E. coli chromosomal-driven expression of NADK2 from A. thaliana: A preferable alternative to plasmid-driven expression for challenging proteins

Matthieu Goussé, Elisa Dell'aglio, Gilles Curien, Stéphanie Borland, Sébastien Renoud, Caroline Ranquet, Alexia Chandor-Proust

Title

***E. coli* chromosomal-driven expression of NADK2 from *A. thaliana*: a preferable alternative to plasmid-driven expression for challenging proteins**

Authors

Matthieu Goussé[1], Elisa Dell'Aglio[2], Gilles Curien[2], Stéphanie Borland[1], Sébastien Renoud[1], Caroline Ranquet[1] and Alexia Chandor-Proust[1]*

Addresses

[1]BGene, 7 rue des Arts et Métiers 38000 Grenoble, France

[2]Univ. Grenoble Alpes, CNRS, CEA, INRAe, IRIG-LPCV, Grenoble, France.

*Corresponding author: Alexia Chandor-Proust alexia.chandor@bgene-genetics.com

Highlights

- Chromosomal-driven expression systems improves bacterial growth after IPTG induction of recombinant proteins with respect of plasmid-driven expression systems
- Chromosomal-driven expression successfully leads to production of the large and complex *Arabidopsis thaliana* NADK2 protein in *E. coli*
- The activity recovery of NADK2 is improved by chromosomal-driven expression

Author contribution

Matthieu Goussé: conceptualization, methodology, validation, investigation, data curation, visualization, writing; **Elisa Dell'Aglio**: methodology, investigation, visualization, writing; **Gilles Curien**: methodology, investigation, writing; **Stéphanie Borland**: visualization, writing; **Sébastien Renoud**: Formal analysis; **Caroline Ranquet**: conceptualization, supervision, project administration, data curation, writing.; **Alexia Chandor-Proust**: conceptualization, supervision, project administration, data curation, writing.

Abstract

The expression and purification of large recombinant proteins or protein complexes is an issue for a number of biotechnology laboratories. Indeed, it is often difficult to obtain sufficient amounts of active proteins to perform biological characterization or reach commercialization when large proteins or protein complexes are expressed in *E. coli* via the popular T7-based plasmid-driven expression system.

There is also an industrial demand to decrease our dependence on plasmid-driven expression, because of its various drawbacks, such as: *i)* the common use of antibiotics to maintain the plasmid, *ii)* the issue of plasmid copy number, and *iii)* the risk of overloading the expression system. For all this, alternative methods such as gene integration in the bacterial chromosome might be an attractive solution, but this strategy is still rarely employed and its advantages are still a matter of debate.

Plant plastidial NAD kinases (NADK; ATP:NAD 2'-phosphotransferase, EC 2.7.1.23) constitute a classic example of proteins with high molecular weight, difficult to express and purify with the traditional T7-based technology. We therefore compared plasmid-driven and chromosomal-driven expression of the *Arabidopsis thaliana* NADK2 protein, thanks to a proprietary counter-selection tool, COLIBELT®, allowing scar-free and marker-free chromosomal modifications.

Here we show that chromosomal-driven expression allowed a better production and the recovery of more active NADK2 protein than classic T7 expression systems, thus confirming that expression from one single chromosomal copy is preferable to plasmid-driven expression and might be appealing for both basic and applied research.

Figure list :

Keywords

Expression system

E. coli

T7-based expression

Chromosomal insertion

NAD kinase

Introduction

The most common and easy strategy for recombinant protein production is plasmid-driven gene expression in a microbial system. *Escherichia coli* (*E. coli*) is one of the main microbial factories used for recombinant protein expression, laboratory investigations, initial development, and industrial production [1]. Initially developed by Studier and Moffat in 1986, the T7-based system has been widely used for the expression of recombinant proteins. It relies on the expression of the T7 *gene 1*, coding for an RNA polymerase from the T7 bacteriophage that is highly active and specific for T7 phage promoters [2]. In the well-known *E. coli* strain BL21(DE3), the transcription of this gene is under control of the *lac*UV5 promoter, therefore allowing repression by the LacI regulator and induction with isopropyl-b-D-1-thiogalactopyranoside (IPTG). The genes encoding the proteins of interest are cloned downstream of a T7 promoter, often into pET plasmids. Over the last decade, the T7 expression system has been highly improved thanks to the development of highly efficient *E. coli* strains (Shuffle T7, Nova Blue DE3, Origami….), that allow the recovery of high amounts of soluble proteins [3].

Nevertheless, many proteins of interest remain recalcitrant to production or end up in insoluble or inactive forms, which often result in cell toxicity [4,5] and affecting the recombinant protein yield. Furthermore, the metabolic burden of plasmid replication can lead to reduced final biomass and plasmid stability [6]. Finally, plasmid maintenance often requires the addition of antibiotics to the medium. Although feasible at the laboratory scale, the use of antibiotics at the industrial scale increases the production cost and raises health concerns due to the spread of antibiotic resistance genes. To avoid these disadvantages, stable insertion of recombinant genes into the chromosome of the bacterial host may be desirable and has been already attempted in synthetic biology approaches. For example, chromosomal gene integration was successfully used to produce valuable compounds and chemicals such as 2'-fucosyllactose and melanin [7,8]. However, for the moment, the use of chromosomal integration is still not common for expression and purification of recombinant proteins, so its advantages are still understudied.

Nicotinamide adenine dinucleotide phosphate NADP(H) is an essential cofactor of dehydrogenases found in all living organisms. NADP(H) is mostly involved in catabolic reactions including lipid biosynthesis, amino acid biosynthesis or reduction of ribonucleotides to deoxyribonucleotides [9], but also participates in anabolic reactions and defence against oxidative stress[10] . In plants, NADP(H) has a crucial role in photosynthesis as it constitutes the final electron acceptor in the light reactions. NADP(H) is synthesized from NAD and ATP by a class of enzymes called NAD kinases (NADK; ATP:NAD 2'-phosphotransferase, EC 2.7.1.23). In *Arabidopsis thaliana*, four isoforms have been identified and named NADK1 (AT3G21070), NADK2 (AT1G21640), NADK3 (AT1G78590), and NADKc (for Calmodulin-dependent NADK, AT1G04280). While NADK1 and NADK3 are localized in the cytosol, NADKc is anchored at the mitochondrial membrane, and NADK2 is the only one located in the chloroplast [11–15]. In agreement with the crucial role of NADK2 for the photosynthetic process, NADK2 mutants have been shown to be drastically impaired in plant development and growth, and display high levels of photodamage in standard growth conditions, with lower production of chlorophyll and overaccumulation of xanthophylls, suggesting that NADK2 is the main source of NADP that sustains photosynthesis [13,16–19].

The *Arabidopsis thaliana* NADK2 (At1g21640) is a large protein of 998 amino acids. The first 61 residues correspond to a predicted transit peptide needed for addressing the protein to the chloroplast [20,21]. Amino acids 678 to 990 constitute the catalytic domain, in common with NADK1 and NADK3

(IPR002504). The region between the transit peptide and the catalytic domain is a long amino-acid sequence of unknown function. This region is conserved in plastidial NADK2 homologues of other plant species [22] and therefore is likely to have a conserved role in the regulation of the enzyme activity and/or protein interaction. It had been hypothesized that this domain would be an anchoring point for calmodulin, a putative activator of NADK2 [22,23], but with the discovery of the Calmodulin-dependent NADKc, whose mutants are totally devoid of CaM-dependent NAD kinase activity, this hypothesis has been disproven [15,24]. Therefore, the role of the NADK2 long N-terminal region remains unknown. However, all previous attempts in protein purification that might shed light on the enzymatic regulation and the role of the N terminal domain have yielded low protein amounts [22,23,25], which is likely due to the high molecular weight of NADK2 proteins (around 100 kDa), thus complicating functional and interaction studies, and preventing crystallography attempts.

In this study, we therefore constructed and improved an *E. coli* plasmid-free expression strain capable of biosynthesizing an active and soluble NADK2 enzyme in higher quantities compared to standard, plasmid-based procedures. To overcome the problem of overproduction through the pET system, the gene encoding the NADK2 enzyme was inserted in one single copy directly into the chromosome of the production strain BL21(DE3). Two loci of insertion were tested as the gene expression yield could be rather different following the location on the chromosome [26]. Our results show that an alternative chromosomal-driven expression system can be advantageously used for the expression of challenging proteins.

Materials and methods

Bacterial strains, plasmids, and growth conditions

All strains and plasmids used in this study are listed in **Erreur ! Source du renvoi introuvable.**. *E. coli* 10β was used as the host to construct and store recombinant and suicide plasmids, and *E. coli* BL21(DE3) was used for overproduction of recombinant proteins. Solid media were supplemented with agar 1.5 % w/v. If necessary, the media were supplemented with kanamycin (Roth) at a final concentration of 50 μg/ml.

Cultures were performed in LB medium (Roth) at 30°C (permissive temperature) or 37°C (for expression assays) with rotation shaking at 200 rpm. To induce the gene expression, cultures at an OD_{600} ~ 0.8-1 were supplemented with 0.4 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma) during 15-16 hours at 20°C with rotation shaking at 200 rpm. At the end of induction, cultures were centrifuged for 10 minutes at 5,000 rpm at 4°C.

Table 1 : *Bacterial strains and plasmids used in this study.*

	Relevant genotype	**Source**
Strains		
E. coli 10β	Host for cloning, Δ*(ara-leu) 7697 araD139 fhuA* Δ*lacX74 galK16 galE15 e14- φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str$_R$) *rph spoT1* Δ*(mrr-hsdRMS-mcrBC)*	New England biolabs
E. coli BL21	*F⁻ ompT gal dcm lon hsdS$_B$($r_B^-m_B^-$) [malB$^+$]$_{K-12}$(λ^S)*	GE Healthcare
E. coli BL21(DE3)	*F⁻ ompT gal dcm lon hsdS$_B$($r_B^-m_B^-$) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB$^+$]$_{K-12}$(λ^S)*	Novagen
E. coli BL21(DE3) *ybfP* :: WT T7 *nadk2*	*F⁻ ompT gal dcm lon hsdS$_B$($r_B^-m_B^-$) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB$^+$]$_{K-12}$(λ^S) ΔybfP* :: WT T7 *nadk2*	this study
E. coli BL21(DE3) *yifB* :: WT T7 *nadk2*	*F⁻ ompT gal dcm lon hsdS$_B$($r_B^-m_B^-$) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB$^+$]$_{K-12}$(λ^S) ΔyifB* :: WT T7 *nadk2*	this study
Plasmids		
pET30	Kn^R	Novagen
pET30 *NADK2*	pET30, *A. thaliana nadk2*, Kn^R	this study
pBGT7_P	*Kn^R, suicide plasmid for insertion into ybfP (oriTs, Colibelt®®, P_{T7} nadk2*)	this study
pBGT7_B	*Kn^R, suicide plasmid for insertion into yifB (oriTs, Colibelt®®, P_{T7} nadk2*)	this study

Table 2 : *Primers used in this study.*

Primers	**Sequence**
NADK2fp	5'-GATTTGTGCATATGGCGCAGCTCTCTGAAGC-3'
NADK2rp	5'-CTAAAATGAGCTCATCAATCATCAGAGAGCC-3'
P_{up}_F	5'- AGCGCAGGGGCTTCCTTCC -3'

P_{up}_R	5’- TGGGATTTCCTTGAGCTACTCAGTTTG -3’
P_{down}_F	5’- AATAGTGAGTCCGGGAATGAAACC -3’
P_{down}_R	5’- CCTTCTTGACGAGTTCTTCTAATGAATCAGTTAGTAAATACTGAAGATTCAGC -3’
B_{up}_F	5’- TAAGGCAGTAGTTTTTCACCGGCAG -3’
B_{up}_R	5’- AAAAAGGGCATTTCGCCCTTTTTATTAATCG -3’
B_{down}_F	5’- AAGCCCTCCTTGAGTCACCATTATG -3’
B_{down}_R	5’- CCTTCTTGACGAGTTCTTCTAAAAACCGGCATAATTGCGCCACC -3’
ACPWTP_F	5’- GAGTAGCTCAAGGAAATCCCATAATACGACTCACTATAGGGGAATTGTGAG -3’
ACPWTP_R	5’- TTTCATTCCCGGACTCACTATTATCCGGATATAGTTCCTCCTTTCAGC -3’
ACPWTB_F	5’- AGGGCGAAATGCCCTTTTTTAATACGACTCACTATAGGGGAATTGTGAG -3’
ACPWTB_R	5’- GTGACTCAAGGAGGGCTTATCCGGATATAGTTCCTCCTTTCAGC -3’
P_{insP}_F	5' - AGCTTCCTCGGTGAAGAACA- 3'
P_{insP}_R	5' - CATGTTCAGGTCCGGGTTTC- 3'
P_{insB}_F	5' - AGCCAGTGACAGAAGCAG- 3'
P_{insB}_R	5' - TGAATGTTGTTCCCTTCC- 3'

Construction of plasmids and chromosomal integration

All primers used are listed into Table 2. Standard Plasmid DNA from *E. coli* was isolated using MACHEREY-NAGEL kits according to the manufacturer instructions (Macherey-Nagel, Düren, Germany). All standard DNA manipulations, digestions, ligation reactions and transformation in *E. coli* were done according to established protocols (Sambrook et al., 1989 add this reference to the list) or according to the manufacturer’s recommendations. Enzymes were purchased from NEB (New England Biolabs France, Evry, France). All PCR products were checked by sequencing (Eurofins, Ebersberg, Germany). For the construction of the pET30 NADK2 plasmid, the gene encoding NADK2 from *A. thaliana* was obtained as a fragment of 3,110 bp by PCR with primers NADK2fp and NADK2rp from a cDNA library. The DNA fragment was then digested with restriction enzymes NdeI and SacI and ligated into plasmid pET30 digested with the same enzymes.

Chromosomal integrations were performed with the help of a suicide plasmid. Suicide plasmids were first designed by a proprietary software (Cad4Bio®) and assembled by the Gibson Assembly’s method [27] with a HiFi kit from NEB, following the manufacturer instructions (New England Biolabs France, Evry, France). Both suicide plasmids designed were made of a thermosensitive replication origin OriTs, an efficient counter-selection cassette (Colibelt®), and the gene to be inserted was flanked by two homology regions of the target locus (patent FR3050997). Two pairs of primers for each site (P_{up}_F/P_{up}_R, P_{down}_F/P_{down}_R for insertion in *ybfP* and B_{up}_F/B_{up}_R, B_{down}_F/B_{down}_R for insertion in *yifB*), were then used to amplify approximatively 700-pb fragments upstream and downstream of *ybfP* and y*ifB* loci from the BL21(DE3) chromosome. Two pairs of primers were used to clone the T7 promoter with the *nadk2* gene into the suicide plasmid for integration in *ybfP* (ACPWTP_F/ ACPWTP_R) or *yifB* (ACPWTB_F/ ACPWTB_R) loci. All the PCR fragments were joined by Gibson assembly to generate the suicide plasmids pBGT7_P and pBGT7_B.

The suicide plasmids were transformed in BL21(DE3) by electroporation following routine protocols [28]. Transformants were selected on LB + Kanamycin plates at a permissive temperature allowing the replication of the plasmid. Briefly, to integrate the suicide plasmid into the targeted locus, clones were further grown at a non-permissive temperature whilst second allelic exchange was stimulated by growing the cells at a permissive temperature in the absence of antibiotic selection pressure. Clones that had the plasmid excised were further counter-selected with the Colibelt®

cassette. The correct integration of the genes of interest were verified by PCR with primers flanking the region of *ybfP* (P_{insP}_F/ P_{insP}_R) or *yifB* gene (P_{insB}_F/P_{insB}_R).

NADK2 expression assay

Cell lysis was done by sonication. Samples were resuspended in 3 mL of lysis buffer containing 50 mM Hepes-KOH (Euromedex), 200 mM KCl (Euromedex), 1 mM EDTA (Euromedex), 10% glycerol (Roth), 1 mM benzamidine (Sigma), 5 mM aminocaproic acid (Sigma), 1 mM NAD^+ (Sigma). To precipitate DNA, 0.2 % of sulfate streptomycin was added and samples were centrifuged for 25 min at 13,000 x g at 4°C. The supernatant sample, corresponding to soluble extract, was frozen in liquid nitrogen and conserved at -80°C.

Protein analysis

The abundance of NADK2 protein was analysed by SDS-PAGE with molecular weight marker (Roti®- Mark 10-150; Roth), allowing a separation between 10 and 150 kDa. Commercial gels were used (Roti®-Page gradient (4-20 %); Roth). Samples (approx. 5 µg) were supplemented with 6X Laemmli buffer (375 mM Tris-HCl pH 6.8, 9% SDS, 50% glycerol, 0.03% bromophenol blue). Migration was performed at 125V in a migration buffer (25 mM Tris-Base, 192 mM glycine, 0.1% SDS).

Western blots were performed on nitrocellulose Hybond ECL membrane (Bio-Rad). The transfer was performed in 90 min at 200 mA in a buffer containing 25 mM Tris-Base, 192 mM glycine, 0.1% SDS, 20% ethanol.

At the end of the transfer, the membrane was washed in TBST (20 mM Tris-Base pH 7.5, 150 mM NaCl and 0.05% (v/v) Triton) and saturated during 1h at room temperature in TBST-milk (milk: 5% (p/v)). After saturation, the membrane was washed three times in TBST and incubated with primary antibody raised against the NADK2 N-terminal domain of unknown function (Guinea pig polyclonal anti-NADK2, 1/5000) overnight at 4°C in TBST-milk. The following day, three washes of 5 min were carried out in TBST. The membrane was incubated with the secondary antibody coupled with horseradish peroxidase (HRP) (Rabbit polyclonal anti-guinea pig, 1/40000) for 1h30 at room temperature and washed again with TBST. The presence of proteins recognized by antibodies was revealed by the detection of HRP activity using a chemiluminescent reagent kit (BioRad) and pictures were recorded on an Image Quant device (Image Quant LAS 4000). A GAPDH antibody was use as loading control.

NADK enzymatic assay

NADK activity was determined spectrophotometrically using a coupled enzymatic assay with glucose-6-phosphate dehydrogenase (G6PDH). The assay was performed in two steps. The first reaction, (NADP production, called “R1”), contained 62.5 mM Hepes-KOH, pH 7.5, 6.25 mM of ATP (mol/ mol NA_2CO_3; Sigma), 12,5 mM MgCl2, 1,25 mM NAD^+ (Sigma), for a final volume of 90 µL. To normalize the test, 10 µg of soluble bacterial extract were added allowing the reaction:

NAD^+ + Mg-ATP → $NADP^+$ + ADP, catalyzed by NAD Kinase

The mix was incubated at 25°C during 10 min. At the end of the reaction, the mix was heated for 5 min at 95°C to denature the NADK2 protein. Samples were kept on ice up to the second reaction.

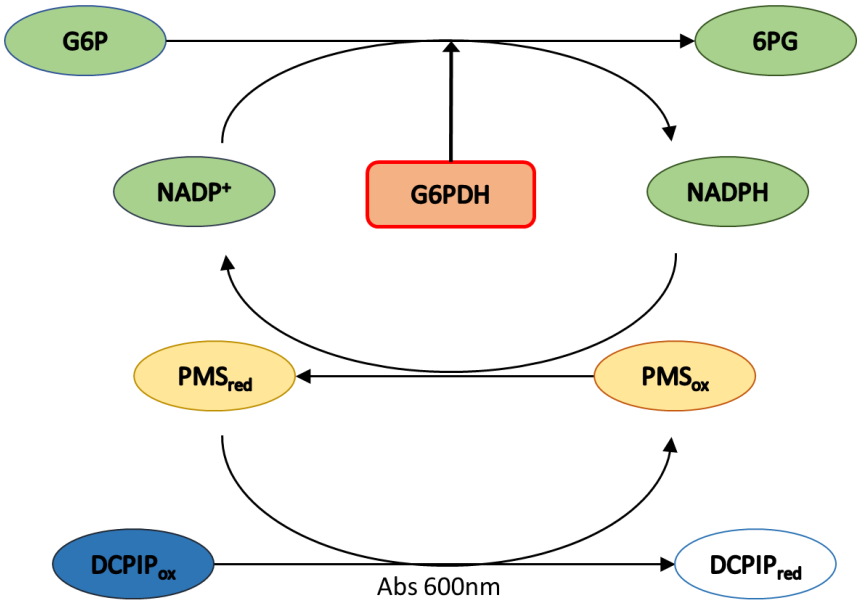

Figure 1 : Reaction of G6PDH enzyme used to quantify NADP products by NAD Kinase*. G6P: Glucose-6-Phosphate; 6PG: 6-Phospogluconate; G6PDH: Glucose-6-Phosphate Dehydrogenase; PMSred: Phenazine Methosulphate reduced; PMSox: Phenazine Methosulphate oxidized; DCPIPox: 2,6-Dichlorophenolindophenol oxidized; DCPIPred: 2,6-Dichloroindophenol reduced.*

The second reaction (NADP detection) was performed with a mix ("R2") containing 125 mM of Hepes-KOH pH 7.5, 15 mM Glucose-6-Phosphate (G6P; Sigma), 0.5 mM Phenazine Methosulphate (PMS; Sigma), 1 mM of 2,6-Dichlorophenolindophenol (DCIP; Sigma), for a final volume of 40 µL. Finally, 30 µL of R1 were added to the R2 mix as well as 30 µL of Glucose-6-Phosphate Dehydrogenase (G6PDH; Sigma) at 1 mU/µL extemporaneously to carry out the second reaction at 25°C (**Figure 1**). The G6PDH uses the NADP produced by the first reaction to dehydrogenate Glucose-6-Phosphate and produces 6-Phosphogluconate. The NADP is reduced to NADPH, which is regenerated by the oxidized PMS. Once reduced, the PMS transfers its electron to the DCPIP. The DCPIP is a blue compound and becomes clear when reduced leading a decrease of absorbance at 600 nm. A kinetics was performed over 10 minutes to monitor the reduction of DCPIP at 600 nm, allowing to correlate this decrease with the amount of NADP produced and thus the activity of NADK2 thanks to a standard curve made with pure $NADP^+$ (Sigma). Measurements were carried out in microtiter plates and UV-vis spectrophotometer (Spark, Tecan). Statistical tests were carried out using the statistical software R [29,30].

Results

Obtaining stable NADK2 expression strains

The *nadk2* gene, downstream of the T7 promoter, was integrated as described above, into two different chromosomal loci in *E. coli*, replacing *ybfP* or *yifB* ORFs. These two genes were chosen because they are non-essential in *E. coli* [31] and located at two different loci in the chromosome of *E. coli* (map position: ybfP [715,412 -> 715,906] ; yifB [3,948,449 <- 3,949,969]). These coordinates were chosen to highlight any possible impact of the chromosomal position for the expression and activity of the enzyme, as it has been demonstrated that the integration position has a direct impact on the gene expression [32]. Moreover, the number of copies of a gene during the replication was higher for genes located near the replication of origin compared to genes located near termini, and that this copy number could affect the expression of the target protein [33–35].

The strain *E. coli* BL21(DE3) was transformed with the suicide plasmids pBGT7_P or pBGT7_B, and insertion of the *nadk2* gene under the control of T7 promoter was performed as described in the "Materials and Methods" section. Integration into the two loci was confirmed by PCR **(Figure 2)** and by Sanger sequencing.

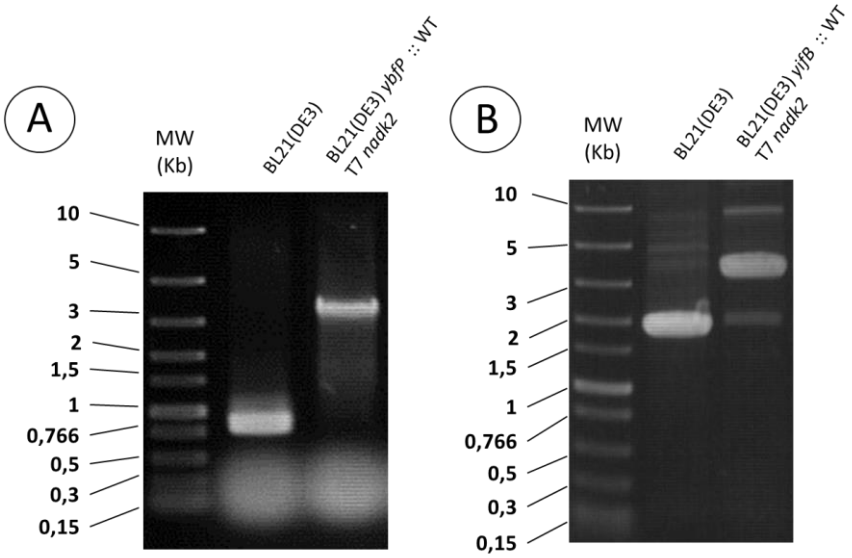

***Figure 2** : Integration of nadk2 gene into ybfP (A) or yifB (B) loci. In both panels, the bands represent the PCR product of each locus, before (left band) and after (right band) integration of the nadk2 gene sequence.*

Chromosomal-driven expression improves bacterial growth in the presence of IPTG

To estimate the advantage of expressing the 110 kDa *Arabidopsis thaliana* NADK2 protein by chromosomal-driven expression, we firstly compared the growth rate of our newly constructed NADK2-containing strains with non-transformed BL21(DE3) bacteria or BL21(DE3) containing a pET30 plasmid expressing the *nadk2* gene (pET30 *NADK2*). No difference in growth rate was observed between strains containing the pET30 *NADK2* plasmid and strains with *nadk2* gene inserted into the chromosome without IPTG, all reaching a plateau at 2.5 OD_{600} after ten hours of growth. On the contrary, the IPTG induction led to a strong delay in the bacterial growth for pET30 *NADK2*-containing strain, to the point that these strains did not reach a plateau during the course of the experiment (total growth time: 15 hours). Notably, growth of the pET30 *NADK2* was even more delayed in the presence of kanamycin, which prevents the plasmid loss and demands additional cost for producing the kanamycin resistance (Fig. 3A).

These results strongly suggest that the overexpression of the *nadk2* gene from a high copy number plasmid is toxic for *E. coli* (**Figure 3**). Since the growth difference is only detected when the protein expression is induced, it can be hypothesized that the toxicity is not due to the high plasmid copy number but rather to the metabolic cost of protein overexpression.

In strong contrast with these observations, NADK2 chromosomal-driven expression from the *ybfP* or the *yifB* locus did not alter the bacterial growth rate, which was identical to the growth rate of BL21(DE3) bacteria, with or without IPTG-dependent induction of NADK2 (Fig. 3B).

These results lead to conclude that chromosomal-driven expression improves bacterial growth after the induction of recombinant protein production with IPTG, thus allowing the recovery of a higher bacterial biomass.

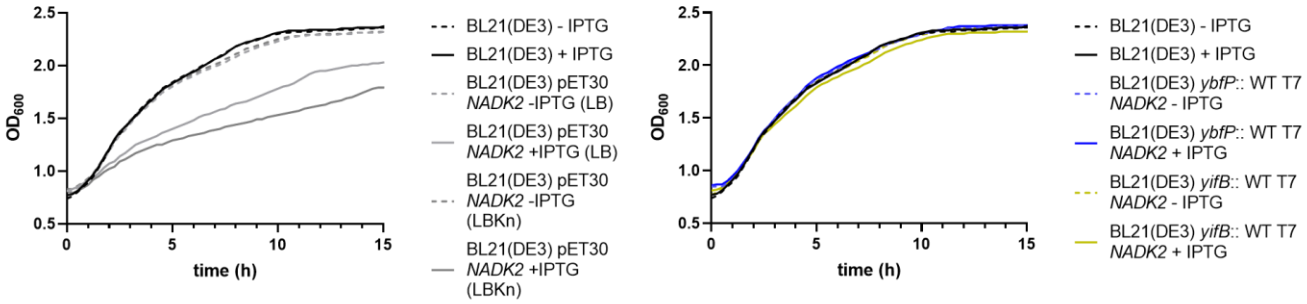

***Figure 3 :** Growth of E. coli strains harbouring the nadk2 gene under the control of WT T7 promoter either A) on plasmid (pET30 NADK2), in the presence (dark grey) or absence (light grey) of kanamycin, orB) on chromosome, with integration in the ybfP (blue) or the yifB (citron), in comparison with the non- transformed BL21(DE3) bacteria (black). The growth was performed with (continuous lines) or without (dashed lines) NADK2 induction at 0.4 mM of IPTG in LB media at 25°C.*

Chromosomal-driven expression improves soluble protein yield

To compare the production of NADK2 according to plasmid-driven or chromosomal-driven expression, NADK2 expression was performed overnight in 100 ml cultures, which were subsequently lysed by sonication, and centrifuged to recover the soluble extract which was resuspended in 3 ml final volume. Protein concentrations of the various soluble extracts ***(Figure 4)*** ranged from 5 to 10mg/mL for the control BL21(DE3) strain and the pET30 *NADK2*-containing strain. Consistent with the higher biomass produced by strains with the *nadk2* gene integrated in the chromosome (figure 3), total protein concentration was much higher in the latter, ranging from 12 to 16 mg/ml .

The integration of the *nadk2* gene into the chromosome then confers a clear advantage compared to plasmid-driven expression for soluble protein yield. This advantage could be due to a lower amount of NADK2 protein (by a slightly reduced but smoother protein production, thus not saturating the general bacterial expression and refolding systems) , and/or to the metabolic cost for plasmid replication which might lower the expression efficiency, as proposed by Striedner et al [36].

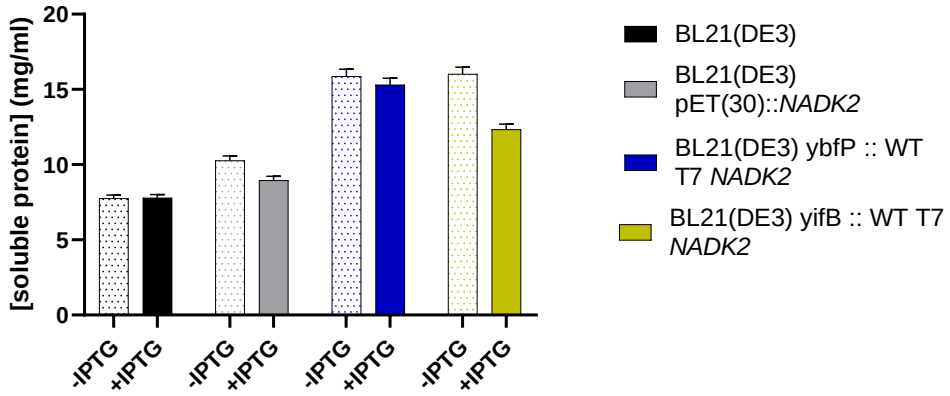

***Figure 4 :** Concentration of soluble proteins in 3 ml resuspension volume recovered from strains harbouring the nadk2 gene under the control of the T7 promoter on plasmid (pET30 NADK2, grey) or on chromosome in the ybfP (blue) or yifB (citron) loci, with (filled bars) or without (dotted bars) induction by 0.4 mM IPTG. The control condition (BL21(DE3) non-transformed bacteria) is shown in black.*

Chromosomal-driven induction allows efficient recombinant NADK2 expression

The expression of the NADK2 protein in IPTG-induced cultures was verified by SDS-PAGE acrylamide gels followed by western blot. The western blot showed that the NADK2 protein was well recovered at the expected size of 110 kDa ***(*****Figure 5*****)*** for strains that expressed the gene either on plasmid or on chromosome after induction with IPTG. A little expression leakage in the absence of IPTG was observed when NADK2 was expressed from the *ybfP* locus. The NADK2 signal observed when the protein was expressed from the chromosome was comparable, with respect to GAPDH, to the plasmid-driven expression, whatever the chromosomal locus, *ybfP or yifB*. In addition, a thin band appeared at a lower mass than the NADK2 signal in all the IPTG-induced strains *E. coli* BL21 (DE3), which may correspond to proteolytic degradation and its presence in all IPTG-induced cultures suggests that this is a drawback independent on the expression system

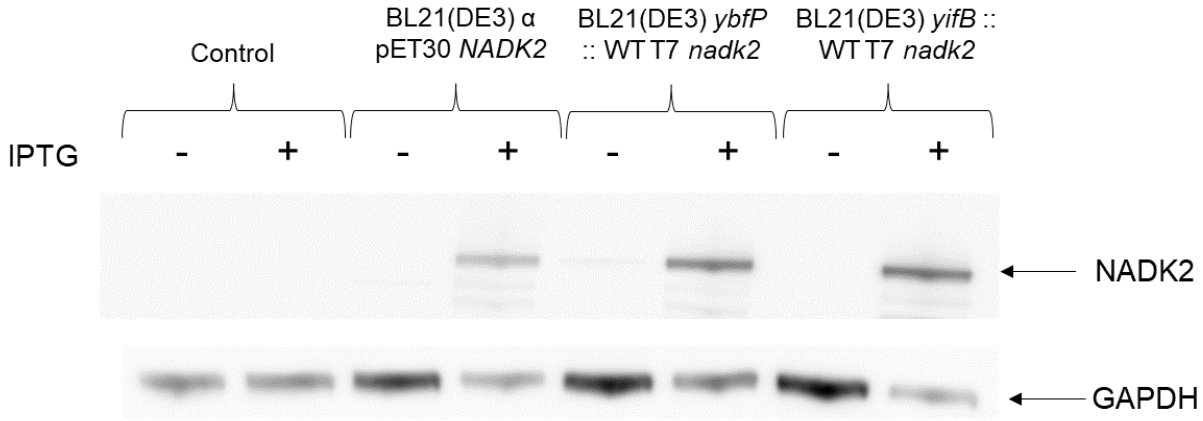

Figure 5 : Western blot showing NADK2 expression from pET30 NADK2 plasmid or from the two chromosomal loci ybfP and yifB, with or without IPTG induction.

Chromosomal-driven expression allows higher recovery of NADK activity with respect to plasmid-driven expression systems

Activity assays were subsequently carried out to check the functionality of the NADK2 enzyme in the bacterial soluble extracts, according to the protocol described in the "Materials and Methods", which was inspired by Turner et al. [22] and adapted to our conditions. For each strain, at least 6 replicates were performed under the same conditions. Since the activity was measured in bacterial soluble extracts, we detected both the activity of the endogenous bacterial NADK enzyme and the activity of NADK2. The bacterial activity is similar in the negative control where NADK2 is absent, both in presence and absence of IPTG induction. Similar activity levels are measured in the absence of IPTG for plasmid-driven expression and chromosomal-driven expression from the *yifB* locus. The NAD kinase activity recovered from the non IPTG induced cultures-from the strain with NADK2 integration in the *ybfP* locus was significantly higher than from the pET30 NADK2-containing strain, and this might be explained by the promoter leakage that was also noticeable, for *ybfP*, in the western blot (**Figure 5**).

In the presence of IPTG induction, higher levels of NADK activity coming from the recombinant gene was detected in both the plasmid-driven and the two chromosomal-driven expression systems (**Figure 6**), leading to a production of around 14 pmoles NADP/s/µg of protein extract. Notably, the activity

values obtained from the strain with NADK2 integration in the *yifB* locus are significantly higher with respect to the pET30 NADK2 – induced strain.

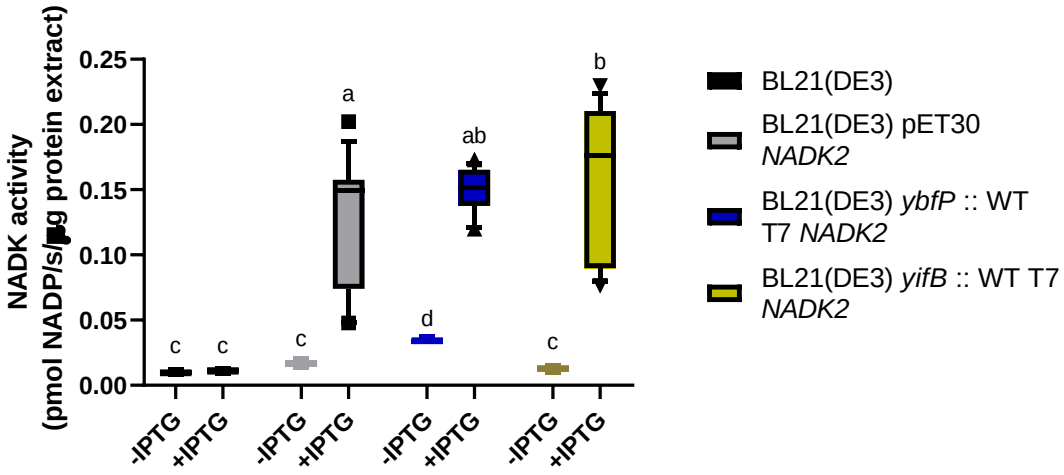

***Figure 6 :** NADK activity detected in bacterial soluble extracts expressing NADK2 either from the pET30 NADK2 plasmid (grey bars) or on chromosome in the ybfP (blue) or yifB (citron) loci. Filled bars: activity form IPTG-induced strains; dotted bars: activity from non-induced strains. The BL21(30) control strain is depicted in black. Error bars represent standard error of the mean. Statistics: Kruskal-Wallis non-parametric test.*

As a consequence of the higher NADK activity recovered per mg of protein (figure 6) and of the surplus of soluble proteins obtained through chromosomal-driven expression (figure 4) it was possible to recover twice as much NADK activity when NADK2 expression was driven from the bacterial chromosome as in the *yifB* locus than from the pET30 plasmid (**Figure 6**).

In conclusion, chromosomal-driven expression is a powerful way to double the recovery of active, hard to express, large recombinant proteins in *E. coli*.

Discussion

Chromosomal-driven expression of recombinant proteins in bacteria eliminates the risk of plasmid loss, the additional metabolic burden for plasmid replication and all possible deleterious effect of the antibiotic pressure on the culture. However, the presence of a single copy of the target protein might counterbalance the possible benefits.

We therefore compared the chromosomal-driven expression with the classic, plasmid-driven expression of a long and tricky protein, NADK2, in *E. coli*, by integrating the gene in two loci close to the origin of replication, which is known to increase the gene copy number [32]. Our results confirm that chromosomal-driven expression in *E. coli* improves bacterial growth, soluble protein yield, and activity per mg of protein. As a consequence, the chromosomal driven expression appears preferable to recover higher quantities of soluble, active protein.

These results are in agreement with previous tests showing that a too high translation rate leads to inclusion bodies or proteolytic degradation and thus a weak production [37]. Likewise, previous attempts of protein expression from the bacterial chromosome were successful in showing improvement in productivity of the heterologous protein [7,38,39].

In our study we have tested the NADK2 expression from two different chromosome loci, namely *ybfP* and *yifB*. The first locus codes for an ORFan gene [40], while the second one is probably a member of AAA+ ATPases [41]. Both genes were known to be not lethal when knocked down and did not lead to growth impairments in normal conditions. The latter was judged as particularly suitable to delete, so that we can bypass the quality control machinery to optimize the strain for recombinant proteins production .

In our case, both tested loci led to comparable results in terms of bacterial growth. However, the *ybfP* locus displayed low levels of expression leakage in the absence of IPTG. In our case, this did not affect bacterial growth, but it may be a problem in case of production of toxic proteins, such as toxin/antitoxin systems [42] where induction must be carefully timed. For this, and for the higher NADK activity recovered, we can conclude that the *yifB* is preferable over *ybfP*. Other loci still need to be considered to widen the range of targets for chromosomal insertions.

In further attempts, chromosomal-driven expression systems will likely allow to overcome expression issues for other difficult targets. For example, protein complexes of several units are uneasy to express from a single plasmid in a bioproduction host. We can cite operons or clusters for Polyketide Synthase heterologous expression [43], exopolysaccharides like xanthan gum biosynthesis [44], or cellulosome complexes for cellulose metabolism [45]. Increasing the number of plasmids in a strain is an issue, as they must be compatible, and it increases the use of antibiotics. Chromosomal driven expression is an answer to improve the expression of these kinds of complexes that are more and more used in metabolic engineering and bioproduction in *E. coli* or other chassis..

Acknowledgments

Competing interests

The authors declare they have no conflict of interest.

References

[1] R. Chen, Bacterial expression systems for recombinant protein production: E. coli and beyond, Biotechnol. Adv. 30 (2012) 1102–1107. https://doi.org/10.1016/j.biotechadv.2011.09.013.

[2] F.W. Studier, B.A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, J. Mol. Biol. 189 (1986) 113–130. https://doi.org/10.1016/0022-2836(86)90385-2.

[3] J.C. Samuelson, Recent developments in difficult protein expression: a guide to E. coli strains, promoters, and relevant host mutations, Methods Mol. Biol. Clifton NJ. 705 (2011) 195–209. https://doi.org/10.1007/978-1-61737-967-3_11.

[4] A. de Marco, Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli, Microb. Cell Factories. 8 (2009) 26. https://doi.org/10.1186/1475-2859-8-26.

[5] S. Schlegel, E. Rujas, A.J. Ytterberg, R.A. Zubarev, J. Luirink, J.-W. de Gier, Optimizing heterologous protein production in the periplasm of E. coli by regulating gene expression levels, Microb. Cell Factories. 12 (2013) 24. https://doi.org/10.1186/1475-2859-12-24.

[6] F. Silva, J.A. Queiroz, F.C. Domingues, Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli, Biotechnol. Adv. 30 (2012) 691–708. https://doi.org/10.1016/j.biotechadv.2011.12.005.
[7] F. Baumgärtner, L. Seitz, G.A. Sprenger, C. Albermann, Construction of Escherichia coli strains with chromosomally integrated expression cassettes for the synthesis of 2'-fucosyllactose, Microb. Cell Factories. 12 (2013) 40. https://doi.org/10.1186/1475-2859-12-40.
[8] A. Sabido, L.M. Martínez, R. de Anda, A. Martínez, F. Bolívar, G. Gosset, A novel plasmid vector designed for chromosomal gene integration and expression: use for developing a genetically stable Escherichia coli melanin production strain, Plasmid. 69 (2013) 16–23. https://doi.org/10.1016/j.plasmid.2012.04.005.
[9] L. Agledal, M. Niere, M. Ziegler, The phosphate makes a difference: cellular functions of NADP, Redox Rep. 15 (2010) 2–10. https://doi.org/10.1179/174329210X12650506623122.
[10] S. Spaans, R. Weusthuis, J. Van Der Oost, S. Kengen, NADPH-generating systems in bacteria and archaea, Front. Microbiol. 6 (2015) 742. https://doi.org/10.3389/fmicb.2015.00742.
[11] W.L. Turner, J.C. Waller, B. Vanderbeld, W.A. Snedden, Cloning and Characterization of Two NAD Kinases from Arabidopsis. Identification of a Calmodulin Binding Isoform, Plant Physiol. 135 (2004) 1243–1255. https://doi.org/10.1104/pp.104.040428.
[12] W.L. Turner, J.C. Waller, W.A. Snedden, Identification, molecular cloning and functional characterization of a novel NADH kinase from *Arabidopsis thaliana* (thale cress), Biochem. J. 385 (2005) 217–223. https://doi.org/10.1042/BJ20040292.
[13] M.-F. Chai, Q.-J. Chen, R. An, Y.-M. Chen, J. Chen, X.-C. Wang, NADK2, an Arabidopsis Chloroplastic NAD Kinase, Plays a Vital Role in Both Chlorophyll Synthesis and Chloroplast Protection, Plant Mol. Biol. 59 (2005) 553–564. https://doi.org/10.1007/s11103-005-6802-y.
[14] M.-F. Chai, P.-C. Wei, Q.-J. Chen, R. An, J. Chen, S. Yang, X.-C. Wang, NADK3, a novel cytoplasmic source of NADPH, is required under conditions of oxidative stress and modulates abscisic acid responses in Arabidopsis, Plant J. 47 (2006) 665–674. https://doi.org/10.1111/j.1365-313X.2006.02816.x.
[15] E. Dell'Aglio, C. Giustini, A. Kraut, Y. Couté, A. Costa, G. Decros, Y. Gibon, C. Mazars, M. Matringe, G. Finazzi, G. Curien, Identification of the Arabidopsis Calmodulin-Dependent NAD+ Kinase That Sustains the Elicitor-Induced Oxidative Burst, Plant Physiol. 181 (2019) 1449–1458. https://doi.org/10.1104/pp.19.00912.
[16] H. Takahashi, A. Watanabe, A. Tanaka, S. Hashida, M. Kawai-Yamada, K. Sonoike, H. Uchimiya, Chloroplast NAD Kinase is Essential for Energy Transduction Through the Xanthophyll Cycle in Photosynthesis, Plant Cell Physiol. 47 (2006) 1678–1682. https://doi.org/10.1093/pcp/pcl029.
[17] S. Hashida, H. Takahashi, H. Uchimiya, The role of NAD biosynthesis in plant development and stress responses, Ann. Bot. 103 (2009) 819–824. https://doi.org/10.1093/aob/mcp019.
[18] K. Takahara, I. Kasajima, H. Takahashi, S. Hashida, T. Itami, H. Onodera, S. Toki, S. Yanagisawa, M. Kawai-Yamada, H. Uchimiya, Metabolome and Photochemical Analysis of Rice Plants Overexpressing Arabidopsis NAD Kinase Gene, Plant Physiol. 152 (2010) 1863–1873. https://doi.org/10.1104/pp.110.153098.
[19] L. Sun, Y. Li, W. Miao, T. Piao, Y. Hao, F.-S. Hao, NADK2 positively modulates abscisic acid-induced stomatal closure by affecting accumulation of H2O2, Ca2+ and nitric oxide in Arabidopsis guard cells, Plant Sci. 262 (2017) 81–90. https://doi.org/10.1016/j.plantsci.2017.06.003.
[20] M. Tomizioli, C. Lazar, S. Brugière, T. Burger, D. Salvi, L. Gatto, L. Moyet, L.M. Breckels, A.-M. Hesse, K.S. Lilley, D. Seigneurin-Berny, G. Finazzi, N. Rolland, M. Ferro, Deciphering thylakoid sub-compartments using a mass spectrometry-based approach, Mol. Cell. Proteomics MCP. 13 (2014) 2147–2167. https://doi.org/10.1074/mcp.M114.040923.
[21] B. Zybailov, H. Rutschow, G. Friso, A. Rudella, O. Emanuelsson, Q. Sun, K.J. van Wijk, Sorting signals, N-terminal modifications and abundance of the chloroplast proteome, PloS One. 3 (2008) e1994. https://doi.org/10.1371/journal.pone.0001994.

[22] W.L. Turner, J.C. Waller, B. Vanderbeld, W.A. Snedden, Cloning and characterization of two NAD kinases from Arabidopsis. identification of a calmodulin binding isoform, Plant Physiol. 135 (2004) 1243–1255. https://doi.org/10.1104/pp.104.040428.
[23] E. Dell'Aglio, C. Giustini, D. Salvi, S. Brugière, F. Delpierre, L. Moyet, M. Baudet, D. Seigneurin-Berny, M. Matringe, M. Ferro, N. Rolland, G. Curien, Complementary biochemical approaches applied to the identification of plastidial calmodulin-binding proteins, Mol. Biosyst. 9 (2013) 1234–1248. https://doi.org/10.1039/c3mb00004d.
[24] E. Dell'Aglio, S. Boycheva, T.B. Fitzpatrick, The Pseudoenzyme PDX1.2 Sustains Vitamin B6 Biosynthesis as a Function of Heat Stress, Plant Physiol. 174 (2017) 2098–2112. https://doi.org/10.1104/pp.17.00531.
[25] E. Dell'Aglio, The regulation of plastidial proteins by calmodulins, phdthesis, Université de Grenoble, 2013. https://tel.archives-ouvertes.fr/tel-01423919 (accessed December 2, 2021).
[26] M. Valens, S. Penaud, M. Rossignol, F. Cornet, F. Boccard, Macrodomain organization of the Escherichia coli chromosome, EMBO J. 23 (2004) 4330–4341. https://doi.org/10.1038/sj.emboj.7600434.
[27] D.G. Gibson, L. Young, R.-Y. Chuang, J.C. Venter, C.A. Hutchison, H.O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, Nat. Methods. 6 (2009) 343–345. https://doi.org/10.1038/nmeth.1318.
[28] J. Sambrook, D.W. Russell, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 2001. https://books.google.fr/books?id=YTxKwWUiBeUC.
[29] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2011. http://www.R-project.org/.
[30] agricolae: Statistical Procedures for Agricultural Research version 1.3-5 from CRAN, (n.d.). https://rdrr.io/cran/agricolae/ (accessed October 19, 2021).
[31] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, H. Mori, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection, Mol. Syst. Biol. 2 (2006) 2006.0008. https://doi.org/10.1038/msb4100050.
[32] J.A. Bryant, L.E. Sellars, S.J.W. Busby, D.J. Lee, Chromosome position effects on gene expression in Escherichia coli K-12, Nucleic Acids Res. 42 (2014) 11383–11392. https://doi.org/10.1093/nar/gku828.
[33] K. Nordström, S. Dasgupta, Copy-number control of the Escherichia coli chromosome: a plasmidologist's view, EMBO Rep. 7 (2006) 484–489. https://doi.org/10.1038/sj.embor.7400681.
[34] J. Slager, J.-W. Veening, Hard-Wired Control of Bacterial Processes by Chromosomal Gene Location, Trends Microbiol. 24 (2016) 788–800. https://doi.org/10.1016/j.tim.2016.06.003.
[35] R. Dryselius, K. Izutsu, T. Honda, T. Iida, Differential replication dynamics for large and small Vibrio chromosomes affect gene dosage, expression and location, BMC Genomics. 9 (2008) 559. https://doi.org/10.1186/1471-2164-9-559.
[36] G. Striedner, I. Pfaffenzeller, L. Markus, S. Nemecek, R. Grabherr, K. Bayer, Plasmid-free T7-based Escherichia coli expression systems, Biotechnol. Bioeng. 105 (2010) 786–794. https://doi.org/10.1002/bit.22598.
[37] L.G. Horga, S. Halliwell, T.S. Castiñeiras, C. Wyre, C.F.R.O. Matos, D.S. Yovcheva, R. Kent, R. Morra, S.G. Williams, D.C. Smith, N. Dixon, Tuning recombinant protein expression to match secretion capacity, Microb. Cell Factories. 17 (2018) 199. https://doi.org/10.1186/s12934-018-1047-z.
[38] A.H. Kachroo, M. Jayaram, P.A. Rowley, Metabolic engineering without plasmids, Nat. Biotechnol. 27 (2009) 729–731. https://doi.org/10.1038/nbt0809-729.
[39] K.E.J. Tyo, P.K. Ajikumar, G. Stephanopoulos, Stabilized gene duplication enables long-term selection-free heterologous pathway expression, Nat. Biotechnol. 27 (2009) 760–765. https://doi.org/10.1038/nbt.1555.
[40] H.P. Narra, M.H.J. Cordes, H. Ochman, Structural features and the persistence of acquired proteins, PROTEOMICS. 8 (2008) 4772–4781. https://doi.org/10.1002/pmic.200800061.
[41] L.M. Iyer, D.D. Leipe, E.V. Koonin, L. Aravind, Evolutionary history and higher order classification of AAA+ ATPases, J. Struct. Biol. 146 (2004) 11–31. https://doi.org/10.1016/j.jsb.2003.10.010.

[42] S.J. Unterholzner, B. Poppenberger, W. Rozhon, Toxin–antitoxin systems, Mob. Genet. Elem. 3 (2013) e26219. https://doi.org/10.4161/mge.26219.
[43] M. Cummings, A.D. Peters, G.F.S. Whitehead, B.R.K. Menon, J. Micklefield, S.J. Webb, E. Takano, Assembling a plug-and-play production line for combinatorial biosynthesis of aromatic polyketides in Escherichia coli, PLOS Biol. 17 (2019) e3000347. https://doi.org/10.1371/journal.pbio.3000347.
[44] F.-J. Vorhölter, S. Schneiker, A. Goesmann, L. Krause, T. Bekel, O. Kaiser, B. Linke, T. Patschkowski, C. Rückert, J. Schmid, V.K. Sidhu, V. Sieber, A. Tauch, S.A. Watt, B. Weisshaar, A. Becker, K. Niehaus, A. Pühler, The genome of Xanthomonas campestris pv. campestris B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis, J. Biotechnol. 134 (2008) 33–45. https://doi.org/10.1016/j.jbiotec.2007.12.013.
[45] A.L. Demain, M. Newcomb, J.H.D. Wu, Cellulase, Clostridia, and Ethanol, Microbiol. Mol. Biol. Rev. 69 (2005) 124–154. https://doi.org/10.1128/MMBR.69.1.124-154.2005.