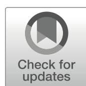

Chapter 13

Auto-DRRAFTER: Automated RNA Modeling Based on Cryo-EM Density

Haiyun Ma, Phillip Pham, Bingnan Luo, Ramya Rangan, Kalli Kappel, Zhaoming Su, and Rhiju Das

Abstract

RNA three-dimensional structures provide rich and vital information for understanding their functions. Recent advances in cryogenic electron microscopy (cryo-EM) allow structure determination of RNAs and ribonucleoprotein (RNP) complexes. However, limited global and local resolutions of RNA cryo-EM maps pose great challenges in tracing RNA coordinates. The Rosetta-based "auto-DRRAFTER" method builds RNA models into moderate-resolution RNA cryo-EM density as part of the Ribosolve pipeline. Here, we describe a step-by-step protocol for auto-DRRAFTER using a glycine riboswitch from *Fusobacterium nucleatum* as an example. Successful implementation of this protocol allows automated RNA modeling into RNA cryo-EM density, accelerating our understanding of RNA structure–function relationships. Input and output files are being made available at https://github.com/auto-DRRAFTER/springer-chapter.

Key words RNA structure, Cryo-EM, Auto-DRRAFTER, Rosetta computational modeling

1 Introduction

RNAs can fold into intricate three-dimensional (3D) structures to perform diverse biological functions in the presence or absence of proteins [1, 2]. Single particle cryogenic electron microscopy (cryo-EM) is routinely used now to determine RNA and ribonucleoprotein (RNP) structures at near-atomic resolution [3–5]. However, limited resolution in flexible global and local regions of RNAs prevents accurate RNA modeling and poses challenges in elucidating the RNA structure–function relationship [6, 7].

Computational modeling and de novo design of RNA structures is in early stages compared to protein modeling, but continues to improve due to expanding experimental methods using X-ray crystallography [8], NMR [9], small-angle X-ray scattering (SAXS) [10], cryo-EM [11], and development of RNA modeling software

Jienyu Ding et al. (eds.), *RNA Structure and Dynamics*, Methods in Molecular Biology, vol. 2568,
https://doi.org/10.1007/978-1-0716-2687-0_13,
© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2023

including Assemble2 [12], MC-sym [13], RACER [14], Phenix [15], Rosetta [16], VfoldLA [17], RNAComposer [18], NAST [19], SimRNA [20], and iFoldRNA [21]. Driving progress, the RNA-Puzzles are held routinely to provide rigorous evaluations on different aspects of existing RNA modeling software, including homology modeling, fragment assembly, and de novo prediction [22]. Recent advances in the application of machine learning algorithms to RNA modeling may also improve the accuracy of de novo RNA structure prediction [23, 24], though a scarcity of RNA 3D structures compared to protein structures may delay dramatic advances from machine learning methods, which depend on large data sets.

Development of single particle cryo-EM has allowed determination of RNA and RNP complexes at increasing throughput. However, in work so far, the majority of RNA structures are resolved at moderate to low resolution (~ 4–10 Å), which limits the use of currently available de novo modeling algorithms based on cryo-EM density at near-atomic (better than 4 Å) resolution [5, 16] and also makes manual tracing of RNA coordinates tedious or even intractable. We previously developed DRRAFTER to build RNA coordinates de novo in biologically important RNP complexes, in which RNAs were resolved at relatively low resolution [25]. DRRAFTER contains an initial manual setup that requires users' knowledge and experience in RNA modeling, which may crucially affect the performance and final results. In the recent Ribosolve pipeline, we developed an updated version of DRRAFTER, named auto-DRRAFTER, to automatically build RNA models into moderate-resolution cryo-EM maps [11]. The input to auto-DRRAFTER includes cryo-EM density maps and secondary structure information that are generated from mutate-and-map read out by next generation sequencing (M2-seq). The three techniques M2-seq, cryo-EM, and auto-DRRAFTER comprise the Ribosolve pipeline. All-atom models are generated in auto-DRRAFTER with the accuracy predicted by overall modeling convergence. Uncertain regions are identified by comparing the per-residue-convergence and real-space cross correlation (CC) between the map and the model (Fig. 1). Using auto-DRRAFTER we determined 11 previously unknown protein-free RNA structures ranging from 119 to 388 nucleotides (nt), demonstrating its ability to rapidly and accurately model protein-free RNA structures [11].

2 Materials

2.1 Recommended Computing Requirement

Typical use of auto-DRRAFTER requires a high-performance computing cluster. While runs may be executed interactively, most auto-DRRAFTER runs benefit from using a job scheduler such as SLURM or Torque.

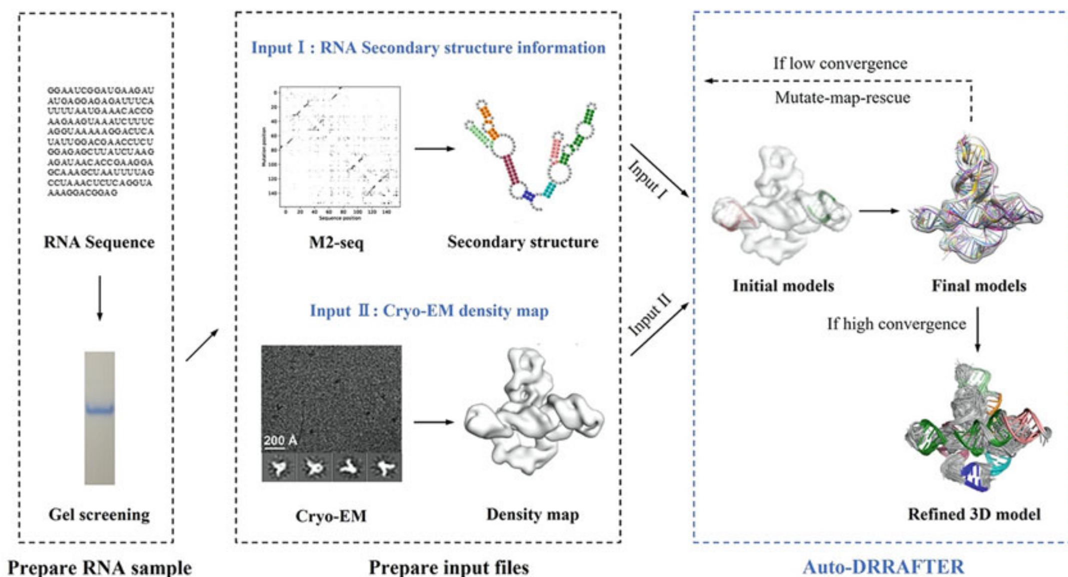

Fig. 1 Auto-DRRAFTER requires RNA secondary structure and cryo-EM density map as input to generate all-atom RNA models

All jobs were executed on SKLB Duyu high performance computing center at Sichuan University and Sherlock 2.0 high performance computing cluster at Stanford University. One RNA model normally takes 2–5 auto-DRRAFTER rounds to converge to the final model building task that takes another two rounds. Each auto-DRRAFTER round contains 150 jobs per helix placement of one end node (described below), with each job using one processor core available on either Duyu or Sherlock (Intel 6230, Intel E5-2640v4, Intel 5118 or AMD 7502). In the case of *F. nucleatum* glycine riboswitch, it took 144 computing hours (a total of six rounds with 24 h per round) on 600 cores (a total of four end nodes with 150 jobs per end node).

2.2 Installing Software

2.2.1 Rosetta (auto-DRRAFTER)

1. The auto-DRRAFTER software is freely available to academic users as part of the Rosetta software suite at www.rosettacommons.org. (*see* **Note 1**).
2. Documentation is available at https://www.rosettacommons.org/docs/latest/application_documentation/rna/auto-drrafter.
3. Demo is available at https://www.rosettacommons.org/demos/latest/public/auto-drrafter/README.
4. A limited version of the software is also freely available through an online ROSIE server at https://rosie.rosettacommons.org/auto-drrafter.

5. Install Rosetta RNA tools. See instructions and documentation here: https://new.rosettacommons.org/docs/latest/application_documentation/rna/RNA-tools.

2.2.2 Python

Make sure that you have Python version 2.7.16 installed, then install networkx (https://networkx.org) and mrcfile (https://mrcfile.readthedocs.io). To check whether the installation is successful, run the "*pip install networkx mrcfile*"command.

2.2.3 UCSF Chimera

Download and install Chimera version 1.1.4 at http://www.cgl.ucsf.edu/chimera/download.html.

2.2.4 EMAN2

EMAN2 is used by auto-DRRAFTER under the hood to process cryo-EM maps. Download and install EMAN2 version 2.31 (Version 2.2 is also available) at https://blake.bcm.edu/emanwiki/EMAN2/Install. And run the "*e2proc3d.py -h*" and "*e2Segment3d.py -h*" commands to check whether e2proc3d.py and e2Segment3d.py are successfully installed.

2.3 Prepare Input Files

The RNA 3D model is derived from the sequence (fasta.txt) and secondary structure in dot-bracket format (secstruct.txt), with cryo-EM density (map.mrc) as additional 3D experimental restraints (*see* Table 1 and files deposited at https://github.com/auto-DRRAFTER/springer-chapter). The RNA secondary structure can be obtained from literature analyses on the target molecule that take into account covariance or, if such data are unavailable, prediction methods such as CONTRAfold (http://contra.stanford.edu/contrafold/), RNAFold (available at http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi), or RNAstructure (available at https://rna.urmc.rochester.edu/RNAstructure.html). More accurate predictions can be achieved through experimental restraints from chemical probing data using selective hydroxyl acylation analyzed by primer extension (SHAPE) [26] or dimethyl sulfate (DMS) [27] mapping, which can be used as inputs in secondary structure prediction algorithms. In the Ribosolve pipeline, RNA secondary structure files are generated from M2-seq, a method that monitors changes in chemical reactivity at each residue as every other residue is mutated. Compared to conventional chemical probing approaches, M2-seq provides more accurate secondary structures as well as uncertainties from nonparametric bootstrapping [28, 29]. We note that high certainty secondary structures are currently a prerequisite to achieving confident tertiary structures from cryo-EM with resolutions worse than 4 Å. Cryo-EM maps are obtained with or without Volta phase plate on a 200 kV or 300 kV microscope. Here we describe a detailed protocol of the Rosetta-based auto-DRRAFTER that utilizes the above mentioned input files to automatically generate all-atom RNA models.

Table 1
A brief explanation of the input files

Input files	Brief explanation
fasta.txt (This example uses FNG_riboswitch.fasta)	The FASTA file, which lists the full sequence of your RNA molecule. It should contain one line that starts with ">" and lists the chain and residue numbers for the sequence. The RNA sequence should be specified with lower-case letters. Currently auto-DRRAFTER can only handle single chain RNAs.
secstruct.txt (This example uses FNG_riboswitch.txt)	A file containing the secondary structure of the complex in dot-bracket notation. The secondary structure should be the same length as the sequence found in the fasta file. This secondary structure will be enforced during the auto-DRRAFTER modeling. We recommend verifying predicted secondary structures biochemically with M2-seq.
map.mrc (This example uses FNG_riboswitch.mrc)	The density map file in MRC format. It can be obtained on 200 kV or 300 kV microscopes by collecting cryo-EM images with or without the use of a Volta phase plate.
job_submission_template.sh	This job submission script uses the conventional directives that specify the resources used to run each round of auto-DRRAFTER and should reflect the job scheduler used on the cluster you are planning to run your job on (*e.g.*, SLURM or TORQUE). If running interactively, the *job_submission_template.sh* file can be an empty file.

3 Methods

3.1 Auto-DRRAFTER Overview

The auto-DRRAFTER program starts by importing the secondary structure information of RNA (Fig. 2a) and the cryo-EM density map (Fig. 2b). Firstly, auto-DRRAFTER identifies end nodes in the density map – regions in which the map contains terminal hairpin loops. To identify these nodes, the density map is low-pass filtered to 20 Å and manually adjusted to the appropriate threshold so that the double-helix regions look like rods with their ends easier to identify. Stem loop regions in the secondary structure are automatically selected as candidates that will be placed at each end node (Fig. 2c), which are labeled as red spheres in the low-pass filtered cryo-EM map (Fig. 2d). Next, a single end node is randomly selected (circled in Fig. 2d) as a starting point to place each stem loop candidate from the secondary structure (black circle regions in Fig. 2c). An idealized A-form helix is generated for each stem loop candidate probe helix (probe helix colored black in Fig. 2e) and automatically placed in the selected end node of the original cryo-EM map (boxed in Fig. 2e). The resulting probe helices in the original cryo-EM maps are used as initial models in the subsequent fragment-based RNA folding method in auto-DRRAFTER to generate all-atom RNA models (Fig. 2f). Hundreds of models are generated in the iterative RNA fragment assembly and refinement

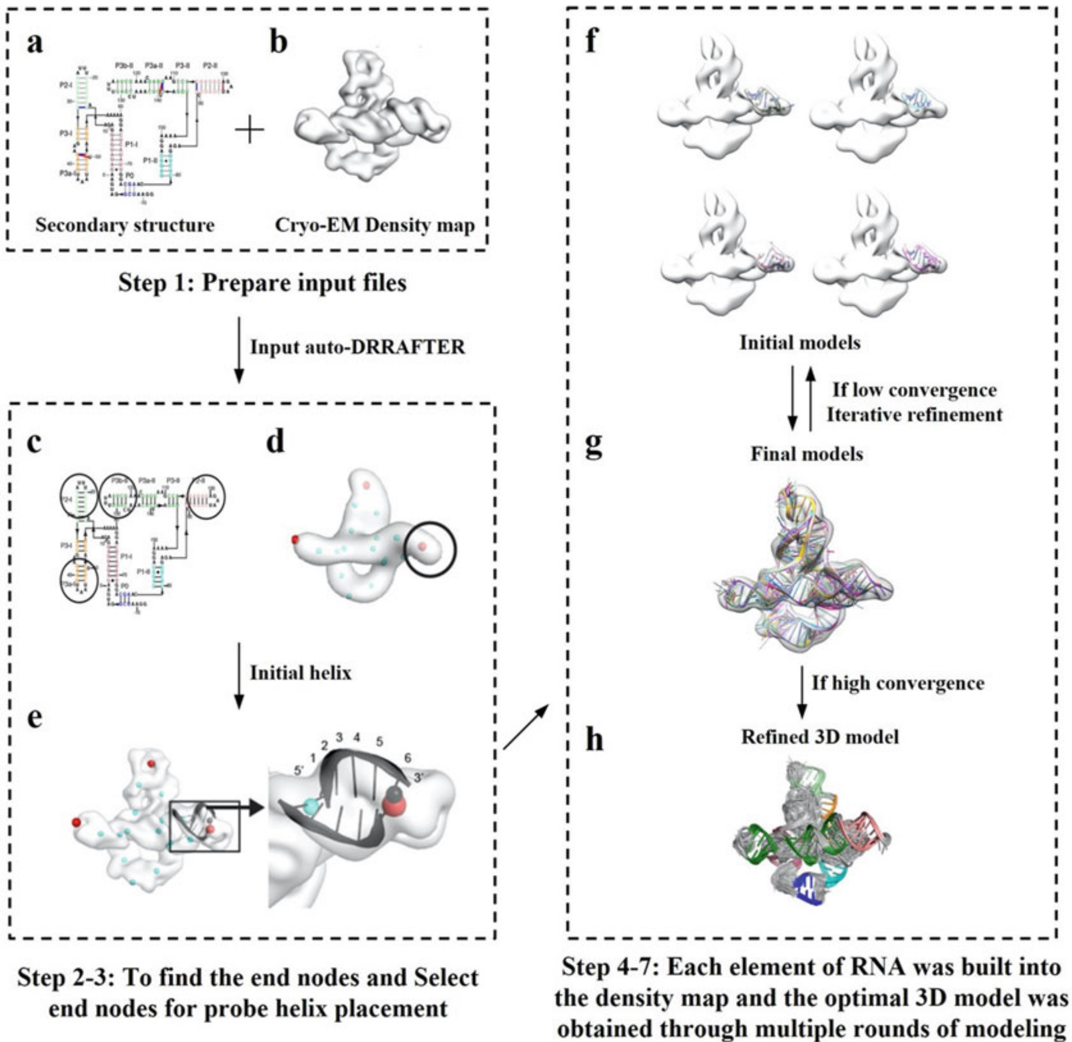

Fig. 2 Auto-DRRAFTER workflow. *F. nucleatum* glycine riboswitch is shown here as an example. RNA secondary structure (**a**) and the cryo-EM map (**b**) are input files for auto-DRRAFTER. Next, end node candidates in the secondary structure (circled in **c**) and end nodes in the low-pass filtered cryo-EM map (red spheres in (**d**) are identified; blue spheres are used to label stems. Then, the end node candidate stem model (probe helix) is placed in the randomly selected end node in the original cryo-EM map and optimized by detecting the distance between the C1′ atom of nucleotide 6 (black sphere) and the end node (black box and black helix in (**e**). All four probe helices serve as starting point to iteratively generate all-atom RNA models till convergence (**f**, **g**). The model with highest overall score yields the final model (**h**)

rounds, and these models are sorted according to a score function that includes both physics-based score terms and the fit between model fragments and the cryo-EM map. Areas with sufficient structural consensus are fixed in the next iteration. This automated process continues until the entire model is confidently built (Fig. 2f, g). Finally, the top models obtained from the last round of modeling is optimized to yield the final model set (Fig. 2h).

3.2 Step-by-Step Method Details

3.2.1 Prepare Input Files Including RNA Secondary Structure (Fig. 2a) and Cryo-EM Density Map (Fig. 2b)

1. The dot-bracket RNA secondary structure file (FNG_riboswitch.txt), which can be generated from prediction software and websites, or from M2-seq protocol that has been previously described [28, 29].
2. Cryo-EM RNA density map (FNG_riboswitch.mrc). Detailed protocol of cryo-EM data processing is described in another chapter.

3.2.2 Low-Pass Filter the Cryo-EM Density Map and Determine an Appropriate Threshold Level (Fig. 2d) to Find End Nodes (Fig. 2c, Black Circles)

1. Go to the project directory and type the command shown below (*see* Table 2):

```
python $ROSETTA/main/source/src/apps/public/DRRAFTER/auto-DRRAFTER_setup.py -map_thr 20 -full_dens_map input_files/FNG_riboswitch.mrc -full_dens_map_reso 10.0 -fasta input_files/FNG_riboswitch.fasta -secstruct input_files/FNG_riboswitch.txt -out_pref FNG_riboswitch -rosetta_directory $ROSETTA/main/source/bin/ -nstruct_per_job 200 -cycles 30000 -fit_only_one_helix -rosetta_extension .static.linuxgccrelease -just_low_pass
```

Table 2
A brief explanation of flags on the typed command

Command flags	Brief explanation
-map_thr	The density threshold at which the detection of optimal helix placement locations will take place. This value is a placeholder for now until we find the best value to use in a later step
-full_dens_map	The path to our density map.
-full_dens_map_res	The resolution of the density map in Å.
-fasta	The path to our fasta file listing the sequence of our RNA molecule.
-secstruct	The path to our secondary structure of the RNA molecule in dot-bracket notation.
-out_pref	The arbitrary prefix for output files generated by auto-DRRAFTER.
-rosetta_directory	The location of the Rosetta executables.
-rosetta_extension	The extension of your Rosetta executables. This extension depends on how Rosetta was installed (look in your Rosetta directory to confirm the extension for your executables).
-repeats	The number of independent attempts to place the helices. 10 is usually a good number for this setting.
-shift_center	This is a setting that is required if you are using a simulated density map that was generated with EMAN2. It should not be necessary in other situations, such as our example in this chapter.

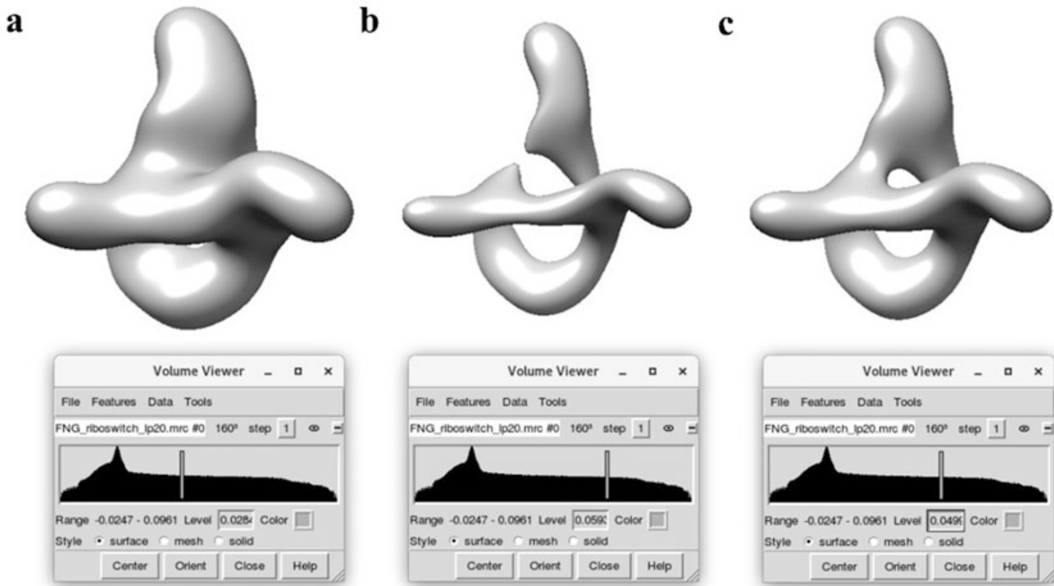

Fig. 3 Use the Volume Viewer tool in Chimera to determine the appropriate threshold level. (**a**) The density threshold shown is lower (0.0284). Although we can see that the end of the spiral may be on this map, we may go to a higher threshold and these areas will be easier to identify. (**b**) The density threshold (0.0593) is too high, and the whole map is no longer connected. (**c**) The threshold (0.0499), works well since the map is still fully connected, and we can clearly see where the ends of the helix/hairpin will be located. Therefore, we will use 0.0499 as the threshold value

Time: Running this command takes less than 20 s (on Intel Gold 6230 2.10 GHz).

2. This will create a single file: FNG_riboswitch_lp20.mrc, indicating that the density map is low-pass filtered to 20 Å, which will be used to figure out the end nodes for candidate placements of probe helices.
3. The appropriate threshold level can be determined using the Volume Viewer Tool in Chimera (Fig. 3). Open the FNG_Riboswitch_lp20.mrc file in Chimera and choose Tools > Volume Data > Volume Viewer. Find the "end node" in the density map that can be clearly identified by changing the threshold value of the map; generally, this value corresponds with the highest threshold value that does not break the continuity of the map. Record the threshold value.
4. This threshold value is only used for placing the probe helix and not for subsequent steps. Here, we used 0.0499 as the threshold value.

3.2.3 Select End Nodes for Probe Helix Placement (Fig. 2d, e)

1. Set up the auto-DRRAFTER run by typing the following command (*see* **Note 2**):

```
python $ROSETTA/main/source/src/apps/public/DRRAFTER/auto-DRRAFTER_setup.py -map_thr 0.0499 -full_dens_map input_files/FNG_riboswitch.mrc -full_dens_map_reso 10.0 -fasta input_files/FNG_riboswitch.fasta -secstruct input_files/FNG_riboswitch.txt -out_pref FNG_riboswitch -rosetta_directory $ROSETTA/main/source/bin/ -nstruct_per_job 200 -cycles 30000 -fit_only_one_helix -rosetta_extension .static.linuxgccrelease
```

Time: about 3–4 min (on Intel Gold 6230 2.10 GHz).

2. After the above command is executed, the information in Fig. 4 will be entered on the screen and a series of files (*see* Table 3) will be generated.
3. This command will generate an *FNG_riboswitch_init_points.pdb* file (*see* Table 3) that consists of spheres fitted throughout the cryo-EM map (Fig. 2d); each sphere has a numeric label. The user can manually open the *FNG_riboswitch_init_points.pdb* file in Chimera or PyMOL and locate the spheres as selected end nodes. For illustration purpose, we colored the potential end node sphere in red and spheres in the middle of

```
Using only one end node!
Low-pass filtering the map to 20A.
Converting density map to graph.
Possible end nodes in the map:
1 5 11
You can visualize the end nodes in FNG_riboswitch_init_points.pdb
You can specify which of these end nodes you'd like to use with -use_end_node
Converting secondary structure to graph.
Mapping secondary structure to density map.
Setting up DRRAFTER runs.
Making full helix H7
Making full helix H4
Making full helix H10
Making full helix H15
```

Fig. 4 The printed information after the command is executed. Firstly, the input density is low-pass filtered to 20 Å and points are placed throughout the density map. These points are then converted into a graph to identify possible end nodes. These end nodes are printed out to the console, which here are nodes 1, 5, and 11. All the points that were placed into the density map are written to a PDB file (here *FNG_riboswitch_init_points.pdb*) that we can visualize in Chimera or PyMOL. Each point has a different number, which corresponds to a node number. Loading the PDB with these points into Chimera (or any other molecular visualization software), residue 1, 5, and 11 can be inspected to see which nodes auto-DRRAFTER has identified as potential end nodes

Table 3
Output files after the command is executed

Output files	Brief explanation
settings_FNG_riboswitch.txt	This file lists all of the settings that were used to set up this auto-DRRAFTER run.
fasta_FNG_riboswitch.txt	The fasta file for the DRRAFTER runs. This is basically the same as the input fasta file, but the numbering has been changed to start at 0 – this is just an auto-DRRAFTER convention. A final step at the end of the auto-DRRAFTER modeling will map these residues numbers back to the input residue numbers.
all_aligned_FNG_riboswitch_0.REORDER.pdb all_aligned_FNG_riboswitch_1.REORDER.pdb all_aligned_FNG_riboswitch_2.REORDER.pdb all_aligned_FNG_riboswitch_3.REORDER.pdb	These are four possible hairpin placements in the density map (visualized above). Each model will be used in a separate DRRAFTER run.
command_FNG_riboswitch_*_R1	The commands for the two DRRAFTER runs for the four possible alignments of the helices into the density map.
flags_FNG_riboswitch_*_R1	These files contain all the flags that will be used for the four DRRAFTER runs for the different helix alignments in the density map.
FNG_riboswitch_H*.pdb	Ideal A-form helices for helix.
FNG_riboswitch_H*_full.out.1.pdb	The same ideal A-form helices from above with placeholder coordinates added for the hairpins.
FNG_riboswitch_auto_fits.txt	This file lists all of the different possible alignments of helices into the density map. Each alignment is numbered and listed on a separate line in this file.
FNG_riboswitch_init_points.pdb	This PDB file contains all of the points that were placed into the density map in order to convert the density map into a graph. This is the file that we visualized later in the chapter to check the assignment of end nodes.
secstruct_FNG_riboswitch.txt	The secondary structure file for the DRRAFTER runs.

stems in blue (Fig. 5). Auto-DRRAFTER randomly chooses one of these end nodes (red spheres) to place the probe helices; however, if a specific end node is preferred, the user may append the flag "*-use_end_node*" at the end of above command.

4. This command also converts the RNA secondary structure (that we specified) to a graph in which helices are represented as edges and junctions and loops are represented as nodes. This secondary structure graph is then mapped onto the graph for the density map. The location of the probe helix was optimized

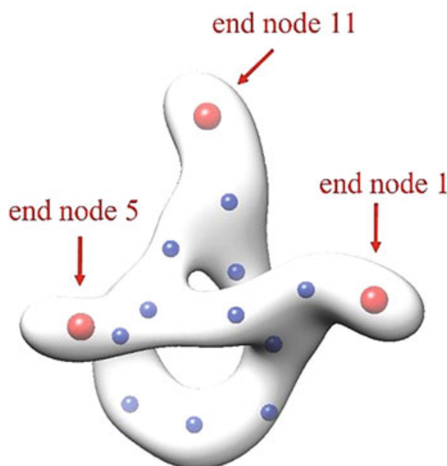

Fig. 5 Visualize all spheres placed in the density map in Chimera. Each point has a numerical label, which corresponds to a node number. We enlarged and colored the end nodes red and other spheres in the middle of stems blue

by minimizing the distance between the C1′ atom of 6 nucleotide in the probe helix and the end node sphere (Fig. 2e). Four possible hairpin placements can be checked in Chimera by opening the density map along with the helix coordinates which are in PDB files with the prefix "*all_aligned_FNG_riboswitch*" (Fig. 6). It is essential to ensure that the helices are placed well in the density map. If the fitting is not appropriate, the remaining steps will likely not yield reasonable results.

3.2.4 Build Initial Models with Different End Nodes (Fig. 2f)

1. Auto-DRRAFTER then continues to build a complete RNA model by stochastically placing the remaining helices and nucleotides from the secondary structure into the cryo-EM density map.
2. Type and run the following command (*see* Table 4):

```
python $ROSETTA/main/source/src/apps/public/DRRAFTER/submit_jobs.py -out_pref FNG_riboswitch -curr_round Ri -njobs 150 -template_submission_script input_files/job_submission_template.sh -queue_command sbatch
```

Time: Typically, one round of modeling takes 12–48 h, depending on the number of models built for each job and how much computing resources are used (*see* Subheading 2.1, Recommended Computing Requirement).

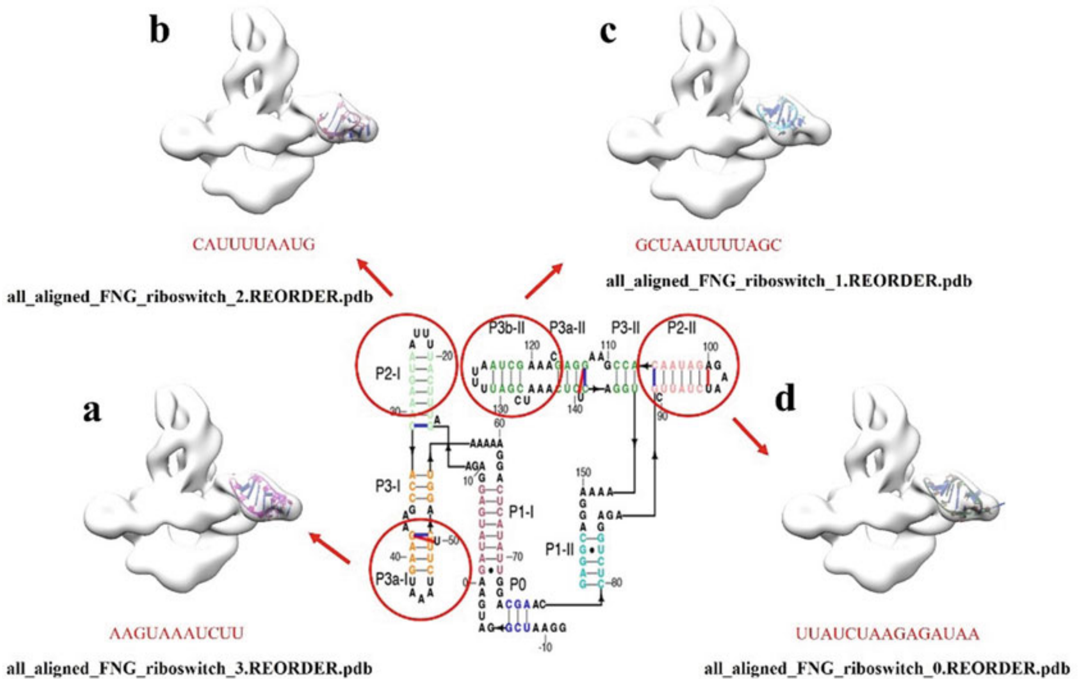

Fig. 6 Visualize the density map of helices in Chimera to check that helices are properly positioned. Four hairpins in the secondary structure are mapped to the end node (we use end node 1 as an example in the figure.). All candidate hairpin placements in the end node will be used to generate models, and an optimized placement will be selected later. (**a**) The corresponding sequence is: AAGUAAAUCUU. (**b**) The corresponding sequence is: CAUUUUAAUG. (**c**) The corresponding sequence is: GCUAAUUUUAGC. (**d**) The corresponding sequence is: UUAUCUAAGAGAUAA

3. This will create the following output directories and files:

job_files: This is a directory that contains all the job submission files.

out_FNG_riboswitch_n_Ri: This directory contains all the DRRAFTER models built starting from probe helix n, where n is an integer to indicate the probe helix number; i indicates the round number.

3.2.5 Iterative Refinement of RNA Models (Fig. 2f, g)

1. Collect all models built in the previous step and calculate overall convergence using the command shown below (*see* **Note 3**):

```
python $ROSETTA/main/source/src/apps/public/DRRAFTER/auto-DRRAFTER_setup_next_round.py -out_pref FNG_riboswitch -curr_-round Ri -rosetta_directory $ROSETTA/main/source/bin/ -conver-gence_threshold 10 -rosetta_extension .static.linuxgccrelease
```

Table 4
A brief explanation of flags in the command above

Flags	Brief explanation
-out_pref	The prefix used for all output files from this run. This should be the same –out_pref that was used in the setup command in Subheadings 3.2.2 and 3.2.3.
-curr_round	This is the round of modeling that is currently being performed. We have not done any modeling yet, so this is round 1. This should always be "R" followed by a number that indicates the round number.
-njobs	This is the number of jobs that will be run. The number of jobs submitted by "-njobs" depends on the real modeling scenario, for a very simple system we used two jobs, but for larger systems we typically use more than 100 jobs. Each job will build the number of models that was specified in the setup command above (−nstruct_per_job), in this case 200 structures per job. We will therefore build 30,000 models in total (200 structures per job × 150 jobs).
-template_submission_script	This flag points to the path of the job submission script, which specifies the resources that auto-DRRAFTER jobs will have access to.
-queue_command	This is the command that will be used to submit the job files to a cluster queuing system. The value for "-queue_command" flag depends on the job scheduler used. For TORQUE computing clusters, the value for the flag would be "qsub" while for SLURM-based clusters, "sbatch" would be used instead. Conversely, "source" is used to run a command interactively. In this example, we use "sbatch" since our computing cluster is based on SLURM.

2. After executing the above command, "Overall Convergence" and "Density Threshold" (*see* **Note 3** and Fig. 7a) are displayed on the screen as well as output files generated (*see* Table 5). This step collects all of the models from the previous step and calculates the convergence of the overall top ten scoring models (across all alignments). Then the next round of modeling is set up based on the models that were built from the previous round. Regions of the models that are well converged will be kept fixed in the next round of modeling.
3. This step outputs several files (*see* Table 5), of which the "FNG_riboswitch_all_models_all_fits_Ri.out.*" PDB files are the overall best scoring models from all of the different alignments in the first round of modeling. We should pay attention to the overall convergence score and how these models fit in the density map (Fig. 7a, b). The first round models do not need to look perfect, but they should roughly fit in the density map.
4. If the overall convergence value is larger than 10 Å, multiple iterations (Subheadings 3.2.4 and 3.2.5) need to be performed

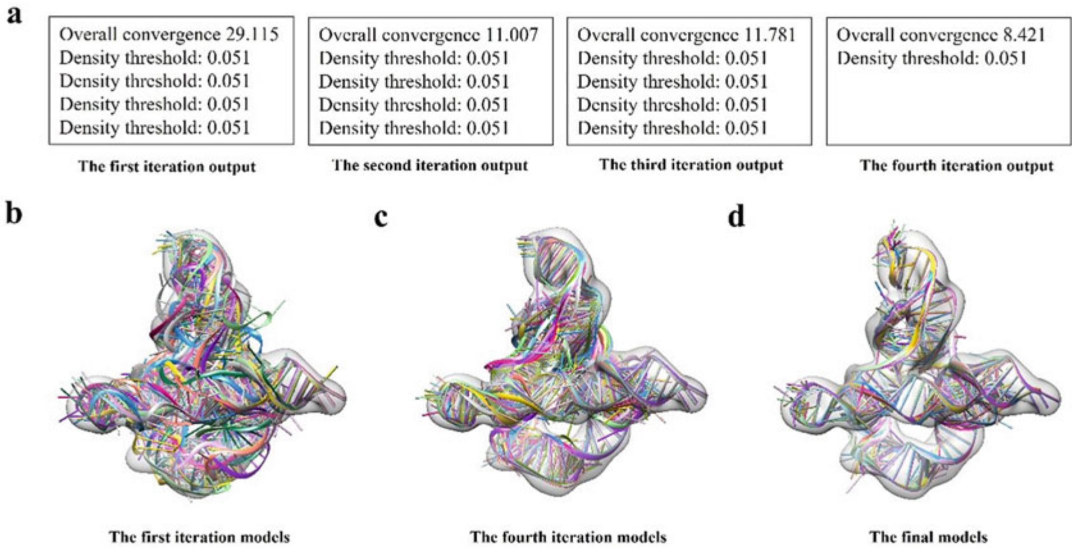

Fig. 7 The main output results in the six rounds of modeling. (**a**) The Overall Convergence and Density Threshold values printed after each iteration. (**b**) Visualize the 10 models generated in the first round of modeling in Chimera to check if models are properly positioned. (**c**) The 10 models generated in round 4 modeling. (**d**) The final models generated in round 6 modeling

Table 5
The output file after the command is executed

Output files	Brief explanation
FNG_riboswitch_all_models_all_fits_Ri.out.*.pdb	These are the overall best scoring models from this round of modeling. It is recommended that we visualize these models in the density map after each round.
command_FNG_riboswitch_Ri	This is the command file for the next round of modeling
flags_FNG_riboswitch_Ri	This file contains all the flags that will be used for each DRRAFTER run for the different helix alignments in the density map.
all_fit_FNG_riboswitch_Ri.REORDER.pdb	This PDB file contains initial helix placements for the next round of modeling.
convergence_FNG_riboswitch_all_models_Ri.txt	This file lists the convergence for the overall top scoring models.
FNG_riboswitch_Ri_CAT_ALL_ROUNDS.out.*.pdb	These are the top scoring models from each of the possible initial helix placements.

until the convergence value is less than 10 Å. In each of these modeling iterations, regions of the RNA models that have converged are kept fixed to focus sampling on segments that remain highly variable.

5. In the FNG_riboswitch example, four iterations are needed by modifying i = 1,2,3 and 4. After four iterations, the overall convergence value was 8.421, and the Density threshold value was 0.051 (Fig. 7a). Visualize the FNG_riboswitch_all_models_all_fits_R4.out.*.pdb file in Chimera to check the 10 models generated (Fig. 7c). Please note that all numbers here (*e.g.*, convergence score, number of iterations) may vary in different trials of the exact same modeling task.

3.2.6 Final Two Rounds of Modeling (Fig. 2g)

1. When the overall convergence of models is less than 10 Å, the user can proceed to the final two rounds of modeling (*see* **Note 4**).
2. In the first final round, all regions of the RNA are allowed to move from their initial positions to achieve further refinement of the modeled coordinates. Run the command below to begin the first final round of modeling (*see* **Note 5**):

```
python $ROSETTA/main/source/src/apps/public/DRRAFTER/submit_jobs.py -out_pref FNG_riboswitch -curr_round FINAL_R5 -njobs 150 -template_submission_script input_files/job_submission_template.sh -queue_command sbatch
```

3. Set up the last round of modeling with the following command. This command extracts converged regions from the models with highest scores in the previous iteration, setting up the last round of modeling (*see* **Note 6**):

```
python $ROSETTA/main/source/src/apps/public/DRRAFTER/autoDRRAFTER_setup_next_round.py -out_pref FNG_riboswitch -curr_round FINAL_R5 -rosetta_directory $ROSETTA/main/source/bin/ -rosetta_extension .static.linuxgccrelease
```

4. In the second final round, converged regions from the previous round are held fixed during fragment assembly steps of auto-DRRAFTER, and they are only allowed to move during high-resolution refinement steps. Run the final round of modeling with the following command:

```
python $ROSETTA/main/source/src/apps/public/DRRAFTER/submit_jobs.py -out_pref FNG_riboswitch -curr_round FINAL_R6 -njobs 150 -template_submission_script input_files/job_submission_template.sh -queue_command sbatch
```

5. Retrieve the top-scoring final models using the command below:

```
python $ROSETTA/main/source/src/apps/public/DRRAFTER/auto-DRRAFTER_setup_next_round.py -out_pref FNG_riboswitch -curr_round FINAL_R6 -rosetta_directory $ROSETTA/main/source/bin/ -rosetta_extension .static.linuxgccrelease
```

6. The output of this command is: DONE building models for FNG_riboswitch.

3.2.7 Completing the Modeling Procedure

1. Type the command shown below (*see* **Note 7**):

```
python $ROSETTA/main/source/src/apps/public/DRRAFTER/finalize_models.py -fasta input_files/FNG_riboswitch.fasta -out_pref FNG_riboswitch -final_round FINAL_R6
```

2. This should print "Done finalizing models" to the screen, indicating that the modeling is complete. This creates final models (Fig. 7d): *FNG_riboswitch_all_models_all_fits_FINAL_R6.out.*.pdb.*
3. The final models should be carefully examined and validated. If there are large parts of the model that are outside the density map, or large regions of the map have missing coordinates, the model is likely not properly built. In this case, some additional attempts may be useful, such as specifying a different end node as the initial helical probe for remodeling, verifying the secondary structure of RNA through further experiments, or modeling by using the more definitive structure as an initial position and so on. In addition, the obtained model can be verified in a variety of ways, with additional biochemical or functional experiments, or through comparison to homologous structures or any available crystal structures of subregions.
4. We can use the linear relationship between modeling convergence and predicted root mean square deviation (RMSD) from the native structure (estimated accuracy = $0.61\times$ convergence +2.4) to estimate model accuracy using convergence value. The auto-DRRAFTER modeling convergence, defined as the average pairwise RMSD of the top ten scoring models, should be correlated with the model RMSD. In this case, the overall convergence of the final ten best models after the last round (FINAL_R6) of modeling is 2.833 Å, yielding the final estimated accuracy of 4.128 Å.

4 Notes

1. Rosetta is free for academic users, but you need permission before you can download it (https://els2.comotion.uw.edu/product/rosetta). auto-DRRAFTER is available in the Rosetta

weekly releases after 2019.47 (it is not available in 2019.47) and files for this example are available in releases after 2020.03. auto-DRRAFTER is not available in Rosetta 3.11.

2. This is the same command we used in the previous step, with the -just_low_pass flag removed and the "-map_thr" updated with the value we recorded in Subheading 3.2.2. Depending on where the helix appears, we may need to try different thresholds. In this particular case, "-map_thr 0.0555" also yielded a similar result.

3. The default and recommended value for "convergence_threshold" is 10. To try out the Ribosolve workflow for the first time, a more lenient threshold of 100 may be used to reduce runtime and assess whether the installation is successful. The "Density threshold" is a critical value that is determined automatically, which is used to determine which regions of the RNA are sufficiently converged to be kept fixed during the next round of modeling. The value that is printed to the screen is the value that was determined automatically. Sometimes, particularly for higher-resolution maps (better than about ~6 Å resolution) or for maps that are relatively noisy, this automatically selected value is too high. This can make it so that regions that are well converged and do fit reasonably well in the density map are not kept fixed in subsequent rounds of modeling, making it much more challenging for the models to converge overall. You can check whether this might be a problem in your case by examining the top scoring models from each fit from the last round (FNG_riboswitch_Ri_CAT_ALL_ROUNDS.out.*.pdb) as well as the PDB files containing the regions that will be kept fixed for the subsequent round of modeling (all_fit_FNG_riboswitch_Ri.REORDER.pdb). If there are regions of your models that look sufficiently converged, but do not show up in the "all_fit_FNG_riboswitch_Ri.REORDER.pdb" files, then you might want to rerun this step and use the flag "-dens_thr" to select a lower value for the density threshold. It often makes sense to try reducing the density threshold by ~50%. If you still see a similar problem, you can reduce the density threshold even further.

4. When the convergence has reached below 10 Å, the auto-DRRAFTER_setup_next_round.py command will produce a command file with the FINAL keyword, indicating that the user can proceed to the final two rounds of modeling. If the convergence remains over 10 Å and no longer improves, the auto-DRRAFTER program will skip to the final modeling.

5. This is basically the same command from Subheading 3.2.4 except "–curr_round" has changed to FINAL_R5. Note that if the modeling has not reached the FINAL stages, "R5" should

be used in place of "FINAL_R5." This creates files very similar to those from Subheading 3.2.5. Again, we should check the models that we built (FNG_riboswitch_all_models_all_fits_FINAL_R5.pdb).

6. In this example, the output of this command is: Overall convergence 4.484, density threshold: 0.051. Convergence was 4.0 in the original Ribosolve pipeline [11].
7. -final_round is the round of modeling that was just completed, in this case FINAL_R6.

Acknowledgments

Computational tasks were performed on SKLB Duyu high performance computing center in Sichuan University and Sherlock 2.0 high performance computing cluster in Stanford University. This work was supported by Ministry of Science and Technology of China (MoST) 2021YFA1301900, Natural Science Foundation of China (NSFC) 82041016 and 32070049, and Sichuan University start-up funding 20822041D4057 to Z.S., the National Science Foundation (NSF) Graduate Research Fellowship award DGE-1656518 to R.R., the Gerald J. Lieberman Fellowship to R. R., the Rosetta REU NSF award 1950697 to P.P., and the National Institutes of Health R35 grant GM122579 to P.P. and R.D.

References

1. Atkins JF, Gesteland R, Cech T (2011) RNA worlds: from life's origins to diversity in gene regulation. Cold Spring Harbor Laboratory Press, New York
2. Cech Thomas R, Steitz, Joan A (2014) The noncoding RNA revolution – trashing old rules to forge new ones. Cell 157:77–94
3. Blundell TL, Chaplin AK (2021) The resolution revolution in X-ray diffraction, Cryo-EM and other technologies. Prog Biophys Mol Biol 160:2–4
4. Watson ZL, Ward FR, Meheust R et al (2020) Structure of the bacterial ribosome at 2 A resolution. elife 9:e60482
5. Su Z, Zhang K, Kappel K et al (2021) Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 A resolution. Nature 96(7873):603–607
6. Schlick T, Pyle AM (2017) Opportunities and challenges in RNA structural modeling and design. Biophys J 113:225–234
7. Sripakdeevong P, Beauchamp K, Das R (2012) Why can't we predict RNA structure at atomic resolution? In: Leontis N, Westhof E (eds) RNA 3D structure analysis and prediction. Springer, Berlin/Heidelberg
8. Zhang J, Ferre-D'amare AR (2014) New molecular engineering approaches for crystallographic studies of large RNAs. Curr Opin Struct Biol 26:9–15
9. Zhang H, Keane SC (2019) Advances that facilitate the study of large RNA structure and dynamics by nuclear magnetic resonance spectroscopy. Wiley Interdiscip Rev RNA 10:e1541
10. Hura GL, Menon AL, Hammel M et al (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods 6:606–612
11. Kappel K, Zhang K, Su Z et al (2020) Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. Nat Methods 17:699–707
12. Jossinet F, Ludwig TE, Westhof E (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. Bioinformatics 26:2057–2059

13. Parisien M, Major F (2008) The MC-fold and MC-Sym pipeline infers RNA structure from sequence data. Nature 452:51–55
14. Xia Z, Bell DR, Shi Y et al (2013) RNA 3D structure prediction by using a coarse-grained model and experimental data. J Phys Chem B 117:3135–3144
15. Liebschner D, Afonine PV, Baker ML et al (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol 75:861–877
16. Das R, Baker D (2007) Automated *de novo* prediction of native-like RNA tertiary structures. Proc Natl Acad Sci USA 104:14664–14669
17. Xu X, Zhao C, Chen SJ (2019) VfoldLA: a web server for loop assembly-based prediction of putative 3D RNA structures. J Struct Biol 207:235–240
18. Popenda M, Szachniuk M, Antczak M et al (2012) Automated 3D structure composition for large RNAs. Nucleic Acids Res 40:e112
19. Jonikas MA, Radmer RJ, Laederach A et al (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. RNA 15:189–199
20. Boniecki MJ, Lach G, Dawson WK et al (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. Nucleic Acids Res 44:e63
21. Sharma S, Ding F, Dokholyan NV (2008) iFoldRNA: three-dimensional RNA structure prediction and folding. Bioinformatics 24:1951–1952
22. Miao Z, Adamiak RW, Antczak M et al (2020) RNA-puzzles round IV: 3D structure predictions of four ribozymes and two aptamers. RNA 26:982–995
23. Li B, Cao Y, Westhof E et al (2020) Advances in RNA 3D structure modeling using experimental data. Front Genet 11:574485
24. Townshend RJL, Eismann S, Watkins AM et al (2021) Geometric deep learning of RNA structure. Science 373:1047–1051
25. Kappel K, Liu S, Larsen KP et al (2018) De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. Nat Methods 15:947–954
26. Merino EJ, Wilkinson KA, Coughlan JL et al (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc 127:4223–4231
27. Peattie DA, Gilbert W (1980) Chemical probes for higher-order structure in RNA. Proc Natl Acad Sci USA 77:4679–4682
28. Cordero P, Kladwang W, Vanlang CC et al (2014) The mutate-and-map protocol for inferring base pairs in structured RNA. Methods Mol Biol 1086:53–77
29. Cheng CY, Kladwang W, Yesselman JD et al (2017) RNA structure inference through chemical mapping after accidental or intentional mutations. Proc Natl Acad Sci USA 114:9876–9881