Molecular tool development for conditional protein degradation in the model bacterium Escherichia coli

Hobel, Tonja

Publication date:
2018

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

Novo Nordisk Foundation
Center for Biosustainability
Technical University of Denmark

www.biosustain.dtu.dk

Tonja Hobel

Molecular tool development for conditional protein degradation in the model bacterium *Escherichia coli*

Molecular tool development for conditional protein degradation in the model bacterium *Escherichia coli*

Tonja Hobel
PhD thesis
May 2018

DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

Molecular tool development for conditional protein degradation in the model bacterium *Escherichia coli*

PhD Thesis

Tonja Hobel

Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark

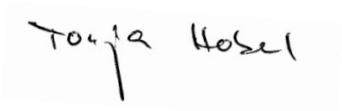
May 2018

Si les moustiques étaient des abeilles,
ils ramèneraient du sang à la ruche
et la reine ferait du boudin.

PHILIPPE GELUCK
Belgian comedian, humorist and cartoonist

PREFACE

This thesis is written as a partial fulfilment of the requirements to obtain a PhD degree at the Technical University of Denmark. The practical work of this thesis was carried out between January 2014 and March 2018 at the Novo Nordisk Foundation Center for Biosustainability in the laboratory of Senior Scientist Morten HH Nørholm. Funding was provided by the Novo Nordisk Foundation.



Tonja Hobel

Kongens Lyngby

May 2018

TABLE OF CONTENTS

ABSTRACT

Protein turnover is the balance between protein biosynthesis and protein degradation in living cells. Protein degradation is thereby not only a final destiny but also an essential and carefully regulated mechanism to maintain cell fitness. Even though the key players for protein degradation are known for long, only few aspects of protein recognition, unfolding and degradation have been characterised mechanistically in bacteria.

A dynamic regulation of protein stability is a key interest in applications in synthetic biology, biotechnology, metabolic engineering and fundamental research. However, only a few tools for conditional protein degradation have been developed in bacteria so far. This thesis presents two attempts to improve the available tool box for conditional degradation of soluble proteins as well as membrane proteins in *E. coli*.

DANSK SAMMENFATNING

Proteinomsætning er balancen mellem proteinsyntese og proteinnedbrydning i levende celler. Proteinnedbrydning er dermed ikke kun en endelig skæbne men også en essentiel og omhyggeligt reguleret mekanisme til at opretholde celle fitness. Selvom nøglemekanismerne for proteinnedbrydning har været kendt længe er kun få aspekter af protein genkendelse, denaturering og nedbrydning blevet karakteriseret mekanistisk i bakterier.

En dynamisk regulering af proteinstabilitet er en hovedinteresse i implementeringen af syntetisk biologi, bioteknologi, teknisk metabolisme og grundvidenskab. Men for nuværende er der kun blevet udviklet få værktøjer til konditionel proteinnedbrydning i bakterier. Denne afhandling præsenterer to forsøg på at forbedre den tilgængelige værktøjskasse til konditionel nedbrydning af både cytoplasmiske og membran associerede proteiner i *E. coli*.

LIST OF PUBLICATIONS

This thesis is partly based upon the following publication (Study I):

1 Virginia Martínez*, Ida Lauritsen*, **Tonja Hobel**, Songyuan Li, Alex Toftgaard Nielsen and Morten H. H. Nørholm (2017) CRISPR/Cas9-based genome editing for simultaneous interference with gene expression and protein stability. *Nucleic Acids Research*. Vol. 45: No. 20 e171 (DOI 10.1093/nar/gkx797)[#]

Published work that is not included in this thesis:

2 Ulla Christensen*, Dario Vazquez-Albacete*, Karina M. Søgaard, **Tonja Hobel**, Morten T. Nielsen, Scott James Harrison, Anders Holmgaard Hansen, Birger Lindberg Møller, Susanna Seppälä, Morten H. H. Nørholm (2017) De-bugging and maximizing plant cytochrome P450 production in *E. coli* with C-terminal GFP fusions. *Applied Microbiology and Biotechnology*. 101:4103–4113 (DOI 10.1007/s00253-016-8076-5)

* These authors contributed equally to this work.

[#] A reproduction of the published paper was made with permission from the publisher.

ACKNOWLEDGMENTS

I would like to express my greatest appreciation and deepest gratitude to the following people who have been helping and supporting me for the completion of this thesis:

- First of all, to the three persons I owe the most. To my husband Cédric, who has been my unconditional backbone, shoulder and merciless mirror whenever needed, ever since we met. And to Lina and Rémi whose innocent joy, endless fantasy and curiosity have been the best to take my mind off things. Something beyond price to come home to every day!
- To the bigger part of the family with their caring and supportive interest even from afar and who were there whenever needed to keep us running.
- To my friends from near and far who took care of well-appreciated smaller or larger changes of scenery, as well as of husband and kids during the busiest desk jobs at home.
- To Morten Nørholm, who gave me the possibility to pursue my scientific career in his laboratory.
- To Susanna Seppalä, who never got tired of answering burning question even during time-lagging discussions over the big Atlantic pond.
- To Sheila Ingemann Jensen for being so reliably helpful in methodical and scientific matters. Sheila, I will miss our endless discussions in the lab in the middle of the night!
- To my co-supervisor Solvej Siedler, whom I deeply thank for our seldom but precious meetings, yet continuous constructive support, discussions on the point and guiding input!
- To Alexey Dudnik, for being such a source of knowledge, reliable reference point in time of doubts and great lab bench neighbour.
- To all lab mates, technicians and “my” students at CfB for their company and who have all one way or the other contributed to the fun in science.
- To the non-scientific part of CfB encompassing the administrative and IT team, purchase support and lab coordination in the second floor for their unconditional and prompt support in every urgent or less urgent matters, the ladies from the lab support in first floor for her positive attitude towards any extra little wish, Susanne Ploughheld Winther from the project administration for her great helpfulness throughout my time at CfB and Rebeca Thostrup from CfB PhD administration for her kind support during “the end”.
- And of course, to all those who I did not list specifically.

Thank you!

ABBREVIATIONS AND GLOSSARY

aa	amino acid(s)
AAA+ proteins	ATPases associated with diverse cellular activities; protein superfamily
araBp	promotor sequence for the *araBAD* operon; DNA level
*Av*GFP	native green fluorescent protein (GFP) from *Aequorea victoria*
bp	base pair(s)
cAMP	3',5'-cyclic adenosine monophosphate
CreiLOV	engineered fluorescence reporter (FR) protein domain of the Light-Oxygen-Voltage (LOV) protein family, derived from *Chlamydomonas reinhardtii*
C-terminal	carboxy-terminal
ClpA	caseinolytic protease A; ATP-binding unfoldase subunit of ClpAP proteasome
ClpP	caseinolytic protease P; proteolytic subunit of proteasomes ClpAP and ClpXP
degradation motif	a degradation motif on a protein usually consists of two degrons: a tether-degron, that regulates protein substrate recognition; and an initiation site-degron, where unfolding/degradation begins (details in text). Both degron moieties are essential to initiate proteasome-dependent protein degradation.
degron	a degron is degradation signal: a specific sequence, conformational fold or amino acid in a polypeptide chain that contributes unmodified (primary degron) to a degradation motif (see details in the text); when a degron is either N- or C-terminally located on a peptide chain, it would be termed either N-degron or C-degron, respectively. Definition modified from VARSHAVSKY (2011) [1].
dg	degradation sequence motif
dgFtsH	native N-terminal degradation sequence motif recognised by the FtsH proteasome [2]
dgClpAP	synthetic N-terminal degradation sequence motif recognised by the ClpAP/ClpS proteasome [3]
EcFbFP	engineered fluorescence reporter (FR) protein domain of the Light-Oxygen-Voltage (LOV) protein family, derived from *Bacillus subtillis*
EnvZ	protein product encoded by the *envZ* gene; *env* is a mnemonic for envelope; EnvZ is described as an osmosensor protein situated in the bacterial inner membrane (IM)
FR	fluorescent reporter, fluorescence reporter
frGFP	GFP folding reporter; a variant of the *A. victoria* green fluorescent protein (*Av*GFP) [4]
FtsH	filament temperature-sensitive protein H; ATP-dependent; membrane-bound proteasome in *E. coli*
ΔG	free energy of unfolding ($\Delta G = G_{final} - G_{initial}$); ΔG is the energy difference in between the unfolded and folded state of a protein
g	unit for relative centrifugal force (RCF)
GFP	green fluorescent protein from the GFP family derived from *A. victoria*
GST	glutathione S-transferase
IM	inner membrane
IMP	inner membrane protein; membrane protein (MP) of the inner membrane
IPTG	isopropyl-β-D-thiogalactoside
lacIo	operator sequence recognised and bound by the LacI repressor; DNA level

LOV protein	photoactivatable protein of the Light-Oxygen-Voltage protein family
LPS	lipopolysaccharide
µL	microgram
µL	microliter
mM	millimolar
MP	membrane protein
MQ water	Milli-Q™ water; ultrapure water
n	count of independent experiment(s)
NarX	protein product encoded by the *narX* gene; *nar* is a mnemonic for nitrate reductase or nitrate regulation; NarX is described as a nitrate sensor-transmitter protein situated in the inner membrane protein
NavZ1	inactive membrane protein chimera consisting of the N-terminal and C-terminal moieties of osmolarity sensor histidine kinases NarX and EnvZ from *E. coli*, respectively [5]
N-terminal	amino-terminal
OM	outer membrane
OMP	outer membrane protein; membrane protein of the outer membrane
ORF	open reading frame
phiLOV2.1	engineered fluorescence reporter (FR) protein domain of the Light-Oxygen-Voltage (LOV) protein family, derived from *Arabidopsis thaliana*
PL	phospholipid
POI	protein of interest
PpiD	peptidyl-prolyl cis-trans isomerase D, inner membrane protein (IMP) of *E. coli*
PQC	protein quality control
pro-degron	precursor of a degron; generally, a pro-degron is a specific sequence, conformational fold or amino acid in a polypeptide chain that can be cleaved or otherwise modified to produce a secondary or primary degron (see details in the text); if upon modification a N-terminal or C-terminal degron is released, it would be termed either pro-N-degron or pro-C-degron, respectively. Definition modified from VARSHAVSKY (2011) [1].
PROTi	protein interference
r_B	count of biological replicate(s)
r_T	count of technical replicate(s)
rpm	revolutions per minute; unit for the frequency of rotation
rsTEVP(X)	restriction site of the tobacco etch virus protease (TEVP) with amino acid X in amino acids position seven of rsTEVP; X becomes N-terminally released upon TEVP cleavage; peptide sequence
SD sequence	Shine-Dalgarno sequence; mRNA level
SecYEG translocon	protein translocation complex in the bacterial inner membrane (IM)
sfGFP	superfolder GFP; variant of the *A. victoria* green fluorescent protein (*Av*GFP) [6]
SRP	signal recognition particle
T	temperature
TEV	tobacco etch virus
TEVP	tobacco etch virus (TEV) protease
TIR	translation initiation region; mRNA level

TM	transmembrane
TMH	transmembrane helix
ts	temperature-sensitive
Ub	refers to free ubiquitin as well as to a ubiquityl moiety bound/fused to a protein
UPS	Ubiquitin (Ub)-proteasome system
U-SM	uracil-excision based site-directed mutagenesis [7]
YFP	yellow fluorescent protein; variant of the *A. victoria* green fluorescent protein (*Av*GFP)
YfgM	protein product encoded by the *yfgM* gene; inner membrane protein (IMP) of *E. coli*
YhaI	protein product encoded by the *yhaI* gene; inner membrane protein (IMP) of *E. coli*

All amino acids and their respective three- or one-letter symbol as suggested by the IUPAC-IUB COMMISSION ON BIOCHEMICAL NOMENCLATURE (1968) [8] are not part of this list.

NOMENCLATURE

The strain nomenclature as well as genetic nomenclature used in this thesis are as described in *Instructions to Authors* of the *Journal of Bacteriology* [9] and *Linkage map of E. coli K-12, Edition 10: The traditional map* by Berlyn [10] with following additional specifications inspired by *Help: Genetic nomenclature* of *EcoliWiki* [11].

Molecular Genetic Constructs

Coding sequences or open reading frames (ORFs) are separated from their respective promoters by a slash symbol (/). Example: *T5p*/*malE-gfp*, the *malE-gfp* gene fusion is under the transcriptional control of the *T5* promoter.

Proteins products of modified genes

As suggested for genes [9], superscript " $^{+}$" and " ' " post-fix' are also used for protein products to demonstrate a complete or truncated protein product, respectively. In case for a terminal truncation, the modified terminus is indicated by using superscript "N" or "C". Example: $X^{N'}$ or $X^{C'}$ exemplify an N- or C-terminally truncated protein X.

Plasmids

Derivatives of natural plasmids or plasmid vectors are indicated with a lowercase "p" pre-fix (e.g. pACYC). For genes on a plasmid and any other gene that are introduced into a strain by experiment (see below), the same nomenclature as for chromosomal genes is used (see ASM (2018) [9]), but placed within square brackets ([]). Example: pACYCΩ[*T5p*/*malE-gfp*] indicates that the *malE-gfp* coding sequence is fused to the *T5* promoter which has been inserted into plasmid pACYC.

Oligonucleotides or primers

Oligonucleotides, often shortened as oligos or also termed primers, are indicated here below with a lowercase "o" pre-fix. Example: oMSB1234 stands for oligo with number 1234 from the collection of author with the initials MSB.

THESIS OUTLINE

Errors happen constantly at any time in nature. This also refers to misfolded proteins *in vivo* du to genomic mutations, to errors during transcription or translation, or to various stress conditions [12]. Misfolded proteins represent a dosage-dependent fitness cost for the cell, sometimes leading to its death [13]. Proteins' fate and therefore also a cell's fitness is determined by the repetitive trial-and-error chaperone-assisted refolding attempts. If misfolded proteins are unable to refold or to be refolded correctly, they are targeted for processive proteolysis [12]. Processive proteolysis is not only a fate but can also be an essential regulatory mechanism for the levels of fully functional proteins. Under certain conditions, a cell favours the dynamic regulation of specific protein levels by proteolysis rather than their transcriptional control [14]. Processive proteolysis usually occurs in two ways: constitutively or conditionally [15]. In either way, it is a carefully regulated mechanism *in vivo* that affects many cellular functions [16,17].

The dynamic regulation of protein abundances is equally a central interest in synthetic biology [18]. Synthetic biology aims to create robust, orthogonal molecular systems with new, valuable functions for industrial and medical applications [19]. Those systems need to be subject to unconditional control for protein biosynthesis as well as for protein stability. Regulated processive proteolysis is especially advantageous in case of long-living proteins which would otherwise persist for many generations [20]. Practically, most native globular proteins are relatively stable *in vivo* [17,21,22] which might represent an adaptation to secure their functionality [21].

For bacterial cells, there are many tools to modulate specific protein levels by altering the rate of protein biosynthesis. However, a versatile tool box for bacteria that is able to tune the stability of specific proteins by their destabilisation (or re-stabilisation) is only just about to be developed [18]. Tools for intracellular proteolysis are of general value for a wide variety of biotechnological applications, metabolic engineering and synthetic biological circuits but also for fundamental studies of essential proteins and biological networks [18,23]. Here, the additional presence of an engineered, conditional element which can trigger or disable targeted intracellular proteolysis has been proven to be an essential part for tool regulation. The nature of this conditional element needs to be adapted on the need of each study.

As for all Gram-negative bacteria, processive proteolysis in *E. coli* is mediated by five ATP-dependent proteasomes which are all located in the cytoplasm: ClpAP, ClpXP, HslUV (*alias* ClpYQ), Lon and FtsH [24]. Only proteins that reveal the right combination of specific degradation signals, or degrons, are (1) recognized by the proteasome, (2) processively unfolded and (3) subsequently degraded [25]. To any synthetic biologist's distress, an overwhelming number of factors affect how proteins are targeted to a proteasome and how they are subsequently degraded [25]. Only few aspects of the bacterial proteasomes' multistep mechanisms encompassing protein recognition, mediation, unfolding and degradation have been characterised

mechanistically, though most of it is still unknown. Additionally, the vast variety of protein substrate classes, conformations, architectures, topologies and their individual intracellular localisation adds up to the complexity of understanding proteolysis.

Nevertheless, with the given knowledge, a few tools for conditional protein destabilisation have been developed in bacteria until now [18,20,23,26]. Each of them is able to target a particular fraction of the cells proteome, which is defined by the protein terminus (i.e. N- or C-terminus) where degradation is initiated by those tools. For some proteins, both termini may be used for degradation. However, all the available conditional tools have been shown suitable for cytoplasmic globular proteins. Their applicability on membrane proteins (MPs) was only tested in rare cases until now. In brief, there is much room for improvement in general and a particular need to expand the available tool box for conditional degradation of MPs.

The practical work conducted during this thesis aimed at expanding the range of proteins that are targetable for proteolysis in a conditional manner. To achieve this, known synthetic, heterologous and endogenous proteic building blocks were rearranged to design a synthetic platform for targeted proteolysis *de novo*. This thesis is structured into three interdependent main parts. The first past (INTRODUCTION) represents an introductory preface which is important for a general understanding of the thesis. The second part (STUDY I) details the development of a bipartite synthetic tool for the conditional control of probably predominantly cytosolic globular proteins in form of a published article. And finally, the third part (STUDY II) describes an unfinished attempt to transplant the synthetic set-up from STUDY I into the context of conditional MP degradation.

INTRODUCTION

For a general understanding of this thesis, the following sections intend to provide insights into the complexity of native processive proteolysis in eukaryotes as well as prokaryotes since there are many parallels between those two domains; however, focus will be kept on prokaryotes. Additionally, an overview is given about synthetic strategies already available to manipulate protein stability. Finally, details about the proteic and molecular building blocks relevant for this thesis are provided.

1 NATIVE PROTEOLYSIS IN BACTERIA

The genome of *E. coli* is predicted to harbour approximately 4,300 protein-coding genes but only about one-fourth of the protein products seem to be abundant at various growth conditions [27,28]. Differences in protein abundance were found to correlate with the proteins' function(s) in the cell [28]. The ability to control protein availability is essential for cell homeostasis, that itself is critical for cell viability [22]. Besides an often complex regulation of protein biosynthesis, the availability of a given protein is also controlled upon the regulation of its stability [14,22].

In bacteria, intracellular protein degradation or *processive proteolysis*, plays a role, on one hand, in the elimination of misfolded, damaged as well as incompletely synthesised proteins; on the other hand, it also regulates levels of key proteins responsible for intracellular signalling as well as gene transcription control [17]. Therefore, processive proteolysis can affect a variety of diverse functions [17]: it plays a major role in the general protein quality control (PQC) [17], can affect many stress response pathways [17,22], has an impact in the regulation of pathogenesis [29] or can trigger transitions between different physiological stages in certain bacteria [30].

General proteolysis is usually catalysed by different proteases available in the cell. In contrast to most regulatory mechanisms, processive proteolysis is inherently irreversible and may seem wasteful [31] since the restoration of a degraded protein requires its re-synthesis [22]. In order to avoid random, uncontrolled degradation, some intracellular proteases developed into ATP-dependent unfoldase-protease complexes, called *proteasomes* (discussed for *E. coli* in INTRODUCTION section 1.1), and evolved processive proteolysis into a carefully regulated process [24]. Generally, proteasomes ensure unfolding of their target proteins with subsequent cleavage of the unfolded polypeptide at multiple sites by releasing oligopeptides that can be further degraded into single amino acids for recycling [32]. Proteasome-mediated proteolysis as such is usually regulated by a two-part degradation motif on the protein substrate. It consist of two inter-dependent specific degradation signals, termed *degrons*, that are responsible for the substrate recognition (recognised by the proteasome directly or by specific adaptor proteins [33]) and the provision of an initiation site for unfolding [25].

Alternatively, other but ATP-independent proteases abundant in cytosol [34] and periplasm [35] are responsible for another form of regulated proteolysis: the limited proteolysis or *proteolytic processing* [22] (not to be confused with the processive proteolysis performed by proteasomes [14]). It involves cleavage at a limited number of specific sites in the protein resulting in its activation or maturation [22]. Alike processive proteolysis, proteolytic processing underlies highly specific mechanisms [14] (not discussed in this thesis). Originally the degron term referred to proteasome-specific degradation motifs [36] but is recently also used for degradation signals from ATP-independent proteases due to an apparent similar complexity [37–39]. Interestingly, proteolytic processing activities from ATP-independent proteases involved in periplasmic PQC (e.g. DegP) seems to be regulated not only by degradation signals abundant on their target proteins but also by the apparent availability of adaptor proteins [37–39]. It seems that proteolytic processing is a similar regulated process as processive proteolysis by proteasomes. Finally, proteolytic processing and processive proteolysis can also cooperate: a site-specific cleavage in a protein might increase its susceptibility for processive proteolysis [14,22].

1.1 Proteasomes in *E. coli*

As for all Gram-negative bacteria, processive proteolysis in *E. coli* is carried out by five ATP-dependent proteasomes which are all located in the cytoplasm: ClpAP, ClpXP, HslUV (*alias* ClpYQ), Lon and FtsH [24] (Figure 1). In general, those proteasomes are composed of two functional moieties, an unfoldase and a protease [40]. The unfoldase and protease moieties are either located on separate polypeptides (ClpAP, ClpXP and HslUV) or both moieties are folded from a single polypeptide chain (Lon and FtsH) [40].

Without exception, all unfoldases belong to the AAA^+ (ATPase associated with diverse cellular activities) superfamily and form an hexameric ring [33,40]. Unfoldase rings can interact as mono- (ClpX, HslU) or double layer (ClpA) with the protease moieties [41] and thereby activate the latter [34]. Biochemical experiments suggested that a double layer of unfoldase rings increase the unfolding capacity to process targeted proteins with high intrinsic thermodynamic stability [42]. Furthermore, unfoldase rings can also bind to both sides of the protease moiety to form a doubly capped protease complex [43].

Generally, the unfoldase unfolds target proteins and translocates them through a central axial pore in an ATP-dependent manner into the protease moiety [33]. Unfolded polypeptides entering the protease moiety are bound by substrate-binding domains of the protease and are subsequently cleaved into oligopeptides in an ATP-independent fashion [24,43]. For ClpP and HslV, two oligomeric protease rings form a compartmentalised *proteolytic chamber* [24]. General access of protein substrates to this proteolytic chamber is severely limited; this blockage is only relieved upon binding with the unfoldase [33]. Non-specific degradation in the proteolytic chamber is limited to small oligopeptides when the unfoldase moiety is not bound [43]. Interestingly for the two-peptide proteasomes, the mono-ring structure for HslV is a homohexamer whereas the ClpP ring is formed as homoheptamer [43]. Consequently, the ClpP ring interacts

with its unfoldase partners, ClpA or ClpX (both homohexamer rings), in a mismatched docking fashion [43]. The mechanistic basis behind this asymmetric mismatch for ClpAP and ClpXP activities are still debated in literature [33,43]. In contrast, HslUV works without mismatch, as do the unfoldase and proteolytic homohexamer-rings of Lon and FtsH which are fused in one polypeptide anyway [33].

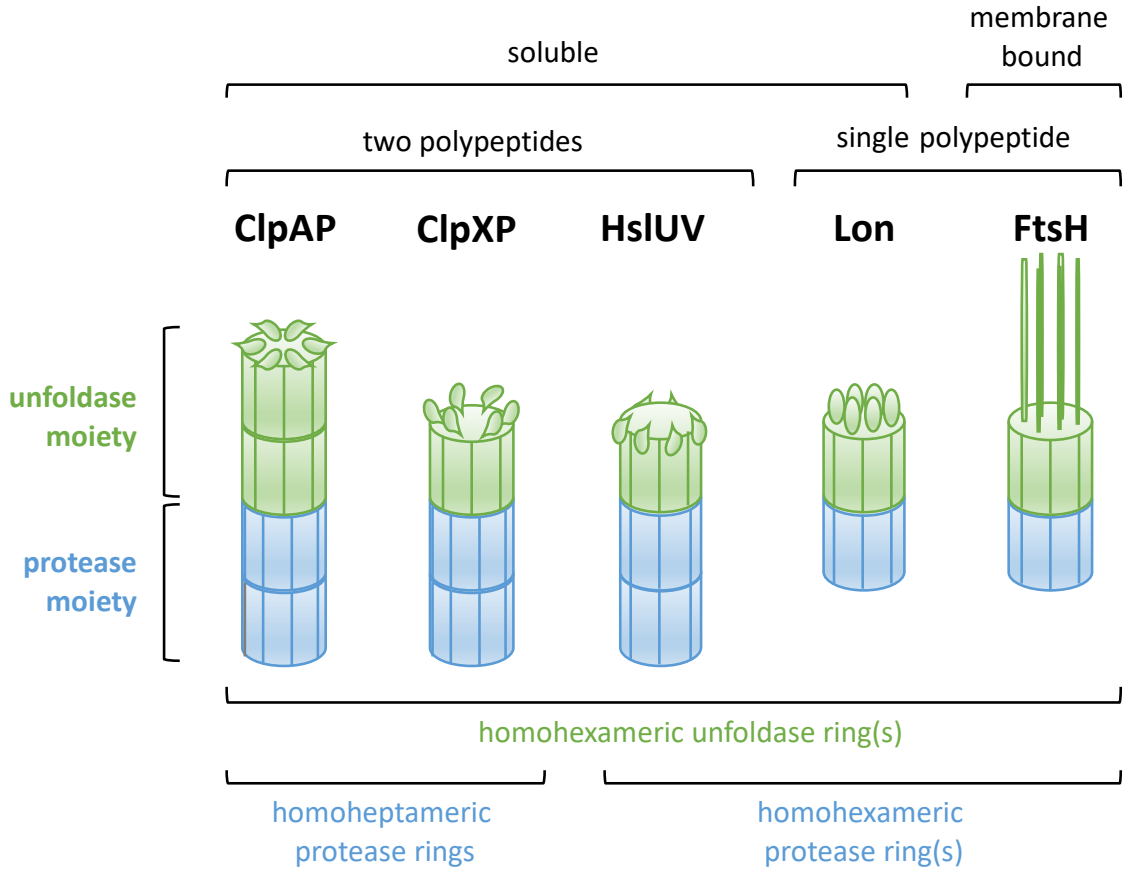


Figure 1: **Overall composition and structure of proteasomes in *E. coli***. ClpAP, ClpXP, HslUV (*alias* ClpYQ) and Lon are all soluble in the cytoplasm whereas FtsH is bound to the IM. The AAA$^+$ unfoldase moieties and auxiliary domains are depicted here in green and the protease moieties are in blue. The auxiliary domains for each AAA$^+$ unfoldase moiety are family-specific and often serve as substrate recognition sites directly, or as docking sites for adaptor proteins. In FtsH, auxiliary domains anchor the proteasome to the IM [33]. Mono- and double layer arrangements of the unfoldase rings and protease rings are indicated and their distinct oligomeric mono-ring structures are given for each complex; the graph also describes whether the unfoldase or protease rings are formed from a single or from two polypeptide chains. The possible binding of two independent unfoldase rings on either side of a given proteolytic chamber is not shown. Figure modified from KIRSTEIN *et al.* (2009) [44].

Although the mechanistic steps for protein substrate entry are well established, the pathway for product release from proteasomes seems to be widely unknown [45]. Two possibilities are discussed in literature: (1) the unfoldase rings (e.g. ClpX) function as gatekeeper not only for substrate entry but also for product release [43]; (2) peptide products are released through dynamic side pores in the proteolytic chamber at the interface between the two protease rings, as discussed for ClpP for instance [17,45]. Once released by the proteasome, oligopeptides have a very brief half-life: within seconds they are rapidly digested into amino

acids by the cytosolic endopeptidases and aminopeptidases involved in protein re-synthesis or energy metabolism [31].

1.2 Degrons – A general overview

Most protein substrates are recognised by direct interaction with the unfoldase moiety of a proteasome. Other proteins require specific substrate recruiting factors, known as *adaptor proteins*, to direct them to a proteasome for conditional processive proteolysis [40]. To date, four adaptor proteins have been identified in *E. coli*, three of which, SspB, UmuD and RssB, modulate ClpXP specificity, whereas a single adaptor protein, ClpS, alters the specificity of ClpAP [44]. There are proteins that have recognition sites for numerous proteasomes, adaptor proteins as well as ATP-independent proteases [20]. In case for processive proteolysis, the proteasomes themselves or the adaptor proteins contain specific domains for substrate recognition which bind selectively to degradation signals on their protein substrates (Figure 2) [46].

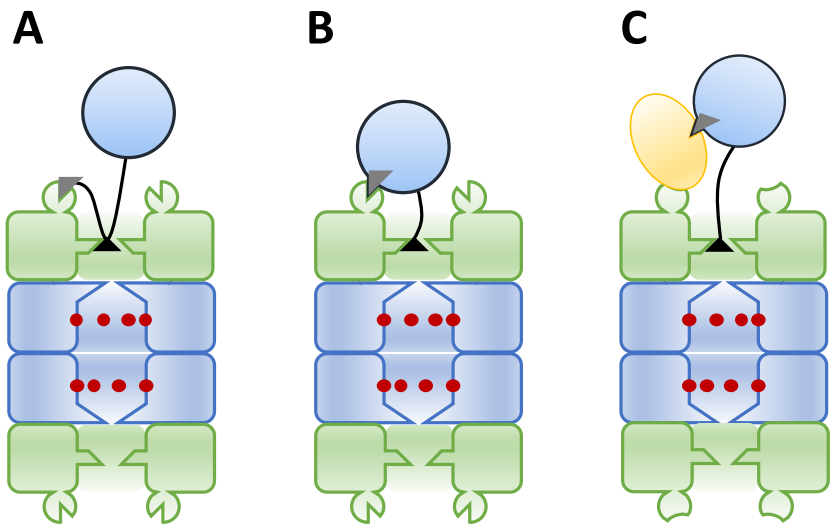


Figure 2: **Minimal models for protein substrate recognition for processive proteolysis**. Protein substrate recognition is regulated by the tether-degron (◣). An exposed tether-degron, terminally (A) or intrinsically (B, C) on a protein, is bound either directly by the auxiliary domains of the AAA$^+$ unfoldase moiety (green; A, B) or through adaptor proteins (yellow; C) or regulatory subunits (not shown). ATP-dependent unfolding of the protein substrate by the unfoldase moiety, translocation to the proteolytic chamber and subsequent degradation are initiated upon interaction of the initiation site-degron (◣) with the axial pore of the AAA$^+$ unfoldase moiety (green; A, B, C). Proteolytic active sites in the proteolytic chamber are depicted as red dots. Details in the text. Figure modified from OLIVARES *et al.* (2016) [43].

Those degradation signals have been originally coined *degrons* [47]. Since their discovery in 1986, degrons, by definition, represent protein features that confer metabolic instability [15,48]. The present knowledge postulates that an active degradation motif that finally leads to processive proteolysis is, however, an interplay of minimum two degrons. One degron is responsible for tethering the substrate protein efficiently

to the proteasome, coined the *tether-degron*, and the other one represents the initiation site where unfolding/degradation begins [25] (*initiation site-degron*). This binary structure has been shown for the prokaryotic SsrA-tag recognised by a variety of proteasomes [49] and the N-terminal degradation sequence motif processed by the prokaryotic N-end rule pathway [50] (Figure 4). The eukaryotic Ub-proteasome system (UPS) is even regulated by three inter-dependent degradation signals [51] that lead to an active UPS degradation motif: (1) a *tether-degron I* which consist of an amino acid residue (terminal or internal) that, once recognised, initiates substrate ubiquitination at (2) one or more internal Lys' residues. Each Lys residue can be characterised, functionally spoken, as *pro-degron* (see Table 1) *of a tether degron II* which is represented by the attached poly-Ub-chain [51]; and finally, (3) an *initiation site-degron* that is located on a structurally disordered segment of a target protein [51]. Thus, the actual tether-degron responsible for proteasome recognition is the Ub-chain (tether degron II) which is dependent on the availability of (1) and (2). Interestingly, it was shown that the Ub-chain and the initiation site are spatially separated and can be either located *in-cis* on one single protein or *in-trans* on different proteins in a protein complex [52] (Figure 3c). In either case, the initiation site-degron is always acting *in-cis* on a protein, thus, only the initiation site-bearing proteins get degraded (*cis*-acting degron) [1]. The tether-degron can be described as *cis*- or *trans*-acting depending on its localisation [1,25,53–55].

Since it was proposed the first time by VARSHAVSKY (1991) [47], the degron term has been used very inconsistently in the literature [51] due to the discovery of apparent increasing structural complexity. A term re-definition for standardised use seems imperative. In this thesis, the term degron will refer to any degradation signal that contributes to an active degradation motif which is able to initiate processive proteolysis. Thereby, a single degron is assumed to be essential but not sufficient to embody an active degradation motif. Instead, an interplay of minimum two degrons is needed to regulated proteasome recognition as well as unfolding initiation. Those will be termed tether- and initiation site-degron, respectively. Furthermore, the degron term will be only used in context with processive proteolysis (not for proteolytic processing). In case there is a higher level of regulation prior to proteasome recognition and unfolding, as in the case for an active UPS degradation motif, the respective degrons or its precursors (pro-degrons) will be characterised accordingly, including their hierarchical order, if known.

Speaking of hierarchy, degrons have been functionally organised into hierarchical levels: primary, secondary and tertiary degrons (Table 1) [17,36,56]. A primary degron can contribute directly to the completion of an active degradation motif that initiates proteolysis (constitutive degron [15]). Alternatively, it can be generated from secondary or tertiary degrons (conditional degrons [15]). Following the logic of VARSHAVSKY's definition of a pro-degron [1], secondary or tertiary degrons are precursors of a primary degron and therefore pro-degrons. Structurally, a pro-degron is a specific sequence, conformational fold or amino acid residue in a polypeptide chain or protein that can be cleaved or otherwise modified to produce a primary degron or primary pro-degron (equivalent to a secondary degron). Accordingly, secondary and tertiary degrons (or primary and secondary pro-degrons, respectively) need one or two modification steps, respectively, to be

turned into a primary degron. In case a (pro-)degron is located directly at the N- or C-terminus, it would be termed (pro-)N-degron or (pro-)C-degron, respectively. However, the recognition and modification of known tertiary degrons, which are exclusively tertiary N-degrons (secondary pro-N-degrons), and their conversion into secondary degrons (primary pro-N-degrons) have been only described in eukaryotes until now [1,17].

Table 1: Hierarchy of degrons and pro-degrons.

DEGRON		PRO-DEGRON	FUNCTIONALITY	AFFILIATION
Primary degron			constitutive	Eukaryotes, prokaryotes
Secondary degron	=	Primary pro-degron	conditional	Eukaryotes, prokaryotes
Tertiary degron	=	Secondary pro-degron	conditional	Eukaryotes

1.2.1 Nature and location of (pro-)degrons

The nature of (pro-)degrons can be manifold:

- intrinsically unstructured sequence motifs (e.g. the eukaryotic PEST motif [57,58]; the initiation site-degron of an UPS degradation motif [51]),
- conformational folds (e.g. the eukaryotic IκBα-degron [59]) and
- terminal/intrinsic exposed amino acids (e.g. the primary tether-degron or Lys residues of an active UPS degradation motif [51]; the N-terminal tether-degron recognised by the prokaryotic N-end rule pathway [50]) [52,60].

Some proteins even contain (1) multiple (pro-)degrons which contribute to the proteolysis by the same degradation pathway(s) (e.g. multiple eukaryotic PEST motifs [57,58]; multiple Lys residues as part of an active UPS degradation motif [51]), (2) overlapping degrons attracting different degradation pathways upon exposure (e.g. the prokaryotic SsrA-tag [49]) or (3) inactive cis-/trans-acting tether-(pro-)degrons that are missing their partnering initiation site-degrons on the same protein in order to form an active degradation motif but which turn active upon complex formation with an initiation site-bearing protein (see Figure 3c; e.g. intrinsic Lys residues, as primary pro-degrons of the Ub-chain tether-degrons, which can contribute *in-cis* or *in-trans* to an active eukaryotic UPS degradation motif [1,25,53–55]).

(Pro-)degrons seem to be located anywhere along the peptide chain, such as at internal sites [58] yet often nearby the N- or C-termini [61]. In folded proteins, (pro-)degrons can be either exposed or buried within the protein structure itself (cryptic degrons); they can also be buried upon complex formation with other protein(s) which is described as cooperative stability (Figure 3b) [14,22,62]. In contrast, buried degrons within a fully folded protein can only become accessible to proteolytic recognition systems upon conformational change (Figure 3e and f), (partial) unfolding of protein substrates (Figure 3g) or specific proteolytic cleavage

(Figure 3i) [22]. Thus, besides the appearance of degrons or pro-degrons and their function as well as given proteasome-specificity, the turnover of a protein is also regulated by its correct folding and/or complex formation with other proteins. Exposed degrons on folded proteins might result in the constitutive and rapid degradation of the latter (Figure 3a). However, rapid degradation combined with stringently regulated protein production can result in very defined protein levels at specific time points [14].

1.2.2 Conditional protein modifications for generating (pro-)degrons or active degradation motifs *in vivo*

There is a whole variety of intracellular or extracellular events which induces the formation or the exposure of (pro-)degrons. On one hand, their release might be the result of conformational changes, unfolding or misfolding of the respective protein. On the other hand, they can be generated by modifications of exposed pro-degrons, proteolytic processing events or direct covalent attachment of a degron. In special cases, even a whole active degradation motif can be attached. An exhaustive overview of (pro-)degron-generating events as well as attachments of active degradations signals in prokaryotes and eukaryotes is provided in Figure 3. A detailed insight on proteolytic processing (Figure 3i) and on intracellular protein modification events including amino acid modifications, direct covalent attachment of a primary degron and attachment of a whole active degradation motif (Figure 3l) is also given.

(Pro-)degron formation by amino acid modification

As reported in the literature there is a whole variety of amino acid modifications that generate degrons with different hierarchies and significance. An overview – differentiating eukaryotic and prokaryotic modifications – is given here below.

Active degradation motifs for the eukaryotic UPS can come in multiple flavours which, in turn, trigger one of the individual degradation pathways within UPS. An active UPS degradation motif is tripartite [51]. It consist of (1) the *tether-degron I* (specific terminal or intrinsic amino acids) that initiates ubiquitination at (2) one or more internal Lys residues which are pro-degrons of the *tether-degron II's*, the poly-Ub-chains, and (3) the *initiation site-degron* that is located on a structurally disordered segment of the protein substrate [51]. Thus, the (pro-)degron nature for (1) and (2) can be summarised as single amino acids. Their appearance can be modulated *in vivo* to turn them into functional primary degrons.

A *tether-degron I* can initially be a secondary pro-degron (tertiary degron). Locally it seems to be restricted to the N-terminus. Two modification steps are needed in order to convert it into a primary degron. The first step is its transfer into a primary pro-degron (secondary degron) which might be derived by

- post-translational enzymatic deamidation of an N-terminal Asn or Gln or
- post-translational chemical oxidation of an N-terminal Cys [1,63].

A *tether-degron I*, that is a primary pro-degron (secondary degron) needs a single modification step before it can initiate ubiquitination. This can be accomplished by

- post-translational phosphorylation of specific intrinsic Ser or Thr,
- post-translational hydroxylation of a specific intrinsic Pro,
- post-translational mono-methylation of a specific intrinsic Lys and
- co-translational acetylation of specific N-terminal amino acids (Ac/N-end rule pathway) [52,64–66]. Another strategy is the
- post-translational aminoacylation of N-terminal Asp (*alias* deaminated Asn), Glu (*alias* deaminated Gln) or oxidised Cys. The aminoacyl-tRNA-protein transferase Ate1 recognises those N-terminal amino acids and conjugate them directly with Arg, a primary degron [52,63].

The internal Lys residue(s) are functionally spoken primary pro-degrons. Upon post-translational poly-ubiquitination [15], they become ubiquitinated. During ubiquitination, the small regulatory Ub protein (76 aa) is covalently conjugated to the exposed Lys residue(s) which finally results in a multi Ub-chain with $\geq$2 Ub moieties [15]. Finally, the poly-Ub-chain act as *tether-degron II* of the active UPS degradation motif responsible for its direct proteasomal recognition.

In prokaryotes, the only mechanism described to create a primary degron based on an amino acid modification is the post-translational aminoacylation of N-terminal Lys, Arg, Asp, Glu and Met[a], similar to eukaryotes [52,63,67]. Depending on the species, one or two of the aminoacyl-tRNA-protein transferases Aat, Bpt and ATEL1[b] recognise those N-terminal amino acids and conjugate them directly with the primary degrons Lys or Phe [52,63]. The N-terminal Lys or Phe functions directly as tether-degrons for the prokaryotic Leu/N-end rule pathway [1]. In an individual case, Met was recently identified as novel N-terminal residue that is prone to aminoacylation [67]. As a side note, N-terminal Met in bacteria is usually formylated (fMet) [68,69]. However, the study does not discuss if there is a deformylating reaction prior to the aminoacylation of Met.

[a] Only Lys, Arg and Met in *E. coli*.

[b] Aat is the only aminoacyl-tRNA-protein transferase in *E. coli*.

Proteolytic processing of pro-degrons

The cooperation between ATP-independent proteases and ATP-dependent proteasomes has been long speculated [14,17,22,48,70]. In essence, a protease with endoproteolytic or exoproteolytic activity was hypothesised to reveal a terminal (pro-)degron upon cleavage and thereby contributing to the generation of substrate recognitions sites for processive proteolysis. Demonstrating the genuineness of this phenomenon took however a long time [71]. Until today, only individual cases in eukaryotes and prokaryotes are described in the literature.

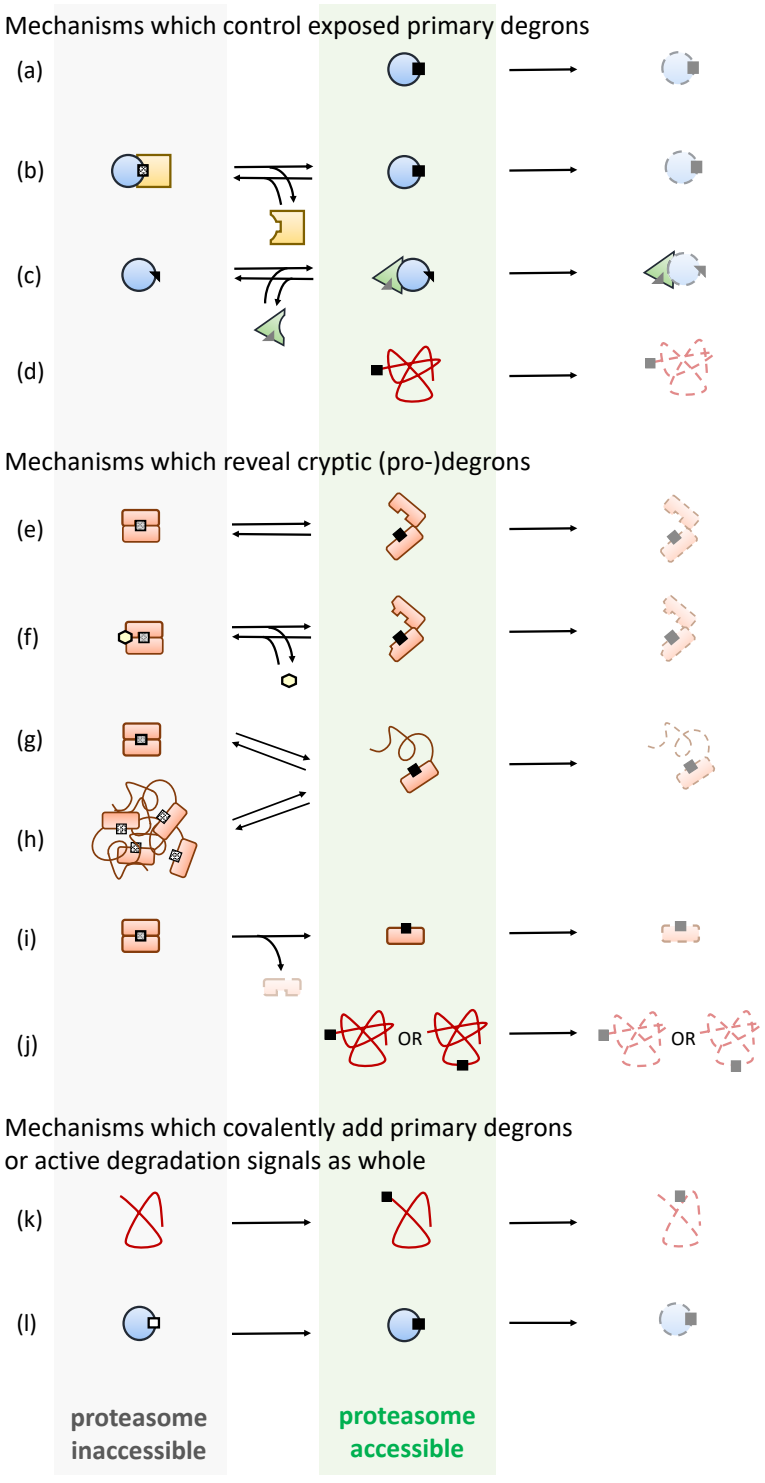


Figure 3: **Regulatory principles of processive proteolysis as part of the general protein quality control.** Proteolysis can either be (a) constitutive or (b-l) conditionally regulated by a whole variety of stimuli or mechanisms. Major scenarios are presented although variations and combinations are possible *in vivo*. (a) Efficient constitutive proteolysis due to an exposed primary degron is usually combined with conditional protein biosynthesis resulting in very defined windows of protein levels [14]. (b) A primary degron buried upon complex formation with other protein(s) (cryptic degron) can be released upon spatial resolution of the protein complex. (c) *In-trans* proteolysis occasionally occurs when the tether- and the initiation site-degron are physically separated on two different proteins. Upon complex formation only the initiation site-bearing protein gets degraded [17]. (d) fMet can act directly as tether-degron of the active prokaryotic formyl-N-degradation motif prior to folding, if the co-translational de-formylation of the N-terminal fMet fails [68,69]. Changes in protein conformation are releasing cryptic degrons due to (e) intracellular stimuli or (f) the loss of a stabilising factor (e.g. AAI[c] molecule stabilises the conformation of TraR in *Agrobacterium tumefaciens* [72]). Alternatively, (g) environmental stresses cause partial protein unfolding and hence the release of a cryptic degron. (Partially) unfolded proteins get either proteolysed directly, undergo chaperone-mediated refolding or (h) reversely form insoluble aggregates. Re-solubilisation of the aggregated proteins occur with the help of chaperones [24]. In contrast, (i) a conditional proteolytic cleavage event of a pro-degron by ATP-independent proteases can make the protein susceptible for proteolysis by releasing a degron. Furthermore, (j) protein misfolding due to genetic mutations, errors during transcription or translation, and stress can cause the disclosure of primary degrons that would otherwise be cryptic. Events that covalently add primary degrons or even active degradation motifs directly to a protein encompass: (k) the co-translationally attachment of the SsrA-tag, an active degradation motifs, to the C-terminus of incompletely synthesised proteins [33]; (l) post-translational amino acid modifications which include, firstly, the attachment/removal of functional groups to/from primary or secondary pro-degrons as well as the oxidation of already existing functional groups [52,63,65,66]; secondly, the terminal addition of a primary degron directly to a primary pro-degron (e.g. the amino acids Phe or Lys, as tether-degron of the prokaryotic N-degron, attached to the prokaryotic secondary degron residues Lys, Arg [33] or Met [67]); and finally, post-translational attachments of small regulatory proteins (e.g. the poly-Ub-chain as tether-degron of the eukaryotic N-degron [15,48,60,73]; the Pup protein as primary degron in mycobacteria [74,75]; poly-SAMP-chain as putative primary degron in archaea [76–78]). Legend: Primary degron, ■; cryptic degron, ■; pro-degron, □; tether-degron, ◣; initiation site-degron, ◣. See text for details. Figure inspired by JENAL AND HENGGE-ARONIS (2003) [14], GOTTESMAN AND MAURIZI (1992) [22], SAUER AND BAKER (2011) [33].

Direct N-terminal (pro-)degron exposure due to endoproteolytic cleavage events

In yeast, the SCC1 protein was shown to be cleaved at an internal site resulting in the release of the C-terminal fragment. This fragment bears a primary tether-degron on its released N-terminus which can be recognised and degraded by the UPS [71].

The discussion whether endoproteolytic cleavage events exist also in prokaryotes might be supported by recent findings of SEKAR *et al.* (2016) [18]. By using a synthetic 113 aa N-terminal degradation tag (Ntag [18]), the authors demonstrated the release of an as-yet unspecified N-terminal (pro-)degron due to a spontaneous cleavage event within the tag-sequence. The specificity of this spontaneous endoproteolytic activity was equally unknown [18]. Pending the clarification regarding the involved protease(s) and their physiological relevance, the aforementioned concept might have found its first evidence in bacteria.

[c] TraI-synthesized autoinducer N-3-oxooctanoyl-l-homoserine lactone is an effector molecule in *A. tumefaciens*.

Direct or indirect N-terminal (pro-)degron exposure due to exoproteolytic cleavage events

The conditional exoproteolytic removal of the N-terminal Met (^{1}Met) to reveal an N-terminal tether-degron has been recently suggested for playing both a direct as well as an indirect role for the fate of a protein [17]. Generally, the first amino acid which initiates protein translation is always a Met [79,80] (more specifically, an ordinary Met in eukaryotes [69] and a formyl-Met (fMet) in prokaryotes as well as eukaryotic mitochondria and chloroplasts [68,69]). Yet, despite being the first amino acid during translation, ^{1}Met predominance is not implicit at all N-termini of functional proteins. In *E. coli*, co-translational removal of ^{1}Met by Met-aminopeptidases was demonstrated in cases where Ala, Gly, Ser, Cys, or Pro were encoded as second amino acid - occasionally also the bulky Val or Thr [68,81]. However, the resulting N-terminally exposed amino acids are widely inconsistent with the known destabilising N-terminal residues of the N-end rule which are predominantly bulky and hydrophobic [52]. Based on aforementioned general observations and individual case studies, three hypothesis are discussed in literature of ^{1}Met-cleavage playing a role in protein destabilisation [17].

The first hypothesis is based on the exceptional yet proven cases where ^{1}Met was cleaved despite the presence of a bulky amino acid in second position in *E. coli* [68,81]. Consequently, the removal of Met followed by an bulky N-end-degron residue is considered as highly speculative but not impossible [17]. In this context, the possible abundance of endopeptidases other than Met-aminopeptidases has been speculated [68].

The second hypothesis describes an indirect role of ^{1}Met-removal in processive proteolysis since ^{1}Met has been identified as secondary destabilising residue [67]. In the unique case – so far – of the prokaryotic cytoplasmic PATase, Met-aminopeptidases were shown unable to remove ^{1}Met. Instead, an aminoacyl-tRNA-protein transferase was discovered to attach directly a primary degron residue to ^{1}Met *in vitro* and *in vivo* [67] (see prokaryotic amino acylation of amino acids here above). It remains to be determined whether this is a general mechanism to regulate proteolysis [24].

The third, more recently formulated hypothesis discusses fMet (*alias* ^{1}Met) as primary degron in bacteria, similarly to N-terminal acetylated amino acids of the eukaryotic Ac/N-end rule [69] (see eukaryotic acetylation of amino acids here above). Generally, the removal of fMet encompasses two steps: (1) co-translationally to the growth of the nascent peptide chain during translation, the formyl-group of fMet is removed by a ribosome-associated deformylase; (2) the resulting deformylated Met can be then conditionally removed by the Met-aminopeptidase [17,69]. Since de-formylation is expected to fail frequently *in vivo* in bacteria, it has been suggested that (non-deformylated) fMet in bacteria can act as N-terminal degron, termed fMet/N-degron, thereby contributing to co-translational quality control of nascent proteins [68,69]. Furthermore, growing polypeptide chains with this (non-deformylated) fMet occur to be largely co-translationally destructed by a novel branch of the bacterial N-end rule pathway, termed the fMet/N-end rule pathway [69]. FtsH is suggested to act as putative processive proteolytic protease in this pathway [68,69].

Attachment of an active degradation motif

According to the literature, there are three special cases where an active degradation motif gets directly attached to a protein or a nascent polypeptide chain (Figure 3k and l). All those cases have been discovered in prokaryotes. They encompass a co-translational attachment of a unique oligopeptide on one hand (SsrA-tag), and on the other hand, two post-translational attachments of small regulatory proteins which have been described as ubiquitination-analogues (pupylation and SAMPylation). All attachments seem to harbour both, the tether-degron as well as the initiation site-degron. These three cases will be described here in more detail.

The SsrA-tag represents a special case for an active degradation motif attachment [61] due to its unique mechanism and key role in intracellular PQC in bacteria [82]. The 11-residue peptide (-AANDENYALAA) is co-translationally attached to the C-termini of nascent polypeptide chains whose biosynthesis has stalled or has been interrupted [82]. The last ten residues of the tag are encoded by the *ssrA* gene that is highly conserved throughout bacteria [82,83]. Its transcript is a stable 362-nucleotide tmRNA [83] with both tRNA-like (charged with Ala) and mRNA-like properties to modify incomplete polypeptides and to mark them for degradation [82]. SsrA-tagged cytoplasmic polypeptides are recognised and degraded *in vivo* by C-terminal-specific proteases which are active in the cytoplasm (ClpAP, ClpXP, Lon [84,85] and FtsH [86]; discussed in INTRODUCTION section 1.1) or the periplasm (Tsp [82,83], DegP and DegQ [38]). Interestingly, recombinant C-terminal SsrA-tagging of specific proteins with subsequent proteolysis has been proven to be a powerful tool [82,84,87,88]. SsrA-mediated proteolysis has been shown essentially for soluble proteins [49,82,84,89,90], but recently also for MPs with an N_x-C_{in} topology. In detail, a solubilised multidomain MP (featuring 12 TMHs) was shown to be degraded *in vitro* by ClpXP [87] and another MP (featuring 3 TMH) was degraded by both FtsH and ClpXP *in vitro* but only by FtsH *in vivo* [88].

The pupylation in mycobacteria has been recently described as a mechanism for the attachment of an active degradation motif in form of a protein [74,75]. Thereby, a single moiety of the modifier protein Pub is C-terminally conjugated to an intrinsic Lys residue of a protein substrate prior to degradation [74].

SAMPylation is a similar process found in archaea whereby the Ub-like proteins SAMP1/2 form polymeric chains on substrate proteins [76–78]. The poly-SAMP-chain is suggested acting as active degradation motif [76–78] even though the whole degradation mechanism needs fundamental clarification.

1.3 Native ClpAP-dependent proteolysis in *E. coli*

In the ClpAP proteasome (Figure 1), the protease moiety ClpP partners with the AAA^+ unfoldase moiety ClpA [43]; ClpA is responsible for unfolding and translocating protein substrates through an axial pore into ClpP [45]. The protease moiety ClpP is a barrel-shaped cylinder composed of two homoheptameric symmetric rings stacked back-to-back forming the catalytic chamber [91]. Protein substrates usually only access the proteolytic chamber upon binding with ClpA, otherwise the two axial pores on both ends of ClpP are

blocked [33]. Among the AAA$^+$ unfoldase moieties in *E. coli*, ClpA is unique as it consist of two homohexameric rings, D1 and D2, which form two distinct stacked layers on top of ClpP [33,41,43]. Obviously, the hexamer-heptamer interaction in ClpAP complexes creates a symmetry mismatch [33]. The mechanistic basis for this interaction still remains to be determined [33,43].

Nevertheless, ClpA is a powerful unfoldase that can efficiently and rapidly process protein substrates of high thermodynamic stability [42]. A tighter grip on the substrate, due to a larger interaction surface, is hypothesised being responsible for the superior and faster unfolding ability in comparison to the unfoldase moiety ClpX that has only a single ring architecture [41]. In comparison to other bacteria, the wide range of proteasome complexes found in E. coli suggests distinct biological roles for each of them [41]. While ClpXP is considered being the primary recipient of SsrA-tagged proteins with unstable folding [49,92], ClpAP is associated with the efficient degradation of intact key proteins [41,93].

1.3.1 Regulation of ClpA specificity

While ClpA can recognise substrate proteins itself (e.g. SsrA-tagged proteins [49]), it also interacts with the ClpA-specific adaptor protein ClpS that can alter the proteasome's specificity depending on its availability [44]: upon binding to ClpA, ClpS prevents degradation of SsrA-tagged protein substrates [93] and simultaneously inhibits ClpA susceptibility towards auto-degradation [92]. At the same time, ClpS delivers N-end rule substrates to ClpAP for proteolysis (INTRODUCTION section 1.3.2) [3,93]. However, ClpS is not only crucial for tethering N-end rule substrates to ClpA [94], it also seems to mediate ClpA-substrate binding (Figure 2C). ClpS is thus essential for substrate unfolding/degradation and thereby ClpAP function in general [44].

1.3.2 N-end rule substrates recognised by ClpS

All bacterial N-end rule substrates are characterised by a short N-terminal sequence whereby the most decisive element is represented in the outermost N-terminal residue. Generally, the identity of this N-terminal amino acid dictates the stability of a bacterial N-end rule substrate protein [36].

However, there are three important components which are essential for efficient ClpS recognition as well as ClpA binding (Figure 4). Firstly, the aforementioned first N-terminal amino acid as well as the nature of its α-amino group are crucial for recognition by ClpS (tether-degron; Figure 4) [50]. Typically, the bulky hydrophobic amino acids Phe, Leu, Trp and Tyr have been identified as primary N-degrons [33,52,56]. Secondly, a hydrophobic element, located downstream of the N-degron, is responsible for ClpA binding (initiation site-degron, Figure 4). Finally, an unstructured linker region of at least four amino acids (aa) between the tether- and the initiation site-degron facilitates ClpA binding [17]. The functional steps are roughly described as follows: after tether-degron-dependent substrate recognition and binding by ClpS, ClpS binds subsequently at the auxiliary domains of ClpA. A second, yet unknown, ClpS-ClpA binding step is suggested

to be necessary for the interaction between the initiation site degron and ClpA, which finally results in the translocation of the substrate [17]. The linker region seems to be necessary to bridge the distance between the ClpS-ClpA binding site and the substrate-ClpA interaction [3,17,50]. Additionally, residues adjacent to the N-terminal tether-degron influence the efficiency of ClpS/ClpAP interactions [50].

Although, the recognition of all N-end rule protein substrates by ClpS has been determined to be essential for proteolysis *via* the bacterial N-end rule pathway [3,94] only two proteins have been identified as true N-end rule substrates to date (PATase and $Dps^{N'}$, discussed in DOUGAN *et al.* (2010) [17]). Nevertheless, a small list of other proteins have been recently identified as ClpS-interacting proteins in two independent studies [67,95] but without having associated any as substrates for ClpAP. If so, their individual degradation mechanisms remain to be addressed [17].

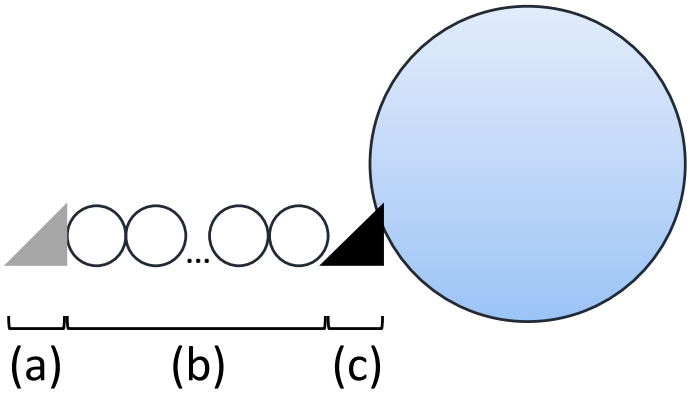


Figure 4: **Components of a prokaryotic N-terminal active degradation motif**. (a) ClpS recognition depends on the identity of the first N-terminal residue (◣, tether-degron) of a protein substrate (blue) and the nature of its α-amino group (not shown). Primary destabilising residues are Phe, Leu, Trp and Tyr. (c) A hydrophobic element is responsible for ClpA interaction (◣, initiation site-degron). (b) An unstructured linker region of minimal 4 aa is crucial to sterically bridge (a) and (c). Furthermore, neighbouring amino acids (○) of (a) modulate ClpS recognition with decreasing strength the further they are distant to (a). Details in the main text. Figure modified from WANG *et al.* (2008) [50] and DOUGAN *et al.* (2010) [17].

Considering the metabolic functions of the two confirmed N-end rule substrates and the ones of the remaining ClpS-interacting proteins identified to date [67,95], the N-end rule pathway does not seem to be linked to a specific metabolic pathway or stress response system [17]. Instead, it is speculated to modulate the levels for specific proteins with distinct metabolic functions [17]. In essence, there is not much known about the physiological function of the N-end rule pathway in bacteria [56]. However, the mechanistic understanding of the functional steps of substrate recognition, ClpS-mediated delivery and degradation has increased dramatically lately (for details, the reader is referred to recent reviews and research articles [3,17,33,41,43,45,94]).

1.4 Native FtsH-dependent proteolysis in *E. coli*

In contrast to other proteasomes in *E. coli*, at the exception of Lon, the unfoldase and protease moieties in FtsH reside on a single polypeptide. Once folded, the FtsH monomers complex into a barrel-shaped homohexamer (Figure 1) [96,97]. Like all proteasomes, a protein substrate is unfolded in the unfoldase moiety in an ATP-dependent fashion and reaches the protease moiety for degradation through the axial pore of the barrel. Substrate access to the IM-adjacent unfoldase moiety might be sterically facilitated by associated, membrane-bound regulatory factors or complexes such as HflKC [98].

There are additional features that make FtsH unique among the other proteasomes in *E. coli* (recently reviewed [99–102]). Firstly, the N-terminal auxiliary domains of the unfoldase moiety are anchoring FtsH in the IM while the two large active moieties are facing the cytoplasm [103] (2 TMHs, N_{in}-C_{in} topology). This makes FtsH the only membrane-bound proteasome. Secondly, FtsH requires catalytic Zn^{2+} ions for functionality of its protease moiety which makes it the only metalloproteasome. Finally, FtsH affects the outer membrane structure which is composed of lipopolysaccharides (LPS) and phospholipids (PL) by regulating the abundance of key enzymes in lipopolysaccharide biosynthesis [104,105]. The balance of LPS' and PL's decides whether a cell should live or die. Thus, FtsH's impact therein makes it the only essential proteasome.

In short, FtsH is suspected to be involved in PQC of assembled and unassembled IMPs [2,88,99,101,106–110] as well as cytosolic proteins [86], in the regulation of lipopolysaccharides biosynthesis [104,105], in the control of lysogeny/lysis cycle of an abundant λ prophage [111] and in the regulation of heat shock response [112].

1.4.1 Substrates recognised by FtsH

At present, twenty-one FtsH substrates have been discovered ranging from cytoplasmic proteins to integral inner membrane proteins (IMPs) [100]. Those substrates – as determined yet – seem to exhibit a wide spectrum of active degradation motifs: they encompass unstructured sequence motifs as well as conformational folds that are located randomly either internally or at the N-/C-termini [2,100]. Until now, a common degradation motif or degradation pathway has not been uncovered since most substrate-dependent mechanisms for recognition and degradation are still unknown [100]. Instead, it seems that most FtsH substrates have unique recognition and degradation mechanisms [2].

Nevertheless, there are few unique transferable terminal degradation motifs (tags) that can be recognised and degraded by FtsH *in vivo*:

- the native, non-polar, C-terminal SsrA-tag effective for soluble [86,103] and IMPs [88,103] (INTRODUCTION section 1.2.2),
- the synthetic, FtsH-specific, non-polar, C-terminal pentapeptide cI108 (-SLLWS [113]) effective for soluble proteins [86,103] as well as one IMP [103] and
- the native, FtsH-specific, N-terminal degradation motif from YfgM (MEIYENENDQVEAVKRFF-; INTRODUCTION section 3.2.2) with a putative specificity for IMPs [2].

Interestingly, for protein substrates tagged with SsrA or cI108, FtsH was shown to have a weak unfoldase activity [103]. This suggests that it would only degrade intact proteins with low intrinsic thermodynamic stability and/or unfolded proteins - independent of their protein class [103]. Generally, FtsH seems to have a high substrate selectivity itself [109] and/or substrate specificity is highly regulated by adaptor proteins [2,101]. Nevertheless, regarding the few deciphered degradation motifs and the function of known protein substrates mentioned above, FtsH seems to degrade aberrant proteins as well as intact key regulatory proteins [100,103].

1.4.2 FtsH-dependent quality control of inner membrane proteins and putative mechanisms of their degradation

As mentioned here above, FtsH is the first enzyme that has been associated with PQC of IMPs. On one hand, it can degrade unassembled IMP moieties of functional IMP complexes, e.g. F_0a subunit of the H^+-ATPase [108] (5 TMHs, N_{in}-C_{out} topology [114]), SecY of the SecYEG translocon [109] (10 TMHs, N_{in}-C_{in} topology [115]) and PspC of the phage-shock-protein (Psp) response system [110] (1 TMH, N_{in}-C_{out} topology [116]). On the other hand, one IMP substrates has been shown to be assembled in a protein complex but physically blocked e.g. the jammed SecY of SecYEG translocon [107]. Furthermore, FtsH does also regulate IMPs associated with itself that carry a FtsH-modulating function e.g. YccA [107] (7 TMHs, N_{in}-C_{out} topology).
FtsH-mediated degradation is described as processive with subsequent release of short oligopeptides into the cytoplasm [99]. However, the specific mechanisms of FtsH-dependent MP substrate recognition, degradation and its regulation are still unknown. Nevertheless, there seem to be two plausible options for the degradation mechanism. On one hand, FtsH is thought to degrade MPs following the pulling model [117] (Figure 5A). FtsH thereby recognises preferably cytoplasmic N- or C-termini with a minimal length of 20 aa [118] and extracts the MP substrates from the IM by dislocating their TMHs and periplasmic domains towards the proteolytic moiety at the cytoplasmic side of the IM [99]. Extended cytosolic loops have also been shown to serve as direct initiation sites for degradation [2,100,118]. Considering the poor unfoldase activity of FtsH [103], FtsH is also thought to participate in proteolytic processing. A domain of high intrinsic thermodynamic stability would be able to abort FtsH-dependent processive proteolysis [103]. As result, undigested yet stable hydrophobic MP fragments would accumulate in the IM [99,118,119]. Alternatively, the release of remaining hydrophilic stable domains in the cytoplasm is equally conceivable. On the other hand, a cooperative interaction of FtsH with ATP-independent IM-bound proteases (e.g. YheL and HtpX with their active site located on the cytoplasmic side of the IM) was proposed [99,120] (cooperative shedding-pulling model, Figure 5B). Protease(s) might cleave cytoplasmic loops of MP substrates and generate thereby new cytoplasmic N- or C-termini that could be recognised and finally dislocated by FtsH [98].

Another conceivable way for MP degradation, however, mainly discussed in eukaryotes but to my knowledge not considered yet for FtsH in bacteria, is an additional cooperative interaction with

intramembrane proteases [121,122]. Intramembrane proteases of the IM [123], could cut cleavage sites within TMHs resulting in the destabilisation of the cleaved TMH fragments and their subsequent secretion into either of the hydrophilic compartment adjacent to the IM [121,122].

Additionally, an cooperative outsourcing mechanism for the dislocation of MPs is also discussed in eukaryotes. MP substrates are hereby recognised and dislocated over the membrane by a membrane-bound protein complex that directs the dislocated MP to a cytoplasmic adjacent proteasome [122]. In reference to the poor unfoldase activity of FtsH [103] the likelihood of similar interaction partners in bacteria is reasonable. Considering the above, one can then suspect that FtsH-mediated MP degradation at the bacterial IM must be far more complex than the existing models proposed in literature in order to maintain the integrity of the diverse IM proteome and thereby securing cell viability under any condition.

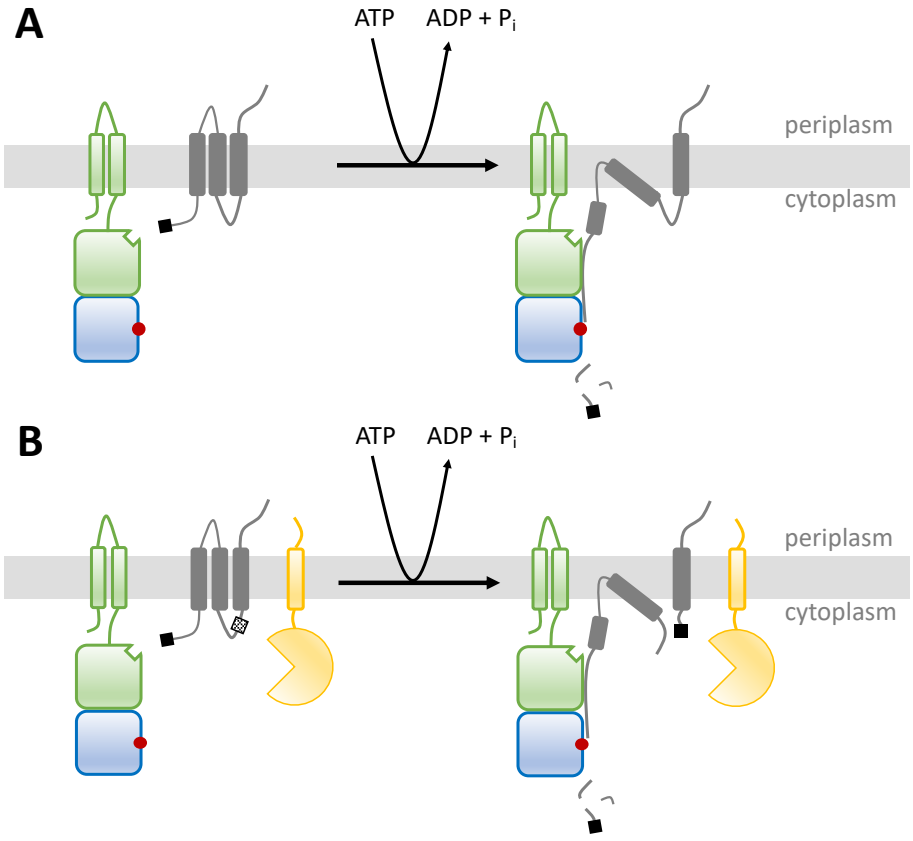


Figure 5: **Proposed models for FtsH-dependent proteolysis of IMPs**. (A) Pulling model. FtsH itself (green-blue; depicted as monomer for simplification) recognises, dislocates and degrades a MP substrate (brown) from a cytoplasmic N- and/or C-terminus. (B) Cooperative shedding-pulling model. A membrane-bound proteases (yellow) [or FtsH itself] cleaves cytoplasmic loops of MP substrates and reveals a FtsH-recognisable cryptic degron. Primary degron, ■; cryptic degron, ■; Proteolytic active site of FtsH, ●. Details in the main text. Figure modified from LANGER (2000) [117].

2 SYNTHETIC STRATEGIES TO MANIPULATE PROTEIN STABILITY

Manipulating intracellular protein levels by engineering protein stability has many applications ranging from studying individual protein functions or whole biological networks, controlling protein levels for metabolic engineering or those in synthetic biological circuits, improving recombinant protein production for purification and enhancing biological therapy for medical use. The purpose dictates the desired intracellular turnover rates of the proteins of interest (POIs) – from high to low or *vice versa* – which itself needs to be engineered but also engineerable. The latter relies on several factors including the tool(s) applied, the type of biological host and its characteristics, the growth conditions and, last but not least, the POI itself (protein class, folding characteristics, location of the active site, protein production pattern and regulation). Until now, many of the tools effecting protein destabilisation have been engineered in eukaryotic systems whereas tool development in prokaryotes has just started [18] (INTRODUCTION section 2.2). A special feature for protein destabilising set-ups is the conditional re-stabilisation of destabilised proteins as a rescue of protein function technology (INTRODUCTION section 2.3). Such tools have been developed only in eukaryotes until now. Furthermore, despite very few exceptions [2,88,103], most of the available tools for protein destabilisation have been shown suitable for soluble proteins more than for MPs. In addition, besides one recently developed exception [18], all destabilising tools in bacteria were focused on C-terminal degradation [20,23,26].

To summarise, there is much room to increase the variability of protein destabilising tools particularly in bacteria, on one hand. On the other hand, there is a general need to expand or improve the already available tools for conditional MP degradation.

2.1 Degradation motifs suitable for synthetic biology

In most studies for engineered protein destabilisation, a known degradation motif is fused to the N- or C-terminus of a protein of interest (POI). Usually the tagged POI is stable until degradation is induced resulting in the rapid decrease in POI concentration [18]. Accordingly, a degradation motif is suitable for engineering set-ups when it can efficiently initiate processive proteolysis without the requirement of additional protein substrate information [61]. In all degradation motifs used for engineered protein destabilisation so far, the minimum requirement for a degradation motif, the tether-degron and initiation site-degron, seemed to be situated closely together on a small, intrinsically unstructured terminal sequence. This contributes to its transferability to a large number of different POIs which is an equal important factor for its application in synthetic biology set-ups [61,65].

However, besides the abundance of a whole degradation motif, success and efficiency of engineered POI degradation rely on several factors: (1) amino acids and/or folds in the neighbourhood of degrons modulating the degron's strength (e.g. prokaryotic N-terminal degradation sequence motif [50]; Figure 4);

and (2) the steric accessibility of degrons to the binding sites of adaptor proteins or proteasomes [18]. There are also (3) overlapping tether-degrons on a single degradation motif whose degradation efficiency is regulated by a competing recognition by proteasomes or adaptor proteins (e.g. SsrA-tag) [20]. Furthermore, it is essential that (4) the engineered degradation motif exhibits no POI-destabilising character on folding and that (5) it does not significantly effect POI activity or function [18]. Additional roles play (6) the concentration ratio between the POI and the recognising proteasome or adaptor protein [19], (7) the intrinsic thermodynamic stability, class as well as topology of the POI and (8) the efficiency as well as robustness of the recognising proteasome(s) [18,20].

Owing to genetic engineering and synthetic biology, many of the mentioned factors affecting strength, specificity and efficiency of a degradation system can be tuned to attain a robust and predictable tool to modulate protein stability [19].

2.1.1 Finding a transferable degradation motif

The identification of a native or synthetic, transferable degradation motif usually includes a minimum of two inter-dependent screening steps. First, a putative degradation motif candidate can act stabilising by turning it into a neutral sequence/fold [2,20] or by removing it from the native protein/POI to prove its destabilising character and/or decipher its relation to specific adaptor proteins or proteasomes. Second, it can be fused or grafted to a known stable POI [2,3,124] to prove its independence and transferability. In all cases levels of targeted proteins are observed over time and compared to those of the unaltered proteins. If the chosen destabilising motif is indeed a degradation motif, the degradation rate of the protein substrate is expected to be decreased during the first screening step and increased during the second.

2.1.2 Tuning the strength of degradation motifs

Degradation motifs may also be modified to create libraries with different degradation rates. This had been especially done by randomisation of whole native/synthetic degradation motifs or degron moieties which consisted of an intrinsically unstructured sequence or amino acid. A good examples are here: the randomisation of the N-terminal tether-degrons belonging either to the prokaryotic or eukaryotic N-terminal degradation motif processed by the respective N-end rule pathways [3,52,56]; the C-terminal prokaryotic SsrA-tag [20,125]; the N-terminal degradation sequence motif from YfgM in *E. coli* [2] and internal eukaryotic PEST sequences [126].

An extreme example for tuning the degradation rate is the stabilisation of a POI by turning a degradation motif or associated degron into a dysfunctional or stabilising motif. Examples from literature are again the N-terminal tether-degrons belonging an N-terminal degradation motif processed by either the prokaryotic [124] or eukaryotic [127] N-end rule pathways and the N-terminal degradation sequence motif from YfgM in *E. coli* [2].

2.2 Strategies for conditional destabilising of target proteins

With the help of a precisely regulated mechanism, an engineered biological set-up catalysing the removal of a POI can be turned into a valuable synthetic tool. The nature of the regulation mechanism can be manifold depending on the engineered degradation set-up, the degradation pathway, the POIs, the host and the final application. A non-exhaustive overview of tools for conditional protein destabilisation in eukaryotes and prokaryotes is provided in the following sections.

2.2.1 Tools for conditional regulation of eukaryotic ubiquitin-dependent proteolysis

The multi-component UPS in eukaryotes has been successfully manipulated on different levels. The tethering of a substrate protein either to the 26S proteasome or to enzymes that mediate poly-ubiquitination have been proven to be sufficient for increased degradation efficiency. Since the early 1990's, Ub-conjugating enzymes (E2s) were known to accomplish the Ub-transfer alone to the substrate protein whereby other Ub transfer reactions required the help Ub protein ligases (E3s) [128,129]. Later, it was shown that the former seems to be the exception and the latter the norm [130] since substrate specificity is usually conferred by E3s [15,130,131] but may be influenced by E2s [130]. Those findings were picked up in a couple of studies to engineer UPS with increased specificity for target proteins and therefore their efficient degradation.

In one study (demonstrating the exception), E2s were fused C-terminally with various POI-specific recognition domains to pre-define POI specificity for targeted Ub transfer. Subsequent ATP-dependent proteolysis was observed for one POI out of five *in vitro* [132]. Alternatively, in other studies (supporting the rule), POIs were tethered to E3s to increase ubiquitination efficiency by replacing the native substrate recognition sites in E3 with a POI binding domain. As a result, successful targeted POI degradation was observed *in vivo* [131,133]. The production of the aforementioned E2 or E3 chimeric proteins (with the re-defined recognition domains) was inducible so that POI-degradation was conditional.
In contrast, JANSE *et al.* (2004) used compound-inducible dimerisation domains to tether a POI in direct vicinity to the proteasome [134]. Finally, another less invasive approach is the Protacs (proteolysis-targeting chimeric molecules)-mediated tethering of POIs to an endogenous E3 which was demonstrated being successfully applicable *in vitro* [135] and *in vivo* [136].

2.2.2 Tools for conditional regulation of the C-terminal SsrA-mediated proteolysis

Until recently, the few tools available for conditional proteolysis in prokaryotes focused on degradation from the C-terminus; all engineered synthetic set-ups were based on a C-terminally attached SsrA-tag as active degradation motif [20,23,26]. Yet, the conditional elements responsible for inducing proteolysis differed from study to study.

The native SsrA-tag in *E. coli* harbours competing recognition motifs for the unfoldases ClpX and ClpA as well as SspB (adaptor protein for ClpXP) [49]. Moreover, SsrA-tagged proteins are also recognised by other proteasomes including Lon [85] and FtsH [86], as well as the proteases Tsp [82,83], DegP and DegQ, [38]. Based on this, McGinness at al. (2006) designed modified SsrA*-tags with recognition motifs almost inefficient for ClpX and ClpA binding but increased for SspB interaction [20]. Upon induced production of endogenous SspB in an *E. coli sspB*$^{-}$ strain, the SspB-dependent ClpXP degradation was increased for cytosolic proteins carrying the modified SsrA-tag [20].

In a similar study, Davis *et al.* (2011) used one of the modified SsrA*-tags from McGinness at al. (2006) [20], but designed a dual split-SspB protein as the conditional element (in an E. coli *sspB*$^{-}$ strain). Both SspB moieties were constitutively produced and dimerised upon supply of a small compound [23]. The set-up resulted in a conditional, specific and efficient tethering of the SsrA*-tagged POI to ClpXP.

By far the most elegant system for conditional protein degradation in bacteria was reported by Cameron and Collins (2014) [26]. The authors introduced successfully an orthogonally functional Lon protease from *Mesoplasma florum* (*Mf*Lon) with its cognate SsrA-tag (*Mf*SsrA-tag) into *E. coli* and *Lactococcus lactis* [26]. *Mf*SsrA -specific proteolysis was then regulated by induced protein production of *Mf*Lon *in-trans* and resulted in highly specific degradation [26]. Their system did not require the disruption or modification of the hosts proteolysis pathways [26], nor does the conditional element require a physical stress event for activation; in contrary, their system was characterised by high specificity, modularity and transferability to other bacteria as well as eukaryotes.

2.2.3 Tools for the conditional exposure of a site-specific N-terminal residue

To date, the synthetic conditional exposure of an N-terminal residue in a site-specific manner has been engineered in two different ways for the purpose of inducible proteolysis starting from the N-terminus. An endoproteolytic cleavage event (proteolytic processing) is part of both.

The classical way to release spontaneously an N-terminal residue is by using the Ub-fusion-technique, discovered in eukaryotes [127,137]. In yeast, a Ub-domain fused to the N-terminus of a POI became co-translationally cleaved by endogenous deubiquitylases after the last ^{76}Gly residue of Ub in a spontaneous fashion [52]. The cleavage was completely unbiased of the nature of the subsequent amino acid which became finally exposed [137]. Interestingly, the Ub-fusion-technique was used by Bachmair *et al.* (1986) in yeast for their elementary work on the eukaryotic N-end rule pathway in which the stability of a substrate protein was shown to be dependent on the identity of the N-terminal amino acid [48]. However, as mentioned before, the set-up as such is spontaneous in eukaryotes and is not subject to any conditional regulation. Only later, Dohmen *et al.* (1994) presented a eukaryotic system in which an N-terminal residue was conditionally exposed in two phases with the help of the Ub-fusion-technique and a conditional element. Firstly, an N-terminally fused Ub provoked a co-translational spontaneous Ub cleavage which in turn released an

N-degron of the N-end rule. The latter was the N-terminus of a temperature sensitive (ts)-protein domain that masked the N-degron at permissive temperatures (cryptic N-degron). Secondly, upon increased temperature and conformational destabilisation of the ts-protein domain, the cryptic N-degron became finally susceptible for degradation [138]. Until the early 1990s, the Ub-fusion technique had been inapplicable in bacteria [56] due to two reasons: (1) bacteria lack Ub and Ub-specific enzymes [56], and (2) the enzyme mapping within the family of Ub-specific proteases had just started in eukaryotes [139–141]. However, TOBIAS AND VARSHAVSKY (1991) isolated the UBP1 protease from *Saccharomyces cerevisiae* which cleaves off single Ub moieties in natural or engineered Ub-POI fusions irrespective of the size of the POI [140]. The same authors finally successfully applied UBP1 in *E. coli* in an inducible fashion to decipher the N-end rule criteria in bacteria: when produced, UBP1 efficiently deubiquitinated the Ub-X-POI construct and released the varying N-terminal residue X attached to the POI [56]. Finally, the nature of X regulated the degradation efficiency of the POI which finally let to the establishment of half-life values for the POI depending on X in *E. coli* [56].

The conceptual idea to use the highly specific protease from tobacco etch virus (TEVP) for the exposure of site-specific N-terminal residues was first mentioned by KAPUST *et al.* (2002) [142] and was finally realised by TAXIS *et al.* (2009) [143]. The authors developed a technique based on TEVP-induced protein instability (TIPI) which is universally applicable [143]. TIPI takes advantage of TEVPs high sequence specificity for amino acids one to six in its recognition site (rsTEVP) and its seemingly high tolerance towards amino acid seven (any amino acid with the exception of Pro) [142]. Upon TEVP cleavage, amino acid seven (*alias* amino acid in P1 position of rsTEVP [142]) becomes exposed as new N-terminal residue [143].

2.3 Strategies for conditional re-stabilising of target proteins

Only few tools for conditional re-stabilisation of artificially destabilised proteins do exist and all were developed in eukaryotic systems. The general goal for those tools is the efficient blocking - sometimes cleavage - of an attached degradation motif to a POI that otherwise directs proteolysis.

STANKUNAS *et al.* (2003), for example, made use of the FRB-FKBP-rapamycin[d] heterodimer formation system to stabilise an otherwise destabilised protein product [144]. The authors fused an intrinsically unstable variant of FRB (FRB*) to a POI. FRB* was able to heterodimerise with FKBP, another yet stable protein, depending on the availability of the small molecule rapamycin. Due to the instability of FRB*, the monomeric FRB*-POI fusion protein was rapidly degraded. However, the FRB*-POI fusion could be stabilised upon inducible dimerisation with FKBP *in vivo* [144]. However, this system could be used in both, eukaryotes and prokaryotes.

[d] Rapamycin, an antifungal antibiotic macrolide, binds simultaneously to the 12-kDa FK506 binding protein (FKBP) and the FKBP-rapamycin binding (FRB) domain and mediates their tight heterodimer formation. The FRB-FKBP-rapamycin heterodimer formation system is one of the most useful dimerization systems for protein engineering [303].

This inducible FRB*/ FKBP heterodimer system was picked up as conditional element by PRATT *et al.* (2007). The authors developed a tool based on split-Ub to expose a stabilising N-terminal residue in eukaryotes. Specifically, the authors fused FRB* and the C-terminal moiety of Ub (UbC) in-line to the N-terminus of a POI (FRB*-UbC-POI). The N-terminal moiety of Ub was fused to FKBP (UbN-FKBP) and produced constitutively. Upon the supply with rapamycin FRB*/FKBP and UbN/UbC could dimerise and complement, respectively. Consequently, the whole complementing protein complex FKBP-UbN/FRB*-UbC-POI got stabilised and became spontaneously cleaved after the complemented Ub by eukaryotic endogenous deubiquitylases (Ub-fusion-technique; INTRODUCTION section 2.2.3). The released N-terminal stabilising residue finally led to a stabilised POI. The tool was coined 'split ubiquitin for rescue of function' (SURF [145])[e].

[e] The conceptional idea of SURF could also be modified, as an idea, to release site-specific destabilising N-terminal residues in a conditional manner in eukaryotes, similar to the tools presented in INTRODUCTION section 2.2.3. However, the FRB*, attached to UbC-POI, would need to be replaced by a stable variant. Instead of a physical stress event for induction – as in the tool developed by DOHMEN *et al.* (1994) [138] (INTRODUCTION section 2.2.3) – this set-up would only require the addition of rapamycin for induction. A methodological advantage over TIPI [143] might depend on the application. Both systems have, from my point of view, advantages and disadvantages.

3 PROTEIC BUILDING BLOCKS FOR CONDITIONAL DESTABILISING OF TARGET PROTEINS IN *E. COLI* IN THE PRESENTED STUDIES

As discussed in INTRODUCTION section 1.2.2, one strategy nature has found to reduce specific protein levels is to generate degrons by revealing a degradation motif *via* proteolytic processing (*alias* proteolytic cleavage). In essence, a conditional endoproteolytic or exoproteolytic cleavage event releases a (pro-)degron which then can contribute to an active degradation motif for subsequent processive proteolysis.

In Studies I and II, we aimed to engineer conditional synthetic tools that were able to mimic the biological design of combined proteolytic processing and processive proteolysis. The difference between both studies were the targeted proteasomes, namely ClpAP (INTRODUCTION section 1.3) and FtsH (INTRODUCTION section 1.4) for Study I and II, respectively. This specificity was defined by the chosen N-terminal degradation motifs. In turn, the pool of targetable protein substrates was defined by the transferability of the respective N-terminal degradation motifs as well as the unfoldase capacity of the targeted proteasome. Based on the actual understanding about mechanism and efficiency of ClpAP-dependent proteolysis, the targetable protein pool for Study I was expected to encompass cytoplasmic globular proteins even with high intrinsic thermodynamic stability. Membrane proteins have, to my knowledge, never been published as substrate proteins for ClpAP/ClpS-mediated proteolysis. In contrast, the targetable protein pool for Study II was, to the greatest extend, unknown due to a very limited understanding about the FtsH-dependent N-terminal degradation sequence motif (dgFtsH; INTRODUCTION section 3.2.2) and its degradation pathway. However, dgFtsH seemed to exhibit a specificity for IMPs [2]; for FtsH accessibility, it appeared albeit imperative that the putative IMP targets exhibit a cytosolic N-terminal (N_{in}) topology. Nevertheless, the substrate specificity of FtsH is at the moment poorly defined, although it is expected that IMPs degraded thereby have a low intrinsic thermodynamic stability due to FtsH's poor unfoldase activity [103].

The feasibility and applicability of a conditional synthetic tool for targeted protein destabilisation are influenced by the following factors: (1) a conditional element that triggers efficiently the proteolytic processing and is easy to manipulate, (2) an active degradation motif that confers highest specificity and efficient degradation of a POI, (3) a spectrum of POIs with broad structural characteristics and (4) an easy and preferably non-invasive read-out.

The proteic building blocks which contributed to the tool development attempts in Study I and II are described in detail in the following sections. Overviews about characteristic details on the proteins touched upon and fluorescence properties of the chosen reporter proteins are given in the APPENDIX Table S 3 and Table S 2, respectively.

3.1 TEV protease and its recognition site

As described in INTRODUCTION section 2.2.3, the technology for Tobacco etch virus protease (TEVP)-induced protein instability (TIPI) [143] is a universal approach to conditionally expose almost any N-terminal residue that is part of an N-terminal degradation motif. Study I and II, described in this thesis, were inspired by this TIPI technology [143]. Essentially, two N-terminal degradation sequence motifs (Table 2) were accordingly masked with a short peptide sequence including the first 6 aa of the optimum recognitions site of TEVP (rsTEVP; Table 2) in order to develop short N-terminal conditional degradation tags for ClpAP- or FtsH-specific proteolysis (cryptic degradation motifs; Table 3). As described for TIPI [143], the TEVP was conditionally produced *in-trans* from a plasmid. In short, upon conditional production of TEVP, the cryptic N-terminal degradation motifs were released and accessible for the respective proteomes for subsequent degradation.

TEVP (EC 3.4.22.44) is a 27 kDa protease from the tobacco etch virus (TEV) [142,146,147]. It is part of the natural TEV precursor polyprotein and also responsible for its cleavage into functional viral proteins [142,146–149]. TEVP has a high sequence specificity for amino acids within the first 6 aa in its 7 aa long recognition site (rsTEVP) and cleaves specifically after ^{6}Q [142,147]. The consensus recognition site from all rsTEVPs is represented as EXXYXQ↓G/S whereas ENLYFQ↓S/G has been shown to be optimal for TEVP recognition [149]. This makes it a unique endopeptidase that has been largely exploited for multiple biotechnological applications [147,149]. TEVP exerts a high tolerance (however not entirely unaffected) to the nature of amino acid seven in rsTEVP [142,147] which is most relevant for TIPI-dependent release of N-terminal destabilising residues. Furthermore, TEVP is monomeric in solution [150].

The TEVP variant used in Study I (Addgene ID 19978) consisted of residues 1–236 of the Nuclear Inclusion a protease (NIa) from TEV [148] including a S219D mutation known for moderating resistance towards auto-proteolysis without compromising its catalytic activity [142,146]. The influence of amino acid seven on cleavage efficiency for this TEVP mutant (referred to as TEVP for simplification; Figure 6f top) has been examined recently [142]. In this study, any amino acid in position seven (with the exception of Pro) became exposed as new N-terminal residue upon TEVP cleavage with similar efficiency rates *in vivo* and *in vitro* [143] - yet seemingly with highest efficiency for any amino acid with short, unbranched polar or nonpolar side chain (Gly, Ser, Ala, Met, Cys). Additionally, its enzymatic activity was determined to be maximal at about 30 °C with a gradual decrease above 34 °C, most probably due to unfolding effects [151]. Equally important, enzymatic activity at 4 °C was shown to be very low but existent [151]. Even though limited heterologous production of active TEVP has been reported due to poor solubility (< 1 mg/mL) [146,152], preliminary experiments under the conditions for Study I showed satisfying activity (even at 37 C) in comparison to TEVP variants that were N-terminally stabilised with MBP (Addgene ID 8827 and 8835, data not shown).
In Study II, the TEVP mutant described above was electively used with a C-terminal polyhistidine (His_8)-tag (referred to as TEVP-His; Figure 6f bottom). An active TEVP variant with a C-terminal poly-arginine (RRRRR)

tail on the C-terminus had been recently published [146]. Therefore, the C-terminal (His_8)-tag was not expected to impair the activity of TEVP-His but still experimentally tested (STUDY II section 3.3). Both, TEVP and TEVP-His were produced each from a pSEVA33 backbone (pBBR1 replicon, CmR) under the control of the inducible rhamnose promoter *rhaBp* (INTRODUCTION section 4.1.3).

3.2 Degradation motifs

A necessary prerequisite for the chosen transferable degradation motifs used in Studies I and II was their N-terminal localisation. The idea was to mask each with an N-terminally fused oligopeptide harbouring the recognitions site for the TEV protease (rsTEVP) which would enable their conditional release and susceptibility for the respective proteasomes upon TEVP cleavage (details in INTRODUCTION section 3.1). Additionally, the degradation motifs were preferred to be short peptide sequences whose respective short coding sequence would be advantageous for efficient genome integration or optional sequence randomisation by single-stranded DNA (ssDNA) recombineering for the generation of protein libraries.

3.2.1 Synthetic ClpAP-specific degradation motif

The N-terminal degradation motif for ClpAP-specific degradation used in Studies I and II (referred to as unstable dgClpAP or short dgClpAP; Table 2B) derived originally from a synthetic oligopeptide developed by ERBSE *et al.* (2006) [3] (Table 2-B). This oligopeptide sequence (12 aa) is based on the modified 11 N-terminal residues of native *Av*GFP (*Av*GFP_{aa1-11}, Table 2B) including a short α-helix motif (GEELF, annotated in UniProt ID#P42212) [3]. ERBSE *et al.* (2006) substituted the N-terminal ^{1}Met in *Av*GFP_{aa1-11} with Phe-Arg [3]. This synthetic Phe-Arg N-terminus is mimicking the result of an *in vivo* post-translational aminoacylation of Arg in *E. coli* (secondary degron); hereby aminoacyl-tRNA-transferase Aat would covalently attach the primary degron residue Phe [52]. As a consequence, proteins beginning with the N-terminal primary destabilising Phe residue (equivalent a primary tether-degron of an prokaryotic N-terminal degradation motif; Figure 4) are recognised with highest efficiency by ClpS [33,52].

As an important side note, the final unstable dgClpAP used in Study I and II showed one unintended amino acid substitution in position 8 at the end of the short α-helix motif, with respect to the oligopeptide from ERBSE *et al.* (2006) (see analysis in Table 2B). The ^{8}Phe[TTC]→^{8}Val[GTC] substitution is most probably a genetic design-flaw which occurred at an early stage of Study I. The impact of the ^{8}Phe→^{8}Val substitution on the short α-helix motif (GEELF) and eventual strength of the whole degradation motif for ClpAP-specific degradation, with respect to the original oligopeptide, is unclear. Both amino acids are non-polar and show similar helix propensity values [153], thus, we speculate that ^{8}Val is still part of the helix, although the four amino acids GEEL solely might be enough to fold into an α-helix alone [154]. Furthermore, the hydrophobic element (equivalent to the initiation site-degron of a prokaryotic N-terminal degradation motif; Figure 4) is

speculated to withhold its ClpA interacting properties despite the substitution (probably ^{7}Leu-^{8}Phe which turned upon substitution into ^{7}Leu-^{8}Val). Nevertheless, a general destabilising effect on proteins by the unstable dgClpAP was demonstrated in Study I.

In order to contrast the effect of the unstable dgClpAP on a protein, a stable variant of dgClpAP (dg$^{\#}$ClpAP) was engineered and used in Studies I and II. Thereby, the N-terminal Phe was replaced with Ala. Ala is known as N-terminal stabilising residue and is accordingly not recognised by ClpS (dysfunctional tether-degron). Thus, proteins N-terminally tagged with dg$^{\#}$ClpAP are supposed to have a considerably longer half-life [56] (Table 2B).

Table 2: **Building blocks for the design of conditional N-terminal degradations tags for ClpAP- or FtsH-mediated proteolysis**. The Amino acid properties are depicted for all sequences. (A) Short peptide sequence (rsTEVP[X]) that is designed to mask the N-terminal degron variants for their conditional release. The first 6 aa of the optimum rsTEVP are underlined. The TEVP cleavage site is annotated with an arrow. X represents the first N-terminal residue of a cryptic N-terminal degradation motifs that becomes released upon TEVP cleavage. (B) Stable and unstable ClpAP-specific degradation sequence motifs and multiple sequence alignment to their parental peptide sequences (see main text for explanation). (C) Unstable FtsH-specific degradation sequence motif. Multiple sequence alignment and determination of amino acid properties were done with Clustal Omega [155] and the Peptide Property Calculator (GenScript), respectively. Acidic (red), basic (blue), hydrophobic uncharged (green) and other residues (dark grey) are coloured as indicated. Consensus symbols in the multiple sequence alignment like fully conserved residues (*) and residues with conserved groups of weakly similar properties (.) are indicated accordingly. Amino acids annotated for the first α-helix motif in *Av*GFP are underlined.

NAME	N-TERMINAL SEQUENCE	REFERENCE
(A)		
rsTEVP(X)	MVENLYFQ↓X–	Study I
(B)		
*Av*GFP$_{aa1-11}$ (parental sequence)	MSKGEELFTGV–	[156]
oligopeptide from ERBSE *et al.* (parental sequence)	FRSKGEELFTGV–	[3]
unstable dgClpAP (dgClpAP)	FRSKGEELVTGV–	Study I
stable dgClpAP (dg$^{\#}$ClpAP)	ARSKGEELVTGV–	Study I
	******.***	
(C)		
unstable dgFtsH (dgFtsH)	MEIYENENDQVEAV–	[2]

3.2.2 Native FtsH-specific degradation motif from YfgM

YfgM has recently been identified as substrate for the FtsH proteasome in *E. coli* [157]. YfgM is a bitopic MP located in the IM. It consists of a short cytoplasmic N-terminal region, a single transmembrane domain, and a longer C-terminus located in the periplasm (1 TMH, N_{in}-C_{out} topology) [2,158]. It seems to play multiple roles in the two cellular compartments it is adjacent to: (1) in the periplasm, it serves as chaperon as it interacts with the chaperone PpiD and with the SecYEG translocon in *E. coli* [100,158,159]; and (2) it acts as negative

regulator for RcsB in the cytoplasm [2] which is in turn a regulator within the envelope Rcs stress response system [160]. There is substantial evidence that YfgM is degraded by the FtsH proteasome in a growth phase-dependent fashion, most efficiently during stationary-phase stress conditions [2,100,157]. Thus, YfgM has been stated to be the first MP degraded by FtsH in a conditional manner; the presence of an adaptor protein is suspected [2]. The degradation motif for FtsH-specific degradation of YfgM (dgFtsH) has been recently determined as being its first 14 N-terminal cytoplasmic residues (Table 2-C) by BITTNER *et al.* (2015) [2]. Thus, degradation efficiency is sequence-dependent which is unique among all known FtsH substrates so far [2]. An indication for a conditionally abundant adaptor protein is given in the importance of different amino acids within the dgFtsH peptide sequence during different growth phases (Table 2-C): E2 and I3 are essential for degradation during stationary phase, while Y4 and E5 are important for stabilisation during early growth phases [2]. Residues 6 to 14 seem to play a minor role in YfgM degradation [2]. A small library of 26 systematic point-mutations and deletions let to this result and YfgM degradation was visualised by immunoblotting [2]. Alternatively, PIATKOV *et al.* (2015) suspect ^{1}Met in dgFtsH being an FtsH-dependent fMet/N-degron (see details in INTRODUCTION section 1.2.2) by referring to unpublished data and postulate also an interplay with a putative adaptor protein [69]. Interestingly, from a synthetic biologist's point of view, dgFtsH was graftable onto the cytoplasmic N-terminus of another IMP with similar topology (PpiD) and thereby conferring the same characteristic growth phase-dependent degradation profile as shown for YfgM [2]. In contrast, dgFtsH grafted onto soluble glutathione S-transferase (GST) did not convert the latter into an FtsH substrate, which indicated a specificity for MPs [2].

In Study II, dgFtsH was used as N-terminal degradation motif for FtsH-specific degradation and grafted onto the membrane proteins PpiD, YhaI and NavZ1.

3.2.3 Engineering N-terminal degradation tags for the conditional release of N-terminal degradation sequence motifs

In Studies I and II we designed three N-terminal degradation tags. Inspired by the TIPI technology [143], we masked N-terminal degradation motifs specific for ClpAP- or FtsH-dependent degradation (Table 2B and C), with a short oligopeptide including the first 6 aa of the optimal TEVP recognition site (rsTEVP; Table 2A) . The release of the N-terminal degradation motifs is directed by the conditional production of TEVP and subsequent TEVP-dependent cleavage within rsTEVP(X) (INTRODUCTION section 3.1). X represents the first N-terminal residue of a degradation motif that gets released. An overview of the resulting degradation tags is given in Table 3.

Table 3: **Overview about the N-terminal degradation tags for the conditional release of unstable or stable N-terminal degradation sequence motifs for ClpAP- or FtsH-mediated proteolysis used in Study I and II**. rsTEVP(X) (orange); Remaining peptide sequence after TEVP cleavage (underlined); α-helix (red highlighted). The N-terminal amino acid residue X which corresponds to the seventh amino acid in rsTEVP(X) is emphasised in brackets.

CONDITIONAL DEGRADATION TAG	PEPTIDE SEQUENCE	AFFILIATION
Unstable rsTEVP(F)-dgClpAP	MVENLYFQFRSKGEELVTGV–	Study I, Study II
Stable rsTEVP(A)-dg#ClpAP	MVENLYFQARSKGEELVTGV–	Study I, Study II
Unstable rsTEVP(M)-dgFtsH	MVENLYFQMEIYENENDQVEAVKRFFA–	Study II

3.3 Proteins of interest

Variants of the green fluorescent protein from *A. victoria* (*Av*GFP) have been proven to be prime candidates to demonstrate the utility of experimental set-ups for protein degradation of soluble proteins [26,50,103]. *Av*GFP variants are soluble, easy to produce to high levels in the cytoplasm of *E. coli* [161], known to have a long half-life [125] and have a modifiable N-terminus [162]. They embody their own reporter system due to their ability to exhibit, if functional, a bright green fluorescence as response to light in the UV-blue light range [163]. Upon induced proteolysis, the fluorophore which is composed of modified amino acid residues within the polypeptide chain [163,164] should become destabilised, thereby affecting the proteins' fluorescence intensity. *Av*GFP variants represent therefore an easy read-out to analyse degradation events in the cytosol. In Study I, we chose the GFP folding reporter (frGFP; INTRODUCTION sections 3.3.1 and 3.4.1) as POI to test our set-up for conditional ClpAP-mediated proteolysis.

In contrast, choosing candidates for conditional FtsH-dependent proteolysis (Study II) was not trivial since there are no membrane proteins (MP) that also exhibit a fluorescent signal. The monitoring of MP proteolysis needed to be reconsidered and we opted for designing fusion proteins using fluorescent reporter (FR) proteins tagged on model MPs. We defined three criteria we believed essential for our MP candidates. First, their N-terminus needed to be cytoplasmic in order to make it accessible by the FtsH proteasome. Second, successful overproduction and integration in *E. coli*'s IM (preferably also as C-terminal fusion to a peptide/protein) should have been reported in the literature. Third, an excess of more than three TMHs was not desired to limit the complexity in preliminary experiments. Good matches were found for the three inner membrane proteins PpiD, NavZ1 and YhaI (see Figure 6b, c, d and INTRODUCTION sections 3.3.2, 3.3.3, 3.3.4, respectively).

Table 4: **N-terminal peptide sequences of all proteins of interest used in Studies I and II**. The N-terminal peptide sequences from all soluble or IM-integrated POIs include the sequence until the end of the first N-terminal β-sheet or until the end of the first transmembrane helix (TMH), respectively. Annotated α-helix (red highlighted); Annotated β-strand (light blue highlighted); Predicted TMH (grey highlighted); Position at which the N-terminal sequence was replaced with an unstable or stable conditional degradation tag from Table 3 (◆). YfgM is only indicated for comparison. Secondary structure annotations for frGFP were obtained from UniProt (P42212 [*Av*GFP]). TMH prediction was performed with ΔGpred [165].

POI	N-TERMINAL PEPTIDE SEQUENCE	REFERENCE
frGFP	M◆SKGEELFTGVVPILVELDGDV-frGFP$_{aa23-238}$	[4,166]
YfgM	MEIYENENDQVEAVKRFFA-ENGKALAVGVILGVGALIGWRYW-YfgM$_{aa43-206}$	[2]
PpiD	MMDSLRTAANSL◆VLKIIFGIIIVSFILTGVSGYL-PpiD$_{aa35-623}$	[2]
NavZ1	MLKRCLS◆PLTLVNQVALIVLLSTAIGLAGMA-NavZ1$_{aa32-467}$	Study II
YhaI	MQWYLSVLKNYVGFS◆GRARRKEYWMFTLINAIVGAIINVIQLIL-YhaI$_{aa45-118}$	Study II

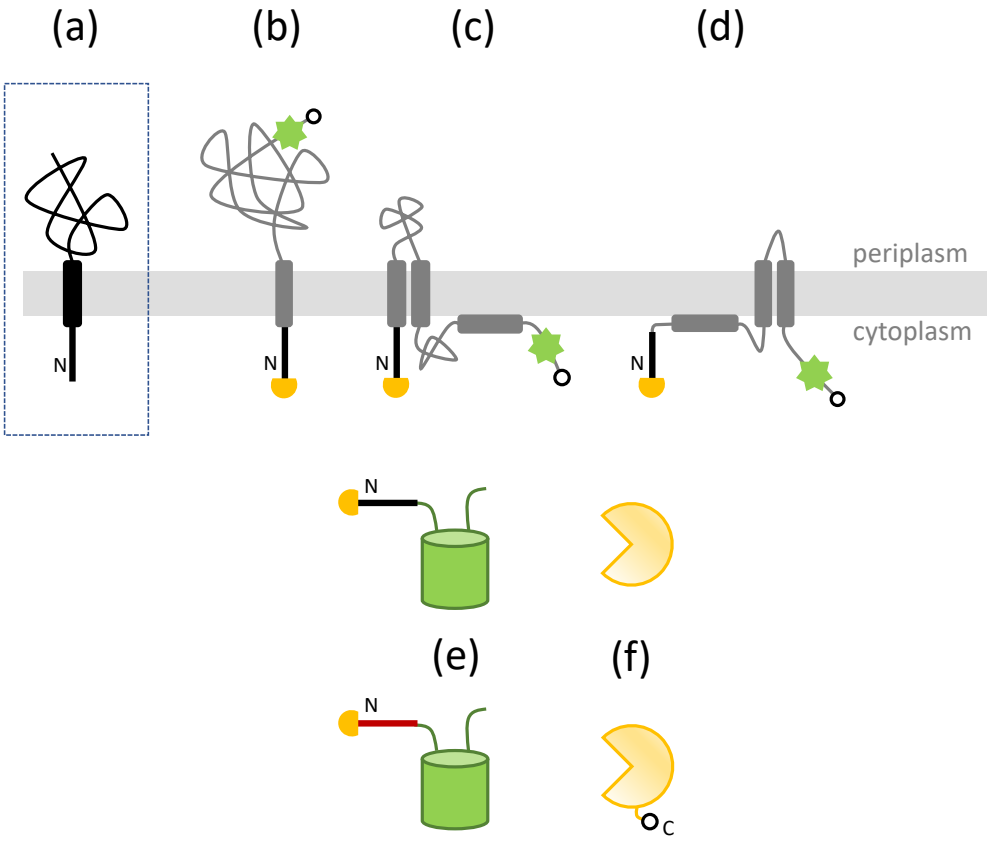


Figure 6: **Schematic illustration of the general structure and localisation of the POIs and TEV protease**. PpiD (b), NavZ1 (c), YhaI (d) as well as the variants for frGFP (e) and TEV protease (f) are presented. YfgM (a) is shown for purpose of topological comparison. All N-termini in (b-e) are shown with their conditional degradation tags. The N- or C-terminus are selectively indicated with an N or C, respectively. Important N- or C-terminal protein features are indicated in colour or symbols: synthetic dgClpAP or dg[#]ClpAP (red bold line) and native dgFtsH (black bold line) for ClpAP- and FtsH-dependent proteolysis, respectively; the rsTEVP(X) capping (◖) as part of the conditional element together with TEVP (f), varying fluorescence reporter including sfGFP, EcFbFP, phiLOV2.1 and CreiLOV are summarised in one symbol (✸); His_8-tag (O) as epitope tag.

3.3.1 The globular frGFP

The GFP folding reporter (frGFP) [4] is used simultaneously as cytoplasmic FR as well as soluble POI in Study I. frGFP is a full-length *Av*GFP variant with six characteristic amino acid substitutions which are discussed in detail in INTRODUCTION section 3.4.1. Until now, frGFP has no own Protein Data Bank (PDB) entry [167] but shows 97.5 % sequence similarity to *Av*GFP as calculated by ExPASy SIM [168]. Thus, frGFP may most certainly exhibit the same structure as *Av*GFP that is a globular α/β-protein [169] folded as a β-barrel around an axial α-helix [163].

For our Studies I and II, the aa_{2-238} of frGFP (Table 4) were N-terminally fused with either the unstable degradation tag rsTEVP(F)-dgClpAP-frGFP or the stable tag variant rsTEVP(A)-dg#ClpAP (Table 3). The resulting protein constructs were then (see Figure 6e):

- rsTEVP(F)-dgClpAP-frGFP
- rsTEVP(A)-dg#ClpAP-frGFP

Upon TEVP cleavage in the degradation tag sequences, Phe (F) or Ala (A) are released as N-terminal destabilising or stabilising residues for ClpAP-dependent proteolysis, respectively. The coding sequences for both frGFP fusions have been independently integrated in the phage 186 integration site within the genomic *tRNIleY* gene of *E. coli* K-12 MG1655 and expressed by the IPTG-inducible *T5* promoter.

3.3.2 The membrane protein PpiD

PpiD is a bitopic MP in the IM of *E. coli* (1 TMH, N_{in}-C_{out} topology); its single N-terminal TM helix is followed by three predicted domains facing the periplasm [170,171]. Only the second periplasmic domain has been structurally determined so far [171]. The first and the third periplasmic domains are putative chaperone domains, whereas the second domain is predicted to have *cis-trans* isomerase activity [171] for the catalysis of peptidyl-prolyl bonds which is, beside disulphide bond formation, another rate-limiting reaction in native protein folding [172].

There are flagrant conflicting debates in literature regarding putative functions of PpiD, namely (a) being essential for growth, (b) exhibiting a specific folding function in outer membrane (OMP) biogenesis or (c) exhibiting a compensating function for SurA[f]; often those three hypotheses are supported by controversial results ([170] versus [173], [174] versus [175], and [170] versus [175], respectively). For now, the overall function(s) of PpiD remain(s) unclear [176,177]. Nevertheless, considering its co-localisation together with SecYEG, PpiD has been proposed as periplasmic gatekeeper and assisting in the initial folding steps of newly translocated proteins prior to their release into the periplasm as chaperone [174,178]. If so, its substrates are more likely envelope proteins [175] and less probably OMP's [174].

[f] SurA is a periplasmic chaperone which is known to prevent aggregation of unfolded OMPs [304].

Due to its topological resemblance (N_{in}-C_{out}) and its apparent complex formation *in vivo* with YfgM (1 TMH, N_{in}-C_{out} topology; INTRODUCTION section 3.2.2) in the IM [158], PpiD has already been used in the fundamental work of conditional proteolysis of IMPs by BITTNER *et al.* (2015). The authors replaced the initial twelve N-terminal residues of PpiD with the N-terminal $aa_{1\text{-}19}$ of YfgM which finally led to FtsH-mediated destabilisation in the stationary phase in E. coli K-12 W3110; $YfgM_{aa1\text{-}14}$ were determined to represent the degradation signal [2] (*alias* dgFtsH; INTRODUCTION section 3.2.2). The coding sequence for the dgFtsH-$PpiD^{N'}$ chimera was expressed from the pASK-IBA5plus derivative pBO3575 (ColE1 replicon, AmpR; personal communication with IBA GmbH, Germany and Lisa-Marie Bittner) by the anhydrotetracycline (AHT)-inducible *tet* promoter. Production of the dgFtsH-$PpiD^{N'}$ MP chimera was verified from whole cells by immunoblotting using an α-PpiD antibody [2]. The latter detection method gives evidence about quantitative NavZ1 abundance, yet it does not differentiate in between folded and aggregated MP [179].

The study of BITTNER *et al.* (2015) embodies the foundation for the presented Study II. In contrast to BITTNER *et al.* (2015), we replaced the N-terminal $aa_{1\text{-}12}$ of PpiD with the unstable conditional degradation tag rsTEVP(M)-dgFtsH (Table 3) for FtsH-dependent degradation. Additionally, the FR proteins sfGFP or EcFbFP (INTRODUCTION section 3.4.2 or 3.4.3, respectively) were attached directly to PpiD's C-terminus including an additional KLAA-$(His)_8$-tag. Hence, the resulting PpiD chimera were then in short (see Figure 6b):

- rsTEVP(M)-dgFtsH-PpiD-EcFbFP-His
- rsTEVP(M)-dgFtsH-PpiD-sfGFP-His

The respective coding sequences were integrated in a pCOLA vector (ColA replicon, KanR) and expressed from the IPTG-inducible *T7* promoter in *E. coli* MG1655 (DE3) in Study II.

3.3.3 The membrane protein NavZ1

The synthetically engineered NavZ1-W0 protein (2 TMHs, N_{in}-C_{in} topology [180]), referred to as NavZ1 below, is a functionally impaired polytopic MP chimera from the two functionally and topologically similar osmolarity sensor histidine kinases NarX and EnvZ [5]. The two N-terminal TMHs and the periplasmic sensor domain derived from the $aa_{1\text{-}217}$-moiety of NarX while the cytoplasmic TM-free C-terminal part derived from the $aa_{222\text{-}450}$-moiety of EnvZ [5]. Two TMHs and an N_{in}-C_{in} topology were determined for NarX and EnvZ by an experimental and in silico analysis on the N- and C-termini [180]. The same topology has been determined experimentally for the chimeric NavZ1; furthermore, NavZ1 dimerises but N- and C-termini were expected being modifiable and the N-terminus being exposed enough to be accessible by a cytoplasmic protein (personal communication with Roger Draheim).

NavZ1 was considered as an ideal MP candidate due to its cytoplasmic N-terminus and its impaired activity to prevent additional physiological stress to the cells throughout IMP-overproduction. Additionally, its

topology with two TMs was thought to be an advantage since this represents an increased structural complexity in comparison to PpiD. NavZ1 had previously been successfully produced from a modified pBR322 vector, pEnvZ-V5 (pMB1 replicon, AmpR, *lacZ* promoter), with a C-terminal V5 epitope tag in *E. coli* ($NavZ1^{+}$-aa_7-V5); protein production was monitored from whole cells by WESTERN blot using an α-V5 antibody [5]. One can speculate that the analysis use here is not enough to demonstrate successful integration of the folded protein into the membrane.

Similar to PpiD, the first 7 aa at the N-terminus of NavZ1 were replaced with the degradation tag rsTEVP(M)-dgFtsH (Table 3) for conditional FtsH-dependent degradation in Study II. Furthermore, NavZ1's C-terminus was directly fused with the FR proteins CreiLOV, phiLOV2.1 and sfGFP combined with a KLAA-$(His)_8$-tag. Accordingly, the resulting NavZ1 chimera (Figure 6c) were then in short:

- rsTEVP(M)-dgFtsH-NavZ1-CreiLOV-His
- rsTEVP(M)-dgFtsH-NavZ1-phiLOV2.1-His
- rsTEVP(M)-dgFtsH-NavZ1-sfGFP-His

Equally to PpiD, the respective coding sequences were integrated in a pCOLA vector (ColA replicon, KanR) and expressed from the IPTG-inducible *T7* promoter in *E. coli* MG1655 (DE3) in Study II.

3.3.4 The membrane protein YhaI

YhaI has been assigned so far as a MP of the IM [180,181]. It is the protein product of the *yhaI* gene which is yet a functionally uncharacterised and unnamed ORF that maps at 81 minutes on the genetic map of *E.coli* K-12 MG1655 [182]. Therefore, the *yhaI* gene still belongs to the number of true y-genes which make still approximately 30% of the 4,377 annotated genes in *E. coli* K-12 MG1655 [181].

In retrospect, YhaI does not seem to be a perfect choice for Study II due to uncertainties in its topology. A THM analysis with ΔGpred [165] predicts three TM helices for YhaI (analysis in APPENDIX Table S 4). This was confirmed by a TMHMM [183] analysis which simultaneously predicted an N_{in}-C_{out} topology (analysis in APPENDIX Figure S 2). In contrast, DALEY *et al.* (2005) determined experimentally the C-terminus to be cytosolic and predicted a N_{in}-C_{in} topology with three predicted THMs [180]. Thus, further experimental analyses are needed to firmly establish the topology and structure of this protein. However, the five positively charged amino acid residues (Lys, Arg) on the N-terminus (Table 4) could represent a strong argument for the cytosolic localisation, as needed for Study II, since basic residues have been reported being more abundant on cytosolic connecting loops than on periplasmic ones [184]. Thus, besides a confirmed C_{in} topology, the cytoplasmic localisation of the N-terminus seems probable but still speculative. Similar speculative is the orientation of the three predicted TMH towards the membrane. In contrast to ΔGpred-predicted TMH2 and TMH3 (analysis in APPENDIX Table S 4), the ΔG^{prep} value for the predicted TMH1 is positive (0.283) which is an indication of lower hydrophobicity and also that it would need stabilising helices around it for an efficient insertion into the membrane [165]. Based on this, one possible scenario is that TMH1

of YhaI can be amphipathic which might be emphasised in the one-sided arrangement of its hydrophobic residues in a helical wheel display (analysis in APPENDIX Figure S 1) generated with Membrane Protein Explorer [185].

In essence, YhaI has been chosen for Study II due to its assumed cytoplasmic N-terminus, its small size and its topology with three helices associated with/in the IM. Here too, the predicted structure was thought to represent an increased complexity in comparison to PpiD and NavZ1. Additionally, despite its unknown function, YhaI has already been seemingly, successfully produced as fusion protein in *E. coli* in three instances: (a) from a modified pET28a(+) vector pWaldo (pBR322 replicon, *T7p*) [4] as YhaI^{+}-linker-rsTEVP(G)-linker-S11-linker-His construct (Morten HH Nørholm, unpublished data, YhaI abundance was confirmed based on fluorescence analysis of splitGFP complementation on whole cells); (b) from modified pBR322 vectors, pHA-1 and pHA-4 (pMB1 replicon, *araBp*), as YhaI^{+}-linker-PhoA$^{N'}$, whereby cytoplasmic located PhoA activity was measured *in vivo* [180]; and (c) from the modified pET28a(+) vectors pWaldo (pBR322 replicon, *T7p*) [4] or pGFPe (pBR322 replicon, *T7p*) [186] as YhaI^{+}-linker-GFP$^{N'}$ or YhaI^{+}-linker-rsTEVP(G)-GFP$^{N'}$-His, whereby cytosolic GFP abundance was measured as normalised emitted GFP fluorescence/OD_{600} [180]. For all genes in examples (b), (c) - and most probably also in (a) - the DNA-coding region encompassing the region coding for the SD sequence until the ATG codon of the coding region, which only partially represents the translation initiation region (TIR[g]), was always the same [180]. One should note, that fluorescence alone is not enough to demonstrate that the protein was folded correctly and integrated into the membrane [187].

In Study II, the first 15 aa of the N-terminus were replaced with the degradation tag rsTEVP(M)-dgFtsH (Table 3) for conditional FtsH-dependent degradation. Furthermore, equally to NavZ1, YhaI's C-terminus was directly fused with the FR proteins CreiLOV, phiLOV2.1 and sfGFP combined with a KLAA-$(His)_8$-tag. Accordingly, the resulting YhaI chimera were in short (Figure 6d):

- rsTEVP(M)-dgFtsH-YhaI-CreiLOV-His
- rsTEVP(M)-dgFtsH-YhaI-phiLOV2.1-His
- rsTEVP(M)-dgFtsH-YhaI-sfGFP-His

Equally to PpiD and YhaI, the respective coding sequences were integrated in a pCOLA vector (ColA replicon, KanR) and expressed from the IPTG-inducible *T7* promoter in *E. coli* MG1655 (DE3) in Study II.

[g] The translation initiation region (TIR) is the nucleotide sequence around the START codon on the mRNA which is responsible for the efficiency of translation initiation and therefore protein biosynthesis. In detail, TIR comprises the Shine-Dalgarno (SD) sequence, a linker region in between the SD sequence and the translational START codon (^{1}Met), approximately the first five codons after ^{1}Met and optional translational enhancers such as A/U-rich sequences upstream of the SD sequence [305,306].

3.4 Fluorescent reporter proteins

An easy and non-invasive read-out suitable for real-time observation in vivo of biological processes like protein degradation are genetically encoded fluorescence reporters (FRs). Genetically encoded FRs are protein sequences that can be fused to a POI to render it fluorescent (recently reviewed [188]). At best, upon unfolding and processive proteolysis of the FR, the fluorophore becomes unprotected resulting in a loss of function. Proteasomal processivity is, among others, dependent on the unfolding power of the proteasome as well as on the intrinsic thermodynamic stability of the protein substrate [41,189]. In contrast to eukaryotes that have only one highly processive proteasome [190], *E. coli* is equipped with five different proteasomes which apparently vary not only in their substrate preferences but also in their unfolding power [33,41,103,191]. ClpAP is known for having a strong unfoldase activity [41,42] whereas the unfolding power of FtsH has been rated as weak [103]. Thus, in order to visualise FtsH-dependent degradation, the FR choice for Study II needed careful considerations to avoid incompletely degraded FR that might be still (partially) fluorescent [103]. In the ideal case, any chosen FR should be sufficiently thermodynamically unstable to match the unfoldase activity of the targeted proteasome.

Since the discovery of *Av*GFP and the determination of its primary and tertiary structure in 1962 [192], 1992 [156] and 1996 [166,193], respectively, FRs from the GFP family and its modified variants have become the prime choice for scientific fluorescent applications. They belong to the class of intrinsically fluorescent proteins which become fluorescent after folding without addition of a fluorophore under aerobic conditions [188]. In the past, *Av*GFP variants have been successfully used to study cytosolic ClpAP-dependent degradation in *E. coli* [3,41,42]. Crucial characteristics for this application were its bright green fluorescence when exposed to blue/UV light, its photostability and long half-life and its modifiability at the C- and/or N-terminus. Furthermore, *Av*GFP variants can be produced to high levels in the cytosol of *E. coli*. Thus, it was obvious to use a similar *Av*GFP variant - frGFP (INTRODUCTION section 3.4.1), to be precise – for the tool development part in Study I.

In order to fold correctly, *Av*GFP variants usually need the reducing environment of the cytoplasm to form disulphide bridges [194,195]. The amino acid substitutions in superfolder GFP (sfGFP; INTRODUCTION section 3.4.2), however, contribute that this *Av*GFP variant is still able to become stably fluorescent in the highly oxidising periplasmic environment under aerobic conditions [6,194–196]. Despite the apparent weak unfoldase activity of FtsH [103], sfGFP was integrated as C-terminal FR-tag in Study II in order to shed some more light on FtsH-dependent IMP degradation in general. We hypothesised that in case sfGFP were not or only partially degraded, it would remain as entity in periplasm/cytoplasm and could be detected with immunoblotting.

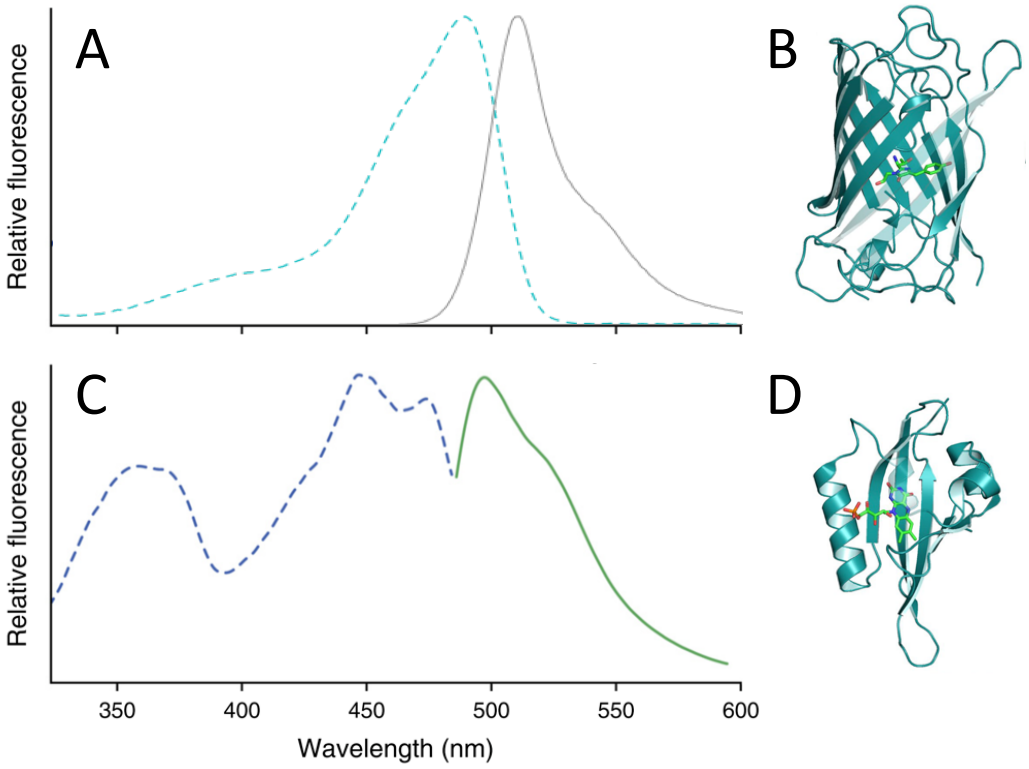


Figure 7: **Fluorescence properties and structures of GFP and LOV protein variants**. Excitation (dashed lines) and emission (solid lines) spectra of (A) EGFP, an *Av*GFP variant, and (C) iLOV, the parental LOV protein from phiLOV2.1. Almost all LOV proteins exhibit a characteristic excitation spectra with a prominent three band feature in the blue spectral region and a maximum near 450 nm; the respective emission spectra are also spectrally very similar with a specific fluorescence maximum around 495 nm and a prominent shoulder in the range of 525–540 nm [197–200]. Structural ribbon diagrams of (B) *Av*GFP, as an intrinsically fluorescent protein (PDB ID 1EMA [193]) and (D) iLOV, the parental LOV protein from phiLOV2.1, as an extrinsically fluorescent protein that binds FMN as the fluorophore (PDB ID 4EES [201]). All protein backbones are shown as cartoon, and the fluorophore is drawn as a stick model (green, carbon; blue, nitrogen; red, oxygen; orange, phosphorus). In some cases, parts of the protein cartoon have been made transparent to better show the fluorophore. (A) has been generated with Spectrumviewer (BD Biosciences); (C) is reproduced from BUCKLEY *et al.* (2015) [202]; (B) and (C) are a reproduction from THORN (2017) [188].

A new emerging generation of genetically encoded FRs for bioimaging are chromophore-binding domains from plant and bacterial photoreceptor proteins; in particular the flavin-binding light-oxygen-voltage (LOV) domains [202]. Their discovery and first application as FRs took place in 1999 [203] and 2007 [199], respectively. LOV domains or LOV proteins are native protein sensors [197] and belong to the extrinsically green fluorescent proteins that can bind endogenous FMN as a fluorophore [188]. Although considerably less bright than *Av*GFP [188] (see APPENDIX Table S 2), LOV proteins have few advantages over green *Av*GFP variants, namely a smaller size (ca. 10-15 kDa), improved pH stability and their independence towards oxygen [188,202]. LOV protein have been shown suitable for applications in both cytoplasm and periplasm of *E. coli*. The latter was particularly important to Study II since a FR was needed to fit the C-termini of the different MP candidates which were either cytoplasmic or periplasmic located. We finally chose to work with the LOV proteins EcFbFP, phiLOV2.1 and CreiLOV. The determining factors for the choice were brightness, predominantly

monomeric state or documented application in the periplasm. Besides, equally important were their similar excitation and emission wavelength peaks in comparison to sfGFP (see Figure 7 and APPENDIX Table S 2).

However, a LOV protein feature that, in retrospective, slipped our attentions were their thermal stability [198,200,204]. From a thermodynamic point of view, the thermostability of a protein is represented by the melting temperature (T_m) at which a protein is completely unfolded and inactive after heat denaturation ($\Delta G[T_m] = 0$; APPENDIX Figure S 3) [205,206]. The bigger the T_m of a protein, the more resistant is a protein's fold/structure towards heat. In contrast, from a functional point of view, a thermostable protein can also describe a protein that can unfold reversibly. More precisely, upon having passed a protein's T_m due to heat shock, a fully unfolded thermostable protein can resume refolding into a fully functional protein upon decreasing temperatures [24,207]. Importantly, EcFbFP and CreiLOV have been shown to be thermodynamically as well as functionally thermostable FRs [198,204]. Furthermore, a proteins T_m seems to be on average positively correlated with its free energy change of unfolding ($\Delta G[T]$) hence its intrinsic thermodynamic stability [205]. Thus, one should bear in mind that LOV proteins, against all initial assumptions, might belong to the group of thermodynamic stable proteins which might counteract with FtsH's poor unfoldase activity. To present, the free energy changes ($\Delta G[T]$) for the chosen LOV proteins have not been experimentally determined yet.

3.4.1 The *Av*GFP variant frGFP

frGFP (APPENDIX Table S 3) has been previously engineered as conventional C-terminal folding reporter for overproduction of normally aggregating proteins [4]. In regards to *Av*GFP, frGFP includes the following amino acid substitutions: F64L, S65T, Q80R, F99S, M153T and V163A.

Substitution F64L is increasing folding efficiency at 37°C [208,209]. Besides increasing fluorescence and photostability, amino acid substitution S65T causes a 6-fold higher amplitude of the excitation peak and a shift of the latter to 488 nm in comparison to *Av*GFP [208,210]. F64L and S65T are also known as *enhanced GFP mutations* [6]. Q80R is stated to be a neutral mutation possibly resulting from a PCR error [211]. Finally, F99S, M153T and V163A, known as the so-called *cycle 3 (C3) mutations* (originally reported as F100S, M154T, V164A), are believed to result in reduced aggregation and increased chromophore activation [212]. frGFP carries none of so-called *monomerising GFP mutations* (A206K, L221K, F223R) and is therefore prone to dimerise in solution at high concentrations [213].

frGFP had been chosen for pragmatic reasons: its sequence availability and good practical experience during other projects [214]. In the Studies I and II, frGFP was used as soluble fluorescent target protein for conditional ClpAP-dependent degradation.

3.4.2 The *Av*GFP variant sfGFP

Superfolder GFP (sfGFP; APPENDIX Table S 3) is an *Av*GFP variant and engineered for robust folding [6]. Compared to *Av*GFP, sfGFP's primary sequence carries the following mutations: S2R, S30R, Y39N, F64L, S65T, S72A, F99S, N105T, Y145F, M153T, V163A, I171V, and A206V.

S2R is thought as silent mutation which happened apparently accidentally by PCR during the development of GFPmut3*, an ancestral *Av*GFP variant of sfGFP [215]. S72A is a mutation on the chromophore-bearing helix [216]. Either alone or in combination with other mutations e.g. F64L, S72A gives greater brightness over *Av*GFP at warmer temperatures e.g. 37°C [163]. Furthermore, the same S72A/F64L combination can enhance GFP folding at 37°C [163,208,209]. S65T is increasing fluorescence and photostability, and causes a 6-fold higher amplitude of the excitation peak and a shift of the latter to 488 nm in comparison to *Av*GFP [208,210]. The *cycle 3 (C3)* mutations F99S, M153T and V163A (originally reported as F100S, M154T, V164A), are thought to reduce aggregation and to increase chromophore activation [212]. Finally, the so-called *superfolder GFP mutations* S30R, Y39N, N105T, Y145F, I171V and A206V have been proven to increase maturation speed, folding kinetics and protein stability in comparison to the parent frGFP variant [6,194]. Those sfGFP mutations enable sfGFP folding and maturation in the oxidising periplasm under aerobic conditions [194–196]. Furthermore, sfGFP is documented to be monomeric [6].

In Study II, sfGFP is fused to the C-termini of the inner membrane proteins PpiD, NavZ1 and YhaI (Figure 6b, c and d) and functions as FR located in cytoplasm or periplasm.

3.4.3 The LOV protein EcFbFP

EcFbFP (APPENDIX Table S 3) is an *E. coli* codon-optimised version of fluorescence reporter BsFbFP that is derived from the LOV domain of *B. subtillis* blue-light photoreceptor YtvA (UniProt ID O34627) [199]. EcFbFP encompasses almost unchanged the first 135 aa of YtvA besides the substitution C62A that substantially increases fluorescence intensity [199]. EcFbFP is a thermostable protein: it remains stably active until a temperature of 50 °C and shows also rapid as well as fully fluorescence recovery after a short thermal deflavination and denaturation step of up to 90 °C *in vitro* [204]. Additionally, EcFbFP is pH-stable over a broad pH range. Furthermore, EcFbFP is known to form stable homodimers in solution [199]. In comparison to other LOV proteins, EcFbFP is relatively bright and was even shown to outperform YFP during real-time measurements when overproduced alone [217]. Interestingly, EcFbFP has been used recently as N- or C-terminal translational fusion tag for the heterologous overproduction of soluble proteins in *E. coli* [218,219]. Using EcFbFP as N-terminal tag could even increase protein stability of its soluble fusion partner resulting from higher-order oligomerisation probably due to dimerisation of the FbFP tag [218].

EcFbFP has been published and patented as first LOV protein suitable for anaerobic and periplasmic applications (transport *via* SecYEG pathway) with relative high fluorescence intensities *in vivo* [199,220]. For Study II, EcFbFP was therefore used as C-terminal fusion tag for PpiD, an IMP with periplasmatic C-terminus (Figure 6b).

3.4.4 The LOV protein phiLOV2.1

phiLOV2.1 (APPENDIX Table S 3) is derived from the improved LOV protein (iLOV) [201]. iLOV, in turn, was engineered from *A. thaliana* blue-light receptor Phot2 (UniProt ID P93025) [201,202]. Accordingly, phiLOV2.1 consists of the $aa_{387-496}$ of *At*Phot2 including the following substitutions relative to the wild-type sequence: the iLOV substitutions C426A, S409G, F470L, S394T, I452T, M475V and six additional substitutions N390S, N401Y L422M, Q468R, E471G and D491V [201]. The latter six substitutions increased phiLOV2.1's photostability and photobleaching resistance compared to iLOV while brightness was decreased by half [201,221]. Importantly, phiLOV2.1 has been used as C-terminal FR-tag to a cytosolic globular protein in the obligate anaerobe *Clostridium difficile* in order to track its transport over the membrane [222]. Furthermore, it was C-terminally tagged to a soluble protein as well as a MP for translocations studies in *E. coli* and *Shigella flexneri* [223].

In Study II, phiLOV2.1 is fused to the C-terminus of NavZ1 and YhaI which are IMPs with their C-termini located in the cytoplasm (Figure 6c and d).

3.4.5 The LOV protein CreiLOV

CreiLOV (APPENDIX Table S 3) is a more recently discovered and engineered LOV protein. Besides being thermo- as well as photostable, fast-maturing and monomeric, CreiLOV even outperforms existing FbFPs in brightness and operational pH range (APPENDIX Table S 2) [198]. CreiLOV is the aa_{16-133}-fragment from *C. reinhardtii* blue-light photoreceptor Phot (UniProt ID Q8LPE0). Equally to EcFbFP, it carries a single amino acid substitution (C57A) in the FMN-binding pocket to improve fluorescence stability [198]. As demonstrated for EcFbFP, CreiLOV exhibits a high degree of functional thermal stability: under denaturising condition of 70 °C for 1 h, CreiLOV fluorescence rapidly recovered 60 % of its original fluorescence upon cooling to RT *in vitro* [198]. Due to its robustness and enhanced brightness among all actual LOV protein candidates *in vitro*, CreiLOV is suggested being one of the most promising LOV protein candidates for biotechnological applications [198,200]. However, an exemplary application *in vivo* remains to be proven.

Accordingly, we incorporated CreiLOV in the list of FRs in Study II and fused it to the cytoplasmic C-termini of the inner membrane proteins NavZ1 and YhaI (Figure 6c and d).

4 MOLECULAR BUILDING BLOCKS OF THE PRESENTED STUDIES

Promoters are designed to recruit the transcriptional machinery from the expression host and provide the platform for gene transcription initiation. The classical elements for bacterial promoters include the -35 and -10 site, as regulatory binding motives for the RNA polymerase [224]. Besides, various abundant regulatory elements are located within the promoter region and can co-regulate gene expression initiation, promoter strength, promoter tightness in the repressed state and translation initiation [225]. Promoters are therefore one of the centrepieces for recombinant gene expression from the genome as well as from a plasmid. The regulatory set-up and function of the promoters used in Studies I and II are described here in more detail.

4.1.1 The synthetic *T7* promoter

The bacteriophage T7 expression system was introduced for recombinant protein production in 1986 [226] and became soon part of many popular bacterial expression systems, including the pET system (Novagen) [227]. In the pET system, gene expression is driven from the bacteriophage *T7* promoter (*T7p*) which requires the abundance of bacteriophage T7 RNA polymerase (T7 RNAP) [227]. The *T7p* from the pET system is actually the renamed equivalent of T7*lac*, an engineered fusion of the bacteriophage T7 *Φ10* promoter and the *lac* operator site (*lacIo*) [228] (Figure 8). T7 RNAP is a very active and processive enzyme, due to its high selectivity to *T7p* [229]. It is usually produced from the genome of pET system host strains [227]. Its production in E.coli hosts is regulated by a *lac* promoter derivative, the *L8-UV5 lac* promoter (*lacUV5p*) [227] which itself is driven by the endogenous *E. coli* RNAP [230]. *lacUV5p* can be induced with the supply of lactose or its synthetic analogue isopropyl β-D-1-thiogalactopyranoside (IPTG) [231] which interact with the operator site-binding lac repressor (LacI) and thereby enable transcription [227].

In the wild type *lac* promoter, efficient transcription initiation requires the additional presence of the activated cyclic AMP receptor protein (CRP). CRP become activated upon binding cyclic AMP (cAMP). Complexed CRP/cAMP is called CAP complex (CAP) [227]. CAP can bind upstream from the -35 region and is able to promote transcription by endogenous *E. coli* RNAP directly [227]. Full induction of *lacUV5p* is only achieved by the presence of both, inducer and CAP [227]. *lacUV5p* has three point mutations in comparison to the wild type *lac* promoter: two point mutations in the –10 region increase promoter strength and decrease its dependence on CAP stimulation for full activation; the third-point mutation is located in the CAP binding site and decreases the affinity for CAP [227]. However, although the *lacUV5* promoter is reported to be well repressed in the absence of inducer, it still exhibit a detectable basal activity which lead a basal production of T7 RNAP [227]. This might be problematic when the gene of interest, driven by T7 RNAP-dependent *T7p* from a plasmid, actually encode for a toxic protein [227]. However, the basal gene expression level of *T7p* can be reduced by supplying small but constitutive amounts of T7 lysozyme [232], a natural inhibitor of T7 RNAP, using the plasmids pLysS and pLysE [233].

4.1.2 The synthetic *T5* promoter

The historical development of the *T5* promoter (*T5p*) had its origin as fusion between the strong T5 bacteriophage promoter P_{N25} [234] and the *E. coli lacIo* in the late 1980s; this construct was first termed *T5* P_{N25}/O [235]. P_{N25} is dependent on the *E. coli* RNA polymerase (RNAP) and exhibits a sequence with remarkably high translation initiation rates *in vivo* and *in vitro* in *E. coli* [234–237]. Upon fusion with *lacIo*, P_{N25} became negatively controlled by the endogenous LacI repressor which binds *lacIo* [235]. LacI repression is released upon supply with lactose or IPTG. In a following developmental step, another *lacIo* was introduced upstream of the first one, right between the -35 and -10 sites to ensure an increased repression level, leading to *T5* promoter variant *PN25/O3/O4* [238] (Figure 8). This promoter was finally picked, renamed *T5* promoter, and commercialised in the pQE vector collection from Qiagen [237]. Sufficient negative regulation of dual operator sequence can only be achieved in the presence of high levels of LacI [237]. Since the pQE vectors do not carry the lacI gene, excess LacI supply is suggested in two ways: (1) with an *E. coli* host strain that carries one of the *lacIq* genes ($lacI^q$ or $lacI^{q1}$) for increased LacI production; or (2) from a co-propagated medium copy plasmid (p15A replicon) from Qiagen that harbours the constitutively expressed *lacI* gene [237,239].

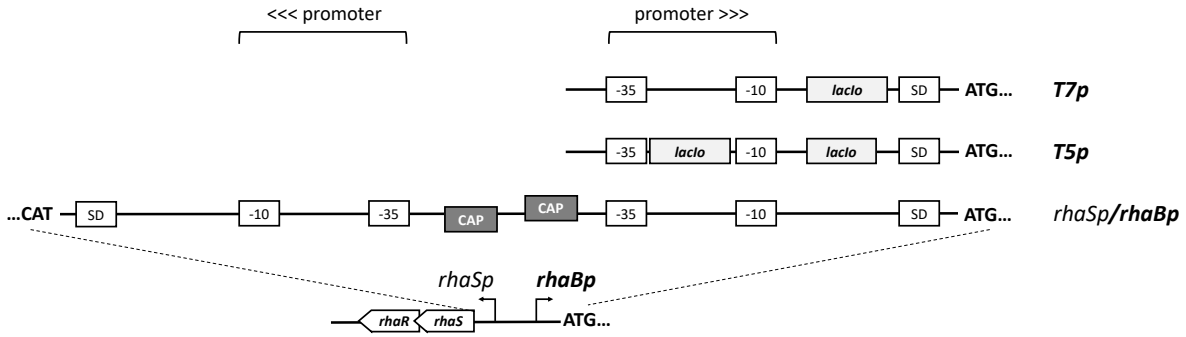


Figure 8: Structure of the promoter systems used for recombinant gene expression in *E. coli*. Promoter regions for *T7p* [228], *T5p* [238] and *rhaBp* [240,241] and some regulatory elements for transcription (-35 and -10 sites; *lacIo*; CAP) and translation (SD) are illustrated. Binding sites for sigma factors which can interact with RNAP (-35 and -10 sites) [242]; operator site for LacI binding (*lacIo*); binding site for CAP complex (CAP); sequence representing the Shine-Dalgarno sequence on mRNA level (SD).

4.1.3 The native *rhaB* promoter from *E. coli*

The *E. coli rhaB* promoter system (*rhaBp*) is originally derived from the *rhaBAD* operon. In nature, *rhaBAD* together with the *rhaT* gene and the *rhaSR* operon encode for all genes necessary for the rhamnose metabolism, the rhamnose-H^+ symporter and the regulatory proteins that control expression from *rhaBAD* and *rhaT*, respectively [243]. *rhaBAD*, *rhaSR* and *rhaT* have all individual promoters which are activated in the presence of L-rhamnose [243].

However, the *rhaBp* system for recombinant gene expression only encompasses the *rhaSR* operon with its promoter *rhaSp* and the adjacent *rhaBp* (originally the promoter for the *rhaBAD* operon) which acts in the

opposite direction (Figure 8). RhaR, an transcriptional activator, is always present at low levels [244] but becomes only activated upon binding of L-rhamnose and induces efficiently the expression of its own encoding *rhaSR* operon [240]. Accumulated RhaS, another transcriptional regulator, can then contribute to the activation of *rhaBp* which leads to the expression of the downstream genes in a L-rhamnose dependent manner [245]. However, gene expression from *rhaSp* as well as *rhaBp* is also subject to catabolite repression. In both promoters, there is a CAP binding site located upstream of the -35/-10 sites [240,241]. Thus, in addition to RhaS, CAP is needed for full activation of *rhaBp* [246]. The activation status of CAP is thereby positively regulated by cAMP, which in turn is upregulated at low glucose concentration or down-regulated at high glucose abundance [247].

Consequently, by being controlled by positive activation and catabolite repression, the *rhaB* promoter system is tightly regulated [248] and tuneable [249]. As for any other sugar-inducible promoter system, there are drawbacks by using the *rhaB* promoter system for recombinant gene expression. On one hand, monosaccharides other than L-rhamnose can activate *rhaBp* gene expression to a certain extend [243]. On the other hand, L-rhamnose concentration will decrease over time followed by decreasing gene expression and eventually its omission in a L-rhamnose-metabolising host [243]. An L-rhamnose catabolism deficient mutant would not only avoid decreasing gene expression, it would also enable gene expression in a truly concentration-dependent manner [249].

STUDY I – TOOL DEVELOPMENT FOR CONDITIONAL DEGRADATION OF SOLUBLE PROTEINS IN *E. COLI*

1 INTRODUCTION

Generating peptide or protein libraries and then screening those peptides or proteins for improved or modified characteristics has been proven to be an useful application in fundamental research as well as protein engineering [250]. While peptide libraries can be to date routinely generated by automated synthesis [251], the most straightforward method of constructing a library of protein variants is still to construct a library of protein encoding DNA sequences from which the protein library can be subsequently translated [250]. Many various methods have been developed until now for gene library construction with improved process efficiency and ever-increasing library size. Additionally, recent advances in screening methodologies allow screening of growing library sizes. A popular application for protein libraries has been the area of protein stability and protein-protein interaction [250]. In the late 1980s and early 1990s, the N-end-rule for protein degradation in eukaryotes [48] and prokaryotes [56] was deciphered by only using 15 and 19 cloned plasmid constructs, respectively, which only differed in the codon that encoded the outermost N-terminal residue of a POI. The nature of this N-terminal residue, once it became released, was shown to determine the proteins half-life in both domains [48,56]. Protein degradation was visualised by pulse-chase analysis (detection of radiolabelled proteins in SDS-PAGE) [48,56,252]. While the ClpAP proteasome, the key player for N-end rule degradation in bacteria, was soon discovered [56], its cognate adaptor protein ClpS was not confirmed until 2006 [3]. In order to systematically analyse the specificity of ClpS for proteic N-terminal residues, a library of 508 oligopeptide pools (13-mer) were generated and c-terminally attached to a cellulose membrane in a chessboard pattern [3]. The peptide-bound ClpS was thereby visualised by immunoblotting leading to mosaic-like pattern representing ClpS-binding efficiency [3,253]. A new area for randomising protein encoding DNA sequences emerged with the introduction of single-stranded oligonucleotide mutagenesis (ssDNA recombineering [254]). The novelty was, that an automated synthesis of randomised oligonucleotides allowed complete control over library size, the nature and the position of the randomisation [250]. Coupled with Flow-Seq, which combines Fluorescence Assisted Cell Sorting (FACS) and high-throughput sequencing [255], it became a powerful method to assess comprehensive data sets for increased understanding of physiological principles as demonstrated recently [256].

With the aim for an in-depth understanding of the N-end-rule in bacteria including the modulating effects of neighbouring amino acids on degradation efficiency, we integrated fr*gfp* on the *E. coli* K-12 MG1655 genome including a 5' in-frame insertion encoding for an N-terminal conditional degradation tag. The latter

consisted of an N-terminal degradation sequence motif for ClpAP-mediated degradation (dgClpAP) masked by the restriction site for the TEV protease (rsTEVP), similar to the published TIPI system [143]: rsTEVP(F)-dgClpAP-tag (described in INTRODUCTION section 3.2.3). We used ssDNA recombineering [254] to change the sequence coding for the first five aa of dgClpAP. This resulted in a protein library of theoretically 3,125 combinations. Upon TEVP-specific cleavage the randomised N-terminus of frGFP became conditionally accessible to ClpAP/ClpS resulting in a broad fluorescing library as visualised by FACS (unpublished data from Virginia Martínes, Tonja Hobel and Morten HH Nørholm). We finally pooled cells with the least fluorescence levels and analysed with Flow-Seq. The results only confirmed existing knowledge on N-end rule destabilising residues (unpublished data from Virginia Martínes, Tonja Hobel and Morten HH Nørholm).

However, motivated about the efficacy of the experimental set-up, we then developed PROTi (protein interference). Referring to the literature, PROTi is the second technology in bacteria for controlled N-terminal degradation of soluble proteins in a ClpAP-mediated fashion [18]. The idea was to engineer target proteins by fusing rsTEVP(F)-dgClpAP (PROTi-tag) to their N-terminus. The conditional element for PROTi is the inducible production of TEVP from a plasmid. Furthermore, the tool was upgraded in a second step by combining PROTi with CRISPRi (CRISPR interference [257]), this combined technology was named CRiPi. This combination makes it a unique tool to control protein levels by targeting both protein biosynthesis and degradation. Thus, in parallel to PROTi function, an RNA-guided dCas9 endonuclease could be targeted specifically towards the DNA sequence coding for the PROTi-tag. For CRiPi, the production of TEVP as well as dCas9 was independently inducible from a plasmid. The sCas9-cognate gRNA was expressed constitutively from another plasmid. The author of this thesis takes only credits for the development of PROTi.

2 PUBLICATION I

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

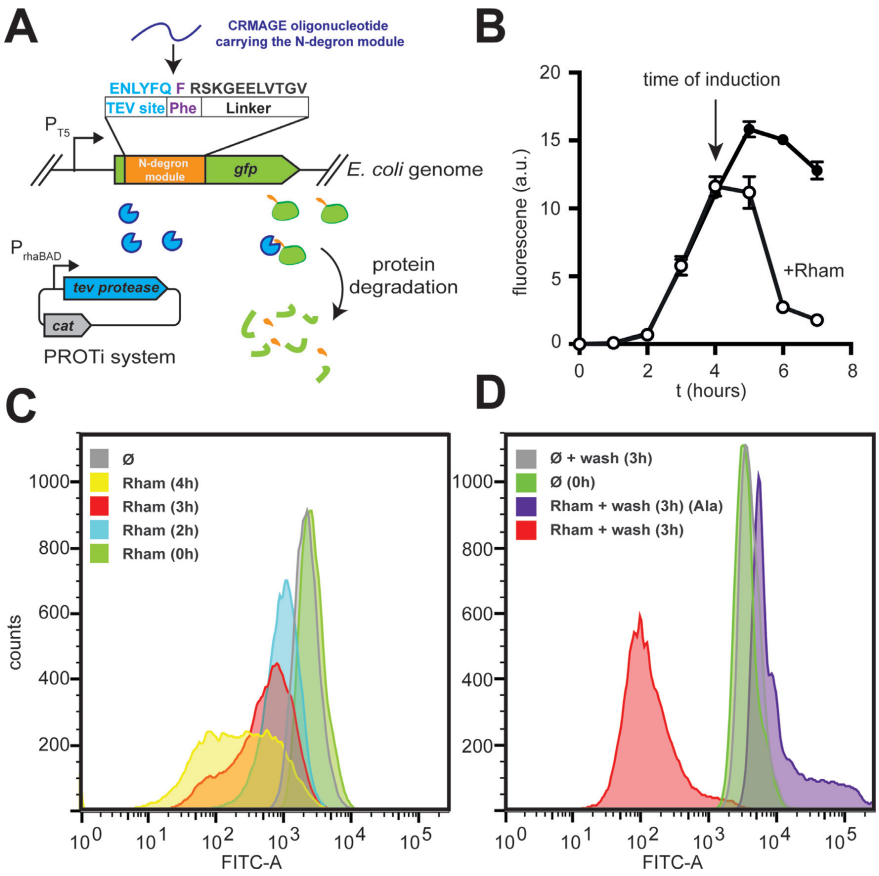


Figure 1. PROTi**:** Development and characterization of the PROTi system to control protein abundances. (**A**) Schematic illustration of the PROTi system. The coding sequence for the pro-N-degron module (orange) is integrated by CRMAGE at the 5′-end of a genomically integrated *gfp* (green). *gfp* 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

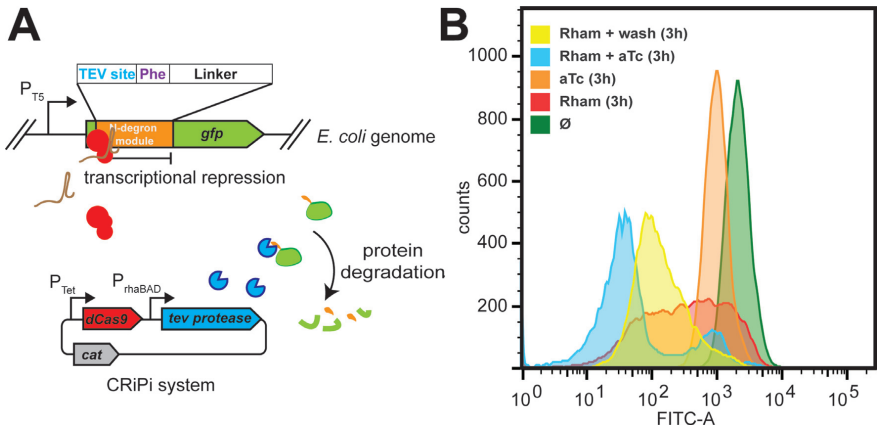


Figure 2. CRiPi**:** Schematic illustration of the CRISPRi-PROTi (CRiPi) system. (**A**) Cellular depletion of the targeted protein can be accelerated by simultaneous expression of *dCas9* (red) from the Ptet promoter and the TEV protease (blue) from the PrhaBAD promoter. Here shown for *gfp* as an example. 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-

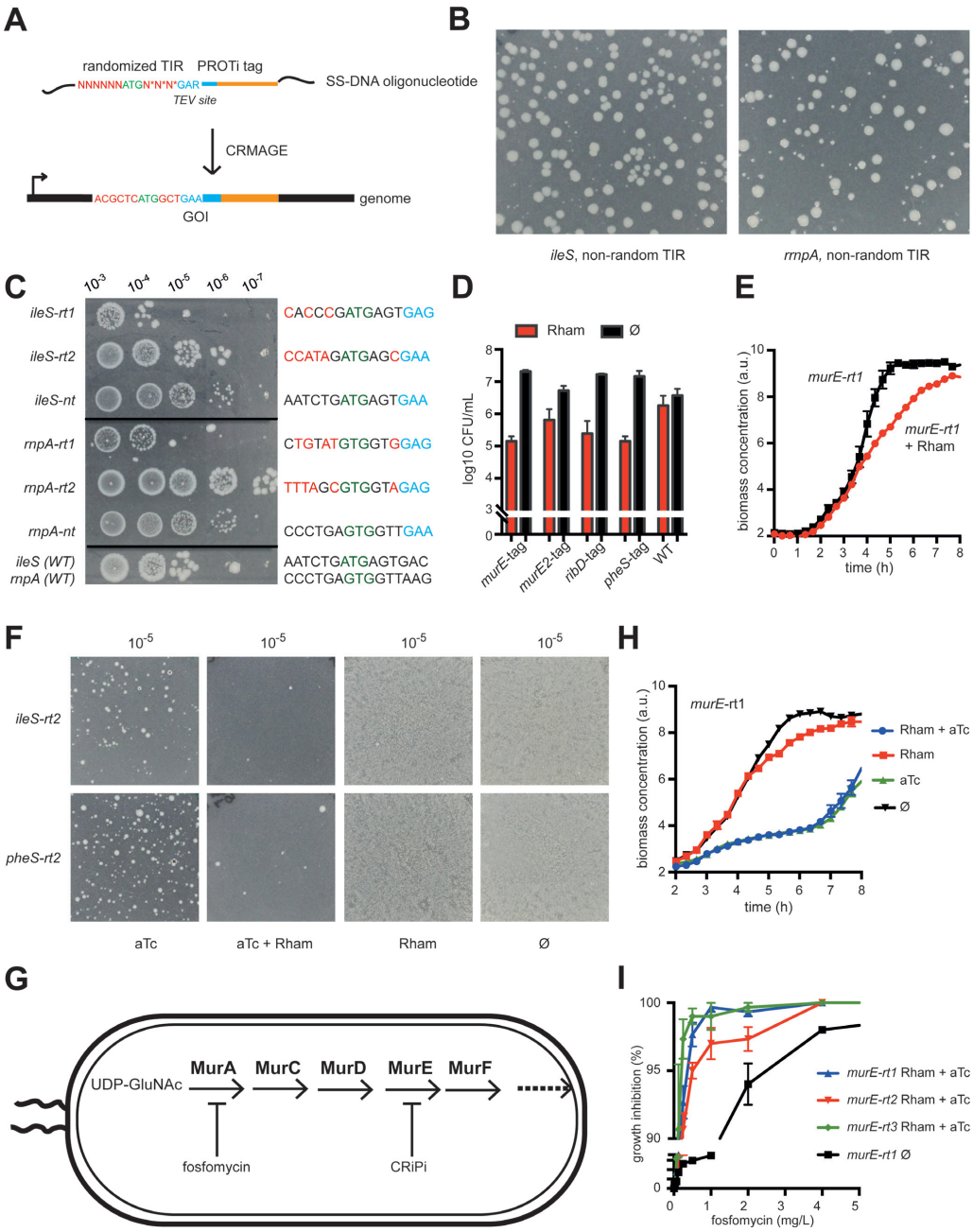


Figure 3. PROTi, CRiPi and TIR randomization can control protein levels and produce antibiotic hypersensitive bacteria. (**A**) Randomization of the translation initiation region (TIR, red font) in a ssDNA oligonucleotide enables CRMAGE-based insertion of the PROTi tag (orange) containing the TEV protease 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 material is to be found in APPENDIX section 3.

STUDY II – TOOL DEVELOPMENT FOR CONDITIONAL DEGRADATION OF MEMBRANE PROTEINS IN *E. COLI*

1 INTRODUCTION

In the early of Study I, ssDNA recombineering [254] and Flow-Seq [255] have been successfully implemented as a combinatorial methodology for a comprehensive understanding on N-end rule degradation criteria in bacteria. Even though no new insights could be gained, existing criteria of the bacterial N-end rule [52,56] based on the outermost N-terminal peptide sequence composition of ClpAP/ClpS substrate proteins could be validated in a comprehensive manner (unpublished data from Virginia Martínes, Tonja Hobel and Morten HH Nørholm). The discovery of the apparent MP-specific N-terminal degradation sequence motif of YfgM ($YfgM_{aa1\text{-}14}$ *alias* dgFtsH; INTRODUCTION section 3.2.2) by BITTNER *et al.* (2015) [2] seemed ideal to apply thereon the combinatorial ssDNA recombineering/Flow-Seq as a proof-of-concept study. The aim was to shed more light on the physiological principles of IMP proteolysis, in general, by deciphering sequence-specific physiological principles of FtsH-mediated IMP proteolysis, taken as first main objective upon starting this study. The second objective of this study was the development of a tool for conditional FtsH-dependent degradation of IMPs by using dgFtsH as transferable N-terminal degradation sequence motif.
However, it was crucial to test first whether and to what extend the PROTi technology from Study I was applicable to FtsH-mediated membrane protein degradation, and second, if a screening methodology with FR proteins was suitable in this context. Unfortunately, due to the accumulation of recurring technical obstacles, only marginal aspects of the project's objectives could be touched upon. Therefore, the following section will focus primarily on exposing the first attempts to clarify the methodological basis of the project. On a side note, our data demonstrated most surprisingly that L-rhamnose, as inducer for *rhaBp* was able to de-repress – to a certain extent – LacI repression of LacI-based promoter systems. This peculiarity has been seemingly neglected by the scientific community until now, supporting that all caution should be exercised when combining both L-rhamnose- and LacI-regulated promoters in a single cell.

2 MATERIAL AND METHODS

2.1 Strains, media and growth conditions

All *E. coli* strains used in this study are listed in Table 5. *E. coli* NEB 5-α and NEB Turbo (both from New England Biolabs (NEB), Ipswich, MA, USA) were used for cloning and propagation of the plasmids. The K-12 strains ssp. MG1655 and its λDE3 lysogen, MG1655 (DE3), were used unmodified or engineered for expression experiments (Table 5). Chemical competent cells of all strains were obtained as described elsewhere [258]. Electro-competent cells were prepared with sequential pelleting and resuspension of the cells with ice-cold MQ water. The strains were generally cultivated in lysogeny broth (LB) [259,260] with shaking (250 rpm) or on LB agar plates at 30 °C or 37 °C depending on the application. LB and LB agar were supplemented with antibiotics when required. If not stated otherwise, MG1655 (DE3) strains were cultivated on LB plates supplemented with 0.5 % (w/v) glucose when transformed with plasmids harbouring LacI-dependent promoters. Similarly, the LB media for such strains also contained 0.5 % (w/v) glucose during pre-culturing (overnight) prior to expression conditions. LB media and agar plates contained antibiotics if required.

Table 5: Overview about strains and genotypes used in Study II.

STRAIN NAME	GENOTYPE	COMMENT	REFERENCE
NEB 5-α (NEB)	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15*	Cloning strain. DH5α™ (Invitrogen) derivative. K-12 strain.	NEB
NEB Turbo (NEB)	F' *proA⁺B⁺ lacIᵠ ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)Tet^S endA1 thi-1 Δ(hsdS-mcrB)5*	Cloning strain. K-12 strain.	NEB
MG1655	F⁻ *λ⁻ ilvG⁻ rfb-50 rph-1*	Expression strain. Wild-type K-12 strain which was sequenced.	[181]
MG1655 (DE3)	F⁻ *λ⁻ ilvG⁻ rfb-50 rph-1 Ω[lacUV5p/T7 gene 1]*	Expression strain with genome-integrated λDE3 cassette.	
MG1655Ω[*T5p*/rsTEVP(A)-dg#ClpAP-frGFP]	F⁻ *λ⁻ ilvG⁻ rfb-50 rph-1 tRNIleY::[T5p/rsTEVP(A)-dg#ClpAP-frgfp]*	Expression strain with genomic integration of the coding sequence of a frGFP fusion variant (stable frGFP).	Study I
MG1655Ω[*T5p*/rsTEVP(F)-dgClpAP-frGFP]	F⁻ *λ⁻ ilvG⁻ rfb-50 rph-1 tRNIleY::[T5p/rsTEVP(F)-dgClpAP-frgfp]*	Expression strain with genomic integration of the coding sequence of a frGFP fusion variant (unstable frGFP).	Study I

If applied, antibiotics were used in the following concentrations: kanamycin (50 μg/mL) and chloramphenicol (50 μg/mL). All pre-cultures (overnight) were inoculated from fresh colonies and cultivated shaking at 37 °C. The volumes of the precultures depended on the following application and ranged 0.5 to 10 mL. Generally, expression cultures were inoculated with a calculated OD_{600nm} of 0.01 which

corresponded approximately to a dilution of 1:100 of the preculture. Furthermore, all expression experiments were performed by using either re-usable polypropylene square 24-deep well (DW) microplates or polypropylene square 96-DW microplates and their re-usable cognate sandwich covers (EnzyScreen, MB Heemstede, Netherlands).

2.2 Plasmids and plasmid construction

Plasmid construction was performed essentially using PCR-based uracil excision (USER) cloning for multi-fragment assembly [261]. Uracil-excision based site-directed mutagenesis (U-SM) [7] was applied for site-specific integration of short sequences into plasmid DNA. All oligonucleotides (APPENDIX Table S 1) for PCR were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA). The PCR fragments for USER cloning were generated with the Phusion U Hot Start polymerase (Thermo Fisher Scientific, Waltham, MA, USA). All plasmids were isolated using QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) and the PCR fragments were purified using NucleoSpin® Gel and PCR clean-up Kit (Macherey and Nagel, Düren, Germany). Each constructs' identity and sequence integrity upon cloning were checked by sequencing of the full sequence between both the 5' and 3' ends.

An overview about the plasmids used in this study is given in Table 6. A detailed description of the constructed plasmids will follow below.

Table 6: Cloning and expression plasmids used in Study II. The protein which can be produced from each plasmid is shown in bold letters within the brackets (excluding N-terminal or C-terminal attachments). STOP codon (*); gBlock endings (◄/►).

PLASMID	ABBREVIATION	RELEVANT CHARACTERISTICS	COMMENT	REFERENCE
pSEVA47Ω[*T5p*/rsTEVP(*)-frGFP]		pSC101, Sm/SpR	template	Camilla Krogsgaard
pMA-TΩga[*T7p*/His-rsTEVP(S)-sfGFP]	pMSB0881	ColE1, AmpR	template	Sara Petersen Bjørn
◄rsTEVP(M)-dgFtsH-CreiLOV►	pMSB1226	gBlock (IDT)	template	Study II
◄rsTEVP(M)-dgFtsH-phiLOV►	pMSB1227	gBlock (IDT)	template	Study II
◄reTEVP(M)-torA-EcFbFP►	pMSB1228	gBlock (IDT)	template	Study II
pEnvZΩga[lacZp/NavZ1-V5]		ColE1, AmpR, *lacI*	template	Roger Draheim
pSEVA33Ω[*rhaBp*/**TEVP**]	pPROTi	pBBR1, CmR		Study I
pSEVA33Ω[*rhaBp*/**TEVP**-His]	pPROTi -His	pBBR1, CmR		Study II
pSEVA33Ω[*T7p*/MCS]		pBBR1, CmR	control	Daniel C. Volke
pCOLADuet™-1	pCOLA	ColE1, KanR, *lacI*	control	Novagen
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**sfGFP**-His]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**CreiLOV**-His]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**phiLOV2.1**-His]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**EcFbFP**-His]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**navZ1-CreiLOV**-His]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**navZ1-sfGFP**-His]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**yhaI-phiLOV2.1**-His]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**yhaI-CreiLOV**-His]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**yhaI-sfGFP**-his]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**ppiD-EcFbFP**-His]		ColE1, KanR, *lacI*		Study II
pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-**ppiD-sfGFP**-His]		ColE1, KanR, *lacI*		Study II

Construction of the pPROTi plasmids

Two TEVP encoding plasmids were used. The pPROTi plasmid, described in Study I (5622bp) consisted of the high-copy pSEVA33 backbone (pBBR1 ori, CmR). It carried the TEVP-S219D-encoding *tev* gene under the control of the L-rhamnose-inducible promoter *rhaBp*. A second plasmid was constructed as follows: a sequence coding for a [$(GGS)_3$GG-$(His)_8$]-tag was inserted as gene-fusion at the 3´-end in frame with *tev* in pPROTi via whole plasmid synthesis and U-SM [7] by using the primer pair oMSB3268/oMSB3269, resulting in the pPROTi-His plasmid.

Construction of the pCOLAΩ[T7p/rsTEVP(M)-dgFtsH-MP-FR-His] plasmids

The cloning strategy was re-adjusted multiple times. The initial cloning of the MP-FR fusions (including a N-terminal FtsH-specific degradation tag and a C-terminal $(His)_8$ epitope) in the low-copy pSEVA47 (pSC101 replicon, Sm/SpR) combined with the IPTG-inducible *T5p* failed systematically even in an *E. coli* NEB Turbo background (Table 5), due to apparent leakiness of the *T5p* and possible protein toxicity. Eventually cloning was successful in NEB 5-α cells with a pCOLA backbone (ColA replicon, KanR, *lacI*) artificially stripped of its promoter (Table 5). We created here a small library of pCOLAΩ[-/MP-FR-His] plasmids using PpiD, NavZ1 or YhaI (as MP target proteins) and sfGFP, CreiLOV, phiLOV2.1 or EcFbFP (as FR-tags), respectively. This was done in multiple steps where we re-used primers and PCR fragments from previous cloning attempts.

(1) Creating a pCOLA template that incorporates the N-terminal rsTEVP of the degradation tag and the C-terminal His-epitope

The pCOLA backbone, harbouring the replicon and the resistance cassette, and a small fragment encoding for rsTEVP ([rsTEVP(*)-frGFP]) were amplified from pCOLADuet^{TM}-1 (Novagen) and pSEVA47Ω[*T5p*/rsTEVP(*)-frGFP] (Camilla Krogsgaard, unpublished) with the primer pairs oMSB3151/oMSB3152 and oMSB3257/oMSB3258, respectively. The two PCR fragments were assembled by USER cloning, resulting in pCOLAΩ[-/rsTEVP(*)-frGFP-His], referred to as template1.

(2) Integration of the encoding sequences for [dgFtsH-FR] modules

The DNA sequences coding for dgFtsH-EcFbFP, dgFtsH-phiLOV2.1 and dgFtsH-CreiLOV were ordered as gBlock® gene fragments from IDT (Coralville, IA, USA). The nucleotide sequence for EcFbFP has been published recently [220] and those for phiLOV2.1 and CreiLOV were codon-optimized for *E. coli* from their respective peptide sequence [198,262] using JCat [263]. The respective gBlocks® are listed as pMSB1228, pMSB1227 and pMSB1226, respectively, in Table 6. Besides, pMSB0881 (Sara Petersen Bjørn, unpublished) was used as template for the amplification of s*fgfp*. The coding sequences for phiLOV2.1, CreiLOV, EcFbFP and sfGFP were amplified from pMSB1227, pMSB1226, pMSB1228 and pMSB0881 with the primer pairs oMSB3290/oMSB2951, oMSB3289/oMSB2951, oMSB2953/oMSB3291 and oMSB3292/oMSB3293, respectively. The PCR fragments were USER-assembled with a pCOLA backbone PCR-fragment which resulted from an amplification of template1 with oMSB2233/oMSB3288. The [(*)-frGFP]-region from

template1 was thereby substituted with the different [dgFtsH-FR] fragments. This generated four new templates for subsequent integration of the MP-coding sequences: pCOLAΩ[-/rsTEVP(M)-dgFtsH-phiLOV2.1-His] (template3), pCOLAΩ[-/rsTEVP(M)-dgFtsH-CreiLOV-His] (template4), pCOLAΩ[-/rsTEVP(M)-dgFtsH-EcFbFP-His] (template5) and pCOLAΩ[-/rsTEVP(M)-dgFtsH-sfGFP-His] (template6).

(3) Integration of the membrane protein coding genes

For the integration of the MP encoding genes, template3, template4, template5 and template6 were then amplified by PCR with the primer pairs oMSB3045/oMSB3054, oMSB3045/oMSB3053, oMSB3045/oMSB3055 and oMSB3045/oMSB3332 to create the pCOLA backbone (bb) fragment3, 4, 5 and 6, respectively. The *navZ1* gene was amplified from pEnvZΩ[*lacZp*/NavZ1-V5] [5] with either oMSB3061/oMSB3063, oMSB3061/oMSB3062 or oMSB3061/oMSB3334 which created the compatible USER-overhangs for an USER assembly with the pCOLA bb fragment3, 4, and 6, respectively. Furthermore, the *yhaI* gene was amplified from genomic DNA of *E. coli* MG1655 (Table 5) with either oMSB3058/oMSB3060, oMSB3058/oMSB3059 and oMSB3058/oMSB3335 which created in turn USER-fragments compatible with pCOLA bb fragment3, 4, and 6, respectively. Finally, genomic DNA of *E. coli* MG1655 served also as template for the *ppiD* gene. With oMSB3056/oMSB3057 and oMSB3056/oMSB3333 we amplified PCR fragments thereof which were in turn USER-compatible with pCOLA bb fragment5 and 6, respectively. The result was an intermediate plasmid library of eight pCOLAΩ[-/rsTEVP(M)-dgFtsH-MP-FR-His] variants without promoter.

(4) Integration of T7p

In a final step, the *T7p* was inserted in each resulting construct of (2) and (4) *via* U-SM [7] with the generic primer pair oMSB3336/oMSB3337. NEB 5-α (Table 5) cells were used for plasmid propagation. Eleven out of twelve constructs showed a correct assembly after sequencing (Table 6). pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-NavZ1-phiLOV2.1-His] could not be obtained. Due to time pressure, the experimental part was initiated in absence of the 12th construct. One should note that with *T7p*, the Shine Dalgarno sequence and the 5' coding region for [rsTEVP(M)-dgFtsH] of all genes, all constructs had the same TIR.

2.3 Methods and technical details

If not indicated otherwise all chemicals were ordered from Sigma Aldrich (St. Louis, MO, USA).

Measurement of cell biomass and fluorescence

Cell biomass (optical density, OD, 600nm) and fluorescence data (λ_{Ex} = 485nm, λ_{Em} = 528nm; gain 100) were collected as end-point measurements from 200 μL aliquots of culture broth in optical 96-well microplates (flat-bottom; Greiner Bio-One, Kremsmünster, Austria) in a MX plate reader (Biotek, Winooski, VT, USA)

without path length correction. If needed, cultures were diluted with peptone saline solution [264] consisting of 0.1 % casein peptone and 0.85 % (w/v) NaCl.

Replacing the growth medium of bacterial cultures

24-DW plates with bacterial cultures were spun at 4818 x g and 4 °C for minimum of 20 min in a swing-rotor. The plates were then quickly inverted to remove the supernatant; the plate edges were cleaned with 70 % (v/v) ethanol to prevent cross-contamination (plates up-side-down). Subsequently, the bacterial pellets were resuspended on ice with the appropriate medium for re-cultivation. The volume of the new medium corresponded to the medium volume prior to spinning. Plates were kept on ice during all intermediate steps, e.g. between cultivation, centrifugation, medium replacement.

Plaque assay with top agar for phage detection

An overnight culture was prepared from colony of a non-lysogenic *E. coli* K-12 MG1655 (DE3) strain (host strain) exhibiting the same plasmid-based resistances as cultures suspected for phage lysis (CamR, KanR). The overnight culture was diluted subsequently 1:30 with warm 0.5% (w/v) LB top agar supplemented with kanamycin and chloramphenicol and poured onto a pre-warmed LB agar plate without antibiotics. A volume of 5 μL of supernatants of suspected phage-infected cultures, such as the respective supernatants of the pre-cultures, but also media and supplements used during the experiment were spotted onto the cooled and dried top-agar surface. The plate was incubated at 30 °C overnight mimicking the growth conditions of the problematic experiments.

2.4 Experimental details

Besides relevant information for the experimental performance, only details will be noted that were differed from the general growth conditions and procedures described above.

Controlling basal expression from T7p in E. coli (DE3) strains

MG1655 (DE3) was transformed either with pCOLA or with pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-sfGFP-his]. Additionally, empty NEB 5-α and MG1655 (DE3) were run as control strains. One single clone ($r_B = 1$) of each strain was used for the experiment. Overnight cultures were supplemented with 1 % D-glucose (w/v) as recommended elsewhere [227]. The expression experiment was performed in a volume of 1 mL LB medium cultivated in 96-DW plates at 37 °C and 250 rpm. The LB medium was supplemented with kanamycin (for plasmid stability if needed) and different concentrations of D-glucose ranging from 0 to 4 % (w/v). Eight technical replicates incubated in parallel for each strain and condition ($r_T = 8$). Cell biomass and fluorescence development was measured after 20 h.

The functionality of the TEVP variants and finding a sugar candidate to complement growth advantages due to L-rhamnose supply

MG1655Ω[*T5p*/rsTEVP(A)-dg$^{\#}$ClpAP-frGFP] (stable frGFP) and MG1655Ω[*T5p*/rsTEVP(F)-dgClpAP-frGFP] (unstable GFP) were transformed either with pPROTi or pPROTi-His. Three clones of each strain without technical replicates (r_B = 3, r_T = 1) were used for the experiment. The expression experiment was performed in a volume of 4 mL LB cultivated in 24-DW plates at 30 °C and 250 rpm. The LB medium was supplemented with chloramphenicol and MQ water or 0.1 mM IPTG. The IPTG induced the production of stable or unstable frGFP right from the beginning of the experiment. After 6h of growth the growth medium was replaced (procedure described in Study II 2.3) with LB medium containing either 5 mM L-rhamnose, 5 mM D-galactose or 5 mM L-fucose, resulting in six growth conditions. Cell biomass and fluorescence were continuously measured over 20 h.

Fluorescence intensities of the selected LOV proteins versus sfGFP

pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-FR-his] was co-transformed with pPROTi into MG1655 (DE3). FR in pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-FR-his] corresponded to the coding sequences for either sfGFP, EcFbFP, CreiLOV or phiLOV. Thus, the transformation resulted in four different strains. Three clones of each strain without technical replicates (r_B = 3, r_T = 1) were used for the experiment. The expression experiment was performed in a volume of 4 mL LB cultivated in 24-DW plates at 30 °C and 250 rpm. The LB medium was supplemented with kanamycin and chloramphenicol as well as and MQ water or 0.1 mM IPTG. The IPTG induced the production of the N- and C-terminally modified FRs right from the beginning. After 6h of growth the growth medium was replaced (procedure described in Study II 2.3) with LB medium containing either 5 mM L-rhamnose, or 5 mM L-fucose, resulting in four growth conditions. Cell biomass and fluorescence were continuously measured over 20 h.

Optimising membrane protein production with different inducer concentrations and inducing membrane protein's susceptibility towards FtsH-specific proteolysis

All pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-MP-FR-His] plasmid constructs were individually co-transformed with pPROTi-His (or pPROTi, depending on the experiment) into *E. coli* K-12 MG1655 (DE3). MP-FR in pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-MP-FR-His] corresponded to the coding sequences for either ppiD-EcFbFP, ppiD-sfGFP, navZ1-CreiLOV, navZ1-sfGFP, yhaI-CreiLOV, yhaI-phiLOV2.1 or yhaI-sfGFP. Thus, this transformation resulted in seven different strains. In parallel, empty pCOLA was co-transformed with either pPROTi-His or empty pSEVA33Ω[*T7p*/MCS] into MG1655 (DE3), resulting in the control strains 1 and 2. Three clones of each strain without technical replicates (r_B = 3, r_T = 1) were used for the experiment. The expression experiment was performed in a volume of 4 mL LB cultivated in 24-DW plates at 30 °C and 250 rpm. The LB medium was supplemented with kanamycin and chloramphenicol.

In order to optimised the *T7* promoter strength for IMP-FR production, IPTG was supplemented in a concentration range in between 0 to 1.6 mM after 4h of growth. IPTG induced the production of the N- and C-terminally modified IMP-FRs. Cell biomass and fluorescence were continuously measured over 20 h.

For testing the susceptibility of produced IMP-FR constructs towards FtsH-specific degradation, cell cultures were grown for 4 h when IMP-FR production was induced with 0.1 mM IPTG (or MQ water). After further 6 h the growth, LB/±IPTG medium was replaced (procedure described in Study II 2.3) with LB medium containing either 5 mM L-rhamnose, or 5 mM L-fucose, resulting in four growth conditions. Cell biomass and fluorescence were continuously measured over 20 h.

3 RESULTS AND DISCUSSION

3.1 Plasmid cloning for the production of membrane proteins N-terminal fused to a conditional degradation tag

The genetic set-up for Study II was initially designed alike the one in Study I. However, due to increased complexity we opted for a plasmid-based system which provided more flexibility for genetic re-engineering. Low copy plasmids are generally preferred for the overproduction of recombinant membrane proteins (MPs) [265] since this strategy would avoid the saturation of co-translational integration of nascent MP polypeptides into the membrane by the *E. coli* Sec translocon [266]. Accordingly, we chose to work with the pSEVA47 backbone (pSC101 replicon, Sm/SpR) from the pSEVA collection [267] harbouring the pSC101 replicon which ensures an average plasmid copy number of 3-5 per chromosome [268,269] - this was first thought to be an acceptable substitute to a genome-based gene expression system.
As used for Study I, we kept the synthetic IPTG-inducible *T5p* (INTRODUCTION section 4.1.2) for the expression of the engineered genes encoding for the PpiD, NavZ1 or YhaI (INTRODUCTION section 3.3) – all N-terminally tagged with the FtsH-specific conditional degradation tag (rsTEVP(M)-dgFtsH-; INTRODUCTION section 3.2.3) and C-terminally fused to a variety of FR-polyhistidine (His)-tags (INTRODUCTION section 3.4); the resulting gene product can be sketched as: rsTEVP(M)-dgFtsH-MP-FR-His.

However, a mutation-free integration of the MP-encoding genes into a pSEVA47Ω[*T5p*/FR-His] backbone proved to be difficult even in an *E. coli* NEB Turbo (NEB) background [270]. NEB Turbo (NEB) harbours the endogenous *lacI* gene driven by a mutated *lacI* promoter (*lacI*^{q} gene) which is responsible for higher transcription rates of *lacI* and therefore increased levels of the LacI repressor within cells (Table 5). The cloning result was a repeated overflow of missense, nonsense and frameshift mutations in the respective coding regions after *T5p*, as documented by gene sequencing (data not shown) [h]. Two possible conclusions were drawn from these results: (1) the increased endogenous LacI levels from NEB Turbo were apparently

[h] This phenomenon was also observed when integrating the same MP coding genes into a similar pSEVA47Ω[*T5p*/S11-his, J23106/gfp$_{opt1-10}$] backbone but not when the genes encoded for soluble proteins [270].

insufficient to repress *T5p* expression *in-trans*, which contrasts with the recommendations from Qiagen regarding plasmid-based *T5p* usage [237]; and (2), the leaky *T5p* expression lead to the basal production of protein products toxic to the host. Conclusion (1) could be easily addressed, whereas conclusion (2) was interesting in the sense that PpiD, YhaI and NavZ1 had been previously successfully overproduced in *E. coli* even from high copy plasmids (see details in INTRODUCTION sections 3.3.2, 3.3.4 and 3.3.3, respectively); in those studies gene expression was driven by promoters dependent either of the bacteriophage T7 RNAP (*T7p*) or of the endogenous *E. coli* RNAP (*tetp*, l*acZp*). The suspicion of protein toxicity raised the additional question regarding which impact the N-terminal FtsH-specific degradation tag or its coding sequence might have on transcription, translation or MP integration into the IM. Generally, bacteria encounter metabolic burden with genetic plasmid reorganisation which might either lead to loss of function of the protein [271] or plasmid instability, allowing the cells to survive [272]. However, the basis of this phenomenon seems to be still poorly understood [273].

Nevertheless, in order to address conclusion (1), we switched the plasmid backbone to pCOLADuet™-1 (Novagen; ColA replicon, KanR), termed pCOLA for simplification, and pursued all intermediate cloning steps in an *E. coli* NEB 5-α (NEB) background (Table 5). pCOLA directly harbours the wildtype *E. coli lacI* gene which constitutively provides an elevated constant pool of LacI additionally to the endogenous one from the NEB 5-α genome. Nevertheless, mutations still persisted (data not shown) [270]. This was again surprising since pCOLA has been designed for dual gene expression with two separated multi cloning sites (MCS) starting each with a *T7p*; every *T7p* contains one *lacIo* for LacI regulation. *T5p*, with its two *lacIo* sites, provides an equal amount of LacI binding sites. Thus, one might think the additional LacI pool derived from pCOLA should also be sufficient for *T5p* repression.
We finally carried out a multiple sequence alignment of all *lacIo* sites from the original *T7* promoters on pCOLA and of those from *T5p* (APPENDIX Figure S 4). According to SASMORE AND BETZ (1990), the optimal binding site for LacI is a symmetric DNA sequence consisting of two 10 bp operator halves (highlighted in grey in APPENDIX Figure S 4) [274]. Interestingly, this optimal LacI binding site can be found as a whole in both *T7p*-located *lacIo* sites on pCOLA but only in the first *lacIo* in *T5p*. The second *lacIo* in *T5p* consists of a shorter palindromic sequence (underlined in APPENDIX Figure S 4) which shortens the optimal binding site for LacI. Thus, it might be probable that in order to get a similar LacI repression, as for the two *T7p* in pCOLA, *T5p* would need a larger LacI pool to complement for the reduced LacI binding affinity to the second *lacIo*.

We resumed cloning with the pCOLA backbone but without promoter and progressed rapidly, bypassing thereby all considerations regarding promoter repression. As a final step, we inserted the coding region for *T7p* in all of the pCOLA constructs to restore them as pET-like vectors. *T7p* has the advantage that in the complete absence of T7 RNAP (INTRODUCTION section 4.1.1) leaky expression is naturally prevented. Eventually, all cloning steps were carried out successfully in a NEB 5-α background (all obtained pCOLA constructs are listed in Table 6), keeping in mind that mutations might re-occur after transformation into

an expression strain that carries the gene for the T7 RNAP (DE3 strain), such as *E. coli* K-12 MG1655 (DE3). Strict precautions were taken at least during the overnight growth of pre-cultures. It is however impossible to rule out that such mutations did not occur during the expression studies.

3.2 Controlling basal expression from promoter *T7p* in *E. coli* (DE3) strains

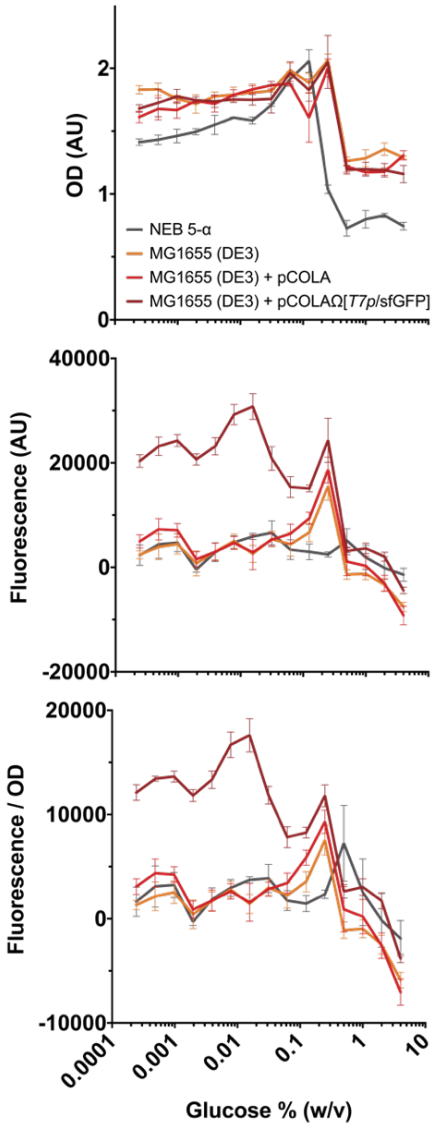


Figure 9: **Different glucose concentrations regulating basal *T7p* gene expression from pCOLA and medium acidification which affect growth of *E. coli* DE3 strains grown to stationary phase**. Cell biomass (OD) and fluorescence end point measurements of *T7p* expression cultures were performed after 20h of growth. Different *E. coli* strains were grown in LB medium supplemented with kanamycin (except empty K-12 MG1655 [DE3] and NEB 5-α) and different glucose concentrations ranging from 0 to 4 % (w/v). Empty K-12 MG1655 (DE3) and NEB 5-α were included as negative controls. $r_B = 1$, $r_T = 8$, $n = 1$.

E. coli K-12 MG1655 (DE3) is an engineered λDE3 lysogen (DE3 strain) which harbours the gene encoding for the bacteriophage T7 RNAP (bacteriophage T7 *gene 1* [275]) in its genome. As stated earlier (INTRODUCTION

section 4.1.1), *gene 1* transcription is under the control of *lacUV5p* which requires an inducer (lactose or IPTG) and the CRP/cAMP complex (CAP) for full induction [227]. However, although being well repressed in the absence of inducer, T7 RNAP can be produced from *lacUV5p* in small amounts [227]. Consequently, basal levels of T7 RNAP can induce *T7p*-driven genes prior to the actual induction. Interestingly, cAMP levels, and therefore also CAP levels, are negatively affected by the presence of glucose [227,247]. In other words, when the glucose level gets low, as it is for DE3 strains grown to stationary phase in complex medium, increased basal *lacUV5p* and consequently also *T7p* expression can be prevented by supplementing glucose [227,276]. Due to additional glucose supply, an initial rapid growth phase can be monitored [227]. Afterwards however, glucose is responsible for a rapid acidification of the extracellular medium leading to inhibited cell growth and death [227,277]. The medium acidification is a result of increased efflux of organic acids from the cells, specifically lactic and acetic acids [277], which are breakdown products of the glucose metabolism [227,277]. There does not seem to be a coherent opinion in literature about the recommended glucose supply that prevents basal T7 RNAP production from *lacUV5p* and yet minimises the acidification of the medium. This might be related to both, different levels of leakiness depending on the respective promoters [278] and strains exhibiting different acid tolerances [279].

In an effort to reduce the risk of mutations prior to induction as experienced during cloning (STUDY II section 3.1), we tested systematically the impact of different glucose concentrations on overnight cell growth and *T7p*-dependent basal sfGFP production from pCOLA in *E. coli* K-12 MG1655 (DE3) grown in LB medium. Accordingly, *E. coli* K-12 MG1655 (DE3) was transformed either with pCOLAΩ[*T7p*/rsTEV-dgFtsH-sfGFP-his] or empty pCOLA plasmid. The obtained strains were grown overnight in LB medium with different concentrations of glucose ranging from 0 to 4 % (w/v), whereby 1 % (w/v) glucose is equivalent to 55 mM. Empty K-12 MG1655 (DE3) and NEB 5-α[i] were grown in parallel as negative controls for basal GFP-production and for the purpose of comparability to another K-strain, respectively. The end point measurements for cell biomass and fluorescence after 20 h of growth are summarised in Figure 9.

Up to 0.25% or 0.125% (w/v) glucose could be supplied to all K-12 MG1655 (DE3) strains or NEB 5-α, respectively, before cell growth was dramatically inhibited due to medium acidification (Figure 9a). This indicated that K-12 MG1655 (DE3) generally exhibits a higher acid tolerance than NEB 5-α. Basal sfGFP-production can be seen by comparing the K-12 MG1655 (DE3) strains harbouring either pCOLAΩ[*T7p*/rsTEV-dgFtsH-sfGFP-his] or empty pCOLA plasmid (Figure 9b and c). A minimum of 0.5% (w/v) glucose was needed to decrease sfGFP-derived fluorescence levels produced from uninduced K-12 MG1655 (DE3) with pCOLAΩ[*T7p*/rsTEV-dgFtsH-sfGFP-his] even down to the auto-fluorescence levels seen for K-12 MG1655 (DE3) with empty pCOLA or empty K-12 MG1655 (DE3). We don't have any plausible explanation for the mountain silhouette-shaped sfGFP fluorescence ranging from 0 to 0.5% (w/v) glucose, nor for the

[i] *E. coli* NEB 5-α (Novagen) is a DH5-α derivative which is known to be a K-strain [307].

fluorescence peaks at circa 0.25% (w/v) and 0.5% (w/v) glucose for all K-12 MG1655 (DE3) strains and NEB 5-α, respectively. The latter might be the result of an impaired take up and metabolising of fluorescent components from the medium which is reflected in higher auto-fluorescence levels at those conditions. The sfGFP-fluorescence landscape, however, might be due to effects of a superior gene expression regulation. Nevertheless, the results show that glucose is necessary to repress promoter leakiness but too much is harmful for the cells due to medium acidification which leads to impaired cell growth and death. In the favour for a condition with upmost *lacUV5p* tightness, we determined the glucose concentration for K-12 MG1655 (DE3) overnight cultures in LB medium to 0.5% (w/v) which corresponds to 28 mM.

3.3 Functionality of the TEVP variants and finding a sugar candidate to complement growth advantages due to L-rhamnose supply

The mandatory requirement to cleave N-terminal conditional degradation tags in our MP constructs to induce FtsH-mediated proteolysis was the presence of active TEPV. For that purpose, we re-used the TEVP producing pPROTi plasmid (pSEVA33Ω[*rhaBp*/tev]) from Study I. Additionally, we cloned a pPROTi variant in which we simply inserted the coding sequence of an $(His)_8$-tag at the 3' end of the *tev* gene and termed it pPROTi-His (pSEVA33Ω[*rhaBp*/tev-His]. To ensure that TEVP activity was unaltered regardless of the C-terminal $(His)_8$-tag, we transformed pPROTi and pPROTi-His into the two *E. coli* K-12 strains MG1655Ω[*T5p*/rsTEVP(F)-dgClpAP-frGFP] and MG1655Ω[*T5p*/rsTEVP(A)-dg$^{\#}$ClpAP-frGFP] which were constructed for pPROTi characterisation in Study I (Figure 10). Both strains harboured the coding sequence for frGFP variants with N-terminal conditional degradation tags for ClpAP proteolysis. The difference between both was the N-terminal residue which would be released upon TEVP cleavage. A released Phe (F) has a destabilising effect on the whole frGFP construct since it directly triggers ClpAP degradation whereas Ala (A) acts as a stabiliser for frGFP (referred to as *unstable frGFP* and *stable frGFP*, respectively). We induced the production of unstable and stable frGFP with IPTG (or MQ water) at the starting point of the experiment. Within the exponential phase we removed the LB/IPTG medium and replaced it with fresh LB supplied with L-rhamnose for the induction of TEVP. Plain *E. coli* K-12 MG1655 (referred to as *no frGFP*) was equally transformed with pPROTi or pPROTi-His and served as control.

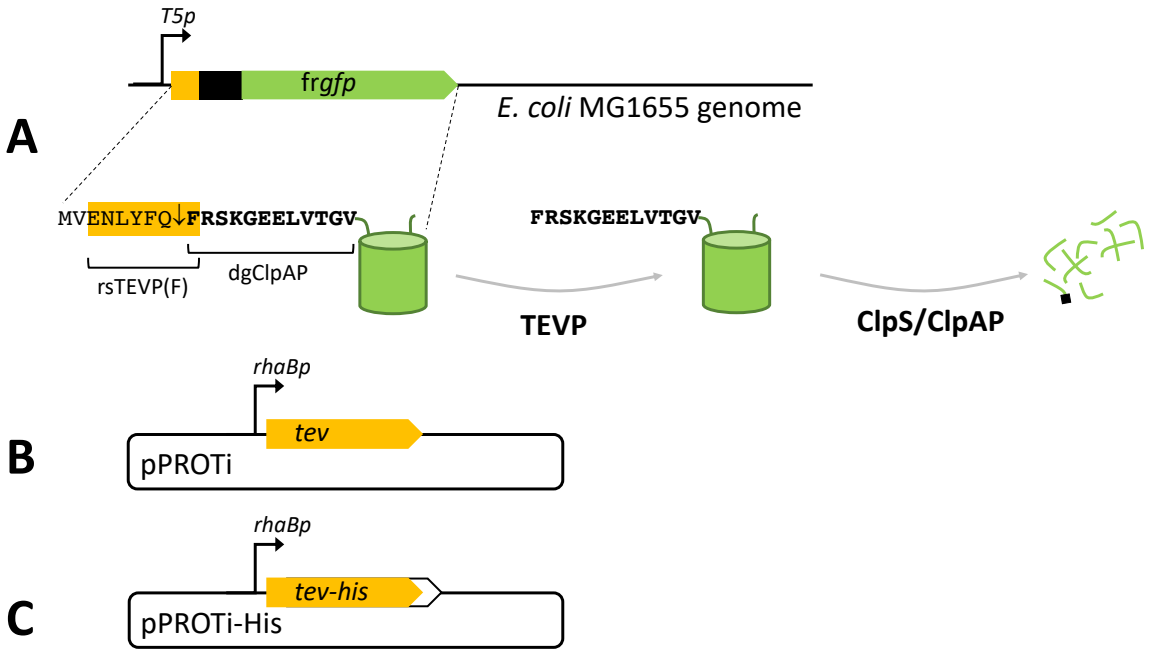


Figure 10: **Experimental set-up for testing the functionality of TEVP and TEVP-His**. (A) The coding sequence of frGFP, N-terminally tagged with the ClpAP-specific conditional degradation tag, was integrated into the genome of E. coli MG1655. Electively, the strain was co-transformed with either (B) pPROTi or (C) pPROTi-His plasmid for TEVP or TEVP-his production, respectively. The example illustrates the ClpS/ClpAP-mediated degradation of the unstable frGFP variant upon TEVP(-His) cleavage (stable frGFP not shown). Details in the main text.

In previous experiments we experienced growth advantages of *E. coli* strains grown in LB medium supplemented with L-rhamnose compared with strains that grew under the same conditions but without the sugar supplement (data not shown). To correct artificial growth divergences, we explored the substitution of L-rhamnose with a sugar candidate with the following characteristics: (1) a hexose monosaccharide (C6 sugar) and *E. coli* metabolite (such as L-rhamnose) whose (2) transport is independent from a phosphotransferase system (PTS) which might be coupled to endogenous cAMP levels and therefore regulate *lacUV5p* -driven T7 RNAP production *via* the CRP/cAMP (CAP) complex in *E. coli* DE3 strains. This would become important for later experiments in DE3 strains. Finally, (3) the sugar candidate had to be a non-inducer for the promoter *rhaBp*. We chose to test D-galactose and L-fucose (Figure 11): after being taken up by their specific transporters, L-fucose (alike L-rhamnose) is metabolised in the cytosol in two enzymatic steps before it enters the glycolysis as two di-hydroxy-acetone-phosphate moieties (C3 sugar) [280,281]; in contrast, D-galactose undergoes five to six enzymatic modification reactions after uptake, before it enters the second step of glycolysis as glucose-6-phosphate (C6 sugar) [282]; both sugars have been reported as non-PTS carbohydrates (like L-rhamnose) [247] and as non-inducers for *rhaBp* (unlike L-rhamnose) [243]. Cell biomass and fluorescence were monitored over 20 h. The results are illustrated in Figure 12. The complexity of the results was unexpected.

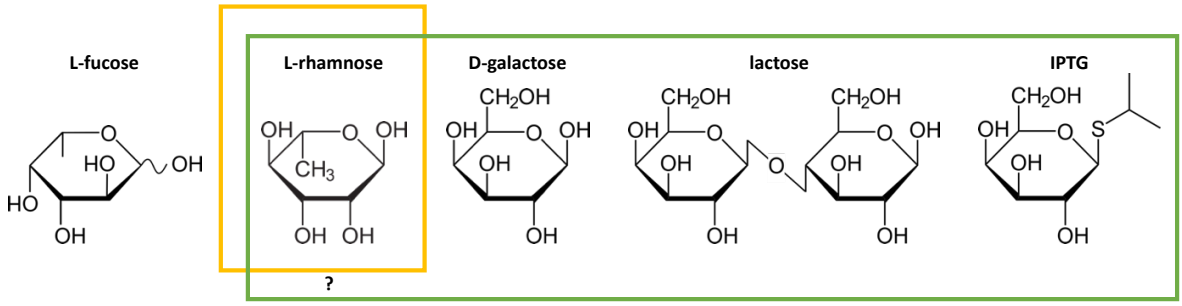


Figure 11: **Chemical structures of synthetic IPTG and *E. coli* sugar metabolites.** Molecules are classified as *rhaBp*-inducing (orange frame) and as effector for LacI (green frame). L-fucose is neither of the two. L-rhamnose as putative effector for LacI (marked with a question mark, ?) is discussed in the main text.

Both TEVP variants are equally active

First of all, the growth curves for all strains and conditions were almost identical. Thus, growth was generally not impaired independently of the nature and variety of the proteins produced. Additionally, *E. coli* K-12 MG1655 harbouring pPROTi or pPROTi-His (controls) showed under all conditions almost constant auto-fluorescence levels, as expected. The activity of TEVP was not impaired by the C-terminally His-tag since all data sets from cultures with pPROTi or pPROTi-His were identical. Furthermore, both TEVP variants indicated highest cleavage efficiency: at time point 10 h (4 h after L-rhamnose supply), fluorescence/OD levels for the unstable frGFP dropped significantly and were comparable to auto-fluorescence levels (no frGFP) whereas stable frGFP signals were highest. However, fluorescence/OD levels of stable and unstable frGFP deviated substantially after supplying the different sugars, seemingly as an effect of the sugars themselves. This phenomenon was evaluated furthermore.

D-galactose and L-rhamnose are effector molecules for LacI

Even though unexpected, there is a significant difference between the fluorescence/OD levels of IPTG-negative conditions for stable frGFP in comparison to the plain auto-fluorescence/OD levels for no frGFP. The supply with D-galactose and L-rhamnose seemed to increase basal production of stable frGFP until time point 10 h (4 h after sugar supply) whereby the effect triggered by D-galactose appeared more enhanced. In contrast, the fluorescence/OD levels of the IPTG-negative/L-fucose condition remained stable within the same time frame and beyond.

It is already known since the early 1950s, that besides allolactose and IPTG, there are many more effector molecules that can alter the affinity of LacI for its DNA binding site [283–285]. Those molecules can mimic the effect of the conventional lactose/IPTG induction of the endogenous lac operon [285] and accordingly of heterologous LacI-regulated expression systems. D-galactose is listed as weak effector molecule for LacI

[283,284]. While IPTG, as most efficient effector, can decrease LacI affinity for *lacIo* 1000-fold [286], D-galactose[j] only shows 1/1000 of the efficiency of IPTG [287]. Other galactosides, tested as functional inducers, showed efficiencies in the same range [286,287] which suggested a common mechanism of interaction with LacI [285]. It seems that a hydroxyl-group on C6 of the sugar ring plays an essential role for effector/LacI interaction; however, it is not sufficient alone to induce enough conformational change in LacI to reduce its affinity to *lacIo* [285]. To accomplish the latter, it appears that the substituent on the C1 position also needs to provide optimal effector/LacI binding [285]. Fucosides, like L-fucose, do not possess the hydroxyl group in C6 and are therefore considered as non-inducers [285] which would underline our findings. In contrast to D-galactose and L-fucose, we do not have any explanation for the enhancing effect of L-rhamnose on basal production of stable frGFP: (1) upon screening the literature, L-rhamnose has neither been associated nor tested as effector for LacI; and (2) it does not possess the C6 hydroxyl-group. Nevertheless, it remains possible that L-rhamnose can interact structurally with LacI in a different way than reported until now.

A similar pattern of differentiated fluorescence/OD levels for the stable frGFP can be seen for the IPTG-positive conditions after replacement of the LB/IPTG media with the LB/sugar media: cultures supplemented with D-galactose and L-rhamnose showed significantly increased levels in fluorescence/OD until timepoint 10 h (4h after sugar supply) compared to those of L-fucose. It is regrettable, that we missed to include the condition IPTG-positive/sugar-negative in the experiment. This way, we would have achieved maximal production levels for stable frGFP only due to IPTG. Yet, we did not expect increased production rates from *T5p* due to D-galactose or L-rhamnose acting as effector molecules for LacI. Having said so, maximal production levels for stable frGFP from IPTG alone are obtained in the IPTG-positive/L-fucose-positive condition since L-fucose is considered being a LacI-non-effector [285].
After timepoint 10 h, the effective levels for stable frGFP in all conditions decreased slowly until the end of the experiment. This might be due to two – potentially synergetic – root causes: (1) the supplied sugar levels were metabolised by the host, which tightened LacI/*lacIo* interaction and therefore stabilised the protein levels of stable frGFP; (2) the cells are entering the stationary phase but continued growing slowly until the end of the experiment, thus, the stabilised levels for stable frGFP were just slowly diluted over time.

Leaky expression from genomic T5p in the absence of an increased LacI pool

As stated earlier, the presence of L-fucose did not affect the basal production of stable frGFP. However, the baseline of the fluorescence/OD level of stable frGFP (IPTG-negative/ L-fucose) was still higher as the auto-

[j] Enhanced basal expression from pET expression systems was observed in host strains with impaired cytosolic D-galactose metabolism like *E. coli* BL21 strains (*galK*$^-$) grown in complex media such as LB [233]. LB contains among others enzymatic digest of milk protein casein [233]. Since milk is rich in lactose, variable amounts of residual lactose may be present in different lots of enzymatic digests [233]. Thus, when D-galactose is well transported into the cell (in the absence of glucose) and accumulating in the cytoplasm, it can contribute to an increased basal expression from LacI-dependent promoters [233]. This effect is usually redundant in D-galactose-metabolising hosts [233].

fluorescence/OD levels measured under any conditions for no GFP. This might be interpreted as leaky production of stable frGFP from pT5. Thus, the endogenous LacI pool appeared not enough for sufficient repression of a single genome-integrated *T5p*, as assumed from Qiagen (personal communication). Consequently, even pT5 as single copy per chromosome would need a bigger LacI pool than *E. coli* can provide by itself: either produced from a genomic *lacI*^{q} or *lacI*^{q1} gene or from a medium-copy plasmid harbouring the *lacI* gene [237]. This had not been tested yet from Qiagen (personal communication) and would need therefore experimental validation. This finding has no impact on later experiments in Study II but might be relevant for future applications with genomic *T5p* expression systems.

D-galactose and L-fucose are true non-inducer for rhaBp

The sugar-specificity of *rhaBp* can be indirectly observed by comparing the fluorescence/OD levels of stable frGFP with those of unstable frGFP under IPTG-positive/sugar-positive conditions. What can be seen directly is the ClpAP-dependent proteolysis of unstable frGFP once its N-terminal destabilizing degradation sequence motif became released upon TEVP cleavage. A discussed earlier, L-rhamnose supply led to a prompt production of active TEVP which finally led to the rapid and complete degradation of unstable frGFP in comparison to stable frGFP. Interestingly the effective protein levels for unstable frGFP and stable frGFP behaved exactly similarly upon supply with D-galactose or L-fucose which disqualified both sugars as inducers for *rhaBp*, like published recently [243].

D-galactose as sugar candidate to complement growth advantages due to L-rhamnose supply in E. coli strains with Rha^{+} phenotype

Independent of the nature of supplemented sugar all growth curves for all strains and conditions were almost identical. This makes both D-galactose and L-fucose suitable candidates to replace L-rhamnose and to complement growth advantaged on first glance. On second glance, by having uncovered L-fucose as non-effector for LacI as well as non-inducer for *rhaBp*, L-fucose would make the ideal sugar candidate we initially looked for. However, by considering (1) that D-galactose and L-rhamnose are apparently both effectors for LacI with similar efficiency and (2) that D-galactose is a non-inducer for *rhaBp*, then D-galactose would become best choice at third glance (Figure 11).

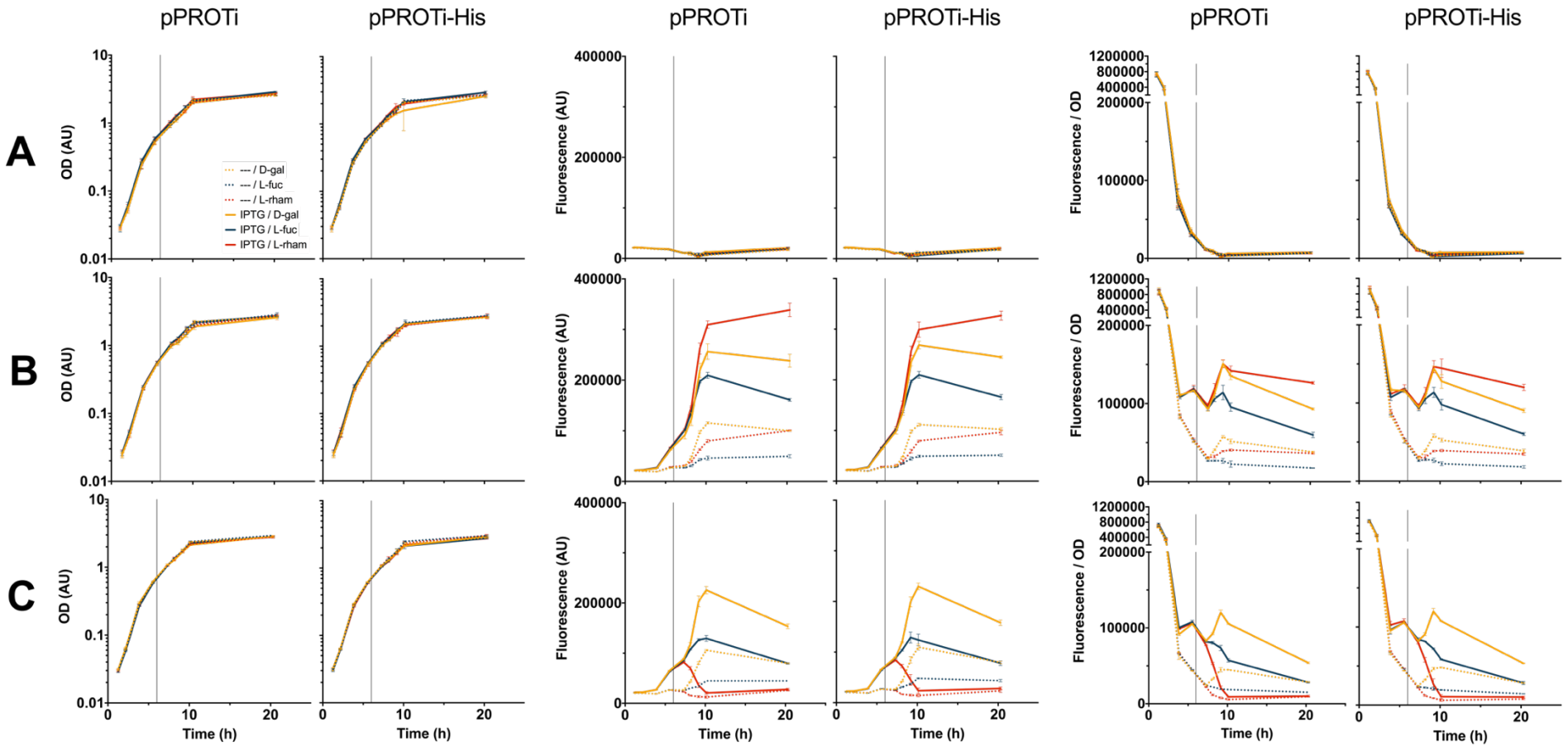


Figure 12: **Functionality of TEVP or TEVP-His on the stability of frGFP variants**. The general cytoplasmic fusion construct produced from the genome was rsTEVP(X)-dgClpAP-frGFP. X represented hereby the outermost N-terminal residue of the dgClpAP degradation sequence motif which would become released upon TEVP cleavage. A released Phe (F) has a destabilising effect on frGFP whereas Ala (A) is stabilising, referred to as unstable frGFP and stable frGFP, respectively. (A) unmodified *E. coli* K-12 MG155; (B) *E. coli* K-12 MG155 Ω[*T5p*/ rsTEVP(A)-dgClpAP-frGFP]; (C) *E. coli* K-12 MG155 Ω[*T5p*/ rsTEVP(F)-dgClpAP-frGFP]. All strains harboured either pPROTi or pPROTi-His. The line at 6 h represents the time point of media change from LB/±IPTG to LB/sugar in order to induce TEVP production. Details in the main text. $r_B = 3$, $r_T = 1$, n = 1

3.4 Fluorescence intensities of the selected LOV proteins versus sfGFP and their degradability as soluble proteins by the FtsH proteasome

In order to visualise MP degradation, we designed IMP fusions with different FR proteins. The idea was, that if FtsH-mediated IMP degradation follows the pulling model (INTRODUCTION section 1.4.2; Figure 5A) the C-terminal FR i (translocated), unfolded and degraded, similar to a globular IMP domain. The FR candidates were sfGFP, CreiLOV, phiLOV2.1 and EcFbFP. sfGFP, CreiLOV, phiLOV2.1 were used for applications in cytoplasm; sfGFP and EcFbFP in the periplasm.

The fluorophore in *Av*GFP variants such sfGFP is not a separate, synthesized prosthetic group but is composed of modified amino acid residues within the polypeptide chain [163,164]. Moreover, fluorophore formation in *Av*GFP variants relies on successful protein folding and the presence of oxygen. With the exception of sfGFP, *Av*GFP variants usually need the reducing environment of the cytoplasm to form disulphide bridges necessary for correct folding [194,195]. The amino acid substitutions in sfGFP, however, contribute to robust folding of the protein (without disulphide bridges) even in the highly oxidising periplasm of *E. coli*; the fluorophore is formed subsequently under aerobic conditions [6,194,195]. Therefore, we still expected a bright periplasmic sfGFP signal in an aerated (shaking) *E. coli* culture, as published previously [195].

In contrast, maturation of LOV proteins is independent of the presence of oxygen [288] but need the non-covalently binding of flavin mononucleotide (FMN) as chromophore [197,262]. FMN is an endogenous cytosolic metabolite in *E. coli* [204], shown to be sufficient for cytosolic LOV protein applications *in vivo* [199]. The use of complex media, including LB medium, which contain naturally certain amounts of FMN, [289] has been proven to be sufficient as FMN source for the production LOV protein fusions in the periplasm [290]. Similarly to sfGFP, we expected cytoplasm-located CreiLOV and phiLOV2.1 and periplasm-located EcFbFP to fluoresce upon cell cultivation in LB medium, as reported [199,220].

Due to its apparent weak unfoldase activity, FtsH was shown unable to degrade soluble GFP-SsrA or GFP-cI108 substrates whereas other soluble protein substrate exhibiting lower intrinsic thermodynamic stability and C-terminally fused with the same degradation tags were degraded [103] (INTRODUCTION section 1.4.1). The latter is an issue that we particularly kept in mind for the IMP-sfGFP fusions and considered those as negative controls. In contrast, the three LOV protein variants were initially thought to have a lower intrinsic thermodynamic stability than most *Av*GFP variants. We expected therefore the LOV proteins to be fully degraded granted FtsH-mediated degradation followed the pulling model (INTRODUCTION section 1.4.2; Figure 5A). If not, the IMP-FR fusions were otherwise expected to be partially degraded, whereas sfGFP or the LOV proteins alone would remain as fluorescing entities in the periplasm or cytoplasm. Immunoblotting would be the method of choice to observe these different phenomena.

According to the literature, the relative fluorescence brightness of the LOV proteins EcFbFP and CreiLOV are approximately in the same range as for the enhanced blue (eBFP) and cyan (eCFP) fluorescent variants of *Av*GFP (APPENDIX Table S 2). The brightness of the two *Av*GFP variants correspond to 1/7 or 1/5 of the relative fluorescence brightness measured for sfGFP, respectively (APPENDIX Table S 2), although one should note that all data provided in APPENDIX Table S 2 were obtained at the respective optimal excitation and emission wave lengths of those proteins. However, in order to standardize and compare data generated during different experiments and in different equipment (e.g. fluorescence plate reader, FACS), we performed all fluorescence measurements with the excitation and emission at 485 nm and 528 nm, respectively. These wavelengths are rather optimal for sfGFP but suboptimal for LOV protein detection (see individual fluorescence characteristics in APPENDIX Table S 2). Thus, it was essential to compare the fluorescence intensities of all selected FRs under the given conditions. We chose to test the latter with constructs producing only the soluble FR entities in the cytosol of *E. coli*.

Furthermore, the specificity of dgFtsH (INTRODUCTION section 3.2.2) had been only proven for a single case; soluble dgFtsH-GST was not degraded by FtsH [2]. Thus, as an additional feature of this experiment, soluble sfGFP, CreiLOV, phiLOV2.1 and EcFbFP were each fused N-terminally to the FtsH-specific conditional degradation tag (rsTEVP(M)-dgFtsH-). Upon induced TEVP cleavage we would enable putative FtsH-mediated degradation of the four soluble target proteins.
Accordingly, we induced the production of all four rsTEVP(M)-dgFtsH-FR constructs at the starting point of the incubation with IPTG or used MQ water as negative control. Within the exponential phase we removed the LB/IPTG medium and replaced it with fresh LB supplied either with L-rhamnose for the induction of TEVP, or with L-fucose[k] as negative control for TEVP production and balance of L-rhamnose-related growth advantages in the control cultures (discussed in STUDY II section 3.3). Cell biomass and fluorescence were monitored over 20 h. The results are summarised in Figure 13.

Soluble LOV proteins CreiLOV and EcFbFP affect cell growth

Even though all cell cultures grew, at some point, to the same cell density in the stationary phase, the growth rates of the exponential phase upon production of the soluble CreiLOV and EcFbFP variants were significantly reduced and reached the stationary phase only after approximately 20 h. In contrast, cell growth was not affected by the production of phiLOV2.1 and sfGFP variants nor by the conditional abundance of TEVP (in comparison to the control conditions); those cultures reached the stationary phase after approximately 12 h. Nevertheless, we do not have an explanation for the negative effect on cell growth due to the production of soluble CreiLOV and EcFbFP variants.

[k] We initially mis-interpreted the data in 3.3. D-galactose would have been the better sugar candidate to complement for L-rhamnose growth advantages in *E. coli* strains with Rha$^+$ phenotype (see discussion in 3.3).

Fluorescence brightness of EcFbFP highest among LOV proteins

The fluorescence intensities (fluorescence/OD levels) among all soluble LOV proteins were in the same low range but significantly elevated in comparison to auto-fluorescence levels. The fluorescence intensities for EcFbFP were highest among the selected LOV proteins which agrees with the literature [202,204]. In contrast, fluorescence intensities of soluble sfGFP outperformed significantly those of the LOV proteins. Most importantly, the fluorescence data was obtained from cytoplasmic FRs.

No fluorescence intensities were measured from FRs located in the periplasm although sfGFP and EcFbFP would locate into the periplasm when fused to the C-terminus of PpiD (INTRODUCTION section 3.3.2). Given the robust folding ability of sfGFP in the periplasm [6] and the aerated culture condition, as well as the evidence of EcFbFP fluorescence in the periplasm [220] combined with the presence of FMN from the complex medium we still expected robust fluorescence signals for both reporter proteins.

CreiLOV, phiLOV2.1, EcFbFP and sfGFP are no FtsH substrates under the given condition

None of the rsTEVP(M)-dgFtsH-FR constructs exhibited drops in relative fluorescence intensities (fluorescence/OD levels) upon media change from LB/IPTG to LB/L-rhamnose or LB/L-fucose (negative control). Thus, it would appear that none of the rsTEVP(M)-dgFtsH-FR constructs became susceptible towards FtsH-specific degradation upon TEVP production. This result can be interpreted in three different ways: (1) dgFtsH is essential but not sufficient alone for FtsH-mediated degradation, and needs membrane-anchoring to become a complete degradation motif, as suggested previously (MP-specificity of dgFtsH) [2]; (2) the intrinsic thermodynamic stability of all FRs are too high for the weak unfoldase activity of FtsH [103] which was expected for sfGFP [103] and, with retrospect, also assumed for the LOV proteins (see discussion in INTRODUCTION section 3.4); and/or (3), due to the medium change, necessary for TEVP induction from the pPROTi plasmid, the cells were artificially re-shifted into a second or prolonged exponential phase which is seemingly unfavourable for FtsH-mediated degradation of dgFtsH-tagged target proteins [2]. While conclusion (1) would reflect a unique and interesting characteristic of dgFtsH from a synthetic biologist point of view, (2) and (3) would actually be methodological flaws in our experimental set-up.

L-rhamnose as effector for LacI is confirmed

The relative fluorescence intensities (fluorescence/OD levels) of sfGFP in the IPTG-negative conditions showed an enhanced basal level after the replacement of the initial growth medium to LB/L-rhamnose in comparison to LB/L-fucose. This confirmed earlier observations, that L-rhamnose (but not L-fucose) might act as an effector for LacI and de-repress LacI-regulated promoters (STUDY II section 3.3). This effect was particularly enhanced for the relatively bright sfGFP.

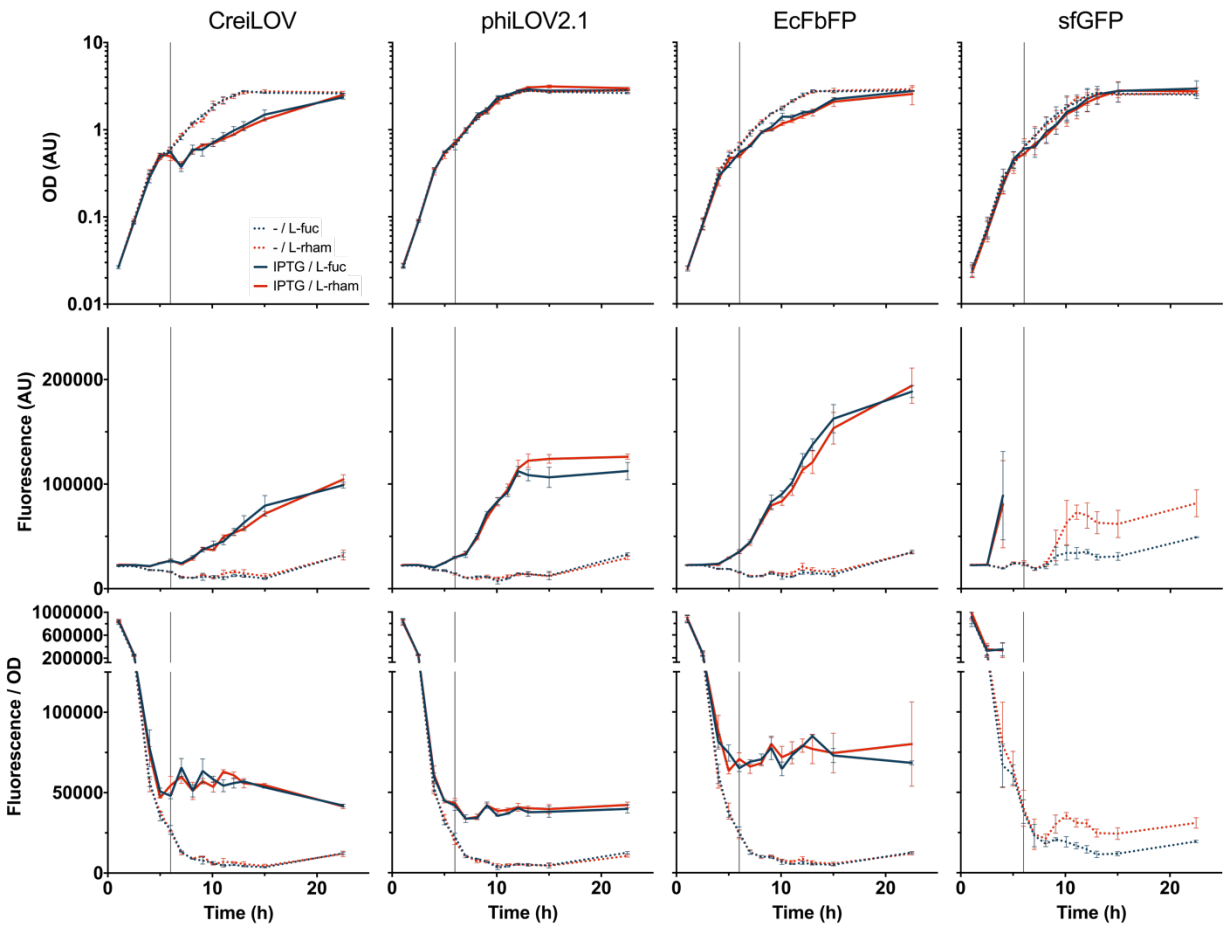


Figure 13: **Comparison of the fluorescence intensities for selected LOV proteins versus sfGFP and their susceptibility towards FtsH-specific degradation**. The general cytoplasmic fusion construct produced from a pCOLA backbone and *T7p* was rsTEVP(M)-dgFtsH-FR-His in *E. coli* K-12 MG1655 (DE3). FR was thereby either CreiLOV, phiLOV2.1, EcFbFP or sfGFP. All strains contained also the pPROTi-His plasmid for the induction of TEVP production. The line at 6 h represents the time point for media change from LB/±IPTG to LB/sugar. The growth conditions are indicated in colours. The development of cell biomass and fluorescence were followed over 22h. Details in the text. $r_B = 3$, $r_T = 1$, n = 1.

3.5 Optimising membrane protein production with different inducer concentrations

The biogenesis of integral membrane proteins (MPs) from their nascent polypeptide chains involves many steps starting with the targeting of nascent MP to the membrane, the insertion and folding therein and finally, in case of polytopic membrane proteins, the assembly [291,292]. Each of these steps requires distinct components and interactors whose activities are inter-dependently balanced to maintain the dynamic integrity of the whole process [179,291,292]. Due to an oversaturation of the aforementioned steps, the overproduction of MPs targeted in *E. coli* for its inner membrane (IM) often results in insoluble protein aggregates (inclusion bodies) in the cytosol [179,293]. Whether an overproduced MP ends up in inclusion bodies or successfully becomes integrated in the IM is not possible to predict [293]. This depends on several

parameters such as the expression strain, induction temperature, growth medium and promoter strength [179].

GFP fused to the C-terminus of MPs has been successfully introduced as folding indicator to distinguish between IM-insertion versus inclusion body formation [187] where the proper folding of the GFP depended on the correct folding of its MP fusion partner [179]. However, it is emphasised that C-terminal GFP is only an indicator and does not guarantee for successful IMP-integration. Moreover, inclusion bodies of MP-GFP fusions may stay fluorescent and the GFP moiety may also be clipped off by proteases *in vivo* [187]. Immunoblot protocols that can differentiate between integrated and aggregated MP have been developed to eliminate doubts [179,187].

Provided that we could use the fluorescence signal of CreiLOV, phiLOV2.1, EcFbFP and sfGFP as an indication for folding/integration of our IMP-FR constructs, we tested different IPTG concentrations ranging from 0 to 1.6 mM IPTG, to regulate their production from the promoter *T7p* located on the pCOLA plasmid. After 4 h of growth, IMP-FR production was finally induced in the exponential phase. We integrated two controls to evaluate the levels for auto-fluorescence. Cell growth and fluorescence were monitored over 24 h. The results are summarised in Figure 14.

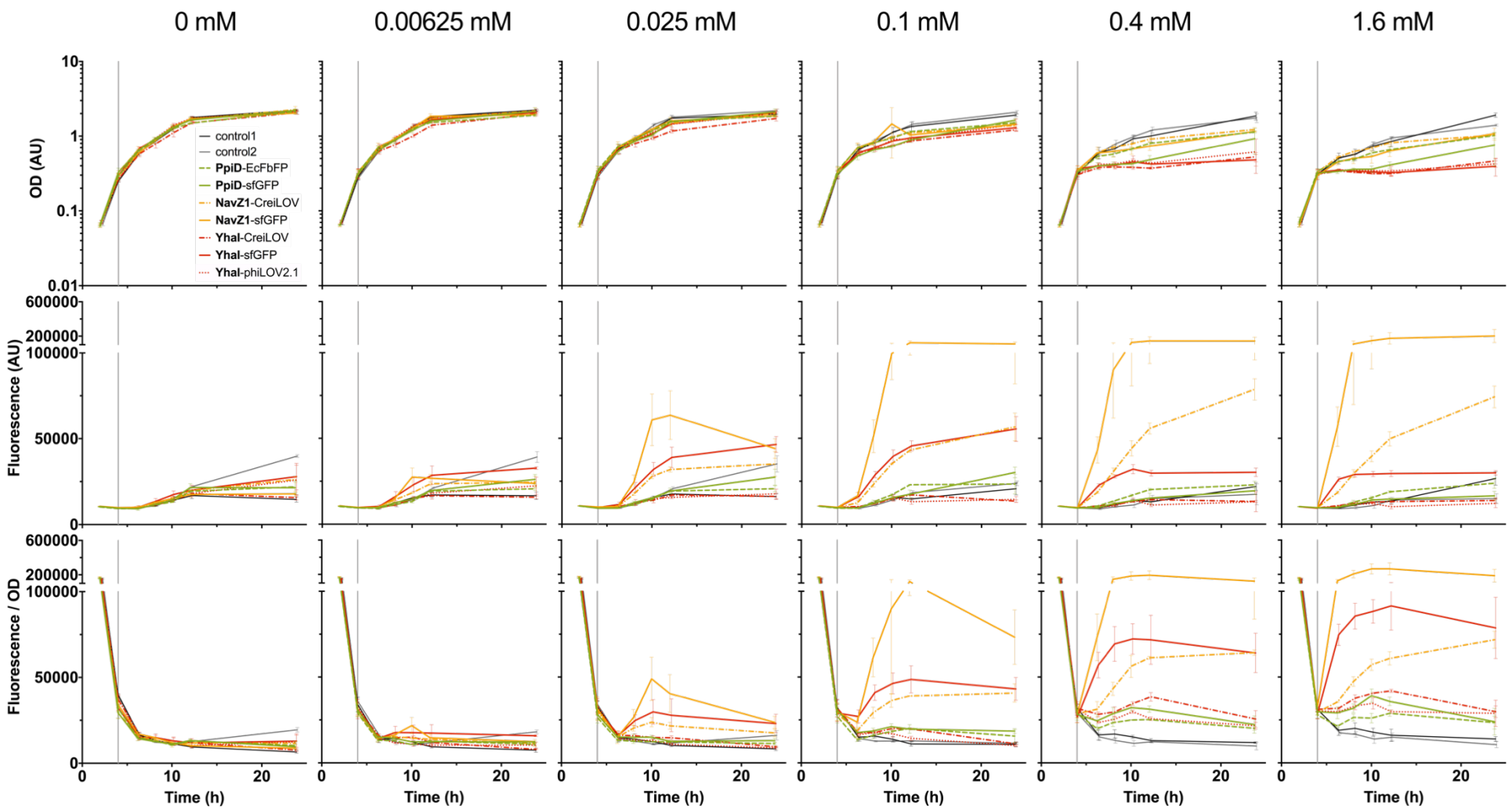


Figure 14: **Optimising the promoter strength of *T7p* for rsTEVP(M)-dgFtsH-MP-FR-His production from pCOLA by taking FR fluorescence as indicator for functional IMP integration in *E. coli* K-12 MG1655 (DE3)**. The colour codes for the IMP-FR constructs are indicated in the legend. The line at 4 h represents the time point of IPTG supply (or MQ water). The final IPTG concentrations are specified. $r_B = 3$, $r_T = 1$, n = 1.

In this experiment, we observed pronounced physiological stress related to IPTG toxicity [294] (especially seen in the controls) and likely related to metabolic load upon the production of the different IMPs; the YhaI constructs showed the most prominent negative effects on growth. The maximum relative fluorescence levels (fluorescence/OD) increased proportionally with increasing IPTG concentrations for all constructs, and the levels were maximal 6 h after IPTG supply (time point 10 h). This time point was then chosen for TEVP induction with rhamnose, hypothesising that the fluorescence pool was highest for all constructs and thus its putative decay would be easier to follow during proteolysis. A dose of 0.1 mM IPTG was chosen in order find a compromise between IPTG toxicity and maximum fluorescence during growth.

3.6 Inducing membrane protein's susceptibility towards FtsH-specific proteolysis

After having defined the optimal experimental conditions, we carried out the main experiment addressing the IMP-FR production in *E. coli* followed by L-rhamnose-triggered TEVP cleavage to induce the IMP's susceptibility towards FtsH degradation (Figure 15). Simultaneously, we planned to sample cell cultures for immunoblotting to visualise the abundance of successful integrated IMP-FRs, on one hand, and, on the other hand, putative degradation events directly on protein level.

Rapidly during the first attempt, aberrant cell density data were observed (Figure 16B): within the first 1 to 15 h of the experiment, the cell density no longer followed the trend observed during the initial test experiment (Figure 16A) and cells seemed to die during cultivation. The experiment was repeated three more times with the similar outcome. This was followed by thorough troubleshooting activities.

The problematic growth curves seemed randomly distributed yet apparently related to specific deep well plates used during the experiments. A physiological explanation for the symptoms was ruled out, thanks to all previous experiments demonstrating the non-lethal effect(s) of the many different parameters chosen. More reasonable, although somehow far-fetched, seemed a phage contamination.

A collection of samples – from media, ingredients to supernatants of culture broths of the shown experiment – were tested in a plaque assay. To our surprise, the supernatant of the culture broth samples that had exhibited the unusual growth behaviour showed a clear plaque upon incubation (Figure 17). Here, phages were lysogenic on growing *E. coli* K-12 MG1655 (DE3) – a similar behaviour we have observed in the exponential phase during the growth experiment. The plaque assay was negative for all media and ingredients as well as for the pre-culture of the tested clone. None of lab equipment seemed to be a cause of infection. The only likely explanations were therefore, (1) a bacteriophage was responsible for the death of the cells during the experiments and (2) the only reasonable source for a phage contamination were the

re-usable 24-deep-well plates. Due to lack of time, this essential experiment could not be repeated once again with newly ordered consumables.

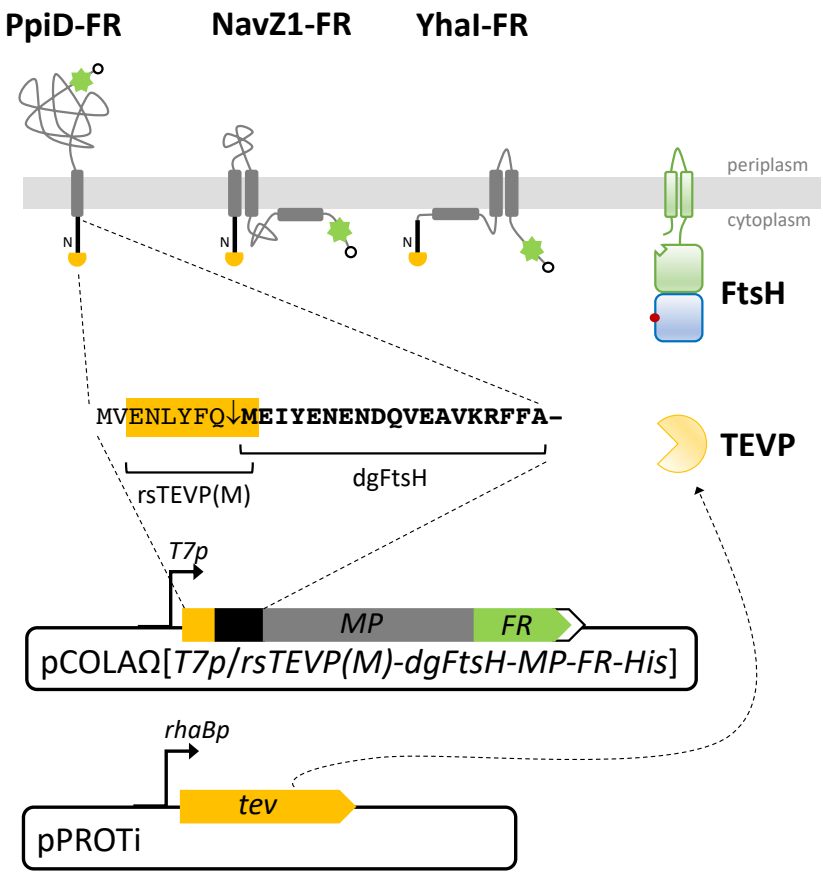


Figure 15: **Experimental set-up for testing the susceptibility of all inner MP-FR candidates towards FtsH-mediated proteolysis**. pCOLAΩ[T7p/rsTEVP(M)-dgFtsH-MP-FR-His] encoding for PpiD-FR, NavZ1-FR and YhaI-FR fusion proteins were each co-transformed with pPROTi into E. coli MG1655 (DE3). All MP-FR fusion were N-terminally tagged with the FtsH-specific conditional degradation tag which was assumed to turn them into FtsH substrates upon TEVP cleavage. Legend: Important N- or C-terminal protein features are indicated in colour or symbols: native dgFtsH (black bold letters or line) for FtsH-dependent proteolysis; the rsTEVP(M) capping (orange highlighted or ◖) as part of the conditional element together with TEVP, varying FR including sfGFP, EcFbFP, phiLOV2.1 and CreiLOV are summarised in one symbol (✸); His_8-tag (O) as epitope tag.

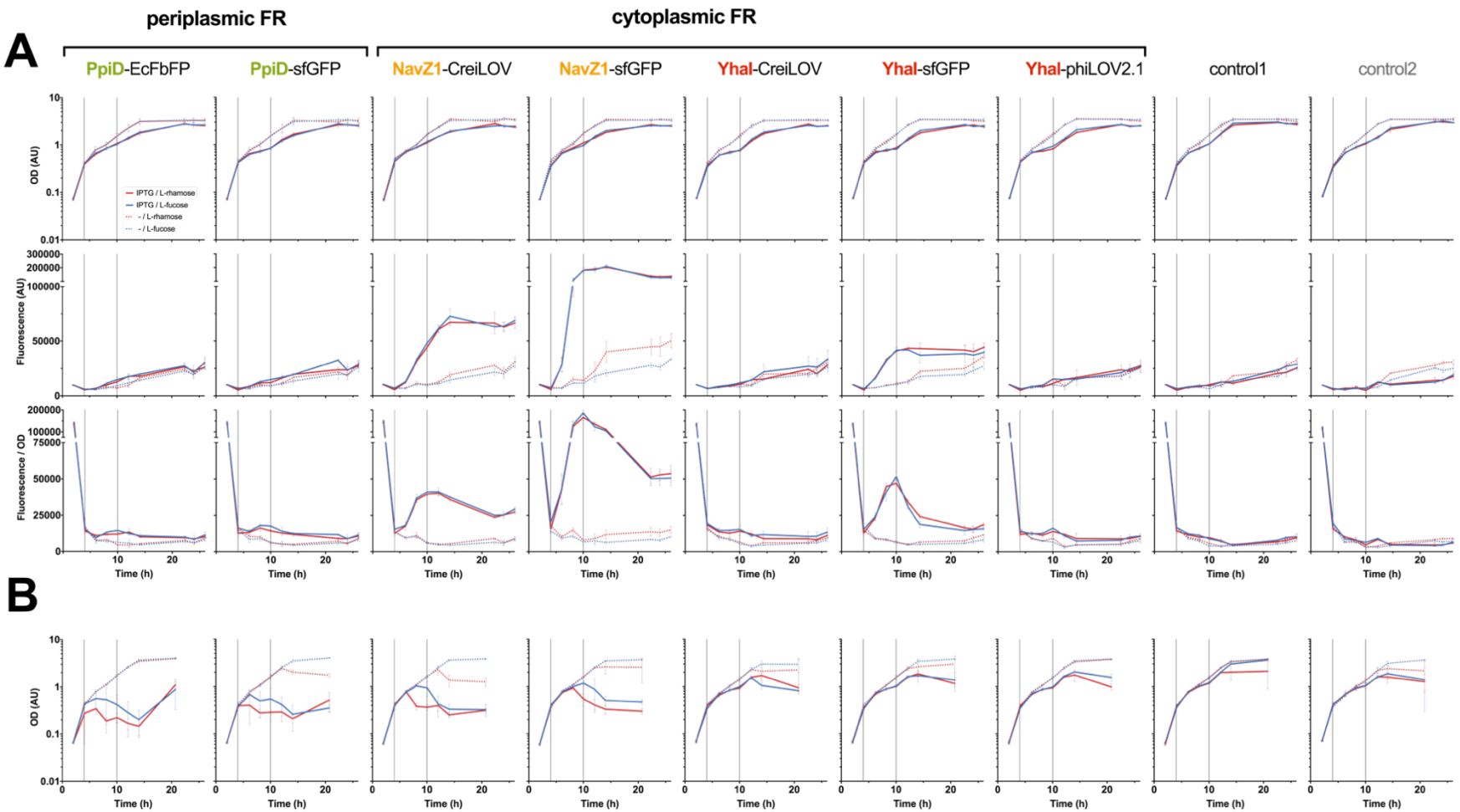


Figure 16: **Testing the susceptibility of the overproduced rsTEVP(M)-dgFtsH-MP-FR-His constructs towards FtsH proteolysis upon TEVP cleavage**. (A) Test experiment. All E. coli MG1655 (DE3) strains harboured a pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-MP-FR-His] construct and the pPROTi-His plasmid. No cultures sampled for immunoblotting. r_B = 3, r_T = 1, n = 1. (B) Representative growth curves of phage-infected expression experiments: Here, the strains were co-transformed with a pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-MP-FR-His] and the pPROTi plasmid. r_B = 3, r_T = 1, n = 4. Legend: The respective MP-FR fusions are specified above the graphs. The line at 4 h and 10 h represent the time point of IPTG supply (or MQ water) or media change from LB/±IPTG to LB/sugar, respectively. The colour codes for the different conditions and MP-FR fusions were kept from previous experiments and are indicated in the legends.

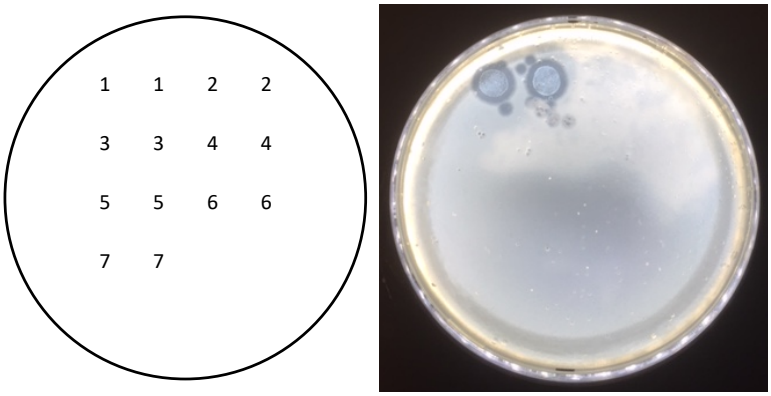


Figure 17: **Plaque assay with top agar shows typical plaque formation only in culture with suspicious growth phenotype**. All media, supplements and respective ON cultures were tested in parallel: (1) supernatant of two cultures[I] with suspicious growth phenotype from experiment; (2) respective ON culture; (3) LB medium supplemented with 0.5 % (w/v) D-glucose used for the pre-culture; (4) LB stock medium from which all LB variants were prepared; (5) fresh LB medium supplemented with 0.1 mM IPTG from the stock solution used for the experiment; (6) LB medium supplemented with 5 mM L-fucose and (7) LB medium supplemented with 5 mM L-rhamnose, both used during the experiment. Plate was incubated at 30 °C overnight. *E. coli* MG1655 (DE3) with plasmid-derived CamR and KanR resistance was used as host in the top agar. $r_B = 2$, $r_T = 1$, n = 1.

However, some statements can be provided on the outcome of the test experiment. The relative fluorescence levels for PpiD fused to its periplasmic FRs did not significantly increase upon IPTG supply compared to the control condition (IPTG-negative). It is conceivable that PpiD-EcFbFP and PpiD-sfGFP could not be integrated in the IM or have not even been produced. Furthermore, even though EcFbFP as well as sfGFP have been both shown active as part of a soluble fusion protein in the periplasm after SecYEG transport, it is possible that both FRs were here inactive due incorrect folding. The same applies for YhaI-CreiLOV and YhaI-phiLOV2.1. Nevertheless, a significant increase for relative fluorescence levels upon IPTG addition was observed for both NavZ1 constructs and YhaI-sfGFP. Interestingly, for none of these three protein fusions the relative fluorescence dropped upon induction of TEVP production, which might be interpreted as the following: (1) those IMP fusion constructs could not be turned into a FtsH substrate by the present method which might also be related to their higher topology grade or (2), similar to an assumption in STUDY II section 3.4, the medium change, necessary for TEVP induction from the pPROTi plasmid, caused a prolonged exponential growth phase which is unfavourable for FtsH-mediated degradation of dgFtsH-tagged proteins [2]. Only an immunoblot could have provided insights on the abundance and quality of IMP-FR integration and the occurrence of FtsH-mediated proteolysis *via* the developed conditional degradation tag (rsTEVP(M)-dgFtsH-).

[I] One clone each of E. coli MG1655 (DE3) co-transformed with either pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-PpiD-EcFbFP-His] or pCOLAΩ[*T7p*/rsTEVP(M)-dgFtsH-PpiD-EcFbFP-His] and pPROTi cultured in condition IPTG/L-rhamnose during the experiment.

3.7 Concluding remarks

The transfer of the PROTi technology on membrane proteins was an ambitious project which was unfortunately hampered by succession of technical challenges. With the present data, it is not possible to conclude if the project failed to deliver its results – the unfortunate bacteriophage incident combined with lack of time did not allow to carry out the last, most crucial experiment including immunoblotting of IMP-FR integration into the IM and putative degradation events. Albeit optimised during the preliminary trials, we believe that our experimental setup may still not have been suitable to match the intrinsic biological nature of the different components involved.

For a start, TEVP, as part of TIPI or PROTi, may be limited – to a certain extend – by type of amino acid in the 7th position of its rsTEVP (discussed in INTRODUCTION section 3.1) and therefore this makes this endopeptidase unfavourable when creating libraries of random degradation tags since TEVP will skew the results towards its preferred residues. Another protease with even lower or no specificity for the released amino acid, as eventually UBP1 [56] (discussed in INTRODUCTION section 2.2.3) would be preferable. However, if the N-terminal residue to be released remains identical throughout experiments (except Pro) then the use of TEVP can be of advantage due to its high specificity and for its relatively small recognition site.

Secondly, with retrospect, we identified several pitfalls related to the choice of promoters used during these studies. The use of an IPTG-inducible promoter for target protein production was not suitable when combined with TEVP induction. The medium change (for the removal of residual IPTG and to stop POI gene expression) was not only laborious but also increased the risk for cross-contamination in our deep-well plate setup. The use of a metabolite alternatives to IPTG for LacI-regulated promoters, for instance lactose, could have been an alternative option (see APPENDIX Figure S 5) although such a strategy would increase significantly in terms of complexity. Indeed, there is a great need to maximize the level of POI gene expression while the inducer should be reduced to the lowest possible prior to triggering TEVP production. Another type of promoter could have alleviated the need for removing the inducing molecule, for instance light-driven and temperature driven promoters – there would still be risks that such promoters may not be as tight.

Speaking of tightness, we have surprisingly observed increased basal expression from LacI-regulated promoters upon the addition of galactose or rhamnose; abolishing LacI repression should be by all means avoided when considering the production of toxic proteins or when certain protein activities are unfavoured at a given time namely in an engineered metabolic pathway. The combination of IPTG-driven protein production with rhamnose-driven TEVP production should therefore be re-evaluated such as both promoter types to be switched. By doing so, one would avoid interrupting the experiment by a medium change, including adding a substantial physiological stress to the cells, although a certain degree of *T7p* leakiness may not be avoided; the co-expression of T7 lysozyme [295] which binds T7 RNAP might help improving the leakiness of *T7p* indirectly. Moreover, the medium change is thought to lead to a second or

prolonged exponential phase which is unfavourable for FtsH proteolysis from dgFtsH-tagged POIs, whose activity peaks during the stationary phase [2]. It would therefore be crucial to reach the stationary phase as rapidly as possible.

In addition, it would have been preferable to use a L-rhamnose catabolism deficient strain for our studies due to two reasons: (1) L-rhamnose would not confer any physiological advantages thus no replacing sugar would be needed during growth (discussed in STUDY II section 3.3) and (2) the *rhaB* promoter would become directly titratable, as recently published [249]. It is obvious that there is no ideal combination of promoters in our case and one would have to design a substantially more complex set up in order to achieve full promoter tightness, no interruption of growth and optimal production of all desired targeted proteins.

Thirdly, a more focused study using only one, well characterized model of membrane protein might have increased our chances to meet our goals. To start the cloning of all proteins of interest was unexpectedly troublesome and some immunoblot data (data not shown) might indicate that most of the proteins were not even produced. As stated in the INTRODUCTION (sections 3.3.2, 3.3.3 and 3.3.4), PpiD, NavZ1 and YhaI were little or poorly characterized in their natural, physiological context. Among other there aren't any information about the respective mechanisms of insertion into the IM, including also which topological determinants are responsible for this insertion. Hypothetically, upon replacing parts of the wild-type N-terminal sequence of those proteins with the FtsH-specific conditional degradation tag (rsTEVP(M)-dgFtsH-) we might have disturbed the natural signals involved with membrane insertion and assembly. There aren't any certainties either that the fusion with our FRs did not alter the insertion properties of the different target membrane proteins. It is yet still unclear why all the mutations in the ORFs of the different plasmids occurred when cloning in presence of active promoters (*T5p* in non-(DE3) cloning strains) – those mutations may very well be the root cause for defective or unsuccessful protein production.

Finally, there are to date only few reporter fluorescent proteins that are active in the periplasm so the number of molecules to choose from in our study was limited. The nature of the FR should match the proteasome's unfoldase abilities but should, of course, still allow for a clear read-out. The choice of LOV proteins as alternatives to GFP appeared considerate enough but this may have backfired in the sense that they may exhibit a higher intrinsic thermodynamic stability (discussed in INTRODUCTION section 3.4) – and hence being a more reluctant substrate for FtsH - than initially assumed.

FUTURE PERSPECTIVES

The work presented in this thesis is ambivalent. We have successfully addressed conditional protein degradation of soluble proteins (Study I) yet failed to obtain the same results for membrane proteins (Study II). Thanks to the synergy between PROTi with CRISPRi, we anticipate that our technology would enable faster discoveries in the fields of fundamental biology but also applied biotechnology; one example could be the discovery of new antibiotics.

Addressing our work in Study II: despite the outcome of our trials, we believe that this project would still be able to deliver its promises. As such, mechanisms for degradation of membrane proteins are still vastly unknown. A simplified study encompassing a thorough re-examination of the items discussed here above might be able to uncover initial aspects of membrane protein proteolysis, and by extension paving the way towards new tools in synthetic biology.

REFERENCES

1. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **20,** 1298–1345 (2011).
2. Bittner, L. M., Westphal, K. & Narberhaus, F. Conditional proteolysis of the membrane protein YfgM by the FtsH protease depends on a novel N-terminal degron. *J. Biol. Chem.* **290,** 19367–19378 (2015).
3. Erbse, A. *et al.* ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **439,** 753–756 (2006).
4. Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17,** 691–695 (1999).
5. Lehning, C. E. Allosteric coupling of ligand-binding to signal output within rationally designed chimeric receptors. (2013).
6. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).
7. Nørholm, M. H. H. A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol.* **10,** (2010).
8. IUPAC-IUB Commission on Biochemical Nomenclature (CBN). A One-Letter Notation for Amino Acid Sequences. *Eur. J. Biochem.* **5,** 151–153 (1968).
9. ASM. Instructions to authors. *J. Bacteriol.* 1–20 (2018). doi:10.1093/llc/8.3.189
10. Berlyn, K. B. M. Linkage Map of Escherichia coli K-12, Edition 10: The Traditional map. *Microbiol. Mol. Biol. Rev.* **62,** 814–984 (1998).
11. Hu, J. Help: Genetic nomenclature. *EcoliWiki* Page ID 110187 (2007).
12. Amm, I., Sommer, T. & Wolf, D. H. Protein quality control and elimination of protein waste: The role of the ubiquitin-proteasome system. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843,** 182–196 (2014).
13. Geiler-Samerotte, K. A., Dion, M. F., Budnik, B. A., Wang, S. M. & Hartl, D. L. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfloded protein response in yeast. *PNAS* **108,** 680–685 (2010).
14. Jenal, U. & Hengge-Aronis, R. Regulation by proteolysis in bacterial cells. *Curr. Opin. Microbiol.* **6,** 163–172 (2003).
15. Varshavsky, A. in *Encyclopedia of Genetics* 2091–2093 (2001). doi:10.1006/rwgn.2001.1348
16. Bojkowska, K. *et al.* Measuring in vivo protein half-life. *Chem. Biol.* **18,** 805–815 (2011).
17. Dougan, D. A., Truscott, K. N. & Zeth, K. The bacterial N-end rule pathway: Expect the unexpected. *Mol. Microbiol.* **76,** 545–558 (2010).
18. 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,** e0149746 (2016).
19. Stikane, A. Degrons. *iGEM Team RewirED* (2014). Available at: http://2014.igem.org/Team:Edinburgh/project/degrons.
20. McGinness, K. E., Baker, T. A. & Sauer, R. T. Engineering Controllable Protein Degradation. *Mol. Cell* **22,** 701–707 (2006).
21. Taverna, D. M. & Goldstein, R. A. Why are proteins marginally stable? *Proteins Struct. Funct. Genet.* **46,** 105–109 (2002).
22. Gottesman, S. & Maurizi, M. R. Regulation by proteolysis: energy-dependent proteases and their targets. *Microbiol. Rev.* **56,** 592–621 (1992).
23. Davis, J. H., Baker, T. A. & Sauer, R. T. Small- Molecule Control of Protein Degradation Using Split Adaptors. *ACS Chem. Biol.* **6,** 1205–1213 (2011).
24. Gur, E., Biran, D. & Ron, E. Z. Regulated proteolysis in Gram-negative bacteria — how and when? *Nat Rev Microbiol* **9,** 839–848 (2011).
25. Schrader, E. K., Harstad, K. G. & Matouschek, A. Targeting proteins for degradation. *Nat. Chem. Biol.* **5,** 815–822 (2009).
26. Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nat. Biotechnol.* **32,** 1276–1283 (2014).
27. Corbin, R. W. *et al.* Toward a protein profile of Escherichia coli: Comparison to its transcription profile. *Proc. Natl. Acad. Sci.* **100,** 9232–9237 (2003).
28. Ishihama, Y. *et al.* Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics* **9,** 1–17 (2008).
29. Butler, S. M., Festa, R. A., Pearce, M. J. & Darwin, K. H. Self-compartmentalized bacterial proteases and pathogenesis. *Mol. Microbiol.* **60,** 553–562 (2006).

30. Jenal, U. The role of proteolysis in the Caulobacter crescentus cell cycle and development. *Res. Microbiol.* **160,** 687–695 (2009).
31. Lecker, S. H. Protein Degradation by the Ubiquitin-Proteasome Pathway in Normal and Disease States. *J. Am. Soc. Nephrol.* **17,** 1807–1819 (2006).
32. Blanchard, J. S. Old approach yields new antibiotic. *Nat. Med.* **11,** 1045–1046 (2005).
33. Sauer, R. T. & Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **80,** 587–612 (2011).
34. Maurizi, M. R. Proteases and protein degradation in Escherichia coli. *Experientia* **48,** 178–201 (1992).
35. Merdanovic, M., Clausen, T., Kaiser, M., Huber, R. & Ehrmann, M. Protein quality control in the bacterial periplasm. *Annu. Rev. Microbiol.* **65,** 149–68 (2011).
36. Varshavsky, A. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12142–12149 (1996).
37. Raivio, T. L. Envelope stress responses and Gram-negative bacterial pathogenesis. *Mol. Microbiol.* **56,** 1119–1128 (2005).
38. Kim, S. S., Song, I., Eom, G. & Kim, S. S. A small periplasmic protein with a hydrophobic C-terminal residue enhances DegP proteolysis as a suicide activator. *J. Bacteriol.* **200,** JB.00519-17 (2017).
39. Isaac, D. D., Pinkner, J. S., Hultgren, S. J. & Silhavy, T. J. The extracytoplasmic adaptor protein CpxP is degraded with substrate by DegP. (2005).
40. Gur, E., Ottofueling, R. & Dougan, D. A. in *Regulated Proteolysis in Microorganisms* (ed. Dougan, D. A.) 3–33 (Springer Dordrecht Heidelberg New York London, 2013). doi:10.1007/978-94-007-5940-4_1
41. Olivares, A. O., Nager, A. R., Iosefson, O., Sauer, R. T. & Baker, T. A. Mechanochemical basis of protein degradation by a double-ring AAA + machine. *Nat. Struct. Mol. Biol.* **21,** 871–875 (2014).
42. Kress, W., Mutschler, H. & Weber-ban, E. Both ATPase Domains of ClpA Are Critical for Processing of Stable Protein Structures. *J. Biol. Chem.* **284,** 31441–31452 (2009).
43. Olivares, A. O., Baker, T. A. & Sauer, R. T. Mechanistic insights into bacterial AAA + proteases and protein-remodelling machines. *Nat. Rev. Microbiol.* **14,** 33–44 (2015).
44. Kirstein, J., Molière, N., Dougan, D. a & Turgay, K. Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases. *Nat. Rev. Microbiol.* **7,** 589–99 (2009).
45. Sprangers, R., Gribun, A., Hwang, P. M., Houry, W. A. & Kay, L. E. Quantitative NMR spectroscopy of supramolecular complexes : Dynamic side pores in ClpP are important for product release. *PNAS* **102,** 16678–16683 (2005).
46. Tasaki, T. *et al.* The substrate recognition domains of the N-end rule pathway. *J. Biol. Chem.* **284,** 1884–1895 (2009).
47. Varshavsky, A. Naming a targeting signal. *Cell* **64,** 13–15 (1991).
48. Bachmair, A. *et al.* In Vivo Half-Life of a Protein is a Function of its Amino-Terminal Residue. *Science (80-.).* **234,** 179–186 (1986).
49. Flynn, J. M. *et al.* Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. (2001).
50. Wang, K. H., Oakes, E. S. C., Sauer, R. T. & Baker, T. a. Tuning the strength of a bacterial N-end rule degradation signal. *J. Biol. Chem.* **283,** 24600–7 (2008).
51. Guharoy, M., Bhowmick, P., Sallam, M. & Tompa, P. Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system. *Nat. Commun.* **7,** 1–13 (2016).
52. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **20,** 1298–1345 (2011).
53. Johnson, E. S., Gonda, D. K. & Varshavsky, A. Cis-trans recognition and subunit-specific degradation of short-lived proteins. *J. Colloid Interface Sci.* **346,** 287–291 (1990).
54. Prakash, S., Inobe, T., Hatch, A. J. & Matouschek, A. Substrate selection by the proteasome during degradation of protein complexes. *Nat. Chem. Biol.* **5,** 29–36 (2009).
55. Cha-Molstad, H. *et al.* Modulation of SQSTM1/p62 activity by N-terminal arginylation of the endoplasmic reticulum chaperone HSPA5/GRP78/BiP. *Autophagy* **12,** 426–428 (2016).
56. Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, a. The N-end rule in bacteria. *Science* **254,** 1374–7 (1991).
57. Rogers, S. *et al.* Amino Acid Sequences Common to Rapidly Degraded Proteins : The PEST Hypothesis Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/1697772 REFERENCES Linked references are available on JSTOR for this. *Science (80-.).* **234,** 364–368 (1986).
58. Singh, G. P., Ganapathi, M., Sandhu, K. S. & Dash, D. Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins Struct. Funct. Genet.* **62,** 309–315 (2006).
59. Fortmann, K. T., Lewis, R. D., Ngo, K. A., Fagerlund, R. & Hoffmann, A. A Regulated, Ubiquitin-Independent

Degron in IκBα. *J. Mol. Biol.* **427,** 2748–2756 (2015).
60. Suzuki, T. & Varshavsky, A. Degradation signals in the lysine – asparagine sequence space. *EMBO J.* **18,** 6017–6026 (1999).
61. Sauer, R. T. *et al.* Sculpting the proteome with AAA+ proteases and disassembly machines. *Cell* **119,** 9–18 (2004).
62. Veitia, R. A. A Fresh Look at 'Aging' Proteins. *Trends Biochem. Sci.* **42,** 86–89 (2017).
63. Graciet, E. *et al.* Aminoacyl-transferases and the N-end rule pathway of prokaryotic–eukaryotic specificity in a human pathogen. (2006). doi:10.1073/pnas.0511224103
64. Holt, L. J. Regulatory modules: Coupling protein stability to phopshoregulation during cell division. *FEBS Lett.* **586,** 2773–2777 (2012).
65. Ravid, T. & Hochstrasser, M. Diversity of degradation signals in the ubiquitin-proteasome system. *Nat. Rev. Mol. Cell Biol.* **9,** 679–689 (2008).
66. Yang, Y. & Bedford, M. T. Titivated for Destruction: The Methyl Degron. *Mol. Cell* **48,** 487–488 (2012).
67. Ninnis, R. L., Spall, S. K., Talbo, G. H., Truscott, K. N. & Dougan, D. A. Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in Escherichia coli. *EMBO J.* **28,** 1732–1744 (2009).
68. Dohmen, J. Starting with a degron: N-terminal formyl-methionine of nascent bacterial proteins contributes to their proteolytic control. *Microb. Cell* **2,** 356–359 (2015).
69. Piatkov, K., Vu, T., Hwang, C.-S. & Varshavsky, A. Formyl-methionine as a degradation signal at the N-termini of bacterial proteins. *Microb. Cell* **2,** 376–393 (2015).
70. Joshi, a & Rao, B. J. MutS recognition: multiple mismatches and sequence context effects. *J. Biosci.* **26,** 595–606 (2001).
71. Rao, H., Uhlmann, F., Nasmyth, K. & Varshavsky, A. Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. *Nature* **410,** 955–959 (2001).
72. Zhu, J. & Winans, S. C. The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 1507–12 (2001).
73. Hill, C. P., Johnston, N. L. & Cohen, R. E. Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 4136–4140 (1993).
74. Burns, K. E. & Darwin, K. H. Pupylation versus ubiquitylation: Tagging for proteasome-dependent degradation. *Cell. Microbiol.* **12,** 424–431 (2010).
75. Samanovic, M. I., Li, H. & Darwin, K. H. in 267–295 (2013). doi:10.1007/978-94-007-5940-4_10
76. Humbard, M. A. *et al.* Ubiquitin-like small archaeal modifier proteins (SAMPs) in Haloferax volcanii. *Nature* **463,** 54–60 (2010).
77. Darwin, K. H. & Hofmann, K. SAMPyling proteins in archaea. *Trends Biochem. Sci.* **35,** 348–351 (2010).
78. Maupin-Furlow, J. A. in *Regulated Proteolysis in Microorganisms* (ed. Dougan, D. A.) 297–327 (Springer, Dortrecht, 2013). doi:10.1007/978-94-007-5940-4_11
79. Sherman, F., Stewart, J. W. & Tsunasawa, S. Methionine or not methionine at the beginning of a protein. *Bioessays* **3,** 27–31 (1985).
80. Touriol, C. *et al.* Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol. Cell* **95,** 169–178 (2003).
81. Frottin, F. *et al.* The Proteomics of N-terminal Methionine Cleavage. *Mol. Cell. Proteomics* **5,** 2336–2349 (2006).
82. Karzai, A. W., Roche, E. D. & Sauer, R. T. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat. Struct. Biol.* **7,** 449–455 (2000).
83. Keiler, K. C. & Sauer, R. T. Sequence determinants of C-terminal substrate recognition by the Tsp protease. *J. Biol. Chem.* **271,** 2589–2593 (1996).
84. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12,** 1338–1347 (1998).
85. Choy, J. S., Aung, L. L. & Karzai, A. W. Lon protease degrades transfer-messenger RNA-tagged proteins. *J. Bacteriol.* **189,** 6564–6571 (2007).
86. Herman, C., Thévenet, D., Bouloc, P., Walker, G. C. & D'Ari, R. Degradation of carboxy-terminal-tagged cytoplasmic proteins by the Escherichia coli protease HflB (FtsH). *Genes Dev.* **12,** 1348–1355 (1998).
87. Chai, Q., Wang, Z., Webb, S. R., Dutch, R. E. & Wei, Y. The ssrA-Tag Facilitated Degradation of an Integral Membrane Protein. *Biochemistry* **55,** 2301–2304 (2016).
88. Hari, S. B. & Sauer, R. T. The AAA+ FtsH protease degrades an ssrA-tagged model protein in the inner membrane of Escherichia coli. *Biochemistry* **55,** 5649–5652 (2016).
89. Kim, Y.-I., Burton, R. E., Burton, B. M., Sauer, R. T. & Baker, T. A. Dynamics of Substrate Denaturation and

Translocation by the ClpXP Degradation Machine. *Mol. Cell* **5,** 639–648 (2000).
90. Keiler, K. C., Waller, P. R. & Sauer, R. T. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271,** 990–993 (1996).
91. Wang, J., Hartling, J. A. & Flanagan, J. M. Resolution The Structure of ClpP at 2.3 A Suggests a Model for ATP-Dependent Proteolysis. *Cell* **91,** 447–456 (1997).
92. Maglica, Ž., Striebel, F. & Weber-Ban, E. An Intrinsic Degradation Tag on the ClpA C-Terminus Regulates the Balance of ClpAP Complexes with Different Substrate Specificity. *J. Mol. Biol.* **384,** 503–511 (2008).
93. Dougan, D. A. *et al.* ClpS, a Substrate Modulator of the ClpAP Machine. *Mol. Cell* **9,** 673–683 (2002).
94. Schuenemann, V. J. *et al.* Structural basis of the N-end rule substraterecognition in Escherichia coli by the ClpAP adaptor protein ClpS. *EMBO Rep.* **10,** 508–514 (2009).
95. Schmidt, R., Zahn, R., Bukau, B. & Mogk, A. ClpS is the recognition component for Escherichia coli substrates of the N-end rule degradation pathway. **72,** 506–517 (2009).
96. Krzywda, S. *et al.* The Crystal Structure of the AAA Domain of the ATP-Dependent Protease FtsH of Escherichia coli at 1.5 A Resolution. *Structure* **10,** 1073–1083 (2002).
97. Langklotz, S., Baumann, U. & Narberhaus, F. Structure and function of the bacterial AAA protease FtsH. *Biochim. Biophys. Acta - Mol. Cell Res.* **1823,** 40–48 (2012).
98. Akiyama, Y. Quality Control of Cytoplasmic Membrane Proteins in Escherichia coli. *J. Biochem.* **146,** 449–454 (2009).
99. Okuno, T. & Ogura, T. in 53–69 (2013). doi:10.1007/978-94-007-5940-4_3
100. Bittner, L. M., Arends, J. & Narberhaus, F. When, how and why? Regulated proteolysis by the essential FtsH protease in Escherichia coli. *Biol. Chem.* **398,** 625–635 (2017).
101. Ito, K. & Akiyama, Y. Cellular Functions, Mechanism of Action, and Regulation of Ftsh Protease. *Annu. Rev. Microbiol.* **59,** 211–231 (2005).
102. Schumann, W. FtsH - a single-chain charonin? *FEMS Micribiol. Rev.* **23,** 1–11 (1999).
103. Herman, C., Prakash, S., Lu, C. Z., Matouschek, A. & Gross, C. A. Lack of a robust unfoldase activity confers a unique level of substrate specificity to the universal AAA protease FtsH. *Mol. Cell* **11,** 659–669 (2003).
104. Ogura, T. *et al.* Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in Escherichia coli. *Mol. Microbiol.* **31,** 833–844 (1999).
105. Katz, C. & Ron, E. Z. Dual Role of FtsH in Regulating Lipopolysaccharide Biosynthesis in Escherichia coli. *J. Bacteriol.* **190,** 7117–7122 (2008).
106. van Bloois, E. *et al.* Detection of cross-links between FtsH, YidC, HflK/C suggests a linked role for these proteins in quality control upon insertion of bacterial inner membrane proteins. *FEBS Lett.* **582,** 1419–1424 (2008).
107. van Stelten, J., Silva, F., Belin, D. & Silhavy, T. J. Effects of Antibiotics and a Proto-Oncogene Homolog on Destruction of Protein Translocator SecY. *Science (80-.).* **325,** 753–757 (2009).
108. Akiyama, Y., Kihara, A. & Ito, K. Subunit a of proton ATPase F0 sector is a substrate of the FtsH protease in Escherichia coli. *FEBS Lett.* **399,** 26–28 (1996).
109. Akiyama, Y., Kihara, A., Tokuda, H. & Ito, K. FtsH (HflB) Is an ATP-dependent Protease Selectively Acting on SecY and Some Other Membrane Proteins. *J. Biol. Chem.* **271,** 31196–31201 (1996).
110. Singh, S. & Darwin, A. J. FtsH-Dependent Degradation of Phage Shock Protein C in Yersinia enterocolitica and Escherichia coli. *J. Bacteriol.* **193,** 6436–6442 (2011).
111. Shotland, Y. *et al.* Proteolysis of the phage lambda CII regulatory protein by FtsH (HflB) of Escherichia coli. *Mol. Microbiol.* **24,** 1303–1310 (1997).
112. Gamer, J. *et al.* A cycle of binding and release of the DnaK , DnaJ and GrpE chaperones regulates activity of the Escherichia coli heat shock transcription factor & 2. **15,** 607–617 (1996).
113. Parsell, A., Silber, K. R. & Sauer, R. T. Carboxy-terminal determinants of intracellular protein degradation. *Genes Dev.* **4,** 277–286 (1990).
114. Long, J. C., Wang, S. & Vik, S. B. Transmembrane Topography of Subunit a in the Escherichia coli F1F0 ATP Synthase as Determined by Labeling of Unique Cysteine Residues. *J. Biol. Chem.* **273,** 16241–16247 (1998).
115. Collinson, I., Corey, R. A., Allen, W. J. & Collinson, I. Channel crossing : how are proteins shipped across the bacterial plasma membrane? *Philos. Trans. R. Soc. London. Ser. B* **370,** 20150025 (2015).
116. Jovanovic, G., Engl, C., Mayhew, A. J., Burrows, P. C. & Buck, M. Properties of the phage-shock-protein (Psp) regulatory complex that govern signal transduction and induction of the Psp response in Escherichia coli. *Microbiology* **156,** 2920–2932 (2010).
117. Langer, T. AAA proteases: Cellular machines for degrading membrane proteins. *Trends Biochem. Sci.* **25,** 247–251 (2000).

118. Chiba, S., Akiyama, Y. & Ito, K. Membrane Protein Degradation by FtsH Can Be Initiated from Either End. *J. Bacteriol.* **184,** 4775–4782 (2002).
119. Kihara, A., Akiyama, Y. & Ito, K. Dislocation of membrane proteins in FtsH-mediated proteolysis. **18,** 2970–2981 (1999).
120. Shimohata, N., Chiba, S., Saikawa, N., Ito, K. & Akiyama, Y. The Cpx stress response system of Escherichia coli senses plasma membrane proteins and controls HtpX , a membrane protease with a cytosolic active site. 653–662 (2002).
121. Müller, S. A., Scilabra, S. D. & Lichtenthaler, S. F. Proteomic Substrate Identification for Membrane Proteases in the Brain. *Front. Mol. Neurosci.* **9,** 96 (2016).
122. Avci, D. & Lemberg, M. K. Clipping or Extracting: Two Ways to Membrane Protein Degradation. *Trends Cell Biol.* **25,** 611–622 (2015).
123. Rather, P. Role of rhomboid proteases in bacteria. *Biochim. Biophys. Acta - Biomembr.* **1828,** 2849–2854 (2013).
124. Martínez, V. *et al.* CRISPR/Cas9-based genome editing for simultaneous interference with gene expression and protein stability. *Nucleic Acids Res.* **45,** e171 (2017).
125. Andersen, J. B. *et al.* New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria. *Appl. Environ. Microbiol.* **64,** 2240–2246 (1998).
126. Li, X. *et al.* Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter. *J. Biol. Chem.* **273,** 34970–34975 (1998).
127. Park, E. C., Finley, D. & Szostak, J. W. A strategy for the generation of conditional mutations by protein destabilization. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 1249–1252 (1992).
128. Jentsch, S. The ubiquitin-conjugation system. *Annu. Rev. Genet.* **26,** 179–207 (1992).
129. Vierstra, R. D. Protein degradation in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44,** 385–410 (1993).
130. Pickart, C. M. & Eddins, M. J. Ubiquitin : structures , functions , mechanisms. *Biochim. Biophys. Acta - Rev.* **1695,** 55–72 (2004).
131. Zhou, P., Bogacki, R., McReynolds, L. & Howley, P. M. Harnessing the ubiquitination machinery to target the degradation of specific cellular proteins. *Mol. Cell* **6,** 751–756 (2000).
132. Gosink, M. M. & Vierstra, R. D. Redirecting the specificity of ubiquitination by modifying ubiquitin-conjugating enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 9117–9121 (1995).
133. Su, Y., Ishikawa, S., Kojima, M. & Liu, B. Eradication of pathogenic beta-catenin by Skp1/Cullin/F box ubiquitination machinery. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12729–34 (2003).
134. Janse, D. M., Crosas, B., Finley, D. & Church, G. M. Localization to the proteasome is sufficient for degradation. *J. Biol. Chem.* **279,** 21415–21420 (2004).
135. Sakamoto, K. M. *et al.* Protacs: Chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl. Acad. Sci.* **98,** 8554–8559 (2001).
136. Schneekloth, J. S. *et al.* Chemical Genetic Control of Protein Levels: Selective in Vivo Targeted Degradation. *J. Am. Chem. Soc.* **126,** 3748–3754 (2004).
137. Taxis, C. & Knop, M. in *Ubiquitin Family Modifiers and the Proteasome: Reviews and Protocols, Methods in Molecular Biology* (eds. Dohmen, R. J. & Scheffner, M.) **832,** 611–626 (Springer Science+Business Media, LLC, 2012).
138. Dohmen, R. J., Wu, P., Varshavsky, A., Gold, A. & Altman, S. Heat-Inducible Degron : A Method for Constructing Temperature-Sensitive Mutants. *Science (80-.).* **263,** 1273–1276 (1994).
139. Miller, H. I. *et al.* Cloning and Expression of a Yeast Ubiquitin-Protein Cleaving Activity in Escherichia Coli. *Bio/Technology* **7,** 698–704 (1989).
140. Tobias, J. W. & Varshavsky, A. Cloning and functional analysis of the ubiquitin-specific protease gene UBP1 of Saccharomyces cerevisiae. *J. Biol. Chem.* **266,** 12021–12028 (1991).
141. Baker, R. T., Tobias, J. W. & Varshavsky, A. Ubiquitin-specific proteases of Saccharomyces cerevisiae: Cloning of UBP2 and UBP3, and functional analysis of the UBP gene family. *J. Biol. Chem.* **267,** 23364–23375 (1992).
142. Kapust, R. B., Tözsér, J., Copeland, T. D. & Waugh, D. S. The P1' specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **294,** 949–55 (2002).
143. Taxis, C., Stier, G., Spadaccini, R. & Knop, M. Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. *Mol. Syst. Biol.* **5,** 267 (2009).
144. Stankunas, K. *et al.* Conditional Protein Alleles Using Knockin Mice and a Chemical Inducer of Dimerization. *Mol. Cell* **12,** 1615–1624 (2003).
145. Pratt, M. R., Schwartz, E. C. & Muir, T. W. Small-molecule-mediated rescue of protein function by an

inducible proteolytic shunt. *Proc. Natl. Acad. Sci.* **104,** 11209–11214 (2007).
146. Kapust, R. B. *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14,** 993–1000 (2001).
147. Renicke, C. *et al.* A Tobacco Etch Virus Protease with Increased Substrate Tolerance at the P1' position. *PLoS One* **8,** e67915 (2013).
148. Dougherty, W. G. & Dawn Parks, T. Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: Identification of an internal cleavage site and delimitation of VPg and proteinase domains. *Virology* **183,** 449–456 (1991).
149. Cesaratto, F., Burrone, O. R. & Petris, G. Tobacco Etch Virus protease: A shortcut across biotechnologies. *J. Biotechnol.* **231,** 239–249 (2016).
150. Phan, J. *et al.* Structural Basis for the Substrate Specificity of Tobacco Etch Virus Protease *. *J. Biol. Cdhemistry* **277,** 50564–50572 (2002).
151. Nallamsetty, S. *et al.* Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein Expr. Purif.* **38,** 108–115 (2004).
152. Cabrita, L. D. *et al.* Enhancing the stability and solubility of TEV protease using in silico design. *Protein Sci.* **16,** 2360–7 (2007).
153. Pace, C. N. & Scholtz, J. M. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **75,** 422–427 (1998).
154. Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci.* **37,** 205–211 (1951).
155. Chojnacki, S., Cowley, A., Lee, J., Foix, A. & Lopez, R. Programmatic access to bioinformatics tools from EMBL-EBI update : 2017. *Nucleic Acids Res.* **45,** W551 (2017).
156. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **111,** 229–233 (1992).
157. Westphal, K., Langklotz, S., Thomanek, N. & Narberhaus, F. A trapping approach reveals novel substrates and physiological functions of the essential protease Ftsh in Escherichia coli. *J. Biol. Chem.* **287,** 42962–42971 (2012).
158. Maddalo, G. *et al.* Systematic analysis of native membrane protein complexes in Escherichia coli. *J. Proteome Res.* **10,** 1848–1859 (2011).
159. Götzke, H. *et al.* YfgM is an ancillary subunit of the secyeg translocon in escherichia coli. *J. Biol. Chem.* **289,** 19089–19097 (2014).
160. Sikdar, R., Simmons, A. R. & Doerrler, W. T. Multiple envelope stress response pathways are activated in an Escherichia coli strain with mutations in two members of the deda membrane protein family. *J. Bacteriol.* **195,** 12–24 (2013).
161. Chew, F. N., Tan, W. S., Boo, H. C. & Tey, B. T. Statistical optimization of green fluorescent protein production from Escherichia coli BL21(DE3). *Prep. Biochem. Biotechnol.* **42,** 535–550 (2012).
162. Snapp, E. Design and Use of Fluorescent Fusion Proteins in Cell Biology. *Curr. Protoc. Cell Biol.* **Chapter,** 1–17 (2005).
163. Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67,** 509–544 (1998).
164. Zimmer, M. Green Fluorescent Protein (GFP): Applications, Structure, and Related Photophysical Behavior. *Chem. Rev.* **102,** 759–781 (2015).
165. Hessa, T. *et al.* Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450,** 1026–1030 (2007).
166. Yang, F., Moss, L. G., Phillips, G. N., Phillips Jr., G. N. & Phillips, G. N. The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **14,** 1246–1251 (1996).
167. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28,** 235–242 (2000).
168. Huang, X. & Miller, W. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **12,** 337–357 (1991).
169. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. for the Investigation of Sequences and Structures. *J. Mol. Biol.* **247,** 536–540 (1995).
170. Dartigalongue, C. & Raina, S. A new heat-shock gene, ppiD, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in Escherichia coli. *EMBO J.* **17,** 3968–3980 (1998).
171. Weininger, U., Jakob, R. P., Kovermann, M., Balbach, J. & Schmid, F. X. The prolyl isomerase domain of PpiD from Escherichia coli shows a parvulin fold but is devoid of catalytic activity. *Protein Sci.* **19,** 6–18 (2010).
172. Miot, M. & Betton, J.-M. Protein Quality Control in the Bacterial Periplasm. *Microb. Cell Fact.* **3,** 1–13 (2004).
173. Justice, S. S. *et al.* Periplasmic Peptidyl Prolyl cis-trans Isomerases Are Not Essential for Viability, but SurA

Is Required for Pilus Biogenesis in Escherichia coli. *J. Bacteriol.* **187,** 7680–7686 (2005).
174. Antonoaea, R., Fürst, M., Nishiyama, K. I. & Müller, M. The periplasmic chaperone PpiD interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry* **47,** 5649–5656 (2008).
175. Matern, Y., Barion, B. & Behrens-kneip, S. PpiD is a player in the network of periplasmic.pdf. (2010).
176. Ricci, D. P. & Silhavy, T. J. The Bam machine: A molecular cooper. *Biochim. Biophys. Acta - Biomembr.* **1818,** 1067–1084 (2012).
177. Goemans, C., Denoncin, K. & Collet, J. F. Folding mechanisms of periplasmic proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843,** 1517–1528 (2014).
178. Sachelaru, I., Petriman, N. A., Kudva, R. & Koch, H. G. Dynamic interaction of the Sec translocon with the chaperone PpiD. *J. Biol. Chem.* **289,** 21706–21715 (2014).
179. Geertsma, E. R., Groeneveld, M., Slotboom, D.-J. & Poolman, B. Quality control of overexpressed membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 5722–5727 (2008).
180. Daley, D. ~O. & et al. Global topology analysis of the Escherichia coli inner membrane proteome. *Science (80-.).* **308,** 1321–1323 (2005).
181. Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science (80-.).* **277,** 1453–1462 (1997).
182. Rudd, K. E. Linkage map of Escherichia coli K-12, edition 10: the physical map. *Microbiol. Mol. Biol. Rev.* **62,** 985–1019 (1998).
183. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).
184. von Heijne, G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* **5,** 3021–3027 (1986).
185. Snider, C., Jayasinghe, S., Hristova, K. & White, S. H. MPEx: A tool for exploring membrane proteins. *Protein Sci.* **18,** 2624–2628 (2009).
186. Rapp, M. *et al.* Experimentally based topology models for E. coli inner membrane proteins. *Protein Sci.* **13,** 937–945 (2004).
187. Drew, D. E., Heijne, G., Nordlund, P. & Gier, J. W. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Lett* **507,** 220–224 (2001).
188. Thorn, K. Genetically encoded fluorescent tags. *Mol. Biol. Cell* **28,** 848–857 (2017).
189. Mclendons, G. & Radany, E. Is Protein Turnover Thermodynamically Controlled? *J. Biol. Chem.* **253,** 6335–6337 (1978).
190. Kraut, D. A. *et al.* Sequence- and Species-Dependence of Proteasomal Processivity. *ACS Chem. Biol.* **7,** 1444–1453 (2012).
191. Koodathingal, P. *et al.* ATP-dependent Proteases Differ Substantially in Their Ability to Unfold Globular Proteins. *J. Biol. Chem.* **284,** 18674–18684 (2009).
192. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan Aequorea. *J. Cell. Comp. Physiol.* **59,** 223–239 (1962).
193. Ormö, M. *et al.* Crystal Structure of the Aequorea victoria Green Fluorescent Protein. *Science (80-.).* **273,** 1392–1395 (1996).
194. Dammeyer, T. & Tinnefeld, P. Engineered fluorescent proteins illuminate the bacterial periplasm. *Comput. Struct. Biotechnol. J.* **3,** e201210013 (2012).
195. Aronson, D. E., Costantini, L. M. & Snapp, E. L. Superfolder GFP Is Fluorescent in Oxidizing Environments When Targeted via the Sec Translocon. *Traffic* **12,** 543–548 (2011).
196. Dinh, T. & Bernhardt, T. G. Using superfolder green fluorescent protein for periplasmic protein localization studies. *J. Bacteriol.* **193,** 4984–4987 (2011).
197. Wingen, M. *et al.* The photophysics of LOV-based fluorescent proteins--new tools for cell biology. *Photochem Photobiol Sci* **13,** 875–883 (2014).
198. Mukherjee, A. *et al.* Engineering and characterization of new LOV-based fluorescent proteins from chlamydomonas reinhardtii and vaucheria frigida. *ACS Synth. Biol.* **4,** 371–377 (2015).
199. Drepper, T. *et al.* Reporter proteins for in vivo fluorescence without oxygen. *Nat Biotechnol* **25,** 443–445 (2007).
200. Mukherjee, A. & Schroeder, C. M. Flavin-based fluorescent proteins: Emerging paradigms in biological imaging. *Curr. Opin. Biotechnol.* **31,** (2015).
201. Christie, J. M. *et al.* Structural tuning of the fluorescent protein iLOV for improved photostability. *J. Biol. Chem.* **287,** 22295–22304 (2012).
202. Buckley, A. M., Petersen, J., Roe, A. J., Douce, G. R. & Christie, J. M. LOV-based reporters for fluorescence imaging. *Curr. Opin. Chem. Biol.* **27,** 39–45 (2015).

203. Christie, J. M., Salomon, M., Nozue, K., Wada, M. & Briggs, W. R. LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): Binding sites for the chromophore flavin mononucleotide. *Proc. Natl. Acad. Sci.* **96,** 8779–8783 (1999).
204. Mukherjee, A., Walker, J., Weyant, K. B. & Schroeder, C. M. Characterization of Flavin-Based Fluorescent Proteins: An Emerging Class of Fluorescent Reporters. *PLoS One* **8,** (2013).
205. Rees, D. C. & Robertson, A. D. Some thermodynamic implications for the thermostability of proteins. *Protein Sci.* **10,** 1187–1194 (2001).
206. Pucci, F., Dhanani, M., Dehouck, Y. & Rooman, M. Protein Thermostability Prediction within Homologous Families Using Temperature- Dependent Statistical Potentials. *PLoS One* **9,** e91659 (2014).
207. Diamant, S. & Goloubinoff, P. Temperature-Controlled Activity of DnaK - DnaJ - GrpE Chaperones: Protein-Folding Arrest and Recovery during and after Heat Shock Depends on the Substrate Protein and the GrpE Concentration. *Biochemistry* **37,** 9688–9694 (1998).
208. Patrick, M. Plasmids 101: Green Fluorescent Protein (GFP). *Addgene's Blog* (2014). Available at: http://blog.addgene.org/plasmids-101-green-fluorescent-protein-gfp.
209. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173,** 33–38 (1996).
210. Heim, R., Cubitt, A. B. & Tsien, R. Y. Improved green fluorescence. *Nature* **373,** 663–664 (1995).
211. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prashert, D. C. Green Fluorescent Protein as a Marker for Gene Expression Author (s): Martin Chalfie , Yuan Tu , Ghia Euskirchen , William W . Ward , Douglas C . Prasher Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.or. *Science (80-.).* **263,** 802–805 (1994).
212. Crameri, a, Whitehorn, E. a, Tate, E. & Stemmer, W. P. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14,** 315–319 (1996).
213. Zhang, J., Campbell, R. E., Ting, A. Y. & Tsien, R. Y. Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* **3,** 906–918 (2002).
214. Christensen, U. *et al.* De-bugging and maximizing plant cytochrome P450 production in Escherichia coli with C-terminal GFP fusions. *Appl. Microbiol. Biotechnol.* **101,** 4103–4113 (2017).
215. Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64,** 2240–2246 (1998).
216. Teerawanichpan, P., Hoffman, T., Ashe, P., Datla, R. & Selvaraj, G. Investigations of combinations of mutations in the jellyfish green fluorescent protein (GFP) that afford brighter fluorescence, and use of a version (VisGreen) in plant, bacterial, and animal cells. *Biochim. Biophys. Acta - Gen. Subj.* **1770,** 1360–1368 (2007).
217. Drepper, T. *et al.* Flavin Mononucleotide-Based Fluorescent Reporter Proteins Outperform Green Fluorescent Protein-Like Proteins as Quantitative In Vivo Real-Time Reporters. *Appl. Environ. Microbiol.* **76,** 5990–5994 (2010).
218. Scholz, K. E. *et al.* Fusion of a flavin-based fluorescent protein to hydroxynitrile lyase from arabidopsis thaliana improves enzyme stability. *Appl. Environ. Microbiol.* **79,** 4727–4733 (2013).
219. Scholz, K. E. *et al.* Synthesis of Chiral Cyanohydrins by Recombinant Escherichia coli Cells in a Micro-Aqueous Reaction System. *Appl. Environ. Microbiol.* **78,** 5025–5027 (2012).
220. Drepper, T., Jaeger, K.-E. & Potzkei, J. USE OF FMN-BINDING FLUORESCENCE PROTEINS (FBFP) AS NEW TYPES OF SECRETION MARKERS. US 2014/0227698 Al (2014).
221. Shcherbakova, D. M., Shemetov, A. a., Kaberniuk, A. a. & Verkhusha, V. V. *Natural Photoreceptors as a Source of Fluorescent Proteins, Biosensors, and Optogenetic Tools*. *Annual review of biochemistry* **84,** (2014).
222. Buckley, A. M. *et al.* Lighting Up Clostridium Difficile: Reporting Gene Expression Using Fluorescent Lov Domains. *Sci. Rep.* **6,** 1–11 (2016).
223. Gawthorne, J. A. *et al.* Visualizing the Translocation and Localization of Bacterial Type III Effector Proteins by Using a Genetically Encoded Reporter System. **82,** 2700–2708 (2016).
224. Djordjevic, M. Redefining Escherichia coli σ70Promoter Elements: -15 Motif as a Complement of the -10 Motif. *J. Bacteriol.* **193,** 6305–6314 (2011).
225. 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.* **25,** 1203–10 (1997).
226. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189,** 113–130 (1986).
227. Novy, R. & Morris, B. Use of glucose to control basal expression in the pET System. *Innovations* **13,** 13–15 (2001).
228. Dubendorf, J. W. & Studier, F. W. Controlling basal expression in an inducible T7 expression system by

blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* **219,** 45–59 (1991).
229. Tabor, S. Expression Using the T7 RNA Polymerase/Promoter System. *Curr. Protoc. Mol. Biol.* 1–11 (2001). doi:10.1002/0471142727.mb1602s11
230. Wanner, B. L., Kodaira, R. & Neidhardt, F. C. Physiological Regulation of a Decontrolled *lac* Operon. *J. Bacteriol.* **130,** 212–222 (1977).
231. Lim, H. K., Lee, S. U., Chung, S. Il, Jung, K. H. & Seo, J. H. Induction of the T7 promoter using lactose for production of recombinant plasminogen kringle 1-3 in Escherichia coli. *J. Microbiol. Biotechnol.* **14,** 225–230 (2004).
232. Saida, F., Uzan, M., Odaert, B. & Bontems, F. Expression of Highly Toxic Genes in E. coli: Special Strategies and Genetic Tools. *Curr. Protein Pept. Sci.* **7,** 47–56 (2006).
233. Studier, F. W. in *Structural Genomics: General Applications, Methods in Molecular Biology* (ed. Chen, Y. W.) **1091,** 17–32 (Spinger Science+Business Media, 2014).
234. Gentz, R. & Bujard, H. Promoters recognized by Escherichia coli RNA polymerase selected by function : highly efficient promoters from bacteriophage T5. *J. Bacteriol.* **164,** 70–77 (1985).
235. Bujard, H. *et al.* A T5 promoter-based transcription-translation system for the analysis of proteins in Vitro and in Vivo. *Methods Enzymol.* **155,** 416–433 (1987).
236. Kawe, M., Horn, U. & Plückthun, A. Facile promoter deletion in Escherichia coli in response to leaky expression of very robust and benign proteins from common expression vectors. *Microb. Cell Fact.* **8,** 8 (2009).
237. QIAGEN. The QIAexpressionist 5th Edition. A handbook for high-level expression and purification of 6xHis-tagged proteins. *Qiagen GmbH, Düsseldorf, Germanys* 1–126 (2003).
238. Stüber, D., Matile, H. & Garotta, G. in *Immunological Methods* (ed. Lefkovits, I.) **IV,** 122–152 (Academic Press Inc., 1990).
239. Samuelson, J. in *Production of Membrane Proteins: Strategies for Expression and Isolation, 1* (ed. Robinson, A. S.) 11–35 (2011).
240. Wickstrum, J. R. & Egan, S. M. Amino Acid Contacts between Sigma 70 Domain 4 and the Transcription Activators RhaS and RhaR. *J. Bacteriol.* **186,** 6277–6285 (2004).
241. Brautaset, T., Lale, R. & Valla, S. Positively regulated bacterial expression systems. *Microb. Biotechnol.* **2,** 15–30 (2009).
242. 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).
243. Kelly, C. L. *et al.* Synthetic Chemical Inducers and Genetic Decoupling Enable Orthogonal Control of the *rhaBAD* Promoter. *ACS Synth. Biol.* **5,** 1136–1145 (2016).
244. Haldimann, A., Daniels, L. L. & Wanner, B. L. Use of New Methods for Construction of Tightly Regulated Arabinose and Rhamnose Promoter Fusions in Studies of the Escherichia coli Phosphate Regulon Use of New Methods for Construction of Tightly Regulated Arabinose and Rhamnose Promoter Fusions in Studi. (1998).
245. Egan, S. M. & Schleif, R. F. A Regulatory Cascade in the Induction of rhaBAD copy.pdf. *J. Mol. Biol.* **234,** 87–98 (1993).
246. Wickstrum, J. R., Santangelo, T. J. & Egan, S. M. Cyclic AMP Receptor Protein and RhaR Synergistically Activate Transcription from the L-Rhamnose-Responsive *rhaSR* promoter in *Escherichia coli*. *J. Bacteriol.* **187,** 6708–6718 (2005).
247. Shimizu, K. Regulation Systems of Bacteria such as Escherichia coli in Response to Nutrient Limitation and Environmental Stresses. *Metabolites* **4,** 1–35 (2013).
248. Giacalone, M. *et al.* Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques* **40,** 355–364 (2006).
249. Hjelm, A. *et al.* Tailoring Escherichia coli for the L-Rhamnose P BAD Promoter-Based Production of Membrane and Secretory Proteins. *ACS Synth. Biol.* acssynbio.6b00321 (2017). doi:10.1021/acssynbio.6b00321
250. Neylon, C. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: Library construction methods for directed evolution. *Nucleic Acids Res.* **32,** 1448–1459 (2004).
251. Wang, Y.-C. & Distefano, M. D. Synthesis and screening of peptide libraries with free C-termini. *Curr Top Pept Protein Res* **15,** 1–23 (2014).
252. Simon, E. & Kornitzer, D. *Pulse-chase analysis to measure protein degradation*. *Methods in Enzymology* **536,** (Elsevier Inc., 2014).
253. Rüdiger, S., Germeroth, L., Schneider-Mergener, J. & Bukau, B. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* **16,** 1501–1507 (1997).

254. 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).
255. Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **110,** 14024–14029 (2013).
256. Bonde, M. T. *et al.* Predictable tuning of protein expression in bacteria. *Nat. Methods* **13,** 233–236 (2016).
257. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (2013).
258. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **96,** 23–28 (1990).
259. BERTANI, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J. Bacteriol.* **62,** 293–300 (1951).
260. Bertani, G. Lysogeny at mid-twentieth century:P1, P2, and other experimental systems. *J Bacteriol* **186,** 595–600 (2004).
261. Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T. & Nørholm, M. H. H. Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth. Biol.* **4,** 1042–1046 (2015).
262. Christie, J. M. LOV domains as in vivo fluorescent reporters of protein expression. *Curr. Protoc. Essent. Lab. Tech.* **2012,** 1–11 (2012).
263. Grote, A. *et al.* JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **33,** 526–531 (2005).
264. King, W. L. & Hurst, A. A Note on the Survival of some Bacteria in Different Diluents. *J. Appl. Bacteriol.* **26,** 504–506 (1963).
265. Massey-Gendel, E. *et al.* Genetic selection system for improving recombinant membrane protein expression in E. coli. *Protein Sci.* **18,** 372–383 (2009).
266. Wagner, S., Bader, M. L., Drew, D. & de Gier, J.-W. Rationalizing membrane protein overexpression. *Trends Biotechnol.* **24,** 364–371 (2006).
267. 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,** 1–10 (2013).
268. Stoker, N. G., Fairweathe, N. F. & Spratt, B. G. Versatile low-copy-number plasmid vectors for cloning in Escherichia coli. *Gene* **18,** 335–341 (1982).
269. Løbner-Olesen, a. Distribution of minichromosomes in individual Escherichia coli cells: implications for replication control. *EMBO J.* **18,** 1712–1721 (1999).
270. Krogsgaard, C. Detection of triggered membrane protein degradation in Escherichia coli. (2017).
271. 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).
272. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* **5,** 1–17 (2014).
273. Suter-Crazzolara, C. & Unsicker, K. Improved Expression of Toxic Proteins in E. coli. *Biotechniques* **19,** 202–204 (1995).
274. Sasmor, H. M. & Betz, J. L. Symmetric lac operator derivatives: effects of half-operator sequence and spacing on repressor affinity. *Gene* **89,** 1–6 (1990).
275. Borgeaud, S. & Blokesch, M. Overexpression of the tcp Gene Cluster Using the T7 RNA Polymerase/Promoter System and Natural Transformation-Mediated Genetic Engineering of Vibrio cholerae. *PLoS One* **8,** 1–10 (2013).
276. Grossman, T. H., Kawasaki, E. S., Punreddy, S. R. & Osburne, M. S. Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene* **209,** 95–103 (1998).
277. Solé, M., Rius, N. & Lorén, J. Rapid extracellular acidification induced by glucose metabolism in non-proliferating cells of Serratia marcescens. *Int. Microbiol.* **3,** 39–43 (2000).
278. Chen, Y. *et al.* Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. *Nat. Commun.* **9,** 1–8 (2018).
279. Šeputiene, V., Daugelavičius, A., Sužiedelis, K. & Sužiedeliene, E. Acid response of exponentially growing Escherichia coli K-12. *Microbiol. Res.* **161,** 65–74 (2006).
280. Boronat, A. & Aguilar, J. Rhamnose-induced propanediol oxidoreductase in Escherichia coli: Purification, properties, and comparison with the fucose-induced enzyme. *J. Bacteriol.* **140,** 320–326 (1979).
281. Muiry, J. a *et al.* Proton-linked L-rhamnose transport, and its comparison with L-fucose transport in Enterobacteriaceae. *Biochem. J.* **290 (Pt 3,** 833–842 (1993).

282. Møller, T., Franch, T., Udesen, C., Gerdes, K. & Valentin-Hansen, P. Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. *Genes Dev.* **16,** 1696–1706 (2002).
283. Müller-Hill, B., Rickenberg, H. V. & Wallenfels, K. Specificity of the induction of the enzymes of the lac operon in Escherichia coli. *J. Mol. Biol.* **10,** 303–318 (1964).
284. Monod, J. & Cohen-Bazire, G. Sur la biothèse de la beta-galactosidase (lactase) chez Escherichia coli. La specificité de l'induction. *Biochem. Biophys. Acta* **7,** 585 (1951).
285. Daber, R., Stayrook, S., Rosenberg, A. & Lewis, M. Structural Analysis of Lac Repressor Bound to Allosteric Effectors. *J. Mol. Biol.* **370,** 609–619 (2007).
286. Barkley, M. D., Jobe, A., Bourgeois, S. & Riggs, A. D. Interaction of Effecting Ligands with Lac Repressor and Repressor-Operator Complex. *Biochemistry* **14,** 1700–1712 (1975).
287. Riggs, A. D., Newby, R. F. & Bourgeois, S. lac repressor-Operator interaction. II. Effect of galactosides and other ligands. *J. Mol. Biol.* **51,** 303–314 (1970).
288. Piekarski, T. *et al.* Genetic tools for the investigation of Roseobacter clade bacteria. *BMC Microbiol.* **9,** 265–276 (2009).
289. Benson, R., Meyer, R., Zaruba, M. & McKhann, G. Cellular autofluorescence Is it due to flavins? *J. Histochem. Cytochem.* **27,** 44–48 (1979).
290. Gawthorne, J. A. *et al.* Express Your LOV: An Engineered Flavoprotein as a Reporter for Protein Expression and Purification. *PLoS One* **7,** 1–6 (2012).
291. Dalbey, R. E., Wang, P. & Kuhn, A. Assembly of bacterial inner membrane proteins. *Annu. Rev. Biochem.* **80,** 161–187 (2011).
292. Drew, D., Fröderberg, L., Baars, L. & De Gier, J. W. L. Assembly and overexpression of membrane proteins in Escherichia coli. *Biochim. Biophys. Acta - Biomembr.* **1610,** 3–10 (2003).
293. Drew, D. E., von Heijne, G., Nordlund, P. & de Gier, J. W. L. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Lett.* **507,** 220–224 (2001).
294. Dvorak, P. *et al.* Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway. *Microb. Cell Fact.* **14,** 1–15 (2015).
295. Studier, F. W. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219,** 37–44 (1991).
296. Becktel, W. J. & Schellman, J. A. Protein stability curves. *Biopolymers* **26,** 1859–1877 (1987).
297. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
298. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat. Methods* **2,** 905–909 (2005).
299. Wachter, R. M., Elsliger, M. A., Kallio, K., Hanson, G. T. & Remington, S. J. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* **6,** 1267–1277 (1998).
300. Patterson, G., Day, R. N. & Piston, D. Fluorescence protein spectra. *J. Cell Sci.* **114,** 837–838 (1998).
301. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–607 (2005). doi:10.1385/1592598900
302. Eguchi, Y. *et al.* Angucycline antibiotic waldiomycin recognizes common structural motif conserved in bacterial histidine kinases. *J. Antibiot. (Tokyo).* **70,** 251–258 (2017).
303. Inobe, T. & Nukina, N. Rapamycin-induced oligomer formation system of FRB-FKBP fusion proteins. *J. Biosci. Bioeng.* **122,** 40–46 (2016).
304. Raivio, T. L. Regulation of Proteolysis in the Gram-Negative Bacterial. *J. Bacteriol.* **200,** e00639-17 (2018).
305. Vimberg, V., Tats, A., Remm, M. & Tenson, T. Translation initiation region sequence preferences in Escherichia coli. *BMC Mol. Biol.* **8,** 1–13 (2007).
306. Osterman, I. A., Evfratov, S. A., Sergiev, P. V. & Dontsova, O. A. Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res.* **41,** 474–486 (2013).
307. Phue, J. N., Sang, J. L., Trinh, L. & Shiloach, J. Modified Escherichia coli B (BL21), a superior producer of plasmid DNA compared with Escherichia coli K (DH5α). *Biotechnol. Bioeng.* **101,** 831–836 (2008).

APPENDIX AND SUPPORTING INFORMATION

1 FIGURES

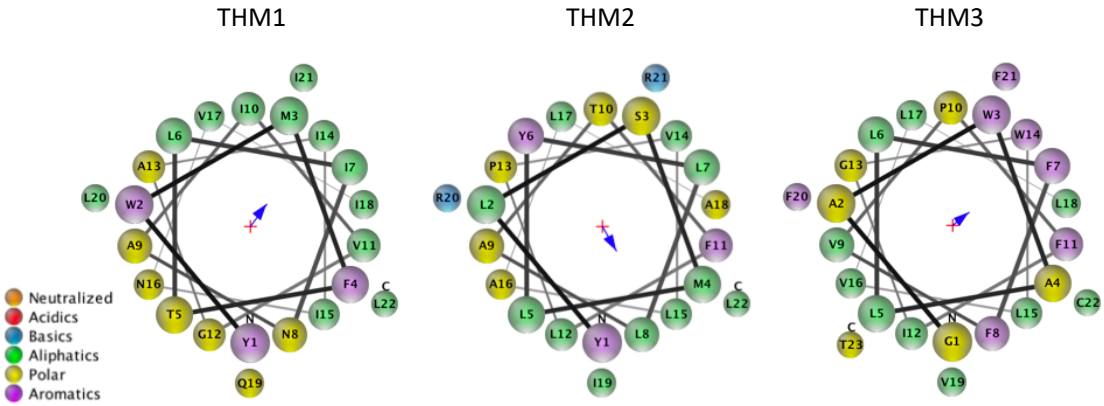


Figure S 1: **Helical wheel display and amino acids characteristics of the predicted transmembrane helices from YhaI**. Analysis was done with MPEx [185] version 3.2.15. Details in text.

A

```
# YhaI Length: 118
# YhaI Number of predicted TMHs:  3
# YhaI Exp number of AAs in TMHs: 63.53852
# YhaI Exp number, first 60 AAs:  33.34606
# YhaI Total prob of N-in:        0.94105
# YhaI POSSIBLE N-term signal sequence
YhaI    TMHMM2.0       inside      1    22
YhaI    TMHMM2.0       TMhelix    23    45
YhaI    TMHMM2.0       outside    46    48
YhaI    TMHMM2.0       TMhelix    49    71
YhaI    TMHMM2.0       inside     72    77
YhaI    TMHMM2.0       TMhelix    78   100
YhaI    TMHMM2.0       outside   101   118
```

B

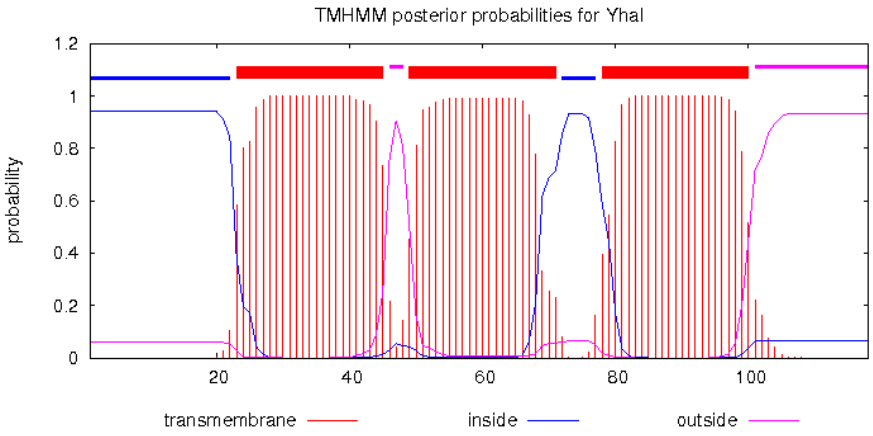


Figure S 2: **Prediction of transmembrane helices and their topology in YhaI.** The analysis was performed by using TMHMM Server v. 2.0. (A) Statistics and a list of the location of the predicted TMHs and the predicted location of the intervening loop regions. (B) Plot illustrates the probable topology of the TMHs inside the membrane releative to the cytosol (inside).

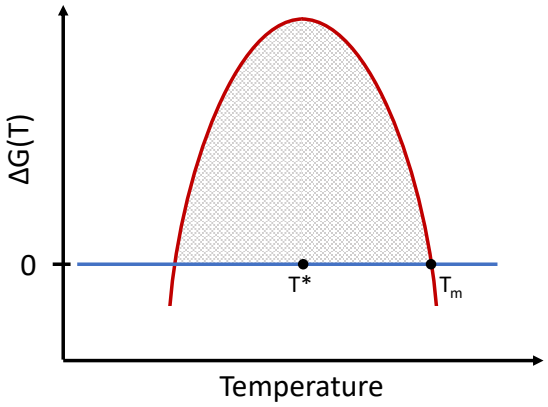


Figure S 3: **Schematic representation of the stability curve illustrating the temperature dependence of the free energy of unfolding (ΔG)**. The temperatures of maximal stability (T*) and heat denaturation (T_m), are indicated. Figure is modified from REES AND ROBERTSON (2001) [205], BECKTEL AND SCHELLMAN (1987) [296].

```
pCOLA_T7p-1_laco      GGAATTGTGAGCGGATAACAATTCC    25
pCOLA_T7p-2_laco      GGAATTGTGAGCGGATAACAATTCC    25
T5p_laco-2            TCAATTGTGAGCGGATAACAATTTC    25
T5p_laco-1            TGCTTTGTGAGCGGATAACAATTAT    25
                        ******************
```

Figure S 4: **Multiple sequence alignment of symmetric *lacIo* sites in the two *T7* promoters on pCOLA and the two in *T5p***. The multiple sequence alignment was performed with the Clustal Omega tool (EMBL-EBI). The perfect palindromic *lac* operator sequences are underlined. The nucleotide which represents the center of symmetry is in bold. Optimal binding site for LacI according to SAMOR AND BETZ (1990) [274] are highlighted in grey.

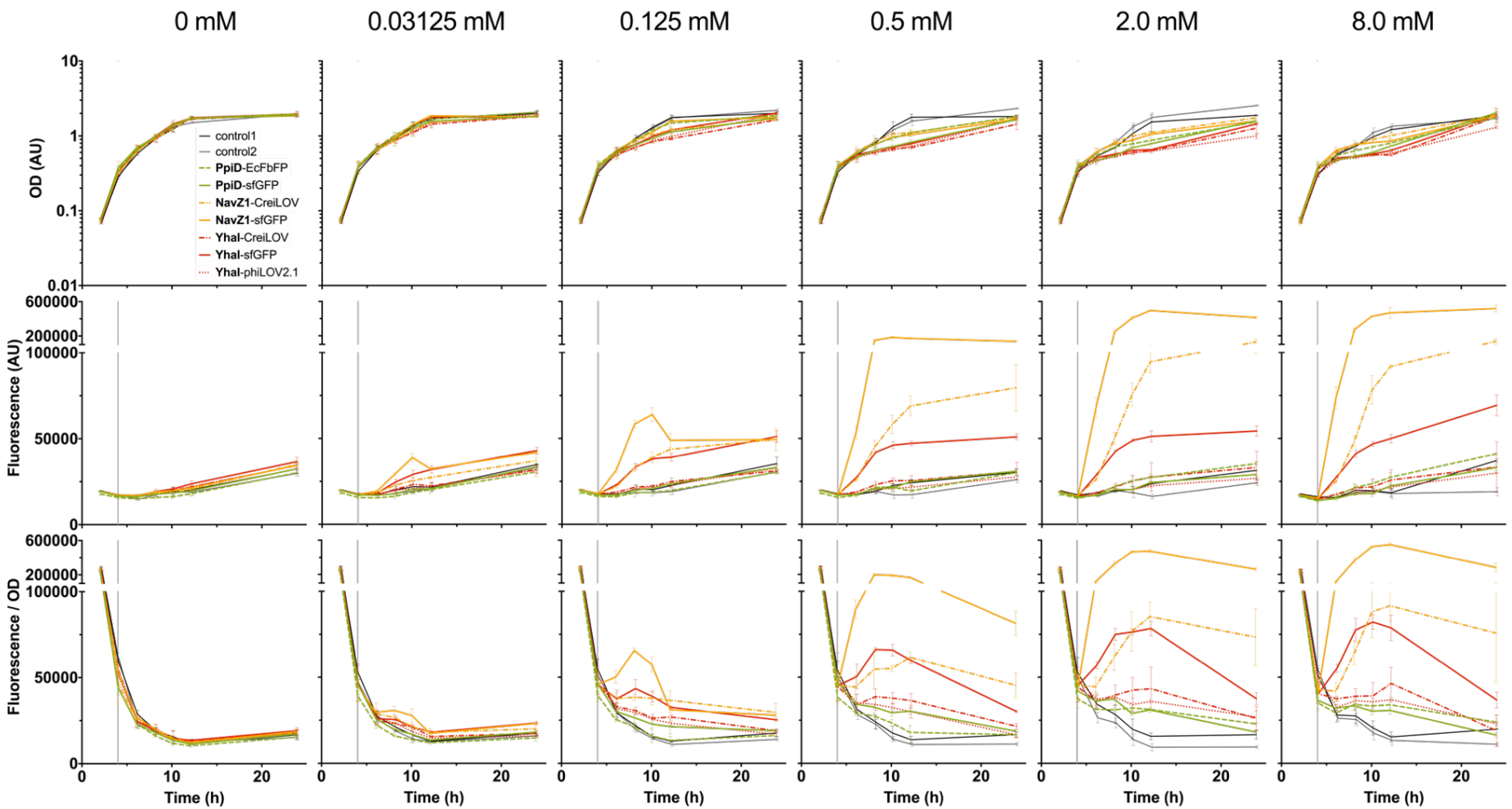


Figure S 5: **Optimising the *T7* promoter strength with lactose for rsTEVP(M)-dgFtsH-MP-FR-His production from pCOLA by taking FR fluorescence as indicator for functional IMP integration in *E. coli* K-12 MG1655 (DE3)**. The colour codes for the IMP-FR constructs are indicated in the legend. The line at 4 h represents the time point of IPTG supply (or MQ water). The final lactose concentrations are specified. 8 mM lactose corresponds to 0.28 % (w/v). For comparison, Studier (2005) suggested, besides 0.05 % (w/v) glucose, 0.2 % (w/v) lactose for usage in auto-inducing media [297]. $r_B = 3$, $r_T = 1$, n = 1.

2 TABLES

Table S 1: Overview about the oligonucleotides used in Study II.

oMSB#	SEQUENCE
2951	ACTTCCAGAUGGAAATTTACGAGAAC
2953	ACTTCCAGAUGGAAATTTACGAGAACGAAAACGACCAGGTAGAAGCGGTTAAACGCTTTTTTGCTGCGTCGTTCCAGTCGTTC
3045	AAGCGTTUAACCGCTTCTAC
3053	ACCTTCGUTGTTGCTGAC
3054	ATCGAAAAAUCTTTCGTTATCACCGAC
3055	AAGGCGCUGGATCACG
3056	AAACGCTUTTTTGCTGTGCTCAAGATTATTTTCGGTATCATTATC
3057	AGCGCCTUCTTGATGACTTCCAGCTGGCCCGGGATGCCGAACGACTGGAACGACGCTTGCTGTTCCAGCGCATC
3058	AAACGCTUTTTTGCTGGTCGCGCACG
3059	ACGAAGGUGTGACGCAGACCAGCATTTGAACCAAACTTCGGATCG
3060	ATTTTTCGAUATTTGAACCAAACTTCGGATCGT
3061	AAACGCTUTTTTGCTCCGCTCACCCTG
3062	ACGAAGGUGTGACGCAGACCAGCCCCTTCTTTTGTCGTGC
3063	ATTTTTCGAUCCCTTCTTTTGTCGTGCC
3151	AAACGCTUTTTTGCTATGAAAATCGAAGAAGGTAAACTGG
3152	ATCCAGGUACCGAAGTCTGCGCGTCTTTC
3257	ATGCGACUCCTGCATTAGGAAATATGGTGGAAAACCTGTACTTC
3258	AGCCTAGGUTAATCAGTGGTGGTGGTGGTGATGATGATGGGCCGCAAGCTTTTTGTAGAGCTCATCCATGCC
3268	ACCACCAUTAAACUAGTCTTGGACTCCTGTTG
3269	ATGGTGGUGGTGATGATGATGGTGACCGCCAGAACCGCCAGAACCGCCAGAACCGCCATTCATGAGTTGAGTCGCTTC
3288	ACCACTGAUTAACCTAGGCTG
3289	ATCAGTGGUGGTGGTGGTGATGATGATGGGCCGCAAGCTTAGCCAGAGCTTTACCTTCGG
3290	ATCAGTGGUGGTGGTGGTGATGATGATGGGCCGCAAGCTTAACGTGGTCAGAACCAACC
3291	ATCAGTGGUGGTGGTGGTGATGATGATGGGCCGCAAGCTTCAGCTTTTCATATTCCTTCTGCTTGG
3292	ACTTCCAGAUGGAAATTTACGAGAACGAAAACGACCAGGTAGAAGCGGTTAAACGCTTTTTTGCTCGTAAAGGCGAAGAGCTG
3293	ATCAGTGGUGGTGGTGGTGATGATGATGGGCCGCAAGCTTTTTGTACAGTTCATCCATACCATG
3294	ATCTGGAAGUACAGGTTTTCCACCATCT
3332	AGAGCTGTUCACTGGTGTC
3333	AACAGCTCUTCGCCTTTACGTTGCTGTTCCAGCGCATCG
3334	AACAGCTCUTCGCCTTTACGCCCTTCTTTTGTCGTGCC
3335	AACAGCTCUTCGCCTTTACGATTTGAACCAAACTTCGGATCG
3336	AGGGGAATUGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCCTAATGCAGGAGTCG
3337	AATTCCCCUCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGACTCGAGGATGGTGGAAAACCTGTACTTC

Table S 2: **Fluorescence properties of selected FR proteins**. Two numbers separated by a dash indicate a range of estimates from different authors working under slightly different conditions. Maximum excitation and emission wave length, λ_{Ex} and λ_{Ex}, respectively; extinction coefficient, E; quantum yield, QY. Table supplemented and modified from WINGEN *et al.* (2014) [197], BUCKLEY *et al.* (2015) [202] and THORN (2017) [188] with supplements from other sources.

FLUORESCENT PROTEIN	ORGANISM	MOLECULAR MASS (kDa)	OLIGOMERIC STATE	λ_{Ex} (nm)	λ_{Em} (nm)	E ($mM^{-1}*cm^{-1}$)	QY	BRIGHTNESS E*QY	PHOTO-STABILITY $t_{50\%}$ (min)	COMMENTS	REF
*Av*GFP	*A. victoria*	26.8	Weak dimer	395-397	504-508	25-30	0.79	19.8-23.7		Native protein from *A. victoria*	163,166
eGFP	*A. victoria*	26.8	Weak dimer	488-489	507-509	56	0.6	33.6	2.9	*Av*GFP variant	166,298
YFP	*A. victoria*	26.8	Weak dimer	514 (2^{nd} peak)	528					*Av*GFP variant	299
eBFP	*A. victoria*	26.8	Weak dimer	383	445	31	0.25	7.75		*Av*GFP variant	199,300
eCFP	*A. victoria*	26.8	Weak dimer	434	477	26	0.40	10.4		*Av*GFP variant	199,300
sfGFP	*A. victoria*	26.8	Monomer	485	510	83	0.65	54.1		*Av*GFP variant	6
EcFbFP	*B. subtilis*	15.1	Dimer	448	495	12.5-13.9	0.34-0.44	4.3-6.1	2.8	BsFbFP codon-optimised for *E. coli*; thermostable until 50 °C	197–199,204
phiLOV2.1	*A. thaliana*	12.1	Monomer	447-450	496-497		0.2		13	iLOV variant with increased photostability and enhanced photobleaching resistance	197,201,221
CreiLOV	*C. reinhardtii*	13.0	Monomer	450	495	12	0.51	6.4		Enhanced brightness, photostability, thermal and pH tolerance *in vitro*	198

Table S 3: **Overview about selected characteristics of the unmodified proteins relevant for Study II**. Theoretical average MW calculated with the ExPASy Compute pI/Mw tool; If not found in literature the theoretical molecular weight (MW) was calculated with ExPASy Compute pI/MW tool [301]; in regards to membrane protein topology, the orientation of the peptide chain N- and C-termini are given relative to the cytosol.

PROTEIN	SIZE (aa)	MW (kDa)	UNIPROT ID	TMHS	PROTEIN CLASS	MP TOPOLOGY	RCSB PDB ID	GENE	ECOGENE ID	ORGANISM	REFERENCE
NarX	598	67.1	P0AFA2	2	Membran protein	Polytopic, N_{in}-C_{in}	3EZH, 3EZI (partially)	*narX*	EG10646	*E. coli*	[5,180,302]
EnvZ	450	50.3	P0AEJ4	2	Membran protein	Polytopic, N_{in}-C_{in}	1BXD, 5XGA (partially)	*envZ*	EG10269	*E. coli*	[180]
NavZ1-W0	467	51.0	-	2	Membran protein	Polytopic, N_{in}-C_{in}		*navZ1-w0*	-	Gene chimera of *narX* and *enzZ* (*E. coli*)	[5]
YhaI	118	13.5	P64592	3?	Membran protein	Polytopic, N_{in}-C_{in}		*yhaI*	EG12748	*E. coli*	[180]
PpiD	623	68.1	P0ADY1	1	Membran protein	Bitopic, N_{in}-C_{out}	2KGJ (partially)	*ppiD*	EG13249	*E. coli*	[171]
YfgM	206	22.2	P76576	1	Membran protein	Bitopic, N_{in}-C_{out}		yfgM	EG14209	*E. coli*	[2]
*Av*GFP	238	26.9	P42212	-	Globular protein	-	1GFL	*gfp*	-	*A. victoria*	[166]
frGFP	238	26.8	-	-	Globular protein	-		fr*gfp*	-	*A. victoria*	[4]
sfGFP	238	26.8	-	-	Globular protein	-	2B3P	sf*gfp*	-	*A. victoria*	[6]
TEVP	237	27.0	P04517 (NIa)	-	Globular domain	-	1LVM (TEVP-S219D)	*tev*	-	Tobacco etch virus	[150]
phiLOV2.1	110	12.7	A0A178URG7 (Phot2)	-	Globular domain	-	4EEF		-	*A. thaliana*	[201,202]
EcFbFP	135	15.4	O34627 (YtvA)	-	Globular domain	-	2PR5 (BsYtvA)		-	*B. subtilis*	[199,220]
CreiLOV	119	12.9	Q8LPE0 (Phot)	-	Globular domain	-	1N9L		-	*C. reinhardii*	[198]

Table S 4: **Predicted transmembrane helices of POIs used in Study II**. Predictions are a result from a full protein scan using ΔGpred [165]. In regards to membrane protein topology, the orientation of the peptide chain N- and C-termini are given relative to the cytosol.

PROTEIN	POSITION	LENGTH	PREDICTED ΔG	SEQUENCE	COMMENT
YfgM					N_{in}-C_{out} topology.
	23-42	20	-0.709	KALAVGVILGVGALIGWRYW	
PpiD					N_{in}-C_{out} topology.
	12-34	23	-2.221	LVLKIIFGIIIVSFILTGVSGYL	
YhaI					N_{in}-C_{in} topology.
	23-44	22	0.283	YWMFTLINAIVGAIINVIQLIL	In theory, TMH1 is unlikely to exist but not impossible. Amphiphatic character possible.
	50-71	22	-1.407	YLSMLYLLATFLPVLALAIRRL	
	78-100	23	-3.187	GAWALLFFVPFIGWLVLLVFFCT	
NarX					N_{in}-C_{in} topology.
	9-31	23	-0.225	LTLVNQVALIVLLSTAIGLAGMA	
	157-179	23	-4.020	RVMAVFMALLLVFTIIWLRARLL	
EnvZ					N_{in}-C_{in} topology.
	12-34	23	-2.928	FARTLLLIVTLLFASLVTTYLVV	
	160-181	22	-3.239	LFRYTLAIMLLAIGGAWLFIRI	
	395-413	19	2.926	SARTISGTGLGLAIVQRIV	In theory, TMH3 is unlikely to exist but not impossible. Practically it is not existent.
NavZ1-W0					N_{in}-C_{in} topology.
	9-31	23	-0.225	LTLVNQVALIVLLSTAIGLAGMA	
	157-179	23	-4.020	RVMAVFMALLLVFTIIWLRARLL	
	391-409	19	2.926	SARTISGTGLGLAIVQRIV	In theory, TMH3 is unlikely to exist but not impossible. Practically it is not existent.

3 SUPPLEMENTARY MATERIAL OF PUBLICATION I

See next page.

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

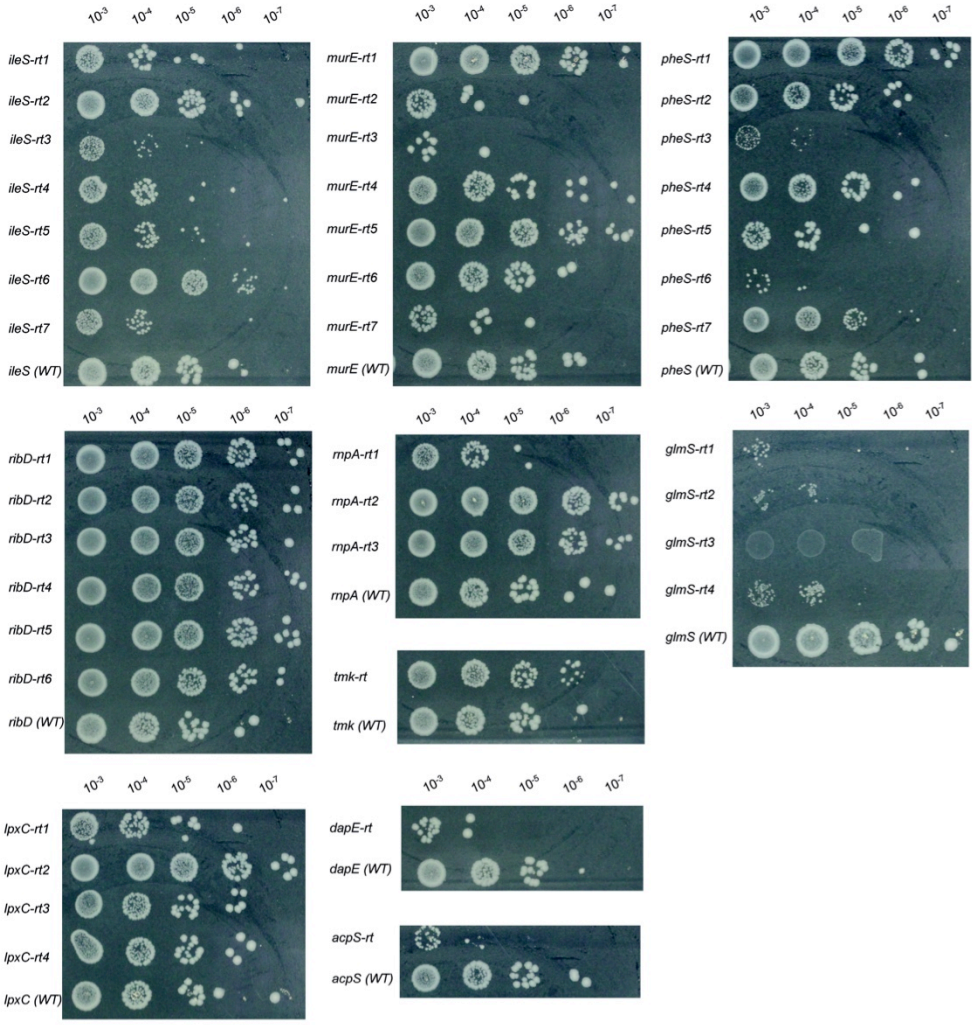


Figure S1. Growth drop tests of isolated variants of ten essential genes tagged with TIR variation. 10-fold serial dilutions were performed from overnight cultures and 10 µL were plated on LB agar. *E. coli* K12 MG1655 serves as control strain (WT).

Figure S2

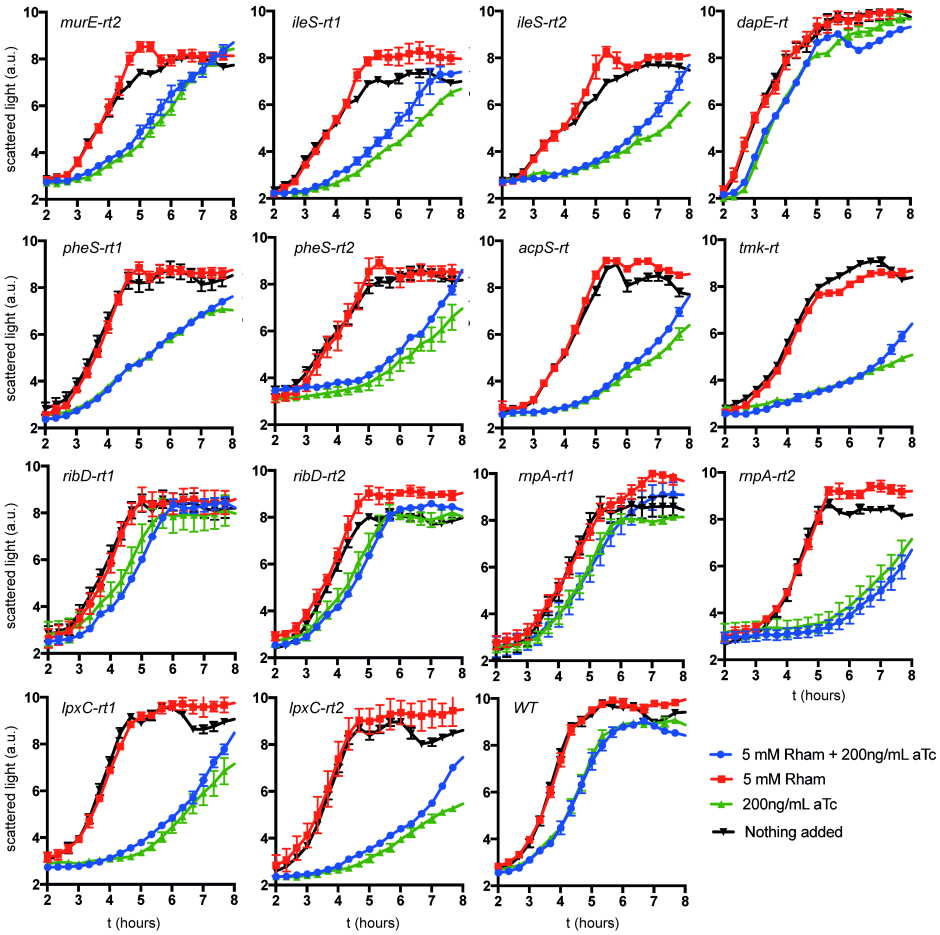


Figure S2. Growth profiles of wild type *E. coli* K12 MG1655 and 14 different tagged essential gene variants when inducing CRiPi with aTc, rhamnose or both. Growth was monitored over time in a BioLector® where scattered light represents biomass concentration. Rhamnose (red) aTc (green), or rhamnose and 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

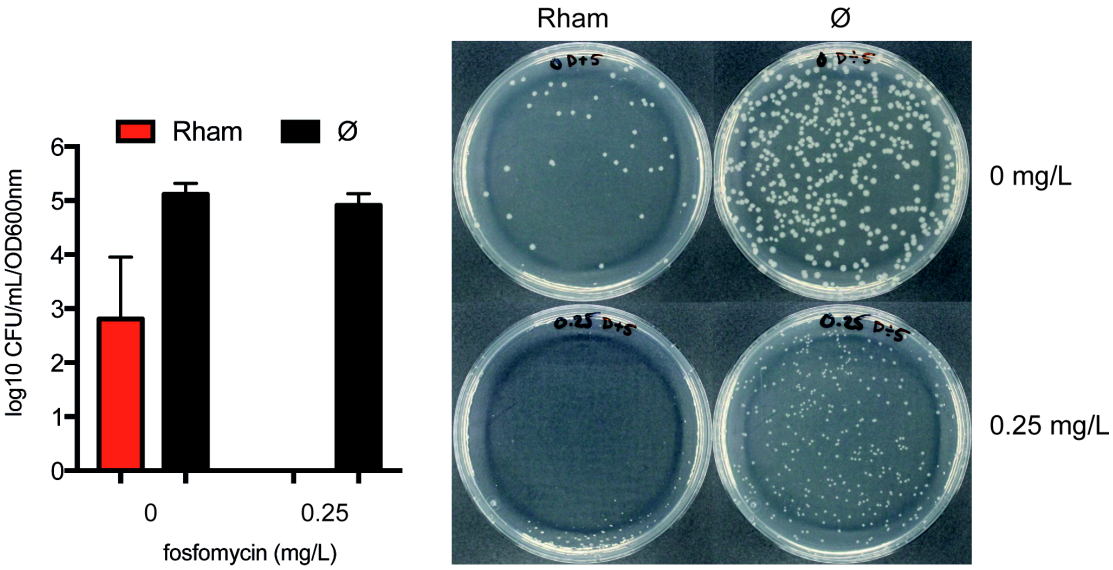


Figure S3. Fosfomycin sensitivity upon PROTi induction in *murE-rt*1-tagged clone or no induction (Ø). Cells carrying the CRiPi system were grown for 4 hours and plated on LB agar with different fosfomycin concentrations. All values are the averages of five biological replicates and bars show standard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

Figure S4

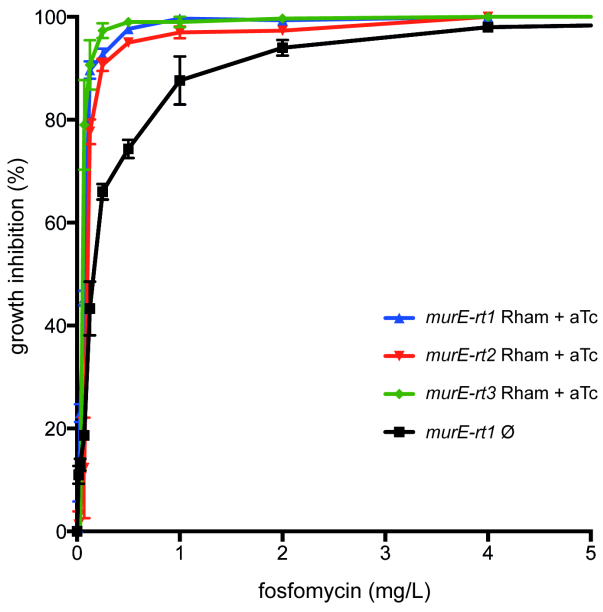


Figure S4. Fosfomycin sensitivity upon CRiPi induction in *murE*-tagged clones or no induction (Ø). Growth was tested in increasing concentrations of fosfomycin after 18 h incubation. All values are the averages of three biological replicates and bars show standard error. Same data as in Figure 3I.

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 tgcgggggggt gccgcaatga gctgttgcgt acccccccttt 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

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 tccttttttcg 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.