

## Structural bioinformatics

# Fast, clash-free RNA conformational morphing using molecular junctions

Amélie Héliou<sup>1,\*</sup>, Dominik Budday<sup>2</sup>, Rasmus Fonseca<sup>3,4</sup>  
and Henry van den Bedem<sup>4,\*</sup>

<sup>1</sup>LIX, Ecole Polytechnique, CNRS, Inria, Université Paris-Saclay, 91128, Palaiseau, France, <sup>2</sup>Chair of Applied Dynamics, University of Erlangen-Nuremberg, Erlangen, Germany, <sup>3</sup>Department of Molecular and Cellular Physiology, Stanford University, Stanford, CA, USA and <sup>4</sup>Biosciences Division, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA 94025, USA

\*To whom correspondence should be addressed.

Associate Editor: Cenk Sahinalp

Received on September 22, 2016; revised on February 12, 2017; editorial decision on February 27, 2017; accepted on March 11, 2017

## Abstract

**Motivation:** Non-coding ribonucleic acids (ncRNA) are functional RNA molecules that are not translated into protein. They are extremely dynamic, adopting diverse conformational substates, which enables them to modulate their interaction with a large number of other molecules. The flexibility of ncRNA provides a challenge for probing their complex 3D conformational landscape, both experimentally and computationally.

**Results:** Despite their conformational diversity, ncRNAs mostly preserve their secondary structure throughout the dynamic ensemble. Here we present a kinematics-based procedure to morph an RNA molecule between conformational substates, while avoiding inter-atomic clashes. We represent an RNA as a kinematic linkage, with fixed groups of atoms as rigid bodies and rotatable bonds as degrees of freedom. Our procedure maintains RNA secondary structure by treating hydrogen bonds between base pairs as constraints. The constraints define a lower-dimensional, secondary-structure constraint manifold in conformation space, where motions are largely governed by molecular junctions of unpaired nucleotides. On a large benchmark set, we show that our morphing procedure compares favorably to peer algorithms, and can approach goal conformations to within a low all-atom RMSD by directing fewer than 1% of its atoms. Our results suggest that molecular junctions can modulate 3D structural rearrangements, while secondary structure elements guide large parts of the molecule along the transition to the correct final conformation.

**Availability and Implementation:** The source code, binaries and data are available at <https://simtk.org/home/kgs>.

**Contact:** [amelie.heliou@polytechnique.edu](mailto:amelie.heliou@polytechnique.edu) or [vdbedem@stanford.edu](mailto:vdbedem@stanford.edu)

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

## 1 Introduction

Cellular activity depends on molecular interactions. While proteins are generally considered the workhorses of the cell, it is increasingly recognized that noncoding ribonucleic acids (ncRNA) play a major role in a variety of biological processes (Esteller, 2011; Mercer and Mattick, 2013). Like proteins, their function is closely related to

their three-dimensional (3D) structure; an RNA molecule can adopt a multitude of conformations to interact with other molecules and perform its function (Kim *et al.*, 2014; Leulliot and Varani, 2001; van den Bedem and Fraser, 2015). Their highly flexible nature often demands representation with a dynamic ensemble of structures (Shi *et al.*, 2016).

Despite the significance of RNA's conformational diversity for function, computational techniques to explore the 3D conformational landscape and examine structural transitions between substates remains relatively underdeveloped (Boniecki *et al.*, 2015). The majority of efforts in 3D computational modeling of RNA is directed at structure prediction (Cruz *et al.*, 2012; Das *et al.*, 2010; Dawson and Bujnicki, 2016; Magnus *et al.*, 2014). Sampling-based strategies to explore conformational space can provide rapid access to native states (Frellsen *et al.*, 2009; Parisien and Major, 2008), or even sparsely populated substates (Fonseca *et al.*, 2014). For large RNA molecules, hierarchical approaches greatly increase the efficiency and conformational diversity of samples (Fonseca *et al.*, 2014; Sim *et al.*, 2012).

Few methods exist to morph between conformational substates separated by a significant energy barrier. Linear interpolation (Flores *et al.*, 2006; Krebs and Gerstein, 2000) and Elastic Network Models (ENM) (Castellana *et al.*, 2013; López-Blanco *et al.*, 2014) are efficient morphing methods, but ignore non-native, transient interactions. Targeted molecular dynamics (TMD) simulation (Noy *et al.*, 2007) adds a distance restraint to the energy function to cross energy barriers, but is prone to get stuck in local minima. Umbrella sampling (Yildirim *et al.*, 2013) can overcome this limitation and additionally report on kinetics, but requires substantial computational resources.

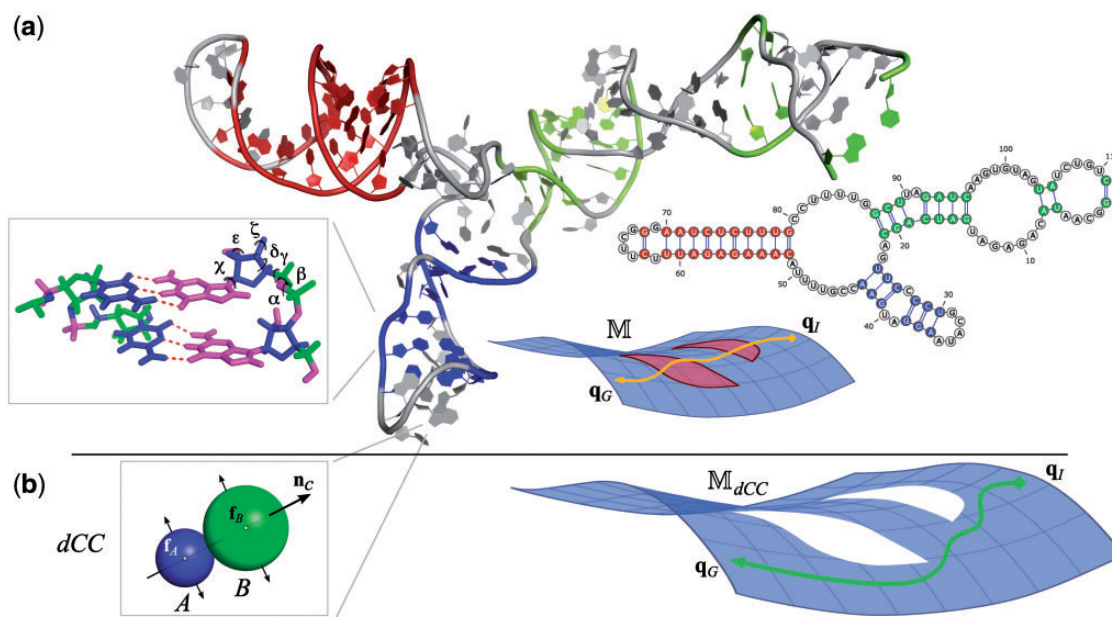
RNA secondary structure is determined by helical stems, which are linked by single-stranded motifs of unpaired nucleotides, e.g. bulges, internal loops and higher order molecular junctions (Fig. 1a). The junctions are highly flexible, in part owing to the lack of Watson–Crick pairing, and play a critical role in determining RNA tertiary structure (Bailor *et al.*, 2010; Butcher and Pyle, 2011; Laing *et al.*, 2013; Mustoe *et al.*, 2012, 2014)). For example, by promoting co-axial stacking of helical domains (Cruz and Westhof, 2009). Importantly, RNA small molecule binding

partners preferentially target these interhelical junctions (Thomas and Hergenrother, 2008), suggesting additional, functional roles for these motifs.

This emerging view of flexible junctions modulating tertiary structure suggests an intimate relation between RNA secondary structure and 3D conformational ensembles (Fig. 1a). However, to what degree sequence and structure of junctions modulate helical arrangements, and how rearrangements affect binding partners remains poorly understood.

Here, we exploit the insight that flexible molecular junctions modulate tertiary structure to design an efficient morphing procedure for determining clash-free transition pathways between conformations. In previous work, we encoded RNA molecules as a kinematic linkage with groups of atoms as links and rotatable bonds as joints (Fig. 1a, Fonseca *et al.* (2014, 2015)). Watson–Crick pairs are the main determinants of secondary structure in RNA. We include non-covalent interactions between Watson–Crick pairs as explicit constraints in the structural representation of RNA, which define a lower-dimensional *secondary-structure constraint manifold*  $\mathbb{M}$  in conformation space (Fig. 1a). This secondary-structure constraint manifold is exactly the molecule's 3D accessible conformation space when secondary structure is maintained. Flexibility on the secondary-structure constraint manifold is largely governed by bulges, loops, and higher order junctions. Thus, conformational transitions on the manifold can rapidly explore the relationship between junction sequence, structure and 3D helical arrangement.

Our contributions are twofold. First, we designed an efficient, all-atom algorithm to compute motions of conformational change *directly* and *exactly* in the secondary structure constraint manifold between two conformers. Second, we ensured that the motions avoid clashes between any pair of atoms, including hydrogens. For each pair of near-colliding atoms, we introduce a temporary 1D distance constraint between them. These dynamic, Clash-avoiding



**Fig. 1.** (a) 3D and 2D representations of the yeast U2/U6 snRNA complex. Helices are color-coded and flexible single stranded regions are shown in gray. The inset on the left shows one nucleotide's degrees of freedom, and the subdivision into individually colored rigid bodies. Red dashed lines show hydrogen bonds of Watson–Crick interactions. The blue plane represents the secondary-structure constraint manifold,  $\mathbb{M}$ , with red patches corresponding to sterically unfeasible regions. Moving directly from the initial conformation,  $q_I$ , to the goal,  $q_G$ , will frequently enter colliding regions. (b) The inset shows two colliding atoms, resulting in a dynamic Clash-avoiding Constraint (dCC) that prevents A and B from moving in closer contact, but allows motion in directions orthogonal to  $n_c$  or jointly along  $n_c$ . The clash regions are absent in the resulting lower-dimensional manifold  $\mathbb{M}_{dCC}$  (cut-out), and the path between  $q_I$  and  $q_G$  is clash-free (Color version of this figure is available at *Bioinformatics* online.)

Constraints (dCC) instantaneously redirect the motion onto a (locally) 'clash-free', lower-dimensional submanifold of the secondary-structure constraint manifold (Fig. 1b). Thus, energetics in molecular transitions are modulated by constant bonded interactions (bond lengths and angles) and a hard-core, all-atom potential for non-bonded interactions.

The reduced dimension of the constraint manifold enables efficient exploration of high-dimensional molecular conformation spaces, while the dynamic, clash-avoiding constraints help navigate complex energy landscapes. This leads naturally to a hierarchical morphing procedure where initial, large conformational changes are guided by a small number of 'marker atoms', while the final stage takes all heavy atoms into account. This is important, as many experimental techniques, including Nuclear Magnetic Resonance (NMR) Residual Dipolar Couplings (RDC) or Double Electron Electron Resonance (DEER) spectroscopy provide the position or interatomic orientation or distance between only a handful of atoms. Our technique can help determine all-atom conformations and collective motions from those sparse experiments: Structural ensemble representations are frequently obtained with the help of conformational sampling and subsequent fitting, starting from, for example, a crystal structure (Fonseca *et al.*, 2014; Pachov *et al.*, 2016). Transitions between the substates can report directly on structural rearrangements and molecular mechanisms (Lipfert *et al.*, 2007; Zhang *et al.*, 2007), which, in turn, could inform rational design principles for RNA-targeting ligands (Bernat and Disney, 2015).

In the remainder we first derive expressions for clash-free conformational transitions in the constraint manifold. We evaluated our procedure on a large benchmark set of 78 RNA molecules determined by NMR spectroscopy. In a direct comparison with a state-of-the-art Normal Mode Analysis (NMA)-based morphing procedure for RNA (iMOD, (López-Blanco *et al.*, 2014)), we found that our procedure approached the goal conformation closer in heavy atom Root Mean Square Deviation (RMSD) than iMOD for all but one out of 78 cases. Next, we tested how sparsity of structural information affects our procedure by using only one marker atom for every five residues as goal positions. Strikingly, while the marker atoms represent fewer than 1% of the molecule, 67% of conformations in the benchmark set improved their *heavy atom* RMSD to the target by at least 20%. This suggests that the constraint manifold is an extremely efficient reduction of conformation space. We found that junction conformation only weakly correlates with helical arrangements. Their elevated conformational variability could serve to accommodate multiple binding partners.

Finally, we applied our procedure to the *Synechococcus elongatus* L-glutamine riboswitch, which undergoes a large transition between its free and ligand binding conformations (Ren *et al.*, 2015).

## 2 Materials and methods

We obtain an RNA conformational transition by balancing three antagonistic objectives along the pathway: maintaining the secondary structure, reaching the goal conformation of marker atoms and avoiding inter-atomic clashes. We implemented the procedure in our KGS (Kino-Geometric Sampling) framework (Budday *et al.*, 2015; Yao *et al.*, 2012). The first step of our procedure aligns the initial and the goal conformation. Next, we deform the molecule using the algorithm described below.

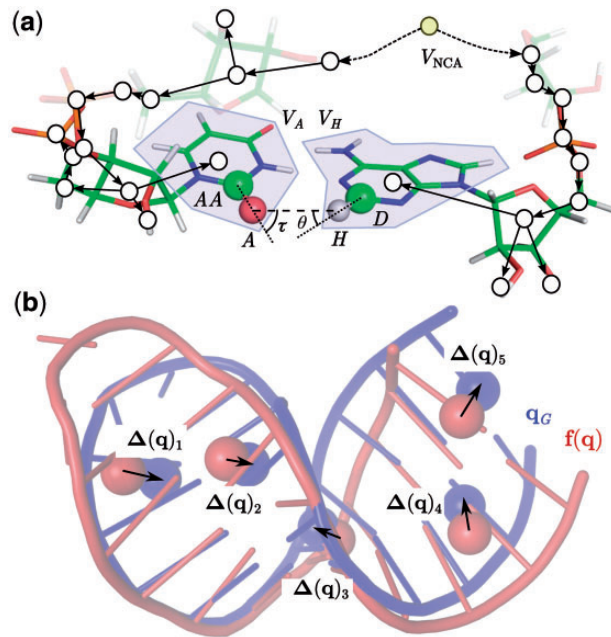
### 2.1 Tree representation

We start from an all-atom, structural representation of RNA molecules that includes hydrogen atoms. We assume that bond lengths and angles are fixed at their initial values. The rotatable covalent bonds ( $\alpha, \beta, \gamma, \delta, \epsilon, \zeta$ ) of the RNA backbone and the  $\chi$  angle around the glycosidic bond are the dihedral degrees of freedom (Fig. 1a). The  $\delta$  dihedral angle is part of the ribose ring, which modulates puckered conformations.

As in previous KGS applications, we use a molecular geometry-based rigidity-analysis approach (Budday *et al.*, 2015; Fonseca *et al.*, 2014, 2015). Despite their simplified representation, similar, but graph-based rigidity-analysis algorithms (Fulle and Gohlke, 2008) have, for example, predicted thermodynamic properties of biomolecules with remarkable accuracy (Radestock and Gohlke, 2011). We represent an RNA molecule by a rooted, directed spanning tree  $T(V, E)$ . Each vertex of  $V$  is a rigid body, i.e. a group of atoms linked by non-rotatable bonds. An edge  $E$  is a rotatable bond, i.e. a degree of freedom (Fig. 2). A molecule is completely specified by an  $n$ -dimensional vector,  $\mathbf{q} \in \mathbb{T}^n$ , where  $n$  denotes the cardinality of  $E$ . Each coordinate of  $\mathbf{q}$  corresponds to the rotation angle around the corresponding edge. Non-covalent interactions are represented by constraints.

### 2.2 Cycle constraints

Our goal is to transform an initial RNA conformation into a goal conformation without disturbing its secondary structure. The secondary structure of RNA molecules is determined by the hydrogen bonding network between donor and acceptor atoms, contributed by pairs of nucleotides. We include Watson-Crick pairs, which form



**Fig. 2.** (a) Part of a kinematic tree, with rigid bodies shown as circles and edges as arrows. A Watson-Crick pair is highlighted, where  $V_A$  is the rigid body of the acceptor and acceptor base atoms (A and AA) and  $V_H$  is the rigid body of the hydrogen and donor atom (H and D). The common ancestor ( $V_{NCA}$ ) of  $V_A$  and  $V_H$  is in yellow. Angles  $\theta$  and  $\tau$  are between the covalent and hydrogen bond axes. (b) 3D representation of an intermediate ( $f(\mathbf{q})$ , tan) conformation and the goal conformation (blue). From the goal conformation,  $\mathbf{q}_G$ , only the positions of marker atoms (blue spheres) are used to drive the transition. Black arrows indicate directions  $\Delta(\mathbf{q})_i$  from the intermediate to the goal position of marker atom  $i$  (Color version of this figure is available at *Bioinformatics* online.)

stable interactions from  $-6$  to  $-4$  kcal/mol (Vendeix *et al.*, 2009) as constraints to our representation. For C-G base pairs, we include hydrogen bonds O6-N4, N1-N3 and N2-O2. For A-U base pairs, we include N6-O4 and N1-N3. These non-covalent interactions form closed kinematic cycles, which require coordinated changes of the dihedral angles to maintain cycle closure or can even completely rigidify dihedral angles. Note that while we limit constraints to Watson-Crick base pair interactions in this study, any interaction (e.g. distance constraints) between arbitrary pairs of atoms can be included. Constraints can be excluded too.

Let  $V_A$  be the vertex of the acceptor and  $V_H$  the vertex of the hydrogen atom of a hydrogen bond, with  $V_{NCA}$  denoting the common ancestor rigid body of  $V_A$  and  $V_H$  in the tree. Maintaining the hydrogen bond requires adjusting the degrees of freedom between  $V_{NCA}$  and  $V_A$  (for example, along the 'left' branch of the cycle) and  $V_{NCA}$  and  $V_H$  (along the 'right' branch of the cycle) in a coordinated fashion. We briefly summarize the resulting constraint equations also described in Budday *et al.* (2015). Let  $\mathbf{f} = \mathbf{f}(\mathbf{q}) \in \mathbb{R}^3$  be the forward endpoint map for the hydrogen atom  $H$  and the acceptor atom  $A$  with respect to the left ( $L$ ) and right ( $R$ ) branch of the cycle from their common ancestor rigid body (Fig. 2a), then

$$\frac{1}{2} \left[ \left( \mathbf{f}_H^L + \mathbf{f}_A^L \right) - \left( \mathbf{f}_H^R + \mathbf{f}_A^R \right) \right] = 0, \quad (1)$$

where  $\frac{1}{2}(\mathbf{f}_H + \mathbf{f}_A)$  is the midpoint of the hydrogen bond. Note that (1) corresponds to three constraints for the Cartesian coordinates of the midpoint. We further constrain the relative orientation of coordinate frames at  $H$  and  $A$  by keeping angles  $\theta$  and  $\tau$  between the bond axis and adjacent covalent bonds constant (Fig. 2a). In practice, slight fluctuations in  $\theta$  and  $\tau$  occur owing to our first order model. Let  $\mathbf{f}_D$  and  $\mathbf{f}_{AA}$  denote the forward maps for the donor atom  $D$  and the acceptor base atom  $AA$ , then

$$(\mathbf{f}_A^R - \mathbf{f}_H^R)^T (\mathbf{f}_H^L - \mathbf{f}_D^L) - c_{\theta,ini} = 0 \quad (2)$$

$$(\mathbf{f}_H^L - \mathbf{f}_A^L)^T (\mathbf{f}_A^R - \mathbf{f}_{AA}^R) - c_{\tau,ini} = 0 \quad (3)$$

where

$$c_{\theta,ini} = (\mathbf{f}_{A,ini}^R - \mathbf{f}_{H,ini}^R)^T (\mathbf{f}_{H,ini}^L - \mathbf{f}_{D,ini}^L) \quad (4)$$

$$c_{\tau,ini} = (\mathbf{f}_{H,ini}^L - \mathbf{f}_{A,ini}^L)^T (\mathbf{f}_{A,ini}^R - \mathbf{f}_{AA,ini}^R) \quad (5)$$

are constants determined by the initial configuration. Note that rotation around the  $H$ -to- $A$ -axis is permitted.

The instantaneous consistency condition  $\mathbf{J}\dot{\mathbf{q}} = 0$  ensures that constraints hold at all times, where  $\mathbf{J}$  is the  $5m \times n$  Jacobian matrix of all  $m$  hydrogen bond constraints. Thus, admissible velocities satisfy  $\dot{\mathbf{q}} \in \ker \mathbf{J}$ , where  $\dim \ker \mathbf{J} = n - r$  and  $r$  the rank of  $\mathbf{J}$ . Generally, both over- and underconstrained scenarios are possible, i.e.  $5m < n$  or  $5m > n$ , with  $r \leq \min(5m, n)$ . In practice, however, hydrogen bonds often add redundant constraints, such that  $r < n$  leading to  $n - r$  remaining internal degrees of freedom in the nullspace. We obtain a basis  $\mathbf{N} \in \mathbb{R}^{n \times (n-r)}$  for the nullspace of  $\mathbf{J}$  from the right-singular vectors with vanishing singular values in the SVD decomposition  $\mathbf{J} = \mathbf{U}\Sigma\mathbf{V}^T$ .

A vector  $\dot{\mathbf{u}} \in \mathbb{R}^{n-r}$  of generalized velocities leads to an admissible velocity  $\dot{\mathbf{q}} \in \ker \mathbf{J}$  via

$$\dot{\mathbf{q}} = \mathbf{N}\dot{\mathbf{u}}, \quad (6)$$

which does not violate the constraints. Finally, we observe that perturbing a conformation  $\mathbf{q}$  with a vector  $\delta\mathbf{q} \in \ker \mathbf{J}$ , i.e.  $\mathbf{q}' = \mathbf{q} + \delta\mathbf{q}$

will maintain the hydrogen bonds and cycles, provided  $|\delta\mathbf{q}|$  is sufficiently small. The Jacobian matrix and its nullspace are updated at every step.

## 2.3 Directed conformational change in the secondary structure constraint manifold

Minimizing the RMSD distance between an initial and a target conformation without constraining degrees of freedom would instantly violate an RNA's secondary structure. Furthermore, iteratively applying these conformational changes would nearly inevitably lead to clashes between atoms. We therefore derive equations to reach a target conformation from an initial conformation using only perturbations in clash avoiding submanifolds of the constraint manifold.

### 2.3.1 Minimizing distance with generalized velocities in the nullspace

The target conformation is represented by the coordinates of any set of  $k$  given marker atoms  $A_i^G, i = 1, \dots, k$ . Let  $\Delta$  be the  $3k$ -dimensional vector of the directions between the marker atom positions in conformation  $\mathbf{q}$  and in the target conformation (Fig. 2b).

$$\Delta(\mathbf{q}) = \begin{pmatrix} A_1^G - \mathbf{f}(\mathbf{q})_{A_1} \\ \vdots \\ A_k^G - \mathbf{f}(\mathbf{q})_{A_k} \end{pmatrix}. \quad (7)$$

An infinitesimal perturbation  $\delta\mathbf{q}$  of a conformation  $\mathbf{q}$  leads to a displacement of atom positions  $\delta\mathbf{f}_{A_i} = \sum_{j=0}^{n-1} \frac{\partial \mathbf{f}_{A_i}}{\partial \mathbf{q}_j} \delta\mathbf{q}_j$ . We define  $\mathbf{M} \in \mathbb{R}^{3k \times n}$  as follows:

$$\mathbf{M} = \begin{pmatrix} \frac{\partial \mathbf{f}_{A_1}}{\partial q_0} & \frac{\partial \mathbf{f}_{A_1}}{\partial q_1} & \cdots \\ \vdots & \vdots & \ddots \\ \frac{\partial \mathbf{f}_{A_k}}{\partial q_0} & \frac{\partial \mathbf{f}_{A_k}}{\partial q_1} & \cdots \end{pmatrix}. \quad (8)$$

Let  $\mathbf{M}\delta\mathbf{q}$  denote an infinitesimal displacement of the marker atoms. Our objective is to obtain a new conformation with marker atoms as close as possible to their goal positions. This is equivalent to minimizing the difference between the displacement of marker atoms  $\mathbf{M}\delta\mathbf{q}$  and  $\Delta(\mathbf{q})$ , or

$$\min_{\delta\mathbf{q}} \|\mathbf{M}\delta\mathbf{q} - \Delta(\mathbf{q})\|_2, \quad (9)$$

at each step. To satisfy constraints,  $\delta\mathbf{q}$  should also be an admissible perturbation, i.e.  $\delta\mathbf{q} = \mathbf{N}\delta\mathbf{u}$ . Substituting into (9) we obtain:

$$\min_{\delta\mathbf{u}} \|\mathbf{MN}\delta\mathbf{u} - \Delta(\mathbf{q})\|_2. \quad (10)$$

A minimum norm solution to (10) is given by  $\delta\mathbf{u}_s = (\mathbf{MN})^\dagger \Delta(\mathbf{q})$ , where  $(\mathbf{MN})^\dagger$  is the pseudo-inverse. While the minimizing direction, i.e. the infinitesimal perturbation along  $\delta\mathbf{u}_s$  is calculated exactly, our first-order approximation to the constraint manifold dictates an iterative approach along this direction using small steps. This adds a complication, as  $\mathbf{M}$ ,  $\mathbf{N}$ ,  $\Delta(\mathbf{q})$  and therefore  $\delta\mathbf{u}_s$  must be updated each iteration.



## 2.4 Is the goal conformation reachable in the secondary-structure constraint manifold?

Interestingly, (10) contains information whether the goal conformation can be reached exactly, given the constraints. The Rouché-Capelli theorem in linear algebra gives

$$\begin{aligned} \min_{\delta \mathbf{u}} \|\mathbf{MN}\delta \mathbf{u} - \Delta(\mathbf{q})\|_2 = 0 &\iff \\ \text{rank}([\mathbf{MN}|\Delta(\mathbf{q})]) &= \text{rank}(\mathbf{MN}). \end{aligned} \quad (11)$$

The matrix  $[\mathbf{MN}|\Delta(\mathbf{q})]$  is  $\mathbf{MN}$  augmented with the vector  $\Delta(\mathbf{q})$  as a last column. Note that  $\text{rank}([\mathbf{MN}|\Delta(\mathbf{q})]) - \text{rank}(\mathbf{MN})$  is either 1 or 0 because the two matrices differ by only one column. A difference of 0 means that the added column, the distance to goal, is a linear combination of the other columns, and therefore it can be reached. If (11) is satisfied, a solution that reaches the goal exactly exists in the tangent space to the constraint manifold. If it is not satisfied, a least squares approximation in the tangent space is the optimal solution. Clearly, this depends on the selection of marker atoms. Therefore, we can generally distinguish two different scenarios:

1. Goal positions are consistent with the constraints (i.e. Watson-Crick pairs and fixed length of covalent bonds and angles). The condition (11) is satisfied and the goal positions are reachable.
2. The condition (11) is not satisfied. The optimal solution is a least-squares approximation in the tangent space. However, if we remove constraints, the dimension of the nullspace generally increases, adding independent columns to  $\mathbf{MN}$ , with an opportunity to satisfy (11). Alternatively, if we reduce the number of goal positions, removing rows,  $\mathbf{MN}$  is more likely to be full row rank, i.e.  $\text{rank}(\mathbf{MN}) = m$ , again with an opportunity to satisfy (11).

## 2.5 Clash-avoiding conformational change

We define a clash between a pair of atoms whenever the distance between their centers is less than the sum of their Van der Waals radii, scaled by a parameter  $c_f$ . To avoid a clash between a pair of atoms  $A$  and  $B$ , we redirect their motion by adding temporary, one-dimensional constraints to conformation  $\mathbf{q}$  according to

$$\mathbf{n}_c^T \left( \frac{\partial \mathbf{f}_B}{\partial \mathbf{q}} - \frac{\partial \mathbf{f}_A}{\partial \mathbf{q}} \right) \dot{\mathbf{q}} = 0. \quad (12)$$

The constraint ensures that the joint motion along a direction  $\mathbf{n}_c = \|\mathbf{p}_B - \mathbf{p}_A\|$  through the atom centers  $\mathbf{p}_A$  and  $\mathbf{p}_B$  maintains their distance along  $\mathbf{n}_c$  (Fig. 1b). Note that the constraint also permits independent motion of  $A$  and  $B$  in the plane perpendicular to  $\mathbf{n}_c$ . The temporary, additional constraints define a submanifold of the secondary constraint manifold; motions in this submanifold avoid clashes. Each perturbation of a conformation starts without temporary constraints, and they are added as necessary. We call the additional constraints dynamic Clash-avoiding Constraints (dCC); each adds one row to the Jacobian matrix  $\mathbf{J}$ .

## 2.6 Putting it all together: clash-free conformational morphing

We used secondary-structure-constraint-manifold-modulated conformational changes and clash-avoiding moves in KGS to implement an RNA conformational morphing procedure (Supplementary Fig. S1). The procedure starts with the initial  $\mathbf{q}_I$  and goal  $\mathbf{q}_G$  conformations. Constraints between Watson-Crick pairs are determined by RNaView (Yang *et al.*, 2003). Next, a minimizing perturbation  $\delta \mathbf{u}$  from  $\mathbf{q}_0 = \mathbf{q}_I$  to  $\mathbf{q}_G$  is computed from (10) and scaled to  $\|\delta \mathbf{u}\|_\infty = 0.01$  to obtain a sufficiently small step size that maintains hydrogen

bonds. We perturb  $\mathbf{q}_0$  to obtain  $\mathbf{q}'_1 = \mathbf{q}_0 + \delta \mathbf{q}$ , and check for clashes. The following two cases can occur:

1. The new conformation  $\mathbf{q}'_1$  does not have clashes. It is accepted into the sampling pool.
2. The new conformation  $\mathbf{q}'_1$  has clashes. We do not accept  $\mathbf{q}'_1$ , but instead append clash-avoiding constraints to the Jacobian. We repeat the procedure to compute a new conformation  $\mathbf{q}''_1$ , and iterate a fixed number of times. If, in rare cases, the same atoms are still clashing, but no additional clashes occurred, the dCC is violated owing to our first order approximation, and the desired direction is sterically inaccessible. Therefore, we perform a random move in  $\mathbb{M}_{\text{dCC}}$  instead of moving towards marker atom goal positions. Note that clash constraints cannot be added indefinitely, since they gradually rigidify the molecule.

Finally, a new clash-free conformation  $\mathbf{q}_1$  is accepted, the clash constraints are removed and a new optimal perturbation is computed starting from  $\mathbf{q}_1$ . Unless otherwise stated, the procedure is repeated until the RMSD distance to the goal is less than 0.5 Å, or 0.01 Å when we omit clash avoidance. The Van der Waals scaling parameter is fixed at  $c_f = 0.6$ , allowing some degree of overlap. Stereochemical constraints are often relaxed in conformational sampling, balancing computational speed and structure quality (Bender *et al.*, 2016).

## 2.7 Benchmark dataset

We selected a dataset of RNA molecules from the Protein Data Bank (PDB, Berman *et al.*, 2000) with one chain and between 15 and 200 nucleotides in length (Supplementary Table S1). All were solved by Nuclear Magnetic Resonance (NMR) spectroscopy, and all had more than one model with identical atoms. RNA molecules with nucleotides other than A, U, C and G were excluded. We selected the first model as the initial conformation, and selected the model with the highest RMSD (with Biopython (Cock *et al.*, 2009)) from the first model as the goal conformation. We limited our dataset to RNAs with an RMSD between initial and goal greater than 2 Å. This resulted in a dataset of 78 multi-model molecules, with average heavy atom RMSD between the initial and goal conformations of 4.14 Å (maximum RMSD 28.80 Å, minimum RMSD 2.01 Å).

## 3 Results and discussion

Avoiding clashes markedly complicates simulating conformational transitions. We therefore first examined the performance of our procedure for computing a conformational transition between substates on the secondary structure constraint manifold while ignoring clashes for the 78 RNA structures in our dataset. For comparison, we used the conformational morphing module of iMOD (López-Blanco *et al.*, 2014). iMOD minimizes the RMSD between all pairs of heavy atoms in the initial and goal conformation, also ignoring clashes.

We then evaluated the performance for computing a clash-avoiding conformational transitions, using a hierarchical approach. We first selected a small subset of atoms as marker atoms to guide the initial to the goal conformation, and further optimize the conformational transition using all heavy atoms.

Next, we used our procedure to understand to what degree bulges and higher-order junctions govern 3D helical arrangement, and, finally, we applied our algorithm to a challenging example; conformational exchanges of an L-glutamine riboswitch (Ren *et al.*, 2015).

### 3.1 KGS approaches goal conformation closer than iMOD

The iMOD morphing tool uses torsion angle normal mode analysis (NMA) to iteratively minimize the distance between all pairs of heavy atoms in the initial and goal conformations. We used the iMOD default parameter settings, except that we selected to use 90% instead of 10% of modes, since that generally led to lower RMSD to the goal conformation. Conformational transitions computed with iMOD are not restricted to the secondary structure constraint manifold. Instead, all non-bonded atoms are interconnected by harmonic springs within a 10 Å radius (López-Blanco *et al.*, 2011). iMOD does not check for clashes, but its Elastic Network Model limits the introduction of new clashes for small amplitude motions. To perform a direct comparison with iMOD we disabled clash avoidance in KGS, while instructing it to minimize the distance between all pairs of heavy atoms. We executed KGS and iMOD transitions on each of the 78 RNA molecules. We terminated KGS after 1000 iterations or when the heavy-atom RMSD was under 0.01 Å.

KGS dramatically reduced the RMSD to the goal conformation. On average by 86%, from 4.14 Å RMSD to 0.51 Å, compared to 69% for iMOD (Fig. 3). That represents an average improvement of 0.55 Å RMSD of KGS over iMOD. Strikingly, for nearly every RNA molecule, 77 out of 78 in the benchmark dataset, KGS approached the goal conformation closer in heavy atom RMSD than iMOD.

In many cases, KGS performed significantly better than iMOD. For example, the initial RMSD for the highly dynamic 23-nucleotide ID3 stem loop of domain 1 of the ai5y group II intron (PDB ID 2M12), is 3.98 Å. KGS came within 0.55 Å of the goal conformation, 1.64 Å closer than iMOD. In particular, iMOD struggled to fit the large amplitude motions of hairpin loop bases that do not participate in Watson–Crick interaction.

In one out of 78 cases iMOD approached the target conformation 0.29 Å closer than KGS (PDB ID 1QC8). However, KGS had not converged within 1000 steps, and did get closer than iMOD in 2000 steps.

### 3.2 Clash-free conformational transitions

As helical structure is mostly preserved between RNA substates, we expected that the dimensionality reduction encoded in our

secondary structure constraint manifold would be effective in guiding large fragments of the molecule to their goal conformation. We exploited this insight to avoid over-constraining the system, adopting a hierarchical approach for clash-avoiding transitions. Equation (11) signifies a small number of marker atoms may be an effective strategy to reach the goal conformation. We selected the C5 atom of every fifth nucleotide to guide the initial to the goal conformation. In NMR experiments,  $^1\text{DC5H}_5$  RDCs are often collected. The C5 atoms therefore emulate a sparse experimental dataset – fewer than 1% of the total number of atoms. Then, starting from the best conformation at this stage, in a second stage we selected all heavy atoms as markers to further reduce the distance to the goal conformation.

#### 3.2.1 Distance to marker atom goal positions using every fifth C5 atom

We first evaluated the RMSD between the final marker atom positions and their corresponding goal positions. Out of the 78 RNA structures in our dataset, 70 (90%) reached the goal positions of marker atoms extremely close, to within 0.5 Å (Table S1). For the remaining eight RNA molecules that did not reach their goal positions to within 0.5 Å, clashes could not be fully resolved, resulting in an excess of random moves.

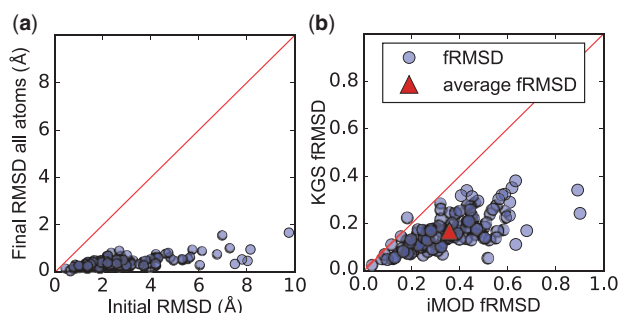
#### 3.2.2 Maintaining secondary structure between substates

We then verified that KGS properly maintained helical structure between clash-avoiding conformational transitions. Our mathematical model maintains distance constraints exactly by calculating admissible velocities in the constraint manifold (Section 2.3). However, the finite step size of the perturbation respects distance constraints only to first order approximation. We therefore monitored deformation of the hydrogen bond distances in conformational transitions. In 91% of the RNA molecules the maximal hydrogen bond length deformation was below 10% (Fig. 4a). These values are well within the range of expected fluctuations, and suggest that our conformational transitions stay close to the secondary structure constraint manifold. This is important as breaking Watson–Crick base pairs is energetically unfavorable, making it less likely to occur in a trajectory.

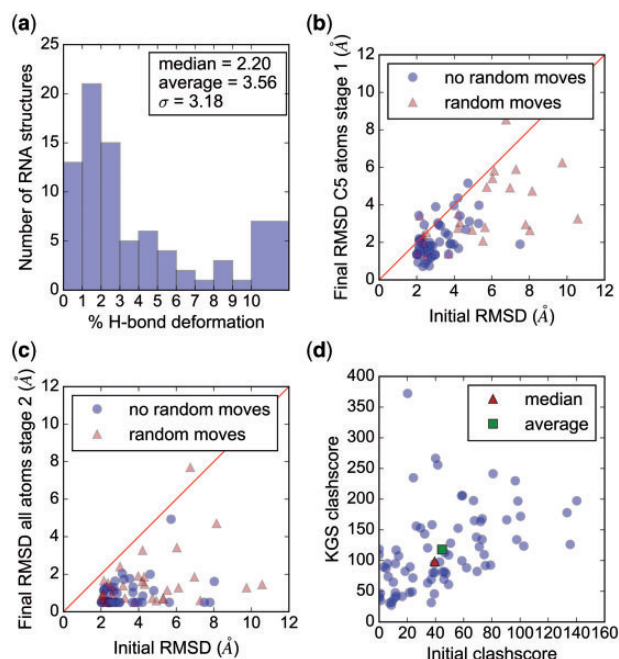
An analysis with DSSR (Lu *et al.*, 2015) confirmed that A-form helices were maintained along the trajectory. Out of 1275 Watson–Crick base pairs among the initial conformations, 486 were classified as A-form, 6 as B-form and 783 could not be classified (Supplementary Table S2). For the KGS final conformation, 440 were A-form, 7 B-form and 840 could not be classified. Just over 9% of Watson–Crick pairs could no longer be classified as A-form, despite maintaining their constraints. Two pairs picked up A-form classification. We did not observe A-form to B-form or reverse transitions.

#### 3.2.3 Distance to the heavy atom goal positions

To understand to what extent RNA secondary structure guides their all-atom 3D structure, we calculated the heavy atom RMSD between the best conformation determined by the marker goal positions and the (true) goal conformation. Figure 4b shows that KGS morphing significantly reduced the heavy atom RMSD to the goal conformation. Strikingly, using fewer than one percent of atoms as markers, the average heavy atom RMSD from the best KGS conformation to the goal conformation was 2.83 Å, compared to 4.14 Å between the initial and goal conformations. 66% of the benchmark set improved the heavy atom RMSD by at least 20%. In some cases, the improvement is dramatic. For example, the heavy atom RMSD between the initial (tan) and goal (blue) conformation



**Fig. 3.** Conformational transitions on the secondary structure constraint manifold, using all heavy atoms as markers but ignoring clashes. (a) The initial RMSD in Å (x-axis) versus the final RMSD (y-axis) to the goal conformation obtained by KGS. Data points below the red diagonal indicate improvement towards the goal conformation. (b) Comparison of the performance of iMOD and KGS. The x-axis shows the fractional RMSD (fRMSD: the RMSD between the final conformation and the goal conformation divided by the RMSD between initial and goal conformations) obtained by iMOD. The y-axis shows the corresponding ratio for KGS. Data points below the red diagonal indicate a lower fRMSD and superior performance for KGS than for iMOD. The red triangle indicates the average fRMSD value over all 78 data points (Color version of this figure is available at *Bioinformatics* online.)



**Fig. 4.** Hydrogen bond deformation and heavy atom RMSD minimization in KGS clash-free conformational transitions. (a) Distribution of largest acceptor-to-hydrogen distance deformation for WC-constraints in the benchmark set. (b) Heavy atom RMSD improvement for morphing with clash-avoidance and one C5 marker atom every five residues (first stage). Triangles indicate RNAs where morphing required random moves. (c) Heavy atom RMSD improvement after switching to all heavy atoms (second stage). PDB ID 1S9S is omitted. It has an initial RMSD of 28.80 Å, its first stage RMSD is 21.20 Å and its second stage RMSD is 20.17 Å. (d) Distribution of Molprobit clashscore for final KGS structures, compared to the initial structure (Color version of this figure is available at *Bioinformatics* online.)

for the 62-nucleotide VS ribozyme II-III-VI three-way junction is reduced from 7.50 Å to 1.90 Å (Supplementary Fig. S2). The final marker atoms RMSD is less than 0.01 Å.

We then selected all heavy atoms as marker atoms to further refine the distance to the goal conformation. We used an RMSD threshold of 0.5 Å. The final, average RMSD to the goal conformation was reduced to 1.47 Å (from 4.14 Å, fRMSD = 0.34). Figure 4c and Supplementary Table S1 summarize the results. By contrast, computing the conformational transition using heavy atoms as markers directly resulted in a final RMSD of 1.90 Å, corresponding to an fRMSD of 0.41. With the hierarchical approach, 64% of the benchmark set improved the heavy atom RMSD by at least 60% compared to 53% using heavy atoms directly. To evaluate the quality of the final KGS structures, we calculated the change in the Molprobit clashscore compared to the initial structure (Chen *et al.*, 2010). The Molprobit clashscore counts atoms per one thousand that overlap by more than 0.4 Å, which is a more stringent criterion than  $c_f = 0.6$ . We found a moderate rise in clashes; on average just shy of a threefold increase (Fig. 4d).

Our results suggest that the constraint manifold efficiently reduces dimensionality of conformation space, while retaining key structural information that is largely accessible through a sparse set of marker atoms. A hierarchical approach for conformational morphing capitalizing on this insight is more efficient than a direct approach.

### 3.2.4 Intra-junction motions modulate 3D helical arrangements

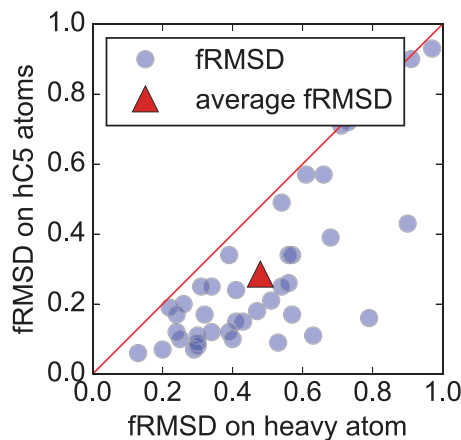
Next, to probe the role of bulges and higher-order junctions in 3D helical arrangement, we used only 'helical' C5 (hC5) atoms as

markers, i.e. all C5 atoms in Watson–Crick base pairs. The hC5 initial RMSD reports on differences in helical arrangement between conformations of our benchmark set. We selected a subset of 41 structures from our benchmark set for which helices of the initial and goal conformations were separated by at least 2 Å. The average hC5 RMSD is 4.28 Å. Clash-avoiding morphing reduced the hC5 RMSD to, on average, 1.69 Å, corresponding to an fRMSD of 0.29. Note that these clash-free motions take place on the secondary structure constraint manifold  $\mathcal{M}_{\text{dCC}}$ , which is precisely the accessible conformation space when helical structures are preserved. The small values of helical C5 fRMSD therefore suggest that single-stranded motifs play a significant role in modulating 3D helical arrangements through intra-junction, coordinated motions on  $\mathcal{M}_{\text{dCC}}$ .

However, hC5 helical arrangement leaves a substantial fraction of heavy atom fRMSD unaccounted for. To examine the flexibility of the single-stranded motifs, we also calculated the overall heavy atom RMSD to the goal conformation. We found that, using only the hC5 atom markers to rearrange helices, in half of the cases in the benchmark set the overall heavy atom RMSD was reduced by at least 57%. This suggests that conformational variability of intra-junction regions remained elevated for a substantial fraction of the benchmark set, and that their conformations only weakly correlate with helical arrangements (Fig. 5). Junctions often bind ligands and are responsible for most function in RNA. While they modulate 3D helical arrangements, these regions must remain flexible to exchange between conformational substates and interact with multiple partners.

### 3.3 Conformational exchange of riboswitch *glnA*

Riboswitches are non-coding domains of messenger RNA, often located in untranslated regions, that bind to a partner and regulate transcription and/or translation (Tucker and Breaker, 2005). The L-glutamine-binding riboswitch *glnA* is a three-stem junction, with helices P1, P2 and P3. L-glutamine-binding riboswitches are implicated in nitrogen metabolism (Ames and Breaker, 2011), and found exclusively in cyanobacteria and marine metagenomic sequences. In a 61 nucleotide construct of the *Synechococcus elongatus glnA* sensing domain, binding of L-glutamine results in a large, 13.11 Å conformational change from a 'tuning-fork'-shaped, free conformation to an 'L'-shaped bound conformation (Ren *et al.*, 2015). The bound



**Fig. 5.** Conformational transitions using all hC5 (C5 atoms in Watson-Crick base pairs) atoms as markers with clash-avoidance. The distribution of heavy atoms fRMSD (x-axis) versus hC5 fRMSD (y-axis) suggest heavy atoms fRMSD is only weakly correlated to 3D helical arrangement. The red triangle indicates the average fRMSD value over all 41 data points (Color version of this figure is available at *Bioinformatics* online.)



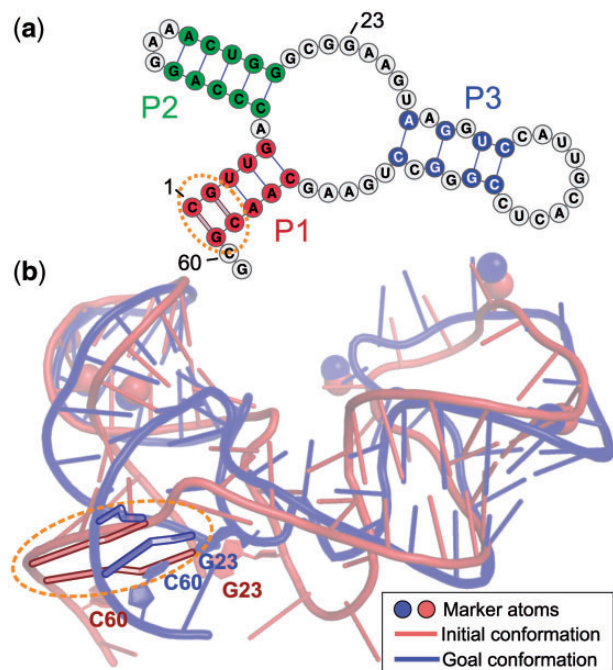
form is stabilized by a long-distance base pair G23-C60. In the absence of L-glutamine, the riboswitch samples the free conformation and a minor conformation similar to the bound conformation. The exchange between the free and ligand-bound conformations is accompanied by partial melting of the P1 helix, which may be required to access the bound conformation.

To test the significance of partial P1 melting for exchanging between substates, we computed clash-free conformational transitions with and without its two dissociating Watson–Crick base pairs (C1–G59 and G2–C58) as constraints (Fig. 6a).

Once more, we took a hierarchical approach. First we selected six C5 atoms as markers, three each in the apical loops of helices P2 and P3 (Fig. 6b). Using these marker atoms, the goal marker positions were reached to within 0.01 Å with and without C1–G59 and G2–C58 as constraints. The heavy atom fRMSD was 0.52 when C1–G59 and G2–C58 were included as constraints, and 0.47 when they were not.

For the second stage, we selected all C5 atoms (one per nucleotide) as markers. When C1–G59 and G2–C58 were omitted as constraints, our procedure found a clash-free conformational transition to within 1.73 Å heavy atom RMSD of the bound conformation. Bases G23 and C60 approach each other in this less-constrained transition, suggesting a base-pair could form. By contrast, when C1–G59 and G2–C58 were included, the RMSD remained elevated at 4.44 Å. Interestingly, these constraints were among the most deformed during the transition, with strain exceeding 5%.

Our results suggest that partial melting of P1 is required to adopt the ligand-bound state. While *glnA* samples a minor conformation similar to the bound conformation, binding of L-glutamine may lower the free-energy barrier to make the bound state accessible.



**Fig. 6.** (a) Secondary structure representation of the L-glutamine riboswitch. Helix P1 is shown in red; the two Watson–Crick pairs C1–G59 and G2–C58 involved in partial melting are encircled in orange. (b) Superposition of the 3D conformations of the L-glutamine riboswitch. The initial conformation (tan) is the unbound state while the goal conformation (blue) is the bound state. The marker atoms are shown in both conformations as spheres. An animation of the motion with partial melting of P1 can be found at <https://youtu.be/BhJVV0DXLR4> (Color version of this figure is available at Bioinformatics online.)

### 3.4 Run time

The run time of our procedure depends on the size of the molecule (Supplementary Fig. S3, top, and the number of constraints, bottom), as these values determine the size of the Jacobian matrix. While the singular value decomposition of the Jacobian is highly optimized by parallel solvers (Intel’s Math Kernel Library, MKL), it has a significant computational complexity and often dominates run time.

Without dCC, the vast majority of transitions finish within seconds (Supplementary Fig. S3, left). Adding dCC clash-avoidance, but keeping the size of the Jacobian small by using a sparse set of marker atoms results in reasonable run times for all systems (Supplementary Fig. S3, right). dCC clash-avoidance in combination with heavy atom marker atoms results in long run times, up to several hours for some RNA in our benchmark set.

## 4 Conclusions

We presented an efficient, computational procedure to obtain clash-free structural transition pathways between conformational sub-states of RNA. Our procedure is mathematically rooted in preserving RNA secondary structure by calculating *exact* admissible velocities in a lower-dimensional constraint manifold. The secondary structure constraint manifold helps guide the initial conformation to the goal conformation from few marker atoms. Stable tertiary motifs, e.g. kissing loops or pseudo-knots, that further govern conformational flexibility are trivially accommodated as constraints in our framework. Despite the near deterministic nature of the move sets, our algorithm performed extremely well. Nonetheless, coupling to more sophisticated motion planners to explore the conformational landscape (Novinskaya et al., 2015; Roth et al., 2016) will likely increase the performance, and result in a broader ensemble of transition pathways.

Why is it important to efficiently explore helical rearrangements? RNA receptors can accommodate a diverse set of small molecules binding to interhelical junctions (Thomas and Hergenrother, 2008), which often selectively stabilize helical arrangements (Bailor et al., 2010). Our results suggest that helical rearrangement poorly correlates with junction conformation, indicating functional roles for junctions beyond governing tertiary structure. Mapping sequence identity and conformational variability of junctions to helical arrangements and their interconversions can help us understand the molecular mechanisms of RNA–ligand interactions. Our algorithm is uniquely positioned to start addressing these important questions. In contrast to MD simulation, conformational sampling-based approaches can quickly interrogate the effect of junction sequence identities, conformations, mutations or insertions for thousands of structures.

Our L-glutamine-binding riboswitch *glnA* example suggests that certain transitions may require Watson–Crick pairs to break. Likewise, Watson–Crick pairs could (transiently) form. Our approach currently provides two mechanisms to facilitate evolving constraints: (i) Watson–Crick pairs can be excluded or included explicitly in the input files for the software, and (ii) In practice, we often find that many singular values of the constraint matrix are very small ( $\sim 10^{-4}$ ), but non-vanishing. Selecting right-singular vectors corresponding to these small, non-vanishing singular values as additional basis vectors for the ‘null-space’ allows for small violations of constraints. In fact, there is a direct correspondence between the magnitude of the singular value and the magnitude of the constraint violations it allows. The singular value cut-off is a



user-defined parameter for the software. The *glnA* application suggests small constraint violations occur even when working in the true nullspace, owing to first order approximation. Furthermore, dCC constraints could act like transient Watson–Crick pairs. While dCC constraints would unlikely result in fully formed Watson–Crick pairs, it is conceivable that a (one-dimensional) dCC forms between GC or AU pairs. Thus, monitoring constraint formations (dCC) or violations along a pathway could report on Watson–Crick pairs that form or break to facilitate conformational transitions.

Our procedure can have applications in helping to interpret sparse data, or exploring conformational landscapes of dynamic ensembles. Conformational transitions can be used as starting models for detailed, but expensive MD simulations. Adapting our procedure to rely on pair-wise distances between atoms instead of absolute distances will make it a valuable tool for interpreting, for example, DEER or FRET experiments.

**Conflict of Interest:** none declared.

## References

- Ames, T.D. and Breaker, R. (2011) Bacterial aptamers that selectively bind glutamine. *RNA Biol.*, **8**, 82–89.
- Bailor, M.H. *et al.* (2010) Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science*, **327**, 202–206.
- Bender, B.J. *et al.* (2016) Protocols for molecular modeling with Rosetta3 and RosettaScripts. *Biochemistry*, **55**, 4748–4763.
- Berman, H.M. *et al.* (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
- Bernat, V. and Disney, M.D. (2015) RNA structures as mediators of neurological diseases and as drug targets. *Neuron*, **87**, 28–46.
- Boniecki, M. *et al.* (2015) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.*, **44**, e63.
- Budday, D. *et al.* (2015) Geometric analysis characterizes molecular rigidity in generic and non-generic protein configurations. *J. Mech. Phys. Solids*, **83**, 36–47.
- Butcher, S.E. and Pyle, A.M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.*, **44**, 1302–1311.
- Castellana, N. *et al.* (2013) MORPH-PRO: a novel algorithm and web server for protein morphing. *Algorithms Mol. Biol.*, **8**, 19.
- Chen, V.B. *et al.* (2010) Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Cryst.*, **D66**, 12–21.
- Cock, P.J. *et al.* (2009) Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
- Cruz, J.A. and Westhof, E. (2009) The dynamic landscapes of RNA architecture. *Cell*, **136**, 604–609.
- Cruz, J.A. *et al.* (2012) RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*, **18**, 610–625.
- Das, R. *et al.* (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, **7**, 291–294.
- Dawson, W.K. and Bujnicki, J.M. (2016) Computational modeling of RNA 3D structures and interactions. *Curr. Opin. Struct. Biol.*, **37**, 22–28.
- Esteller, M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874.
- Flores, S. *et al.* (2006) The Database of Macromolecular Motions: new features added at the decade mark. *Nucleic Acids Res.*, **34**, D296–D301.
- Fonseca, R. *et al.* (2014) Characterizing RNA ensembles from NMR data with kinematic models. *Nucleic Acids Res.*, **42**, 9562–9572.
- Fonseca, R. *et al.* (2015). KGSrna: efficient 3D kinematics-based sampling for nucleic acids. In: Przytycka, T.M. (ed.) *Research in Computational Molecular Biology: 19th Annual International Conference, RECOMB 2015, Warsaw, Poland, April 12–15, 2015, Proceedings*, pp. 80–95. Springer International Publishing, Cham.
- Frellsen, J. *et al.* (2009) A probabilistic model of RNA conformational space. *PLoS Comput. Biol.*, **5**, e1000406.
- Fulle, S. and Gohlke, H. (2008) Analyzing the flexibility of {RNA} structures by constraint counting. *Biophys. J.*, **94**, 4202–4219.
- Kim, H. *et al.* (2014) Protein-guided RNA dynamics during early ribosome assembly. *Nature*, **506**, 334–338.
- Krebs, W.G. and Gerstein, M. (2000) The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res.*, **28**, 1665–1675.
- Laing, C. *et al.* (2013) Predicting helical topologies in RNA junctions as tree graphs. *PLoS One*, **8**, e71947.
- Leulliot, N. and Varani, G. (2001) Current topics in RNA–protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry*, **40**, 7947–7956.
- Lipfert, J. *et al.* (2007) Structural transitions and thermodynamics of a glycine-dependent riboswitch from *Vibrio cholerae*. *J. Mol. Biol.*, **365**, 1393–1406.
- López-Blanco, J.R. *et al.* (2011) iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics*, **27**, 2843–2850.
- López-Blanco, J.R. *et al.* (2014) imods: internal coordinates normal mode analysis server. *Nucleic Acids Res.*, **42**, W271–W276.
- Lu, X.-J. *et al.* (2015) DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.*, **43**, e142.
- Magnus, M. *et al.* (2014) Computational modeling of RNA 3D structures, with the aid of experimental restraints. *RNA Biol.*, **11**, 522–536.
- Mercer, T.R. and Mattick, J.S. (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.*, **20**, 300–307.
- Mustoe, A.M. *et al.* (2012) New insights into the fundamental role of topological constraints as a determinant of two-way junction conformation. *Nucleic Acids Res.*, **40**, 892–904.
- Mustoe, A.M. *et al.* (2014) Coarse grained models reveal essential contributions of topological constraints to the conformational free energy of RNA bulges. *J. Phys. Chem. B*, **118**, 2615–2627.
- Novinskaya, A. *et al.* (2015). Improving protein conformational sampling by using guiding projections. In: *IEEE Intl. Conf. on Bioinformatics and Biomedicine Workshops (BIBMW)*, pp. 1272–1279.
- Noy, A. *et al.* (2007) Theoretical study of large conformational transitions in DNA: The B→A conformational change in water and ethanol/water. *Nucleic Acids Res.*, **35**, 3330–3338.
- Pachov, D.V. *et al.* (2016) Coupled motions in  $\beta$  2AR:G  $\alpha$ s conformational ensembles. *J. Chem. Theory Comput.*, **12**, 946–956.
- Parisien, M. and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
- Radestock, S. and Gohlke, H. (2011) Protein rigidity and thermophilic adaptation. *Proteins Struct. Funct. Bioinf.*, **79**, 1089–1108.
- Ren, A. *et al.* (2015) Structural and dynamic basis for low-affinity, high-selectivity binding of l-glutamine by the glutamine riboswitch. *Cell Rep.*, **13**, 1800–1813.
- Roth, C.-A. *et al.* (2016) Hybridizing rapidly exploring random trees and basin hopping yields an improved exploration of energy landscapes. *J. Comput. Chem.*, **37**, 739–752.
- Shi, X. *et al.* (2016) The solution structural ensembles of RNA kink-turn motifs and their protein complexes. *Nat. Chem. Biol.*, **12**, 146–152.
- Sim, A.Y.L. *et al.* (2012) Modeling and design by hierarchical natural moves. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 2890–2895.
- Thomas, J.R. and Hergenrother, P.J. (2008) Targeting RNA with small molecules. *Chem. Rev.*, **108**, 1171–1224.
- Tucker, B.J. and Breaker, R.R. (2005) Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.*, **15**, 342–348.
- van den Bedem, H. and Fraser, J.S. (2015) Integrative, dynamic structural biology at atomic resolution – it’s about time. *Nat. Methods*, **12**, 307–318.
- Vendeix, F.A. *et al.* (2009) Free energy calculation of modified base-pair formation in explicit solvent: A predictive model. *RNA*, **15**, 2278–2287.
- Yang, H. *et al.* (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.
- Yao, P. *et al.* (2012) Sampling-based exploration of folded state of a protein under kinematic and geometric constraints. *Proteins Struct. Funct. Bioinf.*, **80**, 25–43.
- Yildirim, I. *et al.* (2013) A dynamic structural model of expanded RNA CAG repeats: A refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations. *J. Am. Chem. Soc.*, **135**, 3528–3538.
- Zhang, Q. *et al.* (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, **450**, 1263–1267.