SYNTHETIC REGULATION AND GENETIC CONTROL OF ECOLOGICAL STRATEGY

Thesis by

Travis Scott Bayer

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2009

(Defended 19 December 2007)

© 2009

Travis Scott Bayer

All Rights Reserved

ACKNOWLEDGEMENTS

It is quite rare to have the good fortune to be surrounded by so many excellent people.

First, I would like to thank Christina Smolke for advising, employing, tolerating, and encouraging me over the past years. I appreciate the environment that she has created, which allowed me to spend hours at the whiteboard (and hoods, and windows) drawing things that would probably never work, coming up with insanely ambitious plans, and giving me enough rope to hang myself with (which happened often).

I am grateful for the guidance and good science my committee (Frances Arnold, Paul Sternberg, and Michael Elowitz) has taught me, whether by gentle persuasion, critical reading, or challenging questions. One of the first experiences I had at Caltech was Frances encouraging me to come here, saying "this is a place to get science done." Caltech has not disappointed.

It is always a memorable experience starting in a lab, and I believe more so when the lab itself is starting. I would like to thank the "original crew" – Stephanie Culler, Kristy Hawkins, Maung Win, and Andrew Babiskin for allowing a biologist into the lab, for racking tips by hand, and for teaching each other how to clone. We have all come a long way. Thanks to all the members of the Smolke group for making the lab a productive, functioning unit.

I have been incredibly fortunate to remain friend and colleague to the Ellington group at UT-Austin. Nine years ago, Andy Ellington convinced me to abandon the philosophy degree I was working on and take up science, and has remained an advisor and mentor ever since. Matt Levy and Amos Yan are the types of scientists and people I aspire to be: kind, conscientious, knowledgeable, technical, visionary.

I would like to acknowledge the support and wisdom I have received from everyone in our field. Drew Endy, Chris Voigt, Zach Simpson, and Wendell Lim have helped me develop ideas, test hypothesis, troubleshoot projects, and think big. Chase Beisel taught me that models demand one's respect, but reward amply; and that you can watch the World Cup and talk science at the same time. Steph Culler and Kristy Hawkins teach by example: they excel at everything they do, and still produce great work.

Jeff Tabor and Kevin Hoff have always been my second and third advisors, at least in spirit. I am grateful for the countless coffee breaks and happy hours I have spent talking science, writing, complaining, and thinking with them. Jeff showed me that synthetic biology can and should be fun, and that we should take a hacker approach to it. Kevin has pushed me as much as anyone to do good work, always think big, and always be the toughest critic of your own work.

I would not be writing this thesis today if not for my (extended) family: Scott, Linda, Tyler, Andrea, Robert, Lynn, and Erika. They have listened, advised, supported, cajoled, and convinced me when to stick with things and when to take breaks. It is all too easy to

lose perspective of things at Caltech, and all of them have helped me stay focused on the fact that having a full life should be balanced with having a full lab notebook.

Finally, this work is as much Laura's as it is mine – she knows about every failed experiment, every late night, and every frustration that I went through. It is difficult to predict whether investments of time and effort will payoff in the end, which is especially true in science. I am certain that every day I have spent with her has been worthwhile.

ABSTRACT

The construction of synthetic gene regulatory circuits inside living cells has illuminated how organisms process environmental signals, and has suggested that biological systems can be engineered for useful purposes. However, these lines of inquiry are limited by a lack of technologies for programming gene expression and an understanding of the adaptive or ecological consequences of manipulating gene expression. Here, I describe the design of noncoding RNA regulators of gene expression in *Saccharomyces cerevisiae*. These regulators are able to regulate gene expression in response to a small molecule ligand, which offers the ability to tailor control devices for a variety of applications. In light of this, an open question is the dependence of organism fitness on the levels of a regulator, which has seldom been measured. I found that the expression level of a transcriptional regulator of nitrogen metabolism mediates a trade-off between growth in resource abundant and resource limited environments in *S. cerevisiae*. Redundancy in the metabolic pathways of ammonia assimilation allowed noise, or random fluctuations in the amount of protein present, to dictate whether cells specialized in maximizing fitness in abundant or limiting environments. These results show how gene expression may be programmed via noncoding RNA regulators, and that the manipulation of regulator levels can affect the strategy by which organisms adapt to their environments.

TABLE OF CONTENTS

Chapter 1.

Synthetic Approaches to Understanding Biology

The construction of synthetic circuits has illuminated how cells process and respond to environmental signals, and how cells can be engineered to perform useful functions. Recent progress in the regulatory functions of noncoding RNAs suggests that nucleic acids can be employed as a design substrate for programming cellular function. However, as discussed below, the construction of increasingly complex systems will require an appreciation and understanding of trade-offs between biological functions. An understanding of how organisms tolerate and manage trade-offs in function between different environments could provide design principles for building robust systems. In addition, such work could lead to an understanding of whether the architecture of genetic circuits can shape the organization of larger-scale ecological networks.

1. Synthetic strategies for understanding genetic regulation

1.1 Reconstruction of genetic circuits to understand regulation

The complex functions of biological systems such as organisms, metabolic and developmental pathways, and proteins are ordinarily studied by analysis of genetic and biochemical perturbations. The modularity of biological components like genes and proteins enables a complementary approach: one can construct and analyze synthetic systems such as genetic circuits, organisms, and proteins with unnatural monomers inspired by their natural counterparts. Synthetic biology is the construction of existing or novel biological function from constituent components. Researchers may desire to do this for several reasons – one is to manipulate or measure existing biological systems in more sophisticated ways, leading to a greater understanding of biology. The study of how the structure of synthetic circuits relates to their behavior can potentially illuminate Nature's "design principles," or rules for how evolution has solved the adaptive challenges faced by an organism, although very few studies have yet linked network architecture to adaptive strategy for a given organism and environment. However, several examples of building genetic circuits to understand the biological significance of these architectures are noted below.

An examination of the adaptive and functional significance of network architecture is a daunting task given the complexity and diversity of examples from the natural world. One strategy to tackle this problem is not to break down the complexity of natural

networks, but to construct such systems from the bottom-up. The construction of biological regulatory circuits is at the heart of synthetic biology, and can be envisioned as a physical model of circuit architecture for testing hypotheses about the information processing, adaptive, and functional significance of such architectures.

Observations of the structure of genetic regulatory networks in yeast, bacteria, and other organisms have shown that these networks are comprised of repeated patterns of interaction between genes, knows as motifs.[1] Negative feedback, multi-input motifs, and feedforward loops are among the motifs that are observed far more often than would be expected for a random network. Several groups have tested the functional and adaptive significance of these motifs using a combination of theory, modeling, and construction of synthetic motifs. One example comes from the work on transcriptional feed-forward loops from the Alon lab. Feed-forward loops (FFLs) are three-gene motifs, where a first gene regulates a second, and the two in turn co-regulate a third gene. Two widely observed types are the *coherent* and *incoherent* FFL. Coherent FFLs consist of two activators that regulate a third component, while incoherent FFLs consist of one activator and one repressor.[2] Modeling these motifs with simple differential equations suggested that they have different information processing capabilities: signals that activate the expression of the first gene (the "input") are propagated with different temporal dynamics and strengths. For example, coherent FFLs act as signal delays and are able to filter transient inputs from affecting output gene expression. Incoherent FFLs can act as signal accelerators, showing faster induction of output gene expression than simple regulation by a single transcription factor. To test these models, Mangan and Alon examined the

arabinose utilization system of *Escherichia coli*, which shows FFL topology.[3] They found that the motif displayed "sign-sensitive" kinetics: the induction of motif output was slower than the signal decay after input was removed. This function could be employed by cells to filter noisy inputs (to avoid spurious gene expression, which is metabolically costly), while still being able to turn genes off rapidly when the input signal is removed. Thus, experimental evidence confirms the functional and potentially adaptive significance of a highly abundant feature of network organization.

Another example of the biological insight gained by constructing synthetic circuits comes from post-translational signaling pathways. Eukaryotes utilize mitogen-activated protein kinases (MAPKs) for a variety of signaling functions. In yeast, environmental signals initiate a cascade of phosphorylation events between kinases, ultimately resulting in changes in gene expression of a number of pathways, such as osmolarity responses and mating pathways. The MAPKs are associated with "scaffold" proteins of uncertain function: does the scaffold merely bring kinases in close physical proximity, or does it regulate, activate, and otherwise provide an additional point of control on the cascade? To test this, the Lim group replaced native protein-protein interactions between the kinase and scaffold with heterologous interaction domains from other proteins and found that the cascade was able to function properly.[4] The researchers extended these findings by creating synthetic scaffolds that brought together kinases that are not naturally associated. The synthetic scaffolds possessed novel input-output properties, demonstrating that scaffold proteins could be useful in evolving or engineering new cellular behaviors. In addition, these results convincingly showed that simple tethering is

sufficient to explain MAPK cascade behavior. This work demonstrates how synthetic approaches can complement and extend analytical approaches for biological discovery.

One underlying question in each of these cases is whether the observed network features represent functional units of information processing in the cell, or are merely evolutionary "artifacts" without functional significance. As discussed in section 4.3, general and global network topology and architecture can often be mistakenly ascribed to adaptive origins. A challenge for researchers is to test the functional and adaptive significance of observed patterns in genetic networks.

1.2 Construction of genetic circuits as an engineering discipline

Synthetic genetic circuits are valuable tools for understanding the natural world, and also have shown great utility in the engineering of biology. One advantage to a theoretical and practical understanding of biological circuit behavior is the ability to both design and evolve applications to many pressing technological problems such as therapeutics, energy, bioremediation, and material synthesis. Several examples of synthetic biology as an engineering discipline are described below.

Therapeutics and Human Disease

The diversity of chemical compounds made by biological systems found in Nature is staggering. Many natural compounds have historically been utilized as antibiotics and

other therapeutics.[5] Several groups have engineered microbes to produce fine chemicals for therapeutic purposes.[6] The Keasling group at UC Berkeley has engineered *Saccharomyces cerevisiae* to produce arteminisic acid, a precursor to artemisinin, a potent anti-malarial naturally found in the plant *Artemisia annua*.[7] Currently, the chemical synthesis of artemisinin is cost prohibitive to the population of the Third World, where it is needed most. The researchers were able to use an engineered mevalonate pathway, an amorphadiene synthase, and a cytochrome P450 monooxygenase from *A. annua* to produce artemisinic acid. The engineered yeast produced higher artemisin yields that *A. annua*, although the authors note that industrial scale-up and optimization will be required to make this route to production cost-effective.

In addition to engineering microbes to produce therapeutics, several groups are exploring the use of live cells *as* therapeutics.[6, 8, 9] Microbes in their natural state are endowed with many functions that could be utilized to discriminate between healthy and disease states (such as receptors and environmental sensing components) and act in a therapeutic manner (such as synthesizing therapeutic proteins, invading disease cells, or synthesizing chemicals, as above). Towards these aims, the Voigt group at UC San Francisco engineered *E. coli* to sense and destroy cancer cells by environment-dependent control of invasion.[10] The authors used invasin from the pathogen *Yersinia pseudotuberculosis* as an output that allowed *E. coli* to invade mammalian cells. To render the bacteria cancer cell-specific, invasin expression was controlled by several heterologous sensors: the *Vibrio fisheri* quorum sensing circuit, the hypoxia responsive fdhF promoter, or the arabinose-responsive araBAD promoter. Each of these is designed to induce invasin expression and

bacterial invasion only in the presence of tumor cell environments (for example, tumors are highly hypoxic) or via external, researcher-inducible control. The authors were able to demonstrate invasion of several cancer-derived cell lines with these engineered bacteria, demonstrating that cells can be programmed with sensors and outputs to achieve therapeutic functionality. Taken together, this work shows that natural functions of bacteria can be re-engineered and augmented to construct useful functions.

Biological Pattern Formation

The coordinated organization of cells in specific patterns is a classic example of complex function in many organisms and is central to the development of multicellular organisms from a single fertilized oocyte. Pattern formation typically involves signaling and communication between cells, processing of these signals, and modulation of the expression of a variety of genes. The ability to design pattern formation could hold great utility in applications such as tissue engineering and biomaterials. Towards these aims, several groups have explored how collections of cells can be programmed to form user specified patterns. In one example, researchers enabled *E. coli* cells to communicate with each other using the quorum sensing system from *Vibrio cholerae*.[11] Engineered "receiver cells" were designed to express fluorescent proteins based on the concentration of the signaling molecule synthesized by a "sender" cell. The receiver cells were designed such that they were responsive only to a defined range of signaling molecules, analogous to a bandpass filter. The range of signaling molecule the receiver cells was responsive to could be tuned by changing the kinetic parameters of the underlying

information processing circuit. The authors were able to use combinations of sender and receiver cells to create two-dimensional patterns on a lawn of cells such as a bulls-eye, ellipses, and clovers. Thus, the engineering of underlying functionality (i.e., quorum sensing) and gene circuits allowed spontaneous pattern formation in a population of cells.

Other approaches have been used to program pattern formation that are inspired more from lithographic and printing techniques rather than developmental pattern formation. In a stunning example of engineering synthetic functions into organisms, the Voigt and Ellington groups constructed a strain of bacteria that could sense red light and control gene expression accordingly.[12] To accomplish this, the researchers constructed a chimeric two-component photorhodopsin system from the cyanobacteria *Synechocystis* in *E. coli*. When coupled to the expression of a chemical output, this function allowed a lawn of bacteria to act as a photographic film – projection of an image onto the lawn results in the recording of a high-definition two-dimensional chemical image at resolutions up to 100 megapixels per square inch. The control of pattern formation in living cells could have important applications in constructing complex patterned biomaterials, tissue engineering, and parallel biological computation. In an extension of this work, Tabor and Voigt (personal communication) have enabled massively parallel 'edge detection' of a projected image such that cells communicate to discriminate and delineate boundaries between cells sensing light and dark regions of the image. These efforts hold great potential to explore how biology uses large numbers of computational elements (in this case, cells) to compute complex problems and to combine "top-down" lithographic-style patterning with "bottom-up" parallel computation to specify patterns.

1.3 Nucleic acids as a substrate for designing circuits

The success of rational design of desired function in an engineering discipline is largely determined by the design substrate(s) available. One strategy to design component-level function in biology is to use nucleic acids as a substrate. The recognition that RNA plays a central role as not only an information carrier but as a catalyst of biochemical reactions and of genetic regulation has driven the development of diverse strategies for using RNA and DNA in multiple applications.

Roles of RNA in contemporary biology and in the evolution of life

RNA is pervasive in fundamental biological processes. As an information-rich molecule, it is responsible for carrying information between DNA and the translational machinery and guiding the processing and editing of ribosomal RNA (rRNA). In addition, RNA primes the process of DNA replication and individual nucleotides are used as cofactors in enzymatic reactions. Although RNA naturally is composed of only 4 nucleotides (in contrast to the 20 amino acids that make up proteins), it is able to fold into diverse tertiary structures that can display binding and catalytic activity. The crystal structure of the ribosome revealed that the catalyst of protein synthesis is in fact RNA – a function that likely doomed the RNA world to extinction, replaced by the protein universe. In 1980, Cech and Inoue discovered that the splicing reaction of a rRNA intermediate in *Tetrahymena* was able to proceed in the absence of a protein enzyme via a specific

tertiary structure in the RNA able to perform the catalysis, which they termed a ribozyme.[13] Several other naturally occurring cleavase ribozymes have been discovered in central roles in biological functions, including RNA processing and viral genome replication. The biochemical functionality of RNA is highlighted by the recent discovery of RNA regulatory elements in prokaryotic metabolic genes.[14] Breaker and co-workers discovered conserved RNA structures in the 5' untranslated regions (UTR) of genes involved in cofactor metabolism in *Bacillus subtilis*[15]. These structures were able to directly bind the small molecule cofactor, and by virtue of a conformational change upon binding, occlude the ribosome-binding site (RBS) of the mRNA and control translation. These riboswitches have now been observed in prokaryotes, archaea, and eukaryotes and have been found to regulate translation, transcript stability, and alternative splicing in response to metabolite effectors.

In the past 10 years, the discovery and characterization of a highly conserved RNA-based regulatory mechanism has fundamentally altered the conception of the biological function of RNA in cells. RNA interference (RNAi) was first described in *Caenohabditis elegans* and later shown to be present in many eukaryotes such as fission yeast and mammalian cells[16]. In RNAi, a small (21 nt) double-stranded RNA effector guides the sequence-specific silencing of genes. RNAi silencing is guided either by exogenously delivered small interfering RNAs (siRNAs) or by endogenously produced microRNAs (miRNAs). These RNAs silence the expression of target genes in several ways: one is by base-pairing complementarity to a target transcript, mediated by a multi-protein complex known as RISC (RNA-induced silencing complex). Catalysis is performed by the

endonuclease Argonaute 2 in the complex. RNAi has been used as a tool for targeted gene knockdown in mammalian cells, but has also been shown to be a significant mechanism for cellular regulation. miRNAs are able to fine-tune gene expression during differentiation and development by recognizing the 3' UTR of target genes. The loss of miRNA-mediated regulation in mutants has been linked to oncogenesis as well as developmental defects.[17]

Taken together, these discoveries clearly demonstrate the importance of RNA in biological systems far beyond the original conception that RNA was merely an information carrier. One way that the powerful biochemical, regulatory, and informational properties of RNA can be expanded upon is by directed evolution in the laboratory - recapitulating Darwinian selection on populations of functional RNAs.

Evolving function from populations of RNA

Contemporary theories on biological origins suggest that life arose via a self-replicating molecule or assembly of self-replicating molecules. Because of the almost unique complementarity inherent in nucleobases, it has been suggested that nucleic acids or nucleic acid-like molecules were the first self-replicators, although some researchers hold that self-replicating peptides or lipid systems may have preceded or emerged in parallel with nucleic acids.[18] An early nucleic acid replicator is also tempting because of the prevalence of molecular fossils in modern cells that are related to nucleic acids, such as ATP and other cofactors. Given the fact that the ribosome is at its core a ribozyme, it

seems highly likely that there was a complex RNA world that preceded the modern world of protein catalysts.[19] This RNA world may have in turn descended from an early nucleic acid replicator by duplication, parasitism, and diversification.

Researchers have recapitulated the Darwinian evolution of RNA self-replicators as well as other functional RNAs.[20] One of the earliest examples of extracellular evolution of molecules is the work done by the Spiegelman group.[21] In this work, the researchers discovered that the bacteriophage Qβ utilized a RNA-dependent RNA polymerase to replicate its RNA genome. In this simple system (containing replicase, template RNA, and nucleotides) the researchers found that the system showed autocatalytic kinetics, indicating that self-propagation of the viral genome was occurring and could do so *in vitro* with the necessary components. Furthermore, the Qβ system was used to set-up an extracellular evolution experiment where mutant templates could "compete" with one another for the limited pool of replicases and nucleotides, eventually selecting for template sequences that were able to replicate faster than the original (parent) template[22].

The strategy of competing populations of diverse RNA molecules (analogous to Darwinian evolution) was used by Ellington and Szostak to evolve ligand-binding species. The authors termed the evolved ligand-binding RNAs "aptamers" (from the Greek *aptus*, to fit). This work[23] demonstrated how randomized pools of molecules could be selected to bind organic dyes. Further work demonstrated that aptamers could be evolved to bind other small molecules, proteins, oligosaccharides, as well as cells, tissues, and organisms (by virtue of molecular recognition on the cellular surface).[24]

Aptamers rival protein antibodies in binding specificity and affinity, and have been shown to elicit minimal immunological responses[25]. These properties have led to the development of aptamer therapeutics[25]. The first such therapy to gain FDA approval is an aptamer that binds the vascular endothelial growth factor (VEGF), used in the treatment of age-related macular degeneration.[26]

The success of aptamer selections has enabled several groups to explore the unique structural biochemistry of binding between aptamer and ligand. Because aptamers are evolved solely for binding functionality, the resulting structures are not optimized for other functionality and can provide insight into the universe of possible chemical solutions for RNA-ligand binding. Three-dimensional structural analyses have provided insights into the nature of recognition by nucleic acid-aptamer complexes.[24] The enclosure of large regions of the ligand by the nucleic acid is the basis for specific recognition of the ligand in aptamer complexes. Multiple intermolecular contacts between the nucleic acid and ligand provide specificity. For example, steric occlusion of a methyl group prevents binding of caffeine to the theophylline aptamer.[27] Specific hydrogen bonding creates an interaction with the ligand similar to a base-pair in the AMP aptamer.[28, 29] Aminoglycoside antibiotics are bound by their aptamers by a combination of both electrostatic and hydrogen bonds that create complementarity between ligand and aptamer.[30] Aptamers selected to bind peptides and proteins from HIV-Rev and other viruses often involve stacking, electrostatic contacts, and induced fit of both the peptide and the aptamer.[31, 32] In nucleic acid - ligand binding, the structurally similar nucleotides are limited in the number of ways that they can be structured around a given ligand. The

interaction of ligands into aptamer binding sites often displays imperfect complementarity, which can be compensated for by adaptive recognition – structural rearrangement of the ligand and/or the aptamer.[33, 34] Furthermore, many aptamers contain disordered loop regions that acquire an ordered structure upon ligand binding. Adaptive recognition and the formation of specific contacts between aptamer and ligand are responsible for a general trade-off between binding affinity and an inability of several aptamers to bind variant forms of their specified ligand. For example, aptamers selected to bind the HIV Rev peptide with high affinity are unable to bind even single mutant forms of the peptide, limiting their use as therapeutics for the rapidly mutating HIV virus.[35]

In parallel with the evolution of binding functionality in aptamers, several groups developed schemes for the evolution of catalytic ability from randomized nucleic acid pools. Catalytic RNAs are found throughout biology, as in the protein translation machinery (the ribosome), as well as controlling gene expression[36] and carrying out splicing.[37] Using similar methods to the *in vitro* selection of aptamers (creating pool diversity, partitioning higher fitness molecules, and amplification) researchers have been able to discover RNAs that catalyze a wide range of reactions, including ligation of other nucleic acids,[38] cleavage of nucleic acids,[39] and even a ribozyme that catalyzes an alcohol dehydrogenase reaction,[40] showing that ribozymes could perform redox chemistry. Catalytic nucleic acids have allowed structural insights to nucleic acid function and have played roles in engineering new functions, as described below.

The crystal structures of several natural and *in vitro* selected ribozymes have been solved, giving insights on structural design principles for RNA catalysis.[41] One of the more general themes that emerges from this body of work is that ribozymes catalyze reactions in the same ways that proteins do: by forming substrate-binding sites to decrease the entropic cost of attaining the transition state, having more favorable interactions with the transition state structure versus the ground state, enabling chemistries that involve the movement of protons, and raising the energy of the bound substrate relative to the transition state. Ribozymes are not limited to using only metal ions as functional groups in catalysis (as was proposed in the infancy of ribozyme research), but can use nucleotide bases, sugar hydroxyls and phosphate backbone to accomplish chemical functions[41]. Due to limitations in the diversity of side chains, ribozymes are not as adept at catalyzing the wide variety of reactions that protein enzymes can, which may be one reason that the protein-based biology we observe today triumphed over its RNA-based ancestors.

A recently published ribozyme structure highlights the ability of directed evolution to find novel biochemical "solutions" to a particular catalytic "problem". The L1 ligase was isolated by *in vitro* selection to catalyze the ligation of two RNA molecules.[42] The ribozyme catalyzes a nucleophilic attack by a 3'-hydroxyl group on the phosphorus of the ribozyme's 5'-triphosphate, creating a new phosphodiester linkage and releasing the pyrophosphate. The ligase creates a catalytic pocket by a unique triple base interaction, with contributing residues from invariant nucleotides on each of three helices[43]. This catalytic core juxtaposes the ends of the ligation reactants with a Mg^{2}+ ion. Although this structure and mechanism share some features with that of the hammerhead ribozyme, the

nucleotide geometry involved in catalysis in this ligase has not been observed before, but was "discovered" by directed evolution of ligase function from a random pool. This suggests that even with an incomplete understanding of biophysical principles, directed evolution remains a powerful approach to creating functional nucleic acids. The directed evolution of nucleic acid functionality has enabled the construction of useful biotechnological tools.

Harnessing evolution and design of RNA to engineer function

RNA has two properties that enable researchers to engineer and evolve novel and useful functions for biotechnological applications. One is the above noted biochemical functionality of RNA, such as binding and catalysis. Another is the informational functionality of RNA, the ability of the primary sequence of nucleotides to be "read off", amplified, and manipulated with available enzymes and techniques in a test tube. While all biological molecules can be said to possess information content in the sequences of monomers or chemical structures composing the molecule, the ability to transform or amplify this chemical information (such as protein sequence or metabolite functional groups) has not been demonstrated. In contrast, the information for a particular ribozyme-catalyzed reaction is embodied in the sequence of the RNA molecule, which can be reverse transcribed and PCR amplified to create millions of copies of the ribozyme information. Researchers have exploited this dual informational/biochemical property to develop strategies for biosensing, genetic regulation, cell-specific targeting, and other applications.

An example of combining directed evolution with rational design to create functional nucleic acids is work done by the Sullenger lab on reversible aptamer-based anticoagulants. Sullenger and colleagues first selected an aptamer to bind the clotting factor Ixa.[44] Blood plasma based screening revealed that the aptamer was a potent anticoagulant that acts by blocking further activation of the clotting factor zymogen cascade. The authors then introduced "antidote control" to the aptamer through addition of a complementary oligonucleotide that hybridizes to the aptamer. Thus, the design of an aptamer-antidote pair is intuitive: each aptamer sequence inherently carries the prescription for its antidote. The authors further demonstrated the power of nucleic acid engineering by demonstrating the use of this anticoagulant – antidote pair in a mouse model.[45] To overcome the inherent instability of nucleic acids in the bloodstream, chemically modified nucleotides were introduced together with a pendant cholesterol group to improve bloodstream retention time.

In another example of the dual roles of selection and design in creating functional nucleic acids, several groups have created cell-targeting RNA molecules towards the goal of disease-cell specific drug delivery. The Levy group at Albert Einstein College of Medicine used aptamers selected to the known cancer biomarker prostate-specific membrane antigen (PSMA) that displayed high affinity and specificity.[46] They then conjugated these aptamers to the drug gelonin – due to receptor cycling to and from the membrane, the aptamer-drug conjugate bound PSMA on the cancer cell surface and was internalized, killing the cell through the action of gelonin.[47] The aptamer-drug conjugates

showed 600-fold increased potency compared to cells that did not express PSMA. These results demonstrate how selection can be used to create "magic bullet" therapeutics that target disease cells while leaving healthy cells unharmed, lowering the dose of drug needed and mitigating side-effects to the patient. In separate studies the same PMSA aptamer was used to deliver gene-targeting siRNAs to prostate cancer cells,[48, 49] extending the therapeutic possibilities of that aptamer delivery method. One promising possibility for the creation of cell targeting conjugates is the advent of whole-cell selection procedures. In the above example, PSMA, a known cancer marker, was purified to select aptamers against. However, in many cases extracellular markers of disease are not known. In these cases, selection of aptamers to living cells can be used to discriminate differences on the cell surface. In a recent study,[50] aptamers were generated by selecting against entire leukemia cells. Specificity was achieved by performing a "negative selection" against healthy cells. A negative selection partitions and discards aptamers that bind the negative target (in this case healthy cells) after the selection step (in this case, partitioning of aptamers that bind leukemia cells). Cell-based aptamer selections can be used to evolve aptamers to discriminate between any cell types, and can show utility as cell-specific therapeutics when coupled with drugs or toxins as above.

Information processing with nucleic acids

The dual functions of nucleic acids observed in biology of information carrier and biochemical actuator have been coupled to demonstrate how sets of biomolecules can perform elementary computation with molecular inputs and outputs. A seminal paper by

Adelman demonstrated that nucleic acids can be used not only to encode information but could perform elementary computational operations that could theoretically scaled to yield massive computational power.[51] These efforts can aid in creating complex circuits for nanoscale computation and in understanding how biological systems process information given the constraints of physics at the molecular scale. Several examples are described below.

Stojanovic and Stefanovic used deoxyribozymes to demonstrate the computational abilities of nucleic acids could interpret multiple simultaneous inputs by designing a DNA automaton that could play a game of "tic-tac-toe" against a human player.[52] The outputs of the device were deoxyribozyme cleavage events that yielded single-stranded oligonucleotides. The user-supplied inputs were also single-stranded DNA. Boolean logic linking inputs to outputs was accomplished by engineering allostery into the deoxyribozymes – upon addition of input oligos, the catalysts cleave, yielding an output.[53] The authors were able to incorporate allosteric domains in such a way that the higher order logic functions required to play the game against a human were accomplished. The automaton strategy is thus "hard-wired" in the structures and arrangements of the deoxyribozymes. While this work required a large amount of empirical tuning (to achieve reasonable signal-to-noise ratios, for example), it remains as one of the first examples of building a complex, predictable network from programmable enzymes.

In recent work, Winfree and colleagues demonstrated how cascades of logic functions can be implemented using strand displacement of DNA, without catalysis.[54] The authors were able to show that DNA-based logic gates using their design (based completely on base-pair hybridization) showed the hallmarks of digital abstraction employed in silicon-based electronics such as logic, cascading, signal restoration, fan-out, and modularity. Because of these features, the authors were able to build a complex device using 11 logic gates and performing complex computational tasks such as signal amplification, restoration, and threshold behavior. This work establishes design principles for information processing by nucleic acids, and could be used to control nanoscale devices *in vitro*, to analyze complex chemical samples, or to interface with existing biological circuits inside living cells.

Genetic regulation: information processing and biochemical functionality inside cells

Several groups have begun to use the ease of RNA directed evolution and rational RNA design to construct components for the regulation of gene expression programs inside cells. Researchers have used the ability to rationally design interaction energetics between RNA molecules (via base-pairing) to construct a post-translational regulation system in *E. coli*.[55] Translation in prokaryotes requires the interaction of the ribosome with a (RBS), and as such, translation can be inhibited by occluding the RBS. The researchers used a *cis* stem-loop structure that binds the RBS to silence gene expression. In an elegant demonstration of the power of base-pairing specificity and binding energetics between RNA molecules, the authors then used a *trans* RNA molecule to

activate gene expression. The *trans* RNA targets the *cis* repressive stem-loop via base-pairing, causing a conformational change that activates translation. The interactions of *cis* and *trans* elements are highly specific due to the combinatorial nature of base-pairing, and as such the same design schemes could be used to construct a large number of orthoganal *cis-trans* riboregulator pairs to build increasingly complex networks.

Several groups have utilized the ligand-binding properties of RNA as well as sequence specificity to engineer novel behaviors in living cells. The theophylline-binding aptamer was inserted in the 5' UTR of a reporter gene in *S. cerevisiae* to achieve ligand-dependent regulation of the target gene.[56] Another group then used this scheme to control gene expression in *E. coli* and evolved the dynamic range of this system, from 12-fold difference in gene expression with and without theophylline to a 36-fold difference.[57] The evolved riboswitch was then cloned upstream of the CheZ gene which led to control of bacterial motility in the presence of theophylline[58]. Cells were able to "follow" a trail of theophylline arranged specifically on a plate. Although the engineered chemotactic system differs from the normal control mechanism in several ways (for example, regulation is based on slow changes in gene expression rather than the wildtype protein phosphorylation cascade), the "phenotype level" function is the same – cells sense and move towards a user-specified chemo-attractant. These studies show how RNA can be designed, evolved, and engineered to meet the functional demands for engineering complex behaviors and functions in living cells.

Finally, several qualitative and quantitative differences between regulation of gene expression with noncoding RNA and regulation with proteins have recently been elucidated. Hwa and colleagues combined theoretical and experimental approaches to study regulation by a class of *trans*-acting noncoding RNAs in *E. coli*, small RNA (sRNA). There are over 70 identified sRNAs in *E. coli* that have been implicated in regulating diverse functions such as osmotic response, quorum sensing oxidative stress, DNA damage, and others[59]. This work found that regulation by sRNAs involved a "threshold-linear" response where repression is tight under a threshold of sRNA synthesis rates and is linear depending on target gene and sRNA synthesis above the threshold. In addition, compared with protein regulation, sRNA regulation displayed characteristic noise resistance and a capability for "hierarchical cross-talk" - many sRNAs bind several targets, and because target mRNAs can titrate sRNA by binding them, two target genes can thus have an indirect affect on each other. These results demonstrate that in addition to "programmability" and ease of design, engineering genetic regulation with noncoding RNAs may allow engineers to build novel quantitative and qualitative modes of regulation in living cells.

2. Constraints and trade-offs with increasing complexity

As we begin to assemble components to create functional devices, a new understanding of potential trade-offs and design criteria for optimization will be necessary. Trade-offs and constraints can bias the design criteria of natural and engineered systems. The

engineering of more complex systems will require an understanding and appreciation of the trade-offs between biological functions.

2.1 Trade-offs shape circuit design and function

An understanding of constraints and trade-offs in system function is critical for understanding how to construct more complex biological devices. Design principles will hold the greatest utility if underlying complexity can be effectively "hidden" from users at higher levels. An understanding of the foundations of trade-offs and constraints will enable biological engineers to have a set of "rules" for composing modules and a guideline for expected behavior. Recent work in circuit analysis suggests that trade-offs between system functions may require a greater appreciation of trade-offs at the cellular scale.

Recent work has described how the architecture of the *E. coli* heat-shock response is able to balance a trade-off between temporal response and efficiency. Heat shock causes the rapid unfolding and aggregation of many proteins in the cell, compromising normal function. Cells respond to heat shock by inducing the expression of heat-shock proteins (Hsp), chaperones that enable unfolded proteins to fold to their native structure, as well as proteases that degrade misfolded protein that are beyond repair.[60] Because the effects of heat shock occur on the order of seconds, the heat shock response must be activated quickly. However, the spurious induction of this massive cellular response is

metabolically costly for the cell, requiring tight regulation of its induction. Thus, the system must be optimized for several functionalities such as robustness, efficiency, response speed, and rejection of noise. Given limited cellular resources and energy, these demands are often contradictory – for example, high turnover rates in heat shock regulator production are necessary for a fast response, but come at a cost to metabolic efficiency.[61] Recent work has demonstrated how the architecture of the heat shock system balances such trade-offs.[62] For example, numerous feedback and feed-forward loops were identified that couple the induction dynamics, steady-state levels, and decay of multiple regulators of the heat shock response. Interestingly, the researchers note that the evolution of control architectures to balance these trade-offs results in new fragilities and constraints in the system, a recurring theme in the analysis of biological and man-made systems. The inherent trade-offs and constraints between biological functions and traits are also a significant determinant in shaping how organisms adapt and evolve.

2.2 Trade-offs in life history evolution

The recognition that trade-offs and intrinsic constraints can play a large role in determining the composition of systems has a lengthy precedent in evolutionary biology and ecology. A central assumption in many theories of trait evolution is the existence of trade-offs between functions. Trade-offs are based in fundamental mechanical, physiological, or thermodynamic constraints in metabolism, behavior, morphology, and other functions. One clear example of such constraints is the widely observed trade-off is the trade-off between fertility and survival. To take one case, the survival versus fertility

trade-off in the guppy *Trinidata sp.* is determined by fluid mechanics of swimming.[63] Because guppies are viviparous (give birth to live offspring rather than laying eggs) individuals with large litter/clutch sizes tend to be larger and more rotund than individuals with small litter sizes. The difference in size creates more drag in the water, making it more difficult for the individual to avoid predators. This trade-off can shape adaptive strategies available to a population and can reveal "signatures" of adaptation in the field due to selection determining the optimal balance of traits along the trade-off (from high fertility to high survival). Thus, populations that experience low predation would tend to optimize fertility over survivability, while populations experiencing high predation would tend to optimize survivability. The knowledge of mechanistic determinants of a trade-off can thus illuminate the ecological and evolutionary history of populations in the wild.

A related and prominent trade-off theory in biology is that of *r* versus *K* selection.[64] Organisms under *K* selection are predicted to optimize utilization of resources, such as when the population is near its carrying capacity and resources are scarce. *r* selection occurs when resources are abundant and organisms maximize per capita growth rate. For example, organisms generally face a trade-off between two modes of ATP production, respiration and fermentation – fermentation is faster but less efficient than respiration. Environmental conditions can select for organisms displaying a rate or yield strategy in this case, as is detailed below.[65] Although a given organism may employ more than one ecological strategy, the hypothesis of a general trade-off between growth rate and yield

(the efficiency of resource utilization) remains central to theories regarding the evolution of cooperation and evolution of generalists versus specialists.

Similar, but importantly distinct, trade-offs are well established in metabolic function. The laws of thermodynamics imply a trade-off between the rate of ATP production (moles ATP / time) and the yield (moles ATP / moles substrate) for any catabolic reaction.[66] This trade-off has been experimentally demonstrated in *E. coli* and *S. cerevisiae.*[65] One example is the rate-yield trade-off in two ubiquitous modes of sugar degradation, respiration and fermentation. Respiration is rapidly saturated at high levels of substrate or limiting oxygen, such that organisms can increase the rate of ATP synthesis by fermentation in addition to respiration. However, the yield for fermentation is much lower than respiration (2 moles ATP / mol glucose for fermentation, 32 moles ATP / mol glucose for respiration). Thus, cells can either "choose" to produce ATP rapidly (fermentation) or to extract a higher yield of ATP from a given amount of substrate (respiration). Pfeiffer and Bonhoeffer[66] showed that high yield ATP production can be viewed as cooperation between cells and can evolve in spatially structured environments.

An understanding of potential design constraints comes from an intimate knowledge of the (often) conflicting strategies of coping with environmental challenges. One recent example that has been elucidated is the trade-off between multiplication inside a host and survival outside a host in *E. coli* infecting bacteriophage.[67] Here, the authors found that phage mortality rate outside the bacterial host is inversely correlated with multiplication

rate for the viruses inside the host. Survival time outside of a host is a critical phase of the life cycle for parasites transmitted via the environment. Coliphages can have significant variability in the period between hosts – for example, a phage may be able to infect other *E. coli* cells in the same animal gut (short time between infection), or may be excreted into the external environment (long time between infection). The authors were able to propose and support a mechanism for the trade-off between reproduction and survival. The major determinant of viral survival is the stability of the capsid shell. This study found that the surfactic mass of the capsid as well as the density of the packaged phage genome was positively correlated with survival[67]. However, the energetic and temporal demands for stronger and more densely packed capsids resulted in slower virion production rates inside the host. In this case, evolving population are forced to optimize and balance a trade-off between fitness benefits in two different environments. Human engineers designing a system such as a bacteriophage would need a mechanistic understanding of capsid thickness and genome packaging to determine which function to optimize (reproduction or survival) or how to balance functionality in their engineered system.

3. Genome-scale organization

As we move towards an understanding of how constraints, trade-offs, and optimization drives the evolution of biological function, an open question is whether there are overarching principles or themes that accurately depict the organization of entire biological systems. Such “systems design principles” would be valuable in genome-scale

engineering. The capability to chemically synthesize large DNA sequences is rapidly becoming faster, cheaper, and more reliable such that the synthesis of entire genomes could be feasible for academic or industrial labs.[68] At this point, the rate-limiting step of engineering biology will not be fabrication (i.e., cloning), but will be design. In addition, the existence of fundamental genome-scale design themes could enable a deep understanding of adaptation at the systems level.

3.1 Organizing biology: engineering perspectives

One paradigm for designing biological systems is the use of an abstraction hierarchy for managing complexity.[68] Abstraction hierarchies are widely used in software engineering. For example, high-level programming languages (such as C++) enable programmers to read and write code in a form that is understandable to humans. The high level language is translated into machine-level instructions, which are in turn translated into bit strings for interpretation by the machine. Thus, software engineers can work on specific parts of the hierarchy such as high-level languages or machine-level instructions.

The effect is two-fold, and has many potential similarities in biology and evolution. The first is that the abstraction allows ease of design. Specialists in designing instruction sets (machine-level) do not have to create novel machine-level methods for each new high-level application, whereas high-level programmers do not need to concern themselves with the detail of machine-level operations when writing new code. Another effect of an abstraction hierarchy is the promotion of diversity in high-level applications. Software

engineers are able to use, re-use, and combine high-level functions (specified at the machine-level) to create a myriad of applications. The concept of the recombination of functional modules to create novel systems has parallels in protein evolution. Recent work with eukaryotic signaling proteins has shown how the recombination of a common set of domains can create novel input-output functionality such as allosteric gating, signal integration, and ultrasensitivity.[69]

Several abstraction hierarchies for synthetic biology have been proposed. One is the "parts-devices-systems" hierarchy.[70] In this concept, at the lowest level of complexity, are parts – discrete sequences of DNA that encode for functional units such as (but not limited to) genes, RBS, transcriptional terminators, and promoters. These parts can be characterized and assembled into devices. Devices include multiple parts with a specified function, such as a transcriptional inverter, a signal amplifier, or a toggle switch. At a higher level, devices can be assembled into functional systems. Some examples include oscillators, pattern-forming cells, and tumor-invading bacteria described previously. One goal of synthetic biology is to explore the utility of such abstraction hierarchies in the pursuit of engineering biological function.

3.2 Optimality principles in network organization

Optimal function is a fundamental and shared idea between engineering and evolutionary biology. One open question is toward what functions evolution has optimized metabolic networks. Organisms face multiple challenges from their environments, including

fluctuating nutrient sources, a battery of stress conditions, and potentially deleterious (and beneficial) interactions with other organisms. The optimal organization of metabolic networks would shed insight on the evolutionary forces that have shaped networks as well as provide "design principles" for the forward engineering of organisms for metabolic engineering. The Palsson group has shown that the "object" of metabolic network organization (the "objective function" in their parlance) is to maximize growth, and as such, potential fluxes through a metabolic network can be derived.[71] For example, microbes have evolved to efficiently convert carbon and energy into biomass, the creation of more cells. In an elegant demonstration of this objective function, Fong and Palsson grew *E. coli* on glycerol (a non-preferred carbon source for this organism) and observed that metabolic fluxes did not operate according to the optimality principle.[72] However, after applying selective pressure by repeated growth on glycerol, the network evolved to maximize its growth rate on this substrate according to the predictions described by an *in silico* model. The utility of this simple evolutionary principle has been validated by its use to predict essential genes in metabolic networks, as well as its use to optimize ethanol production in *S. cerevisiae*.

However, alternative selective pressures may result in metabolic networks that are optimized for other functions besides growth. The Sauer group analyzed metabolic gene knockouts in *Bacillus subtilis* and found that several mutants grew faster than wildtype, showing that bacteria does not operate solely according to maximized biomass production.[73] The authors found that *B. subtilis* has suboptimal metabolism because it invests a significant amount of resources in anticipation of changing environmental

conditions, which the authors term a "stand-by mode". For example, most identified knockout mutants displaying increased biomass production were regulators of alternative phenotypic states of *B. subtilis*, such as sporulation and competence. These developmental pathways are activated in starvation conditions and are repressed in nutrient rich environments. Thus, network organization is a compromise between rapid growth in resource abundant environments and the anticipation of environmental change. These results demonstrate how multiple selective pressures can shape metabolic networks and suggest that human engineers will similarly have to balance multiple functions in designing useful biological systems.

3.3 Game theoretic strategies for maximizing competitive ability

Optimality principles such as those described above often make the implicit assumption that populations gradually ascend peaks in a fitness landscape, and as such, gradually increase in fitness over time. However, fitness is often determined and dependent on the fitness of other members of an organism's population or on other interacting species. For example, the selective advantage of a particular tree height depends on the height of surrounding trees. The evolution of a successful immune response to a pathogen will elicit selection pressure leading to enhanced variants of the pathogen. Thus, while optimization theory is useful in describing gradual adaptation, game theoretic explanations have been utilized to describe evolution among interdependent agents. Game theory originated to describe interactions among independent economic agents, and was quickly applied to evolutionary biology. The success of a given strategy in a

game theoretic setting is dependent on the strategies of other players. The biological analog is that the success of a given genome is dependent on the make-up of the population.

The classic example of game theory is the Prisoner's Dilemma.[74] In this game, players receive a benefit if they cooperate (W), but a larger benefit if they act selfishly (X, the so-called "temptation to defect") when the other player cooperates (with Y being the penalty for cooperating when the other player defects). When both players defect, each receives a penalty Z. The Prisoner's Dilemma arises when the payoffs for each action are ranked

$$X > W > Z > Y$$

Thus, strategies that always cooperate can be exploited by strategies that always defect. However, populations of defectors are less fit than populations of cooperators. The Prisoner's Dilemma highlights the difficulty of understanding how cooperation arises from populations of organisms and is captured in the "tragedy of the commons" – how can groups of cooperating individuals resist invasion by selfish individuals? In recent work, MacLean and Gudelj demonstrate mechanisms that can enable cooperators to coexist with selfish individuals: spatial and temporal heterogeneity can allow either coexistence or competitive exclusion, depending on trade-offs in metabolism.[65]

Interestingly, observations of classical Prisoner's Dilemma situations in Nature are lacking. One clear example comes from the work of Turner and Chao on the bacteriophage ϕ-6.[75] The authors were able to measure the payoff matrix associated with cooperation (the manufacture of diffusible products inside the bacterium during infection)

and defection (the sequestration of these products). In this study, selfishness (defection) evolved in viral populations with high multiplicity of infection (MOI, the number of independent viruses infecting a single bacterium) and actually lowered the fitness of the evolved viruses relative to the ancestral virus. Spatial segregation of viral genomes via infection at low MOI evolved clones that showed high fitness and cooperative behavior. Thus, the outcome of simple evolution experiments was heavily dependent on the frequency of other viral genomes in the environment, showing how game theory can complement optimization theory for describing adapting populations.

The fitness of a given network feature may in fact be frequency-dependent: like the success of rare alleles, specific regulatory features may be selected for depending on their relative abundance in a population. An intriguing hypothesis is that regulatory and metabolic architectures are organized to promote cooperation among individuals. Investigations along these lines are only now beginning to be explored.

3.4 Robustness as an organizing principle

One potential recurring design theme in biology may be the organization of cellular networks to be robust to perturbation (both genetic and non-genetic perturbation). Robustness has a long precedent in both engineering disciplines as well as biology. As far back as 50 years ago, Waddington described the robustness of morphological features of organisms to perturbations during development, and ascribed this robustness to natural selection operating to produce organisms that were optimized to some intermediate form

(i.e., stabilizing selection).[76] The idea that robustness is a dominant organizing principle in biological networks is intriguing and continues to be investigated today.

Robustness to non-genetic change

Evolving populations undoubtedly face perturbation and uncertainty from their environments. Organisms have evolved countless mechanisms to maintain function during changes in external (and internal) conditions. For example, the maintenance of osmotic balance, metabolite concentrations, and other homeostatic mechanisms can be considered mechanisms for robustness to non-genetic change, as can thermoregulation in endothermic organisms, flight stabilization in birds, and predator avoidance behavior in higher organisms. The robustness of cellular networks has been explored as well. Barkai and Leibler used a simple two-state model to show how the connectivity of the components of the bacterial chemotaxis network confers robust adaptation.[77] Chemotaxis is a process by which bacteria are able to move toward or away from certain chemicals in the environment by a series of "smooth runs" of motility in one direction punctuated by events of "tumbling" where the direction for a next run is chosen randomly. By adjusting tumbling frequency, the cell is able to bias this random walk behavior and direct its motion toward or away from chemical gradients. Adaptation in this system refers to the invariance of tumbling frequency in environments with homogenous chemical gradients. This adaptation is robust to variations in the environment as well as the levels of chemotaxis regulators and the biochemical parameters of interactions between those

components.[78] This model and further experimental observations show that robustness in adaptation is a direct outcome of the organization of the network.

Recent work has highlighted the importance of robustness to changes in the internal environment of the cell. The presence of noise and the fundamental limits of deterministic behavior at the molecular level suggest that biological systems have evolved to cope with and exploit stochastic behavior in gene expression. Noise in gene expression is a ubiquitous feature of the natural world at the molecular scale and has been demonstrated to arise from the small numbers of molecules involved in the process.[79] Noise intrinsic to gene expression is thought to be dictated by fluctuations in mRNA levels, which may arise from fluctuations in promoter states or the random births and deaths of mRNAs themselves, and has also been shown to result from fluctuations in factors extrinsic to the genes themselves (including pathway specific and global factors of gene expression such as the levels of transcription factors, nucleic acid polymerases, and ribosomes[80]). Noise has been shown to be critical in several biological processes, including determination of competence in *B. subtilis*[81], eye color-vision development in *Drosophila melanogaster*[82], and in viral latency.[83] One future challenge is to demonstrate whether the control of noise is used by biological systems to adapt to specific environments.

Several organizational features of biological networks have been suggested to enable cells to cope with or exploit noise in their internal environment. Negative feedback loops have been demonstrated theoretically and experimentally to reduce noise in gene

expression by damping large fluctuations above the mean.[84] Recently, the Serrano group demonstrated that self-repression reduces noise compared to an unregulated gene, and was effective in reducing both intrinsic and extrinsic noise.[85] The organization of the genome could play a role in reducing gene expression noise as well. Swain showed that genes organized in operons strongly reduced variation between the genes via simulations and analytical derivations.[86] Noise attenuation by operons was effective even in the presence of multiple RBS. Later, experimental work by Tabor and Ellington at the University of Texas at Austin (personal communication) was able to confirm that the organization of genes in operons was an effective means to reduce variation between two genes. Nature, as well as engineers, may be able to utilize genomic organization of this kind to design systems with lower variability between two genes. The importance of these types of noise control with respect to biological function is a central question for future study.

Robustness to mutation

The robustness of the phenotype to underlying mutational change can be central to shaping the emergence of traits and phenotypes in evolution as well as disease, especially in the context of a "genetic capacitor". Mechanisms that confer robustness can act as capacitors for phenotypic change by masking the expression of genetic variation and suddenly revealing this variation when the robustness mechanism is impaired.[87, 88] Recent work in *Candida albicans* has demonstrated how the chaperone Hsp90 acts as a molecular buffering mechanism and can affect the evolution of drug resistance by

enabling novel mutations to have immediate physiological consequences. In addition, the effects of mutant Src kinases can be masked by Hsp90 in mammalian cells, leading to oncogenic phenotypes when Hsp90 is impaired. Thus, Hsp90 can act as a capacitor for phenotypic diversity from genetic variance.

Recent studies have suggested that buffering and genetic capacitance may be a general feature of complex regulatory networks[87]. Several patterns have emerged from studies of gene regulatory networks such as scale-free architecture, small-world structure, and the abundance of regulatory motifs – small "building-blocks" of larger networks. One central question for the engineering of biological systems is if these concepts of robustness can be extended into a set of design principles for specifying the degree of robustness to genetic and non-genetic change. Further work is needed to explore whether such systems-level features in biology are emergent properties of the organization of simpler units (devices and components). In addition, any systems-level property or design themes of regulatory architecture will undoubtedly be shaped by trade-offs between different functions. One example is the trade-off between robustness and fragility in biological and technological systems: the observation that a system robust to one class or type of perturbation is often sensitive or fragile with respect to another. The general occurrence of trade-offs between robustness and fragility [89] has been suggested in diverse complex systems such as the Internet,[90] the immune system,[91] and diseases such as metabolic syndrome [92] and cancer. [93] The "robust yet fragile" nature of biological systems may be a unifying force shaping the evolution of network architecture across the molecular, cellular, organismal, and ecological scales.

Robustness and power-law distributions: a cautionary tale

As discussed above, many biological networks display power-law distributions of connections per node. Albert et al.[94] found that networks with power-law distributions are robust to random perturbations: upon removal of random nodes, the average path length (the distance between any two nodes) changes very little. Power-law networks thus tend to retain overall topology by virtue of alternate paths between nodes, and act to minimize the cascading effects of node removal.[95] These and other theoretical examinations of power-law networks have led to the hypothesis that networks have evolved power-law distributions precisely because of this robustness to node removal and perturbation, and the ubiquity of power-law distributions in biological networks represents a deep and shared evolutionary pressure towards robust function.[95] However, several bodies of evidence contradict this intriguing hypothesis. An examination of networks in biological and physical systems and across spatial scales found that power-law structures exist in a wide spectrum of networks, including networks that have never been under any sort of natural or functional selection[96]. For example, the chemical reaction networks of planetary atmospheres (including Earth, Venus, and Jupiter) found power-law distributions in these systems. The power-law distribution may thus be a general feature of networks (both natural and human-engineered), much like the Gaussian distribution is a general feature of distributions in Nature. While this observation is fascinating, it explains little about the evolution of robustness or the adaptive significance of such global features of cellular network architecture. Thus, the construction and analysis of the

ecological and evolutionary consequences of regulatory mechanisms and architectures is a central area of study for systems and synthetic biology.

4. Open questions: from cellular to ecological networks

Understanding how regulatory architecture affects fitness and how network structure has adapted to cope with noise highlights central themes in systems biology. One current challenge is to understand population genetic parameters (such as fitness and diversity) in terms of cellular network architecture and dynamics, or how metabolic, regulatory, and molecular interaction networks combine to produce the phenotypes observed in Nature. A complementary goal is to understand the adaptive significance of network architecture and how genetic regulation in single cells may shape higher-order ecological interactions.

Open questions and outline of work

Several lines of inquiry at increasing levels of organization (molecular to ecological) will enable researchers to better engineer biology.

(1) How can we use the biochemical and informational functions of nucleic acids to regulate cellular circuits and networks?

(2) How do regulatory circuits regulate trade-offs between cellular functions? Can circuits be engineered to produce new optimal phenotypes, (for example, in terms of fitness in new environments)?

(3) How can ecology-scale interactions between organisms and their environment be understood in terms of regulatory networks? Are there systems-level design principles that suggest adaptation to given ecological strategies?

Progress towards answering these questions should provide insight into how adaptation has shaped the organization of genetic regulatory systems, and illuminate strategies for designing and constructing useful biological technologies as well. Ideally, this and other work will contribute to emerging themes for understanding the emergence of biological function in complex systems, from molecules to organisms.

5. References

1. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet 31, 64-8 (2002).
2. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci U S A 100, 11980-5 (2003).
3. Mangan, S., Zaslaver, A. & Alon, U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. J Mol Biol 334, 197-204 (2003).
4. Dueber, J. E., Yeh, B. J., Chak, K. & Lim, W. A. Reprogramming control of an allosteric signaling switch through modular recombination. Science 301, 1904-8 (2003).
5. Chin, Y. W., Balunas, M. J., Chai, H. B. & Kinghorn, A. D. Drug discovery from natural sources. Aaps J 8, E239-53 (2006).
6. Hughes, E. H. & Shanks, J. V. Metabolic engineering of plants for alkaloid production. Metab Eng 4, 41-8 (2002).
7. Ro, D. K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940-3 (2006).
8. Castagliuolo, I. et al. Engineered E. coli delivers therapeutic genes to the colonic mucosa. Gene Ther 12, 1070-8 (2005).
9. Critchley, R. J. et al. Potential therapeutic applications of recombinant, invasive E. coli. Gene Ther 11, 1224-33 (2004).
10. Anderson, J. C., Clarke, E. J., Arkin, A. P. & Voigt, C. A. Environmentally controlled invasion of cancer cells by engineered bacteria. J Mol Biol 355, 619-27 (2006).
11. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. Nature 434, 1130-4 (2005).
12. Levskaya, A. et al. Synthetic biology: engineering Escherichia coli to see light. Nature 438, 441-2 (2005).
13. Inoue, T., Sullivan, F. X. & Cech, T. R. Intermolecular exon ligation of the rRNA precursor of Tetrahymena: oligonucleotides can function as 5' exons. Cell 43, 431-7 (1985).
14. Tucker, B. J. & Breaker, R. R. Riboswitches as versatile gene control elements. Curr Opin Struct Biol 15, 342-8 (2005).
15. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419, 952-6 (2002).
16. Dykxhoorn, D. M., Novina, C. D. & Sharp, P. A. Killing the messenger: short RNAs that silence gene expression. Nat Rev Mol Cell Biol 4, 457-67 (2003).
17. Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet 39, 673-7 (2007).
18. Saghatelian, A., Yokobayashi, Y., Soltani, K. & Ghadiri, M. R. A chiroselective peptide replicator. Nature 409, 797-801 (2001).

19. Gesteland, R. F., Cech, T. & Atkins, J. F. The RNA world : the nature of modern RNA suggests a prebiotic RNA world (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2006).
20. Joyce, G. F. Forty years of in vitro evolution. Angew Chem Int Ed Engl 46, 6420-36 (2007).
21. Haruna, I. & Spiegelman, S. Autocatalytic synthesis of a viral RNA in vitro. Science 150, 884-6 (1965).
22. Kacian, D. L., Mills, D. R., Kramer, F. R. & Spiegelman, S. A replicating RNA molecule suitable for a detailed analysis of extracellular evolution and replication. Proc Natl Acad Sci U S A 69, 3038-42 (1972).
23. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818-22 (1990).
24. Hermann, T. & Patel, D. J. Adaptive recognition by nucleic acid aptamers. Science 287, 820-5 (2000).
25. Yan, A. C., Bell, K. M., Breeden, M. M. & Ellington, A. D. Aptamers: prospects in therapeutics and biomedicine. Front Biosci 10, 1802-27 (2005).
26. Ng, E. W. et al. Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. Nat Rev Drug Discov 5, 123-32 (2006).
27. Zimmermann, G. R., Wick, C. L., Shields, T. P., Jenison, R. D. & Pardi, A. Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. Rna 6, 659-67 (2000).
28. Lin, C. H. & Patel, D. J. Structural basis of DNA folding and recognition in an AMP-DNA aptamer complex: distinct architectures but common recognition motifs for DNA and RNA aptamers complexed to AMP. Chem Biol 4, 817-32 (1997).
29. Nonin-Lecomte, S., Lin, C. H. & Patel, D. J. Additional hydrogen bonds and base-pair kinetics in the symmetrical AMP-DNA aptamer complex. Biophys J 81, 3422-31 (2001).
30. Jiang, L. & Patel, D. J. Solution structure of the tobramycin-RNA aptamer complex. Nat Struct Biol 5, 769-74 (1998).
31. Tao, J. & Frankel, A. D. Electrostatic interactions modulate the RNA-binding and transactivation specificities of the human immunodeficiency virus and simian immunodeficiency virus Tat proteins. Proc Natl Acad Sci U S A 90, 1571-5 (1993).
32. Puglisi, J. D., Chen, L., Frankel, A. D. & Williamson, J. R. Role of RNA structure in arginine recognition of TAR RNA. Proc Natl Acad Sci U S A 90, 3680-4 (1993).
33. Tan, R. & Frankel, A. D. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. Biochemistry 33, 14579-85 (1994).
34. Das, C., Edgcomb, S. P., Peteranderl, R., Chen, L. & Frankel, A. D. Evidence for conformational flexibility in the Tat-TAR recognition motif of cyclin T1. Virology 318, 306-17 (2004).
35. Bayer, T. S., Booth, L. N., Knudsen, S. M. & Ellington, A. D. Arginine-rich motifs present multiple interfaces for specific binding by RNA. Rna 11, 1848-57 (2005).

36. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428, 281-6 (2004).
37. Fedor, M. J. & Williamson, J. R. The catalytic diversity of RNAs. Nat Rev Mol Cell Biol 6, 399-412 (2005).
38. Robertson, M. P. & Ellington, A. D. Design and optimization of effector-activated ribozyme ligases. Nucleic Acids Res 28, 1751-9 (2000).
39. Breaker, R. R. & Joyce, G. F. A DNA enzyme that cleaves RNA. Chem Biol 1, 223-9 (1994).
40. Tsukiji, S., Pattnaik, S. B. & Suga, H. An alcohol dehydrogenase ribozyme. Nat Struct Biol 10, 713-7 (2003).
41. Doherty, E. A. & Doudna, J. A. Ribozyme structures and mechanisms. Annu Rev Biophys Biomol Struct 30, 457-75 (2001).
42. Robertson, M. P. & Ellington, A. D. In vitro selection of nucleoprotein enzymes. Nat Biotechnol 19, 650-5 (2001).
43. Robertson, M. P. & Scott, W. G. The structural basis of ribozyme-catalyzed RNA assembly. Science 315, 1549-53 (2007).
44. Rusconi, C. P. et al. RNA aptamers as reversible antagonists of coagulation factor IXa. Nature 419, 90-4 (2002).
45. Rusconi, C. P. et al. Antidote-mediated control of an anticoagulant aptamer in vivo. Nat Biotechnol 22, 1423-8 (2004).
46. Lupold, S. E., Hicke, B. J., Lin, Y. & Coffey, D. S. Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. Cancer Res 62, 4029-33 (2002).
47. Chu, T. C. et al. Aptamer:toxin conjugates that specifically target prostate tumor cells. Cancer Res 66, 5989-92 (2006).
48. Chu, T. C., Twu, K. Y., Ellington, A. D. & Levy, M. Aptamer mediated siRNA delivery. Nucleic Acids Res 34, e73 (2006).
49. McNamara, J. O., 2nd et al. Cell type-specific delivery of siRNAs with aptamer-siRNA chimeras. Nat Biotechnol 24, 1005-15 (2006).
50. Shangguan, D. et al. Aptamers evolved from live cells as effective molecular probes for cancer study. Proc Natl Acad Sci U S A 103, 11838-43 (2006).
51. Adleman, L. M. Molecular computation of solutions to combinatorial problems. Science 266, 1021-4 (1994).
52. Stojanovic, M. N. & Stefanovic, D. A deoxyribozyme-based molecular automaton. Nat Biotechnol 21, 1069-74 (2003).
53. Stojanovic, M. N., Mitchell, T. E. & Stefanovic, D. Deoxyribozyme-based logic gates. J Am Chem Soc 124, 3555-61 (2002).
54. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. Science 314, 1585-8 (2006).
55. Isaacs, F. J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol 22, 841-7 (2004).
56. Werstuck, G. & Green, M. R. Controlling gene expression in living cells through small molecule-RNA interactions. Science 282, 296-8 (1998).

57. Desai, S. K. & Gallivan, J. P. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. J Am Chem Soc 126, 13247-54 (2004).
58. Topp, S. & Gallivan, J. P. Guiding bacteria with small molecules and RNA. J Am Chem Soc 129, 6807-11 (2007).
59. Levine, E., Zhang, Z., Kuhlman, T. & Hwa, T. Quantitative characteristics of gene regulation by small RNA. PLoS Biol 5, e229 (2007).
60. Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. Science 295, 1664-9 (2002).
61. El-Samad, H., Kurata, H., Doyle, J. C., Gross, C. A. & Khammash, M. Surviving heat shock: control strategies for robustness and performance. Proc Natl Acad Sci U S A 102, 2736-41 (2005).
62. Kurata, H. et al. Module-based analysis of robustness tradeoffs in the heat shock response system. PLoS Comput Biol 2, e59 (2006).
63. Ghalambor, C. K., Reznick, D. N. & Walker, J. A. Constraints on adaptive evolution: the functional trade-off between reproduction and fast-start swimming performance in the Trinidadian guppy (Poecilia reticulata). Am Nat 164, 38-50 (2004).
64. MacArthur, R. H. & Wilson, E. O. The theory of island biogeography (Princeton University Press, Princeton, N.J.,, 1967).
65. MacLean, R. C. & Gudelj, I. Resource competition and social conflict in experimental populations of yeast. Nature 441, 498-501 (2006).
66. Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. Science 292, 504-7 (2001).
67. De Paepe, M. & Taddei, F. Viruses' Life History: Towards a Mechanistic Basis of a Trade-Off between Survival and Reproduction among Phages. PLoS Biol 4, e193 (2006).
68. Endy, D. Foundations for engineering biology. Nature 438, 449-53 (2005).
69. Dueber, J. E., Mirsky, E. A. & Lim, W. A. Engineering synthetic signaling proteins with ultrasensitive input/output control. Nat Biotechnol 25, 660-2 (2007).
70. Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2, 2006 0028 (2006).
71. Edwards, J. S., Ibarra, R. U. & Palsson, B. O. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 19, 125-30 (2001).
72. Fong, S. S. & Palsson, B. O. Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. Nat Genet 36, 1056-8 (2004).
73. Fischer, E. & Sauer, U. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. Nat Genet 37, 636-40 (2005).
74. Nowak, M. A. & Sigmund, K. Evolutionary dynamics of biological games. Science 303, 793-9 (2004).
75. Turner, P. E. & Chao, L. Prisoner's dilemma in an RNA virus. Nature 398, 441-3 (1999).

76. Waddington, C. H. & Robertson, E. Selection for developmental canalisation. Genet Res 7, 303-12 (1966).
77. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. Nature 387, 913-7 (1997).
78. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. Nature 397, 168-71 (1999).
79. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. Science 309, 2010-3 (2005).
80. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. Nat Genet 31, 69-73 (2002).
81. Suel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440, 545-50 (2006).
82. Samoilov, M. S., Price, G. & Arkin, A. P. From fluctuations to phenotypes: the physiology of noise. Sci STKE 2006, re17 (2006).
83. Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P. & Schaffer, D. V. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122, 169-82 (2005).
84. Hooshangi, S. & Weiss, R. The effect of negative feedback on noise propagation in transcriptional gene networks. Chaos 16, 026108 (2006).
85. Dublanche, Y., Michalodimitrakis, K., Kummerer, N., Foglierini, M. & Serrano, L. Noise in transcription negative feedback loops: simulation and experimental analysis. Mol Syst Biol 2, 41 (2006).
86. Swain, P. S. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. J Mol Biol 344, 965-76 (2004).
87. Bergman, A. & Siegal, M. L. Evolutionary capacitance as a general feature of complex gene networks. Nature 424, 549-52 (2003).
88. Rutherford, S., Hirate, Y. & Swalla, B. J. The Hsp90 capacitor, developmental remodeling, and evolution: the robustness of gene networks and the curious evolvability of metamorphosis. Crit Rev Biochem Mol Biol 42, 355-72 (2007).
89. Doyle, J. & Csete, M. Rules of engagement. Nature 446, 860 (2007).
90. Doyle, J. C. et al. The "robust yet fragile" nature of the Internet. Proc Natl Acad Sci U S A 102, 14497-502 (2005).
91. Kitano, H. & Oda, K. Robustness trade-offs and host-microbial symbiosis in the immune system. Mol Syst Biol 2, 2006 0022 (2006).
92. Kitano, H. et al. Metabolic syndrome and robustness tradeoffs. Diabetes 53 Suppl 3, S6-S15 (2004).
93. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J., 3rd & Doyle, J. Robustness of cellular functions. Cell 118, 675-85 (2004).
94. Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance of complex networks. Nature 406, 378-82 (2000).
95. Kim, B. J., Yoon, C. N., Han, S. K. & Jeong, H. Path finding strategies in scale-free networks. Phys Rev E Stat Nonlin Soft Matter Phys 65, 027103 (2002).
96. Wagner, A. Robustness and evolvability in living systems (Princeton University Press, Princeton, N.J., 2005).

Chapter 2.

Programmable Ligand Controlled Riboregulators of Eukaryotic Gene Expression

Parts reproduced from: Bayer TS and Smolke CD, **Programmable ligand-controlled riboregulators of eukaryotic gene expression.** *Nature Biotechnology*. 23(3):337-43. (2005).

Recent studies have demonstrated the importance of non-coding RNA elements in regulating gene expression networks.[1, 2] We describe the design of a novel class of small trans-acting RNAs that directly regulate gene expression in a ligand-dependent manner. These allosteric riboregulators, which we call antiswitches, are fully tunable and modular by rational design and offer uniquely flexible control strategies by self-regulating to active or inactive forms in response to ligand binding, depending on the platform design. Antiswitches offer "programmable" genetic control and can be tailor-made to control the expression of target transcripts in response to different cellular effectors. Coupled with *in vitro* selection technologies for generating nucleic acid ligand binding species,[3, 4] antiswitches present a powerful platform for designing targeted regulators to program cellular behavior and genetic networks with respect to cellular state and environmental stimuli.

In recent years, *cis* and *trans* RNA elements have become well recognized as important regulators of gene expression. Cells use diverse non-coding RNA-based elements to regulate complex genetic networks such as those involved in developmental timing and circadian clocks.[1, 2] Antisense RNAs are small trans-acting RNAs (taRNAs) that bind to complementary segments of a target messenger RNA (mRNA) and regulate gene expression through mechanisms such as targeting decay, blocking translation, and altering splicing patterns.[5-7] MicroRNAs (miRNAs), small taRNAs that affect either translation or RNA decay by interacting with complementary sequences in mRNA and the genome, are likely widespread in metazoan gene regulation.[8] Small interfering RNAs (siRNAs) and double-stranded RNAs (dsRNAs) are able to precisely target mRNAs and inhibit their expression through the RNA interference (RNAi) pathway in metazoans, and are thought to be part of the cell's host defense system.[9] Ribozymes are RNA molecules exhibiting catalytic function and have been shown to be used by viruses to regulate gene expression.[10] Riboswitches, cis-acting metabolite binding structures in mRNAs, control gene expression by modulating translation initiation, disruption of transcriptional termination, or cleavage of mRNA by ribozyme mechanisms.[11-13] Recent studies have demonstrated the prevalence of these RNA-based regulators across diverse groups of organisms from prokaryotes to humans.[14-16]

Researchers have taken advantage of the relative ease with which RNA libraries can be generated and searched to create synthetic RNA-based molecules with novel functional properties. Aptamers are nucleic acid binding species that interact with high

affinity and specificity to selected ligands. These molecules are generated through iterative cycles of selection and amplification known as *in vitro* selection or systematic evolution of ligands by exponential enrichment (SELEX).[3, 4] Aptamers have been selected to bind diverse targets such as dyes, proteins, peptides, aromatic small molecules, antibiotics, and other biomolecules.[17] High-throughput methods and laboratory automation have been developed to generate aptamers in a rapid and parallel manner[18]. Researchers have demonstrated that aptamers can impart allosteric control properties onto other functional RNA molecules. Such allosteric control strategies have been employed to construct and select *in vitro* signaling aptamers, *in vitro* sensors, and *in vitro* allosterically controlled ribozymes.[19-21]

In addition to the widespread occurrence of RNA-based regulator elements in natural systems, researchers have recently described engineered riboregulator systems. Cis-acting RNA elements were described that regulate relative expression levels in *Escherichia coli* from a two gene transcript by controlling RNA processing and decay.[22] In another example, a combined cis/trans riboregulator system was described in *E. coli* in which cis-acting RNA elements mask the ribosome binding site of a transcript, thereby inhibiting translation, and trans-activating RNAs bind to the cis-acting elements to allow translation.[23] Cis-acting elements were recently described that control gene expression in mammalian cells and mice by acting through RNA cleavage and whose activity can be regulated by a small molecule drug and antisense oligonucleotides.[24] Finally, an allosteric aptamer construct was recently described that upon binding the dye tetramethylrosamine, interacts with protein-based transcriptional activators to induce transcription.[25]

Riboregulators present powerful tools for flexible genetic regulation. However, there is a need to couple the ability of RNA-based regulators that can directly target transcripts with allosteric control typically associated with protein-based regulators. We have engineered ligand responsive riboregulators in *Saccharomyces cerevisiae*. These riboregulators, which we call antiswitches, utilize an antisense domain to control gene expression[6] and an aptamer domain to recognize specific effector ligands. Ligand binding at the aptamer domain mediates a change in the conformational dynamics of these molecules that allows the antisense domain to interact with a target mRNA to affect translation. Antiswitches act as programmable genetic switches, affecting target transcripts only in the presence of a specific ligand. We have developed a modular, tunable class of small RNAs that can be used to achieve sensor-based gene expression control. Because antiswitches are designed on a modular platform, in principle these riboregulators can be tailor-made to regulate the expression of any target transcript in response to any ligand.

Antisense technologies have been widely utilized to regulate gene expression.[26, 27] We sought to engineer allosteric regulatory functionality by designing a platform on which ligand binding structures were appended to the antisense molecule. In this platform, the antisense domain is sequestered in an "antisense stem" in the absence of ligand. Ligand binding to the aptamer domain mediates a change in the conformational dynamics of the antisense stem that results in the antisense domain being in a more single-stranded form (**Fig. 1a**). Such mechanisms have been described in the construction of signaling aptamers and other allosterically-controlled RNAs.[28]

We constructed an initial antiswitch, s1, using a previously selected aptamer that binds the xanthine derivative theophylline with high affinity (K_d = 0.29 μM) and specificity.[29] The antisense RNA domain is designed to base pair with a 15 nucleotide region around the start codon of a target mRNA encoding green fluorescent protein (GFP). The stem of the theophylline aptamer is redesigned so that the antisense portion base pairs in a stable stem, the antisense stem, in the absence of ligand, but so that another, overlapping stem forms upon ligand binding, the "aptamer stem", forcing the antisense portion into a more single-stranded state (**Fig. 1a, b**). The aptamer stem and antisense stem are designed such that the antisense stem is slightly more stable than the aptamer stem. Previous work has demonstrated that the sequence of the lower theophylline aptamer stem is not critical for ligand binding,[30] and this sequence was altered to interact with the antisense stem upon ligand binding. It is anticipated that these molecules will function through alterations in conformational dynamics, such that in the absence of ligand and presence of target transcript, the stem sequestering the antisense is more likely to form; whereas in the presence of both ligand and target transcript, the free energy associated with binding of theophylline (approximately 8.9 kcal/mol[31]) and RNA stabilization in the aptamer structure enables the aptamer stem to form, freeing the antisense domain to bind its target transcript. RNAstructure[32] was used to predict the stability of the RNA secondary structures formed. Due to the dual-stem design of the antiswitch, it is anticipated that the free energies of the aptamer binding to its ligand and the antisense binding to its target mRNA will contribute in a cooperative manner to the structural switching of the antiswitch molecule.

The expression of antiswitches in *S. cerevisiae* was accomplished using a novel non-coding RNA (ncRNA) expression construct similar to a previously described system[33] (**Fig. 1c**). Briefly, the RNA to be expressed is cloned between two hammerhead ribozymes known to self-cleave *in vivo.*[34] This dual hammerhead construct can be placed under the control of Pol II promoters, and when transcribed the flanking hammerhead ribozymes cleave out from the desired RNA at an efficiency greater than 99% (**Table 1**). The construct enables creation of ncRNAs with defined 5' and 3' ends that are free of potentially interfering flanking sequences. Antiswitch s1 was expressed in this construct under control of a galactose-inducible (GAL1) promoter in yeast cells. A plasmid containing a yeast enhanced GFP (yEGFP)[35] under the control of a GAL1 promoter was transformed into the same cells (**Fig. 1a**).

Results from protein expression assays demonstrate ligand specific *in vivo* activity of s1 (**Fig. 2a**). Expression of antiswitch s1 in the absence of theophylline decreases GFP expression from control levels by approximately 30%, where addition of greater than 0.8 mM theophylline decreases expression to background levels. The antisense and aptamer domains were expressed separately as controls and had expected effects on GFP expression levels. It is interesting to note the rapid change in expression levels between 0.75 mM and 0.8 mM theophylline. The antiswitch s1 displays binary, on/off behavior rather than linearly modulating expression over a range of theophylline concentrations. This response supports the anticipated cooperative mechanism of structural switching dependent on both ligand and target mRNA. It has been previously demonstrated that the aptamer used in this antiswitch does not bind caffeine,[29] which differs from theophylline by a single methyl group. The addition of caffeine does not change expression levels

from those of an inactive switch, demonstrating that specific ligand-aptamer interactions are necessary to activate the antiswitch and free the antisense domain to decrease gene expression of GFP.

Quantitative real-time PCR (qRT-PCR) was performed on antiswitch s1 and target mRNA extracted from cells grown under different conditions to determine relative RNA levels (**Table 1**). Relative levels of target transcript did not change significantly between cells harboring s1 grown in the absence and high levels of theophylline, indicating that antiswitches function through translational inhibition rather than affecting target RNA levels. In addition, the steady-state relative level of s1 was approximately 1,000-fold that of target levels, although both antiswitch and target were expressed from the same promoter. This indicates that antiswitch molecules may have higher intracellular stabilities than mRNA potentially due to stabilizing secondary structures or are synthesized more efficiently. The temporal response of antiswitch regulation was determined by inducing antiswitch activation by the addition of theophylline to cells expressing steady-state levels of GFP and s1 in the "off" state (**Fig. 2b**). GFP levels began decreasing shortly after the addition of theophylline at a rate corresponding to a half-life of approximately .5 to 1 hour, which is consistent with the half-life of the GFP variant used in these experiments.[35] This data supports that antiswitch molecules act rapidly to inhibit translation from their target mRNAs in the presence of activating levels of effector and that the time required for target protein levels to decrease is determined by the protein's half-life.

In vitro characterization studies were conducted to examine antiswitch ligand affinity and conformational changes associated with antiswitch response. Gel shift experiments were conducted in the presence of equimolar amounts of a short target transcript (200 nucleotides), containing regions upstream and downstream of the start codon, and labeled s1 and varying concentrations of theophylline to examine antiswitch ligand affinity (**Fig. 2c**). A sharp shift in antiswitch mobility is detected between 2 and 10 μM theophylline, presumably due to binding of both theophylline and target. Nuclease mapping in the presence of ligand alone was also conducted to investigate antiswitch conformational changes (**Fig. 2d**). This data supports that antiswitch molecules exhibit conformational changes at much higher concentrations of ligand than in the presence of ligand and target (between 200 μM and 2 mM versus 2 μM and 10 μM), supporting the cooperative effects of ligand and target on antiswitch conformational dynamics. The *in vivo* data report the concentration of effector molecule in the media and it is anticipated that the intracellular concentration of these molecules will be much lower due to transport limitations across the membrane. One study reported over a 1,000-fold drop in theophylline concentration across the *E. coli* membrane[36]. The *in vitro* experiments indicate that ligand binding and structural switching occur over narrow concentration ranges, much lower than the extracellular concentrations reported in the *in vivo* studies. This data indicates that in the presence of target *in vitro* antiswitch conformational changes display a sharp binary response to ligand concentrations in the low micromolar range, which is probably indicative of the intracellular concentrations of theophylline in these studies.

The switching behavior of the antiswitch platform is dependent on conformational dynamics of the RNA structures; therefore it is possible to tune switching behavior in a straightforward manner by altering thermodynamic properties of the antiswitch. It is anticipated that the absolute and relative stabilities of the antisense stem and the aptamer stem will be important design parameters in tuning the switch behavior of an antiswitch. To explore the dynamic range of switch behavior, we created several antiswitches (s2–s4) with varying antisense and aptamer stem stabilities (**Fig. 3a**). It was anticipated that these altered antiswitches would expand the concentration range over which the switch in gene expression was observed and increase the dynamic range of GFP expression.

In general, it was observed that increasing antisense stem stability by the addition of base pairs created switches that required higher concentrations of theophylline to affect a switch, whereas decreasing stem stabilities created switches that inhibit GFP expression at lower theophylline concentrations. For example, antiswitch s2 differs from antiswitch s1 by a single nucleotide (A21 to C) (**Fig. 3a**). This mutation introduces a mismatched pair in the antisense stem so that in the absence of ligand, the construct is less thermodynamically stable. As a result, s2 exhibits altered switching dynamics: theophylline concentrations greater than 0.2 mM inhibit gene expression, compared to 0.8 mM for construct s1 (**Fig. 3b**). Alternately, increasing the stability of the antisense stem creates a switch that requires higher concentrations of theophylline to inhibit expression. Antiswitch s3 is designed with an antisense stem five nucleotides longer than s1 and an aptamer stem with 3 bp of the lower stem formed, increasing the absolute stem stabilities. As a result of this increased stability, s3 switches from GFP expression to inhibition of

GFP at approximately 1.25 mM theophylline (**Fig. 3b**), roughly 1.5-fold the concentration required to switch s1 and 6-fold of that required to switch s2. Furthermore, s3 exhibits higher levels of GFP expression in the "off" state, 10% versus 30% inhibition from full expression. Antiswitch s4 was constructed to examine the effects of further destabilizing the antisense stem. This antiswitch includes an altered loop sequence (U18 to C), which further destabilizes the antisense stem from s2. Assays indicate that s4 further expands the dynamic switching behavior of the antiswitch construct, exhibiting switching at 0.1 mM theophylline (**Fig. 3b**).

To demonstrate the modularity of the antiswitch design platform, we constructed and characterized several different antiswitch molecules by swapping in different aptamer domains (**Fig. 4**). These changes in the aptamer domain were designed to keep the antisense stem and the switching aptamer stem identical to previous designs since the target transcript was kept the same, while swapping out the remainder of the aptamer module. To further explore the range of ligand responsiveness in designed antiswitches, we constructed a switch s5 employing a previously characterized aptamer exhibiting lower affinity to theophylline.[29] This aptamer has a K_d approximately ten-fold higher than the aptamer used in s1–s4. In addition, the response of this antiswitch was tuned by destabilizing the antisense stem in a manner identical to s2, creating s6. To further test the modularity of this platform, an antiswitch was also constructed with a previously characterized aptamer to tetracycline.[37] This aptamer has an affinity to tetracycline similar to that of the theophylline aptamer used in s1–4 (K_d = 1 _M).

The data in **Figure 3** support the modularity of the antiswitch platform to different aptamer domains. The modified theophylline aptamers exhibit an altered response to ligand concentrations from s1–4. As expected, the switching for s5 and s6 occurs at higher theophylline concentrations (**Fig. 5a**). Significantly, s5, which contains an aptamer domain with a 10-fold higher K_d than the aptamer domain in s1, switches at approximately a 10-fold higher theophylline concentration. In addition, the tetracycline antiswitch s7 shows similar switch dynamics as s1–4, suggesting that the response curve observed is a general feature of designed antiswitches (**Fig. 5b**).

To further examine the flexibility of the antiswitch platform, we redesigned the platform in an attempt to construct an "on" antiswitch from the aptamer and antisense domains used in the design of s1. An antiswitch s8 that inhibits expression in the absence of theophylline, but allows expression in the presence of theophylline, was constructed using similar design principles. This switch displays its antisense domain in the absence of ligand, leaving it free to interact with the target mRNA, while sequestering the antisense in the aptamer stem when ligand is present (**Fig. 6a**). s8 displays similar dynamic behavior to s1 (switching around 1 mM theophylline), as is expected due to similar base pairing energetics **(Fig. 6b).** This functional "on" switch demonstrates the flexibility of the antiswitch platform and the generality of the design themes.

The modular nature of the antiswitch platform allows for systems exhibiting combinatorial control over gene expression. To illustrate this, we introduced into cells two switches each responsive to a different effector molecule and each regulating the protein expression of a different mRNA target: s1, a theophylline responsive GFP

regulator, and s9 (see **Fig. 7a** online), a tetracycline responsive yellow fluorescent variant protein (Venus)[38] regulator (**Fig. 7b**). Changes in the targeting capabilities of these molecules were made by swapping out the antisense stem and switching aptamer stem while keeping the remainder of the aptamer module the same. Concurrent expression of these two antiswitches with a plasmid carrying both GFP and Venus allowed for an assay of the simultaneous regulation of gene expression by modular antiswitch design. As shown in **Fig.7c**, addition of theophylline decreased expression of GFP, while Venus expression remained unaffected and addition of tetracycline decreased Venus while not affecting GFP. Furthermore, the addition of both ligands decreased expression of both GFP and Venus. This simple system illustrates the potential of building more complex genetic circuits that are precisely regulated by multiple antiswitch constructs.

This work demonstrates that engineered, ligand controlled antisense RNAs, or antiswitches, are powerful, allosteric regulators of gene expression. The general design of an antiswitch is based on conformational dynamics of RNA folding to create a dual stem molecule comprised of an antisense stem and an aptamer stem. These stems are designed such that in the absence of ligand, the free energy of the antisense stem is lower than that of the aptamer stem. Ligand and target act cooperatively to alter the conformational dynamics of these molecules and stabilize the formation of the aptamer stem and the binding of the antisense domain to its target transcript. The antiswitch platform is flexible, enabling both positive and negative regulation. The "on" switch is designed using the same energetics on an altered platform such that in the absence or low levels of ligand, the antisense domain is free to bind to the target; however, ligand binding changes

the conformational dynamics of these molecules so that the antisense domain is bound in the aptamer stem.

The switching dynamics of antiswitch regulators are amenable to tuning by forward engineering design strategies based on thermodynamic properties of RNA. Altering the free energy of the antisense domain alters the conformational dynamics of these molecules in a predictable fashion. Specifically, decreasing the stability of the antisense stem decreases the ligand concentration necessary to switch the antiswitch conformation, and increasing the stability of the antisense stem increases the ligand concentration necessary to switch the conformation as well as shifts the dynamics to favor the "off" state at low ligand levels.

In addition, the antiswitch platform is fully modular, enabling ligand response and transcript targeting to be engineered by swapping domains within the antiswitch molecule. The ligand detection capability of antiswitches is designed separately from the targeting capability by swapping only the aptamer domain. Likewise, the targeting capability of these molecules can be designed separately from the ligand detection capability by swapping both the antisense stem and the switching aptamer stem to recognize a different target sequence, while not affecting the aptamer domain.

Antiswitch molecules are novel, RNA-based, allosteric regulators of gene expression that can potentially function across a diverse range of organisms, from prokaryotes to humans, making them extremely useful in many different applications. Their design provides a foundation upon which to build other ligand controlled riboregulators for different systems. This type of allosteric riboregulator presents a

powerful tool for gene therapy applications, where one would like to target specific transcripts in response to specific cellular environments that are indicative of a diseased state[39]. One can also anticipate exogenously delivered antiswitches acting as therapeutic molecules, similar to exogenously delivered antisense oligonucleotides, thereby extending the functionality of current antisense therapies by introducing cell-specific action to an already highly targeted therapy. Antiswitch technology can be used to engineer novel regulatory pathways and control loops for applications in metabolic engineering[40] and synthetic circuit design[41] by enabling the cell to sense and respond to intracellular metabolite levels and environmental signals. Finally, antiswitches present new tools for cellular imaging, measuring, and detection strategies, enabling programmable concentration-specific detection of intracellular molecules. Antiswitches offer a unique platform to create tailor-made cellular sensors and "smart" regulators that potentially can target any gene in response to any target ligand, creating new avenues for cellular control and engineering.

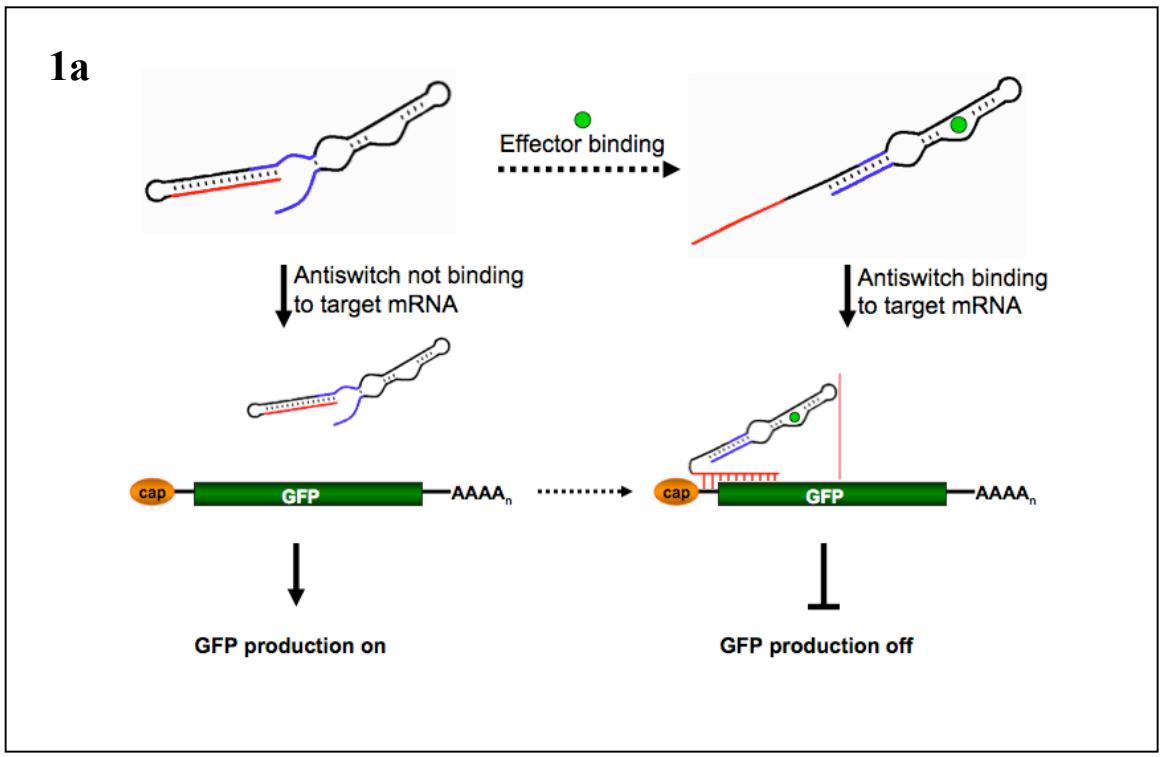
References

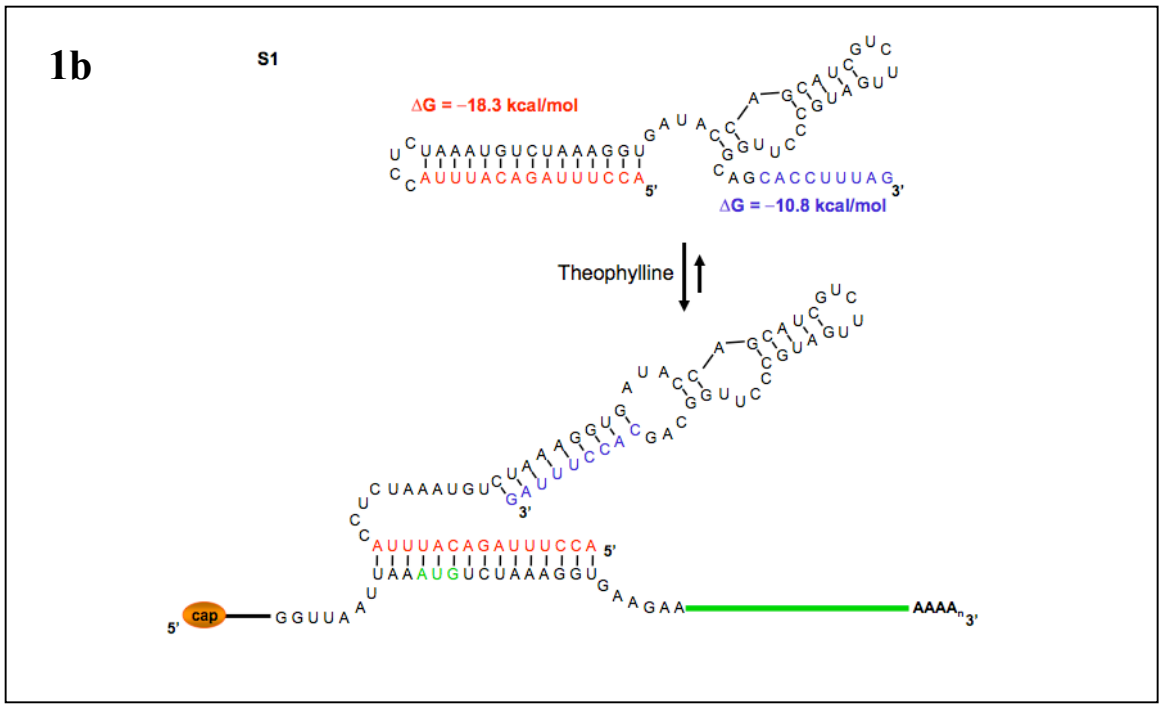
1. Banerjee, D. & Slack, F. Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. Bioessays. **24**, 119-129 (2002).
2. Kramer, C., Loros, J.J., Dunlap, J.C. & Crosthwaite, S.K. Role for antisense RNA in regulating circadian clock function in Neurospora crassa. Nature. **421**, 948-952 (2003).
3. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. Nature. **346**, 818-822 (1990).
4. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science. **249**, 505-510 (1990).
5. Good, L. Diverse antisense mechanisms and applications. Cell. Mol. Life Sci. **60**, 823-824 (2003).
6. Good, L. Translation repression by antisense sequences. Cell. Mol. Life Sci. **60**, 854-861 (2003).

7. Vacek, M., Sazani, P. & Kole, R. Antisense-mediated redirection of mRNA splicing. Cell. Mol. Life Sci. **60**, 825-833 (2003).
8. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. **116**, 281-297 (2004).
9. Scherer, L. & Rossi, J.J. Recent applications of RNAi in mammalian systems. Curr. Pharm. Biotechnol. **5**, 355-360 (2004).
10. Lilley, D.M. The origins of RNA catalysis in ribozymes. Trends Biochem. Sci. **28**, 495-501 (2003).
11. Mandal, M. & Breaker, R.R. Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat. Struct. Mol. Biol. **11**, 29-35 (2004).
12. Winkler, W., Nahvi, A. & Breaker, R.R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature. **419**, 952-956 (2002).
13. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. & Breaker, R.R. Control of gene expression by a natural metabolite-responsive ribozyme. Nature. **428**, 281-286 (2004).
14. Barrick, J.E. et al. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc. Natl. Acad. Sci. USA. **101**, 6421-6426 (2004).
15. Yelin, R. et al. Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. **21**, 379-386 (2003).
16. Lavorgna, G. et al. In search of antisense. Trends Biochem. Sci. **29**, 88-94 (2004).
17. Hermann, T. & Patel, D.J. Adaptive recognition by nucleic acid aptamers. Science. **287**, 820-825 (2000).
18. Cox, J.C. et al. Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. Nuc. Acids Res. **30**, e108 (2002).
19. Jhaveri, S., Rajendran, M. & Ellington, A.D. In vitro selection of signaling aptamers. Nat. Biotechnol. **18**, 1293-1297 (2000).
20. Roth, A. & Breaker, R.R. Selection in vitro of allosteric ribozymes. Methods Mol. Biol. **252**, 145-164 (2004).
21. Stojanovic, M.N. & Kolpashchikov, D.M. Modular aptameric sensors. J. Am. Chem. Soc. **126**, 9266-9270 (2004).
22. Smolke, C.D., Carrier, T.A. & Keasling, J.D. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. Appl Environ Microbiol. **66**, 5399-5405 (2000).
23. Isaacs, F.J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. Nat. Biotechnol. **22**, 841-847 (2004).
24. Yen, L. et al. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. Nature. **431**, 471-476 (2004).
25. Buskirk, A.R., Landrigan, A. & Liu, D.R. Engineering a ligand-dependent RNA transcriptional activator. Chem. Biol. **11**, 1157-1163 (2004).
26. Weiss, B., Davidkova, G. & Zhou, L.W. Antisense RNA gene therapy for studying and modulating biological processes. Cell. Mol. Life Sci. **55**, 334-358 (1999).
27. Scherer, L.J. & Rossi, J.J. Approaches for the sequence-specific knockdown of mRNA. Nat. Biotechnol. **21**, 1457-1465 (2003).
28. Nutiu, R. & Li, Y. Structure-switching signaling aptamers. J. Am. Chem. Soc. **125**, 4771-4778 (2003).

29. Zimmermann, G.R., Wick, C.L., Shields, T.P., Jenison, R.D. & Pardi, A. Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. RNA. **6**, 659-667 (2000).
30. Zimmermann, G.R., Jenison, R.D., Wick, C.L., Simorre, J.P. & Pardi, A. Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. Nat. Struct. Biol. **4**, 644-649 (1997).
31. Gouda, H., Kuntz, I.D., Case, D.A. & Kollman, P.A. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. Biopolymers. **68**, 16-34 (2003).
32. Mathews, D.H. et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. USA. **101**, 7287-7292 (2004).
33. Taira, K., Nakagawa, K., Nishikawa, S. & Furukawa, K. Construction of a novel RNA-transcript-trimming plasmid which can be used both in vitro in place of run-off and (G)-free transcriptions and in vivo as multi-sequences transcription vectors. Nuc. Acids Res. **19**, 5125-5130 (1991).
34. Samarsky, D.A. et al. A small nucleolar RNA:ribozyme hybrid cleaves a nucleolar RNA target in vivo with near-perfect efficiency. Proc. Natl. Acad. Sci. USA. **96**, 6609-6614 (1999).
35. Mateus, C. & Avery, S.V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. Yeast. **16**, 1313-1323 (2000).
36. Koch, A.L. The metabolism of methylpurines by Escherichia coli. I. Tracer studies. J. Biol. Chem. **219**, 181-188 (1956).
37. Berens, C., Thain, A. & Schroeder, R. A tetracycline-binding RNA aptamer. Bioorg. Med. Chem. **9**, 2549-2556 (2001).
38. Nagai, T. et al. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat. Biotechnol. **20**, 87-90 (2002).
39. Watkins, S.M. & German, J.B. Metabolomics and biochemical profiling in drug discovery and development. Curr. Opin. Mol. Ther. **4**, 224-228 (2002).
40. Khosla, C. & Keasling, J.D. Metabolic engineering for drug discovery and development. Nat. Rev. Drug Discov. **2**, 1019-1025 (2003).
41. Kobayashi, H. et al. Programmable cells: interfacing natural and engineered gene networks. Proc. Natl. Acad. Sci. USA. **101**, 8414-8419 (2004).
42. Sambrook, J. & Russell, D.W. Molecular cloning: A laboratory manual, Edn. 3rd. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; 2001).
43. Sikorski, R.S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics. **122**, 19-27 (1989).
44. Gietz, R. & Woods, R. in Guide to Yeast Genetics and Molecular and Cell Biology, Part B, Vol. 350. (eds. C. Guthrie & G. Fink) 87-96 (Academic Press, San Diego; 2002).

45. Caponigro, G., Muhlrad, D. & Parker, R. A small segment of the MAT alpha 1 transcript promotes mRNA decay in Saccharomyces cerevisiae: a stimulatory role for rare codons. Mol. Cell. Biol. **13**, 5141-5148 (1993).


1a
Effector binding
Antiswitch not binding to target mRNA
Antiswitch binding to target mRNA
cap
GFP
AAAA$_n$
GFP production on
GFP production off


1b
S1
ΔG = −18.3 kcal/mol
ΔG = −10.8 kcal/mol
Theophylline
cap
AAAA$_n$

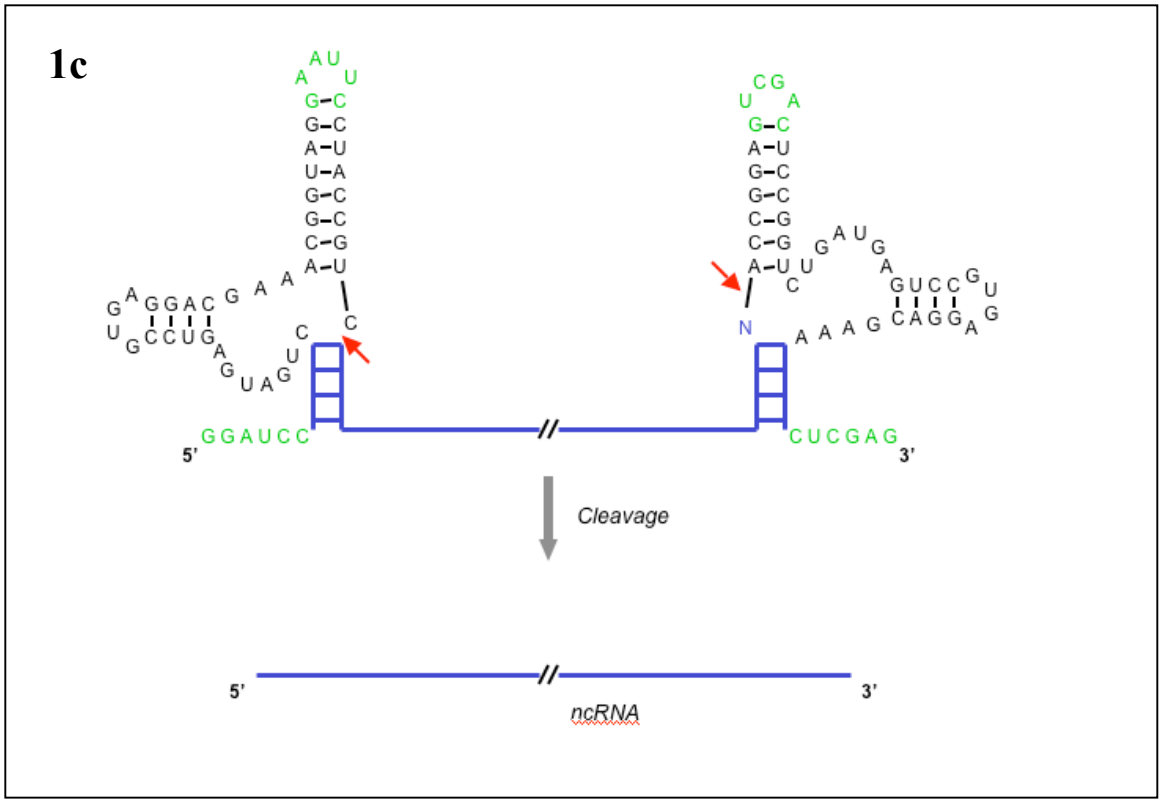
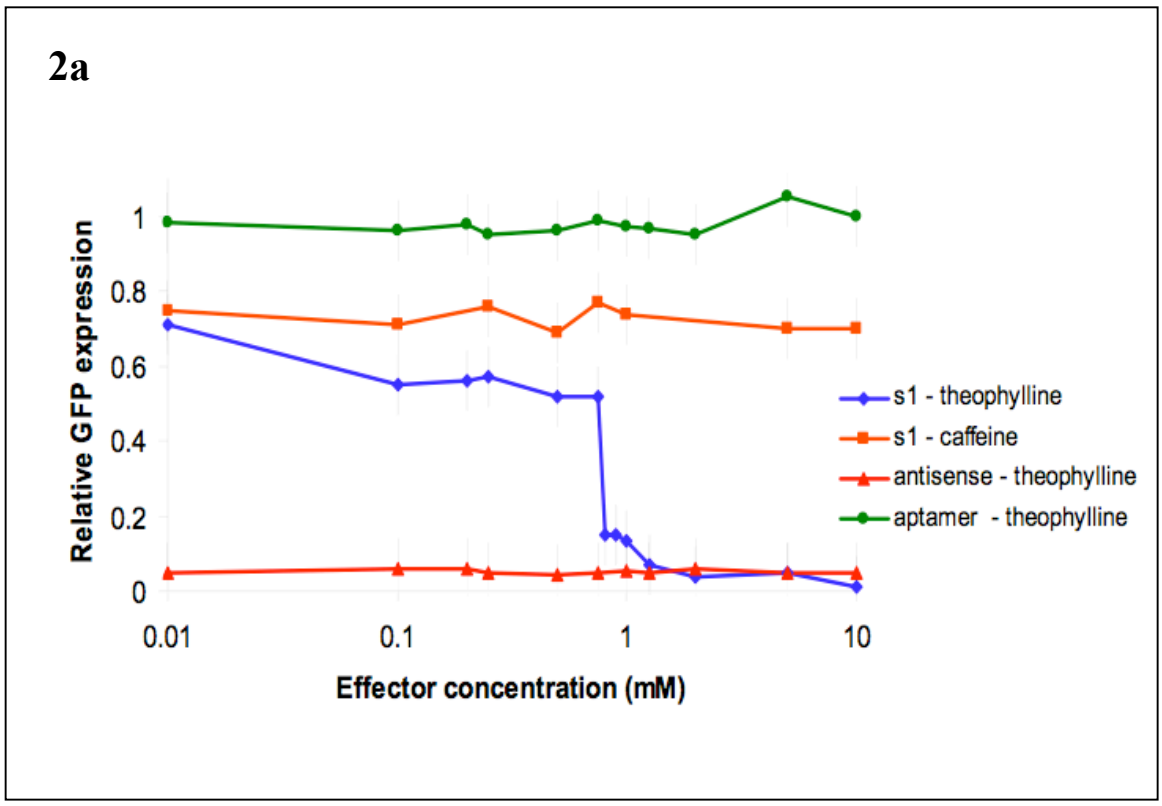
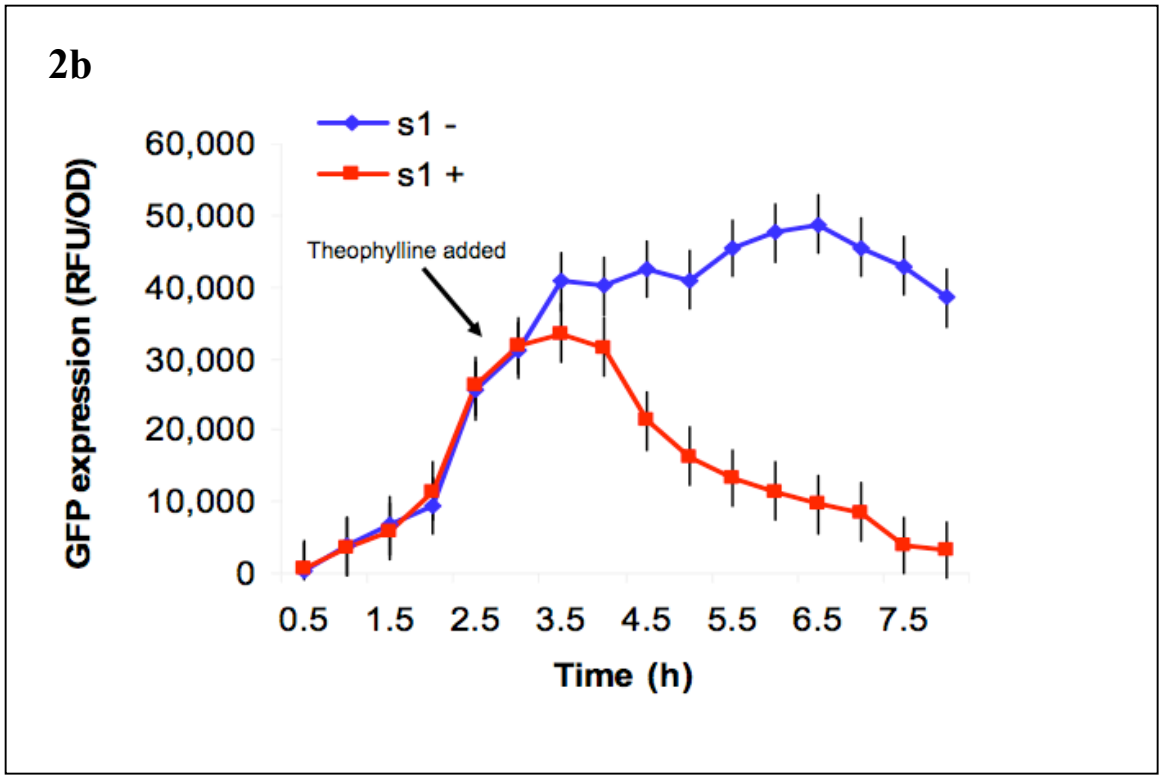


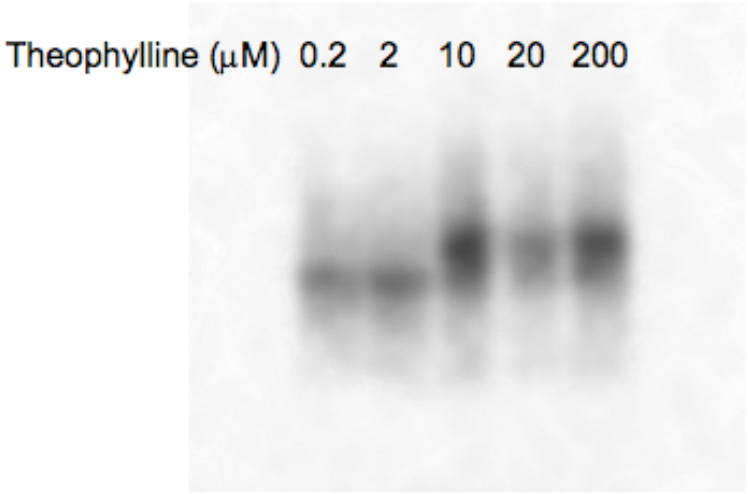
Figure 1 Design and function of a novel antiswitch regulator. (a) General illustration of the mechanism by which an antiswitch molecule acts to regulate gene expression *in vivo*. The antisense sequence is indicated in red; switching "aptamer stem" is shown in blue. In the absence of effector, the antisense domain is bound in a double-stranded region of the RNA referred to as the "antisense stem" and the antiswitch is in the "off" state. In this state the antiswitch is unable to bind to its target transcript, which encodes a *gfp* coding region, and as a result, GFP production is on. In the presence of effector, the antiswitch binds the molecule, forcing the aptamer stem to form, switching its confirmation to the "on" state. In this state the antisense domain of the antiswitch will bind to its target transcript and through an antisense mechanism turn the production of GFP off. (**b**) Sequence and predicted structural switching of a theophylline-responsive

antiswitch, s1, and its target mRNA. On s1, the antisense sequence is indicated in red; switching aptamer stem sequence is indicated in blue, the stability of each switching stem is indicated. On the target mRNA, the start codon is indicated in green. (**c**) Sequence and cleavage mechanism of ncRNA expression construct. The expression construct enables cloning of general sequences between two hammerhead ribozyme sequences through unique restriction sites *Bam*HI, *Eco*RI, *Sal*I, and *Xho*I (indicated in green). Predicted cleavage sites are indicated by red arrows. General ncRNA insert is indicated by a blue line or lettering. Following cleavage, the resulting ncRNA has defined 3' and 5' ends.

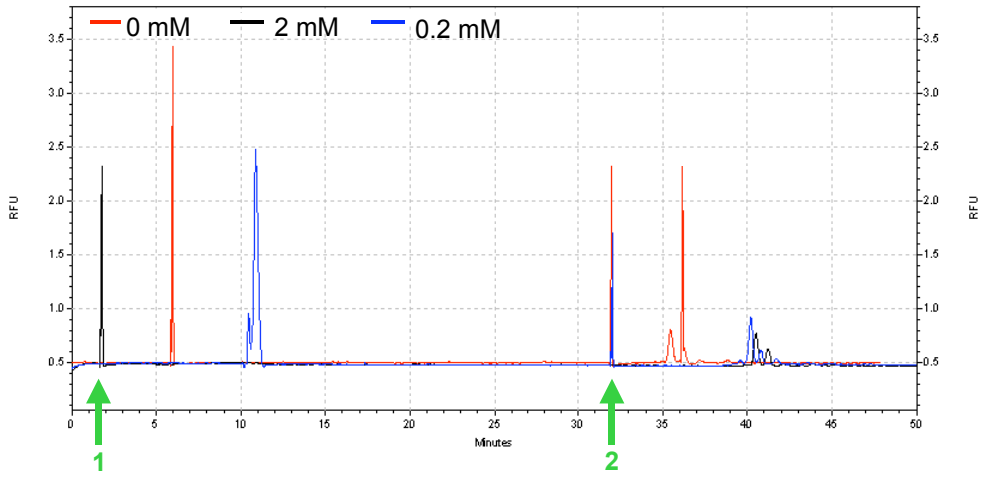

2a
Relative GFP expression
1
0.8
0.6
0.4
0.2
0
0.01
0.1
1
10
Effector concentration (mM)
s1 - theophylline
s1 - caffeine
antisense - theophylline
aptamer - theophylline


2b
GFP expression (RFU/OD)
60,000
50,000
40,000
30,000
20,000
10,000
0
0.5
1.5
2.5
3.5
4.5
5.5
6.5
7.5
Time (h)
s1 -
s1 +
Theophylline added

2c


Theophylline (μM) 0.2 2 10 20 200

2d


0 mM
2 mM
0.2 mM
RFU
Minutes
1
2

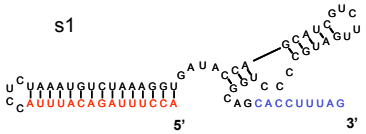
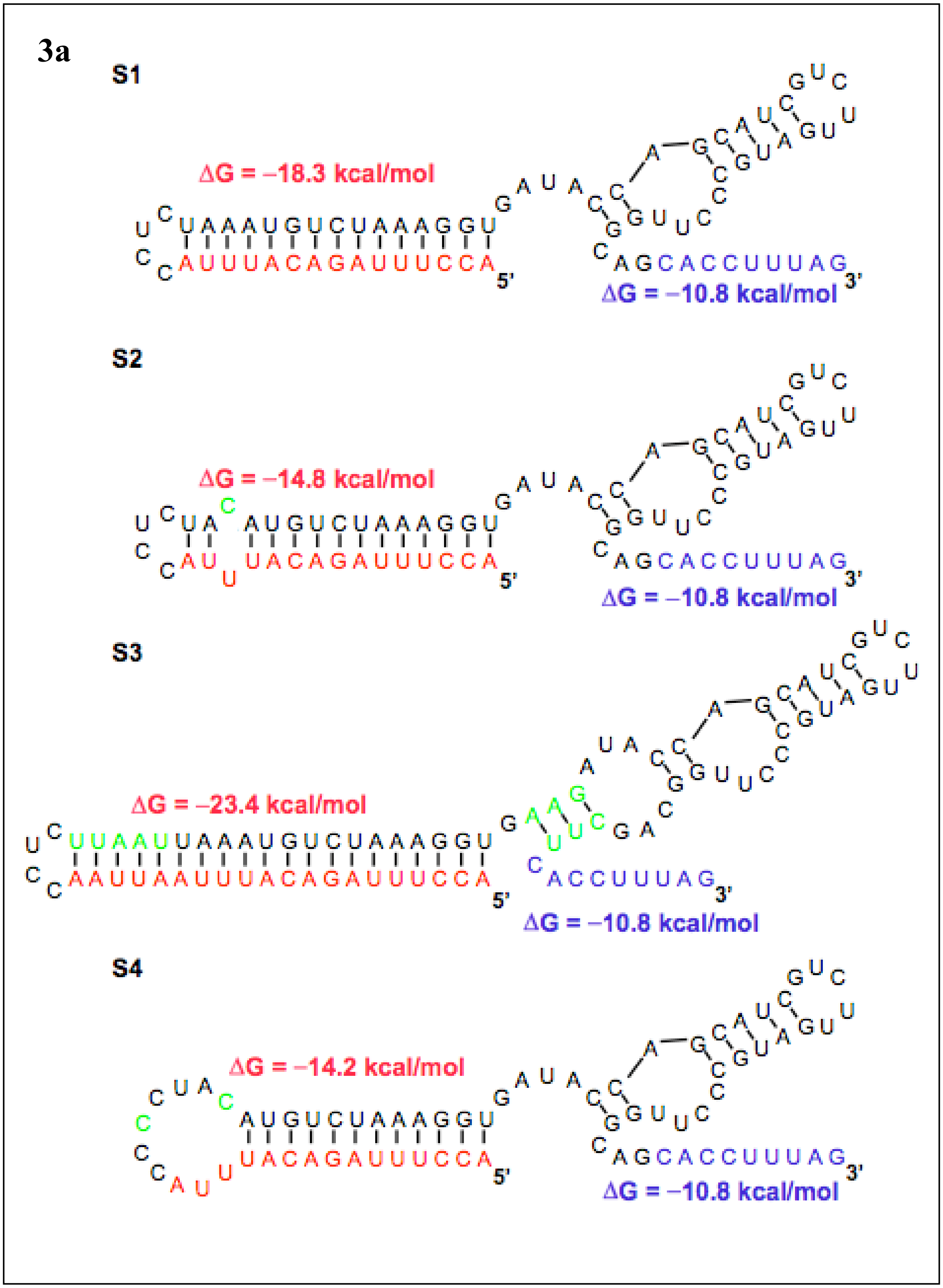

s1
5'
3'

Figure 2 (**a**) *In vivo* GFP regulation activity of s1 and controls across different effector concentrations: aptamer construct (negative control) in the presence of theophylline (green); antisense construct (positive control) in the presence of theophylline (red); s1 in the presence of caffeine (negative control, orange); s1 in the presence of theophylline (blue). Data is presented as relative, normalized GFP expression in cells harboring these constructs against expression levels from induced and uninduced cells harboring only the GFP expression construct. **(b)** *In vivo* temporal response of s1 inhibiting GFP expression upon addition of effector to cells that have accumulated steady-state levels of GFP and antiswitch s1: no theophylline (blue); 2 mM theophylline (red). **(c)** *In vitro* affinity assays of s1 to target and effector molecules. The mobility of radiolabeled s1 was monitored in the presence of equimolar concentrations of target transcript and varying concentrations of theophylline as indicated. (**d**) Structural probing of antiswitch s1 through nuclease mapping. Samples correspond to fluorescently labeled s1 incubated in the presence of RNase T1 and varying concentrations of theophylline. Fragments generated by RNase T1 cleavage 3' of single-stranded G's were separated by capillary electrophoresis. Peak 1 corresponds to the antisense domain, and peak 2 corresponds to the switching aptamer stem. In both the absence of theophylline and 200 μM theophylline, the switching aptamer stem is cleaved (peak 2), indicating that this domain is in a single-stranded form, accessible to the nuclease. In 2 mM theophylline this peak is absent, indicating that the aptamer stem is protected in a double-stranded stem. Furthermore, in 2 mM theophylline the disappearance of peak 2 occurs simultaneously with the appearance of peak 1, indicating that the antisense domain is in a single-stranded form accessible to the

nuclease. This peak is not present in lower levels of theophylline, supporting a change in accessibility of this region of the antiswitch under these concentrations. Unlabeled peaks between 1 and 2 correspond to cleavage within the region connecting the antisense and aptamer stems. Peaks after 2 correspond to full-length constructs.


3a
S1
ΔG = –18.3 kcal/mol
ΔG = –10.8 kcal/mol
S2
ΔG = –14.8 kcal/mol
ΔG = –10.8 kcal/mol
S3
ΔG = –23.4 kcal/mol
ΔG = –10.8 kcal/mol
S4
ΔG = –14.2 kcal/mol
ΔG = –10.8 kcal/mol

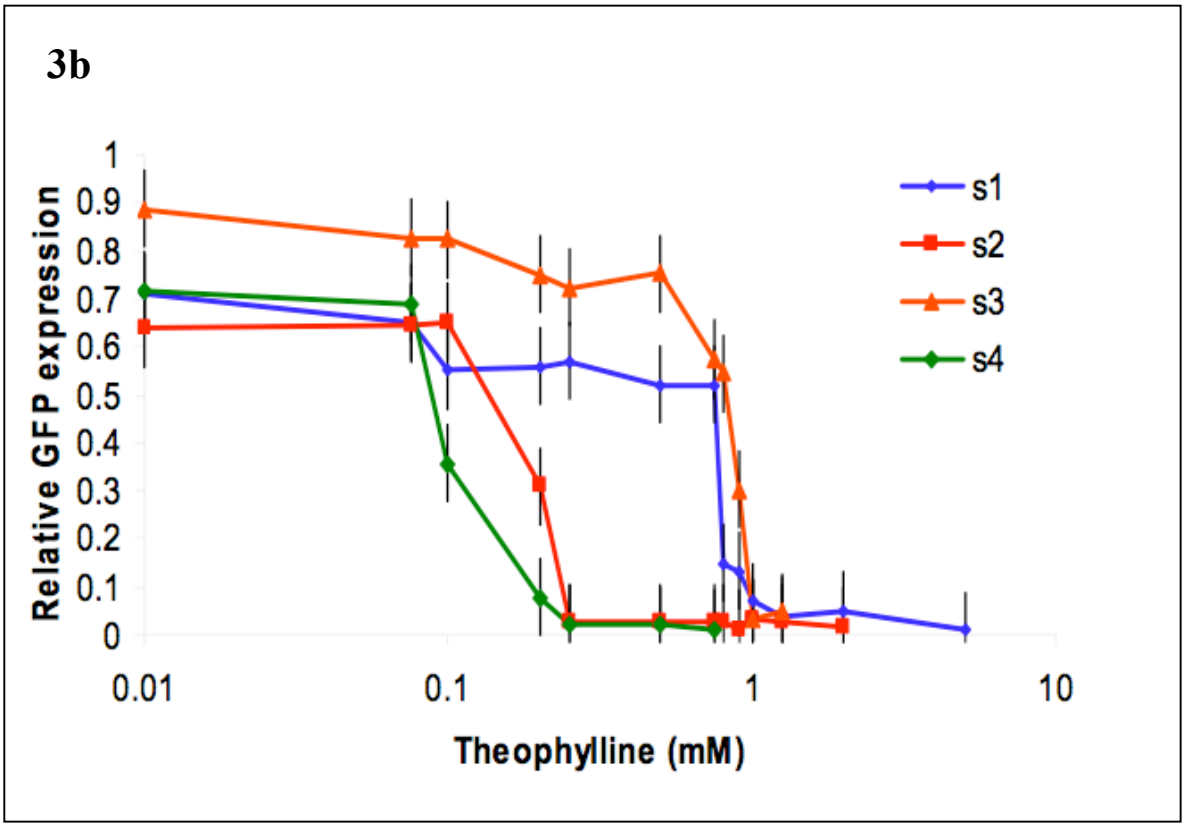


Figure 3 Tuning and expanding the switch response of an antiswitch regulator. (**a**) Predicted structures of tuned antiswitches (s2–s4), based on s1, in the absence of theophylline binding. The antisense sequences are indicated in red; switching aptamer stem sequences are indicated in blue; modified sequences are indicated in green, the stability of each switching stem is indicated. (**b**) *In vivo* GFP regulation activity of s1s4 across different theophylline concentrations: s1- initial antiswitch construct (blue); s2- destabilized antiswitch construct (red); s3- stabilized antiswitch construct (orange); s4- destabilized antiswitch construct (green).

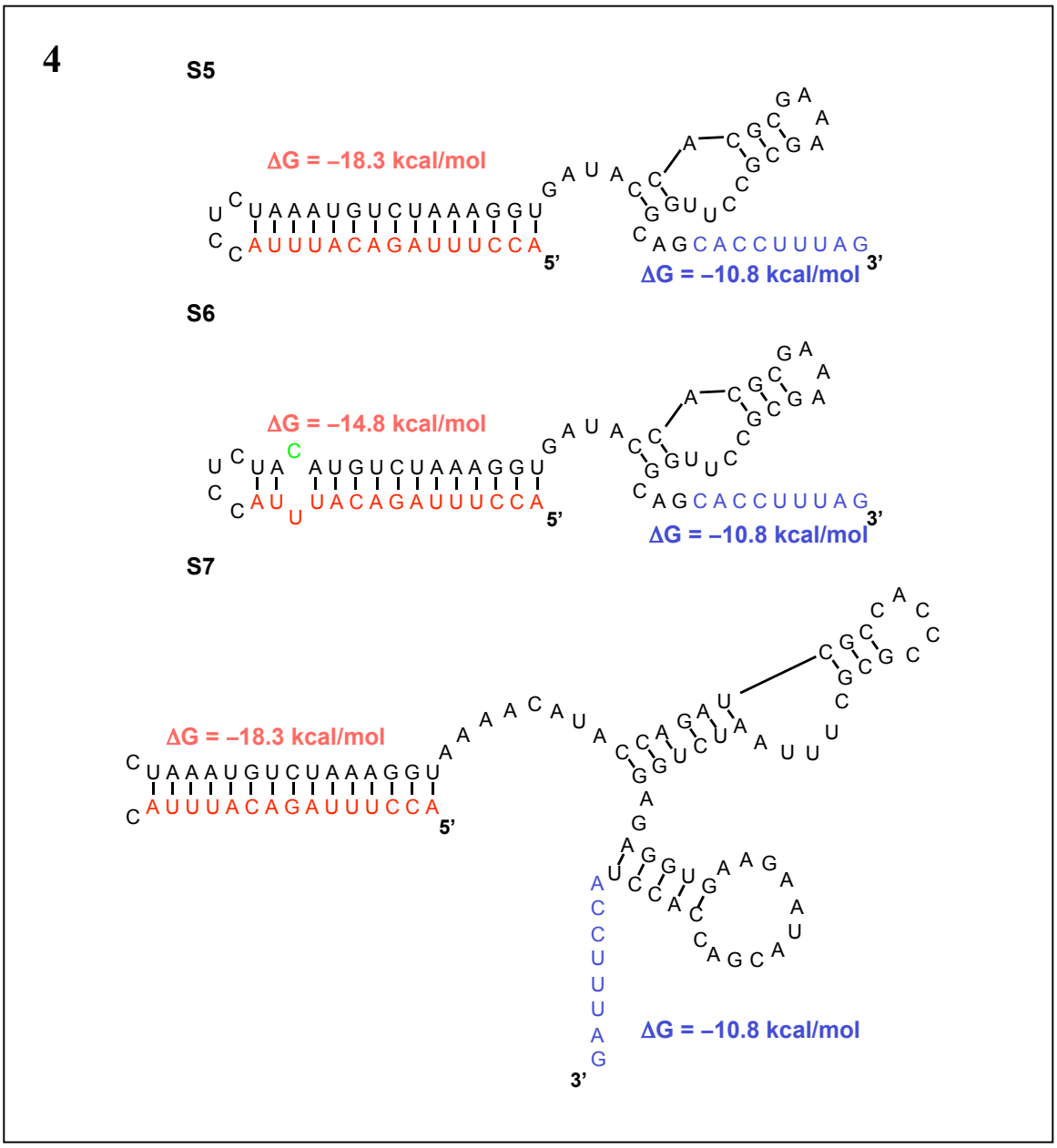
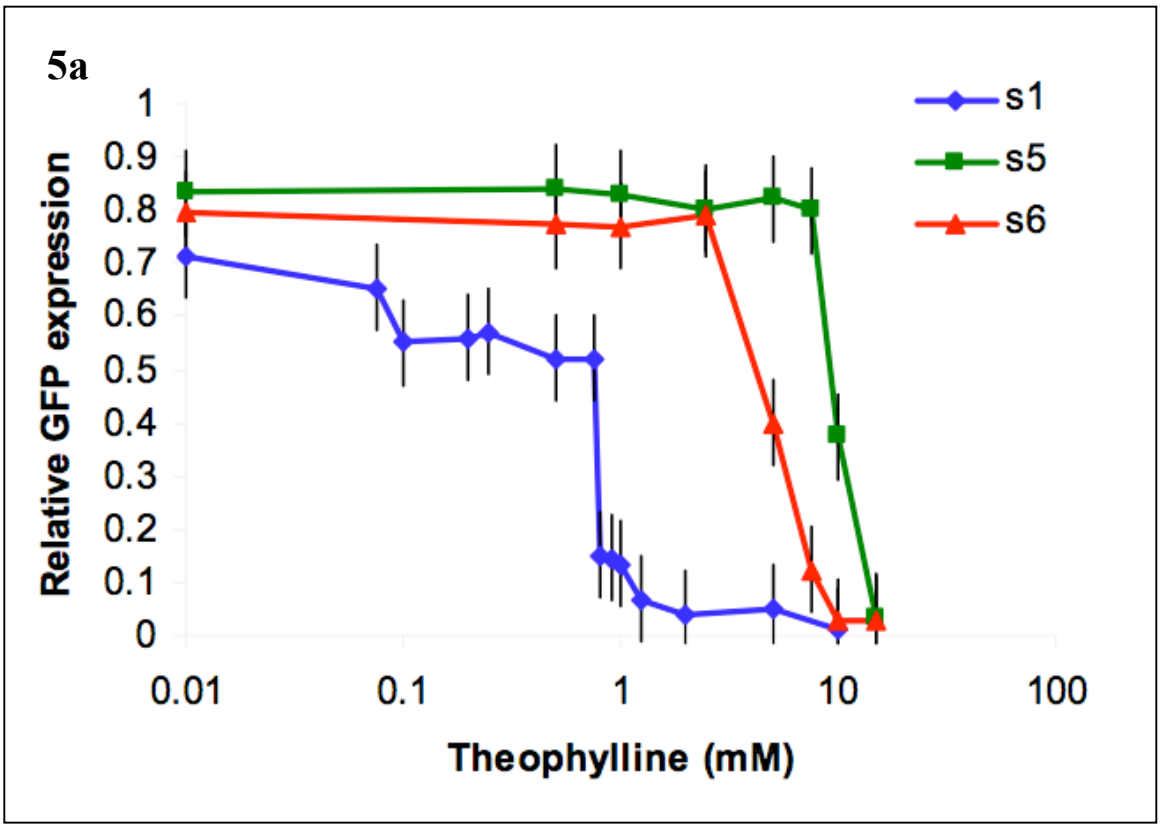


Figure 4. Sequences and predicted structures of antiswitches s5, s6, and s7 in the absence of ligand binding. The antisense sequences are indicated in red; switching aptamer stem sequences are indicated in blue; modified sequences are indicated in green; the stability of each switching stem is indicated: s5- modified theophylline aptamer antiswitch based on s1; s6- destabilized modified theophylline aptamer antiswitch; s7- tetracycline aptamer antiswitch based on s1.


5a
Relative GFP expression
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.01
0.1
1
10
100
Theophylline (mM)
s1
s5
s6

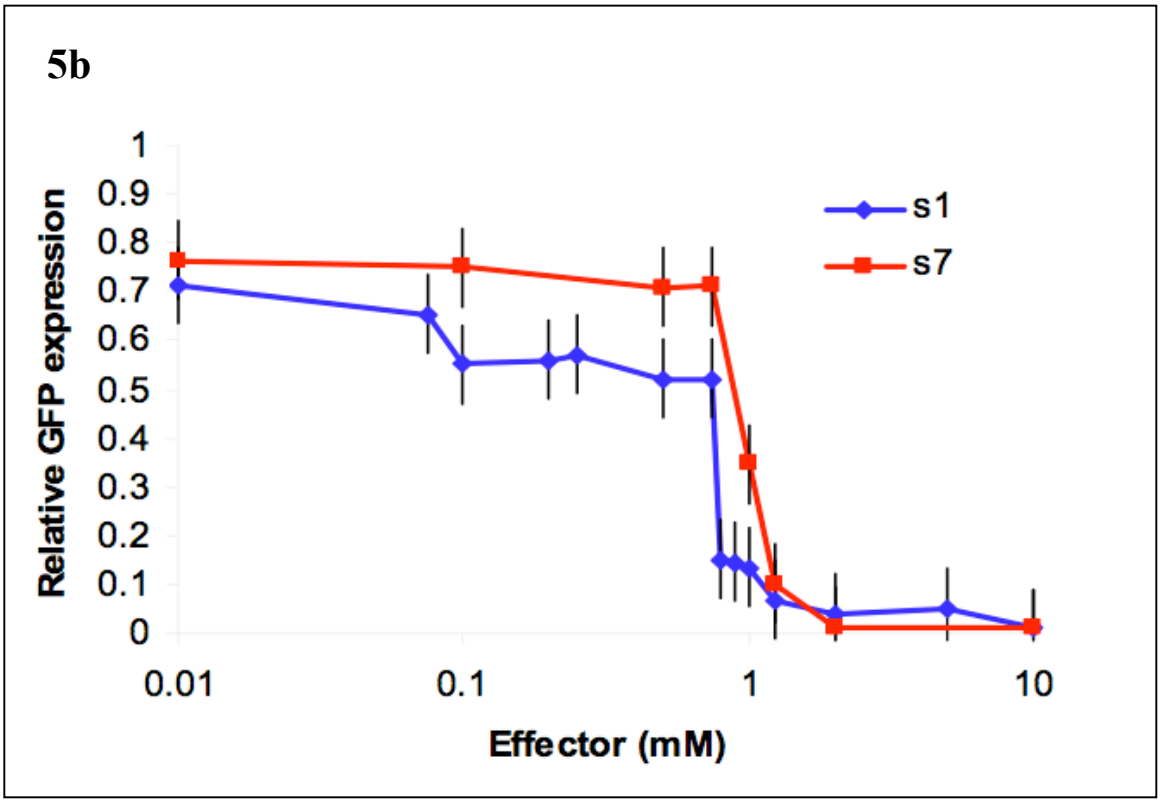
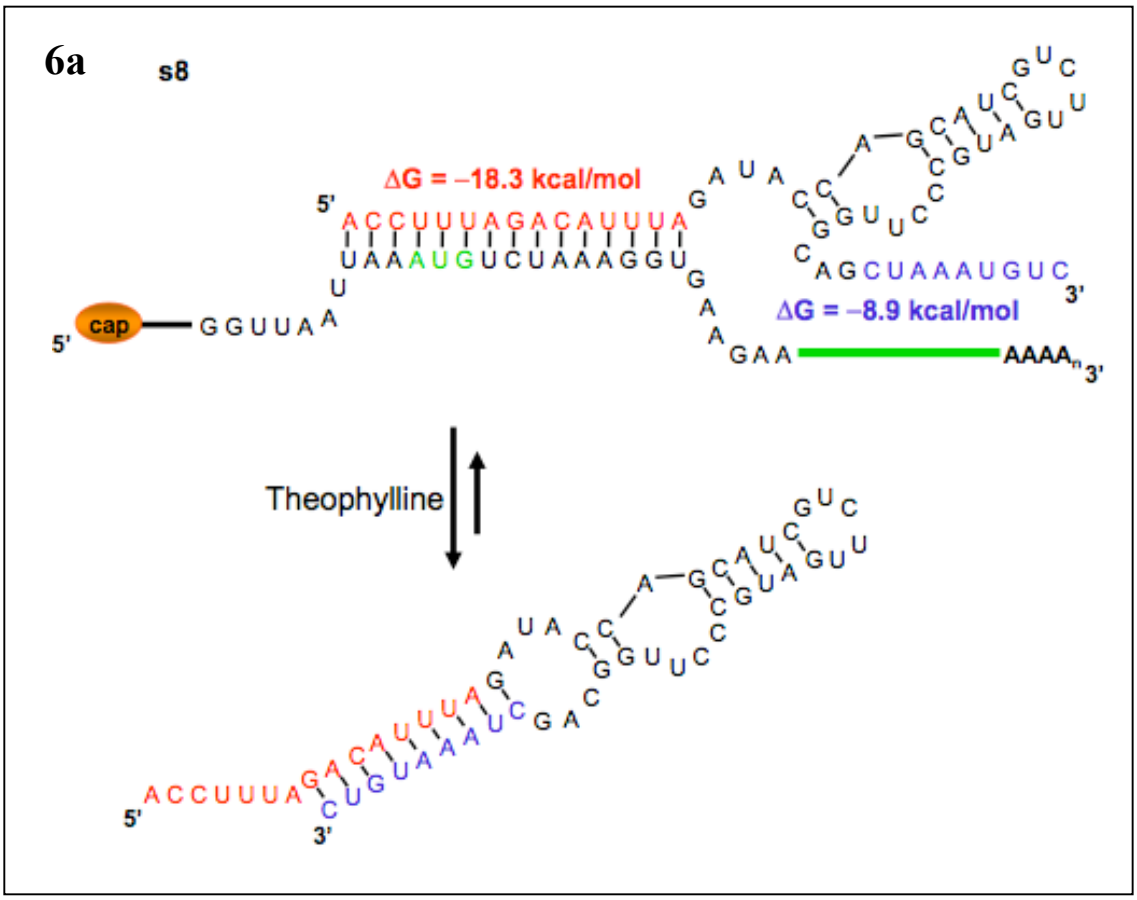

5b
Relative GFP expression
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.01
0.1
1
10
Effector (mM)
s1
s7

Figure 5. (a) *In vivo* GFP regulation activity of modified aptamer-antiswitch constructs (s5–s6) across different theophylline concentrations: s1- initial antiswitch construct (blue); s5- antiswitch construct with an aptamer domain having 10-fold lower affinity to theophylline than that used in s1 (green); s6- destabilized modified aptamer-antiswitch construct, based on s5 (red). **(b)** *In vivo* GFP regulation activity of antiswitch constructs responsive to different small molecule effectors (s1, s7) across different effector concentrations: s1- initial antiswitch construct responsive to theophylline (blue); s7- antiswitch construct modified with a tetracycline aptamer domain, based on s1, responsive to tetracycline (red).


6a
s8
ΔG = –18.3 kcal/mol
5'
ACCUUUAGACAUUUA
UAAAUGUCUAAAGGU
U
cap
GGUUAA
5'
GAA
AAAA
n
3'
ΔG = –8.9 kcal/mol
AGCUAAAUGUC
3'
Theophylline
ACCUUUA
5'
3'

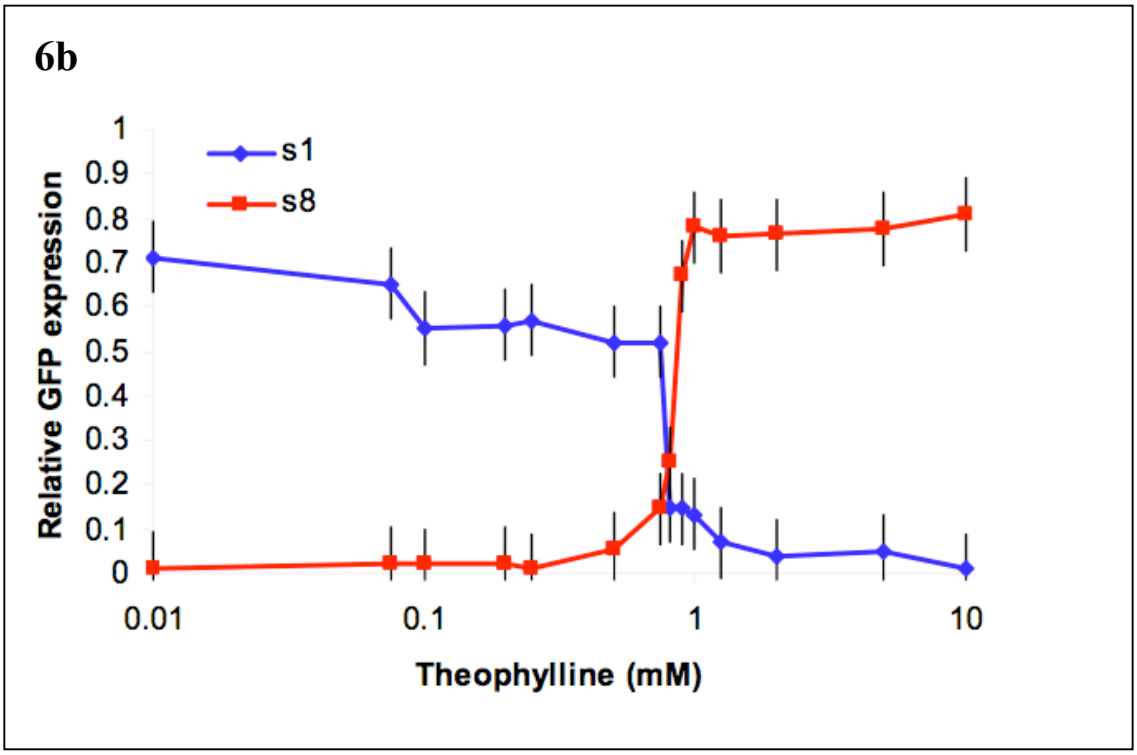
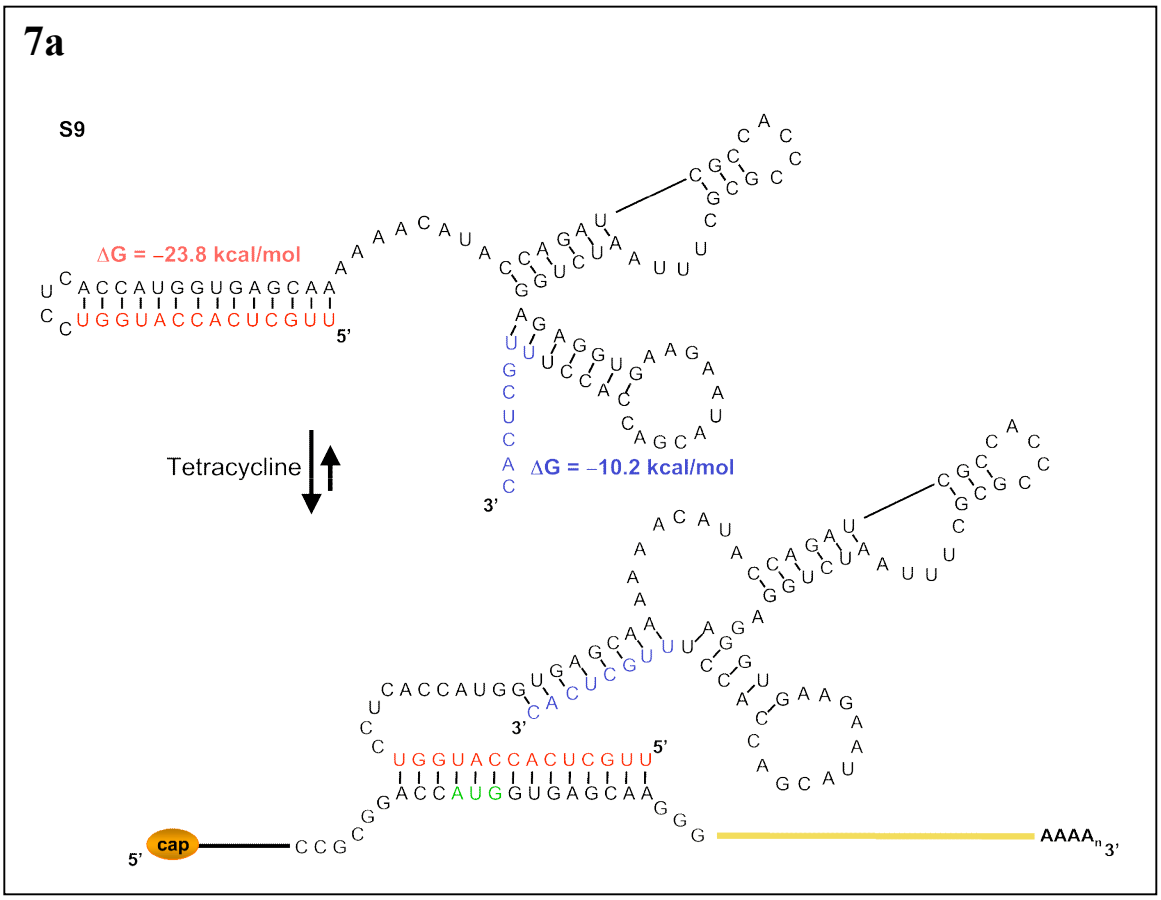

6b
Relative GFP expression
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
s1
s8
0.01
0.1
1
10
Theophylline (mM)

Figure 6 Redesign and characterization of a novel "on" antiswitch regulator. **(a)** Sequence and structural switching of an "on" antiswitch regulator (s8) responsive to theophylline. The antisense sequence is indicated in red; switching aptamer stem sequence is indicated in blue; the stability of each switching stem is indicated. On the target mRNA, the start codon is indicated in green. s8 is designed such that in the absence of theophylline, the antiswitch is "on" or the antisense domain is free to bind to its target. In the presence of theophylline, the antiswitch undergoes a conformational change to the "off" state such that the antisense domain is bound in a double-stranded RNA stem that is part of the aptamer stem. (**b)** *In vivo* GFP regulation activity of "on" and "off" antiswitch constructs across different theophylline concentrations: s1- initial 'off' antiswitch construct (blue); s8- redesigned "on" antiswitch construct, based on s1 (red).


7a
S9
ΔG = –23.8 kcal/mol
Tetracycline
ΔG = –10.2 kcal/mol
cap
AAAA$_n$
7b
Addition of theophylline
Addition of tetracycline
cap
GFP
AAAA$_n$
cap
YFP
AAAA$_n$
GFP production off
YFP production off

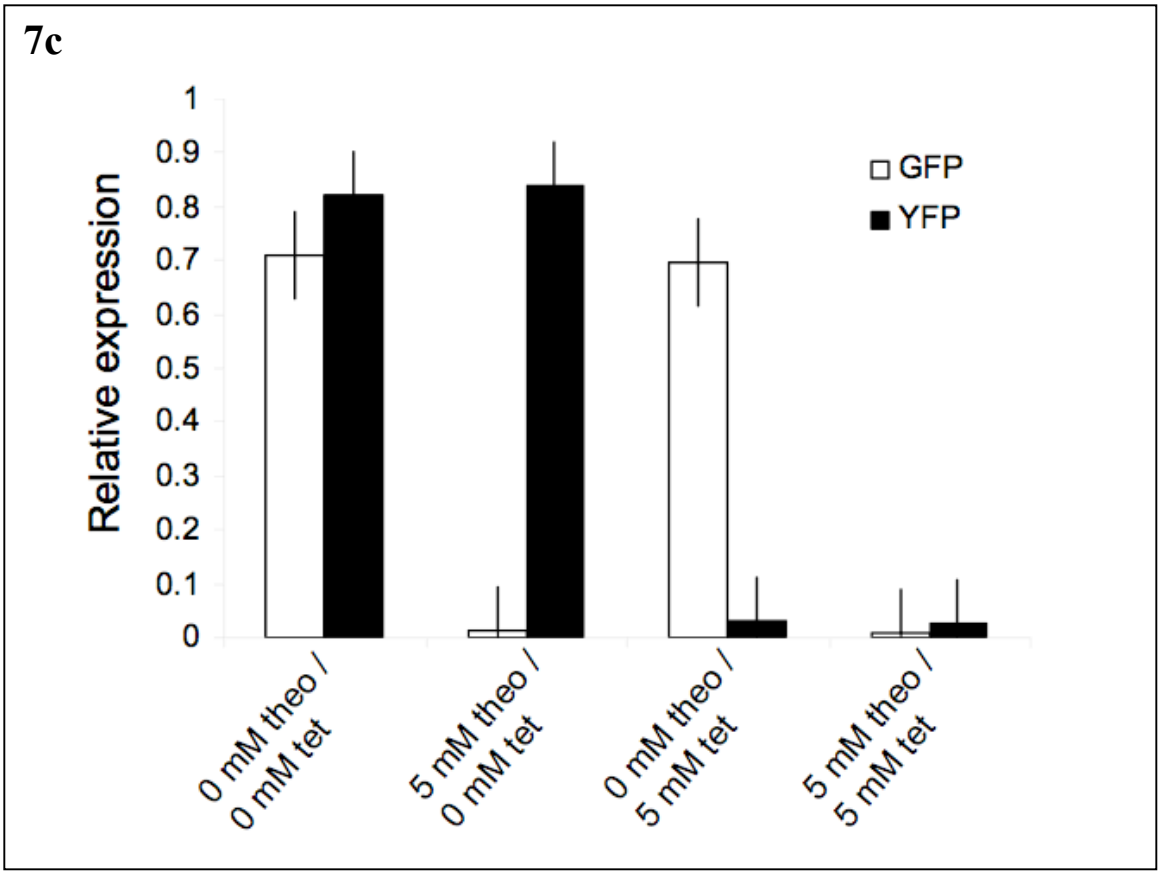


Figure 7. (**a**) Sequence and structural switching of a tetracycline-responsive Venus (YFP) regulator, s9, and its target mRNA. On s9, the antisense sequence is indicated in red; switching aptamer stem sequence is indicated in blue, the stability of switching stem is indicated. On the target mRNA, the start codon is indicated in green. (**b)** Illustration of the mechanism by which two independent antiswitch molecules act to regulate the expression of multiple target genes *in vivo*. In the absence of their respective effectors, the antiswitches are in the "off" state and are unable to bind to their target transcripts. In this state, both GFP and YFP production is on. In the presence of theophylline, one antiswitch switches its conformation to the "on" state and turns off GFP production. In the presence of tetracycline, the second antiswitch switches its conformation to the "on"

state and turns off YFP production. These antiswitches act independently of each other to provide combinatorial control over genetic circuits. **(c)** *In vivo* regulation activity of two antiswitch constructs (s1, s9) against their respective targets (GFP, YFP) in the presence or absence of their respective effector molecules (theophylline, tetracycline). Relative YFP expression (black); relative GFP expression (white).

Table 1 Relative RNA levels of target mRNA and antiswitch s1. Relative levels are normalized to GFP mRNA levels in the absence of theophylline.

RNA	0 mM theophylline	2 mM theophylline
GFP mRNA	1±0.048	1.1±0.052
Antiswitch s1	990±46.2	971±47.1
Uncleaved hammerhead	0.158±0.009	0.149±0.008

Antiswitch construct	RNA sequence
s1	ACCUUUAGACAUUUACCUCUAAAUGUCUAAAGGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCUUUAG
s2	ACCUUUAGACAUUUACCUCUACAUGUCUAAAGGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCUUUAG
s3	ACCUUUAGACAUUUAAUUAACCUCUUAAUUAAAUGUCUAAAGGUGAAGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCUUCACCUUUAG
s4	ACCUUUAGACAUUUACCCCUACAUGUCUAAAGGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCUUUAG
s5	ACCUUUAGACAUUUACCUCUAAAUGUCUAAAGGUGAUACCACGCGAAAGCGCCUUGGCAGCACCUUUAG
s6	ACCUUUAGACAUUUACCUCUACAUGUCUAAAGGUGAUACCACGCGAAAGCGCCUUGGCAGCACCUUUAG
s7	ACCUUUAGACAUUUACCUCUAAAUGUCUAAAGGUAAAACAUACCAGAUCGCCACCCGCGCUUUAAUCUGGAGAGGUGAAGAAUACGACCACCUACCUUUAG
s8	ACCUUUAGACAUUUAGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCUAAAUGUC
s9	UUGCUCACCAUGGUCCUCACCAUGGUGAGCAAAAAACAUACCAGAUCGCCACCCGCGCUUUAAUCUGGAGAGGUGAAGAAUACGACCACCUUUGCUCAC
GFP antisense	ACCUUUAGACAUUUA
Theophylline aptamer	AGGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCU
s1.qpcr.fwd	ACCAGACAACCCAAAGCAA
s1.qpcr.rev	CTAAAGGTGCTGCCAAGGG
s1/2ham.qpcr.fwd	TAGCGGATCCAGGTCTGATGAGTCCGTGAGGACG
gfp.qpcr.fwd	ATTTTGGTTGAATTAGATGGTGA
gfp.qpcr.rev	CTGGCAATTTACCAGTAGTACAAA

Plasmid	Description	Parent plasmid
pTARGET1.gfp	pRS314-Gal expressing yEGFP	pRS314-Gal
pTARGET2.gfp/V	pRS314-Gal expressing yEGFP and Venus	pRS314-Gal
pSWITCH1.s1	GAL1 / ribozyme construct expressing s1	pRS316-Gal
pSWITCH1.s2	GAL1 / ribozyme construct expressing s2	pRS316-Gal
pSWITCH1.s3	GAL1 / ribozyme construct expressing s3	pRS316-Gal
pSWITCH1.s4	GAL1 / ribozyme construct expressing s4	pRS316-Gal
pSWITCH1.s5	GAL1 / ribozyme construct expressing s5	pRS316-Gal
pSWITCH1.s6	GAL1 / ribozyme construct expressing s6	pRS316-Gal
pSWITCH1.s7	GAL1 / ribozyme construct expressing s7	pRS316-Gal
pSWITCH1.s8	GAL1 / ribozyme construct expressing s8	pRS316-Gal
pSWITCH2.s1/9	GAL1 / ribozyme construct expressing s1 and s9	pRS316-Gal
pSWITCH1.anti	GAL1 / ribozyme construct expressing GFP antisense	pRS316-Gal
pSWITCH1.aptamer	GAL1 / ribozyme construct expressing theophylline aptamer	pRS316-Gal

Chapter 3.

The Regulation of Noise in a Metabolic Gene Determines an Ecological Strategy in Yeast

Parts reproduced from:

Travis S. Bayer, Kevin G. Hoff, Chase L. Beisel, Jack J. Lee, and Christina D. Smolke, **The regulation of noise in a metabolic gene determines an ecological strategy in yeast,** In review.

Evolutionary theory suggests that genetic regulatory circuits optimize protein expression levels to maximize fitness.[1, 2] However, the dependence of fitness on levels of a regulator protein across varying environmental conditions has seldom been measured. Here, we found that varying the expression of a transcriptional regulator of nitrogen metabolism, Dal80p[3], mediates a trade-off in fitness between resource-abundant and resource-limiting environments in *Saccharomyces cerevisiae* by modulating noise in the expression of a nitrogen metabolic enzyme, glutamate dehydrogenase (Gdh1p). Redundancy in the metabolic pathways and the regulatory network structure of ammonia assimilation allowed noise rather than abundance of Gdh1p to determine a classic dichotomy in ecological strategies: whether to specialize in maximizing fitness in resource abundant (rate strategy), or to specialize in maximizing fitness in resource limiting environments (yield strategy).[4, 5] Our results suggest that the optimization of protein noise may be as important as the optimization of protein expression levels for crafting ecological strategies to environmental demands.

The ability to assimilate and utilize nitrogen is a significant component of fitness in *S. cerevisiae*, and yeast display considerable strain-to-strain variation in the utilization of this key resource.[6] Nitrogen metabolism is largely controlled by a complex network of auto- and cross-regulation of four transcriptional regulators: Gln3p, Gat1p, Gzf3p, and Dal80p[7] (**Fig. 1**). We replaced the endogenous promoter of *DAL80* with the *GAL1-10* promoter by chromosomal integration (**Fig. 2a**) to achieve galactose-tunable control[8] of Dal80p (**Fig. 2b**)**.** We then measured fitness of the P_{GAL}-*DAL80* strain at various Dal80p levels across a range of ammonia concentrations (spanning near growth limiting to near toxic conditions)[9] by direct competition with a reference strain.[10] At low expression of Dal80p, the engineered strain displayed lower fitness than the parent strain at low ammonia concentrations, and higher fitness with increasing ammonia concentrations (**Fig. 3a**). Conversely, high Dal80p expression led to high relative fitness of the engineered strain at low ammonia and progressively lower fitness as ammonia concentration increased. To parameterize the fitness effects, we defined an environment-dependent fitness term, W_{env}, as the ratio of fitness in high ammonia (556 mM) to fitness in low ammonia (8.6 mM). W_{env} values greater than 1 indicate strains that are more competitive at high ammonia, where values less than 1 indicate strains more competitive at low ammonia (**Fig. 3b**).

Depending on the level of Dal80p expression, strains are either superior competitors in high or low ammonia concentrations, demonstrating a trade-off in fitness across environments. This fitness trade-off is specific to ammonia as a nitrogen source (**Fig. 4a**) and is dependent on all three ammonia assimilation pathways (**Fig. 4b**). Trade-offs have been demonstrated between traits such as reproduction and growth, longevity

and fecundity, and competitive ability and resistance to invasion.[11-13] One of the most prominent trade-off theories in biology is that of *r* versus *K* strategists.[4] Organisms displaying a *K* strategy are predicted to optimize utilization of resources, such as when the population is near its carrying capacity and resources are scarce, while *r* strategists are predicted to dominate when resources are abundant. These trade-offs are often underpinned by trade-offs in cellular biochemistry such as rate and yield of enzymatic reactions[14] and substrate uptake and affinity of resource transport.[15] We next examined the effect of changing Dal80p levels on the primary route of ammonia assimilation in yeast, glutamate dehydrogenase (Gdh1p). Analysis of single-cell expression of a Gdh1p:GFP fusion protein via flow cytometry in the P_{GAL}-*DAL80* strain revealed that increasing levels of Dal80p had little effect on mean Gdh1p abundance (**Fig. 5a**), but changed the noise in Gdh1p expression. Noise, or stochastic fluctuations in the abundance of proteins, can be enhanced or attenuated by regulatory circuits[16] and has been shown to be critical in biological functions such as determining viral latency[17] and competence in *Bacillus subtilis*.[18] In our engineered strain, low levels of Dal80p resulted in higher noise in Gdh1p expression (15% higher than the parent), while high levels of Dal80p reduced noise relative to the parent strain (20% lower than the parent) (**Fig. 5c**). Mean Gdh1p abundances remained relatively constant across all Dal80p levels (**Fig. 5d**).

To test whether noise in Gdh1p expression was correlated with the observed fitness trends independently of other Dal80p targets or galactose inducer effects, we generated a set of mutants with varying Gdh1p abundance and noise values by mutating the *GDH1* promoter. We identified sets of mutants having similar abundances and variable noise in Gdh1p expression (**Fig. 6a**), such that the contribution of either noise or

abundance could be parsed. We measured W_{env} in these mutant sets and observed a stronger positive correlation with noise in Gdh1p expression (correlation coefficient = 0.83, $R^2 = 0.69$) than abundance (correlation coefficient = 0.079, $R^2 = 0.0062$) (**Fig. 6b**).

To examine whether stochastic fluctuations in the expression of an enzyme can affect the total rate of product formation, we used the Gillespie algorithm to perform a stochastic simulation of the expression of an enzyme that converts a substrate into a single product (**Fig. 7a**). The simulation results show a classic hyperbolic enzyme titration curve (**Fig. 7b**). To examine the effect of noise on this system, we repeated the simulations, keeping the mean abundance of the enzyme constant while varying noise in enzyme expression (**Fig. 8a**). We then performed a series of simulations for different enzyme abundance values and calculated the noise dependence of the effective rate for each mean enzyme value (**Fig. 8b**). Noise dependence passes through a maximum value in these simulations, corresponding to the "cusp" of the enzyme titration plateau (**Fig. 7b**), indicating that there is a region of enzyme abundance where the system is most susceptible to noise. One qualitative prediction of this simulation is that noise will have a lower impact on product formation rates at high enzyme levels.

The above simulations and data suggest that noisy enzyme expression can decrease the rate of product formation from Gdh1p. Thus, a strain with lower rates of Gdh1p catalysis should show similar fitness trends as a high noise strain, whereas a strain with higher rates of Gdh1p catalysis should show fitness trends similar to a low noise strain. To test this, we replaced the endogenous copy of Gdh1p with previously characterized Gdh1p rate-enhanced and rate-deficient mutants. The D150H mutant was shown to have a 1000-fold lower rate measured *in vitro*, while the C313S mutant showed

a 0.4-fold increase in rate.[19] The catalytic rate-deficient D150H mutant was more competitive than the parent strain in high concentrations of ammonia (NH_4^+ > 278 mM) and was lower fitness than the parent strain in low ammonia concentrations (**Fig. 9**). In contrast, the catalytic rate-enhanced C313S mutant was less fit than the parent strain in high ammonia environments and displayed higher fitness in low ammonia concentrations.

Taken together, the simulations and data suggest that Gdh1p rates are correlated with fitness in different ways in high and low ammonia environments. For example, at low ammonia concentrations, fitness is positively correlated with Gdh1p rate: the catalytic mutant with high rate showed higher fitness as observed with strains with low noise, which confers higher effective rates according to simulations. A positive correlation of metabolic pathway rates with growth rate has been highlighted in ATP synthesis in microbes.[20] Conversely, in rich ammonia environments, Gdh1p rate is negatively correlated with fitness: the catalytic mutant with high rate showed low fitness, whereas the mutant with lower catalytic rate showed higher fitness. We speculate that the accumulation of downstream metabolites may have deleterious effects on fitness, as has been observed in the perturbation of AdoMet synthesis and methionine hyperaccumulation in yeast.[21]

Regardless of the mechanism through which Gdh1p rate affects fitness, the catalytic point mutants show that the rate of Gdh1p catalysis can impact fitness in different ammonia environments, while the simulations demonstrate how noise can affect the effective rate of reaction, drawing a causal link between Gdh1p noise and fitness trends. However, one would also expect changes in mean Gdh1p abundance to affect effective catalytic rates. Yeast possess an alternate route through which to synthesize

glutamate from ammonia, the NAD-dependent glutamate synthase Glt1p[22]. Glt1p has been shown to be upregulated 3-fold in Gdh1p deletion strains,[22] and titration studies with Dal80p and Gdh1p indicate that *GLT1* transcript levels are inversely correlated with Gdh1p abundance but not noise (**Figs. 10a,b**). These results indicate that regulatory networks controlling levels of Glt1p may provide a mechanism to buffer large-scale changes in Gdh1p expression. Deletion of *GLT1* may remove this "balancing" and impact how W_{env} trends correlate with Gdh1p abundance and noise.

We measured the relationship between noise, abundance, and W_{env} in the absence of this redundant pathway by constructing a *GDH1* promoter library (as before) in a *GLT1Δ* background strain. For these mutants, W_{env} shows stronger negative correlation with Gdh1p abundance (**Fig. 11a**, correlation coefficient = -0.41, $R^2 = 0.48$), than Gdh1p noise (**Fig. 11b**, correlation coefficient = 0.02, $R^2 = 10^{-5}$). These trends are in contrast to those observed previously (**Fig. 6b**) and to a similar experiment with wildtype *GLT1* (**Fig. 12a**, W_{env} versus noise: correlation coefficient = 0.52, $R^2 = 0.27$; **Fig.12b**, W_{env} versus abundance: correlation coefficient = 0.19, $R^2 = 0.03$). In the *GLT1Δ* background, mutants with low Gdh1p levels show high W_{env} values, as would be expected from canonical enzyme titration (less enzyme results in lower product formation rates, showing similar fitness trends to the rate-deficient D150H mutant). The presence or absence of Glt1p, an alternate route for ammonia assimilation, determines whether noise (in the presence of Glt1p) or abundance (in the absence of Glt1p) of Gdh1p determines phenotypic behavior in a sampling of mutants (**Fig. 13**). As an additional control for this model, we overexpressed Glt1p in a set of *GDH1* promoter mutants. Increasing the amount of this enzyme should push the system into a regime where rates (and, therefore,

fitness) are not as sensitive to Gdh1p noise, according to the above simulation. We found that *GLT1* overexpression diminished the effect Gdh1p noise had on W_{env} (**Fig. 14**), supporting that wildtype Gdh1p and Glt1p abundances are in a regime where fitness is susceptible to noise.

Because noise in Gdh1p expression determines an ecological strategy, then environmental conditions should select for members of a population showing high or low noise depending on the favored strategy. To test this hypothesis, we competed the *GDH1* promoter library in two batch culture ammonia environments (139 mM and 556 mM) to impose different selection pressures. Here, higher ammonia is predicted to have stronger selection pressure for rate strategists.[5] After 36 generations of competition, a sampling of individuals from each environment revealed populations that were largely clustered around the noise and W_{env} values of the initial library (**Fig. 15a, b**). At 60 generations, both populations show enrichment for rate strategists ($W_{env} > 1$) versus the initial library (**Fig. 15c**). Interestingly, each population is composed of a mixture of rate and yield strategists. Importantly, the 556 mM environment enriched for mutants with higher average noise in Gdh1p expression than the parent strain ($CV^2_{initial} = 0.61$, $CV^2_{evolved} = 0.65$, $P = 0.007$), as well as rate strategists ($W_{env, initial} = 0.91$, $W_{env, evolved} = 1.69$, $P = 0.01$). The 139 mM environment enriched for mutants displaying lower noise than the parent strain ($CV^2_{initial} = 0.61$, $CV^2_{evolved} = 0.56$, $P = 1.5 \times 10^{-6}$). No such enrichment was observed when comparing mean Gdh1p abundance from each evolved population (**Fig. 16**, $p_{139mM} = 0.85$, $p_{556mM} = 0.86$, $P = 0.72$). Thus, because noise in Gdh1p expression is linked with the *r* and *K* phenotypes, it can be shaped by environments that select for those

phenotypes. To our knowledge, this is the first demonstration that gene expression noise is a selectable trait.

One of the more concrete definitions of *r* and *K* strategies is the notion of density-dependent selection: *r* strategists show fitness advantage at low population density, while *K* strategists show advantage at high density. To test if strains differing only in Gdh1p noise displayed this behavior, we designated two clones that displayed high and low noise values from the evolved populations (at 60 generations) but similar abundance values as r and K, respectively (**Fig. 17a**). We then competed these strains versus a reference strain as above at varying initial densities for 24 hours to simulate a "season" of competition.[23] Competitions were performed in the lowest (8.6 mM) and highest (556 mM) ammonia concentrations measured to simulate a resource-poor and resource-abundant environment, respectively. The fitness of the high noise r strategist is negative density-dependent, implying that this strain is a better competitor in low-density environments (**Fig. 17b**). The density-dependence in the high ammonia environment is not as severe, suggesting that the abundance of ammonia determines how stringent competition will be for a given population density. Conversely, the K strategist is more competitive than the r strain at high population densities, demonstrating that strains differing in Gdh1p expression noise can recapitulate canonical ecological strategies.[24]

By measuring an environment-dependent fitness as a function of a regulator level, we were able to uncover a *r*/*K* trade-off in ammonia metabolism that is modulated by noise in the expression of a metabolic gene, *GDH1*, and not mean abundance of this enzyme by virtue of cross-regulation with a redundant pathway. The expression of Dal80p itself is known to be sensitive to nitrogen catabolite repression (NCR), where

expression is upregulated in nitrogen starvation, environments with non-preferred nitrogen sources, or addition of the small molecule rapamycin.[3] The strategy of Gdh1p noise regulation in varying nitrogen environments is an interesting topic for future study in that strains that differed only in Gdh1p expression noise showed canonical *r* and *K* density-dependent behavior, indicating that noise is able to shape an ecological strategy in nitrogen assimilation. The results presented here may be relevant to other metabolic pathways and organisms, as metabolic redundancy is widespread throughout Nature. While it may be advantageous to harbor duplicate and/or redundant genes and pathways to buffer the effects of gene loss, redundant pathways may also present additional opportunities for the regulation of metabolism and fitness. Regulatory mechanisms that buffer noise in gene expression, such as negative feedback, may be indicative of ancestral adaptation towards specific environments. An exploration of the network architectures regulating such systems will reveal whether the emergence of alternative routes for metabolic processes has presented evolution with a design opportunity for modulating ecological strategy. Taken together, these results illustrate that regulatory networks may optimize noise in gene expression in addition to protein levels to fashion adaptive solutions to environmental challenges, and suggest a link between networks of genetic regulation and networks of ecological interactions.

Methods Summary

Strains and media. All manipulations were performed with derivatives of the S288c background from the University of California San Francisco GFP-tag collection

(Invitrogen).[25] Yeast were grown in synthetic complete media (1.7 g/L yeast nitrogen base, nitrogen source as specified).

Competitions and fitness assays. Fitness was assayed by direct competition versus a common reference strain[10]. The competitor and reference strain constitutively express different fluorescent proteins (GFP and CFP, respectively) from the *ADH1* promoter integrated into the chromosome. The frequency of competitor and reference strain were quantitated before and after the growth period by counting the numbers of GFP expressing cells to non-GFP expressing cells. Fitness (w) of the competitor strain is reported as the natural log of the change in frequency of the strain versus the reference strain during the competitive growth period over the change in frequency of the parent strain versus the reference strain over the same growth period:

w = ln (fold change of the engineered strain after competition with reference strain / fold change of the parent strain after competition with reference strain)

Calculation of noise. Noise was calculated as the square of the coefficient of variation of the distribution[26].

Mutagenic PCR and construction of promoter libraries. A construct comprising the region approximately 500 nucleotides upstream of the *GDH1* coding region was subjected to error-prone PCR, assembled with a selectable marker, and transformed into the specified strain background.

Gillespie simulations. Simulations were performed as previously described,[27] with the probability of reaction occurring in a given time interval proportional to the reaction rate. Noise was introduced by varying the ratio of mRNA decay rate to protein synthesis rate

(burst size of protein production). Compensatory changes in mRNA synthesis rate were adjusted to simulate expression with varying noise and similar abundance.

Experimental evolution. Aliquots from the *GDH1* promoter library were diluted into 2 mL synthetic complete with either 139 mM ammonia or 556 mM ammonia. Populations were grown in batch culture and diluted 10^3-fold into respective fresh media every 24 hours.

Statistical analysis. Pearson correlation coefficients were calculated with either Gdh1p noise or abundance as the dependent variable and W_{env} as the independent variable, respectively. Significance of population averaged W_{env} and noise values in evolution experiments was calculated with a two-tailed t-test. Data was tested for normality using the Kolmogorov-Smirnov test. The mean ±s.d. from at least three independent experiments is shown for all data.

Detailed methods can be found in Appendix B.

References

1. Rosen, R. Optimality principles in biology (Butterworths, London,, 1967).
2. Orr, H. A. The genetic theory of adaptation: a brief history. Nat Rev Genet 6, 119-27 (2005).
3. Cunningham, T. S., Rai, R. & Cooper, T. G. The level of DAL80 expression down-regulates GATA factor-mediated transcription in Saccharomyces cerevisiae. J Bacteriol 182, 6584-91 (2000).
4. MacArthur, R. H. & Wilson, E. O. The theory of island biogeography (Princeton University Press, Princeton, N.J.,, 1967).
5. Pianka, E. R. R-Selection and K-Selection. American Naturalist 104, 592-& (1970).
6. Homann, O. R., Cai, H., Becker, J. M. & Lindquist, S. L. Harnessing natural diversity to probe metabolic pathways. PLoS Genet 1, e80 (2005).

7. Cooper, T. G. Transmitting the signal of excess nitrogen in Saccharomyces cerevisiae from the Tor proteins to the GATA factors: connecting the dots. FEMS Microbiol Rev 26, 223-38 (2002).
8. Hawkins, K. M. & Smolke, C. D. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. J Biol Chem 281, 13485-92 (2006).
9. Hess, D. C., Lu, W., Rabinowitz, J. D. & Botstein, D. Ammonium toxicity and potassium limitation in yeast. PLoS Biol 4, e351 (2006).
10. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet 4, 457-69 (2003).
11. Ghalambor, C. K., Reznick, D. N. & Walker, J. A. Constraints on adaptive evolution: the functional trade-off between reproduction and fast-start swimming performance in the Trinidadian guppy (Poecilia reticulata). Am Nat 164, 38-50 (2004).
12. Rose, M. & Charlesworth, B. A test of evolutionary theories of senescence. Nature 287, 141-2 (1980).
13. Weitz, J. S., Hartman, H. & Levin, S. A. Coevolutionary arms races between bacteria and bacteriophage. Proc Natl Acad Sci U S A 102, 9535-40 (2005).
14. Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. Science 292, 504-7 (2001).
15. Gudelj, I., Beardmore, R. E., Arkin, S. S. & MacLean, R. C. Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments. J Evol Biol 20, 1882-9 (2007).
16. Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. Nature 420, 231-7 (2002).
17. Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P. & Schaffer, D. V. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122, 169-82 (2005).
18. Suel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440, 545-50 (2006).
19. Wang, X. G. & Engel, P. C. Identification of the reactive cysteine in clostridial glutamate dehydrogenase by site-directed mutagenesis and proof that this residue is not strictly essential. Protein Eng 7, 1013-6 (1994).
20. Pfeiffer, T. & Bonhoeffer, S. Evolution of cross-feeding in microbial populations. Am Nat 163, E126-35 (2004).
21. Lu, P. et al. Global metabolic changes following loss of a feedback loop reveal dynamic steady states of the yeast metabolome. Metab Eng 9, 8-20 (2007).
22. Valenzuela, L., Ballario, P., Aranda, C., Filetici, P. & Gonzalez, A. Regulation of expression of GLT1, the gene encoding glutamate synthase in Saccharomyces cerevisiae. J Bacteriol 180, 3533-40 (1998).
23. MacLean, R. C. & Gudelj, I. Resource competition and social conflict in experimental populations of yeast. Nature 441, 498-501 (2006).
24. Sutherland, W. J. From individual behaviour to population ecology (Oxford University Press, Oxford ; New York, 1996).

25. Huh, W. K. et al. Global analysis of protein localization in budding yeast. Nature 425, 686-91 (2003).
26. Paulsson, J. Summing up the noise in gene networks. Nature 427, 415-8 (2004).
27. Gillespie, DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81, 2350 (1977).

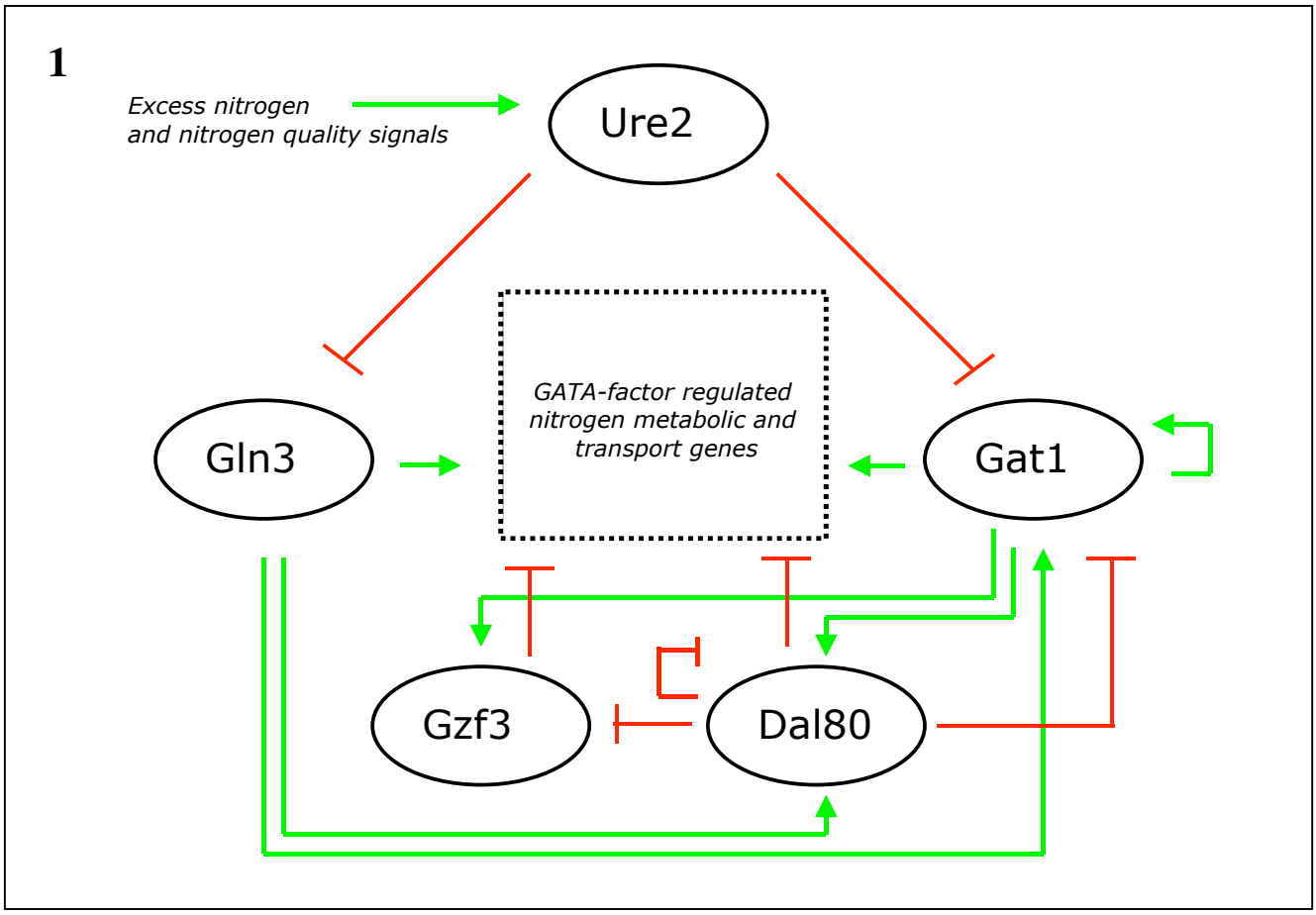
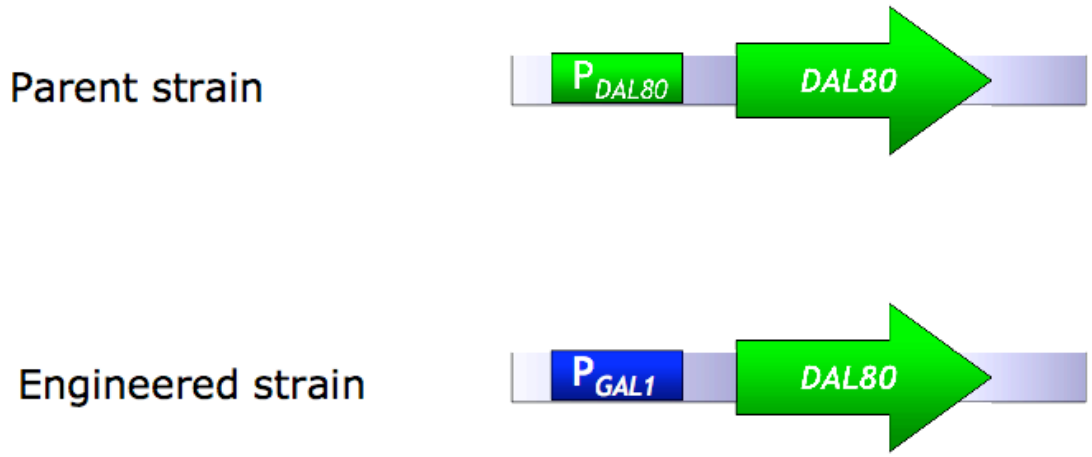


Figure 1. Schematic of nitrogen regulation circuit in yeast. Diagram of known interactions between four transcriptional regulators of nitrogen metabolism: Gln3p, Gat1p, Dal80p, and Gzf3p. Upstream kinases and other signals (from both the environment and inside the cell) of excess nitrogen and high nitrogen quality repress Gln3p and Gat1p by sequestering the transcription factors outside the nucleus with the Ure2p protein. In response to changes in nitrogen availability or quality, Gln3p and Gat1p induce the expression of Gzf3p and Dal80p, which in turn cross-regulate and auto-regulate the other factors as shown. These four regulators in turn regulate the set of nitrogen utilization genes (approximately 500). Due to complex and combinatorial interactions in the circuit as illustrated, as well as at the individual promoters of nitrogen utilization genes, expression can be up-regulated or down-regulated depending on the environment, and can be adjusted to meet environmental demands.

2a


Parent strain
P DAL80
DAL80
Engineered strain
P GAL1
DAL80

2b

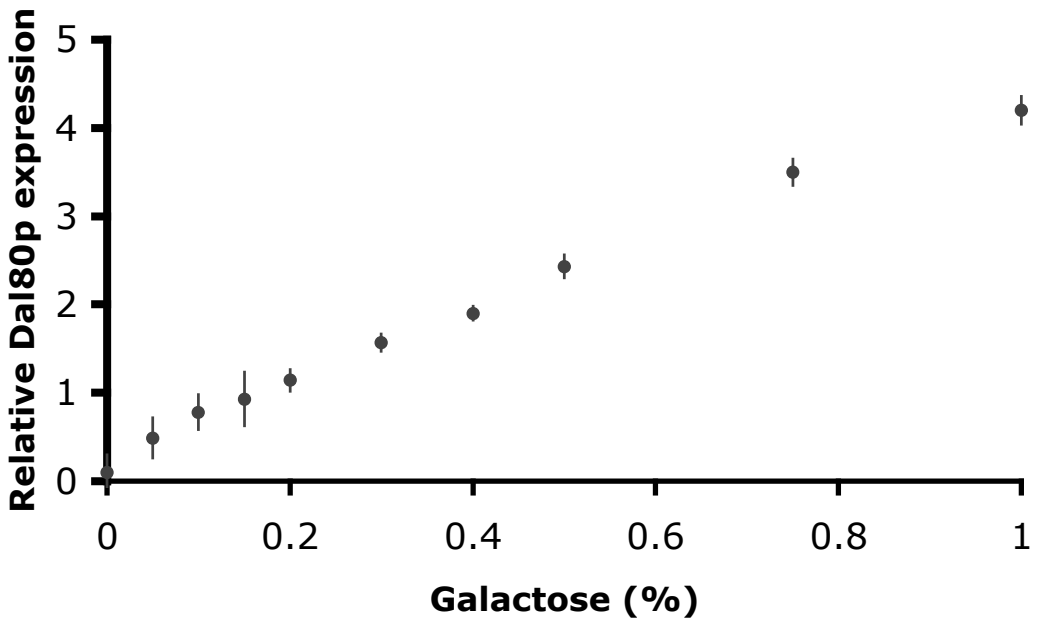
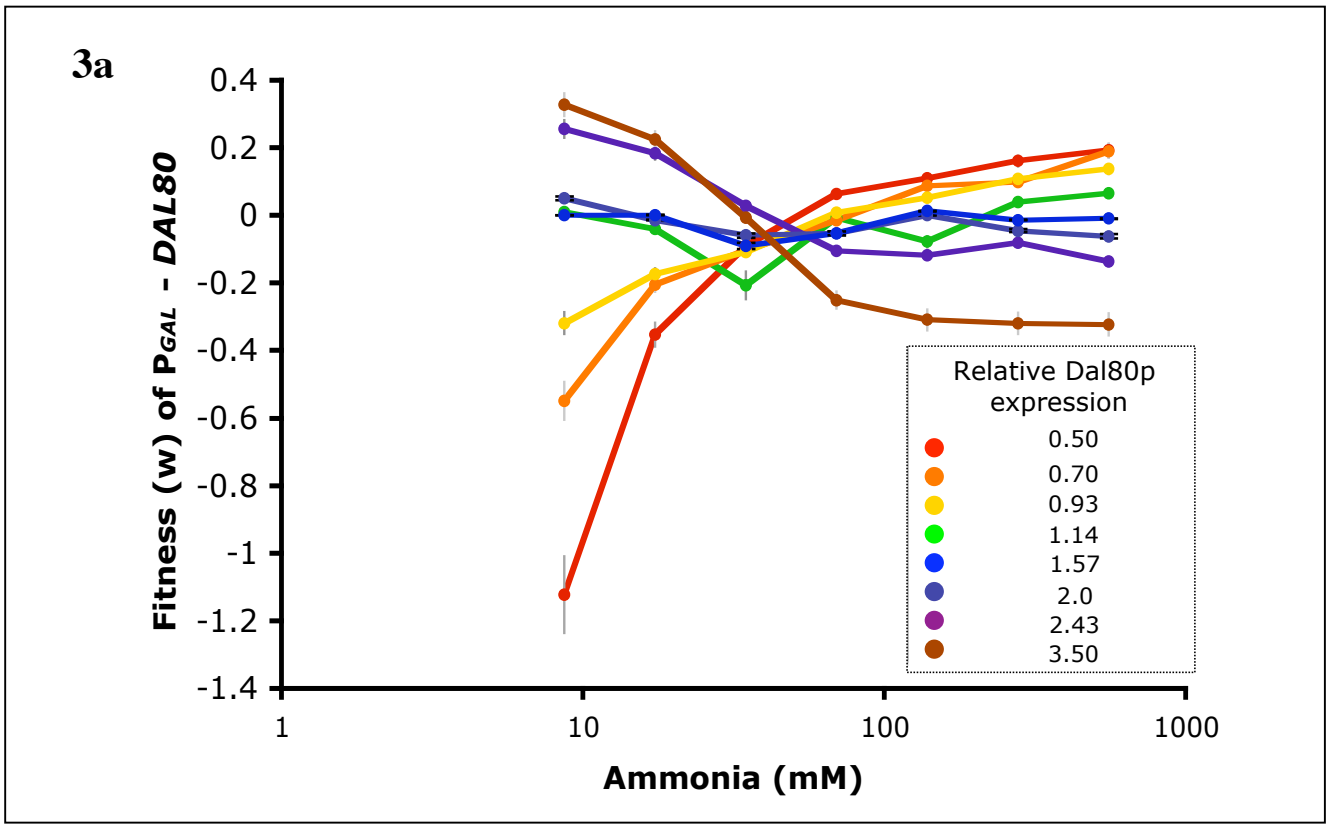
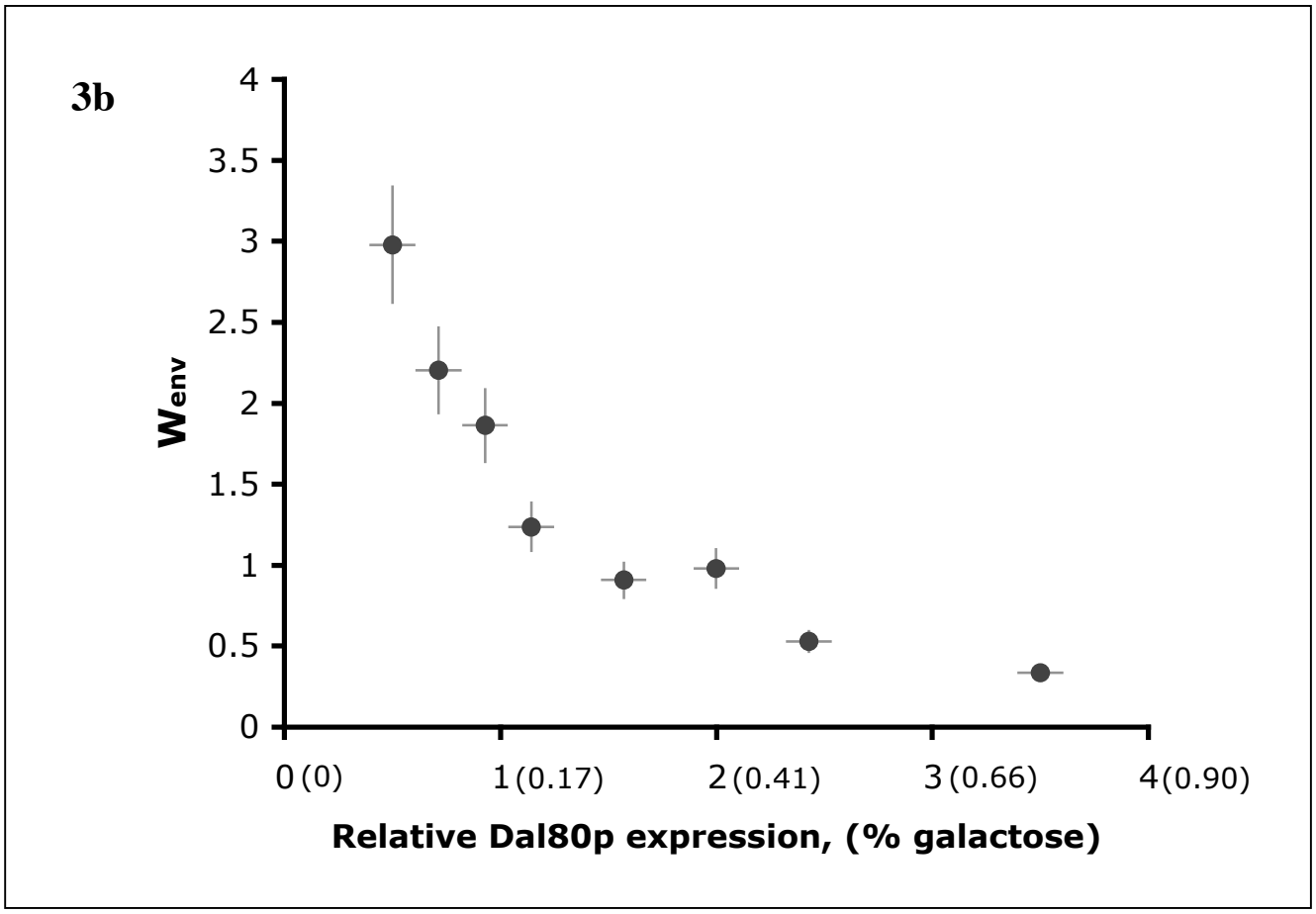

Relative Dal80p expression
0
1
2
3
4
5
0
0.2
0.4
0.6
0.8
1
Galactose (%)

Figure 2 Expression level of a transcriptional regulator determines competitive ability in varying ammonia concentrations. a, The parent strain and engineered P_{GAL}-*DAL80* strain, where the endogenous *DAL80* promoter is replaced with the *GAL1-10* promoter. **b**, Tunable Dal80p expression as a function of galactose concentration for the engineered P_{GAL}-*DAL80* strain. Cells were grown overnight in non-inducing/non-repressing media (synthetic complete with 2% (wt/vol) sucrose, 1% raffinose), diluted 50-fold in the specified galactose concentration, and grown for 6 hours. Cells were harvested and total RNA was extracted as specified. Relative transcript levels were measured by reverse transcription and quantitative PCR (qRT-PCR). Relative *DAL80* transcript levels were normalized to relative *ACT1* transcript levels for each sample and are reported relative to the parent strain. The mean ±s.d. from at least three independent experiments is shown.


3a
Fitness (w) of PGAL - DAL80
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
1
10
100
1000
Ammonia (mM)
Relative Dal80p expression
0.50
0.70
0.93
1.14
1.57
2.0
2.43
3.50


3b
Wenv
4
3.5
3
2.5
2
1.5
1
0.5
0
0 (0)
1 (0.17)
2 (0.41)
3 (0.66)
4 (0.90)
Relative Dal80p expression, (% galactose)

3a, Fitness of the engineered strain across varying ammonia concentrations at different Dal80p expression levels. Dal80p expression was varied by altering the concentration of galactose in the media and measured by qRT-PCR, and is reported relative to parent Dal80p levels for each set of fitness data. Equal amounts of the reference and P_{GAL}-*DAL80* or parent strains were mixed and grown in the indicated ammonia and galactose concentrations. Numbers of each strain were quantitated through flow cytometry and fitness of the P_{GAL}-*DAL80* strain is reported as the natural log of the change in frequency over the growth period relative to the parent strain. **b,** Environment-dependent fitness parameter, W_{env}, of the P_{GAL}-*DAL80* strain as a function of Dal80p expression. W_{env} is calculated as the ratio of fitness in high ammonia (556 mM) to fitness in low ammonia (8.6 mM) relative to the parent strain.

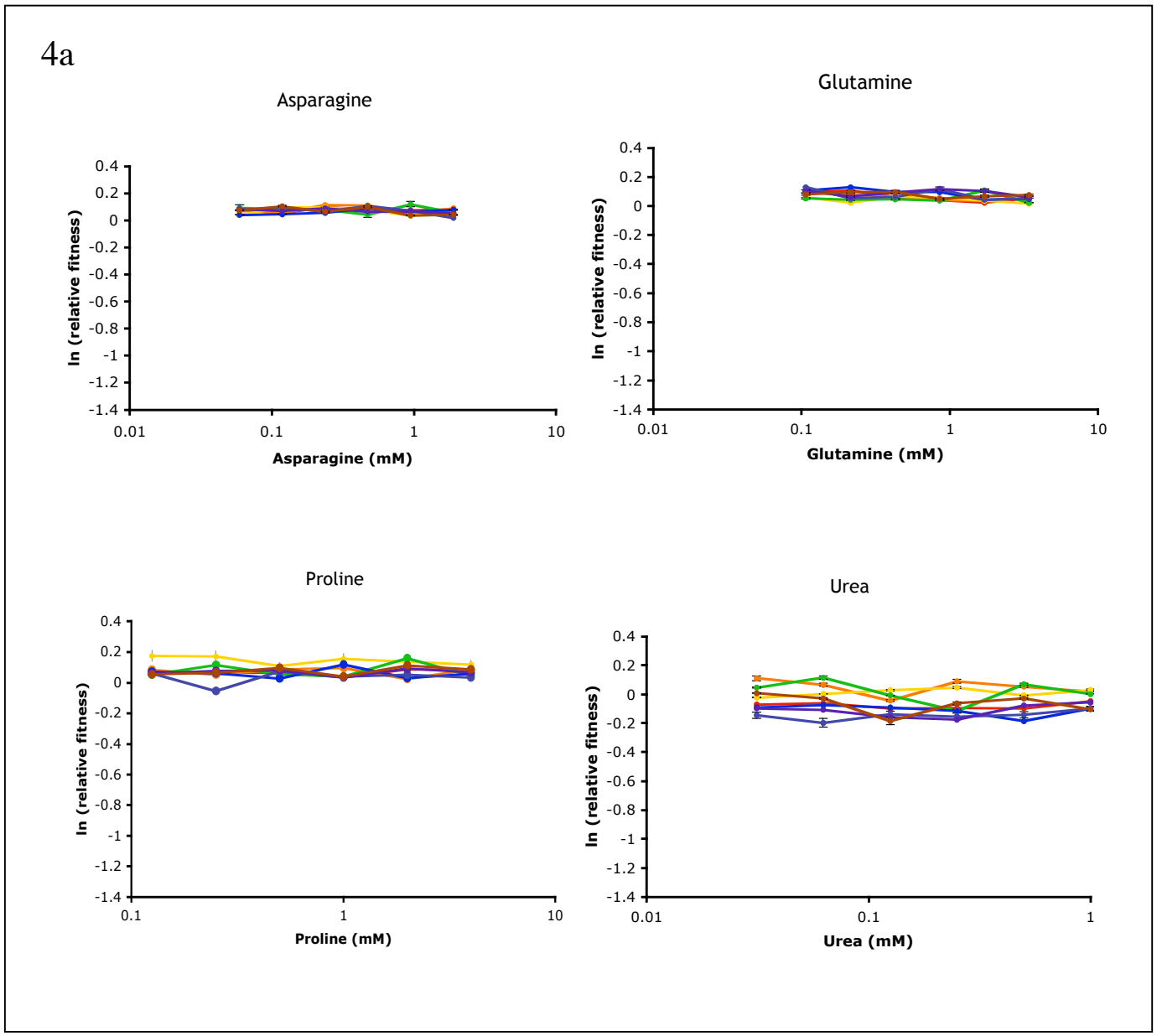


Figure 4a Fitness trends are specific to ammonia as the nitrogen source. Relative fitness of the engineered P_{GAL}-*DAL80* strain in alternative nitrogen sources at different Dal80p expression levels. Fitness trends are reported across varying concentrations of preferred nitrogen sources (asparagine and glutamine) and non-preferred nitrogen sources (proline and urea). Colors represent relative Dal80p expression levels as indicated in Figure 1b. Competitive fitness shows little change relative to the parent strain in either preferred or non-preferred nitrogen sources. The mean ±s.d. from at least three independent experiments is shown for all data.

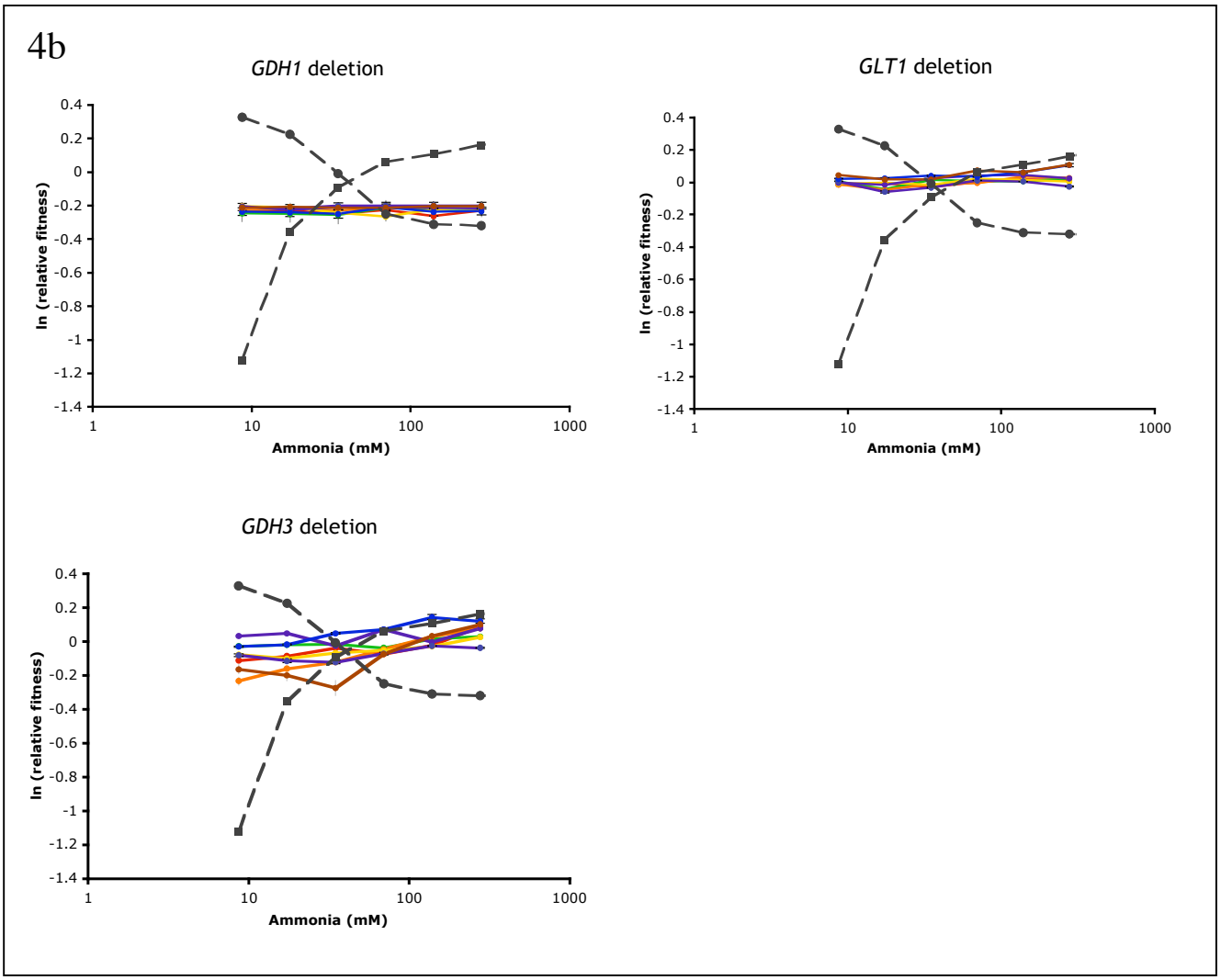
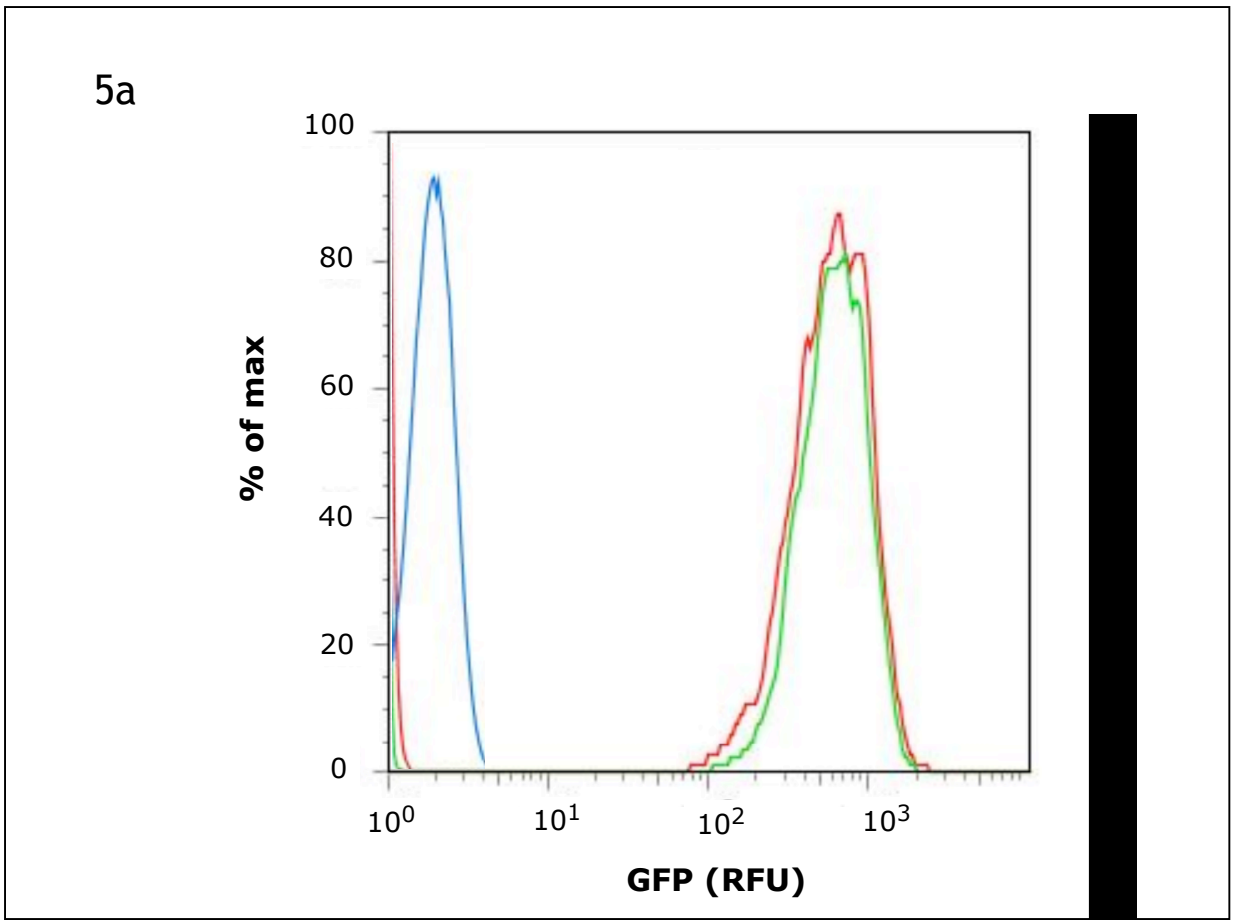
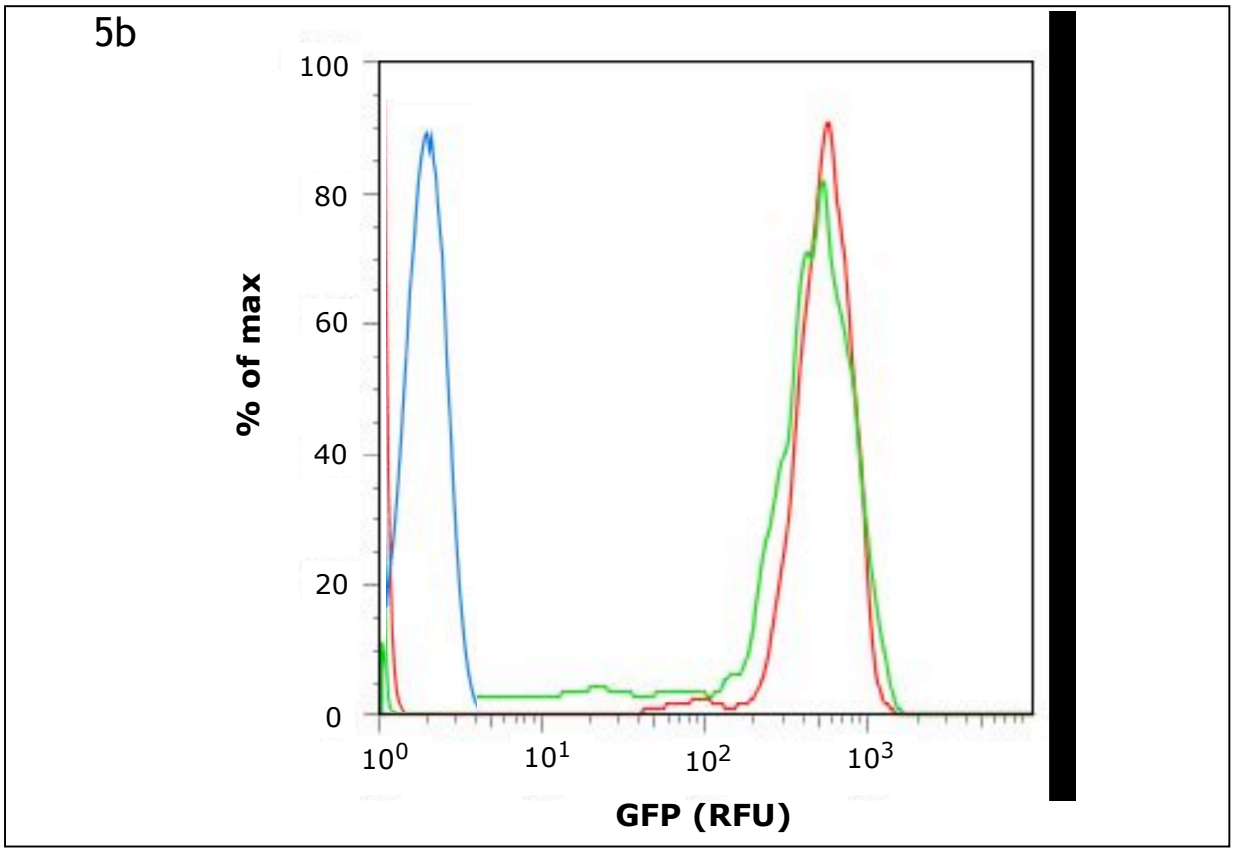
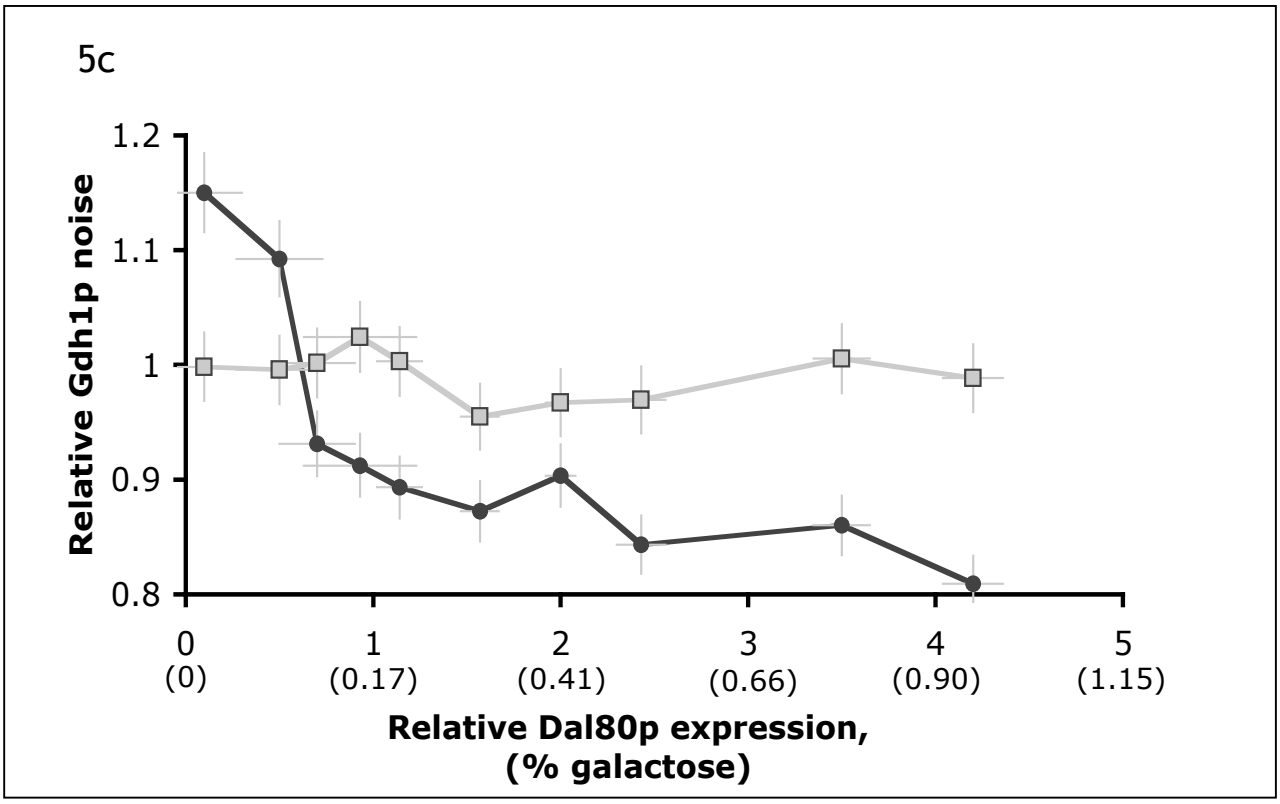


Figure 4b Fitness trends require all three ammonia assimilation pathways. Fitness trends for the P_{GAL}-*DAL80 gdh1Δ,* P_{GAL}-*DAL80 glt1Δ,*and P_{GAL}-*DAL80 gdh3Δ* strains across varying ammonia concentrations at different Dal80p expression levels. Dashed lines represent relative fitness of the engineered strain (P_{GAL}-*DAL80*) across varying ammonia concentrations for low Dal80p (squares, 0.5-fold parent strain) and high Dal80p (circles, 3.5-fold parent strain) expression levels for comparison. The observed fitness trends are abolished in the P_{GAL}-*DAL80 gdh1Δ* strain, and fitness across all ammonia concentrations is lower than the parent strain. Fitness trends are absent in the P_{GAL}-*DAL80 glt1Δ* strain similar to the Gdh1p deletion. Fitness values slightly increase with ammonia concentration in the P_{GAL}-*DAL80 gdh3Δ* strain, although fitness trends are similarly abolished. The mean ±s.d. from at least three independent experiments is shown for all data.


5a
100
80
60
40
20
0
% of max
10^0
10^1
10^2
10^3
GFP (RFU)


5b
100
80
60
40
20
0
% of max
10^0
10^1
10^2
10^3
GFP (RFU)


5c
Relative Gdh1p noise
1.2
1.1
1
0.9
0.8
0
(0)
1
(0.17)
2
(0.41)
3
(0.66)
4
(0.90)
5
(1.15)
Relative Dal80p expression,
(% galactose)

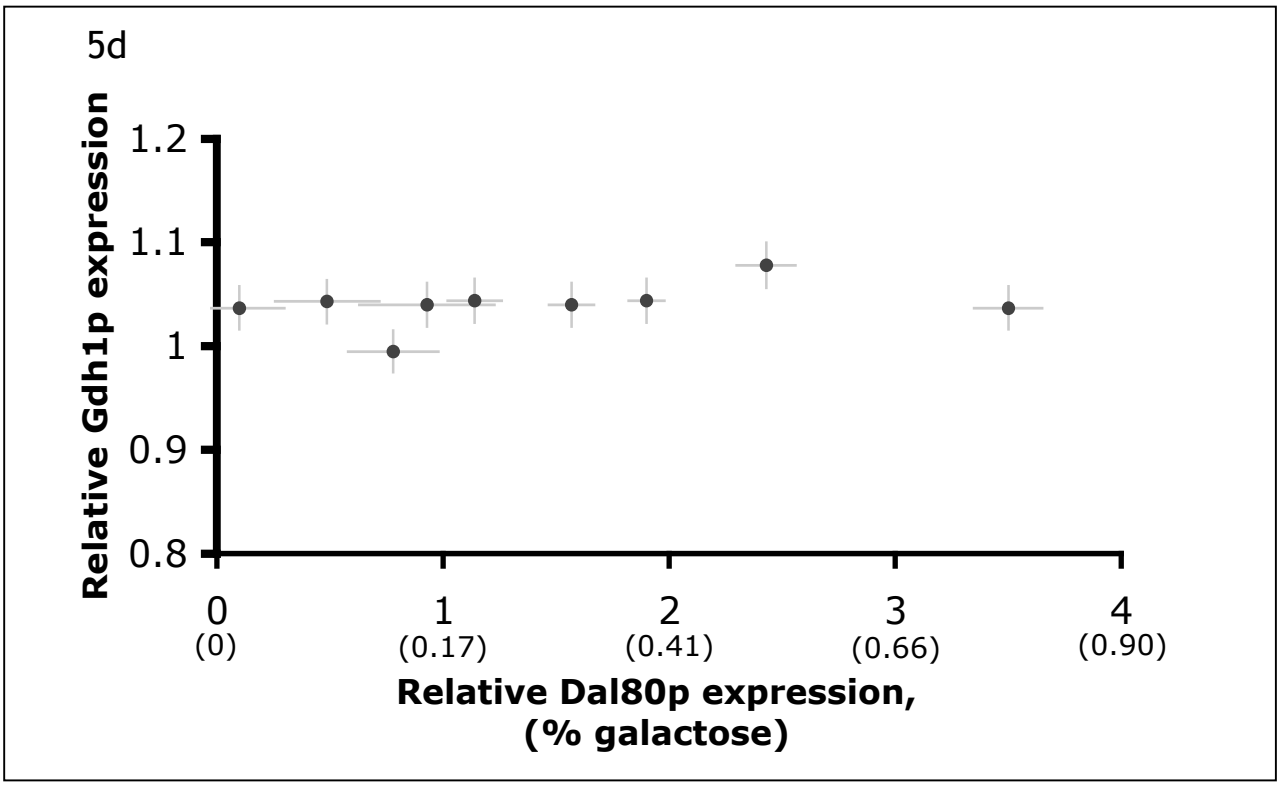
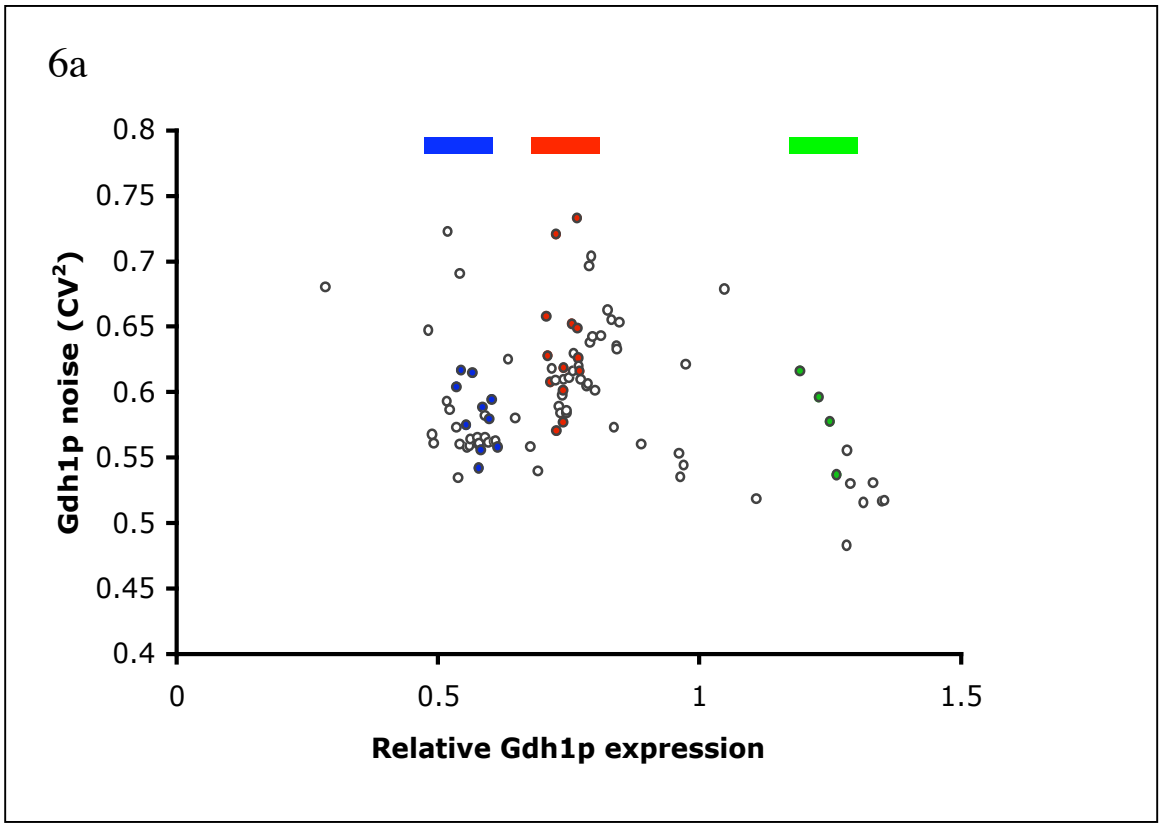

5d
Relative Gdh1p expression
1.2
1.1
1
0.9
0.8
0
(0)
1
(0.17)
2
(0.41)
3
(0.66)
4
(0.90)
Relative Dal80p expression,
(% galactose)

Figure 5a, Representative histogram of Gdh1p expression at 0% in wildtype (green), P_{GAL}-*DAL80* cells at 0% galactose (red), and cellular autofluorescence (W303α, blue). **b,** Representative histogram of Gdh1p expression at 0% in wildtype (green), P_{GAL}-*DAL80* cells at 0% galactose (red), and cellular autofluorescence (W303α, blue). **c,** Noise in Gdh1p:GFP expression in the P_{GAL}-*DAL80* strain (black) and parent strain (grey) with varying Dal80p levels. Noise values are reported relative to the parent strain at 0% galactose. **d,** Gdh1p expression levels in the P_{GAL}-*DAL80* strain with varying Dal80p levels. The arithmetic mean of the fluorescence population distribution relative to the parent strain is shown, and displays little change as Dal80p levels are increased.


6a
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
Gdh1p noise (CV²)
0
0.5
1
1.5
Relative Gdh1p expression

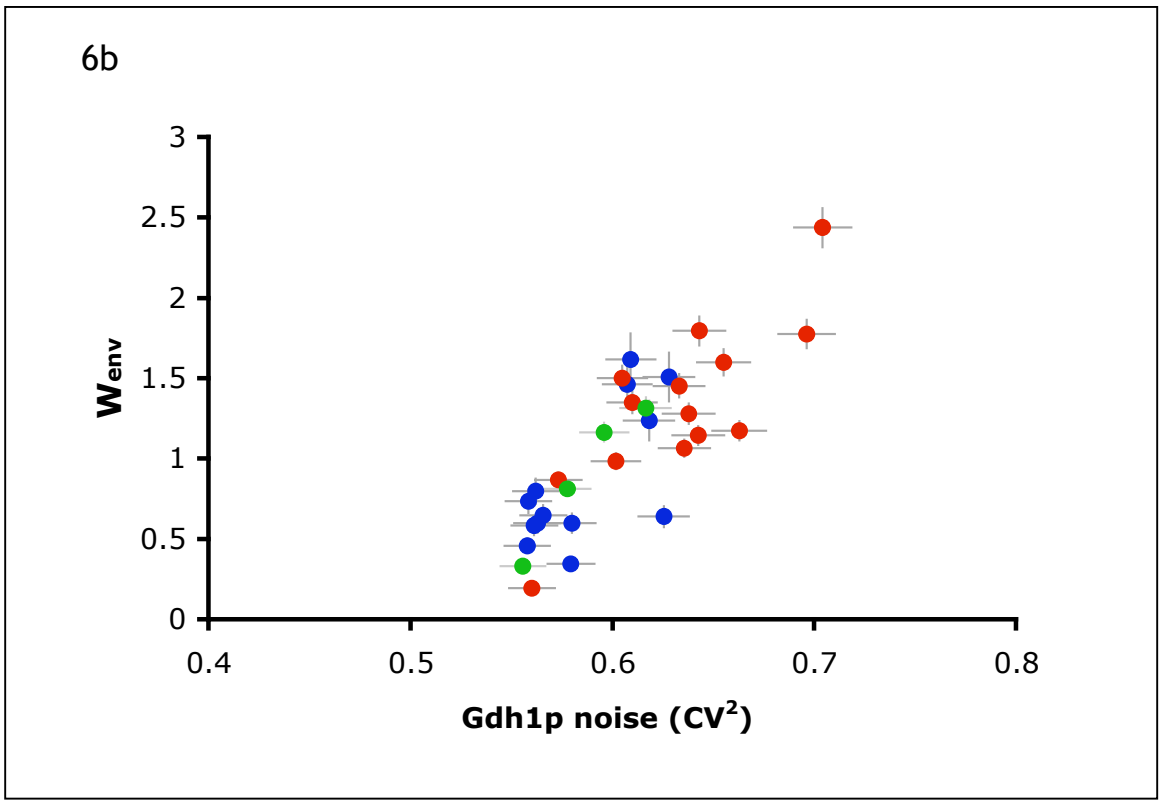
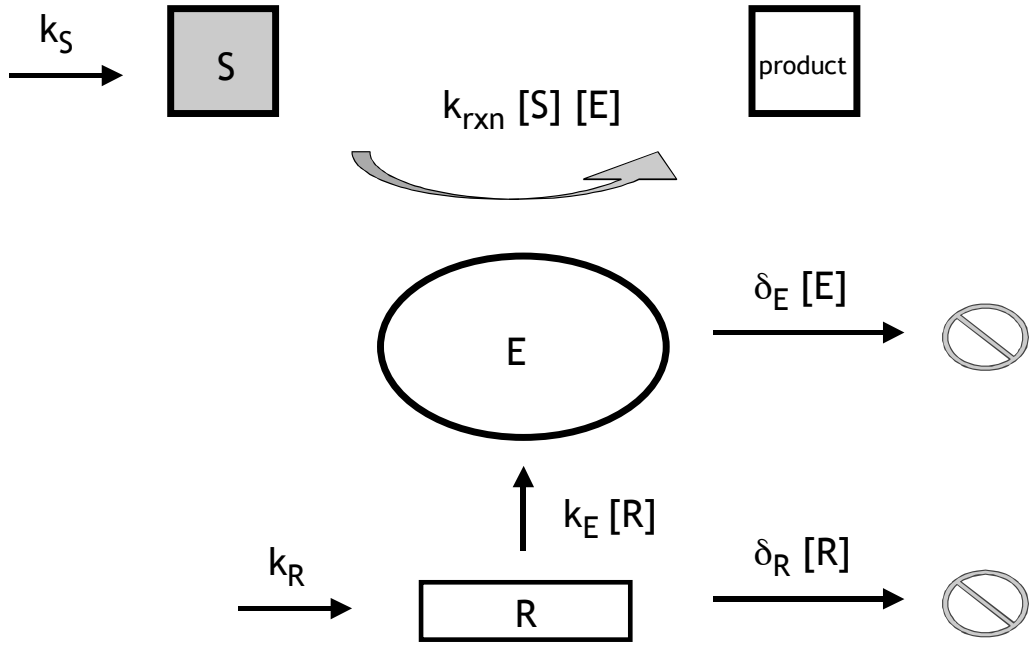

6b
3
2.5
2
1.5
1
0.5
0
Wenv
0.4
0.5
0.6
0.7
0.8
Gdh1p noise (CV²)

Figure 6 *GDH1* promoter mutants show a range of Gdh1p abundance versus noise values. a, Noise in Gdh1p expression versus mean Gdh1p abundance for a set of randomly selected *GDH1* promoter mutants (n = 91). Red, blue, and green bars indicate mutant sets having similar Gdh1p abundances (low, medium, and high, respectively) over a range of noise values. All errors are within 5% of the reported values. **b,** W_{env} versus noise in Gdh1p expression for the highlighted mutant sets. Mutant sets with low (red), medium (blue), and high (green) Gdh1p abundance levels are indicated.

7a


k_S
S
product
k_{rxn} [S] [E]
E
δ_E [E]
k_E [R]
k_R
R
δ_R [R]

7b

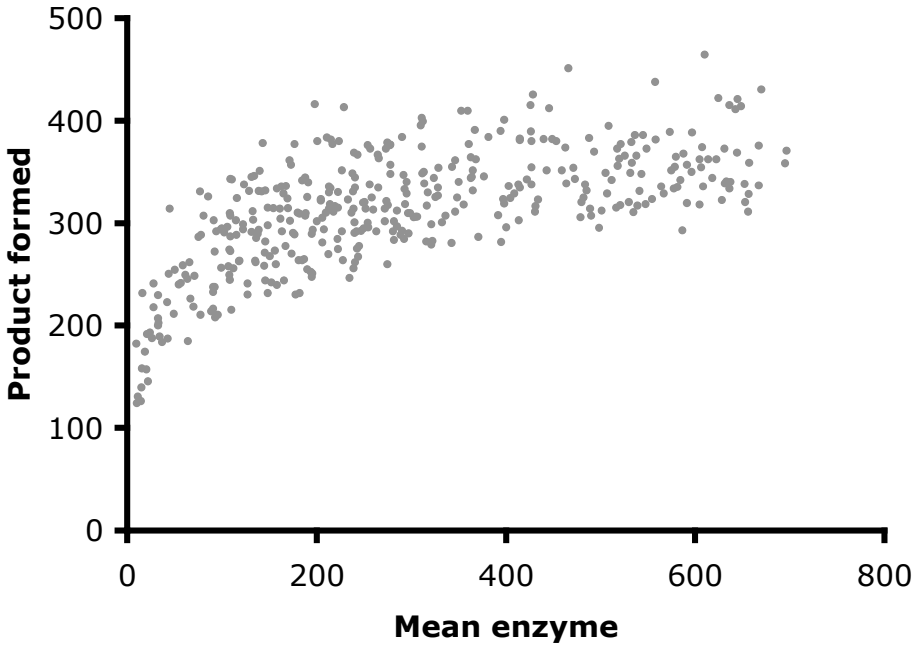
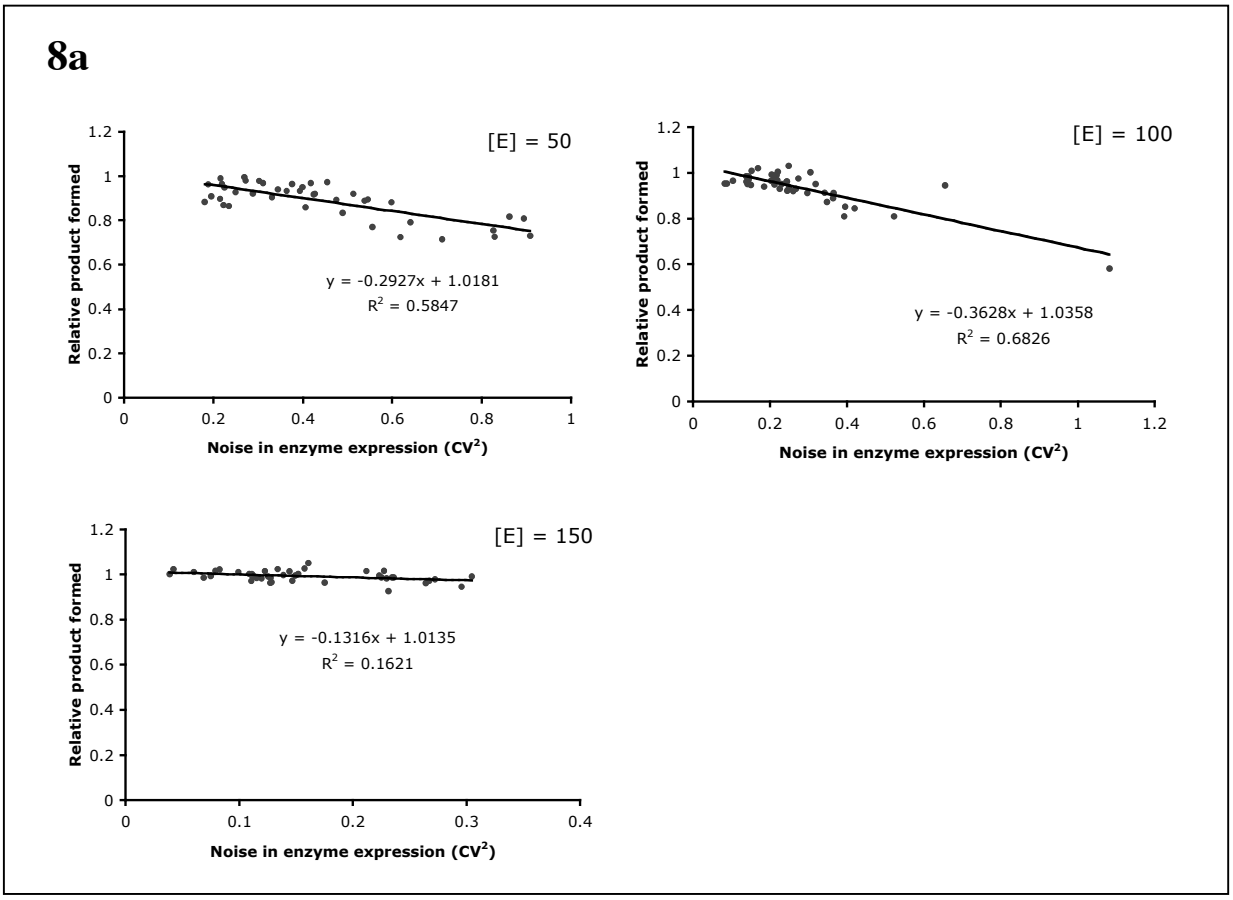

Product formed
500
400
300
200
100
0
0
200
400
600
800
Mean enzyme

Figure 7 Schematic of enzyme reaction simulation. a, schematic for the simulated enzyme reactions is shown. In the specified system, mRNA is produced at a constant rate and decays at a rate proportional to its concentration. Enzyme is produced at a rate proportional to the mRNA concentration and decays at a rate proportional to its concentration. Substrate is imported at a constant rate and is converted to product at a rate dependent on the substrate and enzyme concentrations. The rate of product formation in this system can be determined from the amount of product formed over the run time of the simulation. **b,** Simulation results showing product formation as a function of enzyme abundance with other parameters held constant. Product formation shows canonical hyperbolic dependence on enzyme levels with a plateau region at mean enzyme levels greater than 80.


8a
[E] = 50
y = -0.2927x + 1.0181
R2 = 0.5847
Relative product formed
Noise in enzyme expression (CV2)
[E] = 100
y = -0.3628x + 1.0358
R2 = 0.6826
Relative product formed
Noise in enzyme expression (CV2)
[E] = 150
y = -0.1316x + 1.0135
R2 = 0.1621
Relative product formed
Noise in enzyme expression (CV2)

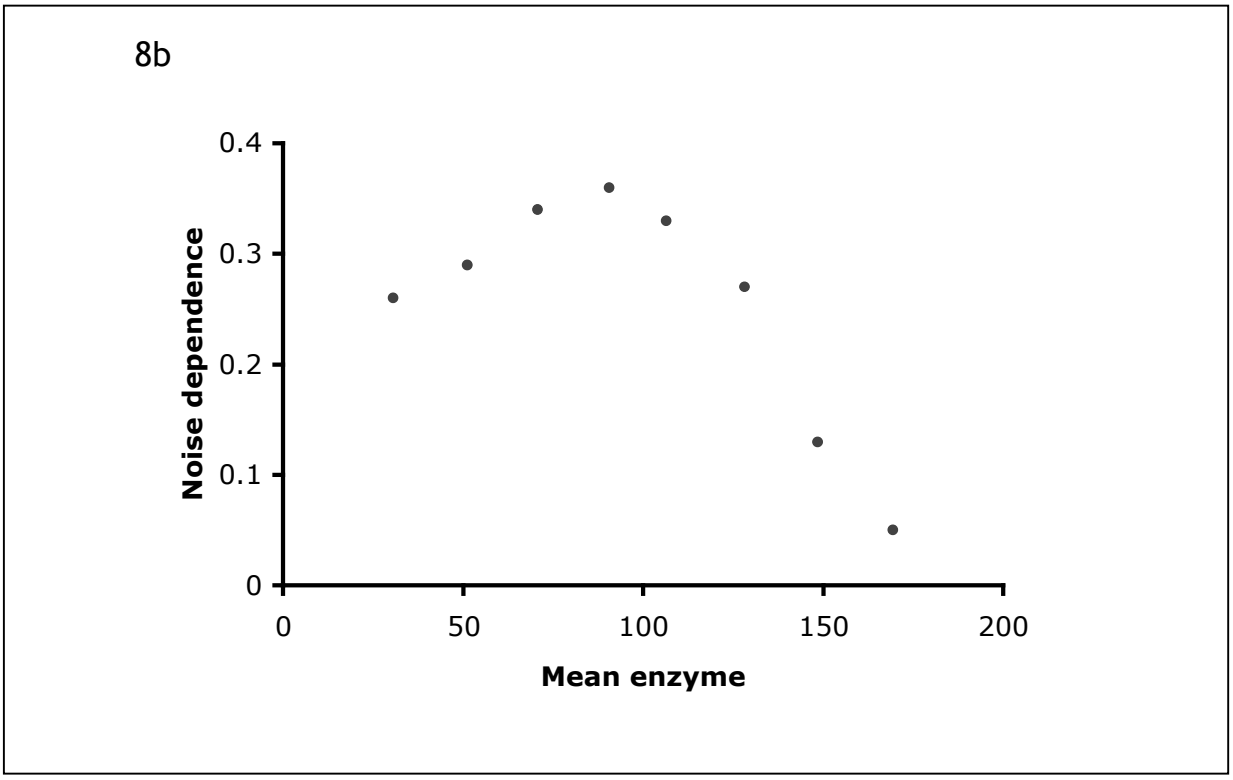

8b
Noise dependence
Mean enzyme
0
0.1
0.2
0.3
0.4
0
50
100
150
200

Figure 8 The dependence of product formation on the noise in enzyme expression in a simulated enzyme reaction system. **a**, the dependence of product formation on noise in enzyme expression at different enzyme abundances for the simulated system is shown. Noise was varied by changing the ratio of δR (mRNA decay rate) to kE (enzyme translation rate), which varies the average number of proteins translated from a single mRNA (burst size), and mRNA synthesis rates were adjusted accordingly to have similar enzyme abundances between simulations. The slope of the product formation versus noise trend represents the noise susceptibility at this enzyme concentration. Simulations are shown for three mean enzyme abundances ([E] = 50, 100, and 150). **b** , Noise dependence of product formation as a function of enzyme expression for the simulated enzyme system. Noise was varied by changing the average number of proteins translated from a single mRNA (burst size), while compensating the rate of RNA production to retain similar mean abundances between simulations. The slope of the product formed versus noise trend for each abundance level represents the noise dependence at this enzyme abundance.

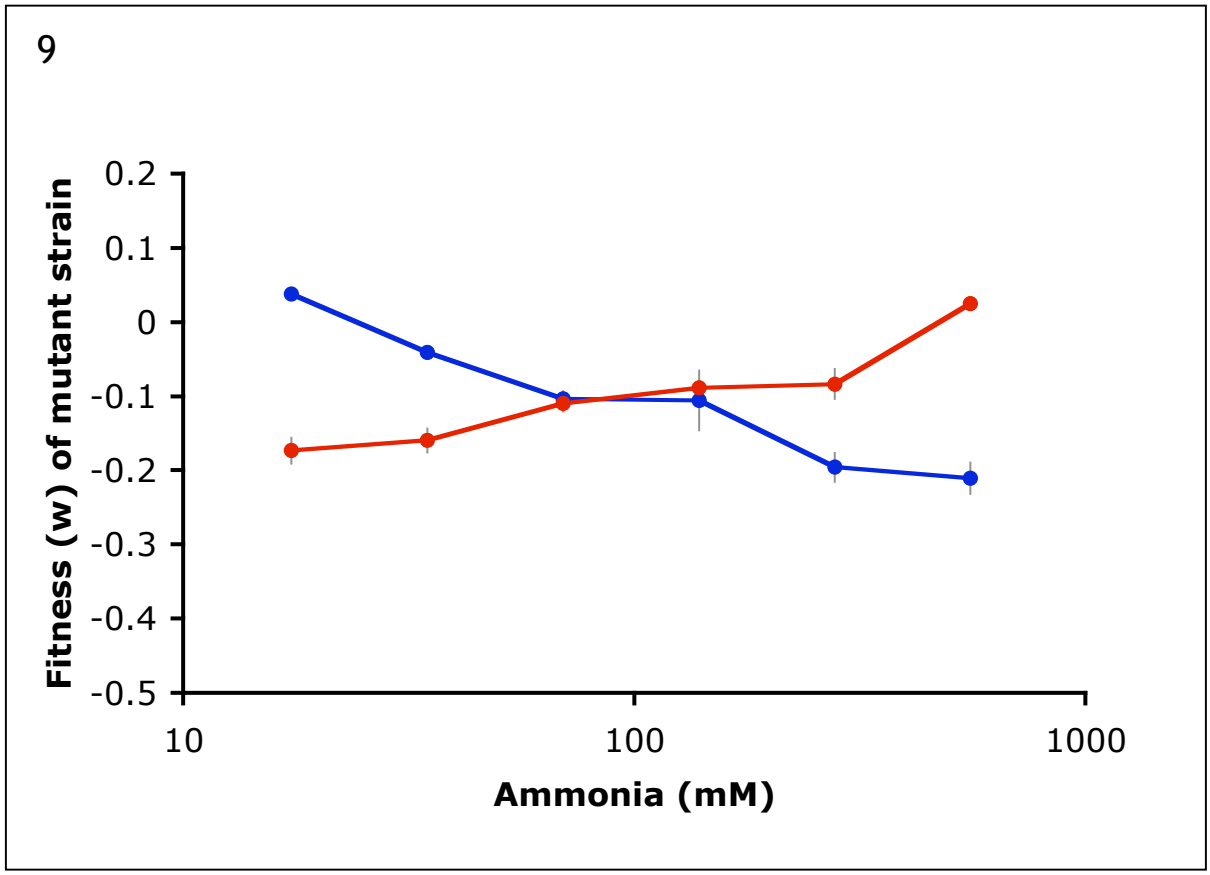
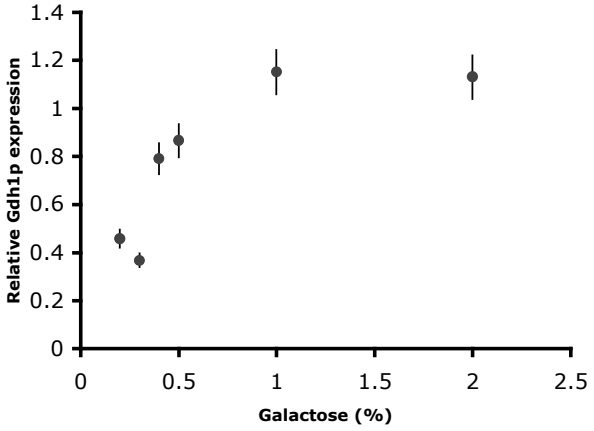
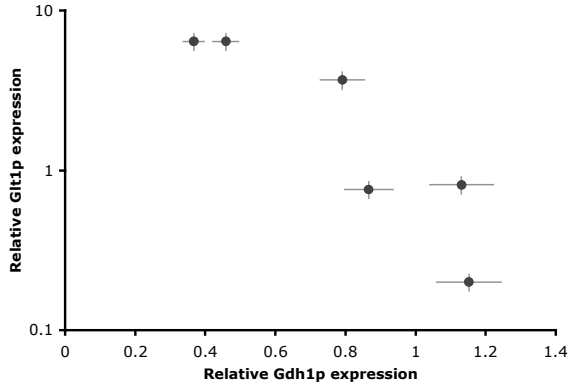


Figure 9. Fitness of the rate-deficient D150H (red) and rate-enhanced C313S (blue) strains relative to the parent strain across varying ammonia concentrations. Fitness was assayed as above.

10a


Relative Gdh1p expression
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
2.5
Galactose (%)

10b


Relative Glt1p expression
10
1
0.1
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Relative Gdh1p expression

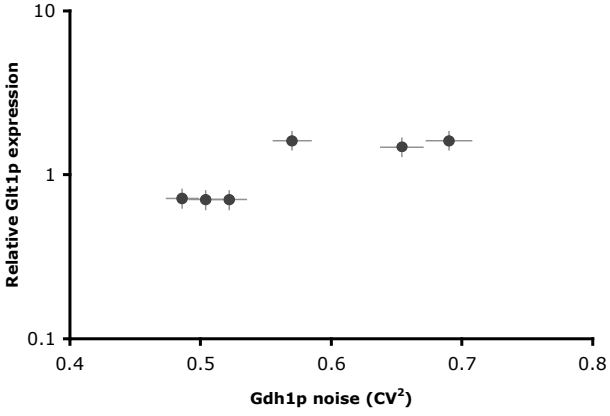
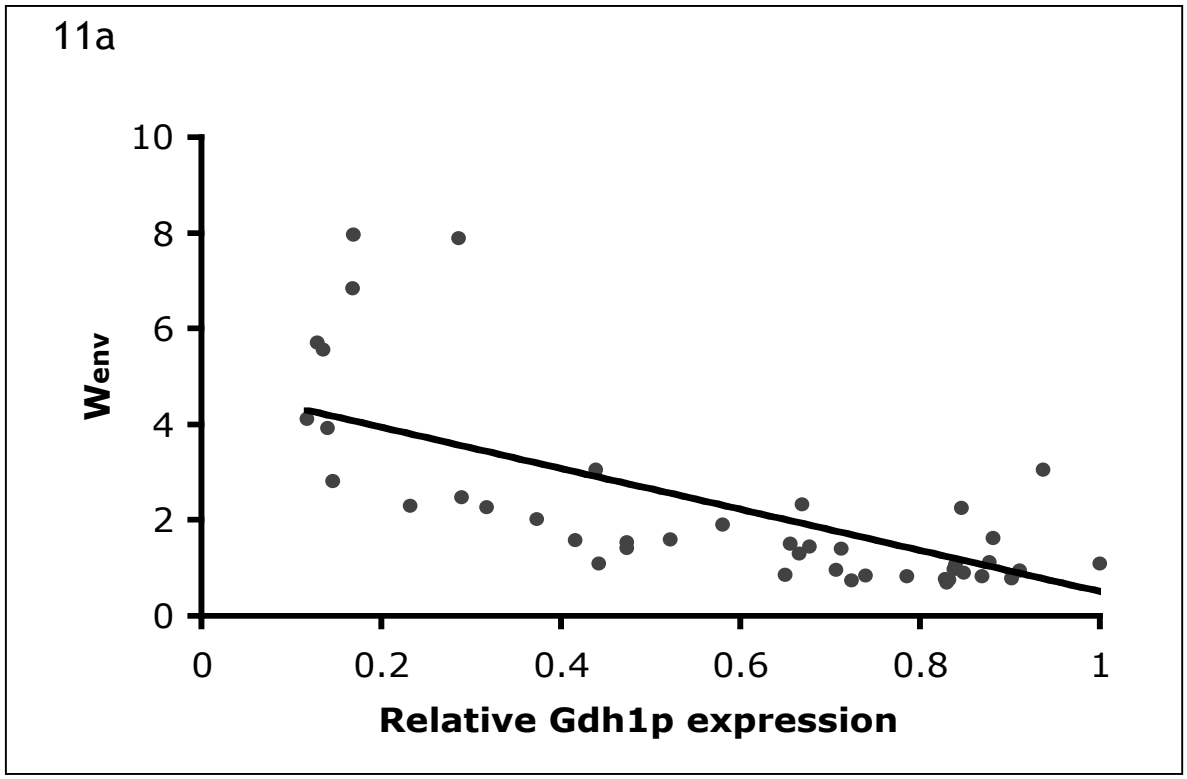

Relative Glt1p expression
10
1
0.1
0.4
0.5
0.6
0.7
0.8
Gdh1p noise (CV²)

Figure 10a Tunable Gdh1p expression as a function of galactose concentration for the engineered P_{GAL}-*GDH1* strain. Relative Gdh1p:GFP levels were measured as before and are reported relative to the parent strain. The mean ±s.d. from at least three independent experiments is shown. **b**, **Glt1p expression varies with Gdh1p abundance rather than noise. Top,** Abundance of *GLT1* mRNA as a function of Gdh1p abundance. The endogenous *GDH1* promoter was replaced with the *GAL1-10* promoter to achieve galactose tunable expression of Gdh1p. Cells were grown in varying galactose concentrations and Gdh1p:GFP expression was measured by flow cytometry as above. Cells were then harvested and *GLT1* mRNA was measured by qRT-PCR. *GLT1* transcript levels show an inverse relationship with Gdh1p abundance, suggesting that cellular regulatory mechanisms act to balance the expression of Gdh1p and Glt1p. **Bottom,** Abundance of *GLT1* mRNA as a function of noise in Gdh1p expression. The engineered galactose tunable Dal80p strain (P_{GAL}-*DAL80*), was used to modulate noise in Gdh1p expression while keeping abundance levels relatively unchanged. *GLT1* transcript levels show little change with varying Gdh1p noise. The mean ±s.d. from at least three independent experiments is shown.


11a
10
8
6
4
2
0
Wenv
0
0.2
0.4
0.6
0.8
1
Relative Gdh1p expression

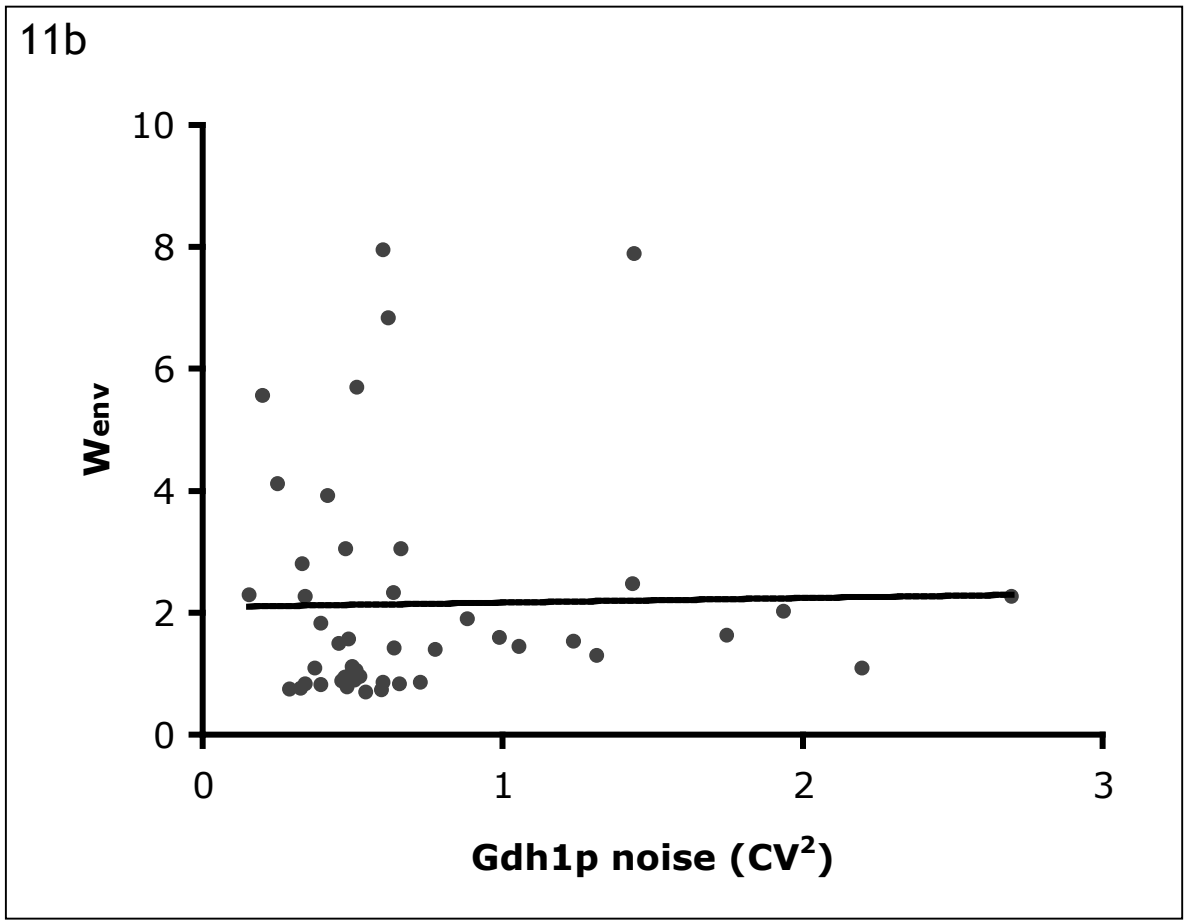
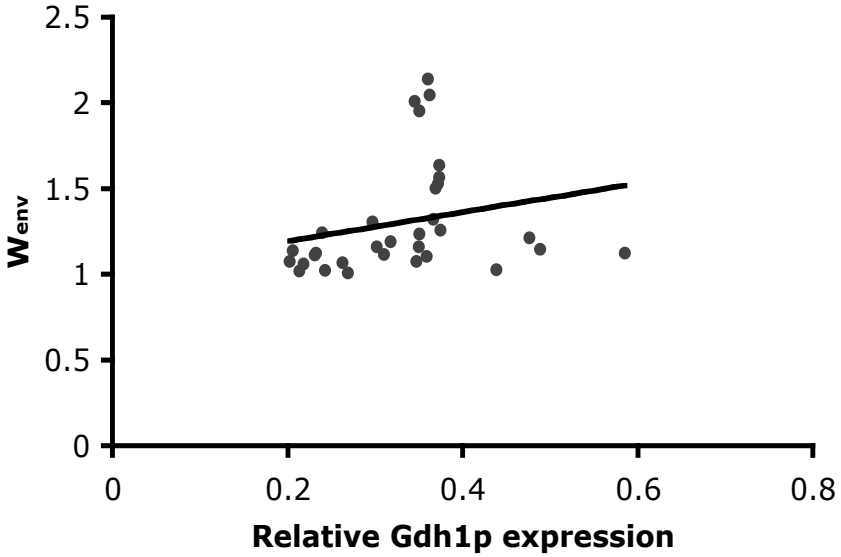

11b
10
8
6
4
2
0
Wenv
0
1
2
3
Gdh1p noise (CV2)

Figure 11 The alternate assimilation enzyme Glt1p determines the effect of Gdh1p noise and abundance on fitness. a, Abundance in W_{env} versus Gdh1p expression for a sampling of strains harboring the *GDH1* promoter library in a Gdh1p:GFP *GLT1Δ* background. W_{env} shows a negative correlation with abundance. **b,** W_{env} versus noise in Gdh1p expression for the same mutant set. W_{env} shows little correlation with noise in this mutant set.

12a


2.5
2
1.5
1
0.5
0
W_{env}
0
0.2
0.4
0.6
0.8
Relative Gdh1p expression

12b

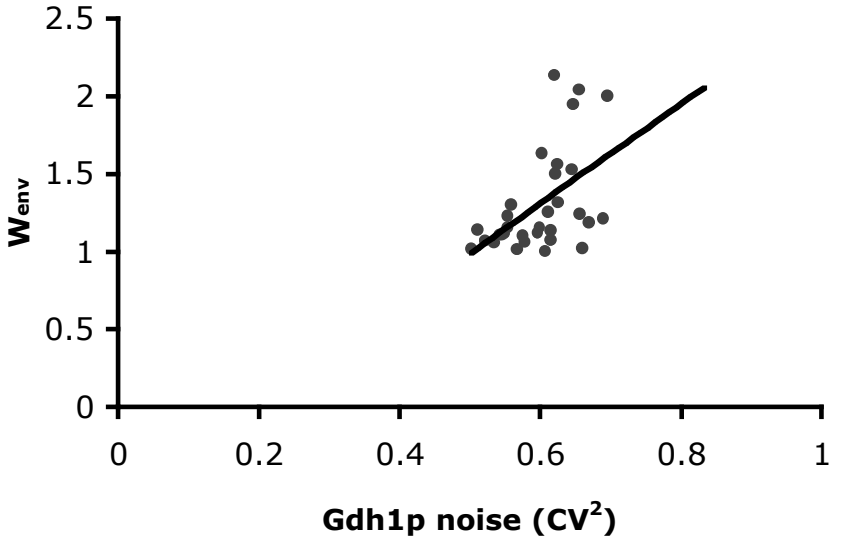

2.5
2
1.5
1
0.5
0
W_{env}
0
0.2
0.4
0.6
0.8
1
Gdh1p noise (CV^2)

Figure 12 W_{env} versus noise and abundance for *GDH1* promoter mutants in the wildtype background. a, Abundance in W_{env} versus Gdh1p expression for a sampling of strains harboring the *GDH1* promoter library in a Gdh1p:GFP background. W_{env} shows little correlation with abundance. The mean ±s.d. from at least three independent experiments is shown. **b,** W_{env} versus noise in Gdh1p expression for the same mutant set. W_{env} shows stronger correlation with noise in this mutant set.

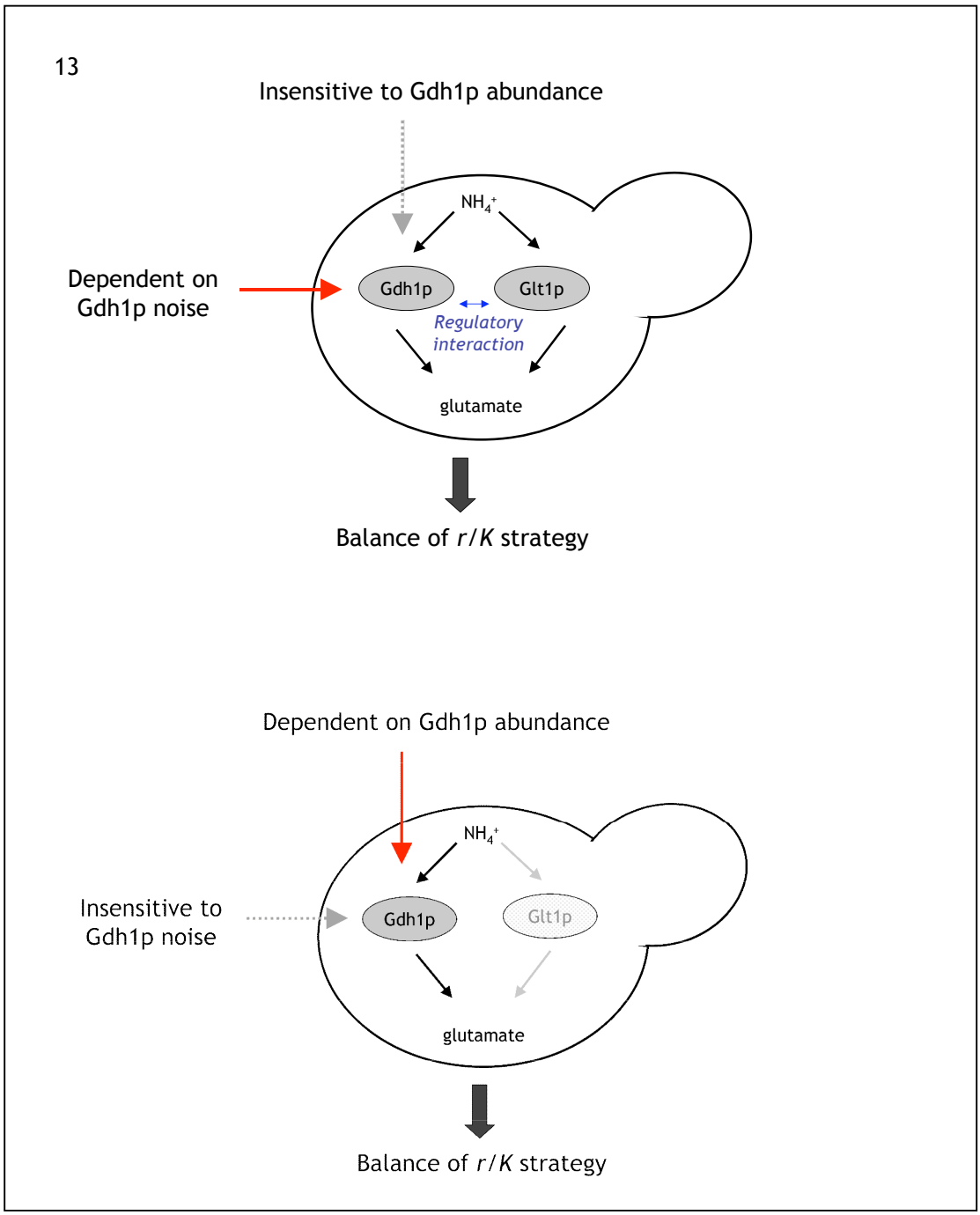


Figure 13. Representation of the effects of Gdh1p noise and abundance in the presence or absence of Glt1p. In the presence of both pathways noise in Gdh1p affects fitness because (uncharacterized) regulatory networks enable Glt1p to compensate for changes in Gdh1p abundance. In the absence of Glt1p the mean abundance of Gdh1p determines fitness.

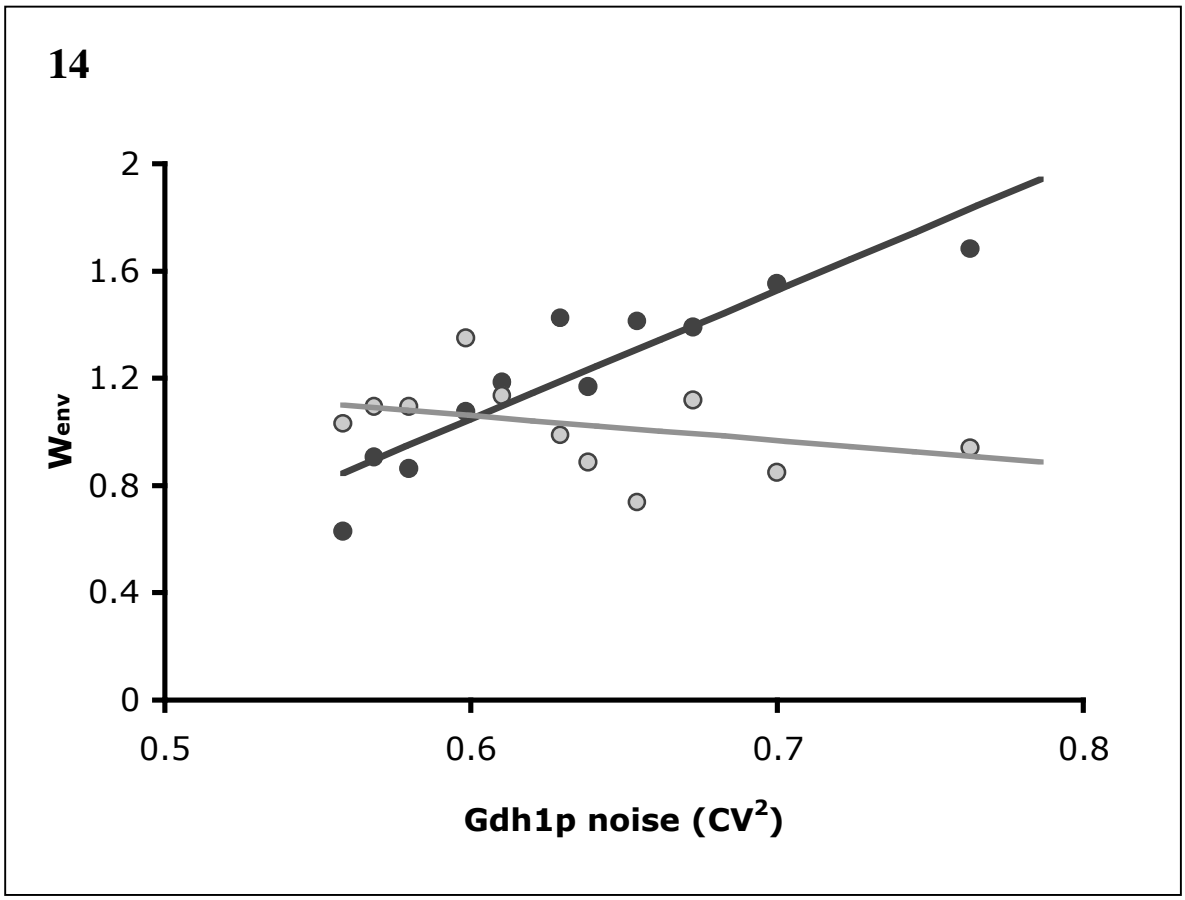
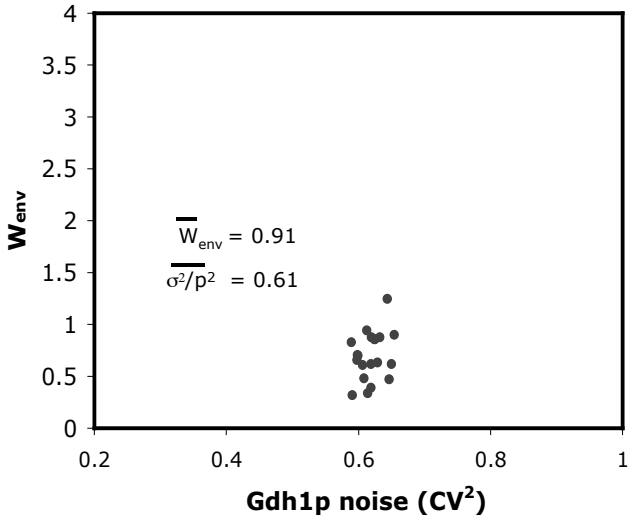
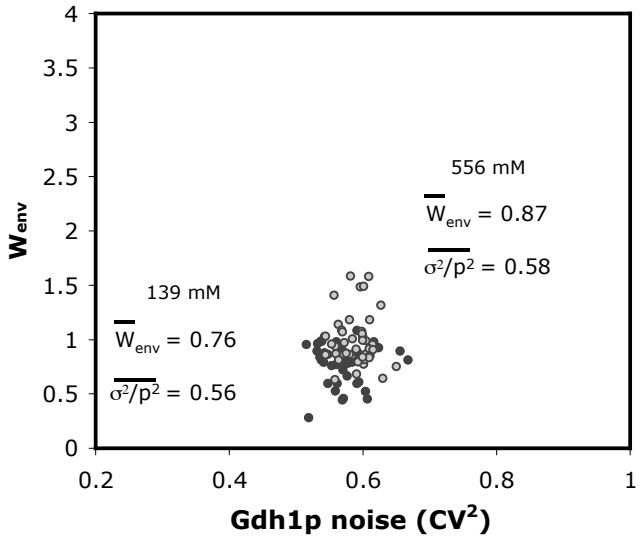


Figure 14 Overexpression of Glt1p lowers W_{env} dependence on noise. W_{env} as a function of Gdh1p noise in a set of *GDH1* promoter mutants transformed with a Glt1p overexpression plasmid (grey). This mutant promoter set transformed with an empty plasmid is shown for comparison (black). The trend of W_{env} with noise observed in the mutant set under wildtype Glt1p expression levels is diminished under Glt1p overexpression. The mean ±s.d. from at least three independent experiments is shown.

15a


W_{env}
$\overline{W_{env}}$ = 0.91
$\overline{\sigma^2/p^2}$ = 0.61
Gdh1p noise (CV^2)

15b


W_{env}
556 mM
$\overline{W_{env}}$ = 0.87
$\overline{\sigma^2/p^2}$ = 0.58
139 mM
$\overline{W_{env}}$ = 0.76
$\overline{\sigma^2/p^2}$ = 0.56
Gdh1p noise (CV^2)

15c

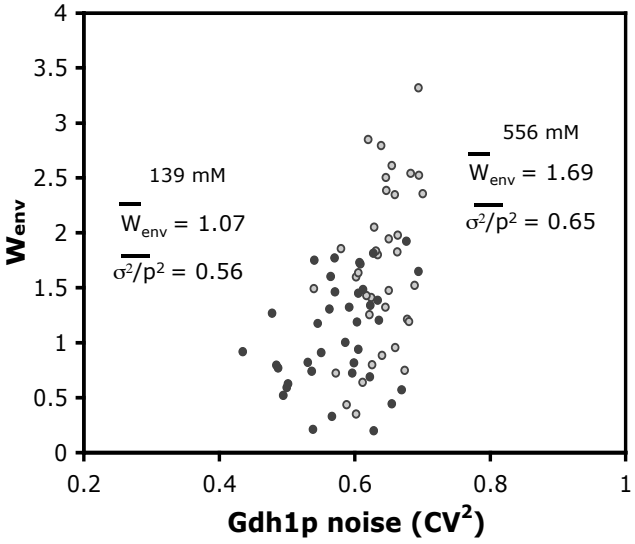

W_{env}
556 mM
$\overline{W_{env}}$ = 1.69
$\overline{\sigma^2/p^2}$ = 0.65
139 mM
$\overline{W_{env}}$ = 1.07
$\overline{\sigma^2/p^2}$ = 0.56
Gdh1p noise (CV^2)

Figure 15 Environmental selection pressure shapes Gdh1p noise in adapted populations. a, W_{env} versus noise for the initial library (t = 0). Mean W_{env} and Gdh1p noise values for the population are shown (n = 30). **b,** W_{env} versus noise for the Day 3 population (~36 generations) from the 139 mM ammonia (grey, n = 48) and 556 mM ammonia (black, n = 45) selection conditions. **c,** Noise versus W_{env} for the Day 5 population (~60 generations, n = 48 for both conditions). All errors are within 5% of the reported values.

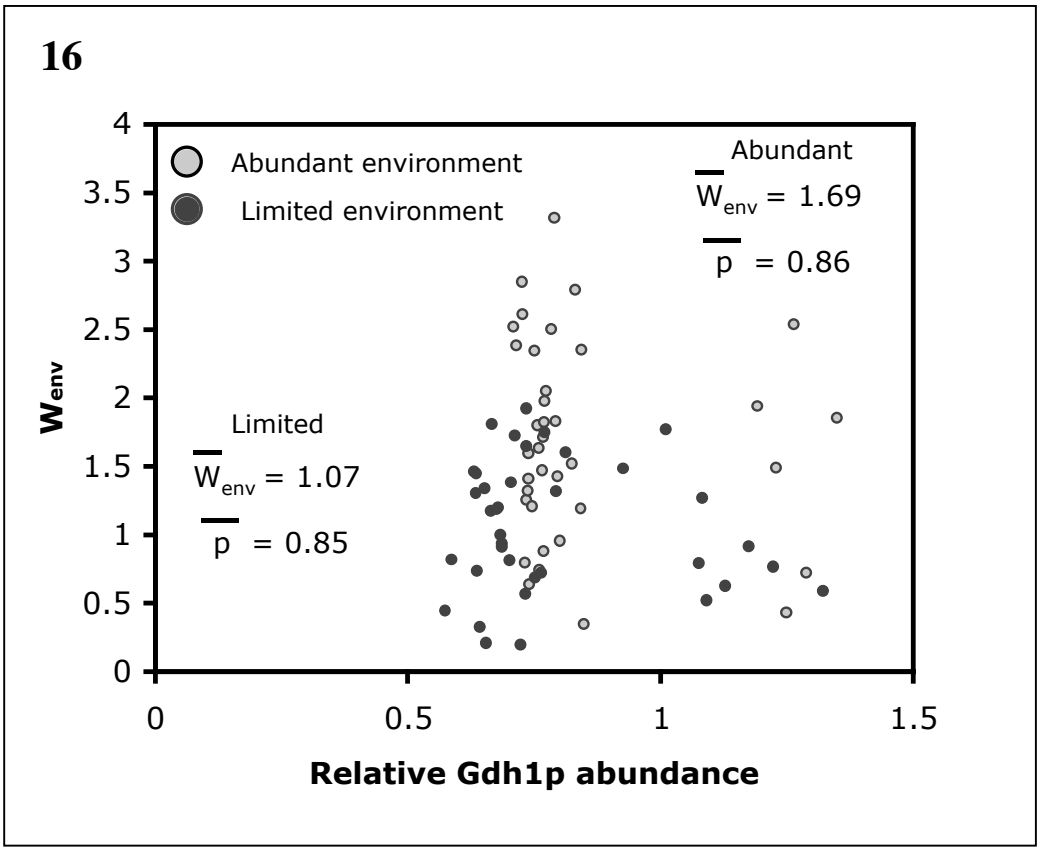
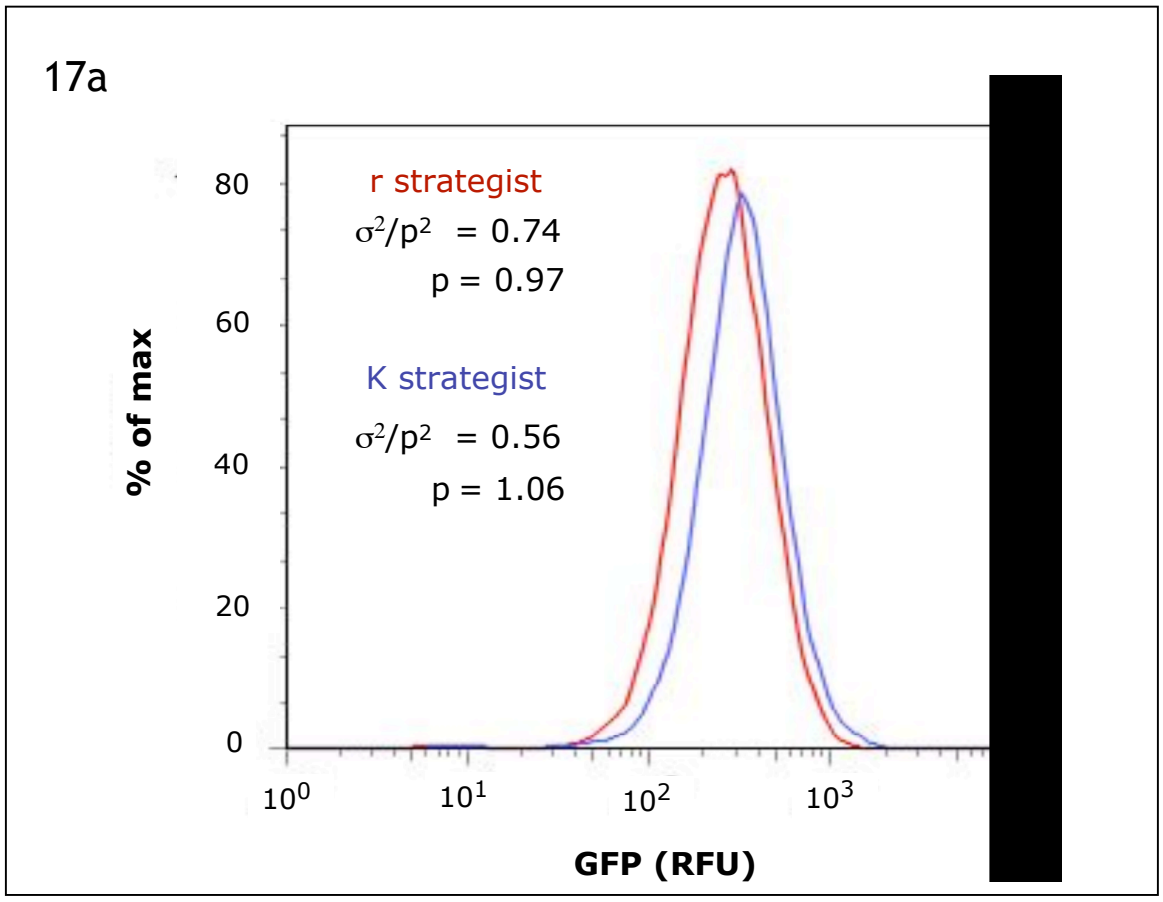


Figure 16 Environmental selection pressure does not affect Gdh1p abundance. Mean Gdh1p abundance versus W_{env} for the Day 5 adapted populations (~60 generations) selected in ammonia-abundant (grey) and ammonia-limited (black) conditions. The population-averaged Gdh1p abundance for individual clones from each environment is shown. Differences in average Gdh1p abundance for each population are not significant (mean Gdh1p = 0.86 relative to the parent strain for the abundant environment, mean Gdh1p = 0.86 relative to the parent strain for the limited environment, P = 0.72), in contrast to the differences observed in average Gdh1p noise. Each point was measured in triplicate and error was within 5% of the reported value.


17a
r strategist
σ²/p² = 0.74
p = 0.97
K strategist
σ²/p² = 0.56
p = 1.06
% of max
80
60
40
20
0
10⁰
10¹
10²
10³
GFP (RFU)

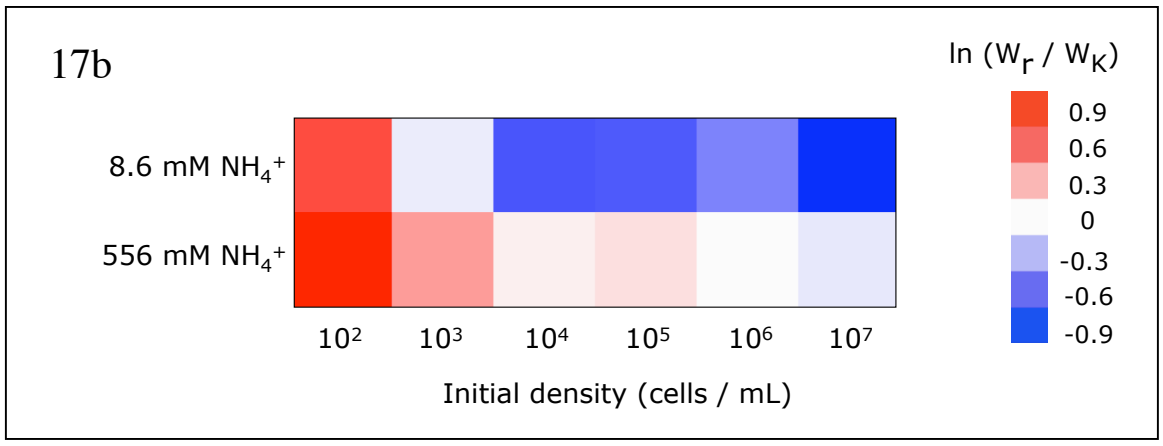

17b
ln (W_r / W_K)
0.9
0.6
0.3
0
-0.3
-0.6
-0.9
8.6 mM NH4+
556 mM NH4+
10²
10³
10⁴
10⁵
10⁶
10⁷
Initial density (cells / mL)

Figure 17. a, Representative histogram of Gdh1p expression in the r (red) and K (blue) strains. Abundance is reported relative to the parent strain. **b,** Density-dependent fitness for the r and K strains in poor and abundant ammonia environments. Fitness is reported as the natural log ratio of r strain fitness to K strain fitness and is represented as red and blue shading. Experiments were performed in triplicate and error is within 5% of the reported values.

Chapter 4.

Yeast Use a Tit-For-Tat Strategy in Ammonia Metabolism to Establish Cooperation

Explaining the emergence of cooperation is a major goal of evolutionary biology.[1-3] Most explanations rely on kin selection,[4] spatial isolation,[5] or rational policing.[6] However, recent theoretical work[7] has suggested that cooperation can evolve via the game theoretic tit-for-tat (TFT) strategy,[8-12] which is able to invade a population of cheaters and is itself resistant to invasion. Here, we found that yeast use a strategy resembling tit-for-tat in regulation of ammonia assimilation.[13] We first identified a tradeoff between maximum growth rate and ammonia utilization efficiency, which creates an opportunity for social conflict in microbial populations.[14-18] Efficient use of resources with a correlated tradeoff in growth rate is regarded as cooperation in microbes, while inefficient use of resources with high growth rate is regarded as cheating.[19, 20] We found that yeast use ammonia efficiently when ammonia is abundant (e.g., they cooperate when resources are abundant, which would indicate cooperation from other cells) and switch to inefficient growth in low ammonia (e.g., they defect when resources are limited, which indicates that other cells may be competing for ammonia). Competition experiments in batch culture with a cheater mutant confirmed that no special conditions (such as spatial isolation) is needed for the TFT strain to invade a cheater population. This data shows that the TFT strategy is a viable mechanism for the emergence of cooperation, even in simple organisms. In addition, this is one of the first demonstrations that microbes use genetic regulation of metabolism to play a game theory strategy.

Cooperation is widespread in nature, yet it remains difficult to explain how it might have emerged in populations via natural selection. This is demonstrated in the "tragedy of the commons": efficient use of a common resource at a cost to an individual can benefit selfish individuals (who incur no costs). Game theory has modeled this situation as the Prisoner's Dilemma, where a player is faced with either cooperation with another player or defection. Both obtain a payoff for mutual cooperation and a lower payoff for mututal defection. If one player defects while the other cooperates, the defector (also known as the cheater) receives the highest payoff while the cooperator receives the lowest payoff (the "sucker's payoff"). Thus, the matrix of fitness payoffs in a Prisoner's Dilemma dictates that defection will overcome cooperation in a population with both strategies. Alternative strategies become possible in the repeated Prisoner's Dilemma, yielding a variety of cooperation and defection decisions depending on the strategy encountered.

While cooperation and social interactions are usually considered in rational agents, microbes have proven valuable model systems for understanding how the use of resources may (or may not) lead to cooperation.[15, 16, 18, 21] Thermodynamic first principles dictate that organisms generally face a tradeoff between rate and yield in metabolic pathways.[15, 18, 22] For example, Pfeiffer and Bonhoeffer have described the tradeoff between the rate of adenosine triphosphate (ATP) production and the yield of ATP production in heterotrophic organisms, and how organisms that produce ATP efficiently can be considered altruistic cooperators. The ATP rate/yield tradeoff is apparent in many microorganisms that use both fermentation and respiration to metabolize glucose. Fermentation of glucose proceeds faster than respiration, but yields less

ATP per glucose (2 versus 32 ATP), meaning that inefficient fermenting strains will be able to outcompete slower but altruistic respiring strains. MacLean and Gudelj[15] have used fermentation and respiration mutants in *Saccharomyces cerevisiae* to demonstrate how competition and cooperation between strain can be influenced by the spatial and temporal parameters of the environment. These and other experimental studies have shown that tradeoffs between rate and efficiency of resource metabolism drive the emergence of metabolic strategies such as cooperation or defection. However, more complex strategies have not been seen observed, which is somewhat surprising given the complexity of metabolic regulatory circuits.[23] As noted by MacLean and Gudelj, the ability to regulate metabolic pathways according to environmental conditions could allow for more complex competitive strategies to arise.

We sought to investigate how the ability to switch metabolic strategies depending on the environment could affect fitness in competition with other strategies. We had previously characterized the fitness of *S. cerevisiae GDH1* promoter mutants that showed tradeoffs between fitness in abundant ammonia and fitness in limiting ammonia. GDH1 is a glutamate dehydrogenase that is responsible for the majority of ammonia assimilation in yeast.[13] We hypothesized that the wildtype strain could use genetic regulation to optimize metabolism for specific ammonia environments. We first examined maximum growth rates (μ_{max}) in continuous culture[24, 25] for the wildtype laboratory strain and two mutants: a strain that showed high fitness in abundant ammonia (denoted as A), and one that showed low fitness in abundant ammonia (denoted as B). In abundant ammonia (> 2.5 g/L), the wildtype and B strain showed similar growth rates, while the growth rate of the A strain was several fold higher (**Fig.1a**). Growth rate

decreased with ammonia concentration for all strains, although the wildtype strain switched from a B-like to A-like rate as ammonia decreased.

We next examined the efficiency of ammonia utilization for each strain at different ammonia concentrations. Utilization efficiency is calculated as the one over the amount of ammonia consumed per unit biomass (Methods) such that high efficiency values indicate efficient use of ammonia per organism. We found that the efficiency of all strains increased with decreasing ammonia concentrations, although the A strain was consistently lower efficiency than the B strain (**Fig.1b**). At high ammonia concentrations the wildtype strain showed an efficiency similar to the B strain. At low ammonia concentrations the wildtype strain switched to relatively low efficiency, similar to the A strain. The tradeoff between growth rate and resource utilization efficiency is a clear situation for social conflict.[15, 18] The A strain shows the hallmarks of defector (or cheater) strains in that it has a high growth rate at the expense of efficiency. The B strain displays cooperator characteristic in that it uses ammonia efficiently (to the benefit of other cells) at a cost to itself (lower growth rate). We will thus refer to the A strain as the defector strain and the B strain as the cooperator strain.

As was previously observed, assays of Gdh1p gene expression variability[26-29] (noise) suggest a mechanistic link to growth rate-efficiency phenotypes. The cooperator strain showed low Gdh1p noise across each ammonia concentration, while the defector strain showed high noise (**Fig.1c**). The wildtype strain varied noise in Gdh1p expression according to ammonia concentration – in abundant ammonia it showed low noise, with increasing noise as external ammonia decreased.

This data further suggests that the wildtype strain switches from cooperator-similar growth to defector-similar growth.

We were surprised that the wildtype strain did not show a growth rate and efficiency profile similar to the defector strain. In studies of glucose metabolism, yeast are found to ferment any excess glucose to achieve high rates of ATP production in spite of ATP yield, likely to outcompete neighboring cells.[30-32] In contrast, the wildtype strain in abundant ammonia utilizes ammonia with high efficiency at a cost to growth rate, indicating cooperative behavior. According to evolutionary game theory, the existsence of this cooperation should be overwhelmed by the emergence of defecting mutants (such as the defector strain here). We propose that the wildtype strategy is analogous to the tit-for-tat (TFT) strategy in the Prisoner's Dilemma. TFT players cooperate in the first round of the Prisoner's Dilemma, and for each subsequent round does whatever its opponent did such that cooperation is met with cooperation and defection is met with defection (**Fig.2a**). Although several superior strategies have since been described,[11, 33] TFT remains a primary model for understanding reciprocal altruism. The wildtype strain cooperates when ammonia is abundant (>2.5 g/L), which could be an indicator that either there are no competitors in the environment or that there are other cooperators in the environment using ammonia efficiently (**Fig.2b**). The wildtype strain defects when ammonia is low (< 1.25 g/L) which could indicate that other strains are rapidly consuming ammonia. We note several caveats, such as the fact that yeast in this context are not engaged in a pair-wise contest and that there is a continuum of growth rates and efficiencies instead of a binary division between cooperation and defection. However, because microbes are rarely involved in pair-wise

competitions the wildtype strain and mutants may be a valuable system for understanding how cooperation can persist in populations.

There are several general theoretical predictions for the TFT strategy in competition with alternative strategies: that the strategy is resistant to invasion by a population of defectors, and that TFT is able to invade defectors in finite populations.[7] In particular, the ability of a TFT strategy to invade a population of defectors would be a clear demonstration that TFT is a route for the emergence of cooperation.[7] We were able to experimentally test these predictions by examining frequency dependent selection in batch culture. We chose a batch culture (or "seasonal"[15]) environment so that the dynamic fitness (as ammonia is consumed) could be assessed. We inoculated varying frequencies of wildtype (without the *GDH1:GFP* fusion) and defector strain at low density (10^3 cfu/mL) in batch culture and allowed the culture to reach stationary phase (48 hours of growth). We quantitated the frequencies of the wildtype and defector strain by plating the cultures on solid media and assaying for fluorescence (Methods). We found that at high initial wildtype frequencies (> 0.5) the defector strain showed little ability to invade, evidenced by the nearly neutral wildtype fitness (w) observed in these competitions (**Fig.3**). At low initial wildtype frequencies (< 0.3), the wildtype strain showed positive fitness values, indicating that it was able to invade the population of defectors. These conditions are analogous to the immigration or emergence of a small subpopulation of TFT players into a population of defectors, and show that TFT can indeed invade a population of individuals selfishly using resources.

The above data shows that yeast are able to play a TFT-like strategy by adjusting the growth rate and ammonia utilization efficiency according to external ammonia concentrations. This strategy is notable because the wildtype strain does not have the optimal growth rate (compared to the defector mutant) or the optimal efficiency (compared to the cooperator mutant) for a wide range of ammonia environments. Instead, the wildtype strain has a regulatory scheme well suited for competing with alternative metabolic strategies in dynamic fitness landscapes.[34] We believe that this is the first demonstration of a game theoretic strategy being played in a microbial population. In addition, this work suggests that the control of metabolism in response to environmental conditions is a route for the emergence of cooperation.

Methods Summary

Strains and media. All strains were derivatives of the *GDH1:GFP* fusion strain of the S288C background. Cells were grown in synthetic complete media with 2% glucose and the indicated amount of ammonia by addition of ammonium sulfate. Construction and selection of the low-noise and high-noise *GDH1* mutants were described previously. Briefly, primers flanking 500 nucleotides upstream of the *GDH1* coding region (1043500 - 1043050, chromosome XV) were used to amplify the fragment from yeast genomic DNA. The fragment was diluted into mutagenic PCR buffer[35]. The *GDH1* fragment was assembled with a *LEU2* gene fragment transformed into yeast strains using a standard lithium acetate procedure.[36]

Continuous growth conditions and growth rate assay. Cells were inoculated in synthetic complete media with the appropriate ammonia concentration in a well-stirred vessel with a

working volume of 250mL maintained at 30 degrees. Cells were allowed to stabilize for 12 hours at a dilution rate of 0.2 hr^{-1}. To measure maximum growth rates (μ_{max}) the washout method[24, 25] was used: when the dilution rate of the chemostat is greater than μ_{max} the cell number decreases by the expression $\ln X = (\mu_{max} - D)t + \ln X_0$; where X is the cell number after time t, X_0 is the initial cell number and D is the dilution rate. We increased the dilution rate to 4.0 hr^{-1} and collected samples at regular time points. Cell number was quantified by OD_{600} and by serial dilution and plating on YPD-agar.

Ammonia utilization assays. Cells were grown in continuos culture as above at low dilution (0.1 hr^{-1}) to standardize growth rates. Cells were collected from the outflow spun down. The supernatant was decanted into 14mL tubes, capped with a rubber stopper, and incubated at room temperature for 30 minutes. Ammonia was quantitated by gas chromatography – mass spectrometry (GC-MS) which can be used for accurately assaying volatile compounds such as ammonia.[37] The GC-MS system consisted of a model 6850 Series II Network GC system (Agilent) and model 5973 Network mass selective system (Agilent). Oven temperature was programmed from 50 degrees (1 min) to 70 degrees (10 degrees / min). 100 μL of culture headspace was withdrawn through the rubber stopper with a syringe and manually injected into the GC-MS. Samples were confirmed as ammonia by comparison with commercially obtained standard, which had a retention time of 1.50 minutes. Ammonia in the headspace was correlated to ammonia in the supernatant by a standard curve. Efficiency is reported as one over the milligrams of ammonia consumed per 10^6 cells.

Measurement of abundance and noise values through flow cytometry. Two gates were used to standardize each cell population. The first gate isolated cells displaying regular morphology based on electronic volume and side-scatter, while the second gate removed non-fluorescent cells from the distribution. This gating method was compared against other methods previously described and the abundance and noise trends observed were consistent between methods.[38, 39] Noise was calculated as the square of the coefficient of variation (σ^2/p^2) of the distribution[26]. Abundance was calculated as the mean of the distribution. 50,000 events were analyzed to calculate noise for each sample. Noise trends were similar when calculated as the coefficient of variation (σ/p) and the variance (σ^2).

Competition assays and fitness. The defector strain and a wildtype S288c strain without the *GDH1:GFP* fusion were grown overnight and diluted to 10^3 cells/mL. The *GDH1:GFP* construct was found to have no fitness effect (data not shown) Cultures were mixed in varying ratios in 2mL of synthetic complete media with 5 g/L ammonium sulfate. Cultures were incubated at 30 degrees with 250 rpm shaking for 48 hours. Cultures were diluted and plated onto YPD-agar and grown for 48 hours. Individual colonies were resuspended in 100μL media and GFP fluorescence was assayed using a Tecan plate reader. 96 colonies were assayed for each competition. Fitness of the wildtype strain is reported as the natural log of the ratio of its final frequency to its initial frequency, $w = \ln(f_{final}/f_{initial})$,[15, 40] such that a values > 0 imply that the wildtype strain increased in frequency versus the defector strain over the competition, while values < 0 imply that it decreased in frequency.

REFERENCES

1. Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. The evolution of cooperation. Q Rev Biol 79, 135-60 (2004).
2. Trivers, R. L. Evolution of Reciprocal Altruism. Quarterly Review of Biology 46, 35-& (1971).
3. Dugatkin, L. Cooperation Among Animals (Oxford University Press, Oxford, UK, 1997).
4. Diggle, S. P., Griffin, A. S., Campbell, G. S. & West, S. A. Cooperation and conflict in quorum-sensing bacterial populations. Nature 450, 411-4 (2007).
5. Hauert, C. Spatial effects in social dilemmas. J Theor Biol 240, 627-36 (2006).
6. Frank, S. A. Mutual policing and repression of competition in the evolution of cooperative groups. Nature 377, 520-2 (1995).
7. Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. Nature 428, 646-50 (2004).
8. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. Science 211, 1390-6 (1981).
9. Axelrod, R. M. The evolution of cooperation (Basic Books, New York, 1984).
10. Milinski, M. TIT FOR TAT in sticklebacks and the evolution of cooperation. Nature 325, 433-5 (1987).
11. Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature 364, 56-8 (1993).
12. Nowak, M. & Sigmund, K. Chaos and the evolution of cooperation. Proc Natl Acad Sci U S A 90, 5091-4 (1993).
13. Magasanik, B. Ammonia assimilation by Saccharomyces cerevisiae. Eukaryot Cell 2, 827-9 (2003).
14. Gudelj, I., Beardmore, R. E., Arkin, S. S. & MacLean, R. C. Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments. J Evol Biol 20, 1882-9 (2007).
15. MacLean, R. C. & Gudelj, I. Resource competition and social conflict in experimental populations of yeast. Nature 441, 498-501 (2006).
16. Novak, M., Pfeiffer, T., Lenski, R. E., Sauer, U. & Bonhoeffer, S. Experimental tests for an evolutionary trade-off between growth rate and yield in E. coli. Am Nat 168, 242-51 (2006).
17. Pfeiffer, T. & Bonhoeffer, S. An evolutionary scenario for the transition to undifferentiated multicellularity. Proc Natl Acad Sci U S A 100, 1095-8 (2003).
18. Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. Science 292, 504-7 (2001).
19. Rankin, D. J., Bargum, K. & Kokko, H. The tragedy of the commons in evolutionary biology. Trends Ecol Evol 22, 643-51 (2007).
20. MacLean, R. C. The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. Heredity 100, 233-9 (2008).
21. Turner, P. E. & Chao, L. Prisoner's dilemma in an RNA virus. Nature 398, 441-3 (1999).
22. Stucki, J. W. The optimal efficiency and the economic degrees of coupling of oxidative phosphorylation. Eur J Biochem 109, 269-83 (1980).
23. Ihmels, J., Levy, R. & Barkai, N. Principles of transcriptional control in the metabolic network of Saccharomyces cerevisiae. Nat Biotechnol 22, 86-92 (2004).

24. Molin, G. Measurement of the Maximum Specific Growth-Rate in Chemostat of Pseudomonas Spp with Different Abilities for Biofilm Formation. European Journal of Applied Microbiology and Biotechnology 18, 303-307 (1983).
25. Pirt, S. Principles of microbe and cell cultivation (Blackwell Scientific Publications, London, 1975).
26. Paulsson, J. Summing up the noise in gene networks. Nature 427, 415-8 (2004).
27. Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. Nature 420, 231-7 (2002).
28. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. Science 309, 2010-3 (2005).
29. Samoilov, M. S., Price, G. & Arkin, A. P. From fluctuations to phenotypes: the physiology of noise. Sci STKE 2006, re17 (2006).
30. Otterstedt, K. et al. Switching the mode of metabolism in the yeast Saccharomyces cerevisiae. EMBO Rep 5, 532-7 (2004).
31. Frick, T. & Schuster, S. An example of the prisoner's dilemma in biochemistry. Naturwissenschaften 90, 327-31 (2003).
32. Aledo, J. C., Perez-Claros, J. A. & Esteban del Valle, A. Switching between cooperation and competition in the use of extracellular glucose. J Mol Evol 65, 328-39 (2007).
33. Imhof, L. A., Fudenberg, D. & Nowak, M. A. Tit-for-tat or win-stay, lose-shift? J Theor Biol 247, 574-80 (2007).
34. Pfeiffer, T. & Schuster, S. Game-theoretical approaches to studying the evolution of biochemical systems. Trends Biochem Sci 30, 20-5 (2005).
35. Cadwell, R. C. & Joyce, G. F. Mutagenic PCR. PCR Methods Appl 3, S136-40 (1994).
36. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol 350, 87-96 (2002).
37. Blunden, J., Aneja, V. P. & Lonneman, W. A. Characterization of non-methane volatile organic compounds at swine facilities in eastern North Carolina. Atmospheric Environment 39, 6707-6718 (2005).
38. Newman, J. R. et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441, 840-6 (2006).
39. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. Science 304, 1811-4 (2004).
40. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet 4, 457-69 (2003).

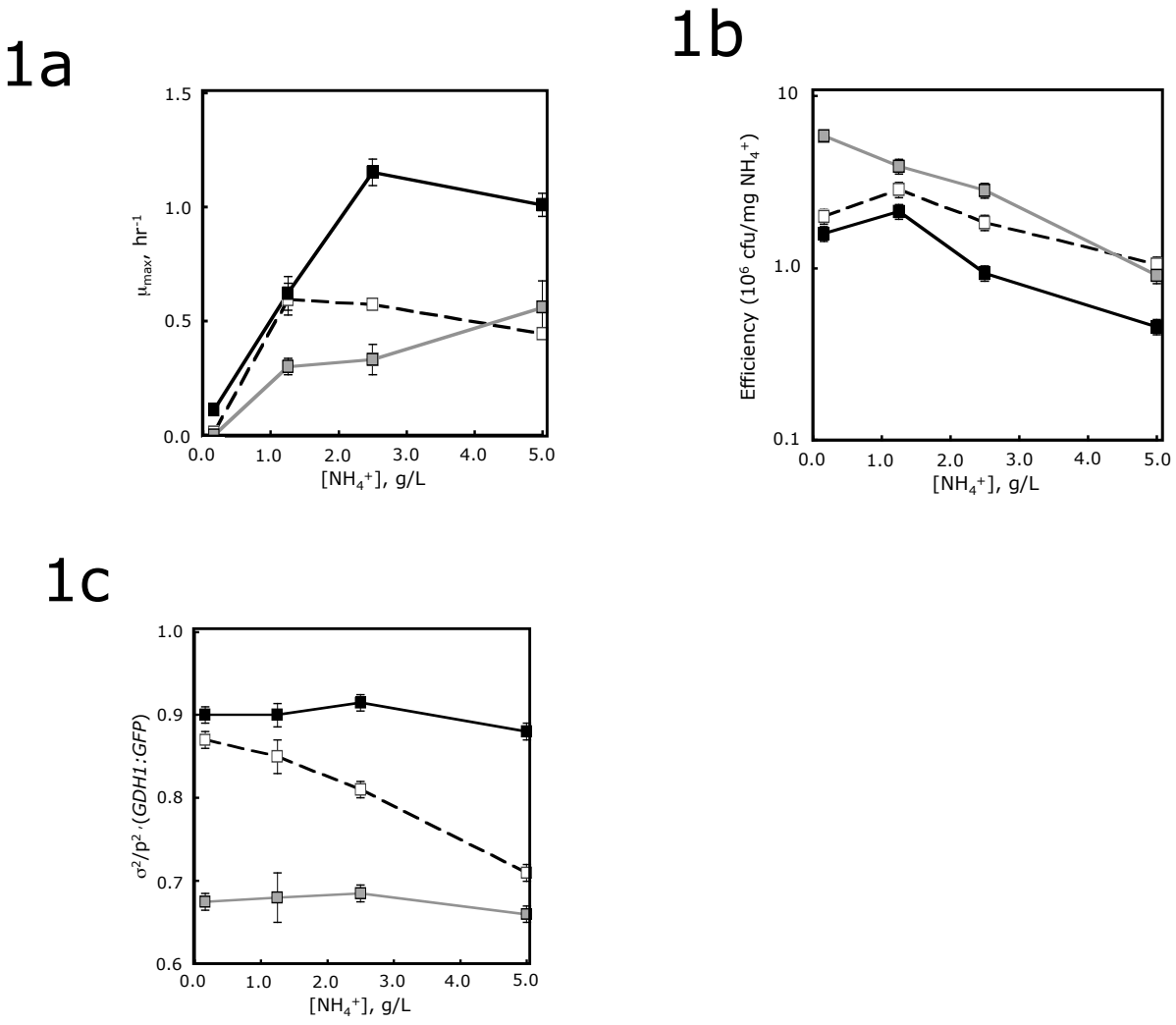


Figure 1. Characterization of wildtype and mutant strains. a, growth rate in continuous culture for the wildtype strain (open squares), A strain (black squares, previously selected for high fitness in 5 g/L ammonia), and B strain (gray squares, selected for low fitness in 5 g/L ammonia). Growth was measured using the washout method[24, 25] at each ammonia concentration. The A strain showed the highest growth rate in all concentrations while the B strain was consistently low. The wildtype strain showed rates similar to the B strain at high ammonia and rates similar to the A strain at low ammonia. **b**, ammonia utilization efficiency, as measured by ammonia consumption per biomass. Data is shown on a log scale for clarity. The A strain and B strain show low and high efficiencies, respectively, while the wildtype strain switches between

high and low efficiency as ammonia decreases. **c**, assays of Gdh1p gene expression noise in each strain. The A strain and B strain display high and low noise, respectively, while the wildtype strain displays low noise at high ammonia and high noise at low ammonia. All measurements were performed at least three times and s.d. is shown.

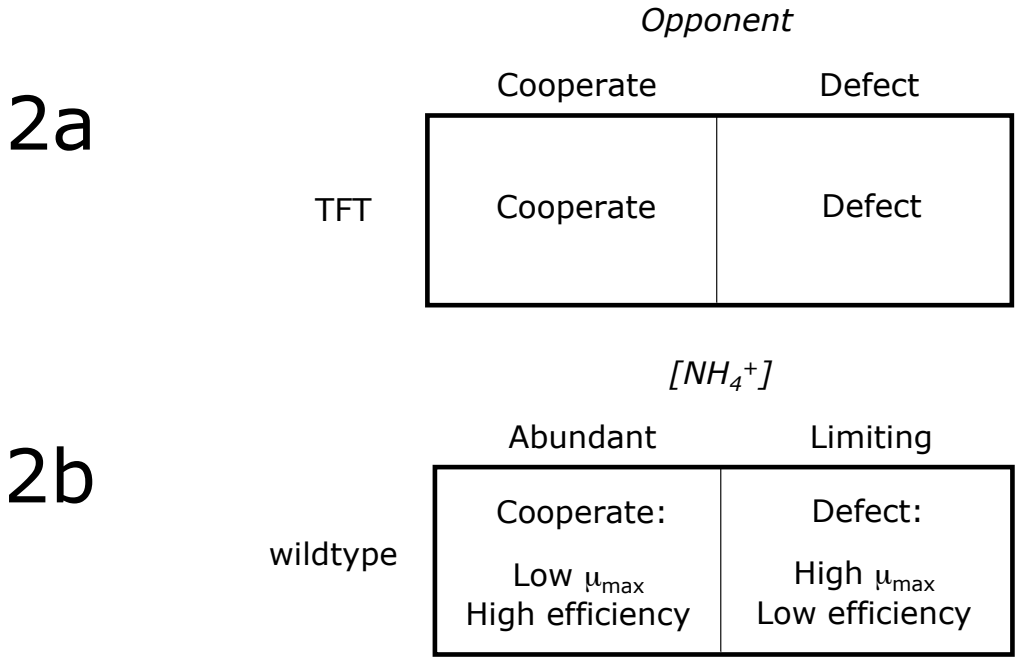


Figure 2. Conceptual model of TFT-like strategy in the wildtype strain. a, TFT strategy in the idealized Prisoner's Dilemma game. After cooperating in the first round, the TFT player does exactly as its opponent did in the last round. **b**, growth and ammonia efficiency strategy in the wildtype strain described here. In high ammonia the strain shows altruistic behavior with high utilization efficiency at expense of growth rate. In low ammonia the strain shows cheater behavior with high growth rate and low efficiency of ammonia utilization.

3

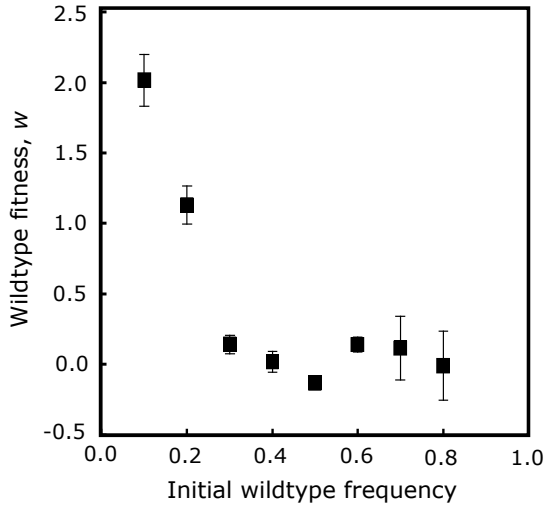


Figure 3. Direct competitions between the wildtype and defector strains. The frequency-dependence of the wildtype strain in competition with the defector strain (B strain) was measured by direct competition in batch culture. Relative frequencies were measured after competition and fitness is reported as w, the natural log of the ratio between final and initial wildtype frequency (Methods). At low initial frequencies (< 0.3) the wildtype strain showed w > 1 indicating that it was able to invade the defector population. In contrast, the defector strain was unable to invade a large population of the wildtype strain (wildtype frequency > 0.5) indicated by the non-negative w values for wildtype in those competitions. All measurements were performed in triplicate and s.d. is shown.

Appendix A.

Riboregulator Methods and Materials

All references are in Chapter 2.

Plasmid construction, cell strains, reagents. Standard molecular biology techniques were employed to construct all plasmids.[42] Four different plasmid constructs were generated by cloning into the pRS314-Gal and pRS316-Gal shuttle plasmids.[43] Genes and antiswitch constructs were cloned into multi-cloning sites, downstream of a GAL1 promoter. These plasmids contain an *E. coli* origin of replication (f1) and selection marker for ampicillin resistance, as well as a *S. cerevisiae* origin of replication (CEN6-ARSH4) and selection markers for tryptophan (TRP1-pRS314) and uracil (URA3-pRS316) biosynthetic genes in order to select cells harboring these plasmids in synthetic complete media supplemented with the appropriate amino acid dropout solution.[42] In the first plasmid system, pTARGET1, yEGFP was cloned into the multi-cloning site and is located between a GAL1 promoter and ADH1 terminator. In the second plasmid system, pSWITCH1, various antiswitches were cloned between two hammerhead ribozymes which are located between a GAL1 promoter and ADH1 terminator. In the third plasmid system, pTARGET2, a P_{GAL}-Venus-$ADH1_{term}$ construct was cloned downstream of the P_{GAL}-yEGFP-$ADH1_{term}$ construct in pTARGET1. Therefore, pTARGET2 produces two target transcripts when induced with galactose. In the fourth system, pSWITCH2, a P_{GAL}-antiswitch-$ADH1_{term}$ construct was cloned downstream of the P_{GAL}-antiswitch-$ADH1_{term}$ construct in pSWITCH1. Therefore, pSWITCH2 produces two antiswitch constructs when induced by the presence of galactose. Two sets of plasmids, pTARGET1 and pSWITCH1 or pTARGET2 and pSWITCH2, were transformed into *S. cerevisiae* simultaneously and maintained with the appropriate nutrient selection pressure. In these

two plasmid sets, expression of antiswitch constructs and their targets was induced upon the addition of galactose to the media. Oligonucleotide primers were purchased from Integrated DNA Technologies. All genes and antiswitches were PCR amplified in a Dyad PCR machine (MJ Research) with Taq DNA polymerase (Roche). The *yegfp* gene was obtained from pSVA15,[35] and the *venus* gene was obtained from pCS2/Venus[38]. All antiswitch sequences were obtained using custom oligonucleotide design.

All plasmids were constructed using restriction endonucleases and T4 DNA ligase from New England Biolabs. Plasmids were screened by transforming into an electrocompetent *E. coli* strain, DH10B (Invitrogen; F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80d*lacZ*ΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ^- *rps*L *nup*G), using a Gene Pulser Xcell System (BioRAD) according to manufacturer's instructions. Subcloning was confirmed by restriction analysis. Confirmed plasmids were then transformed into the wild-type W303α *S. cerevisiae* strain (*MATα his3-11,15 trp1-1 leu2-3 ura3-1 ade2-1*) using standard lithium acetate procedures.[44] *E. coli* cells were grown on Luria-Bertani media (DIFCO) with 100 μg/ml ampicillin (EMD Chemicals) for plasmid selection, and *S. cerevisiae* cells were grown in synthetic complete media (DIFCO) supplemented with the appropriate dropout solution (Calbiochem). Plasmid isolation was done using Perfectprep Plasmid Isolation Kits (Eppendorf).

Protein expression assays. Yeast cells were inoculated into synthetic complete media supplemented with the appropriate drop out solution and sugar source (2% raffinose, 1% sucrose) and grown overnight at 30 °C. Cells were back diluted into fresh media to an OD_{600} of 0.1 and grown at 30 °C. For assaying antiswitch activity, this fresh media

contained appropriate concentrations of theophylline (Sigma), caffeine (Sigma), tetracycline (Sigma), or water (negative control), and expression was induced to a final concentration of 2% galactose, or an equivalent volume of water was added (noninduced control). After growing for 3 hours, the GFP and Venus levels were assayed on a Safire (Tecan) fluorescent plate reader set to the appropriate excitation (GFP- 485 nm; Venus- 515 nm) and emission (GFP- 515 nm; Venus- 508) wavelengths. For assaying the antiswitch temporal response, cells were back-diluted into fresh media containing 2% galactose. After growing in inducing media for 3 h, theophylline or water was added and fluorescence was monitored over time. Fluorescence was normalized for cell number by dividing relative fluorescence units (RFUs) by the OD_{600} of the culture.

RNA quantification. Yeast cells were grown according to methods detailed in protein expression assays. Total RNA was extracted using standard acid phenol extraction procedures.[45] Briefly, cells were pelleted and frozen in liquid nitrogen. Pellets were resuspended in a 50 mM NaOAc (pH 5.2) and 10 mM EDTA buffer. Cells were lysed by the addition of SDS to a final concentration of 1.6% and equal volume of acid phenol. Solutions were kept at 65 °C with intermittent vortexing for 10 min. Following cooling on ice, the aqueous phase was extracted, and further extraction was carried out with an equal volume of chloroform. RNA samples were ethanol precipitated and resuspended in water. Total RNA was quantified by OD_{260} readings. RNA samples were DNased (Invitrogen) according to manufacturer's instructions. cDNA was synthesized using gene-specific primers and Superscript III reverse transciptase (Invitrogen) according to the manufacturer's instructions. qRT-PCR was performed on this cDNA using an iCycler iQ system (BioRAD). Samples were prepared using the iQ SYBR green supermix and

primer pairs specific for different templates on dilution series of the cDNA, according to the manufacturer's instructions. Data was analyzed using the iCycler iQ software.

***In vitro* antiswitch affinity experiments.** Antiswitch and target sequences were amplified with primers containing a T7 polymerase promoter. RNA was transcribed using Ampliscribe T7 transcription kits (Epicentre) according to manufacture's instructions, except that transcription was carried out at 42 °C, and for gel-shift assays, antiswitches were radiolabeled by the addition of [α-^{32}P]-UTP to the transcription mix. The RNA was purified on a 15% denaturing gel, eluted, ethanol precipitated, and resuspended in water. RNA was quantified by OD_{260} readings. For nuclease mapping, antiswitches were 5' end labeled with fluorescein (Molecular Probes) by incubating 25 μg of RNA with phosphate reactive label in labeling buffer (0.12 M methylimidazole pH 9.0, 0.16 M EDAC) for 4 hours, according to manufacturer's instructions. Labeled RNA was purified by ethanol precipitation and run on a 12% denaturing gel. Fluorescent bands were excised from the gel, eluted into water for 3 hours at 37 °C, and ethanol precipitated.

For gel shift assays, equimolar amounts (5 nM) of radiolabeled antiswitches and target RNA were incubated in varying concentrations of theophylline at room temperature for 30 min in 15 μL buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$). Following the incubation, 10% glycerol was added to the RNA-target-ligand mixtures, and RNA complexes were separated from free RNA by electrophoresis at 125 V on an 8% polyacrylamide gel in 1X Tris-borate buffer at room temperature for several hours. Gels were dried, and antiswitch mobility was imaged on a FX phosphorimager (BioRAD).

For nuclease mapping, fluorescein-labeled antiswitch RNA was resuspended in buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$), denatured at 65 °C for 3 min, and allowed to slow cool to room temperature. Antiswitch RNA was incubated with varying concentrations of theophylline at room temperature for 15 min. RNase T1 (Ambion) was added to the antiswitch-ligand mixture and incubated at room temperature for 15 min. Cleavage products were visualized using laser-induced fluorescence capillary electrophoresis on a P/ACE MDQ machine (Beckman) using a single-stranded nucleic acid analysis kit (Beckman) according to manufacturer's instructions.

RNA free energy calculations. RNA free energy was calculated with RNAstructure version 3.71.[32]

Appendix B.

Ecological Strategy Methods and Materials

PCR, transformations, and DNA extraction. Integration and deletion constructs were PCR amplified using KOD polymerase as per manufacturer's instructions (Novagen). Standard lithium acetate transformations for homologous recombination were performed as previously described.[28] Integration of the *GAL* promoter was performed by amplifying the *GAL1-10* promoter sequence from pRS314-Gal.[29] This fragment was PCR assembled with the kanamycin resistance gene from pFA6a-ZZ-TEV-S-kanMX6[30] along with flanking homologous regions to the *DAL80* upstream region (506000 - 504030 and 506500 - 506530 on chromosome XI). The construct was transformed, and colonies were selected on 400 ng/mL G418 YPD-agar plates. Integration was confirmed by colony PCR with primers flanking and internal to the integrated construct. Yeast DNA extraction was performed as previously described using the "bust n' grab" method.[31] Primer sequences are available upon request. Similar techniques were employed to integrate the *GAL1-10* promoter upstream of the *GDH1* coding sequence. Gene deletions were performed by amplifying the kanamycin resistance gene from pFA6a-ZZ-TEV-S-kanMX6[30] along with flanking homologous regions for the entire coding region of *GDH1, GDH3*, or *GLT1*. The D150H and C313S catalytic rate mutants were constructed by amplifying genomic DNA from the *GDH1* coding region with primers carrying the appropriate nucleotide substitution, assembled with the *LEU2* gene from pRS315,[29] and transformed as above. The *GLT1* overexpression plasmid was constructed by cloning the *GLT1* coding region in front of an *ADH1* promoter on a 2μm plasmid.

Quantitative RT-PCR. Cells were pelleted and frozen in liquid nitrogen. Pellets were resuspended in a 50 mM NaOAc (pH 5.2), 10 mM EDTA buffer. Cells were lysed by the

addition of SDS to a final concentration of 1.6% and an equal volume of acid phenol. Solutions were kept at 65 °C with intermittent vortexing for 10 min. After cooling on ice, the aqueous phase was extracted, and further extraction was carried out with an equal volume of chloroform. RNA was further isolated and concentrated by use of RNeasy columns (Qiagen) according to manufacturer's instructions. Total RNA was quantified by OD_{260} readings. RNA samples were treated with DNase (Invitrogen) according to manufacturer's instructions. cDNA was synthesized using gene-specific primers and Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. qRT-PCR was carried out on this cDNA using an iCycler iQ system (BioRAD). Samples were prepared using the iQ SYBR green supermix and primer pairs specific for different templates. Data were analyzed using the iCycler iQ software.

Competitions and fitness assays. All competitor strains were derivatives of the S288C background, while the reference strain was derived from the W303 background. These strains showed different electronic volume versus side-scatter distributions that was also used to quantitate population numbers, in good agreement with the values obtained from fluorescent measurements. Equal amounts of competitor and reference strain were mixed and grown in indicated liquid media for 3 generations (approximately 6 hours). The frequency of competitor and reference strain were quantitated before and after the growth period by counting the numbers of GFP expressing cells to non-GFP expressing cells by flow cytometry using a Quanta SC flow cytometer (Beckman Coulter) equipped with the MPL system. Samples were excited with a 488 nm laser, and GFP fluorescence was detected with a 525 nm bandpass filter. A gate was set above the non-GFP expressing

cells in the Quanta analysis software to partition fluorescent from non-fluorescent cells. Samples of only reference or competitor strains and serial dilutions of ratios of competitor to reference strains were run in parallel as quantitation controls. 5,000 events were collected per sample.

Flow cytometry and calculation of noise. Two gates were used to standardize each cell population for analysis using "magnetic gating" in FlowJo flow cytometry analysis software (Tree Star, Inc.). The first gate isolated cells displaying regular morphology based on electronic volume and side-scatter, while the second gate removed non-fluorescent cells from the distribution. This gating method was compared against other methods previously described and the abundance and noise trends observed were consistent between methods.[32,33] Noise was calculated as the square of the coefficient of variation (σ^2/p^2) of the distribution.[26] 50,000 events were analyzed to calculate noise for each sample. Noise trends were similar when calculated as the coefficient of variation (σ/p) and the variance (σ^2).

Mutagenic PCR and construction of promoter libraries. To construct mutant libraries of the *GDH1* promoter, primers flanking 500 nucleotides upstream of the *GDH1* coding region (1043500 - 1043050, chromosome XV) were used to amplify the fragment from yeast genomic DNA using KOD polymerase. The fragment was then diluted into mutagenic PCR buffer[34] (7 mM $MgCl_2$, 0.5 mM $MnCl_2$, 50mM KCl, 10mM Tris pH 8.3 with 1mM dGTP, 0.2 mM dCTP, 0.2 mM dTTP, 0.2 mM dATP) and further amplified using Taq polymerase (Roche). Separately, the kanamycin resistance gene was amplified

from pFA6a-ZZ-TEV-S-kanMX6[30] using KOD polymerase. The kan^r gene fragment and the promoter library were then ethanol precipitated, resuspended in water, and PCR assembled together by virtue of overlapping primer sequences. The resulting large fragment was then transformed into yeast strains using a standard lithium acetate procedure. Transformants were selected in liquid YPD media supplemented with 400 ng/mL G418. The resulting library was grown to stationary phase and frozen in 15% glycerol at –80 °C.

Gillespie simulations. The reactions shown in Supplementary Fig. S6 were simulated using the Gillespie algorithm as described previously.[27] The probability of a reaction occurring in a given amount of time was proportional to the reaction rate. Briefly, birth and death rates of mRNAs are denoted by k_R and δ_R, respectively; enzyme is made with translation rate k_E and degraded with decay rate δ_E; substrate is created with rate k_S and product was allowed to accumulate. Noise was introduced by varying the "burst size" of translation events by modulating the ratio of k_E and δ_R while adjusting k_R to yield similar enzyme abundance between simulations.

Experimental evolution. A 50 μL aliquot from the *GDH1* promoter library freezer stock was inoculated into liquid YPD and allowed to acclimate overnight. 2 μL of the acclimated population was diluted into 2 mL synthetic complete media with two ammonia concentrations at 139 mM ammonia and 556 mM ammonia. Populations were grown in batch culture and diluted 10^3-fold into respective fresh media every 24 hours. Aliquots from the competitions were diluted 4,000-fold, and 50 μL of this dilution was

plated on YPD-agar plates. Single colonies were inoculated in synthetic complete media and grown overnight for further analysis.

Density-dependent fitness assays. Strains were grown overnight in YPD media. Serial 10-fold dilutions of each strain were performed and cell density of the overnight culture was measured using flow cytometry as described above. Competitor and reference strains were mixed in equal ratios in synthetic complete media at the density specified in Figure 4e. Cells were grown for 24 hours and fitness was assayed as described above. 5,000 events were collected for each sample.

Nitrogen utilization efficiency assay. Cells were back diluted and grown to mid-log phase (OD_{600} ~ 0.5), spun down, dried and weighed.[35] The amount of ammonia in the media before and after growth was quantitated via enzymatic assay (Megazyme).

Methods References

28. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol*. 350, 87-96 (2002).
29. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* 122, 19-27 (1989).
30. Longtine, M. S. et al. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* 14, 953-61 (1998).
31. Harju, S., Fedosyuk, H. & Peterson, K. R. Rapid isolation of yeast genomic DNA: Bust n' Grab. *BMC Biotechnol*. 4, 8 (2004).

32. Newman, J. R. et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* 441, 840-6 (2006).
33. Bar-Even, A. et al. Noise in protein expression scales with natural protein abundance. *Nat. Genet.* 38, 636-43 (2006).
34. Cadwell, R. C. & Joyce, G. F. Mutagenic PCR. *PCR Methods Appl.* 3, S136-40 (1994).