Expanding the Boolean logic of the prokaryotic transcription factor XylR by functionalization of permissive sites with a protease-target sequence

by

Belen Calles and Víctor de Lorenzo*

Systems Biology Program. Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco, 28049, Madrid (Spain)

Keywords: Sigma 54, XylR, logic gates, genetic circuits, transcriptional factors, biodegradation, environmental biotechnology

Running Title: The logic of prokaryotic enhancer-binding proteins

* Correspondence to: Víctor de Lorenzo
Centro Nacional de Biotecnología-CSIC
Campus de Cantoblanco, Madrid 28049, Spain
Tel.: 34- 91 585 45 36; Fax: 34-91 585 45 06
E-mail: vdlorenzo@cnb.csic.es

ABSTRACT

The σ^{54}-dependent prokaryotic regulator XylR implements a one-input / one-output actuator that transduces the presence of the aromatic effector *m*-xylene into transcriptional activation of the cognate promoter *Pu.* Such a signal conversion involves the effector-mediated release of the intramolecular repression of the N-terminal A domain on the central C module of XylR. On this background, we set out to endow this regulator with additional signal-sensing capabilities by inserting a target site of the viral protease NIa in permissive protein locations that once cleaved *in vivo* could either terminate XylR activity or generate an effector-independent, constitutive transcription factor. To find optimal protein positions to this end we saturated the *xylR* gene DNA with a synthetic transposable element designed for randomly delivering in-frame polypeptides throughout the sequence of any given protein. This Tn*5*-based system supplies the target gene with insertions of a selectable marker that can later be excised leaving behind the desired (poly) peptides grafted into the protein structure. Implementation of such knock-in / leave behind (KILB) method to XylR was instrumental to produce a number of variants of this TF that could compute *in vivo* two inputs (*m*-xylene and protease) into a single output following a logic that was dependent on the site of the insertion of the NIa target sequence in the TF. Such NIa-sensitive XylR specimens afforded the design of novel regulatory nodes that entered protease expression as one of signals recognized *in vivo* for controlling *Pu.* This approach is bound to facilitate the functionalization of TFs and other proteins with new traits, especially when their forward engineering is made difficult by, for example, the absence of structural data.

INTRODUCTION

Promoters are the basic molecular devices that translate given physicochemical signals into decision to start transcription of specific DNA sequences into mRNA[1]. Regulation of this process in bacteria is typically mediated by transcriptional factors that either trigger (activators) or inhibit (repressors) the action of RNA polymerase on DNA motifs that are bound on the basis of the sigma factor included in the enzyme[2]. The many possibilities of interplay between different TFs, the RNAP, and the target DNA originate a considerable plasticity both in terms of the input/output logic of the regulatory nodes at stake and its kinetic properties. Both the logic structure and the parameters embodied in each singular

promoter often appear connected to other regulatory devices of the kind to form complex genetic networks which ultimately rule the lifestyle of the bacteria that host them[3].

Virtually all known prokaryotic promoters can be described with Boolean formalisms under which each regulatory event results from the action of one or more binary gates which compute up to two inputs into a single output with a pre-fixed logic[1]. Similarity of such logic circuits to electronic networks has stimulated the design of gates artificially assembled with prokaryotic regulatory parts that can process specific signals and can be combined with others for implementing simple computations[4]. The repertoire of such regulatory devices is typically limited to existing TFs and cognate promoters. The latter can be easily engineered to contain binding sites in positions which make transcriptional output to follow different outcomes depending on the signal-responsive properties of the transcription factors employed in the design[5]. Interestingly, most prokaryotic promoters compute signals on the mere basis of binding (or lack of it) of cognate TFs to DNA[2]. In contrast, extant TFs do not perform any binary computation by themselves, but they simply transduce one signal (e.g. effector binding) into another (e.g. a conformational change) that may result in productive attachment to the target promoter. Activators thus intrinsically implement a YES gate while repressors execute a NOT gate[1]. Dependency of such activities on small effector molecules allows their connection for the sake of growingly complex gates and circuits. Yet, the question at stake is whether one could artificially make single TFs not just to transduce single signals but to compute two inputs with a predetermined logic –thus converting the TF itself (and not its binding to DNA) in the executor of the desired logic operation[6,7]. But what TF or TF family could be optimal to this end? In this work, we advocate prokaryotic activators that depend on the alternative sigma factor σ^{54} as the platform of choice[8] for artificially endowing new-to-nature possibilities to the logic of bacterial promoters.

TFs that act in concert with σ^{54} (also known as prokaryotic enhancer-binding proteins or NtrC-type regulators) have a distinct modular structure that includes an amino-terminal, signal-reception region (A domain), the *hinge* B domain which places the A domain in a position that allows or not transcriptional activation, the central C domain responsible for binding and hydrolysis of the ATP and interactions with σ^{54} and the C-terminal D domain, which binds DNA[9]. In a group of such TFs, the A domain represses the ATPase activity of the TF in the absence of the activating signal (typically a small effector molecule). TFs of this type are involved in different physiological processes, e.g. metabolism of aromatic

compounds (XylR, DmpR, HbpR, TbuT and PhhR), formate metabolism (FhlA), nitrogen fixation (NifA), acetoin catabolism (AcoR), transport systems (DctD), and others[10]. In the case of the XylR regulator of the TOL pathway of *Pseudomonas putida* mt-2[10,11], the A domain interacts directly with the aromatic effector *m*-xylene, an event that results in the release of the intramolecular repression (or anti-activation) caused by the A domain itself on the rest of the protein. As a consequence, XylR variants deleted of the A module (XylRΔA) are constitutively active[12,13]. XylR plus *m*-xylene (or XylRΔA) then activates the target σ^{54}-promoter *Pu* in concert with a number of DNA binding proteins that endows the regulatory node with a complex logic[14]. However, XylR acts in this system only as a mere one-input/ one-output actuator that translates the presence of *m*-xylene into a protein form able to activate transcription. Inspection of the XylR domain structure and its activation mechanism (Fig. 1) suggested that it would be possible to produce TF variants with an expanded logic repertoire if the protein could be conditionally cleaved in a fashion that either destroyed its activity altogether or deleted the A domain and originated an effector-independent, constitutively active regulator.

The results below describe the design and implementation of a new molecular tool for functionalization of target proteins (e.g. XylR) with novel properties brought about by insertion of purposeful polypeptides at otherwise permissive sites of its primary sequence. The tool is based on the *in vitro* saturation of the TF-coding DNA with a synthetic transposon that, after insertion and selection, can be excised leaving behind an in-frame functional sequence of choice (for example, a specific protease cleaving site), which can be tested for permissiveness *in vivo.* Application of this tool to XylR originated TF variants that responded either positively or negatively to expression of such protease, which could then be entered as one of the inputs of the system in live cells. The resulting TFs implemented by themselves a suite of non-natural logic actions that have no precedents in extant prokaryotic regulators and thus expand the repertoire of prokaryotic devices available for engineering logic circuits. Since XylR originates in a system for catabolism of *m*-xylene, its functionalized variants have an especial value for programming bacteria aimed at bioremediation of environmental pollutants.

RESULTS AND DISCUSSION

Rationale for creating logic gates based on XylR. The domain structure and the mechanism of action of XylR on its cognate promoters *Pu* and P_R of the TOL plasmid pWW0 of *P. putida* mt-2 are sketched in

Fig. 1. Three features of the process are worth considering for the sake of this work. First, unlike most prokaryotic TFs, this regulatory proteins is clearly composed of 3 distinct domains: the N-terminus module, which interacts directly with the aromatic effector *m*-xylene (or some structural analogues), the central C domain contacts and activates the sigma factor σ^{54} of RNAP for recognition and eventual formation of an open complex at the -12/-24 DNA motif that is typical of this type of promoters, and the C-terminal helix-turn-helix part (D domain) for binding upstream sequences[12,15]. The A and C domains are connected by a small hinge B sequence. XylR is thus a complete actuator that transforms an input signal (*m*-xylene) into eventual motion of the RNAP. The other components necessary for transcription initiation (promoter DNA, ATP, IHF, and additional nucleoid-associated proteins) can be considered not to vary and thus can be abstracted with a default value[16]. The second unique feature of XylR and other TFs of its class is that the mechanism of activation by *m*-xylene involves the release of an intra-molecular occlusion exerted by the effector-binding A domain on the C domain[12,13]. This makes deletion of the N-terminus of XylR to produce an effector-independent constitutive variant, which –for the sake of *Pu* activation is equivalent to the wild-type protein in the presence of *m*-xylene. Finally, XylR can also act as a repressor of its own synthesis, because it binds also sequences of the TOL plasmid that overlap the σ^{70} promoter P_R for transcription of the *xylR* gene[17].

The logic structure of such a regulatory device of the TOL plasmid is shown in Fig. 1. Perusal of the primary sequence of XylR immediately suggested that it would be possible to enter an additional input to the system by inserting specific protease-cutting sites at strategically located spots of the protein structure, provided that they did not alter TF activity in the absence of cleavage. While many locations could be predicted to terminate XylR function upon proteolysis, those able to excise the A domain from the rest of the protein could in fact activate this TF with a different mechanism than that caused by exposure to *m*-xylene. These scenarios open the possibility of having the same TF responding to two entirely independent inputs (*m*-xylene and protease) and the output to have an opposite sign reliant on the site of the XylR structure subject to cleavage. This would expand considerably the number of logic gates that could be derived from XylR-targeted promoters and similar σ^{54}-dependent TFs. Yet, the technical bottleneck for this endeavor is the identification of such permissive sites for implantation of a functional target for a specific protease within protein structure. The sections below describe the design of a synthetic tool tailored precisely to this end and its application to generate XylR variants endowed with the desired signal-processing capacities.

Genetic grafting of protease-cleaving sites through the XylR structure. Since the permissiveness of protein structures to insertions of extra amino acid sequences is often difficult to predict upfront, we set out to develop a general molecular tool for searching such sites in any protein of interest to be grafted with any other functional polypeptide. To address this, we exploited the known mechanism of transposition of Tn*5*[18] for designing a high-efficiency mobile DNA segment that could first be delivered to the target DNA, selected for insertions and then excised to leave behind the grafted sequence. The organization of the synthetic mobile element engineered to this end, which we have termed mTn*5* [GFP•NIa1], is sketched in Fig. 2. A detailed description of its functional parts and its performance *in vivo* and *in vitro* can be found in the Supporting Information. Once the method for *in vitro* transposition of mTn*5* [GFP•NIa1] into any target sequence was in place we carried on to generate a large library of insertions of this element through the *xylR* gene born by plasmid pBCL4. This was then followed by excision of much of the transposon length to leave behind a sequence scar encoding the short amino acid sequence cleaveable by the viral protease NIa. The workflow for generating such knock-in-leave-behind (KILB) libraries is sketched in Fig. 3. The transposition reaction is predicted to introduce the mobile element throughout the whole plasmid i.e. both inside and outside the *xylR* sequence. Predictably, digestion of the transposition mix with enzymes BamHI and XbaI generated four restriction bands, which could be easily separated by means of electrophoresis in agarose gels (Supplementary Fig. S1). The product of the size of *xylR* plus one mTn*5* [GFP•NIa1] insertion (3541 bp) was recovered and re-cloned in the same sites of the pUC18-SbfI plasmid pre-digested with BamHI and XbaI). This simple procedure allowed the recovery of the inserted *xylR* sequences only, as it discards transposition events occurring *in vitro* beyond the sequence of interest in the pBCL4 plasmid. The ligation pool was then transformed in *E. coli,* followed by selection on media with Ap^R Km^R. The whole of transformants were pooled again and the total plasmid contents extracted from the mixed population. The plasmidic material was then digested with either NotI or SbfI and the digestion products re-ligated. Owing to the design of the synthetic transposon (Fig. 2), such an excision of the internal NotI or SbfI segments of mTn*5* [GFP•NIa1] followed by religation leaves *xylR* DNA with in-frame fit-in insertions of either GFP or the NIa target polypeptide, respectively. One out of 6 of these inserts was predicted to create sandwiched gene fusions between *xylR* and either GFP or the proteolyzable peptide. If the sites of start and end of such grafted polypeptides in XylR happen to be structurally permissive we would then expect to have this TF artificially added in its structure with a new trait i.e. either fluorescence (because of the

sandwiched GFP) or sensitivity to the NIa protease (due to the insertion of a cognate target site). XylR variants of both types were screened for functionality by transforming each pool in *E. coli* CC118 *Pu-lacZ*. This strain has a chromosomal insertion of a reporter β-galactosidase gene to the σ^{54} promoter *Pu* that is activated by XylR in the presence of the aromatic inducer[19]. We in fact obtained a number of both XylR derivatives that were fluorescent and able to activate the cognate σ^{54} promoter *Pu* and others that were responsive to the NIa protease. The sections below, however, focus exclusively on the last category, as they are the ones that change the input/output logic of the regulator, as pursued in this work (see above).

Analysis of NIa-tagged XylR variants. The negligible level of basal transcription of the *Pu* promoter under non-induced conditions (i.e. without XylR or with XylR but not *m*-xylene) made strain *E. coli* CC118 *Pu-lacZ* a phenomenal tool for examining the effect of the genetic grafts discussed above on XylR properties. The reference conditions for such functionality tests are shown in Supplementary Fig. S2. The lawns of plasmid-less *E. coli* CC118 *Pu-lacZ* (or the same strain transformed with insert-less vectors) are colorless when spotted on LB-Xgal plates. The same is true for *E. coli* CC118 *Pu-lacZ* transformed with the reference *xylR*$^{+}$ plasmid pBCL4, which encodes the wild-type sequence of this TF –provided that the plates are not exposed to *m*-xylene. Exposure to this aromatic makes the lawns of *E. coli* CC118 *Pu-lacZ* (pBCL4) to turn intense blue. These visual phenotypes match exactly the levels of β-galactosidase that can be measured in liquid cultures of the same strains, as shown in Supplementary Fig. S2. Reporter readout in this system thus faithfully describes the functionality of XylR as an *m*-xylene responsive TF.

Once the conditions to measure XylR activity were standardized, the KILB library of NIa-target insertions born by plasmid pBCL4 was transformed in *E. coli* CC118 *Pu-lacZ,* plated on LB-Ap^{R} and the resulting colonies exposed to saturating vapors of *m*-xylene as described in Methods. Out of a whole library of 2.7x10^{3} clones, approximately 45 % turned blue under such conditions, suggesting that the extra in-frame polypeptide left in the protein structure by the KILB transposon had hit permissive sites of the protein structure. DNA sequencing of a randomly picked subset of ~ 50 clones indicated that not all permissive insertions had the proper orientation and/or the correct reading frame to generate productive NIa recognition sites within XylR.Finally, only four *xylR* clones inserted with NIa-sites were selected for further phenotypic analyses. Three of these NIa-site insertions were found at various places of the N-

terminal signal reception A module of the XylR protein (M75, G154 and D210) whereas a fourth one (E499) was located in the short linker that connects the central activation module C of the protein and the DNA-binding D domain. As shown in Fig. 4, insertions M75 and G154 were competent for induction of the *Pu-lacZ* fusion of the host, but originated lower β-galactosidase levels than the wild-type XylR when exposed to *m*-xylene. In contrast the NIa-target insertion at the very end of the A domain (D210) fashioned a XylR variant with a higher activity when induced with the same aromatic effector. A similar result was obtained with the NIa-targeted E499 XylR variant, which displayed a significantly higher *Pu* output when exposed to the protease *in vivo* (Fig. 4).

The wild type-like behavior of insertions D210 and E499 did however change when the host reporter strain was made to express the NIa protease by means of plasmids encoding the cognate PPV gene. In the first case, insertion of the NIa recognition site at the end of the A domain of XylR (D210) led to *Pu* induction irrespective of the presence or the absence of the XylR inducer (*m*-xylene) when it was expressed along with the protease. This phenotype is consistent with that expected of a XylRΔA protein, as previously described[12,13]. That XylR^{D210} was cleaved by NIa in vivo could be visualized by means of a Western blot assay of protein extracts of the corresponding cells (Fig. 4b, lanes 7 and 8). Note that antibodies used to detect XylR were raised against the XylRΔA protein[20] and therefore they do not recognize the A domain. Results equivalent to those of Fig. 4b were obtained when the Western blot test was made in the presence of *m*-xylene, i. e. the NIa protease appeared to proteolyze the XylR variants under examination with the same efficiency. These data thus accredited that XylR^{D210} can be converted into a TF form able to activate *Pu* by either exposure to *m*-xylene or by expression of the NIa protease or by both. This notion was further verified by reconstructing a XylR variant which had been deleted exactly of the same portion of the A domain that is predicted to be lost upon cleavage of XylR^{D210} with NIa (see below).

A quite different behavior was found in the XylR variant inserted with a NIa site in position E499. In that case, expression of the protease translated in a virtually inactive TF regardless of whether *m*-xylene was present in the medium (Fig. 4). Western blots of the protein extracts as before confirmed that NIa indeed cleaved XylR^{E499} *in vivo* (Fig. 4b, lines 9 and 10). Since such a cleavage must result in the deletion of the DNA binding domain of XylR, it makes sense that the TF factor loses activity altogether. This last experiment also provided a sidelight in the mechanism of activation of *Pu* by XylR, since it

makes clear that at least part of the D domain of the protein is essential not only for DNA binding but also for maintaining a form of the protein able to activate transcription from solution[19, 21, 22]. Finally, NIa target insertions at sites M75 and G154 resulted in XylR variants that could be cleaved *in vivo* as well (Fig. 4b, lines 3 to 6) but such site-specific proteolysis changed little the corresponding phenotypes regarding *Pu* induction. It is possible that such variants that were identified in the first visual screening (see above) are in fact defective or only transiently active TFs.

The novel Boolean logic of XylR^{D210} and XylR^{E499}. As shown in Fig. 5a, insertion of NIa target sequences in D210 and E499 sites of XylR endowed this TF with the capacity to compute two signals (*m*-xylene and protease) instead of the one-input/one-output observed in the naturally occurring regulator. In one case, XylR^{D210} brings about strong activation of the *Pu* promoter whether cells are exposed to the aromatic inducer, to NIa or both. This state of affairs can be formalized as a Boolean OR gate (Fig. 5b). It is noteworthy that promoter activity caused by cleavage of XylR^{D210} is noticeably higher than that of *m*-xylene and that the first overrides the second when the two are entered together (e.g., compare ß-galactosidase levels of cognate assays in Fig. 4c). This makes sense in view of the mechanism of activation of XylR by aromatic inducers[12]: the loss of the A domain leaves the TF unhindered for interacting with the σ^{54}-dependent transcription machinery. A different logic gate was created by the insertion of a NIa site in XylR^{E499}. In this instance, expression of the protease abolishes activation of the TF by *m*-xylene (Fig. 5). For *Pu* to be transcribed cells thus need to face the aromatic effector but must not be exposed to any proteolysis caused by NIa. The logic is therefore that of a Boolean ANDN gate[1] in which one specific input must be present and the other absent to have a positive outcome of the computation. Note, however, that in the case discussed here, the inputs are not equivalent and their order of appearance makes a difference. In any case, the above manipulations of XylR expand the logic repertoire of this TF to additional signals that can result in either positive or negative outcomes.

***Pu* promoter anti-activation: engineering a cleavable variant of XylRΔA.** The inhibitory action of NIa on XylR^{E499} raised still one more possibility to develop a different logic gate based on this TF. Since the *in vivo* deletion of the D domain leads to an entirely inactive regulator (Fig. 4), we wondered whether introducing directly the NIa site in the constitutively active protein XylRΔA could reverse the action of this TF on *Pu* upon expression on the protease *in vivo*. To examine this possibility we produced a series

of XylRΔA variants that carry various sequences at their N and C termini as shown in Fig. 6 (see details on the protein ends in Supplementary Fig. S3). The collection included as controls the original XylRΔA2 protein of reference[12,13] named SP1 in Fig. 6a) and a faithful reconstruction of the truncated product that is predicted to be released upon cleavage of XylR^{D210} with NIa (SP3 in Fig. 6). Each of these was then engineered with protease-cutting sites at position E499, originating cleavable protein variants SP2 and SP4 respectively (both designated as XylRΔA*). Finally, we recreated the polypeptide that could result from excision of the XylR protein at both D210 and E499 sites, which encompasses the whole C domain of the TF. Plasmids encoding each of these XylR variants were passed to *E. coli Pu-lacZ* strains expressing or not NIa and the production of the regulator examined in each condition. As shown in Supplementary Fig. S3, control variants SP1, SP3 and SP5 were not affected by NIa, while SP2 and SP4 were cleaved as expected. When the same strains were patched on Xgal plates, the change of color of variants SP2 and SP4 in the cells producing NIa became evident. These visual phenotypes are consistently reflected in the actual levels of the reporter product displayed by each of the constructs with and without protease as shown in Fig. 6c. The most dramatic change was delivered by the SP4 variants, which passed from a high β-galactosidase level in the absence of protease (~2000 Miller units) to virtually undetectable in the strain that expressed NIa from plasmid pPPV1. Note that unlike full-length XylR, the default action of XylRΔA is activation of *Pu* in the absence of any effector (a YES gate, Fig. 6d) and the effect of the protease is to defeat this event. The consequence of NIa expression is therefore to revert activation and thus suppress *Pu* activity. If expression of the cleavable XylRΔA variant is given a digital value of 1 then proteolysis can be formalized as an inverter in which NIa is the sole input. But if expression of XylRΔA* is also variable, then the resulting regulatory device becomes an ANDN gate with both NIa and the engineered TF as inputs (Fig. 6d). To the best of our knowledge, this is the first case of either a naturally occurring or an engineered biological inverter that is implemented through an anti-activation mechanism. Although the logic of such NOT device is the same than that brought about by a repressor[1], the biological basis of the inversion is entirely different, what will surely be reflected in the parameters that govern the process *in vivo*. While such parameterization of this and the other regulatory devices described above will be the subject of future work, we expect these new gates based on XylR to enrich the choices available for construction of complex genetic and metabolic circuitry.

Conclusion. The application of Boolean logic to a large number of biological phenomena has allowed both formalization of intricate occurrences in live systems[16] and the engineering of genetic and metabolic devices for programming new-to-nature properties. The biological parts available for such engineering include transcriptional factors and cognate promoters[1,5,6], recombinases[23,24], metabolic reactions[25-27], small molecules[4,7,26] single cells[28] and even multicellular networks[29]. The modularity of logic gates allows the buildup of a degree of multi-scale complexity that is limited only by the biological compatibility of the corresponding inputs and outputs[6,28]. On this basis, contemporary Synthetic Biology claims a similarity between genetic networks and electronic circuits that include not only discrete decision-making modules, but also whole operating systems[30,31]. Logic devices based on regulatory parts are typically implemented by combinations of transcriptional factors and small molecules that act as inputs in given promoters. DNA binding (or not) is, mechanistically, the event that mediates the corresponding computation. We show above that one family of prokaryotic TFs that act in concert with the σ^{54}-containing form of RNAP can be functionalized with protease-cleaving sites in a fashion that makes the TF itself –not its binding to DNA, the performer of the binary computation. Prokaryotic TFs that process two equally effective inputs are thus far unknown in the transcription literature. Some regulators may use intermediate metabolites as allosteric effectors[32,33] but their effects are mild as compared to the drastic change in *Pu* promoter output caused by the XylR variants described above. Moreover, we have not overlooked that the genetic tools described above for implementing the KILB insertion saturation procedure (transposons mTn*5* [GFP•NIa1], mTn*5* [GFP•NIa2] and mTn*5* [GFP•NIa3]) can be tailored *á la carte* for grafting functional sequences in permissive sites of virtually any other protein of interest. While the random insertion approach for *sandwiching* foreign polypeptides in existing proteins is not without precedents[34-37], the work reported here is the first time that the concept is applied to transcriptional factors with a view on changing its regulatory behavior. In this respect, although the data presented in this paper deal only with the ability of XylR to activate *Pu,* Fig. 1 shows also that the same TF represses its own promoter, P_R. It is thus conceivable that the logic of the new gates based on XylR^{D210}, XylR^{E499} and XylRΔA* (Fig. 5b and Fig. 6d) is reverted when the target promoter is P_R instead of *Pu*. Alas, the degree of repression of P_R by XylR is not strong enough to grant a performance as stringent as the one observed with *Pu*[17]). Still the binding of XylR to P_R can be artificially improved, an issue that is currently under investigation. In sum, we argue the value of combining σ^{54}-dependent TFs, cognate promoters, small molecules and proteases as a way of

increasing the toolbox of logic devices that are necessary to build genetic and metabolic circuits of growing complexity e.g. for *in situ* biodegradation of toxic pollutants[38].

METHODS

Strains, plasmids, media and growth conditions. The relevant properties of the strains and constructions used in this work are listed in Supplementary Table S1. *E. coli* DH10B, DH5α and CC118 strains were used for general procedures. The reporter strain *E. coli* CC118 *Pu-lacZ* was used for assessing XylR activity. Bacteria were grown routinely at 37 ºC in LB (10 g l^{-1} of tryptone, 5 g l^{-1} of yeast extract and 5 g l^{-1} of NaCl). When required, ampicillin (Ap, 150 μg/ml), kanamycin (Km, 75 μg/ml) or chloramphenicol (Cm, 30 μg/ml) was added to the culture media. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added where indicated to a final concentration of 0.1 mM. The *Pu-lacZ* fusion was induced by exposing cells either on plates or in liquid cultures to saturating vapors of *m*-xylene. When required, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal) was added at 40 μg ml^{-1} for visualization of β-galactosidase activity.

DNA constructs. General methods for DNA manipulation were performed as described[39]. Oligonucleotides used in polymerase chain reaction experiments (PCR) are listed in Supplementary Table S2. Construction of a transposition target plasmid encoding *xylR* gene, involved two steps. First, the single SbfI site of pUC18 was eliminated by digestion with PstI followed by T4 DNA polymerase treatment and religation, resulting in vector pUC18-SbfI. Next, the DNA sequence of the *xylR* gene was amplified from strain *P. putida* mt-2 with oligos xylR-BamHI (containing an optimal RBS and a *Bam*HI restriction site) and xylR-XbaI (which adds an XbaI site). The resulting fragment was cloned into a pGEM-T (Promega), excised with BamHI and XbaI and ligated into the corresponding sites of pUC18-SbfI. This produced plasmid pBCL4, which was subsequently used as the target DNA in transposition experiments. The DNA segments that compose the KILB transposon used in this work were synthesized (Life Technologies, Regesnburg, Germany) and combined with a Km resistance gene amplified from plasmid pBAM1 cassette with primers Km-SwaI-Fan dKm-PshAI-R, which generate terminal *Swa*I and *Psh*AI sites. The resulting segment, assembled in plasmid pGA-BCL1 (Supplementary Table S1) bears the mini-Tn*5* transposon named mTn*5* [GFP•NIa1], the structure of which is drawn in Fig. 2a. For *in vivo* transposition experiments, the DNA spanning the whole mobile

element was cloned as a PvuII fragment in the corresponding sites of plasmid pBAM1, thereby originating pBAM1-GFP (GenBank HQ908072). Two more versions of the same transposon were constructed -bearing either AscI or PmeI restriction sites in lieu of the SbfI sequences, thereby generating mTn5 [GFP•NIa2] and mTn5 [GFP•NIa3]. Details of their DNA assembly steps are available upon request. Plasmids expressing different XylRΔA truncated variants were constructed as follows. DNA segments encoding SP1 and SP2 -both deleted of their N-terminal domains as described for XylRΔA2[13] were amplified with primers DeltaA2F and M13 (-40) universal-F from plasmids pBCL4 (wt *xylR* gene) and pBCL4-E499 (*xylR*^{E499} variant), respectively. The resulting DNAs were then digested with BamHI and XbaI and cloned into the corresponding sites of pUC18, giving rise to pBCL4-SP1 and pBCL4-SP2. Other XylRΔA variants were made with an N-terminus that mimics the result of the cleavage of XylR^{D210} with the NIa protease. For SP3, the insert of plasmid pBCL4-D210 (encoding the *xylR*^{D210} variant obtained by KILB) was amplified with primers D210F and M13 (-40) universal-F, the resulting DNA digested with BamHI and XbaI and ligated into the corresponding sites of pUC18, raising pBCL4-SP3. In the case of SP4 and SP5, two PCR reactions were run in each case to obtain separate 5´ and 3´ends in each case, followed by a second overlapping reaction using products from the first PCR as templates. The 5´ region, which was common to both SP4 and SP5 was amplified from pBCL4-D210 with primers D210F and XylR-Sol.R. The 3´ portions were obtained by PCR of pBCL4-E499 (encoding the *xylR*^{E499} variant obtained by KILB) with primers XylR-Sol.F and M13 (-40) universal-F -in the case of SP4 and XylR-Sol.F and E499stop-R in the case of SP5. Equivalent amounts of the 5´ DNA fragment together with each of the 3´ segments were used as templates for a second PCR reaction with primers D20F1 and M13 (-40) universal-F for full-length amplification of SP4 and D210F1 and E499stop-R for the same in SP5. The DNAs resulting from this reaction were then digested with BamHI and XbaI and ligated into the corresponding sites of pUC18, thereby originating pBCL4-SP4 and pBCL4-SP5.

***In vitro* transposition and construction of knock-in-leave-behind (KILB) insertion libraries.** A hyperactive variant of the Tn*5* transposase was purified from plasmid pGRTYB35 (kindly provided by W.S. Reznikoff) as described[40]. The donor DNA segment spanning the mTn*5* [GFP•NIa1] transposon was amplified from pGA-BCL1 with primers Tn5ME-F and Tn5ME-R. The amplified fragment was then gel purified with NucleoSpin Extract II kit (Macherey-Nagel), and kept until use. In vitro transposition experiments were set up as described[41]. The reactions were assembled in a volume of 10 µl of

transposition buffer containing 0.1 μM purified transposase (0.1) and an equimolar amount of transposon and target DNA (ratio transposase:transposon:target DNA = 5:1:1). Reactions were incubated at 37ºC for two hours and then halted with 1 ml of stop solution (1% SDS), mixed and heated at 70ºC for 10 minutes. Next, the mixtures were dialyzed againstMilliQ water and electroporated into *E. coli* DH10B. The transformation mixture was then plated on LB Km (75 μg/ml) to select cells with plasmids that had acquired the mTn*5* [GFP•NIa1] transposon (Fig. 3). The efficiency of the transposition reaction was measured as CFUs per pMol of mTn*5* [GFP•NIa1] DNA. Next, the Km^R clones were pooled, the whole plasmid DNA extracted and digested with BamHI and XbaI. This generated four restriction products that were separated with electrophoresis in agarose gels (Supplementary Fig. S1a). The band corresponding to the *xylR* gene with transposon insertions was recovered, re-cloned in pUC18-SbfI and retransformed in *E. coli* DH10B. Clones were pooled again, plasmid DNA extracted and separately digested with either NotI or SbfI and then religated (Supplementary Fig. S1b). As explained in Fig. 2, NotI digestion/religation creates in-frame sandwich GFP fusions, while the same with SbfI leaves the target gene sequence densely punctuated with in-frame insertions of the NIa protease target peptide (plus adjacent sequences inherited from the Tn*5* ends, Fig. 2). The corresponding plasmid pool was recovered and transformed in reporter strain *E. coli* CC118 *Pu-lacZ* for XylR activity assays as explained next.

Monitoring promoter activity *in vivo.* The ability of XylR and its variants to activate transcription from the σ^{54} promoter *Pu* was measured by quantifying the β–galactosidase accumulation driven by a *Pu-lacZ* fusion engineered in the chromosome of *E. coli* CC118[19]. This reporter strain was transformed with the plasmids encoding *xylR* variants described above along, where indicated with plasmid pPPV1[42] or pPPVs20[43] encoding the NIa protease. For the assays, cultures were pregrown overnight at 37ºC in LB medium with appropriate antibiotics, then diluted in fresh medium to an OD_{600} = 0.1 and grown with vigorous shaking up to mid-exponential phase (OD_{600} = 0.4-0.5). At that point flasks were added 0.1 mM IPTG and the incubation continued up to OD_{600} ~ 1.0. Cultures under scrutiny were then exposed to saturating vapors of the XylR effector (*m*-xylene) in airtight flasks and incubated further for 3 hours. β-galactosidase levels were then determined in cells permeabilized with chloroform and SDS as described by[44]. The results shown represent a minimum of 3 experiments per each condition.

Western blot analyses of XylR expression. The performance of the NIa protease to cleave XylR variants *in vivo* was diagnosed in bacteria from the cultures grown as described in the previous section. To this end, cells recovered by centrifugation were directly disrupted by boiling them for 7 min in a denaturing sample buffer containing 2% SDS and 5% β-mercaptoethanol. Samples were then run through 10% SDS-PAGE gels. Purified full-length $_{6xhis}$XylR and XylRΔA proteins kindly provided by C.A. Carreño and[17], respectively, were used as controls. Polyacrylamide gels were subsequently blotted onto a polyvinylidene difluoride (PVDF) Inmobilon-P membrane (Millipore) and probed with 1:2000 dilutions of an anti-XylR recombinant phage antibody PhaB B7[20]. XylR bands were detected with anti-M13 peroxidase conjugates as described and their location revealed by reaction with BM Chemiluminiscence Blotting Substrate (POD) from Roche (Mannheim, Germany).

AUTHOR INFORMATION

Corresponding Author: E-mail: vdlorenzo@cnb.csic.es

Author Contributions: BC performed experiments and drafted the manuscript. VdL directed the project and wrote the paper.

Notes: The authors declare no conflict of interest

ACKNOWLEDGMENTS

This study was supported by the BIO and FEDER CONSOLIDER-INGENIO Program of the Spanish Ministry of Science and Innovation, the MICROME, ST-FLOW and ARYSIS Contracts of the EU, and the PROMT Project of the CAM.

SUPPORTING INFORMATION AVAILABLE

This information is available free of charge via the Internet at http://pubs.acs.org/.

REFERENCES

(1) Silva-Rocha, R., and de Lorenzo, V. (2008) Mining logic gates in prokaryotic transcriptional regulation networks. *FEBS Lett 582*, 1237-1244.

(2) Lee, D. J., Minchin, S. D., and Busby, S. J. (2012) Activating transcription in bacteria. *Annu Rev Microbiol 66*, 125-152.

(3) Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R. Y., Alon, U., and Margalit, H. (2004) Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc Natl Acad Sci USA 101*, 5934-5939.

(4) Goñi-Moreno, A., Redondo-Nieto, M., Arroyo, F., and Castellanos, J. (2010) Biocircuit design through engineering bacterial logic gates. *Nat Comput 10*, 119-127.

(5) Wang, B., Kitney, R. I., Joly, N., and Buck, M. (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat Commun 2*, 508.

(6) Hunziker, A., Tuboly, C., Horvath, P., Krishna, S., and Semsey, S. (2010) Genetic flexibility of regulatory networks. *Proc Natl Acad Sci USA 107*, 12998-13003.

(7) Tamsir, A., Tabor, J. J., and Voigt, C. A. (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature 469*, 212-215.

(8) Wigneshweraraj, S., Bose, D., Burrows, P. C., Joly, N., Schumacher, J., Rappas, M., Pape, T., Zhang, X., Stockley, P., Severinov, K., and Buck, M. (2008) Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. *Mol Microbiol 68*, 538-546.

(9) Zhang, X., Chaney, M., Wigneshweraraj, S. R., Schumacher, J., Bordes, P., Cannon, W., and Buck, M. (2002) Mechanochemical ATPases and transcriptional activation. *Mol Microbiol 45*, 895-903.

(10) Shingler, V. (1996) Signal sensing by sigma 54-dependent regulators: derepression as a control mechanism. *Mol Microbiol 19*, 409-416.

(11) Shingler, V. (2011) Signal sensory systems that impact sigma 54-dependent transcription. *FEMS Microbiol Rev 35*, 425-440.

(12) Fernández, S., de Lorenzo, V., and Pérez-Martin, J. (1995) Activation of the transcriptional regulator XylR of *Pseudomonas putida* by release of repression between functional domains. *Mol Microbiol 16*, 205-213.

(13) Pérez-Martín, J., and De Lorenzo, V. (1995) The amino-terminal domain of the prokaryotic enhancer-binding protein XylR is a specific intramolecular repressor. *Proc Natl Acad Sci USA 92*, 9392-9396.

(14) Silva-Rocha, R., de Jong, H., Tamames, J., and de Lorenzo, V. (2011) The logic layout of the TOL network of *Pseudomonas putida* pWW0 plasmid stems from a metabolic amplifier motif (MAM) that optimizes biodegradation of *m*-xylene. *BMC Syst Biol 5*, 191.

(15) Ramos, J. L., Marques, S., and Timmis, K. N. (1997) Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators. *Annu Rev Microbiol 51*, 341-373.

(16) Silva-Rocha, R., Tamames, J., dos Santos, V. M., and de Lorenzo, V. (2011) The logicome of environmental bacteria: merging catabolic and regulatory events with Boolean formalisms. *Environ Microbiol 13*, 2389-2402.

(17) Bertoni, G., Marqués, S., and de Lorenzo, V. (1998) Activation of the toluene-responsive regulator XylR causes a transcriptional switch between σ^{54} and σ^{70} promoters at the divergent *Pr/Ps* region of the TOL plasmid. *Mol Microbiol 27*, 651-659.

(18) Reznikoff, W. S. (2008) Transposon Tn5. *Ann Rev Genet 42*, 269-286.

(19) de Lorenzo, V., Herrero, M., Metzke, M., and Timmis, K. N. (1991) An upstream XylR- and IHF-induced nucleoprotein complex regulates the sigma 54-dependent *Pu* promoter of TOL plasmid. *EMBO J 10*, 1159-1167.

(20) Fraile, S., Roncal, F., Fernandez, L. A., and de Lorenzo, V. (2001) Monitoring intracellular levels of XylR in *Pseudomonas putida* with a single-chain antibody specific for aromatic-responsive enhancer-binding proteins. *J Bacteriol 183*, 5571-5579.

(21) Dworkin, J., Jovanovic, G., and Model, P. (1997) Role of upstream activation sequences and integration host factor in transcriptional activation by the constitutively active prokaryotic enhancer-binding protein PspF. *J Mol Biol 273*, 377-388.

(22) North, A. K., and Kustu, S. (1997) Mutant forms of the enhancer-binding protein NtrC can activate transcription from solution. *J Mol Biol 267*, 17-36.

(23) Siuti, P., Yazbek, J., and Lu, T. K. (2013) Synthetic circuits integrating logic and memory in living cells. *Nature Biotech* doi: 10.1038/nbt.2510

(24) Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., and Endy, D. (2013) Amplifying Genetic Logic Gates. *Science* doi: 10.1126/science.1232758

(25) Manesh, K. M., Halamek, J., Pita, M., Zhou, J., Tam, T. K., Santhosh, P., Chuang, M. C., Windmiller, J. R., Abidin, D., Katz, E., and Wang, J. (2009) Enzyme logic gates for the digital analysis of physiological level upon injury. *Biosens Bioelectron 24*, 3569-3574.

(26) Niazov, T., Baron, R., Katz, E., Lioubashevski, O., and Willner, I. (2006) Concatenated logic gates using four coupled biocatalysts operating in series. *Proc Natl Acad Sci USA 103*, 17160-17163.

(27) Zhou, J., Arugula, M. A., Halamek, J., Pita, M., and Katz, E. (2009) Enzyme-based NAND and NOR logic gates with modular design. *J Phys Chem B 113*, 16065-16070.

(28) Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., and Voigt, C. A. (2012) Genetic programs constructed from layered logic gates in single cells. *Nature 491, 249-53.*

(29) Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellen, J., Peeters, T., Hohmann, S., de Nadal, E., Posas, F., and Sole, R. (2011) Distributed biological computation with multicellular engineered networks. *Nature 469*, 207-211.

(30) Danchin, A. (2009) Bacteria as computers making computers. *FEMS Microbiol Rev 33*, 3-26.

(31) Yan, K.-K., Fang, G., Bhardwaj, N., Alexander, R. P., and Gerstein, M. (2010) Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *Proc Natl Acad Sci USA 107*, 9186-9191.

(32) Shinar, G., Milo, R., Martinez, M. R., and Alon, U. (2007) Input output robustness in simple bacterial signaling systems. *Proc Natl Acad Sci USA 104*, 19931-19935.

(33) McFall, S. M., Abraham, B., Narsolis, C. G., and Chakrabarty, A. M. (1997) A tricarboxylic acid cycle intermediate regulating transcription of a chloroaromatic biodegradative pathway: fumarate-mediated repression of the *clcABD operon*. *J Bacteriol 179*, 6729-6735.

(34) Ross-Macdonald, P., Sheehan, A., Roeder, G. S., and Snyder, M. (1997) A multipurpose transposon system for analyzing protein production, localization, and function in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA 94*, 190-195.

(35) Biondi, R. M., Baehler, P. J., Reymond, C. D., and Veron, M. (1998) Random insertion of GFP into the cAMP-dependent protein kinase regulatory subunit from *Dictyostelium discoideum*. *Nucleic Acids Res 26*, 4946-4952.

(36) Merkulov, G. V., and Boeke, J. D. (1998) Libraries of green fluorescent protein fusions generated by transposition *in vitro*. *Gene 222*, 213-222.

(37) Sheridan, D., Berlot, C., Robert, A., Inglis, F., Jakobsdottir, K., Howe, J., and Hughes, T. (2002) A new way to rapidly create functional, fluorescent fusion proteins: random insertion of GFP with an *in vitro* transposition reaction. *BMC Neuroscience 3*, 7.

(38) de Lorenzo, V. (2008) Systems biology approaches to bioremediation. *Curr Opin Biotechnol 19*, 579-589.

(39) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning : a laboratory manual.* J. Sambrook, E.F. Fritsch, T. Maniatis. New York : Cold Spring Harbor Laboratory Press.

(40) Bhasin, A., Goryshin, I. Y., and Reznikoff, W. S. (1999) Hairpin Formation in Tn*5* Transposition. *J Biol Chem 274*, 37021-37029.

(41) Goryshin, I. Y., and Reznikoff, W. S. (1998) Tn*5 in vitro* transposition. *J Biol Chem 273*, 7367-7374.

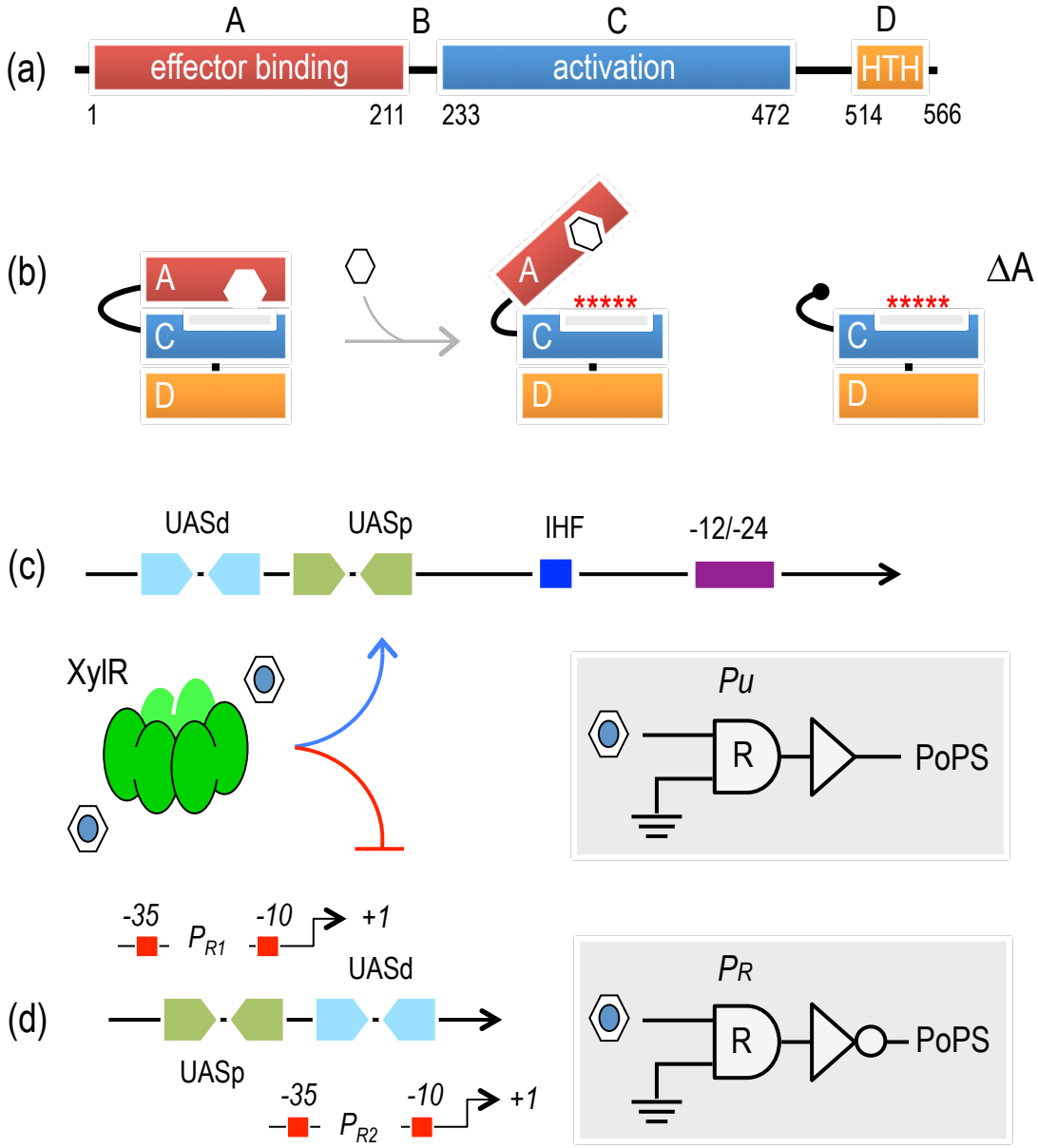
(42) Perez-Martin, J., Cases, I., and de Lorenzo, V. (1997) Design of a solubilization pathway for recombinant polypeptides *in vivo* through processing of a bi-protein with a viral protease. *Protein Eng 10*, 725-730.

(43) Garcia, J. A., Riechmann, J. L., Martín, M. T., and Laín, S. (1989) Proteolytic activity of the plum pox potyvirus NIa-protein on excess of natural and artificial substrates in *Escherichia coli. FEBS Letters 257*, 269-273.

(44) Miller J, H. (1972) *Experiments in Molecular Genetics.* New York : Cold Spring Harbor Laboratory Press.

FIGURES

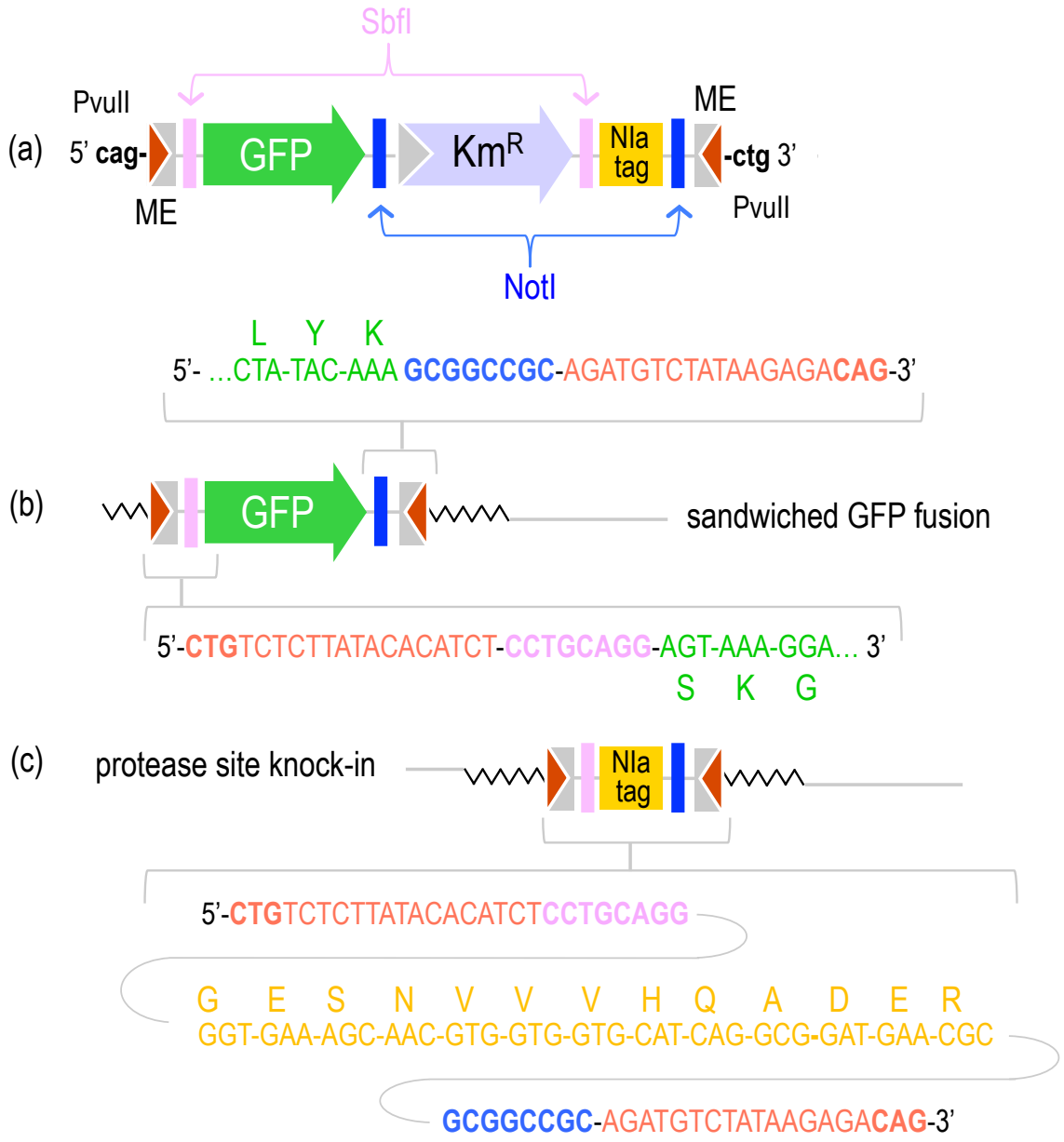
Figure 1. Functional organization and mode of action of the *m*-xylene responsive σ^{54}-dependent regulator XylR.



(a) Functional domains of XylR. The organization of the modules that compose this TF is shown with indication of the amino acid residues that define the limits between the functional domains and the localization of the relevant functions within the protein sequence: A (signal reception and inducer binding), B (interdomain linker region), C (binding and hydrolysis of ATP and contacts with the σ^{54}-

RNAP) and D (including a helix turn helix motif, for binding to the UAS of the target promoter DNA). (b) Activation of XylR by *m*-xylene. The drawing sketches how the TF folds such that the N-terminal A domain hinders an activation surface of the regulator. Effector binding to the A domain releases such a intramolecular repression and XylR becomes then competent for interacting with the σ^{54}-RNAP bound further downstream in *Pu* and activating transcription. The same XylR surface can be presented to the σ^{54}-RNAP by deleting the whole A domain, thereby creating a effector-independent and constitutively active variant XylRΔA. **(c)** The *Pu* promoter region. The DNA segment of interest is expanded, showing the location of relevant sequences, including distal and proximal upstream binding sites for the XylR oligomer (UASd and UASp, respectively), the -12/-24 motif recognized by σ^{54}-RNAP, and one integration host factor (IHF) binding site located within the intervening region. The logic of such an arrangement is an AND gate (inputs *m*-xylene and XylR) followed by a YES operator. If XylR has a default value of 1, then the regulatory node becomes a factual YES gate with *m*-xylene as input and transcription initiation as output (*polymerase per second* or PoPS). **(d)** The P_R promoter region. XylR auto-regulates activity of this σ^{70}-promoter which includes two overlapping initiation sites (P_{R1} and P_{R2}). P_R is repressed by XylR because the UAS of a second divergent σ^{70} promoter (*Ps*) overlap the two -10/-35 sequences that drive divergent transcription of the *xylR* gene. The logic is thus the opposite of that of *Pu*: an AND gate followed by an inverter. As before, if XylR is present throughout, the node becomes a NOT gate with *m*-xylene as input and PoPS as output.

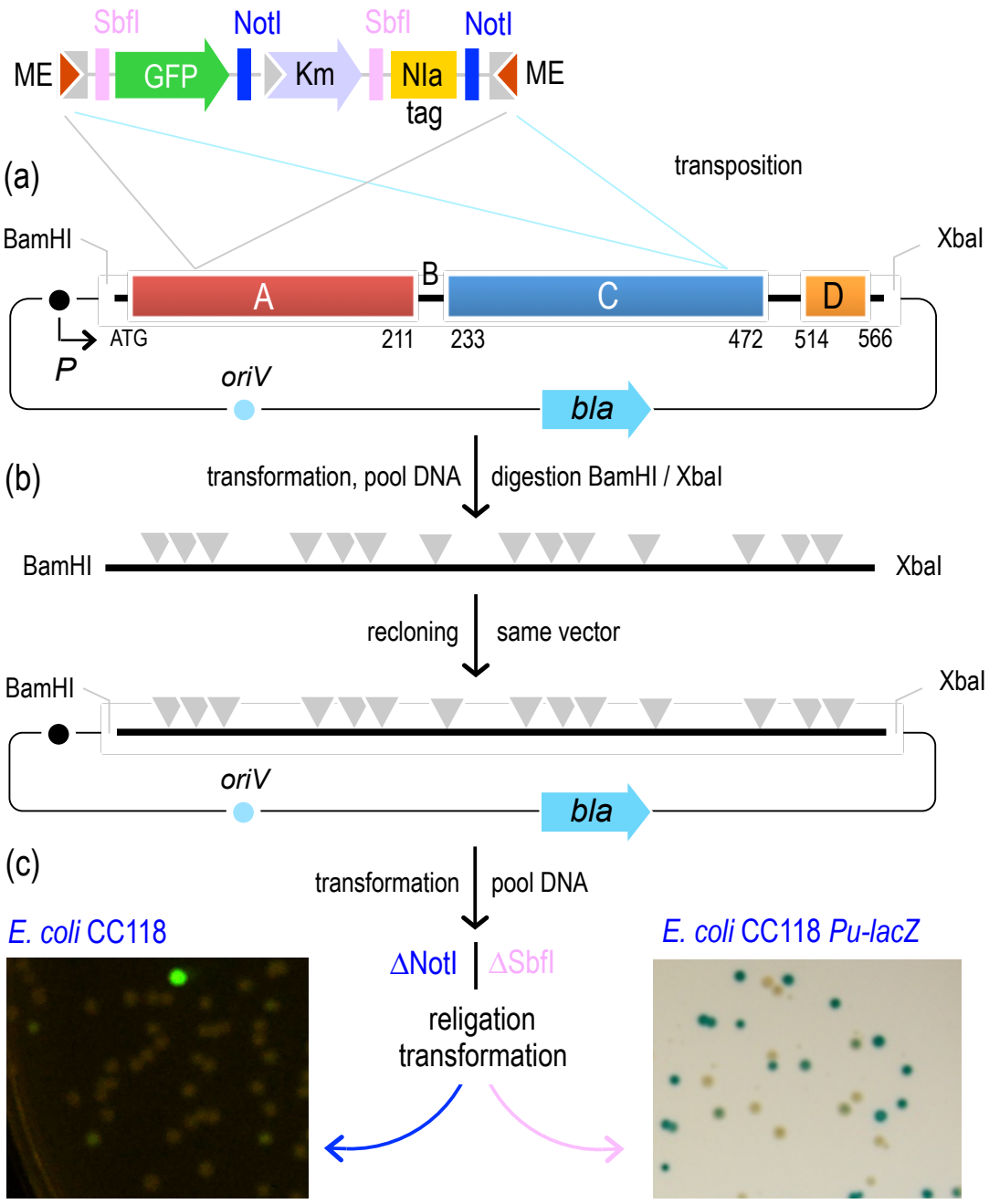
Figure 2. Design and properties of synthetic transposon mTn*5* [GFP•Nla1].



(a) Physical and functional organization. This mobile element is composed by an array of DNA segments which are bracketed by the so-called Tn*5* mosaic ends (ME), i.e. 19 terminal inverted repeats, optimized for hyperactive transposition and both concluding in half PvuII sites. The sequences (left to right in the sketch) between the two MEs ends include [i] a *gfp* gene (GFP) devoid of start and stop codons and bound by restriction sites for the 8-bp cutters SbfI and NotI, [ii] a kanamycin resistance cassette (Km^R) flanked by unique restriction sites SwaI and PshAI (not shown) plus another SbfI and [iii] a 39 pb sequence encoding the peptide that is specifically recognized by the viral NIa protease followed

by one more NotI site. Note the correlation between the two alternate SbfI and Not sites. **(b)** Generation of sandwiched in-frame GFP fusions. Digestion/religation of the transposon-inserted DNA with NotI deletes the Km resistance gene and the NIa target sequence, thereby generating a fusion with both the 5' and the 3' ends of the *gfp* sequence, the boundaries of which are blown up in the sketch. **(c)** Knocking-in target peptides for the NIa protease. Digestion/religation of the same transposon-inserted DNA with SbfI excises the internal GFP/Km segment of mTn*5* [GFP•NIa1] and leaves behind an in-frame addition of the extended amino acid sequence recognized by NIa (in yellow).

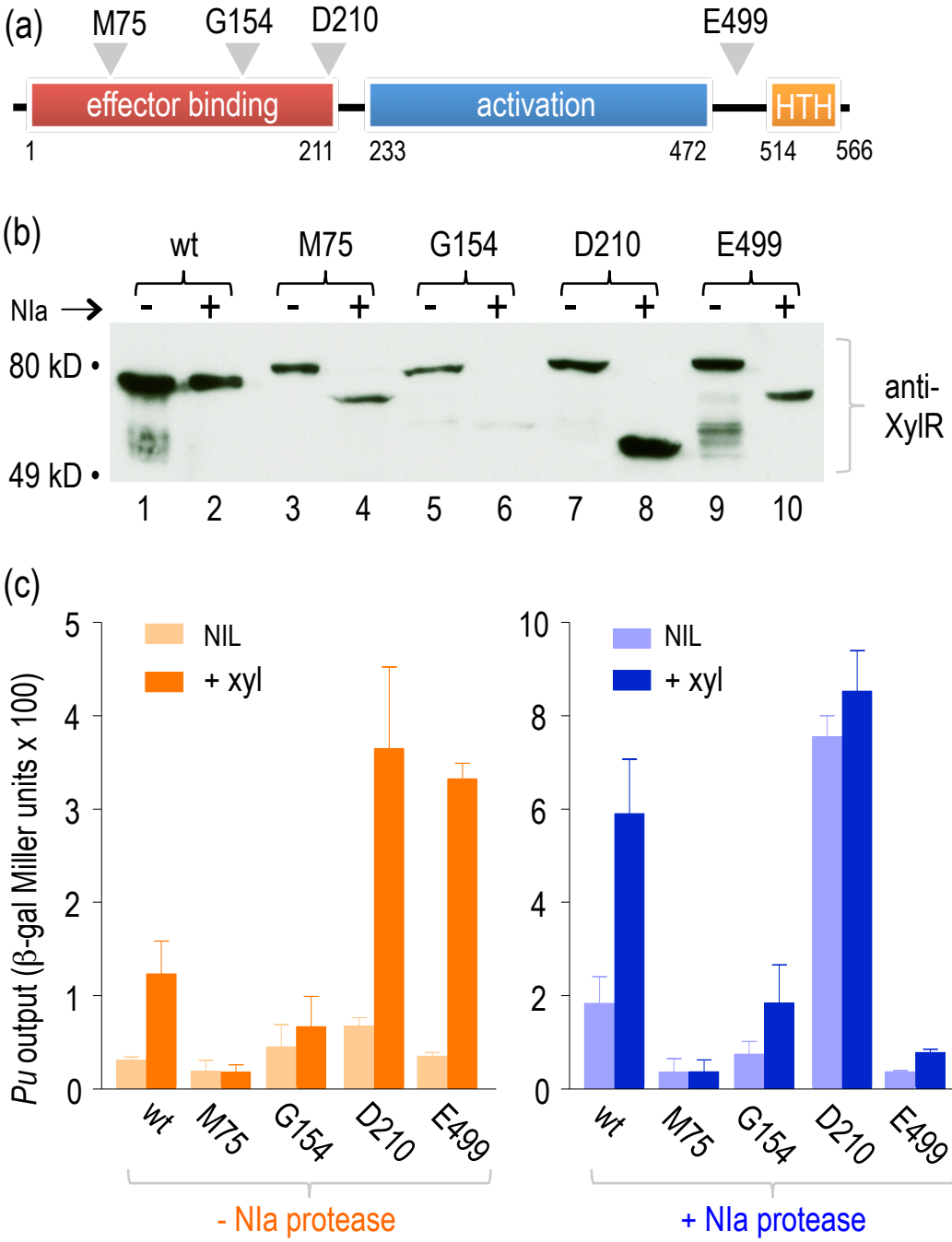
Figure 3. Generation of *knock-in-leave-behind* (KILB) libraries.



(a) *In vitro* mutagenesis. The target gene is cloned in a plasmid as a BamHI-XbaI insert (in this example, sequences corresponding to the functional domains of xylR are indicated) and the DNA is used as the substrate of an in vitro mutagenesis reaction with mTn*5* [GFP•NIa1] as detailed in the Materials and Methods section. **(b)** Recovery of inserted target sequences. The products of the transposition reaction are transformed in *E. coli,* Km^R clones selected, pooled, their plasmids extracted

and the DNA digested with BamHI and XbaI, what allows recovery of a pool of DNA segments with the *xylR* gene inserted randomly with mTn*5* [GFP•NIa1]. This pool (see Supplementary Fig. S1) is then recloned in the BamHI / XbaI sites of the same vector, so that only inserts in the gene of interest are retained. **(c)** Generation of in-frame gene fusions. The ligation mixture is re-transformed and processed in *E. coli* as before (Supplementary Fig. S1b) and the plasmid pool digested and religated with either NotI (thereby creating in-frame sandwich GFP fusions) or with SbfI, that leaves a sequence *scar* that can be cleaved by the NIa protease. The successful production of such knocked-in protein variants can then be tested by transforming the plasmid pool in either plain *E. coli* CC118 and examining the plates with blue light (for GFP expression) or the reporter strain *E. coli* CC118 *Pu-lacZ,* the colonies of which turn blue upon exposure to vapors of the XylR effector *m*-xylene.

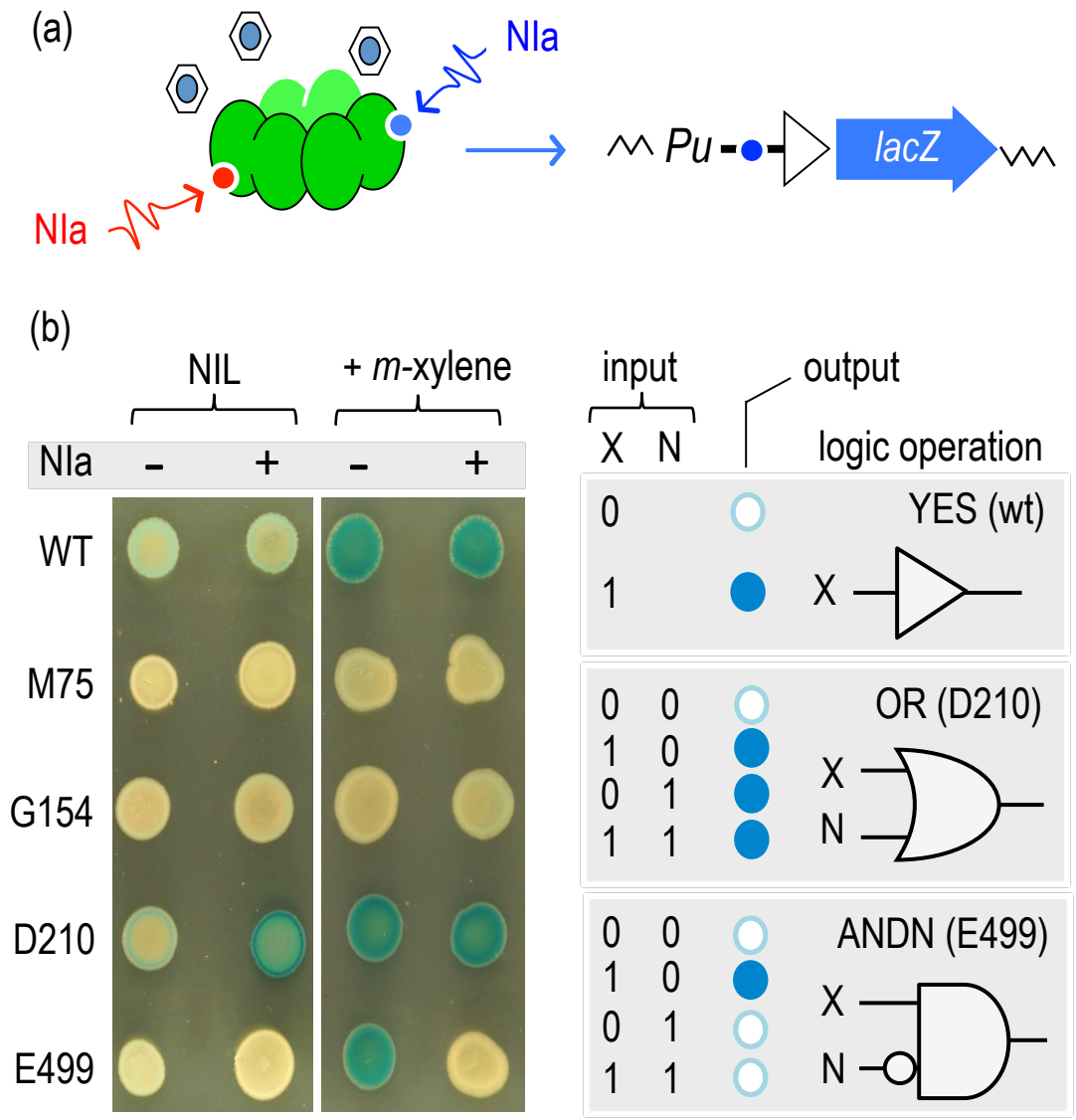
Figure 4. XylR variants knocked-in with NIa protease target sites.



(a) Insertion points of the NIa tag through the protein sequence. The modular organization of XylR is sketched with indication of the permissive locations where the peptide containing the NIa cleavage sites was delivered by the KILB procedure. **(b)** Expression and sensitivity to NIa protease of XylR variants in vivo. Equivalent amounts of crude protein extracts from *E. coli* cultures expressing the XylR types indicated along with NIa (or without protease, as specified) were run in a denaturing gel, blotted and

developed with anti-XylRΔA antibodies. **(c)** Quantification of the activity of NIa-cleavable XylR variants. Cultures of *E. coli* CC118 *Pu-lacZ* strain with plasmids encoding each of the XylR types and the NIa protease were grown and exposed to *m*-xylene as explained in the Materials and Methods section. The diagram plots the accumulation of β-galactosidase after 3 hours of induction with or without the protease as indicated.

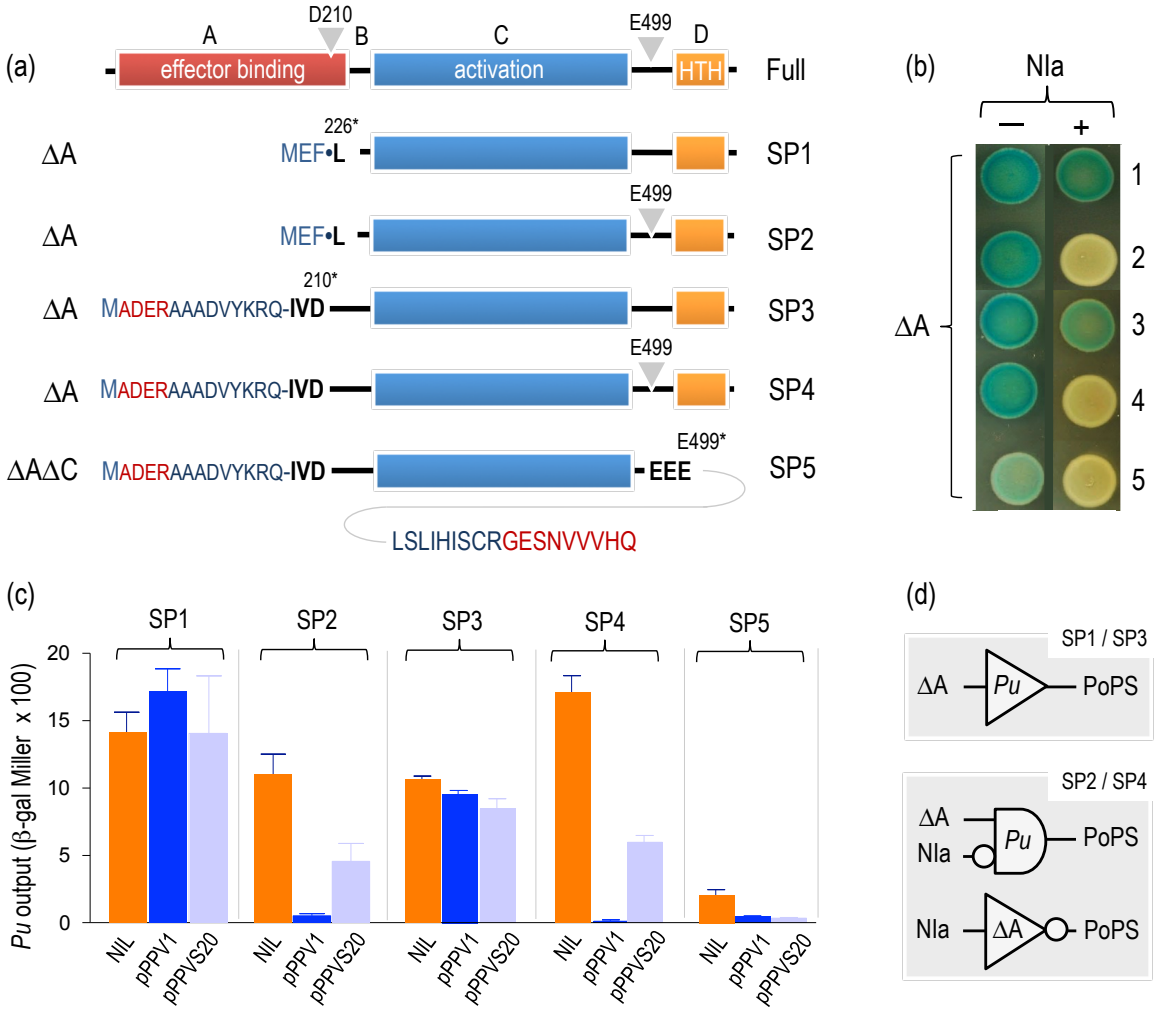
Figure 5. The logic of protease-cleavable XylR variants.



(a) The two inputs of proteolyzable XylR. The drawing represents how cleavage of XylR in alternative sites of the regulator's structure is propagated into the transcriptional activity of the reporter *Pu-lacZ* fusion. **(b)** Visual display of *Pu* activation by NIa-cleavable XylR variants. The left part shows the growth of *E. coli* CC118 *Pu-lacZ* expressing the XylR types labeled to the side and NIa, spotted on LB plates with Xgal and exposed to saturating vapors of *m*-xylene as indicated. The logic gates brought about by XylR versions D210 (cleavage in position 210 of the amino acid sequence, deleting the A domain) and E499 (split by NIa in 499 and excising the D domain) are shown to the right. Wild-type XylR operates as a YES (buffer) gate with *m*-xylene (X) as the only input. $XylR^{D210}$ produces an OR gate with both *m*-

xylene and NIa protease (N) as inputs. Finally, XylR^{E499} generates an ANDN device, where *Pu* activity is on only when one of the inputs is present (X) and the other is absent (N).

Figure 6. The logic or anti-activation of *Pu* by XylRΔA* variants.



(a) Organization of the ΔA versions of XylR in respect to the full-length protein and its NIa-cleavable forms. The upper sketch shows a reference with the sites of the two NIa cutting sites at positions D210 and E499. The synthetic ΔA proteins (SP) below are aligned in respect to such reference with indication of the amino acids that lead the N-terminus and the presence or not of an engineered E499 site. The amino acid sequence of the C-terminus of the SP5 protein variant (XylRΔAΔC) is blown up as well (see Supplementary Fig. S3 for more details on the amino acid termini of each protein). **(b)** *E. coli* CC118 *Pu-lacZ* expressing each of the ΔA XylR types plus minus NIa as indicated and spotted on LB plates with Xgal. **(c)** Quantification of the activity of ΔA XylR variants. *E. coli* CC118 *Pu-lacZ* with plasmids encoding each of the XylR types were grown and NIa expression induced with IPTG. The protease was expressed through two alternative plasmids (pPPV1 and pPPSV20, Supplementary Table S1) as indicated. The graph shows accumulation of β-galactosidase after 3 hours of induction (see expression

and cleavage of each of the XylRΔA variants in vivo in Supplementary Fig. S3). **(d)** Formalization of the regulatory behavior of NIa-cleavable XylRΔA variants as a digital gate. The logic of *Pu* activation by XylRΔA is a YES gate where the TF is the input and PoPS the output. In contrasts, SP3 and SP3 versions of the same regulator generate an ANDN device, where *Pu* activity is on when the ΔA protein is present and the protease is absent. If such ΔA TFs are given a default value of 1, the same device becomes an inverter in which the only input is NIa.

SUPPORTING INFORMATION TO THE ARTICLE

Expanding the Boolean logic of the prokaryotic transcription factor XylR by functionalization of permissive sites with a protease-target sequence

by

Belen Calles and Víctor de Lorenzo*

Systems Biology Program. Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco, 28049, Madrid (Spain)

SUPPLEMENTARY INFORMATION

Structure and modus operandi of the synthetic KILB transposon mTn5 [GFP•NIa1]

The mobile element employed in this work for the knock-in-leave-behind (KILB) procedure adopted in this work has a total size of 1774 bp and allows generation of comprehensive libraries of either in-frame, sandwiched fusions to the green fluorescent protein (GFP) or specific cleavage sites recognized by the plant viral protease NIa (García et al, 1989a). Note that such NIa target sequence (NVVVHQA) is absent from the proteome of E. coli and therefore the duo NIa protease-NIa tagging peptide can be considered an orthogonal device. The engineering of the two cargoes in the same DNA segment allows an estimation of the efficiency of the transposition process as discussed below. The salient features of mTn5 [GFP•NIa1] include the following characteristics. First, the mobile element is flanked by optimized 19 bp inverted repeats (mosaic ends or ME) that are recognized by the Tn5 transposase (Goryshin et al, 1998). Such recognition catalyzes random insertion of mTn5 [GFP•NIa1] into target DNA through a process that results of duplication of the 9 bp adjacent to the transposition site (Reznikoff, 2008). Next to the ME left sequence (ME-L), mTn5 [GFP•NIa1] contains a leaderless gfp gene variant that is optimized for prokaryotic gene fusions to GFP (Miller and Lindow, 1997). Since the gfp sequence lacks a start codon, it should only be translated when inserted within another protein coding sequence in the right orientation and reading frame. mTn5 [GFP•NIa1] insertions that fulfill these conditions can be easily identified for fluorescence emission with the naked eye under blue light. Note also that gfp gene of mTn5 [GFP•NIa1] is flanked by an upstream SbfI site and a downstream NotI site. This is followed by a kanamycin resistance (Km^R) gene (aminoglycoside phosphotransferase, aphA), which is instrumental for selecting transposition events in a wide range of Gram-negative bacteria. The DNA sequence of aphA (the expression of which is driven by its own promoter) was edited to improve codon usage and to eliminate naturally occurring SmaI and HindIII sites that could interfere with subsequent cloning procedures (Martinez-Garcia et al, 2011a). Furthermore, aphA was also flanked by restriction sites that follow the Standard European Vector Architecture (SEVA) format (Silva-Rocha et al, 2013)in such a way that this module can be easily exchanged by any other available marker of the collection (e.g. bla, cat, aadA, tet, aacC1). Finally, the aphA gene born by mTn5 [GFP•NIa1] is followed by a short DNA sequence that encodes the core peptide NVVVHQA that is specifically recognized by the NIa protease of a plant potyvirus (García et al, 1989b) added with three flanking residues at each side (Laín et al,

1989). This was arranged in such a way that the NIa site was delimited by an upstream SbfI site and a downstream NotI site, next to which the mTn5 [GFP•NIa1] ends with the right ME sequence (ME-R) of the mobile element. The alternative tandem arrangement of framed SbfI and NotI sites in mTn5 [GFP•NIa1] allows later excision of much of the inserted transposon after delivery of the mobile element to the target DNA sequence. This leaves in-frame, fit-in GFP or NIa- sequences sandwiched throughout the substrate DNA as described below. Finally, note that either boundary of mTn5 [GFP•NIa1] ends with a half PvuII site within the most external ME sequences. These are intended to ease the cloning and cutting out of the mobile segment in/from the cloning vectors (Supplementary Table S1) as needed for in vivo or in vitro transposition reactions. Two variants of the GFP•NIa -containing transposon were also added to the toolbox (see Materials and Methods section in main text) in which both SbfI sites were replaced by either AscI (resulting in mTn5 [GFP•NIa2]) or by PmeI (same, mTn5 [GFP•NIa3]). The validation of these tools and their exploitation for creating conditionally proteolizable variants of XylR is explained next.

Testing mTn5 [GFP•NIa1] transposition

The transposon described above was first verified for the functionality of all parts embedded in its design. To this end, the DNA sequence spanning the whole mTn*5* [GFP•NIa1] segment was assembled as a PvuII fragment in the backbone of the suicide transposon delivery plasmid pBAM1 (Martinez-Garcia *et al,* 2011a) that encodes both conjugal transfer functions and the Tn*5* transposase protein TnpA. *In vivo* mobilization and transposition assays (Martinez-Garcia *et al,* 2011a) followed by selection of Km^R exconjugants and inspection of green fluorescent colonies suggested the synthetic transposon to work at a frequency of $2.6 \pm 0.1 \times 10^{3}$, creating productive translational GFP fusions at a rate of $1.17 \pm 0.1 \times 10^{-3}$ (not shown). Delivering of the same mTn*5* [GFP•NIa1] segment to *E. coli* by transformation of a preassembled transposome (Goryshin *et al,* 2000) increased the frequency of both events (insertion and production of chromosomal GFP fusions) by 10-fold. While these results accredited the performance of the engineered transposon, the relatively low numbers discouraged its use *in vivo* and advised instead adoption of an all-*in vitro*, high insertional density alternative. For this we set up a method with three purified components i.e. the mTn*5* [GFP•NIa1], a hyperactive variant of the TnpA transposase (purified from expression plasmid pGRTYB35, see Materials and Methods) and the target DNA. The last is plasmid pBCL4, consisting of plasmid vector pUC18deleted of the SbfI site and

carrying the *xylR* sequence as a BamHI-XbaI restriction fragment. *In vitro* reactions were optimized for maximizing the efficiency of transposon insertions. For this, the mTn*5* [GFP•NIa1] element was entered in the reaction mix either as DNA fragment spanning exclusively the sequences bound by the Tn*5* ends or as part of the supercoiled or else linearized pBAM1-GFP plasmid (Martinez-Garcia *et al,* 2011a). Similarly, the target DNA was brought into the system either as a supercoiled or a linearized plasmid (in this last case, followed by a ligation of the transposition products, see below). The efficiency of each of the combinations could be easily quantified by transforming the transposition mixture in *E. coli* and selecting for Ap^R Km^R. These tests revealed that the best insertion rates were brought about by using the very mTn*5* [GFP•NIa1] DNA segment as the transposon donor to the *xylR*-encoding supercoiled plasmid (5 x 10^5 Ap^R Km^R c.f.u / pmol DNA mix), much above using pBAM1-GFP (4.3 x 10^4 c.f.u / pmol DNA) or its linearized version (8.2x10^4 c.f.u / pmol DNA). The scenario was thus set for the generation of high-density knock-in / leave behind (KILB) libraries of protease-target sites through the sequence of *xylR* along the lines explained in the rationale above.

SUPPLEMENTARY TABLES

Supplementary Table S1. Strains and plasmids used in this study

Strain or plasmid	Description	References and/or source
E. coli strains		
DH10B	F- *endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-*	Invitrogen
DH5α	*supE44, ΔlacU169, (ϕ80 lacZΔM15), hsdR17, (rk^-mk^+), recA1, endA1, thi1, gyrA, relA*	Invitrogen
CC118	F-, Δ(*ara-leu*)7697, *araD*139, Δ(*lac*)X74, *phoA*Δ20, *galE, galK, thi, rpsE, rpoB, argE*(Am), *recA1*	Manoil and Beckwith, 1985
CC118 *Pu-lacZ*	CC118 strain with a chromosomal *Pu-lacZ* transcriptional fusion assembled within a streptomycin resistance mini-transposon.	de Lorenzo *et al,* 1991
P. putida strains		
mt-2 (ATCC33015)	Wild-type *P. putida* strain carrying TOL plasmid pWW0	Williams and Murray, 1974; Worsey and Williams, 1975
Plasmids		
pGEM-T	Ap^R, Cloning vector	Promega
pUC18	Ap^R, *oriV* ColE1, cloning vector	Vieira and Messing, 1982; Yanisch-Perron *et al,* 1985
pUC18-SbfI	pUC18 derivative lacking PstI/SbfI sites	This work
pBCL4	pUC18-SbfI derivative with BamHI/XbaI fragment encoding *xylR* gene	This work

pMS-RQ	Sp^R, *ori*ColE1 cloning vector	GeneArt®
pGA-BCL1	pMS-RQ derivative harboring mTn*5* [GFP•NIa1] transposon at SfiI site	This work
pBAM1	Ap^R, *ori*R6K, *ori*T, *tnp*A, mini-Tn5 carrying a Km^R gene. GenBank HQ908071	Martinez-Garcia *et al,* 2011b
pBAM1-GFP	pBAM1 derivative harboring mTn*5* [GFP•NIa1] transposon inserted in PvuII site. GenBank HQ908072	Martinez-Garcia *et al,* 2011b
pGA-BCL2	pMS-RQ derivative harboring mTn*5* [GFP•NIa2] transposon at SfiI site	This work
pGA-BCL3	pMS-RQ derivative harboring mTn*5* [GFP•NIa3] transposon at SfiI site	This work
pBCL4-M75	pBCL4 with *xylR* gene inserted with NIa sequence after M75 residue ($xylR^{M75}$) by transposition with mTn*5* [GFP•NIa1]	This work
pBCL4-G154	pBCL4 with *xylR* gene inserted with NIa sequence after G154 residue ($xylR^{G154}$) by mTn*5* [GFP•NIa1] transposition	This work
pBCL4-D210	pBCL4 with *xylR* gene inserted with NIa sequence after D210 residue ($xylR^{D210}$) by mTn*5* [GFP•NIa1] transposition	This work
pBCL4-E499	pBCL4 with *xylR* gene inserted with NIa sequence after E 499 residue ($xylR^{E499}$) by mTn*5* [GFP•NIa1] transposition	This work
pBCL4-SP1	pUC18 derivative containing BamHI/XbaI fragment enconding SP1 XylR variant	This work
pBCL4-SP2	pUC18 derivative containing a *Bam*HI/*Xba*I fragment enconding SP2 XylR variant	This work
pBCL4-SP3	pUC18 derivative containing a *Bam*HI/*Xba*I fragment enconding SP3 XylR variant	This work
pBCL4-SP4	pUC18 derivative containing a *Bam*HI/*Xba*I fragment enconding SP4 XylR variant	This work

pBCL4-SP5	pUC18 derivative containing a BamHI/XbaI fragment enconding SP5 *xylR* variant	This work
pPPVs20	pSU8 derivative containing the SalI-PstI fragment of PPV cDNA consisting of the 3´terminal region from nt 3627	Garcia *et al,* 1989
pPPV1	pVTR-B plasmid containing 0.6 Kb StuI-HindIII fragment encoding PPV NIa protease from pPPVs20 plasmid	Perez-Martin *et al,* 1997

1

2

3

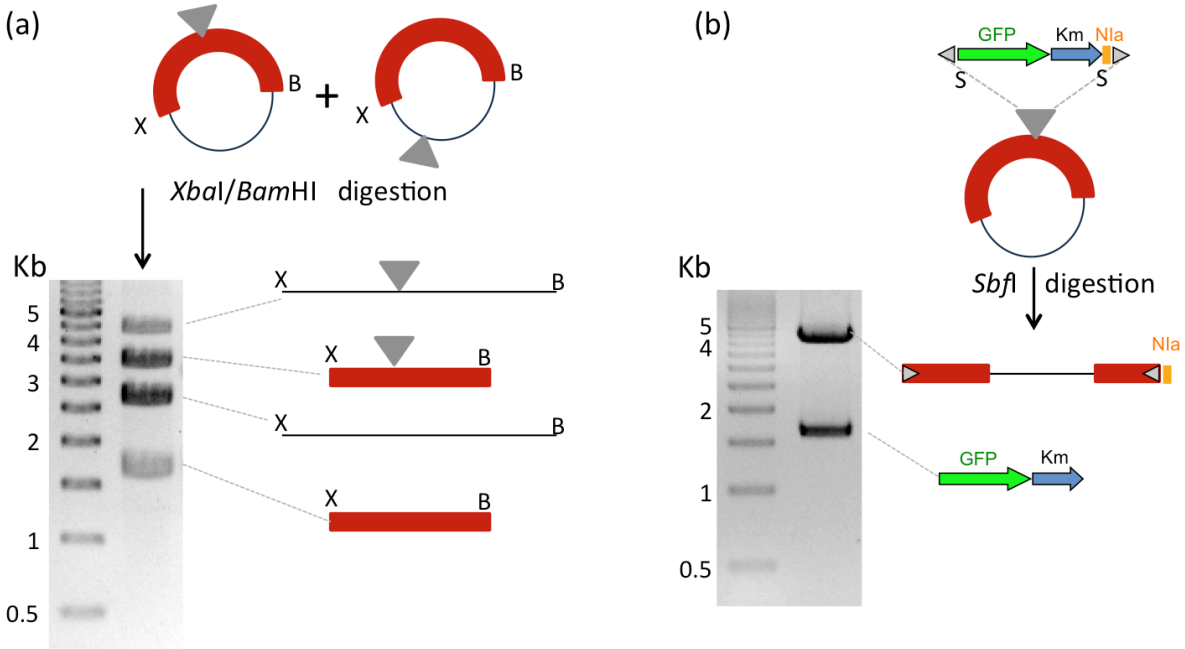
Supplementary Table S2: primers used in the PCR reactions [a)]

Oligonucleotide	Sequence
Tn5ME-F	CTGTCTCTTATACACATCTCCTG
Tn5ME-R	CTGTCTCTTATACACATCTGCGG
xylR-BamHI	ATGGATCCAAGAGGAAAACAAATGTCGC
xylR-XbaI	GTTCTAGACTATCGGCCCATTGCTTTC
Km-SwaI-F	CGCGCGATTTAAATTTGTGTCTCAAAATCTCTGATGTTA
Km-PshAI-R	CGCGCGGACCGCGGTCCAATTAATTATTAGAAAAATT
KpnI-AscI(1)-F	CGCATGGTACCCAGCTGTCTCTTATACACATCT*GGCGCGCC*AGTAAAGGAGAAGAACTTTTCAC
KpnI-PmeI(1)-F	CGCATGGTACCCAGCTGTCTCTTATACACATCT*GTTTAAAC*AGTAAAGGAGAAGAACTTTTCAC
AscI(2)-F	CTAATAATTAATTGGACCGCGGTCCGCGCG*GGCGCGCC*GGTGAAAGCAACGTGGTGGTG
PmeI(2)-F	CTAATAATTAATTGGACCGCGGTCCGCGCG*GTTTAAAC*GGTGAAAGCAACGTGGTGGTG
ApaLI-R	GGGTTCGTGCACACAGCCCAGCTTGGAGCGAAC
DeltaA2-F	CCCGGGGATCCAAGAGGAAAACAAATGGAATTTCTGAAGCAGTACGATGGGCAG
D210F	GATAGGATCCAAGAGGAAAACAAATGGCGGATGAACGCGCGGCCGCAG
D210F1	GATAGGATCCAAGAGGAAAACAAATG
XylR-Sol.F	GAGCGCGGGGTGATTCTTACCGAGAG
XylR-Sol.R	CTCTCGGTAAGAATCACCCCGCGCTC
E499stop-R	GATATCTAGACTACGCCTGATGCACCACCACGTTGCTTTC

[a)] Sequences of oligonucleotides employed in this study. Restriction sites entered for cloning purposes are underlined. Sequences corresponding to mutations are indicated in italics.

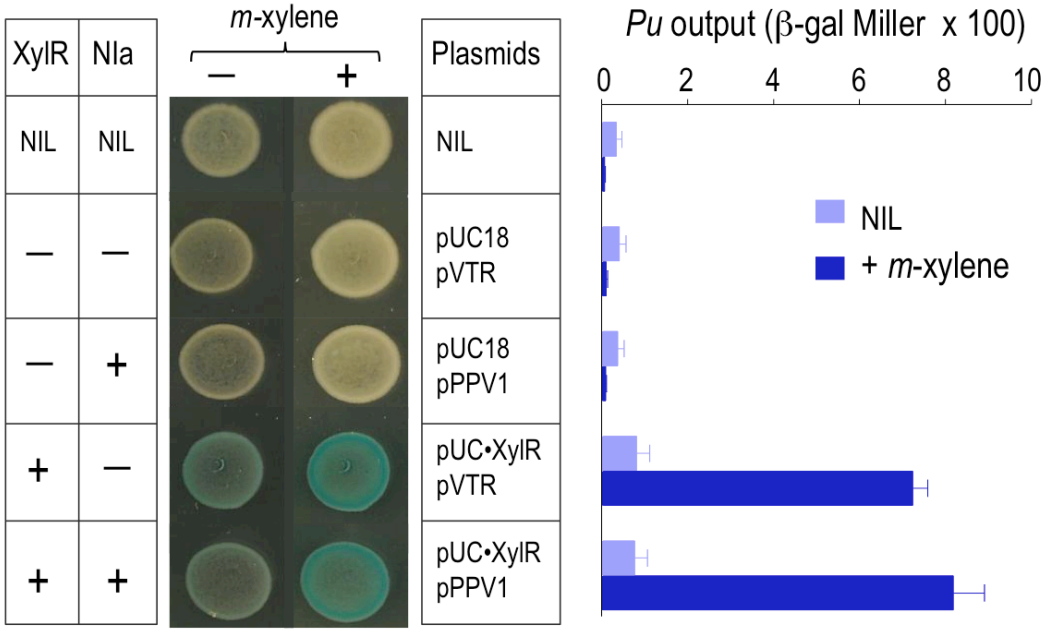
SUPPLEMENTARY FIGURES

Supplementary Figure S1. Agarose gel analysis of DNA intermediates during KILB insertion saturation of a targer gene with mTn*5* [GFP•NIa1].



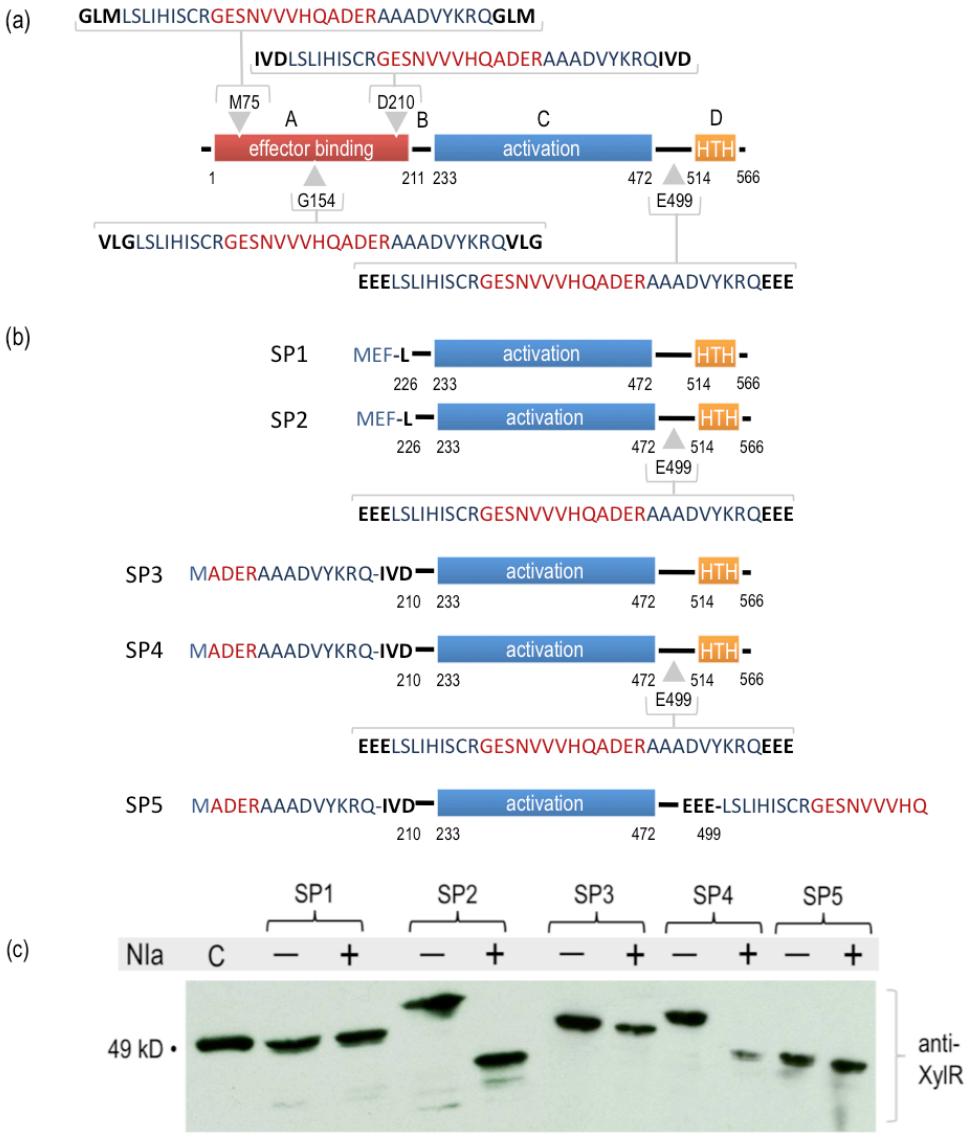
The figure shows two key steps of the process. Panel **(a)** illustrates the patter of the 4 DNA products that result from digesting a plasmid with the gene of interest (*xylR* in this case) with XbaI and BamH (the sites that flank the cloned gene) following saturation insertions with mTn*5* [GFP•NIa1] *in vitro* and recovery of plasmids (see main text for explanation). The pool of same-size DNA fragments with the *xylR* gene densely punctuated with transposon insertions appears clearly separated from the rest of the DNA segments: inserted plasmid backbone, non-inserted counterpart and non-inserted *xylR*. Panel **(b)** shows the excision of most of the mTn*5* [GFP•NIa1] insert from the xylR sequences by digestion of the cognate plasmids with SbfI. Religation of the DNA of the upper DNA band of the gel generates an in-frame library of XylR variants inserted with cutting sites for protease NIa. Alternative digestion with NotI would have similarly created a library of GFP sandwiched in-frame fusions (not shown).

Supplementary Figure S2. Regulatory phenotypes of XylR-encoding and NIa protease-encoding plasmids and cognate insertless vectors.



(a) *E. coli* CC118 *Pu-lacZ* expressing XylR and NIa through the plasmids indicated (Supplementary Table S1) and spotted on LB plates with Xgal and exposed or not to saturating vapors of m-xylene. **(b)** Reference values of *Pu* activity. Liquid cultures of *E. coli* CC118 *Pu-lacZ* transformed with plasmids encoding XylR and NIa protease and their vectors were grown and exposed to *m*-xylene as explained in the Materials and Methods section. The diagram shows the levels of β-galactosidase detected after 3 hours of induction.

Supplementary Figure S3. Organization and *in vivo* expression of NIa-cleavable XylR and XylRΔA variants.



(a) Detail of the amino acid sequences delivered to permissive sites of the XylR structure upon insertion saturation of the *xylR* gene with the KILB procedure. The scheme shows the location of the various NIa cutting sites with an indication of the new sequences introduced in each case: native XylR amino acids in bold, transposition scar sequences entered by the trace of the Tn5 termini in plain (capital) letters and the peptide targeted by the NIa protease in red. **(b)** Blowup of the C- and N-termini of synthetic ΔA proteins (SP) engineered or not with a NIa cleavage site. The amino acid sequence of the C-terminus of the SP5 protein variant (XylRΔAΔC) is blown up as well (color and letter codes same than before). **(c)** Expression and sensitivity to NIa protease of SP XylR variants in vivo. Equivalent amounts of crude

protein extracts from *E. coli* cultures expressing the proteins indicated along with NIa (or without protease, as specified) were run in a denaturing gel, blotted and developed with anti-XylRΔA antibodies.

SUPPLEMENTARY REFERENCES

de Lorenzo V, Herrero M, Metzke M, Timmis KN (1991) An upstream XylR- and IHF-induced nucleoprotein complex regulates the sigma 54-dependent *Pu* promoter of TOL plasmid. *EMBO J* **10:** 1159-1167

García JA, Riechmann J, Lain S (1989a) Proteolytic activity of the plum pox potyvirus Nla-like protein in *Escherichia coli. Virology* **170:** 362-369

García JA, Riechmann JL, Laín S (1989b) Artificial cleavage site recognized by plum pox potyvirus protease in Escherichia coli. *J Virol* **63:** 2457-2460

Garcia JA, Riechmann JL, Martín MT, Laín S (1989) Proteolytic activity of the plum pox potyvirus NIa-protein on excess of natural and artificial substrates in *Escherichia coli. FEBS Letters* **257:** 269-273

Goryshin IY, Jendrisak J, Hoffman LM, Meis R, Reznikoff WS (2000) Insertional transposon mutagenesis by electroporation of released Tn*5* transposition complexes. *Nat Biotech* **18:** 97-100

Goryshin IY, Miller JA, Kil YV, Lanzov VA, Reznikoff WS (1998) Tn5/IS50 target recognition. *Proc Natl Acad Sci USA* **95:** 10716-10721

Laín S, Riechmann J, García JA (1989) The complete nucleotide sequence of plum pox potyvirus RNA. *Virus Res* **13:** 157-172

Manoil C, Beckwith J (1985) Tn*phoA*: a transposon probe for protein export signals. *Proc Natl Acad Sci USA* **82:** 8129-8133

Martinez-Garcia E, Calles B, Arevalo-Rodriguez M, de Lorenzo V (2011) pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol* **11:** 38

Miller WG, Lindow SE (1997) An improved GFP cloning cassette designed for prokaryotic transcriptional fusions. *Gene* **191:** 149-153

Perez-Martin J, Cases I, de Lorenzo V (1997) Design of a solubilization pathway for recombinant polypeptides *in vivo* through processing of a bi-protein with a viral protease. *Protein Eng* **10:** 725-730

Reznikoff WS (2008) Transposon Tn5. *Annual Review of Genetics* **42:** 269-286

Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, de las Heras A, Páez-Espino AD, Durante-Rodríguez G, Kim J, Nikel PI, Platero R, de Lorenzo V (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucl Acids Res* **41:** D666-D675

Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19:** 259-268

Williams PA, Murray K (1974) Metabolism of benzoate and the methylbenzoates by *Pseudomonas putida (arvilla)* mt-2: Evidence for the existence of a TOL plasmid. *J Bacteriol* **120:** 416-423

Worsey MJ, Williams PA (1975) Metabolism of toluene and xylenes by *Pseudomonas putida (arvilla)* mt-2: evidence for a new function of the TOL plasmid. *J Bacteriol* **124:** 7-13

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33:** 103-119