Reengineering orthogonally selective riboswitches

Neil Dixon,1 John N. Duncan,2 Torsten Geerlings,2 Mark S. Dunstan,3 John E. G. McCarthy,2 David Leys4, and Jason Micklefield1

1School of Chemistry and 2Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom.

Edited by Dinshaw J. Patel, Memorial Sloan-Kettering Cancer Center, New York, NY, and approved December 22, 2009 (received for review September 29, 2009)

The ability to independently control the expression of multiple genes by addition of distinct small-molecule modulators has many applications from synthetic biology, functional genomics, pharmaceutical target validation, through to gene therapy. Riboswitches are relatively simple, small-molecule–dependent, protein-free, mRNA genetic switches that are attractive targets for reengineering in this context. Using a combination of chemical genetics and genetic selection, we have developed riboswitches that are selective for synthetic “nonnatural” small molecules and no longer respond to the natural intracellular ligands. The orthogonal selectivity of the riboswitches is also demonstrated in vitro using isothermal titration calorimetry and x-ray crystallography. The riboswitches allow highly responsive, dose-dependent, orthogonally selective, and dynamic control of gene expression in vivo. It is possible that this approach may be further developed to reengineer other natural riboswitches for application as small-molecule responsive genetic switches in both prokaryotes and eukaryotes.

Results & Discussion

Selection of Mutant Riboswitches that Accept New Heterocyclic Ligands. The add A-riboswitch was chosen as the basis for this study because it functions at the translational level via a relatively simple mechanism that could potentially operate in any bacterial system (18, 19) (Fig. 1). Also, inexpensive or easily synthesised purine and other heterocyclic analogues, which possess desirable physicochemical properties, could potentially be used as ligands (20, 21). In previous studies, a number of other well-characterized purine riboswitches have also been identified. However, these function through modulation of transcription, by a more complex mechanism (22–25), which may be more organism-specific (26).

Here, we have applied a combination of chemical genetics and genetic selection to identify new riboswitches that allow orthogonal, selective, tuneable, and dose-dependent control of gene expression in response to nonnatural synthetic small molecules. Our strategy involved mutagenesis of the aptamer domain of the adenine-sensing add A-riboswitch, found in the 5′-UTR of the adenine deaminase encoding gene of Vibrio vulnificus (4).
riboswitch M6 (U47C, U51C), was shown to give an optimal response to ammeline 3 (Fig. 1C). The selected M6 riboswitch regulated eGFP expression dose-dependently over a broad dynamic range upon addition of ammeline 3 (Fig. 2A). The degree to which the M6 riboswitch controls gene expression, referred to as the induction factor, was determined from the ratio of the maximum ligand-induced protein expression over the basal protein expression level. For example, the mutant M6 exhibits a maximal induction factor of circa (ca.) 6.5 at a 500 μM dose of ammeline 3 (Fig. 2A and B). This compares with an induction factor of ca. 11 observed for the parental riboswitch upon addition of adenine 1 (500 μM). Amino acid 3 was further shown to have no effect upon eGFP expression under control of the parental riboswitch, and most significantly adenine 1 had no effect upon eGFP expression under control of the mutant riboswitch M6 (Fig. 2A). These observations indicate that orthogonal control of gene expression is possible, simply by introducing two point mutations into the parent riboswitch aptamer domain. These mutations effectively block recognition of the natural ligand but allow the binding of the selected nonnatural ammeline ligand 3, which in turn has no effect upon the parent riboswitch. These results are consistent with the earlier structural studies which show that nucleotides U47 and U51 confer selectivity within the purine riboswitches, by recognition of the N9-N3-N2 face of the purine ligand (Fig. 1B).

Given that the double mutant M6 (U47C, U51C) retains its functionality in vivo, we sought to explore further the structure-activity relationship by mutating U74 to cytosine to give the triple mutant M6C (U47C, U51C, U74C). Through Watson-Crick base pairing, U74 confers adenine versus guanine selectivity in the natural purine riboswitches (4, 25, 28) (Fig. 1). Despite this, the M6 and M6C mutants were surprisingly found to have partially overlapping ligand-response profiles, with the ammeline 3 also inducing eGFP expression with the M6C mutant albeit at a reduced induction factor of ca. 3.6 (Fig. 2C). The library of ca. 80 ligands (Fig. S2) was then screened against M6C-eGFP, and azacytosine 4 was identified, which displayed a similar induction factor of ca. 2.8 (Fig. 2C). Moreover, neither adenine 1 or guanine 5 induce eGFP expression over the basal level, under the control of the M6C riboswitch. Also, azacytosine 4, like ammeline 3, has no effect on eGFP expression under control of the parental add A-riboswitch, demonstrating that the M6-3 and M6C-4 pairs both show orthogonal selectivities with respect to the parental riboswitch.

**Mutations Within the P2 Stem Increase Absolute Reporter Gene Expression Levels.** Significant induction factors were observed for M6 with ammeline 3 (of ca. 6.5), and for M6C with ammeline 3 (of ca. 3.6) and azacytosine 4 (of ca. 2.8). However, the maximum level of fluorescent protein produced per cell, as determined by comparing the normalized fluorescent units (fu/OD620) at saturation (500 μM ligand), were 35–70% lower for the mutants in comparison to the parental switch (Fig. 2B and C). For example, riboswitches M6 and M6C both became saturated in the presence of ammeline 3 (500 μM) leading to maximum eGFP levels of ca. 3.4 × 106 and 6.7 × 106 fu/OD620, respectively, whereas the parental riboswitch in the presence of adenine (500 μM) reached a maximum eGFP level of ca. 10.4 × 106 fu/OD620. In order to improve the dynamic range of expression levels that can be accessed using these new riboswitch control elements, we first sought to subject M6 to further mutagenesis. Computational analysis using the mFold server (29) suggested that the mutations (U47C, U51C) introduced into the J2/3 loop of mutant M6 can potentially hybridize to G42 of a G-U wobble within the P2 stem (Fig. 1). To examine this misfolding hypothesis, mutants M6′ (G42A, U47C, U51C) and M6′′ (U28G, G42C, U47C, U51C) were prepared (SI Text). These mutants avoid possible misfolding caused by potential additional interactions between G42 and C51 and also strengthen the P2 stem through increased base pairing between nucleotides 28 and 42. The P2 stem mutations resulted in no significant change in the induction factor of 6.5 for the mutants M6, M6′, and M6′′ with ammeline 3 (Fig. 2B), as both the basal and ligand-induced levels of eGFP expression increased proportionally. Indeed, as the strength of nucleobase contacts between positions 28 and 42 increased (from U-G, U-A to G-C), so did the absolute maximal levels of eGFP expression (Fig. 2B). We rationalize these observations using a kinetic partitioning mechanism (30, 31), whereby increasing P2 stem strength increases the population of correctly folded riboswitches available to perform gene regulation, by minimizing the occurrence of the misfolded mRNA transcripts. As a consequence of increasing the P2 stem strength within the M6′ and M6′′
mutants, there also appears to be a change in the relative order of helix stability (Table S1). It has also been postulated that the formation of a correctly folded P2 stem inhibits the formation of alternative incorrectly folded structures for the phuE A-riboswitch (32).

The ability of the mutant M6′′ to control eGFP expression over its increased dynamic range was further tested by varying the concentration of ammeline 3 from 0.5–500 μM. From this it can be seen that M6′′ permits excellent dose-dependent control of eGFP expression down to 5.0 μM (Fig. 2D). Additionally, the excellent orthogonality with respect to adenine, observed for the M6 mutant, was conserved within both the M6′ and M6′′ riboswitches. To corroborate these observations, the same mutations were introduced into the P2 stem of the mutant M6C riboswitch to give mutants M6C′′ (G42A, U47C, U51C, U74C) and M6C′′ (U28G, G42C, U47C, U51C, U74C) (SI Text) (Fig. 1A). This resulted in no change to the observed induction factors of ca. 3 for M6C, M6C′, and M6C′′ with both ammeline 3 and azacytosine 4 (Fig. 2C). However, as predicted the absolute maximum eGFP production levels increased significantly with an increase in P2 stem strength. Indeed, eGFP production levels for M6C′′ with either ammeline 3 (15.5 × 10^6 fu/OD) or azacytosine 4 (14.1 × 10^6 fu/OD) at 500 μM exceed the maximum expression limit of 10.4 × 10^6 fu/OD for the parental riboswitch in the presence of adenine at 500 μM (Fig. 2B and C). It was also observed that the M6C′′ displays excellent orthogonal selectivity toward both ammeline 3 and azacytosine 4, with adenine 1 and guanine 5 again exhibiting little or no induction of eGFP expression (Fig. S3). However, it was noted that guanosine (rG) and 2′-deoxyguanosine (dG) do elicit 1.9- and 2.5-fold induction of eGFP expression, respectively (Fig. S4). This may be due to some similarity between the aptamer domains of M6C′′ and the natural dG-responsive riboswitch from Mesoplasma florum (33). These findings above demonstrate that rational modifications of base pairs peripheral to the ligand binding site, which alter the P2 stem strength, allow the expression landscape to be tuned, resulting in riboswitches that induce low (M6 and M6C), medium (M6′ and M6C′), or high (M6′′ and M6C′′) protein production in response to various nonnatural ligands (Fig. 2B and C).

**New Mutant Riboswitch Pairings Display In Vitro Binding Orthogonality.** To probe the ligand–RNA interactions, we produced the aptamer domains (13–83 nt) of the mutant and parental riboswitches by in vitro transcription and studied ligand binding with isothermal titration calorimetry (ITC). In line with the expression results, orthogonal binding selectivity was observed for the aptamer domain of the M6′′ riboswitch, which binds ammeline 3 with an equilibrium dissociation constant (K_d) of 1.19 μM (Fig. 3A and E), whereas no evidence of binding was observed with the M6′ aptamer and adenine 1 (Fig. 3B). The parental add A-aptamer domain bound to adenine 1 with an affinity similar to previous reports (21) and showed no observable binding to add a-aptamer domain bound to adenine 1 with an affinity similar to previous reports (21) and showed no observable binding to

---

**Fig. 2.** (A) Ligand-dependent eGFP expression versus time under control of parental and M6 riboswitches: Parental riboswitch (black) on addition of adenine 1 (500 μM (black circle) and ammeline 3 (500 μM (black diamond)), M6 riboswitches (green) in the presence of either ammeline 3 (500 μM (green triangle)) and adenine 1 (500 μM (green circle)). The eGFP fluorescence signal is normalized for cell density (fu/OD_{620}) and displayed as an induction factor, which is determined by taking the ratio of normalized eGFP signal in the presence and absence of ligand. (B) eGFP induction factors (left axis indicated by the gray bars) and absolute normalized eGFP fluorescence signal (right axis) in the absence of ligand (open shape) and in the presence of ligand (closed shape) for: parental add A-riboswitch in the presence of 500 μM adenine (black circle) and absence of adenine (open circle); M6, M6′, and M6′′ ribsoswitches in the presence of 500 μM ammeline 3 (black diamond), or absence of ammeline 3 (open diamond). (C) eGFP induction factors and absolute eGFP expression for M6C, M6C′, and M6C′′ in the presence of 500 μM ammeline 3 (black diamond), or absence of ammeline 3 (open diamond); in the presence of 500 μM azacytosine 4 (black square), or absence of azacytosine 4 (open square). (D) Dose-dependent eGFP expression under control of mutant M6′′ (green square) and parental add (black square) riboswitches showing induction factors over a range of ammeline 3 or adenine 1 ligand concentrations.
ammeline 3 (Fig. 3C and D). These results can be rationalized through the binding of ammeline 3, in its keto tautomeric form, which is able to make six hydrogen bonding contacts with the M6′ aptamer (Fig. 3F) through Watson–Crick-like pairing with U74 and three further constructive hydrogen bond contacts to C51. This mode of binding would be consistent with the binding models proposed for purine riboswitches (18, 21, 23, 24, 34). ITC measurements (Fig. 3E) also reveal that ammeline 3 and azacytosine 4 bind to the M6C′′ aptamer domain with $K_d$ values of 1.02 and 1.00 μM, respectively.

Although the ITC results (Fig. 3) are consistent with the observed in vivo orthogonal selectivity, it is apparent that the mutant riboswitches have lower affinities ($K_d$ 1–1.2 μM range) for the selected ligands than the wild-type purine riboswitches, which exhibit $K_d$ values in the low-to-mid-nM range with their cognate purine ligands (21). There is thus room for improving the affinities and potentially the in vivo induction factors, either by more extensive mutagenesis remote from the ligand binding site or by rational design of new ligands based upon the azacytosine 4 structural framework. In agreement with previous studies, the binding of adenine 1 and diaminopurine 2 to the parental riboswitch aptamer domain is highly entropically unfavorable and is driven entirely by enthalpy (21) (Table S2). In contrast, the binding of ammeline 3 to mutant riboswitches M6′′ and M6C′′ aptamer domains is both enthalpically and entropically favorable, whereas binding of azacytosine 4 to M6C′′ aptamer domain is largely entropically driven (Table S2).

Crystal Structure of the M6C′′ Aptamer in Complex with 4. Given the structural similarity between ammeline 3 and azacytosine 4, it is perhaps not surprising that both exhibit similar binding affinity for the M6C′′ aptamer domain. However, unlike the proposed binding mode of ammeline 3 with M6′′ (Fig. 3F), the interaction of ligands 3 and 4 with M6C′′ would require a significant change in structure of the ligand binding pocket within the aptamer domain to avoid unfavorable clashes. To explore this, the M6C′′ aptamer (13–83 nt) complexed to azacytosine 4 was crystallized, under conditions similar to those used for the wild-type add...
A-aptamer-adenine complex (18, SI Text), and the structure was solved to 1.7 Å (Fig. 4A) using the molecular replacement method with the add A structure (1Y26) (rmsd 1.1 Å) (Table S3). Azacytosine 4 could be clearly identified in the electron density and binds through a Watson–Crick-like base pairing to C74. Further H-bonding is observed to C51 between N5(4) and N4(C51) as well as between N4(4) and N3(C51) (Figs. 4B and 4F). An additional H-bond is also made between the 2′OH of U22 and N1 of ligand 4 (Fig. S5). To accommodate these key interactions, several changes in the architecture of the ligand binding site can be observed with respect to the add A structure. Firstly, azacytosine 4 appears to be shifted with respect to adenine 1, causing the C74 to shift toward the major groove (Fig. 4C). Additionally, the backbone of the J2/3 loop appears to allow lateral movement of C51 away from C47, by 1.8 Å, toward the central core of the ligand binding site. Lateral movement of C51 has previously been observed within a U51C mutant riboswitch that recognises 2′-deoxyguanosine (33). Unlike the previous studies, however, the nucleotide at position 47 in the present structure is not dramatically perturbed and does not point away into the solvent phase (Fig. 4C). Finally, coordination to Mg2+ is also shown to induce a kink in the phosphodiester backbone between U22 and A23, enabling the 2′OH of U22 to engage with N1 lone pair of 4 (Fig. S5).

Conclusions

Using a combination of chemical genetics and genetic selection a riboswitch double mutant (M6) was selected from a total pool of close to 1200 riboswitch–ligand pairings. Notably, this mutant and the related triple mutant (M6C) control in vivo gene expression in response to nonnatural ligands ammeline 3 or azacytosine 4, respectively, in a precise dose-dependent fashion. These riboswitches display excellent orthogonal selectivity with respect to the parent riboswitch. Thus, M6 and M6C do not respond to adenine or guanine, nor does the parent riboswitch interact with ammeline 3 or azacytosine 4. Furthermore, by introducing mutations into the P2 stem, remote from the ligand binding site, it was shown to be possible to fine-tune the accessible gene expression landscape, providing riboswitches (M6C′ and M6C′′) that cover a broad dynamic range of gene expression levels. ITC binding studies were also used to confirm the orthogonal selectivity of the mutant and parent riboswitches in vitro. Finally, the x-ray crystal structure of the orthogonal riboswitch M6C′′ demonstrates the inherent flexibility of the ligand binding site.

Taken together the results presented here form the basis for an approach to reengineer natural riboswitches, providing new genetic switches that allow dose-dependent, orthogonal selection, and dynamic control of gene expression and further exemplify the functional and structural versatility of RNA. Reengineered riboswitches are potentially useful orthogonal components for synthetic biology and could also find application for gene functional analysis in vivo, where simultaneous differential control of multiple genes would be desirable.

Materials and Methods

Construct Design, Selection, and Reporter Assay. A synthetic construct containing the lac promoter/operator upstream of the add A-riboswitch from Vibrio vulnificus (4, 18), optimized to give an improved induction factor within E. coli, was purchased (Yorkshire Bioscience) and subcloned with EcoRI and HindIII into the pMOD3 vector. The CAT selection gene and eGFP reporter gene were PCR amplified, digested using EcoRV and NotI, and ligated into the pMOD3 vector, which had been digested with NcoI, Klenow filled, then digested with NotI to give CAT and eGFP parental vectors (SI Text Fig. S1).

Mutants M1–15 were introduced into the CAT vector template, using the multi-site-directed mutagenesis kit (Stratagene). The resulting plasmids were used to transform E. coli, and single colonies were picked and checked by sequencing. Mutants were initially screened on LB agar plates containing 170 μg/ml chloramphenicol, 1 mM IPTG, and 250 μM ligand from the library of 80 ligands (Fig. S2). Colonies that survived in the presence of a particular

![Fig. 4. Crystal structure of mutant M6C′ bound to ligand 4. (A) M6C′ (Blue) overlayed with the add A (Red). Mutated residues are shown in green. (B) Azacytosine 4 bound to the M6C′ displayed with mFo–DFc sigma a-weighted omit map, contoured at 3.5σ. (C) Overlay of ligand binding sites of add A (Red), and M6C′ (Blue).](https://www.pnas.org/content/116/50/11545/F4)

Dixon et al.
lignant, were selected for further characterization. Growth assays were conducted using cells grown overnight in LB at 37 °C, which were diluted 1/20 into fresh M9 medium containing 1 mM IPTG, and then grown at 37 °C for 1 h. 170 μg/ml Chloramphenicol and 500 μM of the selected ligand were added, and bacterial cell growth was monitored by UV absorbance at 620 nm using an Anthos Zenyth 3100 plate reader.

The mutants identified from the CAT selection were cloned into the eGFP parental vector (using BamHI and BsiI restriction sites) and sequenced. The U74C mutation was introduced into the M6-eGFP expression vector via site-directed mutagenesis to afford the M6c-eGFP construct. The P2 stem mutants, G42A (′), and U28G, G42C (′), were again produced by site-directed mutagenesis and checked by sequencing. Expression assays were conducted using cells grown overnight in LB at 37 °C, which were diluted 1/20 into fresh M9 medium with 1 mM IPTG and grown at 37 °C for 1 h. The selected ligand (5 mM DMSO stock) was added, and fluorescence intensity units (using 485/535 filters) normalized for cell density (620nm) were measured using the Anthos Zenyth 3100 plate reader.

**In Vitro RNA Production, ITC, and X-Ray Crystallography.** In vitro transcribed RNA (aptamer domains 13–83 nt) of identical length to that used in previous studies (18, 21), were prepared from double-strand DNA templates, using standard procedures (SI Text). The transcripts were then purified under native conditions by size exclusion chromatography, and assessed for purity under denaturing conditions (8.0 M Urea, 12% polyacrylamide gel).

RNA samples for ITC, were exhaustively dialyzed overnight at 4 °C against 50 mM K HEPES pH 7.5, 20 mM MgCl₂, 100 mM KCl and deaged. ITC was performed with the concentration of the injectant 10 times higher than the cell concentration. Experiments were performed at 25 °C using a VP-ITC microcalorimeter (Microcal, Inc) with 20 injections of 15.0 μL with a reference power of 2.0 kcal/s.

For crystallography, the purified M6C′-4 complex was prepared at 0.52 mM in 50 mM K HEPES buffer pH 7.5, 5.0 mM MgCl₂, 100 mM KCl and recrystallized under conditions similar to those previously described (18). Diffraction data (Table S3) were collected at DLS, Oxfordshire. Coordinate and structure factors have been deposited into the Protein Data Bank (PDB access code 3LAS).

**ACKNOWLEDGMENTS.** This work was supported by the Biotechnology and Biological Sciences Research Council through the Selective Chemical Intervention in Biological Systems Initiative (Grant BB/D005612/1) and a PhD studentship award for J.N.D. We also thank Sri A. Lestari for technical support and Dr. Graeme L. Conn for helpful advice.