nature chemistry

SUPPLEMENTARY INFORMATION

DOI: 10.1038/NCHEM.1176

# A two-dimensional mutate-and-map strategy for non-coding RNA structure

Wipapat Kladwang[1], Christopher C. VanLang[2], Pablo Cordero[3], and Rhiju Das[1,3,4]*

*Departments of Biochemistry[1] & Chemical Engineering[2], Program in Biomedical Informatics[3], and Department of Physics[4], Stanford University, Stanford CA 94305*

** To whom correspondence should be addressed: rhiju@stanford.edu. Phone: (650) 723-5976. Fax: (650) 723-6783.*

## Index





© 2011 Macmillan Publishers Limited. All rights reserved.

## Supporting Methods

*Preparation of model RNAs*

The DNA templates for each RNA (Table S1) consisted of the 20-nucleotide T7 RNA polymerase promoter sequence (TTCTAATACGACTCACTATA) followed by the desired sequence. Double-stranded templates were prepared by PCR assembly of DNA oligomers up to 60 nucleotides in length (IDT, Integrated DNA Technologies, IA) with Phusion DNA polymerase (Finnzymes, MA). For each mutant, an automated MATLAB script was used to determine which primers required single-nucleotide changes and to generate 96-well plate spreadsheets for ordering and guiding pipetting for PCR assembly reactions.

Assembled DNA templates were purified in 96-well Greiner microplates with AMPure magnetic beads (Agencourt, Beckman Coulter, CA) following manufacturer's instructions. Sample concentrations were measured based on UV absorbance at 260 nm measured on Nanodrop 100 or 8000 spectrophotometers. Verification of template length was accomplished by electrophoresis of all samples and 10-bp and 20-bp ladder length standards (Fermentas, MD) in 4% agarose gels (containing 0.5 mg/mL ethidium bromide) and 1x TBE (100 mM Tris, 83 mM boric acid, 1 mM disodium EDTA).

In vitro RNA transcription reactions were carried out in 40 μL volumes with 10 pmols of DNA template; 20 units T7 RNA polymerase (New England Biolabs, MA); 40 mM Tris-HCl (pH 8.1); 25 mM $MgCl_2$; 2 mM spermidine; 1 mM each ATP, CTP, GTP, and UTP; 4% polyethylene glycol 1200; and 0.01% Triton-X-100. Reactions were incubated at 37 °C for 4 hours and monitored by electrophoresis of all samples along with 100–1000 nucleotide RNA length standards (RiboRuler, Fermentas, MD) in 4% denaturing agarose gels (1.1% formaldehyde; run in 1x TAE, 40 mM Tris, 20 mM acetic acid, 1 mM disodium EDTA), stained with SYBR Green II RNA gel stain (Invitrogen, CA) following manufacturer instructions. RNA samples were purified with MagMax magnetic beads (Ambion, TX), following manufacturer's instructions; and concentrations



© 2011 Macmillan Publishers Limited. All rights reserved.

were measured by absorbance at 260 nm on Nanodrop 100 or 8000 spectrophotometers.

*SHAPE measurements*

Chemical modification reactions consisted of 1.2 pmols RNA in 20 μL with 50 mM Na-HEPES, pH 8.0, and 10 mM $MgCl_2$ and/or ligand at the desired concentration (see Table S1); and 5 μL of SHAPE modification reagent. The modification reagent was 24 mg/ml N-methyl isatoic anhydride freshly dissolved in anhydrous DMSO. The reactions were incubated at 24 °C for 15 to 60 minutes, with lower modification times for the longer RNAs to maintain overall modification rates less than 30%. In control reactions (for background measurements), 5 μL of deionized water was added instead of modification reagent, and incubated for the same time. Reactions were quenched with a premixed solution of 5 μL 0.5 M Na-MES, pH 6.0; 3 μL of 5 M NaCl, 1.5 μL of oligo-dT beads (poly(A) purist, Ambion, TX), and 0.25 μL of 0.5 mM 5´-rhodamine-green labeled primer (AAAAAAAAAAAAAAAAAAAGTTGTTGTTGTTTCTTT) complementary to the 3´ end of the MedLoop RNA [also used in our previous studies (*1, 2*)], and 0.05 μL of a 0.5 mM Alexa-555-labeled oligonucleotide (used to verify normalization). The reactions were purified by magnetic separation, rinsed with 40 μL of 70% ethanol twice, and allowed to air-dry for 10 minutes while remaining on a 96-post magnetic stand. The magnetic-bead mixtures were resuspended in 2.5 μL of deionized water.

The resulting mixtures of modified RNAs and primers bound to magnetic beads were reverse transcribed by the addition of a pre-mixed solution containing 0.2 μL of SuperScript III (Invitrogen, CA), 1.0 μL of 5x SuperScript First Strand buffer (Invitrogen, CA), 0.4 μL of 10 mM each dnTPs [dATP, dCTP, dTTP, and dITP (*3*)], 0.25 μL of 0.1 M DTT, and 0.65 μL water. The reactions (5 μL total) were incubated at 42 °C for 30 minutes. RNA was degraded by the addition of



© 2011 Macmillan Publishers Limited. All rights reserved.

5 μL of 0.4 M NaOH and incubation at 90 °C for 3 minutes. The solutions were neutralized by the addition of 5 μL of an acid quench (2 volumes 5 M NaCl, 2 volumes 2 M HCl, and 3 volumes of 3 M Na-acetate). Fluorescent DNA products were purified by magnetic bead separation, rinsed with 40 μL of 70% ethanol, and air dried for 5 minutes. The reverse transcription products, along with magnetic beads, were resuspended in 10 μL of a solution containing 0.125 mM Na-EDTA (pH 8.0) and a Texas-Red-labeled reference ladder (whose fluorescence is spectrally separated from the rhodamine-green-labeled products). The products were separated by capillary electrophoresis on an ABI 3100 or ABI 3700 DNA sequencer. Reference ladders for wild type RNAs were created using an analogous protocol without chemical modification and the addition of, e.g., 2´-3´-dideoxy-TTP in an amount equimolar to dTTP in the reverse transcriptase reaction.

*Data processing*

The HiTRACE software(*4*) was used to analyze the electropherograms. Briefly, traces were aligned by automatically shifting and scaling the time coordinate, based on cross correlation of the Texas Red reference ladder co-loaded with all samples. Sequence assignments to bands, verified by comparison to sequencing ladders, permitted the automated peak-fitting of the traces to Gaussians. Data were normalized so that, within each mutant, the mean band intensity was unity for all nucleotides except the 20 nucleotides closest to the 5´ and 3´ ends. Individual replicate data sets, including aligned electropherograms and quantified band intensities, are being made publically available in the Stanford RNA Mapping Database (http://rmdb.stanford.edu).

For each data set, Z-scores were calculated as follows. Let the observed band intensities be $s_{ij}$ with $i = 1, 2, \ldots N$ indexing the band numbers, and $j = 1, 2, \ldots M$ indexing the nucleotides that were mutated. Then, the mean band intensities $\mu_i$ and standard deviations $\sigma_i$ were computed using their standard definitions:



© 2011 Macmillan Publishers Limited. All rights reserved.

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} s_{ij}$$
$$\sigma_i = \left[\frac{1}{N}\sum_{j=1}^{N}(s_{ij}-\mu_i)^2\right]^{1/2} \quad (1)$$

And the Z-scores were computed as:

$$Z_{ij} = [s_{ij} - \mu_i]/\sigma_i \quad (2)$$

Only data with $Z_{ij} > 0.0$ and associated with bands with mean intensity $\mu_i$ less than a cutoff $\mu_i^{MAX} = 0.8$ were kept, since the mutate-and-map approach seeks to identify site-specific release of nucleotides that are protected in the starting sequence and most variants. [Varying $\mu_i^{MAX}$ from 0.5 to 1.0 gave indistinguishable results for models.] To avoid introducing additional noise, we did not correct for attenuation of longer reverse transcription products; because this effect should be similar for all mutants (and was observed to be such in the data), it scales $s_{ij}$, $\mu_i$, and $\sigma_{ij}$ identically (at a given nucleotide *i*) and did not affect the final $Z_{ij}$ scores in (2). Further, to again avoid introducing unnecessary noise, we did not explicitly subtract background measurements, as they should also subtract out of the Z-score expression (2). Nevertheless, control measurements for all RNAs without SHAPE modification were carried out. For a small number (<0.1%) of nucleotides in specific mutants, weak mutant-specific background bands were observed (likely due to sequence-specific reverse transcriptase stops). An analogous Z-score was carried out for these control background measurements; nucleotides with "background Z-scores" greater than 6.0 were identified as anomalous and set to zero in analyzing $Z_{ij}$ for mutate/map measurements. For data sets with more than one independent replicate (the *add* adenine-sensing riboswitch, the P4-P6 domain, and the *F. nucleatum* glycine-



© 2011 Macmillan Publishers Limited. All rights reserved.

sensing riboswitch), $Z_{ij}$ values were averaged across the replicates. The analysis is available as a single MATLAB script *output_Zscore_from_rdat.m* within the HiTRACE package.

*Inference of contacts through sequence-independent clustering*

The Z-scores $Z_{ij}$ [see above, eq. (2)] define possible long-range contacts in each RNA. As in prior work (*1, 2*), mutate/map pairs with statistically strong signals ($Z_{ij} > Z_{min}$; $Z_{min}$= 1.0) were considered. The pair (*i*,*j*) was defined as neighboring any strong signals at (*i*–1, *j*), (*i*+1, *j*), (*i*, *j*–1), (*i*, *j*+1), or the symmetry partner (*j*,*i*); strong signals were then grouped by single-linkage clustering. Final selection of clusters used simple but stringent filters. Clusters with at least 8 pairs, involving at least three independent mutations, and including at least one pair of symmetry partners were taken as defining long-range interactions with strong support. For this selection, mutations involved in more than 5 such clusters were omitted as potentially being associated with cooperative, extended structural effects beyond the disruption of a single base pair, helix, or tertiary contact. The analysis is available as a single MATLAB script *cluster_z_scores.m* within the HiTRACE package.

*Three-dimensional modeling with Rosetta: command-lines and example files*

Three-dimensional models were acquired using the Fragment Assembly of RNA with Full Atom Refinement (FARFAR) methodology(*5*) in the Rosetta framework.

First, ideal A-form helices were created with the command line:

```
rna_helix.exe -database <path to database> -nstruct 1 -fasta
stem2_add.fasta -out:file:silent stem2_add.out
```

where the file `stem2_add.fasta` contains the sequence of the P2 helix, as determined by the mutate-and-map data:

```
>stem2_add.fasta
uccuaauuggga
```



© 2011 Macmillan Publishers Limited. All rights reserved.

Then, for each RNA loop or junction motif that interconnects these helices, 4,000 models were created with FARFAR. For example, in the adenine riboswitch, two loops (L2 & L3) and the adenine-binding junction are the non-helical motif portions. The command line for building L2 onto the P2 helix is:

```
rna_denovo.<exe> -database  <path to database> -native
motif2_1y26_RNA.pdb -fasta motif2_add.fasta -params_file
motif2_add.params -nstruct 100 -out:file:silent motif2_add.out -cycles
5000 -mute all -close_loops -close_loops_after_each_move -minimize_rna
-close_loops -in:file:silent_struct_type rna -in:file:silent
stem2_add.out -chunk_res  1-6 16-21
```

Here, the optional "-native" flag, inputting the crystallographic structure for the motif, permits rmsd calculations but is not used in building the model. The file `motif2_add.params` defines the P2 stem within this motif-building run:

```
STEM    PAIR 6 16 W W A PAIR 5 17 W W A PAIR 4 18 W W A PAIR 3 19 W W A
PAIR 2 20 W W A PAIR 1 21 W W A
OBLIGATE PAIR 1 21 W W A
```

Finally, the models of separately built motifs and helices are assembled through the FARNA Monte Carlo procedure:

```
rna_denovo.<exe> -database  <path to database> -native 1y26_RNA.pdb -
fasta add.fasta -in:file:silent_struct_type binary_rna  -cycles 10000 -
nstruct 200 -out:file:silent add_assemble.out -params_file
add_assemble.params -cst_file
add_mutate_map_threetertiarycontacts_assemble.cst -close_loops  -
in:file:silent  stem1_add.out stem2_add.out stem3_add.out
motif1_add.out motif2_add.out motif3_add.out -chunk_res   1-9 63-71 13-
18 28-33 42-47 55-60 1-18 28-47 55-71 13-33 42-60
```

In the above command-line, the helix and loop definitions are given by `add_assemble.params`:

```
CUTPOINT_CLOSED  9 18 47
STEM  PAIR 1 71 W W A  PAIR 2 70 W W A  PAIR 3 69 W W A  PAIR 4 68 W W
A  PAIR 5 67 W W A  PAIR 6 66 W W A  PAIR 7 65 W W A  PAIR 8 64 W W A
PAIR 9 63 W W A
OBLIGATE PAIR 9 63 W W A

STEM  PAIR 13 33 W W A  PAIR 14 32 W W A  PAIR 15 31 W W A  PAIR 16 30
W W A  PAIR 17 29 W W A  PAIR 18 28 W W A
OBLIGATE PAIR 18 28 W W A
```



© 2011 Macmillan Publishers Limited. All rights reserved.

```
STEM  PAIR 42 60 W W A  PAIR 43 59 W W A  PAIR 44 58 W W A  PAIR 45 57
W W A  PAIR 46 56 W W A  PAIR 47 55 W W A
OBLIGATE PAIR 47 55 W W A
```

The constraint file `add_mutate_map_threetertiarycontacts_assemble.cst` encodes regions in tertiary contact (here including the short two-base-pair helix) inferred from the mutate-and-map data:

```
[ atompairs ]
N3 10 N3 39 FADE -100 10 2 -20.0
N3 10 N1 40 FADE -100 10 2 -20.0
N1 11 N3 39 FADE -100 10 2 -20.0
N1 11 N1 40 FADE -100 10 2 -20.0
N1 12 N1 40 FADE -100 10 2 -20.0
N3 10 N3 39 FADE -100 10 2 -20.0
N1 25 N3 49 FADE -100 10 2 -20.0
N1 25 N3 50 FADE -100 10 2 -20.0
N1 26 N3 48 FADE -100 10 2 -20.0
N1 26 N3 49 FADE -100 10 2 -20.0
N1 26 N3 50 FADE -100 10 2 -20.0
N3 27 N3 49 FADE -100 10 2 -20.0
N3 27 N3 50 FADE -100 10 2 -20.0
N3 35 N1 40 FADE -100 10 2 -40.0
N1 34 N3 41 FADE -100 10 2 -40.0
```

These constraints give a bonus of –20.0 kcal/mol if the specified atom pairs are within 8 Å; the function interpolates up to zero for distances by a cubic spline beyond 10.0 Å. Note that the Rosetta numbering here starts with 1 for the first nucleotide of the 71-nucleotide adenine binding domain, and is offset by 12 from the numbering in the crystal structure 1Y26. 5000 models of the full-length RNA were generated, and the lowest-energy conformation was taken as the final model. (For the adenine riboswitch, the next nine lowest energy models were within 2 Å RMSD of this model, indicating convergence.) Example files for carrying out the calculation are being distributed with Rosetta release 3.4 in rosetta_demos/RNA_Assembly/, including setup script *setup_rna_assembly_jobs.py*.



© 2011 Macmillan Publishers Limited. All rights reserved.

## Table S1. Benchmark for the mutate-and-map strategy for noncoding RNA base pair inference.

| RNA, source | Solution conditions[a] | Off–set[b] | PDB[c] | Sequence & Secondary Structure[d] |
|---|---|---|---|---|
| Adenine riboswitch, *V. vulnificus* (*add*) | Standard + 5 mM adenine | –8 | **1Y26**<br>1Y27<br>2G9C<br>3GO2<br>… | ggaaaggaaagggaaagaaaCGCUUCAUAUAAUCCUAAUGAUAUGGUUUGGG<br>AGUUUCUACCAAGAGCCUUAAACUCUUGAUUAUGAAGUGaaaacaaaacaaa<br>gaaacaacaacaacaac<br>.....(((((((((...((((((.........)))))<br>)........((((((.......))))))..))))))))).....<br>..... |
| tRNA$^{phe}$, *E. coli* | Standard | –15 | **1L0U**<br>1EHZ | ggaacaaacaaaacaGCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAG<br>AUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCAaaaccaaagaaac<br>aacaacaacaac<br>.............(((((((..((((........)))).((((.....<br>..))))).....(((((.......)))))))))).....<br>..... |
| P4–P6 domain, *Tetrahymena* ribozyme | Standard, 30% methyl-pentanediol[e] | 89 | **1GID**<br>1L8V<br>1HR2<br>2R8S | ggccaaaacaacgGAAUUGCGGGAAAGGGGUCAACAGCCGUUCAGUACCAAG<br>UCUCAGGGGAAACUUUGAGAUGGCCUUGCAAAGGGUAUGGUAAUAAGCUGAC<br>GGACAUGGUCCUAACCACGCAGCCAAGUCCUAAGUCAACAGAUCUUCUGUUG<br>AUAUGGAUGCAGUUCAaaaccaaaccaaagaaacaacaacaacaac<br>.............(((((...((((((.....(((.((((.(((..(<br>(((((((....)))))))))..((.......)).....))).....)))))<br>))....))))))..))).)))))(((...(((...(((((((((...)))))))<br>))..))))...)).....<br>..... |
| 5S rRNA, *E. coli* | Standard | –20 | **3OFC**<br>3OAS<br>3ORB<br>2WWQ<br>… | ggaaaggaaagggaaagaaaUGCCUGGCGGCCGUAGCGCGGUGGUCCCACCU<br>GACCCCAUGCCGAACUCAGAAGUGAAACGCCGUAGCGCCGAUGGUAGUGUGG<br>GGUCUCCCCAUGCGAGAGUAGGGAACUGCCAGGCAUaaaacaaaacaaagaa<br>acaacaacaacaac<br>.....(((((((((.....(((((((((....(((((<br>((.............))))..)))...)))))).))..((.......(((((<br>((...)))))))).......))...)))))))).....<br>..... |
| Cyclic di-GMP riboswitch, *V. cholerae* (*VC1722*) | Standard + 10 μM cyclic diguanosine monophosphate | 0 | **3MXH**<br>3IWN<br>3MUV<br>3MUT<br>… | ggaaaaauGUCACGCACAGGGCAAACCAUUCGAAAGAGUGGGACGCAAAGCC<br>UCCGGCCUAAACCAGAAGACAUGGUAGGUAGCGGGGUUACCGAUGGCAAAAU<br>Gcauacaaaccaaagaaacaacaacaacaac<br>.........((((......((...(((((((....)))))))...)).((<br>.(((((((((..((.........)))))))..))))))...)).))<br>..... |
| Glycine riboswitch, *F. nucleatum* | Standard + 10 mM glycine | –10 | **3P49** | ggacagagagGAUAUGAGGAGAGAUUUCAUUUUAAUGAAACACCGAAGAAGU<br>AAAUCUUUCAGGUAAAAAGGACUCAUAUUGGACGAACCUCUGGAGAGCUUAU<br>CUAAGAGAUAACACCGAAGGAGCAAAGCUAAUUUUAGCCUAAACUCUCAGGU<br>AAAAGGACGGAGaaaacacaacaaagaaacaacaacaacaac<br>.........(((((((.....((((((....)))))).(((...((((.<br>....)))).)..)))........)))))))..........((((.....((((<br>(.....)))))..(((...((((....((((....)))).....)))))..)))<br>.....)))).....<br>..... |

[a] Standard conditions are: 10 mM $MgCl_2$, 50 mM Na–HEPES, pH 8.0 at 24 °C.

[b] Number added to sequence index to yield numbering used in previous biophysical studies, and in Figs. 1-5 of the main text.

[c] The first listed PDB ID was the source of the assumed crystallographic secondary structure; other listed IDs contain sequence variants, different complexes, or different crystallographic space groups and confirm this structure.

[d] In the sequence, lowercase symbols denote 5´and 3´ buffer sequences, including primer binding site (last 20 nucleotides). In all cases, designs were checked in RNAstructure and ViennaRNA to give negligible base pairing between added sequences and target domain. Structure is given in dot–bracket notation, and here denotes base pairs for which there is crystallographic evidence. A long-range two-base-pair helix [25–50, 26–49] in the adenine riboswitch is involved in an extensive non-canonical loop-loop interaction and is not included.

[e] 2-methyl-2,4,-pentanediol (MPD) was included due to reports that its presence in crystallization buffer can change SHAPE reactivity of the P4-P6 RNA (Vicens et al. (2007) RNA 13, 536–48). Mutate/map measurements without MPD gave different reactivities in the P5abc region and ambiguous modeling results in the P5c region; the crystallographic helix or a helix with single-nucleotide register shift was observed in models from different bootstrap replicates.



© 2011 Macmillan Publishers Limited. All rights reserved.

**Table S2. Base-pair-resolution analysis of the helices recovered by the mutate-and-map method.**

| RNA | Crystallographic | Correctly recovered | Missed[a] | | | Extra base pairs[a] | | |
|---|---|---|---|---|---|---|---|---|
| | | | A-U | G-U | G-C | A-U | G-U | G-C |
| Adenine riboswitch[b] | 21 | 21 | 0 | 0 | 0 | 0 | 0 | 0 |
| $tRNA^{phe}$ | 20 | 20 | 0 | 0 | 0 | 1 | 0 | 0 |
| P4-P6 RNA | 48 | 47 | 1 | 0 | 0 | 4 | 1 | 0 |
| 5S rRNA | 34 | 34 | 0 | 0 | 0 | 0 | 0 | 0 |
| c-di-GMP riboswitch[b] | 25 | 23 | 0 | 0 | 0 | 2 | 0 | 0 |
| Glycine riboswitch[b] | 40 | 40 | 0 | 0 | 0 | 1 | 2 | 0 |
| *Total* | 188 | 185 | 1 | 0 | 0 | 8 | 3 | 0 |

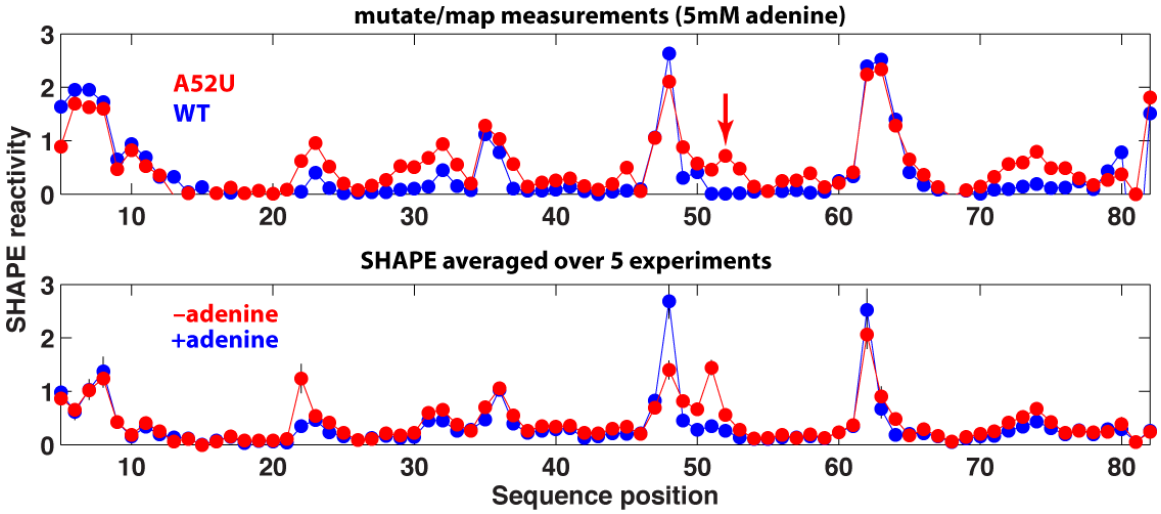
[a] Number of missed or extra base-pairs within helices that match crystallographic helices. (The only crystallographic helix not recovered in the mutate-and-map models is a short stem with two G-C base pairs in the cyclic di-GMP riboswitch.)

[b] Ligand-binding riboswitches were probed in the presence of small-molecule partners (5 mM adenine, 10 μM cyclic di-guanosine-monophosphate, or 10 mM glycine). All experiments were carried out with 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0.



© 2011 Macmillan Publishers Limited. All rights reserved.

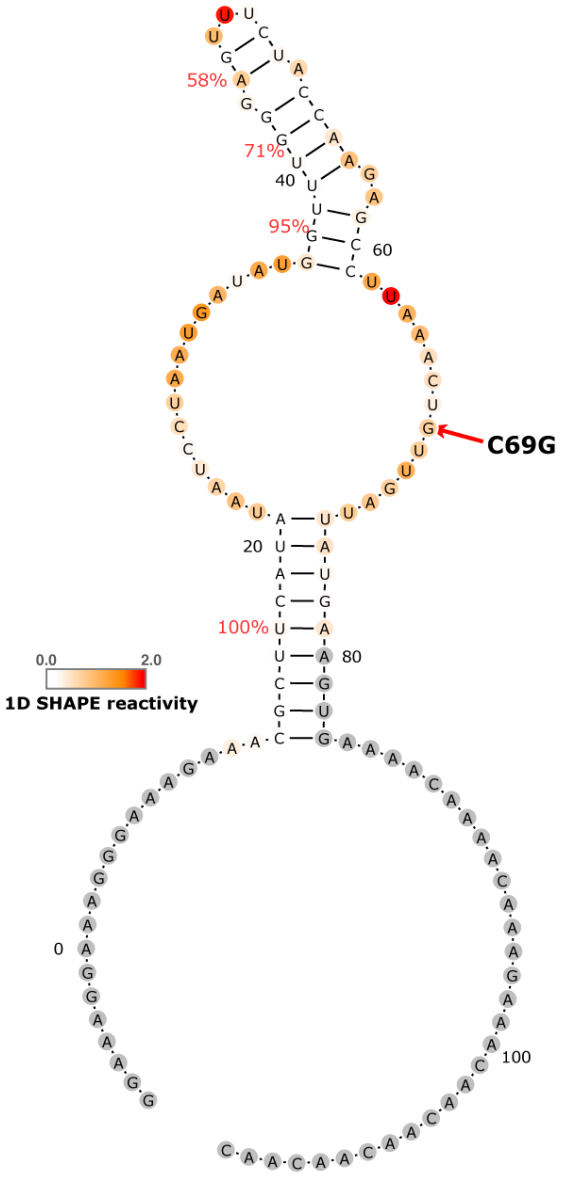
**Figure S1. Example of large-scale change in RNA structure induced by single mutation: loss of ligand binding.** In the mutate-and-map data set of main text Fig. 1, the mutation A52U leads to perturbations throughout the adenine riboswitch ligand-binding domain. (A) Quantified, background-subtracted areas for A52U compared to the wild type sequence ('WT') in 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, and 5 mM adenine. Site of mutation is marked with a red arrow. (B) Perturbations from the A52U mutation are similar to differences between the adenine-free and adenine-bound state of the riboswitch. Data shown are 'gold-standard' measurements (*6*) averaged over five independent replicates for the wild type riboswitch without (red) and with (blue) 5 mM adenine.






© 2011 Macmillan Publishers Limited. All rights reserved.

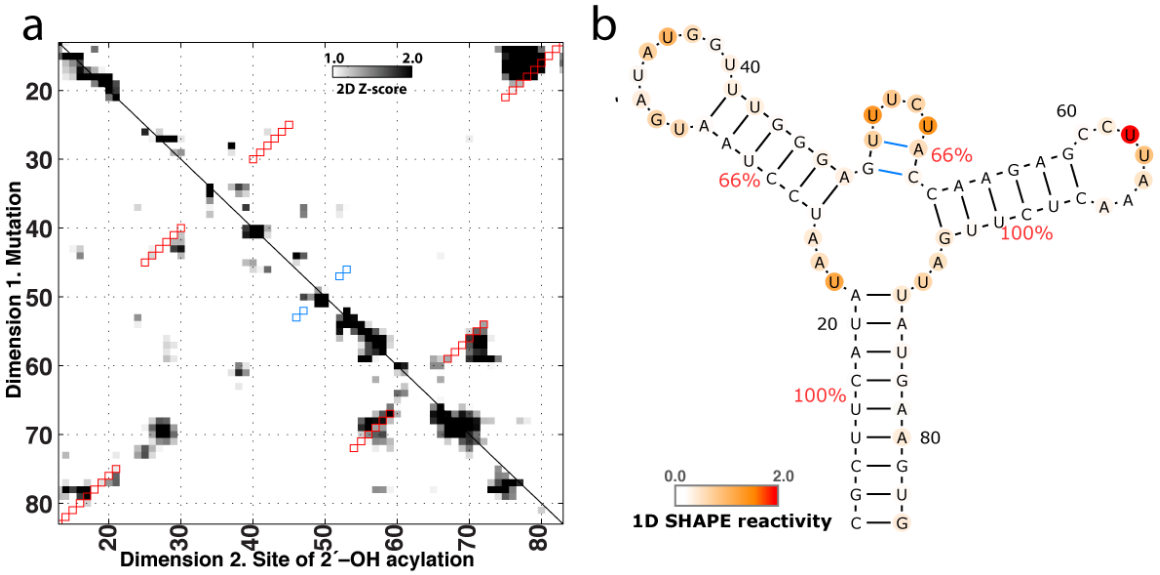
**Figure S2. Example of large-scale change in RNA structure induced by single mutation: change in secondary structure.** Background-subtracted SHAPE reactivities for the adenine riboswitch ligand-binding domain with C69G mutation are shown as coloring from white to red. The mutation site 69 normally lies in the P3 helix of the wild type RNA; the mutation leads to a large-scale change in SHAPE reactivity. The new reactivity is explained well by a change in secondary structure from the adenine-binding structure (cf. Fig. 1d in main text, and also SI Fig S3). The new model was estimated from *RNAstructure* guided by these 1D data (*7*); bootstrap values (*6*) given as percentages.






© 2011 Macmillan Publishers Limited. All rights reserved.

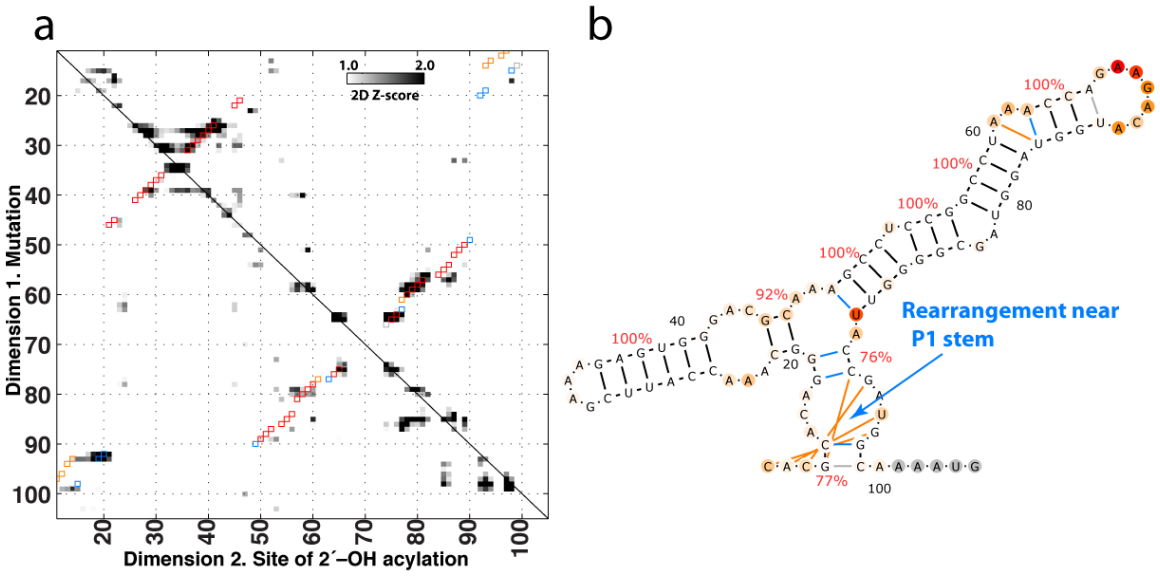
**Figure S3.** Mutate-and-map analysis of a partially ordered state of the adenine riboswitch. (a) Mutate-and-map data (Z-scores) are given in gray-scale for the adenine-binding domain from the *add* riboswitch, *V. vulnificus*, without adenine present. Red squares mark crystallographic secondary structure of the RNA in its adenine-bound form. (b) Dominant secondary structure for the ligand-free adenine riboswitch, inferred from mutate-and-map data, is not distinguishable from the ligand-bound form (see Main Text Fig. 1c-d). Cyan lines mark an 'extra' helix that is also seen in the ligand-bound state; the helix corresponds to neighboring Watson-Crick base pair and non-Watson-Crick base pair seen in the crystallographic ligand-bound model. Bootstrap confidence estimates for each helix are given in red.






© 2011 Macmillan Publishers Limited. All rights reserved.

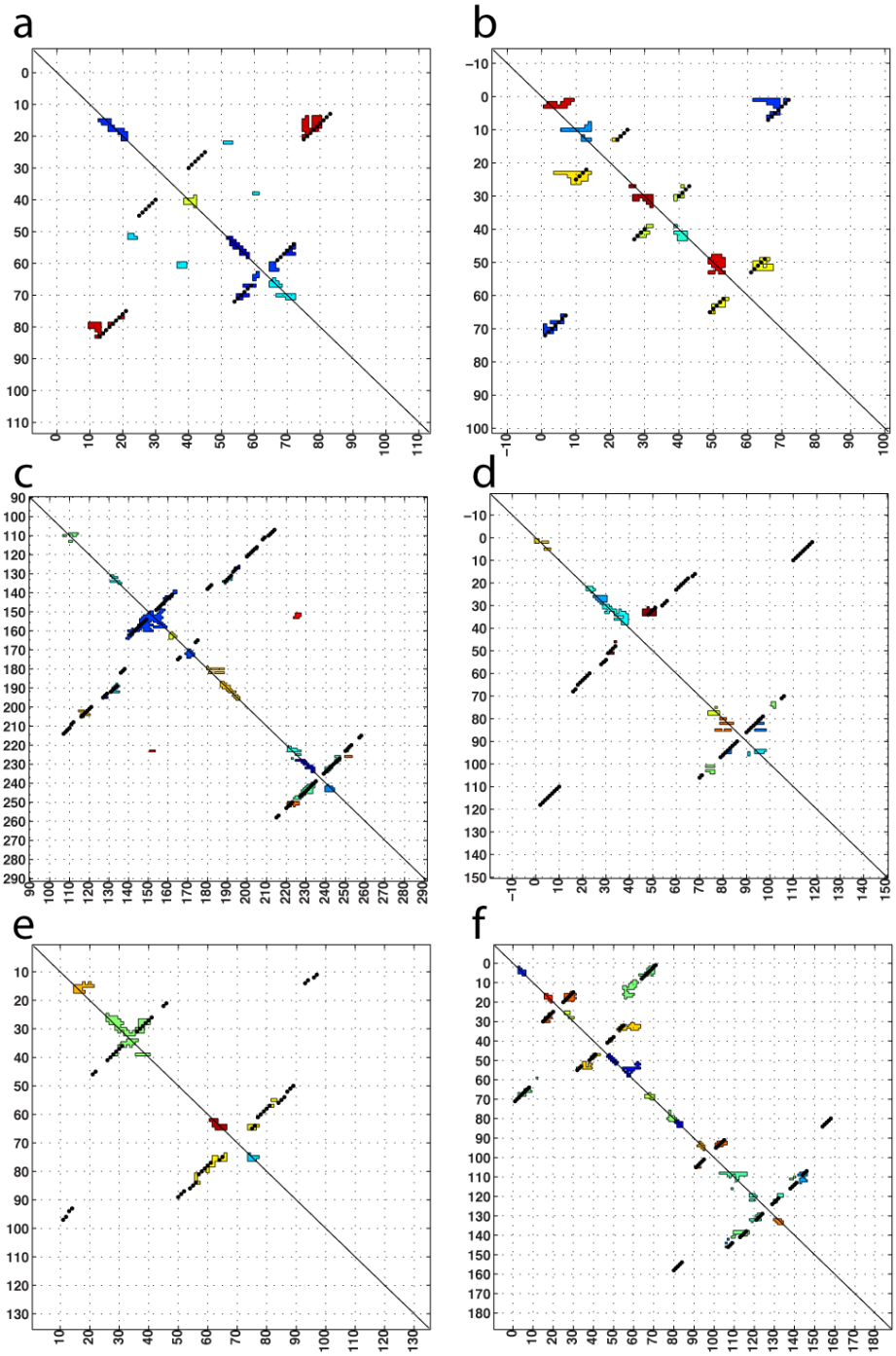
**Figure S4.** Mutate-and-map analysis of a partially ordered state of a cyclic di-guanosine monophosphate (c-di-GMP) riboswitch. (a) Mutate-and-map data (Z-scores) are given in gray-scale for the c-di-GMP-binding domain from the *VC1722* riboswitch, *V. cholerae*, without c-di-GMP present. Red squares mark crystallographic secondary structure of the RNA in its c-di-GMP-bound form. (b) Secondary structure models for this ligand-free c-di-GMP riboswitch, inferred from mutate-and-map data, is different from models for the ligand-bound state near the P1 stem and c-di-GMP binding region. Cyan squares (a) and lines (b) are mutate-and-map base pairs not present in (ligand-bound) crystal structure; orange annotations are crystallographic base pairs not present in the mutate-and-map model. Bootstrap confidence estimates for each helix are given in red.






© 2011 Macmillan Publishers Limited. All rights reserved.

**Figure S5. Accurate inference of contacting regions in structured non-coding RNAs through sequence-independent analysis of mutate-and-map data**. Cluster analysis of *Z*-scores, using filters for signal strength, number of independent mutants, and symmetry of features (see Methods); final clusters are shown in different, randomly chosen colors. Base pairs from crystallographic secondary structure are marked as black symbols. RNAs are (a) adenine riboswitch, (b) tRNA(phe), (c) P4-P6 RNA, (d) 5S rRNA, (e) c-di-GMP riboswitch, and (f) glycine riboswitch. Riboswitch data were collected with ligands present.






© 2011 Macmillan Publishers Limited. All rights reserved.

## References for Supporting Information


1. Kladwang, W., and Das, R. (2010) A mutate-and-map strategy for inferring base pairs in structured nucleic acids: proof of concept on a DNA/RNA helix, *Biochemistry 49*, 7414-7416.
2. Kladwang, W., Cordero, P., and Das, R. (2011) A mutate-and-map strategy accurately infers the base pairs of a 35-nucleotide model RNA, *RNA 17*, 522-534.
3. Mills, D. R., and Kramer, F. R. (1979) Structure-independent nucleotide sequence analysis, *Proc Natl Acad Sci U S A 76*, 2232-2235.
4. Yoon, S., Kim, J., Hum, J., Kim, H., Park, S., Kladwang, W., and Das, R. (2011) HiTRACE: high-throughput robust analysis for capillary electrophoresis, *Bioinformatics 27*, 1798-1805.
5. Das, R., Karanicolas, J., and Baker, D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure, *Nat Methods 7*, 291-294.
6. Kladwang, W., VanLang, C. C., Cordero, P., and Das, R. (2011) Understanding the errors of SHAPE-directed RNA modeling, *Biochemistry*, in press.
7. 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.




© 2011 Macmillan Publishers Limited. All rights reserved.