

# Locked and proteolysis-based transcription activator-like effector (TALE) regulation

Jan Lonzarić<sup>1,2,3</sup>, Tina Lebar<sup>1,2,3</sup>, Andreja Majerle<sup>1</sup>, Mateja Manček-Keber<sup>1,2</sup> and Roman Jerala<sup>1,2,\*</sup>

<sup>1</sup>Laboratory of Biotechnology, National Institute of Chemistry, Ljubljana 1000, Slovenia, <sup>2</sup>EN-FIST Centre of Excellence, Ljubljana 1000, Slovenia and <sup>3</sup>Graduate School of Biomedicine, University of Ljubljana, Ljubljana 1000, Slovenia

Received June 03, 2015; Revised December 22, 2015; Accepted December 26, 2015

## ABSTRACT

Development of orthogonal, designable and adjustable transcriptional regulators is an important goal of synthetic biology. Their activity has been typically modulated through stimulus-induced oligomerization or interaction between the DNA-binding and activation/repression domain. We exploited a feature of the designable Transcription activator-like effector (TALE) DNA-binding domain that it winds around the DNA which allows to topologically prevent it from binding by intramolecular cyclization. This new approach was investigated through noncovalent ligand-induced cyclization or through a covalent split intein cyclization strategy, where the topological inhibition of DNA binding by cyclization and its restoration by a proteolytic release of the topologic constraint was expected. We show that locked TALEs indeed have diminished DNA binding and regain full transcriptional activity by stimulation with the rapamycin ligand or site-specific proteolysis of the peptide linker, with much higher level of activation than rapamycin-induced heterodimerization. Additionally, we demonstrated reversibility, activation of genomic targets and implemented logic gates based on combinations of protein cyclization, proteolytic cleavage and ligand-induced dimerization, where the strongest fold induction was achieved by the proteolytic cleavage of a repression domain from a linear TALE.

## INTRODUCTION

Recent advances in synthetic biology have introduced new generations of designable DNA-binding proteins, such as zinc fingers (1), transcription activator like effectors (TALE) (2–4), and most recently the CRISPR/Cas system that relies on guide RNA to determine binding specificity

(5). The crystal structure of a TALE DNA binding domain (DBD) (6,7) shows that it forms a superhelical structure wound around the target DNA, with interactions between the nucleobases and the recognition repeat variable diresidues (RVDs) of each TALE repeat. With the targeting specificity problem thus mostly solved, the next important challenge of designable transcription factors is regulation of their activity by selected chemical or physical signals. In nature, inducible DNA-binding proteins often depend on post-translational modifications (phosphorylation) or allosteric control, which are difficult to engineer; however, small molecule-dependent transcription factors found in bacteria (e.g. tetR (8)) can be adapted as orthogonal mediators for use in mammalian cells but cannot target arbitrary DNA sequences. Small molecule-inducible systems have been implemented previously through heterodimerization of the transcription activation or repressor domain with the DNA-binding domain of TALEs (9) and zinc fingers (10), or based on the reconstitution of a split Cas9 protein upon ligand binding (11). The drawback of systems that rely on heterodimerization is often their low level transcriptional activation or repression as compared to single molecule transcriptional effectors. Additionally there are only a small number of well characterized, reliable, orthogonal ligand-dependent heterodimerization domains.

Another tool that has recently been widely adopted by synthetic biologists is protein splicing by inteins (recently reviewed by Li (12)). Inteins are small protein domains that undergo autocatalytic splicing, which enables the excision of the intein from the mature protein, while a new peptide bond is formed between the extein parts N- and C-terminally of the intein. Several naturally (13,14) and artificially split inteins (15,16) are known to undergo bimolecular complementation and splicing *in trans*, allowing posttranslational formation of a peptide bond between two proteins bound to the intein fragments. Inteins have been adapted as self-cleavable tags for protein isolation (17,18), split inteins can be used for peptide and protein cyclization, where cyclic proteins are more stable than their linear counterparts and

\*To whom correspondence should be addressed. Tel: +386 1 4760 335; Fax: +386 1 4760 300; Email: roman.jerala@ki.si

can retain their structure and function (19–24) and split inteins have been used to reconstitute both zinc finger (25) and TALE (26) designable effectors in mammalian cells.

Here, we explore an approach to regulation of transcription factors that relies on protein cyclization of the DNA-binding domain. We realized that the winding of TALE domains around DNA opens a prospect of topological control of TALE–DNA interaction and thereby a new mechanism of regulation. The locking of the TALE DBD into a circular conformation could impose a topological constraint prohibiting loading and unloading of the locked protein onto and from DNA, thus either preventing binding or additionally stabilizing it, if the circularization occurs after the protein is already bound to DNA. We show that this type of regulation resulted in a higher absolute level of activation in comparison to regulation based on heterodimerization of DNA-binding and effector domains. Furthermore, combinations of heterodimerization and locking were used to generate effectors that emulate logical functions, integrating chemical inducer- and proteolysis-triggered signals.

## MATERIALS AND METHODS

### Recombinant DNA constructs

The TALE(N) DNA binding domain was cloned from the NicTAL effector (27) (synthesized by Life Technologies). TALE(B) and TALE(D) DNA binding domains were cloned from TALEN1297 (Addgene plasmid # 32279) and TALEN1295 (Addgene plasmid number # 32283) from Keith Joung (28). TALE(A), TALE(C) and TALE(E) DNA binding domains were constructed using the TALEN Assembly Kit (Labomatics). DmrC, DmrA and DmrD domains were obtained from plasmids pC4-RHE, pC4EN-F1 and pC4EN-FM2E, respectively (previously ARIAD, now pHet-1, pHet-Nucl and prHom-1 respectively, Clontech). IntN and IntC domains of the Npu intein were obtained from the pSKDuet01 (Addgene plasmid # 12172) and pSKBAD2 (Addgene plasmid # 15335) from Hideo Iwai (29). The KRAB domain was obtained from pLVPT-rtTR-KRAB-2SM2 (Addgene plasmid # 11652) from Patrick Aebischer & Didier Trono (30) and the VP16 domain from the vector pSGVP, which was provided by Professor Mark Ptashne (31) (Memorial Sloan Kettering Cancer Center, New York, NY, USA). All cloning of the TALEs was performed using Gibson assembly (32) and the various linkers, tags, the nuclear localization sequence and the TEVp cleavage site were inserted as part of PCR amplification primers and overlapping sequences for the Gibson assembly. All TALEs were expressed from the CMV promoter of the pcDNA3 vector. A list of all effector constructs is provided in Supplement Table S1 and detailed protein sequences of all effector constructs are provided in Supplement Table S2.

TALE-binding sites were synthesized by Life Technologies, PCR amplified with a primer containing a minimal promoter (5'-TAGAGGGTATATAATGGAAGCTCGACTTCCAG-3') and cloned upstream of the firefly luciferase gene in pGL4.16 (Promega), replacing the original inducible promoter. Additionally they were cloned upstream of the CMV promoter of the pcDNA3 plasmid with the firefly luciferase gene from pGL4.16 downstream of the promoter.

The promoter and TALE-binding sequences of all reporter plasmids used are provided in Supplement Table S3. The transfection control plasmid used in the dual luciferase assay was phRL-TK (Promega).

The TEVp gene was synthesized by Life Technologies as split N and C fragments and the Dmr-TEVp fusions as well as the complete TEVp sequence were cloned using Gibson assembly with primers encoding the nuclear localization sequence and inserted into the pcDNA3 vector. BFP was obtained from pTagBFP-N (Evrogen), fused to intein domains using Gibson assembly and cloned into the pcDNA3 vector.

### Cell culture and dual luciferase assay

The human embryonic kidney (HEK) 293T cell line was cultured in DMEM medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS) (BioWhittaker, Walkersville, MD, USA) at 37°C in a 5% CO<sub>2</sub> environment.  $2 \times 10^4$  cells per well were seeded in a Costar White 96-well plate (Corning). At 70% confluence, they were transfected with a mixture of jetPEI (Polyplus transfection) and plasmids as indicated in individual figures together with the constitutive *Renilla* luciferase transfection control. Where appropriate, cells were stimulated immediately after transfection with an addition of rapamycin to the final concentration of 2  $\mu$ M and DMSO with a final concentration of 0.2% (v/v). Two days after transfection, cells were harvested and lysed with 30  $\mu$ l of 1 $\times$  Passive lysis buffer (Promega). For the reversibility experiments, every 24 h cells were washed and supplied with fresh medium containing 0.2% DMSO (v/v) and, where appropriate, 2  $\mu$ M rapamycin and then harvested at the indicated timepoints. Firefly luciferase and *Renilla* luciferase activity was measured using an Orion II microplate reader (Berthold Technologies, Pforzheim, Germany). Relative luciferase activity (RLU) was calculated by normalizing each sample's firefly luciferase activity with the constitutive *Renilla* luciferase activity determined within the same sample. The results represent the mean and standard deviation of four measurements in separate wells. The data are representative of at least two independent experiments.

### Pull-down assay

For immunoblotting and precipitation experiments,  $5 \times 10^5$  HEK293T cells per well were seeded in Costar 6-well plates (Corning) and at 70% confluence transfected with a mixture of jetPEI (Polyplus transfection) and plasmids. Three days after transfection, cells were harvested, washed with phosphate buffered saline and lysed in 150  $\mu$ l of 1 $\times$  passive lysis buffer (Promega) supplemented with complete protease inhibitor cocktail tablets (Roche). Lysed cells were centrifuged at 14 000 rpm for 30 min at 4°C and only the supernatant was used for subsequent experiments.

For precipitation experiments, one volume of binding buffer (5 mM EDTA, 5% glycerol, 0.1% Tween-100, 50 mM NaCl, 20 mM Tris, pH 8) and two volumes of 2 $\times$  DNA blocking buffer (1 $\times$  binding buffer, 2 mg/ml Herring sperm DNA (Sigma), 2 $\times$  protease inhibitor cocktail (Sigma)) were added to the cell lysate supernatants. One half of this solution was treated with ProTEV Plus protease (Promega)

and both halves were incubated at room temperature for 2 h while magnetic beads were prepared. DNA with the appropriate TALE-binding sequences was PCR amplified from the reporter vector pcDNA3.12xN.CMV\_mCit (TALE(N)-binding sites) or pcDNA3.12xD.CMV\_mCit (TALE(D) binding sites) using biotin labeled primers (5'-biotin-TCTCAGTACAATCTGCTCTG-3' and 5'-Biotin-GTAACGCGGAAGTCCATATATGGGC-3') and purified using FastGene Gel/PCR extraction kit (Nippon Genetics). 100  $\mu$ l of Dynabeads MyOne Streptavidin C1 magnetic beads (Life Technologies) were washed three times with 2 $\times$  W&B buffer (1 mM EDTA, 2 M NaCl, 10 mM Tris, pH 7.5) on a magnetic rack. 4  $\mu$ g of biotinylated DNA in 100  $\mu$ l of 1 $\times$  W&B buffer and additional 100  $\mu$ l of BSA blocking buffer (1 $\times$  W&B buffer, 0.5% bovine serum albumin (Sigma)) were added to the beads and incubated at 4°C for 90 min with 1 krpm shaking. Unbound DNA was removed by washing the beads twice with W&B buffer and twice with binding buffer. Finally, the beads were aliquoted into two equal parts and 200  $\mu$ l of the treated cell lysates were added. The beads were incubated at 4°C overnight with 1 krpm shaking. The unbound fraction was removed and used for PAGE analysis, while the beads were washed five times with binding buffer and finally the bound proteins eluted with 45  $\mu$ l of Laemmli buffer (2% SDS, 10% glycerol, 5% 2-mercaptoethanol, 0.002% bromophenol blue, 62.5 mM Tris, pH 6.8) at 85°C for 10 min, appropriately diluted (to match the dilution of the unbound fraction) and used for PAGE analysis.

### Electrophoresis and immunodetection

Cell lysates or supernatants from the pull-down assay were mixed with 4 $\times$  Laemmli buffer (to the final concentration of 1 $\times$  Laemmli buffer), incubated at 85°C for 10 min and separated on a 10% polyacrylamide gel in SDS running buffer (190 mM glycine, 0.1% SDS, 25 mM Tris, pH 8.3). Proteins were transferred to Hybond ECL membranes (GE Healthcare) in transfer buffer (190 mM glycine, 20% methanol, 25 mM Tris, pH 8.3) at 350 mA for 2 h. Membranes were washed with deionized water and blocked overnight at 4°C in a solution of 0.2% I-Block (Life Technologies) in PBS with 0.1% Tween-20 (PBS-T). Blocked membranes were incubated for 1.5 h with a solution of primary antibodies (rabbit anti-FLAG (Sigma) in 0.2% I-Block, washed four times with PBS-T, incubated with a solution of HRP conjugated secondary antibodies (Goat Anti-Rabbit IgG H&L (Abcam)) and again washed four times with PBS-T. Detection was performed with SuperSignal West Pico Chemiluminescent Substrate (Life Technologies) in the G:BOX (Syngene) detector and the figures processed using the ImageJ software.

### RNA isolation, reverse transcription and qPCR

For genomic target activation  $1.5 \times 10^5$  HEK293T cells per well were seeded in Costar 24-well plates (Corning) and at 70% confluence transfected with a mixture of jetPEI (Polyplus transfection) and plasmids. Three days after transfection, cells were harvested, washed with phosphate buffered saline and total RNA was isolated

with the high-pure RNA isolation kit (Roche). DNA was digested using RQ1 RNase-Free DNase (Promega) and reverse transcription was performed with a random mix of primers using the High capacity cDNA reverse transcription kit (Applied Biosystems). The obtained cDNA was diluted 100 $\times$  and qPCR was performed with LightCycler480 SYBR green I Master mix (Roche) on the LC480 LightCycler (Roche) using primers 5'-CCGAGAGAGTTTCCCTACGTATACCCTG-3' and 5'-CTTTCAGTGTGGTGATTACGACGTTAGC-3' for CDH1 and primers 5'-CCAGGGCTGCTTTTAACTCTGGTAAAGTGG-3' and 5'-ATTTCCATTGATGACAAGCTTCCCGTTCTC-3' for GAPDH internal control. pcDNA3 transfected cells were used as normalization control. Relative mRNA levels were calculated using the formula  $\text{mRNA} = 2^{-\Delta[\text{Cp}(\text{CDH1, control}) - \text{Cp}(\text{GAPDH, control})]}$  ( $\text{Cp}(\text{CDH1, sample}) - \text{Cp}(\text{GAPDH, sample})$ ).

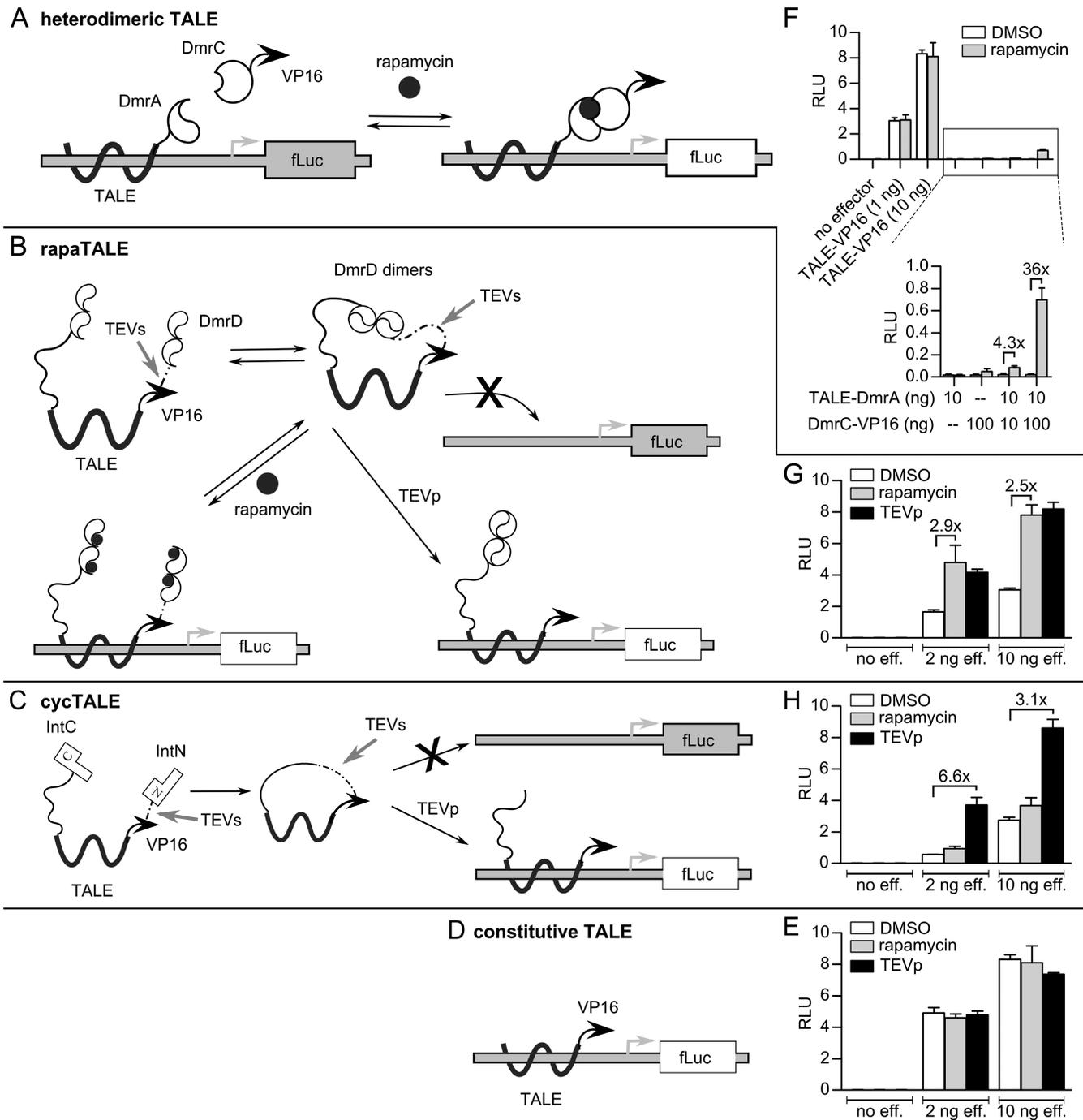
## RESULTS

### Design of locked TALEs

Transcription regulation can be achieved by chemical regulation of protein domain association when the TALE DNA-binding domain and the VP16 activation domain are expressed as separate proteins fused to rapamycin binding domains DmrA and DmrC (33) (Figure 1A). As an alternative, we designed a new method of transcriptional effector regulation by control of DNA binding through circular locking of the TALE DBD. The lock was designed to be inducible and reversible by attaching two copies of the rapamycin binding domain DmrD to each end of the protein via a flexible linker (final construct denoted rapaTALE, Figure 1B). DmrD has been previously described to form homodimers in the absence of its ligand, while the addition of rapamycin or its analogues disrupts dimer formation (34). A TALE designed in this way is expected to form circular structures that open in the presence of rapamycin, allowing the molecule to load onto the DNA (Figure 1B). On the other hand, the strongest type of cyclization is via formation of a covalent bond. Covalently locked TALEs were designed by the fusion of split intein fragments to the termini of the protein (denoted cycTALE, Figure 1C). We selected the Npu intein because it splices with high efficiency and speed (14). TALEs designed in this way were expected to lock before the DNA binding has occurred and the only way of topologically relaxing them to allow binding to DNA would be through the proteolytic cleavage of the linker between the termini of the folded protein domain. We therefore encoded a tobacco etch virus protease (TEVp) proteolytic cleavage site (35) (TEVs) in the linker between the transcription activation domain (VP16) and the dimerization domain (either DmrD or split intein fragment) enabling both the inducible and the covalently locked TALE to be irreversibly unlocked by the TEV protease (Figure 1B and C).

### Activation of TALEs by a chemical inducer and protease

To establish a reference point for our experiments, we first tested the activity of a noninducible TALE activator TALE-VP16 (Figure 1D and E) and a heterodimeric bimolecular



**Figure 1.** Design and function of inducible TALEs. (A) Schematic representation of the heterodimerization approach. TALE is fused to DmrA domain and VP16 is fused to the DmrC domain. The TALE domain can freely bind to the target DNA. Addition of rapamycin triggers interaction between DmrA and DmrC. (B) Schematic representation of the reversibly locked approach. rapaTALE is composed of a TALE-VP16 fused to homodimerizing DmrD domains on both ends. The locked molecule is topologically inhibited from binding to DNA. In the presence of rapamycin, DmrD domains are inhibited from dimerizing and the linear effector binds to the target DNA. The molecule is also unlocked by proteolysis with TEVp, which removes the DmrD domains from one side of the molecule. (C) Schematic representation of the covalently locked approach. cycTALE is composed of a TALE-VP16 fused to intein fragments that cyclize the molecule through splicing. The locked molecule is topologically inhibited from binding to DNA. TEVp can linearize the effector and allow DNA-binding and reporter gene expression. (D) Schematic representation of the uninducible TALE-VP16 direct fusion approach. (E) Activation of reporter gene expression with TALE-VP16. This construct functions as a positive control and is unresponsive to rapamycin or TEVp presence. The data in G, H and E are from the same experiment, representative of at least three repetitions. Error bars represent standard deviation of four biological replicates; numbers above the bars represent fold induction. (F) Activation of reporter gene expression with the heterodimerization approach as compared to the TALE-VP16 direct fusion. The bottom chart shows a magnification of the top chart. Activity of heterodimeric effectors was much lower than the activity of TALE-VP16 direct fusion. (G) Activation of reporter gene expression with L,M-rapaTALE(N). The presence of either rapamycin or TEVp upregulated reporter gene expression. (H) Activation of reporter gene expression with L,M-cycTALE(N). Only the presence of TEVp upregulated reporter gene expression. As expected, rapamycin has no effect on cycTALEs.

rapamycin inducible TALE activator (Figure 1A and F). In rapamycin-stimulated cells expressing the bimolecular activator, reporter expression exhibited a relatively high, up to 36-fold induction. However, the activity of the bimolecular activator was less than 10% of the activity of the direct TALE-VP16 fusion, even with a high excess of the plasmid for the activation domain, which is highly relevant for its application in designed biological systems. This low absolute degree of activation agrees with our results on heterodimeric activators with light induced dimerization domains, such as CRY and CIB (results not shown).

We next evaluated the transcriptional activity of the covalently and noncovalently locked TALEs. We show that reporter expression was induced after the addition of rapamycin to cells expressing the noncovalently locked activator rapaTALE (Figure 1G) while rapamycin, as expected, had no effect on cells expressing the covalently locked activator cycTALE (Figure 1H). Coexpression of TEVp increased reporter activity in both rapaTALE and cycTALE since they both contained the proteolytic cleavage site for the TEVp (Figure 1G and H). This confirmed that proteolytic cleavage can release the inhibition of a locked protein and effectively induce the transcriptional activation. The maximum activity of the TEVp-processed locked activator was comparable to the activity of the linear single-molecule fusion control (TALE-VP16) (Figure 1E) and much higher than the activity of the bimolecular activator. The activity of cycTALE was also made inducible by small molecule by expressing TEVp as a split protein (36) with rapamycin-induced complementation (37) (Supplement Figure S1).

Locked TALEs should form either a rigid circle or a circular spring like an elastic band wrapping around DNA. We expected that this might be determined by the size and flexibility of the peptide linkers between the individual protein domains, where longer or more flexible linkers might allow a locked protein to bind to DNA regardless of topological constraints (Figure 2A), and thereby increase background activity. We therefore replaced the linker between the N-terminal dimerization domain (either DmrD or intein fragment) and the TALE domain (N-terminal linker) with linkers ranging in size from 3 to 31 amino acid residues and the linker between the VP16 domain and the C-terminal IntN fragment (C-terminal linker) in cycTALEs with a longer linker of 62 amino acid residues. Notations for these linkers (S, M, L and X for the N-terminal linker and M and Y for the C-terminal linker, Figure 2B) are prefixed to the rapaTALE and cycTALE names (e.g. S,M-cycTALE). The linker M in either location corresponds to the TEVp cleavage site.

We found that the N-terminal linker length had only a minor effect on the background and induced activities of the designed locked constructs (Figure 2C and D). Some differences in background activities were observed; however the effect size was small and not consistently significant across experiments. On the other hand, we suspected that a long C-terminal linker is required to prevent the steric inhibition of VP16 (Figure 2E). To investigate this effect we tested the activity of constructs that contain only one split intein fragment. These proteins cannot cyclize or dimerize and should therefore exhibit the same transcriptional activation as the simple TALE-VP16 fusion unless VP16 is sterically

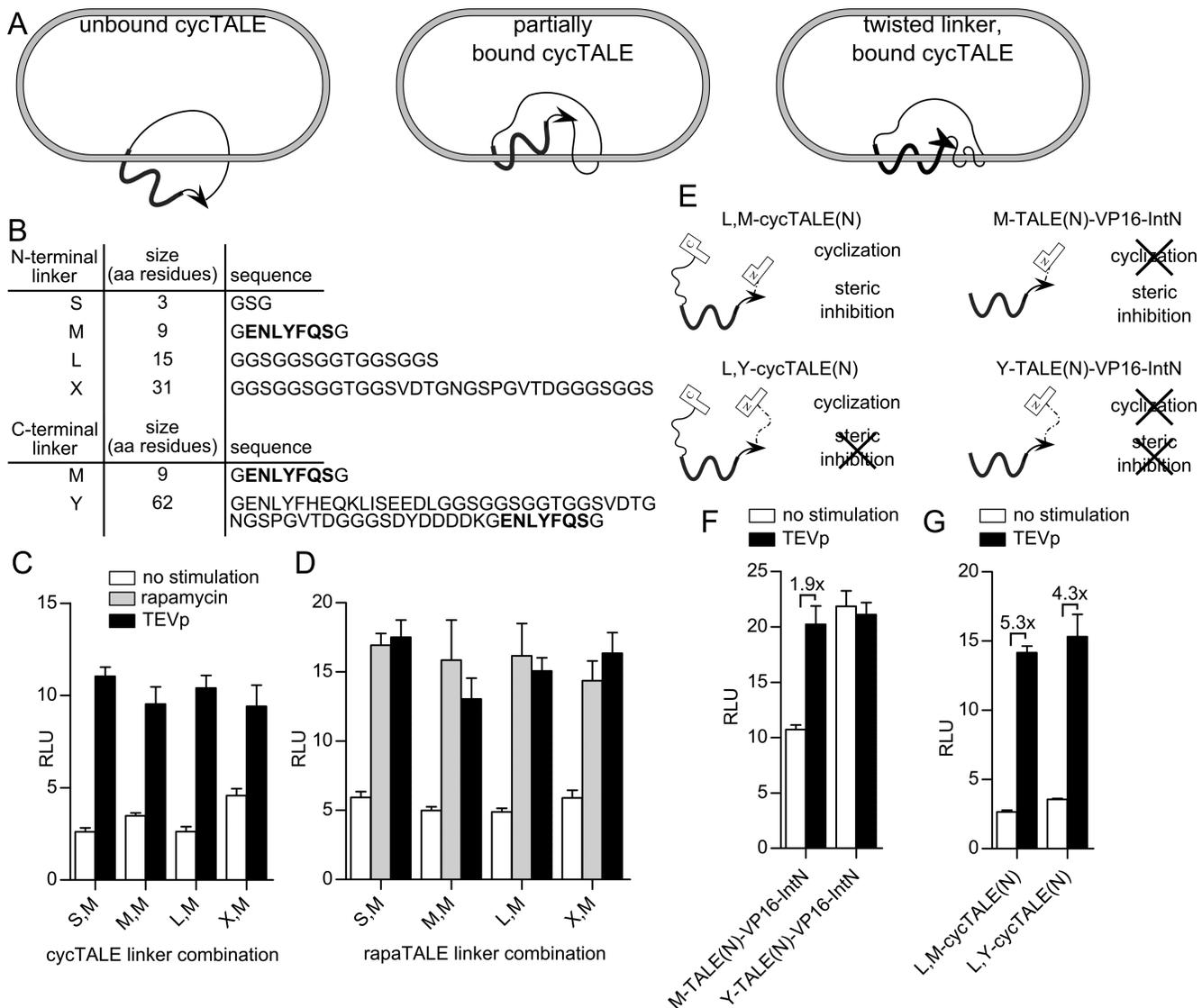
inhibited by the presence of IntN. The construct with just the IntN fragment and the short C-terminal linker M indeed displayed a diminished activity in the absence of TEVp cleavage (Figure 2F), suggesting there is some steric inhibition of the VP16 in case of a short linker. In a construct with the longer TEV cleavage site-containing linker Y this steric effect disappeared with only a slight fold induction decrease for the complete locked effector (Figure 2G). All subsequent experiments were therefore performed using the linker Y to retain only the topological locking effect.

We additionally tested several different TALE DNA-binding domains to determine the effect of the DBD size (number of RVDs) and orthogonality (Supplement Figure S2). M,Y-cycTALEs A, B and C were constructed so that they all bind to the same DNA target (binding site ABC) but have different numbers of RVDs (12, 17 and 19, respectively). The maximum activity of the proteolytically unlocked cycTALEs was comparable to the maximum activity of their respective constitutive TALE-VP16 controls (Supplement Figure S2A) and the highest activation was observed with cycTALE(C) that had the highest number of RVDs. However, only minor effects on the fold induction with cycTALEs were observed, indicating that DBD length and specificity primarily affects unlocked TALE binding and not cycTALE locking. An additional TALE DBD was constructed for a third DNA target (binding site D) and as expected, we observed no cross activation with orthogonal DBDs with binding sites ABC, D and N (Supplement Figure S2B).

### Locking TALEs impairs their DNA binding

The designed effectors were expressed in HEK293T cells and their size analyzed by western blotting. The noncovalently locked effectors had the expected size and were effectively cleaved by TEVp (Figure 3A and B). The mobility of covalently locked TALEs, however, corresponded to a dimer (Figure 3A). Coexpression with TEVp reduced the protein to the expected monomeric size. As the cyclic proteins have been consistently shown to have a higher electrophoretic mobility than linear proteins (19,20,22–24), this result suggests that our setup of intein fragments favors intermolecular dimerization with cyclization of dimers rather than monomeric intramolecular cyclization (Figure 3C). Variation in the linker length also did not affect the cycTALE dimerization or TEVp processing (Supplement Figure S3A and B), although it is unclear why some effectors were expressed at lower levels than others while retaining the same activity (Figure 2C and D).

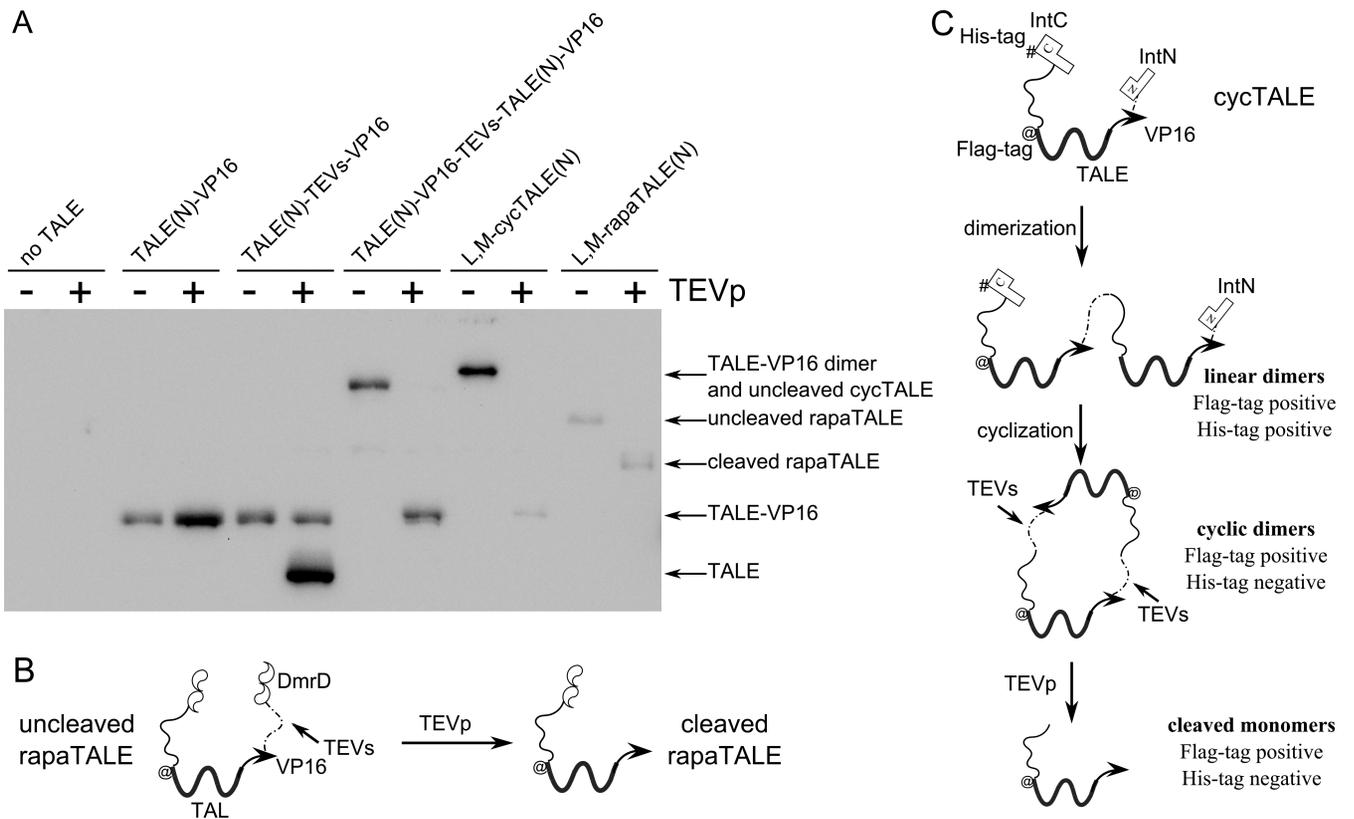
To confirm that dimerization does not interfere with the designed locking mechanism, we introduced a His<sub>6</sub> tag at the N-terminus which should be excised by cyclization, but not by linear dimerization. Indeed, these dimers were not detected with anti-His antibodies (results not shown). Moreover, the fact that reporter gene activation was inducible by TEVp while a dimeric fusion TALE-VP16-TEVs-TALE-VP16 was active even in absence of TEVp (Supplement Figure S3C) suggests that these dimers are cyclic and are topologically inhibited by the cyclization lock in spite of their dimeric state. Further evidence that intermolecular cyclization is indeed favored over intramolec-



**Figure 2.** Locked effectors with different linkers. **(A)** Schematic representation of a possible binding mechanism of locked TALEs. Left: cycTALE and plasmid DNA as overlaid but unlinked circles. Middle: cycTALE partially wraps around plasmid DNA by folding back upon itself. Right: As the TALE DBD continues wrapping around DNA, the flexible linker is also twisted around DNA. All three states are topologically equivalent to unlinked circles. **(B)** A table listing the linkers, their sizes and sequences. TEVp cleavage sites are shown in bold. **(C, D)** Reporter gene expression with cycTALEs and rapaTALEs with different N-terminal linkers. No major effects of the linkers on TALE activity were observed. Error bars represent standard deviation of four biological replicates, numbers above the bars represent fold induction. **(E)** Schematic representation of the influence of the C-terminal linker. Partial constructs with only one intein domain do not undergo splicing and cyclization, but a short linker can nevertheless impair activation through steric hindrances. Increasing the length of the C-terminal linker removes this steric inhibition while retaining locking functionality in the complete construct. **(F)** Reporter gene activation with TALE-VP16-IntN with different linkers. The short linker M allowed inhibition by IntN whereas IntN had no effect with the longer linker Y. **(G)** Reporter gene activation with locked TALEs with different C-terminal linkers. The long C-terminal linker Y decreased fold induction of the cycTALE only slightly. Error bars represent standard deviation of four biological replicates, numbers above the bars represent fold induction.

ular splicing is provided by *trans*-splicing competition. A cycTALE was coexpressed with a blue fluorescent protein (BFP) fused to an intein fragment in order to investigate competition with an intein segment on another polypeptide chain. This setup enabled formation of an intein-mediated fusion protein between the TALE and the BFP with a single intein fragment remaining on the other end, incapable of splicing into a transcriptionally inactive circular protein (Supplement Figure S3D). Indeed, the presence of a competing intein fragment increased the transcriptional activity without the need for proteolysis (Supplement Figure S3E).

Binding of the locked TALEs to the target DNA was analyzed further by a pull-down assay. A DNA-segment containing 12 copies of the TALE binding site was PCR amplified using biotin-labeled primers. This target DNA was bound to magnetic beads conjugated with streptavidin and used to pull out the DNA-binding proteins from lysates of cells expressing cycTALEs, either in the presence or in the absence of TEVp (Figure 4A). We were able to detect a strong band corresponding to the spliced monomer protein in the eluted fractions of the covalently locked TALE only in the presence of TEV protease, whereas in the absence of



**Figure 3.** Locked TALE expression, processing and cleavage by TEVp. **(A)** Expression and processing of different TALEs in the presence or absence of a TEVp coding plasmid, detected by western blot with anti-Flag antibodies. The arrows show the expected sizes of the processed and cleaved proteins. **(B)** A schematic representation of rapaTALE cleavage by TEVp. **(C)** A schematic representation of cycTALE dimerization, cyclization and cleavage. # and @ denote the His and Flag tags, respectively.

TEVp or in the controls with an orthogonal biotin-labeled DNA binding motif no band or only a fainter band was detected (Figure 4B), confirming that circularization prevents efficient binding of locked TALEs to DNA regardless of linker size or TALE specificity. Together these results suggest that topological inhibition of DNA binding due to cyclization is indeed the key mechanism that controls the function of the designed locked proteins.

### Reversibility and endogenous gene targeting by locked TALEs

The covalently locked cycTALEs essentially functioned as TEVp sensors while rapaTALEs functioned as a rapamycin sensor. As the rapamycin locking system is based on noncovalent interactions, we tested the reversibility of this system by removal and subsequent addition of rapamycin to the media of transfected cells. As expected, reporter gene expression diminished after rapamycin removal and increased again after subsequent rapamycin addition (Figure 5A), confirming that rapaTALEs act as reversible rapamycin sensors.

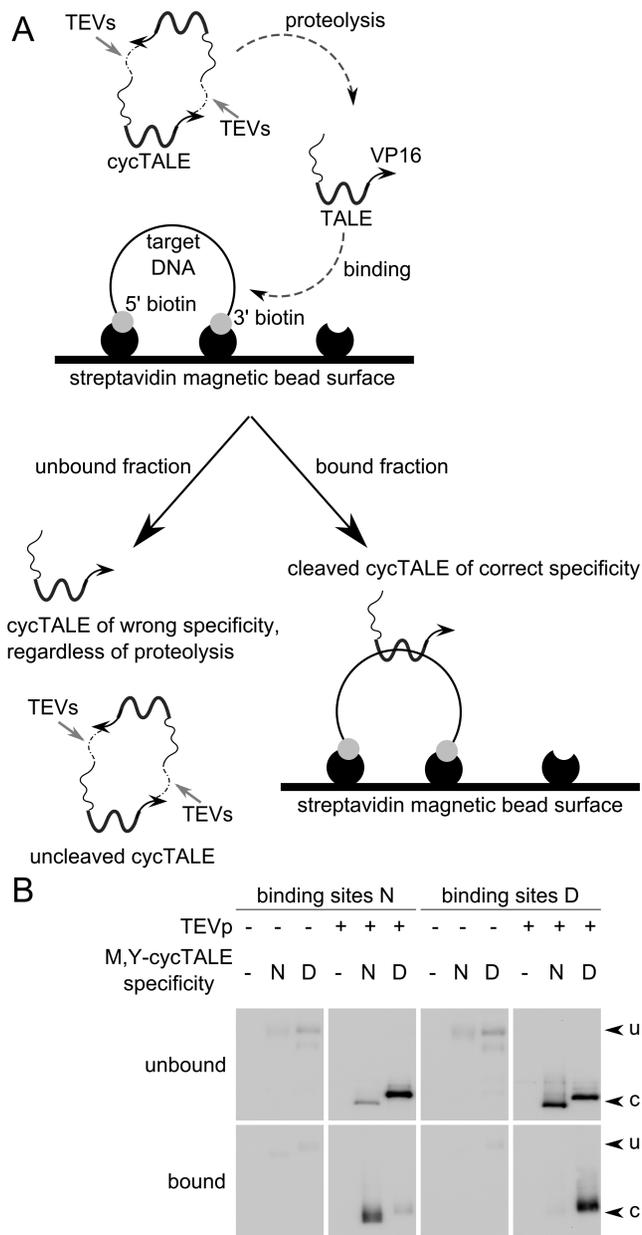
To characterize the cycTALEs and expand their potential applications, we measured activation of a reporter gene from a strong CMV promoter. Similar as described above for the minimal promoter we found that cycTALEs caused high activation also from the strong CMV promoter (Fig-

ure 5B). To extend the potential use of the described platform to genomic targets we further investigated activation of an endogenous CDH1 gene by a cycTALE. The DNA-binding domain TALE(E) that targets the promoter area of the human CDH1 gene (38) was linked to the VP64 domain to augment transcriptional activation and inserted between the split intein fragments as L,Y-cycTALE(E):VP64. This allowed us to use cleavage by the TEV protease to elicit endogenous gene activation (Figure 5C).

An additional type of a transcriptional regulator was designed by linking both VP16 and KRAB to the TALE DNA-binding domain with the TEVp recognition site between them, which also resulted in a protease-regulated transcriptional effector. The presence of the VP16 activator and the KRAB repressor in the same molecule resulted in the domination of the repressor with very low background transcription. After the proteolytic cleavage, the KRAB domain dissociated, and the TALE exhibited transcriptional activation with as much as 28-fold induction of the CMV promoter and 270-fold of the minimal promoter (Figure 5D).

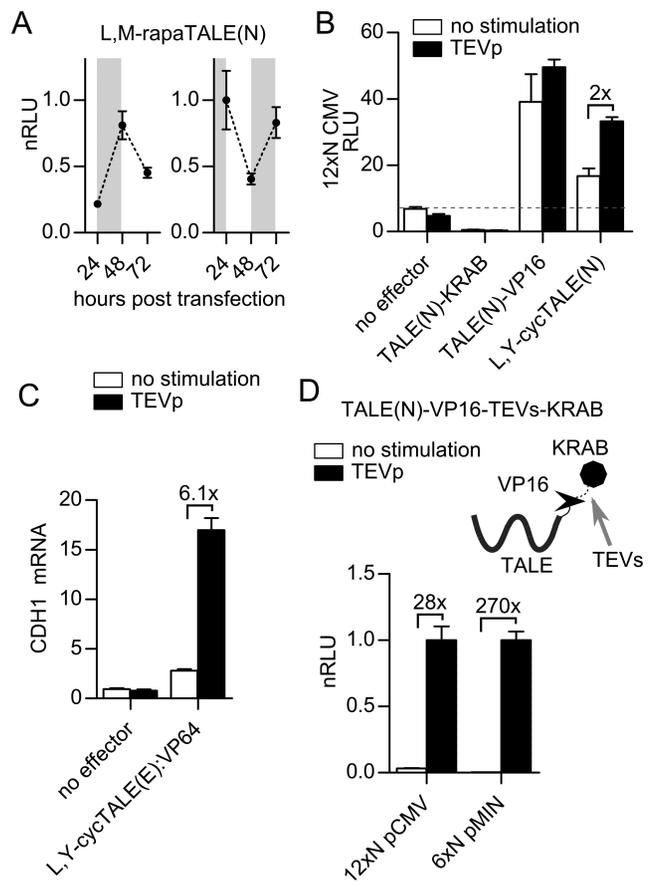
### Construction of logic gates implementing locked TALEs

The designed locked TALEs can be regarded as logic gates using rapamycin and TEVp as the two input signals. We show here the design and implementation of several two-



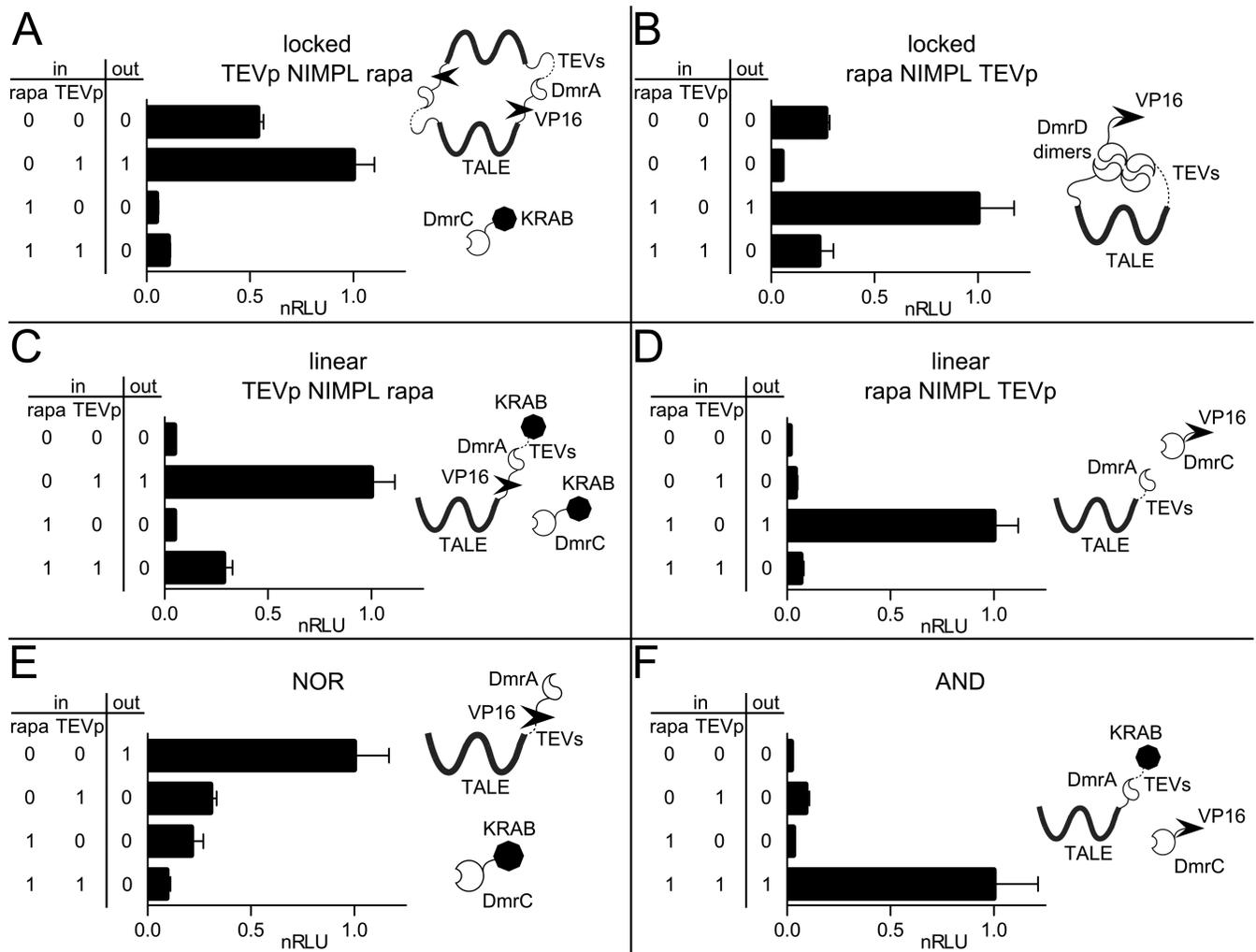
**Figure 4.** Pull-down of processed TALEs from cell lysates with biotinylated DNA bound to streptavidin magnetic beads. (A) Experimental outline. Biotinylated target DNA is bound to streptavidin beads and then mixed with lysed cells expressing TALEs in the presence or absence of TEV protease. Only unlocked TALEs of correct specificity are expected to bind. (B) Western blot of bound and unbound fractions in the presence or absence of TEVp, detected with anti-Flag antibodies. The arrows on the right show approximate expected sizes of uncleaved (u) and cleaved (c) proteins. The DNA-binding domains TALE(N) and TALE(D) showed orthogonality in binding to target DNA sites N and D.

input logic gates, including AND, NOR and nonimplication functions (Figure 6). Introduction of a DmrA domain within the covalently locked TALE activator and implementation of the KRAB domain in combination with the heterodimerization domains enabled construction of non-implication functions (Figure 6A–D) and decreased the background activity of the covalently locked effector in the



**Figure 5.** Additional characterization of locked TALEs for potential applications. (A) Reversibility of L,M-rapaTALE(N) activation upon rapamycin removal and reintroduction. Shaded areas represent rapamycin presence in the culture medium. nRLU values represent reporter gene expression under depicted circumstances normalized to reporter gene expression in constant presence of L,M-rapaTALE(N) and rapamycin. Error bars represent standard deviation of three biological replicates. (B) Reporter gene activation from the strong CMV promoter. Activity of cycTALE(N) was increased with TEVp cleavage. The dotted line represents baseline reporter gene expression from CMV promoter for comparison. Error bars represent standard deviation of four biological replicates; numbers above the bars represent fold induction. (C) Activation of endogenous CDH1 gene by TEVp-regulated cyclic TALEs. TALE(E) binding domain targets the promoter of the CDH1 gene and the bars represent relative CDH1 mRNA levels normalized to the internal GAPDH control averaged over three biological replicates. Error bars represent the minimum and maximum measured value of three biological replicates and numbers above the bars represent fold induction. (D) Schematic representation of alternative protease-inducible regulator TALE(N)-VP16-TEVs-KRAB and normalized reporter gene expression with this regulator. TEVp cleavage resulted in a high fold induction, especially with the minimal promoter. nRLU values are relative to the highest luciferase activity from the denoted promoter. Error bars represent standard deviation of three biological replicates, numbers above bars represent standard fold induction.

presence of rapamycin (Figure 6A). A linear design implementing heterodimerization and proteolysis resulted in slightly improved performance of the nonimplication functions (Figure 6C and D) due to a reduced background activity of the transcription regulators, and enabled construction of NOR and AND functions (Figure 6E and F).



**Figure 6.** Schematic representation and normalized reporter gene expression for combinations of proteolytic cleavage and rapamycin-mediated dimerization that resulted in two-input functions. (A) Cyclic variation of TEVp NIMPL rapamycin. Cleavage with TEVp unlocks the molecule, but presence of rapamycin allows the binding of a repressor domain. (B) Locked variation of rapamycin NIMPL TEVp. The presence of rapamycin unlocks the molecule, but cleavage with TEVp removes the activation domain. (C) Linear variation of TEVp NIMPL rapamycin. Cleavage with TEVp removes the repressors domain, but presence of rapamycin allows for its renewed binding. (D) Linear variation of rapamycin NIMPL TEVp. Rapamycin allows binding of DmrA to DmrC, but cleavage with TEVp removes DmrA from the DNA-binding domain. (E) NOR function. Cleavage with TEVp removes the activation domain from the protein while rapamycin triggers binding of a repressor domain. (F) AND function. Cleavage with TEVp removes the repressor domain from the protein while rapamycin triggers binding of an activation domain. nRLU are relative to the highest activity with the denoted constructs. Error bars represent standard deviation of three biological replicates, numbers above bars represent fold induction.

## DISCUSSION

Circularly locking a protein domain in order to prevent its interaction with DNA represents a new principle of regulation of protein-DNA interaction. Circular locking topologically prevents winding of the DBD around the DNA. Although we expected this mechanism to be robust, particularly when the circularization lock is based on the formation of a covalent bond, we nevertheless observed some leaky activation. The most likely explanation for this is the large size and therefore high flexibility of the observed dimeric circles or the flexibility of the linkers connecting individual protein domains, allowing the TALE DNA-binding domain to wrap itself around DNA while inducing compensatory torsion into the rest of the protein. As the N-terminal part of a TALE DBD has been shown to play an important role

in the initiation of DNA binding (39), transcriptional leakage could also be due to this partial binding to only a fraction of the DNA target motif without complete wrapping of the protein around the DNA. A low level of background DNA binding could also be due to incomplete or late locking, with a fraction of the protein molecules remaining in a linear conformation or adopting the locked conformation only after binding to DNA.

The implementation of inducible inteins might also be used to trigger the cyclization of our proteins after they have already bound to DNA, thus locking them onto the target molecule and increasing their transcriptional activation or repression. This remains to be tested, as we found the rapamycin inducible VMA intein (40,41) to have a high spontaneous *trans*-splicing activity even in the absence of rapamycin. Other advances in intein based design, such as

the recent development of light inducible inteins through introduction of nonstandard amino acid residues (42), might help circumvent this problem.

We additionally describe a combination of locking, heterodimerization and proteolysis that serves as an implementation of logic gates, where protease activity and small molecule binding are processed as inputs for AND, NOR or nonimplication functions. As the proteolysis substrate can be adjusted for other, e.g. virus-specific proteases (43) their implementations might be used to detect viral infection to trigger the appropriate response. Two- and three-input AND logic gates have been previously described based on TALE reconstitution through intein *trans*-splicing (26). The complete range of 16 two-input TALE logic gates based on multiple levels of NOR logic (27) has been described as well. However, our approach is similar to that described by Calles and de Lorenzo based on the prokaryotic XylR transcription factor (44), where disinhibition of a transcription activator is achieved either by a small molecule binding to an allosteric inhibitory domain or the cleavage of this domain by a protease. The advantage of such a Boolean processor is a single-layer system in which different functions can be implemented simply by rationally rearranging domains, and the input signals are not limited to the presence of another transgenic protein, but can easily be rerouted to other physiologically relevant signals (proteases, small molecules). Our system functions in eukaryotic cells and its DNA targeting can be fully designable by the selection of the TALE DNA-binding domain, making it useful for the targeting of endogenous genes as exemplified by *cyc*TALE(E) targeting the human CDH1 gene.

We observed that the absolute level of activation was much higher for a transcriptional activator implemented as a single molecule in contrast to the inducer-based heterodimer formation. We found that the highest difference between the inactive and active state was achieved by a design of a single molecule transcriptional regulator that comprised both a repressor and an activator domain. In this case the KRAB repressor domain was functionally dominant and resulted in a very efficient repression, while its cleavage by a protease led to a high activation, with an observed fold activation of 270 with a minimal promoter. This type of modular transcription regulators could be used to sense and to trigger a designed response to the activity of a protease with high proteolytic target specificity.

This novel approach can, in combination with other synthetic biology tools, not only expand the repertoire of implementable genetic circuits but also aid in the research of DNA-protein interactions. The specificity and thermodynamics of the interaction between TALEs and DNA has recently been the subject of molecular modeling (45,46) but is not yet fully understood. The locking system might be used to shed light on the conformational changes of both DNA and TALE during binding. The system might also be extended to other DNA-binding proteins such as polydactyl zinc finger proteins, and be used in the study and modeling of cell divisions, where naturally circular protein complexes surround DNA to control chromatid distribution (47). On the other hand covalent locking of the protein to DNA might be used to prevent replication by inhibiting separa-

tion of strands or for other interesting molecular biological or nanotechnological applications.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors are grateful to Duško Lainšček for the help with qPCR and Dr Rok Gaber and Dr Anže Smole for many insightful discussions. J.L. and T.L. are PhD students at the Graduate School of Biomedicine at the University of Ljubljana.

## FUNDING

Slovenian Research Agency [P4–0176, J1–6740]; EN-FIST Centre of Excellence. Funding for open access charge: Slovenian Research Agency.

*Conflict of interest statement.* None declared.

## REFERENCES

- Liu, Q., Segal, D.J., Ghiara, J.B. and Barbas, C.F. (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 5525–5530.
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
- Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
- Deng, D., Yan, C., Wu, J., Pan, X. and Yan, N. (2014) Revisiting the TALE repeat. *Protein Cell*, **5**, 297–306.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., Shi, Y. and Yan, N. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, **335**, 720–723.
- Mak, A.N.S., Bradley, P., Cernadas, R.A., Bogdanove, A.J. and Stoddard, B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, **335**, 716–719.
- Gossen, M. and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 5547–5551.
- Li, Y., Moore, R., Guinn, M. and Bleris, L. (2012) Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci. Rep.*, **2**, 897.
- Pollock, R., Giel, M., Linher, K. and Clackson, T. (2002) Regulation of endogenous gene expression with a small-molecule dimerizer. *Nat. Biotechnol.*, **20**, 729–733.
- Zetsche, B., Volz, S.E. and Zhang, F. (2015) A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.*, **33**, 139–142.
- Li, Y. (2015) Split-inteins and their bioapplications. *Biotechnol. Lett.*, **37**, 2121–2137.
- Evans, T.C., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X.Q. and Xu, M.Q. (2000) Protein trans-splicing and cyclization by a naturally split intein from the *dnaE* gene of *Synechocystis* species PCC6803. *J. Biol. Chem.*, **275**, 9091–9094.
- Zettler, J., Schütz, V. and Mootz, H.D. (2009) The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett.*, **583**, 909–914.
- Sun, W., Yang, J. and Liu, X.Q. (2004) Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J. Biol. Chem.*, **279**, 35281–35286.

16. Mootz, H.D., Blum, E.S., Tyszkiewicz, A.B. and Muir, T.W. (2003) Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo. *J. Am. Chem. Soc.*, **125**, 10561–10569.
17. Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A. *et al.* (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, **192**, 271–281.
18. Ramirez, M., Valdes, N., Guan, D. and Chen, Z. (2013) Engineering split intein DnaE from *Nostoc punctiforme* for rapid protein purification. *Protein Eng. Des. Sel.*, **26**, 215–223.
19. Zhao, Z., Ma, X., Li, L., Zhang, W., Ping, S., Xu, M.Q. and Lin, M. (2010) Protein cyclization enhanced thermostability and exopeptidase-resistance of green fluorescent protein. *J. Microbiol. Biotechnol.*, **20**, 460–466.
20. Volkmann, G., Murphy, P.W., Rowland, E.E., Cronan, J.E., Liu, X.Q., Blouin, C. and Byers, D.M. (2010) Intein-mediated cyclization of bacterial acyl carrier protein stabilizes its folded conformation but does not abolish function. *J. Biol. Chem.*, **285**, 8605–8614.
21. Williams, N.K., Liepinsh, E., Watt, S.J., Prosser, P., Matthews, J.M., Attard, P., Beck, J.L., Dixon, N.E. and Otting, G. (2005) Stabilization of native protein fold by intein-mediated covalent cyclization. *J. Mol. Biol.*, **346**, 1095–1108.
22. Jeffries, C.M., Graham, S.C., Stokes, P.H., Collyer, C.A., Guss, J.M. and Matthews, J.M. (2006) Stabilization of a binary protein complex by intein-mediated cyclization. *Protein Sci.*, **15**, 2612–2618.
23. Iwai, H., Lingel, A. and Pluckthun, A. (2001) Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfuI intein from *Pyrococcus furiosus*. *J. Biol. Chem.*, **276**, 16548–16554.
24. Siebold, C. and Erni, B. (2002) Intein-mediated cyclization of a soluble and a membrane protein in vivo: function and stability. *Biophys. Chem.*, **96**, 163–171.
25. Lohmueller, J.J., Armel, T.Z. and Silver, P.A. (2012) A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res.*, **40**, 5180–5187.
26. Lienert, F., Torella, J.P., Chen, J.H., Norsworthy, M., Richardson, R.R. and Silver, P.A. (2013) Two- and three-input TALE-based AND logic computation in embryonic stem cells. *Nucleic Acids Res.*, **41**, 9967–9975.
27. Gaber, R., Lebar, T., Majerle, A., Šter, B., Dobnikar, A., Benčina, M. and Jerala, R. (2014) Designable DNA-binding domains enable construction of logic circuits in mammalian cells. *Nat. Chem. Biol.*, **10**, 203–208.
28. Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K. and Yeh, J.R. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.*, **29**, 697–698.
29. Iwai, H., Züger, S., Jin, J. and Tam, P.H. (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett.*, **580**, 1853–1858.
30. Szulc, J., Wiznerowicz, M., Sauvain, M.O., Trono, D. and Aebischer, P. (2006) A versatile tool for conditional gene expression and knockdown. *Nat. Methods*, **3**, 109–116.
31. Sadowski, I., Ma, J., Triezenberg, S. and Ptashne, M. (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature*, **335**, 563–564.
32. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.
33. Rivera, V.M., Clackson, T., Natesan, S., Pollock, R., Amara, J.F., Keenan, T., Magari, S.R., Phillips, T., Courage, N.L., Cerasoli, F. *et al.* (1996) A humanized system for pharmacologic control of gene expression. *Nat. Med.*, **2**, 1028–1032.
34. Rollins, C.T., Rivera, V.M., Woolfson, D.N., Keenan, T., Hatada, M., Adams, S.E., Andrade, L.J., Yaeger, D., van Schravendijk, M.R., Holt, D.A. *et al.* (2000) A ligand-reversible dimerization system for controlling protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 7096–7101.
35. Carrington, J.C. and Dougherty, W.G. (1988) A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 3391–3395.
36. Wehr, M.C., Laage, R., Bolz, U., Fischer, T.M., Grünewald, S., Scheek, S., Bach, A., Nave, K.A. and Rossmner, M.J. (2006) Monitoring regulated protein-protein interactions using split TEV. *Nat. Methods*, **3**, 985–993.
37. Williams, D.J., Puhl, H.L. and Ikeda, S.R. (2009) Rapid modification of proteins using a rapamycin-inducible tobacco etch virus protease system. *PLoS One*, **4**, e7474.
38. Hu, J., Lei, Y., Wong, W.K., Liu, S., Lee, K.C., He, X., You, W., Zhou, R., Guo, J.T., Chen, X. *et al.* (2014) Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res.*, **42**, 4375–4390.
39. Meckler, J.F., Bhakta, M.S., Kim, M.S., Ovadia, R., Habrian, C.H., Zykovich, A., Yu, A., Lockwood, S.H., Morbitzer, R., Elsässer, J. *et al.* (2013) Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res.*, **41**, 4118–4128.
40. Brenzel, S., Kurpiers, T. and Mootz, H.D. (2006) Engineering artificially split inteins for applications in protein chemistry: biochemical characterization of the split Ssp DnaB intein and comparison to the Sce VMA intein. *Biochemistry*, **45**, 1571–1578.
41. Selgrade, D.F., Lohmueller, J.J., Lienert, F. and Silver, P.A. (2013) Protein scaffold-activated protein trans-splicing in mammalian cells. *J. Am. Chem. Soc.*, **135**, 7713–7719.
42. Binschik, J., Zettler, J. and Mootz, H.D. (2011) Photocontrol of protein activity mediated by the cleavage reaction of a split intein. *Angew. Chem. Int. Ed. Engl.*, **50**, 3249–3252.
43. Majerle, A., Gaber, R., Benčina, M. and Jerala, R. (2015) Function-based mutation-resistant synthetic signaling device activated by HIV-1 proteolysis. *ACS Synth. Biol.*, **4**, 667–672.
44. Calles, B. and de Lorenzo, V. (2013) Expanding the boolean logic of the prokaryotic transcription factor XylR by functionalization of permissive sites with a protease-target sequence. *ACS Synth. Biol.*, **2**, 594–603.
45. Lei, H., Sun, J., Baldwin, E.P., Segal, D.J. and Duan, Y. (2014) Conformational elasticity can facilitate TALE-DNA recognition. *Adv. Protein Chem. Struct. Biol.*, **94**, 347–364.
46. Wan, H., Hu, J.P., Li, K.S., Tian, X.H. and Chang, S. (2013) Molecular dynamics simulations of DNA-free and DNA-bound TAL effectors. *PLoS One*, **8**, e76045.
47. Murayama, Y. and Uhlmann, F. (2014) Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature*, **505**, 367–371.