8276–8289 *Nucleic Acids Research, 2020, Vol. 48, No. 15* *Published online 31 July 2020*
doi: 10.1093/nar/gkaa607

IPANEMAP: integrative probing analysis of nucleic acids empowered by multiple accessibility profiles

Afaf Saaidi[1,†], Delphine Allouche[2,†], Mireille Regnier[1], Bruno Sargueil[2] and Yann Ponty[1,*]

[1]CNRS UMR 7161, LIX, Ecole Polytechnique, Institut Polytechnique de Paris, 1 rue Estienne d'Orves, 91120 Palaiseau, France and [2]CNRS UMR 8038, CitCoM, Université de Paris, 4 avenue de l'observatoire, 75006 Paris, France

Received July 16, 2019; Revised July 03, 2020; Editorial Decision July 06, 2020; Accepted July 29, 2020

ABSTRACT

The manual production of reliable RNA structure models from chemical probing experiments benefits from the integration of information derived from multiple protocols and reagents. However, the interpretation of multiple probing profiles remains a complex task, hindering the quality and reproducibility of modeling efforts. We introduce IPANEMAP, the first automated method for the modeling of RNA structure from multiple probing reactivity profiles. Input profiles can result from experiments based on diverse protocols, reagents, or collection of variants, and are jointly analyzed to predict the dominant conformations of an RNA. IPANEMAP combines sampling, clustering and multi-optimization, to produce secondary structure models that are both stable and well-supported by experimental evidences. The analysis of multiple reactivity profiles, both publicly available and produced in our study, demonstrates the good performances of IPANEMAP, even in a mono probing setting. It confirms the potential of integrating multiple sources of probing data, informing the design of informative probing assays.

INTRODUCTION

Historically used as a validation assays (1), enzymatic and chemical probing is increasingly used in combination with computational methods to inform a rational prediction of secondary structure models for RNA (2). Such an integrated approach to structure modeling has led to sizable improvements in prediction accuracy (3) and is currently at the core of successful modeling strategies (4). However, the interpretation of probing data, to inform structure prediction, is challenged by a number of factors, including structural heterogeneity, experimental errors, structural dynamics and the potential variability of reactivity measurements across replicates.

Reagents used within selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) (5) protocols represent a popular class of probes. They react with the 2′-hydroxyl of flexible ribose (6), although the exact properties observed by SHAPE remain the object of ongoing investigations (6–11). As ribose flexibility is proportional to the degree of freedom of the nucleotide, it is assumed that SHAPE reagents discriminate nucleotides involved in stable interactions. Different SHAPE reagents are endowed with different dynamics that can help differentiate Watson–Crick base pairs from more dynamic tertiary contacts (11–14). Other chemical probes, harbouring different chemical reactivities, have been developed. Among the most popular, DiMethyl Sulfate (DMS) is a small molecule that methylates Adenines and Cytosines if not involved in a hydrogen bond, while CMCT reacts with the Watson–Crick face of unpaired Guanosines and Uracils (15,16). These two reagents not only reveal Watson-Crick base pairing, but also other types of contacts involving the same edge. The diversity of probes, some of which are usable *in vivo* (17,18), not only increases coverage of the different positions and structural contexts, but also provides different qualitative information. Modeling can also benefit from a joint analysis of reactivity profiles of single-point mutants, assuming structural homology (19). Integration of multiple sources of probing to improve structure prediction has thus been widely used since the very early days of RNA structure probing (15,16,20–25) but, to the best of our knowledge, has never been fully automated within a soft constraint framework (26).

Computationally, the reactivity of a nucleotide is typically used as a proxy to assess the unpaired nature of individual nucleotides. The past couple of decades have nevertheless seen a series of paradigm shifts in the ways probing information is integrated, somehow mirroring the evolution of *ab initio* methods for secondary structure prediction. The seminal work of Mathews (2) used cutoffs to transform reactivity values into hard constraints. Depending on the used reagent/enzyme, significantly unreactive or reactive positions were forced to remain paired or un-

*To whom correspondence should be addressed. Tel: +33 1 77578095; Email: yann.ponty@lix.polytechnique.fr
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© The Author(s) 2020. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

paired within predicted models. However, such *hard constraints* can be overly sensitive to the choice of the cutoff, leading to artificially unpaired predictions or unsatisfiable sets of constraints. With the advent of SHAPE probing (27), new methods (3,28,29) lifted the requirement of a threshold, supplementing Turner nearest-neighbor energy model (30) with pseudo-energies derived from the reactivities. They then performed a (pseudo) energy minimization, optimizing a tradeoff between the thermodynamic stability and its compatibility with the reactivity profile. Popular packages for secondary structure prediction, such as **RNAStructure** (2) or the **Vienna** package (31), now routinely accept SHAPE-like reactivity profiles as input of their main structure prediction methods. Besides free-energy optimization, such methods notably include the joint folding of aligned RNAs (32), partition function, statistical sampling and Maximal Expected Accuracy predictions (33).

However, independent computational predictions for the same RNA using probing data obtained with different probes often yield substantially different models, none of which are fully consistent with all the probing data. It follows that, while theoretically informative, a multiprobing strategy often leaves a user with different models, from which one cannot objectively decide. In order to identify the best fit with all the data, researchers resort to very intuitive and manual methods such as projecting the results obtained with one probe onto the different models obtained using the constraints from another probe (34–36). At the end of the modeling process, the modeler is often left with several alternatives, all of which may appear equally consistent with the probing data. Thus, we have developed a new modeling procedure that jointly takes into account multiple probing data, and ultimately yields a small collection of secondary structure models.

MATERIALS AND METHODS

The IPANEMAP method

We introduce the integrative probing analysis of nucleic acids empowered by multiple accessibility profiles (**IPANEMAP**), a novel approach that integrates the signals produced by various probing experiments to predict one or several secondary structure models for a given RNA. It takes as input one or several reactivity profiles produced in various *conditions*, broadly defined to denote the conjunction of a reagent, a probing technology, ionic concentrations and, in extreme cases, structurally homologous mutants.

It performs a structural clustering across multiple sets of structures sampled in different experimental conditions, and ultimately returns a set of structures representing dominant conformations supported across conditions. Its underlying rationale is that the prominent presence of a stable secondary structure within the (pseudo-)Boltzmann ensembles induced by multiple experimental conditions should increase its likelihood to be (one of) the native structure(s) for a given RNA. It is thus hoped that integrating several reactivity profiles may be used to promote the native structure as one of the dominant structures within the multi-ensemble, and help circumvent the limitation of pseudo-energies derived from single reactivity profile, which are generally not sufficient to elect the native structure as its minimum (pseudo)-free energy candidate. In other words, combining ensembles of structures generated using multiple probing experiments is likely to denoise the Boltzmann (multi-)ensemble, and thus mitigate systematic biases induced by experimental conditions and reagents.

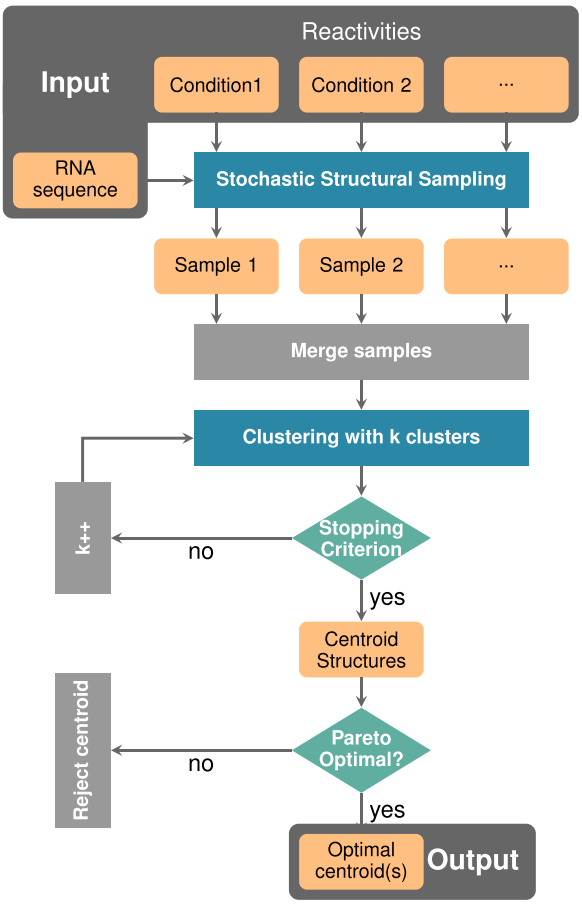

Figure 1. **IPANEMAP** workflow: **IPANEMAP** takes as input an RNA sequence with profiling data, denoted by reactivities, from various experimental conditions. **IPANEMAP** proceeds, first, with a stochastic sampling that results into samples of predicted secondary structure. The data-driven predicted structures are then gathered in one sample, serving as input for the clustering step. **IPANEMAP** proceeds, then, with an iterative clustering that ends once a stopping criterion is reached. This step allows to identify the adequate number of clusters k to be considered. The k resulting clusters are then represented by their centroid structures. Clusters figuring on the 2D-Pareto frontier are considered to be optimal and subsequently their corresponding centroids are reported as the predicted structure through **IPANEMAP**.

Sampling the pseudo-Boltzmann ensembles. Our method, summarized in Figure 1, takes as input a set $\mathcal{D}$ of probing experiments, each materialized by a reactivity profile. It starts by producing (multi)sets of representative structures for each of the reactivity profiles using a SHAPE-directed variant of the classic Ding–Lawrence algorithm (37). Following Deigan *et al.* (28), *soft constraints* are used to complement the free energy contributions of the classic Turner energy model with pseudo energy contributions resulting from the reactivity derived from a probing experiment. Given a reactivity r_i for a position i in a probing d, we as-

sociate a free-energy bonus to unpaired positions, defined as

$$\Delta G_d(i) = m \log(r_i + 1) + b$$

using $m = 1.3$ and $b = -0.4$. Those values were halved in comparison to those recommended by Deigan *et al.* (28), following a grid search optimization on the Cordero dataset, and based on the rationale that lower absolute values for pseudo-energy bonuses increase the expected overlap between pseudo-Boltzmann ensembles. Those pseudo energy contributions effectively guide our predictions towards a subset of structures that are in good agreement with probing data. For each condition d in $\mathcal{D}$, we use the soft constraints framework (26) of the **RNAfold** software (31) to produce a random (multi)set $\mathcal{S}_d$ of M structures in the pseudo-Boltzmann ensemble.

Clustering across conditions. In order to infer recurrent conformations across sampled sets, **IPANEMAP** agglomerate structure (multi)sets while keeping track of their condition of origin, and *clusters* with respect to the *base-pair distance*, the number of base pairs differing between two structures. A clustering algorithm then partitions the (multi)set of sampled structures into *clusters*, (multi)sets of structures such that the accumulated sum of distances over clusters is minimized.

Among the many available options, we chose the Mini Batch k-Means algorithm (38) (MBkM), implemented in the **scikit-learn** package (39), which requires less computational resources than the classic k-means algorithm, yet performed similarly in preliminary studies as an extensive collection of both agglomerative (affinity propagation) and hierarchical (Ward, Diana, McQuitty) clustering algorithms. A dissimilarity matrix, presenting the pairwise base-pair distance between structures, is precomputed and fed to the clustering algorithm.

Any cluster C output by the clustering is a multiset of structures, each labeled with its origin condition of $\mathcal{D}$. The *cluster probability of a structure feature f* (base pair or unpaired base) within a cluster C is then defined as

$$\mathbb{P}_C(f) = \frac{\sum_{\substack{S \in \mathcal{C} \\ \text{s.t. } f \in S}} e^{-E(S)/RT}}{\sum_{S \in \mathcal{C}} e^{-E(S)/RT}}$$

where R represents the Boltzmann constant, $E(S)$ is the Turner free-energy, and $\mathcal{C}$ is the (non-redundant) set of structures in C. From those probabilities we define the *centroid structure* of a cluster as its Maximum Expected Accuracy (MEA) structure, computed efficiently following Lu *et al.* (40).

Moreover, define the (pseudo-)*Boltzmann condition probability* of a structure S, generated for a probing condition d as part of a sampled set $\mathcal{S}_d$, as

$$\mathbb{P}_d(S) = \frac{e^{-E_d(S)/RT}}{\mathcal{Z}_d^*}, \text{ with } \mathcal{Z}_d^* := \sum_{S \in \mathcal{S}_d} e^{-E_d(S)/RT}$$

where $E_d(S)$ is the pseudo free-energy, including the Turner free-energy, assigned to the structure S within the probing condition d. The *stability* of a cluster C denoted its accumulated pseudo-Boltzmann probability across conditions, computed as

$$\text{Stability}(C) = \sum_{d \in \mathcal{D}} \sum_{S \in C \cap \mathcal{S}_d} \mathbb{P}_d(S).$$

A cluster is deemed *significantly populated* if its stability exceeds a predefined threshold ϵ. We set $\epsilon = |\mathcal{D}|/3$ by default, such that at most three clusters are deemed significantly populated, and used as our primary candidates. Finally, we consider two clusters to be *highly similar* if their centroid structures differ by at most δ base pairs ($\delta = 1$ by default), allowing the identification of clusters in the presence of minor variations.

The targeted number of clusters is a critical parameter of the MBkM algorithm. It should, at the same time, remain small enough to ensure reproducibility, while being sufficiently large to discriminate outliers and ensure consistency within each cluster. We determine an *optimal number of clusters k* using an iterative heuristic, gradually increasing the number of clusters until a significantly-populated cluster is split into two similar clusters, or a poorly populated (outlier) cluster is created. Namely, our iterative heuristic consists in running MBkM over an increasing number k of clusters, starting with $k = 2$, until the following *stopping criterion* is met: (i) two significantly populated clusters have associated centroids which are highly similar; or (ii) centroid structures of significantly populated clusters from the previous iteration are highly similar to those of the current iteration.

Filtering the promising conformer(s). Finally, we *select the most promising cluster(s)*, and return their centroid(s). While the final number of clusters may potentially be large, only a handful of clusters are expected to represent structures that are both stable, and supported by a large number of conditions. The remaining clusters are indeed probably artifacts of the clustering method, but nevertheless useful to filter out *noisy* structures.

We postulate that a perfect cluster should have large stability, as defined above, and be representative of several conditions. In the presence of a set of experimental conditions $\mathcal{D}$, we consider that a cluster C supports a given condition d when its probability within d exceeds a given threshold τ. The number of conditions supporting a cluster C is defined as

$$\text{Support}(C) = \left| \left\{ d \in \mathcal{D} \mid \left(\sum_{S \in C \cap \mathcal{S}_d} \mathbb{P}_d(S) \right) \geq \tau \right\} \right|.$$

The value of τ is set to $1/(k + 1)$ with k the final number of clusters, ensuring at least one supporting cluster for each condition.

IPANEMAP evaluates the Stability and Support metrics for each cluster, and filters out any cluster that is dominated by some other with respect to both metrics. The remaining ones are *Pareto optimal*, a classic concept in multi-objective optimization (41). **IPANEMAP** computes and returns the MEA centroid (40) of the Pareto-optimal clusters as its final prediction(s).

Pairwise comparison of structural ensembles induced by reactivity profiles

We want to compare the structural ensembles induced by reactivity profiles, produced across diverse experimental conditions. To that purpose, we simply consider the base pair probability matrices, or dot-plots, resulting from supplementing the Turner energy model with pseudo energy terms. Dot plots can be computed efficiently in the presence of pseudo-energy terms using a variant of the McCaskill algorithm (42).

As a measure of the *ensemble distance* Dist induced by probing data, we consider the dot-plots associated with experimental conditions d and d', and compute the squared Euclidean distance, such that

$$\mathrm{Dist}(d, d') = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left(\mathbb{P}(i, j \mid d) - \mathbb{P}(i, j \mid d')\right)^2 .$$

with n the length of the RNA sequence, $\mathbb{P}(i, j \mid d)$ the Boltzmann probability of forming a base pair (i, j) in the pseudo-Boltzmann ensemble associated with condition d.

Individual dot-plots were computed using the **RNAfold** software in the **Vienna Package 2.2.5**, using the **-p** option in combination with the pseudo-energy terms introduced by Deigan *et al.* (28).

Datasets

To validate our computational method quantitatively, we consider several datasets, depending on the availability of probing data for one or several reagents, restricted to the wild type or produced for several point-wise mutants. Each dataset consists of sequences and individual reactivities to one or several probes, at each position in the RNA, completed with one or several functionally-relevant secondary structures.

Hajdin dataset. A dataset was gathered by Hajdin *et al.* (43) to validate the predictive capacities of probing data-driven predictions. It consists of 24 RNA sequences with known secondary structures for which a single chemical probing reactivity profile (1M7-SHAPE) was made available. This dataset includes sequences originating from a variety of organisms, and spans lengths ranging from 34 nts to 530 nts, with a focus on riboswitches and complex RNA architectures (full list in Supplementary Table S3).

Cordero dataset. Probing data were downloaded from the RMDB (19) on July 2017. In the RMDB, reactivity scores are reported for all nucleotides, including those that are not expected to react with a given reagent. Thus, for the DMS (resp. CMCT) probing, we restricted reactivities to positions featuring nucleotides A and C (resp. G and U), setting the pseudo-energy term to 0 kcal mol^{-1} for other positions. This allowed to decrease the noise generated by reactivities associated with non-targeted nucleotides, leading to more accurate predictions (data not shown).

Didymium structural model and probing data (DiLCrz dataset). We considered the 188 nucleotides Lariat capping ribozyme from *Didymium iridis*, resolved 3.85 Å resolution using X-ray cristallography (PDB: 4P8Z) (44). We annotated the secondary structure elements using the **DSSR** software from the 3DNA suite (45). Non-canonical base pairs were removed, and a non-pseudoknotted secondary structure was extracted as the maximum subset of non-crossing base pairs (46).

Probing data were experimentally generated, as described in the next section, using a comprehensive set of conditions covering some of the popular probing reagents and SHAPE technologies. We also considered the presence/absence of Mg^{2+}, both to assess the capacity of **IPANEMAP** to recover tertiary interactions, and to assess the induced discrepancy on probing profiles and pseudo-Boltzmann ensembles.

Cheng dataset. Starting from the assumption that a functional structure should be preserved during evolution, we wanted to assess the agreement that might exist between probing data profiles for a set of RNA mutants. We considered DMS probing data, generated by (47) through systematic point-wise mutations, for the Lariat-capping ribozyme (equivalent to DMS-MAP^{mg}_{ilu} in the nomenclature below). We renormalized each reactivity profile following the method introduced by Deigan *et al.* (28), restricted to the primer-free sequence: values greater than 1.5 times the interquartile range were discarded, and remaining values were divided by the mean of the top 10% reactivities. Overall, this constitutes a collection of 188 sequences, each having its associated reactivity profile.

Experimental probing protocols

To systematically assess the potential of multiple sources of probing, we considered a difficult example, the *Didymium iridis* Lariat Capping ribozyme (DiLCrz). The native structure of DiLCrz, shown in Figure 4 is highly complex, and features two pseudoknots which cannot be explicitly modeled by most computational methods, making DiLCrz a challenging target for secondary structure prediction.

DiLCrz was probed with different SHAPE reagents: 1M7 (1-methyl-7 nitrosatoic anhydre), NMIA (*N*-methyl isatoic anhydre), BzCN (benzoyl cyanide) and NAI (2-methylnicotinic acid imidazolide) in presence and absence of Mg^{2+}. DiLCrz was also probed with DMS (dimethylsulfate) and CMCT, in presence of Mg^{2+}, resulting in a total of 16 *probing conditions*, a term we use in the following to denote a combination of probing technology, reagent and presence/absence of Mg^{2+} (+ sequencing technology). For each probing condition, three experiments were performed in presence/absence of the reagent, and in a denatured context, following classic SHAPE protocols (5,27).

As a preliminary experiment, we verified that DiLCrz generally adopts a single global architecture. To this end, we subjected DiLCrz to a standard denaturation/renaturation protocol (80°C for 2 min in H_2O, addition of 40 mM of HEPES at 7.5 pH, 100 mM of KCl, 5 mM of $MgCl_2$, followed by 10 min at room temperature, and 10 min at 37°C), and observed the production of a single band on a non-denaturing PAGE, strongly suggesting the adoption of a single conformation.

Stops-inducing probing protocol (SHAPE-CE). 6 pmol of RNA were resuspended in 18 µl of water, denatured at 80°C

for 2 min and cooled down at room temperature during 10 min in the probing buffer (40 mM HEPES at 7.5 pH, 100 mM KCl, in presence or absence of 5 mM $MgCl_2$. After a 10 min incubation at 37°C, RNAs were treated with 2 mM of SHAPE reagent or DMSO (negative control) and incubated for 2 (BzCN), 5 (1M7), 30 (NMIA) or 60 (NAI) minutes at 37°C. Modified or unmodified RNAs were purified by ethanol precipitation and pellets were resuspended in 10 μl of water. Modifications were revealed by reverse transcription using 5′ fluorescently labeled primers (D2 or D4 WellRED, Sigma Aldrich 5′ CTG-TGA-ACT-AAT -GCT-GTC-CTT-TAA 3′) and M-MLV RNAse (H-) reverse transcriptase (Promega) as previously described (48), with only minor modification of the originally described SHAPE protocol (5). cDNAs were separated by capillary electrophoresis (Beckman Coulter, Ceq8000). Data were analyzed using the **QuSHAPE** (49) software. RNA probing was performed in triplicate with distinct RNA preparations. The corresponding data sets are named after the probe followed by CE, and MG if probed in presence of $MgCl_2$.

Mutations-inducing probing (SHAPE-MaP). Probing was conducted as described in Smola *et al.* (27), except we used SuperScript III reverse transcriptase (ThermoFisher) at 50°C for 3 h using the following specific primer:

5′ CTT-CAT-AGC-CTT-ATG-CAG-TTG-CTT -TTT-TTT-TTT TTT-TTT-TTT-GAT-TGT-CTT -GGG-ATA-CCG-GAT 3′

For 1M7-MAP^{mg}_{il}-3D, the denaturation step was repeated three times : After a 1 min incubation at 95°C, 10 mM of 1M7 was added and the mix was incubated 1 min at 95°C. This step was repeated three times for the samples labeled 3D. For NMIA-MAP_{it} and NMIA-MAP^{mg}_{it}, experiments were conducted as described above, except that the NGS library preparation was adapted to Ion Torrent sequencing. Sequences were mapped and analyzed with **ShapeMapper2** (50) for 1M7-MAP^{mg}_{il}-3D and 1M7-MAP^{mg}_{il}-3D. For the other conditions they are mapped with a script based on **ShapeMapper** (27), but adapted to Ion Torrent output files. The corresponding data sets are named after the probe followed by MAP. The presence of MG as a superscript indicates the presence of $MgCl_2$, while the subscript IL (Illumina) or IT (Ion Torrent) specifies the sequencing technology.

DiMethylSulfate (DMS) and CMCT probing. DMS and CMCT probing were conducted essentially as described in (35,36). Succinctly, 6 pmol of RNA were resuspended in 18 μl of water, denatured at 80°C for 2 min and cooled down at room temperature for 10 min in a probing buffer for DMS probing (40 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$) or in CMCT (50 mM of potassium borate pH 8, 100 mM KCl and 5 mM $MgCl_2$). For DMS, RNA was then treated with 1 μl of DMS (1:12 in ethanol) or 1 μl of ethanol for 5 min at 37°C (mock reaction), the reaction was stopped by addition of 400 mM of Tris at 7.5 pH and immediately put on ice. For CMCT, RNA was treated with 10 μl of CMCT (42 mg/mL) or 10 μl H_2O for 10 min at 37°C (mock reaction). Modified RNA were ethanol-precipitated and resuspended in 10 μl of water. Modifications were mapped as described above for SHAPE-CE experiments.The corresponding data sets are named after the probe, followed by CE and MG.

Benchmarking methodology

The Matthews Correlation Coefficient (MCC) is a classic metric for assessing the quality of a predicted structure S, identified by a set of base pairing positions, compared to an accepted reference (native) structure $S^\star$. It represents a compromise between the main metrics derived from the confusion matrix, and is defined as

$$\mathrm{MCC}(S \mid S^\star) = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}},$$

where TP, FP, TN, FN represent the correctly/erroneously predicted and correctly/erroneously omitted base pairs in S with respect to $S^\star$. MCC is a stringent metric, taking values between –100% and 100%, 0% being the expected MCC of a 'coin tossing' random predictor.

For the sake of direct comparison with some competing methods (33), we also report the Geometric Mean (GM) metric, defined as:

$$\mathrm{GM}(S \mid S^\star) = \sqrt{\mathrm{Sens}(S \mid S^\star) \times \mathrm{PPV}(S \mid S^\star)}$$

where $\mathrm{Sens}(S|S^\star)$ represents the proportion of base pairs in $S^\star$ that are in S, and $\mathrm{PPV}(S|S^\star)$ is the proportion of base pairs in S that are also in $S^\star$. This quality metric has very high correlation with the MCC, to which it is equivalent for large values of n, as shown by Gorodkin *et al.* (51).

Assessment of statistical significance

Following Xu *et al.* (52), we assessed the statistical significance of observed changes in predictive performances using a two-tailed Student's t-test (unequal variance) with usual type I error rate $\alpha = 5\%$. When comparing performances within a given dataset, we used the paired version of the test.

Intuitively, the t-test estimates the probability that an observed change in performance can be attributed to chance alone. It considers the hypothesis H_0 that the two distributions have equal mean and computes a P-value, i.e. the likelihood of the observed data under H_0. If the P-value is lower than α, then H_0 is deemed *refuted*, and the observed difference in performances is considered significant.

RESULTS

Implementation

IPANEMAP was implemented in **Python 2.7+**, and mainly depends on the **RNAsubopt** software in the Vienna package (31), and **scikit-learn** (39). **IPANEMAP** is free software distributed under the terms of the MIT license, and can be freely downloaded, along with a collection of helper utilities, from: https://github.com/afafbioinfo/IPANEMAP.

Considering multiple probing conditions appears to improve the quality of predictions

Since **IPANEMAP** supports any number of probing profiles, we considered the Cordero *et al.* (19) dataset, further de-

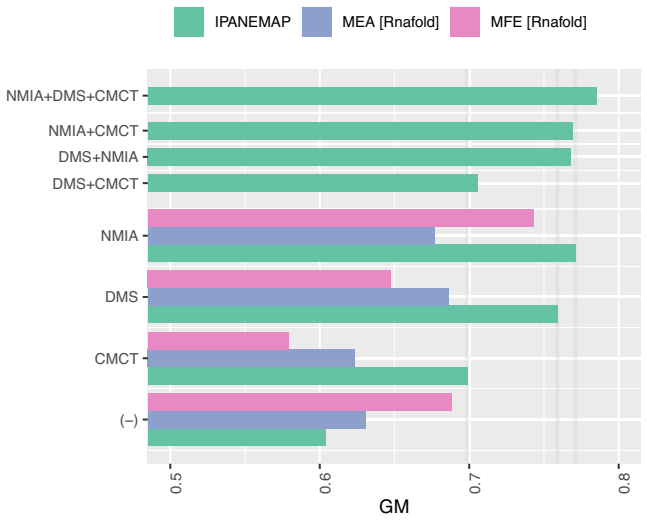

Figure 2. Predictive performances (GM) of **IPANEMAP** on the Cordero (19) dataset consisting of reactivities associated with the CMCT, NMIA and DMS probing for 6 reference RNAs. (Pseudo) MFE and MEA structures, predicted using **RNAfold** (31), are included as references in the case of a single condition, and in the absence of probing.

scribed in the previous section. It consists of six RNAs of known structures, for which reactivity profiles are available for the CMCT, NMIA and DMS reagents. We executed **IPANEMAP** with default parameters on each sequence and any subset of the three available conditions. The centroid secondary structure associated with the largest probability cluster was considered as the final prediction. Predictions were evaluated in term of the geometric mean (GM) metric, for the sake of comparison with previous studies (33). As a control experiment, we also report RNAfold predictions in presence/absence of probing data, both in energy minimization (MFE) and maximum expected accuracy (MEA) modes. Figure 2 shows the averaged GM values for all combinations of tools and probing data (detailed results in Supplementary Tables S1 and S2).

In the absence of probing data, MFE predictions are generally dominant on this dataset, trailed by the MEA, and followed by **IPANEMAP** which, in this setting, devolves into the classic Ding-Lawrence algorithm (37). However, whenever probing reactivities are available, the single centroid returned by **IPANEMAP** always achieves higher GM values (Avg: 70%) than both MEA (Avg: 62%) and MFE (Avg: 58%), whose relative performances depend on the probing reagent. Interestingly, the quality of MFE and MEA predictions does not systematically benefit from additional probing data. Indeed, for half of the reagents and methods, the average GM obtained in the presence of a single reagent is lower than in the absence of probing data. Also, the impact of single probing data varies greatly across the three reagents, and the GE values of predictions respectively informed by CMCT, DMS and NMIA are ordered increasingly for all approaches, except for a minor reversal of DMS and NMIA in MEA mode.

The joint analysis of pairs of probing conditions appears to average the quality of predictions. Any combination of probing data yields a GM value that is always greater than the worst-performing condition in the pair, yet worse than the best-performing alone. Interestingly, the addition of the worst performing condition (CMCT), does not equally affect the performances of DMS (76% → 70.5%) and NMIA (77.1% → 76.9%), despite the latter conditions inducing similar GM values when considered alone. Indeed, CMCT+DMS yields GM values that are only remotely better than the worst-performing CMCT alone (70% → 70.5%), while CMCT+NMIA greatly outperforms CMCT alone (70% → 76.9%), almost matching the performance of the NMIA alone (77.1%). It is also worth mentioning that NMIA+CMCT, combining the best and worst conditions, achieves a better combined performance than DMS+NMIA, the two best mono-probing conditions.

Remarkably, the combination of the three conditions leads to the best overall predictions, averaging 78.5% GM. This seemingly improves by 8.5%, 2.7% and 1.4% over the average predictions achieved using CMCT, DMS and NMIA respectively. Table 1 illustrates the incremental refinement of **IPANEMAP** predictions for a glycine riboswitch upon increasing the number of considered probing conditions. However, these observations merely suggest an improvement, as statistical significance cannot be established on such a small dataset. Still, they suggests some level of complementarity between probing conditions, as already suggested by recent analyses (13,14,16,54) and supported by further analyses in this paper.

IPANEMAP performs comparably to state-of-the-art predictive methods when informed by a single profile

We assessed the predictive capacities of **IPANEMAP** in a classic setting where a single probing condition is available, and compared its performances against **RSample** software recently introduced by Spasic *et al.* (33), **RSample** relies on a sampling/clustering method developed independently from the current work. It was shown to perform comparably as a comprehensive collection of state-of-the-art methods in probing-guided structure prediction, including **RME** (55), **RNAprob** (56), **RNAprobing** (3), **RNAsc** (29) and the **fold** utility from the **RNAStructure** suite (57). We considered the Hajdin dataset (43), which consists of 24 RNAs of lengths ranging from 47 to 500 nts, all believed to fold into a unique documented conformation.

We used the default numbers of sampled structures for **RSample** and **IPANEMAP**, namely 10 000 to compute correction factors and 1 000 for the clustering phase. For all RNAs, a single structure was returned by both software. Note that the Pareto-optimality always implies a single returned cluster/structure for **IPANEMAP** in a mono probing setting. We computed and report in Figure 3 the individual Geometric Means (GM) of predicted structures. Sensitivity and PPV can be found in Supplementary Table S4, while Supplementary Table S5 reports the stability of predictions over 10 independent runs of **IPANEMAP**.

IPANEMAP appears to perform similarly, or even slightly favorably in comparison with **RSample**. It averages 80.1% GM, and reports more accurate structural models than its competitor for 14 RNAs in the Hajdin dataset (43), returns structures equally as good on two RNAs, and is

Table 1. Dot-bracket representations of the native structure and **IPANEMAP** predictions for the glycine riboswitch sequence of the Cordero *et al.* (53) data set, supplemented with an increasing collection of probing conditions. Please note that, throughout the manuscript, by 'conditions' we mean a combination of the nature of the probe, the probing technology, the ions present, the temperature, or any other 'environmental' variation Positions with green and red backgrounds indicate correctly and incorrectly predicted base pairs

Sequence	GAUAUGAGGAGAGAUUUCAUUUUAAUGAAACACCGAAGAAGUAAAAUCUUUCAGGUAAAAAGGACUCAUAUUGGACGAACCUCUGGAGAGCUUAUCUAAGAGAUAACACCGAAGGAGCAAAGCUAAUUUUAGCCUAAACUCUCAGGUAAAAGGACGGAG	GM
∅	(((((((((((((((((.(((((.....)))))..)))))))))))))))).....(((((.....)))))).....(((((.....)))))...(((.....)))...(((.....)))).....(((((.....))))).....((((.....)))).....((((.....)))).....((((.....)

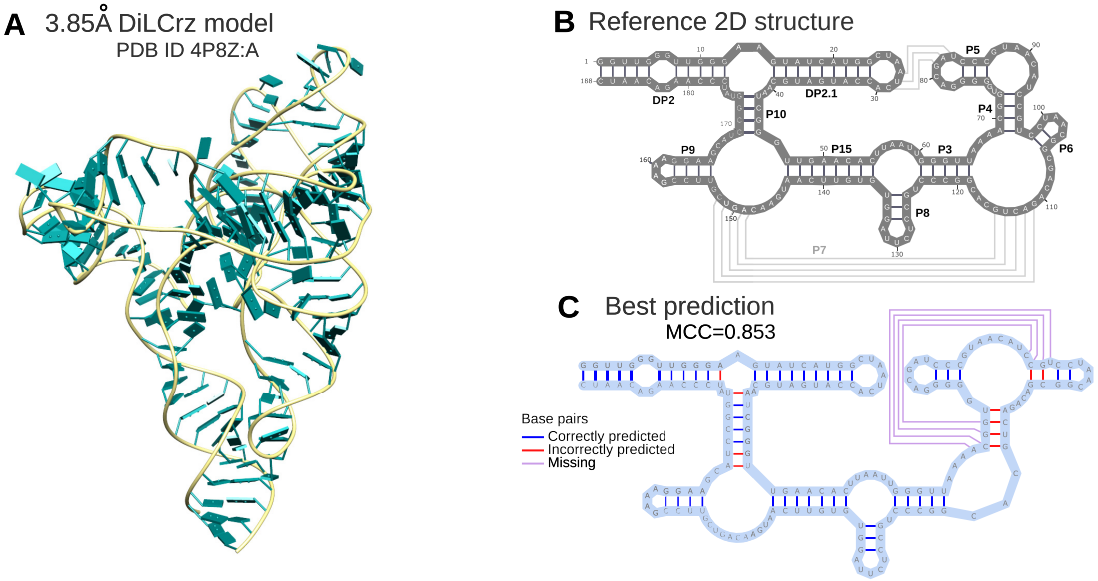

Figure 4. Model of the Lariat capping ribozyme from *Didymium iridis* (**A** – PDB 4P8Z:A by Meyer *et al.* (44)), secondary structure, as annotated by **DSSR** (45) (**B**) and **IPANEMAP** first-ranking prediction (**C**), obtained from the integration of eight maximally-diverging probing conditions (NMIA-MAP$^{mg}_{it}$, 1M7-MAP$_{il}$-3D, NAI-CE^{mg}, 1M7-CE, CMCT-CE^{mg},1M7-MAP$_{il}$-3D, and 1M7-CE^{mg}).

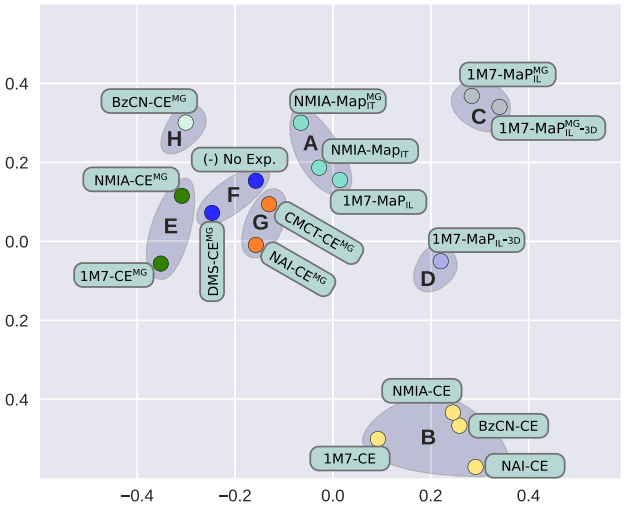

Figure 5. 2D spatial representation of conditions, computed by principal component analysis (PCA) to optimally reflect pairwise ensemble distances between conditions. Colors and grayed areas indicate clusters of conditions.

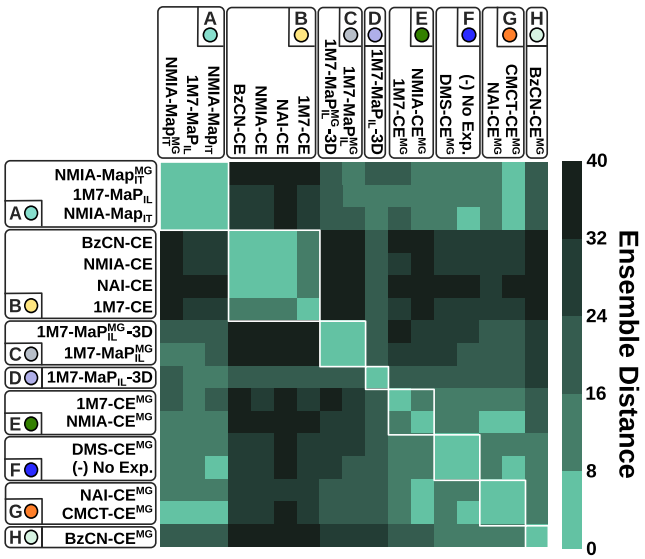

Figure 6. Discretized ensemble distance induced by conditions. Conditions ordered by spectral biclustering. Conditions regrouped within clusters by *k*-mean clustering appear as blocks on the x and y axis.

Consequently, we will single out this cluster within our detailed analysis of the multi-probing performances.

Assessment of predictions informed by individual probing conditions (mono probing). For each probing condition, we executed **IPANEMAP** on a single reactivity profile, using a sample size of 1000 structures. For the sake of reproducibility, we executed **IPANEMAP** 10 times for each condition, and report in Table 2 the average MCC values. We also report for reference the MCCs obtained by running **RNAfold** with default parameters in energy minimization (MFE) and Maximum Expected Accuracy (MEA) modes.

The predictive performances of **IPANEMAP** averages a 70.5% MCC, compared to 68.5% and 68% average MCC for MEA and MFE predictions respectively. This suggests modest, barely significant, improvements compared to classic prediction paradigms (*P*-values of 2% and 14% respectively). The stability of predictions across conditions seem comparable for all **IPANEMAP**, MEA and MFE-driven predictions (std. dev. of 8%). Excluding the outliers of cluster B○ improved the average MCC to 74%.

Across the 16 conditions, we observed a large discrepancy in the capacity of the experimental setup to inform pre-

Table 2. Predictive performances (MCC) of **IPANEMAP**, averaged over 10 runs, for the DiLCrz dataset, consisting of 16 reactivity profiles. (Pseudo) MFE and MEA structures, predicted using **RNAfold** (31), are included for reference. The structure was probed in presence (●) or absence (◯) of Mg^{2+}

Condition	Clust.	Mg^{2+}	MCC% IPANEMAP	MFE	MEA
1M7-MaP$_{IL}^{MG}$	C	●	85	82	82
1M7-MaP$_{IL}$	A	○	84	82	84
NMIA-MaP$_{IT}^{MG}$	A	●	81	80	80
NMIA-MaP$_{IT}$	A	○	80	69	69
1M7-MaP$_{IL}^{MG}$-3D	C	●	74	75	75
NMIA-CE^{MG}	E	●	73	72	73
NAI-CE^{MG}	G	●	73	68	70
BzCN-CE^{MG}	H	●	71	68	54
1M7-CE^{MG}	E	●	70	69	70
CMCT-CE^{MG}	G	●	70	68	73
1M7-MaP$_{IL}$-3D	D	○	64	56	61
DMS-CE^{MG}	F	●	63	66	67
NMIA-CE	B	○	61	60	60
1M7-CE	B	○	60	61	61
BzCN-CE	B	○	60	59	59
NAI-CE	B	○	60	59	59
Avg Technology		–	78	74	75
		–	66	65	65
Avg -/+ Mg^{2+}		○	67	64	65
		●	73	72	71.5
Average overall		–	70.5	68	68.5

dictions, with MCCs ranging from >80% (NMIA-MaP$_{it}$, NMIA-MaP$_{it}^{mg}$, 1M7-MaP$_{il}$ and 1M7-MaP$_{il}^{mg}$) to around 60% (NAI-CE, BzCN-CE, 1M7-CE and NMIA-CE). Interestingly the MCC value cannot be trivially anticipated from the compatibility of the reactivity scores with the predicted structure, or even with the native structure (see Supplementary Table S1). These predictive capacities clearly outperformed those achieved in the absence of probing data (51% MCC), and can be interpreted as indicative of predictions of good overall quality. The lowest MCC value of 60% is, in particular, equally consistent with Sensitivity/PPV values of 60%/60%, 80%/46% or 45%/82%.

Unsurprisingly, the presence/absence of Magnesium ions during the probing was observed to impact the predictions, with an observed drop from an average 73% MCC in the presence of Mg^{2+} to 67% MCC in the absence of Mg^{2+}. However, while the poorly performing conditions in Cluster B○ were probed in the absence of Mg^{2+}, some of the most informative profiles (1M7-MaP$_{il}$ and NMIA-MaP$_{it}$) were obtained in absence of Mg^{2+}. Even more drastic changes of performances were observed when comparing the MCCs of SHAPE-MaP experiments (avg 78%) with the historical SHAPE CE technology (avg 66%).

Bi-probing mitigates the influence of outliers, but does not significantly improve prediction quality. Next, we turned to a systematic exploration of the predictive capacities of bi-probing analyses, based on pairs of probing profiles, and attempted to quantify their impact on **IPANEMAP** predictions. For each pair and triplets of conditions, we executed **IPANEMAP** using 1000 structures per condition, and considered the first returned structure. We then computed the associated MCC, and compared it with the MCC of the worst performing condition (Min), best-performing condition (Max) and average MCC over the single conditions experiments. A summary of the results over pairs is reported in Figure 7 (details in Supplementary Table S6).

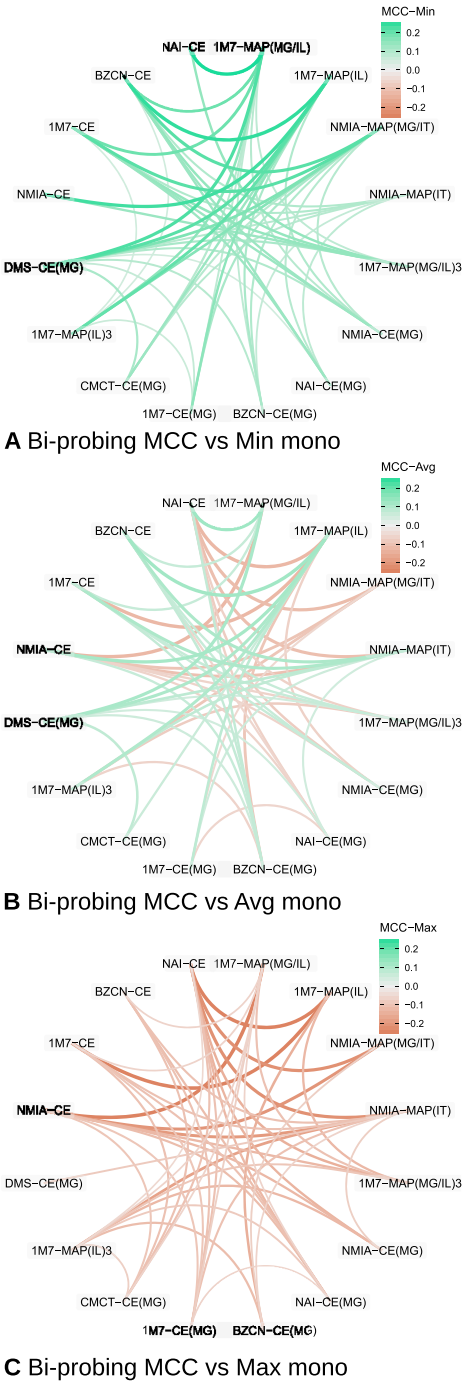

Figure 7. Shift in **IPANEMAP** predictive capacity (MCC) observed upon considering two conditions, compared to the worst (**A**), average (**B**) and best (**C**) predictions achieved for mono-probing conditions. Green and red edges indicate enhanced and degraded predictions respectively compared to the reference (Min, Max or Average of conditions). Thickness indicates absolute shift value (only values >5% reported for readability).

A superficial inspection of the resulting MCCs appears to confirm the conclusions reached on the Cordero *et al.*

(53) dataset. Indeed, considering two conditions led to predictions whose quality fall between the worst and the best one, while being generally close to the average. More precisely, the average MCC value of pair-informed predictions remained at 70.7%, improving only by 0.2% over the average MCC of mono-probing predictions. Pair-informed predictions were marginally more reliable, with the median MCC increasing to 72.5% MCC from 70.5% for mono-probing predictions. However, outliers strongly impacted the overall picture, and ignoring the conditions in cluster B🟡 increased the average MCC to 75.1%, while unsurprisingly decreasing the standard deviation (8.6% → 6.5%).

Compared to the minimum MCC of the pair, the MCC of bi-probing predictions was improved by 5.3% on average over the MCC of the worst-performing conditions. Pair MCCs exceeded their associated min. MCC by at least 5% in 47 pairs out of the 120 possible pairs, while never being dominated by >5% MCC.

Bi-probing seemed to perform similarly as the average of the two mono-probing conditions, with an average MCC improvement of 0.2%. However, for 56 out of the 120 possible pairs, the bi-probing conditions exhibited an improvement of at least 1% over the average, while a decay of at least 1% was observed for 46 out of the remaining 64 pairs. Such a decay can partly be attributed to the disruptive effect of the outliers from cluster B🟡, involved in 35 of the 46 observed loss of predictive performance. Moreover, removing this cluster induced an average MCC improvement of 1.1%.

Predictions informed by two conditions remained, however, generally dominated by the best performing condition of the pair. On average, the MCC of the best condition is 4.9% higher than the one achieved by bi-probing analysis. Outliers in cluster B🟡 are partly responsible for this situation, and their removal reduces the average MCC decay to 3.3%. Moreover, for nine pairs of conditions, the bi-probing analysis produced MCCs that are at least 1% better than the best of its two conditions, but these encouraging examples were generally dominated by 66 examples where the pair performs >1% worse than the best mono-probing condition.

Considering three conditions yields improvement over the average of mono conditions within the same triplet. Increasing the number of conditions to three boosted the average MCC to 72.3% and even reached 76.5% over triplets that do not contain conditions from cluster B🟡. The MCC gain over the worst condition in the triplet reached 9.1% on average. Predictions were also more consistently good, with a median MCC of %73.4. Disregarding conditions in cluster B🟡 did stabilize the quality of prediction (std. dev. 8.4% → 5.5%), while increasing the average MCC to 77.4%.

Compared to bi-probing, MCC values achieved by integrating triplets of conditions indicated a clear gain over the average performances of individual conditions, supporting a notion of complementary between experiments. Indeed tri-probing experiments induced an average +1.7% MCC improvement (median +2.6) over the average of mono-probing prediction in the triplet (P-value=1.4×10^{-9}). The gain of quality was widespread, and observed for 65% of the triplets. Finally, 39 triplets were found to induce MCC values greater than 83%.

Table 3. Predictions do not benefits from the joint consideration of similar conditions. For each cluster, all subsets of conditions in the cluster are considered, and the MCC of the resulting prediction ('Joint') is compared to the average MCC of predictions performed with individual conditions independently ('Mono'), revealing little improvement

	Conditions	MCC% Joint	MCC% Mono
Cluster A 🟢	NMIA-MAP$^{MG}_{IT}$, 1M7-MAP$_{IL}$, NMIA-MAP$_{IT}$	80	82
	NMIA-MAP$^{MG}_{IT}$, 1M7-MAP$_{IL}$	82	82
	1M7-MAP$_{IL}$, NMIA-MAP$_{IT}$	82	82
	NMIA-MAP$^{MG}_{IT}$, NMIA-MAP$_{IT}$	80	80
Cluster B 🟡	NAI-CE, BZCN-CE, 1M7-CE, NMIA-CE	60	60
	NAI-CE, BZCN-CE, 1M7-CE	60	60
	NAI-CE, BZCN-CE, NMIA-CE	61	60
	NAI-CE, 1M7-CE, NMIA-CE	61	60
	BZCN-CE, 1M7-CE, NMIA-CE	61	60
	NAI-CE, BZCN-CE	60	60
	NAI-CE, 1M7-CE	59	60
	NAI-CE, NMIA-CE	60	60
	BZCN-CE, 1M7-CE	61	60
	BZCN-CE, NMIA-CE	60	60
	1M7-CE, NMIA-CE	62	61
Cluster C ⚫	1M7-MAP$^{MG}_{IL}$, 1M7-MAP$^{MG}_{IL}$-3D	82	80
Cluster E 🟢	1M7-CE^{MG}, NMIA-CE^{MG}	73	72
Cluster G 🟠	NAI-CE^{MG}, CMCT-CE^{MG}	73	72

The enhanced performances of tri-probing over bi-probing was deemed statistically significant (P-value=2×10^{-2}), and consistent with the *voting* principle underlying the clustering used in **IPANEMAP**. Indeed, in a bi-probing setting, a single outlier (e.g. NAI-CE) may entirely determine the final structure due to its Boltzmann ensemble being very tightly concentrated around a, presumably erroneous, centroid structure. In the presence of three or more conditions, however, clusters resulting from an outlier condition are typically expected to be dominated in the Pareto front, by compatible structures originating from alternative conditions. It follows that the influence of outliers over the final prediction is mitigated.

Similar conditions do not contribute to the quality of predictions. Our distance assessment and clustering on conditions revealed groups of conditions that were highly compatible in their conclusions, while others seemed to include highly diverging structural information. While some outlier conditions, such as those of cluster B🟡, clearly stood out as erroneous, the results of our systematic analysis of pairs and triplets supported a notion of *complementarity between conditions*. Under this assumption, downstream analyses would be more likely to benefit from the presence of highly diverging probing experiments, than the accumulation of similar (arguably redundant) profiles.

To test this hypothesis, we executed **IPANEMAP** on every subset of conditions within clusters, and report in Table 3 the quality (MCC) of predictions for any subset of similar conditions (see Supplementary Figure S5 for associated models). Note that such an analysis does not use knowledge of the native structure, since our clustering only relies on properties of the pseudo-Boltzmann ensemble.

As expected, compared to the average MCC over associated mono-probing analyses, a joint consideration of similar conditions only induced limited progress over single

Table 4. Predictive performances of conditions across clusters. A single condition is chosen from each cluster except for B. Clusters consisting of a single condition (D, F and H) contribute additional conditions: $\mathcal{M}$={1M7 − MAP$_{IL}$ − 3D, BzCn − CE^{MG}, DMS − CE^{MG}} . The MCC of the predicted structure is reported ('Joint' columns) and compared to the average of mono conditions ('Mono' columns), in presence/absence of NAI-CE

Conditions in individual clusters (+$\mathcal{M}$)				MCC%			
				-NAI		+NAI	
A	C	E	G	Joint	Mono	Joint	Mono
NMIA-MAP$_{IT}^{MG}$	1M7-MAP$_{IL}^{MG}$	1M7-CE^{MG}	NAI-CE^{MG}	84	72	83	71
NMIA-MAP$_{IT}^{MG}$	1M7-MAP$_{IL}^{MG}$	1M7-CE^{MG}	CMCT-CE^{MG}	75	72	82	71
NMIA-MAP$_{IT}^{MG}$	1M7-MAP$_{IL}^{MG}$	NMIA-CE^{MG}	NAI-CE^{MG}	79	73	82	71
NMIA-MAP$_{IT}^{MG}$	1M7-MAP$_{IL}^{MG}$	NMIA-CE^{MG}	CMCT-CE^{MG}	82	72	60	71
NMIA-MAP$_{IT}^{MG}$	1M7-MAP$_{IL}^{MG}$-3D	1M7-CE^{MG}	NAI-CE^{MG}	70	71	60	70
NMIA-MAP$_{IT}^{MG}$	1M7-MAP$_{IL}^{MG}$-3D	1M7-CE^{MG}	CMCT-CE^{MG}	70	70	72	69
NMIA-MAP$_{IT}^{MG}$	1M7-MAP$_{IL}^{MG}$-3D	NMIA-CE^{MG}	NAI-CE^{MG}	72	71	74	70
NMIA-MAP$_{IT}^{MG}$	1M7-MAP$_{IL}^{MG}$-3D	NMIA-CE^{MG}	CMCT-CE^{MG}	75	71	60	70
1M7-MAP$_{IL}$	1M7-MAP$_{IL}^{MG}$	1M7-CE^{MG}	NAI-CE^{MG}	82	73	75	71
1M7-MAP$_{IL}$	1M7-MAP$_{IL}^{MG}$	1M7-CE^{MG}	CMCT-CE^{MG}	84	72	60	71
1M7-MAP$_{IL}$	1M7-MAP$_{IL}^{MG}$	NMIA-CE^{MG}	NAI-CE^{MG}	73	73	60	72
1M7-MAP$_{IL}$	1M7-MAP$_{IL}^{MG}$	NMIA-CE^{MG}	CMCT-CE^{MG}	82	73	82	71
1M7-MAP$_{IL}$	1M7-MAP$_{IL}^{MG}$-3D	1M7-CE^{MG}	NAI-CE^{MG}	84	71	60	70
1M7-MAP$_{IL}$	1M7-MAP$_{IL}^{MG}$-3D	1M7-CE^{MG}	CMCT-CE^{MG}	70	71	84	70
1M7-MAP$_{IL}$	1M7-MAP$_{IL}^{MG}$-3D	NMIA-CE^{MG}	NAI-CE^{MG}	74	72	60	70
1M7-MAP$_{IL}$	1M7-MAP$_{IL}^{MG}$-3D	NMIA-CE^{MG}	CMCT-CE^{MG}	74	71	75	70
NMIA-MAP$_{IT}$	1M7-MAP$_{IL}^{MG}$	1M7-CE^{MG}	NAI-CE^{MG}	82	72	75	71
NMIA-MAP$_{IT}$	1M7-MAP$_{IL}^{MG}$	1M7-CE^{MG}	CMCT-CE^{MG}	83	72	82	70
NMIA-MAP$_{IT}$	1M7-MAP$_{IL}^{MG}$	NMIA-CE^{MG}	NAI-CE^{MG}	73	73	59	71
NMIA-MAP$_{IT}$	1M7-MAP$_{IL}^{MG}$	NMIA-CE^{MG}	CMCT-CE^{MG}	82	72	82	71
NMIA-MAP$_{IT}$	1M7-MAP$_{IL}^{MG}$-3D	1M7-CE^{MG}	NAI-CE^{MG}	73	71	82	69
NMIA-MAP$_{IT}$	1M7-MAP$_{IL}^{MG}$-3D	1M7-CE^{MG}	CMCT-CE^{MG}	70	70	73	69
NMIA-MAP$_{IT}$	1M7-MAP$_{IL}^{MG}$-3D	NMIA-CE^{MG}	NAI-CE^{MG}	74	71	68	70
NMIA-MAP$_{IT}$	1M7-MAP$_{IL}^{MG}$-3D	NMIA-CE^{MG}	CMCT-CE^{MG}	74	71	73	69
			Average	77	72	72	70

probing analyses (between –1% and +1%), leading to an modest average improvement (+0.2% average MCC). This is consistent with the notion that supplementing a probing-based modeling with compatible conditions provides very little additional information, leading to a limited contribution of similar conditions.

Considering diverse conditions improves and stabilizes the quality of predictions. Having established the redundant nature of similar conditions, and their limited contribution to downstream modeling, we turn to an analysis of conditions across clusters. We executed **IPANEMAP** to assess the quality of predictions informed by conditions chosen across the eight clusters. We considered any possible combination, except for the outlier cluster B, for which we only considered NAI-CE as a representative.

The results, summarized in Table 4, demonstrate a clear advantage of including diverse conditions, with an average MCC value of 77% in the absence of condition from cluster B. Those performances compare favorably against the 72% MCC achieved on average by individual conditions (P-value=3×10^{-5}), and reveal a remarkable stability (5% std. dev.), and a degradation in only two out the 24 tested cases (∼1% loss of MCC in both cases). A similar trend can be observed in the disruptive presence of NAI-CE, inducing an 72% average MCC, albeit with an increased standard deviation of 9%.

Finally, considering the 16 conditions together induced a 80% MCC value, increasing to 83% in the absence of cluster B. Overall, these results support the notion that the strategy used by **IPANEMAP** is able to exploit the complementary information provided by multiple conditions, contributing to better predictions.

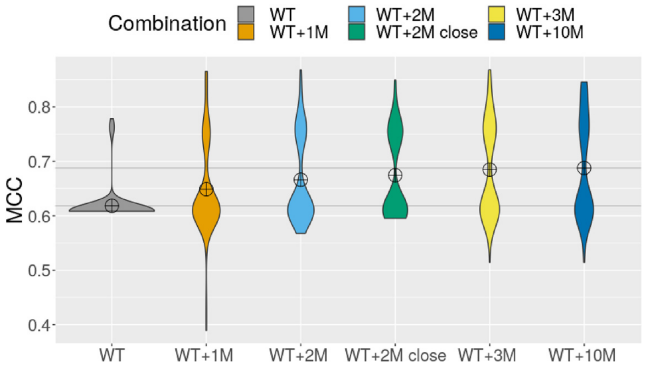

Figure 8. Distributions of predictive accuracies (MCC) of structures predicted by **IPANEMAP**, informed by Mutate-and-Map data, for DiLCrz (Cheng dataset (47)). Predictions were repeated 100 times, informed both by the wild-type sequence and reactivity profile (WT), and with randomly-selected reactivity profiles associated with 1, 2, 3 or 10 single-point mutants (`WT+`k`M`). Pairs of mutants restricted to the 20 most similar to the WT were also considered (`WT+2M close`).

Supplementing wild-type profiles with random single mutants increases prediction quality

In this final analysis, we tested the capacity of **IPANEMAP** to extract and exploit complementary information produced by the systematic probing of single-point mutants using the Mutate-and-Map (MaM) protocol (23). The joint analysis of several mutants may appear error-prone, as single point mutants are expected to adopt different structures than the wild-type sequence (WT), an expectation that is at the core of downstream analyses (23). However, such changes are typically local, and the robustness of **IPANEMAP** to outlier conditions nourishes the hope that the benefits of including independently-produced conditions will outweigh the cost of noise introduced by local changes.

We generated 100 uniformly-distributed subsets consisting of 1, 2, 3 and 10 mutants extracted from the Cheng dataset (47), which we each supplemented with the WT sequence. For each set of probing profiles (DMS-MAP$_{ilu}^{mg}$), we executed **IPANEMAP** with default parameters, using a sample size of 1000 structures. We also analyzed the WT sequence alone, reproducing the analysis 100 times to compare the variability across sets of mutants to the one induced by stochastic aspects of our method. Finally, we performed a restricted analysis focusing on the WT, supplemented with 2 probing profiles selected from the 20 most similar conditions to the WT, as assessed by the ensemble distance metric. By comparing our results with those obtained for unrestricted pairs of mutants, we assess the influence of structure variability on the quality of our predictions. We report in Figure 8 the distribution of MCC values associated with the first prediction returned by **IPANEMAP** (see Supplementary Figure S6 for details).

For the WT alone, the predictions of **IPANEMAP** were associated with 61.8% MCC, comparable to the worst condition of our previous dataset. Our predictions were generally stable and, in 94/100 runs, resulted in a structure hav-

ing MCC between 60.5% and 61%. The remaining six runs showed improved MCC values, ranging from 76% to 78% and suggesting the existence of an alternative conformation in the pseudo-Boltzmann ensemble.

When the WT and a single random mutant (WT+1M) were jointly considered, the dispersion of the MCCs increased, and values ranging from 38.9% to 85% were observed. The average MCC in the presence of a single mutant increased to 64.9%, suggesting a significant positive contribution from the additional mutant profile (P-value=1×10^{-3}).

This trend is confirmed when two random mutants were considered in addition to the WT (WT+2M), with an average MCC across runs increasing to 66.6%. Interestingly, the dispersion of the MCC was lower for WT+2M than for WT+1M, with lower MCC values >57%. This results suggest adding a second mutant profile has a role of *tie-breaker*, allowing to mitigate the deleterious effect of some outlier. However, outliers still played a disruptive role, and restricting chosen mutants to the 20 reactivity profiles most similar to the WT further improved the average MCC 2.8% to 67.7%, a significant improvement compared to WT+1M (P-value $= 2.4 \times 10^{-2}$).

The average MCC again increased in the presence of three mutants, to reach an average MCC of 69.3%, a substantial improvement that nevertheless fails to reach the level of statistical significance (P-value $= 11 \times 10^{-2}$). For 10 mutants, the average MCC and overall distribution remains highly similar to that obtained by considering three mutants, suggesting a plateau in the performances. Overall, these results support a notion of complementarity between the probing data produced across reasonably similar mutants, leading to a gradual decrease of the signal to noise ratio. The bimodal nature of the distribution also suggests the existence of two conformations, equally supported by the WT and the mutant profiles.

CONCLUSION AND DISCUSSION

We introduce **IPANEMAP**, an integrative method for the prediction of secondary structure models from multiple probing profiles. Based on the simple premise that good structural models should be thermodynamically stable and supported by most probing experiments, it uses a combination of stochastic backtrack in the pseudo-Boltzmann ensemble, coupled with a structural clustering to elect dominant conformations. This strategy is fast, generally stable, and leads to predictions that clearly benefit from the availability of multiple probing profiles, as demonstrated by detailed analyses performed of four datasets.

Our analyses reveal that integrating multiple profiles of at least three conditions allows to mitigate the effect of outlier conditions and, at times, to produce better secondary structure models than the one inferred only from the most accurate profile. A comparison of the joint performances to the average appears fair since no single reagent appears to systematically dominate in term of predictive capacity. It follows that a modeler cannot objectively favor *a priori* a probing condition/reagent (54). In other words, multiple experimental probing profiles, even combined in a pairwise

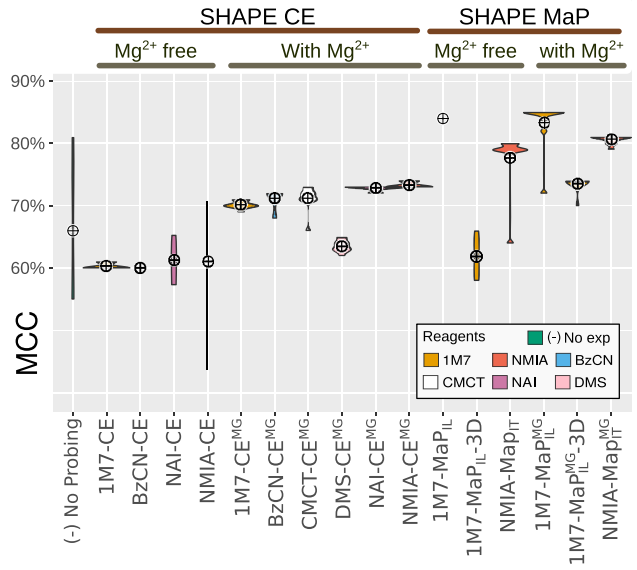

Figure 9. Stochastic variability of **IPANEMAP** predictions for DiLCrz dataset. MCC of predicted structures over 10 runs of **IPANEMAP**, in presence/absence of Mg^{2+}, using both SHAPE CE and MaP protocols for a collection of reagents.

fashion, can be used jointly to mitigate the empirical risk of a misprediction.

To the best of our knowledge, **IPANEMAP** currently represents the only available method for a joint automated consideration of multiple probing profiles. While hard constraints (2) could in principle support multiple sources of probing data, their derivation from reactivities usually require a careful choice of cutoffs, making their automation virtually impossible. Overly liberal cutoffs will often induce contradictory constraints (i.e. no compatible structure) across conditions, while conservative cutoffs will lead to sparse constraints, practically wasting most of the probing-derived information. In fact the tediousness, and suboptimality, of the manual determination of those cutoffs while modeling HIV-1 structural elements (48), was one of the main motivation behind the present work.

Current limitations of **IPANEMAP** include the lack of an explicit support for pseudoknots, which will be addressed in a future version, using polynomial-time dynamic programming (58), or a iterative heuristic alternative (43). The inherently stochastic nature of the sampling scheme may also be challenging for certain types of analysis. However, despite its stochastic foundations, **IPANEMAP** is typically stable in its predictions, as can be seen in Figure 9 for all probing datasets produced over our DiLCrz case study. Interestingly, the most notable exception to the general stability is observed for the probing-free prediction, consistent with the widespread interpretation of SHAPE-induced pseudo-potentials as *focusing* the predictions onto a subset of the Boltzmann ensemble. A similar behavior is observed for the other datasets considered in our study (see Supplementary Tables S2 and S5). Finally, the current clustering method induce a growth of time and memory requirements that scales quadratically with the number of samples. While this al-

lows a joint analysis of a dozen of reactivity profiles in less than an hour on a personal computer, we expect the ability to push the number of samples will lead to better, even more reproducible, predictions. To that purpose, we will explore embedding techniques and linear-time community detection algorithms as alternatives to MBkM (38).

Future extensions also include addressing the dilemma of whether or not to include Mg^{2+} ions during the probing, in the perspective of a downstream computational prediction. On the one hand, RNA reaches its native 3D structure in the presence of Mg^{2+} ions (1). However, such native structures include non WC pairings, and complex topological motifs such as pseudoknots, kissing loops, tetraloop interactions, that classic structure prediction algorithms do not explicitly capture. In contrast, in the presence of only monovalent ions, RNA is expected to form its helical domains, but leave its tertiary contacts unstable, with the notable exception of the G quadruplex. Therefore, probing the RNA without Mg^{2+} could provide a signal that is less deceptive, and thus more informative, to predicting algorithms. Unfortunately, destabilizing tertiary contacts increase the probability of misfolding, or the coexistence in solution of multiple conformations. In this work we probed DiLCrz, which contains many tertiary motifs, in the presence and absence of Mg^{2+}. Remarkably, the most accurate predictions were obtained with 1M7 and NMIA irrespectively of the presence/absence of Mg^{2+}, and despite the fact that reactivity profiles substantially diverged between the two conditions (see Supplementary Figure S4). However, for other reagents/protocols, predictions appeared positively impacted (+10% average MCC) by the presence of Mg^{2+}, painting a mixed picture that would deserve future investigations.

DATA AVAILABILITY

IPANEMAP is freely downloadable a https://github.com/afafbioinfo/IPANEMAP.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to: Ronny Lorenz for developing a probing-informed version of **`RNAsubopt`**, allowing to sample the pseudo-Boltzmann ensemble; Nathalie Chamond for helpful suggestions; Benoit Masquida for the gift of DiLCrz plasmid and helpful discussions about protocols and DiLCrz structure; and an anonymous reviewer for helping us detect an error in our execution of **`RSample`**.

Authors contributions: A.S., B.S. and Y.P. designed the computational method. A.S. and Y.P. implemented the computational method. D.A. and B.S. planned and carried out all probing experiments. A.S. and Y.P. planned and carried out the computational experiments. A.S., D.A., B.S. and Y.P. contributed to the interpretation of the results. All authors provided critical feedback and helped shape the research, analysis and manuscript

FUNDING

Research in both B.S. and Y.P. labs, including the PhD fellowships of A.S. and D.A., have been supported by La Fondation pour la Recherche Medicale [FRM DBI20141423337 to B.S. and Y.P.]. Funding for open access charge: Agence Nationale de la Recherche [ANR-19-CE30-0021].

Conflict of interest statement. None declared.

REFERENCES

1. Knapp,G. (1989) Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.*, **180**, 192–212.
2. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 7287–7292.
3. Washietl,S., Hofacker,I.L., Stadler,P.F. and Kellis,M. (2012) RNA folding with soft constraints: Reconciliation of probing data and thermodynamic secondary structure prediction. *Nucleic Acids Res.*, **40**, 4261–4272.
4. Miao,Z., Adamiak,R.W., Antczak,M., Batey,R.T., Becka,A.J., Biesiada,M., Boniecki,M.J., Bujnicki,J.M., Chen,S.-J., Cheng,C.Y. *et al.* (2017) RNA-puzzles round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*, **23**, 655–672.
5. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
6. McGinnis,J.L., Dunkle,J.A., Cate,J. H.D. and Weeks,K.M. (2012) The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc.*, **134**, 6617–6624.
7. Sexton,A.N., Wang,P.Y., Rutenberg-Schoenberg,M. and Simon,M.D. (2017) Interpreting reverse transcriptase termination and mutation events for greater insight into the chemical probing of RNA. *Biochemistry (Mosc.)*, **56**, 4713–4721.
8. Hurst,T., Xu,X., Zhao,P. and Chen,S.-J. (2018) Quantitative understanding of SHAPE mechanism from RNA structure and dynamics analysis. *J. Phys. Chem. B*, **122**, 4771–4783.
9. Mlýnský,V. and Bussi,G. (2018) Molecular dynamics simulations reveal an interplay between SHAPE reagent binding and RNA flexibility. *J. Phys. Chem. Lett.*, **9**, 313–318.
10. Frezza,E., Courban,A., Allouche,A., Sargueil,B. and Pasquali,S. (2019) The interplay between molecular flexibility and RNA chemical probing reactivities analyzed at the nucleotide level via an extensive molecular dynamics study. *Methods*, **162–163**, 108–127.
11. Busan,S., Weidmann,C.A., Sengupta,A. and Weeks,K.M. (2019) Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry (Mosc.)*, **58**, 2655–2664.
12. Gherghe,C.M., Shajani,Z., Wilkinson,K.A., Varani,G. and Weeks,K.M. (2008) Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. *J. Am. Chem. Soc.*, **130**, 12244–12245.
13. Steen,K.-A., Rice,G.M. and Weeks,K.M. (2012) Fingerprinting noncanonical and tertiary RNA structures by differential SHAPE reactivity. *J. Am. Chem. Soc.*, **134**, 13160–13163.
14. Rice,G.M., Leonard,C.W. and Weeks,K.M. (2014) RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA*, **20**, 846–854.
15. Ehresmann,C., Baudin,F., Mougel,M., Romby,P., Ebel,J.P. and Ehresmann,B. (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res.*, **15**, 9109–9128.
16. Brunel,C. and Romby,P. (2000) Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol.*, **318**, 3–21.
17. Zaug,A.J. and Cech,T.R. (1995) Analysis of the structure of Tetrahymena nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA*, **1**, 363–374.

18. Paillart,J.-C., Dettenhofer,M., Yu,X.-F., Ehresmann,C., Ehresmann,B. and Marquet,R. (2004) First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.*, **279**, 48397–48403.
19. Cordero,P., Lucks,J.B. and Das,R. (2012) An RNA mapping data base for curating RNA structure mapping experiments. *Bioinformatics*, **28**, 3006–3008.
20. Moazed,D., Stern,S. and Noller,H.F. (1986) Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J. Mol. Biol.*, **187**, 399–416.
21. Romaniuk,P.J., de Stevenson,I.L., Ehresmann,C., Romby,P. and Ehresmann,B. (1988) A comparison of the solution structures and conformational properties of the somatic and oocyte 5S rRNAs of Xenopus laevis. *Nucleic Acids Res.*, **16**, 2295–2312.
22. Butcher,S.E. and Burke,J.M. (1994) Structure-mapping of the hairpin ribozyme. Magnesium-dependent folding and evidence for tertiary interactions within the ribozyme-substrate complex. *J. Mol. Biol.*, **244**, 52–63.
23. Cordero,P. and Das,R. (2015) Rich RNA structure landscapes revealed by mutate-and-map analysis. *PLoS Comput. Biol.*, **11**, e1004473.
24. Somarowthu,S., Legiewicz,M., Chillón,I., Marcia,M., Liu,F. and Pyle,A.M. (2015) HOTAIR forms an intricate and modular secondary structure. *Mol. Cell.*, **58**, 353–361.
25. Gross,L., Vicens,Q., Einhorn,E., Noireterre,A., Schaeffer,L., Kuhn,L., Imler,J.-L., Eriani,G., Meignin,C. and Martin,F. (2017) The IRES 5′UTR of the dicistrovirus cricket paralysis virus is a type III IRES containing an essential pseudoknot structure. *Nucleic Acids Res.*, **45**, 8993–9004.
26. Lorenz,R., Luntzer,D., Hofacker,I.L., Stadler,P.F. and Wolfinger,M.T. (2016) SHAPE directed RNA folding. *Bioinformatics*, **32**, 145–147.
27. Smola,M.J., Rice,G.M., Busan,S., Siegfried,N.A. and Weeks,K.M. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.*, **10**, 1643–1669.
28. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 97–102.
29. Zarringhalam,K., Meyer,M.M., Dotu,I., Chuang,J.H. and Clote,P. (2012) Integrating chemical footprinting data into RNA secondary structure prediction. *PLoS One*, **7**, e45160.
30. Turner,D.H. and Mathews,D.H. (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.*, **38**, D280–D282.
31. Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorith. Mol. Biol.: AMB*, **6**, 26.
32. Lavender,C.A., Lorenz,R., Zhang,G., Tamayo,R., Hofacker,I.L. and Weeks,K.M. (2015) Model-free RNA sequence and structure alignment informed by SHAPE probing reveals a conserved alternate secondary structure for 16S rRNA. *PLoS Comput. Biol.*, **11**, e1004126.
33. Spasic,A., Assmann,S.M., Bevilacqua,P.C. and Mathews,D.H. (2017) Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res.*, **46**, 314–323.
34. Herbreteau,C.H., Weill,L., Décimo,D., Prévôt,D., Darlix,J.-L., Sargueil,B. and Ohlmann,T. (2005) HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nat. Struct. Mol. Biol.*, **12**, 1001–1007.
35. James,L. and Sargueil,B. (2008) RNA secondary structure of the feline immunodeficiency virus 5′UTR and Gag coding region. *Nucleic Acids Res.*, **36**, 4653–4666.
36. Weill,L., Louis,D. and Sargueil,B. (2004) Selection and evolution of NTP-specific aptamers. *Nucleic Acids Res.*, **32**, 5045–5058.
37. Ding,Y. and Lawrence,C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.*, **31**, 7280–7301.
38. Sculley,D. (2010) Web-scale k-means clustering. In: *Proceedings of the 19th International Conference on World Wide Web (WWW'10)*. pp. 1177–1178.
39. Pedregosa,F., Varoquaux,G., Gramfort,A., Michel,V., Thirion,B., Grisel,O., Blondel,M., Prettenhofer,P., Weiss,R., Dubourg,V. *et al.* (2012) Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
40. Lu,Z.J., Gloor,J.W. and Mathews,D.H. (2009) Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA*, **15**, 1805–1813.
41. Mattson,C.A. and Messac,A. (2005) Pareto frontier based concept selection under uncertainty, with visualization. *Optim. Eng.*, **6**, 85–115.
42. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
43. Hajdin,C.E., Bellaousov,S., Huggins,W., Leonard,C.W., Mathews,D.H. and Weeks,K.M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 5498–5503.
44. Meyer,M., Nielsen,H., Olieric,V., Roblin,P., Johansen,S.D., Westhof,E. and Masquida,B. (2014) Speciation of a group I intron into a lariat capping ribozyme. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 7659–7664.
45. Lu,X.J., Bussemaker,H.J. and Olson,W.K. (2015) DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.*, **43**, e142.
46. Smit,S., Rother,K., Heringa,J. and Knight,R. (2008) From knotted to nested RNA structures: a variety of computational methods for pseudoknot removal. *RNA*, **14**, 410–416.
47. Cheng,C.Y., Kladwang,W., Yesselman,J.D. and Das,R. (2017) RNA structure inference through chemical mapping after accidental or intentional mutations. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 9876–9881.
48. Deforges,J., de Breyne,S., Ameur,M., Ulryck,N., Chamond,N., Saaidi,A., Ponty,Y., Ohlmann,T. and Sargueil,B. (2017) Two ribosome recruitment sites direct multiple translation events within HIV1 Gag open reading frame. *Nucleic Acids Res.*, **45**, 7382–7400.
49. Karabiber,F., McGinnis,J.L., Favorov,O.V. and Weeks,K.M. (2013) QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA*, **19**, 63–73.
50. Busan,S. and Weeks,K.M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA*, **24**, 143–148.
51. Gorodkin,J., Stricklin,S.L. and Stormo,G.D. (2001) Discovering common stem-loop motifs in unaligned RNA sequences. *Nucleic Acids Res.*, **29**, 2135–2144.
52. Xu,Z., Almudevar,A. and Mathews,D.H. (2012) Statistical evaluation of improvement in RNA secondary structure prediction. *Nucleic Acids Res.*, **40**, e26.
53. Cordero,P., Kladwang,W., Vanlang,C.C. and Das,R. (2012) Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry (Mosc.)*, **51**, 7037–7039.
54. Yu,A.M., Evans,M.E. and Lucks,J.B. (2018) Estimating RNA structure chemical probing reactivities from reverse transcriptase stops and mutations. bioRxiv doi: https://doi.org/10.1101/292532, 30 March 2018, preprint: not peer reviewed.
55. Wu,Y., Shi,B., Ding,X., Liu,T., Hu,X., Yip,K.Y., Yang,Z.R., Mathews,D.H. and Lu,Z.J. (2015) Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Res.*, **43**, 7247–7259.
56. Deng,F., Ledda,M., Vaziri,S. and Aviran,S. (2016) Data-directed RNA secondary structure prediction using probabilistic modeling. *RNA*, **22**, 1109–1119.
57. Reuter,J.S. and Mathews,D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinf*, **11**, 129.
58. Janssen,S. and Giegerich,R. (2015) The RNA shapes studio. *Bioinformatics (Oxford, England)*, **31**, 423–425.