

# Topological constraints: using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation

Maximillian H Bailor, Anthony M Mustoe, Charles L Brooks III and Hashim M Al-Hashimi

Accompanying recent advances in determining RNA secondary structure is the growing appreciation for the importance of relatively simple topological constraints, encoded at the secondary structure level, in defining the overall architecture, folding pathways, and dynamic adaptability of RNA. A new view is emerging in which tertiary interactions do not define RNA 3D structure, but rather, help select specific conformers from an already narrow, topologically pre-defined conformational distribution. Studies are providing fundamental insights into the nature of these topological constraints, how they are encoded by the RNA secondary structure, and how they interplay with other interactions, breathing new meaning to RNA secondary structure. New approaches have been developed that take advantage of topological constraints in determining RNA backbone conformation based on secondary structure, and a limited set of other, easily accessible constraints. Topological constraints are also providing a much-needed framework for rationalizing and describing RNA dynamics and structural adaptation. Finally, studies suggest that topological constraints may play important roles in steering RNA folding pathways. Here, we review recent advances in our understanding of topological constraints encoded by the RNA secondary structure.

## Address

Department of Chemistry and Biophysics, University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109-1055, United States

Corresponding authors: Brooks, Charles L ([brookscsl@umich.edu](mailto:brookscsl@umich.edu)) and Al-Hashimi, Hashim M ([hashimi@umich.edu](mailto:hashimi@umich.edu))

Current Opinion in Structural Biology 2011, 21:296–305

This review comes from a themed issue on  
Nucleic acids  
Edited by Anna Marie Pyle and Zippi Shakked

Available online 14 April 2011

0959-440X/\$ – see front matter

© 2011 Elsevier Ltd. All rights reserved.

DOI [10.1016/j.sbi.2011.03.009](https://doi.org/10.1016/j.sbi.2011.03.009)

## Introduction

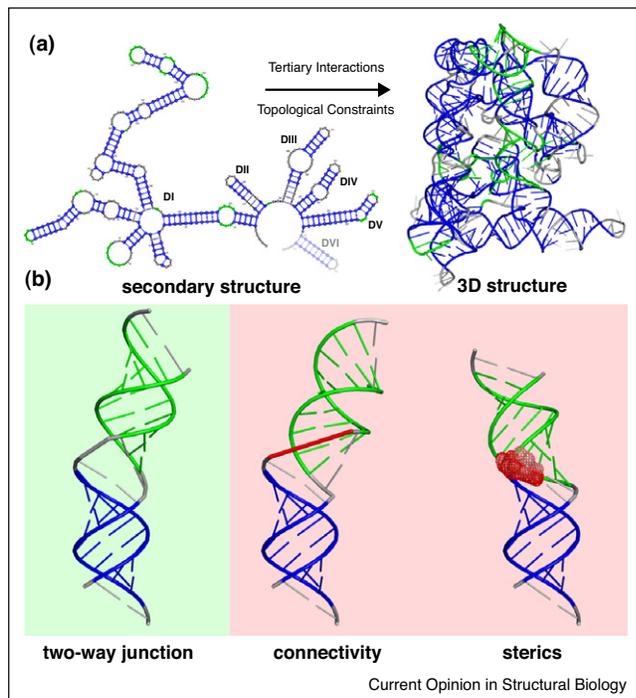
The strongest forces that define the structures of macromolecules are also the simplest ones; atoms must satisfy stereochemical bonding constraints and avoid steric collisions. For example, these constraints alone restrict torsion angles in nucleic acids to <5% of all possible

conformations [1]. At longer length scales, RNA secondary structure consists of A-form helical domains that carry various receptors involved in tertiary interactions and that are linked together by a variety of junctions, including bulges, internal loops, three-way and other higher-order junctions, and single strands. If we think of A-form helices as objects with variable shape and size, and junctions as linkers with variable stereochemical properties, higher-order topological constraints come to existence that severely restrict, and in some cases define, the allowed range of RNA conformations. Recent biophysical studies are providing insights into these constraints while others are exploiting them to make predictions about RNA structure and dynamic adaptation.

## Overview of topological constraints

We begin discussion by defining what we mean by topological constraints. For this purpose, we consider the secondary structure of an RNA in its native folded state, which can be determined with ever-increasing accuracy using experimental-computational approaches (Figure 1a). Even in the absence of tertiary interactions, and ignoring electrostatic interactions, the RNA secondary structure alone imposes significant topological constraints on the allowed range of conformations. In particular, stable A-form helices exist in great abundance and their backbone conformation can to a very good approximation be modeled assuming an A-form helix geometry [2]. These helical elements are in turn tethered together at two positions on either side of the helix perimeter by single strands that are short (~5 Å per nt) in comparison to the local diameter of the A-form helix (~17 Å across a base-pair). This gives rise to *connectivity* constraints, that is, reorientation and translation of one helix relative to another, is severely restricted by the short length of the two tethers on either side of the helix (Figure 1b). The specific nature of these connectivity constraints will vary depending on the junction topology, including junction order (two-way, three-way, etc.), relative lengths of individual strands, and whether bases in junctions adopt a looped in or looped out conformation. In addition, *steric* constraints between RNA helices and other elements, further restrict the allowed translation and reorientation of helices with respect to one another (Figure 1b). These steric collisions can be long-range in nature, occurring between helices that are far apart in sequence. Together, the connectivity and steric constraints define the topological constraints encoded by an RNA secondary structure.

Figure 1



Topological constraints encoded by the RNA secondary structure. **(a)** Secondary and tertiary structure of group IIC intron from *O. iheyensis* (PDB 3IGl) [56]. **(b)** Connectivity and steric constraints in a two-way junction.

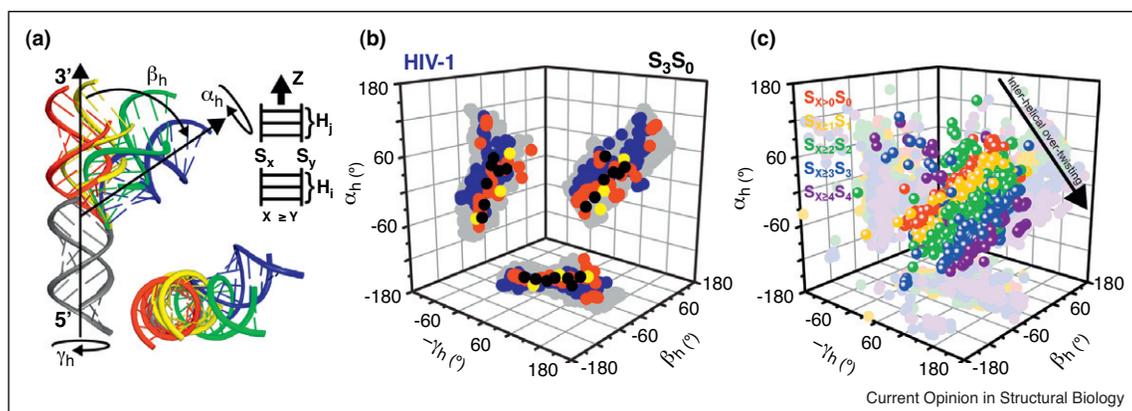
## Inter-helical junctions

Using electrophoretic gel mobility measurements, fluorescence resonance energy transfer, transient electric birefringence, and electron microscopy, Lilley [3–5], Hsieh and Griffith [6], and Zacharias and Hagerman [7,8] showed more than two decades ago that bulges and

internal loops induce directional bends in DNA and RNA duplexes by amounts dependent on the length, asymmetry, and sequence of the junction as well as on ionic strength. Longer bulges induced greater bends but only up to a certain length of  $\sim 5$ – $7$  nucleotides. U-rich bulges induced smaller bends compared to A-rich bulges and were more affected by the addition of  $Mg^{2+}$  ions. Subsequent studies employing single molecule techniques [9] and NMR spectroscopy [10] confirmed that junctions do not induce rigid bends, but rather, a dynamic range of conformations and that this flexibility is important for allowing adaptive conformational changes to take place during recognition [11] and catalysis [12].

Recent studies on simple two-way junction model systems are helping explain these early observations, resulting in the establishment of a framework for quantitatively predicting certain aspects of inter-helical junction behavior based on secondary structure alone. Chu *et al.* [13] used stochastic dynamic simulations to understand how junction topologies define the allowed orientation sampled by helices. They used a duplex model system consisting of two DNA helices adjoined by either one or two polyethylene glycol (PEG) tethers as a mimic for single and double stranded junctions, respectively. The PEG junctions allowed the authors to focus on topological constraints and avoid contributions from sequence-specific interactions. In these simulations, the helices were coarse grained thus significantly increasing the efficiency with which the inter-helical space could be sampled. The simulations revealed that the double stranded PEG junction significantly confines the allowed inter-helical orientations along a hinge-like pathway and that the confinement was less severe and directional for the single stranded junction (Figure 2a). Thus, the ani-

Figure 2



Topological constraints confine the allowed inter-helical orientations across bulges. **(a)** Three inter-helical Euler angles ( $\alpha_h, \beta_h, \gamma_h$ ) used to define the relative orientation of 5' and 3' helices across two-way junctions containing  $X$  and  $Y$  ( $X \geq Y$ ) single stranded (S) residues. **(b)** 3D inter-helical orientation maps showing individual 2D projections along each plane of a 3D cube with associated correlation coefficients ( $R$ ) between the inter-helical twist angles for three-nucleotide bulged RNA TAR ( $S_3S_0$ , HIV-1 TAR). The ligand bound, NMR RDC, NMR-MD, PDB-derived, and topologically computed inter-helical distributions are shown in black, yellow, blue, red, and grey, respectively. **(c)** The PDB-derived inter-helical distributions for various families of Y-strand length junctions.

sotropic kinking of helices is simply the result of the junction topology and does not require the invocation of any sequence-specific interactions.

Concurrently, Bailor *et al.* [14\*\*] showed that topological constraints severely confine the global orientation of helices across two-way junctions. The authors developed and standardized three inter-helical Euler angles ( $\alpha_h$ ,  $\beta_h$ ,  $\gamma_h$ ) for describing the orientation of helices across junctions [15];  $\alpha_h$  and  $\gamma_h$  specify a twist around the axis of each of the two helices and  $\beta_h$  specifies an inter-helical bend angle (Figure 2b). The authors measured the inter-helical Euler angles for all two-way junctions in the protein databank (PDB) and found that the range of orientations sampled was restricted to ~1–4% of the total space (Figure 2a). Increasing the length of the internal loop resulted in an increase in the range of orientations sampled, whereas increasing the asymmetry resulted in a systematic shift in the inter-helical twist angle ( $\alpha_h + \gamma_h$ ) by an amount comparable to the helical periodicity (~33°) (Figure 2a). The authors rationalized these observations by computing the allowed inter-helical orientations subject to two trivial topological constraints: helices cannot collide with one another and cannot assume conformations that cannot be satisfactorily linked based on the length of the junction. The computed space sampled only 4–20% and on average 7% (for internal loops <4 nucleotide long) of the total space, yet it accommodated ~85% of all orientations observed in the PDB (Figure 2a). The systematic ~33° shift in  $\alpha_h + \gamma_h$  was rationalized by the tendency of junction residues to maximize non-canonical base-pairing, which in turn leads to increased inter-helical over twisting.

Recent follow-up studies by Mustoe *et al.* [16] showed that nucleotides within two-way junctions have a strong bias towards looping into the junction to maximize stacking interactions. Manual annotation of several hundred RNA structures revealed that the majority of these junctional residues form non-canonical base-pairs. This makes it possible to assign a ‘default’ bulge topology for any two-way junction, and to thereby compute the topologically allowed inter-helical space for any two-way junction based on secondary structure alone. Notable exceptions to this ‘default’ behavior are cases in which tertiary interactions stabilize looped out junctional nucleotides [16]. Residues in symmetric internal loops tend to form non-canonical base-pairs and give rise to a conformational behavior similar to that observed for a continuous A-form helix.

One can expect topological constraints similar to those mentioned above to manifest themselves in higher-order junctions. Early studies on a small number of model systems showed that helices in such junctions tend to adopt a limited set of conformations often involving coaxial alignment of one or two pairs of helices [5]. The ever-growing number of X-ray structures of

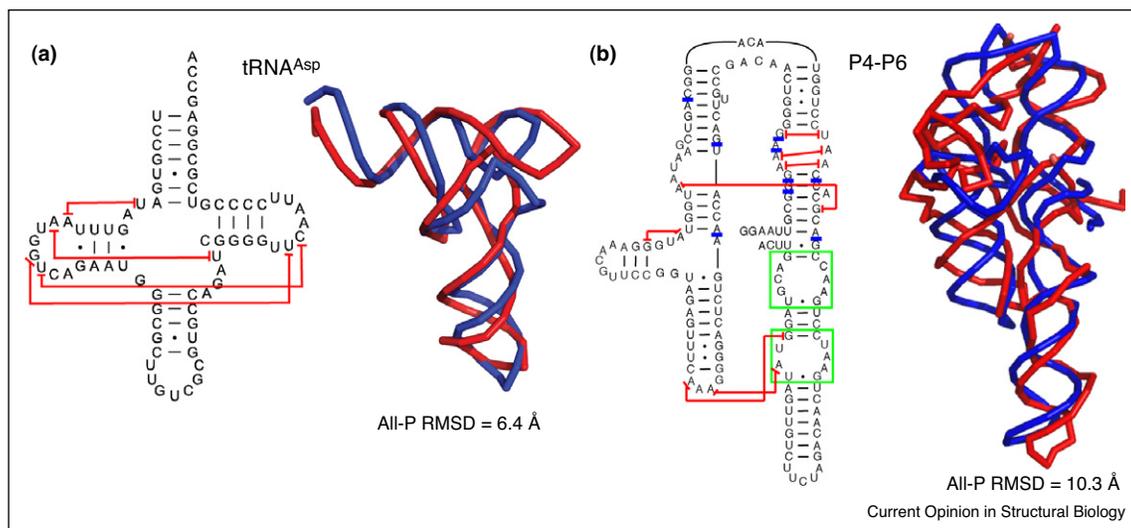
higher-order junctions provides support for these early observations. For example, Lescoute and Westhof [17\*] showed that three-way junctions are dominated by only three classes of global structure and suggested that global orientation may be predicted based on secondary structure alone. Laing and Schlick [18] analyzed four-way junctions and showed correlations between global structure and secondary structure, where helices with short single-stranded connectors are more likely to coaxially stack. Interestingly, if one defines the helices as H1, H2, H3, and H4 according to the order they are encountered when following the RNA strand from the 5' → 3' direction, one finds a strong tendency for H1 to stack on H4 and H2 to stack on H3. The authors speculate that this is due to the right-handedness of A-form helices. Laing *et al.*'s follow-up studies on higher-order junctions also revealed similar helical motifs [19].

A deeper appreciation of these topological constraints is making it possible to better interpret mutagenesis data. A salient example is a recent study by Zhang *et al.* [20\*] on a long five nucleotide pyrimidine rich J2ab bulge in human telomerase. Using RDC NMR analysis, the authors found that the bulge induces a sharp and flexible bend of ~90° in a manner that appears to be relatively independent of sequence. Decreasing the bulge length resulted in a marked decrease in telomerase activity. The authors conclude that the bulge is necessary for both inducing a highly bent structure and allowing for dynamic adaptation needed to guide telomerase activity.

### Interplay between junction-included topological constraints and other interactions

Recent studies are providing insights into the interplay between topological constraints encoded by RNA inter-helical junctions and other interactions. Most notably, it is now becoming apparent that secondary structure encoded topological constraints play a prominent role specifying the geometric positioning of groups involved in key tertiary contacts. For example, simulations by Raymond *et al.* [21] show that constraining the HCV ribozyme to its secondary structure results in a global fold that places distant loops, which participate in a tertiary pseudo-knot motif, in close proximity to one another. This suggests that the topology of the RNA secondary structure is optimized to stabilize specific tertiary interactions. Indeed, prior studies by the Herschlag group on model PEG tethered duplexes [13\*\*] showed that junction-encoded topological constraints can geometrically influence the ability of an RNA to form tertiary contacts. Specifically, the authors showed that the probability of forming hypothetical tertiary contacts between two helices depends on the specific geometrical position of the contact relative to the inter-helical junction and that this dependence weakens when linking helices by a single rather than double stranded PEG tether (Figure 3). The study by Mustoe *et al.* [16] mentioned

Figure 3



Determining backbone global conformation of (a) *tRNA<sup>Asp</sup>* and (b) *Tetrahymena* ribozyme P4–P6 domain based on secondary structure constraints and limited tertiary contacts. Predicted structures (in red) are superimposed onto the native X-ray structure (in blue). Adapted from [35\*,30\*] and (Flores, Sherman, Bruns, Eastman, Altman: *Trans Comput Biol Bioinform*, 2010, in press). Secondary structure motifs within green boxes correspond to motif environments that were subjected to constraints derived from comparison to independent crystal and NMR structures. Blue bars between bases correspond to imposed stacking interactions and red connectors denote imposed tertiary contact constraints.

above showed that the reverse scenario can also occur. Tertiary contacts can modify junction topology by inducing the looping out of residues that would otherwise participate in non-canonical base-pairing. The modified junction topology can in turn allow access to inter-helical conformations that would be inaccessible to the ‘default’ junction topology.

While topological constraints can help define a range of inter-helical orientations that can be sampled by a given junction, other interactions can selectively stabilize specific conformers from this allowed range. One example is sequence-specific stacking interactions. By replacing a flexible AU base-pair flanking a trinucleotide bulge in HIV-1 TAR with a stronger GC base-pair, Stelzer *et al.* [22\*] engineered a TAR mutant in which the helices coaxially stack in a rigid conformation, rather than adopt the bent and flexible conformation of wild-type TAR. Another example is interactions with the ion atmosphere [5,7,23]. In their same study on PEG linked-DNA duplexes, Chu *et al.* [13\*\*] showed that at low ionic strength, electrostatic repulsion between the helices limits the range of accessible conformations, but that increasing the ionic strength allows access to greater regions of the topologically allowed space. Other examples from recent work completed independently by the Woodson and Draper groups reveal that molecular crowding and some osmolytes lead to stabilization of more compact RNA conformations [24,25], presumably by selecting more compact conformations from the topologically allowed distribution.

### From RNA secondary structure to 3D conformation

A growing number of computational modeling studies are taking advantage of topological constraints in making useful inferences about RNA backbone conformation based primarily on secondary structure information. These are exciting developments considering recent advances that make it possible to determine secondary structures of very large, complex, and transient RNAs that do not lend themselves to high-resolution structure determination by NMR spectroscopy and X-ray crystallography [26]. In what follows, we review such applications; not reviewed are other approaches for atomic-level RNA structure prediction, which so far have primarily been applied to smaller (<50 nts) systems [27,28].

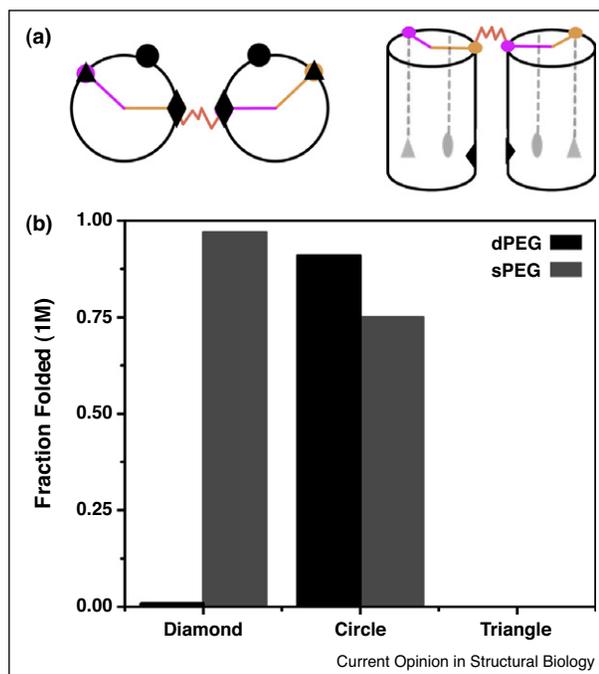
Weeks and coworkers [29\*\*] explored the extent to which secondary structure constraints can help define RNA backbone conformation. They performed a series of computational simulations using a coarse-grained discrete molecular dynamic model to generate structural ensembles for RNAs ranging in size between 27 and 161 nucleotides. Inclusion of secondary structure constraints alone was enough to reduce the ensemble’s mean pairwise root-mean-square-deviation (rmsd) from ~30 to ~15 Å for a 100 nt system. Without tertiary restraints the structure of the P4–P6 domain (160 nt) from the *Tetrahymena* ribozyme (TR) is ill-defined and Flores and Altman’s RNABuilder software [30\*] returns a ‘predicted’ structure with 45 Å all P rmsd from the native structure [30\*]. Similarly, structures of TR P4–P6 pre-

dicted using FARNAs [31] and secondary structure constraints superimpose upon the X-ray structure with all-C4' rmsds ranging between 12 Å and 55 Å, and on average ~30 Å [32<sup>••</sup>]. Even for the smaller example of tRNA<sup>Asp</sup> (75 nt), 3D structure predictions using the DMD software [33] with secondary structure alone produce an 11.0 Å rmsd from the X-ray structure when considering phosphorus atoms of base-paired nucleotides only [34].

Interestingly, much of the remaining uncertainty in defining RNA backbone conformation can be lifted using a limited number of strategically well chosen 'other' readily accessible constraints. Examples of these include tertiary contacts deduced by footprinting experiments, thermodynamic calculations, covariation/mutational analysis [30<sup>•</sup>,35<sup>•</sup>] as well as SAXS [36], Cryo-EM and NMR RDC data [37]. The addition of such constraints has been shown to improve the precision of candidate models for all phosphorus rmsds to <10 Å. For example, by including constraints from ~5 tertiary contacts, Lavender *et al.* [35<sup>•</sup>] built models for tRNA<sup>Asp</sup> (75 nt) and *Thermus thermophilus* group I intron P4–P6 domain (158 nt) with all phosphorus atom rmsds relative to the X-ray structure of 6.4 Å and 11.3 Å, respectively. The agreement improves to 3.6 Å and 5.4 Å for the smaller HHR (67 nt) and CrPV (49 nt) RNA, respectively. Likewise, using constraints inferred from a limited number of tertiary contacts, stacking interactions, and NMR data, Flores and Altman [30<sup>•</sup>] generated models for tRNA<sup>Phe</sup> (76 nt) and the *Tetrahymena* group I intron P4–P6 domain (160 nt) with all phosphorus atom rmsds of 9.6 Å and 10.0 Å, respectively. As shown in Figure 4, these predictions accurately capture the global conformation of the molecules and represent the best predictions to date that have been achieved for these systems with less than ~10 tertiary constraints. Additional results obtained from Yang *et al.* [36] and Jonikas *et al.* [38], while achieving lower accuracy, demonstrate the robustness of this approach across varied sets of prediction algorithms and tertiary constraints. As might be expected, increasing the density of other constraints can improve the precision of models of all phosphorus rmsds to <5 Å. For example, Wang *et al.* [39] used a combination of secondary structure, SAXS data, and NMR RDCs to construct a 3D model of the riboA RNA (71 nt) that superimposed with 3.0 Å backbone-rmsd of the X-ray structure.

As such, new biochemical approaches have been and are being developed to rapidly obtain distance constraints. In the so-called MOCHA method developed by Das *et al.* [32<sup>••</sup>], modified nucleotides are randomly incorporated into an RNA molecule during transcription, tethered to Fe(II)-EDTA moieties, and then allowed to generate random cleavages that occur within 25 Å radius of the modified nucleotide. Two dimensional gel electrophoresis separation is used to locate the site of cleavage of the modified nucleotide, providing pairwise distance

Figure 4



Folding simulations of model HJH motifs demonstrate the formation of tertiary contacts is selected based on contact location linker — single PEG linker (sPEG) and double PEG linkers (dPEG) — properties. (a) Schematic representation of the sPEG HJH motif in a side-by-side helical configuration. Tertiary contacts, highlighted by diamond, circle, and triangle, were placed in three separate locations and investigated computationally. (b) Comparison of the fraction folded sPEG (grey) and dPEG (black) in 1 M monovalent salt concentration for three tertiary contact. The two junction topologies have different conformational preferences (sPEG prefers the diamond position, dPEG prefers circle) and folding specificities (sPEG is more receptive to tertiary contact location) which results in differences of folded fractions. Adapted from [13<sup>••</sup>].

constraints of ~25 Å. Using this approach, the authors were able to measure 77 distance constraints in *Tetrahymena* group I intron P4–P6 domain and thereby generate a structure model with an all-C4' atom rmsd accuracy of 13.1 Å. Gherghe *et al.* [34] used a sequence-specific intercalating agent, MPE, bound to Fe(II)-EDTA moieties to perform cleavage experiments that provide distance constraints between the engineered intercalation site and nucleotides within ~30 Å. Application of this method to tRNA<sup>Asp</sup> yielded a total of ~200 distance constraints. These were combined with secondary structure as inputs to their DMD model, giving a prediction accurate to 4.0 Å all-P rmsd relative to the native structure.

Studies have also been recently reported of models of larger RNAs for which high-resolution X-ray or NMR structures are currently not available. As an example, Lavender *et al.* [35<sup>•</sup>] used a secondary structure along with six tertiary constraints to build a model for the 120 nucleotide HCV IRES pseudo-knot domain that is

consistent with hydroxyl radical cleavage patterns and cryo-EM electron density maps. Lipfert *et al.* [40] were able to construct a model of the 192 nt VS ribozyme using manual fitting of secondary structure into a SAXS electron density profile with incorporation of tertiary contacts known to exist from biochemical experiments. Zhang *et al.* used a combination of secondary structure and RDCs with the MC-Sym software to build a model for the human telomerase core domain [20<sup>\*</sup>]. At an even larger scale, the Harvey group has been utilizing the YAMMP software along with hypothetical secondary structures to predict the structures viral genomes [41].

Another particularly interesting application of topological constraints, which involves the use of secondary and tertiary structure constraints to produce 3D RNA structures was demonstrated by the Perreault group [21] and the Herschlag group [32<sup>\*\*</sup>]. Using prior experimental data on the tertiary contacts present in different intermediates along the folding pathway of the HDV ribozyme Raymond *et al.* were able to use the MC-Sym software to make predictions about the 3D conformation of these intermediates and infer the 3D structural changes that occur during folding. Similarly, Das *et al.* [32<sup>\*\*</sup>] used MOCHA pairwise distance constraints obtained at different salt concentrations to build NAST models of the unfolded and non-native, compact states of the P4-P6 domain of the TR. Although molecular dynamic simulations could theoretically provide information on structures of unfolded states (and to higher resolution), the computation at this time is infeasible. These studies highlight the power of using secondary structure along with tertiary constraints to address questions concerning RNA structure that would be otherwise intractable. With accurate secondary structures now obtainable for molecules as large as viral genomes [42,43], we anticipate the further development and application of these methods towards addressing 3D structural characteristics of truly large genomic RNA structures.

### Folding on frustrated energy landscapes: role of topology

The formation of competing helices during RNA co-transcriptional folding is expected to give rise to topological constraints that may in turn steer folding pathways [44]. So-called Gō models are proving to be a useful tool to explore these aspects of RNA folding. Gō models have been utilized in protein folding studies to delineate dominant folding pathways in cases where the formation of elements of structure, for example, domains and secondary structural components compete to give rise to complex folding kinetics [45–47]. These models are parameterized to only include attractive non-bonded forces that are present in the native structure; all other non-bonded interactions are treated as repulsive Lennard–Jones potentials. Simulations with these models are especially well suited to studying how the formation of

secondary structure elements, and the topological constraints they induce, in differential orders impacts overall folding of the RNA.

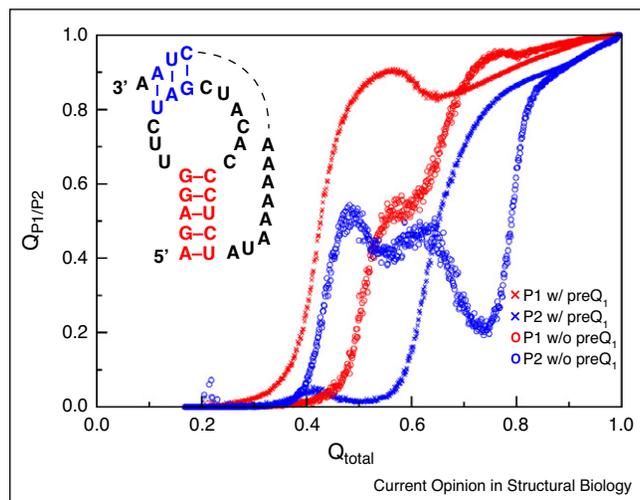
For instance, Thirumalai and coworkers used coarse-grained Gō models to explore the folding of three RNA pseudo-knots from the viral genomes of MMTV and SRV-1 and from human telomerase [48]. By combining thermodynamic based studies with kinetic folding studies Cho *et al.* [48], observe that the relative stabilities of the isolated helices control the folding mechanisms. The kinetic simulations they performed corroborate inferences drawn from the free energy profiles, suggesting that MMTV folds by a hierarchical mechanism with parallel pathways wherein the formation of one helix nucleates the assembly of the remainder of the structure. In the SRV-1 pseudo-knot they observe a more cooperative folding process in which preformed helices simultaneously consolidate the tertiary structure. Folding occurs by multiple pathways in the hTR pseudo-knot, and the potential for topological frustration is evident in the observation of a competing interaction pathway involving differing parts of the tertiary structure.

Moreover, Gō models have also been successfully applied to studies of riboswitches, a class of RNA molecules found on the 5' end of mRNAs that have several competing secondary structures where the choice of one structure or the other governs gene expression. Studies by Lin and Thirumalai [49<sup>\*\*</sup>] indicate that folding of the three-helical A-riboswitch aptamer proceeds hierarchically as determined by the stability of the helical elements. Furthermore, ligand binding can act to stabilize a meta-stable state on a time frame long enough to commit the RNA to a given secondary structure thereby preventing it from forming alternative, competing structures.

Recently, the preQ<sub>1</sub> riboswitch has also been examined using an all-atom Gō model by Feng *et al.* [50]. The authors found that folding was highly cooperative in the presence of metabolite and involved the sequential formation of helix P1 followed by the docking of helix P2 (Figure 5) [50]. These observations are consistent with the idea of local stability determining folding order as observed by Cho *et al.* [48] and Lin and Thirumalai [49<sup>\*\*</sup>]. However, in the absence of preQ<sub>1</sub>, formation of P1 is less facile and competition for the formation of helical interactions in P2 creates a topological trap for structure formation, where premature formation of P2 requires prior unfolding of P2 for assembly of the P1 helix (Figure 5). Clearly, preQ<sub>1</sub> interactions with the P1 helix further stabilize this helix and nucleate the formation of P1 before the establishment of interactions involving the P2 helix.

Similar to the preQ<sub>1</sub> system, all-atom Gō models studies of the S-adenosylmethionine-1 (SAM-1) riboswitch found that formation of the nonlocal helix (P1) is rate-limiting in

Figure 5



Formation of helical elements in the preQ1 riboswitch. All-atom Go model folding simulations of the preQ1 riboswitch showing formation of the P1 (red) and P2 (blue) helix contacts as a function of the total native contacts in the presence and absence of preQ1 ligand. Results are averaged over 78 kinetic folding simulations. Adapted from Feng *et al.* [50].

aptamer domain formation [51<sup>•</sup>]. In the absence of metabolite the 3' end of the P1 helix is sequestered by involvement in an alternative helix. Binding of SAM-1 induced the switching of the conformational states and provides a mechanistic explanation of expression platform regulation. The authors also observe an interesting interplay between the formation of helix P1 and a neighboring pseudo-knot. In a trend that is particularly apparent in their kinetic simulations they find that partial unfolding of the pseudo-knot domain is correlated with the formation of helix P1, which then itself must partially melt to allow cooperative refolding of the two elements. This surprising result suggests that the prior formation of the pseudo-knot results in a topological trap or knot that must be undone in order to form the final folded structure.

### Dynamics and conformational adaptation

Most regulatory RNAs not only have to fold into specific 3D structures, but also have to retain a degree of flexibility to allow their structure to change adaptively during ribonucleoprotein assembly, recognition, catalysis, and signaling. Recent approaches that employ NMR [52] and SAXS [53] are providing insights into the nature of RNA dynamic ensembles and topological constraints and providing a framework for rationalizing these observations.

For example, domain-elongation RDC NMR studies by Zhang *et al.* [54] and subsequent studies combining NMR and MD simulations by Stelzer *et al.* [55] showed that two helices in HIV-1 TAR RNA dynamically sample a limited range of the total inter-helical space forming a distri-

bution in which the helices twist in a correlated manner. Decreasing the bulge length resulted in a more confined distribution and greater motional correlations. These aspects of the inter-helical distributions could readily be rationalized based on the topologically allowed inter-helical space [14<sup>••</sup>]. The computed space closely coincided with that of the dynamic distribution, implying that the two helices efficiently explore all topologically allowed inter-helical orientations (see Figure 2b).

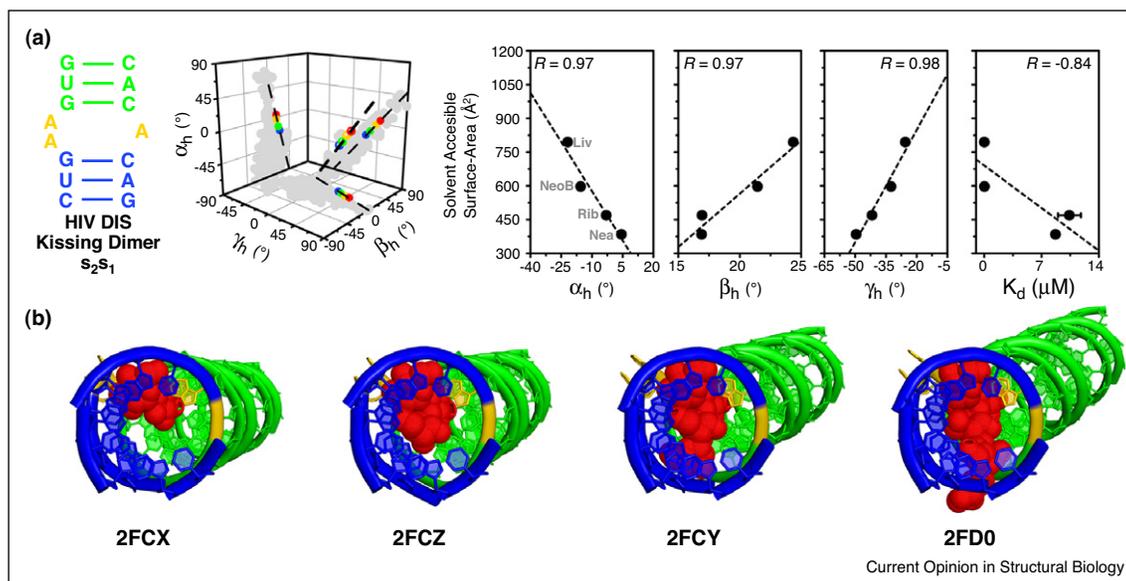
A number of other studies are also revealing a link between topological constraints and conformational adaptation. By changing the sequence-identity of Watson–Crick base-pairs neighboring the HIV-1 TAR bulge, Stelzer *et al.* [22<sup>•</sup>] engineered an HIV-1 TAR mutant that favors a coaxial conformation similar to that observed for TAR when bound to ligand mimics of its cognate target Tat. By biasing the inter-helical distribution to the bound conformation, the authors showed that the TAR mutant bound to a Tat ligand mimic with slightly increased affinity without affecting the resulting ligand bound TAR conformation.

Finally, a study by Bailor *et al.* [14<sup>••</sup>] suggests that topology helps define not only dynamic aspects of RNA global structures, but also, which structures are selected and stabilized upon binding to specific ligands. Comprehensive analysis of RNA structures bound to small molecules that differ in size, shape, and charge, revealed that small changes in the chemical structure of small molecules leads to stabilization of slightly different RNA structures within the allowed topological space. In particular, a linear correlation was observed between the small molecule size and the inter-helical angles describing the bound RNA inter-helical conformation, with larger small molecules stabilizing more bent and twisted conformations (Figure 6). Conversely, changing other aspects of the small molecule, such as its charge, without affecting size, did not lead to sizeable changes in the bound RNA conformation. Close inspection of X-ray structures of RNA complexes reveals that small molecules tend to act as a wedge between helices, thus inducing bending along the topologically allowed pathway by amounts dependent on their size (Figure 6). These results suggest that different RNA conformations from a topologically allowed distribution can be selected by the added topological constraints arising from the interactions with small molecule ligand.

### Future perspective

On the one hand, studies are providing fundamental new insights into topological constraints and how they vary with RNA secondary structure. As a result, we are now in a much better position to appreciate why a given RNA secondary structure conserves aspects of topology such as length of helices and asymmetry of junctions. Likewise, we are in a better position to make and interpret

Figure 6



Size-encoded adaptation of RNA by conformational selection using small molecules. **(a)** Inter-helical orientation maps of small-molecule bound RNA crystal structures for HIV DIS kissing dimers (2FCX, 2FCY, 2FCZ, 2FD0) according to solvent accessible surface area (SAS) of the ligand. Predicted junction-allowed topological distributions (in grey) for corresponding two-way junctions of type 2-1 in the 'free' state are shown in the cube plot. Shown to the right are correlation plots between each small molecule's SAS and the corresponding bound RNA's inter-helical angles ( $\alpha_h, \beta_h, \gamma_h$ ) and  $K_d$ . **(b)** X-ray structures of two-way RNA junction from (a) are shown bound to their small molecules in order of increasing size. Aminoglycosides are shown in red.

mutations. Although we are far from it, it seems that a *complete* and *quantitative* understanding of junction behavior is within our reach. The major challenges are to better understand, and indeed predict, sequence-specific interactions involving junction residues and flanking base-pairs, and of course, to properly take into account interactions involving the ion atmosphere and bound metals. Studies should pursue a deeper understanding of these basic interactions while also making interesting and bold predications about RNA conformation. Further studies are also needed to explore the potentially very significant role topological constraints can play in steering the co-transcriptional folding pathways of RNA.

On the other hand, studies are already taking advantage of topological constraints encoded by the RNA secondary structure to make predictions about 3D conformation and dynamic adaptation. Recent results show unequivocally that this approach has merit. Looking ahead, we can anticipate that low-resolution models of RNA backbone conformation will routinely be depicted, alongside secondary structure, regardless of its size and complexity. Such models will provide a starting framework for interpreting data and for informing study designs. These approaches will likely be the first to unveil new principles of RNA architecture that manifest at much larger length scales and to provide the first insights into higher-order genomic structure. A major challenge will be to bridge the scale back to the atomic level. Here, advances in RNA

atomic-resolution structure prediction will undoubtedly play an important role.

But perhaps the most important and gratifying aspect of topological constraints is that we can look at a given RNA secondary structure and better understand, and indeed appreciate, how evolutionary pressures on folding, structure, and dynamic adaptation led to its selection.

### Acknowledgements

AMM acknowledges support from a Graduate Research Fellowship from the National Science Foundation. CLB acknowledges support for the Center for Multi-scale Modeling Tools in Structural Biology (MMTSB) from the NIH through grant number RR012255. HMA acknowledges support NSF CAREER award (MCB 0644278) and an NIH grant (R01GM089846).

### References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Murthy VL, Srinivasan R, Draper DE, Rose GD: **A complete conformational map for RNA.** *J Mol Biol* 1999, **291**:313-327.
  2. Musselman C, Pitt SW, Gulati K, Foster LL, Andricioaei I, Al-Hashimi HM: **Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings.** *J Biomol NMR* 2006, **36**:235-249.
  3. Riordan FA, Bhattacharyya A, McAteer S, Lilley DM: **Kinking of RNA helices by bulged bases, and the structure of the human immunodeficiency virus transactivator response element.** *J Mol Biol* 1992, **226**:305-310.

4. Lilley DM, Clegg RM, Diekmann S, Seeman NC, Von Kitzling E, Hagerman PJ: **A nomenclature of junctions and branchpoints in nucleic acids.** *Nucleic Acids Res* 1995, **23**:3363-3364.
5. Lilley DM: **Structures of helical junctions in nucleic acids.** *Q Rev Biophys* 2000, **33**:109-159.
6. Hsieh CH, Griffith JD: **Deletions of bases in one strand of duplex DNA, in contrast to single-base mismatches, produce highly kinked molecules: possible relevance to the folding of single-stranded nucleic acids.** *Proc Natl Acad Sci U S A* 1989, **86**:4833-4837.
7. Zacharias M, Hagerman PJ: **The influence of symmetric internal loops on the flexibility of RNA.** *J Mol Biol* 1996, **257**:276-289.
8. Zacharias M, Hagerman PJ: **Bulge-induced bends in RNA: quantification by transient electric birefringence.** *J Mol Biol* 1995, **247**:486-500.
9. Kim HD, Nienhaus GU, Ha T, Orr JW, Williamson JR, Chu S: **Mg<sup>2+</sup>-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules.** *Proc Natl Acad Sci U S A* 2002, **99**:4284-4289.
10. Al-Hashimi HM, Gosser Y, Gorin A, Hu W, Majumdar A, Patel DJ: **Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings.** *J Mol Biol* 2002, **315**:95-102.
11. Williamson JR: **Induced fit in RNA-protein recognition.** *Nat Struct Biol* 2000, **7**:834-837.
12. Lilley DM: **Structure, folding and mechanisms of ribozymes.** *Curr Opin Struct Biol* 2005, **15**:313-323.
13. Chu VB, Lipfert J, Bai Y, Pande VS, Doniach S, Herschlag D: **Do conformational biases of simple helical junctions influence RNA folding stability and specificity?** *RNA* 2009, **15**:2195-2205.  
A study combining molecular dynamic simulations and SAXS measurements on B-DNA duplexes linked by PEG tethers provides fundamental insights into the nature of topological constraints, their interplay with ion atmosphere, and their role in directing the specificity of tertiary interactions.
14. Bailor MH, Sun X, Al-Hashimi HM: **Topology links RNA secondary structure with global conformation, dynamics, and adaptation.** *Science* 2010, **327**:202-206.  
A simple topological model is used to rationalize the anisotropic confinement of inter-helical orientations across two-way junctions in the PDB as well as the differential adaptive changes in inter-helical orientations that follow binding of related ligands.
15. Bailor MH, Musselmann C, Hansen AL, Gulati K, Patel DJ, Al-Hashimi HM: **Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings.** *Nat Protoc* 2007, **2**:1536-1546.
16. Mustoe AM, Bailor M, Brooks III CL, Al-Hashimi HM: **On the role of topological constraints in defining RNA backbone conformation;** submitted for publication.
17. Lescoute A, Westhof E: **Topology of three-way junctions in folded RNAs.** *RNA-A* 2006, **12**:83-93 [Publication of the Rna Society].  
A survey of three-way junctions in the PDB shows that there is a limited set of global helix conformations adopted by these higher-order junctions.
18. Laing C, Schlick T: **Analysis of four-way junctions in RNA structures.** *J Mol Biol* 2009, **390**:547-559.
19. Laing C, Jung S, Iqbal A, Schlick T: **Tertiary motifs revealed in analyses of higher-order RNA junctions.** *J Mol Biol* 2009, **393**:67-82.
20. Zhang Q, Kim NK, Peterson RD, Wang Z, Feigon J: **Structurally conserved five nucleotide bulge determines the overall topology of the core domain of human telomerase RNA.** *Proc Natl Acad Sci U S A* 2010, **107**:18761-18768.  
Experimental characterization of the J2a/b five nucleotide bulge in human telomerase shows that the number of nucleotides in the bulge defines global structure and dynamics and is intimately linked to determining telomerase activity.
21. Reymond C, Levesque D, Bisailon M, Perreault JP: **Developing three-dimensional models of putative-folding intermediates of the HDV ribozyme.** *Structure* 2010, **18**:1608-1616.
22. Stelzer AC, Kratz JD, Zhang Q, Al-Hashimi HM: **RNA dynamics by design: biasing ensembles towards the ligand-bound state.** *Angew Chem Int Ed Engl* 2010, **49**:5731-5733.  
Changing the identity of a base-pair flanking a bulge junction biases the average inter-helical structure and dynamics towards a ligand bound state, resulting in higher affinity ligand binding.
23. Casiano-Negroni A, Sun X, Al-Hashimi HM: **Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition.** *Biochemistry* 2007, **46**:6525-6535.
24. Kilburn D, Roh JH, Guo L, Briber RM, Woodson SA: **Molecular crowding stabilizes folded RNA structure by the excluded volume effect.** *J Am Chem Soc* 2010, **132**:8690-8696.
25. Lambert D, Leipply D, Draper DE: **The osmolyte TMAO stabilizes native RNA tertiary structures in the absence of Mg<sup>2+</sup>: evidence for a large barrier to folding from phosphate dehydration.** *J Mol Biol* 2010, **404**:138-157.
26. Weeks KM: **Advances in RNA structure analysis by chemical probing.** *Curr Opin Struct Biol* 2010, **20**:295-304.
27. Das R, Karanicolas J, Baker D: **Atomic accuracy in predicting and designing noncanonical RNA structure.** *Nat Methods* 2010, **7**:291-294.
28. Parisien M, Major F: **The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data.** *Nature* 2008, **452**:51-55.
29. Hajdin CE, Ding F, Dokholyan NV, Weeks KM: **On the significance of an RNA tertiary structure prediction.** *RNA-A* 2010, **16**:1340-1349 [Publication of the Rna Society].  
Simulations show that inclusion of secondary structure constraints greatly improves the quality of RNA tertiary structure predictions.
30. Flores SC, Altman RB: **Turning limited experimental information into 3D models of RNA.** *RNA-A* 2010, **16**:1769-1778 [Publication of the Rna Society].  
These papers (see also Refs. [35\*,38]) produce accurate global tertiary structure predictions for RNAs as large as 160 nts utilizing secondary structure and <10 inferred tertiary structure constraints.
31. Das R, Baker D: **Automated de novo prediction of native-like RNA tertiary structures.** *Proc Natl Acad Sci U S A* 2007, **104**:14664-14669.
32. Das R, Kudaravalli M, Jonikas M, Laederach A, Fong R, Schwans JP, Baker D, Piccirilli JA, Altman RB, Herschlag D: **Structural inference of native and partially folded RNA by high-throughput contact mapping.** *Proc Natl Acad Sci U S A* 2008, **105**:4144-4149.  
The use of secondary structure and distance constraints obtained from cleavage experiments is able to produce an accurate global tertiary structure of the P4-P6 domain of the Tetrahymena ribozyme. The authors are also able to produce models of different structural intermediates using the same procedure at different salt conditions.
33. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV: **Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms.** *RNA* 2008, **14**:1164-1173.
34. Gherghe CM, Leonard CW, Ding F, Dokholyan NV, Weeks KM: **Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics.** *J Am Chem Soc* 2009, **131**:2541-2546.
35. Lavender CA, Ding F, Dokholyan NV, Weeks KM: **Robust and generic RNA modeling using inferred constraints: a structure for the hepatitis C virus IRES pseudoknot domain (vol. 49, p. 4931, 2010).** *Biochemistry* 2010, **49**:5968-5968.  
See annotation for Ref. [30\*].
36. Yang SC, Parisien M, Major F, Roux B: **RNA structure determination using SAXS data.** *J Phys Chem B* 2010, **114**:10039-10048.
37. Wang J, Zuo X, Yu P, Xu H, Starich MR, Tiede DM, Shapiro BA, Schwieters CD, Wang YX: **A method for helical RNA global structure determination in solution using small-angle X-ray scattering and NMR measurements.** *J Mol Biol* 2009, **393**:717-734.

38. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB: **Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters.** *RNA-A* 2009, **15**:189-199 [Publication of the Rna Society].
39. Wang JB, Zuo XB, Yu P, Xu H, Starich MR, Tiede DM, Shapiro BA, Schwieters CD, Wang YX: **A method for helical RNA global structure determination in solution using small-angle X-ray scattering and NMR measurements.** *J Mol Biol* 2009, **393**:717-734.
40. Lipfert J, Ouellet J, Norman DG, Doniach S, Lilley DMJ: **The complete VS ribozyme in solution studied by small-angle X-ray scattering.** *Structure* 2008, **16**:1357-1367.
41. Devkota B, Petrov AS, Lemieux S, Boz MB, Tang L, Schneemann A, Johnson JE, Harvey SC: **Structural and electrostatic characterization of pariacoto virus: implications for viral assembly.** *Biopolymers* 2009, **91**:530-538.
42. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM: **Architecture and secondary structure of an entire HIV-1 RNA genome.** *Nature* 2009, **460**:711-716.
43. Deigan KE, Li TW, Mathews DH, Weeks KM: **Accurate SHAPE-directed RNA structure determination.** *Proc Natl Acad Sci U S A* 2009, **106**:97-102.
44. Thirumalai D, Lee N, Woodson SA, Klimov D: **Early events in RNA folding.** *Annu Rev Phys Chem* 2001, **52**:751-762.
45. Karanicolas J, Brooks CL III: **Integrating folding kinetics and protein function: biphasic kinetics and dual binding specificity in a WW domain.** *Proc Natl Acad Sci U S A* 2004, **101**:3432-3437.
46. Hills RD Jr, Brooks CL III: **Subdomain competition, cooperativity, and topological frustration in the folding of CheY.** *J Mol Biol* 2008, **382**:485-495.
47. Hills RD Jr, Kathuria SV, Wallace LA, Day IJ, Brooks CL III, Mathews CR: **Topological frustration in beta alpha-repeat proteins: sequence diversity modulates the conserved folding mechanisms of alpha/beta/alpha sandwich proteins.** *J Mol Biol* 2010, **398**:332-350.
48. Cho SS, Pincus DL, Thirumalai D: **Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures.** *Proc Natl Acad Sci U S A* 2009, **106**:17349-17354.
49. Lin JC, Thirumalai D: **Relative stability of helices determines the folding landscape of adenine riboswitch aptamers.** *J Am Chem Soc* 2008, **130**:14080-14081.  
Application of coarse-grained Gō models to adenine riboswitches demonstrates that ligand binding can stabilize a meta-stable secondary structure long enough to commit it to a given folding pathway.
50. Feng J, Walter NG, Brooks CL III: **Cooperative and directional folding of the preQ<sub>1</sub> riboswitch aptamer domain.** *J Am Chem Soc* 2011, **133**:4196-4199.
51. Whitford PC, Schug A, Saunders J, Hennelly SP, Onuchic JN, Sanbonmatsu KY: **Nonlocal helix formation is key to understanding S-adenosylmethionine-1 riboswitch function.** *Biophys J* 2009, **96**:L7-9.  
All-atom Gō model simulations show that helix formation can act as a conformational switching mechanism. Kinetic simulations also reveal that the formation of secondary structures can result in topological frustration where neighboring helical elements must first unfold to allow assembly of downstream helices.
52. Frank AT, Stelzer AC, Al-Hashimi HM, Andricioaei I: **Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition.** *Nucleic Acids Res* 2009, **37**:3670-3679.
53. Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY, Batey RT: **Free state conformational sampling of the SAM-I riboswitch aptamer domain.** *Structure* 2010, **18**:787-797.
54. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM: **Visualizing spatially correlated dynamics that directs RNA conformational transitions.** *Nature* 2007, **450**:1263-1267.
55. Stelzer AC, Frank AT, Bailor MH, Andricioaei I, Al-Hashimi HM: **Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs.** *Methods* 2009, **49**:167-173.
56. Toor N, Keating KS, Taylor SD, Pyle AM: **Crystal structure of a self-spliced group II intron.** *Science* 2008, **320**:77-82.