorithms for the study of rare events. Annual Review of Physical Chemistry, 61, 391–420.

Egloff, S., C, S., & Kiss, T. (2018). 7sk small nuclear rna, a multifunctional transcriptional regulatory rna with gene-specific features. Transcription, 9, 95–101.

Fay, M. M., Lyons, S. M., & Ivanov, P. (2017). RNA G-quadruplexes in biology: Principles and molecular mechanisms. Journal of Molecular Biology, 429(14), 2127–2147.

Ferreiro, D. U., Komives, E. A., & Wolynes, P. G. (2014). Frustration in biomolecules. Quarterly Review in Biophysics, 47, 285–363.

Fiorini, G., Klein, M. L., & Hénin, J. (2013). Using collective variables to drive molecular dynamics simulations. Molecular Physics, 111, 3345–3362.

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

564 Garst, A. D., Edwards, A. L., & Batey, R. T. (2011). Riboswitches: Structures and
565 mechanisms. Cold Spring Harbor Perspectives in Biology, 3(6).
566 Gavory, G., Symmons, M. F., Ghosh, Y. K., Klenerman, D., & Balsubramanian, S. (2006).
567 Structural analysis of the catalytic core of human telomerase rna by fret and
568 molecular modeling.
569 Biochemistry, 45(44), 13304–13311.
570 Gillet, N., Bartocci, A., & Dumont, E. (2021). Assessing the sequence dependence of
571 pyrimidine–pyrimidone (6–4) photoproduct in a duplex double-stranded dna: A
572 pitfall for microsecond range simulation. The Journal of Chemical Physics, 154,
573 135103.
574 Guito, J., & Lukac, D. M. (2012). KSHV RTA promoter specification and viral reactivation.
575 Frontiers in Microbiology, 3, 30.
576 Halvorsen, M., Martin, J. S., Broadaway, S., & Laederach, A. (2010). Disease-associated
577 mutations that alter the rna structural ensemble. PLoS Genetics, 6(8), e1001074.
578 Jaeger, L., Verzemnieks, E., & Geary, C. (2009). The UA_handle: A versatile submotif in
579 stable rna architectures. Nucleic Acids Research, 37(1), 215–230.
580 Joseph, J. A., Röder, K., Chakraborty, D., Mantell, R. G., & Wales, D. J. (2017). Exploring
581 biomolecular energy landscapes. Chemical Communnications, 53, 6974–6988.
582 Kim, N.-K., Zhang, Q., Zhou, J., Theimer, C. A., Peterson, R. D., & Feigon, J. (2008). Solution
583 structure and dynamics of the wild-type pseudoknot of human telomerase rna.
584 Journal of Molecular Biology, 384(5), 1249–1261.
585 Kolesnikova, S., & Curtis, E. A. (2019). Structure and function of multimeric g-
586 quadruplexes. Molecules, 24(17), 3074.
587 Kührová P., Best, R. B., Bottaro, S., Bussi, G., Šponer, J., Otyepka, M., & Banáš, P. (2016).
588 Computer folding of rna tetraloops: Identification of key force field deficiencies.
589 Journal of Chemical Theory and Computation, 12(9), 4534–4548.
590 Lazzeri, G., Micheletti, C., Pasquali, S., & Faccioli, P. (2022). Rna folding landscapes from
591 explicit
592 solvent all-atom simulations. arXiv:2205.12603.
593 Lebars, I., Martinez-Zapien, D., Durand, A., Coutant, J., Kieffer, B., & Dock-Bregeon, A.-C.
594 (2010). HEXIM1 targets a repeated GAUC motif in the riboregulator of
595 transcription 7SK and promotes base pair rearrangements. Nucleic Acids
596 Research, 38(21), 7749–7763.
597 Leontis, N. B., & Westhof, E. (2001). Geometric nomenclature and classification of rna base
598 pairs. RNA, 7(4), 499–512.
599 Leopold, P. E., Montal, M., & Onuchic, J. N. (1992). Protein folding funnels: a kinetic
600 approach to the sequence-structure relationship. Proceedings of the National
601 Academy of Sciences USA, 89(18), 8721–8725.
602 Li, J., & Chen, S.-J. (2021). RNA 3d structure prediction using coarse-grained models.
603 Frontiers in Molecular Biosciences, 8, 720937.

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

Li, P. T., Vieregg, J., & Tinoco, I. (2008). How rna unfolds and refolds. Annual Review of Biochemistry, 77, 77–100.

Lightfoot, H. L., Hagen, T., Tatum, N. J., & Hall, J. (2019). The diverse structural landscape of quadruplexes. FEBS Letters, 593(16), 2083–2102.

Liu, N., Zhou, K. I., Parisien, M., Dai, Q., Diatchenko, L., & Pan, T. (2017). N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. Nucleic Acids Research, 45(10), 6051–6063.

Lorenz, R., Wolfinger, M. T., Tanzer, A., & Hofacker, I. L. (2016). Predicting rna secondary structures from sequence and probing data. Methods, 103, 86–98.

Ma, Y., Monsavior, A., Pasquali, S., & Röder, K. (2021). Comparison of coarse-grained and all-atom representations by explicit energy landscape explorations. chemRxiv, PrePrint.

Maffucci, I., Laage, D., Sterpone, F., & Stirnemann, G. (2020). Thermal adaptation of enzymes: Impacts of conformational shifts on catalytic activation energy and optimum temperature. Chemistry-A European Journal, 26(44), 10045–10056.

Maffucci, I., Laage, D., Stirnemann, G., & Sterpone, F. (2020). Differences in thermal structural changes and melting between mesophilic and thermophilic dihydrofolate reductase enzymes. Physical Chemistry Chemical Physics, 22(33), 18361–18373.

Martinez-Zapien, D., Legrand, P., McEwen, A. G., Proux, F., Cragnolini, T., Pasquali, S., & DockBregeon, A.-C. (2017). The crystal structure of the 5′ functional domain of the transcription riboregulator 7SK. Nucleic Acids Research, 45(6), 3568–3579.

Mazzanti, L., Alferkh, L., Frezza, E., & Pasquali, S. (2021). Biasing rna coarse-grained folding simulations with small-angle x-ray scattering data. Journal of Chemical Theory and Computation, 17(10), 6509–6521.

Mlýnský, V., & Bussi, G. (2018). Exploring rna structure and dynamics through enhanced sampling simulations. Current Opinion in Structural Biology, 49, 63–71.

Mlýnský, V., Janeček, M., Kührová, P., Fröhlking, T., Otyepka, M., Bussi, G., Banáš, P., & Šponer, J. (2022). Toward convergence in folding simulations of rna tetraloops: Comparison of enhanced sampling techniques and effects of force field modifications. Journal of Chemical Theory and Computation, 18(4), 2642–2656.

Nguyen, V. T., Kiss, T., Michels, A., & Bensaude, O. (2001). 7sk small nuclear rna binds to and inhibits the activity of cdk9/cyclin t complexes. Nature, (414), 322–325.

Nonin-Lecomte, S., Felden, B., & Dardel, F. (2006). Nmr structure of the aquifex aeolicus tmrna pseudoknot pk1: New insights into the recoding event of the ribosomal trans-translation. Nucleic Acids Research, 34(6), 1847–1853.

Olson, S. W., Turner, A.-M. W., Arney, J. W., Saleem, I., Weidmann, C. A., Margolis, D. M., Weeks, K. M., & Mustoe, A. M. (2022). Discovery of a large-scale, cell-state-responsive allosteric switch in the 7sk rna using dance-map. Molecular Cell, 82(9), 1708–1723.

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

644 Onsager, L., & Machlup, S. (1951). Fluctuations and irreversible processes. The Physical
645 Review, 91, 1505.
646 Paci, E., & Karplus, M. (1999). Forced unfolding of fibronectin type 3 modules: An analysis
647 by biased molecular dynamics simulations. Journal of Molecular Biology, 288(3),
648 441–459.
649 Papoian, G. A. (2018). Coarse-grained modeling of biomolecules. CRC Press.
650 Pasquali, S., & Derreumaux, P. (2010). HiRE-RNA: A high resolution coarse-grained energy
651 model for RNA. Journal of Physical Chemistry B, 114(37), 11957–11966.
652 Pasquali, S., Frezza, E., & Barroso da Silva, F. (2019). Coarse-grained dynamic rna titration
653 simulations. Interface Focus, 9, 20180066.
654 Perez, A., Marchan, I., Svozil, D., Šponer, J., Cheatham, T., Laughton, C., & Orozco, M. (2007).
655 Refinement of the amber force field for nucleic ncids: Improving the description
656 of alpha/gamma conformers. Biophysical Journal, 92, 3817–3829.
657 Rissone, P., Bizarro, C. V., & Ritort, F. (2022). Stem-loop formation drives rna folding in
658 mechanical unzipping experiments. Proceedings of the National Academy of
659 Science USA, 119(3), e2025575119.
660 Röder, K., Barker, A. M., Whitehouse, A., & Pasquali, S. (2022). Investigating the structural
661 changes due to adenosine methylation of the kaposi's sarcoma-associated herpes
662 virus orf50 transcript. PLoS Computational Biology, 18(5), e1010150.
663 Röder, K., & Wales, D. J. (2018). Evolved minimal frustration in multifunctional
664 biomolecules. Journal of Physical Chemistry B, 122(49), 10989–10995.
665 Röder, K., Joseph, J. A., Husic, B. E., & Wales, D. J. (2019). Energy landscapes for proteins:
666 From single funnels to multifunctional systems. Advanced Theory and
667 Simulations, 2, 1800175.
668 Röder, K., & Pasquali, S. (2021). Rna modelling with the computational energy landscape
669 framework. In L. Ponchon (Ed.), Rna scaffolds (pp. 49–66). Humana.
670 Röder, K., Stirnemann, G., Dock-Bregeon, A.-C., Wales, D. J., & Pasquali, S. (2020). Structural
671 transitions in the rna 7sk 5′ hairpin and their effect on hexim binding. Nucleic
672 Acids Research, 48(1), 373–389.
673 Saldi, T., Riemondy, K., Erickson, B., & Bentley, D. L. (2021). Alternative RNA structures
674 formed during transcription depend on elongation rate and modify RNA
675 processing. Molecular Cell, 81(8), 1789–1801.
676 Sato, K., & Kato, Y. (2022). Prediction of rna secondary structure including pseudoknots
677 for long
678 sequences. Briefings in Bioinformatics, 23(1), 1–9.
679 Schlick, T., Zhu, Q., Jain, S., & Yan, S. (2021). Structure-altering mutations of the sars-cov-2
680 frameshifting rna element. Biophysical Journal, 120(6), 1040–1053.
681 Šponer, J., Bussi, G., Krepl, M., Banáš, P., Bottaro, S., Cunha, R. A., Gil-Ley, A., Pinamonti,
682 G.,Poblete, S., Jurečka, P., Walter, N. G., & Otyepka, M. (2018). RNA structural
683 dynamics as captured by molecular simulations: A comprehensive overview.
684 Chemical Reviews, 118(8), 4177–4338.

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

685 Stadlbauer, P., Mazzanti, L., Cragnolini, T., Wales, D. J., Derreumaux, P., Pasquali, S., &
686 Šponer, J.(2016). Coarse-grained simulations complemented by atomistic
687 molecular dynamics provide new insights into folding and unfolding of human
688 telomeric G-quadruplexes. Journal of Chemical Theory and Com 12(12), 6077–
689 6097.
690 Stirnemann, G., & Sterpone, F. (2015). Recovering protein thermal stability using all-atom
691 hamiltonian replica-exchange simulations in explicit solvent. Journal of Chemical
692 Theory and Computation, 11(12), 5573–5577.
693 Stirnemann, G., & Sterpone, F. (2017). Mechanics of protein adaptation to high
694 temperatures. Journal of Physical Chemistry Letters, 8(23), 5884–5890.
695 Stombaugh, J., Zirbel, C. L., Westhof, E., & Leontis, N. B. (2009). Frequency and isostericity
696 of rna base pairs. Nucleic Acids Research, 37(7), 2294–2312.
697 Sugita, Y., & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for
698 protein folding. Chemical Physics Letters, 314(1), 141–151.
699 Theimer, C. A., Blois, C. A., & Feigon, J. (2005). Structure of the human telomerase rna
700 pseudoknot
701 reveals conserved tertiary interactions essential for function. Molecular Cell,
702 17(5), 671–682.
703 Tiana, G., & Camilloni, C. (2012). Ratcheted molecular-dynamics simulations identify
704 efficiently the transition state of protein folding. Journal of Chemical Physics,
705 137(23), 235101.
706 Tinoco, I., & Bustamante, C. (1999). How rna folds. Journal of Molecular Biology, 293(2),
707 271–281. Vicens, Q., & Kieft, J. S. (2022). Thoughts on how to think (and talk) about rna
708 structure.
709 Proceedings of the National Acaddemy of Science USA, 119(17), e2112677119.
710 Wales, D. J. (2002). Discrete path sampling. Molecular Physics, 100(20), 3285–3305.
711 Wales, D. J. (2004). Some further applications of discrete path sampling to cluster
712 isomerization. Molecular Physics, 102(9-10), 891–908.
713 Wales, D. J., & Salamon, P. (2014). Observation time scale, free-energy landscapes, and
714 molecular symmetry. Proceedings of the National Academy of Sciences USA,
715 111(2), 617–622.
716 Wang, L., Friesner, R. A., & Berne, B. (2011). Replica exchange with solute scaling: A more
717 efficient version of replica exchange with solute tempering (rest2). Journal of
718 Physical Chemistry B, 115(30), 9431–9438.
719 Wang, L., Liu, Y., Zhong, X., Liu, H., Lu, C., Li, C., & Zhang, H. (2019). Dmfold: A novel method
720 to predict rna secondary structure with pseudoknots based on deep learning and
721 improved base pair maximization principle. Frontiers in Genetics, 10, 143.
722 Wassarman, D. A., & Steitz, J. A. (1991). Structural analyses of the 7SK ribonucleoprotein
723 (RNP), the most abundant human small RNP of unknown function. Molecular and
724 Cellular Biology, 11(7), 3432–3445.
725 Yan, S., Zhu, Q., Jain, S., & Schlick, T. (2022). Length-dependent motions of sars-cov-2
726 frameshifting rna pseudoknot and alternative conformations suggest avenues for

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press

frameshifting suppression. Research square, Preprint. https://doi.org/10.21203/rs.3.rs-1160075/v1

Yang, Z., Zhu, Q., Luo, K., & Zhou, Q. (2001). The 7sk small nuclear rna inhibits the cdk9/cyclin t1 kinase to control transcription. Nature, (414), 317–322.

Yu, A. M., Gasper, P. M., Cheng, L., Lai, L. B., Kaur, S., Gopalan, V., Chen, A. A., & Lucks, J. B. (2021). Computationally reconstructing cotranscriptional rna folding from experimental data reveals rearrangement of non-native folding intermediates. Molecular Cell, 81(4), 870–883.

Zaccara, S., Ries, R. J., & Jaffrey, S. R. (2019). Reading, writing and erasing mRNA methylation. Nature Reviews Molecular Cell Biology, 20(10), 608–624.

Zgarbová, M., Otyepka, M., Šponer, J., Mládek, A., Banáš, P., Cheatham, T. E., & Jurečka, P. (2011).˘ Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. Journal of Chemical Theory and Computation, 7(9), 2886–2902.

Zhao, Q., Zhao, Z., Fan, X., Yuan, Z., Mao, Q., & Yao, Y. (2021). Review of machine learning methods for rna secondary structure prediction. PLoS Computational Biology, 17(8), e1009291.

https://doi.org/10.1017/qrd.2022.19 Published online by Cambridge University Press