In vivo diversification of target genomic sites using processive T7 RNA polymerase-base deaminase fusions blocked by RNA-guided dCas9

Beatriz Álvarez[1], Mario Mencía[2], Víctor de Lorenzo[3] and Luis Ángel Fernández[1]*

(1) *Department of Microbial Biotechnology* and (3) *Systems Biology Program, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CNB-CSIC), Darwin 3, Campus UAM Cantoblanco, 28049 Madrid, Spain.*

(2) Centro de Biología Molecular "Severo Ochoa" (Consejo Superior de Investigaciones Científicas – Universidad Autónoma de Madrid), Nicolas Cabrera 1, Campus UAM Cantoblanco, 28049 Madrid, Spain

Running Title: Processive base deaminase fusions blocked by dCas9

Keywords: base deaminases, dCas9, directed evolution, in vivo mutagenesis, T7 RNA polymerase

---

* **Corresponding author**: Dr. Luis Ángel Fernández

E-mail: lafdez@cnb.csic.es

Phone: +34 91 585 4854

Fax: +34 91 585 4506

Abstract

Diversification of specific DNA segments typically involve *in vitro* generation of large sequence libraries and their introduction in cells for selection. Alternative *in vivo* mutagenesis systems on cells often show deleterious offsite mutations and restricted capabilities. To overcome these limitations, we have developed an *in vivo* platform to diversify specific DNA segments based on protein fusions between various base deaminases (BD) and the T7 RNA polymerase (T7RNAP) that recognizes a cognate promoter oriented towards the target sequence. The transcriptional elongation of these fusions generates transitions C to T or A to G on both DNA strands and in long DNA segments. To delimit the boundaries of the diversified DNA, the catalytically dead Cas9 (dCas9) is tethered with custom-designed crRNAs as a “roadblock” for BD-T7RNAP elongation. While the efficiency of this platform is demonstrated in *E. coli,* the system can be adapted to a variety of bacterial and eukaryotic hosts.

Introduction

Directed evolution enables the selection of protein variants with improved properties as therapeutics and biocatalysts [1, 2, 3]. The generation of genetic variability followed by a screening process are the essential steps of directed evolution [4, 5]. *In vitro* mutagenesis techniques (e.g., error-prone PCR) can quickly produce large number of variants of the target gene but their selection requires, in most cases, cloning and transformation into a host cell for expression (e.g., *E. coli*). These steps are time-consuming and labor-intensive, especially when iterative cycles of mutagenesis and selection are needed. Cell-free selection methods are also feasible [6], but they are technically demanding and functional expression of complex proteins (e.g. membrane proteins, multimeric enzymes) is often difficult to achieve. Hence, *in vivo* mutagenesis systems are preferred because the generation of genetic variants, their expression and selection can done in a continuous process, which accellerates directed evolution [5, 7].

Long-established *in vivo* mutagenesis methods (e.g., chemical mutagens, radiations) do not target specific genes and are deleterious for the host cell due to accumulation of random mutations in the genome [8]. Similarly, "mutator" host strains do not allow the concentration of the mutagenic activity on the target gene, inducing the accumulation of unwanted mutations in the host genome [9, 10, 11, 12]. A few *in vivo* mutagenesis systems with targeted specificity have been reported for *E. coli* and yeast cells, which are the preferred hosts for cloning and expression of gene libraries [5]. For example, an error-prone variant of *E. coli* DNA polymersase I enables the mutagenesis of cloned genes in ColE1 plasmids, although it concentrates the mutations close to the origin of replication [13]. A different approach involves the transformation of *E. coli* with mutant oligonucleotide libraries, targeting one or multiple loci in the chromosome, which induce mutation during DNA replication *in vivo* [14, 15, 16]. These systems allow multiloci genome engineering, but also imply iterative cycles of transformation with oligonucleotide libraries followed by high-throughput screenings, often including massive DNA sequencing steps, which are labor-intensive and demand sophisticated equipment. In yeast, generation variability can be achieved by cloning the target DNA segments in retrotransposons having an error-prone

retrotranscriptase [17], but this process is limited to the DNA size that can be cloned in the retrotransposon (<5 kb).

A more versatile mutagenic system for different hosts is based on the tethering of base deaminases (BDs), such as cytosine deaminasaes (CDs) and adenosine deaminases (ADs), to a target DNA using catalytically dead Cas9 (dCas9) or analogous enzymes targeted with a CRISPR RNA (crRNA) or guide RNA (gRNA) [18, 19, 20, 21]. Previous studies had shown that expression of CDs, such as human AID and orthologs from rat (rAPOBEC1) and lamprey (pmCDA1), induce random C to T mutations *in vivo*, both in *E. coli* and yeast. These CDs increase the frequency of C:G to T:A base pair transitions in DNA, especially when uracil DNA N-glycosilase (UNG) activity is inactived by gene deletion or by the specific inhibitor UGI [22, 23, 24, 25]. This is because cytosine deamination produces uracil in DNA that can be eliminated by UNG, generating an abasic DNA that is a substrate of the base-excision repair system [26, 27]. When a CD is fused to dCas9, its mutagenic activity is tethered to the target DNA sequence hybridized by the crRNA (or gRNA) allowing edition of specific bases in the genome [18, 19, 20, 21]. In addition, mutations of A to G (inducing A:T to G:C base pairs transitions) have been generated by fusing dCas9 to an engineered AD named TadA*, derived from the endogenous RNA-dependent AD TadA of *E. coli* [28]. Fusions of these BDs to dCas9 provides precise molecular tools for edition of specific bases in the genome, but its lack of processivity limits its potential for directed evolution of complete genes and operons. An interesenting alternative is the use of a nickase Cas9 (nCas9) fused to an error-prone DNA polymerase that is able to introduce mutations in a DNA segment of up to 350 bp [29], which is still limited for long genes and operons.

Hence, despite the above-mentioned advancements, the development of *in vivo* mutagenesis systems with high specificity and processivity are of great interest. In this study we report a novel approach that fulfill these criteria enabling *in vivo* mutagenesis of a target gene using different highly processive protein fusions of BDs and the bacteriophage T7 RNA polymerase (T7RNAP). The specificity of the mutagenesis is provided by a distinct T7RNAP promoter (P_{T7}) driving transcription through the target DNA. By placing the P_{T7} at the 3'-end of the target gene, in

115 reverse orientation, expression of the target gene can be preserved from its endogenous 5'-
116 promoter recognized by the host RNAP. We report the mutagenic action of fusions of T7RNAP
117 with different BDs (i.e., AID, rAPOBEC1, pmCDA1 and TadA*) on a target genomic DNA
118 segment and show that the DNA-bound crRNA/dCas9 complex hinders elongation BD-T7RNAP
119 hybrids, protecting the downstream DNA. Given the demonstrated functionality of BDs, T7RNAP
120 and dCas9 in different hosts (e.g., bacteria, yeast, mammalian cells) [18, 19, 21, 30, 31, 32], this system
121 has the potential to be implemented in diverse organisms other than *E. coli*.
122
123 ## Results
124 *An E. coli reporter strain of the mutagenic and transcriptional activity of BD-T7RNAP fusions*. A
125 scheme of the overall strategy followed in this study is shown in Fig. 1. To measure both the
126 mutagenic and transcriptional activity of BD-T7RNAP fusions, we designed a GFP-URA3
127 genetic cassette comprising two gene reporters in reverse orientation: a promoter-less *gfp* gene
128 and the URA3 gene from *Saccharomyces cerevisiae* (Fig 2a). Transcription of the URA3 gene
129 was placed under the Ptac promoter recognized by *E. coli* RNAP. Yeast URA3 encodes the
130 enzyme orotidine 5'-phosphate decarboxylase involved in the synthesis of uridine
131 monophosphate (UMP) [33]. The activity of URA3, and that of the *E. coli* orthologue *pyrF*, allows
132 cell growth in the absence of uracil in the medium (positive selection). In addition, URA3
133 expression makes yeast and *E. coli* cells sensitive to 5'-Fluororotic acid (FOA) allowing selection
134 of null mutants (negative selection or counterselection) [34, 35]. To enable specific recruitment of
135 BD-T7RNAP fusions, the promoter P_{T7} was placed downstream of URA3, in reversed orientation
136 to the coding sequence of URA3, but in the same orientation that the promoter-less *gfp* gene
137 (Fig. 2a). Thus, expression of GFP acts as a reporter of the transcriptional activity of BD-
138 T7RNAP fusions. The GFP-URA3 cassette was flanked by transcriptional terminators (T1 and
139 T0). The whole genetic construct was cloned in an integrative suicide vector carrying an
140 apramycin-resistance marker ($Apra^R$) and flanking homology regions of the *flu* gene of *E. coli* K-
141 12, encoding Antigen 43 [36]. Integration of genetic constructs replacing *flu* does not affect
142 bacterial growth and viability [37, 38]. The *E. coli* K-12 strain used for integration was derived from
143 the reference strain MG1655 [39, 40] by correcting using recombineering a natural mutation that

reduces the expression of *pyrE* [41] (Supplementary Fig. 1a). The *pyrE* gene encodes the enzyme orotate phosphoribosyltransferase required for the biosynthesis of UMP and the incorporation of FOA to produce the toxic 5-FUMP. The strain with the corrected *pyrE* allele (named MG1655*) showed higher sensitivity to FOA than the parental MG1655 strain (Supplementary Fig 1b). Deletion of *pyrF* in MG1655* makes bacteria resistant to FOA (Supplementary Fig. 1b). The GFP-URA3 cassette was integrated in the chromosome of MG1655*Δ*pyrF* generating the reporter strain MG*-URA3, which grows well in mineral media (M9) lacking uracil and is highly sensitive to FOA (Supplementary Fig. 1b). Lastly, an *ung* deletion mutant was obtained in this strain (MG*-URA3Δ*ung*).

Expression and activity of T7RNAP fusions to human AID. We first investigated the tolerance of T7RNAP (~99 kDa) for N- and C-terminal protein fusions using the human CD AID (~24 kDa) [22, 24]. AID was fused at the N- and C-terminal ends of T7RNAP using a flexible peptide linker of Gly and Ser (G3S)7, generating AID-T7RNAP and T7RNAP-AID fusions, respectively (Fig. 2b). In addition, variants of AID-T7RNAP were constructed with different N-terminal tags: thioredoxin 1 (TrxA; ~11 kDa) and a cytosolic version of the maltose binding protein (MBPc; ~40 kDa), generating TrxA-AID-T7RNAP and MBPc-AID-T7RNAP, respectively (Fig. 2b). N-terminal TrxA and MBPc are reported to increase the solubility and expression level of protein fusions in *E. coli* [42, 43, 44, 45]. Gene constructs encoding T7RNAP and AID fusions were placed under the control of the tetracycline-regulated promoter (TetR-PtetA) [46, 47] in a low copy-number plasmid (pSEVA221) [48]. MG*-URA3Δ*ung* bacteria harboring these plasmids, and pSEVA221 as negative control, were grown at 37 °C in LB for 2 h (OD600 ~1.0), and induced with anhydrotetracycline (aTc) for additional 1 h. All cultures grew to a similar final optical density (OD600 ~2.5), except bacteria expressing native T7RNAP (OD600 ~1.2) suggesting some toxicity of T7RNAP expression. Whole-cell protein extracts from these cultures were analyzed by Western blot with a monoclonal antibody (mAb) anti-T7RNAP (Fig. 2c). Protein bands corresponding to the expected size of full-length fusions and T7RNAP were detected in bacteria expressing all N-terminal AID fusions, albeit higher levels were found with MBPc-AID-T7RNAP. In the case of the C-terminal fusion T7RNAP-AID, multiple protein bands were visible corresponding to

T7RNAP and truncated polypeptides derived from the fusion, but none corresponding to the full-length polypeptide (Fig. 2c). GFP expression was detected by flow cytometry in bacteria with native T7RNAP and all N-terminal AID fusions, but not in bacteria expressing the C-terminal fusion of AID or carrying the empty vector (Fig. 2d). GFP fluorescence was ca. 2-fold higher in bacteria expressing native T7RNAP than in bacteria expressing any of the N-terminal AID fusions (Fig. 2d). Therefore, N-terminal fusions to T7RNAP produce a transcriptionally active polypeptide in *E. coli*, whereas C-terminal fusions are not stable and transcriptionally inactive.

To test the mutagenic activity of the transcriptionally active fusions, cultures of MG*-URA3Δ*ung* bacteria carrying plasmid constructs encoding none (empty vector control), AID-T7RNAP, TrxA-AID-T7RNAP, and MBPc-AID-T7RNAP, were grown and induced as above. After induction, bacteria were plated on M9+uracil and M9+uracil+FOA to determine colony forming units (CFU/ml) in each media. The mutation frequency of URA3 was determined for each bacterial strain in three independent experiments as the ratio of FOA^R CFU/ml *vs*. total CFU/ml (Fig. 2e). These data revealed a significant increase (~1000-fold) in the frequency of FOA^R bacteria (URA3 mutants) in cultures expressing AID fusions (~10^{-3}) compared to negative control bacteria (~10^{-6}). Interestingly, expression of native T7RNAP alone increased ~20-fold (~$2x10^{-5}$) the frequency of URA3 mutants, suggesting some mutagenic activity caused by high-level transcription of URA3 and/or by the toxicity of T7RNAP overexpression, which also caused a ~10-fold reduction in the total CFU/ml (Fig. 2f). In contrast, bacterial cultures expressing AID fusions did not show any significant change in total CFU/ml, with values similar to the control strain (~10^9 CFU/ml) (Fig. 2f). Hence, the high-frequency of URA3 mutants found in bacteria with AID fusions strongly suggests a mutagenic activity of the fusion polypeptides.

To evaluate the specificity of the mutagenic AID-T7RNAP fusions, bacteria from the above cultures were also plated on rifampicin (Rif)-containing plates. Rif^R colonies of *E. coli* are known to contain mutations in *rpoB*, encoding the β-subunit of *E. coli* RNAP [49, 50]. The frequency of mutation in *rpoB* was determined for each culture as the ratio of Rif^R CFU/ml vs. total CFU/ml (Fig. 2g). These data revealed a ~5-fold increase in the frequency of *rpoB* mutants in bacteria

expressing AID fusions (~10^{-6}) compared to the negative control or bacteria expressing native T7RNAP (~$2x10^{-7}$) (Fig. 2g). These data are in accordance with previous work showing that expression of AID slightly increases the mutagenesis of non-specific loci in *E. coli* Δ*ung* strains [24]. However, this low non-specific mutagenic activity of AID-T7RNAP fusions is clearly insufficient to explain the ~1000-fold increase in URA3 mutants, suggesting a strong specificity of AID-T7RNAP fusions for the mutagenesis of URA3.

Mutagenic activity of different CDs fused to the N-terminus of T7RNAP. Since all fusions having AID at the N-terminal of T7RNAP showed similar transcriptional and mutagenic activities (Fig. 2), we chose the AID-T7RNAP fusion, lacking any additional protein partner, to continue our work. We constructed similar N-terminal fusions with other CDs, namely pmCDA1 and rAPOBEC1 (Fig. 3a). As for AID, pmCDA1 and rAPOBEC1 fusions were also cloned in pSEVA221 under the control of TetR-PtetA. MG*-URA3Δ*ung* bacteria carrying pSEVA221 (negative control) or plasmids encoding T7RNAP, AID-T7RNAP, pmCDA1-T7RNAP, and rAPOBEC1-T7RNAP, were induced with aTc. Western blot analysis of whole-cell protein extracts from these cultures revealed similar expression levels of the fusion proteins, and higher expression of native T7RNAP (Fig. 3b), as before. Flow cytometry analysis also showed similar levels of GFP expression in bacteria encoding AID, pmCDA1 and rAPOBEC1 fusions, with MFI values roughly half of those found in bacteria expressing native T7RNAP (Fig. 3c). Hence, as for AID-T7RNAP, fusions pmCDA1-T7RNAP and rAPOBEC1-T7RNAP were transcriptionally active in *E. coli*.

Next, we determined the frequency of mutagenesis in URA3 and *rpoB* loci upon induction of bacterial cultures carrying pSEVA221 (empty vector) or derivatives encoding native T7RNAP, AID-, pmCDA1- and rAPOBEC1- fusions, in three different *E. coli* strains: MG*-URA3 (*ung*$^+$), MG*-URA3Δ*ung*, and MG*-URA3Δ*ung*ΔP_{T7} (lacking the P_{T7} in URA3). The ratio of FOA^R CFU/ml and Rif^R CFU/ml *vs*. total CFU/ml for each induced culture was determined in three independent experiments (Fig. 3d). This analysis showed a significative increase in the mutagenesis of the URA3 locus for all CD fusions in Δ*ung* bacteria and in the presence of P_{T7}. As expected from

previous data (Fig. 2), the frequency of URA3 mutants in MG*-URA3Δ*ung* bacteria was ~10^{-3} for AID-T7RNAP, ~10^{-5} for native T7RNAP, and ~10^{-6} for the empty vector, but increased dramatically to ~5x10^{-2} for rAPOBEC1-T7RNAP and to ~10^{-1} for pmCDA1-T7RNAP. Thus, rAPOBEC1 and pmCDA1 fusions have a higher mutagenic activity than AID fusion (~50- to 100-fold, respectively). The higher activity of rAPOBEC1 and pmCDA1 was also partially reflected in a slighlty higher "off-target" mutagenesis in *rpoB* (Fig. 3d). Importantly, the frequency of URA3 mutants for all CD fusions dropped to levels close to those of the empty vector in the strain lacking the T7 promoter (MG*-URA3Δ*ung*ΔP_{T7} strain). Hence, these data clearly indicate that the strong mutagenic activity of CD-T7RNAP fusions in URA3 requires the presence of P_{T7}. The requirement of P_{T7} is also seen for the mutagenic activity of native T7RNAP in URA3 (Fig. 3d). Lastly, we found that all CD fusions have a lower mutagenic activity in the *ung*$^{+}$ strain whereas the mutagenic activity of native T7RNAP is independent of the presence of UNG (Fig. 3d). Not surprisingly, the highly active pmCDA1 fusion also showed the highest mutagenic frequency of URA3 in the *ung*$^{+}$ strain, with a 100-fold increase compared to the control with the empty vector (frequency ~10^{-4} vs. ~10^{-6}). A moderate increase of ~5 to 10-fold in the frequency of URA3 mutants was found in the *ung*$^{+}$ strain for AID and rAPOBEC1 fusions, respectively (Fig. 3d).

Mutagenic activity of an adenosine deaminase (AD) fusion to T7RNAP. To broaden the capacity of the mutagenesis system to adenosine bases, a T7RNAP fusion was constructed with the modified AD TadA7.10 (TadA*) [28] (Fig. 4a). TadA* was shown to deaminate adenines in DNA generating inosines that leads to A:T > G:C transitions. The fusion TadA*-T7RNAP was expressed in MG*-URA3Δ*ung* at higher levels than AID-T7RNAP, as determined by Western-blot (Fig. 4b). The TadA*-T7RNAP fusion was also transcriptionally active, producing slightly higher levels of GFP than AID-T7RNAP (Fig. 4c). The potential mutagenic capacity of TadA*-T7RNAP was evaluated in different genetic backgrounds, using AID-T7RNAP as a positive mutagenesis control and pSEVA221 vector as a negative control (Fig. 4d). These experiments revealed that expression of TadA*-T7RNAP generates URA3 mutants with a frequency of ~2-5x10^{-4} (~100-fold higher than the negative control) in both MG*-URA3(*ung*$^{+}$) and MG*-URA3Δ*ung* strains, indicating that TadA* fusion is mutagenic and, contrary to AID-T7RNAP,

independent of the presence of the enzyme UNG (Fig. 4d). In *E. coli* K-12 the gene *nfi* encodes endonuclease V, which is involved in inosine elimination [51, 52]. When TadA*-T7RNAP was expressed in *E. coli* strains lacking *nfi* (MG*-URA3Δ*nfi* and MG*-URA3Δ*ung*Δ*nfi*), the frequency of URA3 mutants further increased to ~10^{-3}, a level similar to that generated by AID-T7RNAP in the Δ*ung* mutant (Fig. 4d). In contrast, deletion of *nfi* had no effect over the mutagenesis frequency of AID-T7RNAP (Fig. 4d). In addition, expression of TadA*-T7RNAP did not produce any significant increase in the levels of off-target mutagenesis in *rpoB* in these strains compared to the negative control (pSEVA221) (Fig. 4d). Lastly, we confirmed that the mutagenic activity of TadA*-T7RNAP in URA3 was dependent on the presence of P_{T7} promoter since its mutagenesis frequency dropped to the baseline levels in the strain MG*-URA3Δ*ung*ΔP_{T7} (Fig. 4e). Altogether these data demonstrate that TadA*-T7RNAP fusion has a specific mutagenic activity for the target DNA having P_{T7}. This mutagenic activity is independent of UNG, being moderately increased 2- to 5-fold when endonuclease V is absent.

Characterization of the mutations produced by BD-T7RNAP fusions. We randomly picked 30 FOA^R colonies (URA3 mutants) from the MG*-URA3Δ*ung* strains expressing native T7RNAP, AID-, pmCDA1- and rAPOBEC1- fusions to T7RNAP, and 30 additional FOA^R colonies grown from MG*-URA3Δ*ung*Δ*nfi* strain expressing TadA*-T7RNAP. The URA3 alleles from these colonies were amplified as PCR fragments including the 5' Ptac promoter, the complete URA3 gene and downstream P_{T7}, and their DNA sequence was determined. As an additional control, the URA3 alleles from 13 FOA-sensitive colonies from MG*-URA3Δ*ung* strain with plasmid pSEVA221 were amplified in parallel and sequenced, which did not show any mutation from the wild type sequence of URA3 allele. In contrast, in all FOA^R colonies analyzed from strains expressing CD-T7RNAP fusions we found multiple transitions C:G to T:A in both DNA strands along the Ptac and URA3 gene, but not in the P_{T7} sequence (Fig. 5a). In the case of FOA^R colonies from MG*-URA3Δ*ung*Δ*nfi* strain with TadA*-T7RNAP, all alleles contained transitions A:T to G:C in both DNA strands of the URA3 gene, except one different mutation, a C:G to T:A transition (Fig. 5b). No other type of DNA mutations, deletions, or insertions, were observed in any of the URA3 alleles of FOA^R colonies analyzed expressing BD-T7RNAP fusions. This was

not the case in FOA^R colonies derived from the expression of native T7RNAP, which contained URA3 alleles with more types of mutations, including transitions (A:T to G:C and C:G to T:A), transversions (A:T to C:G and A:T to T:A), deletions, and insertions (Supplementary Fig. 2). Showing correlation to the mutagenic capacity of the three CD-T7RNAP fusions, the highest total number of mutations was found with pmCDA1-T7RNAP (426) followed by rAPOBEC1-T7RNAP (95) and AID-T7RNAP (42) (Fig 5c). The total number of mutations found in the 30 URA3 alleles from TadA*-T7RNAP was 37, similar to that of AID-T7RNAP (Fig. 5c). The average number of mutations per clone presented the same hierarchy: pmCDA1-T7RNAP (14.2) > rAPOBEC1-T7RNAP (3.2) > AID-T7RNAP (1.4) > TadA*-T7RNAP (1.2) (Fig. 5d). It is worth noting that for all CD-T7RNAP fusions, transitions G to A were detected more frequently than C to T in the URA3 coding strand (Figs. 5a and 5c). This indicates that CD-T7RNAP fusions mutates more frequently Cs in the non-coding strand of URA3, which corresponds to the non-template strand for the CD-T7RNAP fusions (Supplementary Fig. 3). This mutagenesis bias to the non-template strand is less pronounced for AID-T7RNAP (62%) than for rAPOBEC1-T7RNAP (74%) or pmCDA1-T7RNAP (91%) (Fig. 5c). In the case of TadA*-T7RNAP, we also found a bias favoring T to C mutations in the coding strand of URA3 (84%), corresponding to A to G mutations in the non-template strand for the TadA*-T7RNAP fusion (Fig. 5c and Supplementary Fig. 3). Therefore, DNA sequencing of URA3 mutant clones demonstrates that CD and AD fusions induce the expected mutations with a variable bias towards the non-template strand depending on the specific BD employed.

For a further analysis of the mutagenic process a 284 bp PCR fragment of the URA3 gene was amplified from a culture of MG*-URA3Δ*ung* expressing AID-T7RNAP and was subjected to massive next-generation sequencing (NGS; Materials and Methods). As a control, the same region was amplified from an induced culture of MG*-URA3Δ*ung* carrying the empty plasmid pSEVA221 and also subjected to NGS. The results of the variant call analysis from the two samples (ca. $1x10^6$ reads/sample) were compared and only the transition C:G to T:A appeared in an statistically significant higher number of reads in bacteria expressing AID-T7RNAP (Supplementary Fig. 4). Using an identical experimental approach, we analyzed the mutations

caused in URA3 by TadA*-T7RNAP in MG*-URA3Δ*ung*Δ*nfi* compared with a culture of the same strain carrying pSEVA221. In this case, only transitions A:T to G:C were detected in a statistically significant higher number of reads in the sample expressing TadA*-T7RNAP (Supplementary Fig. 5). Hence, massive DNA sequencing data is consistent with the DNA sequencing results of individual URA3 mutants and confirms that the AID- and TadA*-T7RNAP generates only their expected transitions.

Protection from mutagenesis of downstream regions by blockade with dCas9. The BD-T7RNAP fusions were able to transcribe the *gfp* gene downstream of URA3 and, presumably, creating mutations in this gene and beyond. To confirm this, we inserted in the GFP-URA3 cassette the *sacB* gene from *Bacillus subtilis* as an additional counter-selection system [53]. The gene *sacB* codes for exoenzyme levansucrase, which uses sucrose to produce levan that accumulates in the periplasm of *E. coli* killing the bacteria. A Ptac-*sacB* fusion was inserted dowstream *gfp* in the GFP-URA3 cassette (Fig. 6a), and the new cassette was integrated in the *flu* locus of MG*Δ*pyrF*Δ*ung*Δ*nfi*. The resulting strain, named MG*-SacB-URA3Δ*ung*Δ*nfi*, was sensitive to sucrose with a frequency of spontaneous mutants of ca. $5x10^{-6}$. When AID-T7RNAP was expressed in this strain the mutagenesis frequency of *sacB* increased to ~$6x10^{-4}$ (Supplementary Fig. 6). The mutagenesis frequency of URA3 in this strain with AID-T7RNAP (~$1.7x10^{-3}$) was similar to that observed in MG*-URA3Δ*ung*Δ*nfi* (Fig. 4d). These results confirm that BD-T7RNAP fusions are able to mutate downstream regions to the target gene.

To delimit the mutagenesis of BD-T7RNAP fusions within the target gene, we investigated the possibility of blocking elongation of T7RNAP with dCas9. dCas9 can bind a target DNA sequence using a crRNA or gRNA and has been successfully used as transcriptional repressor for the endogenous *E. coli* RNAP [54]. Importantly, co-expression of two gRNAs targeting the same gene enhanced the transcription repression of *E. coli* RNAP by dCas9 [55]. To study whether dCas9 can block elongation of BD-T7RNAP fusions we designed three crRNAs (Ta, Tb, Tc) against the non-template strand of *gfp* (Figs. 6a and 6b) and generate double (Tb.a) and triple (Tb.a.c) crRNAs. The plasmid pdCas9 [54] was used for the constitutive expression of dCas9, the

transactivating CRISPR RNA (tracrRNA) and the designed crRNAs. The strain MG*-SacB-URA3Δ*ung*Δ*nfi* was transformed with plasmid pdCas9 (without crRNAs) and derivatives pdCas9-Tb.a and pdCas9-Tb.a.c. Then, these strains were transformed with the plasmid encoding AID-T7RNAP or the empty vector pSEVA221. After growth and aTc induction the levels of GFP expression were determined by flow cytometry (Fig. 6c). The expression of the double crRNAs Tb.a repressed GFP levels to ~30% and with the triple crRNAs Tb.a.c the levels of GFP were further repressed to ~20%. The basal level of GFP expression in strains carrying pSEVA221 was consider 0%. The protection of *sacB* from the BD-T7RNAP mutagenesis was assessed by the ratio of mutation frequencies in *sacB* vs. URA3 in each of these strains. This ratio was normalized as 1 for the strain expressing AID-T7RNAP and carrying pdCas9 without crRNAs (Fig. 6d). In concordance with the reduction of GFP expression detected by flow cytometry, the mutagenesis of *sacB* dropped ~10 to ~14-fold when the double (Tb.a) and triple (Tb.a.c) crRNAs were expressed, respectively (Fig. 6d). These results demonstrate that the dCas9 blockade can be used to limit the mutagenesis activity of BD-T7RNAP fusions mostly to a target gene (or gene segment), significantly reducing the mutagenesis of downstream DNA regions.

Discussion

In this work we have reported an *in vivo* mutagenesis system with high specificity for a target gene based on the tethering of different BDs (AID, rAPOBEC1, pmCDA1 and TadA*) fused to T7RNAP, which selectively recognizes its specific promoter (P_{T7}) in reverse orientation at the 3'-end and that transcribes thorughout the target gene. Expression of the target gene is maintained from its endogenous 5'-end promoter. The use of different BDs confers flexibility to the system due to their different mutagenesis profile. Although the expression level and transcriptional activity of CD-T7RNAP fusions were shown to be similar, their mutagenic capacity varied greatly ranging from the least active fusion with AID (URA3 mutagenesis frequency of ca. 10^{-3}) to the most active fusion with pmCDA1 (URA3 mutagenesis frequency of ca. 10^{-1}). The fusion bearing rAPOBEC1 produced an intermediate URA3 mutagenesis frequency of ca. 10^{-2}. This means that the mutagenesis frequency of a particular target gene can be modulated using different CD-

T7RNAP fusions. As expected, the mutations found in the target gene (URA3) by the CD-T7RNAP fusions were transitions C:G to T:A corresponding to the CD activity. In order to broad the mutation spectrum to A:T base pairs in the target gene, we also constructed the fusion TadA*-T7RNAP using the modified adenosine deaminase TadA7.10 (TadA*) [28]. This fusion was expressed at higher levels than the CD-T7RNAP fusions, and was proved to generate bp transitions A:T to G:C in the target gene (URA3) with a mutagenesis frequency of ca. 10^{-3}, similar to that elicited by AID-T7RNAP fusion. The average number of mutations in URA3 (ca. 1 kb) per clone agreed with the frequency of FOA^R mutants produced by the expression of the different fusions, being the fusion with pmCDA1 the enzyme introducing the highest number of mutations (ca. 14) whereas AID and TadA* fusions introduced the lowest number of mutations per clone (ca. 1.2 - 1.4). A single mutant isolated from TadA*-T7RNAP expression contained an unexpected bp transition (C:G to T:A), which could be caused by a non-specific base deaminase activity of TadA* or by the use of a common *E. coli* Δ*ung* host strain for expression of all BD-T7RNAP in these experiments. Nevertheless, massive DNA sequencing of a short amplicon of URA3, obtained after induction of AID and TadA* fusions without selection process, detected only the expected bp transitions C:G to T:A for AID and A:T to G:C for TadA*. Importantly, no other type of mutations (e.g. deletions, insertions) were found among the mutants isolated after expression of any of the BD-T7RNAP fusions. This result contrasts with the various types of mutations found after expression of T7RNAP alone. Our data revealed that transcription caused by T7RNAP alone increases the mutation frequency of the target gene by a mechanism that is independent of BDs, but that might be related to high exposure of ssDNA in the target gene and/or with conflicts of transcription with other cellular machineries (e.g. DNA polymerases during replication) [56].

Elimination of the DNA repair enzymes UNG and endonuclease V (*nfi*) was important to obtain optimal mutagenesis frequencies. It is well-known that UNG has a key role in reverting the mutations caused by CDs, so this enzymatic activity was inhibited in other mutagenesis systems based on CDs [18, 19]. In our case, when UNG was present the mutagenesis frequency dropped between 500-fold for AID and 3000-fold for pmCDA1 and rAPOBEC. We deleted *ung* in our host

405 strains to ensure complete abrogation of UNG activity, but an interesting alternative to *ung*
406 deletion is its transient inhibition by expression of UGI (uracil N-glycosylase inhibitor) from
407 bacteriophage PBS2 [25]. For the mutagenesis with TadA*-T7RNAP, the effect of the
408 endonuclease V was very mild. Only a decrease of 2- to 5-fold in the mutagenesis frequency
409 with TadA*-T7RNAP was detected in bacteria with endonuclease V. Therefore, it seems that
410 the removal of inosines generated by TadA* was much less efficient than removal of uracils
411 generated by CDs in the genomic DNA of *E. coli*. This was also observed in eukaryotic cells
412 subjected to mutagenesis with TadA* [28]. It is worth to mention that UNG did not affect the TadA*-
413 T7RNAP mutagenesis nor did endonuclease V affect the mutagenesis of AID-T7RNAP fusion.
414
415 Another observation to be noted is that the mutagenic capacity of the BD-T7RNAP fusions
416 presented a bias towards the non-template strand. This phenomenon is less pronounced for
417 AID-T7RNAP but it is detected with all of them. The reason of this strand preference is unclear
418 but it may be caused by a better exposure of the cytosines or adenines on this strand in the
419 transcription bubble. The template strand could be less accesible due its insertion in the catalytic
420 core of the T7RNAP and the formation of the ssDNA:RNA hybrid [57].
421
422 A key property of a targeted mutagenesis system is the specificity towards the on-target
423 sequence, keeping as low as possible the "off-target" mutagenesis to avoid the generation of
424 deleterious mutations in other parts of the genome. The specifity of the system was assesed by
425 two different approaches: monitoring mutations in the *rpoB* gene that conferred rifampicin
426 resistance and monitoring mutations in the URA3 gene in absence of the T7 promoter. We found
427 very high on-target vs. off-target ratios for all BD-T7RNAP fusions (ca. ≥ 10^3). Nevertheless,
428 some off-target mutagenesis is observed by the expression of these fusions. Compared to
429 mutations frequencies found in a host strain (e.g. Δ*ung*) without BD-T7RNAP fusion, the "off-
430 target" mutagenesis increased moderately (ca. 2 to 5- fold) for AID- and TadA*-fusions and more
431 noticiable (ca. 10 to 20-fold) for pmCDA1- and rAPOBEC-fusions.
432

433 A different situation is caused by "off-target" mutations in downstream sequences (in relation to
434 the T7 promoter) of the target gene due to the high processivity of the BD-T7RNAP fusions. To
435 restrict the mutagenesis of BD-T7RNAP fusions to the target gene, we used dCas9 directed with
436 crRNAs to block transcriptional elongation. This strategy has been previously used to repress
437 gene expression in *E. coli* by blocking the endogenous RNA polymerase [54, 55] . It has been
438 reported that dCas9 blocks the *E. coli* RNAP when the crRNAs bind to the non-template strand
439 and the presence of two targeting crRNAs enhance the repression [55]. Keeping this in mind, we
440 designed three crRNAs targeting the non-template strand. We have shown that the 3
441 crRNAs/dCas9 complexes are able to inhibit elongation of BD-T7RNAP fusions and reduce ca.
442 14-fold the mutagenesis of adjacent downstream DNA (with respect to T7RNAP elongation).
443 Therefore, the protection from mutagenesis of dowstream regions was significant, yet it may be
444 improved with dCas9 variants having increased blockade activity [58].
445
446 One interesting feature of the reported method is that it could be adapted for its use in different
447 hosts since BDs, T7RNAP and dCas9 have been expressed in different bacteria, yeast and
448 mammalian cells [18, 19, 21, 30, 31, 32]. For modulation of the DNA repair systems, the expression of
449 UGI would avoid the need of delete *ung* in host that are not *E. coli*. Elimination of *nfi* is not
450 essential because the mutagenesis with TadA*-T7RNAP is still significant in the presence of the
451 endonuclease V. However, the modified TadA* pairs with the endogenous TadA of *E. coli* to
452 render a full active enzyme [28]. Hence, for mutagenesis in hosts that lacks endogenous TadA, a
453 heterodimeric construct fusing TadA and TadA* is needed [28]. In mammalian cells incorporation
454 of a nuclear localization signal (NLS) may be also necessary for targeting these protein fusions
455 to the cell nucleus.
456
457 The system described in this work can be used for molecular evolution of different genes and
458 operons of interest by simply replacing URA3 by these new target sequences. Alternatively the
459 T7 promoter can be integrated downstream of genes or operons in the genome of *E. coli* or
460 other cell host. The integration of multiple T7 promoters in different parts of the genome (e.g.,
461 downstream of genes involved in a metabolic pathway or in antibiotic resistance) would render

462 in a multiplex genome editing for directed evolution of a cell phenotype. Alternatively, a plasmid
463 that carries the target gene/s and the T7 promoter can be used. A recent independent study
464 reported that a fusion between rAPOBEC1 and T7RNAP was able to generate mutations in
465 antibiotic resistance genes located in plasmids carrying the T7 promoter [59]. The use of multicopy
466 plasmids can ease the initial utilization of the system but it has the drawback that creates multiple
467 variants of the target gene/s in one cell. For selection of a particular variant, plasmid isolation
468 and re-transformation of cells would be needed. When the genetic determinants are integrated
469 in single copy in the genome, the mutagenesis and selection of the variants with the desired
470 properties can be done in sequential iterative cycles without extensive manipulation of the
471 culture as a continuos evolution process. We have demonstrate that different BDs-T7RNAP
472 fusions can be used for the mutagenesis, and in particular that AID has lower off-target activity
473 and mutagenic bias to the non-template strand than rAPOBEC1. Finally, our study also
474 demonstrates that dCas9 with crRNAs can be used to limit the mutagenic activity within a target
475 gene or gene segment without the introduction of multiple transcriptional terminators [59].
476
477 In sum, the BD-T7RNAP-based method hereby documented poses a simple workflow that could
478 be applied for continuos molecular evolution of genes or operons with minimun manipulation.
479 The extension of the mutagenized DNA can be delimited with dCas9 directed with designed
480 crRNAs. Hence, *in vivo* mutagenesis with BD-T7RNAP fusions has the potential to be applied
481 for the molecular evolution of biotechnologically relevant proteins, metabolic engineering of
482 enzymatic pathways, diversification of gene libraries and other applications such as the study of
483 the potential evolution pathways of antibiotic resistance genes.
484
485 ## Methods

486 **Bacterial strains, media and growth conditions.** A full list of the bacterial strains used in this
487 study can be found in Supplementary Table 1. For plasmid propagation and cloning the *E. coli*
488 strains used were DH10BT1R [60] and BW25141 (for the *pir*-dependent suicide plasmid
489 derivatives) [61] . All the strains were grown in Lysogeny Broth (LB) at 37 ºC and shaking at 250
490 rpm [62]. For solid medium, 1.5 % (w/v) agar was added to LB. The following antibiotics were

491 added to the medium when needed: 50 μg/ml kanamycin (Km), 30 μg/ml chloramphenicol (Cm),
492 50 μg/ml apramycin (Apra) and 50 μg/ml rifampicin (Rif). Antibiotics were obtained from
493 Duchefa-Biochemie. All other chemical reagents were obtained from Merck-Sigma unless
494 indicated otherwise. For monitoring the mutagenic process, minimal medium M9 plates were
495 used [62]. This medium contains: 1x M9 salts (1 g/l NH_4Cl, 3 g/l KH_2PO_4 and 6 g/l Na_2HPO_4), 2
496 mM $MgSO_4$, 0.4 % (w/v) glucose, 0.0005 % (w/v) thiamine and 1.6 % (w/v) agar for solidification.
497 When required, the minimal medium was supplemented with 20 μg/ml uracil and 250 μg/ml 5-
498 fluororoic acid (FOA) (Zymo Research) or 60 g/l sucrose (counter-selection with *sacB*).
499
500 **Plasmids and cloning procedures**. A list of the plasmids used in this study can be found in
501 the Supplementary Table 2. The plasmids pGE*pyrF*, pGE*ung* and pGE*nfi* derived from *pir*-
502 dependent plasmid pGE (Km^R, R6K origin of replication) were constructed to delete the genes
503 *pyrF*, *ung* and *nif*, respectively. Thermosensitive plasmids derivatives of pGETS (Km^r, pSC101-
504 ts origins of replication) were used to integrate the mutagenesis reporter cassettes in the non-
505 essential locus of *flu*. The homology regions of the genes *pyrF*, *ung*, *nif* and *flu* were amplified
506 by PCR using as template genomic DNA from *E. coli* MG1655. The gene gfp^{TCD} [63] coding for
507 GFP was obtained from the plasmid pGEyeeJPtac-gfp [38], the URA3 gene was a synthetic
508 version optimized for *E. coli* (GeneCust), and the *sacB* gene was amplified from genomic DNA
509 from *E. coli* T-SACK [64]. The plasmid pSEVA221 (Km^r, RK2-origin) [65] was used for expression of
510 the T7RNAP fusions under the control of the *tetR*-P_{tetA} promoter [46]. The *tetR*-P_{tetA} and the DNA
511 fragments coding for the base deaminase (BD) enzymes fused to the linker peptide $(G_3S)_7$ were
512 chemically synthesized with codon optimization for expression in *E. coli* (GeneART,
513 ThermoFisher Scientific). The following BD enzymes were synthesized: human AID (Activation-
514 induced cytidine deaminase), rAPOBEC1 (rat apolipoprotein B mRNA editing enzyme) [66] ,
515 pmCDA1 (lamprey cytidine deaminase 1) [67], and TadA* (*E. coli* adenine deaminase variant
516 TadA7.10) [28]. The proofreading DNA polymerase Herculase II Fusion (Agilent Technologies)
517 was used to amplify DNA fragments for cloning purposes. The plasmid pdCas9 [54] was used for

518 the constitutively expression of the catalytically "dead" Cas9 (dCas9), the trans-activating crRNA
519 RNA (tracrRNA) and the CRISPR RNAs (crRNA). The double (Tb.a) and triple (Tb.a.c) spacer
520 arrays were cloned into the *Bsa*I site of pdCas9 using hybridized complimentary
521 oligonucleotides, and following the one-step scheme CRATES [68] (Supplementary Methods).
522 Cloning procedures were performed following standard protocols of DNA digestion with
523 restriction enzymes and ligation [62]. All DNA constructs were sequenced by the chain-termination
524 Sanger method (Macrogen) and the sequences are deposited in GenBank (Supplementary
525 Table 2).
526
527 **Generation of the reporter strains**. The reporter strains for the mutagenesis system were
528 derived from the reference of *E. coli* K-12 strain MG1655 [39]. The strain MG1655* with higher
529 sensitivity to FOA was generated by an oligo-mediated allelic replacement method [16]
530 (Supplementary Methods). All the successive modifications were done over the MG1655*
531 genetic background. The genes *pyrF*, *ung* and *nfi* were deleted using the plasmids pGE*pyrF*,
532 pGE*ung* and pGE*nfi*, respectively, by a marker-less genome edition strategy [37]. This strategy is
533 based homologous recombination and resolution of the cointegrant promoted by the expression
534 of the restriction enzyme I-*Sce*I and the λRed from the plasmid pACBSR [69]. The URA3 cassettes
535 were inserted in the *flu* locus using plasmid derivatives of pGETS, as reported previously [38].
536
537 **Induction of the mutagenesis system**. A colony of freshly transformed reporter strain with the
538 indicated plasmid derivative of pSEVA221, was grown overnight (O/N) in LB with Km at 37 ºC
539 with shaking (250 rpm). The next day, the culture was diluted 1:100 in fresh media and incubated
540 under the same conditions for 2 h. Then anhydrotetracycline (aTc, 200 ng/ml) (TOKU-E) was
541 added for induction and the cultures were incubated for 1 h. After that, a 500 μl-aliquot of each
542 culture was washed with 1X phosphate-buffered saline (PBS) and resuspended in the same
543 volume of PBS. A series of ten-fold dilutions of the cell suspension was prepared, and aliquots

544 of 100 μl were plated in duplicates on different media: M9 + uracil; M9 + uracil and FOA; M9 +
545 uracil and sucrose; and LB + Rif (see above for media composition).
546
547 **Flow cytometry analysis**. Levels of expression of *gfp* from induced cultures were determined
548 by flow cytometry analysis as follows: the volume corresponding to one unit of optical density
549 (O.D.) at 600 nm of the induced cultures was collected by centrifugation (3300xg, 5 min) and
550 resuspended in 500 μl 1X PBS. The cell suspension was diluted transferring 200 μl to a tube
551 with 1200 μl of 1X PBS, and its fluorescence levels was determined using a Gallios FC500 flow
552 cytometer (Beckman Coulter).
553
554 **Western blot analysis**. To detect the expression of the fusion proteins in cell extracts of induced
555 cultures, Western blot analysis was performed. From induced cultures, 0.5 O.D. was collected
556 by centrifugation 3300 g 5 min, washed once with 500 μl 1X PBS, resuspended in 60 μl of PBS
557 and mixed with 15 μl of 5X SDS-PAGE sample buffer [70]. The samples were boiled 10 min before
558 loaded into 8% polyacrylamide SDS gels, and electrophoresis was done using the Miniprotean
559 III system (Bio-Rad) during 1 h 30 min at 170 V. The gels were then transferred to a
560 polyvinylidene difluoride membrane (PVDF, Immobilon-P, Millipore) by means of O/N wet
561 transfer (Bio-Rad) 4 ºC at 30 V. The membranes were blocked with PBS 0.1% (v/v) Tween 20
562 with 3 % (w/v) skim milk powder and successively incubated with monoclonal mouse anti-T7
563 RNA polymerase antibodies (Novagen, Merck) and POD labelled goat anti-mouse antibodies
564 (Sigma). Membranes were developed by chemiluminiscence using the Clarity Western ECL
565 Substrate kit (Bio-Rad) and images were acquired using a ChemiDoc Touch system (Bio-Rad).
566
567 **DNA sequencing and analysis**. To determine the DNA sequence of the URA3 alleles in the
568 FOA resistant colonies, a DNA fragment of 1191 bp was amplified using the pair of primers
569 F_GFPseq / R_T0ter (Supplementary Table S6) with the GoTaq Flexi DNA polymerase
570 (Promega) by colony PCR following manufacturer's instructions. The resulting amplicon was

sequenced by the Sanger chain-termination method (Macrogen) using the same primers. The resulting 2 reads per colony were mapped against the Ptac-URA3-P_{T7} reference sequence (942 bp) to detect variants using the program SeqMan Pro (DNAstar).

For deeper analysis of the variations in URA3, NGS sequencing analysis of a 200 bp region was performed. To do this, genomic DNA was extracted from ~4.5 ml of each induced culture (~5 x 10^9 bacteria) using the GNOME DNA Kit (MP Biomedicals). One hundred ng of total genomic DNA was used as template in a PCR reaction to amplify 284 bp of URA3 with the pair of primers F_CS1_URA3 / R_CS2_URA3 that includes Illumina tags CS1 and CS2 (Supplementary Table S6). The DNA amplification was carried out in 50 μl reactions using Herculase II Fusion DNA polymerase (Agilent Technologies) following manufacturer's instructions. The amplicons were sent to the Genomic Unit of the Madrid Scientific Park to be sequenced by the NGS platform Illumina Miseq with paired-end (length > 2 x 300 bp) to acquire ~1,000,000 reads per sample. These reads were processed with the program Bbmap (Bushnell et al., 2017, PMID: 29073143) for merging the paired-end reads. The resulting merged files were pilep up against the reference sequence using the program Samtools (Li et al., 2009, PMID: 19505943), and the variants were obtained with the program VarScan (Koboldt et al., 2012, PMID: 22300766) with the following parameters: --min-coverage 1 --min-reads2 1 --min-avg-qual 40 --min-var-freq 0.000001 --p-value 0.99. The sequences of the oligonucleotides used for amplification were discarded from the analysis since they may contain variations due to chemical synthesis.

Statistics. Means and standard errors of experimental values were calculated using Prism 5.0 (GraphPad software Inc). Statistical analyses comparing groups in pairs were performed using Mann Whitney test (Fig. 3d, Supplementary Figs. S4 and S5) and two-tailed Student's t-test (Figs. 4d and 4e, Fig. 6d) from at least three independent experiments. A value of $p<0.05$ was considered significant.

Data availability.

Data that support the findings of this work can be found in the main manuscript and in the Supplementary information. Materials and additional data are available from the corresponding author upon request. The sequences of the constructs built for this study are deposited in GenBank with the accession numbers listed in Supplementary Table 2.

References

1. Tobin PH, Richards DH, Callender RA, Wilson CJ. Protein engineering: a new frontier for biological therapeutics. *Curr Drug Metab* **15**, 743-756 (2014).

2. Porter JL, Rusli RA, Ollis DL. Directed Evolution of Enzymes for Industrial Biocatalysis. *Chembiochem* **17**, 197-203 (2016).

3. Davis AM, Plowright AT, Valeur E. Directing evolution: the next revolution in drug discovery? *Nature Reviews Drug Discovery* **16**, 681 (2017).

4. Packer MS, Liu DR. Methods for the directed evolution of proteins. *Nature reviews Genetics* **16**, 379-394 (2015).

5. Simon AJ, d'Oelsnitz S, Ellington AD. Synthetic evolution. *Nature Biotechnology* **37**, 730-743 (2019).

6. Jijakli K*, et al.* The in vitro selection world. *Methods* **106**, 3-13 (2016).

7. Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503 (2011).

8. Foster PL. In Vivo Mutagenesis. *Methods in enzymology* **204**, 114-125 (1991).

9. Badran AH, Liu DR. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat Commun* **6**, 8425 (2015).

10. Muteeb G, Sen R. Random mutagenesis using a mutator strain. *Methods Mol Biol* **634**, 411-419 (2010).

11. Nguyen AW, Daugherty PS. Production of randomly mutated plasmid libraries using mutator strains. *Methods Mol Biol* **231**, 39-44 (2003).

12. Irving RA, Kortt AA, Hudson PJ. Affinity maturation of recombinant antibodies using *E. coli* mutator cells. *Immunotechnology* **2**, 127-143 (1996).

13. Camps M, Naukkarinen J, Johnson BP, Loeb LA. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proceedings of the National Academy of Sciences* **100**, 9727-9732 (2003).

14. Wang HH*, et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009).

15. Bonde MT, *et al.* Direct mutagenesis of thousands of genomic targets using microarray-derived oligonucleotides. *ACS synthetic biology* **4**, 17-22 (2015).

16. Nyerges A, *et al.* A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc Natl Acad Sci U S A* **113**, 2502-2507 (2016).

17. Crook N, Abatemarco J, Sun J, Wagner JM, Schmitz A, Alper HS. In vivo continuous evolution of genes and pathways in yeast. *Nat Commun* **7**, 13051 (2016).

18. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).

19. Nishida K, *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, (2016).

20. Hess GT, *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods* **13**, 1036-1042 (2016).

21. Li X, *et al.* Base editing with a Cpf1-cytidine deaminase fusion. *Nat Biotechnol* **36**, 324-327 (2018).

22. Lada AG, *et al.* Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry (Mosc)* **76**, 131-146 (2011).

23. Poltoratsky VP, Wilson SH, Kunkel TA, Pavlov YI. Recombinogenic phenotype of human activation-induced cytosine deaminase. *J Immunol* **172**, 4308-4313 (2004).

24. Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99-103 (2002).

25. Wang ZG, Smith DG, Mosbaugh DW. Overproduction and characterization of the uracil-DNA glycosylase inhibitor of bacteriophage PBS2. *Gene* **99**, 31-37 (1991).

26. Schormann N, Ricciardi R, Chattopadhyay D. Uracil-DNA glycosylases-Structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Science* **23**, 1667-1685 (2014).

27. Krokan HE, Bjoras M. Base excision repair. *Cold Spring Harbor perspectives in biology* **5**, a012583 (2013).

28. Gaudelli NM, *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464 (2017).

29. Halperin SO, Tou CJ, Wong EB, Modavi C, Schaffer DV, Dueber JE. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248-252 (2018).

30. Bolukbasi MF, Gupta A, Wolfe SA. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat Meth* **13**, 41-50 (2016).

31. Dower K, Rosbash M. T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export. *RNA* **8**, 686-697 (2002).

32. Lieber A, Sandig V, Strauss M. A mutant T7 phage promoter is specifically transcribed by T7-RNA polymerase in mammalian cells. *Eur J Biochem* **217**, 387-394 (1993).

33. Jones ME. Orotidylate decarboxylase of yeast and man. *Curr Top Cell Regul* **33**, 331-342 (1992).

34. Boeke JD, LaCroute F, Fink GR. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genet* **197**, 345-346 (1984).

35. Meng X, Smith RM, Giesecke AV, Joung JK, Wolfe SA. Counter-selectable marker for bacterial-based interaction trap systems. *Biotechniques* **40**, 179-184 (2006).

36. van der Woude MW, Henderson IR. Regulation and function of Ag43 (flu). *Annu Rev Microbiol* **62**, 153-169 (2008).

37. Piñero-Lambea C, Bodelón G, Fernández-Periáñez R, Cuesta AM, Álvarez-Vallina L, Fernández LA. Programming Controlled Adhesion of *E. coli* to Target Surfaces, Cells, and Tumors with Synthetic Adhesins. *ACS synthetic biology* **4**, 463-473 (2015).

38. Ruano-Gallego D, Álvarez B, Fernández LÁ. Engineering the Controlled Assembly of Filamentous Injectisomes in *E. coli* K-12 for Protein Translocation into Mammalian Cells. *ACS synthetic biology* **4**, 1030-1041 (2015).

39. Blattner FR*, et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1462 (1997).

40. Hayashi K*, et al.* Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* **2**, 2006 0007 (2006).

41. Jensen KF. The *Escherichia coli* K-12 "wild types" W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. *Journal of Bacteriology* **175**, 3401-3407 (1993).

42. Salema V, Fernández LÁ. High yield purification of nanobodies from the periplasm of *E. coli* as fusions with the maltose binding protein. *Protein Expression and Purification* **91**, 42-48 (2013).

43. Jurado P, de Lorenzo V, Fernández LA. Thioredoxin Fusions Increase Folding of Single Chain Fv Antibodies in the Cytoplasm of *Escherichia coli*: Evidence that Chaperone Activity is the Prime Effect of Thioredoxin. *J Mol Biol* **357**, 49-61 (2006).

44. Kern R, Malki A, Holmgren A, Richarme G. Chaperone properties of *Escherichia coli* thioredoxin and thioredoxin reductase. *Biochem J* **371**, 965-972 (2003).

45. Bach H*, et al. Escherichia coli* Maltose-binding Protein as a Molecular Chaperone for Recombinant Intracellular Cytoplasmic Single-chain Antibodies. *J Mol Biol* **312**, 79-93. (2001).

46. Bertram R, Hillen W. The application of Tet repressor in prokaryotic gene regulation and expression. *Microbial Biotechnology* **1**, 2-16 (2008).

47. Berens C, Hillen W. Gene regulation by tetracyclines. Constraints of resistance regulation in bacteria shape TetR for application in eukaryotes. *European Journal of Biochemistry* **270**, 3109-3121 (2003).

48. Silva-Rocha R*, et al.* The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res* **41**, D666-675 (2013).

49. Ovchinnikov YA*, et al.* RNA polymerase rifampicin resistance mutations in *Escherichia coli*: Sequence changes and dominance. *Molecular and General Genetics MGG* **190**, 344-348 (1983).

50. Severinov K, Soushko M, Goldfarb A, Nikiforov V. RifR mutations in the beginning of the *Escherichia coli* rpoB gene. *Molecular and General Genetics MGG* **244**, 120-126 (1994).

51. Vik ES*, et al.* Endonuclease V cleaves at inosines in RNA. *Nat Commun* **4**, 2271 (2013).

52. Guo G, Ding Y, Weiss B. nfi, the gene for endonuclease V in *Escherichia coli* K-12. *J Bacteriol* **179**, 310-316 (1997).

53. Gay P, Le Coq D, Steinmetz M, Berkelman T, Kado CI. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* **164**, 918-921 (1985).

54. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**, 7429-7437 (2013).

55. Qi Lei S*, et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173-1183 (2013).

56. Jinks-Robertson S, Bhagwat AS. Transcription-Associated Mutagenesis. *Annual Review of Genetics* **48**, 341-359 (2014).

57. Durniak KJ, Bailey S, Steitz TA. The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* **322**, 553-557 (2008).

58. Cebrian-Serrano A, Davies B. CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm Genome* **28**, 247-261 (2017).

59. Moore CL, Papa LJ, 3rd, Shoulders MD. A Processive Protein Chimera Introduces Mutations across Defined DNA Regions In Vivo. *J Am Chem Soc*, (2018).

60. Durfee T*, et al.* The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* **190**, 2597-2606 (2008).

61. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645. (2000).

62. Sambrook J, Russel DW. *Molecular cloning. A laboratory manual*, 3rd edn. Cold Spring Harbor Laboratory Press (2001).

63. Corcoran CP, Cameron AD, Dorman CJ. H-NS silences gfp, the green fluorescent protein gene: gfpTCD is a genetically Remastered gfp gene with reduced susceptibility to H-NS-mediated transcription silencing and with enhanced translation. *J Bacteriol* **192**, 4790-4793 (2010).

64. Li X-t, Thomason LC, Sawitzke JA, Costantino N, Court DL. Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in *Escherichia coli*. *Nucleic Acids Research* **41**, e204 (2013).

65. Martinez-Garcia E, Aparicio T, Goni-Moreno A, Fraile S, de Lorenzo V. SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res* **43**, D1183-1189 (2015).

66. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol* **9**, 229 (2008).

67. Rogozin IB*, et al.* Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat Immunol* **8**, 647-656 (2007).

68. Liao C*, et al.* Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nature Communications* **10**, 2948 (2019).

69. Herring CD, Glasner JD, Blattner FR. Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene* **311**, 153-163 (2003).

70. Jurado P, Ritz D, Beckwith J, de Lorenzo V, Fernández LA. Production of functional single-chain Fv antibodies in the cytoplasm of *Escherichia coli*. *J Mol Biol* **320**, 1-10. (2002).

Acknowledgements

We thank Dr. David Bickard (Institute Pasteur, France) and Drs. Esteban Martínez, Yamal Al-ramahi, and Tomás Aparicio (CNB-CSIC) for providing some materials used in this work. The excellent technical work in massive DNA sequencing of the Genomic Unit of "*Parque Científico de Madrid*" is greatly appreciated. We are grateful to Dr. Alejandra Bernardini (Hospital 12 de Octubre Madrid, Spain) for assistance with massive DNA sequencing data analysis. This work is supported by the Grants BIO2017-89081-R (Agencia Española de Investigación AEI/MICIU/FEDER, EU) to LAF and ERC-2012-ADG_20120314 (European Research Council,EU) to VdL.

Author contributions

LAF, MM and VdL conceived the study. BA and LAF designed the experiments and analysed the results. BA performed the experiments. All authors interpreted the data. BA and LAF wrote the initial manuscript and prepared figures. All the authors revised and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests

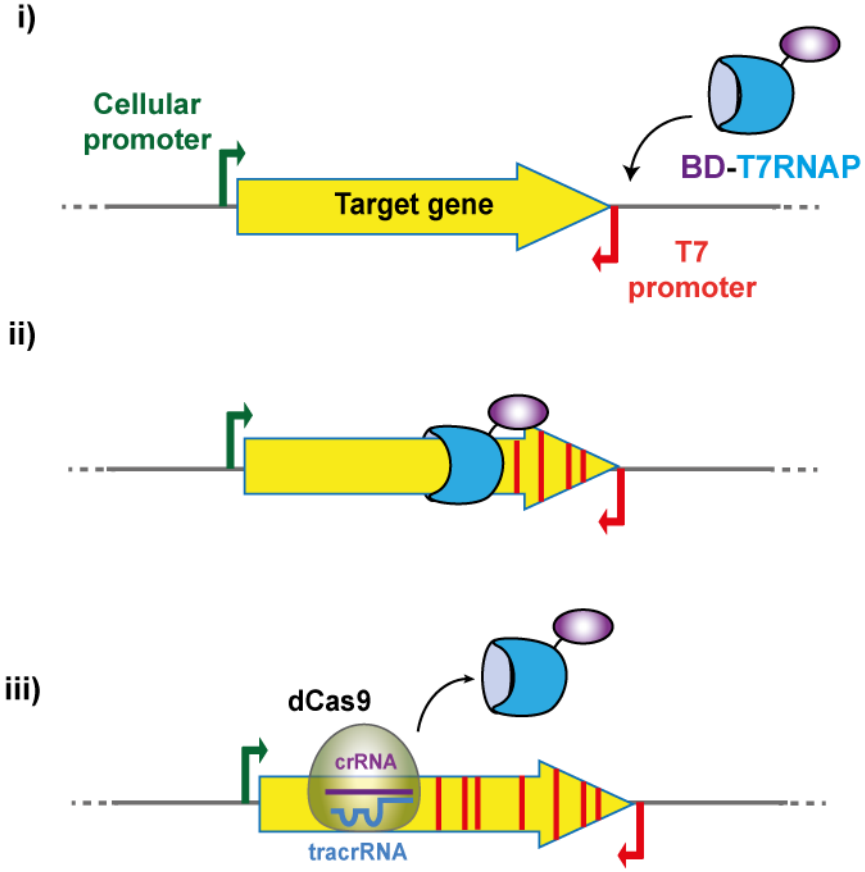

Figure 1.Schematic representation of the mutagenic process. The T7 RNA polymerase of the fusion specifically binds the T7 promoter (i), initiating the transcription and moving along the target gene carrying the base deaminase (BD) that creates mutations (red stripes) in this gene (ii). The enzymatic fusion stops and detaches from the DNA when encounters a dCas9 molecule bound to a specific sequence determined by the crRNA (iii).

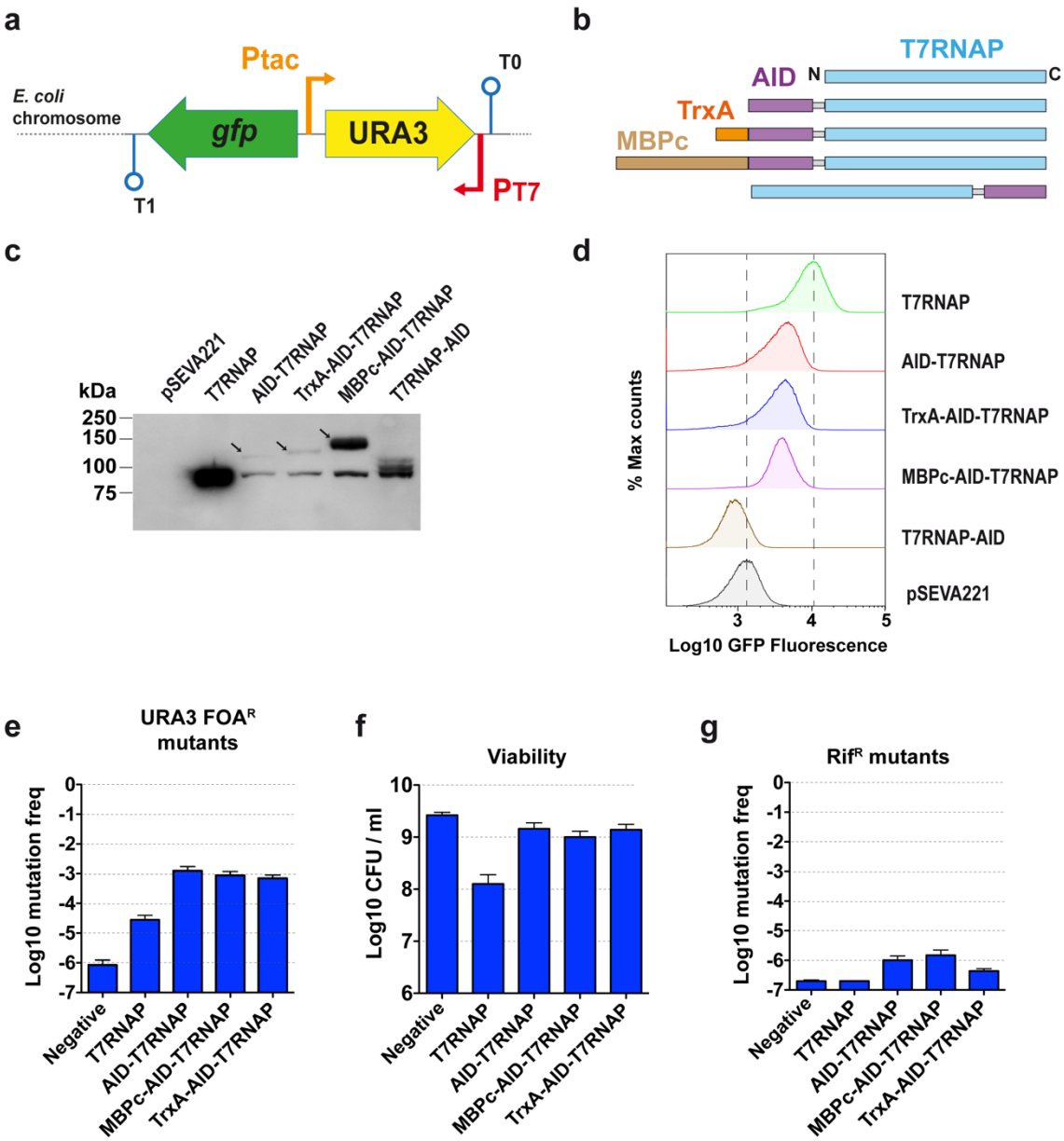

Figure 2. Expression and activity of the different variants of the fusion AID-T7RNAP. (**a**) Representation of the chromosomally integrated reporter cassette to test the mutagenesis system. Thin arrows indicate the promoters tac (P_{tac}) and T7 (P_{T7}), lollipops indicate terminators T0 and T1. (**b**) Representation of the different AID fusion variants. (**c**) Expression of the different variants determined by Western blot analysis of the cell extracts from induced cultures of the strain MG*-URA3Δ*ung* transformed with the different plasmids. The arrows indicate the bands corresponding to the full-length fusions. (**d**) Processivity of the fusions assessed by flow cytometry analysis to detect expression of *gfp* in the induced cultures. (**e**) Mutation frequency of URA3 as the ratio of FOA^R CFU/ml *vs*. total CFU/ml. (**f**) Viability as Log_{10} CFU/ml. (**g**) Mutation frequency of *rpoB* as the ratio of Rif^R CFU/ml *vs*. total CFU/ml. The histograms (**e**, **f** and **g**) represent the means and standard errors of three independent experiments (n=3).

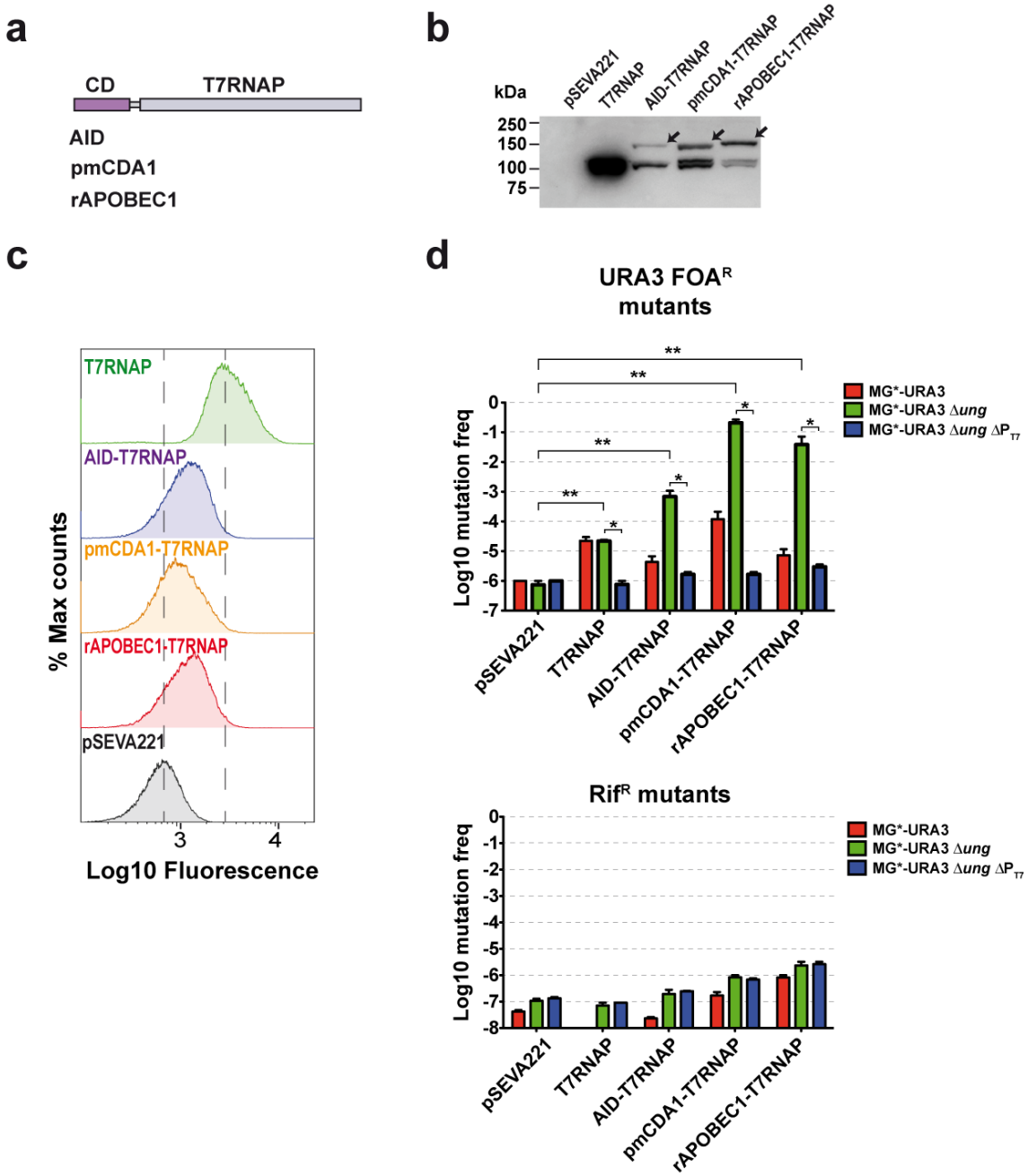

Figure 3. Expression and mutagenic activity of AID-T7RNAP, pmCDA1-T7RNAP and rAPOBEC1-T7RNAP. (**a**) Scheme of the different CDs fused to the T7 RNA polymerase by the linker $(G_3S)_7$. (**b**) Expression of the different fusions determined by Western blot analysis of the cell extracts from induced cultures of the strain MG*-URA3Δ*ung*. (**c**) Processivity of the fusions assessed by flow cytometry analysis to detect expression of GFP in the induced cultures. (**d**) Mutagenic activity in URA3 and *rpoB* of the different fusions using as hosts MG*-URA3, MG*-URA3Δ*ung* and MG*-URA3Δ*ung*ΔP_{T7}. The histograms represent the means and standard errors of at least three independent experiments (n≥3). The statistical analysis was done using Mann Whitney test. Asterisks indicate p-value < 0.05 (*) and p-value < 0.01 (**).

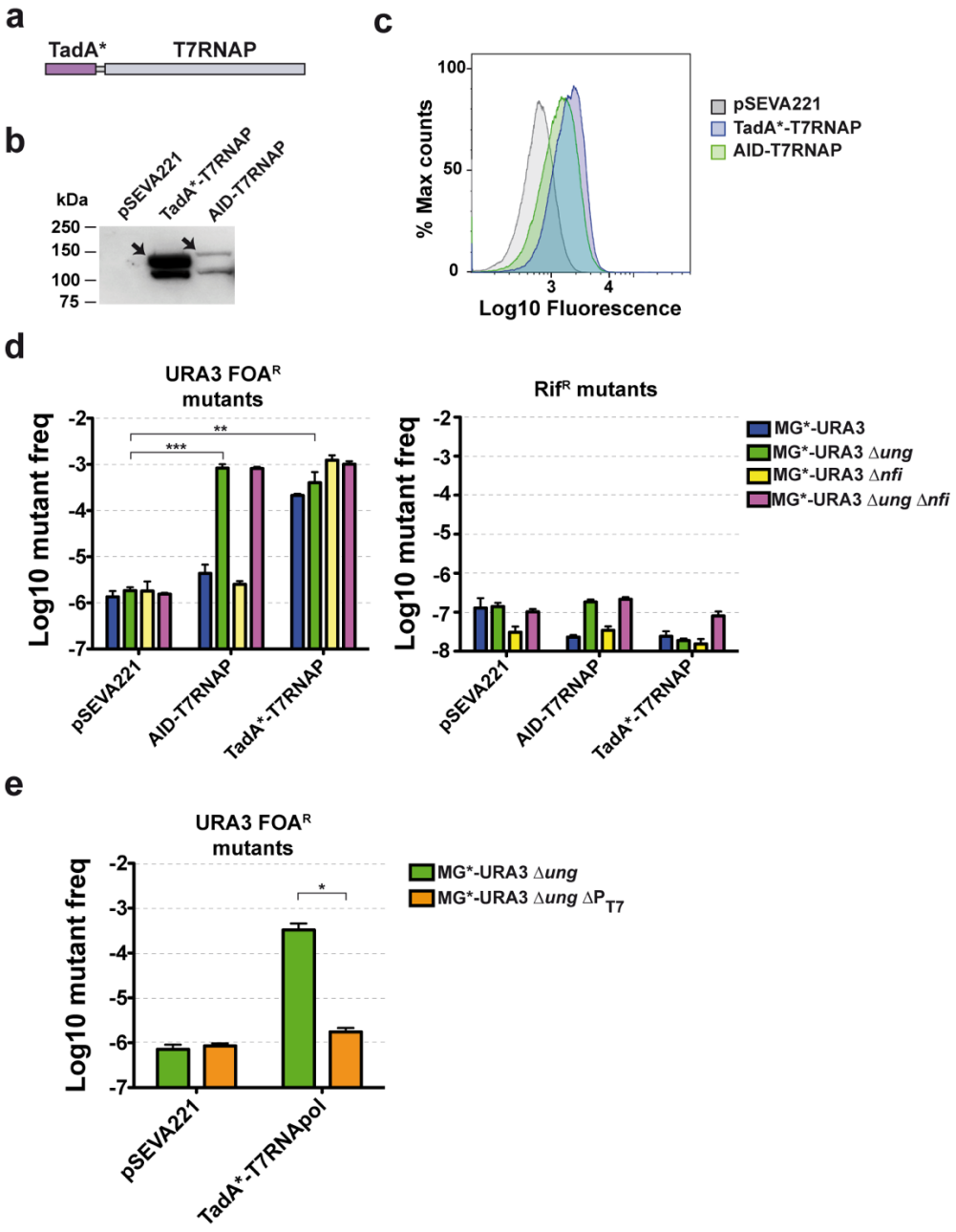

Figure 4. Expression and activity of TadA*-T7RNApol. (**a**) Scheme of the fusion TadA*-T7RNAP with the linker $(G_3S)_7$. (**b**) Expression of the fusion TadA*-T7RNAP in comparison to AID-T7RNAP determined by Western blot analysis of cell extracts from induced cultures. (**c**) Processivity of the fusions assessed by flow cytometry analysis to detect expression of GFP in the induced cultures. (**d**) Mutagenic activity of the AID- and TadA*-T7RNAP fusions in URA3 and *rpoB* using as hosts MG*-URA3, MG*-URA3Δ*ung*, MG*-URA3Δ*nfi* and MG*-URA3Δ*ung*Δ*nfi*. The histograms represent the means and standard errors of three independent experiments (n=3). (**e**) Mutagenesis frequency in URA3 when TadA*-T7RNAP is expressed in MG*-URA3Δ*ung* and MG*-URA3Δ*ung*ΔP_{T7} The histograms represent the means and standard errors of at least four independent experiments (n≥4). Asterisks indicate p-value < 0.05 (*), p-value < 0.01 (**) and p-value < 0.001 (***).

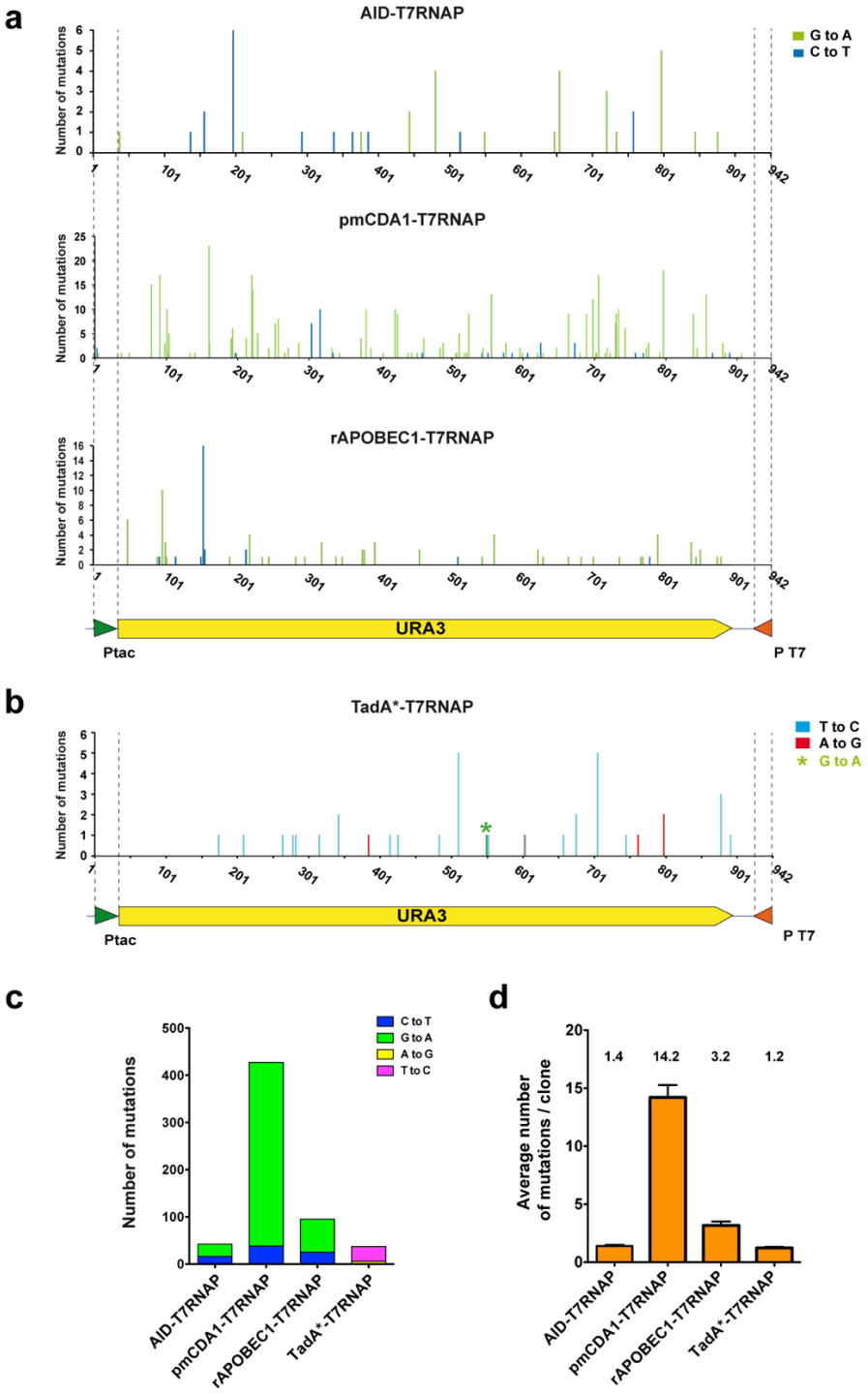

Figure 5. Characterization of URA3 mutations found in FOA^R colonies expressing BD-T7RNAP fusions. (**a**) Number of mutations per nucleotide identified in the URA3 locus from 30 FOA^R-colonies isolated from each MG*-URA3Δ*ung* strain expressing the indicated CD-T7RNAP fusions and (**b**) from MG*-URA3Δ*ung*Δ*nfi* strain expressing TadA*-T7RNAP fusion. The promoters Ptac and T7 are shown with arrow heads and delimited by dashed lines. The indicated base changes correspond to the coding sequence of URA3. Different base substitutions found are labeled with the color codes on the right. A single G to A transition found with TadA*-T7RNAP is labelled with an asterisk. (**c**) Total number of mutations for each BD-T7RNAP fusion indicating the base substitutions found. (**d**) Average number of mutations per clone found in the FOA^R-colonies analyzed for each of the indicated BD-T7RNAP fusions.

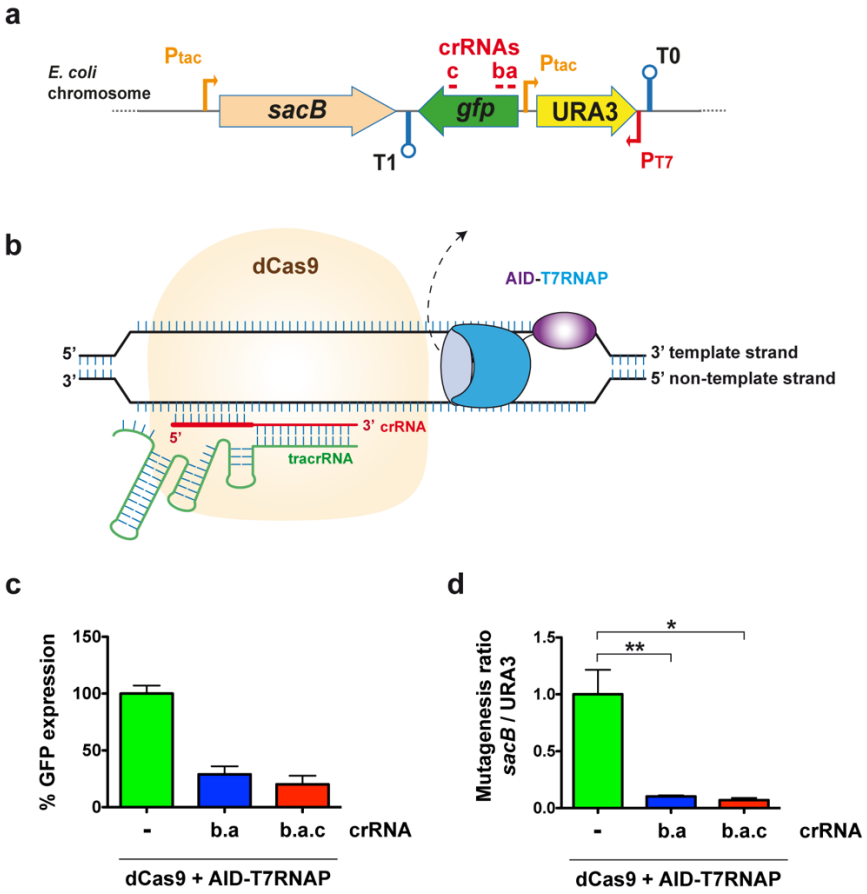

Figure 6. Blocking AID-T7RNAP elongation and mutagenic activity by dCas9 and crRNAs. **(a)** Scheme of the reporter cassette with *sacB* integrated in MG*-SacB-URA3Δ*ung*Δ*nfi* strain. Thin arrows indicate the tac (P_{tac}) and T7 (P_{T7}) promoters, lollipops indicate terminators T0 and T1, and red lines mark targeting sequences of the crRNAs a, b and c. **(b)** Representation of the dCas9 blocking activity showing one crRNA (in red) targeting the non-template strand relative to T7RNAP transcription. The mutagenic protein fusion is displaced from the transcription bubble (dashed arrow) by bound dCas9/crRNA. **(c)** Relative GFP levels measured by flow cytometry of bacteria from strain MG*-SacB-URA3Δ*ung*Δ*nfi* expressing AID-T7RNAP and dCas9 in the absence (-) or presence of crRNA arrays b.a and b.a.c. The histogram shows the percentages of mean flurorescence intensities (MFI) for each condition relative to the strain lacking crRNAs. Background GFP fluorescence signals from this strain with pdCas9 and the empty vector pSEVA221 are subtracted from all values. **(d)** Ratio of mutagenesis of *sacB* vs. URA3 in bacteria MG*-SacB-URA3Δ*ung*Δ*nfi* expressing AID-T7RNAP and dCas9 in the absence (-) or presence of crRNA arrays b.a and b.a.c. The ratio found in bacteria lacking crRNAs are considered 1. For (**c**) and (**d**), the histograms represent the relative means and standard errors from at least three independent experiments (n≥3). The statistical analysis was done using two-tailed Student t test. Asterisks indicate p-value < 0.05 (*), p-value < 0.01 (**).