Cell Factory Stability and Genetic Circuits for Improved Strain Development

Rugbjerg, Peter

Publication date: 2016

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Cell Factory Stability and Genetic Circuits for Improved Strain Development

Peter Rugbjerg
PhD thesis
March 2016

The Novo Nordisk Foundation Center for Biosustainability, DTU
Summary

Development of new chemical-producing microbial cell factories is an iterative trial-and-error process, and to screen candidate cells at high throughput, genetic biosensor systems are appealing. Each biosensor has distinct biological parameters, making modular tuning networks attractive. However, all synthetic gene systems - including the target metabolic pathways themselves - represent a possible fitness burden to the cell and thus constitute a threat to strain stability.

In this thesis, several studies served to develop genetic systems for optimizing cell factory development and understanding the common error modes leading to loss of stable metabolic productivity during long-term microbial fermentation.

A molecular buffer system in *Saccharomyces cerevisiae* was designed and engineered to tune the signals of a known tetracycline-responsive RNA switch (riboswitch). Generalizable and based on split transcription factors, the system e.g. allowed shift of ligand sensitivity and inversion of the output signal from OFF to ON - without changing the riboswitch or output gene. The system was i.a. characterized by green fluorescent protein (GFP), for which a recombination-stabilized multimeric GFP was developed. Overcoming cellular autofluorescence, this multimer enabled detection of weak promoter signals in *S. cerevisiae*. The concept of split transcription factors was further applied in *S. cerevisiae* as a tool to enable selection for three DNA fragments under a single selectable phenotype. This enabled quick introduction of a three-step polyketide pathway and may also serve to challenge the current paradigm of “one selectable trait – one selection gene”, as was demonstrated in plasmid and chromosomal gene introduction.

Despite of genetic selection, the load of all synthetic systems can challenge the stability of strain designs. A metabolite-producing *Escherichia coli* strain was long-term cultured to study production stability and the dynamic effects of mutations within the cell population. A genetic error landscape of pathway disruptions was identified including particular, recurring error modes. Driven by a gain in fitness, these errors within 70 generations led to a transformation of the strain to a population of genetic non-producer cells. Knowledge about these mechanisms and the applied simple mathematical model may likely serve to realize more stable microbial cell factories in future.
Sammenfatning


I denne afhandling tjente flere studier til at udvikle genetiske systemer til at optimere cellefabrikudvikling og forstå de almindelige fejltilstande, som fører til tab af stabil metabolisk produktivitet under langvarige mikrobielle fermenteringer.


Table of Contents

Summary .............................................................................................................................................. i

Sammenfatning (summary in Danish) ............................................................................................... ii

Abbreviations .................................................................................................................................... v

Preface ................................................................................................................................................ vi

Outline of PhD thesis ......................................................................................................................... viii

1 Synthetic biology for signal detection in microbial cell factory design ........................................ 1
   1.1 The inputs: Sensing the environment and intracellular state ....................................................... 3
   1.2 Developing new sensors ................................................................................................................ 7
   1.3 The outputs: to match the circuit and application ......................................................................... 8
   1.4 Balancing a synthetic circuit: The saturation of sensor with ligand ............................................. 11

2 The genetic error mechanisms and productivity loads driving instability of microbial cell factories ............................................................................................................................. 19
   2.1 Genetic error mechanisms in microbial cell factories ................................................................. 20
      2.1.1 DNA replication errors ........................................................................................................... 21
      2.1.2 Homologous recombination errors ...................................................................................... 22
      2.1.3 Slipping strand errors .......................................................................................................... 24
      2.1.4 Mobile element errors ........................................................................................................... 24
      2.1.5 Plasmid loss errors .............................................................................................................. 26
      2.1.6 Environment- and pathway-induced genetic errors ............................................................ 27
   2.2 Causes of productivity loads in microbial cell factories ............................................................. 29
      2.2.1 Target product cytotoxicity .................................................................................................... 31
      2.2.2 Intermediate or by-product cytotoxicity .............................................................................. 32
      2.2.3 Drain on endogenous metabolite pools .................................................................................. 33
      2.2.4 The protein load .................................................................................................................. 34
      2.2.5 The plasmid load ................................................................................................................ 37
   2.3 Recent synthetic anti-escape strategies ....................................................................................... 38

3 Molecular buffers permit robust transduction, sensitivity tuning and inversion of riboswitch signals .................................................................................................................................................. 47
   3.3 Results ........................................................................................................................................... 51
3.3.1 A synthetic molecular signal buffer to effectively relay the predicted riboswitch signal. 52
3.3.2 Constructing the synthetic signal buffer network................................................................. 54
3.3.3 Molecular inversion of sensor signals .................................................................................... 56
3.3.4 Using the synthetic buffer network for modular tuning of signal sensitivity .................... 56
3.4 Discussion ............................................................................................................................. 59
Supporting information ............................................................................................................... 68

4 Flexible metabolic pathway construction using modular and divisible selection gene regulators ................................................................................................................................. 96
4.3 Results ...................................................................................................................................... 99
4.3.1 Design of divisible selection................................................................................................. 99
4.3.2 Detailed design composition............................................................................................... 100
4.3.3 Selection for three different plasmids activating a single selection phenotype .............. 100
4.3.4 Modular exchange of selection phenotype from Ura+ to His+ ........................................... 100
4.3.5 Divisible selection with copy number-fluctuating plasmids (2-micron) ......................... 100
4.3.6 Three simultaneous chromosomal integrations selected under one selection phenotype ........................................................................................................................ 101
4.3.7 Proof of principle construction of fungal polyketide pathway ........................................ 102
4.4 Discussion ................................................................................................................................ 102
Supplementary material ................................................................................................................. 106

5 Recombination-stable multimeric green fluorescent protein for detection of weak promoter outputs in Saccharomyces cerevisiae ................................................................. 120
Results and discussion .................................................................................................................... 124
Supporting information .................................................................................................................... 129

Conclusions .................................................................................................................................... 175
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetyl-CoA</td>
<td>Acetyl-Coenzyme A</td>
</tr>
<tr>
<td>AD</td>
<td>Activation domain</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>CAT</td>
<td>Chloramphenicol acetyltransferase</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>DR</td>
<td>Direct repeat</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
</tr>
<tr>
<td>FOA</td>
<td>Fluoro orotic acid</td>
</tr>
<tr>
<td>FP</td>
<td>Fluorescent protein</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GOI</td>
<td>Gene of interest</td>
</tr>
<tr>
<td>HMG-CoA</td>
<td>3-hydroxy-3-methyl-glutaryl-Coenzyme A</td>
</tr>
<tr>
<td>HMGS</td>
<td>HMG-CoA synthase</td>
</tr>
<tr>
<td>IS</td>
<td>Insertion sequence</td>
</tr>
<tr>
<td>MMR</td>
<td>Methyl-directed mismatch repair</td>
</tr>
<tr>
<td>NADH</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NADPH</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>ppGpp</td>
<td>Guanosine 3′,5′-bispyrophosphate</td>
</tr>
<tr>
<td>RBS</td>
<td>Ribosome binding site</td>
</tr>
<tr>
<td>SC</td>
<td>Synthetic complete</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>tHMGR</td>
<td>Truncated HMG-CoA reductase</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>YFP</td>
<td>Yellow fluorescent protein</td>
</tr>
</tbody>
</table>
Preface

The following PhD thesis was conducted from 15th December 2012 to 14th March 2016, and builds on work performed by me in collaboration with colleagues. Initial parts of the study presented in chapter 3 were carried out during my master thesis (as specified in the joint author statement).

The PhD was funded through an internal stipendium from the Technical University of Denmark (DTU). General financial support for the studies was further given through the Bacterial Cell Factories section at the Novo Nordisk Foundation Center for Biosustainability, DTU, funded by the Novo Nordisk Foundation.

The supervisor for the PhD study was professor Morten Otto Alexander Sommer and co-supervisor was professor Jochen Förster.

Their affiliation was Novo Nordisk Foundation Center for Biosustainability DTU.

Peter Rugbjerg, 14th March 2016

Novo Nordisk Foundation Center for Biosustainability,
DTU
Acknowledgements and thanks

I feel exceptionally thankful to a long list of colleagues and friends without whom this thesis would not have become as it is. I am also immensely thankful for the entire opportunity to write this thesis and learning so much in the process, both in all the experiments that never worked or made it to this thesis, and through important tough prioritization.

I have in particular been grateful for working closely together with a number of people who deserve special highlight. Firstly, thanks to Hans Jasper Genee who supervised me as MSc thesis student and developed me and our early ideas on systems for riboswitch signal tuning.

I was very lucky to work with and supervise Kristian Jensen in the laboratory as a BSc thesis student and in subsequent individual course work. Kristian is most of everything a modeling talent and very rewarding to work and brainstorm with.

I have been very fortunate to also supervise Kira Sarup-Lytzen and Nils Myling-Petersen as BSc/MSc students - in close collaborations that both lasted for 2-2.5 years. Kira and Nils’s enthusiasm, high ambition, ideas and input made it a pleasure to coach them and at the same time move our joint projects forward.

Along these lines, I am also very grateful for so many good moments and clever inspiration from the other members of the Sommer laboratory. I want to devote special thanks to Lejla Imamovic, Andreas Porse, Christian Munck and Eric van der Helm for helpful and perspectival discussions. The researchers of the Bacterial Cell Factory Section at Novo Nordisk Foundation Center for Biosustainability in particular, but also in the other interacting branches of this center.

I also want to thank Adam Feist for hosting me a couple of months (2015) at UC San Diego, though the results did not finish in time for this thesis. I further send thanks to my co-supervisor Jochen Förster for helpful advice especially at our review meetings.

Finally, a great thank you to my supervisor, Morten Otto Alexander Sommer, for talented support, supervision, eye for important problems, guidance and ability to stimulate and make place for motivation.
Overview of scientific articles published or in preparation as result of this thesis:


3. Rugbjerg, P; Genee, HJ; Jensen, K; Sommer, MOA (2016). Molecular buffers permit robust transduction, sensitivity tuning and inversion of riboswitch signals (manuscript under review). Included in PhD thesis as chapter three.

4. Rugbjerg, P; Myling-Petersen; Sommer, MOA (2016), Recurring pathway disruptions shape the genetic error landscape of long-term cultured mevalonic acid-producing *Escherichia coli*. Included in PhD thesis as chapter six.

In addition, the thesis work lead to developing the following patent applications (does not form part of the PhD thesis):

1. Improved biosensor-based detection system. 2014. Rugbjerg, P; Genee, HJ; Sommer, MOA. WO2015044456A1

2. General system for improved fermentation titers. 2015 Rugbjerg, P; Sarup-Lytzen, K; Sommer, MOA. EP15187150.6
The following oral presentations at international scientific conferences were given to disseminate the results of the PhD thesis:

1. Synthetic selection systems for metabolic engineering
   2\textsuperscript{nd} Symposium on Applied Synthetic Biology in Europe, Málaga, Spain
   November 2013

2. Divisible selection: Multiplying selection gene utility in \textit{Saccharomyces cerevisiae}
   Nordic Yeast Research Community Meeting 2014, Copenhagen, Denmark
   September 2014
1 Synthetic biology for signal detection in microbial cell factory design

Metabolic engineering builds on the cell factory concept of fermenting microorganisms, a solid industrial success story with roots in the 1920es. Cultured microorganisms overproduce simple and complex compounds such as amino acids, penicillin and citric acid, following often drastic optimization by mutagenesis and elaborate screens (Hong and Nielsen, 2012; Ikeda, 2003). Since its defining inauguration as a field in the 1990es, metabolic engineering has grown into a discipline of ever-increasing metabolite product targets, genetically optimizing cells for the three performance parameters: yield (product/substrate), productivity (product/time) and titer (product/reactor volume).

Metabolic engineering represents the directed improvement of cell metabolism towards the production goals. There, the metabolic effects of random or (semi-)rational genetic manipulations are often evaluated in a repetitive design cycle. An overall mission is to enable cost-effective biosynthesis of molecules not possible by chemical synthesis routes, or to replace chemical synthesis by more sustainable production methods (Burk and Van Dien, 2015). The development of microbes capable of converting cheap, sustainable or waste resources thus both has an economical and societal motivation.

The current costs are high when engineering a cell factory for a new metabolic product at an economically feasible performance. The process has been cited to commonly require in the range of 3-6 years and beyond 100 researcher years (Hong and Nielsen, 2012), but fortunately its progress is subject to an ever-developing set of more efficient biotechnological methods.

The design of cell factories is likely to substantially leverage recent biotechnological leaps within genetic editing, synthetic biological circuits and computational target predictions: The annotated DNA sequence space expands, alongside the rising computational toolbox for
metabolic analysis (King et al., 2015). Gene synthesis including custom-designed libraries simultaneously drop in price (Rogers and Church, 2016). Diversity-generating technologies such as multiplex automatable genome engineering (MAGE) permit high-efficiency introduction of specific genetic changes throughout libraries of cells (Bonde et al., 2014). Another biotechnological breakthrough likely to accelerate metabolic engineering is the discovery and engineering of CRISPR-Cas9 systems. CRISPR-Cas9 systems improve the ease and flexibility of performing genetic edits by highly programmable site-specific genome cleavage even in organisms previously hard to engineer. The CRISPR-Cas9 systems also provide a faster editing platform for microbial cells (Jakočiūnas et al., 2015). The current common standard for integrating numerous genes in the chromosome of a microbe relies on iterative introduction and subsequent loop out of a selection gene – previously a significant bottleneck in the development process. Transcriptional activation or repression using cleavage-inactive dCas9 also constitutes a potent technology e.g. to direct metabolic pathway flux (Gilbert et al., 2013; Zalatan et al., 2014) and thus generate new types of libraries to assay.

This wide progress in diversity generation amplifies the demand for tools to evaluate the libraries at equally high throughput, which has spurred a broad range of activity in the field of genetically encoded biosensors (Rogers and Church, 2016; Zhang and Keasling, 2011; Zhang et al., 2015). When fully operational, biosensors allow the interrogation of large enzyme/clone libraries in short time, by linking metabolite presence to a change in reporter gene activity (will be discussed in more detail, section 1.1). It is therefore becoming important to develop the limited set of existing sensors (section 1.2), assay and interlink such metabolite signals of microbial cells and assure they are optimally connected to assayable output genes (section 1.3).

One of the other major challenges for cell factory development to become cheaper is the robust and more predictable scale-up, which is sometimes neglected (Lee and Kim, 2015). Even though strains perform well at shake flask or 2 L lab scale, they do not necessarily stably transfer to 300,000 L production scale. It is likely that more detailed consideration of the typical genetic error modes and selective drivers will be important (will be the focus of chapter two).
1.1 The inputs: Sensing the environment and intracellular state

Cells constantly assay and adapt to their environment, a unique skill that cell factory designs might utilize better. Presence of chemicals is sensed and processed into genetic signals, leading to induction or repression of gene activities involved in the behavior of the cell. Nutrient-sensing transcription factors (TFs) and RNA switches (riboswitches) are both examples of prokaryotic and eukaryotic sensing elements helping to direct the cell metabolism towards the available resources. The same genetically-encoded metabolite sensors can be transferred or de novo developed to improve the metabolic engineering cycle by regulation of an output gene that by example selects for ligand-producing biocatalysis (van Sint Fiet et al., 2006; Yang et al., 2013a).

TFs can bind to response elements in the output gene promoter in a ligand-dependent manner, whereas riboswitches control output gene expression through ligand-dependent RNA folding (Fig. 1.1AB). Due to the long-term proven abilities for de novo riboswitch development and other advantages, these will be reviewed in slight more detail (section 1.1.1)

Intracellular states resulting from a particular pathway product or intermediate (e.g. stress) can also be sensed via endogenous promoters responsive via endogenous signaling systems without need for characterizing the responsible regulation system (Dahl et al., 2013) (Fig. 1.1C).

These three categories of sensors can be manipulated to function with different output genes, whereas a fourth biosensor category functions by a direct output signal upon ligand-binding e.g. through fluorescence resonance energy transfer or combined domain proteins (Fig. 1.1D) (Michener et al., 2012). Unable to change output, this category will not be of further focus here.

Figure 1.1 Types of genetically encoded biosensors relevant in metabolic engineering. A) A metabolite-responsive transcription factor (TF) controls an output gene promoter, B) a metabolite-responsive riboswitch controls translation of output gene mRNA, C) a cell state or metabolism-dependent endogenous promoter controls an output gene, D) Binding of a metabolite to a protein results in directly detectable output.

p. 3. Chapter 1
1.1.1 Regulatory mechanisms of RNA switches (riboswitches)

RNA switches are a diverse group of natural and synthetic RNA elements that mediate a change in gene expression as response to binding of a small molecule ligand. Most often RNA switches therefore operate at the level of translational control, as part of the 5’- or 3’-untranslated region (UTR) of the output gene messenger RNA. However, riboswitches can also be functional at transcriptional levels, some involving catalytically active ribozyme activity, and some controlling pre-messenger RNA splicing (Table 1.1).

Core to a riboswitch is its aptamer domain, which binds the ligand. Aptamers can be developed synthetically or be found in natural gene regulation systems. The aptamer requires delicate integration with the output gene, in what becomes the “expression platform” (Berens and Suess, 2015). The flexibility of RNAs permits the specific binding to many small molecules. Upon this RNA folding, a new secondary structure results with a regulatory consequence: Aptamer binding e.g. reveals or sequesters a ribosomal binding site (RBS) of a regulated, downstream output gene sequence (Nudler, 2004).

In bacterial synthetic riboswitches, such control of RBS availability is a common strategy (Berens et al., 2015). Perhaps due to the different translational initiation dynamics of eukaryotes, synthetic riboswitches in this domain of life utilize a different strategy in which a translational “roadblock” is formed in the 5’-UTR as a ligand binds the riboswitch (Wittmann and Suess, 2012).

Riboswitches have also been developed as splicing controllers, giving rise to two ligand-dependent splice forms, while another class of riboswitch mechanisms controls self-cleaving ribozyme hammerheads. It is sometimes possible to transfer aptamer regions from one expression platform to another with the theophylline aptamer being a good example of this modularity type (Sharma et al., 2008; Suess et al., 2004). Different aptamers have also been inserted in the same expression platform, which gave promise to RNA as a modular design entity (Berens et al., 2015).
Table 1.1 Exemplary types of riboswitch control mechanisms, and origin of aptamer with mechanism implemented.

<table>
<thead>
<tr>
<th>Control level</th>
<th>Ligand</th>
<th>Aptamer origin</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation initiation</td>
<td>Tetracycline</td>
<td>Synthetic</td>
<td>(Hanson et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>Theophylline</td>
<td>Synthetic</td>
<td>(Sharma et al., 2008; Suess et al., 2004)</td>
</tr>
<tr>
<td></td>
<td>Neomycin</td>
<td>Synthetic</td>
<td>(Weigand et al., 2008)</td>
</tr>
<tr>
<td></td>
<td>Thiamine pyrophosphate</td>
<td>Natural</td>
<td>(Muranaka et al., 2009)</td>
</tr>
<tr>
<td></td>
<td>Adenosyl-cobalamin</td>
<td>Natural</td>
<td>(Nou and Kadner, 2000)</td>
</tr>
<tr>
<td>Pre-mRNA splicing</td>
<td>Tetracycline</td>
<td>Synthetic</td>
<td>(Weigand and Suess, 2007)</td>
</tr>
<tr>
<td>mRNA self-cleavage hammerhead</td>
<td>Tetracycline</td>
<td>Synthetic</td>
<td>(Beilstein et al., 2014)</td>
</tr>
<tr>
<td>Transcriptional termination</td>
<td>Theophylline</td>
<td>Synthetic</td>
<td>(Wachsmuth et al., 2013)</td>
</tr>
<tr>
<td>Transcriptional activation</td>
<td>Tetramethylrosamine</td>
<td>Synthetic</td>
<td>(Buskirk et al., 2004)</td>
</tr>
</tbody>
</table>

Impressively, RNA switches were engineered from synthetic aptamers before their existence in nature was discovered in 2002.

An important characteristic of riboswitches is their cis type regulation\(^1\) of the output gene, and its implications in synthetic networks. Cis control eases target gene regulation, since the riboswitch can be expressed within the same construct (Berens et al., 2015). Another implication is that the number of sensor molecules (riboswitches) per cell must follow the expression strength of the output gene. This aspect is not shared by the trans-acting transcription factors and means that the riboswitch must be able to function equally predictable at low and high copy numbers, to suit the requirements of different output gene types.

\(^1\) A perhaps rare exception is a trans-acting translation repression observed with the S-adenosyl methionine riboswitches of *Listeria monocytogenes* (Serganov and Nudler, 2013).
1.1.2 Important sensor parameters

Biological sensors are not electronic switches, but their characteristics still benefit from a consistent terminology, even though it sometimes varies in the literature.

The sensitivity range of a sensor is a measure of the affinity for ligand and thus describes the ligand concentration range in which the sensor dynamically saturates with ligand. A classical dissociation constant ($K_d$) for the affinity can be determined from the equilibrium between the bound and unbound sensor.

The sensitivity range is often measured in vivo where the range may be more application-relevant but also subject to many other variables. A certain part of the sensitivity range theoretically has a near linear relationship between ligand concentration and sensor saturation (details in SI of chapter three). Thus some studies use the term linear detection range.

The terms sensitivity and response curve steepness both describe the change in output per change in input (Ang et al., 2013; Dietrich et al., 2010). Accordingly, highly sensitive sensors are described as ultrasensitive (more details in section 1.4.1).

The output range describes the signal change in output gene activity conferred from no binding to maximum binding of the sensor with ligand. The change following addition of ligand can be positive (ON sensor) or negative (OFF sensor), and it can be reported as a fold change. The output range is not confined by the sensitivity range, but is rather an intricate result of the interconnection between sensor and output gene. The riboswitch output range may be improved through optimization of the riboswitch, e.g. by multimerization of the tetracycline riboswitch controlling eukaryotic translation (Kötter et al., 2009). Similarly, TF output ranges may be extended by multimerization of the response element (operator) copy number in the output gene promoter (Ang et al., 2013). Modular sensors in principle show the same characteristics irrespective of the output gene. Sometimes however, the specific output gene may interact, e.g. through folding with the upstream riboswitch.

Studies frequently use the term ‘dynamic range’ strictly for the sensor output range.
1.2 Developing new sensors

Sensors already functional in nature can be implemented for synthetic use. Candidate natural RNA switches can be identified in the 5’ or 3’ untranslated regions of open reading frames within gene clusters or operons that are expected to be associated with a metabolite to be detected e.g. seen for the B_1, B_2 and B_12 vitamin biosynthetic operons (Serganov and Nudler, 2013). Analogously, a candidate metabolite-responsive TF may be identified in clusters of relevant biosynthesis or metabolic export (Lange et al., 2012).

Central to the paradigm of biosensing in metabolic engineering is the availability of biosensors to suit the demands of metabolic engineers to detect other molecules than those measured in natural organisms. Development of new synthetic sensors is therefore an important, yet nontrivial process.

Engineering of affinities for new ligands into existing (hybrid) TFs is one option. Mutagenesis has resulted in changed ligand affinity for a handful of TFs, in a process that can involve computational protein design, saturation mutagenesis of the ligand binding pocket and use of fluorescence outputs in screening for resulting new ligand sensitivities (Taylor et al., 2016). A different recent TF approach fuses a ligand-binding domain with e.g. a transcriptional activation domain, and employs destabilizing mutations to achieve gene activation strictly in presence of the ligand (Feng et al., 2015).

Engineering of new riboswitches begins with the aptamer. Systematic evolution of ligands by exponential enrichment (SELEX) is an iterative procedure in which immobilized ligand on a column is used to selectively enrich a library of short RNA or DNA aptamer candidate sequences (Stoltenburg et al., 2007). Following e.g. 15 rounds of enrichment, sequences with high affinity for the ligand may be recovered, a process that has e.g. yielded aptamers for theophylline, tetracycline and neomycin. These can then undergo a process to become gene-regulating riboswitches, e.g. through screening for ligand-dependent ‘roadblocks’ (Suess et al., 2003; Weigand et al., 2008) and RBS sequestering (Suess et al., 2004; Topp and Gallivan, 2010). Such synthetic riboswitch development importantly underscores that it is possible to de novo-engineer new input sensors for synthetic biology.
1.3 The outputs: To match the circuit and application

A second aspect of successful biosensing setups is the output gene. In a metabolic engineering case, the output gene can be divided in three types: Fluorescence, selection and dynamic pathway control (examples in Table 1.2). These each provide different advantages. Modular biosensors can theoretically be transferred to the most application-suitable output.

Table 1.2 Selected sensors applied for detection of intracellular biosynthesis in metabolic engineering context and their various sensor outputs.

<table>
<thead>
<tr>
<th>Sensed compound</th>
<th>Output</th>
<th>Sensor type</th>
<th>Organism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theophylline</td>
<td>GFP</td>
<td>Riboswitch</td>
<td><em>S. cerevisiae</em></td>
<td>(Michener and Smolke, 2012)</td>
</tr>
<tr>
<td>NADPH</td>
<td>YFP</td>
<td>TF (SoxR)</td>
<td><em>E. coli</em></td>
<td>(Siedler et al., 2014)</td>
</tr>
<tr>
<td>L-valine</td>
<td>YFP</td>
<td>TF (Lrp)</td>
<td><em>C. glutamicum</em></td>
<td>(Mahr et al., 2015)</td>
</tr>
<tr>
<td>L-lysine</td>
<td>YFP</td>
<td>TF (Lrp)</td>
<td><em>C. glutamicum</em></td>
<td>(Mustafi et al., 2014)</td>
</tr>
<tr>
<td>Fatty acid ethyl ester</td>
<td>Dynamic</td>
<td>TF (FadR)</td>
<td><em>E. coli</em></td>
<td>(Zhang et al., 2012)</td>
</tr>
<tr>
<td>Acetyl-P</td>
<td>Dynamic</td>
<td>TF (GlnG)</td>
<td><em>E. coli</em></td>
<td>(Farmer and Liao, 2000)</td>
</tr>
<tr>
<td>Mevalonic acid</td>
<td>LacZ</td>
<td>TF (AraC*)</td>
<td><em>E. coli</em></td>
<td>(Tang and Cirino, 2011)</td>
</tr>
<tr>
<td>Glucosamine-6-P</td>
<td>FCY1</td>
<td>Riboswitch</td>
<td><em>S. cerevisae</em></td>
<td>(Lee and Oh, 2015)</td>
</tr>
<tr>
<td>Benzoic acid</td>
<td>TetA</td>
<td>TF (NahR)</td>
<td><em>E. coli</em></td>
<td>(van Sint Fiet et al., 2006)</td>
</tr>
<tr>
<td>Lysine</td>
<td>TetA</td>
<td>Riboswitch</td>
<td><em>E. coli</em></td>
<td>(Yang et al., 2013b)</td>
</tr>
<tr>
<td>Naringenin</td>
<td>TolC</td>
<td>TF (TtgR)</td>
<td><em>E. coli</em></td>
<td>(Raman et al., 2014)</td>
</tr>
</tbody>
</table>

1.3.1 Fluorescence outputs

Being independent of substrates or cofactors, fluorescent proteins have an advantage as output, and improvements in brightness, folding and maturation of fluorescent proteins are reported with yearly intervals (Miyawaki, 2011). Numerous biosensor setups have taken advantage of fluorescent proteins and sorting with fluorescence-activated cell sorting (FACS) machines, which can reach quite high daily throughputs of $10^8$ cells (Rogers and Church, 2016). Expression noise can limit the throughput by creating false-positive (and potentially
false-negative) signals. Such noise may be reduced with use of chromosomal integration of the output gene and/or standardization to a constitutively expressed fluorescent protein with different emission spectrum (Liang et al., 2012; Michener and Smolke, 2012). Overall, ON sensors appear to be preferred with fluorescent proteins, perhaps for the reason that such gain of function is more suitable to positively distinguish a ligand-producing cell.

Green fluorescent protein (GFP) is commonly used as a linear reporter of gene expression, similar to enzymatic reporter genes. Correlation between protein expression and fluorescence has been established, e.g. in *Escherichia coli* (Albano et al., 1998). This may be important when using fluorescence as modular measure to evaluate biosensors and e.g. to avoid signal attenuation due to other output-limiting factors of the cell.

**Limitations of fluorescent proteins** as true reporters of biosensor performance may be reflected when operating with a low expression level; microbial cells naturally auto-fluoresces, which gives rise to a minimum detection level under which any fluorescence outputs are masked (Billinton and Knight, 2001). Subtraction of background fluorescence is a solution often employed, but will not reveal fully concealed signals. The technique may also produce large (artificial) effects when the signal is close to the background level, a situation that inescapably also increases the standard error of the background-subtracted results. Even with bright, modern fluorescent proteins, a single protein/cell is too weak to shine brighter than the auto-fluorescence of the cell. Random movements (diffusion) in the cytosol may be the reason for this, as tethering yellow fluorescent protein (YFP) to the *E. coli* membrane produces enough fluorescence for detection of a single molecule (Yu et al., 2006).

While auto-fluorescence of the growth medium can be limited by washing with non-fluorescent buffer or single cell flow cytometry, cell auto-fluorescence is less easily avoided. Cell auto-fluorescence might however be minimized by operating at wavelengths where the cell auto-fluorescence is lower. Chapter five focuses on a multimerization strategy for amplifying weak fluorescence signals.
1.3.2 Selection outputs

Selection output genes enable ligand-dependent growth, a useful phenotype in enrichment cultures or for colony formation assays (van Sint Fiet et al., 2006; Yang et al., 2013a). Characterized by simple lab setups, colony formation assays can reach high throughputs of $>10^9$ per day (Dietrich et al., 2010) if engineered well. The common principle utilizes a medium condition (toxin), e.g. antibiotic, to disfavor growth in absence of the active selection (antidote) gene. Positive selection genes such as antibiotic resistance genes and prototrophic genes thus work well with ON sensors to generate ligand-dependent growth, whereas OFF sensors must be combined with negative selection genes, which are less frequent. Dual positive and negative selection genes exist, such as *E. coli* tetA and tolC and *Saccharomyces cerevisiae* URA3 and TRP1, which enable selection in opposite direction upon change of medium conditions.

The dose-response of selection genes to the toxin is important if the biosensor sensitivity range is to be utilized in a graded fashion and cells distinguished by specific ligand concentration. Some resistance (antidote) genes such as *cat* encoding chloramphenicol acetyltransferase (CAT) are dose-responsive, i.e. increased antidote expression (CAT) can counterpart an increased toxin (chloramphenicol) concentration (Maxwell et al., 1999). Prototrophic selection genes lack this ability, since they operate without a toxin.

Using the tetracycline (antiporter) resistance gene *tetA*, (van Sint Fiet et al., 2006) scored *tetA* inducer concentration by colony size, indicating another useful type of dose-response.

1.3.3 Dynamic pathway control

Biosensors can also be applied to take direct control of the metabolic pathway, as a method to improve the flux of the pathway by recognizing that the cellular ability to accomodate the pathway may be dynamic. A dynamically activated pathway can e.g. minimize toxic intermediate formation (Dahl et al., 2013). In an early study, Farmer and Liao optimized the lycopene pathway in *E. coli* by controlling pathway expression using a promoter responsive to the acetyl-phosphate pool, which appeared to yield a more soft, late-phase onset of pathway expression, both improving growth and productivity (Farmer and Liao, 2000). Intermediates of the lycopene pathway such as isopentenyl pyrophosphate was indeed later found to be cytotoxic (Martin et al., 2003).
1.4 Balancing a synthetic circuit: The saturation of sensor with ligand

Use of synthetic circuits can fulfill the function of interlinking sensors with more complex cellular actuators. In design of synthetic circuits it is very important to consider and balance the signal transfer throughout the network to avoid signal loss. Signals typically propagate with changes in the saturation of a receptor (output) by its ligand (input). This can concern many types of basic receptor-ligands such as riboswitch-ligand, DNA-transcription factor, or transcription factor-inducer. Circuits may not functionally transfer signal if changes in input concentration does not change the degree of receptor saturation.

The saturation of receptor with ligand is dependent on the concentration of both elements and their mutual affinity (chapter three, supporting information). The binding of ligand to the receptor follows an equilibrium reaction between their dissociated and associated states. In simple cases without cooperativity this relationship will undertake Michaelis-Menten dynamics, meaning that the fraction of receptors bound by ligand will vary following a standard graded saturation curve. This relationship also sets the theoretical boundaries for the sensitivity range of a sensor: In a simple binding equilibrium, it theoretically requires an 81-fold increase in ligand concentration to saturate the sensor from 10 % to 90 % (without cooperative binding). Nevertheless, the reported sensitivity range of sensors is usually not as wide, perhaps due to difficulties in measuring intracellular ligand levels, limited uptake mechanisms, or ligand binding to other intracellular targets. In certain applications, more steep response curves may be beneficial, e.g. to yield a more robust response.

To achieve good input/output relations in a circuit, the sensitivity and output ranges of the parts must be matched, for example by ensuring the right initial expression level and degradation level. DNA-binding repressors can e.g. have a very high affinity for their DNA-binding sites (e.g. $K_d = 1 \text{ nM}$ for LexA-binding sites (Zhang et al., 2010)). This affinity amounts to a sensitivity range of repressor DNA-binding within very few (around 1) repressor molecules per *E. coli* cell. As an example, if the repressor is expressed at high copy number of e.g. average 1000 copies per cell and regulated with an OFF sensor at an output range of 8-fold, this regulation still results in an average 125 repressors per cell. Such concentration is likely still above the sensitivity range of the binding between repressor and DNA (at $K_d = 1 \text{ nM}$), which remains saturating the DNA, and thus the sensor signal does not propagate.
through the circuit. Lower repressor expression level or binding affinity could theoretically improve the transfer of signal.

Another risk when operating at low copy number is the possible elevation in expression noise. Noise is the fundamental variation in copy number of cellular components due to the stochastic nature of biological systems (Eldar and Elowitz, 2010) and can be quantified as the coefficient of variance for a population of single cells. Noise generally increases at low copy numbers (Silva-Rocha and de Lorenzo, 2010), and may thus additionally challenge the predictability of synthetic biological systems.
1.4.1 Ultrasensitivity-generating systems

Ultrasensitivity in an input-output relation describes the situation where a signal is amplified and a small input change results in a larger output change. Such relationships are known in buffered situations such as pH-controlled human blood, where changes in acid concentration are buffered by a corresponding base, thus maintaining pH homeostasis. In TF synthetic biology, the graded response of a TF controlling an output gene has been converted into an ultrasensitive (digital-like) response using molecular sequestration/titration with an high-affinity inhibitor of the TF that prevents signal relay (Buchler and Cross, 2009). The resulting ultrasensitive response curve (exemplified Fig. 1.2) is less prone to small fluctuations in input (TF) concentration, due to buffering by the inhibitor as long as the inhibitor is in excess to the TF. By variation of the inhibitor concentration, Buchler and Cross were also able to shift the response curve horizontally, effectively changing the sensitivity range of the system (Buchler and Cross, 2009).

![Ultrasensitivity diagram](image)

**Figure 1.2** Ultrasensitivity in TF regulation of an output promoter can be introduced using a high-affinity TF inhibitor. A) Direct binding of TF to output DNA results in a graded output, whereas B) introduction of a high-affinity inhibitor titrates away TF from the system, giving rise to C) an ultrasensitive response curve due to the inhibitor functioning as ‘signal’ sink by molecular titration (drawn example of response curves).
Acknowledgements

Hans J. Genee is thanked for comments to the chapter.
References


p. 17. Chapter 1


2 The genetic error mechanisms and productivity loads driving instability of microbial cell factories

The current design of metabolically engineered microbes usually focuses on high yield, productivity and titer, evaluated from a lab cell culture after a limited number of cell divisions. In some cases following e.g. 25 generations from single cell to a saturated 250 mL culture, and in some cases after more elaborate 2 L fed-batch experiments. Metabolically engineered cell factories represent a very application-close branch of synthetic biology. In fact many synthetic biology designs may also be intended at a much more long-term scale than lab tests, e.g. for devices meant to one day have a medical functionality in the gut of a patient (if ethically and safety-wise acceptable), be released to a (potentially) contained environment for sensing or cleanup of pollution, applied in a 300,000 L fermentation tank for manufacture of a valuable biochemical etc. Such long-term operation, usually involves a significantly increased number of cell divisions, all at which there is a risk for genetic damage to the system.

Various factors can lead to genetic errors. The mechanisms and their rates of successfully taking a cell factory system out of operation (loss of function) will be described in section 2.1, ranging from around $10^{-2}$ to $10^{-10}$ generation$^{-1}$ (Table 2.1). Today, a major instability factor, plasmid loss, is readily addressed (section 2.1.5), yet many other mechanisms provide genetic escape in synthetic systems, though most errors also have possible biotechnological solutions.

The rate of mutations leading to system escape would not have been as critical if synthetic systems did not confer a fitness defect (growth inhibition) to the host. However, productive cells are easily growth-inhibited, resulting in strong selective forces against strain stability, which will be reviewed in section 2.2.

Combined, the genetic errors and their selection can lead to significant loss of system functionality, as has been be characterized experimentally in more detail in chapter six.
2.1 Genetic error mechanisms in microbial cell factories

A microorganism is at constant risk of mutation, and as such its genome is the impressive result of the occurrence of new gene variants and - following classical Darwinian evolution - their possible selection to suit the habitat. The sequence space that such a cell can test out on its trajectory towards a fit state is enormous, and thus the ways to reach a particular fitness optimum are manifold and several mutational mechanisms can unfold depending on the environment and stimuli to the cell (Table 2.1). Each mechanism presents a risk to the stability of humanly designed synthetic gene systems, such as a metabolic pathway. In this section, genetic errors will be reviewed in special relation to cultured *E. coli* populations carrying recombinant metabolic pathways or similar synthetic systems, with additional highlights given from *S. cerevisiae*.

**Table 2.1** Overview of genetic errors that may occur in *E. coli* K-12 and their common reported rates.

<table>
<thead>
<tr>
<th>Type</th>
<th>Cause</th>
<th>Rate</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(generation⁻¹)</td>
<td></td>
</tr>
<tr>
<td>Transposition</td>
<td>Mobile element (e.g. of IS)</td>
<td>10⁻⁶⁻⁻¹⁻⁸ cycA⁻¹</td>
<td>(Durfee et al., 2008)</td>
</tr>
<tr>
<td>SNPs</td>
<td>DNA polymerase III</td>
<td>10⁻¹⁰ bp⁻¹</td>
<td>(Tippin et al., 2004)</td>
</tr>
<tr>
<td></td>
<td>- Without MMR (<em>mutS</em>)</td>
<td>+ 10¹ x</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Without editing-repair (<em>mutD</em>)</td>
<td>+ 10⁻¹⁻¹⁻³ x</td>
<td></td>
</tr>
<tr>
<td>Plasmid loss</td>
<td>Unbalanced segregation</td>
<td></td>
<td>(Summers, 1991)</td>
</tr>
<tr>
<td></td>
<td>- 5 copies per cell</td>
<td>6 10⁻²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 15 copies per cell</td>
<td>6 10⁻⁵</td>
<td></td>
</tr>
<tr>
<td>Tandem duplication/deletion</td>
<td>DNA polymerase slipping</td>
<td>-</td>
<td>(Darmon and Leach, 2014)</td>
</tr>
<tr>
<td>Large deletions</td>
<td>Recombination (<em>recA</em>)</td>
<td>10⁻⁵, &gt;25bp repeat⁻¹</td>
<td>(Lovett et al., 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10⁻⁵, &gt;200bp repeat⁻¹</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recombination (non-<em>recA</em>)</td>
<td>10⁻⁸, &gt;25bp repeat⁻¹</td>
<td>(Lovett et al., 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10⁻⁷, &gt;200bp repeat⁻¹</td>
<td></td>
</tr>
</tbody>
</table>
2.1.1 DNA replication errors

The replication of the *E. coli* genome is under normal laboratory growth driven by DNA polymerase III as part of the replication fork. To aid the fidelity of the polymerization reaction, a number of processes serve to detect and correct possible mutations, making replication a precise activity. The otherwise, uncorrected incorporation of a wrong nucleotide, a single nucleotide polymorphism (SNP), can have dramatic phenotypic consequences: In diverse, protein-coding sequences, a random amino acid change leads to functional inactivation in 20-65% of the instances (Drummond et al., 2005).

First in line, 3’ exonuclease activity immediately attempts to correct wrong incorporation of a nucleotide (mispairing). This act of proof reading (editing repair) provides the DNA polymerase with another attempt upon mispairing. In *E. coli*, the activity is encoded by *mutD*, and improves the polymerization fidelity by a 100-fold (Snyder and Champness, 2007; Tippin et al., 2004).

In addition, *E. coli* possesses another replication error correction system, monitoring the mismatches of the newly synthesized nucleic acid strand. This methyl-directed mismatch repair (MMR) system detects incongruence between two strands with guidance from the ubiquitous CTAG (dam) methylation pattern. The CTAG pattern is methylated a while after DNA replication, and in cases of uncorrected mismatches, the lack of dam methylation state identifies the new, and thus erroneous, strand from its correct template. MMR involves the activity of *mutS*, *mutL* and *mutH* and reduces the mutation rate by a 1000-fold (Snyder and Champness, 2007). Together with *mutD*, these mechanisms ensure a wildtype *E. coli* SNP rate of around $10^{-10}$ bp\(^{-1}\) generation\(^{-1}\), similar to the level in *S. cerevisiae* (Drake et al., 1998).

During the course of bacterial evolution experiments, ‘hypermutable’ subpopulations, characterized by significantly higher mutation rates, are commonly observed (LaCroix et al., 2015). The underlying mechanistic basis can be a disruption of MMR e.g. loss of *mutS* (Fehér et al., 2012). Whereas mutation-prone cells normally would appear as a disadvantage, the ability to spin out mutators within an evolving population may help adapting to sudden environmental changes, while posing a negligible cost to the cell in constant environments under which it is preferable to maintain the genotype. In laboratory evolution experiments, *E. coli* is commonly exposed to new specific environments with the objective of studying its adaptation. Such new environments provide a space of new genetic trajectories towards improved fitness, and thus cells with an elevated mutation rate may be favored. Similarly, such mutator strains are also likely to occur under long-term cultivations cell factory strains.
2.1.2 Homologous recombination errors

*E. coli* and many other microorganisms possess other error correction and diversity-generating systems that permit homologous or identical nucleotide stretches of down to around 25 bp to be efficiently recombined (Lovett et al., 2002). Recombination can generate different structural variation depending on the context of the two recombining DNA regions, and while helpful in natural DNA error correction, the process becomes a common threat to synthetic system stability.

Recombination between two direct repeats in the same DNA can result in “loop-out” and thereby deletion of the DNA in between, or tandem duplication (Fig. 2.1). Tandem duplications will be obvious recombination targets themselves, which thus may restore the original sequence if not providing a selective benefit (e.g. via an amplified gene copy number)(Darmon and Leach, 2014). If the repeats instead are inverted relative to each other, the result of the recombination event is an inversion, which does not lead loss of genetic material, and in fact might invert back. If the repeats are positioned completely exterior to the inverted region, phenotypic effects may even be undetectable.

![Figure 2.1 Recombination between direct repeats (DR) can lead to tandem duplications or deletions.](image)

In multiploid organisms, homologous recombination is key in ensuring meiotic generation of genetic diversity between sister chromosomes, while its function in bacteria may be limited to synthesis error correction. Recombination can assist a replication fork bypass a DNA area in which a lesion or damaged nucleotide prevents ordinary read-through (Bichara et al., 2006). Thus, recombination easily challenges the genetic stability of synthetic designs due to the limited diversity and frequent repetitive use of biological sequences, such as promoters, terminators and genes encoding fluorescent proteins and rate-limiting pathway enzymes.

Homologous recombination is a process involving many proteins. In *E. coli* RecA is a central...
protein, i.a. involved in the sensing of single-stranded (damaged) DNA. The *S. cerevisiae* homolog of RecA is RAD51. *E. coli* recombination can initiate at exposed single-stranded DNA due to damage. Double-stranded DNA breaks are loaded to RecA by the RecBCD complex, whereas the RecFOR proteins assist RecA at single-strand breaks. A subsequent Holiday junction where all four strands meet and exchange finally resolves and the new DNA arrangement is stabilized through ligation (Lovett et al., 2002).

**Solutions to homologous recombination errors:**

Due to its role in recombination, *recA* has been deleted in several *E. coli* K-12 “cloning strains” (e.g. *E. coli* DH10B, XL1 and DHS-alpha). Other *rec* genes have also been targeted such as *recB* (e.g. *E. coli* SOLR). While *recA* deficiency lowers the rate of recombination, *recA*-independent recombination can still take place at considerable rates (Table 2.1), and should ideally be considered in the genetic design.

Diversification of the nucleotide code using synonymous codons may be another method to reduce recombination when expressing protein-coding repeats (Chapter five, Rugbjerg et al., 2015), as well as e.g. use of different promoters on the same DNA entity (Naesby et al., 2009). Employment of chromosomal integration sites separated by essential genes may also aid stability towards loop-out of identical integrated genes, such as devised in *S. cerevisiae* (Mikkelsen et al., 2012), though such systems still may not prevent chromosomal rearrangements.
2.1.3 Slipping strand errors

Slipping strand mutations may happen during replication if DNA polymerase temporarily stalls, separating the nascent strand from the template strand. Upon realignment at sites containing short repeat sequences, misalignments may happen resulting in the deletion or introduction of repeats in the newly synthesized strands (Moxon et al., 2006; Renda et al., 2014). Presence of secondary structures such as hairpins may further promote the stalling of the DNA polymerase (Bichara et al., 2006). Such strand slippage events may generate frameshifts and thereby significantly mutate the strand. Bacteria in some instances exploit these short simple repeats as a type of switch, in which the affected gene may interchange between different biological forms, and resulting in the term contingency locus (Moxon et al., 2006).

Natural “contingency loci” can consist of many repetitive repeats such as (AGTC)_{38} of \textit{Haemophilus influenzae} (Bichara et al., 2006). Synthetically designed DNA can easily contain short repeat sequences and palindromic sequences provide secondary folding: DNA-binding sites for transcription factors are e.g. commonly multimerized to amplify the signaling strengths, serving as substrate for slipping strand mutations (author’s observation). Another source of tandem repeats may be codon-optimized sequence from simple algorithms consequently applying the highest-frequency codon. Even triplet repeats are prone to slipping strand mutagenesis (Bichara et al., 2006).

Diversification of repeat sequences at functionally insensitive loci might limit slipping strand errors, as might the limitation of secondary structures, though no known reports have addressed the issue.

2.1.4 Mobile element errors

Mobile elements are genetic regions, which autonomously move or copy themselves within the genome in a process known as transposition. Transposition into a non-essential gene is highly suited for completely taking such “dispensable” genes out of action, making transposable elements threatening to the stability of many synthetic systems. The simplest bacterial forms are known as \textit{insertion sequences} (ISs) and frequently span 0.7-2.6 kbp in size, encoding only the transposase necessary for its transposition (Darmon and Leach, 2014). Larger transposable elements are formed of two exterior IS elements while a middle region carry auxiliary genes e.g. encoding antibiotic resistance genes, suitable for horizontal gene transfer. The genome of the \textit{E. coli} K-12 strain MG1655 contains 19 different IS
elements, at different copy numbers (43 copies in total). The derived daughter strain DH10B however hosts 63 copies, which appeared to cause a higher mutation rate in DH10B (Durfee et al., 2008). ISs show some degree of target site selectivity (Craig, 1998). The insertion of an IS involves the duplication of a short target region (usually 3-9 bp) to both exterior sides of the IS (Fig. 2.2). ISs can restore the locus when they transpose to a new, however interventions from endogenous DNA repair systems in the process can paradoxically leave sequence scars (Darmon and Leach, 2014).

Unlike SNPs, transposition of an IS into a gene is highly likely to cause complete disruption of that gene functionality. ISs can contain outwards pointing promoters and thus also influence the expression of neighboring, but intact genes (Fehér et al., 2012). Furthermore, multiple IS copies may serve as a substrate for structural variation via recombination. Conversely the duplicated target site regions well below 25 bp provide the IS protection from such direct repeat recombinatory loop out.

The existence of numerous IS types and copies in bacteria has been a cause for speculation to their role in otherwise very compact genomes. ISs can both be seen as beautiful examples of selfish DNA, solely fulfilling the role of self-maintenance parasiting on the host microorganisms, or they can indeed be observed symbiotically as a specific mutagenic instrument possessing the ability of rapidly knocking out genes that burden the host (Schneider and Lenski, 2004; Tenaillon et al., 2004).

![Diagram](https://via.placeholder.com/150)

**Figure 2.2 Insertion of an IS, resulting in the duplication of short target sites at the exterior of both IS junctions**

IS elements are also active in the stationary growth phase, and transposition of around $10^6$ per cell per hour in *E. coli* agar stabs have even been detected (Schneider and Lenski, 2004). The *S. cerevisiae* genome hosts a number of transposable Ty elements. With a transposition
rate of $10^{-7}$ to $10^{-5}$ generation, their importance in evolution and synthetic system stability is possibly similar to ISs in *E. coli* (Kazazian, 2004). The target site selectivity of the Ty elements appear to be more functionally linked than ISs, and Ty3 element is particularly interesting from the perspective of synthetic systems; Ty3 appears to target few nucleotides upstream of open reading frames, sharing DNA-binding affinity with RNA polymerase III transcriptional initiation factors (Chalker and Sandmeyer, 1992; Craig, 1998).

**Solutions to mobile element errors:**

The *E. coli* reduced-genome project constructed a set of *E. coli* strains devoid of various non-essential genes, including all IS elements (*E. coli mds42*) (Pósfai et al., 2006). The application of an IS-free bacterium became part of granted patent (Blattner et al., 2012), and longer stability of IS-eliminated *E. coli* has been demonstrated for stable metabolite (L-threonine) and recombinant protein production (Lee et al., 2009; Park et al., 2014). In a similar approach, removal of ISs in *Corynebacterium glutamicum* has improved pathway stability in this host (Choi et al., 2015).

### 2.1.5 Plasmid loss errors

Plasmids have frequently served as a simple rapid vehicle for introduction or overexpression of genes. The (nonselective) mechanisms preventing plasmid loss are manifold. In some instances it involves active partitioning to equally segregate the plasmids at division, while high copy numbers automatically reduce the risk of random mis-segregation in which no plasmids are transferred to the daughter cell (Summers, 1991). As described by Summers, the probability per division of such loss depends on the copy number $n$ and can be calculated by simple binomial probability:

$$p = 2 \cdot 0.5^n$$

The rate of *S. cerevisiae* ARS/CEN plasmid is an example of a synthetic plasmid stabilized by inclusion of centromeric sites to utilize the centromere segregation mechanism, and its loss rate is around $10^2$ per generation (Hieter et al., 1985).
Solutions to plasmid loss:
Use of selection genes (typically prototrophy or antibiotic resistance genes) is the main solution to plasmid loss, despite the special requirements to the medium. Some natural plasmids employ post-segregational growth arrest mechanisms composed of genetically encoded polycistronic toxins and cognate neutralizing antitoxins (TA systems). Among the targets of these toxins are DNA replication, ATP synthesis and messenger RNA translation effectively leading to growth arrest if not neutralized by antitoxin. Due to a higher instability of the antitoxin, the toxin will prevail in situations where expression of their operon ceases, effectively punishing cells that just lost the plasmid (Darmon and Leach, 2014). The stabilizing effect of TA systems has also been utilized biotechnologically, e.g. by encoding the hok/sok TA system from an unstable plasmid (Wu and Wood, 1994). The use of plasmids may be alleviated through chromosomal integration, e.g. in several copies to reach equal gene dosage.

2.1.6 Environment- and pathway-induced genetic errors
Physical and chemical factors such as ultraviolet (UV) light and mutagenic compounds cause damages in the DNA, and stimulate a specific set of cellular rescue reactions known as the SOS response. Chemical- and UV-generated DNA lesions can trigger error-prone rescue mechanisms, possibly shared with those of stress factors from high-level metabolic production: While UV light and mutagens are usually absent during industrial fermentations, the SOS pathway can still be activated in recombinant protein production, though the exact trigger is unknown (Hoffmann and Rinas, 2004). However, at least presence of misfolded (recombinant) proteins can stimulate the SOS response (Fahnert et al., 2004). The SOS pathway induces a set of more than 40 genes via breakdown of the LexA repressor, a process aided by DNA-bound RecA (signaling damaged DNA) (Foster, 2005; Tippin et al., 2004). The SOS response is also responsible for activating nucleotide excision repair, which is involved in removing damaged bulky DNA (Janion, 2008) that normally stalls replicative DNA polymerase. Therefore, the SOS response also induces a set of E. coli error-prone DNA polymerases II, IV and V, which can ignore the translesions (e.g. inserting a random nucleotide or frameshift) (Tippin et al., 2004). The SOS response thus leads to an elevated mutation rate, a situation the cell likely only accepts when the integrity of the DNA is at bay. The error of these polymerases - at a rate down to $10^{-1}$ to $10^{-3}$ bp$^{-1}$ for undamaged DNA - are significantly more frequent than the $10^{5}$-$10^{7}$ bp$^{-1}$ of polymerase III with editing repair.
In reaction to physical and chemical stresses such as nutrient-starvation, oxidative species, abnormal pH, temperature and osmolarity, *E. coli* activates the “general stress” response of *rpoS* (Battesti et al., 2011), which elevates the mutation rate, possibly through same mechanisms as the SOS response (Lombardo et al., 2004). The RpoS sigma factor influences the expression level of around 10 % of all genes, and causes a state of cross-protection against its stimulatory stressors.

In an industrial fermentation, osmotic, pH and starvation stress may readily result from the metabolic engineering of the cell along with the state of high cell density, thus readily stimulating *rpoS*. Oxidative damage can result from unfolded proteins or metabolism (Polizzi and Kontoravdi, 2015), and may directly lead to DNA strand breaks, while low pH forms a basis for DNA damage by removal of purine bases and even double-stranded lesions (Choi et al., 2000).

RpoS induces error-prone DNA polymerase IV, while also repressing *mutS* of the MMR system (Lombardo et al., 2004), thus also forming the basis for an elevated mutation rate. A possible general stimulation of IS transposition by the SOS or *rpoS* pathways has not been shown, yet *E. coli* IS10 transposition has been linked to UV-generated stress, in a manner dependent on the *lexA* and *recA* pathways (Eichenbaum and Livneh, 1998). Overall evidence across biological kingdoms appear to suggest that transposition in general is induced during stressful situations (Levin and Moran, 2011), a condition where genetic diversity is more evolutionarily favorable.

**Solutions to stress-induced genetic errors:**

In one example of putrescine overproduction in *E. coli*, Qian and colleagues observed growth inhibition and induction of *rpoS*, which led the authors to delete *rpoS* from the genome, causing a better yield on glucose and 7 % higher titer, albeit without further mechanistic insights (Qian et al., 2009). *rpoS* might not be causative of the stress but also seen as providing a beneficial stress tolerance response, and general applicability of this strategy remains to be seen.

Working with the IS-deleted strain *E. coli* mds42, Csorgo and colleagues further deleted the error-prone DNA polymerases II, IV and V, which resulted in a close to 50 % reduction in the spontaneous mutation rate observed, indicating also that the activities of these stress-related polymerases contribute substantially when cells were grown to “early stationary phase” (Csorgo et al., 2012).
2.2 Causes of productivity loads in microbial cell factories

High-performance metabolic production in cell factory strains is presumably a cellular state far from the natural fitness optimum of the host strain. Yet the limited ability of current tools to match the cell resources may challenge how well the designs suit the host physiology. Stress factors, metabolic burdens and biochemical toxicities therefore reduce the growth rates of productive cells.

If not addressed in the strain design process, these factors might inhibit the cells unnecessarily. Most fitness defects specifically affect faithfully producing cells, while sparing competing non-producer sub-populations in the fermentation broth. These effects, selective against the strain design, will collectively be termed productivity loads. Productivity loads may rise from various sources such as the toxicities of intermediate, by and end products besides metabolic loads (metabolite drains, and enzyme and nucleotide synthesis) (Fig. 2.3). Different productivity loads examples are listed (Table 2.2), while in some cases growth inhibitions were observed, but not readily explained (Table 2.3).

By analogy, other synthetic systems than metabolic pathways can confer loads too, but the specifics of these factors will not be addressed here.

Figure 2.3 Metabolic pathway sources to productivity load - a situation where producer cells are selectively growth-inhibited when competing against mutated, non-producing progeny.
In this section, focus will be devoted to the loads from introducing heterologous genes or overexpressing native genes (Table 2.2), as these loads are more mutation prone (i.e. loss-of-function mutations). Different types of strain engineering, such as deletion of chromosomal genes may easily retard the growth rate as well e.g. in limiting fluxes. However, compared to productivity loads of introduced genes, such loads are not easily reverted, unless unsuitably engineered e.g. through simple point mutation.

### Table 2.2 Examples of productivity loads and stressors affecting growth of microbial strains.

<table>
<thead>
<tr>
<th>Type of productivity load</th>
<th>Metabolite and/or enzyme affecting fitness</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate or byproduct toxicity</td>
<td>Mevalonic acid: <strong>HMG-CoA</strong></td>
<td>(Pitera et al., 2007)</td>
</tr>
<tr>
<td></td>
<td>1,3-propanediol: <strong>sn-Glycerol-3-phosphate</strong></td>
<td>(Zhu et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>1,3-propanediol: <strong>Methyglyoxal</strong></td>
<td>(Zhu et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>1,3-propanediol: <strong>3-hydroxypropionaldehyde</strong></td>
<td>(Barbirato et al., 1996)</td>
</tr>
<tr>
<td></td>
<td>Methanol carbon source: <strong>Formaldehyde</strong></td>
<td>(Yurimoto et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>Vanilin glucoside: <strong>Vanilin</strong></td>
<td>(Hansen, 2009)</td>
</tr>
<tr>
<td></td>
<td>Taxadiene: <strong>Indole</strong></td>
<td>(Ajikumar et al., 2010)</td>
</tr>
<tr>
<td></td>
<td>Amorphadiene: <strong>Farnesyl pyrophosphate</strong></td>
<td>(Martin et al., 2003)</td>
</tr>
<tr>
<td>End-product toxicity</td>
<td><strong>3-hydroxypropionic acid</strong></td>
<td>(Kildegaard et al., 2014)</td>
</tr>
<tr>
<td></td>
<td>Limonene</td>
<td>(Dunlop, 2011)</td>
</tr>
<tr>
<td></td>
<td>Butanol</td>
<td>(Dunlop, 2011)</td>
</tr>
<tr>
<td></td>
<td>Farnesyl hexanoate</td>
<td>(Dunlop, 2011)</td>
</tr>
<tr>
<td></td>
<td>Isobutanol</td>
<td>(Atsumi et al., 2010)</td>
</tr>
<tr>
<td></td>
<td>Styrene</td>
<td>(McKenna and Nielsen, 2011)</td>
</tr>
<tr>
<td>Enzyme insolubility (misfolding)</td>
<td><strong>Orotidine-5′-phosphate decarboxylase</strong></td>
<td>(Geiler-Samerotte et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>α-glycosidase</td>
<td>(Jürgen et al., 2000)</td>
</tr>
<tr>
<td>Drain of endogenous substrate or co-factor pool</td>
<td><strong>P450 overexpression: Heme pool</strong></td>
<td>(Michener et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>Spider silk protein: <strong>tRNA^Glu_ pool</strong></td>
<td>(Xia et al., 2010)</td>
</tr>
<tr>
<td></td>
<td>L-lysine: <strong>TCA intermediate pools</strong></td>
<td>(Koffas et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>Lycopene: <strong>Dimethylallyl diphosphate pool</strong></td>
<td>(Yoon et al., 2006)</td>
</tr>
<tr>
<td>Protein synthesis burden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasmid disturbance of metabolism</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

p. 30. Chapter 2: Genetic errors and productivity loads
Table 2.3 Cell factory strains with growth inhibition and no immediate explanation suggested.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Valine</td>
<td>E. coli</td>
</tr>
<tr>
<td>L-serine</td>
<td>C. glutamicum</td>
</tr>
<tr>
<td>L-leucine</td>
<td>C. glutamicum</td>
</tr>
<tr>
<td>Putrescine</td>
<td>E. coli</td>
</tr>
<tr>
<td>L-Tryptophan</td>
<td>B. subtilis</td>
</tr>
</tbody>
</table>

2.2.1 Target product cytotoxicity

The target titers for economically feasible metabolic cell factories commonly reach double-digit gram per liter values, easily equivalent of the 100 mM range. Such concentrations of otherwise in natura-tolerated molecules can put a significant specific growth-inhibition on the producing organisms, e.g. through solvent, osmotic or pH stress (Nicolau et al., 2010; Varela et al., 2004; Warnecke and Gill, 2005). Cells may also endogenously convert the product to a more toxic molecule, as suspected for 3-hydroxypropionic acid produced in S. cerevisiae (Kildegaard et al., 2014). Cells can cope with osmotic and possibly carboxylic acid stress by production of osmolite molecules such as trehalose (Nicolau et al., 2010). But still, efficient extracellular export mechanisms are of importance not only as end-product metabolic sinks, but also to relieve faithful producers from elevated cytotoxicity (a productivity load since mutant non-producers would be less affected). Thus, how severe the toxicity is to strain stability depends on how well the product is exported. End product cytotoxicity is quite commonly characterized in metabolic engineering studies, e.g. through extracellular addition of the product to the host cell organism. Due to the assumption of free membrane transport (which is not always correct), such tolerance values might in fact be over-estimated i.e. producing cells might tolerate less than determined by extracellular addition. Nevertheless, a common method determines IC$_{50}$ values based on end cell density (Beekwilder et al., 2007). Some metabolite toxicities may also stimulate genetic response systems e.g. as introduced in section 2.1.6, amplifying their impact on strain stability.

2.2.1.1 Substrate cytotoxicity

Substrate cytotoxicity represents a special case, which theoretically may be selectively beneficial to faithfully producing cells (they posses an intracellular metabolic drain).
Bioremediation of toxic substrates could constitute an example of such advantageous, negative productivity loads.

### 2.2.2 Intermediate or by-product cytotoxicity

The toxicity of pathway by and intermediate products is a very problematic threat to strain stability. It potentially causes significant productivity loads due to its intracellular nature that specifically impacts the producing cells.

Remarkable impacts from cytotoxic intermediates in metabolic pathways have been reported. In an *E. coli* pathway towards 1,3-propanediol, problematic toxicity was observed from the intermediate and by products sn-glycerol-3-phosphate, methylglyoxal and 3-hydroxypropionaldehyde (Barbirato et al., 1996; Zhu et al., 2002, 2001). In the fission yeast *Schizosaccharomyces pombe* the cytotoxicity of vanillin prompted metabolic engineers to aim for further conversion into the glycosylated vanillin-β-D-glucoside form (Hansen et al., 2009). In this solution, the authors may also have devised a clever nature-inspired design strategy for handling such toxic metabolic products.

While the fluxes through pathway intermediate metabolite pools are high ideally, the actual pool size of intermediate metabolites does not need to be. In a well-balanced metabolic pathway, the toxicity of intermediate metabolites is less of a concern due to immediate consumption in the following catalytic step. One important strategy is therefore the balancing of enzyme concentrations to match the substrate pool. Similarly, toxic by-product formation should ideally be eliminated, or alleviated through conversion to non-toxic substances.

In the engineering of the 1,3-propanediol pathway, the toxic elevated concentration of methylglyoxal associated with glycerol consumption was addressed through the introduction of a glyoxalase enzyme, which directly elevated the resulting 1,3-propanediol titer by a 50 % improvement likely due to a similar relative increase in the biomass formed (Zhu et al., 2001). The improved end cell density further indicated a significant, concomitant reduction of the productive load.

In a study of a mevalonic acid pathway in *E. coli*, Pitera and colleagues discovered and addressed the toxicity of the HMG-CoA intermediate. In a strategy of stepwise introduction of the mevalonic acid pathway enzymes, significant toxicity was observed with the introduction of the HMG-CoA synthase (HMGS) (Pitera et al., 2007). Precursor depletion (drain of the acetyl-CoA pool) was ruled out as cause, since supplementation with the direct precursor did not alleviate the inhibition. Finally, the authors discovered that the toxicity
from HMGS expression was relieved by expression of the reductase consuming HMG-CoA (HMGR), supporting that intracellular HMG-CoA was toxic. Further functional genomics studies suggested HMG-CoA toxicity to arise from inhibition of fatty acid biosynthesis and facilitated a counter strategy by fatty acid medium supplementation (Kizer et al., 2008). Knowledge of the molecular mechanism behind the toxicity is not necessarily required. Build-up of the toxic intermediate farnesyl pyrophosphate (FPP) in the amorphadiene pathway in *E. coli* was addressed dynamically by FPP-responsive promoters in an impressive demonstration (Dahl et al., 2013). The authors analyzed the transcriptional response to FPP accumulation and this way isolated endogenous FPP stress-responsive promoters. Without knowing the basis for FPP toxicity, the use of these promoters to control the downstream pathway dynamically, improved growth and amorphadiene production. Underscoring the improvement by dynamic pathway control, seven static promoters of the Anderson collection were compared, and despite their strengths spanning three orders of magnitude, they all resulted in much lower amorphadiene production (Dahl et al., 2013). Compared to the number of metabolic engineering studies that report toxicities of the end product, the impacts of pathway intermediate metabolites appear considerably more rare in the literature, but may be a factor when the target product concentration does not explain an observed growth inhibition.

2.2.3 Drain on endogenous metabolite pools

Competition with endogenous metabolite fluxes is a metabolic burden that may be observed in unbalanced metabolic pathways. Many pathways consume substrates of the central carbon metabolism, energy and cofactors such as acetyl-CoA, ATP, GTP, NADPH and NADH. Their unbalanced depletion may cause a growth-retarding drain. Too heavy drain on the heme pool was discovered in a global transcriptional analysis of *S. cerevisiae* expressing a cytochrome P450 to produce theophylline (Michener et al., 2012), whereas formaldehyde by-production as another potential stressor was ruled out based on non-differential expression of an associated stress gene.

Heterologous production of glycine-rich (45%) spider silk proteins in *E. coli* conferred a significant growth-inhibition and limitation in production due to depletion of the glycyl-tRNA pool, as supported by a proteomic analysis (Xia et al., 2010). Co-expression of glycyl-tRNA from another plasmid both improved growth 30-50 % and the production of spider silk protein.
2.2.4 The protein load

A protein burden causes productivity loads associated with the transcription, translation and folding of the enzymes active in the metabolic pathway, whereas an increased degree of macromolecular crowding also may play a role.

2.2.4.1 Protein synthesis – resource drain and at limited capacity

Industrial metabolic pathways require very different enzyme expression levels. Well-optimized metabolic cell factories may employ a generous overexpression with multiple gene copies and high transcriptional strength and high-affinity ribosome binding sites (using e.g. the S. cerevisiae TEF1 promoter or E. coli recA promoter). However, the GTP- and ATP-requiring tRNA charging and polymerization of amino acids poses a metabolic burden to the cell (Glick, 1995), which the limited availability of less abundant amino acids can amplify (Bonomo and Gill, 2005; Ramirez and Bentley, 1993).

Further, recombinant protein synthesis also takes up capacity from protein synthesis related to growth. Examining the fitness cost of unnecessarily expressing the E. coli lac operon, (Stoebel et al., 2008) found the cost proportional to the amount of protein synthesized, rather than their downstream metabolic activities. This of course does not exclude further metabolic costs under conditions where the protein is “necessary” and/or enzymatically active. Protein synthesis requires a matching ribosomal capacity: As freshly inoculated E. coli transits from stationary phase to exponential growth, the number of ribosomes available for protein synthesis can become limiting. Studying the influence of inducing lacZ to the simultaneous synthesis of a number of other proteins, (Vind et al., 1993) found a rapid decrease in their translation rates 1-3 minutes after lacZ induction, suggesting a general limitation in the free ribosomes (in total 15,000 ribosomes/E. coli cell in glycerol-based minimal medium). An internal response to protein overproduction appears to compensate this under-capacity: Working from the observations of limited ribosomal capacity, (Shachrai et al., 2010) could describe how the growth inhibition from their model protein (GFP) was highest in the first three generations following inoculation from stationary phase. After limited ribosomal capacity for three generations, the cells appeared to reduce the cost of synthesizing the protein. The authors hypothesized the reason to be feedback-balancing of the resources between synthesis of ribosomal and non-ribosomal protein via the ppGpp (stringent) response (Shachrai et al., 2010). The stringent response is also activated by depletion of charged tRNAs, and induces heat shock proteins including proteases, possibly also resulting in increased protein turnover (Ramirez and Bentley, 1995).
However, limitations of the transcriptional machinery are also described in *S. cerevisiae* and *E. coli* when pushing promoter activities towards the maximum limits. In early studies overexpressing the strong yeast transcriptional activator GAL4, (Gill and Ptashne, 1988) observed a possible titration effect (coined “squelching”). Upon squelching, transcription from other promoters was inhibited upon activation of the GAL4 promoters. In *E. coli* similar competition for limited RNA polymerase has been indicated, e.g. when studying the battle of the σ70 and σf factors as the cells enter stationary phase (Farewell et al., 1998). A maximum protein capacity also exists. Studies following the *E. coli* foreign protein synthesis have described a limiting expression level of 30 % of the total protein content, after which growth and protein synthesis cease (Kurland and Dong, 1996). In pure protein production schemes, a stabilization strategy is to induce protein synthesis at the end of the growth phase (Glick, 1995). While late-exponential or stationary-phase promoters thus appear attractive (Holtz and Keasling, 2010), they may not be as feasible in metabolite-producing strains where production is growth-coupled (Feist et al., 2010).

### 2.2.4.2 Macromolecular crowding

Microbial cells are packed with protein at an intracellular level comparing to 300-400 g/L (Ellis, 2001). This macromolecular “crowding” and its congruent effect of “excluded volume” may further be significantly amplified in cell factory settings overexpressing certain enzymes. It is possible that these effects limit e.g. the effective diffusion of pathway metabolites in the cytosol. In *E. coli*, growth-restricting effects from the limited tRNA diffusion due to molecular crowding has been suggested (Klumpp et al., 2013). While the phenomenon and impact to cell factory settings appears under-explored, it is certainly possible that protein crowding also impacts the fitness of cell factory strains due to enzyme overexpression.

Co-localization of pathway enzymes using scaffolds in *E. coli* has been demonstrated as an advantage in the mevalonic acid pathway (Dueber et al., 2009), where the positive effects were attributed to minimized intermediate toxicity. It can be speculated whether scaffolds might also alleviate possible negative crowding effects.
2.2.4.3 Protein misfolding and inclusion body formation

A related productivity load of enzyme expression can arise from misfolded pathway enzymes and the process of the cell to quality control and correct it. Usually unable to function correctly, misfolded proteins not only represents wasted synthesis capacity and substrate, studies have also shown an increased cytotoxicity due to the misfolding products even in microbes (Drummond and Wilke, 2009; Geiler-Samerotte et al., 2011; Villaverde and Carrió, 2003). Misfolding cytotoxicity might even be a factor describing why more highly expressed proteins are less evolved across all kingdoms of life (Drummond et al., 2005).

Mistranslation is another more generic misfolding source. Protein synthesis is an error-prone process, and given mistranslation rates as high as $10^{-3}$ to $10^{-4}$ per amino acid, proteins with at least a single incorrect amino acids are not unexpected in the E. coli cell (Drummond and Wilke, 2009). The rate further increases if rare aminoacyl-tRNAs become limited (Glick, 1995). Prokaryotic cells lack much of the post-translational decoration machinery of eukaryotes, further challenging the proper folding process of these proteins. As misfolded proteins typically expose insoluble, hydrophobic residues towards the exterior (into the hydrophilic cytosol), the molecular chaperone machinery dynamically functions to detect and re-fold misfolded proteins, however at a limited capacity. Induction of the heat shock response including chaperones and proteases is thus often observed along with expression of heterologous proteins (Hoffmann and Rinas, 2004).

Misfolded proteins can aggregate into insoluble “inclusion bodies” of functionally inactive protein in eukaryotes and prokaryotes. The inclusion body is not a static element however and forms a dynamic part in the cell from which the chaperones are able to extract misfolded aggregates, refolding it to functional, soluble protein, or, alternatively proteases can enable the reuse of the amino acids (Sørensen and Mortensen, 2005; Villaverde and Carrió, 2003). Therefore, inclusion body formation can also be a burden in a strain even despite detection of flux through the catalyzed reaction of that enzyme.

Misfolded protein can cause substantial fitness costs. In a study in S. cerevisiae, Geiler-Samerotte and colleagues characterized the cost of misfolded proteins using wild-type and misfolding mutants of the uracil biosynthesis enzyme orotidine-5′-phosphate decarboxylase (URA3) in growth competition assays (Geiler-Samerotte et al., 2011). Indicating an elevated fitness cost from the response to misfolded proteins, the authors isolated a considerable 3 % fitness cost of misfolded protein (constituting only 0.1 % of the total protein steady-state concentration) compared to approx. 1.4 % cost of expressing the same amount of folded
protein.
A number of strategies are directed towards improvement of protein folding. Enzyme mutagenesis and tagging the enzyme with a fluorescent protein or chloramphenicol acetyltransferase is one strategy due to the advent of tag function if the tagged protein folds. Other factors that can aid protein folding are the reduction of growth temperature, tagging with folding enhancer tags and overexpression of chaperones and tRNAs (Sørensen and Mortensen, 2005).

2.2.5 The plasmid load
The costs of plasmid propagation not only relates to nucleotide synthesis. Biotechnological plasmids can also significantly interfere with the host metabolism (Birnbaum and Bailey, 1991). A major culprit in both E. coli and S. cerevisiae systems may be the common selection genes associated with them. In a recent S. cerevisiae-based systematic evaluation, fitness cost varied considerably (up to a striking 25 %) due to origin type, promoter, but in particular the selection gene (Karim et al., 2012). In E. coli, significant selective loads of the p15A-origin plasmid pACYC184 has e.g. been isolated using a mathematical model to account for the segregational loss rate (Lenski and Bouma, 1987). Cellular resources assigned to the replication likely also causes an elevated metabolic stress and energy drain upon use of very high-copy number plasmids in E. coli such as the artificially elevated copy numbers of the pUC family (Summers, 1991).

In S. cerevisiae, the autonomously replicating, centromeric ARS/CEN plasmids exploit the centromere CEN sequence of a native S. cerevisiae chromosome to more evenly segregate the plasmids during cytokinesis. However, toxic effects associated with use of the same CEN sequence on five different (selected for) plasmids have been observed, such as reduced growth, viability and abnormal cell formation, likely due to interference with the segregation of the endogenous chromosomes (Futcher and Carbon, 1986). The improved methods for S. cerevisiae chromosomal integrations using CRISPRs and streamlined cloning systems have made it easier to perform stable edits in a timely fashion on the chromosomes (Jakočiūnas et al., 2015; Jensen et al., 2013), altogether speaking against plasmids as pathway vectors.
2.3 Recent synthetic anti-escape strategies

The fact that synthetic systems, such as metabolic pathways or biosensor systems, easily escape requires careful attention. Limitation of the productivity load (or system load of a biosensor system) together with minimization of the escape rate will be key to stabilizing future systems.

Use of redundancy in the critical parts is one method for backing up synthetic functionality, as performed in two biosensor selection-coupled strategies (Genee and Sommer, 2013; Raman et al., 2014). There, two different (redundant) selection output genes reduce the rate of cells surviving selection pressure due to system error and not ligand activation of the biosensor. In selecting improved pathways from a diversified *E. coli* genome, Raman et al. experienced that the high rate of system mutations compromised the efficiency of the system. They therefore used the counter-selectable *tolC* to punish growth of system mutants, but a such strategy also requires the possibility to conditionally repress expression of the clone library (Raman et al., 2014).

In a biocontainment strategy, synthetic auxotrophy was engineered into *E. coli* by assigning the UAG codon in a recoded genome to a non-standard amino acid to make cells auxotrophic for it (Mandell et al., 2015). Here essential proteins were redesigned to require incorporation of the non-standard amino acid. System redundancy helped reduce the escape rate (cells growing without supplementation), as cells escaped e.g. via point mutation or mobile element transposition.

In a recent study, (Mahr et al., 2015) applied a fluorescence-coupled L-valine biosensor to guide adaptive laboratory evolution in *C. glutamicum* in rounds where clones were transferred guided by biosensor activity using FACS. This process identified genomic mutations that not only improved the L-valine titer, but also reduced an associated growth inhibition. Making it an attractive method of reducing productivity loads, alternative adaptive evolution without the biosensor would have been challenged by the likelihood that most mutations simply abolish productivity (studied more in chapter six). Indeed adaptive laboratory evolution experiments usually aims to optimize metabolic tolerance (Dragosits and Mattanovich, 2013) and thus not the pathway.
Acknowledgements

Hans J. Genee is thanked for comments to the chapter.
References


p. 40. Chapter 2: Genetic errors and productivity loads

p. 41. Chapter 2: Genetic errors and productivity loads

p. 42. Chapter 2: Genetic errors and productivity loads


3 Molecular buffers permit robust transduction, sensitivity tuning and inversion of riboswitch signals

Authors:
Peter Rugbjerg, petru@biosustain.dtu.dk
Hans Jasper Genee, hjg@biosustain.dtu.dk
Kristian Jensen, krisj@biosustain.dtu.dk
Morten Otto Alexander Sommer*, msom@bio.dtu.dk

Affiliation:
Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark.
Kogle Allé 6, DK-2970 Hørsholm, Denmark

* Corresponding author
3.1 Abstract

Predictable integration of foreign biological signals and parts remains a key challenge in the systematic engineering of synthetic cellular actuations, and general methods to modulate signal stability and sensitivity are needed. To address this problem we modeled and built a molecular signal buffer network in *Saccharomyces cerevisiae* inspired by chemical pH buffer systems. The molecular buffer system can convert a leaky riboswitch into a robust sensor enabling synthetic control of colony formation and modular signal manipulations. The riboswitch signal is relayed to a transcriptional activation domain of a split transcription factor, while interacting DNA-binding domains mediate the transduction of signal and form a stabilizing molecular buffer. The molecular buffer system enables modular signal inversion through integration with repressor modules. Further, tuning of input sensitivity was achieved through perturbation of the buffer pair ratio guided by a mathematical model. Such buffered signal tuning networks will be useful for domestication of RNA-based sensors enabling robust outputs and library-wide selections for drug discovery and metabolic engineering.
3.2 Introduction

Synthetic circuits rely on predictable and robust signal processing at the interface of biological input sensors and output actuators. However, constraints in matching signal input/output (I/O) currently limit the possible functions that can be designed (Jensen and Keasling, 2014). Accordingly, the integration of subtle inputs with general stabilizing and modulating networks is needed to successfully actuate complex biological programs. Such approaches may further help provide stability towards signal errors and context effects inherent in many human-designed biological systems, including multi-component synthetic computation networks, gene-therapeutic dosage control and metabolic biosensors constitute systems that otherwise require precise and robust signal transmission (Ausländer and Fussenegger, 2013; Benenson, 2012; Khalil and Collins, 2010; Michener and Smolke, 2012; Sprinzak and Elowitz, 2005; Wang and Katz, 2010; Wang et al., 2013; Weber and Fussenegger, 2006).

A wide variety of natural input sensor types exist, including protein transcription factors (Eggeling et al., 2015; Galvão and Lorenzo, 2006), chemotaxis systems (Baker et al., 2006) and RNA switches (Berens et al., 2015; Mandal and Breaker, 2004). Despite an apparent abundance of candidate input sensors, only a modest number of such regulators are routinely used to build most synthetic circuits, namely protein-based input sensors such as LacI, LuxR and GAL4 (Cantone et al., 2009; Kaern et al., 2003; Sprinzak and Elowitz, 2005; Stanton et al., 2014; Xie et al., 2011; Yokobayashi et al., 2002). RNA switches hold the potential advantage that their ligand-recognizing part, the aptamer, can be tailored synthetically for virtually any molecule using the SELEX technology (Stoltenburg et al., 2007). Sophisticated sensor modules have been constructed on the basis of modular assembly of synthetic (Win and Smolke, 2008) and programmable de novo RNA switches (Rodrigo et al., 2012). Yet, despite progress in the design and understanding of RNA regulators (Davidson and Ellington, 2007; Isaacs et al., 2006; Lucks et al., 2011; Soulière et al., 2013; Xie et al., 2011) the switches are still limited in a number of aspects: input sensitivity, output ON/OFF direction and predictability.

Specific adaptation of the regulatory properties of a RNA switch to fit the needs of a conceived genetic circuit can be made. For instance, change of sensitivity is possible
through rational mutagenesis of the switch (Soulière et al., 2013), and riboswitch mutants with inverted outputs have been identified through elaborate screenings, (Muranaka et al., 2009). However, these adaptations are challenging and time consuming (Berens and Suess, 2015). In contrast, the response curve of transcription factor-based input sensors can be manipulated modularly, e.g. by deploying synthetic signal drains, such as competitive inhibitors (Ang et al., 2013; Buchler and Cross, 2009), which render the signal ultrasensitive and can change the input trigger thresholds. Furthermore, modular signal inversion can be achieved based on translational fusions of the sensor to repressor/activator domains (Bellí et al., 1998; Zalatan et al., 2014).

To minimize the need for tailoring riboswitch sensors to their specific conceived actuations, a modular signal-processing system offering control interfaces without need for changing the actual input and output parts is needed. In this study we set out to stabilize and transform ineffective responses of a tetracycline-responsive riboswitch by directing its output signals through a molecular signal buffer network to enable tunable population-wide selection. Our idea is inspired by the concept of pH buffers where acid/base fluctuations are curtailed by cognate buffer molecules. Such signal processing adds beneficial modular control points for switch-independent changes of input sensitivity and output direction. We therefore adapt the buffer concept to riboswitch regulation through the use of split transcription factors expressed at uneven ratios as buffer pairs.
3.3 Results

Direct control of complex cellular actuations, such as cell survival by riboswitches can sometimes be challenging to achieve due to leaky expression or noise although the same riboswitch with luminescent or fluorescent outputs yields quantitative and repeatable readouts. In spite of these challenges synthetic control of cell survival is required in order to enable library selections for specific phenotypes or to control cell populations in a variety of applications. To assess these challenges we engineered genetic constructs comprising a tetracycline-responsive riboswitch (Kötter et al., 2009) that down-regulates translation of the classical yeast selection gene URA3 in Saccharomyces cerevisiae. URA3 mediates 5-fluoroorotic acid (FOA) sensitivity (FOA\(^5\)), which permits negative selection. Despite a substantial down-regulation capacity of the riboswitch at 37-fold (Kötter et al., 2009), addition of tetracycline input resulted only in a limited improvement of growth with about 10-fold more colonies appearing on FOA\(^R\) selection plates (Fig. 3.1).

In selections, background cells that incorrectly form colonies in the absence of the input are false positives. Equally undesirable are the false-negative cells that fail to form colonies despite receiving input. Both error types limit the possible throughput e.g. when assaying libraries. We wanted to ensure that URA3 expression was within the dynamic range of cell death and survival. Too high basal URA3 expression would mean that even the down-regulated URA3 expression is too high to cause survival due to leakiness (false negatives). Oppositely, too low basal expression would constitutively

![Genetic circuit](image_url)

**Figure 3.1** Colony formation in response to direct riboswitch-sensed input. Direct relay of the riboswitch signal to URA3 caused poor control of colony formation (PRd5). Colony formation responses were determined using spot assays with ten-fold serial dilution of equally dense cultures (Materials and methods) and representative examples of triplicate tests shown. Plates were SC –leu –trp, with 0.09 % (w/v) FOA in selective medium. 150 μM tetracycline was used as ligand.
cause survival (false positives). Since URA3 toxicity also depends on the FOA concentration, we used this to fine-tune the range. We reduced the concentration of the FOA to a level where increased false-negative cells were seen compared to wild-type expression of URA3 (0.09 % (w/v) FOA) (Supporting Info). Yet, the strain would still only form roughly 10-fold more colonies in the presence of tetracycline, i.e. many cells unintentionally failed colony formation (Fig. 3.1). This simultaneous occurrence of false-positive and false-negative cells indicated a fundamentally poor signal relay.

3.3.1 A synthetic molecular signal buffer to effectively relay the predicted riboswitch signal

Biological and chemical systems maintain pH homeostasis by providing a surplus of inter-convertible acid/base species. These react with fluctuating molecules to render them dysfunctional, rather than affecting pH, unless a specific threshold equivalence point is reached; e.g. human blood is buffered stably to pH 7.4. We hypothesized that a simple, protein-based signal buffer can be engineered in a similar fashion.

One embodiment of such synthetic signal buffer would be a genetic network employing split transcription factors (Rugbjerg et al., 2015b). Here, the input-sensing riboswitch controls the translation of a hybrid activation domain (AD) from a transcription factor such as GAL4. At the same time, a separate, cognate GAL4 DNA-binding domain (DBD) is transcribed at equal levels to the AD transcript (Fig. 3.2). Thus, the DBD will be present in high numbers relative to the low number of DNA-binding sites positioned upstream of the output gene. The DBD thereby functions both as a mediator for the signal-correlated AD when bound to DNA, and as a surplus buffer molecule when DBD is not bound to DNA. Thus, the output would only be driven by the fraction of DBD\textsubscript{DNA} bound to an AD molecule. This network architecture should buffer against a few AD proteins incorrectly translated due to leakage or instability in the riboswitch signals, since such molecules would be more likely to bind to a free DBD rather than DBD\textsubscript{DNA}. 
Figure 3.2 Design of the molecular buffer network to stabilize against riboswitch leaks. The signal-buffer design is composed of equal-level expression of the two independent GAL4 transcription factor domains, the riboswitch-regulated AD (blue) and the DBD (red), both fused to mutually interacting domains. The output gene of interest (GOI) is expressed from a minimal promoter featuring upstream binding sites for the DBD. When adding riboswitch ligand as input to the system, the translation of AD mRNA is inhibited, leading to a reduced expression of the GOI. By employing a vast surplus of inactive free DBD relative to the active DNA-bound DBD, erred fluctuations in signal molecules (AD) are buffered by binding to inactive DBD.

We constructed a simple mathematical model for the conceived buffer network to guide the design. The model was built by expressing receptor:ligand saturation fractions for each molecular layer, while assuming simple Michaelis-Menten dynamics for the different functional interactions: ligand-riboswitch, AD-DBD and DBD-DNA (Supporting Info). According to the model, the specific tuning of the buffer molecule levels significantly impacts the function, and full utilization of the riboswitch regulation potential will be achieved with roughly equal expression of DBD and AD transcripts (Fig. SI 3).
3.3.2 Constructing the synthetic signal buffer network

To set up the buffer network in *S. cerevisiae*, cognate GAL4 AD and DBD proteins were expressed with equal strength from *ADH1* promoters with the tetracycline-responsive riboswitch down-regulating translation of AD (Fig. 3.2). Repeats of the cognate DNA-binding sites were positioned upstream of a minimal promoter from *SPO13*. This promoter features a natural UME6 repressor binding site to reduce system-independent expression, such as previously engineered for a classical yeast two-hybrid strain (SPAL10) (Vidal et al., 1996). After relaying the riboswitch signal through the buffer network, robust, population-level control of the *URA3* phenotypes became possible in an *S. cerevisiae* strain deficient of wild type *GAL4* and *GAL80* (Fig. 3.3A). The signal-buffered strains acquired an ability to link colony formation to presence of the riboswitch input using classical FOA\(^9\) selection where false-positive cells were first observed when spotting $10^5$ cells (Fig. 3.3A) in the buffered strain. Importantly, both false-positive and false-negative colonies were reduced at the same time using the molecular buffer compared to the strain directly linking the signal from the riboswitch sensor to the actuator (Fig. 3.1).
Figure 3.3 Colony formation in response to riboswitch-sensed inputs using the engineered circuits. A) Signal-buffered riboswitch relay provided full ligand-controlled colony formation (PRa22). B) Relay of the riboswitch through the single-protein wild type transcription factor did not allow ligand-dependent colony formation (PRa28). C) Modular inversion of the riboswitch signal by regulation via the GAL80 repressor (PRa84).

Colony-formation responses were determined using spot assays with ten-fold serial dilution of the cultures (Materials and methods) and representative examples of triplicate tests shown. Plates were SC –leu –trp, with 0.09 % (w/v) FOA in selective medium. 150 μM tetracycline was used as input ligand.

We also tested the importance of introducing buffer molecules compared to simply passing the signal through a transcription factor layer employed at the same transcriptional strength but without buffering. Likely due to the high expression level and concomitant constitutive DNA saturation, use of a single-protein, wild type GAL4 transcription factor resulted in no control of colony formation (Fig. 3.3B), despite possessing the same DNA-binding domain (Fig. SI 4). As indicated in a riboswitch...
selection study, such direct riboswitch control of a full transcription factor likely requires use of a much weaker expression level (driven by CYC1 promoter) (Klauser et al., 2015).

3.3.3 Modular inversion of sensor signals
The need for ON or OFF switches in genetic circuits depends on the desired actuations and outputs. To allow the combination of OFF riboswitches with gain-of-function genes, we wanted to demonstrate the ease by which riboswitch signals can be treated modularly with the synthetic buffer network and a repressor module. Engineering of robust repression by adding repressor-binding sites within synthetic/hybrid promoters can be challenging. Instead we took advantage of the constructed, robust activation modules and inverted the signal modularly by relaying the OFF signal to the GAL80 anti-activating repressor, which binds and inhibits GAL4 AD (Traven et al., 2006). In this way signal inversion results from expressing the genes encoding DBD and AD with the same transcriptional strength as a gene encoding GAL80 controlled by the tetracycline riboswitch. With these strains using the exact same sensor and output modules, the opposite colony formation behavior was observed (Fig. 3.3C) compared to the first buffered strain, although the frequency of false positives increased by 10-fold, indicating that the expression level of GAL80 could be further optimized relative to DBD and AD.

3.3.4 Using the synthetic buffer network for modular tuning of signal sensitivity
Riboswitch ligand sensitivity is dependent on the particular RNA-ligand interactions, which are non-trivial to engineer. The graded saturation of riboswitch copies by the ligand may provide a window to assay for increased ligand concentrations over an isogenic population, as average reporter outputs would respond to these. However, when scaling to heterogeneous populations of a diverse library, it becomes difficult to isolate these effects to single cells as desired in high-throughput assays and plate based selections. Instead, our model suggested that the sensitivity could be stably shifted by perturbing the transcriptional ratios of AD to DBD (AD:DBD) (Supporting Info). By introducing more free AD molecules relative to DBD in the system, a higher number of input molecules (i.e. larger AD reduction) would be required to lead to the same output.
Thus, an increased ligand concentration would be required in order to produce an AD:DBD protein ratio that outputs a survival response (Fig. 3.4A).

![Image](image-url)

**Figure 3.4** Tuning the effective sensitivity for the ligand by modulating the transcriptional ratio of AD and DBD in the buffered network with fluorescence and colony formation as output.

A) Model showing the effect on output (URA3 expression) by changing the AD:DBD transcriptional ratio (AD$_0$:DBD) or adding riboswitch ligand.

B) Dynamic range tuning measured in *S. cerevisiae* with 6xGFP as output. Background-subtracted relative fluorescence units (RFU) are shown responding to added riboswitch ligand (tetracycline) for two strains AD$_1$ and AD$_2$ with perturbed AD$_0$:DBD, error bars denote std. error (n = 3). The output is shown with the fit to the model with linear x-axis and in the small window with logarithmic x-axis.

C) Shifted trigger point for colony formation of the two perturbed strains spotted in 10-fold serial-dilution spot assays of equally dense cultures on SC--leu--trp + 0.09 % FOA with indicated riboswitch ligand concentrations (tetracycline) (one representative experiment shown from triplicates).

To test this experimentally, we perturbed the ratios by introducing the weaker CUP1 promoter to drive DBD expression as a low-expression alternative to the ADH1 promoter (Barbour et al., 2000) otherwise used. The response curves of the two perturbed systems measured with 2x3 tandem green fluorescent protein (2x 3vGFP)(Rugbjerg et al., 2015a) displayed this response shift and the changed curvature was fitted to our model (Fig. 3.4B) with high confidence ($R^2 = 0.98$ and 0.96). The increase of AD:DBD transcriptional ratio resulted in a vertical response increase, probably due to a higher degree of AD-binding to DNA-bound DBD. More interestingly, it also introduced a horizontal right-shift, thus increasing the number of ligand molecules needed to produce the same absolute expression level and produce the same relative down-regulation from...
the maximum. This horizontal shift could also be relayed to colony-formation output (Fig. 3.4C). The increased AD:DBD transcription ratio effectively shifted the ligand sensitivity threshold to trigger survival at an increased ligand concentration. Whereas the strain AD1 with ADH1-based expression of AD and DBD required 50 μM tetracycline to trigger the survival, the strain AD2 with increased AD:DBD transcription ratio required 150 μM tetracycline to trigger survival (Fig. 3.4C). As predicted by the model, the low threshold concentration could be reconstituted by simply reducing the absolute transcription levels for AD as much as for DBD, hence reestablishing the 1:1 transcriptional AD:DBD ratio (Fig. SI 5). Much like pH buffers, the output did not change notably in response to absolute changes in the concentration of the interacting buffer pairs when their mutual ratio was kept the same.
3.4 Discussion

In next-generation synthetic systems, biological signal interfaces that improve parts’ interoperability are needed to meet the challenge of designing diverse biological functions using diverse biological parts. Recently, such progress has been attained using spatial insulators to limit the negative impact from the genetic context of the combined parts (Lou et al., 2012) or using modular signal transduction scaffolds with auto-inhibition (Whitaker et al., 2012). In other systems, better I/O coherence has been obtained through use of directed evolution approaches or extensive tuning libraries sampling the functional circuit space (Egbert and Klavins, 2012; Yokobayashi et al., 2002). Another powerful method for signal improvement is the use of protein sequestration to generate ultrasensitivity and to transform graded signals into binary forms (Buchler and Cross, 2009; Palani and Sarkar, 2011). Ultrasensitivity can result if the buffering molecules have higher affinity for the input signal than the output relay has (Buchler and Louis, 2008), whereas in our demonstration, the buffering agent is the same protein as the output relay. Molecular buffering may be a natural signal stabilization strategy. In fact, buffering of noise in some natural systems has been predicted as a result of the order of dimer transcription factor binding, which produces a pool of signal stabilizing, inactive monomers (Bundschuh et al., 2003). In this study, we demonstrated that design principles of molecular buffers can be reconstructed synthetically without ultrasensitivity to effectively stabilize and utilize the signal response of a riboswitch, allowing new cellular actuations. Adapted from pH-stabilizing buffers, this protein-based buffer device allowed robust gene regulation driven by a riboswitch sensing-module. Similarly, the input trigger threshold for shifting the output phenotype could be tuned by changing the ratio of the buffer proteins, analogous to how pH buffer ratios affect the stabilized pH.

Signal modulation has been described employing different pools of “unfunctional” response mediators such as anti-activators and shunt DNA-binding sites to change GFP-based outputs (Buchler and Cross, 2009; Chen and Arkin, 2012; Daniel et al., 2013). Introducing signal computation based on protein-protein interactions, this concept alleviates issues with tuning the expression level of DNA-binding proteins such as repressors at levels of a few molecules per cell where unintended, constitutive oversaturation of the binding sites will result in loss of signal. Using split transcription
factors, we instead rely on customized protein-protein interaction of hybrid proteins. Lower binding affinity between these parts allows responses to transmit at higher signal molecule concentrations, which may provide stability towards fluctuations. These provide an easy protein-protein interaction control point for inversion of the signal direction, which is often important as many output genes work only with e.g. ON signals.

These different generic interaction interfaces of the network may serve to further incorporate multi-component signal schemes comprising e.g. sub-buffer systems by engineering specific conditional DBDs and specific protein-interaction domains without module cross talk. We also anticipate that the signal-stabilizing network engineered in this study could be reconstructed in quite different synthetic embodiments by implementation of other buffer molecules cognate to an otherwise fluctuating signal. By taking advantage of the natural concept of molecular signal buffers, these systems will aid the large-scale domestication of wild type or synthetic input sensors for more predictable, customized cell re-programming.
3.5 Methods and materials

Standard methods for strain construction and molecular biology in \textit{S. cerevisiae} and \textit{Escherichia coli} were used. All plasmids and chromosomal deletion substrates cloned in this study were constructed using uracil-excision cloning (Nour-Eldin et al., 2006) by assembly of PCR fragments as described in Supporting Info. Complete cloning and strain construction methods, strain lists and plasmid lists are given in Supporting Info.

3.5.1 Colony formation response assays

4 mL synthetic complete (SC) medium (2 \% glucose, pH = 5.6) lacking leucine (leu) and tryptophan (trp) was inoculated with a single colony of the strain and split into two halves for pre-culturing with/without the riboswitch ligand (150 \( \mu \)M tetracycline) for 18 hours at 30 degrees C, 175 rpm horizontal shaking. Each culture was ten-fold serial diluted in a 96-well plate, such that each dilution contained 100 \( \mu \)L of volume. 5 \( \mu \)L of each dilution (both pre-cultures) was spotted onto SC–leu–trp plates (pH = 4.5) and the respective assay plates, supplemented with 0.09 \% (w/v) FOA and the relevant concentration of tetracycline. Preparation of FOA-containing plates is further described in Supporting Info. Plates were incubated in darkness at 30 degrees C for three days. For spot assays, equal cell concentrations between strains and conditions were controlled by evaluation of the spots on SC -leu -trp plates. Photographs were taken with a ColonyDoc-It (UVP).
3.5.2 GFP response assays
Pre-cultures of the strains PRa74 (background), PRa78 (AD₁) and PRa79 (AD₂) were inoculated from a single colony of the strain in SC medium (2 % glucose, pH = 5.6) lacking leu, trp and histidine. Following 18 hours of cultivation at 30 degrees C, 175 rpm horizontal shaking, 200 μL microtiter main cultures were inoculated from these in 75 % SC medium (diluted with milliQ water and back-standardized to 2 % glucose) with the relevant concentrations of tetracycline added. The cultures were sealed with a gas-permeable Breathseal (Greiner bio-one) and plastic lid and were cultured in a horizontal shaker (Innova) at 30 degrees C, 300 rpm shaking. Following 16 hours of cultivation, the cultures were measured by flow cytometry on a BD LSRFortessa Cell Analyzer using a FITC filter with collection limit set to 10,000 cells. The mean FITC intensity for each sample was reported. The measurements from the GFP-devoid PRa74 strain were used for background-subtraction.

3.5.3 Mathematical model
The mathematical model is based on ordinary differential equations describing formation of RNA and protein for the different molecular components of the system. The model is described in detail in Supporting Info.
3.6 Acknowledgements

We kindly thank Marius Boulos Faza and Luisa Gronenberg for helpful comments and Niels Bjerg Jensen for helpful suggestions on gene deletion.

This research was funded by the Novo Nordisk Foundation and the European Union Seventh Framework Programme (FP7-KBBE-2013-7-single-stage) under Grant agreement no. 613745, Promys.
3.7 References


Palani, S., Sarkar, C.A., 2011. Synthetic conversion of a graded receptor signal into a tunable,


Supporting Information for:

Molecular buffers permit robust transduction, sensitivity tuning and inversion of riboswitch signals

By Peter Rugbjerg, Hans J. Genee, Kristian Jensen and Morten O. A. Sommer
S1 Increased false negative cells

The occurrence of falsely colony-forming *S. cerevisiae* that posses wild type *URA3* on FOA-containing medium has been described in literature and is around $10^{-7}$ for 0.1 % (w/v) FOA. By plating *S. cerevisiae* CEN.PK113-3C on SC plates containing 0.090 % (w/v) FOA we found a slightly elevated frequency at around $10^{-6}$. Similarly, we verified the complete resistance to FOA on the same plate type by plating *S. cerevisiae* CEN.PK113-5D devoid of functional *URA3*.

Given these outer boundaries of *URA3* selection, the observed in-between frequency of colony-forming *S. cerevisiae* cells bearing the direct riboswitch-regulated *URA3* gene (PRd5) shows that the tested basal *URA3* level was in its phenotypic dynamic range. *URA3* was thus not simply expressed too weakly or strongly to allow control of colony formation: If the observed lack of ability to correctly form colonies in presence of the riboswitch ligand could be explained by too leaky *URA3* expression, then the basal *URA3* expression would be so high that the simultaneous false positive colonies cannot be explained by too low basal *URA3* expression (in absence of the riboswitch ligand).
S2 Introduction to mathematical model

The model we developed to describe the two-component buffer system is based on calculating dissociation equilibria for the interactions involved: Tetracycline and the riboswitch, activation domain (AD) and DNA-binding domain (DBD), and DBD and DNA binding sites.

General binding scheme

Each of these interactions can be described by a general dissociation reaction $AB \rightleftharpoons A + B$, where $A$ is a ligand, $B$ is a receptor and $AB$ is the complex formed upon interaction between the two.

The equilibrium constant for such a reaction is given as $K_d = \frac{[A][B]}{[AB]}$. Brackets denote the actual concentration of each species.

The actual concentrations of $A$ and $B$ can also be expressed by their formal concentrations: $[A] = c_A - [AB]$ and $[B] = c_B - [AB]$.

Substituting these in the equilibrium equation gives:

$$K_d = \frac{(c_A - [AB]) \cdot (c_B - [AB])}{[AB]}$$

This equation can be used to derive an expression for how large a fraction of $B$ is bound to $A$.

This fraction is defined as $f = \frac{[AB]}{c_B}$

Solving this equation for $[AB]$ gives:

$$[AB] = \frac{K_d + c_A + c_B - \sqrt{K_d^2 + 2 \cdot K_d \cdot c_A + 2 \cdot K_d \cdot c_B + c_A^2 + c_B^2 - 2 \cdot c_A \cdot c_B}}{2}$$

This is then divided by the formal concentration of $B$ to give $f$:

$$f = \frac{[AB]}{c_B} = \frac{K_d + c_A + c_B - \sqrt{K_d^2 + 2 \cdot K_d \cdot c_A + 2 \cdot K_d \cdot c_B + c_A^2 + c_B^2 - 2 \cdot c_A \cdot c_B}}{2 \cdot c_B}$$

The ratio $\frac{[AB]}{c_B}$ is the fraction, $f$, of the total $B$ forming complex with the ligand $A$, given formal concentrations of $A$ and $B$, and the dissociation constant $K_d$, which is an inverse measure of the binding affinity between $A$ and $B$.

For any pair of ligand $A$ and receptor $B$, $f$ is the saturation quotient of the receptor.

The expression for $f$ can be used to describe both layers of regulation in the buffer device.

Graphs showing how $f$ varies with ligand concentration, exemplified by tetracycline binding to the riboswitch, are shown in Supplementary Info figure S1.

p. 70. Chapter 3: Molecular buffers
The formal concentration of receptor influences the shape of the curve.

**Supplementary Info figure S1:** Graphs depicting the riboswitch-ligand binding fraction, $f$, as function of tetracycline concentration for an mRNA concentration of 2 nM. The two graphs show the same relation plotted from a linear and logarithmic scale respectively. $K_d$ is here set to 0.77 nM $^2$.

**Riboswitch regulation**

In the riboswitch regulation of a transcript for e.g. AD, the model receptor is the riboswitch and its ligand is tetracycline. Thus $f$ is the fraction of riboswitches that are bound to tetracycline at a given time, or the probability that a single given riboswitch is bound to tetracycline at a given time. The value of $f$ impacts the rate of translation of the AD transcripts.

When $f = 0$, all the riboswitch transcripts are in the “open” conformation, giving the highest possible rate of translation.

When $f = 1$, all the riboswitch transcripts are in the “closed” conformation, giving the lowest possible rate of translation, describing the inherent leakiness of the regulation, caused by translation from tetracycline-bound riboswitch transcripts.

For values of $f$ in between, we assume that the translation rate varies linearly between the two extremes.

Based on this we calculate the AD rate of translation for given tetracycline concentrations:

$$
\text{transl} = \text{transl}_{\text{max}} - (\text{transl}_{\text{max}} - \text{transl}_{\text{min}}) \cdot f.
$$

The minimum and maximum translation values are model parameters matching the observed regulation values for the tc3 tetracycline-responsive riboswitch $^3$.

p. 71. Chapter 3: Molecular buffers
Buffer-mediated regulation of URA3

The general expression for $f$ can also be used to describe the transcriptional regulation of URA3 by cognate AD and DBD proteins.

The two interactions in this regulation can both be described as a relationship of a receptor and ligand. DBD acts as a ligand to the DNA and AD acts as a ligand to DBD.

Thus, assuming simple binding dynamics, the general expression for $f$ can both be used to model the fraction of DNA sites that is bound to DBD and the fraction of DBD that is bound to AD. Since we assume that the two interactions are independent of each other, we can multiply the two fractions to report the fraction of DNA sites that is bound by an AD-DBD complex, or the fraction of time that a given DNA site will be bound by an AD-DBD complex.

We expect that the affinity between DNA and DBD is much higher than between AD and DBD. Thus, at the relevant protein levels, the DNA-DBD binding fraction becomes very close to 1.

The fraction of DNA sites bound to an AD-DBD complex influences the transcription rate of URA3.

When the total binding fraction is 0, none of the DNA sites are bound to a functional complex, giving the lowest possible URA3 transcription.

When $f = 1$, all of the DNA sites are bound to a functional complex, giving the highest possible URA3 transcription.

For values of $f$ in between, we assume the expression varies linearly between these two extremes.

Since the total binding fraction is the product of the AD-DBD and DBD-DNA binding fractions, the total fraction varies proportionally with the AD-DBD fraction as long as the DBD-DNA fraction is held constant.

The modeled correlation between URA3 transcription and AD concentration is shown in Supplementary Info figure S2.
Supplementary Info figure S2: URA3 transcription increases with increasing levels of AD. The dotted lines represent the assumed minimum and maximum transcription rates when no DNA sites or all DNA sites respectively are bound to an AD-DBD complex. Transcription units are arbitrary. AD concentration is shown in multiples of the AD-DBD $K_d$. DBD concentration is set to 10 times the AD-DBD $K_d$ value.

By introducing transcription, translation and degradation rates, as well as the relevant dissociation constants, a dynamic ODE-based model of the protein concentration levels of AD, DBD and URA3 could be constructed. By incorporating the expression principles stated above, the full signal-buffer device was modeled in Python (full script in Supplementary Info appendix). The individual ODEs used to construct the model can be seen in Table S1.
Table S1: Overview of the ordinary differential equations used in the dynamic model.

<table>
<thead>
<tr>
<th>Species</th>
<th>Differential equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD mRNA</td>
<td>( \frac{d[ADmRNA]}{dt} = V_{\text{form}} - k_{\text{degrRNA}} \cdot [ADmRNA] )</td>
</tr>
<tr>
<td>DBD mRNA</td>
<td>( \frac{d[DBDmRNA]}{dt} = V_{\text{form}} - k_{\text{degrRNA}} \cdot [DBDmRNA] )</td>
</tr>
<tr>
<td>URA3 mRNA (driven by DBD:AD)</td>
<td>( \frac{d[URA3mRNA]}{dt} = GAL_{\text{min}} + (GAL_{\text{max}} - GAL_{\text{min}}) \cdot f_{\text{ADDBD}} \cdot f_{\text{DBDDNA}} - k_{\text{degrRNA}} ) \cdot [URA3mRNA]</td>
</tr>
<tr>
<td>AD protein</td>
<td>( \frac{d[AD]}{dt} = (\text{ribo}<em>\text{max} - (\text{ribo}</em>\text{max} - \text{ribo}<em>\text{min}) \cdot f</em>{\text{ribo}}) \cdot [ADmRNA] - k_{\text{degrprot}} \cdot [AD] )</td>
</tr>
<tr>
<td>DBD protein</td>
<td>( \frac{d[DBD]}{dt} = k_{\text{form}} \cdot [DBDmRNA] - k_{\text{degrprot}} \cdot [DBD] )</td>
</tr>
<tr>
<td>GAL4 mRNA</td>
<td>( \frac{d[GAL4mRNA]}{dt} = V_{\text{form}} - k_{\text{degrRNA}} \cdot [GAL4mRNA] )</td>
</tr>
<tr>
<td>GAL4 protein</td>
<td>( \frac{d[GAL4]}{dt} = (\text{ribo}<em>\text{max} - (\text{ribo}</em>\text{max} - \text{ribo}<em>\text{min}) \cdot f</em>{\text{ribo}}) \cdot [GAL4mRNA] - k_{\text{degrprot}} ) \cdot [GAL4]</td>
</tr>
<tr>
<td>URA3 mRNA (driven by GAL4)</td>
<td>( \frac{d[URA3mRNA]}{dt} = GAL_{\text{min}} + (GAL_{\text{max}} - GAL_{\text{min}}) \cdot f_{\text{GALADNA}} - k_{\text{degrRNA}} ) \cdot [URA3mRNA]</td>
</tr>
<tr>
<td>URA3 protein</td>
<td>( \frac{d[URA3]}{dt} = k_{\text{form}} \cdot [URAmRNA] - k_{\text{degrprot}} \cdot [URA3mRNA] )</td>
</tr>
</tbody>
</table>

The model parameters represent biochemical constants and conditions, e.g. individual transcription rates and dissociation constants and were generated by fitting the model to the GFP-based output while assuming biologically relevant values. The model is still likely to function best from a qualitative point of view. For example, even though the genes encoding AD and DBD are transcribed from identical promoters, it is not certain that their equilibrium concentrations are equal, because their individual translation rates might differ. The riboswitch in particular might reduce AD translation efficiency by an unknown factor.

We have used the equations from the dynamic model to calculate steady state concentrations of the six species, for different concentrations of riboswitch ligand and for different transcription levels of DBD. This was used to produce the heat-map of the URA3 concentrations (Figure 3.4A). Steady state concentrations were calculated by setting the ODEs equal to 0 and solving for the concentrations.

To predict the optimal AD\(_0\):DBD ratio in terms of providing the largest riboswitch-dependent regulation of URA3, we plotted the DBD saturation fraction by AD as a function of DBD
concentration for two different, but fixed AD levels. The two AD levels corresponded to full AD expression and down-regulated AD expression respectively. This plot is shown in Supplementary info figure S3.

The optimal DBD concentration is the concentration that results in the largest difference between saturation fractions for the two AD concentrations. It is seen from Supplementary Info figure S3 that the difference is largest when the DBD concentration roughly equals the unregulated AD concentration. This difference maximum is closer to the unregulated AD concentration, the higher the AD and DBD concentrations become in comparison to the AD-DBD $K_d$ value.

![Graph showing URA3 transcription rates](image)

**Supplementary info figure S3:** The solid graphs show the *URA3* transcription rate for increasing DBD levels, at fixed AD concentrations. For the red graph, AD concentration is 100 $K_d$, for the blue graph AD concentration is 100/37 $K_d$, assuming the 37-fold riboswitch regulation reported $^3$. The dotted green line represents the difference between the two transcription levels. The optimal DBD concentration for the given AD concentration can be found by locating the maximum of the difference.
Using the model it is possible to predict levels of the different proteins following a change in tetracycline concentration.

To reach biologically relevant, qualitative predictions, we made assumptions regarding the values of several parameters while fitting to the actual GFP data obtained. The value of each parameter can be seen in Table S2.
Table S2 Parameters found by fitting model to GFP output.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA degradation rate</td>
<td>$0.1 \text{ min}^{-1} \ (T_{1/2} \approx 7 \text{ min})$</td>
</tr>
<tr>
<td>Protein degradation rate</td>
<td>$0.05 \text{ min}^{-1} \ (T_{1/2} \approx 13 \text{ min})$</td>
</tr>
<tr>
<td>Constitutive transcription rate (AD and DBD)</td>
<td>$1 \text{ min}^{-1}$</td>
</tr>
<tr>
<td>Translation coefficient</td>
<td>$1 \left( \text{min} \cdot \text{RNA molecule} \right)^{-1}$</td>
</tr>
<tr>
<td>Actual riboswitch $K_d$</td>
<td>5</td>
</tr>
<tr>
<td>Riboswitch translation down-regulation fold</td>
<td>37</td>
</tr>
<tr>
<td>Basal riboswitch translation reduction fold</td>
<td>4</td>
</tr>
<tr>
<td>AD-DBD $K_d$</td>
<td>2</td>
</tr>
<tr>
<td>DBD-DNA $K_d$</td>
<td>0.02</td>
</tr>
<tr>
<td>GAL4 minimum transcription</td>
<td>0.6</td>
</tr>
<tr>
<td>GAL4 maximum transcription</td>
<td>31</td>
</tr>
<tr>
<td>Tetracycline concentrations</td>
<td>0 to 150 $\mu$M</td>
</tr>
</tbody>
</table>
To investigate whether the output regulatory effect of the buffer device could be achieved by using a complete transcription factor with the same DNA-binding affinity instead of the split-domain protein, we modeled this situation. With the same parameters representing promoter strengths and dissociation constants our simulations showed no regulation of *URA3*, in agreement with our experimental results of the same constructs (Fig. 3.3B). This lack of regulation can be explained by the assumed high binding affinity between the complete transcription factor and the DNA binding site, which lets *URA3* be transcribed at almost full strength even at very low GAL4 concentrations. Supplementary Info figure S4 shows how *URA3* levels nearly not respond to the addition of tetracycline in a model of systems where the riboswitch regulates complete GAL4 transcription factor.

Supplementary Info figure S4 Mathematically computed response curve of the target *URA3* gene with tetracycline input, for a system relaying the riboswitch signal to a full length GAL4 transcription factor with same DNA-binding affinity as DBD in the buffer network.
S4 Modeling tuning of detection threshold (sensitivity shift)

The heat map in Figure 3.4A was constructed by calculating steady state concentrations of URA3 at different tetracycline concentrations and transcriptional AD\(_0\):DBD ratios. The AD\(_0\):DBD ratio was varied by changing the DBD transcription rate while keeping AD transcription rate constant. The curves in Figure 4B were generated by plotting graphs for fixed values of AD\(_0\):DBD in the heat map corresponding to the two experimentally chosen ligand concentrations for which GFP-based response data was available.
S5 Re-establishing AD:DBD transcription ratio re-establishes response

To tune the riboswitch ligand detection level and test our model, we perturbed the relative transcription levels for AD and DBD. The effect of decreasing the AD:DBD ratio (PRA45) was an increase in false-positive colony formation on plates without the riboswitch ligand compared to the optimal strain PRA22 with a 1:1 transcriptional ratio of AD and DBD (Supplementary Info figure S5). The decrease of the ratio was engineered by expression of AD with the weak promoter CUP1, while still maintaining the strong ADH1 promoter for DBD. The effective gene regulation performance of PRA22 could be re-established by reducing the AD expression level, thus re-establishing the ratio of the transcription level of both genes (PRA46).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Promoter for AD</th>
<th>Promoter for DBD</th>
<th>AD:DBD</th>
<th>Colony formation response 0.1 % 5-FOA</th>
<th>Colony formation response 0.1 % 5-FOA, 100 µM tetracycline</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRA22</td>
<td>ADH1</td>
<td>ADH1</td>
<td>1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>PRA45</td>
<td>CUP1</td>
<td>ADH1</td>
<td>&lt;1</td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>PRA46</td>
<td>CUP1</td>
<td>CUP1</td>
<td>1</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Supplementary Info figure S5 Four different S. cerevisiae strains expressing the cognate ADs and DBDs of the buffer device by the indicated promoters. PRA22 expressed AD and DBD from strong promoters at 1:1 ratios, whereas this ratio was reduced in PRA45 leading to increased false positives. When re-establishing the 1:1 ratio of AD:DBD using weak promoters, the colony formation response was re-established. Refer to strain list for full strain genotypes. Each strain was spotted in ten-fold serial dilutions to SC-leu-trp plates supplemented with the indicated molecules. Photos show representative example of triplicates and were taken following incubation for four days at 30 degrees C.
Detailed methods and materials

Reagents
All reagents were purchased from Sigma-Aldrich unless otherwise specified. All oligonucleotides were purchased from Integrated DNA Technologies. Restriction enzymes were purchased from Thermo Scientific, USER enzyme was purchased from New England Biolabs. Throughout, milliQ water from a Barnstead Nanopure was used. Synthetic complete medium was based on Sigma-Aldrich Yeast Synthetic Drop-out Media Supplements without histidine, leucine, tryptophan and uracil (catalog no Y2001)

Plasmids
Plasmids used in the study are listed in Table S3. Plasmid maps for the plasmids pEXP22.1 and pEXP32.4 are shown in Supplementary Info figure S6. Details on the construction of plasmids are given below. Sequences for all plasmids are available upon request.
Table S3 Plasmids used in the study.

<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Relevant features</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pEXP22-RalGDSwt</td>
<td>pADH1-NLS-GAL4AD-RalGDS-tADH1, ARS/CEN, TRP1, AmpR</td>
<td>Life Technologies</td>
</tr>
<tr>
<td>pEXP32-Krev1</td>
<td>pADH1-GAL4DBD-Krev1-tADH1, ARS/CEN, LEU2, GenR</td>
<td>Life Technologies</td>
</tr>
<tr>
<td>pEXP22.1</td>
<td>pADH1-tc3-NLS-GAL4AD-RalGDS-tADH1, ARS/CEN, TRP1, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pEXP32.4</td>
<td>pADH1-GAL4DBD-Krev1-tADH1, pSPAL10-URA3-tURA3, ARS/CEN, LEU2, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pEXP22.2</td>
<td>pADH1-tc3-GAL4-tADH1, ARS/CEN, TRP1, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pEXP22.1CUP</td>
<td>pCUP1-tc3-NLS-GAL4AD-tADH1, ARS/CEN, TRP1, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pEXP32.4CUP</td>
<td>pCUP1-GAL4DBD-Krev1-tADH1, pSPAL10-URA3-tURA3, ARS/CEN, LEU2, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pRP4-URA3</td>
<td>pCUP1-tc3-URA3-tADH1, ARS/CEN, LEU2, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pEXP32.5XX</td>
<td>pSPAL10-2x3vGFP-tURA3, ARS/CEN, HIS3, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pEXP42.2</td>
<td>pADH1-tc3-NLS-GAL80-tADH1, ARS/CEN, HIS3, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pRS413-empty</td>
<td>HIS3, ARS/CEN, AmpR</td>
<td>4</td>
</tr>
<tr>
<td>pRS414-empty</td>
<td>TRP1, ARS/CEN, AmpR</td>
<td>4</td>
</tr>
<tr>
<td>pRS415-empty</td>
<td>LEU2, ARS/CEN, AmpR</td>
<td>4</td>
</tr>
<tr>
<td>pRS416-empty</td>
<td>URA3, ARS/CEN, AmpR</td>
<td>4</td>
</tr>
<tr>
<td>pESC-LEU</td>
<td>tADH1, 2 μ, LEU2, AmpR</td>
<td>Agilent Technologies</td>
</tr>
<tr>
<td>pESC-HIS</td>
<td>tADH1, 2 μ, HIS3, AmpR</td>
<td>Agilent Technologies</td>
</tr>
<tr>
<td>pUG6</td>
<td>loxp flanked kanMX, AmpR</td>
<td>5</td>
</tr>
<tr>
<td>pSH47</td>
<td>Cre recombinase, URA3, AmpR</td>
<td>5</td>
</tr>
<tr>
<td>pADH1-tc3-GFP</td>
<td>pADH1-tc3-GFP, URA3, AmpR</td>
<td>Euroscarf</td>
</tr>
<tr>
<td>pCU2</td>
<td>3vGFP, AmpR</td>
<td>6</td>
</tr>
</tbody>
</table>
Plasmid maps

Supplementary Info figure S6. Plasmid maps showing annotated features of the two plasmids pEXP22.1 and pEXP32.4.

Transformation of S. cerevisiae strains

All S. cerevisiae transformations were performed following standard lithium-acetate procedures.

S. cerevisiae strains

Parent strains

Saccharomyces cerevisiae MaV203 (MATa, leu2-3,112, trp1-901, his3Δ200, ade2-101, gal4Δ, gal80Δ, SPAL10::URA3, GAL1::lacZ, HIS3::HIS3@LYS2, can1R, cyh2R). Purchased from Life Technologies.

Saccharomyces cerevisiae PRa18 (MATa, leu2-3,112, trp1-901, ura3Δ, his3Δ200, ade2-101, gal4Δ, gal80Δ, GAL1::lacZ, HIS3::HIS3@LYS2, can1R, cyh2R). Derived from MaV203 as described in detail elsewhere.

Saccharomyces cerevisiae PRa26 (MATa, leu2-3,112, trp1-901, ura3Δ, his3Δ200, ade2-101, gal4Δ, gal80Δ, GAL1::lacZ, KanMX, can1R, cyh2R). Derived from PRa18 as described in detail elsewhere.

Saccharomyces cerevisiae CEN.PK113-3C (MATa, trp1-289, MAL2-8c, SUC2). Obtained from Euroscarf.

Saccharomyces cerevisiae CEN.PK113-5D (MATa, ura3-52, MAL2-8c, SUC2). Obtained from Euroscarf.
Derived strains
From the parent *S. cerevisiae* strains, the derived strains in Table S4 were constructed by transformation with the plasmids listed in Table S3.

Table S4 *S. cerevisiae* strains derived from the indicated parent strains in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Plasmid #1</th>
<th>Plasmid #2</th>
<th>Plasmid #3</th>
<th>Parent strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRa22</td>
<td>pEXP22.1</td>
<td>pEXP32.4</td>
<td>-</td>
<td>PRa18</td>
</tr>
<tr>
<td>PRa28</td>
<td>pEXP22.2</td>
<td>pRS415-empty</td>
<td>-</td>
<td>MaV203</td>
</tr>
<tr>
<td>PRa45</td>
<td>pEXP22.1CUP</td>
<td>pEXP32.4</td>
<td>-</td>
<td>PRa18</td>
</tr>
<tr>
<td>PRa46</td>
<td>pEXP22.1CUP</td>
<td>pEXP32.4CUP</td>
<td>-</td>
<td>PRa18</td>
</tr>
<tr>
<td>PRa47</td>
<td>pEXP22.1</td>
<td>pEXP32.4CUP</td>
<td>-</td>
<td>PRa18</td>
</tr>
<tr>
<td>PRd5</td>
<td>pRP4-URA3</td>
<td>pRS414-empty</td>
<td>-</td>
<td>PRa18</td>
</tr>
<tr>
<td>PRa78 “AD₁”</td>
<td>pEXP22.1</td>
<td>pEXP32.4</td>
<td>pEXP32.5XX</td>
<td>PRa26</td>
</tr>
<tr>
<td>PRa79 “AD₂”</td>
<td>pEXP22.1</td>
<td>pEXP32.4CUP</td>
<td>pEXP32.5XX</td>
<td>PRa26</td>
</tr>
<tr>
<td>PRa84</td>
<td>pEXP22-RalGDSwt</td>
<td>pEXP32.4</td>
<td>pEXP42.2</td>
<td>PRa26</td>
</tr>
</tbody>
</table>

Construction of plasmids
All plasmids cloned in this study were constructed using uracil-excision cloning \(^8\) of PCR-amplified fragments into plasmids i.e. cloning was performed without obligatory scars and into a USER cassette. For each constructed plasmid, all respective fragments are listed in Table S5 with details of the primers and DNA template used. Primer sequences are given in Table S6. All PCRs were run using the uracil-compatible Pfu DNA polymerase mutant X7 \(^9\) with HF buffer (Thermo Scientific). The reaction was transformed into XL1 Blue competent cells (Stratagene) or DH5-alpha competent cells (Life Technologies) and correctly cloned plasmids were subsequently identified.

Crude extraction of *S. cerevisiae* gDNA
*S. cerevisiae* gDNA for cloning was extracted by dissolving a colony in a 100 μL solution of 200mM lithium-acetate, 1 % SDS and incubated 5 min at 70 degrees C. 300 μL 96 % ethanol
was added, the mixture vigorously vortexed and spun down at 15,000 g, 3 min. The pellet was washed with 400 μL 70 % ethanol, and after spin-down dissolved in 100 μL milliQ water. The solution was spun down at 15,000 g for 15 seconds after which the supernatant was used as 50x gDNA-containing PCR template.

Table S5 List of PCR fragments that were assembled into the respective plasmids in this study. For each PCR fragment an oligo pair and template DNA is given.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Oligo pair</th>
<th>Template DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRP2-URA3</td>
<td>P9+P10</td>
<td>pRS416-empty</td>
</tr>
<tr>
<td></td>
<td>P2+P23</td>
<td>CEN.PK113-3C gDNA</td>
</tr>
<tr>
<td></td>
<td>P11+P3</td>
<td>pADH1-tc3-GFP</td>
</tr>
<tr>
<td></td>
<td>P21+P24</td>
<td>pESC-LEU</td>
</tr>
<tr>
<td>pRP4-URA3</td>
<td>P13+P16</td>
<td>pRP2-URA3</td>
</tr>
<tr>
<td></td>
<td>P14+P15</td>
<td>pRS414-empty</td>
</tr>
<tr>
<td>pEXP22.1</td>
<td>P39+P40</td>
<td>pADH1-tc3-GFP</td>
</tr>
<tr>
<td></td>
<td>P13+P38</td>
<td>pEXP22-RalGDSwt</td>
</tr>
<tr>
<td></td>
<td>P14+P42</td>
<td>pEXP22-RalGDSwt</td>
</tr>
<tr>
<td>pEXP32.4</td>
<td>P57+P56</td>
<td>CEN.PK113-3C gDNA</td>
</tr>
<tr>
<td></td>
<td>P13+P53</td>
<td>pEXP22-RalGDSwt</td>
</tr>
<tr>
<td></td>
<td>P55+P54</td>
<td>MaV203 gDNA</td>
</tr>
<tr>
<td></td>
<td>P14+P52</td>
<td>pEXP32-Krev1</td>
</tr>
<tr>
<td>pEXP22.2</td>
<td>P40+P81</td>
<td>pEXP22.1</td>
</tr>
<tr>
<td></td>
<td>P79+P80</td>
<td>CEN.PK113-3C gDNA</td>
</tr>
<tr>
<td>pEXP22.1CUP</td>
<td>P132+P39</td>
<td>pEXP22.1</td>
</tr>
<tr>
<td></td>
<td>P131+P130</td>
<td>CEN.PK113-3C gDNA</td>
</tr>
<tr>
<td>pEXP32.4CUP</td>
<td>P54+P135</td>
<td>pEXP32.4</td>
</tr>
<tr>
<td></td>
<td>P133+P134</td>
<td>CEN.PK113-3C gDNA</td>
</tr>
<tr>
<td>pEXP32.5XX</td>
<td>P54+P55</td>
<td>pEXP32.4</td>
</tr>
<tr>
<td></td>
<td>P84+P219</td>
<td>pCU2</td>
</tr>
<tr>
<td></td>
<td>P209+P218</td>
<td>pCU2</td>
</tr>
<tr>
<td></td>
<td>P60+P15</td>
<td>pEXP32.4</td>
</tr>
<tr>
<td></td>
<td>P16+P214</td>
<td>pESC-HIS</td>
</tr>
<tr>
<td>pEXP42.2</td>
<td>P249+P250</td>
<td>CEN.PK113-3C gDNA</td>
</tr>
<tr>
<td>Oligo ID</td>
<td>Sequence (5’-3’)</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>ATGACAGAUAUTATGTGATGATTGATTGATTGATTG</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>ATCTGTCAUAATTATCTACTTTAAGGCCTAAAAACAT</td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td>CGTGCAGAUTTAGTTTTGCTGGGCGCA</td>
<td></td>
</tr>
<tr>
<td>P10</td>
<td>ACATATGUCGAAAGCTACATATAAGGAACGTG</td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>ACATATGUTCTCGAGGCCTAGGTGGT</td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>ACCCAGGUGGCACTTTTCGGGGAAAT</td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>ACCTGGGUCCTTTTCATACGTGCA</td>
<td></td>
</tr>
<tr>
<td>P15</td>
<td>AAGACGGGAUGCCTTGCTGCTAACCAC</td>
<td></td>
</tr>
<tr>
<td>P16</td>
<td>ATCCGTCUAAAGAAAAACCATTATTTATCATGACAT</td>
<td></td>
</tr>
<tr>
<td>P21</td>
<td>ACTAAAUCGCGACGATTCGAGCTCT</td>
<td></td>
</tr>
<tr>
<td>P23</td>
<td>ATCCATUACCAGACATTTTGGGGCGCA</td>
<td></td>
</tr>
<tr>
<td>P24</td>
<td>AATGGAUUCGCGAATTTTCTGAGT</td>
<td></td>
</tr>
<tr>
<td>P38</td>
<td>AGCTTGAAAUAATTGTGCAGAAA</td>
<td></td>
</tr>
<tr>
<td>P39</td>
<td>ATTTCAAGCGCGCGCGC</td>
<td></td>
</tr>
<tr>
<td>P40</td>
<td>ATATGTUCTCGAGGCGCTAGGTG</td>
<td></td>
</tr>
<tr>
<td>P41</td>
<td>AACATAUGAAGCTACTGTCTTCTATCGA</td>
<td></td>
</tr>
<tr>
<td>P42</td>
<td>AACATAUGCCAAAGAGAAGCCG</td>
<td></td>
</tr>
<tr>
<td>P52</td>
<td>AGGACGGGAUCGAAAGAAAAGATGGTAA</td>
<td></td>
</tr>
<tr>
<td>P53</td>
<td>ATCGGAUUCTGAGGGGG</td>
<td></td>
</tr>
<tr>
<td>P54</td>
<td>ATCCGTCUUGGAAGTCTCATGGAGATT</td>
<td></td>
</tr>
<tr>
<td>P55</td>
<td>ATATTCCUCGACTCACTTCAATC</td>
<td></td>
</tr>
<tr>
<td>P56</td>
<td>AGAATAATUATGTCGAAAGCTACATATAAGGAA</td>
<td></td>
</tr>
<tr>
<td>P57</td>
<td>ATCCGGAUUCTGCCATTTTACATG</td>
<td></td>
</tr>
<tr>
<td>P58</td>
<td>AGAATAAATUGTCAGGGCGCTAAACAT</td>
<td></td>
</tr>
<tr>
<td>P59</td>
<td>AGTTTTTTAGTTTTTGGCTGGGCGCA</td>
<td></td>
</tr>
<tr>
<td>P60</td>
<td>AAAAACUGTATTATAAGGAATGCATG</td>
<td></td>
</tr>
</tbody>
</table>

Table S6: Oligonucleotides used to clone the plasmids constructed in the study.
| P79  | AACATAUGAAGCTACTGCTTCTATCGAAC |
| P80  | AGCGGCCATUACTCTTTTTTGGGTGTTGG |
| P81  | AATGGGCGCUAAAGTAAGTAAGACGTC  |
| P84  | AGAATAATUATGTCTAAAGGTGAAGAATTATTCATG |
| P87  | ACCCATCUUGTCGGGTATTTTCAC |
| P88  | AGATGCGUAAGGAGAAAATACCG |
| P110 | ACCTTCCGCTUTCTTCTTGGGC |
| P130 | AGCTTTGAAAUTATGTGATGATTGATTGATTG |
| P131 | ATCCGCCAUTACCGACATTG |
| P132 | ATGGCGGAUUGGTACCAGCTTTTGTCC |
| P133 | AGGACGGAUCCATTACCGACAT |
| P134 | ATGTGATGATGATTGATTGATTGATGAT |
| P135 | ATCATCACAUAGCTTGAAGCAAGCCT |
| P209 | AGTTTTTUATTGTACAGTTCATCCATACCA |
| P212 | AAGGAGGUATGGCTAGCAAAGGAGAAGAA |
| P213 | ACCTCCUTTGTAACATCATTCCATCCAC |
| P214 | AGGACGGAUATCGCACGCATTCGAGC |
| P218 | AAGGGAUGTCTAAAGGTGAAGAATTATTC |
| P219 | ATCCCTUTGTACAGTTCATCCATACCAT |
| P249 | AGCGGAAGGUACATGGACTCAACAAAGAGACT |
| P250 | AGCGGCCCATUAATAACTATAATGCGAGATATTGC |
Preparation of assay plates with FOA

To prepare FOA assay plates, synthetic complete medium in double concentration lacking leucine and tryptophan was supplemented with 0.090 % (w/v) 5-fluoroorotic acid (Thermo Scientific) followed by sterile filtration. Before addition of FOA, pH was adjusted to 4.5. To cast plates, the double concentration medium was mixed in equal volumes with a melted 4 % agar solution. To assay plates further supplemented with tetracycline were added the respective volume of tetracycline hydrochloride from a fresh-made 20 mM stock solution. We noticed some batch variation presumably on the synthetic complete drop-out medium, which meant that with a particular batch, adjustments of FOA content was necessary to recreate the observations, perhaps due to differences in induction strength of the CUP1 promoter.
Supporting Information References


#!/usr/bin/python

import numpy as np
import matplotlib.pyplot as plt

# --- Functions ---------------------

def saturation(ligand, receptor, kd):
    '''Calculates a saturation fraction for a receptor, given formal
concentrations
of the ligand and receptor, as well as their mutual Kd-value'''
    f = (kd + ligand + receptor - ((kd + ligand + receptor)**2 -
4*ligand*receptor)**(0.5))/(2.0*receptor)
    return f

    if receptor > 0:
        f = (kd + ligand + receptor - ((kd + ligand + receptor)**2 -
4*ligand*receptor)**(0.5))/(2.0*receptor)
    else:
        f = 0
    return f

def ribokform(ligand, ribo, kd, min, max):
    '''Calculates the translation kform value for an off-riboswitch, at
given
concentrations of ligand and mRNA'''
    kform = max - (max - min) * saturation(ligand, ribo, kd)
    return kform

def galvform(gal, dna, kd, min, max):
    '''Calculates the mRNA vform value for a GAL4 regulated gene, at given
concentrations
of AD and DBD'''
    vform = min + (max - min) * saturation(gal, dna, kd)
    return vform

def y2hvform(ad, dbd, kd1, kd2, min, max):
    '''Calculates the mRNA vform value for a GAL4 regulated gene, at given
concentrations
of AD and DBD'''
    vform = min + (max - min) * saturation(ad, dbd, kd1) * saturation(dbd,
1, kd2)
    return vform

def calcURA3(TET, DBDvform):
    '''Calculates the URA3 output at a given tetracycline concentration and
DBD transcription
rate.'''
    admrna = vformad/kdegrrna
    dbdmrna = DBDvform/kdegrrna
    dbd = kform*dbdmrna/kdegrprot
    ad = ribokform(TET, admrna, ribokd, ribomin, ribomax)*admrna/kdegrprot
    ura3mrna = y2hvform(ad, dbd, galdnk, dnakmin, galmax)/kdegrrna
    ura3 = kform*ura3mrna/kdegrprot
    return ura3

def calcURA3_fullLengthGal(TET):
    '''Calculates the URA3 output at a given tetracycline concentration'''
    gal4rna = vformad/kdegrrna
    gal4 = ribokform(TET, gal4rna, ribokd, ribomin, ribomax)*gal4rna/kdegrprot
    ura3mrna = galvform(gal4, l, dnakd, galmin, galmax)/kdegrrna
    ura3 = kform*ura3mrna/kdegrprot
    return ura3

# --- Script ------------------------
# Model parameters:

# System constants:
kdgrrna = 0.1  # The rna degradation coefficient
kdegrprot = 0.05 # The protein degradation coefficient
vformad = 1 # The transcription rate (ad)
kf = 1.0 # The translation coefficient

# Riboswitch constants:
ribokd = 5  # Dissociation constant
ribomin = kform/(4*37)  # The riboswitch minimum kform
ribomax = kform/4  # The riboswitch maximum (basal) kform

# GAL4 constants:
galkd = 2 # Dissociation constant AD-DBD
dnakd = 0.02 # Dissociation constant DBD-DNA
galmin = 0.6 # GAL4 minimum transcription
galmax = 31.0 # GAL4 maximum transcription

# Arrays of tet concentrations and DBD transcription rates:
Tetconcs = np.arange(0,150,0.1)
transcription_ratios = np.arange(0.01,15,0.01)
Vformdbds = vformad/transcription_ratios

# Meshgrids of the variables:
tetconcs,vformdbds = np.meshgrid(Tetconcs,Vformdbds)

# Experimental tet concentrations:
exptc = np.array([0,0.5,1,5,15,50,100,150])

# Experimental GFP values:
a79gfp = np.array([5244,5238.67,5102.33,3269.67,2792.67,1172,746.67,573.33]).astype('float')
a79gfpe = np.array([101.06,206.40,86.75,297.94,75.47,144.33,200.34,63.64]).astype('float')
a78gfp = np.array([2745.67,2630.33,2358,1625.67,1201.67,286.67,345.67,284.33]).astype('float')
a78gfpe = np.array([85.70,190.26,108.41,135.88,37.42,143.72,223.73,78.48]).astype('float')

# Indices corresponding to the dbd transcription rates
slice1 = 170 # (DBD transcription rate: vformad/1.71)
slice2 = 440 # (DBD transcription rate: vformad/4.41)
maxtet = tetconcs.max()+10
surv_thresh = 230

# URA3 steady state simulation:

# Equilibrium concentrations are calculated by solving the
# mass balance for each substance
admrna = vformad/kdegrrna
dbdmrna = vformdbds/kdegrrna
db = kform*dbmrna/kdegrprot
ad = ribokform(tetconcs,admrna,ribokd,ribomin,ribomax)*admrna/kdegrprot
ura3mrna = y2hvform(ad,db,galkd,dnakd,galmin,galmax)/kdegrrna
ura3 = kform*ura3mrna/kdegrprot

# Plotting:

# Fig1: Heat map with colorbar:
plt.imshow(ura3,cmap='gist_rainbow',origin='lower',extent=[Tetconcs.min(),Tetconcs.max(),transcription_ratios.min(),transcription_ratios.max()],aspect='auto')
plt.colorbar()
plt.xlabel("Tet concentration")
plt.ylabel("AD transcription / DBD transcription")

# plt.savefig("heatmap.svg")

# Fig2: Graphs of 'slices', corresponding to particular DBD transcription rates:
fig = plt.figure()
plt.suptitle("Comparison of experimental results and model predictions")

# Subplot 1:
ax2 = fig.add_subplot(221)
plt.errorbar(exptc,a79gfp, yerr = a79gfpe, zorder=1, fmt='ko')
plt.errorbar(exptc,a78gfp, yerr = a78gfpe, zorder=1, fmt='ko')
plt.plot(tetconcs[slice2, :], ura3[slice2, :], 'b-', [0, maxtet], [surv_thresh, surv_thresh], 'y--')
plt.plot(tetconcs[slice1, :], ura3[slice1, :], 'b-', [0, maxtet], [surv_thresh, surv_thresh], 'y--')
plt.plot(exptc, (a79gfp), 'bo')
plt.ylim([0, 6000])
plt.xlim([0, maxtet])
plt.xlabel("Tet concentration")
plt.ylabel("URA3 concentration (steady state)"
plt.title("Linear scale")

# Subplot 2:
ax3 = add_subplot(222)
plt.errorbar(exptc,a79gfp, yerr = a79gfpe, zorder=1, fmt='ko')
plt.errorbar(exptc,a78gfp, yerr = a78gfpe, zorder=1, fmt='ko')
plt.plot(tetconcs[slice2, :], ura3[slice2, :], 'b-', [0.3, maxtet], [surv_thresh, surv_thresh], 'y--')
plt.plot(tetconcs[slice1, :], ura3[slice1, :], 'b-', [0.3, maxtet], [surv_thresh, surv_thresh], 'y--')
plt.plot(exptc, (a79gfp), 'bo')
plt.ylim([0, 6000])
plt.xlim([0.3, 200])
plt.xlabel("Tet concentration")
plt.xscale('log')
plt.title("Log scale")

# plt.savefig("response_curves.svg")

# Fig3: Simulated graph of the response curve of a circuit with a full-length GAL4 TF.
fig3 = plt.figure()
plt.plot(np.arange(0, maxtet, 0.1), calcURA3_fullLengthGal(np.arange(0, maxtet, 0.1)))
plt.ylim([0, 6500])
plt.title("Response curve GAL4 regulated by riboswitch")
plt.xlabel("Tet concentration")
plt.ylabel("URA3")

# plt.savefig("full_length_gal.svg")

# :::Calculation of goodness of fit measures:::

# RMSD:
a78rmse = 0
a79rmse = 0
for i, tetconc in enumerate(exptc):
a78rmse += ((a78gfp[i]) / 1 - calcURA3(tetconc, vformdbds[slice1, 0])) ** 2
a79rmse += ((a79gfp[i]) / 1 - calcURA3(tetconc, vformdbds[slice2, 0])) ** 2
a78rmse = a78rmse ** (0.5)
a79rmse = a79rmse ** (0.5)
print "a78 rmsd is:", a78rmsd
print "a79 rmsd is:", a79rmsd

# R-squared:
a78mean = a78gfp.mean()
a78ssres = 0
a78sstot = 0
a79mean = a79gfp.mean()
a79ssres = 0
a79sstot = 0

for i, tetconc in enumerate(exptc):
a78ssres += (a78gfp[i] - calcURA3(tetconc, vformdbds[slice1, 0]))**2
a78sstot += (a78gfp[i] - a78mean)**2
a79ssres += (a79gfp[i] - calcURA3(tetconc, vformdbds[slice2, 0]))**2
a79sstot += (a79gfp[i] - a79mean)**2

print "a78 R^2:", 1-a78ssres/a78sstot
print "a79 R^2:", 1-a79ssres/a79sstot

plt.show()
4 Flexible metabolic pathway construction using modular and divisible selection gene regulators

Authors:

Peter Rugbjerg – petru@biosustain.dtu.dk

Nils Myling-Petersen – nimyp@bio.dtu.dk

Morten O. A. Sommer – msom@bio.dtu.dk

a) Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark
Kogle Allé 6, DK-2970 Hørsholm, Denmark

* Corresponding author
Flexible metabolic pathway construction using modular and divisible selection gene regulators

Peter Rugbjerg, Nils Myling-Petersen, Morten O.A. Sommer *

Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, DK-2970 Hørsholm, Denmark

A R T I C L E   I N F O

Article history:
Received 15 December 2014
Received in revised form
10 August 2015
Accepted 11 August 2015
Available online 21 August 2015

Keywords:
Pathway construction
Selection system
Chromosomal integration
Synthetic biology
Split transcription factor

A B S T R A C T

Genetic selections are important to biological engineering. Although selectable traits are limited, currently each trait only permits simultaneous introduction of a single DNA fragment. Complex pathway and strain construction however depends on rapid, combinatorial introduction of many genes that encode putative pathway candidates and homologs. To triple the utility of existing selection genes, we have developed divisible selection in Saccharomyces cerevisiae. Here, independent DNA fragments can be introduced and selected for simultaneously using a set of split hybrid transcription factors composed of parts from Escherichia coli LexA and Herpes simplex VP16 to regulate one single selectable phenotype of choice. Only when co-expressed, these split hybrid transcription factors promote transcription of a selection gene, causing tight selection of transformants containing all desired DNA fragments. Upon transformation, 94% of the selected colonies resulted strictly from transforming all three modules based on ARS/CEN plasmids. Similarly when used for chromosome integration, 95% of the transformants contained all three modules. The divisible selection system acts dominantly and thus expands selection gene utility from one to three without any genomic pre-modifications of the strain. We demonstrate the approach by introducing the fungal rubrofusarin polyketide pathway at a gene load of 11 kb distributed on three different plasmids, using a single selection trait and one yeast transformation step. By tripling the utility of existing selection genes, the employment of divisible selection improves flexibility and freedom in the strain engineering process.

© 2015 International Metabolic Engineering Society. Published by Elsevier Inc. All rights reserved.

1. Introduction

Metabolic engineering and synthetic biology research necessitates frequent introduction of several pieces of heterologous DNA to host strains at ever-increasing pace (Bornscheuer et al., 2012; Boyle and Silver, 2012; Keasling, 2012; Tyo et al., 2007). The rising complexity of reconstructed metabolic pathways demands high flexibility and freedom in these genetic manipulations to characterize the full space of variables.

Various approaches have been developed for combinatorial assembly of single or several pathway steps using yeast homologous recombination, software-guided cloning or simply large random clone assemblies (Genee et al., 2014; Gibson et al., 2010; Naesby et al., 2009; Shao et al., 2009). Systematic sampling of such biological variables as expression strength, codon usage, truncations and homologs is a strategy for developing metabolic pathways, where optimally performing strains may result from particular combinatorial clones. Freedom and flexibility in gene introduction methods are thus of importance.

Irrespective of gene introduction strategy, the selection genes that aid transformation with foreign DNA still follow the same principle of “one selectable trait, one selection gene”. Consequently, selection genes are iteratively removed to allow recycled use when multi-gene systems are constructed, since selective traits are limited. Indeed, many strain construction methods are based on different approaches to recycling, using loop-out mechanisms based on recombinases or endogenous homologous recombination in increasingly streamlined implementations (Da Silva and Srikrishnan, 2012; Hegemann and Heick, 2011; Jensen et al., 2013; Mikkelsen et al., 2012; Siddiqui et al., 2014; Wingler and Comish, 2011). Still, it would increase speed if more independent gene introductions were possible using the limited number of available selectable traits. Indeed, new selection genes are routinely identified to further expand strain construction freedom (Regenberg and Hansen, 2000; Solis-Escalante et al., 2013). Whilst some are dominant in action, the recessive nature of most auxotrophic selection genes requires pre-modification of the receiving host strain by deletion of complementing prototrophic genes. The converse dominant selection genes typically involve antibiotics for maintenance of selection pressure, including harmful, costly agents such as bleomycin and hygromycin. Present-day pathway construction can
thus yield strains requiring a mixture of selection pressures (Nielsen et al., 2014).

_Saccharomyces cerevisiae_ is an important host for production of biofuels and commodity chemicals as well as structurally more advanced secondary metabolites, and the pathway complexities can only be expected to rise (Nielsen et al., 2013; Siddiqui et al., 2011).

As a modular and conceptually novel approach to manipulate the utility of each selectable trait in _S. cerevisiae_, we here design and develop divisible selection based on split transcription factors (TFs). With divisible selection, three independent DNA fragments can be introduced simultaneously at the load of a single selection phenotype of choice. This is possible by dividing the regulation of selectable phenotypes into modules composed of hybrid split TFs. Exclusively when co-expressed, these reconstitute activation of a tight selection gene promoter, yielding efficient selection for all associated DNA fragments. The split TFs of divisible selection are designed to function dominantly and therefore expand the utility of an already functional selection gene from one to three, and possibly more, DNA fragments without any strain pre-modifications. Split divisible phenotypes could therefore take many applications within biological engineering. As a proof-of-principle for the system, we here employ divisible selection within metabolic engineering to reconstruct the _Fusarium graminearum_ three-step polyketide pathway to rubrofusarin in _S. cerevisiae_ by expression of the biosynthetic genes from three individual plasmids under one selectable trait.

2. Methods and materials

2.1. Strains

All characterization of the system efficiencies was performed in _S. cerevisiae_ CEN.PK2-1C (MATα; _ura3-52_; _trp1-289_; _leu2-3,112_; _his3Δ 1_; _MAL2-88_; _SUC2_). The three-step rubrofusarin polyketide pathway was introduced into _S. cerevisiae_ CEN.PK2-1C pre-transformed with pRS413-npgA (CEN.PK2-1C-npgA) yielding the strains listed in Table 1.

All molecular cloning was performed using transformation into _Escherichia coli_ XL1 chemically competent cells (Stratagene).

2.2. Materials

Unless otherwise stated, reagents were purchased from Sigma-Aldrich. Synthetic complete (SC) medium was prepared with 1.4 g/L synthetic complete drop-out mix lacking uracil, tryptophan, leucine and histidine (Y2001), 6.7 g/L yeast nitrogen base without amino acids (YP0626) and 20 g/L D-glucose, pH standardized to 5.6. 

Cu^{2+} for gene induction was added from a stock solution of 100 mM CuSO4. When SC was supplemented with additional amino acids, 60 mg/L Leucine, 20 mg/L uracil, 20 mg/L histidine-HCl and 20 mg/L tryptophan was added. 

Oligonucleotides were purchased from Integrated DNA Technologies.

Table 1

<table>
<thead>
<tr>
<th>Strain</th>
<th>Pathway plasmids</th>
<th>Selective gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>cl-rub</td>
<td>pRS416-PKS12</td>
<td>URA3</td>
</tr>
<tr>
<td></td>
<td>pRS414-aufZ</td>
<td>TRP1</td>
</tr>
<tr>
<td></td>
<td>pRS415-aufJ</td>
<td>LEU2</td>
</tr>
<tr>
<td>ds-rub</td>
<td>pDS1-PKS12</td>
<td>ds1-URA3</td>
</tr>
<tr>
<td></td>
<td>pDS2-aufZ</td>
<td>ds2-DBD</td>
</tr>
<tr>
<td></td>
<td>pDS3-aufJ</td>
<td>ds3-AD</td>
</tr>
</tbody>
</table>

2.3. Construction of plasmids

Plasmids were assembled by uracil-excision (USER) cloning of purified PCR fragments, except for pDS2-aufZ, which was cloned by Gibson assembly of a purified PCR fragment into a pDS2 vector linearized with SmaI. Gibson assembly was performed using 2 × Gibson assembly master mix (New England Biolabs) according to the manufacturer’s protocol. The general method for USER cloning was based on agarose gel-purification of the PCR products amplified with DNA polymerase X7 (Nørholm, 2010). These were mixed and incubated in an equimolar 20 μL reaction with 0.5 μL USER enzyme (New England Biolabs) and 0.5 μL DpnI FastDigest (Thermo Scientific) in FastDigest buffer at 37°C for 1–2 h. Following 25 subsequent minutes at room temperature, 2.5 μL reaction was transformed into chemically competent _E. coli_. Correctly cloned plasmids were identified using restriction analysis and DNA sequencing. The detailed use of oligonucleotides for assembly of all plasmids and origin of parts are listed in supplementary material.

2.4. Plasmids

All plasmids developed and characterized for divisible selection are listed in Table 2 and all plasmids for production of rubrofusarin pathway are listed in Table 3.

2.5. Transformation of _S. cerevisiae_

_S. cerevisiae_ was transformed using lithium acetate–type transformation based on a high-efficiency protocol (Gietz and Schiestl, 2007) with a few minor adjustments: a single colony was pre-cultured overnight at 30°C, 250 rpm shaking in 10 mL yeast peptone dextrose (YPD) medium. On the day of transformation, individual main cultures of 10 mL YPD were each inoculated to OD₆₀₀ = 0.20 and cultured for 4 h at 30°C, 250 rpm shaking. The cell pellet of each was used per transformation. For chromosomal integration, to liberate 3 μg of each integration construct from its vector, 5 μg of each the integration vector were first linearized with Smal FastDigest (Thermo Scientific) in a 40 μL reaction with 1 × Fastdigest buffer for 2 h at 37°C and purified with a PCR clean-up kit (Macherey-Nagel) with elution into 20 μL H₂O. The exact PEG₃₅₀₀ concentration was critical for good transformation efficiencies, and test titrations of every 50% PEG₃₅₀₀ batch were made to assure this by varying the volume of the batch ± 10% in 2% point increments. The resultant optimum was used through compensation by adding less/more H₂O to the transformation mixture. Gentle final resuspension of the transformed cells in H₂O was also important for the transformation efficiency: after removal of the transformation mix supernatant and addition of H₂O, the sample was left to recover at room temperature for ten minutes prior to gentle resuspension using a pipette, first mildly tapping on the cell pellet before slowly pipetting it up and down two-three times. For chromosomal integrations this recovery was extended to twenty minutes. Each transformation was plated on one or two plates of synthetic complete (SC) medium lacking the appropriate selection nutrient. Colonies were counted following 65 h of incubation at 30°C. All transformations described were carried out in three or six replicates and no replicates were conducted in parallel.

2.6. Reconstruction of fungal polyketide pathway by divisible selection

_S. cerevisiae_ CEN.PK2-1C-npgA was first constructed through transformation of _S. cerevisiae_ CEN.PK2-1C with pRS413-npgA. Next, this strain was transformed with the plasmids pDS1U-PKS12,
Table 2
Plasmids for use in divisible selection and conventionally selected plasmids (pRS41x) used for comparison. All vectors propagate in S. cerevisiae using the indicated method and in E. coli using pUC origin of replication and ampicillin resistance. NLS abbreviates nuclear localization signal. Unless otherwise stated, gene source is S. cerevisiae.

<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Selection gene promoter</th>
<th>Selection gene ORF</th>
<th>Selection gene terminator</th>
<th>Propagation in yeast by</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDS1U</td>
<td>80pHm-lexA-SPO13</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS1Unm</td>
<td>80pHm-lexA-SPO13</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS1Ux2</td>
<td>80pHm-lexA-SPO13</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS1H</td>
<td>80pHm-lexA-SPO13</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS2</td>
<td>A. gossypii TEF1</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS2m</td>
<td>A. gossypii TEF1</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS2m2</td>
<td>A. gossypii TEF1</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS2-X3</td>
<td>A. gossypii TEF1</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS3</td>
<td>A. gossypii TEF1</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS3m</td>
<td>A. gossypii TEF1</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS3m2</td>
<td>A. gossypii TEF1</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS3-X4</td>
<td>A. gossypii TEF1</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS1U2</td>
<td>80pHm-lexA-SPO13</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pDS1U3</td>
<td>80pHm-lexA-SPO13</td>
<td>HIS3</td>
<td>-</td>
<td>CEN/ARS</td>
<td>This study</td>
</tr>
<tr>
<td>pRS414-empty</td>
<td>TRP1</td>
<td>LEU2</td>
<td>-</td>
<td>CEN/CEN</td>
<td>(Sikorski and Hieter, 1989)</td>
</tr>
<tr>
<td>pRS415-empty</td>
<td>TRP1</td>
<td>LEU2</td>
<td>-</td>
<td>CEN/CEN</td>
<td>(Sikorski and Hieter, 1989)</td>
</tr>
<tr>
<td>pRS416-empty</td>
<td>TRP1</td>
<td>LEU2</td>
<td>-</td>
<td>CEN/CEN</td>
<td>(Sikorski and Hieter, 1989)</td>
</tr>
</tbody>
</table>

All plasmids constructed in this study will be deposited at the Addgene repository.

plDS2-auRZ and plDS3-auRZ (Table 2) in a single transformation step as described in Section 2.5. After heat shock, the transformed cells were allowed to recover in YPD for 1 h at room temperature prior to removal of YPD and plating. The exact influence of recovery based on a CEN/ARS origin.

Table 3
Divisible selection plasmids developed to reconstruct the F. graminearum (Fg) rubrofusarin polyketide pathway in S. cerevisiae. CEN.PK2-1c-npgA based on uracil prototrophy and benchmark plasmids with the same pathway genes and classical auxotrophic selection genes. The respective pathway open reading frames (ORF) were expressed with the indicated promoter and terminators from S. cerevisiae. All plasmids propagate in S. cerevisiae based on a CEN/ARS origin.

<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Pathway cassette</th>
<th>Selective gene</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDS1U-PKS12</td>
<td>Pcyt1-fgPKS12-fgADH1</td>
<td>d1-Ura3</td>
<td>This study</td>
</tr>
<tr>
<td>pDS2-auRZ</td>
<td>Pcyt1-fgau1-Z-TEND2</td>
<td>d2-DBD</td>
<td>This study</td>
</tr>
<tr>
<td>pDS3-auRZ</td>
<td>Pcyt1-fgau1-J-VC1</td>
<td>d3-AD</td>
<td>This study</td>
</tr>
<tr>
<td>pRS416-PKS12</td>
<td>Pcyt1-fgPKS12-fgADH1</td>
<td>Trp1</td>
<td>(Rugbjerg et al., 2013)</td>
</tr>
<tr>
<td>pRS414-auRZ</td>
<td>Pcyt1-fgau1-Z-TEND2</td>
<td>LEU2</td>
<td>(Rugbjerg et al., 2013)</td>
</tr>
<tr>
<td>pRS413-npgA</td>
<td>Pcyt1-fgau1-J-VC1</td>
<td>HIS3</td>
<td>(Rugbjerg et al., 2013)</td>
</tr>
</tbody>
</table>

All plasmids constructed in this study will be deposited at the Addgene repository.

3. Results
3.1. Design of divisible selection

To develop effective divisible selection in S. cerevisiae, we identified five important system criteria. 1) Stringency: selection should be highly efficient for all associated parts and disallow false selection at incomplete uptake of all parts. 2) Modularity: the ability to divide a selection phenotype into multiple co-dependent entities should be independent of the actual selection phenotype i.e. also function when plugging in a different selection gene ORF. 3) Dominance: pre-adaptation of the host strain should not be necessary in order to use divisible selection instead of the corresponding single selection gene. 4) Stability: the system should use few endogenous parts from the host strain to minimize risks of recombination and crosstalk. 5) Scalability: the concept should be scalable towards wider division, such as split of TFs into additional components or simultaneous use of multiple, orthogonal TF networks.

To meet these criteria, we designed divisible selection as three individually expressed modules (Fig. 1) that we could link physically to three independent DNA fragments (plasmids). The modules consist of: DS1) A selection gene ORF driven by a promoter, tightly responsive to: DS2) A DNA-binding protein fused to a protein-interacting domain, cognate to: DS3) another protein-interacting domain fused to a transcription activation domain. The TF-based approach would ensure modularity since other selection gene ORFs can be swapped in and out. To achieve high stringency, we engineered a synthetic promoter based on the SP013 promoter, which features a mitotic UME6 repressor-binding site to suitably abolish leaky expression as previously utilized (Vidal et al., 1996). UME6 represses SP013 promoter activity in mitotic cells (Mitchell, 1994), thus rendering the TF response tighter. This is important since even very low expression of auxotrophic selection genes typically reconstitute prototrophy.
For expression of the hybrid protein, a heterologous URA3 terminator from E. coli was used to avoid interactions with endogenous repressors or promoters. Finally, we expressed the split TF modules from promoters not present in the S. cerevisiae genome to limit unintended recombination.

3.2. Detailed design composition

The divisible selection module ds1-URA3 was engineered with the classical URA3 gene under control of a 157 bp SPO13 promoter fragment (Fig. 2A). Upstream of the promoter fragment, we inserted four natural and synthetic DNA-binding sites specific for E. coli LexA at various in vitro-based Kd values down to 0.8 · 10⁻¹⁰ M (Zhang et al., 2010) to achieve high saturation as soon as few LexA proteins enter the cell nucleus. As terminator, the URA3 terminator from Klyveromyces lactis was used.

The ds2-DBD module consists of the E. coli LexA DNA-binding repressor in an N-terminal fusion to a nucleus localization sequence (NLS). Using an RSNQTSYYKAGSAAAPFT linker, LexA is C-terminally fused to a Krev-1 (Rap1A) protein, known for its sequence (NLS). Using an RSNQTSLYKKAGSAAAPFT linker, LexA is repressor in an N-terminal fusion to a nucleus localization sequence to interact with endogenous repressors or promoters. Finally, we expressed the split TF modules from promoters not present in the S. cerevisiae genome to limit unintended recombination.

3.3. Selection for three different plasmids activating a single selection phenotype

To characterize the performance of divisible selection, all three modules ds1, ds2 and ds3 were simultaneously transformed into the ura3-deficient S. cerevisiae CEN.PK2-1C, and an average of 220 transformants per 400 ng of each plasmid formed colonies after 3 days incubation on SC-ura (Fig. 3). Theoretically, false activation of the ds1 module in absence of ds2 (DBD) and/or ds3 (AD) has potential to cause false-positive selection. Thus, the vectors containing ds1 and ds3 were used as negative transformation control to specifically characterize possible transcriptional activity from the hybrid DBD protein and selection gene ORF. Following 3 days, an average of 4 colonies were observed after transformation with ds1-URA3 and ds2, whereas 12 colonies were observed with ds1-URA3 and ds3, indicating that the selection system was tight with a 94% occurrence of true-positive transformants (Fig. 3). The few false-positive colonies were generally smaller than the true-positive colonies. The efficiency was further compared to transformation of three plasmids (pRS414, pRS415 and pRS416) selected for using the classical auxotrophic selection genes LEU2, TRP1 and URA3 to benchmark the general co-transformation efficiency. These plasmids resulted in an average 3-fold higher colony count at 707 transformants per 400 ng of each three plasmids. The reason for this advance in efficiency was not apparent, but we suspected suboptimal transcriptional activation of ds1 to be responsible. To modify this, it would be important to not simultaneously increase the frequency of false-positive transformants and we concluded the resulted transformation efficiency to be satisfactory for most applications.

A few preliminary constructs of the ds1 module were also evaluated in slightly modified pRS41x-type vectors, testing fewer lexA binding sites and a heterologous K. lactis URA3 sequence instead of the S. cerevisiae URA3. Remarkably, while no difference in colony formation was seen at 3 days, a considerable number of false-positive transformants formed colonies when using fewer binding sites and the K. lactis URA3 following prolonged incubation at 30 °C for 6 days (Supplementary material).

3.4. Modular exchange of selection phenotype from Ura+ to His+

To test the modularity of divisible selection and its ability to expand other selectable traits, we exchanged the URA3 ORF in ds1 with the ORF of HIS3 from S. cerevisiae. Transformation with the three modules ds1-HIS3, ds2 and ds3 resulted in an average of 468 transformants after 3 days incubation. 17 transformants resulted from the control transformation with only ds1-HIS3 and ds3, relative to 15 colony-forming cells with ds1-HIS and ds2, indicating that the system robustly could shift to a different prototroph (Fig. 3). Thus, the simple exchange of the selection phenotype demonstrates the versatility of the concept, and indicates that more selectable traits could be modularly divided.

3.5. Divisible selection with copy number-fluctuating plasmids (2-micron)

Elevation of gene copy number can lead to increased formation of the metabolic products. To this end, the multi-copy 2-micron yeast plasmid is sometimes utilized, although its copy number varies considerably per cell (Da Silva and Srikrishnan, 2012; Jensen et al., 2013). To test how this plasmid type functions with the divisible selection system, we cloned ds1U, ds2 and ds3 into three plasmids with 2-micron origins. The plasmids were transformed simultaneously, while omitting the ds2 plasmid as a control for false-positive transformants.

Following three days of incubation, an average of 327 transformants (std. dev. = 85) were visible, relative to the false-positive control omitting the ds2 module, which reached 30 visible transformants (std. dev. = 4). However, on day four more transformants appeared on especially the false-positive control plates, and on day five this number reached a proportion of around 50% the number of true-positive transformants. This high occurrence of false-positive transformants could be a consequence of selecting for amplification of the ds1U plasmid copy number, allowed by the 2-micron origin over time. Such an effect is similar to plasmid amplification methods that utilize weakly expressed selection
3.6. Three simultaneous chromosomal integrations selected under one selection phenotype

Chromosomal integration of genes is frequently used in metabolic engineering due to the copy number stability offered, e.g. alleviating selection for maintenance. Thus, to test whether the utility of a selection phenotype also could be tripled for chromosomal integration, we cloned the three ds modules ds1U, ds2 and ds3 into three integration vectors from the Easyclone concept (Jensen et al., 2013). The Easyclone vectors target three distinct loci on S. cerevisiae chromosome X (Mikkelsen et al., 2012) through homologous recombination of 0.5 kb upstream and downstream flanking sequences (Fig. 4A). Such simultaneous integration at three distinct loci is rarely reported, which may also be due to the fact that a single cell should both take up three different DNA fragments and integrate each of them correctly. Since high transformation efficiency would be required to obtain three simultaneous integrations, a generous amount of DNA (3 μg of each ds module) was applied during transformation. Following three days' incubation, 20 transformants were isolated on SC plates lacking uracil. First, we assayed the selection stringency of the resulting clones. PCR on gDNA extracted from the transformants revealed the presence of ds1 and ds2 in all 20 colonies,
whereas ds3 could be confirmed in 19, leading to an overall estimated efficiency of 95% (Fig. 4B). Further, the single transformant lacking ds3 appeared to grow much slower when re-streaked to selective medium. The efficiency of chromosomal targeting in yeast not only depends on the performance of the selection mechanism, but may also depend on the specific integration constructs and target loci. Next, we tested this targeting efficiency by PCR using locus-specific primers. In 40% of our transformants, site-specific insertion of all three ds modules could be verified by PCR. Simultaneous gene targeting at three loci has previously been shown with 44% targeting efficiency using classical auxotrophic selection genes (Jensen et al., 2013), which thus agreed very well.

3.7. Proof-of-principle construction of fungal polyketide pathway

To prove the divisible selection principle for introduction of a multi-gene metabolic pathway, we chose a fungal secondary metabolic pathway as test case. Into each of the three pDS vectors, we inserted one of the three genes known to reconstruct the pathway leading from endogenous acetyl CoA and malonyl CoA to rubrofusarin PKS12, aurZ and aurJ (Rugbjerg et al., 2013). These genes encode respectively an iterative type I polyketide synthase, dehydrogenase and O-methyltransferase (Fig. 5A). Since the host yeast strain harbored a gene encoding a general polyketide synthase-activating phosphopantetheinyl transferase NpgA expressed from a fourth CEN/ARS plasmid (pRS413-npgA), successful gene introduction would result in heterologous rubrofusarin production. This tricyclic polyketide is visible as a distinct orange-brown pigment. Following transformation, six transformants formed colonies. These were evaluated for production by plating on an induction agar plate with 100 μM Cu^{2+}, and yellow pigmentation indicated the successful activity of the metabolic pathway as compared to a classically selected pathway strain and a negative control strain (Fig. 5B). Further, to verify introduction of all three plasmids, specific PCRs were performed on extracted DNA from three randomly chosen transformants and the resulting bands confirmed the presence of all plasmids (example Fig. 5C). Accumulation of the two pathway intermediates resulted in yellontoned though distinguishable pigmentation, for which reason PCR validation was particularly important as final proof of successful gene introduction.

Since five simultaneously propagated centromeric plasmids can cause considerable cytotoxicity (Futcher and Carbon, 1986), this was further evaluated in a strain with four empty CEN/ARS plasmids. Without selection, this strain did show loss of plasmids in long-term cultures, but normal cell morphology and exponential growth rate (Supplementary material). With selection, the stability of the four CEN/ARS plasmids carrying the rubrofusarin pathway was further tested in ten randomly picked colonies following a 30-generation liquid cultivation. PCR on unique plasmids was confirmative for all four plasmids in all 10 colonies, whereas maintained pathway function was indicated through visual inspection of cell pigmentation (Supplementary material).

4. Discussion

Divisible selection is a new modular concept for selection of several physically independent DNA fragments using the same selective trait. This yields increased strain construction freedom since multiple, independent DNA can be introduced utilizing only one selectable phenotype. Such divisible selection is especially relevant in strains with limited selectable traits. Due to the modularity of the concept, flexibility is added to the strain...
engineering process in several aspects. Specifically, we have shown that the particular phenotype can be modularly swapped by only exchanging the selection gene ORF of the system (ds1). Further, while all three modules are essential, they need not all carry a pathway gene. For example, it is possible to utilize only the ds2 and ds3 modules (containing the hybrid DBD and AD parts) for production genes and co-transform ds1 as phenotypic decider to control selection phenotype conveniently without sub-cloning. Such flexibility could become important in saving time needed to make pathway genes compatible with new host strains e.g. when testing performance in genetics-limited industrial strains or combining two individually tested segments of a long metabolic pathway.

ds1 is the critical component to the performance of the system by carrying the full selection gene ORF. As shown, there is a high probability that a single picked clone will be a correct transformant, since false-positive transformants on average occurred at 4–6% of the positively transformed clones. This low percentage could have been overestimated slightly from the assay method of omitting one of the split hybrid TF modules: due to the high co-transformation efficiency in S. cerevisiae, it is possible that a fraction of these false-positive cells would also take up additional plasmids if available. Whereas the plasmid copy number is maintained stably through cell divisions by the CEN sequence (Clarke and Carbon, 1985), it is possible that the few false-positive transformants may have arisen from uptake of multiple ds1 copies during transformation. In contrast the copy number variation of 2-micron plasmids led to an increased number of false positives. This episomal plasmid fluctuates considerably in copy number due to its propagation method (Mead et al., 1986) and thus also appears less ideal for metabolic engineering in some regards (Jensen et al., 2013). With targeted chromosomal integration, the introduction of a single copy of each ds module likely resulted in the high selection stringency observed and similar to that observed with the CEN/ARS plasmids.

The proof of principle by introduction of a three-gene fungal polyketide pathway demonstrates the potential of divisible selection to introduce metabolic pathways under a single selection phenotype. The polyketide pathway was chosen as a relatively large extreme in terms of size (total of 11 kb) and the size mainly resulted from the 7 kb polyketide synthase cassette. Since divisible selection depends on co-transformation of all DNA units in a single step, good transformation efficiency is important. For more extreme sizes, it is therefore possible that the use of simple lithium-acetate transformation methods will prove limiting, calling for methods suited for large DNA fragments. Ideally, future systems should improve the transformation colony formation rate compared to classical selection genes, e.g. through manipulation of the output promoter to increase responsiveness to the TF.

An alternative approach to transcriptional division of selection genes would be to split the individual ORFs encoding the selection genes to form individually unfunctional heteromultimers as demonstrated in E. coli (Schmidt et al., 2012). At a cost of modularity however, such approaches require that the splitting points of each individual selection gene are developed. For use in pathway construction, detailed characterization of false-positive ratios in transformation is also important to understand the risks for picking these. Still, split TF-based divisible selection could be combined with such approaches to yield more co-dependent selection units. Alternative to transcriptional splits, longer nutrient biosynthesis pathways could be utilized for co-dependent selection, e.g. by building up selection upon several uracil biosynthesis genes. The disadvantage of such approaches, however, would be less flexibility and the requirement for pre-adapting strains, which limits strain compatibility with existing selection procedures.

By depending on split TF reconstitution, it may further be possible to break down the different modules to allow additional independent DNA fragments to be selected under the same phenotype. For CEN/ARS-based propagation however, the limit for stable maintenance may lie below five plasmids per strain (Futcher and Carbon, 1986). This thus favors chromosomal integration for multiple gene introductions. Similarly constraining could be the co-transformation efficiency or number of possible TF heteromers. The split of phage T7 RNA polymerase into multiple units (Segall-Shapiro et al., 2014) is e.g. of relevance for future E. coli systems. High specificity of such DNA and protein interactions should enable orthogonal use of multiple co-selecting modules, i.e. through further co-dependent hybrid TFs, which independently control different selection gene ORFs. Such orthogonally divided selections could possibly extend the number of gene introductions possible before selection gene recycling.

New modular selection approaches that rethink selection are likely to significantly aid the speed of multi-gene evaluation. Introduction strategies such as the homologous recombination-
guided DNA assembler method (Shao et al., 2009) provides advantages in terms of speed, but reliance on homologous recombination requires the absence of interfering recombination-prone sequences. Recent gene editing methods relying on clustered regularly interspaced short palindromic repeats (CRISPRs) are also likely to help speed up cell factory development (DiCarlo et al., 2013; Jakočiūnas et al., 2015), especially if off-target effects or intellectual property concerns (Sherkow, 2015) are not significant or relevant. Selection schemes further have metabolic engineering applications within synthetic screens and other high-throughput assays (Dietrich et al., 2010).

Finally, divisible selection acts as a multi-component AND gate in the sense that all modules are required to transmit the signal (of DNA uptake in this case). Since selection is a powerful method for enhancing multi-component screening, it is possible that the general split hybrid TF network could take other applications, e.g. adapted to other synthetic signal-processing schemes requiring AND logics.

In conclusion, a divisible selection concept based on split hybrid TFs was developed to extend the single utility of selection genes to three. Such systems could be particularly well suited for utilizing the benefits of chromosomal gene integration.

Competing interests
The authors declare no competing interests.

Author contributions
PR and MOAS conceived the study. N.MP and PR performed the experiments. PR, NMP and MOAS wrote the manuscript.

Acknowledgments
The authors wish to thank Irina Borodina for providing the plasmids pCIB393, pCIB394 and pCIB395. The research leading to these results has received funding from the Novo Nordisk Foundation, Denmark and from the European Union Seventh Framework Programme (FP7-KBBe-2013-7-single-stage) under Grant agreement no. 613745. Promys. The funding sources had no involvement in the conduct of the research or the preparation of the article.

Appendix A: Supplementary Information
Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ymben.2015.08.004.

References


Supplementary material for

Flexible metabolic pathway construction using modular and divisible selection gene regulators

Authors:

Peter Rugbjerg, Nils Myling-Petersen, Morten O. A. Sommer

Table of Contents

S1 Construction of plasmids ........................................................................................................ 107
S2 Evaluation of alternative ds1 modules ................................................................................. 112
S3 Stability of four CEN6-propagated plasmids ...................................................................... 114
S4 Plasmid maps .................................................................................................................... 118
S1 References .......................................................................................................................... 119
S1 Construction of plasmids

Plasmids were all assembled from individual PCR fragments. Table S1 lists the PCR fragment composition (by primers and DNA template) leading to the indicated plasmid. Further detail on PCR templates is given in Table S2, while the respective oligonucleotide sequences are specified in Table S3.

Table S1 PCR fragments used to assemble the listed plasmids using the indicated primer pairs and DNA template. Sequences of linear DNA fragments used as template are given in text below, and overview of template plasmids is given in Table S2. All plasmids were assembled through uracil excision cloning with the exception of pDS2-aurZ.

<table>
<thead>
<tr>
<th>Plasmid ID</th>
<th>Primer pair</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDS1 U</td>
<td>P312/P313</td>
<td>pRS416</td>
</tr>
<tr>
<td></td>
<td>P273/P258</td>
<td>pUG72</td>
</tr>
<tr>
<td></td>
<td>P305/P56</td>
<td>S. cerevisiae gDNA</td>
</tr>
<tr>
<td></td>
<td>P55/P302</td>
<td>SPO13 hybrid DNA fragment</td>
</tr>
<tr>
<td></td>
<td>P255/ID399</td>
<td>pUG72</td>
</tr>
<tr>
<td>pDS1 H</td>
<td>P285/P55</td>
<td>pDS1 U</td>
</tr>
<tr>
<td></td>
<td>P205/P284</td>
<td>S. cerevisiae gDNA</td>
</tr>
<tr>
<td>pDS2</td>
<td>P313/P312</td>
<td>pRS416</td>
</tr>
<tr>
<td></td>
<td>P261/P273</td>
<td>pUG75</td>
</tr>
<tr>
<td></td>
<td>P192B/P192</td>
<td>E. coli gDNA</td>
</tr>
<tr>
<td></td>
<td>P193/P283</td>
<td>pEXP32</td>
</tr>
<tr>
<td></td>
<td>P260/ID399</td>
<td>pUG75</td>
</tr>
<tr>
<td>pDS3</td>
<td>P312/P313</td>
<td>pRS416</td>
</tr>
<tr>
<td></td>
<td>P261/P273</td>
<td>pUG75</td>
</tr>
<tr>
<td></td>
<td>P42/P112</td>
<td>VP16 AD DNA fragment</td>
</tr>
<tr>
<td></td>
<td>P202/P259</td>
<td>pEXP22</td>
</tr>
<tr>
<td></td>
<td>P260/ID399</td>
<td>pUG75</td>
</tr>
<tr>
<td>pS1.2</td>
<td>P55/P267</td>
<td>pDS1 U</td>
</tr>
<tr>
<td></td>
<td>P252/P273</td>
<td>pUG72</td>
</tr>
<tr>
<td></td>
<td>ID399/P255</td>
<td>pUG72</td>
</tr>
<tr>
<td></td>
<td>P271/P272</td>
<td>pRS416</td>
</tr>
<tr>
<td>pS1.3</td>
<td>P273/P302</td>
<td>pDS1 U</td>
</tr>
<tr>
<td></td>
<td>P255/P272</td>
<td>pS1.2</td>
</tr>
<tr>
<td>Vector</td>
<td>Forward Primer</td>
<td>Reverse Primer</td>
</tr>
<tr>
<td>--------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>pDS1U-PKS12</td>
<td>P273/ID399</td>
<td>P312/P313</td>
</tr>
<tr>
<td>pDS2-aurZ</td>
<td>P273/ID399</td>
<td>Smal-digested plasmid</td>
</tr>
<tr>
<td>Gibson assembly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pDS3-aurJ</td>
<td>P273/ID399</td>
<td>P312/P313</td>
</tr>
<tr>
<td>SmaI-digested plasmid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pDS1Um</td>
<td>P116/P119</td>
<td>P117/P118</td>
</tr>
<tr>
<td>pDS2m</td>
<td>P116/P119</td>
<td>P117/P118</td>
</tr>
<tr>
<td>pDS3m</td>
<td>P116/P119</td>
<td>P117/P118</td>
</tr>
<tr>
<td>pDS1U-X2</td>
<td>P468/ID399</td>
<td>P467/P469</td>
</tr>
<tr>
<td>pDS2-X3</td>
<td>P468/ID399</td>
<td>P467/P469</td>
</tr>
<tr>
<td>pDS3-X4</td>
<td>P468/ID399</td>
<td>P467/P469</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Features (not complete)</td>
<td>Reference</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>pDS1.2U</td>
<td>pUC origin, f1, ARS/CEN, ampR, 80P\textsubscript{PstI}-URA3</td>
<td>This study</td>
</tr>
<tr>
<td>pDS1.3U</td>
<td>pUC origin, f1, ARS/CEN, ampR, 40P\textsubscript{PstI}-KURA3</td>
<td>This study</td>
</tr>
<tr>
<td>pUG72</td>
<td>pUC origin, loxP\textsubscript{pKURA3}-KURA3-loxP\textsubscript{pKURA3}-ampR</td>
<td>(Hegemann and Heick, 2011)</td>
</tr>
<tr>
<td>pUG75</td>
<td>pUC origin, loxP\textsubscript{pKTEF1}-hphMX-ampR</td>
<td>(Hegemann and Heick, 2011)</td>
</tr>
<tr>
<td>pEXP22</td>
<td>RalGDS-GAL4AD, pUC origin, f1, ARS/CEN, TRP1, ampR</td>
<td>Life Technologies - Proquest</td>
</tr>
<tr>
<td>pEXP32</td>
<td>Krev1-GAL4DBD, pUC origin, f1, ARS/CEN, LEU2, ampR</td>
<td>Life Technologies - Proquest</td>
</tr>
<tr>
<td>pRS413</td>
<td>pUC origin, f1, ARS/CEN, HIS3, ampR</td>
<td>(Sikorski and Hieter, 1989)</td>
</tr>
<tr>
<td>pRS414</td>
<td>pUC origin, f1, ARS/CEN, TRP1, ampR</td>
<td>(Sikorski and Hieter, 1989)</td>
</tr>
<tr>
<td>pRS415</td>
<td>pUC origin, f1, ARS/CEN, LEU2, ampR</td>
<td>(Sikorski and Hieter, 1989)</td>
</tr>
<tr>
<td>pRS416</td>
<td>pUC origin, f1, ARS/CEN, URA3, ampR</td>
<td>(Sikorski and Hieter, 1989)</td>
</tr>
<tr>
<td>pRS416-PKS12</td>
<td>p\textsubscript{CLUP1}-PKS12-t\textsubscript{ADH1}, pUC origin, f1, ARS/CEN, URA3, ampR</td>
<td>(Rugbjerg et al., 2013)</td>
</tr>
<tr>
<td>pRS414-aurZ</td>
<td>p\textsubscript{TEF1}-aurZ-t\textsubscript{ADH2}, pUC origin, f1, ARS/CEN, TRP1, ampR</td>
<td>(Rugbjerg et al., 2013)</td>
</tr>
<tr>
<td>pRS415-aurJ</td>
<td>p\textsubscript{GPD1}-aurJ-t\textsubscript{CYC1}, pUC origin, f1, ARS/CEN, LEU2, ampR</td>
<td>(Rugbjerg et al., 2013)</td>
</tr>
<tr>
<td>pESC-HIS</td>
<td>2-micron, HIS3, ampR</td>
<td>Agilent Technologies</td>
</tr>
<tr>
<td>pCfB393</td>
<td>X-2 targeting flanks, pTEF-CFP-tADH1, ampR</td>
<td>(Jensen et al., 2013)</td>
</tr>
<tr>
<td>pCfB394</td>
<td>X-3 targeting flanks, pTEF-RFP-tADH1, ampR</td>
<td>(Jensen et al., 2013)</td>
</tr>
<tr>
<td>pCfB395</td>
<td>X-4 targeting flanks, pTEF-YFP-tADH1, ampR</td>
<td>(Jensen et al., 2013)</td>
</tr>
</tbody>
</table>
Nucleotide sequence of SPO13 hybrid DNA fragment:

>SP013 hybrid DNA fragment
TTCCGACCTGCACTGTATGTACATAACGCTACTGTATGTACATAAGCAGCATGCAAGCTC
TAGAGTCTCCGTCTTAGTCAGCTACCTTGGCAGGAAATGTTAATATTAGGAGTATATTGAGAAATA
GCCGCCGCACAAAAAGGAAAGTCCTCATAAAAGTGTCTAACAGACAATTAAGCGCAATAAAGAAAGAAAAC
GGATTGAAGTGTACGTGAGAAATAATT

Nucleotide sequence encoding VP16 AD, synthesized by Life Technologies GeneArt:

>VP16 AD DNA fragment
TCGACGGCCCGCCGACCGATGTCAGCCTGGGGGACGAGCTCCACTTAGACGCGAAGAGCTGCGATG
GCGCATGCCGACGCGCTAGACGATTTCGATCTGGACATGTTGGGGGACGGGGATTCCCCGGGGCCGGGA
TTTACCCCGACGACTCCGCCCTACGGCGCTCTGCTGATAATGCGCGACTTCGAGTTGAGCAGATGTTT
ACCGATGCCCTTGGAAATCGAGACTACGGTGGG

Table S3 Sequence of oligonucleotides used in the study

<table>
<thead>
<tr>
<th>Primer ID</th>
<th>Oligonucleotide sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P35</td>
<td>GACAGCTTCTCTGATCGGAAGG</td>
</tr>
<tr>
<td>P42</td>
<td>AACATAUGCCCAAGAAGACCGG</td>
</tr>
<tr>
<td>P55</td>
<td>AATTATTTCUCGACTCAACTTAAATC</td>
</tr>
<tr>
<td>P56</td>
<td>AGAAATAATGTGCGAAAGCTACATAAAGAAA</td>
</tr>
<tr>
<td>P111</td>
<td>AGCGGAAGGUCAGCTTTCGACGGGCCCCC</td>
</tr>
<tr>
<td>P112</td>
<td>ACCCACCGUACTCGTCAATTCC</td>
</tr>
<tr>
<td>P116</td>
<td>AGGTGGCACUTTTCGGG</td>
</tr>
<tr>
<td>P117</td>
<td>AGTGCACCGUGAAGCAGG</td>
</tr>
<tr>
<td>P118</td>
<td>ATGTTTTCTUTAGATGATCCAAATATC</td>
</tr>
<tr>
<td>P119</td>
<td>AGAAAACCAUTATTATCATGACATTAAC</td>
</tr>
<tr>
<td>P191</td>
<td>AGCGGAAGGUGUCCAGCATGAAGCGCTTAAAC</td>
</tr>
<tr>
<td>P192</td>
<td>AGCCAGUCGGCGGTTTGG</td>
</tr>
<tr>
<td>P192B</td>
<td>AACATAUGCCCAAGAAGACCGGAGGCTGAGCATGAAAGCGCTTAAACGGCCAG</td>
</tr>
<tr>
<td>P193</td>
<td>ACTGGCGUGGATCGCAATCAACCAAGGTTTGT</td>
</tr>
<tr>
<td>P202</td>
<td>ATGTGCAUGCAAACAAAGGTTGTA</td>
</tr>
<tr>
<td>P205</td>
<td>AGAATAATUTAGCAGACAGACGAAAAGCC</td>
</tr>
<tr>
<td>P230</td>
<td>ACCCCTTCUCCGCTCTTTTGGAGAAAGATCATAGTTGGA</td>
</tr>
<tr>
<td>P252</td>
<td>ATATAGUCACAAATGCATATACCA</td>
</tr>
<tr>
<td>P255</td>
<td>AGAGCUUTTTCGAGAAACC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>P258</strong></td>
<td>AGCTTAUACAGGAAACTTAATAGAACAAATC</td>
</tr>
<tr>
<td><strong>P259</strong></td>
<td>ACCGCTUCTTTTAGGATG</td>
</tr>
<tr>
<td><strong>P260</strong></td>
<td>AAGCGGUAACCTCGAGGACAATAAAAAAG</td>
</tr>
<tr>
<td><strong>P261</strong></td>
<td>ATATGTUCTTTTATAGGATATGTG</td>
</tr>
<tr>
<td><strong>P267</strong></td>
<td>AGCTCGUTCCGACCTCGACTCTC</td>
</tr>
<tr>
<td><strong>P271</strong></td>
<td>ACCCAAUGCATCGAGAAATTGTAAACG</td>
</tr>
<tr>
<td><strong>P272</strong></td>
<td>AAGCTGUUGGTATGGTGCA</td>
</tr>
<tr>
<td><strong>P273</strong></td>
<td>ACAGCTUTCGACAACCTTTATATAACTTC</td>
</tr>
<tr>
<td><strong>P283</strong></td>
<td>ACCGCTUCTAGAGCAGACATGATTTTC</td>
</tr>
<tr>
<td><strong>P284</strong></td>
<td>ATTACAUAAAGAACACCTTTGGTG</td>
</tr>
<tr>
<td><strong>P285</strong></td>
<td>ATGTAUATATACGAAACCTTTATAGAACAATC</td>
</tr>
<tr>
<td><strong>P302</strong></td>
<td>AGCTCGUCTGTATATATACAGATCTGTATATATACACCTCCGACCTGACGTACCTG</td>
</tr>
<tr>
<td><strong>P377</strong></td>
<td>ACTGAGAUAACCTACAGCGTGGAGCTAT</td>
</tr>
<tr>
<td><strong>ID399</strong></td>
<td>ATTGGGUGCATAGGCCACTAGTGGATCTG</td>
</tr>
<tr>
<td><strong>P467</strong></td>
<td>ACGCAGAUCTTGCAGCGTCCCCAAAACC</td>
</tr>
<tr>
<td><strong>P468</strong></td>
<td>ATCGCGUCATAACAGCCTTTCGACAAACCC</td>
</tr>
<tr>
<td><strong>P469</strong></td>
<td>ACCCAAUTCGCCTATAGGAGTGCG</td>
</tr>
</tbody>
</table>
S2 Evaluation of alternative ds1 modules

To investigate a possible advantage of utilizing a heterologous nucleotide sequence of 
**URA3**, we designed a mutant ds1 module with a heterologous selection gene ORF the 
*K. lactis URA3* (ds1.3-KIURA3), which has limited identity to potential host- 
inactivated **URA3** genes in *S. cerevisiae* laboratory strains.

At prolonged incubation times following transformation (30 °C for 6 days), this 
plasmid alone however permitted formation of small false-positive colonies (Fig S1).

We therefore tested the *S. cerevisiae URA3* in a new construct to limit these false-
positive colonies and possibly improve true positives by simultaneously introducing 
extra high-affinity LexA binding sites. Together with ds2 and ds3, the ds1.2 module 
(pDS1.2U) produced an average of 126 transformants following three days incubation 
and alone 2 false-positive transformants on average, while the second ds1.3 module 
(pDS1.3U) with two additional LexA binding sites and *S. cerevisiae URA3* generated 
the same average of false- and true-positive colonies (Fig. S1). Interestingly however, 
the use of **URA3** from *S. cerevisiae* did not lead to elevated false-positive colony 
formation at prolonged incubation times. This indicated a possible lower background 
expression or activity of this URA3 enzyme and also that the few false-positive 
colonies did not result from recombination at the inactivated **ura3** locus of the 
transformed CEN.PK2-1C strain.
Figure S1. Characterization of $ds1$ modules at prolonged incubation. Transformation efficiencies (average CFU per 400 ng of each plasmid) in *S. cerevisiae* CEN.PK2-1C transformed with incomplete or complete divisible selection modules ($ds1$, $ds2$ and $ds3$) under selection on SC–ura. CFUs counted following three or six days at 30 °C. Black points show values from individual replicate transformations (n = 3).
The viability of strains propagating four CEN6/ARS plasmids in a *S. cerevisiae* CEN.PK2-1C strain was assayed in comparison to strains transformed with respectively zero and one CEN6/ARS plasmid (Table S4).

To test if a cost of propagating the number of plasmids was conferred to the strains, their growth rates were first measured. In order to avoid potential influences of the different prototrophic genes on growth rate (Pronk, 2002), which would not be a measure of CEN6/ARS plasmid toxicity, the evaluation of the growth rates was performed in non-selective YPD medium, which would also allow loss of plasmids. Empty plasmids were used to exclude fitness costs associated with expressing various numbers of pathway enzymes. Pre-cultures of the strains grown in selective SC medium were used to inoculate 180 µL microtiter cultures in YPD medium, at 30 °C, 300 rpm horizontal shaking (New Brunswick Innova 44R). Their growth rates were calculated by exponential regression of OD600 values versus time obtained in the same range of the exponential growth phase and including at least seven data points. The resulting similar growth rates in non-selective YPD medium (Table S4) indicated that four different CEN6/ARS plasmids did not constitute a significant growth disadvantage in the CEN.PK2-1C strain without selection for plasmid maintenance. Loss of plasmids over time would however occur (Clarke and Carbon, 1985).

<table>
<thead>
<tr>
<th>Number of different CEN6/ARS plasmids</th>
<th>Growth rate (hr⁻¹) +/- standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.27 +/- 0.0032</td>
</tr>
<tr>
<td>1</td>
<td>0.26 +/- 0.0080</td>
</tr>
<tr>
<td>4</td>
<td>0.27 +/- 0.0061</td>
</tr>
<tr>
<td>4</td>
<td>0.26 +/- 0.0095</td>
</tr>
</tbody>
</table>

The degree of plasmid loss was subsequently indicated through plating of serially diluted spots of a four CEN6/ARS plasmid strain on various, selective SC media (SC, SC -histidine, SC –uracil -histidine) from non-selective YPD cultures grown at 30 °C, for respectively 20 and 30 generations through passing of 15 mL cultures to fresh
medium (Figure S2). Whereas maintenance of only one CEN6/ARS plasmid appeared stable without selection (plated on SC – histidine), a significant number of cells had lost at least one of the four plasmids, as seen when selecting for all simultaneously (SC – histidine, -uracil) (Figure S2). This implies that selection for four plasmids constitutes a negative factor on the growth rate of strains harboring four CEN6/ARS plasmids (overexpression of their respective selection genes could be another factor). In expanding divisible selection concepts beyond three centromeric plasmids, these effects would be important to characterize in depth.

<table>
<thead>
<tr>
<th>Plate medium</th>
<th>10-fold serial diluted spot assays</th>
<th>Generations cultured without selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC no selection</td>
<td><img src="image1" alt="Image" /></td>
<td>20</td>
</tr>
<tr>
<td>SC - histidine</td>
<td><img src="image2" alt="Image" /></td>
<td>30</td>
</tr>
<tr>
<td>Selection for 1 plasmid: pRS413</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>SC - histidine, -uracil</td>
<td><img src="image3" alt="Image" /></td>
<td>20</td>
</tr>
<tr>
<td>Selection for 4 plasmids: pRS413, pDS1U, pDS2, pDS3</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

**Figure S2 Loss of plasmids following long-term incubation in non-selective medium.** 10-fold serial dilution spot assays of *S. cerevisiae* CEN.PK2-1C transformed with four plasmids, on the indicated plates selecting for the number of CEN6/ARS plasmids shown to evaluate the number of colony-forming cells following incubation for the indicated number of generations in non-selective YPD medium. Plates incubated at 30 °C for 4 days.

The cell morphology in cultures grown in selective SC medium at 30 °C was finally inspected by microscopy to assess the health of the four-plasmid cells. Visual comparison showed no clear morphological differences nor long filament-like cell structures (Figure S3) as observed in strains with five CEN3/ARS plasmids in the *S. cerevisiae* BF307-10 strain grown in selective YNB medium (Futcher and Carbon, 1986). The size of the cells containing four CEN6/ARS plasmids could however appear slightly larger than of those containing zero.
Figure S3 Cell morphology. Exponentially growing S. cerevisiae cells in selective SC medium transformed with respectively A) zero CEN6/ARS plasmids and B) four different CEN6/ARS plasmids.

To test the ability to maintain production with a biosynthetic pathway, cultures of divisibly selected rubrofusarin strains (ds-rub) were grown in selective SC medium (-uracil, -histidine) for 30 generations and plated on selective plates. The stability of each of the four CEN6/ARS plasmids was then tested in ten randomly picked colonies through PCR of unique elements (Table S5). Specific PCR products from all four plasmids could be detected in all ten colonies. The corresponding ability to maintain production was also indicated by their pigmentation through re-streak of the same ten colonies to plates with induction of the pathway (Figure S4). The yellow nuance of the rubrofusarin pathway precursors (YWA1 and nor-rubrofusarin) however means that the PCR-based evaluation is important to verify maintenance of all plasmids.

Table S5 PCR confirmation of plasmid maintenance in ten colonies of rubrofusarin-pathway S. cerevisiae following 30-generation liquid cultivation. For each plasmid, the oligonucleotide pairs used in PCR are shown with their results.

<table>
<thead>
<tr>
<th>Plasmid to be verified</th>
<th>Oligonucleotides used</th>
<th>Colonies with plasmid confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRS413-npgA</td>
<td>P35 + P377</td>
<td>10/10</td>
</tr>
<tr>
<td>pDS1U-PKS12</td>
<td>P55 + 1D399</td>
<td>10/10</td>
</tr>
<tr>
<td>pDS2-auroZ</td>
<td>P191 + P283</td>
<td>10/10</td>
</tr>
<tr>
<td>pDS3-auroJ</td>
<td>P111 + P230</td>
<td>10/10</td>
</tr>
</tbody>
</table>
Figure S4 Confirmation of pathway product accumulation of individual colonies picked following long-term culture. Individual colonies streak to SC–uracil, -histidine plates with 100 µM Cu²⁺, incubated for 4 days at 30 °C. Photograph brightness was increased 20 percent.
S4 Plasmid maps

pDS1U (X-2)
6,969bp

pDS2 (X-3)
7,680bp

pDS3 (X-4)
6,938bp

p. 118. Chapter 4
SI References


Sikorski, R.S., Hieter, P., 1989. A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in *Saccharomyces cerevisiae*. 
Recombination-stable multimeric green fluorescent protein for characterization of weak promoter outputs in *Saccharomyces cerevisiae*

Authors:

Peter Rugbjerg - petru@biosustain.dtu.dk

Christoph Knuf - chrik@biosustain.dtu.dk

Jochen Förster - jfor@biosustain.dtu.dk

Morten O. A. Sommer - msom@bio.dtu.dk

a) Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark Kogle Allé 6, DK-2970 Hørsholm, Denmark

* Corresponding author
RESEARCH ARTICLE

Recombination-stable multimeric green fluorescent protein for characterization of weak promoter outputs in Saccharomyces cerevisiae

Peter Rugbjerg, Christoph Knuf, Jochen Förster and Morten O. A. Sommer*

Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, DK-2970 Hørsholm, Denmark

*Corresponding author: Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, DK-2970 Hørsholm, Denmark. Tel: +45 2151 8340; E-mail: msom@bio.dtu.dk

One sentence summary: By fusing three different green fluorescent proteins, the molecular brightness becomes high enough to detect low copy numbers inside a yeast cell in a stable manner.

Editor: Pascale Daran-Lapujade

ABSTRACT

Green fluorescent proteins (GFPs) are widely used for visualization of proteins to track localization and expression dynamics. However, phenotypically important processes can operate at too low expression levels for routine detection, i.e. be overshadowed by autofluorescence noise. While GFP functions well in translational fusions, the use of tandem GFPs to amplify fluorescence signals is currently avoided in Saccharomyces cerevisiae and many other microorganisms due to the risk of loop-out by direct-repeat recombination. We increased GFP fluorescence by translationally fusing three different GFP variants, yeast-enhanced GFP, GFP+ and superfolder GFP to yield a sequence-diverged triple GFP molecule 3vGFP with 74–84% internal repeat identity. Unlike a single GFP, the brightness of 3vGFP allowed characterization of a weak promoter in S. cerevisiae. Utilizing 3vGFP, we further engineered a less leaky Cu2+-inducible promoter based on CUP1. The basal expression level of the new promoter was approximately 61% below the wild-type CUP1 promoter, thus expanding the absolute range of Cu2+-based gene control. The stability of 3vGFP towards direct-repeat recombination was assayed in S. cerevisiae cultured for 25 generations under strong and slightly toxic expression after which only limited reduction in fluorescence was detectable. Such non-recombinogenic GFPs can help quantify intracellular responses operating a low copy number in recombination-prone organisms.

Keywords: signal amplification; synthetic biology; promoter engineering; protein multimerization

INTRODUCTION

Green fluorescent protein (GFP) is an invaluable tool for real-time visualization of intracellular proteins. Since the initial cloning, numerous improvements, variants and applications have been developed (Snapp 2009; Miyawaki 2011). GFP is particularly useful for quantification of intracellular events, localizations and populations at single-cell resolution. However, a minimal expression level is required such that the fluorescent output exceeds the cell autofluorescence and produces detectable signals. Still, biologically important processes occur through the interaction of a few molecules per cell, which is hard to quantify using existing fluorescent proteins and non-specialized experimental setups (Raj and van Oudenaarden 2009; Li and Xie 2011; Gahlmann and Moerner 2014). Further, the engineering of synthetic cell functionalities can depend on fine characterization and balancing of low gene expression levels (Ajikumar et al. 2010; Harton, Wingler and Cornish 2013).
The strategies for improving fluorescent output signals include the design of new GFP variants such as GFP+, yeast-enhanced GFP (yEGFP) and superfolder GFP (sfGFP) (Cormack, Bertram and Egerton 1997; Scholz et al. 2000; Pédelaq et al. 2006). Still, monitoring of single-molecule events such as chromosome movements in Escherichia coli has e.g. required multimerization of 96 DNA-binding sites to localize enough fluorescent protein to produce a distinguishable signal (Xie et al. 2008). Artificial tethering of a bright yellow fluorescent protein (Venus YFP) to the inside of E. coli cell membrane allowed a microscope-detectable signal from a single YFP-tagged protein (Yu et al. 2006). Thus, without techniques for single-molecule GFP sensitivity, the full-genome mapping of subcellular protein localization in Saccharomyces cerevisiae (yeastGFP) did not produce signals above background for 361 open reading frames (8 pct. of total) otherwise shown to be expressed in the growth phase assayed (Ghaemmaghami et al. 2003; Huh et al. 2003). Equivalently, the issue of not detecting all low-expressing S. cerevisiae proteins was also observed when the GFP library was applied to flow cytometry (Newman et al. 2006).

In some contexts, simple overexpression may shed light over the lacking information, but since the location of many proteins is a result of interactions with other cell components, a radical change in copy number could easily result in artificial observations. In other situations, the target output is the activity of specific weak promoters, e.g. in synthetic biological circuits, fluorescence-coupled biosensors or when developing promoter libraries. Several technologies permit the engineering of new promoters, e.g. responsive to other inducer molecules by hybridizing with upstream TF-binding sites (Blazeck and Alper 2013) or tuned to match fine, desirable transcription levels through mutagenesis of a strong native promoter (Nevoigt et al. 2006). Difficulties in GFP detection may have been a limitation in these developments for weaker promoter levels, though low expression may be phenotypically important for a wide range of synthetic biology purposes. In synthetic circuit designs, any concealed information on the shape of dose-response curves inhibits the analysis of mechanistic clues otherwise given by the response curvature (Ang, Harris and Hussey 2013). In applications of metabolite biosensors, background-covered signal levels means that the full regulatory capability cannot be utilized, e.g. limiting subsequent fluorescence-activated cell sorting (FACS). Ultimately, such autofluorescence could conceal properly functional GFP completely (Billinton and Knight 2001).

The efforts aimed at reducing the autofluorescence target two phenomena: simple medium autofluorescence arises from measuring fluorescence without isolating cells from medium, e.g. in continuously growing cultures. These effects can be reduced by the choice of medium or spectral unmixing by correcting for autofluorescence from a wavelength representing effects of the culture medium (Lichten et al. 2014). However, the cell autofluorescence is a more central issue, i.e. resulting from the fluorescence of flavins and NAD(P)H (Billinton and Knight 2001). Cellular autofluorescence also impacts techniques such as flow cytometry and microscopy and the weak signal intensity must be amplified intrinsically to the cell.

Previous studies in mammalian cell lines have tackled the obstacle of cell autofluorescence using directly repeated GFPs typically fused three to six times in tandem using a small translational linker (Genové, Glick and Barth 2005). By such approaches, it has been possible to achieve good linear increments in fluorescence signals. However, tandem repeats are problematic in organisms with proficient homologous recombination such as E. coli or S. cerevisiae where recombination between DNA can happen within windows of identity at around 25 nucleotides (Ahn et al. 1988). This could explain why tandem GFP methods are avoided in these organisms. However, even slight sequence divergence between repeats substantially decreases the rate of recombination as seen in the case of recombination between 350 bp inverted repeats, which was 4600-fold reduced when sequence identity was reduced from 100% to 74% in S. cerevisiae (Datta et al. 1997). Similar effects occur in E. coli where up to 1000-fold reduction was observed following a reduction in repeat identity to 80% (Rayssiguier, Thaler and Radman 1989).

Thus, in this study we present a simple methodology to take advantage of the ability to add sequence divergence to tandem proteins while maintaining function through variation in amino acid sequence as well as synonymous codon usage. By fusing three different GFP variants that vary mainly at nucleotide level, we produce a new triple tandem GFP (3vGFP) stabilized towards direct-repeat recombination. We demonstrate the utility of 3vGFP through a genetically triggered promoter (ON/OFF) and developing and characterizing a new version of a Cu²⁺-responsive promoter with reduced leakiness. Application of 3vGFP allowed visualization of weak signals that could not be separated from autofluorescence levels using the brightest individual GFP variant, superfolder GFP. Lastly, we test the stability towards recombination after culturing of the strain harboring 3vGFP through 25 generations.

**MATERIALS AND METHODS**

**Materials**

Unless otherwise stated, reagents were purchased from Sigma-Aldrich. Synthetic complete (SC) medium was prepared from 1.4 g L⁻¹ SC drop-out mix lacking uracil, tryptophan, leucine and histidine (Y2001), 6.7 g L⁻¹ yeast nitrogen base without amino acids (Y0626) and 20 g L⁻¹ D-glucose, pH standardized to 5.6. When SC was supplemented with additional amino acids, 60 mg L⁻¹ leucine, 20 mg L⁻¹ uracil, 20 mg L⁻¹ histidine-HCl and 20 mg L⁻¹ tryptophan was added. Yeast Peptone Dextrose medium contained 20 g L⁻¹ D-glucose.

Oligonucleotides were purchased from Integrated DNA Technologies.

**Plasmids**

The plasmids employed in this study are listed in Table 1.

**Strains**

The strains analyzed in this study are listed in Table 2. The following background strains were used to construct the strains:

Saccharomyces cerevisiae MaV203 (MATα, leu2-3,112, trp1-901, his3Δ200, ade2-101, gal4Δ, gal80Δ, SPAL10::URA3, GAL1::lacZ, HIS3/ΔYS2, can1Δ, cyh2Δ) (Purchased from Life Technologies).

Saccharomyces cerevisiae PrA18 (MATα, leu2-3,112, trp1-901, his3Δ200, ade2-101, gal4Δ, gal80Δ, GAL1::lacZ, can1Δ, cyh2Δ). Derived from S. cerevisiae MaV203.

Saccharomyces cerevisiae PrA26: MATα, leu2-3,112, trp1-901, his3Δ200, ade2-101, gal4Δ, gal80Δ, GAL1::lacZ, rad16::KanMX, can1Δ, cyh2Δ. Derived from S. cerevisiae PrA18.

Saccharomyces cerevisiae CIB1010 (MATa; ura3-52; his3Δ1; leu2-3/112; MAL2-8Δ; SUC2; ade2ΔloxP-KanMX; X-3::HMG1-PER1- FAD1::AtATR2). Derived from S. cerevisiae CEN.PK 102-5B.
Table 1. Plasmids employed in this study, describing whether they lead to chromosomal integration or propagate autonomously in S. cerevisiae.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Expression cassette (promoter-ORF-terminator)</th>
<th>Maintenance in S. cerevisiae through</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pPR4-3vGFP</td>
<td>pSPAL10-3vGFP-URA3</td>
<td>CEN/ARS, HIS3</td>
<td>This study</td>
</tr>
<tr>
<td>pPR4-sfGFP</td>
<td>pSPAL10-sfGFP-URA3</td>
<td>CEN/ARS, HIS3</td>
<td>This study</td>
</tr>
<tr>
<td>pCU2-3vGFP</td>
<td>pCUP1dim-3vGFP-URA3</td>
<td>CEN/ARS, URA3</td>
<td>This study</td>
</tr>
<tr>
<td>pCIB258-CUP1-3vGFP</td>
<td>pCUP1-3vGFP-tCYC1</td>
<td>Chromosomal integration</td>
<td>This study</td>
</tr>
<tr>
<td>pCIB258-CUP1-SP015-3vGFP</td>
<td>pCUP1dim-3vGFP-tCYC1</td>
<td>Chromosomal integration</td>
<td>This study</td>
</tr>
<tr>
<td>pDS1U-X2-3vGFP</td>
<td>pTEF1-3vGFP</td>
<td>Chromosomal integration</td>
<td>This study</td>
</tr>
<tr>
<td>pEXP22</td>
<td>pADH1-GAL4AD-RalGDS-TADH1</td>
<td>TRP1</td>
<td>Life Technologies</td>
</tr>
<tr>
<td>pEXP32</td>
<td>pADH1-GAL4AD8D-Krev1-TADH1</td>
<td>LEU2</td>
<td>Life Technologies</td>
</tr>
<tr>
<td>pRS413</td>
<td>–</td>
<td>HIS3</td>
<td>(Sikorski and Hieter 1989)</td>
</tr>
<tr>
<td>pRS415</td>
<td>–</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Saccharomyces cerevisiae strains analyzed in this study, indicating which plasmids or chromosomal integrations were introduced into the respective parental strains.

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Promoter</th>
<th>GFP</th>
<th>Plasmid #1</th>
<th>Plasmid #2</th>
<th>Plasmid #3</th>
<th>Integrative plasmid</th>
<th>Parent strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRa106</td>
<td>ON</td>
<td>3vGFP</td>
<td>pPR4-3vGFP</td>
<td>pEXP32</td>
<td>pEXP22</td>
<td>–</td>
<td>PRa26</td>
</tr>
<tr>
<td>PRa107</td>
<td>OFF</td>
<td>3vGFP</td>
<td>pPR4-3vGFP</td>
<td>pRS415</td>
<td>pEXP22</td>
<td>–</td>
<td>PRa26</td>
</tr>
<tr>
<td>PRa108</td>
<td>–</td>
<td>–</td>
<td>pRS413</td>
<td>pRS415</td>
<td>pEXP22</td>
<td>–</td>
<td>PRa26</td>
</tr>
<tr>
<td>PRa109</td>
<td>ON</td>
<td>sfGFP</td>
<td>pPR4-sfGFP</td>
<td>pEXP32</td>
<td>pEXP22</td>
<td>–</td>
<td>PRa26</td>
</tr>
<tr>
<td>PRa110</td>
<td>OFF</td>
<td>sfGFP</td>
<td>pPR4-sfGFP</td>
<td>pRS415</td>
<td>pEXP22</td>
<td>–</td>
<td>PRa26</td>
</tr>
<tr>
<td>CK24</td>
<td>pCUP1</td>
<td>3vGFP</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>pCFB258-CUP1-3vGFP</td>
<td>CB1010</td>
</tr>
<tr>
<td>CK28</td>
<td>pCUP1dim</td>
<td>3vGFP</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>pCFB258-CUP1-SP013-3vGFP</td>
<td>CB1010</td>
</tr>
<tr>
<td>PRa114</td>
<td>pTEF1</td>
<td>3vGFP</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>pDS1U-X2-3vGFP</td>
<td>PRa18</td>
</tr>
</tbody>
</table>

Construction of 3vGFP plasmids

Plasmids were constructed by uracil excision (USER) cloning. The general method for USER cloning was based on agarose gel purification of the PCR products amplified using DNA polymerase X7 (Nørholm 2010). These were mixed in an equimolar 20 μL reaction with 0.5 μL USER enzyme (New England Biolabs) and 0.5 μL DpnI FastDigest (Thermo Scientific) in FastDigest buffer at 37°C for 1-2 h. Following 25 min at room temperature, 2.5 μL reaction was transformed into E. coli. Correctly cloned plasmids were identified using restriction analysis and DNA sequencing. The detailed use of oligonucleotides for assembly of all plasmids is described in Supplementary data.

Construction of strains

Plasmids and DNA for chromosomal targeting was introduced in S. cerevisiae by methods described previously (Gietz and Schiestl 2007). The PRa18 strain was constructed from the MaV203 strain by deletion of SPAL10::URA3 through replacement with a kanMX gene deletion cassette flanked by loxP recombination sites from the pUG6 plasmid as described before (Güldener et al. 1996). DNA flanks to direct homologous recombination of the cassette to the chromosomal locus were generated by PCR on S. cerevisiae MaV203 gDNA spanning a fragment from 5′-CCATTCAACTAAGTACACC to 5′-CTTCCGATAAAATAGGCC (upstream flank) and from 5′-TTCAGAAATAGGTCC to 5′-CCATTCAACTAAGTACACC (downstream flank). These flanks were cloned to the kanMX gene deletion cassette and transformed into yeast. The kanMX cassette was looped out by heterologous expression of Cre recombinase from the pSH47 plasmid (Güldener et al. 1996). To construct PRa26 subsequently, the chromosomal HIS3 gene within the rad16 locus was deleted using the same kanMX approach. The targeting flanks spanned regions from 5′-AGTTGTTGAGATTTAGAACG to 5′-AAACCATAGGGATACCGGAGAAG (upstream flank) and 5′-TGACATCCGGGAAAGAACG to 5′-GATTATGTGTTACCATGTCGA (downstream flank).

To construct PRa18, the pTEF1-3vGFP construct was chromosomally integrated into the PRa18 strain using divisible selection (Rugbjerg, Myling-Petersen and Sommer 2015). DNA fragments for integration were liberated from the vector pDS1U-X2-3vGFP by digestion with Smal and transformed into yeast along with empty divisible selection plasmids pDS2 and pDS3 in order to reconstitute the selectable Ura gal phenotype.

To construct respectively CK24 and CK28 from the CB1010 strain, the pCUP1-3vGFP and pCUP1dim-3vGFP were chromosomally integrated by cloning into the EasyClone integrative vectors (Jensen et al. 2014). The DNA fragments for integration were obtained through NotI digestion of the vectors pCFB258-CUP1-3vGFP and pCFS258-CUP1-SP013-3vGFP, respectively, followed by agarose gel purification.

Estimation of TEF1-3vGFP fitness cost

Microtiter cultures of 200 μL YPD were inoculated by 100x back-dilution of overnight YPD pre-cultures of PRa114 and PRa108, each inoculated from single colonies. The cultures were cultivated in a 96-well plate at 30°C and continuous shaking in an ELx808 plate reader (BioTek), set to measure optical density every 15 min at OD630. The plate was covered with a BreathSeal...
(Greiner Bio-one) and plastic lid. Growth rates were calculated for all three biological replicates by exponential regression between OD<sub>630</sub> and time (hours) during the same OD<sub>630</sub> span of exponential growth phase. All OD<sub>630</sub> values were initially standardized to the time zero reading to account for differences in seal absorbance.

**Cultivations for stability tests**

The PRa114 strain was cultured from a single colony inoculated in 25 mL YPD medium and cultured at 30°C and 250 rpm horizontal shaking in three parallel lineages. By measuring OD<sub>600</sub>, the number of generations passed was calculated. Approximately 2% of the culture was passed to fresh medium and grown again until total 25 generations had passed. For comparison between cultured population and reference strain, approximately 25 μL of each cell population was inoculated in YPD medium at the same time and cultured at 30°C for 16 h with 250 rpm horizontal shaking.

**Fluorescence measurements**

Pre-cultures in selective SC medium were inoculated from single colonies and cultures overnight at 30°C. From these, 200 μL microtiter cultures of selective SC medium were inoculated and cultured at 30°C with 300 rpm horizontal shaking in an Innova shaking incubator for 16 h. As cover, the microtiter plates were covered with a BreathSeal (Greiner Bio-one) and a plastic lid.

The cell cultures were diluted approximately 1:100 in FACS flow buffer (BD Biosciences) and analyzed on a LSR Fortessa flow cytometer (BD Biosciences) equipped with a blue laser (488 nm) and set to measure 10,000 cells within a gate defined by forward and side scatter to capture all yeast cells. A FITC filter (530/30 nm) was used to measure GFP fluorescence reporting the area of the measured peaks. The laser voltage was adjusted to optimally utilize the dynamic range of detection. Data were processed and visualized as histograms with FlowJo version 10 (default settings) by overlaying the populations for each particular comparison.

**Sequence alignment**

Simple nucleotide and protein sequence alignment was performed using the ClustalO algorithm (Sievers et al. 2011).

**RESULTS AND DISCUSSION**

**Amplification of fluorescence by tandems of differently encoded GFPs**

To amplify the fluorescence signal of a GFP molecule while keeping transcription strength constant, the new 3vGFP protein was

![Figure 1](https://example.com/figure1.png)

*Figure 1.* Increased GFP fluorescence signal above autofluorescence level by triple tandem GFP (3vGFP). (A) Internal organization of individual GFP molecules fused as 3vGFP. 3vGFP consists of yeast-enhanced GFP (yEGFP), GFP+ and superfolder GFP. (B) The *S. cerevisiae* strains carrying 3vGFP allowed the capture of the weak, ON/OFF promoter pSPAL10 unlike strains carrying a single sfGFP. The ON levels with single sfGFP corresponded to the background level of the empty control strain without GFP. The strains are described in detail in Table 2. Error bars depict standard error from biological replicates (*n* = 3).
engineered by fusion of nucleotide sequences encoding yEGFP, GFP+ and sfGFP (Cormack, Bertram and Egerton 1997; Pédelecq et al. 2006) (Fig. 1A). Two glycine residues were introduced as translational linker in each junction. The fluorescence of 3vGFP was evaluated when expressed from a weak S. cerevisiae hybrid promoter (pSPLA10) (Vidal et al. 1996) based on pSPO13 to mimic low-expression applications (Huang and Schreiber 1997; Horton, Wingler and Cornish 2013). The low-level strength of pSPLA10 is attained by utilizing the UME6 repressor-binding site naturally present within the SPO13 promoter, which allows very low expression levels e.g. useful for control of cell growth. Further, GAL4-binding sites fused 179 bp upstream of start codon provide an upstream activating sequence (UAS), allowing transcription factor-based ON/OFF inputs.

The output fluorescence was first evaluated with single sfGFP (Fig. 1B), which is the individually brightest of the three GFPs tested. However, the fluorescence levels could not be distinguished from the control strain devoid of genes encoding GFP (PRA108). In contrast, the fluorescence of a strain (PRA106) carrying the gene encoding 3vGFP controlled by the same promoter was 3-fold higher than the background level and thus the level of the single sfGFP strain (Fig. 1B).

To test the utility of 3vGFP as output signal in a synthetic biology setting, we constructed versions of the strain with the pSPLA10 promoter turned OFF. The promoter is activated (ON) when a hybrid GAL4 activation domain binds a cognate hybrid GAL4 DNA-binding domain, which interacts with GAL4-binding sites of pSPLA10. The protein–protein interaction domains were based on the known Krev1 and RalGDS interaction domains (Herrmann et al. 1996). However, omitting the DNA-binding domain prevents reconstitution of a functional transactivator (OFF). These ON/OFF effects of present DNA-binding domain remained hidden below the background levels of the sfGFP strains, while observable in strains with 3vGFP as output (Fig. 1B).

Stability towards recombination

Direct-repeat recombination in mitotic S. cerevisiae is reported to occur at rates between $5.8 \times 10^{-5}$ and $12 \times 10^{-5}$ per cell generation for repeats of several kilo base pair identity (Dornfeld and Livingston 1992). This recombination rate is linearly dependent on identity length at such long segments; however, the rate drops rapidly below the minimal efficient processing segment (MEPS) length at around 250 bp in S. cerevisiae (Jinks-Robertson, Michelitch and Ramcharan 1993). While internal identity of 3vGFP ranges 74–84% (Fig. 2B), the identical segments are maximally at a 10-fold shorter length than the MEPS.

To test the recombination stability of 3vGFP, we wanted to measure whether the fluorescence levels originating from 3vGFP would attenuate following repeated culturing. While the 3vGFP molecule is engineered to limit direct-repeat recombination, long-term cultivation could potentially still lead to this especially if favored by a concurrent fitness advantage. To test stability at high expression level, we therefore also chromosomally integrated 3vGFP under control of the strong promoter from TEF1 i.e. at a level surpassing the intended use of 3vGFP. Expressing 3vGFP from the TEF1 promoter caused a considerable cost in fitness of approximately 15% in YPD, reducing the growth rate from an average of $0.35–0.30$ h$^{-1}$ compared to the negative control strain PRA108. Following culturing by serial passing (2%) of liquid cultures for 25 generations of three parallel lineages, single-cell level analysis revealed that the average fluorescence level of the cell population had diminished by 7%, perhaps due to spontaneous direct-repeat recombination. The single cell-level visualization indicated a slight left shift of the population (Fig. 2A). These results exemplify that direct-repeat recombination can occur within 3vGFP in S. cerevisiae and if selected for, these effects can become significant. However, since 3vGFP is intended for use at levels of low expression, a fitness advantage is not likely to further drive diminished fluorescence at a typical utility of 3vGFP.

Application of 3vGFP to construct an inducible promoter with reduced leakiness

Inducible promoters are important for development of e.g. synthetic genetic circuits, but the leakiness levels can be problematic in certain uses. To demonstrate the utility of 3vGFP, we therefore wanted to use it as output for genetic reengineering of the popular Cu$^{2+}$-responsive promoter of S. cerevisiae CUP1. pCUP1 has been employed in many different biotechnological cases (Labbé and Thiele 1999; Scholz et al. 2000; Rugbjerg et al. 2013), but displays considerable baseline activity (leakiness). pCUP1 induction results from elevated Cu$^{2+}$ concentrations mediated through binding of Cu$^{2+}$ to the ACE1 transcription factor, which in turn binds to UAS elements of pCUP1 (Huibregtse 1989; Evans, Engelke and Thiele 1990) (elements schematically depicted in Fig. 3A). The leakage level of pCUP1 measured with 3vGFP corresponded to 2.5-fold the cell autofluorescence (Fig. 3B). Based on the regulatory mechanism of ACE1, we anticipated that trace levels of Cu$^{2+}$ in the growth medium did not cause this leakiness, but rather assumed this basal transcriptional activity to be ACE1 independent. Accordingly, as strategy we hypothesized that swapping the promoter region downstream of ACE1 UASs for a transcriptionally repressed promoter could
Figure 3. Development of weak Cu$^{2+}$-responsive promoter through characterization with 3vGFP. (A) Organization of DNA-binding sites for the Cu$^{2+}$-responsive ACE1 activator and UME6 repressor in the wild-type CUP1 promoter and the new dimmed, hybrid promoter pCUP1dim. (B) OFF-level fluorescence measured in absence of Cu$^{2+}$ demonstrating the lower activity of the new hybrid promoter as captured with 3vGFP. Error bars depict standard error from biological replicates ($n = 3$). (C) Fluorescence of strain populations in response to addition of Cu$^{2+}$. Flow cytometry of representative example shown. Each sample contained 10 000 cells. The maxima of the samples are standardized to an equal top point.

provide attenuation, while maintaining the response to ACE1-dependent induction. We therefore combined the upstream region of pCUP1 (−149 to −454) containing three ACE1-binding sites, with part of the S. cerevisiae pSPO13 (−1 to −157) including its UME6 repressor-binding site (Fig. 3A). This new promoter (pCUP1dim) controlling 3vGFP resulted in fluorescence that was reduced approximately 61% (before background subtraction) to levels close to the cell autofluorescence (Fig. 3B), while the promoter remained responsive to addition of Cu$^{2+}$ (Fig. 3C).

The recombination-stabilized tandem GFP described in this study can enable characterization of minimally expressed genes in recombination-efficient organisms such as S. cerevisiae and other yeasts. As shown in this study, 3vGFP allowed characterization of the activation of a weak promoter and accordingly characterization of manipulations taking place at such low expression levels. Further, this particular approach of recombination-stabilizing GFPs with different protein and nucleotide sequences can be scaled in number. Recent brighter fluorescent proteins could be applied such as mNeonGreen (Shaner et al. 2013).

In principle, sequence divergence could be generated strictly at nucleotide level through codon optimization of segments encoding the same protein. Codon optimization can however introduce significant effects on the translation efficiencies (Goodman, Church and Kosuri 2013). Another concern may be spurious promoter/RBS activities, which could theoretically cause transcription and translation initiation from locations within the tandem GFP, thus producing truncated tandem proteins. Such situations would complicate the isolation of promoter responses and might require alleviation of the second and third GFP start codon.

An alternative method for assessment of promoter activities could be the use of the fluorescent RNA of the Spinach family, which bypasses the step of translation since the RNA forms
the fluorescent signal (Paige et al. 2012; Pothoulakis et al. 2014). However, while the technology has potential for synthetic biological use, its general applicability remains to be seen, such as the detection limits for low expression levels. Further relevant, fluorescent in situ hybridization for RNA (RNA FISH) is a technique allowing sensitive detection of transcripts at single-cell level (Zenkulsen, Larson and Singer 2008). This alleviates genetic engineering, but entails more sample treatment than for detection of GFP fluorescence.

In this study, a new simple strategy for engineering tandem fluorescent proteins was employed to produce brighter GFP signals with improved stability towards loop-out recombination. GFPs with sequence variation mainly at nucleotide level were translationally linked to form a recombination-stabilized tandem GFP molecule 3vGFP. Such GFPs could be useful for characterizing promoter activities in the range where normal single GFP signals fall below the cell autofluorescence levels. We specifically applied the 3vGFP molecule to characterize the ON/OFF levels of a weak promoter, which was not possible using a single sfGFP, and to develop a new hybrid Cu²⁺-responsive promoter pCUP1dim with lower leakiness level. The plasmid pCU2-3vGFP encompassing the nucleotide sequence of 3vGFP and pCUP1dim will be deposited at the Addgene repository.

SUPPLEMENTARY DATA
Supplementary data are available at FEMSYR online.

ACKNOWLEDGEMENTS
George Church is acknowledged for sfGFP encoded on pJ251-GERC (AddGene plasmid 47441).

FUNDING
This work was supported by the Novo Nordisk Foundation, the European Union Seventh Framework Programme (FP7-KBBE-2013-7-single-stage) under grant agreement no. 613745, Promys, and Deutsche Bundesstiftung Umwelt.

Conflict of interest. None declared.

REFERENCES


Supporting information for

Recombination-stable multimeric green fluorescent protein for characterization of weak promoter outputs in *Saccharomyces cerevisiae*

Authors:

Peter Rugbjerg – petru@biosustain.dtu.dk

Christoph Knuf – chrik@biosustain.dtu.dk

Jochen Förster – jfor@biosustain.dtu.dk

Morten O. A. Sommer - msom@bio.dtu.dk
**Construction of plasmids**

Plasmids were all assembled from individual PCR fragments using uracil-excision cloning. Table S1 lists the PCR fragment composition (by primers and DNA template) leading to the indicated plasmid. Further detail on PCR templates is given in Table S2, while the respective oligonucleotide sequences are specified in Table S3.

<table>
<thead>
<tr>
<th>Table S1 PCR fragments used to assemble the listed plasmids using the indicated oligonucleotide pairs and DNA template. Overview of template plasmids is given in Table S2.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid ID</td>
</tr>
<tr>
<td>pPR4-3vGFP</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>pPR4-sfGFP</td>
</tr>
<tr>
<td>pCU2-3vGFP</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>pCU3-3vGFP</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>pDS1U-X2-3vGFP</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table S2 Plasmids used as PCR template for plasmid construction (listed in Table 1).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
</tr>
<tr>
<td>pFA6a-yeGFP3-CLN2PEST-natMX6</td>
</tr>
<tr>
<td>pADH1-tc3-GFP</td>
</tr>
<tr>
<td>pl251-GERC/Addgene plasmid 47441</td>
</tr>
<tr>
<td>pEXP22</td>
</tr>
<tr>
<td>Vector</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>pRS416-PKS12</td>
</tr>
<tr>
<td>pDS1U-X2</td>
</tr>
<tr>
<td>pESC-HIS</td>
</tr>
</tbody>
</table>
Table S3 Sequence of all oligonucleotides used in the study

<table>
<thead>
<tr>
<th>ID</th>
<th>Oligonucleotide sequence (5’→)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P15</td>
<td>AGGACGGAUCGCTTGCTGTAACCTTAC</td>
</tr>
<tr>
<td>P16</td>
<td>ATCCGTCCUAAGAAACCATTATTATCATGACAT</td>
</tr>
<tr>
<td>P53</td>
<td>ATCGGAUACTCGAGGGGG</td>
</tr>
<tr>
<td>P54</td>
<td>ATCCGTCCUGGAAGTCTCATGGAGATT</td>
</tr>
<tr>
<td>P55</td>
<td>ATATTTCUCGACTCAACTTCAATC</td>
</tr>
<tr>
<td>P57</td>
<td>ATCCGAUGCCATCTTTACAGTCC</td>
</tr>
<tr>
<td>P60</td>
<td>AAAAAACUGTATTATAAGTAAATGCATG</td>
</tr>
<tr>
<td>P84</td>
<td>AGAAATAATGATCTAAAGGTTGAGGAATTTATCATG</td>
</tr>
<tr>
<td>P90</td>
<td>AGGTGCUCATCGCACGC</td>
</tr>
<tr>
<td>P91</td>
<td>AGCGACCUGAATAACCCCTACAAAGGG</td>
</tr>
<tr>
<td>P134</td>
<td>ATGTGATGAUTGATTTGAGATTGACAGT</td>
</tr>
<tr>
<td>P211</td>
<td>ACCCTTGUAAGAGCTCATCCATGC</td>
</tr>
<tr>
<td>P212</td>
<td>AGGAGGUATGGCTAGCAAAGGAGAAGAA</td>
</tr>
<tr>
<td>P213</td>
<td>ACCTCCTUTGTACGATCCATCCATACCAT</td>
</tr>
<tr>
<td>P214</td>
<td>AGGACGGAUATCCACACGTCCATAGCCC</td>
</tr>
<tr>
<td>P219</td>
<td>ATCCCTUTGTACGATCCATCCATACCAT</td>
</tr>
<tr>
<td>P376</td>
<td>ATCTCAGUTCGGTAGGTTACGTCGTC</td>
</tr>
<tr>
<td>P377</td>
<td>ACTGAGAUACCTACACGTGAGCTAT</td>
</tr>
<tr>
<td>P378</td>
<td>AGTTCCGUTTAGCTAGTTAGTACCTTTG</td>
</tr>
<tr>
<td>P379</td>
<td>ACACACUGCACAATCCATATTGCCTTG</td>
</tr>
<tr>
<td>P540</td>
<td>AATTATTUTTTTGAATTTAAACTTGATTAGATTGC</td>
</tr>
</tbody>
</table>
SI references

