# Comparison of coarse-grained and all-atom representations by explicit energy landscape explorations

Yuxin Ma,[†] Antoine Monsavoir,[‡] Samuela Pasquali,[‡,¶] and Konstantin Röder[*,†]

†*Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK*

‡*Laboratoire Cibles Thérapeutiques et Conception de Médicaments, CNRS UMR 8038,*

*Faculté des Sciences Pharmaceutiques et Biologiques, Université de Paris, Paris, France*

¶*Equipe Modélisation Computationnelle des Interactions Protéine-Ligand, Laboratoire*

*Biologie Fonctionnelle et Adaptative, CNRS UMR 8251, Université de Paris, Paris, France*

E-mail: kr366@cam.ac.uk

## Abstract

Computational studies of large molecular systems are often hindered by resource constraints, such as the available computational time. A common approach to reduce the computational cost is to use a coarse-grained description instead of an all-atom representation. However, such a simplification requires careful consideration of the coarse-graining scheme to identify potential artefacts introduced and the limitations of the model. In this contribution, we use the computational energy landscape framework to explicitly explore the energy landscapes for a coarse-grained (HiRE-RNA) and an all-atom potential (AMBER) for an example system, the *Aquifex aeolicus* tmRNA pseudoknot PK1. The method provides insight into structural, thermodynamic and kinetic properties within a common framework, and allows for a comparison of a variety of commonly computed observables, demonstrating the usefulness of this approach.

1

For the specific case study, we observe that both potentials exhibit a number of common features, highlighting that the coarse-grained model captures essential physical features of the system. Nonetheless, we observe shortcomings, and we demonstrate how our approach allows us to improve the model based on the insight obtained from the computational modelling.

# 1 Introduction

The functional importance of RNA in transcription for protein synthesis in living cells is well known. However, RNA molecules are not limited to this important role, but they are also involved in many regulatory processes.[1–3] While the function fulfilled by RNA molecules is strongly related to their three-dimensional structure,[4] RNA molecules exhibit intrinsically a structural plasticity, which leads to a number of possible configurations of comparable energy. This observation is strongly linked to the underlying energy landscape, which in these cases consists of multiple funnels, a feature strongly associated with multifunctionality.[5] This relationship has been shown explicitly for a range of systems, including RNA tetraloops[6] and 7SK RNA.[7,8]

While this multifunctionality proves highly important in our understanding of the role non-coding RNA plays in nature, it leads to a fundamental problem for experimental and computational studies. The multifunnel character of the underlying energy landscapes exhibits high energy barriers, which result in slow transition processes and kinetic trapping. This observation is often framed in terms of the kinetic partition mechanism,[9] and lies at the bottom of many seemingly contradictory observations for example as reported for G-quadruplexes.[10] *In silico*, the key consequence of this behaviour is an increased computational cost and the requirement to use enhanced sampling algorithms to obtain a faithful representation of the thermodynamic and kinetic behaviour of the molecules studies.

A common strategy employed to overcome these challenges is the use of coarse-grained (CG) potentials in the study of RNAs. The advantage of this approach is twofold: Firstly,

the reduced number of particles lowers the computational cost required, allowing for the investigation of larger systems. Secondly, the coarse-graining smooths the energy landscape, allowing the sampling to converge faster. The popularity of this approach can be seen in the large number of available CG models for RNAs.[11–16]

Usually these CG representations are designed with a particular application in mind, and there is no unique way to derive a CG representation for RNA.[17] Given an appropriate representation, CG models are successful in providing insight into multifunctional systems.[18] Nonetheless, how faithful a CG model represents a system will strongly depend on a number of factors,[17] and careful consideration should be given before CG representations are used. Given that many such models are based on specific experimental data used to derive the necessary parameters, how well a CG representation matches the underlying physics is not obvious.[19]

Of course this problem is universal for parameter based potentials, and applies for all atom force fields as well. However, the coarse-graining may introduce artefacts due to the required dimensionality reduction, or rely on coarse-graining parameters without clear correspondence to an all-atom representation.[19] A careful evaluation of CG models is therefore required.

The potential energy landscape for any molecular system contains all necessary information to calculate thermodynamic, kinetic and structural properties. Due to the unique topography associated with the complex interactions encountered in biomolecules the landscape in itself provides an interesting method of characterisation for a molecular system. Through the potential energy landscape framework[20,21] we have direct access to the energy landscape and its topography, as well as thermodynamic, kinetic and structural observables. We therefore propose that explicit explorations of energy landscapes of RNA molecules using CG and all-atom representations will provide a unique way of evaluating the validity of CG models. While all-atom force fields also suffer from shortcomings, they are much better studied, and the physical motivation behind them is better understood.[17]

In this study, we apply this approach to a small RNA pseudoknot using an AMBER

all-atom force field in comparison with the coarse-grained representation given by HiRE-RNA.[11,12] We observe that the proposed comparison methodology provides us with a large range of observables to compare, and we further report a good agreement between the CG and all-atom models for the chosen example system for some key observables, and indicate how future development of the CG force field may be driven by such a comparison of energy landscapes. In particular, we discuss a variation of HiRE-RNA, where we alter the stacking and torsional potential through the introduction of additional parameters to overcome some shortcomings observed in the CG model.

## 2  Methodology

### 2.1  The example system

The example system chosen for this work is the Pseudoknot1 (PK1) from the thermophilic bacterium *A.aeolicus*, which is predicted to be the smallest of all transfer-messenger RNA (tmRNA) with a size of 21 nucleotides, and has been studied by NMR experiments.[22] PK1 is involved in trans-translation, a ribosomal rescuing mechanism in bacteria.[23] There are a number of reason for our choice of this system. RNA pseudoknots are a commonly encountered motif, but the motif is not sequence dependent and therefore poses a challenge to RNA structure prediction.[24-26] At the same time, pseudoknots are involved in a large number of important processes,[1] such as regulation of gene expression,[27] splicing,[28] telomerase activity,[29] frameshifting,[30-32] and ribozyme catalysis.[33] This importance has lead to a large number of studies, providing us with experimental insight into the behaviour of RNA pseudoknots, which can be applied in analysing our results. Last but not least, the small size of PK1 means we can explore the coarse-grained and all-atom landscapes at low computational cost.

## 2.2 Force field specifications

The AMBER force field employed is the ff99[34] force field with the Barcelona $\alpha/\gamma$ backbone modification[35] and the $\chi$ modification tuned for RNA[36,37] with implicit solvent ($igb = 2$). The HiRE-RNA model describes each nucleotide by 6 or 7 beads – a single particle each for the phosphate and group and backbone oxygen, three beads for the sugar, and a single bead for pyrimidine bases and two particles for purine bases. The CG representation implicitly accounts for solvation and ions. The model is specifically designed to represent non-standard RNA structures, including high accuracy of base stacking and canonical as well as non-canonical base pairings.[11,12]

## 2.3 Exploration of the energy landscapes

The exploration of the energy landscapes is based on the computational potential energy landscape framework,[20,21] employing discrete pathsampling (DPS)[38,39] to derive kinetic transition networks.[40,41] Low energy minima to initiate the landscape explorations were located using basin-hopping global optimisation[42–44] For both potentials, searches were initiated from sequence and from the available PDB entry (2G1W).[22] Transition state candidates were located with the doubly-nudged elastic band (DNEB) algorithm,[45–47] employing the quasi-continuous interpolation (QCI) scheme.[48,49] The candidate structures were converged to true transition states using hybrid eigenvector-following (HEF).[50]

The energy landscapes are represented using disconnectivity graphs.[51,52]

## 2.4 Parameter optimisation for the extended CG scheme

HiRE-RNA force-field consists of local interactions (bond, angles and torsions), electrostatics, excluded volume, base pairing and staking interactions. All these terms contain a specific scaling constant and their sum gives the overall energy of the system. Considering all possible interactions accounted for, the original HiRE-RNA model has 25 external parameters

that can be varied upon optimisation, with a single energy parameter to optimise torsional energies, and a generic stacking potential independent of the nature of the nucleotide.

After a first comparison between the energy landscapes obtained from AMBER and the one from HiRE-RNA highlighted some shortcomings of the CG model, we introduced a new parametrisation scheme of the coarse-grained model that accounts in more detail torsions and staking. We then introduced a new scheme including 47 energy parameters in which the energies associated to each torsion can be varied independently upon optimisation, and stacking parameters are different for purine-purine, purine-pyrimidine, and pyrimidine-pyrimidine interactions. We will refer to the original HiRE-RNA as HiRE-RNA-25, and to the new version ad HiRE-RNA-47.

HiRE-RNA-47 was optimised using a genetic algorithm maximising the difference in total energy between native structures and decoys on a set of 16 RNA motifs extracted from the RNA motifs atlas,[53] including stems, internal and apical loops. For each structure we generated 20 decoys both preserving base pairing (non-optimal folded structures) and partially unfolded. For each new set of parameters during the optimisation, energy evaluation was performed after quenching each decoy in order to reach a local minimum and have a structure effectively competing with the native. The genetic algorithm ran with 80 genes and with up to 100 generations, though often convergence was reached before.

# 3    Results and discussion

In this section, we will present and discuss the results of the comparison of the energy landscapes, followed by an account of how changes introduced to the CG potential improve the properties of the structural ensembles observed.

## 3.1 The AMBER energy landscape for PK1

The free energy disconnectivity graph for AMBER is shown in Fig. 1. The global minimum is the correctly folded hairpin, with all native contacts formed. Higher energy configuration in the same funnel exhibit some degree of unfolding, with the 3′-end unfolding first. Higher in energy we observe funnels with partially folded structures, and unfolded structures highest in energy. The unfolding pathways consist of unfolding of the 3′-end base pairs, followed by unfolding of the 5′-end base pairs. This order of unfolding events is expected, as the unfolding order will be determined by individual stabilities of the helical segments,[54] and the 5′-end helix contains more base pairs than the 3′-end helix. The formation of H-type pseudoknots like PK1 has been described before,[1] and such hierarchical folding is observed in other pseudoknots as well.[30]

Overall, we find that the energy landscape for the AMBER potential represents the system well, and our findings are in good agreement with experiment. Hence, it is a reasonable model to compare with the CG representation. The one shortcoming of the energy landscape exploration in this case is the undersampling of the unfolded region, but the folded and partly folded ensembles, as well as the relative energies are representative.

## 3.2 The HiRE-RNA-25 energy landscape for PK1

The free energy disconnectivity graph for the HiRE-RNA model with 25 parameters is presented in Fig. 2. Significant differences with the all-atom model are observed, while some features are preserved.

One point of agreement between AMBER and CG model are the correct folding of the global minimum. We again observe the native structure as the global minimum, and its structure and the contacts formed are in good agreement with experimental structures. As the HiRE-RNA force field was developed using MD simulations with the aim to obtain folded native structures, it is perhaps not surprising that the native structure is found accordingly. Nonetheless, this highlights that the CG correctly stabilises the global minimum, and we do
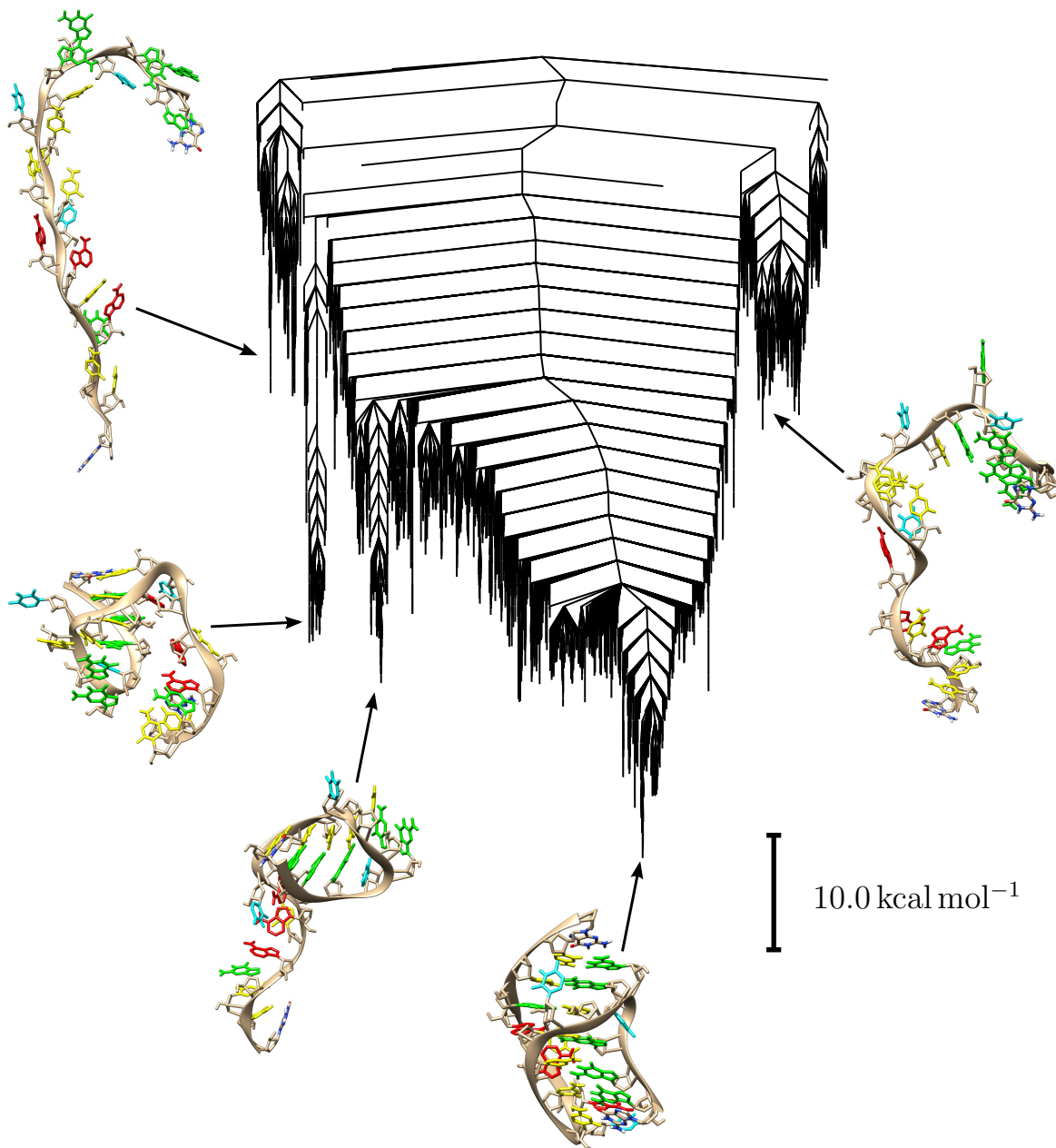
Figure 1: Free energy disconnectivity graph for the PK1 pseudoknot at 310 K using the all-atom AMBER potential.

not observe alternative low energy structures with different base pairing patterns. Around the native fold we observe a large number of structures with small deviations, indicating the start of the unfolding process. These structures, in agreement with the AMBER energy landscape and experiment, are showing some unfolding at the 3′-end.

However, the higher energy regions of the energy landscape are poorly represented, and not consistent with the energy landscape observed for AMBER. On the energy scale presented, we do not observe any fully unfolded structures. We located some such structures in our simulations, but their energies were off the scale. The unfolding structures we located at lower energies[1] maintain hydrogen-bonded base pairs, while we observe large distortions of the backbone. Interestingly, the CG model explores also molecules with an inverted chirality that are found in the high-energy funnel on the left of Fig. 2. However, contrary to what one may expect, these structures are still connected to the native funnel without adopting unfolded transition states. Instead we observe that the molecule switches conformation while preserving most base pairing while losing stacking an straining torsions.

Both of these observation, the very high energy of unfolded structures and the persistence of the hydrogen bonding patterns, hint at a relative overstabilisation of the hydrogen bonding compared to other energy contributions. While in AMBER such interactions are a result of the interplay between van der Waals and electrostatic interactions, in the HiRE-RNA potential, there are explicit terms for base pairing and stacking.

## 3.3   Improving the CG representation

While the low energy description of the system is in agreement between all atom and CG representations, clearly the CG model fails in the exploration of higher energy structures corresponding to the partially folded and unfolded ensembles. From the comparison between the energy landscapes of AMBER and of HiRE-RNA-25 it is clear that base pairing interactions are too strong compared to all other terms and possibly torsions are unbalanced, given they allow the switch in chirality. We therefore looked more carefully into these 3 terms and two main problems became evident. The first one, concerning torsion, was the fact that the strength of all torsions was adjusted as a whole in the optimisation of HiRE-RNA-25. The model accounted for different weights for different dihedral angles, but in the optimisation

---

[1]Some of these structures are shown with their location in the disconnectivity graph.
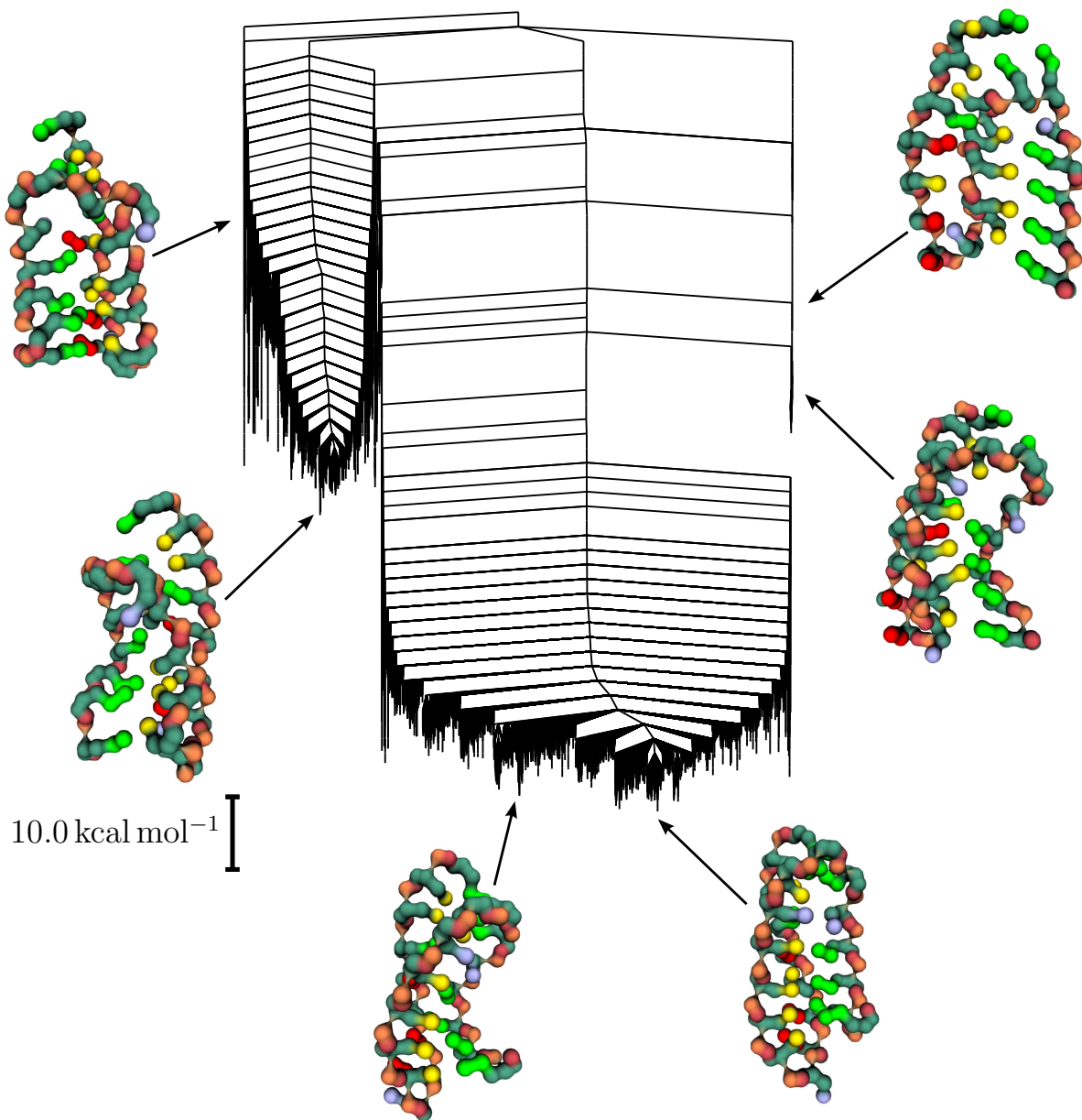
9

Figure 2: Free energy disconnectivity graph for the PK1 pseudoknot at 310 K using the coarse-grained HiRE potential with 25 scale parameters.

process their weights altered together uniformly. This coupling of the strength of all torsions resulted in some torsions being too weak while others are too strong. The second problem concerned stacking and the fact that we did not make any distinction between residues composed by one or two aromatic rings. During optimisation, this simplification also resulted in a compromise between a stronger interaction when two purines are stacked and a weaker

interaction when pyrimidines are staked.

These factors combines likely result in the observed overstabilisation of the base pairing. In HiRE-RNA-25, the base pairing description includes a more detailed description of various different possible interactions, and distinguishes between canonical and non-canonical interactions, as well as between base pairing formed by different nucleobase pairs. When optimising, the structures used to find good parameters will exhibit different base pairing and stacking, but only the former is detailed enough to allow the stabilisation of distinct interaction patterns over others. In other words, the difference between a U-U and a A-U base pair is known to the model, but the the difference between pyrimidine-purine and purine-purine stacking is not. As a result, more weight will be given to the base pairing, as it explains more of the structural variations.

To alleviate this problems, we introduced the HiRE-RNA-47 version where energy parameters of all torsions are independent from one another and three stacking terms, for purine-purine, purine-pyrimidine, pyrimidine-pyrimidine, replace the generic term for all nucleotides. Importantly, these additional parameters also allow for a better balancing between stacking, torsions and base pairing, once the parameters have been optimised.

### 3.3.1 Changes in the structural ensembles and energy contributions

To compare the details of the energetic contributions of HiRE-RNA-25 and HiRE-RNA-47 we consider the structures of the low-energy funnel from HiRE-RNA-25 energy landscape and, for each one, we compute the different terms of the potential energy using both HiRE-RNA-25 force field and HiRE-RNA-47 force field. The comparison is shown in Fig. 3. Structures are organised horizontally from lowest total energy to highest. The first remark is that for HiRE-RNA-25 the energy spread between torsional, base-pairing and stacking energy is much larger than for HiRE-RNA-47. This confirms the overstabilization of base pairing and points to an excessive rigidity of torsions as well. In HiRE-RNA-47 base-paring and stacking energies have similar magnitudes and torsional energy is significantly reduced. A
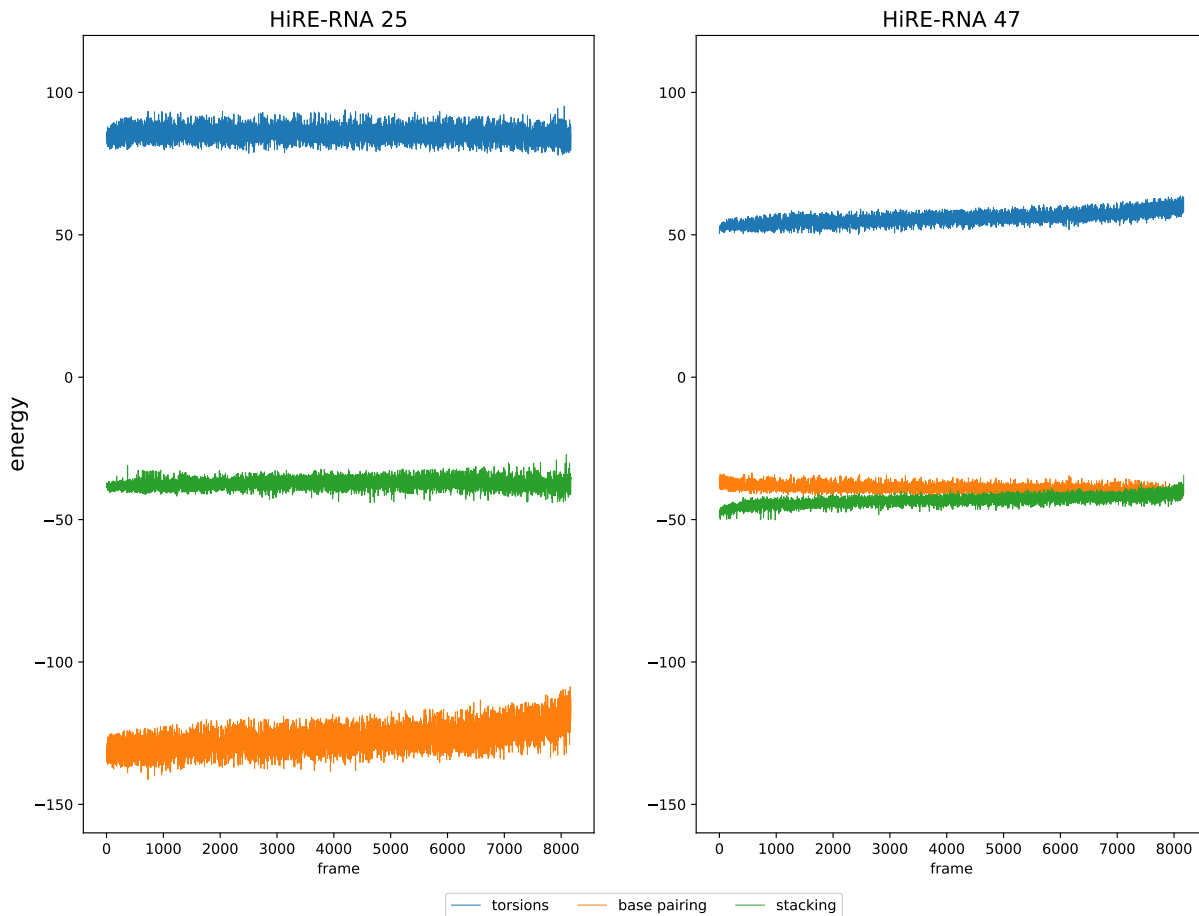
Figure 3: Torsional, base pairing and stacking energies computed on the structures of the lowest energy funnel obtained from HiRE-RNA 25 simulations, computed using HiRE-RNA 25 parameters set and HiRE-RNA 47 parameters set. Numbering of the structures in the funnel is done according to increasing total energy: frame 1, lowest energy, frame 8171 highest energy.

second observation concerns the fluctuations of each energy term across structures that are all very similar (less than 4Å RMSD). In HiRE-RNA-25 there is a large spread in base-pairing energy and, to a lesser extent, this is also the case for torsions. This variation is indicative of a poorly balanced term as one would expect the energies to change only minimally when structural changes do not involve the breaking of base pairing and significant rearrangements of the backbone. If we give fluctuations as percentage of the average value for each term, we compute 25% fluctuations for torsional energy for HiRE-RNA-47 to be compared to 7% for AMBER and to 65% for HiRE-RNA-25, 21% for base pairing energy for HiRE-RNA-47 to be compared to 25% HiRE-RNA-25 and with 14% of fluctuations in Van der Waals energy

for AMBER, and 22% for stacking energy for HiRE-RNA-47 vs. 27% HiRE-RNA-25. Even though values for HiRE-RNA-47 are still higher than those of AMBER, the modifications introduced to the model go in the right direction. A third observation is that for HiRE-RNA-47 the loss in energy between the structures in the funnel is mainly due to variations in stacking, as shown by the slight decrease in staking energy going from lowest-energy to highest-energy structure, while base-pairing remains roughly constant. This observation is consistent with having conformations with the same secondary structure. On the contrary, for HiRE-RNA-25 the loss in total energy is clearly due to a decrease in base pairing energy, despite the fact that there is no change in secondary structure, i.e. base pairs become looser but the bond does not break.

When we repeat the landscape exploration with HiRE-RNA-47, we observe the same trends in the minima around the native structure. Furthermore, we do not observe the chiral structures any longer, indicating a better balance in the torsions. Nonetheless, the energy landscape is still showing artefacts, in particular unfolded and partially folded structures are still far too high in energy. This result is not surprising, as no unfolded and partially folded targets where used in the reoptimisation procedure, but only native structures.

As previously mentioned, HiRE-RNA was developed to give an accurate description of folded molecules, so we did not expect it to match results from AMBER for high energies. The modification in the force field introduced from HiRE-RNA-25 to HIRE-RNA-47 were motivated by observations on the overall landscape, including high-energy structures. Nonetheless, the benefits of these modifications are also clearly noticeable in the low energy structures closely resembling the native fold. This improvement can be observed analysing the local behaviour of the ensemble of structures in the low energy funnels obtained with AMBER, with HIRE-RNA-25 and with HIRE-RNA-47, as shown in Fig. 4. For every structure in the funnel we compute the overall number of staking interactions, of canonical and of non-canonical base pairs. These values are computed using the program Barnaba[55] for atomistic structures, and using HiRE-RNA energies for CG structures. We can observe that
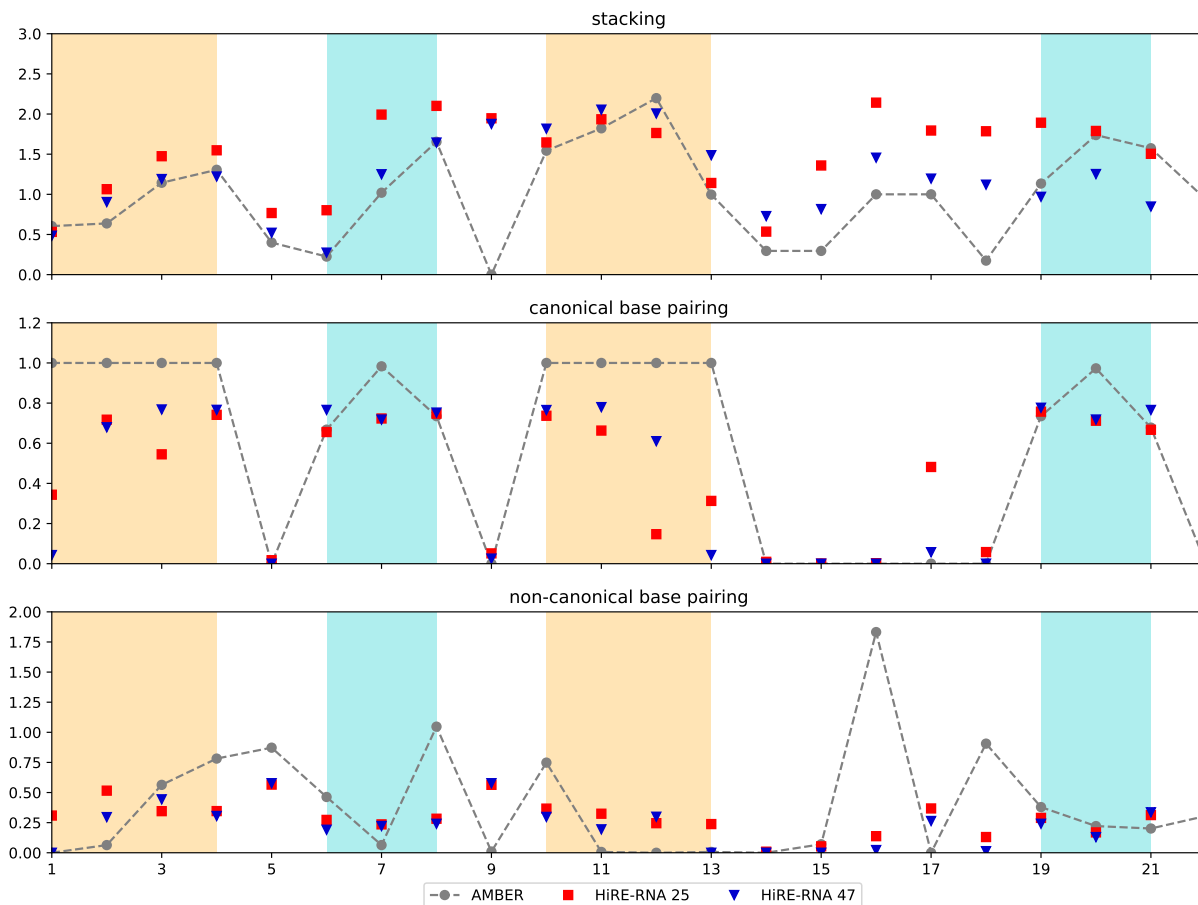
Figure 4: Comparison of overall staking, canonical base pairing and non-canonical base pairing computed for each nucleotide for AMBER, HiRE-RNA 25 and HiRE-RNA 47, for structures included in the lowest energy funnels for each simulation (folded structures). Shaded backgrounds indicate the regions forming the two stems in the native structure.

stacking of each nucleotide by HiRE-RNA-47 has a better correspondence to AMBER than HiRE-RNA-25, with similar values and similar trends. The one noticeable exception is for nucleotide 9, which is a base exposed to the solvent in the native state and in AMBER funnel, but that gets packed inside a loop in the CG models. Similarly, canonical base pairing of HiRE-RNA-47 corresponds better to that of AMBER with the two stems more consistently present than for HiRE-RNA-25. However, the most external base pair, between nucleotides 1 and 13 is less stable in the CG model than in the atomistic. Both HiRE-RNA models behave similarly with respect to non-canonical base paring.

The detailed comparison between the folded structures of the CG model and those of

14

the atomistic model shows the margin of improvement still required by the coarse-grained force field. While the CG model overstabilises base-pairing, this is however not specific enough to stabilise stems to the extent observed in the atomisic structures. Some spurious non-canonical interactions contributing to the destabilisation of canonical base pairs should be eliminated in a future refinement of the force field. It is important to notice that these observations are possible only when looking at ensembles of structures, where it is possible to extract statistically relevant behaviours.

# 4   Conclusions

In this study we have shown how an explicit exploration of energy landscapes across different scales of representations allows insights into how faithful structural features are preserved across all structural ensembles. Not only can we identify missing ensembles and ensembles with too much or too little stability, but we can further identify the features of the underlying model that lead to such behaviour. We demonstrated that it is possible to use this insight to improve coarse-grained representations, achieving significant improvements with modest, but well-motivated modifications.

The method presented is agnostic to the nature of the potentials and representations compared, and allows readily to compare and contrast computational models with each other and with experiment. As the energy landscapes allow detailed insight into structural features and transition mechanism, as well as thermodynamic and kinetic observables, the presented approach provides a useful tool in the development of reliable and faithful coarse-grained models.

As the approach considers all structural ensembles on the energy landscape, fine balancing between competing structures is possible, with the ability to assign changes in relative stability to the underlying model and potential energy functions used.

The availability of multiple structural ensembles representing folded, partially folded and

unfolded structures furthermore provides a future avenue to improve parameter optimisation. Ideally, we would like to optimise our model for folded, partially folded and unfolded structures, but often only folded structures are available from experiment and more traditional simulation methods. This limitation means that optimisation for example in a genetic algorithm stabilises the native fold compared to decoys. Given the access of all ensembles through the energy landscape, it is possible to optimise with respect to energy differences between states, without decoys, enabling better relative stabilities, and providing correspondence between energy scales of different models. This approach will also improve the energies of transition states, making kinetic information accessible through the coarse-grained scheme.

## Acknowledgement

## References

(1) Staple, D. W.; Butcher, S. E. Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* **2005**, *3*, e213.

(2) Ponting, C. P.; Oliver, P. L.; Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **2009**, *136*, 629–641.

(3) Siomi, M. C.; Sato, K.; Pezic, D.; Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 246–258.

(4) Cruz, J. A.; Westhof, E. The dynamic landscapes of RNA architecture. *Cell* **2009**, *136*, 604–609.

(5) Röder, K.; Wales, D. J. Evolved minimal frustration in multifunctional biomolecules. *J. Phys. Chem. B* **2018**, *14*, 10989–10995.

(6) Chakraborty, D.; Collepardo-Guevara, R.; Wales, D. J. Energy landscapes, folding mechanisms, and kinetics of RNA tetraloop hairpins. *J. Am. Chem. Soc.* **2014**, *136*, 18052–18061.

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

(8) Röder, K.; Stirnemann, G.; Dock-Bregeon, A.-C.; Wales, D. J.; Pasquali, S. Structural transitions in the RNA 7SK 5′ hairpin and their effect on HEXIM binding. *Nucleic Acids Res.* **2020**, *48*, 373–389.

(9) Thirumalai, D.; Klimov, D. K.; Woodson, S. A. Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theor. Chem. Acc.* **1997**, *96*, 14–22.

(10) Šponer, J.; Bussi, G.; Stadlbauer, P.; Kührová, P.; Banáš, P.; Islam, B.; Haider, S.; Neidle, S.; Otyepka, M. Folding of guanine quadruplex molecules–funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies. *BBA-Gen. Subjects* **2017**, *1861*, 1246–1263.

(11) Pasquali, S.; Derreumaux, P. HiRE-RNA: A high resolution coarse-grained energy model for RNA. *J. Phys. Chem. B* **2010**, *114*, 11957–11966.

(12) Cragnolini, T.; Laurin, Y.; Derreumaux, P.; Pasquali, S. Coarse-grained HiRE-RNA model for ab initio RNA folding beyond simple molecules, including noncanonical and multiple base pairings. *J. Chem. Theory Comput.* **2015**, *11*, 3510–3522.

(13) Denesyuk, N. A.; Thirumalai, D. Coarse-grained model for predicting RNA folding thermodynamics. *J. Phys. Chem. B* **2013**, *117*, 4901–4911.

(14) Šulc, P.; Romano, F.; Ouldridge, T. E.; Doye, J. P. K.; Louis, A. A. A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.* **2014**, *140*, 235102.

(15) Boniecki, M. J.; Lach, G.; Dawson, W. K.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K. M.; Bujnicki, J. M. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **2015**, *44*, e63.

(16) Bell, D. R.; Cheng, S. Y.; Salazar, H.; Ren, P. Capturing RNA folding free energy with coarse-grained Molecular Dynamics simulations. *Sci. Rep.* **2017**, *7*, 45812.

(17) Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyepka, M. RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chem. Rev.* **2018**, *118*, 4177–4338.

(18) Stadlbauer, P.; Mazzanti, L.; Cragnolini, T.; Wales, D. J.; Derreumaux, P.; Pasquali, S.; Šponer, J. Coarse-grained simulations complemented by atomistic molecular dynamics provide new insights into folding and unfolding of human telomeric G-quadruplexes. *J. Chem. Theory Comput.* **2016**, *12*, 6077–6097.

(19) Wagner, J. W.; Dama, J. F.; Durumeric, A. E. P.; Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys.* **2016**, *145*, 044108.

(20) Joseph, J. A.; Röder, K.; Chakraborty, D.; Mantell, R. G.; Wales, D. J. Exploring biomolecular energy landscapes. *Chem. Commun.* **2017**, *53*, 6974–6988.

(21) Röder, K.; Joseph, J. A.; Husic, B. E.; Wales, D. J. Energy landscapes for proteins: From single funnels to multifunctional systems. *Adv. Theory Simul.* **2019**, *2*, 1800175.

(22) Nonin-Lecomte, S.; Felden, B.; Dardel, F. NMR structure of the Aquifex aeolicus

tmRNA pseudoknot PK1: new insights into the recoding event of the ribosomal trans-translation . *Nucleic Acids Res.* **2006**, *34*, 1847–1853.

(23) Nameki, N.; Tadaki, T.; Himeno, H.; Muto, A. Three of four pseudoknots in tmRNA are interchangeable and are substitutable with single-stranded RNAs. *FEBS Lett.* **2000**, *470*, 345–349.

(24) Lescoute, A.; Leontis, N. B.; Massire, C.; Westhof, E. Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.* **2005**, *33*, 2395–2409.

(25) Kucharík, M.; Hofacker, I. L.; Stadler, P. F.; Qin, J. Pseudoknots in RNA folding landscapes. *Bioinformatics* **2016**, *32*, 187–194.

(26) Antczak, M.; Popenda, M.; Zok, T.; Zurkowski, M.; Adamiak, R. W.; Szachniuk, M. New algorithms to represent complex pseudoknotted RNA structures in dot-bracket notation. *Bioinformatics* **2018**, *34*, 1304–1312.

(27) Peselis, A.; Serganov, A. Structure and function of pseudoknots involved in gene expression control. *Wiley Interdiscip. Rev. RNA* **2014**, *5*, 803–822.

(28) Adams, P. L.; Stahley, M. R.; Kosek, A. B.; Wang, J.; Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **2004**, *430*, 45–50.

(29) Theimer, C. A.; Blois, C. A.; Feigon, J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* **2005**, *17*, 671–682.

(30) Shen, L. X.; Tinoco, I. J. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.* **1995**, *247*, 963–978.

(31) Michiels, P. J.; Versleijen, A. A.; Verlaan, P. W.; Pleij, C. W.; Hilbers, C. W.; Heus, H. A. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J. Mol. Biol.* **2001**, *310*, 1109–1123.

(32) Nixon, P. L.; Rangan, A.; Kim, Y.-G.; Rich, A.; Hoffman, D. W.; Hennig, M.; Giedroc, D. P. Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J. Mol. Biol.* **2002**, *322*, 621–633.

(33) Ke, A.; Zhou, K.; Ding, F.; Cate, J. H. D.; Doudna, J. A. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **2004**, *429*, 201–205.

(34) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(35) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of $\alpha/\gamma$ conformers. *Biophys. J.* **2007**, *92*, 3817–3829.

(36) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M. Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836–3849.

(37) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.

(38) Wales, D. J. Discrete path sampling. *Mol. Phys.* **2002**, *100*, 3285–3305.

(39) Wales, D. J. Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys.* **2004**, *102*, 891–908.

(40) Noé, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struc. Biol.* **2008**, *18*, 154–162.

(41) Wales, D. J. Energy landscapes: some new horizons. *Curr. Opin. Struc. Biol.* **2010**, *20*, 3–10.

(42) Li, Z.; Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6611–6615.

(43) Li, Z.; Scheraga, H. A. Structure and free-energy of complex thermodynamic systems. *J. Mol. Struct.* **1988**, *48*, 333–352.

(44) Wales, D. J.; Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Chem. Phys. A* **1997**, *101*, 5111–5116.

(45) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978–9985.

(46) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(47) Trygubenko, S. A.; Wales, D. J. A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.* **2004**, *120*, 2082–2094.

(48) Carr, J. M.; Wales, D. J. Global optimization and folding pathways of selected alpha-helical proteins. *J. Chem. Phys.* **2005**, *123*, 234901.

(49) Röder, K.; Wales, D. J. Predicting pathways between distant configurations for biomolecules. *J. Chem. Theory Comput.* **2018**, *14*, 4271–4278.

(50) Munro, L. J.; Wales, D. J. Defect migration in crystalline silicon. *Phys. Rev. B* **1999**, *59*, 3969–3980.

(51) Becker, O. M.; Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.* **1998**, *106*, 1495–1517.

(52) Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal energy landscapes. *Nature* **1998**, *394*, 758–760.

(53) Petrov, A.; Zirbel, C.; Leontis, N. Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA* **2013**, *19*, 1327–1340.

(54) Cho, S. S.; Pincus, D. L.; Thirumalai, D. Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17349–17354.

(55) Bottaro, S.; Bussi, G.; Pinamonti, G.; Reißer, S.; Boomsma, W.; Lindorff-Larsen, K. Barnaba: software for analysis of nucleic acid structures and trajectories. *RNA* **2019**, *25*, 219–231.

# Graphical TOC Entry

9 cm by 3.5 cm