

RESEARCH

Open Access



# Resource-aware whole-cell model of division of labour in a microbial consortium for complex-substrate degradation

Eliza Atkinson, Zoltan Tuza, Giansimone Perrino, Guy-Bart Stan<sup>\*†</sup> and Rodrigo Ledesma-Amaro<sup>\*†</sup>

## Abstract

**Background:** Low-cost sustainable feedstocks are essential for commercially viable biotechnologies. These feedstocks, often derived from plant or food waste, contain a multitude of different complex biomolecules which require multiple enzymes to hydrolyse and metabolise. Current standard biotechnology uses monocultures in which a single host expresses all the proteins required for the consolidated bioprocess. However, these hosts have limited capacity for expressing proteins before growth is impacted. This limitation may be overcome by utilising division of labour (DOL) in a consortium, where each member expresses a single protein of a longer degradation pathway.

**Results:** Here, we model a two-strain consortium, with one strain expressing an endohydrolase and a second strain expressing an exohydrolase, for cooperative degradation of a complex substrate. Our results suggest that there is a balance between increasing expression to enhance degradation versus the burden that higher expression causes. Once a threshold of burden is reached, the consortium will consistently perform better than an equivalent single-cell monoculture.

**Conclusions:** We demonstrate that resource-aware whole-cell models can be used to predict the benefits and limitations of using consortia systems to overcome burden. Our model predicts the region of expression where DOL would be beneficial for growth on starch, which will assist in making informed design choices for this, and other, complex-substrate degradation pathways.

**Keywords:** Microbial consortia, Division of labour, Resource-aware whole-cell modelling, Consolidated bioprocesses, Synthetic biology

## Background

### Agricultural waste as low-cost feedstock for biotechnology

Existing methods for biotechnology primarily use processed sugar feedstocks, which are high-cost and non-sustainable. Recycling agricultural waste for use as cheap and sustainable feedstocks for microbes will enable more cost-effective bioproduction, capable of competing with

traditional industrial practices. To achieve this requires organisms engineered to hydrolyse and metabolise the complex biomolecules found in plants, such as lignin, cellulose and starch. Many attempts have been made to engineer desirable production into naturally lignocellulolytic or saccharolytic hosts, or to engineer substrate hydrolysis into optimised industrial production hosts [1, 2]. However, both methods have failed to achieve the high yields needed for a viable commercial process in a single host monoculture [3, 4]. One of the primary limitations to a consolidated bioprocess in a single host has been attributed to the finite energy pool and high interference between the components and with the host [5–7].

<sup>†</sup>Guy-Bart Stan and Rodrigo Ledesma-Amaro contributed equally to this work

\*Correspondence: g.stan@imperial.ac.uk; r.ledesma-amaro@imperial.ac.uk

Department of Bioengineering and Imperial College Centre for Synthetic Biology, Imperial College London, London SW72AZ, UK



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

### Burden and division of labour

The energetic and resource cost of expressing heterologous proteins is known as gene expression burden. In the model organism *Escherichia coli* this burden is primarily attributed to the energetic cost of translation elongation and inefficient use of translational resources [8, 9]; for example, ribosomes wastefully sequestered on transcripts during ribosomal traffic jams. The result is a limited expression level of heterologous genes that can be achieved in a single cell before the burden causes decreased growth rates and reduced product yields [10, 11].

One method to reduce the burden of heterologous gene expression is division of labour (DOL) between members of a microbial consortium (Fig. 1) [12, 13]. Since fewer proteins need to be expressed in each cell, competition between heterologous and endogenous genes for shared cellular resources is decreased [14, 15]. DOL has been used for the production of a wide range of useful molecules, including anthocyanins [16], flavonoids [17], *n*-butanol [18], 3-amino-benzoic acid [19] and tryptamine [20].

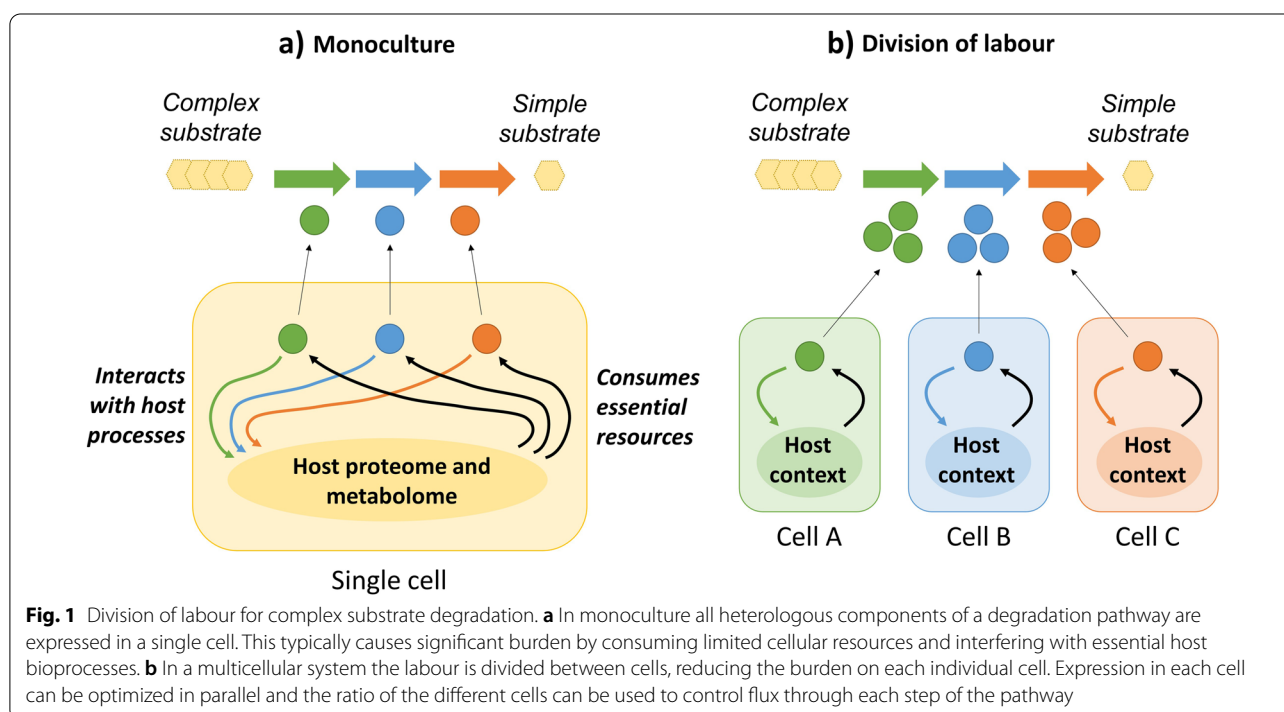
Importantly, DOL may assist in the design of consolidated bioprocesses. Efforts to engineer utilization of lignocellulosic biomass into desirable bioproduction hosts have been complicated by the number and complexity of the degradation enzymes required. DOL can overcome the limitations of these design efforts by reducing the burden imposed on each cell and enabling an alternative

means of optimising the expression of each gene in parallel. For example, Tsai, Goyal & Chen engineered four different strains of *Saccharomyces cerevisiae*, each expressing a different subunit of the cellulosome. Altering the ratio of the four strains in the consortium allowed them to optimise the processes of cellulose hydrolysis and ethanol production [4, 21, 22].

### DOL and computational modelling

Computational models can assist in the design of DOL strategies. A broad range of approaches are used for this, including genome-scale metabolic models (GEMs) [23] and agent-based models [24, 25]. Of interest here, are coarse-grained modelling approaches called host-aware models, which consider the relationship between heterologous protein expression and host cell growth. These models have been used to identify key factors enabling co-culture to outperform monoculture [26, 27]. For example, modelling has assisted in the design of a glucose-acetate cross-feeding consortium by identifying the range of burden within which the consortium is stable [27, 28]. Tsoi et al. [29] used a phenomenological approach to model 24 different pathway architectures for DOL, which allowed them to identify when burden is significantly limiting in the monoculture and when the cost of transporting intermediates becomes a limiting factor for consortia.

Most coarse-grained models of consortia use simple phenomenological relationships between increasing



expression and decreasing growth, treating the mechanisms behind this relationship as a black/grey box. In contrast, more mechanistic host-aware models for monocultures have been proposed which explicitly model the intracellular trade-offs and resource allocations that characterise burden. This includes the dynamics for shared cellular resource pools such as ribosomes [30], global cellular feedback to regulate protein production [31], and external regulation from varying environments [32]. Coarse-grained models which account for cellular resource-pathway interactions within the cell can predict non-intuitive effects of heterologous expression on the host. These resource-aware models may offer advantages over the current phenomenological models used for consortium growth predictions but have yet to be applied to contexts involving more than one host strain.

Weiß et al. published a host-aware whole-cell model for *E. coli* which links transcription and translation to the allocation of cellular energy, ribosomes, and the proteome [33]. This model has been widely used to predict burden for genetic circuits [34, 35]. Here, we adapt this model to predict the growth rates achievable by a simple consortium designed to perform complex-substrate degradation. Many complex polysaccharide substrates found in plants, such as starch, cellulose, xylan or inulin, require two enzymes for efficient degradation: an endohydrolase to cleave bonds within the molecule into smaller chunks and an exohydrolase to release sugars from the end of polysaccharide chains. This simple pathway can be expressed with two genes in a single cell or in a two-strain consortium where each strain expresses one of the genes. Here we use the example of starch hydrolysis to investigate a consortium of two strains of *E. coli*, expressing separately endoamylase and exoamylase. Cultured together, these strains should be able to grow on starch as the sole-carbon source. This model allows exploration of the impact of burden on this two-strain consortium and comparison with a monoculture co-expressing both hydrolytic enzymes simultaneously.

## Results

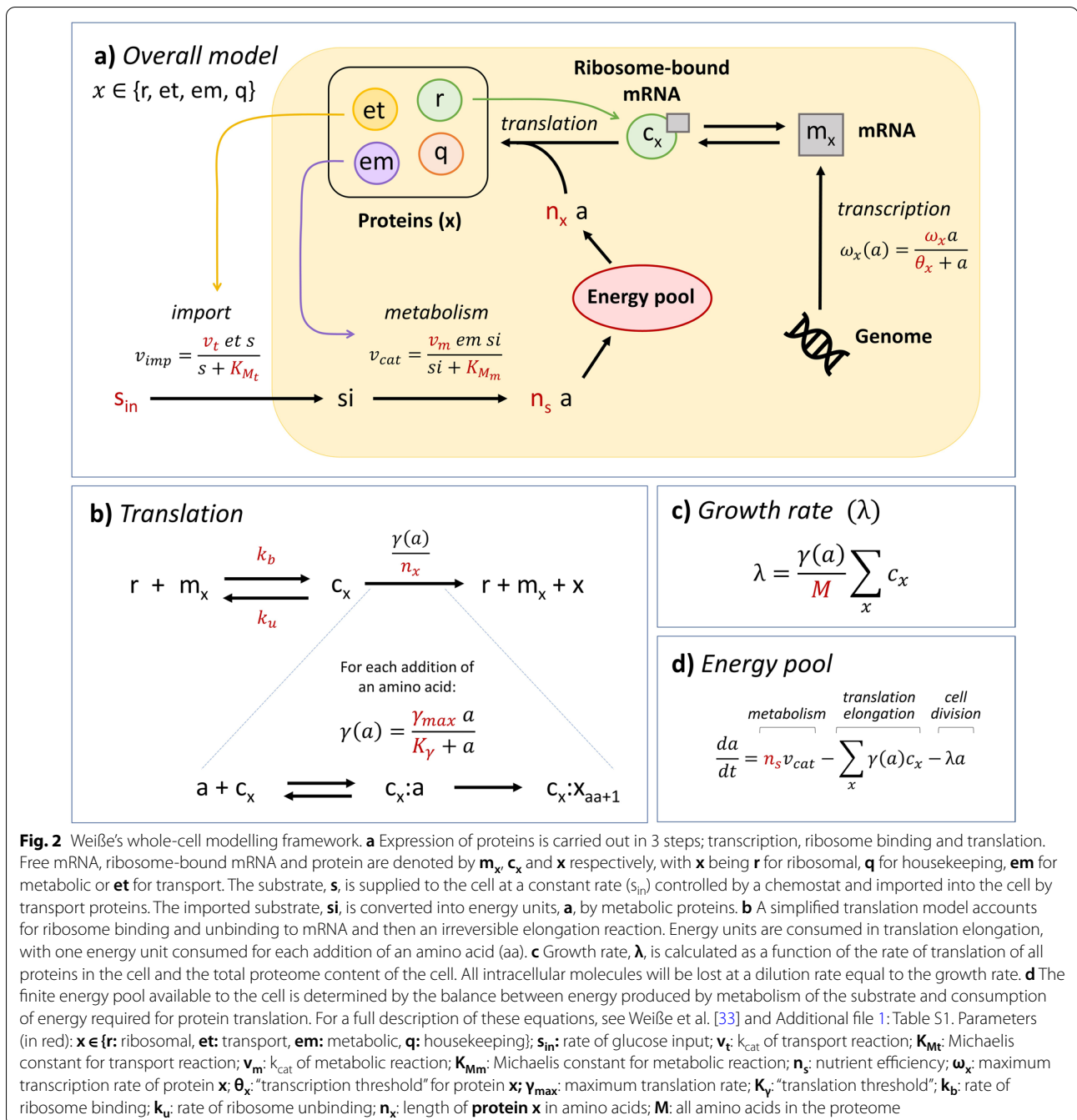
### A resource-aware whole-cell model for burden

Firstly, a whole-cell model for burden was adapted to predict cellular resource allocation and growth rate when expressing heterologous proteins in a DOL system. The ordinary differential equations (ODEs) describing this whole-cell model are derived from those published by Weiß et al. [33] (Fig. 2, Additional file 1: Table S1). The full details of the assumptions made and how the model equations are obtained can be found in Weiß et al. [33], but the key features of the model are:

1. The proteome is roughly divided into 4 categories, ribosomal proteins (r), transport proteins (et), metabolic proteins (em) and housekeeping proteins (q) as described in Fig. 2a. Each of these 4 categories of proteins have ODEs for 3 species of intracellular molecules; the mRNA ( $m_x$ ), the mRNA:ribosomal complex ( $c_x$ ) and the protein (x), where  $x \in \{r, et, em, q\}$ . It is assumed that all types of mRNA degrade at the same rate, and that the proteins are not degraded.
2. A further 2 ODEs model the imported substrate (si) and the energy molecules (a), equivalent to ATP. The primary substrate (s) is provided at a constant rate. Extracellular substrate is imported into the cell by transport proteins. The imported substrate is then converted into energy molecules by metabolic proteins. The number of energy units produced from each substrate is dependent on the nutrient efficiency ( $n_x$ ) (Fig. 2d).
3. The model assumes that translation accounts for the total energy use in the cell. To model translation, each free mRNA binds to a ribosome from the finite pool of free ribosomes, and each addition of an amino acid requires the use of one unit of energy (Fig. 2b).
4. At each time point, growth rate is calculated based on a set of linear growth laws described by Scott et al. [9] (Fig. 2c).

This model was adapted to a two-strain consortium; where one cell (Cell A) expresses enzyme A (ea) and the other (Cell B) expresses enzyme B (eb). For each cell type (A or B) an ODE model was created capturing the dynamics of the 14 intracellular molecules considered in Weiß's framework, plus 3 new ODEs capturing the dynamics of the mRNA, the mRNA:ribosome complex and the protein for each heterologous gene. The equations for ribosomes and growth rate were adjusted to account for the new heterologous proteins as outlined in Nikolados et al. [36]. To enable comparison with the monoculture case, a model was similarly built by adding 6 ODEs to the model of Weiß et al. to capture co-expression of both enzymes A (ea) and B (eb) in a single cell. The same parameters employed by Weiß et al. for an *E. coli* host cell were used [33].

An important consideration when adapting the model to the consortium was the necessity to provide substrate for both Cell A and Cell B. The rate of uptake of glucose from the environment into the cells,  $v_{uptake}(et, s)$ , was adjusted to account for the number of each cell type in the consortia, (see Eq. 1), or in the monoculture (see Eq. 2). For the purpose of this model, we assume that the populations of both cell types in the consortium can be tightly controlled. This could be achieved by a number of



different methods; by computationally-controlled addition of each cell type in a turbidostat, or by further engineering of the cells with autoinhibition or cross-feeding such that the populations will autoregulate themselves to remain at a stable ratio [37–40].

$$v_{uptake}(et, s) = \frac{v_{et} s (N_a et_{cella} + N_b et_{cellb})}{s + K_{met}} \quad (1)$$

$$v_{uptake}(et, s) = \frac{v_{et} s Net}{s + K_{met}} \quad (2)$$

Here  $N_a$  and  $N_b$  are the number of Cell A and Cell B respectively, and  $N$  is the number of cells in the monoculture population. For the case study presented here we assume that the two cell types in the consortium remain stable in equal ratio and that the populations can be normalised so the ratio of Cell A to Cell B is 1:1. For a fair

comparison between the consortium and monoculture cases, we assume there is an equal total number of cells in both cases. Therefore, if the value of  $N_a$  and  $N_b$  are each 1, an equivalent monoculture will have  $N$  equal to 2. In this framework,  $et$  is the number of transport proteins per cell, which may vary across Cell A and Cell B, while  $v_{et}$  is the  $k_{cat}$  of the transport reaction catalysed by all the transport proteins in the cell, and  $K_{met}$  is the Michaelis constant for the same transport reaction.

The amount of glucose available to the cells,  $s$ , will depend on the glucose input,  $s_{in}$ , the amount of glucose consumed by the all the cells in the population,  $v_{uptake}(et, s)$ , and the dilution rate,  $d_s$ , of the chemostat, modelled using Eq. 3:

$$\frac{ds}{dt} = s_{in} - v_{uptake}(et, s) - d_s s \quad (3)$$

All parameter values used in the model are summarised in Additional file 1: Table S2.

### DOL reduces resource redistribution to heterologous genes

The whole-cell model we briefly introduced above can be used to investigate burden when two heterologous proteins are expressed using DOL and compare the simulation results with those obtained from co-expressing all heterologous proteins in a single cell. Running a simulation with glucose as the primary substrate shows that DOL in a two-strain consortium composed of Cells A and B achieves faster growth rates than the monoculture case (Fig. 3a and b). As the maximum transcription rate ( $\omega$ ) of the expressed heterologous proteins increases, the growth rate decreases. In the monoculture, growth rate is thus dependent on the transcription rate of both enzyme A and enzyme B (Fig. 3c), highlighting competition for resources between heterologous proteins and host processes. In the consortia, however, the growth rate of each cell (Cell A or Cell B) depends on the transcription rate of the single heterologous protein it expresses.

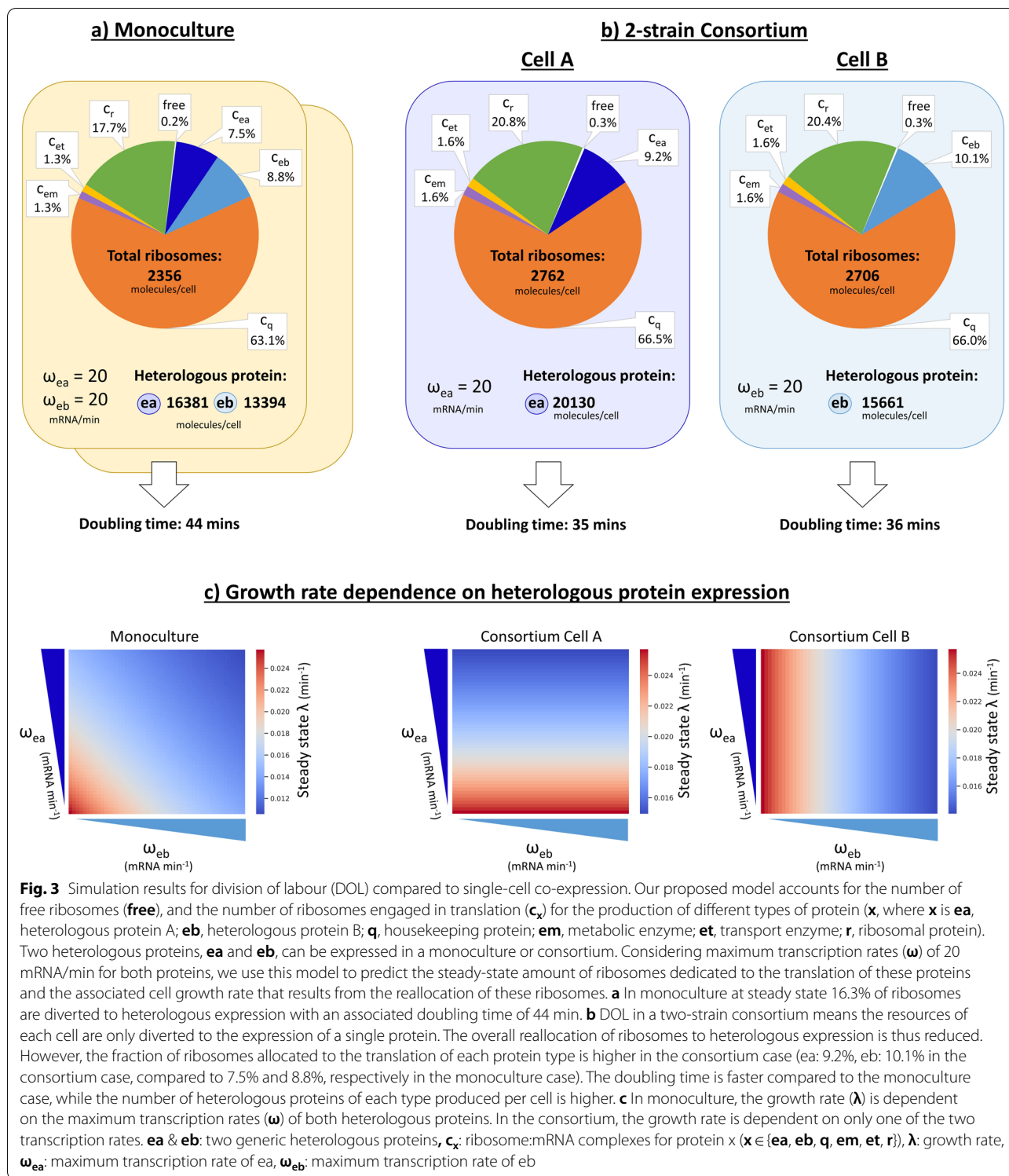
The model predicts that DOL will improve the amount of heterologous protein that can be produced per cell. Cell A produces 20,130 molecules/cell of enzyme A (7.95% of total protein in Cell A; Additional file 1: Table S3), while Cell B produces 15,661 molecules/cell of enzyme B (6.34% of total protein in Cell B). In contrast a single cell in the monoculture produces 16,381 enzyme A and 13,394 enzyme B (11.94% of total protein). While the total number of proteins in the Weisse model may be underestimated [41, 42], the values for heterologous gene expression as a percentage of total protein lie within an acceptable range as *E. coli* has been shown to tolerate up to 30% of their proteome dedicated to heterologous expression [43].

While in the consortia, each cell type, A or B, can produce more of its respective heterologous protein per cell than the monoculture, this may not result in higher heterologous protein yield overall. In a monoculture of 1000 cells, all 1000 will express both proteins. In a consortium with 1000 cells, 500 will express protein A, and 500 will express protein B. Assuming that each cell in these populations produces the amount of protein predicted by the whole-cell model; a 1000-cell monoculture will produce around 16 million enzyme A and 13 million enzyme B. While a 1000-cell consortium will only produce around 10 million enzyme A and 8 million enzyme B. Therefore, while the monoculture case has slower growth and produces fewer heterologous proteins per cell, it may produce more heterologous proteins in total than the equivalent consortia case.

### Starch-degradation

Starch was identified as a desirable complex substrate for the application of division of labour. Starch is the second most abundant source of carbon in plants after cellulose and a common by-product of the food industry. Starch granules are made of two polymers of glucose, amylose and amylopectin, which differ in their structure and glycosidic bonds. Efficient degradation of starch requires both endo- and exo- amylase activities to break down the two types of glycosidic bonds found in its structure.  $\alpha$ -amylase (EC 3.2.1.1) is an endo-amylase which binds randomly in the sugar chain and catalyses hydrolysis of  $\alpha$ -1,4 glycosidic bonds between glucose units, creating variable-length maltodextrins [44]. Glucoamylase (EC 3.2.1.3) is an exo-amylase which releases glucose from the non-reducing ends of a sugar chain by cleaving either  $\alpha$ -1,4 or  $\alpha$ -1,6 bonds [44]. These two enzymes are relatively well characterized for several different starch-sources and maltodextrins.

Mathematically modelling the degradation of starch by amylase can be complex when the parallel activity of both amylases and the stochasticity of bond cleavage are taken into account [45, 46]. Therefore this model utilized a simplified set of equations proposed by Fujii and Kawamura [47]. When starch is abundant, its degradation is modelled using a two-step process. The  $\alpha$ -amylase-dependent first reaction produces an intermediate which can act as a substrate for the glucoamylase-dependent second reaction. The reasoning for this schema is that at the beginning of starch hydrolysis there are relatively few non-reducing ends within the whole molecule that glucoamylase can act on, so the dominant reaction is the  $\alpha$ -amylase hydrolysis reaction. Similarly, once a chain is degraded to a certain length, or limited by branches,  $\alpha$ -amylase affinity for the substrate decreases. If a large proportion of starch is degraded to short chains, then the



$\alpha$ -amylase reaction becomes negligible and the glucoamylase rate parameters will change based on the average chain length of the polysaccharide, as described in Fujii and Kawamura [47]. Both  $\alpha$ -amylase and glucoamylase

show Michaelis–Menten kinetics in terms of the rate of hydrolysis. Equations 4 and 5 below show the rate of change for starch,  $s_0$ , and the intermediate,  $s_1$ , respectively. When starch is externally supplied as the primary

substrate, glucose,  $s$ , will vary based on Eq. 6, replacing the previously stated Eq. 3 which is only applicable when glucose is supplied as the primary substrate.

$$\frac{ds_0}{dt} = s_{0in} - N_a \frac{v_{ea} ea s_0}{s_0 + K_{m_{ea}}} - d_{s_0} s_0 \tag{4}$$

$$\frac{ds_1}{dt} = N_a \frac{v_{ea} ea s_0}{s_0 + K_{m_{ea}}} - N_b \frac{v_{eb} eb s_1}{s_1 + K_{m_{eb}}} - d_{s_1} s_1 \tag{5}$$

$$\frac{ds}{dt} = N_b \frac{v_{eb} eb s_1}{s_1 + K_{m_{eb}}} - v_{uptake}(et, s) - d_s s \tag{6}$$

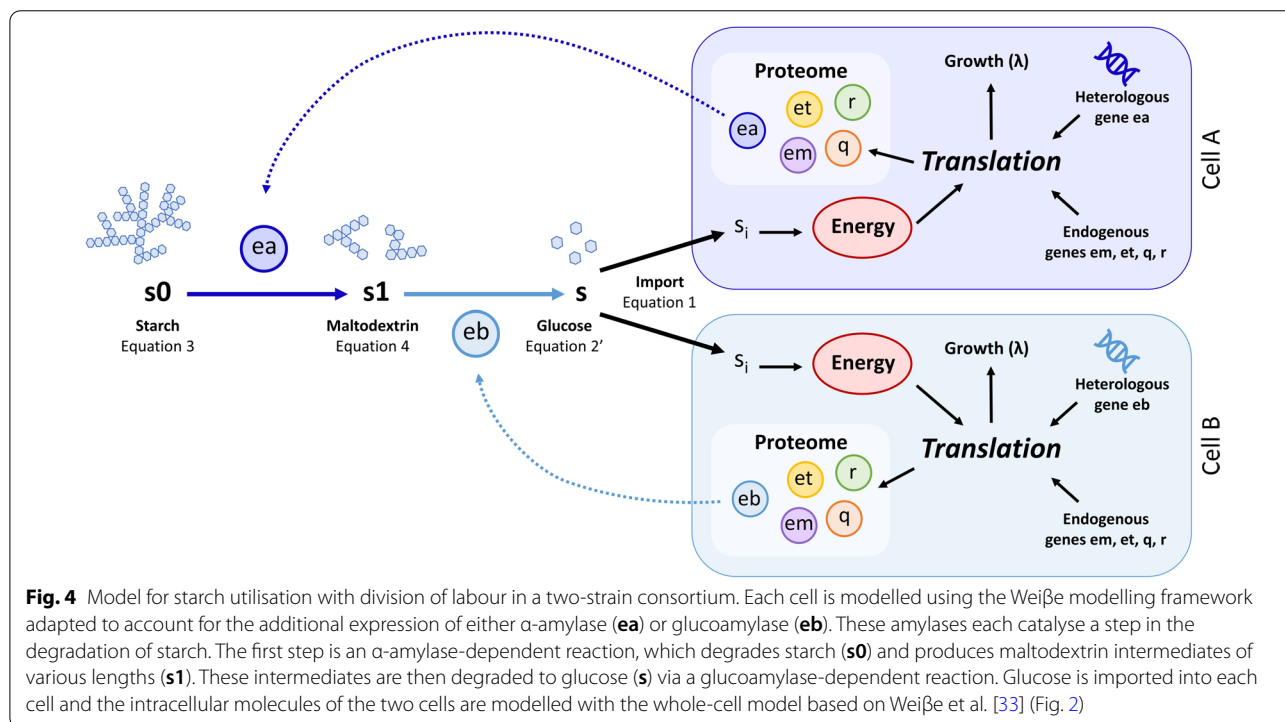
Here  $ea$  represents the concentration of  $\alpha$ -amylase, while  $eb$  denotes that of glucoamylase,  $v_{ea}$  and  $v_{eb}$  represent the  $k_{cat}$  rates for  $ea$  and  $eb$ , respectively, and  $K_{m_{ea}}$  and  $K_{m_{eb}}$  are the Michaelis constant for each enzyme. The rate of each step is also dependent on the number of cells expressing  $\alpha$ -amylase (i.e. the population of Cell A ( $N_a$ )) and the number of cells expressing glucoamylase (i.e. the population of cell B ( $N_b$ )). For a monoculture system all cells will express both, therefore  $N_a$  and  $N_b$  in Eqs. 4, 5 and 6 will be replaced with  $N$  (Additional file 1: Table S4). In this example, we assume that each population type can be tightly controlled and that the proportion of Cell A to Cell B can be robustly maintained at a 1:1 ratio. Under this assumption, without loss of generality, we consider  $N_a = 1$ ,  $N_b = 1$  and  $N = 2$ , so that the monoculture and

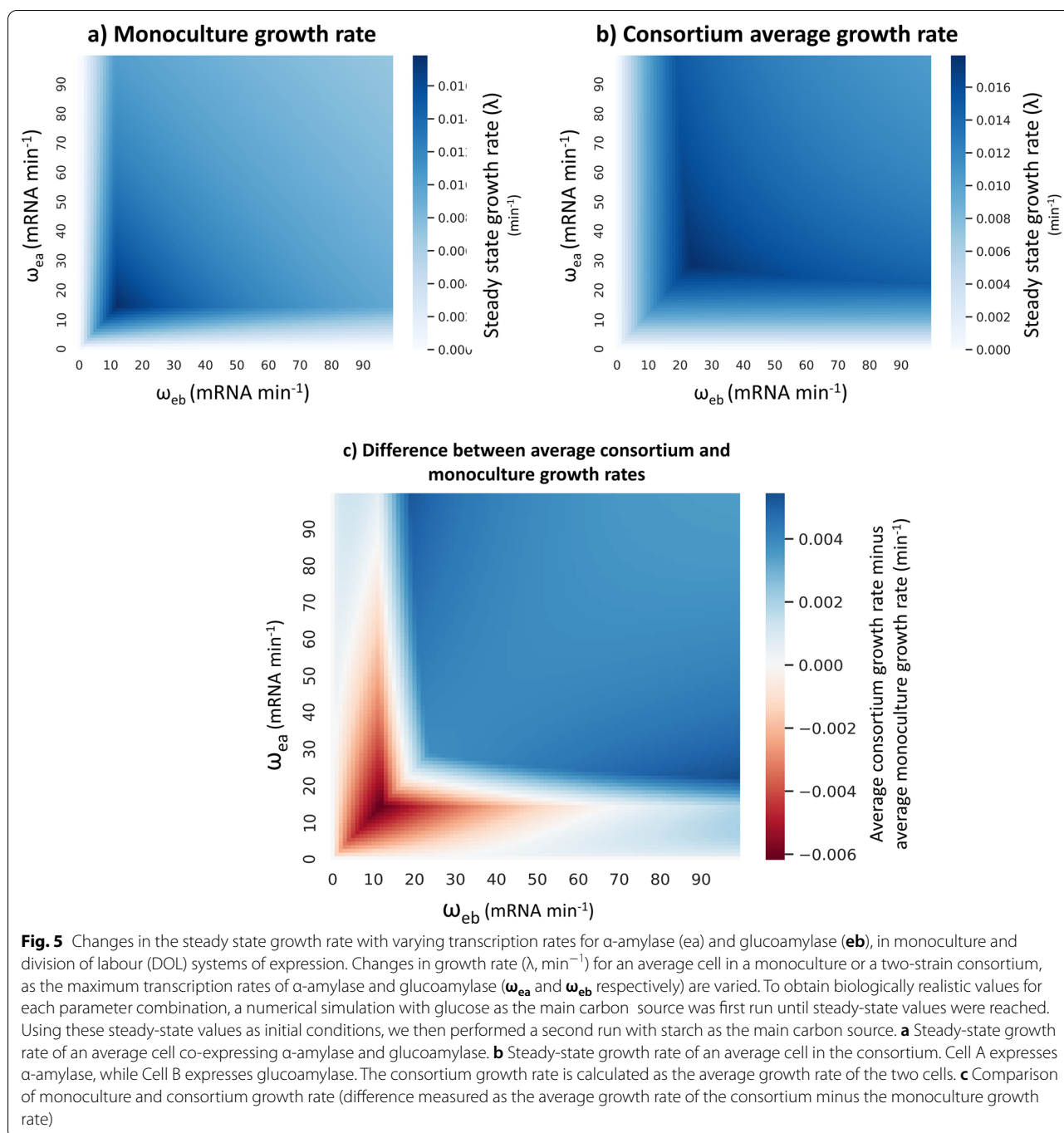
consortium have equal total number of cells. The rate of starch input,  $s_{0in}$ , and the rate of loss of each substrate,  $d_{s_0}$ ,  $d_{s_1}$  and  $d_s$ , can be adjusted for the desired bioreactor setup. These equations were combined with the ODEs for Cell A and Cell B as shown in Fig. 4.

### DOL can achieve higher growth rate on starch than an equivalent monoculture system

To investigate the relationship between amylase expression, starch degradation and burden when starch is used as the primary carbon source, we use the model in Eqs. 1–5 to numerically simulate growth rate for a range of expression levels of amylases.

Figure 5a shows the growth rate for a monoculture single cell as a function of the maximum transcription rate ( $\omega$ ) of  $\alpha$ -amylase ( $ea$ ) and glucoamylase ( $eb$ ). Similarly, Fig. 5b shows the average growth rate of the cells in the consortium, where one cell (Cell A) expresses only  $\alpha$ -amylase, and the other (Cell B) expresses only glucoamylase. The maximum growth rate achieved by the monoculture was  $0.01784 \text{ min}^{-1}$ , or a doubling time of 38.85 min, which was achieved at  $\omega_{ea} = 15$  and  $\omega_{eb} = 12 \text{ mRNA min}^{-1}$ . Different ratios of the two cell types in the consortium were tested. Approximately equal ratios showed the best average growth rate for the consortium (Additional file 1: Fig. S1). For a 1:1 ratio of the cell types the maximum average growth rate achieved by the consortium was  $0.01793 \text{ min}^{-1}$ , or a doubling time of





38.66 min, which was achieved at  $\omega_{ea}=29$  and  $\omega_{eb}=24$  mRNA  $\text{min}^{-1}$ . At this maximum, Cell A is growing at  $0.01779 \text{ min}^{-1}$  (doubling time 38.96 min), while Cell B is growing at  $0.01807 \text{ min}^{-1}$  (doubling time 38.36 min).

Considering the cell types in the consortium separately, the fastest growth rate achieved for Cell A in the consortium was  $0.01915 \text{ min}^{-1}$  (doubling time 36.20 min), at  $\omega_{ea}=22$  mRNA/min and  $\omega_{eb}=99$  mRNA/min (Cell B

at these values has a growth rate of  $0.01004 \text{ min}^{-1}$  or a doubling time of 69.04 min). The fastest growth rate for Cell B was  $0.01924 \text{ min}^{-1}$  (doubling time 36.03 min), at  $\omega_{ea}=99$  mRNA/min and  $\omega_{eb}=19$  mRNA/min (Cell A at these values has a growth rate of  $0.01098 \text{ min}^{-1}$  or a doubling time of 63.13 min). This shows that the growth of each cell type in the consortium is maximal when its consortium partner's transcription rate is high.



Figure 5c shows the difference between the average consortium growth rate and the monoculture growth rate. At lower transcription rates the monoculture has a faster growth rate than the consortium (red region in Fig. 5c). We assume a minimal population of 2 cells in both the monoculture and the consortium. In the monoculture, we assume each of the 2 cells produces both the heterologous proteins in the amount predicted by the monoculture whole-cell model. In the consortium, however, assuming a 1:1 ratio of the two cell types, each cell only produces one of the heterologous proteins. Based on the whole-cell model predictions a 2-cell monoculture will produce a higher number of amylases in total than a 2-cell consortium (Additional file 1: Fig. S2). Interrogating the different regions in Fig. 5c reveals that the parameter values of the red region, at which the monoculture grows better than the consortium, correspond to the parameter values at which the monoculture is consuming all the glucose provided by the amylases (Additional file 1: Fig. S3). This indicates that when glucose production is the limiting factor for growth (rather than heterologous gene expression burden) the monoculture system has the advantage over the DOL system.

At higher transcription rates, the consortium outperforms the monoculture (blue region in Fig. 5c). At these parameter values the monoculture is more affected by burden than the consortium since it co-expresses two heterologous proteins which divert resources away from essential endogenous genes. At the higher transcription levels, this allows the consortium to achieve a higher maximum growth rate than the monoculture. Burden must be sufficiently limiting to growth, which will occur at the higher transcription levels, for DOL to be an appropriate strategy to improve growth on starch. Changing the ratio of the two cell types will shift the threshold at which the consortium outperforms the monoculture (Additional file 1: Fig. S4).

## Discussion

Whole-cell models for predicting burden have been underutilised in the design of DOL in consortia. Here, we have demonstrated that these models can be adapted to observe and predict the difference in growth when expressing heterologous genes by DOL versus in a single cell. A consortium of two cells each expressing a single heterologous protein achieves a higher growth rate than a monoculture co-expressing two heterologous proteins; this benefit of DOL is a phenomenon often observed in natural microbiomes. By explicitly modelling the resource allocations of translation this model reflects the biological mechanism behind this observation; the redirection of ribosomes toward heterologous gene expression is greater in the monoculture

cell than in either of the consortium cells. It is important to note that, in the monoculture, the growth rate and production of both enzymes is dependent on both genes' expression level, which reflects a design challenge when trying to optimise either of the two genes in a one-factor-at-a-time approach. In the DOL case, when grown on glucose, each cell is only dependent on the transcription rate of its respective protein, meaning both cells could be optimised independently and in parallel. Because of the reduced burden, either cell of the consortium, Cell A or Cell B, produces more per cell of their respective enzyme A or B than the amount produced per cell in the monoculture. However, at a population level, when the monoculture and the consortium both have the same total number of cells, the monoculture will produce a higher total amount of heterologous proteins than the consortium. Therefore, it is important to consider the end-point application of such DOL. If the aim is to produce the maximum of enzyme A or enzyme B then the monoculture may be preferable, whereas if the aim is to provide optimal growth rates, perhaps for production of an endogenous molecule of interest, then DOL may be preferable.

Here, the application of interest is in dividing the labour of complex-substrate degradation, with starch used as a specific example. Whether or not DOL is suitable to improve growth of engineered *E. coli* on starch depends on the balance between useful expression of the amylases to produce glucose, and non-productive expression which creates unnecessary burden. The model identifies the parameter space within which the monoculture or the DOL system is preferable. Where starch degradation to glucose is the limiting factor to growth, a monoculture system would be preferable. A 2-cell monoculture performs better than a 2-cell consortium at low transcription rates because total amylase production is higher when 2 cells are co-expressing both amylases, compared to a duoculture consortium in which each cell is expressing a single type of amylase. However, there is a threshold beyond which amylase expression is no longer beneficial to growth and instead becomes detrimental. Monoculture cells reach this threshold at a lower transcription rate than the consortium cells. For all transcription rates above this threshold the consortium achieves a higher growth rate than the monoculture. Therefore, where burden is the limiting factor to growth, a DOL system is preferable. The consortium could also achieve a higher maximum growth rate than the monoculture if the parameters are optimised precisely. By identifying the parameter space within which DOL can provide benefits, this model will inform design choices for complex-substrate degradation and aid in attaining the optimum expression level for maximal growth.

When grown on glucose as the primary carbon source, the difference in growth achieved by the consortia compared to the monoculture is much more distinct, because the expressed heterologous protein is disconnected from the substrate production. Currently the only output is growth on starch, however for a full consolidated bioprocess, the cells should also biosynthesise desirable products. Heterologous biosynthesis genes will not have the positive connection to growth seen by amylases. Therefore, DOL may be more beneficial to a bioproduction pathway than to a substrate degradation pathway.

One important limitation of this model is that it assumes a stable ratio of cell types in the consortium and therefore does not consider the effect of asymmetric growth rates on population dynamics. In the monoculture the primary substrate is shared equally, as all cells grow at the same rate. But, in consortia, if one cell type has a faster growth rate it may consume more substrate than its consortium partner. The current version of the two-strain consortium model indicates that if one protein transcription is set very high and the other very low there may be a significant difference in growth rate between the two cell types. The lower-expressing cell will have a survival advantage and could dominate a co-culture. However, this model assumes that populations can be robustly controlled at a stable fixed ratio and therefore does not give any perspective on the population dynamics. Simplified host-aware coarse-grained models or agent-based model approaches may be preferable for achieving both cellular-level and population-level dynamics.

Another limitation of the model in its current state is the lack of experimental validation for the quantitative results. This model demonstrates the reduced metabolic burden that is seen when DOL is implemented. This advantage of microbial consortia is supported by experimental data for DOL in biosynthesis pathways, allowing higher productivity or production of molecules that is impossible to achieve by single cells [16, 17, 48]. Further, there are many experimental studies that support the benefits of using consortia for degradation of lignocellulosic biomass [21, 49–51], including combining  $\alpha$ -amylase-expressing species with glucoamylase-expressing species [52]. However, experimental studies which directly compare a consortium expressing different degradation enzymes with an equivalent co-expressing monoculture are currently an underexplored area of investigation. So, while this study provides a theoretical prediction for how such a consortium would behave compared to monoculture, these predictions remain to be verified experimentally.

Currently, our proposed approach only considers gene expression burden caused by the energetic cost and reallocation of translational resources. However, burden can

encompass other costs incurred by a host cell due to the expected or unexpected functions of the synthetic pathway [53]. When introduced to a new proteome and metabolome, heterologous proteins may be promiscuous [54]. This can cause toxicity burden, and includes the production of unexpected toxic byproducts, induction of unfolded-protein stress response, or obstruction of endogenous protein function [11]. On the other hand, the heterologous protein may function as desired but inadvertently produce role-based burden, such as depleting the intracellular resources for native metabolic pathways [53]. Therefore, some pathways that could benefit from DOL due to these other burdens will not accurately be predicted by our proposed model.

## Conclusions

In conclusion, we have proposed a resource-aware whole-cell model that highlights the interesting relationships that exist between DOL and burden for a system where primary substrate catabolism is dependent on heterologous protein production. A two-strain *E. coli* consortium expressing  $\alpha$ -amylase and glucoamylase singularly in different cell types can achieve a higher growth rate on starch than a monoculture co-expressing both genes. However, this is only achievable within a particular region of expression, identified by the model, where amylase production is not limiting the glucose supplies to the cell. This model may be reparametrized to other host organisms where translation plays the primary role in heterologous expression burden. Further, adapting this, or similar, resource-aware coarse-grained models may be useful in the future for other DOL applications, such as distributed biosynthesis pathways or multicellular genetic circuits, in predicting burden and informing design choices.

## Methods

### Hardware and software

Data was generated on a Windows 10 PC with Intel® Core™ i7-1065G7 CPU @ 1.30 GHz, 16 GB RAM using Python 3.7 with a Spyder 4.1 (as part of Anaconda) IDE. Where multiple processors were available, the computational labour was divided between processors using the `dask.distributed` package (`distributed.dask.org`) for Python.

### ODE solver and simulation methods

The `scipy.integrate.solve_ivp` function from the `scipy` package was used to solve ODEs, with the LSODA method (Adams/BDF method with automatic stiffness detection and switching). In all simulations the time span was set to 0–20,000 min, and the absolute tolerance (`atol`) and relative tolerance (`rtol`) were set to  $1e^{-12}$ . Steady state

was taken as the value for each intracellular molecule at the end of a numerical simulation of 20,000 min.

For simulations where glucose is supplied externally the ODEs were initialised with 1000 energy units ( $\alpha$ ) and 10 ribosomes ( $r$ ), as in the original Weiße model. For simulations on starch, a simulation was first run for growth on glucose, then the steady-state values reached were used as initial conditions for a simulation on starch.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12934-022-01842-0>.

**Additional file 1.** Additional Tables and figures.

## Acknowledgements

G.-B.S. gratefully acknowledges the support of the Royal Academy of Engineering via the RAE Chair in Emerging Technologies programme (Grant Agreement C1ET1819\5).

## Author contributions

Conceptualization: RL and G-BS; Model creation: EA, G-BS and ZT; Simulation & analysis: EA; Interpretation of data: EA, ZT, GP, G-BS, RL; Supervision: G-BS, ZT and RL; Writing—original draft: EA; Writing—review and editing: G-BS, RL and GP. All authors read and approved the final manuscript.

## Funding

This work received funding from BBSRC (BB/R01602X/1), 19-ERACoBioTech- 33 SyCoLim BB/T011408/1 and BBSRC BB/T013176/1) British Council 527429894, Newton Advanced Fellowship (NAF/R1\201187), Yeast4Bio Cost Action 18229, European Research Council (ERC) (DEUSBIO - 949080) and the Bio-based Industries Joint (PERFECOAT- 101022370) under the European Union's Horizon 2020 research and innovation programme. E.A. was funded by the EPSRC as part of the Biodesign Engineering Centre for Doctoral Training.

## Availability of data and materials

The model generated during this study is available publicly on the GitHub repository: <https://github.com/elizaatkinson/Resource-aware-whole-cell-model-of-DOL-in-2-strain-consortium-for-complex-substrate-degradation>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 28 January 2022 Accepted: 30 May 2022

Published online: 14 June 2022

## References

1. Olguin-Maciel E, Singh A, Chable-Villacis R, Tapia-Tussell R, Ruiz HA. Consolidated bioprocessing, an innovative strategy towards sustainability for biofuels production from crop residues: an overview. *Agronomy*. 2020;10:1834.
2. Fan Z. Consolidated bioprocessing for ethanol production. *Biorefineries Integr Biochem Process Liq Biofuels*. 2014. p. 141–60.
3. Brethauer S, Studer MH. Consolidated bioprocessing of lignocellulose by a microbial consortium. *Energy Environ Sci*. 2014;7:1446–53.
4. Minty JJ, Lin XN. Engineering synthetic microbial consortia for consolidated Bioprocessing of lignocellulosic biomass into valuable fuels and chemicals. *Direct Microb Convers Biomass to Adv Biofuels*. Elsevier; 2015. p. 365–81.
5. Brenner K, You L, Arnold FH. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol*. 2008;26:483–9.
6. Verbeke TJ, Zhang X, Henrissat B, Spicer V, Rydzak T, Krokhin OV, et al. Genomic evaluation of *Thermoanaerobacter* spp. for the construction of designer co-cultures to improve lignocellulosic biofuel production. *PLoS ONE*. 2013;8:e59362. <https://doi.org/10.1371/journal.pone.0059362>.
7. Tsai SL, Oh J, Singh S, Chen R, Chen W. Functional assembly of minicellulosomes on the *Saccharomyces cerevisiae* cell surface for cellulose hydrolysis and ethanol production. *Appl Environ Microbiol*. 2009;75:6087–93.
8. Scott M, Hwa T. Bacterial growth laws and their applications. *Curr Opin Biotechnol*. 2011;22:559–65.
9. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of cell growth and gene expression: origins and consequences. *Science*. 2010;330:1099–102.
10. Borkowski O, Ceroni F, Stan G-B, Ellis T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr Opin Microbiol*. 2016;33:123–30.
11. Glick BR. Metabolic load and heterologous gene expression. *Biotechnol Adv*. 1995;13:247–61.
12. Chen T, Zhou Y, Lu Y, Zhang H. Advances in heterologous biosynthesis of plant and fungal natural products by modular co-culture engineering. *Biotechnol Lett*. 2019;41(1):27–34.
13. Zhang H, Wang X. Modular co-culture engineering, a new approach for metabolic engineering. *Metab Eng*. 2016;37:114–21.
14. Bernstein HC, Paulson SD, Carlson RP. Synthetic *Escherichia coli* consortia engineered for syntrophy demonstrate enhanced biomass productivity. *J Biotechnol*. 2012;157:159–66.
15. Zhang H, Pereira B, Li Z, Stephanopoulos G, Demain AL. Engineering *Escherichia coli* coculture systems for the production of biochemical products. *Proc Natl Acad Sci USA*. 2015;112:8266–71.
16. Jones JA, Vernacchio VR, Collins SM, Shirke AN, Xiu Y, Englaender JA, et al. Complete biosynthesis of anthocyanins using *E. coli* polycultures. *MBio*. 2017;8:e00621.
17. Jones JA, Vernacchio VR, Sinkoe AL, Collins SM, Ibrahim MHA, Lachance DM, et al. Experimental and computational optimization of an *Escherichia coli* co-culture for the efficient production of flavonoids. *Metab Eng*. 2016;35:55–63.
18. Saini M, Hong Chen M, Chiang CJ, Chao YP. Potential production platform of *n*-butanol in *Escherichia coli*. *Metab Eng*. 2015;27:76–82.
19. Zhang H, Stephanopoulos G. Co-culture engineering for microbial biosynthesis of 3-amino-benzoic acid in *Escherichia coli*. *Biotechnol J*. 2016;11:981–7. <https://doi.org/10.1002/biot.201600013>.
20. Wang X, Policarpio L, Prajapati D, Li Z, Zhang H. Developing *E. coli*-*E. coli* co-cultures to overcome barriers of heterologous tryptamine biosynthesis. *Metab Eng Commun*. 2020;10:e00110.
21. Tsai SL, Goyal G, Chen W. Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production. *Appl Environ Microbiol*. 2010;76:7514–20.
22. Goyal G, Tsai SL, Madan B, DaSilva NA, Chen W. Simultaneous cell growth and ethanol production from cellulose by an engineered yeast consortium displaying a functional mini-cellulosome. *Microb Cell Fact*. 2011;10:89. <https://doi.org/10.1186/1475-2859-10-89>.
23. Thommes M, Wang T, Zhao Q, Paschalidis IC, Segrè D. Designing metabolic division of labor in microbial communities. *mSystems*. 2019;4:e00263.
24. Gutiérrez M, Gregorio-Godoy P, Pérez Del Pulgar G, Muñoz LE, Sáez S, Rodríguez-Patón A. A new improved and extended version of the multi-cell bacterial simulator gro. *ACS Synth Biol*. 2017;6:1496–508.
25. Matyjaszkiewicz A, Fiore G, Annunziata F, Grierson CS, Savery NJ, Marucci L, et al. BSim 2.0: an advanced agent-based cell simulator. *ACS Synth Biol*. 2017;6:1969–72.
26. Rollié S, Mangold M, Sundmacher K. Designing biological systems: systems engineering meets synthetic biology. *Chem Eng Sci*. 2012;69:1–29.

27. Harvey E, Heys J, Gedeon T. Quantifying the effects of the division of labor in metabolic pathways. *J Theor Biol.* 2014;360:222–42.
28. Mauri M, Gouzé JL, de Jong H, Cinquemani E. Enhanced production of heterologous proteins by a synthetic microbial community: conditions and trade-offs. *PLoS Comput Biol.* 2020;16:e1007795.
29. Tsoi R, Wu F, Zhang C, Bewick S, Karig D, You L. Metabolic division of labor in microbial systems. *Proc Natl Acad Sci USA.* 2018;115:2526–31.
30. Gorochowski TE, Avciilar-Kucukgoze I, Bovenberg RAL, Roubos JA, Ignatova Z. A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes. *ACS Synth Biol.* 2016;5:710–20.
31. Liao C, Blanchard AE, Lu T. An integrative circuit-host modelling framework for predicting synthetic gene network behaviours /631/553/552/631/553/2695 article. *Nat Microbiol.* 2017;2:1658–66.
32. Sickler JJ, Ni C, Shen D, Wang Z, Jin M, Lu T. Integrative circuit-host modeling of a genetic switch in varying environments. *Sci Rep.* 2020;10:1–9. <https://doi.org/10.1038/s41598-020-64921-5>.
33. Weiße AY, Oyarzún DA, Danos V, Swain PS. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc Natl Acad Sci USA.* 2015;112:E1038–47.
34. Darlington APS, Kim J, Jiménez JI, Bates DG. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat Commun.* 2018;9:1–12.
35. Nikolados EM, Weiße AY, Ceroni F, Oyarzún DA. Growth defects and loss-of-function in synthetic gene circuits. *ACS Synth Biol.* 2019;8:1231–40. <https://doi.org/10.1021/acssynbio.8b00531>.
36. Nikolados E-M, Weiße AY, Oyarzún DA, Oyarzún O. Prediction of cellular burden with host-circuit models. 2020. <http://arxiv.org/abs/2004.00995>
37. Annunziata F, Matyjaszkiewicz A, Fiore G, Grierson CS, Marucci L, Di Bernardo M, et al. An orthogonal multi-input integration system to control gene expression in *Escherichia coli*. *ACS Synth Biol.* 2017;6:1816–24. <https://doi.org/10.1021/acssynbio.7b00109>.
38. Losoi PS, Santala VP, Santala SM. Enhanced population control in a synthetic bacterial consortium by interconnected carbon cross-feeding. *ACS Synth Biol.* 2019;8:2642–50. <https://doi.org/10.1021/acssynbio.9b00316>.
39. McCardell RD, Huang S, Green LN, Murray RM. Control of bacterial population density with population feedback and molecular sequestration. *bioRxiv.* 2017. <https://doi.org/10.1101/225045v1>.
40. Jiang W, Yang X, Gu F, Li X, Wang S, Luo Y, et al. Construction of synthetic microbial ecosystems and the regulation of population proportion. *ACS Synth Biol.* 2022. <https://doi.org/10.1021/acssynbio.1c00354>.
41. Soufi B, Krug K, Harst A, Macek B. Characterization of the *E. coli* proteome and its modifications during growth and ethanol stress. *Front Microbiol.* 2015;6:103.
42. Milo R. What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays.* 2013;35:1050.
43. Dong H, Nilsson L, Kurland CG. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J Bacteriol.* 1995;177:1497–504. <https://doi.org/10.1128/jb.177.6.1497-1504.1995>.
44. El-Fallal A, Abou M, El-Sayed A, Omar N. Starch and microbial  $\alpha$ -amylases: from concepts to biotechnological applications. *Carbohydrates.* 2012. <https://doi.org/10.5772/51571>.
45. Bijttebier A, Goesaert H, Delcour JA. Amylase action pattern on starch polymers. *Biologia.* 2008;63:989–99.
46. Azhari R, Lotan N. Enzymic hydrolysis of biopolymers via single-scission attack pathways: a unified kinetic model. *J Mater Sci Mater Med.* 1991;2:9–18.
47. Fujii M, Kawamura Y. Synergistic action of  $\alpha$ -amylase and glucoamylase on hydrolysis of starch. *Biotechnol Bioeng.* 1985;27:260–5. <https://doi.org/10.1002/bit.260270308>.
48. Zhang H, Li Z, Pereira B, Stephanopoulos G. Engineering *E. coli-E. coli* cocultures for production of muconic acid from glycerol. *Microb Cell Fact.* 2015;14:1–10. <https://doi.org/10.1186/s12934-015-0319-0>.
49. Kogo T, Yoshida Y, Koganei K, Matsumoto H, Watanabe T, Ogihara J, et al. Production of rice straw hydrolysis enzymes by the fungi *Trichoderma reesei* and *Humicola insolens* using rice straw as a carbon source. *Bioresour Technol.* 2017;233:67–73.
50. Li X, He Y, Zhang L, Xu Z, Ben H, Gaffrey MJ, et al. Discovery of potential pathways for biological conversion of poplar wood into lipids by co-fermentation of Rhodococci strains. *Biotechnol Biofuels.* 2019;12:1–16. <https://doi.org/10.1186/s13068-019-1395-x>.
51. Chen L, Du JL, Zhan YJ, Li JA, Zuo RR, Tian S. Consolidated bioprocessing for cellulosic ethanol conversion by cellulase–xylanase cell-surfaced yeast consortium. *Prep Biochem Biotechnol.* 2018;48:653–61. <https://doi.org/10.1080/10826068.2018.1487846>
52. Wang S, Tang H, Peng F, Yu X, Su H, Xu P, et al. Metabolite-based mutualism enhances hydrogen production in a two-species microbial consortium. *Commun Biol.* 2019;2:1–11.
53. Ellis T. Predicting how evolution will beat us. *Microb Biotechnol.* 2019. <https://doi.org/10.1111/1751-7915.13327>.
54. Cardinale S, Arkin AP. Contextualizing context for synthetic biology—identifying causes of failure of synthetic biological systems. *Biotechnol J.* 2012;7:856–66.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

