Improving the robustness of engineered bacteria to nutrient stress using programmed proteolysis

Klara Szydlo[1], Zoya Ignatova[1,*], Thomas E. Gorochowski[2,*]

[1] Institute of Biochemistry and Molecular Biology, University of Hamburg, Hamburg, Germany

[2] School of Biological Sciences, University of Bristol, Bristol, United Kingdom

* Correspondence should be addressed to T.E.G. (thomas.gorochowski@bristol.ac.uk) and Z.I. (zoya.ignatova@uni-hamburg.de)

Keywords: proteolysis, protein degradation, genetic circuit, resource recycling, burden

Abstract

The use of short peptide tags in synthetic genetic circuits allows for the tuning of gene expression dynamics and the freeing of amino acid resources through targeted protein degradation. Here, we use elements of the *Escherichia coli* and *Mesoplasma florum* transfer-messenger RNA (tmRNA) ribosome rescue systems to compare endogenous and foreign proteolysis systems in *E. coli*. We characterize the performance and burden of each and show that while both greatly shorten the half-life of a tagged protein, the endogenous system is approximately seven times more efficient. Based on these results, we then show how proteolysis can be used to improve cellular robustness through targeted degradation of a reporter protein in auxotrophic strains, providing a limited secondary source of essential amino acids that help partially restore growth when nutrients become scarce. These findings provide avenues for controlling the functional lifetime of engineered cells once deployed and increasing their tolerance to fluctuations in nutrient availability.

Introduction

Prokaryotic protein degradation is an essential quality control mechanism in cells and plays a crucial role in eliminating damaged and/or non-functional proteins [1–3]. It is enabled by a network of ATP-dependent proteases and adaptors that recognize specific motifs in misfolded proteins, or degrons [4,5]. Protein degradation in bacteria is mediated by the prokaryotic transfer-messenger RNA (tmRNA) ribosome rescue system, whereby an SsrA peptide tag is added C-terminally to nascent polypeptides, targeting them for degradation by several endogenous proteases [6]. These include ClpXP, ClpAP, FtsH and Lon, with ClpXP and ClpAP being the most active in *Escherichia coli,* degrading over 90% of SsrA-tagged proteins [1,3,7]. The tagging of proteins for degradation has gained interest in the field of synthetic biology, as it allows for specific and controllable protein degradation, which has been used to modulate protein turnover rates, investigate protein function by reducing intracellular concentrations, and as a means to tune dynamic processes, e.g., the period of genetic oscillators [8–11].

The SsrA peptide-tag system is conserved across prokaryotic species, but the tags vary in their amino acid composition and length [8,12–14]. The *E. coli* SsrA tag is the most extensively characterized, and its last three amino acids, ‘LAA’, determine the tag strength and the rate of tagged protein degradation [8]. Variants of these critical residues such as ‘LVA’, ‘AAV’ and ‘ASV’ result in different degradation rates, with ‘LAA’ and ‘LVA’ rendering tagged-GFP much more unstable than the ‘AAV’ or ‘ASV’ variants [8]. The growing knowledge of *E. coli* proteases and their dependency on auxiliary adaptor proteins has also allowed for controllable modulation of protein half-lives and degradation [2,15,16]. For example, the degradation of proteins tagged with an *E. coli* tag variant ‘DAS’ is mediated by the induction of the SspB adaptor protein in *Bacillus subtilis* [14].

Using SsrA tags from distinct species offers another level of control over protein degradation. The simultaneous use of multiple tags in parallel supports the construction of more complex systems where degradation of multiple proteins can be independently controlled. Several SsrA tags from other species have been characterized [13,14,17], including that of *Mesoplasma florum* [12]. This is targeted by the efficient *M. florum* Lon protease that acts orthogonally to the endogenous *E. coli* system, making it possible to use both simultaneously in *E. coli* cells [12]. Previous studies have identified regions of the *M. florum* tag which are crucial for recognition by *E. coli* and *M. florum* proteases, leading to the development of variants of the *M. florum* tag through deletion of non-essential regions or replacement of residues with other amino acids [10,18]. Furthermore, the specificity of the endogenous *M. florum* Lon protease to the cognate *M. florum* SsrA tag has enabled the development of inducible orthogonal protein degradation systems in *E. coli* with diverse applications, including controlling synthetic circuits such as toggle switches [10–12,18].

Whilst targeted protein degradation has seen wide use in tuning the function of genetic circuits, much less attention has been placed on its use in a more native context. For example, using protein degradation to help recycle essential amino acid resources when nutrient stress occurs[19,20]. Although such capabilities are less important when cells are grown in the rich and carefully controlled conditions of the lab, when deploying an engineered system into more realistic real work environments like your gut or the soil, high variability in nutrient availability is inevitable[21–23]. Therefore, having systems to help buffer cells from these effects is an important area that warrants further research.

Here, we attempt to directly address this by exploring how endogenous and heterologous protein degradation systems can be used to manage reservoirs of amino acids that are locked up in stable proteins that can then be subsequently released when needed. We explore the suitability of endogenous and heterologous systems for implementing this type of system and show using auxotrophic strains how targeted release of amino acids from a reporter protein enables the partial recovery of growth when an essential amino acid becomes scarce. This proof-of-concept offers a starting point for developing new cellular chassis that are more robust to nutrient fluctuations, as well as opening avenues to constrain the functional "shelf-life" of a cell by providing an internal amino acid reservoir with a limited capacity.

Results

Assessing the proteolytic activities of E. coli and M. florum SsrA tags

To build our system, we began by comparing the activities of the *E. coli* and *M. florum* proteolysis systems by assembling genetic constructs where an *eGFP* (GFP) reporter gene was tagged with one of our two proteolysis tags. Specifically, we used the *E. coli* (Ec; AANDENYALAA) and *M. florum* (Mf; AANKNEENTNEVPTFMLNAGQANYAFA) SsrA tag sequences which were codon optimized for expression in *E. coli* (**Materials and Methods**) and fused to the C-terminus of *GFP* whose expression was under the control of an IPTG-inducible promoter (P_{lac}). In this way, GFP was synthesized bearing one of two peptide tags, targeting it for proteolytic degradation by each of our chosen systems (**Figure 1A**). Because the Mf tag is specifically recognized by its cognate Lon protease from *M. florum* (Mf-Lon) which is not present in *E. coli* [12], we also constructed a separate plasmid where a codon-optimized *lon* gene from *M. florum* [12] was expressed under the control of an arabinose-inducible promoter (P_{BAD}).

To assess the performance of the two tags, we expressed untagged GFP, GFP-Ec, or GFP-Mf alone and simultaneously with Mf-Lon in *E. coli* BL21(DE3) cells. We observed almost no fluorescence in cells expressing GFP-Ec compared to cells expressing untagged GFP (2.8% at 6 h), indicating that the *E. coli* tag was effective in targeting the tagged protein for

degradation by endogenous proteases. By contrast GFP-Mf, when expressed alone, saw reduced, though nevertheless substantial levels of GFP, suggesting that most, but not all, of this protein escaped the endogenous *E. coli* proteases (**Figure 1B**). As expected, further induction of Mf-Lon protease caused a 76% drop in GFP-Mf fluorescence, supporting the notion that the Mf tag is specifically recognized (**Figure 1B**). The fluorescence observed from cells expressing the untagged GFP remained largely the same upon induction of the Mf-Lon protease, further indicating the specificity of the protease for the Mf tag (**Figure S1**). It should be noted, that the fluorescence of cells expressing only GFP-Mf was lower than that of cells expressing untagged GFP. To assess whether this resulted from off-target degradation by endogenous proteases, or low levels of leaky Mf-Lon expression, we compared the fluorescence levels of cells co-transformed with GFP-Mf and Mf-Lon and cells transformed with only GFP-Mf (**Figure S2**). The expression levels of GFP-Mf when expressed alone in both cases were comparable, corroborating previous findings that off-target degradation by *E. coli* proteases does occur for the Mf-tag sequence, most likely due to residues homologous to the *E. coli* tag [10,18].

To further compare the efficiency of the Ec and Mf tags, we induced the expression of untagged GFP, GFP-Mf, or GFP-Ec and after 5 hours removed the inducer, having allowed the GFP to reach high levels within the cells and surpass its maturation time[24]. Consequently, we monitored the degradation rate of each GFP variant and calculated their half-lives (**Figure 1D**; **Materials and Methods**). The fluorescence levels of cells containing GFP-Ec remained very low throughout, indicating that even strong expression rates could not overcome the endogenous protein degradation. The half-lives of untagged GFP, GFP-Ec and GFP-Mf were calculated from this data to be 718 min, 18 min and 129 min, respectively. These numbers support the high efficiency of the endogenous *E. coli* proteases (almost seven times faster than when using the *M. florum* system). However, the Mf-tag also caused a large increase in the turnover rate, exhibiting a half-life less than a fifth of the untagged GFP.

Dynamic and targeted control of protein degradation using the M. florum SsrA system

A potential advantage of using the *M. florum SsrA* tag system in *E. coli* for the recycling of amino acids is the ability to dynamically control its expression to coincide with an increased demand by the cell (e.g., during starvation conditions). This reduces the strength at which tagged proteins acting as a reservoir of amino acids need to be expressed as their turn-over rate can be kept low ensuring long-term stability. Such a method is not possible with the endogenous system as it is continually active. Therefore, stronger and continual expression of the tagged protein is necessary to maintain a similar sized pool.

To test whether induced expression of Mf-Lon would cause the rapid degradation of tagged proteins, we performed several time-course experiments where GFP-Mf expression

was induced at *t* = 0 and Mf-Lon was either not present, simultaneously induced with GFP-Mf, or induced 1 or 2 hours after GFP-Mf induction (**Figure 2**). We found that only simultaneously inducing Mf-Lon with GFP-Mf resulted in strong degradation of GFP-Mf, with a 32 % drop in fluorescence after 6 hours. This compared to drops of 4% and 3% when Mf-Lon was induced 1 and 2 hours after GFP-Mf, respectively.

This result was unexpected given that Mf-Lon has been shown to function efficiently in *E. coli*[10,18], but could have been due to the varying expression strengths of the GFP-Mf reporter and Mf-Lon protease, which reside on different plasmids and are driven by different promoters (**Figure 1**). To test if this might be the case, we carried out additional experiments where Mf-Lon expression was induced 2 hours before induction of GFP-Mf (**Figure 2**). This would provide sufficient time for Mf-Lon to reach a high concentration before expression of the reporter and allow us to assess the maximal degradation rate that could be achieved by our system. We found that the initial increase in fluorescence when Mf-Lon was induced simultaneously with GFP-Mf, was negated when Mf-Lon was induced 2 hours prior, suggesting that expression and maturation of Mf-Lon occurs quickly, and a higher level of GFP degradation could occur. Nevertheless, the rate of fluorescence increase from 3 h after GFP-Mf induction was almost identical (**Figure 2B**), indicating that the concentration of Mf-Lon achieved when expressed from a P_{BAD} promoter and low-copy plasmid (p15A origin; ~10 copies per cell) is unable to significantly impact the amount of GFP-Mf in the longer term, making this system unsuitable for rapid dynamic and inducible control of GFP degradation and recycling.

Recovering cell growth by controlled amino acid recycling

A major challenge when developing genetic circuits is managing the burden they place on shared cellular resources [25,26]. The expression of a genetic construct will sequester key cellular machinery like ribosomes and may exhaust amino acid supplies, which in turn can impact overall the cell physiology[27,28], alter translation dynamics[26] and/or trigger stress responses[29,30]. A reason for this large impact is that often circuit components are strongly expressed and highly stable, locking away a large portion of the cell's resources from endogenous processes. It has also been observed that supplementing cells expressing recombinant proteins with amino acids, enhanced cell growth and protein production[30]. Consequently, we hypothesized that by increasing amino acid turnover of heterologous protein products through targeted proteolysis, less stress would be imparted on the cell under these conditions.

To test this hypothesis, we measured the growth rate of cells expressing tagged and untagged GFP under the control of the same strong P_{lac} promoter (**Figure 1A**). We reasoned that the expression of the tagged GFP would place less of a burden on the host compared to

non-tagged version, due to increased recycling of amino acids [25,30,31]. We found that compared to untransformed cells, the expression of any GFP protein, both tagged and untagged, reduced the cell growth rate (**Figure 3**). However, the reduction in growth rate (within the first three hours of induction), was smaller for cells expressing GFP-Ec (41%), or GFP-Mf and Mf-Lon (37%) compared to cells expressing untagged GFP (51%). This suggests that while the overall expression of a recombinant protein will always cause a burden on the cell, this burden is partially alleviated by more effective recycling of amino acids so that they are more readily available for endogenous processes.

Next, we asked whether the potential benefit of using tagged proteins might be amplified when the host cell experienced nutrient related stress. It is known that protein degradation is elevated under various stresses, possibly as a way to increase the availability of amino acids for synthesis of stress-related proteins [19,32]. Furthermore, as part of the *E. coli* stringent response to nutrient limitations, there is an increase in the level of amino acid biosynthesis enzymes, to meet the demand for amino acids [33]. Therefore, we reasoned that increased recycling of a heterologous pool of proteins could benefit a host cell where nutrients to synthesize amino acids had become scarce in the environment.

To explore this idea, we developed a simple mathematical model to capture the key flows of a hypothetical essential resource in the cell (e.g. an amino acid) and its impact on cell growth (**Figure 4A**). The model consisted of three ordinary differential equations that track the concentrations of the shared resource that is either available for use within the cell (N_c), is actively in use by endogenous proteins (P_e), or is locked up in foreign heterologous proteins (P_f):

$$\frac{dN_c}{dt} = r_i N_e + r_r P_f - N_c(r_f + r_e + \mu), \quad (1)$$

$$\frac{dP_e}{dt} = r_e N_c - P_e \mu, \quad (2)$$

$$\frac{dP_f}{dt} = r_f N_c - r_r \mu. \quad (3)$$

Here, N_e is the external resource concentration outside the cell with a cellular import rate of r_i, r_e and r_f are the rates that available resources within the cell are converted into endogenous or heterologous proteins, respectively, and r_r is the recycling rate of the heterologous proteins (e.g., due to targeted proteolysis). Cellular growth and the associated dilution (by cell division) of all resource pools was captured by $\mu = 0.1P_e$. Parameters were chosen such that overall growth rate of the cell was consistent with *E. coli* data (i.e., having a division time ~25 min) and that relative internal transport, production and degradation rates were biologically realistic (**Materials and Methods**).

Using this model, we simulated cells expressing tagged and untagged proteins (**Figure 4B**), and cells exposed to several external environmental shifts to limit the resources available

(**Figure 4C**). In the first shift, we removed all resources from the environment at 500 min, and in the second, at the same time point, we applied an oscillating external nutrient concentration (for more details see **Materials and Methods**). In both cases, we compared cells not producing any heterologous protein (i.e., $r_f = 0$) to those producing a recombinant protein that is subsequently recycled for reuse by the cell. We then measured their response in terms of growth rate normalized to when the external nutrient was continually present (i.e., the steady state growth rate when $N_e = 1$). In both cases, the model showed a reduction in the relative impact to changes in environmental availability (**Figure 4C**), demonstrating the ability for a recycled internal reservoir of a heterologous resource to act as a backup source that can help buffer the cell temporarily from environmental change. It should be noted that inclusion of a heterologous resource pool and its recycling does have an impact on cellular growth rate. However, for some applications (e.g., excitable systems that are sensitive to even minor fluctuations in cellular behaviors), it may be preferable to have a more consistent performance when faced with environmental variability.

To test the model predictions, we used auxotrophic *E. coli* strains RF10[34] (Δ*lysA*) and ML17[35] (Δ*glnA*), which are unable to synthesize lysine and glutamine, respectively. This allowed us to tightly control endogenous amino acid levels by modulating the external supply in the media. Furthermore, lysine and glutamine are amongst the most abundant amino acids in our GPF reporter (8.4% and 6.7% of the total amino acid composition, respectively) offering suitable reservoirs of these key resources. We grew each of the strains expressing untagged GFP and GFP-Ec in nutrient-rich media to allow for a buildup of the recombinant protein. Following this, cells were switched to minimal medium, effectively removing the source of all external amino acids, and for our auxotrophic strains, completely removing access to lysine and glutamine, respectively. Agreeing with model predictions, we found that both strains expressing GFP-Ec exhibited a higher growth rate than cells expressing untagged GFP; 0.068 and 0.044 h^{-1} for GFP-Ec versus 0.044 and 0.022 h^{-1} for GFP-nt for the Δ*lysA* and Δ*glnA* strains, respectively (**Figure 5A-B**). We suspect the higher growth rates are attributed to the degradation of GFP-Ec, indicated by the lower fluorescence yields (**Figure 5C**). Addition of lysine or glutamine (7 mM) to the medium for the respective auxotrophic strains saw a marked increase in cell growth rate from 0.068 to 0.096 h^{-1} for the Δ*lysA* strain with lysine present, and from 0.022 to 0.09 h^{-1} for the Δ*glnA* strain with glutamine present when expressing untagged GFP (**Figure 5B**). This indicates that indeed glutamine and lysine were the major limiting factors for cell growth and that recycling of the internal heterologous protein reservoir was able to partially buffer this impact (25% and 24% recovery for Δ*lysA* and Δ*glnA*, respectively).

We observed an increased growth rate of 35% and 50%, for the Δ*lysA* and Δ*glnA* strains respectively, between cells expressing untagged GFP and those expressing GFP-Ec. The greater increase in growth rate for the Δ*glnA* strain, despite it being less abundant in GFP,

indicates that in this case the abundance levels of lysine and glutamine in GFP did not influence the increase in growth rate when cells were forced to use their internal amino acid supplies. Interestingly, we saw the Δ*glnA* strain grew slower in general than the Δ*lysA* strain (**Figure 5B**), which could be attributed to the fact that glutamine is more common in the *E. coli* proteome than lysine [36]. Therefore, a lack of endogenous glutamine would have a greater effect on cellular growth when external nutrients were limited, than a lack of endogenous lysine.

Together, these results show how targeted degradation of heterologous proteins can be beneficial to cells experiencing severe amino acid shortage and be used to buffer fluctuations in intracellular levels of amino acids and facilitating growth under amino acid limitations.

Discussion

In this work, we have characterized the performance of the endogenous *E. coli* proteolysis system and a similar heterologous system from *M. florum* within *E. coli* cells (**Figure 1**) [8,12]. When targeting identically expressed fluorescent reporter proteins, we found both systems were able to increase degradation rate. However, the endogenous system was approximately seven times more effective than the *M. florum* system, shortening the half-life of untagged GFP almost 40-fold. We also observed some crosstalk between these systems, with the reporter protein containing the *M. florum* tag also seeing increased degradation compared to an untagged reporter when the cognate Mf-Lon protease was not present. While characterization of these systems has been performed independently [8,10,11,18,37], we believe this study to be the first that directly compares these systems working on an identical target protein in the same host context.

In addition, we explored the option to switch on targeted degradation by inducing expression of the *M. florum* system dynamically over time. Unfortunately, we found that expression of the Mf-Lon protease in our system was unable to strongly degrade the targeted reporter, unless it was induced before the reporter was expressed. This was likely due to the expression rate of Mf-Lon being lower than for the reporter, due to the use of different plasmid backbones with different plasmid copy numbers, and thus the protease being unable to fully degrade the additional reporter being produced (**Figure 2**).

We also studied the effect of proteolysis tags in genetic circuits on the host cell, by assessing cell growth (**Figure 3**). As was previously observed, we found that a metabolic burden is placed on cells expressing recombinant proteins, whether tagged or untagged, indicated by a reduction in growth rate. Interestingly though, this effect was smaller in cells expressing tagged GFP than in cells expressing the untagged variant, providing evidence for

the benefits of using proteolysis tags when expressing recombinant proteins. This beneficial effect is likely due to increased amino acid recycling, freeing up these resources for use by endogenous processes connected to growth. Based on these findings, we developed a model that further illustrated these benefits, specifically under nutrient stress, where we reasoned that the benefits would be further increased due to external amino acid shortages. (**Figure 4**). Finally, by using auxotrophic strains, we were able to show that recycled heterologous proteins could act as a limited reservoir of amino acid resources, helping buffer the cell from fluctuations in nutrient availability and partially recover cell growth (**Figure 5**).

As our ambitions in synthetic biology grow and we begin to consider the construction of entire cellular systems, understanding how resources flow and are recycled in these systems will become crucial. Demonstrating the novel benefit of proteolysis tags as an amino acid recycling mechanism leads the way for further being able to control these nutrient fluxes within cells, and could aid in the construction of more complex synthetic systems, which include resource recycling. This work subsequently provides a perspective on the use of internal pools of heterologous proteins that can be released when needed to alleviate potential environmental nutrient fluctuations. Furthermore, this approach can be used to help buffer the cell and our engineered genetic systems from unavoidable variability that is present within real-world environments and paves the way for creating more reliable and robust host cell biosystems.

Materials and Methods

Bacterial strains, media, and cloning

The *E. coli* strain DH5α (*φ80dlacZ ΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17rK-mK+ phoA supE44 λ- thi-1*) was used for plasmid construction and cloning, and the strain BL21(DE3) (*F – ompT hsdSB (rB– mB–) gal dcm (DE3)*) used for characterisation of our genetic systems. Cells were grown in Luria-Bertani (LB) media (Roth, #X968.4), or minimal media (12.8 g/l $Na_2HPO_4.7H_2O$, 3 g/l KH_2PO_4, 1 g/l NH_4Cl, 2mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% glucose). 100 mg/ml ampicillin (Sigma Aldrich, #A9393), 50 mg/ml kanamycin (Sigma Aldrich #K1377), or 34 mg/ml chloramphenicol (Sigma Aldrich, #C0378) were used as selection markers for cloned plasmids. Enhanced green fluorescent protein (eGFP) in the pET16b vector under the IPTG-inducible *Lac* promoter system was C-terminally tagged with the *E. coli* (Ec) tag through site directed mutagenesis: overlap PCR primers were designed which contained the Ec tag sequence. These were phosphorylated and used for PCR with the plasmid backbone. The product was digested with *DpnI* (NEB, #R0176S) overnight, and the resulting product PCR purified. A ligation was carried out to circularise the vector, using 10-50 ng of DNA and T4 DNA ligase (Thermo fisher, #EL0011), according to the manufacturer's

instructions. The resulting plasmid was transformed into competent DH5α cells. The *M. florum* tag was codon optimized for expression in *E. coli* by selecting the most highly abundant codons in *E. coli* for the corresponding amino acids (codon sequence: GCT GCA AAC AAG AAC GAG GAA AAC ACC AAC GAA GTA CCG ACC TTC ATG CTG AAC GCA GGC CAG GCT AAC TAT GCA TTC GCA), and GFP was C-terminally tagged with the Mf tag using a digest-and-ligate approach: oligonucleotides were designed to contain the Mf tag sequence, and annealed to create double-stranded DNA fragments, then phosphorylated. The pET16b-eGFP vector was digested with fast digest *BsrgI* and *XhoI* (NEB, #R0102S and #R0146S) and used in a ligation reaction with the inserts (3:1 ratio) using T4 DNA ligase (Thermo Fisher, #EL0011), at 22 °C for 4-6 h. Competent DH5α cells were transformed with the resulting product. The *M. florum* Lon protease from the pBAD33 vector (a gift from Robert Sauer; Addgene plasmid #21867) was subcloned into the *pSB3C5* plasmid under the *araBAD* promoter using Golden Gate assembly. The primers for the PCR reaction were designed to flank the *Mf-Lon* with *BsmBI* restriction sites and include them into the vector (pSB3C5). The Golden Gate assembly reaction was set up, which included insert:vector in a 4:1 ratio, *BsmBI* (NEB, # R0739S), and 1µl T4 DNA Ligase (Thermo Fisher, #EL0011). The following conditions were used for the reaction: 60 cycles of 42°C for 3 min then 16°C for 4 min, followed by 50°C for 5 min and 80°C for 5 min. The resulting product was transformed into *E. coli* DH5α cells.

Proteolysis tag activity assays

Overnight cultures of BL21(DE3) cells transformed with pET16b-eGFP-no tag, pET16b-eGFP-Ec, or pET16b-eGFP-Mf and pSB3C5-mfLon were grown for 12-16 h at 37°C 250rpm, then re-suspended in minimal media with appropriate antibiotics for selection. The cultures were grown to an OD_{600} of 0.4–0.6, then induced with 0.5 mM IPTG (Roth, #2316.3) or 0.2% (w/v) arabinose (Roth, #5118.2). 1% glucose was added to the cultures expressing untagged proteins, to prevent basal expression from the pET16b vector[39]. For degradation assays, cells were pelleted 5 h after induction, washed twice in 1X phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10mM Na_2HPO_4, 1.8 mM KH_2PO_4) then re-suspended in minimal medium containing the relevant antibiotics, without IPTG. In cultures co-transformed with two plasmids, IPTG induction was stopped, but the second inducer, 0.2% (w/v) arabinose, was added to the medium to induce expression of Mf-Lon. 200 µl of the cultures were grown in a 96-well flat-bottom black plate with clear bottom (Corning, Sigma Aldrich #CLS3603-48EA) at 37°C with orbital shaking in a multimode microplate reader (Tecan Spark). Optical density (at 600 nm) and fluorescence measurements (excitation and emission wavelengths of 472 nm and 517 nm, respectively, with a gain of 50) were recorded at discrete intervals. Fluorescence was normalised to the OD_{600} value (a.f.u./OD_{600}). Untransformed BL21(DE3) cells were used as a negative control and their normalised autofluorescence values (a.f.u/OD_{600}) were

subtracted from the normalised fluorescence values (a.f.u./OD_{600}) of the cells in different conditions.

Auxotrophic strain and starvation assays

The RF10 (Δ*lysA*) and ML17 strains (Δ*glnA*) (a gift from Robert Gennis & Toshio Iwasaki Addgene plasmids #62076 and #61912) were transformed with the plasmids developed in this work and grown in LB to an OD_{600} of ~0.3. GFP expression was then induced with 0.5 mM IPTG for 1 h. After this, cells were pelleted, washed in 1X PBS, and re-suspended in minimal medium containing appropriate antibiotics for plasmid selection with additional 0.5 mM IPTG to maintain GFP expression. Additionally, 7 mM of lysine (Sigma Aldrich, #L5501) or glutamine (Serva, #22942) was added to positive control samples. The OD_{600} value and fluorescence were then measured as described above using a microplate reader every 10 min over 12 h.

Data analysis

Python version 3.9.5 and packages matplotlib version 3.3.2, NumPy version 1.19.2, and SciPy version 0.13 were used to fit the degradation data to a first order decay function of the form, $N(t) = 100e - \lambda t$, where $N(t)$ is the percentage of remaining fluorescence at time t post the halt of GFP production, and λ the decay constant. The half-live of GFP was then given by $t_{1/2} = \ln(2)/\lambda$. When investigating the dynamics of the Mf-tag system, the rate of GFP production (F) was calculated as the gradient of fluorescence values normalized to OD_{600} between 3 and 7 h after induction (**Figure 2A**). To obtain values for the growth rate of cells expressing tagged or untagged GFP, the slope of a linear fit to the growth curve (OD_{600} measurements) was calculated between 1 - 4 h (**Figure 3**). The growth rate of auxotrophic strains was calculated in the same way, between 5 and 12 h (**Figure 5B**). To compare whether the growth rates of the auxotrophic strains were statistically significantly different, a 2-sample t test was used. $P<0.05$ was considered as statistically significant. The statistical analysis was performed, and all plots and slopes of best fit were generated, using OriginLab Pro software (2019 version 64 bit).

Model parameterization and simulation

Parameters for the model of resource allocation and use were selected based on the assumption that an external resource concentration of $N_e = 1$ would lead to a realistic cell doubling time (~25 min) and that internal cellular rates would have biologically feasible relative values. This resulted in simulations with foreign protein recycling present being simulated with parameters: $r_i = 0.015$, $r_e = 0.02$, $r_f = 0.2$, $r_r = 0.01$, and with $\mu = 0.1 \times P_e$. In all simulations, initial conditions for all states were set to 0, and the dynamics simulated for 500 min with $N_e = 1.0$ for the system to reach a steady state before any environmental fluctuations occurred. The

model was simulated using Python version 3.9 and the SciPy version 0.13. The code for all simulations is available as **Supplementary Data 1**.

Supporting Information

The specificity of the Lon protease for the Mf tag (**Figure S1**); Off-target degradation by *E. coli* proteases on the Mf tag (**Figure S2**). The effects of Ec and Mf proteolysis tags on cell fluorescence and growth (**Figure S3**)

Acknowledgements

We thank Irem Avcilar-Kucukgoze for creating the original GFP-expressing plasmid, Robert Sauer for providing us with the pBAD33-mf-lon plasmid, and Robert Gennis and Toshio Iwasaki for providing us with the ML17 and RF10 *E. coli* strains. This work was supported by European Union's Horizon 2020 research and innovation program as part of the SynCrop ETN under the Marie Skłodowska-Curie grant 764591 (Z.I.), BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre grant BB/L01386X/1 (T.E.G.) and a Royal Society University Research Fellowship grant UF160357 (T.E.G.)

Author contributions

Z.I. and T.E.G. conceived the study. K.S. performed experiments and analyses. T.E.G. developed the mathematical model. Z.I. and T.E.G. supervised the work and discussed the data. All authors contributed to the writing of the manuscript.

Conflict of interest statement

None declared.

References

1. Keiler, K. C., Waller, P. R. H. & Sauer, R. T. (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science.* **271**(5251), 990–993
2. Farrell, C. M., Grossman, A. D. & Sauer, R. T. (2005) Cytoplasmic degradation of ssrA-tagged proteins. *Mol. Microbiol.* **57**(6), 1750–1761 3.
3. Maurizi, M. R. (1992) Production of abnormal proteins in *E. coli* stimulates transcription of lon and other heat shocks genes. *Trends biochem. Science.* **48**(10)
4. Gottesman, S. (2003) Proteolysis in Bacterial Regulatory Circuits. *Annual Review of Cell and Developmental Biology*. **19,** 565–587.
5. Chung, C. H. & Goldberg, A. L. (1981)The product of the lon (capR) gene in *Escherichia coli* is the ATP-dependent protease, protease La. *Proc. Natl. Acad. Sci. U. S. A.* **78**(81), 4931–4935
6. Moore, S. D. & Sauer, R. T. (2007) The tmRNA system for translational surveillance and ribosome rescue. *Annual Review of Biochemistry*. **76,** 101–124.
7. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes Dev. **12**, 1338-1347
8. Andersen, J. B. *et al.* (1998) New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria. *Applied and Environmental Microbiology.* **64**(6), 2240-2246
9. Stricker, J. *et al.* (2008) A fast, robust and tunable synthetic gene oscillator. *Nature* **456**(7221), 516–519
10. Cameron, D. E. & Collins, J. J. (2014) Tunable protein degradation in bacteria. *Nat. Biotechnol.* **32**(12), 1276–1281.
11. Chan, C. T. Y., Lee, J. W., Cameron, D. E., Bashor, C. J. & Collins, J. J. (2016) 'Deadman' and 'Passcode' microbial kill switches for bacterial containment. *Nature Chemical Biology*. **12**, 82–86
12. Gur, E. & Sauer, R. T. (2008) Evolution of the ssrA degradation tag in Mycoplasma: Specificity switch to a different protease. *Proc. Natl. Acad. Sci. U. S. A.* **105**(42), 16113–16118.
13. Ge, Z. & Karzai, A. W. (2009) Co-evolution of multipartite interactions between an extended tmRNA tag and a robust Lon protease in Mycoplasma. *Mol. Microbiol.* **74**(5), 1083–1099.
14. Griffith, K. L. & Grossman, A. D. (2008) Inducible protein degradation in Bacillus subtilis using heterologous peptide tags and adaptor proteins to target substrates to

the protease ClpXP. *Mol. Microbiol.* **70**(4), 1012–1025.

15. Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T. & Baker, T. A. (2000) Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol. Cell* **5**(4), 639–648.
16. McGinness, K. E., Baker, T. A. & Sauer, R. T. (2006) Engineering Controllable Protein Degradation. *Mol. Cell* **22**(5), 701–707.
17. Ahlawat, S. & Morrison, D. A. (2009) ClpXP degrades SsrA-tagged proteins in streptococcus pneumoniae. *J. Bacteriol.* **191**(8), 2894–2898.
18. Lv, L., Wu, Y., Zhao, G. & Qi, H. (2019) Improvement in the Orthogonal Protein Degradation in Escherichia coli by Truncated mf-ssrA Tag.*Trans. Tianjin Univ.* **25**(4), 357–363.
19. Jozefczuk, S. *et al.* (2010) Metabolomic and transcriptomic stress response of Escherichia coli. *Mol. Syst. Biol.* **6**(364), 1–16.
20. Mahmoud, S. A. & Chien, P. (2018) Regulated Proteolysis in Bacteria. *Annual Review of Biochemistry*. **87** 677–696.
21. Kotula, J. W. *et al.* (2014) Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci. U. S. A.* **111**(13), 4838–4843.
22. Saltepe, B., Kehribar, E. Ş., Su Yirmibeşoğlu, S. S. & Şafak Şeker, U. Ö. (2018) Cellular Biosensors with Engineered Genetic Circuits. *ACS Sensors* **3**(1), 13–26.
23. Corbisier, P. *et al.* (1999) Whole cell- and protein-based biosensors for the detection of bioavailable heavy metals in environmental samples. *Analytica Chimica Acta*. **387**(3) 235–244.
24. Macdonald, P. J., Chen, Y. & Mueller, J. D. (2012) Chromophore maturation and fluorescence fluctuation spectroscopy of fluorescent proteins in a cell-free expression system. *Anal Biochem* **421**(1), 291–298.
25. Glick, B. R. (1995) Metabolic load and heterologous gene expression. *Biotechnology Advances*. **13**(2), 247–261.
26. Gorochowski, T. E., Avcilar-Kucukgoze, I., Bovenberg, R. A. L., Roubos, J. A. & Ignatova, Z. (2016) A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes. *ACS Synth. Biol.* **5**(7), 710–720.
27. Ceroni, F., Algar, R., Stan, G. B. & Ellis, T. (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **12**(5), 415–418.
28. Kurland, C. G. & Dong, H. (1996) Bacterial growth inhibition by overproduction of protein. *Molecular Microbiology*. **21**(1) 1–4.
29. Ceroni, F. *et al.* (2018) Burden-driven feedback control of gene expression. *Nat.*

475 *Methods* **15**(5), 387–393.
476 30. Kumar, J., Chauhan, A. S., Shah, R. L., Gupta, J. A. & Rathore, A. S. (2020) Amino
477 acid supplementation for enhancing recombinant protein production in E . coli.
478 *Biotechnology and Bioenginering*. **117**(8), 2420–2433
479 31. Muhamadali, H. *et al.* (2016) Molecular BioSystems Interfacing chemical biology with
480 the-omic sciences and systems biology Metabolomic analysis of riboswitch containing
481 E. coli recombinant expression system. *Mol. BioSyst* **12**, 350.
482 32. Sanchez-Vazquez, P., Dewey, C. N., Kitten, N., Ross, W. & Gourse, R. L. (2019)
483 Genome-wide effects on Escherichia coli transcription from ppGpp binding to its two
484 sites on RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **116**(17), 8310–8319.
485 33. Irving, S. E., Choudhury, N. R. & Corrigan, R. M. (2021) The stringent response and
486 physiological roles of (pp)pGpp in bacteria. *Nature Reviews Microbiology*. **19**(4), 256–
487 271.
488 34. Lin, M. T. *et al.* (2015) Escherichia coli Auxotroph Host Strains for Amino Acid-
489 Selective Isotope Labeling of Recombinant Proteins. *Methods in Enzymology*. **565**,
490 45–66.
491 35. Lin, M. T. *et al.* (2011) A rapid and robust method for selective isotope labeling of
492 proteins. *Methods* **55**(4), 370–378.
493 36. Spahr, P. F. (1962) Amino acid composition of ribosomes from Escherichia Coli. *J.*
494 *Mol. Biol.* **4**(5), 395–406.
495 37. Butzin, N. C. & Mather, W. H. (2018) Crosstalk between Diverse Synthetic Protein
496 Degradation Tags in Escherichia coli. *ACS Synth. Biol.* **7**(1), 54–62.
497 38. Weiße, A. Y., Oyarzún, D. A., Danos, V. & Swain, P. S. (2015) Mechanistic links
498 between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. U. S.*
499 *A.* **112**(9), 1038–1047.
500 39. Grossman, T. H., Kawasaki, E. S., Punreddy, S. R. & Osburne, M. S. (1998)
501 Spontaneous cAMP-dependent derepression of gene expression in stationary phase
502 plays a role in recombinant expression instability. *Gene* **209**(1-2), 95–103.
503

Figures and captions

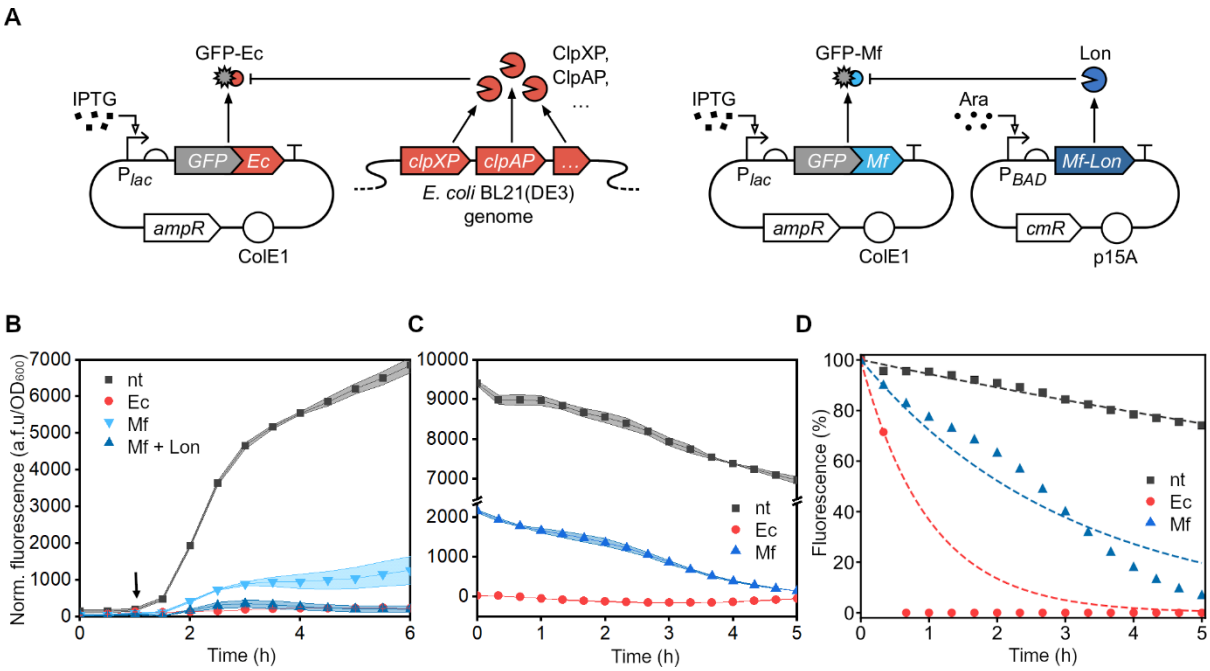

Figure 1: *E. coli* and *M. florum* proteolysis systems used for targeted protein degradation in *E. coli*. (**A**) Schematic of the proteolysis systems. GFP is expressed with Ec or Mf SsrA tags, which mark it for degradation by endogenous proteases, or the orthogonal plasmid-borne Mf-Lon protease, respectively. (**B**) GFP fluorescence normalized to cell density of *E. coli* BL21(DE3) cells expressing non-tagged GFP (nt), GFP-Ec (Ec) or GFP-Mf without and with the co-expression of Mf-Lon (Mf and Mf + Lon, respectively). Arrow indicates timepoint of GFP induction. (**C**) GFP fluorescence normalized to cell density of cells expressing untagged GFP (nt), GFP-Ec (Ec), or GFP-Mf (Mf) after removal of inducer, whilst maintaining Mf-Lon expression in the case of GFP-Mf. (**D**) % Fluorescence normalized to the time of removal of the inducer of cells expressing untagged GFP (nt), GFP-Ec (Ec), or GFP-Mf (Mf). Curves are fitted to first order exponential decay. Data are means ± SD (n = 3 independent biological replicates).

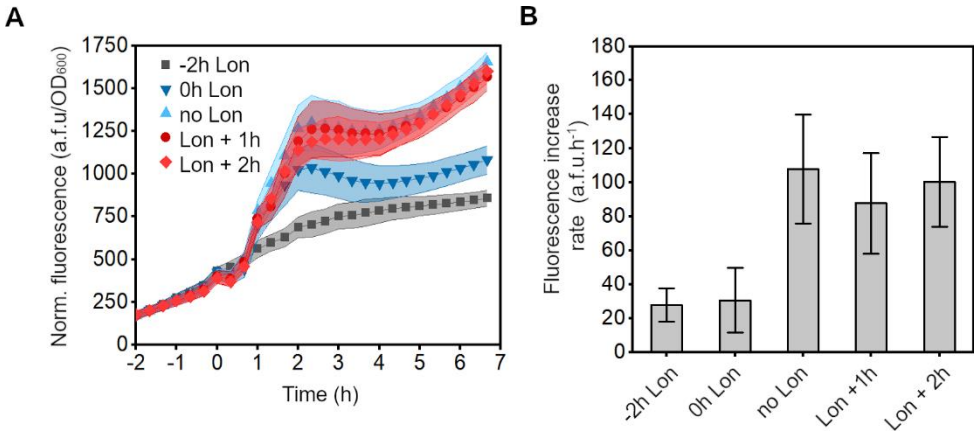

Figure 2: The dynamics of the *M. florum* proteolysis system. (**A**) GFP fluorescence normalized to cell density of cells expressing GFP-Mf without/with Mf-Lon induced at different time points. GFP-Mf expression was induced with 0.5mM IPTG at t = 0. Mf-Lon expression was induced 2 h before, simultaneously, or 1 or 2 h after GFP-Mf induction, with 0.2%(w/v) arabinose. Data are means ± SD (**B**) Fluorescence increase rate of cells expressing GFP-Mf without/with Mf-Lon induced at different time points in relation to GFP-Mf induction. Data are means ± SE (*n* = 5 independent biological replicates).

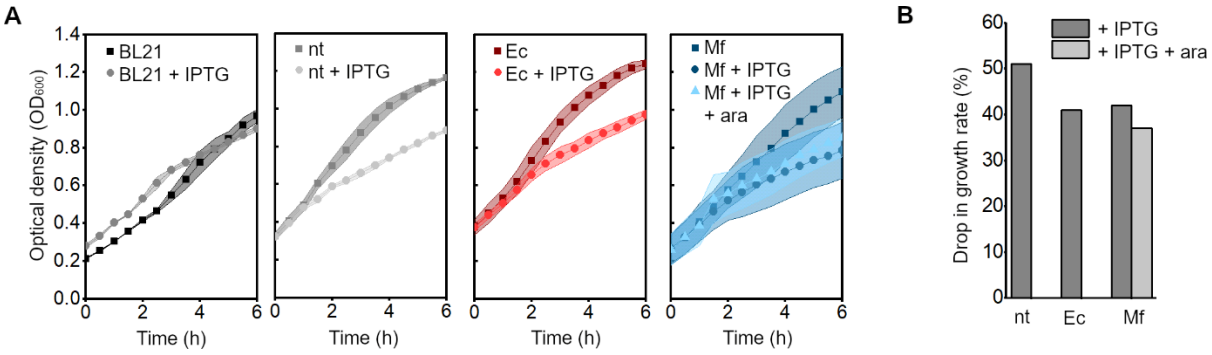

Figure 3: Recombinant protein expression has a negative effect on cell growth. (A) The effect of IPTG inducer (0.5mM) on the growth of untransformed BL21 cells, cells transformed with pet16b-GFP (nt), cells transformed with pet16b-GFP-Ec (Ec), or the combined effect of IPTG (0.5 mM) and arabinose (0.2 %(w/v)) on cells co-transformed with pet16b-GFP-Mf and pSB3C5-Mf-Lon. Cells were induced at 1 h. **(E)** Drop in growth rate (%), between 1 and 4 h, of cells induced to express either GFP-nt, GFP-Ec, GFP-Mf (+ IPTG) or GFP-Mf and Mf-Lon (+ IPTG + ara), compared to uninduced cells. Data are means ± SD (*n*=3 independent biological replicates).

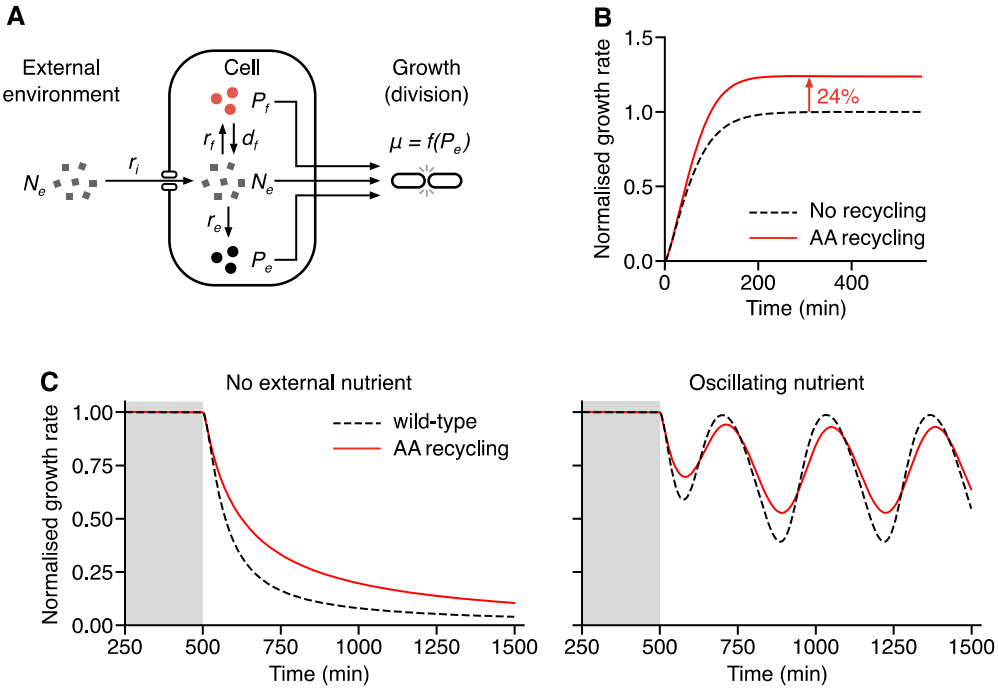

Figure 4: Model capturing the benefits of amino acid recycling. (**A**) Model overview. N_e denotes the external resource concentration, N_c, P_e and P_f denote the concentration of a key resource (ie. amino acid), available within the cell, locked up in endogenous or in heterologous proteins, respectively. r_i denotes the cellular import rate of resources, which can be divided into r_e, the rate at which resources are converted into endogenous proteins, and r_f, the rate at which resources are converted into foreign recombinant proteins. $\mu = f(P_e)$ captures cell growth and dilution of resources by cell division. (**B**) Simulation of the normalized cell growth rate in a strain expressing recombinant proteins, i.e. with no amino acid recycling, and in a strain expressing tagged foreign proteins, i.e. with amino acid recycling. (**C**) Simulation of the effects of nutrient stress on normalized cell growth rate expressing tagged proteins, i.e. with amino acid recycling (AA recycling), and in a strain expressing no heterologous protein (wild type). The external resource (nutrient) is continually present for the first 500 min, then either removed completely, or oscillating nutrient levels are applied after this time. In all cases, growth rate is normalized to the control growth with present external resource, i.e. $N_e = 1$.

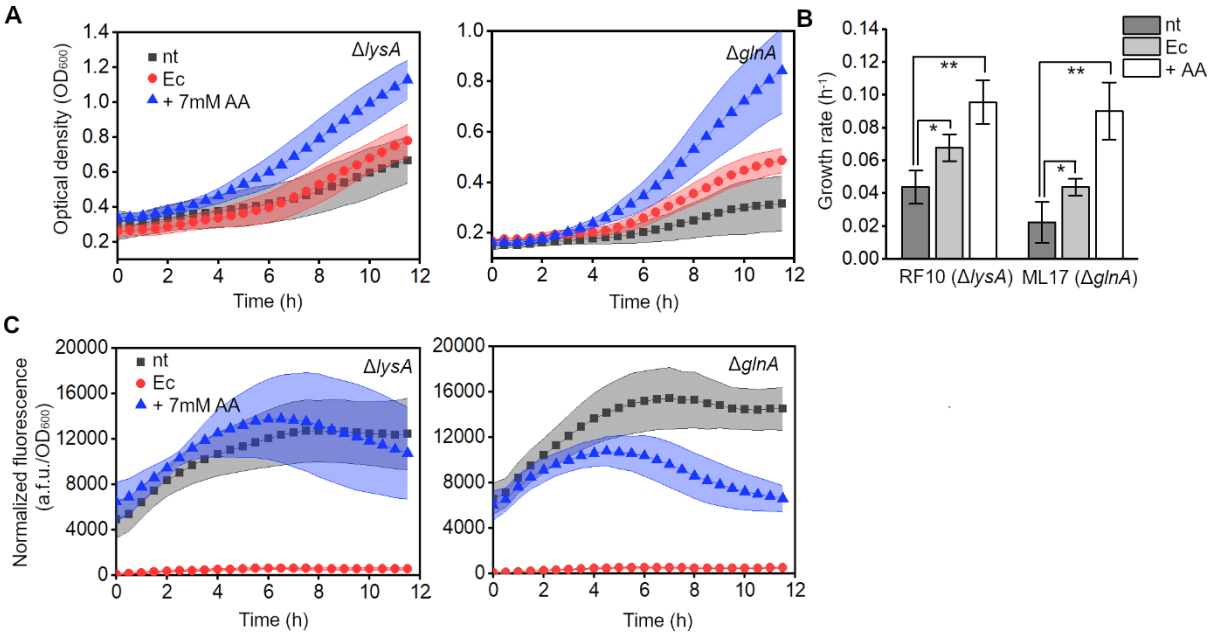

Figure 5: Targeted GFP degradation provides amino acids to auxotrophic strains upon nutrient limitation. (**A**) Growth of the RF10 (Δ*lysA*) and ML17 (Δ*glnA*) strains, expressing untagged GFP (nt) or GFP-Ec (Ec) with the addition of 7 mM lysine or glutamine supplement (+7mM AA). Data are means ± SD (**B**) Quantification of the growth rates, μ, of cells, between 5–12 h of growth. (*$p < 0.05$, **$p < 0.005$, as compared to nt condition for each strain, with 2-sample t test). Data are means ± SE (**C**) GFP fluorescence normalized to cell density of the RF10 (Δ*lysA*) and ML17 (Δ*glnA*) strains, expressing untagged GFP (nt) or GFP-Ec (Ec) with the addition of 7 mM lysine or glutamine supplement (+7mM AA). Data are means ± SD (*n* = 5 independent biological replicates).