nature REVIEWS MOLECULAR CELL BIOLOGY

SUPPLEMENTARY INFORMATION

In format as provided by the authors

The roles of structural dynamics in the cellular functions of RNAs

Laura R. Ganser, Megan L. Kelly, Daniel Herschlag and Hashim M. Al-Hashimi

https://doi.org/10.1038/s41580-019-0136-0

Supplementary Box 1: Experimental techniques used to determine various aspects of dynamic RNA ensembles.

Experimentally-informed RNA ensembles are crucial to understanding RNA function. Although the RNA ensemble contains millions of conformations, it is often possible to gain insights into biological function by visualizing a subset of conformers within local energetic minima along the free-energy landscape. The ability to model thermodynamic quantities such as the binding affinity between an RNA and partner molecule crucially depends on the quality of the ensemble in terms of how well related conformations can be resolved and assigned a specific population. The quality and information content of experimental data used in the ensemble determination will dictate the quality of the ensemble and what information it will provide. Here, we review experimental techniques that have been applied to generate RNA ensembles. Each technical approach provides unique types of structural information that report on dynamics at specific timescales. We discuss the pros and cons of each technique and provide examples of how the techniques have been applied to generate RNA ensembles. All the discussed techniques report on equilibrium dynamics; non-equilibrium time-resolved techniques that also provide information regarding RNA ensembles and free energy landscapes are not described here.

1. Chemical probing:

How it works: A chemical probe is used to chemically modify nucleotides in a structure-specific manner. Two major classes of chemical probes are typically used: (1) DMS, which attacks the base pairing face with reaction rates dependent on solvent exposure and hence single-stranded population; (2) SHAPE like reagents, which attack the 2'OH and thus depend on sugar flexibility (sugar pucker), which is again proportional to single-strandedness. Sites of chemical modification are detected by reverse transcription, either as a stop at the modification site or an insertion, deletion or point mutation in the cDNA. Sequencing can be used to locate where the RNA was modified, allowing calculation of the reactivity of each nucleotide. Typically, this data is fit to a single conformation, which can lead to biased results for dynamic RNAs. However, computational methods have been developed to fit structure-probing data to an ensemble of secondary structures rather than a single conformation[1,2].

Information provided: Chemical probing measures the reactivity of each nucleotide, which in turn depends on the extent of base pairing. The reactivity is an ensemble-weighted average over all structures of the RNA. Note that there are some caveats here — for SHAPE reagents, certain sugar puckers lead to very high reactivities. Such nucleotides are single-stranded, but

chemical reactivity per se is not purely related to base pairing in this case. Likewise for DMS, in certain structural contexts, the reaction rate seems to be almost catalyzed relative to pure single stranded nucleotides.

Timescale sensitivity: Fastest chemical probing has averaging timescale of 2-3 seconds (Benzyl cyanide). For the common 1M7 SHAPE reagent, averaging occurs on ~1 min timescale (Tier 0).

Pros: Data can be collected *in vivo* and at transcriptome-wide scales. Compared to other methods, it requires small amounts of sample (endogenous expression levels are frequently suitable).

Cons: Structure is never directly observed, rather inferred based on consistency with chemical reactivity profile. Thus, inferred models can be inaccurate or biased by computational algorithms. Provides limited information regarding 3D structure.

Example application: Computational methods, including structure landscape explorer and quantifier (SLEQ)[2] and Rsample[1] have been developed to fit structure-probing data to an ensemble of secondary structures rather than a single conformation. These methods were able to recover the structural heterogeneity of RNAs known to have multiple structures, including riboswitches.

2. Cryogenic electron microscopy (cryo-EM)

How it works: Electrons are accelerated through a specimen cooled at cryogenic temperatures and the scattered electrons are focused by the electromagnetic lenses of the microscope. Computational algorithms are used to separate small conformational changes in single molecule images using single particle analysis or sub-tomogram averaging techniques. Cryo-electron tomography, single-particle cryo-electron microscopy, and electron crystallography are different modalities of cryo-EM.

Information provided: Low energy conformations in the ensemble distribution are captured as snapshots of different conformational classes under solution conditions prior to flash freezing.

Timescale sensitivity: NA

Pros: Cryo-EM requires smaller amounts of sample relative to NMR and can be applied to large RNP complexes, virus particles, cells and tissue sections. The maximum observable object size is essentially only limited by the specimen thickness that can be penetrated with the electron beam.

Cons: Does not provide kinetic information regarding the rates of interconversion between different snapshots. Also, there is a lower molecular weight cutoff ~ 40 kDa and therefore few applications to non-coding RNAs that are not part of ribonucleoprotein complexes.

Example application: Ensemble analysis unveiled conformational states of the ribosome during translation that play essential roles in the decoding mechanisms[3]

3. Electron paramagnetic resonance (EPR)

How it works: EPR monitors transitions between the energy levels of unpaired electron spin(s). When coupled with site-directed spin labeling, in which spin labels containing stable unpaired electron(s) (most commonly a nitroxide) are attached at specific sites of the target molecule, EPR provides information such as rotational motions of a label (derived using continuous-wave EPR (cw-EPR)) and distances between pairs of labels (5 to 20 Å with cw-EPR or 20 to ~100 Å with pulsed EPR techniques).

Information provided: Conformations of parent macromolecule(s) from EPR measured inter-label distance distributions and/or pattern of individual label behaviors (e.g., nitroxide rotational dynamics), as well as site-specific dynamics information from label rotational motions and relaxation behaviors.

Timescale sensitivity: From nanoseconds to seconds (Tiers 0-2). Provides rate information and in some cases the order of conformational transitions.

Pros: Can be applied to high-molecular weight complexes in solution under physiological conditions using a small amount of samples. Distances can be measured from a pair of chemically identical spin labels, and a number of spin labels (e.g., nitroxides) are smaller and potentially less perturbing than most fluorophores. Measurements have been demonstrated *in vivo*.

Cons: Macromolecular information extracted from EPR measurement is influenced by the orientation and dynamics of spin labels with respect to the parent molecule, and various degrees of approximations often have to be made in data analyses and interpretation.

Example application: Investigation of nucleic acid-dependent conformational changes in CRISPR-Cas9[4].

4. Mutate-and-Map:

How it works: In mutate-and-map experiments, every nucleotide of the RNA of interest is mutated and its effect on the rest of the RNA is determined with chemical probing (Reviewed in [5]). Most mutations only cause a local effect on reactivity that can be used to determine the

pairing partner of that nucleotide. However, some mutations lead to large-scale reactivity differences that are proposed to reveal alternative secondary structures of the RNA. When integrated with computer modeling, these methods can provide an ensemble of secondary structures. REEFIT (RNA ensemble extraction from footprinting insights technique) was the first method developed for this purpose[6]. In mutate-map-rescue, discovery of additional compensatory mutations that rescue the effects of each originally perturbing single mutation provide tests of each base pair, including those present at low frequencies in alternative helices. In lock-mutate-map-rescue, a four-dimensional chemical mapping method, each candidate helix is locked through designed mutations and the rise or depletion of the base pairs of other helices are assayed by compensatory mutagenesis, allowing for a detailed dissection of allostery in highly complex secondary structure ensembles[7]. Going beyond secondary structure, MOHCA-seq (Multiplexed •OH Cleavage Analysis with paired-end sequencing) incorporates hydroxyl radical sources to the RNA and reads out cleavage from these radical at other positions, giving a map of nucleotide-nucleotide proximities at nanometer resolution; the method is sensitive to transient tertiary contacts even in RNA states that form heterogeneous tertiary ensembles[8].

Information provided: Chemical probing data for alternative structures accessed by single point mutations that map to an alternative secondary structure; validated helices and allosteric correlations between helices, from the multiple mutation approaches; and tertiary contacts in heterogenous ensembles from MOHCA-seq.

Timescale sensitivity: (see 'Chemical Probing').

Pros: Compensatory rescue enables independent validation of the "alternative structures".

Cons: No kinetic information.

Example application: A mutate-map-rescue approach uncovered a 20% populated alternative conformation of 16S rRNA accessible by a single nucleotide register shift[9]. This conformation agreed with crystallographic structures, suggesting that differences in conditions redistribute the population distribution of this RNA ensemble[9].

5. Nuclear magnetic resonance (NMR)

General Pros: Can be performed under a variety of solution conditions. Ultra high spatial-temporal resolution to local and global dynamics spanning twelve orders of magnitude in time; uniquely capable of visualizing low abundance conformational states.

General Cons: Typically limited to RNAs <70 nucleotides with applications becoming significantly more challenging for larger RNA. Resonance assignments and data collection and analysis can be time-consuming.

Chemical Shifts

How it works: The NMR chemical shift measured for individual ^{1}H, ^{13}C, ^{15}N, and ^{31}P nuclei depend on the local electronic distribution which in turn depends on local aspects of the structure.

Information provided: The chemical shifts are exquisitely sensitive to bond lengths, bond angles, dihedral angles, hydrogen bonding, protonation and tautomeric states as well as ring current effects arising due to circulation of π-electrons in the aromatic nucleobases, magnetic anisotropy, and other electrostatic effects. For a dynamic ensemble, the chemical shifts will correspond to a population weighted average over all conformations provided that the rates of interconversion between conformers is larger than the corresponding differences in chemical shifts.

Timescale sensitivity: Motionally averaged over timescales faster than milliseconds (tier 1 and 2). Also sensitive to slow tier 0 dynamics which results in the appearance of distinct resonances for the different conformational states.

Pros: Exquisitely sensitive to local electronic structure including ionization and tautomerization broadly across the base sugar and backbone. One of the easiest parameters to measured by solution NMR. Powerful approach to compare ensembles and assess ensemble redistribution.

Cons: Insensitive to global structure. The accuracy with which chemical shift data can be computed from 3D structure remains sub-optimal for determining ensembles, but can be used to test ensembles determined using other methods.

Example applications: Used to cross-validate and test ensembles of the HIV-1 TAR[10].

Residual dipolar couplings (RDCs)

How it works: Partial alignment (1 in 10^3-10^5 molecules are aligned) of RNA molecules in ordering media (typically filamentous bacteriophage[11,12]) results in incomplete averaging of dipolar couplings which manifests as splitting of resonances. The resulting residual dipolar coupling (RDC) depends on the cube of the distance separating the two nuclei, which is the bond length for directly bonded spins, as well as on the angle between the bond vector and molecule-fixed alignment tensor frame describing overall alignment of the RNA.

Information provided: Reports ensemble averaged trigonometric functions of the angle between bond vectors (typically directed bonded C-H and N-H vectors) and a molecule-fixed alignment tensor frame describing overall alignment of the molecule average over <millisecond timescales.

Timescale sensitivity: Motionally averaged over dynamics occurring <milliseconds. Sensitive to Tier 1 and 2 dynamics but not Tier 0.

Pros: Exquisitely sensitive to global and local aspects of ensembles. Broad sensitivity to timescales < milliseconds. Limited timescale sensitivity simplifies ensemble determination. RDCs can be accurately predicted for a given conformation based on its 3D structure.

Cons: Limited distance sensitivity. Does not provide timescale information.

Example application: Ensemble of apo- TAR RNA reveals conformations similar to those observed in seven distinct ligand bound states[10].

Relaxation dispersion and ZZ-exchange

How it works: Stochastic conformational transitions at the micro-to-second timescale result in dephasing of NMR magnetization (relaxation dispersion) or cross-peaks in 2D spectra (ZZ-exchange) in a manner dependent on the population and kinetic rates of the transition as well as on the chemical shift signatures of the different conformation sampled.

Information provided: The population and kinetic rate constants for conformational transitions involving more than two states as well as chemical shift signatures that report on the 3D structure of the states.

Timescale sensitivity: Depends on differences in chemical shifts and signal to noise but generally sensitive to motions occurring on the micro-to-second timescale.

Pros: Can be used to detect exceptionally low populated (populations as low as 0.01%) short-lived (lifetimes as low as a few microseconds) conformational states in RNA ensembles. Provides comprehensive kinetic, thermodynamic, and structural information at atomic resolution. Can be used to characterize complex kinetic networks and topologies.

Cons: Can be particularly time-consuming to collect and analyze data. Determining structure of the low-abundance state not always straightforward.

Example application: Studies of the fluoride riboswitch uncover a low-abundance short-lived transient state that exposes the RNA for strand invasion during co-transcriptional folding thus directing folding toward the OFF state.

Exact nuclear Overhauser effects (NOEs)

How it works: Quantitative measurements of buildup-derived NOEs across different short mixing times allows determination of accurate ensemble-averaged inter-proton distances. A sufficiently dense network of these distances allows determination of an ensemble of structures[13].

Information provided: Time-averaged inter-proton distances averaged over several milliseconds. The information content is complementary to angular restraints obtained from residual dipolar couplings and scalar couplings.

Timescale sensitivity: Less than several milliseconds (depends on the longest mixing time, typically ~200 ms).

Pros: Can be applied without any modifications to the sample and to unlabeled samples for small < 20 nucleotides. Yields rich density of data (several 10s per nucleotide).

Cons: Deviations from isotropic diffusion can complicate data analysis. Correction of spin diffusion is highly dependent on initial model. Distance averaging can be different depending on timescale of motion; therefore, timescale information needed to rigorously interpret data. Restricted to RNA size amenable to NMR spectroscopy.

Example applications: Used to determine a two-state ensemble for a 14-mer UUCG tetraloop[14].

6. RNA-MaP:

How it works: Leverages Illumina sequencing to allow massively parallel and quantitative measurements of the formation of $>10^6$ model RNA heterodimers referred to as 'tectoRNA'[15]. The tectoRNA 'host-guest' system is a heterodimer composed of two RNAs with two intermolecular tetraloop–tetraloop-receptor (TL–TLR) tertiary contacts and non-contacting RNA elements, both of which can be varied (see Fig 4b). RNA-MaP provides direct fluorescence-based binding curves and multiple replicates per chip gives high precision and rigorous statistics. As the same tertiary elements are present when various RNA elements such as helices and junctions are used, any difference in binding is a consequence of the ensemble properties of the inserted elements. Conversely, when exploring properties of tertiary elements, connecting helices and junctions are held constant.

Information provided: The likelihood of forming the tectoRNA assembly reflects the inherent ensemble properties of its constituent RNA elements (helices, junctions, and tertiary contacts) that compose the assembly. RNA-MaP therefore provides binding 'thermodynamic fingerprints' that can be used to compare and constrain models for the conformational ensembles of embedded elements.

Timescale sensitivity: Provides data on kinetics of binding which in principle also carry information regarding ensemble kinetics.

Pros: $>10^6$ measurements can routinely be carried out from a single sequencing chip on a time scale of <1 week. Can be used to obtain millions of thermodynamic measurements and

quantitative 'thermodynamic fingerprints' for a vast number of RNA helix, junction, and tertiary contact elements.

Cons: Deconvoluting ensemble properties from junctions and tertiary contacts may not always be straightforward. The tecto host-guest system is limited to specific RNA motifs and probes strained conformations; new host-guest systems are currently under development. Translating thermodynamic finerprints into ensembles requires a model for assembly energetics.

Example applications: Binding data has been used to obtain thermodynamic fingerprints for thousands of two-way HJH motifs and A-form helices[15]. This study exposed sequence-dependent ΔG^0_{redist} penalties involving redistribution of the A-form helix ensemble that can impact the energetics of RNA tertiary assembly[1]

7. Small angle X-ray scattering (SAXS)-based approaches

General Pros: Can be performed under a variety of solution conditions.

General Cons: Limited resolution for local changes in the ensemble. Does not provide kinetic information regarding the rates of interconversion.

SAXS

How it works: Elastic scattering of X-rays is measured at very low angles (typically 0.1º-10º), providing information about the overall shape and size of biomolecules that are 5 nm to 25 nm in size.

Information provided: Population-weighted distribution of distances for all atom-pairs within the biomolecule.

Timescale sensitivity: NA

Pros: Electron rich phosphate groups provide favorable contrast in nucleic acid applications.

Cons: Need to avoid aggregation. Calculations can be computationally expensive, and approximations are often introduced in order to accelerate the calculations.

Example application: Systematic studies of different riboswitches reveals diverse degrees of ensemble redistribution following ligand binding[16].

Au-SAXS

How it works: Provides precise distance distributions for gold nanocrystals that are attached at specific positions on the nucleic acid[17]. The scattering profile between two attached gold nanocrystals is isolated from the intra-nanocrystal, intra-nucleic acid, and nucleic acid-nanocrystal interference pattern by subtracting the scattering profile measured for isolated gold

nanocrystals, isolated nucleic acid, and singly-labeled nucleic acid from the scattering profile measured for the double-labeled nucleic acid.

Information provided: Distance distributions between gold nano-particles that are attached to the RNA.

Timescale sensitivity: NA

Pros: Precise determination of distance distribution. Multiple pairwise distance distribution can often be measured. Insensitivity to timescales of the motion can simplify determination of the thermodynamic ensemble.

Cons: Motions of the gold nanocrystal relative to the nucleic acid can influence the scattering profile and these contributions need to be minimized. Tethering of the linker to RNA itself should not affect the intrinsic dynamics.

Example application: Studies of the k-turn motif reveal ensemble redistribution following binding to protein L7Ae[18].

8. Single molecule Förster resonance energy transfer (smFRET)

How it works: FRET relies on measuring the transfer of energy from one donor fluorophore to another acceptor fluorophore in a distance dependent manner. FRET is observed when the fluorophores are within 2 to 10 nm. Changes in FRET signal can be used to identify dynamic transitions as well as calculate relative distances between the fluorophores based on the efficiency of the energy transfer.

Information provided: Distance information between fluorophores that are specifically attached to the RNA. Measures transitions within a single molecule and provides information about the underlying conformations and their rates of interconversion.

Timescale sensitivity: Allows access to timescales slower than 100-μs.

Pros: In bulk studies, the measured signal is averaged over an ensemble of molecules. This can make it difficult to deconvolute the kinetics and thermodynamics of multiple elementary steps. Single-molecule approaches such as smFRET can overcome these challenges, since the signal from individual molecules are analyzed. The approach also bypasses the need to synchronize molecules as is done in traditional bulk kinetic studies.

Cons: Data also depend on the orientation and dynamics of the fluorophores, and approximations often have to be made to extract distance information[19-21]. Need to rule out that fluorescent dyes perturb RNA ensemble.

Example application: smFRET was used to measure the thermodynamics and kinetics of TL–TLR tertiary contact formation in the P4-P6 domain of the *Tetrahymena* group I ribozyme providing support for ensemble modularity and the reconstitution model[22].

References:

1 Spasic, A., Assmann, S. M., Bevilacqua, P. C. & Mathews, D. H. Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic acids research* **46**, 314-323 (2018).

2 Li, H. & Aviran, S. Statistical modeling of RNA structure profiling experiments enables parsimonious reconstruction of structure landscapes. *Nature communications* **9**, 606 (2018).

3 Abeyrathne, P. D., Koh, C. S., Grant, T., Grigorieff, N. & Korostelev, A. A. Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome. *Elife* **5**, e14874 (2016).

4 Vazquez Reyes, C. *et al.* Nucleic acid-dependent conformational changes in CRISPR-Cas9 revealed by site-directed spin labeling. *Cell biochemistry and biophysics* **75**, 203-210 (2017).

5 Tian, S. & Das, R. RNA structure through multidimensional chemical mapping. *Quarterly reviews of biophysics* **49**, e7 (2016).

6 Cordero, P. & Das, R. Rich RNA structure landscapes revealed by mutate-and-map analysis. *PLoS computational biology* **11**, e1004473 (2015).

7 Tian, S., Kladwang, W. & Das, R. Allosteric mechanism of the *V. vulnificus* adenine riboswitch resolved by four-dimensional chemical mapping. *Elife* **7**, e29602 (2018).

8 Cheng, C. Y. *et al.* Consistent global structures of complex RNA states through multidimensional chemical mapping. *Elife* **4**, e07600 (2015).

9 Tian, S., Cordero, P., Kladwang, W. & Das, R. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA*, **20**, 1815-1826 (2014).

10 Salmon, L., Bascom, G., Andricioaei, I. & Al-Hashimi, H. M. A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: the basis for interhelical motions revealed. *Journal of the American Chemical Society* **135**, 5457-5466 (2013).

11 Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature structural biology* **5**, 1065-1074 (1998).

12 Clore, G. M., Starich, M. R. & Gronenborn, A. M. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod- shaped viruses. *Journal of the American Chemical Society* **120**, 10571-10572 (1998).

13 Vögeli, B. The nuclear Overhauser effect from a quantitative perspective. *Progress in nuclear magnetic resonance spectroscopy* **78**, 1-46 (2014).

14 Nichols, P. J. *et al.* High-resolution small RNA structures from exact nuclear Overhauser enhancement measurements without additional restraints. *Communications biology* **1**, 61 (2018).

15 Denny, S. K. *et al.* High-throughput investigation of diverse junction elements in RNA tertiary folding. *Cell* **174**, 377-390 (2018).

16 Zhang, J., Jones, C. P. & Ferre-D'Amare, A. R. Global analysis of riboswitches by small-angle X-ray scattering and calorimetry. *Biochimica et biophysica acta* **1839**, 1020-1029 (2014).

17 Mathew-Fenn, R. S., Das, R. & Harbury, P. A. Remeasuring the double helix. *Science* **322**, 446-449 (2008).

18 Shi, X., Huang, L., Lilley, D. M., Harbury, P. B. & Herschlag, D. The solution structural ensembles of RNA kink-turn motifs and their protein complexes. *Nature chemical biology* **12**, 146-152 (2016).

19 Bokinsky, G. & Zhuang, X. Single-molecule RNA folding. *Accounts of chemical research* **38**, 566-573 (2005).

20 Zhuang, X. Single-molecule RNA science. *Annu Rev Biophys Biomol Struct* **34**, 399-414 (2005).

21 Qin, P. Z. & Dieckmann, T. Application of NMR and EPR methods to the study of RNA. *Current opinion in structural biology* **14**, 350-359 (2004).

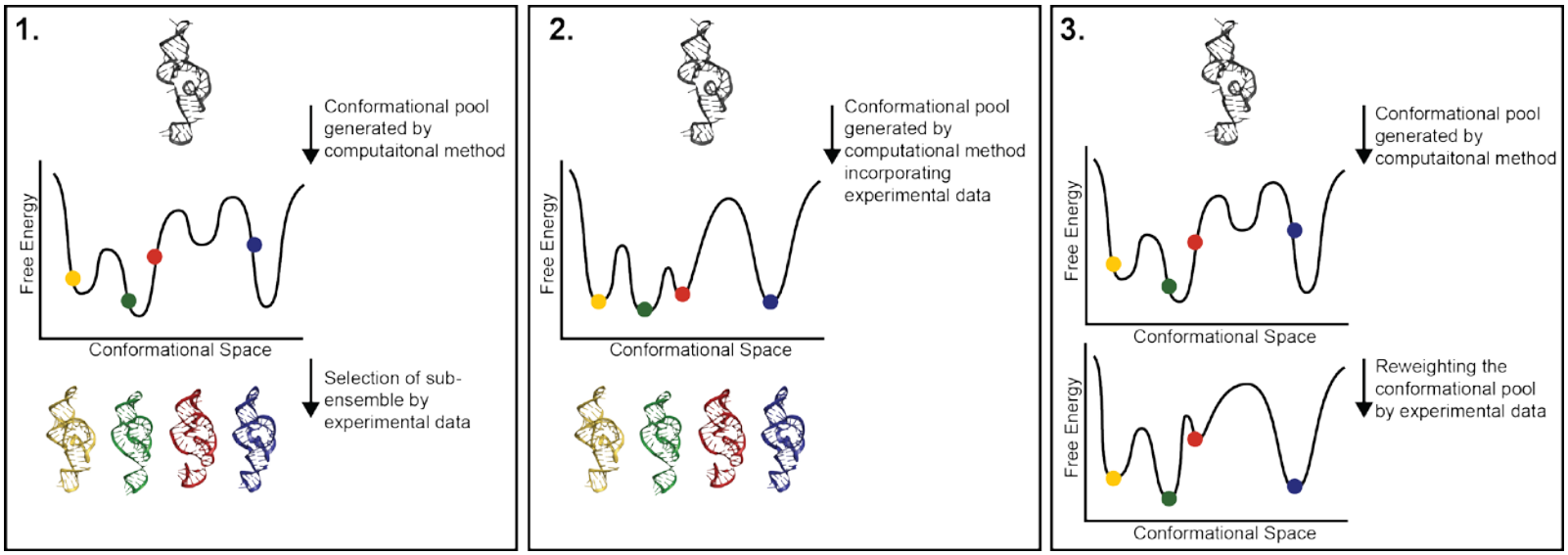
22 Bisaria, N. *et al.* Kinetic and thermodynamic framework for P4-P6 RNA reveals tertiary motif modularity and modulation of the folding preferred pathway. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E4956-4965 (2016).

Supplementary Box 2: Determining dynamic ensembles of RNA and computing the energy cost of re-distribution (ΔG^0_{redist})

Determining dynamic ensembles of RNA

Three general approaches (see the figure) have been used to construct atomic-resolution ensembles based on various types of methods (Supplementary Box 1). All approaches rely on some input from computational modeling or computational force fields to fill the 'data gap', as many more parameters are needed to specify an ensemble than to obtain a single structure. In one approach (see the figure, part 1), experimental data guides selection of conformations from a pool that is typically generated using computational methods, such as molecular dynamics simulations[1]. This increases the ease of integrating many different types of data, however, the quality of the ensemble is strongly dependent on the conformational pool. In a second approach (see the figure, part 2), experimental constraints are included as penalty functions or pseudo-energies in the conventional force field[2,3]. The experimental pseudo-potential can bias sampling towards conformations that may otherwise be disfavored by the force field. This approach is only practical with data that can be efficiently computed during the course of the trajectory. A disadvantage is that over-fitting of data can result in artifactual conformational deformation. Also, the experimental pseudo-potential can in some cases also induce non-physical perturbations to the free energy landscape. Finally, in a third approach (see the figure, part 3), the experimental data is used to minimally perturb a prior distribution (e.g. obtained from MD simulation) using either maximum entropy or Bayesian[4] approaches. This approach can use a variety of experimental inputs and yields high-resolution descriptions of the ensembles. However, the accuracy of the ensembles generated is highly dependent on the quality of the initial conformational pool.

When using all three approaches, it is very important to assess the uncertainty in the ensemble, because there are usually many different ensembles that can reproduce the experimental observables. Methods used to assess RNA ensembles include cross-validation by predicting data not used in the ensemble determination[1], using synthetic data to determine how well experimental data can define specific aspects of an ensemble[1,5], Monte Carlo based approaches to assess the accuracy and the precision of a given model or parameterization[1], and Bayesian approaches to assess ensemble uncertainty[4].



Computing ΔG^0_{redist}

Consider an RNA that transitions from ensemble **A** (free) to ensemble **B** (bound) upon binding to a partner ligand (L)

$$A + L \leftrightarrow BL$$

At equilibrium, the observed free energy of binding $\Delta G^{\circ}_{bind,obs}$ is given by

$$\Delta G^{\circ}_{bind,obs} = -RTln(K_d) = -RTln\left(\frac{(\frac{[BL]}{[BL]_0})}{(\frac{[A]}{[A]_0})(\frac{[L]}{[L]_0})}\right)$$

where K_d is the equilibrium constant for the binding process, R is the universal gas constant, T is the temperature of the system, and the quantities in the square brackets are the concentrations of the molecular species, with the subscript of 0 denoting the standard state.

The net binding reaction can be written as the sum of two constituent reactions involving (1) the redistribution of the ensemble **A** to ensemble **B** in the absence of L

$$A \leftrightarrow B$$

(2) the binding of L to ensemble B

$$B + L \leftrightarrow BL$$

with the associated free energy changes $\Delta G^{\circ}_{redist}$ and $\Delta G^{\circ}_{bind,B}$ respectively. Thus

$$\Delta G^{\circ}_{bind,obs} = \Delta G^{\circ}_{redist} + \Delta G^{\circ}_{bind,B}$$

$\Delta G^{\circ}_{bind,B}$ is driven by enthalpic and entropic contributions to the binding process. To determine the free energy term, $\Delta G^{\circ}_{redist}$, we can represent the conformational probability distribution of ensemble **A** as $p_A(x)$ for and $p_B(x)$ for ensemble **B**. In this notation, x is a geometric descriptor of the ensemble conformational space. However, we note that this is a simplification since, in principle, ensembles are distributions over multi-dimensional phase space (3N, where N is the number of atoms in the system). The energetic cost, $\Delta G^{\circ}_{redist}$, to re-weight or change the conformational probability distribution from $p_A(x)$ to $p_B(x)$ is given by $\Delta G^{\circ}_{redist}$[6]

$$\Delta G^{\circ}_{redist} = -RT \int p_B(x) ln\left(\frac{p_A(x)}{p_B(x)}\right) dx$$

Deriving an expression for $\Delta G^{\circ}_{redist}$requires knowledge of the changes in enthalpy and entropy accompanying ensemble redistribution. The entropy change for the conformational redistribution $\Delta S^{\circ}_{redist}$ is given by

$$\Delta S^{\circ}_{redist} = -R \int p_B(x) ln\big(p_B(x)\big) - p_A(x)\, ln\big(p_A(x)\big) dx$$

Assuming that the volume change accompanying a shift in conformational ensemble from **A** to **B** is negligible, the enthalpy change for the conformational transition $\Delta H^{\circ}_{redist}$is equal to $\Delta U^{\circ}_{redist}$ and is given by

$$\Delta H^{\circ}_{redist} = \Delta U^{\circ}_{redist} = U^{\circ}_B - U^{\circ}_A = \int (p_B(x) - p_A(x)) E_A(x) dx$$

From the definition of the population $p_A(x)$, we have

$$p_A(x) = e^{-(E_A(x)/RT)}/Z$$

where Z is the partition function. On substituting for $E_A(x)$ and solving for $\Delta H^{\circ}_{redist}$, we get

$$\Delta H^{\circ}_{redist} = \Delta U^{\circ}_{redist} = -RT \int (p_B(x) - p_A(x)) \ln(p_A(x)) dx$$

Thus,

$$\Delta G^{\circ}_{redist} = \Delta H^{\circ}_{redist} - T\Delta S^{\circ}_{redist} = -RT \int p_B(x) ln\left(\frac{p_A(x)}{p_B(x)}\right) dx$$

A common misconception is that conformational penalties only come in the form of entropy; rather, both entropic and enthalpic contributions are important because the penalty will depend on the enthalpic stabilities of newly populated conformers.

References:

1 Salmon, L., Bascom, G., Andricioaei, I. & Al-Hashimi, H. M. A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: the basis for interhelical motions revealed. *Journal of the American Chemical Society* **135**, 5457-5466 (2013).

2 Borkar, A. N., De Simone, A., Montalvao, R. W. & Vendruscolo, M. A method of determining RNA conformational ensembles using structure-based calculations of residual dipolar couplings. *Journal of chemical physics* **138**, 215103 (2013).

3 Borkar, A. N., Vallurupalli, P., Camilloni, C., Kay, L. E. & Vendruscolo, M. Simultaneous NMR characterisation of multiple minima in the free energy landscape of an RNA UUCG tetraloop. *Physical chemistry chemical physics* **19**, 2797-2804 (2017).

4 Bottaro, S., Bussi, G., Kennedy, S. D., Turner, D. H. & Lindorff-Larsen, K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Science advances* **4**, eaar8521 (2018).

5 Yang, S., Salmon, L. & Al-Hashimi, H. M. Measuring similarity between dynamic ensembles of biomolecules. *Nature methods* **11**, 552-554 (2014).

6 Qi, Y. *et al.* Continuous interdomain orientation distributions reveal components of binding thermodynamics. *Journal of molecular biology* **430**, 3412-3426 (2018).