Received: 11 September 2022 | Accepted: 26 January 2023

DOI: 10.1111/1751-7915.14233

MICROBIAL BIOTECHNOLOGY Applied Microbiology International

RESEARCH ARTICLE

Engineering resource allocation in artificially minimized cells: Is genome reduction the best strategy?

Elisa Marquez-Zavala[1,2] | **Jose Utrilla**[2]

[1]Department of Biotechnology and Food Science, Norwegian University of Science and Technology, Trondheim, Norway

[2]Synthetic Biology Program, Center for Genomic Sciences, Universidad Nacional Autónoma de México, Cuernavaca, Mexico

Correspondence
José Utrilla, Synthetic Biology Program, Center for Genomic Sciences, Universidad Nacional Autónoma de MéxicoAv. Universidad sn Col. Chamilpa, Cuernavaca, Morelos, Mexico.
Email: utrilla@ccg.unam.mx

Funding information
DGAPA- UNAM- PAPIIT, Grant/Award Number: IN213420; CONACyT Ciencia de Frontera, Grant/Award Number: 319352

Abstract

The elimination of the expression of cellular functions that are not needed in a certain well-defined artificial environment, such as those used in industrial production facilities, has been the goal of many cellular minimization projects. The generation of a minimal cell with reduced burden and less host-function interactions has been pursued as a tool to improve microbial production strains. In this work, we analysed two cellular complexity reduction strategies: genome and proteome reduction. With the aid of an absolute proteomics data set and a genome-scale model of metabolism and protein expression (ME-model), we quantitatively assessed the difference of reducing genome to the correspondence of reducing proteome. We compare the approaches in terms of energy consumption, defined in ATP equivalents. We aim to show what is the best strategy for improving resource allocation in minimized cells. Our results show that genome reduction by length is not proportional to reducing resource use. When we normalize calculated energy savings, we show that strains with the larger calculated proteome reduction show the largest resource use reduction. Furthermore, we propose that reducing highly expressed proteins should be the target as the translation of a gene uses most of the energy. The strategies proposed here should guide cell design when the aim of a project is to reduce the maximum amount or cellular resources.

INTRODUCTION

Biological complexity is a challenge to understand and model, many synthetic biology projects are based on the idea that to improve production hosts or chassis we need to reduce the inherent biological complexity. Starting from a simplified organism has enormous potential to reduce detrimental host-function interaction, therefore, there have been many efforts to reduce microbial genomes (Fredens et al., 2019; Michalik et al., 2021) potentially also reducing the complexity of microbial production hosts. The idea of a simple cell with the minimal set of functions needed to grow and to perform a programmed synthetic biology function has been placed as the cornerstone of the establishment of a quasi-universal synthetic biology chassis. Many complexity reduction approaches are mainly based on genome reduction, focused on finding the minimal set of genes that are able to sustain life, however many of such strains have shown growth defects (Choe et al., 2019).

Building bottom-up synthetic genomes should provide us with the ability to better understand and design a minimal cell, however to date those projects have shown to be challenging. In the last iteration of a bottom-up synthetic genome, many genes with unknown functions had to be added to the minimal cell in order to produce a viable cell (Hutchison et al., 2016). Top-down projects have mainly focused on reducing parts of the genome using different approaches, however, the cellular resource consumption of eliminated genes-proteins-functions has not been considered.

In this work, we aim to compare bacterial complexity reduction approaches using *Escherichia coli* as a model organism. We formalize the calculations of the saved

resources of eliminating genes when those genes are transcribed and translated (to a median value). We calculated the theoretical liberated resources in terms of energy and proteome liberation for a defined growth environment. We show that resource reallocation efforts can be optimized if they are focused on a few genes producing highly expressed dispensable proteins. We then propose that resource allocation reduction strategies should be focused on the proteome rather than on the genome.

EXPERIMENTAL PROCEDURES

Data processing of strains

We processed the information of nine strains with a minimized genome (Hashimoto et al., 2005; Hirokawa et al., 2013; Karcagi et al., 2016; Mizoguchi et al., 2008; Park et al., 2014; Pósfai et al., 2006) and a strain with a minimized proteome (Lastiri-Pancardo et al., 2020) (Table 1). We obtained the eliminated genes of the minimized genome strains from their most recent articles. If they reported genome elimination ranges, these coordinates were mapped against the information from their parental strain in the GenBank database (Sayers et al., 2019; the *Escherichia coli* str. K-12 substr. MG1655 genome, version NC_000913.3 and the *Escherichia coli* str. K-12 substr. W3110 genome, version AP009048.1).

Calculation of the proteome reduction of each analysed strain

To calculate the amount of proteome that a certain strain is reducing we assumed that the removed protein-coding genes produce the amount of protein measured in the absolute proteomics data set in the same growth condition (Schmidt et al., 2016). The amount of saved proteome was calculated in femtograms per cell, however, it was expressed in percentage of the total measured proteome to facilitate comparison amongst strains and conditions. The used data set comprises 95% of proteome coverage by mass (around 2300 proteins in each condition). Proteins with no data in a certain condition were not considered by these analyses since their proteome contribution may be considered negligible.

Estimation of replication, transcription and translation costs

To have a normalized basis for comparison amongst different cellular process we performed calculations on a gene with an average length of 950 bp as defined by Lynch & Marinov, 2015. We calculate the energetic cost of the replication, transcription or translation of a gene of 950 pb with the iJLE1678-ME model of *E. coli*. In order to perform these calculations, we simulated the change in biomass composition by increasing the genome, the transcriptome or the proteome size by a unit equivalent of a gene of 950 pb (or its gene product). Once the biomass composition was modified in the model we used fixed the growth rate, glucose uptake rate and oxygen uptake rate to simulate batch growth on a minimal medium and we used ATP maintenance (ATPM) as the objective function to calculate the differences on unaccounted for energy by changing the biomass composition of each cellular process. The calculated change in ATPM is the cost of producing DNA, RNA or protein. As the abundance of proteins can span several orders of magnitude and in order to set the translation level of this average gene, we used the median contribution to the proteome of a gene. This is that a 950 pb gene would be translated to reach 0.042% of the proteome and 22.4 Kbp of the genome are needed to produce 1% of the proteome.

Analysis of the cellular resources in terms of ATP from each eliminated gene with the ME model and proteomic data

To calculate total replication, transcription and protein production costs for each minimized strain, first, we used the amount of genome, transcriptome and calculated proteome corresponding to each reduced strain. We used the cellular composition reported in Bremer and Dennis 1996 (Dennis & Bremer, 2008) to compensate for growth-dependent changes in cellular composition, we also compensated the genome equivalents considering the genome equivalents at a given growth rate (e.g. at 0.69 h^{-1} we estimate 1.63 genomes per cell). The iLE1678-ME model was used by fixing the growth rate, the glucose uptake rate and the oxygen uptake rate, we used the ATP maintenance (ATPM) as the objective function for the proteome simulation, whereas for the other simulations, we maintained the default objective of dummy protein production. Then, changing the amount of DNA in the cell, changes in the number of transcribed genes, and changes in protein production respectively. ME-model simulations calculated the changes in ATP consumption, then it was converted to cost per Mb depending on the length of the gene (data obtained from NCBI) as well as the percentage of proteome they represented in a specific condition from the quantitative proteome data of Schmidt et al., 2016. Detailed computational simulations can be reproduced using the code and data provided in the accompanying github repository.

TABLE 1 Genotype and phenotypic traits of the analysed strains.

Strain	Background	Eliminated genome (mb)	Eliminated genome (%)	Eliminated genes (Number)	Phenotype	Reference
Δ16	MG1655	1.3	29.70	1227	• Lower growth rate and changes in cell shape in some media (thicker and elongated) • Increased cell volume likely as a result of deteriorated peptidoglycan integrity and/or by impaired osmotic regulation	Hashimoto et al. (2005)
MS56	MG1655	1.07	23	1072	• Higher stability in the production of recombinant proteins • Possible reduction in the production of the foreign genes (plasmids) • Similar growth rates in rich media, but either faster or slower growth in minimal media • May be sensitive to small differences in cultural conditions • Imbalances in some cofactors and disrupted metabolism in minimal media	Park et al. (2014), Choe et al. (2019)
MDS69	MG1655	0.94	20.30	965	• Growth rate is 17% lower (probably related to large fluctuations in transcriptome caused by genome reduction) • Reduced competitive fitness in LB medium • Decreased nitrogen utilization efficiency • Growth defects in 14.3% of environments where the parental has no issues	Pósfai et al. (2006), Karcagi et al. (2016), Vernyik et al. (2020)
MDS42	MDS/ MG1655	0.66	14.3	704	• Increased L-threonine production by 83% • Higher electroporation efficiency • Similar growth rates in minimal and rich media, with one study showing a 20% lower growth rate • Similar mutation rates and reduced genome stability under minimal media • Growth defects in 8.8% of environments where the parental has no issues	Pósfai et al. (2006), Lee et al. (2009), Ying et al. (2013), Karcagi et al. (2016), Yuan et al. (2017), Ying & Yama (2018). Vernyik et al. (2020)
MDS12	MDS/ MG1655	0.37	8.11	423	• Higher density in stationary phase • Similar growth rates in minimal and rich media • Similar doubling times, electroporation and transformation efficiencies • Marginally outgrown when cocultured	Kolisnychenko et al. (2002), Pósfai et al. (2006), Karcagi et al. (2016), Vernyik et al. (2020)
MGF01	MGF/W3110	1.03	22.2	1081	• Increased L-threonine production (2.4-fold) • Higher glucose consumption (1.44-fold) and acetate accumulation (0.09-fold) • Similar or slightly higher rate in rich media • Either a higher or a lower growth rate and cell density in minimum media • Higher mutation rate	Hashimoto et al. (2005), Mizoguchi et al. (2007, 2008), Hirokawa et al. (2013), Nakayasu et al. (2020)
MGF02		1.1	25	~1200	• Higher growth rate	Hashimoto et al. (2005), Hirokawa et al. (2013), Nakayasu et al. (2020)

TABLE 1 (Continued)

Strain	Background	Eliminated genome (mb)	Eliminated genome (%)	Eliminated genes (Number)	Phenotype	Reference
DGF298	DGF/W3110	1.67	35.80	~1700	• Better growth rate and cell yield in rich medium • More stable genome • No auxotrophy phenotype • Non-productive mutants of protodeoxyviolacein as it was too burdensome	Hashimoto et al. (2005), Hirokawa et al. (2013), Nakayasu et al. (2020)
DGF327		1.38	30.90	~1400	• Better growth rate and cell yield in rich medium • No auxotrophy phenotype	Hashimoto et al. (2005), Hirokawa et al. (2013), Nakayasu et al. (2020)
PFC	BW25113	~1.67 x 10^{-3}	0.004	3	• Higher proteomic budget • Increased production of violacein (18%) and pDNA (33% and 53%)	Lastiri-Pancardo et al. (2020), de la Cruz et al. (2020)

RESULTS

In this work, we aimed to analyse genome or proteome reduction projects with a resource allocation approach. First, to have a basis for comparison we normalized DNA replication, RNA transcription and protein translation into ATP equivalents. Then, as our results confirmed what has been shown before: translation is the main driver of cellular resource consumption (Lynch & Marinov, 2015; Wagner, 2007), we analysed how much of the proteome was reduced assuming that the eliminated genes were expressed at the same magnitude as in the absolute proteomics data set (Schmidt et al., 2016) in the same growth condition (Figure 1).

In order to have a normalized comparison amongst the analysed cellular processes we used two different calculations that rendered a similar result. We used the Lynch & Marinov, 2015 equations, which calculate the energetic cost, measured in units of ATP hydrolyses, associated with the replication, transcription, and translation of an average gene of 950 bp length as previously defined (Lynch & Marinov, 2015). We also simulated the resulting decrease in unaccounted for energy (ATPM) with a ME-model (Lloyd et al., 2018) when we increase the genome, the transcriptome or the proteome by an equivalent of a gene of 950 bp. For calculation basis, we set the median contribution to the proteome by a 950 pb gene to be 0.042% of the proteome (see experimental procedures). Therefore, from ME-model simulations, we obtain (that 1% of proteome represents a cost of 3.76×10^{-4} and of 3.16×10^{-3} mmol ATP gDW h^{-1}) for replication and transcription, respectively. Whereas translation represents 1.65×10^{-1} mmol ATP gDW h^{-1}.

As we show above, the normalized energy cost of the production of the proteome is three and two orders of magnitude higher than the production of the genome or transcriptome respectively. In order to compare the reduction in resource allocation amongst several genome or proteome-reduction approaches we took the information of nine genomes reduced and one proteome-reduced strain of *E. coli* (Table 1). The MDS (Multiple-Deletion Series) set of strains (Karcagi et al., 2016; Pósfai et al., 2006) are derived from MG1655. In that set of strains, genes were eliminated when they are not present in close relatives, mobile elements and also genes with unknown and non-essential functions were eliminated (such as flagella). The Δ16 strain and MS56 were designed to minimize the MG1655 genome as much as possible aiming to obtain a simple and highly controllable cell with non-essential genes and unneeded genome fragments removed (Hashimoto et al., 2005; Park et al., 2014). The MGF (Minimum Genome Factory) strains, derived from W3110, were designed to eliminate genomic regions that do not hybridize with other genomes and regions that were eliminated in other *E. coli* reduction projects. They also eliminated genes with unknown functions

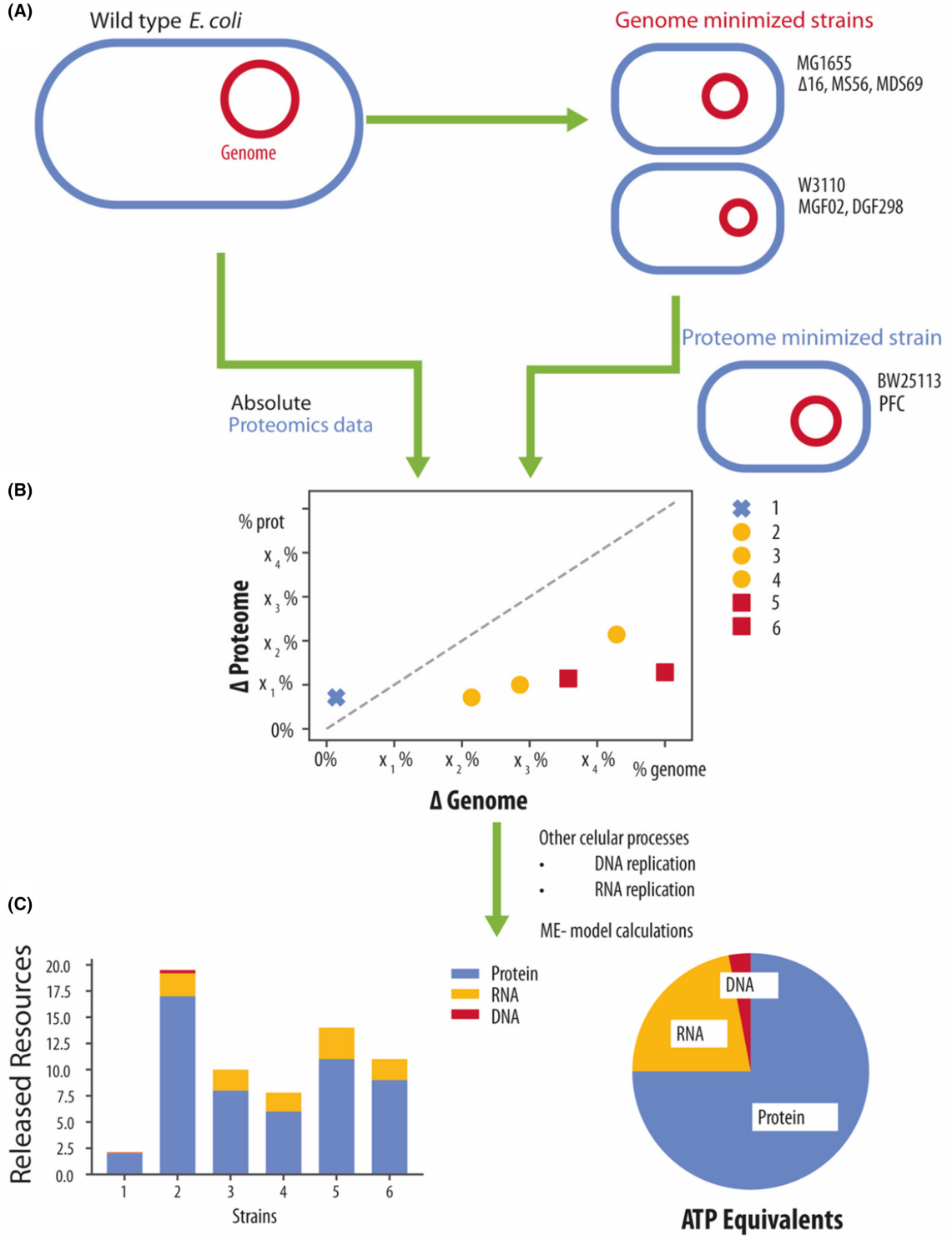

FIGURE 1 Flow chart of the study. (A) The data included for this analysis came from five MG1655 and W3110 derivatives with their genomes minimalized. In addition, data from a proteome-minimized *Escherichia coli* K12 BW25113 derivative was used. (B) Comparison of the amount of proteome that corresponds to the deleted genes in the minimized strains, based on the absolute proteomics data set (Figure 3 for full detail). (C) Calculation of the equivalent ATP consumption in the wild type (WT) provided by the genes that were removed from the minimized genome strains using a metabolism and expression model of *E. coli* and proteomics data. The consumption was evaluated for protein, RNA and DNA production. A pie chart displays the average consumption for each category of the genome-minimized strains.

(Mizoguchi et al., 2008). Based on MGF strains, DGF strains (Designed Genome Factory) were generated by removing insertion sequences and toxin-antitoxin systems. Previously removed regions from MGF strains, causing growth defects such as auxotrophies were restored in this project (Hirokawa et al., 2013). For

better visualization of the results, we selected the strain with the highest genome reduction as a representative of the project the strain with the highest reduction. The MDS69 strain represented the MDS strains, the DGF298 strain represented the DGF strains, and the MGF02 strain represented the MGF strains. Finally, we compare genome-reduced strains to PFC a proteome-reduced strain created by our group (Lastiri-Pancardo et al., 2020). In this project, we applied the ReProMin method, which identifies the minimal set of genetic interventions that maximize the savings in cell resources (Figure 1B).

We used the Schmidt et al., 2016 absolute proteomic data set to calculate the amount of proteome that would be reduced if we assume that the eliminated genes are being expressed at the same magnitude as in the measured proteomes (Figure 2A). Genes with no proteomic information in the data set are not considered since they belong to the 5% of the proteome that is not measured, therefore, we assume that they represent a small contribution to the proteome. In order to set a fixed condition for our comparisons, we used glucose M9 minimal medium as the main analysed growth condition.

All the studied strains from genome reduction projects range from 8 to 35% genome reduction, however, strains were designed and constructed with different approaches, and the per cent of genome reduction is not always proportional to the number of reduced genes (Table 1, Figure 3). As mentioned above, here we focused on the potential resource savings. First, we need to notice the uneven distribution of the contribution of each gene to the total proteome (Figure 2B). The plotted mass contribution of the measured proteins to the proteome in femtograms per cell (fg/cell) spans seven

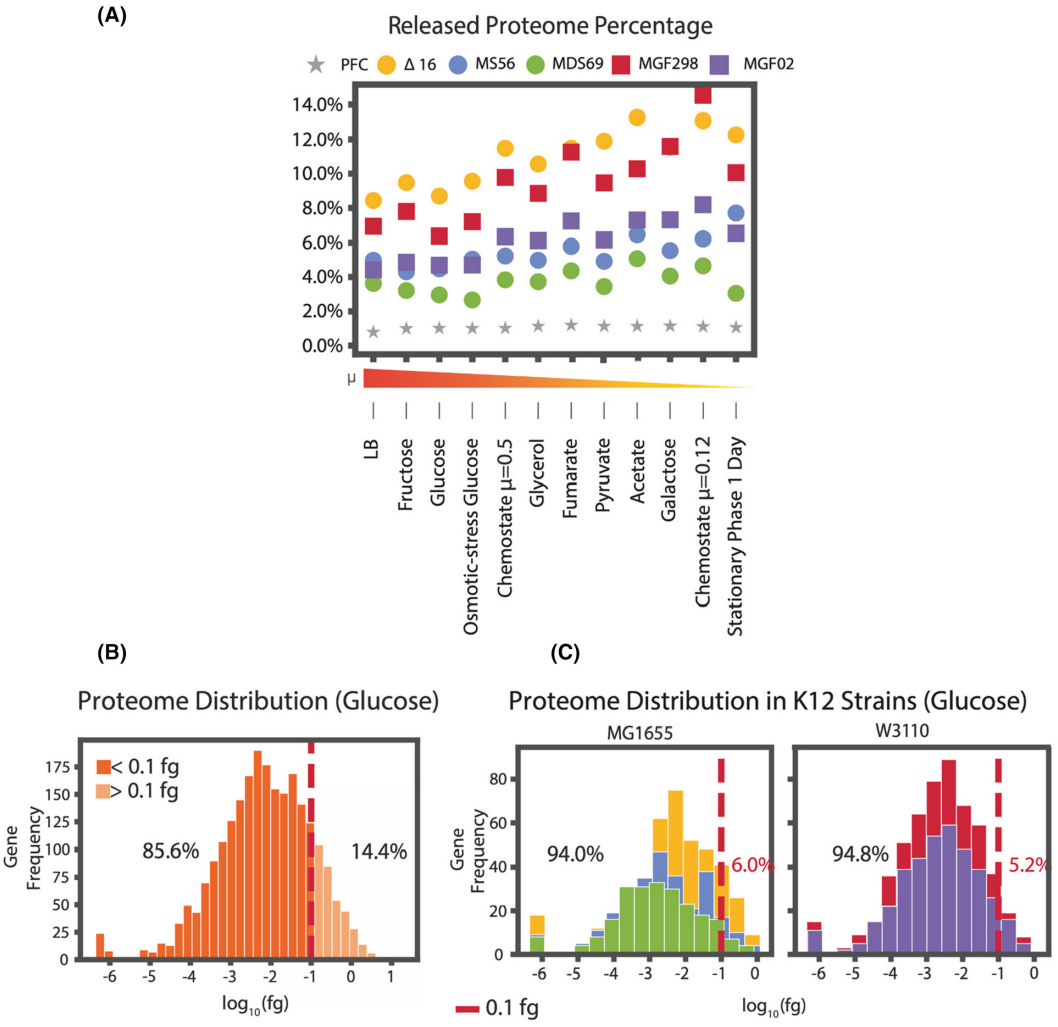

FIGURE 2 Proteome calculations on genome-reduced strains. We used the strains with the smallest genome of each project, and the Schmidt et al., 2016 proteome data set. (A) Percentage of the released proteome in 12 conditions, sorted by specific growth rate. This is the percentage released if we assume that the eliminated genes are being expressed at the same magnitude as in the data set. (B) Proteome mass distribution ($\log_{10}$ fg) per gene on glucose minimal medium growth condition. On the right of the dotted line are the per cent of proteins that have a mass over 0.1 fg. (C) Proteome mass distribution ($\log_{10}$ fg) per from the deleted genes of the genome reduced strains in glucose minimal medium growth condition. On the right of the dotted line are the per cent of proteins that have a mass over 0.1 fg.

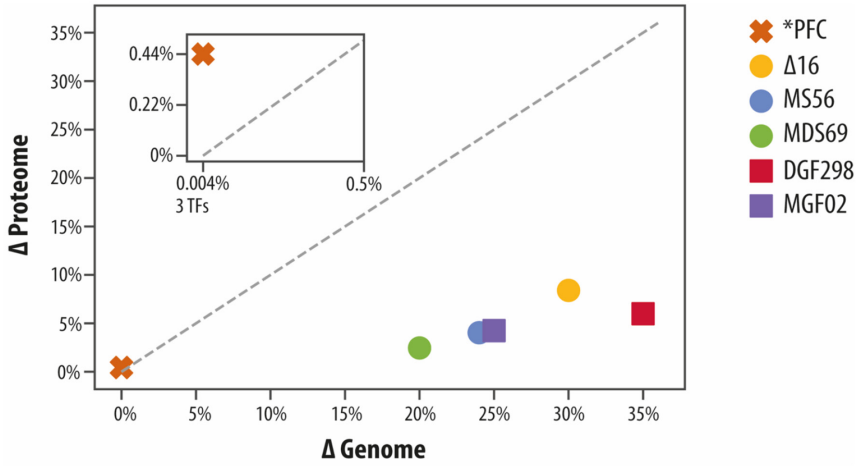

FIGURE 3 Representative strains with their genome reduction in comparison to their calculated proteome reduction in the M9 glucose growth condition. The insert represents the proteome-reduced PFC strain.

orders of magnitude (from 1^{e-6} to 1 fg/cell). Only 14.4% of the measured genes (344) code for a protein that contributes more than 0.1 fg/cell (many of those code for highly expressed essential genes, such as *tufA* or *metE*). This means that few highly expressed genes will be responsible for the greater proteome reduction, therefore it demonstrates that the relation between genome size reduction and potential resource savings is not proportional (Figure 3).

As has been previously stated, our main focus condition is glucose minimal medium, in this condition the calculated proteome reduction of the analysed strains spans from 2.45% to 8.40%. The Δ16 strain showed the largest calculated proteome savings (8.4%) and the MDS69 strain was the one with the smallest calculated proteome savings (2.45%). It is worth mentioning that the Δ16 strain is not the most genome-reduced one, it is the DGF298, however, in that specific condition, the Δ16 strain presents the highest potential proteome reduction. In another remarkable case, for chemostat growth at the 0.12 h^{-1} dilution rate, the DGF298 presents a larger proteome reduction than the Δ16 strain and the largest calculated here (14%, Figure 2A).

We mentioned above that translation (the biosynthesis of the proteome) takes 96% of the energy, when compared on a per gen (950 bp) basis, whereas transcription takes 3.9% and replication just around 0.1% of the total energy (Table 2). Therefore, translation is the main focus of this work. Here, we were also interested in comparing the differences in resource allocation—in terms of normalized flux of ATP in mmol ATP•g DW^{-1} •h—when we calculate the amount of saved energy equivalents. To that end, we used the iLE-1678-ME model to calculate the cost of producing the actual normalized amount of genomic DNA, the transcripts (mRNAs) and the proteins eliminated in each analysed strain. In Figure 4, we show the contribution of each cellular process to the total of saved energy for each strain. These calculations confirmed that translation is the main driver of energy savings ranging from 54 to 84% of saved ATP equivalents. We also showed that

TABLE 2 Calculated ATP cost for each cellular process to produce a protein from a 950 bp gene.

Process	ATP	Per cent
Replication	4180.58	0.23
Transcription	71,359.04	3.93
Translation	1,737,983.23	95.83
Total	1,813,522.86	100

the strains, that according to our calculations, have the greater amount of saved proteome also save the greatest amount of energy. Comparing genome reduction approaches to a proteome reduction-oriented project we show that a strain with only three transcription factors eliminated that resulted in a theoretical proteome reduction of 0.5%, presents a similar calculated resource reduction use than the MSD12 strain with an 8% genome reduction (Figure 4). The proteome-reduced strain showed increases around 20% in the production of fluorescent protein reporters and in violacein from a heterologous metabolic pathway (Lastiri-Pancardo et al., 2020). These results show that if we compare resource savings by the number of removed nucleotides in each project, the proteome reduction approach is more effective than the genome reduction approaches. However, it is also noteworthy that most of the genome-reduced strains have greater savings than the only proteome-reduced strain analysed here.

DISCUSSION

Reduction of cellular complexity is a complicated endeavour that needs to be tackled from many perspectives. The bottom-up approach has proven to be challenging as the definition of minimal gene sets still escapes from our full understanding (Glass et al., 2017). In that regard, top-down approaches may be easier to adopt, therefore more widely used. Considering the

resource use of the genes to eliminate will improve the success of the minimized strains, since reducing less genes with higher contribution to resource use should result in higher budget benefits with less phenotypic defects.

Here, we formalize the calculations of the saved resources of eliminating genes when those genes are transcribed and translated (to a median value). Gene expression is highly variable in terms of transcript and proteins produced per gene, in order to have a basis for calculation we used median values of gene expression. This certainly introduces a large bias in our calculations, however, we provide a coarse estimate of the amount of saved resources for each process. Also, the order of magnitude of change of the resource expenditure amongst the different processes analysed here, highlight that besides those large possible differences in gene expression, the amount of saved resources is much larger for translation than for transcription or DNA replication. Moreover, with less gene deletions, the minimized cell behaviour will become more predictable, and our proteomic data-based calculations should be more accurate, as several gene deletions can dramatically alter protein concentrations. Using a genome-scale model of metabolism and expression (ME-model), we are able to account for "distant" processes such as the need of macromolecular machinery (such as ribosomes) that carry out each process thus incurring a cost.

Many analyses have shown that translation is the major driver of resource consumption of the cell (Kepp, 2020; Lynch & Marinov, 2015; Wagner, 2007). Here, we show that genome reduction does not have a direct correspondence into resource consumption reduction. Our presented findings show that if a cell minimization project aims to reduce the use of resources of the non-used cellular functions then the proteome should be the target. We show that the protein concentration in *E. coli* spans seven orders of magnitude, finding the most expressed non-essential proteins is a pretty simple endeavour that should yield a significant reduction in resource consumption with a less complicated genome editing strategy.

The idea that a minimized genome will yield a minimized cell may not be totally accurate since the mere bearing of a gene does not mean that it will be expressed, and the translation level of a gene can be highly variable. The use of absolute proteomics data sets has proven to be very informative in resource optimization, unfortunately, quantitative absolute proteomics is a quite challenging technique with a large equipment investment cost, therefore those proteome-wide data sets are scarce.

Using a data set with 95% of proteome coverage by mass (2300 genes approximately) leaves many genes outside of our analysis. Although those genes may have a reduced expression and therefore not a great contribution to the total saved resource, those genes may be of great importance to cellular fitness and should not be considered dispensable (Price et al., 2018). In addition, the minimization design should consider the trade-off between strain robustness and resource reduction, by accounting for the fitness contribution of the candidate genes to eliminate, such as the quasi-essential genes on the environment of interest. Genome reduction projects have been carried out in other *E. coli* backgrounds (non-K12), however, the lack of absolute proteomic information limits the application of our method to those strains (Umenhoffer et al., 2017). Absolute proteomic data sets, gene essentiality, cost and fitness contribution of each gene in a genome may provide the data needed to better design minimized strains with a resource allocation approach for *E. coli* strains of different lineage and other microbes of interest.

Improving the understanding of mRNA degradation can play a critical role in optimizing metabolic engineering strategies (Roux et al., 2021). Targeting mRNA degradation has been shown to enhance protein production without competing for cellular resources in the host strain (Mao & Inouye, 2012; Venturelli et al., 2017; Wu et al., 2020). However, due to the variability of mRNA degradation rates between genes, it can be challenging to model this process for an "average gene". Despite this, considering the impact of mRNA degradation on resource utilization is important in order to fully understand the effects of gene modification on cellular behaviour and potentially improve our calculation methods. Overall, mRNA degradation has great potential as a target for metabolic engineering to improve the performance of bacterial cell factories.

Large-scale fermentations generate stressful conditions, such as substrate availability gradients. The regulatory response to those stressful conditions has a cost, as it demands protein expression and it has been shown to increase ATP maintenance demands by 40%–50% (Löffler et al., 2016). A successful approach has been to target those specific stress responses that consume resources. The resulting minimized strains showed a lower maintenance coefficient and increase the production yield of GFP in simulated large-scale bioprocesses (Ziegler et al., 2021).

In this work, we provide a straightforward path to pursue when designing a strain with a higher potential to divert resources to a function of interest whilst performing fewer genetic interventions. Those resource-minimized strains should perform better as microbial production hosts than their non-improved counterparts (de la Cruz et al., 2020; Mizoguchi et al., 2008; Park et al., 2014).

AUTHOR CONTRIBUTIONS

Elisa Marquez-Zavala: Conceptualization (supporting); data curation (lead); investigation (lead); methodology

(equal); writing – original draft (equal); writing – review and editing (supporting). **José Utrilla:** Conceptualization (lead); formal analysis (supporting); funding acquisition (lead); methodology (equal); project administration (lead); supervision (lead); writing – original draft (equal); writing – review and editing (lead).

ACKNOWLEDGMENTS

Funding from DGAPA- UNAM- PAPIIT project IN213420 and CONACyT Ciencia de Frontera (grant no. 319352). Technical support of Jose Alfredo Hernandez and Victor del Moral. EMZ acknowledges the Programa de Maestría y Doctorado en Ciencias Bioquímicas UNAM and the master's scholarship 900742 from CONACyT.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Github repository with the code and source data to reproduce this analysis: https://github.com/utrillalab/Cell_Resource_Minimization_Strategies

ORCID

Elisa Marquez-Zavala https://orcid.org/0000-0001-8096-5280
Jose Utrilla https://orcid.org/0000-0003-3048-9241

REFERENCES

Choe, D., Lee, J.H., Yoo, M., Hwang, S., Sung, B.H., Cho, S. et al. (2019) Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nature Communications*, 10, 935.

de la Cruz, M., Ramírez, E.A., Sigala, J.-C., Utrilla, J. & Lara, A.R. (2020) Plasmid DNA production in proteome-reduced *Escherichia coli*. *Microorganisms*, 8, 1444.

Dennis, P.P. & Bremer, H. (2008) Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3, 1–49.

Fredens, J., Wang, K., de la Torre, D., Funke, L.F.H., Robertson, W.E., Christova, Y. et al. (2019) Total synthesis of *Escherichia coli* with a recoded genome. *Nature*, 569(7757), 514–518.

Glass, J.I., Merryman, C., Wise, K.S., Hutchison, C.A. & Smith, H.O. (2017) Minimal cells—real and imagined. *Cold Spring Harbor Perspectives in Biology*, 9, a023861.

Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K. et al. (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Molecular Microbiology*, 55, 137–149.

Hirokawa, Y., Kawano, H., Tanaka-masuda, K., Nakamura, N., Nakagawa, A., Ito, M. et al. (2013) Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli*. *Journal of Bioscience and Bioengineering*, 116, 52–58.

Hutchison, C.A., Chuang, R.-Y.R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H. et al. (2016) Design and synthesis of a minimal bacterial genome. *Science*, 351, aad6253.

Karcagi, I., Draskovits, G., Umenhoffer, K., Fekete, G., Kovács, K., Méhi, O. et al. (2016) Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining. *Molecular Biology and Evolution*, 33, 1257–1269.

Kepp, K.P. (2020) Survival of the cheapest: how proteome cost minimization drives evolution. *Quarterly Reviews of Biophysics*, 53, e7.

Kolisnychenko, V., Plunkett, G., Herring, C.D., Fehér, T., Pósfai, J., Blattner, F.R. et al. (2002) Engineering a reduced *Escherichia coli* genome. *Genome Research*, 12, 640–647.

Lastiri-Pancardo, G., Mercado-Hernández, J.S., Kim, J., Jiménez, J.I. & Utrilla, J. (2020) A quantitative method for proteome reallocation using minimal regulatory interventions. *Nature Chemical Biology*, 16, 1026–1033.

Lee, J.H., Sung, B.H., Kim, M.S., Blattner, F.R., Yoon, B.H., Kim, J.H. et al. (2009) Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. *Microbial Cell Factories*, 8, 1–12.

Lloyd, C.J., Ebrahim, A., Yang, L., King, Z.A., Catoiu, E., O'Brien, E.J. et al. (2018) COBRAme: a computational framework for genome-scale models of metabolism and gene expression. *PLoS Computational Biology*, 14, e1006302.

Löffler, M., Simen, J.D., Jäger, G., Schäferhoff, K., Freund, A. & Takors, R. (2016) Engineering E. coli for large-scale production–strategies considering ATP expenses and transcriptional responses. *Metabolic Engineering*, 38, 73–85.

Lynch, M. & Marinov, G.K. (2015) The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 15690–15695.

Mao, L. & Inouye, M. (2012) Use of *E. coli* for the production of a single protein. *Methods in Molecular Biology*, 899, 177–185.

Michalik, S., Reder, A., Richts, B., Faßhauer, P., Mäder, U., Pedreira, T. et al. (2021) The Bacillus subtilis minimal genome compendium. *ACS Synthetic Biology*, 10, 2767–2771.

Mizoguchi, H., Mori, H. & Fujio, T. (2007) *Escherichia coli* minimum genome factory. *Biotechnology and Applied Biochemistry*, 46, 157–167.

Mizoguchi, H., Sawano, Y., Kato, J.I. & Mori, H. (2008) Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. *DNA Research*, 15, 277–284.

Nakayasu, E.S., Chazin-Gray, A.M., Francis, R.M., Eaton, A.M., Auberry, D.L., Muñoz, N. et al. (2020) Resource reallocation in engineered *Escherichia coli* strains with reduced genomes. *bioRxiv*, 1–35. https://doi.org/10.1101/2020.10.19.346155

Park, M.K., Lee, S.H., Yang, K.S., Jung, S.C., Lee, J.H. & Kim, S.C. (2014) Enhancing recombinant protein production with an *Escherichia coli* host strain lacking insertion sequences. *Applied Microbiology and Biotechnology*, 9815(98), 6701–6713.

Pósfai, G., Plunkett, G., Fehér, T., Frisch, D., Keil, G.M., Umenhoffer, K. et al. (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science*, 312, 1044–1046.

Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Liu, H. et al. (2018) Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature*, 557, 503–509.

Roux, C., Etienne, T.A., Hajnsdorf, E., Ropers, D., Carpousis, A.J., Cocaign-Bousquet, M. et al. (2021) The essential role of mRNA degradation in understanding and engineering *E. coli* metabolism. *Biotechnology Advances*, 54, 107805.

Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D. & Karsch-Mizrachi, I. (2019) GenBank. *Nucleic Acids Research*, 47, D94–D99.

Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L. et al. (2016) The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology*, 34, 104–110.

Umenhoffer, K., Draskovits, G., Nyerges, Á., Karcagi, I., Bogos, B., Tímár, E. et al. (2017) Genome-wide abolishment of Mobile genetic elements using genome shuffling and CRISPR/Cas-assisted MAGE allows the efficient stabilization of a bacterial chassis. *ACS Synthetic Biology*, 6, 1471–1483.

Venturelli, O.S., Tei, M., Bauer, S., Chan, L.J.G., Petzold, C.J. & Arkin, A.P. (2017) Programming mRNA decay to modulate

synthetic circuit resource allocation. *Nature Communications*, 8, 1–11.

Vernyik, V., Karcagi, I., Tímár, E., Nagy, I., Györkei, Á., Papp, B. et al. (2020) Exploring the fitness benefits of genome reduction in *Escherichia coli* by a selection-driven approach. *Scientific Reports*, 10, 1–12.

Wagner, A. (2007) Energy costs constrain the evolution of gene expression. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 308, 322–324.

Wu, J., Bao, M., Duan, X., Zhou, P., Chen, C., Gao, J. et al. (2020) Developing a pathway-independent and full-autonomous global resource allocation strategy to dynamically switching phenotypic states. *Nature Communications*, 11, 1–14.

Ying, B.W., Seno, S., Kaneko, F., Matsuda, H. & Yomo, T. (2013) Multilevel comparative analysis of the contributions of genome reduction and heat shock to the *Escherichia coli* transcriptome. *BMC Genomics*, 14, 1–13.

Ying, B.W. & Yama, K. (2018) Gene expression order attributed to genome reduction and the steady cellular state in *Escherichia coli*. *Frontiers in Microbiology*, 9, 2255.

Yuan, X., Couto, J.M., Glidle, A., Song, Y., Sloan, W. & Yin, H. (2017) Single-cell microfluidics to study the effects of genome deletion on bacterial growth behavior. *ACS Synthetic Biology*, 6, 2219–2227.

Ziegler, M., Zieringer, J., Döring, C.L., Paul, L., Schaal, C. & Takors, R. (2021) Engineering of a robust *Escherichia coli* chassis and exploitation for large-scale production processes. *Metabolic Engineering*, 67, 75–87.

How to cite this article: Marquez-Zavala, E. & Utrilla, J. (2023) Engineering resource allocation in artificially minimized cells: Is genome reduction the best strategy? *Microbial Biotechnology*, 16, 990–999. Available from: https://doi.org/10.1111/1751-7915.14233