

# I-Block: a simple *Escherichia coli*-based assay for studying sequence-specific DNA binding of proteins

Sarolta Szentes<sup>1,†</sup>, Nikolett Zsibrita<sup>1,2,†</sup>, Mihály Koncz<sup>1,2</sup>, Eszter Zsigmond<sup>1,2</sup>, Pál Salamon<sup>1</sup>, Zita Pletl<sup>1</sup> and Antal Kiss<sup>1,\*</sup>

<sup>1</sup>Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, 6726 Szeged, Hungary and <sup>2</sup>Doctoral School of Biology, Faculty of Science and Informatics, University of Szeged, 6726 Szeged, Hungary

Received July 03, 2019; Revised November 26, 2019; Editorial Decision January 02, 2020; Accepted January 05, 2020

## ABSTRACT

We have developed a simple method called I-Block assay, which can detect sequence-specific binding of proteins to DNA in *Escherichia coli*. The method works by detecting competition between the protein of interest and RNA polymerase for binding to overlapping target sites in a plasmid-borne *lacI* promoter variant. The assay utilizes two plasmids and an *E. coli* host strain, from which the gene of the Lac repressor (*lacI*) has been deleted. One of the plasmids carries the *lacI* gene with a unique NheI restriction site created in the *lacI* promoter. The potential recognition sequences of the tested protein are inserted into the NheI site. Introduction of the plasmids into the *E. coli*Δ*lacI* host represses the constitutive β-galactosidase synthesis of the host bacterium. If the studied protein expressed from a compatible plasmid binds to its target site in the *lacI* promoter, it will interfere with *lacI* transcription and lead to increased β-galactosidase activity. The method was tested with two zinc finger proteins, with the lambda phage cI857 repressor, and with CRISPR-dCas9 targeted to the *lacI* promoter. The I-Block assay was shown to work with standard liquid cultures, with cultures grown in microplate and with colonies on X-gal indicator plates.

## INTRODUCTION

Sequence-specific DNA–protein interactions are at the core of many biological processes (1). During characterization of DNA binding proteins it is often important to test whether

the protein can bind to a particular DNA sequence and to explore how alterations in the protein and/or in the DNA affect the binding strength and/or specificity (for reviews see (2–4)). There is a large variety of *in vitro* methods for studying binding affinity and specificity of DNA-binding proteins (5–15). A drawback of these methods is that they either require purified proteins and/or are technically demanding.

There are *Escherichia coli*-based techniques suitable to investigate sequence-specific DNA–protein interactions *in vivo* without the need of protein purification. These one- and two-hybrid methods were designed to determine the binding site of a protein or to select protein variants showing high affinity to a particular DNA sequence. The techniques involve fusing the protein of interest to a subunit of RNA polymerase and cloning the assumed binding site or an oligonucleotide library potentially containing the binding site closely upstream of the promoter of a selectable reporter gene, so that binding of the chimeric protein to the target site can enhance the transcription rate of the reporter gene (16–23). Because of the need to construct protein fusions, the one- and two-hybrid approaches are laborious.

In the course of our work with C<sub>2</sub>H<sub>2</sub> zinc finger proteins, we needed to test whether the proteins can bind to a specific DNA sequence in *E. coli*. Because the published *E. coli*-based methods appeared unnecessarily complicated for answering such relatively simple question, we decided to develop a simpler technique, which would be able to detect binding of a protein to its target site without the requirement of creating protein fusions. The assay reported here is based on the well-characterized regulatory mechanism of the *E. coli lac* operon and works by blocking transcription of the Lac repressor gene (*lacI*) resulting in elevated β-galactosidase activity. We describe construction of the assay and show its application with two zinc finger proteins, with

\*To whom correspondence should be addressed. Tel: +36 62 599 630; Fax: +36 62 433 506; Email: kiss.antal@brc.mta.hu

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Present addresses:

Sarolta Szentes, Aktogen Limited, Unit 1 Highlode Industrial Estate, Stocking Fen Road, Huntingdon, UK.

Mihály Koncz, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands.

Eszter Zsigmond, Institute of Genetics, Biological Research Centre of the Hungarian Academy of Sciences, 6726 Szeged, Hungary.

Pál Salamon, Institute of Chemistry, Faculty of Science, University of Pécs, Pécs, Hungary.

Zita Pletl, SeqOmics Biotechnology, Ltd., Vállalkozók útja 7, Mórahalom, Hungary.

the catalytically inactive dCas9 variant of the Cas9 protein and with the temperature-sensitive cI857  $\lambda$  phage repressor. We demonstrate application of the technique with cells grown in liquid cultures and with colonies grown on the surface of X-gal indicator plates. With reference to its mechanism of blocking production of the LacI repressor protein, we suggest to call the method I-Block assay.

## MATERIALS AND METHODS

### Bacterial strains and growth conditions

*Escherichia coli* DH10B ( $F^-$  *endA1 recA1 galE15 galK16 nupG rpsL  $\Delta$ lacX74 80dlacZ $\Delta$ M15 araD139  $\Delta$ (*ara leu*)7697 *mcrA  $\Delta$ (*mrr-hsdRMS-mcrBC*) relA1 spoT1  $\lambda$ -)* (24) was used as general cloning host. The *E. coli* strain CJ236 *F $\Delta$ (HindIII)::cat (Tra+ Pil+ CamR)/ung-1 relA1 dut-1 thi-1 spoT1 mcrA*, was used to generate uracil-containing single-stranded DNA for site-directed mutagenesis (25). The *E. coli* ER1821  $\Delta$ *lacI* strain was constructed in this work from *ER1821 F- glnV44 e14-(McrA-) rfbD1? relA1? endA1 spoT1? thi-1  $\Delta$ (mcrC-mrr)114::IS10* (26) by deleting the chromosomal *lacI* gene using the method described in (27).*

Bacteria were grown in LB liquid medium (28) in screw-cap flasks, in test tubes or in wells of microtiter plates at 30°C. Expression of zinc-finger proteins was induced by adding 0.1% L-arabinose (Sigma) to the medium. Ampicillin (Ap), kanamycin (Kn) and chloramphenicol (Cm) were used at 100, 50 and 25  $\mu$ g/ml concentration, respectively. Colonies producing  $\beta$ -galactosidase were screened on LB agar plates containing reduced concentration of yeast extract (2.5 g per liter) and 40  $\mu$ g/ml 5-bromo-4-chloro-3-indolyl  $\beta$ -D-galactoside (X-gal).

### Plasmids

Plasmids are listed in Supplementary Table S1. The plasmids pOK12 (29), pACYC184 (30), pUA66 (Kn<sup>R</sup>) (31) and pBAD24 (Ap<sup>R</sup>) (32) were used as cloning vectors. Construction of plasmids is described in Supplementary Data (Supplementary Figure S1).

The pLacI plasmids expressing the Lac repressor were named to indicate the vector (pOK-, pAC-, pUA- or pdUA-), the *lacI* promoter variant, the cloning site(s) created in the *lacI* promoter region and (in parentheses) the site of insertion with the inserted target sequence of the tested protein. The PvuI, NheI, MunI and EcoRI cloning sites are denoted by the letter P, N, M and E, respectively. For example, the plasmid pUA-lacI<sub>PN</sub>(N-6ZA) is based on the vector pUA66, contains the *lacI* gene with engineered unique PvuI and NheI sites in the *lacI* promoter, and carries the 6ZA zinc finger protein binding site inserted into the NheI site.

The plasmids pB6ZA and pB6ZB contain the genes of the 6ZA or the 6ZB zinc finger proteins (original names: 6-ZFP-A and 6-ZFP-B (33)) cloned in pBAD24. Plasmid pcI857, a derivative of pORTMAGE-2 (34) constitutively expresses the temperature-sensitive cI857 mutant of the lambda phage repressor. Plasmid pdCas9, a gift from Luciano Marraffini (Addgene plasmid # 46569; <http://n2t.net/addgene:46569>; RRID:Addgene\_46569) encodes the catalytically inactive D10A, H840A double mutant

of the Cas9 nuclease, expresses the tracrRNA and carries a minimal CRISPR array with BsaI sites for insertion of oligonucleotide duplexes defining the target sequence (35). The plasmids pdCas9-376, pdCas9-385, pdCas9-458, pdCas9-516, pdCas9-518 and pdCas9-520 were constructed by cloning the double-stranded oligonucleotides AK376-377, AK385-386, AK458-459, AK516-517, AK518-519 and AK520-521, respectively (Supplementary Table S2) between the BsaI sites of pdCas9. The numbers added as extensions in the names of the pdCas9- plasmids indicate the oligonucleotides corresponding to the top strand of the targeted sequence (Figure 2, Supplementary Tables S1 and 2).

### $\beta$ -galactosidase assay

To determine  $\beta$ -galactosidase activity, cells were grown in the presence or absence of arabinose overnight or for 4–8 h at 30°C. Cells were permeabilized and  $\beta$ -galactosidase activity was measured by the original (36) or by a simplified version (37) of the Miller protocol. In cells grown in microplate wells  $\beta$ -galactosidase activity was determined in Synergy<sup>TM</sup> HT or Synergy 2 Multi-Mode Microplate Readers (Bio-Tek). Details are described in Supplementary Data. Colonies producing  $\beta$ -galactosidase were identified on X-gal containing agar plates (28,36).

### Other methods

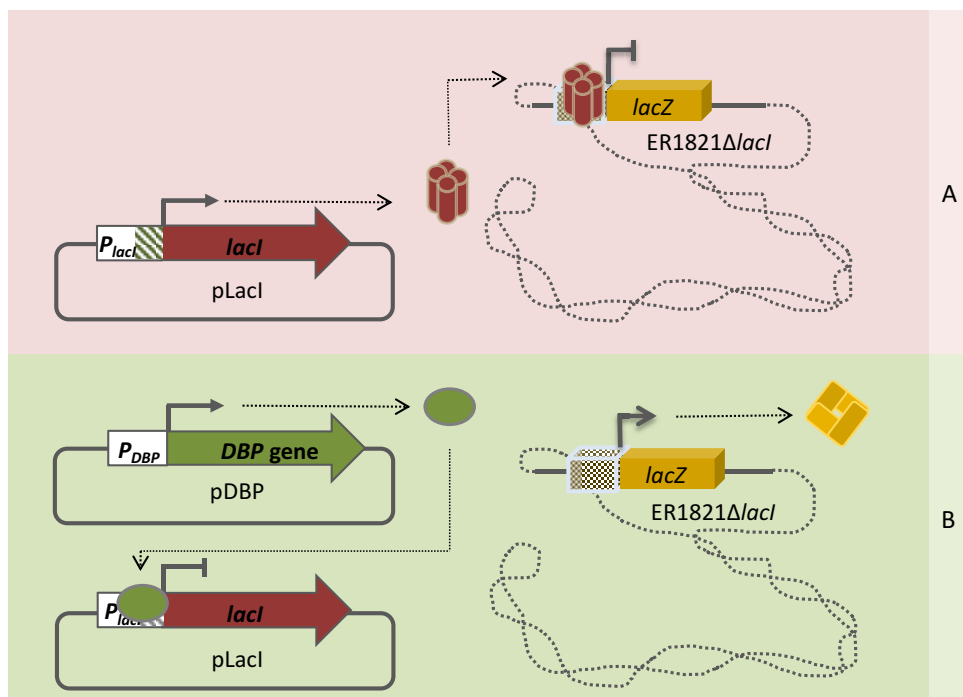
Cloning of DNA fragments was done by standard methods (28). Enzymes were purchased from Thermo Scientific or New England Biolabs. Site-directed mutagenesis was performed either by the Kunkel method (25) or by a variant of the QuikChange method as described in Supplementary Data. Conditions of measuring expression of the reporter green fluorescent protein fused to *lacI* promoter variants are described in Supplementary Data. Deoxyoligonucleotides were synthesized in the Nucleic Acid Synthesis Laboratory of this institute and are listed in Supplementary Table S2. All other chemicals were analytical grade commercial products. Bar diagrams were created and statistical analysis was performed with the GraphPad Prism program (GraphPad Software, Inc.).

## RESULTS

### Principle of the assay

The assay is based on the regulatory mechanism of the *E. coli lac* operon. In the wild-type arrangement the *lacI* gene encoding the Lac repressor is located on the bacterial chromosome very closely to the *lac* operon and is transcribed from the weak constitutive promoter *P<sub>lacI</sub>*. In the absence of inducer the Lac repressor binds to the *lac* operator and shuts off transcription of the *lacZYA* genes (38,39).

We envisaged a technique, which would detect binding of a protein to DNA by interference with transcription of the *E. coli lacI* gene resulting in the induction of  $\beta$ -galactosidase synthesis. The scheme of the assay is shown in Figure 1. An *E. coli* host from which the chromosomal copy of the *lacI* gene had been deleted contains two compatible plasmids. One of the plasmids (generic name: pLacI)



**Figure 1.** Principle of the I-Block assay. An *Escherichia coli* host strain from which the chromosomal copy of the *lacI* gene has been deleted, carries two compatible plasmids (generic names pLacI and pDBP). The plasmid pLacI encodes the Lac repressor (LacI), whereas pDBP carries the gene of the protein of interest (DNA Binding Protein, DBP). The assumed binding site of the tested DBP is inserted into the promoter of the *lacI* gene. (A) In the absence of DBP the *lacI* gene on pLacI is transcribed, and the synthesized Lac repressor blocks transcription of the chromosomal *lacZ* gene. (B) When the DBP is synthesized, it binds to its cognate site in the *lacI* promoter and interferes with *lacI* transcription leading to production of  $\beta$ -galactosidase.

expresses the *E. coli* Lac repressor, whereas the other plasmid (generic name: pDBP) expresses the DNA Binding Protein (DBP) of interest. Nucleotide sequence of the 5'-end of the wild-type *lacI* gene is shown in Figure 2. The assumed recognition sequence of the DNA binding protein is inserted into the *lacI* promoter in a position, where it does not severely impair *lacI* transcription. We expected that constitutive  $\beta$ -galactosidase synthesis of the *lacI* deletion host strain would be repressed by the Lac repressor synthesized from pLacI, but binding of the tested protein to its target site on pLacI would interfere with *lacI* transcription and lead to induction of  $\beta$ -galactosidase production (Figure 1).

### Construction of the assay system and pilot experiments

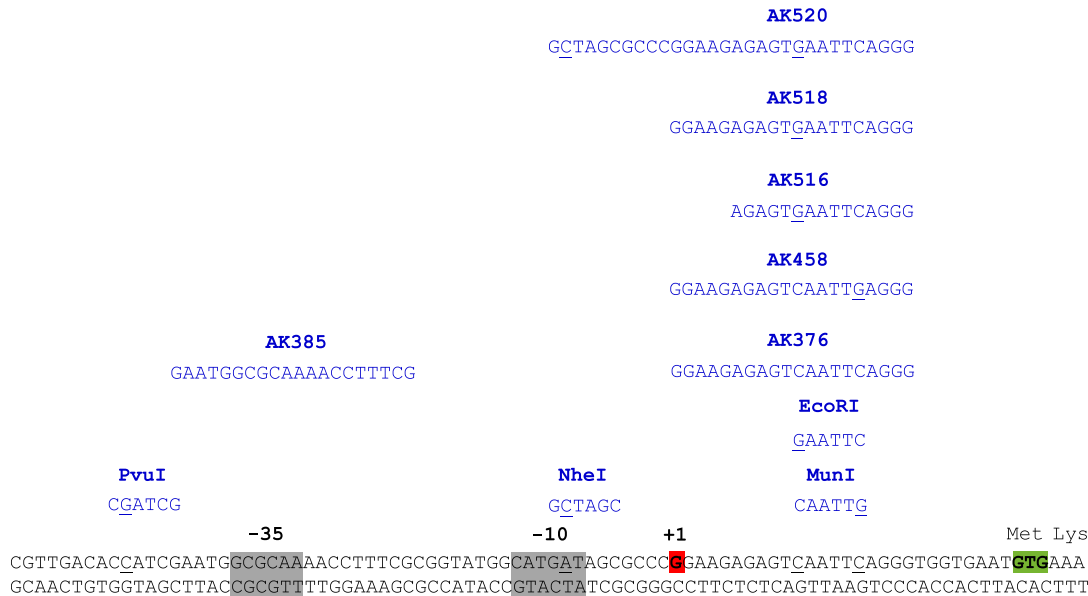
The LacI-deficient *E. coli* host bacterium was created by deleting the *lacI* gene from *E. coli* ER1821. The 1189 bp deletion removed the genomic segment between coordinates 366349–367537 of *E. coli* MG1655. The deletion strain *E. coli* ER1821 $\Delta$ *lacI* produces  $\beta$ -galactosidase constitutively.

Conditions of the assay were first tested with two C<sub>2</sub>H<sub>2</sub> zinc finger proteins (6-ZFP-A and 6-ZFP-B) whose binding specificity was known from *in vitro* experiments. The 6-ZFP-A protein recognizes the sequence 5'-GCC-GGG-GCT-GGG-GGA-GGG, whereas 6-ZFP-B recognizes 5'-GGA-GTT-GGG-GGA-GTG-AGT (33). In this paper, for brevity, 6-ZFP-A and 6-ZFP-B will be named 6ZA and 6ZB, respectively. To express the two zinc finger proteins in inducible fashion, the 6ZA and 6ZB genes were cloned

in the expression vector pBAD24 to obtain pB6ZA and pB6ZB, respectively. In these plasmids transcription of the 6ZA and 6ZB genes is controlled by the *araBAD* promoter and can be induced with arabinose.

Details of constructing and testing the different pLacI plasmids are described in Supplementary Data. The first pLacI versions were based on the medium copy number vectors pOK12 (29) and pACYC184 (30), contained the strong *lacI*<sup>Q</sup> promoter (40–42) and had a MunI cloning site created in the 5'-untranslated region between the *lacI*<sup>Q</sup> promoter and the GTG start codon (Figure 2). When the plasmids pAC-lacI<sup>Q</sup><sub>M</sub>-(M-6ZA) or pAC-lacI<sup>Q</sup><sub>M</sub>-(M-6ZB) [i.e. pAC-lacI<sup>Q</sup><sub>M</sub> containing the 6ZA or the 6ZB binding site inserted into the MunI site; for plasmid nomenclature see 'Materials and Methods' section] were introduced into *E. coli* ER1821 $\Delta$ *lacI*, the high  $\beta$ -galactosidase activity of the host was repressed, but the expected increase of  $\beta$ -galactosidase activity upon induction of the synthesis of the cognate zinc finger protein was not observed. We hypothesized that the failure to detect binding by the two zinc finger proteins had two possible reasons. One of the possibilities was that the multiple plasmid-borne copies of the *lacI*<sup>Q</sup> gene and the elevated transcription rate characterizing the *lacI*<sup>Q</sup> promoter (40–42) resulted in Lac repressor concentrations, which were too high for achieving even partial induction of the *lac* operon. The second possibility was that proteins bound to the 5'-untranslated region cannot efficiently block the transcribing RNA polymerase.

To find the right conditions for the assay, several pLacI plasmids were constructed (Supplementary Data). Design



**Figure 2.** Nucleotide sequence of the *Escherichia coli lacI* promoter region (41,44–45). The –35 and –10 motifs are highlighted in gray, the transcriptional start site in red and the GTG start codon in green. The PvuI, NheI, MunI and EcoRI sites created by site directed mutagenesis are shown above the sequence. The mutated nucleotides are underlined. The AK376, AK385, AK458, AK516, AK518 and AK520 oligonucleotides shown above the sequence define target sites for CRISPR-dCas9.

of the plasmids was guided by two goals: lowering the concentration of the Lac repressor and finding a site within the *lacI* promoter, where binding proteins can efficiently interfere with *lacI* transcription. The pACYC184 vector was replaced with the pSC101-based low copy number plasmid vector pUA66 (31) and the strong *lacI*<sup>Q</sup> promoter with the weaker (41) wild-type *lacI* promoter. Two unique restriction sites (PvuI and NheI) were created in the wild-type *lacI* promoter by site-directed mutagenesis. The PvuI site is located a few bp upstream of the –35 conserved motif, whereas the NheI site partially overlaps the –10 conserved element. Creation of the NheI site introduced an A to C change in the conserved –10 element (Figure 2). The *lacI* gene variants carrying the MunI, PvuI or the NheI site were designated *lacI*<sub>M</sub>, *lacI*<sub>P</sub> and *lacI*<sub>N</sub>, respectively. The changes described above led to partial success. The pUA66-based plasmids pUA-lacI<sub>N</sub>-(N-6ZA) and pUA-lacI<sub>P</sub>-(N-6ZA) containing the 6ZA target sequence inserted into the NheI cloning site repressed, and expression of the 6ZA protein restored 50–60% of the original β-galactosidase activity (Supplementary Data and Figure S2C and E). However, expression of the 6ZB protein in the cell containing the cognate pUA-lacI<sub>N</sub>-(N-6ZB) plasmid did not increase β-galactosidase activity (not shown).

The assay system was also tested with the non-zinc finger DNA binding protein dCas9. Targeting the constitutively expressed dCas9 to the sequence between positions –41 and –22 (AK385, Figure 2) did not induce β-galactosidase production (Supplementary Figure S3A–E), whereas targeting to the sequence between positions +1 to +20 (AK376, Figure 2) led to a significant increase of β-galactosidase activity, but only when the *lacI* promoter contained the NheI site (Supplementary Figure S3I and J). Because in the ex-

periments with dCas9 no foreign sequence was in the NheI site, we concluded that the ability to induce β-galactosidase synthesis with the *lacI*<sub>N</sub> and *lacI*<sub>P</sub> variants (Supplementary Figure S3I and J) was due to the A to C mutation associated with the creation of the NheI site in the *lacI* promoter (Figure 2). A possible explanation was that the mutation weakened the *lacI* promoter and lowered the Lac repressor concentration to a range, where the competing effect of the 6ZA and dCas9 proteins became detectable. A comparison of the promoters fused to the green fluorescent protein (gfp) reporter gene showed that the *lacI*<sub>N</sub> promoter was indeed weaker than the wild-type *lacI* promoter (Supplementary Figure S4).

### Use of the binding assay with standard liquid cultures, microtiter plates and colonies

The data described above identified two essential conditions for the functional assay: low copy number pLacI plasmid and the weak *lacI*<sub>N</sub> promoter. Moreover, the results of the initial experiments suggested that the NheI site overlapping the –10 element could be suitable for inserting the tested target sequences.

Because efficient interference with *lacI*<sub>N</sub> transcription was also observed with dCas9 targeted to the first twenty transcribed nucleotides (Supplementary Figure S3I and J), and insertion of different foreign (target) sequences into the 5'-untranslated region was thought to cause less variation in the expression of the *lacI*<sub>N</sub> gene than insertion of these sequences immediately downstream of the conserved –10 element, we assumed that the MunI site could be better for the assay than the NheI site. Use of the MunI site in combination with the low copy number vector pUA66 and the



weak *lacI<sub>N</sub>* promoter could not be tested because pUA66 contained three *MunI* sites, which made the use of *MunI* as cloning enzyme impractical. To obtain a pUA66-based pLacI plasmid with a unique cloning site in the *lacI* 5'-untranslated region, an *EcoRI* site was created at the position exactly corresponding to the *MunI* site (Figure 2). To optimize the assay conditions, a family of pLacI plasmids containing different combinations of the *PvuI*, *NheI* and *EcoRI* cloning sites with the inserted 6ZA and/or 6ZB binding sites were created as described in Supplementary Data. The vector part of these plasmids (denoted by the prefix pdUA-) slightly differed from pUA66 in that it carried a deletion removing the gene of the green fluorescent protein and an *EcoRI* site originally present in pUA66. These plasmids were used in the rest of the work.

To simplify the assay and to make it more reliable, two changes were introduced. In the experiments described above addition of the inducer (and carbon source) arabinose increased the growth rate and led to substantially higher cell densities than those of the uninduced cultures making comparison of  $\beta$ -galactosidase activities difficult. To maintain comparable growth rates, uninduced cultures were supplemented with 0.1% glucose at the same time when arabinose was added to the induced cultures. Another difference compared to previous experiments was that the traditional Miller protocol, which uses chloroform and sodium dodecyl sulfate for permeabilization of the cells before measuring  $\beta$ -galactosidase activity (36), was replaced with a simpler method employing hexadecyltrimethylammonium bromide (CTAB) and sodium deoxycholate for permeabilization (37).

Measurements carried out under the new conditions confirmed previous results (Supplementary Figure S2C and E) with pUA-lacI<sub>N</sub> carrying the 6ZA binding sequence in the *NheI* site, but could also demonstrate binding of the 6ZB zinc finger protein (Figure 3A–D), which was not detectable under the conditions of the classical Miller assay (see above). In general, in our hands the simplified version of the Miller assay (37) gave more reproducible results than the original protocol (36). The observed difference between the  $\beta$ -galactosidase activities elicited by 6ZA and 6ZB presumably indicates the difference between the binding strengths of the two zinc finger proteins to their respective target sites. Induction of  $\beta$ -galactosidase synthesis by 6ZA binding occurred only with the cognate target site [pdUA-lacI<sub>N</sub>-(N-6ZA)], and no increase of  $\beta$ -galactosidase activity was observed with pdUA-lacI<sub>N</sub>(N-6ZB), which carried the 6ZB recognition sequence (Figure 3E) showing that the protein–DNA interaction detected by the assay was sequence-specific.

The new conditions of the binding assay were validated with dCas9 and with the temperature-sensitive cI857 mutant of the  $\lambda$  phage repressor. Confirming previous results (Supplementary Figure S3I and J) obtained with the original Miller method, targeting dCas9 to the 5'-untranslated region of the *lacI<sub>N</sub>* gene containing the *NheI* site in the –10 element of the promoter resulted in high  $\beta$ -galactosidase activity (Figure 3F and Supplementary Figure S5). In cells carrying the OR1 and OR2 lambda phage operator inserted into the *NheI* site of pdUA-lacI<sub>N</sub> expression of the temperature-sensitive cI857 repressor resulted in increased

$\beta$ -galactosidase activity at the permissive temperature 30°C, but not at 42°C (Figure 3G).

With the goal to adapt the assay to processing multiple samples, cultures were grown and assayed in microplate wells of a BioTek Synergy instrument as described in Supplementary Data. The ability to grow several samples in parallel allowed easy comparison of different pdUA-lacI-based plasmids containing the 6ZA or the 6ZB recognition sequence inserted in the *PvuI*, *NheI* or the *EcoRI* site. In accordance with previous results, the assay worked only in cases when the recognition sequence of the tested protein was in the *NheI* site, and did not work when it was in the *PvuI* or in the *EcoRI* site (Figure 4). Plasmid combinations, which were tested in cultures grown in flasks (Figure 3) as well as in microplates (Figure 4), showed similar  $\beta$ -galactosidase activities.

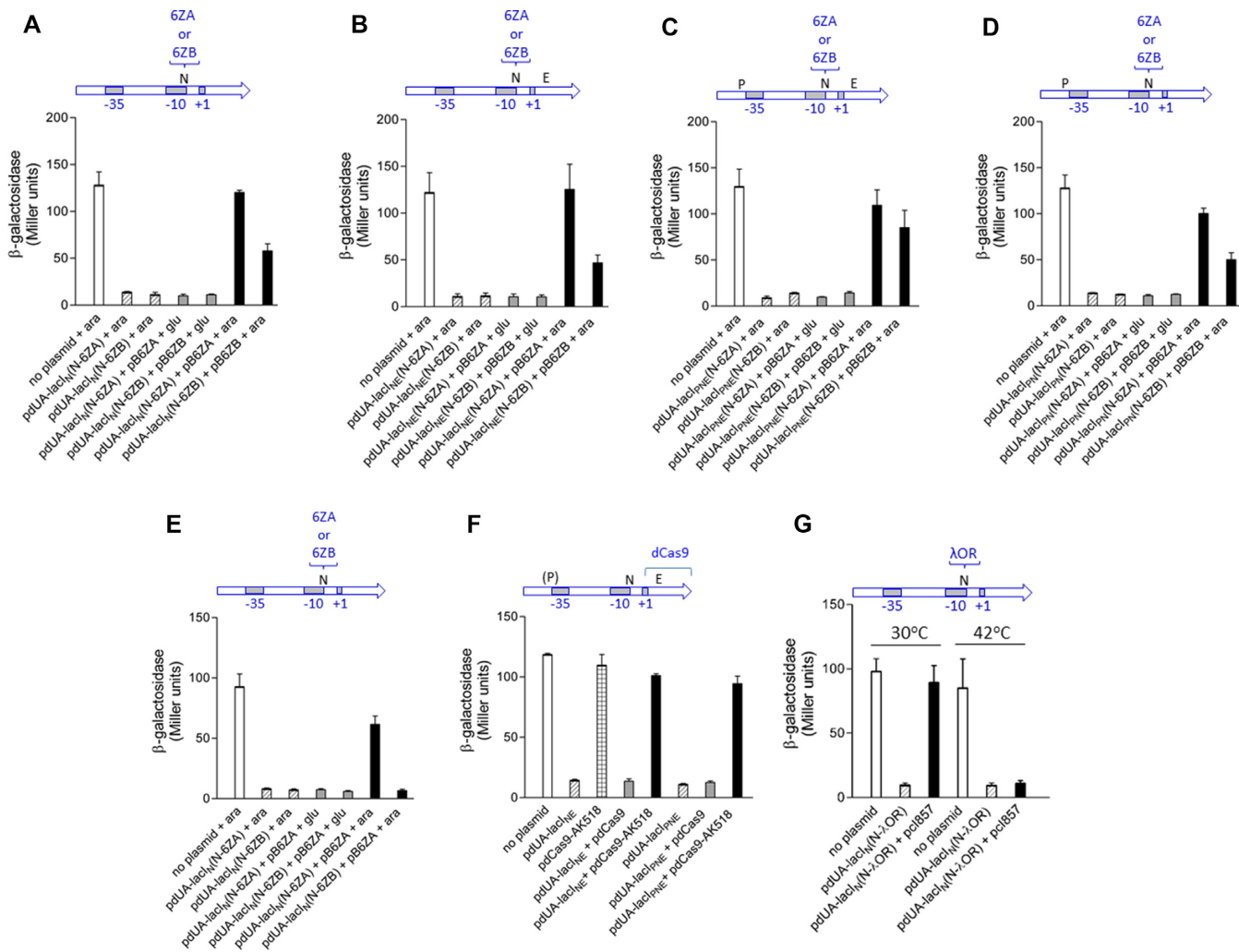
Plasmids with the engineered *PvuI* and *EcoRI* cloning sites and with the 6ZA recognition sequence inserted in the *EcoRI* site showed a phenotype that differed from that of all other pdUA-lacI plasmids tested: pdUA-lacI<sub>PE</sub>(E-6ZA) repressed  $\beta$ -galactosidase synthesis of the host bacterium only partially, and pdUA-lacI<sub>PNE</sub>(E-6ZA) did not repress it at all (Figure 4) suggesting that the combined effect of the changes introduced with the *PvuI*, *NheI* sites and with the 6ZA sequence was a substantial decrease of the *lacI* transcription rate. The detrimental effect of the mutations associated with the *PvuI* and *NheI* sites is not surprising because of their position relative to the conserved elements of the *lacI* promoter. The mechanism of the effect of the 6ZA target site inserted into the nonconserved sequence of the first transcribed nucleotides is less clear, a possible explanation can be that it inhibits promoter escape by the RNA polymerase (43).

For many applications screening a large number of clones is essential. To test whether the method described above can be used to screen colonies on solid medium, ER1821 $\Delta$ *lacI* cells harboring either pdUA-lacI<sub>N</sub>(N-6ZA) plus pB6ZA or pdUA-lacI<sub>N</sub>(N-6ZB) plus pB6ZA were spread on the surface of LB/Ap/Kn/arabinose/X-gal agar plates and incubated overnight at 37°C. Colonies containing the pdUA-lacI<sub>N</sub> plasmid with the 6ZA recognition sequence in the *NheI* site turned blue, whereas those with the 6ZB recognition site remained white (Figures 5 and Supplementary Figure S6).

## DISCUSSION

In this work we have converted the *E. coli lac* operon into a new genetic circuit suitable to test whether a protein can bind to a particular DNA sequence. The technique utilizes the regulatory mechanism of the *lac* operon and works by detecting competition between the studied protein and *E. coli* RNA polymerase for binding to overlapping target sites in the plasmid-borne *lacI<sub>N</sub>* promoter (Figure 1). With reference to its working mechanism we suggest to call the method I-Block assay. The method was tested with two zinc finger proteins recognizing 18 bp sequences, with CRISPR-dCas9 targeted to the *lacI* promoter and with the cI857 temperature-sensitive  $\lambda$  phage repressor.

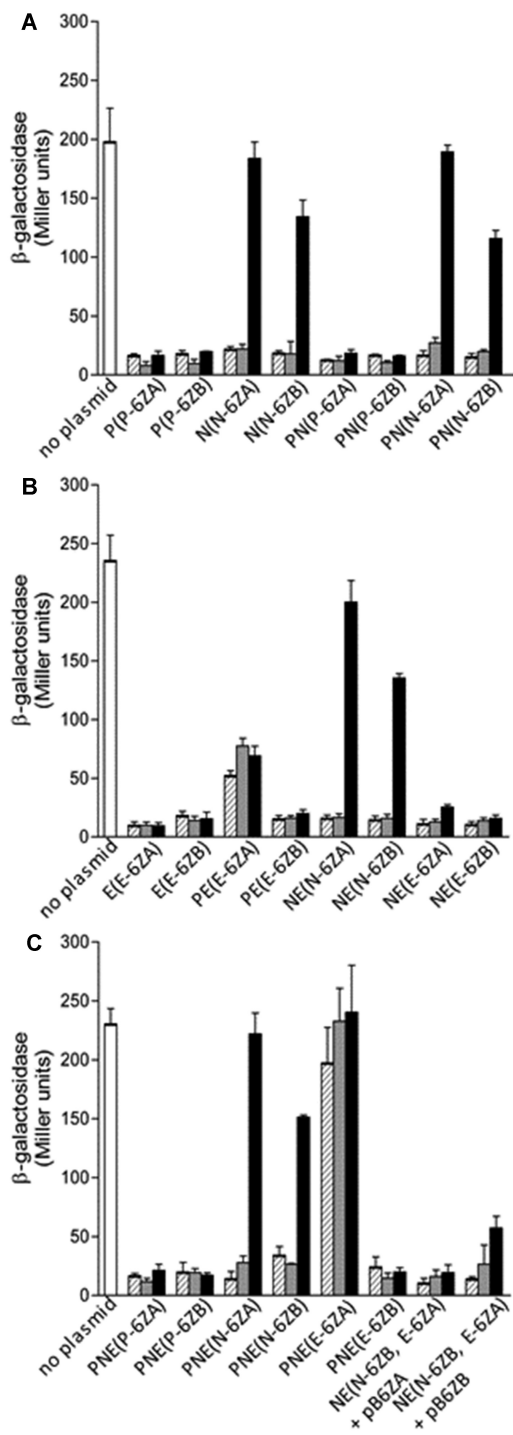
Establishment of the method turned out to be more difficult than we had originally expected (see Supplementary



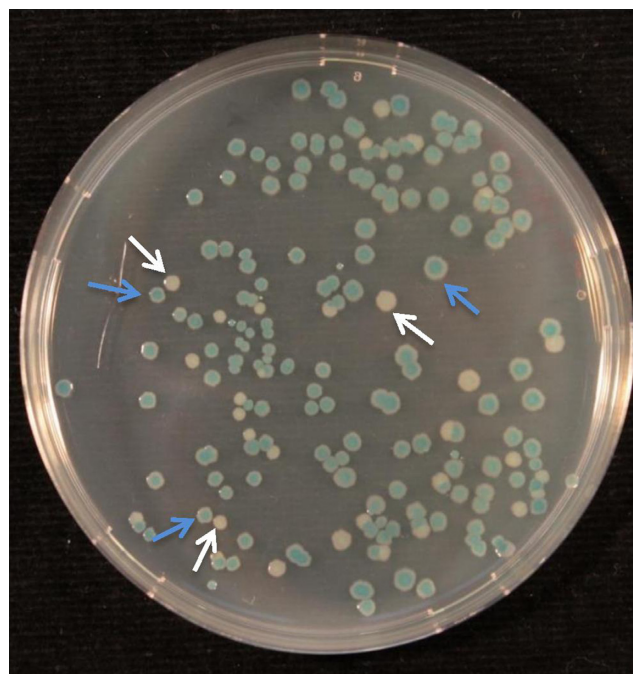
**Figure 3.** Detection of sequence-specific DNA–protein binding in *Escherichia coli* by the I-Block assay using samples of standard liquid cultures. The *E. coli* ER1821Δ*lacI* host contained the indicated plasmids. β-galactosidase activities were determined from overnight cultures using the simplified Miller assay and are the average of three independent experiments. Error bars indicate standard error of the mean values (for clarity only above the bars). The p*l*UA- plasmids carry the *lacI* gene with PvuI, NheI or EcoRI acceptor site(s) engineered in the *lacI* promoter region (see Figure 2). The presence of the PvuI, NheI or EcoRI acceptor sites is indicated by the letters P, N and E, respectively in lower index. (A–E) In the names of the plasmids the extensions in parentheses indicate the presence of the 6ZA or 6ZB binding site cloned in the NheI acceptor site. The plasmids pB6ZA and pB6ZB express, upon arabinose induction, the 6ZA and the 6ZB zinc finger protein, respectively. ara, arabinose; glu, glucose added to the culture. (F) The p*d*Cas9-AK518 plasmid constitutively expresses the d*C*as9 protein and the guide RNA determining the d*C*as9 binding site (AK518, Figure 2). (G) λOR, λ phage OR1 and OR2 operator cloned in the NheI acceptor site. The p*c*I857 plasmid constitutively expresses the temperature-sensitive λ cI857 repressor. The schemes above the graphs show the *lacI* promoter variants with the approximate position of the 6ZA, 6ZB, d*C*as9 or λ repressor binding sites. P values derived from one-way ANOVA and subsequent Tukey’s multiple comparison tests are shown in Supplementary Data.

Data). Three conditions, the right concentration of the Lac repressor, the position of the inserted target site and the simplified method of measuring β-galactosidase activity proved crucial for the functional assay. The optimal Lac repressor concentration was found to be dependent on three factors (low copy number p*L*acI plasmid, wild-type *lacI* gene instead of *lacI<sup>Q</sup>* and a mutation in the conserved –10 element of the *lacI* promoter). The relatively low LacI concentration determined by these conditions was sufficiently high for efficient repression of β-galactosidase synthesis in the absence of a tested DNA binding protein, but suitably low to allow detection of small changes in the rate of *lacI* transcription as a result of the tested protein’s interference.

At the beginning of the work we assumed that binding of a protein between the promoter and the start codon of the *lacI* gene would efficiently block *lacI* transcription, thus the 5′-untranslated region would be suitable for inserting the target sites of the tested proteins. However, when the target site was inserted into the MunI or into the EcoRI site located in the 5′-untranslated sequence, binding of the two zinc finger proteins had no detectable effect on *lacI* transcription. Similarly, the PvuI site located a few bp upstream of the –35 element, proved unsuitable as acceptor site for target sequences. Zinc finger protein binding was only detectable in cases when the target sequence was inserted into the artificially created NheI site, which partially overlaps the



**Figure 4.** (A–C) Detection of sequence-specific DNA–protein binding by the I-Block assay using cultures grown in wells of microtiter plates. The *Escherichia coli* ER1821  $\Delta lacI$  host contained the indicated pdUA-lacI-based plasmids. For nomenclature of the plasmids and for experimental conditions see the legend of Figure 3. Empty bars, no plasmid, culture grown in the presence of arabinose; gray bars, pdUA-lacI variant, grown in the presence of arabinose; striped bars, pdUA-lacI variant plus the plasmid (pB6ZA or pB6ZB) expressing the cognate zinc finger protein, grown in the presence of glucose; black bars, pdUA-lacI variant plus the plasmid (pB6ZA or pB6ZB) expressing the cognate zinc finger protein, grown in the presence of arabinose. *P*-values derived from one-way ANOVA and subsequent Tukey’s multiple comparison tests are shown in Supplementary Data.



**Figure 5.** Detection of sequence-specific DNA–protein binding by the I-Block assay in colonies grown on X-gal indicator plates. Cells of *Escherichia coli* ER1821  $\Delta lacI$  harboring pdUA-lacI<sub>N</sub>(N-6ZA) + pB6ZA or pdUA-lacI<sub>N</sub>(N-6ZB) + pB6ZA were mixed in 4 to 1 ratio, and plated on the surface of X-Gal plates containing arabinose to induce expression of the 6ZA zinc finger protein. White arrows indicate colonies containing plasmids with the non-cognate 6ZB binding site, whereas blue arrows indicate colonies containing plasmids with the cognate 6ZA target site.

–10 conserved element. An important message of our data is that in *E. coli* a DNA-binding protein can more efficiently interfere with transcription by competing with binding of RNA polymerase to the promoter than by obstructing the elongating enzyme.

Creation of the NheI site, which meant an A to C change in the –10 element, turned out to have two-fold advantage for the assay. Firstly, it created an acceptor site at a position, where a protein can efficiently interfere with RNA polymerase binding to the *lacI* promoter. The second effect of the A to C mutation was that the assay system became sensitive to subtle changes in *lacI* transcription. The NheI site overlaps with the –10 box only partially, thus cloning of foreign sequences into the NheI site does not lead to changes in the conserved elements of the *lacI*<sub>N</sub> promoter, which makes comparison of different target sites more reliable. Of the pLacI plasmid variants tested, pdUA-lacI<sub>N</sub> appears to be the best for the assay.

The I-Block assay was shown to work with standard liquid cultures, with cells grown in microplate wells as well as with colonies on X-gal indicator plates. A great advantage of the method is simplicity, it only requires basic techniques of molecular biology. In contrast to published approaches designed to detect specific DNA–protein interactions in *E. coli* (16–23), the I-Block assay does not require construction of protein fusions, which makes it easier to use and avoids potential influence of the fusion partner on the structure and function of the tested protein.



We wish to develop the I-Block assay into a semi-quantitative method suitable to estimate the DNA-binding affinities of proteins in *E. coli*. One of the potential problems hindering quantitative of the assay is the variability inherent to the *lacI<sub>N</sub>* gene and the gene of the tested protein residing on different replicons. We have preliminary results showing that the method can work in an arrangement where the two genes are on the same plasmid (Supplementary Figure S7). For quantitative data on binding strength, it will be essential to measure the intracellular concentration of the studied protein, which can be, due to differences in transcription, translation and protein stability, very different for different proteins. For the envisaged version of the assay we plan to explore the possibilities of using fluorescent tags fused to the studied protein.

Here we demonstrated that the I-Block assay can be used to test whether a protein can bind to a particular DNA sequence in *E. coli*. We envisage application of the method for comparing affinities of different proteins to the same DNA sequence, or for comparing affinities of a protein to different potential target sites. The I-Block assay, in its present form, does not match the capacity of the powerful one-hybrid and two-hybrid systems employing direct selection (16–23) to determine the binding site of a protein or to find the best binding protein variant from large sequence libraries, but it can potentially be developed into a direct selection technique. In the envisaged high throughput version of the assay, to avoid using the carbon source arabinose, the studied protein would be expressed from a plasmid constitutively. To find the target site of the studied protein, a library of randomized oligonucleotides would be cloned into the *NheI* site of *pdUA-lacI<sub>N</sub>*. Double transformants harboring *pdUA-lacI<sub>N</sub>* with the inserted oligonucleotide library and the plasmid constitutively expressing the protein of interest would be spread on the surface of minimal agar plates containing the non-inducing  $\beta$ -galactosidase substrate phenyl- $\beta$ -D-galactoside (P-Gal) as the only carbon source (36). Under such conditions only cells producing  $\beta$ -galactosidase are expected to form colonies.

DNA-binding proteins differ widely in their binding modes, and can have, for example, enthalpy-driven or entropy-driven binding mechanism (for references see (1)). Here we showed that the I-Block assay worked with proteins representing three different classes of DNA-binding proteins (zinc finger, Cas9 and lambda phage cI repressor), which suggests that the technique will be generally applicable.

The *E. coli lac* operon has been a source of inspiration for concepts and countless techniques for 60 years. We believe that the I-Block assay described here will be a useful addition to the  $\beta$ -galactosidase based toolbox of molecular biology.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Marianne G. Rots for the plasmids carrying the genes of the 6ZA and 6ZB proteins, Luciano Marraffini

for *pdCas9* (Addgene plasmid # 46569), Ákos Nyerges for *pORTMAGE2* (Addgene plasmid #72677), Edit Tímár for the plasmid *pSEVA3513* and advice on the construction of the *E. coli* host with *lacI* deletion, Krystyna Ślaska-Kiss for experiments at the beginning of this project and Pál Venetianer for suggestions on the manuscript.

## FUNDING

National Research, Development and Innovation Office [NN109787]; Ministry of Finance of Hungary [GINOP-2.3.2-15-2016-00001]; European Union [H2020-MSCA-ITN-2014, 642691 to M.K.]; Erasmus+K103 Programme of the European Union (to P.S). Funding for open access charge: Ministry of Finance of Hungary [GINOP-2.3.2-15-2016-00001].

*Conflict of interest statement.* None declared.

## REFERENCES

- Stormo, G.D. (2013) *Introduction to Protein-DNA Interactions. Structure, Thermodynamics, and Bioinformatics*. Cold Spring Harbor Laboratory Press, NY.
- Xie, Z., Hu, S., Qian, J., Blackshaw, S. and Zhu, H. (2011) Systematic characterization of protein-DNA interactions. *Cell. Mol. Life Sci.*, **68**, 1657–1668.
- Dey, B., Thukral, S., Krishnan, S., Chakrobarty, M., Gupta, S., Manghani, C. and Rani, V. (2012) DNA-protein interactions: methods for detection and analysis. *Mol. Cell. Biochem.*, **365**, 279–299.
- Stormo, G.D. and Zhao, Y. (2010) Determining the specificity of protein-DNA interactions. *Nat. Rev. Genet.*, **11**, 751–760.
- Jones, O.W. and Berg, P. (1966) Studies on the binding of RNA polymerase to polynucleotides. *J. Mol. Biol.*, **22**, 199–209.
- Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.*, **9**, 3047–3060.
- Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
- Liu, J. and Stormo, G.D. (2005) Combining SELEX with quantitative assays to rapidly obtain accurate models of protein-DNA interactions. *Nucleic Acids Res.*, **33**, e141.
- Gustafsdottir, S.M., Schlingemann, J., Rada-Iglesias, A., Schallmeiner, E., Kamali-Moghaddam, M., Wadelius, C. and Landegren, U. (2007) *In vitro* analysis of DNA-protein interactions by proximity ligation. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 3067–3072.
- Vickers, T.A. and Croke, S.T. (2016) Development of a quantitative BRET affinity assay for nucleic acid-protein interactions. *PLoS One*, **11**, e0161930.
- Hampshire, A.J., Rusling, D.A., Broughton-Head, V.J. and Fox, K.R. (2007) Footprinting: A method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods*, **42**, 128–140.
- Brand, L.H., Henneges, C., Schussler, A., Kolukisaoglu, H.U., Koch, G., Wallmeroth, N., Hecker, A., Thurov, K., Zell, A., Harter, K. et al. (2013) Screening for protein-DNA interactions by automatable DNA-protein interaction ELISA. *PLoS One*, **8**, e75177.
- Fischer, S.M., Boser, A., Hirsch, J.P. and Wanke, D. (2016) Quantitative analysis of protein-DNA interaction by qDPI-ELISA. *Methods Mol. Biol.*, **1482**, 49–66.
- Bulyk, M.L., Huang, X., Choo, Y. and Church, G.M. (2001) Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 7158–7163.
- Ho, S.-W., Jona, G., Chen, C.T.L., Johnston, M. and Snyder, M. (2006) Linking DNA-binding proteins to their recognition sequences by using protein microarrays. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 9940–9945.



16. Meng, X., Brodsky, M.H. and Wolfe, S.A. (2005) A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat. Biotechnol.*, **23**, 988–994.
17. Durai, S., Bosley, A., Abulencia, A.B., Chandrasegaran, S. and Ostermeier, M. (2006) A bacterial one-hybrid selection system for interrogating zinc finger-DNA interactions. *Comb. Chem. High Throughput Screen.*, **9**, 301–311.
18. Meng, X. and Wolfe, S.A. (2006) Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid method. *Nat. Protoc.*, **1**, 30–45.
19. Meng, X., Thibodeau-Beganny, S., Jiang, T., Joung, J.K. and Wolfe, S.A. (2007) Profiling the DNA-binding specificities of engineered Cys2His2 zinc finger domains using a rapid cell-based method. *Nucleic Acids Res.*, **35**, e81.
20. Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H. and Wolfe, S.A. (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell*, **133**, 1277–1289.
21. Christensen, R.G., Gupta, A., Zuo, Z., Schriefer, L.A., Wolfe, S.A. and Stormo, G.D. (2011) A modified bacterial one-hybrid system yields improved quantitative models of transcription factor specificity. *Nucleic Acids Res.*, **39**, e83.
22. Joung, J.K., Ramm, E.I. and Pabo, C.O. (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 7382–7387.
23. Hurt, J.A., Thibodeau, S.A., Hirsh, A.S., Pabo, C.O. and Joung, J.K. (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12271–12276.
24. Durfee, T., Nelson, R., Baldwin, S., Plunkett, G. 3rd, Burland, V., Mau, B., Petrosino, J.F., Qin, X., Muzny, D.M., Ayele, M. *et al.* (2008) The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J. Bacteriol.*, **190**, 2597–2606.
25. Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.*, **154**, 367–382.
26. Waite-Rees, P.A., Keating, C.J., Moran, L.S., Slatko, B.E., Hornstra, L.J. and Benner, J.S. (1991) Characterization and expression of the *Escherichia coli* Mrr restriction system. *J. Bacteriol.*, **173**, 5207–5219.
27. Fehér, T., Karcagi, I., Györfy, Z., Umenhoffer, K., Csörgö, B. and Pósfai, G. (2008) Scarless engineering of the *Escherichia coli* genome. *Methods Mol. Biol.*, **416**, 251–259.
28. Sambrook, J. and Russell, D.W. (2006) *The Condensed Protocols From Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
29. Vieira, J. and Messing, J. (1991) New pUC-derived cloning vectors with different selectable markers and DNA replication origins. *Gene*, **100**, 189–194.
30. Chang, A.C. and Cohen, S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.*, **134**, 1141–1156.
31. Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G. and Alon, U. (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods*, **3**, 623–628.
32. Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, **177**, 4121–4130.
33. Gommans, W.M., McLaughlin, P.M., Lindhout, B.I., Segal, D.J., Wiegman, D.J., Haisma, H.J., van der Zaal, B.J. and Rots, M.G. (2007) Engineering zinc finger transcription factors to downregulate the epithelial glycoprotein-2 promoter as a novel anti-cancer treatment. *Mol. Carcinog.*, **46**, 391–401.
34. Nyerges, Á., Csörgö, B., Nagy, I., Bálint, B., Bihari, P., Lázár, V., Apjok, G., Umenhoffer, K., Bogos, B., Pósfai, G. *et al.* (2016) 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.
35. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F. and Marraffini, L.A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.*, **41**, 7429–7437.
36. Miller, J.H. (1992) *A Short Course in Bacterial Genetics. A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor Laboratory Press, NY.
37. Zhang, X. and Bremer, H. (1995) Control of the *Escherichia coli* *rrnB* P1 promoter strength by ppGpp. *J. Biol. Chem.*, **270**, 11181–11189.
38. Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, **3**, 318–356.
39. Wilson, C.J., Zhan, H., Swint-Kruse, L. and Matthews, K.S. (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol. Life Sci.*, **64**, 3–16.
40. Müller-Hill, B., Crapo, L. and Gilbert, W. (1968) Mutants that make more lac repressor. *Proc. Natl. Acad. Sci. U.S.A.*, **59**, 1259–1264.
41. Calos, M.P. (1978) DNA sequence for a low-level promoter of the *lac* repressor gene and an ‘up’ promoter mutation. *Nature*, **274**, 762–765.
42. Glascock, C.B. and Weickert, M.J. (1998) Using chromosomal *lacIQ1* to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene*, **223**, 221–231.
43. Hsu, L.M. (2002) Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta*, **1577**, 191–207.
44. Farabaugh, P.J. (1978) Sequence of the *lacI* gene. *Nature*, **274**, 765–769.
45. Steege, D.A. (1977) 5'-Terminal nucleotide sequence of *Escherichia coli* lactose repressor mRNA: features of translational initiation and reinitiation sites. *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 4163–4167.