



Published in final edited form as:

*Nat Chem Biol.* 2017 March ; 13(3): 295–301. doi:10.1038/nchembio.2278.

## Recurrent RNA motifs as scaffolds for genetically encodable small molecule biosensors

Ely B. Porter<sup>a</sup>, Jacob T. Polaski, Makenna M. Morck<sup>b</sup>, and Robert T. Batey<sup>\*</sup>

Department of Chemistry and Biochemistry, University of Colorado at Boulder, Campus Box 596, Boulder Colorado 80309-0596

### Abstract

Allosteric RNA devices are increasingly viewed as important tools capable of monitoring enzyme evolution, optimizing engineered metabolic pathways, facilitating gene discovery and regulators of nucleic acid-based therapeutics. A key bottleneck in the development of these platforms is the availability of small molecule binding RNA aptamers that robustly function in the cellular environment. While aptamers can be raised against nearly any desired target by *in vitro* selection, many cannot be easily integrated into devices or do not reliably function in a cellular context. Here, we describe a new approach using secondary and tertiary structural scaffolds derived from biologically active riboswitches and small ribozymes. Applied to neurotransmitter precursors 5-hydroxytryptophan and 3,4-dihydroxyphenylalanine, this approach yields easily identifiable and characterizable aptamers predisposed for coupling to readout domains to engineer nucleic acid sensory devices that function *in vitro* and in the cellular context.

The means to generate synthetic RNA elements with new regulatory and sensing abilities is powerfully enabled by *in vitro* selection<sup>1</sup> and the pool of available synthetic aptamers is currently large<sup>2</sup>. However, only a few small molecule binding RNA aptamers have transitioned into effective and widely used intracellular biosensors or other devices<sup>3–5</sup>. This discrepancy between *in vitro* binding and intracellular activity is problematic, suggesting current selection strategies cannot easily access small molecule binding RNA aptamers capable of functioning robustly in the cellular environment<sup>6</sup>. While *in vivo* selection

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>\*</sup>Correspondence and request for materials should be addressed to R.T.B. Corresponding Author: Tel: (303)-735-2159, Fax: (303)-492-8425, [robert.batey@colorado.edu](mailto:robert.batey@colorado.edu).

<sup>a</sup>Current address: Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

<sup>b</sup>Current address: Department of Biochemistry, Stanford University, Palo Alto, California 94305

### Accession codes

Coordinates and structure factors have been deposited in the RSCB Protein Data Bank under the accession code 5KPY.

### Author contributions

E.B.P. conceived of this project and either performed or participated in all experiments of this project and in writing of the manuscript. J.T.P. performed the 5HTP and L-DOPA sensor construction, assay, and interpretation. M.M.M. participated in the GR/SSIII selection, screened for crystals of the 5GR-II/5HTP complex and collected crystallographic data. R.T.B. oversaw the project including planning experiments, interpretation of data and writing of the manuscript.

### Competing financial interests

The authors declare no competing financial interests.

Any supplementary information, chemical compound information and source data are available in the online version of the paper.

strategies for small molecule binding RNAs could generate cell-capable sensors, current methods have proven to be limited in diversity and application<sup>6</sup>. Thus, engineering of RNA devices responsive to small molecules continues to rely upon a protracted workflow incorporating traditional *in vitro* selection with tandem, application-specific selections for enhanced function<sup>7</sup>.

Unlike synthetic aptamers, small molecule binding domains of natural riboswitches have evolved in the context of the cell and incorporate additional features extending beyond the ligand binding site that include high fidelity folding and an ability to communicate with downstream regulatory switches to elicit their required function<sup>8</sup>. These aptamers are highly modular and robust, observed in a broad spectrum of bacterial species<sup>9</sup> and interface with diverse regulatory domains acting on transcription, translation, alternative splicing and mRNA stability. Thus, they are highly flexible with regards to mechanisms of communication with adjacent domains or sequences that elicit an output (e.g., gene regulation). Not surprisingly, these aptamers are highly successful in synthetic applications and the current validation workhorse of novel synthetic RNA tools<sup>10–12</sup>.

Here, we propose that recurrent structural folds found in natural RNA aptamers and small nucleolytic ribozymes can be reprogrammed using *in vitro* selection to host a broad spectrum of small molecule binding sites while preserving the robust folding and highly stable architectural properties of the parent. The logic of this approach is analogous to that applied in directed protein evolution where biologically recurrent “privileged” scaffolds such as the ( $\beta\alpha$ )<sub>8</sub> or immunoglobulin fold are used as starting points for the evolution of novel enzymes<sup>13</sup>. Using partially structured RNA libraries in the selection of small molecule binding aptamers has been previously employed<sup>5,14</sup>, but these simple hairpins and helices do not have the potential to form higher-order structure akin to natural aptamers.

Using scaffolds derived from two different riboswitch aptamer domains and a ribozyme, we have obtained a diverse set of aptamers that selectively bind two precursor molecules of the neurotransmitters serotonin and dopamine: 5-hydroxytryptophan and 3,4-dihydroxyphenylalanine, respectively. Biochemical characterization of a subset of 5-hydroxytryptophan binding aptamers reveals that while each of the scaffolds provides unique solutions for molecular recognition, they converge on similar binding affinities and discriminate against chemically related molecules. These aptamers are predisposed by the structural scaffold for coupling to widely used “Spinach” fluorogenic modules<sup>15</sup> to create biosensors that function *in vitro* and bacterial cells. The diversity of aptamers readily achieved using this approach enables more flexible strategies with less *in vitro* characterization required to implement practical RNA devices.

## RESULTS

### Libraries with information from biological RNAs

Examination of small biological RNAs with multi-helix packing suggests two recurrent architectures that are privileged scaffolds. The first is the H-type pseudoknot, which is broadly found in biological RNAs as well as *in vitro* selected aptamers<sup>16</sup> but from a design perspective is difficult to engineer. The other is the three-way junction (3WJ) supported by a

remote tertiary interaction that organizes helical arrangement around the junction<sup>17</sup>. This fold is more suitable for development of RNA devices as it positions a designable helical element (called the P1 helix) proximal to the functional site typically housed in the junction. Practical RNA biosensors using both natural and synthetic aptamers employing the 3WJ fold have already been developed<sup>10,18,19</sup>. Thus, the 3WJ fold is an ideal scaffold for selection of small molecule binding aptamers.

Within the 3WJ fold group, there is a large selection of potential candidates that can be used to scaffold an initial library of sequences for *in vitro* selection. For this study, we chose three: the aptamer domain of the *Bacillus subtilis xpt-pbuX* guanine riboswitch (referred to as “GR”)<sup>20</sup>, the aptamer domain of the *Vibrio cholerae* Vc2 cyclic di-GMP riboswitch (“CG”)<sup>21</sup>, and the *Schistosoma mansoni* hammerhead ribozyme (“HR”)<sup>22</sup> (Fig. 1). In each of these parental RNA scaffolds, the junction hosts the key biological activity proximal to the P1 helix that will serve as the secondary structural bridge to a readout domain.

Starting libraries were designed based on available crystal structures to preserve the overall secondary and tertiary structure of the scaffold while randomizing a sufficient number of nucleotides in the junction to ensure adequate pool diversity such that winners emerge<sup>23</sup>. All nucleotides in the joining strands of the junction were randomized as well as at least one base pair in each helix proximal to the junction (Fig. 1). For the GR scaffold this yielded an initial library of 23 randomized nucleotide positions equating to a library size of  $\sim 7 \times 10^{13}$  sequences; the CG and HR libraries contained similar levels of diversity. This number of sequences is theoretically fully represented in the initial pool of RNA with at least 5-fold redundancy. While this is substantially lower diversity than what is recommended for deep selection<sup>24</sup>, aptamers have been attained from starting pools with an even more limited sampling of sequence space<sup>25</sup>.

The three scaffolds were integrated into a library cassette with specific design features. The P1 helix of each scaffold containing the initial and terminal bases of the scaffold sequences was replaced in all libraries with a designed helix containing structured amplification cassettes based upon those developed for “SHAPE” chemical probing<sup>26</sup> (Supplementary Results, Supplementary Fig. 1). This ensured that the constant regions necessary for replication are structured and less likely to be incorporated as a necessary element for the selected aptamer. To further minimize the potential for the constant regions to participate in formation of the ligand-binding site, the P1 helix was extended to at least ten base pairs. Full sequences of DNA templates encoding the initial starting libraries are given in Supplementary Table 1.

A further complicating issue for scaffolded selection is the low fidelity of viral reverse transcriptases (RTs). Mutations in conserved sequences of the scaffold that disrupt secondary and/or tertiary structure yield RNAs that are amplified more efficiently by RT in the replication step, likely contributing to the “tyranny of small motifs” phenomenon observed in selection<sup>27</sup>. To address this, we adopted a recently characterized RT derived from a mobile group II intron from the thermophile *Geobacillus stearothermophilus*, GsI-IIc, that is more thermostable and has inherently higher fidelity than MMLV-derived RTs<sup>28</sup>. For comparison, the GR scaffold selection was performed with a commonly used RT derived

from MMLV (SuperScript III or “SSIII”) along with GsI-IIC (see Online Methods for complete details).

### Scaffolded selection against two aromatic amino acids

The targets for selection were 5-hydroxy-L-tryptophan (5HTP) and 3,4-dihydroxy-L-phenylalanine (L-DOPA), the immediate biosynthetic precursors of serotonin and dopamine, respectively (chemical structures of all key compounds are given in Supplementary Fig. 2). The amino acids were immobilized on a solid matrix via the carboxylate group and seven rounds of selection with each library were carried out with increasingly stringent washing procedures in later rounds. Counterselection against L-tryptophan or L-tyrosine was performed to ensure selectivity of the final aptamers against related amino acids that are abundant in the cell. In the SSIII selection with 5HTP a conventional SELEX protocol was adopted in which the affinity column was extensively washed in early rounds prior to competitive elution to remove nonbinding RNAs<sup>24</sup>. Competitive elution was initially observed in round 4 and peaked at >50% of total input RNA in round 6. In order to encourage sequence diversity in the GsI-IIC selections with 5HTP and L-DOPA, the concentration of MgCl<sub>2</sub> was lowered to 2.5 mM from 10 mM and a less stringent washing protocol was adopted where the final ~10% of total RNA left on the column under competitive elution was collected for amplification in the initial 4 rounds to preserve sequence diversity in the pool before increasing wash stringency<sup>27,29</sup>. Details of selections are given in the Online Methods and Supplementary Table 2.

In preserving sequence diversity and minimizing early stochastic events, we relied on a combination of next generation sequencing (NGS) and downstream bioinformatic analysis to reveal potential aptamers and elucidate key features of the selection. For each selection, >200,000 reads were obtained from the final round and resultant sequences were clustered and maximum-likelihood trees generated. Comparison of the SSIII and GsI-IIC selections using the GR scaffold revealed several important features. A distance matrix of the GR/SSIII selection clearly showed only a few isolated clusters, and within each cluster the sequences had a high degree of internal relatedness (Fig. 2a). The majority (>80%) of sequences clustered into three distinct sequence-related families, referred to as 5GR-I, -II, and -III, with the remaining clustering into small, difficult to interpret populations. This is typical of a traditional selection experiment where single isolates are often identified and further mutagenesis and selection are necessary to obtain covariation information. In contrast, the GR/GsI-IIC selection yielded more diverse clusters with a higher sampling of the sequences populating regions between major clusters (Fig. 2b). Consensus secondary structures of six major clusters from the two selections against 5HTP are shown in Fig. 2c. The CG and HR scaffolded selections with GsI-IIC were similarly diverse in their sequence space with many potential aptamers (Supplementary Fig. 3). While traditional selection approaches often rely upon overselection to facilitate finding an aptamer with limited sequence information, preservation of the diversity of winners and sequence analysis by NGS allows for a more thorough analysis of conservation and covariation patterns, aiding in determining consensus aptamer sequences<sup>30,31</sup>. Similar results were observed in the GR-scaffolded L-DOPA selection (Supplementary Fig. 4), indicating that these results are not idiosyncratic to 5HTP.

A subset of 250 sequences from each of the clusters that yielded validated 5HTP and L-DOPA aptamers is given as Supplementary Data.

In addition to the limited sequence diversity of the SSIII selection, a heavy accumulation of deletions and point mutations was observed such that no sequences retaining the full identity of the constant regions of the scaffold were recovered in the final round. Two of the clusters, 5GR-I and 5GR-III, have deletions in L2 or L3 of the scaffold essential for tertiary structure. In addition, 5GR-III members acquired several point deletions resulting in an alternative minimal free energy (MFE) secondary structure, and covariation analysis of this cluster indicates a consensus sequence consistent with the L-tryptophan aptamer<sup>32</sup>. 5GR-II is the only major cluster maintaining sequence requirements necessary for the tertiary structure designed into the library and was the only abundant sequence shared between the SSIII and GsI-IIC selections. In contrast to the selection using SSIII, the GsI-IIC selections exhibited low amounts of mutations accruing in the scaffold's constant region (Supplementary Fig. 5).

To identify sequences with high aptamer potential, the ten most populous clusters from each pool were individually aligned, MFE structures predicted and covariation models generated<sup>33,34</sup>. This allowed for an information-rich view of the major consensus sequences presented by the selections. Since the GR/SSIII experiment was highly overselected, the most abundant sequence from each of the three major clusters was chosen for further validation. For the GsI-IIC selections, the dominant sequences from one or more clusters whose consensus MFE structure is consistent with the parent scaffold were selected.

### Scaffolded aptamers bind 5HTP with high selectivity

The structural scaffold greatly facilitates validation of the structural and interaction features of resultant 5HTP aptamers. Chemical probing of RNA structure using *N*-methylisatoic anhydride (NMIA, a technique referred to as "SHAPE"<sup>35</sup>) reveals whether the secondary and tertiary architecture of the parental scaffolds were preserved as well as ligand-dependent structural changes in the aptamer. In the GR/SSIII selection, the 5GR-I and 5GR-II aptamers had localized changes in the NMIA reactivity patterns in the presence of ligand within the 3WJ elements, consistent with this being the ligand-binding site (Fig. 3a; Supplementary Fig. 6). 5GR-III, however, showed changes outside of J2/3 in the constant regions, consistent with the predicted structure and tryptophan binding site of a previously described aptamer<sup>32</sup>. Preservation of the GR scaffold was assessed using a unique, ligand-independent NMIA reactivity signature in L3 that is only present when it interacts with L2<sup>36</sup>. 5GR-II is the only sequence from the three clusters of the SSIII selection exhibiting this feature; conversely, all tested sequences from the GR/GsI-IIC selection had this tertiary structure signature (Supplementary Fig. 7). These data strongly indicated that the GR/SSIII selection yielded three distinct aptamers with only 5GR-II preserving the structural scaffold, whereas the GR/GsI-IIC selection produced multiple solutions maintaining the scaffold. While the 5HTP-dependent signatures for the GR/GsI-IIC isolates were weaker than those of 5GR-II and the parental aptamer, quantification revealed that they localized to the junction in an analogous manner (Fig. 3b). SHAPE characterization of a representative sequence from the most populous cluster from the CG/GsI-IIC and HR/GsI-IIC selections (5CG-I and 5HR-I,

respectively) showed ligand-dependent changes for the new classes of aptamers and an overall reactivity pattern resembling the parental scaffold (Supplementary Figs. 8, 9).

The affinity and selectivity of these aptamers for 5HTP and a set of chemically similar compounds was assessed by isothermal titration calorimetry (ITC). Importantly, for all of the tested aptamers, the 5' - and 3' -cassette sequences were not required for 5HTP binding, indicating the successful design of neutral sequences (Supplementary Fig. 10). Several trends emerged from this analysis. First, both aptamers that do not preserve the parent scaffold (5GR-I, -III) do not discriminate between 5HTP and L-tryptophan (Table 1), a crucial requirement for cell-based applications. Second, the majority of the aptamers preserving the GR scaffold had higher affinities for 5HTP than the aptamers with disrupted scaffolds and all strongly discriminated against L-tryptophan. This suggested that the architecture of the scaffold was important for creating a selective binding pocket and was able to maintain affinities comparable to other synthetic and natural aptamers that bind amino acids<sup>23,37</sup>. In addition to binding the positively charged  $\alpha$ -methyl-5-hydroxy-L-tryptophan (Me-5HTP), which serves as a close approximation to the column due to the  $\alpha$ -ester, they also bound serotonin, the decarboxylation product of 5HTP. This suggested a lesser requirement for the main chain atoms in binding. Most striking from the GsI-IIC selections was that the dominant aptamers from each selection, despite having different scaffold architectures, converged on highly similar binding affinities and selectivity profiles, revealing that 3WJs were a robust fold for hosting 5HTP binding pockets. Together, these data show that distinct 3WJ architectural variants were able to find robust solutions for 5HTP recognition and yielded comparable affinities to more simplistic selection strategies that are encumbered by low complexity.

### A recurrent RNA motif is used for 5HTP binding

To further demonstrate that the scaffolded selection strategy preserved the fold of the parental RNA and to elucidate how RNA recognized 5HTP, the structure of 5GR-II complexed with 5HTP was determined at 2.0 Å resolution (Fig. 3c; representative electron density maps are shown in Supplementary Fig. 11 and crystallographic statistics are given in Supplementary Table 3). This structure globally superimposed with the parental guanine riboswitch aptamer<sup>38</sup> with an r.m.s.d. of 6.5 Å over all backbone atoms in residues 19–77, with the main sources of deviation produced by a different angle for P1 in relation to the binding pocket and the varied junction region (Supplementary Fig. 12). Within the L2-L3 tertiary interaction the pattern of base-base interactions and backbone geometry was almost identical between the two RNAs (r.m.s.d. 0.96 Å over all atoms in residues 31–39, 61–67). Thus, the GR scaffold remained intact, both globally and locally, during the selection process.

The ligand binding pocket of 5GR-II resides within the 3WJ and has a radically different local structure from the parent RNA. Direct ligand contacts were primarily mediated by nucleotides in J2/3 using a common RNA structural module, the T-loop (Fig. 3d). The first five nucleotides of J2/3 formed a canonical T-loop structure superimposing almost perfectly with a tRNA<sup>Phe</sup> T-loop (r.m.s.d. 0.49 Å for backbone residues). Stabilization of position 3 in the tRNA T-loop by long range Watson-Crick pairing with the D-loop is critical for

activity<sup>39</sup>; 5GR-II possessed a similar interaction between G47 of the T-loop and C75 of J3/1. The 5GR-II T-loop hosted 5HTP stacked between positions 4 and 5 in a manner analogous to how the tRNA T-loop hosts an intercalating purine from the D-loop. Similarly, thiamine pyrophosphate (TPP) is recognized by its cognate riboswitch using the T-loop motif<sup>40,41</sup> (Fig. 3e). While the T-loop was directly responsible for recognition of 5HTP, nucleotides from all three randomized regions were involved in local structure aiding in the formation of a compact junction that stabilized the T-loop.

The crystal structure of 5GR-II yielded additional insights into 5HTP recognition by other scaffolded aptamers. The most abundant cluster in the GR/GsI-IIC selection, 5GR-IV, also contained the UUGAA signature of the T-loop. The motif was 3'-shifted by a single nucleotide, likely leading to an alternative orientation within the 3WJ as suggested by significant differences in the composition of J1/2 and J3/1 between 5GR-II and 5GR-IV. In the HR selection, the most abundant sequence of the most populous cluster (5HR-I) also contained the conserved UUGAA sequence of the T-loop in J2/3. Sequence variation analysis of this region of the 5HR-I aptamer revealed a pattern of conservation matching that of the biological T-loops<sup>39,42</sup> with only slight deviations (Supplementary Fig. 13).

### Scaffolded aptamers yield robust small molecule sensors

To create functional synthetic RNA biosensors from scaffolded aptamers a conceptually simple and established strategy of linking a small molecule binding aptamer to a fluorophore-binding module via a short helical element referred to as the communication module (CM) was employed (Fig. 4a)<sup>5,43</sup>. This fusion employed the biologically active fluorogenic RNA aptamer, “Broccoli”<sup>15</sup>, and was further coupled to a tRNA scaffold to stabilize the biosensor for cell-based applications<sup>44</sup>. Four different GR-scaffolded 5HTP aptamers were coupled to four communication modules of differing lengths (Fig 4a) and each biosensor tested for the ability to fluoresce in a ligand-dependent fashion. Each sensor was assessed for their ligand-dependent fold change in fluorescence and maximal brightness relative to the isolated Broccoli aptamer both *in vitro* (Fig. 4b; Supplementary Tables 4 and 5) and in *Escherichia coli* (Figure 4c; Supplementary Tables 6 and 7). These data revealed three sensors (5GR-II, -IV, and -V) that can detect 5HTP and/or serotonin both *in vitro* and in the cellular context, with 5GR-II demonstrating the best performance with respect to combined fold increase in fluorescence and maximal brightness.

The above 5HTP biosensors were designed with knowledge from biochemical and biophysical analysis of select aptamers. However, an optimal workflow for rapid development of biosensors would be able to use information derived only from the computational analysis of the selection to design candidate RNAs. To demonstrate that scaffolded aptamers incorporate design principles that enable biosensor engineering in the absence of experimental characterization, the above biosensor strategy was employed for four aptamers derived from the L-DOPA selection. None of these aptamers were validated in any fashion prior to their incorporation into allosteric fluorogenic sensors. Screening of the resultant biosensors with L-DOPA and dopamine *in vitro* (Fig. 4d; Supplementary Tables 4 and 5) and in *E. coli* (Fig. 4e; Supplementary Tables 6 and 7) revealed two aptamers (DG-I and DG-II) that function in both contexts.

To further demonstrate the potential of scaffolded aptamers, live cell imaging was used to visualize the uptake of 5HTP by *E. coli* using the 5GR-II/CM-4 biosensor. Imaging of single cells revealed a rapid induction of fluorescence upon addition of 2 mM 5HTP to *E. coli* growing in a rich chemically defined medium, with ~80% of bacteria displaying an observable response within 20 minutes (Fig. 5a, b). The fluorescence signal was completely dependent upon the RNA device binding 5HTP; no detectable signal gain was observed when L-tryptophan is included in the media (Supplementary Fig. 14) or when the sensor contained a point mutation (A48U) in the T-loop module that ablates ligand binding to the isolated aptamer (Fig. 5c, d). The observed increase in relative fluorescence in the presence of 5HTP is comparable to robust cyclic dinucleotide sensors based upon natural riboswitch aptamer domains in live cells<sup>10,45</sup>.

## Discussion

A strategy has been presented that exploits the secondary and tertiary structural architecture of naturally evolved riboswitches and ribozymes to scaffold small molecule binding pockets raised through *in vitro* selection. One key strength of the described approach is the use of multiple scaffolds in parallel selections to obtain a suite of aptamers. This differs significantly from traditional selections where the same subset of solutions is reproducibly generated from a simple randomized pool which significantly constrains sensor diversity and development<sup>23</sup>. While the aptamers derived from different scaffolds have similar affinities for 5HTP and selectivity against L-tryptophan, they clearly have distinct characteristics with respect to their ability to communicate with a readout domain via the P1 helix. We hypothesize that differences in sensor performance across different aptamers is in part due to variation in the spatial relationship between the ligand and the interdomain (P1) helix, a feature that cannot be fully controlled in the selection. However, unlike deep selections, the scaffolded selection approach strongly biases selections towards a favorable ligand/P1 orientation by constraining the possible ligand position.

With a suite of aptamers, combinatorial approaches can be employed to rapidly screen for sensors without extensive aptamer characterization or device optimization as exemplified by the dopamine sensor development herein. Development of an RNA device from aptamers derived from deep selections requires thorough characterization along with broad screening of communication modules while leaving the sensory aptamer as a fixed, unalterable node due to the lack of diversity. With the scaffolded selection approach, a set of distinct aptamers can be combinatorially coupled to a set of communication modules and rapidly screened for variants with the desired activity, as demonstrated with the L-DOPA selection. In this fashion, our approach should facilitate the expedient development of RNA devices and sensors by easing a key bottleneck. Notably, while we focused upon only the most populous clusters in each selection for characterization and/or sensor design, within each selection there are many clusters containing alternative sequences that could further enrich the initial pool of aptamers for developing downstream applications.

A second powerful advantage of this selection strategy is the potential for robust folding in the cellular context provided by the tertiary interaction of the 3WJ architecture. Each of these aptamers has a fold that has undergone extensive biological evolution. Further, distal



tertiary interactions that organize the 3WJ core can be highly stable; both the L2-L3 interaction of the purine riboswitch and the tetraloop-tetraloop receptor of the cyclic di-GMP riboswitch scaffold stably form outside the context of other RNA structure<sup>46,47</sup>. The presence of robust secondary and tertiary structure in the scaffold enables these elements to potentially guide the folding of all members of the initial library. In contrast, RNA misfolding during selection and or the presence of multiple MFE structures in the final aptamers is often a significant problem for traditional deep selection. Since there is no significant selection pressure for high fidelity folding in a typical selection protocol, providing this information in the starting library can be a path towards robust folding RNAs.

While 3WJ scaffolds were chosen as the focus of this study, the diversity of natural riboswitches and ribozymes can provide further feedstock for this approach. Within the 3WJ family, there is a broad array of sequences that vary the orientation of the three helices, size of the joining regions, and the nature of the distal tertiary interaction that may provide superior scaffolds for a particular ligand or sensor<sup>17</sup>. Furthermore, other folds may be predisposed to bind a target small molecule based on the nature of the cognate ligand. For example, larger ligands may be more easily recognized by flavin mononucleotide or cobalamin riboswitch-derived scaffolds, while dinucleotides such as NADH may be readily accommodated by one of the di-cyclic nucleotide aptamers. As natural RNA aptamers have been discovered to recognize chemically diverse small molecules, exploiting their architectures towards the selection of novel aptamers has the potential to facilitate the development of powerful new tools for monitoring and responding to small molecules in the cellular environment across a broad range of applications.

## Online Methods

### Library construction

For each scaffold, nucleotides within an 8 Å shell surrounding the ligand binding site or active site of the parent RNA were identified from their crystal structure (GR, PDB ID 4FE5; CG, PDB ID 3IWN; HR, PDB ID 3ZD5). The corresponding positions were randomized in a DNA ultramer that spanned the entire aptamer domain with conserved flanking sequences for reverse transcription and amplification (Integrated DNA Technologies; sequences of all nucleic acids used in this study are provided in Supplementary Table 1). ssDNA was converted into dsDNA templates for transcription using standard Taq PCR conditions in which  $\sim 2 \times 10^{-12}$  mol DNA (corresponding to  $\sim 10^{12}$  individual sequences) was used in each 100 µL PCR reaction and amplified for 15 cycles. Approximately  $1 \times 10^{14}$  sequences were transcribed in a 12.5 mL transcription reaction containing 40 mM Tris-HCl, pH 8.0, 25 mM DTT, 2 mM spermidine, 0.01% Triton X-100, 4 mM each rNTP pH 8.0, 0.08 units inorganic phosphatase (Sigma-Aldrich, lyophilized powder), and 0.25 mg/mL T7 RNA polymerase and incubated at 37 °C for 4 hours. Transcription samples were then precipitated in 75% ethanol at -20 °C, pelleted, and reconstituted in a solution of 300 µL formamide, 3 mL 8 M Urea, and 300 µL 0.5 M EDTA pH 8.0. Full length RNA was purified with a denaturing 8%, 29:1 acrylamide:bisacrylamide gel. Product RNA was excised from the gel after visualizing by UV shadowing and eluted in 0.3 M NaOAc pH 5.0 before exchange and storage in 0.5× TE.

### Synthesis of 5HTP affinity column matrix

For the derivatized columns, 3 mL bed volume of EAH Sepharose 4B (GE Healthcare) was dehydrated with dimethylformamide (DMF). 10  $\mu$ moles of Fmoc-5-hydroxy-L-tryptophan and 10  $\mu$ moles of benzotriazol-1-yl-oxytrityrrolidinophosphonium hexafluorophosphate (PyBOP) were dissolved in 1 mL of DMF and added to the dehydrated column with 20  $\mu$ moles N,N-diisopropylethylamine (DIPEA) and incubated with agitation for 2 hours at room temperature. The column matrix was then drained and washed extensively with DMF. Unreacted sepharose amines were acetylated by adding 1 mmole of acetic anhydride and 1 mmole DIPEA in approximately 1 mL DMF and mixed at room temperature for 1 hour. The column was drained of the acetylating mixture and washed with DMF prior to Fmoc deprotection using 20% v/v piperidine/DMF. Amino acid concentration on the column was determined by measuring the concentration of Fmoc in the deprotection fractions ( $A_{301\text{ nm}} = 8000\text{ M}^{-1}\text{ cm}^{-1}$ ). This method generated approximately 0.5–1 mM deprotected amino acid per mL resin. For counter selection, EAH sepharose was prepared exactly the same except omitting the ligand coupling step resulting in acetylated sepharose.

### *In vitro* selection

For the GR scaffold selection using the SuperScript III reverse transcriptase (GR/SSIII), 350  $\mu$ L acetylated sepharose was equilibrated in selection buffer (10 mM Na-HEPES, pH 7.0, 250 mM NaCl, 50 mM KCl, 10 mM  $\text{MgCl}_2$ , 0.1 mg/mL tRNA) and 1 nmol of library RNA in 350  $\mu$ L of selection buffer was incubated at room temperature for 30 minutes with agitation. The applied solution was removed and the column matrix washed once with 350  $\mu$ L of selection buffer. The pooled flow through and wash (750  $\mu$ L total) was added to pre-equilibrated 5HTP-derivatized sepharose 4B column and incubated for 45 minutes. The column was then drained and washed three times with selection buffer before elution with 10 mM 5HTP in selection buffer (two 1 hour incubations in 350  $\mu$ L; total 700  $\mu$ L eluted volume). The eluted fractions were then concentrated to 50  $\mu$ L in a 0.5 mL Ultracel 10kD MWCO filter (Millipore) and ethanol precipitated in 0.3 M sodium acetate (pH 5.0), 5  $\mu$ g glycogen, and brought to a final concentration of 75% ethanol before storage at  $-70\text{ }^\circ\text{C}$  for 30 minutes. Details of the conditions of each cycle are provided in Supporting Table 2.

To convert the competitively eluted RNA into a new population of RNA, the elution fractions were ethanol precipitated, pelleted at  $13000 \times g$  at  $4\text{ }^\circ\text{C}$ , decanted, and dried under vacuum. The dried pellet was reconstituted with 0.7 mM each dNTP, 7  $\mu$ M RT-PCR primer, and brought to a total volume of 14  $\mu$ L before heating to  $65\text{ }^\circ\text{C}$  for 5 minutes and incubation on ice for 10 minutes. The solution was then brought up to  $1\times$  SuperScript III first-strand buffer ( $5\times$ : 250 mM Tris-HCl, pH 8.3, 375 mM KCL, 15 mM  $\text{MgCl}_2$ ) with 5 mM DTT and 200 units SuperScript III (Life Technologies) in a total volume of 20  $\mu$ L before a 15 minute extension at  $54\text{ }^\circ\text{C}$ . The entire 20  $\mu$ L reverse transcription solution was PCR amplified in a total volume of 500  $\mu$ L using standard Taq DNA polymerase conditions. The amplified pool was then transcribed by adding 100  $\mu$ L of the PCR reaction to a 1 mL transcription reaction containing 40 mM Tris-HCl, pH 8.0, 25 mM DTT, 2 mM spermidine, 0.01% Triton X-100, 4 mM each rNTP pH 8.0, 0.08 units inorganic phosphatase, and 0.25 mg/mL T7 RNA polymerase and incubated at  $37\text{ }^\circ\text{C}$  for 2 hours. A 100  $\mu$ L transcription reaction for  $^{32}\text{P}$  labeled RNA was performed under similar condition with the exception that the rNTPs were

lowered to 2 mM for UTP, CTP, and GTP while ATP was reduced to 200  $\mu$ M and  $\sim$ 100  $\mu$ Ci  $^{32}$ P-ATP. Transcription samples were gel purified as described above with the gel loading conditions scaled accordingly.

Selections using the GsI-IIC reverse transcriptase were performed as described above with the following changes. The buffer for selection contained a reduced magnesium concentration and more physiologically relevant monovalent cations: 25 mM Na-HEPES, pH 7.0, 150 mM KCl, 50 mM NaCl, 3 mM MgCl<sub>2</sub>. GsI-IIC reverse transcriptase was used in place of SuperScript III. GsI-IIC reverse transcriptase was expressed in *E. coli* and purified as described by Mohr et al.<sup>28</sup>. The precipitated RNA pellet was brought up in 1.25 mM dNTPs and 20  $\mu$ M RT-PCR primer prior to denaturation at 65 °C, annealing at 4 °C, and equilibration at 60 °C. The solution was then brought up to 1 $\times$  GsI-IIC buffer conditions (10 mM NaCl, 1 mM MgCl<sub>2</sub>, 20 mM TrisCl, pH 7.5, 1 mM DTT) in 20  $\mu$ L total volume and sufficient enzyme was added for extension at 60 °C. PCR was as performed as described above.

### High throughput sequencing and bioinformatic analysis

Standard PCR was conducted to append the Illumina hybridization sequences necessary for annealing to the flow cell. Each library was amplified with the forward sequencing primer and a unique reverse primer containing a distinguishing 12 nucleotide barcode (sequences are given in Supplemental Table 1). The samples were sequenced using a v3 reagents kit for 150 cycles on a MiSeq (Illumina) with custom read and indexing primers.

The resulting sequences were demultiplexed, trimmed, and quality filtered using scripts from QIIME<sup>48</sup>. All sequence information outside of the P1 stem was trimmed and only sequences containing a Phred score  $\geq$  20 for each nucleotide were used in the analysis. The resulting fasta format files for each library were then subjected to clustering by USEARCH<sup>49</sup>, which generated seed sequences that were clustered at 90% identity; any clusters containing a single sequence were discarded. The top ten populous clusters were then mapped back to their original sequence file, and 250 individual sequences were randomly taken as a representative sample of each cluster for further analysis. Sequences in each cluster were aligned using MUSCLE<sup>50</sup> and the resultant alignment analyzed using CMfinder<sup>33</sup>. R2R<sup>34</sup> was run at its default settings to generate figures of the sequence conservation mapped onto the minimum free energy (MFE) secondary structure.

### NMIA chemical probing

RNA was prepared as described previously<sup>51</sup>. Structure cassettes flanking the 5' and 3' ends of the RNA were added to facilitate reverse transcription and NMIA modification was performed using the established protocols<sup>26</sup> at 25 °C. RNA was probed at 100 nM in 100 mM Na-HEPES, pH 8.0, 100 mM NaCl, and 6 mM MgCl<sub>2</sub>. Ligand concentration was 500  $\mu$ M where indicated. Gel images were analyzed by SAFA<sup>52</sup> and ImageJ (NIH).

### Isothermal Titration Calorimetry (ITC)

All RNAs tested were exchanged into the SSIII selection buffer (10 mM Na-HEPES, pH 7.0, 250 mM NaCl; 50 mM KCl; 10 mM MgCl<sub>2</sub>) and washed three times in 10 kD MWCO filter

(EMD Millipore). The ligand was brought up from a dry solid directly into the binding buffer and concentration established on a NanoDrop 2000 (Thermo Scientific) using an extinction coefficient at 275 nm of  $8000 \text{ M}^{-1} \text{ cm}^{-1}$  for the 5-hydroxyindole moiety. The RNA was diluted to between 50–100  $\mu\text{M}$  and the ligand was titrated at roughly 10 times the concentration of RNA. Titrations were performed at 25 °C using a MicroCal iTC200 microcalorimeter (GE Healthcare) using established protocols<sup>53</sup>. Data was analyzed and fitting was performed with the Origin 5.0 software suite (Origin Laboratories).

### Structure determination of the 5GR-II/5HTP complex

RNA for crystallization was prepared as previously described<sup>51</sup>. The RNA was concentrated in an Amicon Ultra 15 10k MWCO filter (EMD Millipore, Inc.) and exchanged into 0.5× T.E. buffer. Diffraction quality crystals were obtained by mixing 2  $\mu\text{L}$  RNA:ligand complex (1:1) and 3.5  $\mu\text{L}$  mother liquor (8–14% 2-methyl-2,4-pentanediol, 40 mM sodium cacodylate pH 5.5, 4 mM  $\text{MgCl}_2$ , 12 mM NaCl, 80 mM KCl, and 4–9 mM cobalt hexamine), micro-seeding, and incubation at 22 °C for 1–3 days. The crystals needed no further cryoprotection and were flash frozen in liquid nitrogen before data collection. Data was collected with a Rigaku R-Axis IV image plate system using  $\text{CuK}\alpha$  radiation (1.5418 Å) at 100 K, and was indexed and scaled using D\*TREK<sup>54</sup>. Data on a heavy atom derivative made by replacing cobalt hexamine with 1–11 mM iridium hexamine was also collected on the home x-ray source. Phases were determined using the single isomorphous replacement with anomalous scattering (SIRAS) method. AutoSol<sup>55</sup> was used to find 12 iridium atoms that were then used to calculate phases. The resulting experimental density map displayed unambiguous features of the RNA backbone and helices and was used for building the model.

The initial model was iteratively built without the ligand in Coot<sup>56</sup> between rounds of refinement in PHENIX<sup>55</sup>. The RNA model was brought through several rounds of refinement and simulated annealing before 5HTP was built into the model. At this point of building, there was clear ligand density in the binding pocket that allowed for the confident placement and orientation of the ligand. The placement of the ligand and bases was validated by a composite omit map (Supplementary Figure 4b). Water placement was automated in final rounds of refinement after ligand placement based on peak size in the  $F_o - F_c$  difference map. The resultant model had good geometry as judged using MolProbity<sup>57</sup> and final model statistics ( $R_{\text{work}}$  and  $R_{\text{free}}$  are 21.9% and 26.2%, respectively). All crystallographic data and model statistics are given in Supplementary Table 3.

### *In vitro* Broccoli sensor assays

Engineered sensors were synthesized as G-blocks (sequences of sensors given in Supplementary Table 1; Integrated DNA Technologies) and cloned between the XbaI and BlnI sites in pET30b using standard molecular cloning techniques; all resultant plasmids were sequence verified. For T7 RNA polymerase transcription reactions, a DNA template was generated by PCR using 1  $\mu\text{M}$  outer primers (5': GGCCGTAATACGACTCACTATAGGAGCCCGGATAGCTCAGTCGGTAGAGCAG, 3': TGGCGCCCGAACAGGGACTTGAACCCTGGA) using a standard PCR reaction. Templates were directly added to an *in vitro* transcription reaction (see above) and RNA

synthesis was allowed to proceed for 2 hours at 37 °C. RNA from the above transcription reaction was directly used in assays without further purification.

The activity of each sensor was monitored in a 100  $\mu$ L reaction containing 50  $\mu$ L of *in vitro* transcription reaction, 10  $\mu$ L of 10 $\times$  Survey Buffer (1 $\times$ : 50 mM K-HEPES, pH 7.5, 10 mM MgCl<sub>2</sub>, 150 mM KCl, 50 mM NaCl), 30  $\mu$ M DFHBI-1T and 2 mM ligand (for plus ligand reactions). Reactions were incubated at room temperature for 30 minutes and DFHBI fluorescence was measured by placing 90  $\mu$ L reaction volume in a Greiner 96-well flat bottom black fluorescence plate (Thermo Scientific) and reading in a Tecan Infinite M200 PRO plate reader. Samples were excited at 460 nm and fluorescence emission measured as the average signal between 506 and 510 nm. For all experiments, a positive control of a tRNA-scaffolded Broccoli aptamer was performed in the presence and absence of ligand, which was also used as a reference for relative brightness. Fold induction was calculated by dividing the fluorescence values for the DFHBI-1T plus ligand reaction by the fluorescence value for the DFHBI-1T condition alone. All experiments were performed in triplicate and quantified data reported with the standard error of the mean (s.e.m.)

### ***In vivo* Broccoli sensor assays**

*E. coli* One Shot<sup>®</sup> BL21 Star (DE3) cells (Thermo Fisher) were transformed with a pET30b-derived plasmid containing a sensor under inducible control, plated onto LB agar supplemented with 50  $\mu$ g/mL kanamycin and incubated at 37 °C for ~16 hours. Individual colonies were picked and grown overnight (~16 hours) in 5 mL of LB supplemented 50  $\mu$ g/mL kanamycin to allow the culture to reach saturation. For screening experiments, 5  $\mu$ L of the saturated overnight culture was added to 5 mL of LB supplemented with 50  $\mu$ g/mL kanamycin and grown to mid-log phase (OD<sub>600</sub> ~0.4–0.6) at 37 °C. To induce expression of the broccoli aptamer alone or the broccoli/riboswitch aptamer fusion constructs, IPTG was added to a final concentration of 1 mM in each culture, which were then grown for an additional 2 hours at 37 °C. Cells were then pelleted by centrifugation and washed once with 5 mL of 1 $\times$  M9 salts supplemented with MgSO<sub>4</sub> at a final concentration of 5 mM and kanamycin at a final concentration of 50  $\mu$ g/mL. After washing, cells were pelleted by centrifugation, resuspended in 250  $\mu$ L of the above M9 medium and split into two 100  $\mu$ L aliquots. In half of the aliquots, DFHBI-1T was added to a final concentration of 50  $\mu$ M in a final volume of 110  $\mu$ L. In the other half of the aliquots, DFHBI-1T was added to a final concentration of 50  $\mu$ M and the ligand (5HTP, 5HP or dopamine) was added to a final concentration of 1 mM in a final volume of 110  $\mu$ L. Cells were then incubated at 37 °C for 30 minutes to allow for uptake of each compound. Following the 30 minute incubation, 100  $\mu$ L of each aliquot was pipetted into a Greiner 96-well black microplate and chilled on ice for 30 minutes. For fluorescence measurements, DFHBI-1T was monitored at an excitation wavelength of 472 nm and a 520 nm emission wavelength. Quantified data represent the average fluorescence values  $\pm$  standard error of the mean (s.e.m.) from three biological replicates, which were background corrected using a pET30b empty vector control. Fold induction was calculated by dividing the average fluorescence values of cells exposed to ligand by the average fluorescence of cells without ligand.

## Intracellular fluorescence imaging of 5HTP

DNA and cultures were prepared as previously described<sup>58</sup>. Briefly, the tRNA/Broccoli fusion sequence was cloned into pET30b between the XbaI and BlnI sites downstream of an inducible T7 promoter. The sequence verified plasmid was transformed into BL21(DE3)-STAR cells (Invitrogen) and single colonies were grown up overnight in Luria Broth (LB) supplemented with 50 µg/mL kanamycin. The overnight culture was used to inoculate fresh LB/kanamycin medium at a 1:1000 dilution and the culture grown at 37 °C to an OD<sub>600</sub> = 0.4–0.6 before induction with 1 mM IPTG and growth at 37 °C for 2–4 hours. 200 µL of the resultant culture was centrifuged, decanted, and resuspended in 2 mL of M9 minimal salts medium supplemented with 50 µg/mL kanamycin, 5 mM MgSO<sub>4</sub>, and 1 mM IPTG. 200 µL of the resuspended culture was transferred to 96-well poly-D-lysine coated glass bottom plates (MatTek) and incubated at 37 °C for one hour. The media was then removed and the wells washed with M9/kanamycin/1 mM IPTG medium before adding 200 µL of M9 media, 1 mM IPTG, and 400 µM DFHBI-1T (Lucerna). The live fluorescence images were taken with an Andor iXon3 897 EMCCD using a 60× oil objective, an excitation filter 472/30, dichroic mirror 490 (long pass) and emission filter 520/40 on a Nikon Ti-E microscope and analyzed with FIJI<sup>59</sup>.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors would like to thank Profs. Beatrix Suess and Alexander Heckel for the kind gift of α-methyl-5-hydroxy-L-tryptophan, Prof. Rob Knight and Joshua Shorenstein for useful discussions about bioinformatics analysis of selections, Prof. Alan Lambowitz for the generous gift of the GsI reverse transcriptase, Prof. Michael Yarus for the amino acid column labeling protocol, and Prof. David McKay and Dr. Jeremiah Trausch for crystallographic support and Alexandra Young and Prof. Amy Palmer for assistance with microscopy. Prof. Jeffrey Kieft, Dr. Jennifer Pfingsten and Michal Matyjasik gave critical feedback on the manuscript. This work was supported by grants to R.T.B. from the National Science Foundation (NSF 1150834) and the National Institutes of Health (R01 GM073850) and a predoctoral training grant in Signaling and Cellular Regulation (T32 GM008759).

## References

1. Darmostuk M, Rimpelova S, Gbelcova H, Ruml T. Current approaches in SELEX: An update to aptamer selection technology. *Biotechnol Adv.* 2015; 33:1141–1161. [PubMed: 25708387]
2. McKeague M, Derosa MC. Challenges and opportunities for small molecule aptamer development. *J Nucleic Acids.* 2012; 2012:748913. [PubMed: 23150810]
3. Berens C, Thain A, Schroeder R. A tetracycline-binding RNA aptamer. *Bioorganic & Medicinal Chemistry.* 2001; 9:2549–2556. [PubMed: 11557342]
4. Jenison RD, Gill SC, Pardi A, Polisky B. High-Resolution Molecular Discrimination by RNA. *Science.* 1994; 263:1425–1429. [PubMed: 7510417]
5. Paige JS, Wu KY, Jaffrey SR. RNA mimics of green fluorescent protein. *Science.* 2011; 333:642–646. [PubMed: 21798953]
6. Berens C, Groher F, Suess B. RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol J.* 2015; 10:246–257. [PubMed: 25676052]
7. Sharma V, Nomura Y, Yokobayashi Y. Engineering complex riboswitch regulation by dual genetic selection. *J Am Chem Soc.* 2008; 130:16310–16315. [PubMed: 18998646]

8. Garst AD, Edwards AL, Batey RT. Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol.* 2011; 3
9. Barrick JE, Breaker RR. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* 2007; 8:R239. [PubMed: 17997835]
10. Kellenberger CA, Wilson SC, Sales-Lee J, Hammond MC. RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. *J Am Chem Soc.* 2013; 135:4906–4909. [PubMed: 23488798]
11. Ketterer S, Gladis L, Kozica A, Meier M. Engineering and characterization of fluorogenic glycine riboswitches. *Nucleic Acids Res.* 2016; 44:5983–5992. [PubMed: 27220466]
12. Su Y, Hickey SF, Keyser SG, Hammond MC. In Vitro and In Vivo Enzyme Activity Screening via RNA-Based Fluorescent Biosensors for S-Adenosyl-L-homocysteine (SAH). *J Am Chem Soc.* 2016; 138:7040–7047. [PubMed: 27191512]
13. Romanini DW, Peralta-Yahya P, Mondol V, Cornish VW. A Heritable Recombination system for synthetic Darwinian evolution in yeast. *ACS Synth Biol.* 2012; 1:602–609. [PubMed: 23412545]
14. Davis JH, Szostak JW. Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc Natl Acad Sci U S A.* 2002; 99:11616–11621. [PubMed: 12185247]
15. Filonov GS, Moon JD, Svensen N, Jaffrey SR. Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J Am Chem Soc.* 2014; 136:16299–16308. [PubMed: 25337688]
16. Peselis A, Serganov A. Structure and function of pseudoknots involved in gene expression control. *Wiley Interdiscip Rev RNA.* 2014; 5:803–822. [PubMed: 25044223]
17. de la Pena M, Dufour D, Gallego J. Three-way RNA junctions with remote tertiary contacts: a recurrent and highly versatile fold. *RNA.* 2009; 15:1949–1964. [PubMed: 19741022]
18. Nakayama S, Luo Y, Zhou J, Dayie TK, Sintim HO. Nanomolar fluorescent detection of c-di-GMP using a modular aptamer strategy. *Chem Commun (Camb).* 2012; 48:9059–9061. [PubMed: 22854718]
19. Wittmann A, Suess B. Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. *Mol Biosyst.* 2011; 7:2419–2427. [PubMed: 21603688]
20. Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell.* 2003; 113:557–586.
21. Sudarsan N, et al. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science.* 2008; 321:411–413. [PubMed: 18635805]
22. Khvorova A, Lescoute A, Westhof E, Jayasena SD. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat Struct Biol.* 2003; 10:708–712. [PubMed: 12881719]
23. Majerfeld I, Puthenvedu D, Yarus M. RNA affinity for molecular L-histidine; genetic code origins. *J Mol Evol.* 2005; 61:226–235. [PubMed: 15999244]
24. Marshall KA, Ellington AD. In vitro selection of RNA aptamers. *Methods Enzymol.* 2000; 318:193–214. [PubMed: 10889989]
25. Illangasekare M, Turk R, Peterson GC, Lladser M, Yarus M. Chiral histidine selection by D-ribose RNA. *RNA.* 2010; 16:2370–2383. [PubMed: 20940341]
26. Wilkinson KA, Merino EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc.* 2006; 1:1610–1616. [PubMed: 17406453]
27. Joyce GF. Directed evolution of nucleic acid enzymes. *Annu Rev Biochem.* 2004; 73:791–836. [PubMed: 15189159]
28. Mohr S, et al. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA.* 2013; 19:958–970. [PubMed: 23697550]
29. Carothers JM, Goler JA, Kapoor Y, Lara L, Keasling JD. Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. *Nucleic Acids Res.* 2010; 38:2736–2747. [PubMed: 20159999]

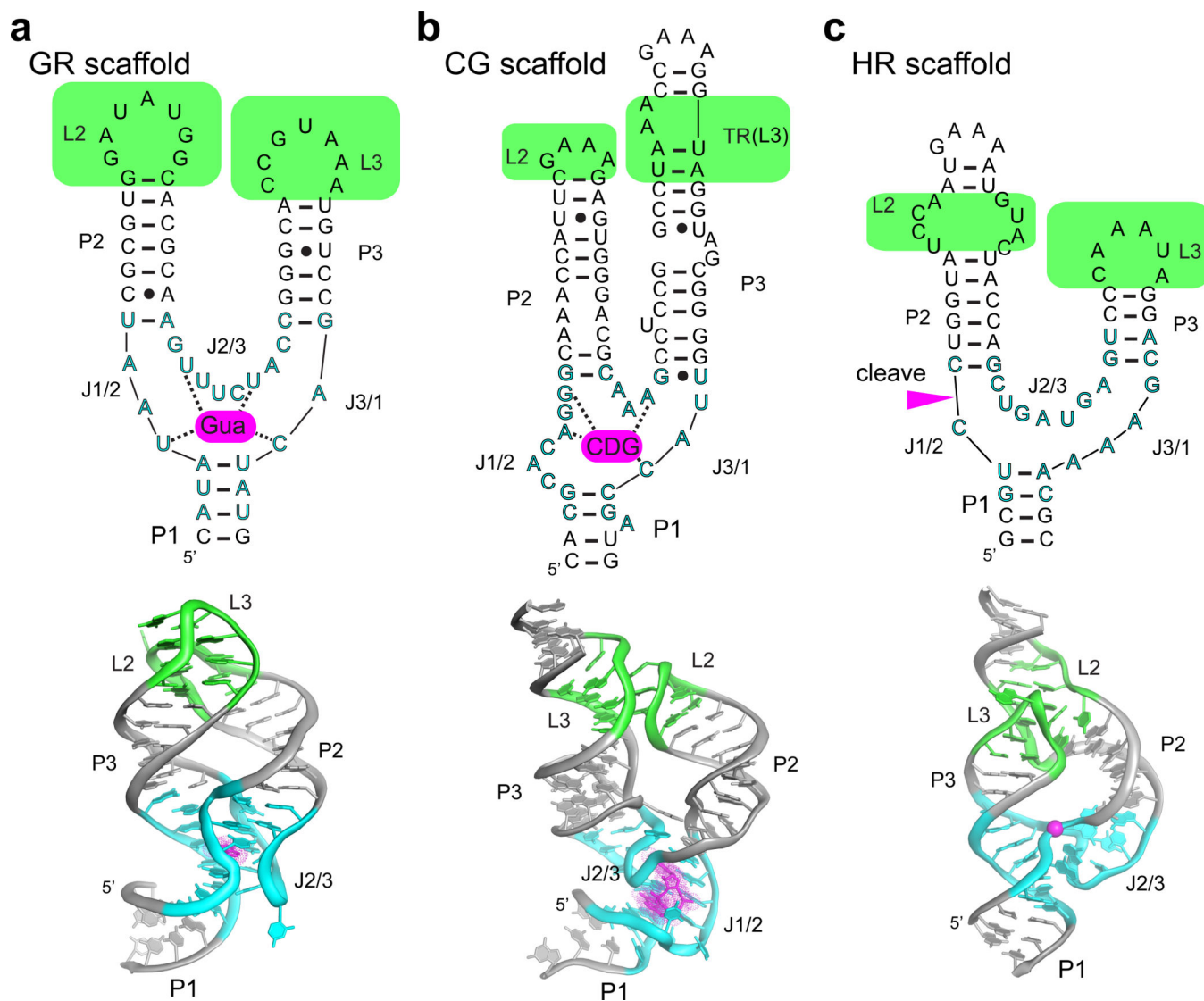
30. Ditzler MA, et al. High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase. *Nucleic Acids Research*. 2012; 41:1873–1884. [PubMed: 23241386]
31. Schutze T, et al. Probing the SELEX process with next-generation sequencing. *PLoS One*. 2011; 6:e29604. [PubMed: 22242135]
32. Majerfeld I, Yarus M. A diminutive and specific RNA binding site for L-tryptophan. *Nucleic Acids Res*. 2005; 33:5482–5493. [PubMed: 16186130]
33. Yao Z, Weinberg Z, Ruzzo WL. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics*. 2006; 22:445–452. [PubMed: 16357030]
34. Weinberg Z, Breaker RR. R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics*. 2011; 12:3. [PubMed: 21205310]
35. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc*. 2005; 127:4223–4231. [PubMed: 15783204]
36. Stoddard CD, Gilbert SD, Batey RT. Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA*. 2008; 14:675–684. [PubMed: 18268025]
37. Garst AD, Porter EB, Batey RT. Insights into the regulatory landscape of the lysine riboswitch. *J Mol Biol*. 2012; 423:17–33. [PubMed: 22771573]
38. Batey RT, Gilbert SD, Montange RK. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*. 2004; 432:411–415. [PubMed: 15549109]
39. Krasilnikov AS, Mondragon A. On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA*. 2003; 9:640–643. [PubMed: 12756321]
40. Edwards TE, Ferre-D'Amare AR. Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure*. 2006; 14:1459–1468. [PubMed: 16962976]
41. Serganov A, Polonskaia A, Phan AT, Breaker RR, Patel DJ. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature*. 2006; 441:1167–1171. [PubMed: 16728979]
42. Nagaswamy U, Fox GE. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA*. 2002; 8:1112–1119. [PubMed: 12358430]
43. Stojanovic MN, Kolpashchikov DM. Modular aptameric sensors. *J Am Chem Soc*. 2004; 126:9266–9270. [PubMed: 15281816]
44. Ponchon L, Dardel F. Recombinant RNA technology: the tRNA scaffold. *Nat Methods*. 2007; 4:571–576. [PubMed: 17558412]
45. Kellenberger CA, Chen C, Whiteley AT, Portnoy DA, Hammond MC. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic di-AMP. *J Am Chem Soc*. 2015; 137:6432–6435. [PubMed: 25965978]
46. Buck J, Noeske J, Wohnert J, Schwalbe H. Dissecting the influence of Mg<sup>2+</sup> on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res*. 2010; 38:4143–4153. [PubMed: 20200045]
47. Hodak JH, Downey CD, Fiore JL, Pardi A, Nesbitt DJ. Docking kinetics and equilibrium of a GAAA tetraloop-receptor motif probed by single molecule FRET. *Proc Natl Acad Sci U S A*. 2005; 102:10505–10510. [PubMed: 16024731]

## Online Methods References

48. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010; 7:335–336. [PubMed: 20383131]
49. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010; 26:2460–2461. [PubMed: 20709691]
50. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32:1792–1797. [PubMed: 15034147]

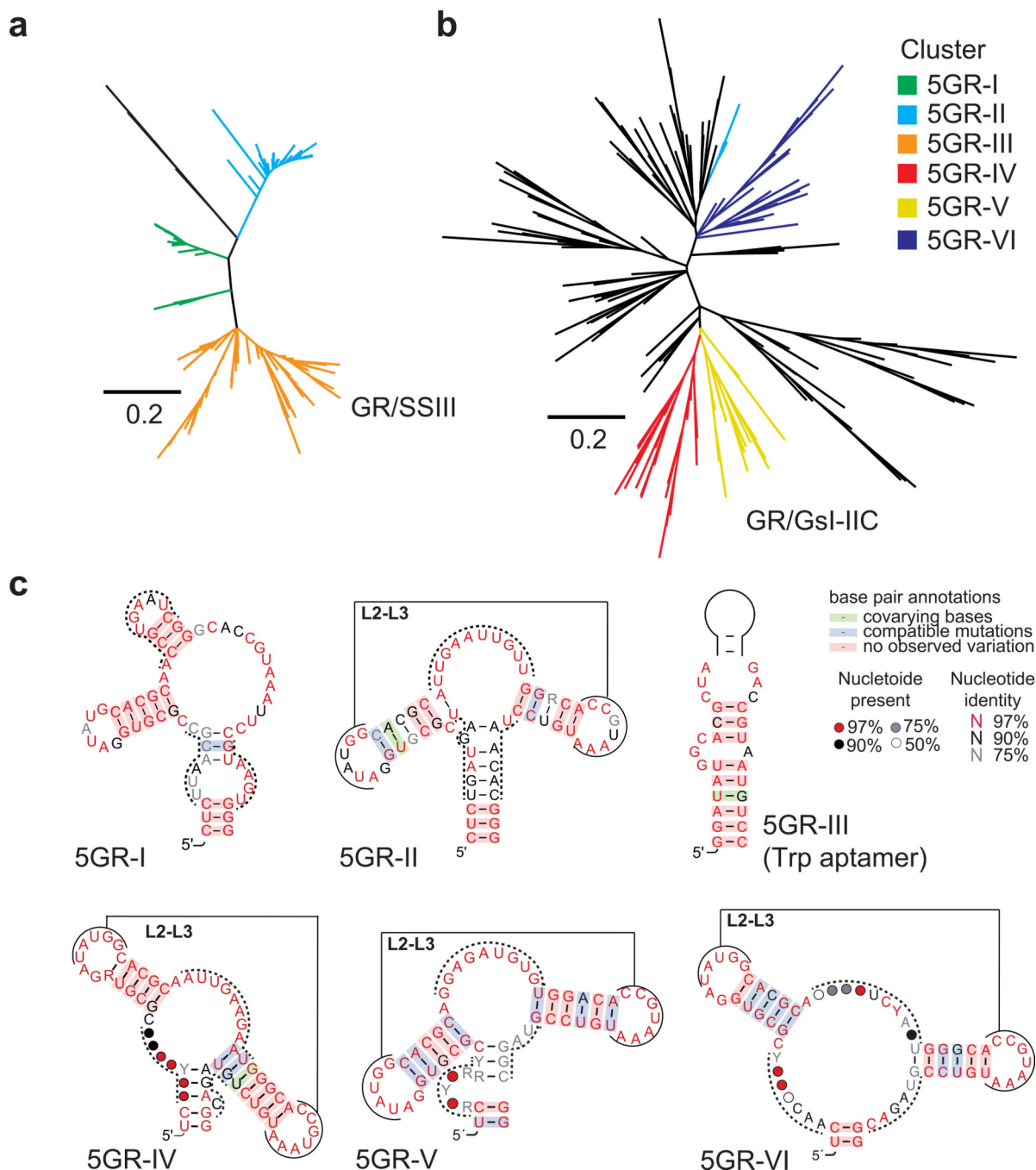


51. Edwards AL, Garst AD, Batey RT. Determining structures of RNA aptamers and riboswitches by X-ray crystallography. *Methods Mol Biol.* 2009; 535:135–163. [PubMed: 19377976]
52. Das R, Laederach A, Pearlman SM, Herschlag D, Altman RB. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA.* 2005; 11:344–354. [PubMed: 15701734]
53. Gilbert SD, Batey RT. Monitoring RNA-ligand interactions using isothermal titration calorimetry. *Methods Mol Biol.* 2009; 540:97–114. [PubMed: 19381555]
54. Pflugrath JW. The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr.* 1999; 55:1718–1725. [PubMed: 10531521]
55. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:213–221. [PubMed: 20124702]
56. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–2132. [PubMed: 15572765]
57. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:12–21. [PubMed: 20057044]
58. Paige JS, Nguyen-Duc T, Song W, Jaffrey SR. Fluorescence imaging of cellular metabolites with RNA. *Science.* 2012; 335:1194–1194. [PubMed: 22403384]
59. Schindelin J, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012; 9:676–682. [PubMed: 22743772]



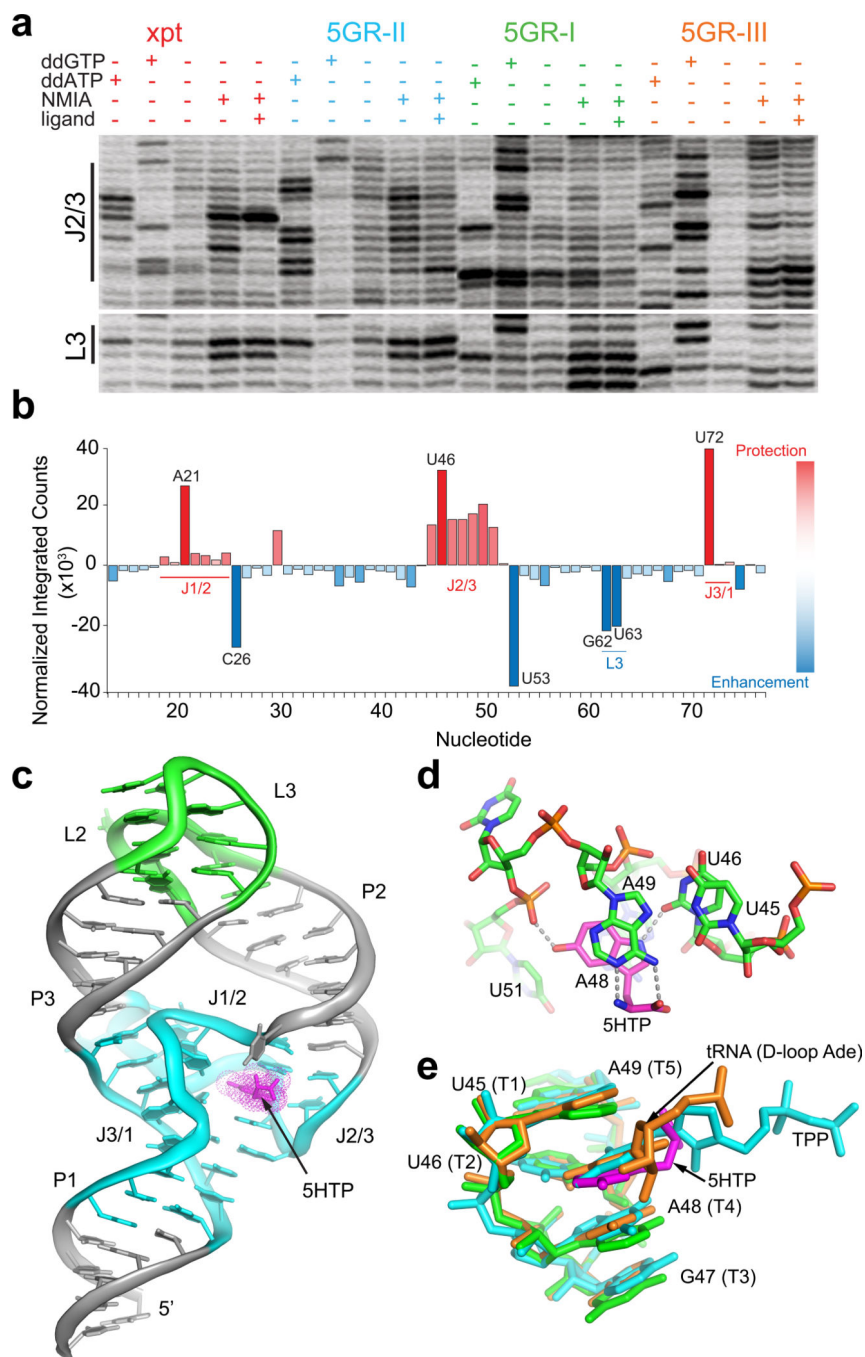
**Figure 1. RNA scaffolds used for selection**

(a) The GR scaffold is derived from the aptamer domain of the *B. subtilis xpt-pbuX* guanine riboswitch. The aptamer is comprised of three paired (P) regions connected by the joining (J) regions of the 3WJ that contains the guanine (Gua, magenta) binding site (dashed lines represent direct RNA-ligand interactions). Nucleotides outlined in cyan are those that were randomized for selection. The terminal loops of P2 and P3 (L2 and L3, green box) participate in a tertiary interaction that organizes the 3WJ. Below is the three dimensional structure of the RNA (PDB ID 4FE5) with the same coloring scheme emphasizing the spatial relationship between the ligand binding site and the randomized nucleotides. (b) The secondary (top) and tertiary structure (bottom) of the CG scaffold derived from the aptamer domain *V. cholera Vc2* cyclic *d*-*GMP* riboswitch (PDB ID 3IWN). (c) The secondary (top) and tertiary structure (bottom) of the HR scaffold derived from the *S. mansoni* *hammerhead* ribozyme (PDB ID 3ZP8). The labeling and coloring scheme in panels (b) and (c) are the same as in (a).



**Figure 2. Selection of scaffolded aptamers that selectively bind 5-hydroxytryptophan (5HTP)**  
 (a) Unrooted phylogenetic tree representation of the distance matrix of sequences derived from round 7 of the GR/SSIII selection. Sequences are grouped into three main clusters, which are independently colored. The distance is expressed as the maximum likelihood estimate (MLE) of the number of substitutions occurring per site between two nodes of the tree (bar shown for scale). (b) Unrooted phylogenetic tree representation of the distance matrix of sequences derived from round 7 of the GR/GsI-IIC selection. The four clusters from which representative sequences were analyzed are shown in independent colors (legend

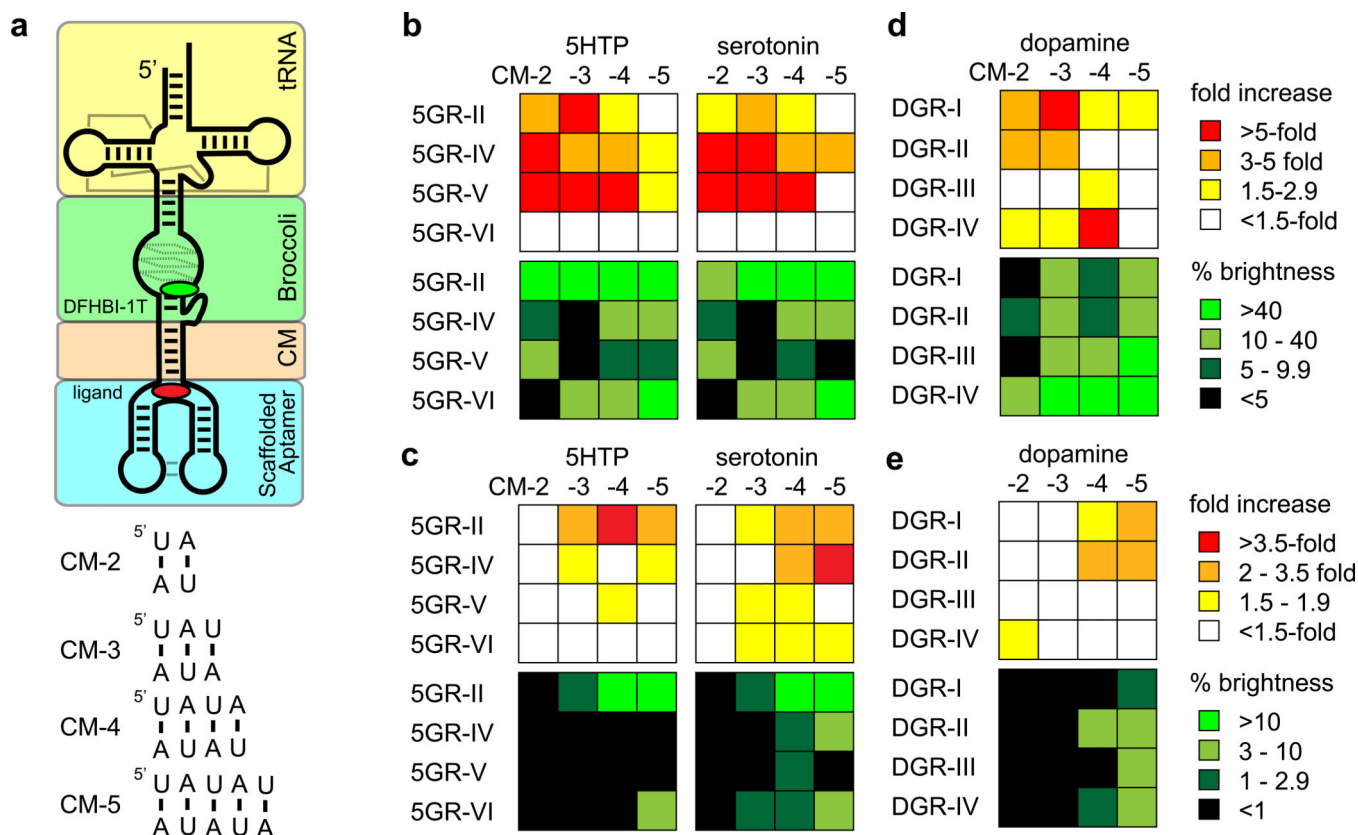
shown to right); black region represents unanalyzed clusters and regions of the tree. (e) Covariation model of six observed clusters derived from the GR selections. The dashed lines correspond to the regions of the scaffold that were randomized and the line connecting L2 and L3 denote clusters where the sequences were maintained that would support the tertiary interaction.



**Figure 3. Analysis of potential 5HTP aptamers**

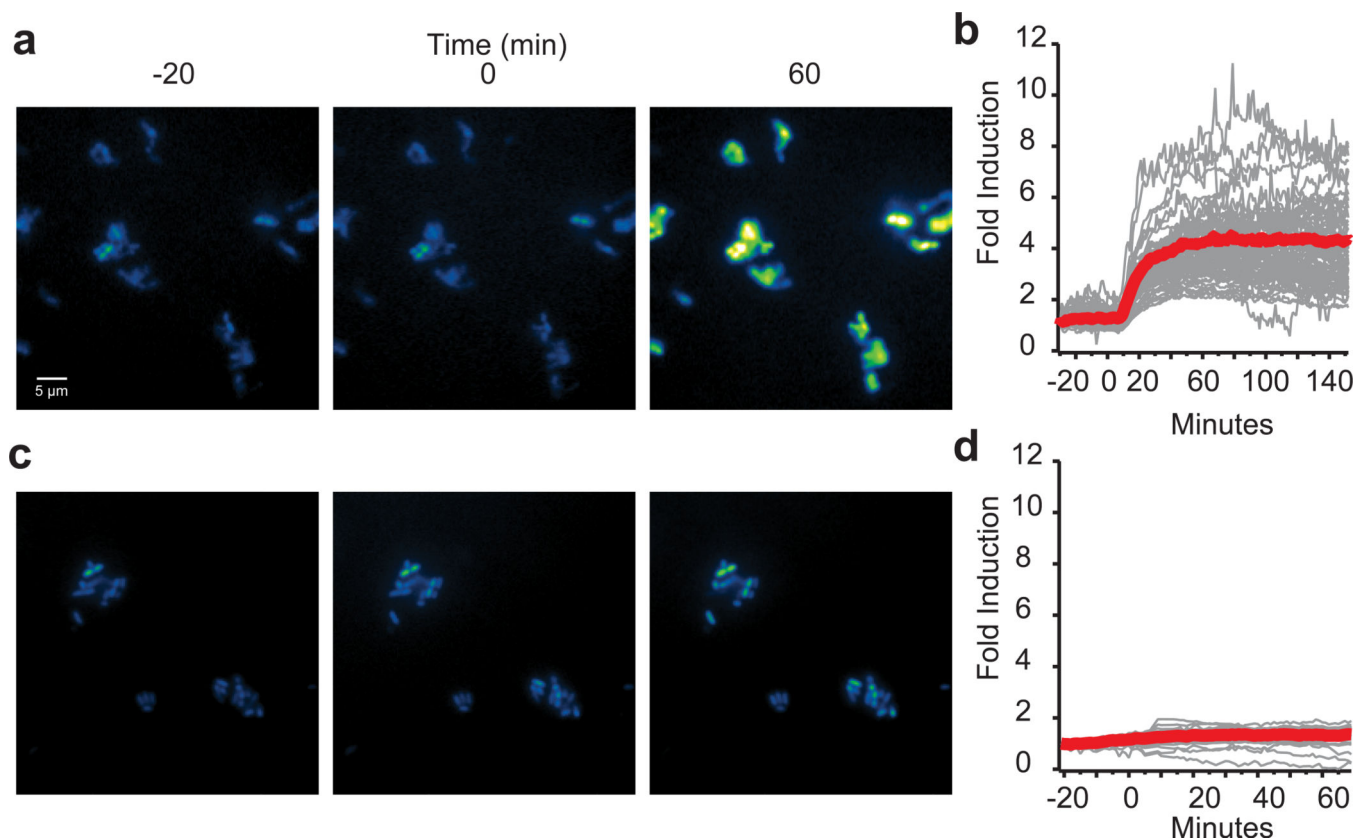
(a) "SHAPE" chemical probing of three sequences (5GR-I, -II, and -III) in comparison to the native *xpt* guanine riboswitch. (full gel shown in Supplementary Fig. 5). (b) Quantitation of the ligand-dependent differential intensities of SHAPE probing of 5GR-II in the presence and absence of 5HTP demonstrates that reactivity localizes to the junctions and an enhancement in L3 suggests a reinforced tertiary structure upon ligand binding. (c) Crystal structure of the 5GR-II aptamer in complex with 5HTP (magenta) with sequence elements colored as in Figure 1a. (d) The 5GR-II binding pocket within the 3WJ forms a set of

hydrogen bonding interactions that engages every polar functional group in 5HTP with the exception of an  $\alpha$ -oxygen which was presented as column-coupled amide during the selection. The complex is also stabilized by stacking interactions between the hydroxyindole ring of 5HTP and adenine bases (A48 and A49) in J2/3. (e) The core of the binding pocket of the 5GR-II aptamer (green) is a T-loop that superimposes almost perfectly with the T-loops from tRNA<sup>Phe</sup> (orange) and the thiamine pyrophosphate (TPP) riboswitch (cyan). In each example, the space between purines T4 and T5 enables intercalation of an aromatic ring (T1–5 numbering indicates the nucleotide position within the T-loop motif).



**Figure 4. In vitro and intracellular performance of 5HTP and L-DOPA sensors**

(a) Schematic of the secondary structure of genetically encodable biosensors of 5HTP and L-DOPA in which a GR-scaffolded aptamer (cyan) is coupled to a fluorogenic aptamer (“Broccoli”, green) via a communication module (orange, CM; sequences on bottom) and stabilized *in vivo* with the tRNA scaffold (yellow). (b, c) Heat maps depicting the fold change in fluorescence observed upon the addition of ligand (top) and the prospective sensor’s percent brightness of a tRNA/Broccoli control in the presence of ligand (bottom) for a series of GR-scaffolded aptamers coupled to Broccoli with CMs from 2 to 5 base pairs in length. (d, e) Heat maps of the performance of the same sensors in *E. coli*.



**Figure 5. A 5HTP aptamer-based biosensor functions in *E. coli***

(a, b) The wild-type 5GR-II aptamer specifically activates the fluorescence of the Broccoli reporter in the presence of 5HTP, but not in the presence of L-tryptophan. (c) A single point mutation in the 5HTP-binding pocket of the 5GR-II aptamer (A48U) also ablates fluorescence in the presence of the fluorophore. (d–f) Single cell traces of fluorescence induction for the wild type 5GR-II-Broccoli sensor in the presence of 5HTP, L-tryptophan, and the binding incompetent 5GR-II A48U construct in the presence of 5HTP. At t=0 minutes, either 2 mM 5HTP or 5 mM L-tryptophan was added to the media.



Table 1

Affinity of aptamers for 5HTP and related compounds

Selection (scaffold/RT)	aptamer	5HTP $K_D$ , $\mu\text{M}^a$	L-Trp $K_D$ , $\mu\text{M}$	Serotonin $K_D$ , $\mu\text{M}$	Me-5HTP $K_D$ , $\mu\text{M}$
GR/SSIII	5GR-I	33 $\pm$ 1	41 $\pm$ 1	>1000	----
	5GR-II	3.9 $\pm$ 0.1	280 $\pm$ 30	38 $\pm$ 8	26 $\pm$ 9
	5GR-III	38 $\pm$ 14	20 $\pm$ 5	>1000	----
GR/Gsl-IIC	5GR-IV	8.8 $\pm$ 1.5	520 $\pm$ 90	4.7 $\pm$ 0.3	1.3 $\pm$ 0.1
	5GR-V	11 $\pm$ 1	170 $\pm$ 10	16 $\pm$ 4	6.6 $\pm$ 0.4
	5GR-VI	60 $\pm$ 15	N.D. <sup>b</sup>	16 $\pm$ 2	25 $\pm$ 8
CG/Gsl-IIC	5CG-I	9.3 $\pm$ 0.3	N.D.	1.2 $\pm$ 0.1	2.1 $\pm$ 0.4
HR/Gsl-IIC	5HR-I	7.3 $\pm$ 2.8	N.D.	1.2 $\pm$ 0.2	2.5 $\pm$ 0.5

<sup>a</sup> All measurements take at 25 °C in 10 mM MgCl<sub>2</sub> containing buffer<sup>b</sup> not detectable