Overcoming Long-Standing Challenges in Recombinant Protein Expression

by

Romel I. Menacho Melgar

Department of Biomedical Engineering
Duke University

Date:_____

Approved:

_____
Michael D. Lynch, Advisor

_____
Michael R. Tadross

_____
Fan Yuan

_____
Joel Collier

_____
Michael C. Fitzgerald

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biomedical Engineering in the Graduate School
of Duke University

2020

ABSTRACT

Overcoming Long-Standing Challenges in Recombinant Protein Expression

by

Romel I. Menacho Melgar

Department of Biomedical Engineering
Duke University

Date:_____
Approved:

_____
Michael D. Lynch, Advisor

_____
Michael R. Tadross

_____
Fan Yuan

_____
Joel Collier

_____
Michael C. Fitzgerald

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biomedical Engineering in the Graduate School of
Duke University

2020

Abstract

The market for proteins has been constantly increasing for many different applications, such as protein-based drugs. However, protein production in *E. coli*, the most widely used recombinant host, has remained fundamentally the same since its inception. Here, we developed from the bottom up, a new expression platform using synthetic biology. The resulting platform reaches high protein titer across scales and allows for autoinduction and autolysis. More importantly, we are able to decouple growth from protein production and show that this platform is easily adaptable to other applications such as (i) reduction of essential housekeeping proteases to enable production of "hard-to-express" proteins and (ii) complete *in vivo* post-translational modifications.

Contents

List of Tables

List of Figures

Acknowledgements

I would first like to thank my former and current lab mates. I treasure all the good times we have had as well as the many things you have taught me during these past seven years. Specially, I would like to thank Adim whose friendship I find invaluable. Your jokes have made hard times a lot easier to bear and your work ethic has pushed me to become a better scientist.

I would also like to thank my advisor, Michael Lynch, for the tremendous amount of time that you have invested in mentoring me. Thanks to you, I am now more able to make strategic decisions to move forward, not only in the lab, but in my decisions to pursue my career goals. I have truly valued every piece of advice you have given me.

También quiero agradecer a mis amigos en Perú, Jesús, Eduardo, Edwin, Álvaro y Luis, y a mi primo Listter. Me hacen reír constantemente y verlos es una de las cosas que más espero todos los años. A mi mamá, por enseñarme a meterle empuje. A mi padre, por enseñarme a reír en malos momentos. A mi otra madre, Tere, por enseñarme humildad. A mi hermana, en cuyos recuerdos de infancia encuentro refugio.

Gracias a mi esposa que siempre tiene un abrazo para mí cuando lo necesito a cambio de un vaso de agua. Gracias por tenerme paciencia y aprender de mí, así como yo aprendo de ti. Y en especial, a mi hijo. Me alegras el día cada vez que sonríes impulsivamente al verme. Éstas últimas semanas han sido bastante complicadas y me habrían costado mucho más sin tus sonrisas.

1. Introduction

1.1 Introduction

Proteins including industrial enzymes, laboratory reagents and pharmaceuticals, make up the bulk of biotechnological products. Industrial enzymes are used in a wide variety of applications from chemicals manufacturing to food processing. As drugs, they offer high specificities due to their intricate structures, and are being used to treat a myriad of diseases with lower safety concerns than small molecule drugs. (Fosgerau & Hoffmann, 2015) The market for biopharmaceuticals alone is over $300 billion and continues to grow. (Ade et al., 2019)

There are multiple cellular hosts routinely used for the production of recombinant protein. Perhaps, the most widely used is still the Gram negative bacteria *Escherichia coli*. *E. coli* is a widely studied laboratory work horse, with numerous tools available for genetic manipulation. Additional advantages of *E. coli* are its fast growth rate (with a doubling time of about 20 minutes) even in inexpensive minimal media and its ability to grow to high cell densities (in theory, up to 200 gCDW)(Märkl et al., 1993). *E. coli* is currently used to produce about 30% of FDA-approved protein drugs (Baeshen et al., 2015) and about 70% of proteins in a research setting. (Bill, 2014) Currently, the most popular *E. coli* strains are BL21(DE) and derivatives which are protease deficient and express the highly processive and orthogonal T7 RNA polymerase. (F. W. Studier & Moffatt, 1986)

However, *E. coli* cannot naturally perform many post-translational modifications, such as glycosylation, which is beneficial for many pharmaceuticals. (Ding et al., 2017)

In addition, in drug applications, endotoxin must be removed from proteins produced in *E coli* increasing downstream purification costs. (Mamat et al., 2015) A lack of secretion capacity also limits applications where protein secretion reduces downstream costs, such as in the production of industrial enzymes. Heterologous proteins expressed *E. coli* can also be more prone to form inclusion bodies, or not be expressed at all. Many of these challenges can currently be overcome by an alternative expression host.

The Gram-positive *Bacillus subtilis* is routinely used to produce secreted proteins and unlike *E. coli*, does not produce endotoxin.(Simonen & Palva, 1993) Despite a better safety profile, *B. subtilis* is mostly used for industrial enzyme production, producing ~ 60% of commercial enzymes, (Contesini et al., 2018) where its secretory capacity is leveraged. However, like *E. coli*, *B subtilis* is unable to produce many "difficult to express" proteins and also cannot perform protein glycosylation.

Eukaryotic hosts such as yeasts are capable of glycosylation, although they do not exactly replicate human glycosylation patterns. Protein expression in yeast currently leverages two primary hosts: *Saccharomyces cerevisiae* and *Pichia pastoris*. Both strains are able to reach high cell densities (>100 gCDW) and grow fairly fast, with doubling times of ~90 minutes and 1-3 hours respectively. (Kastilan et al., 2017; Salari & Salari, 2017) Mammalian cells are another alternative for recombinant protein expression. Although they grow slowly (CHO and HEK cells have a doubling time of less than 24 h),(Hunter et al., 2019) and they are the most expensive and difficult to grow, mammalian hosts are able to glycosylate proteins in a human-like fashion. For that reason, mammalian cells account for the production of ~70% of biopharmaceuticals, most of

which are antibodies. (Tytgat et al., 2019) Unfortunately, most advances enabled by high throughput approaches have focused on microbial cells, which will be the focus of this review.

Despite the differences between these various expression hosts, they all share key commonalities. Fundamentally, heterologous expression requires promoters and vectors to drive transcription, the translation of mRNA and post translational processing of newly made proteins In this review, we will discuss recent advances in our fundamental understanding and engineering of these processes, as well as advances in host engineering to improve protein production. We will further discuss how high-throughput methodologies could further be applied to engineer better hosts to address current post-translational processing challenges.

1.2 High throughput studies on Transcription & Translation

The initial step in recombinant protein expression is often the design of an expression construct enabling heterologous transcription and translation. Understanding of design principles enabling efficient expression has grown dramatically over the past decade in numerous expression hosts. Codon bias (which refers to natural codon occurrence frequencies) and the stability of transcript secondary structures have long been thought to be strongly correlated with protein expression.(Behloul et al., 2017; Pellizza et al., 2018) In recent years, fundamentals underlying optimal transcription and translation have been investigated using newer high-throughput methodologies. These results, and the models they enable, have identified important variables as well as gaps in our mechanistic understanding of transcription and translation.

For any transcript to be well expressed optimal promoters are needed. Fortunately, various sets of promoters enabling heterologous expression in numerous hosts have been characterized, and while there is always ongoing work in developing new promoters, this area is relatively mature compared to others. We will refer the reader to the following reviews on promoters. (Blazeck & Alper, 2013; Cos et al., 2006; Terpe, 2006)

Methods to optimize specific components of expression constructs to obtain higher protein titers have also been described and include many computational methods. For example, computational methods utilizing first principles have been developed to predict and optimize ribosome binding sites for improved protein expression, with some predictive success. (Salis, 2011)) Computational frameworks aimed at optimizing translation have been complemented by numerous experimental approaches. For example, antibiotic resistance was coupled to higher protein expression to successfully select for improved translation initiation sequences, from a library of over 30,000 variants.(Rennig et al., 2018) Similarly, FACS and next-generation sequencing enabled the analysis of over 200,000 GFP reporter constructs to correlate sequence features with expression levels. These results show that the first elongation steps (codons 3-5) are key in determining translational efficiencies independently of codon content, mRNA structures or translation initiation.(Verma et al., 2019) Additionally, the expression of over 6000 genes from 171 different organisms in *E. coli*, demonstrated that codon content content (which refers to which codon is used in an expression construct, unrelated to codon occurrence frequencies) has a larger effect on translation than codon bias. Codon

bias only negatively affected expression for infrequently used codons AUA, CGG, CGA and CUA (with AGA contribution remaining inconclusive due to experimental constraints). On the other hand, stable secondary structures at the beginning of the transcript (first 16 codons) were confirmed to lead to lower levels of expression.(Boël et al., 2016) In another study, large scale DNA synthesis coupled with design of experiments enabled the thorough analysis of the expression of 244,000 different sequences. Although codon bias and transcript structure were again observed as the most important parameters, statistical modelling could only account for 50% of the experimental variance.(Cambray et al., 2018)

These results point to significant remaining unknown or poorly-described factors that affect protein expression.(Cambray et al., 2018) As a case in point, Gamble *et al*, using 35,000 GFP reporters investigated the effect of codon pairs, as opposed to single codons, on translation and found 17 codon pairs that are inhibitory to expression in *S. cerevisae*. (Gamble et al., 2016) A similar phenomenon has also been demonstrated in *E. coli*. (Navon et al., 2016) Additionally, to date, solutions to problems that may arise from long range relationships in coding sequences have so far been excluded from high throughout studies. For example, clusters of slowly translated codons, rather than individual codons, may have a larger impact on translation.(Alexaki et al., 2019) Similarly, as suggested by Verma *et. al.*, highly processive elongation followed by stalling may lead to ribosome collisions and activation of quality control mechanisms.(Verma et al., 2019) The nascent peptide chain can also modulate translation

rates, but this modulation has only been well-characterized for specific motifs. (Nakatogawa & Ito, 2002)

The high-throughput studies described above have been invaluable in our understanding of transcription and translation. As described in Figure 1b, in general these all agree that eliminating stable mRNA structures at the 5' of the transcript and hydrophilic amino acids favor high expression levels. And while they all agree that codon optimization is important, there is no consensus as to which is more important codon content or bias.

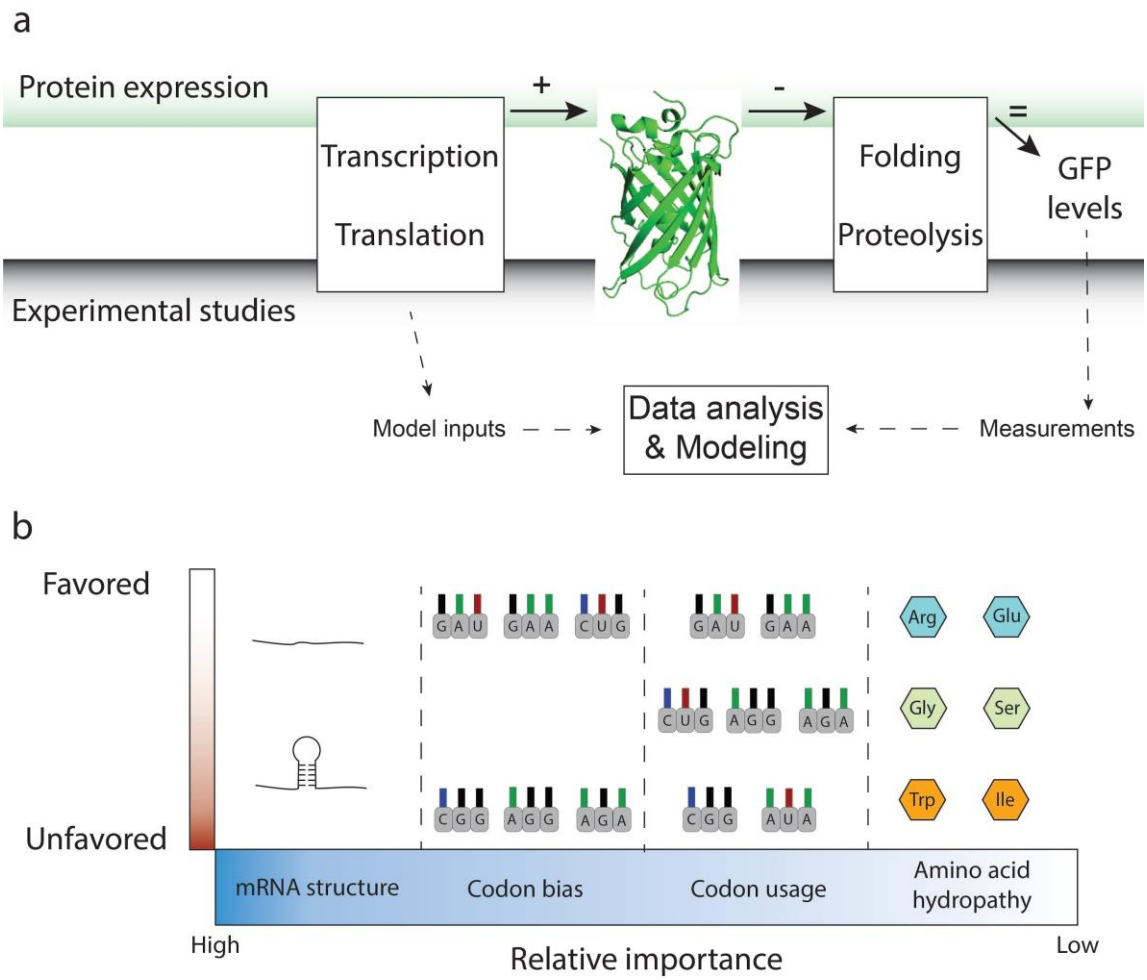

Figure 1: High-throughput studies on transcription and translation. These have (a) not addressed post-translational protein processing such as folding and proteolysis and only consider parameters related to transcription and translation. (b) Stability of mRNA structures are the most important factor in determining protein levels with unstable structures favored for high expression, followed by codon bias or usage. While controversy remains on how codons affect protein levels, there is consensus in their importance.

1.3 Current methodologies to improve heterologous protein production

It is important to note that the above studies have primarily focused on end-point protein abundances while post-translational processing was not investigated. As illustrated in Figure 1a, many times, key challenges to the high level expression of any particular protein occur after both transcription and translation have occurred. Post-

translational processing including: proper folding, proteolytic degradation, membrane insertion and modifications such as glycosylation can all impact final protein levels, protein solubility and quality The poor expression of certain protein classes can be due to issues in post translational processing. For example, poor expression of mammalian glycoenzymes in E. coli(Moremen et al., 2018) remains a bottleneck to a variety of novel prototype glycosylated biopharmaceuticals. In this case, E. coli is prefered over yeast strains because E. coli lacks endogenous glycosylation machinery whereas yeast can produce immunogenic glycan structures. (Laukens et al., 2020; Piirainen et al., 2016) While lack of progress has been ameliorated by cell-free expression systems, complex human-like glycosylation patterns are still not possible. (Kightlinger et al., 2019)

Current workflows to improve protein folding and solubility still often include brute force approaches including varying promoters, media and induction conditions, knocking-out specific proteases, the co-expression of different chaperones and the screening of solubility tags.(Gutiérrez-González et al., 2019; Reyes et al., 2017) Unfortunately, these efforts are only applicable to specific proteins. For example, efforts in E. coli have been mainly focused on reducing proteolysis, such as knocking out the periplasmic proteases Tsp and degP to improve production of Fabs. (Ellis et al., 2017) Likewise, some metabolic engineering strategies are dependent on the amino acid composition of the target protein. In E. coli, the coexpression of cysK, encoding cysteine synthase A, alleviated toxicity associated with the expression of serine rich proteins leading to increased production (M.-J. Han et al., 2003) and elevating the glycyl-tRNA

pools enabled increased expression of spider silk which is over 40% glycine. (Xia et al., 2010)

Also, more complicated to characterize is metabolic burden induced by heterologous protein expression; which can not only inhibit growth, but also limit protein levels. (REFs) Frumklin et. al. characterized gene features that minimize metabolic burden using ~14,000 GFP reporters and a competition assay coupled with deep-sequencing. This study showed that lower RNA levels do not necessarily produce less protein, but decrease metabolic burden (Frumkin et al., 2017) due to the metabolic cost of transcription. (Hausser et al., 2019) They found that gene features that reduced metabolic burden correlated with increased expression levels. These included hydrophilic amino acids and high initiation rates. (Frumkin et al., 2017)

Another important variable is cellular regulation, which has also been investigated with higher throughput methods. Using ribosome profiling in S. cerevisiae, Shiber et al. demonstrated that proper folding of proteins may not only depend on dedicated chaperones or proper translational rates, but on transient interactions with other proteins such as other members of the same protein complex or a metabolic pathway. (Shiber et al., 2018) Importantly, translation rates of an heterologous gene may be hard to be conserved in a protein production host as evidenced by enrichment of different amino acids at similar protein structures on different organisms, such as at the boundary of two secondary structures. (Saunders & Deane, 2010)

A couple of recent studies aimed at engineering post translational processing highlight how little has the genome of recombinant hosts have been explored to obtain

non-traditional and non-obvious ways to improve these protein levels. In E. coli, mutations in yidC, a gene involved in membrane insertion of membrane proteins, and hslV, a processive protease, led to improved expression of 3 mammalian membrane proteins. (Hatahet et al., 2015) In another study, deletion of non-essential ribosomal proteins in P. pastoris reduced the translation elongation speed causing an almost 3-fold improvement in expression of phytase, a protein prone to form aggregates due to a high content of beta β-sheet structures. (X. Liao et al., 2019) Importantly, extension of these methodologies to a broader panel of proteins has yet to be realized.

1.4 Discussion & Future Outlook

Lower costs in DNA and sequencing have enlarged the variants that can be tested for a particular experiment. However these have not been fully exploited to improve protein production as evidenced by *E coli*, which has fundamentally remained the same since the inception of T7-driven protein expression even though it is the most widely used host,. (Rosano et al., 2019) In fact, it is estimated that currently, less than 50% of bacterial proteins and less than 15% of non-bacterial proteins are expressible in *E. coli*. (T. Wang et al., 2018)

Advances to improve recombinant protein production are still focused on case-by-case optimization workflows. General methods, or engineered strains, that can be used to maximize success rates and titers of a wide range of proteins have not yet been developed, thoroughly characterized or widely adapted. Most new methods to optimize expression, still rely on screening of a few conditions(Hjelm et al., 2017) and are not

warranted to work since their applicability has not been fully described, limiting their generalized use.

A few exceptions where high-throughput methodologies were used to engineer recombinant protein hosts are worth noting. For example, in *E. coli*, using multiplexed RNA-seq, Ceroni et al identified genes that become upregulated with induction of heterologous protein expression. (Ceroni et al., 2018) With this information, a set of circuits that control protein expression upon sensing metabolic burden were designed and tested.(Ceroni et al., 2018) These circuits resulted in reduction of burden ranging from ~50% while maintaining about a third of protein expression to a ~10% reduction while maintaining 90% of protein expression. Next, in *S. cerevisiae*, a genome-wide transcriptional analysis of a group of strains with different levels of protein secretion capacity resulted in guidelines for future host engineering, including downregulation of redundant glycolytic enzymes and upregulation of genes involved in protein synthesis and trafficking.(M. Huang et al., 2017) These results were later corroborated using a high-throughput RNAi approach with over 240,000 knockdows to identify key genes that affect protein production in yeast. (G. Wang et al., 2019) A subset of these genes were then tuned to obtain a maximal 2.2-fold improvement in the expression of amylase. (G. Wang et al., 2019) Finally, in *P. pastoris*, a library of ~3000 different strains with rewired metabolic networks, found that downregulation of the metabolism of unfavorable nitrogen sources improved expression of GFP, insulin, an opiod receptor, spider silk protein and an immunotoxin. (Windram et al., 2017)

On the other hand, high-throughput studies on determining what factors are important to transcription and translation rates have been more numerous and have had more successful outcomes despite conflicts about what factors may be more important. Conflicting results may be due to differences in experimental design or lack of parameters that include post-translational protein processing. For example, Boel *et. al.* found that codon content, as opposed to codon bias, was an important parameter to determining protein titers, while Cambray *et. al.* found that codon bias was the more significant feature. Importantly, Boel *et. al.* was analyzing non-codon optimized sequences from different organisms while Cambray *et. al.* was analyzing smaller sequences at the N-terminus of a fluorescent reporter. Because of this, the sequences analyzed by Boel may have had a higher variation on protein sequences, structures and folding. Besides, folding is encoded in translation rates which are not expected to be conserved when inserting native DNA sequences to a recombinant host, as discussed above. (Saunders & Deane, 2010) Thus, it is not unlikely that, to some extent, both heterologous protein sequences and incorrect translation rates result in exposed protease-prone regions leading to protein degradation, and that the codon optimization by Boel *et. al.* may have favored codons that prevent proteolysis. One alternative to decouple transcription and translation from post-translational processes may be the use of cell-free expression systems, and even though the throughput may be limited due to cost, efforts to improve such systems are ongoing. (Villarreal et al., 2018)

A more viable alternative to better understand recombinant protein production may be studies that are focused on post-translational processing and host engineering

using high-throughput methodologies, or at the very least, that integrate these with transcription and translation studies. For example, genome-wide screening using dCas9 in *E. coli* have been used to identify essential genes (Rousset et al., 2018), but could also be used to identify genes that affect protein solubility. High-throughput methods that assay protein solubility using a split fluorescent reporter have been described and may be used in such studies. (Listwan et al., 2009) Another problem that can also be addressed is expression of membrane proteins. As discussed above, *E. coli* proteins involved in membrane insertion can affect the expression of membrane proteins(Hatahet et al., 2015) and high-throughput screening of metabolic networks may unveil other and more general mechanisms by which expression of this protein class can be improved. This can be coupled with electrophoresis capillary assays to measure expression of membrane proteins(Welner et al., 2017) or FACS if a fluorescent reporter is fused to the membrane proteins. These same high-throughput methods can be used to discover mechanisms to reduce proteolysis.

Future high-throughput studies may be further benefited from two-stage production methodologies that decouple growth from production. Two-stage approaches enable the implementation of engineering strategies that are difficult to implement in growing cells and reduce the potential impact of metabolic burden on protein expression.(Burg, Cooper, Ye, Reed, & Moreb, 2016; Michael D. Lynch, 2016) For example, simultaneous controlled inhibition of *E. coli* RNA polymerase with protein induction led to a 2-5 fold improvement in the titer of a couple of hard-to-express glycosyltransferases proteins.(Lemmerer et al., 2019) A different approach using

phosphate limitation to simultaneously stop growth and auto induce protein expression has also been described. (Menacho-Melgar, Moreb, et al., n.d.; Menacho-Melgar, Ye, et al., n.d.) Phosphate levels in the media are tuned to define different target biomass levels at which protein expression occurs, making this strategy readily scalable.(Menacho-Melgar, Ye, et al., n.d.) Using this approach, GFP titers of > 8g/L have been demonstrated with biomass levels of only 30 gCDW/L. Moreover, screening for improved hosts genotypes using two-stage production methodologies should avoid confounding effects that may arise from growth-associated production methodologies where improved protein titers are inevitably correlated with cell growth. Importantly, finding ways to control post-translational processing may be utilized to reduce so far unexplained variances in high-throughput transcription and translation studies so that a quantitative consensus on important parameters is reached. Finally, future studies aiming at engineering improved expression hosts should include various test proteins and as structurally diverse as feasible (within the scope of a particular study) to ensure the new methodologies are generalizable.

In summary, a more integrated study of protein expression with novel methodologies that couple high-throughput and aim at minimizing confounding effects while maximizing the variation studied will offer new mechanistic insights and offer new, improved, standard protein expression methodologies.

2. Scalable two-stage, autoinduction of recombinant protein expression in *E. coli* utilizing phosphate depletion

2.1 Introduction

Heterologous protein expression is a standard workflow common in numerous fields of biology and *E. coli* is the workhorse microbe for routine protein production in academia and industry.(McKenna et al., 2019; Robinson et al., 2015; Skretas et al., 2012; Zhou et al., 2016) *E. coli* based processes are used for the production of over 30% of protein based drugs that are on the market today, (Baeshen et al., 2015; Jozala et al., 2016; Sanchez-Garcia et al., 2016) and pET based expression in *E. coli* strain BL21(DE3) and its derivatives is a mainstay of heterologous expression in many labs.

Standard protocols rely on easily prepared media (LB and or TB) but require culture monitoring to optimize induction in exponential phase. (Glazyrina et al., 2010; Neidhardt et al., 1990) Auto-induction protocols removing the need for manual additions have been developed, most notably by Studier, and require the use of multiple carbon substrates, such as glucose and lactose. After glucose depletion the consumption of lactose induces heterologous expression.(F. William Studier, 2005, 2014) Significant recent work has been done in developing new protocols enabling auto-induction systems focused on using novel auto inducing promoters that respond to a variety of signals from cell density to oxygen limitation.(Anilionyte et al., 2018; Baez et al., 2014; Ben et al., 2016; Briand et al., 2016; Nocadello & Swennen, 2012) Despite simplifying expression protocols, many of these approaches still result in relatively low biomass and protein levels and have not been validated in multiple culture systems including instrumented

bioreactors. Scale up of protein expression to higher cell density fermentations remains a nontrivial task. The use of BL21 and its derivatives can be further complicated by heterogenous induction, resulting from lactose based inducers (Khlebnikov et al., 2000; Labhsetwar et al., 2013; Novick & Weiner, 1957), as well as the accumulation of acetic acid in fermentations with excess carbon source, which can have toxic effects on both cell growth and protein expression. (De Mey et al., 2007; Eiteman & Altman, 2006; Shiloach & Fass, 2005; H. Wang et al., 2014; Wong et al., 2008)

There remains a need for auto-inducible protein expression methods with tightly controlled expression, minimal overflow metabolism, and a high level of protein expression. Ideally new methods will be adaptable to numerous workflows and culture volumes, from high throughput screening approaches in microtiter plates and be readily scalable to higher cell densities in larger instrumented bioreactors, in commercially relevant media.

We report the development of a facile protocol for the routine high level expression of proteins. The method relies on a promoter that is induced by phosphate depletion, where protein expression is induced at the entry into stationary phase. While the expression of heterologous proteins during stationary phase may seem counterintuitive and at odds with maximal production, stationary phase cells can maintain significant metabolic activity and produce high levels of protein.(Burg, Cooper, Ye, Reed, Moreb, et al., 2016; Chubukov & Sauer, 2014; M. D. Lynch, 2011; Michael D. Lynch, 2016; Ye, Z., Lynch,M.D., Trahan, A.D., Rodriguez, D.L., Cooper, C.B. Bozdag, A., 2015) Specifically, phosphate depletion has been used routinely for heterologous

protein expression.(Balzer et al., 2013; Lübke et al., 1995; Song et al., 2017) In addition, it has been shown that phosphate depletion can be used to amplify the expression of heterologous proteins using the pET based T7 promoters in *E. coli.(Huber et al., 2011)* Phosphate dependent promoters are used in an engineered strain of *E. coli* with minimal acetate production, and near optimal growth rates and yields, offering tightly controlled expression. These strains and plasmids can be used in minimal media in instrumented bioreactors as well as with an optimized autoinduction broth enabling high level batch expression, in cultures as small as 20 μL in 384 well plates, to 100 mL in larger shake flasks.

2.2 Materials & Methods

2.2.1 Reagents and Media

Unless otherwise stated, all materials and reagents were of the highest grade possible and purchased from Sigma (St. Louis, MO). Luria Broth, lennox formulation with lower salt was used for routine strain and plasmid propagation and construction and is referred to as LB below. All media formulations including stock solutions are described in Supplemental Materials, Section 4, Tables S4.1 through S4.6. Working antibiotic concentrations were as follows: kanamycin (35 µg/mL), chloramphenicol (35 µg/mL), ampicillin (100 µg/mL), tetracycline (5 µg/mL), apramycin (100 µg/mL). Polypropylene glycol (MW of 2000) and casamino acids were obtained from VWR international (Suwanee, GA), product numbers E278-500G and 90001-740, respectively. Yeast extract and MOPS (3-(N-morpholino)propanesulfonic acid) were obtained from Biobasic (Amherst, NY)., product numbers G0961 and MB0360, respectively.

2.2.2 Strains and Strain Construction:

E. coli strains BL21(DE3) (Catalogue # C2527) and BL21(DE3) pLysS (Catalogue # C3010) were obtained from New England BioLabs, Ipswich, MA. Strain BW25113 was obtained from the Yale *E. coli* Genetic Stock Center (https://cgsc.biology.yale.edu/, (Baba et al., 2006; Datsenko & Wanner, 2000). Strain BWapldf was a gift from George Chen (Tsinghua University).(Jian et al., 2010) Chromosomal modifications were made using standard recombineering methodologies (Sharan et al., 2009) through scarless tet-sacB selection and counterselection, strictly following the protocols of Li et al. (Li et al., 2013) The recombineering plasmid pSIM5 and the tet-sacB selection/counterselection marker cassette were kind gifts from Donald Court (NCI, https://redrecombineering.ncifcrf.gov/court-lab.html). Briefly, the tet-sacB selection/counterselection cassette was amplified using the appropriate oligos supplying ~50 bp flanking homology sequences using Econotaq (Lucigen Middleton, WI) according to manufacturer's instructions, with an initial 10 minutes denaturation at 94 ℃, followed by 35 cycles of 94 ℃, for 15 seconds, 52 ℃ for 15 seconds, and 72 ℃ for 5 minutes. Cassettes used for "curing" of the tet-sacB cassette were obtained as gBlocks from (Integrated DNA Technologies, Coralville, Iowa, USA). The *ompT* protease gene was deleted using standard recombineering methods by selection for an apramycin selectable marker obtained from the pMDIA plasmid. (J. Yang et al., 2014) pMDIAI was a gift from Sheng Yang (Addgene plasmid # 51655; http://n2t.net/addgene:51655 ; RRID:Addgene_51655). Primers and DNA sequences are given in Supplemental Materials Section 3, Tables S3.1 and S3.2. Chromosomal modifications were confirmed

by PCR amplification and sequencing (Genewiz, NC) using paired oligonucleotides, either flanking the entire region.

2.2.3 Plasmids

pETM6, and pETM6-mCherry was a gift from Mattheos Koffas (Addgene plasmids ## 49795 and # 66534). pLysS was obtained from New England Biolabs (NEB, Ipswich, MA). Plasmids made in this study were constructed using G-blocks™ and/or PCR products and assembled using NEBuilder® HiFi DNA Assembly Master Mix following manufacturer's protocol (NEB, Ipswich, MA). Polymerase chain reactions were performed with Q5 DNA Polymerase (NEB, Ipswich, MA). pSMART-HC-Kan (Lucigen, WI), pTWIST-Chlor-Medium Copy (Twist Biosciences San Francisco, CA), pTWIST-Kan-High Copy (Twist Biosciences San Francisco, CA) and pCDF (derived from pCDF-1b, EMD Millipore, Burlington, MA) were used as a backbone vectors in these studies. Sequences of all oligos and synthetic DNA are given in Supplemental Materials Section 3, Tables S3.1 and S3.2. All plasmid sequences were confirmed by DNA sequencing (Genewiz, NC). Sequences and maps are available with Addgene. Refer to Table 1 for Addgene numbers. pCDF was constructed from pCDF-1b by first amplifying the vector with primers pCDF-1b-ampl1 and pCDF-1b-ampl2, to remove the lacI gene, followed by DNA assembly with pCDF-MCS. All genes were codon optimized for expression in *E. coli* using IDT's codon optimization tool (https://www.idtdna.com/CodonOpt). pHCKan-yibDp-GFPuv, pHCKan-yibDp-mCherry, pSMART-Ala1, pHCKan-yibDp-GFP and pHCKan-yibDp-GST were constructed by DNA assembly with linearized pSMART-HC-Kan obtained from lucigen

with yibDp-GFPuv, yibDp-mCherry, yibDp-ald*, yibDp-GFP, yibDp-GST and yibDp-GFP-β20cp6 G-blocks™ respectively. pCDF-yibDp-matB was obtained by DNA assembly of a G-block™ (yibDp-matB) with pCDF-ev which was amplified by PCR with SR2_rc and SL1_rc. pCDF-yibDp-mdlC-his was constructed from the assembly of 2 synthetic G-blocks™ yibDp-mdlC-his1 and yibDp-mdlC-his2. pTCmc-yibDp-SBS-mCherry, pTKhcan-yibDp-cimA3.7, pHCKan-yibDp-GFP-β20cp6 and pHCKan-yibDp-Nef were constructed at TWIST Biosciences (San Francisco, CA) using the pTWIST-Chlor-Medium-Copy, pSMART-HC-Kan, pSMART-HC-Kan and pTWIST-Kan-High-Copy vectors respectively. pHCKan-yibDp-GFP-cp6 was constructed by Q5 mutagenesis or "Around the World" PCR of plasmid pHCKan-yibDp-GFP-β20cp6, to remove the Lon degron tag, with primers GFP_cp6_F and GFP_cp6_R followed by DpnI treatment phosphorylation and self ligation, using KLD reaction mix obtained from NEB. (Ipswich, MA). pHCKan-yibDp-CBD-hGLY was constructed by DNA assembly with 2 PCR products amplified from (i) a plasmid coding hGLY under a T7 promoter (pHCKan-T7-CBD-hGLY, Addgene #134940, constructed at TWIST Biosciences) using primers pS-yibD-hGLY_F and pS-yibD-hGLY_R; and (ii) a plasmid containing the yibDp promoter, (pHCKan-yibDp-ald*-alaE, Addgene # 134939) using primers pS-yibDp-FOR and pS-yibDp-REV).

Table 1. Plasmids and strain used in Chapter 2

Plasmid	Insert	promoter	ori	Res	Addgene	Source
pSMART-HC-Kan	None	None	colE1	Kan	NA	Lucigen
pLysS	T7 lysozyme	NA	p15a	Cm	NA	NEB
pHCKan-yibDp-GFPuv	GFPuv	yibDp	colE1	Kan	127078	This study
pHCKan-yibDp-mCherry	6xhis-mCherry	yibDp	colE1	Kan	127058	This study
pETM6	none	T7 (pET)	colE1	Amp	49795	[a]
pETM6-mCherry	mCherry	T7 (pET)	colE1	Amp	66534	[a]
pCDF	None	None	cloDF13	Sm	89596	This study
pTCmc-yibDp-SBS-mCherry	SBS-mCherry	yibDp	p15a	Cm	134598	This study
pTKhc-yibDp-GFP-β20cp6	GFP-β20-cp6	yibD	colE1	Kan	127060	This study
pTKhc-yibDp-GFP-cp6	GFP-cp6	yibD	colE1	Kan	134938	This study
pCDF-yibDp-matB	matB	yibDp	cloDF13	Sm	134597	This study
pSMART-Ala1	AlaDh(D196A/L197R)	yibDp	colE1	Kan	65814	This study
pCDF-yibDp-mdlC-his	mdlC-6xhis	yibDp	cloDF13	Sm	134590	This study
pHCKan-yibDp-GST	GST-6xHis	yibD	colE1	Kan	134592	This study
pHCKan-yibDp-Nef	Nef	yibD	colE1	Kan	134593	This study
pTKhc-yibDp-	cimA3.7	yibDp	colE1	Kan	134595	This

cimA3.7						study
pHCKan-yibDp-CBD-hGLY	hGLYAT2	yibD	colE1	Kan	134596	This study
Strains used in this study						
Strain	**Genotype**					**Source**
BL21(DE3)	F- *ompT gal dcm* Lon *hsdS*$_B$(*r*$_B$*m*$_B$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*]$_{K-12}$(λ$_S$)					NEB
BWapldf	F-, λ-, Δ*(araD-araB)567, lacZ4787(del)(::rrnB-3) , rph-1, Δ(rhaD-rhaB)568, hsdR51, ΔackA-pta, ΔpoxB, ΔpflB, ΔldhA, ΔadhE*					[a]
DLF_R002	BWapldf, *ΔiclR, ΔarcA*					This study
DLF_R003	DLF_R002, *ΔompT::apmR*					This study
NEB - New England BioLabs, Res - resistance marker, Sm - spectinomycin, Cm- chloramphenicol, Kan - kanamycin, Amp – ampicillin						

2.2.4 BioLector™ Experiments

Growth and fluorescence measurements were obtained in a Biolector (m2p labs, 11 Baesweiler, Germany) using a high mass transfer FlowerPlate (CAT#: MTP-48-B, m2p-labs, Biolector settings were as follows: RFP gain=40, GFP gain=20, Biomass gain 20, shaking speed 1300 rpm, temperature 37 ℃, humidity 85%. Single colonies of each strain were inoculated into 5 mL LB with appropriate antibiotics and cultured at 37 ℃, 150 rpm overnight. Overnight cultures OD_{600nm} was measured and normalized to OD_{600nm} = 25.8 µL of normalized overnight culture was inoculated into 792 µL of the appropriate medium with appropriate antibiotics and transferred into wells of the FlowerPlate. Every strain was analyzed in triplicate.

2.2.4 Microtiter Plate Based Growth and Expression

Plasmids were transformed into host strains using standard protocols. Glycerol stocks were prepared for each strain plate by adding equal volume of overnight LB culture with sterile 20% glycerol. 3 μL of glycerol stocks were used to inoculate overnight culture in 150 μL LB medium with appropriate antibiotics. 96 well and 384 well plates used in these studies were obtained from Genesee Scientific (San Diego, CA, Cat #: 25-104) and VWR ((Suwanee, GA, Cat #: 10814-224). Plates were covered with sandwich covers (Model # CR1596, 96 well plates) (Model # CR1384, 384 well plates) obtained from EnzyScreen, Haarlam, The Netherlands). These covers ensured minimal evaporative loss during incubation. Microtiter plates were cultured at 37 °C, 300 rpm for 16 hours, shaker orbit is 50 mm. This combination of orbit and minimal shaking speed is required to obtain needed mass transfer coefficient and enable adequate culture oxygenation. After 16 hours of growth, a 1% volume of overnight culture was inoculated into autoinduction media plus the appropriate antibiotics. Plates were again covered with sandwich covers and grown at 37 °C, 300 rpm for 24 hours at which point samples were harvested for analysis, ie SDS-PAGE, fluorescence and optical density readings. To test the expression level of the protein panel, a volume of 100 μL of AB media per well was used.

2.2.5 Autoinduction Media Development

The autoinduction media was developed using DoE definitive screening designs and JMP software (SAS, Cary, NC). 1X trace metal mix contains 0.01 mL/L of concentrated H_2SO_4, 0.0012 g/L $CoSO_4*7H_2O$, 0.001 g/L $CuSO_4*5H_2O$, 0.0012 g/L

$ZnSO_4*7H_2O$, 0.0004 g/L $Na_2MoO_4*2H_2O$, 0.0002 g/L H_3BO_3, and 0.0006 g/L $MnSO_4*H_2O$. 0.25 g/L citric acid, 68 mM $(NH_4)_2SO_4$, 0.16 mM $FeSO_4$, 10 mM $MgSO_4$, 0.0625 mM $CaSO_4$, 2X trace metal mix, 2.5 g/L yeast extract, and 2.5 g/L casamino acid were used as the starting center point, 4X and 1/4X of the center point values were used as the upper and lower concentration ranges. Definitive screening design was performed in 5 iterations. Center point, upper and lower concentration ranges for future iterations were determined based on DoE results from the previous iteration. For testing all 212 media from the DOE, one mL of each media was prepared in deep well 96-well plates. Media were prepared from sterilized liquid stocks: $(NH_4)_2SO_4$ (3 M), Citric Acid (25 g/L), $FeSO_4$ (20 mM), $MgSO_4$ (1 M), $CaSO_4$ (5 mM), Trace Metals (250 X), Yeast Extract (100 g/L), Casamino Acids (100 g/L), Thiamine HCl (50 g/L), MOPS (1 M), and glucose (500 g/L). As all media contained equal amounts of Thiamine HCl, MOPS, glucose, and Kanamycin, these were added to each media first, followed by water. Then worklists were prepared to add the remaining media components using Tecan Evo for liquid handling. In between addition of media components, plates were shaken in a Benchmark Incu-Mixer^{TM} MP at 1500 rpm to ensure proper mixing and prevent media precipitation. Once completed, 148.5 uL of media was distributed to triplicate 96 well plates and each well was inoculated with 1.5uL of overnight LB culture. The plates were covered with EnzyScreen covers and shaken at 300 rpm at 37 ℃. After 24 hours, OD and fluorescence were measured.

2.2.6 Shake Flask Growth and Expression

Glycerol stocks were used to inoculate overnight cultures in 5 mL of LB media, with appropriate antibiotics. After 16 hours of growth, a 1% volume of overnight culture was inoculated into autoinduction media plus the appropriate antibiotics. Flasks cultures were grown at 37 ℃, 150 rpm in baffled 250 ml Erlenmeyer flasks for 24 hours at which point samples were harvested for analysis.

2.2.7 Fermentation Seeds

Single colony from transformation plate was inoculated into 5 mL LB with appropriate antibiotics and cultured at 37 ℃, 150 rpm for 16 hours. 200 µL of the LB culture was inoculated into 20 mL SM10+ media with appropriate antibiotics in 250 ml shaker flasks. The culture was incubated at 37 ℃ with a shaking speed of 150 rpm for 16 hours, at which time OD_{600nm} is usually between 6 and 10. The culture was harvested by centrifugation at 4000 rpm for 15 min, the supernatant was discarded and the cell culture was normalized to OD_{600nm} = 10 using FGM10 media. Seed vials were prepared by adding 1.5 mL of 50% glycerol to 6.5 mL of normalized OD_{600nm} = 10 culture in cryovials, and stored at -60 ℃.

2.2.8 1L Fermentations

An Infors-HT Multifors (Laurel, MD, USA) parallel bioreactor system was used to perform 1L fermentations. Vessels used had a total volume of 1400 mL and a working volume of up to 1 L. Online pH and pO2 monitoring and control were accomplished with Hamilton probes. Offgas analysis was accomplished with a multiplexed Blue-in-One

BlueSens gas analyzer (BlueSens. Northbrook, IL, USA). Culture densities were continually monitored using Optek 225 mm OD probes, (Optek, Germantown, WI, USA). The system used was running IrisV6.0 command and control software and integrated with a Seg-flow automated sampling system (Flownamics, Rodeo, CA, USA), including FISP cell free sampling probes, a Segmod 4800 and FlowFraction 96 well plate fraction collector. Tanks were filled with 800 mL of either FGM10 or FGM30 medium, which has enough phosphate to target a final *E. coli* biomass concentration ~ 10 gCDW/L or 30 gCDW/L respectively. Antibiotics were added as appropriate. Phosphate, glucose, thiamine and antibiotics were added after cooling the tank vessel containing the rest of the media components. Frozen seed vials were thawed on ice and 8 mL of seed culture was used to inoculate the tanks. After inoculation, tanks were controlled at 37 ℃ and pH 6.8 using 14.5 M ammonium hydroxide and 1 M hydrochloric acid as titrants. The following oxygen control scheme was used to maintain the desired dissolved oxygen set point. Initial air flow rate was set to 0.3 vvm and initial agitation of 300 rpm. In order to maintain a dissolved oxygen concentration of 25%, airflow was increased to 1 vvm, then agitation was increased to a maximum of 1200 rpm, and finally oxygen was increased in the gas mixture from 0% additional to a maximum of 80% additional.

Glucose feeding was as follows. For 10 gCDW/L fermentations, starting batch glucose concentration was 25 g/L. A constant concentrated sterile filtered glucose feed (500 g/L) was added to the tanks at 0.5 g/h once dissolved oxygen concentration dropped from 100% to 25% and ramped up to 1 g/h, once increased agitation was required to maintain the dissolved oxygen setpoint. For 30 gCDW/L fermentations, starting batch

glucose concentration was 25 g/L. Concentrated sterile filtered glucose feed (500 g/L) was added to the tanks at an initial rate of 3.5 g/L-h when the rate of agitation increased above 1000 rpm. This rate was then increased exponentially, doubling every 1.083 hours (65 min) until 40 g total glucose had been added, at which point the feed was set to 1g/L-hr.

2.2.9 Organic Acid Quantification

Two orthogonal methods were used to quantify organic acids including lactate, acetate, succinate, fumarate, pyruvate, malate and others. The first method was a reverse phase UPLC method. Chromatographic separation was performed using a Restek Ultra AQ C18 column (150 mm × 2.1 i.d., 3 μm; CAT#: 9178362, Restek Corporation, Bellefonte, PA) at 30 ℃. 20 mM phosphoric acid was used as the eluent. The isocratic elution rate was at 0.8 mL/min, run time was 1.25 min. Sample injection volume was 10 μL. Absorbance was monitored at 210 nm. The second method relied on ion exchange chromatography and refractive index detection. A Phenomenex Rezex™ ROA-Organic Acid H+ (8%) (30 x 4.6 mm; CAT#: 00A-0138-E0, Phenomenex, Torrance, CA) was used for a 30 minute isocratic separation using a mobile phase of 5 mM H_2SO_4, at a flow rate of 0.5 mL/min. Again sample injections were 10 μL. Organic acid elution times were as follows: Pyruvate 13.3 min, Citramalate 13.75 min, Citrate 10.9 min, Lactate 17.5 min and Acetate 20.3 min.

2.2.10 Glucose Quantification

Similarly, two methods were used to quantify glucose. The first was identical to the second organic acid method, utilizing the Resex column for ion exchange linked to refractive discussed above, wherein glucose eluted at 12.5 minutes. The second method was a similar UPLC method also relying on ion exchange and refractive index detection. Chromatographic separation was performed using a Bio-Rad Fast Acid Analysis HPLC Column (100 x 7.8 mm, 9 μm particle size; CAT#: #1250100, Bio-Rad Laboratories, Inc., Hercules, CA) at 65 ℃. 5 mM sulfuric acid was used as the eluent. The isocratic elution was as follows: 0–0.1 min, flow rate increased from 0.4 mL/min to 0.42 mL/min, 0.1–12 min flow rate at 0.48 mL/min. Sample injection volume was 10 μL.

2.2.11 Determination of Strain Dry Weight

Culture samples (5 ml, n=3) were taken and washed 2X with deionized water via centrifugation and resuspension. After wash steps the OD of the samples were determined at 600 nm. Subsequently, samples were filtered over pre-weighed nitrocellulose filters (pore size, 0.45 μm). Filters were washed extensively with demineralized water and dried in a microwave oven for 2 min and weighed to determine correlation of OD_{600nm} and gDCW, which was 0.35.

2.2.12 Phosphate Quantification

Phosphate concentrations were determined using the BioMOL Green colorimetric assay from Enzo Life Sciences (Farmingdale, NY) according to manufacturer's instructions.

2.2.13 Fluorescence measurements

Optical densities and fluorescent were measured using a Tecan Infinite 200 plate reader. Measurements were performed using 200 uL in black 96 well plates (Greiner Bio-One, Reference Number: 655087). Optical density was read at 600 nm (Filter from Omega Optical, Part Number: 3019445) and adjusted by subtracting a blank, followed by correction for pathlength and dilutions. For GFP fluorescence, samples were excited at 412 nm (Omega Optical, Part Number: 3024970) and emission was read at 530 nm (Omega Optical, Part Number: 3032166) using a gain of 60. Fluorescence readings were then adjusted for dilution.

2.2.14 SDS-PAGE and GFP quantification

The OD_{600nm} of culture samples to be analyzed was measured before harvesting the cells by centrifugation, which was done at 4000 rpm for 15 minutes. The cells were resuspended in 50 µl of phosphate-buffered saline with protease inhibitors (ThermoFisher Scientific, MA, product number A32965) and 5 mM EDTA. 25 µl of the resuspended cells were mixed with 25 µl of 2x Laemmli sample buffer (Biorad, CA) and boiled for 5 minutes at 95 ℃. The boiled samples were centrifuged at 14,000 rpm for 10 minutes and 20 µg of total protein per sample was then loaded into a 4-15% gradient Mini-Protean TGX precast protein gel (Biorad, CA) and ran at 140 V. The volume loaded per sample was calculated as volume=$100/OD_{600nm}$ The gels were stained using Coomassie Brilliant Blue R-250. Gels were imaged using a UVP PhotoDoc-It™ Imaging System (Analytik Jena, CA) and expression levels were quantified using ImageJ (NIH, MD). To correlate GFPuv fluorescence with grams of GFPuv, samples were taken wherein both (i)

fluorescence was measured as described above and (ii) expression level was calculated as described above. Total cellular protein was estimated at 500 mg/gDCW or 50% of dry cell weight (Long & Antoniewicz, 2014). In these comparison, 3.24 e 9 relative fluorescent units corresponded to 1 gram of GFPuv. This correlation was also used to calculate GFPuv titers across all experiments.

2.2.15 Cytometry

BL21(DE3) pLys bearing pETM6 (negative control) or pETM6-mCherry as well as DLF_R002 bearing pSMART-HC-Kan (negative control) or pHCKan-yibDp-mCherry were grown in 5 ml of LB overnight at 37 °C, 150 rpm. After 16 hours, 1% volume of overnight BL21(DE3)pLys or DLF_R002 cultures were used to inoculate 20 ml of LB or AB media in 250 baffled Erlenmeyer and incubated at 37 °C, 150 rpm. BL21(DE3) cultures were induced at OD_{600nm} ~0.3 with 1 M IPTG solution to a final 1 mM IPTG concentration. Samples for BL21(DE3) were collected 20 hours after induction with IPTG. Samples for DLF_R002 were collected after 24 hours of inoculation. Samples were serially diluted 1000-fold with sterile DI water before analyzing them in a Thermo Attune NXT flow cytometer (ThermoFisher Scientific, MA). Samples were run at a 12.5 µl/min flow rate. Fluorescence measurements were taken from the 620/15 band pass filter after exciting the cells with a yellow laser at 561 nm. Forward scatter vs time plot was monitored during the run to ensure no clogging occurred. The forward scatter height vs forward scatter area plot was also monitored to ensure no cell clumping. Forward scatter height vs side scatter height plots were analyzed using DLF_R002 bearing pSMART-HC-Kan to determine the appropriate gating to exclude small particles from being

counted as events. A forward scatter height of 10,000 and a side scatter of 2,500 were used for gating for all samples. Data was analyzed using FlowJo v10.6.1 (BD, NJ).

2.2 Results

2.2.1 Initial Characterization of Phosphate Induction with the yibDp gene promoter

To move from IPTG based induction to autoinduction via phosphate depletion, we leveraged a previously reported *phoB* regulated promoter, specifically a modified promoter of the *E. coli yibD* (*waaH*) gene, referred to herein as yibDp, and constructed a plasmid enabling the induction of mCherry upon phosphate depletion (pHCKan-yibDp-mCherry, Table 1; refer to Appendix A, Table 6 for the promoter sequence). (Baek & Lee, 2006; Huerta & Collado-Vides, 2003; Lipscomb et al., 2018; M. D. Lynch, 2011; Michael D Lynch, Ryan T Gill, Tanya EW Lipscomb, 2013; Yamada et al., 1989) We initially evaluated the expression of this construct in BL21(DE3), BL21(DE3) with pLysS and a well characterized *E. coli* K12 derivative : BW25113. (Baba et al., 2006; Grenier et al., 2014) The accessory plasmid (pLysS) expressing T7 lysozyme, is routinely used to reduce leaky induction in pET-based systems. (F. W. Studier, 1991) Unexpectedly, significant basal expression was observed in BL21(DE3) (Appendix A, Figure 29). In contrast, no significant basal expression was observed in BW25113. These results indicate the potential ability of T7 RNA polymerase to recognize this promoter, differential *phoB* regulation between these strains, or yet additional unknown regulatory differences.

2.2.2 Host Strain Engineering

With BL21(DE3) demonstrating baseline heterogeneous leaky expression with the yibDp promoter, and in light of other routine issues encountered in using BL21 and its derivatives, such as accumulation of acetic acid, we turned to engineering a BW25113 derivative for optimal growth and minimal byproduct formation. We began with a previously reported derivative, strain BWapldf, with deletions in genes leading to common mixed acid fermentation products, such as lactic and acetic acid.(Jian et al., 2010) BWapldf has deletions in the following genes: ackA-pta, pflB, adhE, ldhA, and poxB reducing the rates of production of acetate, formate, lactate and ethanol from overflow metabolites. Deletions of the two global regulators iclR and arcA were next incorporated into this strain. These mutations have been shown to improve biomass yield and reduce overflow metabolism in K12 derivatives. Together these mutations increase flux through the citric acid cycle and glyoxylate bypass and reduce overflow metabolism by increasing i) the rate of oxidation of excess carbon to carbon dioxide and increasing ATP supply.(Waegeman et al., 2011, 2012)

These strains, as well as a BL21(DE3) pLysS control were initially evaluated in controlled fed batch fermentations, using a defined minimal media (FGM10 media, refer to Materials and Methods) wherein phosphate concentrations limit biomass levels. Growth rates, biomass and byproducts, including acetic acid, were measured. Results are given in Figure 2. In these studies organic acid byproducts, other than acetic acid were not observed. As expected, BL21(DE3) produced acetic acid during growth (Figure 2e) . Interestingly, strain BWapldf, despite having numerous deletions had a significantly

decreased biomass yield and increased acetic acid production compared to BW25113 (Figure 2f and 2g). The deletion of the two global regulators, arcA and iclR, (strain DLF_R002) recovered biomass yield and virtually eliminated acetic acid production in this host (Figure 2h). Refer to Appendix A Figure 30 for maximal growth rates.

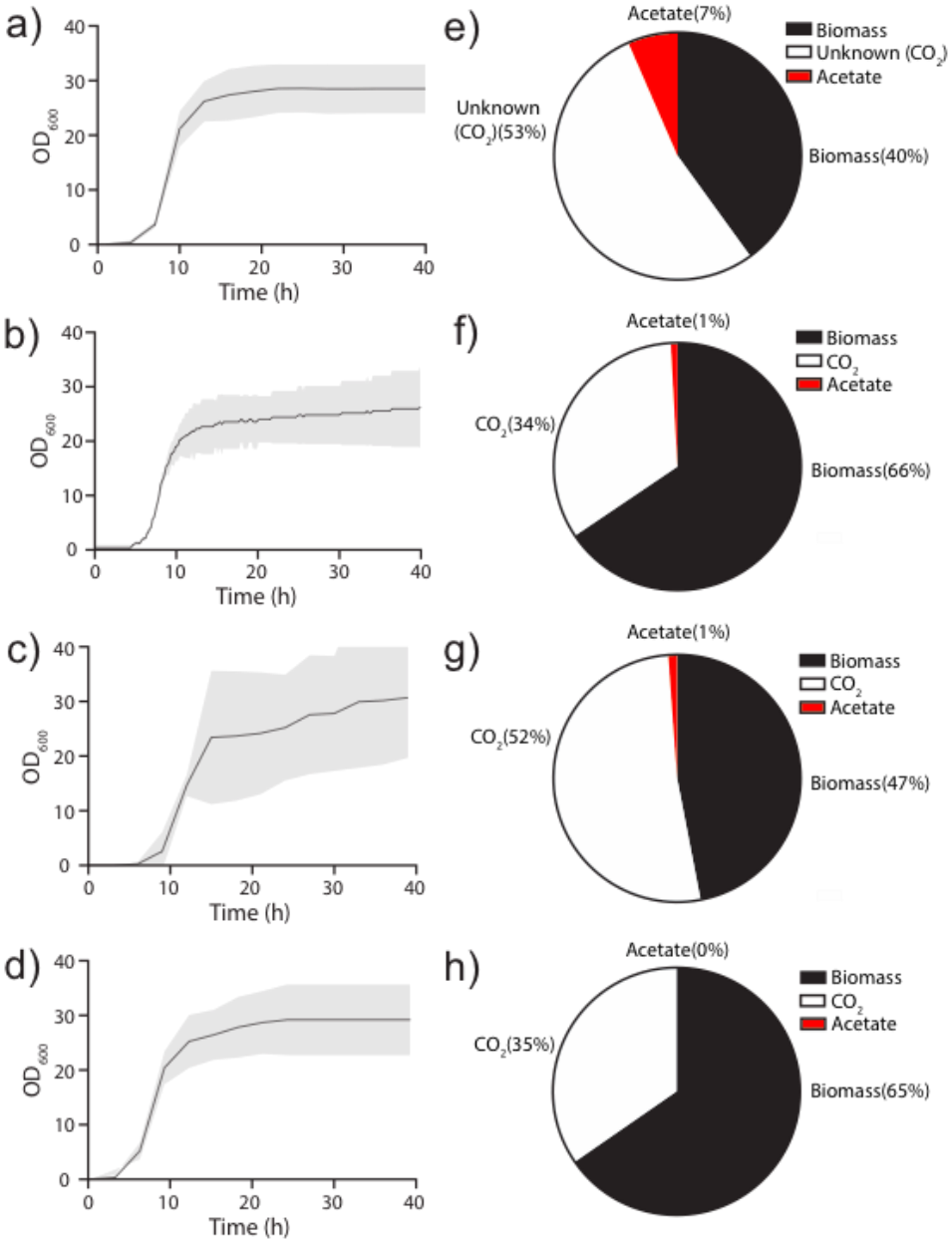

Figure 2: Growth and byproduct formation of *E. coli* strains in minimal media fermentations. Biomass levels as a function of time for (a) BL21(DE3)pLys (b) BW25113, (c) BWapldf and (d) DLF_R002 respectively. (e-h) Distribution of glucose utilized during growth in minimal medium fermentations: (e) BL21(DE3)pLys, (f) BW25113, (g) BWapldf and (h) DLF_R002 respectively. Results are averages of duplicate fermentations. CO_2 was explicitly measured via off-gas analysis for strain BW25113, BWapldf and DLF_R002. In the case of BL21(DE3) pLys, CO_2 is included in unknown products required to account for glucose consumption.

Using strain DLF_R002 we next turned to evaluate protein expression in bioreactors using FGM10 media. As mentioned biomass levels supported by FGM10 media are limited by phosphate, and phosphate depletion occurs when biomass levels reach an optical density of ~ 30-35 or ~ 10 gCDW/L. In this case we constructed an additional plasmid with GFPuv driven by the yibDp promoter (pHCKan-yibDp-GFPuv, Table 1). (Crameri et al., 1996) Results are given in Figure 3a. Biomass levels reached ~10 gCDW/L producing final GFPuv titers of ~ 2.7 g/L or 270 mg/gCDW. With the success in low cell density we turned to develop a process with higher cell density utilizing FGM30 media with enough phosphate to support three times the biomass (~ 30 gCDW/L). Results are given in Figure 3b in this case using DLF_R003. DLF_R003 is a derivative of DLF_R002 with a deletion in the outer membrane *ompT* protease, which has proteolytic activity even under denaturing conditions, creating issues with purification of recombinant proteins. (Gill et al., 2000; White et al., 1995) Specific expression levels were maintained at higher biomass levels resulting in a three-fold improvement in GFPuv titer reaching ~ 8.1g/L.

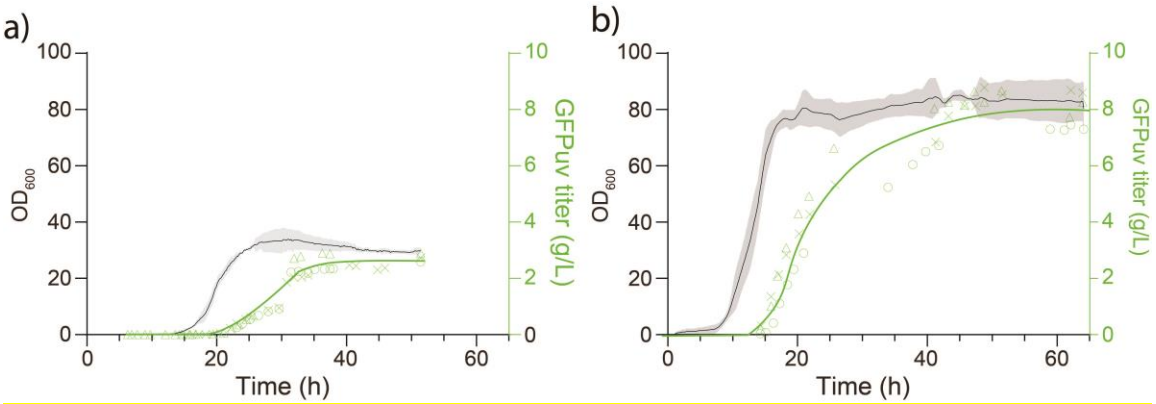

Figure 3: Autoinduction of GPFuv expression in bioreactors. Triplicate 1L bioreactors, with FGM10 minimal media were inoculated with DLF_R002 bearing plasmid pHCKan-yibDp-GFPuv. Optical density (black lines) and GFPuv were measured over time. Shaded area is standard error of triplicate growth profiles. X's, triangles and circles are normalized GFPuv fluorescence units from each of the three fermentations. Green line is the best fit of these three expression profiles.

2.2.3 Development of Phosphate Limited Media for Auto-induction

We next turned to the optimization of media formulations for more routine autoinduction via phosphate depletion. Importantly, the fermentations discussed above (Figure 2) were performed with defined minimal media, which while lower in cost in larger scale production, can lead to significant lags when cells transition from a richer cloning and propagation media such as LB. In order to overcome this, seed cultures are often used to adapt the cells to a more minimal media (as they were in this case, refer to Methods) prior to inoculation of bioreactors. For routine lab scale protein expression, media adaptation is not desirable, and rather protocols enabling direct inoculation of production flasks from overnight LB cultures is preferred. As a result, we developed batch autoinduction broth with more complex nutrient sources including yeast extract and casamino acids. Media formulations were developed using standard Design of Experiments methodology (DoE) and evaluated in 96 well plates. These experiments were performed using strain DLF_R002 bearing plasmid pHCKan-yibDp-GFPuv, described above. Briefly, overnight LB cultures were used to inoculate various media in 96 well plates. Biomass and GFPuv levels were measured after 24 hours. Importantly, no phosphate was added to these media, as adequate batch phosphate is supplied in the

complex nutrient sources (yeast extract and casamino acids). Results are given in Figure 4. Models built based on these results did not predict significant improvements in expression over the best performing experimentally tested formulations Figure 4a. The media formulation producing the most GFPuv (as measured by relative fluorescence), was renamed AB (autoinduction broth) and used in subsequent studies. To evaluate the time course of growth, phosphate depletion and autoinduction in AB, DLF_R002 pHCKan-yibDp-GFPuv, was grown in AB in the Biolector™ Microreactor. Results are given in Figure 4b.

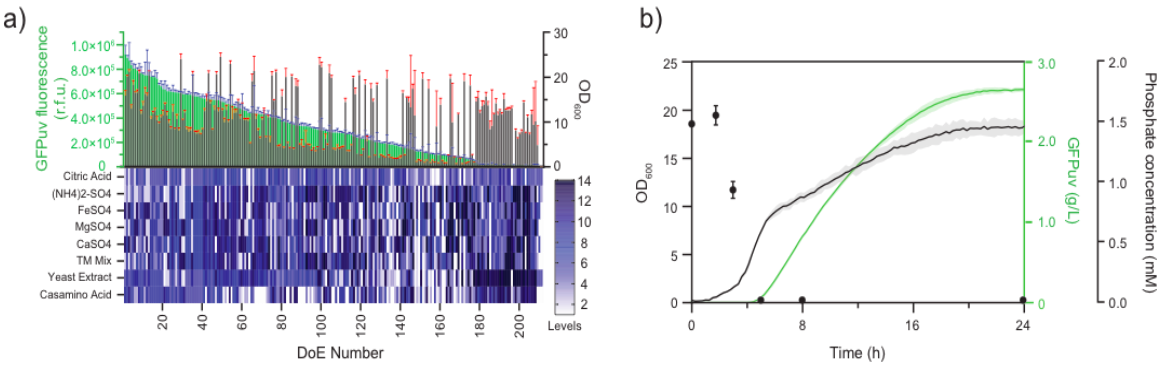

Figure 4: Media Development using Design of Experiment Methodology. 212 media formulations were evaluated for autoinduction based on phosphate depletion, each comprising different "levels" of casamino acids, yeast extract, trace metals (TM Mix), calcium sulfate ($CaSO_4$), magnesium sulfate ($MgSO_4$), iron(II) sulfate ($FeSO_4$), ammonium sulfate ($(NH_4)_2SO_4$) and citric acid. a) Upper panel: GFP (green bars) and OD_{600nm} (gray bars) rank ordered plot for all media formulations. Standard deviations are from triplicate experiments. Lower panel: Nutrient concentration levels for all media (Refer to Supplemental Materials Section 4, for specific concentrations for each level). Strain DLF_R002 with plasmid pHCKan-yibDp-GFPuv was used for all experiments. b). GFP fluorescence (green line), phosphate levels (black circles) and OD_{600nm} (black line) for strain DLF_R002 with plasmid pHCKan-yibDp-GFPuv in media #36 (Autoinduction Broth, AB) media. Standard deviations (shaded regions) are from triplicate experiments.

2.2.4 Comparison with current approaches.

With the successful development of an optimal autoinduction broth, we turned to a head to head comparison of this approach with the traditional protocols based in LB media as well as the lactose based autoinduction system as developed by Studier. (F. William Studier, 2005, 2014) Due to the availability of a pET-mCherry plasmid (Table

1) mCherry was used as the reporter for this comparison. Specifically, induction of mCherry in BL21(DE3) with pLysS and pETM6-mCherry, using either IPTG based induction in LB media, or lactose autoinduction media was compared to strain DLF_R002 with plasmid pHCKan-yibDp-mCherry in AB. To monitor not only endpoint expression but the dynamics of growth and auto-induction, these studies were performed in the Biolector™. Results are shown in Figure 5. As expected, using *E. coli* BL21(DE3) and pET-based expression, lactose based autoinduction media enabled higher cell densities and higher expression levels of mCherry than induction with IPTG (Figure 5a). Phosphate based autoinduction using strain DLF_R002 enabled a further 40% increase in final mCherry levels at 24 hrs over BL21(DE3) (Figure 5b). Refer to Appendix A, Figure 31 for a comparison of BL21(DE3) vs BL21(DE3) plus pLys. Cytometry was used to further characterize these two expression systems (Figure 5c). Phosphate based autoinduction not only had more homogeneous induction but also more expression per cell. Additionally, one of the major potential reported advantages of BL21(DE3) and related strains is reduction in Lon protease activity.(Ratelade et al., 2009) To investigate the impact of Lon activity in these strains, a previously reported fluorescent Lon substrate was used to monitor the impact of this protease. Specifically, a circular permutation variant of GFP with a Lon degradation tag (GFP-β20-cp6) was used.(Wohlever et al., 2013) Results are given in Figure 5d. The expression level of the Lon substrate was significantly reduced compared to a non-Lon substrate for both strains, but at least with this specific reporter, no significant difference in cell specific Lon activity was observed between BL21(DE3) and DLF_R002.

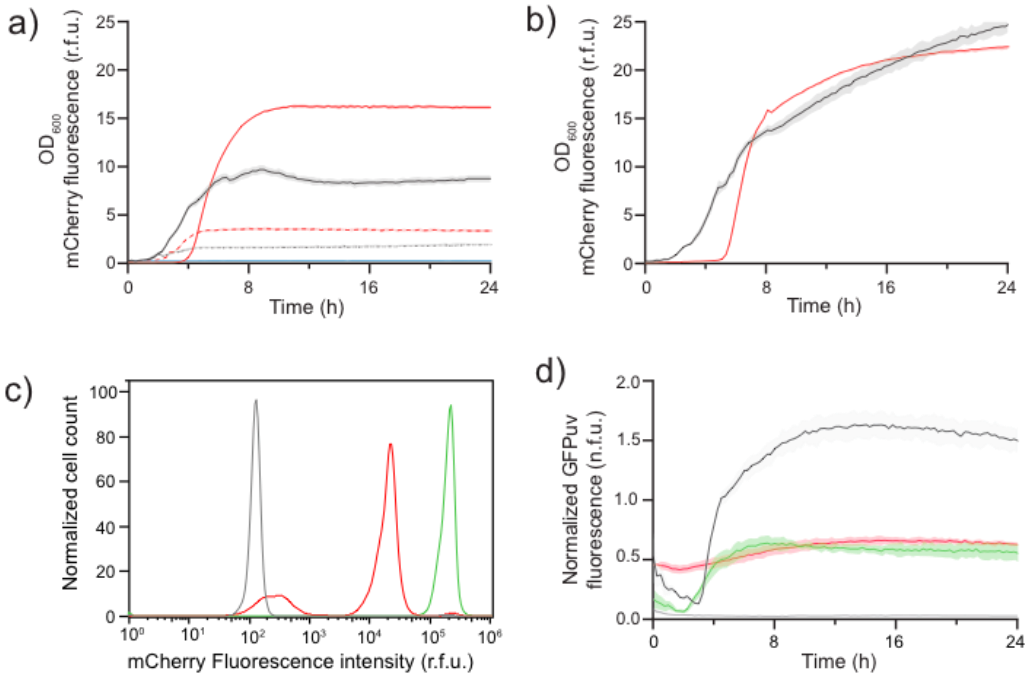

Figure 5: Head to head comparison of autoinduction via phosphate depletion with pET based expression in BL21(DE3). a) pET based mCherry expression in BL21(DE3) with pLysS. mCherry (red lines) and biomass levels (OD_{600nm}, black lines) over time. Solid lines- lactose based autoinduction. Dashed lines- IPTG induction in LB media. b) yibDp based mCherry expression in DLF_R002 in AB media mCherry (red lines) and biomass levels (OD_{600nm}, black lines). c) Cytometry of induced populations (gray- empty vector control, red- pET-mCherry in BL21(DE3) + pLysS, green - yibDp-mCherry in DLF_R002). d) expression of the Lon substrate (GFP-β20-cp6) in BL21(DE3) and DLF_R002. Normalized fluorescence is relative fluorescence normalized to optical density. Black line- GFP control (non Lon substrate) in DLF_R002. Red line - BL21(DE3) expressing GFP-β20-cp6. Green line - DLF_R002 expressing GFP-β20-cp6. Shaded areas are standard deviations of at least three replicates.

2.2.5 Development of Shake Flask Protocols

For any expression protocol to be widely applicable, it cannot rely on controlled bioreactors and/or specialty plate systems, but be accessible to the average laboratory. Toward this goal, we turned to the optimization of the protocol in shake flask cultures. As mentioned above, one primary difference between bioreactor experiments and shake flask cultivation is oxygen transfer. While instrumented bioreactors and micro-reactors such as the Biolector^{TM} can easily meet these mass transfer targets, standard shake flask have

reported oxygen transfer rates anywhere from 20 mmoles/L-hr (for unbaffled flasks) to 120 mmoles/L-hr for baffled glassware .(Running & Bansal, 2016) A key potential consequence of shake flasks is oxygen limitation and reduced growth rates and expression. As a consequence we sought to evaluate the optimal culture conditions to achieve maximal expression in shake flask cultures with a focus on baffled 250 mL Erlenmeyer flasks and 2.8 L Fernbach flasks. As seen in Figure 6, again culture volume plays a key role in optimal protein expression, with 20mL or lower being optimal in baffled 250 mL Erlenmeyer flasks and 100 mL or lower being optimal in 2.8 L Fernbach flasks. These results were obtained in shakers where an adhesive mat is used to hold flasks and shaking speeds are limited to 150 rpm. Using clamps, higher shaking speeds may enable optimal expression using larger shake flask fill volumes.

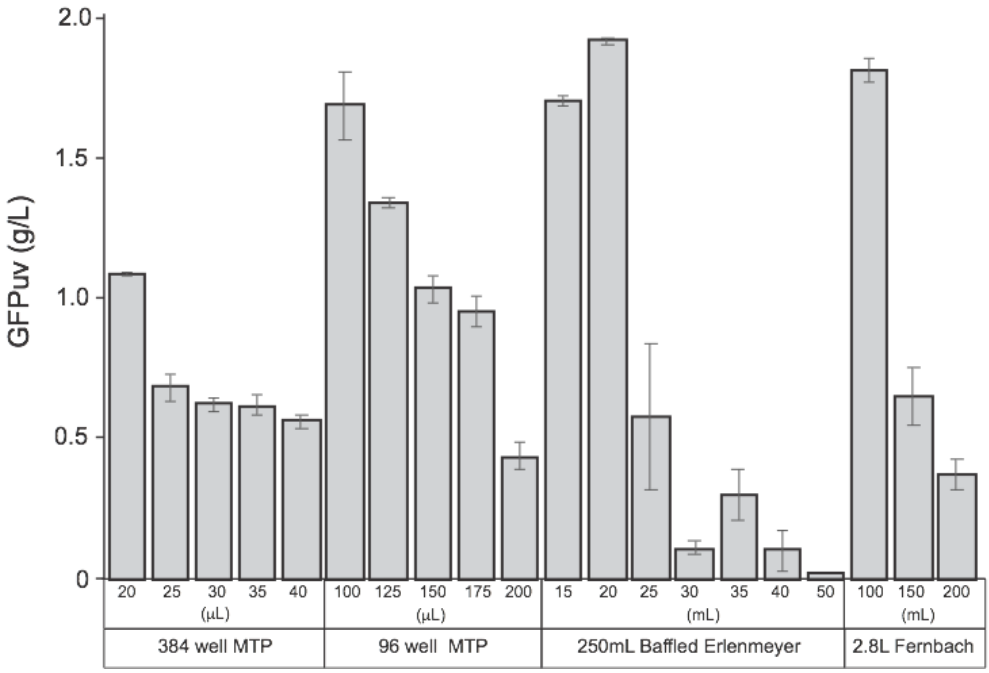

Figure 6: Optimization of autoinduction in batch cultures at various scales. Impact of various fill volumes on expression in AB. Varying fill volumes in 384 and 96 well plates as well as 250 mL baffled Erlenmeyer and 2.8 L Fernbach flasks. When using 384 well plates, 0.05% polypropylene glycol (2000MW) was added to the media. DLF_R002 with plasmid pHCKan-yibDp-GFPuv was used for all experiments.

2.2.6 Utility with a diverse group of recombinant proteins

All results discussed to this point relied on easily quantified reporter proteins (GFPuv and mCherry), which are easily expressed to high levels in most expression hosts. In order to evaluate the broader applicability of the approach, the expression of a group of other diverse proteins was evaluated in several vector backbone contexts in the phosphate autoinduction protocol. These included: a borneol diphosphate synthase, a terpene synthase with a C-terminal mCherry tag (Whittington et al., 2002; Wise et al., 1998), a mutant alanine dehydrogenase (Lerchner et al., 2016), a malonyl-CoA synthetase (An & Kim, 1998), a benzoylformate decarboxylase (Tsou et al., 1990), glutathione S-transferase (Oakley, 2011), HIV-1 nef protein (Pereira & daSilva, 2016), a mutant citramalate synthase (Atsumi & Liao, 2008), and a human glycine acyltransferase with an N-terminal chitin binding tag (Waluk et al., 2010). (Refer to Table 1 for construct details.) As can be seen in Figure 7, expression levels ranged from ~ 10 % of total protein for a large terpene synthase to 55% in the case of alanine dehydrogenase, achieving maximal protein concentrations of 275 mg/gCDW in the best case.

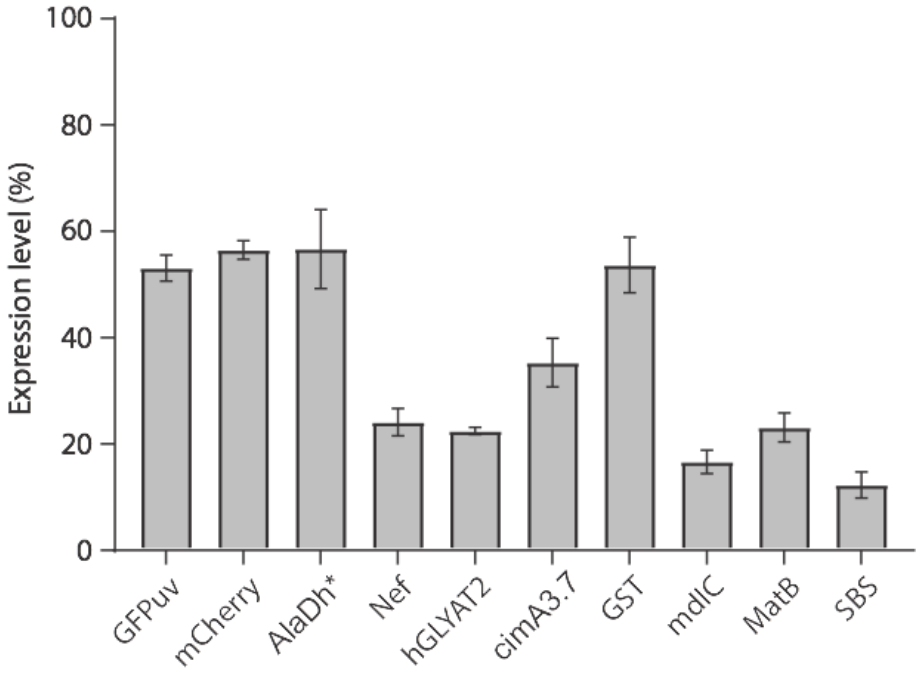

Figure 7: Autoinduction in AB in 96 well plates for a diverse set of recombinant proteins. These include: GFPuv, mCherry, AlaDh* (a mutant alanine dehydrogenase) , Nef (HIV-1 Nef protein), hGLYAT2 (human glycine acyltransferase-2 an N-terminal chitin binding tag), cimA3.7 (a mutant citramalate synthase), GST, mdlC (benzylformate decarboxylase), matB (malonyl-CoA synthetase), and SBS (bornyl-diphosphate synthase with a C-terminal mCherry tag). Percent of total expression is given for three replicates. Refer to Supplemental Figure S7 for an example SDS-PAGE result.

2.3 Discussion

Two-stage expression induced upon phosphate depletion enables a facile and versatile approach to routine high-level recombinant protein production. In the case of GFPuv, protein titers approaching 2 g/L in batch microtiter plates and shake flasks. These titers correspond to protein yields of 20 μg of protein per well in 384 well plates, 170 μg per well in 96 well plates, and 40 mg and 180 mg of protein in 250 mL Erlenmeyer and Fernbach flasks, respectively. Importantly, current results also support homogenous expression using phosphate depletion. Expression levels will of course vary as a function

of the protein and expression construct, but initial testing with additional proteins supports expression levels from 10 to 55 percent of total cellular protein, which at the high end is ~ 275 mg/gCDW of recombinant protein and represents significant improvements in heterologous protein expression in *E. coli (Chew et al., 2012; Striedner et al., 2010)*. More work is needed to better understand the mechanisms unexpectedly high expression levels observed in this system. Initial adaptation to instrumented bioreactors, enabled GFP titers as high as 2.7 g/L, 270 mg/gCDW and 55% expression, with 10 gCDW/L. Process intensification to increase biomass levels to ~ 30 gCDW/L, while maintaining specific expression levels, resulting in a ~3 fold improvement in GFP titers reaching 8.1 g/L. Further optimization of bioreactor protocols may enable much higher cell density cultures. If truly high cell density fermentations (from 50-100 gCDW/L of biomass) can be developed with equivalent expression levels, protein titers in the range of 15-30 g/L or higher in some cases can be expected.

There are however several challenges with the existing protocol. Firstly, proteins of interest must be cloned into a plasmid with the yibDp promoter. Screening of additional phosphate (*phoB*) regulated promoters may yield improved or varied expression. Adaptation of the system for use with existing pET-based plasmids would also be of utility for proteins that are already cloned into these standard vectors. Secondly, preparation of AB media is more complicated than making routine LB media. Lastly, this system is not designed to address challenges with difficult to express proteins common to other standard heterologous expression systems.

Despite these potential limitations, the development of strains, plasmids and protocols for autoinduction based on phosphate depletion not only enables improved expression, with impressive protein titers, but also a scalable methodology. A single host and plasmid can be used in high throughput screening of initial expression constructs or mutant variants all the way through to instrumented bioreactors. These results support the biosynthetic potential of phosphate depleted stationary phase cultures of *E. coli*. Decoupling growth from production also has the potential to enable future studies focused on key remaining limitations in protein biosynthesis in this well characterized host.

3. Improved two-stage protein expression and purification via autoinduction of both autolysis and auto DNA/RNA hydrolysis conferred by phage lysozyme and DNA/RNA endonuclease

3.1 Introduction

E. coli is a mainstay for routine expression of recombinant proteins. Recent estimates indicate that over 70% of laboratory studies, reliant on heterologous proteins, utilize *E. coli*. This microbe is commonly used in workflows ranging from high throughput screens, to routine shake flask expression and larger scale fermentations. (Bill, 2014; Rosano & Ceccarelli, 2014b) In addition, *E. coli* is also used for the manufacturing of proteins at large scale, including the production of over 30% of protein based drugs. (Ferrer-Miralles et al., 2009; Sanchez-Garcia et al., 2016) A key challenge to the use of *E. coli* as well as other expression systems where proteins are not secreted, is the recovery of protein from the cell, which routinely requires cell lysis. Common laboratory methods for lysis include: chemical (base or detergents), biochemical (lysozyme) as well as mechanical methods (cell disruptors, french press or sonication), which can not only be tedious and time consuming but yield inconsistent results. (Asenjo, 1990; Cull & McHenry, 1990; Foster, 1992) Certain proteins may not tolerate the use of chemical lysis buffers and mechanical methods can lead to incomplete lysis and release of target proteins. In addition, mechanical methods are not amenable to certain workflows such as high throughput screening. (Aharoni et al., 2005; Cai et al., 2008; Chien & Lee, 2006; Foster, 1992; Reymond, 2006) At larger scales, homogenizers are often used to enable more consistent cell lysis, but these units are both costly and add additional steps

to commercial processes. (Doran, 1995; Harrison, 2019) Significant efforts have been made in developing methods for rapid, consistent cell lysis. (Aharoni et al., 2005; Asenjo, 1990; Dhawan et al., 2002; Di Carlo et al., 2003; F. Han et al., 2003), including engineering of *E. coli* strains for autolysis, usually upon induction of one more proteins with lytic activity including: lysozyme, D-amino acid oxidase, muramidase and bacterial phage lysis proteins, which are induced in parallel with proteins of interest and activated after cells are harvested. (Cai et al., 2008; Chien & Lee, 2006; Jechlinger et al., 1999; Kloos et al., 1994; M. Morita et al., 2001; Tanji et al., 1998; Vasala et al., 1999)

Key remaining challenges with many of these approaches include additional process steps, incomplete lysis or additional induction procedures or vectors. (Cai et al., 2008; Cárcel-Márquez et al., 2019; Didovyk et al., 2017; Halfmann & Lubitz, 1986; Tamekou Lacmata et al., 2017; Xu et al., 2006; H. Zhang et al., 2019) In addition, previous efforts have been focused on cell wall lysis and protein release without consideration of lysate clarification to remove oligonucleotide contamination as well as reduce lysate viscosity. In commercial production after cell lysis, nucleases such as the non specific DNA/RNA endonuclease from *S. marcescens* (Benzonase ®) or alternatives are often used to remove nucleotide contaminants and reduce lysate viscosity to enable easier follow on purification. (Ball et al., 1987; Benedik & Strych, 1998; Biedermann et al., 1989) the non specific DNA/RNA endonuclease from *S. marcescens* is a small nonspecific extracellular nuclease that is routinely used to hydrolyze contaminating nucleotides during protein purification, and has activity with both double stranded and single stranded DNA as well as RNA. (Olszewski & Filipkowski, 2009) An engineered

strain of *E. coli* has been reported with periplasmic expression of this endonuclease which auto-hydrolyzes host nucleic acids upon cell lysis, (Cooke et al., 2003) but autolysis and autohydrolysis have yet to be combined.

We recently reported strains, plasmid and protocols for the autoinduction of protein expression in stationary phase upon batch phosphate depletion, enabling high protein titers in a very simplified protocol, with no leaky expression. (Menacho-Melgar, Ye, et al., n.d.) To build upon this system, in this work we have further engineered strains with a autolysis and autohydrolysis "module" consisting of lambda phage lysozyme (Lambda R gene) and endonuclease (encoded by the *Serratia marcescens nucA* gene).(Ball et al., 1987; Benedik & Strych, 1998; Biedermann et al., 1989) The expression of a protein of interest as well as expression of the autolysis module are induced upon phosphate depletion coincident with entry into stationary phase, via phosphate regulated promoters (Figure 8a). (Baek & Lee, 2006; Huerta & Collado-Vides, 2003; Lipscomb et al., 2018; M. D. Lynch, 2011; Michael D Lynch, Ryan T Gill, Tanya EW Lipscomb, 2013; Yamada et al., 1989) These two genes are integrated as an operon into the chromosome in the *ompT* locus (Figure 8b), also deleting this protease, which can lead to improved protein yields. (Gill et al., 2000; White et al., 1995) Tightly controlled (non-leaky) expression and autoinduction media enable a greatly simplified single step process for both high levels of expression and lysate preparation prior to further purification.

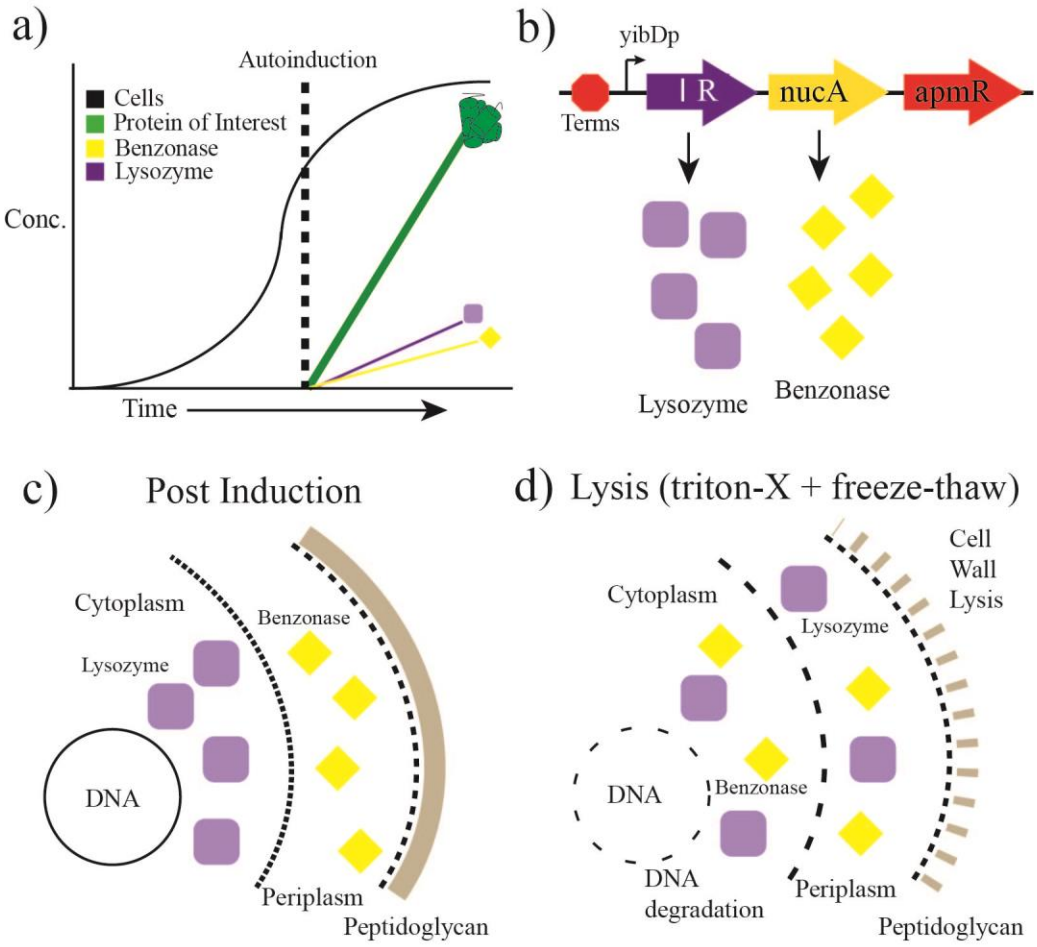

Figure 8: Overview of 2-stage autoinducible autolysis/hydrolysis. a) Cells are grown in optimized media until phosphate is consumed and cells enter a productive stationary phase, simultaneously phosphate regulated promoters are used to induce expression of i) the protein of interest as well as ii) a cytosolic lambda lysozyme and periplasmic endonuclease which are b) integrated as an operon into the chromosome at the ompT locus along with an apramycin resistance cassette (apmR) c) both during and post induction lysozyme is separated from the peptidoglycan cell wall by the cell membrane, similarly endonuclease is isolated from cellular DNA. d) Upon the addition of detergents or mechanical stress to disrupt the cellular membrane (in this case triton-X100) the endonuclease and lysozyme can access their respective substrates leading to combined autolysis and autohydrolysis of DNA and RNA.

3.2 Materials & Methods

3.2.1 Reagents and Media

Unless otherwise stated, all materials and reagents were of the highest grade possible and purchased from Sigma (St. Louis, MO). Luria Broth, lennox formulation with lower salt was used for routine strain and plasmid propagation and construction and is referred to as LB below. Working antibiotic concentrations were as follows: kanamycin (35 µg/mL) and apramycin (100 µg/mL). Autoinduction Broth (AB) and FGM10 media were prepared as previously reported. (Menacho-Melgar, Ye, et al., n.d.)

3.2.2 Strains and Plasmids:

Strain Xjb(DE3) was obtained from Zymo Research (Irvine, CA). *E. coli* strains DLF_R002 and DLF_R003 were constructed as previously reported. (Menacho-Melgar, Ye, et al., n.d.) The autolysis/autohydrolysis strain: DLF_R004 was constructed using synthetic DNA. Briefly, linear DNA (gBlock, IDT Coralville, IA) was obtained with the Lamba lysozyme and endonuclease operon driven by a yibDp phosphate controlled promoter, preceded by a strong transcriptional terminator and followed by an apramycin resistance marker (Figure 8b). (Mairhofer et al., 2015; J. Yang et al., 2014) The *nucA* reading frame included its native N-terminal secretory signal (‘MRFNNKMLALAALLFAAQAS’). (Ball et al., 1987; Benedik & Strych, 1998) These sequences were flanked by homology arms targeting the deletion of the *ompT* protease . This cassette (sequence supplied in Supplemental Materials) was directly integrated into the genome of strain DLF_R002 via standard recombineering methodology. (Sharan et al., 2009) The recombineering plasmid pSIM5 was a kind gift from Donald Court (NCI,

https://redrecombineering.ncifcrf.gov/court-lab.html). *OmpT* deletion and autolysis/autohydrolysis operon integration was confirmed by PCR amplification and sequencing (Genewiz, NC). Plasmid pHCKan-yibDp-GFPuv (Addgene #127078) was constructed as previously reported. (Menacho-Melgar, Ye, et al., n.d.)

3.2.3 Cell Growth & Expression

Shake flask cultures, BioLector™ studies, microfermentations (microtiter plate cultivations) and 1L instrumented fermentations were performed as described in Menacho-Melgar et al.. (Menacho-Melgar, Ye, et al., n.d.) Briefly, batch cultures utilized autoinduction broth (AB Media) and fermentations were performed using FGM10 media. Shake flask expression were performed at 150 rpm in baffled 250 mL Erlenmeyer flasks, with 20 mL of culture.

3.2.4 Lysis Measurements

DLF_R003 and DLF_R004 strains bearing plasmid pHCKan-yibDp-GFPuv were grown in LB overnight and later used to inoculate 250 mL shake flasks containing AB Media. After 24 hours, cells were harvested by centrifugation at 4000 rpm at 4 °C and resuspended in lysis buffer. Cultures were aliquoted in 1 mL samples. Lysis buffer consisted of either Buffer 1 or Buffer 2. Buffer 1 was used when hydrolysis was not needed and Buffer 2 for autohydrolysis. Buffer 1: phosphate buffer saline pH 7.4 (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4) supplemented with 0.1% Triton-X100 and 1x Halt Protease inhibitors (ThermoFisher Scientific, Waltham, MA). Buffer 2: 20 mM Tris, pH 8.0, 2mM $MgCl_2$ supplemented with 0.1% Triton-X100 and 1x

Halt Protease inhibitors. Cells were resuspended in 1/10 to ½ the original culture volume (in the case of MTPs). To lyse, cells were incubated in ice (0 °C experiments), preheated heat blocks (25 and 37 °C experiments) or prechilled tube racks (-20 °C and -60 °C experiments) for the indicated time. After lysis, samples were centrifuged at 4 °C at 13 000 rpm for one minute. Fluorescence readings were performed using a Tecan Infinite 200 plate reader in black 96 well plates (Greiner Bio-One, reference 655087) using 200 µl. Samples were excited at 412 nm (OmegaOptical, Part Number 3024970) and emission was read at 530 nm (OmegaOptical, Part Number 3032166) using a gain of 60. Fluorescence values were normalized to complete soluble protein release as obtained from sonicating one sample of each flask using a needle sonicator at 50% power output and 10s/30s on/off cycles for 20 minutes. Under these conditions, we found no more protein release with further sonication.

3.2.5 DNA Hydrolysis

DLF_R004 and DLF_R004 plus pHCKan-yibDp-GFPuv strains were grown overnight in LB. Overnight cultures were used to inoculate 20 mL of AB , in a 250 mL Erlenmeyer flask at 1% v/v in triplicate. Antibiotics were added as appropriate. Cultures were grown for 24 hours at 37 °C and 150 rpm. Cells were harvested by centrifugation and resuspended in 2 mL of Lysis/Hydrolysis Buffer (20mM Tris, pH 8.0, 2mM $MgCl_2$, 0.1% Triton-X100, with or without 50mM EDTA). After resuspension, cells were subjected to a single freeze thaw at -20 °Celsius. Following freeze thaw samples were incubated at 37 degrees Celsius. Samples were taken every 20 minutes, and in the no EDTA reaction, EDTA was added to a final concentration of 50mM. In the case of

DLF_R004, samples were clarified by centrifugation and the supernatants analyzed via agarose gel electrophoresis. In the case of DLF_R004 plus pHCKan-yibDp-GFPuv, as GFPuv is also visualized under UV light used to visual agarose gels, after initial lysate clarification and supernatant sampling for GFPuv release, samples were heat denatured at 95 °Celsius for 5 minutes, and then clarified again by centrifugation and the supernatants analyzed via agarose gel electrophoresis.

3.2.6 Microtiter Plate Expression and Autolysis

96 well plate expression studies again utilized AB media according to Menacho-Melgar et al. using 100 µl of culture volume. (Menacho-Melgar, Ye, et al., n.d.) After 24 hours of growth in AB, cells were harvested using a Vpsin (Agilent) plate centrifuge for 8 minutes at 3000 rpm. Supernatant was removed using a Biotek Plate washer/filler. 50 µl of Lysis Buffer (Buffer 1 above) was added, cells were resuspended by shaking, and placed at -60 °C for 30 minutes. After freezing cells were thawed for 10 minutes at 25 °C, and lysates clarified by centrifugation again using a Vpsin plate centrifuge for 8 minutes at 3000 rpm. 5 µl lysate (supernatant) was collected and diluted 40-fold for analysis of GFPuv levels.

3.2.7 Strain Stability Measurements

DLF_R003, DLF_R004 and Xjb(DE3) strains were grown overnight in LB. Overnight cultures were used to inoculate 5 mL of LB at 2% v/v in triplicate. The new cultures were grown for 2-3 hours at 37 °C and 150 rpm until 0.6-0.8 OD600 was reached. At this point, samples were taken, diluted 250,000-fold and 50 µL were plated in

LB agar plates. The rest of the cells were made electrocompetent by washing twice with 1 mL ice-cold water and once with 1 mL ice-cold glycerol. Cells were then frozen for 2 hours at -60 °C. After thawing, samples were again diluted and plated as described above. Colonies were counted after incubating the agar plates overnight.

3.3 Results

3.3.1 Impact of Autolysis/Hydrolysis Modules on Growth and Protein Expression

After the construction of a modified strain (DLF_R004), with integrated, phosphate regulated lysozyme and endonuclease (Figure 8b), we sought to first evaluate any negative impact these modifications may have on growth and autoinduction of heterologous protein expression. Toward this aim, we evaluated our autolysis/hydrolysis strain as well as its parent lacking any lysozyme or endonuclease for growth and protein expression in autoinduction broth, in the MP2 Labs BioLector ™ (where biomass and protein expression can be monitored). (Menacho-Melgar, Ye, et al., n.d.) Specifically, cells of either strain DLF_R004 (our autolysis/hydrolysis strain) or its parent DLF_R003, were transformed with plasmid pHCKan-yibDp-GFPuv enabling the low phosphate induction of GFP. As can be seen in Figures 2a/b, no significant difference in growth and/or protein expression was observed when the autolysis/hydrolysis module was present. We then further investigated the impact of this module in instrumented bioreactors in minimal autoinduction media, wherein active agitation results in increased shear stresses compared to smaller scale systems. Again, as can be seen in Figure 9c, no

significant difference in growth and/or expression was observed indicating strain stability at least to this level of shear.

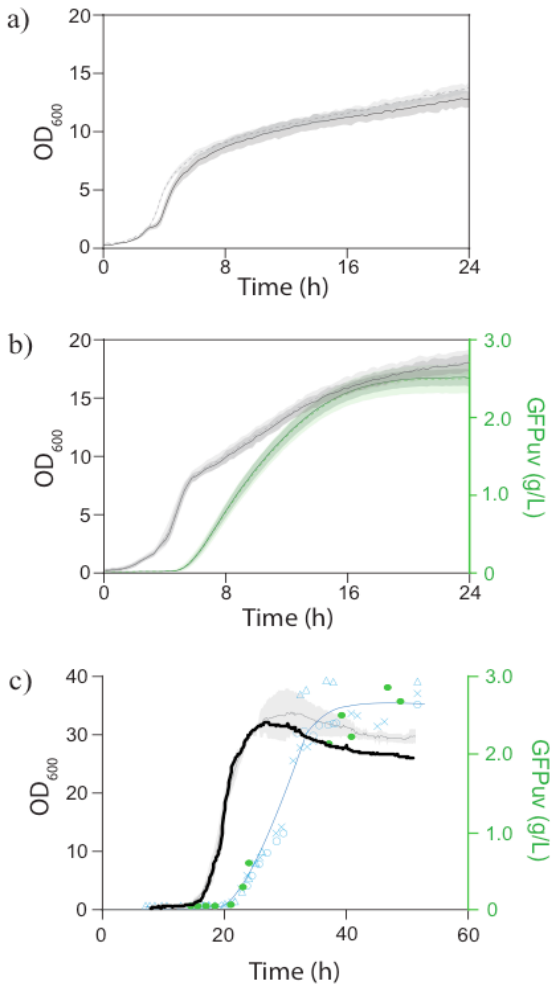

Figure 9: Growth and autoinduction of the autolysis/hydrolysis strain (DLF_R004) compared to a non-autolytic control (DLF_R003).Black and gray lines indicate biomass levels (OD 600nm) and blue and green lines and markers indicate GFP levels. Shaded areas indicate the standard error of replicates. a) growth of strains DLF_R004 and DLF_R003 in autoinduction broth in the M2P Labs BioLector™. DLF_R003 - dashed line, DLF_R004 - solid line. Results are of 6 replicates. b) growth and autoinduction of strains DLF_R003 and DLF_R004 both carrying the autoinducible GFP reporter plasmid pHCKan-yibDp-GFPuv, in autoinduction broth in the M2P Labs BioLector™.

3.3.2 Autolysis

After demonstrating equivalent expression with no significant growth defects, we next turned to validate the autolysis behavior of our engineered strain as shown in Figure 10 below. Shake flask cultures were started in autoinduction broth (AB), and the cells were harvested by centrifugation post cell growth and GFP autoinduction. Cell pellets were washed, and triton-X100 was added at 0.1%. GFP release was measured over time by centrifugation and measurement of fluorescence in the supernatant (Figure 10a). The addition of 0.1% triton-X100 was found to be sufficient for the release of ~ 55% of the total GFP in about an hour. No GFP release was observed either in the control strain or in our autolysis/hydrolysis strain without triton-X100 addition. Increasing triton-X100 levels did not impact protein release (Appendix B, Figure 36). To further optimize protein release, we evaluated the impact of a freeze-thaw cycle (Figure 10b) on autolysis. Freeze-thaw is well known to provide cell wall and membrane disruption. As can be seen in Figure 10b, a single 30 minute freeze-thaw after the addition of 0.1% at -20 degrees Celsius, led to >90% release of GFP. Without triton-X100 addition, the same freeze thaw led to only ~ 10% GFP release (Appendix B, Figure 37).

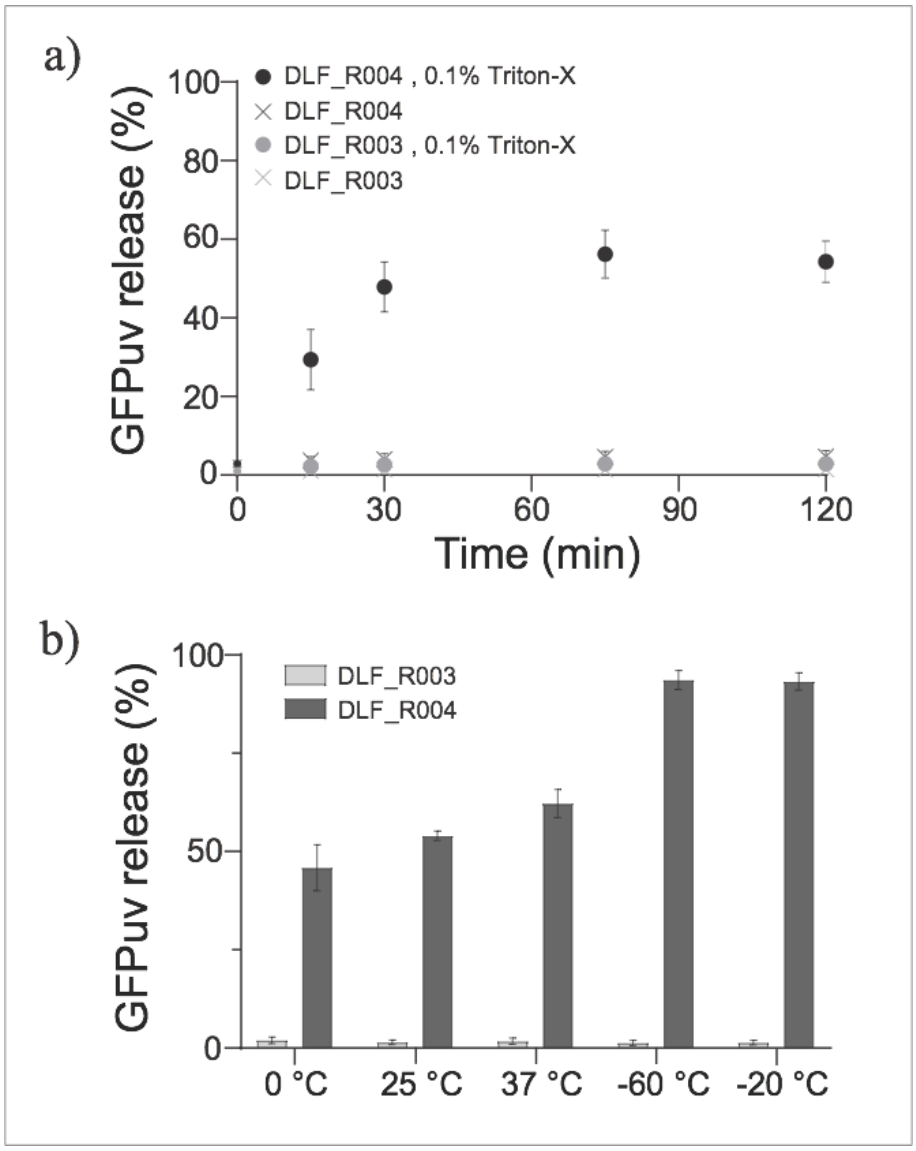

Figure 10: Autolysis and protein release of strains DLF_R004 (autolysis/hydrolysis strain) and DLF_R003 (control). a) Autolysis and GFPuv release as a function of time after the addition of Triton-X100, cells were incubated at room temperature (25 Celsius). b) Autolysis and GFPuv after the addition of 0.1% Triton-X100 and incubation for 30 minutes, on ice (0 Celsius), room temperature (25 Celsius), 37 Celsius, and a 30 minute freeze thaw at either at -60 or -20 Celsius.

3.3.3 Autohydrolysis

We next turned to validate the autohydrolysis conferred by the endonuclease in our autolysis/hydrolysis strain. To accomplish this we measured DNA/RNA hydrolysis as a function of time during cell autolysis. In the case of hydrolysis, cell lysates were more concentrated to be able to measure differences in DNA concentrations. Cell pellets were resuspended in 1/10th culture volume of 20mM Tris buffer (pH=8.0), plus 2mM $MgCl_2$. As a control, EDTA (50mM) was optionally added prior to freeze thawed pellets to inhibit endonuclease. Cell pellets were treated with 0.1 % triton-X100 followed by a single 30 minute freeze thaw. After freeze thaw samples were incubated at 37 °Celsius and samples taken to evaluate hydrolysis. 50mM EDTA was added to samples to inhibit nuclease activity before analysis. Relative levels as well as the size of DNA/RNA were measured both by agarose gel electrophoresis. Results are given in Figure 11a and 11b below. DNA hydrolysis, occurs in parallel with autolysis, and visible DNA/RNA was gone within 60 minutes of initiating autolysis. This protocol (with more concentrated lysate) was then evaluated for protein release using GFPuv, results of which are given in Figure 11c, (Refer to Appendix B, Figure 38 for DNA analysis) leading to a recommended routine expression and autolysis/hydrolysis protocol for shake flask cultures (outlined in Appendix B, Figure 39). Refer to Appendix B, Figure 40 for an example lysate generated using this protocol.

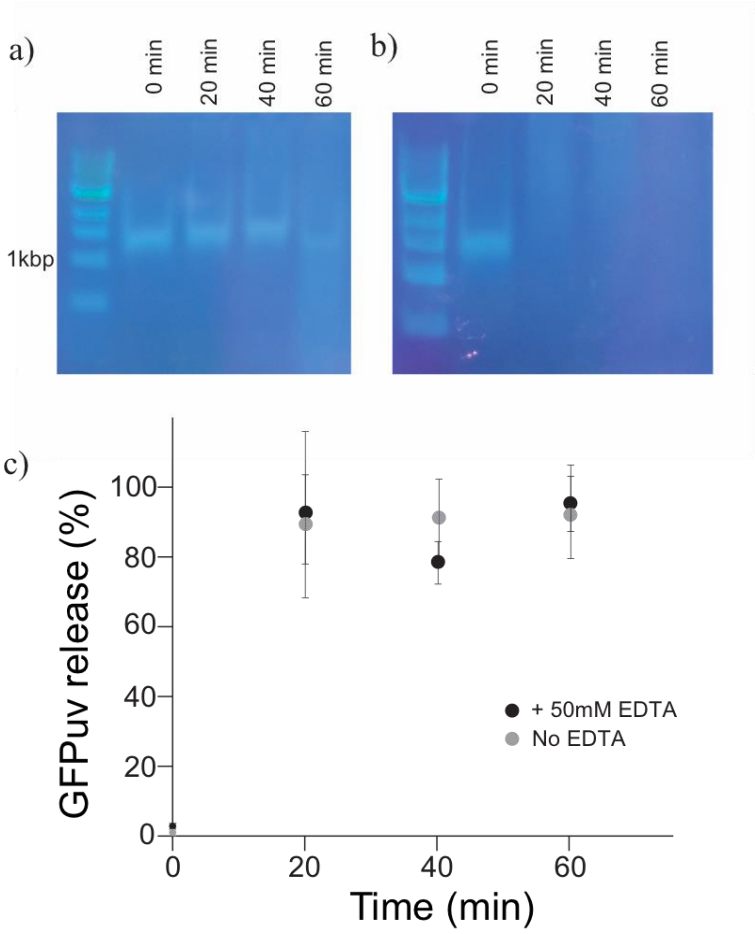

Figure 11: (a,b) Autohydrolysis of DNA/RNA of strain DLF_R004. Time course of DNA/RNA hydrolysis with (a) or without (b) EDTA, which inhibits endonuclease by chelating Mg+2. The distinct "band" at ~ 1500 bp is presumably rRNA and/or sheared genomic DNA. c) Time course of autolysis and GFPuv release in under autohydrolysis conditions using strain DLF_R004 bearing plasmid pHCKan-yibDp-GFPuv.

3.3.4 High throughput Autolysis/Hydrolysis

To build upon the successful autolysis and autohydrolysis observed in cells harvested from shake flask cultures, we moved to validate this approach with high throughput microtiter-based expression. Autolysis/hydrolysis in microtiter plates has the potential to greatly simplify high throughput screening of proteins in crude lysates as well

as proteins purified from crude lysates. As illustrated in Figure 12, autolysis and protein release successfully scaled down to MTPs.

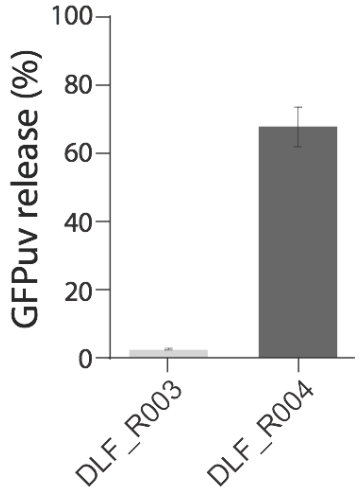

Figure 12: Autolysis and protein release in 96 well microtiter plates.

3.3.5 Stability of Uninduced Cells

A challenge with several current autolysis strains is sensitivity to free thaw during routine workflows, presumably due to leaky expression of the lysis proteins. And while DLF_R004 has demonstrated stability in autoinduction cultures, we wanted to confirm that autolysis did not occur during routine freeze thaw cycles such as those used in preparing electrocompetent cells where not only are cells frozen and thawed but also thoroughly washed to remove ions including magnesium ions. We tested the stability of electrocompetent cells for both DLF_R004 as well as another well known, readily available autolytic strain of E. coli , strain Xjb(DE3) from ZymoResearch. Xjb(DE3) relies on arabinose induction to induce lysozyme and autolytic behavior. (Jia et al., 2011) In addition, the manufacturer recommends that excess magnesium is added to routine cultures to stabilize the cell wall of these cells, which is not feasible when preparing

electrocompetent cells. Results of these competent cell studies are given in Figure 13. While strain Xjb(DE3) suffered from unwanted lysis in these studies, DLF_R004, with tight control over expression of lysozyme and endonuclease had increased stability during this process.

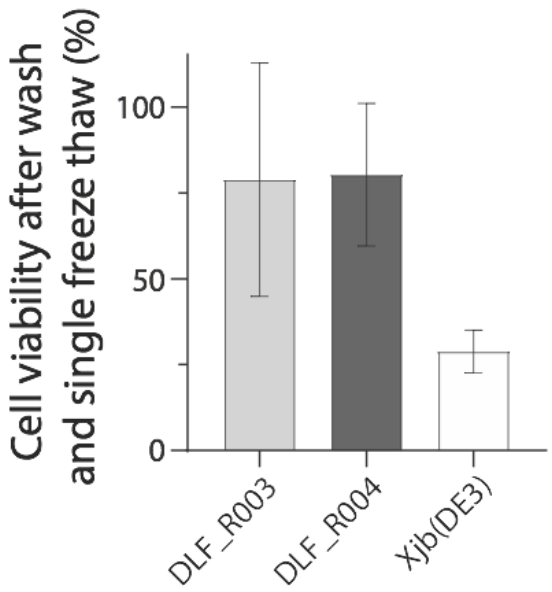

Figure 13: Stability of uninduced strain DLF_R004 (autolysis/hydrolysis strain), DLF_R003 (control) and autolysis strain E. coli Xjb. Percent viability was measured after washing with ice-cold water (twice), ice-cold 10% glycerol (once) and a single freeze thaw. Viability was measured as colony forming units after the freeze thaw normalized to colony forming units before freeze thaw, multiplied by 100%.

3.4 Discussion

In this study we have demonstrated the development of an improved strain of *E. coli* for not only autoinduction of protein expression but also of lysozyme and endonuclease enabling combined autolysis and auto DNA/RNA hydrolysis. To our knowledge this is the first combination of these two mechanisms to improve cellular lysis and DNA removal, and an example of the potential benefits of two stage production. (Burg, Cooper, Ye, Reed, Moreb, et al., 2016; M. D. Lynch et al., 2017; Michael D.

Lynch, 2016) This system enables > 95% lysis and hydrolysis. Due to tightly controlled expression these strains are stable to shear forces in stirred tank bioreactors and even when subjected to freeze thaw cycles in deionized water, with 10% glycerol. Complete autolysis/hydrolysis allows for simplified liquid handling automation, useful in high throughput screening protocols. The mild detergents (0.1% triton-X100) used are also compatible with high throughout SDS-PAGE alternatives including capillary electrophoresis systems, (Rodionova et al., n.d.) and while an optimal buffer was used for optimal autohydrolysis, most buffers are compatible with lysozyme activity and protein release. In commercial production, the autoinduction of endonuclease can remove the need to purchase nucleases for DNA removal and simplify purification and reduce costs.

There are some remaining challenges to the use of the current system in several key applications, due to the use of endonuclease. endonuclease is difficult to inactivate and only denatured under conditions that most likely will impact the activity of any protein of interest. (Z. Yang et al., 2013) As a result, subsequent purification is needed to remove endonuclease. This is likely not an issue for routine shake flask expression or commercial scale production where additional downstream purification steps are expected, but can complicate high throughput screens where lysate assays can be performed and complete purification may not otherwise be needed. This is likely only an issue with high throughput evaluation of enzymes, where retained nuclease activity would be a problem, such as for DNA/RNA modifying enzymes. In these applications additional purification will be needed. Despite these limitations, the method is well suited

for routine shake flask expression and protein purification, as well as larger scale production. In addition, the approach may have applicability to the production of other intracellular products beyond proteins including polyhydroxyalkanoates (PHAs). (Andreessen et al., 2014; Borrero-de Acuña et al., 2017)

4. Robust Protein Expression Enabled by Dynamic Control over Host Proteases

4.1 Introduction

Recombinant protein expression is critical in many aspects of basic and applied research in numerous areas of biotechnology.(Baeshen et al., 2015; Jozala et al., 2016; Sanchez-Garcia et al., 2016) Additionally, together industrial and clinical proteins comprise a growing multi-billion dollar annual global market. (Baeshen et al., 2015; Jozala et al., 2016; Sanchez-Garcia et al., 2016) Numerous efforts over the past several decades have been focused on engineering improved production hosts for both specific proteins as well as proteins in general, including *E. coli*, a mainstay for protein expression. Despite many advances and the ease of many cloning and expression protocols, numerous proteins are still difficult to express in *E. coli*. (Saccardo et al., 2016) As a result, significant effort can be spent optimizing expression protocols, such as medium, induction time, level and temperature to enable adequate expression. (Gutiérrez-González et al., 2019; Kaur et al., 2018) In many cases, it is often easier to try to engineer the protein (Hussain et al., 2017; Saccardo et al., 2016) or to give up on *E. coli* and screen several additional expression hosts such as *Bacillus Saccharomyces, Streptomyces,* or mammalian cell lines such as Chinese hamster ovary (CHO) or human embryonic kidney (HEK), or even *in vitro* cell free expression. (Alves & Dobrowsky, 2017; Carlson et al., 2012; Cui et al., 2018; Gomez-Escribano & Bibb, 2012; Jin & Hong, 2018; Thoring et al., 2017) While these additional expression systems have their own advantages and disadvantages, and are not conceptually more complicated than expression in *E. coli*, they

can add significant cost and time to obtaining purified protein. These alternative hosts are often impractical for a lab without prior experience and can necessitate significant investment in developing working expertise with a new expression platform. Improved *E. coli* hosts that overcome challenges with "hard to express" proteins are of use to many labs working with alternative expression systems.

While it can be difficult to define a "hard to express" protein, for the sake of this discussion we will define a few classes of proteins that have proven difficult in *E. coli*: toxic proteins, proteins that are slow to fold, and large proteins. (Rosano & Ceccarelli, 2014a) Toxic protein can be defined as those those whose induction causes significant growth arrest leading to reduced growth and expression.(Giacalone et al., 2006; Saïda et al., 2006) Proteins that are slow to fold, may normally rely on chaperones or cofactors for rapid folding in native hosts, and when expressed in *E. coli* may lead to partially folded states that are recognized and degraded by *E. coli* proteases prior to maturation. (Jones et al., 2002) Large proteins as the name implies are proteins nominally larger than 100 kDa that can be difficult to express, and may make up a special class of slow folders, wherein protein size alone reduces rates of maturation, or where other structures may lead to recognition and unfolding and or degradation. Large proteins are of particular interest in the fields of natural products wherein many polyketide synthases and non-ribosomal peptide synthases are large and can be difficult to express in *E. coli*. (Liu et al., 2017; Weber, 2019)

Most studies looking at protein expression have been complicated by induction of heterologous protein expression during the growth phase, such as in mid-exponential

phase. While this traditional methodology would seem to leverage a highly productive cellular state for protein synthesis it leads to competition between cellular growth and heterologous production as well as potentially to cellular toxicity and decreased growth rate. We have recently demonstrated the tightly controlled production of heterologous proteins in stationary phase, where expression is auto-induced by phosphate depletion, leading to relatively high protein titers. (Menacho-Melgar, Ye, et al., n.d.) These initial results confirm the utility of stationary phase expression. In addition, in these studies we evaluated the relative impact of Lon protease activity on the expression of a Lon substrate, when two stage expression was compared to expression using BL21(DE3), a known Lon deficient strain. To our surprise, although we did observe expression of Lon substrates, the levels were far from optimal when compared to non Lon substrate controls. These results indicated that there was room to further reduce the activity of Lon, and protease activity more generally, to improve the expression of heterologous proteins that may be substrates of native housekeeping proteases.

4.2 Materials and Methods

4.2.1 Reagents and Media

Unless otherwise stated, all materials and reagents were of the highest grade possible and purchased from Sigma (St. Louis, MO). Luria Broth, lennox formulation with lower salt was used for routine strain and plasmid propagation and construction and is referred to as LB below. Working antibiotic concentrations were as follows:

kanamycin (35 µg/mL) and apramycin (100 µg/mL). Autoinduction Broth (AB) and FGM10 media were prepared as previously reported. (Menacho-Melgar, Ye, et al., n.d.)

4.2.2 Strains and Plasmids

Strains DLF_R002 and DLF_R003 were constructed as previously reported. (Menacho-Melgar, Ye, et al., n.d.) Strains DLF_R0010, DLF_R0011 and DLF_R0012 were obtained using linear DNA that have homologies to the target genes (lon, ftsh and hslUV, respectively) and encode a das4 tag for lon and ftsh with antibiotic markers (DLF_R0010 and DLF_R0011, respectively; linear DNA purchased from IDT Coralville, IA, as gBlocks) or an antibiotic marker for tet for a clean knock-out (DLF_R0012, linear DNA was generated using PCR). DLF_R0012 was made from DLF_R0011 which was made from DLF_R0010. These cassettes (sequences supplied in Supplemental Materials) were directly integrated into the genome via standard recombineering methodology. (Sharan et al., 2009) The recombineering plasmid pSIM5 was a kind gift from Donald Court (NCI, https://redrecombineering.ncifcrf.gov/court-lab.html). Plasmids pHCKan-yibDp-GFPuv (Addgene #127078), pTKhc-yibDp-GFP-β20cp6 (Addgene #127060) and pTKhc-yibDp-GFP-cp6 (Addgene #134938) were constructed as previously reported. (Menacho-Melgar, Ye, et al., n.d.) pBT1-mcherry and pBT1-mcherry-das4 plasmids were assembled using Gibson Assembly. mcherry was ordered as gblock from IDT, pBT1 backbone was PCR amplified using SR2F and SL1R primers using Q5 master mix from NEB following manufacturer's protocols. Plasmids pCASCADE-proD, pCASCADE-lon, pCASCADE-ftsHp1, pCASCADE-ftsHp2, pCASCADE-ftsHp3 and pCASCADE-ftsHp4 were obtained using "Around the World" PCR of plasmid pCASCADE-EV.

Strains and plasmids were confirmed by PCR amplification and sequencing (Genewiz, NC).

4.2.3 Cell Growth & Expression

Shake flask cultures, BioLector™ studies, microfermentations (microtiter plate cultivations) and 1L instrumented fermentations were performed as described in Menacho-Melgar et al.. (Menacho-Melgar, Ye, et al., n.d.) Briefly, batch cultures utilized autoinduction broth (AB Media) and fermentations were performed using FGM10 media. Shake flask expression was performed at 150 rpm in baffled 250 mL Erlenmeyer flasks, with 20 mL of culture. BioLector™ studies were performed with GFP filter at gain 20 (ex: 488 nm, em: 520 nm) and a RFP filter at gain 60 (ex: 580 nm, em: 610 nm). Fluorescence levels from microfermentations were measured using a Tecan Infinite 200 plate reader. Measurements were performed using 200 uL in black 96 well plates (Greiner Bio-One, Reference Number: 655087). Samples were excited at 412 nm (Omega Optical, Part Number: 3024970) and emission was read at 530 nm (Omega Optical, Part Number: 3032166) using a gain of 60.

4.2.4 SDS-PAGE and mirA quantification

The OD_{600nm} of culture samples to be analyzed was measured before harvesting the cells by centrifugation, which was done at 4000 rpm for 15 minutes. The cells were resuspended in 50 µl of phosphate-buffered saline with protease inhibitors (ThermoFisher Scientific, MA, product number A32965) and 5 mM EDTA. 25 µl of the resuspended cells were mixed with 25 µl of 2x Laemmli sample buffer (Biorad, CA) and boiled for 5

minutes at 95 ℃. The boiled samples were centrifuged at 14,000 rpm for 10 minutes and 20 µg of total protein per sample was then loaded into a 4-15% gradient Mini-Protean TGX precast protein gel (Biorad, CA). The gels were stained using Coomassie Brilliant Blue R-250. Gels were imaged using a UVP PhotoDoc-It™ Imaging System (Analytik Jena, CA) and expression levels were quantified using ImageJ (NIH, MD).

4.3 Results

4.3.1 Two-Stage Synthetic Metabolic Valves

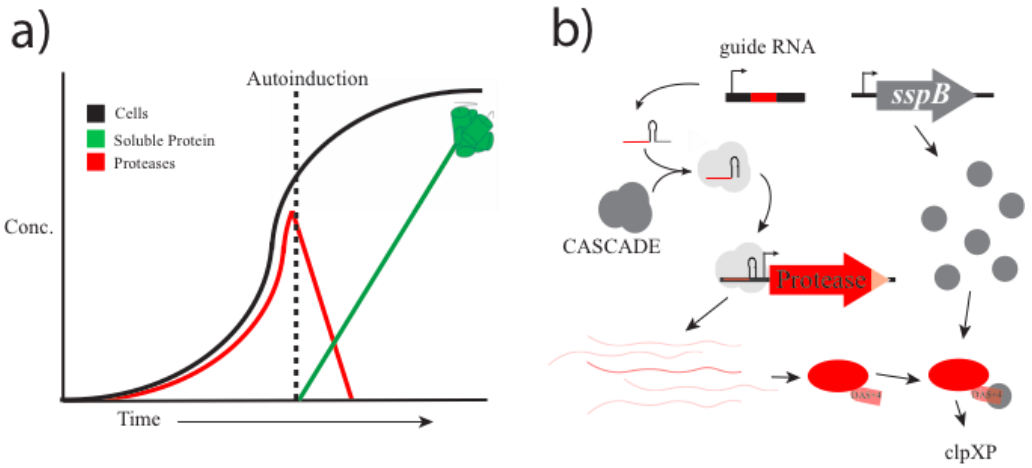

Figure 14: 2-stage dynamic control over host machinery such as proteases. a) Time course of 2-stage dynamic control, cells (black) line grow until a limiting nutrient is depleted, in this case phosphate. As phosphate is depleted autoinduction of recombinant protein expression (green) occurs in parallel with reduction in key enzymes such as housekeeping proteases. b) Dynamic turnover of housekeeping proteases can be achieved through i) CRISPR mediated gene silencing and controlled proteolysis alone and or in combination. gRNA(s) are induced upon nutrient depletion and along with the native nuclease deficient Cascade complex of E. coli (with a deletion of cas3 the nuclease component of the Cascade complex) bind to targeted promoters reducing gene expression. In parallel, the protein of interest for turnover is modified to have a C-terminal degron tag. Upon nutrient depletion induction of the sspB gene produces a chaperone binding to the degron tag and targeting the protein for clpXP mediated proteolysis.

We first developed synthetic metabolic valves (SMVs), which are capable of the dynamic reduction of protein levels in a 2-stage process. These SMVs rely on controlled

proteolysis or CRISPRi/Cascade based gene silencing or both proteolysis and silencing in combination (Burg, Cooper, Ye, Reed, & Moreb, 2016; Luo et al., 2015; Michael D. Lynch, 2016; McGinness et al., 2006; Menacho-Melgar, Ye, et al., n.d.; Qi et al., 2013) to reduce levels of key metabolic enzymes (Figure 14b). Cell growth and dynamic metabolic control is implemented using phosphate depletion as an environmental trigger, in autoinduction broth as previously described. (Menacho-Melgar, Ye, et al., n.d.) It has recently been shown that deleting the *cas3* gene and expressing guide RNAs can lead to gene silencing. (Luo et al., 2015) Also, targeted proteolysis was controlled by linking the expression of the chaperone *sspB* to phosphate deprivation. *SspB*, when induced, binds to C-terminal DAS+4 (DAS4) peptide tags on any target protein and causes degradation by the clpXP protease of *E. coli* (Figure 14b). (McGinness et al., 2006) These systems were implemented in a host strain we previously reported, which was engineered for minimal byproduct formation and high biomass yields and growth rates, (Menacho-Melgar, Ye, et al., n.d.) by first deleting the native *sspB* gene, followed by replacing the cas3 gene with a low phosphate inducible *sspB* construct, followed by a constitutive promoter to drive expression of the downstream Cascade operon. These efforts resulted in an *E. coli* host strain (DLF_Z0025) capable of wild type growth rates, high biomass yields, and the low phosphate induction of proteolytic and CRISPR/Cascade machinery. Refer to Table 2, for a complete list of strains and plasmids used in this study. Using this system, as Figure 15 demonstrates, protein levels can be controlled in a 2-stage processes using controlled proteolysis alone, as exemplified by turning “ON’ GFP and “OFF” mCherry fluorescent proteins with phosphate depletion in autoinduction broth. Unfortunately, while, in this

case, gene silencing results in the lowest levels of mcherry degradation (Figure 15c), leaky transcription of silencing guides results in constitutive silencing as opposed to a 2-stage process, as evidenced by low mcherry fluorescence before induction. Efforts to suppress leaky silencing and their conversion to a 2-stage process are on-going. Moreover, the impact of each approach (proteolysis or silencing) will vary dependent on the target gene/enzyme and its specific natural turnover rates and expression levels. (Grünenfelder et al., 2001; Hintsche & Klumpp, 2013)

Table 2: Plasmids and strain used in Chapter 4

Plasmid	**Insert**	**promoter**	**ori**	**Marker**	**Addgene**	**Source**
pSMART-HC-Kan	None	None	colE1	Kan	NA	Lucigen
pHCKan-yibDp-GFPuv	GFPuv	yibDp	colE1	Kan	127078	(Menacho-Melgar, Ye, et al., n.d.)
pBT1-proDp-mCherry	mCherry	proD	BBR1	Amp		This study
pTKhc-yibDp-GFP-β20cp6	GFP-β20-cp6	yibD	colE1	Kan	127060	This study
pTKhc-yibDp-GFP-cp6	GFP-cp6	yibD	colE1	Kan	134938	This study
pCASCADE-EV	None	ugpBp1	p15a	Cm	65821	This study
pCASCADE-proD	proDp gRNA	ugpBp1	p15a	Cm	65820	This study
pCASCADE-lon	lon	ugpBp1	p15a	Cm		This study
pCASCADE-ftsHp1	ftsH	ugpBp1	p15a	Cm		This study
pCASCADE-ftsHp2	ftsH	ugpBp1	p15a	Cm		This study
pCASCADE-ftsHp3	ftsH	ugpBp1	p15a	Cm		This study

pCASCADE-ftsHp4	ftsH	ugpBp1	p15a	Cm		This study
Strains used in this study						
Strain	**Genotype**					**Source**
DLF_R002	F-, λ-, Δ(araD-araB)567, lacZ4787(del)(::rrnB-3) , rph-1, Δ(rhaD-rhaB)568, hsdR514, ΔackA-pta, ΔpoxB, ΔpflB, ΔldhA, ΔadhE, ΔiclR, ΔarcA					(Menacho-Melgar, Ye, et al., n.d.)
DLF_Z001	DLF_R002, ΔsspB::frt					This study
DLF_Z002	DLF_R002, Δcas3:: pro-casA					This study
DLF_Z0025	DLF_Z001, Δcas3:: ugpBp-sspB-pro-casA					This study
DLF_R0010	DLF_Z001, Δcas3:: ugpBp-sspB-pro-casA, lon-DAS4:zeoR					This study
DLF_R0011	DLF_Z001, Δcas3:: ugpBp-sspB-pro-casA, ftsH-DAS4:ampR					This study
DLF_R0012	DLF_Z001, Δcas3:: ugpBp-sspB-pro-casA, lon-DAS4:zeoR, ftsH-DAS4:ampR, ΔhslUV::tetR					This study
Res - resistance marker, ori- origin of replication,, Cm- chloramphenicol, Kan - kanamycin, Amp - ampicillin, tet - tetracycline, zeo- zeocin						

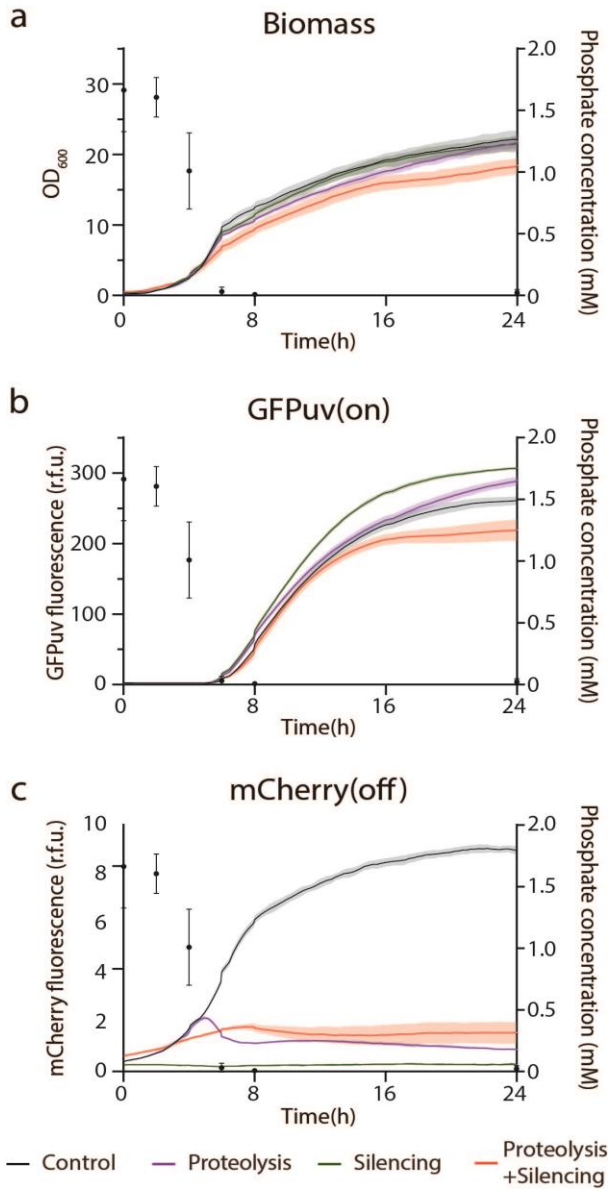

Figure 15: 2-stage dynamic control over fluorescent proteins in autoinduction broth. A series of DLF_Z0025 derivatives with autoinducible GFP were evaluated in autoinduction broth for 2-stage dynamic control. These included: a no mCherry control (Empty, gray line), an mCherry with an empty gRNA silencing vector (Cont. black line), an mCherry with a degron tag (Proteolysis, purple line), a gRNA silencing the mCherry promoter (Silencing, green line) as well as an mCherry with a degron tag and a gRNA silencing the mCherry promoter (S+P, red line). a) Growth, b) GFP expression and c) mCherry levels over time. Phosphate levels are shown as dots with corresponding error bars, for each plot.

4.3.2 Dynamic control over Lon protease activity

With the demonstration of dynamic protein turnover in a 2-stage process with autoinduction, we turned to investigate the dynamic control of the key housekeeping protease Lon. Although not strictly essential, Lon has important regulatory roles in cell division, capsule synthesis, recovery from the SOS response and general stress tolerance and protein quality control, making complete deletions nonideal. (Aertsen & Michiels, 2005; Bershtein et al., 2013; Cho et al., 2015; Dopazo et al., 1987; García-Fruitós et al., 2007; Higashitani et al., 1997; J.-H. Liao et al., 2012; Sakr et al., 2010; Schoemaker et al., 1984; Torres-Cabassa & Gottesman, 1987) Strains were constructed with either modifications to induce proteolysis or gene silencing alone or in combination. These strains were then evaluated using a previously reported fluorescent Lon substrate , **β**20-cp6-GFP, as shown in Figure 16c. (Wohlever et al., 2013) In this case, protein (fluorescence) should be increased when Lon activity is decreased, as Lon degrades the protein being induced. Results are given in Figure 16. Silencing of *lon* expression had the largest impact of protein levels, with a ~2 fold improvement. Interestingly, unlike the case of mCherry (Figure 15), proteolysis had a minimal impact of Lon activity and the combination of silencing and proteolysis was not an improvement over silencing alone (Figure 16c). This could be due to the DAS4 tag altering the natural turnover, or activity of Lon. The DAS4 may also not be accessible to the sspB chaperone or alter transcript stability and basal expression levels with the chromosomal addition of the DAS4 degron tag. Alternatively, perhaps basal Lon activity is required for optimal expression and induction. In any event, this requires further investigation. As silencing alone had a

significant impact on the expression levels of Lon substrate, future work should leverage silencing to reduce Lon activity. Importantly, silencing of Lon improved protein levels equivalent to a Lon knock-out and by ~6.5 fold over the widely used BL21(DE3) strain which is regarded as Lon deficient (Figure 16e).

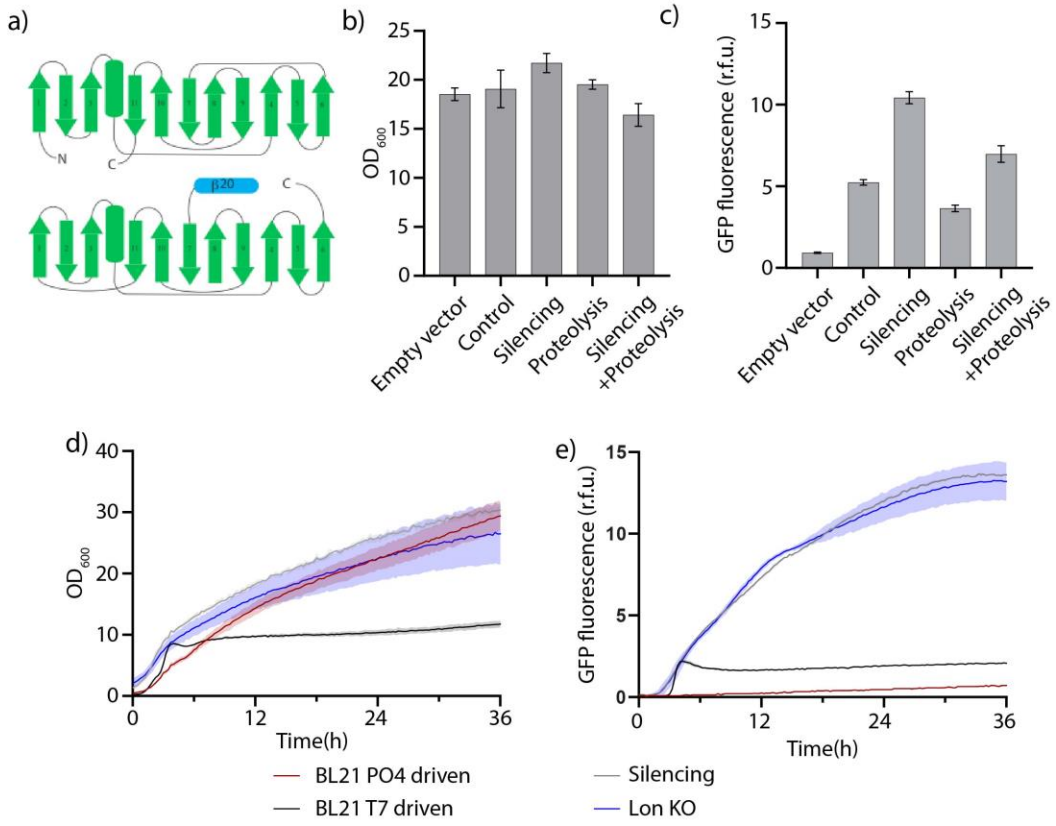

Figure 16: 2-stage dynamic control over Lon protease activity. (a) a schematic of superfolder GFP (top), and a schematic of the Beta-20 tagged circularly permuted GFP (β20-cp6-GFP) (bottom), which is a substrate for the Lon protease. Reduction of Lon protease levels was evaluated in a series of DLF_Z0025 strains by measuring fluorescence of the Lon substrate after 24 hours, we show (b) biomass levels and (c) fluorescence values. Then, we compared Lon substrate protein levels and dynamics of lon silencing (grey) to the BL21(DE3) strain bearing a plasmid for the induction of the GFP Lon substrate under phophate limitation (red) and T7-driven induction (blue), as well as a lon knock-out (blue); for which (d) biomass levels and (e) fluorescence levels are shown.

4.3.3 Dynamic control over FtsH protease activity

We next turned to the dynamic control over another key and essential, membrane anchored protease in *E. coli*, FtsH, which also serves regulatory as well as protein quality control functions. (Ito & Akiyama, 2005) FtsH is essential in *E. coli*, due to its regulation of lipid A synthesis and deletions can only be made with a suppressor mutation in the *fabZ* gene. (Katz & Ron, 2008; Ogura et al., 1999) As in the case of Lon, we engineered strains to introduce a C-terminal degron tag to FtsH, however for the *ftsH* gene, there are four known or anticipated promoters, and as a result 4 different silencing gRNAs were constructed (Table 2). (Herman et al., 1995; Huerta & Collado-Vides, 2003) These strains were then evaluated for changes in FtsH activity by analyzing the expression of miraculin, a small plant protein previously shown to be a substrate for FtsH. Results are given in Figure 17a below. In the case of FtsH, both silencing and proteolytic control improved micraculin expression, with silencing of the second *ftsH* promoter combined with proteolytic degradation of FtsH, having a four-fold improvement over the control and reaching expression levels of ~ 12%. However, we found that dynamic control over FtsH gave inconsistent results (data not shown) due to impaired growth when plasmids bearing FtsH substrates are present (Figure 17b). Impaired growth can result in poor expression since phosphate levels in the media may not get depleted, thereby induction is inconsistent. This may be due to small levels of basal FtsH substrate expression that may saturate or impair FtsH activity, which is essential for growth. Further characterization of this phenomena is needed.

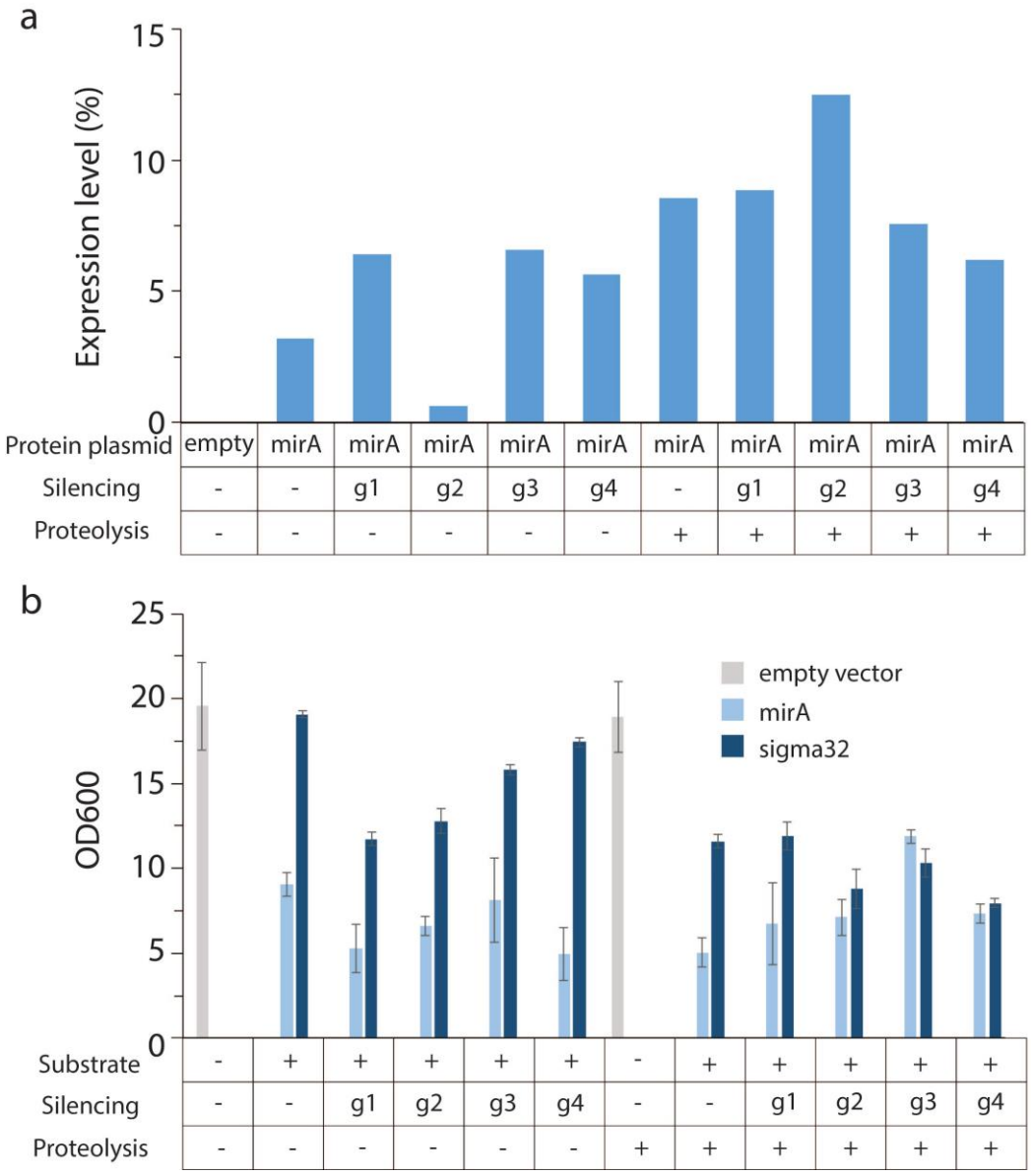

Figure 17: 2-stage dynamic control over FtsH protease activity. a) 2-stage dynamic control over the autoinducible miraculin substrate was evaluated in a series of DLF_Z0025 derivatives. These included: DLF_Z0025 (no proteolysis), DLF_R0011 a strain with a DAS4 degron tag on FtsH (+ Proteolysis), an empty vector (no miraculin) or a plasmid containing the mirA gene encoding miraculin and plasmids with either an empty gRNA (-) or silencing gRNA targeting each of the four *ftsH* gene promoters (g1, g2, g3 and g4).

4.3.4 Improved Expression of a "Difficult to Express" Protein

While further strain engineering is needed to optimize this system, as proof of concept, we constructed a strain where we added a degron tag to FsH and Lon, and deleted the nonessential HslUV, arriving at strain DLF_R0012 (Table 2). Using this strain, we expressed a large difficult to express 220kDa PKS/NRPS, salinosporamide synthase encoded by the s*alABE* genes of *Salinospora tropica.* Results are given in Figure 18b below. Dynamic control over proteases resulted in significant increases in expression to ~ 8% of total cell protein.

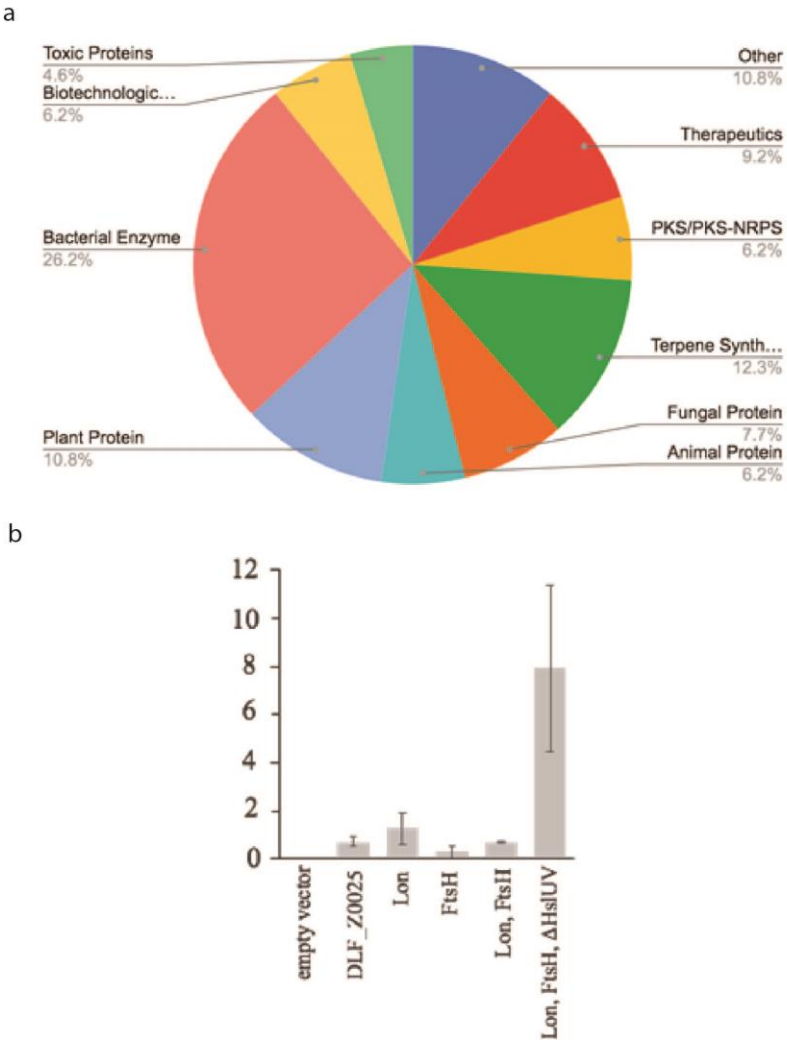

Figure 18: Improved protein expression with dynamic control over host proteases. a) A group of 60 diverse difficult to express proteins, including 6 therapeutics, 4 PKS or PKS-NRPS enzymes, 8 terpene synthases, 5 fungal proteins, 4 animal proteins, 7 plant proteins (non terpene synthases), 17 bacterial enzymes, 3 known toxic proteins and 7 additional proteins. b) 2-stage dynamic control over protease activity improves expression of the 220kDa PKS-NRPS, salinosporamide synthase from S. *tropica*. Expression of the 220kDa protein, in the control lacking dynamic control over proteases (DLF_Z0025) was compared to strains with dynamic control in Lon, FtsH, both Lon and FtsH as well as the deletion of the third major housekeeping protease, HslUV.

4.4 Discussion

Further efforts to obtain an optimized strain with reduced proteolytic activity requires the elimination of basal silencing. Re-evaluation of Lon proteolysis and silencing in such a system should be performed. Additionally, better characterization over FtsH dynamic control, as well as the possible toxicity of FtsH substrate expression is needed.

While reducing the levels of other proteases may be beneficial, it is important to note that the other two major housekeeping proteases in *E. coli*, are the ClpAP and ClpXP proteases. The ClpXP protease is instrumental for our controlled proteolysis, and so these were not manipulated in this study.

Finally, once an optimized strain is obtained, a panel of a "difficult" to express proteins will be evaluated for improved expression. As illustrated in Figure 18a, this panel consists of a group of bacterial, fungal, plant, animal proteins, as well as known toxic proteins, natural product PKS or PKS/NRPS enzymes, cytochrome P450s, terpene synthases, therapeutic molecules, and proteins of biotechnological interest such as Cas9. Each of these proteins was cloned behind the auto inducible yibDp promoter. In addition in several cases pET (T7) based expression constructs were constructed or obtained to enable a head to head evaluation with standard BL21(DE3) production.

5. A Review of Lipidation in the Development of Advanced Protein and Peptide Therapeutics

This chapter was published in Journal of Controlled Release, 295, Romel Menacho-Melgar, John S. Decker, Jennifer N. Hennigan, Michael D. Lynch, A review of lipidation in the development of advanced protein and peptide therapuetics, Pages 1-12, Copyright (2019).

5.1 Introduction

The use of biologics in the treatment of numerous diseases has increased steadily over the past several years. The biologics market grew at an annual rate of 8% in 2014, doubling the annual growth rate of conventional pharma (Otto et al., 2014). In fact, pharmaceutical companies are increasingly shifting their focus towards biologics: biologics-based revenue is expected to account for over 32% of total revenue by 2023, from 22% in 2013 (Otto et al., 2014; The Economist, 2014).

Biologics have grown as a class of therapies due to the high specificities afforded by their complex structures, often leading to reduced toxicity and side-effects (Fosgerau & Hoffmann, 2015). This specificity has allowed biologics to be approved at consistently higher rates than small molecules in recent years (Smietana et al., 2016). For example, from 2012 to 2014, biologics had twice the approval rate (18%) of small molecules (Smietana et al., 2016). Biologics also have had a lower number of withdrawals due to safety concerns at only 2% (0% if antibodies are excluded) from 1982 to 2013, compared to 3% for small molecules (Kinch, 2015).

Despite the success and future promise of biologics, peptide and protein-based drugs in general continue to face many challenges not shared with small molecules, as illustrated in Figure 19. Because peptides and proteins generally cross epithelia less readily than small molecules and are more prone to host metabolism, most biologic drugs are administered using either subcutaneous or intravenous injections. Compared to non-invasive and self-administered modes of delivery, this leads to greater costs and lower patient compliance (Figure 19a) (Cutfield et al., 2011; Hooven, 2017). Without special modifications to increase their half-lives, most peptides and proteins are cleared from the serum in the range of minutes to hours by intra- and extracellular host metabolism and renal elimination (Figure 19b) (Kontermann, 2011). This short serum half-life, coupled with slow uptake into target tissues, demands increased dose size and frequency for therapeutic effect and therefore tends to increase cost, reduce compliance and narrow the therapeutic window. An additional major limitation on the therapeutic window and efficacy of biologics is their potential for immunogenicity (Figure 19c). Protein and peptide drugs elicit anti-drug antibodies (ADAs) much more effectively than small molecules, both because their larger and more complex structures are more effective at stimulating B cells and because they are uniquely able to stimulate T cells. In turn, these ADAs and other anti-drug immune responses can reduce drug efficacy and may cause severe adverse health effects (Yin et al., 2015). Finally, biologic drugs tend to be poorly permeable across cell membranes, and routes of cellular uptake that are available often lead to drug degradation in lysosomes (Figure 19d) (Varkouhi et al., 2011).

Consequently, the vast region of potential drug target space that lies inside cells is virtually inaccessible to biologics at present.

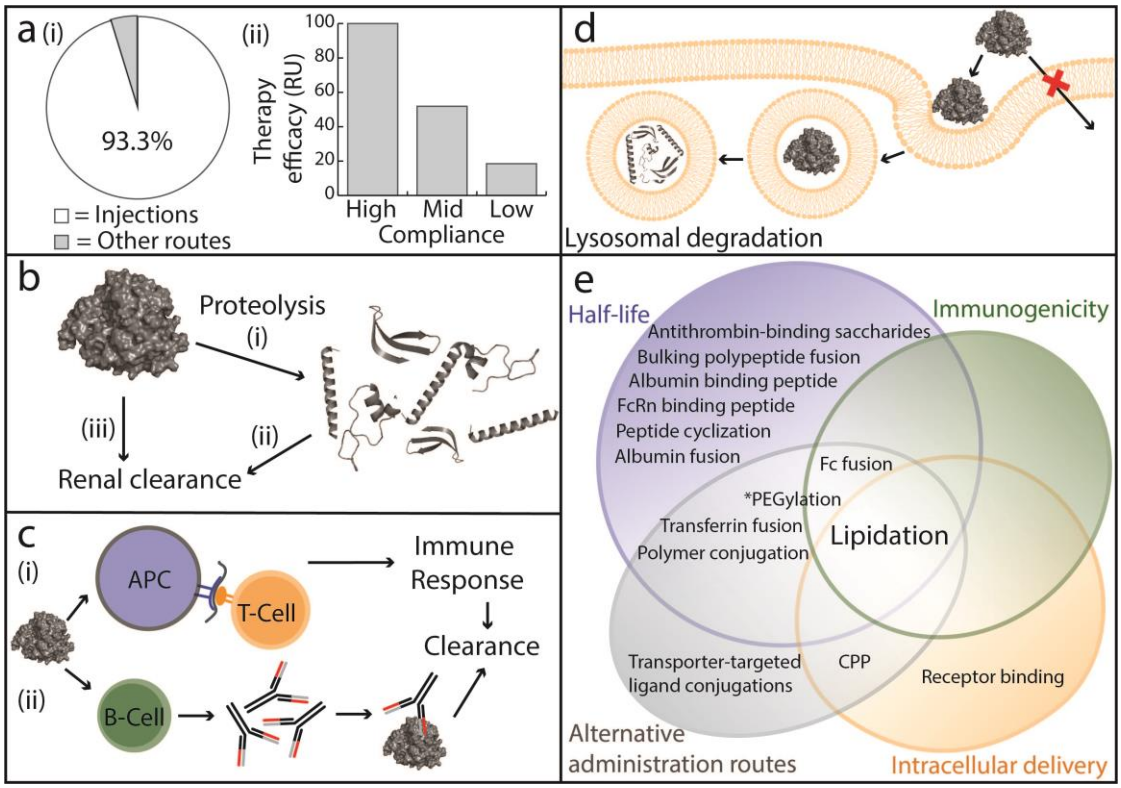

Figure 19: Current challenges and solutions facing biologic therapeutics. (a) (i) The majority of biologics are administered through injections while alternative routes including pulmonary, oral, and topical administration remain underused. (ii) Injections can lower therapy effectiveness through increased patient non-compliance (Data from (Cutfield et al., 2011)). (b) Half-life of biologic therapeutics are shortened by (i) by proteolysis and (ii & iii) renal clearance from the bloodstream. (c) Biologics can also induce immune reactions. (i) Antigen presenting cells (APC) can present peptide fragments to T-cells inducing an immune response through cytokine release that leads to biologic rapid clearance. (ii) B-cell receptors recognize protein antigens and secrete antigen specific immunoglobulin into the serum which can neutralize the therapeutic activity and mark therapeutics for clearance. (d) Biologics mainly enter cells through endocytosis. Endocytosis requires endosomal escape for the therapeutic to gain access to the cytosol without which the biologic is degraded in the lysosome. (e) Challenges and current solutions for biologic therapeutics. Approaches listed in overlapping regions address multiple challenges. *denotes controversial data on PEGylation that supports both reductions and increases in immunogenicity of biologic therapeutics.

One means of addressing these key challenges has been to covalently modify the pharmacologically active peptide or protein with other groups that confer new properties. Examples can be found of approved or investigational drugs covalently modified to overcome each of the aforementioned challenges, and several general strategies have emerged as depicted in Figure 19e. Success in non-invasive administration has mostly come from recombinant fusions with groups that enable epithelial transport by receptor-mediated trafficking (Dumont et al., 2005; Vallee et al., 2012), or with groups that facilitate endocytosis and translocation, such as cell-penetrating peptides (Gautam et al., 2016; Kauffman et al., 2015; Patel et al., 2009). Increased half-life has been achieved with both chemical and recombinant modifications and has relied on three primary mechanisms: (i) increased hydrodynamic radius to reduce renal elimination, (ii) steric shielding from proteases, and (iii) receptor-mediated endosomal escape and recycling. All three mechanisms may be employed by fusions to Fc(Ducore et al., 2014), albumin(Golor et al., 2013), or groups that bind to albumin (Adabi et al., 2017; Kim et al., 2017; Langenheim & Chen, 2009; Miyakawa et al., 2011). Modification with the polymer polyethylene glycol (PEG), by far the most widely used covalent modification for biologics to date (Harris et al., 2001; Swierczewska et al., 2015; C. Zhang et al., 2014), relies on the first two of these mechanisms. PEGylation is also a potential means of reducing immunogenicity, though anti-drug antibodies have remained a significant concern even for some PEGylated therapeutics (Gefen et al., 2013; Mima et al., 2015). Other approaches such as fusion to tolerogenic Fcs have been proposed as potential platform technologies in this area (Levin et al., 2015). Finally, cytosolic uptake of peptide

and protein drugs has been improved by fusions that enable specific receptor-mediated endocytosis (e.g. by antibodies against human insulin receptor(Giugliani et al., 2018) or the transferrin receptor(Sonoda et al., 2018)) or cell-penetrating peptides enabling non-specific endocytosis or translocation(Guidotti et al., 2017).

Though there has been much success with covalent modifications of biologics to date, there are still significant problems that are not well addressed by any of the above well-explored strategies. For example, the function of protein and peptide drugs is exquisitely sensitive to their tertiary structure, which is easily disturbed by fusions and poorly-placed modifications. With respect to cytosolic uptake, most current strategies rely on endocytosis and thus face the challenge of subsequent escape from the endosome, as opposed to direct translocation or membrane permeabilization. Finally, because many biologic drug candidates face several of the major challenges outlined above, it is desirable to find modification strategies that can simultaneously improve performance in as many of these areas as possible. Lipidation, or acylation, is a clinically-proven strategy with the potential to address each of the key challenges in protein and peptide drug development while overcoming many of the challenges of current alternative methods.

Lipidation involves the transfer of a lipid group to a protein (Lanyon-Hogg et al., 2017). In nature, lipidation is used by the cell to control the function and intracellular location of several proteins (Guan & Fierke, 2011; Lanyon-Hogg et al., 2017). When used to modify therapeutics, lipidation can greatly increase their half-life and cell membrane permeability (Lau et al., 2015; Trier et al., 2015). Moreover, lipidation has

been shown, in some cases, to have superior performance to PEGylation, the current standard post-translational modification (Bellmann-Sickert et al., 2011). Currently, there are several FDA-approved lipidated therapeutics with more in development, as summarized in Table 3. Potential advantages of lipidation can primarily be attributed to several key effects, including improving half-life, enhanced delivery, enabling alternative routes of administration, improving pharmacologic potency and reducing immunogenicity.

Table 3: Current acylated therapeutic peptide and protein based drugs

Drug	**Parent molecule**	**Indication**	**Mode of delivery**	**Stage**	**Half-life (hours)**
Daptomycin	N/A	Bacterial infections	Injection	Approved	8.6(Kullar et al., 2011)
Polymixin B	N/A	Bacterial infections	Topical/Injection	Approved	4-6
Tesamorelin	Growth hormone releasing factor	HIV-associated lipodystrophy	Injection	Approved	0.4-0.6(Stanley et al., 2011)
Detemir	Insulin	Diabetes	Injection	Approved	5-7(Danne et

					al., 2003)
Degludec	Insulin	Diabetes	Injection	Approved	25(Haahr & Heise, 2014)
Liraglutide	GLP-1	Diabetes	Injection	Approved	13(Knudsen et al., 2000)
		Obesity	Injection	Approved	
Semaglutide	GLP-1	Diabetes	Injection	Approved	168(L. Jensen et al., 2017)
			Oral	Phase 3	
		Obesity	Injection	Phase 3	
		Non-alcoholic fatty liver disease	Injection	Phase 2	
Somapacitan	Human growth hormone	Growth disorders	Injection	Phase 3	34-45(Battelino et al., 2017)
MEDI0382	Glucagon/GLP-1	Diabetes and obesity	Injection	Phase 2	9.5-12(Ambery et al., 2018)

LY3298176	GLP-1/ Gastric inhibitory polypeptide (GIP)	Diabetes	Injection	Phase 2	116.7(Coskun et al., 2018)
Tri-agonist 1706	Glucagon/GLP-1/GIP	Obesity		Phase 1	

Despite the several potential advantages of lipidation, this post-translational modification remains underutilized for developing therapeutics. We will discuss the mechanisms by which lipidation alters the properties of peptide and protein therapeutics as well as highlight important variables that may impact therapeutic development and efficacy.

5.2 Improvements in drug half-life

Lipidated molecules have demonstrated significant improvements in half-life ranging from ~10-fold increase for liraglutide, a lipidated GLP-1 analogue indicated for diabetes and obesity (13 hours half-life vs. 1.5 hours for native GLP-1), to an over 100-fold improvement for semaglutide (168 hours half-life), a next-generation lipidated GLP-1 analogue (L. Jensen et al., 2017; Knudsen et al., 2000). Another recent advance is the development of a lipidated human growth hormone analogue(Ramírez-Andersen et al.,

2018), which can be administered once weekly thanks to its ~40 hour half-life (Johannsson et al., 2018).

Lipidation improves drug half-life primarily by enabling albumin binding, as shown for the drugs in Table 1 and other leads in development (Hamilton-Wessler et al., 1999; J. Wang et al., 2003, 2002; Yuan et al., 2005). Albumin, a 66.5 kDa protein and the most abundant protein in blood at 40 mg/ml (Andersen et al., 2014), has a very long half-life (20 days) because it can bind to neonatal Fc receptors (FcRn) in growing endosomes which recycle albumin by sorting it away from lysosomes (Chaudhury et al., 2003). Albumin's main role is to transport hydrophobic molecules in the bloodstream, including fatty acids (Quinlan et al., 2005). When incubated *in vitro* with plasma at clinically relevant concentrations, a large fraction (>95%) of lipidated biologics bind to albumin (Hamilton-Wessler et al., 1999; Plum et al., 2013). Albumin binding improves half-life (Markussen et al., 1996; Matthews et al., 2008) by limiting the access of proteases to the drug (Jacobsen et al., 2016), greatly reducing renal clearance (Malm-Erjefalt et al., 2010; J. Wang et al., 2003; Yuan et al., 2005), and leveraging FcRn-mediated recycling (Figure 20a). Because albumin binding is reversible (Figure 20), lipidated therapeutics are free to elicit their effect once unbound (Hamilton-Wessler et al., 1999). Importantly, under normal physiological conditions only 2 out of 7 binding sites on albumin are occupied with fatty acids, eliminating the threat of lipidated therapeutics competing for binding sites with albumin's natural cargo (Fujiwara & Amisaki, 2013; Kurtzhals et al., 1995; Markussen et al., 1996).

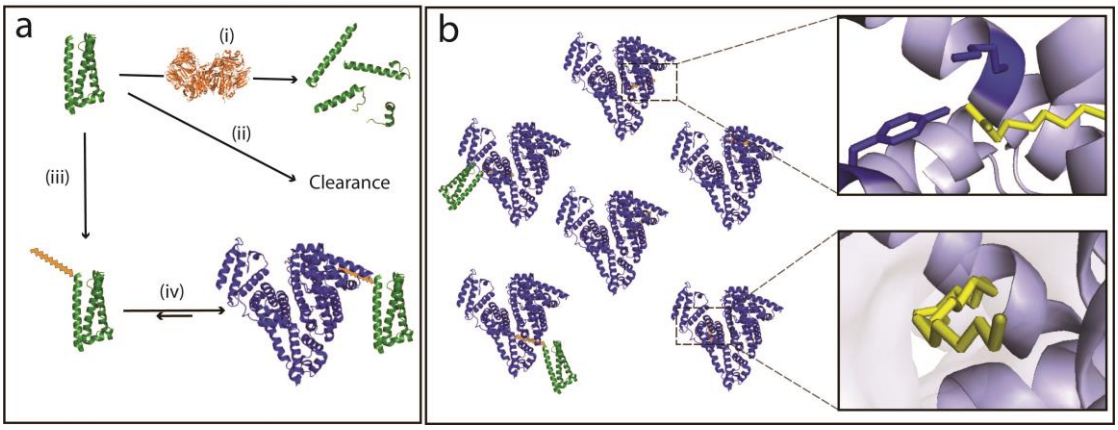

Figure 20: Half-life extension of lipidated therapeutics through albumin binding. (a) non-lipidated molecules have increased rates of (i) proteolysis and (ii) renal clearance. (iii) Lipidation enables reversible albumin binding (iv) reducing renal clearance and degradation by proteases, such as dipeptidyl peptidase 4 (relevant for GLP-1 analogues; PDB ID: 4APD). (b) Seven sites are available on albumin for medium and long-chain fatty acids. Interactions between the lipid carboxylate and polar amino acids at some of these sites (top close-up, dark blue) along with increased nonpolar interactions resulting from longer lipids (bottom close-up) can be leveraged for tighter albumin binding (PDB ID: 1E7H).

Affinity to albumin's fatty acid binding sites results from both electrostatic interactions of the lipid carboxylate group and polar side-chains as well as key binding of the nonpolar fatty acid chain as shown in Figure 20b (Bhattacharya et al., 2000; Spector, 1975). Increasing the length of the attached fatty acid on lipidated drugs generally leads to tighter albumin binding and increased therapeutic half-life (Knudsen et al., 2000; Markussen et al., 1996; J. Wang et al., 2002; Y. Wang et al., 2016). Recent efforts to increase lipidated therapeutic binding to albumin have resulted in the use of longer lipids (18 or 20 carbons) and the transition from monocarboxylic acids to dicarboxylic acids (Lau et al., 2015; Ramírez-Andersen et al., 2018). The improved half-life associated with the use of these lipids has so far been attributed to tighter albumin binding (Lau et al., 2015; Ramírez-Andersen et al., 2018). Linear fatty acids also tend to outperform bulkier

albumin binders since the latter category results in either decreased therapeutic potency or half-life (Madsen et al., 2007). Also, unsaturated bonds do not appear to significantly affect albumin binding (Madsen et al., 2007). Importantly, the lipid attachment site within the lipidated therapeutic does not affect albumin binding (Knudsen et al., 2000).

5.3 Enhanced Drug Delivery

Another key advantage of lipidation is that it enables enhanced drug delivery. First, lipidation enables a depot offering delayed drug absorption after subcutaneous injection. Second, lipidation enables improved cellular permeabilization and cellular drug uptake.

Lipidated therapeutics are routinely delivered subcutaneously and they have been shown to reversibly form multimers at the injection site as illustrated in Figure 21. These multimers act like a depot that delays absorption from the injection site to the bloodstream (Figure 21a) (Y. Wang et al., 2016). Although lipidated therapeutics are still able to bind albumin at the interstitium and this contributes to some absorption delay, the formation of multimers has a greater contribution (Dea et al., 2002; Havelund et al., 2004; Jacobsen et al., 2016; Plum et al., 2013; Y. Wang et al., 2016). For example, the formation of a subcutaneous depot almost doubles the half-life of liraglutide as compared to intravenous injections of the same molecule (Figure 21b) (Elbrond et al., 2002). Once the therapeutic has entered the bloodstream, albumin binding takes over as the primary mechanism for extending half-life.

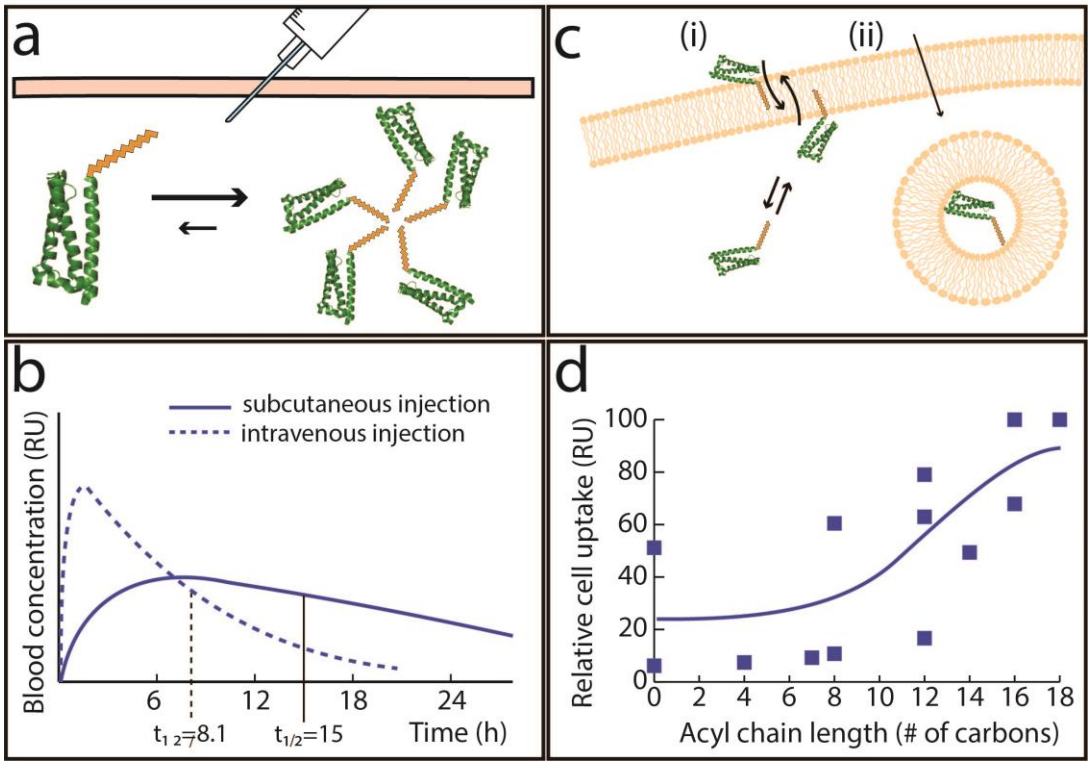

Figure 21: Enhanced drug delivery due to lipidation. (a) Lipidated therapeutics are routinely injected subcutaneously where they form a drug depot through reversible self-association causing (b) a delay of drug absorption into the body, which accounts, for example, for almost doubling the half-life of liraglutide, a GLP-1 analogue used to treat diabetes and obesity (data taken adapted from (Elbrond et al., 2002)). (c) Lipidation also increases cell uptake through (i) a flip-flop mechanism or (ii) endocytosis. (d) Increases in lipid length results in increased cell uptake (data taken adapted from (Bode et al., 2012; Oh et al., 2014; Trier et al., 2014)).

Another major consequence of lipidation is an increase in cell membrane permeability of peptide drugs. The lipid groups increase affinity towards the cell membrane and can enable insertion into the membrane (D'Errico et al., 2008; Y. Wexler-Cohen & Shai, 2009). Increased affinity towards the cell membrane also promotes higher cellular uptake with potential implications for intracellular delivery (J. Wang et al., 2002). Generally, increasing the lipid length leads to increased cell uptake (Figure 21d) (Bode et al., 2012; Oh et al., 2014; Trier et al., 2014). However the peptide portion can

also affect the magnitude of cell internalization and can sometimes shift the optimal lipid length for maximal cell uptake to shorter lipids (Lee & Tung, 2010; Nelson et al., 2007). For example, lipidation of nona-arginine (9 arginines) with laurate (12 carbons) resulted in a 1.5- to 3-fold increase in uptake when compared to lipidation with myristate (14 carbons) or palmitate (16 carbons), respectively (Lee & Tung, 2010). It has been hypothesized that the formation of stable multimers, a common occurrence for lipidated molecules, can lead to decreased cell membrane interactions (Trier et al., 2015). In these cases, elongating the lipid chain may lead to more stable multimers due to increased non-polar interactions. The competing equilibria between multimer formation and cell membrane interaction may alter the specific impact of lipid length on cellular uptake.

Studies on the uptake of lipidated peptides have been performed and point to both endocytosis and direct translocation as pertinent mechanisms (Figure 21c) (Lee & Tung, 2010; Missirlis et al., 2009; Nelson et al., 2007). Intracellular delivery is a highly sought-after goal, as only one third of the proteome is accessible from outside cells (Guidotti et al., 2017; Miersch & Sidhu, 2016). The direct translocation of lipidated peptides would eliminate the need for endosomal escape in intracellular delivery. However, uptake through direct translocation has been highly disputed, in part due to studies using fixed cells in confocal microscopy experiments (Richard et al., 2003). Cell fixation causes an artifact where peptides are observed in the cytoplasm after uptake (suggesting direct translocation or endosomal escape), contrary to what is observed in non-fixed cells (Missirlis et al., 2009; Richard et al., 2003). Still, experiments looking at quenching

kinetics of fluorescently labeled lipidated peptides loaded into artificial vesicles indicate lipidated peptides can flip-flop between the inner and outer membrane of the vesicles (Eisele et al., 2002). Moreover, using FRET probes, it was observed that a family of lipidated therapeutics, pepducins, can enter cells and localize in the inner leaflet of the cell membrane (Tsuji et al., 2013).

Direct translocation is further supported by the observation that lipidated peptides accumulate in the cytoplasm after incubation at 4 °C, at which temperature endocytosis is stopped (Bode et al., 2012; Lee & Tung, 2010; Nelson et al., 2007). While these experiments are well-established and easy to perform, they erroneously assume that direct translocation, which can be an energy-dependent process, is not significantly affected at 4 °C; thus, they likely underestimate the magnitude of direct translocation at elevated temperatures (Nelson et al., 2007). In addition, membrane fluidity is also reduced at temperatures below 20 °C (Missirlis et al., 2009; Nelson et al., 2007). The latter concern is especially relevant to lipidated peptides since they cross cell membranes in fluid-phase domains after perturbing membrane packing (Swiecicki et al., 2015; Trier et al., 2015). In fact, the use of unsaturated lipids as opposed to saturated ones can increase relative rates of direct translocation because they can interfere with membrane packing to a higher extent (Swiecicki et al., 2015).

5.4 Routes of administration

The increased cell membrane permeability of lipidated therapeutics also has implications for alternative routes of administration, including oral, topical and

pulmonary delivery (J. Wang et al., 2003). Current research is ongoing to utilize lipidation to enhance the oral delivery of peptide-based drugs. Studies using Caco-2 cell monolayers, a model for drug intestinal absorption, have shown that lipidation increases translocation of the lipidated therapeutic through the monolayer (Trier et al., 2014, 2015). Notably, the optimal lipid length for monolayer translocation was shorter (12 carbons) than the optimal length for cellular uptake (16 carbons), probably due to higher retention of longer chain lipidated biologics in cell membranes(Bode et al., 2012; Trier et al., 2014, 2015). Lipidated therapeutics can also act as permeation enhancers on Caco-2 monolayers in a detergent-like manner (Trier et al., 2015). Higher concentrations and lower pH favor increased permeability (Trier et al., 2015). In addition to lipid length, lipid attachment site also influences changes in permeability, with lipid length trends even reversing for different attachment positions (Trier et al., 2015). It was hypothesized that a dependence on lipidation site may be due to the formation of different multimers where more stable multimers would lead to decreased interactions with the cell monolayer and thus decreased permeability (Trier et al., 2015). This is in agreement with previous observations that only monomers, as opposed to multimers, are internalized by cells (Missirlis et al., 2009). Thus, dissociation of single lipidated peptides from multimers seems to be a necessary step prior to uptake.

An important recent development that highlights the potential of lipidation to improve biologic-based therapies has been a novel oral formulation for semaglutide, as depicted in Figure 22. Semaglutide is currently approved for once-weekly administration

through subcutaneous injections, and its new daily oral formulation is currently undergoing Phase 3 clinical trials (Aroda et al., 2018). The oral semaglutide formulation needs at least 20 times more lipidated GLP-1 analogue (20 mg) than an injection of the same molecule (1 mg) to achieve the same efficacy (Aroda et al., 2018; Davies et al., 2017). Absorption into the bloodstream happens in the gastric epithelium requiring patients to be fasting before ingesting the pill (Figure 22a) (Connor et al., 2017). The permeability of semaglutide alone through the gastric epithelium does not offer sufficient bioavailability, so semaglutide is supplemented with the hydrophobic permeation enhancer sodium N-8(-[2-hydroxylbenzoyl]amino) caprylate (SNAC) (Figure 22b, 4c)(Bjerregaard et al., 2012). SNAC has also been used to increase the efficacy of oral insulin and peptide YY administration (Karsdal et al., 2015). In addition to enhancing the permeability of semaglutide, SNAC also increases the local pH around the tablet in the stomach, protecting the therapeutic against degradation (Connor et al., 2017).

Semaglutide maintains a very long half-life after oral delivery (~153-160 h compared to 168 h after injection) (Granhall et al., 2017), but as discussed its bioavailability is drastically reduced, requiring higher daily doses. Future lipidated therapeutics intended for oral delivery may benefit from lipidation with shorter lipids (12 carbons) than used in semaglutide (18 carbons) which may increase bioavailability. Lipidated therapeutics that do not form stable multimers in the stomach may also have higher permeability (and bioavailability) due to increased interactions with the membrane (Figure 22d-f). Generally, more efficacious lipidated therapeutics intended for oral

delivery have the potential to be designed specifically for that route of administration where higher bioavailability is balanced with improvements in half-life.

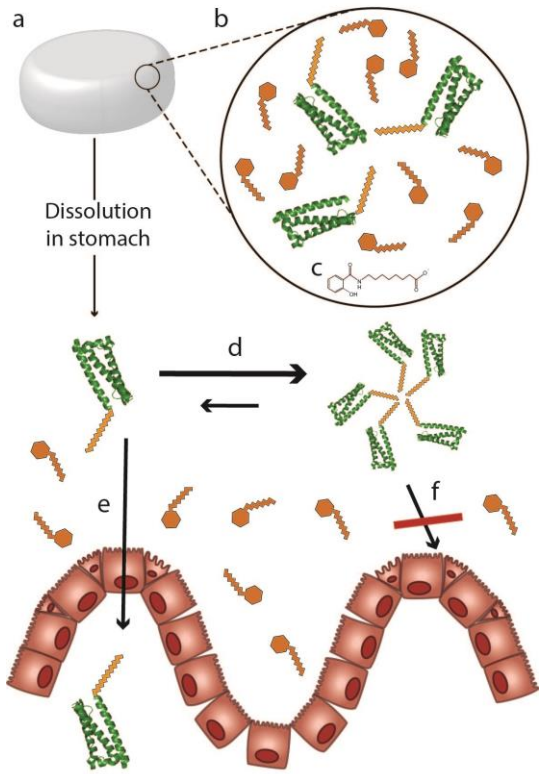

Figure 22: Oral delivery of peptides enabled by lipidation. (a) Tablet formulations are now possible for lipidated therapeutics, such as semaglutide, a GLP-1 analogue used to treat diabetes. (b) Besides the active ingredient (eg semaglutide), permeation enhancers are needed to increase therapeutic bioavailability such as (c) N-[8-(2-hydroxybenzoyl) amino] caprylate (SNAC) for semaglutide's oral formulation. (d) Self-association of lipidated therapeutics may negatively affect permeation of lipidated therapeutics through the gastric endothelium as (e) monomers may be able pass, (f) but not multimers.

Lipidation has also been widely used for topical delivery. Specifically, lipidation of small peptides increases skin permeability (Y. L. Choi et al., 2014; Rocco et al., 2016), where increasing the fatty acid chain length increases permeability, but also retention of the peptides in the first skin layer, the stratum corneum (Yamamoto et al., 2003).

However, lipidated peptides still have poor half-lives when applied on the skin(Y. L. Choi et al., 2014) and further studies looking to optimize dermal delivery using lipidation have not been extensively performed.

Among non-invasive routes of administration for peptides and proteins, the lungs have an exceptional potential for rapid and complete systemic uptake of lipidated molecules. The alveolar epithelium in particular is very thin, extensive, covered only by a small volume of fluid, and generally contains far lower levels of drug-metabolizing enzymes than either the skin or the gastrointestinal tract (Patton et al., 2004). However, approvals for inhaled biologics have been rare despite many years of extensive efforts (Morales et al., 2017). The use of lipidation in pulmonary delivery efforts is unexplored, but promising. In particular, drug stabilization and transport across the alveolar epithelium might be supported by binding of lipidated drugs to albumin either *in situ* or in the formulated drug product, because albumin is present in the alveolar epithelial lining fluid and is actively transported into the capillaries at significant rates (Buchackert et al., 2012).

Despite the route of administration, lipidated therapeutics have a unique biodistribution. Lipidation primarily results in increased accumulation in the liver when compared to non-lipidated counterparts (Bellmann-Sickert et al., 2011; Hamilton-Wessler et al., 1999; J. Wang et al., 2003; Yuan et al., 2005), though the mechanisms behind this accumulation are unclear. Liver targeting may be an advantage for certain therapies, but accumulation in the liver may also be undesirable if other tissues are targeted, due to

drug unavailability or liver toxicity. Although, to date, lipidated therapeutics have not been observed to cause liver damage and importantly, dosage modification has not been required on patients with hepatic impairment (Flint et al., 2010; Lene Jensen et al., 2018), future potential therapies such as cytotoxic cancer drugs may face a different reality. Targeting moieties that have been previously shown to lessen hepatic accumulation of lipidated agents, may be of future use to alter biodistribution (Chekhonin et al., 1991).

5.5 Changes in Pharmacologic Potency

Lipidation, like other drug modifications, can have a profound effect on the potency and pharmacologic activity of the target therapeutic. The impact can be a function of the length of the lipid, as well as the site of lipidation, the linker used, and any structural interactions between the lipid or linker and the therapeutic or its target. As a result, the impact of lipidation on potency can be variable, unpredictable and drug-specific, as illustrated in Figure 23.

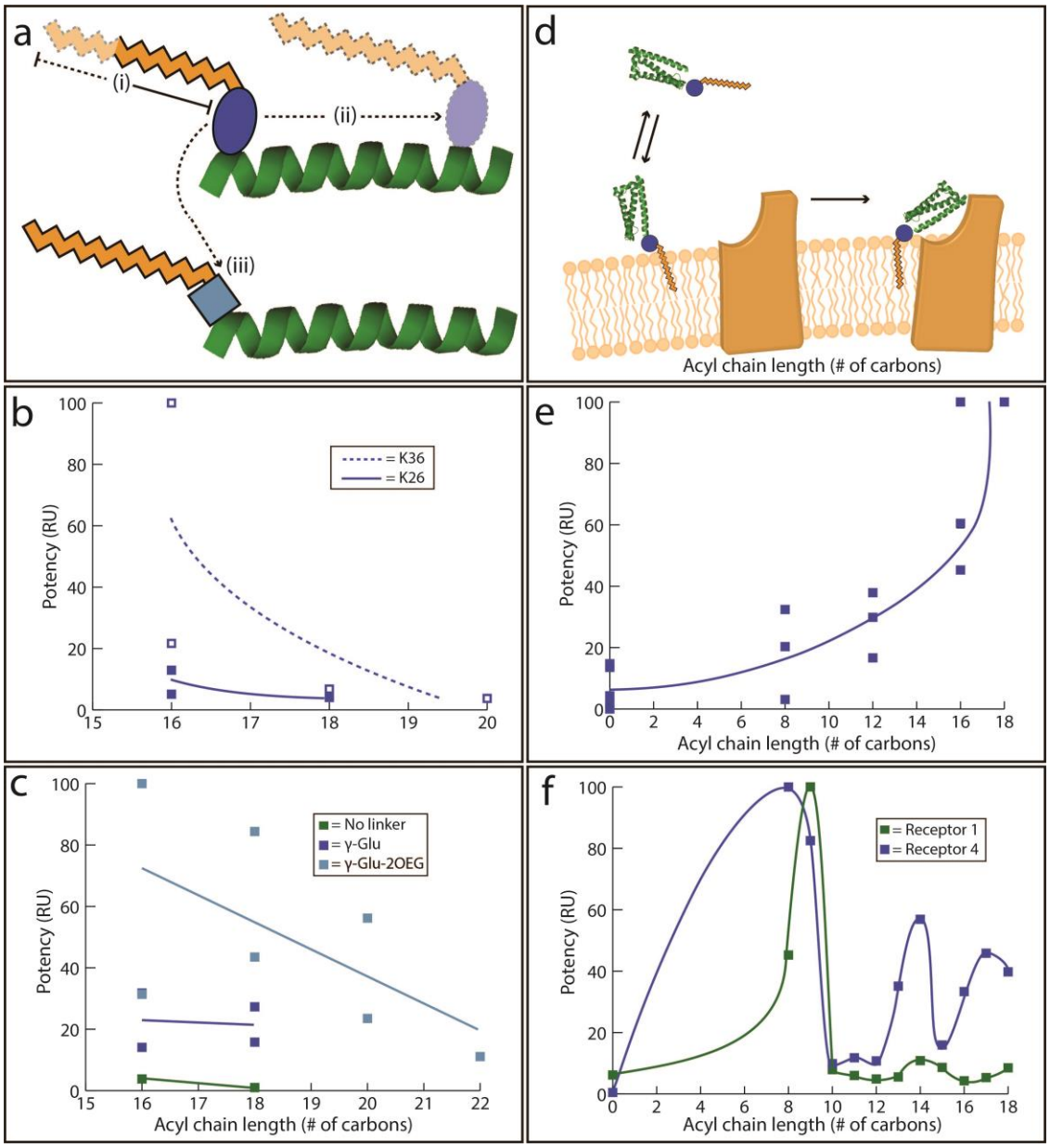

Figure 23: Changes in drug potency due to lipidation. (a) (i) Variations in lipid length, (ii) site of lipid attachment and (iii) type of linker used to attach the lipid influence resulting drug potency. (b) In some cases, increasing lipid length leads to decreased potency where lipidation at a particular site may result in a series of lipidated analogues with higher potency than at a different site (eg for GLP-1 analogues, lipidation at K36 results in analogues with higher potency than lipidation at K26) (Data adapted from (Knudsen et al., 2000)). (c) Longer, hydrophilic linkers allow the attachment of longer lipids at lower potency loss such as the use of a γ-glutamyl linker elongated with amino-3,6-dioxaoctanoic acid (OEG) which results in more potent GLP-1 analogues when compared to the use of only the γ-glutamyl linker or no linker (Data adapted from (Lau et al., 2015)). (d) In some

other cases, because lipidation increases affinity towards the cell membrane, it can lower the energy barrier for the formation of therapeutic-target complex (e) which results in increased drug potency with increasing lipid length (Data adapted from (Doyle et al., 2014; Skovbakke et al., 2015; Y. Wexler-Cohen & Shai, 2009)). (f) In some other cases, increasing the lipid length results in oscillating potencies with maxima and minima (Data adapted from (Todorovic et al., 2005)).

Lipid length can be a powerful modifier of drug potency. Lipidation has been shown to either (i) decrease potency with increasing lipid length (Figure 23b-c), (ii) increase potency with increasing lipid length (Figure 23e) or (iii) result in optimal potency at intermediate lipid lengths (Figure 23f). Furthermore, altering the lipidation site can result in molecules with varying potencies (Knudsen et al., 2000; Y. Wang et al., 2016), and the use of different lipids at one site may result in a series of analogues with higher potency than at a different site (Figure 23b). As such, it is important to evaluate different lipid attachment sites and lipid lengths to optimize therapeutic efficacy (Knudsen et al., 2000; Ramírez-Andersen et al., 2018; Y. Wang et al., 2016).

The linker used to attach the lipid moiety to the target therapeutic can also impact potency (Knudsen et al., 2000; Lau et al., 2015; J. Wang et al., 2002). For molecules whose potency decreases with increasing lipid length, the use of a long hydrophilic linker for attachment of longer fatty acids can minimize the potency cost of lipidation (Figure 23c) (Knudsen et al., 2000; Lau et al., 2015). Conformational changes resulting from lipidation can modify the affinity of the therapeutic with its target and modify potency, and increasing the linker length may minimize the extent of these changes by increasing the distance between the attached lipid and the drug. In fact, screening different linkers

can completely compensate for the potency loss of attaching long fatty acids (Lau et al., 2015; Ramírez-Andersen et al., 2018). In general, however, it is unpredictable how linker choice may affect potency (Lau et al., 2015), as linker length has also been associated with affecting half-life, as exemplified by a 2-fold reduction in half-life when shorter linkers are used in lipidated human growth hormone analogues (Ramírez-Andersen et al., 2018).

Recently, better-performing linkers have been designed. The use of γ-Glu has been most successful, complemented with further linker elongation using repeats of ethylene groups (PEG) (eg. semaglutide, a GLP-1 analogue used for diabetes) or amino-3,6-dioxaoctanoic acid (OEG) (eg. somapacitan, a growth hormone analogue used for growth disorders) (Dahl et al., 2013; Henderson et al., 2016; Lau et al., 2015; Ramírez-Andersen et al., 2018; Schultz et al., 2018) . The hydrophilicity of these linkers may aid in the synthesis and formulation since solubility issues are common for lipidated therapeutics (van Witteloostuijn et al., 2017). Importantly, initial clinical studies have shown no significant safety concerns for these PEG-based and OEG-based linkers (Ikushima et al., 2018; Johannsson et al., 2018), though their long-term effects are yet to be evaluated (Bech et al., 2018). It is unclear what specific features of these new linkers (besides length) favor higher drug potencies. More in-depth structure-activity relationship studies may be required on a case-by-case basis, including crystallographic studies of different linkers and lipids, although these have proven difficult in the past (Ramírez-Andersen et al., 2018).

If the therapeutic target of the lipidated biologic is membrane-bound, insertion into the cell membrane aided by the lipid group can reduce the overall energy requirements for the formation of ligand-receptor complexes (Figure 23d) (Todorovic et al., 2005). For these cases, increasing the lipid length elevates the affinity for the cell membrane and thus increases potency (Figure 23e) (Doyle et al., 2014; Skovbakke et al., 2015; Todorovic et al., 2005; Y. Wexler-Cohen & Shai, 2009). Interestingly, lipidated therapeutics have also been shown to preferentially bind to membranes of cell types expressing their target receptors, which may lower concerns related to the retention of lipidated therapeutics at non-target tissues (Chekhonin et al., 1991; Y. Wexler-Cohen & Shai, 2009).

However, as discussed above, the relationship between lipid length and drug-target affinity can be complex, as shown in Figure 23f, and even small changes to the overall structure can greatly impact potency. The large impact of these small changes in lipid length can be exploited to improve drug specificity, such as in targeting specific melanocortin receptors (Todorovic et al., 2005). For example, Figure 23f illustrates a maximum local potency against Melanocortin Receptor 1 for a drug with a lipid length of 17 carbons, which also corresponds to a local affinity minimum for an alternative receptor: Melanocortin Receptor 4 (Todorovic et al., 2005). These results again highlight the need for experimentation, and perhaps better structural characterization, to identify lipid and linker design principles that may very well be drug and target specific.

A better understanding of design principles may allow for the use of very long chain fatty acids (20 carbons or more), a virtually untapped therapeutic space with few examples (Ramírez-Andersen et al., 2018). Studies have shown that while very long fatty acid chains are able to bind to albumin, they preferentially partition to a greater extent with lipoproteins (J. K. Choi et al., 2002; Shafrir et al., 1965). The effect of lipoprotein binding on half-life has not been investigated, but selective binding to lipoproteins as opposed to albumin could be leveraged to use them as drug carriers or to treat lipoprotein-related conditions (Stoekenbroek et al., 2015; Thaxton et al., 2016; Wolfrum et al., 2007).

5.6 Reduced Immunogenicity

Lipidation has also been shown to decrease the immunogenicity of potential biologics, which has been a major hurdle in the clinical translation of protein-based therapeutics (Armstrong et al., 2007; Baker et al., 2010). As of 2015, 89% of all approved biologics have reported some immunogenic response and 55% of these reactions have had an impact on therapy efficacy(Y. Wang et al., 2016). These responses include life-threatening autoimmunity, allergic reactions and accelerated blood clearance due to formation of anti-drug antibodies (Baker et al., 2010; Mima et al., 2015). Furthermore, some post-translational modifications intended to increase therapeutic half-life are themselves immunogenic, limiting efficacy (Mima et al., 2015; P. Zhang et al., 2016).

Patients treated with lipidated therapeutics have also been found to develop low levels of anti-drug antibodies, although to date the presence of these antibodies has not

been shown to reduce therapy efficacy (Bartley et al., 2008; Buse et al., 2011). On the other hand, lipidated peptides have been shown in some cases to be less immunogenic than their non-lipidated counterparts, as depicted in Figure 24 (Cloake et al., 2014; Schultz et al., 2018). Of course, this excludes highly immunogenic bacteria-inspired lipid epitopes (eg. $Pam_{2/3}Cys$), which are known to activate immune responses and are used for vaccine development (Jackson et al., 2004; Zeng et al., 2011).

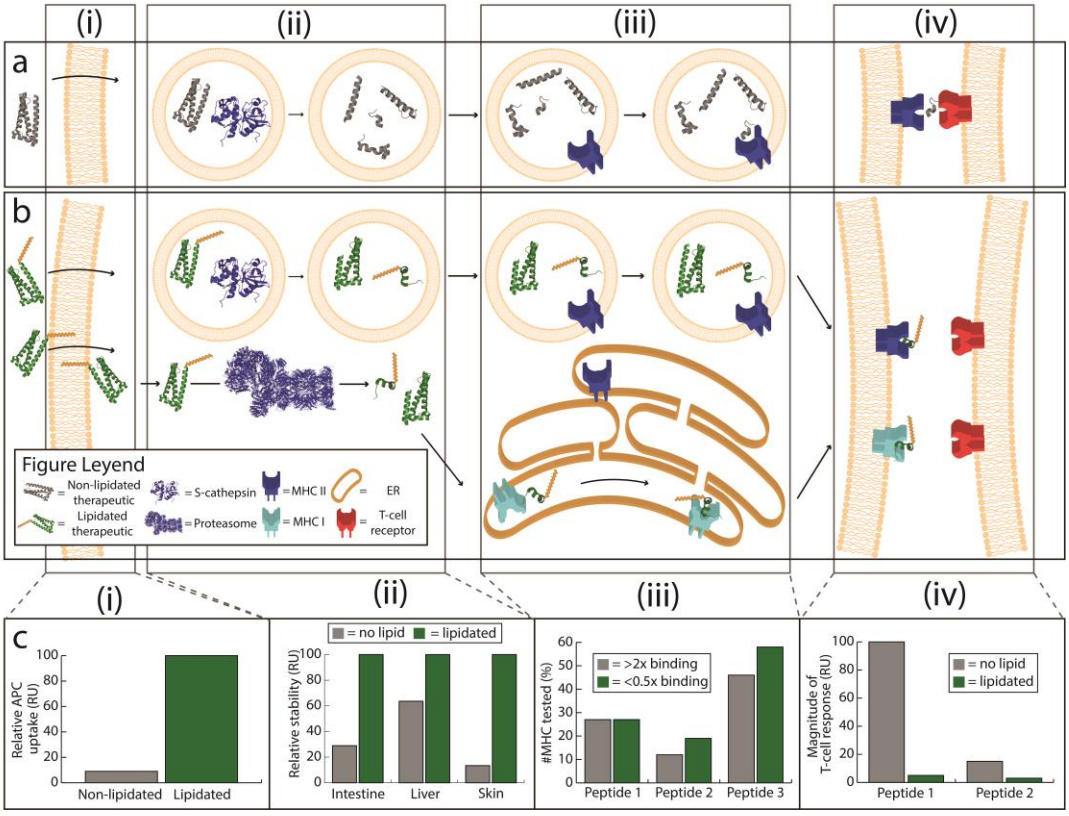

Figure 24: Reduced immunogenicity due to lipidation. (a) Peptide and protein-based drugs can become immunogenic through the MHC II antigen presentation pathway. (b) Lipidated therapeutics can potentially be presented through both MHC I and MHC II antigen presentation pathways but have been shown to have lower immunogenic potential than non-lipidated counterparts. (c) Comparison of non-lipidated with lipidated therapeutics at different steps in the antigen presentation pathway. (i) Lipidation causes increased antigen uptake by antigen presenting cells (APC) (Data adapted from (N. Pfender et al., 2003)). (ii) Lipidation have been shown to be resistant to proteolytic degradation,

though this has not been shown in APCs specifically (Data adapted from (Asada et al., 1994; Y. L. Choi et al., 2014; Yuan et al., 2008)). Lipidation causes an overall decrease in binding to MHC II molecules (Data adapted from (Schultz et al., 2018)). (iv) Lipidation can greatly decrease the magnitude of T-cell responses (Data adapted from (Schultz et al., 2018)).

One key determinant of the immunogenicity of lipidated drugs is the stability of the bond between the lipid and the peptide. An early report showed that use of a thioester bond (which is labile in the intracellular environment), resulted in significantly higher immunogenicity compared to the use of an amide bond, which is more stable (Beekman et al., 1997). Another study showed that lipidating a peptide using an amide linker could convert an immunogenic peptide into a non-immunogenic peptide whereas the use of a thioester linker would do the opposite (Cloake et al., 2014).

To explain this phenomenon, it was suggested that lipidation increased the uptake of the peptide by antigen presenting cells (APCs), increasing immunogenicity, but also that the attached lipid could interfere with antigen processing or T cell activation, decreasing immunogenicity (Beekman et al., 1997; N. A. Pfender et al., 2008). Lipidation does increase cellular uptake regardless of the nature of the bond (N. Pfender et al., 2003), however if the bond is stable, it is hypothesized that the lipid remains attached inside APCs and interference with immune activation occurs, outweighing increased APC uptake and lowering overall immunogenicity (Cloake et al., 2014). A recent study by Schultz *et. al.* determined that lipidation can interfere with T cell activation because it can decrease binding to MHC II molecules and significantly reduce T cell responses (Figure 24) (Bueno et al., 2004; Schultz et al., 2018). The exact mechanisms for decreased MHC II binding and T-cell responses are still under investigation, but lipidated

peptides are known to be protease-resistant (Asada et al., 1994; Y. L. Choi et al., 2014; Yuan et al., 2008), and impaired antigen proteolysis may prevent downstream processes from happening (Antoniou et al., 2000). It has been observed that increasing the lipid chain length decreases immunogenicity (Schultz et al., 2018). The lipid attachment site also affects any decrease in immunogenicity, with several cases showing complete immune response evasion (Schultz et al., 2018). These changes are not necessarily additive, as lipidation at multiple sites does not further reduce immunogenicity and, in some cases, can increase it (Deliyannis et al., 2002; Schultz et al., 2018).

Lipidated peptides should also be capable of being presented to the immune system through the MHC I pathway since they can access the cytoplasm after cellular uptake. MHC I molecules that bind lipidated peptides have been described and shown to bind shorter peptides (5-mer lipidated peptides) than regular MHC I molecules (9-mer peptides) (D. Morita et al., 2016). The shorter peptide length is a result of the lipid portion also interacting with the antigen binding groove of the MHC I molecule (D. Morita et al., 2016). The impact of these shorter peptides on MHC I presentation and immune activation is still unknown, and the effects of lipidation on MHC I pathway presentation are still under investigation. Studying how lipidation interferes with proper antigen processing may also help explain why some long (>20 amino acids) amide-bond lipidated peptides derived from pathogens, used for vaccine development, become immunogenic upon lipidation (BenMohamed et al., 2004). Lipidation in these cases may increase APC uptake but may not confer sufficient proteolysis resistance, leading to

proper antigen processing, presentation and immune stimulation. The sequence-specific degradation of lipidated peptides may also explain why lipidation at multiple sites is not beneficial, since multiple lipid groups do not provide additional protection against proteolysis (Asada et al., 1994).

5.7 Discussion and Future Outlook

While the use of lipidation as a post-translational modification in biologics has seen significant success and shows great promise in future drug development, there are some applications were its potential benefits may be limited. For example, in short-term anticoagulation therapies, used to prevent post-operative complications(Gunaratne et al., 2018), rapid drug clearance may be desirable as opposed to half-life extension. In addition, lipidation of monoclonal antibodies, the largest group of biologic therapeutics to date, may provide limited advantages. Antibodies already possess a long half-life on the order of days, due to natural FcRn-mediated recycling (Keizer et al., 2010) and absorption of subcutaneously injected antibodies into the bloodstream is already a slow process (Zhao et al., 2013). While lipidation of antibodies may increase therapy potency (Chekhonin et al., 1991) and has been used to enable intracellular targeting(Cruikshank et al., 1997), in general antibody lipidation remains understudied, with perhaps limited advantages.

In some applications, the use mixtures of lipidated drugs with their non-lipidated counterparts, may present some advantages. These mixtures could be used to balance potency and extensions in half-life. For example, in wound healing applications(X.

Huang & Brazel, 2001) a high initial therapeutic effect could be provided by a potent, short-lived non-lipidated therapeutic, and followed by a lower sustained effect from longer-lived, less potent, lipidated therapeutics.

Furthermore, even in applications where lipidation may offer a significant advantage, many challenges remain to the generalized use of lipidation. These challenges include: (i) a lack of mechanistic understanding of the effect of lipidation on cell uptake and epi/endothelium permeability, (ii) a lack of mechanistic understanding of the effect of lipidation on APC processing and immunogenicity and (iii) a lack of methodologies for the rapid synthesis of lead compounds.

A key understudied aspect of lipidated peptides and proteins is the effect of structure on cell uptake and permeability. There remains a need for studies that better quantify cellular uptake of lipidated therapeutics and elucidate the underlying uptake mechanisms at biologically relevant concentrations. Methodologies developed to study cell penetrating peptides can readily be translated to study lipidated drugs. These methods would allow for the precise quantification of uptake by direct translocation as well as endocytosis and the kinetics of such processes in both artificial vesicles (lacking endocytosis machinery) and live cells (Bechara & Sagan, 2013; Henriques et al., 2007). Future studies need to determine the properties that affect uptake through direct translocation (vs. endocytosis) such as degree of unsaturation, peptide hydrophobicity, charge and polarity, amphiphilicity, and concentration. Although this discussion has focused on the cellular uptake of peptides, it would also be interesting to determine any

molecular weight cut-offs for direct translocation that may impact the development of larger protein based lipidated drugs.

Importantly, the specific role of multimers in permeability and uptake also needs further study. Lipidated therapeutics need to be in monomeric form to enter cells (Missirlis et al., 2009). Thus, formation of stable structures may decrease, rather than increase, cell uptake (Trier et al., 2015). However, it has also been observed that increasing the lipid length increases cell uptake despite the formation of structures with longer lipids (Bode et al., 2012). The key to understanding this relationship may be the stability of the multimers formed and the nature of the multiple equilibria involved, as opposed to whether or not structures are formed. Future studies on cell uptake and membrane permeability need to be coupled with robust structural characterization studies. Formation of multimers also protects against proteases (Asada et al., 1994), but it is unknown whether this can occur *in vivo* inside APCs at biologically relevant concentrations.

Further investigations on the effects of lipidation on immunogenicity are also needed. Thus far, clinically successful lipidated therapeutics have been based on human proteins (e.g., GLP-1) with low intrinsic immunogenicity, as seen in Table 3, where all therapeutics, besides a couple of antibiotics, are human protein-based. However, expanding the repertoire of attainable lipidated therapeutics may include the use of intrinsically more immunogenic non-human-based proteins (e.g., peptide agonists to therapeutic targets). While there is an indication that lipidation can lower

immunogenicity, the exact mechanisms remain unclear. Future immunological studies should evaluate the presentation through MHC I (non-endocytic) pathways and subsequent T-cell activation. The effect of lipidation on antigen processing and MHC antigen presentation (not just binding) are also unknown.

As discussed above, one challenge with all post-translational modifications, including lipidation, is that the effect of the modification on potency and/or immunogenicity may be unpredictable. (Armstrong et al., 2007; Ehrlich et al., 2013; Sundy et al., 2011) . These unpredictable effects are often dealt with on a case-by-case basis and overcome by screening many compounds for optimized properties (Lau et al., 2015; Ramírez-Andersen et al., 2018). This often requires changing (i) the attached lipid moeity, (ii) the lipid attachment site, and (iii) the linker used for the attachment in order to optimize the potency as well as the pharmacokinetic and pharmacodynamic properties of the lead molecule. This non-trivial case-by-case optimization requires the synthesis of a panel of compounds, which is generally first screened for improved half-life while minimizing potency losses. To date, the synthesis of these panels has relied on purely chemical methods that have been limited in the number and nature of compounds that can be generated.

Site-selectivity is a major restriction in current chemical synthesis approaches, where only a few types of amino acids can be used for conjugation. Small lipidated peptides can generally be synthesized using solid phase synthesis wherein a unique amino acid can then lipidated, most commonly lysine (Knudsen et al., 2000) or where lipidated

amino acid building blocks are premade and incorporated during synthesis (Mejuch & Waldmann, 2016). For example, this strategy has been used to introduce a free cysteine into human growth hormone which, after production in *E coli*, allows site-specific lipidation (Ramírez-Andersen et al., 2018). Site-selectivity becomes a more significant challenge when lipidating proteins because this can require mutations to either introduce or reduce the number of sites available for conjugation, which may carry a risk of immunogenicity. Another strategy is to produce a truncated protein that can later be ligated with a chemically synthesized lipidated peptide. These ligation reactions include native chemical ligation, MIC ligation and sortease-mediated ligation, but all of these have sequence restrictions on the lipid attachment site and on the peptides/proteins that can be ligated. These approaches have been recently reviewed by Mejuch and Waldmann (Mejuch & Waldmann, 2016). Microbial transglutaminase has also been used to site-specifically lipidate proteins, but also has substrate specificity restrictions (Abe et al., 2011; Ramírez-Andersen et al., 2018). Purely biosynthetic routes are attractive and have the potential to enable the synthesis of a large diversity of non-natural lipidated peptides or proteins. But, current biological methodologies rely on media supplementation with free fatty acids which are then used for lipidation of natural protein substrates by existing post-translational lipidation machinery (Glück et al., 2010). To broaden the applicability of this approach, significant improvements in enzymes are needed because the substrate specificity of the lipidation machinery is restricted. Altogether, the production of next-generation lipidated biologics will require the development of new synthesis technologies that expand the array of readily obtained compounds.

In summary, new lipidation methodologies coupled with a better understanding of the mechanisms underlying the efficacy of lipidated therapeutics will lead to next-generation lipidated drugs with more diverse and complex properties, tailored to specific applications, accelerating the development of new therapies for a large variety of diseases.

6. Improved *in vivo* Protein N-Myristoylation through Dynamic Metabolic Control.

6.1 Introduction

The use of biologics to treat a myriad of conditions is on the rise. (Kinch, 2015) It is not surprising that the $200 Billion USD annual biologics market is continuing to grow at a rapid pace, with an ever increasing percentage of the population benefiting from these drugs. (Walsh, 2018) To date, the translation of candidate proteins to effective clinical therapies has often required PEGylation, which has been necessary to increase the stability of these drugs, but which can also result in a heterogeneous population of molecules. (Andrews et al., 2015; Hakem et al., 2013; Veronese, 2001) Protein lipidation, which has been successfully utilized clinically, is an attractive alternative to PEGylation. (Garber, 2011; Knudsen et al., 2000; Menacho-Melgar et al., 2019; Ryan & Hardy, 2011; Stephenson et al., 2014; Ward et al., 2013)

Importantly, lipidation has emerged as a strategy to increase the half-life of peptide therapeutics. (Menacho-Melgar et al., 2019; L. Zhang & Bulaj, 2012) This approach exploits fatty acid affinity towards albumin, which is the most abundant protein in serum and has multiple fatty acid binding sites. (Andersen et al., 2014; Knudsen et al., 2000) Lipidation of biologics results in a drug reservoir of drug-bound albumin in serum and is a clinically proven strategy that has resulted in eight currently FDA-approved therapeutics, with more agents currently in development. (Menacho-Melgar et al., 2019) The mechanisms by which lipidation can extend the half-life of biologics have been well characterized. (Menacho-Melgar et al., 2019)

The development of novel lipidated drug candidates is currently limited by lipidation methodologies that rely on traditional chemistry, that often require multiple steps, have low specificity, require purification, and result in low yields. (Mejuch & Waldmann, 2016; Wu et al., 2004) Furthermore, these methodologies have constrained the array of lipidated biologics that can be translated into the clinic to small peptides that may require additional mutagenesis to prevent product heterogeneity when chemically derivatized. (Menacho-Melgar et al., 2019) It is often necessary to reduce the number of chemically reactive sites, eg. lysine, to a single reactive site to enable product specificity. (Knudsen et al., 2000; Lau et al., 2015; Mejuch & Waldmann, 2016) An alternative approach is the direct biological synthesis of lipidated proteins. While specifically, myristoylation of proteins in *E. coli* has been previously demonstrated, these approaches have relied on the native ability of *E. coli* to uptake and activate free fatty acids from the culture medium. (De Cotiis et al., 2008; Duronio et al., 1990; Glück et al., 2010; Knoll & Gordon, 1993) This approach not only requires the addition of detergent solubilized fatty acids to the medium, but also the use of rich media, as fatty acid uptake and activation is repressed in many minimal media.(Tao et al., 1999) These limitations reduce the utility of this approach for commercial therapeutics production, and to date these approaches have resulted in only low production at 20 mg/L (Glück et al., 2010) in the case of purified HIV Nef protein and >40 mg/L (Luginbuhl et al., 2017) in the case of purified unstructured elastin like proteins (ELPs).

E. coli, which is still used for the commercial production of over a third of current biologics, is readily amenable to both metabolic engineering strategies as well as

heterologous protein production, making it an ideal choice as for further platform development. (Menacho-Melgar, Ye, et al., n.d.) Numerous prior studies have demonstrated the ability to engineer *E. coli* to specifically produce key metabolites, including numerous fatty acyl-CoAs and derivatives, at high levels. (M. Lynch et al., 2019; Peralta-Yahya et al., 2012; Rude et al., 2011; Schirmer et al., 2010; Steen et al., 2010) In addition, the use of *E. coli* as a host for directed enzyme evolution is well vetted. (Firnberg & Ostermeier, 2012) We report steps towards the development of a platform *E. coli* strain for the single pot, facile production of N-terminally myristoylated proteins and peptides, reliant on *in vivo* N-terminal lipidation as illustrated in Figure 24. (M. Lynch et al., 2019; Peralta-Yahya et al., 2012; Rude et al., 2011; Schirmer et al., 2010; Steen et al., 2010) This approach includes (i) the expression of a target protein to be myristoylated (ii) the generation of fatty acids and their activation into fatty acyl-CoAs and (iii) a myristoyl transferase that catalyzes the lipidation reaction. In addition, (iv) the deletion of key proteases that degrade lipidated proteins are required to maximize production.

Specifically, we have leveraged previous strains and plasmids for 2-stage biosynthesis, wherein myristoylated proteins are produced in a productive phosphate limited stationary phase. (Burg, Cooper, Ye, Reed, & Moreb, 2016; Michael D. Lynch, 2016) This includes strains and methods for the production of recombinant protein upon phosphate depletion. (Menacho-Melgar, Ye, et al., n.d.) as well as the use of phosphate depletion in two-stage dynamic metabolic control to induce dynamic changes in metabolism, by reducing levels of metabolic enzymes, leveraging controlled proteolysis. (Ye, Z., Lynch,M.D., Trahan, A.D., Rodriguez, D.L., Cooper, C.B. Bozdag, A., 2015)

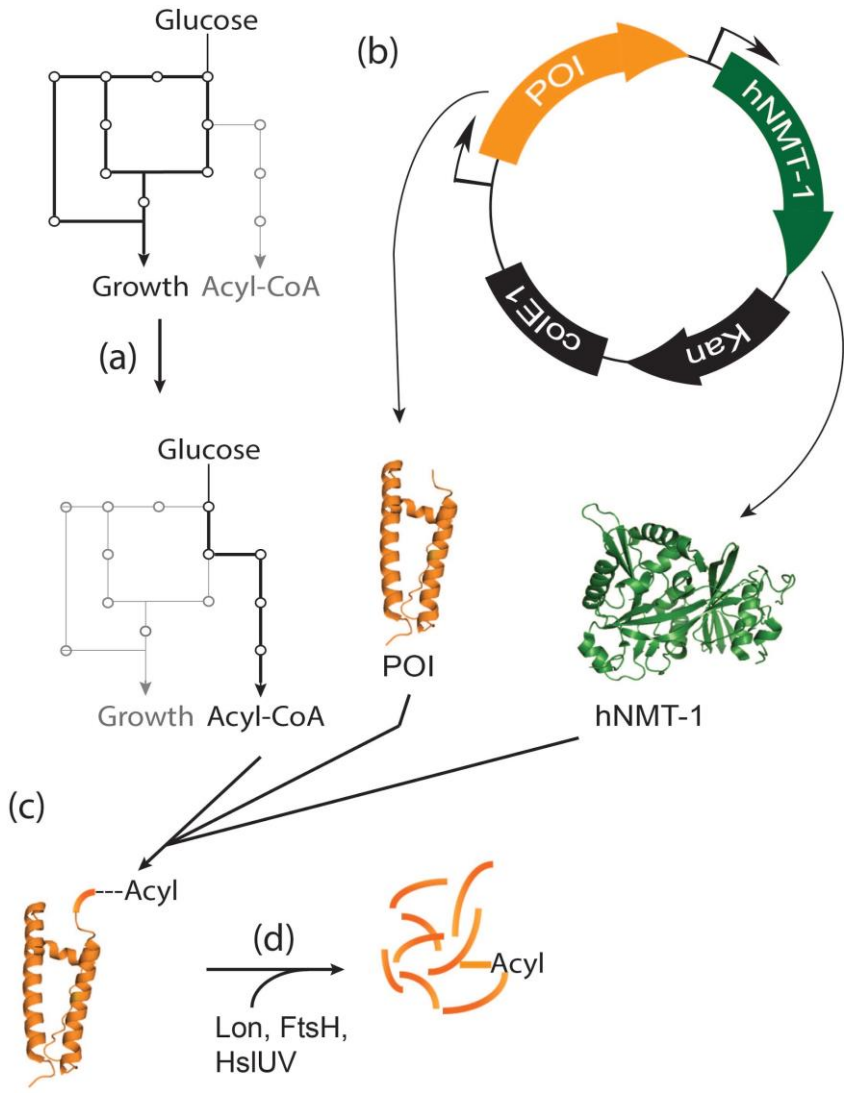

Figure 25: Complete in vivo protein lipidation. (a) Increased production of acyl-CoAs is achieved through dynamic metabolic control, upon phosphate medium depletion. (b) A protein/peptide of interest (POI) bearing an N-terminal myristoylation sequence is expressed along with hNMT-1. (c) hNMT-1 transfers acyl groups from acyl-CoA to the target protein/peptide. (d) Decreased levels of Lon, Ftsh and HslUV increase lipidated Nef levels.

6.2 Materials & Methods

6.2.1 Materials

Unless otherwise stated, all materials and reagents were of the highest grade possible and purchased from Sigma (St. Louis, MO). Luria Broth (LB) was used for strain and plasmid propagation and construction. Working antibiotic concentrations were as follows: kanamycin (35µg/mL), chloramphenicol (Cm. 35µg/mL), spectinomycin (Sp. 100µg/mL), zeocin (Zeo. 50 µg/mL), gentamicin (Gent. 10µg/mL), blasticidin (Bsd. 100µg/mL), tetracycline (Tet. 5µg/mL). Luria broth with low salt (Lennox formulation) was used to select for zeocin and blasticidin resistant clones.

6.2.2 Strain Construction

Chromosomal modifications were made using recombineering methodologies [63] either with direct antibiotic cassette integration, through scarless tet-sacB selection and counterselection, adapted from Li et al. (Li et al., 2013) The recombineering plasmid pSIM5 and the tet-sacB selection/counterselection marker cassette were kind gifts from Donald Court (NCI, https://redrecombineering.ncifcrf.gov/court-lab.html). Oligonucleotides and synthetic linear DNA used for strain construction (Gblocks™), given in Table 4, were obtained from Integrated DNA Technologies (IDT, Coralville, IA). Briefly, the tet-sacB selection/counterselection cassette was initially amplified with tetA_F and sacB_R. Subsequently, PCR amplification using the appropriate oligos was performed to add ~50 bp flanking homology sequences to target specific genes (*ompT* a membrane associated protease) using Econotaq (Lucigen Middleton, WI) according to manufacturer's

instructions, with an initial 10 minutes denaturation at 94 ℃, followed by 35 cycles of 94 ℃, for 15 seconds, 52 ℃ for 15 seconds, and 72 ℃ for 5 minutes. A cassette for "curing" the tet-sacB cassette and simultaneous deletion of *ompT* was obtained as Gblocks™ from IDT. To change the native promoter of the gene *fadD* to the low phosphate inducible *yibD* gene promoter and replace the *fadE* to with *Fatb1*, synthetic linear DNA coding the desired chromosomal changes, along with ~50 bp of flanking homology sequences targeting *fadD* or *fadE* and an antibiotic cassette for direct selection (gentamicin for *fadD* and zeocin for *fadE*) were obtained as Gblocks^{TM}, and introduced by standard recombineering methods. Chromosomal modifications were confirmed by PCR amplification and sequencing (Eton Biosciences) using paired oligonucleotides as given in Table 4, flanking the entire region. Protease knock out strains were made using recombineering plasmid pSIM5 and either (i) a tetA cassette obtained through PCR to introduce ~50 bp homology regions to target genes or (ii) Gblocks^{TM} with ~50 bp homology regions to target genes that introduces a das-4 tag at the C-terminus and has an ampicillin resistance marker for selection.

Table 4: Oligonucleotides and Synthetic DNA utilized in Chapter 6

Name	**Sequence**
tetA_F	TCCTAATTTTTGTTGACACTCTATC
sacB_R	ATCAAAGGGAAAACTGTCCATATGC

ompT_tetA	AGATATAAAAAATACATATTCAATCATTAAAACGATTGAATGGA GAACTTTTTCCTAATTTTTGTTGACACTCTATC
ompT_sacB	GAAATGGCTAGTTATTCCCCGGGGCGATTTTCACCTCGGGGAAAT TTTAGTTGATCAAAGGGAAAACTGTCCATATGC
ompT_500up	AACGGATAAGACGGGCATAAAT
ompT_500dn	AGATTAAGGGATGAAGGAACGTC
fadD_500up	ATTTCCGTGGACAACTGGTTA
fadD_500dn	GGACGGCTTCACACAAAG
fadD_seq_F	AACTGAATAATTGCTTGTTTTT
fadD_seq_R	GCTCAAACATATCTACCAGAGA
fadE_500up	ATGACTAACGTCAGAAATAGC
fadE_500dn	CGGGAGGAATGATGTTTAAG
fadE_seq_F	ATGTTTTTACATCCACTACAACC
fadE_seq_R	ATCCGGATGGCTTTAATTT

fadE_mid	ATGGGTACGTTTGACCACC
SL1	CAGTCCAGTTACGCTGGAGTC
SR2	GGTCAGGTATGATTTAAATGGTCAGT
pCDF_piece1_F	GACGAATTCTCTAGATATCGC
pCDF_piece1_R	GCAGGTATCTTCGAGCCA
pCDF_piece2_F	TGGCTCGAAGATACCTGC
pCDF_piece2_R	GGAAACCGTTGTGGTCTC
ClpX_250up	TGGCGCTTCATACGGGTCAA
ClpX_250dn	GGATAAACCACCACATCGCG
ClpX_tet_F	CATTTGCGTCGTCGTGTGCGGCACAAAGAACAAAGA AGAGGTTTTGACCC TCCTAATTTTTGTTGACACTCTATC
ClpX_tet_R	TACTGCTTGGTCAGGGCGTTTTTCGGCTCTTTGAGGATCTGA ATCAGAGCATCATTTGGTGACGAAATAACTAAG

ClpP_500up	CTGCTGCTGGGCGAAGTTAT
ClpP_500dn	TCGGACCGATCAGCAGAATG
ClpP_tet_F	AGGTTACAATCGGTACAGCAGGTTTTTTCAATTTTATCCAGG AGACGGAATCCTAATTTTTGTTGACACTCTATC
ClpP_tet_R	CTGGATAAGTATAGCGGCACAGTTGCGCCTCTGGCAtcaATT ACGATGGGATCATTTGGTGACGAAATAACTAAG
HslUV_500up	CTAAACCGGCTTCTTCTCAG
HsIUV_500dn	TAGGTGATAGCGGCAATGAT
HsIUV_tet_F	TACTTTTGTACGGGGTTTGTACTCTGTATTCGTAACCAAG GGGTCAGCTC TCCTAATTTTTGTTGACACTCTATC
HsIUB_tet_R	AAACAATGATGAAAATGATTGAACGCGAttaTAGGATAAAAC GGCTCAGAATCATTTGGTGACGAAATAACTAAG
lon_500up	CGCCCTGACCAAGCAGTATC
lon_500dn	ACTGGGGACAACGCTTAGTT

lon_tet_F	CAGTCGTGTCATCTGATTACCTGGCGGAAATTAAACTAAGA GAGAGCTCT TCCTAATTTTTGTTGACACTCTATC
lon_tet_R	AAGCCCGAATTAGCCTGCCAGCCCTGTTTTTATTAGTGCATT TTGCGCGAATCATTTGGTGACGAAATAACTAAG
PepD_250up	TTCTGCGCGTTTCAGCACTT
PepD_250dn	CGATAACGCATGGGCATTAA
pepD_tet_F	CTCAAACATATCTCGCAAGCCTGTCTTGTGTTGACAACAT TTTCTGCTAATCCTAATTTTTGTTGACACTCTATC
pepD_tet_R	CAACCCGCGACGGACATTTACAGGGATGGCTTTTATCGAAG GATAATGAAATCATTTGGTGACGAAATAACTAAG
PtrB_500up	ATCGACCAGCAATTGCGTGG
PtrB_500dn	GTCGCCGATAAACCGTGGA
ptrB_tet_F	GGTGAGTTTTGCCACCCTTATAAGATGTTTCAACCAGAAA GAACAATAACTCCTAATTTTTGTTGACACTCTATC
ptrB_tet_R	GCATGAGCCCGGAGCTGCGTTTAACACTGAAACATTATCTG

	GAAAATACATCATTTGGTGACGAAATAACTAAG
prlC_500up	AACCGGGGTAGTAACGCAAC
prlC_500dn	GCGGGGTAATATCAGTCAGC
prlC_tet_F	TGAAATTCACTACACTTAACCCCATGCTACACACATTATG TAAAGCGCCTTCCTAATTTTTGTTGACACTCTATC
prlC_tet_R	GCCCTTAATGCCGTAATGCTCCAGCATCGCATCCAGCTGCG GTTCACGACATCATTTGGTGACGAAATAACTAAG
pepN_500up	TTCATCATCCTGCAGGCGCA
pepN_500dn	TTAAAGGCCAGTGGCGCGAT
pepN_tet_F	ACAGGAATAGACTGAACACCAGACTCTATAAAAGATGCTA AAGGTTATTTTCCTAATTTTTGTTGACACTCTATC
pepN_tet_R	TTATTTAtcaAGCCAGTGCTTTAGTTATCTTCTCGTACAGATC GCCAGAG ATCATTTGGTGACGAAATAACTAAG
HtpX_500up	CATCATGAAGGATATCGCGC

HtpX_500dn	TTACTTCTGACACCGTGGCC
HtpX_tet_F	ATGTGGGTATCGCATATTGCGTTTTGTTAAACTGAGGTAAA AAGAAAATTTCCTAATTTTTGTTGACACTCTATC
HtpX_tet_R	GACGCGCTTTTTAGTATTTACTTCATAttaCTTCAGGTATTCAC CCGTAATCATTTGGTGACGAAATAACTAAG
sppA _500up	GCATCCGGCAGACATTTCCT
sppA_500dn	CCGGTTTACCGAGATTCTCG
sppA_tet_F	GACAGGTGTGACCTTAAGTTGGGAGAATACatgCGAACCCTT TGGCGATTTCCTAATTTTTGTTGACACTCTATC
sppA_tet_R	CTGTCGGCCACTCAGTACAAGACttaACGCATGTTGGCGCAG GTCAGGCAATCATTTGGTGACGAAATAACTAAG
sohB_500up	TCCAGACGCGGATAATTAACGG
sohB_500dn	GCGCTATAGCACGACGGAATTT
sohB_tet_F	GCCTGTTGTAAACTGTGAGCCAAAGCGTTGTTTAACCAAGG TGGGGACTCTCCTAATTTTTGTTGACACTCTATC

sohB_tet_R	CTTAGCCTCGCGTTTGTCTTttaCATCAATGGCTTTTGACCCCG CTGCCAATCATTTGGTGACGAAATAACTAAG
degP_500up	GTGTTTAGCCATCCAGATGT
degP_500dn	ATTTCAGCCGTCATGCAGAC
degP_tet_F	TCTGAAGAACACAGCAATTTTGCGTTATCTGTTAATCGAG ACTGAAATACTCCTAATTTTTGTTGACACTCTATC
degP_tet_R	GTTTTCAGGAAGGGGTTGAGGGAGAttaCTGCATTAACAGGT AGATGGTGATCATTTGGTGACGAAATAACTAAG
degQ_500up	CTGTGGCCATGCGTTTTGGT
degQ_500dn	ACCGCCAGTGGCATTAATTT
degQ_tet_F	AAACTGTTTTGAATCTCTTTTCTTATCATTCAGGTACGAG AGCAGGAATATCCTAATTTTTGTTGACACTCTATC
degQ_tet_R	TTTTAATACCTTTGCCGCCATCTTTTAGCTGACCATCGCTCA ACGTTGCAATCATTTGGTGACGAAATAACTAAG
degS_500up	AAACCGCTGAATAGCTTTGC

degS_500dn	GAAGTTGAAGTGCCGGTTAT
degS_tet_F	TTAACTCGTGGTATGCTGCTGCCGTTCCCTTTTTTAATGA CGCCTCCATCTCCTAATTTTTGTTGACACTCTATC
degS_tet_R	CAGACTCCGGTTTTTTGTTTTGAGCGCACGACttaATTGGTTG CCGGATAATCATTTGGTGACGAAATAACTAAG
prc_500up	GTGTCGATCTTGACGGCAAC
prc_500dn	ATGGTAGATAGCCACTTGCG
prc_tet_F	TGGTGTTCTGAAACGGAGGCCGGGCCAGGCatgAACATGT TTTTTAGGCTTCCTAATTTTTGTTGACACTCTATC
prc_tet_R	ATTTCTTGTGCCTGATTGATAttaCTTGACGGGAGCGGGTTGT TCCGCGGATCATTTGGTGACGAAATAACTAAG
ClpA_500up	ACAGGAATGTGACAGATGTC
ClpA_500dn	CAACTCCCTATAGTAGCGCC
clpA_tet_F	GGAGGTGCCTatgCTCAATCAAGAACTGGAACTCAGTTTA AATATGGCTTTCCTAATTTTTGTTGACACTCTATC

clpA_tet_R	TTACGGACTTGACCAACCTACCTAACAATCAGAttaATGCGCT GCTTCCGATCATTTGGTGACGAAATAACTAAG
ClpS_500up	TGGACATCAAACTGAACGGA
ClpS_500dn	AGGATCAGAAGACTGTGTCG
clpS_tet_F	GCGTTCTGCCGATAACCGTAACCGAAGATGATAACTGACA atgGGTAAAATCCTAATTTTTGTTGACACTCTATC
clpS_tet_R	CCAGTTCTTGATTGAGCATAGGCACCTCCCCCAATTTTTATG CCTGCATtATCATTTGGTGACGAAATAACTAAG
Synthetic DNA	
yibD-fadD-GentR	
AACTGAATAATTGCTTGTTTTTAAAGAAAAAGAAACAGCGGCTGGTCCGCT GTTTCTGCATTCTTACGGTAAAGATAAAAATAAATAGTGACGCGCTTCGCGAATC CATGTGGGAGTTTATTCTTGACACAGATATTTATGATATAATAACTGAGTAAGCTT AACATAAGGAGGAAAAACATATGTTACGCAGCAGCAACGATGTTACGCAGCAGG GCAGTCGCCCTAAAACAAAGTTAGGTGGCTCAAGTATGGGCATCATTCGCACATG TAGGCTCGGCCCTGACCAAGTCAAATCCATGCGGGCTGCTCTTGATCTTTTCGGTC GTGAGTTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTCCGATTACCT CGGGAACTTGCTCCGTAGTAAGACATTCATCGCGCTTGCTGCCTTCGACCAAGAA GCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCCAAGTTTGAGCAGCCGCGTAG TGAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAGGGCATT GCCACCGCGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTGGTGCTTATGT GATCTACGTGCAAGCAGATTACGGTGACGATCCCGCAGTGGCTCTCTATACAAAG TTGGGCATACGGGAAGAAGTGATGCACTTTGATATCGACCCAAGTACCGCCACCT	

ATGCCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTT TTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCG GGTGGGCCTTTCTGCGTTTATACACAGCTAACACCACGTCGTCCCTATCTGCTGCC CTAGGTCTATGAGTGGTTGCTGGATAACGTGCGTAATTGTGCTGATCTCTTATATA GCTGCTCTCATTATCTCTCTACCCTGAAGTGACTCTCTCACCTGTAAAAATAATAT CTCACAGGCTTAATAGTTTCTTAATACAAAGCCTGTAAAACGTCAGGATAACTTCT ATATTCAGGGAGACCACAACGGTTTCCCTCTACAAATAATTTTGTTTAACTTTCGT AAAGAGGAGAAATACTAGTTGAAGAAGGTTTGGCTTAACCGTTATCCCGCGGACG TTCCGACGGAGATCAACCCTGACCGTTATCAATCTCTGGTAGATATGTTTGAGCAG TCGGTCG
yibD-ΔfadE::CC FatB1-ZeoR
ATGTTTTTACATCCACTACAACCATATCATCACAAGTGGTCAGACCTCCTAC AAGTAAGGGGCTTTTCGTTTTACTAAACTGAACTTTCTGCCGGAATGACGCTGATG CCACGAAAGCTATCGGTAAGTTTCGGACGCCATTCGGTCTTAGCGCGCAGCACTT CGCTACCTCCTTCCAGCTGCAGCAGATGTTCGCAAACCAGACCTGCTTCACTGCTA CCGCCCGATACCGTCGTCAGGCTCTGTAACACACTATCCATGGTACATTCACGAC GATATTCGATGGTAAAGCTACTGATATGATGGCTTTCGAAGATGCTATCTGGCAC GGTTTCCAGAATCCAGTCAACGTATTTGATGTTGTTGACATGTTGGTTGATGTCCA AATCATTCCAGCGCGGGGTAAGTCCCCCTTGGATATAATCGGCGGTGCTATCGTTC AATTTCTGCGGTTTTTTAATTTCCTCATCTTTCACGGCCACGTTGTCAATGAAAGCC GGGCCAATTTCACCACGCACTTCCTCCGGGATTTTAGACAGACGACGGGTACGAG TGTTCATCATAACGCTCAGGCTCGTGCAGCGGGTCAGAATCTCCCCCGTTTTGCAA TCACGCACTAAGAAATCGTGACGACGCCCGTTATTGCCGGACGCCCCTACCCAAC ACTCCACTTCAACTGTGTCGCCCCATGCCGGATAGCGCTCCACGGCAACATGGGT ACGTTTGACCACCCAAATAAGATCACGCTTAGACATTTCTAACGTGGTGCCAAAA CCATCTCCCAGGATACCAACAGATTTGGCATGGTTGAGTGCCGCCTCCTGTAAGTG GTTCATGACTGCCACGATGCTAGTACTACGGTCTGGGCCCACTTCATAAGAACGG ATGGCGAAGGTGCGGCGGAACACCAGACCATGCGGACCGAAATGGTCGTCCAGG AGCTGTGGCGGATTTGGTTTGGGTTTCCACTCCAGATTCGTCCACTGCTTTTCCGC CGCACTAAAAATGGTAGTGATAACCGCGAATAACATGCTCCAATCCGGCAGTTTT TTAAGGGATTCGGTGTAGCTAAATTTGGTACCGTTGATCATTTTCAGGCTGGTCTG TGCATTGCCGGCACGTAACTGTAAGTCAGAGCTACGCGGTTTCATCCCACGTCCAT CGCGAGCCAGCATAACTGCTTTCATCGAACAAAAAGCGCTCGCCAGGGATGTGGT TGCCATCTAGTATTTCTCCTCTTTACGAAAGTTAAACAAAATTATTTGTAGAGGGA AACCGTTGTGGTCTCCCTGAATATAGAAGTTATCCTGACGTTTTACAGGCTTTGTA TTAAGAAACTATTAAGCCTGTGAGATATTATTTTTACAGGTGAGAGAGTCACTTCA GGGTAGAGAGATAATGAGAGCAGCTATATAAGAGATCAGCACAATTACGCACGT TATCCAGCAACCACTCATAGACCTAGGGCAGCAGATAGGGACGACGTGGTGTTAG CTGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACTCAC

TATAGGAGGGCCATCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGC
GCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGA
CTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATC
AGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGTGGGTGC
GCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCG
GGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGA
GTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTTGTGGCAGAGGAGCAG
GACTGAGGATAAGATAACGGAGCCGAAAGGCTCCGTTTCTTTATCCGCTAATTAT
TTAAAATTAAAGCCATCCGGATGGTTTTC

yibD-Δ80-CBD-hNMT-1

CTCACCTGTAAAAATAATATCTCACAGGCTTAATAGTTTCTTAATACAAAG
CCTGTAAAACGTCAGGATAACTTCTATATTCAGGGAGACCACAACGGTTTCCCTCT
ACAAATAATTTTGTTTAACTTTCACAGCTAACACCACGTCGTCCCTATCTGCTGCC
CTAGGTCTATGAGTGGTTGCTGGATAACGTGCGTAATTGTGCTGATCTCTTATATA
GCTGCTCTCATTATCTCTCTACCCTGAAGTGACTCTCTCACCTGTAAAAATAATAT
CTCACAGGCTTAATAGTTTCTTAATACAAAGCCTGTAAAACGTCAGGATAACTTCT
ATATTCAGGGAGACCACAACGGTTTCCCTCTACAAATAATTTTGTTTAACTTTTCA
ATCTAAATTAGTAAGGAGGTAGTCAATGACAAATCCTGGTGTAAGTGCCTGGCAA
GTTAATACCGCATATACCGCTGGGCAGTTAGTCACTTATAACGGCAAGACCTACA
AGTGCTTGCAGCCTCACACATCCTTGGCAGGTTGGGAACCGTCCAATGTACCCGC
CCTTTGGCAACTTCAGGGCTCTGCCGGTAGTGCGGCGGGTTCCGGTGAATTTAACT
CGTTGCCTGCTGAGCGTATCCAAGAAATTCAGAAGGCAATTGAACTGTTTAGCGT
GGGTCAAGGCCCAGCCAAAACGATGGAGGAAGCGAGCAAACGTTCGTATCAGTT
CTGGGATACGCAACCGGTGCCGAAGCTCGGTGAAGTGGTGAATACGCACGGGCCT
GTTGAGCCGGATAAGGACAATATTCGTCAGGAGCCATATACGCTGCCTCAGGGTT
TCACTTGGGACGCCCTGGACCTGGGTGACCGTGGTGTGCTGAAAGAACTGTACAC
CCTGCTTAATGAGAATTATGTAGAAGATGACGACAACATGTTCCGTTTTGACTATA
GCCCGGAATTCCTGTTATGGGCACTCCGTCCGCCGGGTTGGCTGCCGCAGTGGCA
TTGCGGTGTCCGCGTAGTTTCGAGCCGTAAACTCGTAGGTTTCATCAGTGCAATCC
CGGCCAACATTCATATCTATGACACCGAGAAAAAAATGGTAGAAATTAACTTCCT
GTGTGTTCATAAGAAGTTGCGTAGCAAACGCGTAGCGCCTGTCCTCATTCGTGAA
ATCACGCGCCGCGTACATTTAGAAGGTATCTTCCAGGCAGTATATACTGCTGGCGT
CGTGCTCCCAAAACCGGTTGGGACTTGCCGCTATTGGCACCGCTCTCTGAATCCGC
GTAAGCTGATTGAAGTTAAATTTAGCCATTTGTCACGCAACATGACCATGCAGCG
CACCATGAAACTTTACCGTCTGCCGGAAACCCCGAAAACTGCTGGTTTGCGCCCA
ATGGAGACGAAAGATATTCCTGTCGTCCATCAGCTGCTGACGCGTTATTTAAAAC
AGTTTCACTTAACTCCTGTCATGAGCCAGGAAGAGGTTGAACATTGGTTTTATCCG
CAAGAAAACATCATCGACACCTTCGTAGTGGAGAATGCGAATGGCGAAGTCACG

GACTTTTTATCCTTCTATACTTTGCCGAGCACCATCATGAACCATCCGACCCATAA
AAGCCTGAAGGCCGCGTACTCATTTTATAATGTCCACACGCAGACCCCGTTATTGG
ATCTGATGTCTGATGCGTTGGTCCTGGCCAAAATGAAAGGTTTCGACGTTTTTAAT
GCGCTGGACCTGATGGAGAACAAAACCTTTCTGGAAAAATTGAAATTCGGAATTG
GCGATGGTAATCTGCAATACTATCTGTATAATTGGAAATGCCCGTCGATGGGTGC
GGAAAAAGTTGGTCTGGTACTGCAGTAGTAAGACGAATTCTCTAGATATCGCTC

lspA-das4-AmpR

GCAGTGATGATGTATCGCTCGAAGGCCACGCAGAAGCTAAACAATATCGC
TTACGCGCTGATTATTGGCGGCGCGCTGGGCAACCTGTTCGACCGCCTGTGGCAC
GGCTTCGTTGTCGATATGATCGACTTCTACGTCGGCGACTGGCACTTCGCCACCTT
CAACCTTGCCGATACTGCCATCTGTGTCGGTGCGGCACTGATTGTGCTGGAAGGTT
TTTTGCCTTCTAGAGCGAAAAAACAAGCGGCCAACGATGAAAACTATTCTGAAAA
CTATGCGGATGCGTCTTAATGAGTGATCGGCACGTAAGAGGTTCCAACTTTCACC
ATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGA
GCTAAGGAAGCTAAAATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAG
ATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAG
CGGTAAGATCCTTGAGAGTTTACGCCCCGAAGAACGTTTTCCAATGATGAGCACT
TTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCA
ACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCA
CAGAAAAGCATCTCACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCAT
AACCATGAGTGATAACACTGCGGCCAACTTACTTCTGGCAACGATCGGAGGACCG
AAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATC
GTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGA
TGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTAC
TCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGA
TCACTTCTGCGCTCGGCCCTCCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGC
CGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCC
TCCCGCATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAA
ATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAATAAACCCT
GCCGGATGCGATGCTGACGCATCTTATCCGGCCTACAGATTGCTGCGAAATCGTA
GGCCGGATAAGGCGTTTACGCCGCATCCGGCAAAAATCCTTAAATATAAGAGCAA
ACCTGCATGTCTGAATCTGTACAGAGCAATAGCGCCGTCCTGGTGCACTTCACGCT
AAAACTCGACGATGGCACCACCGCCGAGTCTACCCGCAACAACGGTAAACCGGC
GCTGTTCCGCCT

ftsH-das4-AmpR
TGACCGACAATATGGATATTCTGCATGCGATGAAAGATGCTCTCATGAAAT ATGAGACTATCGACGCACCGCAGATTGATGACCTGATGGCACGTCGCGATGTACG TCCGCCAGCGGGCTGGGAAGAACCAGGCGCTTCTAACAATTCTGGCGACAATGGT AGTCCAAAGGCTCCTCGTCCGGTTGATGAACCGCGTACGCCGAACCCGGGTAACA CCATGTCAGAGCAGTTAGGCGACAAGGCGGCCAACGATGAAAACTATTCTGAAA ACTATGCGGATGCGTCTTAATGAGTGATCGGCACGTAAGAGGTTCCAACTTTCAC CATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGG AGCTAAGGAAGCTAAAATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTT TTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAA GATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACA GCGGTAAGATCCTTGAGAGTTTACGCCCCGAAGAACGTTTTCCAATGATGAGCAC TTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGC AACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTC ACAGAAAAGCATCTCACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCA TAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGGCAACGATCGGAGGACC GAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGAT CGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACG ATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTA CTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGG ATCACTTCTGCGCTCGGCCCTCCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAG CCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCC CTCCCGCATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGA AATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAATTCCCGC ATCAGATGACTGTATTTGTACCGAAAACCCCGGGGCGTGCTCCGGGGTTTTTTCTT ATCAATTCATACCAGGGATAACATCATGAAACTCTTTGCCCAGGGTACTTCACTGG ACCTTAGCCATCCTCACGTAATGGGGATCCTCAACGTCACGCCTGATTCCTTTTCG GATGGTGGCACGCATAACTCGCTGATAGATGCGGTGAAACATGCGAATCTGATGA TCAACGCTGGCG

6.2.3 Plasmid cloning

The design and construction of plasmids as discussed above utilized the primers given below in Table 4. Synthetic linear DNA coding Δ80-CBD-hNMT-1 (including an

N terminal chitin binding domain and the first 80 amino acids of wild-type hNMT-1) was obtained as a Gblock™. This included ~20 bp flanking homology regions to clone the gene into pCDF expression vector using NEBuilder® HiFi DNA Assembly Mix (New England Biolabs, Ipswich, MA) according to manufacturer's instructions. Briefly, pCDF was linearized in two pieces using as template a vector already containing the *yibD* gene promoter obtained from plasmid pCDF-IN:yibDp-MatB (Addgene #89631). The PCR reactions used oligonucleotides sets pCDF_piece1_F/R and pCDF_piece2_F/R, with an initial 2 minutes 30 seconds denaturation at 94 ℃, followed by 35 cycles of 94 ℃ for 15 seconds, 59/63 ℃ for 15 seconds (piece 1/piece2), and 72 ℃ for 1 minute. The PCR products were gel purified using a DNA gel recovery kit (Zymo Research, Irvine, CA) and used together with 100 ng of the purchased Gblocks™ to perform a Gibson reaction for 15 minutes at 50℃. pSMART-M-Nef and pSMART-M-Nef_Δ80-CBD-hNMT-1 were purchased from Twist Biosciences (San Francisco, CA).

Further modifications to the plasmids were obtained using Q5® High-Fidelity PCR Mix (New England Biolabs, Ipswich, MA). The N-terminal SUMO tag was removed from the peptide substrate to make pNAP-1-noSUMO using oligonucleotides remove SUMO_F/R with an initial 2 minutes 30 seconds denaturation at 94 ℃, followed by 35 cycles of 94 ℃ for 15 seconds, 61 ℃ for 15 seconds, and 72 ℃ for 2 minutes and 30 seconds. The first N-terminal 80 amino acids of hNMT-1 were deleted from pCDF-His-hNMT-1 to make pCDF-yibD-delta80-His-hNMT-1 using oligonucleotides Δ80_hNMT-1_F/R with an initial 2 minutes 30 seconds denaturation at 94 ℃, followed by 35 cycles of 94 ℃ for 15 seconds, 63 ℃ for 15 seconds, and 72 ℃ for 2 minutes and

30 seconds. The PCR products were gel purified using a DNA gel recovery kit (Zymo Research, Irvine, CA), and circularized during one hour at room temperature, using T4 DNA ligase in 1x T4 DNA ligase reaction buffer, in the presence of T4 polynucleotide kinase (3' phosphatase minus) and DpnI, all purchased from New England Biolabs, Ipswich, MA. All plasmids sequences were confirmed by DNA sequencing (Eton Bioscience, NC). Strains and plasmids used in this study are summarized in Table 5 below.

Table 5: Strains and plasmids used in Chapter 6

Plasmids			
Name	Genes	Addgene #	Source
pSMART-HC-Kan	ColE1, Kan, empty vector	--	Lucigen
pCDF	CloDF13 origin, Sp, empty vector	89596	This Study
pCDF-Δ80-CBD-hNMT-1	CloDF13 origin, Sp, Δ80 human N-myristoyl transferase 1	87683	This Study
pSMART-M-Nef	ColE1, Kan, M-GNAASARR-Nef	N/A	This Study
pSMART-Duet-M-Nef_Δ80-CBD-hNMT-1	ColE1, Kan, M-GNAASARR-Nef, hNMT-1	N/A	This Study
Strains			

Name	Genotype	Plasmid(s)	Source
DLF_Z0025	F-, *λ*-, *Δ(araD-araB)567*, *ΔlacZ4787*(::rrnB-3), *rph-1*, *Δ(rhaD-rhaB)568*, *hsdR514*,BW25113, *ΔldhA::frt*, *ΔpoxB::frt*, *ΔpflB::frt*, *ΔackA-pta::frt*, *ΔadhE::frt*, *ΔiclR*, *ΔarcA*, *ΔsspB*, *Δcas3::ugpBp-sspB-pro-cas*	--	TBD
RLS_001	DLF_Z0025, *yibD-fadD*	--	This Study
RLS_002	DLF_Z0025, *ΔfadE::CC-FatB1*	--	This Study
RLS_003	DLF_Z0025, *ΔfadE::CC-FatB1*, *yibD-fadD*	--	This Study
RLS_003_lonKO	DLF_Z0025, *ΔfadE::CC-FatB1*, *yibD-fadD* *Δlon::tetA*	--	This Study
RLS_003_sppAKO	DLF_Z0025, *ΔfadE::CC-FatB1*, *yibD-fadD* *ΔsppA::tetA*	--	This Study
RLS_003_hslUVKO	DLF_Z0025, *ΔfadE::CC-FatB1*, *yibD-fadD* *ΔhslUV::tetA*	--	This Study
RLS_003_ClpSKO	DLF_Z0025, *ΔfadE::CC-FatB1*, *yibD-fadD*	--	This Study

	ΔClpS::tetA		
RLS_003_pepNKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔpepN::tetA*	--	This Study
RLS_003_degQKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔdegQ::tetA*	--	This Study
RLS_003_ptrBKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔptrB::tetA*	--	This Study
RLS_003_clpXKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔclpX::tetA*	--	This Study
RLS_003_prcKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD Δprc::tetA*	--	This Study
RLS_003_lspA-DAS4	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD lpsA-das4*	--	This Study
RLS_003_degSKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔdegS::tetA*	--	This Study

RLS_003_ftsH-DAS4	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ftsH-das4*	--	This Study
RLS_003_clpPKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔclpP::tetA*	--	This Study
RLS_003_htpXKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔhtpX::tetA*	--	This Study
RLS_003_sohBKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔsohB::tetA*	--	This Study
RLS_003_prlCKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔprlC::tetA*	--	This Study
RLS_003_pepDKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔpepD::tetA*	--	This Study
RLS_003_clpAKO	DLF_Z0025, *ΔfadE::CC-FatB1, yibD-fadD ΔclpA::tetA*	--	This Study
DLF_R0012	DLF_Z0025, *lon-das4, ftsH-das4,*	--	TBD

	ΔhslUV::tetA		

6.2.4 Shake Flask Experiments

Production of myristoyl-CoA, substrate peptides and hNMT-1due to phosphate depletion was performed in vented cap square flasks (Genesee Scientific, San Diego, CA). Spectinomycin and/or kanamycin were added when appropriate. Cell cultures were started from frozen stocks in 10 ml (50 ml for MALDI experiments) of Growth Medium. One liter of Growth Medium consists of 9 g ammonium sulfate, 0.25 g citrate, 2.5 g yeast extract, 45 g glucose, 5 mM of phosphate buffer, 200 mM MOPS buffer (pH=7.4), 2.5 mM magnesium sulfate, 0.06 mM calcium sulfate, 10 mg thiamine-HCl, 0.16 mM iron(II) sulfate, supplemented with 0.2 ml of trace metals (for 1 liter of trace metals: 10 ml sulfuric acid, 0.6 g cobalt (II) sulfate heptahydrate, 5 g copper (II) sulfate pentahydrate, 0.6 g zinc sulfate heptahydrate, 0.2 g sodium molybdate dihydrate, 0.1 g boric acid and 0.3 g manganese (II) sulfate monohydrate), pH 6.8. The cultures were left to grow to 3 OD (600 nm) by incubating at 37 ℃ and 220 rpm overnight. Cells were induced by centrifuging and resuspending them in Induction Medium (Growth Medium without phosphate or yeast extract); and incubating at 37 ℃ and 220 rpm. Cells were harvested 24 hours after induction.

6.2.5 Microfermentations

To test protease deficient strains, these were transformed with pSMART-Duet-M-Nef_Δ80-CBD-hNMT-1 and grown in a 96-well microtiter plate overnight to 3-5 OD (600 nm) by incubating at 37 ℃ and 220 rpm overnight. Cells were induced by centrifuging and resuspending them in Induction Medium (Growth Medium without phosphate or yeast extract); and incubating at 37 ℃ and 220 rpm. Cells were harvested 24 hours after induction.

6.2.6 Analytical Methods

Myristoyl-CoA was extracted and quantified as previously described (Blachnio-Zabielska et al., 2011) with some modifications, noted below. Briefly, cells were pelleted by centrifugation and resuspended in 0.5 ml of freshly made 100 mM potassium phosphate monobasic (pH 4.9) and 0.5 ml of acetonitrile:2-propanol:methanol (3:1:1). Cells were lysed using a Branson 4c15 sonicator (Branson Ultrasonics, Dansbury, CT) with 10 seconds, 30 seconds on/off cycles for 3 minutes at 50% tip amplitude. The cells were centrifuged at 14000 g and the supernatant was collected. The pellet was re-extracted using 0.5 ml of acetonitrile:2-propanol:methanol (3:1:1). After centrifuging again, the new supernatant was combined with the previous one and dried under nitrogen gas. The dried extract was resuspended in 200 ul of methanol:water (1:1). Myristoyl-CoA was separated using an Acquity UPLC (Waters Co, Milford, MA) in a Waters BEH C18 50 mm reverse phase column (Waters Co, Milford, MA). We used a gradient starting at 80% solvent A (15 mM NaOH in water) and 20% solvent B (15 mM NaOH in AcN), decreased it to 60% solvent A over 0.5 minutes, then decreased to 0% solvent A over 1

minute. The flow ran at 0% solvent A for 0.5 minutes, before turning back to the initial 80% solvent A gradient. It ran at 80% solvent A to re-equilibrate the column for the next sample for 1.5 minutes. Samples were detected using a Xevo TQD™ mass spectrometer (Waters Co, Milford, MA). A calibration curve was also performed using purchased myristoyl-CoA at 2, 1.6, 0.4, 0.1 and 0.025 mg/l. Extractions were performed in triplicates before and after phosphate induction. Values were normalized to biomass levels and always measured in the linear range of our calibration curve.

SDS-PAGE was performed using Mini-PROTEAN® TGX™ 4-20% gradient gels (Bio-rad, Hercules, CA) as previously described. (Laemmli, 1970) Samples were mixed 1:1 with 2X Laemmli sample buffer (Bio-rad, Hercules, CA). ~10 µg were loaded for lysates. Protein gels were stained using Coomassie Blue and imaged using a Kodak Gel Logic 200 imaging system (Kodak, Rochester, NY).

6.3 Results

6.3.1 Dynamic Biosynthesis of Acyl-CoAs

We first engineered strains for inducible fatty acyl-CoA accumulation, as illustrated in Figure 26a, this was accomplished by first deleting the native *fadE* gene from the *E. coli* chromosome, and replacing it with an inducible acyl-ACP thioesterase from *Cinnamomum camphorum* with a preference for myristoyl(C14) substrates.(Yuan et al., 1995) Together these modifications eliminate fatty acyl-CoA beta oxidation and enable the conversion of biosynthetic fatty acyl-ACPs into intracellular free fatty acids. Simultaneously, the native *E. coli fadD* gene encoding a fatty acyl-CoA synthetase

(which activates free fatty acids into fatty acyl-CoAs, the substrates for acyl-transfer) is induced by replacement of its native promoter with the low phosphate inducible *yibD* gene promoter. (Lipscomb et al., 2018; M. Lynch et al., 2019; Menacho-Melgar, Ye, et al., n.d.; Ye, Z., Lynch,M.D., Trahan, A.D., Rodriguez, D.L., Cooper, C.B. Bozdag, A., 2015; Youngquist et al., 2013) As seen Figure 26b, together these modifications enable the biosynthesis and accumulation of myristoyl-CoA *in vivo* upon phosphate depletion.

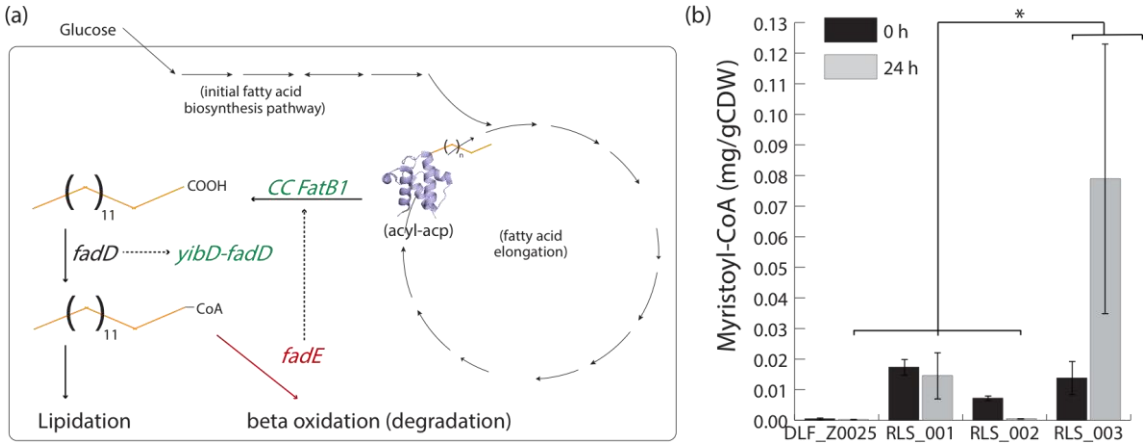

Figure 26: Dynamic accumulation of Acyl-CoAs. (a) Acyl-CoA biosynthesis is a control strain (DLF_Z0025) compared to dynamic acyl-CoA biosynthesis in a modified strain (RLS_011) Heterologous C. camphora acyl-ACP thioesterase (TE) expression leads to increased myristic acid production by cleaving myristate off myristoyl-ACP and alleviating the inhibition of ACCase by acyl-ACPs, this gene was introduced while simultaneously deleting the native fadE gene eliminating the beta oxidative pathway. Overexpression of the fadD (acyl-CoA synthetase) by replacing its native promoter with the low phosphate inducible yibD gene promoter results in accumulation of myristoyl-CoA (b) Myristoyl-CoA (Myr-CoA) synthesis in E. coli. Myr-CoA overproduction in engineered strains of E. coli before (white) and 24 hours after (grey) phosphate depletion. Control strain (DLF_Z0025) has baseline intracellular Myr-CoA. Overexpression of the fadD synthetase (RLS_001) and replacement of the fadE gene with a C. camphora acyl-ACP thioesterase (TE) gene (RLS_002) increase Myr-CoA pools. Deletion of fadE in combination with the overexpression of fadD and the TE lead to large increases in Myr-CoA (RLS_003).

6.3.2 In vivo N-terminal Myristoylation leads to reduced protein levels.

With successful improvements in acyl-CoA levels, we next turned to *in vivo* myristoylation. Toward this goal, as illustrated in Figure 25, we constructed a plasmid capable of expressing both hNMT-1 as well as the HIV Nef protein, a substrate previously myristoylated in *E. coli. (Duronio et al., 1990; Glück et al., 2010)* Introducing this plasmid into both the control strain (DLF_Z025) as well as the RLS_003 (producing increased myristoyl-CoA pools, Figure 26b) , surprisingly led to inconsistent expression and in many cases very low levels of Nef expression, as can be seen in Figure 27 for RLS_003. Expression of Nef and NMT in a control plasmid wherein an N-terminal purification tag of hNMT-1 was removed (eliminating hNMT-1 expression) confirmed that the reduced protein levels were a function of N-myristoylation (Data not shown). Interestingly, these results indicate that even basal levels of myristoyl-CoA in the control strain DLF_Z0025, are adequate in enabling N-myristoylation and reduced protein levels.

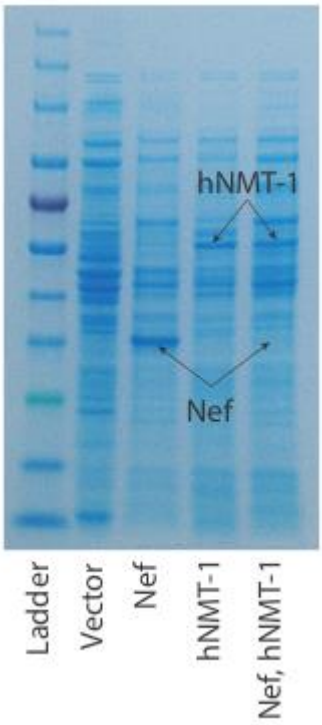

Figure 27: N-myristoylation reduces protein titers. Expression of Nef when expressed alone and/or with co-expression of hNMT-1.

6.3.3 Reduction of proteases rescued Nef expression

As the Nef expression construct leads to reasonable protein levels in the absence of hNMT-1, we hypothesized that the specific degradation of N-myristoylated Nef was the reason for reduced protein titers as opposed to reduced expression levels, and that intracellular proteases were preferentially degrading Nef with a N-terminal lipid group. To test this hypothesis, we constructed a set of strains with deletions in each of 15 protease and/or peptidase genes. In addition, as *ftsH* and *lspA* are essential, strains with a DAS+4 degron tags were used, as previously reported. (Michael David Lynch et al., 2017) This enables dynamic reduction in FtsH and LspA activity (Refer to Chapter 4).

Also, a strain with a deletion in HslUV as well as degron tags on the Lon and FtsH proteases, but lacks the myristoylation machinery was evaluated (DLF_R0012 strain, refer to Chapter 4). Results from whole cell preparations are shown in Figure 28. The only strain visibly increasing Nef expression, in the presence of hNMT-1 was the the DLF_R0012 strain, with reduced levels of FtsH, HslUV and Lon proteases. While this strain lacks the myristoylation machinery, we have also observed absence of Nef production in a DLFZ0025 when co-expressed with hNMT-1, as discussed above. Confirmation that the band corresponding to Nef (Figure 28, last lane) is (i) in the soluble fraction and (ii) myristoylated are pending.

We were surprised to see that decreasing FtsH levels alone did not lead to increased levels of lipidated Nef as FtsH is a processive protease localized in the membrane(Bittner et al., 2015) and lipidated proteins are expected to interact with the cell membrane. (Yael Wexler-Cohen & Shai, 2009) Moreover, FtsH is induced in stationary phase,(Bittner et al., 2015) and thus, co-induced with Nef and hNMT-1. The need to delete additional proteases to rescue expression of lipidated Nef may result from impaired degradation of the heat shock response regulator (σ32). A heat-shock response is commonly induced with recombinant protein expression and upregulates expression of chaperones and proteases alike (Gawin et al., 2019; Kanemori et al., 1994) FtsH specifically degrades σ32, tightly controlling its levels. (Meyer & Baker, 2011) Therefore, reduction of FtsH levels may maintain upregulation of heat-shock induced proteases by not allowing σ32 degradation. Both Lon and HslUV are processive proteases, have broad specificities and are induced during the heat-shock response. Thus,

coupling Ftsh with Lon and HslUV decreased levels may have increased levels of lipidated Nef through synergistic effects.

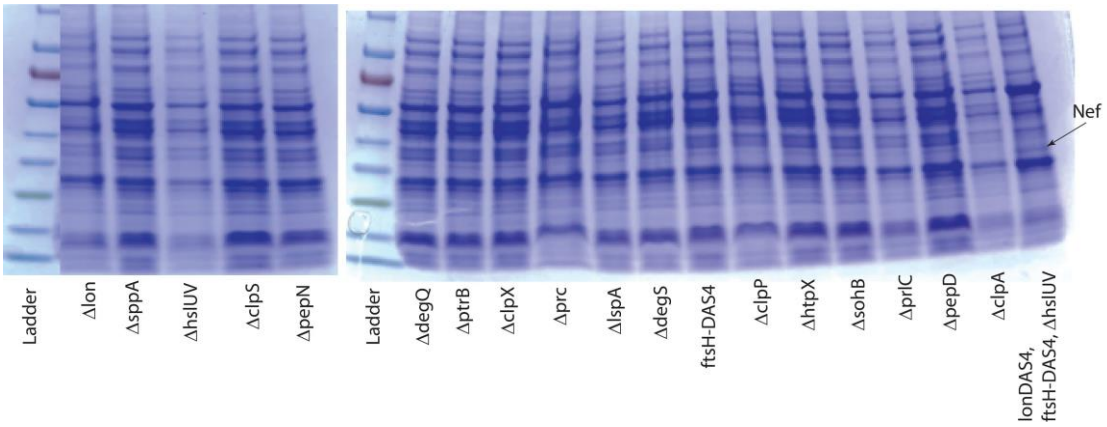

Figure 28: Impact of protease/chaperone deletions on expression of N-myristoylated Nef. Individual gene deletions as well as degron tagging (DAS+4) of lon and/or ftsH we evaluated for Nef expression (when hNMT-1 was coexpressed). Only in the case of a triple protease reduction (last lane on the right), was a significant increase in Nef expression obtained.

6.4 Discussion

We have developed an *E. coli* strain enabling improved *in vivo* enzymatic transfer of acyl groups to the N-terminus of proteins *in vivo* without the need for exogenously added fatty acids, enabling the facile production of myristoylated proteins and peptides.

Unexpectedly, we observed degradation of the target protein when lipidated. Since *in vivo* myristoylation of Nef (the target protein) by lipid media supplementation has previously been reported using T7 growth-associated production (Glück et al., 2010), degradation of myristoylated Nef may be associated with stationary phase, phosphate depletion, or higher expression levels obtained in our 2-stage process. We found that reduced levels of proteases FtsH, Lon and HslUV may rescue lipidated Nef expression

through synergistic effects (final analytical confirmation pending). Future engineering of the specificity of fatty acyl-CoA chain biosynthesis as well the substrate specificity of hNMT-1 have the potential to enable protein lipidation with a variety of acyl chains. These specificities can potentially be tuned to obtain a higher quantity and purity of desired lipidated peptides and proteins with a desired lipid chain length.

This toolkit would represent a next step toward direct, *in vivo,* and site specific post-translational modifications of therapeutic peptides and proteins which represent an enormous largely untapped space of molecular diversity in the discovery and engineering of advanced, next-generation biologics.

7. Conclusions

This work has been motivated by the lack of transformative and readily generalizable platforms for the production of recombinant proteins in E. coli. Despite being the most widely used host for protein production, E. coli has remained fundamentally the same since the inception of BL21(DE3) that combined a T7 RNA polymerase with a protease deficient E. coli strain (BL21).

While many recent strategies to improve protein production have adapted components from the existing standard methodologies (eg improving BL21(DE3) induction profiles or responses to a particular inducer), we have designed from the bottom up a new protein production platform that is highly adaptable to different applications.

Our platform results in high protein titers at different scales enabling, for example, effortless transitions from high-throughput enzyme screening to the large scale production of a particular enzyme in fermenters. Ease of handling is further improved by (i) autoinduction of protein expression bypassing the need to optimize expression of bioprocessing conditions and (ii) autolysis and autohydrolysis that replace mechanical lysis and facilitate high-throughput workflows.

Perhaps, the most important feature of this platform is the decoupling of cellular growth from protein production. This allows for the dynamic implementation of engineering strategies in the production phase, such as proteolysis and/or silencing of an essential gene, that would not be possible during the growth phase due to expected toxicity. Moreover, since growth is not associated with protein production, problems such as

protein toxicity or metabolic burden become easier to address. Importantly, we showed that dynamic control over major housekeeping processive proteases resulted in expression of a large hard-to-express protein. The versatility of this platform is also leveraged by its adaptation to the complete in vivo production of lipidated proteins in E. coli.

Importantly, future work to optimize both production of hard-to-express proteins through reduced proteolysis and production of lipidated proteins may reveal currently unknown mechanisms that negatively affect overall protein production. Moreover, expansion of this platform to genome-wide studies may also lead to potential routes to improve overall protein production, increase titer of specific classes of proteins such as antibodies, or enable specific post-translational processing such as membrane insertion or non-natural post-translational modifications.

Appendix A. Chapter 2 Supplementary Information

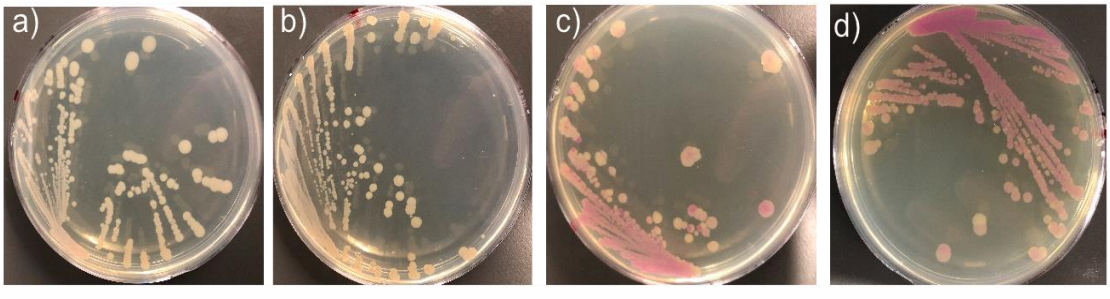

Figure 29: Leaky Expression of yibDp-mCherry in BL21(DE3).Agar plates with growth of plasmid pHCKan-yibDp-mcherry in a) BW25113, b) DLF_R002, c) BL21(DE3) and d) BL21(DE3) pLys

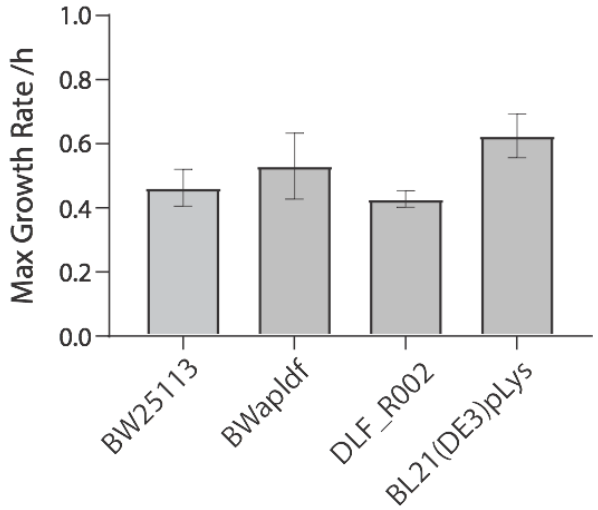

Figure 30: Maximal growth rates of E. coli strains in minimal media fermentations.Results are averages of duplicate fermentations.

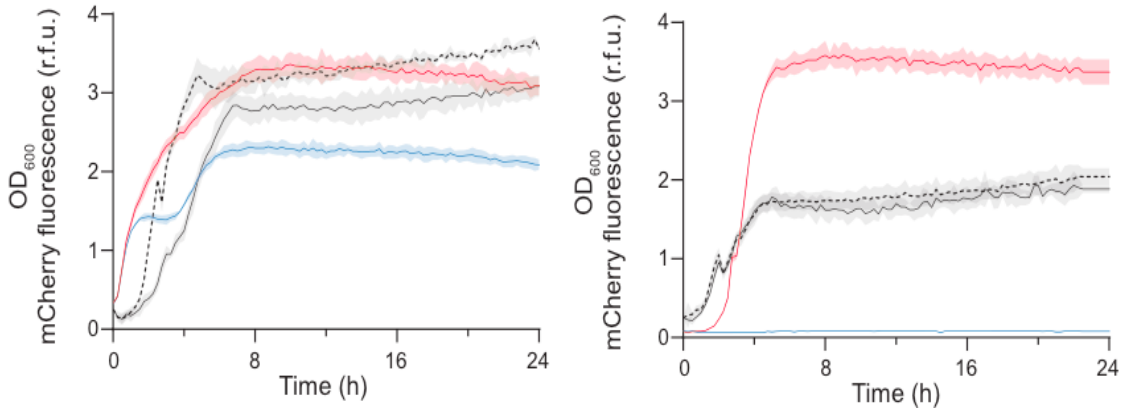

Figure 31: Standard expression results using BL21(DE3).LB media with IPTG based induction. mCherry total fluorescence (colored lines, red induced, blue uninduced) and OD600 (black lines, solid induced, dashed uninduced) are plotted over time. IPTG was added at inoculation. LEFT: BL21(DE3) grown in LB media with IPTG based induction. RIGHT: BL21(DE3) with pLys grown in LB media with IPTG based induction.

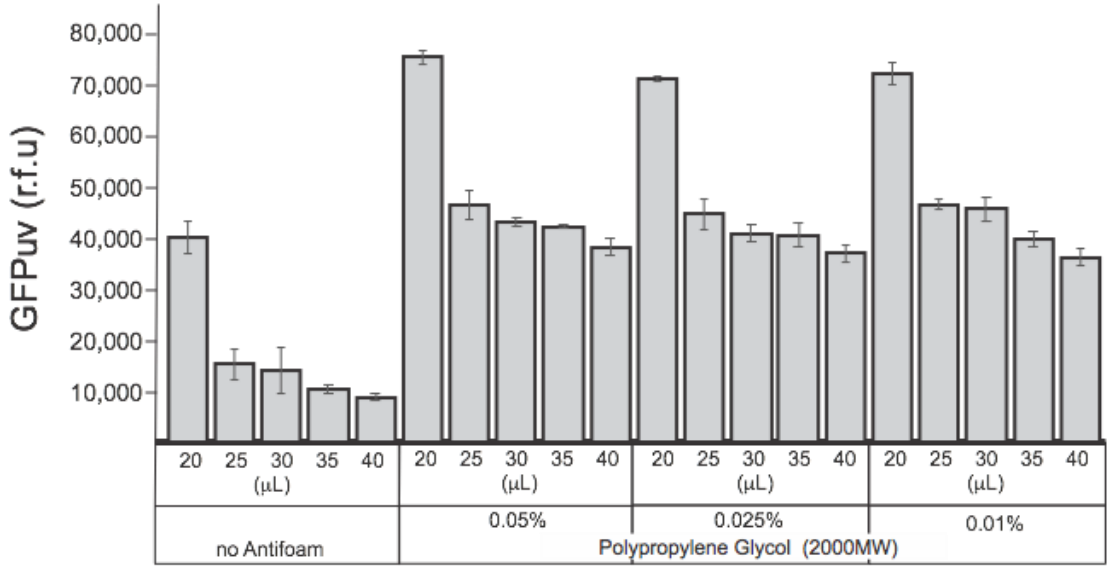

Figure 32: Impact of surfactants on expression in 384 well plates.Relative GFPuv levels in AB as a function of filling volume and antifoam concentration. Polypropylene glycol (MW of 2000) was used. % are in volume (v/v%). DLF_R002 with plasmid pHCKan-yibDp-GFPuv was used for all experiments.

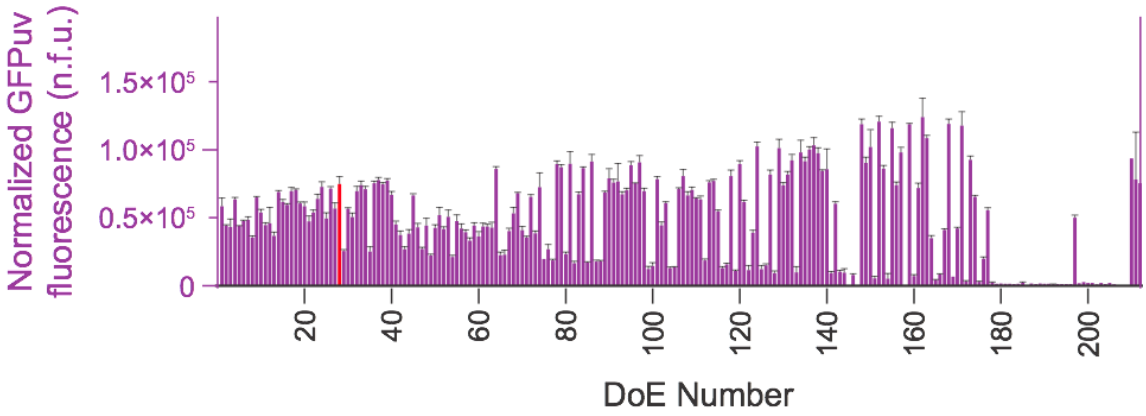

Figure 33: Normalized GFP/ OD600nm for DoE studies.Normalized fluorescence units (n.f.u) for each media formulation is given, which is the relative fluorescence (r.f.u) divided by the optical density at 600 nm. The data order is the same rank order as in Figure 3 in the main text. Media AB-C7 is highlighted in red, and reaches final optical densities from 7-10, or ~ 3 gCDW/L. DLF_R002 with plasmid pHCKan-yibDp-GFPuv was used for all experiments.

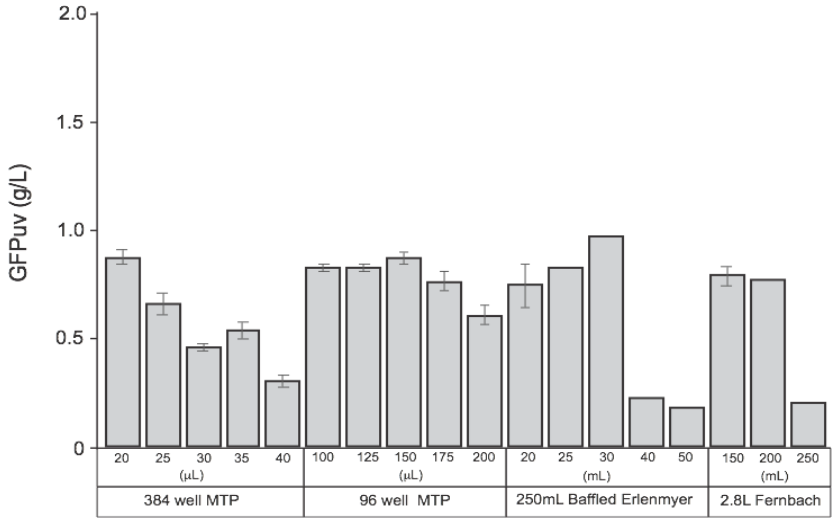

Figure 34: Autoinduction in batch cultures at various scales using AB-C7 media with lower supported biomass levels (~ 3gCDW/L). Varying fill volumes in 384 and 96 well plates as well as 250 mL baffled Erlenmeyer and 2.8 L Fernbach flasks. When using 384 well plates, 0.042% polypropylene glycol (2000MW) was added to the media. DLF_R002 with plasmid pHCKan-yibDp-GFPuv was used for all experiments. Where error bars are present, data are averages of at least triplicate experiments, when absent data are from single studies.

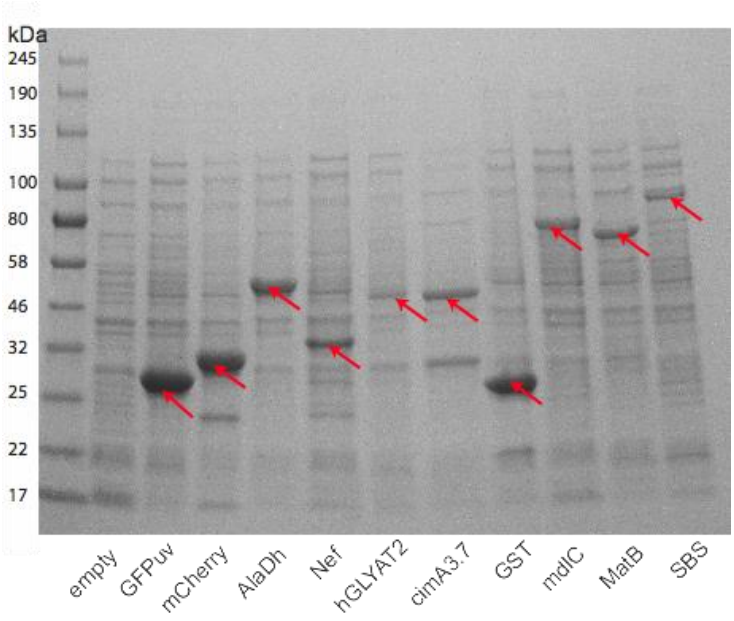

Figure 35: SDS PAGE Results for a diverse set of proteins. Annotated gel image. Samples taken after autoinduction in AB media using strain DLF_R002 and the appropriate plasmid from Table 1. Expression was performed in 96 well plates. Expression levels range from 10% in the case of SBS (a terpene synthase) to 55% in the case of GFPuv, mCherry, alanine dehydrogenase (AlaDh) and GST.

Table 6: yibDp sequence

Promoter Name	Sequence of promoter underlined, including example ribosomal binding site and start codon (green)
yibDp	GTGCGTAATTGTGCTGATCTCTTATATAGCTGCTCTCATTATCTCTCTACCCTGA AGTGACTCTCTCACCTGTAAAAATAATATCTCACAGGCTTAAT AGTTTCTTAATACAAAGCCTGTAAAACGTCAGGATAACTTCTGTGTAGGAGGAT AATCTATG

Table 7: DNA and Oligos used for strain construction used in Chapter 2.

Name	Sequence
tet-sacB Cassette	TCCTAATTTTTGTTGACACTCTATCATTGATAGAGTTATTTTACCACTCC CTATCAGTGATAGAGAAAAGTGAAATGAATAGTTCGACAAAGATCGC ATTGGTAATTACGTTACTCGATGCCATGGGGATTGGCCTTATCATGCCA GTCTTGCCAACGTTATTACGTGAATTTATTGCTTCGGAAGATATCGCTA ACCACTTTGGCGTATTGCTTGCACTTTATGCGTTAATGCAGGTTATCTT TGCTCCTTGGCTTGGAAAAATGTCTGACCGATTTGGTCGGCGCCCAGTG CTGTTGTTGTCATTAATAGGCGCATCGCTGGATTACTTATTGCTGGCTT TTTCAAGTGCGCTTTGGATGCTGTATTTAGGCCGTTTGCTTTCAGGGAT CACAGGAGCTACTGGGGCTGTCGCGGCATCGGTCATTGCCGATACCAC CTCAGCTTCTCAACGCGTGAAGTGGTTCGGTTGGTTAGGGGCAAGTTTT GGGCTTGGTTTAATAGCGGGGCCTATTATTGGTGGTTTTGCAGGAGAG ATTTCACCGCATAGTCCCTTTTTTATCGCTGCGTTGCTAAATATTGTCAC TTTCCTTGTGGTTATGTTTTGGTTCCGTGAAACCAAAAATACACGTGAT AATACAGATACCGAAGTAGGGGTTGAGACGCAATCGAATTCGGTATAC ATCACTTTATTTAAAACGATGCCCATTTTGTTGATTATTTATTTTTCAGC GCAATTGATAGGCCAAATTCCCGCAACGGTGTGGGTGCTATTTACCGA AAATCGTTTTGGATGGAATAGCATGATGGTTGGCTTTTCATTAGCGGGT CTTGGTCTTTTACACTCAGTATTCCAAGCCTTTGTGGCAGGAAGAATAG CCACTAAATGGGGCGAAAAAACGGCAGTACTGCTCGGATTTATTGCAG ATAGTAGTGCATTTGCCTTTTTAGCGTTTATATCTGAAGGTTGGTTAGT TTTCCCTGTTTTAATTTTATTGGCTGGTGGTGGGATCGCTTTACCTGCAT TACAGGGAGTGATGTCTATCCAAACAAAGAGTCATCAGCAAGGTGCTT TACAGGGATTATTGGTGAGCCTTACCAATGCAACCGGTGTTATTGGCC CATTACTGTTTGCTGTTATTTATAATCATTCACTACCAATTTGGGATGG CTGGATTTGGATTATTGGTTTAGCGTTTTACTGTATTATTATCCTGCTAT CGATGACCTTCATGTTAACCCCTCAAGCTCAGGGGAGTAAACAGGAGA CAAGTGCTTAGTTATTTCGTCACCAAATGATGTTATTCCGCGAAATATA ATGACCCTCTTGATAACCCAAGAGCATCACATATACCTGCCGTTCACTA TTATTTAGTGAAATGAGATATTATGATATTTTCTGAATTGTGATTAAAA AGGCAACTTTATGCCCATGCAACAGAAACTATAAAAAATACAGAGAAT GAAAAGAAACAGATAGATTTTTTAGTTCTTTAGGCCCGTAGTCTGCAA ATCCTTTTATGATTTTCTATCAAACAAAAGAGGAAAATAGACCAGTTG CAATCCAAACGAGAGTCTAATAGAATGAGGTCGAAAAGTAAATCGCG CGGGTTTGTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAAAGGGC AAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGGTCTTTTTTTAT TGTGCGTAACTAACTTGCCATCTTCAAACAGGAGGGCTGGAAGAAGCA GACCGCTAACACAGTACATAAAAAAGGAGACATGAACGATGAACATC AAAAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTG CTGGCAGGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAA GCCATATAAGGAAACATACGGCATTTCCCATATTACACGCCATGATAT GCTGCAAATCCCTGAACAGCAAAAAAATGAAAAATATCAAGTTCCTGA GTTCGATTCGTCCACAATTAAAAATATCTCTTCTGCAAAAGGCCTGGAC GTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAAC TATCACGGCTACCACATCGTCTTTGCATTAGCCGGAGATCCTAAAAAT GCGGATGACACATCGATTTACATGTTCTATCAAAAAGTCGGCGAAACT TCTATTGACAGCTGGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGAC AAATTCGATGCAAATGATTCTATCCTAAAAGACCAAACACAAGAATGG

	TCAGGTTCAGCCACATTTACATCTGACGGAAAAATCCGTTTATTCTACA CTGATTTCTCCGGTAAACATTACGGCAAACAAACACTGACAACTGCAC AAGTTAACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTAG AGGATTATAAATCAATCTTTGACGGTGACGGAAAAACGTATCAAAATG TACAGCAGTTCATCGATGAAGGCAACTACAGCTCAGGCGACAACCATA CGCTGAGAGATCCTCACTACGTAGAAGATAAAGGCCACAAATACTTAG TATTTGAAGCAAACACTGGAACTGAAGATGGCTACCAAGGCGAAGAA TCTTTATTTAACAAAGCATACTATGGCAAAAGCACATCATTCTTCCGTC AAGAAAGTCAAAAACTTCTGCAAAGCGATAAAAAACGCACGGCTGAG TTAGCAAACGGCGCTCTCGGTATGATTGAGCTAAACGATGATTACACA CTGAAAAAAGTGATGAAACCGCTGATTGCATCTAACACAGTAACAGAT GAAATTGAACGCGCGAACGTCTTTAAAATGAACGGCAAATGGTACCTG TTCACTGACTCCCGCGGATCAAAAATGACGATTGACGGCATTACGTCT AACGATATTTACATGCTTGGTTATGTTTCTAATTCTTTAACTGGCCCAT ACAAGCCGCTGAACAAAACTGGCCTTGTGTTAAAAATGGATCTTGATC CTAACGATGTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAA AGGAAACAATGTCGTGATTACAAGCTATATGACAAACAGAGGATTCTA CGCAGACAAACAATCAACGTTTGCGCCAAGCTTCCTGCTGAACATCAA AGGCAAGAAAACATCTGTTGTCAAAGACAGCATCCTTGAACAAGGAC AATTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGATATTGA CTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGATTCAGGC TGTCTATGTGTGACTGTTGAGCTGTAACAAGTTGTCTCAGGTGTTCAAT TTCATGTTCTAGTTGCTTTGTTTTACTGGTTTCACCTGTTCTATTAGGTG TTACATGCTGTTCATCTGTTACATTGTCGATCTGTTCATGGTGAACAGC TTTAAATGCACCAAAAACTCGTAAAAGCTCTGATGTATCTATCTTTTTT ACACCGTTTTCATCTGTGCATATGGACAGTTTTCCCTTTGAT
ΔilcR_cure	AAATGATTTCCACGATACAGAAAAAAGAGACTGTCATGGGCAGAATAT TGCCTCTGCCCG CCAGAAAAAG
ΔarcA_cure	CTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAACTCGGCTTTACCACC GTCAAAAAAAAC GGCGCTTTT
ilcR-tetA-F	TAACAATAAAAATGAAAATGATTTCCACGATACAGAAAAAAGAGACT GT CATCCTAATTTTTGTTGACACTCTATC
ilcR_sacB_R	TGCCACTCAGGTATGATGGGCAGAATATTGCCTCTGCCCGCCAGAAAA A GATCAAAGGGAAAACTGTCCATATGC
iclR_500up	CCGACAGGGA TTCCA TCTG
iclR_500dn	TATGACGACCATTTTGTCTACAGTTC
arcA-tetA-F	GGACTTTTGTACTTCCTGTTTCGATTTAGTTGGCAATTTAGGTAGCAAA CT CCTAATTTTTGTTGACACTCTATC
arcA_sacB_R	ATAAAAACGGCGCTAAAAAGCGCCGTTTTTTTTGACGGTGGTAAAGCC G AATCAAAGGGAAAACTGTCCATATGC
arcA_500up	CCTGACTGTACTAACGGTTGAG

arcA_500dn	TGACTTTTATGGCGTTCTTTGTTTTTG
OmpTKO_AprR_F	AGATATAAAAAATACATATTCAATCATTAAAACGATTGAATGGAGAAC TTTTGGCTGACGCCGTTGGATAC
OmpTKO_AprR_R	TTTAAGGGTTAATTGTTACATTGAAATGGCTAGTTATTCCCCGGGGCGA TTCAGCCAATCGACTGGCGAG
OmpT_up	CCTCATGCTATTTTCGCTTATATGC
OmpT_dn	GATTATTATGGTGTCACGCCATCTC

Table 8: Synthetic DNA and Oligos used for plasmid construction in Chapter 2

Name	**Sequence**
pCDF-1b_ampl1	GTCATCGTGGCCGGATCTTG
pCDF-1b_ampl2	ATTAATGCAGCTGGCACGACAG
pCDF-MCS	CTGTCGTGCCAGCTGCATTAATCAGTCCAGTTACGCTGGAGTCCAGTC CAGTTACGCTGGAGTCTGAGGCTCGTCCTGAATGATATCAAGCTTGA ATTCGTTGACGAATTCTCTAGATATCGCTCAATACTGACCATTTAAAT CATACCTGACCGTCATCGTGGCCGGATCTTG
yibDp-GFPuv	TGAGGCTCGTCCTGAATGATATCAAGCTTGAATTCGTTGTGCGTAATT GTGCTGATCTCTTATATAGCTGCTCTCATTATCTCTCTACCCTGAAGTG ACTCTCTCACCTGTAAAAATAATATCTCACAGGCTTAATAGTTTCTTA ATACAAAGCCTGTAAAACGTCAGGATAACTTCTTGTAGGAGGATAAT CTATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATT CTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGT GGAGAGGGTGAAGGTGATGCTACATACGGAAAGCTTACCCTTAAATT TATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCAC TACTTTCTCTTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATG AAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACA GGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTG CTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAA AAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTC GAGTACAACTATAACTCACACAATGTATACATCACGGCAGACAAACA AAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATTGAAG ATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATT GGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAA TCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCT TCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCT CTACAAATAATGAGGATCCCCGGCTTATCGGTCAGTTTCACCTGATTT ACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGGGGCAGAAA GATGAATGACTGTCCACGACGCTATACCCAAAAGAAAGACGAATTCT CTAGATATCGCTCAATACTGA

yibDp-mCherry	GTCTGAGGCTCGTCCTGAATGATATCAAGCTTGAATTCGTTCGTGCGT AATTGTGCTGATCTCTTATATAGCTGCTCTCATTATCTCTCTACCCTGA AGTGACTCTCTCACCTGTAAAAATAATATCTCACAGGCTTAATAGTTT CTTAATACAAAGCCTGTAAAACGTCAGGATAACTTCTGTGTAGGAGG ATAATCTATGGTATCAAAAGGAGAGGAAGATAACATGGCTATTATCA AAGAATTTATGCGCTTCAAAGTTCACATGGAAGGTAGCGTGAACGGT CACGAGTTCGAGATTGAAGGTGAAGGTGAAGGGCGTCCGTACGAAG GTACACAGACCGCTAAACTGAAGGTGACGAAAGGTGGCCCTCTTCCA TTTGCGTGGGATATTCTTAGTCCGCAATTTATGTATGGATCTAAGGCG TATGTCAAGCACCCGGCTGACATCCCAGATTACTTAAAACTTAGCTTC CCAGAGGGATTCAAATGGGAGCGCGTTATGAATTTCGAGGACGGCGG TGTAGTGACCGTCACTCAGGATTCATCACTTCAAGATGGCGAATTTAT CTACAAGGTCAAGCTGCGTGGGACAAATTTTCCGTCGGATGGGCCTG TCATGCAGAAGAAGACAATGGGCTGGGAAGCGTCGTCAGAGCGTAT GTATCCAGAGGACGGAGCGTTAAAAGGGGAAATTAAGCAGCGCCTG AAGTTGAAGGATGGCGGGCATTATGACGCAGAGGTTAAAACCACTTA TAAAGCGAAAAAGCCAGTCCAATTGCCAGGAGCCTACAATGTCAATA TCAAATTAGATATCACAAGTCATAACGAGGATTACACGATCGTCGAA CAATATGAGCGCGCAGAAGGTCGCCATAGTACAGGAGGAATGGACG AACTGTACAAATAATGACTCGAGtctggtaaaactagcatTCGACCTAGCATAA CCCCGCGGGGCCTCTTCGGGGGTCTCGCGGGGTTTTTTGCTGAAAGA AGCTTCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGT TTTATCTGTTGTTTGTCGCTGCGGCCGGGTCAGGTATGATTTAAATGG TCAGTAACGGGTCTTGAGGGGTTTTTTGCAATGGGTTCATCCCGTGGG GACGAATTCTCTAGATATCGCTCAATACTGA

yibDp-matB	TGCCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGC CTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCAC ACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATACACAGCTAACA CCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTGGTTGCTGGAT AACGTGCGTAATTGTGCTGATCTCTTATATAGCTGCTCTCATTATCTCT CTACCCTGAAGTGACTCTCTCACCTGTAAAAATAATATCTCACAGGCT TAATAGTTTCTTAATACAAAGCCTGTAAAACGTCAGGATAACTTCTAT ATTCAGGGAGACCACAACGGTTTCCCTCTACAAATAATTTTGTTTAAC TTTGGAAAAAGGAGATATACCATGCATCATCATCATCATCACTCCAA CCATTTATTCGATGCGATGCGTGCCGCGGCGCCTGGGAATGCGCCGTT CATCCGCATCGACAATACACGCACCTGGACCTACGACGATGCCTTTG CCTTGAGTGGTCGCATTGCGTCAGCTATGGACGCACTCGGCATTCGCC CGGGAGACCGCGTGGCCGTGCAGGTGGAGAAATCTGCCGAAGCACT GATTCTGTATCTTGCGTGTCTGCGTAGCGGTGCCGTATATTTGCCACT GAACACTGCTTATACACTTGCGGAACTGGACTACTTTATTGGTGATGC CGAACCGCGCCTCGTTGTTGTAGCGTCATCCGCCCGTGCAGGTGTGG AAACCATTGCGAAACCGCGCGGGGCCATTGTAGAAACTCTGGATGCA GCTGGCAGCGGAAGCTTGCTGGACTTGGCGCGCGATGAGCCTGCTGA TTTCGTGGACGCTAGTCGCTCGGCGGACGATCTGGCGGCAATTCTTTA TACAAGTGGGACGACAGGGCGTTCTAAGGGGGCAATGCTGACACAC GGCAACCTGCTTTCTAACGCGCTTACATTGCGTGATTTCTGGCGTGTA ACCGCAGGCGATCGTCTGATTCATGCGTTACCGATTTTTCATACACAT GGCCTGTTTGTCGCTACTAATGTCACATTACTGGCCGGGGCCTCTATG TTTCTGTTAAGCAAATTTGATCCGGAAGAGATCCTGTCTTTGATGCCG CAGGCTACCATGCTGATGGGCGTACCGACCTTCTATGTTCGCTTGCTG CAATCACCGCGCCTGGATAAACAGGCAGTAGCGAATATTCGCCTGTT TATTAGTGGGTCCGCACCACTGCTGGCAGAGACACACACTGAATTTC AAGCGCGTACCGGCCATGCCATTCTGGAACGCTACGGAATGACCGAG ACCAACATGAACACCTCAAATCCGTATGAGGGTAAACGTATTGCGGG TACTGTGGGCTTTCCTCTCCCGGATGTCACTGTTCGTGTTACCGACCC GGCAACCGGTCTCGCCTTACCTCCGGAACAGACGGGAATGATCGAAA TTAAAGGTCCGAACGTGTTTAAGGGCTATTGGCGCATGCCCGAGAAG ACCGCTGCCGAATTCACCGCCGATGGTTTCTTTATCAGTGGTGATTTA GGTAAAATCGATCGCGATGGATACGTTCATATTGTGGGGCGCGGGAA AGATCTGGTTATTTCAGGAGGCTATAATATTTATCCGAAAGAAGTTG AGGGGGAAATTGACCAGATTGAAGGGGTGGTTGAATCAGCAGTGATC GGCGTTCCGCACCCGGATTTTGGTGAAGGTGTCACAGCGGTTGTGGTT CGCAAACCAGGGGCGGCTCTGGATGAAAAAGCGATCGTCTCTGCTCT GCAGGACCGTCTGGCTCGTTATAAACAACCGAAACGCATCATTTTTG CTGAAGATCTGCCGCGTAACACAATGGGTAAAGTCCAGAAAAACATC TTGCGCCAGCAGTATGCAGACTTATATACTCGTACGTAGTAAGACGA ATTCTCTAGATATCGCTCAATACTG
yibDp-mdlC-his1	CCTGACGTTTTACAGGCTTTGTATTAAGAAACTATTAAGCCTGTGAGA TATTATTTTTACAGGTGAGAGAGTCACTTCAGGGTAGAGAGATAATG AGAGCAGCTATATAAGAGATCAGCACAATTACGCACTTATTTGCCGA CTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAACT GATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCT GTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCC

	ATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACT GCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATC GCCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTA GCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCC GCCGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGC AAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGC AAGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGC TGGATAACGCCACGGAATGATGTCGTCGTGCACAACAATGGTGACTT CTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAAGCCGAAGTTTCC AAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCAAGCCTTACG GTCACCGTAACCAGCAAATCAATATCACTGTGTGGCTTCAGGCCGCC ATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTTCGA GATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATACTT CGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAAG CATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT TTAGAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGG GCGTGAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACA GATTCCGTGGATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAG GGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGT TCACATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGC TCAACCATGAATCTGACAGTACGGGCGAAACCCGACAGGACTTAAAG ATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCC GACCCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGA AGTGTGGCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTGT AGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAACTCCCCGTTCAG CCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAGTCCAACCCG GAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTAACTGGG AGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGAAGTG TGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTGC TCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTA ACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCA AAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCT TTTCTACTGAACCGCTCTGC
yibDp-mdlC-his2	AGCCTGTAAAACGTCAGGATAACTTCTATATTCAGGGAGACCACAAC GGTTTCCCTCTACAAATAATTTTGTTTAACTTTCGTGTGTAGGAGGAT AATCTATGTATACGGTGGGGGACTACTTGCTTGATCGCCTGCACGAGT TAGGCATCGAGGAAATTTTTGGTGTACCCGGGGACTATAACCTGCAG TTCCTTGATCAGATCATTTCACGTGAGGATATGAAATGGATTGGGAA CGCCAATGAACTTAACGCATCATATATGGCGGATGGATATGCTCGCA CAAAAAAAGCCGCGGCTTTTCTTACAACTTTCGGCGTGGGGGAGTTA AGTGCTATCAATGGATTGGCCGGCTCGTATGCTGAAAATCTGCCCGTT GTAGAAATCGTAGGTAGCCCAACCTCCAAGGTCCAGAACGACGGTAA ATTCGTCCACCACACTTTAGCAGATGGCGATTTTAAGCACTTCATGAA GATGCATGAACCGGTGACAGCTGCCCGCACTCTTTTGACCGCCGAGA ATGCGACTTATGAAATTGATCGTGTCTTAAGTCAACTGCTGAAGGAA CGTAAACCAGTTTACATTAACTTACCCGTCGATGTCGCGGCAGCTAA GGCAGAGAAACCAGCCTTGAGTTTGGAGAAGGAAAGTTCAACCACA AACACCACCGAGCAAGTTATTCTTTCAAAGATTGAAGAGTCCTTGAA GAACGCCCAAAAACCAGTCGTTATTGCCGGTCATGAGGTTATCAGTT TCGGGCTTGAGAAAACAGTCACGCAATTCGTCTCCGAGACCAAACTT

	CCAATCACAACGCTGAATTTCGGCAAGTCTGCGGTCGATGAATCATT ACCTTCGTTTTTGGGGATCTATAATGGTAAACTGAGTGAGATCTCTCT TAAAAACTTTGTCGAATCAGCCGATTTTATCTTAATGCTTGGCGTGAA ATTAACGGACTCATCTACTGGCGCTTTTACCCATCATTTGGATGAAAA TAAAATGATTTCTCTTAATATCGATGAGGGTATTATTTTCAATAAGGT GGTAGAGGATTTCGATTTTCGCGCAGTTGTCTCGTCATTATCAGAACT TAAAGGTATTGAGTACGAAGGACAATATATCGATAAACAGTACGAAG AGTTTATCCCGAGCAGCGCACCACTTTCTCAAGATCGCTTATGGCAAG CAGTGGAGAGCCTGACTCAGTCAAATGAAACTATTGTCGCTGAACAA GGAACGTCTTTTTTTGGTGCCTCTACTATCTTCCTTAAAAGCAACTCG CGTTTCATCGGCCAACCACTGTGGGGGTCAATCGGGTACACGTTCCCC GCTGCTCTTGGGTCTCAGATTGCCGACAAGGAGAGTCGCCATCTGTTA TTCATTGGTGACGGGTCCCTTCAACTGACTGTTCAGGAGTTAGGCCTG TCTATCCGCGAAAAATTGAATCCAATCTGTTTTATCATTAATAATGAC GGTTATACCGTGGAGCGCGAGATCCATGGGCCAACACAAAGCTACAA CGACATTCCCATGTGGAATTATTCGAAGCTTCCGGAAACTTTTGGAGC AACAGAGGACCGTGTTGTAAGCAAGATCGTTCGCACGGAGAATGAGT TCGTATCCGTTATGAAGGAAGCTCAAGCGGACGTTAACCGTATGTAT TGGATTGAACTGGTTTTGGAAAAAGAGGATGCCCCCAAGTTATTGAA AAAAATGGGAAAATTGTTCGCGGAGCAGAATAAGCATCATCACCACC ACCACTGATGTTGACCGCAAAAAACCCCGCTTCGGCGGGGTTTTTTCG CAGAGCGGTTCAGTAGAAA
yibDp-GST	TGAGGCTCGTCCTGAATGATATCAAGCTTGAATTCGTTTGCCCAGGCA TCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTA TCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCAC CTTCGGGTGGGCCTTTCTGCGTTTATACACAGCTAACACCACGTCGTC CCTATCTGCTGCCCTAGGTCTATGAGTGGTTGCTGGATAACGTGCGTA ATTGTGCTGATCTCTTATATAGCTGCTCTCATTATCTCTCTACCCTGAA GTGACTCTCTCACCTGTAAAAATAATATCTCACAGGCTTAATAGTTTC TTAATACAAAGCCTGTAAAACGTCAGGATAACTTCTATATTCAGGGA GACCACAACGGTTTCCCTCTACAAATAATTTTGTTTAACTTTAAAGAG GAGAAATACTAGATGGGGAACGCAGCATCTGCGCGCCGCATGTCCCC GATCCTGGGTTACTGGAAAATCAAAGGGTTAGTGCAGCCAACCCGTC TGTTATTAGAATACCTGGAGGAAAAATACGAGGAACACCTGTACGAG CGCGATGAAGGCGATAAATGGCGCAATAAAAAATTCGAACTCGGGCT GGAATTCCCAAACTTACCCTATTATATTGATGGAGATGTTAAATTGAC CCAGTCTATGGCAATCATTCGCTATATTGCAGATAAACATAACATGTT GGGCGGCTGTCCTAAGGAGCGCGCGGAAATTAGTATGCTGGAAGGCG CGGTGCTGGATATCCGCTATGGTGTTAGCCGCATTGCGTACTCGAAA GATTTTGAGACGCTCAAAGTTGATTTTCTGAGTAAACTGCCTGAAATG TTAAAGATGTTTGAAGATCGCTTGTGTCACAAAACGTATTTAAATGGT GATCATGTCACCCATCCAGACTTTATGCTGTATGATGCGCTTGATGTG GTTTTGTACATGGATCCGATGTGCCTGGATGCCTTTCCGAAGCTGGTC TGTTTCAAAAAACGCATCGAGGCTATTCCGCAAATCGACAAATATCT CAAATCTAGTAAATACATCGCGTGGCCTCTGCAGGGCTGGCAAGCGA CCTTTGGTGGGGGCGATCATCCGCCAAAATGATAAGACGAATTCTCT AGATATCGCTCAATACTGA
pHCKan-T7-CBD-hGLY	CATTGCATAATACGACTCACTATAGGGAGACCACAACGGTTTCCCAC TAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATAAAGAGG

	AGAAATACTAGATGACAAATCCTGGTGTAAGTGCCTGGCAAGTTAAT ACCGCATATACCGCTGGGCAGTTAGTCACTTATAACGGCAAGACCTA CAAGTGCTTGCAGCCTCACACATCCTTGGCAGGTTGGGAACCGTCCA ATGTACCCGCCCTTTGGCAACTTCAGGGCTCTGCCGGTAGTGCGGCG GGTTCCGGTGAATTTATGCTGGTTCTGCACAACAGCCAAAAACTGCA AATTCTGTACAAATCTCTGGAAAAATCTATTCCTGAAAGCATTAAAGT TTATGGTGCGATCTTCAACATCAAGGATAAAAACCCTTTCAATATGG AGGTGCTGGTCGACGCGTGGCCTGATTACCAAATCGTCATCACCCGT CCGCAGAAGCAAGAAATGAAAGATGACCAGGACCACTATACTAACA CCTACCACATCTTCACTAAAGCGCCGGACAAACTGGAAGAAGTTCTG AGCTACTCCAATGTTATCTCCTGGGAACAAACGCTGCAGATTCAAGG TTGCCAGGAAGGTCTGGATGAGGCGATTCGTAAGGTCGCTACCTCTA AGTCCGTTCAAGTCGATTACATGAAAACCATCCTGTTTATCCCGGAGC TGCCGAAGAAACACAAAACCTCCTCTAACGACAAGATGGAGCTGTTC GAAGTAGACGATGACAACAAAGAGGGTAACTTTTCCAATATGTTCCT GGACGCCTCCCACGCCGGTCTGGTTAATGAGCACTGGGCGTTCGGTA AAAACGAACGCTCCCTGAAGTACATTGAGCGTTGTCTGCAGGATTTT CTGGGTTTTGGTGTCCTGGGTCCAGAGGGTCAACTGGTCTCTTGGATC GTTATGGAACAGTCTTGTGAACTGCGTATGGGTTATACTGTGCCAAA ATACCGTCACCAAGGTAATATGCTGCAAATCGGTTATCATCTGGAGA AGTACCTGAGCCAGAAGGAAATCCCGTTCTATTTCCATGTTGCCGATA ATAACGAAAAGAGCCTGCAAGCCCTGAACAATCTGGGCTTCAAGATC TGCCCTTGCGGTTGGCACCAGTGGAAGTGCACGCCTAAAAAGTACTG TGGCGGTGGCCATCATCACCATCACCATTAATGA
SL1_rc	ACTCCAGCGTAACTGGACTG
SR2_rc	ACTGACCATTTAAATCATACCTGACC
GFP_cp6_F	AGCAGCCATCACCATC
GFP_cp6_R	GCCCATATGTATATCTCCTTCTTAAAG
pS-yibD-hGLY_F	CTGGAAAAAGGAGATATACCATGACAAATCCTGGTGTAAGTGCC
pS-yibD-hGLY_R	GTGAGTCGTATTAGAAGAGCTCATTAATGGTGATGGTGATGATGGC
pS-yibDp-FOR	TAATGAGCTCTTCTAATACGACTCACTATAGGG
pS-yibDp-REV	CATGGTATATCTCCTTTTTCCAGAAGTG

Media Stock Solutions

- 10X concentrated Ammonium-Citrate 30 salts (1 L) by mixing 30 g of $(NH_4)_2SO_4$ and 1.5 g Citric Acid in water with stirring, adjust pH to 7.5 with NaOH. Autoclave and store at room temperature (RT).

- 10X concentrated Ammonium-Citrate 90 salts (1 L) by mixing 90 g of $(NH_4)_2SO_4$ and 2.5 g Citric Acid in water with stirring, adjust pH to 7.5 with NaOH. Autoclave and store at RT.
- 3 M Ammonium sulfate solution in water. Autoclave and store at RT.
- 100 g/L citric acid in water. Autoclave and store at RT.
- 1 M Potassium 3-(N-morpholino) propanesulfonic Acid (MOPS), adjust to pH 7.4 with KOH. Filter sterilize (0.2 µm) and store at RT.
- 0.5 M potassium phosphate buffer, pH 6.8 by mixing 248.5 mL of 1.0 M K_2HPO_4 and 251.5 mL of 1.0 M KH_2PO_4 and adjust to a final volume of 1000 mL with ultrapure water. Filter sterilize (0.2 µm) and store at RT.
- 2 M $MgSO_4$ and 10 mM $CaSO_4$ solutions. Filter sterilize (0.2 µm) and store at RT.
- 50 g/L solution of thiamine-HCl. Filter sterilize (0.2 µm) and store at 4°C.
- 500 g/L solution of glucose, dissolving by stirring with mild heat. Cool, filter sterilize (0.2 µm), and store at RT.
- 100 g/L yeast extract, autoclave, and store at RT.
- 100 g/L casamino acid, autoclave, and store at RT.
- 500X Trace Metal Stock: Prepare a solution of micronutrients in 1000 mL of water containing 10 mL of concentrated H_2SO_4, 0.6 g $CoSO_4 \cdot 7H_2O$, 0.5 g $CuSO_4 \cdot 5H_2O$, 0.6 g $ZnSO_4 \cdot 7H_2O$, 0.2 g $Na_2MoO_4 \cdot 2H_2O$, 0.1 g H_3BO_3, and 0.3 g $MnSO_4 \cdot H_2O$. Filter sterilize (0.2 µm) and store at RT in the dark.

- Prepare a fresh solution of 40 mM ferric sulfate heptahydrate in water, filter sterilize (0.2 µm) before preparing media each time.

Media Formulations

Prepare the final working medium by aseptically mixing stock solutions based on the following tables in the order written to minimize precipitation, then filter sterilize (with a 0.2 µm filter).

Table 9: SM10+ Seed Media, pH 6.8:

Ingredient	Concentration Stock	Volume in 1 L (mL)	Final Concentration
Ammonium-Citrate 90 Salts, pH 7.5	10 X	100.0	1 X
Phosphate Buffer, pH 6.8	500 mM	10.0	5.00 mM
Trace Metals	500 X	4.0	2 X
Fe (II) Sulfate	40 mM	4.0	0.16 mM
$MgSO_4$	2 M	1.25	2.50 mM
$CaSO_4$	10 mM	6.25	0.0625 mM
Glucose	500 g/L	90.0	45 .0g/L
MOPS	1 M	200.0	200 mM
Thiamine-HCl	50 g/L	0.2	0.01 g/L
Yeast Extract	100 g/L	10.0	1.0 g/L
Casamino Acids	100 g/L	0	0

Table 10: FGM10 Media, pH 6.8

Ingredient	Concentration Stock	Volume in 1 L (mL)	Final Concentration
Ammonium-Citrate 90 Salts, pH 7.5	10 X	100.0	1 X
Phosphate Buffer, pH 6.8	500 mM	10.0	5.00 mM
Trace Metals	500 X	4.0	2 X
Fe (II) Sulfate	40 mM	4.0	0.16 mM
$MgSO_4$	2 M	1.25	2.50 mM
$CaSO_4$	10 mM	6.25	0.06 mM
Glucose	500 g/L	50.0	25 .0 g/L
Thiamine-HCl	50 g/L	0.2	0.01 g/L
Yeast Extract	100 g/L	0	0
Casamino Acids	100 g/L	0	0

Table 11: AB Autoinduction Broth (aka Awesome Broth):

Ingredient	Concentration Stock	Volume in 1 L (mL)	Final Concentration
Ammonium sulfate	3 M	13.6	40.8 mM
Citric acid	100 g/L	2.5	0.25 g/L
Trace Metals	500 X	5.6	2.8 X
Fe (II) Sulfate	40 mM	2.4	0.096 mM
$MgSO_4$	2 M	4.35	8.7 mM
$CaSO_4$	10 mM	7.08	0.0708 mM
Glucose	500 g/L	90.0	45 .0 g/L
MOPS	1 M	200.0	200 mM
Thiamine-HCl	50 g/L	0.2	0.01 g/L
Yeast Extract	100 g/L	62	6.2 g/L
Casamino Acids	100 g/L	35	3.5 g/L

Table 12: Autoinduction C7 Media

Ingredient	Concentration Stock	Volume in 1 L (mL)	Final Concentration
Ammonium sulfate	3 M	22.67	68 mM
Citric acid	100 g/L	2.5	0.25 g/L
Trace Metals	500 X	4	2 X
Fe (II) Sulfate	40 mM	4	0.160 mM
$MgSO_4$	2 M	5	10 mM
$CaSO_4$	10 mM	6.25	0.0625 mM
Glucose	500 g/L	90.0	45 .0g/L
MOPS	1 M	200.0	200 mM
Thiamine-HCl	50 g/L	0.2	0.01 g/L
Yeast Extract	100 g/L	25	2.5 g/L
Casamino Acids	100 g/L	25	2.5 g/L

Table 13 Nutrient Levels used in the media DoE experiment in Chapter 2

Levels	Citric Acid (g/L)	$(NH_4)_2SO_4$ (mM)	$FeSO_4$ (mM)	$MgSO_4$ (mM)	$CaSO_4$ (mM)	TM Mix (X)	Yeast Extract (g/L)	Casamino Acid (g/L)
1	0.0625	17	0.040	2.5	0.016	0.5	0.625	0
2	0.08	22.67	0.053	3.33	0.021	0.67	0.83	0.625
3	0.125	34	0.080	5	0.031	1	1.25	0.83
4	0.17	40.8	0.096	6.0	0.038	1.2	1.67	1.67
5	0.25	45.33	0.107	6.67	0.042	1.33	2.5	2.5
6	0.375	58.9	0.139	8.7	0.054	1.7	3.5	3.5
7	0.500	68	0.160	10	0.063	2	3.75	5.00
8	0.75	77.1	0.18	11.3	0.071	2.3	5.00	6.2
9	1	95.2	0.22	14.0	0.088	2.8	6.2	7.50
10	1.50	102	0.24	15	0.094	3	7.50	8.8
11		136.00	0.32	20.00	0.125	4.00	8.8	10
12		204.00	0.48	30.00	0.188	6.00	10	11.5
13		272	0.64	40	0.250	8	11.5	15.00

14		408.00	0.96	60.00	0.375	12.00	15.00	

Appendix B Chapter 3 Supplementary Information

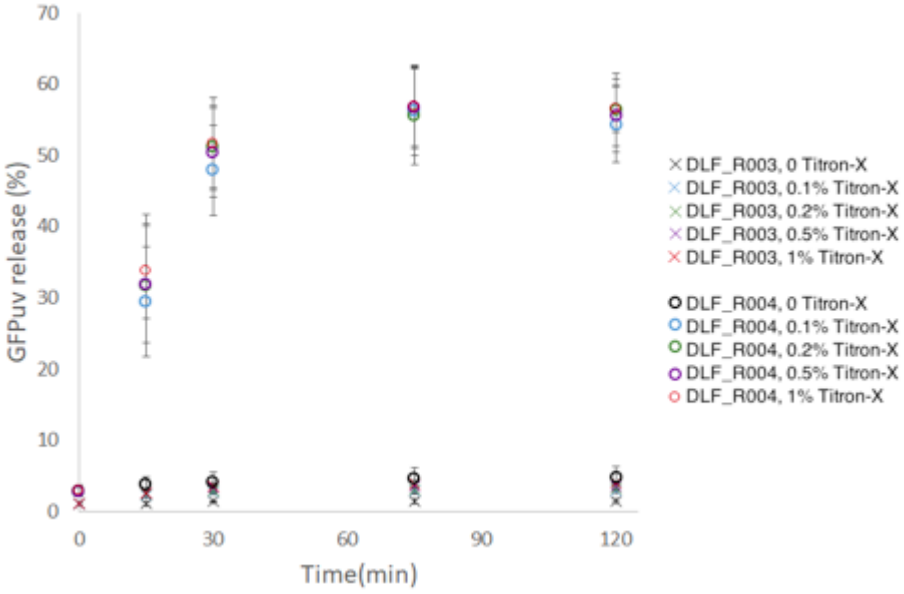

Figure 36: Impact of triton level on lysis and protein release. GFP release in strains DLF_R004 and DLF_R003 with the addition of triton X from 0.1% to 1%.

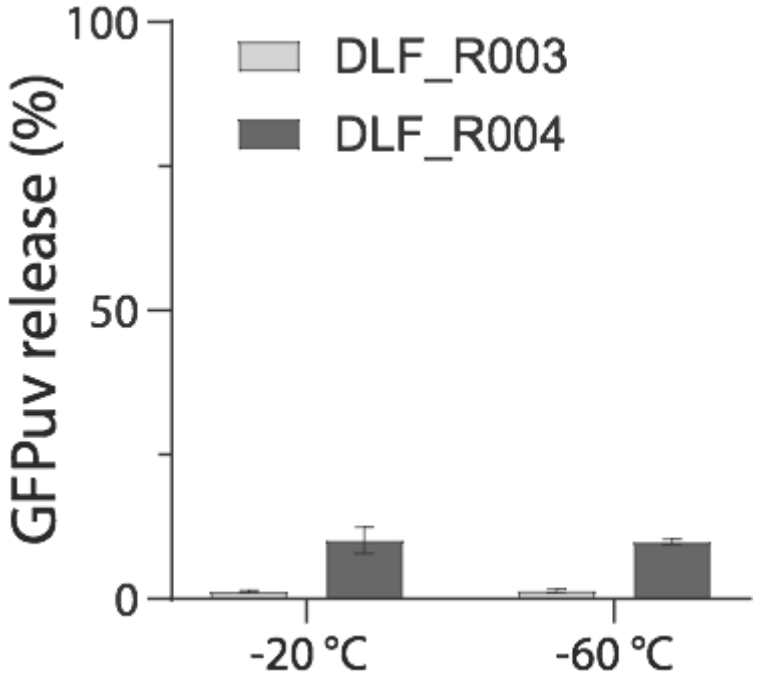

Figure 37: Impact of freeze thaw without triton X addition on protein release.

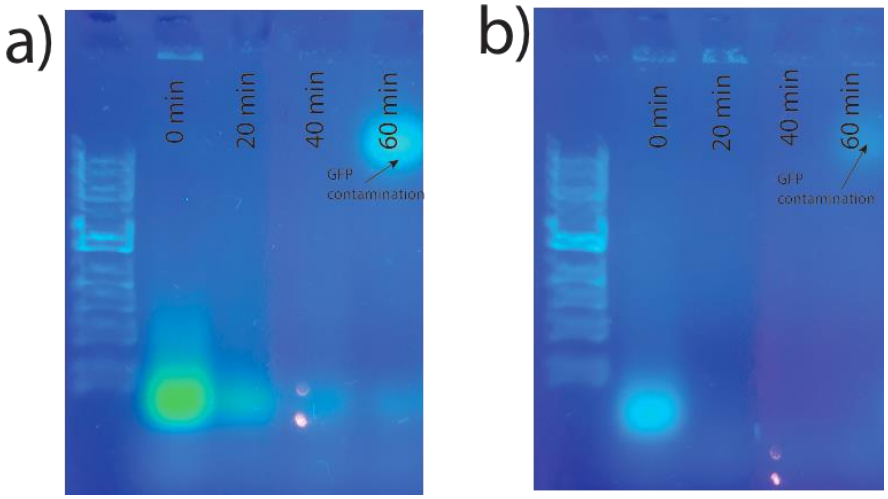

Figure 38: DNA hydrolysis in DLF_R004 + pHCKan-GFPuv. a) Agarose gel electrophoresis of heat denatured lysates with EDTA present from the beginning of lysis. a) Agarose gel electrophoresis of heat denatured lysates with active benzonase.

1) 5mL LB culture

↓ Overnight 37°C ,150 rpm

2) 200 μL into 20 mL of AB (250 mL Erlenmeyer)

↓ >24 hrs 37°C ,150 rpm

3) Harvest Cells / Centrifugation

↓ Overnight 37°C ,150 rpm

4) Resuspend in 2mL Buffer

↓ 20mM Tris, pH 8.0, 2mM $MgCl_2$ 0.1% Triton-X100

5) Freeze 30 minutes -20°C

↓

6) Incubate 60 min, 37°C

↓

7) Centrifugation, Harvest Supernatant (Lysate)

Figure 39: Autoinducible Lysis and Hydrolysis Shake Flask Protocol.

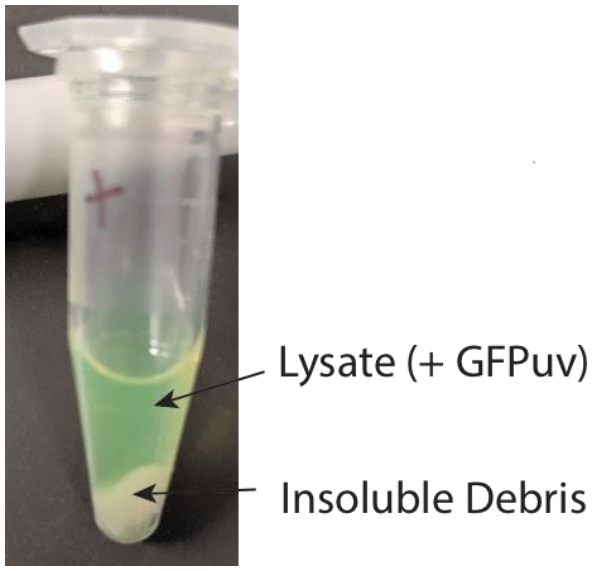

Figure 40: Sample of Cleared Lysate.

Table 14. DNA And Oligonucleotides used in this study

Name	Sequence
delete-ompT-Term-yibDp-Lys-nucA-apmR	AAGAAATTAAAAAACGAAAATACCAGCTATAGCCAGATT GTCACAGAGTGTCGTATGCGTTACGCCGTACAGATGTTATTGATG GATAACAAAAATATCACTCAGGTGGCGCAATTATGTGGCTATAGC AGCACGTCGTACTTTATCTCTGTTTTTAAGGCGTTTTACGGCCTGA CACCGTTGAATTATCTCGCCAAACAGCGACAAAAAGTGATGTGGt gaAGGGCAAAGCGGAAACGGATAAGACGGGCATAAATGAGGAAG AAATGGCTCGACCTAGCATAACCCCGCGGGGCCTCTTCGGGGGTC TCGCGGGGTTTTTTGCTGAAAGAAGCTTCAAATAAAACGAAAGGC TCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGCTG CGGCCGGGTCAGGTATGATTTAAATGGTCAGTAACGGGTCTTGAG GGGTTTTTTGCATATGTGCGTAATTGTGCTGATCTCTTATATAGCT GCTCTCATTATCTCTCTACCCTGAAGTGACTCTCTCACCTGTAAAA ATAATATCTCACAGGCTTAATAGTTTCTTAATACAAAGCCTGTAA AACGTCAGGATAACTTCTGTGTAGGAGGATAATCTTTTGCTGGAA AAAGGAGATATACCATGGTGGAGATCAATAACCAGCGTAAAGCA TTTTTGGATATGTTAGCGTGGTCTGAAGGGACTGACAATGGCCGT CAGAAAACACGTAACCATGGCTATGACGTGATCGTGGGCGGTGA ATTATTCACAGACTATTCGGACCATCCACGCAAATTAGTAACCCT TAATCCGAAGCTGAAGTCGACGGGCGCGGGCCGCTATCAACTGCT TAGCCGTTGGTGGGATGCTTACCGCAAGCAGTTAGGGCTGAAGGA CTTTTCGCCGAAGAGCCAGGATGCTGTCGCTCTGCAACAAATTAA AGAGCGTGGGGCTTTACCAATGATCGACCGTGGTGACATCCGTCA GGCCATTGATCGCTGCTCTAACATCTGGGCATCTCTTCCAGGTGC GGGCTACGGGCAGTTCGAGCACAAAGCCGACTCCCTTATTGCTAA GTTCAAAGAAGCGGGAGGCACGGTTCGCGAGATCGATGTATAAG GATCTAGGAGGGAGATCATATGCGTTTTAATAATAAAATGTTAGC TCTTGCCGCATTATTGTTTGCAGCCCAAGCGTCAGCGGATACGCTT GAATCTATCGATAACTGTGCCGTGGGTTGCCCGACCGGGGGTTCG TCTAATGTAAGCATCGTCCGTCACGCCTATACCCTTAACAATAATT CTACCACGAAATTCGCGAACTGGGTGGCCTACCATATTACCAAAG ATACTCCCGCATCGGGTAAAACACGCAACTGGAAGACTGACCCG GCACTGAATCCAGCCGATACATTGGCTCCGGCAGATTACACAGGT GCAAACGCTGCTCTGAAAGTAGACCGCGGACATCAGGCTCCGCTT GCGAGTCTGGCTGGAGTTAGTGACTGGGAATCACTGAACTATTTG TCGAATATTACCCCACAGAAATCTGACTTGAATCAAGGCGCGTGG GCTCGTTTAGAAGACCAAGAGCGTAAACTTATCGACCGTGCAGAC ATTTCCAGCGTCTATACAGTAACAGGTCCTCTGTACGAGCGCGAC ATGGGAAAGCTTCCTGGTACTCAGAAAGCTCATACAATCCCTTCT GCATATTGGAAGGTTATCTTTATTAATAATAGCCCGGCTGTCAAT CATTATGCCGCCTTCTTATTTGATCAGAACACTCCCAAGGGGGCT GACTTCTGCCAATTCCGCGTCACTGTAGACGAGATTGAGAAGCGC ACGGGGTTAATCATCTGGGCCGGATTACCTGACGATGTGCAGGCT TCCTTGAAATCCAAGCCGGGAGTTCTTCCCGAGTTGATGGGTTGC AAAAACTAATAAACCGTGTTGACAATTAATCATCGGCATAGTATA TCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCatgtcatcagc

	ggtggagtgcaatgtcgtgcaatacgaatggcgaaaagccgagctcatcggtcagcttctcaaccttggggtt acccccggcggtgtgctgctggtccacagctccttccgtagcgtccggcccctcgaagatgggccacttgga ctgatcgaggccctgcgtgctgcgctgggtccgggagggacgctcgtcatgccctcgtggtcaggtctgga cgacgagccgttcgatcctgccacgtcgcccgttacaccggaccttggagttgtctctgacacattctggcgc ctgccaaatgtaaagcgcagcgcccatccatttgcctttgcggcagcggggccacaggcagagcagatcat ctctgatccattgcccctgccacctcactcgcctgcaagcccggtcgcccgtgtccatgaactcgatgggcag gtacttctcctcggcgtgggacacgatgccaacacgacgctgcatcttgccgagttgatggcaaaggttccct atggggtgccgagacactgcaccattcttcaggatggcaagttggtacgcgtcgattatctcgagaatgacca ctgctgtgagcgctttgccttggcggacaggtggctcaaggagaagagccttcagaaggaaggtccagtcg gtcatgcctttgctcggttgatccgctcccgcgacattgtggcgacagccctgggtcaactgggccgagatcc gttgatcttcctgcatccgccagaggcgggatgcgaagaatgcgatgccgctcgccagtcgattggctgagc tcaGAACGCCAACTAAAATTTCCCCGAGGTGAAAATCGCCCCGGGG AATAACTAGCCATTTCAATGTAACAATTAACCCTTAAAATAAACC CAGAAGGTTATTAACTAAATCACATAGAAAACCATCAATTATAGT ATGTATAAAATAGGCGACAGCAACCCAATTACAAATTAATGGTTC CAGAATATCACATCAAAAAAAACGCTGTATAATATTATAATTAAC ATGTAGACAACTTGTAATAAACATTATCAGTCAATTGTTTTGTTTA TTCCATCTGTGACGCCGATTATTTTCTCAAAATAATGAGATGGCGT GACACCATAATAATCTTTAAATGCACATATGAAATATGAAGTACT GTTATAGCCACATTTCTGGGCTACGACATTGATAGAATAAGAGTT TGAAGTTATGAGTTTTTTTGCATACCTCATCCTAGTATCTCTCAAT ATTTCAGTAAATGACGTTCCTTCATCCCTTAATCTTTTTTTTATTAA AC
OmpT_1000bp_up	GCATTGCTTTTTACCGTATTGTCTAAC
DLF_R004_seq_conf_2	ATGGTGGAGATCAATAACCAGCGTAAAGC
DLF_R004_seq_conf_3	TAGCTCTTGCCGCATTATTGTTTGC
DLF_R004_seq_conf_4	CCTTCTTATTTGATCAGAACACTC
OmpT_500bp_dn	GATTATTATGGTGTCACGCCATCTC

References

Abe, H., Goto, M., & Kamiya, N. (2011). Protein Lipidation Catalyzed by Microbial Transglutaminase. *Chemistry - A European Journal*, *17*(50), 14004–14008.

Adabi, E., Saebi, F., Moradi Hasan-Abad, A., Teimoori-Toolabi, L., & Kardar, G. A. (2017). Evaluation of an Albumin-Binding Domain Protein Fused to Recombinant Human IL-2 and Its Effects on the Bioactivity and Serum Half-Life of the Cytokine. *Iranian Biomedical Journal*, *21*(2), 77–83.

Ade, N., Koirala, Y., & Mannan, M. S. (2019). Towards an inherently safer bioprocessing industry: A review. *Journal of Loss Prevention in the Process Industries*, *60*, 125–132.

Aertsen, A., & Michiels, C. W. (2005). SulA-dependent hypersensitivity to high pressure and hyperfilamentation after high-pressure treatment of Escherichia coli lon mutants. *Research in Microbiology*, *156*(2), 233–237.

Aharoni, A., Griffiths, A. D., & Tawfik, D. S. (2005). High-throughput screens and selections of enzyme-encoding genes. *Current Opinion in Chemical Biology*, *9*(2), 210–216.

Alexaki, A., Hettiarachchi, G. K., Athey, J. C., Katneni, U. K., Simhadri, V., Hamasaki-Katagiri, N., Nanavaty, P., Lin, B., Takeda, K., Freedberg, D., Monroe, D., McGill, J. R., Peters, R., Kames, J. M., Holcomb, D. D., Hunt, R. C., Sauna, Z. E., Gelinas, A., Janjic, N., … Kimchi-Sarfaty, C. (2019). Effects of codon optimization on coagulation factor IX translation and structure: Implications for protein and gene therapies. In *Scientific Reports* (Vol. 9, Issue 1). https://doi.org/10.1038/s41598-019-51984-2

Alves, C. S., & Dobrowsky, T. M. (2017). Strategies and Considerations for Improving Expression of "Difficult to Express" Proteins in CHO Cells. In P. Meleady (Ed.), *Heterologous Protein Production in CHO Cells: Methods and Protocols* (pp. 1–23). Springer New York.

Ambery, P. D., Klammt, S., Posch, M. G., Petrone, M., Pu, W., Rondinone, C., Jermutus, L., & Hirshberg, B. (2018). MEDI0382, a GLP-1/glucagon receptor dual agonist, meets safety and tolerability endpoints in a single-dose, healthy-subject, randomized, Phase 1 study. *British Journal of Clinical Pharmacology*, *84*(10), 2325–2335.

Andersen, J. T., Dalhus, B., Viuff, D., Ravn, B. T., Gunnarsen, K. S., Plumridge, A., Bunting, K., Antunes, F., Williamson, R., Athwal, S., Allan, E., Evans, L., Bjoras, M., Kjaerulff, S., Sleep, D., Sandlie, I., & Cameron, J. (2014). Extending serum half-life of albumin by engineering neonatal Fc receptor (FcRn) binding. *The Journal of Biological Chemistry*, *289*(19), 13492–13502.

Andreessen, B., Taylor, N., & Steinbüchel, A. (2014). Poly(3-hydroxypropionate): a promising alternative to fossil fuel-based materials. *Applied and Environmental Microbiology*, *80*(21), 6574–6582.

Andrews, L., Ralston, S., Blomme, E., & Barnhart, K. (2015). A snapshot of biologic drug development: Challenges and opportunities. *Human & Experimental Toxicology*, *34*(12), 1279–1285.

Anilionyte, O., Liang, H., Ma, X., Yang, L., & Zhou, K. (2018). Short, auto-inducible promoters for well-controlled protein expression in Escherichia coli. *Applied Microbiology and Biotechnology*, *102*(16), 7007–7015.

An, J. H., & Kim, Y. S. (1998). A gene cluster encoding malonyl-CoA decarboxylase (MatA), malonyl-CoA synthetase (MatB) and a putative dicarboxylate carrier protein (MatC) in Rhizobium trifolii--cloning, sequencing, and expression of the enzymes in Escherichia coli. *European Journal of Biochemistry / FEBS*, *257*(2), 395–402.

Antoniou, A. N., Blackwood, S. L., Mazzeo, D., & Watts, C. (2000). Control of antigen presentation by a single protease cleavage site. *Immunity*, *12*(4), 391–398.

Armstrong, J. K., Hempel, G., Koling, S., Chan, L. S., Fisher, T., Meiselman, H. J., & Garratty, G. (2007). Antibody against poly(ethylene glycol) adversely affects PEG-asparaginase therapy in acute lymphoblastic leukemia patients. *Cancer*, *110*(1), 103–111.

Aroda, V. R., Rosenstock, J., Terauchi, Y., Jeppesen, O. L. E., Christiansen, E., Hertz, C. L., & Haluzik, M. (2018). Effect and Safety of Oral Semaglutide Monotherapy in Type 2 Diabetes—PIONEER 1 Trial. *Diabetes*, *67*(Supplement 1). http://diabetes.diabetesjournals.org/content/67/Supplement_1/2-LB.abstract

Asada, H., Douen, T., Mizokoshi, Y., Fujita, T., Murakami, M., Yamamoto, A., & Muranishi, S. (1994). Stability of acyl derivatives of insulin in the small intestine: relative importance of insulin association characteristics in aqueous solution. *Pharmaceutical Research*, *11*(8), 1115–1120.

Asenjo, J. A. (1990). Selection of operations in separation processes. *Bioprocess Technology*, *9*, 3–16.

Atsumi, S., & Liao, J. C. (2008). Directed evolution of Methanococcus jannaschii citramalate synthase for biosynthesis of 1-propanol and 1-butanol by Escherichia coli. *Applied and Environmental Microbiology*, *74*(24), 7802–7808.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*(1), 2006.0008.

Baek, J. H., & Lee, S. Y. (2006). Novel gene members in the Pho regulon of Escherichia coli. *FEMS Microbiology Letters*, *264*(1), 104–109.

Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A. I., Saini, K. S., Baeshen, N. A., & Redwan, E. M. (2015). Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives. *Journal of Microbiology and Biotechnology*, *25*(7), 953–962.

Baez, A., Majdalani, N., & Shiloach, J. (2014). Production of recombinant protein by a novel oxygen-induced system in Escherichia coli. *Microbial Cell Factories*, *13*(1), 50.

Baker, M. P., Reynolds, H. M., Lumicisi, B., & Bryson, C. J. (2010). Immunogenicity of protein therapeutics: The key causes, consequences and challenges. *Self/nonself*, *1*(4), 314–322.

Ball, T. K., Saurugger, P. N., & Benedik, M. J. (1987). The extracellular nuclease gene of Serratia marcescens and its secretion from Escherichia coli. *Gene*, *57*(2-3), 183–192.

Balzer, S., Kucharova, V., Megerle, J., Lale, R., Brautaset, T., & Valla, S. (2013). A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in Escherichia coli. *Microbial Cell Factories*, *12*, 26.

Bartley, P. C., Bogoev, M., Larsen, J., & Philotheou, A. (2008). Long-term efficacy and safety of insulin detemir compared to Neutral Protamine Hagedorn insulin in patients with Type 1 diabetes using a treat-to-target basal-bolus regimen with insulin aspart at meals: a 2-year, randomized, controlled trial. *Diabetic Medicine: A Journal of the British Diabetic Association*, *25*(4), 442–449.

Battelino, T., Rasmussen, M. H., De Schepper, J., Zuckerman-Levin, N., Gucev, Z., & Savendahl, L. (2017). Somapacitan, a once-weekly reversible albumin-binding GH derivative, in children with GH deficiency: A randomized dose-escalation trial. *Clinical Endocrinology*, *87*(4), 350–358.

Bechara, C., & Sagan, S. (2013). Cell-penetrating peptides: 20 years later, where do we stand? *FEBS Letters*, *587*(12), 1693–1702.

Bech, E. M., Pedersen, S. L., & Jensen, K. J. (2018). Chemical Strategies for Half-Life Extension of Biopharmaceuticals: Lipidation and Its Alternatives. *ACS Medicinal Chemistry Letters*, *9*(7), 577–580.

Beekman, N. J., Schaaper, W. M., Tesser, G. I., Dalsgaard, K., Kamstrup, S., Langeveld, J. P., Boshuizen, R. S., & Meloen, R. H. (1997). Synthetic peptide vaccines: palmitoylation of peptide antigens by a thioester bond increases immunogenicity. *The Journal of Peptide Research: Official Journal of the American Peptide Society*, *50*(5), 357–364.

Behloul, N., Wei, W., Baha, S., Liu, Z., Wen, J., & Meng, J. (2017). Effects of mRNA secondary structure on the expression of HEV ORF2 proteins in Escherichia coli. *Microbial Cell Factories*, *16*(1), 200.

Bellmann-Sickert, K., Elling, C. E., Madsen, A. N., Little, P. B., Lundgren, K., Gerlach, L. O., Bergmann, R., Holst, B., Schwartz, T. W., & Beck-Sickinger, A. G. (2011). Long-acting lipidated analogue of human pancreatic polypeptide is slowly released into circulation. *Journal of Medicinal Chemistry*, *54*(8), 2658–2667.

Benedik, M. J., & Strych, U. (1998). Serratia marcescens and its extracellular nuclease. *FEMS Microbiology Letters*, *165*(1), 1–13.

BenMohamed, L., Thomas, A., & Druilhe, P. (2004). Long-Term Multiepitopic Cytotoxic-T-Lymphocyte Responses Induced in Chimpanzees by Combinations of Plasmodium falciparum Liver-Stage Peptides and Lipopeptides. In *Infect Immun* (Vol. 72, pp. 4376–4384).

Ben, R., Jiying, F., Jian'an, S., Tu, T. N., Jing, S., Jingsong, Z., Qiuyi, & Yaling, S. (2016). An auto-inducible expression system based on the RhlI-RhlR quorum-sensing regulon for recombinant protein production in E. coli. *Biotechnology and Bioprocess Engineering: BBE*, *21*(1), 160–168.

Bershtein, S., Mu, W., Serohijos, A. W. R., Zhou, J., & Shakhnovich, E. I. (2013). Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Molecular Cell*, *49*(1), 133–144.

Bhattacharya, A. A., Grune, T., & Curry, S. (2000). Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. *Journal of Molecular Biology*, *303*(5), 721–732.

Biedermann, K., Jepsen, P. K., Riise, E., & Svendsen, I. (1989). Purification and characterization of a Serratia marcescens nuclease produced by Escherichia coli. *Carlsberg Research Communications*, *54*(1), 17–27.

Bill, R. M. (2014). Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. In *Frontiers in Microbiology* (Vol. 5). https://doi.org/10.3389/fmicb.2014.00085

Bittner, L.-M., Westphal, K., & Narberhaus, F. (2015). Conditional Proteolysis of the Membrane Protein YfgM by the FtsH Protease Depends on a Novel N-terminal Degron. *The Journal of Biological Chemistry*, *290*(31), 19367–19378.

Bjerregaard, S., Nielsen, F. S., & Sauerberg, P. (2012). Solid compositions comprising a glp-1 agonist and a salt of n-(8-(2-hydroxybenzoyl) amino) caprylic acid. In *Patent*. https://patents.google.com/patent/WO2012080471A1/en%20US4325121.pdf

Blachnio-Zabielska, A. U., Koutsari, C., & Jensen, M. D. (2011). Measuring long-chain acyl-coenzyme A concentrations and enrichment using liquid chromatography/tandem mass spectrometry with selected reaction monitoring. *Rapid Communications in Mass Spectrometry: RCM*, *25*(15), 2223–2230.

Blazeck, J., & Alper, H. S. (2013). Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnology Journal*, *8*(1), 46–58.

Bode, S. A., Thévenin, M., Bechara, C., Sagan, S., Bregant, S., Lavielle, S., Chassaing, G., & Burlina, F. (2012). Self-assembling mini cell-penetrating peptides enter by both direct translocation and glycosaminoglycan-dependent endocytosis. *Chemical Communications* , *57*. https://doi.org/10.1039/C2CC33240J

Boël, G., Letso, R., Neely, H., Price, W. N., Wong, K.-H., Su, M., Luff, J., Valecha, M., Everett, J. K., Acton, T. B., Xiao, R., Montelione, G. T., Aalberts, D. P., & Hunt, J. F. (2016). Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature*, *529*(7586), 358–363.

Borrero-de Acuña, J. M., Hidalgo-Dumont, C., Pacheco, N., Cabrera, A., & Poblete-Castro, I. (2017). A novel programmable lysozyme-based lysis system in Pseudomonas putida for biopolymer production. *Scientific Reports*, *7*(1), 4373.

Briand, L., Marcion, G., Kriznik, A., Heydel, J. M., Artur, Y., Garrido, C., Seigneuric, R., & Neiers, F. (2016). A self-inducible heterologous protein expression system in Escherichia coli. *Scientific Reports*, *6*, 33037.

Buchackert, Y., Rummel, S., Vohwinkel, C. U., Gabrielli, N. M., Grzesik, B. A., Mayer, K., Herold, S., Morty, R. E., Seeger, W., & Vadasz, I. (2012). Megalin mediates transepithelial albumin clearance from the alveolar space of intact rabbit lungs. *The Journal of Physiology*, *590*(20), 5167–5181.

Bueno, C., Lee, K. K., Chau, L. A., Lee-Chan, E., Singh, B., Strejan, G. H., & Madrenas, J. (2004). Mechanism of modulation of T cell responses by N-palmitoylated peptides. *European Journal of Immunology*, *34*(12), 3497–3507.

Burg, J. M., Cooper, C. B., Ye, Z., Reed, B. R., & Moreb, E. A. (2016). Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations. *Current Opinion in*. https://www.sciencedirect.com/science/article/pii/S2211339816300703

Burg, J. M., Cooper, C. B., Ye, Z. X., Reed, B. R., Moreb, E. A., & Lynch, M. D. (2016). Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations. *Current Opinion in Chemical Engineering*, *14*, 121–136.

Buse, J. B., Garber, A., Rosenstock, J., Schmidt, W. E., Brett, J. H., Videbaek, N., Holst, J., & Nauck, M. (2011). Liraglutide treatment is associated with a low frequency and magnitude of antibody formation with no apparent impact on glycemic response or increased frequency of adverse events: results from the Liraglutide Effect and Action in Diabetes (LEAD) trials. *The Journal of Clinical Endocrinology and Metabolism*, *96*(6), 1695–1702.

Cai, Z., Xu, W., Xue, R., & Lin, Z. (2008). Facile, reagentless and in situ release of Escherichia coli intracellular enzymes by heat-inducible autolytic vector for high-throughput screening. *Protein Engineering, Design & Selection: PEDS*, *21*(11), 681–687.

Cambray, G., Guimaraes, J. C., & Arkin, A. P. (2018). Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. *Nature Biotechnology*, *36*(10), 1005–1015.

Cárcel-Márquez, J., Flores, A., Martín-Cabello, G., Santero, E., & Camacho, E. M. (2019). Development of an inducible lytic system for functional metagenomic screening. *Scientific Reports*, *9*(1), 3887.

Carlson, E. D., Gan, R., Hodgman, C. E., & Jewett, M. C. (2012). Cell-free protein synthesis: applications come of age. *Biotechnology Advances*, *30*(5), 1185–1194.

Ceroni, F., Boo, A., Furini, S., Gorochowski, T. E., Borkowski, O., Ladak, Y. N., Awan, A. R., Gilbert, C., Stan, G.-B., & Ellis, T. (2018). Burden-driven feedback control of gene expression. *Nature Methods*, *15*(5), 387–393.

Chaudhury, C., Mehnaz, S., Robinson, J. M., Hayton, W. L., Pearl, D. K., Roopenian, D. C., & Anderson, C. L. (2003). The major histocompatibility complex-related Fc receptor for IgG

(FcRn) binds albumin and prolongs its lifespan. *The Journal of Experimental Medicine*, *197*(3), 315–322.

Chekhonin, V. P., Kabanov, A. V., Zhirkov, Y. A., & Morozov, G. V. (1991). Fatty acid acylated Fab-fragments of antibodies to neurospecific proteins as carriers for neuroleptic targeted delivery in brain. *FEBS Letters*, *287*(1-2), 149–152.

Chew, F. N., Tan, W. S., Boo, H. C., & Tey, B. T. (2012). Statistical optimization of green fluorescent protein production from Escherichia coli BL21(DE3). *Preparative Biochemistry & Biotechnology*, *42*(6), 535–550.

Chien, L.-J., & Lee, C.-K. (2006). Synergistic effect of co-expressing d-amino acid oxidase with T7 lysozyme on self-disruption of Escherichia coli cell. *Biochemical Engineering Journal*, *28*(1), 17–22.

Choi, J. K., Ho, J., Curry, S., Qin, D., Bittman, R., & Hamilton, J. A. (2002). Interactions of very long-chain saturated fatty acids with serum albumin. *Journal of Lipid Research*, *43*(7), 1000–1010.

Choi, Y. L., Park, E. J., Kim, E., Na, D. H., & Shin, Y. H. (2014). Dermal Stability and In Vitro Skin Permeation of Collagen Pentapeptides (KTTKS and palmitoyl-KTTKS). In *Biomol Ther (Seoul)* (Vol. 22, pp. 321–327).

Cho, Y., Zhang, X., Pobre, K. F. R., Liu, Y., Powers, D. L., Kelly, J. W., Gierasch, L. M., & Powers, E. T. (2015). Individual and collective contributions of chaperoning and degradation to protein homeostasis in E. coli. *Cell Reports*, *11*(2), 321–333.

Chubukov, V., & Sauer, U. (2014). Environmental dependence of stationary-phase metabolism in Bacillus subtilis and Escherichia coli. *Applied and Environmental Microbiology*, *80*(9), 2901–2909.

Cloake, N. C., Beaino, W., Trifilieff, E., & Greer, J. M. (2014). Thiopalmitoylation of altered peptide ligands enhances their protective effects in an animal model of multiple sclerosis. *Journal of Immunology* , *192*(5), 2244–2251.

Connor, A., Borregaard, J., Buckley, S. T., Donsmark, M., Hartoft-Nielsen, M. L., Maarbjerg, S. J., Sondergaard, F. L., Vegge, A., Knudsen, L. B., & Baekdal, T. A. (2017). Site of Absorption of an Oral Formulation of Semaglutide. *Diabetes*, *66*(Supplement 1).

Contesini, F. J., Melo, R. R. de, & Sato, H. H. (2018). An overview of Bacillus proteases: from production to application. *Critical Reviews in Biotechnology*, *38*(3), 321–334.

Cooke, G. D., Cranenburgh, R. M., Hanak, J. A. J., & Ward, J. M. (2003). A modified Escherichia coli protein production strain expressing staphylococcal nuclease, capable of auto-hydrolysing host nucleic acid. *Journal of Biotechnology*, *101*(3), 229–239.

Coskun, T., Sloop, K. W., Loghin, C., Alsina-Fernandez, J., Urva, S., Bokvist, K. B., Cui, X., Briere, D. A., Cabrera, O., Roell, W. C., Kuchibhotla, U., Moyers, J. S., Benson, C. T., Gimeno, R. E., D'Alessio, D. A., & Haupt, A. (2018). LY3298176, a novel dual GIP and GLP-1 receptor agonist

for the treatment of type 2 diabetes mellitus: From discovery to clinical proof of concept. *Molecular Metabolism*, *18*, 3–14.

Cos, O., Ramón, R., Montesinos, J. L., & Valero, F. (2006). Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris under different promoters: A review. *Microbial Cell Factories*, *5*(1), 17.

Crameri, A., Whitehorn, E. A., Tate, E., & Stemmer, W. P. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology*, *14*(3), 315–319.

Cruikshank, W. W., Doctrow, S. R., Falvo, M. S., Huffman, K., Maciaszek, J., Viglianti, G., Raina, J., Kornfeld, H., & Malfroy, B. (1997). A lipidated anti-Tat antibody enters living cells and blocks HIV-1 viral replication. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology: Official Publication of the International Retrovirology Association*, *14*(3), 193–203.

Cui, W., Han, L., Suo, F., Liu, Z., Zhou, L., & Zhou, Z. (2018). Exploitation of Bacillus subtilis as a robust workhorse for production of heterologous proteins and beyond. *World Journal of Microbiology & Biotechnology*, *34*(10), 145.

Cull, M., & McHenry, C. S. (1990). Preparation of extracts from prokaryotes. *Methods in Enzymology*, *182*, 147–153.

Cutfield, W. S., Derraik, J. G., Gunn, A. J., Reid, K., Delany, T., Robinson, E., & Hofman, P. L. (2011). Non-compliance with growth hormone treatment in children is common and impairs linear growth. *PloS One*, *6*(1), e16223.

Dahl, K., Schaeffer, L., & Kruse, T. (2013). Amylin analogues and pharmaceutical compositions thereof (USPTO Patent No. 8575091). In *US Patent* (No. 8575091). https://patentimages.storage.googleapis.com/8b/68/18/d54785c0f7218f/US8575091.pdf

Danne, T., Lupke, K., Walte, K., Von Schuetz, W., & Gall, M. A. (2003). Insulin detemir is characterized by a consistent pharmacokinetic profile across age-groups in children, adolescents, and adults with type 1 diabetes. *Diabetes Care*, *26*(11), 3087–3092.

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(12), 6640–6645.

Davies, M., Pieber, T. R., Hartoft-Nielsen, M. L., Hansen, O. K. H., Jabbour, S., & Rosenstock, J. (2017). Effect of Oral Semaglutide Compared With Placebo and Subcutaneous Semaglutide on Glycemic Control in Patients With Type 2 Diabetes: A Randomized Clinical Trial. *JAMA: The Journal of the American Medical Association*, *318*(15), 1460–1470.

Dea, M. K., Hamilton-Wessler, M., Ader, M., Moore, D., Schaffer, L., Loftager, M., Volund, A., & Bergman, R. N. (2002). Albumin binding of acylated insulin (NN304) does not deter action to stimulate glucose uptake. *Diabetes*, *51*(3), 762–769.

De Cotiis, D. A., Woll, M. P., Fox, T. E., Hill, R. B., Levenson, R., & Flanagan, J. M. (2008). Optimized expression and purification of myristoylated human neuronal calcium sensor 1 in E. coli. *Protein Expression and Purification*, *61*(2), 103–112.

Deliyannis, G., Jackson, D. C., Ede, N. J., Zeng, W., Hourdakis, I., Sakabetis, E., & Brown, L. E. (2002). Induction of long-term memory CD8(+) T cells for recall of viral clearing responses against influenza virus. *Journal of Virology*, *76*(9), 4212–4221.

De Mey, M., De Maeseneire, S., Soetaert, W., & Vandamme, E. (2007). Minimizing acetate formation in E. coli fermentations. *Journal of Industrial Microbiology & Biotechnology*, *34*(11), 689–700.

D'Errico, G., D'Ursi, A. M., & Marsh, D. (2008). Interaction of a peptide derived from glycoprotein gp36 of feline immunodeficiency virus and its lipoylated analogue with phospholipid membranes. *Biochemistry*, *47*(19), 5317–5327.

Dhawan, M. D., Wise, F., & Baeumner, A. J. (2002). Development of a laser-induced cell lysis system. *Analytical and Bioanalytical Chemistry*, *374*(3), 421–426.

Di Carlo, D., Jeong, K.-H., & Lee, L. P. (2003). Reagentless mechanical cell lysis by nanoscale barbs in microchannels for sample preparation. *Lab on a Chip*, *3*(4), 287–291.

Didovyk, A., Tonooka, T., Tsimring, L., & Hasty, J. (2017). Rapid and Scalable Preparation of Bacterial Lysates for Cell-Free Gene Expression. *ACS Synthetic Biology*, *6*(12), 2198–2208.

Ding, N., Yang, C., Sun, S., Han, L., Ruan, Y., Guo, L., Hu, X., & Zhang, J. (2017). Increased glycosylation efficiency of recombinant proteins in Escherichia coli by auto-induction. *Biochemical and Biophysical Research Communications*, *485*(1), 138–143.

Dopazo, A., Tormo, A., Aldea, M., & Vicente, M. (1987). Structural inhibition and reactivation of Escherichia coli septation by elements of the SOS and TER pathways. *Journal of Bacteriology*, *169*(4), 1772–1776.

Doran, P. M. (1995). *Bioprocess Engineering Principles*. Elsevier.

Doyle, J. R., Krishnaji, S. T., Zhu, G., Xu, Z. Z., Heller, D., Ji, R. R., Levy, B. D., Kumar, K., & Kopin, A. S. (2014). Development of a membrane-anchored chemerin receptor agonist as a novel modulator of allergic airway inflammation and neuropathic pain. *The Journal of Biological Chemistry*, *289*(19), 13385–13396.

Ducore, J. M., Miguelino, M. G., & Powell, J. S. (2014). Alprolix (recombinant Factor IX Fc fusion protein): extended half-life product for the prophylaxis and treatment of hemophilia B. *Expert Review of Hematology*, *7*(5), 559–571.

Dumont, J. A., Bitonti, A. J., Clark, D., Evans, S., Pickford, M., & Newman, S. P. (2005). Delivery of an erythropoietin-Fc fusion protein by inhalation in humans through an immunoglobulin transport pathway. *Journal of Aerosol Medicine: The Official Journal of the International Society for Aerosols in Medicine*, *18*(3), 294–303.

Duronio, R. J., Jackson-Machelski, E., Heuckeroth, R. O., Olins, P. O., Devine, C. S., Yonemoto, W., Slice, L. W., Taylor, S. S., & Gordon, J. I. (1990). Protein N-myristoylation in Escherichia coli: reconstitution of a eukaryotic protein modification in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(4), 1506–1510.

Ehrlich, G. K., Michel, H., Truitt, T., Riboulet, W., Pop-Damkov, P., Goelzer, P., Hainzl, D., Qureshi, F., Lueckel, B., Danho, W., Conde-Knape, K., & Konkar, A. (2013). Preparation and characterization of albumin conjugates of a truncated peptide YY analogue for half-life extension. *Bioconjugate Chemistry*, *24*(12), 2015–2024.

Eisele, F., Kuhlmann, J., & Waldmann, H. (2002). Synthesis and membrane binding properties of a lipopeptide fragment from influenza virus a hemagglutinin. *Chemistry* , *8*(15), 3362–3376.

Eiteman, M. A., & Altman, E. (2006). Overcoming acetate in Escherichia coli recombinant protein fermentations. *Trends in Biotechnology*, *24*(11), 530–536.

Elbrond, B., Jakobsen, G., Larsen, S., Agerso, H., Jensen, L. B., Rolan, P., Sturis, J., Hatorp, V., & Zdravkovic, M. (2002). Pharmacokinetics, pharmacodynamics, safety, and tolerability of a single-dose of NN2211, a long-acting glucagon-like peptide 1 derivative, in healthy male subjects. *Diabetes Care*, *25*(8), 1398–1404.

Ellis, M., Patel, P., Edon, M., Ramage, W., Dickinson, R., & Humphreys, D. P. (2017). Development of a high yielding E. coli periplasmic expression system for the production of humanized Fab' fragments. *Biotechnology Progress*, *33*(1), 212–220.

Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E., & Villaverde, A. (2009). Microbial factories for recombinant pharmaceuticals. *Microbial Cell Factories*, *8*, 17.

Firnberg, E., & Ostermeier, M. (2012). PFunkel: efficient, expansive, user-defined mutagenesis. *PloS One*, *7*(12), e52031.

Flint, A., Nazzal, K., Jagielski, P., Hindsberger, C., & Zdravkovic, M. (2010). Influence of hepatic impairment on pharmacokinetics of the human GLP-1 analogue, liraglutide. In *British Journal of Clinical Pharmacology* (Vol. 70, Issue 6, pp. 807–814). https://doi.org/10.1111/j.1365-2125.2010.03762.x

Fosgerau, K., & Hoffmann, T. (2015). Peptide therapeutics: current status and future directions. *Drug Discovery Today*, *20*(1), 122–128.

Foster, D. (1992). Cell disruption: breaking up is hard to do. *Bio/technology* , *10*(12), 1539–1541.

Frumkin, I., Schirman, D., Rotman, A., Li, F., Zahavi, L., Mordret, E., Asraf, O., Wu, S., Levy, S. F., & Pilpel, Y. (2017). Gene Architectures that Minimize Cost of Gene Expression. *Molecular Cell*, *65*(1), 142–153.

Fujiwara, S., & Amisaki, T. (2013). Fatty acid binding to serum albumin: molecular simulation approaches. *Biochimica et Biophysica Acta*, *1830*(12), 5427–5434.

Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S., & Grayhack, E. J. (2016). Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. *Cell*, *166*(3), 679–690.

Garber, A. J. (2011). Long-acting glucagon-like peptide 1 receptor agonists: a review of their efficacy and tolerability. *Diabetes Care*, *34 Suppl 2*, S279–S284.

García-Fruitós, E., Martínez-Alonso, M., Gonzàlez-Montalbán, N., Valli, M., Mattanovich, D., & Villaverde, A. (2007). Divergent genetic control of protein solubility and conformational quality in Escherichia coli. *Journal of Molecular Biology*, *374*(1), 195–205.

Gautam, A., Nanda, J. S., Samuel, J. S., Kumari, M., Priyanka, P., Bedi, G., Nath, S. K., Mittal, G., Khatri, N., & Raghava, G. P. S. (2016). Topical Delivery of Protein and Peptide Using Novel Cell Penetrating Peptide IMT-P8. *Scientific Reports*, *6*, 26278.

Gawin, A., Peebo, K., Hans, S., Ertesvåg, H., Irla, M., Neubauer, P., & Brautaset, T. (2019). Construction and characterization of broad-host-range reporter plasmid suitable for on-line analysis of bacterial host responses related to recombinant protein production. In *Microbial Cell Factories* (Vol. 18, Issue 1). https://doi.org/10.1186/s12934-019-1128-7

Gefen, T., Vaya, J., Khatib, S., Harkevich, N., Artoul, F., Heller, E. D., Pitcovski, J., & Aizenshtein, E. (2013). The impact of PEGylation on protein immunogenicity. *International Immunopharmacology*, *15*(2), 254–259.

Giacalone, M. J., Gentile, A. M., Lovitt, B. T., Berkley, N. L., Gunderson, C. W., & Surber, M. W. (2006). Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system. *BioTechniques*, *40*(3), 355–364.

Gill, R. T., DeLisa, M. P., Shiloach, M., Holoman, T. R., & Bentley, W. E. (2000). OmpT expression and activity increase in response to recombinant chloramphenicol acetyltransferase overexpression and heat shock in E. coli. *Journal of Molecular Microbiology and Biotechnology*, *2*(3), 283–289.

Giugliani, R., Giugliani, L., de Oliveira Poswar, F., Donis, K. C., Corte, A. D., Schmidt, M., Boado, R. J., Nestrasil, I., Nguyen, C., Chen, S., & Pardridge, W. M. (2018). Neurocognitive and somatic stabilization in pediatric patients with severe Mucopolysaccharidosis Type I after 52 weeks of intravenous brain-penetrating insulin receptor antibody-iduronidase fusion protein (valanafusp alpha): an open label phase 1-2 trial. *Orphanet Journal of Rare Diseases*, *13*(1), 110.

Glazyrina, J., Materne, E.-M., Dreher, T., Storm, D., Junne, S., Adams, T., Greller, G., & Neubauer, P. (2010). High cell density cultivation and recombinant protein production with Escherichia coli in a rocking-motion-type bioreactor. *Microbial Cell Factories*, *9*, 42.

Gluck, J. M., Hoffmann, S., Koenig, B. W., & Willbold, D. (2010). Single vector system for efficient N-myristoylation of recombinant proteins in E. coli. *PloS One*, *5*(4), e10081.

Golor, G., Bensen-Kennedy, D., Haffner, S., Easton, R., Jung, K., Moises, T., Lawo, J. P., Joch, C., & Veldman, A. (2013). Safety and pharmacokinetics of a recombinant fusion protein linking

coagulation factor VIIa with albumin in healthy volunteers. *Journal of Thrombosis and Haemostasis: JTH*, *11*(11), 1977–1985.

Gomez-Escribano, J. P., & Bibb, M. J. (2012). Streptomyces coelicolor as an expression host for heterologous gene clusters. *Methods in Enzymology*, *517*, 279–300.

Granhall, C., Donsmark, M., Golor, G., Sondergaard, F. L., & Thomsen, M. (2017). Safety, Tolerability, and Pharmacokinetics of Multiple Once-Daily Dosing of Oral Semaglutide in Healthy Males and in Males with T2D. *Diabetes*, *66*(Supplement 1).

Grenier, F., Matteau, D., Baby, V., & Rodrigue, S. (2014). Complete Genome Sequence of Escherichia coli BW25113. *Genome Announcements*, *2*(5). https://doi.org/10.1128/genomeA.01038-14

Grünenfelder, B., Rummel, G., Vohradsky, J., Röder, D., Langen, H., & Jenal, U. (2001). Proteomic analysis of the bacterial cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(8), 4681–4686.

Guan, X., & Fierke, C. A. (2011). Understanding Protein Palmitoylation: Biological Significance and Enzymology. *Science China. Chemistry*, *54*(12), 1888–1897.

Guidotti, G., Brambilla, L., & Rossi, D. (2017). Cell-Penetrating Peptides: From Basic Research to Clinics. *Trends in Pharmacological Sciences*, *38*(4), 406–424.

Gunaratne, R., Kumar, S., Frederiksen, J. W., Stayrook, S., Lohrmann, J. L., Perry, K., Bompiani, K. M., Chabata, C. V., Thalji, N. K., Ho, M. D., Arepally, G., Camire, R. M., Krishnaswamy, S., & Sullenger, B. A. (2018). Combination of aptamer and drug for reversible anticoagulation in cardiopulmonary bypass. *Nature Biotechnology*, *36*(7), 606–613.

Gutiérrez-González, M., Farías, C., Tello, S., Pérez-Etcheverry, D., Romero, A., Zúñiga, R., Ribeiro, C. H., Lorenzo-Ferreiro, C., & Molina, M. C. (2019). Optimization of culture conditions for the expression of three different insoluble proteins in Escherichia coli. *Scientific Reports*, *9*(1), 16850.

Haahr, H., & Heise, T. (2014). A Review of the Pharmacological Properties of Insulin Degludec and Their Clinical Relevance. *Clinical Pharmacokinetics*, *53*(9), 787–800.

Hakem, I. F., Leech, A. M., Bohn, J., Walker, J. P., & Bockstaller, M. R. (2013). Analysis of heterogeneity in nonspecific PEGylation reactions of biomolecules. *Biopolymers*, *99*(7), 427–435.

Halfmann, G., & Lubitz, W. (1986). Differential induction of Escherichia coli autolysis by penicillin and the bacteriophage phi X174 gene E product. *Journal of Bacteriology*, *166*(2), 683–685.

Hamilton-Wessler, M., Ader, M., Dea, M., Moore, D., Jorgensen, P. N., Markussen, J., & Bergman, R. N. (1999). Mechanism of protracted metabolic effects of fatty acid acylated insulin, NN304, in dogs: retention of NN304 by albumin. *Diabetologia*, *42*(10), 1254–1263.

Han, F., Wang, Y., Sims, C. E., Bachman, M., Chang, R., Li, G. P., & Allbritton, N. L. (2003). Fast electrical lysis of cells for capillary electrophoresis. *Analytical Chemistry*, *75*(15), 3688–3696.

Han, M.-J., Jeong, K. J., Yoo, J.-S., & Lee, S. Y. (2003). Engineering Escherichia coli for increased productivity of serine-rich proteins based on proteome profiling. *Applied and Environmental Microbiology*, *69*(10), 5772–5781.

Harris, J. M., Martin, N. E., & Modi, M. (2001). Pegylation: a novel process for modifying pharmacokinetics. *Clinical Pharmacokinetics*, *40*(7), 539–551.

Harrison, R. (2019). *Protein Purification Process Engineering*. Routledge.

Hatahet, F., Blazyk, J. L., Martineau, E., Mandela, E., Zhao, Y., Campbell, R. E., Beckwith, J., & Boyd, D. (2015). Altered Escherichia coli membrane protein assembly machinery allows proper membrane assembly of eukaryotic protein vitamin K epoxide reductase. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(49), 15184–15189.

Hausser, J., Mayo, A., Keren, L., & Alon, U. (2019). Central dogma rates and the trade-off between precision and economy in gene expression. In *Nature Communications* (Vol. 10, Issue 1). https://doi.org/10.1038/s41467-018-07391-8

Havelund, S., Plum, A., Ribel, U., Jonassen, I., Volund, A., Markussen, J., & Kurtzhals, P. (2004). The mechanism of protraction of insulin detemir, a long-acting, acylated analog of human insulin. *Pharmaceutical Research*, *21*(8), 1498–1504.

Henderson, S. J., Konkar, A., Hornigold, D. C., Trevaskis, J. L., Jackson, R., Fritsch Fredin, M., Jansson-Löfmark, R., Naylor, J., Rossi, A., Bednarek, M. A., Bhagroo, N., Salari, H., Will, S., Oldham, S., Hansen, G., Feigh, M., Klein, T., Grimsby, J., Maguire, S., … Coghlan, M. P. (2016). Robust anti-obesity and metabolic effects of a dual GLP-1/glucagon receptor peptide agonist in rodents and non-human primates. In *Diabetes Obes Metab* (Vol. 18, pp. 1176–1190).

Henriques, S. T., Melo, M. N., & Castanho, M. A. (2007). How to address CPP and AMP translocation? Methods to detect and quantify peptide internalization in vitro and in vivo (Review). *Molecular Membrane Biology*, *24*(3), 173–184.

Herman, C., Thévenet, D., D'Ari, R., & Bouloc, P. (1995). Degradation of sigma 32, the heat shock regulator in Escherichia coli, is governed by HflB. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(8), 3516–3520.

Higashitani, A., Ishii, Y., Kato, Y., & Koriuchi, K. (1997). Functional dissection of a cell-division inhibitor, SulA, of Escherichia coli and its negative regulation by Lon. *Molecular & General Genetics: MGG*, *254*(4), 351–357.

Hintsche, M., & Klumpp, S. (2013). Dilution and the theoretical description of growth-rate dependent gene expression. *Journal of Biological Engineering*, *7*(1), 22.

Hjelm, A., Karyolaimos, A., Zhang, Z., Rujas, E., Vikström, D., Slotboom, D. J., & de Gier, J.-W. (2017). Tailoring Escherichia coli for the l-Rhamnose PBAD Promoter-Based Production of Membrane and Secretory Proteins. *ACS Synthetic Biology*, *6*(6), 985–994.

Hooven, M. D. (2017). *Opportunities and Challenges in Biologic Drug Delivery*. https://www.americanpharmaceuticalreview.com/Featured-Articles/345540-Opportunities-and-Challenges-in-Biologic-Drug-Delivery/

Huang, M., Bao, J., Hallström, B. M., Petranovic, D., & Nielsen, J. (2017). Efficient protein production by yeast requires global tuning of metabolism. *Nature Communications*, *8*(1), 1131.

Huang, X., & Brazel, C. S. (2001). On the importance and mechanisms of burst release in matrix-controlled drug delivery systems. *Journal of Controlled Release: Official Journal of the Controlled Release Society*, *73*(2-3), 121–136.

Huber, R., Roth, S., Rahmen, N., & Buchs, J. (2011). Utilizing high-throughput experimentation to enhance specific productivity of an E.coli T7 expression system by phosphate limitation. *BMC Biotechnology*, *11*, 22.

Huerta, A. M., & Collado-Vides, J. (2003). Sigma70 promoters in Escherichia coli: specific transcription in dense regions of overlapping promoter-like signals. *Journal of Molecular Biology*, *333*(2), 261–278.

Hunter, M., Yuan, P., Vavilala, D., & Fox, M. (2019). Optimization of Protein Expression in Mammalian Cells. *Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [et Al.]*, *95*(1), e77.

Hussain, H., Fisher, D. I., & Abbott, W. M. (2017). Use of a protein engineering strategy to overcome limitations in the production of "Difficult to Express" recombinant proteins. *BioTechnology* . https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.26358?casa_token=Rs_uESQTUiUAAAAA:u9WDuC3ogwyK_f0Y5NWKn1Qr6v5uSukoVSFQMqaLJbpDm38RSdYf-BimnhFyJSoYbwq27Qs73Dsg

Ikushima, I., Jensen, L., Flint, A., Nishida, T., Zacho, J., & Irie, S. (2018). A Randomized Trial Investigating the Pharmacokinetics, Pharmacodynamics, and Safety of Subcutaneous Semaglutide Once-Weekly in Healthy Male Japanese and Caucasian Subjects. In *Adv Ther* (Vol. 35, pp. 531–544).

Ito, K., & Akiyama, Y. (2005). Cellular functions, mechanism of action, and regulation of FtsH protease. *Annual Review of Microbiology*, *59*, 211–231.

Jackson, D. C., Lau, Y. F., Le, T., Suhrbier, A., Deliyannis, G., Cheers, C., Smith, C., Zeng, W., & Brown, L. E. (2004). A totally synthetic vaccine of generic structure that targets Toll-like receptor 2 on dendritic cells and promotes antibody or cytotoxic T cell responses. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(43), 15440–15445.

Jacobsen, L. V., Flint, A., Olsen, A. K., & Ingwersen, S. H. (2016). Liraglutide in Type 2 Diabetes Mellitus: Clinical Pharmacokinetics and Pharmacodynamics. In *Clin Pharmacokinet* (Vol. 55, pp. 657–672).

Jechlinger, W., Szostak, M. P., Witte, A., & Lubitz, W. (1999). Altered temperature induction sensitivity of the lambda pR/cI857 system for controlled gene E expression in Escherichia coli. *FEMS Microbiology Letters*, *173*(2), 347–352.

Jensen, L., Helleberg, H., Roffel, A., van Lier, J. J., Bjornsdottir, I., Pedersen, P. J., Rowe, E., Derving Karsbol, J., & Pedersen, M. L. (2017). Absorption, metabolism and excretion of the GLP-1 analogue semaglutide in humans and nonclinical species. *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences*, *104*, 31–41.

Jensen, L., Kupcova, V., Arold, G., Pettersson, J., & Hjerpsted, J. B. (2018). Pharmacokinetics and tolerability of semaglutide in people with hepatic impairment. *Diabetes, Obesity & Metabolism*, *20*(4), 998–1005.

Jian, J., Zhang, S.-Q., Shi, Z.-Y., Wang, W., Chen, G.-Q., & Wu, Q. (2010). Production of polyhydroxyalkanoates by Escherichia coli mutants with defected mixed acid fermentation pathways. *Applied Microbiology and Biotechnology*, *87*(6), 2247–2256.

Jin, X., & Hong, S. H. (2018). Cell-free protein synthesis for producing "difficult-to-express" proteins. *Biochemical Engineering Journal*. https://www.sciencedirect.com/science/article/pii/S1369703X18302614

Johannsson, G., Feldt-Rasmussen, U., Håkonsson, I. H., Biering, H., Rodien, P., Tahara, S., Toogood, A., Rasmussen, M. H., Karges, W., Mann, A., Christiansen, J. S., Hansen, T. K., Andersen, M., Borresen, S., Borson-Chazot, F., Kerlan, V., Cariou, B., Verges, B., Matsuno, A., … Brooke, A. (2018). Safety and convenience of once-weekly somapacitan in adult GH deficiency: a 26-week randomized, controlled trial. In *Eur J Endocrinol* (Vol. 178, pp. 491–499).

Jones, C. H., Dexter, P., Evans, A. K., Liu, C., Hultgren, S. J., & Hruby, D. E. (2002). Escherichia coli DegP protease cleaves between paired hydrophobic residues in a natural substrate: the PapA pilin. *Journal of Bacteriology*, *184*(20), 5762–5771.

Jones, J. A., Vernacchio, V. R., Lachance, D. M., Lebovich, M., Fu, L., Shirke, A. N., Schultz, V. L., Cress, B., Linhardt, R. J., & Koffas, M. A. G. (2015). ePathOptimize: A Combinatorial Approach for Transcriptional Balancing of Metabolic Pathways. *Scientific Reports*, *5*, 11301.

Jozala, A. F., Geraldes, D. C., Tundisi, L. L., Feitosa, V. de A., Breyer, C. A., Cardoso, S. L., Mazzola, P. G., Oliveira-Nascimento, L. de, Rangel-Yagui, C. de O., Magalhães, P. de O., Oliveira, M. A. de, & Pessoa, A., Jr. (2016). Biopharmaceuticals from microorganisms: from production to purification. *Brazilian Journal of Microbiology: [publication of the Brazilian Society for Microbiology]*, *47 Suppl 1*, 51–63.

Kanemori, M., Mori, H., & Yura, T. (1994). Induction of heat shock proteins by abnormal proteins results from stabilization and not increased synthesis of sigma 32 in Escherichia coli. *Journal of Bacteriology*, *176*(18), 5648–5653.

Karsdal, M. A., Riis, B. J., Mehta, N., Stern, W., Arbit, E., Christiansen, C., & Henriksen, K. (2015). Lessons learned from the clinical development of oral peptides. *British Journal of Clinical Pharmacology*, *79*(5), 720–732.

Kastilan, R., Boes, A., Spiegel, H., Voepel, N., Chudobová, I., Hellwig, S., Buyel, J. F., Reimann, A., & Fischer, R. (2017). Improvement of a fermentation process for the production of two PfAMA1-DiCo-based malaria vaccine candidates in Pichia pastoris. In *Scientific Reports* (Vol. 7, Issue 1). https://doi.org/10.1038/s41598-017-11819-4

Katz, C., & Ron, E. Z. (2008). Dual role of FtsH in regulating lipopolysaccharide biosynthesis in Escherichia coli. *Journal of Bacteriology*, *190*(21), 7117–7122.

Kauffman, W. B., Fuselier, T., He, J., & Wimley, W. C. (2015). Mechanism Matters: A Taxonomy of Cell Penetrating Peptides. *Trends in Biochemical Sciences*, *40*(12), 749–764.

Kaur, J., Kumar, A., & Kaur, J. (2018). Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements. *International Journal of Biological Macromolecules*, *106*, 803–822.

Keizer, R. J., Huitema, A. D. R., Schellens, J. H. M., & Beijnen, J. H. (2010). Clinical pharmacokinetics of therapeutic monoclonal antibodies. *Clinical Pharmacokinetics*, *49*(8), 493–507.

Khlebnikov, A., Risa, O., Skaug, T., Carrier, T. A., & Keasling, J. D. (2000). Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *Journal of Bacteriology*, *182*(24), 7029–7034.

Kightlinger, W., Duncker, K. E., Ramesh, A., Thames, A. H., Natarajan, A., Stark, J. C., Yang, A., Lin, L., Mrksich, M., DeLisa, M. P., & Jewett, M. C. (2019). A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nature Communications*, *10*(1), 5404.

Kim, D., Jeon, H., Ahn, S., Choi, W. I., Kim, S., & Jon, S. (2017). An approach for half-life extension and activity preservation of an anti-diabetic peptide drug based on genetic fusion with an albumin-binding aptide. *Journal of Controlled Release: Official Journal of the Controlled Release Society*, *256*, 114–120.

Kinch, M. S. (2015). An overview of FDA-approved biologics medicines. *Drug Discovery Today*, *20*(4), 393–398.

Kloos, D. U., Strätz, M., Güttler, A., Steffan, R. J., & Timmis, K. N. (1994). Inducible cell lysis system for the study of natural transformation and environmental fate of DNA released by cell death. *Journal of Bacteriology*, *176*(23), 7352–7361.

Knoll, L. J., & Gordon, J. I. (1993). Use of Escherichia coli strains containing fad mutations plus a triple plasmid expression system to study the import of myristate, its activation by Saccharomyces …. *The Journal of Biological Chemistry*. http://www.jbc.org/content/268/6/4281.short

Knudsen, L. B., Nielsen, P. F., Huusfeldt, P. O., Johansen, N. L., Madsen, K., Pedersen, F. Z., Thogersen, H., Wilken, M., & Agerso, H. (2000). Potent derivatives of glucagon-like peptide-1 with pharmacokinetic properties suitable for once daily administration. *Journal of Medicinal Chemistry*, *43*(9), 1664–1669.

Kontermann, R. E. (2011). Strategies for extended serum half-life of protein therapeutics. *Current Opinion in Biotechnology*, *22*(6), 868–876.

Kullar, R., Chin, J. N., Edwards, D. J., Parker, D., Coplin, W. M., & Rybak, M. J. (2011). Pharmacokinetics of Single-Dose Daptomycin in Patients with Suspected or Confirmed Neurological Infections. *Antimicrobial Agents and Chemotherapy*, *55*(7), 3505.

Kurtzhals, P., Havelund, S., Jonassen, I., Kiehr, B., Larsen, U. D., Ribel, U., & Markussen, J. (1995). Albumin binding of insulins acylated with fatty acids: characterization of the ligand-protein interaction and correlation between binding affinity and timing of the insulin effect in vivo. *Biochemical Journal*, *312 (Pt 3)*, 725–731.

Labhsetwar, P., Cole, J. A., Roberts, E., Price, N. D., & Luthey-Schulten, Z. A. (2013). Heterogeneity in protein expression induces metabolic variability in a modeled Escherichia coli population. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(34), 14006–14011.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, *227*(5259), 680–685.

Langenheim, J. F., & Chen, W. Y. (2009). Improving the pharmacokinetics/pharmacodynamics of prolactin, GH, and their antagonists by fusion to a synthetic albumin-binding peptide. *The Journal of Endocrinology*, *203*(3), 375–387.

Lanyon-Hogg, T., Faronato, M., Serwa, R. A., & Tate, E. W. (2017). Dynamic Protein Acylation: New Substrates, Mechanisms, and Drug Targets. *Trends in Biochemical Sciences*, *42*(7), 566–581.

Lau, J., Bloch, P., Schaffer, L., Pettersson, I., Spetzler, J., Kofoed, J., Madsen, K., Knudsen, L. B., McGuire, J., Steensgaard, D. B., Strauss, H. M., Gram, D. X., Knudsen, S. M., Nielsen, F. S., Thygesen, P., Reedtz-Runge, S., & Kruse, T. (2015). Discovery of the Once-Weekly Glucagon-Like Peptide-1 (GLP-1) Analogue Semaglutide. *Journal of Medicinal Chemistry*, *58*(18), 7370–7380.

Laukens, B., Jacobs, P. P., Geysens, K., Martins, J., De Wachter, C., Ameloot, P., Morelle, W., Haustraete, J., Renauld, J.-C., Samyn, B., Contreras, R., Devos, S., & Callewaert, N. (2020). Off-target glycans encountered along the synthetic biology route towards humanized N-glycans in Pichia pastoris. In *bioRxiv* (p. 2020.01.28.923631). https://doi.org/10.1101/2020.01.28.923631

Lee, J. S., & Tung, C. H. (2010). Lipo-oligoarginines as effective delivery vectors to promote cellular uptake. *Molecular bioSystems*, *6*(10), 2049–2055.

Lemmerer, M., Mairhofer, J., Lepak, A., Longus, K., Hahn, R., & Nidetzky, B. (2019). Decoupling of recombinant protein production from Escherichia coli cell growth enhances functional expression of plant Leloir glycosyltransferases. *Biotechnology and Bioengineering*, *116*(6), 1259–1268.

Lerchner, A., Jarasch, A., & Skerra, A. (2016). Engineering of alanine dehydrogenase from Bacillus subtilis for novel cofactor specificity. *Biotechnology and Applied Biochemistry*, *63*(5), 616–624.

Levin, D., Golding, B., Strome, S. E., & Sauna, Z. E. (2015). Fc fusion as a platform technology: potential for modulating immunogenicity. *Trends in Biotechnology*, *33*(1), 27–34.

Liao, J.-H., Kuo, C.-I., Huang, Y.-Y., Lin, Y.-C., Lin, Y.-C., Yang, C.-Y., Wu, W.-L., Chang, W.-H., Liaw, Y.-C., Lin, L.-H., Chang, C.-I., & Wu, S.-H. (2012). A Lon-like protease with no ATP-powered unfolding activity. *PloS One*, *7*(7), e40226.

Liao, X., Zhao, J., Liang, S., Jin, J., Li, C., Xiao, R., Li, L., Guo, M., Zhang, G., & Lin, Y. (2019). Enhancing co-translational folding of heterologous protein by deleting non-essential ribosomal proteins in Pichia pastoris. *Biotechnology for Biofuels*, *12*, 38.

Lipscomb, Lynch, & Gill. (2018). *Method for producing 3-hydroxypropionic acid and other products*. https://www.osti.gov/biblio/1489753

Listwan, P., Terwilliger, T. C., & Waldo, G. S. (2009). Automated, high-throughput platform for protein solubility screening using a split-GFP system. *Journal of Structural and Functional Genomics*, *10*(1), 47–55.

Liu, X., Ding, W., & Jiang, H. (2017). Engineering microbial cell factories for the production of plant natural products: from design principles to industrial-scale production. In *Microbial Cell Factories* (Vol. 16, Issue 1). https://doi.org/10.1186/s12934-017-0732-7

Li, X.-T., Thomason, L. C., Sawitzke, J. A., Costantino, N., & Court, D. L. (2013). Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli. *Nucleic Acids Research*, *41*(22), e204.

Long, C. P., & Antoniewicz, M. R. (2014). Quantifying biomass composition by gas chromatography/mass spectrometry. *Analytical Chemistry*, *86*(19), 9423–9427.

Lübke, C., Boidol, W., & Petri, T. (1995). Analysis and optimization of recombinant protein production in Escherichia coli using the inducible pho A promoter of the E. coli alkaline phosphatase. In *Enzyme and Microbial Technology* (Vol. 17, Issue 10, pp. 923–928). https://doi.org/10.1016/0141-0229(94)00130-j

Luginbuhl, K. M., Mozhdehi, D., Dzuricky, M., Yousefpour, P., Huang, F. C., Mayne, N. R., Buehne, K. L., & Chilkoti, A. (2017). Recombinant Synthesis of Hybrid Lipid-Peptide Polymer Fusions that Self-Assemble and Encapsulate Hydrophobic Drugs. In *Angewandte Chemie* (Vol. 129, Issue 45, pp. 14167–14172). https://doi.org/10.1002/ange.201704625

Luo, M. L., Mullis, A. S., Leenay, R. T., & Beisel, C. L. (2015). Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. In *Nucleic Acids Research* (Vol. 43, Issue 1, pp. 674–681). https://doi.org/10.1093/nar/gku971

Lynch. (2011). Compositions and methods for 3-hydroxypropionate bio-production from biomass. In *US Patent 8,048,624*. https://patents.google.com/patent/US8048624B1/en

Lynch, M. D. (2016). Into new territory: improved microbial synthesis through engineering of the essential metabolic network. *Current Opinion in Biotechnology*, *38*, 106–111.

Lynch, M. D., Trahan, A. D., Rodriguez, D., & Ye, Z. (2017). Compositions and methods for rapid and dynamic flux control using synthetic metabolic valves. In *US Patent App. 15*. https://patents.google.com/patent/US20170121707A1/en

Lynch, M. D., Trahan, A. D., Rodriguez, D., Ye, Z., Cooper, C. B., & Bozdag, A. (2017). Compositions and methods for rapid and dynamic flux control using synthetic metabolic valves (USPTO Patent No. 20170121707:A1). In *US Patent* (No. 20170121707:A1). https://patentimages.storage.googleapis.com/61/0c/50/d6de63c7db5027/US20170121707A1.pdf

Lynch, M., Louie, M., Copley, S., Spindler, E., Prather, B., Lipscomb, M., Lipscomb, T., Liao, H., Hogsett, D., & Evans, R. (2019). Microorganisms and methods for the production of fatty acids and fatty acid derived products (USPTO Patent No. 10337038). In *US Patent* (No. 10337038). https://patentimages.storage.googleapis.com/de/d6/14/edeaa6156eb2d7/US10337038.pdf

Madsen, K., Knudsen, L. B., Agersoe, H., Nielsen, P. F., Thogersen, H., Wilken, M., & Johansen, N. L. (2007). Structure-activity and protraction relationship of long-acting glucagon-like peptide-1 derivatives: importance of fatty acid length, polarity, and bulkiness. *Journal of Medicinal Chemistry*, *50*(24), 6126–6132.

Mairhofer, J., Wittwer, A., Cserjan-Puschmann, M., & Striedner, G. (2015). Preventing T7 RNA Polymerase Read-through Transcription—A Synthetic Termination Signal Capable of Improving Bioprocess Stability. In *ACS Synthetic Biology* (Vol. 4, Issue 3, pp. 265–273). https://doi.org/10.1021/sb5000115

Malm-Erjefalt, M., Bjornsdottir, I., Vanggaard, J., Helleberg, H., Larsen, U., Oosterhuis, B., van Lier, J. J., Zdravkovic, M., & Olsen, A. K. (2010). Metabolism and excretion of the once-daily human glucagon-like peptide-1 analog liraglutide in healthy male subjects and its in vitro degradation by dipeptidyl peptidase IV and neutral endopeptidase. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *38*(11), 1944–1953.

Mamat, U., Wilke, K., Bramhill, D., Schromm, A. B., Lindner, B., Kohl, T. A., Corchero, J. L., Villaverde, A., Schaffer, L., Head, S. R., Souvignier, C., Meredith, T. C., & Woodard, R. W. (2015). Detoxifying Escherichia coli for endotoxin-free production of recombinant proteins. *Microbial Cell Factories*, *14*, 57.

Märkl, H., Zenneck, C., Dubach, A. C., & Ogbonna, J. C. (1993). Cultivation of Escherichia coli to high cell densities in a dialysis reactor. *Applied Microbiology and Biotechnology*, *39*(1), 48–52.

Markussen, J., Havelund, S., Kurtzhals, P., Andersen, A. S., Halstrom, J., Hasselager, E., Larsen, U. D., Ribel, U., Schaffer, L., Vad, K., & Jonassen, I. (1996). Soluble, fatty acid acylated insulins bind to albumin and show protracted action in pigs. *Diabetologia*, *39*(3), 281–288.

Matthews, J. E., Stewart, M. W., De Boever, E. H., Dobbins, R. L., Hodge, R. J., Walker, S. E., Holland, M. C., & Bush, M. A. (2008). Pharmacodynamics, pharmacokinetics, safety, and tolerability of albiglutide, a long-acting glucagon-like peptide-1 mimetic, in patients with type 2 diabetes. *The Journal of Clinical Endocrinology and Metabolism*, *93*(12), 4810–4817.

McGinness, K. E., Baker, T. A., & Sauer, R. T. (2006). Engineering controllable protein degradation. *Molecular Cell*, *22*(5), 701–707.

McKenna, R., Noelle Lombana, T., Yamada, M., Mukhyala, K., & Veeravalli, K. (2019). Engineered sigma factors increase full-length antibody expression in Escherichia coli. In *Metabolic Engineering* (Vol. 52, pp. 315–323). https://doi.org/10.1016/j.ymben.2018.12.009

Mejuch, T., & Waldmann, H. (2016). Synthesis of Lipidated Proteins. *Bioconjugate Chemistry*, *27*(8), 1771–1783.

Menacho-Melgar, R., Decker, J. S., Hennigan, J. N., & Lynch, M. D. (2019). A review of lipidation in the development of advanced protein and peptide therapeutics. *Journal of Controlled Release: Official Journal of the Controlled Release Society*, *295*, 1–12.

Menacho-Melgar, R., Moreb, E. A., Efromson, J. P., & Lynch, M. D. (n.d.). *Improved, two-stage protein expression and purification via autoinduction of both autolysis and auto DNA/RNA hydrolysis conferred by phage lysozyme and DNA/RNA endonuclease*. https://doi.org/10.1101/2020.01.09.900753

Menacho-Melgar, R., Ye, Z., Moreb, E. A., Yang, T., Efromson, J. P., Decker, J. S., & Lynch, M. D. (2020). Improved, scalable, two-stage, autoinduction of recombinant protein expression in E. coli utilizing phosphate depletion. In *bioRxiv* (p. 820787). https://doi.org/10.1101/820787

Meyer, A. S., & Baker, T. A. (2011). Proteolysis in the Escherichia coli heat shock response: a player at many levels. *Current Opinion in Microbiology*, *14*(2), 194–199.

Miersch, S., & Sidhu, S. S. (2016). Intracellular targeting with engineered proteins. *F1000Research*, *5*. https://doi.org/10.12688/f1000research.8915.1

Mima, Y., Hashimoto, Y., Shimizu, T., Kiwada, H., & Ishida, T. (2015). Anti-PEG IgM Is a Major Contributor to the Accelerated Blood Clearance of Polyethylene Glycol-Conjugated Protein. *Molecular Pharmaceutics*, *12*(7), 2429–2435.

Missirlis, D., Khant, H., & Tirrell, M. (2009). Mechanisms of peptide amphiphile internalization by SJSA-1 cells in vitro. *Biochemistry*, *48*(15), 3304–3314.

Miyakawa, N., Nishikawa, M., Takahashi, Y., Ando, M., Misaka, M., Watanabe, Y., & Takakura, Y. (2011). Prolonged circulation half-life of interferon gamma activity by gene delivery of interferon gamma-serum albumin fusion protein in mice. *Journal of Pharmaceutical Sciences*, *100*(6), 2350–2357.

Morales, J. O., Fathe, K. R., Brunaugh, A., Ferrati, S., Li, S., Montenegro-Nicolini, M., Mousavikhamene, Z., McConville, J. T., Prausnitz, M. R., & Smyth, H. D. C. (2017). Challenges and Future Prospects for the Delivery of Biologics: Oral Mucosal, Pulmonary, and Transdermal Routes. *The AAPS Journal*, *19*(3), 652–668.

Moremen, K. W., Ramiah, A., Stuart, M., Steel, J., Meng, L., Forouhar, F., Moniz, H. A., Gahlay, G., Gao, Z., Chapla, D., Wang, S., Yang, J.-Y., Prabhakar, P. K., Johnson, R., Rosa, M. D., Geisler,

C., Nairn, A. V., Seetharaman, J., Wu, S.-C., … Jarvis, D. L. (2018). Expression system for structural and functional studies of human glycosylation enzymes. *Nature Chemical Biology*, *14*(2), 156–162.

Morita, D., Yamamoto, Y., Mizutani, T., Ishikawa, T., Suzuki, J., Igarashi, T., Mori, N., Shiina, T., Inoko, H., Fujita, H., Iwai, K., Tanaka, Y., Mikami, B., & Sugita, M. (2016). Crystal structure of the N-myristoylated lipopeptide-bound MHC class I complex. *Nature Communications*, *7*, 10356.

Morita, M., Asami, K., Tanji, Y., & Unno, H. (2001). Programmed Escherichia coli cell lysis by expression of cloned T4 phage lysis genes. *Biotechnology Progress*, *17*(3), 573–576.

Nakatogawa, H., & Ito, K. (2002). The ribosomal exit tunnel functions as a discriminating gate. *Cell*, *108*(5), 629–636.

Navon, S. P., Kornberg, G., Chen, J., Schwartzman, T., Tsai, A., Puglisi, E. V., Puglisi, J. D., & Adir, N. (2016). Amino acid sequence repertoire of the bacterial proteome and the occurrence of untranslatable sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(26), 7166–7170.

Neidhardt, F. C., Ingraham, J. L., & Schaechter, M. (1990). *Physiology of the bacterial cell: a molecular approach* (Vol. 20). Sinauer Associates Sunderland, MA.

Nelson, A. R., Borland, L., Allbritton, N. L., & Sims, C. E. (2007). Myristoyl-based transport of peptides into living cells. *Biochemistry*, *46*(51), 14771–14781.

Nocadello, S., & Swennen, E. F. (2012). The new pLAI (lux regulon based auto-inducible) expression system for recombinant protein production in Escherichia coli. *Microbial Cell Factories*, *11*, 3.

Novick, A., & Weiner, M. (1957). ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON. *Proceedings of the National Academy of Sciences of the United States of America*, *43*(7), 553–566.

Oakley, A. (2011). Glutathione transferases: a structural perspective. *Drug Metabolism Reviews*, *43*(2), 138–151.

Ogura, T., Inoue, K., Tatsuta, T., Suzaki, T., Karata, K., Young, K., Su, L. H., Fierke, C. A., Jackman, J. E., Raetz, C. R., Coleman, J., Tomoyasu, T., & Matsuzawa, H. (1999). 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. *Molecular Microbiology*, *31*(3), 833–844.

Oh, D., Nasrolahi Shirazi, A., Northup, K., Sullivan, B., Tiwari, R. K., Bisoffi, M., & Parang, K. (2014). Enhanced Cellular Uptake of Short Polyarginine Peptides through Fatty Acylation and Cyclization. *Molecular Pharmaceutics*, *11*(8), 2845–2854.

Olszewski, M., & Filipkowski, P. (2009). Benzonase—possibility of practical application. *POSTĘPY BIOCHEMII*, 24.

Otto, R., Santagostino, A., & Schrader, U. (2014). *Rapid growth in biopharma: Challenges and opportunities*. McKinsey&Company. https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/rapid-growth-in-biopharma

Patel, L. N., Wang, J., Kim, K. J., Borok, Z., Crandall, E. D., & Shen, W. C. (2009). Conjugation with cationic cell-penetrating peptide increases pulmonary absorption of insulin. *Molecular Pharmaceutics*, *6*(2), 492–503.

Patton, J. S., Fishburn, C. S., & Weers, J. G. (2004). The lungs as a portal of entry for systemic drug delivery. *Proceedings of the American Thoracic Society*, *1*(4), 338–344.

Pellizza, L., Smal, C., Rodrigo, G., & Arán, M. (2018). Codon usage clusters correlation: towards protein solubility prediction in heterologous expression systems in E. coli. *Scientific Reports*, *8*(1), 10618.

Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B., & Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature*, *488*(7411), 320–328.

Pereira, E. A., & daSilva, L. L. P. (2016). HIV-1 Nef: Taking Control of Protein Trafficking. *Traffic* , *17*(9), 976–996.

Pfender, N. A., Grosch, S., Roussel, G., Koch, M., Trifilieff, E., & Greer, J. M. (2008). Route of uptake of palmitoylated encephalitogenic peptides of myelin proteolipid protein by antigen-presenting cells: importance of the type of bond between lipid chain and peptide and relevance to autoimmunity. *Journal of Immunology* , *180*(3), 1398–1404.

Pfender, N., Guéin, E., Greer, J. M., & Trifilieff, E. (2003). Solid-phase synthesis of a biotin-labelled thiopalmitoylated myelin proteolipid protein epitope and application in the study of uptake of antigen by macrophages. *Letters in Peptide Science: LIPS*, *10*(5), 581–588.

Piirainen, M. A., Boer, H., de Ruijter, J. C., & Frey, A. D. (2016). A dual approach for improving homogeneity of a human-type N-glycan structure in Saccharomyces cerevisiae. *Glycoconjugate Journal*, *33*(2), 189–199.

Plum, A., Jensen, L. B., & Kristensen, J. B. (2013). In vitro protein binding of liraglutide in human plasma determined by reiterated stepwise equilibrium dialysis. *Journal of Pharmaceutical Sciences*, *102*(8), 2882–2888.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183.

Quinlan, G. J., Martin, G. S., & Evans, T. W. (2005). Albumin: biochemical properties and therapeutic potential. *Hepatology* , *41*(6), 1211–1219.

Ramírez-Andersen, H. S., Behrens, C., Buchardt, J., Fels, J. J., Folkesson, C. G., Jianhe, C., Nørskov-Lauritsen, L., Nielsen, P. F., Reslow, M., Rischel, C., Su, J., Thygesen, P., Wiberg, C., Zhao, X., Wenjuan, X., & Johansen, N. L. (2018). Long-Acting Human Growth Hormone Analogue by

Noncovalent Albumin Binding. *Bioconjugate Chemistry*. https://doi.org/10.1021/acs.bioconjchem.8b00463

Ratelade, J., -C. Miot, M., Johnson, E., -M. Betton, J., Mazodier, P., & Benaroudj, N. (2009). Production of Recombinant Proteins in the lon-Deficient BL21(DE3) Strain of Escherichia coli in the Absence of the DnaK Chaperone. In *Applied and Environmental Microbiology* (Vol. 75, Issue 11, pp. 3803–3807). https://doi.org/10.1128/aem.00255-09

Rennig, M., Martinez, V., Mirzadeh, K., Dunas, F., Röjsäter, B., Daley, D. O., & Nørholm, M. H. H. (2018). TARSyn: Tunable Antibiotic Resistance Devices Enabling Bacterial Synthetic Evolution and Protein Production. *ACS Synthetic Biology*, *7*(2), 432–442.

Reyes, L. H., Cardona, C., Pimentel, L., Rodríguez-López, A., & Alméciga-Díaz, C. J. (2017). Improvement in the production of the human recombinant enzyme N-acetylgalactosamine-6-sulfatase (rhGALNS) in Escherichia coli using synthetic biology approaches. In *Scientific Reports* (Vol. 7, Issue 1). https://doi.org/10.1038/s41598-017-06367-w

Reymond, J.-L. (2006). *Enzyme Assays: High-throughput Screening, Genetic Selection and Fingerprinting*. John Wiley & Sons.

Richard, J. P., Melikov, K., Vives, E., Ramos, C., Verbeure, B., Gait, M. J., Chernomordik, L. V., & Lebleu, B. (2003). Cell-penetrating peptides. A reevaluation of the mechanism of cellular uptake. *The Journal of Biological Chemistry*, *278*(1), 585–590.

Robinson, M.-P., Ke, N., Lobstein, J., Peterson, C., Szkodny, A., Mansell, T. J., Tuckey, C., Riggs, P. D., Colussi, P. A., Noren, C. J., Taron, C. H., DeLisa, M. P., & Berkmen, M. (2015). Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. *Nature Communications*, *6*, 8072.

Rocco, D., Ross, J., Murray, P. E., & Caccetta, R. (2016). Acyl lipidation of a peptide: effects on activity and epidermal permeability in vitro. *Drug Design, Development and Therapy*, *10*, 2203–2209.

Rodionova, N., Peng, Z., Miller, E., & Nilsson, M. (n.d.). *Detection of AAV structure: capsid proteins and nucleic acid composition on LabChip® GXII Touch™ Platform*. http://www.perkinelmer.co.jp/Portals/0/resource/products_ls/dnarna/pdf/AP_note_etc/Poster_ProteinEXact_AAV8.pdf

Rosano, G. L., & Ceccarelli, E. A. (2014a). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, *5*, 172.

Rosano, G. L., & Ceccarelli, E. A. (2014b). Recombinant protein expression in microbial systems. *Frontiers in Microbiology*, *5*, 341.

Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019). New tools for recombinant protein production in Escherichia coli : A 5-year update. *Protein Science: A Publication of the Protein Society*, *28*(8), 1412–1422.

Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F., & Bikard, D. (2018). Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. *PLoS Genetics*, *14*(11), e1007749.

Rude, M. A., Baron, T. S., Brubaker, S., Alibhai, M., Del Cardayre, S. B., & Schirmer, A. (2011). Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from Jeotgalicoccus species. *Applied and Environmental Microbiology*, *77*(5), 1718–1727.

Running, J. A., & Bansal, K. (2016). Oxygen transfer rates in shaken culture vessels from Fernbach flasks to microtiter plates. *Biotechnology and Bioengineering*, *113*(8), 1729–1735.

Ryan, G. J., & Hardy, Y. (2011). Liraglutide: once-daily GLP-1 agonist for the treatment of type 2 diabetes. In *Journal of Clinical Pharmacy and Therapeutics* (Vol. 36, Issue 3, pp. 260–274). https://doi.org/10.1111/j.1365-2710.2010.01180.x

Saccardo, P., Corchero, J. L., & Ferrer-Miralles, N. (2016). Tools to cope with difficult-to-express proteins. *Applied Microbiology and Biotechnology*, *100*(10), 4347–4355.

Saïda, F., Uzan, M., Odaert, B., & Bontems, F. (2006). Expression of highly toxic genes in E. coli: special strategies and genetic tools. *Current Protein & Peptide Science*, *7*(1), 47–56.

Sakr, S., Cirinesi, A.-M., Ullers, R. S., Schwager, F., Georgopoulos, C., & Genevaux, P. (2010). Lon protease quality control of presecretory proteins in Escherichia coli and its dependence on the SecB and DnaJ (Hsp40) chaperones. *The Journal of Biological Chemistry*, *285*(30), 23506–23514.

Salari, R., & Salari, R. (2017). Investigation of the Best Saccharomyces cerevisiae Growth Condition. *Electronic Physician*, *9*(1), 3592–3597.

Salis, H. M. (2011). The ribosome binding site calculator. *Methods in Enzymology*, *498*, 19–42.

Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., & Villaverde, A. (2016). Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microbial Cell Factories*, *15*, 33.

Saunders, R., & Deane, C. M. (2010). Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Research*, *38*(19), 6719–6728.

Schirmer, A., Rude, M. A., Li, X., Popova, E., & del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science*, *329*(5991), 559–562.

Schoemaker, J. M., Gayda, R. C., & Markovitz, A. (1984). Regulation of cell division in Escherichia coli: SOS induction and cellular location of the sulA protein, a key to lon-associated filamentation and death. *Journal of Bacteriology*, *158*(2), 551–561.

Schultz, H. S., Ostergaard, S., Sidney, J., Lamberth, K., & Sette, A. (2018). The effect of acylation with fatty acids and other modifications on HLA class II:peptide binding and T cell stimulation for three model peptides. *PloS One*, *13*(5), e0197407.

Shafrir, E., Gatt, S., & Khasis, S. (1965). PARTITION OF FATTY ACIDS OF 20-24 CARBON ATOMS BETWEEN SERUM ALBUMIN AND LIPOPROTEINS. *Biochimica et Biophysica Acta*, *98*, 365–371.

Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., & Court, D. L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols*, *4*(2), 206–223.

Shiber, A., Döring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., Tippmann, F., Kramer, G., & Bukau, B. (2018). Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature*, *561*(7722), 268–272.

Shiloach, J., & Fass, R. (2005). Growing E. coli to high cell density—A historical perspective on method development. *Biotechnology Advances*, *23*(5), 345–357.

Simonen, M., & Palva, I. (1993). Protein secretion in Bacillus species. *Microbiological Reviews*, *57*(1), 109–137.

Skovbakke, S. L., Heegaard, P. M., Larsen, C. J., Franzyk, H., Forsman, H., & Dahlgren, C. (2015). The proteolytically stable peptidomimetic Pam-(Lys-betaNSpe)6-NH2 selectively inhibits human neutrophil activation via formyl peptide receptor 2. *Biochemical Pharmacology*, *93*(2), 182–195.

Skretas, G., Makino, T., Varadarajan, N., Pogson, M., & Georgiou, G. (2012). Multi-copy genes that enhance the yield of mammalian G protein-coupled receptors in Escherichia coli. *Metabolic Engineering*, *14*(5), 591–602.

Smietana, K., Siatkowski, M., & Moller, M. (2016). Trends in clinical success rates. *Nature Reviews. Drug Discovery*, *15*(6), 379–380.

Song, H., Jiang, J., Wang, X., & Zhang, J. (2017). High purity recombinant human growth hormone (rhGH) expression in Escherichia coli under phoA promoter. In *Bioengineered* (Vol. 8, Issue 2, pp. 147–153). https://doi.org/10.1080/21655979.2016.1212137

Sonoda, H., Morimoto, H., Yoden, E., Koshimura, Y., Kinoshita, M., Golovina, G., Takagi, H., Yamamoto, R., Minami, K., Mizoguchi, A., Tachibana, K., Hirato, T., & Takahashi, K. (2018). A Blood-Brain-Barrier-Penetrating Anti-human Transferrin Receptor Antibody Fusion Protein for Neuronopathic Mucopolysaccharidosis II. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *26*(5), 1366–1374.

Spector, A. A. (1975). Fatty acid binding to plasma albumin. *Journal of Lipid Research*, *16*(3), 165–179.

Stanley, T. L., Chen, C. Y., Branch, K. L., Makimura, H., & Grinspoon, S. K. (2011). Effects of a growth hormone-releasing hormone analog on endogenous GH pulsatility and insulin sensitivity in healthy men. *The Journal of Clinical Endocrinology and Metabolism*, *96*(1), 150–158.

Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B., & Keasling, J. D. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, *463*(7280), 559–562.

Stephenson, R., Varamini, P., Butcher, N., Minchin, R., & Toth, I. (2014). Effect of lipidated gonadotropin-releasing hormone peptides on receptor mediated binding and uptake into prostate cancer cells in vitro. *Nanomedicine: Nanotechnology, Biology, and Medicine*, *10*(8), 1799–1808.

Stoekenbroek, R. M., Stroes, E. S., & Hovingh, G. K. (2015). ApoA-I mimetics. *Handbook of Experimental Pharmacology*, *224*, 631–648.

Striedner, G., Pfaffenzeller, I., Markus, L., Nemecek, S., Grabherr, R., & Bayer, K. (2010). Plasmid-free T7-based Escherichia coli expression systems. *Biotechnology and Bioengineering*, *105*(4), 786–794.

Studier, F. W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology*, *219*(1), 37–44.

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234.

Studier, F. W. (2014). Stable expression clones and auto-induction for protein production in E. coli. *Methods in Molecular Biology* , *1091*, 17–32.

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, *189*(1), 113–130.

Sundy, J. S., Baraf, H. S., Yood, R. A., Edwards, N. L., Gutierrez-Urena, S. R., Treadwell, E. L., Vazquez-Mellado, J., White, W. B., Lipsky, P. E., Horowitz, Z., Huang, W., Maroli, A. N., Waltrip, R. W., 2nd, Hamburger, S. A., & Becker, M. A. (2011). Efficacy and tolerability of pegloticase for the treatment of chronic gout in patients refractory to conventional treatment: two randomized controlled trials. *JAMA: The Journal of the American Medical Association*, *306*(7), 711–720.

Swiecicki, J. M., Di Pisa, M., Lippi, F., Chwetzoff, S., Mansuy, C., Trugnan, G., Chassaing, G., Lavielle, S., & Burlina, F. (2015). Unsaturated acyl chains dramatically enhanced cellular uptake by direct translocation of a minimalist oligo-arginine lipopeptide. *Chemical Communications* , *51*(78), 14656–14659.

Swierczewska, M., Lee, K. C., & Lee, S. (2015). What is the future of PEGylated therapies? *Expert Opinion on Emerging Drugs*, *20*(4), 531–536.

Tamekou Lacmata, S., Yao, L., Xian, M., Liu, H., Kuiate, J.-R., Liu, H., Feng, X., & Zhao, G. (2017). A novel autolysis system controlled by magnesium and its application to poly (3-hydroxypropionate) production in engineered Escherichia coli. *Bioengineered*, *8*(5), 594–599.

Tanji, Y., Asami, K., Xing, X.-H., & Unno, H. (1998). Controlled expression of lysis genes encoded in T4 phage for the gentle disruption of Escherichia coli cells. *Journal of Fermentation and Bioengineering*, *85*(1), 74–78.

Tao, H., Bausch, C., Richmond, C., Blattner, F. R., & Conway, T. (1999). Functional Genomics: Expression Analysis ofEscherichia coli Growing on Minimal and Rich Media. *Journal of Bacteriology*, *181*(20), 6425–6440.

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, *72*(2), 211–222.

Thaxton, C. S., Rink, J. S., Naha, P. C., & Cormode, D. P. (2016). Lipoproteins and lipoprotein mimetics for imaging and drug delivery. *Advanced Drug Delivery Reviews*, *106*(Pt A), 116–131.

The Economist,. (2014). *Going Large*. "The Economist." https://www.economist.com/business/2014/12/30/going-large

Thoring, L., Dondapati, S. K., Stech, M., & Wüstenhagen, D. A. (2017). High-yield production of "difficult-to-express" proteins in a continuous exchange cell-free system based on CHO cell lysates. *Scientific Reports*. https://www.nature.com/articles/s41598-017-12188-8

Todorovic, A., Holder, J. R., Bauzo, R. M., Scott, J. W., Kavanagh, R., Abdel-Malek, Z., & Haskell-Luevano, C. (2005). N-terminal fatty acylated His-dPhe-Arg-Trp-NH(2) tetrapeptides: influence of fatty acid chain length on potency and selectivity at the mouse melanocortin receptors and human melanocytes. *Journal of Medicinal Chemistry*, *48*(9), 3328–3336.

Torres-Cabassa, A. S., & Gottesman, S. (1987). Capsule synthesis in Escherichia coli K-12 is regulated by proteolysis. *Journal of Bacteriology*, *169*(3), 981–989.

Trier, S., Linderoth, L., Bjerregaard, S., Andresen, T. L., & Rahbek, U. L. (2014). Acylation of Glucagon-Like Peptide-2: Interaction with Lipid Membranes and In Vitro Intestinal Permeability. In *PLoS One* (Vol. 9).

Trier, S., Linderoth, L., Bjerregaard, S., Strauss, H. M., Rahbek, U. L., & Andresen, T. L. (2015). Acylation of salmon calcitonin modulates in vitro intestinal peptide flux through membrane permeability enhancement. *European Journal of Pharmaceutics and Biopharmaceutics: Official Journal of Arbeitsgemeinschaft Fur Pharmazeutische Verfahrenstechnik e.V*, *96*, 329–337.

Tsou, A. Y., Ransom, S. C., Gerlt, J. A., Buechter, D. D., Babbitt, P. C., & Kenyon, G. L. (1990). Mandelate pathway of Pseudomonas putida: sequence relationships involving mandelate racemase, (S)-mandelate dehydrogenase, and benzoylformate decarboxylase and expression of benzoylformate decarboxylase in Escherichia coli. In *Biochemistry* (Vol. 29, Issue 42, pp. 9856–9862). https://doi.org/10.1021/bi00494a015

Tsuji, M., Ueda, S., Hirayama, T., Okuda, K., Sakaguchi, Y., Isono, A., & Nagasawa, H. (2013). FRET-based imaging of transbilayer movement of pepducin in living cells by novel intracellular

bioreductively activatable fluorescent probes. *Organic & Biomolecular Chemistry*, *11*(18), 3030–3037.

Tytgat, H. L. P., Lin, C.-W., Levasseur, M. D., Tomek, M. B., Rutschmann, C., Mock, J., Liebscher, N., Terasaka, N., Azuma, Y., Wetter, M., Bachmann, M. F., Hilvert, D., Aebi, M., & Keys, T. G. (2019). Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. *Nature Communications*, *10*(1), 5403.

Vallee, S., Rakhe, S., Reidy, T., Walker, S., Lu, Q., Sakorafas, P., Low, S., & Bitonti, A. (2012). Pulmonary administration of interferon Beta-1a-fc fusion protein in non-human primates using an immunoglobulin transport pathway. *Journal of Interferon & Cytokine Research: The Official Journal of the International Society for Interferon and Cytokine Research*, *32*(4), 178–184.

van Witteloostuijn, S. B., Mannerstedt, K., Wismann, P., Bech, E. M., Thygesen, M. B., Vrang, N., Jelsing, J., Jensen, K. J., & Pedersen, S. L. (2017). Neoglycolipids for Prolonging the Effects of Peptides: Self-Assembling Glucagon-like Peptide 1 Analogues with Albumin Binding Properties and Potent in Vivo Efficacy. *Molecular Pharmaceutics*, *14*(1), 193–205.

Varkouhi, A. K., Scholte, M., Storm, G., & Haisma, H. J. (2011). Endosomal escape pathways for delivery of biologicals. *Journal of Controlled Release: Official Journal of the Controlled Release Society*, *151*(3), 220–228.

Vasala, A., Isomäki, R., Myllykoski, L., & Alatossava, T. (1999). Thymol-triggered lysis of Escherichia coli expressing Lactobacillus phage LL-H muramidase. *Journal of Industrial Microbiology & Biotechnology*, *22*(1), 39–43.

Verma, M., Choi, J., Cottrell, K. A., Lavagnino, Z., Thomas, E. N., Pavlovic-Djuranovic, S., Szczesny, P., Piston, D. W., Zaher, H. S., Puglisi, J. D., & Djuranovic, S. (2019). A short translational ramp determines the efficiency of protein synthesis. *Nature Communications*, *10*(1), 5774.

Veronese, F. M. (2001). Peptide and protein PEGylation: a review of problems and solutions. *Biomaterials*, *22*(5), 405–417.

Villarreal, F., Contreras-Llano, L. E., Chavez, M., Ding, Y., Fan, J., Pan, T., & Tan, C. (2018). Synthetic microbial consortia enable rapid assembly of pure translation machinery. *Nature Chemical Biology*, *14*(1), 29–35.

Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nature Biotechnology*, *36*(12), 1136–1145.

Waluk, D. P., Schultz, N., & Hunt, M. C. (2010). Identification of glycine N-acyltransferase-like 2 (GLYATL2) as a transferase that produces N-acyl glycines in humans. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *24*(8), 2795–2803.

Wang, G., Björk, S. M., Huang, M., Liu, Q., Campbell, K., Nielsen, J., Joensson, H. N., & Petranovic, D. (2019). RNAi expression tuning, microfluidic screening, and genome recombineering for improved protein production in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(19), 9324–9332.

Wang, H., Wang, F., Wang, W., Yao, X., Wei, D., Cheng, H., & Deng, Z. (2014). Improving the expression of recombinant proteins in E. coli BL21 (DE3) under acetate stress: an alkaline pH shift approach. *PloS One*, *9*(11), e112777.

Wang, J., Chow, D., Heiati, H., & Shen, W. C. (2003). Reversible lipidization for the oral delivery of salmon calcitonin. *Journal of Controlled Release: Official Journal of the Controlled Release Society*, *88*(3), 369–380.

Wang, J., Wu, D., & Shen, W. C. (2002). Structure-activity relationship of reversibly lipidized peptides: studies of fatty acid-desmopressin conjugates. *Pharmaceutical Research*, *19*(5), 609–614.

Wang, T., Badran, A. H., Huang, T. P., & Liu, D. R. (2018). Continuous directed evolution of proteins with improved soluble expression. *Nature Chemical Biology*, *14*(10), 972–980.

Wang, Y., Lomakin, A., Kanai, S., Alex, R., Belli, S., Donzelli, M., & Benedek, G. B. (2016). The molecular basis for the prolonged blood circulation of lipidated incretin peptides: Peptide oligomerization or binding to serum albumin? *Journal of Controlled Release: Official Journal of the Controlled Release Society*, *241*, 25–33.

Ward, B. P., Ottaway, N. L., Perez-Tilve, D., Ma, D., Gelfanov, V. M., Tschöp, M. H., & DiMarchi, R. D. (2013). Peptide lipidation stabilizes structure to enhance biological function. In *Molecular Metabolism* (Vol. 2, Issue 4, pp. 468–479). https://doi.org/10.1016/j.molmet.2013.08.008

Warnecke, Lynch, & Gill. (2013). Microorganism production of high-value chemical products, and related compositions, methods and systems (USPTO Patent No. 20130122541A1). In *US Patent* (No. 20130122541A1).

Weber, T. (2019). Engineering of cell factories for the production of natural products. *Natural Product Reports*, *36*(9), 1231–1232.

Welner, D. H., Shin, D., Tomaleri, G. P., DeGiovanni, A. M., Tsai, A. Y.-L., Tran, H. M., Hansen, S. F., Green, D. T., Scheller, H. V., & Adams, P. D. (2017). Plant cell wall glycosyltransferases: High-throughput recombinant expression screening and general requirements for these challenging enzymes. *PloS One*, *12*(6), e0177591.

Wexler-Cohen, Y., & Shai, Y. (2009). Membrane-anchored HIV-1 N-heptad repeat peptides are highly potent cell fusion inhibitors via an altered mode of action. *PLoS Pathogens*, *5*(7). https://www.ncbi.nlm.nih.gov/pmc/articles/pmc2699469/

Wexler-Cohen, Y., & Shai, Y. (2009). Membrane-anchored HIV-1 N-heptad repeat peptides are highly potent cell fusion inhibitors via an altered mode of action. *PLoS Pathogens*, *5*(7), e1000509.

White, C. B., Chen, Q., Kenyon, G. L., & Babbitt, P. C. (1995). A novel activity of OmpT. Proteolysis under extreme denaturing conditions. *The Journal of Biological Chemistry*, *270*(22), 12990–12994.

Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R. B., & Christianson, D. W. (2002). Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15375–15380.

Windram, O. P. F., Rodrigues, R. T. L., Lee, S., Haines, M., & Bayer, T. S. (2017). Engineering microbial phenotypes through rewiring of genetic networks. *Nucleic Acids Research*, *45*(8), 4984–4993.

Wise, M. L., Savage, T. J., Katahira, E., & Croteau, R. (1998). Monoterpene synthases from common sage (Salvia officinalis). cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase. *The Journal of Biological Chemistry*, *273*(24), 14891–14899.

Wohlever, M. L., Nager, A. R., Baker, T. A., & Sauer, R. T. (2013). Engineering fluorescent protein substrates for the AAA+ Lon protease. *Protein Engineering, Design & Selection: PEDS*, *26*(4), 299–305.

Wolfrum, C., Shi, S., Jayaprakash, K. N., Jayaraman, M., Wang, G., Pandey, R. K., Rajeev, K. G., Nakayama, T., Charrise, K., Ndungo, E. M., Zimmermann, T., Koteliansky, V., Manoharan, M., & Stoffel, M. (2007). Mechanisms and optimization of in vivo delivery of lipophilic siRNAs. *Nature Biotechnology*, *25*(10), 1149–1157.

Wong, M. S., Wu, S., Causey, T. B., Bennett, G. N., & San, K.-Y. (2008). Reduction of acetate accumulation in Escherichia coli cultures for increased recombinant protein production. *Metabolic Engineering*, *10*(2), 97–108.

Wu, Z., Alexandratos, J., Ericksen, B., Lubkowski, J., Gallo, R. C., & Lu, W. (2004). Total chemical synthesis of N-myristoylated HIV-1 matrix protein p17: structural and mechanistic implications of p17 myristoylation. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(32), 11587–11592.

Xia, X.-X., Qian, Z.-G., Ki, C. S., Park, Y. H., Kaplan, D. L., & Lee, S. Y. (2010). Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14059–14063.

Xu, L., Li, S., Ren, C., Cai, Z., & Lin, Z. (2006). Heat-inducible autolytic vector for high-throughput screening. *BioTechniques*, *41*(3), 319–323.

Yamada, M., Makino, K., Amemura, M., Shinagawa, H., & Nakata, A. (1989). Regulation of the phosphate regulon of Escherichia coli: analysis of mutant phoB and phoR genes causing different phenotypes. *Journal of Bacteriology*, *171*(10), 5601–5606.

Yamamoto, A., Setoh, K., Murakami, M., Shironoshita, M., Kobayashi, T., Fujimoto, K., Okada, N., Fujita, T., & Muranishi, S. (2003). Enhanced transdermal delivery of phenylalanyl-glycine by chemical modification with various fatty acids. *International Journal of Pharmaceutics*, *250*(1), 119–128.

Yang, J., Sun, B., Huang, H., Jiang, Y., Diao, L., Chen, B., Xu, C., Wang, X., Liu, J., Jiang, W., & Yang, S. (2014). High-efficiency scarless genetic modification in Escherichia coli by using lambda red recombination and I-SceI cleavage. *Applied and Environmental Microbiology*, *80*(13), 3826–3834.

Yang, Z., Li, M., Chen, W., & Peters, A. (2013). The Smart Solution for DNA Removal in Biopharmaceutical Production by Benzonase Endonuclease. In *Journal of Applied Virology* (Vol. 2, Issue 1). https://doi.org/10.21092/jav.v2i1.26

Ye, Z., Lynch,M.D., Trahan, A.D., Rodriguez, D.L., Cooper, C.B. Bozdag, A. (2015). Compositions and methods for rapid and dynamic flux control using synthetic metabolic valves (United Kingdom Patent No. GB2528177). In *Patent* (No. GB2528177).

Yin, L., Chen, X., Vicini, P., Rup, B., & Hickling, T. P. (2015). Therapeutic outcomes, assessments, risk factors and mitigation efforts of immunogenicity of therapeutic protein products. *Cellular Immunology*, *295*(2), 118–126.

Youngquist, J. T., Schumacher, M. H., Rose, J. P., Raines, T. C., Politz, M. C., Copeland, M. F., & Pfleger, B. F. (2013). Production of medium chain length fatty alcohols from glucose in Escherichia coli. *Metabolic Engineering*, *20*, 177–186.

Yuan, L., Voelker, T. A., & Hawkins, D. J. (1995). Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(23), 10639–10643.

Yuan, L., Wang, J., & Shen, W. C. (2005). Reversible lipidization prolongs the pharmacological effect, plasma duration, and liver retention of octreotide. *Pharmaceutical Research*, *22*(2), 220–227.

Yuan, L., Wang, J., & Shen, W. C. (2008). Reversible lipidization of somatostatin analogues for the liver targeting. *European Journal of Pharmaceutics and Biopharmaceutics: Official Journal of Arbeitsgemeinschaft Fur Pharmazeutische Verfahrenstechnik e.V*, *70*(2), 615–620.

Zeng, W., Eriksson, E. M., Lew, A., & Jackson, D. C. (2011). Lipidation of intact proteins produces highly immunogenic vaccine candidates. *Molecular Immunology*, *48*(4), 490–496.

Zhang, C., Desai, R., Perez-Luna, V., & Karuri, N. (2014). PEGylation of lysine residues improves the proteolytic stability of fibronectin while retaining biological activity. *Biotechnology Journal*, *9*(8), 1033–1043.

Zhang, H., Chu, W., Sun, J., Liu, Z., Huang, W.-C., Xue, C., & Mao, X. (2019). A novel autolysis system for extracellular production and direct immobilization of a phospholipase D fused with cellulose binding domain. *BMC Biotechnology*, *19*(1), 29.

Zhang, L., & Bulaj, G. (2012). Converting peptides into drug leads by lipidation. *Current Medicinal Chemistry*, *19*(11), 1602–1618.

Zhang, P., Sun, F., Liu, S., & Jiang, S. (2016). Anti-PEG antibodies in the clinic: Current issues and beyond PEGylation. *Journal of Controlled Release: Official Journal of the Controlled Release Society*, *2016/07/03*. https://doi.org/10.1016/j.jconrel.2016.06.040

Zhao, L., Ji, P., Li, Z., Roy, P., & Sahajwalla, C. G. (2013). The antibody drug absorption following subcutaneous or intramuscular administration and its mathematical description by coupling physiologically based absorption process with the conventional compartment pharmacokinetic model. *Journal of Clinical Pharmacology*, *53*(3), 314–325.

Zhou, Y., Liu, P., Gan, Y., Sandoval, W., Katakam, A. K., Reichelt, M., Rangell, L., & Reilly, D. (2016). Enhancing full-length antibody production by signal peptide engineering. *Microbial Cell Factories*, *15*, 47.