Limited specificity of molecular interactions incurs an environment-dependent fitness cost in bacteria

Claudia Igler[1,2], Claire Fourcade[1], Torsten Waldminghaus[3], Florian M. Pauler[1], Balaji Santhanam[4], Gašper Tkačik[1], Călin C. Guet[1,*]

[1] IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria

[2] Institute of Integrative Biology, Universitätsstrasse 16 ETH Zurich, 8092 Zurich, Switzerland

[3] Department of Biology, Technische Universität Darmstadt, Darmstadt, Germany

[3] Centre for Synthetic Biology, Technische Universität Darmstadt, Darmstadt, Germany

[4] Department of Systems Biology, Columbia University,

1150 St. Nicholas Avenue, New York, NY 10032

[*] Corresponding author: Călin C. Guet

Email: calin@ist.ac.at

Competing Interest Statement: There are no competing interests.

Classification: Biological Sciences/ Microbiology

Keywords: Transcription factors, DNA binding, limited binding specificity, binding cooperativity

Abstract

Reliable operation of cellular programs depends crucially on the specificity of biomolecular interactions. In gene regulatory networks, the appropriate expression of genes is determined through the specific binding of transcription factors (TFs) to their cognate DNA sequences. However, the large genomic background likely contains many DNA sequences showing similarity to TF target motifs, potentially allowing for substantial non-cognate TF binding with low specificity. Whether and how non-cognate TF binding impacts cellular function and fitness remains unclear. We show that increased expression of different transcriptional regulators in *Escherichia coli* and *Salmonella enterica* can significantly inhibit population growth across multiple environments. This effect depends upon (i) TF binding to a large number of DNA sequences with low specificity, (ii) TF cooperativity, and (iii) the ratio of TF to DNA. DNA binding due to the limited specificity of promiscuous or non-native TFs can thus severely impact fitness, giving rise to a fundamental biophysical constraint on gene regulatory design and evolution.

Introduction

Biology at all levels crucially depends on the timely recognition and interaction between cognate biomolecules (Box 1A). The importance of specificity of molecular encounters in the cell is highlighted by the intricate mechanisms that ensure appropriate, and thus specific interactions, a classic example being kinetic proofreading in the loading of amino acids onto tRNAs (1). In gene regulation, transcription factors (TFs) determine the expression of genes at the right time and place by binding to their respective operator sites in a highly specific manner. Additionally, non-cognate interactions can occur due to non-specific TF-DNA interactions or due to specific binding (i.e. recognition of a TF-specific DNA motif) at non-cognate sites (Box 1). These interactions are not necessarily detrimental, as non-specific DNA binding was found to speed up the TF search process through sliding along the DNA (called “facilitated diffusion”) (2,3). Upon encounter of the target sequence, conformational changes can then increase binding specificity (2). In prokaryotes, many transcriptional regulators appear to be non-specifically bound to DNA most of the time (3,4) and this state likely plays an important role in gene regulation (5–7).

These non-specific TF-DNA interactions are likely weak and transient. However, non-cognate TF-DNA interactions can also arise from recognition of a certain DNA motif or conformation, which makes them more specific and therefore stronger (Box 1). As TFs have to search for their 'correct' DNA targets among an extensive genomic background, they might encounter many non-cognate sites with sufficient target sequence similarity, thus potentially trapping TF molecules (2,8). TF binding to such non-cognate sites could thus incur substantial fitness consequences for the cell through i) high-specificity binding to few DNA sites with high similarity to the target sequence ; or ii) low-specificity binding to many sites across the genome that accrues to a large overall effect. While in i) the fitness effect is more likely to stem from regulatory interference, in ii) also changes in chromosome structure could play a role. In support of the latter, growth arrest has been found with overexpression of one of the nucleoid-associated proteins (NAPs) of *E. coli* involved in DNA organization, H-NS (9), which binds DNA preferentially at AT-rich and curved regions (10).

In eukaryotes, which have shorter operator sites, low-specificity non-cognate binding of TFs seems to be an integral part of gene regulatory function and evolution (11,12). On the other hand, in a recent theoretical study that explored the impact of non-cognate binding we suggested that genomic low-specificity sites impose the existence of a global biophysical constraint, termed "crosstalk" (13). Some forms of TF cooperativity and combinatorial regulation can limit this problem (13), but additional mechanisms, such as out-of-equilibrium proofreading mechanisms, may be at work as well in eukaryotes (14,15). In prokaryotes, the importance of non-cognate binding with low specificity has rarely been investigated, although it has been acknowledged in regulatory interference (16), however experiments exploring its fitness impact in general are lacking entirely.

Results

Experimental setup

We wanted to test the consequences of non-cognate TF binding on bacterial fitness by expressing TFs lacking known cognate binding sites in native and non-native host cells. For this purpose, we chose phage repressors, which coexist for long periods at low numbers in bacterial hosts during the lysogenic cycle. We expect these repressors to be adapted to their host's genomic background – as temperate phages likely spend most of their existence as lysogens (17) – but potentially not to that of

another host. Therefore, we use repressors from different lambdoid phages, but from the same protein family, three of them native to *E. coli* (*λ cI, 434 cI*, *HK022 cI*) and one to *S. enterica* (*P22 c2*) and explore their effects on growth and fitness in both bacterial hosts. A bacterial repressor, LacI, served as a control in our experiments, as it is known to have only a single cognate target region in *E. coli*, at the *lac* operon (18). Further, LacI uses facilitated diffusion for target search (3), which means it should have very weak non-specific binding tendencies that should not affect fitness. Note, that all four phage transcriptional regulators can function as either repressors or activators depending on the promoter they bind to.

The five repressors were each cloned under the control of an aTc-inducible promoter (P_{tet}) onto a low copy number plasmid (Fig. 1A). The plasmids were then transformed into *E. coli* and *S. enterica* cells, whose genomes do not contain any of the phage operators, nor the three *lac* operators. The phage repressors bind to their cognate operator sites in the phage genome as a dimer, but they can also bind cooperatively to adjacent operators or form short- and long-distance loops (up to 10kbp for λ CI (20)) involving two to four dimers (21) (Fig. 1A). This type of TF cooperativity was previously found to facilitate non-cognate binding (22). Notably, for λ CI significant non-cognate binding has been predicted theoretically (22,23) and substantiated experimentally (24), whereas for P22 C2 binding was lost at few mismatches with the target motif (24), making it a less likely non-cognate binder (Box 1B). LacI on the other hand, is a tetrameric protein and cooperativity stems from the fact that one repressor tetramer can bind to two different DNA sites simultaneously, resulting in looping (25), as opposed to cooperativity resulting from binding of two different repressor molecules (Fig. 1A), which LacI lacks. We compared the impact of potential low-specificity non-cognate interactions of the five different TFs by measuring growth of *E. coli* and *S. enterica* in various environments over a 10h interval as a fitness proxy.

Repressor-induced growth reduction depends on growth medium and induction timing

Reduction in growth was quantified as the normalized difference between growth in the presence and absence of repressor (with 1 indicating complete cessation of growth and 0 representing wildtype growth, see Methods). We observed a wide spectrum of growth behaviors across the five repressors and the two bacterial hosts. For cells grown in minimal media (M9) with glucose (the standard media used if not specified otherwise), the presence of λ CI resulted in a strong reduction of growth in *E. coli*

and an even more substantial reduction in *S. enterica* cells (Fig. 1B-D, Table S1). 434 CI also reduced growth in both hosts, though less than λ CI, and interestingly more in its native host, *E. coli* (Fig. 1B). P22 C2 on the other hand, showed no effect in its native host *S. enterica*, while stopping growth completely when expressed in *E. coli* (Fig. 1D, Table S1). There was no significant impact on growth in either host with HK022 CI or with LacI (Fig. 1B), which was expected for the latter - at least in *E. coli*. Further, no growth defect was seen with our other controls: cells with only plasmid backbone or the control plasmid expressing a fluorescence marker instead of a repressor (Fig. S1A). Thus, four different repressors stemming from the same TF family, but likely having different modes of DNA recognition (26), showed a broad spectrum of growth effects in the two different bacterial host species. We explored these growth effects and their causes further by focusing on the two best-characterized ones, λ CI and P22 C2, which are known to have different propensities for binding at DNA sequences far away from their target motif (24) (Box 1B).

As a next step, we varied the environmental conditions in which bacteria carrying λ CI or P22 C2 were grown. In rich media (LB), growth inhibition was abolished almost entirely in *E. coli* for both repressors, and substantially reduced with λ CI expressed in *S. enterica* (Fig. 2A, Table S1). Minimal media supplemented with casamino acids and glucose resulted in intermediate growth reductions between rich and poor media (Fig. S2A, Table S1). P22 C2 did not affect growth in *S. enterica* in any of the conditions (Fig. 2A, Table S1), which is why this combination is generally not discussed further.

Next, we tested the dependence of the growth reduction on repressor concentration and induction timing. In *E. coli*, decreasing the concentration of either repressor showed a gradual recovery of normal growth (Fig. 2B, Table S2). Conversely, even low expression of λ CI in *S. enterica* resulted in strong growth reductions (Fig. 2B, Table S2). The concentrations used here (see Methods for measurement details) range from 0.5-5 fold of those achieved under physiological lysogen conditions (27). Surprisingly, the induction time point was also an important determinant for λ CI-induced growth reduction, not however for P22 C2-induced ones: while λ CI induction in early- and mid-exponential growth (as opposed to induction during the lag phase) had progressively smaller effects on growth in *E. coli* and *S. enterica* (Fig. 2C, S2B,C, Table S3), this was not the case for P22 C2 in *E. coli*, where growth was always halted ~2h after repressor induction (Fig. S2D, Table S3). Overall, we found a

strong dependence of repressor-induced growth reduction on environmental conditions and repressor concentration.

Increased repressor expression leads to severe fitness reduction

As the severe growth reductions we observed made it difficult to determine meaningful growth rates in our system, we determined the fitness effect of repressor expression in direct competition experiments, which reflect all growth differences between the competitors. As a 'neutral' competitor, we used cells expressing LacI from a plasmid construct that contained an additional YFP-*venus* marker (Fig. 3A). The *venus* marker resulted in a minor fitness cost (selection coefficients for cells without the marker were 0.05 (*E.coli*) and 0.09 (*S.enterica*), see Methods), meaning that an increase in fluorescence (i.e. LacI-carrying cells) indicates an even more pronounced benefit of the LacI-carriers than measured. 1:1 mixtures of cells with phage repressor- and LacI-carrying plasmids were grown in minimal media (Fig. 3A, Methods) and fluorescence was compared between cell mixtures grown without the repressors (no fitness difference; baseline fluorescence) and cell mixtures induced for repressor expression (potential fitness cost of phage repressors over LacI) (28). In accordance with the experimental results from Fig. 1B-D, expression of repressors led to a significant increase in LacI-expressing cells, except for competitions with P22 C2 in *S. enterica* (Fig. 3B,C, Table S4). Growth reductions translated directly into fitness costs as the competition assays were even able to capture the gradual increase in growth reduction with increasing repressor concentration for λ CI in *S. enterica* (Fig. S3, Table S4).

Growth reduction is caused by cooperative, low-specificity binding distributed across the genome

Given the surprisingly detrimental growth effect of the two repressors in several environments, we set out to determine its cause. Transcriptional repressors are DNA-binding proteins and could therefore interfere with the cellular program through DNA binding at various non-cognate sites (29). To determine the role of TF binding in the observed growth reductions, we used the fact that λ CI is one of the best-studied TFs, and thus an exhaustive range of mutants for most of its functions exist. As neither of these mutants have been characterized for any of the other repressors, we only performed these experiments with λ CI.

Specifically, we tested the expression of a mutant that cannot form dimers (30) (as λ CI only binds DNA in its dimeric form (31)), as well as of a mutant defective in DNA binding (32), and found that normal growth (Fig. 4A, Table S3) and fitness (Fig. S3) were almost completely restored in *E. coli* as well as in *S. enterica* cells. Similar results for a λ CI mutant defective in cooperativity between repressor dimers (Fig. 4A, Table S3) suggest an important contribution from DNA looping or some other form of repressor oligomerization. This is intriguing as λ CI cooperativity and oligomerization are thought to increase binding specificity (25,33), but likely lead to a general increase in binding strength, particularly in the absence of specific sites. We ruled out that repressor misfolding or aggregation was responsible for our observations by over-expressing a chaperone gene (*tig*) together with the repressors (Fig. S1B).

Hence, the ability to bind DNA, potentially in a cooperative and motif-dependent manner, seems to be central to repressor-mediated growth effects. We tested this hypothesis by combining λ CI cooperativity with the binding specificity of another repressor, using chimeric TFs. Specifically, we replaced the DNA binding helix of λ CI (see Methods) with: i) that of another phage repressor, 434 CI, which showed some growth defect; and ii) the bacterial repressor, LacI, which showed no growth defect as a wildtype protein (Fig. 1B, Table S5). It has been reported that changes in the geometry of 434 CI cooperativity strongly interfere with its binding affinity and the structure of the TF-DNA complex (21,34), which indeed in our experiments resulted in rescue of growth with the λ-434 CI chimera. In contrast, with the λ CI-LacI chimera the growth reductions were even stronger than with λ CI, leading to growth arrest in *S. enterica* in rich and minimal media (Fig. 4B, Table S5). This opposing behavior of LacI and λ CI-LacI strongly supports our hypothesis that LacI binding affinity and basepair bias are conducive to low-specificity binding (Box 1B), but it is lacking the strong intermolecular cooperativity and oligomerization potential of λ CI (Fig. 1A). The chimera, however, combines these attributes, leading to strong interference with cell growth.

In order to determine if the non-cognate binding effects involved (i) a few essential, or (ii) many distributed, regions of the chromosome we performed ChIP-sequencing for λ CI in *E. coli* and *S. enterica*. In *E. coli* the data did not reveal strong peaks for any genomic site, but rather indicated weak

binding at numerous sites all over the chromosome (Fig. 5A, S4A, Table S6). Note that all of the regions plotted in Fig. 5A are significantly enriched in the presence of λ CI, but they only appear at a more lenient cutoff than typically used for strong binding (see Methods). In *S. enterica* we found both, distributed weak binding as well as a broad peak (indicating substantial binding in several adjacent genes). Interestingly, this broad peak corresponds to prophage regions on the genome that seem to provide binding hotspots for λ CI (Fig. 5B, S4B, Table S6). As such a binding hotspot was absent in *E. coli*, but λ CI still showed a growth defect, we did not consider this finding necessary to qualitatively explain our results (although it could account for the stronger growth defects seen in *S. enterica*). Further, none of the apparent peaks for either genome encoded a gene that is essential or obviously beneficial in minimal media conditions.

Using a simple thermodynamic model to predict λ CI binding across the bacterial genome we found a surprisingly high degree of correlation with the number of reads from ChIP-sequencing (Fig. 5C,D), especially given that these models generally perform poorly for low affinity sites (35). Even more surprising, we found a comparable prediction (Fig. 5C,D, S5) with an energy matrix that conserved only the overall λ CI preference for the basepair composition (see Methods). This result could not be explained by nucleotide composition bias in the ChIP-sequencing experiments (Fig. S6), but shows that a large part of the correlation between predicted binding and ChIP-sequencing reads can be explained by the overall genomic basepair bias. Correct basepair composition bias in the DNA sequences could provide λ CI with sufficient recognition pattern to bind with low specificity. In agreement with this hypothesis, the GC bias of the stronger λ CI cognate operators (O_{R1}, O_{R2}, O_{L1} and O_{L2}) is 52.94%, which is very close to that of the *S. enterica* genome (52.2%) and only slightly higher than that of the *E. coli* genome (50.8%). However, the residual sequence-dependent contribution beyond the basepair composition bias is still highly significant in *E. coli* and weakly significant in S. enterica (as determined by Monte-Carlo permutation tests for significance; Fig. 5C,D, S7). Overall, our results indicate substantial non-cognate binding due to sequence-dependence and basepair bias, which has been also reported, for example, for NAPs (9,10). Non-cognate binding is facilitated by repressor oligomerization (22,36), and distributed over the thousands of low-specificity λ CI binding sites, known to be present in the *E. coli* genome (29). These findings agree with previous studies on non-cognate binding of λ CI and other prokaryotic TFs (2,3,23,37).

Low-specificity binding leads to arrest of cell division

The distributed non-cognate DNA binding demonstrated by the ChIP-sequencing data for λ CI is in agreement with the observation that increasing concentrations of repressor gradually increases the magnitude of the growth reduction seen in Fig. 2B. Additionally, the dependence on growth media and induction timing indicates that DNA concentration – or rather the ratio between repressor and DNA – might play a role. If cell doubling time is slower than the time needed for DNA replication and cell division (~60min. in *E. coli* (38) and ~50min. in *S. enterica* (39), which is close to our observed doubling time in minimal media: ~63min. and ~58min. respectively), each bacterial cell contains on average only one chromosome. At faster growth, replication cycles are overlapping and daughter cells inherit 2-8 origins at birth, together with partially replicated chromosomes (38). Hence, the richer the medium and the faster the growth, the more DNA will be available (~2-fold for rich versus minimal media in our experiments) to titrate away potentially detrimental non-cognate binding TFs. In agreement with previous studies (40), the number of proteins at fast and slow growth were very similar (see Methods: Protein quantification), leading to a decreased protein concentration in rich media as cells become larger (this is particularly true for proteins expressed from plasmids (40)). Similarly, cells that are induced during the lag or early-exponential phase (after 1-2 doublings) will only have on average one chromosome as they did not inherit partially replicated chromosomes from their mothers and grandmothers yet.

Thus we set out to test the titration hypothesis by introducing a high copy number plasmid carrying four cognate λ CI binding sites into *E. coli* cells with inducible λ CI (Fig. S8A), which should reduce the number of free λ CI dimers available for non-cognate binding by about one half (see Methods). Although the expression of λ CI was still detrimental, growth was ~20% faster than for cells without the operators (Fig. S8B). Hence, titration of λ CI alleviates the growth reduction – likely even more so if additional chromosomal DNA is present (e.g. at faster growth), which provides many more potential binding sites with low specificity (29). For P22 C2, which is more discerning in its DNA binding targets (24), partially replicated chromosomes would provide less titration, thus explaining why later induction does not rescue growth. The titration phenomenon is reminiscent of growth bistability in drug resistant bacterial cells, which is caused by feedback between the growth rate and the speed of counteracting

toxic agent (41). In our system, the repressors can be seen as 'toxic agents', which are 'counteracted' by dilution if cells manage to start growing, or are growth-arrested if they are not able to dilute the repressors fast enough.

The titration hypothesis together with our ChIP-sequencing results implies that the overall ratio of chromosomal DNA to repressor proteins is a crucial factor determining the growth effects. This suggests that non-cognate binding might interfere with global cellular functions, like DNA replication or cell division, which we investigated using fluorescence microscopy of *E. coli* cells expressing λ CI. First, we imaged cells expressing a SeqA-GFP fusion protein, which is an indicator of replication fork progression (42). λ CI-expressing cells generally formed long filaments, suggesting an inhibition of cell division in these cells, even though the numerous fluorescent dots revealed ongoing replication (Fig. 6A, S9). Most filamentous cells showed low induction of the stress response promoter *PsulA* (Fig. 6B, S10), which is unlikely to induce sufficient self-cleavage (i.e. inactivation) of the repressor molecules (43) - particularly because λ CI becomes a poor substrate for self-cleavage at higher concentrations (44) - or to inhibit cell division substantially (45). Cell division can also be hindered by the presence of DNA at mid-cell (46). Fluorescence microscopy of cell membrane and DNA shows that in many filamentous cells DNA is located mid-cell, often on top of the established division septum (Fig. 6C, S11). This suggests that it is not FtsZ-ring formation, but a subsequent step in the cell division cascade that is disrupted. However, some filamentous cells manage to divide after growing to substantial length, as FtsZ-ring formation starts to occur at quarter points (47). Indeed, overexpression of FtsZ together with λ CI rescues growth entirely (Fig. S12), likely by forming additional, non-central division septa, which could produce viable cells as filamentous cells often contain additional chromosomes, which are distributed across the cell (Fig. 6C). As the *ftsZ* operon region was not enriched among ChIP-sequencing reads, it is unlikely that λ CI interferes with FtsZ expression directly through transcriptional interference.

Discussion

We investigated the consequences of limited specificity in molecular recognition of DNA by proteins, using four different, but related, phage repressors and a bacterial repressor, which produced a wide range of effects on host cell growth, from high to no fitness costs, and higher costs either in the native

or the non-native host (Fig.1B, 4B). Taking advantage of the rich and well-established genetics and biochemistry of the classic bacteriophage repressor λ CI, we found that its fitness cost results from cooperative, low-specificity binding, which interferes with growth by inhibiting cell division. The abundance of low-specificity binding sites in eukaryotic genomes has been shown to play a critical role in gene regulation, potentially increasing robustness and specificity (11). Our data, however, support the hypothesis that binding strategies of prokaryotic TFs are under selection to avoid low-specificity binding to the genomic background (29) and highlight the fundamental differences in gene regulatory design between prokaryotes and eukaryotes, and therefore differing evolutionary constraints (48).

For prokaryotes – in contrast to eukaryotes – TF target sites are sufficiently long to allow specific recognition of single operators (48). Mismatches with the preferred target sequence lead to progressive loss of binding, but the speed of this loss can vary substantially between TFs (24). λ CI, which shows strong operator binding (offset), low mismatch penalties (energy matrix) and strong cooperativity, is likely to be a rather promiscuous binder (Box 1B, (24)) and indeed induced a high fitness cost due to distributed low-specificity binding. For P22 C2, the lower offset and higher mismatch penalties make it a more specific binder (Box 1B, (24)), producing a significant cost only in non-native host cells. LacI, which shows similar binding characteristics as λ CI, only showed a strong fitness effect when coupled with λ CI's intermolecular cooperativity. Together, the mutant and chimera experiments (Fig. 4) demonstrate a significant contribution of cooperativity - and likely oligomerization - to the potential for low-specificity binding, which supports the theoretical finding that TF cooperativity does not strongly alleviate crosstalk when it stabilizes cognate as well as non-cognate binding (13). Hence, the lack of intermolecular cooperativity with LacI could be a sign of its adaptation to be highly specific, as it is one of the few single-target regulators in *E. coli* (18). Binding cooperativity, offset and TF concentration, can all serve to increase non-cognate binding of a TF (independently of the target motif preference) and the particular interplay between these factors has to be tuned by the cell to avoid fitness costs due to low-specificity interactions. Therefore, considering low-specificity binding is crucial in choosing TFs for synthetic biological systems in order to avoid global toxicity effects, as well as unwanted TF titration, which can affect target gene regulation (16,49).

The magnitude of the fitness cost depends on a repressor's ability to bind non-cognate DNA at low specificity, as well as on a repressor's relative ratio to the total amount of DNA within the cell. Slow cell growth compounds the effect as cells contain less DNA but accumulate more proteins than at fast growth (40). Additionally, stress tolerance could be higher under optimal growth conditions as found in rich media (50). It does not seem likely, however, that media-specific genes are targeted, as ChIP-sequencing generally revealed distributed, low-specificity binding all over the chromosome (Fig. 5). Rather, inhibition of cell division seems to result at least partially from nucleoid localization at mid-cell. Clearance of the division site is impeded if sister chromosomes fail to segregate (46,51), which could be caused by the formation of "bridges" between cooperatively bound repressors, holding the sister chromosomes together. Generally, more compact nucleoids are more efficient at preventing the formation of division septa in the same area (52). Intriguingly, a very similar growth effect has been found with H-NS, one of the NAPs responsible for chromosome organization and compaction in *E. coli* (9,10): H-NS overproduction drastically reduced cell viability, which seemed to be related to the formation of higher-order H-NS oligomers (53) – a state that is favoring its ability to form bridges between DNA regions (54). As cell growth, shape, division and DNA replication are thought to be tightly linked in complex and poorly understood ways (55), a mechanistic explanation of the observed division inhibition is presently not possible, but prior studies on H-NS combined with our findings make the case for the existence of general constraints on DNA-binding proteins.

Our results suggest that the inherent ability of DNA-binding proteins to occupy non-cognate DNA regions can pose, in addition to potential regulatory interference (16), a substantial challenge for host cell fitness overall – particularly considering facilitating conditions like cellular crowding (56), horizontal gene transfer (57) and mutations that alter the binding specificity of a protein. This challenge stems from the fundamental limits to molecular recognition that are set by the biophysics of molecular interactions (13) and could lead to various non-cognate effects on the physiology of the cell. The five transcription factors used here show a variety of non-cognate effects, which could indicate different selection pressures that have been acting on their binding and cooperativity characteristics as well as different potential for being tolerated in a particular environment. There might also be different mechanisms underlying the growth phenotype, as suggested by the ChIP-sequencing results for λ CI: the additional strong peak region found in *S. enterica* could indicate an

additional effect of high-affinity sites and explain the stronger growth defect as compared to *E. coli* (especially at low repressor concentrations). A potential for high fitness costs, even in native environments, as seen here with λ CI and 434 CI, can limit the number and binding affinity of promiscuous DNA-binding proteins in the cell (13). More specific binders such as P22 C2 might only be detrimental in non-native environments. Hence, the influx of foreign genes through horizontal gene transfer could be considerably impaired through non-cognate binding effects, as both are likely to occur under slow growth conditions. As the phage repressors we used originate from temperate phages, interference with host cell growth can limit their potential host range with regard to successfully establishing lysogeny. Considering that phage repressor concentrations are kept low during lysogenic cycles, the selection pressure to reduce low-specificity binding might generally be weak, which would explain the diversity in fitness impacts we observed with related phage repressors. More generally, experimentally uncovering the fundamental biophysical constraints imposed by low-specificity bindings of TFs is difficult, as TFs with many specific binding targets need to recognize a diversity of sequences and by default affect many cellular functions, while single target TFs are a very few (18). This is what ultimately motivated our choice of focusing our experiments on phage repressors and LacI.

We experimentally demonstrated for the first time that low specificity in biomolecular recognition can constitute a limiting factor for cellular function and evolution due to the fundamental biophysical constraints on protein-DNA interactions. However, these costs could be counter-balanced by increased TF robustness to target site mutations or higher evolvability, precisely because interactions can be formed at low specificity (24,58). For example, a TF could co-opt regulation of a non-cognate gene - even if only to a small degree – that provides an advantage in a certain environment, which can subsequently be refined by evolution. This opens up a wider question about the interplay of costs and benefits of low-specificity molecular interactions, especially when these interactions also serve as drivers of evolution.

METHODS

Plasmids and strains

The phage repressors λ CI, P22 C2, 434 CI, HK022 CI or the bacterial repressor LacI were cloned under the control of a $P_{LtetO-1}$ promoter in a low copy number kan^R plasmid (pZS)(59). The plasmids were then transformed into either MG1655 derived *E. coli* cells, which are deleted for the *lac* operon

(strain BW27785, CGSC#: 7881)(60), or into LT2 derived *S. enterica* cells with a tetracycline cassette inserted in the P22 attachment site (LT2 attP22::*tetRA*). In control experiments the phage repressor on the plasmid was replaced by a fluorescence marker gene (*gfp*). The low copy number plasmids used (pZS21) are under stringent replication control, which is linked to chromosome replication (61). Titration of λ CI was tested by transforming *E. coli* cells containing the pZS21- *λ cI* plasmid with a compatible, high-copy number pZE plasmid (50-70 copies)(59), which carries the natural λ CI operators O_{R1}, O_{R2}, O_{L1} and O_{L2}, i.e. 200-280 operators per cell – although the copy number of this plasmid's ori was originally documented at 25-30 copies per cell (62), hence there might only be about 100-120 operators per cell. At 25ng aTc induction, there are about 500 λ CI dimers per cell (see below), which reduces the number of dimers available per cell either by half or by one fifth.
For the competition experiments we introduced a constitutive fluorescent marker *venus-yfp* (63) into the low-copy pZS plasmid carrying *lacI* (for a detailed description see below).
In order to test for misfolding of repressor proteins, we used a high copy number plasmid containing a chaperone gene (*tig* (64)), which is native to *E. coli* and *S. enterica*, under the control of a P_{Lac} promoter from the ASKA(-) library (65).
To monitor induction of the stress response, we used a strain with a fast-maturing yellow fluorescent protein (YFP)(63) fused to the promoter of *sulA* (*PsulA-yfp*), which was placed on the chromosome using lambda red recombineering (66). *SulA* is strongly upregulated as a part of the stress response (67). The *PsulA-yfp* strain was then transformed with the pZS21- *λ cI* plasmid. We checked induction of the reporter by exposing cells to UV light for 30 seconds.
We used a SeqA-GFP translational fusion under the control of the natural *seqA* promoter to monitor replication as SeqA binds hemi-methylated GATC sequences in the wake of the advancing replication fork, marking newly synthesized DNA (42). In order to avoid unnaturally high *seqA* expression that could influence replication progression, the fusion protein was inserted into the HK022 attachment site on the *E. coli* chromosome using CRIM plasmids (68).
Additional expression of FtsZ, the major cell division protein, was performed by cloning *ftsZ* downstream of *λ cI* as a transcriptional fusion (i.e. putting it also under the control of aTc induction).

λ CI mutants

Based on previous studies, we cloned three different λ repressor mutants into the same low copy number plasmid (pZS) under the control of a $P_{LtetO-1}$ promoter: (i) a repressor mutant that is defective in its ability to bind DNA (N52D)(32); (ii) a repressor mutant that cannot form dimers (S228N)(30) and hence not bind DNA effectively anymore; and (iii) a mutant that can dimerize but not form higher-order oligomers, i.e. that cannot bind cooperatively (Y210N)(30). Mutations were introduced using site-directed mutagenesis. The function and stability of the proteins has been shown previously (30,32), as well as that their production levels are not different from wildtype repressor (30).

λ CI chimeras

Chimeric λ CI repressors were constructed based on literature describing a chimeric λ CI-434 repressor (69) and a chimeric 434-P22 repressor (70) by changing the recognition (i.e. DNA binding)

helix of λ CI to either the one of the 434 CI repressor or of the LacI repressor. This was done by introducing the following changes for the λ CI-434 chimera: G44T, S46Q, G49E, A50Q; and for the λ CI-LacI chimera: G44S, Q45Y, S46Q, G49S, A50R. (Note, that Q45 was not changed in the λ CI-434 chimera because both repressors contain the amino acid Q at this position.)

Growth measurements

All cells were grown overnight at 37°C in M9 medium supplemented with 0.2% glucose and 50µg/ml kanamycin (except specified differently). Cultures were used to dilute (1:100) 6 replicates without inducer and 6 replicates with 25ng aTc in 96 well plates and were grown at 37°C under shaking at 220 rpm. Populations were measured (OD_{600}) every 30min or every 60min using Biotek H1 plate reader for 10h. Population growth was also measured in LB, or M9 medium supplemented with 0.5% Casamino acids and either 0.5% glycerol or 0.2% glucose. Where indicated inducer concentration was changed to 1, 2, 3 and 4ng aTc - which was chosen in a way that the concentration range was covered as uniformly as possible, given that the $P_{LtetO-1}$ promoter has a very steep induction curve (59) - and the induction time was varied from the inoculation time point (0h) to 2h or 4h post-inoculation (early- and mid-exponential phase). The chaperone gene was induced using 1mM IPTG and Fis-GFP was expressed by adding 0.1mM IPTG to the medium.

Growth reduction was measured as the normalized difference between the areas under the growth curves (from the point of induction for 8.5h, i.e. including lag phase but not stationary growth) in the presence and absence of repressor ($\frac{AUC_{no\ repressor} - AUC_{repressor}}{AUC_{no\ repressor}}$). Hence, a ratio of 0 corresponds to wildtype growth even in the presence of repressor and a ratio of 1 corresponds to no growth at all in the presence of repressor. We used the area between the growth curves as opposed to growth rates, because growth rates were hard to define under conditions with strong growth reductions: growth did no longer show an exponential increase and the calculated rate was strongly dependent on the specific time points used for determination. Furthermore, growth was generally slowing down over time, and the repressor binding effects were not strictly limited to the 'exponential' growth phase. Hence, a maximal growth rate would not capture all the effects and we chose to use a growth measure that integrates over all of the growth phases and gives an impression of the general decrease in fitness.

DNA quantification

The relative amount of DNA within the cells in rich and minimal media was determined using the Wizard® Genomic DNA Purification Kit. We grew cells in LB or M9 with glucose to an OD of 0.2-0.3 and extracted genomic DNA using the purification kit. The amount of DNA was quantified by nanodrop and normalized by OD, which gave a 2-fold higher amount of DNA in rich media than in minimal media. This is an estimate of the minimal difference in DNA concentration as cells are slightly bigger in LB than in M9- (Fig. S11).

Protein quantification

Repressor concentration was determined using the Promega Nano-Glo HiBiT Lytic Detection System. The HiBiT peptide tag was attached at the N-terminal (which is involved in DNA binding) using a $(GGGS)_2$ linker (sequence: GGTGGTGGTTCTGGTGGTGGTTCT) to assure accessibility of the tag for interaction with the detection reagent. Briefly, cells induced (1ng or 25ng aTc) for expression of wildtype repressor or repressor with the HiBit tag were grown to early exponential phase in minimal media with glucose or rich media (LB), pelleted and frozen. Cells were resuspended in media and supplemented with 0.1 culture volume of PopCulture Reagent (Sigma Aldrich), 10^{-3} culture volume Benzonase Nuclease (Sigma Aldrich) and $0.5*10^{-3}$ culture volume lysozyme (Sigma Aldrich). Cells were lysed for 30min at room temperature and then kept on ice. A protein standard (Promega) was added to the non-tagged cells as a known reference of protein concentration to luminescence output. Samples were mixed 1:1 with HiBit enzyme mixture and measured in white plates after shaking (in the dark) for 15minutes. Dilution series of tagged repressor and protein standard were measured in a Tecan platereader (Spark 10M) with an integration time of 1.5 seconds.

Repressor protein numbers in minimal media with glucose gave about 500 dimers per cell at 25ng aTc and about 50 dimers per cell at 1ng aTc induction (as compared to ~125 dimers of λ CI in lysogenic cells (27)). Similarly, we found around 500 dimers per cell at 25ng aTc in rich media. Note, that the fitness reduction seen for λ CI concentrations at >1ng aTc induction (Fig. 2B) does not directly translate to lysogenic cell fitness as it is masked by phage induction, superinfection and superinfection exclusion (71) effects.

Competition assays

In order to distinguish between strains carrying a phage repressor (showing growth reduction) or the bacterial repressor LacI (not showing growth reduction), we used the pZS plasmid carrying *lacI* (under the control of $P_{LtetO-1}$), together with a constitutively expressed fluorescence marker (*venus*) cloned in the opposite direction upstream of $P_{LtetO-1}$. Venus was expressed from a mutated version of P_R, which abolishes *λ* CI binding affinity in the O_{R1} operator (72,73) (the O_{R1} sequence used is: TGCCTTAATACTGGATA) – and does not contain O_{R2} or O_{R3} - and is therefore constitutive ('$P_{constitutive}$'). The fluorescence marker was placed on the LacI-carrying plasmid because LacI only had a minor growth effect in *S. enterica* and none in *E. coli* (Fig. 4B, Table S5), which did not lead to filamentation and hence did not affect fluorescence due to cell morphology changes. The lack of growth reduction due to LacI expression agrees with previous findings that GalR (and possibly other members of the GalR family like LacI) seem to have evolved to have fewer low-specificity sites across the chromosome (18,29). There was only a slight fitness cost due to the presence of the fluorescence marker on the plasmid (leading to a selection coefficient of 0.05 and 0.09 in *E. coli* and *S. enterica* for the LacI-carrying plasmid without the marker over the one with the marker), which however only strengthens our findings that LacI-expressing cells increase in competition with phage repressor-expressing cells despite this cost.

A single colony for each host strain (*E. coli* or *S. enterica*) – plasmid (pZS21-*lacI,* pZS21- *λ cI*, pZS21-*P22 c2*, pZS21- *λ cI* dimerization mutant, pZS21-*lacI*-$P_{constitutive}$-*venus*) combination was picked from a freshly streaked plate and grown overnight in minimal media supplied with 0.2% glucose and 50µg/ml

kanamycin. Strains containing a phage repressor plasmid were mixed 1:1 with a strain carrying pZS21-*lacI*-$P_{constitutive}$-*venus*, diluted 1:100 into fresh medium and grown in 96-wellplates for 10h. Fluorescence was measured every 30min. and compared between cultures that were induced with aTc (at concentrations as indicated, either 1, 2, 3, 4 or 25ng) and cultures that were not induced. This means that we compared the abundance of fluorescent cells (i.e. abundance of LacI-carrying cells) between cultures expressing and not expressing the repressor.

Selection coefficients were calculated using $\ln[(R^+_t/R^-_t)/(R^+_0/R^-_0)]$, where R^+_t and R^-_t represent fluorescence measurements (as a proxy for relative LacI-expressing cell density) of cells with and without inducer aTc (presence or absence of repressor expression) at time t=10h respectively, and R^+_0 and R^-_0 represent fluorescence measurements at the beginning of the experiment.

Microscope fluorescence measurements

A Nikon Ti-E microscope equipped with a thermostat chamber (TIZHB, Tokai Hit), 100× oil immersion objective (Plan Apo λ, N.A. 1.45, Nikon), cooled CCD camera (ORCA-Flash, Hamamatsu Photonics) and LED excitation light source (DC2100, Thorlabs) was used for the microscopy fluorescence measurements of PsulA-Yfp and SeqA-Gfp. The microscope was controlled by micromanager (https://micro- manager.org). The cells were grown overnight in minimal media with glucose, diluted 1:100 in fresh media and grown to early exponential phase in the presence of the inducer aTc. YFP or GFP fluorescence (where appropriate), RFP fluorescence (for image correction) and phase contrast images were taken simultaneously at 3-min time-lapse intervals. Multiple patches of cells were monitored in a single experiment. A custom macro of ImageJ (http://imagej.nih.gov/ij/) was used for image analysis.

Imaging of cell membranes and DNA positioning was done using a Leica DMI6000B (inverted) microscope with an Andor iXon EM CCD camera (front illuminated, 8x8 square micron pixel size) and a 100x 1,47Na Oil HCX Plan Apo objective, giving an effective pixel size of 64nm/pixel. Images were acquired using 405(20)nm and 561(10)nm laser excitation for blue (Hoechst) and red (NileRed) dyes respectively. The cells were grown overnight in minimal media with glucose or LB, diluted 1:100 in fresh media and grown to early exponential phase in the absence or presence of the inducer aTc. After addition of both dyes (Hoechst at 10ug/mL and NileRed at 1ug/mL), cells were shaken at room temperature for one hour and imaged in drops of the respective growth media. Images were deconvolved using Huygens Professional (version 4.5) and further analyzed using ImageJ.

ChIP-sequencing

To perform ChIP-sequencing experiments, λ CI was cloned with an HA-Tag at the carboxy-terminal end and transformed into both host strains. HA-tagged λ CI showed the same growth phenotype as wildtype in both bacterial strains (Fig. S13). Samples from strains grown in the presence or absence of λ CI were prepared according to Waldminghaus & Skarstad (2010)(74); library preparation and Illumina Sequencing was performed at the VBCF NGS Unit (www.vbcf.ac.at). The obtained data was analyzed using Galaxy and RStudio.

Peak calling was performed using custom R scripts modified from Santhanam et al. (75). Briefly, the genome was computationally partitioned into non-overlapping shorter fragments, typically spanning a few kbs to account for local biases arising from sequence content and immuno-precipitation (76,77). Peak calling was performed within these fragments using partially overlapping (50% overlap) windows of 100bp. For each window, we calculated strand-specific enrichment as the log-ratio of the scaled read coverage between the sample and control ChIP-seq experiments while permitting a maximum of 5 reads to be mapped to the same genomic coordinates. We calculated strand-wise p-values for enrichment by first resampling scaled read coverage within each fragment and then randomly partitioning them to calculate enrichments. Finally, we identified bound regions to be those with positive enrichment scores on both strands with a Benjamini-Hochberg false-discovery rate of less than 30% as we were looking for binding of low specificity and ChIP-binding data was previously found to be highly informative for a wide range of specificity profiles (78). For the regions that showed significant enrichment in this analysis we plotted the read-depth across the genome in Fig. 5 (A,B) and for comparison we plotted the read-depth for not significantly enriched regions in Fig. S4. As control for our ChIP-sequencing procedure and analysis we used antibodies against SeqA, which gave the expected peaks as published previously (74).

We calculated the nucleotide composition of the sequences underlying enriched regions in ChIP-seq data for both bacterial species (Fig. S6). In order to test for sequence composition bias in these enriched regions, we sought to test if the sequence compositions of the enriched regions were significantly different compared to the rest of the genome. To this end, we randomly selected 50 genomic regions with at least 5kbp distance between them. We then calculated the nucleotide composition of these randomly selected regions and by repeating this procedure 1000 times, generated a null distribution for sequence composition of randomly selected genomic regions. Similarly, we calculated the di-nucleotide composition (with 1 bp overlap) of the same randomly selected genomic regions and compared it to that of the enriched regions.

The number of reads within 1000bp windows was compared with the predicted binding by calculating binding energy at each genome position (using a sliding window approach) from the λ CI offset (i.e. the energy difference between the repressor being bound specifically to an operator and being free in solution (79)) and the energy penalty as given by the λ CI energy matrix (73). Smaller energies result in stronger binding, meaning positive energy penalties decrease binding affinity (note that negative penalties could increase binding over the one seen with λ CI wildtype operator sites). Binding strength was calculated using $1/(1+\exp(E-\mu))$, with E being the calculated binding energy, as described above and used in (24), and μ being the chemical potential, which we optimized to give the highest Spearman correlation fit (2.6 in *E. coli* and 2 in *S. enterica*). For comparison with the number of ChIP-sequencing reads, calculated binding strength was summed over the same genomic 1000bp regions (considering binding to both strands). In Fig. 5 (C,D) we plot a non-parametric, non-linear relationship estimate between the predicted binding energy and the ChIP-sequencing reads obtained from a series of conditional medians. To investigate the dependence of the correlation between the affinity predictions and the ChIP-sequencing reads on the structural versus the sequence information contained in the energy matrix, we repeated the analysis with i) a matrix of the same size that

conserves only the ACGT bias of the λ CI energy matrix (each row contains the average value of that row) or ii) matrices that had completely reshuffled entries. For the latter the average correlation was taken over 100 permutations.

To assess the importance of specific sequence information versus nucleotide (GC) bias, we used a Monte-Carlo permutation test: We calculated the difference between Spearman correlations of ChIP reads with binding prediction using the wildtype energy matrix vs binding prediction using the energy matrix that only conserves λ CI basepair bias, for the true ChIP read assignment, and 10^4 random read assignments (null distribution). We found an overall strongly significant difference in *E. coli* and lower significance in *S. enterica* (Fig. 5C,D), even though the effect size was small. This means that while most of the measured ChIP signal can be accounted for by a TF model that predicts binding based on the nucleotide content of genomic fragments alone, there is a small but highly significant residual ChIP binding signal that requires the full binding site preference (energy matrix), not just single nucleotide bias, to be explained. Further, we examined the influence of GC content by repeating the Monte-Carlo permutation test for genomic sequences of a specific GC %. Here, we found only a significant motif contribution for the 49% bin in *E. coli* (Fig. S7).

Additionally we used the offset and energy matrix for LacI (25,80) and P22 C2 (81) to predict binding and calculate the Spearman correlation with the λ CI ChIP-sequencing reads (Fig. S5). Basepair bias of the energy matrices was calculated as the sum of the average A and T preference minus the sum of the average G and C preference.

Statistical analysis

Collected data was tested for normality (Shapiro-test) and subsequently we compared mean OD_{600} or fluorescence expression values using t-tests with FDR correction for multiple comparisons in RStudio. T-tests were performed for four different time points between cultures grown in the presence and absence of inducer aTc (presence or absence of repressors) under indicated conditions. Error bars on growth reductions from AUC differences were obtained as 95% confidence intervals through bootstrapping by resampling the data at each time point 1000 times.

Spearman correlation was calculated for the fit between model predictions of binding strength and the number of obtained ChIP-sequencing reads per 1000bp window. P-values for WT and basepair bias energy matrix predictions were P<.001 and for the average over random matrices P<.01.

ACKNOWLEDGEMENT

We thank T. Bergmiller, R. Chait, K. Jain, M. La Fortezza and the ETH Zurich ScopeM facility for their support with fluorescence microscopy, B. Kavčič for his help with protein concentration measurements, and T. Friedlander and J. Crocker for useful discussions. C.I. is the recipient of a DOC Fellowship of the Austrian Academy of Sciences. Library preparation and sequencing of ChIP seq samples was performed by the Next Generation Sequencing Facility at Vienna BioCenter Core Facilities (VBCF), member of the Vienna BioCenter (VBC), Austria.

AUTHOR CONTRIBUTIONS

C.I., C.F., G.T. and C.C.G. conceived the study together. C.I. designed the experiments and carried them out together with C.F. (growth experiments) and T.W. (ChIP-sequencing). C.I. analyzed the data with input from F.M.P. and B.S.. C.I. wrote the initial draft of the manuscript and revised it together with the rest of the authors.

CONFLICT OF INTEREST

Authors declare no competing financial interests.

REFERENCES

1. Hopfield JJ. Biosynthetic Processes Requiring High Specificity. Pnas. 1974;71(10):4135–9.
2. Mirny L, Slutsky M, Wunderlich Z, Tafvizi A, Leith J, Kosmrlj A. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. J Phys A Math Theor [Internet]. 2009 Oct 30;42(43):434013. Available from: http://stacks.iop.org/1751-8121/42/i=43/a=434013?key=crossref.c3b19cfa3ebb32a80370e612c091028f
3. Elf J, Li G-W, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. Science [Internet]. 2007 May 25;316(5828):1191–4. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1141967
4. Flyvbjerg H, Keatch SA, Dryden DTF. Strong physical constraints on sequence-specific target location by proteins on DNA molecules. Nucleic Acids Res. 2006;34(9):2550–7.
5. von Hippel PH, Revzin A, Gross CA, Wang AC. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. Proc Natl Acad Sci U S A. 1974 Dec;71(12):4808–12.
6. Bakk A, Metzler R. Nonspecific binding of the OR repressors CI and Cro of bacteriophage lambda. J Theor Biol [Internet]. 2004 Dec 21;231(4):525–33. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022519304003285
7. Kao-Huang Y, Revzin A, Butler AP, O'Conner P, Noble DW, Von Hippel PH. Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. Proc Natl Acad Sci U S A [Internet]. 1977;74(10):4228–32. Available from: http://www.pnas.org/content/74/10/4228.short%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=431912&tool=pmcentrez&rendertype=abstract
8. Gerland U, Moroz JD, Hwa T. Physical constraints and functional characteristics of transcription factor-DNA interaction. Proc Natl Acad Sci U S A. 2002;99(19):12015–20.
9. Spurio R, Dürrenberger M, Falconi M, La Teana A, Pon CL, Gualerzi CO. Lethal overproduction of the Escherichia coli nucleoid protein H-NS: ultramicroscopic and molecular autopsy. MGG Mol Gen Genet. 1992;231(2):201–11.
10. Yamada H, Yoshida T, Tanaka K ichi, Sasakawa C, Mizuno T. Molecular analysis of the Escherichia coli has gene encoding a DNA-binding protein, which preferentially recognizes curved DNA sequences. MGG Mol Gen Genet. 1991;230(1–2):332–6.
11. Crocker J, Preger-Ben Noon E, Stern DL. The Soft Touch. In: Current Topics in Developmental Biology [Internet]. 1st ed. Elsevier Inc.; 2016. p. 455–69. Available from: http://dx.doi.org/10.1016/bs.ctdb.2015.11.018
12. Burger A, Walczak AM, Wolynes PG. Abduction and asylum in the lives of transcription factors. Proc Natl Acad Sci [Internet]. 2010;107(9):4016–21. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0915138107
13. Friedlander T, Prizak R, Guet C, Barton NH. Intrinsic limits to gene regulation by global crosstalk. 2015;
14. Cepeda-Humerez SA, Rieckh G, Tkačik G. Stochastic Proofreading Mechanism Alleviates Crosstalk in Transcriptional Regulation. Phys Rev Lett. 2015;115(24):1–5.
15. Grah R, Zoller B, Tkačik G. Nonequilibrium models of optimal enhancer function. Proc Natl Acad Sci U S A. 2020;117(50):31614–22.
16. Sasson V, Shachrai I, Bren A, Dekel E, Alon U. Mode of Regulation and the Insulation of Bacterial Gene Expression. Mol Cell [Internet]. 2012;46(4):399–407. Available from: http://dx.doi.org/10.1016/j.molcel.2012.04.032
17. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J [Internet]. 2017 Jul

14;11(7):1511–20. Available from: http://dx.doi.org/10.1038/ismej.2017.16
18. Shimada T, Ogasawara H, Ishihama A. Single-target regulators form a minor group of transcription factors in Escherichia coli K-12. Nucleic Acids Res. 2018;46(8):3921–36.
19. Ptashne M. Principles of a switch. Nat Chem Biol [Internet]. 2011 Aug [cited 2014 Jun 3];7(8):484–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21769089
20. Priest DG, Cui L, Kumar S, Dunlap DD, Dodd IB, Shearwin KE. Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and repressors. Proc Natl Acad Sci [Internet]. 2014;111(1):349–54. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1317817111
21. Donner AL, Carlson PA, Koudelka GB. Dimerization specificity of P22 and 434 repressors is determined by multiple polypeptide segments. J Bacteriol [Internet]. 1997 Feb [cited 2014 Jul 8];179(4):1253–61. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=178823&tool=pmcentrez&rendertype=abstract
22. Pray TR, Burz DS, Ackers GK. Cooperative non-specific DNA binding by octamerizing λcI repressors: a site-specific thermodynamic analysis. J Mol Biol [Internet]. 1998 Oct [cited 2017 Mar 1];282(5):947–58. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022283698920563
23. Bakk A, Metzler R. In vivo non-specific binding of λ CI and Cro repressors is significant. FEBS Lett. 2004;563(1–3):66–8.
24. Igler C, Lagator M, Tkačik G, Bollback JP, Guet CC. Evolutionary potential of transcription factors for gene regulatory rewiring. Nat Ecol Evol. 2018;2(October).
25. Vilar JM, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Curr Opin Genet Dev [Internet]. 2005 Apr [cited 2014 May 14];15(2):136–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15797196
26. Fattah KR, Mizutani S, Fattah FJ, Matsushiro A, Sugino Y. A comparative study of the immunity region of lambdoid phages including Shiga-toxin-converting phages : molecular basis for cross immunity . 2000;223–32.
27. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. Switches in bacteriophage lambda development. Annu Rev Genet [Internet]. 2005 Jan [cited 2014 May 24];39:409–29. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16285866
28. Schlechter RO, Remus DM, Remus-Emsermann MNP. Constitutively expressed fluorescent proteins allow to track bacterial growth and to determine relative fitness of bacteria in mixed cultures. bioRxiv. 2020;(Figure 1):1–5.
29. Chakrabarti J, Chandra N, Raha P, Roy S. High-affinity quasi-specific sites in the genome: How the DNA-binding proteins cope with them. Biophys J [Internet]. 2011;101(5):1123–9. Available from: http://dx.doi.org/10.1016/j.bpj.2011.07.041
30. Whipple FW, Kuldell NH, Cheatham L a, Hochschild a. Specificity determinants for the interaction of lambda repressor and P22 repressor dimers. Genes Dev [Internet]. 1994 May 15 [cited 2014 Sep 22];8(10):1212–23. Available from: http://www.genesdev.org/cgi/doi/10.1101/gad.8.10.1212
31. Weiss MA, Pabo CO, Karplus M, Sauer RT. Dimerization of the operator binding domain of phage lambda repressor. Biochemistry [Internet]. 1987 Feb 10 [cited 2014 Jun 5];26(3):897–904. Available from: http://dx.doi.org/10.1021/bi00377a034
32. Nelson HCM, Sauer RT. Interaction of mutant λ repressors with operator and non-operator DNA. J Mol Biol. 1986;192(1):27–38.
33. Mazumder A, Batabyal S, Mondal M, Mondol T, Choudhury S, Ghosh R, et al. Specific DNA sequences allosterically enhance protein-protein interaction in a transcription factor through modulation of protein dynamics: Implications for specificity of gene regulation. Phys Chem Chem Phys [Internet]. 2017;19(22):14781–92. Available from: http://dx.doi.org/10.1039/C7CP01193H
34. Guarnaccia C, Raman B, Zahariev S, Simoncsits A, Pongor S. DNA-mediated assembly of weakly interacting DNA-binding protein subunits: In vitro recruitment of phage 434 repressor and yeast GCN4 DNA-binding domains. Nucleic Acids Res. 2004;32(17):4992–5002.
35. Maerkl SJ, Quake SR. A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. Science (80-) [Internet]. 2007 Jan 12;315(5809):233–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17218526
36. Sarkar-Banerjee S, Goyal S, Gao N, Mack J, Thompson B, Dunlap D, et al. Specifically bound lambda repressor dimers promote adjacent non-specific binding. Saiz L, editor. PLoS One

[Internet]. 2018 Apr 2;13(4):e0194930. Available from: http://dx.doi.org/10.1371/journal.pone.0194930

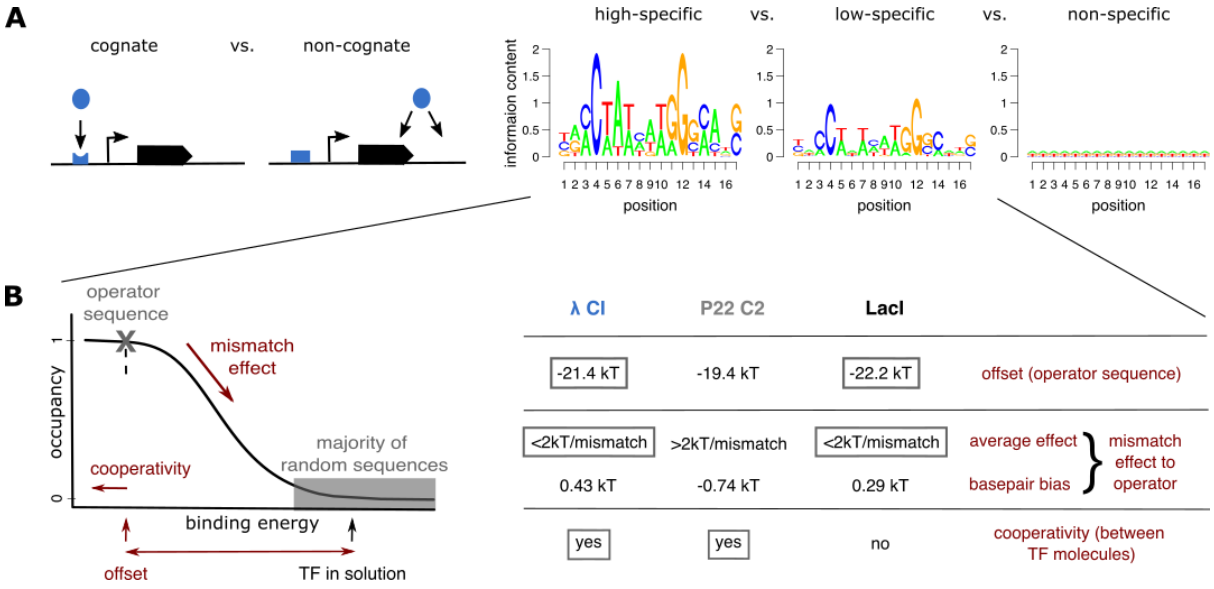
37. Chen Y, Golding I, Sawai S, Guo L, Cox EC. Population Fitness and the Regulation of Escherichia coli Genes by Bacterial Viruses. Waldor M, editor. PLoS Biol [Internet]. 2005 Jun 21;3(7):e229. Available from: http://dx.plos.org/10.1371/journal.pbio.0030229
38. Nordström K, Dasgupta S. Copy-number control of the Escherichia coli chromosome: A plasmidologist's view. EMBO Rep. 2006;7(5):484–9.
39. Cooper S, Ruettinger T. Replication of deoxyribonucleic acid during the division cycle of Salmonella typhimurium. J Bacteriol. 1973;114(3):966–73.
40. Klumpp S, Zhang Z, Hwa T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. Cell [Internet]. 2009 Dec;139(7):1366–75. Available from: http://dx.doi.org/10.1016/j.cell.2009.12.001
41. Deris JB, Kim M, Zhang Z, Okano H, Hermsen R, Groisman A, et al. The Innate Growth Bistability and Fitness Landscapes of Antibiotic-Resistant Bacteria. Science (80-) [Internet]. 2013 Nov 29;342(6162):1237435–1237435. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1237435
42. Waldminghaus T, Weigel C, Skarstad K. Replication fork movement and methylation govern SeqA binding to the Escherichia coli chromosome. Nucleic Acids Res. 2012;40(12):5465–76.
43. Sauer RT, Ross MJ, Ptashne M. Cleavage of the lambda and P22 repressors by recA protein. J Biol Chem. 1982 Apr;257(8):4458–62.
44. Phizicky EM, Roberts JW. Kinetics of recA protein-directed inactivation of repressors of phage λ and phage P22. J Mol Biol. 1980;139(3):319–28.
45. Arends SJR, Weiss DS. Inhibiting Cell Division in Escherichia coli Has Little If Any Effect on Gene Expression. J Bacteriol [Internet]. 2004 Feb 1;186(3):880–4. Available from: http://jb.asm.org/cgi/doi/10.1128/JB.186.3.880-884.2004
46. Cambridge J, Blinkova A, Magnan D, Bates D, Walker JR. A Replication-inhibited unsegregated nucleoid at mid-cell blocks Z-ring formation and cell division independently of SOS and the SlmA nucleoid occlusion protein in Escherichia coli. J Bacteriol. 2014;196(1):36–49.
47. Gullbrand B, Nordström K. FtsZ ring formation without subsequent cell division after replication runout in Escherichia coli. Mol Microbiol [Internet]. 2002 Jan 18;36(6):1349–59. Available from: http://doi.wiley.com/10.1046/j.1365-2958.2000.01949.x
48. Wunderlich Z, Mirny L. Fundamentally different strategies for transcriptional regulation are revealed by analysis of binding motifs. Nat Preced. 2008;2(i).
49. Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R. The transcription factor titration effect dictates level of gene expression. Cell. 2014;156(6):1312–23.
50. Pleška M, Qian L, Okura R, Bergmiller T, Wakamoto Y, Kussell E, et al. Bacterial Autoimmunity Due to a Restriction-Modification System. Curr Biol [Internet]. 2016 Feb 8;26(3):404–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26804559
51. de Boer PAJ. Advances in understanding E. coli cell fission. Curr Opin Microbiol. 2010;13(6):730–7.
52. Sun Q, Margolin W. Effects of perturbing nucleoid structure on nucleoid occlusion-mediated toporegulation of FtsZ ring assembly. J Bacteriol. 2004;186(12):3951–9.
53. Spurio R, Falconi M, Brandi A, Pon CL, Gualerzi CO. The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. EMBO J. 1997;16(7):1795–805.
54. Verma SC, Qian Z, Adhya SL. Architecture of the Escherichia coli nucleoid. Vol. 15, PLoS Genetics. 2019. 1–35 p.
55. Amir A, Männik J, Woldringh CL, Zaritsky A. Editorial: The Bacterial Cell: Coupling between Growth, Nucleoid Replication, Cell Division, and Shape Volume 2. Front Microbiol. 2019;10(September):1–3.
56. Zabet NR, Adryan B. The effects of transcription factor competition on gene regulation. Front Genet. 2013;4(OCT):1–10.
57. Baltrus DA. Exploring the costs of horizontal gene transfer. Trends Ecol Evol [Internet]. 2013;28(8):489–95. Available from: http://dx.doi.org/10.1016/j.tree.2013.04.002
58. Friedlander T, Prizak R, Barton NH, Tkačik G. Evolution of new regulatory functions on biophysically realistic fitness landscapes. Nat Commun. 2017;8(1).
59. Lutz R, Bujard H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res [Internet]. 1997 Mar 15;25(6):1203–10. Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=146584&tool=pmcentrez&rendertype=abstract

60. Datsenko K a, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A [Internet]. 2000 Jun 6;97(12):6640–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18686&tool=pmcentrez&rendertype=abstract
61. Kües U, Stahl U. Replication of plasmids in gram-negative bacteria. Microbiol Rev [Internet]. 1989 Dec;53(4):491–516. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2687680
62. Hershfield V, Boyer HW, Yanofsky C, Lovett MA, Helinski DR. Plasmid ColEl as a molecular vehicle for cloning and amplification of DNA. Proc Natl Acad Sci U S A. 1974;71(9):3455–9.
63. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat Biotechnol [Internet]. 2002 Jan;20(1):87–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11753368
64. Nishihara K, Kanemori M, Yanagi H, Yura T. Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. Appl Environ Microbiol [Internet]. 2000 Mar;66(3):884–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10698746
65. Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, et al. Complete set of ORF clones of Escherichia coli ASKA library (A complete set of E. coli K-12 ORF archive): unique resources for biological research. DNA Res. 2005;12(5):291–9.
66. Costantino N, Court DL. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. Proc Natl Acad Sci U S A [Internet]. 2003;100(26):15748–53. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=307639&tool=pmcentrez&rendertype=abstract
67. Friedberg EC, Walker GC, Siede W, Wood RD. DNA repair and mutagenesis. American Society for Microbiology Press; 2005.
68. Haldimann A, Wanner BL. Conditional-Replication, Integration, Excision, and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria. J Bacteriol [Internet]. 2001 Nov 1;183(21):6384–93. Available from: http://jb.asm.org/cgi/doi/10.1128/JB.183.21.6384-6393.2001
69. Marchetti A, Abril-Marti M, Illi B, Cesareni G, Nasi S. Analysis of the Myc and Max interaction specificity with λ repressor-HLH domain fusions. J Mol Biol. 1995;248(3):541–50.
70. Wharton RP, Ptashne M. Changing the binding specificity of a represser by redesigning an α-helix. Nature. 1985;316(6029):601–5.
71. Lin L, Bitner R, Edlin G. Increased reproductive fitness of Escherichia coli lambda lysogens. J Virol. 1977;21(2):554–9.
72. Flashman SM. NMutational Analysis of the Operators of Bacteriophage Lambda. 1978;73:61–73.
73. Sarai A, Takeda Y. Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. Proc Natl Acad Sci U S A [Internet]. 1989;86(17):6513–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=297874&tool=pmcentrez&rendertype=abstract
74. Waldminghaus T, Skarstad K. ChIP on Chip: surprising results are often artifacts. BMC Genomics [Internet]. 2010;11(1):414. Available from: http://www.biomedcentral.com/1471-2164/11/414
75. Santhanam B, Cai H, Devreotes PN, Shaulsky G, Katoh-Kurasawa M. The GATA transcription factor GtaC regulates early developmental gene expression dynamics in Dictyostelium. Nat Commun. 2015;6(May).
76. Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res. 2008;36(16):5221–31.
77. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9).
78. Tanay A. Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res [Internet]. 2006 Jun 29;16(8):962–72. Available from: http://www.genome.org/cgi/doi/10.1101/gr.5113606
79. Koblan KS, Ackers GK. Site-Specific Enthalpic Regulation of DNA Transcription at Bacteriophage λ OR. Biochemistry. 1992;31(1):57–65.
80. Barne SL, Belliveau NM, Ireland WT, Kinney JB, Phillips R. Mapping DNA sequence to

857 transcription factor binding energy in vivo. PLoS Comput Biol. 2019;15(2):1–29.
858 81. Hilchey SP, Wu L, Koudelka GB. Recognition of Nonconserved Bases in the P22 Operator by
859 P22 Repressor Requires Specific Interactions between Repressor and Conserved Bases *.
860 1997;272(32):19898–905.
861
862

FIGURE LEGENDS



Box 1. (A) In order to elicit an appropriate function, TFs have to recognize their cognate DNA sites (operators) among a large background of non-cognate sites, where binding is non-functional and potentially interferes with cellular programs. Binding of TFs to cognate or non-cognate sites can occur by recognizing a specific motif on the DNA (with specificity ranging from high to low, depending on the overlap of the target site with the TF's consensus sequence (8)) or through generic, non-specific interactions with any DNA sequence. **(B)** Binding with high specificity usually occurs to a TF-specific DNA consensus motif and the offset gives the binding energy (lower energy = stronger binding) to a single operator sequence relative to the unbound state (TF in solution). Random DNA sequences are usually located at the lower end of the sigmoid (gray box) with energy values similar to that of the TF in solution. Basepair mismatches with the operator sequence incur an energy penalty and increase the binding energy (i.e. weaken the binding) based on two features: specific effects of basepair mismatches and overall basepair bias (given here as average AT preference – average GC preference). Hence, for a TF with low offset, which is aided by high cooperativity, and an energy matrix characterized by small mismatch effects, many random sequences will not be far from the rise of the sigmoid, making higher occupancy at low specificity binding sites more likely. The table compares these three criteria for three well-characterized TFs, showing that all of them are fulfilled for λ CI, but only partially for P22 C2 and LacI (24,25,80,81), making λ CI an obvious candidate for substantial low-specificity non-cognate binding.

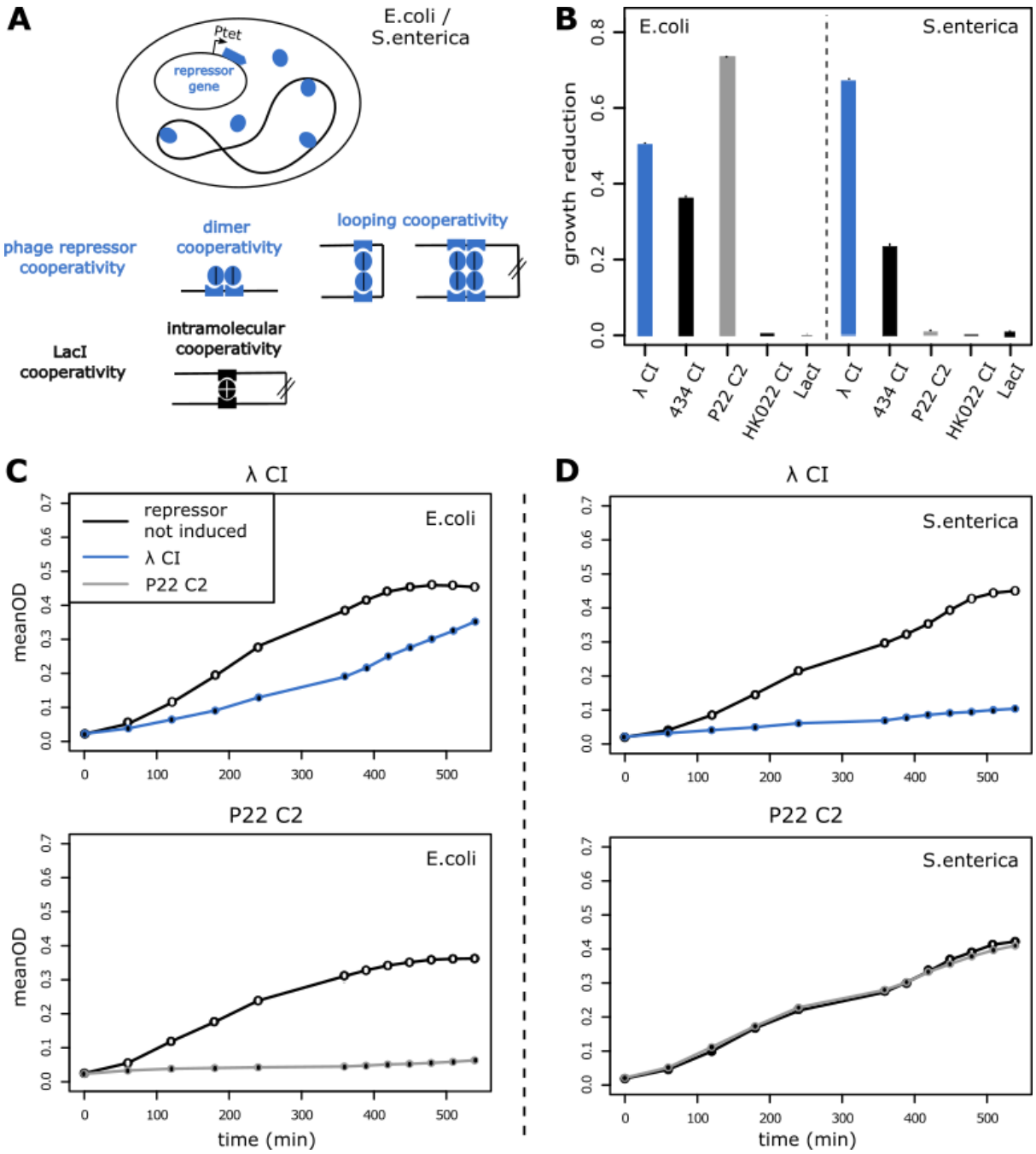


Figure 1. Growth reduction in the presence of repressors in minimal media with glucose. **(A)** The experimental model system with repressors being expressed from a plasmid and their binding cooperativity modes are shown. **(B)** Growth reduction as calculated by the normalized growth difference in the presence and absence of repressor are shown for λ CI, 434 CI, P22 C2, HK022 CI and LacI in *E. coli* and *S. enterica* cells. Error bars show 95% confidence intervals. **(C,D)** Curves show mean OD_{600} for *E. coli* (left) or *S. enterica* (right) cells in the presence (color) or absence (black) of **(C)** λ CI or **(D)** P22 C2; error bars show standard deviation over 6 replicates (black on color and white on black).

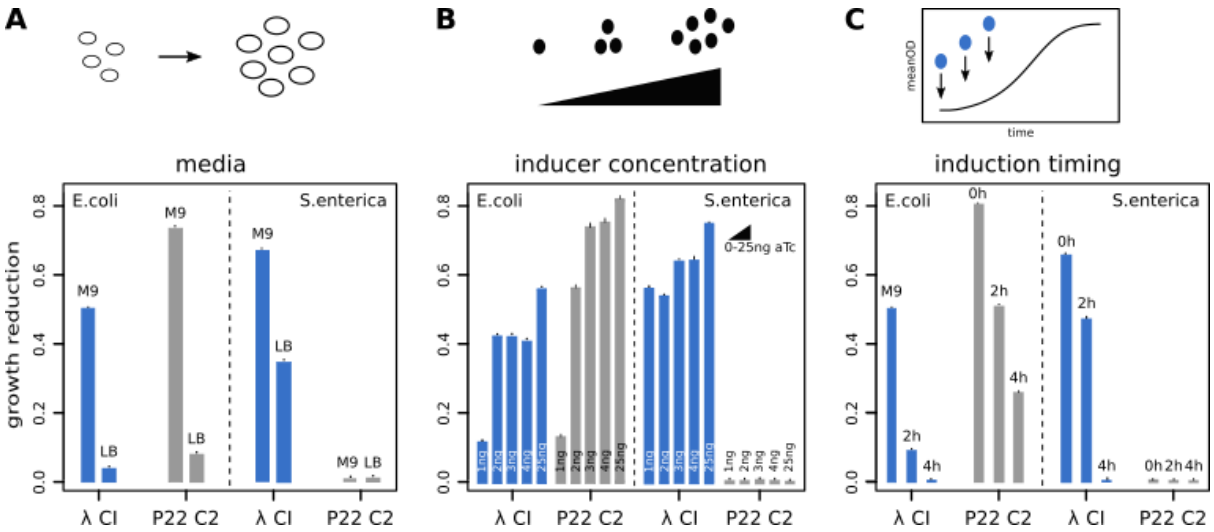


Figure 2. Effect of environment on repressor-dependent growth reduction.

Growth reduction as calculated by the normalized growth difference in the presence and absence of repressor are shown for λ CI (blue) and P22 C2 (grey) in *E. coli* and *S. enterica*. Error bars show 95% confidence intervals. **(A)** Cells were grown in minimal medium with glucose (M9) or rich media (LB) at full induction of repressors. **(B)** Inducer concentrations for repressor expression were varied from 1 to 25ng (very low to full induction (59)). **(C)** Induction time points of repressor expression were varied from lag phase (0h) to early- (2h) and mid- exponential phase (4h).

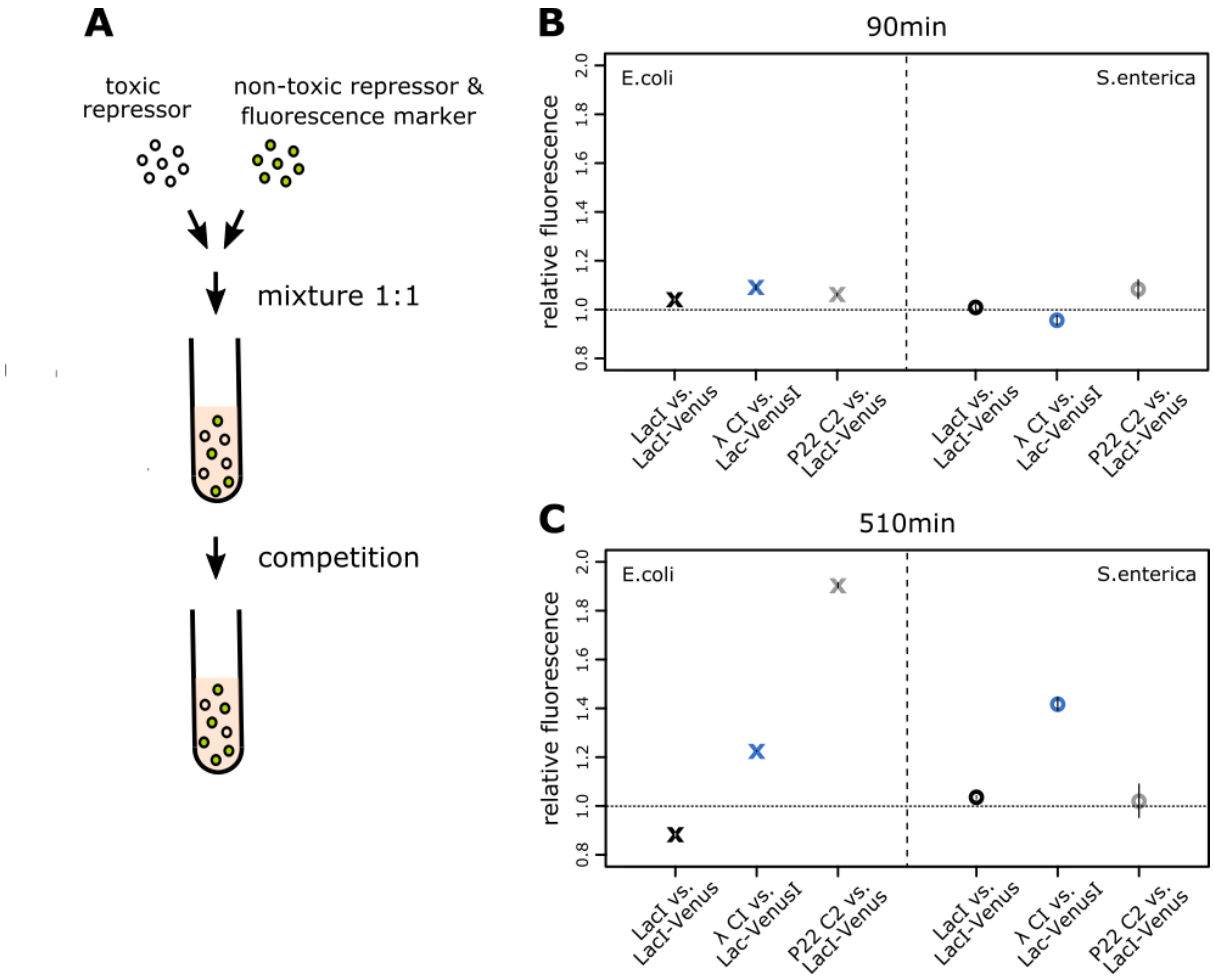


Figure 3. Competition assays reveal fitness cost of repressor expression.

(A) Cells containing plasmids with a repressor affecting growth (λ CI or P22 C2) or with a repressor not affecting growth (LacI) were mixed 1:1 with cells containing a plasmid with the repressor not affecting growth and a separately, constitutively expressed Venus marker (LacI-Venus). Competition was performed in minimal media with glucose and fluorescence was used as a measure of the relative change in the cells carrying LacI-Venus. **(B,C)** Relative fluorescence was calculated for *E. coli* (crosses) or *S. enterica* (circles) between induced and non-induced samples of cell mixtures of LacI-Venus together with LacI (control), λ CI or P22 C2 after **(B)** 90min or **(C)** 510min of competition; error bars show relative errors. Selection coefficients (for calculation see Methods) after 10h were 0.27 (*E. coli*) and 0.29 (*S. enterica*) for λ CI and 0.67 (*E. coli*) and -0.06 (*S. enterica*) for P22 C2.

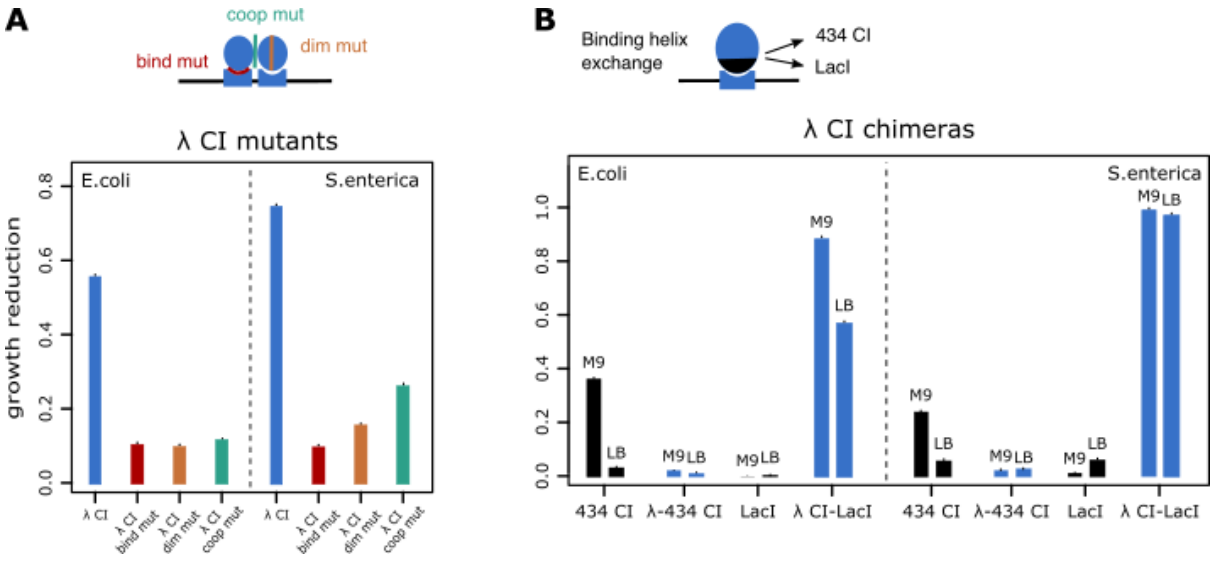


Figure 4. Effect of mutants and chimeric proteins on repressor-dependent growth reduction. Growth reduction as calculated by the normalized growth difference in the presence and absence of repressor are shown in *E. coli* and *S. enterica*. Error bars show 95% confidence intervals. **(A)** Growth reduction for λ CI wildtype (blue) is compared to a λ CI binding mutant (red), a λ CI dimerization mutant (orange) or a λ CI cooperativity mutant (turquoise) in minimal media. **(B)** Growth reduction of the phage repressor 434 CI and the bacterial repressor LacI (black) compared to a chimera of the λ CI protein containing the binding specificity of either 434 CI (λ-434 CI) or LacI (λ CI-LacI) (blue) is shown for minimal (M9) and rich media (LB).

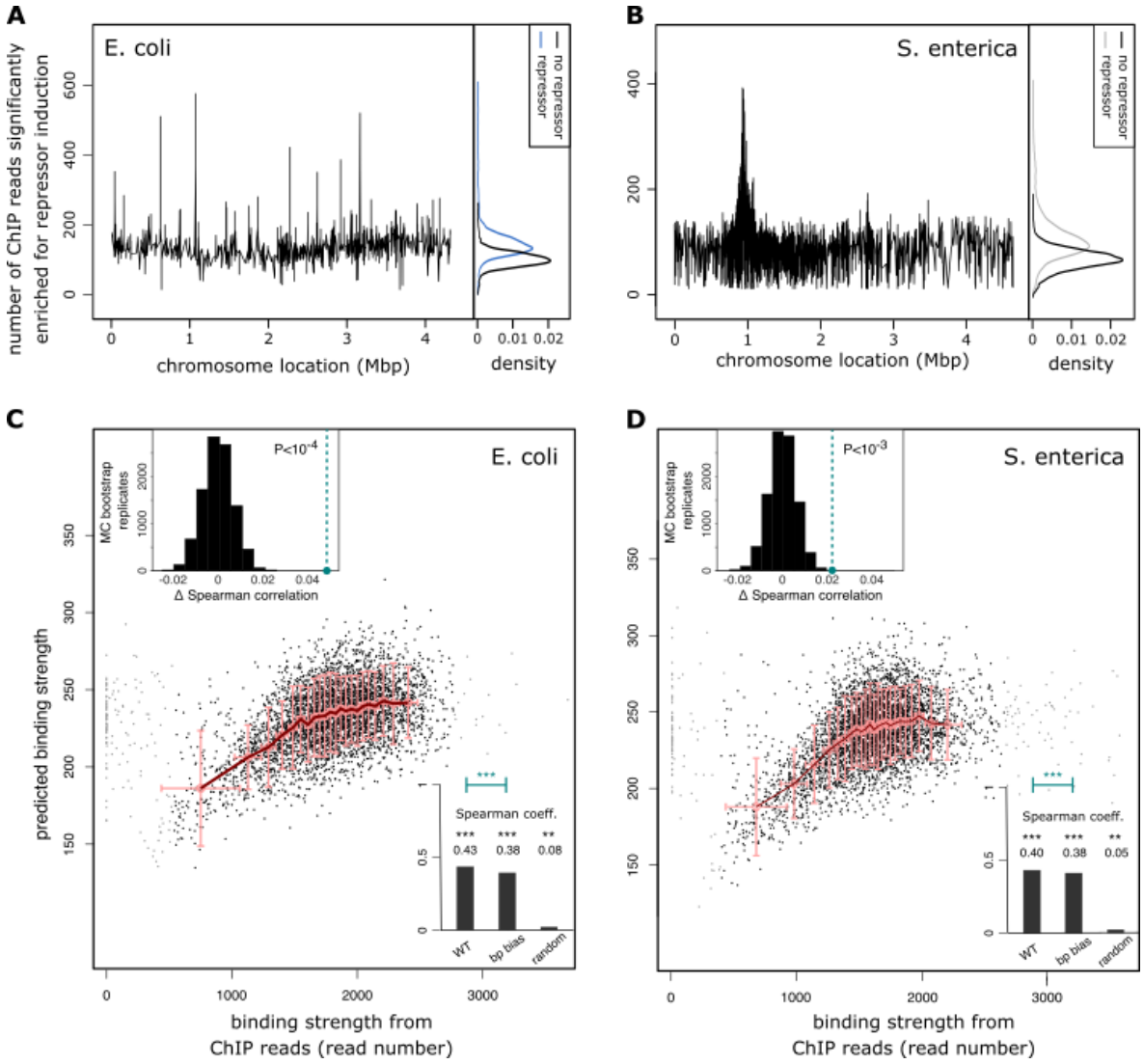


Figure 5. Distributed, low-specificity binding of λ CI across the genomes of *E. coli* and *S. enterica*.

(A,B) Distributions of ChIP-sequencing reads for the regions found to be significantly enriched in the experiment with λ CI over the experiment without λ CI across the **(A)** *E.coli* or **(B)** *S. enterica* genome. On the right density plots of enriched read numbers are given for repressor (color) or no repressor (black) experiments (Number of reads for not significantly enriched regions are shown in Fig. S4 for comparison). None of the apparent peaks in **(A)** or **(B)** encodes for an essential gene, nor one obviously beneficial in minimal media. **(C,D)** Fit between binding strength predictions of a simple thermodynamic model using the energy matrix for λ CI binding and the ChIP-sequencing reads across 1000bp windows along the **(C)** *E. coli* or **(D)** *S. enterica* genome. In **(D)**, 33 out of 4856 data points showed more than 3600 reads and were omitted for clarity (the data is available on request). Lower insets show the calculated Spearman correlations using either the wildtype energy matrix, one that only conserves the λ CI basepair bias (Fits are shown in Fig. S5) or one that has completely reshuffled entries (averaged over 100 permutations). Upper insets: Permutation test for the significance of the difference in Spearman correlation between binding predictions using the wildtype energy matrix (first bar in the lower inset) vs prediction with the energy matrix that only conserves λ CI basepair bias (second bar in

the lower inset). Black histograms represent the Monte-Carlo-derived null distribution (10^4 random reassignments of ChIP reads to genomic regions), green dot and line show the true excess Spearman correlation. The correlations for *S. enterica* were not substantially affected by the strong binding peak in prophage regions shown in **(B)**.

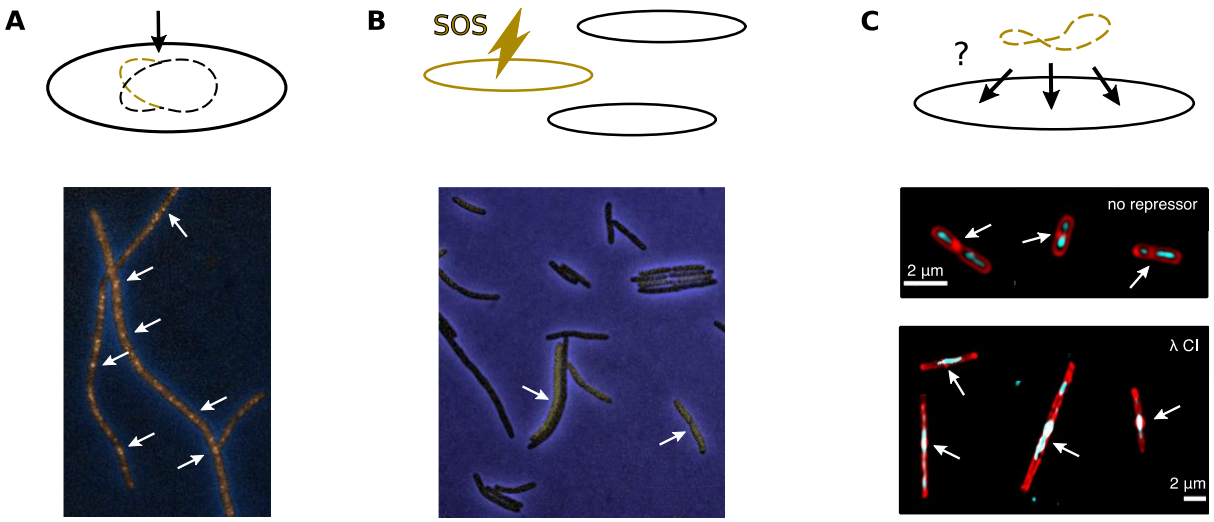


Figure 6. Induction of λ CI interferes with cell division but not DNA replication.
Cells were imaged under the microscope in minimal media with glucose and aTc, either **(A,B)** on an agar pad, or directly **(C)** in liquid media (see Methods). Fluorescence indicates **(A)** ongoing replication (a SeqA-GFP fusion as a replication fork marker) or **(B)** potential induction of a primary stress response promoter (P_{sulA}-*yfp* reporter). **(C)** Chromosome positioning within the cell is shown in blue (Hoechst dye), relative to the cell membrane in red (NileRed). White arrows indicate an overlap of septation spots (thicker red dots) with DNA (blue) for cells containing λ CI.