Article

Energy landscapes, folding mechanisms and kinetics of RNA tetraloop hairpins

Debayan Chakraborty, Rosana Collepardo-Guevara, and David J. Wales

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/ja5100756 • Publication Date (Web): 02 Dec 2014

Energy landscapes, folding mechanisms and kinetics of RNA tetraloop hairpins

Debayan Chakraborty, Rosana Collepardo-Guevara, and David J. Wales*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

E-mail: dw34@cam.ac.uk

*To whom correspondence should be addressed

Abstract

RNA hairpins play a pivotal role in a diverse range of cellular functions, and are integral components of ribozymes, mRNA and riboswitches. However, the mechanistic and kinetic details of RNA hairpin folding, which are key determinants of most of its biological functions, are poorly understood. In this work, we use the discrete path sampling (DPS) approach to explore the energy landscapes of two RNA tetraloop hairpins, and provide insights into their folding mechanisms and kinetics in atomistic detail. Our results show that the potential energy landscapes have a distinct funnel-like bias towards the folded hairpin state, consistent with efficient structure-seeking properties. Mechanistic and kinetic information is analyzed in terms of kinetic transition networks. We find microsecond folding times, consistent with temperature jump experiments, for hairpin folding initiated from relatively compact unfolded states. This process is essentially driven by an initial collapse, followed by rapid zippering of the helix stem in the final phase. Much lower folding rates are predicted when the folding is initiated from extended chains, which undergo longer excursions on the energy landscape before nucleation events can occur. Our work therefore explains recent experiments and coarse-grained simulations, where the folding kinetics exhibit precisely this dependency on the initial conditions.

Introduction

RNA is an active player in biology, acting as an efficient information carrier, a regulator of gene expression, and a catalyst for chemical reactions.[1–4] As structure often dictates biological function, deciphering the details of how RNA adopts particular three-dimensional structures, under physiological conditions, is a key step towards improving our understanding of important cellular processes. RNA folding is hierarchically organized, with clear separation of timescales. Formation of secondary structure elements typically occurs within microseconds, whereas initiation of tertiary contacts is orders of magnitude slower.[5–7] The interplay of competing base-pairing, base-stacking and electrostatic interactions tends to make the RNA energy landscapes topologically frustrated,[7] resulting in a 'kinetic partitioning' of the unfolded state population into slow and fast folders. The

multitude of timescales, arising due to the complexity of the underlying landscape[8–11] has so far precluded a complete understanding of the folding mechanisms and kinetics, despite the upsurge of interest in this direction.[12,13] Recent experiments[14–16] and computational studies[17–19] indicate that even formation of simple secondary structure elements, such as hairpins, is characterized by complex folding kinetics, with the possibility of kinetic trapping in misfolded conformations. These 'rare events' are therefore ideal targets for our computational energy landscapes approach, based on geometry optimization, which have proved to be both efficient and accurate for landscapes featuring broken ergodicity.[20–22] In the present contribution, we investigate the mechanism and kinetics of hairpin formation, as a first step towards decoding the microscopic details and generic features of the RNA folding problem.

Within the RNA subfamily, tetraloop hairpins have attracted the most interest because of their natural abundance, as well as their diverse range of biological functions.[1,23] Tetraloops serve as templates in the building of higher order RNA structures,[2,24,25] and are known to mediate RNA-RNA as well as RNA-protein interactions.[26–30] RNA tetraloops consist of a helical stem with Watson-Crick base pairs, capped by a loop comprising four nucleotides.[23,31] The UNCG (N=any nucleotide) and GNRA (N=any nucleotide; R=purine) loop sequences are the most common within the tetraloop family, and these hairpins have exceptional thermodynamic stability due to the presence of several non-canonical interactions.[32] Because of their structural simplicity and size, tetraloop hairpins often serve as prototypes for understanding the folding of larger RNA molecules.

In recent years, there has been much debate about the folding kinetics of nucleic acid hairpins, and no general consensus has been reached so far. The earliest experiments[33,34] suggested that folding can be described by a classical two-state kinetic model.[35] The emergence of ultrafast techniques, which have enabled probes of hairpin folding at very fine temporal resolution, has given new impetus to this debate. In particular, laser induced temperature jump[14,16,36–38] and fluorescence correlation spectroscopy experiments[15,39,40] have indicated significant deviations from the two-state hypothesis. These experiments argue in favor of a more complex multi-state kinetic

model, with the existence of several intermediates along the folding pathway. Recently, temperature jump in combination with microfluidic mixing techniques have revealed that the relaxation kinetics are strongly dependent on the region of the landscape from which folding is initiated.[41]

Computer simulations have complemented these experimental findings, by providing microscopic insights into the folding mechanisms.[42] From their massive simulation effort using distributed computing, Pande and coworkers have suggested that tetraloops can fold via multiple routes, involving predominantly zipping or compaction mechanisms.[43,44] Subsequent work[18,19,45–50] has corroborated their findings, and has revealed the presence of intermediates and misfolded traps along the pathways.

Even for small nucleic acid hairpins, which exhibit a folding timescale of microseconds, unbiased simulations have generally not been successful in reaching the folded hairpin from the unfolded state. In order to bridge the gap between experimental and simulation timescales, as well as overcome sampling limitations imposed by the complexities of the underlying energy landscapes, most of the all-atom computational studies on RNA tetraloops have employed enhanced sampling techniques, such as replica exchange molecular dynamics (REMD),[18,19,43–47,51] or umbrella sampling[52] along predefined coordinates, as well as high temperature unfolding simulations.[53] However, most enhanced sampling techniques lead to loss of temporal resolution, making it difficult to obtain kinetic information. Insights into the experimental folding kinetics have generally come from Go-like[54] and coarse-grained potentials,[17,55,56] as well as statistical mechanical models.[36,57–62] Kinetic Monte Carlo simulations,[63,64] based on various simplified potentials, have also given estimates of the folding times.

In the present contribution, we address the hairpin folding problem from both a mechanistic and kinetic standpoint. The discrete path sampling (DPS) technique[65,66] is used to map out the underlying energy landscapes, and the folding mechanisms and kinetics are analyzed in terms of kinetic transition networks.[67–69] The DPS method exploits geometry optimization based techniques and provides a complementary approach to molecular dynamics and Monte Carlo simulations. It can potentially probe rare events occurring over any range of timescales, and is largely unaffected by

the presence of kinetic bottlenecks. The RNA tetraloop hairpins considered in the present work were modelled using an all-atom force-field in conjunction with an implicit solvent model. To the best of our knowledge, there is no previous study at the all-atom level that exploits this energy landscape framework,[20] to give a unified view of RNA hairpin folding in both kinetic and mechanistic terms.

Our results provide unprecedented insights into the hairpin folding mechanisms and kinetics, and explain the disparate timescales reported by different experiments.[15,41,70] We find that the free energy landscapes exhibit a distinct folding funnel, and we predict multiple kinetically convergent pathways to the native basin along with the associated time scales. Interestingly, the rates and mechanisms are found to be strongly dependent on the nature of the denatured ensemble from which folding is initiated, in agreement with previous coarse-grained simulations.[17,55,56]

Methods

In the present work we consider both a UUCG (UNCG type) and a GCAA (GNRA type) tetraloop. Each RNA hairpin is ten nucleotides long. The initial coordinates for the UUCG tetraloop hairpin were taken from a previously reported X-ray structure[71] (PDB code 1F7Y, residues 6-15). The coordinates for the GCAA tetraloop hairpin were taken from a high resolution NMR structure[72] (PDB code 1ZIH). We have made the stem compositions of the two tetraloops identical by removing the terminal G-U base pair from the original NMR structure, and mutating the G-C base pair in the middle of the stem to a C-G base pair. The base mutations were accomplished using the 3DNA software.[73]

The bases in the loop region of the UUCG hairpin are designated U_{L1}, U_{L2}, C_{L3} and G_{L4}, respectively. The bases in the loop region for the GCAA hairpin are designated G_{L1}, C_{L2}, A_{L3} and A_{L4}, respectively. The base pairs in the stem for both the sequences are designated $G_{S1}-C_{S1}$, $G_{S2}-C_{S2}$ and $C_{S3}-C_{S3}$.

The hairpins were modelled using a correctly symmetrized[74] version of the all-atom AM-

BER99/bsc0[75] force-field, employing the latest torsional corrections.[76] The solvent effects were included implicitly using a generalized Born model.[77,78] The salt concentration was maintained at 0.1 M using the Debye-Hückel approximation.[79]

Molecular Dynamics

The molecular dynamics simulations were carried out using the AMBER12 code[80] with no cutoff for the nonbonded interactions. Temperature control was maintained using a Langevin thermostat[81] employing a collision frequency of $1\,\mathrm{ps}^{-1}$.

Discrete Path Sampling

The Discrete Path Sampling (DPS)[65,66] method was employed to explore the energy landscapes of the RNA hairpins, and characterize folding mechanisms. The DPS procedure provides a convenient framework to construct kinetic transition networks from stationary point databases, which can be used to analyze the global thermodynamics and kinetics. It has been previously used to study the energy landscapes of proteins and peptides[82,83] as well as atomic and molecular clusters.[84] In this section we will simply highlight the key steps of DPS, as the theory has been presented in detail elsewhere.[65,66]

A discrete path on the potential energy surface (PES) consists of a sequence of minima connected by intervening transition states. A stationary point for which all the non-zero normal mode frequencies are positive is a local minimum, whereas a transition state has one imaginary frequency corresponding to the reaction coordinate.[85] Approximate steepest-descent paths directed parallel and anti-parallel to the reaction coordinate terminate at the adjoining minima, describing the connectivity of a given transition state. The number of steps in a discrete path corresponds to the number of transition states in the connected chain. A modified version of the LBFGS algorithm described by Liu and Nocedal[86] was used for local minimization. The doubly-nudged[87] elastic band method[88,89] was employed to find initial guesses for transition state structures between pairs of local minima. These transition state candidates were further refined using the hybrid eigenvector-

following method.[88,90] Geometry optimizations were deemed to be converged when the root mean square gradient fell below 10^{-6} kcal/mol. The geometry optimizations and the transition state searches were carried out using our OPTIM code[91] via an AMBER9[92] interface.

After an initial discrete path was found between the endpoints of interest, the stationary point databases were further expanded using several database refinement schemes. The SHORTCUT procedure[93] chooses pairs of local minima that are closest together in configuration space, but are separated by a minimum number of steps on the discrete path. Another scheme, SHORTCUT BARRIER,[93] selects pairs of minima on either side of, and an equal number of steps away from, the largest potential or free energy barriers. The latter scheme is efficient in finding alternative paths that circumvent high energy barriers, whilst the SHORTCUT method provides discrete paths with fewer steps. There are several refinement schemes that remove artificial frustration from stationary point databases, in order to make them more faithful representatives of the global kinetics. Here we used UNTRAP,[93] which is based on the ratio of the potential energy barrier to the potential energy difference between pairs of local minima. The final stationary point databases were refined using sequential applications of the three schemes using the PATHSAMPLE[94] program.

The rate constant k_{BA}^{SS} for the A $\rightarrow$ B transition when intervening minima are treated in steady-state, can be written as a sum over all discrete paths, if the dynamics between adjacent local minima is assumed to be Markovian.[65,66] The sum over discrete paths is weighted by the occupation probability of the reactant minimum, as well as the relevant branching probabilities. The discrete paths that make the largest contribution to k_{BA}^{SS} are termed 'fastest paths'. Dijkstra's shortest path algorithm[95] was used to extract the fastest paths from the database between end points of interest using an appropriate edge weight.[65,66] This approach provides a convenient way of visualizing pathways in atomistic detail, and drawing mechanistic insights. We have also considered the possibility of alternate pathways, and used the recursive enumeration algorithm[96,97] to obtain the set of discrete paths that make the most significant contributions to the overall rate constant.

Calculation of Free Energies and Rate Constants

The database of minima obtained from DPS is used to estimate free energies using the harmonic superposition approximation.[98] The molecular partition function is represented as a sum of contributions from different catchment basins (of each minimum).[99,100] The vibrational density of states for each minimum is calculated using the normal mode frequencies.[82,83,93]

The phenomenological rate constants for the folding process were calculated from the databases using the new graph transformation (NGT) approach, as described elsewhere.[69] This method has been implemented within the PATHSAMPLE code. All the rate constants reported in this work were estimated at 298 K, and have been computed using a self-consistent regrouping scheme, which lumps structures separated by free energy barriers below a certain threshold into one macrostate.[83] These approaches are attractive as they exploit the separation of time-scales between the overall folding process, and local equilibration in the reactant and product regions. As a consequence of regrouping, the original end-points representing the reactant and product states are expanded into ensembles of structures in local equilibrium, making direct comparison with experimentally reported time-scales feasible.[101]

Visualization of the Energy Landscapes: Disconnectivity Graphs

Disconnectivity graphs[102–106] were constructed to visualize the potential energy and the free energy landscapes of the RNA hairpins. A disconnectivity graph partitions the energy landscape into disjoint sets of minima known as superbasins.[102] Local minima within each superbasin are mutually accessible via transition states lying below a chosen energy threshold, whereas interbasin transitions involve higher potential or free energy barriers.

Results and Discussion

The native and the unfolded ensembles

To assess the thermodynamic stability of the native hairpins, 200 ns molecular dynamics simulations employing the AMBER all-atom potential were first carried out at 298 K starting from the experimental hairpin structures. For each tetraloop sequence, configurations were saved every 10 ps, and were subsequently locally minimized.

The lowest potential energy minimum (Figure 1) for the UUCG sequence sampled along the MD trajectory is a hairpin structure with a fully formed helix stem. The folded hairpin retains nearly all of the interactions in the loop region of the X-ray structure. The hydrogen-bonding network between the G_{L4} and U_{L1} bases, which results in the formation of a non-canonical G-U *trans*-wobble interaction, is correctly reproduced. The U_{L2} base is flipped out, and exposed to the solvent, consistent with the X-ray structure. The hydrogen-bond between the amino group of C_{L3} and the 2OP1 of U_{L1}, which lends additional stability to the tetraloop, is also retained. We find that, unlike the X-ray structure, C_{L3} is slightly rotated out of plane in the direction of the tetraloop backbone, and does not stack as effectively on top of U_{L1}. Similar loss of stacking due to rotation of C_{L3} is also observed in the NMR structure reported by Allain and Varani.[107]

For the GCAA sequence, the lowest potential energy minimum (Figure 1), is a folded hairpin, with all the Watson-Crick base pairs in the stem intact. However, the topology of the loop region is quite different from the initial energy minimized NMR structure. The G_{L1}-A_{L4} sheared base pair that is observed in the experimental structure is absent, and the extensive base stacking and hydrogen-bonding interactions that stabilize the tetraloop geometry are partially lost. The A_{L3} and A_{L4} bases maintain their stacking interaction, but are flipped out into the solvent, while C_{L2} is looped inside, towards G_{L1}. Conformational clusters having similar structural features have also been reported by De Paul *et al.*[108] from their massively parallel simulation of the GCAA tetraloop. A systematic investigation of the MD trajectory, and structural clustering of the quenched snapshots, reveals that the tetraloop region can adopt several alternate stacking patterns (supporting

information, Figure 1). In most of these structures, the G_{L1} base maintains its stacking interaction with the first base pair in the stem, and is relatively static along the trajectory. The loop dynamics agree with previous simulations [108,109] and ultrafast experiments, [110] which suggest that the GNRA tetraloop family is inherently flexible. Several hairpin structures were also located along the trajectory, within 1.2 kcal/mol of the lowest minimum, which retain the main interactions found in the NMR structure. The sheared G_{L1}-A_{L4} base pair is retained in this ensemble of structures, and C_{L2}-A_{L3}-A_{L4} bases are stacked on the 3′ side of the helix stem. The hydrogen-bond between N7 of A_{L3} and 2′OH of G_{L1} is also correctly reproduced. The lowest energy structure from this ensemble is shown in Figure 1. The marginal destabilization of the near native GCAA hairpin structures, compared to the global minimum, may be due to the lack of explicit water-mediated interactions in the loop region. We have tested this possibility by carrying out 10 ns long molecular dynamics simulations in explicit water starting from the near-native GCAA hairpin structure. We find several water molecules around the loop region that exhibit fairly long residence times. In agreement with previous work, [111] a water molecule within the loop region is found to act as a bridge between the sheared G_{L1}-A_{L4} base pair (supporting information, Figure 2), lending additional stability to the hydrogen-bonding network. Since all these structures are lumped together in the analysis of rates and pathways, these effects do not change our principal conclusions.

A key requirement in the construction of stationary point databases is the characterization of suitable endpoints as reactant and product states. For the UUCG sequence, the lowest energy structure was initially taken as the sole representative of the folded ensemble (the product state). As the GCAA tetraloop was found to be more dynamic, we included both the lowest potential energy minimum as well as the lowest energy minimum structurally closest (in terms of RMSD) to the native hairpin, as initial members of the folded group.

To explore the denatured ensemble, multiple high temperature MD simulations (in the range of 500 to 1000 K), starting from the native hairpin structures, were carried out. Histograms of the radius of gyration R_g, for the conformations sampled along the trajectories (supporting information, Figures 3 and 4), reveals that even at temperatures as high as 1000 K, extended chain-like confor-

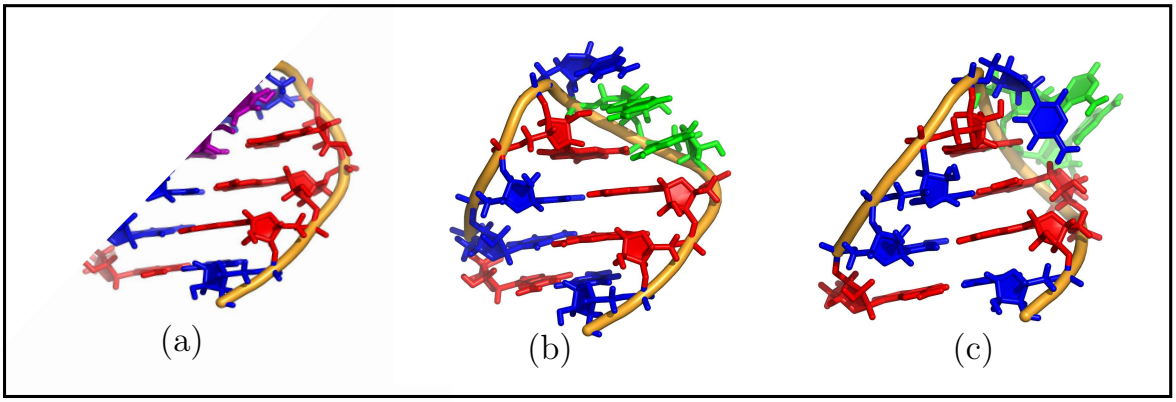

Figure 1. The lowest potential energy minima that are obtained after quenching snapshots sampled along the molecular dynamics trajectories at 298 K. (a) The UUCG hairpin. (b) The GCAA hairpin that is structurally closest to the NMR structure. (c) The lowest potential energy minimum for the GCAA hairpin located along the MD trajectory. The structures were rendered using the Pymol program.[112]

mations are relatively unpopulated compared to more compact, denatured states. This observation is consistent with previous studies,[17] which indicate that the thermally denatured ensemble essentially comprises compact structures, whereas completely stretched conformations are more likely to be populated under force-induced melting conditions.

As highlighted in the introduction, the seemingly contradictory experimental results reported on nucleic acid folding kinetics are probably due to the disparate initial conditions for folding accessed in different experiments.[36] In temperature jump experiments,[14,36] unfolding is initiated by finite jumps in the temperature space, and the regions of the energy landscape explored in these experiments are likely to be mirrored in the canonical distributions from our high temperature MD simulations. The relatively compact unfolded states ($R_g \sim 1.4$ nm) located near the mode of the R_g distribution should be similar to the initial population of unfolded states accessible by temperature jump experiments.[14,36] In contrast, in ion jump experiments, which can effectively probe the top of the folding funnel,[36,113] folding is initiated from highly extended chain-like conformations by rapidly switching to a high counter-ion concentration. These initial conditions correspond to the highly extended conformations ($R_g \sim 1.7$ nm) located near the tail of the R_g distribution. The fold-

ing pathways were explored from these distinct starting points (reactant states). An MD snapshot for which the R_g value was closest to the mode of the distribution, was quenched and chosen as the initial representative of the compact unfolded ensemble. For the starting point corresponding to the extended state, the snapshot exhibiting the largest R_g value along the trajectory was considered.

The database refinement process (see Methods section), was seeded from the initial set of discrete paths found between the different end points. DPS runs were deemed to have converged when the estimated folding rate constants were stable to within an order of magnitude, with respect to adding new stationary points. The final stationary point database included 73708 minima and 84192 transition states for the UUCG hairpin, and 80214 minima and 86599 transition states for the GCAA hairpin.

In the following sections, we analyze topological features of the underlying free energy landscapes, along with the folding mechanisms and kinetics of hairpin formation.

Energy Landscapes of the RNA tetraloops

The free energy landscapes for the UUCG and the GCAA tetraloop hairpins computed at 298 K are depicted in the form of disconnectivity graphs in Figure 2 and Figure 3, respectively. The corresponding potential energy disconnectivity graphs are included in the supporting information (Figures 5 and 6). The free energy graphs were produced by lumping together minima that are separated by free energy barriers less than 4 kcal/mol, employing the self-consistent regrouping scheme.[83] These graphs are qualitatively similar to those generated from the ungrouped databases (data not shown), implying that the regrouping thresholds have been chosen appropriately. A diverse range of conformations appear on the free energy landscapes; some representative snapshots are depicted on the disconnectivity graphs.

The graphs are colored according to the number of native Watson-Crick base pairs. For identifying native contacts within the stem, the criterion discussed in a recent contribution[114] was employed. The energy landscapes have distinct 'palm tree' like features,[20,106] with a prominent funnel-like bias towards the native state. The free energy global minimum is kinetically accessible

from the different basins, as expected for a good structure seeking system.[20]

The selected structural order parameter segregates the different sets of conformations quite well, but the mixing of colors in different regions of the disconnectivity graphs suggest that structural criteria alone are not sufficient to faithfully represent the complexities of the underlying landscape. The use of such order parameters often leads to artificially high folding rates, as minima classified as 'folded' and 'unfolded' may be separated by relatively low free energy barriers.[115] We stress that the use of 'native contacts' in our analysis is merely for visualization purposes, and the construction of the free energy disconnectivity graphs, as well as the estimation of folding rates, is based solely on kinetic criteria employing the recursive regrouping scheme.[83]

Folded hairpins with a complete helix stem are situated at the bottom of the disconnectivity graph, and constitute the native funnel. The free energy global minimum for the UUCG hairpin comprises a homogeneous cluster of hairpins, which have the same global fold and loop topology, and exhibit only minor variations in some of the internal degrees of freedom. In contrast, the lowest free energy ensemble corresponding to the GCAA hairpin exhibits some degree of structural heterogeneity. The corresponding hairpins retain all the native contacts in the stem, but exhibit a diverse range of stacking patterns near the loop region. These structures rapidly interconvert among themselves, highlighting the dynamical nature of the GCAA tetraloop motif, an observation consistent with our initial MD simulations.

Near-native structures, such as frayed hairpins, which have sufficiently well formed loops but lack some of the native contacts in the stem, are also low in free energy, and are separated from the native state by relatively small barriers. Within the ensembles of frayed hairpin structures some additional diversity is observed, which is mainly due to the different native contacts retained in the stem. Frayed hairpin-like states have been identified as putative intermediates along the folding pathways in previous work.[48] Several kinetic models[14,37] have also suggested the emergence of such 'prezipped' states in the final phases of the folding process, and have attributed the fast relaxation phase observed in temperature jump experiments to the rapid dynamics between frayed and native structures. Our mechanistic analysis of the folding process (discussed below) suggests

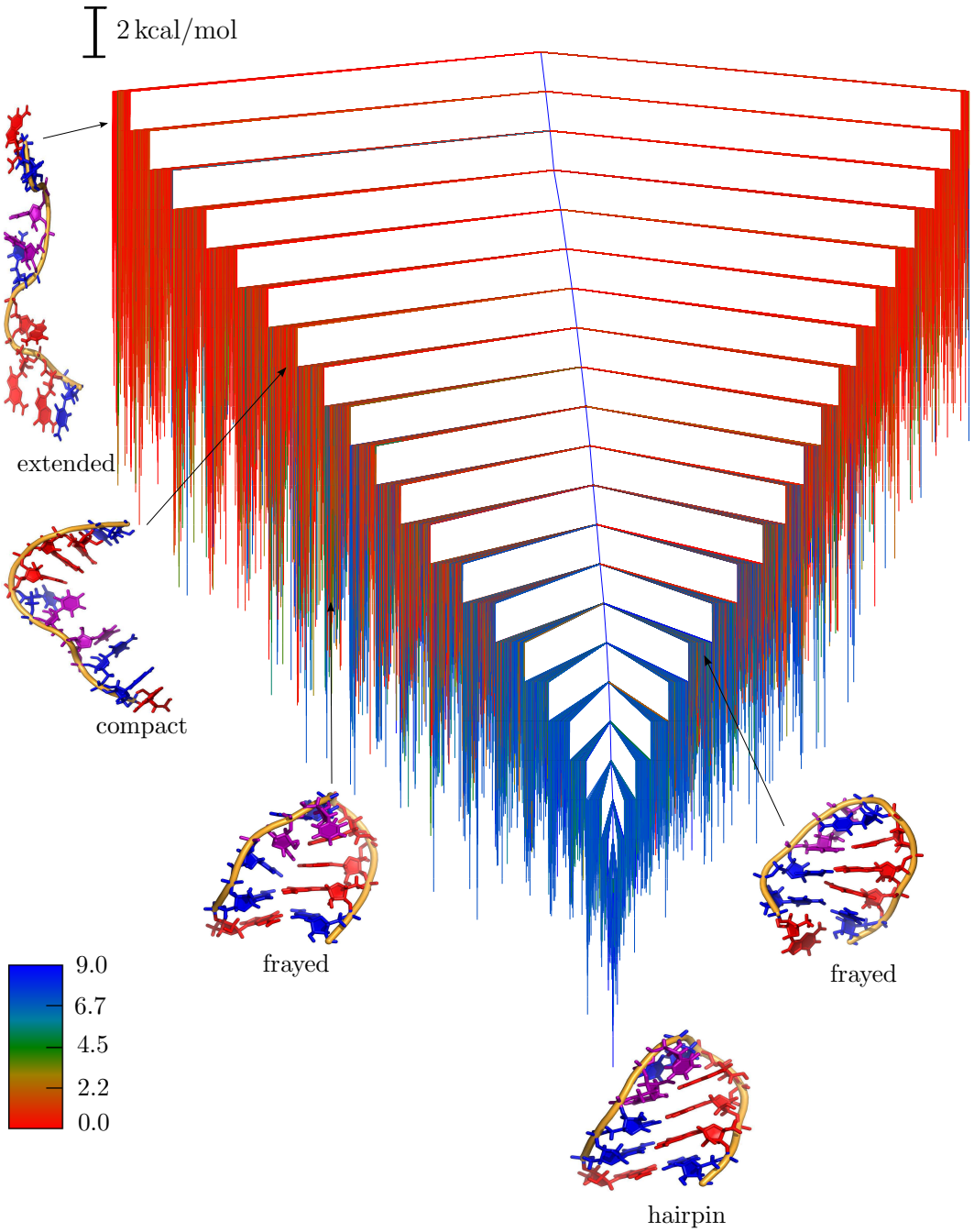

Figure 2. Free energy disconnectivity graph obtained for a regrouping threshold[83] of 4 kcal/mol at 298 K for the UUCG hairpin. The branches are colored according to the number of Watson-Crick base pairs in the helix stem (from blue to red; with blue representating structures having all the native contacts, and red representating structures without any native contacts). Some representative structures from the different conformational ensembles (frayed, compact unfolded/coil like, extended and native) that are found on the free energy landscape are also shown. Two distinct types of frayed hairpin states are depicted, one having a broken terminal base pair, and the other one with a broken closing base pair.

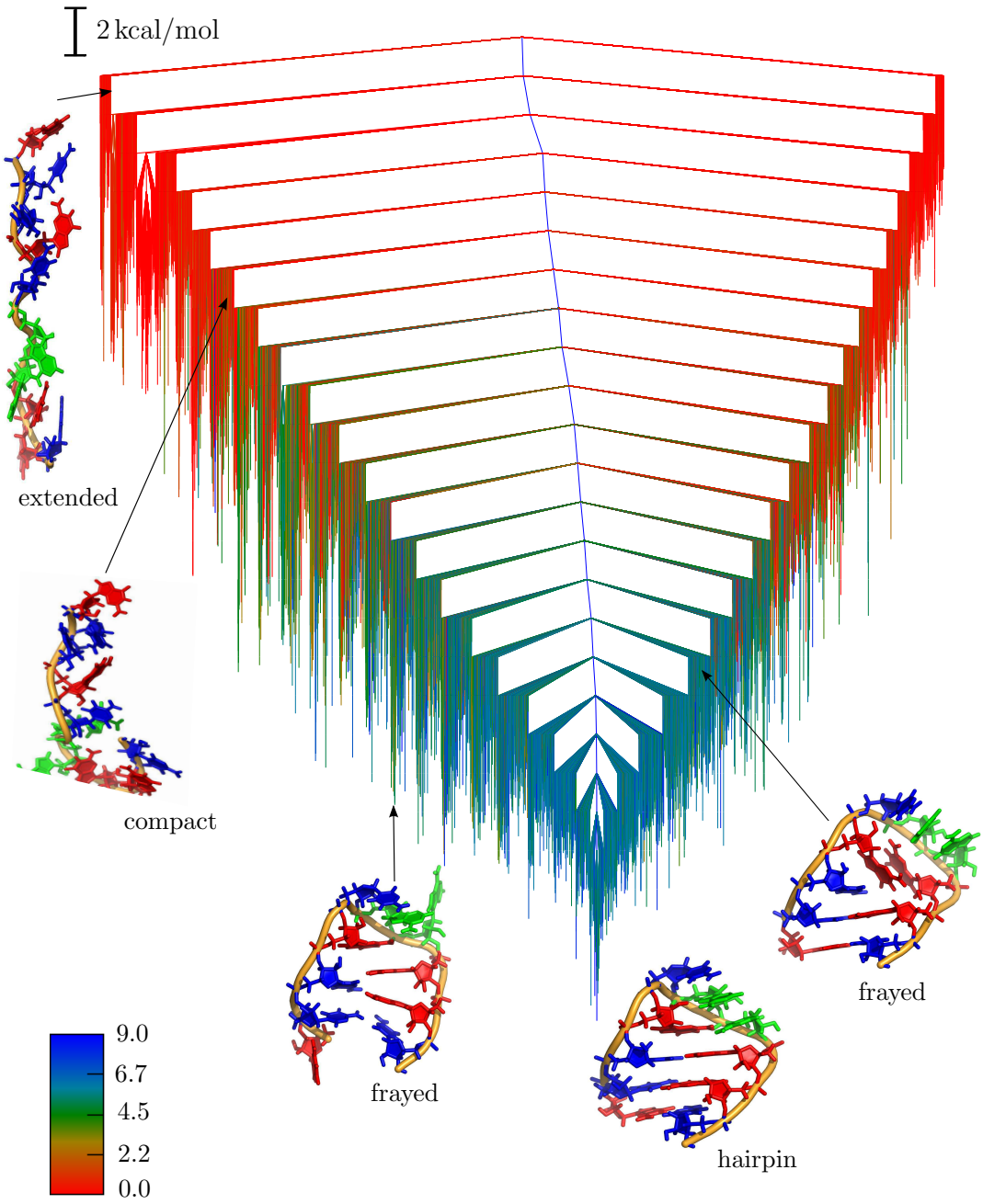

Figure 3. The free energy disconnectivity graph obtained for a regrouping threshold[83] of 4 kcal/mol at 298 K for the GCAA hairpin. The coloring scheme is same as in Figure 2. Some representative structures from the different conformational ensembles (frayed, compact unfolded/coil like, extended and native) that are found on the free energy landscape are also shown. Two distinct types of frayed hairpin states are depicted, one having a broken terminal base pair, and the other one with a broken closing base pair.

a similar scenario where the formation of the first correct base pair (resulting in the frayed hairpin) in the final phase of folding, drives the subsequent fast zipping of the rest of the helix stem.

Relatively compact random coil-like structures and denatured states are located in the intermediate regions of the disconnectivity graphs. These structures constitute clusters that are very heterogeneous in terms of compactness and deviation from the native state. They lack any native contacts, but some of them have residual structure near the loop regions. They are in different stages of collapse, and in some of the structures the two opposing strands are approximately aligned, which can facilitate folding. Extended conformations have very low equilibrium occupation probability at 298 K compared to the rest of the ensembles. They are located in the high energy regions of the disconnectivity graphs. These structures represent major perturbations from the native state and constitute clusters that are quite homogeneous, with large end-to-end distances between the strand terminals, and they differ mainly in the number of base stacking interactions.

The organization of different structural ensembles relative to the native state, along with the associated barrier heights for their interconversion, provide insight into the chain of events that occurs during hairpin folding. The compact denatured states are likely to fold faster, with the emergence of frayed hairpin states along the folding pathways. In contrast, extended conformations are likely to be associated with much longer refolding times, due to the significant structural rearrangements that are necessary to reach the native state.

Folding Mechanism and Kinetics

To draw mechanistic insights into the folding process from the compact denatured as well as the extended states, Dijkstra's shortest path algorithm[95] was used to extract pathways from the stationary point databases that make the largest contributions to the rate constant. The pathways from the compact denatured and the extended states to the native UUCG hairpin are shown in Figure 4. The pathways leading to the GCAA hairpin from the corresponding denatured state are illustrated in Figure 5. Some selected snapshots along the folding pathways are superimposed on the energy profiles. In general we find that, for both the hairpins, folding from the compact unfolded states is dominated by an initial collapse, which brings the two opposing strands into close contact. This collapse is evident from the rapid decrease in the radius of gyration (R_g) during the initial phase

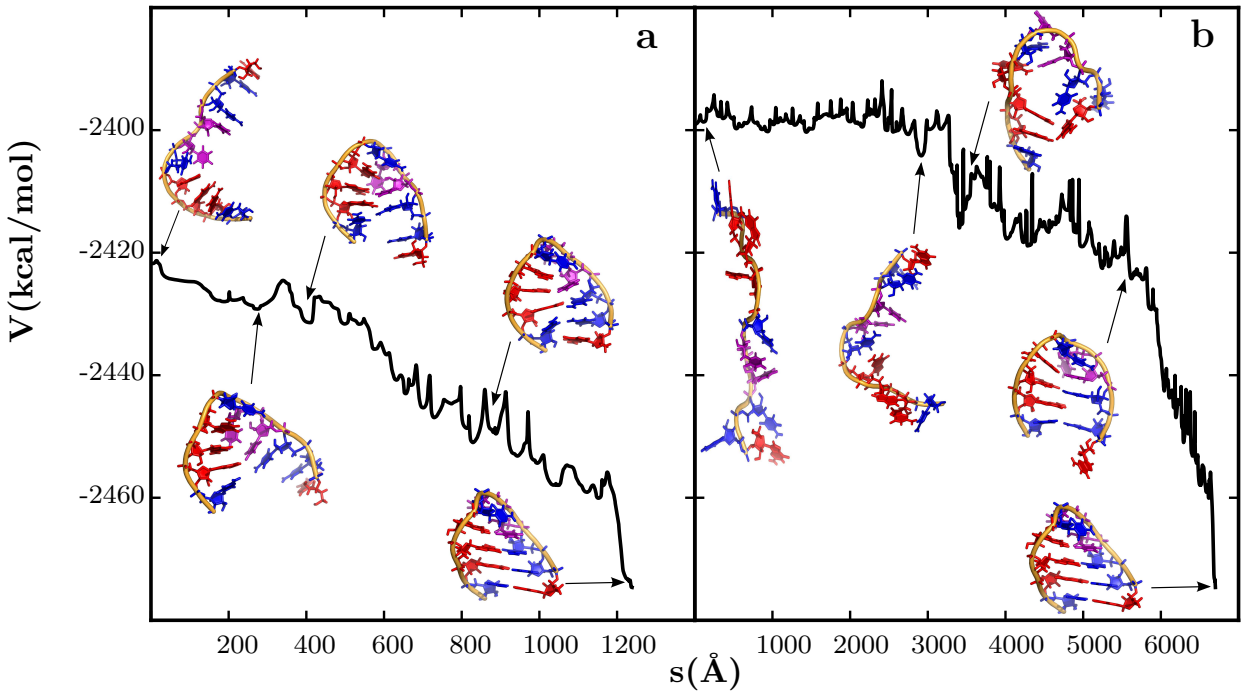

Figure 4. The total potential energy as a function of the integrated path length for the folding pathway that makes the largest contribution to the rate constant for (a) compact unfolded state to the native UUCG hairpin, and (b) extended state to the native UUCG hairpin. Snapshots of some representative structures that are encountered at different stages of the folding process are also shown.

of folding (supporting information, Figure 7). After the rapid collapse various contacts are made within the stem region, most of which are non-native. For the UUCG hairpin, non-native contacts are formed between the U_{L1} and G_{S1} nucleobases, which results in the concomitant formation of non-native base pairs in the rest of the stem. Interestingly, a similar trend is observed for the GCAA sequence, where a base from the loop region (G_{L1}) forms a non-native contact with a base from the stem (G_{S1}), resulting in other non-native contacts within the stem. These non-native contacts persist until the final phases of folding. The disruption of non-native contacts along the pathways is generally associated with energy barriers in the range of 7 to 8 kcal/mol, which appear as local maxima in the energy profiles. The loop regions of the hairpins also attain more optimal packing after the initial collapse phase. The bases in the loop region of the UUCG hairpin adopt an arrangement that is commensurate with the native loop topology, aside from the missing wobble base pair interaction between G_{L4} and U_{L1}. In contrast, in the GCAA hairpin, the loop remains

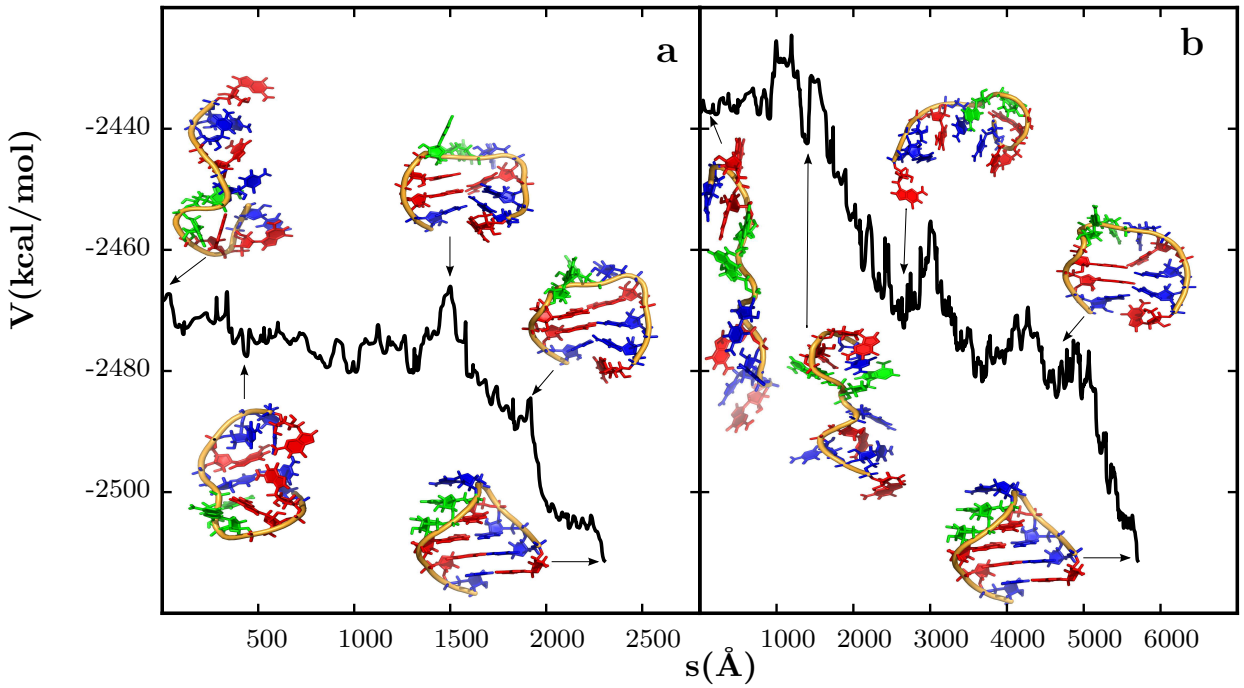

Figure 5. The total potential energy as a function of the integrated path length for the folding pathway that makes the largest contribution to the rate constant for (a) compact unfolded state to the native GCAA hairpin, and (b) extended state to the native GCAA hairpin. Snapshots of some representative structures that are encountered at different stages of the folding process are also shown.

slightly more disordered following initial collapse and thereafter. Although the C_{L2} and A_{L3} bases are engaged in stacking interactions, A_{L4} is completely flipped out of the loop region. It is only in the final phases of the folding that A_{L4} enters this region to form the $C_{L2} - A_{L3} - A_{L4}$ stack on the $3'$ side of stem, and complete the hydrogen-bonding network. The final phase of folding is downhill. During this phase, the formation of the first correct base pair within the stem region (resulting in frayed hairpin-like states), drives the zipping of the rest of the helix in a cooperative fashion. The remaining interactions within the tetraloop are also established concurrently.

Unlike the compact state, folding from the extended conformations requires navigating through a much wider region of the configurational space en route to the native basin, and is characterized by longer pathways. The chain compaction happens in three fairly distinct stages (supporting information, Figure 8). In the initial phase, compaction does not lead to successful contacts (either native or non-native) in the stem region, as the nucleobases are not in the preferred orientations

for engaging in base pairing interactions. Multiple chain collapse and expansion events, including significant structural reorganization within the loop and stem region, occurs during the next phases, resulting in a collapsed state with the nucleobases in favorable orientations for making contacts. Most of the contacts established in the stem are non-native, and these are subsequently broken. The final phase of folding is similar to the pathways from the compact states, where the loop acquires more optimal packing leading to the native tetraloop topology, and the correct contacts are established within the stem region. The emergence of compact states along the folding pathways, supporting non-native stem contacts as well as partial loop structure, seems consistent with the configurational diffusion model proposed by Ansari and coworkers.[36]

To find out if other possible routes exist between the unfolded and folded states, the recursive enumeration analysis[96] method was employed to extract the 1000 discrete paths between the endpoints of interest that make the largest contributions to the phenomenological rate costant. For the UUCG hairpin no significant mechanistic deviations are observed. For the GCAA hairpin, there is some diversity in the pathways from the compact unfolded state, which can be primarily attributed to the dynamics in the loop region. It is possible that pathways exhibiting significant mechanistic differences from those described above also exist in our databases. The presence of multiple frayed hairpin conformations also suggests that pathways proceeding via pure compaction or zipping mechanisms[43,48] cannot be ruled out, but they probably do not feature among the kinetically relevant set of pathways in our databases. In Figure 6 a potential energy disconnectivity graph is depicted using only the stationary points involved in the 1000 fastest discrete paths from the compact unfolded and the extended states. For both the hairpin sequences, the minima that are common to all the pathways are present only in the low energy sections of the disconnectivity graphs, indicating that the folding mechanism from the different endpoints is similar in the final phase.

The rate constants for the folding process were computed using the NGT method at a temperature of 298 K, employing the self-consistent regrouping scheme.[83] The estimated rate constants take into account all possible paths that exist between two reactant and product states. The group-

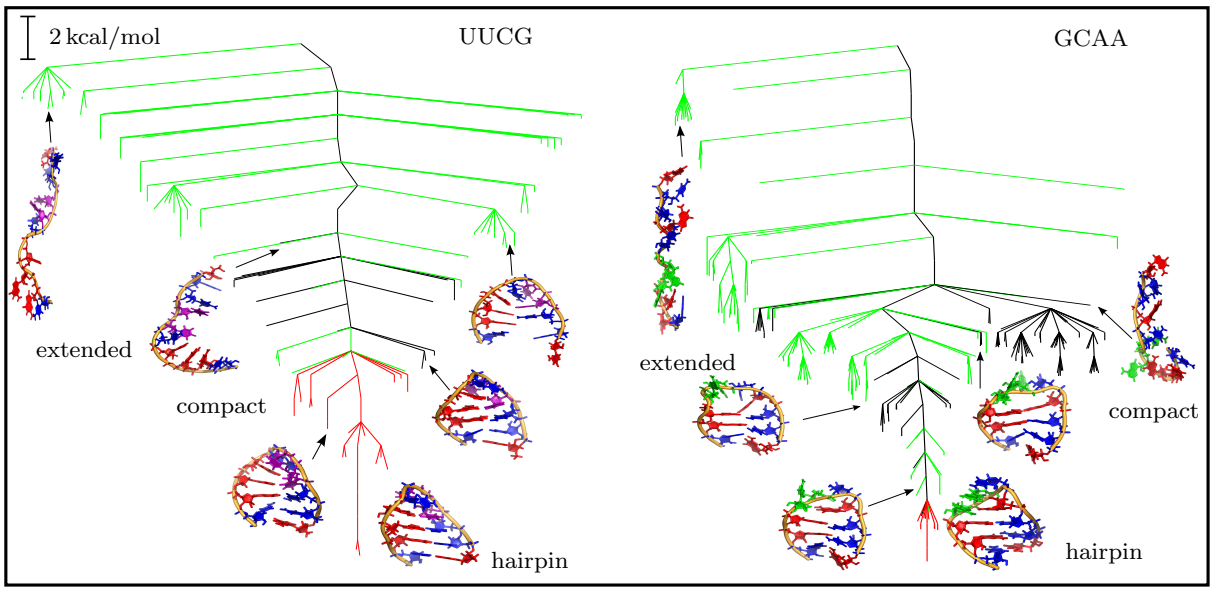

Figure 6. Potential energy disconnectivity graphs constructed using the stationary points on the 1000 fastest-folding pathways from the compact unfolded and extended states to the native UUCG and GCAA hairpins. The green branches lead to the minima that are exclusive to the pathways from the extended to the native state. The black branches lead to the minima that are encountered only along the pathways from the compact to the native state. The red branches lead to the minima that are common to all the pathways. The hairpin structures, as well as the original end points corresponding to the compact and extended states, are labelled in the graphs. Some representative minima that are encountered at different stages along the pathways are also shown.

ing procedure alleviates any bias arising due to the original choice of end points, by expanding single stationary points into ensembles of structures in local equilibrium, and the kinetics are then described in terms of transitions between groups of minima. A threshold of 4 kcal/mol is appropriate for both the sequences, as similar results were obtained in terms of disconnectivity graphs and rate constants for a range of values around this selection. The topological features of the free energy disconnectivity graphs for the two hairpins, computed with this threshold, are described in detail in the previous section.

The folding process from the compact unfolded states to the native hairpin is complete within microseconds, with estimated rate constants of $1.4 \times 10^5\,\mathrm{s}^{-1}$ for the UUCG hairpin, and $2.6 \times 10^5\,\mathrm{s}^{-1}$ for the GCAA hairpin. These estimates are in excellent agreement with values reported for hairpins of similar size in recent temperature jump experiments.[70] The folding rate constants from the extended states are found to be orders of magnitude lower, with estimated values of $1.5 \times 10^1\,\mathrm{s}^{-1}$

and $5.5 \times 10^1\,s^{-1}$ for the UUCG and GCAA hairpins, respectively.

The disparity in the observed timescales, when folding commences from different denatured ensembles on the free energy landscape, is consistent with the recent experimental findings (for DNA hairpins) of Ansari and co-workers.[41] In their study, the authors employed a combination of laser induced temperature jump and microfluidic techniques to initiate folding from different starting conditions, and found the relaxation kinetics to differ by an order of magnitude. Coarse-grained simulations of RNA hairpins[17,56] as well as pseudoknots[116] by Thirumalai and coworkers, have also reported discrepancies between force-quench or ion jump and temperature quench refolding kinetics. In our simulations, the difference of nearly four orders of magnitude (compared to previous results), may be due to the absence of explicit ions, which is known to aid rapid collapse of the highly extended states by counter-ion condensation,[11] and relative undersampling of the high energy regions of the energy landscape. In addition, we observe a variable number of base stacking interactions in the highly extended states. The unphysically strong base stacking interactions within the current AMBER force-fields[18,117] may lead to an overestimation of some of the barriers along the folding pathways from the extended states, and hence lower rate constants. Nonetheless, the resulting upper bound estimates clearly highlight the key role of unfolded state dynamics in the RNA folding problem.

Conclusion

In this work we have investigated the folding mechanisms and kinetics of RNA tetraloop hairpins from an energy landscape perspective. The discrete path sampling method was employed to map out the underlying free energy landscapes, which were visualized in terms of disconnectivity graphs. The landscapes are characterized by a distinct folding funnel at 298 K, leading to native hairpin-like states. The organization of the different structural ensembles is consistent with experimental predictions,[36–38,41] and is likely to result in a scale-free network.[118]

The folding mechanisms and kinetics were analyzed from the databases of stationary points

(kinetic transition networks) for two distinct denatured ensembles. The folding from relatively compact states is characterized by shorter paths and is complete within microseconds. In contrast, extended states, which are only sparsely populated at 298 K, exhibit much longer routes to the native state, and are associated with lower folding rate constants. Our analysis also reveals that the multiple folding routes are predicted to converge during the final stages. The strong dependence of folding rates on the choice of initial conditions is consistent with previous experimental and theoretical work, and suggests that a multi-state kinetic model is required to describe the RNA folding kinetics.

The dynamics in the unfolded state is also likely to be an integral component of the folding problem. Recent work by Pande and coworkers[119] suggests that the RNA folded state behaves as a kinetic hub, with multiple denatured states connected to the native state, interconverting amongst themselves on a timescale much longer than the folding time.

In the future we plan to evaluate the specific role of ions and solvent viscosity in modulating the folding kinetics. The kinetic transition networks described in the present work have already considered such effects at the mean field level, and should be ideal starting points for such studies.

Acknowledgement

We are grateful to Dr David de Sancho, Dr Yassmine Chebaro, Dr Guillem Portella, Dr Chris Whittleston and Dr Joanne M. Carr for helpful discussions. We also thank Mr Boris Fackovec for his comments on an initial version of the manuscript. The work was financially supported by the ERC. D.C. gratefully acknowledges the Cambridge Commonwealth and Overseas Trust for financial support.

The authors declare no competing financial interests.

Supporting Information Available

Figures showing GCAA hairpins with different stacking patterns in the loop region, distribution of radius of gyration (R_g) for the unfolded states sampled in the high temperature MD simula-

tions, potential energy disconnectivity graphs for the ungrouped databases, and figures showing the variation of (R_G) along the folding pathways, are included in the supporting information. This material is available free of charge via the Internet at `http://pubs.acs.org`.

References

(1) Bevilacqua, P. C.; Blose, J. M. *Ann. Rev. Phys. Chem.* **2007**, *59*, 79–103.

(2) Varani, G. *Ann. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379–404.

(3) Svoboda, P.; Di Cara, A. *Cell. Mol. Life Sci.* **2006**, *63*, 901–918.

(4) Chen, S. J. *Ann. Rev. Biophys.* **2008**, *37*, 197–214.

(5) Tinoco, I.; Bustamante, C. *J. Mol. Biol.* **1999**, *293*, 271–281.

(6) Brion, P.; Westhof, E. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113–137.

(7) Thirumalai, D.; Hyeon, C. Theory of RNA Folding: From Hairpins to Ribozymes. In *Non-Protein Coding RNAs*; Walter, N., Woodson, S., Batey, R., Eds.; Springer Berlin Heidelberg, 2009; Vol. 13, pp 27–47.

(8) Thirumalai, D.; Woodson, S. A. *Acc. Chem. Res.* **1996**, *29*, 433–439.

(9) Thirumalai, D.; Klimov, D. K.; Woodson, S. A. *Theor. Chem. Acc.* **1997**, *96*, 14–22.

(10) Lin, J. C.; Hyeon, C.; Thirumalai, D. *J. Phys. Chem. Lett.* **2012**, *3*, 3616–3625.

(11) Thirumalai, D.; Lee, N.; Woodson, S. A.; Klimov, D. K. *Ann. Rev. Phys. Chem.* **2001**, *52*, 751–762.

(12) Onoa, B.; Tinoco, I. *Curr. Opin. Struct. Biol.* **2004**, *14*, 374–379.

(13) Sosnick, T. R.; Pan, T. *Curr. Opin. Struct. Biol.* **2003**, *13*, 309–316.

(14) Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. *J. Am. Chem. Soc.* **2006**, *128*, 1523–1530.

(15) Jung, J.; Van Orden, A. *J. Phys. Chem. B* **2005**, *109*, 3648–3657.

(16) Ma, H.; Wan, C.; Wu, A.; Zewail, A. H. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 712–716.

(17) Hyeon, C.; Thirumalai, D. *Biophys. J.* **2006**, *90*, 3410–3427.

(18) Chen, A. A.; Garcia, A. E. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16820–16825.

(19) Kührová, P.; Banáš, P.; Best, R. B.; Šponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2013**, *9*, 2115–2125.

(20) Wales, D. J. *Energy Landscapes*; Cambridge University Press, U.K., 2003.

(21) Somani, S.; Wales, D. J. *J. Chem. Phys.* **2013**, *139*, 121909–1–121909–17.

(22) Doye, J. P. K.; Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *110*, 6896–6906.

(23) Uhlenbeck, O. C. *Nature* **1990**, *346*, 613–614.

(24) Woese, C. R.; Winker, S.; Gutell, R. R. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8467–8471.

(25) Wolters, J. *Nucleic Acids Res.* **1992**, *20*, 1843–1850.

(26) Marino, J. P.; Gregorian, J., Razmic S.; Csankovszki, G.; Crothers, D. M. *Science* **1995**, *268*, 1448–1454.

(27) Chauhan, S.; Woodson, S. A. *J. Am. Chem. Soc.* **2008**, *130*, 1296–1303.

(28) Jaeger, L.; Michel, F.; Westhof, E. *J. Mol. Biol.* **1994**, *236*, 1271–1276.

(29) Nagai, K. *Curr. Opin. Struct. Biol.* **1992**, *2*, 131–137.

(30) Thapar, R.; Denmon, A. P.; Nikonowicz, E. P. *WIREs RNA* **2014**, *5*, 49–67.

(31) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *J. Mol. Biol.* **1996**, *264*, 968–980.

(32) Sheehy, J. P.; Davis, A. R.; Znosko, B. M. *RNA* **2010**, *16*, 417–429.

(33) Porschke, D. *Biophys. Chem.* **1974**, *1*, 381–386.

(34) Bonnet, G.; Krichevsky, O.; Libchaber, A. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8602–8606.

(35) Ying, L.; Wallace, M. I.; Klenerman, D. *Chem. Phys. Lett.* **2001**, *334*, 145–150.

(36) Ansari, A.; Kuznetsov, S. V.; Shen, Y. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 7771–7776.

(37) Sarkar, K.; Meister, K.; Sethi, A.; Gruebele, M. *Biophys. J.* **2009**, *97*, 1418–1427.

(38) Sarkar, K.; Nguyen, D. A.; Gruebele, M. *RNA* **2010**, *16*, 2427–2434.

(39) Jung, J.; Van Orden, A. *J. Am. Chem. Soc.* **2006**, *128*, 1240–1249.

(40) Kim, J.; Doose, S.; Neuweiler, H.; Sauer, M. *Nucleic Acids Res.* **2006**, *34*, 2516–2527.

(41) Narayanan, R.; Zhu, L.; Velmurugu, Y.; Roca, J.; Kuznetsov, S. V.; Prehna, G.; Lapidus, L. J.; Ansari, A. *J. Am. Chem. Soc.* **2012**, *134*, 18952–18963.

(42) Kara, M.; Zacharias, M. *WIREs Comput. Mol. Sci.* **2014**, *4*, 116–126.

(43) Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *85*, 790–803.

(44) Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2516–2524.

(45) Garcia, A. E.; Paschek, D. *J. Am. Chem. Soc.* **2008**, *130*, 815–817.

(46) Zhuang, Z.; Jaeger, L.; Shea, J. E. *Nucleic Acids Res.* **2007**, *35*, 6995–7002.

(47) Zuo, G.; Li, W.; Zhang, J.; Wang, J.; Wang, W. *J. Phys. Chem. B* **2010**, *114*, 5835–5839.

(48) Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2008**, *130*, 9676–9678.

(49) Yao, Y.; Sun, J.; Huang, X.; Bowman, G. R.; Singh, G.; Lesnick, M.; Pande, V. S.; Guibas, L. J.; Carlsson, G. *J. Chem. Phys.* **2009**, *130*, 144115–1–14115–10.

(50) Lin, M. M.; Meinhold, L.; Shorokov, L.; Zewail, A. H. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4227–4239.

(51) Villa, A.; Widjajakusuma, E.; Stock, G. *J. Phys. Chem. B* **2008**, *112*, 134–142.

(52) Deng, N. J.; Cieplak, P. *Biophys. J.* **2010**, *98*, 627–636.

(53) Sorin, E. J.; Engelhardt, M. A.; Herschlag, D.; Pande, V. S. *J. Mol. Biol.* **2002**, *317*, 493–506.

(54) Nivon, L. G.; Shakhnovich, E. I. *J. Mol. Biol.* **2004**, *344*, 29–45.

(55) Hyeon, C.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6789–6794.

(56) Hyeon, C.; Thirumalai, D. *J. Am. Chem. Soc.* **2008**, *130*, 1538–1539.

(57) Zhang, W.; Chen, S. J. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1931–1936.

(58) Chen, S. J.; Dill, K. A. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 646–651.

(59) Kuznetsov, S. V.; Ansari, A. *Biophys. J.* **2012**, *102*, 101–111.

(60) Zhang, W.; Chen, S. J. *Biophys. J.* **2006**, *90*, 765–777.

(61) Zhang, W.; Chen, S. J. *Biophys. J.* **2006**, *90*, 778–787.

(62) Cao, S.; Chen, S. J. *Biophys. J.* **2009**, *96*, 4024–4034.

(63) Faber, M.; Klumpp, S. *Phys. Rev. E* **2013**, *88*, 052701–1–052701–9.

(64) Sauerwine, B.; Widom, M. *Phys. Rev. E.* **2011**, *84*, 061912–1–061912–8.

(65) Wales, D. J. *Mol. Phys.* **2002**, *100*, 3285–3305.

(66) Wales, D. J. *Mol. Phys.* **2004**, *102*, 891–908.

(67) Wales, D. J. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3–10.

(68) Noe, F.; Fischer, S. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162.

(69) Wales, D. J. *J. Chem. Phys.* **2009**, *130*, 204111(1)–204111(7).

(70) Proctor, D. J.; Ma, H.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. C. *Biochemistry* **2004**, *43*, 14004–14014.

(71) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. *J. Mol. Biol.* **2000**, *304*, 35–42.

(72) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *J. Mol. Biol.* **1996**, *264*, 968–980.

(73) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *17*, 5108–5121.

(74) Malolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S. N.; Wales, D. J. *J. Comput. Chem.* **2010**, *31*, 1402–1409.

(75) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.

(76) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.

(77) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.

(78) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B.* **2000**, *104*, 3712–3720.

(79) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426–434.

(80) Case, D. A.; Darden, T. A.; Cheatham, T.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I. *AMBER 12*, 2012. `http://ambermd.org/`.

(81) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523–535.

(82) Evans, D. E.; Wales, D. J. *J. Chem. Phys.* **2004**, *121*, 1080–1090.

(83) Carr, J. M.; Wales, D. J. *J. Phys. Chem. B* **2008**, *112*, 8760–8769.

(84) Farrell, J. D.; Lines, C.; Shepherd, J. J.; Chakrabarti, D.; Miller, M. A.; Wales, D. J. *Soft Matter* **2013**, *9*, 5407–5416.

(85) Murrell, J. N.; Laidler, K. J. *Trans. Faraday Soc.* **1968**, *64*, 371–377.

(86) Liu, D.; Nocedal, J. *Math. Program.* **1989**, *45*, 503–528.

(87) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082–2094.

(88) Henkelman, G.; Jònsson, H. *J. Chem. Phys.* **1999**, *111*, 7010–7022.

(89) Henkelman, G.; Uberuaga, B. P.; Jònsson, H. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(90) Munro, L. J.; Wales, D. J. *Phys. Rev. B.* **1999**, *59*, 3969–3980.

(91) Wales, D. J. *OPTIM: A program for optimising geometries and calculating pathways*, http://www-wales.ch.cam.ac.uk/software.html.

(92) Case, D. A.; Darden, T. A.; Cheatham, T.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I. *AMBER 9*, 2006. `http://ambermd.org/`.

(93) Strodel, B.; Whittleston, C. W.; Wales, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 16005–16014.

(94) Wales, D. J. *PATHSAMPLE: A program for generating connected stationary point databases and extracting global kinetics*, http://www-wales.ch.cam.ac.uk/software.html.

(95) Dijkstra, E. W. *Numer. Math.* **1959**, *1*, 269–271.

(96) Carr, J. M.; Wales, D. J. In *The Energy Landscape as a Computational Tool*; Conner-ade, J. P., Solov'yov, A., Eds.; Imperial College Press, London, 2008; pp 321–330.

(97) Jimènez, J.; Marzal, A. In *Algorithm Engineering: 3rd International Workshop, WAE'99, London, UK, July 1999*; Springer, Berlin, 1999; pp 15–29.

(98) Strodel, B.; Wales, D. J. *Chem. Phys. Lett.* **2008**, *466*, 105–115.

(99) Hoare, M. R.; McInnes, J. J. *Faraday Discuss. Chem. Soc.* **1976**, *61*, 12–24.

(100) Hoare, M. R. In *Advances in Chemical Physics*; John Wiley and Sons, USA, 1979; Vol. 40, pp 49–129.

(101) Wales, D. J.; Salamon, P. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 617–622.

(102) Becker, O. M.; Karplus, M. *J. Chem. Phys.* **1997**, *106*, 1495–1517.

(103) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14766–14770.

(104) Krivov, S. V.; Karplus, M. *J. Chem. Phys.* **2002**, *117*, 10894–10903.

(105) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2003**, *118*, 3891–3897.

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

(107) Allain, F. H.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333–353.

(108) DePaul, A. J.; Thompson, E. J.; Patel, S. S.; Haideman, K.; Sorin, E. J. *Nucleic Acids Res.* **2010**, 4856–4867.

(109) Zhang, Y.; Zhao, X.; Mu, Y. *J. Chem. Theory Comput.* **2009**, *5*, 1146–1154.

(110) Zhao, L.; Xia, T. *J. Am. Chem. Soc.* **2007**, *129*, 4118–4119.

(111) Sarzynska, J.; Nilsson, L.; Kullinski, T. *Biophys J.* **2003**, *85*, 3445–3459.

(112) Schrödinger, LLC, *The PyMOL Molecular Graphics System, Version 1.4.1*, 2010.

(113) Lapidus, L. J. *Curr. Opin. Struct. Biol.* **2013**, *23*, 30–35.

(114) Portella, G.; Orozco, M. *Angew. Chem. Intl. Ed.* **2010**, *49*, 7673–7676.

(115) Wales, D. J.; Head-Gordon, T. *J. Phys. Chem. B.* **2012**, *116*, 8394–8411.

(116) Biyun, S.; Cho, S. S.; Thirumalai, D. *J. Am. Chem. Soc.* **2011**, *133*, 20634–20643.

(117) Banáš, P.; Mládek, A.; Otyepka, M.; Zgarbová, M.; Juerčka, P.; Svozil, D.; Lankaš, F.; Šponer, J. *J. Chem. Theory Comput.* **2012**, *8*, 2448–2460.

(118) Doye, J. P. K. *Phys. Rev. Lett.* **2002**, *88*, 238701/1–238701/4.

(119) Huang, X.; Yao, Y.; Bowman, G. R.; Sun, J.; Guibas, L. J.; Carlsson, G.; Pande, V. S. *Pac. Symp. Biocomput.* **2010**, *15*, 228–239.