Lean-proteome strains – next step in metabolic engineering

Kaspar Valgepea¹, Karl Peebo¹,², Kaarel Adamberg¹,³ and Raivo Vilu¹,²*

¹ Competence Center of Food and Fermentation Technologies, Tallinn, Estonia
² Department of Chemistry, Tallinn University of Technology, Tallinn, Estonia
³ Department of Food Processing, Tallinn University of Technology, Tallinn, Estonia

*Correspondence: raivo@kkfi.ee

Edited by:
Hilal Taymaz Nikerel, Bogazici University, Turkey

Reviewed by:
Steve Van Dien, Genomatica, USA

Keywords: strain engineering, recombinant cells, absolute quantitative proteomics, genome engineering, Escherichia coli, genome reduction, chassis cells, whole-cell model

Rapid development of high-throughput -omics (e.g., proteomics) and genetic engineering technologies together with an array of new metabolic modeling tools during this century has led to the emergence of new fields of biological research termed systems biology and synthetic biology. The successful exploitation of these developments is evidenced by the creation of increasing number of genetically engineered recombinant cells with superior characteristics (Fantama et al., 2008; Becker et al., 2011) or totally novel functions (Nakamura and Whited, 2003; Yim et al., 2011; Paddock et al., 2013) for diverse sectors such as chemicals and healthcare (Huang et al., 2012; Lee et al., 2012; Sun and Alper, 2014). However, there exists a significant gap in bioprocess performance between studies of the literature and the requirements for an industrially feasible bioprocess for chemical production (Van Dien, 2013). Overall bioprocess performance [productivity (gram/liter/hour), titer (gram/liter) etc.] has to be increased further for successful industrial-scale commercialization to drive the shift from fossil fuel to bioprocess-based chemical production and cost-effective production of novel drugs (Van Dien, 2013). Hence, there is great need for novel approaches addressing these key challenges in chemical and healthcare sectors.

POTENTIAL OF PROTEOME OPTIMIZATION

With this opinion, we propose that a novel approach of proteome optimization carries a substantial potential for addressing the aforementioned challenges in bioprocess development. That potential arises from the fact that cells express proteins not essential (e.g., flagellar, heat or acid stress proteins) for growth under well-controlled optimal conditions, typically realized in biotechnological processes. This leads to non-efficient use of protein synthesis capacity (translation machinery) and energy for bioprocesses. As translation capacity is believed to be one of the growth-limiting factors, at least in the bacterium *Escherichia coli* (Klumpp et al., 2013), synthesis of non-essential proteins sequesters ribosomes potentially lowering the synthesis capacity of target molecule production. Thus removing the expression burden of non-essential proteins, i.e., creation of lean-proteome strains, could enable to specifically manipulate the allocation of ribosomes for higher synthesis of proteins leading to increased target molecule production. Optimization of the cellular genome through experimental testing of strains with optimized expression of non-essential proteins and inclusion of protein synthesis capacity constraints in metabolic modeling could open a new avenue for the creation of superior cell factories.

Initial experimental confirmation of the potential of optimization of the layer of protein synthesis capacity for increasing the maximum specific growth rate (μ*) of cells comes from two studies of *E. coli* investigating the effects of heterologous protein expression on μ (Scott et al., 2010; Biennick et al., 2014). Both studies show for several heterologous proteins (e.g., LacZ, eGFP) that increasing their expression has a linear negative effect on μ. Their data suggest that for expression of every 1% of heterologous protein per dry cell weight, μ decreases by ~3%. It would be sensible to assume that a similar correlation would exist for the opposite case – decreasing the fraction of non-essential proteins by 1% would lead to an increase in μ by ~3%. Our proposal is also supported by two studies of *Bacillus subtilis* showing that reducing the expression load of proteins non-essential under bioprocess conditions by ~9% fraction from the total proteome through the deletion of the flagellar/motility regulator gene sigD leads to a ~30% increase of both μ* and biomass yield (Fischer and Sauer, 2005; Munte et al., 2014). Further support comes from recent experiments of D’Souza et al. (2014), which show that deletion of single amino acid, vitamin, or nucleobase biosynthesis genes from *E. coli* results in higher μ* compared to the wild-type strain when both strains are grown on medium containing the amino acid, vitamin, or nucleobase that the deletion strain was auxotrophic for. These observations are consistent with earlier chemostat studies with *B. subtilis* (Zamenhof and Eichhorn, 1967) and *E. coli* (Dykhuizen, 1978) where mutants impaired in tryptophan biosynthesis demonstrate significant fitness advantages in the presence of tryptophan relative to prototrophic cells. More importantly, D’Souza et al. (2014) show that deleting genes with higher protein expression cost leads to a greater growth advantage.

The results presented above suggest that proteome resource optimization through decreasing the fraction of non-essential proteins could lead to faster growth and thus also to better bioprocess performance.
For instance, target molecule productivity could be increased in growth-coupled production processes by enabling faster growth at the same expression level(s) of target molecule production-related proteins. On the other hand, recombinant protein titers could be significantly elevated by allocating more proteome resources for target protein expression at the expense of lower synthesis of non-essential proteins even at the same μ and/or protein synthesis rate.

REDUCED-GENOME APPROACHES
A conceptually similar approach of creating reduced-genome strains for industrial purposes has been applied in few cases before (Pósfai et al., 2006; Mizoguchi et al., 2008; Unthan et al., 2014; Xue et al., 2014). However, these efforts concentrated on reducing the genome and neglected the effects of gene deletions on the cellular proteome.

The approach of deleting large chunks of the genome, instead of specific genes, based on gene function and not on protein abundance was probably responsible for the observed minor positive effects on cellular growth and target molecule production. While the latter studies focused on large-scale genome reduction, experimental technologies enabling more targeted and accurate engineering of strains with reduced load of gene expression have recently emerged. Hence, now the successful execution of the concept of targeted optimization of the layer of protein synthesis is feasible due to the recent rapid progress in proteome-wide absolute quantitative proteomics (Arike et al., 2012; Ahrné et al., 2013; Wiśniewski et al., 2014) and high-throughput genome engineering technologies [e.g., Multiplexed Automated Genome Engineering (MAGE; Wang et al., 2009), trackable multiplex recombineering (TRMR; Warner et al., 2010)]. Thus, the time is ripe to design and create lean-proteome strains possibly leading to superior bioprocess performance.

CHALLENGES WITH PROTEOME OPTIMIZATION
The main challenge with creating lean-proteome strains is hitting the correct genes/proteins, i.e., genes, which deletion does not lead to detrimental effects. This is a serious concern even in the most studied bacterium E. coli since functions for a third of its proteins are still unknown (Keseler et al., 2013) while only ~300 proteins are considered essential for E. coli (http://coliwiki.net/colipedia/index.php/Essential_genomes). It is important to point out that knowing functions/essentiality for more proteins is not the objective per se – it is actually much more important to know the functions/essentiality of the proteins with the biggest translational burden (abundance × length), as their deletion presumably leads to stronger effects. The good news here is that for many organisms, the proteome mass (a good proxy for length) distribution follows the Pareto principle – ~20% of proteins make up ~80% of the proteome mass (Ghaemmaghami et al., 2003; Maier et al., 2011; Schmidt et al., 2011; Valgepea et al., 2013). Thus, instead of targeting hundreds of genes/genome areas like in the reduced-genome approach described above, one could theoretically greatly increase the key metrics of bioprocess performance (titer, yield, productivity; Van Dien, 2013) by deleting as few as ~10 non-essential genes with the highest translational burden in E. coli (in total 7% of proteome; Valgepea et al., 2013) and substituting the “freed” 7% of the total proteome with target molecule-related proteins. Importantly, current mass-spectrometric techniques of absolute proteome quantification (Arike et al., 2012; Ahrné et al., 2013; Wiśniewski et al., 2014) are accurate enough to determine the proteins with the biggest translational burden on the whole-proteome level.

STRATEGIES OF PROTEOME OPTIMIZATION FOR CREATING LEAN-PROTEOME STRAINS
The first and most important step toward creating lean-proteome strains is absolute quantitative proteome analysis of the initial recombinant strain. Accurate characterization of the full proteome is needed for the compilation of lists of non-essential target proteins with the biggest translational burden. We propose two strategies for creating superior lean-proteome strains by targeting proteins with the biggest translational burden, currently specifically for E. coli:

1. The first strategy targets proteins with known functions and presumably unnecessary under optimal bioprocess conditions, e.g., pH, temperature, oxygen tension control; defined substrate feed; stirring. These could be proteins involved in stress responses (acid, heat, and osmotic shock), alternative substrate transport and catabolism and cellular movement (flagellar).

2. The second strategy targets proteins with unknown functions with the biggest translational burden. Beneficial for both approaches is the growth screen of all the Keio collection single (Baba et al., 2006) and double deletion strains (personal communication with Prof. Hirotada Mori) that can be used to determine the genes/proteins, which should and should not be targeted.

Another important step is the experimental construction of lean-proteome strains and selection for better production strains. Instead of reducing the proteome one protein at a time, one should target tens of genes with an approach similar to MAGE (Wang et al., 2009), which constantly generates genetic heterogeneity in the pool of mutants allowing the generation of thousands of lean-proteome strains within a few days. The challenge of selecting for better production strains could be tackled by combining several screening methods. First, one could screen for fast growth as reduction of non-essential protein expression should lead to faster growth. Second, high-producing strains could be isolated using fluorescence activated cell sorting (FACS) using a sensor system based on a fluorescent readout corresponding to target molecule levels.

POTENTIAL OF METABOLIC MODELING
Lastly, one would greatly benefit from an in silico metabolic model, which would enable quantitative prediction of the effects of removing non-essential proteins on target molecule production. This should be a model, which incorporates the cellular proteome with the two central features of regulation of μ – cell geometry and cell cycle – and ties the latter to the fluxes of flux balance analysis (FBA)-type models for in silico analysis and design of lean-proteome strains. Recently, we have seen serious progress into this direction by the development of a novel single-cell model (Abner et al., 2013), next-generation FBA-type of genome-scale
ACKNOWLEDGMENTS

The financial support for this work was provided by the European Regional Develop-
ment Fund project EU299994 and institu-
tional research (IUT 1927) and personal (G9192) funding of the Estonian Ministry of
Education and Research.

REFERENCES

Abner, K., Aaviksaar, T., Adamberg, K., and Vilu, R. (2014). Critical assessment of proteome-wide
label-free absolute abundance estimation strate-
gies. Proteomics 14, 2577–2578. doi:10.1002/prmc.201300135

Arike, L., Valgepea, K., Peil, L., Nahku, R., Adamberg,
K., and Vilu, R. (2012). Comparison and applica-
tions of label-free absolute proteome quantifi-
cation methods on Escherichia coli. J. Proteomics 75,
5437–5448. doi:10.1016/j.jprot.2012.06.020

Baba, T., Ara, T., Hasugawa, M., Takai, Y., Okumura, Y.,
Baba, M., et al. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants:
The Keio collection. Mol. Syst. Biol. 2, 206.0008.
doi:10.1038/msb4100050

Becker, J., Zelder, O., Hafner, S., Schröder, H., and
Wittmann, C. (2011). From zero to hero – design-based systems metabolic engineering of
Corynebacterium glutamicum for l-lysine produc-
tion. Metab. Eng. 13, 159–168. doi:10.1016/j.
ymben.2011.01.003

Bienick, M. S., Young, K. W., Klesmith, J. R., Detwiler,
E. E., Tornek, K. J., and Whitehead, T. A. (2014).
The interrelationship between promoter strength, gene
expression, and growth rate. PLoS ONE 9:e109105.
doi:10.1371/journal.pone.0109105

D’Souza, G., Waschyna, S., Pande, S., Bohi, K., Kaleta,
C., and Kost, C. (2014). Less is more: selective advantages can explain the prevalent loss of biosyn-
thetic genes in bacteria. Evolution 68, 2559–2570.
doi:10.1111/evol.12468

Dykhuizen, D. E. (1978). Selection for trypto-
phan auxotrophs of Escherichia coli in glucose-
limited chemostats as a test of energy-conservation
hypothesis. Evolution 32, 125–150. doi:10.2307/2407415

Fischer, E., and Sauer, U. (2005). Large-scale in vivo
flux analysis shows rigidity and suboptimal perfor-
man ce of Bacillus subtilis metabolism. Nat. Genet.
37, 636–640. doi:10.1038/ng1555

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson,
R. W., Belle, A., Dephoure, N., et al. (2003). Global
analysis of protein expression in yeast. Nature 425,
737–741. doi:10.1038/nature02466

Huang, C.-J., Lin, H., and Yang, X. (2012). Indus-
trial production of recombinant therapeutics in
Escherichia coli and its recent advancements. J.
Microbiol. Biotechnol. 39, 383–399. doi:10.1007/
s10295-011-1082-9

Jantama, K., Zhang, X., Moore, J. C., Shanmugam, K. T.,
Svornons, S. A., and Ingram, L. O. (2008). Eliminat-
ing side products and increasing succinate yields in
engineered strains of Escherichia coli C. Biotechnol.
Biopros. 101, 881–893. doi:10.1002/bt22005

Karr, J. R., Sanghvi, J. C., Mackin, D. N., Gutschow,
M. V., Jacobs, J. M., Bolival, B., et al. (2012). A whole-
cell computational model predicts pheno-
type from genotype. Cell 150, 389–401. doi:10.
1016/j.cell.2012.05.044

Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-
Zavalaeta, A., Gama-Castro, S., Bonavides-Martínez,
C., et al. (2013). EcoCyc: fusing model organism
databases with systems biology. Nucleic Acids Res.
41, D605–D612. doi:10.1093/nar/gks3027

Klumpp, S., Scott, M., Pedersen, S., and Hwa, T.
(2013). Molecular crowding limits translation and
cell growth. Proc. Natl. Acad. Sci. U.S.A. 110,
16754–16759. doi:10.1073/pnas.1310577110

Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., and
Lee, S. Y. (2012). Systems metabolic engineering
of microorganisms for natural and non-natural
chemicals. Nat. Chem. Biol. 8, 536–546. doi:10.
1038/nchembio.970

Liu, J. K., O’Brien, E. J., Lerman, J. A., Zengler, K.,
Pals, B. O, and Feist, A. M. (2014). Reconstruction,
and modeling protein translocation and compart-
mentalization in Escherichia coli at the genome-
scale. BMC Syst. Biol. 810. doi:10.1186/s12918-
014-0110-6

Maier, T., Schmidt, A., Gießel, M., Kühner, S., Gervin,
A. C., Abeler, R., et al. (2011). Quantification of
mRNA and protein and integration with pro-
tein turnover in a bacterium. Mol. Syst. Biol. 7, 511.
doi:10.1038/msb.2011.38

Mizoguchi, H., Sawano, Y., Kato, J., and Mori, H.
(2008). Superpositioning of deletions promotes
growth of Escherichia coli with a reduced genome.
DNA Res. 15, 277–284. doi:10.1093/dnares/dsn019
Muntel, J., Fromion, V., Goelzer, A., Maass, S.,
Mader, U., Buttner, K., et al. (2014). Compre-
sion, B. Ø, and Feist, A. M. (2014). Reconstruction
of mRNA and protein expression without spike-
ening of strains: from industrial-scale to lab-scale
chemical production. J. Ind. Microbiol. Biotechnol.
37, 32–43. doi:10.1007/s10295-014-1539-8

Unthan, S., Baumgart, M., Radek, A., Herbst, M.,
Siebert, D., Brühl, N., et al. (2014). Chassis organ-
ism from Corynebacterium glutamicum – a top-
down approach to identify and delete irrele-
vant gene clusters. Biotechnol. Biofuels 7, 100.
doi:10.1002/biot.201400041

Valgepea, K., Adamberg, K., Seiman, A., and Vilu,
R. (2013). Escherichia coli achieves faster growth
by increasing catalytic and translation rates of
proteins. Mol. Biosyst. 24, 2354–2358. doi:10.1039/
c3mb07119k

Von Dien, S. F. J. (2013). From the first drop to the
first truckload: commercialization of microbial
processes for renewable chemicals. Curr. Opin.
Biotechnol. 24, 1061–1068. doi:10.1016/j.copbio.
2013.03.002

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z.,
Xu, G., Forest, C. R., et al. (2009). Programming
cells by multiplex genome engineering and acceler-
ated evolution. Nature 460, 894–898. doi:10.1058/
nature08187

Warner, J. R., Reeder, P. J., Karpimoup-Fard, A.,
Woodruff, L. B. A., and Gill, R. T. (2010). Rapid
profiling of a microbial genome using mixtures of
barcoded oligonucleotides. Nat. Biotechnol. 28,
856–862. doi:10.1038/nbt.1653

Wieniewski, J. R., Hein, M. Y., Cox, J., and Mann,
M. (2014). ‘A protocistm rule’ for protein copy num-
er and concentration estimation without spike-
 standoffards. Mol. Cell. Proteomics 13, 3497–3506.
doi:10.1074/mcp.M113.037309

Xue, X., Wang, T., Jiang, P., Shao, Y., Zhou, M., Zhong,
L., et al. (2014). The MEGA (Multiple Essen-
tial Genes Assembling) deletion and replacement
method for genome reduction in Escherichia coli.
ACS Synth. Biol. 3, 1021–1025. doi:10.1021/sb50032ap

Yim, H., Haselbeck, R., Niu, W., Pujel-Baxley, C., Bur-
gard, A., Boldt, J., et al. (2011). Metabolic engi-
neering of Escherichia coli for direct production of
1,4-butanediol. Nat. Chem. Biol. 7, 445–452. doi:10.
1038/nchembio.580
Zamenhof, S., and Eichhorn, H. H. (1967). Study of microbial evolution through loss of biosynthetic functions – establishment of defective mutants. Nature 26, 456–458. doi:10.1038/216456a0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 January 2015; accepted: 22 January 2015; published online: 06 February 2015.

Citation: Valgepea K, Peebo K, Adamberg K and Vilu R (2015) Lean-proteome strains – next step in metabolic engineering. Front. Bioeng. Biotechnol. 3:11. doi: 10.3389/fbioe.2015.00011

This article was submitted to Systems Biology, a section of the journal Frontiers in Bioengineering and Biotechnology.

Copyright © 2015 Valgepea, Peebo, Adamberg and Vilu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.