# Predicting Evolutionary Constraints by Identifying Conflicting Demands in Regulatory Networks

Kogenaru, Manjunatha; Nghe, Philippe; Poelwijk, Frank J.; Tans, Sander J.

**DOI**
10.1016/j.cels.2020.05.004

**Publication date**
2020

**Document Version**
Final published version

**Published in**
Cell Systems

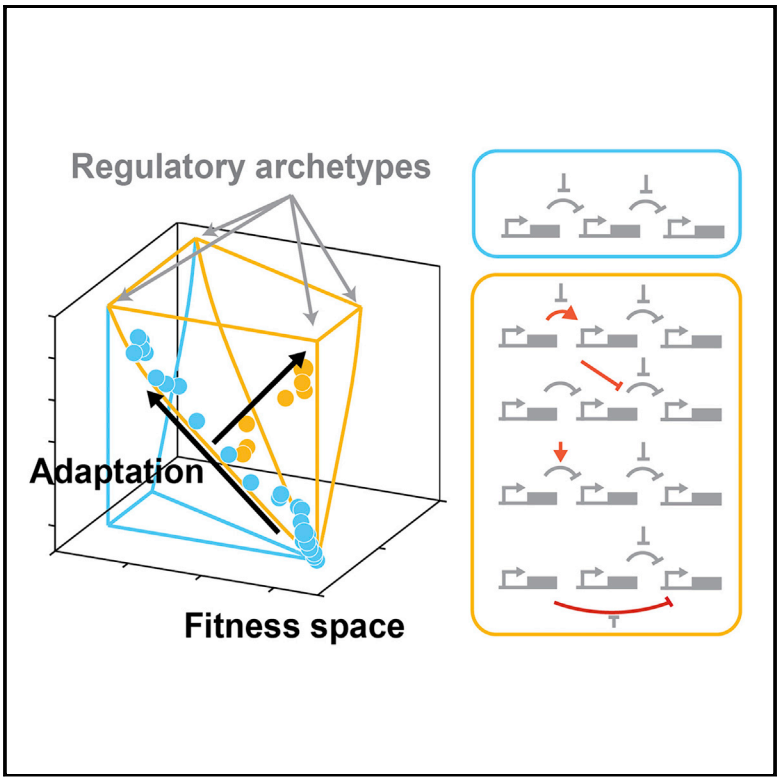
**Citation (APA)**
Kogenaru, M., Nghe, P., Poelwijk, F. J., & Tans, S. J. (2020). Predicting Evolutionary Constraints by Identifying Conflicting Demands in Regulatory Networks. *Cell Systems*, *10*(6), 526-534. https://doi.org/10.1016/j.cels.2020.05.004

Cell Systems

Report

# Predicting Evolutionary Constraints by Identifying Conflicting Demands in Regulatory Networks

## Graphical Abstract




## Highlights

- Genetic networks evolve by tuning interactions first, then by topological innovations
- A network's evolutionary constraints are predictable without genetic information
- A network's evolutionary constraints are predictable without molecular information
- "Partial order" rapidly identifies evolutionary constraints and Pareto front dimensions

## Authors

Manjunatha Kogenaru, Philippe Nghe, Frank J. Poelwijk, Sander J. Tans

## Correspondence

philippe.nghe@espci.psl.eu (P.N.), tans@amolf.nl (S.J.T.)

## In Brief

To predict evolutionary constraints in regulatory networks, we developed a new network approach based on partial order and verified it by evolving a genetic network in *Escherichia coli* in variable environments.

Kogenaru et al., 2020, Cell Systems *10*, 526–534
June 24, 2020 
https://doi.org/10.1016/j.cels.2020.05.004

CellPress

CellPress

Cell Systems

Report

# Predicting Evolutionary Constraints by Identifying Conflicting Demands in Regulatory Networks

Manjunatha Kogenaru,[1,2,6] Philippe Nghe,[1,3,6,*] Frank J. Poelwijk,[4,6] and Sander J. Tans[1,5,7,*]

[1]AMOLF, Science Park 104, Amsterdam 1098 XG, the Netherlands

[2]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

[3]Laboratoire de Biochimie, UMR CBI 8231, ESPCI Paris - PSL, PSL Research University, 10 rue Vauquelin, Paris 75005, France

[4]Department of Data Sciences, Dana-Farber Cancer Institute, 360 Brookline Avenue, Boston, MA 02215, USA

[5]Department of Bionanoscience, Kavli Institute of Nanoscience Delft, Delft University of Technology, Van der Maasweg 9, Delft 2629, the Netherlands

[6]These authors contributed equally

[7]Lead Contact

*Correspondence: philippe.nghe@espci.psl.eu (P.N.), tans@amolf.nl (S.J.T.)

https://doi.org/10.1016/j.cels.2020.05.004

## SUMMARY

Gene regulation networks allow organisms to adapt to diverse environmental niches. However, the constraints underlying the evolution of gene regulation remain ill defined. Here, we show that partial order—a concept that ranks network output levels as a function of different input signals—identifies such constraints. We tested our predictions by experimentally evolving an engineered signal-integrating network in multiple environments. We find that populations: (1) expand in fitness space along the Pareto-optimal front associated with conflicts in regulatory demands, by fine-tuning binding affinities within the network, and (2) expand beyond the Pareto-optimal front through changes in the network structure. Our constraint predictions are based only on partial order and do not require information on the network architecture or underlying genetics. Overall, our findings show that limited knowledge of current regulatory phenotypes can provide predictions on future evolutionary constraints.

## INTRODUCTION

Regulatory networks that integrate multiple environmental cues enable organisms to proliferate in diverse environments (Bell, 2010; Brion et al., 2016). For instance, bacteria can tolerate highly diverse conditions by recognizing specific combinations of stressors, such as pH and osmotic pressure (Hofmann and Todgham, 2010), and plants can elongate above dense canopies by responding to particular combinations of light intensity and wavelength (Ballaré, 2014). However, it is not straightforward to establish whether a particular regulatory network is able to optimally respond to the multiplicity of signals presented by a complex environment (Taute et al., 2014; Poelwijk et al., 2007; Sorrells et al., 2015; Peng et al., 2015; de Vos et al., 2013) and, consequently, how evolutionary constraints limit the range of tolerated environments. Constraints on the adaptation abilities of regulatory networks have been studied experimentally, by targeted mutagenesis or knockout of its constituent components (Sorrells et al., 2015; Mayo et al., 2006; Wray, 2007), and computationally, by varying parameters in kinetic models (Ma et al., 2009; Ciliberti et al., 2007; Payne and Wagner, 2013). The common denominator for these approaches is their need for detailed information on the network topology and the functioning of its parts, which is lacking for many phenotypes. For instance, evolutionary constraints have been mostly studied for regulatory networks that control developmental programs (Jiménez et al., 2015), but they remain poorly explored for networks that are involved in competition and selection in variable environments (Taute et al., 2014; Nghe et al., 2018a). To address these issues, we developed a method to identify constraints in signal integration phenotypes that only requires information on current responses, and subsequently tested the predictions by subjecting synthetic networks in *Escherichia coli* to laboratory evolution.

## RESULTS

### The Order of Expressed Phenotypes

Central to our approach is the concept of *partial order*, which is used in combinatorial optimization applied to task scheduling, algorithmic verification (Nghe et al., 2018b), and decision making (Brüggemann and Carlsen, 2006). We will explain it using the following example: phenotype $P$. $P$ is controlled by two signals, $s_1$ and $s_2$, through a regulatory network (Figure 1A, right), with $S = (s_1, s_2)$ representing both signals. Importantly, the network's response to its input signals is monotonic, that is, $P$ only increases or only decreases as $s_1$ increases, and $P$ only increases

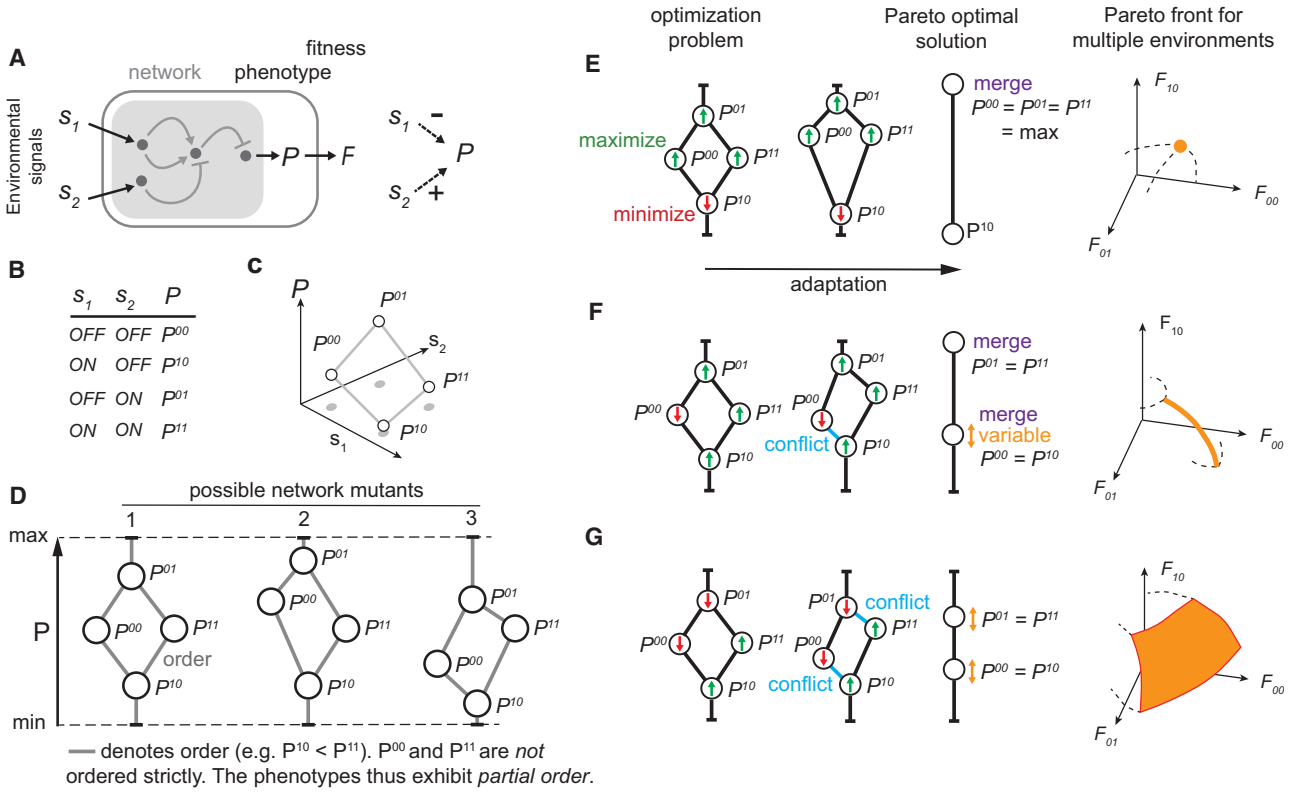



**Figure 1. Partial Phenotype Order Identifies Conflicting Regulatory Objectives and Constraint**

(A) Schematic diagram of studied system. An organism exhibits a phenotype $P$ that controls fitness $F$, in response to two environmental signals $s_1$ and $s_2$ (left). A signal integration network is used for this purpose, but its architecture can be unknown in this framework. The only required information is that $P$ is repressed by $s_1$ and activated by $s_2$ (right). This also implies that dependencies between $s_1$ and $s_2$ may well exist but do not need to be known.

(B) Four environmental conditions that alternate in time. Signals $s_1$ and $s_2$ can be present at a certain concentration ("*ON*") or absent ("*OFF*").

(C) Corresponding magnitudes of $P$. Independently of other details, we know $P^{01}$ is highest because $P$ is not repressed (by $s_1$) but activated (by $s_2$). $P^{10}$ is lowest because $P$ is repressed and not activated.

(D) Hasse diagrams explaining partial order corresponding to (C). Phenotypes are nodes (circles), with their position along the Y axis indicating their magnitude, in between their maximum and minimum possible values. We consider mutations that alter the activation and repression, hence change the positions of the nodes. However, as an example, $P^{01}$ remains higher than $P^{00}$, therefore the two nodes exhibit an invariant order (grey line), which represents a constraint. There is no invariant order between $P^{00}$ and $P^{11}$, because either repression or activation can dominate. Therefore, these phenotypes are not connected by a grey line. Hence, the four phenotypes exhibit a *partial order*.

(E–G) Prediction of conflicts, trade-offs, Pareto-optimal front, and its dimensionality. Nodes indicate the magnitude of $P$ in the four environments in between their maximum and minimum. Up or down arrows in circles indicate selection objectives imposed on $P$ in the four environments. Connected nodes are constrained in their rank order, unconnected nodes are not. Blue lines indicate conflicts between selective objectives and order constraints. Right: dashed lines are cartoons of the edge of fitness domain, accessible given the partial order constraints. Orange indicates the Pareto front of this domain. (E) No conflict case: through mutation and selection, all nodes can meet their objective without conflict. Phenotypes in three environments are maximal and equal, and their nodes merge. This yields a single optimum phenotype, and hence a single point in the multi-environment fitness space (right). (F) One-conflict case. The order indicates $P^{10} < P^{00}$ while $P^{10}$ objective is minimization and $P^{00}$ is maximization. The best solution is then $P^{10} = P^{00}$ (nodes merge), but no particular value of $P$ meets the objectives better than another. This results in a one-dimensional Pareto front in the fitness space (orange line), indicating a trade-off: if $P^{10} = P^{00}$ increases, $F^{00}$ increases at the expense of a decrease in $F^{10}$. (G) Two-conflict case, resulting in a two-dimensional Pareto front (right).

or only decreases as $s_2$ increases. Such monotonic responses are ubiquitous in a wide range of biochemical networks (Figure S1) (Sontag, 2007). In our approach, the only information that is required is whether these signals repress (here $s_1$) or activate $P$ (here $s_2$, see Figure 1A, left). $P$ may, for instance, be the expression level of a gene, but in principle can be of any phenotype on which selection acts. With the two signals, $s_1$ and $s_2$, being either "*OFF*" (0) or "*ON,*" four environments are possible. The magnitudes of the corresponding values of $P$ are denoted as: $P^{00}$, $P^{10}$, $P^{01}$, and $P^{11}$ (Figure 1B). Note that the *OFF* and *ON* values of $P$ are taken for simplicity. However, in principle any number of intermediate values can be considered within this approach.

Key to our approach is to identify the rank order between pairs of these values of $P$—or lack thereof. For instance, here we know $P^{00} < P^{01}$, because the *activating* signal $s_2$ is turned *ON* in $P^{01}$. Similarly, $P^{00} > P^{10}$ because the *repressing* signal $s_1$ is turned *ON* in $P^{10}$ (Figure 1C). Importantly, this ranking order is preserved even upon mutations that change the strength but not the nature of the repression and the activation, as $s_1$ and $s_2$ then still repress and activate—even if they do so differently. However, this preservation of rank order does not hold for all pairs. In particular, $P^{00}$

may be either larger (Figure 1D, middle) or smaller than $P^{11}$ (Figure 1D, right), because here both the *repressing* $s_1$ and the *activating* $s_2$ singles are turned ON in $P^{11}$. If the repression dominates over the activation, then $P^{00} > P^{11}$, but conversely if the activation dominates over the repression, then $P^{00} < P^{11}$. Hence, $P^{00}$ and $P^{11}$ are not strictly ordered.

Thus, all values of $P$ together are bound by a *partial order constraint*, as the rank order between some pairs is constrained but for others it is not (Figure 1D). Note that none of these constraints are absolute, as evolution may extensively alter the network and its components and thus produce any change in order. Indeed, genetic constraints in general are not absolute and can be broken in principle, in contrast to constraints based on physico-chemical laws (Taute et al., 2014; Poelwijk et al., 2007; de Vos et al., 2013). These *partial order constraints* should thus be obeyed for evolutionary adaptation where $s_1$ and $s_2$ maintain their repressive and activating nature (Figure 1A, right).

## Partial Order Constraints and Their Graph Representation

The absence of rank order between some pairs of phenotypic values seems to imply that phenotypic order will not always allow us to identify a constraint. However, the notion of *partial* order provides an approach to still capture the available order information, which can be represented in graphs introduced in Figure 1D, which are also called Hasse diagrams (Davey and Priestley, 2002). The nodes of the graph thus represent the magnitude of the phenotypes $P(S)$ for different environments $S$ (Figure 1D). The node with the highest value of $P$ is displayed at the top, the one with the lowest value of $P$ at the bottom. Any two nodes that are ordered are connected by vertices. Here the resulting Hasse diagram is diamond-shaped (Figure 1D). Mutations affecting characteristics like activation and repression strength can move nodes up and down, but cannot alter the connectivity or topology of the graph. The graph therefore is a *partial order* graph, as it defines both the order and lack of order that is present.

Given a certain evolutionary objective, the partial order graphs become simpler, and indicate the optimal $P(S)$ values. For instance, suppose that the four environments alternate in time, with low $P$ favored only in $S = (1,0)$, and high $P$ in the other environments (green and red arrows in Figure 1E, left). In this situation, all four nodes can optimize to the minimum and maximum possible values of $P$ without encountering conflicts. More specifically, the optimal solution here is that $P^{10}$ takes the minimum value, and $P^{00}$, $P^{01}$, and $P^{11}$ all take the maximum possible value, which also means that these three nodes merge (Figure 1E, middle). Here, the solution is thus one regulatory phenotype (defined by $P^{00}$, $P^{10}$, $P^{01}$, and $P^{11}$), which determines a single point in fitness space (Figure 1E, right).

The situation differs for selective objectives that favor the minimization of $P^{00}$ and maximization of $P^{10}$, because it logically conflicts with the order constraint $P^{00} > P^{10}$ (Figure 1F). In other words, $P^{00}$ cannot minimize and $P^{10}$ maximize because $s_1$ is a repressing signal—turning it *ON* in $P^{10}$ will decrease $P$ (Figure 1A). Rather, the best solution then is for $P^{00}$ to be equal to $P^{10}$. Again the nodes thus merge, but now their objectives (Figure 1F, green and red arrows) are met similarly (poorly) for all $P$ values, as long as $P^{00} = P^{10}$, which thus remains variable within the optimal solution (Figure 1F). Consequently, the system can either optimize $P^{10}$ (by increasing $P^{00} = P^{10}$), or optimize $P^{00}$ (by decreasing $P^{00} = P^{10}$), but not both. Overall, the shape of the Hasse diagrams thus becomes simpler under selective pressure and allows the identification of regulatory conflicts.

## Conflicting Objectives, Trade-Offs, and Pareto Fronts

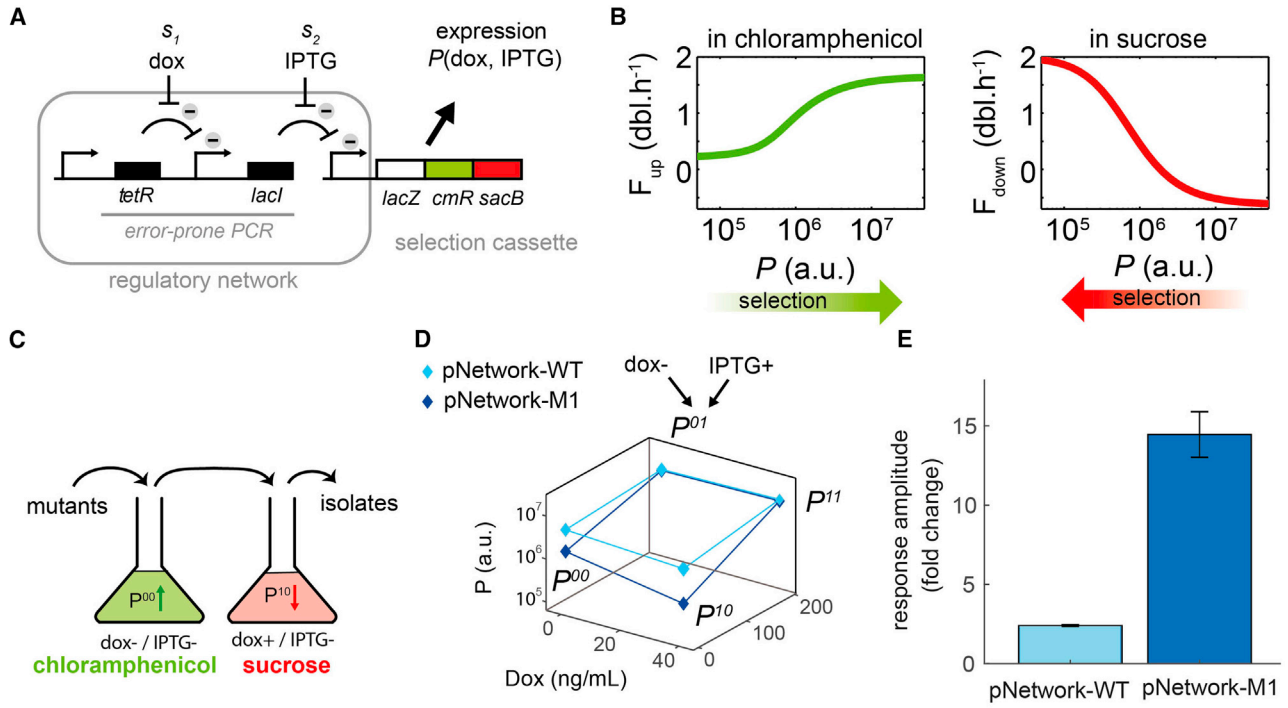
The conflict discussed above reflects a trade-off: at some point (when the nodes merge, Figure 1F), the system can do better at the objective in one environment, but only by doing worse at the objective in another environment. Hence, there are now multiple best regulatory phenotypes which map to a line in fitness space (Figure 1F, right), this line being referred to as the Pareto front (Pareto, 1906), and its curvature depends on the expression-fitness relations (Figure 1F, right). In the same way, the Pareto front becomes a surface for two conflicts (Figure 1G), and so on. We have previously shown theoretically that this approach can reduce many constraints into a Pareto optimal set of much lower dimension, for arbitrary numbers of signals and monotone responses (Nghe et al., 2018b). The algorithm developed for this purpose processes larger-scale Hasse diagrams, and hence similarly as above identifies conflicting regulatory objectives, the number of which predicts the dimensionality of the Pareto front. We illustrate with a few examples how one can find the Pareto optimal set of regulatory phenotypes for more complex networks (Figures S1 and S2).

## Experimental System and Selection Protocol in Variable Environments

To test the partial order predictions experimentally, we constructed a regulatory network responding to two input signals, designed a selection protocol in which mutants compete in different consecutive environments that correspond to a variety of objectives (Figure 2), and tested this protocol by improving the initial network, pNetwork-WT. Specifically, we engineered a genetic network responding to the inducers doxycycline (dox) and isopropyl-β-D-galactopyranoside (IPTG), which define $s_1$ and $s_2$ respectively (Figure 2A; STAR Methods; Table S1). The network controlled a selection cassette (Poelwijk et al., 2011a), whose expression level defined $P$. In this way, the network output $P$ was coupled to the growth rate ($F$), on which selection acts (Figure 2A).

To select for increased $P$, a chloramphenicol-containing medium was used, resulting in a growth rate $F_{up}$. Increased expression of chloramphenicol acetyltransferase (cmR) within the cassette provides resistance and hence faster growth. The relation between $P$ and $F_{up}$ can be quantified (Poelwijk et al., 2011a) (Figure 2B, left), which shows that $F_{up}$ is close to zero when $P$ is low, and increases to about 1.7 doublings per hours (dbl/h) when $P$ is high. To select for decreased $P$, a sucrose-containing medium was used, resulting in a growth rate $F_{down}$. Because the polymerization of sucrose by the levansucrase enzyme (SacB) is toxic, decreased expression of the cassette then produces faster growth. Quantification of the relation between $P$ and $F_{down}$ indeed shows that $F_{down}$ is about 2 dbl/h when $P$ is low and decreases to negative values when $P$ is high (Figure 2B, right) (Poelwijk et al., 2011a).

The regulatory behavior of the initial network (pNetwork-WT) corresponds to the example of Figure 1A (right). Increases in $s_1$




F

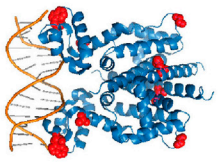
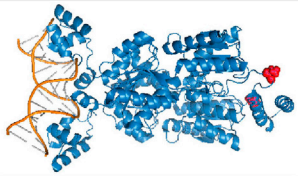
| Network mutant | *tetR*/protein | TetR structure | *lacI*/protein | LacI structure |
|---|---|---|---|---|
| pNetwork-M1 | G128C/Trp43Ser<br>C132T/synonymous<br>C149G/Ala50Gly<br>G214T/Gly72Trp<br>C568A/Leu190Ile |  | G531A/synonymous<br>C1001T/Thr334Met<br>-<br>-<br>- |  |

**Figure 2. Experimental System and Selection Protocol in Variable Environments**

(A) Diagram of engineered genetic network and selection cassette. Expression of selection cassette (*P*) is controlled by two environmental signals dox ($s_1$) and IPTG ($s_2$). Dox represses *P*, owing to the sequence of 3 negative interactions. IPTG is an activating signal for *P*, owing to the sequence of 2 negative interactions. The effect of IPTG depends on dox, and vice versa, the effect of dox depends on IPTG. Our partial order framework does not rely on knowing such dependencies – only that $s_1$ represses and $s_2$ activates *P* (Figure 1). The network is mutated using error-prone PCR, yielding a population of different network variants. The selection cassette contains beta-galactosidase fragment LacZ to measure expression levels and the chloramphenicol resistance gene *cmR* to select for increasing *P*. The levansucrase gene *sacB* confers toxicity in sucrose media, hence selects for decreasing *P*.

(B) Quantification of selection (data from Poelwijk et al., 2011a). Measured growth rates as function of *P*, which here is varied by induction. In a chloramphenicol-containing medium (left), cells with high *P* grow faster and hence are favored, due to their high levels of resistance. In a sucrose-containing medium, cells with lower *P* grow faster and hence are favored, due to reduced toxicity.

(C) Illustration of the experimental evolution protocol (see STAR Methods). A mutant population is successively grown in two environments, which present input signals (dox and/or IPTG) and selection *P* (cm or sucrose).

(D) Expression *P* for pNetwork-WT (light blue) or pNetwork-M1 (dark blue) in presence and absence of dox and IPTG, measured using a $pP_{trc}$-*eYFP* reporter construct. eYFP fluorescence was normalized to the optical density at 550 nm. Data are represented as mean ± SD from two biological replicates (error bars are smaller than the symbols and thus not visible). This shows $P^{10}$ decreased according to selection, while preventing TetR knockouts by also selecting for high $P^{00}$.

(E) Measured increase in the dynamic range (the fold change between highest and lowest value of *P*). Data are represented as mean ± SD from two biological replicates.

(F) Genotype of pNetwork-M1, obtained by Sanger sequencing of corresponding plasmid. Amino-acid substitutions are indicated in red in the TetR and LacI structural models (PDB codes 1QPI and 1Z04).

(dox) should relieve repression by the upstream TetR repressor, of the downstream LacI repressor, which in turn represses the network output *P*. On the other hand, increases in $s_2$ (IPTG) should relieve repression of *P*. Thus, $s_1$ should repress, while $s_2$ should activate *P* (Figure 1A). Note that here the effect of IPTG on *P* should depend on dox, and *vice versa*, the effect of dox should depend on IPTG. The advantage of the partial order framework is that the details of such dependencies do not need to be known (Figure 1). Using an enhanced yellow fluorescent protein (eYFP) reporter, we verified that dox and IPTG were indeed repressing and activating signals, respectively (Figure 2D; Table S1). The dynamic range was small, however: even the

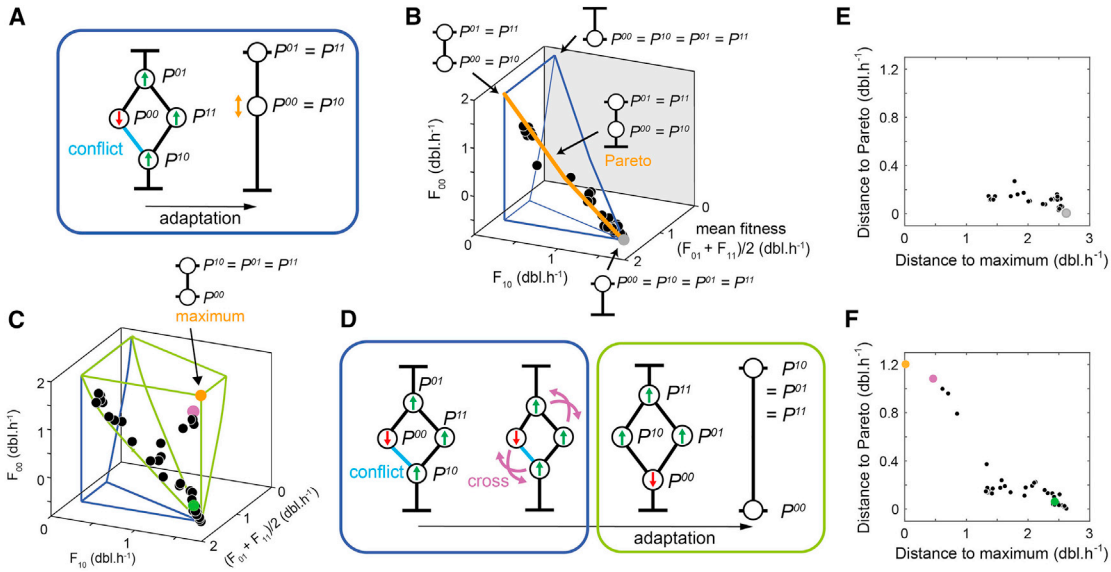



**Figure 3. Experimental Evolution under Conflicting Objectives**

(A) Constraint prediction for experimental network and selective growth. Selection objectives (green and red arrows) for phenotypes *P* in different environments (see Figure 1) predict a single conflict between $P^{10}$ and $P^{00}$ (blue edge). Genetic adaptation in response to this selection should drive $P^{10}$ and $P^{00}$ to become equal (nodes merge). The two objectives cannot be met, given the order constraint. Their magnitude remains variable (orange arrows), because it is not clear, *a priori*, which objective is more important.

(B) Blue and orange lines: boundaries of the accessible fitness domain theoretically predicted from the partial order constraints (A). Corners of this domain are predicted archetypal regulatory phenotypes. Connecting edges are predicted intermediate regulatory phenotypes, which are determined using the partial order graph (A) and the measured relations between *P* and fitness (Figure 2B). Orange line: Pareto front ($P^{10} = P^{00}$ varies along this line) predicted from the optimization objectives and partial order. Grey dot: initial genotype pNetwork-WT, which is close to the archetypal phenotype that has maximum expression in all environments (corner of fitness domain). Black dots: single isolates after first selection round (Figure 2A), whose phenotype was measured. See Figure S4A for 2D projection of same dataset.

(C) Blue lines: as panel B. green lines: fitness domain predicted from partial order constraints in (D). green dot, pNetwork-M2 and founder for the second round of mutagenesis; Orange dot, theoretically predicted optimal phenotype in green domain; Purple dot, pNetwork-M3 isolate closest to orange dot; Black dots, other experimental isolates measured after the second round of selection. See Figure S4B for 2D projection of same dataset.

(D) Suggested change in partial order. Conflict is resolved by crossing the nodes, which alters the order relations. Resulting Hasse diagram (green box) predicts a different accessible fitness domain (green lines C).

(E and F) Euclidian distance to optima, for isolates after first and second selection round. Colors as in (B and C). Data are represented as mean from at least two biological replicates.

lowest *P* ($P^{10}$) was near the maximum of the range of *P* (2.4-fold change, Figure 2E). To validate the experimental evolution protocol, we first aimed to select for a lower $P^{10}$.

The mutation-selection protocol was tested for a network optimization without regulatory conflicts (as in Figure 1E). The network spanning *tetR* to *lacI* including the promoters of respective genes (2,235 base pairs) was randomly mutated by an error-prone PCR (Figure 2A). On average, we observed three mutations per kilobase, or 6–7 mutations for the entire mutagenic region. We subsequently inserted the mutated amplicons into a vector containing intact selection cassette (see STAR Methods). This resulting pool of mutant networks was transformed into *E. coli*, yielding a population size of about $5 \times 10^5$ to $1 \times 10^7$. Next, we performed selection in a medium with chloramphenicol in the absence of both inducers, which corresponds to an upward pressure for $P^{00}$ (Figures 2B and 2C). After this, the population was transferred to a medium containing dox and sucrose but without IPTG, which corresponds to a downward pressure on $P^{10}$ (Figures 2B and 2C; STAR Methods). This dual selection provided a counter-selection for *tetR* knockout mutants. The latter can be achieved by a wide range of mutations that not only decrease $P^{10}$ but also fully abolish the response to dox. Finally, the expression *P* was characterized in the presence and absence of IPTG and dox for isolates of the resulting population, using an eYFP reporter as the network output. Figure 2D shows the results for an evolved network (pNetwork-M1) that displayed the most significant decrease in $P^{10}$ (Figure 2E, 14-fold change of the response after selection), while the response to dox remained intact. Sequencing showed that both TetR and LacI were indeed mutated within pNetwork-M1 (Figure 2F). Overall, these findings indicated that the network could be optimized with this mutation and selection protocol.

## Prediction of Fitness Domains in Presence of Conflicting Objectives

Next, we aimed to study the case of conflicting regulatory objectives (Figures 1F and 3A), using the same genetic network and evolutionary protocol (Figure 2A). This case corresponded to a downward selection of $P^{00}$, an upward selection of $P^{10}$, $P^{01}$, and $P^{11}$ (Figure 3A), and hence a conflict between $P^{00}$ and $P^{10}$. Using the predicted optimum partial order graph (Figure 3A, right) and the measured $F_{down}(P)$ and $F_{up}(P)$ (Figure 2B), we

determined the shape of the one-dimensional Pareto front theoretically (Figure 3B, orange line). We similarly mapped out the accessible four-dimensional fitness space (Figure S3), here visualized in three dimensions by averaging the fitness of $P^{01}$ and $P^{11}$, as they both favor maximization without conflict (Figure 3B). They can similarly be mapped onto a two-dimensional fitness space (Figure S4). In these fitness spaces, each network genotype and its regulatory phenotype (as defined by $P^{00}$, $P^{10}$, $P^{01}$, and $P^{11}$), corresponds to a single point, while the Pareto front represents a collection of networks. The two extreme ends of this Pareto line are two archetypal regulatory responses (Figure 3B), following Shoval et al. (Shoval et al., 2012). These respectively correspond to $P^{00}$ and $P^{01}$ being low while $P^{10}$ and $P^{11}$ are both high, and $P^{00}$, $P^{10}$, $P^{01}$, and $P^{11}$ all being high (Figure 3B). Additionally, the lower and upper limit of $F^{00}$ on the Pareto front in Figure 3B correspond to the minimum and maximum values measured for $F_{down}(P)$ and $F_{up}(P)$, which level off at low and high $P$, as seen in Figure 2B. The other corners of the fitness boundaries in Figure 3B correspond to other archetypal regulatory phenotypes that are not optimal under this selective regime (Figure S3). Altogether, the regulatory archetypes and the line connecting them limit a domain of fitness that is theoretically accessible under the defined partial order constraints, while the space beyond it is predicted to be inaccessible.

## Experimental Evolution with Conflicting Objectives

We tested the above theoretical predictions by experimental evolution. The initial wild-type network (pNetwork-WT, Figure 2A) was found to map to a corner of this fitness domain (Figure 3B, grey dot) and hence is close to an archetypal response. This is consistent with the empirical observation that the expression level $P$ was comparatively high in all four environments (Figure 2D). Next, we mutated the network by error-prone PCR, yielding about $5 \times 10^5$ to $1 \times 10^7$ variants, and performed selection in different environments, with a protocol similar to the one described in Figure 2C, but with the selective pressures as indicated in Figure 3A. The resulting isolated and characterized genotypes were found to remain confined within the fitness domain predicted by the partial order (Figure 3B, each black dot representing one assayed isolate). Moreover, a majority of these isolates scattered along the predicted Pareto front (Figure 3B, black dots along the orange line, Figure S4A). The isolates showed increased $F^{00}$ at the expense of decreased $F^{10}$, while hardly affecting $F^{01}$ and $F^{11}$ (Figure 3B), consistent with the single conflict identified between the nodes $P^{00}$ and $P^{10}$ (Figure 3A). A similar experiment starting from the same mutant but with a selection pressure inducing a conflict between $P^{11}$ and $P^{10}$ also yielded spreading of mutant networks along a Pareto front, though logically a different one (Figure S5), indicating that our method is not limited to one selection protocol.

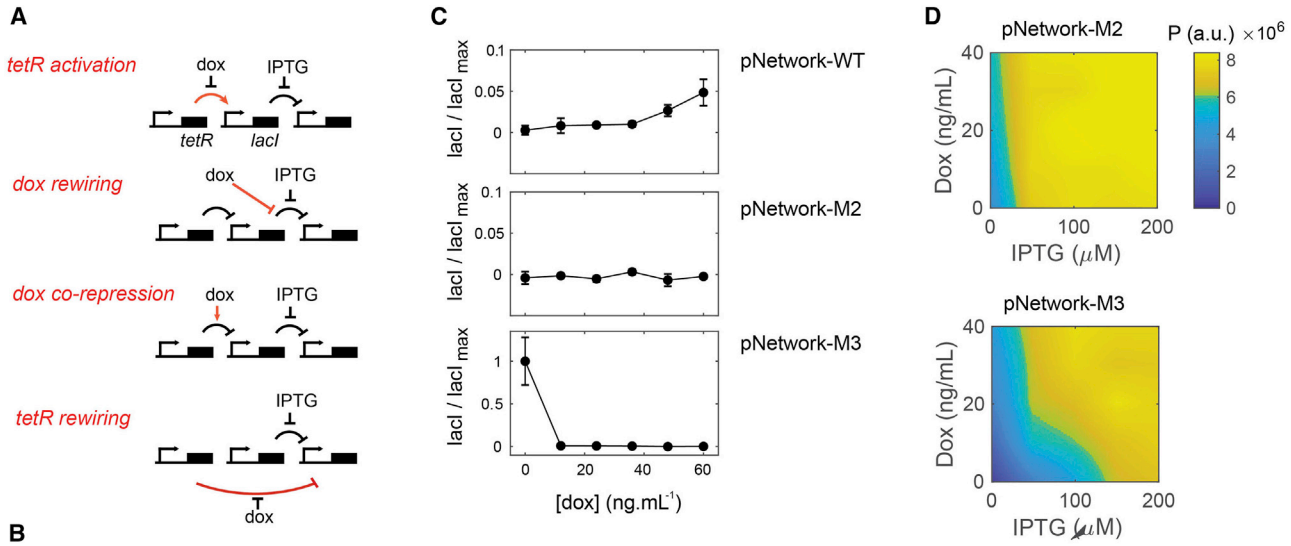
We performed a second round of mutagenesis and selection, with as founder one of the isolates that was selected during the first round (Figure 3C, pNetwork-M2, green dot). Consistent with the previous round, many of the progeny that were present in the population after selection in multiple environments were again found scattered along the one-dimensional Pareto front (Figure 3C, black dots along the orange line). Some of the progeny, however, had moved outside the predicted fitness domain (Figure 3C). These isolates were scattered toward the theoretical fitness optimum (orange dot in Figures 3C, 3F, and S4B), where growth rates are as high as they can be in all environments. One of the isolates, referred to as pNetwork-M3, had a nearly optimal response (Figures 3C and 3F, purple dot).

The departure from the initial fitness domain findings could, at least in principle, indicate a problem with our predictions. Instead, however, as shown in the next section, we found that they rather indicated the emergence of an altered network structure that resolved the conflict between the $P^{00}$ and $P^{10}$ objectives and hence overcame the partial order constraints and allowed population of the fitness domain corresponding to this altered network structure.

## Breaking Order Constraints with Changes in Network Topology

We wondered if our partial order approach was consistent with the observed higher fitness genotypes outside the fitness domain in Figure 3C. Indeed, partial order can change in principle through more extensive network changes. Originally, the objectives of decreasing $P^{00}$ while increasing $P^{10}$ conflicted with the phenotype order, $P^{00} > P^{10}$ (Figure 3A). This conflict would be resolved by a changed order: $P^{10} > P^{00}$, which means that the corresponding nodes cross (Figure 3D). This partial order change can result from several network changes, all representing ways in which $s_1$ (dox) no longer represses $P$ but rather activates it (Figure 4A, top to bottom): (1) the upstream TetR repressor becomes an activator, (2) the downstream LacI becomes sensitive to dox, (3) dox becomes a co-repressor of TetR, and (4) the upstream TetR becomes a direct repressor of the output, like LacI. In these solutions, not only the order between $P^{00}$ and $P^{10}$ changes, but also the order $P^{01} > P^{11}$ reverses to $P^{11} > P^{01}$ (Figure 3D). The new corresponding (green) fitness domain (Figure 3C, green lines) is directly adjacent to the previous (blue) fitness domain, and indeed enclosed the variants that moved away from the initial Pareto constraint.

To distinguish between the different networks (Figure 4A) that could produce the altered partial order (Figure 3D), we sequenced and characterized the responses of the most optimal network variants after rounds 1 and 2 (Figures 4B–4D). The sequences revealed point mutations in the evolved *tetR* and *lacI* coding sequences, but not in the regions controlling DNA or ligand binding. Solutions involving altered ligand binding, which could make LacI sensitive to dox (Figure 4A, second from the top), and altered TetR binding sites on the DNA, which could enable an activator TetR variant (Figure 4A, top), or allow TetR to repress the output directly (Figure 4A, bottom), thus seem unlikely. To further test this hypothesis, we functionally characterized TetR variants from pNetwork-WT, pNetwork-M2, and pNetwork-M3, by measuring the expression of a fluorescent reporter protein that is regulated by TetR directly (Table S1). These data showed that fluorescent reporter protein expression increased with increasing dox for the wild-type TetR, as expected for an induced repressor (Figure 4C). Expression became low and insensitive to dox after the first round (pNetwork-M2). However, after the second round, expression was high for low dox and then decreased with increasing dox (pNetwork-M3). These data suggest that dox has become a co-repressor of TetR (Figure 4A, third from the top). The expression $P$ as a function of dox




B

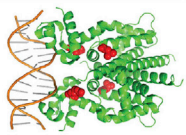
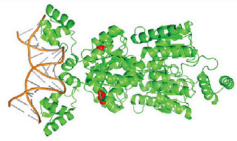
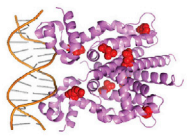
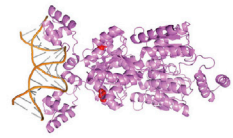
| Network mutant | $P_{N25}$ | *tetR*/protein | TetR structure | $k_{dox}$ (ng/mL) | $K_{TetR}$ | *lacI*/protein | LacI structure | $k_{IPTG}$ (mM) | $K_{LacI}$ |
|---|---|---|---|---|---|---|---|---|---|
| pNetwork-WT | - | - | - | 32 ±6 | 0.011 ±0.004 | - | - | 16 ±1 | 0.015 ±0.005 |
| pNetwork-M2 | -<br>-<br>-<br>-<br>- | C132T/synonymous<br>C145G/Arg49Gly<br>C271Δ/Leu91Δ<br>T272Δ/Leu91Δ<br>A273ΔI/Leu91Δ |  | 25 ±10 | 0.05 ±0.01 | G285A/synonymous<br>G325T/Ala109Ser<br>C900T/synonymous<br>-<br>- |  | 2.5 ±0.5 | 0.04 ±0.01 |
| pNetwork-M3 | A-36G<br>-<br>-<br>-<br>-<br>-<br>- | C132T/synonymous<br>C145G/Arg49Gly<br>C271Δ/Leu91Δ<br>T272Δ/Leu91Δ<br>A273Δ/Leu91Δ<br>T355A/Phe119Ile<br>G549T/Glu183Asp |  | 0.23 ±0.07 | 3.3 ±0.17 | G285A/synonymous<br>G325T/Ala109Ser<br>C900T/synonymous<br>-<br>-<br>-<br>- |  | 2.5 ±0.5 | 0.04 ±0.01 |

**Figure 4. Breaking an Order Constraint with Network Innovations**

(A) Four possible network changes that all correspond to the adjusted partial order, in which the conflict (Figure 3D, blue line) resolved. In all cases, dox now activates *P* whereas it repressed *P* originally (Figure 2A).

(B) Genotypes of pNetwork-M2 and pNetwork-M3, obtained by Sanger sequencing of corresponding plasmid. Amino-acid substitutions are indicated in red in the TetR and LacI structural models (PDB codes 1QPI and 1Z04). Indicated dissociation constants determined by fitting (Methods): $k_{dox}$ , $K_{TetR}$ , $k_{IPTG}$ , $K_{LacI}$ respectively denote binding of dox to TetR, TetR to promotor, IPTG to LacI, LacI to promotor. $K_{TetR}$ and $K_{LacI}$ are unit-less because of normalization to maximum TetR and LacI concentrations.

(C) LacI expression level vs dox concentration for different TetR variants. The LacI expression level is normalized by the highest measured value obtained across all experiments (no dox, lowest panel). LacI expression is measured by fusing it with a fluorescent marker mCherry (Table S1). Top: data for the wild-type network. Dox relieves repression of LacI by TetR, and hence increases LacI expression. Middle: LacI expression level from the first round selection isolate that was transferred to the second round of selection (Figure 3C, green dot). The expression level of LacI becomes insensitive to dox, within detection limit. Bottom: LacI expression after the second round, of the most optimal isolate (Figure 3C, purple dot). Dox now decreases LacI expression. Data are represented as mean ± SD from two biological replicates.

(D) Measured *P* as a function of dox and IPTG, for the isolate transferred from round 1 and the most optimal isolate. These data are consistent with dox having become a co-repressor of TetR (A, third from the top).

and IPTG concentrations (Figure 4D) was indeed improved; from a nearly insensitive response to dox before the second round of selection (Figure 4D, top), to a regulatory phenotype (Figure 4D, bottom) in which $P^{00}$ is minimized and $P^{10}$, $P^{01}$, and $P^{11}$ are maximized. In fact, the observed mutation R49G (Figure 4B) is consistent with a TetR inversion (Scholz et al., 2004). Fitting these data to a cascade model of biochemical rate constants suggests that the dissociation constants of both transcription factors have changed across the rounds of evolution (Figures 4B and S6). Overall, we found that a molecular innovation resolved the conflict and allowed access into a new region of fitness space.

## DISCUSSION

Pareto fronts have long been established as a powerful concept in disciplines ranging from economy to instrument design (Greco et al., 2016; Osiadacz, 1989; Marler and Arora, 2004). They allow one to consider possible solutions and their limits when pursuing multiple objectives. In biology, Pareto fronts are observed when mapping collections of species within multi-dimensional phenotype spaces (Walker et al., 2008; Poelwijk et al., 2011b). It has been proposed that Pareto fronts can be detected by interpolating highly specialized phenotypes (also called archetypes)

(Shoval et al., 2012). They thus provide insight in constraints arising from functional trade-offs in evolutionary adaptation. At the same time, it is a major challenge to mechanistically understand and predict such constraints (Hofmann and Todgham, 2010; Taute et al., 2014; Mayo et al., 2006; Kogenaru et al., 2009), in addition to observing them. Even framing the problem is not straightforward: to identify which aspects could be predictable and which not, and what information about the evolving system is then required, whether it be at the genetic, phenotypic, or fitness level.

Here, we addressed these issues by developing a framework to predict constraints of networks that integrate multiple signals in monotonic fashion. We found that the notion of partial order identifies such constraints. Specifically, it defines the evolutionary limits of a network, in which functional properties such as transcription-factor-binding affinities can be altered, but their activating or repressive nature and the overall network topology remain unchanged. The partial order captures the limited amount of information that is needed, namely whether input signals activate or repress the phenotype in question. Notably, not needed are typically poorly understood details like the actual topology, the number of regulatory proteins, how they function, or which mutations affect their function and how. Owing to its foundations in graph theory, the approach is well suited for more complex environments and regulatory objectives and indeed can reveal the minimal core underlying conflicts between these objectives (for example, see Figure S1) (Nghe et al., 2018b).

The partial order framework may be used to predict the space of accessible phenotypes, provide the dimensionality and shape of the Pareto-optimal front, identify extremal regulatory phenotypes (regulatory archetypes), and allow more targeted network engineering approaches. Interestingly, the dimensionality of the Pareto-optimal front equals the number of conflicts between (regulatory) objectives in different environments. This dimensionality relates to diversity, as a zero-dimensional front indicates a single optimal phenotype, while additional dimensions indicate that diverse phenotypes can be equally optimal. Conceptually, one may see the partial order analysis to apply to regulatory networks in a similar fashion that Flux Balance Analysis (Ibarra et al., 2002) applies to metabolic networks, where the mere knowledge of a graph structure constrains accessible fluxes and optimality is used to predict evolutionary outcomes. We note that the approach is not valid for non-monotonic responses to signals, when unknown signals vary jointly with considered signals, and when dynamical features such as oscillations are central to function and selective advantage. On the other hand, many regulatory responses are monotonic (Sontag, 2007). In addition, it has been shown that any biological regulatory network can be decomposed into monotonic modules (Dasgupta et al., 2007), which may allow further generalization.

The experiments presented here provided a direct test of these concepts, and illustrate which types of functional changes can modify the phenotype order. Experimental evolution in multiple environments revealed two modes of adaptation. In the first mode, solutions that emerge are those that obey the partial order constraints defined by the founding genotype, and are enriched at the predicted Pareto-optimal front. The second mode involves types of mutations that are rarer: those that confer functional innovations capable of altering the partial order, and hence allowing escape from these constraints. This observed escape indicates that partial order represents "soft" genetic constraint that can be overcome in principle, in contrast to "hard" constraints that are for instance set by thermodynamics. The strength of the constraint in turn should depend on various factors such as the mutation rate and the population dynamics. Given the comparatively high mutation rates used here, partial order may be more constraining in nature than in our experiments. Overall, our findings also identify qualitatively distinct evolutionary stages in regulatory strategies.

The ability to define evolutionary constraints of regulatory phenotypes, as we have pursued here, will be central to arriving at a mechanistic understanding of evolution in complex niches. It can provide hypotheses on the compatibility of different regulatory objectives, or lack thereof, and on their evolutionary accessibility, as also illustrated by our data. Such regulatory limitations, and associated tradeoffs when occupying broad spectra of environmental conditions, can promote niche exclusion in the context of competition (Rainey et al., 2000), and hence play a role in species diversity and coexistence.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- METHOD DETAILS
  - Constructs
  - Mutagenesis
  - Selection
  - Measurement of Network Responses
  - Mathematical Formulation of Pareto Optima
  - Application to Signal Integration Networks
  - Fitness Computation
  - Estimation of Network Parameters

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cels.2020.05.004.

## ACKNOWLEDGMENTS

Work in the S.J.T. group is sponsored in part by the Netherlands Organization for Scientific Research (NWO), the Netherlands.

## AUTHOR CONTRIBUTIONS

The research was conceived by M.K., P.N., F.J.P., and S.J.T.; The experiments were performed by M.K. and F.J.P.; Data were analyzed by M.K., P.N., F.J.P., and S.J.T.; Theory was developed by P.N. and S.J.T.; Manuscript was written by P.N., M.K., F.J.P., and S.J.T.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 15, 2019
Revised: March 14, 2020
Accepted: May 17, 2020
Published: June 17, 2020

## REFERENCES

Ballaré, C.L. (2014). Light regulation of plant defense. Annu. Rev. Plant Biol. *65*, 335–363.

Bell, M.A. (2010). Evolution since Darwin : The First 150 Years (Sinauer Associates).

Brion, C., Pflieger, D., Souali-Crespo, S., Friedrich, A., and Schacherer, J. (2016). Differences in environmental stress response among yeasts is consistent with species-specific lifestyles. Mol. Biol. Cell *27*, 1694–1705.

Brüggemann, R., and Carlsen, L. (2006). Partial Order in Environmental Sciences and Chemistry (Springer).

Ciliberti, S., Martin, O.C., and Wagner, A. (2007). Robustness can evolve gradually in complex regulatory gene networks with varying topology. PLoS Comput. Biol. *3*, e15.

Dasgupta, B., Enciso, G.A., Sontag, E., and Zhang, Y. (2007). Algorithmic and complexity results for decompositions of biological networks into monotone subsystems. Biosystems *90*, 161–178.

Davey, B.A., and Priestley, H.A. (2002). Introduction to Lattices and Order (Cambridge University Press).

de Vos, M.G.J., Poelwijk, F.J., and Tans, S.J. (2013). Optimality in evolution: new insights from synthetic biology. Curr. Opin. Biotechnol. *24*, 797–802.

Greco, S., Figueira, J., and Ehrgott, M. (2016). Multiple Criteria Decision Analysis: State of the Art Surveys (Springer).

Hofmann, G.E., and Todgham, A.E. (2010). Living in the now: physiological mechanisms to tolerate a rapidly changing environment. Annu. Rev. Physiol. *72*, 127–145.

Ibarra, R.U., Edwards, J.S., and Palsson, B.O. (2002). Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature *420*, 186–189.

Jiménez, A., Cotterell, J., Munteanu, A., and Sharpe, J. (2015). Dynamics of gene circuits shapes evolvability. Proc. Natl. Acad. Sci. USA *112*, 2103–2108.

Kogenaru, M., De Vos, M.G., and Tans, S.J. (2009). Revealing evolutionary pathways by fitness landscape reconstruction. Crit. Rev. Biochem. Mol. Biol. *44*, 169–174.

Ma, W., Trusina, A., El-Samad, H., Lim, W.A., and Tang, C. (2009). Defining network topologies that can achieve biochemical adaptation. Cell *138*, 760–773.

Marler, R.T., and Arora, J.S. (2004). Survey of multi-objective optimization methods for engineering. Struct. Multidiscip. Optim. *26*, 369–395.

Mayo, A.E., Setty, Y., Shavit, S., Zaslaver, A., and Alon, U. (2006). Plasticity of the cis-regulatory input function of a gene. PLoS Biol. *4*, e45.

Nghe, P., Kogenaru, M., and Tans, S.J. (2018a). Sign epistasis caused by hierarchy within signalling cascades. Nat. Commun. *9*, 1451.

Nghe, P., Mulder, B.M., and Tans, S.J. (2018b). A graph-based algorithm for the multi-objective optimization of gene regulatory networks. Eur. J. Oper. Res. *270*, 784–793.

Osiadacz, A.J. (1989). Multiple criteria optimization; theory, computation, and application, Ralph E. Steuer, Wiley Series in Probability and Mathematical Statistics -Applied, Wiley, 1986, No. of pages 546, Price f5 1.40, $77.10. Optim. Control Appl. Meth. *10*, 89–90.

Pareto, V. (1906). Manuale di Economia Politica.

Payne, J.L., and Wagner, A. (2013). Constraint and contingency in multifunctional gene regulatory circuits. PLoS Comput. Biol. *9*, e1003071.

Peng, W., Liu, P., Xue, Y., and Acar, M. (2015). Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nat. Commun. *6*, 6226.

Poelwijk, F.J., De Vos, M.G., and Tans, S.J. (2011a). Tradeoffs and optimality in the evolution of gene regulation. Cell *146*, 462–470.

Poelwijk, F.J., Heyning, P.D., De Vos, M.G., Kiviet, D.J., and Tans, S.J. (2011b). Optimality and evolution of transcriptionally regulated gene expression. BMC Syst. Biol. *5*, 128.

Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., and Tans, S.J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. Nature *445*, 383–386.

Rainey, P.B., Buckling, A., Kassen, R., and Travisano, M. (2000). The emergence and maintenance of diversity: insights from experimental bacterial populations. Trends Ecol. Evol. (Amst.) *15*, 243–247.

Scholz, O., Henssler, E.M., Bail, J., Schubert, P., Bogdanska-Urbaniak, J., Sopp, S., Reich, M., Wisshak, S., Köstner, M., Bertram, R., and Hillen, W. (2004). Activity reversal of Tet repressor caused by single amino acid exchanges. Mol. Microbiol. *53*, 777–789.

Shoval, O., Sheftel, H., Shinar, G., Hart, Y., Ramote, O., Mayo, A., Dekel, E., Kavanagh, K., and Alon, U. (2012). Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. Science *336*, 1157–1160.

Sontag, E.D. (2007). Monotone and near-monotone biochemical networks. Syst. Synth. Biol. *1*, 59–87.

Sorrells, T.R., Booth, L.N., Tuch, B.B., and Johnson, A.D. (2015). Intersecting transcription networks constrain gene regulatory evolution. Nature *523*, 361–365.

Taute, K.M., Gude, S., Nghe, P., and Tans, S.J. (2014). Evolutionary constraints in variable environments, from proteins to networks. Trends Genet. *30*, 192–198.

Walker, R.S., Gurven, M., Burger, O., and Hamilton, M.J. (2008). The trade-off between number and size of offspring in humans and other primates. Proc. Biol. Sci. *275*, 827–833.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. *8*, 206–216.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Bacterial and Virus Strains | | |
| *Escherichia coli*: strain MC1061 | Avidity Inc | Cat#EVB100; Cat#AVB100 |
| Critical Commercial Assays | | |
| Genemorph II random mutagenesis kit | Agilent | Cat#200550 |
| Recombinant DNA | | |
| Plasmid: pNetwork-WT | This paper | N/A |
| Plasmid: pNetwork-M1 | This paper | NA |
| Plasmid: pNetwork-M2 | This paper | NA |
| Plasmid: pNetwork-M3 | This paper | NA |
| Plasmid: p$P_{lacI}^{Q}$-*lacZ*$\omega$ | Poelwijk et al., 2011a | N/A |
| Plasmid: p$P_{lac}$-*lacZ* | Poelwijk et al., 2011a | N/A |
| Plasmid: p$P_{trc}$-*eYFP* | Nghe et al., 2018a | N/A |
| Plasmid: p$P_{LtetO1}$-*lacI-mCherry* | Nghe et al., 2018a | N/A |
| Software and Algorithms | | |
| Evolutionary Network Analysis Code | This paper | https://hal.archives-ouvertes.fr/hal-02501477/document |

### RESOURCE AVAILABILITY

#### Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Sander Tans (tans@amolf.nl).

#### Materials Availability

Requests for plasmids generated in this study should be sent to tans@amolf.nl.

#### Data and Code Availability

The code generated in this study together with the related data are available on the publicly available French national archive HAL under the reference code hal-02501477 (openURL: https://hal.archives-ouvertes.fr/hal-02501477/document).

### METHOD DETAILS

#### Constructs

We modified a regulatory circuit in which the selection operon consisting of *lacZ*$\alpha$, *cmR*, and *sacB* genes driven by the promoter $P_{trc}$ is under control of the LacI transcriptional repressor(Poelwijk et al., 2011a). Expression of LacI is under the control of TetR repressor via promoter $P_{LtetO1}$. TetR itself is constitutively expressed via promoter $P_{N25}$(Poelwijk et al., 2011a). This network harboring plasmid contains a kanamycin-resistance gene and a medium copy *p15A* origin of replication. Materials are available upon request.

To measure the output from the network variants, various reporter constructs were used (Table S1). The LacZ based assays utilized either constitutively expressed *lacZ*$\omega$ fragment via $P_{lacI}^{Q}$ to measure in *cis,* that constitute a functional LacZ together with *lacZ*$\alpha$ encoded by the selection operon of the network, or the full version of *lacZ* under the promoter $P_{lac}$ to measure in *trans*(Poelwijk et al., 2011a). Whilst the fluorescent protein based readout assays utilized the plasmid encoding either *lacI-mCherry* under the promoter $P_{LtetO1}$ or *eYFP* under the promoter $P_{trc}$ (Nghe et al., 2018a). This reporter plasmid backbone contains an ampicillin-resistance gene and a medium copy *colE1-rop* origin of replication, which is compatible to co-reside with the network harboring plasmid in the same cell (Table S1).

#### Mutagenesis

The mutations were introduced into the regulatory network sequence spanning *tetR* to *lacI* including the promoters of respective genes by error-prone PCR (Stratagene Genemorph II Random Mutagenesis kit). The mutated PCR amplicons digested with DraIII (NEB), ligated into the vector backbone containing the intact selection cassette, and transformed into *E. coli MC1061* strain by electroporation (Avidity EVB100). Routinely we obtained a pool size of half a million to ten million. To determine the mutation rate, DNA

isolated from randomly picked transformants was Sanger sequenced, revealing an average mutation rate of 3.0 mutations per kb (n=9).

### Selection

Cells harboring mutant networks were grown at 37 °C with vigorous shaking in 20-40 mL of EZ Rich Defined medium (Teknova, Cat. M2105) supplemented with 0.2% glucose as a carbon source, 1 mM thiamine hydrochloride and 50 μg/ml kanamycin. The small-molecule inducers doxycycline or isopropyl-β-D-galactopyranoside (IPTG) were added 3 hours prior to the beginning of selection. After this pre-selection growth phase, chloramphenicol (40 μg/ml) or sucrose (0.25% w/v) were added for selection and cells were cultured for an additional 6 hours. This duration of the selection growth phase was chosen to obtain significant enrichment factors (of up to $10^4$), while still maintaining the diversity of the population. During selection the optical density of the culture was monitored at regular intervals and diluted 500 times in pre-warmed medium whenever the optical density (OD) at 550 nm reached 0.1 value.

### Measurement of Network Responses

The output of an isolated network variant was measured by co-transforming with a suitable reporter encoding plasmid (Table S1). In all Figures, the data presented are mean of at least two biological replicates.

For LacZ based assays, 200 μl cultures were grown at 37 °C in EZ Rich Defined medium (Teknova, cat. M2105) with glucose as a carbon source and supplemented with 1 mM thiamine hydrochloride and the appropriate antibiotics, in a 96-well optical-bottom black color micro-titer plate (NUNC, Cat. 165305), using Wallac Perkin Elmer Victor3 plate reader. The OD at 600 nm was recorded at regular intervals of 4 minutes, and evaporation of the cultures were compensated by adding 9 μl of sterile water per well at an interval of 29 minutes. When most of the cultures were grown to 0.05 to 0.2 OD range, cells were fixed by adding 20 μl of fixation solution, which was freshly constituted before use by mixing 109 μM fluorescein-di-β-D-galactopyranoside (FDG, MarkerGene, cat. M0250) substrate, 0.15% formaldehyde and 0.04% DMSO in sterile water. The development of the fluorescence from LacZ activity was measured at regular interval of 8 minutes by excitation at 480 nm and emission at 535 nm, in parallel to the $OD_{600}$ measurement. This data was analyzed as described previously(Poelwijk et al., 2011a).

For the fluorescent protein based assays, the cultures were grown in EZ Rich Defined medium early-exponential growth phase, and then diluted into a final $OD_{550}$ of $1x10^{-4}$ and transferred to a 96-well optical-bottom black color micro-titer plate (NUNC, Cat. 165305) in a total volume of 200 μL per well. The $OD_{550}$ and fluorescence intensities from two distinct fluorescent reporter proteins (mCherry (excitation 580/20, emission 632/45) and eYFP (excitation 500/20, emission 535/25) were monitored in a Wallac Perkin Elmer Victor3 plate reader at regular intervals at 37 °C. The instrument was shaking (double orbital) and replenishing 9 μL of sterile water per well every 27 minutes. This data was analyzed as described previously(Nghe et al., 2018a).

### Mathematical Formulation of Pareto Optima

We consider the environments $E^{(n)}$ (n = 1, ..., N), each corresponding a vector $S^{(n)}$ of inputs signals $(s_1^{(n)}, ..., s_K^{(n)})$ and a monotone regulatory network which transforms each input vector $S^{(n)}$ into a certain level $P^{(n)}$ of the phenotype. Given the monotonicity of the network, the output $P^{(n)}$ is an increasing (resp. decreasing) function of $s_k^{(n)}$ for the indexes k in the subset $I^+$ (resp. $I^-$). The natural partial order on $\mathbb{R}^K$ induces a partial order "$\geq$" between the $P^{(n)}$ values defined as follows: $P^{(u)} \geq P^{(v)}$ if and only if for every k in $I^-$, we have $s_k^{(u)} \leq s_k^{(v)}$, and for every k in $I^+$, we have $s_k^{(u)} \geq s_k^{(v)}$. Additionally, depending on the phenotype being beneficial or deleterious, we consider a fitness optimization objective being respectively the maximization or minimization of $P^{(n)}$.

We have devised a graph algorithm which allows to compute efficiently (in polynomial time) an exact closed form of the Pareto front of this optimization problem under partial order constraints as reported in Nghe et al., 2018b. The steps of the algorithm are exemplified in Figure S2. The problem is represented as a directed graph, where each vertex u corresponds to the phenotype level $P^{(u)}$ in the environment $E^{(u)}$, and an edge between from vertex u to vertex v indicates $P^{(u)} \geq P^{(v)}$. To each vertex, we attribute one of the four following states:

1. "descending" if $P^{(u)}$ should be minimized;
2. "ascending" if $P^{(u)}$ should be maximized;
3. "trade-off" if the vertex results from the fusion of a descending and ascending vertices;
4. "bound" if the vertex results from the fusion of a vertex representing the minimum or the maximum possible value of $P^{(u)}$.

The solution to the problem is found by an algorithm applied to the graph of the partial order between the $P^{(n)}$ values, where the vertex states are attributed according to fitness objectives. At each step of the algorithm, a pair of vertices {U,V} is fused (corresponding to imposing $P^{(u)} = P^{(v)}$), according to the rules described below, applied recursively to the graph, until there is no ascending or descending vertex left:

- Perform a transitive reduction of the graph.
- Find an ascending (or resp. descending) vertex V which is pointed to (resp. which points to) no other vertex of the same state.
- For each vertex U which points to (resp. is pointed by) V, create a new graph resulting from the fusion between U and V, the state of this vertex being "bound" if U is a "bound", or " trade-off" in any other case.

We have shown in (Nghe et al., 2018b) that the terminal graphs of the recursion are exact and minimal parameterizations of each face of the Pareto front. They consists of convex sub-spaces of $\mathbb{R}^N$ defined by the partial order relations (given by a terminal graph)

between sets of variables all equal to each other (the $P^{(u)}$ values which are grouped in a same vertex). The union of all these subspaces is the Pareto front. The number of vertices of each terminal graph provides the dimension of the corresponding face.

### Application to Signal Integration Networks

The full envelope of the accessible fitness, as depicted in main text Figure 3, has been obtained by computing all the extremal regulatory phenotypes compatible with the partial order (Figure S2). Given that the accessible domain of regulatory phenotypes is convex, the envelope of the domains can be computed by interpolating their extremal phenotypes. The envelope in the space of fitness values is then the image $F(P^{(1)},...,P^{(N)}) = (F_1(P^{(1)}),...,F_N(P^{(N)}))$, where each $F_n$ is either the $F^+$ or the $F^-$ function of Figure 2B depending on whether the environment $E^{(n)}$ contains respectively chloramphenicol or sucrose.

### Fitness Computation

The fitness as a function of the expression level $P$ was previously modeled and fitted(Poelwijk et al., 2011a) and take the form:

$$F_{up} = \frac{1.65}{1 + 0.085\left(c - 8 - P\cdot 10^{-5} + \sqrt{1.28\cdot 10^{3} + \left(c - 8 - P\cdot 10^{-5}\right)^{2}}\right)}$$

where $F_{up}$ is the increasing fitness response as a function of increasing expression levels $P$, in the presence of a chloramphenicol concentration $c$ = 40 μg/ml in the medium, and

$$F_{down} = \frac{2.84}{1 + 100\, r^{1.5}} - 0.84$$

with

$$r = 2.7\cdot 10^{-6}\, P \Bigg/ \left(1 + \frac{10}{\left(s - 10 - 2.7\cdot 10^{-6}\, P\right) + \sqrt{40s + \left(s - 10 - 2.7\cdot 10^{-6}\, P\right)^{2}}}\right)$$

where $F_{down}$ is the decreasing fitness response to increasing expression levels P in the presence of a sucrose concentration $s$ = 0.25% (weight fraction) in the medium. Expression levels of the mutants of Figure 3 are reported in the Supplemental Information.

### Estimation of Network Parameters

Dissociation constants were estimated by fitting the responses of the separate components (Figure S5) assuming constant constitutive expression of TetR and using the following forms:

i) TetR induction by dox:

$$\mathrm{LacI}(\mathrm{dox}) = \frac{1}{1 + 1/(K_{\mathrm{TetR}}(1 + \mathrm{dox}/k_{\mathrm{dox}}))}$$

ii) TetR co-repression by dox:

$$\mathrm{LacI}(\mathrm{dox}) = \frac{1}{1 + (1 + \mathrm{dox}/k_{\mathrm{dox}})/K_{\mathrm{TetR}}}$$

iii) YFP output as a function LacI induction by IPTG:

$$\mathrm{P}(\mathrm{dox}, \mathrm{IPTG}) = \frac{P_{\max}}{1 + \mathrm{LacI}(\mathrm{dox})/K_{\mathrm{LacI}}(1/(1 + \mathrm{IPTG}/k_{\mathrm{IPTG}}))}$$

where $K_{\mathrm{TetR}}$ is expressed in units of constitutively expressed TetR concentration, LacI and $K_{\mathrm{LacI}}$ are both normalized to maximum LacI expression. Fitted curves shown in Figure S6. For each regulatory network, the four dissociation constants where fitted to 31 experimentally measured data points which were measured in duplicate.