Elucidation of gene regulatory mechanisms in ageing bacterial colonies and tool development for cell factory optimization

Lauritsen, Ida

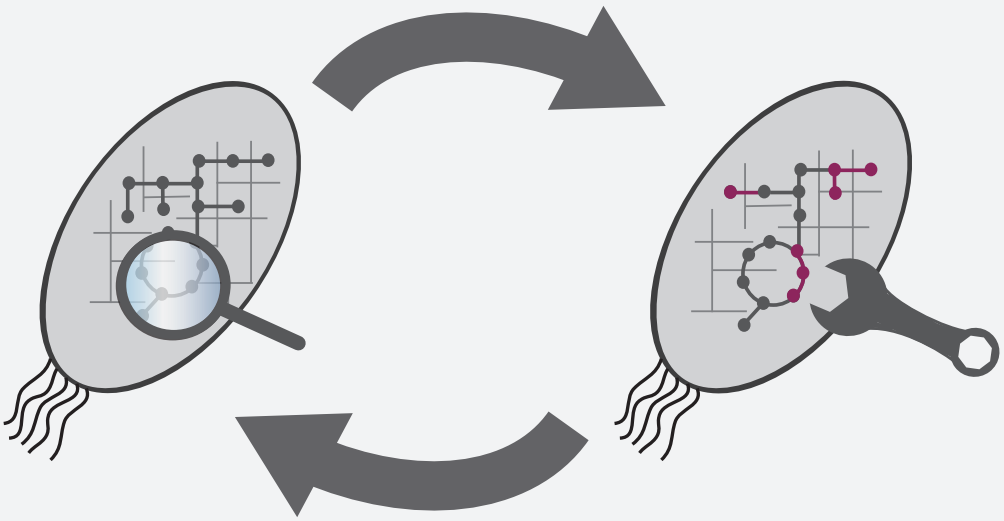
Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Elucidation of gene regulatory mechanisms in ageing bacterial colonies and tool development for cell factory optimization

Ida Lauritsen

PhD thesis
January 2020



DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

Elucidation of gene regulatory mechanisms in ageing bacterial colonies and tool development for cell factory optimization

PhD thesis

Ida Lauritsen

Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

January 2020

Elucidation of regulatory mechanisms in ageing bacterial colonies and tool development for cell factory optimization

PhD thesis written by Ida Lauritsen
Supervised by Senior Scientist Morten H. H. Nørholm

Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark

To Freja and Ludvig,

"Giv mig Vinter og Hunde
– saa maa I for mig gerne beholde alt andet for jer selv, - Kvinderne med!
– Og man smider sig paa sin Slæde – ligeglad med hvor det bærer hen. –
Herre over sin Dag og Herre over sine Hunde!!!"

Knud Rasmussen (1879-1933)
Danish polar explorer and adventurer

Preface

This thesis is written as a partial fulfillment of the requirements to obtain a PhD degree at the Technical University of Denmark. The work presented in this thesis was carried out between November 2016 and January 2020 at the Novo Nordisk Foundation Center for Biosustainability at the Technical University of Denmark, Kgs. Lyngby and was supervised by Senior Scientist Morten H.H. Nørholm. Funding was provided by the Novo Nordisk Foundation.

Ida Lauritsen
Kgs. Lyngby, January 2020

Abstract

Bacteria are indispensible model organisms for studying fundamental biology and their metabolic pathways are being engineered for the purpose of creating high-performance microbial cell factories and moving towards a more sustainable society. However, due to unpredictability, lack of robustness, and complexity of biological systems, more insight into cellular behavior and evolvability in extreme physiological conditions is required in addition to the development of efficient and flexible molecular engineering tools.

This PhD thesis aims at addressing two sides of biology for improving microbial cell factory performance: applied engineering and fundamental biological knowledge. The possibilities to regulate processes that convert DNA to protein and bacterial genome editing technologies are explained as well as regulatory mechanisms of sugar utilization and the role of the well-studied global transcription factor, the cyclic AMP receptor protein (CRP), in the model organism *Escherichia coli (E. coli)*. Cellular conditions in stationary phase and transcriptional responses upon starvation are described due to a recent experimental evolution study that revealed CRP to be a mutational hotspot in dormant and carbon-starved bacterial cells.

The work presented in this thesis covers the development of two CRISPR-Cas9-based synthetic biology tools for microbial cell factory optimization that i) can modulate two steps of the information flow from DNA to functional proteins and ii) selectively remove unwanted plasmids used in molecular biology and engineering. The CRiPi technology enabled interference with uninvestigated essential genes from *E. coli* and the one-step curing platform pFREE demonstrated efficient curing in both *E. coli* and the soil-bacterium *Pseudomonas putida*. In a manuscript of this thesis, CRP activity was inhibited by the pyrimidine nucleosides, cytidine and uridine, that accumulate during carbon starvation. The surprising new connection between CRP and the pyrimidine metabolism suggests an alternative way to reduce global CRP-cAMP-dependent transcription for balancing nitrogen and carbon metabolism in response to carbon deprivation and ageing. We envision that the work presented in this thesis can provide the engineering tools and basis for further investigations of regulatory mechanisms in changing environmental conditions that hopefully can be explored for designing improved and more stable microbial cell factories in the future.

Dansk resumé

Bakterier er uundværlige modelorganismer for at undersøge fundamentale biologiske spørgsmål og deres metaboliske veje bliver ændret for at lave produktive mikrobielle cellefabrikker og bevæge os mod et mere bæredygtige samfund. Forskellige aspekter ved disse cellefabrikker kan være vanskelige såsom uforudsigelighed, manglen på robusthed og stabilitet samt kompleksiteten ved at ændre et biologisk system. Derfor er der et behov for at få indblik i de cellulære mekanismer og tendenser til at mutere for at tilpasse sig ekstreme fysiologiske forhold samt udvikle effektive og fleksible genteknologiske værktøjer og teknologier.

Denne PhD-afhandling har det formål at forbinde to sider af biologien for at optimere mikrobielle cellefabrikker: anvendelsesforskning med henblik på genetiske ændringer for et anvendeligt formål og fundamental biologisk viden i form af grundforskning. Mulighederne for at regulere de processer, som konverterer DNA til funktionelle proteiner og anvendte bakterielle genteknologiske metoder, er forklaret. Derudover beskrives udnyttelsen af sukker som næringsstof samt den regulatoriske rolle af den velstuderede transskriptionsregulator, cyklisk AMP receptor protein (CRP), i modelorganismen *Escherichia coli (E. coli)*. De cellulære forhold i stationærfase og transskriptionelle effekter under sult er også fremlagt, da et tidligere studie baseret på eksperimentel evolution viste, at CRP oftest er muteret i sovende og karbon-sultede bakterieceller.

De projekter, som er præsenteret i denne afhandling, omhandler blandt andet udvikling af to CRISPR-Cas9-baserede syntetisk biologi-værktøjer til at optimere mikrobielle cellefabrikker, der kan i) manipulere to steder i informationsflowet fra DNA til funktionelt protein og ii) selektivt fjerne uønskede plasmider, som bruges inden for molekylærbiologien og genetisk redigering. Med CRiPi-teknologien blev ustuderede essentielle gener undersøgt fra *E. coli* og plasmid-fjernelsesteknologien pFREE demonstrerede effektiv fjernelse af plasmider både i *E. coli* og jordbakterien *Pseudomonas putida*. Et manuskript i denne afhandling viste, at CRPs aktivitet var hæmmet af pyrimidin nukleosiderne cytidine og uridine, som ophobes under sult. Den overraskende nye forbindelse mellem CRP og pyrimidinstofskiftet er en alternativ måde at reducere den globale CRP-cAMP-afhængige transskriptionelle regulering for at balancere nitrogen- og karbonstofskiftet under aldring og under de forhold, hvor der mangler

næringsstoffer. Vi håber, at arbejdet, som er præsenteret i denne afhandling kan bidrage med de genetiske værktøjer og den basisviden, der er nødvendig for at kunne fortsætte undersøgelsen af de regulatoriske mekanismer i skiftende miljøer, som forhåbentlig kan blive udnyttet til at designe forbedrede og mere stabile mikrobielle cellefabrikker i fremtiden.

Acknowledgements

I started as a "miniprepping" student helper in the summer of 2013 at the old CFB in the forest of Hørsholm. Now more than six years later, this PhD thesis is the result of countless hours of work and many good experiences. I want to acknowledge the individuals that have been part of this journey and that I could not have done it without.

Morten, I owe great thanks to you as my supervisor. For allowing me to be part of the research group in the first place, for always having time for discussions and questions with your open door, and for convincing me that basic science is not so bad after all. Not to forgot your infectious enthusiasm for science and all your crazy ideas for new experiments. It has been a true pleasure working with you all these years!

I owe a special thanks to the people that I worked closely with on projects of this thesis. To the *guapa* former postdoc Virginia Martínez, I want to thank for the work on my *CReePy* Russian boyfriend PROTi and for teaching me dirty tricks in the lab to become a true scientist. To Andreas Porse, for the many fun hours working on our pFREE, for being such a passionate and positive scientist that has inspired me, and for always having time for a snack (from your drawer) and a chat. And to Pernille Ott Frendorf, for being my "co-pilot" on our beloved CRP story, for always being up for a coffee drink from the café and a detailed and complex talk about CRP and its involvement in metabolism in the office, and for keeping up the good mood when things were not working in the lab.

I want to thank all former and present members of the Microbial Evolution Synthetic Biology group for being good colleagues, creating a great atmosphere with many laughs and cakes (at snack-o'clock), for help and advice in the lab, and for fruitful discussions and ideas. From the old CFB in the forest: Karina Søgaard, Ulla Christensen, Victor Forman, Emil Fischer, Darío Vázquez Albacete, Sofie Wendel, Mafalda Cavaleiro, Virginia Martínez and Se Hyeuk Kim alias Master Kim. In the new "Gucci-building:" Tonja Hobel, Roberto Ferro, Alexey Dudnik, Silvia Capucci, Maja Rennig, Pernille Ott Frendorf, Kristoffer Bach Falkenberg, Cristina Hernández Rollán, Sophia Heyde, Carolyn Nicole Bayer, Andreas Birk Bertelsen, and Anja Ehrmann. I especially want to acknowledge my two fellow cat lovers Silvia *my Italian mama* and Alexey for spreading joy and good vibes. And Maja, for helping me countless times and

for being the best lab and office companion all these years. You deserve extra credit!

Many people and fun events have contributed to the superb environment at CFB. From the old PhD club, I want to thank Maja, Alicia, Kira, Anne Sofie, Christian (Mette), and Kristian (Paul). The CFB Christmas Party Committee of 2016 that I was part of also deserves to be recognized for all the fun and hard work: Michael, Ruben, Søren, Kristian, Kira, and Anne Sofie. I also want to express my gratitude to other CFB people that have contributed to the great and positive atmosphere: Rosa, Pep, Alicia, Klara, Ruggero, Jennifer, Patricia, Mikkel, Isotta aka Miss Sunshine, Sheila, Hemanshu, Alex, Elleke, Torbjørn, Jenny, Gonzalo, Felipe, David, Carola, Emil (skævøje), and many others. And thanks to Rebeca Thostrup and Susanne Plougheld Winther for helping with the administrative part of my PhD.

I will forever be grateful for my time at Landsbyskolen with the most amazing teachers, Stig Jørgensen, Christina Smidt Christensen, and Torben Falkenberg among others - the best place in the world to learn, where you can truly be yourself. And especially thanks to Stig Jørgensen for opening my eyes to the fascinating field of biology. Benny Mouzinsky, you invited my parents and I to East Greenland CNP – the most beautiful place on each - and was a huge inspiration for us further on. I sincerely wish you could be here today, you for sure would be very proud of me! ;)

At last but not least, I owe the biggest thanks to my family! The human incubator and sniper Móðir, with our shared love for Iceland and the Blue Lagoon, and Wildfire P. Arkinson, my travel campaign by sea and car. Freja (Bubbu), Ludvig (Luddi), Coco Paninnguaq, Baldus, Helle Littlepillow, Oliver, Mikkel, Mikisoq and Maxine, you all mean the world to me.

List of publications

Overview of scientific articles published or in preparation as result of this thesis:

* Denotes equal contribution
# Reprint was made with permission from publisher

1 Martínez, V*., Lauritsen, I*., Hobel, T., Li, S., Nielsen, A. T., & Nørholm, M. H. H. (2017). **CRISPR/Cas9-based genome editing for simultaneous interference with gene expression and protein stability.** *Nucleic Acids Research*, *45*(20), [gkx797]. https://doi.org/10.1093/nar/gkx797#

2 Lauritsen, I., Martínez, V., Ronda, C., Nielsen, A. T., & Nørholm, M. H. H. (2018). **Bacterial Genome Editing Strategy for Control of Transcription and Protein Stability**. In M. Jensen, J. Keasling (eds) *Synthetic Metabolic Pathways* (Vol. 1671, pp. 27-37). Methods in Molecular Biology. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7295-1_3#

3 Lauritsen, I*., Porse, A*., Sommer, M. O. A., & Nørholm, M. H. H. (2017). **A versatile one-step CRISPR-Cas9 based approach to plasmid-curing**. *Microbial Cell Factories*, *16*, [135]. https://doi.org/10.1186/s12934-017-0748-z

4 Lauritsen, I., Kim, S. H., Porse, A., & Nørholm, M. H. H. (2018). **Standardized Cloning and Curing of Plasmids**. In J. Braman (ed.), *Synthetic Biology* (Vol. 1772, pp. 469-476). Methods in Molecular Biology. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7795-6_28#

5 Frendorf, P. O*., Lauritsen, I*., Sekowska, A., Danchin, A., & Nørholm, M. H. H. (2019). **Mutations in the Global Transcription Factor CRP/CAP: Insights from Experimental Evolution and Deep Sequencing.** *Computational and Structural Biotechnology Journal*, *17*, 730 736. https://doi.org/10.1016/j.csbj.2019.05.009

6 Lauritsen, I*, Frendorf, P.O*., Capucci, S*., Wendel, S., Fischer, E.C., Sekowska, A., Danchin, A. & Nørholm, M. H. H. (2020) **Temporal evolution of master regulator Crp identifies pyrimidines as catabolite modulator factors.**
Manuscript submitted

7 Lauritsen, I., Nørholm, M. H. H. (2017). **CRISPR-Cas9 - den revolutionerende gensaks.** *Dansk Kemi*, 98, nr. 8, page 14-17
Popular science article (in Danish)

Publications not included in this thesis

8 Muheim, C., Götzke, H., Eriksson, A. U., Lindberg, S., Lauritsen, I., Nørholm, M. H. H., & Daley, D. O. (2017). **Increasing the permeability of Escherichia coli using MAC13243**. *Scientific Reports*, *7*, [17629]. https://doi.org/10.1038/s41598-017-17772-6

Abbreviations and nomenclature

8-oxo-G	8-oxo-guanine
αCTD	C-terminal domain of RNAP α subunit
α-KG	α –ketoglutarate
αNTD	N-terminal domain of RNAP α subunit
ALE	Adaptive Laboratory Evolution
AR	activating region
cAMP	cyclic adenosine monophosphate
CAP	catabolite activator protein
CCR	carbon catabolite repression
CMF	catabolite modulating factor
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats
CRISPRi	CRISPR interference
CRP	cyclic AMP receptor protein
CyaA	adenylate cyclase
CytR	cytidine regulator
dCas9	dead Cas9 endonuclease
dNTPs	deoxyribonucleotides
Dps	DNA-binding protein from starved cells
dsDNA	double-stranded DNA
E. coli	*Escherichia coli*
GASP	growth advantage in stationery phase
goi	gene of interest
gRNA	guide RNA
HTH	helix-turn-helix
IM	inner membrane
KO	knockout
MAGE	Multiplex Automated Genome Editing
NGS	next-generation sequencing
NTPs	nucleoside triphosphates
OM	outer membrane
P. aeruginosa	*Pseudomonas aeruginosa*
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
ppGpp	guanosine tetraphosphate
P. putida	*Pseudomonas putida*
pppGpp	guanosine pentaphosphate
PTS	phosphoenolpyruvate-carbohydrate phosphotransferase system
RBS	ribosome binding site
Rif	rifampicin
RMF	ribosomal modulation factor
RNAP	RNA polymerase
ROS	reactive oxygen species
rRNA	ribosomal RNA
S. cerevisiae	*Saccharomyces cerevisiae*
SEVA	Standard European Vector Architecture

ssDNA	single-stranded DNA
TCA	tricarboxylic acid
TF	transcription factor
TFBS	transcription factor binding site
TIR	translational initiation region
tRNAs	transfer RNAs
TSS	transcription start site
UMP	uridine monophosphate
Vfr	virulence factor regulator

List of Figures

List of Tables

Table of contents

Introduction and thesis outline

There are more bacteria on earth than stars in the universe[1,2]. Bacteria thrive in various habitats with changing environmental conditions such as soil, water and hot springs, and inside bodies of humans and animals due to their versatile metabolisms[3]. Fundamental studies of bacteria have been conducted to understand general concepts of living organisms and cellular machinery and metabolic pathways are now being engineered and rewired for large-scale productions and biosustainable solutions[4]. The engineering of cell metabolism can lead to unpredictable physiological changes that require further investigations to understand[5].

The work conducted in this PhD thesis aims at understanding cellular conditions in extreme physiological conditions such as starvation in dormant bacteria and developing synthetic biology tools for the engineering of biological systems to improve microbial cell factory performance. The thesis is divided in three main introductory chapters where general concepts of biotechnology and synthetic biology tools, regulation of sugar uptake and utilization, and cellular conditions in stationary phase and upon starvation are described.

Chapter 1 explains the concepts of cell factories, synthetic biology, *Escherichia coli* (*E. coli*) as a model organism, and the importance of fundamental studies for engineering and *vice versa*.

Chapter 2 describes different possibilities to modulate the cellular processes that convert DNA to protein with emphasis on technologies to control protein abundance. Bacterial genome editing approaches such as CRISPR-Cas9 (Clustered Regulatory Interspaced Short Palindromic Repeats and its associated endonuclease, Cas9)-assisted recombineering with subsequently curing of plasmid-borne engineering systems are presented.

Chapter 3 presents the basics of transcriptional initiation with focus on regulation by the global transcription factor cyclic AMP receptor protein (CRP) from *E. coli* and its role in the regulatory mechanism of sugar uptake and utilization. The coordinated events and signals of starvation that lead to growth arrest in stationary phase will be described as well as the application of experimental evolution for studying carbon–deprived ageing cells and a surprising connection to pyrimidine metabolism.

The work of this thesis is presented in the form of submitted manuscripts or published articles and book chapters. Two CRISPR-Cas9-based synthetic biology tools were developed to expand the engineering toolbox. The first technology allows for control of gene expression by inducible repression and stability of proteins by conditional proteolysis (Paper 1 to 2). The other technology offers a one-step curing platform of vectors and plasmids used in molecular biology (Paper 3 to 4). An overview of the applications of the CRISPR-Cas9 system for microbial cell factory optimization was published as a popular science article in Danish (Paper 7).

To deepen the understanding of CRP as a central part of carbon metabolism in *E. coli*, engineered and evolved mutants from studies during the past 50 years were reviewed in Paper 5. Paper 6 presents the newly discovered role of CRP as a sensor of pyrimidine nucleosides that accumulate during growth arrest and carbon starvation.

Chapter 1 - From understanding to engineering biology for saving the world

1.1 Early ages of molecular biology

Historically, biologists have pursued to understand life as it exists. Groundbreaking discoveries have led to the foundation of scientific fields such as the decryption of the DNA helix structure[6], the genetic code[7], and the central dogma that explains the molecular flow from DNA-encoded information to functional proteins[8]. From the study of the *lac operon* in *E. coli*, Jacob and Monod deduced the existence of regulatory circuits that respond to extracellular signals such as nutrients in the environment[9]. We learned from these findings the molecular fundamentals of "reading" DNA and the transcriptional mechanisms behind. The discovery of controlled gene expression became the root of molecular biology and biological engineering[10]. Nowadays, we have moved beyond describing *what is* in traditional biology to engineering it to what it *potentially could be*.

1.1.1 *Escherichia coli*

The central *workhorse* involved in most scientific milestones in biology is the Gram-negative rod-shaped bacterium *E. coli*[11]. This microorganism is a natural constituent of the mammalian gut microbiome. However, due to toxin production and infectious traits by some pathogenic species, the relationship with humans is not solely positive[12]. With a rapid doubling time of 20 minutes[13], ease of handling, the ability to grow happily with and without oxygen as a facultative aerobe, and broad substrate breadth[11], this microbe has been the cornerstone of fundamental findings in bacterial physiology, genetics[14–16], and adaptive evolution[17,18] among many other things. The obtained knowledge and developed molecular tools for investigating and manipulating the biology of *E. coli* in laboratory settings, as well as the more recent characterization and engineering of its genome and metabolism, has made it the most well-studied model organism[12] (Figure 1). As one of the first microorganisms, the complete genome of the laboratory model strain K-12 was sequenced in 1997[19].

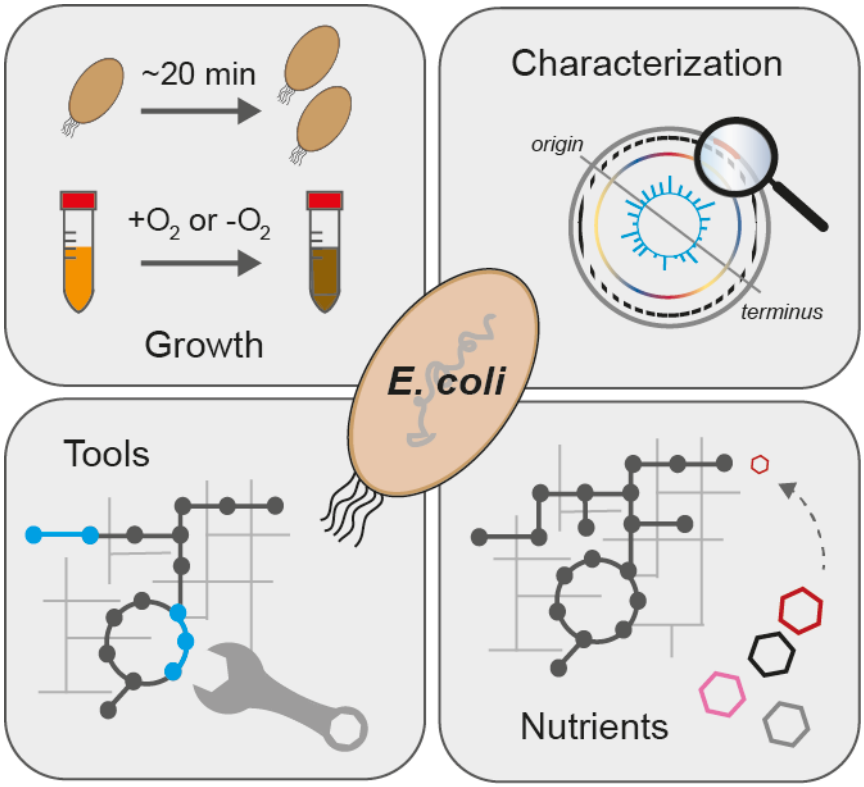


Figure 1 – The model organism *E. coli*. Properties of *E. coli* as model organism include fast doubling time, ability to growth with and without oxygen, characterized and annotated genome and metabolism, available and flexible tools and methods for engineering, and broad substrate range. Inspired by Blattner *et al.*[19] and Petzold *et al.*[20]

1.2 Modern biotechnology

Broadly, the term biotechnology covers the usage of living organisms for the production of valuable products for human purposes. Traditional biotechnology involved e.g. crop breeding, beer brewing with fermenting yeast such as *Saccharomyces cerevisiae* (*S. cerevisiae*), and cheese production aided by lactic acid producing bacteria[10]. With the expansion of molecular technologies throughout the 1970's and 80's such as restriction enzyme based cloning[21,22], the polymerase chain reaction (PCR)[23–25], and Sanger sequencing[26], the basis for more advanced biotechnology with targeted genetic manipulations was founded[27,28]. Cohen and colleagues performed the first restriction enzyme based DNA cloning in 1973[21] and the hormone somatostatin was the first recombinant protein to be produced in *E. coli* in 1977[29].

Currently, automated next-generation sequencing (NGS) and large-scale high-throughput 'omics' datasets of e.g. mapped genomic information (genomics), levels of RNA transcripts (transcriptomics), or proteins abundances (proteomics) create comprehensive pictures of

metabolic states and shed light on unknown biological interactions[28,30]. The obtained data is fed into computational models to enhance predictability of biological systems[31,32]. The revolutionary technology of the CRISPR-Cas9 system has proven to be a biotechnological breakthrough for efficient genome editing by enabling programmable genomic cleavage in various organisms[33–35]. With the CRISPR-Cas9 system, a variable 20-nucleotide sequence of a guide RNA (gRNA) called the protospacer determines the target DNA by sequence complementarity. The target DNA must be located next to a protospacer adjacent motif (PAM) that dictates Cas9 sequence recognition and cleavage[33] (Figure 2). As for restriction enzymes, the CRISPR-Cas9 system derives from bacterial adaptive immune systems and has been hijacked and adapted for engineering purposes due to its specificity and flexibility[36–38].

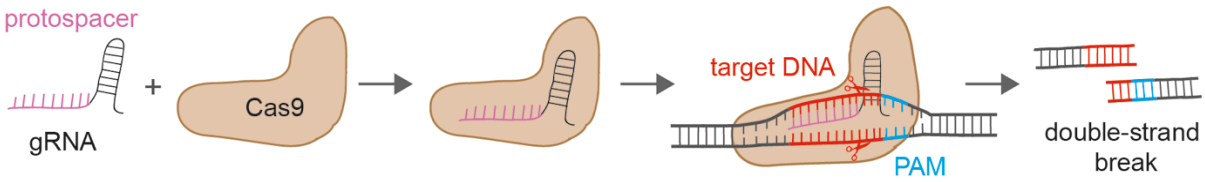


Figure 2 – CRISPR-Cas9 induces double-strand breaks in DNA. The Type II CRISPR-Cas9 system that originates from the bacteria *Streptococcus pyogenes* has been adapted for site-specific DNA double-strand breaks[33]. The Cas9 endonuclease is directed to the target DNA (highlighted in red) by the variable 20-nucleotide region called the protospacer of the gRNA (highlighted in pink) using RNA-DNA complementarity base-pairing rules. The target DNA must be located adjacent to a three-nucleotide PAM sequence of the canonical form 5'-NGG-3' (shown here as blue) that dictates Cas9 cleavage. The gRNA complexes with Cas9 and causes a blunt-ended DNA double-strand break three nucleotides upstream of the PAM sequence[39]. Adapted from Karlgren *et al.*[40]

1.3 Towards a sustainable society one cell factory at a time

The ongoing depletion of stored fossil fuel reserves has a tremendous environmental impact in terms of global warming, ice reservoirs melting, and animal species becoming extinct[41]. Additionally, uncertainties regarding the complexity and biological consequences of emitted chemicals and plastics in natural ecosystems cause growing concerns[42]. These negative environmental changes motivate to transit from petrol and oil-based industries to more sustainable bio-based processes. One approach is the use of microbial cell factories for biosustainable productions, ideally using renewable carbon sources and waste products as substrates[43–45].

The term *cell factory* refers to the use of organisms as a factory unit with the overall goal of producing proteins, biofuels, bulk chemicals, or other compounds of interest. Microorganisms such as *E. coli* and *S. cerevisiae* are especially suited for this task with rapid growth rates, scalability, versatile metabolism, and possibility of making high-density cultures in fermentations for high yield productions[46,47]. The rewiring of native metabolisms by directed genetic engineering to change metabolic fluxes or endowing cells with new abilities can improve productions[48,49]. Cell factory performance can be tested, evaluated, and re-designed in repetitive design-build-test-learn cycles until the optimal performance for the desired purpose has been achieved[50]. Two classic examples of microbial cell factory derived recombinant protein products are human insulin and penicillin, both relevant for the medical industry and beneficial for humankind[51].

1.3.1 Principles of synthetic biology

Synthetic biology, as an emerging scientific discipline, has modified the definition of biological systems, such as cell factories, by interpreting all organisms as biological networks compromised of standardized genetic *parts*[28,52]. In the model organism or *chassis*, these parts are assembled in circuits according to engineering principles and designs. A chassis is here considered the cellular container that provides the general structures and necessary cellular machinery to execute required functions[53]. The advances in DNA assembly methodologies[54–56] and a continuous reduction in DNA synthesis prices have enabled complete tailor-made cell

designs from scratch[57,58]. Two examples are the successful synthesis and assembly of a complete genome in viable *Mycoplasma* cells[59] and the subsequently genome minimization to define the extract essential genetic components for bacterial life[60]. Very recently, a synthetic genome with recoded codon usage was synthesized and functionally implemented in *E. coli*[61]. Collections and sharing of standardized genetic parts or modules are of great value for "plug-and-play" *de novo* constructions and for the vibrant synthetic biology community[28,62,63]. The Standard European Vector Architecture (SEVA) platform is an example of a repository of vectors with standardized compositions that can be received upon request as part of an open-source philosophy and is constantly expanded by the community[64–66].

1.3.2 Build it to understand it or *vice versa*?

The field of biology is expanding from focusing on natural living organisms, shaped by evolution, to the bottom-up development of synthetic ones[4]. Dictated by the concept of synthetic biology, all biological systems can be dissected down to functional modules that can be recombined into improved or new systems with molecular engineering tools[67]. Undefined biological functions and interactions of parts, however, can affect cell factory performance, making predictability difficult[68]. Additionally, long-term stability and robustness of biological circuits can be challenging in changing environments or metabolic conditions such as nutrient starvation, depicted in the accumulation of adaptive mutations[69,70]. Thus, elucidating the behavior and evolvability of biological systems in extreme physiological situations is needed for improved cell factory design and construction (Figure 3).

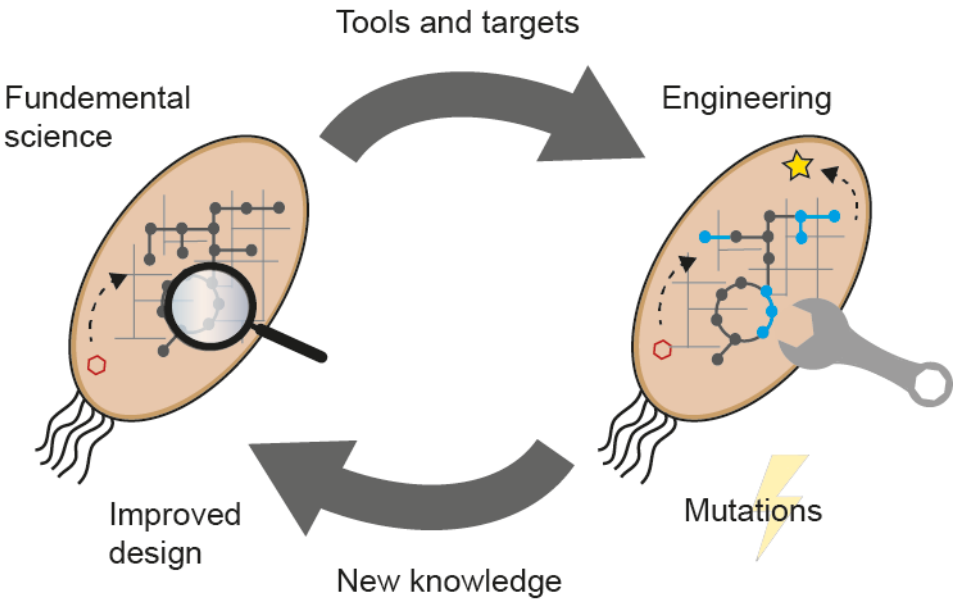


Figure 3 – Connection between basic and applied science. Knowledge of basic cellular components and mechanisms has enabled the development of tools and revealed interesting engineering targets. The engineering of existing metabolic pathways or synthetic *de novo* constructions create biological systems with new abilities that often show lack of long-term stability and accumulation of mutations. Understanding specific mutational hotspots and cellular behavior in different conditions can be applied for improving microbial cell factory design.

As a discipline, synthetic biology bridges basic and applied research. Knowledge of fundamental mechanisms and available tools are needed for the engineering of a biological system and by engineering, constructing and testing, the fundamental understanding of the same system increases[4]. From such studies we can discover that even "old" and well-established understandings of cellular functions are not completely deciphered. In Chapter 2 and 3 of this thesis, the development of technologies and tools that enable engineering of biological systems for cell factory optimization and basic research are presented. Additionally, the fundamentals of carbon utilization and the discovery of the global transcription factor CRP as a starvation stress sensor are addressed.

Chapter 2 - The toolbox for engineering biology

2.1 The biological machinery as targets for engineering

Engineering technologies are fundamental for the genetic manipulation of metabolic networks for improving cell factory performance[71,72] or for fundamental studies e.g. by enabling gene knockout (KO) studies[73]. Pivotal molecular details of biological systems are the basis of most synthetic biology tools and results obtained from engineering studies can aid further tool and technology development[50,52]. A variety of tools and methods allow for controlling different steps of the central dogma, such as transcriptional initiation, the post-transcriptional, translational, and post-translational level, to tune the information flow from DNA to RNA to protein[74].

A traditional method to eliminate the effect of native genes and encoded proteins is by KO approaches through the insertion of an antibiotic resistance cassette and subsequent removal at the genomic locus[73,75]. More sophisticated strategies cover tuning of gene expression levels e.g. via RNAP (RNA polymerase) recruitment or promoter engineering[76–78]. Complete transcriptional inhibition can e.g. be achieved with the CRISPR interference (CRISPRi) technology where a catalytically dead Cas9 endonuclease (dCas9) sterically hinders RNAP binding at promoter regions and turn off genes[79]. The ribosome binding site (RBS) or the translational initiation region (TIR) can be randomized and variants for low, medium, or high protein production can be selected for since translational initiation is a rate-limiting step in protein synthesis[80–82]. At the post-translational level, the presence of degradation tags can enhance protein turnover in a temporal manner[83] (see section 2.1.2). Figure 4 depicts an overview of targets for modulating the information flow from DNA to RNA to protein.

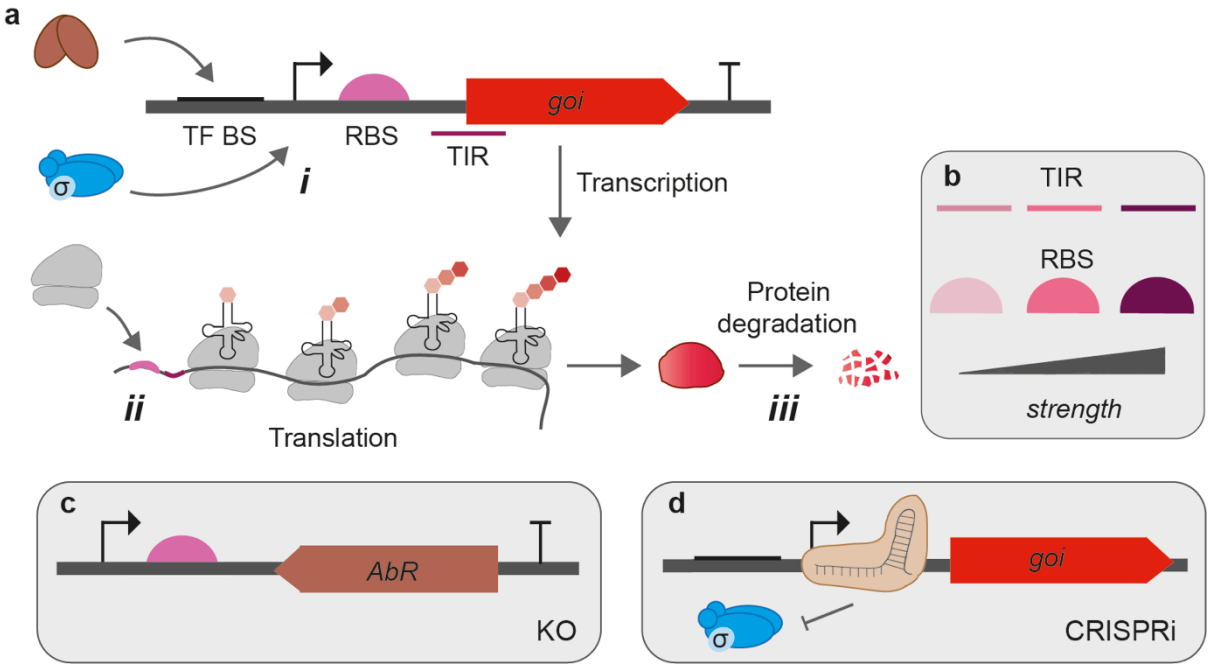


Figure 4 – Targets for modulating the information flow from DNA to protein. a. By modulating the processes involved in transcription, translation, or protein degradation, the levels of functional protein encoded by a gene of interest (goi) can be altered. ***i)*** For transcriptional initiation, RNAP (blue) recruitment at a specific promoter can be adjusted e.g. by manipulating binding of transcription factors (brown) at transcription factor binding sites (TF BS, black). ***ii)*** At the post-transcriptional level, binding of the ribosome (grey) to mRNA can be altered by varying the ribosome binding site regions (RBS, light pink) or translational initiation regions (TIR, dark pink). ***iii)*** Protein turnover can be enhanced by tunable protease expression to expose native degradation signals. **b.** The nucleotide sequence of the RBS region (up to 35 nucleotides upstream of the start codon[84,85]) or the TIR (six nucleotides upstream of the start codon and the two positions following the start codon[86]) can be randomized and variants with different strength can be screened or selected for e.g. by translationally-coupled antibiotic resistance[81]. **c.** KOs can be performed to disrupt gene function with the insertion of a selective antibiotic resistance gene (AbR) that is removed subsequently. **d.** With the CRISPRi technology, transcriptional repression can be performed by dCas9 binding at promoter regions to hinder RNAP recruitment. Adapted from Kent and Dixon[74].

2.1.1 CRISPR-Cas9-assisted λ-Red recombineering

Genomic expression of genes, operons or metabolic pathways provides stable and low-copy expression, which can be desirable for e.g. long-term protein productions as it reduces antibiotic and metabolic burden on the cell[87]. For this purpose, genome editing techniques have been developed to enable precise and direct modifications such as single-nucleotide mutations, deletions, or insertions of small sequences as well as large multi-gene fragments at genome-level[88]. In prokaryotes and especially in *E. coli*, most of these technologies rely on λ-

Red recombineering with the phage proteins Beta, Gam, and Exo via homologous recombination (Figure 5). For the integration of large double-stranded DNA (dsDNA) templates such as PCR products, all three proteins are needed[89]. Single-stranded DNA (ssDNA) is generated by the exonuclease activity of Exo[90,91], Beta proteins are believed to protect the ssDNA from degradation and deliver it to the lagging strand of the replication fork[92,93], and Gam inhibits the endogenous nucleases RecBCD and SbcCD that degrade dsDNA fragments[94,95]. Only the Beta protein is required for simplified ssDNA oligonucleotide-based recombineering as in Multiplex Automated Genome Editing (MAGE) that involves repetitive cycles of recombineering for simultaneous small modifications across multiple genomic loci[96,97].

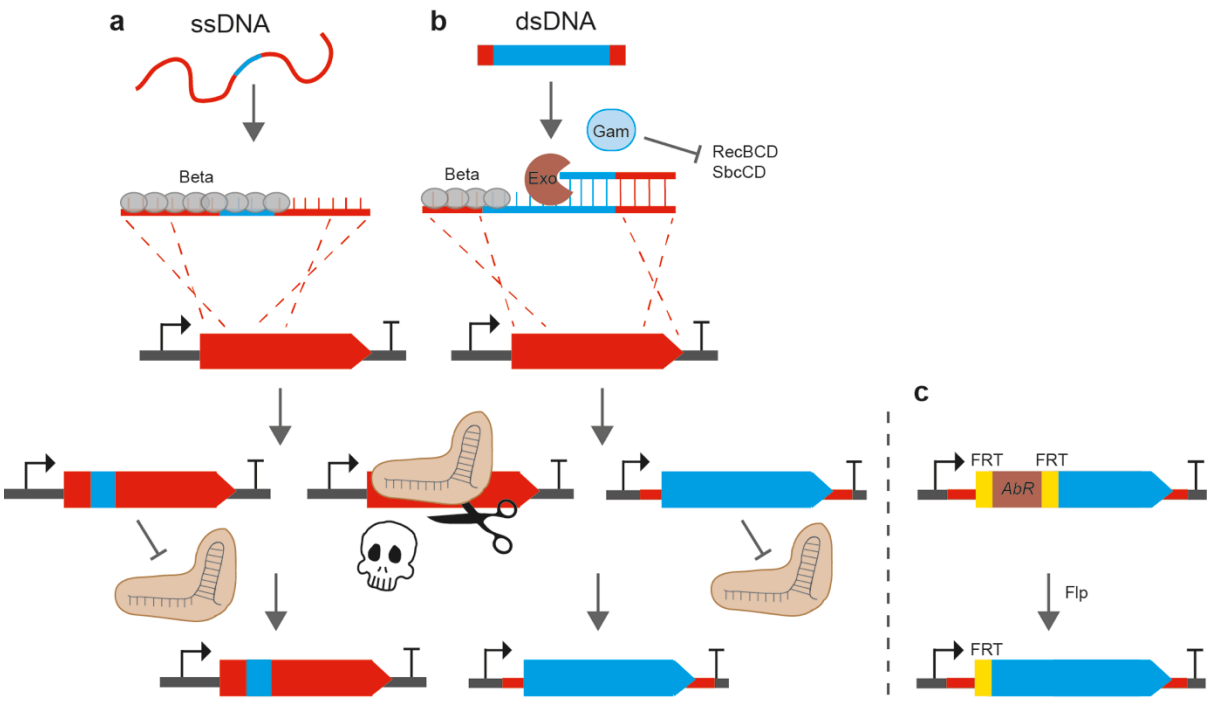


Figure 5 – CRISPR-Cas9-assisted λ-Red recombineering. a. For λ-Red recombineering with ssDNA oligonucleotides as templates, Beta proteins protect the template and guide to the replication fork for integration. This type of recombineering is most efficient for small modifications[96] (here shown as the small blue part of the red gene). **b.** The Beta, Exo, and Gam proteins are required for λ-Red recombineering with dsDNA as templates (here shown as a PCR product encoding an entire gene in blue). After the recombineering process, the endonuclease Cas9 is targeted to the wild-type sequence by the gRNA, leading to lethal double-strand breaks in the genome of the non-engineered clones. The engineered clones have been genomically edited and are not recognized by Cas9. The CRISPR-Cas9 counter-selection method is markerless and increases the likelihood of identifying correct engineered clones. **c.** An antibiotic resistance gene (AbR) flanked by e.g. FRT sites (yellow) can be used for selection of successful integrants. However, after excision with the Flp recombinase, scars are left in the genome. Inspired from Sharan *et al.*[98] and M. Rennig[99]

A major bottleneck in λ-Red recombineering is the identification of successfully modified clones due to low recombination efficiencies[100,101]. To aid the screening process, CRISPR-Cas9 counter-selection can be applied. In *E. coli*, the Cas9 endonuclease mediates lethal site-specific double-strand breaks in the DNA of the non-engineered clones, thereby enriching for the genomically modified cell population[101]. The use of the CRISPR-Cas9 system as a counter-selection method for recombineering offers markerless genomic modifications[102] (Figure 5). In comparison, dsDNA genomic co-integrations of selectable markers such as antibiotic resistance cassettes, which are flanked by recombinase target sites (FRT/loxP) for recombinases (Flp/Cre), leave scars after excision[75,98,103]. The combination of λ-Red recombineering and the CRISPR-Cas9 systems facilitates rapid and efficient genome editing of both large genetic elements such as complete metabolic pathways[102,104,105], and smaller changes such as RBSs, promoters, or single amino acid exchanges at single or multiple genomic loci[96,106–108].

2.1.2 Degron-based protein degradation technologies

Manipulating intracellular protein levels by changing protein stability has many applications such as studying individual protein functions or controlling protein levels in metabolic pathways[109,110]. Depletion of functional proteins relies on natural dilution and slow degradation rates due to long half-lives of most proteins[111]. Furthermore, essential proteins serve vital function in the cell and cannot be studied or eliminated in pathways by a classical reverse-engineering approach without impairing viability[73]. Thus, technologies that allow accurate, inducible, and rapid control of intracellular protein abundance are desired.

Many of the developed tools for targeted proteolysis depend on the attachment of engineered degradation tags to speed up protein depletion[83,110,112–114]. Degradation tags or signals (degrons) are features of a protein that determines its half-life in both eukaryotes and prokaryotes. In *E. coli*, a certain category of degrons (N-degrons) relies on the identity of the N-terminal residue for proteolytic recognition and targeting to the ClpAP proteasome complex by the adaptor protein ClpS in the so-called N-end rule pathway[115–119]. Genomic integration of small engineered degradation tags at the locus of a gene of interest can be achieved with genome editing technolgies such as CRISPR-Cas9-assisted recombineering[106]

(see section 2.1.1). The engineered degradation tag must ideally not affect the native functionality of the protein or destabilize protein before induction of the degradation process[110]. Thus, the degron must initally be masked and exposed upon induction e.g. by protease-cleavage at a protease recogniztion site encoded as part of the engineered degradation tag [83,120].

In Paper 1 of this thesis, an inducible protease-dependent protein degradation technology (PROTi) was combined with CRISPRi-mediated transcriptional repression into the CRiPi system that allows for comprehensive studies of knockdown phenotypes including essential genes. The CRiPi technology was converted into a user-friendly protocol and published as a book chapter for applications in research laboratories (Paper 2).

2.2 Vector-based expression systems

The recent advances in genome-editing technologies enable modifications to regulatory parts of genes or operons in their native genomic context and the exchange of genetic elements or entire pathways. Nevertheless, plasmid vectors are essential for heterologous expression, the basis for DNA cloning and molecular biology, easily modified, and transferable vehicles between bacterial strains[121,122]. Unlike the chromosomal DNA, plasmids are mobile non-essential DNA elements with independent replication abilities that are believed to carry genetic features for short-term adaptation in natural environments[123]. Expression levels can vary dependent on the nature of the plasmid e.g. dictated by the origin of replication with copy-numbers ranging from low, medium to high[124]. Due to the nonessential nature, plasmids often require cellular maintenance with antibiotic resistance marker selection[125]. Impediment of normal cellular and metabolic functions by overexpression of plasmid-borne genes and antibiotic selection can cause toxicity and metabolic overload[65,126].

From an engineering point of view, plasmids are useful shuttles of e.g. biosensor systems for characterization purposes, modules of circuits, or technologies such the λ-Red proteins and the CRISPR-Cas9 system for successful genome editing. In our laboratory, the implementation of the CRISPR-Cas9-assisted recombineering technology facilited efficient genomic modifications and faster identification of correct clones. However, when desired

genetic changes had been obtained, the subsequent removal of the multiple plasmid-borne system used for genome editing was problematic. Even though most plasmids impose a metabolic burden that can make curing favorable, some vectors can be stably maintained under nonselective conditions, making disposal challenging[127–129]. In Paper 3 of this thesis, a CRISPR-Cas9-based tool for plasmid curing was developed, the pFREE system, that targets the most applied vectors in molecular biology. Additionally, Paper 4 of this thesis was published as a step-by-step protocol for plasmid curing with the pFREE system.

Chapter 3 - Sugar starvation in ageing bacteria

3.1 Life as bacteria know it

The construction and engineering of biological systems facilitated by synthetic biology tools allow for studying stressful physiological conditions, can reveal unknown biological functions, and lead to considerable advances in basic science[50]. One of the most fundamental scientific questions concerns the basic requirements for life[58]. Nutrients, such as nitrogen and carbon, are imperative for bacterial physiology and make up the foundation for building blocks and energy for metabolic processes. Carbon sources, in the form of sugars, are taken up from the environment and channeled into metabolic networks for the production of cellular energy currencies, ATP, NADH, and NADPH, that fuel the enzymatic reactions, required for proliferation and maintenance[130–132]. Decades of extensive studies have deciphered underlying mechanisms and transcriptional regulation of sugar uptake as a coordinated event of inhibitory effects of glucose and hierarchical consumption of sugars in the model organism *E. coli*[133,134]. Despite of this, some factors and their involvement in global carbon metabolism remain unknown.

3.1.1 Regulation of transcriptional initiation

Transcription is one of the most fundamental cellular processes as part of the central dogma of molecular biology. In this process, information stored in DNA is "read and copied" into a complementary RNA transcript (mRNA) for the eventual translation into functional proteins. Transcription is accomplished in three steps: initiation, elongation, and termination that require the DNA-dependent RNAP. Since Paper 6 of this thesis aims at understanding the impact of a global transcriptional regulator during carbon deprivation, focus here will only be on the regulatory strategies for the initiation phase in *E. coli*.

The core RNAP enzyme (α2ββ'ω) is built up of five subunits: two α-subunits (*rpoA*), β (*rpoB),* β' (*rpoC*), and ω (*rpoZ*) in *E. coli*[135,136]. The α-subunits bind to the template DNA and are primarily involved in the assembly of the core complex[137,138]. The catalytic center of RNA synthesis is formed by the β and β' subunits whereas the ω subunit enforces structural

stability[139]. The core enzyme is able to catalyze transcription, but an additional subunit, the sigma factor (σ), is required for specific promoter sequence recognition, transcriptional initiation, and correct RNAP positioning at the target promoter and DNA unwinding. The sigma factor subunit associates with the core RNAP, thereby forming the holoenzyme that recognizes the -10 and -35 elements at most promoters[140,141]. -10 and -35 refer to the approximate distances in nucleotides upstream of the +1 transcription start site (TSS). Additionally, an extra AT-rich DNA sequence called the UP element or an extended -10 site can increase promoter activity[135,142,143] (Figure 6). *E. coli* has seven different sigma factors. The main sigma factor σ70, encoded by *rpoD,* equips RNAP to recognize most housekeeping genes. Alternative sigma factors are responsible for responses to stressful conditions: stationary phase/starvation σ38/σS (*rpoS*) (see section 3.3.1), low levels of nitrogen σ54 (*rpoN*), heat shock σ32 (*rpoH*), chemotaxis σ28 (*fliA*), unfolded or denatured proteins σ24/ σE (*rpoE)*, and iron starvation σ19 (*fecl*)[141,144–146].

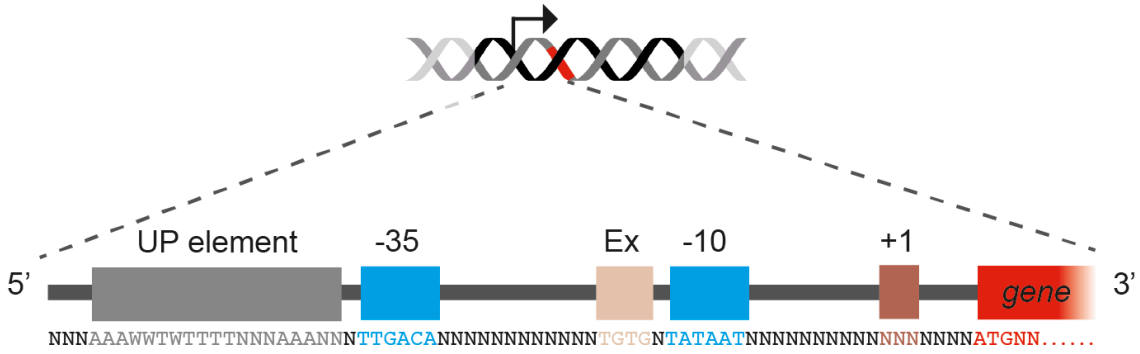


Figure 6 – Schematic illustration of a typical σ70 promoter. The -35 and -10 motifs[147] (blue) are essential for RNAP recognition by the σ70 sigma factor. The AT-rich UP element (grey) and the extended -10 motif (Ex; brown) can improve RNAP activity at some promoters. +1 denotes the transcriptional start site. N: A,C,G,T and W: A or T. The RBS is not shown. Adapted from Gilman and Love[148] and Browning and Busby[149].

Modulation of specific promoter strengths is controlled by transcription factors (TFs)[150,151]. A common characteristic of most TFs is the helix-turn-helix (HTH) motif that facilitates DNA binding[151]. More than 300 TFs have been predicted in *E. coli* with only half experimentally validated[152,153]. TFs can roughly be divided in three types: activators (transcription on), repressors (transcription off), and dual regulators. A correlation between promoter binding site position and regulatory function of TFs has been observed[154]. Most of the transcriptional

activators bind upstream of the +1 TSS to assist RNAP and DNA interaction whereas repressors bind around TSS for hindrance of RNAP binding[150,154]. Activators typically require interaction with ligands to function and inducers can associate with repressors to relieve transcriptional inhibition[151].

Beside the type of regulation function, TFs can be classified according to their regulatory scope[155]. Local TFs have restricted regulons compromising few genes or operons that induce discrete and specific changes like on/off-states[156,157]. Global TFs, on the other hand, cover large regulatory landscapes and networks with genes belonging to different functional groups[158]. Features such as altered DNA structure e.g. bending or looping, specific contact between RNAP and TF, and co-operative functions of TFs can contribute to transcriptional regulation[149,154]. The highly compacted structure of supercoiled chromosomal DNA[159], specific location of target promoters[160], and the influence of nucleoid-associated proteins[161] can alter the accessibility of promoters for RNAP and act as additional control of transcriptional initiation.

In summary, sigma factors, local and global TFs, small effector molecules such as cAMP and (p)ppGpp (see section 3.1.2 and 3.3.2), and DNA organization can influence the transcriptional initiation process and enable transcriptional switching according to environmental signals and metabolic states[150,162].

3.1.2 Carbon catabolite repression

In 1961, Jacob and Monod published the groundbreaking work that explained the transcriptional regulation of the *lac* operon[9]. A few years later in 1970, Bechwith and colleagues identified a key player for *lac* operon expression, called CRP or catabolite activator protein (CAP)[163,164]. The best described role of CRP is its involvement in the *glucose effect* or carbon catabolite repression (CCR), a regulatory phenomenon preventing the expression of enzymes involved in utilization of secondary carbon sources (e.g. lactose) in the presence of a preferred carbon source such as glucose[165–167]. This phenomenon was already manifested in 1942 by Monod in which bacterial cells exhibiting a diauxic shift by hierarchical consumption of lactose and glucose[134].

In CCR, the glucose-specific phosphoenolpyruvate-carbohydrate phosphotransferase system (PTS) is responsible for the uptake of glucose (PTS-sugar). A phosphate group is transferred from phosphoenolpyruvate (PEP) to a series of cytoplasmic PTS proteins - enzyme I, HPr, $EIIA^{Glc}$ and eventually to the membrane transporter $EIIBC^{Glc}$, which simultaneously transports and phosphorylates glucose[168] (Figure 7). The phosphorylation status of the PTS component $EllA^{Glc}$ and the cellular concentration ratio of PEP/pyruvate (intermediates of glycolysis) determine the transport activity[169]. If the PEP/pyruvate ratio is high, the majority of $EllA^{Glc}$ is phosphorylated, enabling activation of adenylate cyclase (CyaA) that converts ATP into cyclic adenosine monophosphate (cAMP). The secondary messenger cAMP complexes with CRP and promotes regulation of various genes, most of them involved in transport and utilization of non-PTS/secondary sugars such as lactose, maltose, and galactose. If the PEP/pyruvate ratio is low and glucose is present, the majority of $EllA^{Glc}$ is in the dephosphorylated form[166,170]. The CCR mechanism ensures maximized growth rates on preferred sugars and improves competitive success in natural environments.

An additional regulatory level of the CCR model involves the inaccessibility of the inducer so-called *inducer exclusion*. As an example, dephosphorylated $EllA^{Glc}$ inhibits LacY (transporter for lactose) by direct binding, thereby preventing induction of the *lac* operon[171] (Figure 7). Regulation of CyaA activity and inducer exclusion is linked together in CCR because any alterations in the phosphorylation status of $EIIA^{Glu}$ will affect both phenomena.

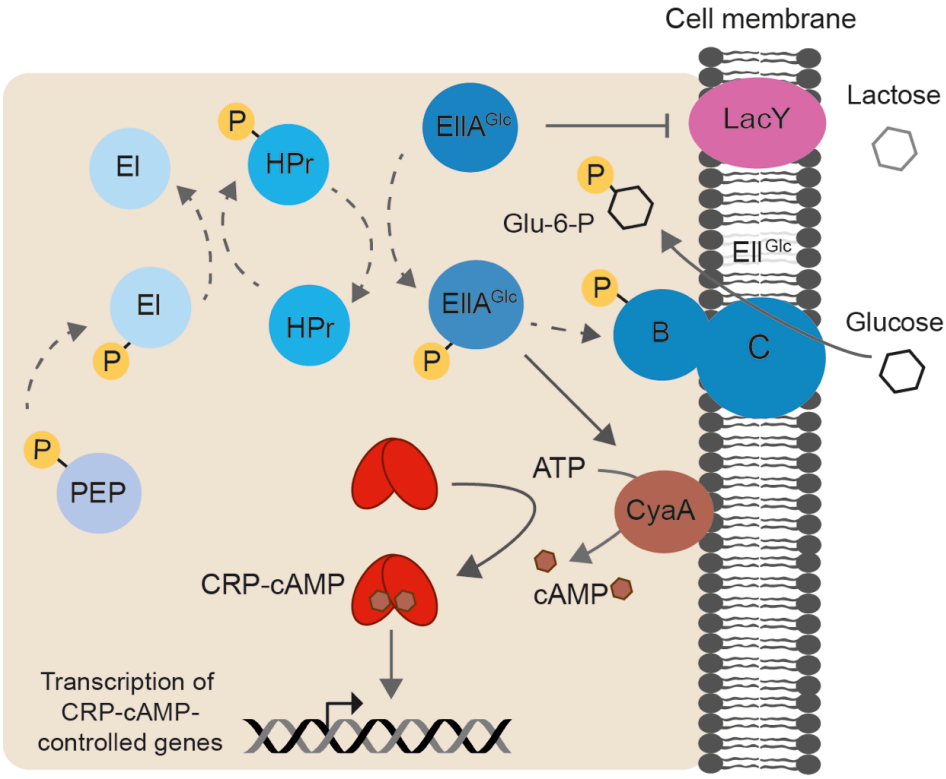


Figure 7 – Carbon catabolite repression. A phosphate group is donated from PEP and transferred through the PTS system (shown in shades of blue) to the membrane-bound EIIBC^{Glc} that transports glucose (black hexagon) across the membrane and phosphorylates it to glucose-6-phosphate (Glu-6-P). Dephosphorylated EIIA^{Glc} inhibits activity of the lactose transporter LacY (pink) and excludes import of the inducer lactose (grey hexagon) that relieves LacI-induced transcriptional repression (inducer exclusion, not shown). Phosphorylated EIIA^{Glc} activates CyaA (brown) and cAMP (brown hexagon) is produced from ATP. CRP (red) binds cAMP and promotes transcription of CRP-cAMP-dependent genes. Dashed lines represent phosphate transfer. Adapted from Görke and Stülke[168]

3.1.3 CRP – more than just a simple transcription factor

CRP belongs to the "exclusive" group of seven global TFs that influence the expression of a large number of transcriptional units in *E. coli*, including many other TFs[158,172] (Figure 8). More than 300 genes have been predicted to be part of the CRP regulon[173,174], including *cyaA* and *crp* that are subject to CRP-mediated transcriptional repression and autoregulation, respectively[175–177]. An overview of CRP structure and a summary of evolved and engineering CRP mutants are presented in Paper 5.

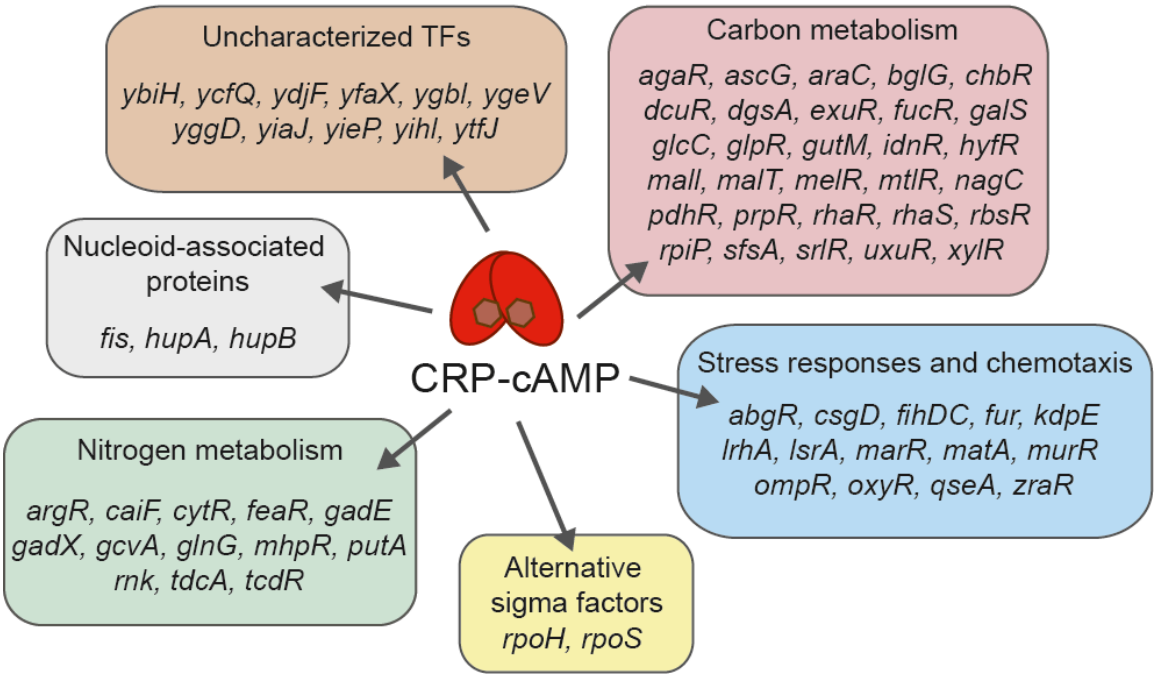


Figure 8 – Transcription factor regulatory network of CRP. CRP directly regulates expression of TFs and regulators that are e.g. involved in carbon metabolism (highlighted in pink), nitrogen metabolism (highlighted in green), stress responses and chemotaxis (highlighted in blue), alternative sigma factors (highlighted in yellow), and nucleoid-associated proteins (highlighted in grey). CRP binding sites have been predicted as part of uncharacterized TFs (highlighted in orange). Autoregulation of *crp* is not depicted. Adapted from Shimada *et al.*[174] with information from the EcoCyc database (Keseler *et al.*[136]) and Martínez-Antonio *et al.*[178]

CRP-cAMP, CRP bound with its allosteric ligand cAMP, binds to the canonical DNA binding motif 5'-TGTGA-N_6-TCACA-3' and functions as a versatile dual regulator with both activating and repressive activities determined by the architecture of specific promoters[174,179–181]. CRP directly interacts with the RNAP holoenzyme through the so-called activating regions (ARs) in the structure of CRP[182] and a HTH motif mediates DNA binding[183]. Three classes of CRP-cAMP-dependent promoters have been assigned. Class I occupies one CRP-

cAMP binding site located upstream of the -35 motif (around -90 to -60) with direct interaction of AR1 and the C-terminal domain of RNAP α-subunit (αCTD). The *lac* promoter, governing the expression from the *lac* operon, is the textbook example of a class I CRP-cAMP-dependent promoter[184–186]. The *malT* promoter also belongs to class I and controls *malT* expression, the transcriptional activator of all maltose utilization genes[187,188]. Like class I, only one CRP-cAMP binding site is present at class II promoters. The binding site overlaps with the -35 element and AR1, AR2, and AR3 are required for RNAP interaction with αCTD, the N-terminal domain of RNAP (αNTD), and the sigma factor[179,189–191]. The best-characterized class II promoter is *galP1*, one of two overlapping promoters that control expression of the galactose operon[192,193]. Class III promoters hold multiple CRP-cAMP binding sites for e.g. synergistic transcriptional activation[150]. The promoter of the mannitol operon repressor, MtlR, is an example of a class III promoter with five CRP binding motifs[194].

CRP-cAMP-based transcriptional activation can be counteracted by *anti-activation.* The repressor, cytidine regulator (CytR), masks AR1 of two DNA-bound CRP dimers and prevents αCTD interaction[195]. The presence of the pyrimidine nucleoside cytidine alleviates CytR-anti-activation and allows again for CRP-cAMP-dependent transcription. Promoters under CytR-control are mainly involved in nucleoside and deoxynucleoside utilization (e.g. *cdd* and *udp*) and transport[196,197]. The expression of *cytR* is induced by CRP-cAMP and repressed by the CytR protein[198]. The CytR-mediated anti-activation mechanism coordinated by cytidine allows for precise adjustment of the transcriptional profile according to DNA/RNA demand and metabolic state of the cell. Pyrimidine nucleotides will be addressed further in section 3.5

Important structural features are shared between CRP-homologues from different bacterial species. However, the physiological roles within each organism and cAMP binding affinities are very diverse due to adaptation to different niches with variations in cAMP availability[199]. CRP from the soil-bacterium *Pseudomonas putida* (*P. putida*) exhibits very high cAMP affinity and is involved in degradation and utilization of certain nitrogen sources[200,201]. The CRP-homologue of the closely related bacterium *Pseudomonas aeruginosa (P. aeruginosa)* called virulence factor regulator (Vfr), functions both dependently and independently of cAMP and the regulatory repertoire involves e.g. virulence factor production, twitching

motility, and global gene expression, not related to CCR[202–205]. Despite unrelated functions in the native host, complementation of CRP from *P. putida* and Vfr in *E. coli* was previously achieved, indicating overlap in recognition pattern and machinery[200,203]. Based on unspecific DNA binding affinity and conservation in other bacteria than *E. coli*, CRP has been suggested to occupy two roles, not only as a global TF but also as a nucleoid-associated protein, involved in chromosomal shaping and organization[206]. The role of nucleoid-associated proteins resembles the function of histones in eukaryotes that associates with DNA for condensed packaging in the nucleus[207].

3.2 Catabolite modulating factor – unknown factor of carbon catabolite repression

Even though CCR and its link to CRP have been studied extensively in the last 50 years, not all contributors and molecular triggers of the glucose-induced repression seem to have been discovered. In 1976, Ullmann and co-workers made the fascinating observation that water-soluble extracts from glucose-grown overnight stationary phase *E. coli* were able to repress the synthesis of β-galactosidase, the main enzyme for the consumption of the secondary/non-PTS sugar lactose[208]. This observation pointed towards an unknown factor or molecule(s) tuning the regulation of carbon-involved enzymes, given the relevant name catabolite modulating factor (CMF).

As a result of using column-dependent purification methods, physical characteristics of the CMF were deduced: low weight, uncharged, and heat and acid/alkaline tolerant. A follow-up study demonstrated that catabolite-sensitive enzymes, such as galactokinase and trypthophanase, exhibited repression by CMF whereas glucose-6-phosphate dehydrogensase and phoshoglucomutase did not[209]. The CMF-based repression was transient, presumably due to cellular degradation of the factor. Most importantly, a mutated cAMP-insensitive *lac* promoter showed CMF-mediated repression, indicating independent actions of cAMP and CMF. A connection between CRP and CMF was suggested when cAMP-independent *crp** mutants depicted less sensitivity to CMF compared to the reference strain. Taken these results together, the authors speculated that *"CMF might act via the CAP-protein"* by preventing interaction at promoter regions during transcription[209].

The initial studies of CMF provided proof of its existence and repressive action on CCR-regulated enzymes. CMF was even introduced into a mathematic model as a *"mediator of negative control"* to describe and predict CRR regulation in relation to lactose utilization[210]. Nevertheless, the chemical nature of the specific factor or factors was never identified or isolated.

3.2.1 α-ketoacids as carbon modulator factors

It is well-established in *E. coli* that carbon usage is coordinated though cAMP signaling in the proposed CCR model where the presence of glucose lead to low levels of cAMP and *vice versa*[168]. However, inconsistencies have been observed in several studies where cells that were grown on non-PTS sugars displayed low cAMP levels that should have promoted high levels of cAMP according to the CCR model[170,211–213]. In parallel, low levels of cAMP were detected during nitrogen limiting conditions[214,215]. Not only does the discordant data imply a cAMP-independent regulatory mechanism, partly explained by inducer exclusion[211], it also points towards a tight connection between carbon and nitrogen utilization channeled by cAMP. But what is the connection?

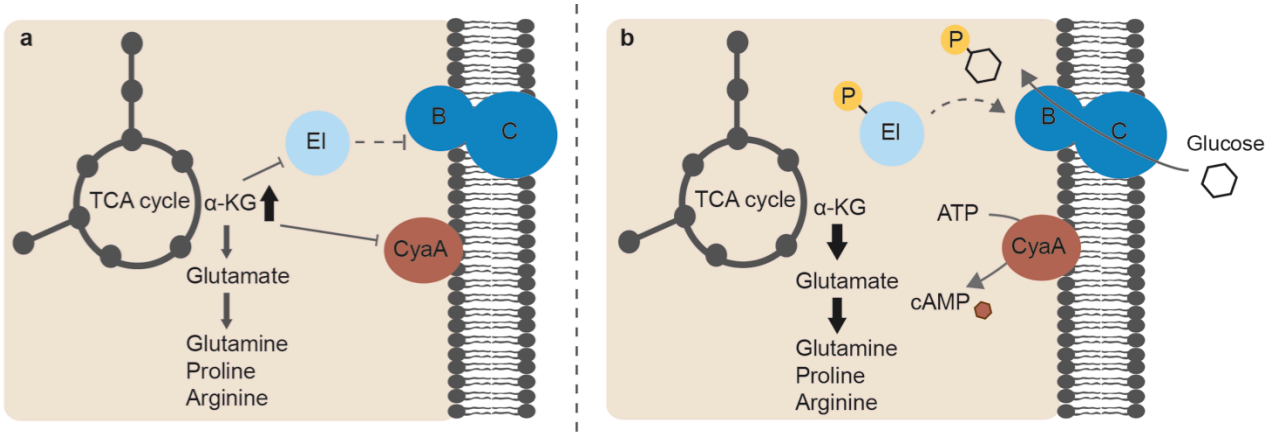


Figure 9 – α–ketoglutarate coordinates carbon and nitrogen metabolism. a. During nitrogen limiting conditions, α-KG accumulates from the TCA cycle (shown as a thick black arrow). α-KG directly inhibits activity of enzyme I (EI) and CyaA. Thus, $EIIBC^{Glc}$ does not take up glucose and cAMP is not produced. **b**. When nitrogen is available, α-KG is converted into the amino acids glutamate, glutamine, proline, and arginine. EI transfers phosphate groups that enable glucose uptake and CyaA can produce cAMP. Dashed lines represent several steps that are not illustrated.

A central player in nitrogen assimilation is the tricarboxylic acid (TCA) cycle intermediate α–ketoglutarate (α-KG), an α–ketoacid, that provides the carbon backbone of certain amino acids. During nitrogen limitation, less flux is directed towards amino acids biosynthesis and consequently α-KG accumulates[216] (Figure 9a). In 1986, Daniel and Danchin found α-KG to inhibit cAMP synthesis and cause repression of β-galactosidase expression. This effect disappeared however in a strain lacking PTS enzyme I[217]. More recently, Doucette *et al.* proved that α-KG abolished the activity of PTS enzyme I by directly binding to it, thereby blocking glucose uptake[218]. To further investigate the coordinated regulation of carbon and nitrogen metabolism, You and colleagues demonstrated that α-KG and other α–ketoacids inhibit CyaA activity and prevent cAMP accumulation in strains with and without PTS[215]. In this way, variations in α-KG levels appear to coordinate carbon and nitrogen metabolism either by directly repressing the activity of CyaA and/or synergistically inhibiting glucose uptake via enzyme I (Figure 9). You and co-workers hinted that α-KG could be the missing piece of the "CMF-puzzle," originally researched almost 40 years earlier. The CMF will be discussed further in section 3.5.3.

3.3 Carbon starvation in ageing bacterial cells

Most free-living bacteria are believed to have a *feast and famine* life cycle, primed for frequently changing environmental settings and nutrient availability[219]. During famine-conditions or starvation, growth is ceased and gene expression programs are adjusted by local and global regulatory systems to ensure prolonged survival in a non-growing state. With limited available resources, there is a trade-off between maintenance and reproduction[220].

3.3.1 The last "growth" phase – becoming old

When grown in laboratory batch cultures i.e. nutrients are only added from the beginning of growth, the bacterial life cycle can be divided into five phases: *lag* (maturing, delay before growth), *exponential (*constant growth), *stationary* (nutrient depletion, growth stopped), *death* (loss of viability), and *long-term stationary* (consistent state for years) phase[221,222] (Figure 10). In this thesis, the focus will be on stationary phase and long-term stationary phase, conditions existing for growth-arrested dormant cells.

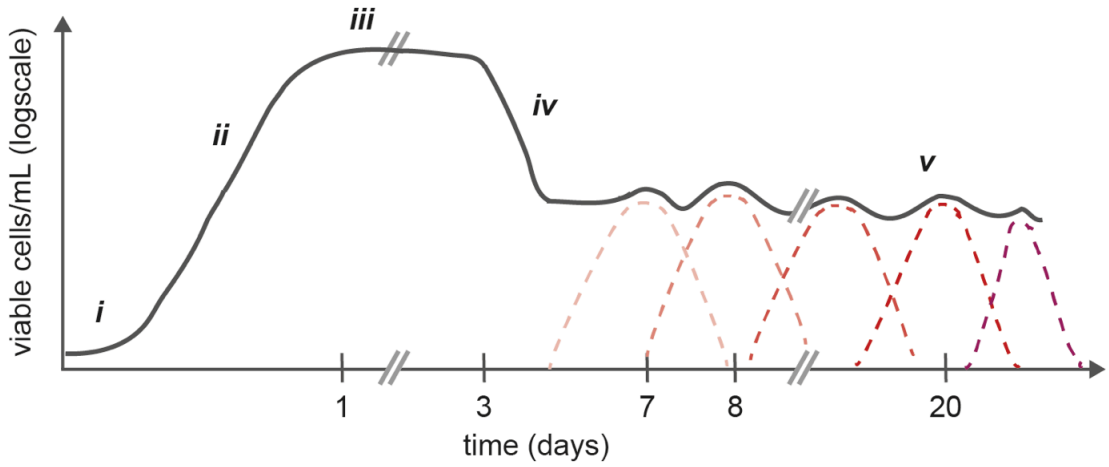


Figure 10 – Growth phases of *E. coli*. *E. coli* cells growing in optimal rich-media conditions at 37°C with aeration. ***i***. lag phase, ***ii.*** exponential phase, ***iii.*** stationary phase, ***iv.*** death phase, and ***v.*** long-term stationary phase. The dashed red colored lines represent cycles of continuous growth and death for different mutants in deep long-term stationary phase. Adapted from Navarro Llorens *et al.*[222]

The main characteristics of bacterial life in stationary phase are growth arrest, reduced metabolic activities, and increased resistance to environmental stresses[223] (Table 1). Nutrient depletion promotes entrance into stationary phase by processes mainly governed by RpoS or

σ38, the stationary-phase and stress sigma factor, inducing *the general stress response*[224]. In stationary phase and during carbon starvation, levels of RpoS increase significantly[225,226]. Low levels of carbon stimulate RpoS synthesis and reduce degradation[222,227]. Expression of *rpoS* is negatively regulated by CRP-cAMP, coordinating the induction of stress signaling with sugar availability[228]. The RpoS sigma factor directly and indirectly controls transcription of up to 10% of the *E. coli* genome[229]. The RpoS regulon includes genes involved in survival and resistance to various stresses e.g. oxidative stress, DNA damage, starvation, heat shock, osmotic stress, and UV- and pH changes[230,231].

Table 1 – Overview of general changes in stationary phase.

	Protein(s)/regulator(s) involved	Change(s)	Reference(s)
Cell shape/morphology	Penicillin binding proteins and β-lactamases, RpoS, BolA	- Smaller cells - More spherical cells due to increased surface/volume ratio - Degradation of cytoplasmic membrane and cell wall (dwarfing) - Reduced fluidity and more resistant envelope	[232–235]
Nucleoid	Dps (DNA-binding protein from starved cells), nucleoid-associated proteins, RpoS	- Condensed nucleoid	[236–238]
Metabolic	RelA, SpoT,(p)ppGpp	- Stringent response - Repression of aerobic metabolism	[239,240]
Transcriptional	(p)ppGpp, RpoS	- Change in sigma factor affinity	[241]
Translational	RMF (Ribosomal modulation factor), (p)ppGpp	- Formation of inactive 100S ribosome dimers - Decrease in protein synthesis - Increase in proteases and peptidases synthesis	[242–244]
Chemical changes e.g. higher pH and oxidative stress	RpoS	- Increased tolerance to stresses	[231,245–247]

With permanent growth arrest for more than three days, 90-99% of an original population of the laboratory model strain *E. coli* K-12 grown in rich medium "dies" and loses ability to form colonies on nutrient agar plates[248]. Consequently, nutrients are released and scavenged by survivors, conferring a *growth advantage in stationery phase* (GASP) phenotype (Figure 10). In this so-called long-term stationary phase, growth and death are balanced out and can be sustained for up to five years[221,249].

3.3.2 Cellular signals and transcriptional responses to nutrient starvation

Elevated cAMP levels have been observed during glucose-scarcity, coupling the CCR response with carbon starvation via cAMP[170,226]. CRP-cAMP induces e.g. transcription of TCA cycle genes and catabolic genes for exploration of alternative carbon sources than glucose[174,250]. When amino acids are deprived, *the stringent response* is initiated, controlled by the hormone-like secondary messenger guanosine tetraphosphate (ppGpp) or guanosine pentaphosphate (pppGpp), here collectively referred to as (p)ppGpp – also known as the magical spot[251–253]. Ribosomal bound-uncharged transfer RNAs (tRNAs) initiate RelA-dependent accumulation of (p)ppGpp that inhibits transcription of ribosomal RNA (rRNA) (part of the translational machinery) and tRNA, thereby tuning down overall protein biosynthesis and growth rate[252,254,255]. (p)ppGpp also interacts directly with RNAP together with a small protein, DksA[256]. Consequently, the regulatory effect of the stringent response via (p)ppGpp coordinates precursor demand of amino acids with precursor supply of available carbon sources[250]. A direct link between (p)ppGpp and utilization of carbon sources through the CRP-cAMP module has been observed in which expression of *crp* was negatively influenced by (p)ppGpp[257]. Additionally, SpoT (bifunctional (p)ppGpp synthase/hydrolase) generates high levels of (p)ppGpp in response to carbon starvation[258,259].

An alteration in relative competitiveness of sigma factors for RNAP – *transcriptional switching* - is believed to be part of the molecular explanation of how environmental cues are translated into signals that control allocation of resources upon growth arrest and nutrient starvation[162]. At the transcriptional level, RNAP is limiting for transcription and sigma factors compete for binding to RNAP[141,260]. (p)ppGpp has been shown to enhance competitiveness of RNAP for alternative sigma factors[241]. In nutrient-rich conditions during growth,

transcription occurs primary of σ70-dependent housekeeping genes, as long as the (p)ppGpp levels are low. In contrast, if levels of (p)ppGpp are elevated as a result of the stringent response, σ38/RpoS-controlled genes are transcribed due to changed sigma factor competition and increased σ38 levels in accordance with cellular demands during physiological stress and stationary phase[162,220,225,261] (Figure 11). Furthermore, the (p)ppGpp/DksA module positively regulates expression of *rpoS,* thereby increasing stress tolerance upon nutrient deprivation[262,263].

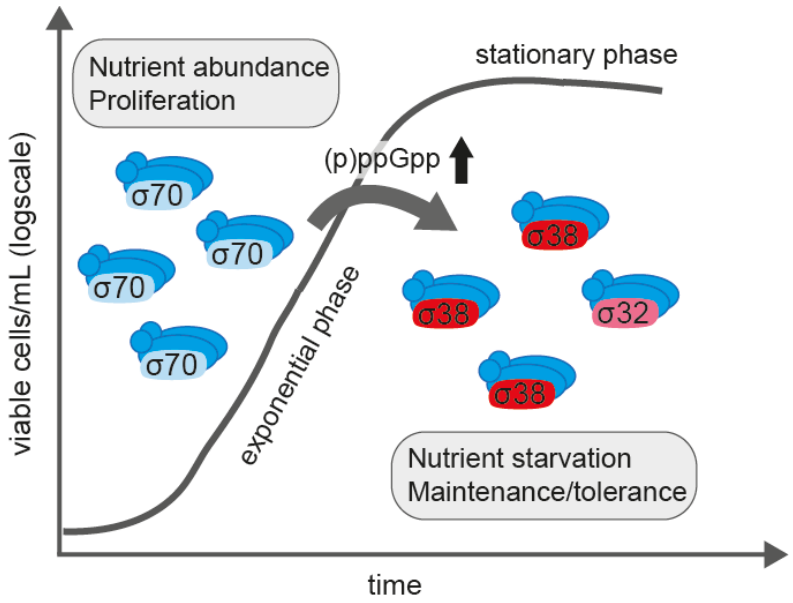


Figure 11 – Transcriptional switching upon nutrient starvation. In nutrient-rich conditions, σ70-dependent housekeeping genes are transcribed that allow for growth. Upon entry to stationary phase due to nutrient depletion, an increase in (p)ppGpp levels (shown as a thick black arrow) changes competitiveness of sigma factors, allowing stationary phase σ38/RpoS-controlled genes or other alternative sigma factors such as σ32/RpoH to be transcribed that are involved in tolerance and maintenance. Inspired from Sharma and Chatterji[162].

3.4 Genetic adaptation to starvation

Bacteria have been recognized as *masters of adaptation* because of their ability to cope with extreme changes in nature such as nutrient depletion. Adaptation can be addressed by altering local and global transcriptional programs and by adaptive evolution[264]. Adaptive evolution refers to processes where genetic mutations appear in bacteria adapting to a new environment under selective conditions[265]. Adaptive mutants can develop from a non-proliferating cell population by relieving selection pressure[266–268]. Selection can either be lethal e.g. rifampicin (Rif) resistance[269,270] or non-lethal e.g. carbon source utilization[18,271].

3.4.1 Retromutagenesis

The specific mechanism and theory behind the appearance of adaptive mutants in non-growing cells is a long-lasting debate among evolutionary biologists[267,272,273]. As phenomena, adaptive evolution implies properties of both Lamarckian ideas of evolution with factors e.g. stress directing specific beneficial mutations and the Darwinian model with environmental conditions driving the natural selection of beneficial traits among randomly generated mutations[274]. Recently published work by our laboratory and others have showed evidence of transcription-associated mutagenesis and the model of retromutagenesis as one evolutionary mechanism in ageing growth-arrested cells, as already suggested by Bridges in 1994[266,275,276]. In this model, the opening of the DNA double helix during transcription likely enhances susceptibility to mutations in both single strands. However, only mutations on the transcribed strand are transferred to mRNA and give rise to mutant proteins that provide beneficial phenotypes. Subsequently, the activity of mutant proteins enables replication and the initial adaptive mutations are fixed on both DNA strands (Figure 12). With this model, mutations are favored on the transcribed strand and the specific nucleotide changes can be traced back when the mutational mechanisms are known[266,276]. Oxidation of guanine to 8-oxo-guanine (8-oxo-G) and cytosine deamination are two dominating mutational events in dormant cells during stationary phase, causing G:C to T:A transversions and G:C to A:T transitions, respectively[277–279].

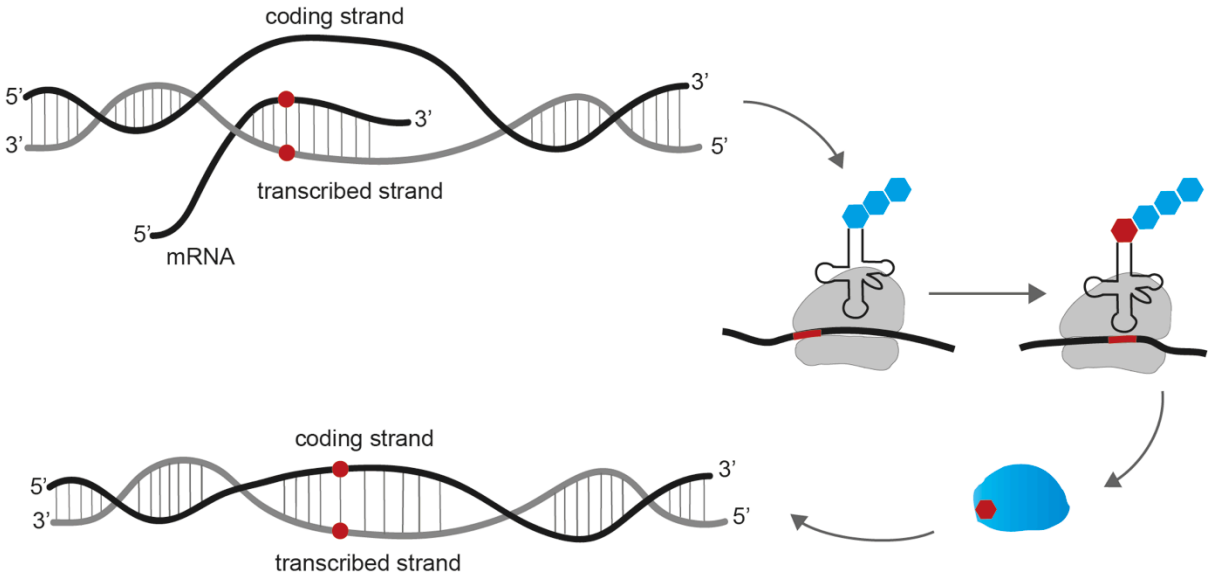


Figure 12 – The model of retromutagenesis. A mutation (here depicted in red) on the transcribed strand is transferred to mRNA and translated into a protein (blue circle). If the mutation allows for growth and replication, it becomes fixed on the coding strand in the next generation.

3.4.2 Experimental evolution for studying carbon-starved cells

The mutational solution space possible in a given evolution experiment is determined by the genetic background of the organism, conditions of the experimental set-up, and the selection pressure applied. Thus, different experimental evolution approaches have been pursued for the study of nutrient-starved bacterial cells. In the study by Sekowska *et al.*, the parental strain Δ*cyaA*, deficient in cAMP synthesis, was plated on MacConkey medium supplied with the CRP-cAMP-dependent carbon source maltose. Small amounts of peptones in the medium allowed for initial colony formation after the first day but due to CRP-cAMP-dependency for maltose utilization, growth was arrested. However, after additional four days of incubation, adapted maltose-fermenting secondary colonies termed *papillae* started to appear that continued for over two months[276] (Figure 13a). The experimental set-up allowed for genetic solutions that conferred beneficial phenotypes for dormant cells stuck in a quiescent state in a carbon-limited environment.

Another strategy for experimental evolution involves continuous growth of a population in well-defined conditions for extended time periods with serial passage of an aliquot to new media at regulator intervals. This type of experiment is often referred to as Adaptive Laboratory Evolution (ALE)[280]. Over time, adaptive mutations that confer beneficial

phenotypes e.g. improved growth rate are selected for and accumulate[281]. The most famous ALE work is the long-term evolution experiment by Richard Lenski that started in 1988 and has now reached over 60.000 generations[282,283]. Improvement of substrate utilization has been attempted in numerous ALE studies[281,284], but only few have investigated adaptive mutations in nutrient-deprived environments[285].

Various physical and chemical factors such as temperature, oxygen, pH, nutrient availability, and osmotic stress can vary according to the type of experimental setting applied e.g. unstructured environments like in agitated liquid cultures as in an ALE-setup or more structured environments such as colonies formed on solid agar plates[286–288]. Differences in the type of environment can impact processes in adaptive mutagenesis, exemplified by Taddei and colleagues who were not able to isolate Rif resistant mutants in ageing liquid cultures but were able to do so on solid agar plates[287]. A very clear age-related trend was observed in the study by Sekowska and co-workers: the frequency of 8-oxo-G mutations in dormant cells on solid agar plates increased over the two months of the experiment, presumably as a result of oxidative damage by reactive oxygen species (ROS)[276,289]. Thus, the environmental settings, chemical conditions, and the temporal development of evolution experiments can shape the mutational solution space.

3.4.3 Mutational hotspots

To elucidate the nature of acquired genetic changes in the evolved strains, NGS was performed on 96 papillae genomes in the work by Sekowska *et al.* and revealed *crp* to be most frequently mutated, mainly producing CRP* mutants enabling cAMP-independent maltose utilization[276] (Figure 13b). An additional 500 *crp* loci were sequenced specifically to analyze the CRP solution space. The most dominating mutation was an alanine to threonine substitution in position 144 (CRPA144T), corresponding to 67% of mutants with *crp* mutations. This is the most frequently isolated CRP mutant in previous CRP studies (reviewed in Paper 5).

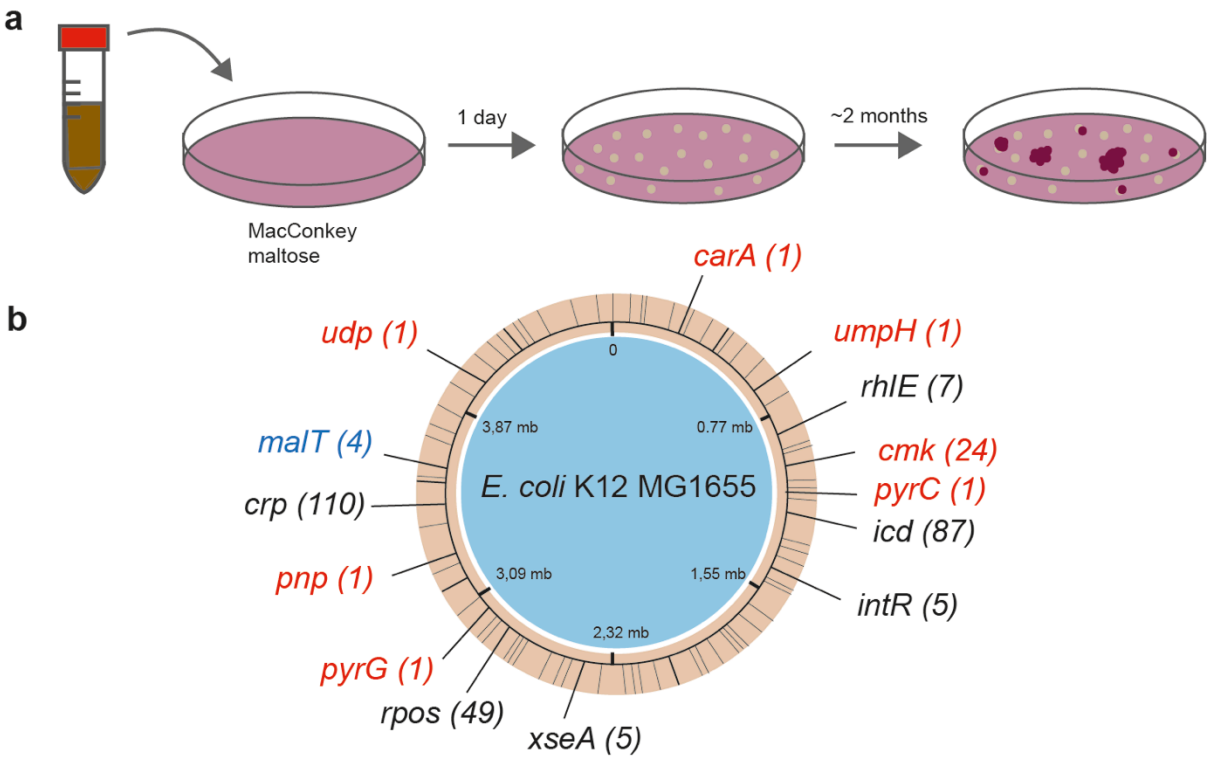


Figure 13 – Experimental evolution of ageing carbon-starved cells and the mutational hotspots. a. The parental strain Δ*cyaA* was plated on MacConkey supplied with maltose that it was not able to metabolize. After one day of incubation at 37°C, small white colonies appeared. For up to two months, papillae mutants (pink colonies) outgrew by using the maltose as carbon source. **b.** NGS of 96 papillae genome revealed mutational hotspots, including genes presumably related to maltose utilization such as *malT* (highlighted in blue) and genes connected to pyrimidine metabolism (highlighted in red). The total number of mutations in each gene is depicted within brackets. Adapted from Sekowska *et al.*[276]

Another frequently observed mutated gene was *cmk* (24 out of 96 genomes), encoding the cytidylate kinase that phosphorylates CMP to CDP[276,290]. The most dominating mutation CmkA216E (17 out of 96) exhibited fermentation on MacConkey-maltose with CRPA144T, indicating a beneficial phenotype by this combination and not wild-type CRP, as depicted in Paper 6. Additionally, it was not possible to isolate the CmkA216E mutant without spontaneous co-occurrence of CRPA144T by genetic editing approaches[276]. Other mutated genes were *carA* (carbamoyl phosphate synthetase α-subunit), *udk* (uridine/cytidine kinase), *pyrG* (CTP synthetase), *pyrC* (dihydroorotase), *nagD/umpH* (UMP phosphatase), and *pnp* (polynucleotide phosphorylase)[136,276], suggesting a relationship between growth-arrested carbon-stressed cells and pyrimidine metabolism (Figure 13b). Mutations were also found in *malT* and its promoter, probable promoting maltose utilization[276] (see Figure 14 for overview

of pyrimidine metabolic pathways and mutated genes). The genetic changes of the sequenced papillae mutants directly or indirectly enabled growth on maltose. Nevertheless, the specific function of the mutations and their effects still remain elusive and has yet to be determined.

Mutations in *rpoS* were also detected in the study by Sekowska and colleagues and these increased in frequency over time[276]. Previously, strains have been observed to inactivate or down-regulate RpoS activity during continuous nutrient limitation such as in stationary phase[291,292]. One explanation of reduced RpoS activity is to unbalance competition between the sigma factors σ70/RpoD and σ54/RpoN that are involved in glucose scavenging and alkaline protection, respectively[245]. RpoS inactivation has also been associated with a *cheater phenotype* that allows for continuous growth by scavenging dying cells upon nutrient deprivation - an example of a GASP phenotype[223]. Thus, it seems that there is a trade-off between stress tolerance conferred by RpoS and growth by acquiring *rpoS* mutations in a nutrient-deprived stationary-phase culture.

3.5 Pyrimidine metabolism

Pyrimidines and purines are two families of nitrogenous bases that together with deoxyribose or ribose sugars and phosphate groups make up the fundamental building blocks of nucleotides in DNA and RNA, respectively. Pyrimidines present in DNA and RNA are cytosine, uracil, and thymine whereas the purines are adenine and guanine[293]. Only pyrimidines will be described in this section in relation to results from Paper 6.

The pyrimidine metabolic pathways can roughly be divided into three parts: i) *de novo* synthesis where the pyrimidine nucleotide uridine monophosphate (UMP) is produced, ii) conversion of UMP into different nucleoside triphosphates (NTPs) and deoxyribonucleotides (dNTPs), and iii) salvage pathways where nucleosides and nucleobases are converted into NMPs[294]. Nucleosides and nucleobases can either be taken up from the external environment or released by intracellular degradation of e.g. RNA[295] (Figure 14).

Transport of uracil and cytosine is mediated by uracil permease (*uraA*) and cytosine permease (*codB*), respectively[296,297] whereas e.g. the nucleoside-specific channel-forming protein (*tsx*) is believed to import nucleosides through the outer membrane[298,299]. Available intracellular uracil can be converted to UMP whereas cytosine first requires deamination to uracil by cytosine deaminase (*codA*)[297]. The nucleoside cytidine can be converted to uridine by cytidine deaminase (*cdd*). High levels of cytidine and CRP-cAMP activate *cdd* expression by relieving CytR-mediated repression[300,301] (see section 3.1.3). The uridine phosphorylase *udp*, that converts uridine into uracil, and the nucleoside-channel *tsx* are both under the transcriptional regulation of CRP-cAMP/CytR/cytidine[302–304] (Figure 14). Thus, CRP and CytR appear to hold regulatory roles in the utilization of pyrimidines as carbon and nitrogen sources coordinated by available cytidine.

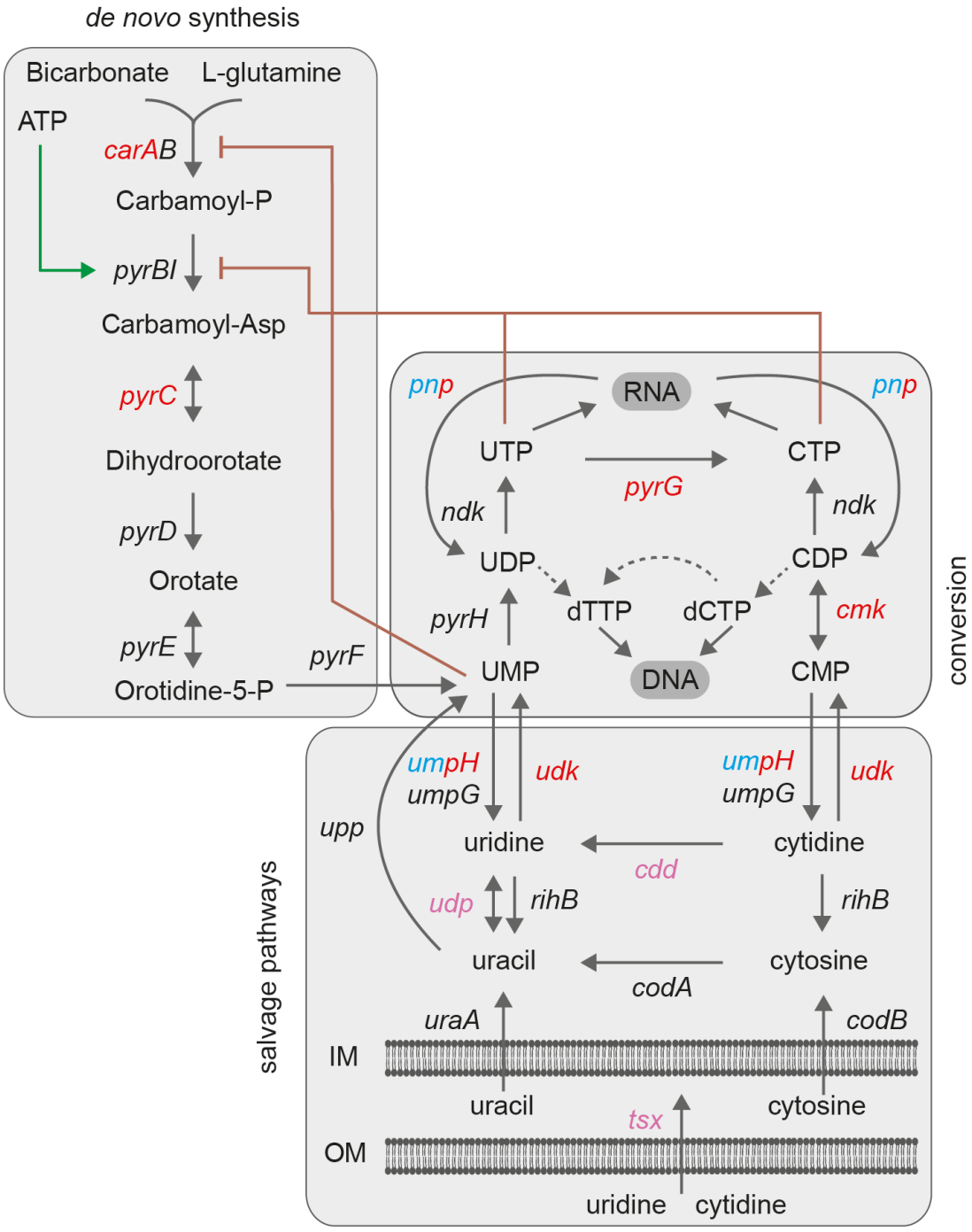


Figure 14 – Overview of the pyrimidine metabolic pathways. *De novo* synthesis involves enzymatic steps that lead to the production of UMP, UMP is converted to UTP, CTP, dTTP, and dCTP for RNA and DNA synthesis, and in the salvage pathways, nucleosides and nucleobases can be released from degradation of e.g. RNA or scavenged from the environment that require uptake across the inner (IM) and outer membrane (OM). Mutational hotspots identified in the work by Sekowska *et al.*[276] are highlighted in red (see Figure 13b). Genes subject to CRP-cAMP-dependent regulation are highlighted in blue and genes under the control of CRP-cAMP/CytR/cytidine are highlighted in pink. *umpH* and *pnp* are both mutational hotspots and regulated by CRP-cAMP. Feedback inhibition by CTP, UTP, and UMP is highlighted in brown whereas activation by ATP is shown in green. Feedback inhibition was deleted in the mutant strain from Reaves *et al.*[305] that showed elevated uracil levels. Dashed lines represent multiple enzymatic steps. ATP, GTP, or co-factor usage is not shown. Adapted

from Reaves *et al.*[305] and Turnbough and Switzer[295] with information from the EcoCyc database (Keseler *et al.*[136]).

3.5.1 Directed pyrimidine overflow

Tight regulation of the nucleotide pathways is employed to sustain constant nucleotide pools to avoid increased DNA errors and mutagenesis[306]. This regulatory control is performed both at the transcriptional and enzymatic level e.g. with feedback inhibition by end products such as UTP and CTP[295]. Another strategy to maintain nucleotide homeostasis is by directed pyrimidine overflow. Here, uracil accumulates and is subsequently eliminated in response to excess nucleotides and low demand of genetic material, a mechanism investigated by Reaves *et al.*[305] The biosynthetic pathway terminal products UTP and most strongly CTP allosterically inhibit the first committed enzymatic step of the pathway and by disrupting this regulation, CTP accumulation was anticipated[305,307,308] (Figure 14). Surprisingly, a strain deficient in regulating excess pyrimidines by end product feedback inhibition, showed accumulation of uracil and not CTP as expected[305]. The uracil build-up was caused by the UMP degradation enzymes UmpH and UmpG to ensure CTP and pyrimidine intermediates at constant concentrations when demand for CTP was low[305] (Figure 14).

Even though the mutant strain depicted an artificial scenario with no feedback inhibition, directed pyrimidine overflow can be considered an alternative strategy for physiological conditions in which genetic demands decreases below the threshold where feedback is effective[309]. Such conditions might be stationary phase-induced growth arrest and starvation in which degradation of RNAs (e.g. mRNA and rRNA) causes release of excess nucleosides or nucleotides[226,310–312].

3.5.2 Pyrimidines during starvation and growth arrest

In the study presented in Paper 6, cytidine and uridine inhibit CRP activity *in vivo*. These results indicate that CRP senses pyrimidine nucleosides and triggers a regulatory effect in stressful conditions where levels of pyrimidines are high such as carbon-starvation and growth arrest. Increased pyrimidine concentrations have been proposed to act as a stress signal in *E. coli* previously[313]. In stationary phase, RNA is degraded and due to low nucleotide

demand in growth-arrested cells, directed pyrimidine overflow causes accumulation of uracil[226,305] and presumably also uridine. Elevated levels of uridine and cytidine levels in non-proliferating cells would consequently inhibit CRP and tune down global transcription of CRP-cAMP-dependent genes. As described in section 3.4.3, *cmk* and six other genes in the pyrimidine metabolic pathways were mutated to accommodate long-term carbon-stress. These results indicate that physiological conditions in carbon-stressed cells promote changes in the pyrimidine metabolism[276].

To elucidate whether nucleotides/nucleosides accumulation occurs as a consequence of starvation and long-term growth arrest, quantification is valuable. Quantification of intracellular nucleotide pools, either by HPLC- or LC/MS-based methods, has been performed in several studies[226,314–317]. In a study by Bennett and co-workers, nucleotide pools were quantified by LC-MS and concentrations of uridine were 1000-fold higher than cytidine in exponential phase[315]. Very recently, Lempp *et al.* detected metabolites in dynamic culture conditions with initial growth in glucose for six hours, carbon starvation for 12 hours, and resumption of growth with glucose for two hours[226]. Throughout the starvation phase, cytidine levels increased up to four-fold compared to levels at the initial growth. Uridine levels were not determined. With resumption of growth, cytidine concentration decreased, presumably caused by CMP and CTP synthesis due to elevated nucleotide demand upon proliferation. Interestingly, throughout the starvation period, CMP, CDP, and CTP levels were low, possibly as a consequence of reduced activity of the energy-dependent enzymes Udk and Cmk due to starvation-induced GTP/ATP-depletion[226] (Figure 14). Overall, cytidine levels were elevated in a starvation phase compared to nutrient-rich conditions in exponential phase, indicating that increased levels of cytidine is common in non-growing starved cells, possibly similar to overflow metabolism.

3.5.3 Pyrimidine nucleosides as catabolite modulating factors

In the original CMF work from the 1970's, the authors suggested CMF to act in concert with or via CRP[209]. The three genes discovered to be subject to CMF-mediated repression, encoding β-galactosidase, galactokinase, and trypthophanase, have all been found to be under transcriptional control of CRP-cAMP, highlighting the link to CRP[136,318]. The same is not the

case for the non-CMF-responsive enzymes: glucose-6-phosphate dehydrogensase and phoshoglucomutase[136] (see section 3.2).

Paper 6 of this thesis proposes the pyrimidine nucleotides cytidine and uridine to be unknown CMF(s) that are sensed by CRP and cause reduction in global CRP-cAMP-dependent transcription during starvation and growth arrest. The presented data fits with several observations from the original CMF-work: CMF is a low-weight molecule and present in extracts of stationary-phase starved overnight cultures since pyrimidine nucleotides accumulate due to carbon scarcity, RNA degradation, and directed overflow metabolism[226,305].

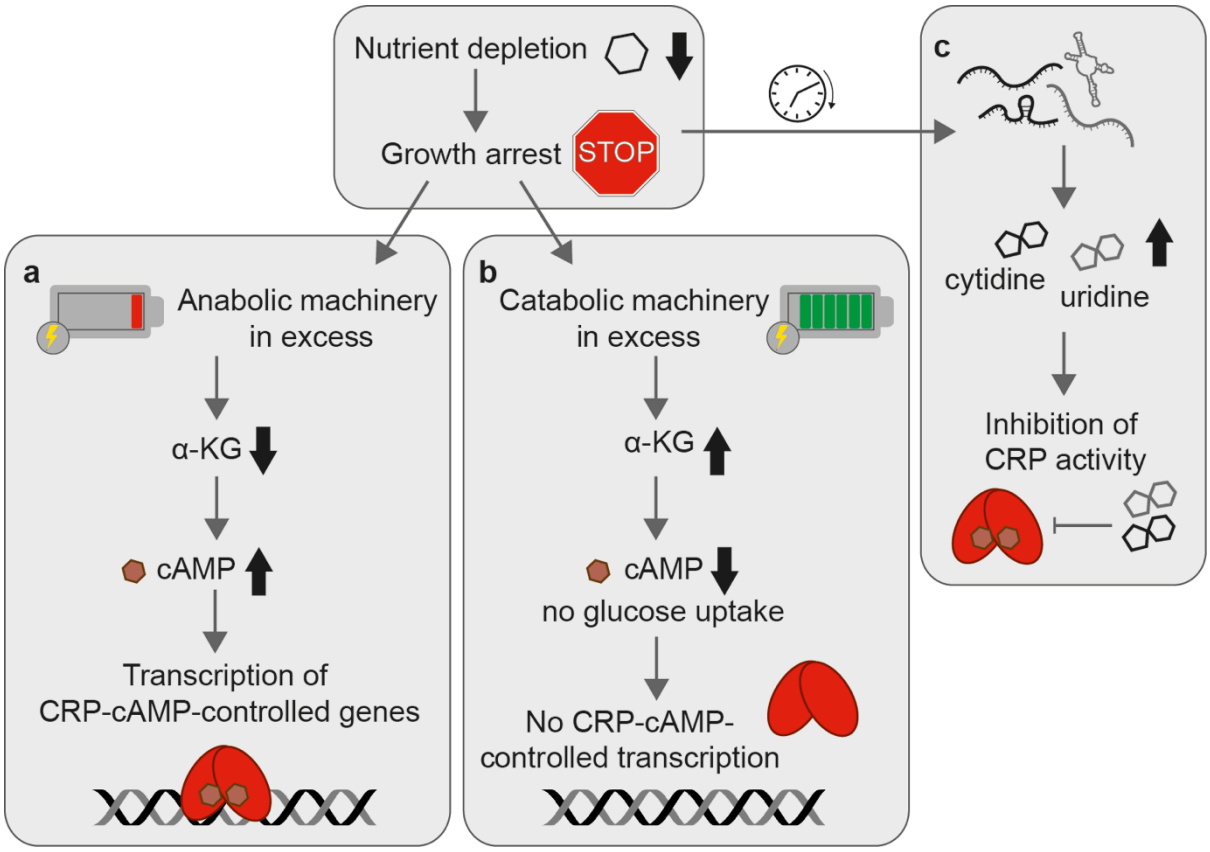


Figure 15 – Schematic model of regulatory mechanisms upon starvation. Nutrient depletion results in growth arrest that promotes entrance into stationary phase. **a.** If the anabolic machinery that uses energy is in excess, levels of α-KG decline, and cAMP is produced. This leads to transcription of CRP-cAMP-dependent genes for e.g. uptake of new nutrients and generation of energy. **b.** If the catabolic machinery that generates energy is in excess, levels of α-KG are elevated and activity of CyaA and EI is inhibited (see section 3.2.1). cAMP is not produced and CRP-cAMP-dependent genes are not transcribed. **c.** Long-term growth arrest and carbon starvation cause RNA degradation and directed overflow metabolism. This increases levels of cytidine and uridine that inhibit CRP activity and global CRP-cAMP-dependent transcription.

α-KG inhibits activity of CyaA and prevents cAMP accumulation and expression of CRP-cAMP-dependent genes, as described in section 3.2.1. If the anabolic machinery that consumes energy is in excess upon growth arrest, α-KG is depleted, cAMP levels rise, and expression of CRP-cAMP-dependent catabolic enzymes that generates energy is induced (Figure 15a). Lempp *et al.* observed this physiological scenario with low levels of α-KG and elevated cAMP levels during carbon starvation[226]. On the other hand, if carbon catabolic machinery is in excess in nitrogen limiting conditions, α-KG levels will increase and cAMP levels will decrease followed by a reduction in global CRP-cAMP expression of catabolic genes[319] (Figure 15b). During long-term carbon deprivation, general down-regulation of energy-dependent cellular responses and RNA degradation cause accumulation of cytidine/uridine, inhibition of CRP activity, and reduction of global CRP-cAMP-dependent expression (Figure 15c). The regulatory mechanism of α-KG via cAMP and the inhibitory effect of pyrimidine nucleotides on CRP might work in concert in which anabolic processes are governed by α-KG and catabolic by pyrimidines through cAMP and CRP.

Concluding remarks and future perspectives

The development of efficient microbial-based cell factories for the production of various compounds of interest and as alternatives to petro-based industries necessitates fundamental understanding of bacterial physiology and adaptive evolution upon extreme conditions such as starvation as well as efficient engineering tools and technologies.

In Paper 1 of this thesis, the CRISPR-Cas9-based tool CRiPi was developed to offer inducible control of two steps in the information flow from DNA to functional proteins: transcriptional initiation with the CRISPRi technology and protein abundance by targeted proteolysis through the N-end-rule pathway. The engineered degradation tag, the N-degron module, was inserted at the genomic locus of a gene of interest with CRMAGE by randomizing the TIR regions in the ssDNA templates to overcome viability impairment. The technology was applied to interfere with a range of essential genes in *E. coli* in which several had never previously been knocked down. The CRiPi technology was also employed for creating antibiotic hypersensitive strains that can be useful for identifying compounds with anti-microbial activity for the development of new antibiotics with the currently emerging antibiotic resistance. Since CRISPRi and the N-end-rule pathway are both cross-kingdom mechanisms, the CRiPi system can be used in various organisms such as yeast in future studies. The study of uncharacterized essential genes with the CRiPi technology is an example of using a synthetic biology tool, developed for applied purposes e.g. by depletion of enzymes to change fluxes in metabolic pathways, for fundamental and basic research.

In Paper 3, the CRISPR-Cas9-based plasmid curing platform pFREE was established for the removal of the most applied plasmid vectors in molecular biology e.g. after genome engineering, bio-sensing, or characterization purposes to obtain plasmid-free strains. Using pFREE, efficient and fast simultaneous curing of single and multiple target plasmids was demonstrated in *E. coli* (40-100% efficiency) and the broad-host version of pFREE, pFREE-RK2, was constructed for plasmid curing in *P. putida* and other bacteria. The pFREE system can even be expanded to eukaryotic model organisms such as *S. cerevisiae*. Ideally, the pFREE system can assist in fighting pathogen bacteria carrying natural and stable plasmids encoding multidrug resistant genes in the future.

The tools developed in Paper 1 and 3 can be applied for the development and optimization of microbial cell factories. Protocols of the two technologies were published as book chapters for usage in other research laboratories (Paper 2 and 4). Additionally, all pFREE constructs, with a variety of resistance markers, were deposited in Addgene for easy public accessibility as part of the open-source synthetic biology mindset.

Nutrient fluctuations in the environment are part of the bacterial lifestyle and regulatory systems cope with these changes by e.g. increase in (p)ppGpp levels and alteration of sigma factor competition for RNAP. However, to study extreme physiological conditions upon long-term carbon deprivation in growth-arrested *E. coli* Δ*cyaA* cells, an experimental evolution approach was applied in a previous study. This work highlighted CRP as a key player in carbon-stressed cells, a connection to pyrimidine metabolism, and retromutagenesis as one possible evolutionary mechanism in ageing bacteria. Knowledge of specific evolutionary mechanisms and the type of nucleotide mutations that occurs during sugar starvation and stationary-phase-conditions can be applied for improving cell factories robustness by redesigning genes to be less prone to mutagenesis. Other applications of experimental evolution in dormant bacteria are the development of more tolerant organisms to e.g. high concentrations of antibiotics, toxic proteins (unpublished observations from our laboratory), or higher temperatures.

In Paper 6 of this thesis *in vivo* CRP activity was inhibited by the pyrimidine nucleosides cytidine and uridine. The mutant CRPA144T, that depicted cAMP-independent utilization of maltose, showed enhanced *in vitro* DNA binding and *in vivo* activity in the presence of cytidine as well as uridine. The inhibitory effect of cytidine on CRP was also demonstrated for CRP-homologues from two related gram-negative bacteria, *P. putida* and *P. aeruginosa,* suggesting a conserved role of pyrimidine nucleosides as stress signals to tune down CRP-cAMP-dependent transcriptional regulation. With the global regulatory repertoire of CRP, the accumulation of pyrimidine nucleosides can be interpreted as a way to regulate carbon and nitrogen metabolism coordinated by metabolic and cellular conditions. Based on the inhibitory effect of cytidine and uridine, these molecules are suggested to be the unknown catabolite modulating factors that were investigated in the 1970's but never identified.

Our discovery of CRP as a sensor of cytidine and uridine adds an additional layer of global transcriptional regulation in physiological situations where nutrients are deprived. Currently, investigations are ongoing in our laboratory to elucidate the binding site of cytidine and uridine in the structure of CRP, to study the connection between RpoS and CRP in ageing bacteria as being two major mutagenic hotspots, and to investigate potential interactions with other regulatory proteins such as the co-regulator Sxy. Even though the role of CRP has been studied extensively for more than 50 years as being a pivotal player in global transcriptional regulation in *E. coli*, we envision this study to open up new doors for investigations.

The understanding of biological systems fuels the urge to engineer and generate stable cell factories and investigations of these can teach us new aspects of fundamental biology. This thesis connects the engineering and the knowledge part of biology with the development of tools and the deciphering of a new role for a central component of transcriptional regulation in *E. coli*. Hopefully, genetic technologies and fundamental biological information can be the foundation for the development of robust and efficient microbial cell factories, leading to advancement in treatment of genetic diseases, production of affordable medicine, and sustainable industries for helping humankind and saving the planet for many future generations to come.

References

1. Microbiology by numbers. *Nat. Rev. Microbiol.* **9**, 628 (2011).
2. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6578–6583 (1998).
3. Rothschild, L. J. & Mancinelli, R. L. Life in extreme environments. *Nature* **409**, 1092–1101 (2007).
4. Elowitz, M. & Lim, W. A. Build life to understand it. *Nature* **468**, 889–890 (2010).
5. Wu, G. *et al.* Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* **34**, 652–664 (2016).
6. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids. *Nature* **171**, 737–738 (1953).
7. Crick, F. H. C., Barnett, L., Brenner, S. & Watts-Tobin, R. J. General nature of the genetic code for proteins. *Nature* **192**, 1227–1232 (1961).
8. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).
9. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
10. Bhatia, S. *History, scope and development of biotechnology*. *Introduction to Pharmaceutical Biotechnology, Volume 1* (2018). doi:10.1088/978-0-7503-1299-8ch1
11. Idalia, V.-M. N. & Bernado, F. Escherichia coli as a Model Organism and Its Application in Biotechnology. *Intech* 253–274 (2017). doi:http://dx.doi.org/10.5772/57353
12. Blount, Z. D. The unexhausted potential of E. coli. *Elife* **4**, 1–12 (2015).
13. Sezonov, G., Joseleau-Petit, D. & D'Ari, R. Escherichia coli physiology in Luria-Bertani broth. *J. Bacteriol.* **189**, 8746–8749 (2007).
14. Neidhart, F. C., Ingraham, J. L. & Schaechter, M. *Physiology of the bacterial cell — A molecular approach*. *Sinauer Associates Inc* (1990).
15. Monod, J. The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).
16. Lehman, I. R., Bessman, M. J., Simms, E. S. & Kornberg, A. Enzymatic Synthesis of Deoxyribonucleic Acid - I. Preperation of substrate and partial purification of an enzyme from Escherichia coli. *J. Biol. Chem.* **237**, 519–525 (1964).
17. Luria, S. E. & Delbrück, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28**, 491–511 (1943).
18. Cairns, J. & Foster, P. L. Adaptive reversion of a frameshift mutation in Escherichia coli. *Genetics* **128**, 695–701 (1991).
19. Blattner, F. R. *et al.* The Complete Genome Sequence of Escherichia coli K-12. *Science (80-.).* **277**, 1453–1462 (1997).
20. Petzold, C. J., Chan, L. J. G., Nhan, M. & Adams, P. D. Analytics for metabolic engineering. *Front. Bioeng. Biotechnol.* **3**, 1–11 (2015).
21. Cohen, S. N., Chang, A. C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3240–4 (1973).
22. Meselson, M. & Yuan, R. DNA restriction enzyme from E. coli. *Nature* **217**, 1110–1114 (1968).
23. Saiki, R. K. *et al.* Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science (80-.).* **230**, 1350–1354 (1985).
24. Mullis, K. *et al.* Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. 1986. *Cold Spring Harb. Symp. Quant. Biol.* **LI**, 263–273 (1986).

25. Mullis, K. B. & Faloona, F. A. Specifc Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reacion. *Methods Enzymol.* **155**, 1–740 (1987).
26. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467 (1977).
27. Srinivasan, S., Gunasekaran, P. & Rajendhran, J. *Chapter 3 - Fundamentals of Molecular Biology*. *Current Developments in Biotechnology and Bioengineering: Foundations of Biotechnology and Bioengineering* (Elsevier B.V., 2017).
28. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**, 381–390 (2014).
29. Itakura, K. *et al.* Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. *Science (80-.).* **198**, 1056–1063 (1977).
30. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).
31. Aizat, W. M., Ismail, I. & Noor, N. M. *Chapter 1 - Recent Development in Omics Studies*. *Springer Nature Swizerland AG 2018 Omics Applications for Systems Biology* (2018). doi:10.1007/978-3-319-98758-3
32. Monk, J. M. *et al.* Multi-omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes. *Cell Syst.* **3**, 238-251.e12 (2016).
33. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337**, 816–822 (2012).
34. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).
35. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2579–2586 (2012).
36. Wright, A. V., Nunez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29–44 (2016).
37. Jansen, R., Van Embden, J. D. A., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
38. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
39. Jiang, F. & Doudna, J. A. CRISPR – Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, 505–531 (2017).
40. Karlgren, M., Simoff, I., Keiser, M., Oswald, S. & Artursson, P. CRISPR-Cas9: A new addition to the drug metabolism and disposition tool box. *Drug Metab. Dispos.* **46**, 1776–1786 (2018).
41. Pörtner, H.-O. *et al. IPCC, 2019: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. (2019).
42. Thompson, R. C., Moore, C. J., Saal, F. S. V. & Swan, S. H. Plastics, the environment and human health: Current consensus and future trends. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 2153–2166 (2009).
43. French, K. E. Harnessing synthetic biology for sustainable development. *Nat. Sustain.* **2**, 250–252 (2019).
44. Erickson, B., Nelson & Winters, P. Perspective on opportunities in industrial biotechnology in renewable chemicals. *Biotechnol. J.* **7**, 176–185 (2012).
45. Gavrilescu, M. & Chisti, Y. Biotechnology - A sustainable alternative for chemical industry. *Biotechnol. Adv.* **23**, 471–499 (2005).

46. Calero, P. & Nikel, P. I. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* **12**, 98–124 (2019).
47. Chen, X. *et al.* Metabolic engineering of Escherichia coli: A sustainable industrial platform for bio-based chemical production. *Biotechnol. Adv.* **31**, 1200–1223 (2013).
48. Pontrelli, S. *et al.* Escherichia coli as a host for metabolic engineering. *Metab. Eng.* **50**, 16–46 (2018).
49. Nielsen, J. & Keasling, J. D. Synergies between synthetic biology and metabolic engineering. *Nat. Biotechnol.* **29**, 693–695 (2011).
50. Budin, I. & Keasling, J. D. Synthetic Biology for Fundamental Biochemical Discovery. *Biochemistry* **58**, 1464–1469 (2019).
51. Demain, A. L. From natural products discovery to commercialization: A success story. *J. Ind. Microbiol. Biotechnol.* **33**, 486–495 (2006).
52. Serrano, L. Synthetic biology: Promises and challenges. *Mol. Syst. Biol.* **3**, 158 (2007).
53. Chi, H. *et al.* Engineering and modification of microbial chassis for systems and synthetic biology. *Synth. Syst. Biotechnol.* **4**, 25–33 (2019).
54. Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* **34**, e122 (2006).
55. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, (2008).
56. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–5 (2009).
57. Hughes, R. A. & Ellington, A. D. Synthetic DNA synthesis and assembly: Putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* **9**, (2017).
58. Schwille, P. Jump-starting life? Fundamental aspects of synthetic biology. *J. Cell Biol.* **210**, 687–690 (2015).
59. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science (80-.).* **329**, 52–56 (2010).
60. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science (80-.).* **351**, 6253–6253 (2016).
61. Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. *Nature* (2019). doi:10.1038/s41586-019-1192-5
62. Endy, D. Foundations for engineering biology. *Nature* **438**, 449–453 (2005).
63. Vilanova, C. & Porcar, M. IGEM 2.0 - Refoundations for engineering biology. *Nat. Biotechnol.* **32**, 420–424 (2014).
64. 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**, 666–675 (2013).
65. Kim, S. H., Cavaleiro, A. M., Rennig, M. & Nørholm, M. H. H. SEVA Linkers: A Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology. *ACS Synth. Biol.* **5**, 1177–1181 (2016).
66. Martínez-Garćía, E., Aparicio, T., Goñi-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–D1189 (2015).
67. Padilla-Vaca, F., Anaya-Velázquez, F. & Franco, B. Synthetic biology: Novel approaches for microbiology. *Int. Microbiol.* **18**, 71–84 (2015).
68. Kwok, R. Five hard truths for synthetic biology. *Nature* **463**, 288–290 (2010).

69. Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Mol. Biosyst.* **10**, 1668–1678 (2014).
70. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. Designing and engineering evolutionary robust genetic circuits. *J. Biol. Eng.* **4**, 12 (2010).
71. Keasling, J. D. Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* **14**, 189–195 (2012).
72. Dai, Z. *et al.* Genetic tool development and systemic regulation in biosynthetic technology. *Biotechnol. Biofuels* **11**, 1–12 (2018).
73. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 1–11 (2006).
74. Kent, R. & Dixon, N. Contemporary Tools for Regulating Gene Expression in Bacteria. *Trends Biotechnol.* **202**, 1–18 (2019).
75. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–5 (2000).
76. Blazeck, J. & Alper, H. S. Promoter engineering: Recent advances in controlling transcription at the most fundamental level. *Biotechnol. J.* **8**, 46–58 (2013).
77. Chen, Y. *et al.* Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. *Nat. Commun.* **9**, 1–8 (2018).
78. Bervoets, I. & Charlier, D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: Opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiol. Rev.* **43**, 304–339 (2019).
79. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
80. Laursen, B. S., Sørensen, H. P., Mortensen, K. K. & Sperling-Petersen, H. U. Initiation of Protein Synthesis in Bacteria. *Microbiol. Mol. Biol. Rev.* **69**, 101–123 (2005).
81. Rennig, M. *et al.* TARSyn: Tunable Antibiotic Resistance Devices Enabling Bacterial Synthetic Evolution and Protein Production. *ACS Synth. Biol.* **7**, 432–442 (2018).
82. Bonde, M. T. *et al.* Predictable tuning of protein expression in bacteria. *Nat. Methods* **13**, 233–236 (2016).
83. Gao, C. *et al.* Programmable biomolecular switches for rewiring flux in Escherichia coli. *Nat. Commun.* **10**, 1–12 (2019).
84. Salis, H. M. The Ribosome Binding Site Calculator. *Methods Enzymol.* **498**, 19–42 (2011).
85. Shine, J. & Dalgarno, L. The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346 (1974).
86. Mirzadeh, K. *et al.* Enhanced Protein Production in Escherichia coli by Optimization of Cloning Scars at the Vector − Coding Sequence Junction. *ACS Synth. Biol.* **4**, 959–965 (2015).
87. Liu, Z., Liang, Y., Ang, E. L. & Zhao, H. A New Era of Genome Integration - Simply Cut and Paste! *ACS Synth. Biol.* **6**, 601–609 (2017).
88. Jeong, J., Cho, N., Jung, D. & Bang, D. Genome-scale genetic engineering in Escherichia coli. *Biotechnol. Adv.* **31**, 804–10 (2013).
89. Murphy, K. C. Use of Bacteriophage λ Recombination Functions To Promote Gene Replacement in Escherichia coli. *J. Bacteriol.* **180**, 2063–2071 (1998).
90. Little, J. W. An Exonucleause Induced by Bacteriophage λ. *J. Biol. Chem.* **242**, 679–686 (1967).

91. Mosberg, J. A., Lajoie, M. J. & Church, G. M. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics* **186**, 791–799 (2010).
92. Li, Z., Karakousis, G., Chiu, S. K., Reddy, G. & Radding, C. M. The beta protein of phage λ promotes strand exchange. *J. Mol. Biol.* **276**, 733–744 (1998).
93. Karakousis, G. *et al.* The beta protein of phage λ binds preferentially to an intermediate in DNA renaturation. *J. Mol. Biol.* **276**, 721–731 (1998).
94. Marsic, N., Roje, S., Stojiljkovic, I., Salaj-Smic, E. & Trgovcevic, Z. In vivo studies on the interaction of RecBCD enzyme and λ gam protein. *J. Bacteriol.* **175**, 4738–4743 (1993).
95. Datta, S., Costantino, N. & Court, D. L. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109–115 (2006).
96. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–8 (2009).
97. Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6742–6 (2001).
98. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4**, 206–223 (2009).
99. Rennig, M. Synthetic biology approaches for protein production optimization in bacterial cell factories. *Kgs. Lyngby Novo Nord. Found. Cent. Biosustainability* (2017).
100. Swaminathan, S. *et al.* Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* **29**, 14–21 (2001).
101. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. a. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–9 (2013).
102. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in Escherichia coli. *Appl. Environ. Microbiol.* **81**, 5103–5114 (2015).
103. Schweizer, H. P. Applications of the Saccharomyces cerevisiae Flp-FRT system in bacterial genetics. *J. Mol. Microbiol. Biotechnol.* **5**, 67–77 (2003).
104. Bassalo, M. C. *et al.* Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synth. Biol.* **5**, 561–568 (2016).
105. Li, Y. *et al.* Metabolic engineering of Escherichia coli using CRISPR–Cas9 meditated genome editing. *Metab. Eng.* **31**, 13–21 (2015).
106. Ronda, C., Pedersen, L. E., Sommer, M. O. A. & Nielsen, A. T. CRMAGE: CRISPR Optimized MAGE Recombineering. *Sci. Rep.* **6**, 19452 (2016).
107. Isaacs, F. J. *et al.* Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science (80-.).* **333**, 348–353 (2011).
108. Jiang, Y. *et al.* Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).
109. Brockman, I. M. & Prather, K. L. J. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnol. J.* **10**, 1360–1369 (2015).
110. Sekar, K., Gentile, A. M., Bostick, J. W. & Tyo, K. E. J. N-Terminal-Based Targeted , Inducible Protein Degradation in Escherichia coli. *PLoS One* **11**, 1–17 (2016).
111. Mosteller, R. D., Goldstein, R. V. & Nishimoto, K. R. Metabolism of individual proteins in exponentially growing Escherichia coli. *J. Biol. Chem.* **255**, 2524–2532 (1980).

112. Davis, J. H., Baker, T. & Sauer, R. Small-molecule control of protein degradation using split adaptors. *ACS Chem. Biol.* 1205–1213 (2011). doi:10.1021/cb2001389
113. Taxis, C. *et al.* Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. *Mol. Syst. Biol.* **5**, 1–7 (2009).
114. Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nat. Biotechnol.* **32**, 1276–1283 (2014).
115. Dougan, D. A., Truscott, K. N. & Zeth, K. The bacterial N-end rule pathway: Expect the unexpected. *Mol. Microbiol.* **76**, 545–558 (2010).
116. Mogk, A., Schmidt, R. & Bukau, B. The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol.* **17**, 165–172 (2007).
117. Schmidt, R., Zahn, R., Bukau, B. & Mogk, A. ClpS is the recognition component for Escherichia coli substrates of the N-end rule degradation pathway. *Mol. Microbiol.* **72**, 506–517 (2009).
118. Rivera-Rivera, I., Román-Hernández, G., Sauer, R. T. & Baker, T. A. Remodeling of a delivery complex allows ClpS-mediated degradation of N-degron substrates. *PNAS* **111**, E3853-9 (2014).
119. Erbse, A. *et al.* ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **439**, 753–756 (2006).
120. Durante-Rodríguez, G., De Lorenzo, V. & Nikel, P. I. A Post-translational Metabolic Switch Enables Complete Decoupling of Bacterial Growth from Biopolymer Production in Engineered Escherichia coli. *ACS Synth. Biol.* **7**, 2686–2697 (2018).
121. Cohen, S. N. DNA cloning: a personal view after 40 years. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15521–9 (2013).
122. Nora, L. C. *et al.* The art of vector engineering: towards the construction of next-generation genetic tools. *Microb. Biotechnol.* **12**, 125–147 (2019).
123. Smalla, K., Jechalke, S. & Top, E. M. Plasmid Detection, Characterization, and Ecology. *Microbiol. Spectr.* **3**, 1–21 (2015).
124. Friehs, K. Plasmid copy number and plasmid stability. *Adv. Biochem. Eng. Biotechnol.* **86**, 47–82 (2004).
125. Corchero, J. L. & Villaverde, A. Plasmid maintenance in Escherichia coli recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol. Bioeng.* **58**, 625–632 (1998).
126. Silva, F., Queiroz, J. A. & Domingues, F. C. Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. *Biotechnol. Adv.* **30**, 691–708 (2012).
127. Trevors, J. T. Plasmid curing in bacteria. *FEMS Microbiol. Lett.* **32**, 149–157 (1986).
128. Wein, T., Hülter, N. F., Mizrahi, I. & Dagan, T. Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nat. Commun.* **10**, 1–13 (2019).
129. Summers, D. K. & Sherratt, D. J. Multimerization of High Copy Number Plasmids Causes Instability: ColE1 Encodes a Determinant Essential for Plasmid Monomerization and Stability. *Cell* **36**, 1097–1103 (1984).
130. Holm, A. K. *et al.* Metabolic and Transcriptional Response to Cofactor Perturbations in Escherichia coli. *J. Biol. Chem.* **285**, 17498–17506 (2010).
131. Varma, A. & Palsson, B. O. Metabolic Capabilities of Escherichia coli: I. Synthesis of Biosynthetic Precursors and Cofactors. *Journal of Theoretical Biology* **165**, 477–502 (1993).
132. Chubukov, V., Gerosa, L., Kochanowski, K. & Sauer, U. Coordination of microbial

metabolism. *Nat. Rev. Microbiol.* **12**, 327–340 (2014).
133. Beisel, C. L. & Afroz, T. Rethinking the Hierarchy of Sugar Utilization in Bacteria. *J. Bacteriol.* **198**, 374–376 (2016).
134. Monod, J. *Recherches sur la Croissance des Cultures Bactériennes. Paris: Herman & Cie* (1942).
135. Murakami, K. S., Masuda, S. & Darst, S. A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science (80-.).* **296**, 1280–1284 (2002).
136. Keseler, I. M. *et al.* EcoCyc: A comprehensive database of Escherichia coli biology. *Nucleic Acids Res.* **39**, 583–590 (2011).
137. Zhang, G. & Darst, Seth, A. Structure of the Escherichia coli RNA polymerase α Subunit Amino-Terminal Domain. *Science (80-.).* **281**, 262–265 (1998).
138. Jeon, Y. H. *et al.* Solution Structure of the Activator Domain of the RNA Polymerase α Subunit. *Science (80-.).* **270**, 1495–1497 (1995).
139. Zhang, G. *et al.* Crystal structure of thermus aquaticus core RNA polymerase at 3.3 å resolution. *Cell* **98**, 811–824 (1999).
140. Burgess, R. R., Travers, A. A., Dunn, J. J. & Bautz, E. K. F. Factor stimulating transcription by RNA polymerase. *Nature* **221**, 43–46 (1969).
141. Ishihama, A. Functional Modulation of Escherichia coli RNA polymerase. *Annu. Rev. Microbiol.* **54**, 499–518 (2000).
142. Ross, W., Ernst, A. & Gourse, R. L. Fine structure of E. coli RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev.* **15**, 491–506 (2001).
143. Sanderson, A., Mitchell, J. E., Minchin, S. D. & Busby, S. J. W. Substitutions in the Escherichia coli RNA polymerase σ70 factor that affect recognition of extended -10 elements at promoters. *FEBS Lett.* **544**, 199–205 (2003).
144. Gruber, T. M. & Gross, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annu. Rev. Microbiol.* **57**, 441–466 (2003).
145. Enz, S., Braun, V. & Crosa, J. H. Transcription of the region encoding the ferric dicitrate-transport system in Escherichia coli: similarity between promoters for fecA and for extracytoplasmic function sigma factors. *Gene* **163**, 13–18 (1995).
146. Arnosti, D. N. & Chamberlin, M. J. Secondary σ factor controls transcription of flagellar and chemotaxis genes in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 830–834 (1989).
147. Hawley, D. K. & Mcclure, W. R. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res.* **11**, 2237–2255 (1983).
148. Gilman, J. & Love, J. Synthetic promoter design for new microbial chassis. *Biochem. Soc. Trans.* **44**, 731–737 (2016).
149. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638–650 (2016).
150. Browning, D. F. & Busby, S. J. W. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2**, 57–65 (2004).
151. Ishihama, A. Prokaryotic genome regulation: Multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol. Rev.* **34**, 628–645 (2010).
152. Perez-Rueda, E. The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Res.* **28**, 1838–1847 (2000).
153. Gao, Y. *et al.* Systematic discovery of uncharacterized transcription factors in Escherichia coli K-12 MG1655. *Nucleic Acids Res.* **46**, 10682–10696 (2018).
154. Babu, M. M. & Teichmann, S. A. Functional determinants of transcription factors in

Escherichia coli: Protein families and binding sites. *Trends Genet.* **19**, 75–79 (2003).
155. Sashasayee, A. S. N., Sivaraman, K. & Luscombe, N. M. *Chapter 2 - An Overview of Prokaryotic Transcription Factors. A Handbook of Transcription Factors* **53**, (2011).
156. Balleza, E. *et al.* Regulation by transcription factors in bacteria: Beyond description. *FEMS Microbiol. Rev.* **33**, 133–151 (2009).
157. Janga, S. C., Salgado, H., Collado-Vides, J. & Martínez-Antonio, A. Internal Versus External Effector and Transcription Factor Gene Pairs Differ in Their Relative Chromosomal Position in Escherichia coli. *J. Mol. Biol.* **368**, 263–272 (2007).
158. Martínez-Antonio, A. & Collado-Vides, J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.* **6**, 482–489 (2003).
159. Dorman, C. J. DNA supercoiling and transcription in bacteria: a two-way street. *BMC Mol. cell Biol.* **20**, 26 (2019).
160. Bryant, J. A., Sellars, L. E., Busby, S. J. W. & Lee, D. J. Chromosome position effects on gene expression in Escherichia coli K-12. *Nucleic Acids Res.* **42**, 11383–11392 (2014).
161. Dame, R. T. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.* **56**, 858–870 (2005).
162. Sharma, U. K. & Chatterji, D. Transcriptional switching in Escherichia coli during stress and starvation by modulation of σ70 activity. *FEMS Microbiol. Rev.* **34**, 646–657 (2010).
163. Zubay, G., Schwartz, D. & Beckwith, J. Mechanism of Activation of Catabolite-Sensitive Genes: A Positive Control System. *Proc. Natl. Acad. Sci.* **66**, 104–110 (1970).
164. Emmer, M., deCrombrugghe, B., Pastan, I. & Perlman, R. Cyclic AMP Receptor Protein of E. coli: Its Role in the Synthesis of Inducible Enzymes. *Proc. Natl. Acad. Sci.* **66**, 480–487 (1970).
165. Magasanik, B. Catabolite Repression. *Cold Spring Harb. Symp. Quant. Biol.* 249–256 (1961). doi:10.1101/SQB.1961.026.01.031
166. Deutscher, J. The mechanisms of carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.* **11**, 87–93 (2008).
167. Makman, R. S. & Sutherland, E. W. Adenosine 3',5'-Phosphate in Escherichia. *J. Biol. Chem.* **240**, 1309–1314 (1965).
168. Görke, B. & Stülke, J. Carbon catabolite repression in bacteria: Many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6**, 613–624 (2008).
169. Hogema, B. M. *et al.* Inducer exclusion in Escherichia coli by non-PTS substrates: The role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIA(Glc). *Mol. Microbiol.* **30**, 487–498 (1998).
170. Epstein, W., Rothman Denes, L. B. & Hesse, J. Adenosine 3':5' cyclic monophosphate as mediator of catabolite repression in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 2300–2304 (1975).
171. Nelson, S. O., Wright, J. K. & Postma, P. W. The mechanism of inducer exclusion. Direct interaction between purified III Glc of the phosphoenolpyruvate:sugar phosphotransferase system and the lactose carrier of Escherichia coli . *EMBO J.* **2**, 715–720 (1983).
172. Gottesman, S. Bacterial regulation: Global regulatory networks. *Ann. Rev. Genet.* **18**, 415–41 (1984).
173. Santos-Zavaleta, A. *et al.* RegulonDB v 10.5: Tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Res.* **47**, D212–D220 (2019).

174. Shimada, T., Fujita, N., Yamamoto, K. & Ishihama, A. Novel Roles of cAMP Receptor Protein (CRP) in Regulation of Transport and Metabolism of Carbon Sources. *PLoS One* **6**, (2011).
175. Aiba, H. Transcription of the Escherichia coli adenylate cyclase gene is negatively regulated by cAMP-cAMP receptor protein. *J. Biol. Chem.* **260**, 3063–3070 (1985).
176. Aiba, H. Autoregulation of the Escherichia coli crp gene: CRP is a transcriptional repressor for its own gene. *Cell* **32**, 141–149 (1983).
177. Hanamura, A. & Aiba, H. A new aspect of transcriptional control of the Escherichia coli crp gene: positive autoregulation. *Mol. Microbiol.* **6**, 2489–2497 (1992).
178. Martínez-Antonio, A., Janga, S. C. & Thieffry, D. Functional organisation of Escherichia coli transcriptional regulatory network. *J. Mol. Biol.* **381**, 238–247 (2008).
179. Kolb, A., Busby, S., Buc, H., Gargas, S. & Adhya, S. Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **62**, 749–795 (1993).
180. Aiba, H., Nakamura, T., Mitani, H. & Mori, H. Mutations that alter the allosteric nature of cAMP receptor protein of Escherichia coli. *EMBO J.* **4**, 3329–3332 (1985).
181. Leuze, M. R., Karpinets, T. V., Syed, M. H., Beliaev, A. S. & Uberbacher, E. C. Binding motifs in bacterial gene promoters modulate transcriptional effects of global regulators CRP and ArcA. *Gene Regul. Syst. Bio.* **2012**, 93–107 (2012).
182. Rhodius, V. A., West, D. M., Webster, C. L., Busby, S. J. W. & Savery, N. J. Transcription activation at class II CRP-dependent promoters: The role of different activating regions. *Nucleic Acids Res.* **25**, 326–332 (1997).
183. Brennan, R. G. Interactions of the helix-turn-helix binding domain. *Curr. Opin. Struct. Biol.* (1991). doi:10.1016/0959-440x(91)90015-l
184. Zhou, Y., Merkel, T. J. & Ebright, R. H. Characterization of the activating region of Escherichia coli catabolite gene activator protein (CAP) II. Role at class I and class II CAP-dependent promoters. *J. Mol. Biol.* (1994). doi:10.1016/0022-2836(94)90035-3
185. Busby, S. & Ebright, R. H. Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* **293**, 199–213 (1999).
186. de Crombrugghe, B. *et al.* Lac DNA, RNA Polymerase and Cyclic AMP Receptor Protein, Cyclic AMP, Lac Repressor and Inducer are the Essential Elements for Controlled Lac Transcription. *Nat. New Biol.* **231**, 139–142 (1971).
187. Chapon, C. & Kolb, A. Action of CAP on the malT promoter in vitro. *J. Bacteriol.* **156**, 1135–1143 (1983).
188. Boos, W. & Shuman, H. Maltose/maltodextrin system of Escherichia coli: transport, metabolism, and regulation. *Microbiol. Mol. Biol. Rev.* **62**, 204–29 (1998).
189. Niu, W., Kim, Y., Tau, G., Heyduk, T. & Ebright, R. H. Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase. *Cell* **87**, 1123–1134 (1996).
190. Lawson, C. L. *et al.* Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol.* **14**, 1–11 (2004).
191. Rhodius, V. A. & Busby, S. J. W. Interactions between Activating Region 3 of the Escherichia coli cyclic AMP receptor protein and region 4 of the RNA polymerase σ70 subunit: Application of suppression genetics. *J. Mol. Biol.* **299**, 311–324 (2000).
192. Aiba, H., Adhya, S. & de Crombrugghe, B. Evidence for two functional gal promoters in intact Escherichia coli cells. *J. Biol. Chem.* **256**, 11905–11910 (1981).
193. Lewis, D. E. A. & Adhya, S. Molecular mechanisms of transcription initiation at gal promoters and their multi-level regulation by GalR, CRP and DNA loop. *Biomolecules* **5**, 2782–2807 (2015).

194. Ramseier, T. M. & Saier, M. H. cAMP-cAMP receptor protein complex: five binding sites in the control region of the Escherichia coli mannitol operon. *Microbiology* **141**, 1901–1907 (1995).
195. Møllegaard, N. E., Rasmussen, P. B., Valentin-Hansen, P. & Nielsen, P. E. Characterization of promoter recognition complexes formed by CRP and CytR for repression and by CRP and RNA polymerase for activation of transcription on the Escherichia coli deoP2 promoter. *J. Biol. Chem.* **268**, 17471–17477 (1993).
196. Valentin-Hansen, P., Søgaard-Andersen, L. & Pedersen, H. A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Mol. Microbiol.* **20**, 461–466 (1996).
197. Pedersen, H., Søgaard-Andersen, L., Holst, B. & Valentin-Hansen, P. Heterologous Cooperativity in Escherichia coli. *J. Biol. Chem.* **266**, 17804–17806 (1991).
198. Gerlach, P., Valentin-Hansen, P. & Bremer, E. Transcriptional regulation of the cytR repressor gene of Escherichia coli: autoregulation and positive control by the cAMP/CAP complex. *Mol. Microbiol.* **4**, 479–488 (1990).
199. Soberón-Chávez, G., Alcaraz, L. D., Morales, E., Ponce-Soto, G. Y. & Servín-González, L. The transcriptional regulators of the CRP family regulate different essential bacterial functions and can be inherited vertically and horizontally. *Front. Microbiol.* **8**, 1–8 (2017).
200. Arce-Rodríguez, A. *et al.* The Crp regulator of Pseudomonas putida: Evidence of an unusually high affinity for its physiological effector, cAMP. *Environ. Microbiol.* **14**, 702–713 (2012).
201. Herrera, M. C., Daddaoua, A., Ana, F. E. & Ramos, J. L. Involvement of the global crp regulator in cyclic amp-dependent utilization of aromatic amino acids by pseudomonas putida. *J. Bacteriol.* **194**, 406–412 (2012).
202. Fuchs, E. L. *et al.* The Pseudomonas aeruginosa Vfr regulator controls global virulence factor expression through cyclic AMP-dependent and -independent mechanisms. *J. Bacteriol.* (2010). doi:10.1128/JB.00363-10
203. West, S. E. H., Sample, A. K. & Runyen-Janecky, L. J. The vfr gene product, required for Pseudomonas aeruginosa exotoxin A and protease production, belongs to the cyclic AMP receptor protein family. *J. Bacteriol.* **176**, 7532–7542 (1994).
204. Suh, S. J. *et al.* Effect of vfr mutation on global gene expression and catabolite repression control of Pseudomonas aeruginosa. *Microbiology* **148**, 1561–1569 (2002).
205. Beatson, S. A., Whitchurch, C. B., Sargent, J. L., Levesque, R. C. & Mattick, J. S. Differential regulation of twitching motility and elastase production by Vfr in Pseudomonas aeruginosa. *J. Bacteriol.* **184**, 3605–3613 (2002).
206. Visweswariah, S. S. & Busby, S. J. W. Evolution of bacterial transcription factors: How proteins take on new tasks, but do not always stop doing the old ones. *Trends Microbiol.* **23**, 463–467 (2015).
207. Weaver, T. M., Morrison, E. A. & Musselman, C. A. Reading More than Histones: The Prevalence of Nucleic Acid Binding among Reader Domains. *Molecules* **23**, 1–25 (2018).
208. Ullmann, A., Tillier, F. & Monod, J. Catabolite modulator factor: A possible mediator of catabolite repression in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* (1976). doi:10.1073/pnas.73.10.3476
209. Dessein, A., Tillier, F. & Ullmann, A. Catabolite modulator factor: Physiological properties and in vivo effects. *MGG Mol. Gen. Genet.* (1978). doi:10.1007/BF00333854
210. Ray, N. G., Vieth, W. R. & Venkatasubramanian, K. Regulation of lac operon

expression in mixed sugar chemostat cultures. *Biotechnol. Bioeng.* **29**, 1003–1014 (1987).
211. Inada, T., Kimata, K. & Aiba, H. Mechanism responsible for glucose-lactose diauxie in Escherichia coli: Challenge to the cAMP model. *Genes to Cells* **1**, 293–301 (1996).
212. Bettenbrock, K. *et al.* Correlation between growth rates, EIIACrr phosphorylation, and intracellular cyclic AMP levels in Escherichia coli K-12. *J. Bacteriol.* **189**, 6891–6900 (2007).
213. Notley-McRobb, L., Death, A. & Ferenci, T. The relationship between external glucose concentration and cAMP levels inside Escherichia coli: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. *Microbiology* **143**, 1909–1918 (1997).
214. Mandelstam, J. The Repression of Constitutive β-galactosidase in Escherichia coli by Glucose and other Carbon sources. *Biochem. J.* **82**, 489–493 (1962).
215. You, C. *et al.* Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**, 301–6 (2013).
216. Bren, A. *et al.* Glucose becomes one of the worst carbon sources for E.coli on poor nitrogen sources due to suboptimal levels of cAMP. *Sci. Rep.* **6**, 2–11 (2016).
217. Daniel, J. & Danchin, A. 2-Ketoglutarate as a possible regulatory metabolite involved in cyclic AMP-dependent catabolite repression in Escherichia coli K12. *Biochimie* **68**, 303–310 (1986).
218. Doucette, C. D., Schwab, D. J., Wingreen, N. S. & Rabinowitz, J. D. A-Ketoglutarate Coordinates Carbon and Nitrogen Utilization Via Enzyme I Inhibition. *Nat. Chem. Biol.* **7**, 894–901 (2011).
219. Koch, A. L. The Adaptive Responses of Escherichia coli to a Feast and Famine Existence. *Adv. Microb. Physiol.* **6**, 147–217 (1971).
220. Nyström, T. Aging in bacteria. *Curr. Opin. Microbiol.* **5**, 596–601 (2002).
221. Finkel, S. E. Long-term survival during stationary phase: Evolution and the GASP phenotype. *Nat. Rev. Microbiol.* **4**, 113–120 (2006).
222. Navarro Llorens, J. M., Tormo, A. & Martínez-García, E. Stationary phase in gram-negative bacteria. *FEMS Microbiol. Rev.* **34**, 476–495 (2010).
223. Vulić, M. & Kolter, R. Evolutionary cheating in Escherichia coli stationary phase cultures. *Genetics* **158**, 519–526 (2001).
224. Hengge, R. *The General Stress Response in Gram-Negative Bacteria*. *Bacterial Stress Responses, Second Edition* (2011). doi:10.1128/9781555816841.ch15
225. Jishage, M., Iwata, A., Ueda, S. & Ishihama, A. Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: Intracellular levels of four species of sigma subunit under various growth conditions. *J. Bacteriol.* **178**, 5447–5451 (1996).
226. Lempp, M. *et al.* Systematic identification of metabolites controlling gene expression in E. coli. *Nat. Commun.* **10**, (2019).
227. Lange, R. & Hengge-Aronis, R. The cellular concentration of the σS subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability. *Genes Dev.* **8**, 1600–1612 (1994).
228. Amores, G. R., De Las Heras, A., Sanches-Medeiros, A., Elfick, A. & Silva-Rocha, R. Systematic identification of novel regulatory interactions controlling biofilm formation in the bacterium Escherichia coli. *Sci. Rep.* **7**, 1–14 (2017).
229. Weber, H. *et al.* Genome-Wide Analysis of the General Stress Response Network in. *Society* **187**, 1591–1603 (2005).
230. Battesti, A., Majdalani, N. & Gottesman, S. The RpoS-Mediated General Stress

Response in Escherichia coli . *Annu. Rev. Microbiol.* **65**, 189–213 (2011).
231. McCann, M. P., Kidwell, J. P. & Matin, A. The putative σ factor KatF has a central role in development of starvation-mediated general resistance in Escherichia coli. *J. Bacteriol.* **173**, 4188–4194 (1991).
232. Reeve, C. A., Bockman, A. T. & Matin, A. Role of protein degradation in the survival of carbon-starved Escherichia coli and Salmonella typhimurium. *J. Bacteriol.* **157**, 758–763 (1984).
233. Nyström, T. Stationary-Phase Physiology. *Annu. Rev. Microbiol.* **58**, 161–181 (2004).
234. Santos, J. M., Lobo, M., Matos, A. P. A., De Pedro, M. A. & Arraiano, C. M. The gene bolA regulates dacA (PBP5), dacC (PBP6) and ampC (AmpC), promoting normal morphology in Escherichia coli. *Mol. Microbiol.* **45**, 1729–1740 (2002).
235. Huisman, G. W., Siegele, D. A., Zambrano, M. & Kolter, R. Morphological and Physiological Changes During Stationary Phase. in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology. American Society for Microbiology, Washington, DC,* 1672–1682 (1996).
236. Almiron, M., Link, A. J., Furlong, D. & Kolter, R. A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli. *Genes Dev.* **6**, 2646–2654 (1992).
237. Nair, S. & Finkel, S. E. Dps protects cells against multiple stresses during stationary phase. *J. Bacteriol.* **186**, 4192–4198 (2004).
238. Altuvia, S., Almirón, M., Huisman, G., Kolter, R. & Storz, G. The dps promoter is activated by OxyR during growth and by IHF and σS in staionary phase. *Mol. Microbiol.* **13**, 265–272 (1994).
239. Cashel, M. & Gallant, J. Two Compounds implicated in the Function of RC gene. *Nature* **221**, 838–841 (1969).
240. Chang, D. E., Smalley, D. J. & Conway, T. Gene expression profiling of Escherichia coli growth transitions: An expanded stringent response model. *Mol. Microbiol.* **45**, 289–306 (2002).
241. Jishage, M., Kvint, K., Shingler, V. & Nyström, T. Regulation of σ factor competition by the alarmone ppGpp. *Genes Dev.* **70**, 1260–1270 (2002).
242. Wada, A., Yamazaki, Y., Fujita, N. & Ishihama, A. Structure and probable genetic location of a 'ribosome modulation factor' associated with 100S ribosomes in stationary-phase Escherichia coli cells. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2657–2661 (1990).
243. Izutsu, K., Wada, A. & Wada, C. Expression of ribosome modulation factor (RMF) in Escherichia coli requires ppGpp. *Genes to Cells* **6**, 665–676 (2001).
244. Groat, R. G., Schultz, J. E., Zychlinsky, E., Bockman, A. & Matin, A. Starvation proteins in Escherichia coli: Kinetics of synthesis and role in starvation survival. *J. Bacteriol.* **168**, 486–493 (1986).
245. Farrell, M. J. & Finkel, S. E. The Growth Advantage in Stationary-Phase Phenotype Conferred by rpoS Mutations Is Dependent on the pH and Nutrient Environment. *J. Bacteriol.* **185**, 7044–7052 (2003).
246. Schurig-Briccio, L. A., Farías, R. N., Rodríguez-Montelongo, L., Rintoul, M. R. & Rapisarda, V. A. Protection against oxidative stress in Escherichia coli stationary phase by a phosphate concentration-dependent genes expression. *Arch. Biochem. Biophys.* **483**, 106–110 (2009).
247. Lange, R. & Hengge-Aronis, R. Identification of a central regulator of stationary-phase gene expression in Escherichia coli. *Mol. Microbiol.* **5**, 49–59 (1991).

248. Gonidakis, S. & Longo, V. D. Assessing Chronological Aging in Bacteria. *Cell Senescence* **965**, 421–437 (2013).
249. Kolter, R. The Stationary Phase of the Bacterial Life Cycle. *Annu. Rev. Microbiol.* **47**, 855–874 (1993).
250. Hardiman, T., Lemuth, K., Keller, M. A., Reuss, M. & Siemann-Herzberg, M. Topology of the global regulatory network of carbon limitation in Escherichia coli. *J. Biotechnol.* **132**, 359–374 (2007).
251. Gentry, D. R. & Cashel, M. Mutational analysis of the Escherichia coli spoT gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. *Mol. Microbiol.* **19**, 1373–1384 (1996).
252. Haurylik, V., Atkinson, G. C., Murakami, K. S., Tenson, T. & Gerdes, K. Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Micobiology* **13**, 298–309 (2015).
253. Potrykus, K. & Cashel, M. (p)ppGpp: Still Magical? *Annu. Rev. Microbiol.* (2008). doi:10.1146/annurev.micro.62.081307.162903
254. Cashel, M. & Gallant, J. Two compounds implicated in the Function of the RC Gene of Escherichia coli. *Nature* **221**, 838–841 (1969).
255. Sørensen, M. A., Fehler, A. O. & Lo Svenningsen, S. Transfer RNA instability as a stress response in Escherichia coli: Rapid dynamics of the tRNA pool as a function of demand. *RNA Biol.* **15**, 586–593 (2018).
256. Ross, W. *et al.* PpGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response. *Mol. Cell* **62**, 811–823 (2016).
257. Johansson, J. *et al.* Nucleoid proteins stimulate stringently controlled bacterial promoters: A link between the cAMP-CRP and the (p)ppGpp regulons in Escherichia coli. *Cell* **102**, 475–485 (2000).
258. Kanjee, U., Ogata, K. & Houry, W. A. Direct binding targets of the stringent response alarmone (p)ppGpp. *Mol. Microbiol.* **85**, 1029–1043 (2012).
259. Xiao, H. *et al.* Residual guanosine 3',5'-bispyrophosphate synthetic activity of relA null mutants can be eliminated by spoT null mutations. *J. Biol. Chem.* **266**, 5980–5990 (1991).
260. Farewell, A., Kvint, K. & Nyström, T. Negative regulation by RpoS: A case of sigma factor competition. *Mol. Microbiol.* **29**, 1039–1051 (1998).
261. Magnusson, L. U., Farewell, A. & Nyström, T. ppGpp: A global regulator in Escherichia coli. *Trends Microbiol.* **13**, 236–242 (2005).
262. Gentry, D. R., Hernandez, V. J., Nguyen, L. H., Jensen, D. B. & Cashel, M. Synthesis of the stationary-phase sigma factor σ(s) is positively regulated by ppGpp. *J. Bacteriol.* **175**, 7982–7989 (1993).
263. Brown, L., Gentry, D., Elliott, T. & Cashel, M. DksA affects ppGpp induction of RpoS at a translational level. *J. Bacteriol.* **184**, 4455–4465 (2002).
264. Bleuven, C. & Landry, C. R. Molecular and cellular bases of adaptation to a changing environment in microorganisms. *Proc. R. Soc. B Biol. Sci.* **283**, (2016).
265. Rosenberg, S. M. Evolving responsively: Adaptive mutation. *Nat. Rev. Genet.* **2**, 504–515 (2001).
266. Morreall, J. *et al.* Evidence for Retromutagenesis as a Mechanism for Adaptive Mutation in Escherichia coli. *PLoS Genet.* **11**, 1–12 (2015).
267. Cairns, J., Overbaugh, J. & MIller, S. The origin of mutants. *Nature* **335**, 142–145 (1988).

268. Hall, B. G. Adaptive mutagenesis: A process that generates almost exclusively beneficial mutations. *Genetica* **102–103**, 109–125 (1998).
269. Taddei, F., Matic, I. & Radman, M. cAMP-dependent SOS induction and mutagenesis in resting bacterial populations. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 11736–11740 (1995).
270. Bjedov, I. *et al.* Stress-Induced Mutagenesis in Bacteria. *Science (80-.).* **300**, 1404–1410 (2003).
271. Bharatan, S. M., Reddy, M. & Gowrishankar, J. Distinct Signatures for Mutator Sensitivity of lacZ Reversions and for the Spectrum of lacI/lacO Forward Mutations on the Chromosome of Nondividing Escherichia coli. *Genetics* **166**, 681–692 (2004).
272. Roth, J. R., Kugelberg, E., Reams, A. B., Kofoid, E. & Andersson, D. I. Origin of Mutations Under Selection: The Adaptive Mutation Controversy. *Annu. Rev. Microbiol.* **60**, 477–501 (2006).
273. Foster, P. L. Adaptive Mutation in Escherichia coli. *J. Bacteriol.* **186**, 4846–4852 (2004).
274. Koonin, E. V. & Wolf, Y. I. Is evolution Darwinian or/and Lamarckian? *Biol. Direct* **4**, 42 (2009).
275. Bridges, B. A. Starvation-associated mutation in Escherichia coli: a spontaneous lesion hypothesis for 'directed' mutation. *Mutat. Res.* **307**, 149–156 (1994).
276. Sekowska, A., Wendel, S., Fischer, E. C., Nørholm, M. H. H. & Danchin, A. Generation of mutation hotspots in ageing bacterial colonies. *Sci. Rep.* **6**, 2 (2016).
277. Hall, B. G. Spectrum of mutations that occur under selective and non-selective conditions in E. coli. *Genetica* (1991). doi:10.1007/BF00116545
278. Bridges, B. Mutations in resting cells: the role of endogenous DNA Damage. *Cancer Surv.* **28**, 155–167 (1996).
279. Brégeon, D., Doddridge, Z. A., You, H. J., Weiss, B. & Doetsch, P. W. Transcriptional mutagenesis induced by uracil and 8-oxoguanine in Escherichia coli. *Mol. Cell* **12**, 959–970 (2003).
280. Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. & Feist, A. M. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* **56**, 1–16 (2019).
281. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution - principles and applications for biotechnology. *Microb. Cell Fact.* **12**, 1–17 (2013).
282. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence durign 2,00 Generations. *Am. Nat.* **138**, 1315–1341 (1991).
283. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).
284. LaCroix, R. A. *et al.* Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. *Appl. Environ. Microbiol.* **81**, 17–30 (2015).
285. Notley-McRobb, L. & Ferenci, T. The generation of multiple co-existing mal-regulatory mutations through polygenic evolution in glucose-limited populations of Escherichia coli. *Environ. Microbiol.* **1**, 45–52 (1999).
286. Cooper, A. L., Dean, A. C. R. & Hinshelwood, C. Factors Affecting the Growth of Bacterial Colonies on Agar Plates. *Proc. Roy. Soc.* **171**, 175–199 (1968).
287. Taddei, F., Halliday, J. A., Matic, I. & Radman, M. Genetic analysis of mutagenesis in aging Escherichia coli colonies. *Mol. Gen. Genet.* **256**, 277–281 (1997).
288. Saint-Ruf, C. *et al.* Massive diversification in aging colonies of Escherichia coli. *J. Bacteriol.* **196**, 3059–3073 (2014).

289. Sekiguchi, M. & Tsuzuki, T. Oxidative nucleotide damage: Consequences and prevention. *Oncogene* **21**, 8895–8904 (2002).
290. Fricke, J., Neuhard, J., Kelln, R. A. & Pedersen, S. The cmk gene encoding cytidine monophosphate kinase is located in the rpsA operon and is required for normal replication rate in Escherichia coli. *J. Bacteriol.* (1995). doi:10.1128/jb.177.3.517-523.1995
291. Notley-Mcrobb, L., King, T. & Ferenci, T. rpoS Mutations and Loss of General Stress Resistance in Escherichia coli Populations as a Consequence of Conflict between Competing Stress Responses. *J. Bacteriol.* **184**, 806–811 (2002).
292. Zambrano, M. M., Siegele, D. A., Almirn, M., Tormo, A. & Koltert, R. Microbial Competition: Escherichia coli Mutants That Take Over Stationary Phase Cultures. *Science (80-.).* **259**, 1757–1760 (1993).
293. Cohen, G. N. *Microbial Biochemistry, Springer*. (2014).
294. Martinussen, J., Willemoës, M. & Kilstrup, M. Nucleotide Metabolism. in *Comprehensive Biotechnology, Second Edition* (2011). doi:10.1016/B978-0-08-088504-9.00010-6
295. Turnbough, C. L. & Switzer, R. L. Regulation of Pyrimidine Biosynthetic Gene Expression in Bacteria: Repression without Repressors. *Microbiol. Mol. Biol. Rev.* **72**, 266–300 (2008).
296. Andersen, P. S., Frees, D., Fast, R. & Mygind, B. Uracil uptake in Escherichia coli K-12: Isolation of uraA mutants and cloning of the gene. *J. Bacteriol.* **177**, 2008–2013 (1995).
297. Danielsen, S., Kilstrup, M., Barilla, K., Jochimsen, B. & Neuhard, J. Characterization of the Escherichia coli codBA operon encoding cytosine permease and cytosine deaminase. *Mol. Microbiol.* **6**, 1335–1344 (1992).
298. Maier, C., Bremer, E., Schmid, A. & Benz, R. Pore-forming activity of the Tsx protein from the outer membrane of Escherichia coli. Demonstration of a nucleoside-specific binding site. *J. Biol. Chem.* **263**, 2493–2499 (1988).
299. Benz, R., Schmid, A., Maier, C. & Bremer, E. Characterization of the nucleoside-binding site inside the Tsx channel of Escherichia coli outer membrane Reconstitution experiments with lipid bilayer membranes. *Eur. J. Biochem.* **176**, 699–705 (1988).
300. Holst, B., Søgaard-Andersen, L., Pedersen, H. & Valentin-Hansen, P. The cAMP-CRP/CytR nucleoprotein complex in Escherichia coli: two pairs of closely linked binding sites for the cAMP-CRP activator complex are involved in combinatorial regulation of the cdd promoter. *EMBO J.* **11**, 3635–3643 (1992).
301. Josephsen, J. & Hammer-Jespersen, K. Fusion of the lac Genes to the Promoter for the Cytidine Deaminase Gene of Escherichia coli K-12. *Mol Gen Genet* **182**, 154–158 (1981).
302. Gavigan, S. A., Nguyen, T., Nguyen, N. & Senear, D. F. Role of multiple CytR binding sites on cooperativity, competition, and induction at the Escherichia coli udp promoter. *J. Biol. Chem.* **274**, 16010–16019 (1999).
303. Gerlach, P. *et al.* The cyclic AMP (cAMP)-cAMP receptor protein complex functions both as an activator and as a corepressor at the tsx-p2 promoter of Escherichia coli K-12. *J. Bacteriol.* **173**, 5419–5430 (1991).
304. Chahla, M., Wooll, J., Laue, T. M., Nguyen, N. & Senear, D. F. Role of protein-protein bridging interactions on cooperative assembly of DNA-bound CRP-CytR-CRP complex and regulation of the Escherichia Coli CytR regulon. *Biochemistry* **42**, 3812–3825 (2003).
305. Reaves, M. L., Young, B. D., Hosios, A. M., Xu, Y. F. & Rabinowitz, J. D. Pyrimidine

homeostasis is accomplished by directed overflow metabolism. *Nature* (2013). doi:10.1038/nature12445
306. Schaaper, R. M. & Mathews, C. K. Mutational consequences of dNTP pool imbalances in E. coli. *DNA Repair (Amst).* **12**, 73–79 (2013).
307. Kantrowitz, E. R. Allostery and cooperativity in Escherichia coli aspartate transcarbamoylase. *Arch. Biochem. Biophys.* **519**, 81–90 (2012).
308. Gerhart, J. C. & Pardee, A. B. The enzymology of control by feedback inhibition. *J. Biol. Chem.* **237**, 891–896 (1962).
309. Cornish-Bowden, A. Curbing the excesses of low demand The handiwork of tinkering. *Nature* **500**, 157–158 (2013).
310. Deutscher, M. P. Degradation of Stable RNA in Bacteria. *J. Biol. Chem.* **278**, 45041–45044 (2003).
311. Liang, W., Malhotra, A. & Deutscher, M. P. Acetylation Regulates the Stability of a Bacterial Protein: Growth Stage-dependent Modification of RNase R. *Mol Cell* **44**, 160–166 (2011).
312. Rinas, U., Hellmuth, K., Kang, R., Seeger, A. & Schlieker, H. Entry of Escherichia coli into stationary phase is indicated by endogenous and exogenous accumulation of nucleobases. *Appl. Environ. Microbiol.* **61**, 4147–4151 (1995).
313. Rossi, E. *et al.* Cellulose production is coupled to sensing of the pyrimidine biosynthetic pathway via c-di-GMP production by the DgcQ protein of Escherichia coli. *Environ. Microbiol.* (2017). doi:10.1111/1462-2920.13918
314. Varik, V., Oliveira, S. R. A., Hauryliuk, V. & Tenson, T. HPLC-based quantification of bacterial housekeeping nucleotides and alarmone messengers ppGpp and pppGpp. *Sci. Rep.* **7**, 1–12 (2017).
315. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* (2009). doi:10.1038/nchembio.186
316. Zborníková, E., Knejzlík, Z., Hauryliuk, V., Krásný, L. & Rejman, D. Analysis of nucleotide pools in bacteria using HPLC-MS in HILIC mode. *Talanta* **205**, (2019).
317. Buckstein, M. H., He, J. & Rubin, H. Characterization of nucleotide pools as a function of physiological state in Escherichia coli. *J. Bacteriol.* **190**, 718–726 (2008).
318. Pastan, I. & Pearlman, R. Cyclic Adenosine Monophosphate in Bacteria. *Science (80-.).* **169**, 339–344 (1970).
319. Rabinowitz, J. D. & Silhavy, T. J. Metabolite turns master regulator. *Nature* **500**, 283–284 (2013).

Publications

Paper 1
CRISPR/Cas9-based genome editing for simultaneous interference with gene expression and protein stability

Paper 2
Bacterial Genome Editing Strategy for Control of Transcription and Protein Stability.

Paper 3
A versatile one-step CRISPR-Cas9 based approach to plasmid-curing

Paper 4
Standardized Cloning and Curing of Plasmids

Paper 5
Mutations in the Global Transcription Factor CRP/CAP: Insights from Experimental Evolution and Deep Sequencing

Paper 6
Temporal evolution of master regulator Crp identifies pyrimidines as catabolite modulator factors

Paper 7
CRISPR-Cas9 - den revolutionerende gensaks (in Danish)

Paper 1

CRISPR/Cas9-based genome editing for simultaneous interference with gene expression and protein stability

Published online 7 September 2017

Nucleic Acids Research, 2017, Vol. 45, No. 20 ***e171***
doi: 10.1093/nar/gkx797

CRISPR/Cas9-based genome editing for simultaneous interference with gene expression and protein stability

Virginia Martínez†, Ida Lauritsen†, Tonja Hobel, Songyuan Li, Alex Toftgaard Nielsen and Morten H. H. Nørholm*

Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, DK-2800, Denmark

Received January 03, 2017; Revised July 06, 2017; Editorial Decision August 26, 2017; Accepted August 30, 2017

ABSTRACT

Interference with genes is the foundation of reverse genetics and is key to manipulation of living cells for biomedical and biotechnological applications. However, classical genetic knockout and transcriptional knockdown technologies have different drawbacks and offer no control over existing protein levels. Here, we describe an efficient genome editing approach that affects specific protein abundances by changing the rates of both RNA synthesis and protein degradation, based on the two cross-kingdom control mechanisms CRISPRi and the N-end rule for protein stability. In addition, our approach demonstrates that CRISPRi efficiency is dependent on endogenous gene expression levels. The method has broad applications in e.g. study of essential genes and antibiotics discovery.

INTRODUCTION

The combined impact of synthesis and degradation dynamically determines protein levels in living cells. There is a growing need for synthetic biology tools that can control the abundance of specific proteins, e.g. for the fine-tuning of enzymes in metabolic pathways or studies of essential genes for which genetic knockouts are lethal (1). Existing methods typically focus either on genetic knockouts, conditional repression of transcription or direct interference with protein function or stability. However, a combination of these approaches is desirable to achieve a more controllable, rapid or stronger repression of the amount of selected proteins in the cell. Furthermore, for conditional removal of proteins, stability is a key factor even in fast-growing microbes such as yeast where the majority of proteins are very long-lived (2).

Given the recent rapid development in synthetic biology and genome editing technologies, we asked to what extent it was possible to harness generic molecular mechanisms for simultaneously controlling both protein synthesis and stability. To this end we first looked for a cross-kingdom protein regulatory mechanism. The N-end rule states that the identity of the N-terminal residue (N-degron) of a protein is a prime determinant of its half-life across all kingdoms of life (3). Conveniently, the tobacco etch virus (TEV) protease can accommodate most amino acid residues in the P1' position following the cleavage site (ENLYFQ↓X, where X denotes all amino acids except proline) and thus this small recognition site can mask an N-degron (4,5), and we noted that the N-terminal location enables simultaneous manipulation of the translational initiation region (TIR) - a region surrounding the start codon that has a major impact on gene expression levels (6,7). This enables manipulation of both protein degradation and translation initiation, which is important when manipulating essential genes as shown later.

The CRISPR–Cas9 system enables cutting of very specific DNA sequences in a wide variety of living organisms and has revolutionized our ability to edit genetic information (8). The system has also been repurposed to regulate transcription—e.g. to activate (9) or repress gene expression in e.g. *Escherichia coli* (10), *Bacillus subtilis* (11), human cells (12) and *Saccharomyces cerevisiae* (13) in a general approach known as CRISPR interference (CRISPRi). CRISPRi is based on a catalytically inactive Cas9 endonuclease (dCas9), which can interfere with transcription by binding to specific DNA sequences with the aid of a guide RNA. However, the efficiency of CRISPRi-based systems is affected by endogenous gene expression levels (9). Thus, concomitant manipulation of endogenous gene expression levels may expand the applicability of CRISPRi. Here, we develop a single workflow that combines conditional protein degradation with CRISPRi and TIR manipulation and apply the system to study essential genes and develop strains hypersensitive to antibiotics.

*To whom correspondence should be addressed. Tel: +45 217 99184; Fax: +45 353 33300; Email: morno@biosustain.dtu.dk
†These authors contributed equally to this work as first authors.

MATERIALS AND METHODS

Plasmid construction

The pPROTi plasmid resulted from a triple PCR-fragment assembly *via* USER cloning, as described previously (14). For further details on all the plasmids described here see Supplementary Table S1. We amplified the pSEVA33 backbone, the L-rhamnose inducible promoter PrhaBAD, and the S219D mutant of TEV protease from pSEVA331 (15), pKS1 (16) and pKM586 (17) with the oligonucleotides oMSB1270/1267, oMSB1268/1269 and oMSB1271/1273, respectively. For further details on all the oligonucleotides described here see Supplementary Table S2. The pMAZ-SK plasmid with different guide RNAs used to PROTi tag essential genes by CRMAGE were constructed by USER cloning (18). This was done by mixing two annealed oligonucleotides that were complement to the amplified pMAZ-SK backbone after USER treatment, as described previously (19). Specifically, for the genes *accD, fabG, ftsZ, glmS, ileS, murE, pheS, ribD, prfB, rnpA, tmk, acpS, ispH, murA, dapE, lpxC* and *ribE*, the oligonucleotide pairs oMSB2565/2566, oMSB2591/2592, oMSB2569/2570, oMSB2571/2572, oMSB2573/2574, oMSB2575/2576, oMSB2577/2578, oMSB2593/2594, oMSB2579/2580, oMSB2583/2584, oMSB2585/2586, oMSB2750/2551, oMSB2740/2741, oMSB2742/2743, oMSB2748/2749, oMSB2744/2745 and oMSB2746/2747 were used, respectively. To construct the pCRiPi plasmid, the pPROTi plasmid was PCR amplified using the oligonucleotide pair oMSB1865/2312, and the *dCas9* gene (including the aTc promoter and the terminator) was amplified from pdCas9 with the oligonucleotides oMSB2313/oMSB1866. For construction of the pgRNA-CRiPi plasmid targeting the PROTi tag for the CRiPi system, the pSLQ1236 (Supplementary Table S1) was used as backbone. The gRNA was changed by standard Gibson assembly (20) with oligonucleotides oSONG145/146. Nucleotide sequences of pPROTi, pCRiPi and pgRNA-CRiPi are provided in the supplementary information.

Bacterial strains, media and growth conditions

The parental strain for all experiments was *E. coli* MG1655 (ATCC 47076). *E. coli* MG1655 with GFP integrated into the genome (21) was initially used to tag *gfp* first by recombineering (22) with the ssDNA MAGE oligonucleotide oMSB1277 and then with the ssDNA MAGE oligonucleotides oMSB1275 (unstable GFP variant with phenylalanine as the N-degron) or oMSB1276 (stable GFP variant with alanine). These two GFP variants were used as PCR templates to add the IPTG inducible T5 promoter with the oligonucleotides oMSB1661 and oMSB1662. After, T5-GFP variants were integrated into the *E. coli* MG1655 genome by clonetegration (23), using pOSIP-KT, and the oligonucleotides oMSB1297 and oMSB1298. FLP-mediated excision was performed as previously described (23). All *E. coli* strains were grown in lysogeny broth (LB) at 37°C shaking at 300 rpm. The antibiotics ampicillin (100 μg/mL), chloramphenicol (30 μg/ml) or kanamycin (50 μg/ml) were added when needed. *gfp* expression was induced with 0.1 mM IPTG; *tev* protease expression was induced with 5 mM rhamnose; and *dCas9* endonuclease expression was induced with 200 ng/ml aTc.

Genomic integration of the PROTi tag

The PROTi tag was inserted after the first codon downstream of the start codon in genes of interest, according to the previously described CRMAGE protocol (19). The starting strain for CRMAGE was MG1655 K-12 harboring the pMA7CR_2.0 and pZS4Int-tetR plasmids. The pMA7CR_2.0 plasmid expresses the Cas9 nuclease, the λ-red β-proteins and the dam protein that represses the mismatch DNA repair system for obtaining higher genome editing efficiency (19) 5μM of ssDNA CRMAGE oligonucleotide and 250 ng of pMAZ-SK plasmid with inserted gRNA were used for electroporation. Cultures were grown at 37°C in Julabo SW22 linear-shaking water-bath at 250 rpm. After aTc addition to express *Cas9* according to the previously published protocol (19), recovery was performed overnight. The ssDNA CRMAGE oligonucleotide contained 35–45 nucleotide end homology on each side of the insertion. For each gene, a PAM sequence (5′-NGG-3′) and the adjacent gRNA (20 nucleotides) were chosen in the coding sequence. The pMAZ-SK plasmids with inserted gRNA were constructed as described above. One nucleotide in the PAM sequence was changed in the ssDNA CRMAGE oligonucleotide to avoid Cas9 recognition of mutants with the inserted PROTi tag. Randomization of the TIR was done by changing the specific nucleotides in the ssDNA CRMAGE oligonucleotide used for insertion of the PROTi tag. For the genes *accD, fabG, ftsZ, glmS, ileS, murE, pheS, ribD, prfB, rnpA, tmk, acpS, ispH, murA, dapE, lpxC* and *ribE*, the ssDNA CRMAGE oligonucleotides oMSB2595, oMSB2596, oMSB2597, oMSB2598, oMSB2599, oMSB2600, oMSB2601, oMSB2651, oMSB2602, oMSB2603, oMSB2604, oMSB2757, oMSB2752, oMSB2753, oMSB2756, oMSB2754 were used, respectively.

PROTi characterization

E. coli MG1655 strains containing the PROTi tagged *gfp* in the genome and harboring the pPROTi plasmid, were inoculated from an overnight culture to OD_{600} 0.01 in LB supplemented with chloramphenicol and IPTG. After 4 h of growing, the cultures were induced by adding 5 mM rhamnose. To wash out IPTG from the culture medium before inducing with rhamnose, the cultures were centrifuged and resuspended in the same volume of LB with chloramphenicol and rhamnose. After 1–4 h, cultures were analyzed in a SynergyMx Microplate reader from Biotek. For GFP fluorescence quantification, emission was detected at 512 nm with the excitation light of 480 nm and 80 level gain.

Flow cytometry

Flow cytometry measurements were performed on a FACS Aria (Becton–Dickinson, San Jose, USA) with 488 nm excitation from a blue solid-state laser. Cells were diluted 1:100 in PBS for analysis. At least 20 000 cells were collected for each measurement. FlowJo (Treestar) was used for data analysis.

Growth profiles

Growth was monitored in the BioLector® from m2p-labs. Each of the tagged PROTi clones were diluted 1:100 in 1 ml LB supplied with appropriate antibiotics in a FlowerPlate (48-well MTP, flower) for the BioLector®. Cultures were induced with rhamnose and aTc from the beginning of growth and grown at 37°C, shaking at 1200 rpm.

CFU assays and drop tests

To determine colony forming units (CFU), cultures of PROTi tagged essential genes harboring the pPROTi plasmid were inoculated in 1 ml LB supplied with chloramphenicol and rhamnose in a 96-well microtiter plate. After 4 h of growth, cultures were plated on LB agar plates in serial dilutions. For the CRiPi system, cultures were inoculated in 1 ml LB supplied with chloramphenicol and ampicillin and grown for 4 h. The CRiPi system was induced with rhamnose and aTc and the cultures were grown for additional 4 h before plating. For drop tests, overnight cultures of tagged essential gene variants were diluted to the same OD and 10-fold serial dilutions were performed. From each dilution, 10 μl were placed on LB agar plates with appropriate antibiotics.

Fosfomycin sensitivity

Fosfosmycin sensitivity was determined with minimum inhibitory concentration (MIC) assays as previously described (24). Stock solution of fosfomycin disodium (Sigma Aldrich) salt was dissolved in MiliQ water (1 mg/ml). Briefly, 2-fold serial dilutions of antibiotic stocks (from 0.015 to 256 g/l) were made in 150 μl LB medium supplemented with the appropriate antibiotics, rhamnose and aTc (for CRiPi-induction) and OD after 18 h was determined and plotted relative to the growth of the same cells in the absence of fosfomycin. For testing fosfomycin sensitivity upon PROTi induction, cultures of *murE1*-tagged cells harbouring the CRiPi system were inoculated in 1 ml LB supplied with chloramphenicol, ampicillin and rhamnose in a 96-well microtiter plate. After 4 h of growth, cultures were plated on LB agar plates supplied with different concentrations of fosfomycin.

RESULTS

We first engineered a pro-N-degron module by incorporating the corresponding nucleotide sequence at the 5′ end of a gene on the *E. coli* chromosome using CRMAGE genome editing (19). The pro-N-degron module consists of a seven amino acid peptide recognition site of the tobacco etch virus (TEV) protease (ENLYFQ↓F) (5) and an eleven amino acid-linker (25). In the presence of TEV protease, the peptide is cleaved and an N-end-rule substrate is generated with phenylalanine as the prime protein-destabilizing factor (Figure 1A). Importantly, the entire coding sequence for the pro-N-degron module is small enough to fit into a standard oligonucleotide compatible with CRMAGE. With this protein interference (PROTi) system, by rhamnose-inducible expression of the TEV protease from a plasmid, the N-degron becomes de-protected and the protein is targeted for proteasomal degradation through the N-end rule pathway.

To characterize the system, the PROTi tag was fused to the N-terminus of GFP by integrating the coding sequence into the *E. coli* genome with a synthetic IPTG-inducible T5 promoter (Figure 1A). In the resulting strain, GFP production was induced by adding IPTG followed by growth for four hours. At this stage, expression of the TEV protease was induced with rhamnose. Three hours after addition of rhamnose, GFP levels showed a strong decrease (83%—measured by whole cell fluorescence) compared to the uninduced control (Figure 1B). Further characterization of the system by flow cytometry revealed a broad population of cells with different fluorescence levels upon induced protein degradation (Figure 1C), which we hypothesized was caused by simultaneous strong GFP synthesis driven by the T5 promoter. Washing out IPTG from the culture medium prior to the induction of the PROTi system with rhamnose confirmed this hypothesis as it resulted in a homogeneous non-fluorescent population (Figure 1D). The reduction in GFP fluorescence can be ascribed to the generated N-degron, since GFP fluorescence remained high in cells harboring a stable PROTi tag variant, with the phenylalanine of the N-degron replaced by an alanine (Figure 1D).

The fact that protein abundance is a function of both synthesis and stability prompted us to turn to the broadly applicable CRISPR–Cas9-based gene regulation technology. With the aim of gaining control over both transcription and protein stability with a single genome-editing step, we developed a CRISPRi–PROTi (CRiPi) system, where dCas9 can be produced together with the TEV protease (Figure 2A), thereby enabling simultaneous inhibition of gene expression as well as degradation of the target protein.

Based on previous studies, the gRNA was designed to bind to the non-template DNA in the 5′ end of the gene, and *dCas9* was expressed from a plasmid with an anhydrous tetracycline (aTc)-inducible promoter (10). Moreover, we designed the gRNA so that it only targets the CRMAGE-inserted sequence, which encodes the TEV protease recognition site and the N-degron-linker, thereby creating a generic gRNA target independent from the site of insertion (Figure 2A).

As shown in Figure 2B, cellular depletion of GFP was rapidly achieved with high efficiency when the CRiPi system was induced with both rhamnose and aTc. Specifically, 76% of the cells showed complete loss of fluorescence after two hours of induction. In contrast, the induction of *dCas9* expression alone caused only a slightly reduced GFP fluorescence, denoting the high stability of GFP (Figure 2B).

The technology described above is particularly useful for analysis of genes that are essential for the organism and thus inaccessible with traditional knockout approaches. Thus, to further demonstrate the functionality of the system, we compared the effectiveness of the PROTi, CRISPRi and CRiPi technologies to control the level of proteins encoded by essential genes in *E. coli*. In a previous approach all essential genes in *E. coli* were individually targeted with a sequence encoding a C-terminal protein degradation tag (*mf-ssrA*) but 67 proteins could not be tagged this way despite repeated attempts (26). We noted that 54 of these 67 'inaccessible' genes were part of operons, making lethal polar effects a likely explanation. Thus, the N-terminal location of

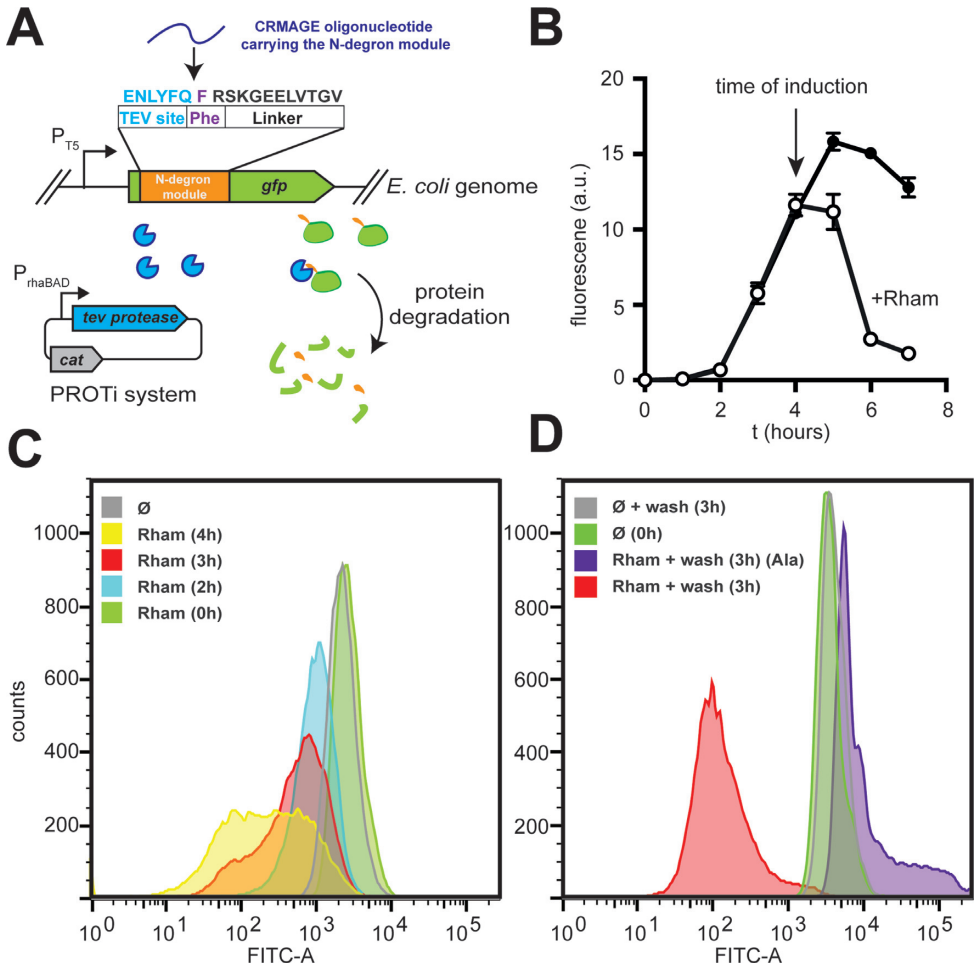


Figure 1. PROTi**:** Development and characterization of the PROTi system to control protein abundances. (**A**) Schematic illustration of the PROTi system. The coding sequence for the pro-N-degron module (orange) is integrated by CRMAGE at the 5′-end of a genomically integrated *gfp* (green). *gfp* is under control of the IPTG-inducible T5 promoter. With PROTi, the TEV protease (blue) is expressed from the PrhaBAD promoter leading to de-protection of the N-degron followed by degradation of GFP (green) through the N-end rule pathway. (**B**) Whole-cell fluorescence measurement of cells expressing GFP tagged with the pro-N-degron—with and without the PrhaBAD inducer rhamnose (Rham). Data represents the average of three biological replicates and bars show the standard error. (**C**) Flow cytometry histogram displaying the fluorescence signals after 0, 2, 3 or 4 h induction with rhamnose or without induction (Ø). (**D**) Fluorescence signals after 3 h of PROTi induction with rhamnose, while removing the IPTG inducer from the culture medium by washing. A stable PROTi tag with alanine (Ala) replacing phenylalanine was included as control.

the PROTi tag could enable the targeting of some of these essential genes and since the inserted module overlaps with the TIR, expression tuning by nucleotide variation in this region could minimize polar effects caused by e.g. changes in translational speed.

Using CRMAGE, we attempted targeting of ten essential genes that were not previously *mf*-ssrA-tagged (*glmS, ileS, murE, pheS, ribD, tmk, accD, prfB, fabG* and *rnpA*) and seven that were previously *mf*-ssrA-tagged (*ftsZ, acpS, ispH, murA, dapE, lpxC* and *ribE*) (26) using the pro-N-degron module designed with a TIR library made of six random nucleotides upstream from the start codon and all synonymous codons sampled in two positions following the start codon (Figure 3A). Remarkably, this way we were able to identify insertions in seven of the 10 genes that were not previously *mf*-ssrA-tagged, despite their location in essential operons—as well as three out of the seven previously *mf*-ssrA-tagged genes (Supplementary Table S3). Not surprisingly, we also noted a high variability in colony size directly after CRMAGE, and in cell viability assays (Supplementary Figure S1), as a directly observable consequence of the TIR variation. This is highly useful both for gauging the success rate of CRMAGE and when searching for variants with wild type gene expression levels. For example, for *ileS*, 11 small colonies were screened by colony PCR and all had the PROTi tag inserted. From 12 big colonies, 11 were negative. For the *rnpA* gene, five out of 12 small colonies were positive, whereas 10 out of 11 of the big colonies were not tagged.

To demonstrate the value of the TIR variation approach, for five of the genes (*murE, pheS, rpnA, ileS* and *ribD*), we attempted to insert the corresponding sequences without TIR variation – preserving the six nucleotides upstream from the start codon from the native gene context. For three of these genes we were unable to isolate tagged clones, indicating a lethal effect, whereas we could isolate clones with the *ileS* and *rnpA* genes tagged (Figure 3B and C). However, these clones clearly exhibited growth defects both directly after

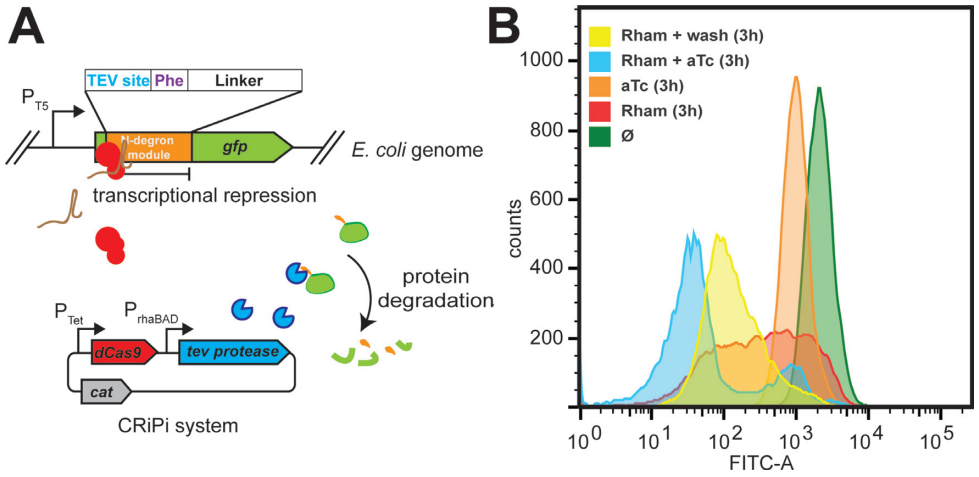


Figure 2. CRiPi**:** Schematic illustration of the CRISPRi-PROTi (CRiPi) system. (**A**) Cellular depletion of the targeted protein can be accelerated by simultaneous expression of *dCas9* (red) from the Ptet promoter and the TEV protease (blue) from the PrhaBAD promoter. Here shown for *gfp* as an example. The dCas9 targets the genomically integrated pro-N-degron encoding sequence with the aid of a guide RNA (gRNA, brown, curved line) and represses transcription. (**B**) Fluorescence after 3 h of PROTi, PROTi and wash, CRISPRi, or CRiPi induction with rhamnose and the Ptet inducer anhydrous tetracycline (aTc), or with no induction (Ø).

CRMAGE and in subsequent viability tests (Figure 3B and C). In contrast, using the TIR randomization approach, we were able to isolate tagged gene variants without any observable growth defects.

After having obtained both growth-affected and unaffected PROTi-tagged gene variants, we moved on to study the CRiPi system, by inducing protein degradation and/or gene silencing. Most of the strains were unaffected in growth in liquid cultures after PROTi induction with rhamnose (Supplementary Figure S2). In fact, rhamnose had a small stimulatory effect on growth in several cases (Supplementary Figure S2). However, the tagged *murE, ribD* and *pheS* strains exhibited 1–2 orders of magnitude decrease in viability when plating the cells after PROTi induction (Figure 3D) and most of the strains showed growth retardation after CRISPRi induction with aTc (Supplementary Figure S2). Only the tagged *murE* strain showed a clear effect when inducing PROTi directly in liquid culture (Figure 3E). Of the two approaches utilized in the CRiPi method, CRISPRi had the strongest overall effect (Supplementary Figure S2). However, for the *ileS* and *pheS*-strains the repression of growth after plating was clearly enhanced by simultaneous targeting of both transcription and protein stability (Figure 3F), thereby showing the versatility and strength of the CRiPi method.

To demonstrate the relevance of CRiPi and PROTi for applied biotechnology, we explored its performance as a tool for creating antibiotic hypersensitive strains for use in antibiotic discovery. When screening large compound libraries it is challenging to supply sufficiently high concentrations of each compound, which leads to false negatives in the screen. Lowering the concentration of essential protein targets, enable high-throughput screening with sub-inhibitory drug concentrations and discovery of combinatorial drugs and targets (27). We focused on MurE, a central enzyme in peptidoglycan biosynthesis (Figure 3G). When CRiPi or PROTi were induced in cells harboring PROTi-tagged MurE, a significant decrease in viability was observed (Figure 3D and H). We next demonstrated *murE* as a potential target for creating hypersensitive strains that can be used for screening compound libraries to identify agents with antibacterial activity. Fosfomycin is an antibiotic that causes specific inhibition of the enzyme MurA, which is involved in the same peptidoglycan biosynthetic pathway as MurE (Figure 3G). By applying fosfomycin to cells with an induced CRiPi or PROTi system targeting MurE, the sensitivity to the antibiotic increased, depicted as complete growth inhibition at lower concentrations of the antibiotic, compared to the non-induced control (Figure 3I and Supplementary Figure S3). Moreover, fosfomycin sensitivity varied markedly in clones with different *murE*-TIR backgrounds upon induction of CRiPi (Figure 3I and Supplementary Figure S4).

DISCUSSION

Our approach has some limitations: N-terminal peptide tagging of essential proteins may not always be allowed as the tag itself could compromise activity. Furthermore, when attached to some proteins, TEV cleavage or subsequent targeting to the ClpP protease may not be efficient. However, PROTi could serve as a complement to other protein destabilizing technologies (e.g. proteins that are compromised by C-terminal tagging). Here, we were able to target 7 out of 10 proteins that previously had failed with a C-terminal destabilizing tag approach. It is possible that an even higher success rate could be obtained with additional screening efforts (and almost certainly for targeting of non-essential genes). Four out of these seven proteins (encoded by *murE, ribD, ileS* and pheS) were sensitive to induction of PROTi.

CRISPRi generally had the highest growth-effect on essential genes, compared to PROTi, possibly due to the reduced affinity of TEV protease with the N-end rule substrate Phe in the P1' position (5) or because essential protein depletion is compensated for by gene expression upregulation. The approach could potentially benefit from increas-

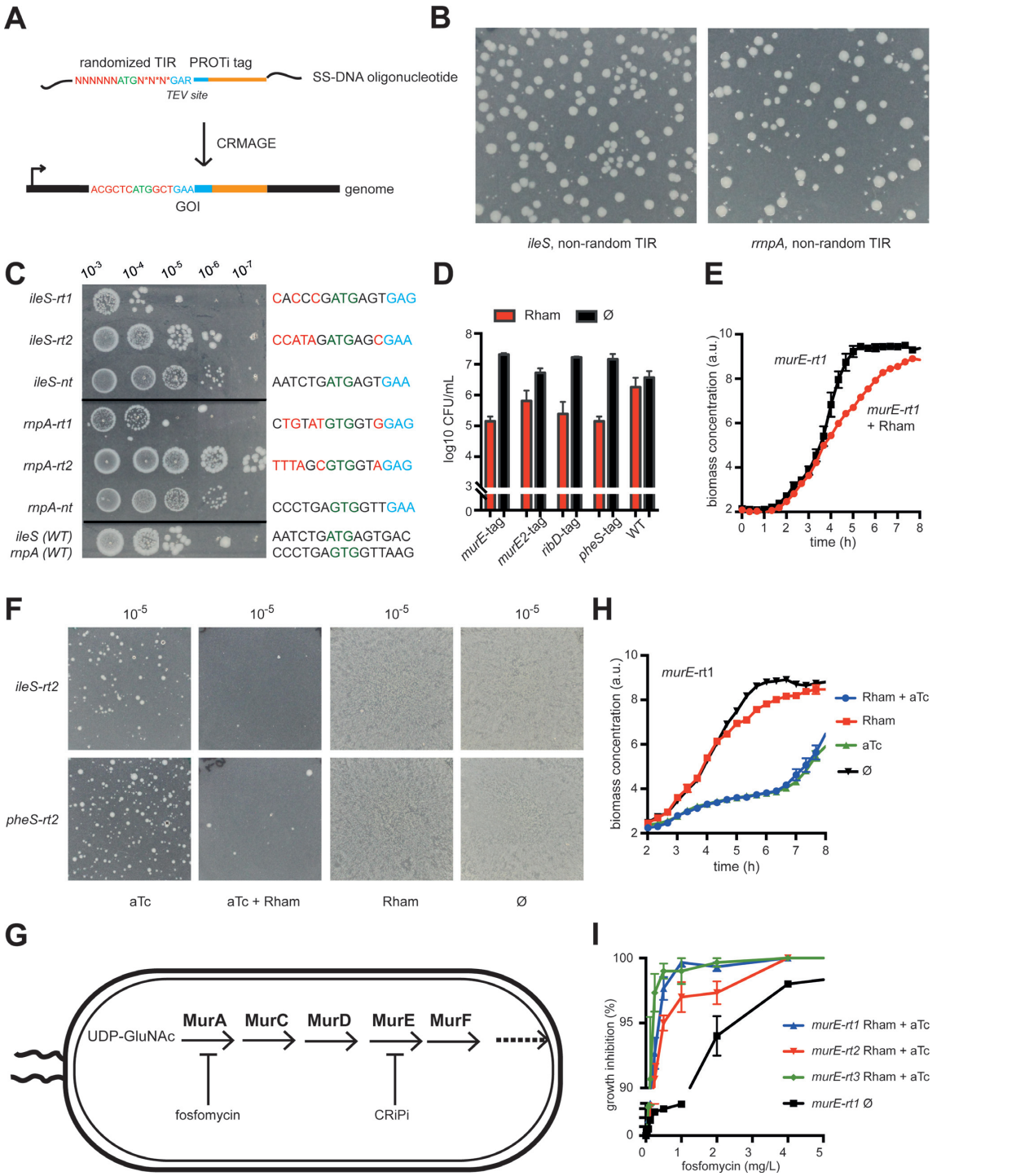


Figure 3. PROTi, CRiPi and TIR randomization can control protein levels and produce antibiotic hypersensitive bacteria. (**A**) Randomization of the translation initiation region (TIR, red font) in a ssDNA oligonucleotide enables CRMAGE-based insertion of the PROTi tag (orange) containing the TEV protease recognition site (blue) in genes of interest (GOI). (**B**) CRMAGE cultures plated directly after insertion of the PROTi tag in *ileS* (left) and *rnpA* (right) genes without TIR variation. Note the presence of both small and large colonies. The small colonies contained the PROTi tagged gene variants, as verified by colony PCR. (**C**) Drop tests of cells with *ileS* and *rnpA* tagged with and without TIR variation (rt: randomized TIR, nt: non-random TIR). TIR region sequences are indicated next to the different drop test lanes - red font indicates nucleotides different from the wild type sequence, blue font highlights the TEV protease recognition site and the start codon is highlighted in green. (**D**) Viability of control strain (WT with the PROTi plasmid) and PROTi variants tagged in the essential *E. coli* genes *murE, ribD* and *pheS* with and without rhamnose (Rham) induction. (**E**) Growth profile of *murE*-rt1 cells carrying the PROTi system with and without Rham induction. (**F**) Agar plates illustrating the effect of *ileS*- and *pheS*-tagged cells with CRISPRi (aTc), PROTi (Rham) or CRiPi (aTc + Rham) induced after 4 h and plated after an additional 4 h of growth. Without inducers (Ø) growing bacteria completely cover the plate. (**G**) Illustration of the Mur enzymes involved in peptidoglycan biosynthesis pathway. (**H**) Growth profile of uninduced *murE*-rt1 cells (Ø) or induced with Rham and/or aTc. (**I**) Fosfomycin sensitivity upon CRiPi induction in *murE*-tagged clones or no induction (Ø). Growth was tested in increasing concentrations of fosfomycin after 18 h incubation. All values are the averages of three biological replicates and bars show standard error.

ing the *in vivo* TEV protease activity, or by increasing the activity of the endogenous ClpA/P/S factors as shown previously in a similar system (29).

We were initially surprised by the almost complete lack of effect of PROTi in log phase cultures whereas subsequent plating resulted in several orders of magnitude reduction in growth for some of the PROTi targeted essential gene products. A similar observation was made recently in a high-throughput targeting of essential genes with CRISPRi in *B. subtilis* (11). There, it was suggested that (essential) protein levels are important for outgrowth from stationary phase, whereas maximal growth rate in log phase is less affected. It is possible that our CRISPRi system is more efficient than the one described for *Bacillus*, whereas the weaker effect of PROTi more resembles the *Bacillus* CRISPRi efficiency.

By inserting the sequence as a TIR variation library it is possible to create expression variants that can be screened with minimal polar effects, mimicking the natural gene expression level. Furthermore, some of the TIR variants, e.g. *rnpA*-rt1 and *rnpA*-rt1 (Supplementary Figure S2) or *murE*-rt1 and *murE*-rt2 (Figure 3H and Supplementary Figure S2) varied significantly in their sensitivity to CRISPRi, providing further support for the relationship between gene expression and CRISPRi efficiency (9). Thus, this type of multi-level reverse genetics tool may further expand the utility of the highly successful CRISPR/Cas system.

Bacterial antibiotic resistance is rapidly exhausting the number of available effective antimicrobial agents. Consequently, there is an urgent need to identify new target-specific inhibitors to develop antimicrobial compounds (28). With our combined CRiPi approach, insertion of a simple and inexpensive oligonucleotide enables subtle tuning of potential antibiotic targets on both the transcriptional and posttranslational level. Furthermore, the system represents a unique and versatile workflow that may enable future in-depth characterization of essential genes located in operons.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Novo Nordisk Foundation; Novo Nordisk and Novozymes (to I.L.). Funding for open access charge: Novo Nordisk Foundation
Conflict of interest statement. None declared.

REFERENCES

1. Brockman,I.M. and Prather,K.L.J. (2015) Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnol. J.*, **10**, 1360–1369.
2. Christiano,R., Nagaraj,N., Fröhlich,F. and Walther,T.C. (2014) Global proteome turnover analyses of the yeasts S. cerevisiae and S. pombe. *Cell Rep.*, **9**, 1959–1966.
3. Bachmair,A., Finley,D. and Varshavsky,A. (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.
4. Taxis,C., Stier,G., Spadaccini,R., Knop,M., Biology,C. and Unit,B. (2009) Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. *Mol. Syst. Biol.*, **5**, 1–7.
5. Kapust,R.B., Tözsér,J., Copeland,T.D. and Waugh,D.S. (2002) The P1' specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.*, **294**, 949–955.
6. Kozak,M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, **361**, 13–37.
7. McCarthy,J.E. and Gualerzi,C. (1990) Translational control of prokaryotic gene expression. *Trends Genet.*, **6**, 78–85.
8. Wright,A.V., Nunez,J.K. and Doudna,J.A. (2016) Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell*, **164**, 29–44.
9. Chavez,A., Scheiman,J., Vora,S., Pruitt,B.W., Tuttle,M., P R Iyer,E., Lin,S., Kiani,S., Guzman,C.D., Wiegand,D.J. *et al.* (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods*, **12**, 326–328.
10. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
11. Peters,J.M., Ph,D., Larson,M.H., Hawkins,J.S., Lu,C.H.S., Koo,B., Shiver,A.L., Whitehead,E.H., Weissman,J.S., Brown,E.D. *et al.* (2016) A Comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell*, **165**, 1493–1506.
12. Kiani,S., Beal,J., Ebrahimkhani,M.R., Huh,J., Hall,R.N., Xie,Z., Li,Y. and Weiss,R. (2014) CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods*, **11**, 723–726.
13. Gilbert,L.A., Larson,M.H., Morsut,L., Liu,Z., Brar,G.A., Torres,S.E., Stern-Ginossar,N., Brandman,O., Whitehead,E.H., Doudna,J.A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442–451.
14. Cavaleiro,A.M., Kim,S.H., Seppälä,S., Nielsen,M.T. and Nørholm,M.H.H. (2015) Accurate DNA assembly and genome engineering with optimized uracil excision cloning. *ACS Synth. Biol.*, **4**, 1042–1046.
15. Silva-Rocha,R., Martínez-García,E., Calles,B., Chavarría,M., Arce-Rodríguez,A., De Las Heras,A., Páez-Espino,A.D., Durante-Rodríguez,G., Kim,J., Nikel,P.I. *et al.* (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.*, **41**, 666–675.
16. Söderström,B., Skoog,K., Blom,H., Weiss,D.S., von Heijne,G. and Daley,D.O. (2014) Disassembly of the divisome in *Escherichia coli*: evidence that FtsZ dissociates before compartmentalization. *Mol. Microbiol.*, **92**, 1–9.
17. Kapust,R.B. and Waugh,D.S. (2000) Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.*, **19**, 312–318.
18. Nour-Eldin,H.H., Hansen,B.G., Nørholm,M.H.H., Jensen,J.K. and Halkier,B.A. (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.*, **34**, e122.
19. Ronda,C., Pedersen,L.E., Sommer,M.O.A. and Nielsen,A.T. (2016) CRMAGE: CRISPR optimized MAGE recombineering. *Sci. Rep.*, **6**, 19452.
20. Gibson,D.G., Young,L., Chuang,R.-Y., Venter,J.C., Hutchison,C.a, Smith,H.O., Iii,C.A.H. and America,N. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.
21. Bonde,M.T., Pedersen,M., Klausen,M.S., Jensen,S.I., Wulff,T., Harrison,S., Nielsen,A.T., Herrgård,M.J. and Sommer,M.O.A. (2016) Predictable tuning of protein expression in bacteria. *Nat. Methods*, **13**, 233–236.
22. Lennen,R.M., Nilsson Wallin,A.I., Pedersen,M., Bonde,M., Luo,H., Herrgård,M.J. and Sommer,M.O.A. (2015) Transient overexpression of DNA adenine methylase enables efficient and mobile genome engineering with reduced off-target effects. *Nucleic Acids Res.*, **44**, e36.
23. St-Pierre,F., Cui,L., Priest,D.G., Endy,D., Dodd,I.B. and Shearwin,K.E. (2013) One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.*, **2**, 537–541.
24. Wiegand,I., Hilpert,K. and Hancock,R.E.W. (2008) Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.*, **3**, 163–175.

25. Erbse,A., Schmidt,R., Bornemann,T., Mogk,A., Zahn,R. and Dougan,D.A. (2006) ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature*, **439**, 753–756.
26. Cameron,D.E. and Collins,J.J. (2014) Tunable protein degradation in bacteria. *Nat. Biotechnol.*, **32**, 1276–1283.
27. Liu,A., Tran,L., Becket,E., Lee,K., Chinn,L., Park,E., Tran,K. and Miller,J.H. (2010) Antibiotic sensitivity profiles determined with an Escherichia coli gene knockout collection: generating an antibiotic bar code. **54**, 1393–1403.
28. Fair,R.J. and Tor,Y. (2014) Antibiotics and bacterial resistance in the 21st century. *Perspect. Med. Chem.*, **6**, 25–64.
29. Sekar,K., Gentile,A.M., Bostick,J.W. and Tyo,K.E.J. (2016) N-terminal-based targeted, inducible protein degradation in Escherichia coli. *PLoS One*, **11**, 1–17.

Supplementary information for

CRISPR/Cas9-based genome editing for simultaneous interference with gene expression and protein degradation

Virginia Martínez[1], Ida Lauritsen[1], Tonja Hobel, Songyuan Li, Alex Toftgaard Nielsen and Morten H. H. Nørholm[*]

Figure S1

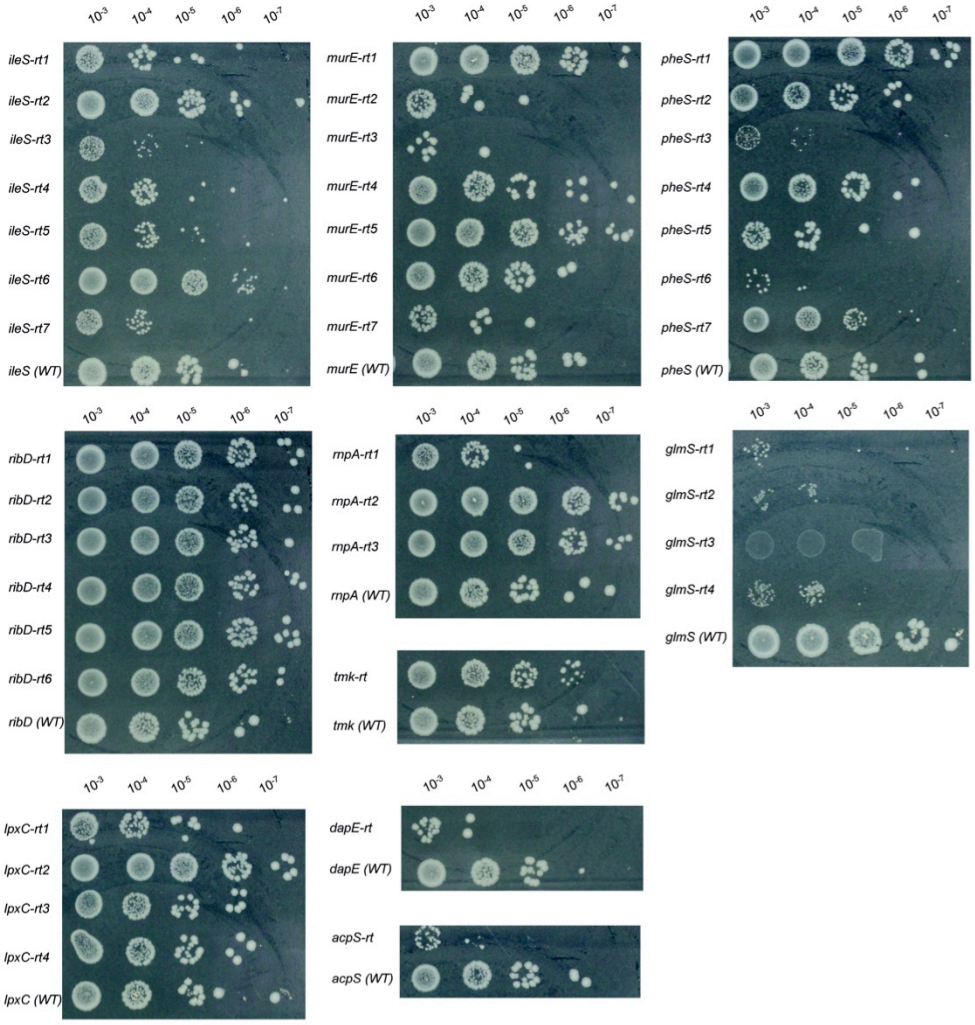


Figure S1. Growth drop tests of isolated variants of ten essential genes tagged with TIR variation. 10-fold serial dilutions were performed from overnight cultures and 10 μL were plated on LB agar. *E. coli* K12 MG1655 serves as control strain (WT).

Figure S2

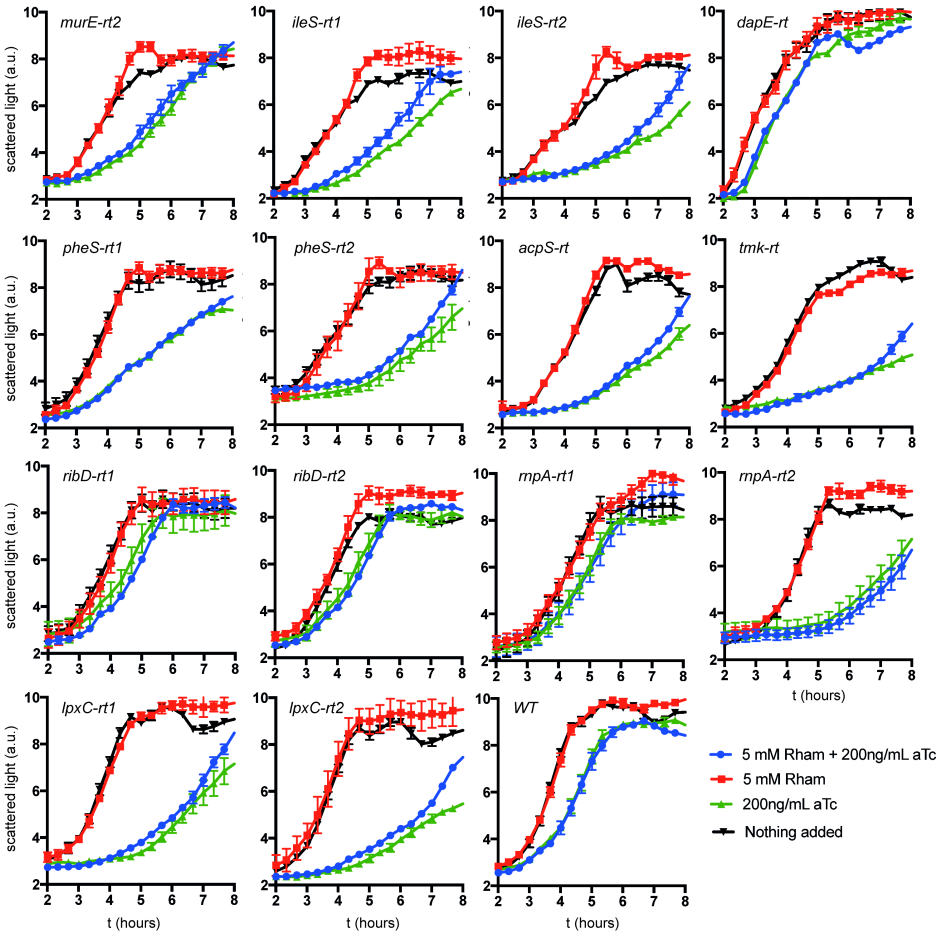


Figure S2. Growth profiles of wild type *E. coli* K12 MG1655 and 14 different tagged essential gene variants when inducing CRiPi with aTc, rhamnose or both. Growth was monitored over time in a BioLector® where scattered light represents biomass concentration. Rhamnose (red) aTc (green), or rhamnose and aTc (blue) were added from the beginning of growth or without inducers (black). The different variants of the same essential gene have different TIR regions. The CRiPi system was not tested for tagged essential gene variants of *glmS* due to growth defects as shown in Figure S1.Results are mean values of biological triplicates with error-bars showing standard error. Rham: rhamnose, aTc: anhydrotetracycline, a.u.: arbitrary unit.

Figure S3

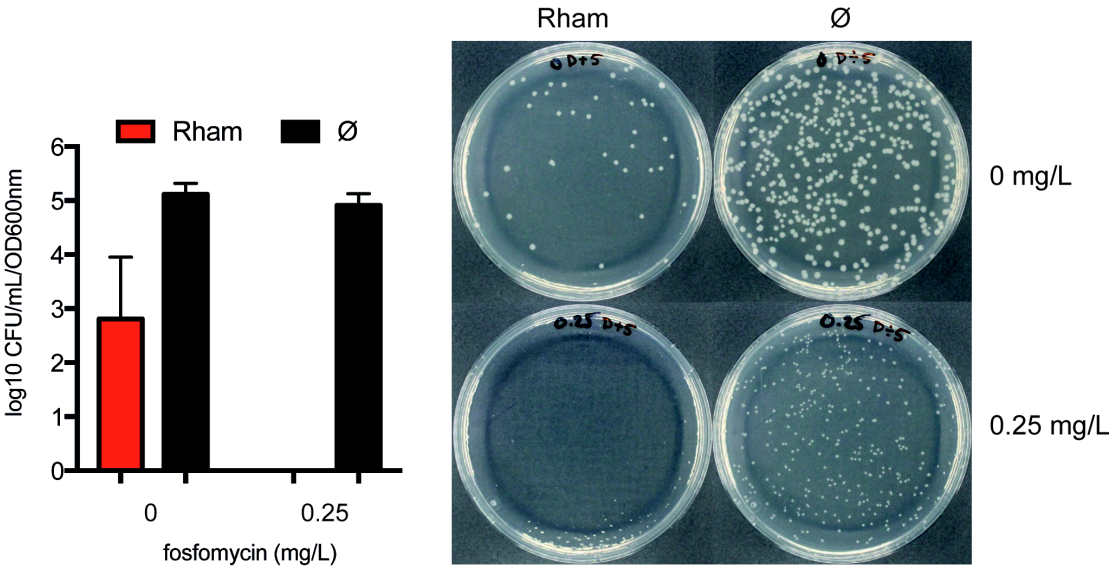


Figure S3. Fosfomycin sensitivity upon PROTi induction in *murE-rt*1-tagged clone or no induction (Ø). Cells carrying the CRiPi system were grown for 4 hours and plated on LB agar with different fosfomycin concentrations. All values are the averages of five biological replicates and bars show standard error. Representative LB agar plates with 0 mg/mL(top pnael) or 0.25 mg/L (bottom panel) of fosfomycin.

Figure S4

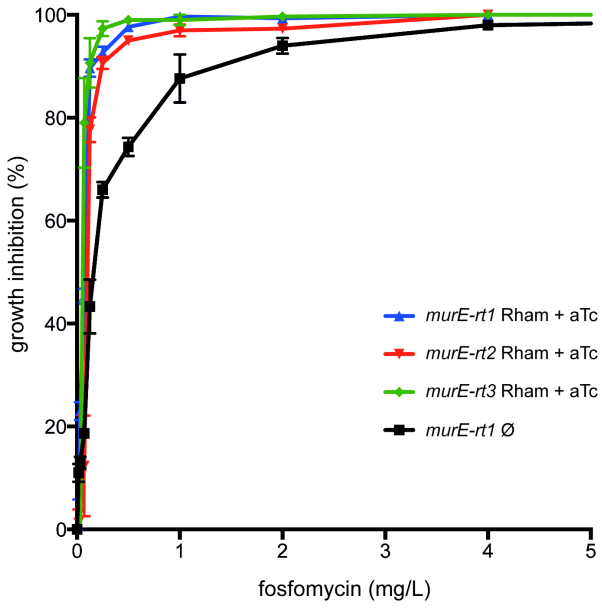


Figure S4. Fosfomycin sensitivity upon CRiPi induction in *murE*-tagged clones or no induction (Ø). Growth was tested in increasing concentrations of fosfomycin after 18 h incubation. All values are the averages of three biological replicates and bars show standard error. Same data as in Figure 3I.

Table S1

Table S1. Plasmids used in this publication

Plasmid ID	Relevant genotype, description	Reference/source
pSEVA331	pBBR1, Cm^R	(1)
pPROTi	pSEVA331 derivative plasmid with *tev* protease under the PrhaBAD promoter	This publication
pdCas9	dCas9 expression plasmid, *dCas9* gene expressed under aTc inducible promoter	Addgene (2)
pCRiPi	pPROTi derivative plasmid with *dCas9* gene under aTc inducible promoter	This publication
pSLQ1236	gRNA expression plasmid, gRNA was expressed under aTc inducible promoter	(3)
pgRNA-CRiPi	pSLQ1236 derivative plasmid with gRNA targeting the PROTi tag under aTc inducible promoter	This publication
pZS4Int-tetR	Expression of tetR for CRMAGE	(4)
pMA7CR_2.0	Expression of λ-Red β -proteins, dam protein and Cas9 for CRMAGE	(4)
pMAZ-SK	Expression of target gRNA for Cas9 recognition for CRMAGE	(4)
pKS1	Expression of mCherry under rhamnose inducible promoter	(5)
pKM586	*tev* protease expression plasmid	Addgene (6)

Table S2

Table S2. Oligonucleotides used in this publication

Oligo ID	Oligo sequence (5' to 3')
oMSB1267	AAAGGCAUCAAATAAAACGAAAGGCTC

oMSB1268	ATGCCTTUAATTAATAAAAATAGGCGTATCACGAGGCCCTTTC
oMSB1269	ATGGTATATUCCTCCTGAATTTCATTACGAC
oMSB1270	ACTAGTCTUGGACTCCTGTTGATAGATC
oMSB1271	AATATACCAUGGGAGAAAGCTTGTTTAAG
oMSB1273	AAGACTAGUTTAATTCATGAGTTGAGTCGCTTCCTTAAC
oMSB1275	AGTGACAAGTTCTTCTCCTTTGCTACGAAACTGGAAGTACAGGTTTT CCACCATCTAG
oMSB1276	AGTGACAAGTTCTTCTCCTTTGCTACGAGCCTGGAAGTACAGGTTTT CCACCATCTAG
oMSB1277	GACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAACTCCAGTGAcAA GTTCTTCTCCTTTGCTACGTTACTGGAAGTACAGGTTTTCCACCATC TAGTATTTCTCCTCTTTAATCTCTAGTAGCTAGCACTGTACCTAGGAC TGAGC
oMSB1661	ATGCATCUTCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTA TAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTA AAGAGGAGAAATTCTAGATGGTGGAAAACCTGTACTTCCAG
oMSB1662	AGAGGGCUTCATTTGTAGAGCTCATCCATGCCATG
oMSB1865	ACGTTGAUACGCCTATTTTTATTAATTAAAGGCATCAAATAAAACGAA A
oMSB1866	ATCAACGUCTCATTTTCGCCAGATATCGACGTC
oSMB2312	AAGAAATUCACCACAATTCAGCAAATTGTGAACATC
oMSB2313	AATTTCTUTATAAACGCAGAAAGGCCCACCCGAAGG
oMSB1297	AGATGCAUGGCGCCTAACC
oMSB1298	AGCCCTCUAGAGGATCCCCGGGTAC
oMSB2595	ACACCCCTTCAGGAATGCTCGCCTTGCGGGTTGGAGTAATGTTGCT TTTAATTCGTTCAATCCAAACTCCAGTGACAAGTTCTTCTCCTTTGCT ACGAAACTGGAAGTACAGGTTYTCRCTCATNNNNNNCCTTTCTGTCT GAACCTGGTTCGATGCCAGT
oMSB2596	CGTTTCAGCAATTGCGCGGCCAATTCCGCGACTTGCACCGGTTACC AGTGCGATTTTTCCTTCAAAAACTCCAGTGACAAGTTCTTCTCCTTTG CTACGAAACTGGAAGTACAGGTTYTCRTTCATNNNNNNCCTCTTTTA AAGCTCGAGCGCCGCTGCC
oMSB2597	CCGCCGACGCCGATGACTTTAATCACCGCGTCATTTGTAAGTTCCAT TGGTTCAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAACTGGA AGTACAGGTTYTCRAACATNNNNNNTCTCCGATTTGTGCCTGTCGC CTGAGGCCGTAATCA
oMSB2598	GTATTCCAGACGACGTAAACCTTCAAGAAGTATTTCTGCTACATCAC GTTGCGCGATCGCGCCAACAATTCCAACTCCAGTGACAAGTTCTTC TCCTTTGCTACGAAACTGGAAGTACAGGTTYTCRCACATNNNNNNTG ATTCCGATTTATATCGTTGTCGGTCAACCTGT
oMSB2599	CCACGCATCGGGAACCCTGTTTCCGGCAAATTCAGTGTTGATTTATA GTCAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAACTGGAAGT ACAGGTTYTCRCTCATNNNNNNCTCGGTTCCGTATTTCGGTTTGATT

ACATAACAGGCTTA

oMSB2600 CGCGAAGGTGCGTCTGGCACCCACGGAGCAAGAAGATCGCGCAAA
TTACGATCAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAACTG
GAAGTACAGGTTYTCNGCCACNNNNNNCCTCGCCTTGATTAATCAC
AAATTCATTTTTATCGC

oMSB2601 ACATCTGACGCCTGGCTAATGGCCGCCTTCGCACTTGCAACCAGTT
CTGCGAGATGAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAAC
TGGAAGTACAGGTTYTCNGACATNNNNNNCCTCATTGTGTCAGTGG
TGACACTGGTTCGTTGGAC

oMSB2651 TGGGATGCGTGGTAAAACGTCCTCGTTGCGCAAGCTTTAGCGCCCG
CGCCATGTAATACTCGTCAACTCCAGTGACAAGTTCTTCTCCTTTGC
TACGAAACTGGAAGTACAGGTTYTCYTGCACNNNNNNGTCCTCCAG
GCGCGCGATCTCTTCGCCAAATTC

oMSB2602 CCCCCTAAGAACGTCGGAGCGTTCCGTGAGTTCCTGAATGCGATTA
TTTACCGGATTAATTTCAACTCCAGTGACAAGTTCTTCTCCTTTGCTA
CGAAACTGGAAGTACAGGTTYTCRAACATNNNNNNATTTCTTTTATT
GAGCTAGTCAAAATGCGGT

oMSB2603 CGTCTGACCGTTTCTAAGTAATAAAGCTAACNNNNNNGTGGTNGAR
AACCTGTACTTCCAGTTTCGTAGCAAAGGAGAAGAACTTGTCACTGG
AGTTAAGCTCGCATTTCCAAGGGAGTTACGCTTGTTAACTCCAAGTC
AATTCACATTCGTCTTCCAGCAGCCA

oMSB2604 ATTACGCGCGGTAGTTTTGCCTGCGCCTTCAAGCCCCTCAATGACG
ATATACTTACTAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAAC
TGGAAGTACAGGTTYTCNCGCATNNNNNNCCTTAAGCACTTTCAGAT
AATCCTGCACAGA

oMSB2757 GGGCGATCACCGCTTCGATGCGAGCGATCTCAACAATATCCGTGCC
TAAACCTAATATAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAA
CTGGAAGTACAGGTTYTCNGCCATNNNNNNCGCGCTTCCAGCATCA
GACGCTTCATTTCTGCC

oMSB2752 AGCGCGGTCTACCCCGGCACAAAAACCACGCGGGTTTGCCAACAG
GATAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAACTGGAAGT
ACAGGTTYTCYTGCATNNNNNNCCTCCAGTGCCGGATCGATTTCCA
GCACTTCAA

oMSB2753 TTTAGCGCCGGAAATTGTGACTTCGCCCTGTAGCTTCGTTGGCCCC
TGAACACGAAATTTAACTCCAGTGACAAGTTCTTCTCCTTTGCTACG
AAACTGGAAGTACAGGTTYTCRTCCATNNNNNNTGTTCTCAGTTAAC
AATTCATATCCGCTACCGGC

oMSB2756 GATTCCAGTTATCAGCAATTTTTCCATGAGGTGNNNNNNATGTCNGA
RAACCTGTACTTCCAGTTTCGTAGCAAAGGAGAAGAACTTGTCACTG
GAGTTTGTCCGGTTATTGAGCTGACACAACAGCTTATTCGCCG

oMSB2754 GACTTTCTTGCCGGTATGTAAACCGACACCAGTCGCCTGAACGATA
CGTTTAAGTGTCCTTTGTTTAACTCCAGTGACAAGTTCTTCTCCTTTG
CTACGAAACTGGAAGTACAGGTTYTCRATCATNNNNNNATCTCGCC
AAATTACCTATCCAACCGAAGTGTA

oMSB2755 CGCGATGGTGATGGCGACGCGAGCGTCCGGTGTAGCAACGTTAGC
TTCAATAATAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAACTG
GAAGTACAGGTTYTCRTTCATNNNNNNCCTTCGGGTTCGAGTATGG
CCCCGCAGGGGGGCG

oMSB2662	CCACGCATCGGGAACCCTGTTTCCGGCAAATTCAGTGTTGATTTATA GTCAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAACTGGAAGT ACAGGTTTTCACTCATCAGATTCTCGGTTCCGTATTTCGGTTTGATTA CATAACAGGCTTA
oMSB2663	CGCGAAGGTGCGTCTGGCACCCACGGAGCAAGAAGATCGCGCAAA TTACGATCAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAACTG GAAGTACAGGTTTTCTGCCACCTGTCCCCTCGCCTTGATTAATCACA AATTCATTTTTATCGC
oMSB2664	ACATCTGACGCCTGGCTAATGGCCGCCTTCGCACTTGCAACCAGTT CTGCGAGATGAACTCCAGTGACAAGTTCTTCTCCTTTGCTACGAAAC TGGAAGTACAGGTTTTCTGACATGGTTTTCCTCATTGTGTCAGTGGT GACACTGGTTCGTTGGAC
oMSB2665	TGGGATGCGTGGTAAAACGTCCTCGTTGCGCAAGCTTTAGCGCCCG CGCCATGTAATACTCGTCAACTCCAGTGACAAGTTCTTCTCCTTTGC TACGAAACTGGAAGTACAGGTTTTCCTGCACGGCTTAGTCCTCCAG GCGCGCGATCTCTTCGCCAAATTC
oMSB2666	CGTCTGACCGTTTCTAAGTAATAAAGCTAACCCCTGAGTGGTTGAAA ACCTGTACTTCCAGTTTCGTAGCAAAGGAGAAGAACTTGTCACTGGA GTTAAGCTCGCATTTCCAAGGGAGTTACGCTTGTTAACTCCAAGTCA ATTCACATTCGTCTTCCAGCAGCCA
oMSB2565	GAGCACAGGAATGCTCGCCTTGCGGGGTTTTAGAGCTAGAAAT
oMSB2566	CTAAAACCCCGCAAGGCGAGCATTCCTGTGCTCAGTATCTCT
oMSB2591	GAGCACCAATTGCGCGGCCAATTCCGGTTTTAGAGCTAGAAAT
oMSB2592	CTAAAACCGGAATTGGCCGCGCAATTGGTGCTCAGTATCTCT
oMSB2569	GAGCACGACTTTAATCACCGCGTCATGTTTTAGAGCTAGAAAT
oMSB2570	CTAAAACATGACGCGGTGATTAAAGTCGTGCTCAGTATCTCT
oMSB2571	GAGCACGACGACGTAAACCTTCAAGAGTTTTAGAGCTAGAAAT
oMSB2572	CTAAAACTCTTGAAGGTTTACGTCGTCGTGCTCAGTATCTCT
oMSB2573	GAGCACCCCTGTTTCCGGCAAATTCAGTTTTAGAGCTAGAAAT
oMSB2574	CTAAAACTGAATTTGCCGGAAACAGGGGTGCTCAGTATCTCT
oMSB2575	GAGCACCTGGCACCCACGGAGCAAGAGTTTTAGAGCTAGAAAT
oMSB2576	CTAAAACTCTTGCTCCGTGGGTGCCAGGTGCTCAGTATCTCT
oMSB2577	GAGCACGCTAATGGCCGCCTTCGCACGTTTTAGAGCTAGAAAT

oMSB2578	CTAAAACGTGCGAAGGCGGCCATTAGCGTGCTCAGTATCTCT
oMSB2593	GAGCACCATGGCGCGGGCGCTAAAGCGTTTTAGAGCTAGAAAT
oMSB2594	CTAAAACGCTTTAGCGCCCGCGCCATGGTGCTCAGTATCTCT
oMSB2579	GAGCACGAACGTCGGAGCGTTCCGTGGTTTTAGAGCTAGAAAT
oMSB2580	CTAAAACCACGGAACGCTCCGACGTTCGTGCTCAGTATCTCT
oMSB2583	GAGCACGAAGACGAATGTGAATTGACGTTTTAGAGCTAGAAAT
oMSB2584	CTAAAACGTCAATTCACATTCGTCTTCGTGCTCAGTATCTCT
oMSB2585	GAGCACGTATATCGTCATTGAGGGGCGTTTTAGAGCTAGAAAT
oMSB2586	CTAAAACGCCCCTCAATGACGATATACGTGCTCAGTATCTCT
oMSB2750	GAGCACAGGTTTAGGCACGGATATTGGTTTTAGAGCTAGAAAT
oMSB2751	CTAAAACCAATATCCGTGCCTAAACCTGTGCTCAGTATCTCT
oMSB2740	GAGCACGGCACAAAAACCACGCGGGTGTTTTAGAGCTAGAAAT
oMSB2741	CTAAAACACCCGCGTGGTTTTTGTGCCGTGCTCAGTATCTCT
oMSB2742	GAGCACCGGAAATTGTGACTTCGCCCGTTTTAGAGCTAGAAAT
oMSB2743	CTAAAACGGGCGAAGTCACAATTTCCGGTGCTCAGTATCTCT
oMSB2748	GAGCACGTTGTGTCAGCTCAATAACCGTTTTAGAGCTAGAAAT
oMSB2749	CTAAAACGGTTATTGAGCTGACACAACGTGCTCAGTATCTCT
oMSB2744	GAGCACTAAACGTATCGTTCAGGCGAGTTTTAGAGCTAGAAAT
oMSB2745	CTAAAACTCGCCTGAACGATACGTTTAGTGCTCAGTATCTCT
oMSB2746	GAGCACGATGGCGACGCGAGCGTCCGGTTTTAGAGCTAGAAAT
oMSB2747	CTAAAACCGGACGCTCGCGTCGCCATCGTGCTCAGTATCTCT
oSONG145	CTTCTCCTTTGCTACGAAACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
oSONG146	GTTTCGTAGCAAAGGAGAAGACTAGTCTTTTCTCTATCACTGATAGGGA

Table S3

Table S3. Essential genes investigated in this study

Essential gene tagged with PROTi tag	Tagged by Cameron et al.	Number of tagged clones	Variants used for droptest	Variants used for growth profile (Biolector)
glmS	No	8	*glmS*-rt1, rt2, rt3, rt4	None (sick phenotype)
ileS	No	12	*ileS*-rt1, rt2, rt3, rt4, rt5, rt6, rt7	*ileS*-rt1, *ileS*-rt2
murE	No	15	*murE*-rt1, rt2, rt3, rt4, rt5, rt6, rt7	*murE*-rt1, *murE*-rt2
pheS	No	18	*pheS*-rt1, rt2, rt3, rt4, rt5, rt6, rt7	*pheS*-rt1, *pheS*-rt2
rnpA	No	3	*rnpA*-rt1, rt2, rt3	*rpnA*-rt1, *rnpA*-rt2
ribD	No	7	*ribD*-rt1, rt2, rt3, rt4, rt5, rt6	*ribD*-rt1, *ribD*-rt2
tmk	No	1	*tmk*-rt	*tmk*-rt
accD	No	0		
fabG	No	0		
prfB	No	0		
ftsZ	yes	0		
acpS	yes	1	*acpS*-rt	*acpS*-rt
ispH	yes	0		
murA	yes	0		
dapE	yes	1	*dapE*-rt	*dapE*-rt
lpxC	yes	4	*lpxC*-rt1, rt2, rt3, rt4	*lpxC*-rt1, lpxC-rt2
ribE	Yes	0		

Nucleotide sequences of pPROTi, pCRiPi and pgRNA-CRiPi in Genbank format

pPROTi

```
LOCUS       Exported                5622 bp ds-DNA     circular
SYN 27-JUN-2017
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 5622)
  AUTHORS   towo
  TITLE     Direct Submission
  JOURNAL   Exported Thursday, Jul 6, 2017 from SnapGene 3.3.4
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..5622
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     terminator      2..96
                     /label=lambda t0 terminator
                     /note="transcription terminator from phage
lambda"
     promoter        144..246
                     /label=cat promoter
                     /note="promoter of the E. coli cat gene"
     CDS             247..906
                     /codon_start=1
                     /gene="cat"
                     /product="chloramphenicol acetyltransferase"
                     /label=CmR
                     /note="confers resistance to
chloramphenicol"

/translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAFL

KTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFSS

LWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNVANM

DNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGGA"
     oriT            1055..1163
                     /note="incP origin of transfer"
     CDS             complement(1174..1836)
                     /codon_start=1
                     /product="replication protein for the broad-
host-range
                     plasmid pBBR1 from Bordetella
bronchiseptica"
                     /label=pBBR1 Rep

/translation="MATQSREIGIQAKNKPGHWVQTERKAHEAWAGLIARKPTAAMLLH

HLVAQMGHQNAVVVSQKTLSKLIGRSLRTVQYAVKDLVAERWISVVKLNGPGTVSAYVV

NDRVAWGQPRDQLRLSVFSAAVVVDHDDQDESLLGHGDLRRIPTLYPGEQQLPTGPGEE
```

```
PPSQPGIPGMEPDLPALTETEEWERRGQQRLPMPDEPCFLDDGEPLEPPTRVTLPRR"
     rep_origin      1837..2608
                     /label=pBBR1 oriV
                     /note="replication origin of the broad-host-
range plasmid
                     pBBR1 from Bordetella bronchiseptica;
requires the pBBR1
                     Rep protein for replication"
     terminator      2723..2809
                     /gene="Escherichia coli rrnB"
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the
E. coli rrnB
                     gene"
     misc_feature    complement(2857..3792)
                     /label=rhaR
     CDS             complement(3779..4612)
                     /codon_start=1
                     /label=rhaS

/translation="MTVLHSVDFFPSGNASVAIEPRLPQADFPEHHHDFHEIVIVEHGT

GIHVFNGQPYTITGGTVCFVRDHDRHLYEHTDNLCLTNVLYRSPDRFQFLAGLNQLLPQ

ELDGQYPSHWRVNHSVLQQVRQLVAQMEQQEGENDLPSTASREILFMQLLLLLRKSSLQ

ENLENSASRLNLLLAWLEDHFADEVNWDAVADQFSLSLRTLHRQLKQQTGLTPQRYLNR

LRLMKARHLLRHSEASVTDIAYRCGFSDSNHFSTLFRREFNWSPRDIRQGRDGFLQ"
     misc_feature    4816..4891
                     /label=PrhaBAD
     misc_feature    4886..4892
                     /label=SDS
     CDS             4900..5613
                     /codon_start=1
                     /label=TEV protease S219D

/translation="MGESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITN

KHLFRRNNGTLLVQSLHGVFKVKNTTTLQQHLIDGRDMIIIRMPKDFPPFPQKLKFREP

QREERICLVTTNFQTKSMSSMVSDTSCTFPSSDGIFWKHWIQTKDGQCGSPLVSTRDGF

IVGIHSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSGWRLNADSVLWGGHKVFMDKP
                     EEPFQPVKEATQLMN"
ORIGIN
        1 ggactcctgt tgatagatcc agtaatgacc tcagaactcc atctggattt
gttcagaacg
       61 ctcggttgcc gccgggcgtt ttttattggt gagaatccag gggtccccaa
taattacgat
      121 ttaaattggc gaaaatgaga cgttgatcgg cacgtaagag gttccaactt
tcaccataat
      181 gaaataagat cactaccggg cgtatttttt gagttatcga gattttcagg
agctaaggaa
      241 gctaaaatgg agaaaaaaat cactggatat accaccgttg atatatccca
atggcatcgt
      301 aaagaacatt ttgaggcatt tcagtcagtt gctcaatgta cctataacca
gaccgttcag
```

361 ctggatatta cggccttttt aaagaccgta aagaaaaata agcacaagtt
ttatccggcc
421 tttattcaca ttcttgcccg cctgatgaat gctcatccgg aatttcgtat
ggcaatgaaa
481 gacggtgagc tggtgatatg ggatagtgtt cacccttgtt acaccgtttt
ccatgagcaa
541 actgaaacgt tttcatcgct ctggagtgaa taccacgacg atttccggca
gtttctacac
601 atatattcgc aagatgtggc gtgttacggt gaaaacctgg cctatttccc
taaagggttt
661 attgagaata tgtttttcgt ctcagccaat ccctgggtga gtttcaccag
ttttgattta
721 aacgtggcca atatggacaa cttcttcgcc cccgttttca ccatgggcaa
atattatacg
781 caaggcgaca aggtgctgat gccgctggcg attcaggttc atcatgccgt
ttgtgatggc
841 ttccatgtcg gcagaatgct taatgaatta caacagtact gcgatgagtg
gcagggcggg
901 gcgtaatttg acttttgtcc ttttccgctg cataaccctg cttcggggtc
attatagcga
961 ttttttcggt atatccatcc tttttcgcac gatatacagg attttgccaa
agggttcgtg
1021 tagactttcc ttggtgtatc caacggcgtc agccgggcag gataggtgaa
gtaggcccac
1081 ccgcgagcgg gtgttccttc ttcactgtcc cttattcgca cctggcggtg
ctcaacggga
1141 atcctgctct gcgaggctgg ccgtaggccg gccctaccgg cgcggcagcg
ttacccgtgt
1201 cggcggctcc aacggctcgc catcgtccag aaaacacggc tcatcgggca
tcggcaggcg
1261 ctgctgcccg cgccgttccc attcctccgt ttcggtcaag gctggcaggt
ctggttccat
1321 gcccggaatg ccgggctggc tgggcggctc ctcgccgggg ccggtcggta
gttgctgctc
1381 gcccggatac agggtcggga tgcggcgcag gtcgccatgc cccaacagcg
attcgtcctg
1441 gtcgtcgtga tcaaccacca cggcggcact gaacaccgac aggcgcaact
ggtcgcgggg
1501 ctggccccac gccacgcggt cattgaccac gtaggccgac acggtgccgg
ggccgttgag
1561 cttcacgacg gagatccagc gctcggccac caagtccttg actgcgtatt
ggaccgtccg
1621 caaagaacgt ccgatgagct tggaaagtgt cttctggctg accaccacgg
cgttctggtg
1681 gcccatctgc gccacgaggt gatgcagcag cattgccgcc gtgggtttcc
tcgcaataag
1741 cccggcccac gcctcatgcg ctttgcgttc cgtttgcacc cagtgaccgg
gcttgttctt
1801 ggcttgaatg ccgatttctc tggactgcgt ggccatgctt atctccatgc
ggtaggggtg
1861 ccgcacggtt gcggcaccat gcgcaatcag ctgcaacttt tcggcagcgc
gacaacaatt
1921 atgcgttgcg taaaagtggc agtcaattac agattttctt taacctacgc
aatgagctat
1981 tgcggggggt gccgcaatga gctgttgcgt accccccttt tttaagttgt
tgatttttaa
2041 gtctttcgca tttcgcccta tatctagttc tttggtgccc aaagaagggc
acccctgcgg
2101 ggttccccca cgccttcggc gcggctcccc ctccggcaaa aagtggcccc
tccggggctt

2161 gttgatcgac tgcgcggcct tcggccttgc ccaaggtggc gctgcccccc
tggaaccccc
2221 gcactcgccg ccgtgaggct cgggggggcag gcgggcgggc ttcgcccttc
gactgccccc
2281 actcgcatag gcttgggtcg ttccaggcgc gtcaaggcca agccgctgcg
cggtcgctgc
2341 gcgagccttg acccgccttc cacttggtgt ccaaccggca agcgaagcgc
gcaggccgca
2401 ggccggaggc ttttccccag agaaaattaa aaaaattgat ggggcaaggc
cgcaggccgc
2461 gcagttggag ccggtgggta tgtggtcgaa ggctgggtag ccggtgggca
atccctgtgg
2521 tcaagctcgt gggcaggcgc agcctgtcca tcagcttgtc cagcagggtt
gtccacgggc
2581 cgagcgaagc gagccagccg gtggccgctc gcggccatcg tccacatatc
cacgggctgg
2641 caagggagcg cagcgaccgc gcagggcgaa gcccggagag caagcccgta
ggggggggcgc
2701 gcccagctgt ctagggcggc ggatttgtcc tactcaggag agcgttcacc
gacaaacaac
2761 agataaaacg aaaggcccag tctttcgact gagcctttcg ttttatttga
tgcctttaat
2821 taataaaaat aggcgtatca cgaggccctt tcgtctatct ttctgcgaat
tgagatgacg
2881 ccactggctg ggcgtcatcc cggtttcccg ggtaaacacc accgaaaaat
agttactatc
2941 ttcaaagcca cattcggtcg aaatatcact gattaacagg cggctatgct
ggagaagata
3001 ttgcgcatga cacactctga cctgtcgcag atattgattg atggtcattc
cagtctgctg
3061 gcgaaattgc tgacgcaaaa cgcgctcact gcacgatgcc tcatcacaaa
atttatccag
3121 cgcaaaggga cttttcaggc tagccgccag ccgggtaatc agcttatcca
gcaacgtttc
3181 gctggatgtt ggcggcaacg aatcactggt gtaacgatgg cgattcagca
acatcaccaa
3241 ctgcccgaac agcaactcag ccatttcgtt agcaaacggc acatgctgac
tactttcatg
3301 ctcaagctga ccaataacct gccgcgcctg cgccatcccc atgctaccta
agcgccagtg
3361 tggttgccct gcgctggcgt taaatcccgg aatcgccccc tgccagtcaa
gattcagctt
3421 cagacgctcc gggcaataaa taatattctg caaaaccaga tcgttaacgg
aagcgtagga
3481 gtgtttatca tcagcatgaa tgtaaaagag atcgccacgg gtaatgcgat
aagggcgatc
3541 gttgagtaca tgcaggccat taccgcgcca gacaatcacc agctcacaaa
aatcatgtgt
3601 atgttcagca aagacatctt gcggataacg gtcagccaca gcgactgcct
gctggtcgct
3661 ggcaaaaaaa tcatctttga gaagttttaa ctgatgcgcc accgtggcta
cctcggccag
3721 agaacgaagt tgattattcg caatatggcg tacaaatacg ttgagaagat
tcgcgttatt
3781 gcagaaagcc atcccgtccc tggcgaatat cacgcggtga ccagttaaac
tctcggcgaa
3841 aaagcgtcga aaagtggtta ctgtcgctga atccacagcg ataggcgatg
tcagtaacgc
3901 tggcctcgct gtggcgtagc agatgtcggg ctttcatcag tcgcaggcgg
ttcaggtatc

```
     3961 gctgaggcgt cagtcccgtt tgctgcttaa gctgccgatg tagcgtacgc
agtgaaagag
     4021 aaaattgatc cgccacggca tcccaattca cctcatcggc aaaatggtcc
tccagccagg
     4081 ccagaagcaa gttgagacgt gatgcgctgt tttccaggtt ctcctgcaaa
ctgcttttac
     4141 gcagcaagag cagtaattgc ataaacaaga tctcgcgact ggcggtcgag
ggtaaatcat
     4201 tttccccttc ctgctgttcc atctgtgcaa ccagctgtcg cacctgctgc
aatacgctgt
     4261 ggttaacgcg ccagtgagac ggatactgcc catccagctc ttgtggcagc
aactgattca
     4321 gcccggcgag aaactgaaat cgatccggcg agcgatacag cacattggtc
agacacagat
     4381 tatcggtatg ttcatacaga tgccgatcat gatcgcgtac gaaacagacc
gtgccaccgg
     4441 tgatggtata gggctgccca ttaaacacat gaatacccgt gccatgttcg
acaatcacaa
     4501 tttcatgaaa atcatgatga tgttcaggaa aatccgcctg cgggagccgg
ggttctatcg
     4561 ccacggacgc gttaccagac ggaaaaaaat ccacactatg taatacggtc
atactggcct
     4621 cctgatgtcg tcaacacggc gaaatagtaa tcacgaggtc aggttcttac
cttaaatttt
     4681 cgacggaaaa ccacgtaaaa aacgtcgatt tttcaagata cagcgtgaat
tttcaggaaa
     4741 tgcggtgagc atcacatcac cacaattcag caaattgtga acatcatcac
gttcatcttt
     4801 ccctggttgc caatggccca ttttcttgtc agtaacgaga aggtcgcgaa
tccaggcgct
     4861 ttttagactg gtcgtaatga aattcaggag gaatatacca tgggagaaag
cttgtttaag
     4921 gggccgcgtg attacaaccc gatatcgagc accatttgtc atttgacgaa
tgaatctgat
     4981 gggcacacaa catcgttgta tggtattgga tttgggccct tcatcattac
aaacaagcac
     5041 ttgtttagaa gaaataatgg aacactgttg gtccaatcac tacatggtgt
attcaaggtc
     5101 aagaacacca cgactttgca acaacacctc atcgatgggc gggacatgat
cattattcga
     5161 atgcctaagg atttcccacc atttcctcaa aagctgaaat ttcgcgagcc
acaacgggaa
     5221 gagcgcattt gtcttgtgac aaccaacttc caaactaaga gcatgtctag
catggtgtca
     5281 gacactagtt gcacattccc ttcatctgat ggcatattct ggaagcattg
gattcaaacc
     5341 aaggatgggc agtgtggcag tccattagta tcaactcgag atgggttcat
tgttggtata
     5401 cactcagcat cgaatttcac caacacaaac aattatttca caagcgtgcc
gaaaaacttc
     5461 atggaattgt tgacaaatca ggaggcgcag cagtgggtta gtggttggcg
attaaatgct
     5521 gactcagtat tgtgggggggg ccataaagtt ttcatggaca aacctgaaga
gccttttcag
     5581 ccagttaagg aagcgactca actcatgaat taaactagtc tt
//
```

pCRiPi

```
LOCUS       Exported                8726 bp ds-DNA     circular
SYN 17-MAR-2017
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    pCRiPi
SOURCE      synthetic DNA construct
  ORGANISM  recombinant plasmid
REFERENCE   1  (bases 1 to 8726)
  AUTHORS   Morten Norholm Group
  TITLE     Direct Submission
  JOURNAL   Exported Thursday, Jul 6, 2017 from SnapGene 3.3.4
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..8726
                     /organism="recombinant plasmid"
                     /mol_type="other DNA"
     misc_feature    complement(7..633)
                     /label=TetR
     misc_feature    634..706
                     /label=tet promoter region
     misc_feature    685..703
                     /label=Tetracycline Response Element (TRE)
     RBS             722..733
                     /label=RBS (strong)
     misc_feature    743..4843
                     /label=Streptococcus pyogenes dCas9
     misc_feature    3257..3258
                     /label=CA in Cas9 wt (His)
     terminator      4862..4990
                     /label=dblTerm
     misc_feature    5057..5132
                     /label=PrhaBAD
     misc_feature    5127..5133
                     /label=SDS
     CDS             5141..5854
                     /codon_start=1
                     /label=TEV protease S219D

/translation="MGESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITN

KHLFRRNNGTLLVQSLHGVFKVKNTTTLQQHLIDGRDMIIIRMPKDFPPFPQKLKFREP

QREERICLVTTNFQTKSMSSMVSDTSCTFPSSDGIFWKHWIQTKDGQCGSPLVSTRDGF

IVGIHSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSGWRLNADSVLWGGHKVFMDKP
                     EEPFQPVKEATQLMN"
     terminator      5865..5959
                     /label=lambda t0 terminator
                     /note="transcription terminator from phage
lambda"
     promoter        6007..6109
                     /label=cat promoter
                     /note="promoter of the E. coli cat gene"
     CDS             6110..6769
                     /codon_start=1
                     /gene="cat"
                     /product="chloramphenicol acetyltransferase"
```

```
                     /label=CmR
                     /note="confers resistance to
chloramphenicol"

/translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAFL

KTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETFSS

LWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNVANM

DNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGGA"
     oriT            6918..7026
                     /note="incP origin of transfer"
     CDS             complement(7037..7699)
                     /codon_start=1
                     /product="replication protein for the broad-
host-range
                     plasmid pBBR1 from Bordetella
bronchiseptica"
                     /label=pBBR1 Rep

/translation="MATQSREIGIQAKNKPGHWVQTERKAHEAWAGLIARKPTAAMLLH

HLVAQMGHQNAVVVSQKTLSKLIGRSLRTVQYAVKDLVAERWISVVKLNGPGTVSAYVV

NDRVAWGQPRDQLRLSVFSAAVVVDHDDQDESLLGHGDLRRIPTLYPGEQQLPTGPGEE

PPSQPGIPGMEPDLPALTETEEWERRGQQRLPMPDEPCFLDDGEPLEPPTRVTLPRR"
     rep_origin      7700..8471
                     /label=pBBR1 oriV
                     /note="replication origin of the broad-host-
range plasmid
                     pBBR1 from Bordetella bronchiseptica;
requires the pBBR1
                     Rep protein for replication"
     terminator      8586..8672
                     /gene="Escherichia coli rrnB"
                     /label=rrnB T1 terminator
                     /note="transcription terminator T1 from the
E. coli rrnB
                     gene"
ORIGIN
        1 gacgtcttaa gacccacttt cacatttaag ttgtttttct aatccgcata
tgatcaattc
       61 aaggccgaat aagaaggctg gctctgcacc ttggtgatca aataattcga
tagcttgtcg
      121 taataatggc ggcatactat cagtagtagg tgtttccctt tcttctttag
cgacttgatg
      181 ctcttgatct tccaatacgc aacctaaagt aaaatgcccc acagcgctga
gtgcatataa
      241 tgcattctct agtgaaaaac cttgttggca taaaaaggct aattgatttt
cgagagtttc
      301 atactgtttt tctgtaggcc gtgtacctaa atgtactttt gctccatcgc
gatgacttag
      361 taaagcacat ctaaaacttt tagcgttatt acgtaaaaaa tcttgccagc
tttccccttc
      421 taaagggcaa aagtgagtat ggtgcctatc taacatctca atggctaagg
cgtcgagcaa
      481 agcccgctta ttttttacat gccaatacaa tgtaggctgc tctacaccta
gcttctgggc
```

541 gagtttacgg gttgttaaac cttcgattcc gacctcatta agcagctcta atgcgctgtt
601 aatcacttta cttttatcta atctagacat cattaattcc taatttttgt tgacactcta
661 tcgttgatag agttatttta ccactcccta tcagtgatag agaaaagaat tcaaaagatc
721 taaagaggag aaaggatcta tggataagaa atactcaata ggcttagcta tcggcacaaa
781 tagcgtcgga tgggcggtga tcactgatga atataaggtt ccgtctaaaa agttcaaggt
841 tctgggaaat acagaccgcc acagtatcaa aaaaaatctt ataggggctc ttttatttga
901 cagtggagag acagcggaag cgactcgtct caaacggaca gctcgtagaa ggtatacacg
961 tcggaagaat cgtatttgtt atctacagga gatttttca aatgagatgg cgaaagtaga
1021 tgatagtttc tttcatcgac ttgaagagtc ttttttggtg gaagaagaca agaagcatga
1081 acgtcatcct atttttggaa atatagtaga tgaagttgct tatcatgaga aatatccaac
1141 tatctatcat ctgcgaaaaa aattggtaga ttctactgat aaagcggatt tgcgcttaat
1201 ctatttggcc ttagcgcata tgattaagtt tcgtggtcat tttttgattg agggagattt
1261 aaatcctgat aatagtgatg tggacaaact atttatccag ttggtacaaa cctacaatca
1321 attatttgaa gaaaacccta ttaacgcaag tggagtagat gctaaagcga ttctttctgc
1381 acgattgagt aaatcaagac gattagaaaa tctcattgct cagctccccg gtgagaagaa
1441 aaatggctta tttgggaatc tcattgcttt gtcattgggt ttgaccccta attttaaatc
1501 aaattttgat ttggcagaag atgctaaatt acagctttca aaagatactt acgatgatga
1561 tttagataat ttattggcgc aaattggaga tcaatatgct gatttgtttt tggcagctaa
1621 gaatttatca gatgctattt tactttcaga tatcctaaga gtaaatactg aaataactaa
1681 ggctccccta tcagcttcaa tgattaaacg ctacgatgaa catcatcaag acttgactct
1741 tttaaaagct ttagttcgac aacaacttcc agaaaagtat aaagaaatct tttttgatca
1801 atcaaaaaac ggatatgcag gttatattga tgggggagct agccaagaag aattttataa
1861 atttatcaaa ccaattttag aaaaaatgga tggtactgag gaattattgg tgaaactaaa
1921 tcgtgaagat ttgctgcgca agcaacggac ctttgacaac ggctctattc cccatcaaat
1981 tcacttgggt gagctgcatg ctattttgag aagacaagaa gacttttatc catttttaaa
2041 agacaatcgt gagaagattg aaaaaatctt gacttttcga attccttatt atgttggtcc
2101 attggcgcgt ggcaatagtc gttttgcatg gatgactcgg aagtctgaag aaacaattac
2161 cccatggaat tttgaagaag ttgtcgataa aggtgcttca gctcaatcat ttattgaacg
2221 catgacaaac tttgataaaa atcttccaaa tgaaaaagta ctaccaaaac atagtttgct
2281 ttatgagtat tttacggttt ataacgaatt gacaaaggtc aaatatgtta ctgaaggaat

2341 gcgaaaacca gcatttcttt caggtgaaca gaagaaagcc attgttgatt tactcttcaa
2401 aacaaatcga aaagtaaccg ttaagcaatt aaaagaagat tatttcaaaa aaatagaatg
2461 ttttgatagt gttgaaattt caggagttga agatagattt aatgcttcat taggtaccta
2521 ccatgatttg ctaaaaatta ttaaagataa agatttttg gataatgaag aaaatgaaga
2581 tatcttagag gatattgttt taacattgac cttatttgaa gatagggaga tgattgagga
2641 aagacttaaa acatatgctc acctctttga tgataaggtg atgaaacagc ttaaacgtcg
2701 ccgttatact ggttggggac gtttgtctcg aaaattgatt aatggtatta gggataagca
2761 atctggcaaa acaatattag atttttgaa atcagatggt tttgccaatc gcaattttat
2821 gcagctgatc catgatgata gtttgacatt taaagaagac attcaaaaag cacaagtgtc
2881 tggacaaggc gatagtttac atgaacatat tgcaaattta gctggtagcc ctgctattaa
2941 aaaaggtatt ttacagactg taaaagttgt tgatgaattg gtcaaagtaa tggggcggca
3001 taagccagaa aatatcgtta ttgaaatggc acgtgaaaat cagacaactc aaaagggcca
3061 gaaaaattcg cgagagcgta tgaaacgaat cgaagaaggt atcaaagaat taggaagtca
3121 gattcttaaa gagcatcctg ttgaaaatac tcaattgcaa aatgaaaagc tctatctcta
3181 ttatctccaa aatggaagag acatgtatgt ggaccaagaa ttagatatta atcgtttaag
3241 tgattatgat gtcgatgcca ttgttccaca aagtttcctt aaagacgatt caatagacaa
3301 taaggtctta acgcgttctg ataaaaatcg tggtaaatcg gataacgttc caagtgaaga
3361 agtagtcaaa aagatgaaaa actattggag acaacttcta aacgccaagt taatcactca
3421 acgtaagttt gataatttaa cgaaagctga acgtggaggt ttgagtgaac ttgataaagc
3481 tggttttatc aaacgccaat tggttgaaac tcgccaaatc actaagcatg tggcacaaat
3541 tttggatagt cgcatgaata ctaaatacga tgaaaatgat aaacttattc gagaggttaa
3601 agtgattacc ttaaaatcta aattagtttc tgacttccga aaagatttcc aattctataa
3661 agtacgtgag attaacaatt accatcatgc ccatgatgcg tatctaaatg ccgtcgttgg
3721 aactgctttg attaagaaat atccaaaact tgaatcggag tttgtctatg gtgattataa
3781 agtttatgat gttcgtaaaa tgattgctaa gtctgagcaa gaaataggca aagcaaccgc
3841 aaaatatttc ttttactcta atatcatgaa cttcttcaaa acagaaatta cacttgcaaa
3901 tggagagatt cgcaaacgcc ctctaatcga aactaatggg gaaactggag aaattgtctg
3961 ggataaaggg cgagattttg ccacagtgcg caaagtattg tccatgcccc aagtcaatat
4021 tgtcaagaaa acagaagtac agacaggcgg attctccaag gagtcaattt taccaaaaag
4081 aaattcggac aagcttattg ctcgtaaaaa agactgggat ccaaaaaaat atggtggttt

4141 tgatagtcca acggtagctt attcagtcct agtggttgct aaggtggaaa aagggaaatc
4201 gaagaagtta aaatccgtta aagagttact agggatcaca attatggaaa gaagttcctt
4261 tgaaaaaaat ccgattgact ttttagaagc taaaggatat aaggaagtta aaaaagactt
4321 aatcattaaa ctacctaaat atagtctttt tgagttagaa aacggtcgta aacggatgct
4381 ggctagtgcc ggagaattac aaaaaggaaa tgagctggct ctgccaagca aatatgtgaa
4441 ttttttatat ttagctagtc attatgaaaa gttgaagggt agtccagaag ataacgaaca
4501 aaaacaattg tttgtggagc agcataagca ttatttagat gagattattg agcaaatcag
4561 tgaattttct aagcgtgtta ttttagcaga tgccaattta gataaagttc ttagtgcata
4621 taacaaacat agagacaaac caatacgtga acaagcagaa aatattattc atttatttac
4681 gttgacgaat cttggagctc ccgctgcttt taaatatttt gatacaacaa ttgatcgtaa
4741 acgatatacg tctacaaaag aagttttaga tgccactctt atccatcaat ccatcactgg
4801 tctttatgaa acacgcattg atttgagtca gctaggaggt gactaactcg agtaaggatc
4861 tccaggcatc aaataaaacg aaaggctcag tcgaaagact gggcctttcg ttttatctgt
4921 tgtttgtcgg tgaacgctct ctactagagt cacactggct caccttcggg tgggcctttc
4981 tgcgtttata aagaaattca ccacaattca gcaaattgtg aacatcatca cgttcatctt
5041 tccctggttg ccaatggccc attttcttgt cagtaacgag aaggtcgcga atccaggcgc
5101 tttttagact ggtcgtaatg aaattcagga ggaatatacc atgggagaaa gcttgtttaa
5161 ggggccgcgt gattacaacc cgatatcgag caccatttgt catttgacga atgaatctga
5221 tgggcacaca acatcgttgt atggtattgg atttgggccc ttcatcatta caaacaagca
5281 cttgtttaga agaaataatg gaacactgtt ggtccaatca ctacatggtg tattcaaggt
5341 caagaacacc acgactttgc aacaacacct catcgatggg cgggacatga tcattattcg
5401 aatgcctaag gatttcccac catttcctca aaagctgaaa tttcgcgagc cacaacggga
5461 agagcgcatt tgtcttgtga caaccaactt ccaaactaag agcatgtcta gcatggtgtc
5521 agacactagt tgcacattcc cttcatctga tggcatattc tggaagcatt ggattcaaac
5581 caaggatggg cagtgtggca gtccattagt atcaactcga gatgggttca ttgttggtat
5641 acactcagca tcgaatttca ccaacacaaa caattatttc acaagcgtgc cgaaaaactt
5701 catggaattg ttgacaaatc aggaggcgca gcagtgggtt agtggttggc gattaaatgc
5761 tgactcagta ttgtgggggg gccataaagt tttcatggac aaacctgaag agccttttca
5821 gccagttaag gaagcgactc aactcatgaa ttaaactagt cttggactcc tgttgataga
5881 tccagtaatg acctcagaac tccatctgga tttgttcaga acgctcggtt gccgccgggc

20

5941 gtttttttatt ggtgagaatc caggggtccc caataattac gatttaaatt ggcgaaaatg
6001 agacgttgat cggcacgtaa gaggttccaa ctttcaccat aatgaaataa gatcactacc
6061 gggcgtattt tttgagttat cgagattttc aggagctaag gaagctaaaa tggagaaaaa
6121 aatcactgga tataccaccg ttgatatatc ccaatggcat cgtaaagaac attttgaggc
6181 atttcagtca gttgctcaat gtacctataa ccagaccgtt cagctggata ttacggcctt
6241 tttaaagacc gtaaagaaaa ataagcacaa gttttatccg gcctttattc acattcttgc
6301 ccgcctgatg aatgctcatc cggaatttcg tatggcaatg aaagacggtg agctggtgat
6361 atgggatagt gttcaccctt gttacaccgt tttccatgag caaactgaaa cgttttcatc
6421 gctctggagt gaataccacg acgatttccg gcagtttcta cacatatatt cgcaagatgt
6481 ggcgtgttac ggtgaaaacc tggcctattt ccctaaaggg tttattgaga atatgttttt
6541 cgtctcagcc aatccctggg tgagtttcac cagttttgat ttaaacgtgg ccaatatgga
6601 caacttcttc gcccccgttt tcaccatggg caaatattat acgcaaggcg acaaggtgct
6661 gatgccgctg gcgattcagg ttcatcatgc cgtttgtgat ggcttccatg tcggcagaat
6721 gcttaatgaa ttacaacagt actgcgatga gtggcagggc ggggcgtaat ttgacttttg
6781 tccttttccg ctgcataacc ctgcttcggg gtcattatag cgatttttc ggtatatcca
6841 tccttttcg cacgatatac aggattttgc caaagggttc gtgtagactt tccttggtgt
6901 atccaacggc gtcagccggg caggataggt gaagtaggcc cacccgcgag cgggtgttcc
6961 ttcttcactg tcccttattc gcacctggcg gtgctcaacg ggaatcctgc tctgcgaggc
7021 tggccgtagg ccggccctac cggcgcggca gcgttacccg tgtcggcggc tccaacggct
7081 cgccatcgtc cagaaaacac ggctcatcgg gcatcggcag gcgctgctgc ccgcgccgtt
7141 cccattcctc cgtttcggtc aaggctggca ggtctggttc catgcccgga atgccgggct
7201 ggctgggcgg ctcctcgccg gggccggtcg gtagttgctg ctcgcccgga tacagggtcg
7261 ggatgcggcg caggtcgcca tgccccaaca gcgattcgtc ctggtcgtcg tgatcaacca
7321 ccacggcggc actgaacacc gacaggcgca actggtcgcg gggctggccc cacgccacgc
7381 ggtcattgac cacgtaggcc gacacggtgc cggggccgtt gagcttcacg acggagatcc
7441 agcgctcggc caccaagtcc ttgactgcgt attggaccgt ccgcaaagaa cgtccgatga
7501 gcttggaaag tgtcttctgg ctgaccacca cggcgttctg gtggcccatc tgcgccacga
7561 ggtgatgcag cagcattgcc gccgtgggtt tcctcgcaat aagcccggcc cacgcctcat
7621 gcgctttgcg ttccgtttgc acccagtgac cgggcttgtt cttggcttga atgccgattt
7681 ctctggactg cgtggccatg cttatctcca tgcggtaggg gtgccgcacg gttgcggcac

```
     7741 catgcgcaat cagctgcaac ttttcggcag cgcgacaaca attatgcgtt
gcgtaaaagt
     7801 ggcagtcaat tacagatttt ctttaaccta cgcaatgagc tattgcgggg
ggtgccgcaa
     7861 tgagctgttg cgtacccccc ttttttaagt tgttgatttt taagtctttc
gcatttcgcc
     7921 ctatatctag ttctttggtg cccaaagaag ggcacccctg cggggttccc
ccacgccttc
     7981 ggcgcggctc cccctccggc aaaaagtggc ccctccgggg cttgttgatc
gactgcgcgg
     8041 ccttcggcct tgcccaaggt ggcgctgccc ccttggaacc cccgcactcg
ccgccgtgag
     8101 gctcgggggg caggcgggcg ggcttcgccc ttcgactgcc cccactcgca
taggcttggg
     8161 tcgttccagg cgcgtcaagg ccaagccgct gcgcggtcgc tgcgcgagcc
ttgacccgcc
     8221 ttccacttgg tgtccaaccg gcaagcgaag cgcgcaggcc gcaggccgga
ggcttttccc
     8281 cagagaaaat taaaaaaatt gatggggcaa ggccgcaggc cgcgcagttg
gagccggtgg
     8341 gtatgtggtc gaaggctggg tagccggtgg gcaatccctg tggtcaagct
cgtgggcagg
     8401 cgcagcctgt ccatcagctt gtccagcagg gttgtccacg ggccgagcga
agcgagccag
     8461 ccggtggccg ctcgcggcca tcgtccacat atccacgggc tggcaaggga
gcgcagcgac
     8521 cgcgcagggc gaagcccgga gagcaagccc gtaggggggg cgcgcccagc
tgtctagggc
     8581 ggcggatttg tcctactcag gagagcgttc accgacaaac aacagataaa
acgaaaggcc
     8641 cagtctttcg actgagcctt tcgttttatt tgatgccttt aattaataaa
aataggcgta
     8701 tcaacgtctc attttcgcca gatatc
//
```

pgRNA-CRiPi

```
LOCUS       Exported                3177 bp ds-DNA     circular
SYN 05-JAN-2017
DEFINITION  synthetic circular DNA
ACCESSION   .
VERSION     .
KEYWORDS    pgRNA-CRiPi : pSLQ1236-Tev for CRiPi - aTc inducible
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 3177)
  AUTHORS   .
  TITLE     Direct Submission
  JOURNAL   Exported Thursday, Jul 6, 2017 from SnapGene 3.3.4
            http://www.snapgene.com
FEATURES             Location/Qualifiers
     source          1..3177
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
     misc_feature    17..643
                     /label=TetR
     misc_feature    258..263
                     /label=XbaI-SpeI scar(1)
     misc_feature    644..716
                     /label=tet promoter region
     misc_feature    665..682
                     /label=Q1, Q2 promoter (tetR)
     misc_feature    695..713
                     /label=TetO2 site
     misc_feature    724..743
                     /label=TEV gRNA
     misc_feature    744..785
                     /label=JAD designed chimera RNA 3'UTR
     misc_feature    752..757
                     /label=NheI-XbaI scar
     misc_feature    777..782
                     /label=NheI-SpeI scar
     misc_feature    790..826
                     /label=Streptococcus pyogenes tracrRNA
terminator
     terminator      827..1194
                     /label=TrrnB
     misc_feature    complement(863..868)
                     /label=BglII-BamHI scar(3)
     misc_feature    863..868
                     /label=BamHI-BglII scar
     misc_feature    914..919
                     /label=BsaI site
     misc_feature    1092..1097
                     /label=BsaI site(1)
     terminator      1120..1187
                     /label=sglTerm
     rep_origin      complement(1380..2062)
                     /direction=LEFT
                     /label=ColE1 origin
     misc_feature    complement(1831..1836)
                     /label=XbaI-SpeI scar
     misc_feature    complement(1978..1983)
                     /label=BglII-BamHI scar(4)
     misc_feature    1978..1983
```

```
                     /label=BamHI-BglII scar(1)
     misc_feature    1989..1994
                     /label=BglII-BamHI scar
     misc_feature    complement(1989..1994)
                     /label=BamHI-BglII scar(4)
     misc_feature    complement(2075..2080)
                     /label=BglII-BamHI scar(5)
     misc_feature    2075..2080
                     /label=BamHI-BglII scar(2)
     misc_feature    2087..2092
                     /label=BglII-BamHI scar(1)
     misc_feature    complement(2087..2092)
                     /label=BamHI-BglII scar(5)
     CDS             complement(2160..2819)
                     /label=AmpR
     misc_feature    complement(2297..2302)
                     /label=BsaI site(2)
     misc_feature    2419..2424
                     /label=NheI-XbaI scar(1)
     misc_feature    complement(2855..2860)
                     /label=BglII-BamHI scar(6)
     misc_feature    2855..2860
                     /label=BamHI-BglII scar(3)
     misc_feature    2872..2877
                     /label=BglII-BamHI scar(2)
     misc_feature    complement(2872..2877)
                     /label=BamHI-BglII scar(6)
     misc_feature    complement(3059..3087)
                     /label=AmpR promoter
ORIGIN
        1 gaattcttca agatctttaa gacccacttt cacatttaag ttgtttttct
aatccgcata
       61 tgatcaattc aaggccgaat aagaaggctg gctctgcacc ttggtgatca
aataattcga
      121 tagcttgtcg taataatggc ggcatactat cagtagtagg tgtttccctt
tcttctttag
      181 cgacttgatg ctcttgatct tccaatacgc aacctaaagt aaaatgcccc
acagcgctga
      241 gtgcatataa tgcattctct agtgaaaaac cttgttggca taaaaaggct
aattgatttt
      301 cgagagtttc atactgtttt tctgtaggcc gtgtacctaa atgtactttt
gctccatcgc
      361 gatgacttag taaagcacat ctaaaacttt tagcgttatt acgtaaaaaa
tcttgccagc
      421 tttccccttc taaagggcaa aagtgagtat ggtgcctatc taacatctca
atggctaagg
      481 cgtcgagcaa agcccgctta ttttttacat gccaatacaa tgtaggctgc
tctacaccta
      541 gcttctgggc gagtttacgg gttgttaaac cttcgattcc gacctcatta
agcagctcta
      601 atgcgctgtt aatcacttta cttttatcta atctagacat cattaattcc
taatttttgt
      661 tgacactcta tcgttgatag agttatttta ccactcccta tcagtgatag
agaaaagact
      721 agtcttctcc tttgctacga aacgttttag agctagaaat agcaagttaa
aataaggcta
      781 gtccgttatc aacttgaaaa agtggcaccg agtcggtgct ttttttgaag
cttgggcccg
      841 aacaaaaact catctcagaa gaggatctga atagcgccgt cgaccatcat
catcatcatc
```

```
      901 attgagttta aacggtctcc agcttggctg ttttggcgga tgagagaaga
ttttcagcct
      961 gatacagatt aaatcagaac gcagaagcgg tctgataaaa cagaatttgc
ctggcggcag
     1021 tagcgcggtg gtcccacctg accccatgcc gaactcagaa gtgaaacgcc
gtagcgccga
     1081 tggtagtgtg gggtctcccc atgcgagagt agggaactgc caggcatcaa
ataaaacgaa
     1141 aggctcagtc gaaagactgg gcctttcgtt ttatctgttg tttgtcggtg
aactggatcc
     1201 ttactcgagt ctagactgca ggcttcctcg ctcactgact cgctgcgctc
ggtcgttcgg
     1261 ctgcggcgag cggtatcagc tcactcaaag gcggtaatac ggttatccac
agaatcaggg
     1321 gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa
ccgtaaaaag
     1381 gccgcgttgc tggcgttttt ccataggctc cgccccctg acgagcatca
caaaaatcga
     1441 cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc
gtttccccct
     1501 ggaagctccc tcgtgcgctc tcctgttccg accctgccgc ttaccggata
cctgtccgcc
     1561 tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta
tctcagttcg
     1621 gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac cccccgttca
gcccgaccgc
     1681 tgcgccttat ccggtaacta tcgtcttgag tccaacccgg taagacacga
cttatcgcca
     1741 ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg
tgctacagag
     1801 ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtatttgg
tatctgcgct
     1861 ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg
caaacaaacc
     1921 accgctggta gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag
aaaaaaagga
     1981 tctcaagaag atcctttgat cttttctacg gggtctgacg ctcagtggaa
cgaaaactca
     2041 cgttaaggga ttttggtcat gagattatca aaaaggatct tcacctagat
ccttttaaat
     2101 taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc
tgacagttac
     2161 caatgcttaa tcagtgaggc acctatctca gcgatctgtc tatttcgttc
atccatagtt
     2221 gcctgactcc ccgtcgtgta gataactacg atacgggagg gcttaccatc
tggccccagt
     2281 gctgcaatga taccgcgaga cccacgctca ccggctccag atttatcagc
aataaaccag
     2341 ccagccggaa gggccgagcg cagaagtggt cctgcaactt tatccgcctc
catccagtct
     2401 attaattgtt gccgggaagc tagagtaagt agttcgccag ttaatagttt
gcgcaacgtt
     2461 gttgccattg ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc
ttcattcagc
     2521 tccggttccc aacgatcaag gcgagttaca tgatccccca tgttgtgcaa
aaaagcggtt
     2581 agctccttcg gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt
atcactcatg
     2641 gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg
cttttctgtg
```

```
     2701 actggtgagt actcaaccaa gtcattctga gaatagtgta tgcggcgacc
gagttgctct
     2761 tgcccggcgt caatacggga taataccgcg ccacatagca gaactttaaa
agtgctcatc
     2821 attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt
gagatccagt
     2881 tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt
caccagcgtt
     2941 tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa agggaataag
ggcgacacgg
     3001 aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta
tcagggttat
     3061 tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaaat
aggggttccg
     3121 cgcacatttc cccgaaaagt gccacctgac gtctaagaaa ccattattat
catgaca
//
```

References

1. Silva-Rocha,R., Martínez-García,E., Calles,B., Chavarría,M., Arce-Rodríguez,A., De Las Heras,A., Páez-Espino,A.D., Durante-Rodríguez,G., Kim,J., Nikel,P.I., *et al.* (2013) The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.*, **41**, 666–675.
2. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–83.
3. Li,S., Jendresen,C.B., Grünberger,A., Ronda,C., Jensen,S.I., Noack,S. and Nielsen,A.T. (2016) Enhanced protein and biochemical production using CRISPRi-based growth switches. *Metab. Eng.*, **38**, 274–284.
4. Ronda,C., Pedersen,L.E., Sommer,M.O.A. and Nielsen,A.T. (2016) CRMAGE: CRISPR Optimized MAGE Recombineering. *Sci. Rep.*, **6**, 19452.
5. Söderström,B., Skoog,K., Blom,H., Weiss,D.S., von Heijne,G. and Daley,D.O. (2014) Disassembly of the divisome in *E scherichia coli* : evidence that FtsZ dissociates before compartmentalization. *Mol. Microbiol.*, **92**, 1–9.
6. Kapust,R.B. and Waugh,D.S. (2000) Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.*, **19**, 312–8.

Paper 2

Bacterial Genome Editing Strategy for Control of Transcription and Protein Stability

Bacterial genome editing strategy for control of transcription and protein stability

Ida Lauritsen[1], Virginia Martínez[1], Carlotta Ronda[2], Alex Toftgaard Nielsen[1] and Morten H. H. Nørholm[1,*]

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Denmark, Kemitorvet B220, 2800 Lyngby Denmark.

[2] Systems and Synthetic Biology, Columbia University Medical Center, 3960 Broadway, New York, NY 10032, USA

*Corresponding author. E-mail: morno@biosustain.dtu.dk

Phone: +45 217-99184; Fax: +45-353-33300

Published in:

M. Jensen, J. Keasling (eds)
Synthetic Metabolic Pathways (2018)
Methods in Molecular Biology 1671, pp. 27-37. Humana Press, New York, NY

Abstract

In molecular biology and cell factory engineering, tools that enable control of protein production and stability are highly important. Here, we describe protocols for tagging genes in *Escherichia coli* allowing for inducible degradation and transcriptional control of any soluble protein of interest. The underlying molecular biology is based on the two cross-kingdom tools CRISPRi and the N-end rule for protein degradation. Genome editing is performed with the CRMAGE technology and randomization of the translational initiation region minimizes the polar effects of tag insertion. The approach has previously been applied for targeting proteins originating from essential operon-located genes and has potential to serve as a universal synthetic biology tool.

Key words: Genome editing, N-end rule pathway, CRISPR interference, CRISPR-Cas9, CRMAGE, essential genes, N-degron, protein stability, PROTi, CRiPi.

1. Introduction

Control over cellular protein levels is a key aspect of basic molecular biology and applied biotechnology. Nucleotide and protein-manipulation technologies are thus extremely important, best exemplified with the recent paradigm shifting CRISPR-Cas9 genome editing. Such technologies are valuable for controlling protein levels of enzymes involved in biosynthetic pathways or to elucidate the physiological functions of essential genes in which genetic knockouts are not an option ***(1)***. Protein abundance can be manipulated by specific peptide degradation signals or "tags" that determine the protein half-life ***(2)***. Several studies and technologies utilize synthetic degradation tags for inducible protein control engineered at the C-terminal end of proteins of interest ***(3–5)***. However, these approaches can be suboptimal for operon-located genes (due to "polar effects" on the expression of downstream genes) and for proteins requiring a free C-terminus for functionality. Alternative degradation signals, named N-degrons, are attached at the N-termini of proteins where stability is mostly dictated by the identity of the first amino acid and a few additional parameters ***(6)***. N-degron residues that lead to protein degradation are called destabilizing and in bacteria, these are divided into two classes: the primary (leucine, phenylalanine and tryptophan) and the secondary (arginine and lysine) ***(7)***. The primary N-degrons directly promotes protein degradation by adapter ClpS recognition, thereby targeting the N-degron to the ClpAP

protease complex. Secondary N-degrons are modified into primary ones by enzymatic addition of the corresponding amino acids ***(8)***. The N-end rule pathway has been identified in bacteria ***(7)***, mammals ***(9)***, yeast ***(6)*** and plants ***(10)***.

Based on the N-end rule pathway, we have recently developed and characterized a bacterial protein degradation technology, the Protein interference system (PROTi). This technology offers conditional protein degradation, involving exposure of an N-degron upon rhamnose-controlled expression of a TEV protease ***(11)***. The N-degron is part of a small degradation tag, the PROTi tag, and the corresponding nucleotide sequence is genomically inserted in the 5'-end of genes of interest. The PROTi tag (54 nucleotides) encodes the TEV recognition site (consensus sequence: ENLYFQ↓X ***(12)***, where X denotes any amino acid except proline), the N-degron phenylalanine (in the X-position of the TEV recognition site sequence) and a linker region, shown to promote degradation by the N-end rule pathway ***(13)*** (**Fig. 1**). With this system, we have been able to deplete e.g. GFP and essential proteins in *Escherichia coli* (*E. coli*) – the latter leading to a significant decrease in viability upon regrowth of colonies on agar plates after PROTi induction ***(11)***. Combining the PROTi system with transcriptional repression by CRISPR interference (CRISPRi), we developed the CRiPi system ***(11)***. The method of CRISPR interference (CRISPRi) enables transcriptional inhibition by use of a dead Cas9 (dCas9), assumed to prevent RNA polymerase binding ***(14)***. By targeting the dCas9 to the DNA encoding the PROTi tag, the CRiPi system enables simultaneous protein degradation and knockdown of transcription of genes with the genomically inserted PROTi tag (**Fig. 1**). With this technology, we successfully knocked down soluble proteins encoded by operon-located essential genes and demonstrated enhanced phenotypes for two essential genes by simultaneous targeting of both transcription and protein stability ***(11)***.

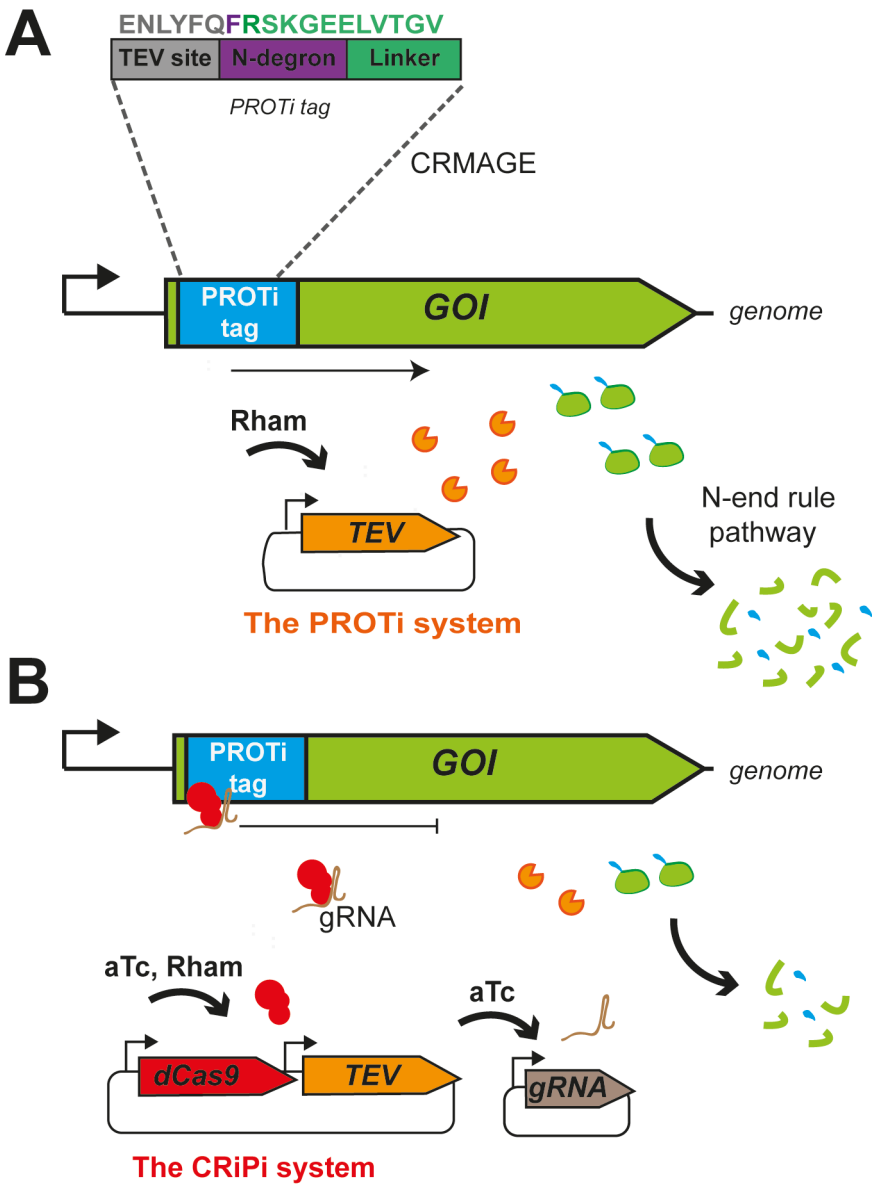


Fig. 1 The PROTi and CRiPi systems. (**A**) The PROTi tag (blue) is genomically integrated using CRMAGE ***(15)***, tagging the gene of interest (GOI) at the 5-end. In the PROTi system, upon rhamnose-inducible expression of the TEV protease (corresponding gene and protein illustrated in orange color) the N-degron becomes exposed leading to protein degradation by N-end rule pathway (gene and protein of interest illustrated in green color). (**B**) Protein depletion can be accelerated by simultaneously implementing the CRISPRi technology. The CRiPi system relies on a combination of controlled protein degradation and repression of transcription. Expression of dCas9 (illustrated in red) can be induced by addition of anhydrotetracycline (aTc). The dCas9 is targeted towards the DNA encoding the PROTi tag by a guide RNA (gRNA).

Genomic insertion of the PROTi tag is performed using CRMAGE genome editing. This technology combines λ-Red recombineering-based multiplex automated genome engineering (MAGE) with selection against the unmodified sequence using the CRISPR-Cas9 system in *E. coli* ***(15)***. The small PROTi tag is encoded in a single-stranded oligonucleotide (ssDNA oligo), compatible with CRMAGE. To apply the CRISRP-Cas9 system for selection, a protospacer adjacent motif (PAM) sequence of the canonical form 5'-NGG-3' is to be removed upon recombineering the ssDNA into the genome ***(15)***. Using this selection method, identification of clones with the genomically inserted PROTi tag is strongly facilitated. Occasionally, the genomic insertion of the PROTi tag can have lethal consequences – e.g. when targeting essential genes located in operons ***(11)***. In that case, insertion of the PROTi tag is facilitated by simultaneous randomization of the translational initiation region (TIR), which has been shown to heavily influence translation rates ***(16–18)***. Specifically, six nucleotides upstream and downstream from the start codon are randomized in our approach and this creates expression libraries that can be screened phenotypically ***(11)***. The approach of TIR randomization has previously enabled tagging of seven operon-located essential genes that were not tagged before, presumably by minimizing polar effects and allowing for identification of clones with near-wild type expression levels ***(11)***. An overview of the workflow for application of the PROTi and CRiPi systems is presented in **Fig. 2.**

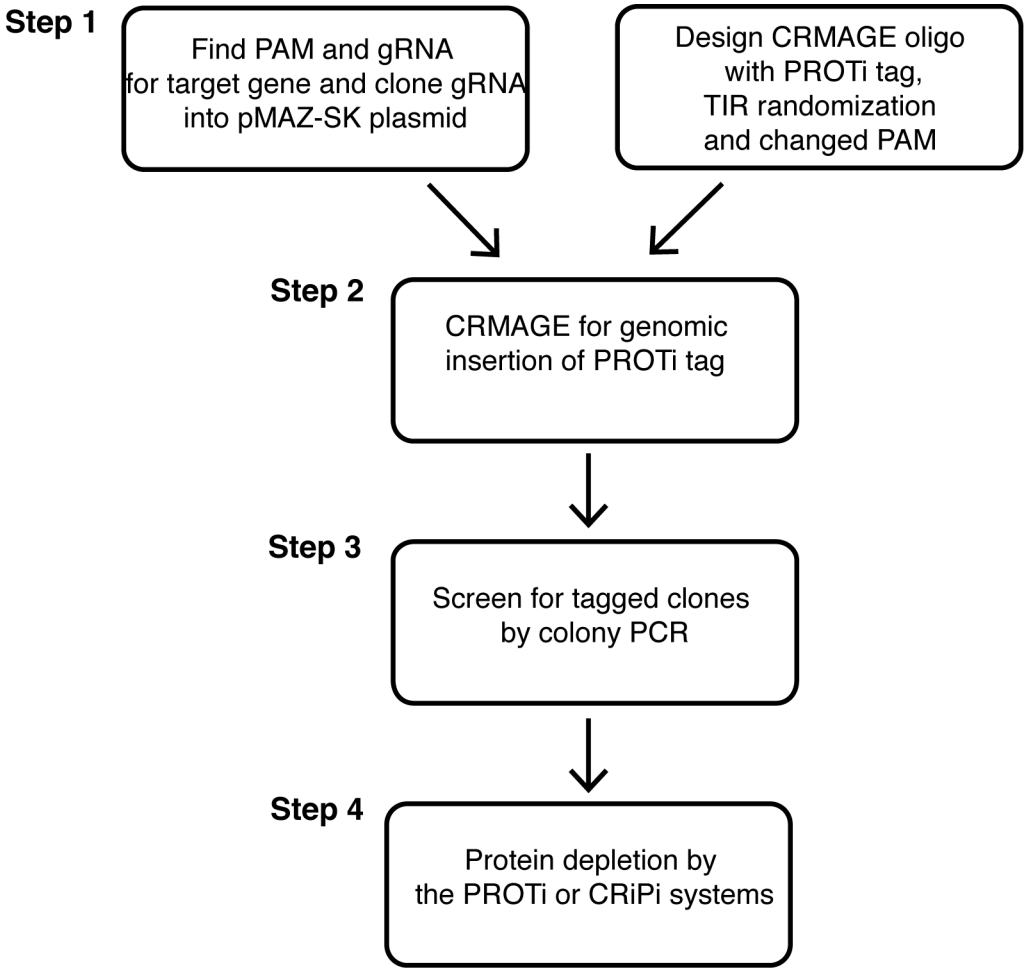


Fig. 2 Workflow for application of the PROTi and CRiPi systems. *Step 1:* Selection of a PAM and gRNA sequence in the coding sequence of the gene of interest. The selected gRNA is cloned into the pMAZ-SK plasmid. Upon expression from this plasmid, the gRNA will guide Cas9 site-specific cleavage, selecting against the unmodified sequence in CRMAGE. Furthermore, a ssDNA CRMAGE oligonucleotide is designed to encode the PROTi tag, a randomized TIR (to minimize effects of tagging) and a changed PAM to circumvent Cas9 cleavage (see Fig. 3 for further details). *Step 2:* The CRMAGE oligonucleotide and the pMAZ-SK-gRNA are used for genomic integration and selection of an inserted PROTi tag in the gene of interest with CRMAGE. *Step 3:* Genomically PROTi-tagged clones are identified by colony PCR upon recovery after CRMAGE. *Step 4:* Inducible depletion of target protein is now possible with the PROTi or CRiPi systems induced with rhamnose and/or anhydrotetracycline.

2. Materials

2.1. CRMAGE medium, CRiPi inducers and buffers

1. LB-lennox: 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl supplemented with 0.5 mM $MgSO_4$ (Sigma-Aldrich).
2. CRiPi inducers: L-rhamnose (0.5 M), L-arabinose (10%) and anhydrotetracycline (200 µg/mL) (Sigma Aldrich).
3. NEBuffer4 (New England Biolabs® Inc.).
4. Antibiotics: Kanamycin (50 mg/mL), chloramphenicol (25 mg/mL) and ampicillin (100 mg/mL).

2.2. Oligonucleotides

1. ssDNA CRMAGE oligos (longer than 100 nucleotides) for PROTi tag insertion
2. Oligos containing the chosen gRNA for the target gene with overhangs that match the pMAZ-SK vector ***(15)***.
3. Screening oligos: For amplification of a PCR product with e.g. the size of 200 bp for verification of inserted PROTi tag.

2.3. Strains, plasmids and polymerases

1. CRMAGE strain: *E. coli* K-12 MG1655 harboring pMA7CR.2.0 and pZS4Int-tetR ***(15)*** (available from Addgene), standard *E. coli* cloning strain (e.g. DH5α).
2. Plasmids: pMAZ-SK for gRNA cloning ***(15)*** (available from Addgene), pPROTi, pCRiPi and pgRNA for CRiPi-induced protein depletion ***(11)***. An overview of plasmids needed is shown in **Table 1.**
3. Polymerase chain reaction (PCR) kit for colony PCR (e.g. OneTaq®), Uracil-Specific Excision Reagent (USER) enzyme for cloning (New England Biolabs® Inc.).

Table 1 – Plasmids for generation of PROTi tagged genes and PROTi and/or CRiPi-induced protein depletion.

Plasmid name	Description	Reference
pMAZ-SK	Expression of target gRNA, aTc-inducible, Km^R	***(15)***
pPROTi	Expression of TEV protease, rhamnose-inducible, Cm^R	***(11)***
pCRiPi	Expression of TEV protease and dCas9, rhamnose and aTc-inducible, Cm^R	***(11)***
pgRNA	Expression of gRNA towards PROTi tag, aTc- inducible, Amp^R	***(11)***

3. Methods

An outline of the workflow is shown in Fig. 2.

3.1. Cloning of target gRNA into pMAZ-SK backbone

1. A PAM sequence in the form of 5'-NGG-3' or 5'-CCN-3' is chosen in the coding sequence, in close proximity to the start codon of the target gene (*see* **Note 1**). The 20 nucleotides next to the PAM sequence are the guide RNA (gRNA). If the PAM is chosen as 5'-NGG-3', the gRNA is upstream of the PAM and if selected as 5'-CCN-3', the gRNA is located downstream. Both the PAM and the gRNA are essential components for Cas9 recognition and cleavage of the target sequence.
2. Overhang sequences matching the pMAZ-SK backbone **(Table 2**) are added to the 20 nucleotides of the chosen gRNA (5'-GAGCAC-N_{20}-GTTTTAGAGCTAGAAAT-3') and the complementary sequence (5'-CTAAAAC-N_{20}-GTGCTCAGTATCTCT-3') according to ***(15)*** and ordered as a forward and a reverse oligo.
3. 10 µl of both the forward and reverse gRNA oligo are mixed at a concentration of 100 µM with 10 µL NEBuffer 4 and 70 µL MiliQ water in a 1.5 mL Eppendorf tube.
4. The sample is incubated at 95°C for 5 min, then cooled to room temperature (20-25°C) for annealing.
5. The pMAZ-SK backbone is amplified by PCR with backbone primers (**Table 2**) and the annealed gRNA is cloned into the pMAZ-SK plasmid by USER cloning as described in ***(19)***. The melting temperature is 21°C for the USER overhangs of the pMAZ-SK backbone. Cloning is performed with 15 min at 37°C, 15 min at 21°C and 10 min. at 10°C.
6. 5 µL of the USER reaction is mixed with 50 µL chemical competent *E. coli* cells (e.g. the DH5α strain) in a 1.5 mL Eppendorf tube, put on ice for 30 min and then heat-shocked at 42°C for 60 seconds.
7. The sample is cooled on ice for 2 min., 500 µL of LB is added and then incubated 1 hour at 37°C with shaking.
8. The sample is plated on LB agar with kanamycin (50 µg/mL) and incubated overnight at 37°C. Positive colonies are screened by PCR followed by standard DNA purification and sequence validation.

Table 2 – Oligos for generation of PROTi tagged genes and PROTi and/or CRiPi-induced protein depletion. The N20 denotes the 20 nucleotides of the gRNA that is changed for the gene of interest.

Oligo name	Description	Sequence (5'→ 3')	Reference
gRNA overhang fw	Overhangs matching pMAZ-SK backbone after USER treatment	GAGCAC-N_{20}-GTTTTAGAGCTAGAAAT	***(15)***
gRNA overhangs rv	Overhangs matching pMAZ-SK backbone after USER treatment	CTAAAAC-N_{20}-GTGCTCAGTATCTCT	***(15)***
pMAZ-SK backbone fw	Amplification of pMAZ-SK backbone for gRNA cloning	AGCTAGAAAUAGCAAGTTAAAATAAGGC	***(15)***
pMAZ-SK backbone rv	Amplification of pMAZ-SK backbone for gRNA cloning	AGTATCTCUATCACTGATAGGGATGTCA	***(15)***

3.2. Design of ssDNA CRMAGE oligo with randomized TIR and changed PAM

An overview of CRMAGE oligo design is illustrated in **Fig. 3**. The CRMAGE ssDNA oligos for each gene target must be designed to target the lagging strand of the replication fork during DNA replication.

1. The online program MAGE Oligonucleotide Design Tool (MODEST) is used to determine the oligo direction and chromosomal position for lagging strand-targeting. The organism *E. coli* str. K12 MG1655, version NC_000913.3 is chosen as settings ***(20)***.
2. Insertion of the PROTi tag is done after the first codon downstream of the start codon of the target gene.
3. To apply CRISPR-Cas9 as negative selection during CRMAGE, the wild type PAM sequence (5'-NGG-3') is changed to avoid Cas9 recognition. This substitution should only cause a synonymous substitution in the coding sequence (*see* **Note 2**).
4. On each side of the chromosomal modifications (insertion of PROTi tag and change of PAM), homology regions are needed with the minimum size of 35 nucleotides (*see* **Note 3**).
5. For randomization of the TIR region, the six nucleotides upstream of the start codon are completely randomized. The downstream six nucleotides are randomized only to allow synonymous substitutions in the coding sequence. Due to the position for

PROTi tag insertion, the first codon of the TEV-recognition site (first part of PROTi tag) is also randomized as part of the TIR region **(Fig. 3).**

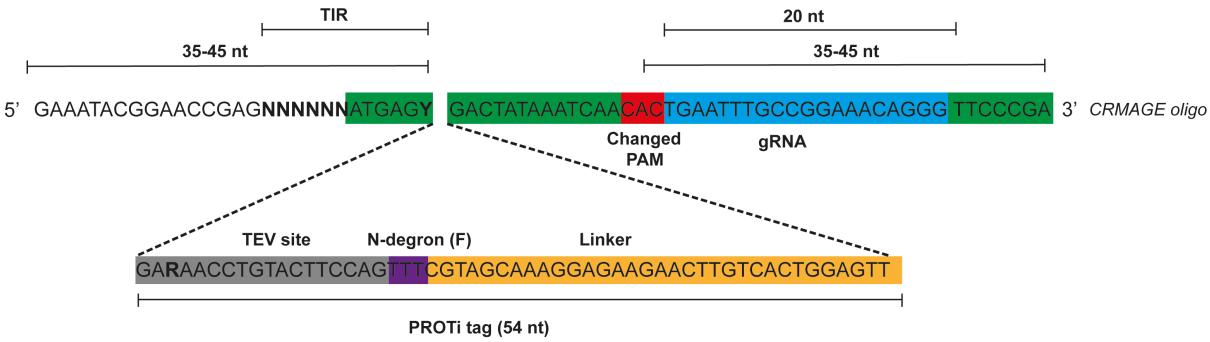


Fig. 3 Illustration of the design of a CRMAGE oligonucleotide. The CRMAGE oligonucleotide encodes the PROTi tag (illustrated in grey, purple and yellow) and contains 35-45 nucleotide end homology to ensure efficient genomic insertion. The PAM sequence (red) is chosen in the target coding sequence (specific nucleotides shown here is a random example) and must be of the canonical form 5'-NGG-3' or 5'-CCN-3'. To avoid Cas9 recognition of clones with inserted PROTi tag, a single-nucleotide substitution is performed in one of the guanine nucleotides (or cytosine) of the PAM (shown here as the second C substituted with A). This substitution should only be a synonymous codon substitution. The translation initiation region (TIR) is completely randomized upstream of the start codon (ATG) (shown here as NNNNNN in bold). In addition, the first codon downstream of the start codon and the first codon of the TEV recognition site (highlighted in bold) are synonymously randomized (Y - cytosine or thymine in the second codon and R - adenine or guanine in the first codon of the TEV sequence). The total length of the ssDNA CRMAGE oligo typically vary from 120 to 200 nucleotides, dependent of the location of the chosen PAM sequence and gRNA. The genomic sequence of the gene *ileS* is shown here as an example.

3.3. Generation of PROTi-tagged strains

A modified protocol for CRMAGE is used ***(15)***. Cultures are grown at 37°C in a water-bath shaking at 250 rpm.

1. The starting strain is *E. coli* K-12 MG1655 harboring pMA7CR.2.0 and pZS4Int-tetR plasmids ***(15)***.
2. An overnight culture is diluted to an OD_{600} of 0.05 in 15 mL LB-lennox with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol to select for pMA7CR.2.0 and pZS4Int-tetR.

3. When the culture reaches an OD_{600} of 0.4-0.5, L-arabinose is added to a final concentration of 0.2% for expression of the λ-Red beta proteins and growth is continued for 15 min.
4. The culture is cooled on ice for 15 min, moved to 50 mL Falcon tubes, harvested by centrifugation at 6.500xg for 7 min at 4°C and washed with 35 mL ice-cold MilliQ water. The harvesting and washing step is repeated.
5. After discarding the supernatant, the culture is resuspended in 1 mL ice-cold MilliQ water and transferred to a 2 mL Eppendorf tube for an additional washing step. The culture is spun at 11.000xg for 1 minute at 4°C in a table-top centrifuge.
6. The sedimented cells are finally resuspended in 0.4 mL ice-cold MilliQ water.
7. For each CRMAGE reaction, 50 µL of prepared cells, 1 µL of the target ssDNA CRMAGE oligo (5 pmol/ µL) and 250 ng of the corresponding pMAZ-SK plasmid with inserted gRNA are mixed in PCR tubes. The samples are electroporated at 1.8 kV, 200 Ω, 25 µF for 5 mS immediately after electroporation. Then 950 µL of LB-lennox containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol are added for recovery and the sample transferred to a new 50 mL Falcon tubes.
8. After 1 hour of recovery, kanamycin is added to a final concentration of 50 µg/mL and incubated for a further 2 hours.
9. Anhydrotetracycline is added to a final concentration of 200 ng/mL and incubated overnight.
10. Cultures are plated on selective LB agar plates and identification of clones with the inserted PROTi tag is performed by colony PCR with screening oligos (Table 2).

3.4. Protein depletion by PROTi and CRiPi

1. Clones with identified PROTi tags are transformed according to standard transformation methods ***(21)*** with the plasmid-borne PROTi or CRiPi systems (Table 1).
2. From overnight cultures, 10 µL of cells are inoculated in 1 mL LB supplied with the appropriate antibiotics in a 96-well plate and incubate at 37°C shaking at 300 rpm.
3. L-rhamnose is added to a final concentration of 5 mM for PROTi-induced protein degradation and anhydrotetracycline to a final concentration of 200 ng/mL for

transcriptional inhibition. For CRiPi induction, both rhamnose and anhydrotetracycline are added (*see* **Note 4**).

4. Dependent on the target protein, the phenotype of protein depletion can be measured by reduction in fluorescence, western blotting, cell viability etc. (*see* **Note 5**).

4. Notes

1. The PAM sequence is chosen around 3-20 nucleotides downstream of the start codon in the coding sequence to avoid interference with upstream located ribosomal binding site, sites for regulatory elements or other genes in operons.
2. Substitution of the second guanine (5'-N**G**G-3') to an adenine (5'-N**A**G-3') in the PAM sequence should be avoided due to weak Cas9 recognition.
3. The position of the chosen PAM sequence and its single-nucleotide substitution to avoid Cas9 recognition influence the total size of the CRMAGE oligo. From the changed single-nucleotide substitution in the PAM and inserted PROTi tag, the CRMAGE oligo must contain a minimum of 35-nucleotide end homology to ensure efficient chromosomal insertion. The total size of the ssDNA CRMAGE oligo should be designed as short as possible due to possible inhibitory secondary structures (e.g. hairpins) that can decrease the efficiency of chromosomal insertion.
4. Dependent on the target gene and its encoding protein and research purpose, protein depletion by the CRiPi system can be induced at different time points e.g. corresponding to different growth stages. Protein depletion can be achieved by inducible PROTi-induced protein degradation, CRISPRi-induced transcriptional repression or by the combined CRiPi system. In our previous work, induced protein degradation of GFP was induced after 4 hours of growth. Applying CRISPRi additionally accelerated GFP depletion. For the essential genes investigated in our previous work, PROTi-induced growth inhibition was observed with rhamnose supplemented from the beginning of growth. CRISPRi-induced transcriptional inhibition for the essential genes was observed when inducing at the beginning of growth and after 4 hours of growth. Synergetic CRiPi-induced protein depletion by combining transcriptional repression with protein degradation was observed for two of the tested essential genes when inducing the CRiPR system after 4 hours of growth ***(11)***.

5. CRiPi-induced knockdown phenotypes can be observed in liquid culture or by plating on solid medium. Protein depletion of essential proteins was observed as a reduction in optical density and in colony forming unit numbers when plating induced culture for visualization of single colonies.

References

1. I.M. Brockman and K.L.J. Prather (2015) Dynamic metabolic engineering: New strategies for developing responsive cell factories, Biotechnology Journal. 10, 1360–1369.
2. E.K. Schrader, K.G. Harstad, and A. Matouschek (2009) Targeting proteins for degradation., Nature chemical biology. 5, 815–22.
3. K.E. McGinness, T.A. Baker, and R.T. Sauer (2006) Engineering Controllable Protein Degradation, Molecular Cell. 22, 701–707.
4. D.E. Cameron and J.J. Collins (2014) Tunable protein degradation in bacteria, Nature Biotechnology. 32, 1276–1283.
5. C. Taxis, G. Stier, R. Spadaccini, et al. (2009) Efficient protein depletion by genetically controlled deprotection of a dormant N-degron, Molecular systems biology. 5, 1–7.
6. A. Bachmair, D. Finley, and A. Varshavsky (1986) In vivo half-life of a protein is a function of its amino-terminal residue., Science. 234, 179–186.
7. J.W. Tobias, T.E. Shrader, G. Rocap, et al. (1991) The N-End Rule in Bacteria, Science. 254, 1374–1377.
8. I. Rivera-Rivera, G. Román-Hernández, R.T. Sauer, et al. (2014) Remodeling of a delivery complex allows ClpS-mediated degradation of N-degron substrates., PNAS. 111, E3853-9.
9. D.K. Gonda, A. Bachmair, I. Wünning, et al. (1989) Universality and structure of the N-end rule, The Journal of biological chemistry. 264, 16700–16712.
10. T. Potuschak, S. Stary, P. Schlögelhofer, et al. (1998) PRT1 of Arabidopsis thaliana encodes a component of the plant N-end rule pathway., Proceedings of the National Academy of Sciences of the United States of America. 95, 7904–8.
11. V. Martínez, I. Lauritsen, T. Wolff, et al. CRISPR/Cas9-based genome editing for simultaneous interference with gene expression and protein stability, In preparation.

12. R.B. Kapust, J. Tözsér, T.D. Copeland, et al. (2002) The P1' specificity of tobacco etch virus protease, Biochemical and Biophysical Research Communications. 294, 949–955.
13. A. Erbse, R. Schmidt, T. Bornemann, et al. (2006) ClpS is an essential component of the N-end rule pathway in Escherichia coli, Nature. 439, 753–756.
14. L.S. Qi, M.H. Larson, L.A. Gilbert, et al. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression., Cell. 152, 1173–83.
15. C. Ronda, L.E. Pedersen, M.O.A. Sommer, et al. (2016) CRMAGE: CRISPR Optimized MAGE Recombineering., Scientific reports. 6, 19452.
16. M. Kozak (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes, Gene. 361, 13–37.
17. J.E. McCarthy and C. Gualerzi (1990) Translational control of prokaryotic gene expression, Trends in genetics. 6, 78–85.
18. K. Mirzadeh, V. Martínez, S. Toddo, et al. (2015) Enhanced Protein Production in Escherichia coli by Optimization of Cloning Scars at the Vector − Coding Sequence Junction, ACS Synthetic Biology. 4, 959–965.
19. A.M. Cavaleiro, S.H. Kim, S. Seppälä, et al. (2015) Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning, ACS Synthetic Biology. 4, 1042–1046.
20. M.T. Bonde, M.S. Klausen, M. V. Anderson, et al. (2014) MODEST: a web-based design tool for oligonucleotide-mediated genome engineering and recombineering, Nucleic Acids Research. 42, W408–W415.
21. J. Sambrook, E.F. Fritsch, and T. Maniatis (1989), Molecular cloning: a laboratory manual.

Paper 3

A versatile one-step CRISPR-Cas9 based approach to plasmid-curing

Lauritsen *et al. Microb Cell Fact (2017) 16:135*
DOI 10.1186/s12934-017-0748-z

Microbial Cell Factories

TECHNICAL NOTES

Open Access

A versatile one-step CRISPR-Cas9 based approach to plasmid-curing

Ida Lauritsen[†], Andreas Porse[†], Morten O. A. Sommer and Morten H. H. Nørholm[*]

Abstract

Background: Plasmids are widely used and essential tools in molecular biology. However, plasmids often impose a metabolic burden and are only temporarily useful for genetic engineering, bio-sensing and characterization purposes. While numerous techniques for genetic manipulation exist, a universal tool enabling rapid removal of plasmids from bacterial cells is lacking.

Results: Based on replicon abundance and sequence conservation analysis, we show that the vast majority of bacterial cloning and expression vectors share sequence similarities that allow for broad CRISPR-Cas9 targeting. We have constructed a universal plasmid-curing system (pFREE) and developed a one-step protocol and PCR procedure that allow for identification of plasmid-free clones within 24 h. While the context of the targeted replicons affects efficiency, we obtained curing efficiencies between 40 and 100% for the plasmids most widely used for expression and engineering purposes. By virtue of the CRISPR-Cas9 targeting, our platform is highly expandable and can be applied in a broad host context. We exemplify the wide applicability of our system in Gram-negative bacteria by demonstrating the successful application in both *Escherichia coli* and the promising cell factory chassis *Pseudomonas putida*.

Conclusion: As a fast and freely available plasmid-curing system, targeting virtually all vectors used for cloning and expression purposes, we believe that pFREE has the potential to eliminate the need for individualized vector suicide solutions in molecular biology. We envision the application of pFREE to be especially useful in methodologies involving multiple plasmids, used sequentially or simultaneously, which are becoming increasingly popular for genome editing or combinatorial pathway engineering.

Keywords: CRISPR-Cas9, Plasmid-curing, pFREE, Replicon analysis, *Pseudomonas putida*, Genome engineering

Background

Since their discovery in the early 1950s, plasmids have played a pivotal role in the advancement of molecular biology, and form the basis for DNA cloning and gene expression in modern biotechnology [1]. While the diversity and applications of cloning vectors have grown dramatically, the vector backbones used today are, for historical reasons, build upon a limited set of parts [2–6].

A central property of a plasmid is its replication machinery that determines the copy-number and ability of plasmids to co-exist [7]. One group of cloning vectors that display a relatively high copy-number is based on the ColE1-like replication machinery, including the pMB1 replicon of pBR322 and its high-copy derivatives found in e.g. pUC18/19, pBluescript® and pJET1.2® [5, 8]. All of the ColE1-derived replicons function via anti-sense RNA for replication control but are able to co-reside to some degree. This group of RNA-controlled ColE-like replicons also contains the widely used p15A replicon that can stably exist together with ColE1-like plasmids and is maintained in fewer copies per cell [3, 9]. A large proportion of naturally occurring plasmids replicate through the use of replication (Rep) proteins that act in a self-inhibitory fashion to control plasmid copy-number [10]. These include the replicons of pBBR1, RK2 and RSF1010 that are found in cloning vectors and function in a broad host context [11]. Similarly, the Rep protein based pSC101 vector was the first to be used for recombinant gene

*Correspondence: morno@biosustain.dtu.dk
[†]Ida Lauritsen and Andreas Porse contributed equally to this work
Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

expression, and is popular due to its relatively high stability in spite of a low copy-number (<8 copies per cell) along with the ability to co-exist with ColE1-like and p15A replicons [1, 12].

While techniques for transfer of plasmid DNA into many bacterial hosts are well established, obtaining plasmid-free cells still poses a significant challenge [13, 14]. In genome and metabolic engineering, the introduction of one or more plasmid-based genetic tools is often required, although a plasmid-free strain is eventually desired [15–18]. For example, sequential steps of plasmid-based genome editing, and the use of screening and characterization tools for strain engineering might involve multiple vectors that need removal prior to final application of the strain [17, 19].

Due to the high copy-number and intrinsic stability of modern cloning vectors, plasmid-curing is often tedious. Traditional methods for plasmid-curing are based on prolonged growth under stressful conditions, such as elevated temperature or the addition of DNA intercalating agents, to interfere with plasmid replication [14]. Other methods based on replicon-incompatibility exploit competition between identical replicons but require precise knowledge of the replication machinery of the target plasmid, as well as subsequent curing of the interfering plasmid [13, 20]. A considerable downside of the existing methods is the variable efficiency, time consumption, and the risk of accumulating unwanted mutations due to prolonged growth regimes and the use of mutagenic curing agents [17, 21].

To accommodate the need for efficient removal of cloning vectors when needed, temperature sensitive plasmid-replicons have been developed [22]. However, the relatively large size, temperature restrictions, low copy-number and little variety of these vectors, complicates cloning procedures and limits their application for multi-plasmid and broad-host purposes. Another way to facilitate the selection of plasmid-free clones is by incorporating a counter-selectable marker into the plasmid backbone [23]. Although this strategy allows for rapid identification of cells lacking the marker gene, these do not actively remove the plasmid and negative selection markers are prone to mutational escape and often have stringent requirements to the growth media and host background [23–25].

With the advent of CRISPR-Cas9 technology, mimicking the natural bacterial defense against plasmid and phage intruders, a powerful and flexible approach to precise DNA targeting is now available for a wide range of organisms [26, 27]. Although CRISPR-Cas9 has been applied for specific targeting of certain plasmid features, a generally applicable platform for quick and efficient curing of cloning vectors will constitute a highly useful tool in molecular biology [28].

Here we exploit the common origin of modern plasmid vectors to develop a broadly applicable CRISPR-Cas9-based curing platform. We show that our system enables fast and efficient curing of all major plasmid replicons used in modern molecular biology laboratories and can be applied in a broad phylogenetic context.

Results

We first explored the distribution of cloning vector replicons by performing a BLAST search of selected replicons against all bacterial plasmids with full nucleotide sequences available in the Addgene plasmid repository [29] (Fig. 1). The ColE1-like (including p15A) and pSC101 replicons accounted for 91% of the plasmids in the Addgene database. The vast majority of these plasmids belonged to the ColE1 family (86.4%), underlining the popularity of these vectors in molecular biology (Fig. 1). Surprisingly, a considerable fraction of vectors annotated with the pBBR1 and RK2 broad host-range replicons also contained full-sized ColE1-like replicon sequences. Including these redundant replicons in our calculations, a plasmid-curing system targeting the ColE1 and pSC101 plasmid groups will cover 93.3% of the (at present 4657) bacterial vectors deposited in Addgene (Additional file 1: Figure S1).

Through sequence alignments of representative replicons from each replicon-group, we identified highly conserved regions between all ColE1-like replicons that were also shared with p15A (Fig. 2a). These regions were used to design CRISPR-Cas9-compatible guide RNA (gRNA) that, upon recognition by Cas9, target all ColE1-like and p15A vectors. Because the protein-based mechanism of

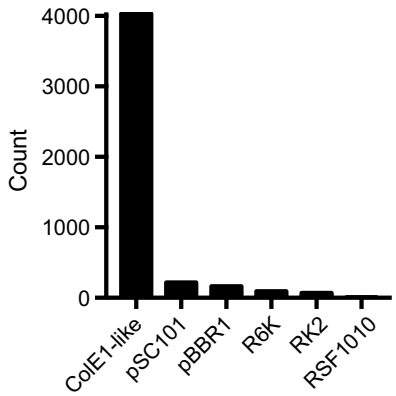


Fig. 1 Frequency of major replicons in bacterial cloning and expression vectors deposited to Addgene. A BLAST search was performed against all complete bacterial vector sequences (4657) in the Addgene database (Feb. 2017). ColE1-like plasmids include the closely related RNA-based replicons of ColE1, pBR322/pMB1, pUC18/19, pJET1.2®, colA and p15A

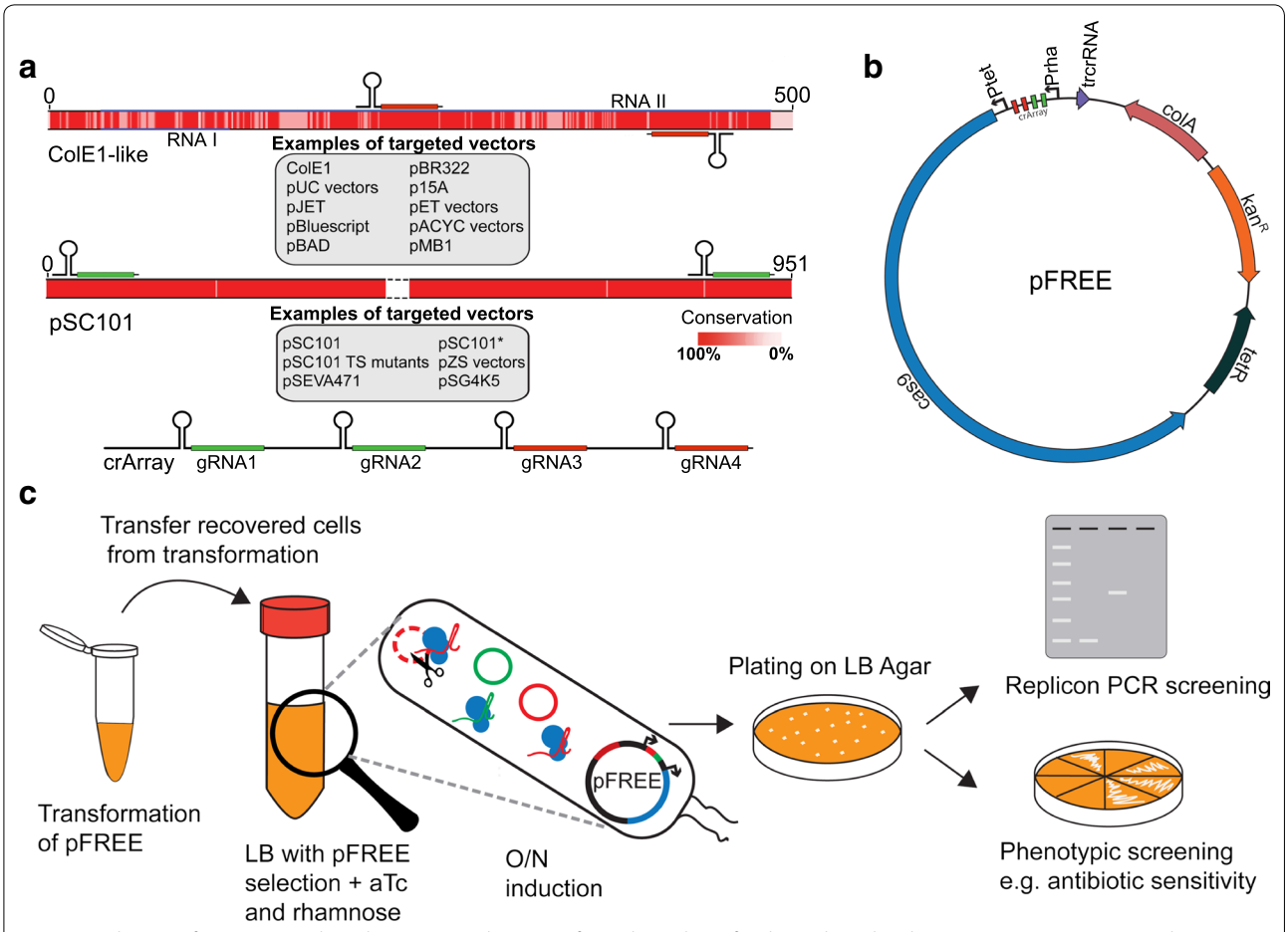


Fig. 2 **a** Selection of gRNA-targets based on conserved regions of popular replicon-families. Selected replicon sequences representing the RNAI and RNAII encoding part of ColE1 and the *repA* encoding part of pSC101 replicon groups, were aligned and gRNA was selected based on the degree of conservation (illustrated as the color intensity). The center part of *repA* was fully conserved and omitted in the depiction. *The plasmid names in grey boxes* are examples of vectors belonging to each replicon family. Two gRNAs were selected for each replicon group, and all four gRNAs were combined into a CRISPR-array (crArray). **b** Plasmid map of pFREE. The pFREE plasmid was constructed by inserting the crArray targeting the ColE1 and pSC101 replicons into a colA vector encoding Cas9 along with other essential modules for CRISPR-Cas9 activity such as trans-activating CRISPR RNA (trcrRNA). The gRNA array and Cas9 nuclease are controlled by the inducible rhamnose (PrhaBAD) and tetracycline (Ptet) promoters to ensure tight regulation of curing functionality. **c** One-step curing workflow using the pFREE system. The pFREE plasmid is transformed into a strain harboring the target plasmids for curing. After transformation recovery, cells from the recovered culture are transferred into medium with pFREE selection, 0.2% rhamnose and 200 ng/mL anhydrotetracycline (aTc) added. The system is induced overnight (O/N) to allow the cleavage of target plasmids (*red and green* respectively) by Cas9 (*blue*), guided by the gRNA expression from pFREE (*black plasmid*). The culture is plated on non-selective agar and cured cells can be identified by replicon PCR (Additional file 1: info S2) or by phenotypic screening e.g. antibiotic sensitivity

pSC101 replication is fundamentally different from that of the ColE1-like replicons, we designed separate gRNA to facilitate curing of the pSC101-based vectors. To increase curing efficiency and counteract the potential for mutational escape, we included two gRNA targets for each replicon group (Table 1).

The four gRNAs were implemented as a CRISPR-array, along with the tracrRNA and Cas9-components and incorporated into a single vector containing all parts necessary to form the fully functional curing system designated "pFREE" (Fig. 2b). An important feature of

Table 1 Selected gRNAs and their target replicons

gRNA	Sequence (5′ to 3′)	Targeted replicon group
gRNA1	ATGAACTAGCGATTA GTCGCTATGACTTAA	*pSC101*
gRNA2	AACCACACTAGAGAA CATACTGGCTAAATA	*pSC101*
gRNA3	GGTTGGACTCAAGAC GATAGTTACCGGATA	*ColE1-like except colA*
gRNA4	GGCGAAACCCGACAG GACTATAAAGATACC	*ColE1-like including colA (self-curing of pFREE)*

a plasmid-curing system is a suicide functionality that renders the resulting cells completely plasmid-free without any additional incubation steps. The pFREE vector is based on the colA replicon that resembles ColE1-like replicons to some degree but colA is only recognized by one of the ColE1-targeting gRNAs. Due to the self-curing feature of pFREE, plasmid-curing can be done in a one-step workflow directly after transformation of pFREE as outlined in Fig. 2c.

Quantification of curing efficiency

In order to test the efficiency of our plasmid-curing system, we constructed three target plasmids by inserting *gfp* under control of a constitutive promoter into similar backbones of the pZ vector system [30]. These three plasmids differ only by their ColE1, p15A or pSC101 replicons and are designated pZE-GFP, pZA-GFP and pZS-GFP. The curing efficiency was quantified at different time points, and the loss of fluorescence reflected plasmid-curing of the *gfp* expressing vectors. The curing rates were comparable between the target plasmids and after 24 h the vast majority of all three populations were cured with 80–90% of the plated cells being plasmid-free (Fig. 3). Non-fluorescent cells were assessed for self-curing of the pFREE plasmid by kanamycin sensitivity, and no pFREE-carrying cells were detected after 24 h. These results clearly demonstrate effective plasmid-curing of vectors with ColE1, p15A and pSC101 replicons, targeted by the crArray of the pFREE system, and efficient self-curing of the pFREE plasmid.

pFREE cures major cloning vector-systems used in *E. coli*

Seven representatives of widely used cloning and expression vector systems were selected to demonstrate the general applicability of the pFREE system to cure commonly used vectors with similar replicons but variable backbone content. The majority of these plasmids contained variations of ColE1 replicons including the pJET1.2®, pUC19 and pBluescript® high copy-number variants as well as a pET-vector most commonly used for

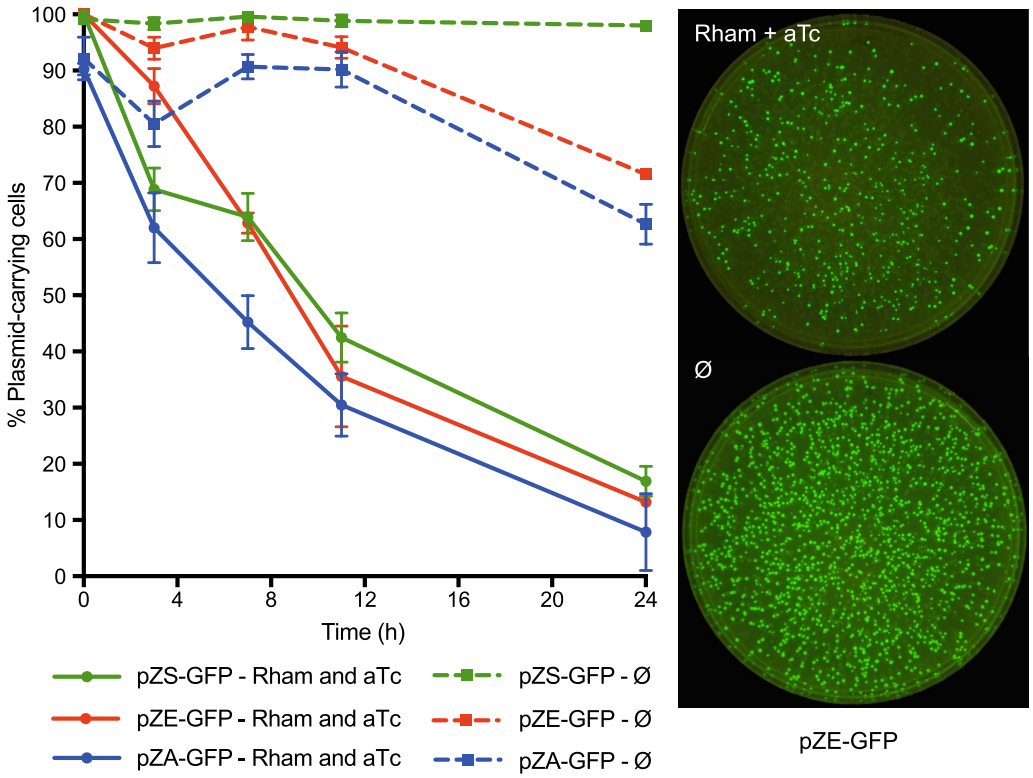


Fig. 3 Time course characterization of the pFREE plasmid-curing system. Curing of pZ-plasmids expressing GFP with either pSC101 (pZS-GFP, *green*), ColE1 (pZE-GFP, *red*) or p15A (pZA-GFP, *blue*) replicon. *The solid lines* indicate induced cultures with rhamnose (Rham) and anhydrotetracycline (aTc), whereas the *dashed lines* refer to non-induced (Ø). Plating was performed at induction time (0) and 3, 7, 11 and 24 h after induction. Between 100 and 150 colony forming units (CFUs) were counted from each replicate and the ratio between fluorescent and non-fluorescent cells were determined. The percentage of plasmid-carrying cells is depicted. Of the non-fluorescent and tested cells, all had lost the pFREE plasmid after 24 h. Data points represent mean value of three biological replicates with *error-bars* showing standard deviation. Representative LB agar plates for pZE-GFP with equal number of cells plated with cultures induced with rhamnose and aTc (*top*) and non-induced (Ø) (*bottom*) of the pFREE system after 24 h

protein production. In addition, the low copy-number pSEVA471 [31] plasmid harboring a pSC101 replicon and the p15A-based pACYC-Duet-1 medium-copy plasmid was also included.

While the pSEVA471 and pACYC-Duet-1 plasmids were cured with similar efficiency to what was observed for the pZ plasmids (Figs. 3, 4), the ColE1-like replicons were cured with efficiencies ranging from 40 to 100%. These results exemplify that, although replicon context does play a role, the pFREE system can be used for efficient curing of the most common commercial plasmid vectors with varying copy-numbers and auxiliary content.

One-step curing of multiple plasmids

To improve the practical application of the pFREE system as a fast and simple curing system, we developed a one-step workflow as displayed in Fig. 2c. Plasmid-curing is induced directly after pFREE-transformation and completely plasmid-free clones (without target and pFREE plasmids) can easily be detected either by phenotypic screening (e.g. antibiotic sensitivity) or faster by the set of universal replicon amplifying PCR oligonucleotides that we developed (Additional file 1: info S2). To test the one-step protocol and to evaluate the performance of the pFREE system for curing multiple plasmids simultaneously, we prepared a strain containing three compatible target plasmids. After transformation of pFREE into this strain, the target plasmids were cured directly from the transformation mix and plated on non-selective LB agar after overnight induction. From the tested cells, 80% were completely cured whereas 10% or less contained one or more plasmids and all cells had lost pFREE (Fig. 5).

Self-curing dynamics of pFREE

To investigate the dynamics of the pFREE self-targeting feature, we quantified the self-curing efficiency of the pFREE plasmid over time. In the absence of plasmid selection, 90% of the cells were cured of pFREE after 7 h of induction whereas 65% of the cells were cured in the presence of plasmid selection (Additional file 1: Figure S2). After 10 h of pFREE induction, all cells were cured for pFREE regardless of the plasmid selection.

pFREE-RK2: a temperature sensitive and broad host-range version of pFREE

The curing efficiency of the pFREE system was between 40 and 100% as depicted in Figs. 3 and 4. Due to the highly efficient self-curing of pFREE, we speculated that over-efficient self-targeting could be a bottleneck preventing complete curing of the target plasmids. In that case, a system allowing self-curing to take place only after the target plasmid-curing has occurred, might increase duration of CRISPR expression and consequently the

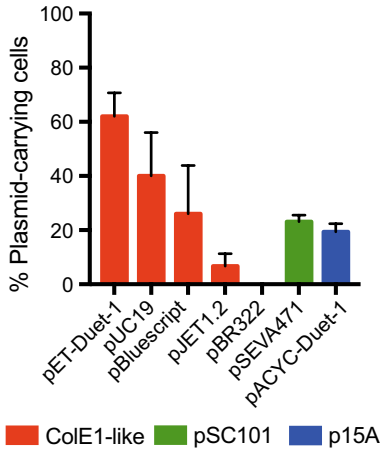


Fig. 4 pFREE-mediated curing of selected widely used cloning vectors with either ColE1-like (*red*), pSC101 (*green*) or p15A (*blue*) replicons. 50 CFUs from each replicate of each target plasmid was checked for antibiotic sensitivity after 24 h of induction of the pFREE system. The percentage of plasmid-carrying cells is depicted. The pFREE plasmid was cured in all colonies tested. Curing of the target plasmids was verified by replicon PCR (Additional file 1: info S2). The *bars* represent mean value of three biological replicates with *error-bars* showing standard deviation

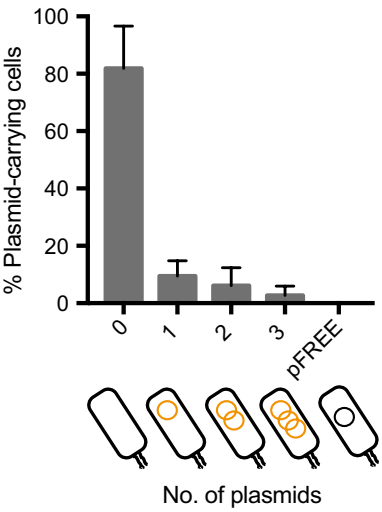


Fig. 5 Curing of multiple co-residing plasmids using the one-step transformation protocol depicted in Fig. 2c. The pFREE plasmid was transformed into a strain harboring the same three pZ plasmids as used in Fig. 3. After recovery, plasmid-curing with the pFREE system was induced and cultures were plated after 24 h of induction. 50 CFUs from each replicate were checked for antibiotic sensitivity on LB agar plates. The percentage of cells carrying 0, 1, 2, 3 (*orange plasmids*) or pFREE (*black plasmid*) is depicted. The *bars* represent mean value of three biological replicates with *error-bars* showing standard deviation

curing efficiency. To compare the effect of self-curing mechanism and copy-number on curing outcomes, we designed a pFREE version with the temperature-sensitive, low-copy RK2 replicon that replicates in a broad representation of Gram-negative bacteria [32] designated pFREE-RK2 (Additional file 1: Figure S3). The RK2 replicon is not targeted by the pFREE crArray, thus omitting the CRISPR-Cas9-based self-curing feature of the pFREE system but carries a *trfA* mutant that allows curing at elevated temperatures instead [33]. The curing efficiency of the pFREE-RK2 system was quantified in the same way as for pFREE and exhibited comparable curing efficiencies of 35–100% (Additional file 1: Figures S4, S5). The highly similar curing efficiencies observed for pFREE and pFREE-RK2, indicates that simultaneous self-curing and curing of target plasmid does not significantly affect the overall curing efficiency of the system. Multi-plasmid curing by the one-step protocol was also tested for pFREE-RK2 and showed comparable efficiency (Additional file 1: Figure S6), demonstrating that a lower copy-number and fundamentally different mechanism of self-curing does not alter CRISPR-Cas9 targeting efficiency.

Curing in *Pseudomonas putida*

To demonstrate the versatility of our plasmid-curing system in a broader phylogenetic context, we set out to test the pFREE-RK2 in an alternative host bacterium supported by the RK2 replicon. We chose the Gram-negative soil bacterium *Pseudomonas putida* as our model host due to its promise as a new and powerful chassis for metabolic engineering and production of fine chemicals [34]. Using the *P. putida* strain KT2440 harboring the *gfp*-expressing pSEVA441-GFP plasmid, we targeted the ColE1-based pRO1600/ColE1 fusion replicon without the need to change any components of the pFREE-RK2 curing plasmid. After overnight induction of the curing system, approximately half of the *P. putida* population (53% SD ± 5.1%, three biological replicates) was cured for pSEVA441-GFP, whereas no detectable curing was observed without pFREE-RK2. These results demonstrate that our CRISPR-based curing system can be applied in a broader host context and that CRISPR-Cas9 technology can be successfully applied in *P. putida*.

pFREE enables precise curing without off-target effects

The curing functionality of pFREE is tightly regulated and the gRNAs were selected to avoid potential off-target effects of CRISPR-Cas9 expression [35]. However, to ensure that the curing activity of pFREE did not induce off-target effects, we whole-genome sequenced three individual isolates of *E. coli* and *P. putida* harboring pFREE or pFREE-RK2 respectively before and after the curing procedure. The sequencing results showed that 24 h of induction with the pFREE system did not cause mutations (SNPs and small INDELS) or larger rearrangements in the host genomes; confirming the orthogonality of pFREE in these hosts.

Discussion

Plasmids are fundamental in all aspects of molecular biology due to their role as genetic scaffolds that are easy to modify and transfer between hosts. However, when plasmids carry functions that are only temporarily necessary, or a clean strain background is needed, limited options are currently available for efficient plasmid-curing of the most widely used cloning vectors in bacteria.

Existing methods for plasmid-curing are based on curing agents or incompatibility mediated plasmid displacement [13, 20]. However, these methods require sequential rounds of growth in stressful or non-selective conditions to promote the appearance of plasmid-free segregants. Such methods increase the chance of accumulating unwanted mutations and are time-consuming. Prior work has demonstrated that plasmids, traditionally considered incompatible can co-exist stably for multiple growth cycles [36], which only complicates incompatibility-based plasmid-curing further; especially for plasmids maintained at high copy-numbers.

To address this methodological bottleneck, we developed the pFREE system as a fast and simple one-step plasmid curing-method based on sequence conservation within replicon groups and CRISPR-Cas9-targeted plasmid cleavage. Using this system, curing of one or multiple target plasmids can be performed directly after transformation of the pFREE plasmid and cured cells can easily be screened for specific phenotypes (e.g. antibiotic resistance) or by the diagnostic PCR developed here (Additional file 1: info S2). In the absence of prior plasmid sequence information, the PCR based replicon identification protocol is also useful for replicon profiling prior to curing (Additional file 1: info S2).

Using the pFREE system, we cured both single and multiple plasmids with an efficiency of 40–100% (Figs. 3, 4). We investigated the dynamics of the pFREE self-curing feature and observed complete curing of pFREE already after 10 h with kanamycin added for pFREE selection. Although the inclusion of pFREE selection during curing reduced the self-curing rate, and allows for the one-step transformation protocol, it also shows that cells that are actively cured during selective culturing are not necessarily killed (Additional file 1: Figure S2; Fig. 2c).

Such persistence may result from slower degradation of the resistance conferring aminoglycoside phosphotransferase enzyme compared to the rate of plasmid-curing, or could be an effect of indirect resistance were the

antibiotic sensitive cells are protected by pFREE-carrying cells [37].

We first speculated that the highly efficient self-curing of pFREE was limiting the trans-curing efficiency of pFREE. However, similar curing efficiencies were observed when the self-targeted colA replicon was replaced with the temperature sensitive RK2 replicon (pFREE-RK2).

Differences in curing efficiencies were observed for the individual plasmids tested here; presumably caused by variations in copy-number, plasmid incompatibility or fitness constrains originating from other factors present in the plasmid backbones. We did not observe a clear correlation between copy-number and curing efficiency, with the extremely high copy vectors of pJET1.2®, pBluescript® and pUC19 displaying curing efficiencies similar to the low and medium-copy-number pSEVA471 and pACYC-Duet-1 (Fig. 4). Although the overall curing efficiency was similar between pFREE and pFREE-RK2, there were small differences in the relative efficiency against the different replicon families (Fig. 3; Additional file 1: Figure S4). Such differences, e.g. the higher efficiency for curing of replicons more similar to the replicon of the curing plasmid, could be a result of partial replicon incompatibility and might explain the higher loss of p15A and ColE1-like RNA-based replicons for pFREE when co-residing without induction of the curing system (Fig. 3; Additional file 1: Figure S4). Surprisingly, the pBR322 plasmid was completely cured for both versions of pFREE, whereas the pET-Duet-1, carrying the exact same replicon, displayed the lowest curing efficiency (40% cured) observed here; indicating a substantial effect of auxiliary plasmid factors on plasmid stability (Fig. 4; Additional file 1: Figure S5). Such differences can be caused by factors such as resistance markers or other genetic cargo that affects plasmid persistence at the population level. The pBR322 is known to inflict a fitness cost on *E. coli* hosts due to the expression of the costly tetracycline efflux pump encoded by *tetA* [38]. If a high fitness benefit of losing the plasmid exists, the expansion of the plasmid-free population will contribute exponentially to the observed plasmid loss and synergistically improve the curing outcome.

Mutations in target plasmids or in the CRISPR platform of pFREE along with biological stochasticity could also explain the non-perfect curing of target plasmids by our system. CRISPR-Cas9 systems are widely used for genome-editing purposes, and other applications have shown similar susceptibility to small subpopulations of "escapers" that avoid targeting [39].

We developed the crArray encoded by pFREE to target replicons belonging to the ColE1/p15A and pSC101 groups based on the distribution of replicons in bacterial vectors deposited in the Addgene database, which agreed with historical trends in cloning vector usage [1, 5, 9]. Additionally, we discovered that the selected gRNAs in pFREE indirectly target vectors with other replicons such as pBBR1 and RK2 due to redundant replicon sequences present in a high proportion of these backbones (Additional file 1: Figure S1). Although we target the majority of replicons used for routine cloning in *E. coli* there are exceptions within the broad host-range vectors and R6K (Additional file 1: Figure S1). The R6K replicon is primarily used as a suicide vector and is of little relevance in a curing perspective [40]. Since the vast majority of plasmid vectors that are used belong to the ColE1-like, p15A or pSC101 replicon groups (Fig. 1), the chance of a target plasmid being covered by the pFREE system is high. Hence, less knowledge about the replicon group of target plasmids is needed prior to curing compared to incompatibility-based curing methods [13] and only a few colonies will have to be screened to identify a cured variant.

Due to the broad functionality of the CRISPR-Cas9 technology in a variety of organisms [27] the pFREE curing system has great potential as a universal plasmid-curing tool in bacteria, as shown here for both *E. coli* and *P. putida,* and can in theory be expanded to eukaryotic organisms such as yeast where plasmids are also employed [41]. With decreasing cost of nucleic acid synthesis, custom crArrays for targeting of other plasmids than the ones included here are easily implemented into the pFREE backbones. It is possible that a similar approach can be used clinically to combat the increasing medical burden of plasmid-encoded multidrug resistance in pathogenic bacteria. Although the diversity of natural plasmid replicons by far exceeds that of cloning vectors, the most endemic plasmid-families encoding virulence and antibiotic resistance factors do share conserved features within their replication, stability, resistance and conjugation modules that could be targeted for future expansion of our plasmid-curing system [42, 43].

Conclusions

We show that all major replicons used for cloning and expression purposes share sequence features that allow for universal CRISPR-Cas9 targeting and use this information to develop a fast and one-step plasmid-curing platform that allows for targeting of the major classes of vectors used in molecular biology. Using our curing protocol, we demonstrate efficient curing of major cloning and expression vectors in biotechnology and perform in-depth characterization of the curing dynamics. To facilitate subsequent identification of plasmid-cured variants, we supply a set of universal primers that allow for rapid PCR screening directly from a culturing plate. Furthermore, we construct a temperature-sensitive and broad

host-range version of pFREE (pFREE-RK2) that provides an efficient curing solution for broad range of Gram-negative bacteria including the upcoming cell factory *Pseudomonas putida*.

Methods

Replicon prevalence and conservation analysis

Bioinformatic analysis was performed using the CLC Main Workbench (QIAGEN Bioinformatics) and R (version 3.3.1) software. Replicons used in multiple sequence alignments and BLAST searches were downloaded from GenBank or Addgene. Sequences with the following accession numbers were included as ColE1-like replicons: ColE1 (GenBank NC_001371), pBR322/pMB1 (GenBank J01749.1), pUC19 (Addgene plasmid #49793), pJET1.2® (GenBank EF694056.1), ColA (Addgene plasmid #73962). In addition, p15A and pSC101 replicons were included: p15A (GenBank V00309.1) and pSC101 *repA* (GenBank K00042.1), temperature sensitive pSC101 *repA* of pKD46 (GenBank AY048746) and pGRG36 (GenBank DQ460223.1). For the BLAST analysis, the R6K (GenBank KX485333.1) and broad host-range replicons of pBBR1 (GenBank U02374.1), *trfA* gene of RK2 (GenBank U05774.1) and RSF1010 (GenBank M28829.1) were also included.

Multiple alignments were used to identify conserved regions in the selected replicon sequences. gRNA was selected based on broad conservation in replicons, as well as the absence of matches to proteobacterial chromosomes in NCBIs RefSeq database; where at least four chromosomal mismatches were present for each gRNA sequence.

We downloaded all (4657) Addgene entries of bacterial plasmids where the full nucleotide sequence was accessible from the search function at https://www.addgene.org/ (accessed Feb. 2017). Replicon frequencies and positive gRNA hits were assessed using BLAST [44]. An e-value cutoff of 1e-130 was used for replicon BLAST and allowed proper classification according to database annotations. For gRNA BLAST searches, only hits with a perfect match to the query replicons were included as positive hits.

Plasmid construction

pFREE was constructed by amplification of the pMAZ-SK backbone [18] using oligonucleotides 1 and 2. See Additional file 1: Table S3 for all plasmids and references and Additional file 1: Table S4 for all oligonucleotides used in this study. The crArray encodes four different gRNAs of 30 nts, separated by direct repeats of 36 nts. The crArray was constructed by PCR using two ultramer oligonucleotides 3 and 4 (size of 200 nts and 173 nts respectively) with an overlapping region of 72 nts mixed with the two uracil-containing oligonucleotides 5 and 6, and cloned into the pMAZ-SK amplified PCR backbone by USER cloning as described previously [45]. Insertion of the tetratracycline repressor (*tetR*) was performed by Gibson assembly as described elsewhere [46] with oligonucleotides 7 and 8 for pMAZ-SK backbone amplification and *tetR* amplification from plasmid pZS4Int-tetR with oligonucleotides 9 and 10. Oligonucleotides 11 and 12 were used to amplify the pFREE backbone and the *Cas9* gene was amplified from pMA7CR_2.0 [18] with oligonucleotides 13 and 14 and cloned into pFREE by Gibson assembly.

The temperature sensitive broad host-range version of pFREE (pFREE-RK2) was constructed by amplification of the temperature sensitive RK2 replicon from pSIM9 [47] using oligonucleotides 15 and 16, including the *trfA* gene and *oriV* regions. The backbone from pFREE was amplified with oligonucleotides 17 and 18 to insert the RK2 replicon into the pFREE backbone via USER cloning [48]. Likewise, versions of pFREE and pFREE-RK2 with ampicillin, chloramphenicol and zeocin resistance genes were constructed and all pFREE constructs are available through Addgene.

Bacterial strains, media and growth conditions

Escherichia coli Top10 (Thermo Fisher Scientific, Waltham, MA, USA) was used for cloning and curing experiments. *E. coli* cultures were grown in lysogeny broth (LB) at 30 °C with shaking at 270 rpm. The antibiotics ampicillin (100 μg/mL), chloramphenicol (34 μg/mL), zeocin (100 μg/mL) and kanamycin (50 μg/mL) were added when needed. *crArray* expression was induced with 0.2% L-rhamnose (w/v) and expression of *Cas9* endonuclease was induced with 200 ng/mL anhydrotetracycline (aTc).

Time course curing dynamics of pFREE and pFREE-RK2

Overnight cultures of *E. coli* Top10 harboring the pZA-GFP, pZE-GFP or pZS-GFP plasmid respectively and pFREE or pFREE-RK2 were diluted 2000-fold in 10 mL LB broth containing 0.2% L-rhamnose, 200 ng/mL aTc and 50 μg/mL kanamycin. Cultures were grown from three randomly picked colonies. Time course assessment of curing efficiency was done by plating on non-selective LB agar. For each time-point, the ratio between fluorescent cells and non-fluorescent cells were determined by quantification of GFP-fluorescent colonies among 100–150 CFUs from each plate. To assess self-curing of pFREE, at least 10 colonies were checked for growth on LB agar plates containing kanamycin (50 μg/mL).

Curing of widely used cloning and expression vectors

The plasmids pET-Duet-1, pBR322, pJET1.2®, pUC19, pACYC-Duet-1, pSEVA471 pBluescript® and pSEVA441-GFP were cured with the protocol described above. To test for plasmid-curing, individual colonies were checked for growth on LB agar plates containing the relevant antibiotic. Plasmid-curing was verified by replicon PCR with oligonucleotides S1, S2, S3, S4 and S5 (Additional file 1: info S2).

One-step curing of multiple plasmids

Escherichia. coli Top10 strain harboring three (pZA-GFP, pZE-GFP and pZS-GFP) plasmids was grown in 5 mL LB containing antibiotics for plasmid selection. The culture was grown to an OD_{600} of 0.3 and made electrocompetent by three steps of washing in MilliQ water. 50 µL of competent cells were transformed with 50 ng of pFREE or pFREE-RK2 by electroporation (1.65 kV, 200 Ohm, 25 µF) and recovered for 2 h in 500 µL SOC medium at 30 °C and shaking (500 rpm). After recovery, 50 µL of the recovered cells were transferred to 10 mL LB medium with 0.2% L-rhamnose, 200 ng/mL aTc and 50 µg/mL kanamycin added. The cultures were plated on non-selective LB agar after 24 h of incubation. 50 colonies of each replicate were checked on relevant antibiotics to assess the curing efficiency.

Assessment of genomic off-target effects

Two parallel cultures were initiated from each of three individual colonies of *E. coli* carrying pFREE and *P. putida* carrying pFREE-RK2. One culture was induced to activate the curing process while the other functioned as a control without induction of pFREE. Both cultures were grown at 30 °C shaking (250 rpm) for 24 h and genomic DNA was purified using the QIAGEN blood and tissue DNA isolation kit. The genomic DNA was prepared for sequencing using the KAPA HyperPlus Kit (Kapa Biosystems) and the resulting libraries were sequenced on an Illumina NextSeq platform. Fastq output files were imported into CLC Genomics Workbench software (QIAGEN) where all analysis was performed. Reads were trimmed and quality filtered before mapping of the reads, originating from the plasmid cured genomes, to the assembled control genomes. SNP and small INDEL variants were detected using quality based variant detection and larger INDELS and structural variants were assessed using the "Structural Variants and InDels" pipeline as well as by manual inspection of read mappings.

Additional file

Additional file 1. Supplementary information.

Authors' contributions

IL, AP, MHHN and MOAS designed the experiments. IL and AP performed the experiments. IL and AP wrote the manuscript with input from MHHN and MOAS. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge Pablo Ivan Nikel and Patricia Maria Calero Valdayo for providing the *P. putida* strain and plasmid.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All constructs are available from Addgene. Whole genome sequence data has been deposited in NCBI's short read achieve and can be accessed from the Bioproject PRJNA393518.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by the EU H2020 ERC-20104-STG LimitMDR (638902) Grant, the Danish Council for Independent Research Sapere Aude programme DFF-4004-00213 and the Novo Nordisk Foundation.

Publisher's Note

Received: 10 May 2017 Accepted: 22 July 2017
Published online: 02 August 2017

References

1. Cohen SN. DNA cloning: a personal view after 40 years. Proc Natl Acad Sci USA. 2013;110:15521–9.
2. Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci USA. 1973;70:3240–4.
3. Chang ACY, Cohen SN. Construction and characterization of amplifiable DNA cloning vectors derived from P15A cryptic plasmid. J Bacteriol. 1978;134:1141–56.
4. Chang ACY, Cohen SN. Genome construction between bacterial species in vitro: replication and expression of *Staphylococcus* plasmid genes in *Escherichia coli*. Proc Natl Acad Sci USA. 1974;71:1030–4.
5. Hershfield V, Boyer HW, Yanofsky C, Lovett MA, Helinski DR. Plasmid ColEl as a molecular vehicle for cloning and amplification of DNA. Proc Natl Acad Sci USA. 1974;71:3455–9.
6. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. 2014;5:1–17.
7. Novick RP. Plasmid incompatibility. Microbiol Rev. 1987;51:381–95.
8. Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heyneker HL, Boyer HW, et al. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene. 1977;2:95–113.
9. Selzer G, Som T, Itoh T, Tomizawa JI. The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. Cell. 1983;32:119–29.
10. del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejas R. Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev. 1998;62:434–64.
11. Jain A, Srivastava P. Broad host range plasmids. FEMS Microbiol Lett. 2013;348:87–96.
12. Tucker WT, Miller CA, Cohen SN. Structural and functional analysis of the par region of the pSC 101 plasmid. Cell. 1984;38:191–201.

13. Hale L, Lazos O, Haines AS, Thomas CM. An efficient stress-free strategy to displace stable bacterial plasmids. Biotechniques. 2010;48:223–8.
14. Trevors J. Plasmid curing in bacteria. FEMS Microbiol Lett. 1986;32:149–57.
15. Lee J, Saddler JN, Um Y, Woo HM. Adaptive evolution and metabolic engineering of a cellobiose- and xylose- negative *Corynebacterium glutamicum* that co-utilizes cellobiose and xylose. Microb Cell Fact. 2016;15:20.
16. Ginesy M, Belotserkovsky J, Enman J, Isaksson L, Rova U. Metabolic engineering of *Escherichia coli* for enhanced arginine biosynthesis. Microb Cell Fact. 2015;14:29.
17. Schlegel S, Genevaux P, de Gier JW. De-convoluting the genetic adaptations of *E. coli* C41 (DE3) in real time reveals how alleviating protein production stress improves yields. Cell Rep. 2015;10:1758–66.
18. Ronda C, Pedersen LE, Sommer MOA, Nielsen AT. CRMAGE: CRISPR optimized MAGE recombineering. Scientific Reports. Nature Publishing Group. 2016; 6:19452.
19. Raman S, Rogers JK, Taylor ND, Church GM. Evolution-guided optimization of biosynthetic pathways. Proc Natl Acad Sci. 2014;111:201409523.
20. Kamruzzaman M, Shoma S, Thomas CM, Partridge SR, Iredell JR. Plasmid interference for curing antibiotic resistance plasmids in vivo. PLoS ONE. 2017;12:e0172913.
21. Crameri R, Davies JE, Hütter R. Plasmid curing and generation of mutations induced with ethidium bromide in streptomycetes. J Gen Microbiol. 1986;132:819–24.
22. Hashimoto-Gotoh T, Franklin FCH, Nordheim A, Timmis KN. Specific-purpose plasmid cloning vectors I. Low copy number, temperature-sensitive, mobilization-defective pSC101-derived containment vectors. Gene. 1981;16:227–35.
23. Reyrat J-M, Pelicic V, Gicquel B. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. Infect Immun. 1998;66:4011–7.
24. Gregg CJ, Lajoie MJ, Napolitano MG, Mosberg JA, Goodman DB, Aach J, et al. Rational optimization of tolC as a powerful dual selectable marker for genome engineering. Nucleic Acids Res. 2014;42:4779–90.
25. Podolsky T, Fong S, Lee B. Direct selection of tetracycline-sensitive *Escherichia coli* cells using nickel salts. Plasmid. 1996;115:112–5.
26. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA—guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–22.
27. Singh V, Braddick D, Dhar PK. Exploring the potential of genome editing CRISPR-Cas9 technology. Gene. 2017;599:1–18 **(Elsevier B. V.)**.
28. Kim JS, Cho DH, Park M, Chung WJ, Shin D, Ko KS, et al. CRISPR/Cas9-mediated re-sensitization of antibiotic-resistant *Escherichia coli* harboring extended-spectrum Beta-lactamases. J Microbiol Biotechnol. 2015;26:394–401.
29. Kamens J. The Addgene repository: an international nonprofit plasmid and data resource. Nucleic Acids Res. 2015;43:D1152–7.
30. Lutz R, Bujard H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res. 1997;25:1203–10.
31. Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, De Las Heras A, et al. The standard european vector architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res. 2013;41:666–75.
32. Karunakaran P, Blatny JM, Ertesvåg H, Valla S. Species-dependent phenotypes of replication-temperature-sensitive trfA mutants of plasmid RK2: a codon-neutral base substitution stimulates temperature sensitivity by leading to reduced levels of trfA expression. J Bacteriol. 1998;180:3793–8.
33. Haugan K, Karunakaran P, Blatny JM, Valla S. The phenotypes of temperature-sensitive mini-RK2 replicons carrying mutations in the replication control gene trfA are suppressed nonspecifically by intragenic cop mutations. J Bacteriol. 1992;174:7026–32.
34. Nikel PI, Chavarría M, Danchin A, de Lorenzo V. From dirt to industrial applications: *Pseudomonas putida* as a synthetic biology chassis for hosting harsh biochemical reactions. Curr Opin Chem Biol. 2016;34:20–9.
35. Schaefer KA, Wu W-H, Colgan DF, Tsang SH, Bassuk AG, Mahajan VB. Unexpected mutations after CRISPR–Cas9 editing in vivo. Nat Publ Gr. 2017;14:547–8.
36. Velappan N, Sblattero D, Chasteen L, Pavlik P, Bradbury ARM. Plasmid incompatibility: more compatible than previously thought? Protein Eng Des Sel. 2007;20:309–13.
37. Nicoloff HH, Andersson DI. Indirect resistance to several classes of antibiotics in cocultures with resistant bacteria expressing antibiotic-modifying or -degrading enzymes. J Antimicrob Chemother. 2016;71:100–10.
38. Lenski RE, Simpson SC, Nguyen TT. Genetic analysis of a plasmid-encoded, host genotype-specific enhancement of bacterial fitness. J Bacteriol. 1994;176:3140–7.
39. Li Y, Lin Z, Huang C, Zhang Y, Wang Z, Tang YJ, et al. Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 mediated genome editing. Metab Eng. 2015;31:13–21 **(Elsevier)**.
40. Martínez-García E, Calles B, Arévalo-Rodríguez M, de Lorenzo V. pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. BMC Microbiol. 2011;11:38.
41. Stearns T, Ma H, Botstein D. Manipulating yeast genome using plasmid vectors. Methods Enzymol. 1990;185:280–97.
42. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Front Microbiol. 2015;6:1–16.
43. Johnson TJ, Wannemuehler YM, Johnson SJ, Logue CM, White DG, Doetkott C, et al. Plasmid replicon typing of commensal and pathogenic *Escherichia coli* isolates. Appl Environ Microbiol. 2007;73:1976–83.
44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
45. Cavaleiro AM, Kim SH, Seppälä S, Nielsen MT, Nørholm MHH. Accurate DNA assembly and genome engineering with optimized uracil excision cloning. ACS Synth Biol. 2015;4:1042–6.
46. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6:343–5.
47. Datta S, Costantino N, Court DL. A set of recombineering plasmids for gram-negative bacteria. Gene. 2006;379:109–15.
48. Nour-Eldin HH, Hansen BG, Nørholm MHH, Jensen JK, Halkier BA. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res. 2006;34:e122.

Supplementary information for

A versatile one-step CRISPR-Cas9 based approach to plasmid-curing

Ida Lauritsen[1§], Andreas Porse[1§], Morten O.A. Sommer[1] and Morten H. H. Nørholm[1,*]

Supplementary info S1 – Addgene vector database

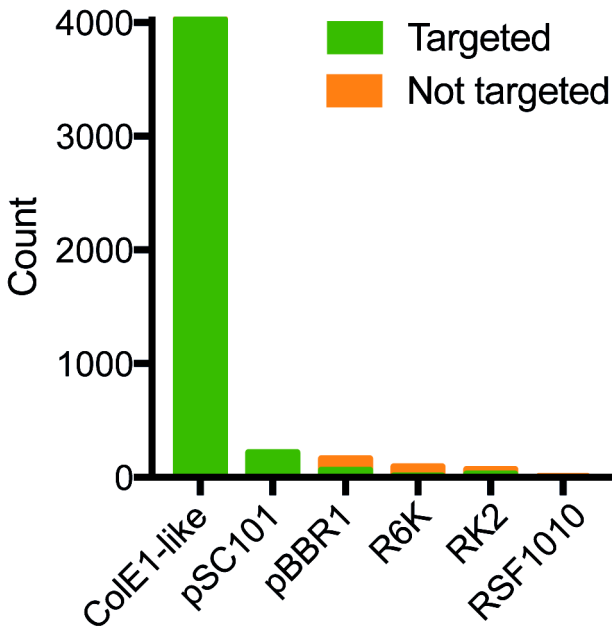
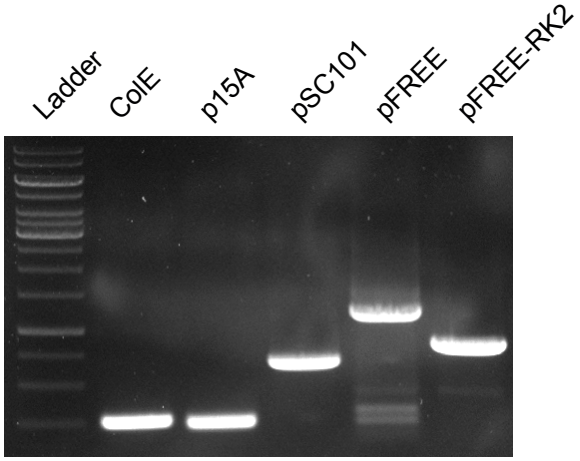


Fig. S1. Frequency of major replicons in bacterial cloning and expression vectors deposited to Addgene. A BLAST search was performed against all complete bacterial vector sequences (4657) in the Addgene database (Feb. 2017). ColE1-like plasmids include the closely related RNA-based replicons of ColE1, pBR322/pMB1, pUC18/19, pJET1.2®, colA and p15A. Green bars represent the ColE1 and pSC101 vector groups that are targeted by the pFREE system. The orange bars represent vector groups not targeted.

Supplementary info S2 - Replicon PCR



Agarose gel showing the results of a colony PCR with replicon specific oligonucleotides. The size of the PCR product identifies the specific replicon. For plasmids harboring ColE1 replicons, the PCR product will be 246 bp and slightly shorter (230 bp) for p15A, 600 bp for pSC101, 1125 bp for pFREE and 779 bp for pFREE-RK2.

Table S1. Oligonucleotides for replicon PCR

Oligo no.	Name	Sequence (5'→ 3')	PCR products
S1	ColE-like_Fw	CCGACAGGACTATAAAGATACC	ColE1 replicon: 246 bp p15A replicon: 230 bp pFREE with pSCFw: 1125 bp
S2	ColE-like_Rv	CTCAAGACGATAGTTACCGG	ColE1 replicon: 246 bp p15A replicon: 230 bp
S3	pSCFw	CCAGTATGTTCTCTAGTGTGG	pSC101 replicon: 600 bp
S4	pSCRv2	TGCCAAGTTCTCAAGCG	pSC101 replicon: 600 bp
S5	RK2Mul_Fw	GGATCGGATTCCACC	pFREE-RK2 with pSCFw: 779 bp

Table S2. PCR program for used for replicon PCR detection with Taq polymerase and colony templates.

Step	Duration	Temp. (°C)	Cycles
1	4 min.	99	
2	20 sec.	99	
3	20 sec.	55	
4	30 sec.	70	Go to step 2 x 29 times
5	5 min.	70	

Supplementary info S3 – Self-curing of pFREE

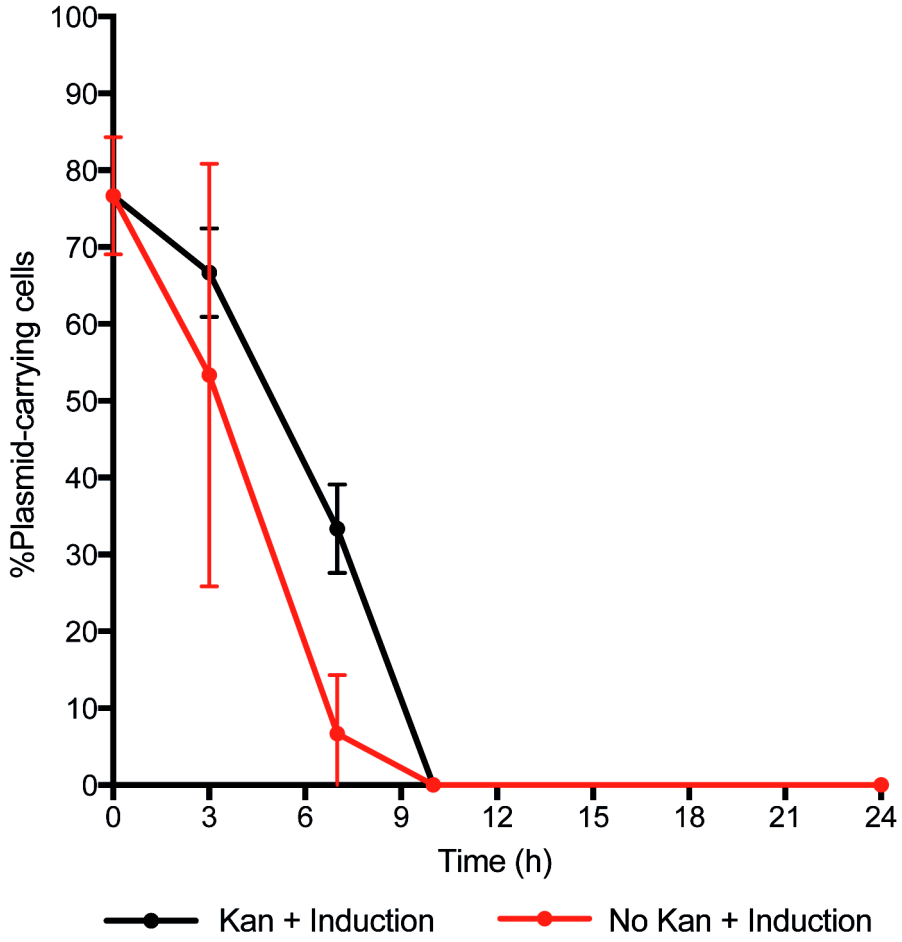


Fig. S2. Self-curing dynamics of the pFREE system with (black) and without 50 µg/mL kanamycin (red) added during curing. For each time-point, 20 CFU from each replicate was checked for growth on plates containing 50 µg/mL kanamycin. The percentage of plasmid-carrying cells is depicted. Data points represent mean value of three biological replicates with error-bars showing standard deviation.

Supplementary info S4 - pFREE-RK2

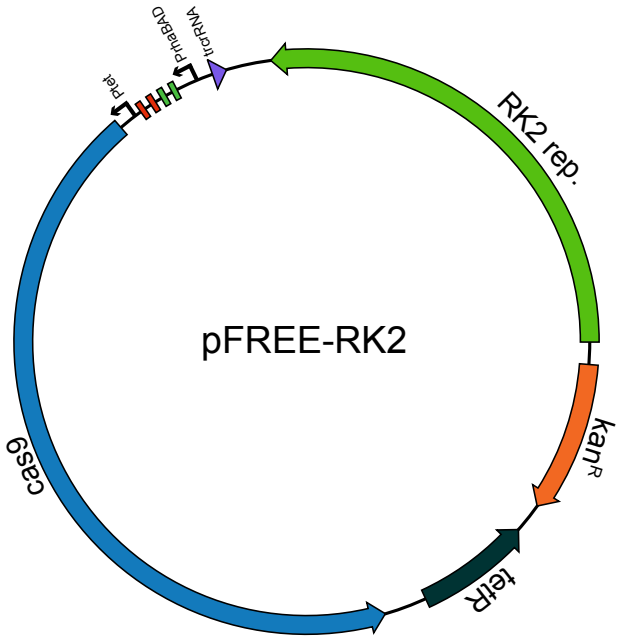


Fig. S3. Plasmid map of pFREE-RK2. pFREE-RK2 was constructed by changing the colA replicon to a temperature-sensitive, broad host range and low-copy RK2 replicon in the pFREE plasmid. The crArray, trans-activating CRISPR RNA (trcrRNA), the tetR repressor and Cas9 are expressed as in pFREE.

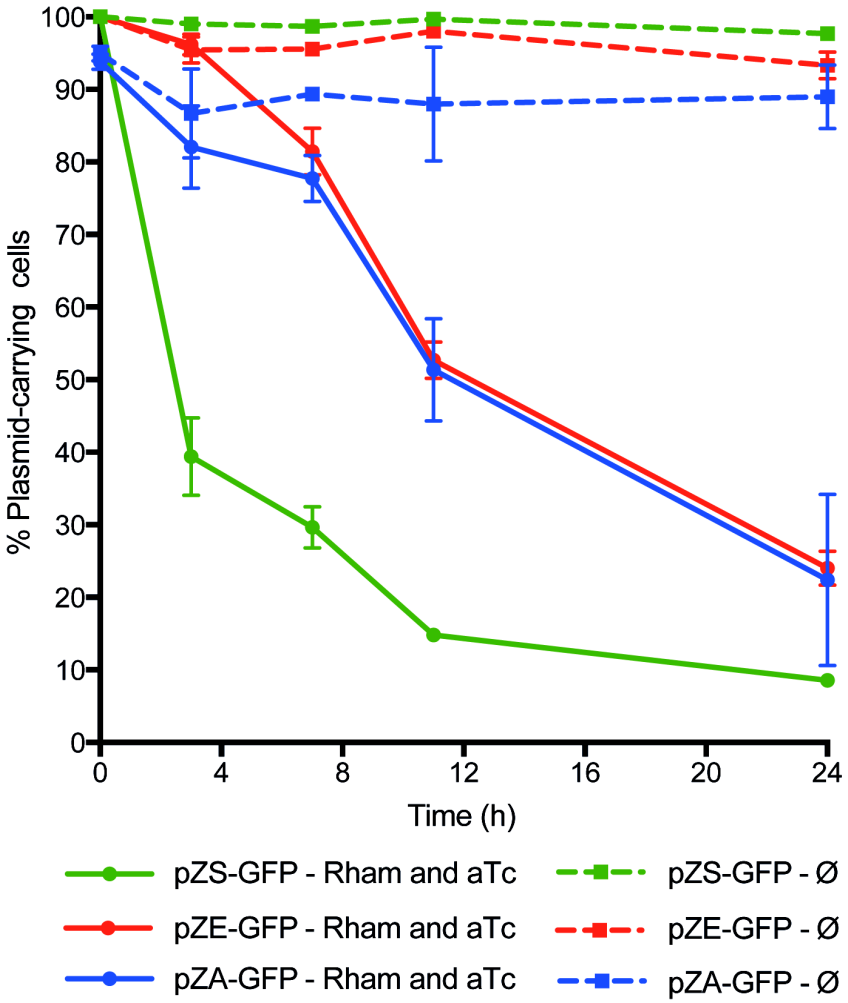


Fig. S4. Time course characterization of the pFREE-RK2 plasmid curing system Curing of pZ-plasmids expressing GFP with either pSC101 (pZS-GFP, green), ColE1 (pZE-GFP, red) or p15A (pZA-GFP, blue) replicon. The solid lines indicate induced cultures with rhamnose (Rham) and anhydrotetracycline (aTc), whereas the dashed lines refer to non-induced (Ø). Plating was performed at induction time (0) and 3, 7, 11 and 24 hours after induction. Between 100-150 CFUs were counted from each replicate and the ratio between fluorescent and non-fluorescent cells were determined. The percentage of plasmid-carrying cells is depicted. Data points represent mean value of three biological replicates with error-bars showing standard deviation.

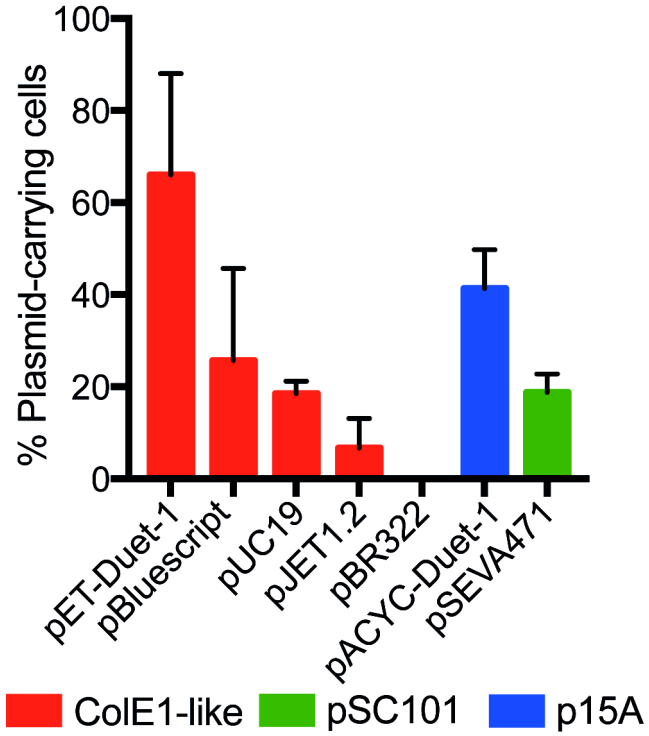


Fig. S5. pFREE–RK2 mediated curing of selected widely used cloning vectors with either ColE1-like (red), pSC101 (green) or p15A (blue) replicons. 50 CFUs from each replicate of each target plasmid was checked for antibiotic sensitivity of LB agar plates after 24 hours of induction of the pFREE-RK2 system. The percentage of plasmid-carrying cells is depicted. The pFREE-RK2 plasmid was cured in all colonies tested. Curing of the target plasmids was verified by replicon PCR. The bars represent mean value of three biological replicates with error-bars showing standard deviation.

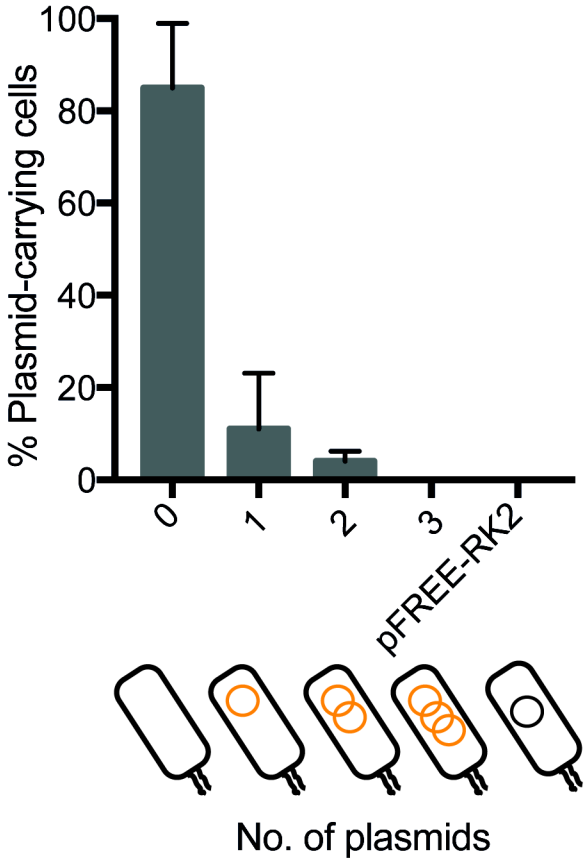


Fig. S6. Curing of multiple co-residing plasmids using the one-step transformation protocol depicted in Fig. 2c. The pFREE-RK2 plasmid was transformed into a strain harboring the same thre pZ plasmids as used in Fig. 3. After recovery, plasmid curing with the pFREE-RK2 system was induced and cultures were plated after 24 hours of induction. 50 CFUs from each replicate were checked for antibiotic sensitivity on LB agar plates. The percentage of cells carrying 0, 1, 2, 3 (orange plasmids) or pFREE-RK2 (black plasmid) is depicted. The bars represent mean value of three biological replicates with error-bars showing standard deviation.

Supplementary info S5 - Plasmids and oligonucleotides

Table S3. Plasmids used in this study

Oligo ID	Description	Reference/source
pMAZ-SK	CRISPR array and tracrRNA expression plasmid, CRISPR array under control of rhamnose-inducible promoter and tracrRNA constitutively expressed, Kan^R	[1] / Addgene #73962
pMA7CR_2.0	Cas9 expression plasmid, *Cas9* gene expressed under control of inducible aTc promoter, Amp^R	[1] / Addgene #73950
pZS4Int-tetR	Tetracycline repressor expression plasmid, constitutively expressed, $Spec^R$	EXPRESSSYS [1]
pFREE	Curing plasmid - aTc inducible expression of Cas9 and rhamnose inducible expression of CRISPR array. Derivative of pMAZ-SK with ColA replicon, Kan^R	This study
pZA-GFP	pZ backbone with p15A replicon, *gfp* constitutively expressed from BBa_J23110 promoter , Amp^R	This study
pZE-GFP	pZ backbone with ColE1 replicon, *gfp* constitutively expressed from BBa_J23110 promoter, Cm^R	This study
pZS-GFP	pZ backbone with pSC101 replicon, *gfp* constitutively expressed from BBa_J23110 promoter, Zeo^R	This study
pSIM9	RK2 temperature sensitive replicon, Cm^R	[2]
pFREE-RK2	Curing plasmid - aTc inducible expression of Cas9 and rhamnose inducible expression of CRISPR array. Derivative of pMAZ-SK with RK2 temperature sensitive replicon, Kan^R	This study
pUC19	High copy ColE1-like replicon, Amp^R	NEB
pBR322	ColE1-like (pMB1) replicon, Amp^R	NEB
pBluescript®	Very high copy ColE1-like replicon, Amp^R	Stratagene
pET-Duet-1	ColE1 (pMB1) replicon, Amp^R	Novogen
pJET1.2®	Very high copy ColE1-like replicon, Amp^R	ThermoFisher Scientific
pACYC-Duet-1	p15A replicon, Cm^R	Novogen
pSEVA471	pSC101 replicon, $Spec^R$	Addgene [3]
pSEVA441-GFP	pRO1600/ColE1 replicon, *gfp* expressed from Psal promoter, $Spec^R$	This study
pFREE_Amp	pFREE, Amp^R	This study
pFREE_Cm	pFREE, Cm^R	This study
pFREE_Zeo	pFREE, Zeo^R	This study
pFREE-RK2_Amp	pFREE-RK2, Amp^R	This study
pFREE-RK2_Cm	pFREE-RK2, Cm^R	This study
pFREE-RK2_Zeo	pFREE-RK2, Zeo^R	This study

Table S4. Oligonucleotides used in this study

Oligo no.	Name	Sequence (5'→ 3')
1	pMAZ-SK_backbone_fw	ATGCTGTUTTGAATGGTCCCAAAACaaaaaaaaacc
2	pMAZ-SK_backbone_rv_2	attcgcgaccutctcgttactgacagg
3	Killing_oligo_fw	tcaggcgctttttagactggtcgtGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACATGAACTAGCGATTAGTCGCTATGACTTAAGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACAACCACACTAGAGAACATACTGGCTAAATAGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGTTGGAC
4	Killing_oligo_rv	AGCTCTAAAACGGTATCTTTATAGTCCTGTCGGGTTTCGCCGTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACTATCCGGTAACTATCGTCTTGAGTCCAACCGTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACTATTTAGCCAGTATGTTCTCTAGTGTGG
5	gRNA_array_fw_2	aggtcgcgaautcaggcgctttttagactggtcgtg
6	gRNA_array_rev	AACAGCAUAGCTCTAAAACGGTATCTTTATAGTCCTGTCG
7	pMAZ_BB_tetR_Fw	aatgtgaaagtgggtcttaaGCAATAGACATAAGCGGC
8	pMAZ_BB_tetR_Rv	gtgtgacccgttggcgtcaaTACCGGTTTATTGACTACC
9	tetR_rbs_Fw	ttgacgccaacgggtcacacgggtacagtgctagcAGAAACATATCCATGAAATCC
10	tetR_Rv	TTAAGACCCACTTTCACATTTAAG
11	cas9_Gib_Fw	CCTAGGCTGCTGCCACCG
12	28_PtetRv	GTGCTCAGTATCTCTATCACTGATAGG
13	Cas9_Fw	AGAAGGAGATTATACATGGATAAGAAATACTCAATAGG
14	cas9_Gib_Rv:	CATGGGATCTCAGTCACC
15	RK2_pSIM_Fw	ATAAATATCUAACACCGTGCGTGTTGAC
16	RK2_pSIM_Rv	AGATGTGUATAAGAGACAGCTGTGCTG
17	pAI_woRep_Fw	ACACATCUCTTCAAATGTAGCACCTGAAGT

		CAG
18	pAI_woRep_Rv	AGATATTTAUTCTGCGCTAGCATGCCTATT TGT

REFERENCES

1. Ronda C, Pedersen LE, Sommer MOA, Nielsen AT. CRMAGE: CRISPR Optimized MAGE Recombineering. Sci. Rep.. Nature Publishing Group; 2016;6:19452.

2. Datta S, Costantino N, Court DL. A set of recombineering plasmids for gram-negative bacteria. Gene. 2006;379:109–15.

3. Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, De Las Heras A, et al. The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res. 2013;41:666–75.

Paper 4

Standardized Cloning and Curing of Plasmids

Running Head: Cloning and curing of plasmids

Standardized cloning and curing of plasmids

Ida Lauritsen[1], Se Hyeuk Kim[1], Andreas Porse[1] and Morten H. H. Nørholm[1,*]

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet B220, 2800 Lyngby, Denmark.

*Corresponding author. E-mail: morno@biosustain.dtu.dk

Phone: +45 217-99184; Fax: +45-353-33300

Published in:

J. Braham. (ed)
Synthetic Biology (2018)
Methods in Molecular Biology 1772, pp. 469-476. Humana Press, New York, NY

Abstract

Plasmids are highly useful tools for studying living cells and for heterologous expression of genes and pathways in cell factories. Standardized tools and operating procedures for handling such DNA vectors are core principles in synthetic biology. Here we describe protocols for molecular cloning and exchange of genetic parts in the Standard European Vectors Architecture (SEVA). Additionally, to facilitate rapid testing and iterative bioengineering using different vector designs, we provide a one-step protocol for a universal CRISPR-Cas9-based plasmid curing system (pFREE) and demonstrate the application of this system to cure SEVA constructs (all available at SEVA/Addgene).

Key words: SEVA, plasmid, DNA assembly, plasmid curing, pFREE, vector backbone exchange, cell factory design

1. Introduction

The historic importance of plasmids for the development of molecular biology cannot be overestimated and despite the recent progress in genome engineering technologies, plasmids continue to be the preferred choice for manipulating or introducing new and useful traits into living cells. A typical plasmid is composed of at least three genetic elements: 1) an origin of replication that determines the plasmid copy number and ensures that copies are available for transfer to other cells, 2) a selection marker such as an antibiotic resistance gene and 3) an additional *cargo* that introduces a specific genetic trait into the cell – typically a regulatory element controlling the expression of one or several genes ***(1)***. With today's synthetic DNA services and repositories at hand, such genetic elements are readily available at a low cost. However, context dependence and the robustness of different genetic elements varies greatly, and can significantly influence the performance of a cell factory ***(2, 3)***. For example, gene expression levels are influenced by the plasmid copy number, and the expression levels often needs to be balanced to obtain optimal production in specific settings and cellular context due to metabolic stress of e.g. overexpression ***(4)***. For construction of a desired plasmid, the first step involves DNA assembly, often followed by exchange of different genetic elements ***(5)***. Simple methods and standardized genetic elements that can facilitate such workflow are highly attractive when constructing biological systems and cell factories ***(6, 7)***. Importantly, such standards facilitate the comparison of data between different laboratories. An example

of a standardized vector system is the Standard European Vector Architecture (SEVA) collection that includes nine origins of replication, six antibiotic resistance markers and hosts different rare restriction sites suitable for exchange of genetic parts ***(8, 9)***.

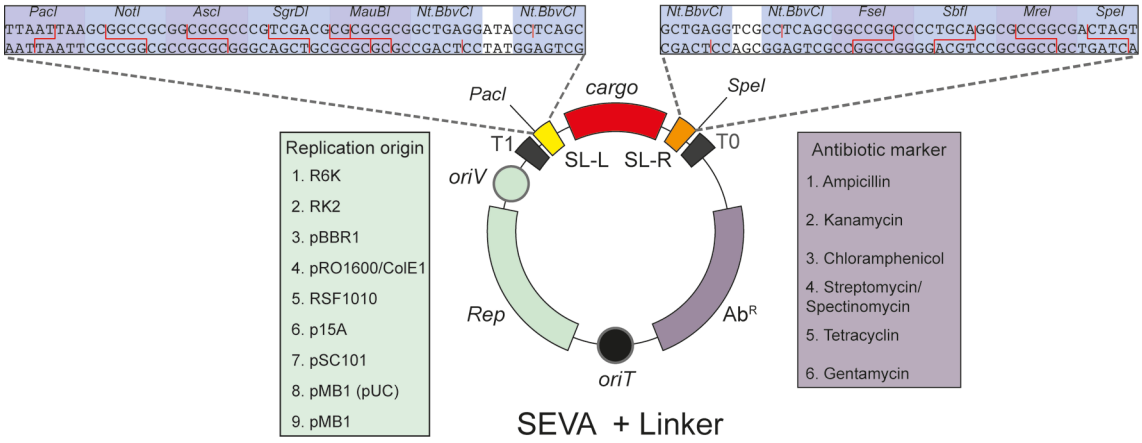


Fig. 1 Illustration of the Standard European Vector Architecture (SEVA) system adapted with the SEVA linkers. The SEVA linkers enable simple, one-pot backbone shuffling by introducing two multifunctional linker sequences (highlighted in yellow and orange) flanking the cargo in the SEVA system. The specific sequence and restriction sites are shown. Based on the SEVA system, the SEVA linker plasmids contain three basic components: origins of replication (Rep, green), antibiotic selection markers (Ab^R, purple) and the cargo (red). Any cargo that is flanked by SEVA linkers can be converted to the SEVA system by utilizing the *PacI* and *SpeI* restriction sites but also by a range of other molecular cloning technologies such as Gibson assembly and uracil excision cloning.

Previously, we designed two multifunctional SEVA linker sequences, adapted to the SEVA collection. These SEVA linkers flank the cargo sequence and introduce DNA assembly standards such as uracil excision cloning ***(10)***, Gibson assembly ***(11)*** as well as extra rare restriction sites to the SEVA ***(12)*** (**Fig. 1**). The SEVA linker sequences can be introduced into plasmids containing a gene or pathway of interest and also enable simple one-tube nicking-enzyme-based backbone exchange within the SEVA plasmid collection; a procedure we term SEVA cloning. 30 standardized SEVA-based plasmid backbones comprising five origins of replication and six antibiotic resistance markers carrying the toxic *ccdb* gene as cargo were constructed previously ***(12)***. By facilitating counter-selection of incorrectly assembled vectors, the toxic *ccdB* gene provides simple one-tube and one-step transfer of "donor" cargo to a

backbone "acceptor" simply by mixing two SEVA linker plasmids with the *Nt.BbvCI* nicking enzyme. The correctly assembled vector constructs can be selected on LB agar plates selective for the acceptor backbone (**Fig. 2**). We tested the performance of 30 SEVA backbones swapped this way for both a four-gene β-carotene biosynthetic pathway and the membrane protein NarK and more than 10-fold and 430-fold difference in production was observed, respectively ***(12)***.

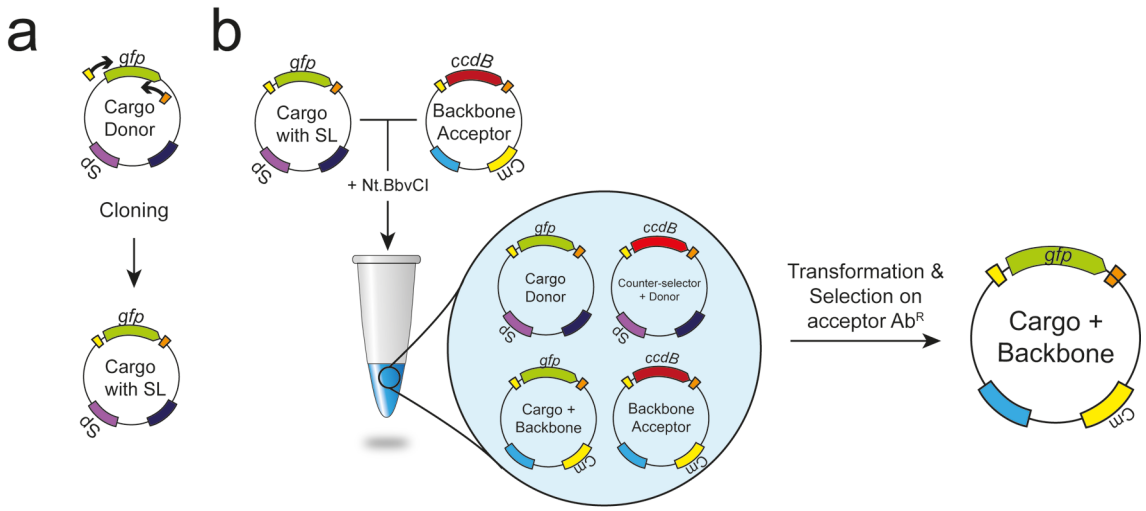


Fig. 2 Introduction of SEVA linkers and SEVA-linker-based backbone exchange (SEVA cloning). **(a)** The cargo, here illustrated as a plasmid-encoded *gfp*, from any cargo donor plasmid can be flanked by SEVA linkers by using state-of-the-art cloning techniques such as Gibson assembly, uracil excision cloning or traditional restriction cloning. **(b)** The two SEVA linkers host altogether nine rare restriction sites and four *Nt.BbvcI* nicking enzyme recognition sites (see in detail in **Fig. 1**). After mixing the SEVA linker flanked *gfp*-plasmid encoded cargo with a backbone acceptor plasmid containing the SEVA linker-flanked toxic *ccdB* gene and the *Nt.BbvcI* enzyme, the recombined cargo can be selected on the antibiotic resistance marker hosted in the acceptor backbone. The expanded view of the reaction shows all assembly possibilities.

Multiple plasmids are often utilized at the same time to manipulate living cells temporarily (e.g. when editing genomes with CRISPR/Cas9), but are unwanted on the long term because they constitute a metabolic burden or can cause off-target genomic mutations ***(13)***. Thus, when different plasmid designs are assessed or plasmid-based tools are employed throughout the cell factory design process, the complete removal of plasmids is necessary when new designs are tested or specific components are no longer needed. Some of the traditional approaches for plasmid curing involve stressing the cells with prolonged growth at high

temperatures, addition of DNA intercalating agents ***(14)*** or replicon-incompatibility ***(15, 16)***. However, these methods are often time demanding, allow the accumulation of unwanted mutations or act on specific replicons only ***(13, 17)***. Based on replicon abundance and sequence conservation analysis of bacterial cloning and expression vectors, we have developed a CRISPR-Cas9-based universal curing system (pFREE) that allows for targeting of all major plasmid replicons in molecular biology, including the SEVA vectors. With this system, we obtained curing efficiencies between 40-100% for the plasmids most widely used in molecular biology ***(18)*** and developed a one-step procedure for identification of plasmid-free clones within 24 hours (**Fig. 3a**). Here, we provide a one-step pFREE curing protocol for efficient curing of two sets of SEVA vectors with different origins of replicon (**Fig. 3b**). Due to the many similarities between different SEVA plasmids, e.g. the *oriT* for conjugation or the SEVA linkers, it should be possible to design a small set of pFREE plasmids for curing of any plasmid in the SEVA collection.

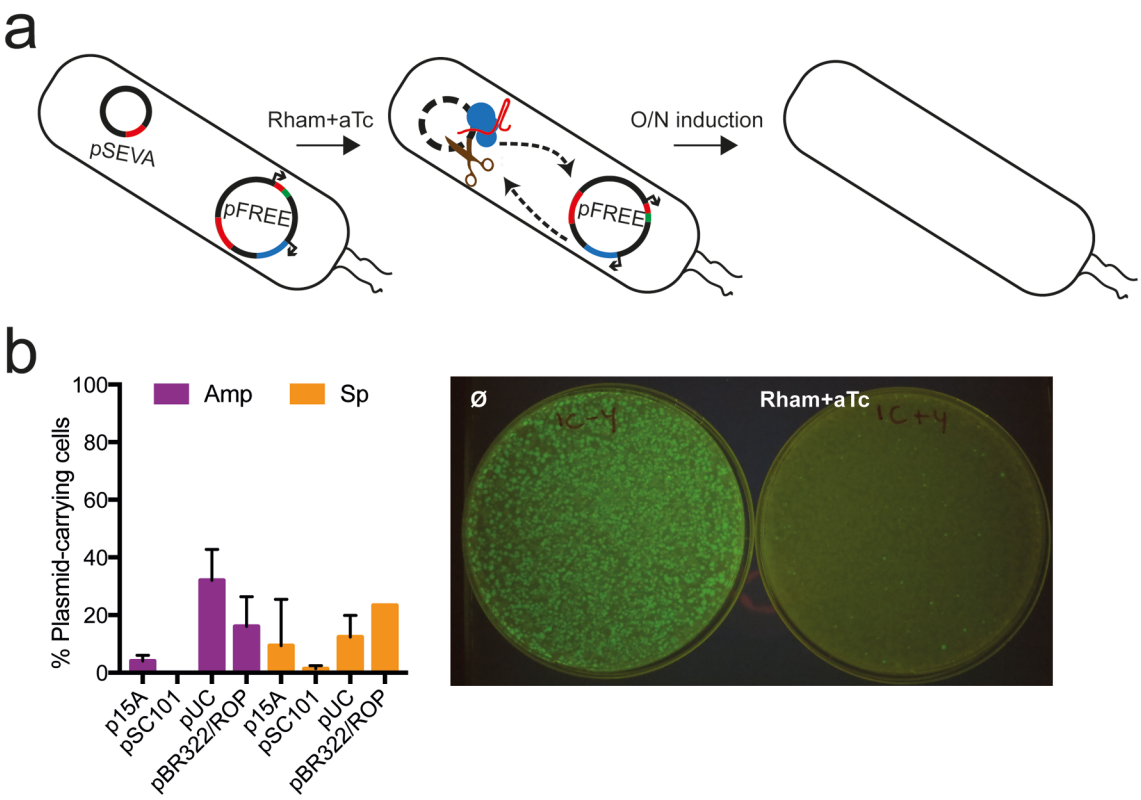


Fig. 3 Curing of SEVA linker vectors. **(a)** The pFREE plasmid is transformed into a strain harboring the target SEVA plasmid (pSEVA) for curing. The pFREE plasmid contains a guide RNA (gRNA) array targeting ColE1-like (red) and pSC101 (green) replicons and the Cas9 nuclease (blue). The addition of 0.2% rhamnose (Rham) and 200ng/mL anhydrotetracycline (aTc) expresses the gRNA array and Cas9 nuclease from pFREE, respectively. Due to a self-curing feature of pFREE, complete curing of the target SEVA and pFREE plasmid can be achieved after over night (O/N) induction. **(b)** pFREE-mediated curing of the pSEVA-T7-*narK-gfp* series with different replicons ***(12)***. The pSEVA vector either contained an ampicillin (purple) or spectinomycin resistance marker (orange). 50 colony-forming units from each replicate of each target plasmid were tested for antibiotic sensitivity after 24 hours of induction of the pFREE system. The pFREE plasmid was self-cured in all colonies tested. The data represents the mean value of three biological replicates with standard deviations. Representative LB agar plates for pSEVA-T7-*narK-gfp* carrying an ampillicin resistance marker and a p15A replicon. An equal number of cells were plated from non-induced (Ø, left) and induced cultures with rhamnose and aTc (right) of the pFREE system after 24 hours.

2. Materials

2.1 Media and pFREE inducers

1. Growth media: SOC (20 g Bacto-Tryptone, 5 g Yeast extract, 10 mM NaCl, 2.5 mM KCl, 20 mM MgSO4, 20 mM Glucose, water up to 1 L); 2×YT (16 g Bacto-Tryptone, 10 g Yeast extract, 5 g NaCl, water up to 1 L); LB: (10 g Bacto-Tryptone, 5 g Yeast extract, 10 g NaCl, water up to 1 L) all reagents can be purchased from Sigma-Aldrich (St. Louis, USA).
2. pFREE curing inducers: L-rhamnose (0.2% w/v), anhydro-tetracycline (aTc) (200 ng/mL) from Sigma Aldrich (St. Louis, MO, USA).

2.2 Molecular biology reagents

1. T4 DNA ligase, *DpnI* and other restriction enzymes from Thermo Fischer Scientific (Waltham, MA, USA).
2. *Nt.BbvCI* and USER enzyme from New England Biolabs (Ipswich, MA, USA).
3. 10 × CutSmart® buffer (New England Biolabs, Ipswich, MA, USA).
4. PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Life Technologies, Foster City, USA) for PCR purification.

2.3 Strains, plasmid vectors and oligonucleotides

1. Bacterial strains: *Escherichia coli* NEB5α (New England Biolabs, Ipswich, MA, USA) for propagation of plasmids, backbone swapping and cloning host; target strain carrying cargo SEVA linker vectors for curing.
2. Plasmid vectors for SEVA cloning: A series of pSEVA with counter-selector cargo, *ccdB*, flanked by SEVA linkers – can be obtained from the SEVA collection (http://seva.cnb.csic.es/).
3. Oligonucleotides for introduction of SEVA linkers into cargo and for amplification of pSEVA backbones.
4. pFREE plasmid for SEVA vector curing (available from Addgene).

3. Methods

*3.1 Introduction of SEVA linkers with uracil excision cloning (**Fig. 2**)*

1. Amplify your gene or pathway of interest with uracil containing forward and reverse oligonucleotides containing the SEVA linker sequence as shown below.
   Forward oligonucleotide: 5'-AGG ATA CCU CAG C-binding sequence-3'
   Reverse oligonucleotide: 5'-AGG CGA CCU CAG C-binding sequence-3'
2. Amplify your backbone with uracil containing forward and reverse oligonucleotides as shown below.
   Forward oligonucleotide: 5'- AGG TCG CCU CAG CGG CCG GCC CCT GCA GGC GCC GGC GAC TAG T-binding sequence-3'
   Reverse nucleotide: 5'- AGG TAT CCU CAG CCG CGC GCG CGT CGA CGG GCG CGC CGC GGC CGC TTA ATT AA-binding sequence-3'
3. *DpnI* treat PCR products for 1 hour at 37°C and purify of each PCR product according to manufacturer's protocol with PureLink™ Quick Gel Extraction and PCR Purification Combo Kit.
4. Clone the cargo PCR product flanked by SEVA linker sequences into the target plasmid vector by USER cloning as described in ***(19)***. Cloning is performed at 37°C for 15 min., at 25°C for 15 min. and 10 min. at 10°C.
5. Mix the USER reaction with 50 µL chemical NEB5α cells in a 1.5 mL Eppendorf tube, put on ice for 30 min. and then expose to heat-shock at 42°C for 60 seconds.
6. Cool the sample on ice for 2 min., add 500 µL of LB and incubate 1 hour at 37°C with shaking.
7. Plate samples on selective LB agar and screen the following day for positive colonies by PCR and sequencing.

3.2 Nicking enzyme mediated one-tube backbone exchange (SEVA cloning)

1. Quantify each plasmid, cargo "donor" and backbone "acceptor", containing SEVA linkers and add 0.06 pmol of each plasmid to a 10 or 20 µl total reaction volume (*see* **Note 1**).
2. Add 10 × CutSmart® buffer (New England Biolabs, Ipswich, MA, USA) and 5 units of *Nt.BbvCI* to mixture.
3. Keep mixture at 37 °C for 1 hour, followed by 15 min. at 25 °C and 10 min. at 10 °C.

4. Add 2.5 U of T4 DNA ligase and buffer (*see* **Note 2**), followed by incubation at room temperature for 15 min. and 5 min. on ice.
5. Transform chemical competent NEB5α cells according to standard transformation method ***(20)*** and plate on selective LB agar plates.

3.3 One-step curing of SEVA vectors

1. From an overnight culture of cells carrying SEVA linker construct(s) for curing, dilute it 100x fold in 5 mL LB.
2. Grow culture until OD_{600} of 0.3-0.4.
3. Make cells electro competent by three steps of washing in ice-cold MilliQ water.
4. Transform 50 µL of competent cells with 50 ng of pFREE by electroporation (1.65 kV, 200 Ohm, 25uF) and recover for 2 hours in 500 µL SOC medium at 30°C shaking at 500 rpm (*see* **Note 3**).
5. Transfer 50 µL of recovered cells to 10 mL LB medium supplied with 0.2% L-rhamnose, 200 ng/mL aTc and 50 µg/mL kanamycin to induce curing (*see* **Note 4**).
6. Plate cultures on non-selective LB agar plates after over night curing (*see* **Note 5**).
7. Screen for plasmid-cured cells by replicon PCR or antibiotic sensitivity (*see* **Note 6**).

4. Notes

1. Molar ratio between cargo "donor" plasmid and backbone "acceptor" plasmid can be varied and 5:1 ratio increases efficiency approximately 2-fold compared to a 1:1 ratio.
2. The addition of T4 DNA ligase can increase efficiency of backbone exchange up to 10-fold ***(12)***. Too long incubation time after addition of T4 DNA ligase can result in incorrect annealing of fragments.
3. After recovery of pFREE transformation, cells can be plated on LB Agar containing 50 µg/mL kanamycin to select for pFREE-carrying cells. Starting the curing procedure from a pFREE-carrying colony or culture has been observed to increase curing efficacy compared to simultaneous selection and curing of transformants.
4. For pFREE plasmids with other resistance markers than kanamycin (available at Addgene), a final antibiotic concentration of 100 µg/mL ampicillin, 34 µg/ml chloramphenicol or 100 µg/mL zeocin must be supplied during curing.

5. Efficiency of plasmid curing with the pFREE system can be increased by inducing cultures for 24 hours as shown in prior work ***(18)***.
6. Plasmid content can be profiled and cured cells identified with a universal replicon PCR assay ***(18)*** or by phenotypic screening e.g. antibiotic sensitivity or loss of fluorescence from *gfp*-expressing vectors.

References

1. T.K. Singha, P. Gulati, A. Mohanty, et al. (2017) Efficient genetic approaches for improvement of plasmid based expression of recombinant protein in Escherichia coli: A review, Process Biochemistry. 55, 17–31.
2. W.E. Bentley, N. Mirjalili, D.C. Andersen, et al. (1990) Plasmid-Encoded Protein : The Principal Factor in the " Metabolic Burden " Associated with Recombinant Bacteria, Biotechnology and bioengineering. 35, 668–681.
3. G.L. Rosano and E.A. Ceccarelli (2014) Recombinant protein expression in Escherichia coli: Advances and challenges, Frontiers in Microbiology. 5, 1–17.
4. R. Grabherr, E. Nilsson, G. Striedner, et al. (2002) Stabilizing plasmid copy number to improve recombinant protein production, Biotechnology and Bioengineering. 77, 142–147.
5. R.E. Cobb, J.C. Ning, and H. Zhao (2014) DNA assembly techniques for next-generation combinatorial biosynthesis of natural products, Journal of Industrial Microbiology and Biotechnology. 41, 469–477.
6. A. Casini, M. Storch, G.S. Baldwin, et al. (2015) Bricks and blueprints: methods and standards for DNA assembly, Nature Reviews Molecular Cell Biology. 16, 568–576.
7. C. Vilanova, K. Tanner, P. Dorado-Morales, et al. (2015) Standards not that standard, Journal of Biological Engineering. 9, 17.
8. E. Martínez-Garćía, T. Aparicio, A. Goñi-Moreno, et al. (2015) SEVA 2.0: An update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities, Nucleic Acids Research. 43, D1183–D1189.
9. R. Silva-Rocha, E. Martínez-García, B. Calles, et al. (2013) The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes, Nucleic Acids Research. 41, 666–675.

10. H.H. Nour-Eldin, B.G. Hansen, M.H.H. Nørholm, et al. (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments, Nucleic Acids Research. 34, e122.
11. D.G. Gibson, L. Young, R.-Y. Chuang, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases., Nature methods. 6, 343–5.
12. S.H. Kim, A.M. Cavaleiro, M. Rennig, et al. (2016) SEVA Linkers: A Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology, ACS Synthetic Biology. 5, 1177–1181.
13. S. Schlegel, P. Genevaux, and J.W. de Gier (2015) De-convoluting the Genetic Adaptations of E.coli C41(DE3) in Real Time Reveals How Alleviating Protein Production Stress Improves Yields, Cell Reports. 10, 1758–1766.
14. J. Trevors (1986) Plasmid curing in bacteria, FEMS Microbiology Letters. 32, 149–157.
15. L. Hale, O. Lazos, A.S. Haines, et al. (2010) An efficient stress-free strategy to displace stable bacterial plasmids, BioTechniques. 48, 223–228.
16. M. Kamruzzaman, S. Shoma, C.M. Thomas, et al. (2017) Plasmid interference for curing antibiotic resistance plasmids in vivo, Plos One. 12, e0172913.
17. R. Crameri, J.E. Davies, and R. Hütter (1986) Plasmid curing and generation of mutations induced with ethidium bromide in streptomycetes., Journal of general microbiology. 132, 819–824.
18. I. Lauritsen, A. Porse, M.O.A. Sommer, et al. (2017) A versatile one-step CRISPR-Cas9 based approach to plasmid-curing,.
19. A.M. Cavaleiro, S.H. Kim, S. Seppälä, et al. (2015) Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning, ACS Synthetic Biology. 4, 1042–1046.
20. J. Sambrook, E.F. Fritsch, and T. Maniatis (1989), Molecular cloning: a laboratory manual.

Paper 5

Mutations in the Global Transcription Factor CRP/CAP: Insights from Experimental Evolution and Deep Sequencing

Computational and Structural Biotechnology Journal 17 (2019) 730–736

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini Review

Mutations in the Global Transcription Factor CRP/CAP: Insights from Experimental Evolution and Deep Sequencing

Pernille Ott Frendorf [a,1], Ida Lauritsen [a,1], Agnieszka Sekowska [b], Antoine Danchin [c], Morten H.H. Nørholm [a,*]

[a] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet B220, DK-2800 Kgs. Lyngby, Denmark
[b] Institut de Cardiométabolisme et Nutrition, CHU Pitié-Salpêtrière, 47 boulevard de l'Hôpital, 75013 Paris, France
[c] Institut Cochin, INSERM U1016, CNRS UMR8104, Université Paris Descartes, 24 rue du Faubourg Saint-Jacques, 75014 Paris, France

ARTICLE INFO

Article history:
Received 10 April 2019
Received in revised form 22 May 2019
Accepted 26 May 2019
Available online 18 June 2019

Keywords:
cAMP receptor protein
Carbon catabolite repression
Global transcriptional regulation
Experimental evolution

ABSTRACT

The *Escherichia coli* cyclic AMP receptor protein (CRP or catabolite activator protein, CAP) provides a textbook example of bacterial transcriptional regulation and is one of the best studied transcription factors in biology. For almost five decades a large number of mutants, evolved *in vivo* or engineered *in vitro*, have shed light on the molecular structure and mechanism of CRP. Here, we review previous work, providing an overview of studies describing the isolation of CRP mutants. Furthermore, we present new data on deep sequencing of different bacterial populations that have evolved under selective pressure that strongly favors mutations in the *crp* locus. Our new approach identifies more than 100 new CRP mutations and paves the way for a deeper understanding of this fascinating bacterial master regulator.

Contents

1. Introduction

Studies of CRP date back to the earliest days of molecular biology, shortly after the model for negative *lac* gene regulation was presented [25]. In the decade following Jacob and Monod's groundbreaking discoveries, several reports on positive gene regulation were published including catabolite activation by CRP [14,15,69]. CRP is mostly known for its global regulatory role in carbon catabolism in the model bacterium *Escherichia coli (E. coli)*: In the absence of readily metabolized carbon sources such as glucose, the enzyme adenylate cyclase is activated, producing cyclic AMP (cAMP) from ATP. cAMP binds and activates CRP, increasing the affinity for DNA, which in many cases activate operons involved in the utilization of alternative carbon sources such as lactose and maltose. However, CRP can also repress gene expression and has been shown to regulate hundreds of genes in the *E. coli* genome, earning it the status of "global" or "master" regulator (Fig. 1) [31]. In fact, beyond the many specific binding sites experimentally validated in *E. coli*, CRP exhibits unspecific DNA binding affinity and together these observations points towards a role more akin to that of a nucleoid-associated protein involved in the organisation of the bacterial chromosome [55].

The first experimentally solved three-dimensional structure of CRP bound to cAMP was published in 1981 [32], but for many years the absence of an experimentally determined structure of apo-CRP hindered the understanding of the conformational changes that occur upon cAMP binding [18,27]. Presently, only one crystal structure [62] and one NMR structure [39] of wildtype apoCRP have been published. Both structures indicate that large structural rearrangements take place for DNA binding to occur. These include reorientation of the

* Corresponding author.
E-mail address: morno@biosustain.dtu.dk (M.H.H. Nørholm).
[1] These authors contributed equally.

https://doi.org/10.1016/j.csbj.2019.05.009

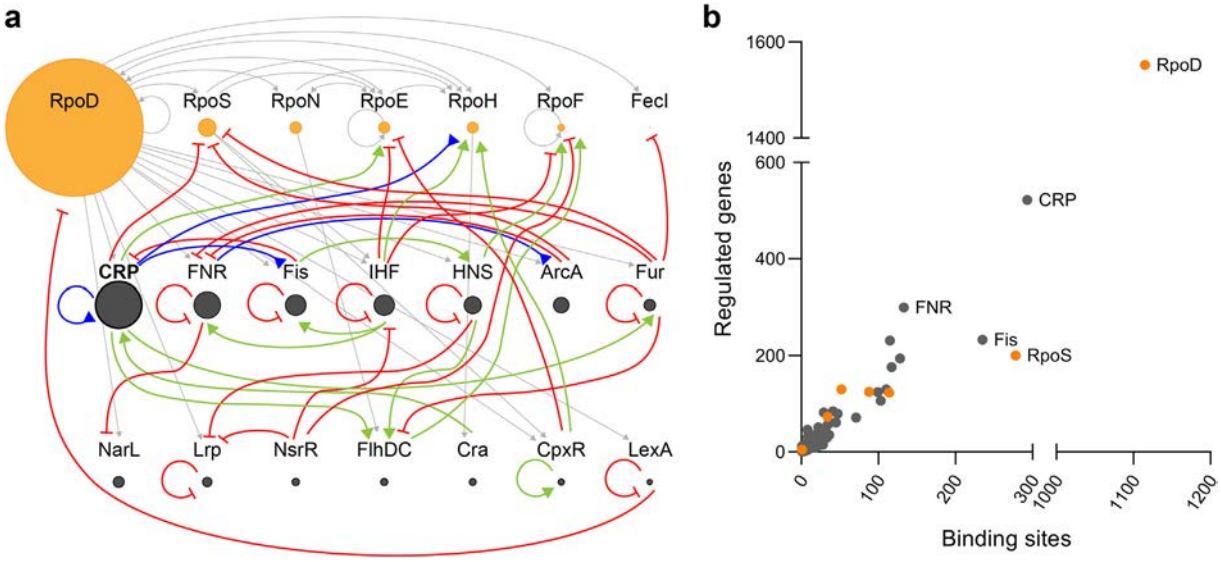


Fig. 1. Network of global transcription factors and sigma factors in *Escherichia coli*. Illustrated is a) The sigma factors (orange circles) and the 14 biggest regulators by regulon size (grey circles). The size of the circles is directly proportional to the sigmulon or regulon size by number of genes directly affected (RegulonDB 04-03-2019) [19]. Arrows designate regulation of regulator expression; sigma factor transcription (grey arrows), activation (green, arrowheads), repression (red, perpendicular line ends) or dual regulation (blue, reverse arrowheads) [26]. b) The sigma factors (orange circles) and 207 regulators (grey circles) of *E. coli* plotted by the number of promoters or binding sites recognized, respectively, and the number of genes directly affected (RegulonDB 04-03-2019) [19].

DNA-binding domain and stabilization of the backbone helix, but the two different structures do not agree on the orientation of the C-terminal domain. This observation, and the limited number of published apo CRP structures, suggests that apoCRP may be unstable due to flexibility of the C-terminal domain [44].

CRP is a 45 kDa homodimer (Fig. 2), with each monomer consisting of 209 amino acids in two separate domains. The larger N-terminal domain (residues 1–138) binds the allosteric effector cAMP in the *anti*-conformation (residues 71, 72, 82, 83, 127 and 128) with reported binding constants in the range of 1–28 μM [2,13,14,22,30,56]. The C-terminal domain (residues 139–209) houses the DNA-recognition helix (residues 181–193) as part of a helix-turn-helix (HTH) motif [6]. The N- and C-terminal domains are connected by a hinge region (residues 135–138). In the absence of ligand, CRP exists in a closed conformation (Fig. 2, left), where the HTH motif is secluded inside the C-terminal domain. When cAMP binds the main binding site, allosteric change stabilizes an open complex (Fig. 2, right), resulting in the HTH motif protruding from the surface of the protein, thereby enabling DNA binding. A secondary effector binding site binds cAMP in its *syn*-conformation (residues 58, 135, 180) with binding constants in the millimolar range likely to be of limited physiological relevance [30,38]. In the presence of excess cAMP, both cAMP binding sites in one monomer are occupied, resulting in a lower DNA-binding affinity [38].

When CRP binds, the DNA is bent 90°, which likely significantly affects protein-protein and protein-DNA interactions in promoter regions [17,28,47,61] and several studies suggest that CRP is involved both in recruiting the RNA polymerase and in post-recruitment regulation [7,29,35,41]. In the case of transcriptional activation, regulation is mediated by direct interactions with the RNA polymerase holoenzyme through so-called activating regions (ARs). AR1 (residues 156–164) interacts with the αCTD domain [36,46,63–66], AR2 (residues 19, 21, 96 and 101) with the αNTD domain [35,58] and AR3 (residues 52–55 and 58) with the sigma factor [4,41,42,58–60].

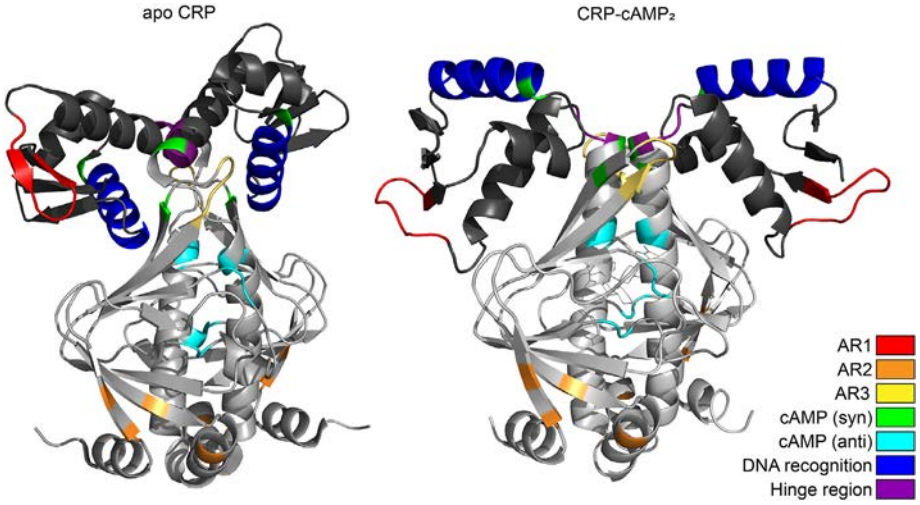


Fig. 2. Illustration of functional domains in a closed and an open conformation of CRP. The locations of functionally important CRP domains in the tertiary structure of apo-CRP (left) and CRP-cAMP$_2$ (right). The protein structures are from Protein Data Bank entries 3FWE (apo-CRP) and 1ZRC (CRP-cAMP$_2$) and were modified using the PyMOL Molecular Graphics System, Version 2.3 Schrödinger, LLC.

Complementary to structural and biochemical studies, evolved and rationally engineered mutations are imperative when elucidating protein structure and function. In this respect, CRP is probably one of the best studied transcription factors. For several decades, mutant generation has been applied to study the physiological role and molecular mechanism of CRP [37,45]. A comprehensive list of mutations identified in CRP was previously assembled [37], but to our knowledge the different approaches towards obtaining these CRP mutants have not been reviewed previously. In light of recent high throughput sequencing (HTS) approaches applied to *crp* mutants [48], we here aim at presenting a short overview on the more than 40 years of exploration of the CRP mutational space and in addition provide complementary new data on deep sequencing of experimentally evolved *crp*, obtained in aging bacterial colonies. An overview of all CRP mutants reviewed here is shown in Tables S1 and S2.

2. Evolved and Engineered Mutations in *crp*

The outcome of an evolution experiment is dependent on the mutational space available under the given experimental conditions, the genotype of the organism and the selection pressure applied. Evolution of CRP has been pursued by inducing mutagenesis using *e.g.* UV radiation or chemical reagents [33,45] or by directly targeting the *crp* gene with error-prone PCR [67], but several studies have also relied on spontaneously arising mutants [9,43,48,50]. In most studies, the CRP variants were generated in a cAMP-deficient production strain (Δ*cya*) and screened for mutations enabling fermentation of a carbon source such as lactose or maltose. Such cAMP suppressor mutations were termed *csm* [33] and CRP variants called CRP* [45] or CRP^{i} [5].

In the 1970's, early after the discovery of CRP, researchers began isolating *csm,* CRP* [5,12,45,51,54] and defective *crp*- mutants [3] mostly under conditions of induced mutagenesis. However, the exact molecular nature of these mutants was unknown for some time due to the lack of DNA sequencing and amplification technologies.

In 1985, Aiba and colleagues published a paper where they had exposed a plasmid-borne *crp* gene to UV-radiation and selected for lactose utilization in a Δ*cya* strain. The obtained *crp* mutations in the isolated strains caused amino acid substitutions in positions 53, 62, 141, 142 and 148 of the CRP protein [1]. It was noted that positions 53 and 62 were in vicinity of the cAMP binding site, but that the phenotypes differed in that only CRP D53H was activated by the alternative cyclic nucleotide cGMP. The role of position 53 being located in the AR3 region, possible interacting with sigma factors, was not discussed in this work as AR3 was unknown at this time. Amino acids 141, 142 and 148 are part of the D-α-helix and it was speculated that they were critical in the allosteric transition, from the N-terminal domain to the DNA binding C-terminal domain, normally caused by binding of cAMP.

Around the same time, Garges and Adhya used *crp*-carrying phages for infection and growth in a mutator *E. coli* strain for CRP mutant generation [21]. Phages carrying mutagenized *crp* variants were isolated as positive lactose utilizing plaques in an Δ*cya* background. The detected mutations were in positions 72, engaged in cAMP binding, and again in D-α-helix residues 141, 142 and 144.

The following year, Harman and colleagues sequenced and characterized three CRP* mutants that previously were selected by different methods [23,33,40,45]. The CRP mutation A144T in the D-α-helix was again identified – in this case from a Δ*cya* strain selected on xylose as carbon source. A T127I mutation in the cAMP binding site was identified in combination with Q170K from a strain that complemented a CRP binding site mutation designated L8 in the *lac* promoter [45]. This double mutant showed a CRP* phenotype and was activated by cGMP, although the physiological relevance of the latter was questioned by Harman and colleagues. The individual effects of the 127 and 170 mutations were not explored further in this work. Finally, the mutation L195R was evolved in the Δ*cya*, *crp T127I, Q170K* mutant background and the extra mutation enabled growth on arabinose in the absence of cAMP [40]. The authors suggested that the increased positive charge of this L195R mutation in the DNA binding domain caused an increase in the affinity for DNA.

In a follow-up study by Garges and Adhya, CRP* suppressor mutants, causing a loss of the G141S and A144T CRP* phenotypes on lactose, were identified as T127A and R169C/E171G, respectively [20]. In case of the G141S CRP* mutant, it is perhaps not surprising that a mutation near the cAMP binding (T127A) can neutralize cAMP independence. Similarly, it was noted that amino acids near the two mutations in positions 169 and 171, identified in the CRP A144T mutant background, were previously suggested to interact with the amino acid Y63 near the cAMP binding site, but also could be in direct interaction with the DNA [57].

Mutations in positions 141, 141 and 144 again occurred in a study that described the selection of CRP* mutants based on growth on lactose in a Δ*cya* mutant background [53]. In this case, two different mutations in position 144, A144T and A144E, were found in combination with T28K. These two combinations were found to be toxic when expressed on a multicopy plasmid, whereas two other D-helix mutations T140K and G141D were tolerated in high copy.

Another broad category of CRP mutations, more generally termed *positive control mutants (pcm)*, were selected for their inability to induce transcription while retaining binding to specific CRP DNA binding sites. The first attempt at creating CRP *pc* mutants introduced the mutations E171Q, E171K and Q170K based on similarity to the lambda repressor, but these caused different effects at different promoters [4,24]. A more clear *pcm* phenotype was observed with the mutation CRP H159L in AR1 and second site revertants was identified as K52N and K52Q in AR3 [4]. Eschenlauer and Reznikoff screened for CRP mutants that repressed the *gal* promoter but had lost their ability to induce the lac promoter [16]. This way, they identified mutations in cAMP binding position 72, and in position 162 in the AR1 region. In a similar study by Zhou and co-workers, plasmid-harbored *crp* genes were mutagenized by error-prone PCR and screened in an engineered Δ*cya* strain for defective ribose fermentation, while retaining the ability to repress a modified *lac* promoter [66]. The identified mutations were in positions 156, 158, 159, and 162 that are all part of AR1. Finally, Niu and co-workers identified mutations in the AR1 region that could not activate transcription of Class I and II CRP-dependent promoters as well as the mutations H19L, H19Y, H21L and K101E in the AR2 region that were only defective in Class II promoter activation [35].

Experimental evolution of CRP has also addressed the complex interplay between CRP and the CytR transcription factor. A CytR-repressed *tsx* promoter construct was screened in combination with a mutagenized *crp* plasmid library for CRP variants that were dominantly activating the promoter. This approach identified mutations in positions 17, 18, 108 and 110 in CRP – all in the vicinity of AR2 [52].

Recent work has again explored CRP mutants that evolve spontaneously under different selection regimes. Sievert and co-workers observed that by growing an *E. coli* strain in high levels of the (CRP-dependent) carbon source xylose, the CRP G141D mutation again was found to evolve, promoting increased xylose utilization and growth rate [50]. In adaptive laboratory evolution for improved fitness in minimal medium supplied with lactate, the CRP mutations L150Q and I165T evolved [9]. These are both located near AR1 and presumably cause a changed interaction with the RNA polymerase.

In summary, adaptive mutations identified in the previous four decades of CRP studies (Fig. 3) occur predominantly in the cAMP binding site, the D-α-helix, and in the RNA polymerase activating domains AR1 and AR2. These three categories are intuitively easy to understand as they likely either directly affect ligand binding, ligand-induced allosteric transitions, or the productive interaction with the core RNA polymerase, respectively. Mutations around

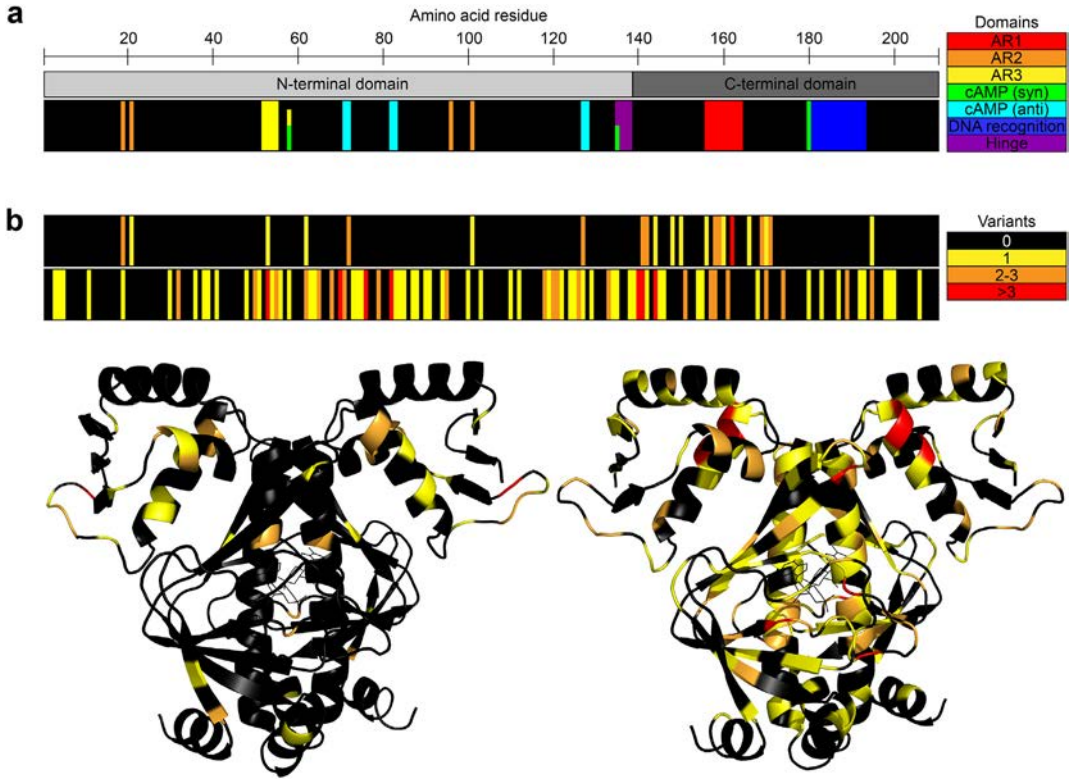


Fig. 3. Illustration of adaptive mutations identified in CRP a) The locations of functionally important CRP domains in the primary amino acid sequence. b) Locations of mutations identified in the primary and tertiary structure of CRP in previous studies (upper black bar and structure to the left) and in our laboratories by deep sequencing (lower black bar and structure to the right). The number of variants identified per residue is colour coded (black: 0 variants, yellow: 1 variant, orange: 2–3 variants, and red: more than three variants).

position 170 are also frequently observed but are more difficult to interpret and have been discussed both to directly affect interactions with cAMP and the DNA.

To supplement *crp* mutants that evolve under selective pressure, the advent of PCR enabled hypothesis-driven site-directed mutagenesis exploring the role of specific amino acid residues in the different domains of the CRP protein. As summarized in Table S2, these are often studied in areas where mutations evolve naturally (the cAMP-*anti* binding pocket, the D-α-helix and the activating regions AR1 and AR2), but have also explored mutations in *e.g.* the hinge region that connects the C- and D-helices, AR3, and the DNA binding domain in the C-terminus.

In an experiment designed to study adaptation when bacteria age and starve over two months, our recent work identified a large number of CRP mutants. By plating a Δ*cya E. coli* strain on MacConkey agar supplied with the CRP-dependent carbon source maltose, mutant red "*papillae*" - or "secondary colonies" - appeared that were able to use the extra carbon source. 96 mutants were selected for genome sequencing based on their different CRP* phenotypes and their temporal appearance. In addition, an additional approximately 500 mutants had their *crp* loci sequenced by PCR amplification and Sanger sequencing [48]. 35 different missense mutations were identified in *crp*. Seven of the identified mutations, S62F, T127I, G141D, G141S, A144T, A144E and L195R were previously observed in adaptive evolution studies and have all been classified as CRP*. Other *crp* mutations have not been identified elsewhere, including P110Q, L134M, T140P/R/K, A144K, G162S and M189K. The CRP mutant S62Y has not been identified before. However, the substitution of serine to phenylalanine (S62F) has been observed by induced mutagenesis and screening on lactose previously [1]. The Q170K mutation was always observed in combination with an additional *crp* mutation such as T140R, A144T/E or M198K [48]. Interestingly, a similar trend was observed in the study by Harman and coworkers in which the Q170K mutation was paired with T127I or T127I and L195R [23].

3. Deep Sequencing of CRP in Aging Bacterial Populations

We reasoned that we could obtain deeper insights into the mutational space of CRP by growing a large number of different bacterial colonies, followed by deep sequencing of the *crp* locus. To this end, we followed the same workflow as previously described [48], but now isolated DNA from more than 500 colonies per plate at different time points, followed by PCR amplification of *crp* and HTS.

The new data presented here identifies more than 100 new *crp* missense mutations (Figs. 3–4, Table S3), although we only observed a significant increase in different mutations towards the end of the 35 days. The HTS approach provides a previously unmatched look at the mutational landscape of CRP, revealing novel insights into the evolutionary response of a strain during a selective event. However, a tradeoff for the sequencing depth is its inability to distinguish between amplification of a single mutation and multiple occurrences of the same mutation. Thus, mutations detected by this HTS approach can only be roughly categorized into those appearing in higher frequencies (likely due to a clear fitness advantage leading to a dominating population) or those that appear in low frequencies. Similar, we cannot distinguish between mutations that occur alone and those that only occur in combination with other mutations such as described above for the mutation Q170K.

The data presented in Figs. 3 and 4 is generally well aligned with previous observations. Several high-frequency mutations are observed in the residues D53, S62, T140, G141, A144, Q170, and relatively few mutations are seen in the C-terminus of CRP. In addition, DNA was isolated from aging bacteria in the presence of two CRP ligands: 0.1 mM cAMP and 0.5 mM cGMP, and with parental strains that already contained the frequently observed CRP* mutants A144T and A144E (Fig. 4). From plates supplemented with the two cyclic nucleotides, binding site mutations again dominate together with mutations in the D-α-helix. Interestingly, although the A144T/E mutations observed by

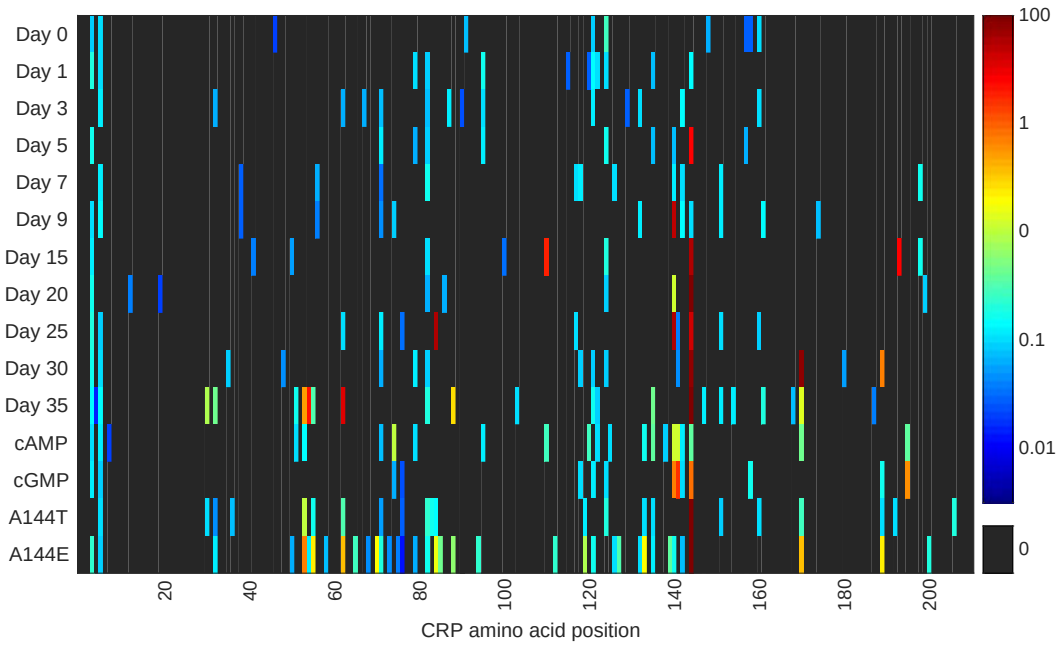


Fig. 4. Heatmap of the mutation frequencies (%) of each CRP residue in an *E. coli* Δ*cyaA* pTIG strain. Strain and evolution experiment set-up were as previously described [48]. Biomass samples of the populations were taken from preculture (day 0) or from one of 20 identical plates during the evolution experiment (days 1–35). Furthermore, samples were taken at day 7 of the Δ*cyaA* strain supplemented with 0.1 mM cAMP or 0.5 mM cGMP, and of *E. coli* Δ*cyaA* strains with CRP A144T or CRP A144E mutations. The *crp* gene was sequenced by amplicon next-generation sequencing and the data processed by CLC Genomics Workbench (Qiagen, Aarhus, Denmark).

Sekowska *et al.* occur in the same residue, they have distinct effects on the mutational landscape of CRP. The A144T mutation causes almost uniformly distributed low frequency mutations, while the A144E mutation causes increased mutagenesis generally concentrated around residues 50–90 (AR3 and the cAMP binding site) and 110–150 (the cAMP binding site and inter-domain stability).

To our knowledge, the data presented here is the most comprehensive overview of the natural mutational landscape of CRP to date. Our approach provides a detailed map of CRP mutations for future in-depth characterization. With the increasingly affordable deep sequencing methodologies, the approach could be generalized to study similar evolutionary tracks in transcriptional regulation. How can we try and extend our knowledge of structural and regulatory features of the CRP protein?

Besides exploring changes in physical growth conditions (altering temperature or osmolarity) coupling with other master regulatory systems should be rewarding. Global regulators such as CRP manage coordination of gene expression in a variety of conditions, that are also coordinated by other regulatory molecules, in particular the alarmone ppGpp. This “magic spot” has been discovered half a century ago as involved in monitoring amino acid availability [8]. Yet its role is far from fully understood and still a matter of considerable research. It is now known that altered levels of this regulatory molecule in *relA spoT* mutants – coding for enzymes controlling the synthesis and turnover of the molecule – resulted in non-optimal resource allocation in *E. coli* [68]. Interestingly, this happened under conditions where it is expected that CRP is involved in the management of ppGpp-mediated effects [49] and CRP-mediated contribution to *relA* expression has been demonstrated [34]. This regulation must match the coupling between the cAMP-CRP regulation and amino acid biosynthesis.

As a case in point indeed, it has long been known that there is an explicit link between ppGpp synthesis and serine/one carbon metabolism. Serine excess resulted in growth inhibition of *relA* mutants, while *relA cya*-defective or *crp*-defective mutants became resistant to excess serine. In *relA cya* strains, sensitivity to serine was restored when the growth medium was supplemented with cAMP, substantiating the serine-mediated interference in the cAMP-CRP control of gene expression [10]. To be sure, this effect was reverted in a *crp** background. It was therefore interesting to isolate secondary mutants that would again be resistant to excess serine in order to better understand how CRP was involved in this regulation. A new class of CRP mutants was identified in *E. coli cya relA crp** strains. These mutants were mapped in the *crp* gene, and their physiological features differed from both the wild type *crp* and the *crp** allele [11]. However, they could not be studied more in-depth at the time. Exploring this selection procedure with the “omics” techniques that are now familiar should allow us to enter a new evolution landscape of the protein.

Similar approaches could be developed to study other global regulators. In general, letting genes that are expressed under stationary conditions evolve should bring about new observations in the unchartered territory of adaptive mutations.

Acknowledgements

We would like to thank the three anonymous reviewers for their positive, insightful and constructive comments. This work was supported by the Novo Nordisk Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2019.05.009.

References

[1] Aiba H, Nakamura T, Mitani H, Mori H. Mutations that alter the allosteric nature of cAMP receptor protein of *Escherichia coli*. EMBO J 1985;4:3329–32. https://doi.org/10.1002/j.1460-2075.1985.tb04084.x.
[2] Anderson WB, Schneider AB, Emmer M, Perlman RL, Pastan IRA. Purification of and properties of the cyclic adenosine monophosphate receptor protein which mediates cyclic gene transcription in *Escherichia coli*. J Biol Chem 1971;246:5929–37.
[3] Artman M, Werthamer S. Effects of cyclic guanosine 3′,5'-monophosphate on the synthesis of enzymes sensitive to catabolite repression in intact cells of *Escherichia coli*. J Bacteriol 1974;120:980–3.
[4] Bell A, Gaston K, Williams R, Chapman K, Kolb A, Buc H, et al. Mutations that alter the ability of the *Escherichia coli* cyclic AMP receptor protein to activate transcription. Nucleic Acids Res 1990;18:7243–50. https://doi.org/10.1093/nar/18.24.7243.

[5] Botsford JL, Drexler M. The cyclic 3′,5′-adenosine monophosphate receptor protein and regulation of cyclic 3′,5′-adenosine monophosphate synthesis in *Escherichia coli*. MGG Mol Gen Genet 1978;165:47–56. https://doi.org/10.1007/BF00270375.
[6] Brennan RG. Interactions of the helix-turn-helix binding domain. Curr Opin Struct Biol 1991. https://doi.org/10.1016/0959-440x(91)90015-l.
[7] Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). J Mol Biol 1999;293:199–213. https://doi.org/10.1006/jmbi.1999.3161.
[8] Cashel M, Gallant J. Two compounds implicated in the function of RC gene. Nature 1969;221:838–41.
[9] Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, et al. Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol 2009;10:1–12. https://doi.org/10.1186/gb-2009-10-10-r118.
[10] Daniel J, Danchin A. Involvement of cyclic AMP and its receptor protein in the sensitivity of *Escherichia coli* K 12 toward serine. Mol Gen Genet MGG 1979;176:343–50. https://doi.org/10.1007/BF00333096.
[11] Daniel J, Joseph E, Ullmann A, Danchin A. CrpX mutants of *Escherichia coli* K-12: selection and physiological properties. FEMS Microbiol Lett 1981;10:389–93. https://doi.org/10.1111/j.1574-6968.1981.tb06277.x.
[12] Dessein A, Schwartz M, Ullmann A. Catabolite repression in *Escherichia coli* mutants lacking cyclic AMP. MGG Mol Gen Genet 1978;162:83–7. https://doi.org/10.1007/BF00333853.
[13] Donoso-Pardo JL, Turner PC, King RW. Cyclic nucleotide binding to cAMP receptor protein from *Escherichia coli* optical and ligand-binding studies. Eur J Biochem 1987. https://doi.org/10.1111/j.1432-1033.1987.tb13470.x.
[14] Emmer M, deCrombrugghe B, Pastan I, Perlman R. Cyclic AMP receptor protein of *E. coli*: its role in the synthesis of inducible enzymes. Proc Natl Acad Sci 1970;66:480–7. https://doi.org/10.1073/pnas.66.2.480.
[15] Englesberg E, Irr J, Power J, Lee N. Positive control of enzyme synthesis by gene C in the L-arabinose system. J Bacteriol 1965;90:946–57.
[16] Eschenlauer AC, Reznikoff WS. *Escherichia coli* catabolite gene activator protein mutants defective in positive control of lac operon transcription. J Bacteriol 1991;173:5024–9. https://doi.org/10.1128/jb.173.16.5024-5029.1991.
[17] Fried MG, Crothers DM. CAP and RNA polymerase interactions with the lac promoter: binding stoichiometry and long range effects. Nucleic Acids Res 1983. https://doi.org/10.1093/nar/11.1.141.
[18] Gallagher DT, Smith N, Kim SK, Robinson H, Reddy PT. Profound asymmetry in the structure of the cAMP-free cAMP receptor protein (CRP) from *Mycobacterium tuberculosis*. J Biol Chem 2009. https://doi.org/10.1074/jbc.C800215200.
[19] Gama-Castro S, Salgado H, Santos-Zavaleta A, Ledezma-Tejeida D, Muñiz-Rascado L, García-Sotelo JS, et al. RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. Nucleic Acids Res 2016. https://doi.org/10.1093/nar/gkv1156.
[20] Garges S, Adhya S. Cyclic AMP-induced conformational change of cyclic AMP receptor protein (CRP): intragenic suppressors of cyclic AMP-independent CRP mutations. J Bacteriol 1988;170:1417–22. https://doi.org/10.1128/jb.170.4.1417-1422.1988.
[21] Garges S, Adhya S. Sites of allosteric shift in the structure of the cyclic AMP receptor protein. Cell 1985;41:745–51. https://doi.org/10.1016/S0092-8674(85)80055-6.
[22] Green J, Stapleton MR, Smith LJ, Artymiuk PJ, Kahramanoglou C, Hunt DM, et al. Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches. Curr Opin Microbiol 2014;18:1–7. https://doi.org/10.1016/j.mib.2014.01.003.
[23] Harman JG, McKenney K, Peterkofsky A. Structure-function analysis of three cAMP-independent forms of the cAMP receptor protein. J Biol Chem 1986;261:16332–9 [doi:3023348].
[24] Irwin N, Ptashne M. Mutants of the catabolite activator protein of *Escherichia coli* that are specifically deficient in the gene-activation function. Proc Natl Acad Sci 1987;84:8315–9. https://doi.org/10.1073/pnas.84.23.8315.
[25] Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 1961;3:318–56. https://doi.org/10.1016/S0022-2836(61)80072-7.
[26] Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, et al. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res 2017. https://doi.org/10.1093/nar/gkw1003.
[27] Kumar P, Joshi DC, Akif M, Akhter Y, Hasnain SE, Mande SC. Mapping conformational transitions in cyclic AMP receptor protein: crystal structure and normal-mode analysis of *Mycobacterium tuberculosis* apo-cAMP receptor protein. Biophys J 2010. https://doi.org/10.1016/j.bpj.2009.10.016.
[28] Liu-Johnson HN, Gartenberg MR, Crothers DM. The DNA binding domain and bending angle of *E. coli* CAP protein. Cell 1986;47:995–1005. https://doi.org/10.1016/0092-8674(86)90814-7.
[29] Liu B, Hong C, Huang RK, Yu Z. Structural basis of bacterial transcription activation. Science 2017;951:947–51 (80-.).
[30] Małecki J, Polit A, Wasylewski Z. Kinetic studies of cAMP-induced allosteric changes in cyclic AMP receptor protein from *Escherichia coli*. J Biol Chem 2000;275:8480–6. https://doi.org/10.1074/jbc.275.12.8480.
[31] Martínez-Antonio A, Collado-Vides J. Identifying global regulators in transcriptional regulatory networks in bacteria. Curr Opin Microbiol 2003;6:482–9. https://doi.org/10.1016/j.mib.2003.09.002.
[32] McKay DB, Steitz TA. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. Nature 1981;290:744–9. https://doi.org/10.1038/290744a0.
[33] Melton T, Snow LL, Freitag CS, Dobrogosz WJ. Isolation and characterization of cAMP suppressor mutants of *Escherichia coli* K12. Mol Gen Genet 1981;182:480–9.
[34] Nakagawa A, Oshima T, Mori H. Identification and characterization of a second, inducible promoter of relA in *Escherichia coli*. Genes Genet Syst 2006;81:299–310. https://doi.org/10.1266/ggs.81.299.
[35] Niu W, Kim Y, Tau G, Heyduk T, Ebright RH. Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. Cell 1996;87:1123–34. https://doi.org/10.1016/S0092-8674(00)81806-1.
[36] Niu W, Zhou Y, Dong Q, Ebright YW, Ebright RH. Characterization of the activating region of *Escherichia coli* catabolite gene activator protein (CAP) I. saturation and alanine-scanning mutagenesis. J Mol Biol 1994. https://doi.org/10.1016/0022-2836(94)90034-5.
[37] Passner JM, Schultz SC, Steitz TA. Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 AÅ resolution. J Mol Biol 2000;304:847–59. https://doi.org/10.1006/jmbi.2000.4231.
[38] Passner JM, Steitz TA. The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. Proc Natl Acad Sci U S A 1997;94:2843–7. https://doi.org/10.1073/pnas.94.7.2843.
[39] Popovych N, Tzeng S-R, Tonelli M, Ebright RH, Kalodimos CG. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. Proc Natl Acad Sci 2009. https://doi.org/10.1073/pnas.0900595106.
[40] Puskas R, Fredd N, Gazdar C, Peterkofsky A. Methylglyoxal-mediated growth inhibition in an *Escherichia coli* cAMP receptor protein mutant. Arch Biochem Biophys 1983;223:503–13.
[41] Rhodius VA, Busby SJW. Interactions between activating region 3 of the *Escherichia coli* cyclic AMP receptor protein and region 4 of the RNA polymerase σ70subunit: application of suppression genetics. J Mol Biol 2000:311–24. https://doi.org/10.1006/jmbi.2000.3737.
[42] Rhodius VA, Busby SJW. Transcription activation by the *Escherichia coli* cyclic AMP receptor protein: determinants within activating region 3. J Mol Biol 2000;299:295–310. https://doi.org/10.1006/jmbi.2000.3736.
[43] Sabourin D, Beckwith J. Deletion of the *Escherichia coli* crp gene. J Bacteriol 1975;122:338–40.
[44] Saha A, Mukhopadhyay J, Datta AB, Parrack P. Revisiting the mechanism of activation of cyclic AMP receptor protein (CRP) by cAMP in *Escherichia coli*: lessons from a subunit-crosslinked form of CRP. FEBS Lett 2015. https://doi.org/10.1016/j.febslet.2014.12.021.
[45] Sanders R, McGeoch D. A mutant transcription factor that is activated by 3′:5′-cyclic guanosine monophosphat. Proc Natl Acad Sci 1973;70:1017–21.
[46] Savery NJ, Lloyd GS, Busby SJW, Thomas MS, Ebright RH, Gourse RL. Determinants of the C-terminal domain of the *Escherichia coli* RNA polymerase α subunit important for transcription at class I cyclic AMP receptor protein-dependent promoters. J Bacteriol 2002. https://doi.org/10.1128/JB.184.8.2273-2280.2002.
[47] Schultz SC, Shields GC, Steitz TA. Crystal-structure of a cap-DNA complex - the DNA is bent by 90-degrees. Science 1991;253:1001–7 (80-.).
[48] Sekowska A, Wendel S, Fischer EC, Nørholm MHH, Danchin A. Generation of mutation hotspots in ageing bacterial colonies. Sci Rep 2016;6:2. https://doi.org/10.1038/s41598-016-0005-4.
[49] Shimizu K. Metabolic regulation and coordination of the metabolism in bacteria in response to a variety of growth conditions. In: Ye Q, Bao J, Zhong J-J, editors. Bioreactor engineering research and industrial applications I: cell factories. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016. p. 1–54. https://doi.org/10.1007/10_2015_320.
[50] Sievert C, Nieves LM, Panyon LA, Loeffler T, Morris C, Cartwright RA, et al. Experimental evolution reveals an effective avenue to release catabolite repression via mutations in XylR. Proc Natl Acad Sci 2017;114:7349–54. https://doi.org/10.1073/pnas.1700345114.
[51] Silverstone AE, Goman M, Scaife JG. ALT: a new factor involved in the synthesis of RNA by *Escherichia coli*. MGG Mol Gen Genet 1972;118:223–34. https://doi.org/10.1007/BF00333459.
[52] Søgaard-Andersen L, Mironov AS, Pedersen H, Sukhodelets VV, Valentin-Hansen P. Single amino acid substitutions in the cAMP receptor protein specifically abolish regulation by the CytR repressor in *Escherichia coli*. Proc Natl Acad Sci 1991;88:4921–5. https://doi.org/10.1073/pnas.88.11.4921.
[53] Tagami H, Inada T, Kunimura T, Aiba H. Glucose lowers CRP*levels resulting in repression of the lac operon in cells lacking cAMP. Mol Microbiol 1995;17:251–8. https://doi.org/10.1111/j.1365-2958.1995.mmi_17020251.x.
[54] Takebe Y, Shibuya M, Science M. A new Extragenic suppressor of cya mutation. J Biochem 1978;1623:1615–23.
[55] Visweswariah SS, Busby SJW. Evolution of bacterial transcription factors: how proteins take on new tasks, but do not always stop doing the old ones. Trends Microbiol 2015;23:463–7. https://doi.org/10.1016/j.tim.2015.04.009.
[56] Weber IT, Steitz TA. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. J Mol Biol 1987;198:311–26. https://doi.org/10.1016/0022-2836(87)90315-9.
[57] Weber IT, Steitz TA. Model of specific complex between catabolite gene activator protein and B-DNA suggested by electrostatic complementarity. Proc Natl Acad Sci 1984;81:3973–7. https://doi.org/10.1073/pnas.81.13.3973.
[58] West D, Williams R, Rhodius V, Bell A, Sharma N, Zou C, et al. Interactions between the *Escherichia coli* cyclic AMP receptor protein and RNA polymerase at class II promoters. Mol Microbiol 1993;10:789–97. https://doi.org/10.1111/j.1365-2958.1993.tb00949.x.
[59] Williams R, Bell A, Sims G, Busby S. The role of two surface exposed loops in tranction activation by the *Escherichia Coli* CRP and FNR proteins. Nucleic Acids Res 1991. https://doi.org/10.1093/nar/19.24.6705.
[60] Williams RM, Rhodius VA, Bell AI, Kolb A, Busby SJW. Orientation of functional activating regions in the *Escherichia coli* CRP protein during transcription activation at class II promoters. Nucleic Acids Res 1996;24:1112–8. https://doi.org/10.1093/nar/24.6.1112.
[61] Wu HM, Crothers DM. The locus of sequence-directed and protein-induced DNA bending. Nature 1984. https://doi.org/10.1038/308509a0.

[62] Yu S, Steitz TA, Sharma H, Wang J, Kong J. Structure of apo-CAP reveals that large conformational changes are necessary for DNA binding. Proc Natl Acad Sci U S A 2009;106:16604–9. https://doi.org/10.1073/pnas.0908380106.
[63] Zhou Y. The functional subunit of a dimeric transcription activator protein depends on promoter architecture. Trends Genet 1995. https://doi.org/10.1016/0168-9525(95)90593-6.
[64] Zhou Y, Busby S, Ebright R. Identification of the functional subunit of a dimeric transcription activator protein by use of oriented heterodimers. Cell 1993;73:375–9. https://doi.org/10.1016/0092-8674(93)90236-J.
[65] Zhou Y, Merkel TJ, Ebright RH. Characterization of the activating region of *Escherichia coli* catabolite gene activator protein (CAP) II. Role at class I and class II CAP-dependent promoters. J Mol Biol 1994. https://doi.org/10.1016/0022-2836(94)90035-3.
[66] Zhou Y, Zhang X, Ebright RH. Identification of the activating region of catabolite gene activator protein (CAP): isolation and characterization of mutants of CAP specifically defective in transcription activation. Proc Natl Acad Sci U S A 1993;90:6081–5.
[67] Zhou Y, Zhang X, Ebright RH. Random mutagenesis of gene-sized DNA molecules by use of CRP with Taq DNA polymerase. Nucleic Acids Res 1991;19:6052. https://doi.org/10.1093/nar/19.21.6052.
[68] Zhu M, Dai X. Growth suppression by altered (p)ppGpp levels results from non-optimal resource allocation in *Escherichia coli*. Nucleic Acids Res 2019;47:4684–93. https://doi.org/10.1093/nar/gkz211.
[69] Zubay G, Schwartz D, Beckwith J. Mechanism of activation of catabolite-sensitive genes: a positive control system. Proc Natl Acad Sci 1970;66:104–10. https://doi.org/10.1101/sqb.1970.035.01.056.

Supplementary information

Materials and Methods

Bacterial strains. The strains applied in this study were *Escherichia coli* K12 MG1655 *ΔcyaA::cat* strains carrying either WT CRP and the pTrc*tig* plasmid (expressing the Tig chaperone), the CRP A144T mutation or the CRP A144E mutation (Sekowska et al., 2016). The strains were cultivated at 37°C in either LB medium (routine experiments) or MacConkey (Macconkey, 1905) medium with 1% maltose (evolution experiments) supplemented with appropriate antibiotics.

Adaptive evolution experiment. The strains in question were streaked on LB plates, from which single colonies were used to inoculate precultures (5 mL LB) for the adaptive evolution experiment. Precultures were grown for exactly 7 hours before being diluted to OD_{600}=1 in LB and then to $OD_{600}=10^{-4}$ in 0.9% NaCl to ensure equal nutrients present in all samples. From the dilutions, 100 µL were plated on MacConkey maltose plates, supplemented if necessary with 0.1 mM cAMP or 0.5 mM cGMP, and the plates were incubated in a box with water to ensure constant humidity throughout the experiment.

Biomass samples. For the time series of the WT CRP background, biomass samples of the populations were taken from preculture (day 0) or from one of 20 identical plates of the adaptive evolution experiment. Furthermore, biomass samples were taken at day 7 of the WT CRP background supplemented with cAMP or cGMP, as well as of the CRP A144T and CRP A144E backgrounds, using 10 identical plates for each condition. The biomass samples were obtained by scraping off all biomass of plates using liquid LB medium and a drigalski spatula. The dissolved biomass was centrifuged and the pellet saved at -80°C. Genomic DNA was purified from the biomass samples using the DNeasy Blood & Tissue Kit (Qiagen).

Amplicon next-generation sequencing of crp. Three amplicons of 293, 282 and 338 bp, respectively, were designed to cover the entire *crp* gene, and were amplified using the adapter oligonucleotides shown in Table X. The amplicons were cleaned up using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) and eluted in 10 mM Tris-HCl pH 8.5. The library preparation was completed using the KAPA HyperPlus Kit (Roche), and the amplicons were sequenced on the MiSeq system (Illumina) using the MiSeq V3 Reagent Kit with 600 cycles (Illumina). The data was demultiplexed in the system before download.

Processing of amplicon sequencing data. The data processing was completed using CLC Genomics Workbench (CLC bio). FastQ files were imported into the workbench as paired reads. The paired reads were then merged by overlapping regions, and non-mergeable reads were discarded. The reads were them trimmed by quality scores (discarding those with p>0.001), ambiguous nucleotides (not allowed), adapters (forward and reverse complement, in case of amplification errors) and length (based on the minimum amplicon length without adapters). Trimmed reads were then mapped to a *crp* reference sequence using global alignment with standard alignment settings (match score = 1, mismatch cost = 2, linear gap cost = 3, length fraction = 0.5, similarity fraction = 0.8). Following this, the reads were examined for structural variants (quality p < 0.0001 and a

maximum number of mismatches = 3 for end breakpoints), and this information was applied on the mapping for local realignment (two iterations, maximum variant length = 200 bp). Nucleotide variants were called and annotated by the software using low-frequency variant detection (no minimum variant count, neighbourhood quality filter = $p<0.0003$ in a 5 base radius, no set ploidy).

The adapter oligonucleotides applied for amplicon generation. Bold = *crp* annealing.

Amplicon	Length (bp)	Oligonucleotides
1	293	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CGTTATCTGGCTCTGGAGAAA**
		GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GTTCCTGGCCCTCTTCAAAC**
2	282	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GGGTGATTTTATTGGCGAAC**
		GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GGGTGAGTCATAGCGTCTGG**
3	338	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**AGAGAAAGTGGGCAACCTG**
		GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GCATAGTTGATATCGGGGTGA**

Table S1. *In vivo* selected CRP mutants.

Position	Mutation	Additional mutation (s)	Strain genotype (+ or Δ for *cya*, *crp)*	*crp* location	Type of mutagenesis	Selection carbon source	Reference(s)
L11	I	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
H17	R		*Δcya, Δcrp*	Plasmid			(Søgaard-Andersen et al., 1991)
C18	R		*Δcya, Δcrp*	Plasmid			(Søgaard-Andersen et al., 1991)
H19	L, Y		*cya+, crpΔ*	Plasmid	PCR	Lactose	(Niu et al., 1996)
H21	L		*cya+, crpΔ*	Plasmid	PCR	Lactose	(Niu et al., 1996)
T28	K	A144T, E	*Δcya, crp+*	Genome		Lactose	(Tagami et al., 1995)
I30	F	A144T, E	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
L39	M	A144T, E	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
D53	H		*Δcya, crp+*	Plasmid	UV	Lactose	(Aiba et al., 1985)
	N	A145T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
E55	D	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	E	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	K	A144T, E	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
S62	F		*Δcya, crp+*	Plasmid	UV	Lactose	(Aiba et al., 1985)
	F	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	Y		*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	Y	A144T, E	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
Y63	F	M189K	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
L64	Q	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
E72	A		*Δcya, crp+*	Plasmid	UV	Lactose	(Garges and Adhya, 1985)
	V		*cya+, crpΔ*	Plasmid	UV	Lactose+ Galactose	(Eschenlauer and Reznikoff, 1991)
R82	S	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
A84	E	A144T, E	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
W85	R	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
K101	E		*cya+, crpΔ*	Plasmid	PCR	Lactose	(Niu et al., 1996)
V108	A		*Δcya, Δcrp*	Plasmid			(Søgaard-Andersen et al., 1991)
P110	Q		*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)

	S		*Δcya, Δcrp*	Plasmid			(Søgaard-Andersen et al., 1991)
	Q	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
Q119	H	T140R, A144E	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
T127	A	G141S	*Δcya, Δcrp*	Plasmid	UV	Lactose	(Garges and Adhya, 1988)
	I	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	I	Q170K	*Δcya, crp+*	Genome	Chemical + UV	Lactose	(Harman et al., 1986; Puskas et al., 1983; Sanders and McGeoch, 1973)
	I	Q170K + L195R	*Δcya, crp+*	Genome	Chemical + UV	Arabinose	(Harman et al., 1986; Puskas et al., 1983)
N133	H	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
L134	M		*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
T140	P, R, K		*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	K		*Δcya, crp+*	Plasmid		Lactose	(Tagami et al., 1995)
	R	Q119H, Q170K, V183A	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	K	A144T, E, M198R	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
G141	D		*Δcya, crp+*	Plasmid	UV	Lactose	(Aiba et al., 1985)
	D		*Δcya, crp+*	Plasmid		Lactose	(Tagami et al., 1995)
	D		*cya+, crp+*	Genome		Xylose	(Sievert et al., 2017)
	D,S		*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	S		*Δcya, crp+*	Plasmid	UV	Lactose	(Garges and Adhya, 1985)
	S	T127A	*Δcya, Δcrp*	Plasmid	UV	Lactose	(Garges and Adhya, 1988)
R142	C		*Δcya, crp+*	Plasmid	UV	Lactose	(Aiba et al., 1985)
	H		*Δcya, crp+*	Plasmid	UV	Lactose	(Garges and Adhya, 1985)
A144	T		*Δcya, crp+*	Plasmid	UV	Lactose	(Garges and Adhya, 1985)
	T		*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	T		*Δcya, crp+*	Genome	Chemical	Xylose	(Harman et al., 1986; Melton et al., 1981)
	T	T28K	*Δcya, crp+*	Genome		Lactose	(Tagami et al., 1995)

	T	R169C + E171G	*Δcya, Δcrp*	Plasmid	UV	Lactose	(Garges and Adhya, 1988)
	T	L11I, I30F, L39M, D53N, E55D,K, E, S62F,Y, L64Q, R82S, A84E, W85R, P110Q, T127I, N133H, T140K, Q145K, T146A, D155N, Q170K, M189I, M189K, Q193K	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	E, K		*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	E	T28K	*Δcya, crp+*	Genome		Lactose	(Tagami et al., 1995)
	E	I30F, L39M, E55K, S62Y, A84E, Q119H, T140K, Q170K, Q174K, M189K	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
Q145	K	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
T146	A	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
L148	R		*Δcya, crp+*	Plasmid	UV	Lactose	(Aiba et al., 1985)
L150	Q		*cya+, crp+*	Genome		Lactate	(Conrad et al., 2009)
D155	N	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
A156	D		*cya+, crpΔ*	Plasmid	PCR	Lactose+ Ribose	(Zhou et al., 1993)
T158	A, I		*cya+, crpΔ*	Plasmid	PCR	Lactose+ Ribose; Lactose	(Niu et al., 1996; Zhou et al., 1993)
H159	L		*cya+, crpΔ*	Plasmid	PCR	Lactose+ Ribose	(Zhou et al., 1993)
	R		*cya+, crpΔ*	Plasmid	PCR	Lactose+ Ribose; Lactose	(Niu et al., 1996; Zhou et al., 1993)

P160	T		*cya+, crpΔ*	Plasmid	PCR	Lactose	(Niu et al., 1996)
G162	D		*cya+, crpΔ*	Plasmid	PCR	Lactose+ Ribose; Lactose	(Niu et al., 1996; Zhou et al., 1993)
	C,N		*cya+, crpΔ*	Plasmid	UV	Lactose+ Galactose	(Eschenlauer and Reznikoff, 1991)
	S		*cya+, crpΔ*	Plasmid	PCR	Lactose+ Ribose	(Zhou et al., 1993)
	V, R		*cya+, crpΔ*	Plasmid	PCR	Lactose	(Niu et al., 1996)
I165	T		*cya+, crp+*	Genome		Lactate	(Conrad et al., 2009)
R169	C	A144T + E171G	*Δcya, crp+*	Plasmid	UV	Lactose	(Garges and Adhya, 1988)
Q170	K	T140R, A144T, E, M198K	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	K	T127I	*Δcya, crp+*	Genome	Chemical + UV	Lactose	(Harman et al., 1986; Puskas et al., 1983; Sanders and McGeoch, 1973)
	K	T127I + L195R	*Δcya, crp+*	Genome	Chemical + UV	Arabinose	(Harman et al., 1986; Puskas et al., 1983)
E171	G	A144T + R169C	*Δcya, crp+*	Plasmid	UV	Lactose	(Garges and Adhya, 1988)
Q174	K	A145E	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
V183	A	T140R	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
M189	K		*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	K	Y63F, T140R, A144T, E, Q170K	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	I	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	R	T140K	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
Q193	K	A144T	*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
L195	R		*Δcya, crp+*	Genome		Maltose	(Sekowska et al., 2016)
	R	T127I + Q170K	*Δcya, crp+*	Genome	Chemical + UV	Arabinose	(Harman et al., 1986; Puskas et al., 1983)

Table S2. Rationally engineered CRP mutants.

Position	Mutation	Additional mutation(s)	CRP function	Rationale behind mutagenesis	Reference(s)
D8	K			Transcriptional activation mechanism	(Zhang et al., 1992)
E12	K			Transcriptional activation mechanism	(Zhang et al., 1992)
H19	Y		AR2	Polymerase interaction	(Rhodius et al., 1997)
	Y	K101E	AR2	Polymerase interaction	(Rhodius et al., 1997)
E34	K			Transcriptional activation mechanism	(Zhang et al., 1992)
E37	K			Transcriptional activation mechanism	(Zhang et al., 1992)
K52	A			Polymerase interaction	(Rhodius and Busby, 2000)
	N		AR3	DNA binding; Ligand binding	(Bell et al., 1990; Dai et al., 2004; Lin et al., 2002; Williams et al., 1996)
	N	H159L	AR3	Ligand binding; Polymerase interaction	(Lin et al., 2002; West et al., 1993; Williams et al., 1996)
	N	E96G	AR3	Polymerase interaction	(Williams et al., 1996)
	N	E96G + H159L	AR3	Polymerase interaction	(Williams et al., 1996)
	N	H159L + E181V	AR3	Polymerase interaction	(Williams et al., 1996)
	N	K101E + H159L	AR3	Polymerase interaction	(Rhodius and Busby, 2000)
	N	E58K + K101E + H159L	AR3	Polymerase interaction	(Rhodius and Busby, 2000)
	N	E58G + K101E + H159L	AR3	Polymerase interaction	(Rhodius and Busby, 2000)
	Q		AR3	DNA binding	(Bell et al., 1990)
	E	H159L	AR3	Polymerase interaction	(West et al., 1993)
D53	H		AR3	Allosteric change; Ligand binding	(Dai et al., 2004; Lin et al., 2002; Tzeng and Kalodimos, 2009)
	K		AR3	Transcriptional activation mechanism	(Zhang et al., 1992)
E54	K		AR3	Transcriptional activation mechanism	(Zhang et al., 1992)

E55	K		AR3	Transcriptional activation mechanism	(Irwin and Ptashne, 1987; Zhang et al., 1992)
E58	H		AR3, cAMP binding (*syn*)	Polymerase interaction	(West et al., 1993)
	K		AR3, cAMP binding (*syn*)	Transcriptional activation mechanism; Polymerase interaction	(Rhodius and Busby, 2000; Zhang et al., 1992)
	K	K52N + K101E + H159L	AR3, cAMP binding (*syn*)	Polymerase interaction	(Rhodius and Busby, 2000)
	G		AR3, cAMP binding (*syn*)	Polymerase interaction	(Rhodius and Busby, 2000)
	G	K52N + K101E + H159L	AR3, cAMP binding (*syn*)	Polymerase interaction	(Rhodius and Busby, 2000)
S62	A			Crystal structure	(Gronenborn et al., 1988)
	F			Allosteric change; Ligand binding	(Lin et al., 2002; Tzeng and Kalodimos, 2009)
D68	K			Transcriptional activation mechanism	(Zhang et al., 1992)
E72	D, Q		cAMP binding (*anti*)	Ligand binding	(Belduz et al., 1993; Moore et al., 1992)
	D, Q	A144T	cAMP binding (*anti*)	Ligand binding	(Belduz et al., 1993; Moore et al., 1992)
	A, P,G,H,R		cAMP binding (*anti*)	Ligand binding	(Moore et al., 1992)
	A, P,G,H,R	A144T	cAMP binding (*anti*)	Ligand binding	(Moore et al., 1992)
	L		cAMP binding (*anti*)	Ligand binding	(Belduz et al., 1993)
	L	A144T	cAMP binding (*anti*)	Ligand binding	(Belduz et al., 1993)
E77	K			Transcriptional activation mechanism	(Zhang et al., 1992)
E78	K			Transcriptional activation mechanism	(Zhang et al., 1992)

E81	K			Transcriptional activation mechanism	(Zhang et al., 1992)
R82	K		cAMP binding (*anti*)	Ligand binding	(Belduz et al., 1993; Zhang et al., 1992)
	K	A144T	cAMP binding (*anti*)	Ligand binding	(Belduz et al., 1993; Zhang et al., 1992)
	A,T,E		cAMP binding (*anti*)	Ligand binding	(Moore et al., 1992)
	A,T,E	A144T	cAMP binding (*anti*)	Ligand binding	(Moore et al., 1992)
	L,Q,H		cAMP binding (*anti*)	Ligand binding	(Belduz et al., 1993)
	L,Q,H	A144T	cAMP binding (*anti*)	Ligand binding	(Belduz et al., 1993)
S83	A		cAMP binding (*anti*)	Crystal struture, Ligand binding	(Gronenborn et al., 1988; Moore et al., 1992)
	A	A144T	cAMP binding (*anti*)	Ligand binding	(Moore et al., 1992)
	K		cAMP binding (*anti*)	Crystal struture	(Gronenborn et al., 1988)
	G,C		cAMP binding (*anti*)	Ligand binding	(Lee et al., 1994)
	T,I,V		cAMP binding (*anti*)	Ligand binding	(Moore et al., 1992)
	T,I,V	A144T	cAMP binding (*anti*)	Ligand binding	(Moore et al., 1992)
E93	K			Transcriptional activation mechanism	(Zhang et al., 1992)
E96	K		AR2	Transcriptional activation mechanism	(Zhang et al., 1992)
	G		AR2	Polymerase interaction	(Williams et al., 1996)
	G	K52N	AR2	Polymerase interaction	(Williams et al., 1996)
	G	K52N + H159L	AR2	Polymerase interaction	(Williams et al., 1996)
	G	H159L + E181V	AR2	Polymerase interaction	(Williams et al., 1996)
	G	H159L	AR2	Polymerase interaction	(West et al., 1993; Williams et al., 1996)
K101	E		AR2	Polymerase interaction	(Rhodius et al., 1997; Rhodius and Busby, 2000)

	E	H19Y	AR2	Polymerase interaction	(Latif et al., 2018; Rhodius et al., 1997)
	E	H159L	AR2	Polymerase interaction	(Latif et al., 2018; Rhodius and Busby, 2000)
	E	K52N + H159L	AR2	Polymerase interaction	(Rhodius and Busby, 2000)
	E	K52N + E58K + H159L	AR2	Polymerase interaction	(Rhodius and Busby, 2000)
	E	K52N + E58G + H159L	AR2	Polymerase interaction	(Rhodius and Busby, 2000)
D111	K			Transcriptional activation mechanism	(Zhang et al., 1992)
R123	A,Q,E,P			Ligand binding	(Moore et al., 1992)
	A,Q,E,P	A144T		Ligand binding	(Moore et al., 1992)
T127	A		cAMP binding (*anti*)	Crystial structure, Ligand binding	(Gronenborn et al., 1988; Gunasekara et al., 2015)
	A	S128V	cAMP binding (*anti*)	Ligand binding	(Gunasekara et al., 2015)
	G		cAMP binding (*anti*)	Ligand binding	(Lee et al., 1994)
	L		cAMP binding (*anti*)	Ligand binding	(Dai et al., 2004; Gorshkova et al., 1995; Lin et al., 2002; Wang et al., 2000)
	L	S128I,V,M,L, A	cAMP binding (*anti*)	Ligand binding	(Youn et al., 2006)
	L	S128L + L61V	cAMP binding (*anti*)	Ligand binding	(Youn et al., 2006)
	L	S128I + 161V	cAMP binding (*anti*)	Ligand binding	(Youn et al., 2006)
	L	S128N	cAMP binding (*anti*)	Ligand binding	(Gunasekara et al., 2015)
	I		cAMP binding (*anti*)	Ligand binding	(Lee et al., 1994)
	I	S128I	cAMP binding (*anti*)	Ligand binding	(Youn et al., 2006)
	M		cAMP binding (*anti*)	Ligand binding	(Gunasekara et al., 2015)
	M	S128I	cAMP binding (*anti*)	Ligand binding	(Youn et al., 2006)

	V	S128I	cAMP binding (*anti*)	Ligand binding	(Youn et al., 2006)
	V	S128T,V, A	cAMP binding (*anti*)	Ligand binding	(Gunasekara et al., 2015)
	C	S128I	cAMP binding (*anti*)	Ligand binding	(Gunasekara et al., 2015)
	C,S		cAMP binding (*anti*)	Ligand binding	(Gunasekara et al., 2015; Lee et al., 1994)
	T	S128N	cAMP binding (*anti*)	Ligand binding	(Gunasekara et al., 2015)
S128	A		cAMP binding (*anti*)	Crystal stucture, Ligand binding	(Cheng et al., 1995; Cheng and Ching Lee, 1998; Gronenborn et al., 1988; Lee et al., 1994; Moore et al., 1996; Wang et al., 2000)
	A	G141Q	cAMP binding (*anti*)	Ligand binding	(Cheng and Ching Lee, 1998)
	A	T127V		Ligand binding	(Gunasekara et al., 2015)
	T		cAMP binding (*anti*)	Ligand binding	(Gunasekara et al., 2015; Lee et al., 1994)
	P		cAMP binding (*anti*)	Ligand binding	(Cheng et al., 1995)
	N		cAMP binding (*anti*)	Ligand binding	(Gunasekara et al., 2015)
	N	T127L		Ligand binding	(Youn et al., 2006)
	V	T127A		Ligand binding	(Youn et al., 2006)
	V	T127V		Ligand binding	(Youn et al., 2006)
	I	T127L,I,M,V, C		Ligand binding	(Youn et al., 2006)
	I	T127L + L61V		Ligand binding	(Youn et al., 2006)
	L	T127L + L61V		Ligand binding	(Youn et al., 2006)
	T	T127V		Ligand binding	(Youn et al., 2006)
E129	K			Transcriptional activation mechanism	(Zhang et al., 1992)
D138	K			Transcriptional activation mechanism	(Zhang et al., 1992)
	A,N,E,Q, G,L,K,F,V			Allosteric change	(Ryu et al., 1993)

G141	Q			Ligand binding, Allosteric change	(Cheng and Ching Lee, 1998; Cheng and Lee, 1994; Dai et al., 2004; Kim et al., 1992; Lin et al., 2002)
	S			Allosteric change	(Tzeng and Kalodimos, 2009)
	S,Y,L,R,K, A,I,V			Allosteric change	(Kim et al., 1992)
	K			Ligand binding	(Lin et al., 2002)
R142	H	A144T		Allosteric change	(Tzeng and Kalodimos, 2009)
A144	T			Ligand binding	(Belduz et al., 1993; Moore et al., 1992)
	S,Q,Y,L,F,V ,C			Allosteric change, CRP* phenotype	(Kim et al., 1992)
	T	E72,D,Q,A,P, G,H,R, R82,K,A,T,E, L,Q,H, S83A,T,I,V, R123A,Q,E,P		Ligand binding, Allosteric change	(Belduz et al., 1993; Moore et al., 1992; Tzeng and Kalodimos, 2009)
L148	R			Allosteric change; Ligand binding	(Lin et al., 2002; Tzeng and Kalodimos, 2009)
D155	K			Transcriptional activation mechanism	(Zhang et al., 1992)
T158	A		AR1	Polymerase interaction	(West et al., 1993)
H159	L		AR1	Polymerase interaction	(Bell et al., 1990; Dai et al., 2004; Latif et al., 2018; Lin et al., 2002; Rhodius and Busby, 2000; West et al., 1993; Williams et al., 1996)
	L	K101E	AR1	Polymerase interaction	(Latif et al., 2018; Rhodius and Busby, 2000)
	L	K52N	AR1	Ligand binding; Polymerase interaction	(Lin et al., 2002; West et al., 1993; Williams et al., 1996)
	L	K52E	AR1	Polymerase interaction	(West et al., 1993)
	L	E96G	AR1	Polymerase interaction	(West et al., 1993; Williams et al., 1996)
	L	K52N + K101E	AR1	Polymerase interaction	(Rhodius and Busby, 2000)

	L	K52N + E96G	AR1	Polymerase interaction	(Williams et al., 1996)
	L	K52N + E181V	AR1	Polymerase interaction	(Williams et al., 1996)
	L	E96G + E181V	AR1	Polymerase interaction	(Williams et al., 1996)
	L	K52N + E58K + K101E	AR1	Polymerase interaction	(Rhodius and Busby, 2000)
	L	K52N + E58G + K101E	AR1	Polymerase interaction	(Rhodius and Busby, 2000)
D161	K		AR1	Transcriptional activation mechanism	(Zhang et al., 1992)
Q170	K, E			Transcriptional activation mechanism	(Breul et al., 1993; Irwin and Ptashne, 1987)
	L	E171K		Transcriptional activation mechanism	(Breul et al., 1993)
E171	L			Transcriptional activation mechanism	(Breul et al., 1993)
	K			Transcriptional activation mechanism	(Bell et al., 1990; Breul et al., 1993; Irwin and Ptashne, 1987; Zhang et al., 1992)
	Q			Transcriptional activation mechanism	(Irwin and Ptashne, 1987)
	K	Q170L		Transcriptional activation mechanism	(Breul et al., 1993)
R180	A, G		cAMP binding *(syn)*	DNA binding	(Zhang and Ebright, 1990)
E181	V			Polymerase interaction	(Williams et al., 1996)
	V	K52N + H159L		Polymerase interaction	(Williams et al., 1996)
	V	E96G + H159L		Polymerase interaction	(Williams et al., 1996)
E191	K		DNA recognition	Transcriptional activation mechanism	(Irwin and Ptashne, 1987; Zhang et al., 1992)
D192	K		DNA recognition	Transcriptional activation mechanism	(Zhang et al., 1992)
L195	R			CRP* phenotype	(Harman et al., 1988)

Table S3. The mutational landscape of *Escherichia coli* K12 MG1655 *ΔcyaA* populations during the adaptive evolution experiment. The *crp* genetic background is either wild-type ('WT', with the pTrc*tig* plasmid), CRP A144T ('A144T') or CRP A144E ('A144E'). Age refers to the age of plates in the evolution experiment from which biomass samples were taken. Supplement refers to the nucleotide supplements to MacConkey maltose plates during the adaptive evolution experiment, where either no nucleotide ('None'), 0.1 mM cAMP ('cAMP') or 0.5 mM cGMP ('cGMP') was supplemented. The frequency corresponds to the percentage of the population, which in the nucleotide position for each mutation contained the mutant variant as opposed to the wild-type variant.

Amino acid	Mutation	*crp* genetic background	Age (days)	Supplement	Frequency (%)
G3	D	WT	0	None	0.090
		WT	1	None	0.098
		WT	5	None	0.088
		WT	9	None	0.108
		WT	20	None	0.178
		WT	25	None	0.209
		WT	30	None	0.171
		WT	35	None	0.148
		WT	7	cAMP	0.097
		WT	7	cGMP	0.127
		A144E	7	None	0.094
		A144E	7	None	0.017
	Silent	WT	1	None	0.098
		WT	5	None	0.088
		WT	15	None	0.019
		WT	15	None	0.103
		A144E	7	None	0.017
		A144E	7	None	0.094
K4	N	WT	35	None	0.016
P5	A	WT	7	cAMP	0.032
	Silent	WT	0	None	0.110
		WT	1	None	0.098
		WT	3	None	0.121
		WT	7	None	0.121
		WT	9	None	0.144
		WT	25	None	0.090
		WT	30	None	0.106
		WT	35	None	0.133
		WT	7	cAMP	0.097
		WT	7	cGMP	0.088
		A144T	7	None	0.088
		A144T	7	None	0.016
		A144E	7	None	0.094
T7	Silent	WT	7	cAMP	0.022

E12	*	WT	20	None	0.042
H19	P	WT	20	None	0.021
I30	F	WT	35	None	0.859
		A144T	7	None	0.096
Q32	K	WT	3	None	0.060
	P	WT	35	None	0.431
		A144T	7	None	0.048
		A144E	7	None	0.129
K35	Silent	WT	30	None	0.082
A36	V	A144T	7	None	0.080
T38	P	WT	7	None	0.030
		WT	9	None	0.027
Y41	F	WT	15	None	0.037
S46	Silent	WT	0	None	0.020
A48	V	WT	30	None	0.049
L50	V	WT	15	None	0.019
		WT	15	None	0.038
	P	A144E	7	None	0.060
I51	L	WT	35	None	0.157
		WT	7	cAMP	0.097
D53	N	WT	35	None	3.003
		A144T	7	None	0.500
		A144E	7	None	4.667
	H	A144E	7	None	0.129
	Y	WT	35	None	0.141
		WT	7	cAMP	0.140
		A144T	7	None	0.193
		A144E	7	None	1.380
	A	A144T	7	None	0.089
		A144E	7	None	0.087
	G	WT	35	None	1.498
		A144T	7	None	0.113
		A144E	7	None	0.831
	V	A144T	7	None	0.024
		A144E	7	None	0.061
E54	K	WT	35	None	21.107
		A144E	7	None	0.120
E55	K	WT	35	None	0.205
		A144T	7	None	0.090
		A144E	7	None	2.013
	G	WT	35	None	0.120
		A144T	7	None	0.082
		A144E	7	None	0.290
G56	V	WT	9	None	0.036
	Silent	WT	7	None	0.063
E58	Q	A144E	7	None	0.070
S62	F	WT	25	None	0.100

		WT	35	None	36.783
		A144T	7	None	0.257
		A144E	7	None	2.963
	Y	A144T	7	None	0.088
		A144E	7	None	0.629
	Silent	WT	3	None	0.060
N65	H	A144E	7	None	0.217
	Y	A144E	7	None	0.063
G67	Silent	WT	3	None	0.044
		WT	3	None	0.018
D68	G	A144E	7	None	0.009
		A144E	7	None	0.026
	E	A144E	7	None	0.009
I70	F	A144E	7	None	1.557
	S	A144E	7	None	0.026
	V	A144E	7	None	0.036
	T	A144E	7	None	0.174
G71	D	A144E	7	None	0.036
	V	WT	5	None	0.023
	Silent	WT	3	None	0.072
		WT	5	None	0.054
		WT	5	None	0.054
		WT	7	None	0.032
		WT	9	None	0.047
		WT	25	None	0.115
		WT	30	None	0.068
		WT	7	cAMP	0.079
		A144T	7	None	0.050
		A144E	7	None	0.054
		A144E	7	None	0.054
L73	V	A144E	7	None	0.044
G74	C	WT	7	cAMP	1.070
	Silent	WT	9	None	0.082
		WT	7	cGMP	0.064
L75	Q	A144E	7	None	0.040
F76	V	WT	7	cGMP	0.022
	C	A144T	7	None	0.029
	L	WT	25	None	0.036
	Silent	A144E	7	None	0.012
G79	A	A144E	7	None	0.059
	Silent	WT	1	None	0.109
		WT	5	None	0.032
		WT	5	None	0.032
		WT	30	None	0.121
		WT	7	cAMP	0.100
R82	G	WT	35	None	0.205
		A144T	7	None	0.117

		A144E	7	None	0.180
	H	A144T	7	None	0.043
	P	WT	1	None	0.093
		WT	3	None	0.077
		WT	5	None	0.094
		WT	7	None	0.159
		WT	15	None	0.102
		WT	20	None	0.061
		WT	30	None	0.093
		A144T	7	None	0.072
S83	G	A144T	7	None	0.175
A84	E	WT	25	None	54.226
		A144T	7	None	0.135
		A144E	7	None	1.514
W85	R	A144E	7	None	0.441
V86	Silent	WT	20	None	0.061
R87	H	WT	3	None	0.116
A88	V	WT	35	None	2.813
		A144E	7	None	0.595
T90	P	WT	3	None	0.026
A91	T	WT	0	None	0.080
V94	L	A144E	7	None	0.236
A95	P	WT	1	None	0.027
	T	WT	1	None	0.080
		WT	5	None	0.126
	Silent	WT	1	None	0.067
		WT	3	None	0.090
		WT	3	None	0.013
		WT	7	cAMP	0.123
K100	Q	WT	15	None	0.034
R103	H	WT	35	None	0.103
P110	Q	WT	15	None	18.110
		WT	7	cAMP	0.270
I112	L	A144E	7	None	0.249
R115	Silent	WT	1	None	0.027
S117	Silent	WT	7	None	0.100
		WT	25	None	0.106
A118	T	WT	30	None	0.093
		WT	7	cGMP	0.109
	V	WT	7	None	0.119
Q119	H	A144T	7	None	0.116
		A144E	7	None	0.177
		A144E	7	None	0.734
M120	L	WT	1	None	0.027
	I	WT	7	cAMP	0.234
		WT	7	cAMP	0.037
A121	E	A144E	7	None	0.047

	G	WT	7	cGMP	0.022
	V	WT	0	None	0.091
		WT	1	None	0.134
		WT	3	None	0.116
		WT	30	None	0.094
		WT	35	None	0.176
		WT	7	cGMP	0.098
		A144E	7	None	0.131
R122	H	WT	1	None	0.108
		WT	35	None	0.088
		WT	7	cAMP	0.099
L124	P	WT	0	None	0.091
		WT	15	None	0.086
		WT	7	cGMP	0.109
	Silent	WT	0	None	0.148
		WT	0	None	0.023
		WT	1	None	0.107
		WT	5	None	0.173
		WT	15	None	0.120
		WT	20	None	0.092
		WT	30	None	0.093
		A144T	7	None	0.203
Q125	K	WT	7	cAMP	0.098
V126	A	WT	7	None	0.100
	D	A144E	7	None	0.107
T127	I	A144E	7	None	0.308
E129	A	WT	3	None	0.026
G132	Silent	WT	3	None	0.103
		WT	9	None	0.116
		A144E	7	None	0.012
		A144E	7	None	0.083
N133	H	WT	7	cAMP	0.111
		A144T	7	None	0.102
		A144E	7	None	1.222
	Y	WT	7	cAMP	0.062
	Silent	A144E	7	None	0.629
A135	V	WT	5	None	0.073
		WT	35	None	0.409
		WT	7	cAMP	0.369
		A144T	7	None	0.111
		A144E	7	None	0.117
	Silent	WT	1	None	0.077
D138	E	WT	7	cAMP	0.093
V139	A	A144E	7	None	0.374
T140	P	WT	7	cGMP	1.083
	S	WT	7	cGMP	0.018
	R	WT	5	None	0.073

		WT	7	None	0.051
		WT	9	None	47.513
		WT	25	None	53.574
		A144E	7	None	0.180
	K	WT	20	None	1.215
		WT	7	cAMP	1.019
		WT	7	cGMP	5.525
	M	WT	7	None	0.076
		WT	7	cAMP	0.090
		WT	7	cGMP	0.049
G141	S	WT	7	cAMP	0.409
		WT	7	cGMP	4.952
	A	WT	7	cAMP	0.019
	D	WT	7	cAMP	0.695
		WT	7	cGMP	9.704
	V	WT	25	None	0.036
	Silent	WT	30	None	0.046
R142	H	WT	3	None	0.136
		WT	7	None	0.110
		WT	9	None	0.139
		WT	7	cAMP	0.119
		WT	7	cGMP	0.098
		A144E	7	None	0.079
A144	Q	A144E	7	None	0.009
	K	A144T	7	None	0.030
		A144E	7	None	1.113
	P	WT	5	None	0.023
	T	WT	1	None	0.138
		WT	5	None	23.550
		WT	9	None	0.111
		WT	15	None	56.514
		WT	20	None	85.331
		WT	25	None	38.476
		WT	30	None	83.973
		WT	35	None	99.732
		WT	7	cAMP	0.159
		WT	7	cGMP	7.142
		A144T	7	None	99.728
		A144E	7	None	0.129
	E	WT	7	cAMP	0.208
		WT	7	cGMP	1.377
		A144T	7	None	0.152
		A144E	7	None	98.723
	G	WT	7	cGMP	0.008
	V	WT	7	cAMP	0.027
		WT	7	cGMP	0.264
L147	Silent	WT	35	None	0.123

L148	Silent	WT	0	None	0.066
A151	T	WT	7	None	0.113
		WT	9	None	0.127
		WT	25	None	0.101
		WT	35	None	0.120
		A144T	7	None	0.090
P154	Q	WT	35	None	0.117
M157	V	WT	5	None	0.059
	I	WT	0	None	0.027
T158	P	WT	0	None	0.055
		WT	7	cGMP	0.064
	S	WT	7	cGMP	0.096
P160	Silent	WT	0	None	0.110
		WT	3	None	0.098
		WT	25	None	0.090
		A144T	7	None	0.096
D161	G	WT	9	None	0.142
	E	WT	35	None	0.039
	Silent	WT	35	None	0.156
T168	S	WT	35	None	0.078
Q170	E	A144T	7	None	0.032
	K	WT	30	None	82.890
		WT	35	None	1.404
		WT	7	cAMP	0.421
		A144T	7	None	0.226
		A144E	7	None	3.669
Q174	P	WT	9	None	0.070
R180	G	WT	30	None	0.055
L187	R	WT	35	None	0.039
M189	R	WT	7	cGMP	0.095
		A144T	7	None	0.097
		A144E	7	None	1.197
	K	WT	30	None	7.072
		WT	7	cGMP	0.063
		A144E	7	None	1.160
D192	A	A144T	7	None	0.097
Q193	K	WT	15	None	24.056
L195	R	WT	7	cAMP	0.065
		WT	7	cGMP	5.566
	Q	WT	7	cAMP	0.324
A198	T	WT	7	None	0.166
		WT	15	None	0.179
H199	D	WT	20	None	0.082
G200	S	A144E	7	None	0.183
Y206	Silent	A144T	7	None	0.193

References

Aiba, H., Nakamura, T., Mitani, H., Mori, H., 1985. Mutations that alter the allosteric nature of cAMP receptor protein of Escherichia coli. EMBO J. 4, 3329–3332. doi:10.1002/j.1460-2075.1985.tb04084.x

Belduz, A.O., Lee, E.J., Harman, J.G., 1993. Mutagenesis of the cyclic AMP receptor protein of Escherichia coli : targeting positions 72 and 82 of the cyclic nucleotide binding pocket. Nucleic Acids Res. 21, 2894-1827–1835. doi:10.1093/nar/22.15.2894

Bell, A., Gaston, K., Williams, R., Chapman, K., Kolb, A., Buc, H., Minchin, S., Williams, J., Busby, S., 1990. Mutations that alter the ability of the Escherichia coli cyclic AMP receptor protein to activate transcription. Nucleic Acids Res. 18, 7243–50. doi:10.1093/nar/18.24.7243

Breul, A., Aßmann, H., Golz, R., von Wilcken-Bergmann, B., Müller-Hill, B., 1993. Mutants with substitutions for Glu171 in the catabolite activator protein (CAP) of Escherichia coli activate transcription from the lac promoter. MGG Mol. Gen. Genet. 238, 155–160. doi:10.1007/BF00279542

Cheng, X., Ching Lee, J., 1998. Interactive and dominant effects of residues 128 and 141 on cyclic nucleotide and DNA bindings in Escherichia coli cAMP receptor protein. J. Biol. Chem. 273, 705–712. doi:10.1074/jbc.273.2.705

Cheng, X., Kovac, L., Ching Lee, J., 1995. Probing the Mechanism of CRP Activation by Site-Directed Mutagenesis: The Role of Serine 128 in the Allosteric Pathway of cAMP Receptor Protein Activation. Biochemistry 34, 10816–10826. doi:10.1021/bi00034a014

Cheng, X., Lee, J.C., 1994. Absolute requirement of cyclic nucleotide in the activation of the G141Q mutant cAMP receptor protein from Escherichia coli. J. Biol. Chem. 269, 30781–30784.

Conrad, T.M., Joyce, A.R., Applebee, M.K., Barrett, C.L., Xie, B., Gao, Y., Palsson, B.T., 2009. Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol. 10, 1–12. doi:10.1186/gb-2009-10-10-r118

Dai, J., Lin, S.-H., Kemmis, C., Chin, A.J., Lee, J.C., 2004. Interplay between Site-Specific Mutations and Cyclic Nucleotides in Modulating DNA Recognition by Escherichia coli Cyclic AMP Receptor Protein † , ‡. Biochemistry 43, 8901–8910. doi:10.1021/bi0499359

Eschenlauer, A.C., Reznikoff, W.S., 1991. Escherichia coli catabolite gene activator protein

mutants defective in positive control of lac operon transcription. J. Bacteriol. 173, 5024–5029. doi:10.1128/jb.173.16.5024-5029.1991

Garges, S., Adhya, S., 1988. Cyclic AMP-induced conformational change of cyclic AMP receptor protein (CRP): intragenic suppressors of cyclic AMP-independent CRP mutations. J. Bacteriol. 170, 1417–1422. doi:10.1128/jb.170.4.1417-1422.1988

Garges, S., Adhya, S., 1985. Sites of allosteric shift in the structure of the cyclic AMP receptor protein. Cell 41, 745–751. doi:10.1016/S0092-8674(85)80055-6

Gorshkova, I., Moore, J.L., McKenney, K.H., Schwarz, F.P., 1995. Thermodynamics of cyclic nucleotide binding to the cAMP receptor protein and its T127L mutant. J. Biol. Chem. 270, 21679–21683. doi:10.1074/jbc.270.37.21679

Gronenborn, A.M., Sandulache, R., Gärtner, S., Clore, G.M., 1988. Mutations in the cyclic AMP binding site of the cyclic AMP receptor protein of Escherichia coli. Biochem. J. 253, 801–807. doi:10.1042/bj2530801

Gunasekara, S.M., Hicks, M.N., Park, J., Brooks, C.L., Serate, J., Saunders, C. V., Grover, S.K., Goto, J.J., Lee, J.-W., Youn, H., 2015. Directed Evolution of the Escherichia coli cAMP Receptor Protein at the cAMP Pocket. J. Biol. Chem. 290, 26587–26596. doi:10.1074/jbc.m115.678474

Harman, J.G., McKenney, K., Peterkofsky, A., 1986. Structure-function analysis of three cAMP-independent forms of the cAMP receptor protein. J. Biol. Chem. 261, 16332–16339. doi:3023348

Harman, J.G., Peterkofsky, A., McKenney, K., 1988. Arginine substituted for leucine at position 195 produces a cyclic AMP-independent form of the Escherichia coli cyclic AMP receptor protein. J. Biol. Chem. 263, 8072–8077.

Irwin, N., Ptashne, M., 1987. Mutants of the catabolite activator protein of Escherichia coli that are specifically deficient in the gene-activation function. Proc. Natl. Acad. Sci. 84, 8315–8319. doi:10.1073/pnas.84.23.8315

Kim, J., Adhya, S., Garges, S., 1992. Allosteric changes in the cAMP receptor protein of Escherichia coli: hinge reorientation. Proc. Natl. Acad. Sci. 89, 9700–9704. doi:10.1073/pnas.89.20.9700

Latif, H., Federowicz, S., Ebrahim, A., Tarasova, J., Szubin, R., Utrilla, J., Zengler, K., Palsson, B.O., 2018. ChIP-exo interrogation of Crp, DNA, and RNAP holoenzyme interactions. PLoS One 13, e0197272. doi:10.1371/journal.pone.0197272

Lee, E.J., Glasgow, J., Leu, S.-F., Belduz, A.O., Harman, J.G., 1994. Mutagenesis of the cyclic AMP receptor protein of Escherichia coli : targeting positions 83, 127 and 128 the

cyclic nucleotide binding pocket. Nucleic Acids Res. 22, 2894–2901. doi:10.1093/nar/22.15.2894

Lin, S.H., Kovac, L., Chin, A.J., Chin, C.C.Q., Lee, J.C., 2002. Ability of E. coli Cyclic AMP receptor protein to differentiate cyclic nucelotides: Effects of single site mutations. Biochemistry 41, 2946–2955. doi:10.1021/bi0119215

Macconkey, A., 1905. Lactose-fermenting bacteria in faeces. J. Hyg. (Lond).

Melton, T., Snow, L.L., Freitag, C.S., Dobrogosz, W.J., 1981. Isolation and Characterization of cAMP Suppressor Mutants of Escherichia coli K12. Mol. Gen. Genet. 182, 480–489.

Moore, J., Kantorow, M., Vanderzwaag, D., McKenney, K., 1992. Escherichia coli Cyclic AMP Receptor Protein Mutants Provide Evidence for Ligand Contacts Important in Activation. J. Bacteriol. 174, 8030–8035. doi:10.1128/jb.174.24.8030-8035.1992

Moore, J.L., Gorshkova, I.I., McKenney, K.H., Schwarz, F.P., Brown, J.W., 1996. Effect of cAMP Binding Site Mutations on the Interaction of cAMP Receptor Protein with Cyclic Nucleoside Monophosphate Ligands and DNA. J. Biol. Chem. 271, 21273–21278. doi:10.1074/jbc.271.35.21273

Niu, W., Kim, Y., Tau, G., Heyduk, T., Ebright, R.H., 1996. Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase. Cell 87, 1123–1134. doi:10.1016/S0092-8674(00)81806-1

Puskas, R., Fredd, N., Gazdar, C., Peterkofsky, A., 1983. Methylglyoxal-Mediated Growth Inhibition in an Escherichia coli cAMP Receptor Protein Mutant. Arch. Biochem. Biophys. 223, 503–513.

Rhodius, V.A., Busby, S.J.W., 2000. Interactions between Activating Region 3 of the Escherichia coli cyclic AMP receptor protein and region 4 of the RNA polymerase σ70subunit: Application of suppression genetics. J. Mol. Biol. 311–324. doi:10.1006/jmbi.2000.3737

Rhodius, V.A., West, D.M., Webster, C.L., Busby, S.J.W., Savery, N.J., 1997. Transcription activation at class II CRP-dependent promoters: The role of different activating regions. Nucleic Acids Res. 25, 326–332. doi:10.1093/nar/25.2.326

Ryu, S., Kim, J., Adhya, S., Garges, S., 1993. Pivotal role of amino acid at position 138 in the allosteric hinge reorientation of cAMP receptor protein. Proc. Natl. Acad. Sci. 90, 75–79. doi:10.1073/pnas.90.1.75

Sanders, R., McGeoch, D., 1973. A Mutant Transcription Factor That is Activated By 3':5'-Cyclic Guanosine Monophosphat. Proc. Natl. Acad. Sci. 70, 1017–1021.

Sekowska, A., Wendel, S., Fischer, E.C., Nørholm, M.H.H., Danchin, A., 2016. Generation

of mutation hotspots in ageing bacterial colonies. Sci. Rep. 6, 2. doi:10.1038/s41598-016-0005-4

Sievert, C., Nieves, L.M., Panyon, L.A., Loeffler, T., Morris, C., Cartwright, R.A., Wang, X., 2017. Experimental evolution reveals an effective avenue to release catabolite repression via mutations in XylR. Proc. Natl. Acad. Sci. 114, 7349–7354. doi:10.1073/pnas.1700345114

Søgaard-Andersen, L., Mironov, A.S., Pedersen, H., Sukhodelets, V. V., Valentin-Hansen, P., 1991. Single amino acid substitutions in the cAMP receptor protein specifically abolish regulation by the CytR repressor in Escherichia coli. Proc. Natl. Acad. Sci. 88, 4921–4925. doi:10.1073/pnas.88.11.4921

Tagami, H., Inada, T., Kunimura, T., Aiba, H., 1995. Glucose lowers CRP*levels resulting in repression of the lac operon in cells lacking cAMP. Mol. Microbiol. 17, 251–258. doi:10.1111/j.1365-2958.1995.mmi_17020251.x

Tzeng, S.R., Kalodimos, C.G., 2009. Dynamic activation of an allosteric regulatory protein. Nature 462, 368–372. doi:10.1038/nature08560

Wang, S., Shi, Y., Gorshkova, I., Schwarz, F.P., 2000. RNA polymerase-cNMP-ligated cAMP receptor protein (CRP) mutant interactions in the enhancement of transcription by CRP mutants. J. Biol. Chem. 275, 33457–33463. doi:10.1074/jbc.M004877200

West, D., Williams, R., Rhodius, V., Bell, A., Sharma, N., Zou, C., Fujita, N., Ishihama, A., Busby, S., 1993. Interactions between the Escherichia coli cyclic AMP receptor protein and RNA polymerase at Class II promoters. Mol. Microbiol. 10, 789–797. doi:10.1111/j.1365-2958.1993.tb00949.x

Williams, R.M., Rhodius, V.A., Bell, A.I., Kolb, A., Busby, S.J.W., 1996. Orientation of functional activating regions in the Escherichia coli CRP protein during transcription activation at class II promoters. Nucleic Acids Res. 24, 1112–1118. doi:10.1093/nar/24.6.1112

Youn, H., Kerby, R.L., Conrad, M., Roberts, G.P., 2006. Study of highly constitutively active mutants suggests how cAMP activates cAMP receptor protein. J. Biol. Chem. 281, 1119–1127. doi:10.1074/jbc.M509421200

Zhang, X., Zhou, Y., Ebright, W., Ebrights, H., 1992. Catabolite Gene Activator Protein (CAP) Is Not an " Acidic Activating Region " Transcription Activator Protein. J. Biol. Chem. 267, 8136–8139.

Zhang, X.P., Ebright, R.H., 1990. Identification of a contact between arginine-180 of the catabolite gene activator protein (CAP) and base pair 5 of the DNA site in the CAP-

DNA complex. Proc. Natl. Acad. Sci. 87, 4717–4721. doi:10.1073/pnas.87.12.4717

Zhou, Y., Zhang, X., Ebright, R.H., 1993. Identification of the activating region of catabolite gene activator protein (CAP): isolation and characterization of mutants of CAP specifically defective in transcription activation. Proc. Natl. Acad. Sci. U. S. A. 90, 6081–5.

Paper 6

Temporal evolution of master regulator Crp identifies pyrimidines as catabolite modulator factors

Temporal evolution of master regulator Crp identifies pyrimidines as catabolite modulator factors

Authors: Ida Lauritsen[1]†, Pernille Ott Frendorf[1]†, Silvia Capucci[1]†, Sofie Wendel[1], Emil C. Fischer[1], Agnieszka Sekowska[2], Antoine Danchin[3], Morten H. H. Nørholm[1]*

Affiliations:

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet B220, DK-2800 Kgs. Lyngby, Denmark

[2] Institut de Cardiométabolisme et Nutrition, CHU Pitié-Salpêtrière, 47 boulevard de l'Hôpital 75013 Paris, France

[3] Institut Cochin, INSERM U1016 - CNRS UMR8104 - Université Paris Descartes, 24 rue du Faubourg Saint-Jacques, 75014 Paris, France

† These authors contributed equally

* Correspondance to morno@biosustain.dtu.dk

Abstract: Genetic information is stored in DNA in the combination of four nucleotides A, T, C, and G. It is unclear why the phosphate-free nucleosides and nucleobases are not synthesized *de novo*, but instead are scavenged from nucleotides, DNA, RNA, and membranes. Here, by studying the evolution of ageing bacterial colonies, we uncover an unexpected role for pyrimidine nucleosides as antagonists of the global transcription factor Crp. Cytidine and uridine act like the catabolite modulating factors whose existence was suggested by Jacques Monod and colleagues in 1976 and proposed to interact directly with Crp. The paradigmatic transcription factor Crp thus appears to serve a dual role in sensing both carbon availability and metabolic flux towards DNA and RNA.

One-sentence summary: Crp/Cap senses the pyrimidine nucleosides cytidine and uridine and connects directed overflow metabolism with global gene regulation in *Escherichia coli*

Main text:

Cells prefer glucose and a few other carbohydrates as a carbon source and regulate their genetic programs accordingly. This phenomenon has been known for more than a century as glucose effects: the presence of glucose or other rapidly metabolizable substrates in growth media elicits a more or less severe but permanent repression of catabolic enzymes. This outcome is split into three processes: transient repression, inducer exclusion, and carbon catabolite repression (CCR) (*1*, *2*). In the model bacterium *Escherichia coli*, the purine nucleotide cyclic AMP plays a major role in CCR: cAMP is produced in the absence of glucose (*3*), and then binds and activates the global transcription factor cAMP receptor protein (Crp, also known as Cap, Fig. 1A) (*4*, *5*).

The first mechanistic details of gene regulation were revealed by the pioneering work by Jacob and Monod on the *lac* operon (*6*), one out of hundreds of operons activated by Crp. Yet, while Crp is a paradigm in positive gene regulation, the full details behind CCR are still not fully elucidated. For example, Ullmann, Monod and co-workers observed that water-soluble extracts of *E. coli* repressed catabolite-sensitive operons such as *lac* (*7*), and it was later speculated that Crp was negatively regulated by direct binding of a catabolite-modulator factor (CMF) to Crp (*8*). Studies suggest that α-ketoacids serve a role in inducing CCR (*9*, *10*) as one type of CMF, however, a physiologically relevant CMF that directly binds and inhibits Crp has never been identified.

In addition to regulating carbon metabolism, living cells maintain nucleotide homeostasis in response to supply and demand of genetic material. This is critical because the concentration and balance of nucleotides affect mutation rates (*11*). The pyrimidine nucleotide uridine monophosphate (UMP) is synthesized *de novo* from central metabolism in multiple enzymatic steps (*12*). The dephosphorylated nucleoside uridine and the nucleobase uracil are not precursors of UMP in this biosynthetic pathway (Fig. 1B). However, enzymes exist that enable scavenging of uridine and uracil from the environment or from turnover of RNA that is converted into UMP. In *E. coli*, two enzymes UmpH and UmpG produce uridine from UMP as a safety mechanism termed directed overflow metabolism to maintain pyrimidine homeostasis when genetic material is in low demand (*13*). The pyrimidine nucleosides and nucleotides cytidine, CMP and CDP are even further disconnected from *de novo* synthesis and are only produced in cells by hydrolysis of RNA, of precursors of complex carbohydrates or phospholipids, or scavenging from the environment (Fig. 1B) (*14*). CDP is

the precursor of deoxy CDP that is required for incorporating cytosine into DNA. It is also a major source of thymine (Fig.1B). This raises fundamental questions: why is the biosynthesis conserved like this and do the pyrimidines serve special roles in living cells? The research described here establishes a direct link between pyrimidine metabolism and CCR via the master transcriptional regulator Crp showing that pyrimidines act as signaling molecules.

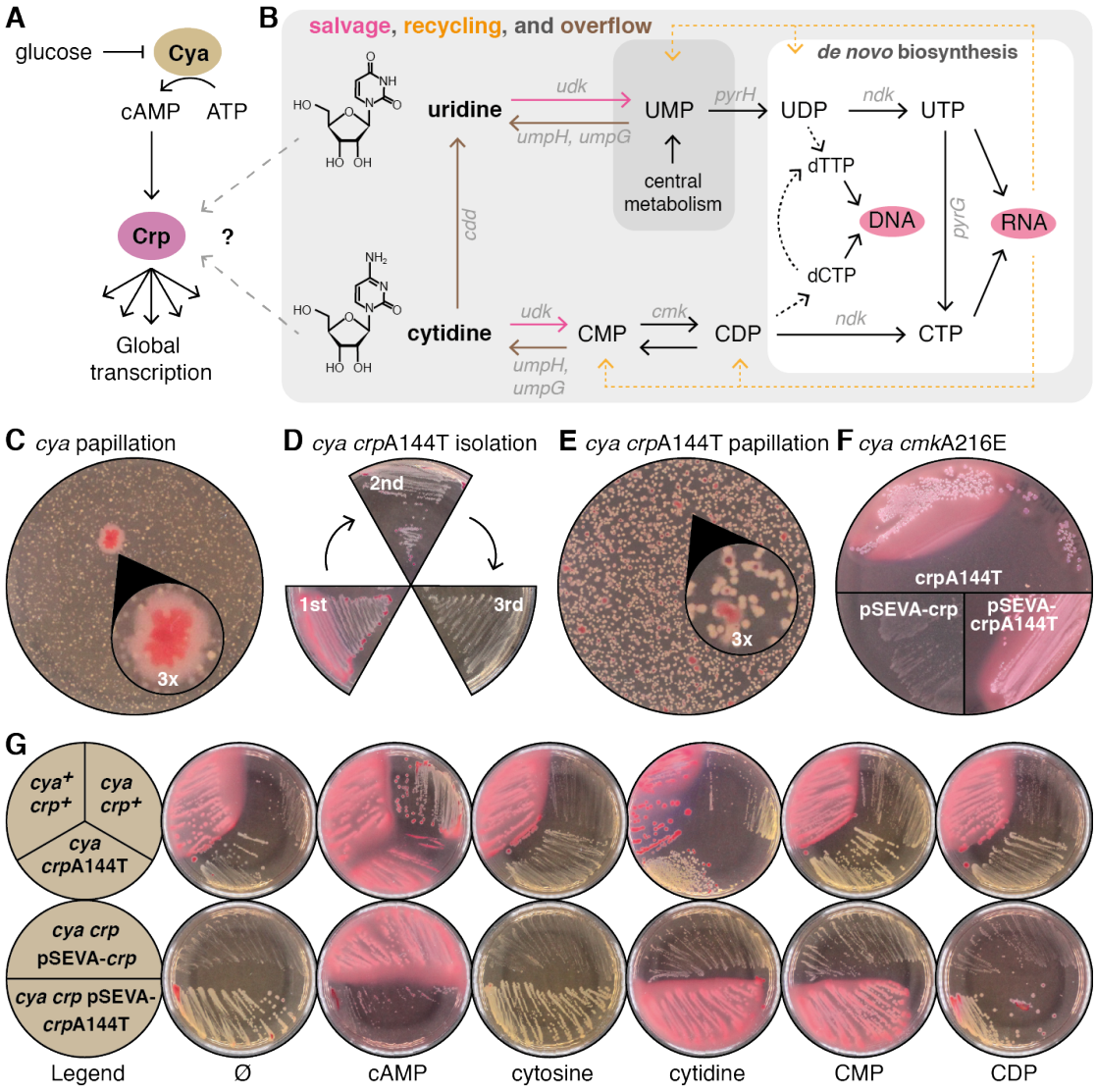


Fig. 1. Crp mutants point to a link between catabolite repression and pyrimidine metabolism. (**A**) In the absence of glucose, the enzyme adenylate cyclase (Cya) produces cyclic AMP that binds and activates the global transcription factor cAMP Receptor Protein (Crp). (**B**) Pyrimidines materialize in living cells by *de novo* synthesis from central metabolites, salvage via uptake from the environment, recycling mainly from RNA, or directed overflow from nucleotides. (**C**) *Escherichia coli cya* mutants grow poorly on maltose MacConkey medium, but upon extended incubation mutant red papillae appear. (**D**) A majority of papillae contain the mutation CrpA144T, but the fermentation phenotype gradually disappears upon re-streaking. (**E**) The *crp*A144T mutation accelerates mutant papillae formation. (**F**) The A216E mutation in the *cmk* gene that encodes cytidylate kinase, increases the

activity of CrpA144T. This is shown in a *cya crp*A144T strain background (upper part of plate) or in a *cya crp* background that is complemented with low-copy plasmids expressing the wildtype *crp* or A144T mutant (lower part of plate). (**G)** Exogenously added cytosine nucleosides and nucleotides stimulate the activity of the CrpA144T mutant, both when expressed from the genome and from a low-copy pSEVA plasmid. cAMP was included as a positive control.

Results:

We previously explored the evolution of ageing *E. coli* colonies using a *cya* genetic background deficient in synthesis of cAMP (*15*). The mutant grows poorly but forms small white colonies on MacConkey agar supplied with maltose. When left in the incubator for up to two months, adaptive mutants appear as red "papillae" or "secondary colonies" (Fig. 1C and movie S1). Many of these have uncovered a cAMP-independent route to maltose fermentation, leading to formation of organic acids and red color due to a pH indicator in the medium (*15*).

A major mutational hotspot identified under these conditions is the *crp* locus and many of these mutants (termed Crp*) were previously shown to activate Crp in the absence of cAMP (*16*–*18*). In a total of 594 clones previously sequenced, either by whole genome re-sequencing or by amplicon sequencing of the *crp* locus, 523 (88%) had mutations in *crp* (*15*) (Table S1). Remarkably, one specific mutation was predominant: an alanine to threonine substitution in position 144, occurring in 349 (67%) of the 523 isolates with mutations in *crp*. The second most abundant mutation identified was in the same position and changes alanine into glutamate (17% of the 523). A144T is a canonical Crp* mutant and together with G141D, it is the most frequently isolated Crp mutation – both have been identified in at least four independent *in vivo* evolution experiments since 1981 (reviewed in (*18*)), and the structure of the A144T mutant has been studied using X-ray crystallography (*19*).

We were driven by the curiosity to understand the mechanistic details of temporal evolution of Crp and a number of observations warranted further exploration: Firstly, while a number of the isolated strains exhibited a clear maltose fermentation phenotype, the *crp*A144T mutation exhibited only a transient phenotype on the selective medium. Mutant papillae typically turn red on MacConkey agar when maltose is efficiently fermented (Fig. 1C), and the A144T mutant grew better than the parental strain (Fig. 1G). However, red color was gradually lost when A144T mutants were re-streaked on fresh medium (Fig. 1D).

Secondly, in the 594 sequenced *crp* loci, some mutations such as Q170K and S62F occur at above average frequency but are only found in combination with other mutations – mainly A144T, but also A144E, and T140R (*15*) (Table S1). In line with this observation, papillae occurred at high frequency when starting with a *crp*A144T strain background (Fig. 1E) and deep sequencing of papillae developing from this background confirmed appearance of mutations Q170K, S62F, and many others (*18*). In contrast, other canonical Crp* mutations such as T140K and G141D develop as single Crp mutations, but at much lower frequency than A144T (*15*). What makes A144T dominant under these selective conditions, and why do second site mutations develop in the A144T background?

Thirdly, in the 96 full genomes sequenced from papillae, 24 had mutations in the gene *cmk* (*15*). *cmk* encodes cytidylate kinase that catalyzes phosphorylation of CMP to CDP (*20*). Among the other 96 genome sequenced isolates, mutations were identified in *carA*, *udk, pyrC, pyrG, pnp,* and *umpH* (*15*), strongly indicating a phenotypic link between pyrimidine metabolism and Crp-dependent growth on maltose. When attempting to introduce the dominating *cmk*A216E mutation into the parental *cya* strain by recombineering, we were unable to isolate strains without A144T mutations spontaneously formed in *crp* (*15*). Furthermore, CrpA144T appears to be more active in a *cmk*A216E background, based on growth on maltose MacConkey agar, both when *crp*A144T is expressed from the genomic locus and from a plasmid (Fig. 1F). What is the connection between Crp, CCR and pyrimidine metabolism?

We hypothesized that a metabolite, perhaps related to pyrimidines, was building up in the ageing colonies, and was affecting the solution space of Crp mutants isolated. The rationale was that upon restoring growth, either by re-streaking on fresh medium or by forming a Crp* mutation in a subpopulation of the colony, the metabolite would be gradually lost together with the maltose fermentation phenotype. At the same time, additional mutations conferring independence of this metabolite would become advantageous.

cmk knockout mutants have been shown to accumulate 30-fold more CMP than a wildtype strain (*20*), and we therefore speculated that CMP, or a closely related metabolite, was playing a role in promotion of cAMP-independent growth. To test this hypothesis, we plated a *cya crp*A144T double mutant, a *cya* single mutant, and a *cya*+ strain on maltose MacConkey agar in the presence of 10 mM cytosine, cytidine, CMP, CDP or 0.5 mM cAMP and monitored growth on maltose MacConkey agar. We observed strong acidification of the

medium by the *cya crp*A144T mutant only when grown in the presence of cAMP and weak acidification when grown with cytidine (Fig. 1G). To substantiate this observation, and because additional mutations occur at higher frequency in this background making it hard to control genetically, we complemented a *cya crp* strain with a low-copy plasmid version of wildtype *crp* or *crp*A144T and observed a clear fermentation phenotype of the A144T mutant when grown in the presence of cytidine or CMP (Fig. 1G). The different phenotypes of the *crp*A144T mutants, observed when expressed from the genome or the low-copy vector, is likely due to differences in *crp* expression levels. This is probably also why *crp*A144T expressed from a plasmid becomes toxic in presence of cAMP, Fig. 1G).

A drawback of this simple *in vivo* screen is that it is not possible to rule out that different nucleosides and their phosphorylated counterparts are transported differently across the inner and outer membrane of *E. coli,* or that phosphorylation or dephosphorylation of the added compounds takes place. To directly test interactions between Crp and different metabolites, we performed *in vitro* label-free biolayer interferometry. Wildtype Crp and A144T were expressed and purified by affinity chromatography, mixed with different ligands, and interactions with biotinylated synthetic DNA encoding P*malT,* a Crp responsive promoter from *E. coli* (*21*), were analyzed using the Octet RED96 system (Fig. 2A). In line with our *in vivo* screen, this analysis showed that cytidine and cAMP activated the DNA binding activity of the A144T mutant (Fig. 2B-C), as did uridine, but not the 2'-deoxy nucleoside thymidine (Fig. 2C). In contrast, only cAMP activated wildtype Crp (Fig. 2D). In competition with cAMP, a 1000-fold surplus of cytidine, uridine, CMP or UMP showed a 23, 14, 10, and 15% decrease in Crp activity, respectively (Fig. 2E). Again, thymidine showed no inhibitory effect, whereas the known (but physiologically irrelevant) inhibitor cyclic GMP showed an 84% decrease in Crp activity. This suggests that Crp interacts directly with pyrimidine nucleosides and that the nucleoside ribose 2'-hydroxyl group is involved in the recognition by Crp.

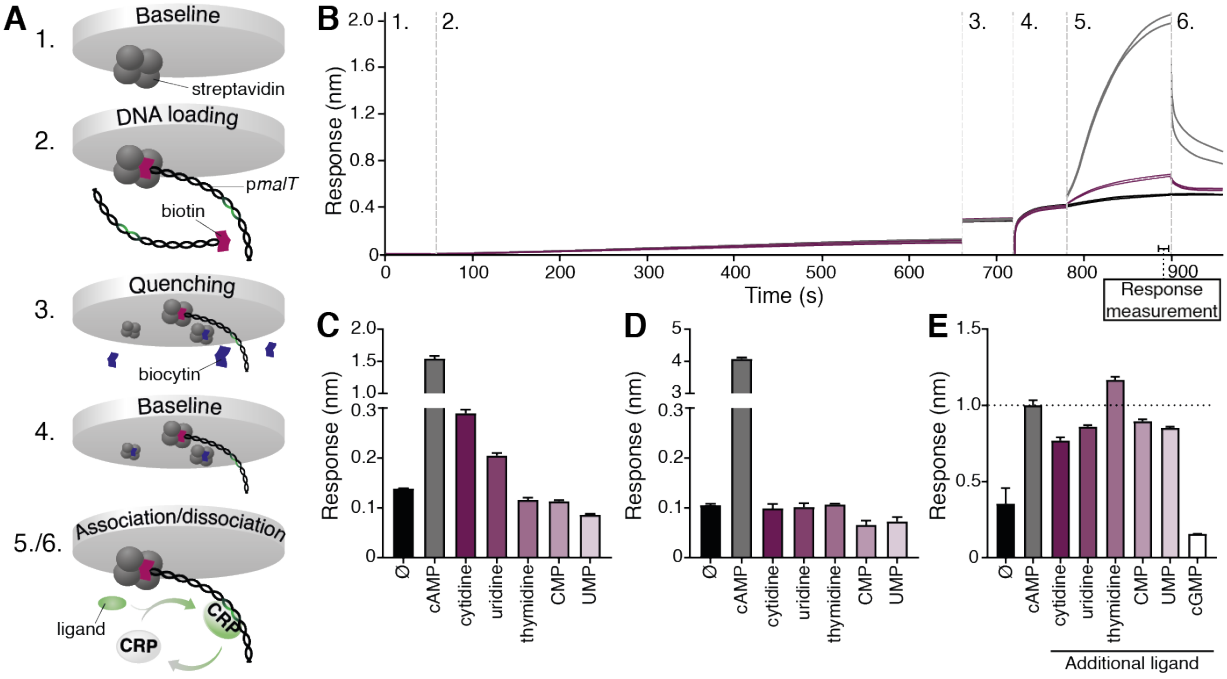


Fig. 2. *In vitro* activity of Crp and CrpA144T. (**A**) Illustration of biolayer interferometry. A streptavidin-coated biosensor (1) associates with biotinylated DNA encoding the *malT* promoter (2), followed by saturation of non-bound streptavidin using biocytin (3). The response is reset (4) before association with Crp and ligand (5) and subsequent dissociation (6). (**B**) Representative output from the biolayer interferometry assay showing progression of response (nm) aligned to baseline. Data corresponds to data series Ø (black line), cAMP (grey line) and cytidine (purple line) in Fig. 2C. (**C**) Association of CrpA144T to P*malT* in the presence of different pyrimidines (10 mM) or the positive control cAMP (0.5 mM). No ligand added (Ø) serves as a negative control. (**D**) Association of wildtype Crp to P*malT* in the presence of different pyrimidines (10 mM) or the positive control cAMP (0.5 mM). (**E**) Association of wildtype Crp to P*malT* in the presence of cAMP (5 µM) in combination with 1000-fold excess of different pyrimidines. The competitive inhibitor cGMP was included as a positive control, and the response is normalized to the effect of cAMP in the absence of other ligands (grey, dotted line). Data represent the average of two replicates with standard deviations.

Cytidine and uridine have not previously been reported to interact with Crp. To explore this in more detail *in vivo*, we developed a sensitive Crp activity assay based on a plasmid carrying a fusion between P*malT* and the fluorescent reporter GFP (Fig. 3A). As intended, the construct responded highly sensitively to different concentrations of cAMP added exogenously in a *cya* background (Fig. 3A). In agreement with both our *in vitro* data, and our qualitative *in vivo* assay on maltose MacConkey, we observed an increase in activity of the CrpA144T mutant in the presence of 10 mM cytidine (Fig. 3B) and uridine (fig. S1). Furthermore, in the wildtype K12 MG1655 strain, the reporter was clearly inhibited by exogenously adding these pyrimidines and we observed the same effects when replacing

P*malT* with P*lac* in the reporter plasmid (Fig. 3B and fig. S1). This suggests that cytidine and uridine act as signaling molecules by inhibiting Crp. The observed effects are stronger *in vivo* than *in vitro*, which may reflect missing components in the simplified *in vitro* system or that pyrimidines affect more than the binding of Crp to DNA, for example Crp-RNA polymerase interactions.

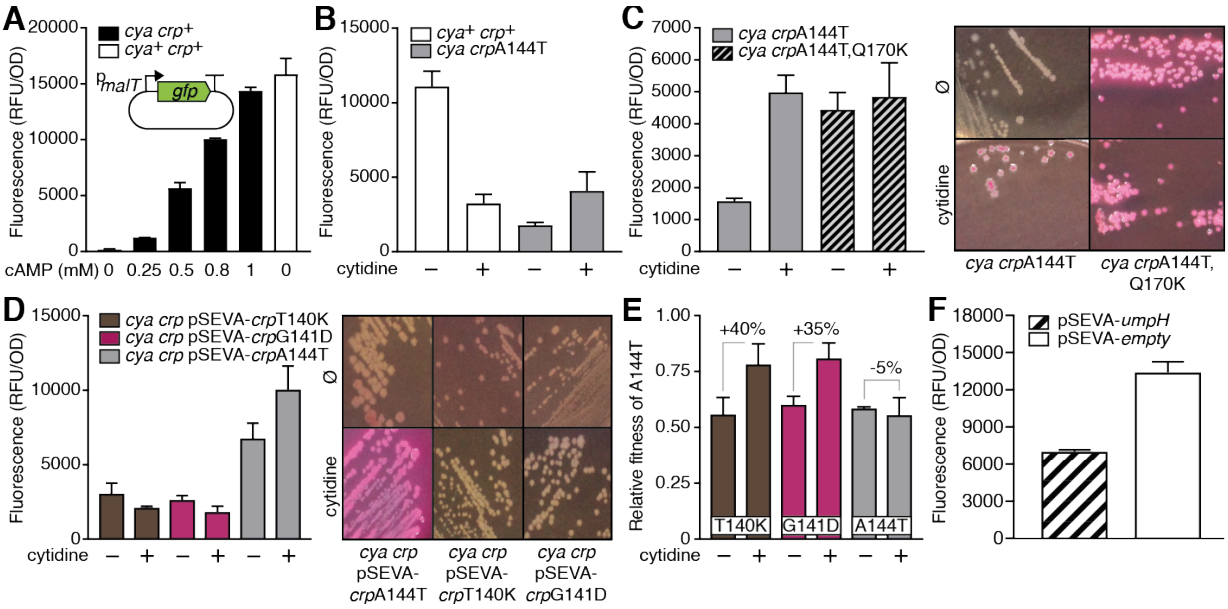


Fig. 3. *In vivo* activity and fitness effects of different Crp mutants. (**A**) An *in vivo* Crp activity reporter, based on a plasmid with the *malT* promoter controlling GFP expression (illustrated), was validated in a *cya* strain by detecting fluorescence in the presence of different concentrations of cAMP. A *cya*+ strain served as positive control. (**B**) The effect of exogenously added cytidine (10mM) on the GFP reporter in a wildtype (*cya*$^+$*crp*$^+$) or mutant (*cya crp*A144T) strain. (**C**) Effect of the additional Crp mutation Q170K in presence or absence of exogenously added cytidine assayed with the GFP reporter (left panel) or on maltose MacConkey agar (right panel). (**D**) Relative activities of Crp* mutations T140K and G141D compared to A144T, expressed from a low-copy SEVA plasmid, in the presence or absence of cytidine assayed with the GFP reporter (left panel) or on maltose MacConkey agar (right panel). (**E**) Relative fitness of Crp* mutation A144T in competition with T140K and G141D by growth on maltose agar plates in the presence or absence of cytidine. (**F**) Effect of *umpH* overexpression, compared to empty vector control, in the wildtype strain assayed with the GFP reporter. Data represent the average of three biological replicates with standard deviations.

Additional mutations such as Q170K frequently occur in the *crp*A144T background and we speculated that this was due to the loss of an activating metabolite. To investigate this hypothesis, we assayed the activity of both the A144T mutant and the A144T Q170K double mutant expressed in the presence or absence of 10 mM cytidine using both the GFP reporter

and maltose MacConkey agar. This showed that Q170K has an activating effect on CrpA144T, much like cytidine, and that the double mutant no longer is affected by cytidine (Fig. 3C). It is thus plausible that A144T is strongly selected for when cytidine accumulates, and that when growth is restored, pyrimidine levels drop, thereby making additional mutations such as Q170K advantageous. In support of this idea, five papillae isolated from the *cya* strain grown in the presence of cytidine all had developed the *crp*A144T mutation.

For a full understanding of the evolutionary mechanisms at play in a specific selective environment, it is important not only to consider the mutations that dominate, but also those that are underrepresented – much akin to Abraham Wald's statistical principles on WWII aircraft survivability (*22*) – survivors dominate the data, but the missing data contains important information: The A144T (**G**CA->**A**CA) and A144E (G**C**A->G**A**A) mutations occur in starving colonies at higher frequency than other mutations such as T140K (A**C**G->A**A**G) and G141D (G**G**C->G**A**C) (*15*), but all have been characterized previously as Crp* mutants (*18*). Further, the mutated residues are neighbors in the same structural domain, and even pairwise represent the same types of mutations (C->A or G->A). The latter is important because the mutations C->A, most likely caused by oxidation of guanosine on the complementary strand, and G->A, most likely caused by cytosine deamination on the complementary strand, are highly dominant under these conditions (*15*, *23*) and mutation bias caused by the available mutational space could limit the observed solution space. This prompted us to assay the activity of these variants in response to different pyrimidine levels. We found that the T140K and G141D mutants were less active than A144T and like wildtype Crp, but in contrast to CrpA144T, were inhibited by cytidine (Fig. 3D). Furthermore, a fitness assay showed an increase in fitness of A144T mutants relative to T141K and G141D mutants in the presence of cytidine when grown on agar plates with maltose (Fig. 3E and fig. S2).

Uridine acts similarly to cytidine by inhibiting Crp and activating CrpA144T. However, in contrast to CMP, UMP is *de novo* synthesized and biosynthesis is feedback regulated, which may explain why *cmk* mutations that likely lead to build-up of both CMP and cytidine are more frequently observed under our experimental conditions. Accumulation of uracil and uridine, catalysed by the enzymes UmpH and UmpG (Fig. 1B), was previously observed in mutant *E. coli* with defective feedback regulation of pyrimidine metabolism (*13*). In nutrient-rich conditions, however, uridine may be a more relevant Crp ligand than

cytidine since it is present in concentrations several orders of magnitude higher than cytidine and cAMP (*24*).

To explore if directed overflow metabolism catalysed by UmpH impacts Crp activity, we expressed UmpH from a plasmid in the presence of the P*malT*-GFP reporter. Indeed, UmpH expression significantly reduced expression from the Crp sensitive reporter (Fig. 3F). In a previous study, global gene regulation was explored by transcriptomics when overflow of uridine and uracil occurred in a pyrimidine feedback dysregulated strain (*13*). In line with our observations, transcriptomics showed that expression of genes known to be positively regulated by Crp (*yfcT, malE, malK, malM, bglB*, and *csgF*) were repressed in this strain, whereas genes known to be downregulated by Crp (*gadA, gadE, gadX*, and *gadW*) were upregulated (*13*).

The Crp transcription factor is widespread in the bacterial kingdom. Curiously, it is found even in bacteria that have little or no cAMP (*25*) and some homologues have been found to work independently of cAMP (*26*, *27*). In light of our findings that cytidine and uridine act as Crp antagonists, is it possible that pyrimidines are more physiologically relevant Crp ligand than cAMP in these bacteria? To explore this hypothesis, we cloned and expressed two Crp homologues from *Pseudomonas aeruginosa* and *Pseudomonas putida* and assayed their activity with the sensitive P*malT*-GFP reporter in the *cya* strain. Compared to the corresponding *E. coli* construct, both of the *Pseudomonas* Crp homologues were highly active in the cAMP-synthesis defective strain, but all three were inhibited when cytidine was added exogenously (Fig. 4 and fig. S3). Thus, it is possible that pyrimidine sensing is a common feature of Crp homologues throughout the bacterial kingdom.

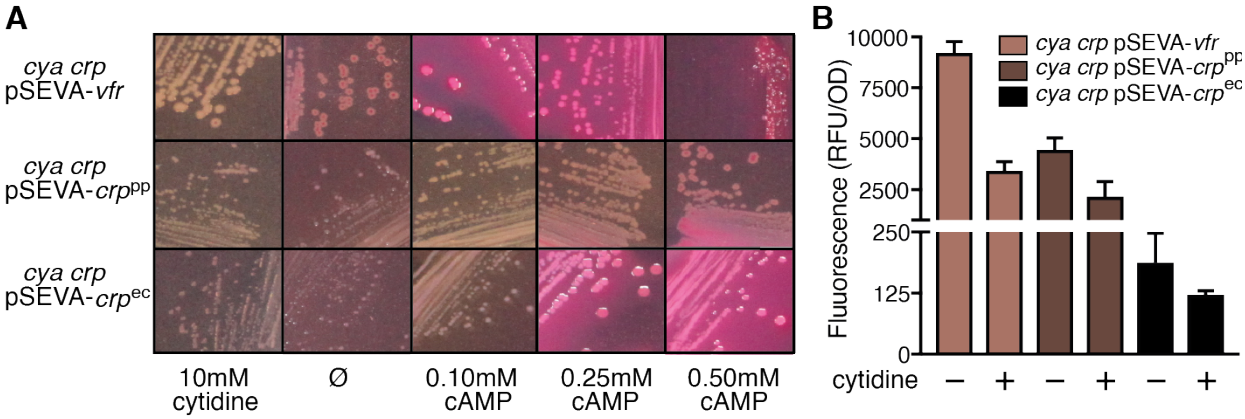


Fig. 4. *In vivo* characterization of the Crp homologues *Pseudomonas aeruginosa* Vfr and *Pseudomonas putida* Crp^{pp}. (**A**) *In vivo* phenotypic characterization of *crp* variants expressed from a low-copy plasmid in a *cya crp* strain plated on maltose MacConkey agar. The agar was supplied different concentrations of cAMP (positive controls), 10mM cytidine, or no additional ligand (Ø). *E. coli crp^{ec}* was expressed and assayed under identical conditions. (**B**) *In vivo* phenotypic characterization of *crp* variants expressed from a low-copy plasmid in combination with a P*malT*-GFP reporter in a *cya crp* strain. Fluorescence activity was assayed after nine hours of growth in liquid culture in the presence or absence of 10mM cytidine. Data represent the average of three biological replicates with standard deviations.

Discussion:

DNA, RNA, and proteins are the principal polymeric components of life and in all kingdoms of life, signaling pathways have evolved to balance their information flow and synthesis. Energy is harvested from carbohydrates and regulated by a process known as carbon catabolite repression. In *E. coli*, cAMP is a central signaling molecule in this process through its binding to the global transcription factor Crp. Two other purine nucleotides, (p)ppGpp, play a central role in the stringent response to nutrient availability and like cAMP impact an array of different microbial phenomena such as biofilm formation, persistence and virulence (*28*). The role of pyrimidines as signaling molecules is less well-described. Pyrimidine biosynthesis flow has been suggested to represent a pivotal sensing mechanism in bacteria, and build-up of pyrimidines may be a general cellular stress signal (*29*), but in contrast to the purines little is known about receptors responsible for pyrimidine signal processing. One exception is the transcription factor CytR that binds cytidine and controls the expression of a small set of genes involved in transport and utilization of nucleosides and deoxynucleosides in *E. coli* (*30*). Interestingly, CytR works in concert with Crp (*31*) and it will be interesting to learn how the cytidine binding sites relate to each other in this complex.

Crp is one the best studied transcription factors - mutations have been studied for almost five decades and more than 100 different mutants have been identified (*18*). The effects of uridine and cytidine on Crp are in the mM-range - a relevant physiological concentration at least for uridine (*24*). Possibly cytidine can reach the same levels during starvation. A physiological relevance is further supported by the negative effect of *umpH* expression on Crp activity. The frequent isolation of A144T mutations in Crp points to a prominent role for cytidine or uridine as Crp ligands, not just under the specific experimental conditions described here. The canonical A144T mutation converts Crp from being inhibited to being activated by cytidine and uridine. Mutations in *cmk* and the second site mutations that occur in *crp*A144T and confer independence from cytidine, support a hypothesis that cytidine builds up in the starving, ageing *cya* colonies. Build-up of nucleobases and nucleosides when DNA and RNA are in low demand is well-documented e.g. when *E. coli* enters into stationary phase (*32*), starve (*33*), and in a phenomenon known as directed overflow metabolism (*13*).

A simple interpretation of the data presented here is that high concentrations of cytidine and uridine downregulate global transcription as a general stress response by binding to Crp, thereby acting as the CMF previously searched for by Magasanik, Ullman, Monod and others (*7*, *8*, *34*). The regulatory mechanism of α-ketoacids via cAMP (*10*) and the inhibitory effect of pyrimidine nucleotides on Crp might work in concert in which sensing of anabolic processes are governed by α-ketoacids and sensing of catabolic processes by pyrimidines. This takes Crp to a new level as a ubiquitous regulator of transcription in prokaryotes, responding not only to carbon sources, but to the pool of available nucleotides. The universally conserved biosynthesis of DNA from RNA via CDP (Fig. 1B), and the finding that cytidine and uridine antagonize Crp homologs in bacteria beyond *E. coli*, indicates that in many living systems pyrimidine nucleosides act as signals rather than building blocks. The findings warrant further exploration of the mechanisms at play and the signaling roles of pyrimidine nucleosides in eukaryotic systems.

References and Notes:

1. F. Dienert, Sur la fermentation du galactose et sur l'accoutumance des levures à ce sucre. *Ann. Inst. Pasteur*. **19**, 139–189 (1900).
2. A. Ullmann, A. Danchin, Role of cyclic AMP in bacteria. *Adv. Cycl. Nucleotides Res.* **15**, 1–53 (1983).
3. R. L. Perlman, B. De Crombrugghe, I. Pastan, Cyclic AMP regulates catabolite and transient repression in E. coli. *Nature* (1969), doi:10.1038/223810a0.
4. M. Emmer, B. deCrombrugghe, I. Pastan, R. Perlman, Cyclic AMP Receptor Protein of E. coli: Its Role in the Synthesis of Inducible Enzymes. *Proc. Natl. Acad. Sci.* **66**, 480–487 (1970).
5. G. Zubay, D. Schwartz, J. Beckwith, Mechanism of Activation of Catabolite-Sensitive Genes: A Positive Control System. *Proc. Natl. Acad. Sci.* **66**, 104–110 (1970).
6. F. Jacob, J. Monod, Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
7. A. Ullmann, F. Tillier, J. Monod, Catabolite modulator factor: A possible mediator of catabolite repression in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 3476–3479 (1976).
8. A. Dessein, F. Tillier, A. Ullmann, Catabolite modulator factor: Physiological properties and in vivo effects. *MGG Mol. Gen. Genet.* (1978), doi:10.1007/BF00333854.
9. J. Daniel, A. Danchin, 2-Ketoglutarate as a possible regulatory metabolite involved in cyclic AMP-dependent catabolite repression in Escherichia coli K12. *Biochimie*. **68**, 303–310 (1986).
10. C. You, H. Okano, S. Hui, Z. Zhang, M. Kim, C. W. Gunderson, Y. P. Wang, P. Lenz, D. Yan, T. Hwa, Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature*. **500**, 301–306 (2013).
11. R. M. Schaaper, C. K. Mathews, Mutational consequences of dNTP pool imbalances in E. coli. *DNA Repair (Amst).* **12**, 73–79 (2013).
12. J. Martinussen, M. Willemoës, M. Kilstrup, in *Comprehensive Biotechnology, Second Edition* (2011).
13. M. L. Reaves, B. D. Young, A. M. Hosios, Y. F. Xu, J. D. Rabinowitz, Pyrimidine homeostasis is accomplished by directed overflow metabolism. *Nature* (2013), doi:10.1038/nature12445.
14. A. Danchin, Comparison between the Escherichia coli and Bacillus subtilis genomes

suggests that a major function of polynucleotide phosphorylase is to synthesize CDP. *DNA Res.* (1997), doi:10.1093/dnares/4.1.9.

15. A. Sekowska, S. Wendel, E. C. Fischer, M. H. H. Nørholm, A. Danchin, Generation of mutation hotspots in ageing bacterial colonies. *Sci. Rep.* **6** (2016), doi:10.1038/s41598-016-0005-4.
16. H. Aiba, T. Nakamura, H. Mitani, H. Mori, Mutations that alter the allosteric nature of cAMP receptor protein of Escherichia coli. *EMBO J.* **4**, 3329–3332 (1985).
17. S. Garges, S. Adhya, Sites of allosteric shift in the structure of the cyclic AMP receptor protein. *Cell.* **41**, 745–751 (1985).
18. P. O. Frendorf, I. Lauritsen, A. Sekowska, A. Danchin, M. H. H. Nørholm, Mutations in the Global Transcription Factor CRP/CAP: Insights from Experimental Evolution and Deep Sequencing. *Comput. Struct. Biotechnol. J.* **17** (2019), doi:10.1016/j.csbj.2019.05.009.
19. I. T. Weber, G. L. Gilliland, J. G. Harman, A. Peterkofsky, Crystal structure of a cyclic AMP-independent mutant of catabolite gene activator protein. *J. Biol. Chem.* **262**, 5630–5636 (1987).
20. J. Fricke, J. Neuhard, R. A. Kelln, S. Pedersen, The cmk gene encoding cytidine monophosphate kinase is located in the rpsA operon and is required for normal replication rate in Escherichia coli. *J. Bacteriol.* (1995), doi:10.1128/jb.177.3.517-523.1995.
21. C. Chapon, *J. Bacteriol.*, in press.
22. M. Mangel, F. J. Samaniego, Abraham Wald's work on aircraft survivability. *J. Am. Stat. Assoc.* (1984), doi:10.1080/01621459.1984.10478038.
23. B. G. Hall, Spectrum of mutations that occur under selective and non-selective conditions in E. coli. *Genetica* (1991), doi:10.1007/BF00116545.
24. B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, J. D. Rabinowitz, Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* (2009), doi:10.1038/nchembio.186.
25. P. Milanesio, A. Arce-Rodríguez, A. Munoz, B. Calles, V. De Lorenzo, Regulatory exaptation of the catabolite repression protein (Crp)-cAMP system in Pseudomonas putida. *Environ. Microbiol.* (2011), doi:10.1111/j.1462-2920.2010.02331.x.
26. V. De Crecy-Lagard, P. Glaser, P. Lejeune, O. Sismeiro, C. E. Barber, M. J. Daniels, A.

Danchin, A Xanthomonas campestris pv. campestris protein similar to catabolite activation factor is involved in regulation of phytopathogenicity. *J. Bacteriol.* (1990), doi:10.1128/jb.172.10.5877-5883.1990.

27. E. L. Fuchs, E. D. Brutinel, A. K. Jones, N. B. Fulcher, M. L. Urbanowski, T. L. Yahr, M. C. Wolfgang, The Pseudomonas aeruginosa Vfr regulator controls global virulence factor expression through cyclic AMP-dependent and -independent mechanisms. *J. Bacteriol.* (2010), doi:10.1128/JB.00363-10.
28. K. Potrykus, M. Cashel, (p)ppGpp: Still Magical? *Annu. Rev. Microbiol.* (2008), doi:10.1146/annurev.micro.62.081307.162903.
29. E. Rossi, S. Motta, A. Aliverti, F. Cossu, L. Gourlay, P. Mauri, P. Landini, Cellulose production is coupled to sensing of the pyrimidine biosynthetic pathway via c-di-GMP production by the DgcQ protein of Escherichia coli. *Environ. Microbiol.* (2017), doi:10.1111/1462-2920.13918.
30. P. Gerlach, P. Valentin-Hansen, E. Bremer, Transcriptional regulation of the cytR repressor gene of Escherichia coli: autoregulation and positive control by the cAMP/CAP complex. *Mol. Microbiol.* (1990), doi:10.1111/j.1365-2958.1990.tb00614.x.
31. K. Hammer-Jespersen, in *Metabolism of nucleotides, nucleosides and nucleobases in microorganisms. Munch-Petersen, A. (ed.). London:Academic Press* (1983), pp. 203–258.
32. U. Rinas, K. Hellmuth, R. Kang, A. Seeger, H. Schlieker, Entry of Escherichia coli into stationary phase is indicated by endogenous and exogenous accumulation of nucleobases. *Appl. Environ. Microbiol.* **61**, 4147–4151 (1995).
33. M. Lempp, N. Farke, M. Kuntz, S. A. Freibert, R. Lill, H. Link, Systematic identification of metabolites controlling gene expression in E. coli. *Nat. Commun.* **10** (2019), doi:10.1038/s41467-019-12474-1.
34. B. Magasanik, Catabolite repression. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 249–256 (1961).
35. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–5 (2000).
36. S. Datta, N. Costantino, D. L. Court, A set of recombineering plasmids for gram-negative bacteria. *Gene.* **379**, 109–115 (2006).
37. A. M. Cavaleiro, S. H. Kim, S. Seppälä, M. T. Nielsen, M. H. H. Nørholm, Accurate

DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth. Biol.* **4**, 1042–1046 (2015).

38. R. Chait, S. Shrestha, A. K. Shah, J. B. Michel, R. Kishony, A differential drug screen for compounds that select against antibiotic resistance. *PLoS One*. **5**, 1–8 (2010).
39. R. Silva-Rocha, E. Martínez-García, B. Calles, M. Chavarría, A. Arce-Rodríguez, A. De Las Heras, A. D. Páez-Espino, G. Durante-Rodríguez, J. Kim, P. I. Nikel, R. Platero, V. De Lorenzo, The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41**, 666–675 (2013).

Acknowledgements: We would like to thank Peter Brodersen and Gunnar von Heijne for comments on the manuscript and Mogens Kilstrup, Joshua Brickman, and Martin Willemoës for feedback and discussions on the work.

Funding: This work was supported by grants from the Novo Nordisk Foundation and the Technical University of Denmark.

Author contributions: Ida Lauritsen: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Pernille Ott Frendorf: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Silvia Capucci: Conceptualization, Methodology, Resources, Investigation, Writing - Review & Editing. Sofie Wendel: Conceptualization, Resources, Investigation, Writing - Review & Editing. Emil C. Fischer: Conceptualization, Resources, Writing - Review & Editing. Agnieszka Sekowska: Conceptualization, Writing - Review & Editing. Antoine Danchin: Conceptualization, Writing - Review & Editing. Morten H. H. Nørholm: Conceptualization, Methodology, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Competing interests: Authors declare no competing interests.

Data and materials availability: All data is available in the main text or the supplementary materials.

Supplementary materials:

Materials and Methods

Table S1 to S3

Figures S1 to S3

Movie S1

Supplementary Materials for

Temporal evolution of master regulator Crp identifies pyrimidines as catabolite modulator factors

Ida Lauritsen, Pernille Ott Frendorf, Silvia Capucci, Sofie Wendel, Emil C. Fischer, Agnieszka Sekowska, Antoine Danchin, Morten H. H. Nørholm

Correspondence to: morno@biosustain.dtu.dk

This PDF file includes:

Materials and Methods
Figs. S1 to S3
Tables S1 to S3
Movie S1

Other Supplementary Materials for this manuscript include the following:

Movie S1

Materials and Methods

Bacterial strains, media and growth conditions

E. coli strains were grown in lysogeny broth (LB) with shaking at 250 rpm or supplemented with agar for growth on plates. Strains were incubated at 37°C unless otherwise mentioned. All strains used in this study are described in Supplementary information Table S2. The antibiotics ampicillin (100 µg/ml), spectinomycin (50 µg/ml), kanamycin (50 µg/ml), or tetracycline (10 µg/ml) were added when needed. MacConkey agar was purchased from BD Diagnostic, Difco MacConkey Agar Base (281810). Simple MacConkey (SM) agar media composition for 1 liter was: Peptone from soybean meat (70178-100G Sigma Aldrich, 5.1g), Protease peptone (P0431-250G Sigma Aldrich, 0.9 g), NaCl (Sigma Aldrich, 1.5 g), and Agar (Sigma Aldrich, 4.05 g).

The *crp* gene was deleted in Escherichia coli K-12 MG1655 *cyaA::cat Δfnr* and in Escherichia coli K-12 MG1655 *cyaA::cat Δfnr cmkA216E* by lambda-red recombineering with pSIM19 as previously described (*35*, *36*)

Plasmid construction

All plasmid constructs were generated by PCR and USER cloning, as described previously (*37*), and verified by sequencing. For further details on all the plasmids described here see Table S2. PCR amplification was performed with either plasmids or genomic DNA as template. All oligonucleotides used for plasmid constructions are shown in Table S3. The P*trc* promoter in pSEVA27-Crp^{ec} and pSEVA-Ptrc-*crp*A144T was exchanged with the P*crp* promoter, amplified from genomic DNA with oligonucleotides #4119 and #4120. For construction of pSEVA27-Crp^{pp} and pSEVA27-Vfr, Crp^{pp} was amplified with #4235 and #4236 from pGEM-Pcrp-Crp^{pp} and Vfr with #4571 and #4572 from genomic DNA isolated from *P. aeruginosa* PA14, provided by Dr. Elio Rossi. For construction of pGEM-PX-hp-sfGFP, the pGEM backbone with sfgfp-ssrA was amplified using #4255 and #4257, and the promoters were amplified with #4256 and #4258 (PmalT) and #4529 and #4530 (Plac).

For Crp protein production using pET52b-*crp*, *crp*A144T, the pET52b backbone was amplified with #2697 and #2698 and crp was amplified from genomic DNA of wildtype *crp* or *crp*A144T mutant strains using #2710 and #2711. The antibiotic marker was exchanged with

an ampicillin resistance marker in pZE21-sfGFP and pZE21-RFP with #4518 and #4519 and backbone amplification with #4516 and #4517.

Strain characterization on MacConkey agar

Strains were first streaked on LB agar supplemented with appropriate antibiotics from −80°C cryo stocks and the next day restreaked on MacConkey agar supplemented with 1% maltose and the appropriate antibiotics. Streaking on MacConkey agar was also performed directly from −80°C cryo. The MacConkey agar plates were further supplemented with either cAMP (0.5mM), cytidine (10mM), cytosine (10mM), cytidine monophosphate (10mM), or cytidine diphosphate (10mM).

Protein production

For *crp* expression, 100 µL chemically competent *E. coli* BL21(DE3) were transformed with 1 µL of pET52b-*crp* or pET52b-*crp*A144T and plated on LB agar. The following day, a colony was inoculated in a preculture of 10 ml LB with 0.04% glucose. The next day, 400 ml LB cultures were inoculated 1:100 with preculture and grown to OD600 0.3-0.5, followed by induction of Crp expression using 1 mM IPTG and 1:800 protease inhibitor. After induction for 4-5 hours, the biomass was pelleted at 4°C (6000g, 10 min), the supernatant discarded, and the pellet saved in stored at −80°C overnight. Lysis buffer (50 mM NaH2PO4, 300 mM NaCl, 10 mM imidazole, pH 8) was mixed with lysozyme, benzonase and protease inhibitor 1:100 immediately before use, and 2 ml was used for resuspension of the cell pellet. The suspension was then put on ice for 2 hours, centrifuged at 4°C (6000g, 30 min), and the supernatant was collected.

The protein was purified from the supernatant using the Ni-NTA Spin Columns (QIAGEN, Germany) according to the manufacturers protocol, with minor modifications. Equilibration was completed using 5 culture volumes (CV) of lysis buffer, bound lysate was washed with 10 CV of lysis buffer and 15 CV of buffer NPI-20, and elution was completed with 5.5 CV of buffer NPI-500. Centrifugation was not applied, and flow through was collected for all steps. An SDS-PAGE was applied to the collected flow through to verify protein production and to determine the concentration of protein, and if necessary, the

protein was concentrated using the Amicon Ultra-0.5 Ultracel-10 Centrifugal Filter Units (Merck, Germany) according to the manufacturers protocol.

Biolayer interferometry

To study Crp interactions, SA biosensors (ForteBio, USA) were applied using the Octet RED96 system (ForteBio, USA) to assay DNA binding. Prior to use, fresh biosensors were hydrated in 200 µL PBS buffer (from PBS (10X), pH 7.4, GibcoTM) for 10 minutes at 30°C and 1000 rpm. Assays were run in 96-well plates with one row per sample and protocol steps in separate columns. Biotinylated DNA for the SA biosensors (P*malT*) was generated by annealing oligonucleotides #3915 and #3916 for 20 minutes at 70°C. To load the DNA, the SA biosensors were subjected to PBS for a baseline (60 s) before the loading of 20 nM biotinylated DNA (600 s) and quenching using 10 µg/ml biocytin (60 s). The saturated SA biosensors were then subjected to the applied protein buffer PBS-TB (PBS with 0.2% Tween-20 and 0.1% bovine serum albumin) for another baseline (60 s), followed by association using Crp (250 nM) with or without additional analytes (60 s) and dissociation in PBS-TB (60 s).

Growth experiments

Cells were streaked from −80°C cryo on LB agar with appropriate antibiotics. Next day, 3 colonies were used for inoculation in 5 ml LB (supplemented with appropriate antibiotics) for biological replicates and grown for 5-7 hours. OD600 was adjusted to 1 and 1 µl was used for inoculation in 149 µl medium supplemented with MilliQ water, 0.5mM cAMP, 10mM cytidine, or 10mM uridine in a 96-well plate (Costar 96 flat bottom) with a breath-seal. The plate was incubated in Synergy H1 Microplate reader from BioTek for 24-48 hours at 37°C orbital shaking continuously at 425 cpm (3mm). Optical density (600 nm) and GFP fluorescence (excitation 485 nm, emission 528 nm, gain 50-100, measured from the bottom) was measured every 10 min.

Competition assays

Competition assays were performed with fluorescently labeled strains by growth on agar plates as described previously (*38*). MG1655 K-12 *cya crp* strains carrying plasmid pSEVA27-crpA144T, T140K, or G141D and pZE21-sfGFP or pZE21-RFP were inoculated overnight in 5 ml LB supplemented with ampicillin, kanamycin and 0.4% glucose. The next day, the cultures were diluted to OD600 2 in LB and plated on SM-Maltose agar supplemented with ampicillin and kanamycin, mixed in ratios of 1:10 and 1:1000. Pure strains were also plated. Dilution of cells before plating was done in 9g/L NaCl water. A 20 µl drop of 0.5 M cytidine was placed in the middle of the plate and plates were incubated overnight at 37°C. For quantification, cells were scraped off, resuspended in 9g/L NaCl water and fluorescence was measured in the Synergy H1 Microplate reader from BioTek. GFP (excitation 485 nm, emission 528 nm) with gain 100 and RFP (excitation 587 nm, emission 610 nm) fluorescence with gain 90 was measured. The fluorescence values derived from plates with colonies from one strain were used as reference for quantification (100%).

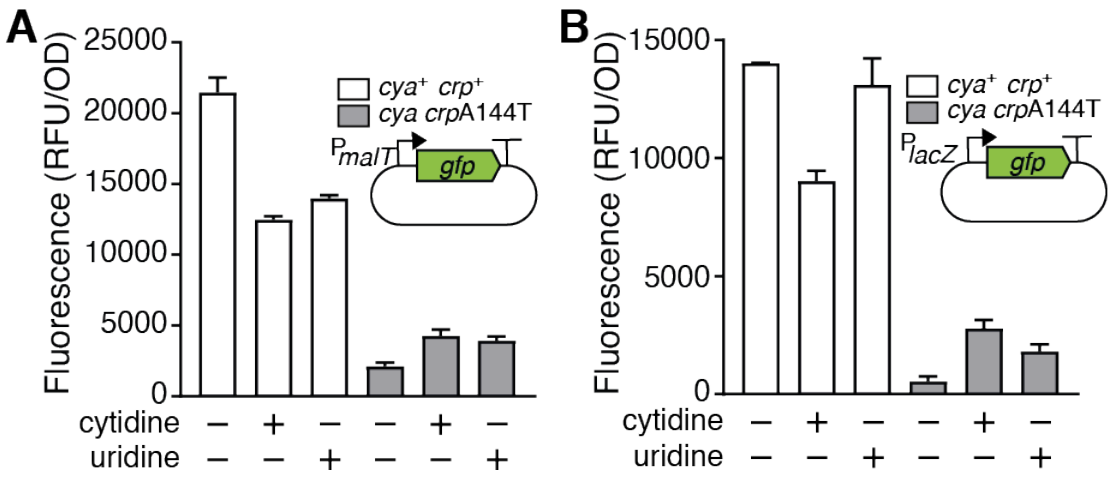


Fig. S1. *In vivo* Crp activity measured with different CRP responsive reporters in the presence of the nucleosides cytidine or uridine. (**A**) Crp activity reporter with the P*malT* promoter controlling expression of *gfp*, or (**B**) P*lacZ*. Crp activity was measured in a wildtype *crp* or *crp*A144T strain background after nine hours of growth. Data represent the average of three biological replicates with standard deviations.

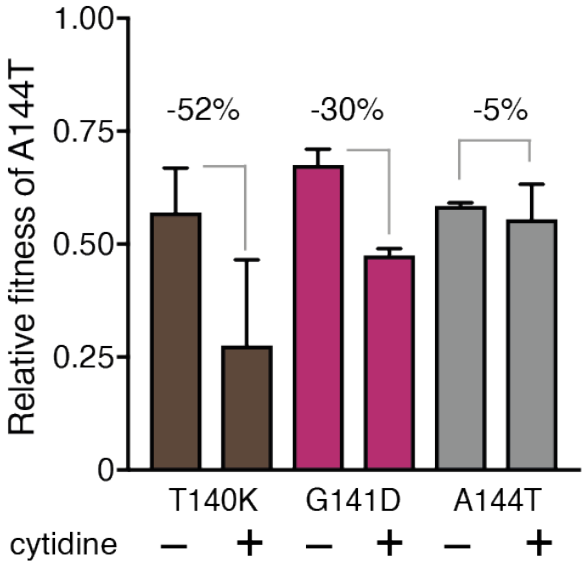


Fig. S2. Competition assays between CrpA144T and CrpT140K or G141D. Relative fitness, measured as RFP fluorescence, of Crp* mutations T140K and G141D competed with A144T by growth on agar plates with maltose in the presence or absence of cytidine. The reporter plasmids expressing either *rfp* or *gfp* of the competing strains are switched compared to data presented in Fig. 3 (**E**).

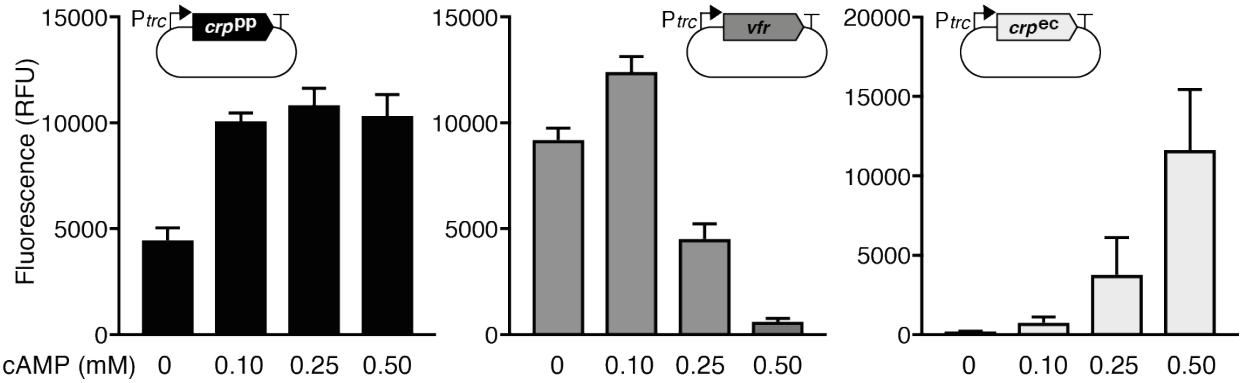


Fig. S3. *In vivo* characterization of Crp homologues. *crp* variants from *Pseudomonas putida* crp^{pp} (black), *Pseudomonas aeruginosa vfr* (grey), and *E. coli* crp^{ec} (white) were expressed from a low-copy plasmid in combination with a P*malT*-GFP reporter in a *cya crp* strain. Fluorescence activity was assayed after nine hours of growth in liquid culture in the presence of different concentrations of cAMP. Data represent the average of three biological replicates with standard deviations.

Table S1. Crp mutants identified in 594 sequenced *crp* loci from ageing bacteria (*15*)

CRP mutation 1	type	CRP paired mutation	type	count
A144T	G->A			307
A144T	G->A	L11I	C->A	1
A144T	G->A	I30F	A->T	8
A144T	G->A	L39M	C->A	2
A144T	G->A	D53N	G->A	1
A144T	G->A	E55K	G->A	1
A144T	G->A	E55D	G->T	1
A144T	G->A	E55E	G->A	1
A144T	G->A	S62F	C->T	3
A144T	G->A	S62Y	C->A	1
A144T	G->A	L64Q	T->A	1
A144T	G->A	R82S	C->A	1
A144T	G->A	A84E	C->A	2
A144T	G->A	W85R	T->A	1
A144T	G->A	P110Q	C->A	1
A144T	G->A	T127I	C->T	2
A144T	G->A	N133H	A->C	1
A144T	G->A	T140K	C->A	2
A144T	G->A	Q145K	C->A	1
A144T	G->A	T146A	A->G	1
A144T	G->A	D155N	G->A	1
A144T	G->A	Q170K	C->A	5
A144T	G->A	M189K	T->A	1
A144T	G->A	M189I	G->A	1
A144T	G->A	Q193K	C->A	1
A144T	G->A	3 aa insertion		1

A144E	C->A			57
A144E	C->A	I30F	A->T	2
A144E	C->A	L39M	G->A	3
A144E	C->A	E55K	C->A	1
A144E	C->A	S62Y	C->A	4
A144E	C->A	A84E	C->A	1
A144E	C->A	Q119H	G->T	1
A144E	C->A	T140K	C->A	7
A144E	C->A	Q170K	C->A	7
A144E	C->A	Q174K	T->A	2
A144E	C->A	M189K	T->A	4
A144E	C->A	22 aa insertion		1
T140K	C->A			7
T140K	C->A	M189R	T->G	1
T140K	C->A	insertion of N between G56 and K57		1
T140R	C->G	T140R		16
T140R	C->G	Q119H	C->G	1
T140R	C->G	Q170K	C->G	1
T140R	C->G	V183A	T->C	1
M189K	T->A			3
M189K	T->A	Y63F	A->T	1
M189K	T->A	Q170K	C->A	1
insertion K57				11
insertion K57		A84E	C->A	2

S62Y	C->A		1
P110Q	C->A		2
L134M	C->A		1
T140P	A->C		1
G141S	G->A		1
G141D	G->A		9
A144K	G->A and C->A		12
L195R	T->G		4
insertion of N between G56 and K57			1
insertion of L between V86 and R87			1
insertion of V between V86 and R87			1
duplication 3 aa 86-88			1
duplication of 20 aa			2
duplication of 28 aa			1
No mutation			71
		SUM	594

Table S2. Strains and plasmids used in this study.

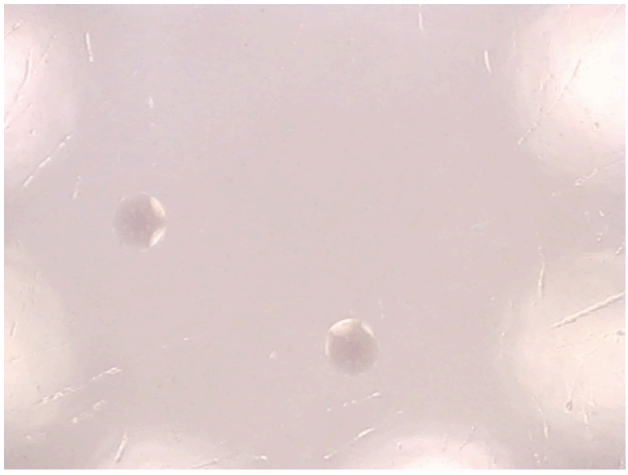
Strains	Description/relevant characteristics	Reference
WT	MG1655 K-12 WT	Our laboratory
cya	K-12 MG1655 *cyaA::cat* Δ*fnr*	(*15*)
crp	K-12 MG1655 *cyaA::cat* Δ*fnr* Δ*crp*	This study
*crp*A144T	MG1655 K-12 *cyaA::cat* Δ*fnr* - *crp*A144T	(15)
*cmk*A216E	MG1655 K-12 *cyaA::cat* Δ*fnr* Δ*crp cmk*A216E	This study
*cmk*A216E-*crp*	*MG1655 K-12 cyaA::cat Δfnr cmk*A216E	(*15*)
*crp*A144T-Q170K	MG1655 K-12 *cyaA::cat* Δ*fnr crp*A144T-Q170K	(*15*)
BL21(DE3)	F– *omp*T *hsdS*$_B$ (r_B–, m_B–) *gal dcm* (DE3)	ThermoFischer Scientific

Plasmids	Description/relevant characteristics	Reference
pSEVA27	pSC101 replicon, Kan^R	(*39*)
pCDF-*crp*	*crp* expression from *trc* promoter, CloDF13 replicon, Sp^R	Our laboratory
pCDF-*crp*A144T	*crp*A144T expression from *trc* promoter, CloDF13 replicon, Sp^R	Our laboratory
pSEVA-Ptrc-*crp*A144T	*crp*A144T, expression from *trc* promoter, pSC101 replicon, Kan^R	This study
pSEVA27-*crp*	*crp* expression from native *crp* promoter, pSC101 replicon, Kan^R	This study
pSEVA27-*crp*A144T	*crp*A144T expression from native *crp* promoter, pSC101 replicon, Kan^R	This study
pSEVA27-*crp*T140K	*crp*T140K expression from native *crp* promoter, pSC101 replicon, Kan^R	This study
pSEVA27-*crp*G141D	*crp*G141D expression from native *crp* promoter, pSC101 replicon, Kan^R	This study
pSEVA27-Crp^{ec}	*crp* from *E. coli,* expression from *trc* promoter, pSC101 replicon, Kan^R	This study
pGEM-Pcrp-Crp^{pp}	*crp* from *Pseudomonas putida*, expression from the *E. coli crp* promoter, ColE1 replicon, Amp^R	This study
pSEVA27-Crp^{pp}	*crp* from *Pseudomonas putida*, expression from *trc* promoter, pSC101 replicon, Kan^R	This study
pSEVA27-Vfr	*Vfr* expression from *trc* promoter, pSC101 replicon, Kan^R	This study
pZE21-sfGFP-Kan	*sfgfp* expression from *tet* promoter, constitutively (no *tetR),* ColE1 replicon Kan^R	Dr. Andreas Porse
pZE21-RFP-Kan	*rfp* expression from *tet* promoter, constitutively (no *tetR),* ColE1 replicon, Kan^R	Dr. Andreas Porse
pZE21-sfGFP	*sfgfp* expression from *tet* promoter, constitutively (no *tetR),* ColE1 replicon Amp^R	This study
pZE21-RFP	*rfp* expression from *tet* promoter, constitutively (no *tetR),* ColE1 replicon, Amp^R	This study
pGEM-Pcrp-crp-sfGFP	Transcriptional fusion of *crp* and *sfGFP* expressed from *crp* promoter ColE1 replicon, Kan^R, Amp^R	Our laboratory
pGEM-PmalT-hp-sfGFP	*sfgfp-ssrA*-pHP14 hairpin, expression from *malT* promoter, strong SD, ColE1 replicon, Kan^R, Amp^R	This study

pGEM-Pcrp-hp-sfGFP	*sfgfp-ssrA*-pHP14 hairpin, expression from *crp* promoter, strong SD, ColE1 replicon, Kan^R, Amp^R	This study
pGEM-Pcrp-negreg-hp-sfGFP	*sfgfp-ssrA*-pHP14 hairpin, expression from *crp* promoter deleted of negative autoregulation, strong SD, ColE1 replicon, Kan^R, Amp^R	This study
pGEM-PlacZ-hp-sfGFP	*sfgfp-ssrA*-pHP14 hairpin, expression from *lacZ* promoter, strong SD, ColE1 replicon, Kan^R,	This study
pGEM-PmalT-hp-sfGFP_noKan	*sfgfp-ssrA*-pHP14 hairpin, expression from *malT* promoter, strong SD, ColE1 replicon, Amp^R	This study
pSIM19	*beta, exo, gam* expression, pSC101 replicon $repA^{ts}$ temperature sensitive, $Spec^R$	(*36*)
pET52-*crp*	*crp* expression from T7 promoter, ColE1 replicon, Amp^R	This study
pET52-*crp*A144T	*crp*A144T expression from T7 promoter, ColE1 replicon, Amp^R	This study

Table S3. Oligonucleotides used in this study

Oligo ID	Name	Sequence (5'→3')
3838	lacI_crp_fw	AAGACTAGUTCACTGCCCGCTTTCCAG
3839	lacI_crp_rev	ATGCCTTUTTAACGAGTGCCGTAAACGACG
1267	pSEVA33-Rev	AAAGGCAUCAAATAAAACGAAAGGCTC
1270	pSEVA33-fwd	ACTAGTCTUGGACTCCTGTTGATAGATC
2697	pET52-rev	ATGGTATAUCTCCTTCTTAAAGTTAAACAAAATTATTTC
2698	pET52-his_fwd	ACCATCAUCACCATCACCACCAC
2710	Crp_fwd	AGATATACCAUGGTGCTTGGCAAACCGCAAAC
2711	Crp_rev	ATGATGGUGACGAGTGCCGTAAACGACGATG
3915	malT_cds	5'Biotin-GTTGTTAATAAAGATTTGGAATTGTGACACAGTGCAAATTCAGACACATAAAAAAACGTC
3916	malT_ts	GACGTTTTTTTATGTGTCTGAATTTGCACTGTGTCACAATTCCAAATCTTTATTAACAAC
4117	pSEVA_Pcrp_fw	ATGGTGCTUGGCAAACCG
4118	pSEVA_Pcrp_rev	AGTCGTATTAAUTTCCTAATGCAGGAGTC
4119	Pcrp_pSEVA_fw	attaatacgacUTTTGCTACTCCACTGCGTCAATTTTC
4120	Pcrp_pSEVA_rev	AAGCACCAUGCGCGGTTATC
4521	Ptrc-pSEVA_ppCRP_fw	AGGCAACCAUAGGTATAtCTCCTTCTTAAAGTTAAACAA
4254	ppCRP_pSEVA_rev2	ACCCGCTAGUAGGCATCAAATAAAACGAAAGGCTC
4255	pGEM_PmalT_fw	AATTGGGATUAGGCAGGGAGGAGTTG
4256	pGEM_PmalT_rv	AATCCCAATUCACTGGCCGTCGTTTTAC
4257	PmalT-php14_rv	AATCTCACUCGAGAGTCGACGTATTAATGGTTATAAGGTCGGCCAGAAAC
4258	php14-RBS-sfGFP_fw	AGTGAGATUGTTGACGGTACCGTATTTTCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGAGCAAAGGCGAAGAGCTGTTCACTG
4569	Ptrc-pSEVA27bb_Paeru_fw	AGCTACCAUAGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAG
4570	Ptrc-pSEVA27bb_Paeru_rev	AGCACCCAUAAAGGCATCAAATAAAACGAAAGGCTC
4235	ppCRP_cds_fw	ATGGTTGCCUCCGCCCTAC
4236	ppCRP_cds_rev	ACTAGCGGGUACCGTGGAC
4571	vfr_Paeru_fw	ATGGTAGCUATTACCCACACACC
4572	vfr_Paeru_rev	ATGGGTGCUGTTCAGCG
4516	pZE21_bb_fw	AAACGAUCCTCATCCTGTCTCTTGATC
4517	pZE21-bb_rev	AGCGGGACUCTGGGGTTC
4529	pGEM-PlacZ-fw	AATTGGGATUGCGCAACGCAATTAATGTGAG
4530	pGEM-PlacZ-rv	AATCTCACUCGAGAGTCGACGTTTCCTGTGTGAAATTGTTATCCGCTC
4518	AmpR_pZE21_fw	AGTCCCGCUTTACCAATGCTTAATCAGTGAGGCAC
4519	AmpR_pZE21_rev	AGGATCGTTUATGAGTATTCAACATTTCCGTGTCGC



Movie S1.Time lapse movie of papillae formation. A *cya crp*A144E strain was plated on MacConkey supplied with maltose. After initial colony formation, pictures of two single colonies were taken from the bottom of the plate every 6 min over a period of 5 days.

Paper 7

CRISPR-Cas9 - den revolutionerende gensaks

CRISPR-Cas9
- den revolutionerende gensaks

De fleste af os har hørt om det - enten i nyhederne, på nettet eller hos kollegaen. CRISPR-Cas9, det mest hypede klippeværktøj i den molekylære værktøjskasse i nyere tid. Men hvad er alt det *"fuzz about"*? Hvad er CRISPR-Cas9, og hvad er det, som gør CRISPR-Cas9 til det mest omtalte, revolutionerende og debatskabende fænomen i vores tidsalder?

Af Ida Lauritsen og Morten H.H. Nørholm, Danmarks Tekniske Universitet, The Novo Nordisk Center for Biosustainability

CRISPR-Cas9 revolutionerer biologien og har muliggjort, at vi nu meget effektivt og præcist kan "klippe og klistre" genetisk materiale ind og ud af et utal af organismer. Denne enestående egenskab stammer oprindeligt fra bakteriers immunforsvar, som beskytter mod fremmed DNA-indtrængen. Da kinesiske forskere i 2015 redigerede det genetiske materiale i menneskeceller fra et tidligt fosterstadie, skabte det stor debat verden over, men åbnede samtidig også muligheder for nye måder at bekæmpe genetiske sygdomme på. CRISPR-bølgen er over os og dette er kun begyndelsen.

"Hacking" af bakteriens immunforsvar

CRISPR-Cas9-systemet er oprindeligt en del af bakteriers immunforsvar og bruges til at genkende og tilintetgøre fremmede genetiske elementer ved at klippe dem i stykker [1]. Disse elementer kan f.eks. stamme fra virus, som angriber bakterier i naturen. Forkortelsen CRISPR står for *"Clustered Regularly Interspaced Short Palindromic Repeats"*, som hentyder til bestemte gentagende DNA-sekvenser, der er blevet observeret i det genetiske materiale fra forskellige bakterier [2]. I 2005 lavede forskere forbløffende opdagelser, da de fandt ud af, at sekvenser som lå mellem disse gentagede sekvenser, matchede DNA-sekvenser fra forskellige vira - og især de såkaldte phager, som angriber bakterier [3]. Derved fandt man ud af, at information omkring angribende fremmede phager - et genetisk mønster - bliver gemt i bakteriens DNA. Hvis phagen angriber igen, vil bakterien genkende den og have sit beredskab klar til forsvar.

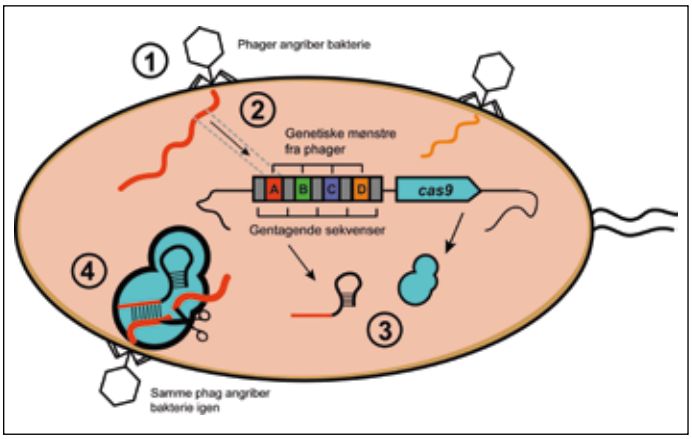
Men hvad er Cas9 så? Cas9 (*CRISPR-associated protein 9*) er det protein, der fungerer som den molekylære saks, der klipper det fremmede genetiske materiale i stykker, hvis bakterien angribes igen. Denne saks er så specifik, at den kun klipper, hvis den genkender et bestemt stykke DNA, som passer sammen med skabelonen i bakteriens DNA [4], figur 1. For at have et velfungerede CRISPR-Cas9 forsvarssystem i bakterien, er det nødvendigt at have to komponenter:
1) Cas9-gensaksen samt et tilhørende hjælperkompleks og
2) en guidesekvens, som leder Cas9 til den DNA-sekvens, som skal klippes.

I 2012 fik en forskergruppe i Californien en idé - hvad hvis man kunne "hacke" denne naturlige forsvarsmekanisme til at styre Cas9 og klippe bestemte DNA-sekvenser [5]?

Dette kunne man f.eks. bruge til at genmodificere organismer, så de kunne få nye, eller optimere, allerede-eksisterende egenskaber. Hurtigt derefter blev systemet testet i forskellige organismer, og det startede en biologisk revolution - en revolution, der muliggjorde, at man nu kunne modificere og ændre det genetiske arvemateriale yderst effektivt og med høj præcision. Og lige siden har det kun taget fart. Nu rider alle forskere inden for life-science med på CRISPR-bølgen.

Universel gensaks

Ud over at kunne klippe hvilken som helst DNA-sekvens, så er en af de mest fascinerende egenskaber ved CRISPR-Cas9-



Figur 1. CRISPR-Cas9-systemet som del af bakteriens immunforsvar.
(1) Phager angriber bakteriecellen og inficerer med deres genetiske elementer (rød og orange).
(2) Phagernes unikke genetiske mønster (A, B, C, D i forskellige farver) bliver indsat i bakteriens DNA med gentagende sekvenser imellem sig (grå). Her i DNA'et findes også genet, som koder for Cas9-proteinet.
(3) De unikke genetiske mønstre fra phagerne bliver lavet til en guidesekvens og samlet i et CRISPR hjælperkompleks. Samtidig bliver Cas9 (blå) proteinet produceret i cellen.
(4) Hvis den samme phag angriber bakterien igen, matcher guidesekvensen phagens inficerende genetiske element, som derved genkendes og klippes af Cas9.

systemet, at det fungerer i et utal af organismer. Dette er blandt andet bakterier som *Escherichia coli*, gær, svampe, planter, dyr - ja selv i menneskeceller [6]. Dette betyder altså, at systemet er universelt - ligesom DNA er. På Center for Biosustainability (CFB) på Danmarks Tekniske Universitet bruger vi dagligt CRISPR-Cas9 til at designe og optimere cellefabrikker. *"Cellefabrikker"* er celler fra forskellige organismer, der bruges som arbejdsheste til at producere forskellige værdifulde stoffer, som f.eks. medicin, brændstoffer eller kemikalier i industriel skala. Ved at udnytte cellefabrikker kan man producere store mængder af stoffer som måske naturligt kun forekommer i små mængder i den oprindelige organisme, under ekstreme eller ikke-miljøvenlige forhold.

Klassiske eksempler på sådanne stoffer er artemisinin og lycopen, som henholdsvis er et anti-malaria medikament og en vigtig antioxidant, der begge oprindeligt er produceret i langsomtgroende planter i begrænsede mængder [7].

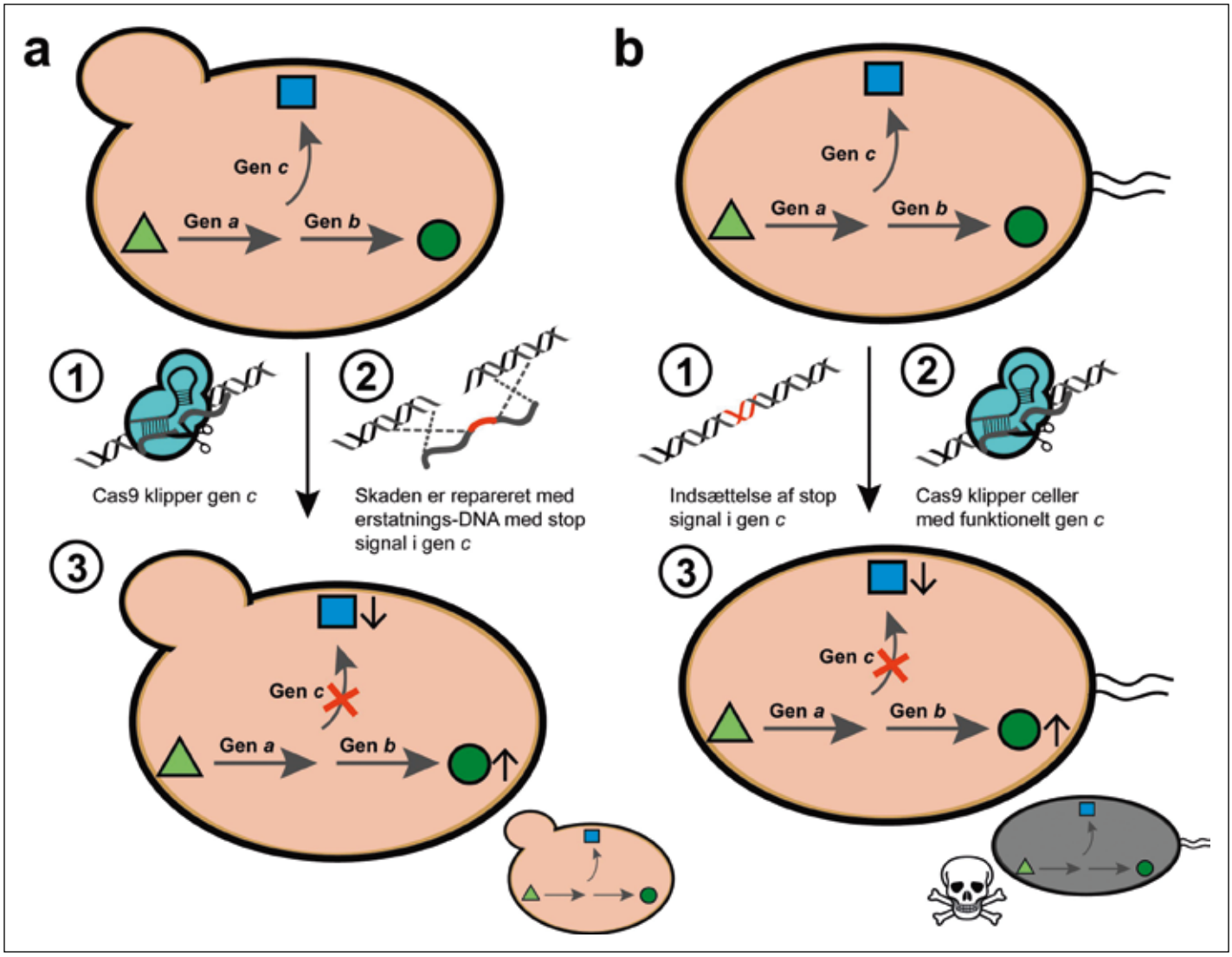
Et klip i DNA'et kan give en produktiv cellefabrik

Det genetiske arvemateriale for alle organismer består af ufattelig mange gener, der koder for proteiner. Disse proteiner kan være enzymer, der kan omdanne ét stof til et andet. For at kunne producere stoffer i cellefabrikker, som ikke naturligt kodes for i en organismes genetiske arvemateriale, må man "tilføje" nye enzymer, redigere eller fjerne de allerede-eksisterende. Dette kræver molekylære værktøjer, som kan modificere de gener, der koder for enzymerne. Værktøjer til at udføre genetisk redigering har eksisteret længe, men det som gør CRISPR-Cas9-systemet så enestående til dette formål, er, at genetisk redigering kan gøres med meget stor præcision og er billigt. Der er ikke behov for dyre maskiner, og teknologien er tilgængelig for alle forskere - da CRISPR-Cas9-komponenterne kan bestilles på internettet.

CRISPR-Cas9-systemet klipper ved en bestemt kort DNA-sekvens, styret af en guidesekvens. Et klip i vores genetiske arvemateriale - DNA'et - forårsaget af CRISPR-Cas9 kan være fatalt for de fleste organismer. Men hvordan udnyttes systemet så til at kunne ændre det genetiske materiale?

Dette afhænger af måden, som cellen reparerer skaden i DNA'et på. Der findes forskellige måder at reparere skaden, hvor den mest velkendte involverer brugen af et DNA-erstatningsstykke, som passer ind i den position, hvor klippet er sket [8]. Gær er en eukaryot (organisme med en cellekerne) og mester til at bruge stykker af erstatnings-DNA til at fikse et klip. Hermed kan vi indsætte en ny DNA-sekvens eller fjerne en sekvens i den position, hvor klippet er sket, figur 2a, side 16. Med CRISPR-Cas9 kan vi derfor "klippe og klistre" de gener, vi gerne vil have ind og ud af organismens DNA - vi kan lave genetisk redigering.

Anvendelsen af CRISPR-Cas9-systemet til genetisk redigering er dog lidt anderledes i de fleste bakterier, som ofte ikke er gode til at reparere deres DNA - de dør af et klip i det genetiske materiale [9]. Dette skyldes blandt andet, at bakterier er prokaryoter (uden cellekerne) og mindre komplekse. Vi kombinerer derfor andre metoder med CRISPR-Cas9 for at kunne lave genetisk redigering i bakterier, hvor CRISPR-Cas9 bruges til at fjerne de celler, som ikke har fået ændret deres genetiske materiale, figur 2b, side 16. På denne måde kan vi ▶



Figur 2. Brugen af CRISPR-Cas9 i cellefabrikker til produktion af værdifulde stoffer. Både gær- og bakteriecellen indeholder generne (*a* og *b*), der koder for enzymer, som omdanner substratet (illustreret som en lysegrøn trekant) til et værdifuldt stof (mørkegrønne cirkel). Dog har begge disse organismer et enzym (kodet af gen *c*), som omdanner substratet til et uønsket biprodukt (blå firkant). For at undgå at producere mindre af det værdifulde stof, må enzymet, kodet af gen *c*, inhiberes. Genetisk redigering bruges derfor til at indsætte et stopsignal i gen *c*, så det ikke kan producere et aktivt enzym.
a. (1) I gærcellen bruges CRISPR-Cas9 til at klippe i gen *c*, (2) og skaden repareres med et stykke erstatnings-DNA, hvor et stopsignal (fremhævet i rød) er indsat. (3) Med et inhiberet gen *c* (rødt kryds) producerer cellen mindre af biproduktet (pil ned ved blå firkant) og mere af det værdifulde stof (pil op ved mørkegrøn cirkel). Gærceller, som ikke har brugt erstatnings-DNA'et til reparation af skaden, kan f.eks. fikse skaden ved at klistre DNA'et sammen igen, men disse producerer lige så meget værdifuldt stof som den oprindelige celle.
b. (1) I bakteriecellen bruges bestemte molekylære metoder til at indsætte et stopsignal i gen *c*, (2) og CRISPR-Cas9 bruges her til at fjerne de celler, som stadig har et funktionelt gen *c*. (3) Cellen med et inhiberet gen *c* producerer mere værdifuldt stof (pil op) og mindre af biproduktet (pil ned). Bakterieceller, som stadig har et funktionelt gen *c*, vil blive klippet i DNA'et og dø på grund af mangel på effektiv DNA-reparation (grå bakteriecelle).

lettere finde de celler, som er blevet ændret. Som eksempel skal man uden CRISPR-Cas9-systemet søge 100 celler for at finde én celle, som er korrekt genetisk modificeret, hvorimod man med CRISPR-Cas9 kun skal lede 10 igennem eller færre [10]. I vores forskningsgruppe arbejder vi med bakterier og bruger CRISPR-Cas9-systemet, når vi laver genetisk redigering for at skabe bestemte bakteriestammer, hvor f.eks. et gen er inaktiveret. Dermed kan vi studere betydningen af enzymet, som netop dette gen koder for. Vi gør dette for at udvikle og teste nye molekylære værktøjer til brug i cellefabrikker samt for at studere gener, som kan være vitale for cellens evne til at mutere sit genetiske materiale for at tilpasse sig et bestemt nyt miljø.

Hvad er næste skridt?

Umiddelbart lyder CRISPR-Cas9-systemet til at have åbnet utallige døre inden for biologien. Systemet er veltestet i både bakterier, dyreceller samt gær, men bliver stadig optimeret og videreudviklet. Nu findes der sågar versioner af systemet, som kan bruges til at op- og nedregulere udtrykkelsen af bestemte gener uden at klippe dem, men hvor man udnytter den specifikke DNA-findende egenskab [11,12]. Udover brugen i cellefabrikker, kan man bruge CRISPR-Cas9-systemet til f.eks. at lave planter, der er mere robuste i ekstreme miljøer som tørke, mere næringsrige grøntsager der smager af mere samt undersøge grundlæggende spørgsmål vedrørende ældning, evolution og udviklingen af kræft [13].

Et af de mest omdiskuterede CRISPR-Cas9-baserede eksperimenter fra 2015 involverede en kinesisk forskningsgruppe, som formåede at fjerne et sygdomsgen i menneskeceller fra et tidligt fosterstadie i et reagensglas [14].

Dette eksperiment startede etiske diskussioner vedrørende brugen af CRISPR-Cas9. Men uanset hvor kontroversielt dette eksperiment var, viste det tydeligt, at teknologien kan bruges til at modificere sygdomsgener i de tidligere stadier af udviklingen, og allerede i 2016 gav det amerikanske medicinske forskningsråd *National Institutes of Health* grønt lys til at påbegynde brugen af CRISPR-Cas9 i menneskeforsøg til bekæmpelse af kræft [15].

Denne udvikling kan potentielt vise sig at have stor betydning for, hvordan vi kommer til at bekæmpe og behandle genetiske sygdomme fremover.

Hvordan fremtiden ser ud vedrørende brugen af CRISPR-Cas9 vides ikke, men hvad der er ganske sikkert, er, at CRISPR-Cas9 har ændret biologien, og hvor lang tid vi fortsat vil ride på CRISPR-bølgen, kan kun tiden vise.

E-mail:
Ida Lauritsen: idla@biosustain.dtu.dk
Morten H. H. Nørholm: morno@biosustain.dtu.dk

Referencer

1. R. Barrangou and J. Van Pijkeren, "Exploiting CRISPR – Cas immune systems for genome editing in bacteria," *Curr. Opin. Biotechnol.*, vol. 37, pp. 61-68, 2016.
2. R. Jansen, J.D.A. Van Embden, W. Gaastra, and L.M. Schouls, "Identification of genes that are associated with DNA repeats in prokaryotes," *Mol. Microbiol.*, vol. 43, no. 6, pp. 1565-1575, 2002.
3. A. Bolotin, B. Quinquis, A. Sorokin, and S. Dusko Ehrlich, "Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin," *Microbiology*, vol. 151, no. 8, pp. 2551-2561, 2005.
4. J.D. Sander and J.K. Joung, "CRISPR-Cas systems for editing, regulating and targeting genomes," *Nat. Biotechnol.*, vol. 32, no. 4, pp. 347–355, 2014.
5. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier, "A Programmable Dual-RNA - Guided DNA Endonuclease in Adaptice Bacterial Immunity," *Science*, vol. 337, no. August, pp. 816-822, 2012.
6. V. Singh, D. Braddick, and P.K. Dhar, "Exploring the potential of genome editing CRISPR-Cas9 technology," *Gene*, vol. 599, pp. 1-18, 2017.
7. J.A. Chemler and M.A. Koffas, "Metabolic engineering for plant natural product biosynthesis in microbes," *Curr. Opin. Biotechnol.*, vol. 19, no. 6, pp. 597-.605, 2008.
8. Z. Liu, Y. Liang, E.L. Ang, and H. Zhao, "A New Era of Genome Integration - Simply Cut and Paste!," *ACS Synth. Biol.*, vol. 6, no. 4, pp. 601-609, 2017.
9. J.M. Peters, M.R. Silvis, D. Zhao, J.S. Hawkins, C.A. Gross, and L.S. Qi, "Bacterial CRISPR: Accomplishments and prospects," *Curr. Opin. Microbiol.*, vol. 27, pp. 121-126, 2015.
10. C. Ronda, L.E. Pedersen, M.O.A. Sommer, and A.T. Nielsen, "CRMAGE: CRISPR Optimized MAGE Recombineering.," *Sci. Rep.*, vol. 6, p. 19452, 2016.
11. L.A. Gilbert, M.A. Horlbeck, B. Adamson, J.E. Villalta, Y. Chen, E.H. Whitehead, C. Guimaraes, B. Panning, H.L. Ploegh, M.C. Bassik, L.S. Qi, M. Kampmann, and J.S. Weissman, "Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation," *Cell*, vol. 159, no. 3, pp. 647-661, 2014.
12. L.S. Qi, M.H. Larson, L.A. Gilbert, J.A. Doudna, J. S. Weissman, A.P. Arkin, and W.A. Lim, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.," *Cell*, vol. 152, no. 5, pp. 1173-83, 2013.
13. V. Lau and J.R. Davie, "The discovery and development of the CRISPR system in applications in genome manipulation," *Biochem. Cell Biol.*, vol. 8, no. October, pp. 1-8, 2016.
14. P. Liang, Y. Xu, X. Zhang, C. Ding, R. Huang, Z. Zhang, J. Lv, X. Xie, Y. Chen, Y. Li, Y. Sun, Y. Bai, Z. Songyang, W. Ma, C. Zhou, and J. Huang, "CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes," *Protein Cell*, vol. 6, no. 5, pp. 363-372, 2015.
15. S. Reardon, "First CRISPR clinical trial gets green light from US panele," *Nature News*, 2016. [Online]. Available: http://www.nature.com/news/first-crispr-clinical-t.