UNIVERSITY OF CALIFORNIA, SAN DIEGO

Enhancing mammalian bioprocessing through engineering Chinese hamster ovary cell metabolism

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy

in

Bioengineering

by

Hooman Hefzi

Committee in charge:

Professor Nathan E. Lewis, Chair
Professor Bernhard Ø. Palsson, Co-Chair
Professor Randal Kaufman
Professor Prashant Mali
Professor Christian Metallo
Professor Maho Niwa

2018

The dissertation of Hooman Hefzi is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California, San Diego

2018

Epigraph

Mathematics would certainly have not come into existence if one had known from the beginning that there was in nature no exactly straight line, no actual circle, no absolute magnitude.

Friedrich Nietzsche

Table of Contents

List of Figures

List of Tables

Acknowledgements

This work would not have been possible without the influence and support of countless people.

First and foremost, I am thankful for the guidance and friendship of Professor Nathan Lewis. I have often heard that the Ph.D. experience can be terrible but can say that under his supervision my studies have been truly enjoyable. Thank you for giving me the independence to explore ‘out there’ ideas and supporting me throughout my time at UCSD. I could not have asked for a better advisor. I am also grateful for Professor Bernhard Palsson. Had I not taken his undergraduate course when I did, I likely would have abandoned science altogether. I will always be appreciative of your ability to provoke deep scientific inquiry. I would be remiss in not thanking the rest of my dissertation committee for their time and agreeing to oversee this work.

In the course of my Ph.D., I was fortunate enough to spend time at the Novo Nordisk Foundation Center for Biosustainability at the Denmark Technical University to conduct experiments. Without the support of Helene Kildegaard and Bjørn Voldborg, this would not have been possible.

While in Denmark I had the pleasure of working with a truly outstanding group of individuals. Much of the work presented here would not be possible without their contributions on both a scientific and personal level.

Stefan Kol and Daniel Ley: My ‘CHO bros’. Thank you for the good times in Copenhagen and Valencia. Being able to work with you has been a pleasure.

Lise Grav: A true savior, without you I’m sure many more of my experiments would have failed. Spending all those summer days wandering the city staring at our phones is a time I will look back on fondly. Even though you couldn’t make it out for a proper stay, I’m glad we got to have at least one US adventure.

Julie Karottki: Pretty sure I would have gone insane in the winter without you. Even though we are basically complete opposites, the fact that Venus orbits the Sun and we do too completely why we're so close. Tak for hygge and everything else.

Lasse Pedersen: Very glad that you were sent to UCSD. From discussions about Danish politics to the Warburg effect to Swoop! there was never a dull moment with you.

In no particular order, I would also like to thank Thomas Amann, Kristina Andersen, Johnny Arnsdorf, Sara Bjørn, Karen Brøndum, Marianne Decker, Karoline Fremming, Tae Kwang Ha, Anders Hansen, Maria Javidi, Kristian Jensen, Thomas Kallehauge, Kim Marquart, Ivan Monge, Sara Pereira, Helle Petersen, Nachon Petersen, Nuša Pristovšek, Lars Schrübbers, Daria Sergeeva, Ankita Singh, Zulfiya Sukhova, and Tune Wulff for an amazing time academically and socially.

The scientific conversations I had with Gyun Min Lee, Lars Nielsen, Helene Kildegaard, and Bjørn Voldborg over the years have also been invaluable. Your insight and expertise is much appreciated.

At UCSD I was part of an incredible lab. All the CHOmigos, old and new, have been incredible to work—and play—with. Moji, Anne, Joanne, Austin, Curtis, Chintan, Isaac, Omar, Vahid, Matt, Bokan, Jeremy, Alex, thanks for being a part of the journey. Ben, your legal based pun will never be forgotten. Jahir, Shangzhong, Phil, we were the original CHOmigos and it has been a fun ride. Thank you for all your help and the Tacos el Gordo adventures. I also had the good fortune of working with many in the Palsson lab. I am especially appreciative of Aarash and Dan for valuable scientific discussions.

I am also indebted all my coauthors on the CHO model paper (without whom this herculean effort would not have been successful) and Soo Min Noh for her work on the Warburg study.

To all my friends, thank you for reminding me that there is more to life than school.

Last but certainly not least, I am deeply grateful for my family. They have always been there for me and my strongest supporters. Mom, Dad, Aunita (and everyone else) thank you for everything.

Chapter 1, in part, is a modified version of the material as it appears in Hefzi, H., Lewis, N.E. From random mutagenesis to systems biology in metabolic engineering of mammalian cells. *Pharmaceutical Bioprocessing*, 2:355-358 (2014). I was the primary author, while the co-author participated in and supervised the drafting of this review.

Chapter 2, in full, is a modified version of the material as it appears in Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., Martinez, V.S., Kyriakopoulos, S., Jimenez, N.E., Zielinski, D.D., Quek, L.E., Wulff, T., Arnsdorf, J., Li, S., Lee, J.S., Paglia, G., Loira, N., Spahn, P.N., Pedersen, L.E., Gutierrez, J.M., King, Z.A., Lund, A.M., Nagarajan, H., Thomas, A., Abdel-Haleem, A.M., Zanghellini, J., Kildegaard, H.F., Voldborg, B.G., Gerdtzen, Z.P., Betenbaugh, M.J., Palsson, B.O., Andersen, M.R., Nielsen, L.K., Borth, N., Lee, D.Y., Lewis, N.E. A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary (CHO) Cell Metabolism. *Cell Systems*, 3, 434-443 (2016). I was a co-first author, while the co-authors provided support in the research underlying the study.

Chapter 3 and 4 contain material from Hefzi, H., Noh, S.M., Monge, I.M., Decker, M., Arnsdorf, J., Kol, S., Pristovšek, N., Hansen, A.H., Bjorn, S.P., Brøndum, K.K., Javidi, E.M., Jensen, K.L., Kallehauge, T.B., Ley, D., Ménard, P., Petersen, H.M., Sukhova, Z., Voldborg, B.G., Nielsen, L.K., Lee, G.M., Kildegaard, H.F., Lewis, N.E. Multiplex genome editing eliminates the Warburg Effect without affecting oxidative metabolism or growth rate. (Submitted). I was the primary author, while the co-authors provided support in the research underlying the study.

Vita

EDUCATION

2011 Bachelor of Science, Bioengineering: Biotechnology | University of California, San Diego

2018 Doctor of Philosophy, Bioengineering | University of California, San Diego

PUBLICATIONS

11. Ley, D., Pereira, S., Pedersen, L.E., Petersen, S.B., Arnsdorf, J., Hefzi, H., Lund, A.M., Ha, T.K., Wulff, T., Kildegaard, H.F., Andersen, M.R. Reprogramming Amino Acid Catabolism in CHO Cells With CRISPR/Cas9 Genome Editing Improves Cell Growth and Reduces By-Product Secretion (In preparation).
10. Hefzi, H., Noh, S.M., Monge, I.M., Decker, M., Arnsdorf, J., Kol, S., Pristovšek, N., Hansen, A.H., Bjorn, S.P., Brøndum, K.K., Javidi, E.M., Jensen, K.L., Kallehauge, T.B., Ley, D., Ménard, P., Petersen, H.M., Sukhova, Z., Voldborg, B.G., Nielsen, L.K., Lee, G.M., Kildegaard, H.F., Lewis, N.E. Multiplex genome editing eliminates the Warburg Effect without affecting oxidative metabolism or growth rate. (Under review at Nature Biotechnology).
9. Abdel-Haleem, A.M., Hefzi, H., Mineta, K., Gao, X., Gojobori, T., Palsson, B.O., Lewis, N.E., Jamshidi, N. Functional interrogation of Plasmodium genus metabolism identifies species- and stage-specific differences in nutrient essentiality and drug targeting. *PLoS Computational Biology*, 14(1): e1005895 (2017).
8. Hastings, J., Mains, A., Sanz, M.A., Bergmann, S., Braeckman, B., Bundy, J., Cabreiro, F., Dobson, P., Ebert, P., Hattwell, J., Hefzi, H., Houtkooper, R., Jelier, R., Joshi, C., Kothamachu, V., Lewis, N.E., Lourenço, A., Nie, Y., Norvaisas, P., Pearce, J., Riccio, C., Rodriguez, N., Santermans, T., Scarcia, P., Schirra, H.J., Sheng, M., Smith, R., Suriyalaksh, M., Towbin, B., Tuli, M.A., van Weeghel, M., Weinkove, D., Zečić, A. ,Zimmerman, J., le Novère N., Kaleta, C., Witting, M., Casanueva, O. WormJam: A consensus C. elegans Metabolic Reconstruction and Metabolomics Community and Workshop Series. *Worm*, 6(2): e1373939 (2017).
7. Brunk, E.*, Chang, R.L.*, Xia, J., Hefzi, H., Yurkovich, J., Kim, D., Buckmiller, E., Wang, H., Yang, C., Palsson, B.O., Church, G.M., Lewis, N.E. Systemic post-translational control of bacterial metabolism regulates adaptation in dynamic environments. bioRxiv 180646 (2017).
6. Hefzi, H.*, Ang, K.S.*, Hanscho, M.*, Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., Martinez, V.S., Kyriakopoulos, S., Jimenez, N.E., Zielinski, D.D., Quek, L.E., Wulff, T., Arnsdorf, J., Li, S., Lee, J.S., Paglia, G., Loira, N., Spahn, P.N., Pedersen, L.E., Gutierrez, J.M., King, Z.A., Lund, A.M., Nagarajan, H., Thomas, A., Abdel-Haleem, A.M., Zanghellini, J., Kildegaard, H.F., Voldborg, B.G., Gerdtzen, Z.P., Betenbaugh, M.J., Palsson, B.O., Andersen, M.R., Nielsen, L.K., Borth, N., Lee, D.Y., Lewis, N.E. A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary (CHO) Cell Metabolism. *Cell Systems*, 3, 434-443 (2016).
5. Swainston, N., Smallbone, K., Hefzi, H., Dobson, P.D., Brewer, J., Hanscho, M., Zielinski, D.C., Ang, K.S., Gardiner, N.J., Gutierrez, J.M., Kyriakopoulos, S., Lakshmanan, M., Li, S., Liu, J.K., Martinez, V.S., Orellana, C.A., Quek, L.E., Thomas, A., Zanghellini, J., Borth, N., Lee, D.Y., Nielsen, L.K., Kell, D.B., Lewis, N.E., Mendes, P. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*, 12:109 (2016).
4. Golabgir, A.*, Gutierrez, J.M.*, Hefzi, H., Li, S., Palsson, B.O., Herwig, C., Lewis, N.E. Quantitative feature extraction from the Chinese Hamster Ovary bioprocess bibliome using a novel meta-analysis workflow. *Biotechnology Advances*, 34(5):621–633 (2016).
3. Hefzi, H., Lewis, N.E. From random mutagenesis to systems biology in metabolic engineering of mammalian cells. *Pharmaceutical Bioprocessing*. 2:355-358 (2014).

2. Hefzi, H., Palsson, B.O., Lewis, N.E. Reconstruction of genome-scale metabolic networks. In “Handbook of Systems Biology” (2012).
1. Ghosh, S., Larson, S., Hefzi, H., Marnoy, Z., Cutforth, T., Dokka, K., Baldwin, K.K. Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature,* (472):217-220 (2011).

* These authors contributed equally

PATENTS

1. **Hefzi, H.**, Lewis, N.E. Mammalian cells devoid of lactate dehydrogenase activity. Patent filed.

Abstract of the Dissertation

Enhancing mammalian bioprocessing through engineering
Chinese hamster ovary cell metabolism

by

Hooman Hefzi

Doctor of Philosophy in Bioengineering

University of California, San Diego, 2018

Professor Nathan E. Lewis, Chair
Professor Bernhard Ø. Palsson, Co-Chair

Chinese hamster ovary (CHO) cells are the dominant host for producing biotherapeutic proteins. Despite a long history of use, the techniques used to generate high quantities of recombinant drugs from this cell line remain largely unchanged. This is—at least partly—due to the complexity of mammalian cell lines. Historically—in contrast to microbial systems—CHO cells

have been viewed as a bioprocessing black box, a system to optimize around rather than engineer. In this dissertation, advancements in gene editing tools and the availability of a genomic sequence are used to establish a framework to move away from that perspective and enable rational engineering of CHO cells for desirable phenotypes. Specifically, a genome-scale model of CHO cell metabolism, iCHO1766, was reconstructed from the genome sequence. iCHO1766 contains the biochemical basis for growth and protein production in CHO cells and predicts growth rates as well as CHO-specific amino acid auxotrophies. This model was also deployed to identify strategies that efficiently redirect resources from growth to protein production. Using the recently characterized CRISPR-Cas9 system, we were able to eliminate the Warburg effect in CHO cells by simultaneously knocking out lactate dehydrogenase and regulators involved in a negative feedback loop that typically inhibits pyruvate conversion to acetyl-CoA. In contrast to long-standing assumptions about the role of aerobic glycolysis, Warburg-null cells maintain the same growth rate as wildtype cells while consuming less glucose without increasing oxygen uptake to compensate for lost glycolytic ATP. The cells produce negligible lactate—allowing prolonged growth to higher cell densities—and remain viable for generating recombinant protein producing cell lines. Introducing this phenotype into a CHO cell line already producing a biotherapeutic antibody maintained protein production and improved glycan galactosylation. Thus, by leveraging accessible gene editing techniques, an improved CHO cell line has been generated. With newly available systems biology models, the tools for further rational engineering toward better and more affordable biotherapeutics are now available.

Chapter 1: Microbial engineering provides a roadmap for engineering biopharmaceutical producing mammalian cells

Global annual sales of pharmaceuticals have nearly crested one trillion dollars[1]. While chemical synthesis of small molecules remains the dominant fraction of sales, biologics represent an ever-growing sector of the industry[2], and are disproportionately represented in the highest grossing pharmaceuticals (Table 1.1).

Table 1.1: Annual sales of top 10 pharmaceutical products in 2016.
Italicized products are biopharmaceuticals produced in a recombinant host (vs. small molecules synthesized chemically). Bolded products are produced in mammalian cells. Data from Lindsley[1].

Pharmaceutical	2016 Sales (US $ BN)
Humira	***16.078***
Harvoni	9.081
Enbrel	***8.874***
Rituxan	***8.583***
Remicade	***7.829***
Revlimid	6.974
Avastin	***6.752***
Herceptin	***6.751***
Lantus	*6.054*
Prevnar 13	*5.718*

This class of compounds is characterized by their complexity—containing orders of magnitude greater atoms than small molecules such as aspirin—thus necessitating the utilization of cellular hosts for their synthesis (Table 1.2).

Table 1.2: Example pharmaceutical compounds.
Pharmaceutical compounds differ greatly in size and complexity, necessitating different modes of production.

Compound	Number of Atoms	Glycosylation Required?	Preferred Mode of Production
Aspirin	21	No	Chemical synthesis
Insulin	~1,700 (110 AA)	No	*E. coli* or yeast
Erythropoietin	>3,000 (193 AA)	Yes	Mammalian
IgG (antibody)	>20,000 (1370 AA)	Yes	Mammalian
Factor VIII	>37,000 (2351 AA)	Yes	Mammalian

While simpler biologics like insulin can be synthesized in microbial hosts such as *E. coli* or yeast, the majority of this class of compound is generated in mammalian cells[3]—largely to ensure that critical post-translational processes (e.g., glycosylation) are preserved, permitting proper protein folding, and compatibility with human patients[4].

A combination of historical fortuity and practicality has led to an ovarian cell line (CHO) derived from *Cricetulus griseus* (the Chinese hamster) being the primary workhorse of the biopharmaceutical industry. The isolation of dihydrofolate reductase (*Dhfr*)-deficient CHO cells in the early 1980s[5] coincided with the development of a *Dhfr*-linked method for expressing heterologous genes in mammalian cells. Adaptation of the Axel patents' method of cotransformation[6] turned CHO into viable hosts for producing complex biologics; their primacy in the field (over other mammalian lines) is a result of several factors:

1. CHO cells were the production line used for the first FDA approved recombinant protein made in a mammalian line, easing the regulatory burden of subsequent endeavors.
2. As a result of (1), there is a strong history of bioprocess development around CHO cells[7].
3. CHO cells exhibit decreased viral susceptibility due to not expressing many genes needed for viral entry or replication[8].
4. In comparison to other non-human mammalian lines, CHO does not produce immunogenic N-glycolylneuraminic acid containing glycan structures[9,10].

Accordingly, the majority of biotherapeutic proteins are produced using CHO cell lines[11].

Ultimately, the processes underpinning biotherapeutic protein production (e.g., translation, glycosylation, cell growth, etc.) are fueled by metabolism (Figure 1.1).

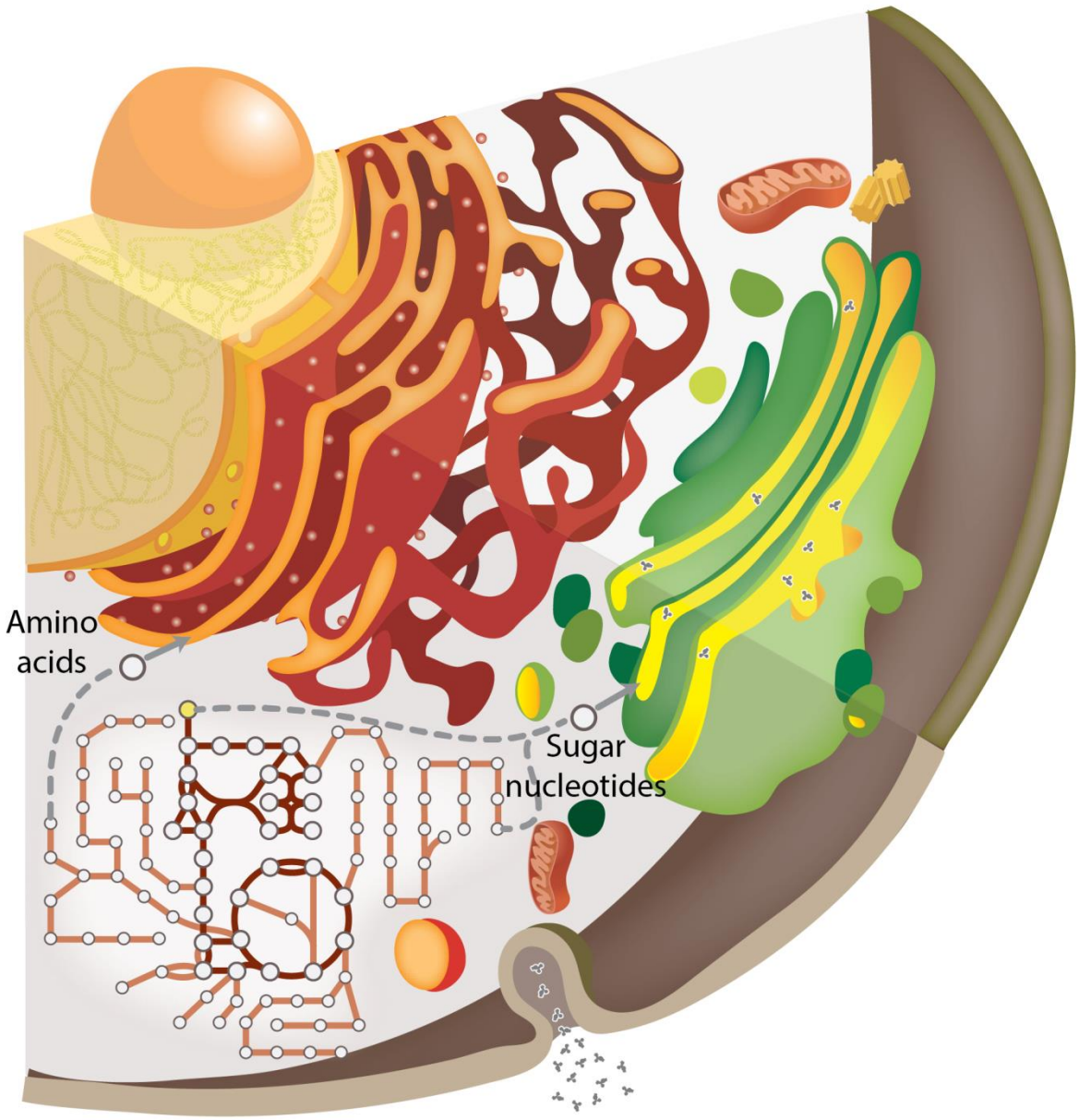


Figure 1.1: Metabolism fuels multiple processes underlying protein secretion.
The metabolic network synthesizes the amino acids and sugar nucleotides used for protein translation and glycosylation, respectively.

While certain parts of the metabolic network, such as glycolysis, the pentose phosphate pathway, and TCA cycle are commonly taught in basic scientific curriculum, these pathways represent only a fraction of the metabolic capabilities present in the vast majority of cells. Indeed, beyond generation of energy molecules (e.g., ATP, NADH, NADPH, etc.), dividing cells must be able to synthesize all components needed for growth, such as lipids, proteins, DNA, and RNA. Thus, the building blocks for these macromolecules—fatty acids, amino acids, deoxynucleotides, and nucleotides, respectively—must be produced or obtained by cells in the course of culture. It is worth nothing that CHO cells cannot synthesize 12 of the 20 amino acids[12–15]. As a result, culture medium will contain—at minimum—glucose, glutamine, and these 12 amino acids, which must be used by metabolic pathways to fuel cellular growth and other activities. The same pathways, especially those involved in amino acid and nucleotide sugar metabolism, must also provide the materials needed to produce recombinant biopharmaceuticals. From a broader perspective then, utilizing CHO cells for protein production is a delicate balancing act of shifting metabolic resources from growth and other 'desirable' (to the cell) activities to production of a heterologous protein without any benefit to the cell—a challenging task where rational metabolic engineering may prove useful.

Metabolic engineering is rapidly developing, with a continuous stream of technological developments that are being employed to expand the portfolio of molecules produced in cell factories. For commodity chemical production (e.g. amino acids, biofuels, etc.), metabolic engineering has progressed through three phases[16]. Initially, biological products were obtained through random mutagenesis of production strains and large screening efforts. Improved microbial strains could be isolated, but mechanisms underlying the desired phenotype were often poorly understood[17]. Diverse molecular biology techniques facilitated the second phase, in which simple, intuitive modifications were made. The third phase now employs systems biology techniques to understand the effect of modifications on all other metabolic pathways and on cell physiology. Thus, we have entered an era in which metabolic engineering aims to improve

microbial strains in a reproducible fashion, using complex designs based on detailed biochemical knowledge and computational model simulations. Here we highlight the historical progression towards using systems biology in microbial metabolic engineering and compare this to the current status of mammalian production cell line development. We also discuss the unique challenges in engineering mammalian cell lines for biotherapeutic production and outline how systems biology can facilitate metabolic engineering efforts for these platforms.

The systems biology approach to metabolic engineering has been enabled by three primary advancements: whole-genome sequencing, gene editing tools, and genome-scale models of cellular metabolism. The completion of the *Escherichia coli* K-12 genome sequencing effort in 1997[18] provided a comprehensive parts list for targeted metabolic engineering and expanded the scope of our understanding of the machinery within this microbe. The further development of efficient genetic modification systems, such as the lambda Red recombination system[19], enabled the deployment of targeted metabolic engineering designs, such as the removal of competing pathways that divert flux away from the formation of a desired product. Predictions of the systemic effects of genetic modifications were enabled when the information in the sequenced genome was harnessed for the development of genome-scale models of metabolism[20]. These models contain all known biochemical reactions in a cell, thus allowing one to predict the overall impact of modifications on phenotypic traits such as growth rate and small molecule secretion.

Systems biology approaches are now important tools in microbial metabolic engineering. Yim et. al.[21] genetically modified *E. coli* to produce 1,4-butanediol (BDO) by introducing heterologous genes to allow for biosynthesis of the non-native molecule. Further strain optimization was carried out by using a genome-scale model to identify genetic modifications that couple production of BDO to cell growth. A quadruple knockout strategy was identified by this *in silico* approach, and was implemented in conjunction with additional enzyme engineering, resulting in a host cell line that could produce high yields of BDO. Following the same principles,

Curran et al.[22] produced muconic acid in *Saccharomyces cerevisiae* by introducing heterologous genes, deleting glucose-6-phosphate dehydrogenase, and overexpressing transketolase. These modifications shunted metabolic flux towards the desired product, resulting in a twofold increase in final titer. These are just two of many examples in which modeling and genome editing are being employed together in metabolic engineering in academic and industrial settings.

While similar approaches have the potential to be applied to biotherapeutic protein production, efforts to do so have been limited. The existing norm in generation of production CHO cell lines for biotherapeutics relies on random integration of a construct consisting of a selection marker along with the gene of interest, followed by amplification of the construct using a drug. Such approaches are effective in generating high-producing cell lines, but random integration and the accumulation of uncharacterized mutations make reproducibility difficult and the elucidation of precise mechanisms behind high productivity nearly impossible. Current metabolic engineering efforts in mammalian cell lines are focused on supplementing the drug-selected strain's properties with intuitive genetic modifications to gain desirable traits. For example, buildup of malate in the medium of CHO culture was alleviated by overexpression of malate dehydrogenase[23]. Similar logic was used in decreasing lactate production by decreasing lactate dehydrogenase and increasing pyruvate dehydrogenase activities[24], but little has been done in understanding this cell line from a systemic perspective.

Fortunately, the same drivers that have allowed systems approaches in microbial metabolic engineering have recently been developed for CHO. The genomic sequence of CHO and the Chinese hamster have been published[25,26] and the accompanying genome annotations are providing a comprehensive catalog of enzymes in CHO. These resources are enabling the construction of a genome-scale metabolic network model, furthering previous efforts built upon murine models[27,28]. Novel genome editing technologies such as transcription activator-like effector nucleases (TALENs), zinc-finger nucleases, and the CRISPR/Cas9 system now allow targeted

genetic modification of mammalian cells lines[29]. Taken together these tools should enable a more systemic approach towards engineering mammalian cell lines.

The leap toward systems biology-based engineering of CHO will be an important one and this effort comes with several opportunities. However, it is important to note that the nature of the desired product (almost exclusively recombinant proteins) and the use of a mammalian production host raise their own unique set of challenges. First, the tradeoff between growth and recombinant protein production is much more significant than the tradeoff for production of small molecules. This is because recombinant proteins directly drain protein synthesis (the largest constituent of cell mass) and make growth-coupled design incredibly difficult. However, systems biology models can potentially elucidate non-intuitive interventions that enhance resource allocation to protein synthesis. Second, microbial engineering often addresses stoichiometric limitations on product formation after ensuring that necessary enzymes are expressed. Biotherapeutic protein production, however, further requires that the cell is capable of properly folding, post-translationally modifying, and secreting the non-native product; each step can be a bottleneck[30]. These additional requirements can be met in large part by the development of models of the protein secretion pathway, and the use of such models to analyze -omics measurements from individual host cell lines that exhibit differences in secreted protein yield. Third, it can be difficult to produce a homogeneous product since multiple glycoforms are often produced for a given recombinant protein. Analyses of glycomic data in the context of models of glycosylation may remedy this as they provide a more mechanistic understanding of the factors leading to diversity among glycoforms. Fourth, metabolic enzymes in mammals are far less defined and studied than their microbial counterparts and cell line properties (e.g. gene expression) that vary from tissue to tissue must be accounted for. However, detailed studies of metabolic enzymes and phenotypic assessment of mutations will help remedy these challenges over time. Finally, while techniques exist to construct and analyze metabolic models for mammalian cell lines—and even account for cell line specific knowledge—the unique properties of recombinant protein production suggest

that some of the exact approaches used in microbial metabolic engineering (e.g. coupling production to growth or elimination of competing pathways) may be difficult to implement. Fortunately, novel approaches complementing existing techniques targeted at mammalian systems are being developed to specifically address these challenges.

There will be certain areas in which established methods in genome-scale modeling will be particularly informative. Indeed, constraint-based modeling with genome-scale models is powerful at predicting growth and production capabilities. A plethora of *in silico* methods have been developed for use with these models[31] and will prove useful for bioprocess optimization. Additionally, methods for expanding these models to account for processes such as transcription/translation[32] or protein secretion[33] have been developed and can guide the expansion of CHO models to allow for analysis of specific areas of importance, including those discussed above. The biochemical knowledge contained within genome-scale models is also informative for understanding the mechanistic basis behind relationships observed in various -omics datasets and should shed light on potential areas for cell line improvement[34]. Ultimately, we envision these models as tools to supplement established techniques for strain development by interpreting existing and newly generated -omic data in order to pinpoint bottlenecks, metabolic or otherwise, and identify actionable solutions.

The remainder of this dissertation will discuss important developments on the path to rationally engineered CHO cell lines. Chapter 2 highlights the construction of a consensus genome-scale model of CHO cell metabolism, which revealed insights into how secretory pathway engineering can *efficiently* shift resource utilization into protein production. Chapter 3 shows an application of CRISPR-Cas9 gene editing to eliminate the Warburg effect, a near universal phenomenon in proliferating mammalian cells—and problem plaguing the field of mammalian bioprocessing for decades. Chapter 4 expands the characterization of Warburg-null CHO cell lines to evaluate their viability for continued use as a biopharmaceutical production host. Finally, the importance of these developments—both for the biopharmaceutical industry and more

broadly—will be reviewed; a logical path forward for further engineering and exploration of Warburg-null cell lines will also be proposed.

Chapter 1, in part, is a modified version of the material as it appears in Hefzi, H., Lewis, N.E. From random mutagenesis to systems biology in metabolic engineering of mammalian cells. *Pharmaceutical Bioprocessing*, 2:355-358 (2014). I was the primary author, while the co-author participated in and supervised the drafting of this review.

Chapter 2: A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism

INTRODUCTION

Since their first commercial use in the late 1980s to produce tissue plasminogen activator, Chinese hamster ovary (CHO) cell lines have remained the platform of choice for producing proteins requiring complex post-translational modifications for therapeutic activity and regulatory approval[34]. Over the years, dramatic increases in product titer have been achieved in CHO cells as the result of bioprocess optimizations that increased cell culture density and longevity[7], resulting in CHO being the dominant host cell line for biotherapeutic production[11]. Despite these achievements, the molecular basis of protein production in CHO cells remains poorly characterized. Recent access to genome sequences[8,25,26] and advances in systems biology[35] now enable the construction of a mechanistic basis for growth and protein production in CHO cells.

Three key cellular processes drive recombinant protein production: transgene expression, metabolism, and protein secretion. Metabolism is particularly important and inexorably linked to the others. For example, metabolic enzymes, including dihydrofolate reductase[36] and glutamine synthetase[37], have served as selection systems for transfecting and amplifying transgenes in CHO cells. Additionally, metabolism provides the building blocks for the protein product and the secretory machinery needed to secrete it. Cell metabolism has been modulated extensively in the enhancement of CHO-based bioprocessing. Specifically, the balance of cellular metabolic demands has been targeted through media optimization to improve cell density, growth, and product yields[38]. Efforts have also reduced the secretion of undesirable byproducts (e.g., lactate and NH_3) to ameliorate the impact on cell growth[39], product quality[40], and the cellular metabolic state[41]. Additionally, metabolism influences product quality attributes (e.g., drug efficacy and compatibility with the human immune system) including glycosylation[42], oxidation, acetylation, and disulfide bridge formation[43]. Intuitive modifications of metabolic enzyme levels have improved

protein production and quality[44]; however, since each enzyme contributes to pathways, imbalances of components and interactions between pathways can yield unexpected results. Thus, a more complete understanding of CHO metabolism is vital to identify metabolic bottlenecks in CHO cell culture and to rationally guide complex cell engineering efforts.

To cope with the complexity of CHO metabolism, computational models have been applied to study CHO under various conditions[27,45–49]. Studies have focused primarily on central metabolism[45] or used models extrapolated from mice[27,50,51]. However, CHO-specific genome-scale metabolic models (GeMs) are now within reach, given the recent sequencing of the CHO-K1 and Chinese hamster genomes[8,25,26]. GeMs[31] contain detailed information about all known biochemical reactions in a specific organism based on its genome and physiological information. Since metabolic pathways synthesize the components necessary for growth and survival, these models link the genetic basis of a cell to phenotypic capabilities, allowing more precise and complex metabolic engineering efforts[22,35].

Here we present a genome-scale metabolic network reconstruction for CHO cells that specifically links the genes encoded by the CHO-K1 and hamster genome to growth and recombinant protein production. This network was constructed and carefully curated by dozens of researchers in the community, and delineates the genetic basis of the metabolic pathways fueling all cell functions in CHO. We further built specific models for the CHO-K1, -S, and -DG44 cell lines and demonstrate that the models accurately predict important phenotypes such as growth rate, and unique metabolic features of CHO cells (e.g., auxotrophies). Using these models, we analyzed the metabolic impact of common bioprocess treatments that aim to increase cell productivity. We found that while the treatments increased product titers by liberating cell resources, the cells failed to efficiently redirect metabolic precursors towards recombinant protein production after treatment. However, targeted engineering efforts showed more efficient redirection. Differences between treatments in redirection efficacy highlight potential avenues for

further engineering efforts. Thus, the genome-scale network of CHO metabolism is an invaluable tool for data analysis, bioprocess optimization, and CHO cell engineering efforts.

RESULTS

A community genome-scale metabolic network reconstruction of Cricetulus griseus

Draft *C. griseus* metabolic models were reconstructed independently by multiple research groups. Subsequently, these models were merged using a systematic reconciliation process to capture all the careful curation that had gone into each model by groups representing diverse areas of expertise (Figure 2.1). As a result, we constructed and present a community consensus model for CHO metabolism and protein secretion.

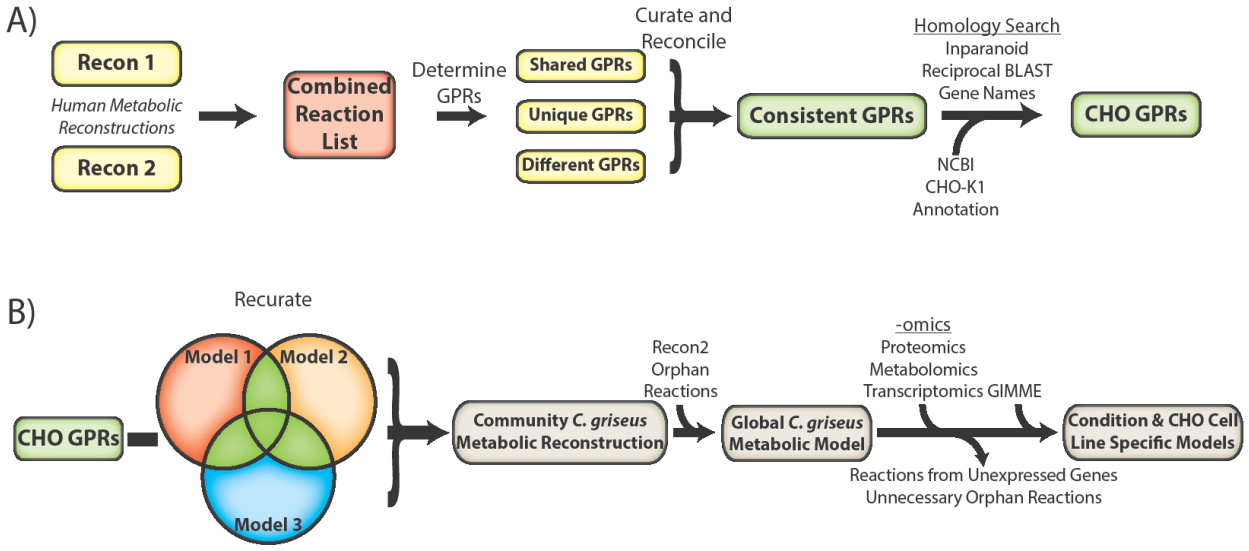
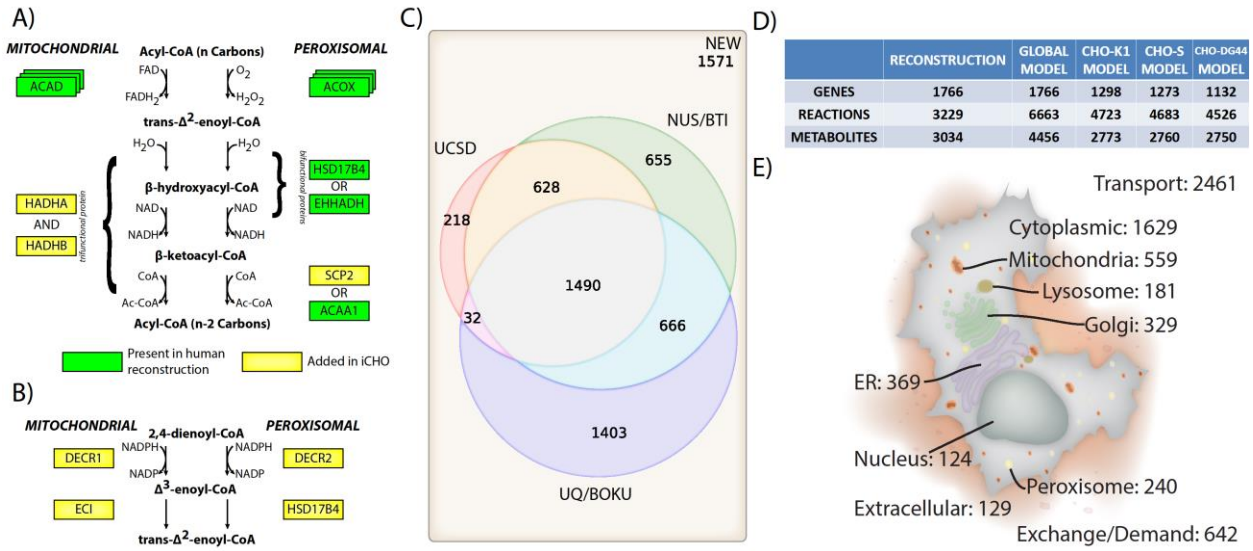


Figure 2.1*:* A multi-step process was used to reconcile a few existing unpublished models and to generate the final community reconstruction and models.
(A) The manually-curated human metabolic network reconstructions (Recon 1[52] and Recon 2[53]) were used to define an initial set of reactions catalyzed in *C. griseus* and the genes and proteins involved in each reaction (GPRs). Specifically, Recon 1 and Recon 2 were combined and all enzyme-catalyzed reactions that differed between the two were manually curated and reconciled to obtain consistent GPRs. *C. griseus* homologs were obtained for each human gene to obtain a set of draft GPRs linked specifically to genes in the CHO-K1 genome annotation. **(B)** The draft CHO GPRs were then compared with the GPRs from three independently-reconstructed and unpublished CHO genome-scale models, thus leveraging the manual curation invested in each input model. By manually verifying all GPRs and adding additional CHO-specific reactions present in the input CHO genome-scale models, we obtained a more comprehensive community reconstruction for *C. griseus*. To enable computation with this network, orphan reactions from Recon 2 were added, and omics data were used to build a global and cell-line specific models.

The reconciliation and reconstruction effort first identified and curated the biochemical relationships linking genes, proteins, and reactions (GPRs) for human metabolism from established genome-scale reconstructions for *Homo sapiens*[52–54]. CHO homologs to human genes were found using three different methods: reciprocal BLAST, the standalone InParanoid program v4.1[55], and gene name matching. Based on this effort, putative CHO GPRs were constructed. These GPRs were then compared against the GPRs from each of the earlier independently-developed genome-scale metabolic network reconstructions. Manual curation of the reactions ensured accuracy of the GPRs, subcellular compartmentalization, and reaction stoichiometry. Additional CHO-specific reactions were included based on literature support and updates were made to mitochondrial and peroxisomal beta oxidation to reflect biochemical GPRs (Figure 2.2A/B).

The resulting community genome-scale metabolic network reconstruction, based on literature support from over 1,300 publications, includes 1,766 genes and 3,229 reactions associated with those genes. As is common when building genome-scale metabolic models, an additional 3,434 reactions without gene associations were added to convert the reconstruction to a computable model. These include boundary reactions defining metabolite uptake and secretion rates, as well as reactions where the specific gene responsible is unknown (e.g., transporters). In total, the global *C. griseus* metabolic model, iCHO1766, contains 6,663 metabolic reactions involving 2,341 unique metabolites (4,456 metabolites when accounting for subcellular compartmentalization). Only 22% (1,490) of the reactions were present in all three input models. Furthermore, almost 25% (1,571) of the reactions were newly added, in that these reactions were not present in any of the initial genome-scale CHO models. A more detailed breakdown of reaction sources is shown in Figure 2.2C.



	RECONSTRUCTION	GLOBAL MODEL	CHO-K1 MODEL	CHO-S MODEL	CHO-DG44 MODEL
GENES	1766	1766	1298	1273	1132
REACTIONS	3229	6663	4723	4683	4526
METABOLITES	3034	4456	2773	2760	2750

Figure 2.2*:* Outcomes from the reconciliation process.
GPRs were updated and novel reactions were added in the reconstruction process, as shown for example in **(A)** mitochondrial and peroxisomal beta oxidation. Specifically, the trifunctional enzyme necessary for the 2nd, 3rd, and 4th steps of mitochondrial beta oxidation has been added to oxidation reactions occurring in the mitochondria while the GPRs of peroxisomal beta oxidation reactions have been updated to reflect additional catalytic activity of the SCP2 protein. **(B)** Additional enzymes necessary for catabolism of various unsaturated fatty acids have been added for both peroxisomal and mitochondrial beta oxidation. **(C)** The starting models had considerable differences in content, and following reconciliation, 1,571 new reactions were added to the model that had not been included in any of the starting CHO models. UCSD, NUS, and UQ/BOKU indicate the various groups contributing models to the initial reconciliation effort (see Methods for additional details for each model). **(D)** The reconstruction refers solely to gene-associated content. To convert the reconstruction into a computable model, a global *C. griseus* model was built by including orphan reactions from Recon 2. These additions enable the activity of known enzymes in *C. griseus* and serve as hypotheses for enzyme discovery in CHO. After removing unexpressed genes, cell-line specific models were constructed for CHO-K1, -S, and -DG44 cells. **(E)** The reactions in the global *C. griseus* model account for pathways in multiple subcellular compartments.

To formulate a GeM that can simulate growth and recombinant protein production, we used known compositional data to define the relative amounts of each metabolite needed for synthesizing all cell components (i.e., biomass[56]) and the recombinant proteins erythropoietin (EPO) and IgG. Due to differences between calculated gross cell composition in nonproducing CHO cell lines and measured values for IgG-producing hybridoma lines (i.e. measured protein fraction was more than 70% of cell dry weight in a producing cell, while calculated to be 55% in a non-producing cell), two separate biomass reactions were formulated (see Methods). Cell-line specific models for CHO-K1, -S, and -DG44 were constructed using the GIMME algorithm[57] and contained 4,718, 4,672, and 4,502 reactions, respectively (Figure 2.2D). The CHO-S and CHO-

K1 models were built using RNA-Seq and proteomic data for both cell lines. In the proteomic data, 1,326 and 3,200 proteins were detected in CHO-S and CHO-K1, respectively. The CHO-DG44 model was built based on microarray data for 13,504 genes (see Methods for data generation protocols, and Supplemental Data S4[58] for data used for model construction). Additional information on reconstruction and model content is available in Supplemental Data S1[58], including supporting literature associated with reactions, metabolites, and/or genes. The model accounts for 9 compartments (cytosol, mitochondria, nucleus, endoplasmic reticulum, Golgi complex, lysosome, peroxisome, mitochondrial intermembrane space, and extracellular space) with the subcellular localization of reactions summarized in Figure 2.2E. The global model and reconstruction are available to browse and download at the BiGG Models Database[59] (http://bigg.ucsd.edu), while the three cell line models and global model are additionally hosted at http://www.CHOgenome.org[60].

CHO cell-line specific models recapitulate experimental growth rates

Genome-scale models can be used to predict phenotypes, including growth rates, when uptake and secretion rates are provided for major metabolites. Thus, to evaluate the models, we simulated growth rates for several different cell lines producing recombinant proteins. Data were acquired from literature[27,46,50,61,62] and from new experiments presented here (see Methods). Measured uptake and secretion rates for major nutrients and recombinant proteins were applied as constraints to the appropriate cell-line specific model, and flux balance analysis was used to predict growth rates by optimizing for flux through the appropriate biomass reaction (see Supplemental Data S3[58] for constraints). These computationally predicted growth rates were—on average—within 25% of the measured growth rates for cells grown in serum free conditions (Figure 2.3). Additional predictions were made for datasets exhibiting high uncertainty in calculated uptake and secretion rates, as well as cells grown in serum containing media (Figures 2.S3-S5).

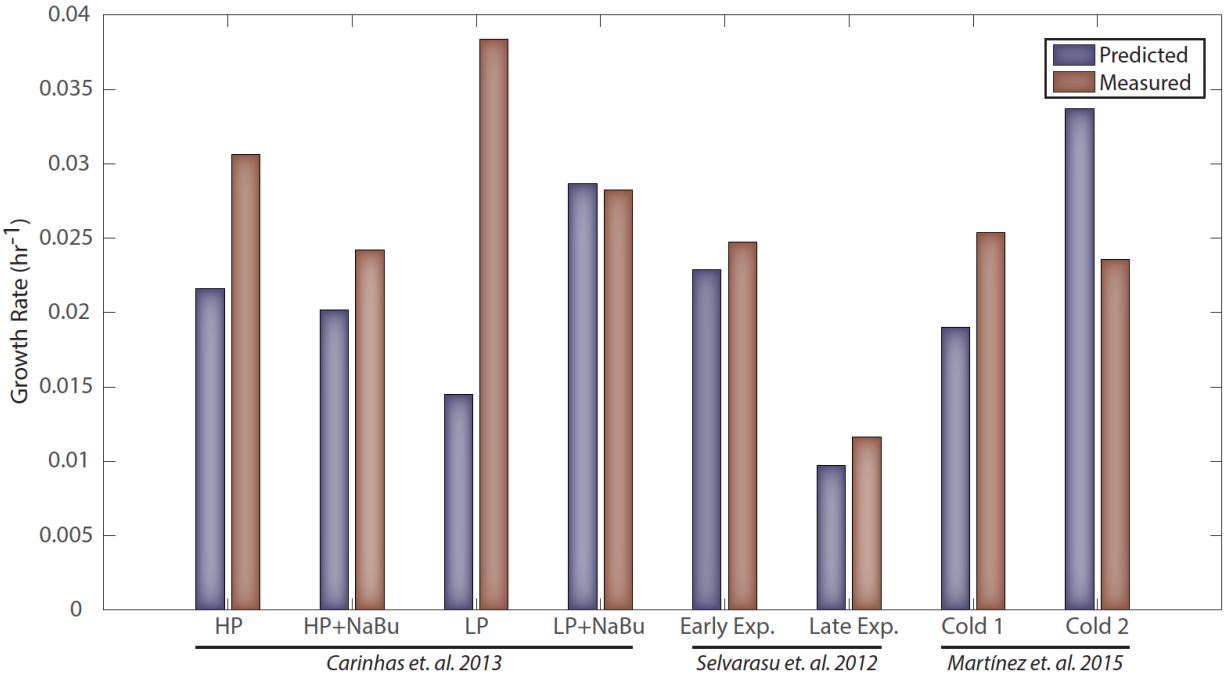


Figure 2.3: The CHO cell-line models can compute growth rates for various IgG-producing cell lines. All cell lines are grown in serum-free media and are producing IgG. HP and LP refer to high and low producers respectively while NaBu indicates the presence of sodium butyrate addition to the media[46]. Early exp./Late exp. refer to the early and late exponential phase, respectively[27]. Two cell lines are from cultures exposed to a temperature shift[50]; however the data points used come from the time period prior to the temperature shift. Simulations for the Selvarasu study utilize the CHO-DG44 cell line model. Other simulations use the CHO-K1 cell line model.

Cell-line specific models recapitulate known amino acid auxotrophies

CHO cells exhibit several amino acid auxotrophies[12–14] beyond the nine essential amino acids in human (His, Ile, Leu, Lys, Met, Phe, Thr, Trp, and Val) and arginine, which is additionally essential in rats for normal growth[63] . The CHO-specific auxotrophies include cysteine, proline, and at least one report of an asparagine auxotrophy[15]. While cysteine, proline, asparagine, and arginine are essential based on experimental evidence, homologs for all genes required for their biosynthesis are in the *C. griseus* genome. Thus, we tested the cell line specific models for agreement with the reported amino acid auxotrophies and investigated the cell line specific transcriptomic and proteomic data (see Supplemental Data S4[58]) to understand the underlying mechanisms.

All the cell line models reproduced the arginine and cysteine auxotrophies. Arginine biosynthesis was inhibited due to low or absent gene expression of one or both arginine biosynthetic genes (Figure 2.4A), and cysteine synthesis was disabled since cystathionine lyase and synthase were not expressed (Figure 2.4B). The asparagine synthase reaction, on the other hand, has strong transcriptomic and proteomic evidence for its presence in CHO cell lines (Figure 2.4D). Therefore, we experimentally tested if our CHO-S and CHO-K1 lines were auxotrophic for asparagine. Consistent with our models and the expression data, the cells could grow without asparagine (Table 2.1). The reported asparagine auxotrophy was likely unique to the CHOK1SV cell line used in the earlier study[15], and not a general characteristic of CHO cells.

Table 2.1: The effect of asparagine removal on CHO-K1 and CHO-S growth.
Data from 72 hours after seeding. Both cell lines are able to grow without asparagine in the culture media, albeit at slightly slower rates.

	Viability (%)	VCD (10^6 cells/mL)
CHO-S: w/Asn (Control 1)	94.4	2.13
CHO-S: w/Asn (Control 2)	93.6	2.18
CHO-S: w/out Asn	93	1.66
CHO-K1: w/Asn (Control 1)	96.5	1.45
CHO-K1: w/Asn (Control 2)	96.2	1.5
CHO-K1: w/out Asn	96.2	1.15

We next analyzed the source of the proline auxotrophy in CHO cells (Figure 2.4C). Mammals can synthesize proline from arginine or glutamate; however, previous reports show that CHO-K1 cells do not incorporate glutamate into proline, and exhibit a drastically decreased (yet present) rate of ornithine transaminase activity (<10% of the activity in *C. griseus* lung cells)[14]. Both biosynthetic pathways normally converge upon glutamate-5-semialdehyde, which is then converted to proline after spontaneous decomposition to 1-pyrroline-5-carboxylate. In the CHO-S and CHO-DG44 models, the pathway from glutamate is missing due to a lack of the necessary protein and/or mRNA expression. The GIMME algorithm incorrectly included the pathway from

arginine to proline (via ornithine) in the CHO-S and CHO-DG44 models (the CHO-K1 model was missing the arginase reaction); however, careful inspection of the transcriptomic and proteomic data demonstrate that this pathway is missing expression of at least one enzyme in all of the cell lines. The gene responsible for the final step in proline biosynthesis, reduction of pyrroline-5-carboxylate, is expressed in all three cell lines, consistent with previous reports showing activity of the enzyme in CHO-K1[12].

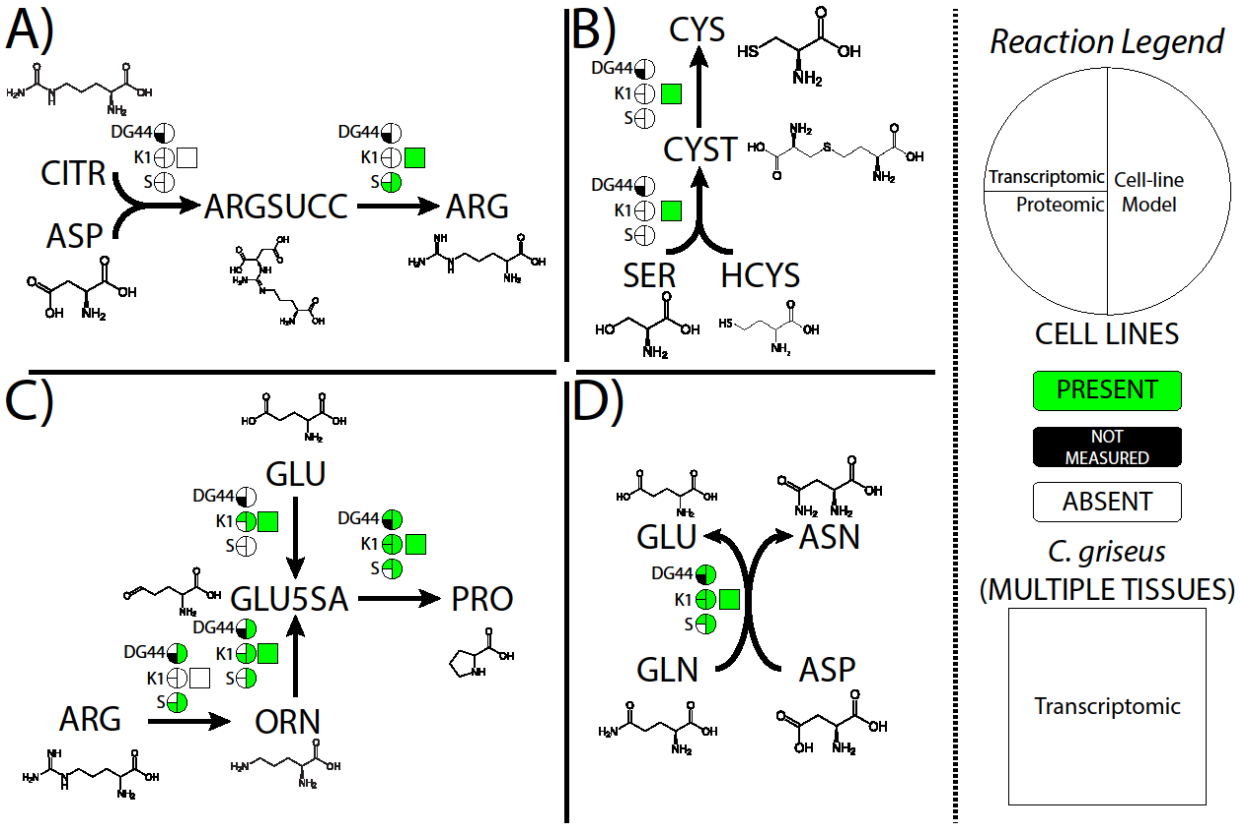


Figure 2.4: The models provide insights into the molecular basis of CHO-specific amino acid auxotrophies.
(A) Arginine, **(B)** Cysteine, **(C)** Proline, **(D)** Asparagine. For each reaction, information in the circles indicates whether the enzyme catalyzing the reaction is seen in transcriptomic (upper left quarter-circle) and/or proteomic (lower left quarter-circle) data, as well as its presence or absence in the cell line specific models generated by GIMME (right semi-circle). The square shows whether the enzyme catalyzing the reaction is seen in transcriptomic data from a mix of *C. griseus* tissues. Data used is available in Supplemental Data S4[58]. Metabolite abbreviations are as follows: CITR-citrulline, ASP-aspartate, ARGSUCC-argininosuccinate, ARG-arginine, SER-serine, HCYS-homocysteine, CYST-cystathionine, ORN-ornithine, GLU-glutamate, GLU5SA-glutamate 5 semialdehyde, PRO-proline, GLN-glutamine, ASN-asparagine.

Thus, the model-predicted proline synthesis capabilities show some inconsistencies with experimental observations due to limitations of the GIMME algorithm (i.e., it erroneously added

the pathway linking arginine to proline); however, by overlaying transcriptomic and proteomic data, the reconstruction provided an explanation for the proline auxotrophy characteristic of CHO cells (Figure 2.4C), guiding further model curation and enabling more accurate model predictions.

An examination of *C. griseus* transcriptomic data from multiple tissues[25] (see Supplemental Data S4[58]) further elucidates which auxotrophies are characteristic of the organism and which are cell line-specific. The enzymes needed for cysteine and asparagine synthesis were expressed in the hamster tissues, consistent with their nonessentiality in rat and human (Figure 2.4B/D). In the hamster tissues, the arginine biosynthetic pathway has high expression in one of two necessary reactions. The second reaction, argininosuccinate synthase, exhibits low expression levels (Figure 2.4A). This is perhaps unsurprising since rats can synthesize arginine, but not in levels required to support normal growth[63]. Finally, synthesis of proline via ornithine shows expression in the hamster tissues, consistent with reports of higher ornithine transaminase expression in *C. griseus* lung cells[14]; the pathway via glutamate also shows expression in the hamster tissues (Figure 2.4C). These data suggest that the cysteine and proline auxotrophies are specific to the CHO cell lines and the arginine auxotrophy is—at least partially—shared between the hamster and the CHO cell lines. Thus, by analyzing transcriptomic and proteomic data in the context of the cellular pathways, we are able to elucidate the functional basis for known amino acid auxotrophies in CHO cells.

Cell engineering enhances the efficiency of resource utilization compared to common bioprocess treatments

Substantial increases have been achieved in recombinant protein yields in CHO-based bioprocessing in part by balancing nutrient concentrations in media. Since nutrients can be used for growth and other purposes, chemical treatments and temperature shifts have been employed to increase product synthesis at the expense of growth. However, it is unclear how well these treatments redirect resources from growth to protein product. To test this, we used the cell-line

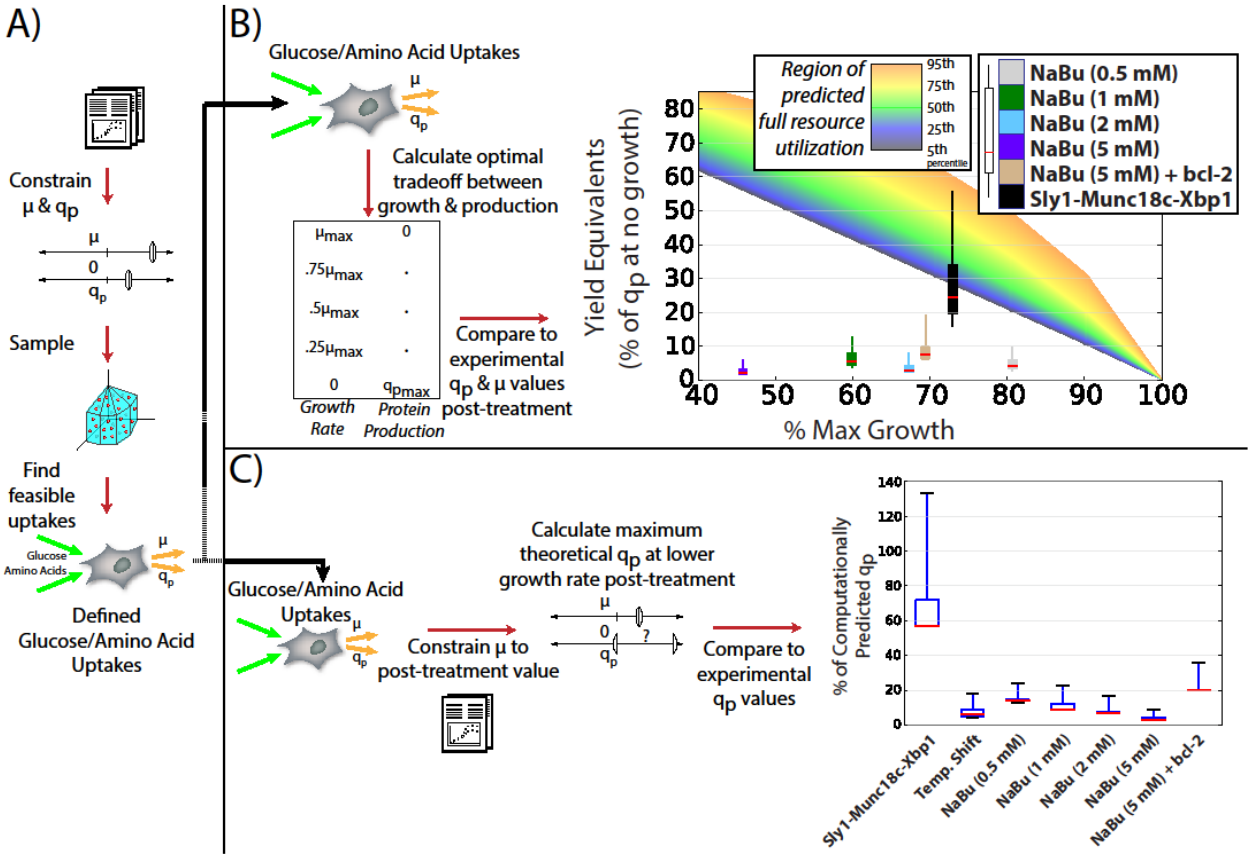


Figure 2.5: Resource utilization efficiency in CHO cell lines is greater after cell line engineering.
(A) Feasible uptake fluxes were generated for nutrient utilization efficiency analyses. Growth rates and specific productivities were obtained and used to constrain the model. By sampling the constrained model, a set of feasible metabolite uptake rates and secretion rates was calculated that support growth and production at the specified values. **(B)** The efficacy of resource utilization following common growth-inhibiting treatments in protein-producing CHO cell lines was quantified. Uptake and secretion rates from **(A)** were used to predict maximum growth (i.e., no protein production) and maximum protein yield (i.e., no growth), as well as yields at various fractions of maximum growth. These were used to predict a range of optimal protein production rates (i.e. making full use of resources) as growth rate decreases (indicated as a region of full resource utilization), thus showing the theoretical maximal protein secretion rates at decreased growth rates. The 5th, 25th, 50th, 75th, and 95th percentile of theoretical maximal protein production rates at each growth rate are indicated by the gradient in color from black to orange. After several cell treatments, the experimentally measured increased protein yield and decreased cell growth rate were compared to the predicted optimal protein secretion rates to assess how successfully each treatment utilized available resources (e.g., amino acids and sugars) for growth and protein production. Boxes span the 25th and 75th percentiles, whiskers represent the 5th and 95th percentiles, and a red line denotes the median value of overall resource utilization efficiency for each treatment, calculated as the ratio of experimentally measured protein production to the theoretical maximum protein yield (i.e., no growth). **(C)** The efficiency of diversion of resources toward protein production following common treatments was assessed. Uptake and secretion rates from **(A)** were used to compute the theoretical maximum specific productivity after cell line or process modifications yielding a range of theoretical optimal q_p values were computed. Experimentally measured production rates were compared to the computational predictions to assess how effectively the cells are able to make use of resources gained from growing slower. Boxes span the 25th and 75th percentiles, whiskers represent the 5th and 95th percentiles, and a red line denoting the median value.

models to simulate cell metabolism and recombinant protein production before and after culture perturbations. This approach allows us to analyze how efficiently cellular resources were redirected to protein synthesis following bioprocess treatments or targeted cell engineering efforts.

Specifically, we studied temperature shifts[64], sodium butyrate (NaBu) addition with[65] and without[66,67] *BCL2* overexpression, and overexpression of secretory pathway genes[30]. In each case, estimates for uptake and secretion rates of major nutrients were predicted for the control culture from each study (Figure 2.5A) by sampling feasible uptake and secretion rates that supported the measured growth and protein secretion rates (see Methods and Supplemental Data S5[58]). These limits on nutrient availability were then applied as constraints to the models exhibiting the lower growth rates and higher protein secretion rates observed in the studies after bioprocess treatments or cell engineering. Using this approach, we calculated the maximum protein production rate as a function of growth rate for each treatment by optimizing for protein production after constraining growth to a fixed value. When compared to experimentally measured changes in growth and protein synthesis following each treatment, we quantified how well the cells made use of resources for both growth and production (Figure 2.5B). Furthermore, we assessed how efficiently the cells use newly available resources at lower growth rates for protein production (Figure 2.5C), as detailed in the following sections.

Bioprocess treatments, such as histone deacetylase inhibitors and temperature shifts can improve specific productivities and product titers. However, our analysis shows that overall resource utilization was inefficient post-treatment, independent of the decrease in growth rate (Figure 2.5B). To quantify the inefficiency, we compared the amount of protein that was produced experimentally to the amount of protein that could be produced if growth was halted (as calculated via model simulation). We simulated maximum protein production across the range of feasible growth rates, and defined a region of optimal resource utilization for protein production—an area where neither additional growth nor protein production is possible with the resources available

(rainbow region in Figure 2.5B). For several treatments, we examined proximity to optimal resource utilization by comparing the measured growth and protein production rates following bioprocess or cell engineering treatments to the simulated maximum growth rate and protein production rate. Cells treated with NaBu produced protein at 2-6% of the predicted maximum rate, irrespective of the decrease in growth rate, and were far from the region of full resource utilization. *BCL2* overexpression (relieving the apoptotic effect of NaBu) still only achieved 8% of the maximum protein production rate. Meanwhile, overexpressing the secretory pathway proteins *XBP1*, *STXBP3* (*MUNC18C*), and *SCFD1* (*SLY1*) allowed protein production at 24% of the maximum possible rate—3 fold higher than the best performance seen with cells overexpressing *BCL2* and treated with NaBu. Thus, yields are low, even following bioprocess treatments. However, the increases seen with targeted engineering, as cells move closer to the region of optimal resource utilization, suggest that there is considerable potential for increasing cell-specific productivity to near optimal levels.

The inherent trade-off between growth and protein production suggests that resources for protein production may be liberated following growth-inhibiting bioprocess treatments. Thus, we simulated how efficiently cells, after suffering a decrease in growth rate from the different treatments, redirected newly available resources to protein production (Figure 2.5C). We found that while culture at lower temperatures (33°C vs. 37°C) often increases final titers, the specific productivity (q_p) only increased to 6% of the theoretically possible level at the measured growth rate. Similarly, NaBu treated cells produced only 3-14% of their computationally predicted q_p. Furthermore, simulating increases in NaBu concentration resulted in even less efficient redirection of resources to protein production. While these bioprocess treatments failed to efficiently redirect resources specifically to protein production, targeted cell engineering efforts improved the shift of resources. For example, overexpressing *BCL2* during NaBu treatment resulted in a modest improvement in resource utilization (roughly 20% of the predicted q_p). The greatest improvement

in resource utilization was observed when the secretory pathway proteins were overexpressed, producing at close to 57% of computationally predicted maximum specific productivity.

DISCUSSION

Here we present a genome-scale metabolic network reconstruction for the Chinese hamster, *C. griseus*. This resource enables the enumeration of true genotype-phenotype links by connecting the functions of more than 1,700 genes to CHO cell phenotypes such as growth, metabolic pathway activity, and protein production. It presents many opportunities for analyzing bioprocess performance and guiding cell engineering. First, the reconstruction provides a platform for managing and interpreting CHO-relevant big data. Second, the models demonstrated potential for improving recombinant protein production by engineering cells to efficiently shift resources to protein production. The reconstruction will guide further strategies for cell line development as advances are made in modeling and relevant techniques. Third, the reconstruction will serve as a resource, in which up-to-date knowledge on the biochemistry of CHO cells can be maintained and made available to the entire CHO community.

To gain greater control of product yields and quality attributes of biopharmaceuticals, efforts to engineer CHO cells must consider the activities of cell pathways and associated genes. Thus, CHO-based bioprocessing is adopting data-rich technologies to quantify the cellular parts, using next-generation sequencing, high-throughput omics techniques, high-content imaging, and online bioprocess measurement techniques. To effectively analyze the data and deploy interventions, powerful statistical and modeling methods are needed[68]. This reconstruction serves as a platform for analyzing many types of molecular and phenotypic data using a variety of algorithms[31]. Furthermore, since this reconstruction provides a mechanistic link between the genotype and phenotype of CHO cells (via enumeration of enzymes underlying metabolic pathways), it allows for the effective integration of orthogonal data types such as metabolomics, transcriptomics, genetic variants, and growth rates[69,70]. We demonstrated this with our cell line

specific models for the CHO-K1, -S, and -DG44 lines. These models were built by integrating transcriptomic, proteomic, and metabolomic data with the global genome-scale metabolic model, validated by predicting growth rates and auxotrophies, and ultimately used to predict protein production capabilities. As the scope and quality of data expands, the CHO genome-scale network reconstruction will continue to enable the diagnosis of the molecular basis of different phenotypes, by serving as a platform to analyze the interplay of diverse data types.

The models provided quantitative evidence that targeted engineering of the secretory pathway allows for more efficient use of liberated resources as growth decreases. The models will enable further analyses of cell lines to help design mutants that provide desired protein quality attributes. The accuracy of such cell designs will further improve as five challenges are addressed. First, different CHO cell lines have accumulated hundreds of mutations in metabolic pathways[25], which may contribute to differences seen in phenotypic traits across cell lines[71]. However, more research is required to identify which mutations impact the activity of pathways. Second, growth and protein yield predictions require accurate measurements of metabolite concentrations, and advances in analytical chemistry will provide improved constraints on simulations with CHO metabolic models (Figure 2.S1). Third, gaps remain in our understanding of mammalian metabolism, but algorithmic developments[72] and biochemical assays will continue to refine the CHO metabolic network. Fourth, our biomass objective functions were defined based on experimental measurements in CHO and hybridoma cells and provided accurate predictions of cell growth; however, biomass composition is not static, which may impact quantitative predictions of growth and protein production. Thus, there is a need for comprehensive measurements of the composition of CHO cells (e.g. protein, DNA, RNA, lipids, etc.) under different conditions to formulate more accurate biomass objective functions. Finally, metabolism is only part of the system that controls the quantity and quality of recombinant proteins. Efforts to model other processes such as glycosylation[73,74] and the secretory pathway[33] can be integrated with metabolic networks for a more comprehensive view of protein production; further addition of

signaling and regulatory networks could reveal mechanisms behind inefficiencies in protein production. These five advances will further improve the predictive power of the genome-scale CHO models.

Lastly, the development of this resource involved the concerted effort of many groups in the community with interest in genome-scale metabolic modeling for mammalian bioprocesses and CHO cells. Together, these combined efforts enabled the careful curation of this community resource. The models are available at http://www.CHOgenome.org[60]. Additionally, the global model can be browsed and downloaded at the BiGG Models database[59] (accessible at http://bigg.ucsd.edu), where literature information and experimental evidence is provided for each gene, reaction, and metabolite in the model. Updates will be made with continued research into CHO cells, including improvements in the *C. griseus* genome sequence and annotation, and experimental tests of model predictions[75]. As such studies accumulate, the network reconstruction, as a community resource, will be maintained and improved over time, thus enabling researchers throughout the CHO cell community to conduct many studies to enhance the value of the primary host for biotherapeutic production.

METHODS

Cultures for Transcriptomic, Proteomic, and/or Metabolomic Data

CHO-S (Life Technologies # A11557-01) cells were cultured in CD CHO medium (Life Technologies # 10743-029) supplemented with 8 mM L-glutamine (Life Technologies # 25030-024), Anti-clumping agent 1:500 (Life Technologies # 0010057 AE), and Pen-Strep 1:100 (Life Technologies # 15140-122) in 2 L Corning shake flasks (Sigma # 431255) with 400 mL medium. All cultures were maintained in an incubator kept at 37°C, 5% CO_2, 70% humidity and 25 mm throw, shaking at 120 rpm.

Adherent CHO-K1 (ATCC CCL-61) cells were grown in F-12K medium, supplemented with 10% FBS, 1% non-essential amino acids, and 2 mM L-glutamine (Gibco) incubated at 37°C

and 5% CO_2. Spent medium was sampled from the culture at regular intervals and metabolite concentrations were measured. Cell-free medium was also sampled at regular intervals to control for metabolite degradation. Samples were taken from the CHO-K1 and cell-free control every 2 hours between 24-36 hours, and every 4 hours subsequently (see Supplemental Data S4[58]).

Cultures for Determining Asparagine Essentiality

CHO-S (Life Technologies) and CHO-K1 (ATCC CCL-61 adapted to grow in suspension, serum-free conditions) cells were grown in CD-CHO media with and without asparagine at a seeding concentration of 0.4-0.5 x 10^6 cells/mL. Asparagine was either added via MEM NEAA solution (Control 1) or as an individual amino acid supplementing MEM NEAA solution lacking asparagine (Control 2). Both cell lines were able to grow successfully in media lacking asparagine, as measured by viability and VCD 72 hours after seeding (Table 2.1). Cell growth and viability were monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec, Allerod, Denmark) based on two fluorescent dyes, acridine orange and DAPI for the total and dead cell populations, respectively.

Input genome-scale model construction

In the preliminary stages of this effort, three research groups contributing to this study independently developed unpublished genome-scale models of CHO cell metabolism. These were supplemented by an additional independent reconstruction of amino acid metabolism.

University of California San Diego metabolic model

The initial draft of the UCSD genome-scale CHO metabolic network was reconstructed using GIMMEp[76]. The algorithm requires an initial metabolic network, a core set of reactions that must be operational, a set of required metabolic functions, and an optional set of additional reactions that may or may not be included in the final model.

First, we used the global human metabolic network (Recon 1[52]) as the initial network for the algorithm to build the CHO metabolic network. Recon 1 served as a template for which metabolic reactions were initially chosen for the CHO metabolic network.

Second, a core set of reactions was defined based on homology between the CHO and human genomes. A bidirectional BLASTP was done by comparing human RefSeq proteins (downloaded on 12/13/2011) against the CHO genome sequence[8]. Metabolic proteins from Recon 1 with a high identity (>70%) and high overlap (>70%) with a CHO protein in the bidirectional BLAST were marked as the core set of homologous proteins. A small set of proteins met the inclusion criteria in only one direction and was subsequently manually curated for inclusion. The gene-protein-reaction rules for Recon 1 reactions were updated with CHO annotations for homologous proteins, and removed if no homologous protein was present. The set of optional reactions provided as an input to GIMMEp was not included in the final model.

Third, a set of metabolic functions was defined including: 1) the ability for the metabolic model to produce biomass, 2) the ability to produce a glycosylated protein (EPO), and 3) the ability to uptake and secrete known metabolites characteristic of CHO-K1 cells. The necessary reactions to accomplish these three tasks were added to the initial model before the model building algorithm was run and were included in the core set of reactions. A biomass objective function[56] was constructed accounting for RNA, DNA, protein content, free amino acids, glycogen, free fatty acids, lipids, phospholipids, triglycerides, and associated ATP maintenance costs using primary literature. The ratio of nucleotide content in DNA was set based on the GC content of the genome[8]. Ratios of nucleotide content in RNA and amino acid content in proteins were estimated based on RNA-Seq data[8]. In addition, metabolic reactions were reconstructed that translated the EPO protein and added the necessary N- and O-glycans, as well as degraded the protein. Also, the exchange reactions of metabolites known to be taken up or secreted by CHO cells were included in the core set of reactions.

The resultant draft model contained all CHO metabolic proteins that had high homology with the human metabolic proteins, while minimizing the number of metabolic reactions with low homology that were added to enable the model to accomplish a set of key metabolic functions including growth, protein production, and uptake/secretion of documented metabolites.

This was followed by manual curation of the model to ensure that gene associations, reaction localization, and other reaction contents were accurate.

National University of Singapore/Bioprocessing Technology Institute metabolic model

The *H. sapiens* reconstruction, Recon 1[52], served as the starting point of the NUS CHO-K1 genome scale metabolic model reconstruction. This was motivated by the high-quality manual curation performed on Recon 1, and the large number of genes (81%) that were shared between the human and CHO-K1 genomes[8]. To create a draft list of reactions for the new model, a protein BLAST search of human genes was conducted against CHO-K1 genes. This allowed the mapping of metabolic genes from human to CHO-K1. Of the 1496 metabolic genes in the Recon 1 model, 1441 (96%) corresponding matches were found in the CHO-K1 genome. Based on the list of 55 missing genes, 39 relevant reactions were removed from the Recon 1 model. The number of reactions removed is less than the number of genes due to reactions associated with isozymes. The 1514 non-gene associated reactions in the original model were retained, since they were necessary to produce a functioning model.

The removal of missing genes introduced gaps within the model that prevented biomass production and removed metabolic functions. By examining the list of removed reactions, SQLEr (squalene epoxidase) and C3STKR2r (C-3 sterol keto reductase zymosterol) were found to be necessary to ensure biomass formation. We also found that the removal of the three reactions catalyzed by 2-oxoisovalerate dehydrogenase prevented the catabolism of branched amino acids (valine, leucine, and isoleucine). While the removal of TPI1 (triosephosphate isomerase 1) strictly did not impair biomass formation, there are strong physiological arguments for its existence, and

there is evidence that the enzyme exists in the ancestral CHO cell line[77]. Based on experimental evidence[27], the GGLUCT (gamma-glutamylcyclotransferase) reaction was not also removed.

As the Recon 1 model was constructed to represent the global human metabolic model, the model includes tissue specific reactions that are not relevant to CHO-K1. Consequently, we removed these reactions from the reaction list. The reactions removed include 26 demand reactions for micronutrients such as vitamins, and 36 duplicate reactions for liver and uterus tissue (e.g., in dolichol biosynthesis). The R Group synthesis reactions were also removed as the lipid requirement for CHO-K1 biomass formation was subsequently reconstructed based on literature sources. Finally, 80 reactions and relevant metabolites with the “hs” suffix were relabeled to the “cho” suffix.

To further refine the model, we utilized the KEGG PATHWAY[78] and BRENDA Enzyme Database[79] databases to identify CHO metabolic reactions based on physiological evidence. Based on curated databases of BRENDA and KEGG, 85 additional reactions were added to subsystems such as arginine and proline metabolism, IMP biosynthesis, and inositol phosphate metabolism. These include two reactions previously removed based on the gene mapping using BLAST.

The model reconstruction was originally conceived for biotechnological applications. Therefore, relevant biomass formation reactions were constructed and added. As the R Group synthesis reactions were removed, 15 reactions were added for the production of fatty acid and cholesterol components of the cellular biomass. Reversible transport via proton symport reactions were also added to facilitate the independent exchange of amino acids. Finally, based on recent evidence obtained by LC-MS experiments[27], 13 reactions were added to account for the production and secretion of the detected amino acid derivatives. The reconstructed CHO-K1 model consists of 3718 reactions and 2774 compartment specific metabolites (1523 non compartment specific).

The University of Queensland/ University of Natural Resources and Life Sciences, Vienna metabolic model

An updated version[53] of the *H. sapiens* reconstruction Recon 2, was used as a starting point for generation of a manually curated CHO specific model by the UQ/BOKU groups. Curation focused on the identification of inconsistencies in the naming conventions, annotations, removal of duplicated reactions and metabolites, as well as correction of the mass and energy balance.

Homologous genes between CHO and *H. sapiens* were identified from RefSeq Release 66 using the standalone InParanoid program v4.1[55]. The identified homologies were then used to convert the human specific Recon 2 metabolic network to a CHO specific network. This provided an initial basis for a CHO-specific metabolic network based on Recon 2, which later was considered in the curation of the community-level network.

Technical University of Denmark amino acid subnetwork

A metabolic network reconstruction of amino acid metabolism in CHO cells was generated at DTU using mouse genomic and biochemical pathway information from the KEGG database as a starting point. To identify orthologous metabolic genes in CHO, a protein BLAST search of amino acid metabolic genes from mouse was manually conducted against the CHO-K1 genome[8], hosted at http://www.CHOgenome.org. The resulting list of CHO genes was manually curated for inclusion based on information from literature.

Reconciliation process

To merge these into one final community reconstruction, we developed a workflow to reconcile similarities and differences among the existing models (Figure 2.1).

The final metabolic network reconstruction was initially based on knowledge from human metabolism, as detailed in Recon 1[52] and an updated version of Recon 2[53] (Figure 2.1A). Additional curation of Recon 2 was conducted to refine the reconstruction[54]. These two

reconstructions were first compared to determine a baseline set of gene-protein-reaction (GPR) relationships for all reactions in the models. If GPRs showed discrepancies between the two reconstructions, manual curation was carried out to determine a consensus GPR. This combined human reconstruction was subsequently used to identify *C. griseus* homologs. Specifically, three different homology methods were used to determine CHO-human protein homologs. First, two-way BLAST was conducted between the CHO proteome and the human proteome. Reciprocal top matches for both *C. griseus* and human proteins were identified and retained. Second, the standalone InParanoid program v4.1[55] was used to find orthologs between human and *C. griseus*. Third, a search was conducted to identify genes with identical gene names.

For each human gene, the union of these three methods was used to map from human GPRs to putative CHO GPRs, resulting in 3701 reactions with gene associations. In this set of reactions, 733 had GPRs that were identical between all 3 preliminary reconstructions and the putative CHO GPR. For the remaining reactions (2968), careful manual curation was carried out to determine the most accurate GPR for each reaction in the final model (Figure 2.1B). In each case, primary literature was searched to find CHO-specific information about the reaction (e.g. substrate specificity, subcellular localization, protein complex composition, gene association, etc.). When such information was unavailable, information from other mammals was used. Through the curation process, additional CHO-specific reactions were identified and added.

The curated gene-associated reactions were then combined with all the non-gene associated reactions from Recon 1 and the updated Recon 2 to give a base CHO metabolic network. Mass imbalances were corrected and the network was tested for biomass functionality. Further refinements included the removal of opposing irreversible reactions (in favor of a single reversible form), the replacement of lumped reactions, and the addition of pathways for synthesizing IgG and EPO.

As a final comparison, the model content was compared against another unpublished and independently reconstructed CHO metabolic reconstruction, based on Recon 1[52] (see below for details on this model's construction) and any differences manually curated.

University of Chile metabolic model

For the generation of a preliminary CHO genome-scale metabolic model we used Pantograph, a tool to reconstruct genome-scale models for eukaryotic organisms[80]. This tool uses a template metabolic model and annotated genomes for both template and objective organisms, automatically rewriting the template's GPRs in terms of the genes of the target organism, inheriting knowledge from the template model and producing a draft metabolic model well suited for manual curation.

Three models were considered as templates for network reconstruction. A hybridoma metabolic reconstruction based on *Mus musculus* genomic and biochemical information[81] was modified and upgraded to include new genomic information available in databases such as KEGG and NCBI-gene by manual curation and by using a script that connects to the KEGG database to download new GPR candidates to be added to the model.

A metabolic reconstruction for *Mus musculus* based on Recon 1[52], iMM1415, was also used as template for the generation of a model for CHO cell metabolism. This model includes 1,415 genes, 2,212 gene-associated reactions and 1,514 non-gene-associated reactions. Finally, the human reconstruction, Recon 1, was also used as a template for the generation of a CHO genome-scale model[52].

Orthologous genes often exhibit similar biological activities. Thus, they can often reliably be used to build novel reconstructions from a template organism. The search for orthologs was performed using the standalone InParanoid program v4.1[55]**,** which finds clusters of orthologous genes based on similarity scores calculated by NCBI-Blast between proteomes of the analyzed species. The protein sequences for CHO were retrieved from http://www.CHOgenome.org and

Ensembl, and these sequences were used to find orthologs between CHO and *Mus musculus* and CHO and *Homo sapiens*.

The template model and ortholog information for *Mus musculus*, *Homo sapiens*, and CHO were used to generate a preliminary genome-scale model for CHO cells using Pantograph. Critical components for biomass synthesis were identified by analyzing metabolic pathways that lead to their synthesis using the COBRA toolbox[82].

GapFind[83] was used to find dead-end metabolites, which were subsequently studied using information from databases such as CHOgenome.org**,** KEGG, and Virtual Metabolic Human in order to fill the gaps present in the initial reconstruction. Model validation was performed using Pantograph[80] which tests the ability of the obtained genome-scale models to replicate experimental data, such as the effect of known gene deletions and use of alternate carbon sources for CHO cells in culture.

While the University of Chile model was not used as a base model in the reconciliation, it was compared to the global model and all discrepancies manually curated. All validated reactions suggested by this secondary curation process were added to the final global model.

Biomass reaction

Two biomass reactions were built here: one for a recombinant-protein producing cell line (biomass_cho_producing) and one for a nonproducing cell line (biomass_cho). The overall cell composition (e.g., protein, DNA, RNA, lipid fraction) of a protein producing cell line was averaged from previously reported values for mouse hybridoma cell lines. The nonproducing cell line overall cell composition was calculated based on literature values for different cellular components in CHO. Both biomass reactions used experimentally determined amino acid compositions of the protein fraction, obtained by averaging the composition of 5 cell lines[27]. The ratios of different nucleotides for RNA and DNA composition were determined from genome and transcript sequences. Phospholipid composition was taken from previously reported values for CHO-K1. A

detailed formulation of the biomass equations (including all associated references) is available in Supplemental Data S2[58].

Choice of objective function

An objective function should encapsulate what the cell is "trying" to do. What that entails remains a matter of great scientific interest. In bacterial systems, the use of a biomass objective function[56] has come to dominate, as experimental work shows cells evolving toward optimal states as predicted by maximization of biomass production[84–86].

Mammalian cells, as a whole, are more difficult to ascribe a one-size-fits-all objective function for, since many (e.g., terminally differentiated cells) do not rapidly proliferate; their 'objective' may be another biological activity such as antibody production (plasma cells), maintaining structural integrity (red blood cells), or generation of energy and signaling molecules (neurons). As CHO cells are proliferative, we selected a biomass objective function for simulation of cell growth. In a study comparing experimental fluxomic data to model flux predictions[87] in core metabolism in *Escherichia coli*, alternative objective functions were evaluated for their accuracy in recapitulating experimental measurements. This study found that maximization of ATP or biomass both led to the highest consistency between FBA predictions of metabolic flux and experimental flux measurements. Subsequent research provided further support for biomass optimization[86], but found that maximization of ATP yield showed less consistency in genome-scale models. The authors noted that "in a genome-scale model, the maximization of ATP selects against the usage of biosynthetic pathways, since the end products are not specified in the objective function". A similar conclusion can be drawn for minimizing redox load as an objective function, which also does not require the use of many biosynthetic pathways, and so such an objective function would not capture the activities of all pathways.

It is possible that in mammalian cell culture, that some combination of objective functions is the most biologically accurate; however, as our line of investigation largely focuses on

predictions of growth rate, a gross phenotypic characteristic which definitionally has an upper limit set by cellular composition (e.g., biomass generation), a simplification to a biomass objective function for growth simulations is a reasonable approximation. Furthermore, the ability to recapitulate experimentally-measured growth rates provides further support for this assumption and that the measurements used the biomass objective functions are within a reasonable range.

Construction of cell-line specific models

We used GIMME[57] to generate cell-line specific models for CHO-S, CHO-K1, and CHO-DG44. Transcriptomic and proteomic data were generated for CHO-S, while existing data were used for CHO-K1 and CHO-DG44 (see Method Details for information on data used or generated). Genes were called as present if their FPKM was greater than 1 or they were found in the proteomic analysis. For microarray data, genes were called as present if their normalized $\log_2$-transformed value was greater than 10. Blocked reactions were removed from the global model and the algorithm was used while optimizing for growth, IgG production, and erythropoietin production. The union of reactions present in these models served as the base cell-line model in each case.

Since algorithmic generation of the cell-line specific models ensures only functionality for user-defined objectives (here we used growth, IgG production, and erythropoietin production), additional manual curation was done to incorporate other biological functionalities important for simulation. We added eight reactions to all models: GLYCt and SUCCt4_3 were added to permit secretion of glycerol and succinate, respectively. HISD, URCN, IZPN, GluForTx, FTCD and DM_trp_L[c] were added to account for histidine and tryptophan uptake rates being significantly higher than needed for growth and protein production. Histidine catabolic reactions were added to permit conversion to glutamate, since evidence for histamine production (from histidine) in CHO could not be found. The tryptophan demand reaction was added as tryptophan is converted to many metabolites with diverse functions (e.g., hormones), at least some of which have been

detected in CHO cell culture[88]; however, we were uncertain whether the production of *specific* tryptophan-derived metabolites is cell line specific or a characteristic of the CHO cell in general, thus the inclusion of a generic demand reaction rather than specific biosynthetic pathways. Three additional reactions were added to the DG44 model, the mitochondrial and endoplasmic reticulum localized phosphatidylethanolamine N-methyltransferase and methionine adenosyltransferase (PETOHMm_cho, PETOHMr_cho, and METAT, respectively) to allow growth in the absence of measured choline uptake. The ORNTArm, GLU5Km, and G5SDym reactions were constrained to be off (lower bound=upper bound=0) to reflect the fact that—even if present—the enzymes do not appear to carry flux, leading to the experimentally observed proline auxotrophy in CHO cell lines[14].

Omic data generation

Transcriptomics and Proteomics

CHO-S

Starting at 72 hours into culture and every 12 hours after, cells were collected for proteomic (5 time points) and transcriptomic analysis (10 time points). 5 x 10^6 cells were harvested for transcriptomic analysis via RNA-Seq. Cells were centrifuged and the pellet resuspended in 2%/98% DTT/RLT buffer and stored at -80°C. RNA was extracted with Qiagen's RNeasy mini kit (Qiagen #74104) according to manufacturer's protocol with on-column DNase digestion. RNA libraries for sequencing were prepared using TruSeq Stranded mRNA Sample prep kit with 96 dual indexes (Illumina, CA, USA) according to the manufacturer's instructions with the following changes. The protocols were automated using an Agilent NGS workstation (Agilent, CA, USA) using purification steps as previously described[89,90]. Libraries were clustered using cBot and sequenced on HiSeq2500 (HiSeq Control Software 2.2.38/RTA 1.18.61) with a 2x101 setup in RapidHighOutput mode. The raw reads are available at GEO (Accession GSE77800). Bcl to

Fastq conversion was performed using bcl2Fastq v1.8.3 from the CASAVA software suite. The quality scale is Sanger / phred33 / Illumina 1.8+.

FastQC[91] was used to assess read quality. Trimmomatic v0.32[92] was used to trim reads with adapters or low quality scores. STAR2.4.0a[93] was used to align trimmed reads to the CHO-K1 genome[8]. Mapping results were stored using SAMtools 1.0[94]. Cufflinks 2.2.1[95] was used to assemble mapped reads and quantify expression levels. FPKM levels for all time points are available in Supplemental Data S4[58].

For proteomic analysis, 1 x 10^7 cells were harvested and centrifuged. Pelleted CHO cells were lysed in 100 µl urea (8M, 75 mM NaCl, 50 mM Tris-HCl pH 8.2). To assist the rupture of the cells, two 3 mm Zirconium oxide beads were added and samples were placed in a mixer mill (Glen Mills Inc, NJ, USA). Two rounds of mixing were applied, with the first 2 min at 25 Hz in the mixer mill followed by 30 min at 4 °C and additional 2 min at 25 Hz in the mixer mill. Between the two rounds, an additional 100 µl of urea were added. Samples were then centrifuged for 15 minutes at 15000 g, after which the supernatant was collected. Then 400 µl 25 mM ammonium bicarbonate was added and the volume reduced to 100 µl using a 3 kDa cutoff filter. Five µl DTT were added to samples containing 100 ug of protein and then kept at 37 °C for 45 minutes, after which 100 µl of iodoacetamide was added and samples were kept in the dark for 45 minutes. Tryptic digestion of the proteins was done for 8 hours at 37 °C. Digestion was terminated with the addition of 10 µl 10% TFA and finally samples were staged tipped using C18 filters (Empore, 3M Company, USA) following an established protocol[96].

Each sample was trapped on a precolumn (Symmetry C18 5µm, 180µm x 20mm, Waters, Manchester UK) and washed for 4 min after which it was loaded on the analytical column. The analytical setup consists of a nanoACQUITY ™ System (Waters, Manchester UK) equipped with a nanoACQUITY ™ BEH130 C18 1.7 µm, 75 µm x 250 mm analytical reversed-phase column (Waters, Manchester UK). The reverse phase elution profile included mobile phase A (0.1 formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile), during which B was increased

from 5-40% over 90 minutes with a flow rate of 250 nL min^{-1} and a column temperature of 35°C. To minimize carry over, the method included a 30 minutes wash phase to clear the column.

Data acquisition was accomplished on a Synapt G2 (Waters, Manchester UK) Q-TOF instrument using ESI with a NanoLock-spray source. The mass spectrometer was operated in positive and resolution mode with continuum spectra being acquired. Data were continuously calibrated using leucine encephalin as lock mass. Data were acquired using MSE, during which the mass spectrometer alternated between low- and high-energy mode using a scan time of 0.8 s for each mode over a 50-2000 Da interval. In the low-energy MS mode, data were collected at constant collision energy of 4 eV. In the elevated-energy MS mode, the collision energy was increased from 15 to 40 eV.

Raw data files for protein identification were obtained by using the ProteinLynxGlobalServer software v2.5.3 (Water corporation) using the in-build MSE search function against the Chinese hamster UniProt proteome database (UP000001075). The search parameters were trypsin as enzyme, carboxamidomethyl on cysteine as a fixed modification and oxidation of methionine as partial modification, while allowing one missed cleavage.

CHO-K1

Additional proteomic and RNA-Seq data were obtained from existing studies[8,97]. Briefly, for RNA-Seq, CHO-K1 cells were grown in F-12K medium (Invitrogen) supplemented with 10% FBS at 37°C with RNA extraction during exponential phase. Sequencing was carried out using Illumina GA2 technology with paired-end reads. Quantification of expression levels was carried out in an identical manner as for CHO-S. For proteomics, CHO-K1 cells were grown in F-12K medium, supplemented with 10% FBS, 1% non-essential amino acids, and 2 mM L-glutamine (Gibco) and gathered at 70-80% confluence for analysis.

CHO-DG44

Microarray data for an IgG-producing CHO-DG44 derivative were obtained from literature[98]. Briefly, the CHO-DG44 cells expressing IgG, known as CHO M250-9, were grown in a proprietary protein free and chemically defined medium. The total RNAs were extracted using the Qiagen RNeasy Plant Mini Prep kit during the exponential phase of cell culture. Subsequently, the gene expression was profiled with a NimbleGen CHO microarray chip containing 135,883 probes corresponding to a total of 13,514 annotated CHO genes. Scanned microarray signals were then analyzed by the NimbleScan V2.6 (Nimblegen, U.S.A.) and quantile normalized using the R package AffyPLM[99].

Metabolomics

CHO-K1

For each media sample, polar extracellular metabolites were analyzed by ultra performance liquid chromatography (UPLC) (Acquity, Waters, Manchester, UK) coupled in line with a quadrupole-time-of-flight hybrid mass spectrometer (Synapt G2, Waters, Manchester, UK) as previously reported[100].

For the analysis of targeted metabolites, data were processed using TargetLynx (Waters) while for untargeted analysis MarkerLynx (Waters) was used to integrate and align MS data points and convert them into exact mass retention time pairs. Extracted ion chromatograms were obtained by using a 0.02 mDa window centered on the expected m/z for each targeted compound. Quantitation was performed by external calibration with reference standards[100].

The identity of each metabolite was established by verifying retention time, accurate mass measurements and collision induced dissociation information against our in-house database and/or online databases, including HMDB and METLIN[101,102].

All materials used in the UPLC-MS experiments were purchased from Sigma-Aldrich (Germany) and were of analytical grade or higher purity.

Model simulation

When optimizing for growth, solutions were obtained by maximizing flux through the biomass_cho or biomass_cho_producing reaction, depending on if the simulation was of a non-producing or producing cell line, respectively. For predicting maximum protein production, solutions were obtained by maximizing flux through the DM_igg[g] or DM_epo[g] reaction, for production of IgG or EPO, respectively.

Growth-rate prediction discrepancies

Serum containing media formulations

When applied as constraints, the uptakes from two CHO-K1 MFA studies were unable to recapitulate the experimentally observed growth rate for the CHO-K1 cells grown in serum. The cause of this discrepancy is due to the low uptake of proline in one study[62] and the production of proline by the cells in another study[61], despite reports of the proline auxotrophy being a hallmark of the cell line. It is possible that the presence of serum (and associated protease activity or peptide uptake and catabolism) may 'mask' a higher uptake rate of proline required to sustain the higher observed growth. However, we also investigated whether spontaneous proline prototroph revertants[12] could also explain the difference in calculated vs. observed growth rates. To do so, we first added an extracellular arginase reaction to the models[103], and then simulated growth with individual or both proline biosynthetic pathways active, compared to simulated proline auxotrophy. Inclusion of active proline biosynthetic pathways in the model greatly improved the consistency of model predictions with measured growth (Figure 2.S2), indicating that increased proline availability, either via biosynthetic routes or via unmeasured uptake from serum components, can explain the discrepancies between model predicted and experimental growth rates.

Metabolomic measurement uncertainty impacts growth rate prediction accuracy

We also examined how variability in uptake and secretion rate calculations impacts the accuracy of model predictions. To do this test, we focused on three data sets in which computed uptake and secretion rates for some metabolites demonstrated large standard deviations. These included three lower quality cultures generated by the authors of this study. In these experiments, the calculated range (between 5^{th} and 95^{th} percentile confidence intervals) for more than half of the metabolite uptake or secretion rates was larger than the predicted best fit value (Figure 2.S1). For completeness, we show this metric compared to two cultures from a recent temperature shift study[50] (labeled “Cold 1” and “Cold 2” in Figure 2.S1) that exhibited lower variability and are the same as those included in the main text (Figure 2.3).

We took the calculated values for amino acids, glucose, lactate, and IgG (if produced) and generated a family of uptake and secretion values (within the 5^{th}-95^{th} percentile confidence intervals) for each metabolite based on its predicted directionality of flow. For example, essential amino acids were forced to be taken up; alanine, traditionally secreted, was forced to be secreted. More formally, we generated 3000 sets of uptake and secretion values satisfying the following criteria. If metabolite is forced to be taken up, we generated random uptake values between 5^{th} percentile value and min(0, 95^{th} percentile value). If metabolite is forced to be secreted, than we generated random secretion values between max(0, 5^{th} percentile value) and 95^{th} percentile value. Otherwise, we generated a random uptake or secretion value between the 5^{th} and 95^{th} percentile values. Each set of values was applied as constraints to the model and used to predict growth, and the results compared to the growth rate prediction obtained using just the calculated best-fit uptake/secretion values (Naive Uptakes bars in Figure 2.S3). This approach led to predictions more in agreement with experimental measurements (Figure 2.S3), highlighting the importance of accurate metabolomic measurements.

Uptake flux generation

Lower limits for amino acids and glucose were defined based on experimental measurements from the studies used for phenotype validation. The growth rate and production rate from the non-treated culture were set as constraints on the producing cell biomass reaction and IgG/EPO production (as appropriate) in the appropriate cell-line specific model. Iteratively, the following procedure was followed. First, an amino acid or glucose was randomly selected. If an uptake rate for the metabolite is not known, then we found the minimum and maximum allowable uptake/secretion rate for the nutrient that permits growth and production at the experimentally determined rate. Then the uptake or secretion of the nutrient was set to a randomly selected value within that range. This was then repeated until all amino acids and glucose had uptake or secretion values. For each study, 3000 sets of uptake fluxes were generated. Each set of uptake fluxes was checked by ensuring that removal of full constraints on protein production or growth resulted in a production/growth rate within 1% of the experimentally measured value. Uptake fluxes which satisfied this criterion were used for further analysis and are available in Supplemental Data S5[58].

Algorithmically generated uptake and secretion rates are consistent with experimental measurements

Results for nutrient uptake and secretion flux generation were validated using data from a previous study[27]. The algorithm was applied at the growth rate and specific productivity measured during early exponential phase and the range of resultant nutrient fluxes (available in Supplemental Data S5[58]) is compared to the experimental values (Figure 2.S4). For 19 out of 20 measured metabolites (all except phenylalanine), the experimental uptake or secretion value was within the bounds of our algorithm-generated values. For phenylalanine, the deviation from the calculated range was very small: approximately 4.4×10^{-4} mmol gDW^{-1} hr^{-1} (2.8% of the median

predicted value). Thus, it is clear that the predicted fluxes were consistent with experimentally measured fluxes.

Metabolite uptake and secretion rates are consistent before and after treatment

The resource redirection efficiency analysis assumes that metabolite uptake and secretion rates do not significantly change after treatment. This assumption was assessed by analyzing data from a study examining treatment of cells with sodium butyrate (NaBu)[46]. In this study, essential amino acid fluxes (which are limiting for growth and protein production) remain fairly stable before and after NaBu treatment (Figure 2.S5). In fact, the majority of changes were actually in the direction of increased metabolite uptake, which—if extrapolated to all simulated uptakes for NaBu treated cell lines—would further decrease the calculated efficiencies and yields for those treatments.

Redirection analysis

For each 'treatment', the model growth rate was constrained to the experimentally measured value while protein production was unconstrained. After applying a set of uptake flux values, the model was simulated while optimizing for maximum protein production. This was repeated for all flux values passing the check for growth/production rate mentioned previously. The experimentally observed production rate was compared to the family of simulated production rates.

For the data stemming from a temperature shift treatment[64], fluxes were scaled by a random factor between 0.517 and 0.798 based on the difference in uptake rates observed for glucose and glutamine between cultures with and without a temperature shift.

Calculation of growth and protein production tradeoff

For each set of uptake fluxes passing the check for growth and production rates, the simulated maximum possible growth rate was calculated by not forcing any protein production and optimizing for biomass production (i.e. flux through biomass_cho_producing). Maximum simulated protein production was calculated by setting the growth rate to 0 and optimizing for protein production. The fraction of maximum protein production at various fractions of maximum growth rate were calculated by constraining growth rate to a fraction of the simulated maximum growth rate for a specific uptake flux and then optimizing for protein production (normalized by maximum protein production for that uptake flux). The growth rates and production rates from the treatment papers were normalized by the family of maximum growth rates and production rates, respectively, to evaluate production efficiency.

Quantification of model constraints

Metabolite uptake and secretion rates were quantified and used as constraints for the model simulations (see Supplemental Data S3[58]). The workflow for quantification and integration of these exchange rates is presented below. Experimental growth rates were calculated by determining the slope of the linear polynomial fit to the natural log of the viable cell densities.

Serum-grown CHO-K1, non-producing (In house)

Uptake and secretion rates were computed by calculating the slope of a linear polynomial fit to metabolite concentrations in the spent and control media vs. the integral of cell concentration (with respect to time) for exponential phase. The sample at 36 h was excluded due to poor quality (i.e., there were spurious jumps in some metabolite concentrations, which was inconsistent with measurements of preceding and subsequent time points). The final exchange rates were calculated by subtracting the control media rate from the spent media rate. Cell dry weight was set at 216.1 pg/cell, based on component weights (see Supplemental Data S2[58]). The metabolite

exchange rates were consistent with the expected rates for growth, except for arginine and cysteine equivalents, which showed a net efflux, despite being essential in CHO[13] (also see Figure 2.S3).

CHO M250-9 cells grown in protein-free, chemically defined (PFCD) media

Previously published data were acquired for the metabolite uptake and secretion rates[27]. Since no information was available on tryptophan uptake, a previously published[46] uptake rate of 0.0032 mmol gDW^{-1} hr^{-1} was used.

Butyrate-treated CHO cells

Previously published uptake and secretion rates were obtained for exponential phase of CHO cells producing high and low quantities of protein under control and butyrate treatment while growing on CD CHO medium[46]. Since cysteine uptake was not measured, an uptake rate of 0.0052 mmol gDW^{-1} hr^{-1} for cysteine was obtained from a previously reported value[27]. This value is qualitatively consistent with a previous report[104] showing that cysteine is taken up.

CHO-K1 from MFA studies

Exchange rates were taken from previous studies[61,62] and supplemented with the values used in the associated metabolic flux analysis (MFA) model simulations for metabolites for which experimental data were not available. As both studies reported uptakes on a per-cell basis, fluxes were calculated after scaling to a cell dry weight of 216.1 pg/cell, 315 pg/cell, and 350 pg/cell since dry weight composition was not measured. These flux values thus cover the range of observed cell weights for CHO[50,105]. As the cells were grown in 10% FBS, an extracellular arginase reaction was added to the models[103].

CHO-K1 derivative producing IgG in serum-free medium

Data were taken from a CHO XL99-Ab2 cell line producing an IgG1 antibody[50]. Uptake and secretion rates were calculated by fitting a linear polynomial to metabolite concentration vs. the integral of cell concentration (with respect to time, prior to temperature shift) and scaled using a cell dry weight of 350 pg/cell. We discuss the control cultures (no temperature shift) in the supplementary information (Figure 2.S1/S3). Since cysteine uptake was not measured, an uptake rate of 0.0052 mmol gDW^{-1} hr^{-1} for cysteine was obtained from a previously reported value[27].

Data and Software Availability

Raw data files for CHO-S RNA sequencing have been deposited in the NCBI Gene Expression Omnibus under accession number GSE77800. All models are available at http://www.CHOgenome.org. Additionally, the global model can be browsed and downloaded at the BiGG Models database (http://bigg.ucsd.edu).

SUPPLEMENTARY FIGURES

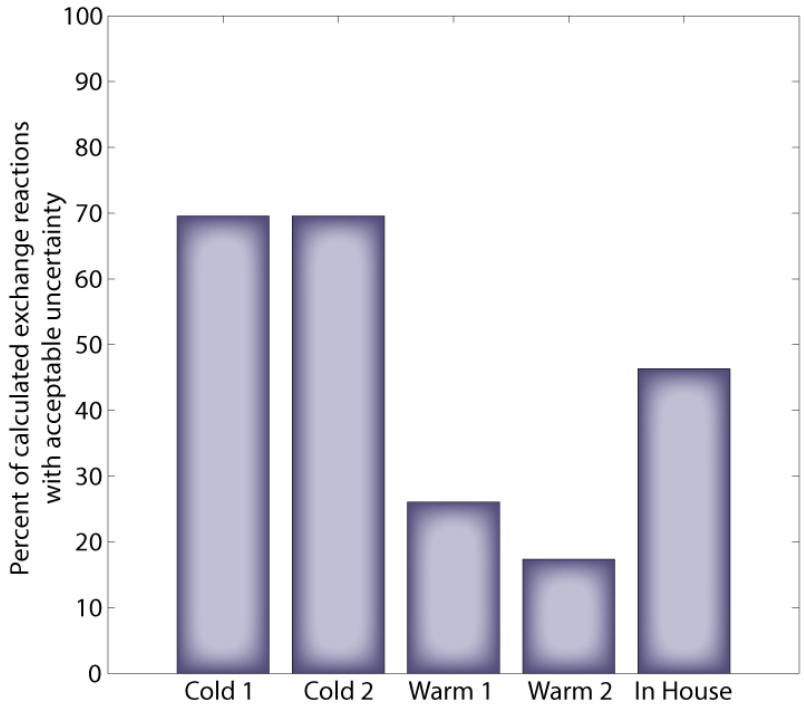


Figure 2.S1: Percentage of calculated exchange reactions for various cultures meeting quality criteria.

Our quality criteria is defined as having a best-fit value larger than the range between the 5th and 95th percentile confidence interval values. Cold 1 and 2 refer to measurements from cultures growing at 32 °C from a prior study[50], albeit prior to the temperature shift, and are included in the main text. Warm 1 and 2 were measurements from cultures grown at 37 °C from the same study.

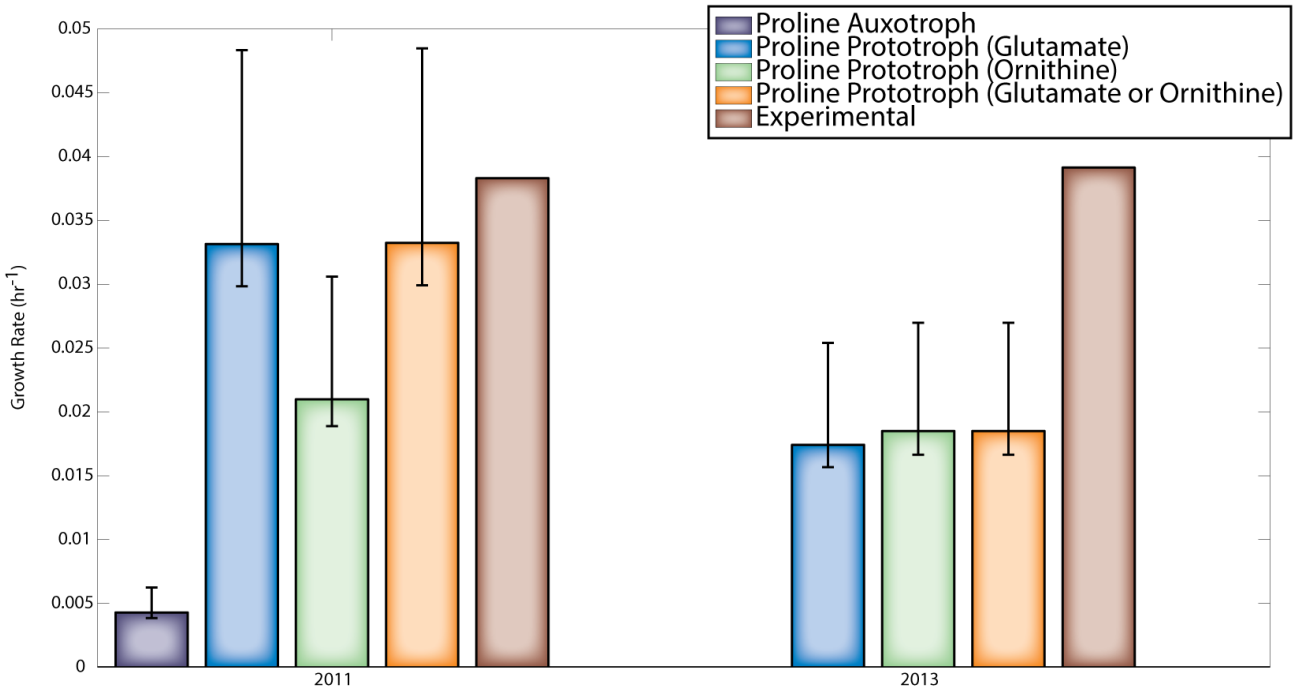


Figure 2.S2: Effect of simulated proline prototrophy on the consistency of model-predicted growth rates.

Simulations done for MFA studies from 2011[62] and 2013[61]. Since the studies did not measure cell sizes, we vary cell dry weight using a range of cell sizes from other studies. The bottom whisker is calculated assuming a dry weight of 350 pg/cell[50], the top whisker at a dry weight of 216 pg/cell (by summing macromolecule masses—e.g. protein, DNA, RNA, etc.—from multiple sources[106–108]), and the bar height at a dry weight of 315 pg/cell reflecting the assumptions made in the MFA studies[61,62]. For the data from 2013, the given uptake and secretion rates were not predicted to be able to support growth in a proline auxotrophic cell line due to observed proline secretion. Proline prototroph simulations were carried out with biosynthetic pathways (via glutamate or ornithine, see Figure 2.4) active individually or together.

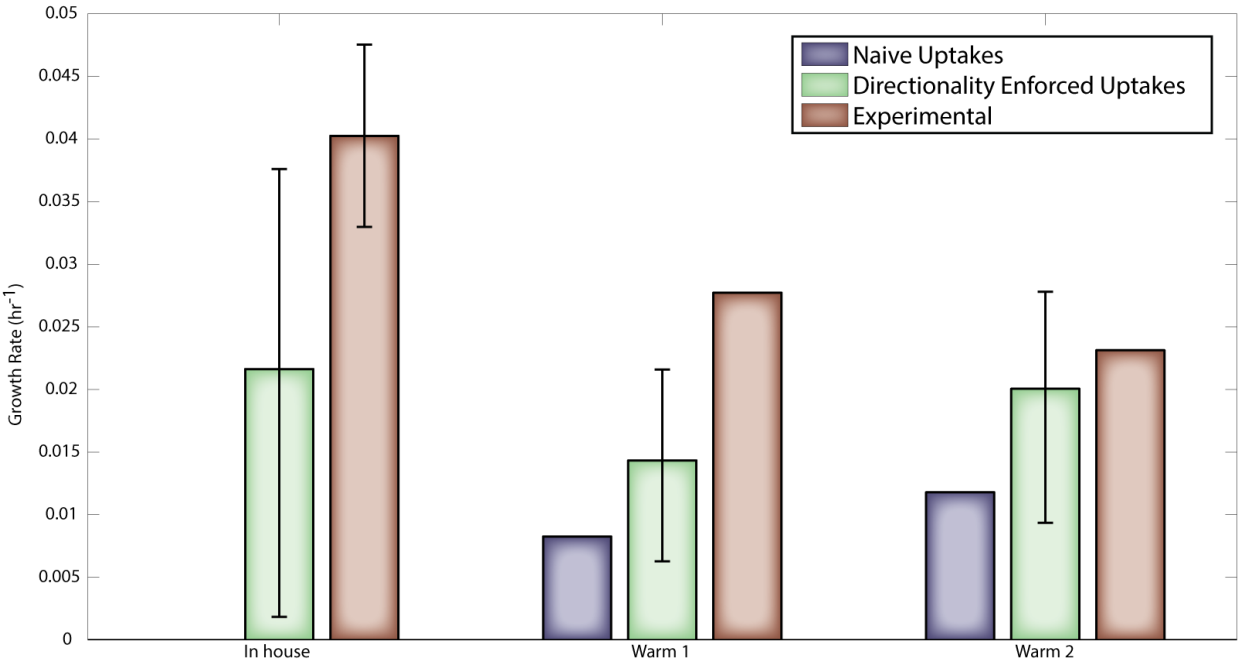


Figure 2.S3: Effect of enforcing uptake and secretion directionality on model-predicted growth rates.

Error bars for the directionality enforced uptake values indicate the 5th/95th percentile confidence intervals for the family of 3000 simulated uptake and secretion rates, while those on the In house experimental measurement reflect 5th/95th percentile confidence intervals for the measured growth rate. Variation in metabolomic measurements that lead to changes in direction of uptake and secretion can be seen to have a pronounced effect on the accuracy of model simulations.

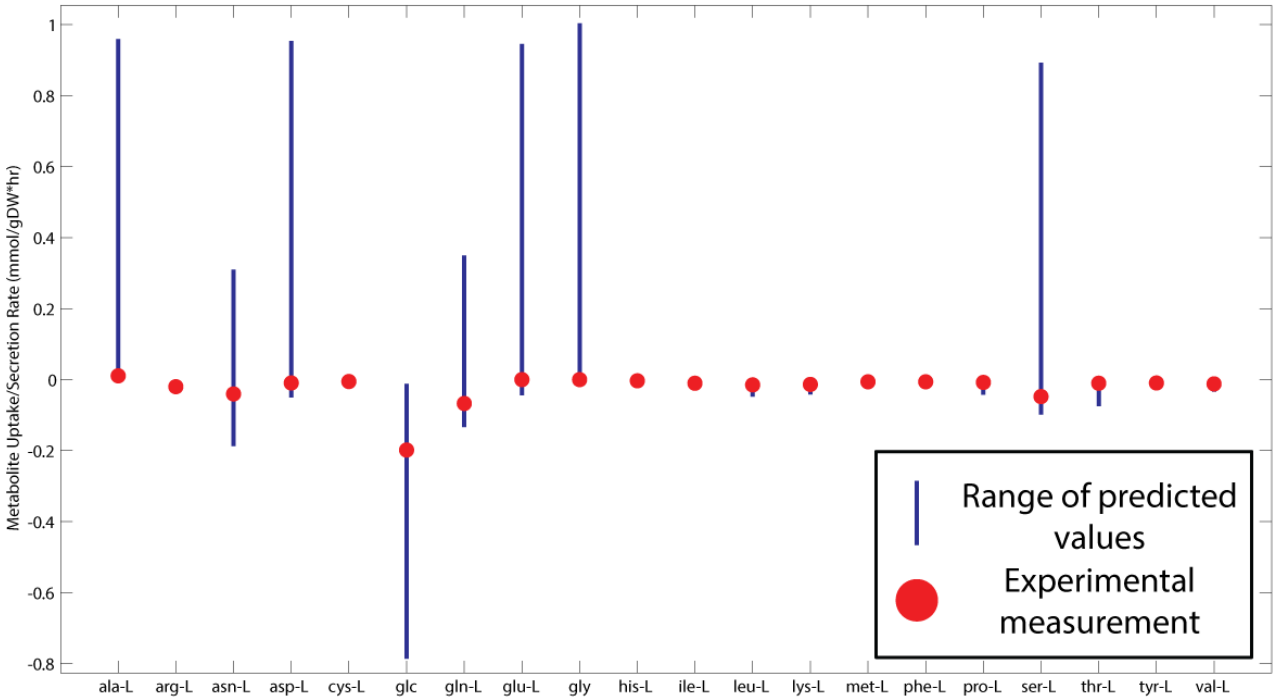


Figure 2.S4: Validation of uptake/secretion flux generation.

The red circles indicate experimental measurements while vertical lines denote the range of uptake/secretion values generated by our algorithm, available in Supplemental Data S5[58]. As shown here, actual experimentally measured uptake and secretion rates of the amino acids are within the range of uptake and secretion values generated by our algorithm.

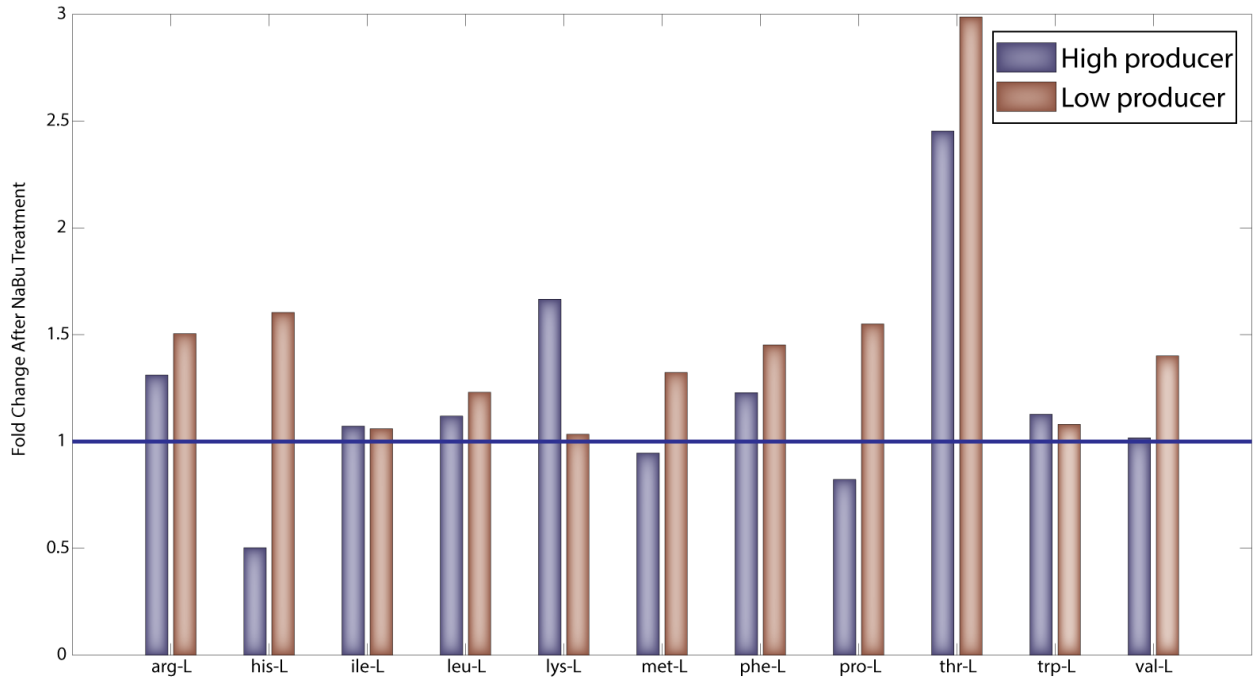


Figure 2.S5: Comparison of essential amino acid fluxes before and after NaBu treatment.
High producer (blue) and low producer (red) fold changes in essential amino acid uptakes after NaBu treatment. No change is indicated by the horizontal line. Data are taken from a study examining the effect of NaBu treatment on cell behavior[46]. Essential amino acids are generally taken up at similar (or higher) rates post-NaBu treatment, meaning simulations with unchanging uptakes provide upper limits to redirection efficiency.

Chapter 2, in full, is a modified version of the material as it appears in Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., Martinez, V.S., Kyriakopoulos, S., Jimenez, N.E., Zielinski, D.D., Quek, L.E., Wulff, T., Arnsdorf, J., Li, S., Lee, J.S., Paglia, G., Loira, N., Spahn, P.N., Pedersen, L.E., Gutierrez, J.M., King, Z.A., Lund, A.M., Nagarajan, H., Thomas, A., Abdel-Haleem, A.M., Zanghellini, J., Kildegaard, H.F., Voldborg, B.G., Gerdtzen, Z.P., Betenbaugh, M.J., Palsson, B.O., Andersen, M.R., Nielsen, L.K., Borth, N., Lee, D.Y., Lewis, N.E. A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary (CHO) Cell Metabolism. *Cell Systems*, 3, 434-443 (2016). I was a co-first author, while the co-authors provided support in the research underlying the study.

Chapter 3: Multiplex genome editing eliminates the Warburg Effect

INTRODUCTION

Lactate is a key metabolite, involved in many important processes. One of its prominent roles is in the Warburg effect[109], in which highly proliferative cells (e.g., cancer, immune, and stem cells) exhibit high rates of glycolytic flux with secretion of lactate even in the presence of oxygen[110–113]. Despite nearly a century of study, many questions remain surrounding its purpose, but it is believed to benefit cell proliferation[114]. However, aerobic glycolysis has posed a considerable challenge in biotherapeutic protein production, since mammalian production cells often secrete high levels of lactate. Lactate accumulation is deleterious for cell growth, viability, protein production, and product quality, both directly from media acidification and indirectly through increased osmolarity as base is added to control culture pH[39,115,116]. Accordingly, a significant amount of work has aimed to control the Warburg effect in bioprocessing[44].

In mammalian cells, lactate is produced by lactate dehydrogenase (Ldh). There are three Ldh isozymes: Ldha, Ldhb, and Ldhc, with different substrate specificities[117], such as Ldhc's preference for catalysis of the reverse reaction, forming pyruvate[118]. Ldha and Ldhb form homo- or heterotetramers with tissue specific distributions[119]. Ldha shows increased activity in many cancers[120] and is present in the Chinese hamster ovary (CHO) Ldh complex[121]. While deficiencies in either LDHA[122–124] or LDHB[125], have been reported clinically in humans and are not lethal, efforts to delete Ldha in mammalian cell bioprocessing have been unsuccessful[126–128], pointing to the essentiality of Ldh-mediated NAD^+ regeneration. Many approaches have been utilized to limit lactate accumulation in culture, including knockdown of Ldh[129], replacement of glucose with alternate sugars[130], controlled feeding strategies[115], and others[44]; however none have proven successful in eliminating Ldh activity entirely.

Underpinning the enzymatic conversion of pyruvate to lactate by Ldh, the Warburg effect in mammalian cells is controlled by complex regulatory mechanisms[131]. Metabolically, there is a negative feedback loop around pyruvate. Glucose is converted to pyruvate via glycolysis and

pyruvate sits at the branch between fermentation to lactate by Ldh or oxidative metabolism starting with the pyruvate dehydrogenase complex (Pdh). A negative feedback loop involving several proteins and metabolites regulates the metabolic fate of pyruvate (Figure 3.1A).

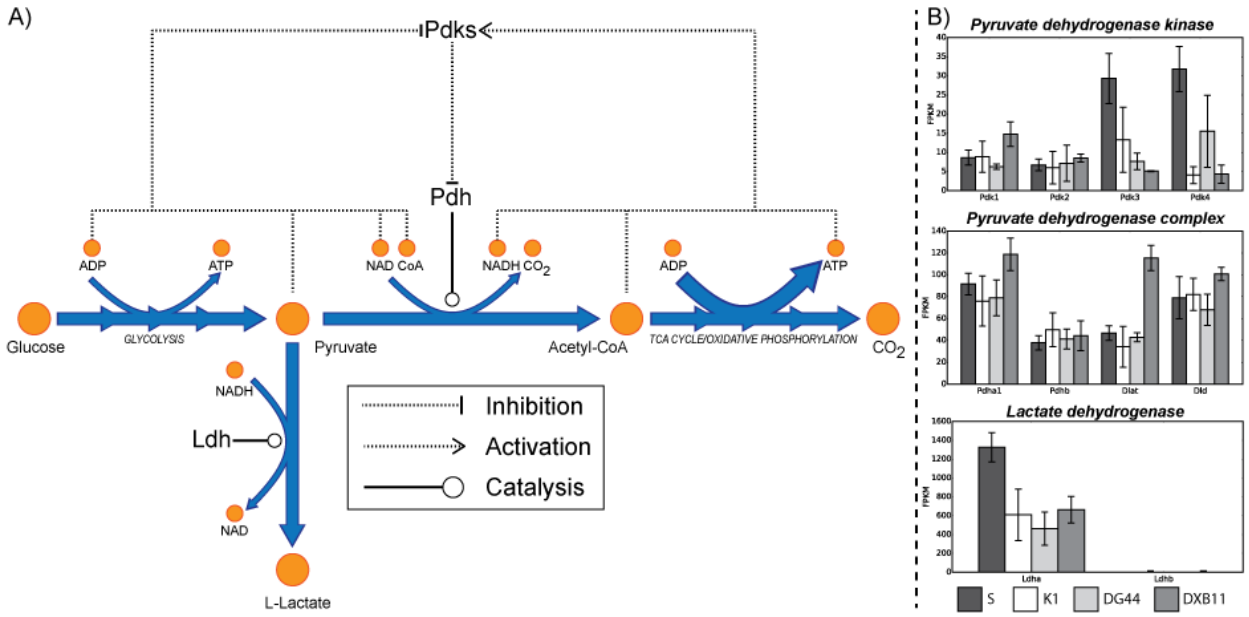


Figure 3.1: The Warburg effect is influenced by a regulatory circuit involving multiple metabolites and proteins.

(A) Pyruvate sits at a branch point between fermentation through lactate dehydrogenase (Ldh) or oxidative metabolism starting with the pyruvate dehydrogenase (Pdh) complex. Pyruvate dehydrogenase kinase isoforms (Pdks) are regulated by the products and substrates, of the Pdh reaction, forming a negative feedback loop that reinforces an increase in lactate secretion when glycolytic flux is high. **(B)** Gene expression data from major CHO cell lines shows that Pdh complex subunits are all expressed, as are all four Pdk isoforms. Ldhb shows near zero expression while Ldha expression is high. All gene expression values are given in fragments per kilobase per million reads (FPKM), and are shown as the mean +/- standard deviation, data sources are provided in Table 3.S1. Gene abbreviations are as folows: Pdk1-pyruvate dehydrogenase kinase 1, Pdk2-pyruvate dehydrogenase kinase 2, Pdk3- pyruvate dehydrogenase kinase 3, Pdk4-pyruvate dehydrogenase kinase 4, Pdha1-pyruvate dehydrogenase E1 alpha 1 subunit, Pdhb-pyruvate dehydrogenase E1 beta subunit, Dlat-dihydrolipoamide S-acetyltransferase, Dld-dihydrolipoamide dehydrogenase, Ldha-lactate dehydrogenase A, Ldhb-lactate dehydrogenase B.

Specifically, the first step to oxidative metabolism, Pdh, can be regulated by four pyruvate dehydrogenase kinases (Pdks). Increases in the ratios of acetyl-CoA / CoA, ATP / ADP, or NADH / NAD^+ lead to the increased activation of Pdk, albeit with different efficacies[132]. Pdks, when activated, phosphorylate and inhibit Pdh. Conversely, pyruvate inhibits Pdk1 and 2--thereby alleviating their inhibition of Pdh--but not Pdk3 or 4[132]. As excess flux through Pdh leads to Pdh inhibition, Ldh-mediated reduction of pyruvate to lactate functions as a release valve for the high

glycolytic activity that occurs in proliferative cells. Thus, even in the presence of oxygen, glucose is not fully oxidized and lactate production is observed.

Here we report that removing the feedback loop around pyruvate permitted the first successful elimination of Ldh activity in a mammalian cell line. We achieved this through the homozygous knockout of Ldha and the four Pdks in CHO cells. The cells remain proliferative while consuming significantly less glucose without increasing oxidative metabolism to compensate for the loss of glycolysis-derived ATP. These results suggest that the Warburg effect is not essential to obtain the necessary ATP and biomass precursors to fuel mammalian cell growth and thus the Warburg effect must serve an alternative purpose.

RESULTS

Ldha can be knocked out when concurrently targeting Pdk1, 2, 3, and 4

In all CHO cell lines we examined, all four Pdk isozymes and core Pdh subunits are expressed. While Ldha is highly expressed, Ldhb is silenced, indicating that the Ldh enzyme is made exclusively of Ldha subunits (Figure 3.1B). Accordingly, CHO cells likely have this feedback loop and indeed produce lactate growing aerobically. Previous efforts to knockout Ldha in CHO cells have been unsuccessful[126–128], possibly because the knockout of Ldha would lead to accumulation of pyruvate, which cannot be fully processed by Pdh due to activation of Pdks inhibiting Pdh.

We thus aimed to remove Ldha and the Pdk-mediated feedback loop with the simultaneous removal of Ldha and all four Pdks. We hypothesized that this would relieve Pdk-mediated inhibition of Pdh and allow uninhibited flux of pyruvate through Pdh. Therefore, using CRISPR/Cas9, we simultaneously targeted all five genes in the CHO-S (WT) cell line (Table 3.S2). Consistent with our hypothesis, we isolated several clones with frameshift mutations in all alleles of the five genes (see Methods for details) (Tables 3.S3/S4).

Further characterization of these clones confirmed that the Warburg effect was eliminated. We measured Ldha by Western blot, showing that protein expression was undetectable (Figure 3.2A, Figure 3.S1). As further support, we tested the cell lysates from knockout clones for Ldh activity and found that it was reduced to near zero (Figure 3.2B). Finally, lactate was undetectable during exponential growth when assayed by the BioProfile 400 (Figure 3.2C). To obtain a more rigorous quantification of the decrease in lactate levels, we employed more sensitive methods in subsequent experiments testing a variety of clones and culture conditions. We found that maximum lactate concentrations obtained were in the sub-millimolar range for clones 0-F4 and 0-F5 while clone 0-B6 reached single millimolar concentrations (Figure 3.S2). When compared to values observed in wildtype or Warburg-positive clones/pools (maximum concentrations in the 10s to low 100s mM lactate) it is apparent that the deletion of Ldha effectively abolished Warburg metabolism in these cells.

Cell growth of Ldha knockout enhanced by knockout of all Pdks

In addition to clones with all five genes knocked out, we found a minority of clones with frameshift mutations in Ldha, while retaining some wildtype Pdk alleles. This was surprising, as previous work targeting Pdk1-3 with siRNAs could not isolate a complete Ldha knockout[127]. Therefore, we conducted four more rounds of CRISPR/Cas9-mediated engineering targeting all combinations of three Pdks alongside Ldha (e.g., Ldha + Pdk 1, 2, and 4) to verify that the Warburg effect could be eliminated without knocking out all Pdks. In total we single-cell sorted 378 clones (93-95 clones per 3-Pdk combination) for genotyping by sequencing.

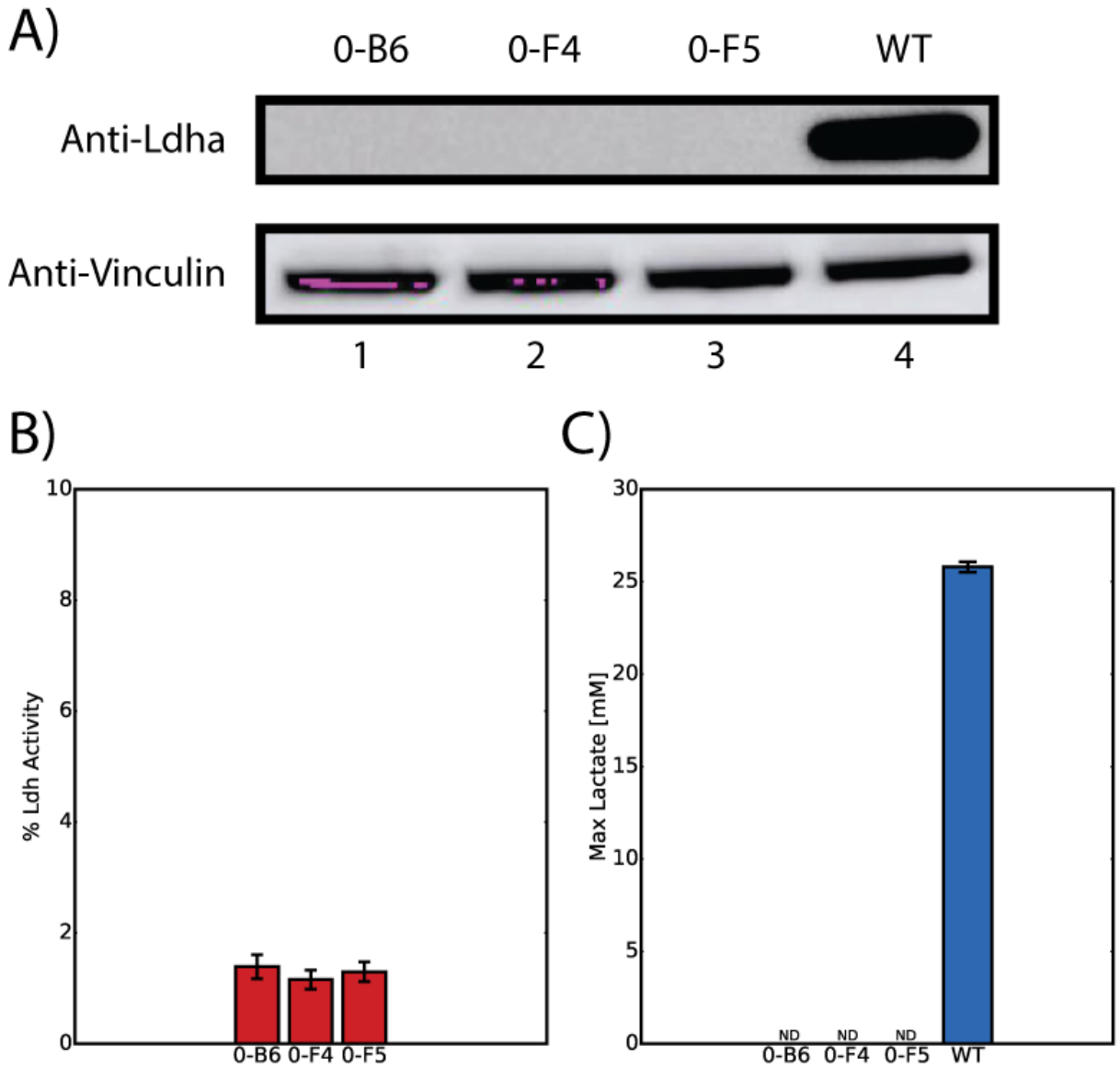


Figure 3.2: Ldha is successfully knocked out and the Warburg effect is eliminated.
(A) Simultaneously targeting of Ldha and the Pdk genes using CRISPR/Cas9, leads to no detectable Ldha expression in any clones, as verified by chemiluminescent Western blot. **(B)** Enzymatic assay verifies a loss of lactate dehydrogenase activity in knockout lines (shown here as percent of wildtype Ldh activity, n=3). **(C)** Measurement of maximum lactate levels during growth in batch culture shows that lactate cannot be detected (n=2). Data shown as mean ± standard deviation (uncertainty due to variance in background in **(B)** accounted for via propagation of error).

Between these clones and those obtained from targeting all Pdks and Ldha simultaneously (515 clones in total), we obtained a diverse set of clones with different combinations of gene knockouts. First, consistent with previous work[126–128], we were unable to obtain stable, viable clones with only homozygous Ldha frameshift mutations. Second, we found

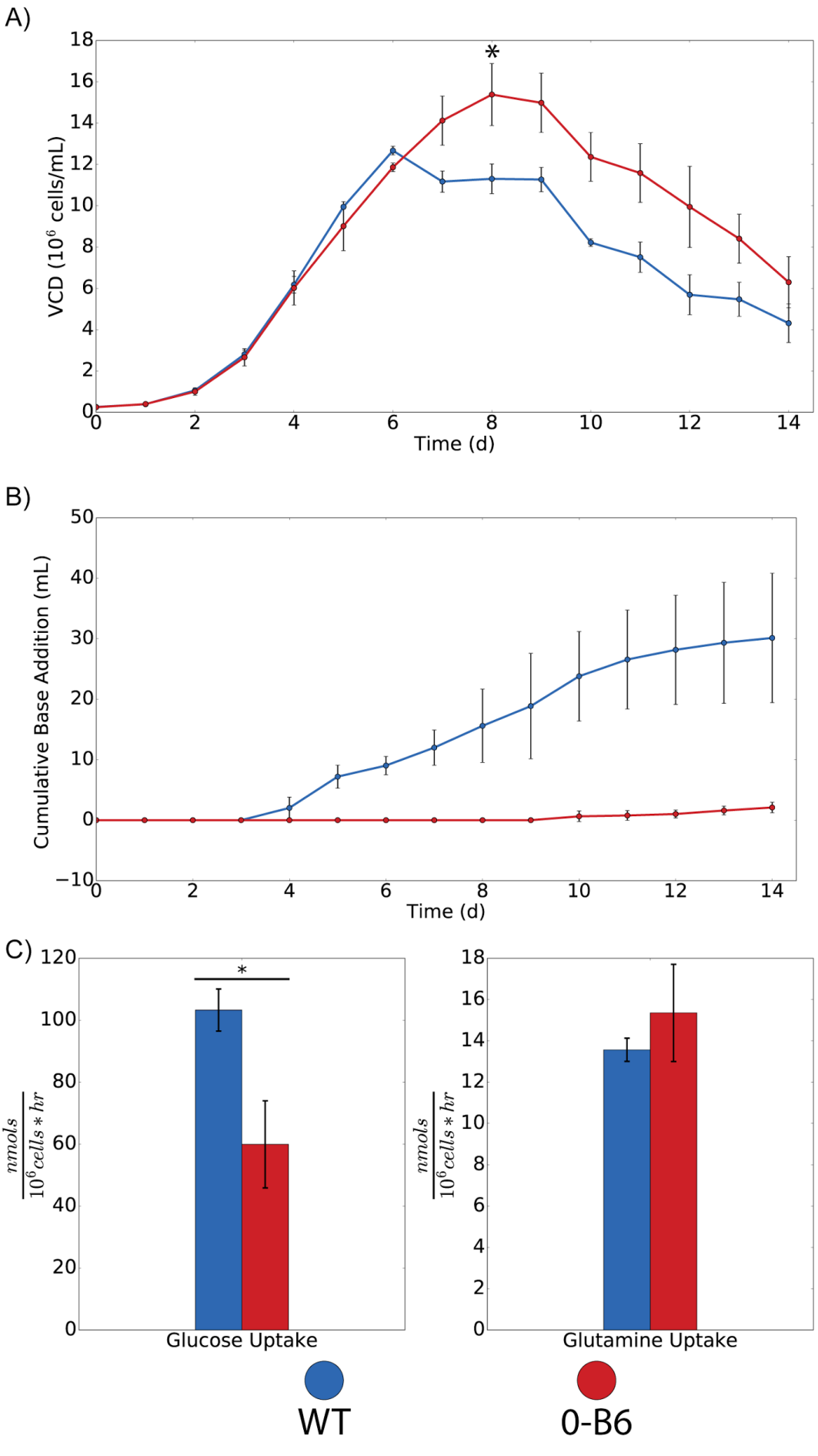
several clones harboring complete Ldha knockouts alongside diverse combinations of Pdk knockouts, including some with only 1, 2, or 3 Pdks eliminated (Tables 3.S2/S3), which also had the Warburg effect eliminated (Figure 3.3). However, the clones with more Pdks knocked out showed a significant trend towards faster growth ($p < 0.0012$; Figure 3.S4), indicating residual Pdk activity may have a mildly deleterious effect in Warburg-null lines. We note, however, that clones remained proliferative even with only one Pdk knocked out alongside Ldha.

The elimination of the Warburg effect enhances proliferation

It has been hypothesized that the Warburg effect enhances cell proliferation by allowing increased production of biomass precursors[114]. Therefore, we further characterized two clones with all Pdks and Ldha knocked out (0-B6 and 0-F5). We cultivated these clones in fed-batch experiments with WT controls to assess if the Warburg-null phenotype slowed growth or proved advantageous in a prolonged culture. Both KO clones exhibited comparable growth rates to the WT, but also a prolonged (additional 1-2 days) growth phase, leading to a higher maximum cell density (Figure 3.3A, Figure 3.S5A). The extended growth phase is not due to different essential amino acid depletion profiles between the WT and KO cultures, since depletion occurs for both cell types days after growth ceased in WT cultures (Figure 3.S6). Rather, entrance to stationary phase for WT cells corresponds with osmolite accumulation from base used to control media pH during lactate accumulation, which is absent in KO cells (Figure 3.3B, Figure 3.S5B). Thus, even though the Warburg effect is thought to facilitate proliferation by providing biomass precursors, we see no change in the growth rate. Furthermore, there is actually an increase in cell density as osmolarity effects are alleviated.

Figure 3.3: Elimination of the Warburg effect results in cells that grow to higher densities and exhibit more efficient glucose metabolism.

Knockout clone 0-B6 (red, n=5) and wildtype (blue, n=3) cells were grown in fed-batch culture. **(A)** The knockout clone exhibits a growth rate similar to the WT but remains in exponential phase until nutrients are depleted (indicated by *, see (Figure 3.S6). **(B)** As a result of not secreting lactate, 0-B6 require less addition of base (1 M $NaHCO_3$) over the course of culture to maintain pH at the appropriate levels (starting culture volume ~270 mL). **(C)** During exponential growth (days 1-5), 0-B6 consumes significantly less glucose compared to the WT, but takes up similar amounts of glutamine. * indicates $p<0.05$ (two-sample two-tailed Wlech's t-test). Data shown as mean ± standard deviation.


A)
VCD (10^6 cells/mL)
Time (d)
*
B)
Cumulative Base Addition (mL)
Time (d)
C)
$\frac{nmols}{10^6 cells * hr}$
Glucose Uptake
*
$\frac{nmols}{10^6 cells * hr}$
Glutamine Uptake
WT
0-B6

Warburg-null cells decrease glucose metabolism without increased oxidative phosphorylation

Another hypothesized purpose of the Warburg effect is to increase ATP generation through substrate-level phosphorylation from high glycolytic flux[114]. Thus, we assessed how Warburg-null cells responded in glycolytic and oxidative metabolism. We found glucose uptake of 0-B6 and 0-F5 clones significantly decreased (compared to WT cells) during the exponential phase in fed-batch culture while glutamine uptake (the other major nutrient taken up at considerable rates in proliferative mammalian cells) was unchanged (Figure 3.3C, Figure 3.S5C). Interestingly, this decreased glucose uptake was observed for all Warburg-null clones grown in batch culture, without differences dependent on the number of Pdk knockouts. (Figure 3.S7)

The decrease in glycolytic metabolism should decrease ATP production. Thus, we wondered if oxidative metabolism increases to make up the ATP deficit. We first estimated Pdh flux (i.e. metabolic flux entering the TCA cycle and oxidative phosphorylation) by calculating the amount of carbon unaccounted for after measuring glucose, pyruvate, alanine, and lactate uptake/secretion rates during exponential growth. This demonstrated there was no change between WT and clone 0-B6 (Figure 3.S8). Furthermore, in a separate experiment, we found that oxygen uptake between 0-F5 and WT remained unchanged during exponential phase (Figure 3.S9), implying that oxidative metabolism is unchanged. Thus, eliminating the Warburg effect reduced glucose uptake significantly, but it surprisingly does not alter oxidative metabolism to compensate for lost ATP, suggesting that a need for increased ATP production cannot explain the Warburg effect in proliferating mammalian cells.

DISCUSSION

Here we present the first reported elimination of the Warburg effect via removal of lactate dehydrogenase and peripheral regulators in a proliferative mammalian cell line. The engineered CHO cells exhibit drastically decreased glucose consumption alongside near zero lactate

secretion. However, oxygen consumption is unchanged, indicative of a lack of compensation of lost ATP, normally generated from high glycolytic flux. Surprisingly, the cells remain rapidly proliferative, and, when grown in fed-batch, exhibit an extended growth phase. This is likely due to avoiding medium hyperosmolarity, since base addition is not needed to neutralize the effect of lactic acid accumulation.

The successful elimination of the Warburg effect in mammalian cells establishes a novel model system in which fundamental questions regarding the Warburg effect can be addressed. That is, the ubiquity of aerobic lactate production in proliferative mammalian cells has led to a rich field of study to understand how and why this phenomenon occurs[114]. However, while there have been numerous studies where lactate production has been inhibited to varying degrees[120,121,127,133,134] (e.g. using siRNA, chemical inhibitors, or CRISPR), the lack of proliferative Warburg-null cell lines (i.e. cells that cannot produce lactate while growing in the presence of oxygen) have hindered full exploration of this phenotype. While we have focused primarily on the biotechnological relevance of the Warburg effect, our study provides interesting insights into outstanding hypotheses and debates in the broader scientific discipline. First, it is commonly assumed that immortalized cells such as CHO and cancer have either dysfunctional mitochondria or insufficient mitochondrial capacity for needed ATP generation and therefore the Warburg effect is needed to support the energetic demands of growth, cell maintenance, and other activities[114,135,136]. However, while the Ldha knockout resulted in decreased glucose uptake and glycolytic flux, there did not seem to be a response to increase ATP generation by enhancing mitochondrial metabolism. Indeed, we were surprised to find oxygen uptake remained stable, indicating oxidative phosphorylation activity is unchanged. Thus, compensation for the loss of ATP generated via high glycolytic flux does not appear to occur in these cells. Second, many studies have suggested that increased glycolytic metabolism is necessary for producing biomass precursors for proliferation seen in cancer, T-cell activation, and embryonic development[114,137,138]. However, we found that near elimination of lactate production does not appear to harm the cells

or hinder their ability to proliferate. While we can therefore say that aerobic lactate production is not needed to meet energetic or biosynthetic demands of proliferative cells, the evolutionary benefit to this ubiquitous phenotype remains ripe for investigation. The genome-editing strategy proposed here provides an approach to develop new cell models to explore such questions.

In conclusion, we found that the elimination of the Warburg effect was feasible in mammalian cells by the elimination of lactate dehydrogenase activity. While prior studies showed that Ldha in CHO cells was essential[126–128], we found that its essentiality in CHO cells was due to its involvement in a regulatory circuit with negative feedback loops and that this circuit could be removed without any apparent negative effects, if multiple genes were simultaneously deleted. Thus, this highlights how seemingly necessary but undesirable traits may be engineered as we unravel their systems-level context[139,140], and that multiplex genome editing strategies[141] and combinatorial genome editing screens[142,143] will be particularly important in such research.

METHODS

Transfection

CHO-S™ cells (Gibco Cat. # A11557-01) were transfected using either FreeStyle MAX reagent (Gibco Cat. # 16447100) or FuGENE HD reagent (Promega Cat. # E2311). For both procedures, at the day prior to transfection, viable cell density was adjusted to $8x10^5$ cells/mL in transfection medium: CD CHO medium (Gibco Cat. #10743-029) supplemented with 8 mM L-glutamine (Lonza Cat. # BE17-605F). With the FreeStyle MAX procedure, at the day of transfection, viable cell density was adjusted to $1x10^6$ cells/mL in an MD6 plate (Falcon Cat. # 351146) containing 3 mL transfection medium per well. For each transfection, 1.9 µg Cas9-2A-GFP plasmid DNA and 1.9 µg gRNA plasmid DNA was diluted in 60 µL OptiPro SFM (Gibco Cat. # 12309019). Separately, 3.8 µL FreeStyle MAX reagent was diluted in 60 µL OptiPro SFM and the two mixtures were incubated for 5 minutes at room temperature. After incubation, the plasmid DNA/OptiPro SFM mixture was added to the FreeStyle MAX/OptiPro SFM mixture and incubated

at room temperature for additional 20 minutes. The resultant 120 µL DNA/lipid mixture was added dropwise to the cells in one well. With the FuGENE HD procedure, at the day of transfection, viable cell density was adjusted to $8x10^5$ cells/mL in an MD6 plate containing 3 mL transfection medium per well. For each transfection, 1.5 µg Cas9-2A-GFP plasmid DNA and 1.5 µg gRNA plasmid DNA was diluted in 75 µL OptiPro SFM. Separately, 9.0 µL FuGene HD reagent was diluted in 66 uL OptiPro SFM. The plasmid DNA/OptiPro SFM mixture was added to the FuGENE HD/OptiPro SFM mixture and incubated at room temperature for 5 minutes and the resultant 150 µL DNA/lipid mixture was added dropwise to the cells in one well. Plasmids were constructed using the uracil-specific excision reagent (USER) cloning method as described previously[144], with the sgRNA1_C plasmid as a backbone.

Single cell sorting and expansion

Transfected cells were single cell sorted 48 hours post transfection, using the FACSJazz, based on green fluorescence with gating determined by comparison to non-transfected cells. Sorting was done into MD384 plates (Corning Cat. # 3542) containing CD CHO medium (Gibco Cat. # 10743-029) supplemented with 8 mM L-glutamine (Lonza Cat. # BE17-605F), 1% antibiotic-antimycotic (Gibco Cat. # 15240-062), and 1.5% HEPES buffer (Gibco Cat. # 15630-056). After 15 days, colonies were transferred to an MD96F plate (Falcon Cat. # 351172) containing CD CHO medium supplemented with 8 mM L-glutamine, and 1% antibiotic-antimycotic.

Clone genotyping

After two days, 50 uL cell suspension from each well was transferred to a MicroAmp Fast 96 well reaction plate (Thermo Cat. # 4346907), along with $5x10^5$ wildtype cells as a control. The plate was centrifuged at 1000 x g for 10 minutes, then the supernatant was removed via rapid inversion. 20 µL of QuickExtract DNA Extraction Solution (Epicentre Cat. # QE09050) (prewarmed

to 65°C) was added to each well and mixed via pipetting. The plate was then processed in the thermocycler (65°C for 15 minutes followed by 95°C for 5 minutes).

Amplicons were generated for each gene of interest per well using Phusion Hot Start II DNA Polymerase (Thermo Cat. # F549L) and verified to be present visually on a 2% agarose gel. Amplicons from each well had unique barcodes, allowing them to be pooled and purified using AMPure XP beads (Beckman Coulter Cat. # A63881) according to manufacturer's protocol, except using 80% ethanol for washing steps and 40 uL beads for 50 uL sample. Samples were indexed using the Nextera XT Index kit attached using 2x KAPA HiFi Hot Start Ready mix (Fisher Scientific Cat. # KK2602). AMPure XP beads were used to purify the resulting PCR products.

DNA concentrations were determined with the Qubit 2.0 Fluorometer and used to pool all indices to an equimolar value and diluted to a final concentration of 10 nM using 10mM Tris pH 8.5, 0.1% Tween 20. The average size of the final library was verified with the Bioanalyzer 2100. The amplicon library was then sequenced on an Illumina MiSeq.

Insertions and deletions were identified by comparison of expected vs. actual amplicon size. Clones with frameshift indels in all alleles of the Ldha gene were selected for expansion.

Batch culture

CHO-S™ cells (Gibco Cat. # A11557-01) and derivative clones were cultured in complete CHO-S medium: CD CHO medium supplemented with 8 mM L-glutamine and 2 mL/L of Anti-clumping agent (Gibco Cat. # 0010057AE). All cultures were maintained in an incubator at 37°C, 5% CO_2, 70-95% humidity and 25 mm throw, while shaking at 120 rpm. Cell growth and viability were monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec, Denmark) based on two fluorescent dyes, Acridine Orange and DAPI for the total and dead cell populations, respectively. Metabolite concentrations were measured using the BioProfile 400 (Nova Biomedical, USA).

Ldh enzymatic assay

Lactate dehydrogenase activity was measured using the Cytotoxicity Detection Kit PLUS (LDH) from Roche (Basel, Switzerland) according to the instructions of the manufacturer. In short, cell pellets ($1x10^6$) were resuspended in 250 µl 1% Triton X-100, 50 mM TRIS-HCl pH 7.5 and lysed by incubation for 30 minutes at 37°C. All samples were clarified by centrifugation for 5 minutes (14290xg) at room temperature. Supernatants were diluted 27-fold with PBS and transferred to a 96 well plate for measurement.

Western blotting

Cell pellets ($1x10^6$ cells) were resuspended in 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1 pill cOmplete/50mL (Roche) and lysed by incubation for 30 minutes on ice. Lysates were clarified via centrifugation at 14290 x g for 10 minutes at room temperature. After separation by SDS-PAGE, proteins were transferred onto PVDF membranes and probed with antibodies against Vinculin (V9131, 1:1000, Sigma, St. Louis, MO, USA) and Ldha (#2012, 1:1000, Cell Signaling Technology, Danvers, MA, USA). Proteins of interest were detected with Cy5 and Cy3 conjugated secondary antibodies (29-0382-78 and 29-0382-75, 1:2500, Amersham, Little Chalfont, UK) and visualized on an AI600 imager (Amersham). Quantification was done using ImageQuant TL 1D v8.1. For nonfluorescent blots, samples were treated as above except a nitrocellulose membrane was used, the antibody against Vinculin was diluted 1:2000, HRP-conjugated donkey anti-rabbit (ab6802, 1:5000, Abcam, Cambridge, UK) or goat anti-mouse (P0447, 1:5000, Dako) were used as secondary antibodies, and detection was with the Amersham ECL detection system.

Oxygen uptake

Cells were grown in CD OptiCHO medium (Gibco Cat. # 12681011) containing 4 mM glutamine, 50 ppm Antifoam C (Sigma A8011), 1.05 mL/L Anti-clumping agent, and 10.5 mL/L

antibiotic-antimycotic. Cultures were grown in DASGIP bioreactors (see Fed-batch culture for set-points and control parameters) without any additional nutrient feeding, except glutamine added on demand.

Oxygen uptake rate (OUR) was determined using the dynamic method[145]. Briefly, every 6-8 hours dissolved oxygen (DO) levels were increased to 60%. Afterwards, DO control and gas supply via sparger were paused while an N_2 flow (0.07 vvm) was introduced into the bioreactor headspace to obtain an oxygen-free gas phase and avoid oxygen transport back to the culture medium. OUR calculations were performed by monitoring the DO profile between 60 and 30% to quantify the respiratory activity of the cells while accounting for oxygen desorption from the liquid to gas phase using an experimentally determined desorption constant, K_{des} (Equation 1). DO control resumed when levels dropped below 30%. Measurements were taken from 2 bioreactors containing wildtype cells and 2 containing Warburg-null clone 0-F5.

$$OUR = \frac{C_{O_2}(t_0) - C_{O_2}(t_f)}{t_f - t_0} + \frac{\int_{t_0}^{t_f} (-K_{des} * C_{O_2}(t))\, dt}{t_f - t_0}$$

For every time interval containing at least 3 data points per bioreactor between t=14 and 84 hr (i.e., during exponential growth), a maximum likelihood fit of a linear mixed-effect model was generated, describing the association between volumetric oxygen uptake rate and viable cell density with a cell type-specific intercept (wildtype or knockout) as fixed effects with random intercepts for each bioreactor to account for potential differences in bioreactor-specific oxygen dissipation/leakage. This additive model was extended with an interaction term (cell type:VCD) to describe the cell type-specific effect on oxygen uptake rate. A likelihood ratio test between the additive and interaction models determined if the inclusion of the cell type-specific uptake term had a significant effect on the best-fit slope (i.e. per-cell oxygen uptake rate) at an FDR of 0.05 as determined by the Benjamini–Hochberg method.

Fed-batch culture

Cells were grown from a seed density of $2.5x10^5$ cells/mL in DASGIP bioreactors with a starting volume of 270 mL at a temperature of 37°C, agitated at 200 rpm using pitched blade impellers. The pH was maintained continuously at 7.10±0.02 with 1 M sodium bicarbonate or CO_2. Dissolved oxygen was maintained at 40% using pure oxygen, air, or a mixture, as needed. A 2222 mM glucose solution and a 200 mM glutamine solution were used to control glucose and glutamine levels at the desired levels via a once-daily feeding based on the measured concentration, doubling time, and calculated consumption rates. Complex feed addition was started when cell concentration was approximately $1.4\text{-}2.2x10^6$ cells/mL and continued for 4 days, feeding increasing volumes each successive day. Cultures were sampled daily for cell growth and viability using the NucleoCounter NC-200 Cell Counter, metabolite concentrations using the BioProfile 400, and rituximab concentration using the OctetRED96 for protein producing cultures. Process variations for the different experiments are detailed below:

Clone 0-B6

Cells were maintained in CD CHO medium supplemented with 8 mM L-glutamine, 1% antibiotic-antimycotic (Life Technologies #15240062), and 1 mL/L anti-clumping agent. 0.04% Pluronic F-68 (Life Technologies #24040032) was added to each reactor prior to inoculation and supplemented as needed over the course of culture. Glucose was maintained between approximately 24-40 mM while additional glutamine was not supplemented over the course of culture. Efficient Feed B (Thermo Fisher #A1024001) was added at the following days/volumes:

Day 3: 4 mL

Day 4: 8 mL

Day 5: 12 mL

Day 6: 16 mL

Clone 0-F5

Cells were maintained in CD CHO medium supplemented with 8 mM L-glutamine, 1% antibiotic-antimycotic, and 1 mL/L anti-clumping agent. Antifoam C was added as needed over the course of culture. Glucose was maintained between 10-24 mM. Glutamine was maintained between 1-3 mM. Efficient Feed B was added at the following days/volumes:

Day 3: 4 mL

Day 4: 8 mL

Day 5: 12 mL

Day 6: 16 mL

SUPPLEMENTARY FIGURES & TABLES

Table 3.S1: Source for gene expression data in Figure 3.1.

Cell Line	Description	Reference
S (n=40)	Wildtype, 40 samples taken between mid-late exponential growth (72-180 hours of cultivation) from 4 biological replicates	Hefzi et. al.[58]
K1 (n=2)	Wildtype, single timepoint	Hefzi et. al.[58]
K1 (n=2)	Wildtype, single timepoint	van Wijk et. al.[146]
K1 (n=2)	Wildtype, exponential and stationary	Lund et. al.[147]
DG44 (n=6)	Wildtype, exponential and stationary phases with and without MEM NEAA in media. Two samples following NaBu treatment (24hr and 48hr post-treatment)	Lund et. al.[147]
DXB11 (n=2)	Wildtype and Mre11 knockdown cells	Kondratova et. al.[148]

Table 3.S2: Oligos used for cloning gRNAs and primers for MiSeq analysis of genes targeted in this study.

Sequences for generating gRNAs and primers used for engineering and amplicon sequencing. Pdk mutations introduce a frameshift upstream of the catalytic kinase domain, thus disrupting function[149]. Ldha mutations disrupt residues critical for substrate recognition and the Rossman fold domain necessary for NAD^+/NADH binding[150].

	Pdk4	Pdk3	Pdk2	Pdk1	Ldha
Forward gRNA	GGAAAGGACGAAACACCTGGCGCTGCGCATCACGAAGGTTTTAGAGCTAGAAAT	GGAAAGGACGAAACACCATTGTGTTCGCCAGTCGTACGTTTTAGAGCTAGAAAT	GGAAAGGACGAAACACCGTCATTGTGCCGGTTCCGGAGTTTTAGAGCTAGAAAT	GGAAAGGACGAAACACCACTTCGACTACATCGCAGTTGTTTTAGAGCTAGAAAT	GGAAAGGACGAAACACCGCTGGGCACTGATACCGACAGTTTTAGAGCTAGAAAT
Reverse gRNA	CTAAAACCTTCGTGATGCGCAGCGCCAGGTGTTTCGTCCTTTCCACAAGATAT	CTAAAACGTACGACTGGCGAACACAATGGTGTTTCGTCCTTTCCACAAGATAT	CTAAAACTCCGGAACCGGCACAATGACGGTGTTTCGTCCTTTCCACAAGATAT	CTAAAACAACTGCGATGTAGTCGAAGTGGTGTTTCGTCCTTTCCACAAGATAT	CTAAAACTGTCGGTATCAGTGCCCAGCGGTGTTTCGTCCTTTCCACAAGATAT
gRNA Sequence	TGGCGCTGCGCATCACGAAGCGG	ATTGTGTTCGCCAGTCGTACAGG	GTCATTGTGCCGGTTCCGGATGG	ACTTCGACTACATCGCAGTTTGG	GCTGGGCACTGATACCGACAAGG
Forward Primer	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCAGGTCTCCTCTCACCAA	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTTTCTCTGTGTTAATCTTAGGGA	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGCTGGGGAGGATAGAGAA	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGGTGTTAGGTAAATGCGC	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGCTCTCCTGTGGAAACATTG
Reverse Primer	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACCCTGAGTCCACAGCG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTGAAGGGCGGTTCAACAA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATGGAGATCCGGCTGAGGTA	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCTCTCATGACACAAGCAGT	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTTTCCATGCTGCCAATCACG
Protein Length	412	415	407	358	332
Cut Site (aa)	3	53	117	142	216

Table 3.S3: Genotypes of isolated Warburg-null clones.
Black box indicates wildtype alleles for the gene, X a full knockout, ½ a partial knockout. Shading of the cell indicates what portion of the observed indels are in-frame, white: all indels frameshift, red: all indels in-frame, yellow: some indels frameshift, some in-frame.

Clone	Parent	Pdk1	Pdk2	Pdk3	Pdk4	Ldha
0-B6	WT	X	X	X	X	X
0-F4	WT	X	X	X	X	X
0-F5	WT	X	X	X	X	X
1-C5	WT	X	X		X	X
1-D6	WT		X		X	X
2-D5	WT	X		½	X	X
2-G1	WT	X			X	X
3-C4	WT		X			X
3-D8	WT		X	X	X	X
3-F1	WT		X		X	X
3-G12	WT		X	X	X	X
3-H9	WT		X		X	X
4-B6	WT		X			X
4-B7	WT	X	X		X	X
4-C8	WT	X	X		X	X
4-C12	WT	½	X		½	X
4-H9	WT	½	X			X

Table 3.S4: Detailed genotypes of isolated Warburg-null clones.
Indel sizes in targeted genes for all characterized clones.

Clone	Pdk1	Pdk2	Pdk3	Pdk4	Ldha
0-B6	1	1	-8/4	-2	-8/-1
0-F4	1	1	1	1/-1	-1
0-F5	1	1	101	1	-1
1-C5	-20/1	1	-	-49/-1/1	-4/-1
1-D6	-	4	-	-17/-5	-4/1
2-D5	1	-	-24	1	-2/1
2-G1	1	-	-	1/4	-10
3-C4	-	-10/1	-15	-	-4/4
3-D8	-	-16/-10	-8	-24/-18	-11/-2
3-F1	-	1	-39	-22/-21/1	1
3-G12	-	1	-1	1	-31/-19
3-H9	-	1	-	-21	-2/2
4-B6	-	1	-	-	-4
4-B7	-1/1	-8/1	-	-30/-17	-88/-1
4-C8	1/12	-16/1	-	-24/-12	-7/-2
4-C12	1	1	-	19	-4/-1
4-H9	1	-11	-	-	-22
P-H6	-	-2/-1/1	1	-13/1	-2/-1
P-G8	-	-3/1	-	-25/-19/-2/8	-73/-10/-1

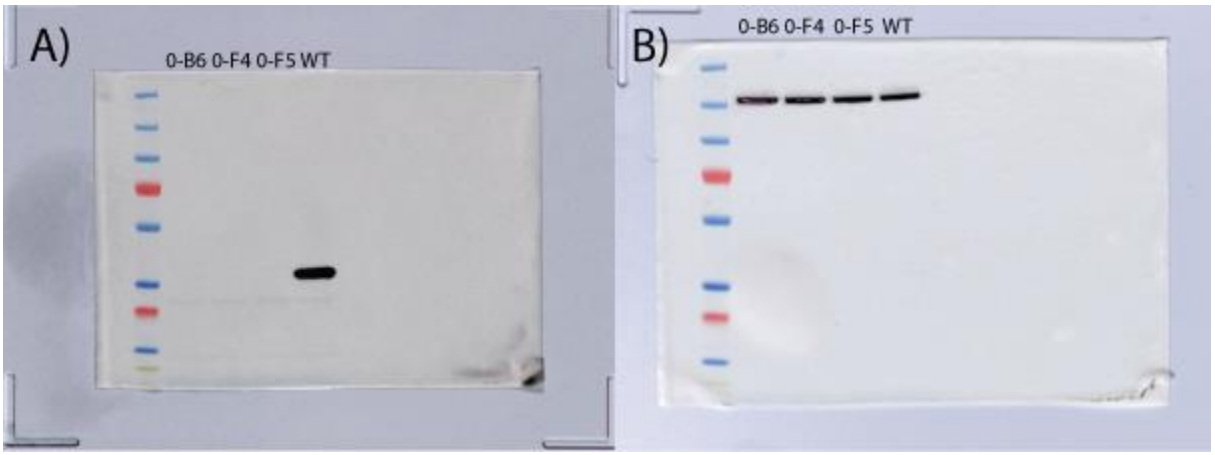


Figure 3.S1: Western blots for Pdk1-4 + Ldha knockout clones.
Blots for Ldha **(A)** and Vinculin **(B)** were performed using chemiluminescent detection.

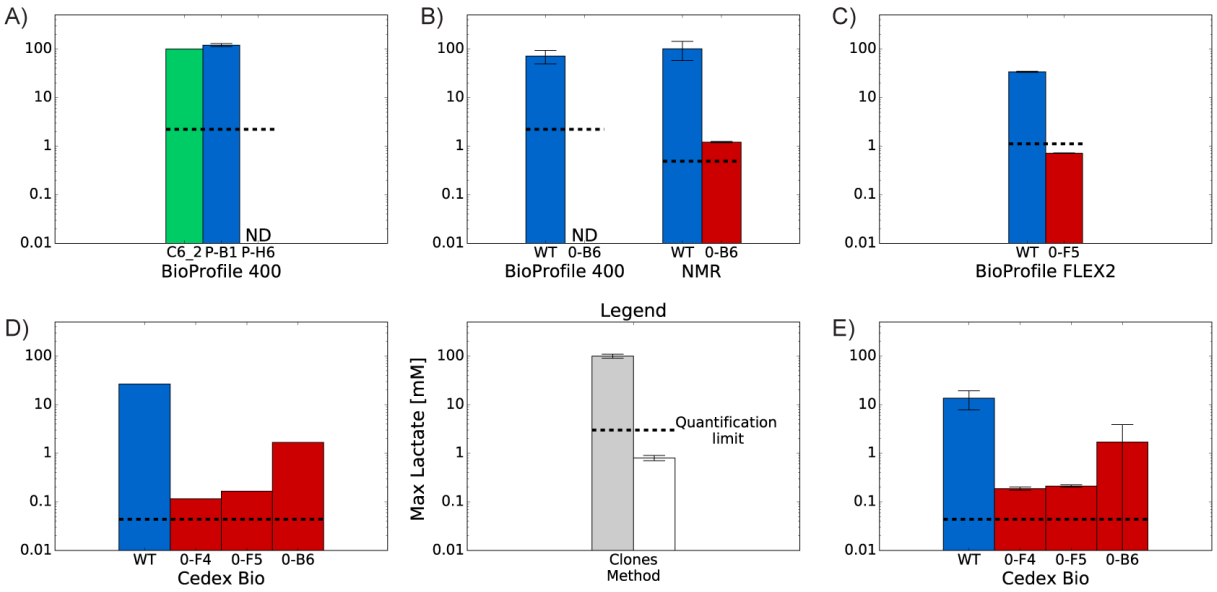


Figure 3.S2: Maximum lactate concentrations for cultures in this study.
Maximum lactate concentrations from fed-batch of **(A)** rituximab-producing cells (see also Figure 4.4), **(B)** non-producing clone 0-B6 (see also Figure 3.3), **(C)** non-producing clone 0-F5 (see also Figure 3.S5), and batch culture of **(D)** non-producing clones 0-B6, 0-F4, and 0-F5 (n=1 for each clone), or **(E)** rituximab producing pools derived from 0-B6, 0-F4, and 0-F5 (see also Figure 4.1). Method of measuring lactate is shown below each plot with the quantification limit denoted by a horizontal dashed bar. ND: not detected (reported lactate 0 mM). Quantification limits for the methods are as follows: BioProfile 400-2.22 mM, BioProfile FLEX2-1.11 mM, NMR-0.49 mM, Cedex Bio-0.044 mM. Data shown as mean ± standard deviation.

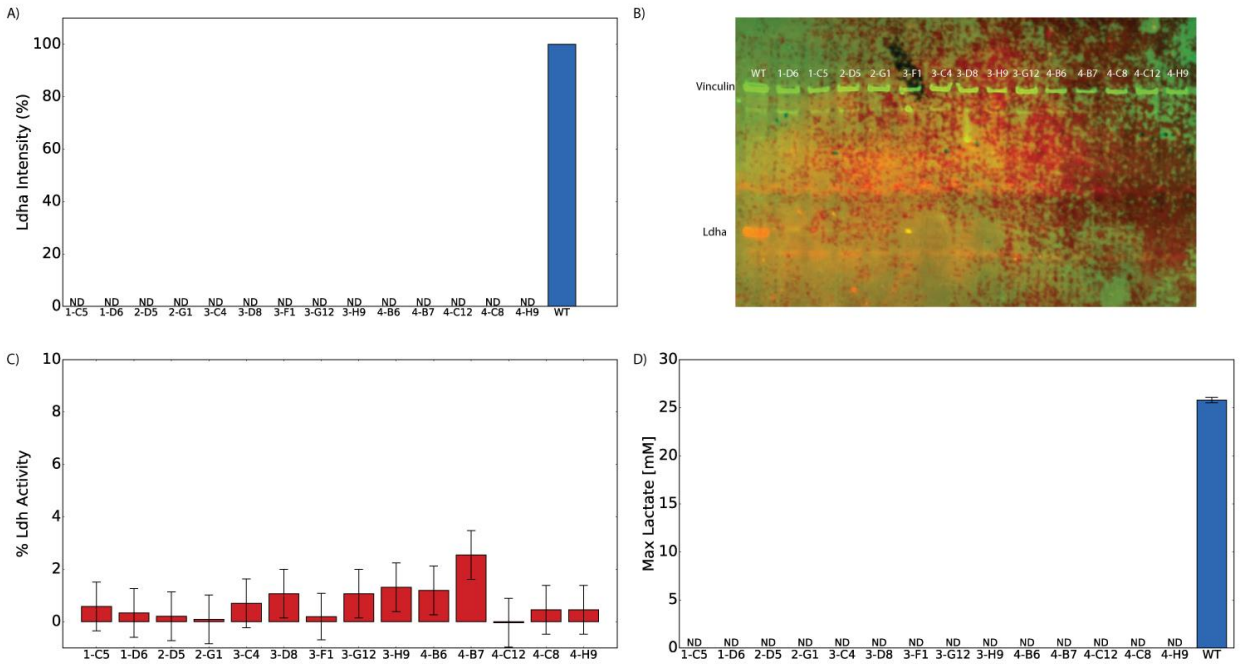


Figure 3.S3: Knockout of sub-combinations of Pdks permits Ldha knockout and elimination of the Warburg effect.

(A,B) Simultaneously targeting of Ldha and Pdk genes using CRISPR/Cas9, leads to no detectable Ldha expression in any clones, as verified by fluorescent Western blot. **(C)** Enzymatic assay verifies a loss of lactate dehydrogenase activity in knockout lines (n=1 per clone). **(D)** Measurement of maximum lactate levels during growth in batch culture shows that lactate cannot be detected (n=2 per clone). Data shown as mean ± error from variance in background (calculated via propagation of error) for **(C)** and mean ± standard deviation for **(D)**.

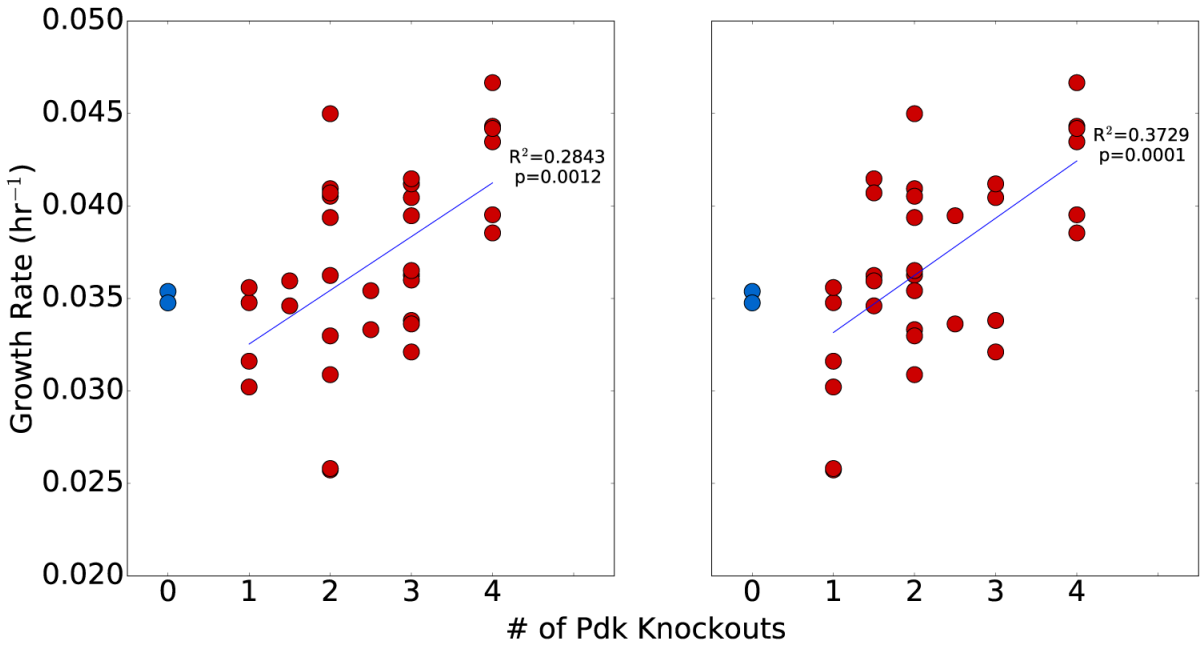
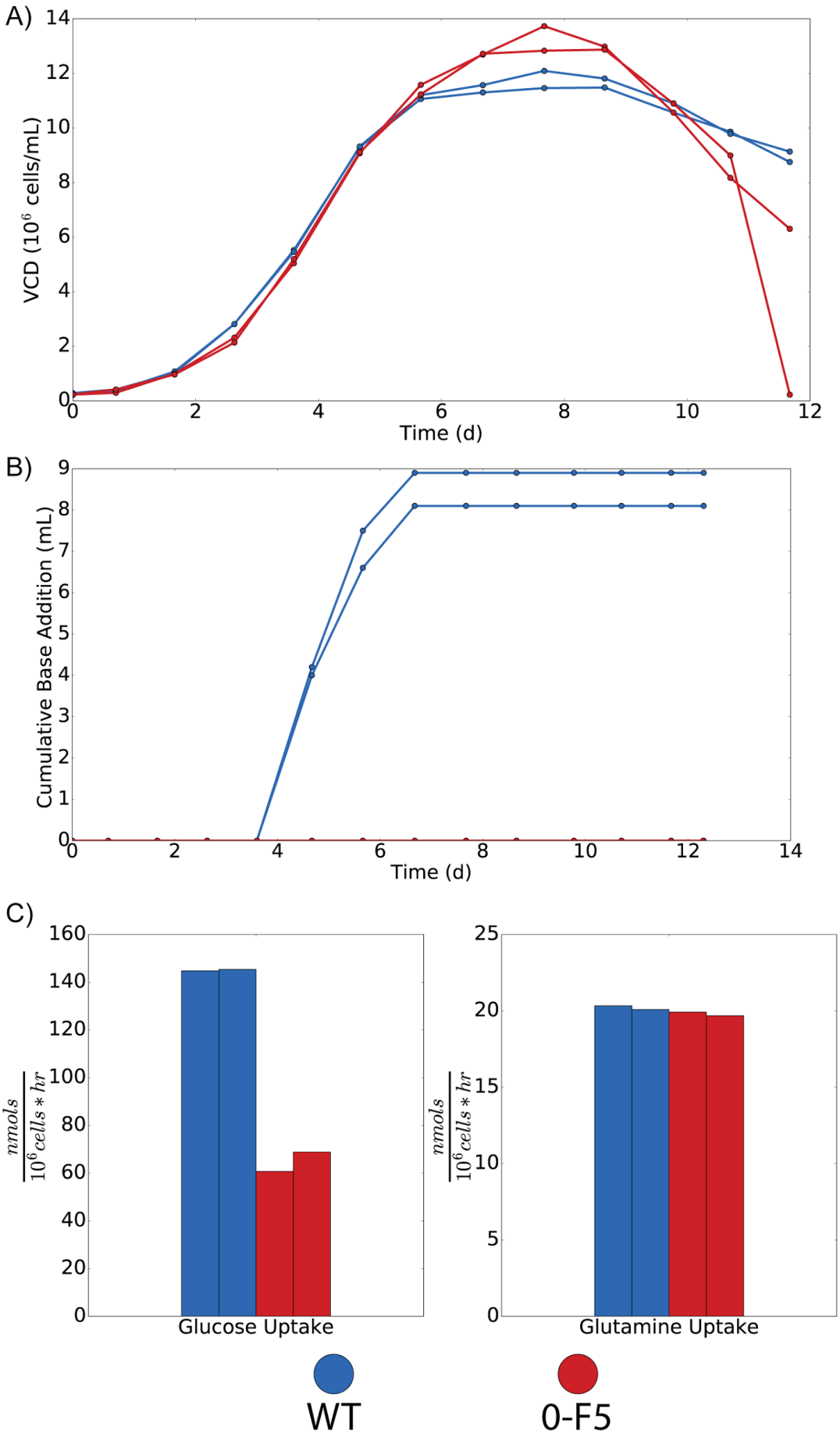


Figure 3.S4: Growth rate of Warburg-null clones.

Warburg-null cell lines show improved growth rate with increasing number of Pdk knockouts. This effect is observed whether in-frame mutations are counted as knockouts **(A)** or wildtype **(B)**. The growth of wildtype CHO-S is shown in blue. Correlation computed across knockout cell lines only.

Figure 3.S5: A fed-batch experiment on a second Warburg-null clone, and demonstrated similar results.

Clone 0-F5 (red, n=2) and wildtype (blue, n=2) were grown in fed-batch culture. **(A)** The knockout clone exhibits a growth rate similar to the WT but remains in exponential phase for a prolonged period of time. **(B)** As a result of not secreting lactate, 0-F5 require less addition of base (1 M $NaHCO_3$) over the course of culture to maintain pH at the appropriate levels. **(C)** During exponential growth (calculated between day 1-5), significantly less glucose is consumed compared to the WT but similar amounts of glutamine.


A)
VCD (10^6 cells/mL)
Time (d)
B)
Cumulative Base Addition (mL)
Time (d)
C)
$\frac{nmols}{10^6 cells * hr}$
Glucose Uptake
Glutamine Uptake
WT
0-F5

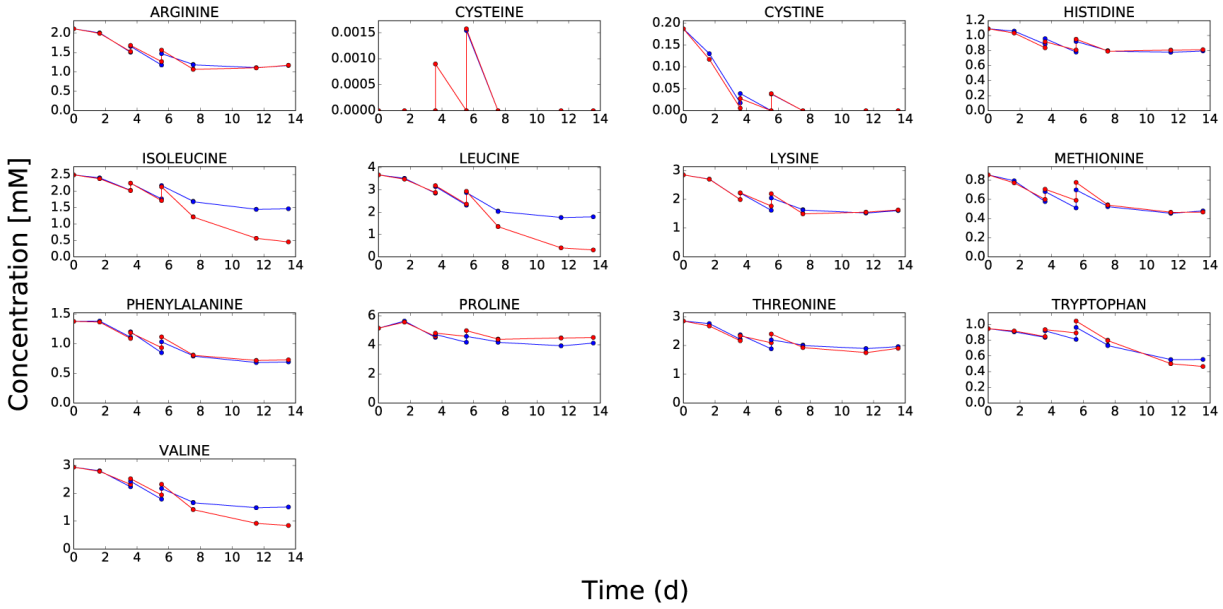


Figure 3.S6: Time-course profiles of essential amino acids in fed-batch culture.
Amino acid concentrations from cultures shown in Figure 3.3 (0-B6: red, WT: blue). Cysteine and cystine are depleted at day 8, at which point cells begin to die. Data shown as average concentrations before and after feed addition.

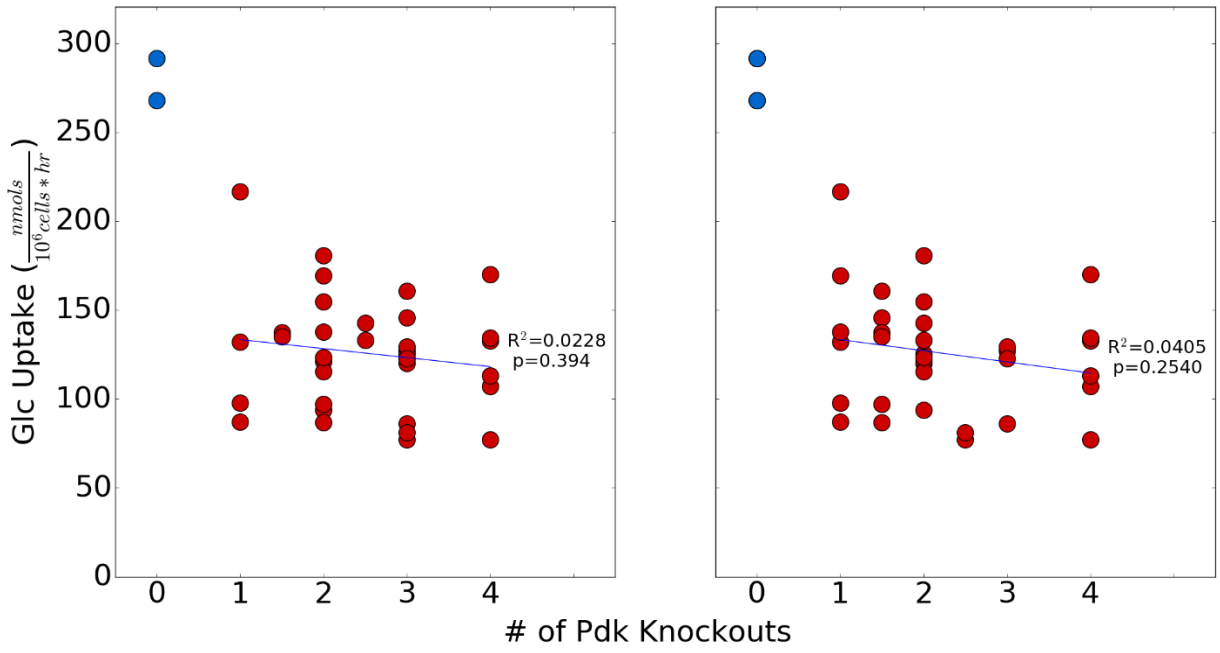


Figure 3.S7: Glucose uptake rates of Warburg-null clones.
Warburg-null cell lines show decreased glucose uptake rate, independent of number of Pdk knockouts. This is observed whether in-frame mutations are counted as knockouts **(A)** or wildtype **(B)**. The glucose uptake rate of wildtype CHO-S is shown in blue. Correlation computed across knockout cell lines only.

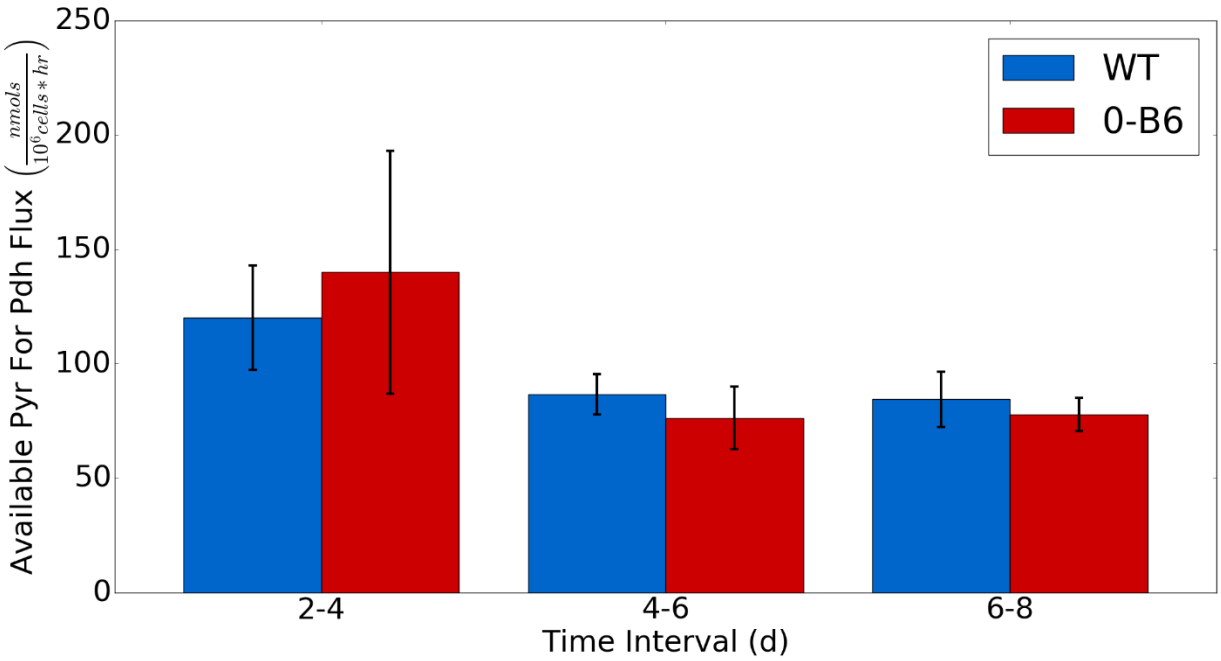


Figure 3.S8: Glucose oxidation is unchanged in Warburg-null lines.

Estimated Pdh flux (available carbon after accounting for glucose, lactate, pyruvate, and alanine uptake/secretion) is largely unchanged for 0-B6 (n=5) compared to WT (n=3) over the course of exponential growth in fed-batch culture. Data shown as mean ± standard deviation.

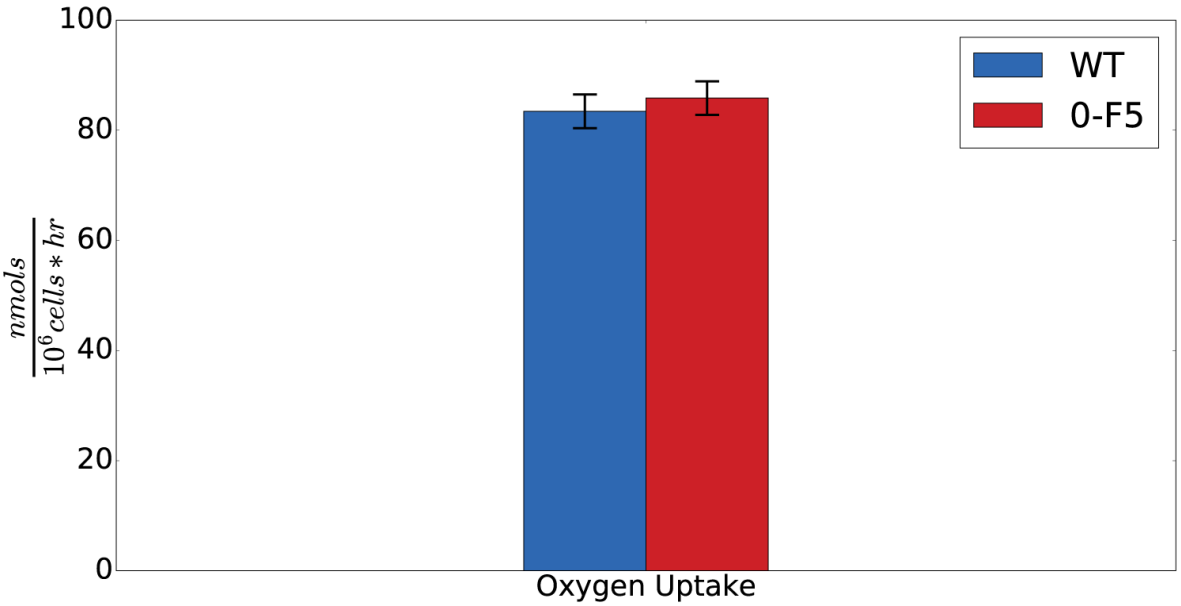


Figure 3.S9: Oxygen uptake is unchanged in Warburg-null lines.
In batch culture (n=2 wildtype, n=2 0-F5), the oxygen uptake rate is unchanged during the exponential growth phase (t=14-84 hr). All pairwise comparisons of oxygen uptakes over time intervals exceeding ~12 hr (i.e. containing at least 3 data points per bioreactor) were not statistically significant, as calculated by a likelihood ratio test comparing linear mixed-effect models including or excluding a cell type-specific interaction term (see Methods). This finding was robust against varying the number of data points needed to build a model for an interval (e.g. the same analysis for time intervals exceeding ~29 hours or 6 data points per bioreactor gave qualitatively similar results). Data shown is the estimate of the slope (cell specific oxygen uptake rate) ± standard error of the estimate.

Chapter 3 contains material from Hefzi, H., Noh, S.M., Monge, I.M., Decker, M., Arnsdorf, J., Kol, S., Pristovšek, N., Hansen, A.H., Bjorn, S.P., Brøndum, K.K., Javidi, E.M., Jensen, K.L., Kallehauge, T.B., Ley, D., Ménard, P., Petersen, H.M., Sukhova, Z., Voldborg, B.G., Nielsen, L.K., Lee, G.M., Kildegaard, H.F., Lewis, N.E. Multiplex genome editing eliminates the Warburg Effect without affecting oxidative metabolism or growth rate. (Submitted). I was the primary author, while the co-authors provided support in the research underlying the study.

Chapter 4: Characterization of Warburg-null cells for biopharmaceutical production

INTRODUCTION

Non-disruptive advancements in the biopharmaceutical industry must aim to improve one or more of the following areas—without adversely affecting the others:

1. Time to generate a production cell line
2. Final product titer
3. Product quality

Simple transfection of a biotherapeutic protein-coding gene into CHO cells is unlikely to result in a useful production cell line. This is due to both the inherent nature of random integration to result in low expression of the gene of interest as well as the origin of the cell line—ovarian tissue cells are not traditionally thought of as “professional secretors” and likely are not natively prepared to synthesize, process, and secrete high quantities of protein. Accordingly, the standard industrial process for generating a high producing cell line has to address both these issues. This process can take more than a year from start to finish and needs to be repeated for each new product. It is thus critical that any engineering strategies do not harm cellular viability for use in this pipeline, described briefly here. To achieve high expression, coamplification of the gene of interest alongside a selectable marker is traditionally used. For example, dihydrofolate reductase (*Dhfr*) catalyzes the first step in regeneration of cofactors needed for *de novo* purine and pyrimidine biosynthesis. When grown in media lacking hypoxanthine and thymidine, *Dhfr* deficient cells will not survive, thus introduction of the gene of interest on the same construct as *Dhfr* enables simple selection of cells expressing both genes. Treatment of surviving cells with a Dhfr inhibitor such as methotrexate (MTX) selects for cells with high expression of *Dhfr*, often those that have integrated or amplified multiple copies of *Dhfr* and the gene of interest. In this way, cells with the highest expression of the desired product can be isolated. In a similar fashion, glutamine synthetase (*Gs*)

can be selected for in glutamine-free medium and inhibited by methionine sulfoximine (MSX). Following this process, a large number (hundreds[151], thousands[152], or more[153]) of clones are single-cell sorted, expanded, and assessed for ability to produce high quantities of product. Finally, a subset of the best performing clones are selected for long-term stability studies and further optimization for use in the final production phase. We show that Warburg-null cell lines remain viable for use in standard industrial processes for generating high-producers.

The final amount of product generated in an industrial bioprocess is a function of the integral of the viable cell density over the course of culture (i.e., process intensity) as well as the specific productivity, traditionally described in picograms per cell per day (pcd) of the cell line (i.e., process efficiency). A cell line with low pcd can still achieve high product titer if it is able to maintain high cell density for a prolonged period. Conversely, cells with high pcd can reach high titers in shorter processes. Historically, improvements in final titer have been driven by improved process intensity rather than process efficiency. For example, a prototypical process in the late 1980s may have occurred over 7 days in batch culture reaching cell densities of $<5 \times 10^6$ cells/mL, while more modern processes occur in fed-batch processes over 2 weeks reaching cell densities of $>30 \times 10^6$ cells/mL. Final titers from these processes have improved from ~100 mg/L to upward of 5 g/L[7,154,155]. This is not to discount the importance of high specific productivity—indeed, standard workflows have isolated clones producing ~90 pcd[155]—however, following integration, amplification, and isolation of clones with high efficiency, further engineering of production cell lines has proven difficult due to lack of tools (now somewhat ameliorated due to CRISPR) and readily identifiable gene/pathway targets. It has been comparatively simpler to optimize the bioprocess, e.g. the medium composition, feeding strategy, and other process parameters, to increase process intensity as a path to increased final titers. We show that Warburg-null cells are amenable to increased process intensification and retain high product titers even in an unoptimized bioprocess.

A final important consideration is the product quality at the end of a process. Indeed, the primary impetus for using mammalian cells such as CHO (rather than *Escherichia coli* or yeast) is to ensure that critical post-translational modifications—such as glycoslyation—are present. Identifying self vs. non-self is a critical ability of multicellular organisms, and recognition of 'foreign' entities is often mediated by recognition of non-human glycans[156]. Consistent with this, glycosylation of biotherapeutics can affects product efficacy, clearance time, and immunogenic response[157]. Thus, it is essential that engineering strategies do not adversely impact the glycoprofile of a product. We show that rituximab produced by Warburg-null cells actually display an improved glycoprofile compared to the parental line.

RESULTS

Warburg-null cells exhibit positive traits for protein production

The elimination of the Warburg effect addresses an outstanding challenge in the mammalian bioprocessing field. Our observations of increased cell density with no negative effects on growth rate highlight a potential value of implementing this strategy in mammalian cells used for protein production. However, the changes in ATP production and other potential changes to the cells raise questions regarding whether a Warburg-null phenotype could negatively impact their utility for protein production. Therefore, we investigated how this impacts the process of cell line development, process intensification (and final titers), as well as product quality.

Warburg-null cells are amenable to transfection and transgene amplification

For decades, the development of highly productive cell lines has involved transgene integration, selection, and amplification using a selectable marker. These result in polyclonal pools producing the transgene, from which stable clones are identified and used for production[155]. Ideally, engineered host cell lines would be compatible with existing workflows, but recent work showed that shRNA-mediated downregulation of Ldha resulted in cells unable to survive the

established selection and amplification process[158]. These results raise concerns that the knockout of Ldha and Pdk-mediated Pdh regulation may result in a similarly unsuitable host cell line. Fortunately, when we subjected the 0-B6, 0-F4, and 0-F5 clones to an identical selection and amplification process, we successfully generated polyclonal pools of mAb-producing cells (Figure 4.1A). These pools had similar growth rates as pools derived from WT cells (Figure 4.1B), and comparable or even improved product titers (Figure 4.1C). Thus, Warburg-null cells are compatible with established cell line development protocols and show suitability as host cell lines.

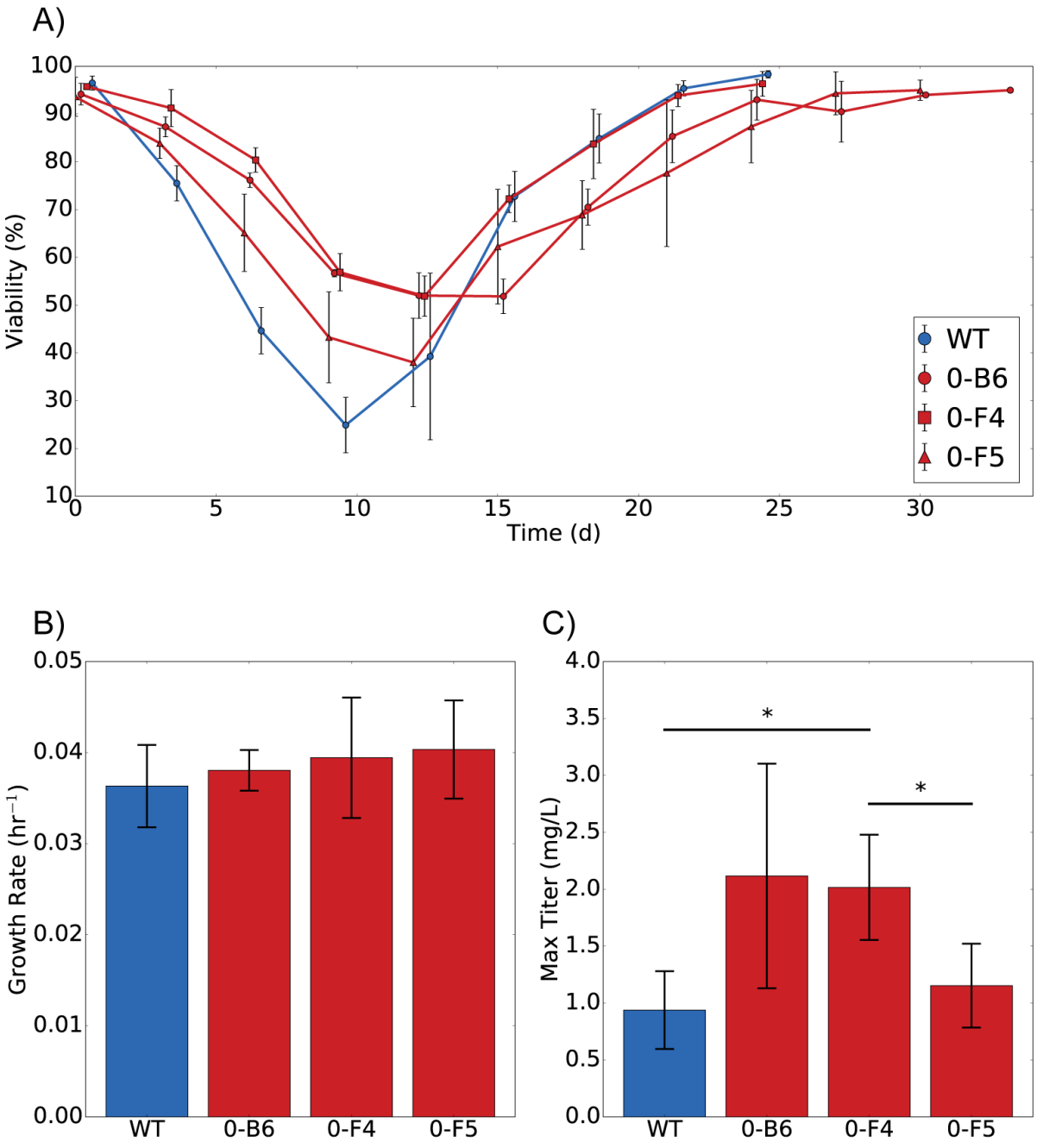
Warburg-null cells exhibit improved behavior in process intensification

Preliminary experiments showed that Warburg-null clones showed an extended growth phase compared to the wildtype (Figure 3.3, Figure 3.S5). There, wildtype clones entered stationary phase prior to depletion of essential nutrients while Warburg-null clones only ceased growth after cysteine and cystine were depleted (Figure 3.S6). We were thus curious to see if optimization of the bioprocess, with respect to medium and feed strategies, could further improve process intensity.

We found that changing the basal medium to OptiCHO and feeding with Cell Boost 7a and 7b for an extended period of time (see Methods for details) was able to again substantially increase proliferation (Figure 4.2). Interestingly, both wildtype and Warburg-null cells ceased growth and began to die prior to the end of the feeding phase of culture, indicating that additional optimization may increase process even more.

Figure 4.1: Eliminating the Warburg effect does not impact polyclonal pool generation.

(A) Warburg-null clones show comparable recovery times to wildtype CHO-S cells when faced with a standard workflow for stable integration of a transgene. The resultant pools have indistinguishable growth **(B)** and protein production **(C)** characteristics. Thus, Ldha knockout clones are suitable for use as starting cell lines for biotherapeutic protein production. * indicates $p<0.05$ as determined by a two-sample two-tailed Welch's t-test. Data shown as mean ± standard deviation. In all panels, n=3 for 0-B6 and n=4 for other lines (see Figure 4.S1). In **(A)**, measurements taken every 3rd day, slight offsets to displayed times for visual clarity.


A)
Viability (%)
Time (d)
WT
0-B6
0-F4
0-F5
B)
Growth Rate (hr^{-1})
WT
0-B6
0-F4
0-F5
C)
Max Titer (mg/L)
*
*
WT
0-B6
0-F4
0-F5

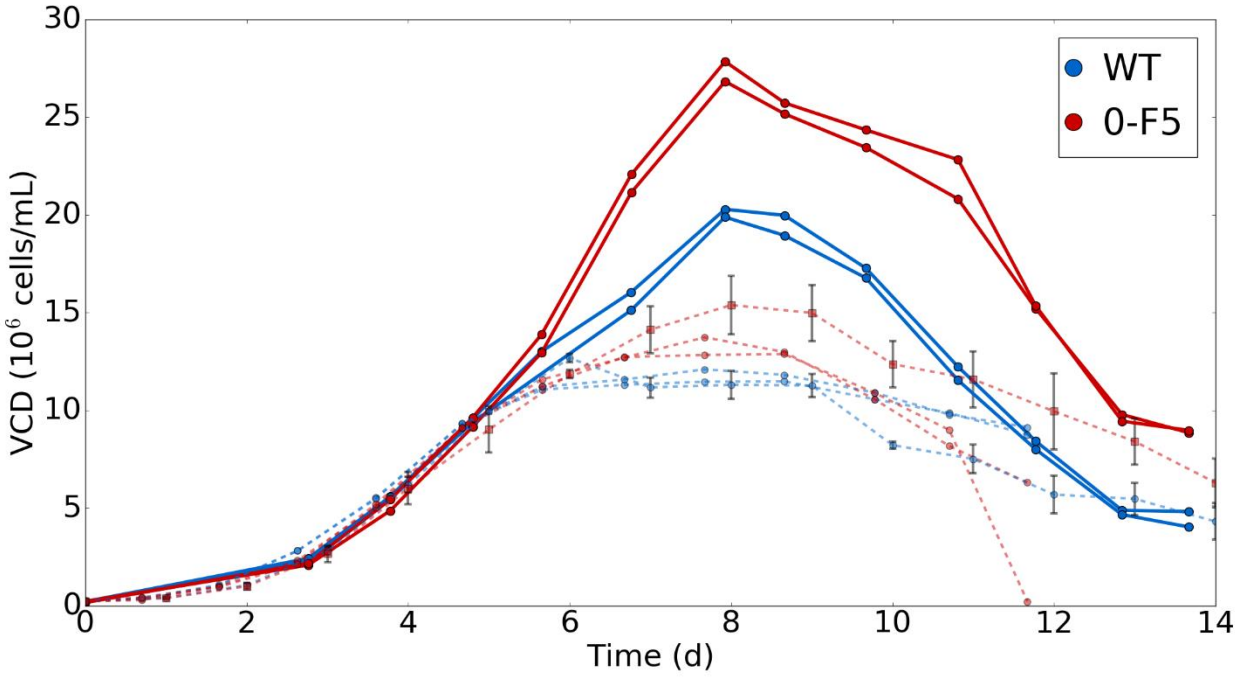


Figure 4.2: Process intensification around the Warburg-null phenotype.
Changing the bioprocess strategy from the ones used in Figure 3.3 and Figure 3.S5 (shown in dotted lines) to prolong feeding and growth results in increased cell concentration during fed-batch culture.

The Ldha deletion can be introduced into mAb-producing lines

Generating high producing cell lines is a time-consuming process, thus it could be advantageous to introduce the Warburg-null phenotype into existing CHO cell lines that are producing biotherapeutics. To test the effect of eliminating the Warburg phenotype on mAb production, we targeted all Pdks and Ldha (Table 3.S2) in a rituximab-producing CHO-S line[159] (C6_2) and isolated two clones (P-H6 and P-G8) with Ldha deleted (Tables 4.S1/S2). In parallel, as controls for the cell engineering process, we isolated five clones (P-B1,-B2,-B3,-B5, and -B6; collectively referred to here as “mock”) derived from transfection with only Cas9.

As clonal variation is a significant concern in CHO cells due to inherent instability of the line[160,161], we first grew all clones in batch culture to identify the best performers for further characterization. Surprisingly, both Warburg-null clones (P-H6 and P-G8) showed significantly worse growth profiles than all mock clones and the parental line (Figure 4.3A). This was likely due

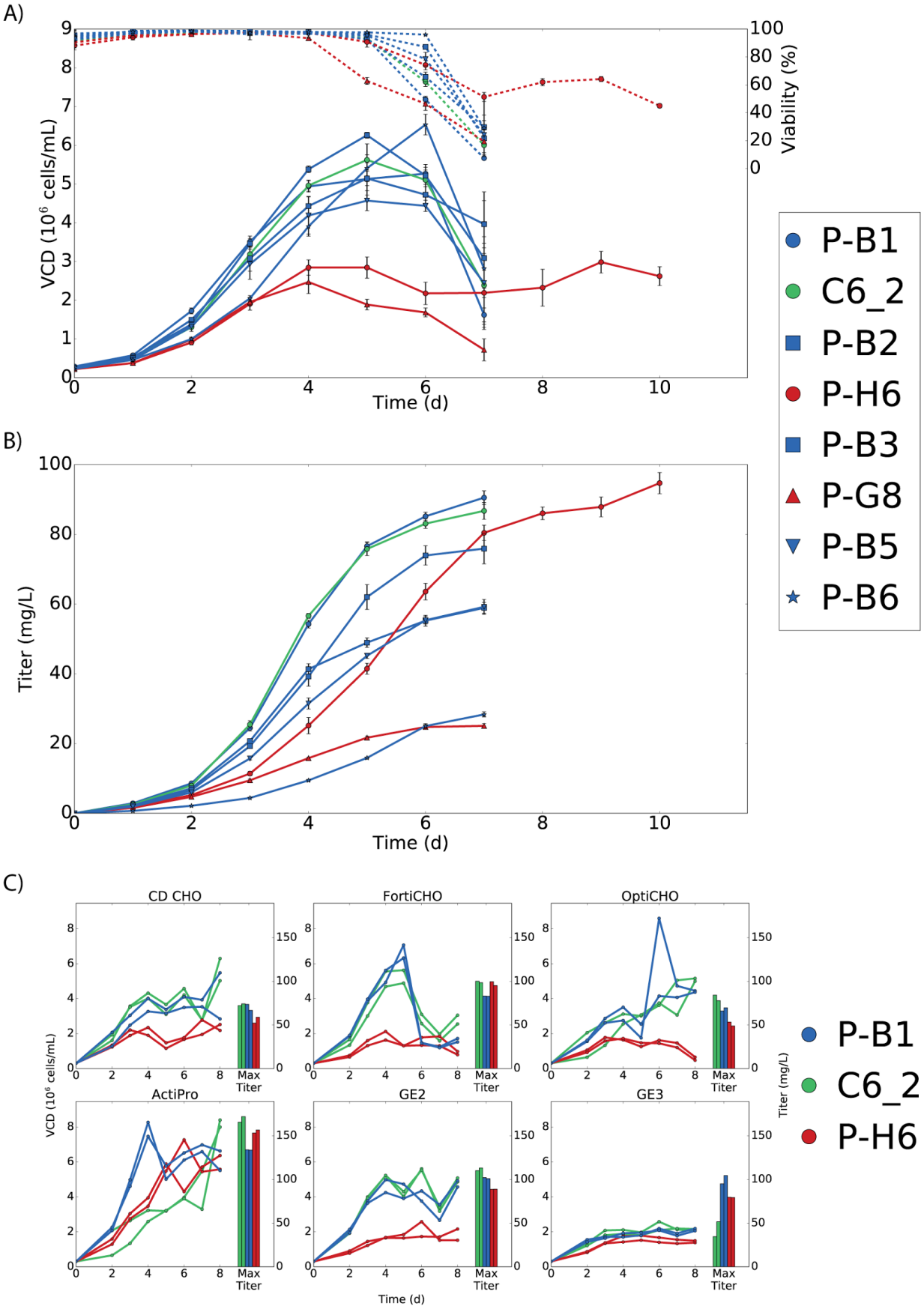
to a high degree of cell aggregation and deposition at the liquid-flask interface for P-H6 and P-G8 (thus ‘removing’ cells from suspension). Although growth profiles were similar for all mock lines, investigation of protein titer revealed stark differences, with the best clone (P-B1) producing more than three times the rituximab of the worst clone (P-H6) while behaving comparably to the parental C6_2 clone (Figure 4.3B).

Knowing that medium can have a significant effect on cell behavior, we took C6_2, P-B1, and P-H6 (the better Warburg-null clone) and carried out a small medium screen (6 different basal mediums in tested in batch culture). ActiPro medium proved to be the consensus best choice for all three lines, both in terms of cell growth and final rituximab titer (Figure 4.3C). We note that the aggregation and deposition observed for P-H6 in CD CHO medium did not occur in ActiPro.

The same clones were then further characterized in fed-batch. Consistent with the non-producing Warburg-null clones (Figure 3.3A, Figure 4.S5A), P-H6 exhibited improved growth behavior compared to both the parental and mock lines (Figure 4.4A). Protein titer was maintained at similar levels to the parental line and significantly increased compared to the control clone (Figure 4.4B). Thus, the Warburg-null genotype can be engineered into biotherapeutic-producing cells. Furthermore, since the medium used here was designed for lactate-producing cells, it is anticipated that further medium optimization would achieve superior titers in Warburg-null cells.

Figure 4.3: Batch-culture characterization and media optimization of clones derived from rituximab-producing line C6_2, following single-cell sorting.

Following single cell sorting and identification of Warburg-null clones, we grew the parental line, Warburg-null clones, and control clones in batch culture for preliminary characterization of the impact on growth and protein production. Significant variation is observed in both **(A)** growth profile and **(B)** rituximab titer both between engineered strains (i.e., KO vs. Mock vs. Parental) and within strains (e.g., P-B1 vs P-B6) when grown in batch culture in CD CHO media (n=3 for all lines). Data shown as mean ± standard deviation. Since they had the highest product titer, we selected P-H6 and P-B1 for further characterization in fed-batch alongside the parental C6_2 line following a media screen to identify optimal media for growth and protein production **(C)**. Media screening was carried out following the same protocol varying only the media. We tested CD CHO, FortiCHO (Gibco Cat. # A1148301), OptiCHO (Gibco Cat. # 12681011), ActiPRO, and 2 GE formulations, labeled here as GE2 (GE Healthcare Cat. # SH30871.02) and GE3 (GE Healthcare Cat. # SH30557.02). All media were supplemented with 1% antibiotic-antimycotic and 1 mL/L anti-clumping agent; all media except GE2 (which contained glutamine) was supplemented with 8 mM L-glutamine. Media screening revealed ActiPro permitted the best growth and protein production for all lines tested.


A)
VCD (10^6 cells/mL)
Viability (%)
Time (d)
B)
Titer (mg/L)
Time (d)
P-B1
C6_2
P-B2
P-H6
P-B3
P-G8
P-B5
P-B6
C)
CD CHO
FortiCHO
OptiCHO
ActiPro
GE2
GE3
Max Titer
VCD (10^6 cells/mL)
Titer (mg/L)
Time (d)
P-B1
C6_2
P-H6

Warburg-null cells achieved more mature mAb glycosylation

To assess the impact of eliminating the Warburg effect on product quality, we quantified the abundance of N-linked glycans on rituximab at three time points (late exponential-day 7, mid stationary-day 10, and just prior to culture termination-day 14). Modest changes were seen in glycosylation during at least the first 10 days of culture. Specifically, the P-H6 clone consistently produced rituximab with increased galactosylation, with increased bigalactosylated (G2F) and decreased agalactosylated glycans (G0F) (Figure 4.4C, Figure 4.S2A/B). Interestingly, by day 14, a slight increase in Man5 (Figure 4.S2C) was observed, indicating a need to optimize the fed-batch process to prolong a healthy cellular state. However, the increased terminal galactosylation, shown to improve CDC and ADCC[162,163], represents an improvement in product quality.

DISCUSSION

As CHO cells are the primary workhorse of the biopharmaceutical industry, advantageous phenotypes must be amenable to existing workflows for generating new protein producing cell lines or be introducible into existing lines without deleterious effects. We demonstrated that the Warburg-null phenotype satisfies both criteria. As a host cell line, polyclonal pools producing recombinant antibody were generated following standard industrial workflows. These pools behaved similarly to pools derived from wildtype CHO-S cells, in contrast to previous work showing Ldha downregulated cells could not survive the standard workflow[158]. Introducing the phenotype into a CHO cell line producing a commercially relevant biopharmaceutical improved growth by prolonging the exponential phase, and maintained or improved product titers while increasing galactosylated glycans. We note that these improvements were seen in processes designed around Warburg positive cells (e.g. glucose rich media where the anticipated entry to

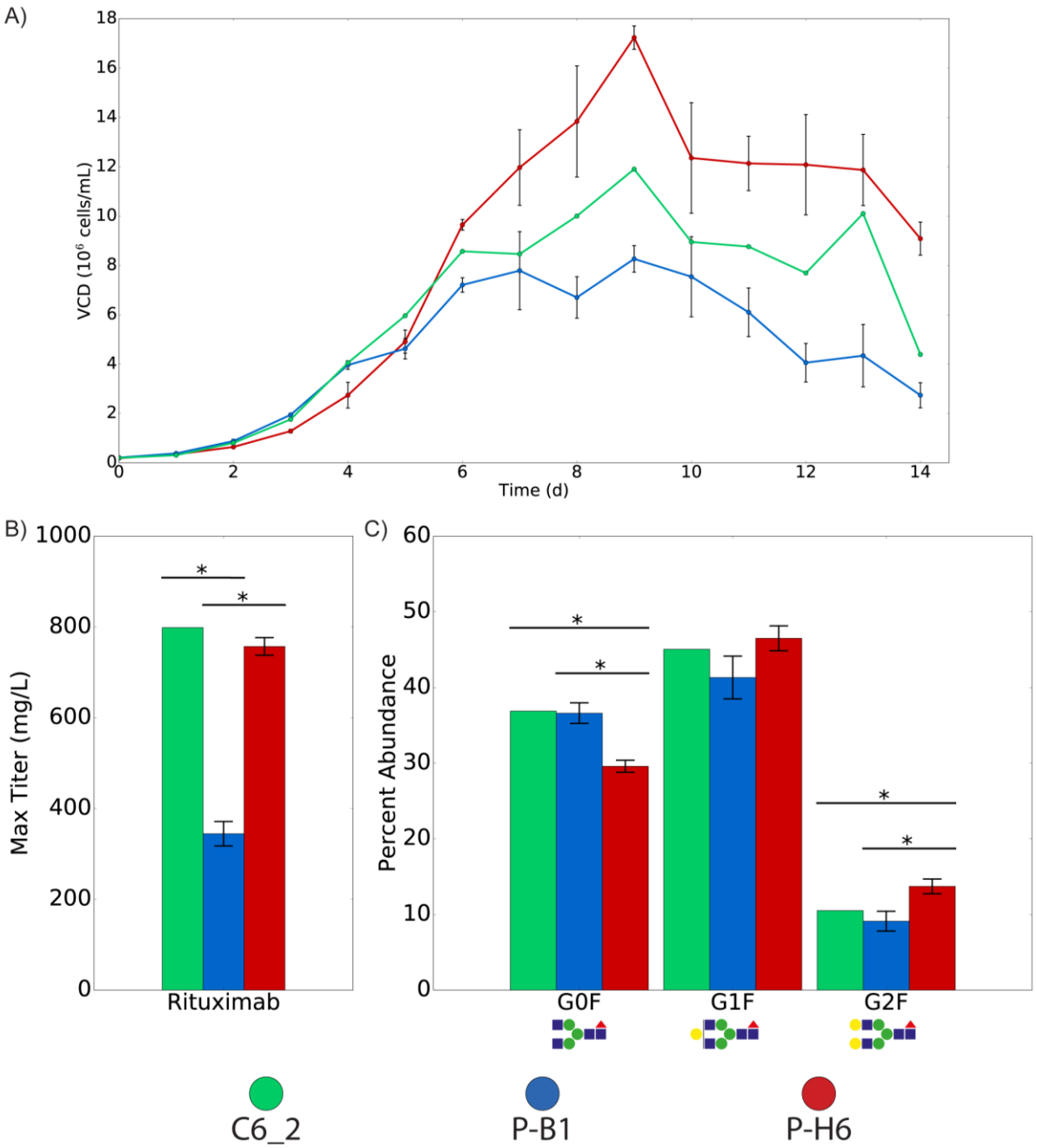


Figure 4.4: Elimination of the Warburg effect in a rituximab-producing cell line improves growth and product quality.

Parental (C6_2, green, n=1), mock control (P-B1, blue, n=3), and Warburg-null (P-H6, red, n=3) cells producing rituximab were grown in fed-batch culture. **(A)** Knockout cells showed a prolonged period of exponential growth compared to both the parental and control line. **(B)** Knockout cells were able to maintain (vs. parental) or improve (vs. control) product titer. **(C)** Warburg-null clones exhibited slight increases in galactosylation of rituximab at day 7 of the culture (see Figure 4.S2 for abundances of all glycans, including samples from other time points). **(B)** * indicates $p<0.05$ as determined by a two-sample (mock vs knockout comparison) or one-sample (parental vs mock or knockout comparisons) two-tailed t-test. (C) * indicates significance at an FDR of 0.05 as determined by the Benjamini–Hochberg method (individual comparisons were made using a two-sample two-tailed Welch's t-test for mock vs knockout comparisons and a one-sample two-tailed t-test for parental vs mock or knockout comparisons). Data shown as mean ± standard deviation.

stationary phase is osmolarity linked). Thus, we anticipate that further bioprocess optimization around the Warburg-null phenotype would further enhance these benefits. For example, without lactate secretion and the associated base addition to control pH, culture osmolarity is less of a constraint, enabling bioreactor runs of longer duration with an appropriately tailored bioprocess. There will be additional opportunity for optimizing media formulation and feeding of amino acids, owing to the interplay between the metabolism of these compounds. For example, excess amino acids are often added, since they have been shown to reduce aerobic glycolysis[164]. However amino acid catabolism generates ammonia and other growth-inhibitory compounds[165]. Without lactate secretion, amino acids could be maintained at lower concentrations and used to primarily feed protein synthesis[165]. However, these are just possibilities and further studies are needed to test the long-term implications of the removal of the Warburg effect.

METHODS

Transfection

CHO-S™ cells (Gibco Cat. # A11557-01) producing rituximab[159] were transfected using either FreeStyle MAX reagent (Gibco Cat. # 16447100). Briefly, the day prior to transfection, viable cell density was adjusted to $8x10^5$ cells/mL in transfection medium: CD CHO medium (Gibco Cat. #10743-029) supplemented with 8 mM L-glutamine (Lonza Cat. # BE17-605F). On the day of transfection, viable cell density was adjusted to $1x10^6$ cells/mL in an MD6 plate (Falcon Cat. # 351146) containing 3 mL transfection medium per well. For each transfection, 1.9 µg Cas9-2A-GFP plasmid DNA and 1.9 µg gRNA plasmid DNA was diluted in 60 µL OptiPro SFM (Gibco Cat. # 12309019). Separately, 3.8 µL FreeStyle MAX reagent was diluted in 60 µL OptiPro SFM and the two mixtures were incubated for 5 minutes at room temperature. After incubation, the plasmid DNA/OptiPro SFM mixture was added to the FreeStyle MAX/OptiPro SFM mixture and incubated at room temperature for additional 20 minutes. The resultant 120 µL DNA/lipid mixture was added dropwise to the cells in one well. Plasmids were constructed using the uracil-specific excision reagent (USER) cloning method as described previously[144], with the sgRNA1_C plasmid as a backbone.

Single cell sorting and expansion

Transfected cells were single cell sorted 48 hours post transfection, using the FACSJazz, based on green fluorescence with gating determined by comparison to non-transfected cells. Sorting was done into MD384 plates (Corning Cat. # 3542) containing CD CHO medium (Gibco Cat. # 10743-029) supplemented with 8 mM L-glutamine (Lonza Cat. # BE17-605F), 1% antibiotic-antimycotic (Gibco Cat. # 15240-062), and 1.5% HEPES buffer (Gibco Cat. # 15630-056). After 15 days, colonies were transferred to an MD96F plate (Falcon Cat. # 351172) containing CD CHO medium supplemented with 8 mM L-glutamine, and 1% antibiotic-antimycotic.

Clone genotyping

After two days, 50 uL cell suspension from each well was transferred to a MicroAmp Fast 96 well reaction plate (Thermo Cat. # 4346907), along with $5x10^5$ wildtype cells as a control. The plate was centrifuged at 1000 x g for 10 minutes, then the supernatant was removed via rapid inversion. 20 µL of QuickExtract DNA Extraction Solution (Epicentre Cat. # QE09050) (prewarmed to 65°C) was added to each well and mixed via pipetting. The plate was then processed in the thermocycler (65°C for 15 minutes followed by 95°C for 5 minutes).

Amplicons were generated for each gene of interest per well using Phusion Hot Start II DNA Polymerase (Thermo Cat. # F549L) and verified to be present visually on a 2% agarose gel. Amplicons from each well had unique barcodes, allowing them to be pooled and purified using AMPure XP beads (Beckman Coulter Cat. # A63881) according to manufacturer's protocol, except using 80% ethanol for washing steps and 40 uL beads for 50 uL sample. Samples were indexed using the Nextera XT Index kit attached using 2x KAPA HiFi Hot Start Ready mix (Fisher Scientific Cat. # KK2602). AMPure XP beads were used to purify the resulting PCR products.

DNA concentrations were determined with the Qubit 2.0 Fluorometer and used to pool all indices to an equimolar value and diluted to a final concentration of 10 nM using 10mM Tris pH 8.5, 0.1% Tween 20. The average size of the final library was verified with the Bioanalyzer 2100. The amplicon library was then sequenced on an Illumina MiSeq.

Insertions and deletions were identified by comparison of expected vs. actual amplicon size. Clones with frameshift indels in all alleles of the Ldha gene were selected for expansion.

Batch culture

CHO-S™ cells (Gibco Cat. # A11557-01) and derivative clones were cultured in complete CHO-S medium: CD CHO medium supplemented with 8 mM L-glutamine and 2 mL/L of Anti-clumping agent (Gibco Cat. # 0010057AE). All cultures were maintained in an incubator at 37°C, 5% CO_2, 70-95% humidity and 25 mm throw, while shaking at 120 rpm. Cell growth and viability

were monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec, Denmark) based on two fluorescent dyes, Acridine Orange and DAPI for the total and dead cell populations, respectively. Metabolite concentrations were measured using the BioProfile 400 (Nova Biomedical, USA).

Pool generation

Host cell lines were transfected with a vector containing the GS and antibody genes using FreeStyleMAX (Invitrogen) according to the manufacturer's protocol. Twenty-four hours after transfection, cells were inoculated at a concentration of $1x10^6$ cells/mL into 125-mL Erlenmeyer flasks containing 50 mL of selection medium (CD CHO supplemented with GSEM, 25 µM MSX, 300 µg/mL zeocin and 2 µL/mL anti-clumping agent). The flasks were subsequently incubated in a Climo-shaking incubator (Kuhner AG, Basel, Switzerland) at 110 rpm in a humidified 5% CO_2/air mixture at 37°C. Every third day, medium was exchanged following centrifugation of cells at 1200 rpm for 5 minutes and reseeded at up to $1x10^6$ cells/mL. When viable cell concentration first exceeded $2x10^6$ cells/mL, cells were reseeded at $5x10^5$ cells/mL; each subsequent reseeding was at $3x10^5$ cells/mL. When viability exceeded 90%, pool selection was considered complete.

Fed-batch culture

Cells were grown from a seed density of $2.5x10^5$ cells/mL in DASGIP bioreactors with a starting volume of 270 mL at a temperature of 37°C, agitated at 200 rpm using pitched blade impellers. The pH was maintained continuously at 7.10±0.02 with 1 M sodium bicarbonate or CO_2. Dissolved oxygen was maintained at 40% using pure oxygen, air, or a mixture, as needed. A 2222 mM glucose solution and a 200 mM glutamine solution were used to control glucose and glutamine levels at the desired levels via a once-daily feeding based on the measured concentration, doubling time, and calculated consumption rates. Complex feed addition was started when cell concentration was approximately $1.4\text{-}2.2x10^6$ cells/mL and continued for 4 days,

feeding increasing volumes each successive day. Cultures were sampled daily for cell growth and viability using the NucleoCounter NC-200 Cell Counter, metabolite concentrations using the BioProfile 400, and rituximab concentration using the OctetRED96 for protein producing cultures. Process variations for the different experiments are detailed below:

Clone 0-F5

Cells were maintained in OptiCHO medium supplemented with 8 mM L-glutamine, 1% antibiotic-antimycotic, and 1 mL/L anti-clumping agent. Antifoam C was added as needed over the course of culture. Glucose was maintained between 10-24 mM. Glutamine was maintained between 1-3 mM. Cell Boost 7a (2% reactor volume) and 7b (0.2% reactor volume) (GE Healthcare) were added from day 4 to 12:

Clone P-H6

Cells were maintained in ActiPro medium (GE Healthcare Cat. # SH31039.02) supplemented with 8 mM L-glutamine, 1% antibiotic-antimycotic, and 1 mL/L anti-clumping agent. Antifoam C was added as needed over the course of culture. Glucose was maintained between 10-24 mM. Glutamine was maintained between 2-4 mM. Cell Boost 7a and 7b were added starting at day 4 at roughly the following volumes:

Day 4: 4 mL 7a + 0.4 mL 7b

Day 5: 8 mL 7a + 0.8 mL 7b

Day 6: 12 mL 7a + 1.2 mL 7b

Day 7: 17.5 mL 7a + 1.6 mL 7b

Rituximab quantification

To quantify rituximab, biolayer interferometry was performed using an Octet RED96 (Pall Corporation, Menlo Park, CA). ProA biosensors (Fortebio 18-5013) were hydrated in PBS and

preconditioned in 10 mM glycine pH 1.7. A calibration curve was prepared using human IgG at 400, 200, 100, 50, 25, 12.5, and 6.3 mg/L. Rituximab and control sample supernatants were collected after centrifugation and association was performed for 120 seconds with a shaking speed of 200 rpm at 30°C. Octet System Data Analysis 7.1 software was used to calculate binding rates and absolute rituximab concentrations.

Glycan analysis

After centrifugation and filtration to remove the cells and cell debris, the secreted mAbs were purified from 5 mL aliquots of culture supernatant. Purification was done using protein A affinity chromatography (MabSelect recombinant protein A agarose, GE Healthcare, Uppsala, Sweden), according to the manufacturer's protocol. Purified mAbs were fluorescently labeled with GlykoPrep Rapid N-Glycan kit (ProZyme, Hayward, CA), according to the manufacturer's protocol. N-linked glycan analysis was performed by LC-MS system using a Thermo Ultimate 3000 HPLC with the fluorescence detector coupled on-line to a Thermo Velos Pro Iontrap MS, as described previously[141].

SUPPLEMENTARY FIGURES & TABLES

Table 4.S1: Genotypes of isolated rituximab-producing Warburg-null clones.
Black box indicates wildtype alleles for the gene, X a full knockout, ½ a partial knockout. Shading of the cell indicates what portion of the observed indels are in-frame, white: all indels frameshift, red: all indels in-frame, yellow: some indels frameshift, some in-frame.

Clone	Parent	Pdk1	Pdk2	Pdk3	Pdk4	Ldha
P-H6	C6_2		X	X	X	X
P-G8	C6_2		X		X	X

Table 4.S2: Detailed genotypes of isolated rituximab-producing Warburg-null clones.
Indel sizes in targeted genes for all characterized clones.

Clone	Pdk1	Pdk2	Pdk3	Pdk4	Ldha
P-H6	-	-2/-1/1	1	-13/1	-2/-1
P-G8	-	-3/1	-	-25/-19/-2/8	-73/-10/-1

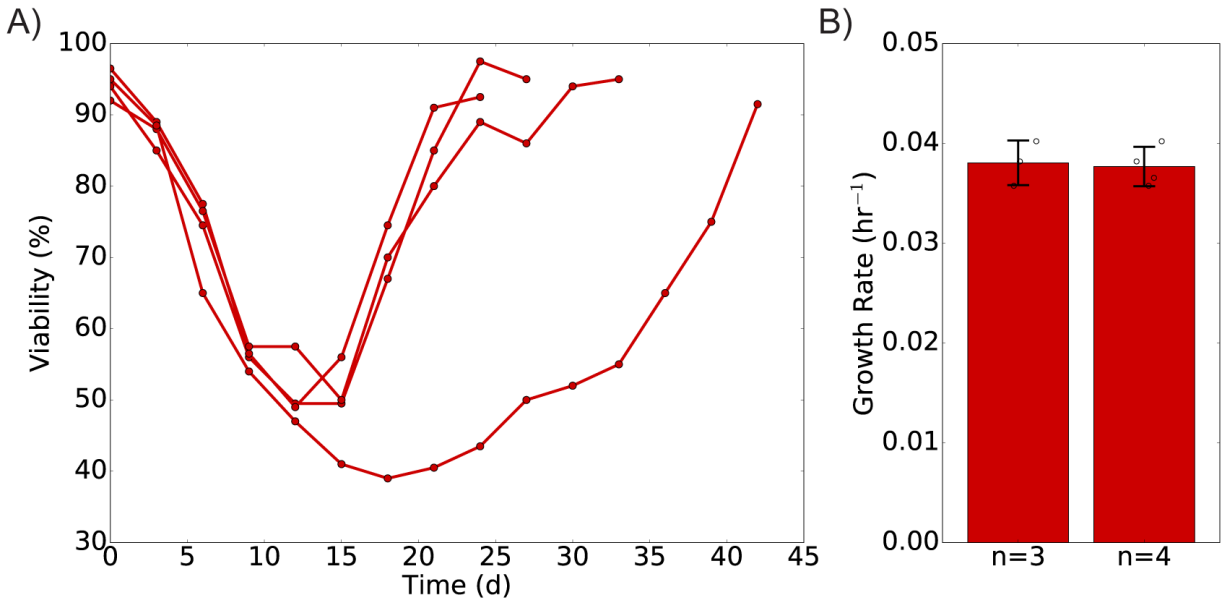
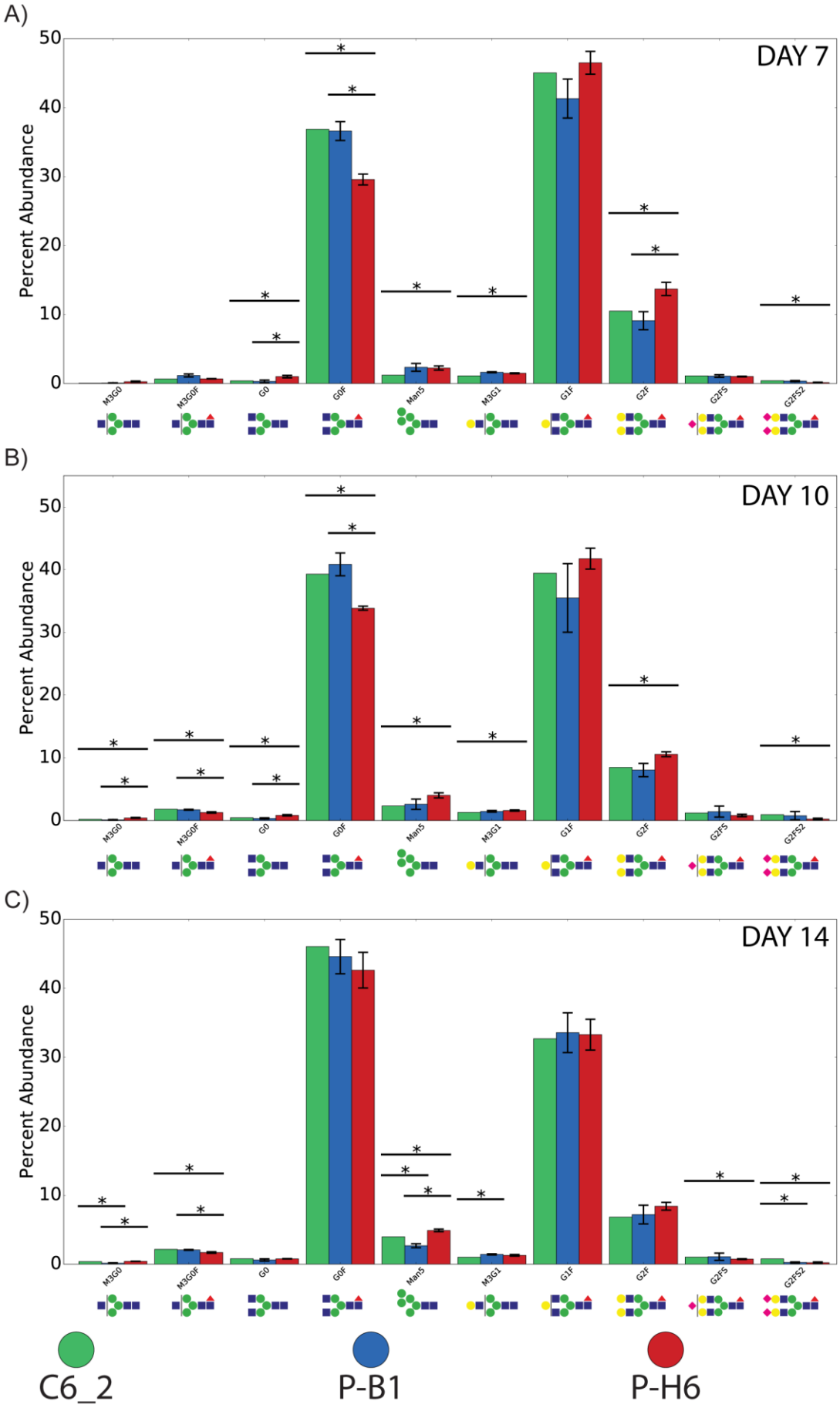


Figure 4.S1: Outlier in pool generation process for clone 0-B6.
During the cell line generation process (Figure 4.1), one biological replicate of clone 0-B6 exhibited a drastically longer recovery time **(A)** for unknown reasons. While it is not included in the main text, preliminary characterization of the resultant pool with respect to growth **(B)** shows that it does not behave differently than any of the other 0-B6 derived clones (n=3, pools characterized in Figure 4.1; n=4, pools characterized in Figure 4.1 as well as the outlier from panel A).

Figure 4.S2: Rituximab glycoprofiles during fed-batch culture of parental and engineered cell lines. Parental (C6_2, green, n=1), mock control (P-B1, blue, n=3), and Warburg-null (P-H6, red, n=3) cells producing rituximab in fed-batch (Figure 4.4) had product glycoprofiles analyzed on day 7 **(A)**, 10 **(B)**, and 14 **(C)** of culture. * indicates significance at an FDR of 0.05 as determined by the Benjamini–Hochberg method: (individual comparisons were made using a two-sample two-tailed Welch's t-test for mock vs knockout comparisons; one-sample two-tailed t-test for parental vs mock or knockout comparisons). Data shown as mean ± standard deviation.


A)
DAY 7
Percent Abundance
50
40
30
20
10
0
M3G0
M3G0F
G0
G0F
Man5
M3G1
G1F
G2F
G2FS
G2FS2
B)
DAY 10
Percent Abundance
C)
DAY 14
Percent Abundance
C6_2
P-B1
P-H6

Chapter 4 contains material from Hefzi, H., Noh, S.M., Monge, I.M., Decker, M., Arnsdorf, J., Kol, S., Pristovšek, N., Hansen, A.H., Bjorn, S.P., Brøndum, K.K., Javidi, E.M., Jensen, K.L., Kallehauge, T.B., Ley, D., Ménard, P., Petersen, H.M., Sukhova, Z., Voldborg, B.G., Nielsen, L.K., Lee, G.M., Kildegaard, H.F., Lewis, N.E. Multiplex genome editing eliminates the Warburg Effect without affecting oxidative metabolism or growth rate. (Submitted). I was the primary author, while the co-authors provided support in the research underlying the study.

Future Perspectives: Towards systems biology-based engineering in a Warburg-null CHO cell line

In the course of this work, we have been able to catalog a metabolic "parts list" for CHO cells and generate a reconstruction of the metabolic network underlying growth and protein production. This is an important development in enabling more complex metabolic engineering as it permits identification of non-intuitive targets based on computational simulations around optimal cellular behavior. I have further identified a genetic engineering strategy to eliminate the Warburg-effect in this cell line, an important advancement for bioprocessing (due to the deleterious effect of lactate accumulation) and a valuable tool for probing fundamental biological questions about the near-ubiquity of the phenotype in proliferative mammalian cells.

Engineering improved growth beyond the Warburg effect

Recent work[165] has shown that when lactate is not allowed to accumulate (via process optimization), a secondary set of growth inhibitory molecules leads to the cessation of growth (later than the cessation caused by lactate/osmolite accumulation). While the mechanism of growth inhibition is unknown, these molecules—the intermediates in various amino acid catabolic pathways—have a marked effect. Process optimization was able to reduce production of these compounds[165], however the method of control—necessitating careful maintenance of amino acid levels at low levels—is a technical challenge. Owing to their peripheral position in the metabolic network, these compounds represent an interesting test case for elimination via model-guided target identification. Indeed, we were able to find genes coding for enzymes responsible for these—and other—amino acid catabolic intermediates. Preliminary characterization of some knockouts in a wildtype background shows a growth benefit (data not shown); introduction and characterization of these knockouts into a Warburg-null cell line is underway.

Model guided target discovery

The model has also proven useful in designing forward genetic screens via the CRISPR-Cas9 system[166,167]. Screening against metabolic genes (those in the model, those with unknown metabolic activity, and associated transcription factors) in an industrially relevant condition has revealed novel genetic targets with improved phenotypes (data not shown). Furthermore, screen results can be used to validate and improve the model based on gene essentiality predictions.

Metabolic phenotypes are complex and rarely the result of a single gene. While combinatorial screens[142,143] permit exploration of the phenotypic solution space, the scale (e.g., more than 3,000,000 unique pairs of metabolic genes from iCHO1766[58]) makes exhaustive characterization infeasible. Model guided design of screen targets, for example selecting genes in specific pathways(s), represents a potential way for targeted investigation of the metabolic flexibility of CHO cells.

Exploring the Warburg effect

The development of a CHO cell line that secretes negligible lactate while remaining proliferative permits more rigorous probing of the Warburg effect by serving as a true negative control for investigatory experiments. The hallmark of the Warburg effect—increased glucose uptake concomitant with lactate secretion—results in increases in energy molecule pools (e.g., ATP, NAD+/NADH ratio, etc.). Probing cellular robustness to various energetic stressors may thus provide insight into the benefit of the Warburg effect. Preliminary analyses around energy molecule levels and gene expression paint the outlines of an intriguing portrait. It is likely that the lack of cytosolic acidification (due to negligible lactate production) necessitates less proton efflux via carbonic anhydrase mediated conversion of HCO_3 to CO_2 and H_2O, consistent with previous work[121]. As HCO_3 can be cotransported into the cell with Na+, the Warburg effect could coincide with higher Na+/K+ ATPase activity (to mediate Na+ efflux), requiring higher ATP generation. We have also seen that the NAD+/NADH ratio is severely decreased in Warburg-null cells producing

rituximab (P-H6, Table 4.S1), with no apparent deleterious effect. It is possible, therefore, that increases in ATP generation and NAD+/NADH ratio are necessary outcomes of the Warburg effect (being needed to deacidify the cytosol and unavoidable with high lactate dehydrogenase flux, respectively) rather than the evolutionary advantage that has made this phenotype so ubiquitous. Experimental validation of this hypothesis is planned and should provide valuable insight.

The work presented here clearly demonstrates the value of following the path of microbial systems biology-based engineering to guide the progression of CHO cell line engineering. With a 'parts list' from the genome sequence of CHO[8] and *C. griseus*[25], we have been able to reconstruct a genome-scale model of CHO cell metabolism. This model identified different efficiencies in redirecting metabolic resources from growth to protein production following various bioprocess or genetic modifications. It has also been deployed to identify gene targets for eliminating growth inhibitory byproducts. Thus, the genome-scale model allows *in silico* evaluation of engineering strategies before experimental testing. By utilizing the CRISPR-Cas9 gene editing tool, we were able to eliminate the Warburg effect, a phenotype that has plagued the biopharmaceutical industry for decades. The work presented here clearly demonstrates that—while still in its infancy—the era of complex systems biology guided engineering in CHO has begun. The ability to rationally engineer this cell line holds great promise for decreasing production costs and ultimately the cost of life-saving biopharmaceuticals.

References

1. Lindsley, C. W. New 2016 Data and Statistics for Global Pharmaceutical Products and Projections through 2017. *ACS Chem. Neurosci.* **8,** 1635–1636 (2017).

2. *Deloitte 2016 Global life sciences outlook*. (2015).

3. Baeshen, N. A., Baeshen, M. N., Sheikh, A., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A. I., Saini, K. S. and Redwan, E. M. Cell factories for insulin production. *Microb. Cell Fact.* **13,** 141 (2014).

4. Durocher, Y. and Butler, M. Expression systems for therapeutic glycoprotein production. *Curr. Opin. Biotechnol.* **20,** 700–707 (2009).

5. Urlaub, G. and Chasin, L. A. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc. Natl. Acad. Sci. U. S. A.* **77,** 4216–20 (1980).

6. Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. Transformation of mammalian cells with genes from procaryotes and eucaryotes. *Cell* **16,** 777–85 (1979).

7. Jayapal, K., Wlaschin, K., Hu, W. S. and Yap, M. Recombinant Protein Therapeutics from Cho Cells - 20 Years and Counting. *CHO Consort. SBE Spec. Ed.* 40–47 (2007).

8. Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M. R., Neff, N., Passarelli, B., Koh, W., Fan, H. C., Wang, J., Gui, Y., Lee, K. H., Betenbaugh, M. J., Quake, S. R., Famili, I., Palsson, B. O. and Wang, J. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* **29,** 735–41 (2011).

9. Muchmore, E. A., Milewski, M., Varki, A. and Diaz, S. Biosynthesis of N-glycolyneuraminic acid. The primary site of hydroxylation of N-acetylneuraminic acid is the cytosolic sugar nucleotide pool. *J. Biol. Chem.* **264,** 20216–23 (1989).

10. Ghaderi, D., Taylor, R. E., Padler-Karavani, V., Diaz, S. and Varki, A. Implications of the presence of N-glycolylneuraminic acid in recombinant therapeutic glycoproteins. *Nat. Biotechnol.* **28,** 863–7 (2010).

11. Walsh, G. Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* **32,** 992–1000 (2014).

12. Kao, F. T. and Puck, T. T. Genetics of somatic mammalian cells. IV. Properties of Chinese hamster cell mutants with respect to the requirement for proline. *Genetics* **55,** 513–24 (1967).

13. Naylor, S. L., Townsend, J. K. and Klebe, R. J. Characterization of naturally occurring auxotrophic mammalian cells. *Somatic Cell Genet.* **5,** 271–7 (1979).

14. Valle, D., Downing, S. J., Harris, S. C. and Phang, J. M. Proline biosynthesis: Multiple defects in Chinese hamster ovary cells. *Biochem. Biophys. Res. Commun.* **53,** 1130–1136 (1973).

15. Duarte, T. M., Carinhas, N., Barreiro, L. C., Carrondo, M. J. T., Alves, P. M. and Teixeira, A. P. Metabolic responses of CHO cells to limitation of key amino acids. *Biotechnol. Bioeng.* **111,** 2095–106 (2014).

16. Park, J. H. and Lee, S. Y. Towards systems metabolic engineering of microorganisms for amino acid production. *Curr. Opin. Biotechnol.* **19,** 454–60 (2008).

17. Backus, M. P. and Stauffer, J. F. The Production and Selection of a Family of Strains in Penicillium chrysogenum. *Mycologia* **47,** 429–463 CR–Copyright © 1955 Mycological So (1955).

18. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. The complete genome sequence of Escherichia coli K-12. *Science* **277,** 1453–62 (1997).

19. Datsenko, K. A. and Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6640–5 (2000).

20. McCloskey, D., Palsson, B. Ø. and Feist, A. M. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Mol. Syst. Biol.* **9,** 661 (2013).

21. Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J. and Van Dien, S. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7,** 445–452 (2011).

22. Curran, K. A., Leavitt, J. M., Karim, A. S. and Alper, H. S. Metabolic engineering of muconic acid production in Saccharomyces cerevisiae. *Metab. Eng.* **15,** 55–66 (2013).

23. Chong, W. P. K., Reddy, S. G., Yusufi, F. N. K., Lee, D. Y., Wong, N. S. C., Heng, C. K., Yap, M. G. S. and Ho, Y. S. Metabolomics-driven approach for the improvement of Chinese hamster ovary cell growth: overexpression of malate dehydrogenase II. *J. Biotechnol.* **147,** 116–21 (2010).

24. Zhou, M., Crawford, Y., Ng, D., Tung, J., Pynn, A. F. J., Meier, A., Yuk, I. H., Vijayasankaran, N., Leach, K., Joly, J., Snedecor, B. and Shen, A. Decreasing lactate level and increasing antibody production in Chinese Hamster Ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. *J. Biotechnol.* **153,** 27–34 (2011).

25. Lewis, N. E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A. M., Rosenbloom, J., Bian, C., Xie, M., Chen, W., Li, N., Baycin-Hizal, D., Latif, H., Forster, J., Betenbaugh, M. J., Famili, I., Xu, X., Wang, J. and Palsson, B. O. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nat. Biotechnol.* **31,** 759–65 (2013).

26. Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., Kofler, R., Romand, S., Hesse, F., Budach, W. E., Galosy, S., Müller, D., Noll, T., Wienberg, J., Jostock, T.,

Leonard, M., Grillari, J., Tauch, A., Goesmann, A., Helk, B., Mott, J. E., Pühler, A. and Borth, N. Chinese hamster genome sequenced from sorted chromosomes. *Nat. Biotechnol.* **31,** 694–5 (2013).

27. Selvarasu, S., Ho, Y. S., Chong, W. P. K., Wong, N. S. C., Yusufi, F. N. K., Lee, Y. Y., Yap, M. G. S. and Lee, D. Y. Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnol. Bioeng.* **109,** 1415–29 (2012).

28. Martínez, V. S., Dietmair, S., Quek, L. E., Hodson, M. P., Gray, P. and Nielsen, L. K. Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. *Biotechnol. Bioeng.* **110,** 660–6 (2013).

29. Kim, H. and Kim, J. S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15,** 321–34 (2014).

30. Peng, R. W. and Fussenegger, M. Molecular engineering of exocytic vesicle traffic enhances the productivity of Chinese hamster ovary cells. *Biotechnol. Bioeng.* **102,** 1170–81 (2009).

31. Lewis, N. E., Nagarajan, H. and Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10,** 291–305 (2012).

32. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. and Palsson, B. Ø. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* **9,** 693 (2013).

33. Feizi, A., Österlund, T., Petranovic, D., Bordel, S. and Nielsen, J. Genome-scale modeling of the protein secretory machinery in yeast. *PLoS One* **8,** e63284 (2013).

34. Kildegaard, H. F., Baycin-Hizal, D., Lewis, N. E. and Betenbaugh, M. J. The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. *Curr. Opin. Biotechnol.* **24,** 1102–7 (2013).

35. Gutierrez, J. M. and Lewis, N. E. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. *Biotechnol. J.* **10,** 939–49 (2015).

36. Kaufman, R. J. and Sharp, P. A. Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase complementary dna gene. *J. Mol. Biol.* **159,** 601–21 (1982).

37. Bebbington, C. R., Renner, G., Thomson, S., King, D., Abrams, D. and Yarranton, G. T. High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. *Biotechnology. (N. Y).* **10,** 169–75 (1992).

38. Castro, P. L., Hayter, P., Ison, A. and Bull, A. Application of a statistical design to the optimization of culture medium for recombinant interferon-gamma production by Chinese hamster ovary cells. *Appl. Microbiol. Biotechnol.* **38,** (1992).

39. Lao, M. S. S. and Toth, D. Effects of ammonium and lactate on growth and metabolism of a recombinant Chinese hamster ovary cell culture. *Biotechnol. Prog.* **13,** 688–91 (1997).

40. Chen, P. and Harcum, S. W. Effects of elevated ammonium on glycosylation gene expression in CHO cells. *Metab. Eng.* **8,** 123–32 (2006).

41. Yang, M. and Butler, M. Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation. *Biotechnol. Bioeng.* **68,** 370–80 (2000).

42. Fan, Y., Jimenez Del Val, I., Müller, C., Wagtberg Sen, J., Rasmussen, S. K., Kontoravdi, C., Weilguny, D. and Andersen, M. R. Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. *Biotechnol. Bioeng.* **112,** 521–35 (2015).

43. Lorendeau, D., Christen, S., Rinaldi, G. and Fendt, S. M. Metabolic control of signalling pathways and metabolic auto-regulation. *Biol. Cell* **107,** 251–72 (2015).

44. Altamirano, C., Berrios, J., Vergara, M. and Becerra, S. Advances in improving mammalian cells metabolism for recombinant protein production. *Electron. J. Biotechnol.* **16,** (2013).

45. Templeton, N., Dean, J., Reddy, P. and Young, J. D. Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. *Biotechnol. Bioeng.* **110,** 2013–24 (2013).

46. Carinhas, N., Duarte, T. M., Barreiro, L. C., Carrondo, M. J. T., Alves, P. M. and Teixeira, A. P. Metabolic signatures of GS-CHO cell clones associated with butyrate treatment and culture phase transition. *Biotechnol. Bioeng.* **110,** 3244–57 (2013).

47. Nolan, R. P. and Lee, K. Dynamic model of CHO cell metabolism. *Metab. Eng.* **13,** 108–24 (2011).

48. Zamorano, F., Wouwer, A. Vande and Bastin, G. A detailed metabolic flux analysis of an underdetermined network of CHO cells. *J. Biotechnol.* **150,** 497–508 (2010).

49. Sengupta, N., Rose, S. T. and Morgan, J. A. Metabolic flux analysis of CHO cell metabolism in the late non-growth phase. *Biotechnol. Bioeng.* **108,** 82–92 (2011).

50. Martínez, V. S., Buchsteiner, M., Gray, P., Nielsen, L. K. and Quek, L. E. Dynamic metabolic flux analysis using B-splines to study the effects of temperature shift on CHO cell metabolism. *Metab. Eng. Commun.* **2,** 46–57 (2015).

51. Smallbone, K. Standardized network reconstruction of CHO cell metabolism. (2013).

52. Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R. and Palsson, B. Ø. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 1777–82 (2007).

53. Quek, L. E., Dietmair, S., Hanscho, M., Martínez, V. S., Borth, N. and Nielsen, L. K. Reducing Recon 2 for steady-state flux analysis of HEK cell culture. *J. Biotechnol.* **184,** 172–8 (2014).

54. Swainston, N., Smallbone, K., Hefzi, H., Dobson, P. D., Brewer, J., Hanscho, M., Zielinski, D. C., Ang, K. S., Gardiner, N. J., Gutierrez, J. M., Kyriakopoulos, S., Lakshmanan, M., Li, S., Liu, J. K., Martinez, V. S., Orellana, C. A., Quek, L. E., Thomas, A., Zanghellini, J., Borth, N., Lee, D. Y., Nielsen, L. K., Kell, D. B., Lewis, N. E. and Mendes, P. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* (2016).

55. Ostlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D. N., Roopra, S., Frings, O. and Sonnhammer, E. L. L. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **38,** D196-203 (2010).

56. Feist, A. M. and Palsson, B. O. The biomass objective function. *Curr. Opin. Microbiol.* **13,** 344–349 (2010).

57. Becker, S. A. and Palsson, B. O. Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* **4,** e1000082 (2008).

58. Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C. A., Baycin-Hizal, D., Huang, Y., Ley, D., Martinez, V. S., Kyriakopoulos, S., Jiménez, N. E., Zielinski, D. C., Quek, L. E., Wulff, T., Arnsdorf, J., Li, S., Lee, J. S., Paglia, G., Loira, N., Spahn, P. N., Pedersen, L. E., Gutierrez, J. M., King, Z. A., Lund, A. M., Nagarajan, H., Thomas, A., Abdel-Haleem, A. M., Zanghellini, J., Kildegaard, H. F., Voldborg, B. G., Gerdtzen, Z. P., Betenbaugh, M. J., Palsson, B. O., Andersen, M. R., Nielsen, L. K., Borth, N., Lee, D. Y. and Lewis, N. E. A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Syst.* **3,** 434–443.e8 (2016).

59. King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O. and Lewis, N. E. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44,** D515-22 (2015).

60. Kremkow, B. G., Baik, J. Y., MacDonald, M. L. and Lee, K. H. CHOgenome.org 2.0: Genome resources and website updates. *Biotechnol. J.* **10,** 931–8 (2015).

61. Ahn, W. S. and Antoniewicz, M. R. Parallel labeling experiments with [1,2-(13)C]glucose and [U-(13)C]glutamine provide new insights into CHO cell metabolism. *Metab. Eng.* **15,** 34–47 (2013).

62. Ahn, W. S. and Antoniewicz, M. R. Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. *Metab. Eng.* **13,** 598–609 (2011).

63. Borman, A., Wood, T. R., Black, H. C., Anderson, E. G., Oesterling, M. J., Womack, M. and Rose, W. C. The role of arginine in growth with some observations on the effects of argininic acid. *J. Biol. Chem.* **166,** 585–594 (1946).

64. Kim, Y. G. and Lee, G. M. Bcl-xL overexpression does not enhance specific erythropoietin productivity of recombinant CHO cells grown at 33 degrees C and 37 degrees C. *Biotechnol. Prog.* **25,** 252–6 (2009).

65. Kim, N. S. and Lee, G. M. Overexpression of bcl-2 inhibits sodium butyrate-induced

apoptosis in Chinese hamster ovary cells resulting in enhanced humanized antibody production. *Biotechnol. Bioeng.* **71,** 184–93 (2000).

66. De Leon Gatti, M., Wlaschin, K. F., Nissom, P. M., Yap, M. and Hu, W. S. Comparative transcriptional analysis of mouse hybridoma and recombinant Chinese hamster ovary cells undergoing butyrate treatment. *J. Biosci. Bioeng.* **103,** 82–91 (2007).

67. Mimura, Y., Lund, J., Church, S., Dong, S., Li, J., Goodall, M. and Jefferis, R. Butyrate increases production of human chimeric IgG in CHO-K1 cells whilst maintaining function and glycoform profile. *J. Immunol. Methods* **247,** 205–16 (2001).

68. Clarke, C., Barron, N., Meleady, P. and Clynes, M. Statistical methods for mining Chinese hamster ovary cell 'omics data: from differential expression to integrated multilevel analysis of the biological system. *Pharm. Bioprocess.* **2,** 469–481 (2014).

69. Cardoso, J. G. R., Andersen, M. R., Herrgård, M. J. and Sonnenschein, N. Analysis of genetic variation and potential applications in genome-scale metabolic modeling. *Front. Bioeng. Biotechnol.* **3,** 13 (2015).

70. Hyduke, D. R., Lewis, N. E. and Palsson, B. Ø. Analysis of omics data with genome-scale models of metabolism. *Mol. Biosyst.* **9,** 167–74 (2013).

71. Golabgir, A., Gutierrez, J. M., Hefzi, H., Li, S., Palsson, B. O., Herwig, C. and Lewis, N. E. Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. *Biotechnol. Adv.* (2016). doi:10.1016/j.biotechadv.2016.02.011

72. Thiele, I., Vlassis, N. and Fleming, R. M. T. fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics* **30,** 2529–31 (2014).

73. Spahn, P. N., Hansen, A. H., Hansen, H. G., Arnsdorf, J., Kildegaard, H. F. and Lewis, N. E. A Markov chain model for N-linked protein glycosylation - towards a low-parameter tool for model-driven glycoengineering. *Metab. Eng.* **33,** 52–66 (2016).

74. Sha, S., Agarabi, C., Brorson, K., Lee, D. Y. and Yoon, S. N-Glycosylation Design and Control of Therapeutic Monoclonal Antibodies. *Trends Biotechnol.* (2016). doi:10.1016/j.tibtech.2016.02.013

75. Chowdhury, R., Chowdhury, A. and Maranas, C. D. Using Gene Essentiality and Synthetic Lethality Information to Correct Yeast and CHO Cell Genome-Scale Models. *Metabolites* **5,** 536–70 (2015).

76. Bordbar, A., Mo, M. L., Nakayasu, E. S., Schrimpe-Rutledge, A. C., Kim, Y. M. M., Metz, T. O., Jones, M. B., Frank, B. C., Smith, R. D., Peterson, S. N., Hyduke, D. R., Adkins, J. N. and Palsson, B. O. Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. *Mol. Syst. Biol.* **8,** 558 (2012).

77. Daar, I. O., Artymiuk, P. J., Phillips, D. C. and Maquat, L. E. Human triose-phosphate isomerase deficiency: a single amino acid substitution results in a thermolabile enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 7903–7 (1986).

78. Kanehisa, M. and Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

79. Chang, A., Scheer, M., Grote, A., Schomburg, I. and Schomburg, D. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res.* **37,** D588-92 (2009).

80. Loira, N., Zhukova, A. and Sherman, D. J. Pantograph: A template-based method for genome-scale metabolic model reconstruction. *J. Bioinform. Comput. Biol.* **13,** 1550006 (2015).

81. Selvarasu, S., Kim, D. Y., Karimi, I. A. and Lee, D. Y. Combined data preprocessing and multivariate statistical analysis characterizes fed-batch culture of mouse hybridoma cells for rational medium design. *J. Biotechnol.* **150,** 94–100 (2010).

82. Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., Kang, J., Hyduke, D. R. and Palsson, B. O. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6,** 1290–1307 (2011).

83. Satish Kumar, V., Dasika, M. S. and Maranas, C. D. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* **8,** 212 (2007).

84. Edwards, J. S., Ibarra, R. U. and Palsson, B. O. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* **19,** 125–130 (2001).

85. Ibarra, R. U., Edwards, J. S. and Palsson, B. O. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420,** 186–189 (2002).

86. Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K., Eils, R., Konig, R., Smith, R. D. and Palsson, B. O. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6,** 390 (2010).

87. Schuetz, R., Kuepfer, L. and Sauer, U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Mol. Syst. Biol.* **3,** 119 (2007).

88. Hiller, G. W. and Mulukutla, B. C. Method of cell culture. (2015).

89. Borgström, E., Lundin, S. and Lundeberg, J. Large scale library generation for high throughput sequencing. *PLoS One* **6,** e19119 (2011).

90. Lundin, S., Stranneheim, H., Pettersson, E., Klevebring, D. and Lundeberg, J. Increased throughput by parallelization of library preparation for massive sequencing. *PLoS One* **5,** e10029 (2010).

91. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).

92. Bolger, A. M., Lohse, M. and Usadel, B. Trimmomatic: a flexible trimmer for Illumina

sequence data. *Bioinformatics* **30,** 2114–20 (2014).

93. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

94. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–9 (2009).

95. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J. and Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–5 (2010).

96. Rappsilber, J., Mann, M. and Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2,** 1896–906 (2007).

97. Baycin-Hizal, D., Tabb, D. L., Chaerkady, R., Chen, L., Lewis, N. E., Nagarajan, H., Sarkaria, V., Kumar, A., Wolozny, D., Colao, J., Jacobson, E., Tian, Y., O'Meally, R. N., Krag, S. S., Cole, R. N., Palsson, B. O., Zhang, H. and Betenbaugh, M. Proteomic analysis of Chinese hamster ovary cells. *J. Proteome Res.* **11,** 5265–76 (2012).

98. Courtes, F. C., Lin, J., Lim, H. L., Ng, S. W., Wong, N. S. C., Koh, G., Vardy, L., Yap, M. G. S., Loo, B. and Lee, D. Y. Translatome analysis of CHO cells to identify key growth genes. *J. Biotechnol.* **167,** 215–24 (2013).

99. Bolstad, B. M., Collin, F., Brettschneider, J., Simpson, K., Cope, L., Irizarry, R. A. and Speed, T. P. Quality Assessment of Affymetrix GeneChip Data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor SE - 3* (eds. Gentleman, R., Carey, V., Huber, W., Irizarry, R. & Dudoit, S.) 33–47 (Springer New York, 2005). doi:10.1007/0-387-29362-0_3

100. Paglia, G., Hrafnsdóttir, S., Magnúsdóttir, M., Fleming, R. M. T., Thorlacius, S., Palsson, B. Ø. and Thiele, I. Monitoring metabolites consumption and secretion in cultured cells using ultra-performance liquid chromatography quadrupole-time of flight mass spectrometry (UPLC-Q-ToF-MS). *Anal. Bioanal. Chem.* **402,** 1183–98 (2012).

101. Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorndahl, T., Perez-Pineiro, R., Eisner, R., Allen, F., Neveu, V., Greiner, R. and Scalbert, A. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res.* **41,** D801-7 (2013).

102. Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R. and Siuzdak, G. METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **27,** 747–51 (2005).

103. Hölttä, E. and Pohjanpelto, P. Polyamine dependence of Chinese hamster ovary cells in serum-free culture is due to deficient arginase activity. *Biochim. Biophys. Acta* **721,** 321–

7 (1982).

104. Gorfien, S. F., Paul, W., Judd, D., Tescione, L. and Jayme, D. W. Optimized Nutrient Additives for Fed-Batch Cultures. *BioPharm Int.* **16,** 34 (2003).

105. Altamirano, C., Illanes, A., Casablancas, A., Gámez, X., Cairó, J. J. and Gòdia, C. Analysis of CHO cells metabolic redistribution in a glutamate-based defined medium in continuous culture. *Biotechnol. Prog.* **17,** 1032–41 (2001).

106. Mackinnon, W. B., May, G. L. and Mountford, C. E. Esterified cholesterol and triglyceride are present in plasma membranes of Chinese hamster ovary cells. *Eur. J. Biochem.* **205,** 827–839 (1992).

107. Bauer, K. D. and Dethlefsen, L. A. Total cellular RNA content: correlation between flow cytometry and ultraviolet spectroscopy. *J. Histochem. Cytochem.* **28,** 493–498 (1980).

108. McMahon, R. J. and Frost, S. C. Glycogen: a carbohydrate source for GLUT-1 glycosylation during glucose deprivation of 3T3-L1 adipocytes. *Am. J. Physiol.* **270,** E640–E645 (1996).

109. Warburg, O. On the Origin of Cancer Cells. *Science (80-.).* **123,** 309–314 (1956).

110. Vander Heiden, M. G., Cantley, L. C. and Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science (80-.).* **324,** 1029–1033 (2009).

111. Pearce, E. J. E. L. and Pearce, E. J. E. L. Metabolic Pathways in Immune Cell Activation and Quiescence. *Immunity* **38,** 633–643 (2013).

112. Abdel-Haleem, A. M., Lewis, N. E., Jamshidi, N., Mineta, K., Gao, X. and Gojobori, T. The Emerging Facets of Non-Cancerous Warburg Effect. *Front. Endocrinol. (Lausanne).* **8,** 279 (2017).

113. Shyh-Chang, N. and Daley, G. Q. Metabolic Switches Linked to Pluripotency and Embryonic Stem Cell Differentiation. *Cell Metab.* **21,** 349–350 (2015).

114. Liberti, M. V. and Locasale, J. W. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem. Sci.* **41,** 211–218 (2016).

115. Gagnon, M., Hiller, G., Luan, Y. T., Kittredge, A., DeFelice, J. and Drapeau, D. High-End pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO Fed-batch cultures. *Biotechnol. Bioeng.* **108,** 1328–1337 (2011).

116. Pacis, E., Yu, M., Autsen, J., Bayer, R. and Li, F. Effects of cell culture conditions on antibody N-linked glycosylation-what affects high mannose 5 glycoform. *Biotechnol. Bioeng.* **108,** 2348–2358 (2011).

117. Krieg, A. F., Rosenblum, L. J. and Henry, J. B. Lactate dehydrogenase isoenzymes a comparison of pyruvate-to-lactate and lactate-to-pyruvate assays. *Clin. Chem.* **13,** 196–203 (1967).

118. Blanco, A., Burgos, C., Gerez de Burgos, N. M. and Montamat, E. E. Properties of the testicular lactate dehydrogenase isoenzyme. *Biochem. J.* **153,** 165–72 (1976).

119. WROBLEWSKI, F. and GREGORY, K. F. Lactic dehydrogenase isozymes and their distribution in normal tissues and plasma and in disease states. *Ann. N. Y. Acad. Sci.* **94,** 912–32 (1961).

120. Le, A., Cooper, C. R., Gouw, A. M., Dinavahi, R., Maitra, A., Deck, L. M., Royer, R. E., Vander Jagt, D. L., Semenza, G. L. and Dang, C. V. Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression. *Proc. Natl. Acad. Sci.* **107,** 2037–2042 (2010).

121. Jeong, D. won, Kim, T. S., Lee, J. W., Kim, K. T., Kim, H. J., Kim, I. H. and Kim, I. Y. Blocking of Acidosis-Mediated Apoptosis by a Reduction of Lactate Dehydrogenase Activity through Antisense mRNA Expression. *Biochem. Biophys. Res. Commun.* **289,** 1141–1149 (2001).

122. Maekawa, M., Sudo, K., Kanno, T. and Li, S. S. Molecular characterization of genetic mutation in human lactate dehydrogenase-A (M) deficiency. *Biochem. Biophys. Res. Commun.* **168,** 677–82 (1990).

123. Kanno, T., Sudo, K., Takeuchi, I., Kanda, S., Honda, N., Nishimura, Y. and Oyama, K. Hereditary deficiency of lactate dehydrogenase M-subunit. *Clin. Chim. Acta.* **108,** 267–76 (1980).

124. Maekawa, M., Sudo, K. and Kanno, T. Immunochemical studies on lactate dehydrogenase A subunit deficiencies. *Am. J. Hum. Genet.* **39,** 232–8 (1986).

125. Kitamura, M., Iijima, N., Hashimoto, F. and Hiratsuka, A. Hereditary deficiency of subunit H of lactate dehydrogenase. *Clin. Chim. Acta.* **34,** 419–23 (1971).

126. Lee, J. S., Kallehauge, T. B., Pedersen, L. E. and Kildegaard, H. F. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.* **5,** 8572 (2015).

127. Yip, S. S., Zhou, M., Joly, J., Snedecor, B., Shen, A. and Crawford, Y. Complete knockout of the lactate dehydrogenase A gene is lethal in pyruvate dehydrogenase kinase 1, 2, 3 down-regulated CHO cells. *Mol. Biotechnol.* **56,** 833–8 (2014).

128. Chen, K., Liu, Q., Xie, L., Sharp, P. A. and Wang, D. I. Engineering of a mammalian cell line for reduction of lactate formation and high monoclonal antibody production. *Biotechnol. Bioeng.* **72,** 55–61 (2001).

129. Kim, S. H. and Lee, G. M. Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. *Appl. Microbiol. Biotechnol.* **74,** 152–9 (2007).

130. Wlaschin, K. F. and Hu, W. S. Engineering cell metabolism for high-density cell culture via manipulation of sugar transport. *J. Biotechnol.* **131,** 168–176 (2007).

131. Cairns, R. A., Harris, I. S. and Mak, T. W. Regulation of cancer cell metabolism. *Nat. Rev.*

Cancer **11,** 85–95 (2011).

132. Roche, T. E., Baker, J. C., Yan, X., Hiromasa, Y., Gong, X., Peng, T., Dong, J., Turkan, A. and Kasten, S. A. Distinct regulatory properties of pyruvate dehydrogenase kinase and phosphatase isoforms. *Prog. Nucleic Acid Res. Mol. Biol.* **70,** 33–75 (2001).

133. Xie, H., Hanai, J. ichi, Ren, J. G., Kats, L., Burgess, K., Bhargava, P., Signoretti, S., Billiard, J., Duffy, K. J., Grant, A., Wang, X., Lorkiewicz, P. K., Schatzman, S., Bousamra, M., Lane, A. N., Higashi, R. M., Fan, T. W. M., Pandolfi, P. P., Sukhatme, V. P. and Seth, P. Targeting Lactate Dehydrogenase-A Inhibits Tumorigenesis and Tumor Progression in Mouse Models of Lung Cancer and Impacts Tumor-Initiating Cells. *Cell Metab.* **19,** 795–809 (2014).

134. Calvaresi, E. C., Granchi, C., Tuccinardi, T., Di Bussolo, V., Huigens, R. W., Lee, H. Y., Palchaudhuri, R., Macchia, M., Martinelli, A., Minutolo, F. and Hergenrother, P. J. Dual Targeting of the Warburg Effect with a Glucose-Conjugated Lactate Dehydrogenase Inhibitor. *ChemBioChem* **14,** 2263–2267 (2013).

135. Epstein, T., Xu, L., Gillies, R. J. and Gatenby, R. A. Separation of metabolic supply and demand: aerobic glycolysis as a normal physiological response to fluctuating energetic demands in the membrane. *Cancer Metab.* **2,** 7 (2014).

136. Epstein, T., Gatenby, R. A. and Brown, J. S. The Warburg effect as an adaptation of cancer cells to rapid fluctuations in energy demand. *PLoS One* **12,** e0185085 (2017).

137. DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G. and Thompson, C. B. The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metab.* **7,** 11–20 (2008).

138. Lunt, S. Y. and Vander Heiden, M. G. Aerobic Glycolysis: Meeting the Metabolic Requirements of Cell Proliferation. *Annu. Rev. Cell Dev. Biol.* **27,** 441–464 (2011).

139. Yizhak, K., Le Dévédec, S. E., Rogkoti, V. M., Baenke, F., de Boer, V. C., Frezza, C., Schulze, A., van de Water, B. and Ruppin, E. A computational study of the Warburg effect identifies metabolic targets inhibiting cancer migration. *Mol. Syst. Biol.* **10,** 744 (2014).

140. Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z. and Bergmann, S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods* **13,** 366–370 (2016).

141. Grav, L. M., Lee, J. S., Gerling, S., Kallehauge, T. B., Hansen, A. H., Kol, S., Lee, G. M., Pedersen, L. E. and Kildegaard, H. F. One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. J.* **10,** 1446–1456 (2015).

142. Shen, J. P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A. N., Sanchez, K. S., Thomas, A., Kuo, C. C., Du, D., Roguev, A., Lewis, N. E., Chang, A. N., Kreisberg, J. F., Krogan, N., Qi, L., Ideker, T. and Mali, P. Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* **14,** 573–576 (2017).

143. Najm, F. J., Strand, C., Donovan, K. F., Hegde, M., Sanson, K. R., Vaimberg, E. W., Sullender, M. E., Hartenian, E., Kalani, Z., Fusi, N., Listgarten, J., Younger, S. T., Bernstein, B. E., Root, D. E. and Doench, J. G. Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* (2017). doi:10.1038/nbt.4048

144. Ronda, C., Pedersen, L. E., Hansen, H. G., Kallehauge, T. B., Betenbaugh, M. J., Nielsen, A. T. and Kildegaard, H. F. Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* **111,** 1604–16 (2014).

145. Lecina, M., Soley, A., Gràcia, J., Espunya, E., Lázaro, B., Cairó, J. J. and Gòdia, F. Application of on-line OUR measurements to detect actions points to improve baculovirus-insect cell cultures in bioreactors. *J. Biotechnol.* **125,** 385–394 (2006).

146. van Wijk, X. M., Döhrmann, S., Hallström, B. M., Li, S., Voldborg, B. G., Meng, B. X., McKee, K. K., van Kuppevelt, T. H., Yurchenco, P. D., Palsson, B. O., Lewis, N. E., Nizet, V. and Esko, J. D. Whole-Genome Sequencing of Invasion-Resistant Cells Identifies Laminin α2 as a Host Factor for Bacterial Invasion. *MBio* **8,** e02128-16 (2017).

147. Lund, A. M., Kaas, C. S., Brandl, J., Pedersen, L. E., Kildegaard, H. F., Kristensen, C. and Andersen, M. R. Network reconstruction of the mouse secretory pathway applied on CHO cell transcriptome data. *BMC Syst. Biol.* **11,** 37 (2017).

148. Kondratova, A., Watanabe, T., Marotta, M., Cannon, M., Segall, A. M., Serre, D. and Tanaka, H. Replication fork integrity and intra-S phase checkpoint suppress gene amplification. *Nucleic Acids Res.* **43,** 2678–2690 (2015).

149. Bowker-Kinley, M. and Popov, K. M. Evidence that pyruvate dehydrogenase kinase belongs to the ATPase/kinase superfamily. *Biochem. J.* **344 Pt 1,** 47–53 (1999).

150. Read, J. A., Winter, V. J., Eszes, C. M., Sessions, R. B. and Brady, R. L. Structural basis for altered activity of M- and H-isozyme forms of human lactate dehydrogenase. *Proteins* **43,** 175–85 (2001).

151. Porter, A. J., Racher, A. J., Preziosi, R. and Dickson, A. J. Strategies for selecting recombinant CHO cell lines for cGMP manufacturing: Improving the efficiency of cell line generation. *Biotechnol. Prog.* **26,** 1455–1464 (2010).

152. Hou, J. J. C., Hughes, B. S., Smede, M., Leung, K. M., Levine, K., Rigby, S., Gray, P. P. and Munro, T. P. High-throughput ClonePix FL analysis of mAb-expressing clones using the UCOE expression system. *N. Biotechnol.* **31,** 214–220 (2014).

153. Roy, G., Miro-Quesada, G., Zhuang, L., Martin, T., Zhu, J., Wu, H., Marelli, M. and Bowen, M. A. Sequential screening by ClonePix FL and intracellular staining facilitate isolation of high producer cell lines for monoclonal antibody manufacturing. *J. Immunol. Methods* **451,** 100–110 (2017).

154. Birch, J. R. and Racher, A. J. Antibody production. *Adv. Drug Deliv. Rev.* **58,** 671–685 (2006).

155. Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian

cells. *Nat. Biotechnol.* **22,** 1393–8 (2004).

156. Marth, J. D. and Grewal, P. K. Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* **8,** 874–87 (2008).

157. Chiang, A. W., Li, S., Spahn, P. N., Richelle, A., Kuo, C. C., Samoudi, M. and Lewis, N. E. Modulating carbohydrate–protein interactions through glycoengineering of monoclonal antibodies to impact cancer physiology. *Curr. Opin. Struct. Biol.* **40,** 104–111 (2016).

158. Noh, S. M., Park, J. H., Lim, M. S., Kim, J. W. and Lee, G. M. Reduction of ammonia and lactate through the coupling of glutamine synthetase selection and downregulation of lactate dehydrogenase-A in CHO cells. *Appl. Microbiol. Biotechnol.* **101,** 1035–1045 (2017).

159. Pristovšek, N., Hansen, H. G., Sergeeva, D., Borth, N., Lee, G. M., Andersen, M. R. and Kildegaard, H. F. Using titer and titer normalized to confluence are complementary strategies for obtaining Chinese hamster ovary cell lines with high volumetric productivity of etanercept. *Biotechnol. J.*

160. Derouazi, M., Martinet, D., Besuchet Schmutz, N., Flaction, R., Wicht, M., Bertschinger, M., Hacker, D. L., Beckmann, J. S. and Wurm, F. M. Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem. Biophys. Res. Commun.* **340,** 1069–1077 (2006).

161. Baik, J. Y. and Lee, K. H. A framework to quantify karyotype variation associated with CHO cell line instability at a single-cell level. *Biotechnol. Bioeng.* **114,** 1045–1053 (2017).

162. Thomann, M., Reckermann, K., Reusch, D., Prasser, J. and Tejada, M. L. Fc-galactosylation modulates antibody-dependent cellular cytotoxicity of therapeutic antibodies. *Mol. Immunol.* **73,** 69–75 (2016).

163. Hodoniczky, J., Zheng, Y. Z. and James, D. C. Control of Recombinant Monoclonal Antibody Effector Functions by Fc N-Glycan Remodeling in Vitro. *Biotechnol. Prog.* **21,** 1644–1652 (2005).

164. González-Mateos, F., Gómez, M. E., García-Salguero, L., Sánchez, V. and Aragón, J. J. Inhibition of glycolysis by amino acids in ascites tumor cells. Specificity and mechanism. *J. Biol. Chem.* **268,** 7809–17 (1993).

165. Mulukutla, B. C., Kale, J., Kalomeris, T., Jacobs, M. and Hiller, G. W. Identification and control of novel growth inhibitors in fed-batch cultures of Chinese hamster ovary cells. *Biotechnol. Bioeng.* **114,** 1779–1790 (2017).

166. Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G. and Zhang, F. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–7 (2014).

167. Wang, T., Wei, J. J., Sabatini, D. M. and Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343,** 80–4 (2014).