Systematic reconstruction of RNA functional motifs with high-throughput microfluidics

Lance Martin[1–3], Matthias Meier[2,3,6], Shawn M Lyons[4], Rene V Sit[2,3], William F Marzluff[4], Stephen R Quake[2,3,5] & Howard Y Chang[1,2]

We present RNA–mechanically induced trapping of molecular interactions (RNA-MITOMI), a microfluidic platform that allows integrated synthesis and functional assays for programmable RNA libraries. The interaction of a comprehensive library of RNA mutants with stem-loop–binding protein precisely defined the RNA structural and sequence features that govern affinity. The functional motif reconstructed in a single experiment on our platform uncovers new binding specificities and enriches interpretation of phylogenetic data.

The Encyclopedia or DNA Elements (ENCODE) project has shown that three-quarters of the human genome can be transcribed[1], whereas only ~1.5% of the genome encodes protein. Noncoding RNAs (ncRNAs), including the 5′ and 3′ untranslated regions of mRNAs, can exert epigenetic, transcriptional and post-transcriptional regulation of gene networks that affect human health[2]. Regulatory function is primarily carried out by interactions between ncRNAs and diverse ligands. Yet it is difficult to pinpoint the sequence or structural motifs in an ncRNA that govern interaction specificity. Moreover, molecular principles that govern the ncRNA interactome are poorly understood[3], and pathogenic mutations in ncRNAs that disrupt their regulatory function are difficult to identify[4].

We developed a general strategy aimed at identifying the contribution of RNA structural and sequence features to affinity for a given ligand (**Supplementary Note 1**). We applied this strategy to the well-characterized interaction between stem-loop–binding protein (SLBP) and the 3′ histone mRNA stem-loop[5] (**Supplementary Table 1**). After designing a library of stem-loop mutants, we adapted the MITOMI[6] microfluidic platform such that the entire RNA library could be simultaneously synthesized and then assayed for ligand-binding affinity (**Supplementary Note 2**).

The MITOMI microfluidic chip used in this study has 640 microchambers, each with a volume ~1 nl. We arrayed DNA oligos, which serve as transcription templates for the RNA library, and overlaid the MITOMI chip onto the array such that each spot was compartmentalized in a unique microchamber. Each microchamber has a back-chamber, which houses the spotted DNA template, and a detection chamber in which the interaction between RNA and ligand can be measured (**Fig. 1a**). After testing two molecular beacon designs (**Supplementary Fig. 1**), we chose to immobilize a 5′ biotin– and 3′ fluorescein (FAM)-labeled poly(T) ssDNA capture probe in each detection chamber (**Supplementary Fig. 2**). We then flowed an *in vitro* transcription mix onto the chip, transcribing RNA from each DNA spot. The transcribed poly(A)-tailed RNA molecules hybridized to the surface immobilized capture probe. After RNA synthesis, we used a quencher probe to quantify RNA capture in each detection chamber (**Supplementary Fig. 3**). Thus, our design allows co-transcriptional RNA folding and quantification of RNA capture without incorporation of fluorophore-modified nucleotides or use of intercalating dyes, which can alter RNA structure or function. After RNA synthesis, our design enables interrogation of the RNA library against a wide diversity of ligands.

Next we incubated N-terminal glutathione *S*-transferase (GST)-tagged SLBP with Texas Red dye conjugated to a GST antibody and flowed the protein across the immobilized RNA library. We quantified both RNA and protein captured in each chamber after equilibration using a microarray scanner (**Fig. 1b**). The wild-type stem-loop RNA (SLWT) and SLRS, an RNA mutant with no binding affinity to SLBP[7], had similar amounts of RNA capture, but SLWT retrieved ~100-fold more SLBP than did the SLRS negative control (**Fig. 1c**), demonstrating that the assay had favorable dynamic range and low background signal owing to nonspecific RNA-protein binding.

We measured relative SLBP binding across our library of single and double RNA mutants at 3 nM (**Fig. 2a**) and 30 nM (**Supplementary Fig. 4**) protein concentrations. On a scatter plot we displayed the binding of each point mutant with respect to the relative binding of each double mutant that rescued RNA structure (**Fig. 2b** and **Supplementary Fig. 5**). The points on the scatter plot fell into three regimes: 15 mutants had little effect on binding, 15 deleterious mutants were rescued by compensatory mutations that restore RNA structure, but three mutants could not be rescued by restoring structure, indicating sequence-based recognition at these positions (**Fig. 2c**).

We used the data set to generate a 'functional motif', which distills sequence and structural requirements for stem-loop function (SLBP binding) at single-nucleotide resolution (**Fig. 3a**). We validated

[1]Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California, USA. [2]Howard Hughes Medical Institute, Stanford University, Stanford, California, USA. [3]Department of Bioengineering, Stanford University, Stanford, California, USA. [4]Program in Molecular Biology and Biotechnology, University of North Carolina, Chapel Hill, North Carolina, USA. [5]Department of Applied Physics, Stanford University, Stanford, California, USA. [6]Present address: Department of Microsystems Engineering and Center for Biological Signaling Studies, University of Freiburg, Freiburg, Germany. Correspondence should be addressed to H.Y.C. (howchang@stanford.edu) or S.R.Q. (quake@stanford.edu).

RECEIVED 16 JUNE; ACCEPTED 25 SEPTEMBER; PUBLISHED ONLINE 11 NOVEMBER 2012; DOI:10.1038/NMETH.2225

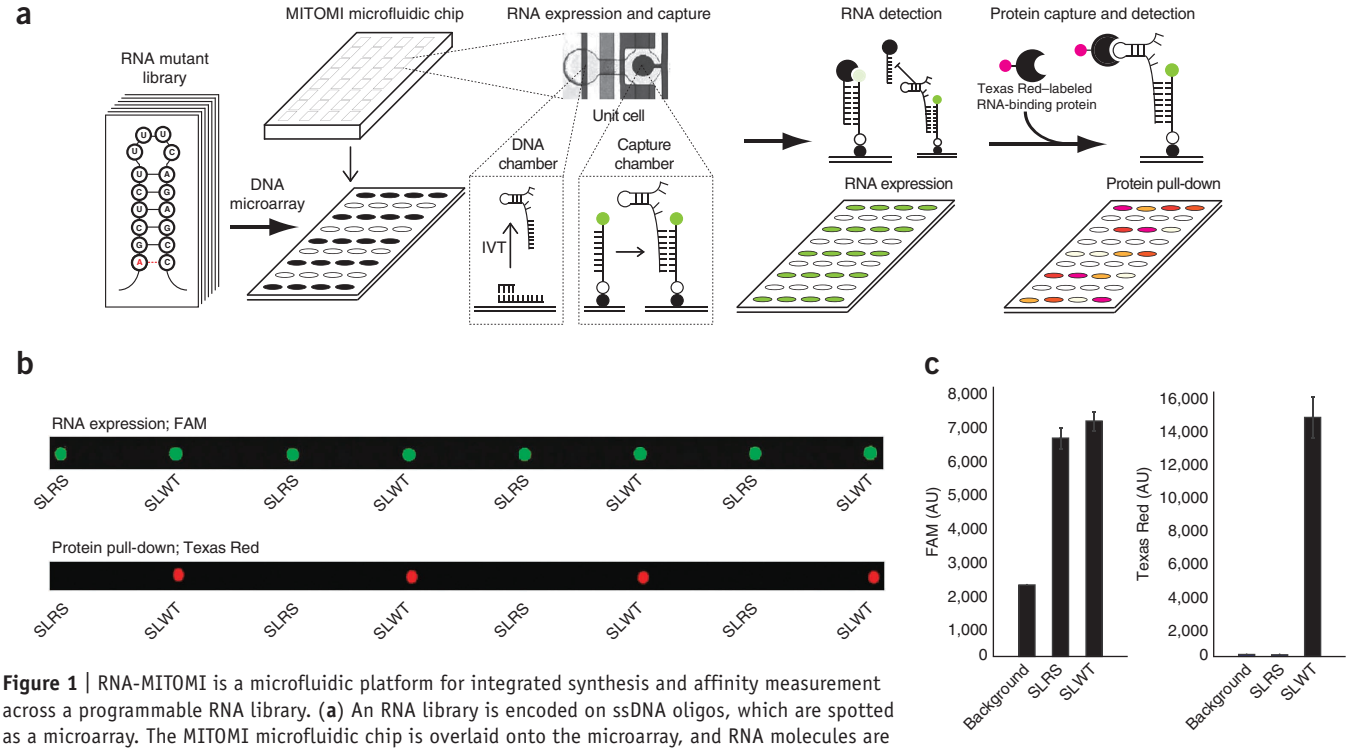

Figure 1 | RNA-MITOMI is a microfluidic platform for integrated synthesis and affinity measurement across a programmable RNA library. (**a**) An RNA library is encoded on ssDNA oligos, which are spotted as a microarray. The MITOMI microfluidic chip is overlaid onto the microarray, and RNA molecules are transcribed from spotted DNA templates in each chamber. IVT, *in vitro* transcription. RNA is captured and quantified with a fluorophore-tagged ligand before protein capture. (**b**) Imaging on orthogonal channels quantifies RNA and protein capture in each cell of the device. (**c**) The imaged spots shown in **b** are averaged for SLWT and SLRS. AU, arbitrary units. Error bars, s.d. (n = 4).

the functional motif in two ways. First, the features identified in the functional motif recapitulated the pattern of phylogenetic conservation of stem-loop sequences[8], suggesting that SLBP binding is the dominant selective constraint on the histone mRNA 3′ end (**Fig. 3b**). Residues (G2 and U9) and structural features (the U6-A11 base pair) that are critical for SLBP binding are conserved from *Tetrahymena* sp. to humans, whereas we observed covariation for base pairs that have less stringent sequence-specificity but need to maintain pairing (G1-C16, C3-G14, U4-A13 and C5-G12). Second, we tested nine point mutants with electrophoretic mobility shift analysis. The shift data agreed with the measurement of binding affinity obtained with RNA-MITOMI (**Supplementary Fig. 6**), confirming that measurements on the MITOMI platform can be recapitulated with conventional biochemical assays. Concordance of the functional motif with the evolutionary conservation of histone 3′ sequences also suggests

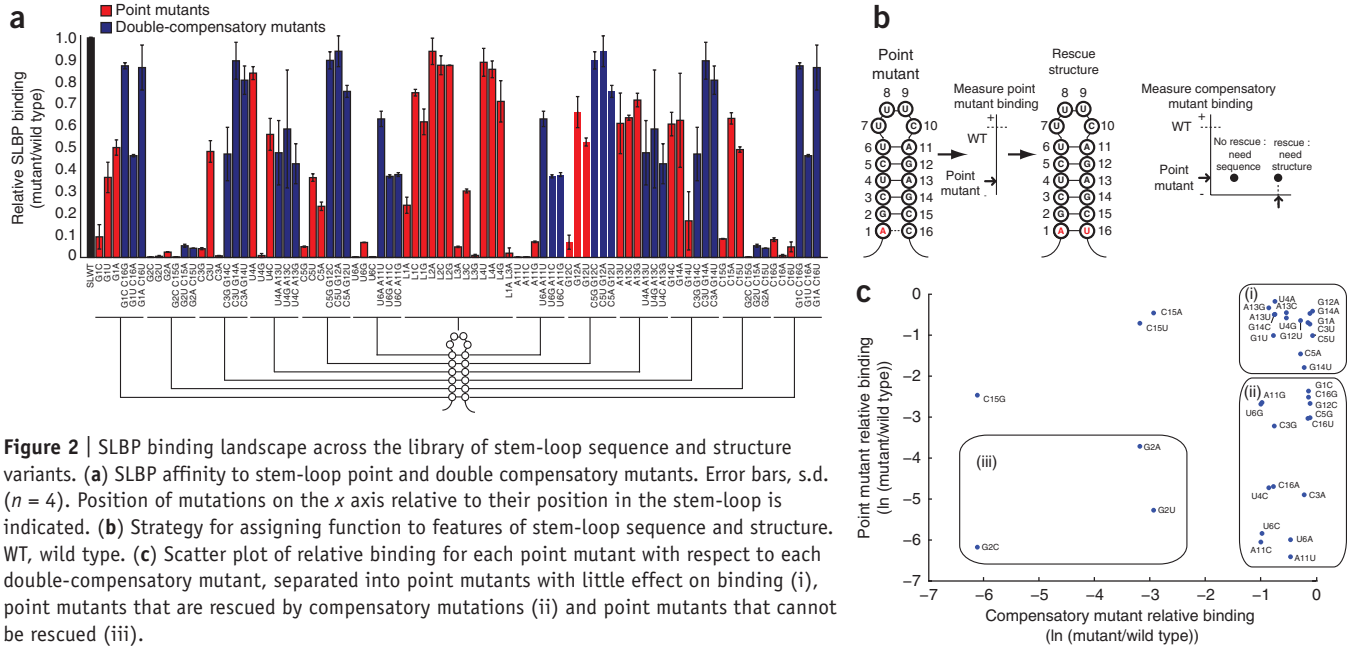

Figure 2 | SLBP binding landscape across the library of stem-loop sequence and structure variants. (**a**) SLBP affinity to stem-loop point and double compensatory mutants. Error bars, s.d. (n = 4). Position of mutations on the x axis relative to their position in the stem-loop is indicated. (**b**) Strategy for assigning function to features of stem-loop sequence and structure. WT, wild type. (**c**) Scatter plot of relative binding for each point mutant with respect to each double-compensatory mutant, separated into point mutants with little effect on binding (i), point mutants that are rescued by compensatory mutations (ii) and point mutants that cannot be rescued (iii).

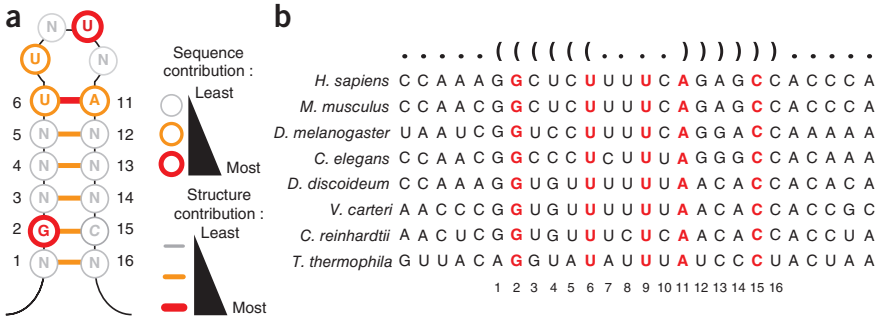

Figure 3 | Reconstruction and validation of the stem-loop functional motif. (**a**) For the functional motif, shown are stem-loop sequence and structural features that are required for function (SLBP binding). (**b**) Phylogenetic analysis showing sequence conservation at features that are required for SLBP binding as well as covariation without sequence preservation at base pairs for which structure, but not sequence, is important. Dot-parenthesis notation (top) indicates RNA secondary structure.

that MITOMI measurements can identify features that are functionally relevant *in vivo*.

Our results show that SLBP recognizes much of the stem via RNA secondary structure rather than sequence. For nine of the 12 bases in the stem, function was not dependent on sequence, as individual base-pair substitutions at these positions were tolerated. In some cases, base substitutions were functionally neutral only if structure was preserved. In particular, Watson-Crick base pairing is required at the U6-A11 position at the top of the stem (**Supplementary Fig. 7**). Nucleotide-specific contacts are also important for SLBP binding. Consistent with prior studies[7], two residues in the loop (U7 and especially U9) as well as G2 in the stem are required for binding (**Supplementary Fig. 8**).

Extending the results from prior studies[7,9], using our panel of mutants we also identified noncanonical base pairs that are tolerated. Though the G2–C15 base pair cannot be replaced by any of the three Watson-Crick base pairs, base substitutions in the position (C15) opposing G2 are tolerated, as mutations that allow wobble base-pairing (C15U) or prevent Watson-Crick base pairing (C15A) had nearly wild-type binding. Similar to the C15A mutation, C5A established a functionally tolerated G•A 'base pair'. In addition, G12A and G1A mutations created a functionally tolerated A•C pair and G14C created a tolerated C•C (**Supplementary Table 1**). Yet these noncanonical pairs can be deleterious if combined (**Supplementary Fig. 9**), suggesting limited tolerance to structure perturbation in the stem. Furthermore, some of these noncanonical pairs were not tolerated at other positions in the stem, such as G•A created by U4G and C16A as well as C•C created by G1C, which may indicate higher sensitivity to structural deformation at these positions.

The functional motif reconstructed by RNA-MITOMI analysis can enrich interpretation of phylogenetic data, as sequence covariation patterns can be viewed through a functional lens. Our results suggest that many stem-loop variants can bind SLBP with near wild-type affinity. Some of the base pairs not found in vertebrates have been found in other organisms, such as *Drosophila*[5]. The fact that many of the other variants have not been observed in nature suggests that they might lack other important functions of the stem-loop, such as formation of a SLBP–3′ hExo ternary complex[10,11] or other regulatory interactions that have a role in histone mRNA metabolism.

RNA-MITOMI analysis is complementary with sequencing-based methods for RNA interactomics[12–15], which can identify thousands of putative RNA-ligand interactions that need to be validated. Furthermore, motifs may be recombined and assayed to support engineering of ncRNAs or untranslated regions with new functions[2,16].

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank R.C. Spitale and C. Chu for helpful discussions, J.C. Liang, R.J. Bloom and C.D. Smolke for helpful discussions and assistance with Biacore, and A. Clore of Integrated DNA Technologies for assistance with probe designs. L.M. was funded by the Department of Defense National Defense Science & Engineering Graduate Fellowship. Our work was supported by US National Institutes of Health R01-HG004361 (to H.Y.C.) and R01GM29832 (to W.F.M.). H.Y.C. and S.R.Q. are funded by the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

L.M., S.R.Q. and H.Y.C. conceived the study. L.M., M.M., S.M.L. and R.V.S. performed experiments. L.M., M.M., S.M.L., W.F.M., S.R.Q. and H.Y.C. discussed the results and analyzed the data. L.M. and H.Y.C. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nmeth.2225.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Djebali, S. *et al. Nature* **489**, 101–108 (2012).
2. Martin, L. & Chang, H.Y. *J. Clin. Invest.* **122**, 1589–1595 (2012).
3. Guttman, M. & Rinn, J.L. *Nature* **482**, 339–346 (2012).
4. Cooper, G.M. & Shendure, J. *Nat. Rev. Genet.* **12**, 628–640 (2011).
5. Marzluff, W.F., Wagner, E.J. & Duronio, R.J. *Nat. Rev. Genet.* **9**, 843–854 (2008).
6. Maerkl, S.J. & Quake, S.R. *Science* **315**, 233–237 (2007).
7. Battle, D.J. & Doudna, J.A. *RNA* **7**, 123–132 (2001).
8. Davila Lopez, M. & Samuelsson, T. *RNA* **14**, 1–10 (2008).
9. Williams, A.S. & Marzluff, W.F. *Nucleic Acids Res.* **23**, 654–662 (1995).
10. Dominski, Z., Yang, X.C., Kaygun, H., Dadlez, M. & Marzluff, W.F. *Mol. Cell* **12**, 295–305 (2003).
11. Yang, X.C., Purdy, M., Marzluff, W.F. & Dominski, Z. *J. Biol. Chem.* **281**, 30447–30454 (2006).
12. Koenig, J., Zarnack, K., Luscombe, N.M. & Ule, J. *Nat. Rev. Genet.* **13**, 77–83 (2011).
13. Zhao, J. *et al. Mol. Cell* **40**, 939–953 (2010).
14. Granneman, S., Kudla, G., Petfalski, E. & Tollervey, D. *Proc. Natl. Acad. Sci. USA* **106**, 9613–9618 (2009).
15. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. *Mol. Cell* **44**, 667–678 (2011).
16. Liang, J.C., Bloom, R.J. & Smolke, C.D. *Mol. Cell* **43**, 915–926 (2011).

ONLINE METHODS

Microarray preparation. Each RNA mutant was encoded on a DNA oligo, which were ordered in antisense orientation relative to the mutant sequence in **Supplementary Table 1**. Equimolar (9 μM) amounts of antisense oligo and sense T7 promoter oligo were annealed for 1 min at 95 °C and allowed to cool at room temperature. After annealing, the oligos were diluted 1,000-fold and resuspended in a final concentration of 10 mg/ml BSA (Sigma, 100 mg/ml stock) for array spotting. The annealed oligos were arrayed in a 384-well plate and loaded into a custom microarry spotter with a 36-pin printhead (Parallel Synthesis 32-pin microarray print head) and three pins (Parallel Synthesis silicon pins with 75 mm × 75 μm tips). Oligos were arrayed from a 384-well plate to an epoxy slide (25 mm × 75 mm VEPO-25 Vantage expoxy slides) slide, resulting in an array orientation shown in **Supplementary Table 2**. Common T7 promoter (AAT TTAATACGACTCACTATAGG) and RNA-capture tail (CCCCA AAAAAAAAAAAAAAAAAAAAAAA) sequences were used for all oligos listed in **Supplementary Table 1**. We chose an unstructured poly(A) sequence to reduce the likelihood of intramolecular interactions between the stem-loop and tail, increasing reliability of RNA capture. We tested nine RNA species that lack the poly(A) tail using electrophoretic mobility shift analysis (EMSA; **Supplementary Fig. 6**) and observed concordant results with RNA-MITOMI, suggesting that the poly(A) tail does not affect the RNA-SLBP interaction.

MITOMI chip fabrication. MITOMI devices were fabricated as described previously[6] (**Supplementary Fig. 10**).

Surface chemistry. After 4 h of thermal bonding between the slide and MITOMI chip at 80 °C, the slide surface was derivatized using biotinylated BSA (bBSA; Thermo Fisher, 2 mg/ml), neutravidin (Thermo Fisher, 1 mg/ml) and HEPES buffer (pH 8.0) as described previously[6]. Derivitization resulted in a region of exposed avidin under each button valve on the MITOMI chip. After testing molecular beacon designs (**Supplementary Fig. 1**), we chose an ssDNA probe (IDT; 5Biosg//iSp18/ TTTTTTTTTTTTTTTTTTTTTTTTGGGGAA/36-FAM/), which was flowed for 30 min after derivtization and was pulled down specifically to the regions with exposed avidin under each button valve. After pull-down of the capture probe and 10 min of wash with buffer, the button valve was depressed. An *in vitro* transcription mix (MEGAshortscript IVT kit, Ambion) was prepared (3 μl water, 2 μl IVT buffer, 8 μl NTP mix, 4 μl enzyme, 1 μl superasin (Ambion) and 2 μl BSA), flowed across the chip for 30 min and finally flowed into the cDNA-containing back-chambers of each unit cell by releasing the neck valve. The neck valve was then depressed again, and transcription proceeded in the back-chambers at 37 °C for 30 min while the chip was flushed with buffer (HBS-EP binding buffer, GE Healthcare). After flushing was complete, the button and neck valves were released, and the sandwich was depressed to allow RNA to be simultaneously transcribed and to hybridize with capture probe in each unit cell for an additional hour at 37 °C. Once *in vitro* transcription was complete, a quencher probe (IDT; /5IAbFQ/ TTCCCCAAAAAAAAAAAAAAA) was used to confirm transcription and pull-down of RNA. The FAM dye on each capture probe slide was imaged using a Tecan LS Reloaded scanner at 488 nm excitation and FAM filter set.

SLBP expression and purification. SF9 insect cells were grown in suspension culture in SF900-II medium supplement with 10% FBS to a density of 2×10^6 cells/ml. Cells were infected with baculovirus produced using standard protocol from Invitrogen Bac-to-Bac system. Infection was allowed to proceed for 3 d. Cells were pelleted and lysed in buffer containing 50 mM Tris-HCl (pH 8.5), 5 mM 2-mercaptoethanol, 1 mM PMSF and 1% Nonidet P-40. Clarified lysate was then bound to Ni-NTA resin and recombinant protein was eluted into buffer containing 20 mM Tris-HCl (pH 8.5), 100 mM KCl, 100 mM imidazole, 5 mM 2-mercaptoethanol and 10% glycerol. Protein aliquots were stored at −80 °C.

SLBP preparation. A 3-μl aliquot of purified N-terminally GST-tagged SLBP at 1.8 μg/μl (~31 μM) was thawed, and 1 μl of SLBP was mixed with 5 μl of 1.0 μg/μl Texas Red–conjugated antibody to GST (Abcam ab34733) in 1,000 μl of 1× buffer (HBS-EP, GE Healthcare). The mix was incubated for 1 h at room temperature. After incubation, 88 μl of protein-antibody complex was mixed with 1 μl of 100 mg/ml BSA (Sigma), 1 μl of 1 mg/ml salmon sperm DNA (Ambion) and 10 μl of 10 mg/ml yeast tRNA (Ambion) for a maximum assayed protein concentration of ~30 nM. SLBP concentration was titrated by adjusting the abundance of protein-antibody complex in the final preparation. The mix was flowed across the chip for 30 min to ensure equilibration between RNA and protein before trapping the bound SLBP with the button valve and imaging on a Tecan LS reloaded scanner at 532 nm with the Texas Red filter set.

Array data analysis. Scanned files were analyzed using GenePix v6.0 software to determine FAM and Texas Red intensity under the button valve. The Texas Red intensity value for the wild-type stem-loop was used to normalize all other variants (**Supplementary Table 1**), resulting in normalized binding data. The natural log of relative intensity values is used for the scatter-plot comparisons between point and compensatory mutants, as relative binding is displayed as negative values relative to 0 (wild-type relative affinity).

Mobility shift experiments. Five femtomoles of uniformly labeled RNA were incubated on ice with varying amounts of recombinant SLBP in 10 mM HEPES (pH 7.6), 50 mM KCl, 0.1 mM EDTA, 10% glycerol, 1 μg/μl yeast tRNA and 0.1 μg/μl BSA on ice for 10 min. Reactions were directly loaded onto 8% native polyacrylamide gel (acrylamide:bisacrylamide was 29:1 in TBE buffer) without loading dyes. Gels were dried and visualized by autoradiography. Images were processed using the gel analysis route in ImageJ software. Quantified data are in **Supplementary Table 3**.

Phylogenetic analysis. Sequences were obtained from ref. 8 and conservation is quantified in **Supplementary Table 4**.