Title

Modulating bacterial function utilizing A knowledge base of transcriptional regulatory modules

Authors

Shin, Jongoh

Zielinski, Daniel C

Palsson, Bernhard O

Publication Date

2024-08-28

DOI

10.1093/nar/gkae742

Nucleic Acids Research, 2024, 1–16
https://doi.org/10.1093/nar/gkae742
Synthetic Biology and Bioengineering

Modulating bacterial function utilizing A knowledge base of transcriptional regulatory modules

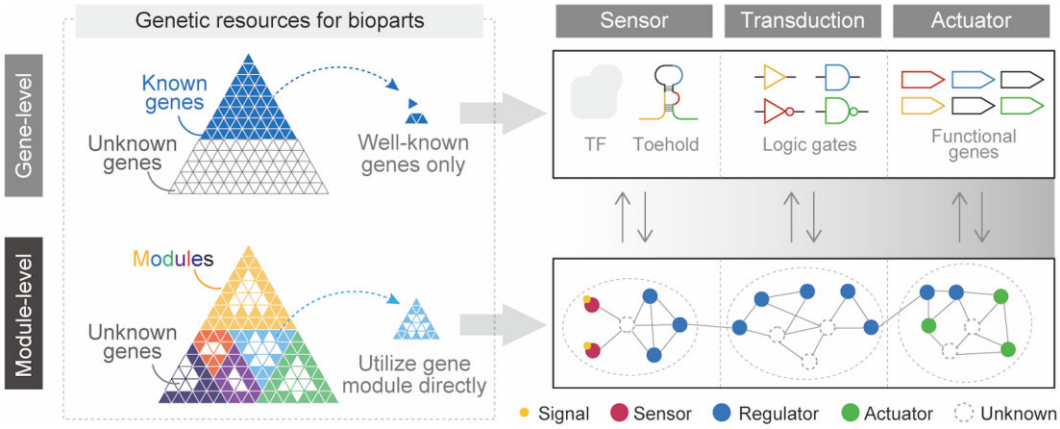
Jongoh Shin[1], Daniel C. Zielinski[1] and Bernhard O. Palsson[1,2,3,*]

[1]Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA
[2]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby 2800, Denmark
[3]Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA
[*]To whom correspondence should be addressed. Tel: +1 858 246 1625; Email: bpalsson@ucsd.edu

Abstract

Synthetic biology enables the reprogramming of cellular functions for various applications. However, challenges in scalability and predictability persist due to context-dependent performance and complex circuit-host interactions. This study introduces an iModulon-based engineering approach, utilizing machine learning-defined co-regulated gene groups (iModulons) as design parts containing essential genes for specific functions. This approach identifies the necessary components for genetic circuits across different contexts, enhancing genome engineering by improving target selection and predicting module behavior. We demonstrate several distinct uses of iModulons: (i) discovery of unknown iModulons to increase protein productivity, heat tolerance and fructose utilization; (ii) an iModulon boosting approach, which amplifies the activity of specific iModulons, improved cell growth under osmotic stress with minimal host regulation disruption; (iii) an iModulon rebalancing strategy, which adjusts the activity levels of iModulons to balance cellular functions, significantly increased oxidative stress tolerance while minimizing trade-offs and (iv) iModulon-based gene annotation enabled natural competence activation by predictably rewiring iModulons. Comparative experiments with traditional methods showed our approach offers advantages in efficiency and predictability of strain engineering. This study demonstrates the potential of iModulon-based strategies to systematically and predictably reprogram cellular functions, offering refined and adaptable control over complex regulatory networks.

Graphical abstract



Introduction

Synthetic biology aims to reveal biological design principles and reprogram cellular functionalities for user-defined purposes by systematically constructing biological genetic circuits (1,2). Despite considerable advancements, the design of functional genetic circuits continues to face challenges related to scalability and predictability (3–6). These challenges stem from limitations in current gene-centric approaches and the abstraction hierarchy (3), which often fail to account for context-dependent performance of genetic components and unpredictable circuit-host interactions (3,4,6).

Several pioneering approaches have been undertaken to address these issues, including enhancing dynamic range and orthogonality (7), automation (8), whole-cell modeling (9), coarse-grained modeling (10,11), integrative circuit–host modeling (12), utilizing orthogonal transcription factors (TFs) (13), and local feedback regulation strategies (14,15). However, these approaches predominantly focus on predictions

Received: February 22, 2024. **Revised:** July 29, 2024. **Editorial Decision:** August 7, 2024. **Accepted:** August 19, 2024

and prevention of unintended results based on known gene functions or gene-centric circuit design. This unintended behavior then requires a redesign, increasing the complexity of the design combinatorially rather than linearly with the number of components, making the process cumbersome and unscalable.

Recently, we have defined transcriptional regulatory modules, known as iModulons (iMs) (16,17), for various cellular functions across bacteria species (17–22) by using independent component analysis (ICA) on organism-specific gene expression compendia. By defining 'weak regulatory linkages' in the whole genetic networks, which have been refined over billions of years in nature (3), we decompose the activity of the transcriptional regulatory network into iMs, which are independently regulated gene sets for specific cellular functions (Figure 1A). This approach is expected to facilitate module-centered system composition prediction across various environmental contexts, offering several advantages in the genome engineering process such as 1) target selection, 2) pre-estimation of module behavior and interactions, and 3) precise strain design at the network level (Supplementary Note S1, Supplementary Figure S1). Unlike conventional approaches—genome-scale metabolic models (GEMs), which rely heavily on current gene annotations; multi-omics analyses, which often result in a large number of differentially expressed genes that are difficult to interpret; and functional screening, which primarily identifies genes related to cell fitness—this method allows us to easily identify gene groups responsible for target cellular functions without fully understanding each gene's function. This not only enables the discovery of new modules responsible for cellular functions but also enhances the scalability of genetic circuits by shifting from gene-centric to module-level design (Figure 1B). Furthermore, this approach enhances our understanding of the behavior of each module and their interactions with other modules across all tested environments (Supplementary Note S1, Supplementary Figure S1). Consequently, these increased scalability and predictability can reduce the number of genome designs needing testing and accelerate the design process.

In this study, we introduce iM-based engineering approaches in *Escherichia coli*, *Vibrio natriegens* and *Pseudomonas putida* as model organisms. These strategies include target selection, estimating module behavior across various conditions, and implementing four types of design, which we term iModulon Discovery, Boosting, Rebalancing, and Rewiring (Figure 1C). (i) Through iModulon Discovery, we identified previously unknown modules that enhanced protein production, heat tolerance, and fructose utilization. (ii) Using an iModulon Boosting module, we incorporated necessary modules into the native network, significantly improving cell growth under osmotic stress and poor carbon source conditions. (iii) Using an iModulon Rebalancing module, we minimized the trade-offs caused by previous adaptive laboratory evolution experiments under oxidative stress, reducing the negative impacts on natural competence activity and antibiotic sensitivity at the network level. (iv) Lastly, iModulon Rewiring enabled the activation of natural competence under general growth conditions by rewiring several iMs comprising 189 genes. This module-based strategy offers easy-to-implement and adaptable control over cellular functions to achieve coherent and minimally disruptive modifications to native transcriptional regulatory networks on a novel scale.

Materials and methods

Bacterial strains and growth conditions

The bacterial strains used and developed in this study are listed in Supplementary Table S1. The wild-type strain *Vibrio natriegens* ATCC 14048 (*V. natriegens*), *Escherichia coli* str. K-12 MG1655 (*E. coli*) and *Pseudomonas putida* KT2440 (*P. putida*) were used throughout this study. *Vn* was routinely cultivated at 30°C in LBv2 media (25 g/l LB Miller broth, 200 mM NaCl, 4.2 mM KCl, 23.14 mM MgCl2), with agitation at 180 RPM. LB and LBv2 were supplemented with 1.5% wt/vol agar for solid media. Unless stated otherwise, chemical reagents for cell culture were primarily obtained from Sigma-Aldrich (Burlington, MA).

E. coli and *P. putida* cells were routinely grown in LB medium (#71753) or M9 medium. The M9 medium contained 47.75 mM Na_2HPO_4, 22.04 mM KH_2PO_4, 8.56 mM NaCl, 18.70 mM NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and trace elements. Trace elements were prepared as a 2000 × concentrated solution, consisting of 100 mM $FeCl_3$, 9.54 mM $ZnCl_2$, 8.41 mM $CoCl_2$, 8.27 mM Na_2MoO_4, 0.75 mM $CaCl_2$, 0.91 mM $CuCl_2$ and 0.5 mM H_3BO_3 in a 3.7% (w/w) hydrochloric acid solution.

For constitutive expression experiments (pUC-PJ23106-SpeAB, pUC-PJ23106-EctABC-ThrA) and iModulon Boosting related to osmolarity stress in *V. natriegens*, NaCl concentrations ranging from 0 to 800 mM were added to M9 medium without NaCl (M9woNa). For constitutive expression experiments (pUC-AhpCF-KatG) and OxyR-iModulon rebalancing in *V. natriegens*, M9 medium with 258.5 mM NaCl and 0.5% (wt/vol) glucose (M9Na) was used. Oxidative stress was induced with H_2O_2 concentrations ranging from 0 to 0.7 mM for the WT, evolved VN-ALE (from a previous study (23)), and OxyRiM-Rebal-Vn strains.

For YjfJ iM experiments in *E. coli*, an M9 medium with 30 mM glucose was used. For acetate iM boosting in *E. coli*, M9 medium with 30 mM glucose, fructose, acetate, or glycerol was used. For OxyR-iModulon rebalancing experiments in *E. coli*, M9 medium with 30 mM glucose was used. Oxidative stress was induced with H_2O_2 concentrations ranging from 0 to 5 mM for the WT, previously evolved EC-ALE (23), and OxyRiM-Rebal-Ec strains. For PP_2260–3 iM experiments in *P. putida*, M9 medium with 30 mM fructose was used.

When necessary, antibiotic selection of *V. natriegens* was conducted using carbenicillin (Carb, 50 μg/ml), chloramphenicol (Cm, 10 μg/ml), kanamycin (Kan, 100 μg/ml), or spectinomycin (Spec, 360 μg/ml). For *E. coli*, ampicillin/carbenicillin (Amp/Carb, 100 μg/ml), chloramphenicol (Cm, 25 μg/ml), kanamycin (Kan, 50 μg/ml), or spectinomycin (Spec, 120 μg/ml) were used. For *P. putida*, kanamycin (Kan, 50 μg/ml) and gentamicin (Gent, 30 μg/ml) were used. For plasmid cloning, NEB® 10-beta Competent *Escherichia coli* (New England BioLabs, Ipswich, MA) was used, which was aerobically cultivated at 37°C in LB Miller broth or on LB agar plates. The appropriate antibiotics were added as follows: ampicillin/carbenicillin (Amp/Carb, 100 μg/ml), chloramphenicol (Cm, 25 μg/ml), kanamycin (Kan, 50 μg/ml), gentamicin (Gent, 20 μg/ml), or spectinomycin (Spec, 120 μg/ml).

iModulon analysis

For the iM-based engineering strategy implementation, we utilized iM data generated through Independent Compo-

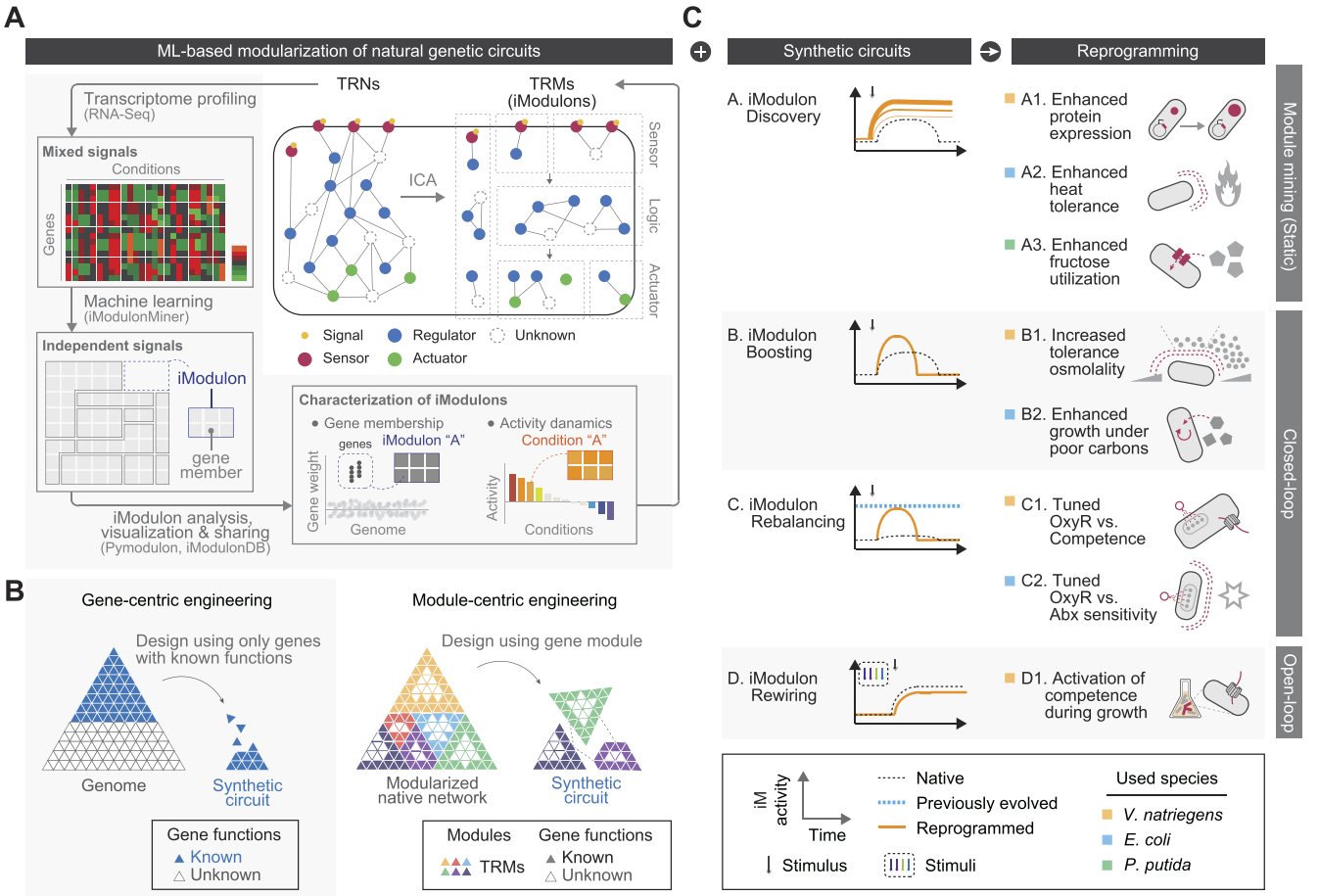


Figure 1. iModulon-based engineering. This diagram illustrates a strategy for engineering cellular functions by integrating iModulon (shortly, iM) and synthetic genetic parts. (**A**) The workflow begins with Independent Component Analysis (ICA) to find and identify transcriptional regulatory modules (TRMs) as independently modulated sets of genes (iMs) using a transcriptome compendium. The characterization of iModulons involves determining gene membership and activity dynamics under different conditions. This approach could identify previously uncharacterized genes and their function in particular cellular processes represented by iMs. iModulonMiner (https://github.com/sbrg/imodulonminer) (24), PyModulon (https://github.com/SBRG/pymodulon), and iModulonDB (https://imodulondb.org) are specialized software tools designed for distinct steps in the analysis process (18,24). The PRECISE-1K (*E. coli*), natPRECISE148 (*V. natriegens*), and putidaPRECISE321 (*P. putida*) iModulonDB were used in this study (https://imodulondb.org) (18). (**B**) Comparison of gene-centric and module-centric engineering. Gene-based engineering involves designing synthetic circuits or genome engineering using only genes with known functions. Module-based engineering leverages gene modules, including both known and unknown genes, from a modularized native network to design circuits or genomes. (**C**) Reprogramming through iModulons. iModulon Discovery leads to enhanced protein expression, heat tolerance, and fructose utilization (A1–A3). iModulon Boosting increases tolerance to osmolarity and improves growth under poor carbon conditions (B1–B2). iModulon Rebalancing fine-tunes genetic network activity, such as OxyR activity versus competence, and adjusts antibiotic sensitivity (C1–C2). iModulon Rewiring reprograms cellular behavior, such as activating competence during growth (D1).

nent Analysis (ICA) based on the natPRECISE148 iModulonDB (detailed at https://imodulondb.org/) (18). Routine iM analysis is able to be performed using the (i) iModulonMiner (https://github.com/sbrg/imodulonminer) (24), (ii) Pymodulon package (https://github.com/SBRG/pymodulon), and iModulonDB (https://imodulondb.org/) (24). Following methodologies described in prior studies (22,24), we modeled differences in iM activities across biological replicates using a log-normal distribution. Our statistical approach involved comparing mean activity differences across all iMs to compute *P*-values. These *P*-values were adjusted using the Benjamini-Hochberg method to address multiple hypothesis testing. iMs demonstrating an FDR change of less than 0.05 under specific conditions were identified as statistically significant.

Plasmid constructions

The oligonucleotides used in this study, listed in Supplementary Table S1, were sourced from Integrated DNA Technologies (Coralville, IA). High-fidelity PCR amplifications and genetic analyses utilized PrimeSTAR GXL DNA Polymerase (Takara Bio, Shiga, Japan) and Q5 High-Fidelity DNA Polymerase (New England Biolabs). PCR products and *E. coli*-derived plasmids were purified using the DNA Clean & Concentrator kit (Zymo Research) and Monarch Plasmid DNA Miniprep Kit (New England Biolabs), respectively. Newly constructed vectors and their assembly methods are detailed in Supplementary Table S1. Small-sized plasmids were assembled via NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs).

Construction of deletion mutants

To delete Arabinose and NarP-1 iM genes in *V. natriegens*, knockout (KO) plasmids were designed to generate transforming DNA (tDNA) following previously established methods (22,25). Homology arms (HAs) of ∼2 kb adjacent to target regions were amplified from *V. natriegens* genomic DNA. These HAs and antibiotic-resistance cassettes ($Carb^R$, $Spec^R$ or Kan^R) were integrated into the pACYC184 plasmid. This process resulted in ptDNA-'Δstrain name' plasmids, with sequences confirmed by Sanger sequencing (Eton Bioscience, San Diego, CA). The genes were then deleted and validated using the natural transformation process of *V. natriegens*, as described below in the 'Natural transformation' section.

P. putida deletion mutant was generated using homologous recombination and *sacB* counterselection using a previously established method (26). In brief, homology fragments, each between 1 and 2 kb upstream and downstream of the target gene, were cloned into allele exchange vector pMQ30. The plasmid was then introduced into *E. coli* S17 by electroporation and subsequently transferred into *P. putida* through conjugation. Transconjugants were selected on LB agar plates containing 30 μg/ml gentamicin and 25 μg/ml chloramphenicol. These conjugants were cultured overnight in LB medium with 30 μg/ml gentamicin and 30 μg/ml chloramphenicol, followed by plating on LB agar without NaCl, supplemented with 10% (wt/vol) sucrose. Potential deletion mutants were restreaked on LB agar without NaCl and supplemented with 10% (wt/vol) sucrose and then screened using PCR with primers flanking the target gene to verify gene deletion.

DNA assembly and cloning for iModulon circuits

The pcBAC15a shuttle vector was constructed for yeast DNA assembly and *V. natriegens* protein expression, refining a design from a previous study (27). It includes components such as bacterial artificial chromosome (BAC), CEN, p15A origin, chloramphenicol resistance, HIS3 marker, and oriT (Supplementary Figure S4A–B).

For large vector assemblies, DNA fragments with 60–70 bp homologies were transformed in *Saccharomyces cerevisiae* VL6-48 using the LiAc/SS carrier DNA/PEG method (28). Yeast clones with accurate bacterial artificial chromosome (BAC) assemblies were identified through colony PCR targeting the junctions of the constructs. These BACs were extracted using the Gentra Puregene Yeast/Bact. Kit (Qiagen), electroporated into *E. coli* NEB10β cells, and verified via Whole Plasmid Sequencing service by Plasmidsaurus (Eugene, OR).

To induce GFP expression in *V. natriegens*, three vector variants were generated: pcBAC-Ptac-GFP, pcBAC-Ptet-GFP, and pcBAC-PBAD-GFP, each inducible by IPTG, anhydrotetracycline, or arabinose, respectively. Additionally, four vectors (pcNarP1iM-1 to -4) based on pcBAC-Ptac-GFP were constructed for NarP-1 iM activation alongside GFP expression (Supplementary Figure S4C). Two vectors (pcGroESL and pcDnaJK) derived from pcNarP1iM-3 were designed for the constitutive expression of chaperone proteins along with GFP (Supplementary Figure S4D). Furthermore, two vectors (pcYjfJ-I and pBb-Ptac-2660iM) were constructed for the inducible expression of YjfJ in *E. coli* and PP_2660–3 iMs in *P. putida* (Supplementary Figures S4E, F).

For Low-osmolarity iM-Boost, Ectoine iM-Boost, and OxyR-Rebalancing module integration into the *V. natriegens* genome, we first created two plasmids, pcBAC-tetR-Boost-GFP and pcBAC-araC-Boost-GFP. Post-validation experiments (Supplementary Figures S6B–C, S6E), pcBAC-araC-Boost-GFP was selected for module development, leading to five shuttle vectors (Supplementary Figures S8A–B, E). These modules, with ∼2 kb HAs, were used for genomic integration. To avoid partial integration, we used pcLOiMBoost-REV and pcEctiMBoost-REV vectors to generate transforming DNA (tDNA) for module integration.

For Acetate iM-Boost, and OxyR-Rebalancing module integration into the *E. coli* genome, we created three shuttle vectors (Supplementary Figures S8C–D, F). These modules, with 50 bp HAs and CRISPR/Cas9 spacer binding sites were used for genomic integration through replicon excision enhanced recombination (REXER) (please see below for details for the REXER step).

For the rewiring of competence iMs within BAC vectors, we initially created three vectors (Supplementary Figure S8G): pcBAC-T (with the *tfoX* gene under the P_{Trc} promoter), pcBAC-TQ-v1 (harboring *tfoX* and *qstR* under the P_{Trc} promoter), and pcBAC-TQUA-v1 (containing *tfoX* and *qstR* under the P_{Trc} promoter, with UC-1 iM and *ahpCF* genes under the P_{BAC} promoter). To reposition *qstR* expression from the P_{Trc} promoter to the P_{BAC} promoter and to separate UC-1 iM and *ahpCF* genes from the earlier constructs, we further developed pcBAC-TQ-v2, pcBAC-TQU-v2, and pcBAC-TQUA-v2 (Supplementary Figure S8H). These revised vectors were then used in experiments related to growth-enabled competence.

Electroporation

Electrocompetent *V. natriegens* cells were prepared following a protocol from a previous study (22). Initially, cells were cultured overnight in the LBv2 medium at 30°C. The following steps involved centrifuging the cells at 5000 × g for 5 min at 4°C, washing them with LBv2, and then inoculating them into fresh LBv2 medium at a 1:100 dilution. When the culture reached an OD_{600} of 0.4, the cells were harvested and washed twice with cold 1M sorbitol. The final cell suspension was prepared in 250 μl of cold 1M sorbitol and aliquoted into 50 μl portions. For electroporation, 100–200 ng of plasmid DNA was mixed with the electrocompetent *V. natriegens* cells. This mixture was then transferred to a 0.2 mm Gene Pulser cuvette (Bio-Rad Laboratories). Electroporation was performed using a Bio-Rad Gene Pulser, set to 800 V, 25 μF and 1000 Ω parameters. Following electroporation, cells were allowed to recover in 1 ml of LBv2 for 1 hour at 30°C with agitation. The recovered cells were then plated on LBv2 agar plates supplemented with the appropriate antibiotics. The plates were incubated for at least 12 hours at room temperature or 8 h at 30°C to allow colony formation.

For electroporation of *E. coli* and *P. putida*, cells were cultured overnight in LB medium at 30°C (for *P. putida*) or 37°C (for *E. coli*). The next day, the cultures were inoculated into fresh LB medium at a 1:1000 dilution. When the cultures reached an OD_{600} of 0.5, the cells were harvested and washed three times with cold 10% (vol/vol) glycerol. For the electroporation step, 100–1000 ng of vector DNA was mixed with the electrocompetent cells, and the mixture was transferred to a 0.2 mm Gene Pulser cuvette. Electroporation was performed using a Bio-Rad Gene Pulser set to 2.5 kV, 25 μF and 100 Ω. After electroporation, the cells were recovered in 1 ml of SOC medium for 1 h at 30°C (for *P. putida*) or 37°C (for *E. coli*) with agitation. The recovered cells were then plated

on LB agar plates containing the appropriate selection markers. For plasmid cloning and introducing BACs from purified yeast DNA into *E. coli*, electro-transformation was performed using *E. coli* NEB10β (New England Biolabs) to facilitate propagation.

Natural transformation

The natural transformation was conducted as described previously (22). Strains harboring competence plasmids (e.g. pACYC-TfoX, pcBAC-T, pcBAC-TQ-v1, pcBAC-TQUA-v1, pcBAC-TQ-v2, pcBAC-TQU-v2, pcBAC-TQUA-v2) for inducible *tfoX* gene expression with IPTG were cultured overnight in LBv2 medium supplemented with 10 μg/ml chloramphenicol and 100 μM IPTG at 30°C in a roller drum. Cultures were adjusted to an OD_{600} of 4.0 for transformation under normal competence conditions. 5 μl of this culture was added to 350 μl of competence buffer (28 g/l Instant Ocean Sea Salt, https://www.instantocean.com) with 500 μM IPTG. 50 ng of tDNA was then introduced, and the mixture was incubated at 30°C for 4 h without agitation. Afterward, the mixture was transferred to 1 ml of fresh LBv2 and recovered for 2 h at 30°C with shaking at 180 rpm. Cells were then centrifuged, resuspended in 200 μl of LBv2, and plated for colony-forming units (CFUs) analysis. Transformation efficiency was assessed using dilution factors of 1:2 and 1:10 on LBv2 agar plates supplemented with 50 μg/ml carbenicillin, 100 μg/ml kanamycin, or 360 μg/ml spectinomycin. For total CFUs, dilutions of 1:10 000 and 1:100 000 were used on non-selective LBv2 agar plates. The efficiency was calculated as the ratio of transformants to total CFUs. tDNA integration was confirmed via PCR using primers listed in Supplementary Table S1.

For gene deletions or insertion of genetic modules, tDNA was prepared using previously described methods (22). For gene deletions, tDNA was PCR-amplified utilizing the 'Δstrain name'_Left_For/Right_Rev primer pair and corresponding knockout (KO) plasmids as templates (Supplementary Table S1). After amplification, we treated the PCR products with DpnI enzyme (New England Biolabs) and purified them using the DNA Clean & Concentrator kit (Zymo Research). To integrate the Low-osmolarity iM-, Ectoine iM-Boost modules, and the OxyR-Rebalancing-module into the *V. natriegens* genome, we amplified each module's respective tDNA using the LowOs_tDNA_F2/R, Ect_tDNA_F/R or Oxy_tDNA_F/R primer sets. The templates used were pcLOiMBoost-REV, pcEctiMBoost-REV and pcOxyRiM-Rebalancing, respectively. The amplified tDNA was purified from agarose gels with the Zymoclean Gel DNA Recovery Kit (Zymo Research). PCR was performed to confirm the module integration using purified genomic DNA with specific primers (Supplementary Figures S6F–G, S11D).

Competence activity under growth conditions was induced using control strains and those with rewired competence iMs. Natural transformation assays using these iMs-rewiring strains were carried out under both competence and growth conditions, as described above. or initiating natural competence in growth conditions, the strains were cultured overnight in LBv2 medium with 10 μg/ml Cm and 100 μM IPTG. The cultures were then centrifuged for 5 min at 5000 × g at 4°C, the supernatant discarded, and the cells resuspended and adjusted to an OD_{600} of 4.0. Then, 5 μl of culture was added to fresh LBv2 with 500 μM IPTG and varying arabinose concentrations. We added 50 ng of tDNA to the cell mixtures and gently combined them. After incubating for 4 h at 30°C with agitation, cells were plated on selective and non-selective LBv2 agar plates to determine transformants and total CFUs. PCR confirmation of tDNA integration used ΔEndA_Val_For/Rev primers and was validated by Sanger sequencing.

Replicon excision enhanced recombination (REXER)

AcetateiM-Boost, Acetate iM-OE, and OxyR-Rebalancing modules were integrated into the *E. coli* genome using a previously established REXER approach (29–31) with minor modification. Initially, a *sacB-cat* dual-selection landing pad was inserted via lambda-red recombination into the MG1655 strain genome, carrying the pREDCas9 plasmid, either below the *aldA* gene (for AcetateiM-Boost and Acetate iM-OE modules) or the *ahpF* gene (for the OxyR-Rebalancing module). Next, BAC DNA containing the respective iM modules (pAceiM_Boost_Ec, pAceiM_OE_Ec, pOxyRiM_Rebal_Ec) was introduced into the strain by electroporation. The recipient cells were then grown in SOB medium with 25 μg/ml chloramphenicol, 100 μg/ml carbenicillin, 120 μg/ml spectinomycin, and 30 mM L-arabinose at 37°C until reaching an OD_{600} of 0.6. Six micrograms of PCR-amplified linear universal guide-RNA construct (31), targeting the BAC, were introduced via electroporation, and cells were screened on NaCl-free LB-agar plates (10 g/l tryptone, 5 g/l yeast extract, and 1.5% (wt/vol) agar) containing 100 μg/ml carbenicillin, 120 μg/ml spectinomycin, and 10% sucrose. The phenotype and genotype of post-REXER colonies were further validated by testing for chloramphenicol sensitivity and performing PCR genotyping (Supplementary Figures S6H, S12E).

Optical density and fluorescence measurement

In flask culture experiments, OD_{600} of bacterial cultures was measured at 600 nm using the BioMate™ 3S Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). For high-throughput assessments, the Infinite 200 Pro Plate Reader (Tecan, Männedorf, Switzerland) was used for OD_{600} measurements in Flat Bottom 96-well plates, each containing 100 μl of culture. For experiments involving Green Fluorescent Protein (GFP), fluorescence was measured alongside OD_{600} at 485/515 nm excitation/emission wavelengths using the same plate reader and Bio-One CELLSTAR μClear™ 96-well plates (Greiner, Kremsmünster, Austria). Plates were incubated at 30°C with orbital shaking, with readings taken at 10 to 15-minute intervals. Relative fluorescence units (RFU) were calibrated by subtracting blanks, which included the medium with and without molecules. Growth dynamics and GFP expression analysis involved calculating lag-time (λ), specific growth rate (μ), and specific GFP rate via linear regression, utilizing the QurvE web tool (32). In some cases, growth rates and GFP rates were normalized against control samples for intuitive comparison.

RNA-sequencing

Biological duplicates of five engineered strains (V, T, TQ-v2, TQU-v2, and TQUA-v2) were prepared for competence iM rewiring studies under growth conditions with 0.001% and 0.1% arabinose. Post-induction of natural competence (Figure 5F), cells were centrifuged at 5000 × g for 10 mins at 4°C, and the supernatant was discarded. RNA-Seq library

construction followed the previously described protocol (22). Briefly, after cell harvesting and supernatant removal, RNA was isolated using the Quick-RNA Fungal/Bacterial Microprep Kit (Zymo Research, Irvine, CA). rRNA and DNA contaminants were removed from the RNA using the RiboRid method (22,33), and the depletion of rRNA was confirmed with the 4150 TapeStation System (Agilent, Santa Clara, CA). These rRNA-depleted samples were then processed into libraries using the KAPA RNA HyperPrep kit (Roche, Basel, Switzerland). Library integrity was verified with the Agilent 4150 TapeStation System, and quantification was conducted using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA). The libraries were then pooled for sequencing on the Illumina NovaSeq 6000 platform at the UC San Diego IGM Genomics Center using a 100 bp single-end protocol.

For data processing and initial RNA-Seq quality control, raw reads were trimmed with Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Aligned reads to the *V. natriegens* reference genome (GCA_001456255.1) (34) were achieved using Bowtie (35), with SAM to BAM file conversion via Samtools (http://www.htslib.org/). Gene read counts were determined using RSeQC (36) and FeatureCounts (37), with all quality control metrics aggregated by MultiQC (https://multiqc.info/) (38). We ensured RNA-Seq data quality by confirming compliance with FASTQC standards, including per_base_sequence_quality, per_sequence_quality_scores, per_base_n_content, and adapter_content. A high correlation was observed within biological replicates ($R^2 > 0.9$) in this RNA-Seq data. The read counts were then normalized and presented as $\log_2$-transformed Transcripts per Million (logTPM) for downstream analysis. To identify differentially expressed genes (DEGs), raw read counts were normalized using the Bioconductor package DESeq2 (39). Significance was determined with an adjusted *P*-value (P_{adj}) < 0.05.

Statistical analysis

Beyond transcriptomic analysis and ICA, further statistical evaluations were conducted using GraphPad Prism v10 software (GraphPad, San Diego, CA, USA). These included Pearson's correlation coefficient, the two-tailed Student's *t*-test, and the two-tailed Wilcoxon–Mann–Whitney test. Statistical significance was established at *P*-values <0.05.

Results

iM-based engineering approaches

Traditional static regulation methods such as gene knockout, substitution, or modulation of gene expression utilize rationally selected target genes with known functions. In contrast, iM-based engineering approaches allow the discovery and direct engineering of modules without requiring a complete understanding of each gene's function (Figure 1, Supplementary Figure S2A). For instance, the 'iModulon Discovery' approach discovers and directly utilizes previously unknown modules to engineer strains.(Figure 1C, Supplementary Figure S1).

Further examples showed that iM-based engineering could replace conventional genetic circuit components, such as sensor, transduction, or actuator modules, with iMs by decomposing whole genetic networks through ICA (Figure 1A, Supplementary Figure S2A). This approach expands traditional genetic circuit design principles with increased scalability (Figure 1B, Supplementary Figure S2B). Unlike conventional genetic circuit modules, iM analysis offers detailed insights into their behavior (e.g. changes in module activity) and interactions (e.g. transcriptional relationship between modules) within the cell across various conditions (Supplementary Note S1, Supplementary Figure S1). This knowledge allows for precise modulation of cellular functions at the network level. Based on these advantages, we utilized iMs as engineering modules to construct general genetic circuits, such as closed-loop (feedback) and open-loop (feedforward) systems (Supplementary Figure S2C). Examples include the implementation of 'iModulon Boosting', 'iModulon Rebalancing' and 'iModulon Rewiring' in endogenous genetic circuits for strain engineering (Figure 1C).

iModulon discovery

Bacterial regulons, typically activated by specific stimuli, initiate cellular responses to neutralize these triggers and subsequently deactivate to conserve resources efficiently (40) (Figure 2A). By decomposing the transcriptome database into independent genetic components, this approach links genes to iMs, identifying unknown modules activated by specific conditions and suggesting their functional roles (22). We applied this 'iModulon Discovery' workflow in *V. natriegens*, *E. coli* and *P. putida* species, enhancing cellular functionality through both constitutive and inducible activation of iMs. This strategy is particularly effective for processes requiring sustained cellular activity, ensuring persistent response and improved output efficiency.

First, we targeted the NarP-1 iM in *V. natriegens*, a component of the NarP regulon previously identified in other *Vibrio* species (Supplementary Figure S3A), comprising the *napABCDF* and *nirBCD* genes (Supplementary Figure S3B). We observed significant activation in all protein overexpression (OE) samples (Figure 2B, Supplementary Figure S3C) in nat-PRECISE148 iModulonDB. Typically responsive to nitrate or nitrite (41), the NarP-1 iM displayed an activation pattern most pronounced in the mid-exponential phase, diminishing in later stages (Supplementary Figure S3D). This suggests its potential involvement in the targeted protein OE process. However, using traditional differential expression analysis, it is challenging to discern the potential role of the NarP regulon due to the large number of DEGs (1182 shared among the protein OE samples) (Supplementary Figure S3E). Gene deletion studies further elucidated the NarP-1 iM's role. Deleting the nitrate reductase genes (Δ*napABCDF*), designated as ΔNarP-1iM-A (Supplementary Figure S3F), resulted in a significant reduction in GFP production (0.56-fold, $P < 0.01$; Supplementary Figure S3G–H) at both microplate and flask levels (0.56-fold, $P = 0.0013$; Supplementary Figure S3I). In contrast, ΔNarP-1iM-B (Δ*nirBCD*) showed no notable reduction of GFP production, which could be due to compensation by an additional nitrite reductase gene cluster (RS19420–RS19425) (Supplementary Figure S3J). These results indicate a potential correlation between NarP-1 iM activation and protein production, warranting further molecular investigation.

To investigate whether sustained activation of the NarP-1 iM could enhance heterologous protein expression, we engineered a plasmid-based expression system incorporating NarP-1 iM genes. Utilizing the pcBAC15a shuttle vector

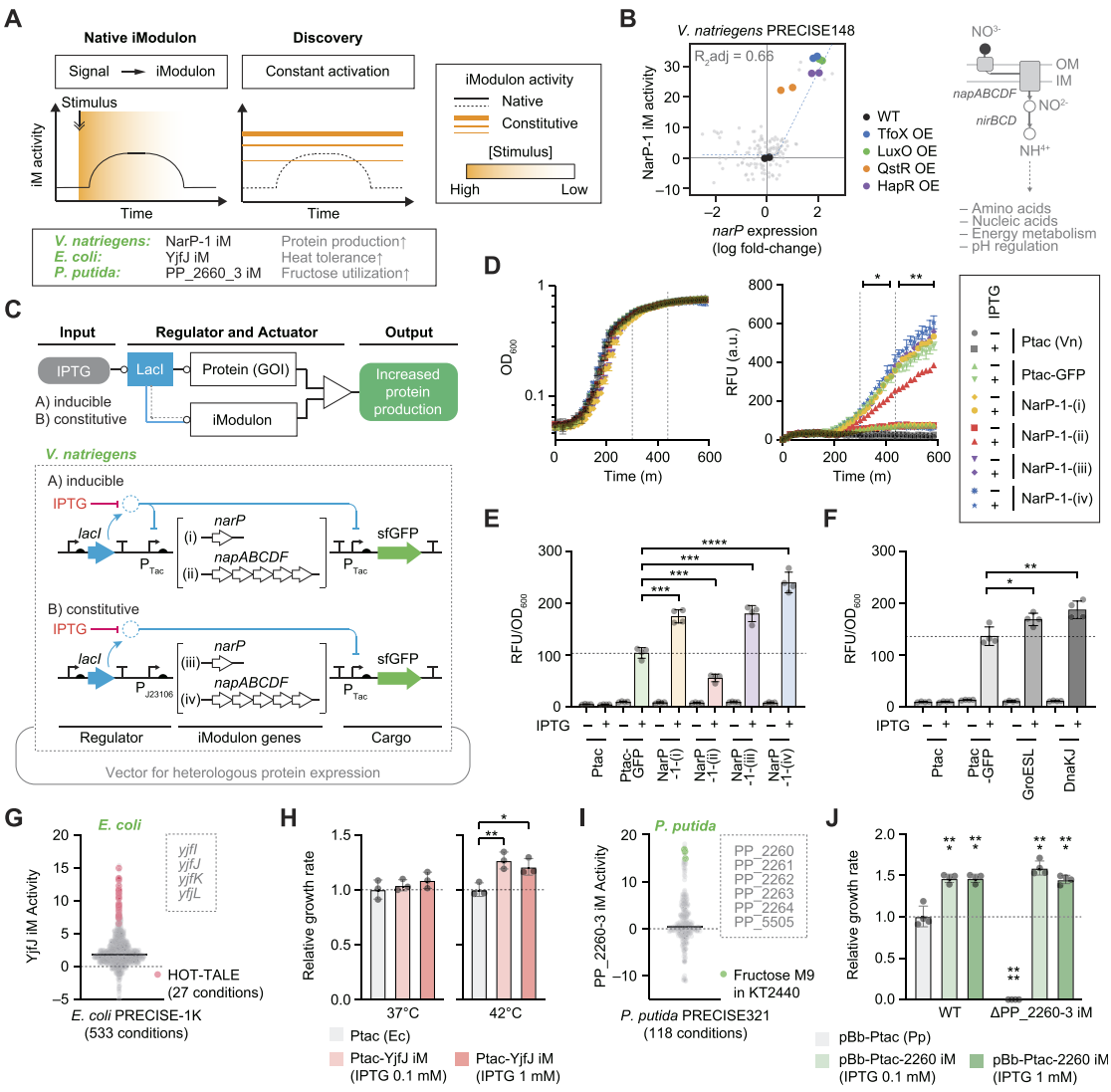


Figure 2. iModulon Discovery in various bacterial species. (**A**) Native iModulon Discovery. The left panel shows general regulation of iM activity in response to stimuli, typically activating pathways (e.g. catabolic) to counteract stimulus. The right panel illustrates engineered iM activity regulation using constitutive or inducible promoters integrated with iM genes for targeted activation. Constant activation of new iMs leads to enhanced protein production, heat tolerance, and fructose utilization across different species. (**B**) NarP-1 iM's activation in protein overexpression samples in *V. natriegens* PRECISE148 iModulonDB (18). The left panel shows the correlation between the NarP-1 iM activation and *narP* transcription factor expression with protein overexpression (OE) samples (colored circles) from a previous study (22). The right panel outlines the gene composition and known enzymatic functions of the NarP-1 iM. (**C**) Design of protein expression system utilizing the NarP-1 iM in *V. natriegens*. Schematic of designed circuits, either inducible (NarP-1-(i) and (ii)) or constitutive (NarP-1-(iii) and (iv)), were incorporated into the pcBAC-Ptac-GFP plasmid. GFP serves as a cargo protein, with the *narP* transcription factor used in constructs (i) and (iii) and *napABCDF* genes in (ii) and (iv). (**D**) Growth and relative fluorescence units (RFU) in engineered NarP-1 iM strains. The strains were cultured in chloramphenicol-supplemented LBv2 in a 96-well plate, and GFP signals were measured for protein synthesis. Ptac and Ptac-GFP strains contain the pcBAC-Ptac and pcBAC-Ptac-GFP plasmids, respectively. The significance of RFU differences between WT (green) and NarP-1-(iv) (blue) under 1 mM IPTG induction was evaluated using Student's *t*-tests (*$P < 0.05$, **$P < 0.01$). Data represent mean ± standard deviation (SD) from four biological replicates. (E, F) GFP expression in controls and engineered strains based on (**E**) iMs or (**F**) existing knowledge at the flask level. GFP signals were measured at stationary phase (OD_{600} 3.6–3.9) and normalized to cell density (OD_{600}) in 250 ml flasks with 50 ml working volume. Ptac and Ptac-GFP strains contain the pcBAC-Ptac and pcBAC-Ptac-GFP plasmids, respectively. The significance of cellular GFP signal (RFU/OD_{600}) between control and engineered strains under IPTG induction was determined by Student's *t*-test (*$P < 0.05$, **$P < 0.01$,***$P < 0.001$, ****$P < 0.0001$). Data represent mean ± SD from four biological replicates. (**G**) YfjJ iM activity in *E. coli*. Scatter plot displaying the activity of the YfjJ iM across all conditions in the PRECISE-1K dataset (https://imodulondb.org) (18), highlighting the specific activation of YfjJ iM in the HOT-TALE project, which involved laboratory evolution of an *E. coli* strain to enhance heat tolerance up to 45.3°C (44). The genes within the dashed line represent the gene membership of YfjJ iM. (**H**) Growth rates of control and constitutively activated strains at different temperatures. Bar plot comparing the relative growth rates of *E. coli* with Ptac and Ptac-YfjJ iM at 37°C and 42°C under varying IPTG concentrations. The significance of relative growth rate differences between control and engineered strains under IPTG induction was determined by Student's t-test (*$P < 0.05$, **$P < 0.01$). (**I**) PP_2260–3 iM activity in *P. putida*. Scatter plot showing the activity of the PP_2260–3 iM across all conditions in the *P. putida* PRECISE321 dataset (https://imodulondb.org) (18), emphasizing the influence of fructose M9 on PP_2260–3 iM activity. The genes within the dashed line represent the gene membership of PP_2260–3 iM. (**J**) Growth rates of control, deleted, and constitutively activated strains under M9 fructose conditions. Bar plot comparing the relative growth rates of wild-type and ΔPP_2260–3 strains with or without constitutive PP_2260–3 iM activation under different IPTG concentrations, demonstrating the impact of the mutation and activation of the iM genes for fructose metabolism. The significance was determined by Student's *t*-test (***$P < 0.001$, ****$P < 0.0001$).

(Supplementary Figure S4A–B), we constructed four plasmid variants (Figure 2C, Supplementary Figure S4C). These variants were designed to enable either IPTG-co-regulated or constitutive expression of the *narP* TF and *napABCDF* genes. Notably, strains with constitutive expression of *narP* (NarP-1-(iii)) and *napABCDF* (NarP-1-(iv)) exhibited significant increases in GFP production rate, over 1.15-fold higher than the Ptac-GFP control ($P < 0.002$), as observed at the microplate level (Figure 2D, Supplementary Figure S3K). However, a concurrent decrease in growth rate was noticed (<0.94-fold, $P < 0.002$). At the flask level, these strains demonstrated enhanced GFP yields, up to 2.3-fold greater compared to the control ($P < 0.0002$) (Figure 2E). This increase surpasses the improvements of previous rational engineering (42,43) by co-expression with chaperone proteins such as GroES or DnaJK (up to 1.4-fold, $P < 0.024$) (Figure 2F, Supplementary Figures S3L–M, S4D).

We applied the iModulon Discovery approach using the *E. coli* PRECISE-1K and *P. putida* PRECISE321 iModulonDB as additional examples. In *E. coli*, we identified the unknown YjfJ iM module, which was specifically highly activated in the HOT-TALE project, involving the adaptive evolution of *E. coli* at 45°C (44) (Figure 2G, Supplementary Figures S5A–B). This suggests that the YjfJ iM module plays a significant role in the high-temperature adaptation of *E. coli*. When we artificially activated this module using a plasmid (Supplementary Figure S4E), we observed a significant increase in growth rate (>1.21-fold, $P < 0.021$) at a high temperature of 42°C (Figure 2H, Supplementary Figure S5C).

Similarly, in *P. putida*, we identified the PP_2260–3 iM, composed of genes encoding an unknown transporter activated under M9 fructose conditions (Figure 2I, Supplementary Figure S5D, E). The high activity of the BkdR iM, responsive to starvation response (19), indicates that the given nutrient condition was not sufficient for cell growth (Supplementary Figure S5F). When this iM was deleted, the strain could not grow under fructose conditions (Figure 2J, Supplementary Figure S5G). However, upon activation of this iM, we observed enhanced growth (>1.45-fold, $P < 0.0005$) under the fructose conditions.

These results demonstrate that the iModulon Discovery approach allows us to infer the roles of modules within the cell based on their dynamics, even without detailed knowledge of each gene's function. This approach enables us to either completely remove or activate these modules to repurpose them for desired outcomes, facilitating strain engineering.

iModulon boosting

Static expression engineering and adaptive laboratory evolution aim to optimize strains for specific environments but can lead to issues like unnecessary gene expression and resource allocation trade-offs (23,45–47). These problems are critical under stressful or resource-limited conditions, requiring expression optimization through promoter or RBS libraries (48,49) or more sophisticated designs (50,51).

To address these challenges, We developed an 'iModulon Boosting' method, which amplifies the activity of specific iModulons while preserving their inherent regulatory mechanisms. By leveraging the natural responsiveness of iMs to specific environmental stimuli, iModulon Boosting allows iMs to act as both sensors and actuators, activating specific cell functions under particular conditions without the need for extensive optimization of each component (Figure 3A, Supplementary Figure S6A). This method reduces the risk of unnecessary gene expression and mitigates resource allocation trade-offs by ensuring that only the necessary modules are activated in response to specific environmental conditions. As a result, this approach effectively improves cell growth under salt stress and poor carbon source conditions.

In *V. natriegens*, we focused on enhancing salt stress tolerance by targeting the Low-osmolarity and Ectoine iMs (Figure 3B), which naturally respond to changes in salt concentration. General DEG analysis makes it difficult to identify key gene sets for salt stress adaptation due to the many DEGs (Supplementary Figure S7A). However, iM analysis easily revealed key iMs for low and high salt stress adaptation (Supplementary Figure S7B). Specifically, the Low-osmolarity iM, including putrescine synthesis genes, is activated in low salt environments (0 mM NaCl), whereas the Ectoine iM, including ectoine synthesis genes, operates in high salt conditions (800 mM NaCl). These iMs highlight the unique adaptations of *V. natriegens* as a marine bacterium to various osmotic environments (52).

To implement iM Boosting, we engineered 'boost modules' with additional iM gene copies paired with a signal amplifier, comprising *araC* and *lacI* repressors, strategically placed adjacent to the original iM genes in the genome (Figure 3C, Supplementary Figure S6A, Supplementary Figure S8A, B). This setup allowed co-regulation with the native system. We confirmed the signal amplifier's functionality in both *E. coli* and *V. natriegens* (Supplementary Figure S6B, C). Removing the Arabinose iM further enhanced the circuit's functionality in *V. natriegens* (Supplementary Figure S6D, E).

To enhance iM activities under varying osmolarity conditions, we integrated booster modules (as an actuator module) downstream of respective iM genes (as a sensor module) (Figure 3C). This resulted in the development of LOiM-Boost and EctiM-Boost strains, incorporating a booster module following their respective iM genes (Figure 3C, Supplementary Figure S6F–G). These strains showed significant improvements in salt stress tolerance under low (0 mM NaCl) and high (800 mM NaCl) salt conditions, as evidenced by increased growth rates (>1.43-fold, $P < 0.0011$) and shorter lag times (<0.46-fold, $P < 0.0022$) compared to control strains (Figure 3D, E). Compared to constitutively expressed genes using strong synthetic promoters (Supplementary Figure S7C–E), this iModulon Boosting approach avoids growth retardation under normal conditions (200 mM NaCl) and effectively enhances cell growth under low and high salt conditions. The results demonstrate that actuator module activation was maximized under specific salt stress conditions.

This strategy was also applied to poor carbon conditions in *E. coli*. Unlike good carbon sources such as glucose, low-glycolytic flux states activate the Acetate iM, promoting acetate reuse as a beneficial nutrient, and enhancing cell growth (53,54) (Figure 3F, Supplementary Figure S9A). Analysis of *E. coli* PRECISE 1K iModulonDB under single nutrient perturbation conditions in M9 medium showed that the Acetate iM activity indicates low-glycolytic flux states when using acetate, glycerol, or fructose (Figure 3G, Supplementary Figure S9B). Using the Acetate iM genes as a sensor module for low-glycolytic flux in poor carbon sources, we constructed a boosting module and integrated it into the genome using REXER (29–31) (Supplementary Figures S6H, S8C–D, S9C). The AceiM-Boost strain contains the genome-integrated

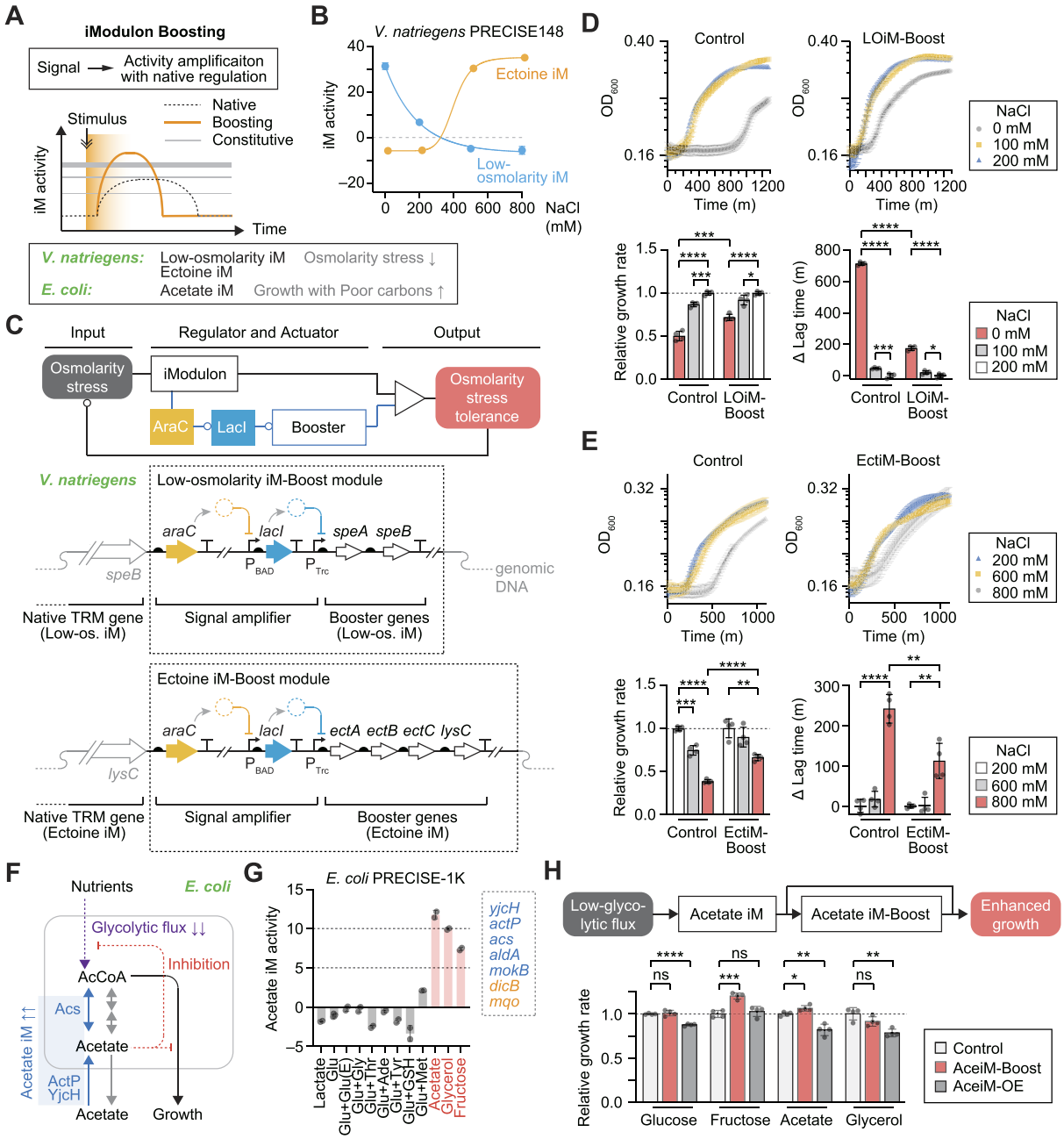


Figure 3. iModulon Boosting to enhance the growth under stressful conditions in *V. natriegens* and *E. coli*. (**A**) iModulon Boosting: transient elevation iM activity through native regulation by incorporating a signal amplifier. The activity of Low-osmolarity and Ectoine iModulons were boosted in *V. natriegens*, and the activity of Acetate iModulon was boosted in *E. coli*. (**B**) iM activities in osmolarity stress conditions show quantitative activation of the Low-osmolarity and Ectoine iMs at low and high salt concentrations, respectively. (**C**) Boosting circuit design and genomic integration. The top panel shows the boosting circuit design. The bottom panel illustrates the genetic integration of the booster module into the genome. The Low-osmolarity iM-Boost module is integrated following the native Low-osmolarity iM gene (*speB*), resulting in the LOiM-Boost strain. The Ectoine iM-Boost module (EctiM-Boost) is added following the native Ectoine iM gene (*lysC*), resulting in the EctiM-Boost strain. These modules are activated under corresponding osmotic stress conditions. For clarity, except for *araC* and its RBS, all constructs are depicted on the same strand as araC, though actual integration is the opposite. (D, E) Growth enhancement of LOiM-Boost (**D**) and EctiM-Boost (**E**) strains. Top: Growth profiles under low salt concentrations (0–200 mM) and high salt concentrations (200–800 mM). Bottom: Growth rate and lag time comparisons. Control strains express ampicillin and spectinomycin resistance genes under the same promoter as the LOiM-Boost and EctiM-Boost strains, integrated into the arabinose catabolism and endA genes, respectively. Growth rate and lag time are relative values for each strain compared to the control condition (200 mM NaCl). Data are mean ± SD from four biological replicates. The significance was determined by Student's *t*-test (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$). (**F**) Schematic diagram illustrating the beneficial function of the Acetate iModulon in central carbon metabolism at low glycolytic flux states. A red dashed line indicates acetate-induced inhibition of glycolysis and TCA cycle enzymes at high glycolytic flux. Acetate is utilized as a beneficial nutrient at a low glycolytic flux (53). (**G**) Activity of the Acetate iModulon under single nutrient variant samples in *E. coli*. The genes within the dashed line represent the gene membership of the Acetate iModulon. Glu, glucose; Glu (**E**), glutamic acid; Gly, glycine; Thr, threonine; Ade, adenine; Tyr, tyrosine; GSH, glutathione; Met, methionine. (**H**) Growth performance of *E. coli* with Acetate iM-Boost and Acetate iM-OE modules. Bar plots showing relative growth rates of control, Acetate iM-Boost, and Acetate iM-OE strains on glucose, fructose, acetate, and glycerol as carbon sources in M9 medium. Significant differences are indicated (Student's *t*-test; *$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$), and non-significant differences are indicated (ns).

Acetate iM boost module as an actuator module below the Acetate iM gene *aldA* as a sensor module. The AceiM-OE strain constitutively expresses the Acetate iM genes at the same genomic location without a controller, resulting in constitutive activation of the Acetate iM function. The AceiM-Boost strain showed improved growth rate under fructose (1.20-fold, $P = 0.0002$) and acetate (1.07-fold, $P = 0.016$) or final cell density under glycerol conditions (1.15-fold, $P < 0.0001$), while the AceiM-OE strain exhibited decreased growth rates (<0.88-fold, $P < 0.0019$) under glucose, acetate, and glycerol conditions (Figure 3H, Supplementary Figure S9D).

These results demonstrate that iM knowledge enables precise and effective engineering, enhancing cellular adaptability and response to specific environmental changes through the iModulon Boosting approach.

iModulon Rebalancing

In our third example of iM-based engineering, termed 'iModulon Rebalancing', we aimed to adjust the dynamic range of OxyR iM activity, which plays a pivotal role in the oxidative stress response (Figure 4A). This approach involves fine-tuning the activity levels of specific iModulons to balance cellular functions and mitigate trade-offs that arise from constant activation. By rebalancing iModulon activity, we can enhance stress tolerance while minimizing adverse effects on other cellular processes. We utilized laboratory-evolved *E. coli* and *V. natriegens* strains that had previously adapted to oxidative stress conditions (23). These evolved strains exhibit constitutively active OxyR iM due to mutations in the *oxyR* transcription factor. However, this constant activation comes with side effects: significantly decreased natural competence activity in *V. natriegens* and increased sensitivity to the antibiotic amoxicillin in *E. coli*.

In *V. natriegens*, analysis of the OxyR iM within the nat-PRECISE148 iModulonDB revealed significant differences between the WT and the evolved VN-ALE strain under oxidative stress conditions (0–0.4 mM H_2O_2) (Figure 4B, Supplementary Figure S10A–B). The WT exhibited weak OxyR iM activation, contributing to low oxidative stress tolerance (Figure 4C). Conversely, VN-ALE showed high OxyR iM activity (36.3-fold at 0.4 mM H_2O_2, compared to the WT) due to a mutation in the *oxyR* gene (designated *oxyR**; Supplementary Figure S10C–D), leading to constant activation. This network-level activation is more effective at enhancing stress tolerance than previously known ROS scavenger expression using strong synthetic promoters (55) (Supplementary Figure S10E).

Continuous OxyR iM activation also results in trade-offs, such as reduced growth rates (0.91-fold, $P = 0.0497$) and decreased competence activity (0.05-fold, $P < 0.00002$) (Figure 4D, Supplementary Figure S10F). Our recent study highlighted the importance of maintaining balanced reactive oxygen species (ROS) levels for optimal natural competence activity, as either insufficient or excessive removal of ROS scavengers significantly reduces competence (22). Transformation efficiency (TE) was significantly diminished in both VN-ALE and ΔoxyR strains due to the disrupted OxyR iM activity balance (Figure 4D). This highlights the need to rebalance its activity to manage oxidative stress tolerance effectively while minimizing impacts on native TE (Figure 4E), an essential consideration for utilizing natural competence in scalable synthetic biology applications of *V. natriegens* (56).

To address these issues, we engineered a module combining a mutated *oxyR** TF from VN-ALE with the developed signal amplifier (Figure 4F, Supplementary Figure S11A). Under oxidative stress, the native *oxyR* weakly activates the OxyR iM. Our rebalancing module activates the expression of the mutated *oxyR**, leading to robust OxyR iM activation. In the absence of stress, increased inactive *oxyR* expression, coupled with *lacI*-mediated repression of *oxyR**, effectively deactivates the OxyR iM (Supplementary Figure S11B). We integrated this rebalancing module into the *V. natriegens* genome below of *dps* gene, resulting in the OxyRiM-Rebal-Vn strain (Supplementary Figures S8F, S11C, D). This strain showed significant improvements in oxidative stress tolerance and maintained stable growth rates across both non-oxidative and oxidative stress conditions (Figure 4G). Additionally, it exhibited enhanced TE (11.4-fold, $P = 0.0014$) and decreased cell viability (0.73-fold, $P = 0.032$) compared to the VN-ALE strain under standard competence-inducing conditions (Figure 4H). These results highlight the effective rebalancing of OxyR iM activity in *V. natriegens* (Figure 4E).

In *E. coli*, we observed a trade-off between OxyR iM activation and amoxicillin sensitivity. Analysis of the OxyR iM within the *E. coli* PRECISE-1K iModulonDB revealed differences between WT and the evolved EC-ALE strain (23) (Supplementary Figure S12A, B). The EC-ALE strain consistently showed high OxyR iM activity due to a mutation in the *oxyR* gene (designated *oxyR**; Supplementary Figure S12C, D). We constructed a rebalancing module combining the mutated *oxyR** from EC-ALE with the signal amplifier and integrated it into the *E. coli* genome through the REXER method, resulting in the OxyRiM-Rebal-Ec strain (Figure 4F, Supplementary Figure S11A, S12E). The ALE strain could resist high levels of oxidative stress (5 mM H_2O_2, $P < 0.0001$) but showed high sensitivity to the minimal inhibitory concentration of amoxicillin (10 μM, $P < 0.0001$) (Figure 4I, Supplementary Figure S12F). In contrast, the OxyRiM-Rebal-Ec strain demonstrated strong oxidative stress tolerance at 5 mM H_2O_2 ($P < 0.0001$) without a significant trade-off in amoxicillin sensitivity ($P > 0.18$, compared to WT).

These results demonstrate that by using the iModulon Rebalancing approach, we can minimize trade-offs at the network level and suggest a viable strategy for managing oxidative stress in various contexts.

iModulon Rewiring

In the fourth example of iM-based engineering, we utilized 'iModulon Rewiring' to manage complex cellular systems involving multiple TFs and over 100 genes (Figure 5A). iModulon Rewiring involves modifying the interactions and activities of iMs to achieve desired cellular functions, thereby providing an alternative to the limitations of high-performance bioparts and genetic circuit predictability. By rewiring iMs, we can integrate multiple regulatory modules to modulate cellular function to specific environmental signals. We specifically applied this method to natural competence in *Vibrio*, which relies on an intricate interplay of environmental signals (57) (Figure 5B). Typically, natural competence in *V. natriegens* is induced by cultivating cells to the stationary phase and transferring them to nutrient-depleted ocean sea salt (OSS) media, supplemented with tDNA and induced TfoX master regulator expression to activate the TfoX and QstR regulons (22,25) (Figure 5B).

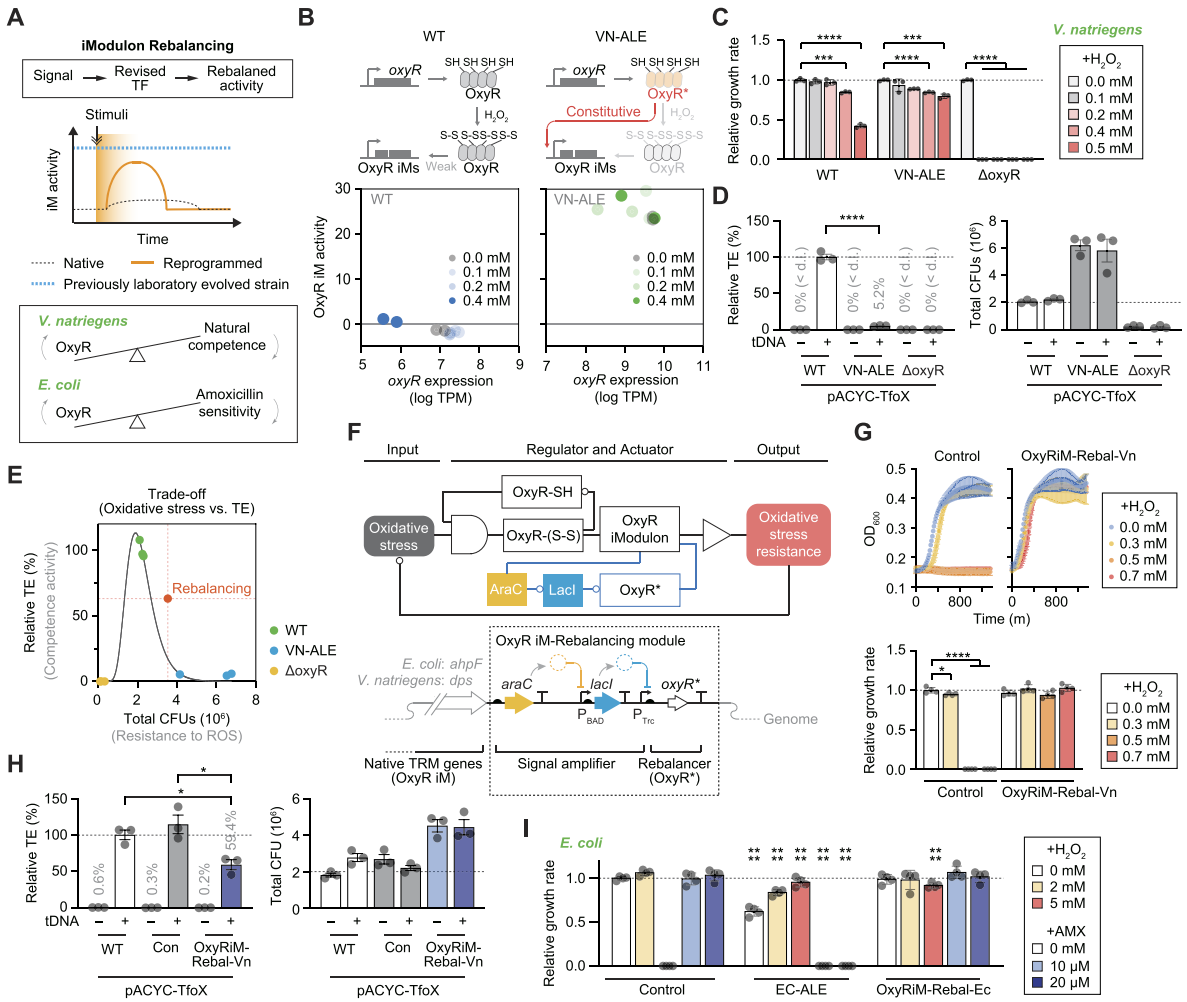


Figure 4. Rebalancing of iModulon activity to minimize the trade-off derived from adaptive laboratory evolution experiments in *E. coli* and *V. natriegens*. (**A**) iModulon Rebalancing: Rebalancing iM activity through integrating a signal amplifier and mutant transcription factor derived from an evolved strain. In the wild-type (WT), iM activity is low and regulated, whereas, in the evolved strain, there is a consistent activation of the OxyR iM due to the mutation in the corresponding transcription factor. This module is applied to *V. natriegens* for rebalancing OxyR iM versus natural competence activity and *E. coli* for rebalancing OxyR iM versus amoxicillin sensitivity. (**B**) OxyR iM activities in WT and evolved *V. natriegens* strains. The top diagram details two states of the OxyR protein in response to oxidative stress, represented by the presence of hydrogen peroxide (H_2O_2). Upon exposure to H_2O_2, free thiol groups (SH) form disulfide bonds (S-S), leading to a conformational change to the OxyR protein, activating it. The activated, oxidized form of OxyR then regulates the expression of various genes involved in OxyR iM. The WT strain exhibits a weak activation at 0.4 mM H_2O_2. In contrast, the VN-ALE strain (23) shows robust expression of *oxyR* and OxyR iM activity under diverse H_2O_2 conditions, indicating dysfunctional negative feedback of OxyR. (**C**) Comparison of the growth rate of WT, VN-ALE and ΔoxyR *V. natriegens* strains under various H_2O_2 concentrations in M9Na medium with 0.5% wt/vol glucose. Relative growth rates were normalized against control conditions (0 mM H_2O_2) for an intuitive explanation. Data are mean ± SD from three biological replicates. Statistical significance was evaluated using Student's *t*-test (***$P < 0.001$; ****$P < 0.0001$). (**D**) Comparison of transformation efficiency (TE) of WT, VN-ALE and ΔoxyR *V. natriegens* strain. TEs of these strains were measured using tDNA-Sm^R, which contains about 2 kb homology of the *endA* gene and spectinomycin resistance gene cassette. The relative TE (rTE) values were normalized against the positive control (WT, pACYC-TfoX). When no transformants were recovered, the TE and rTE values were set to the detection limit and 0%, respectively. Statistical significance was determined by Student's *t*-test (****$P < 0.0001$).< d.l., below detection limit. (**E**) Trade-off between TE and cell viability after natural transformation process. Given the critical need to minimize oxidative stress during the natural competence process (22), the WT TfoX + strain displayed high TE but reduced cell viability (as measured by total CFUs). In contrast, the VN-ALE strain showed lower TE (average 5.2% of WT) but increased viability. The ΔoxyR strain demonstrated no viability in natural transformation. iM rebalancing aims to minimize this trade-off between TE and cell viability (red). (**F**) Rebalancing circuit schematic and genomic integration. Top: Diagram of rebalancing circuit. Bottom: Integration of the OxyR-Rebalancing module following the native OxyR iM gene (*dps*), creating the OxyRiM-Rebal-Vn strain. OxyR* represents mutant OxyR from a previous study (Supplementary Figure S10C) (23). (**G**) Growth of OxyRiM-Rebal-Vn strain. Growth profiles (top) and growth rate comparison (bottom) under various H_2O_2 concentrations (0.0–0.7 mM). The Control strain, expressing ampicillin and spectinomycin resistance genes under the same promoter as the OxyRiM-Rebal-Vn strain, were integrated into arabinose catabolism and *endA* genes, respectively. Specific growth rates compared to the Control strain. Data are mean ± SD from four biological replicates. Statistical significance was determined by Student's t-test (*$P < 0.05$, ****$P < 0.0001$). (**H**) Comparison of TE between WT, Control, and OxyRiM-Rebal-Vn strains. TEs of these strains were measured using tDNA-Km^R, which contains about 2 kb homology of the *endA* gene and kanamycin resistance gene cassette. The rTE values were normalized against the positive control (WT, pACYC-TfoX). Note that *V. natriegens* exhibits low-level inherent resistance to kanamycin, which may result in occasional non-transformant colony formation in negative controls (22). Data are mean ± SD from three biological replicates. Statistical significance was evaluated using Student's *t*-test (*$P < 0.05$). (**I**) Relative growth rates of *E. coli* strains under oxidative stress and antibiotic-treated conditions. Bar plots showing relative growth rates of Control, EC-ALE, and OxyRiM-Rebal-Ec strains under oxidative stress (+H_2O_2 at 2 mM and 5 mM) and antibiotic amoxicillin conditions (+AMX at 10 μM and 20 μM). Significant differences are indicated (****$P < 0.0001$). 10 μM amoxicillin was selected as the previously reported minimal inhibitory concentration in *E. coli* (68).

Our recent work, leveraging the natPRECISE104 iModulonDB (22), highlighted the significance of three competence iMs—TfoX, QstR, and UC-5 (uncharacterized-5) iMs—in transforming DNA uptake during competence, along with the AhpCF (alkyl hydroperoxide reductase) iM's role in oxidative stress management. Updates in the natPRECISE148 iModulonDB revealed minor gene membership adjustments among iMs and the reclassification of the HapR iM (formerly Ttr/Vi iM) and the OxyR iM (formerly AhpCF iM), detailed in Supplementary Note S2 and Supplementary Figure S13. These changes provide a more comprehensive understanding of the regulatory mechanisms behind four key natural competence iMs (Figure 5C).

To investigate the rewiring of these iMs, we constructed BAC vectors incorporating genes from the TfoX and QstR TFs, as well as from UC-1 iM (formerly known as UC-5) genes (Figure 5B), while strategically excluding non-essential genes (22). Initial prototypes included a TQ-v1 strain carrying a v1 rewiring vector with IPTG-induced expression of TfoX and QstR TFs, and a TQUA-v1 strain carrying a v1 rewiring vector with additional arabinose-induced UC-1 iM and *ahpCF* genes (formerly AhpCF iM) (Supplementary Figure S8G, S14A–B). These strains were evaluated for competence activity under both standard (OSS) and growth media (LBv2) conditions (Figure 5D). Notably, under growth conditions, the TQUA-v1 strain exhibited a noticeable, though low, competence activity (average TE = 1.3×10^{-6}) (Supplementary Figure S14C).

Building on these initial results, we refined our rewiring strategy by separating the TfoX and QstR TFs from the same inducible promoter and fine-tuning the arabinose concentrations to better modulate the expression of other competence iMs. This adjustment aimed to align their expression more closely with the TfoX + strain under the standard competence condition (Figure 5C). Additionally, we created a strain with deleted AhpCF genes to achieve ROS balancing during growth conditions. These refinements led to the construction of five distinct v2 rewiring vectors (Supplementary Figure S8H). These vectors were introduced into *V. natriegens*, resulting in V, T, TQ-v2, TQU-v2 and TQUA-v2 strains (Figure 5E).

The TQU-v2 strain demonstrated the highest TE (average TE = 7.0×10^{-6}, 5.3-fold increase compared to the prototype TQUA-V1) at the lowest arabinose concentration (0.001%) under growth conditions (Figure 5F). We validated the designed integration of transforming DNA into the genome (Supplementary Figure S14D). Notably, TQU-v2's TE decreased significantly with increasing arabinose concentrations, as shown in Figure 5F. A similar pattern emerged in the TQU-v2 and TQUA-v2 samples under the standard competence condition, where forced expression of genes other than TfoX led to reduced TE (Supplementary Figure S14E). These results emphasize the need to finely tune each iM's activity levels to optimize performance under both growth and standard competence conditions. Furthermore, the TQ-v2 and TQUA-v2 strains showed significantly lower TEs (0.05-fold, $P < 0.0029$) than TQU-v2 at 0.001% arabinose concentration. These observations, in line with previous research (22), highlight the importance of expressing essential and quasi-essential genes within the UC-1 iM and maintaining a balanced ROS level for effective natural transformation in *V. natriegens*.

To better understand competence iM expression dynamics under growth conditions, we conducted RNA-Seq analysis on five rewired strains at two arabinose concentrations (0.001% and 0.1%). This analysis elucidated the superior TE of TQU-v2 at the 0.001% arabinose condition, where there was significant activation of TfoX and UC-1 iM genes, closely mirroring the activity profile of the TfoX + strain under the standard condition (Figure 5G, Supplementary Figure S14F). Additionally, the RNA-Seq results revealed that higher arabinose concentrations in growth media led to increased *ahpCF* expression, resulting in reduced TfoX iM activity (0.72-fold, $P < 0.0001$ in V strain). Intense activation of UC-1 iM genes, either alone or combined with *ahpCF* genes in TQU-v2 and TQUA-v2 strains, significantly decreased the expression of both TfoX (<0.49-fold, $P < 0.0001$) and QstR (<0.77-fold, $P < 0.013$) iMs compared to T and TQ-v2 (Figure 5G, Supplementary Figure S14F). This pattern underscores the critical need for balanced protein expression, considering proteome allocation constraints (58). These insights pave the way for further optimization of rewiring natural competence-related iMs, enhancing competence activity for synthetic biology applications under diverse growth conditions.

Discussion

Strain engineering typically follows the DBTL (Design, Build, Test, Learn) cycle, beginning with identifying target genes and designing genetic circuits (59). The iM knowledge significantly enhances this process by providing valuable insights for 1) target gene selection, 2) pre-estimation of design, and 3) network-level circuit design, thereby improving scalability and predictability (Supplementary Note S1, Supplementary Figure S1, and Supplementary Table S1). This work is the first to demonstrate the regulation or enhancement of specific functions within strains based on iModulon information, highlighting the novel application of iMs in strain engineering. We validate these advantages through four iM-based engineering approaches in *E. coli*, *V. natriegens*, and *P. putida*.

The iM-based approach offers a scalable module unit for genetic circuit or strain engineering compared to traditional gene-centric designs (Supplementary Note S1). This scalability arises from the ability to modularize genes related to specific cellular functions, facilitating independent module selection. This contrasts with conventional methods like genome-scale metabolic simulations (60,61) and computational pathway design (62), which rely heavily on gene annotations (63). The iM-based process identifies designs more likely to succeed, reduces the number of circuit variants required for testing, and accelerates the design process. Additionally, this approach extends synthetic biology applications beyond a few model strains, such as *E. coli* and *B. subtilis* to non-model organisms by leveraging ICA to identify modules regardless of existing knowledge constraints. This capability allowed us to engineer modules even for genes with unknown functions (iModulon Discovery). Employing iMs for module definition has been key in deciphering and managing complex regulatory networks (iModulon Rewiring). This approach simplifies genome-scale TRN reconstruction and condenses complex networks into manageable sets of iMs. For instance, in *V. natriegens*, we condensed natural competence networks into 189 genes, including 102 previously uncharacterized functions (22). The 'iModulon Rewiring' strategy validated the significance of TfoX, QstR, and UC-1 iMs in activating natural competence under typical growth conditions.

In terms of predictability, iMs, as naturally evolved genetic circuits, exhibit reliable and predictable functions. Predicting the function of arbitrarily assembled synthetic genetic

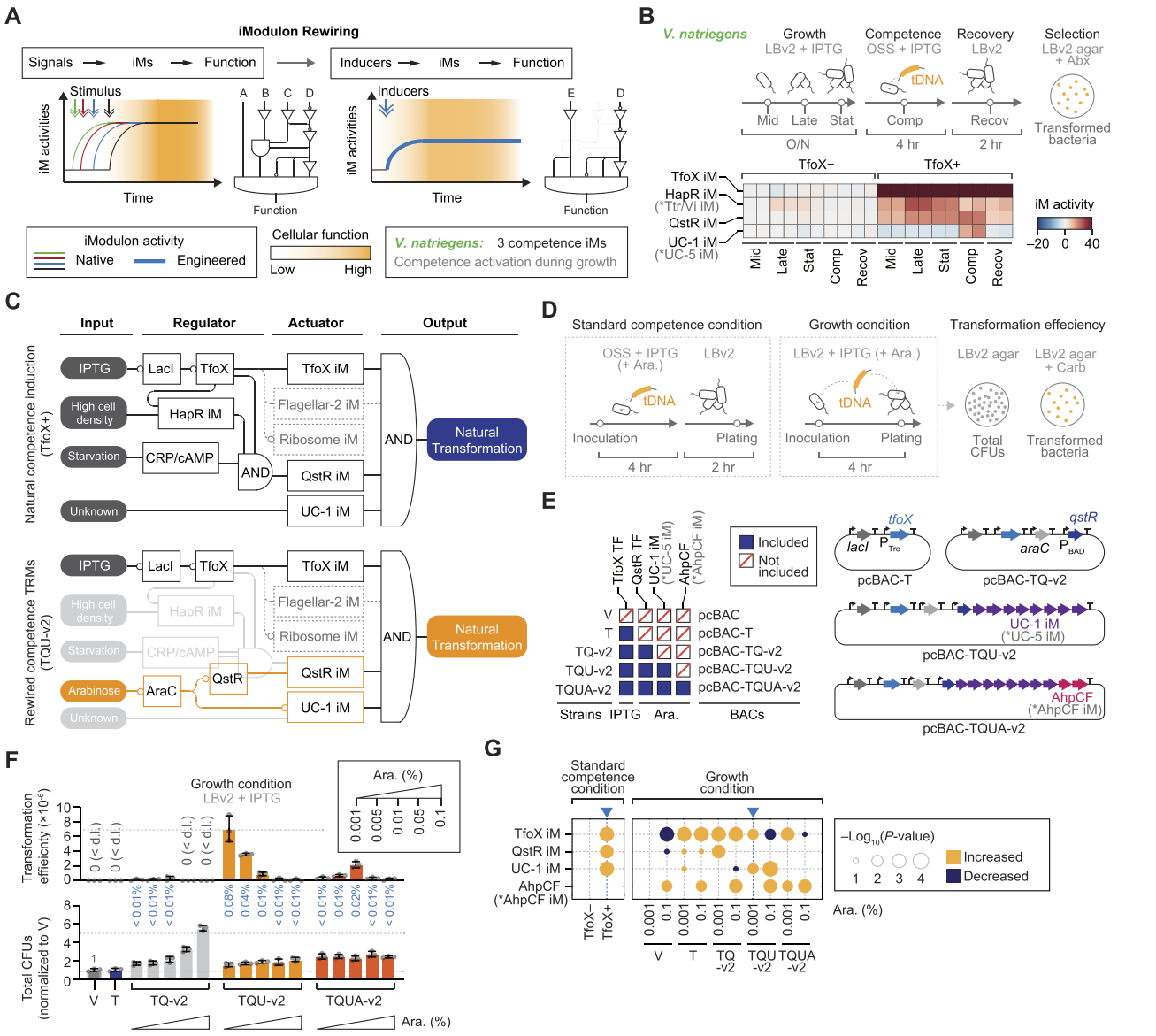


Figure 5. Activation of natural competence during growth through rewiring competence-related iModulons. (**A**) iModulon Rewiring: Rewiring multiple iMs, generally activated by specific stimuli, to forcefully initiate cellular functions. The output of the system, initially complex, is streamlined in the rewired system. (**B**) The activity of competence-specific iMs. This section displays sample stages during cell growth, natural competence, and recovery phases, established in previous studies (22,25). Four competence iMs are specifically activated during these stages. (**C**) Natural vs. rewired competence activation: The upper pathway illustrates natural activation, dependent on environmental cues such as TfoX overexpression, high cell density, and starvation (22,25). The lower pathway depicts the rewired strain (TQU-v2) using arabinose and IPTG to induce competence via AraC and LacI regulators. Dotted lines denote regulatory interactions; solid lines with arrows indicate direct activations. (**D**) Natural transformation under standard competence and growth conditions. For growth-condition competence activation, parameters such as inoculation amount, incubation time, working volume, and temperature are consistent with the standard competence condition. tDNA-Amp^R, containing ~2 kb *endA* gene homology and an ampicillin resistance gene cassette, and IPTG and arabinose, are utilized to induce competence. After tDNA uptake, bacteria are plated on LBv2 agar with or without carbenicillin to calculate transformation efficiency (TE). (**E**) Engineered vectors and strains for competence iM rewiring. The plot details rewired competence strains, including the harbored plasmids (left panel). Each plasmid encompasses known iM TFs or iM genes (if TFs are unknown). BAC constructs for iM rewiring (right panel). Constructs employ IPTG to trigger the P_{Trc} promoter for *tfoX* expression, while arabinose activates the P_{BAD} promoter for *qstR*, UC-1 iM (formerly known as UC-5 iM in natPRECISE 104), and *ahpCF* genes (previously AhpCF iM in natPRECISE 104, now part of OxyR iM in natPRECISE 148; see detail in Supplementary Figure S13a. (**F**) TE and viability in rewired strains under growth conditions. TE was determined using tDNA-Amp^R. Results from similar experiments under the standard competence condition are in Supplementary Figure S14E. Relative TE (rTE, %) is normalized against the native TE (T strain under the standard competence condition, Supplementary Figure S14C) and presented in blue for clarity. Data are mean ± SD from three biological replicates. (**G**) Dot plot analysis of competence-related iMs expression in rewired strains under the growth condition. The size of each dot reflects the statistical significance, while the color denotes the change in expression levels (log_2TPM). The control dataset shows the activation of competence iMs in the TfoX+ strain compared to the TfoX– strain under the standard competence condition (left panel). In contrast, the RNA-Seq data reveal the activation of competence iMs in rewired strains by comparing them to the control V strain (with 0.001% arabinose) under growth conditions (right panel). For comprehensive log_2TPM values and significance levels, see Supplementary Figure S14F. Conditions leading to the highest TE are highlighted with a blue triangle. An asterisk (*) denotes iM names previously identified in natPRECISE104 (Supplementary Figure S13A). Abbreviations: < d.l. (below detection limit), tDNA (transforming DNA), OSS (ocean sea salt medium), Comp (competence condition), and Recov (recovery condition).

circuits is often challenging (64–66), and such circuits often interact with the host in multiple unintended ways (3,4,6). However, iMs can replace traditional genetic circuit components such as sensor, transduction, and actuator modules in synthetic circuits (Supplementary Figures S2A–B). Various visualizations, such as iM activity plots and differential activity plots, illustrate iM levels and interactions, providing insights into potential conflicts and synergies between modules (Supplementary Note S1 and Supplementary Figure S1). Strategies such as iModulon Boosting and iModulon Rebalancing minimize reliance on synthetic circuits while enhancing native regulation. For instance, iModulon Boosting amplified iM activity under only specific conditions, showcasing adaptability and minimizing trade-offs. Rebalancing the OxyR iM achieved increased oxidative stress tolerance while minimizing side effects like reduced competence activity in *V. natriegens* or increased antibiotic sensitivity in *E. coli*. The iModulon Rebalancing strategy offers refined control, addressing trade-offs associated with TF mutations reported in studies, including ALEdb (85). These examples highlight the practicality of our approach and its potential for broader applications in cellular function engineering.

However, this method requires high-quality, large datasets (typically over 50 conditions), which can be resource-intensive to generate. Currently, iM data for 14 widely used model species (totaling 17 datasets and 1762 iMs) are publicly available in the iModulonDB (18). Additionally, researchers can generate iModulonDBs using their own private datasets by following the publicly available workflows (18,24), allowing them to analyze and visualize iMs to better understand each module and regulatory network for even non-model strains. However, one limitation is that this method is currently not suitable for eukaryotic RNA-Seq data due to its complexity.

In summary, this study presents a scalable and efficient strategy for the predictable reprogramming of cellular functions using iMs. By leveraging machine learning for module-level engineering, this approach simplifies access to genes, enhances high-level cellular function annotation, and significantly improves our understanding of these processes, thereby streamlining engineering efforts. The iM-based strategy greatly enhances the predictability and efficiency of cellular engineering, marking a substantial advancement in navigating the complexities of genetic circuits. Moreover, it offers potential for cross-species iM function transfer (67), highlighting its broad applicability and versatility in synthetic biology.

Data availability

The RNA-Seq data generated for this article are available in NCBI SRA at https://www.ncbi.nlm.nih.gov/sra and can be accessed with PRJNA1059255. Analysis pipelines for iModulons are available at the iModulonMiner (https://github.com/sbrg/imodulonminer, also archived at https://doi.org/10.5281/zenodo.13273405) and Pymodulon (https://pymodulon.readthedocs.io/). Metadata and iModulon data of PRECISE-1K (*E. coli*), natPRECISE148 (*V. natriegens*), and putidaPRECISE321 (*P. putida*) used in this study are publicly available at the iModulonDB web portal (https://imodulondb.org/).

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We would like to thank Marc Abrams (Systems Biology Research Group, University of California San Diego) for assistance with paper editing.

Author contributions: Jongoh Shin: Conceptualization, Formal Analysis, Methodology, Visualization, Validation, Writing – original draft. Daniel C. Zielinski: Formal Analysis, Software, Project administration, Writing—review & editing. Bernhard O. Palsson: Supervision, Formal Analysis, Funding acquisition, Writing—review & editing.

Funding

Novo Nordisk Foundation (NNF) Center for Biosustainability (CfB) at the Technical University of Denmark [NNF20CC0035580]; Y.C. Fung Endowed Chair in Bioengineering at the University of California San Diego. Funding for open access charge: Novo Nordisk Foundation (NNF) Center for Biosustainability (CfB) at the Technical University of Denmark.

Conflict of interest statement

None declared.

References

1. Benner,S.A. and Sismour,A.M. (2005) Synthetic biology. *Nat. Rev. Genet.*, **6**, 533–543.
2. Voigt,C.A. and Keasling,J.D. (2005) Programming cellular function. *Nat. Chem. Biol.*, **1**, 304–307.
3. Ilia,K. and Del Vecchio,D. (2022) Squaring a circle: to what extent are traditional circuit analogies impeding synthetic biology?*Gen. Biotechnol.*, **1**, 150–155.
4. Şimşek,E., Yao,Y., Lee,D. and You,L. (2023) Toward predictive engineering of gene circuits. *Trends Biotechnol.*, **41**, 760–768.
5. Sánchez-Osorio,I., Hernández-Martínez,C.A. and Martínez-Antonio,A. (2020) Quantitative modeling of the interplay between synthetic gene circuits and host physiology: experiments, results, and prospects. *Curr. Opin. Microbiol.*, **55**, 48–56.
6. Grunberg,T.W. and Del Vecchio,D. (2020) Modular analysis and design of biological circuits. *Curr. Opin. Biotechnol.*, **63**, 41–47.
7. Moon,T.S., Lou,C., Tamsir,A., Stanton,B.C. and Voigt,C.A. (2012) Genetic programs constructed from layered logic gates in single cells. *Nature*, **491**, 249–253.
8. Appleton,E., Madsen,C., Roehner,N. and Densmore,d.,D. (2017) Design automation in synthetic biology. *Cold Spring Harb. Perspect. Biol.*, **9**, a023978.
9. Karr,J.R., Sanghvi,J.C., Macklin,D.N., Gutschow,M.V., Jacobs,J.M., Bolival,B. Jr, Assad-Garcia,N., Glass,J.I. and Covert,M.W. (2012) A whole-cell computational model predicts phenotype from genotype. *Cell*, **150**, 389–401.
10. Scott,M., Gunderson,C.W., Mateescu,E.M., Zhang,Z. and Hwa,T. (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science*, **330**, 1099–1102.
11. Weiße,A.Y., Oyarzún,D.A., Danos,V. and Swain,P.S. (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E1038–E1047.
12. Liao,C., Blanchard,A.E. and Lu,T. (2017) An integrative circuit-host modelling framework for predicting synthetic gene network behaviours. *Nat. Microbiol.*, **2**, 1658–1666.
13. Meyer,A.J., Segall-Shapiro,T.H., Glassey,E., Zhang,J. and Voigt,C.A. (2019) Escherichia coli 'Marionette' strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol.*, **15**, 196–204.

14. Ceroni,F., Boo,A., Furini,S., Gorochowski,T.E., Borkowski,O., Ladak,Y.N., Awan,A.R., Gilbert,C., Stan,G.-B. and Ellis,T. (2018) Burden-driven feedback control of gene expression. *Nat. Methods*, **15**, 387–393.
15. Huang,H.-H., Qian,Y. and Del Vecchio,D. (2018) A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nat. Commun.*, **9**, 5415.
16. Sastry,A.V., Hu,A., Heckmann,D., Poudel,S., Kavvas,E. and Palsson,B.O. (2021) Independent component analysis recovers consistent regulatory signals from disparate datasets. *PLoS Comput. Biol.*, **17**, e1008647.
17. Sastry,A.V., Gao,Y., Szubin,R., Hefner,Y., Xu,S., Kim,D., Choudhary,K.S., Yang,L., King,Z.A. and Palsson,B.O. (2019) The Escherichia coli transcriptome mostly consists of independently regulated modules. *Nat. Commun.*, **10**, 5536.
18. Rychel,K., Decker,K., Sastry,A.V., Phaneuf,P.V., Poudel,S. and Palsson,B.O. (2021) iModulonDB: a knowledgebase of microbial transcriptional regulation derived from machine learning. *Nucleic Acids Res.*, **49**, D112–D120.
19. Lim,H.G., Rychel,K., Sastry,A.V., Bentley,G.J., Mueller,J., Schindel,H.S., Larsen,P.E., Laible,P.D., Guss,A.M., Niu,W., *et al.* (2022) Machine-learning from Pseudomonas putida KT2440 transcriptomes reveals its transcriptional regulatory network. *Metab. Eng.*, **72**, 297–310.
20. Rajput,A., Tsunemoto,H., Sastry,A.V., Szubin,R., Rychel,K., Sugie,J., Pogliano,J. and Palsson,B.O. (2022) Machine learning from Pseudomonas aeruginosa transcriptomes identifies independently modulated sets of genes associated with known transcriptional regulators. *Nucleic Acids Res.*, **50**, 3658–3672.
21. Rychel,K., Sastry,A.V. and Palsson,B.O. (2020) Machine learning uncovers independently regulated modules in the Bacillus subtilis transcriptome. *Nat. Commun.*, **11**, 6338.
22. Shin,J., Rychel,K. and Palsson,B.O. (2023) Systems biology of competency in Vibrio natriegens is revealed by applying novel data analytics to the transcriptome. *Cell Rep.*, **42**, 112619.
23. Anand,A., Chen,K., Catoiu,E., Sastry,A.V., Olson,C.A., Sandberg,T.E., Seif,Y., Xu,S., Szubin,R., Yang,L., *et al.* (2020) OxyR is a convergent target for mutations acquired during adaptation to oxidative stress-prone metabolic states. *Mol. Biol. Evol.*, **37**, 660–667.
24. Sastry,A.V., Poudel,S., Rychel,K., Yoo,R., Lamoureux,C.R., Chauhan,S., Haiman,Z.B., Al Bulushi,T., Seif,Y. and Palsson,B.O. (2021) Mining all publicly available expression data to compute dynamic microbial transcriptional regulatory networks. bioRxiv doi: https://doi.org/10.1101/2021.07.01.450581, 02 July 2021, preprint: not peer reviewed.
25. Dalia,T.N., Hayes,C.A., Stolyar,S., Marx,C.J., McKinlay,J.B. and Dalia,A.B. (2017) Multiplex genome editing by natural transformation (MuGENT) for synthetic biology in Vibrio natriegens. *ACS Synth. Biol.*, **6**, 1650–1655.
26. Thompson,M.G., Blake-Hedges,J.M., Cruz-Morales,P., Barajas,J.F., Curran,S.C., Eiben,C.B., Harris,N.C., Benites,V.T., Gin,J.W., Sharpless,W.A., *et al.* (2019) Massively parallel fitness profiling reveals multiple novel enzymes in Pseudomonas putida Lysine metabolism. *mBio*, **10**, 17.
27. Jana,B., Keppel,K. and Salomon,D. (2021) Engineering a customizable antibacterial T6SS-based platform in Vibrio natriegens. *EMBO Rep.*, **22**, e53681.
28. Gietz,R.D. and Schiestl,R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, **2**, 31–34.
29. Wang,K., Fredens,J., Brunner,S.F., Kim,S.H., Chia,T. and Chin,J.W. (2016) Defining synonymous codon compression schemes by genome recoding. *Nature*, **539**, 59–64.
30. Robertson,W.E., Funke,L.F.H., de la Torre,D., Fredens,J., Wang,K. and Chin,J.W. (2021) Creating custom synthetic genomes in Escherichia coli with REXER and GENESIS. *Nat. Protoc.*, **16**, 2345–2380.
31. Zürcher,J.F., Kleefeldt,A.A., Funke,L.F.H., Birnbaum,J., Fredens,J., Grazioli,S., Liu,K.C., Spinck,M., Petris,G., Murat,P., *et al.* (2023) Continuous synthesis of E. coli genome sections and mb-scale human DNA assembly. *Nature*, **619**, 555–562.
32. Wirth,N.T., Funk,J., Donati,S. and Nikel,P.I. (2023) QurvE: user-friendly software for the analysis of biological growth and fluorescence data. *Nat. Protoc.*, **18**, 2401–2403.
33. Choe,D., Szubin,R., Poudel,S., Sastry,A., Song,Y., Lee,Y., Cho,S., Palsson,B. and Cho,B.-K. (2021) RiboRid: a low cost, advanced, and ultra-efficient method to remove ribosomal RNA for bacterial transcriptomics. *PLoS Genet.*, **17**, e1009821.
34. Lee,H.H., Ostrov,N., Wong,B.G., Gold,M.A., Khalil,A.S. and Church,G.M. (2019) Functional genomics of the rapidly replicating bacterium vibrio natriegens by CRISPRi. *Nat. Microbiol.*, **4**, 1105–1113.
35. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
36. Wang,L., Wang,S. and Li,W. (2012) RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, **28**, 2184–2185.
37. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
38. Ewels,P., Magnusson,M., Lundin,S. and Käller,M. (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**, 3047–3048.
39. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
40. Battesti,A., Majdalani,N. and Gottesman,S. (2011) The RpoS-mediated general stress response in Escherichia coli. *Annu. Rev. Microbiol.*, **65**, 189–213.
41. Stewart,V. (1993) Nitrate regulation of anaerobic respiratory gene expression in Escherichia coli. *Mol. Microbiol.*, **9**, 425–434.
42. Makino,T., Skretas,G. and Georgiou,G. (2011) Strain engineering for improved expression of recombinant proteins in bacteria. *Microb. Cell Fact.*, **10**, 32.
43. Martínez-Alonso,M., García-Fruitós,E., Ferrer-Miralles,N., Rinas,U. and Villaverde,A. (2010) Side effects of chaperone gene co-expression in recombinant protein production. *Microb. Cell Fact.*, **9**, 64.
44. Rychel,K., Chen,K., Catoiu,E.A., Olson,C.A., Sandberg,T.E., Gao,Y., Xu,S., Hefner,Y., Szubin,R., Patel,A., *et al.* (2024) Laboratory evolution reveals transcriptional mechanisms underlying thermal adaptation of Escherichia coli. bioRxiv doi: https://doi.org/10.1101/2024.02.22.581624, 23 February 2024, preprint: not peer reviewed.
45. Wytock,T.P., Fiebig,A., Willett,J.W., Herrou,J., Fergin,A., Motter,A.E. and Crosson,S. (2018) Experimental evolution of diverse Escherichia coli metabolic mutants identifies genetic loci for convergent adaptation of growth rate. *PLoS Genet.*, **14**, e1007284.
46. Rychel,K., Tan,J., Patel,A., Lamoureux,C., Hefner,Y., Szubin,R., Johnsen,J., Mohamed,E.T.T., Phaneuf,P.V., Anand,A., *et al.* (2023) Laboratory evolution, transcriptomics, and modeling reveal mechanisms of paraquat tolerance. *Cell Rep.*, **42**, 113105.
47. Horinouchi,T., Tamaoka,K., Furusawa,C., Ono,N., Suzuki,S., Hirasawa,T., Yomo,T. and Shimizu,H. (2010) Transcriptome analysis of parallel-evolved Escherichia coli strains under ethanol stress. *Bmc Genomics [Electronic Resource]*, **11**, 579.
48. Hammer,K., Mijakovic,I. and Jensen,P.R. (2006) Synthetic promoter libraries–tuning of gene expression. *Trends Biotechnol.*, **24**, 53–55.
49. Rao,X., Li,D., Su,Z., Nomura,C.T., Chen,S. and Wang,Q. (2024) A smart RBS library and its prediction model for robust and accurate fine-tuning of gene expression in Bacillus species. *Metab. Eng.*, **81**, 1–9.

50. English,M.A., Gayet,R.V. and Collins,J.J. (2021) Designing biological circuits: synthetic biology within the operon model and beyond. *Annu. Rev. Biochem.*, **90**, 221–244.
51. Moser,F., Espah Borujeni,A., Ghodasara,A.N., Cameron,E., Park,Y. and Voigt,C.A. (2018) Dynamic control of endogenous metabolism with combinatorial logic circuits. *Mol. Syst. Biol.*, **14**, e8605.
52. Gregory,G.J. and Boyd,E.F. (2021) Stressed out: bacterial response to high salinity using compatible solute biosynthesis and uptake systems, lessons from Vibrionaceae. *Comput. Struct. Biotechnol. J.*, **19**, 1014–1027.
53. Millard,P., Gosselin-Monplaisir,T., Uttenweiler-Joseph,S. and Enjalbert,B. (2023) Acetate is a beneficial nutrient for E. coli at low glycolytic flux. *EMBO J.*, **42**, e113079.
54. Shin,J., Zielinski,D.C. and Palsson,B.O. (2024) Deciphering nutritional stress responses via knowledge-enriched transcriptomics for microbial engineering. *Metab. Eng.*, **84**, 34–47.
55. Weinstock,M.T., Hesek,E.D., Wilson,C.M. and Gibson,D.G. (2016) Vibrio natriegens as a fast-growing host for molecular biology. *Nat. Methods*, **13**, 849–851.
56. Specht,D.A., Sheppard,T.J., Kennedy,F., Li,S., Gadikota,G. and Barstow,B. (2023) Efficient natural plasmid transformation of Vibrio natriegens enables zero-capital molecular biology. bioRxiv doi: https://doi.org/10.1101/2023.08.11.553013, 13 August 2023, preprint: not peer reviewed.
57. Matthey,N. and Blokesch,M. (2016) The DNA-uptake process of naturally competent vibrio cholerae. *Trends Microbiol.*, **24**, 98–110.
58. Scott,M. and Hwa,T. (2023) Shaping bacterial gene expression by physiological and proteome allocation constraints. *Nat. Rev. Microbiol.*, **21**, 327–342.
59. Kitano,S., Lin,C., Foo,J.L. and Chang,M.W. (2023) Synthetic biology: learning the way toward high-precision biological design. *PLoS Biol.*, **21**, e3002116.
60. Lewis,N.E., Nagarajan,H. and Palsson,B.O. (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.*, **10**, 291–305.
61. Ko,Y.-S., Kim,J.W., Lee,J.A., Han,T., Kim,G.B., Park,J.E. and Lee,S.Y. (2020) Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chem. Soc. Rev.*, **49**, 4615–4636.
62. Kim,D.I., Chae,T.U., Kim,H.U., Jang,W.D. and Lee,S.Y. (2021) Microbial production of multiple short-chain primary amines via retrobiosynthesis. *Nat. Commun.*, **12**, 173.
63. Ghatak,S., King,Z.A., Sastry,A. and Palsson,B.O. (2019) The y-ome defines the 35% of Escherichia coli genes that lack experimental evidence of function. *Nucleic Acids Res.*, **47**, 2446–2454.
64. Lu,T.K., Khalil,A.S. and Collins,J.J. (2009) Next-generation synthetic gene networks. *Nat. Biotechnol.*, **27**, 1139–1150.
65. Purnick,P.E.M. and Weiss,R. (2009) The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.*, **10**, 410–422.
66. Kwok,R. (2010) Five hard truths for synthetic biology. *Nature*, **463**,288–290.
67. Choe,D., Olson,C.A., Szubin,R., Yang,H., Sung,J., Feist,A.M. and Palsson,B.O. (2024) Advancing the scale of synthetic biology via cross-species transfer of cellular functions enabled by iModulon engraftment. *Nat. Commun.*, **15**, 2356.
68. Hoeksema,M., Brul,S. and Ter Kuile,B.H. (2018) Influence of reactive oxygen species on de novo acquisition of resistance to bactericidal antibiotics. *Antimicrob. Agents Chemother.*, **62**,e02354-17.

Received: February 22, 2024. **Revised:** July 29, 2024. **Editorial Decision:** August 7, 2024. **Accepted:** August 19, 2024