A Model for Designing Adaptive Laboratory Evolution Experiments

The occurrence of mutations is a cornerstone of the evolutionary theory of adaptation, capitalizing on the rare chance that a mutation confers a fitness benefit. Natural selection is increasingly being leveraged in laboratory settings for industrial and basic science applications. Despite increasing deployment, there are no standardized procedures available for designing and performing adaptive laboratory evolution (ALE) experiments. Thus, there is a need to optimize the experimental design, specifically for determining when to consider an experiment complete and for balancing outcomes with available resources (i.e., laboratory supplies, personnel, and time). To design and to better understand ALE experiments, a simulator, ALEsim, was developed, validated, and applied to the optimization of ALE experiments. The effects of various passage sizes were experimentally determined and subsequently evaluated with ALEsim, to explain differences in experimental outcomes. Furthermore, a beneficial mutation rate of $10^{-6.9}$ to $10^{-8.4}$ mutations per cell division was derived. A retrospective analysis of ALE experiments revealed that passage sizes typically employed in serial passage batch culture ALE experiments led to inefficient production and fixation of beneficial mutations. ALEsim and the results described here will aid in the design of ALE experiments to fit the exact needs of a project while taking into account the resources required and will lower the barriers to entry for this experimental technique.

IMPORTANCE ALE is a widely used scientific technique to increase scientific understanding, as well as to create industrially relevant organisms. The manner in which ALE experiments are conducted is highly manual and uniform, with little optimization for efficiency. Such inefficiencies result in suboptimal experiments that can take multiple months to complete. With the availability of automation and computer simulations, we can now perform these experiments in an optimized fashion and can design experiments to generate greater fitness in an accelerated time frame, thereby pushing the limits of what adaptive laboratory evolution can achieve.
Antibiotic-Induced Changes to the Host Metabolic Environment Inhibit Drug Efficacy and Alter Immune Function

Bactericidal antibiotics alter microbial metabolism as part of their lethality and can damage mitochondria in mammalian cells. In addition, antibiotic susceptibility is sensitive to extracellular metabolites, but it remains unknown whether metabolites present at an infection site can affect either treatment efficacy or immune function. Here, we quantify local metabolic changes in the host microenvironment following antibiotic treatment for a peritoneal Escherichia coli infection. Antibiotic treatment elicits microbiome-independent changes in local metabolites, but not those distal to the infection site, by acting directly on host cells. The metabolites induced during treatment, such as AMP, reduce antibiotic efficacy and enhance phagocytic killing. Moreover, antibiotic treatment impairs immune function by inhibiting respiratory activity in immune cells. Collectively, these results highlight the immunomodulatory potential of antibiotics and reveal the local metabolic microenvironment to be an important determinant of infection resolution.
A Padawan Programmer's Guide to Developing Software Libraries

With the rapid adoption of computational tools in the life sciences, scientists are taking on the challenge of developing their own software libraries and releasing them for public use. This trend is being accelerated by popular technologies and platforms, such as GitHub, Jupyter, R/Shiny, that make it easier to develop scientific software and by open-source licenses that make it easier to release software. But how do you build a software library that people will use? And what characteristics do the best libraries have that make them enduringly popular? Here, we provide a reference guide, based on our own experiences, for developing software libraries along with real-world examples to help provide context for scientists who are learning about these concepts for the first time. While we can only scratch the surface of these topics, we hope that this article will act as a guide for scientists who want to write great software that is built to last.

Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics

The increasing availability of metabolomics data necessitates novel methods for deeper data analysis and interpretation. We present a flux balance analysis method that allows for the computation of dynamic intracellular metabolic changes at the cellular scale through integration of time-course absolute quantitative metabolomics. This approach, termed "unsteady-state flux balance analysis" (uFBA), is applied to four cellular systems: three dynamic and one steady-state as a negative control. uFBA and FBA predictions are contrasted, and uFBA is found to be more accurate in predicting dynamic metabolic flux states for red blood cells, platelets, and Saccharomyces cerevisiae. Notably, only uFBA predicts that stored red blood cells metabolize TCA intermediates to regenerate important cofactors, such as ATP, NADH, and NADPH. These pathway usage predictions were subsequently validated through 13C isotopic labeling and metabolic flux analysis in stored red blood cells. Utilizing time-course metabolomics data, uFBA provides an accurate method to predict metabolic physiology at the cellular scale for dynamic systems.
Fast growth phenotype of E. coli K-12 from adaptive laboratory evolution does not require intracellular flux rewiring

Adaptive laboratory evolution (ALE) is a widely-used method for improving the fitness of microorganisms in selected environmental conditions. It has been applied previously to Escherichia coli K-12 MG1655 during aerobic exponential growth on glucose minimal media, a frequently used model organism and growth condition, to probe the limits of E. coli growth rate and gain insights into fast growth phenotypes. Previous studies have described up to 1.6-fold increases in growth rate following ALE, and have identified key causal genetic mutations and changes in transcriptional patterns. Here, we report for the first time intracellular metabolic fluxes for six such adaptively evolved strains, as determined by high-resolution 13C-metabolic flux analysis. Interestingly, we found that intracellular metabolic pathway usage changed very little following adaptive evolution. Instead, at the level of central carbon metabolism the faster growth was facilitated by proportional increases in glucose uptake and all intracellular rates. Of the six evolved strains studied here, only one strain showed a small degree of flux rewiring, and this was also the strain with unique genetic mutations. A comparison of fluxes with two other wild-type (unevolved) E. coli strains, BW25113 and BL21, showed that inter-strain differences are greater than differences between the parental and evolved strains. Principal component analysis highlighted that nearly all flux differences (95%) between the nine strains were captured by only two principal components. The distance between measured and flux balance analysis predicted fluxes was also investigated. It suggested a relatively wide range of similar stoichiometric optima, which opens new questions about the path-dependency of adaptive evolution.

General information

State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, ALE Technology & Software Development, Big Data 2 Knowledge, University of Delaware
Authors: Long, C. P. (Ekstern), Gonzalez, J. E. (Ekstern), Feist, A. M. (Intern), Palsson, B. O. (Intern), Antoniewicz, M. R. (Ekstern)
Pages: 100-107
Publication date: 2017
Main Research Area: Technical/natural sciences
Global transcriptional regulatory network for Escherichia coli robustly connects gene expression to transcription factor activities

Transcriptional regulatory networks (TRNs) have been studied intensely for >25 y. Yet, even for the Escherichia coli TRN—probably the best characterized TRN—several questions remain. Here, we address three questions: (i) How complete is our knowledge of the E. coli TRN; (ii) how well can we predict gene expression using this TRN; and (iii) how robust is our understanding of the TRN? First, we reconstructed a high-confidence TRN (hiTRN) consisting of 147 transcription factors (TFs) regulating 1,538 transcription units (TUs) encoding 1,764 genes. The 3,797 high-confidence regulatory interactions were collected from published, validated chromatin immunoprecipitation (ChIP) data and RegulonDB. For 21 different TF knockouts, up to 63% of the differentially expressed genes in the hiTRN were traced to the knocked-out TF through regulatory cascades. Second, we trained supervised machine learning algorithms to predict the expression of 1,364 TUs given TF activities using 441 samples. The algorithms accurately predicted condition-specific expression for 86% (1,174 of 1,364) of the TUs, while 193 TUs (14%) were predicted better than random TRNs. Third, we identified 10 regulatory modules whose definitions were robust against changes to the TRN or expression compendium. Using surrogate variable analysis, we also identified three unmodeled factors that systematically influenced gene expression. Our computational workflow comprehensively characterizes the predictive capabilities and systems-level functions of an organism's TRN from disparate data types.
To the Editor:

Extracting knowledge from the many types of big data produced by high-throughput methods remains a challenge, even when data are from Escherichia coli, the best characterized bacterial species. Here, we present iML1515, the most complete genome-scale reconstruction of the metabolic network in E. coli K-12 MG1655 to date, and we demonstrate how it can be used to address this challenge. Enabling analysis of several data types, including transcriptomes, proteomes, and metabolomes, iML1515 accounts for 1,515 open reading frames and 2,719 metabolic reactions involving 1,192 unique metabolites. The iML1515 knowledgebase is linked to 1,515 protein structures to provide an integrated modeling framework bridging systems and structural biology. We apply iML1515 to build metabolic models of E. coli human gut microbiome strains from metagenomic sequencing data. We then use iML1515 to build metabolic models for E. coli clinical isolates and predict their metabolic capabilities. Finally, we use iML1515 to carry out a comparative structural proteome analysis of 1,122 E. coli strains and identify multi-strain sequence variations.

**iML1515, a knowledgebase that computes Escherichia coli traits**

General information

State: Published

Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, ALE Technology & Software Development, Big Data 2 Knowledge, University of California at San Diego, Nara Institute of Science and Technology


Pages: 904-908

Publication date: 2017

Main Research Area: Technical/natural sciences
Laboratory Evolution to Alternating Substrate Environments Yields Distinct Phenotypic and Genetic Adaptive Strategies

Adaptive laboratory evolution (ALE) experiments are often designed to maintain a static culturing environment to minimize confounding variables that could influence the adaptive process, but dynamic nutrient conditions occur frequently in natural and bioprocessing settings. To study the nature of carbon substrate fitness tradeoffs, we evolved batch cultures of Escherichia coli via serial propagation into tubes alternating between glucose and either xylose, glycerol, or acetate. Genome sequencing of evolved cultures revealed several genetic changes preferentially selected for under dynamic conditions and different adaptation strategies depending on the substrates being switched between; in some environments, a persistent "generalist" strain developed, while in another, two "specialist" subpopulations arose that alternated dominance. Diauxic lag phenotype varied across the generalists and specialists, in one case being completely abolished, while gene expression data distinguished the transcriptional strategies implemented by strains in pursuit of growth optimality. Genome-scale metabolic modeling techniques were then used to help explain the inherent substrate differences giving rise to the observed distinct adaptive strategies. This study gives insight into the population dynamics of adaptation in an alternating environment and into the underlying metabolic and genetic mechanisms. Furthermore, ALE-generated optimized strains have phenotypes with potential industrial bioprocessing applications.

IMPORTANCE Evolution and natural selection inexorably lead to an organism's improved fitness in a given environment, whether in a laboratory or natural setting. However, despite the frequent natural occurrence of complex and dynamic growth environments, laboratory evolution experiments typically maintain simple, static culturing environments so as to reduce selection pressure complexity. In this study, we investigated the adaptive strategies underlying evolution to fluctuating environments by evolving Escherichia coli to conditions of frequently switching growth substrate. Characterization of evolved strains via a number of different data types revealed the various genetic and phenotypic changes implemented in pursuit of growth optimality and how these differed across the different growth substrates and switching protocols. This work not only helps to establish general principles of adaptation to complex environments but also suggests strategies for experimental design to achieve desired evolutionary outcomes.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, ALE Technology & Software Development, University of California, San Diego, University of Tartu
Authors: Sandberg, T. E. (Ekstern), Lloyd, C. J. (Ekstern), Palsson, B. O. (Intern), Feist, A. M. (Intern), Kivisaar, M. (ed.) (Ekstern)
Number of pages: 15
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Applied and Environmental Microbiology
Volume: 83
Issue number: 13
Article number: e00410-17
ISSN (Print): 0099-2240
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 4.08
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 1.891 SNIP 1.308 CiteScore 4.14
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 1.857 SNIP 1.384 CiteScore 4.02
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 1.899 SNIP 1.414 CiteScore 4.25
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
Literature mining supports a next-generation modeling approach to predict cellular byproduct secretion

The metabolic byproducts secreted by growing cells can be easily measured and provide a window into the state of a cell; they have been essential to the development of microbiology, cancer biology, and biotechnology. Progress in computational modeling of cells has made it possible to predict metabolic byproduct secretion with bottom-up reconstructions of metabolic networks. However, owing to a lack of data, it has not been possible to validate these predictions across a wide range of strains and conditions. Through literature mining, we were able to generate a database of Escherichia coli strains and their experimentally measured byproduct secretions. We simulated these strains in six historical genome-scale models of E. coli, and we report that the predictive power of the models has increased as they have expanded in size and scope. The latest genome-scale model of metabolism correctly predicts byproduct secretion for 35/89 (39%) of designs. The next-generation genome-scale model of metabolism and gene expression (ME-model) correctly predicts byproduct secretion for 40/89 (45%) of designs, and we show that ME-model predictions could be further improved through kinetic parameterization. We analyze the failure modes of these simulations and discuss opportunities to improve prediction of byproduct secretion.
Machine learning in computational biology to accelerate high-throughput protein expression

Motivation: The Human Protein Atlas (HPA) enables the simultaneous characterization of thousands of proteins across various tissues to pinpoint their spatial location in the human body. This has been achieved through transcriptomics and high-throughput immunohistochemistry-based approaches, where over 40,000 unique human protein fragments have been expressed in E. coli. These datasets enable quantitative tracking of entire cellular proteomes and present new avenues for understanding molecular-level properties influencing expression and solubility.

Results: Combining computational biology and machine learning identifies protein properties that hinder the HPA high-throughput antibody production pipeline. We predict protein expression and solubility with accuracies of 70% and 80%, respectively, based on a subset of key properties (aromaticity, hydropathy and isoelectric point). We guide the selection of protein fragments based on these characteristics to optimize high-throughput experimentation.

Availability and implementation: We present the machine learning workflow as a series of IPython notebooks hosted on GitHub (https://github.com/SBRG/Protein_ML). The workflow can be used as a template for analysis of further expression and solubility datasets.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, High Throughput Molecular Bioscience, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, University of California, San Diego, KTH - Royal Institute of Technology, Technical University of Denmark
Authors: Sastry, A. (Ekstern), Monk, J. M. (Ekstern), Tegel, H. (Ekstern), Uhlén, M. (Intern), Palsson, B. (Intern), Rockberg, J. (Ekstern), Brunk, E. (Ekstern)
Pages: 2487-2495
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Bioinformatics
Volume: 33
Issue number: 16
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 6.42
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 6.06
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 5.5
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 5.78
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): CiteScore 6.73
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): CiteScore 5.61
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 2
Web of Science (2008): Indexed yes
Web of Science (2007): Indexed yes
Web of Science (2006): Indexed yes
Web of Science (2005): Indexed yes
Web of Science (2004): Indexed yes
Web of Science (2003): Indexed yes
Web of Science (2002): Indexed yes
Web of Science (2001): Indexed yes
Web of Science (2000): Indexed yes
Original language: English
DOIs: 10.1093/bioinformatics/btx207
Source: FindIt
Source-ID: 2356534600
Publication: Research - peer-review › Journal article – Annual report year: 2017

Reliable and efficient solution of genome-scale models of Metabolism and macromolecular Expression
Constraint-Based Reconstruction and Analysis (COBRA) is currently the only methodology that permits integrated modeling of Metabolism and macromolecular Expression (ME) at genome-scale. Linear optimization computes steady-state flux solutions to ME models, but flux values are spread over many orders of magnitude. Data values also have greatly varying magnitudes. Standard double-precision solvers may return inaccurate solutions or report that no solution exists. Exact simplex solvers based on rational arithmetic require a near-optimal warm start to be practical on large problems (current ME models have 70,000 constraints and variables and will grow larger). We have developed a quadruple-precision version of our linear and nonlinear optimizer MINOS, and a solution procedure (DQQ) involving Double and Quad MINOS that achieves reliability and efficiency for ME models and other challenging problems tested here. DQQ will enable extensive use of large linear and nonlinear models in systems biology and other applications involving multiscale data.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, Stanford University, University of California, University of Luxembourg
Authors: Ma, D. (Ekstern), Yang, L. (Ekstern), Fleming, R. M. T. (Ekstern), Thiele, I. (Ekstern), Palsson, B. (Intern), Saunders, M. A. (Ekstern)
Number of pages: 11
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Scientific Reports
Volume: 7
Article number: 40863
ISSN (Print): 2045-2322
Ratings:
BFI (2018): BFI-level 1
BFI (2017): BFI-level 1
Revealing genome-scale transcriptional regulatory landscape of OmpR highlights its expanded regulatory roles under osmotic stress in Escherichia coli K-12 MG1655

A transcription factor (TF), OmpR, plays a critical role in transcriptional regulation of the osmotic stress response in bacteria. Here, we reveal a genome-scale OmpR regulon in Escherichia coli K-12 MG1655. Integrative data analysis reveals that a total of 37 genes in 24 transcription units (TUs) belong to OmpR regulon. Among them, 26 genes show more than two-fold changes in expression level in an OmpR knock-out strain. Specifically, we find that: 1) OmpR regulates mostly membrane-located gene products involved in diverse fundamental biological processes, such as narU (encoding nitrate/nitrite transporter), ompX (encoding outer membrane protein X), and nuoN (encoding NADH: ubiquinone oxidoreductase); 2) by investigating co-regulation of entire sets of genes regulated by other stressresponse TFs, stresses are surprisingly independently regulated among each other; and, 3) a detailed investigation of the physiological roles of the newly discovered OmpR regulon genes reveals that activation of narU represents a novel strategy to significantly improve osmotic stress tolerance of E. coli. Thus, the genome-scale approach to elucidating regulons comprehensively identifies regulated genes and leads to fundamental discoveries related to stress responses.
The aldehyde dehydrogenase, AldA, is essential for L-1,2-propanediol utilization in laboratory-evolved Escherichia coli

Most Escherichia coli strains are naturally unable to grow on 1,2-propanediol (PDO) as a sole carbon source. Recently, however, a K-12 descendent E. coli strain was evolved to grow on 1,2-PDO, and it was hypothesized that this evolved ability was dependent on the aldehyde dehydrogenase, AldA, which is highly conserved among members of the family Enterobacteriaceae. To test this hypothesis, we first performed computational model simulation, which confirmed the essentiality of the aldA gene for 1,2-PDO utilization by the evolved PDO-degrading E. coli. Next, we deleted the aldA gene from the evolved strain, and this deletion was sufficient to abolish the evolved phenotype. On re-introducing the gene on a plasmid, the evolved phenotype was restored. These findings provide experimental evidence for the computationally predicted role of AldA in 1,2-PDO utilization, and represent a good example of E. coli robustness, demonstrated by the bacterial deployment of a generalist enzyme (here AldA) in multiple pathways to survive carbon starvation and to grow on a non-native substrate when no native carbon source is available.
Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation

Maintenance of a properly folded proteome is critical for bacterial survival at notably different growth temperatures. Understanding the molecular basis of thermoadaptation has progressed in two main directions, the sequence and structural basis of protein thermostability and the mechanistic principles of protein quality control assisted by chaperones. Yet we do not fully understand how structural integrity of the entire proteome is maintained under stress and how it affects cellular fitness. To address this challenge, we reconstruct a genome-scale protein-folding network for Escherichia coli and formulate a computational model, FoldME, that provides statistical descriptions of multiscale cellular response consistent with many datasets. FoldME simulations show (i) that the chaperones act as a system when they respond to unfolding stress rather than achieving efficient folding of any single component of the proteome, (ii) how the proteome is globally balanced between chaperones for folding and the complex machinery synthesizing the proteins in response to perturbation, (iii) how this balancing determines growth rate dependence on temperature and is achieved through nonspecific regulation, and (iv) how thermal instability of the individual protein affects the overall functional state of the proteome. Overall, these results expand our view of cellular regulation, from targeted specific control mechanisms to global regulation through a web of nonspecific competing interactions that modulate the optimal reallocation of cellular resources. The methodology developed in this study enables genome-scale integration of environment-dependent protein properties and a proteome-wide study of cellular stress responses.
Whole-Genome Sequencing of Invasion-Resistant Cells Identifies Laminin α2 as a Host Factor for Bacterial Invasion

To understand the role of glycosaminoglycans in bacterial cellular invasion, xylosyltransferase-deficient mutants of Chinese hamster ovary (CHO) cells were created using clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated gene 9 (CRISPR-cas9) gene targeting. When these mutants were compared to the pgsA745 cell line, a CHO xylosyltransferase mutant generated previously using chemical mutagenesis, an unexpected result was obtained. Bacterial invasion of pgsA745 cells by group B Streptococcus (GBS), group A Streptococcus, and Staphylococcus aureus was markedly reduced compared to the invasion of wild-type cells, but newly generated CRISPR-cas9 mutants were only resistant to GBS. Invasion of pgsA745 cells was not restored by transfection with xylosyltransferase, suggesting that an additional mutation conferring panresistance to multiple bacteria was present in pgsA745 cells. Whole-genome sequencing and transcriptome sequencing (RNA-Seq) uncovered a deletion in the gene encoding the laminin subunit α2 (Lama2) that eliminated much of domain L4a. Silencing of the long Lama2 isoform in wild-type cells strongly reduced bacterial invasion, whereas transfection with human LAMA2 cDNA significantly enhanced invasion in pgsA745 cells. The addition of exogenous laminin-a2β1γ1/laminin-a2β2γ1 strongly increased bacterial invasion in CHO cells, as well as in human alveolar basal epithelial and human brain microvascular endothelial cells. Thus, the L4a domain in laminin a2 is important for cellular invasion by a number of bacterial pathogens.

General information

State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, CHO Core, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, University of California, San Diego, Technical University of Denmark, Robert Wood Johnson Medical School, Radboud University Nijmegen
Authors: van Wijk, X. M. (Ekstern), Döhrmann, S. (Ekstern), Hallstrom, B. (Ekstern), Li, S. (Ekstern), Voldborg, B. G. (Intern), Meng, B. X. (Ekstern), McKee, K. K. (Ekstern), van Kuppevelt, T. H. (Ekstern), Yurchenco, P. D. (Ekstern), Palsson, B. (Intern), Lewis, N. (Ekstern), Nizet, V. (Ekstern), Esko, J. D. (Ekstern)
Number of pages: 11
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: mBio
Volume: 8
Issue number: 1
Article number: e02128-16
ISSN (Print): 2161-2129
Ratings:
BFI (2018): BFI-level 1
BFI (2017): BFI-level 1
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 5.79
BFI (2015): BFI-level 1
Scopus rating (2015): CiteScore 4.93
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): CiteScore 4.23
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Why microbes will rule the world – and our industries

Microbes have ruled the world for approximately 4 billion years. But the future actually depends on their dominance, some would argue. Why? Because microbes, as well as mammalian cells, can be engineered into producing high-value chemicals and medicine. Therefore, scientists at The Novo Nordisk Foundation Center for Biosustainability are hard at work developing cell factories to benefit us all.

A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism

Chinese hamster ovary (CHO) cells dominate biotherapeutic protein production and are widely used in mammalian cell line engineering research. To elucidate metabolic bottlenecks in protein production and to guide cell engineering and bioprocess optimization, we reconstructed the metabolic pathways in CHO and associated them with >1,700 genes in the *Cricetulus griseus* genome. The genome-scale metabolic model based on this reconstruction, iCHO1766, and cell-line-specific models for CHO-K1, CHO-S, and CHO-DG44 cells provide the biochemical basis of growth and recombinant protein production. The models accurately predict growth phenotypes and known auxotrophies in CHO cells. With the models, we quantify the protein synthesis capacity of CHO cells and demonstrate that common bioprocess treatments, such as histone deacetylase inhibitors, inefficiently increase product yield. However, our simulations show that the metabolic resources in CHO are more than three times more efficiently utilized for growth or recombinant protein synthesis following targeted efforts to engineer the CHO secretory pathway. This model will further accelerate CHO cell engineering and help optimize bioprocesses.
BiGG Models: A platform for integrating, standardizing and sharing genome-scale models

BiGG Models contain more than 75 high-quality, manually-curated genome-scale metabolic models. On the website, users can browse, search and visualize models. BiGG Models connects genome-scale models to genome annotations and external databases. Reaction and metabolite identifiers have been standardized across models to conform to community standards and enable rapid comparison across models. Furthermore, BiGG Models provides a comprehensive application programming interface for accessing BiGG Models with modeling and analysis tools. As a resource for highly curated, standardized and accessible models of metabolism, BiGG Models will facilitate diverse systems biology studies and support knowledge-based analysis of diverse experimental data.
Characterizing Strain Variation in Engineered E. coli Using a Multi-Omics-Based Workflow

Understanding the complex interactions that occur between heterologous and native biochemical pathways represents a major challenge in metabolic engineering and synthetic biology. We present a workflow that integrates metabolomics, proteomics, and genome-scale models of Escherichia coli metabolism to study the effects of introducing a heterologous pathway into a microbial host. This workflow incorporates complementary approaches from computational systems biology, metabolic engineering, and synthetic biology; provides molecular insight into how the host organism microenvironment changes due to pathway engineering; and demonstrates how biological mechanisms underlying strain variation can be exploited as an engineering strategy to increase product yield. As a proof of concept, we present the analysis of eight engineered strains producing three biofuels: isopentenol, limonene, and bisabolene. Application of this workflow identified the roles of candidate genes, pathways, and biochemical reactions in observed experimental phenomena and facilitated the construction of a mutant strain with improved productivity. The contributed workflow is available as an open-source tool in the form of iPython notebooks.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, Joint Bioenergy Institute, University of California
Number of pages: 13
Pages: 335-346
Publication date: 2016
Main Research Area: Technical/natural sciences

Evaluation of rate law approximations in bottom-up kinetic models of metabolism
Background: The mechanistic description of enzyme kinetics in a dynamic model of metabolism requires specifying the numerical values of a large number of kinetic parameters. The parameterization challenge is often addressed through the use of simplifying approximations to form reaction rate laws with reduced numbers of parameters. Whether such simplified models can reproduce dynamic characteristics of the full system is an important question. Results: In this work, we compared the local transient response properties of dynamic models constructed using rate laws with varying levels of approximation. These approximate rate laws were: 1) a Michaelis-Menten rate law with measured enzyme parameters, 2) a Michaelis-Menten rate law with approximated parameters, using the convenience kinetics convention, 3) a thermodynamic rate law resulting from a metabolite saturation assumption, and 4) a pure chemical reaction mass action rate law that removes the role of the enzyme from the reaction kinetics. We utilized in vivo data for the human red blood cell to compare the effect of rate law choices against the backdrop of physiological flux and concentration differences. We found that the Michaelis-Menten rate law with measured enzyme parameters yields an excellent approximation of the full system dynamics, while other assumptions cause greater discrepancies in system dynamic behavior. However, iteratively replacing mechanistic rate laws with approximations resulted in a model that retains a high correlation with the true model behavior. Investigating this consistency, we determined that the order of magnitude differences among fluxes and concentrations in the network were greatly influential on the network dynamics. We further identified reaction features such as thermodynamic reversibility, high substrate concentration, and lack of allosteric regulation, which make certain reactions more suitable for rate law approximations. Conclusions: Overall, our work generally supports the use of approximate rate laws when building large scale kinetic models, due to the key role that physiologically meaningful flux and concentration ranges play in determining network dynamics. However, we also showed that detailed mechanistic models show a clear benefit in prediction accuracy when data is available. The work here should help to provide guidance.
to future kinetic modeling efforts on the choice of rate law and parameterization approaches.

**General information**

State: Published

Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California

Authors: Du, B. (Ekstern), Zielinski, D. C. (Ekstern), Kavvas, E. S. (Ekstern), Draeger, A. (Ekstern), Tan, J. (Ekstern), Zhang, Z. (Ekstern), Ruggiero, K. E. (Ekstern), Arzumanyan, G. A. (Ekstern), Palsson, B. (Intern)

Number of pages: 15

Publication date: 2016

Main Research Area: Technical/natural sciences

**Publication information**

Journal: BMC Systems Biology

Volume: 10

Article number: 40

ISSN (Print): 1752-0509

Ratings:

- BFI (2018): BFI-level 1
- BFI (2017): BFI-level 1
- Web of Science (2017): Indexed yes
- BFI (2016): BFI-level 1
- Scopus rating (2016): SJR 1.444 SNIP 0.735 CiteScore 2.11
- Web of Science (2016): Indexed yes
- BFI (2015): BFI-level 1
- Scopus rating (2015): SJR 1.47 SNIP 0.912 CiteScore 2.54
- Web of Science (2015): Indexed yes
- BFI (2014): BFI-level 1
- Scopus rating (2014): SJR 1.356 SNIP 0.893 CiteScore 2.55
- Web of Science (2014): Indexed yes
- BFI (2013): BFI-level 1
- Scopus rating (2013): SJR 1.832 SNIP 0.999 CiteScore 3.42
- ISI indexed (2013): ISI indexed yes
- Web of Science (2013): Indexed yes
- BFI (2012): BFI-level 1
- Scopus rating (2012): SJR 1.563 SNIP 1.001 CiteScore 3.32
- ISI indexed (2012): ISI indexed yes
- BFI (2011): BFI-level 1
- Scopus rating (2011): SJR 1.505 SNIP 0.98 CiteScore 3.45
- ISI indexed (2011): ISI indexed yes
- BFI (2010): BFI-level 1
- Scopus rating (2010): SJR 1.816 SNIP 1.061
- Web of Science (2010): Indexed yes
- BFI (2009): BFI-level 1
- Scopus rating (2009): SJR 1.489 SNIP 0.868
- Web of Science (2009): Indexed yes
- BFI (2008): BFI-level 1
- Scopus rating (2008): SJR 1.063 SNIP 0.763
- Web of Science (2008): Indexed yes
- Web of Science (2007): Indexed yes

Original language: English

Electronic versions:


**Bibliographical note**
Evolution of E. coli on [U-13C] Glucose Reveals a Negligible Isotopic Influence on Metabolism and Physiology

13C-Metabolic flux analysis (13C-MFA) traditionally assumes that kinetic isotope effects from isotopically labeled compounds do not appreciably alter cellular growth or metabolism, despite indications that some biochemical reactions can be non-negligibly impacted. Here, populations of Escherichia coli were adaptively evolved for similar to 1000 generations on uniformly labeled 13C-glucose, a commonly used isotope for 13C-MFA. Phenotypic characterization of these evolved strains revealed -40% increases in growth rate, with no significant difference in fitness when grown on either labeled (13C) or unlabeled (12C) glucose. The evolved strains displayed decreased biomass yields, increased glucose and oxygen uptake, and increased acetate production, mimicking what is observed after adaptive evolution on unlabeled glucose. Furthermore, full genome re-sequencing revealed that the key genetic changes underlying these phenotypic alterations were essentially the same as those acquired during adaptive evolution on unlabeled glucose. Additionally, glucose competition experiments demonstrated that the wild-type exhibits no isotopic preference for unlabeled glucose, and the evolved strains have no preference for labeled glucose. Overall, the results of this study indicate that there are no significant differences between 12C and 13C-glucose as a carbon source for E. coli growth.
Global Rebalancing of Cellular Resources by Pleiotropic Point Mutations Illustrates a Multi-scale Mechanism of Adaptive Evolution

Pleiotropic regulatory mutations affect diverse cellular processes, posing a challenge to our understanding of genotype-phenotype relationships across multiple biological scales. Adaptive laboratory evolution (ALE) allows for such mutations to be found and characterized in the context of clear selection pressures. Here, several ALE-selected single-mutation variants in RNA polymerase (RNAP) of Escherichia coli are detailed using an integrated multi-scale experimental and computational approach. While these mutations increase cellular growth rates in steady environments, they reduce tolerance to stress and environmental fluctuations. We detail structural changes in the RNAP that rewire the transcriptional machinery to rebalance proteome and energy allocation toward growth and away from several hedging and stress functions. We find that while these mutations occur in diverse locations in the RNAP, they share a common adaptive mechanism. In turn, these findings highlight the resource allocation trade-offs organisms face and suggest how the structure of the regulatory network enhances evolvability.
MID Max: LC–MS/MS Method for Measuring the Precursor and Product Mass Isotopomer Distributions of Metabolic Intermediates and Cofactors for Metabolic Flux Analysis Applications

The analytical challenges to acquire accurate isotopic data of intracellular metabolic intermediates for stationary, nonstationary, and dynamic metabolic flux analysis (MFA) are numerous. This work presents MID Max, a novel LC–MS/MS workflow, acquisition, and isotopomer deconvolution method for MFA that takes advantage of additional scan types that maximizes the number of mass isotopomer distributions (MIDs) that can be acquired in a given experiment. The analytical method was found to measure the MIDs of 97 metabolites, corresponding to 74 unique metabolite-fragment pairs (32 precursor spectra and 42 product spectra) with accuracy and precision. The compounds measured included metabolic intermediates in central carbohydrate metabolism and cofactors of peripheral metabolism (e.g., ATP). Using
only a subset of the acquired MIDs, the method was found to improve the precision of flux estimations and number of resolved exchange fluxes for wild-type *E. coli* compared to traditional methods and previously published data sets.

**General information**

*State:* Published  
*Organisations:* Novo Nordisk Foundation Center for Biosustainability, University of California, Vanderbilt University  
*Authors:* McCloskey, D. (Ekstern), Young, J. D. (Ekstern), Xu, S. (Ekstern), Palsson, B. (Intern), Feist, A. (Intern)  
*Number of pages:* 9  
*Pages:* 1362–1370  
*Publication date:* 2016  
*Main Research Area:* Technical/natural sciences

**Publication information**

*Journal:* Analytical Chemistry  
*Volume:* 88  
*Issue number:* 2  
*ISSN (Print):* 0003-2700  
*Ratings:*  
*Scopus rating (2016):* CiteScore 6.08  
*Scopus rating (2015):* CiteScore 6  
*Scopus rating (2014):* CiteScore 5.79  
*Scopus rating (2013):* CiteScore 6.01  
*ISI indexed (2013):* ISI indexed no  
*Scopus rating (2012):* CiteScore 5.8  
*ISI indexed (2012):* ISI indexed no  
*Scopus rating (2011):* CiteScore 5.86  
*Original language:* English  
*DOIs:* 10.1021/acs.analchem.5b03887  
*Source:* FindIt  
*Source-ID:* 2289769382  
*Publication:* Research - peer-review › Journal article – Annual report year: 2016

**Modeling Method for Increased Precision and Scope of Directly Measurable Fluxes at a Genome-Scale**

Metabolic flux analysis (MFA) is considered to be the gold standard for determining the intracellular flux distribution of biological systems. The majority of work using MFA has been limited to core models of metabolism due to challenges in implementing genome-scale MFA and the undesirable trade-off between increased scope and decreased precision in flux estimations. This work presents a tunable workflow for expanding the scope of MFA to the genome-scale without trade-offs in flux precision. The genome-scale MFA model presented here, iDM2014, accounts for 537 net reactions, which includes the core pathways of traditional MFA models and also covers the additional pathways of purine, pyrimidine, isoprenoid, methionine, riboflavin, coenzyme A, and folate, as well as other biosynthetic pathways. When evaluating the iDM2014 using a set of measured intracellular intermediate and cofactor mass isotopomer distributions (MIDs),(1) it was found that a total of 232 net fluxes of central and peripheral metabolism could be resolved in the *E. coli* network. The increase in scope was shown to cover the full biosynthetic route to an expanded set of bioproduction pathways, which should facilitate applications such as the design of more complex bioprocessing strains and aid in identifying new antimicrobials. Importantly, it was found that there was no loss in precision of core fluxes when compared to a traditional core model, and additionally there was an overall increase in precision when considering all observable reactions.

**General information**

*State:* Published  
*Organisations:* Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California, Vanderbilt University  
*Authors:* McCloskey, D. (Ekstern), Young, J. D. (Ekstern), Xu, S. (Ekstern), Palsson, B. (Intern), Feist, A. (Intern)  
*Number of pages:* 9  
*Pages:* 3844-3852  
*Publication date:* 2016  
*Main Research Area:* Technical/natural sciences

**Publication information**

*Journal:* Analytical Chemistry  
*Volume:* 88  
*Issue number:* 7  
*ISSN (Print):* 0003-2700  
*Ratings:*
Multi-omic data integration enables discovery of hidden biological regularities

Rapid growth in size and complexity of biological data sets has led to the 'Big Data to Knowledge' challenge. We develop advanced data integration methods for multi-level analysis of genomic, transcriptomic, ribosomal profiling, proteomic and fluxomic data. First, we show that pairwise integration of primary omics data reveals regularities that tie cellular processes together in *Escherichia coli*: the number of protein molecules made per mRNA transcript and the number of ribosomes required per translated protein molecule. Second, we show that genome-scale models, based on genomic and bibliomic data, enable quantitative synchronization of disparate data types. Integrating omics data with models enabled the discovery of two novel regularities: condition invariant *in vivo* turnover rates of enzymes and the correlation of protein structural motifs and translational pausing. These regularities can be formally represented in a computable format allowing for coherent interpretation and prediction of fitness and selection that underlies cellular physiology.

General information

State: Published
Multi-omics Quantification of Species Variation of *Escherichia coli* Links Molecular Features with Strain Phenotypes

*Escherichia coli* strains are widely used in academic research and biotechnology. New technologies for quantifying strain-specific differences and their underlying contributing factors promise greater understanding of how these differences significantly impact physiology, synthetic biology, metabolic engineering, and process design. Here, we quantified strain-specific differences in seven widely used strains of *E. coli* (BL21, C, Crooks, DH5a, K-12 MG1655, K-12 W3110, and W) using genomics, phenomics, transcriptomics, and genome-scale modeling. Metabolic physiology and gene expression varied widely with downstream implications for productivity, product yield, and titer. These differences could be linked to differential regulatory structure. Analyzing high-flux reactions and expression of encoding genes resulted in a correlated
and quantitative link between these sets, with strain-specific caveats. Integrated modeling revealed that certain strains are better suited to produce given compounds or express desired constructs considering native expression states of pathways that enable high-production phenotypes. This study yields a framework for quantitatively comparing strains in a species with implications for strain selection.

**General information**

State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, iLoop, Network Reconstruction in Silico Biology, Big Data 2 Knowledge, Research Groups, University of California
Number of pages: 26
Pages: 238-251
Publication date: 2016
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Cell Systems
Volume: 3
Issue number: 3
ISSN (Print): 2405-4712
Ratings:
Web of Science (2017): Indexed yes
Scopus rating (2016): CiteScore 4.31
Original language: English
Electronic versions:
BSOG_nihms819693.pdf
DoIs: 10.1016/j.cels.2016.08.013
Source: FindIt
Source-ID: 2345699171
Publication: Research - peer-review › Journal article – Annual report year: 2016

**Principles of proteome allocation are revealed using proteomic data and genome-scale models**

Integrating omics data to refine or make context-specific models is an active field of constraint-based modeling. Proteomics now cover over 95% of the Escherichia coli proteome by mass. Genome-scale models of Metabolism and macromolecular Expression (ME) compute proteome allocation linked to metabolism and fitness. Using proteomics data, we formulated allocation constraints for key proteome sectors in the ME model. The resulting calibrated model effectively computed the “generalist” (wild-type) E. coli proteome and phenotype across diverse growth environments. Across 15 growth conditions, prediction errors for growth rate and metabolic fluxes were 69% and 14% lower, respectively. The sector-constrained ME model thus represents a generalist ME model reflecting both growth rate maximization and “hedging” against uncertain environments and stresses, as indicated by significant enrichment of these sectors for the general stress response sigma factor sigma(S). Finally, the sector constraints represent a general formalism for integrating omics data from any experimental condition into constraint-based ME models. The constraints can be fine-grained (individual proteins) or coarse-grained (functionally related protein groups) as demonstrated here. This flexible formalism provides an accessible approach for narrowing the gap between the complexity captured by omics data and governing principles of proteome allocation described by systems-level models.

**General information**

State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, University of California, San Diego, Stanford University
Authors: Yang, L. (Ekstern), Yurkovich, J. T. (Ekstern), Lloyd, C. J. (Ekstern), Ebrahim, A. (Ekstern), Saunders, M. A. (Ekstern), Palsson, B. (Intern)
Number of pages: 8
Publication date: 2016
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Scientific Reports
Volume: 6
Article number: 36734
ISSN (Print): 2045-2322
Ratings:
BFI (2018): BFI-level 1
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 4.63 SJR 1.625 SNIP 1.401
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 2.057 SNIP 1.684 CiteScore 5.3
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 2.103 SNIP 1.544 CiteScore 4.75
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.886 SNIP 1.51 CiteScore 4.06
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.458 SNIP 0.896 CiteScore 2.44
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
ISI indexed (2011): ISI indexed no
Original language: English
Electronic versions:
Principles_of_proteome_allocation.pdf
DOIs:
10.1038/srep36734
Source: FindIt
Source-ID: 2349023007
Publication: Research - peer-review › Journal article – Annual report year: 2016

Quantification and Classification of *E. coli* Proteome Utilization and Unused Protein Costs across Environments
The costs and benefits of protein expression are balanced through evolution. Expression of un-utilized protein (that have no benefits in the current environment) incurs a quantifiable fitness costs on cellular growth rates; however, the magnitude and variability of un-utilized protein expression in natural settings is unknown, largely due to the challenge in determining environment-specific proteome utilization. We address this challenge using absolute and global proteomics data combined with a recently developed genome-scale model of *Escherichia coli* that computes the environment-specific cost and utility of the proteome on a per gene basis. We show that nearly half of the proteome mass is unused in certain environments and accounting for the cost of this unused protein expression explains >95% of the variance in growth rates of *Escherichia coli* across 16 distinct environments. Furthermore, reduction in unused protein expression is shown to be a common mechanism to increase cellular growth rates in adaptive evolution experiments. Classification of the unused protein reveals that the unused protein encodes several nutrient- and stress- preparedness functions, which may convey fitness benefits in varying environments. Thus, unused protein expression is the source of large and pervasive fitness costs that may provide the benefit of hedging against environmental change.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California
Authors: O'Brien, E. J. (Ekstern), Utrilla, J. (Ekstern), Palsson, B. (Intern)
Number of pages: 22
Publication date: 2016
Main Research Area: Technical/natural sciences

Publication information
Journal: PLoS Computational Biology (Online)
Volume: 12
Issue number: 6
Article number: e1004998
ISSN (Print): 1553-7358
Ratings:
BFI (2018): BFI-level 1
Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow

The scientific literature concerning Chinese hamster ovary (CHO) cells grows annually due to the importance of CHO cells in industrial bioprocessing of therapeutics. In an effort to start to catalogue the breadth of CHO phenotypes, or phenome, we present the CHO bibliome. This bibliographic compilation covers all published CHO cell studies from 1995 to 2015, and each study is classified by the types of phenotypic and bioprocess data contained therein. Using data from selected studies, we also present a quantitative meta-analysis of bioprocess characteristics across diverse culture conditions, yielding novel insights and addressing the validity of long held assumptions. Specifically, we show that bioprocess titers can be predicted using indicator variables derived from viable cell density, viability, and culture duration. We further identified a positive correlation between the cumulative viable cell density (VCD) and final titer, irrespective of cell line, media, and other bioprocess parameters. In addition, growth rate was negatively correlated with performance attributes, such as VCD and titer. In summary, despite assumptions that technical diversity among studies and opaque publication practices can limit research re-use in this field, we show that the statistical analysis of diverse legacy bioprocess data can provide insight into bioprocessing capabilities of CHO cell lines used in industry. The CHO bibliome can be accessed at
solveME: fast and reliable solution of nonlinear ME models

Background: Genome-scale models of metabolism and macromolecular expression (ME) significantly expand the scope and predictive capabilities of constraint-based modeling. ME models present considerable computational challenges: they are much (>30 times) larger than corresponding metabolic reconstructions (M models), are multiscale, and growth maximization is a nonlinear programming (NLP) problem, mainly due to macromolecule dilution constraints. Results: Here, we address these computational challenges. We develop a fast and numerically reliable solution method for growth maximization in ME models using a quad-precision NLP solver (Quad MINOS). Our method was up to 45 % faster than binary search for six significant digits in growth rate. We also develop a fast, quad-precision flux variability analysis that is accelerated (up to 60× speedup) via solver warm-starts. Finally, we employ the tools developed to investigate growth-coupled succinate overproduction, accounting for proteome constraints. Conclusions: Just as genome-scale metabolic reconstructions have become an invaluable tool for computational and systems biologists, we anticipate that these fast and numerically reliable ME solution methods will accelerate the wide-spread adoption of ME models for researchers in these fields. Electronic supplementary material The online version of this article (doi:10.1186/s12859-016-1240-1) contains supplementary material, which is available to authorized users.
A pH and solvent optimized reverse-phase ion-paring-LC–MS/MS method that leverages multiple scan-types for targeted absolute quantification of intracellular metabolites

Comprehensive knowledge of intracellular biochemistry is needed to accurately understand, model, and manipulate metabolism for industrial and therapeutic applications. Quantitative metabolomics has been driven by advances in analytical instrumentation and can add valuable knowledge to the understanding of intracellular metabolism. Liquid chromatography coupled to mass spectrometry (LC–MS and LC–MS/MS) has become a reliable means with which to quantify a multitude of intracellular metabolites in parallel with great specificity and accuracy. This work details a method that builds and extends upon existing reverse phase ion-paring liquid chromatography methods for separation and detection of polar and anionic compounds that comprise key nodes of intracellular metabolism by optimizing pH and solvent composition. In addition, the presented method utilizes multiple scan types provided by hybrid instrumentation to improve confidence in compound identification. The developed method was validated for a broad coverage of polar and anionic metabolites of intracellular metabolism.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California
A streamlined ribosome profiling protocol for the characterization of microorganisms

Ribosome profiling is a powerful tool for characterizing in vivo protein translation at the genome scale, with multiple applications ranging from detailed molecular mechanisms to systems-level predictive modeling. Though highly effective,
this intricate technique has yet to become widely used in the microbial research community. Here we present a streamlined ribosome profiling protocol with reduced barriers to entry for microbial characterization studies. Our approach provides simplified alternatives during harvest, lysis, and recovery of monosomes and also eliminates several time-consuming steps, in particular size-selection steps during library construction. Furthermore, the abundance of rRNAs and tRNAs in the final library is drastically reduced. Our streamlined workflow enables greater throughput, cuts the time from harvest to the final library in half (down to 3-4 days), and generates a high fraction of informative reads, all while retaining the high quality standards of the existing protocol.
Computing the functional proteome: recent progress and future prospects for genome-scale models

Constraint-based models enable the computation of feasible, optimal, and realized biological phenotypes from reaction network reconstructions and constraints on their operation. To date, stoichiometric reconstructions have largely focused on metabolism, resulting in genome-scale metabolic models (M-Models). Recent expansions in network content to encompass proteome synthesis have resulted in models of metabolism and protein expression (ME-Models). ME-Models advance the predictions possible with constraint-based models from network flux states to the spatially resolved molecular composition of a cell. Specifically, ME-Models enable the prediction of transcriptome and proteome allocation and limitations, and basal expression states and regulatory needs. Continued expansion in reconstruction content and constraints will result in an increasingly refined representation of cellular composition and behavior.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California, San Diego
Authors: O'Brien, E. J. (Ekstern), Palsson, B. (Intern)
Number of pages: 10
Pages: 125-134
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Current Opinion in Biotechnology
Volume: 34
ISSN (Print): 0958-1669
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 8.55 SJR 3.331 SNIP 2.1
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 3.113 SNIP 2.143 CiteScore 7.99
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 3.271 SNIP 2.068 CiteScore 7.45
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 3.322 SNIP 2.198 CiteScore 7.93
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
Decoding genome-wide GadEWX-transcriptional regulatory networks reveals multifaceted cellular responses to acid stress in Escherichia coli

The regulators GadE, GadW and GadX (which we refer to as GadEWX) play a critical role in the transcriptional regulation of the glutamate-dependent acid resistance (GDAR) system in *Escherichia coli* K-12 MG1655. However, the genome-wide regulatory role of GadEWX is still unknown. Here we comprehensively reconstruct the genome-wide GadEWX transcriptional regulatory network and RpoS involvement in *E. coli* K-12 MG1655 under acidic stress. Integrative data analysis reveals that GadEWX regulons consist of 45 genes in 31 transcription units and 28 of these genes were associated with RpoS-binding sites. We demonstrate that GadEWX directly and coherently regulate several proton-generating/consuming enzymes with pairs of negative-feedback loops for pH homeostasis. In addition, GadEWX regulate genes with assorted functions, including molecular chaperones, acid resistance, stress response and other regulatory activities. These results show how GadEWX simultaneously coordinate many cellular processes to produce the overall response of *E. coli* to acid stress.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California
Authors: Seo, S. W. (Ekstern), Kim, D. (Ekstern), O'Brien, E. J. (Ekstern), Szubin, R. (Ekstern), Palsson, B. O. (Intern)
Number of pages: 8
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Do genome-scale models need exact solvers or clearer standards?

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, New Bioactive Compounds, Yeast Cell Factories, University of California, Norwegian University of Science and Technology, University of Luxembourg, Sinopia Biosciences, Genomatica Inc, Harvard Medical School, Intrexon, Inc., Virginia Commonwealth University, University of Lausanne, Rose-Hulman Institute of Technology, California Institute of Technology, Babraham Institut, Stanford University, University of Toronto, Biological Research Center, University of Virginia, European Molecular Biology Laboratory, Institute for Systems Biology, University of Wisconsin-Madison, Virginia Tech, Lawrence Livermore National Laboratory, Utah State University, University of Queensland, Pennsylvania State University, Centro de Investigaciones Biologicas
Publication information
Journal: Molecular Systems Biology
Volume: 11
Issue number: 10
Article number: 831
ISSN (Print): 1744-4292
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 8.23 SJR 8.366 SNIP 2.15
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 8.731 SNIP 2.395 CiteScore 9.76
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 10.072 SNIP 3.505 CiteScore 11.8
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 9.637 SNIP 2.875 CiteScore 11.84
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 7.904 SNIP 2.417 CiteScore 10.13
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 7.481 SNIP 2.306 CiteScore 8.78
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 6.293 SNIP 2.45
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 6.386 SNIP 2.602
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 5.206 SNIP 2.307
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 3.243 SNIP 1.572
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 2.24
Web of Science (2006): Indexed yes
Original language: English
Electronic versions:
Do_genome_scale_models_need_exact.pdf
DOIs:
10.15252/msb.20156157

Bibliographical note
This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
Genome-wide Reconstruction of OxyR and SoxRS Transcriptional Regulatory Networks under Oxidative Stress in *Escherichia coli* K-12 MG1655

Three transcription factors (TFs), OxyR, SoxR, and SoxS, play a critical role in transcriptional regulation of the defense system for oxidative stress in bacteria. However, their full genome-wide regulatory potential is unknown. Here, we perform a genome-scale reconstruction of the OxyR, SoxR, and SoxS regulons in *Escherichia coli* K-12 MG1655. Integrative data analysis reveals that a total of 68 genes in 51 transcription units (TUs) belong to these regulons. Among them, 48 genes showed more than 2-fold changes in expression level under single-TF-knockout conditions. This reconstruction expands the genome-wide roles of these factors to include direct activation of genes related to amino acid biosynthesis (methionine and aromatic amino acids), cell wall synthesis (lipid A biosynthesis and peptidoglycan growth), and divalent metal ion transport (Mn^{2+}, Zn^{2+}, and Mg^{2+}). Investigating the co-regulation of these genes with other stress-response TFs reveals that they are independently regulated by stress-specific TFs.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California
Authors: Seo, S. W. (Ekstern), Kim, D. (Ekstern), Szubin, R. (Ekstern), Palsson, B. O. (Intern)
Number of pages: 11
Pages: 1289-1299
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Cell Reports
Volume: 12
Issue number: 8
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 8.4 SJR 8.012 SNIP 1.749
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 8.507 SNIP 1.754 CiteScore 8.15
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 8.334 SNIP 1.883 CiteScore 7.88
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 8.046 SNIP 1.677 CiteScore 7.22
ISI indexed (2013): ISI indexed no
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
ISI indexed (2012): ISI indexed no
Original language: English
Electronic versions:
Genome_wide_Reconstruction_of_OxyR_and_SoxRS.pdf
DOIs:
10.1016/j.celrep.2015.07.043

Bibliographical note
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Source: FindIt
Source-ID: 275812827
Publication: Research - peer-review › Journal article – Annual report year: 2015
How to set up collaborations between academia and industrial biotech companies

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, New Bioactive Compounds, Network Reconstruction in Silico Biology, Yeast Cell Factories, Delft University of Technology, Genomatica Inc, BP
Authors: Pronk, J. T. (Ekstern), Lee, S. Y. (Intern), Lievense, J. (Ekstern), Pierce, J. (Ekstern), Palsson, B. (Intern), Uhlén, M. (Intern), Nielsen, J. (Intern)
Number of pages: 4
Pages: 237-240
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Nature Biotechnology
Volume: 33
Issue number: 3
ISSN (Print): 1087-0156
Ratings:
BFI (2018): BFI-level 3
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 13.16 SJR 20.253 SNIP 6.303
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 17.892 SNIP 5.505 CiteScore 11.88
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 16.443 SNIP 5.433 CiteScore 11.4
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 13.849 SNIP 5.416 CiteScore 10.45
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 10.76 SNIP 4.96 CiteScore 8.44
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 11.627 SNIP 6.248 CiteScore 8.21
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 7.763 SNIP 5.607
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 6.046 SNIP 5.07
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 5.039 SNIP 4.588
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 5.74 SNIP 4.596
Scopus rating (2005): SJR 5.151 SNIP 3.832
Scopus rating (2004): SJR 4.673 SNIP 3.635
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 3.804 SNIP 2.947
JSBML 1.0: providing a smorgasbord of options to encode systems biology models

JSBML, the official pure Java programming library for the Systems Biology Markup Language (SBML) format, has evolved with the advent of different modeling formalisms in systems biology and their ability to be exchanged and represented via extensions of SBML. JSBML has matured into a major, active open-source project with contributions from a growing, international team of developers who not only maintain compatibility with SBML, but also drive steady improvements to the Java interface and promote ease-of-use with end users. Source code, binaries and documentation for JSBML can be freely obtained under the terms of the LGPL 2.1 from the website http://sbml.org/Software/JSBML. More information about JSBML can be found in the user guide at http://sbml.org/Software/JSBML/docs/ jsbml-development@googlegroups.com or andraeger@eng.ucsd.edu Supplementary data are available at Bioinformatics online.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, European Bioinformatics Institute, University of Utah, Marquette University, University of Toronto, Boston University, University of Tubingen, Institute of Plant Genetics and Crop Plant Research, California Institute of Technology, University of California
Authors: Rodriguez, N. (Ekstern), Thomas, A. (Ekstern), Watanabe, L. (Ekstern), Vazirabad, I. Y. (Ekstern), Kofia, V. (Ekstern), Gómez, H. F. (Ekstern), Mittag, F. (Ekstern), Matthes, J. (Ekstern), Rudolph, J. (Ekstern), Wrzodek, F. (Ekstern) , Netz, E. (Ekstern), Diamantikos, A. (Ekstern), Eichner, J. (Ekstern), Keller, R. (Ekstern), Wrzodek, C. (Ekstern), Fröhlich, S. (Ekstern), Lewis, N. (Intern), Myers, C. J. (Ekstern), Le Novère, N. (Ekstern), Palsson, B. (Intern), Hucka, M. (Ekstern), Dräger, A. (Ekstern)
Number of pages: 4
Pages: 3383-6
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Bioinformatics
Volume: 31
Issue number: 20
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 6.42
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 6.06
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 5.5
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 5.78
Model-driven discovery of synergistic inhibitors against E. coli and S. enterica serovar Typhimurium targeting a novel synthetic lethal pair, aldA and prpC

Mathematical models of biochemical networks form a cornerstone of bacterial systems biology. Inconsistencies between simulation output and experimental data point to gaps in knowledge about the fundamental biology of the organism. One such inconsistency centers on the gene aldA in Escherichia coli: it is essential in a computational model of E. coli metabolism, but experimentally it is not. Here, we reconcile this disparity by providing evidence that aldA and prpC form a synthetic lethal pair, as the double knockout could only be created through complementation with a plasmid-borne copy of aldA. Moreover, virtual and biological screening against the two proteins led to a set of compounds that inhibited the growth of E. coli and Salmonella enterica serovar Typhimurium synergistically at 100-200 μM individual concentrations. These results highlight the power of metabolic models to drive basic biological discovery and their potential use to discover new combination antibiotics.
Model-driven discovery of underground metabolic functions in *Escherichia coli*

Enzyme promiscuity toward substrates has been discussed in evolutionary terms as providing the flexibility to adapt to novel environments. In the present work, we describe an approach toward exploring such enzyme promiscuity in the space of a metabolic network. This approach leverages genome-scale models, which have been widely used for predicting growth phenotypes in various environments or following a genetic perturbation; however, these predictions occasionally fail. Failed predictions of gene essentiality offer an opportunity for targeting biological discovery, suggesting the presence of unknown underground pathways stemming from enzymatic cross-reactivity. We demonstrate a workflow that couples constraint-based modeling and bioinformatic tools with KO strain analysis and adaptive laboratory evolution for the purpose of predicting promiscuity at the genome scale. Three cases of genes that are incorrectly predicted as essential in *Escherichia coli*—aspC, argD, and gltA—are examined, and isozyme functions are uncovered for each to a different extent. Seven isozyme functions based on genetic and transcriptional evidence are suggested between the genes aspC and tyrB, argD and astC, gabT and puuE, and gltA and prpC. This study demonstrates how a targeted model-driven approach to discovery can systematically fill knowledge gaps, characterize underground metabolism, and elucidate regulatory mechanisms of adaptation in response to gene KO perturbations.

**General information**

**State:** Published

**Organisations:** Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California, San Diego, Russian Academy of Sciences, Lawrence Berkeley National Laboratory
Next-generation genome-scale models for metabolic engineering

Constraint-based reconstruction and analysis (COBRA) methods have become widely used tools for metabolic engineering in both academic and industrial laboratories. By employing a genome-scale in silico representation of the metabolic network of a host organism, COBRA methods can be used to predict optimal genetic modifications that improve the rate and yield of chemical production. A new generation of COBRA models and methods is now being developed, encompassing many biological processes and simulation strategies. Next-generation models enable new types of predictions. Here, three key examples of applying COBRA methods to strain optimization are presented and discussed. Then, an outlook is provided on the next generation of COBRA models and the new types of predictions they will enable for systems metabolic engineering.
Understanding individual variation is fundamental to personalized medicine. Yet interpreting complex phenotype data, such as multi-compartment metabolomic profiles, in the context of genotype data for an individual is complicated by interactions within and between cells and remains an unresolved challenge. Here, we constructed multi-omic, data-driven, personalized whole-cell kinetic models of erythrocyte metabolism for 24 healthy individuals based on fasting-state plasma and erythrocyte metabolomics and whole-genome genotyping. We show that personalized kinetic rate constants, rather than metabolite levels, better represent the genotype. Additionally, changes in erythrocyte dynamics between individuals occur on timescales of circulation, suggesting detected differences play a role in physiology. Finally, we use the models to identify individuals at risk for a drug side effect (ribavirin-induced anemia) and how genetic variation (inosine triphosphatase deficiency) may protect against this side effect. This study demonstrates the feasibility of personalized kinetic models, and we anticipate their use will accelerate discoveries in characterizing individual metabolic variation.
SBMLsqueezer 2: context-sensitive creation of kinetic equations in biochemical networks

Background: The size and complexity of published biochemical network reconstructions are steadily increasing, expanding the potential scale of derived computational models. However, the construction of large biochemical network models is a laborious and error-prone task. Automated methods have simplified the network reconstruction process, but building kinetic models for these systems is still a manually intensive task. Appropriate kinetic equations, based upon reaction rate laws, must be constructed and parameterized for each reaction. The complex test-and-evaluation cycles that can be involved during kinetic model construction would thus benefit from automated methods for rate law assignment.

Results: We present a high-throughput algorithm to automatically suggest and create suitable rate laws based upon reaction type according to several criteria. The criteria for choices made by the algorithm can be influenced in order to assign the desired type of rate law to each reaction. This algorithm is implemented in the software package SBMLsqueezer 2. In addition, this program contains an integrated connection to the kinetics database SABIO-RK to obtain experimentally-derived rate laws when desired.

Conclusions: The described approach fills a heretofore absent niche in workflows for large-scale biochemical kinetic model construction. In several applications the algorithm has already been demonstrated to be useful and scalable. SBMLsqueezer is platform independent and can be used as a stand-alone package, as an integrated plugin, or through a web interface, enabling flexible solutions and use-case scenarios.
Systems biology definition of the core proteome of metabolism and expression is consistent with high-throughput data

Finding the minimal set of gene functions needed to sustain life is of both fundamental and practical importance. Minimal gene lists have been proposed by using comparative genomics-based core proteome definitions. A definition of a core proteome that is supported by empirical data, is understood at the systems-level, and provides a basis for computing essential cell functions is lacking. Here, we use a systems biology-based genome-scale model of metabolism and expression to define a functional core proteome consisting of 356 gene products, accounting for 44% of the *Escherichia coli* proteome by mass based on proteomics data. This systems biology core proteome includes 212 genes not found in previous comparative genomics-based core proteome definitions, accounts for 65% of known essential genes in E. coli, and has 78% gene function overlap with minimal genomes (*Buchnera aphidicola* and *Mycoplasma genitalium*). Based on transcriptomics data across environmental and genetic backgrounds, the systems biology core proteome is significantly enriched in nondifferentially expressed genes and depleted in differentially expressed genes. Compared with the noncore, core gene expression levels are also similar across genetic backgrounds (two times higher Spearman rank correlation) and exhibit significantly more complex transcriptional and posttranscriptional regulatory features (40% more transcription start sites per gene, 22% longer 5'UTR). Thus, genome-scale systems biology approaches rigorously identify a functional core proteome needed to support growth. This framework, validated by using high-throughput datasets, facilitates a mechanistic understanding of systems-level core proteome function through in silico models; it de facto defines a paleome.
The architecture of ArgR-DNA complexes at the genome-scale in *Escherichia coli*

DNA-binding motifs that are recognized by transcription factors (TFs) have been well studied; however, challenges remain in determining the *in vivo* architecture of TF-DNA complexes on a genome-scale. Here, we determined the *in vivo* architecture of *Escherichia coli* arginine repressor (ArgR)-DNA complexes using high-throughput sequencing of exonuclease-treated chromatin-immunoprecipitated DNA (ChIP-exo). The ChIP-exo has a unique peak-pair pattern indicating 5' and 3' ends of ArgR-binding region. We identified 62 ArgR-binding loci, which were classified into three groups, comprising single, double and triple peak-pairs. Each peak-pair has a unique 93 base pair (bp)-long (±2 bp) ArgR-binding sequence containing two ARG boxes (39 bp) and residual sequences. Moreover, the three ArgR-binding modes defined by the position of the two ARG boxes indicate that DNA bends centered between the pair of ARG boxes facilitate the non-specific contacts between ArgR subunits and the residual sequences. Additionally, our approach may also reveal other fundamental structural features of TF-DNA interactions that have implications for studying genome-scale transcriptional regulatory networks.
Use of Adaptive Laboratory Evolution To Discover Key Mutations Enabling Rapid Growth of *Escherichia coli* K-12 MG1655 on Glucose Minimal Medium

Adaptive laboratory evolution (ALE) has emerged as an effective tool for scientific discovery and addressing biotechnological needs. Much of ALE’s utility is derived from reproducibly obtained fitness increases. Identifying causal genetic changes and their combinatorial effects is challenging and time-consuming. Understanding how these genetic changes enable increased fitness can be difficult. A series of approaches that address these challenges was developed and demonstrated using Escherichia coli K-12 MG1655 on glucose minimal media at 37°C. By keeping E. coli in constant substrate excess and exponential growth, fitness increases up to 1.6-fold were obtained compared to the wild type. These increases are comparable to previously reported maximum growth rates in similar conditions but were obtained over a shorter time frame. Across the eight replicate ALE experiments performed, causal mutations were identified using three approaches: identifying mutations in the same gene/region across replicate experiments, sequencing strains before and
after computationally determined fitness jumps, and allelic replacement coupled with targeted ALE of reconstructed strains. Three genetic regions were most often mutated: the global transcription gene rpoB, an 82-bp deletion between the metabolic pyrE gene and rph, and an IS element between the DNA structural gene hns and tdk. Model-derived classification of gene expression revealed a number of processes important for increased growth that were missed using a gene classification system alone. The methods described here represent a powerful combination of technologies to increase the speed and efficiency of ALE studies. The identified mutations can be examined as genetic parts for increasing growth rate in a desired strain and for understanding rapid growth phenotypes.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California
Authors: LaCroix, R. A. (Ekstern), Sandberg, T. E. (Ekstern), O'Brien, E. J. (Ekstern), Utrilla, J. (Ekstern), Ebrahim, A. (Ekstern), Guzman, G. I. (Ekstern), Szubin, R. (Ekstern), Palsson, B. (Intern), Feist, A. (Intern)
Number of pages: 14
Pages: 17-30
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Applied and Environmental Microbiology
Volume: 81
Issue number: 1
ISSN (Print): 0099-2240
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 4.08
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 1.891 SNIP 1.308 CiteScore 4.14
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 1.857 SNIP 1.384 CiteScore 4.02
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 1.899 SNIP 1.414 CiteScore 4.25
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 1.975 SNIP 1.429 CiteScore 4.29
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 1.914 SNIP 1.455 CiteScore 4.12
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 1.887 SNIP 1.436
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 1.972 SNIP 1.528
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 2.156 SNIP 1.572
Web of Science (2008): Indexed yes
Using Genome-scale Models to Predict Biological Capabilities

Constraint-based reconstruction and analysis (COBRA) methods at the genome scale have been under development since the first whole-genome sequences appeared in the mid-1990s. A few years ago, this approach began to demonstrate the ability to predict a range of cellular functions, including cellular growth capabilities on various substrates and the effect of gene knockouts at the genome scale. Thus, much interest has developed in understanding and applying these methods to areas such as metabolic engineering, antibiotic design, and organismal and enzyme evolution. This Primer will get you started.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California
Authors: O’Brien, E. J. (Ekstern), Monk, J. M. (Ekstern), Palsson, B. O. (Intern)
Number of pages: 17
Pages: 971-987
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Cell
Volume: 161
Issue number: 5
ISSN (Print): 0092-8674
Ratings:
BFI (2018): BFI-level 3
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 22.79 SJR 26.947 SNIP 4.959
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 27.696 SNIP 5.314 CiteScore 23.62
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
A model-driven quantitative metabolomics analysis of aerobic and anaerobic metabolism in E. coli K-12 MG1655 that is biochemically and thermodynamically consistent

The advent of model-enabled workflows in systems biology allows for the integration of experimental data types with genome-scale models to discover new features of biology. This work demonstrates such a workflow, aimed at establishing a metabolomics platform applied to study the differences in metabolomes between anaerobic and aerobic growth of Escherichia coli. Constraint-based modeling was utilized to deduce a target list of compounds for downstream method development. An analytical and experimental methodology was developed and tailored to the compound chemistry and growth conditions of interest. This included the construction of a rapid sampling apparatus for use with anaerobic cultures. The resulting genome-scale data sets for anaerobic and aerobic growth were validated by comparison to previous small-scale studies comparing growth of E. coli under the same conditions. The metabolomics data were then integrated with the E. coli genome-scale metabolic model (GEM) via a sensitivity analysis that utilized reaction thermodynamics to reconcile simulated growth rates and reaction directionalities. This analysis highlighted several optimal network usage inconsistencies, including the incorrect use of the beta-oxidation pathway for synthesis of fatty acids. This analysis also identified enzyme promiscuity for the pykA gene, that is critical for anaerobic growth, and which has not been previously incorporated into metabolic models of E coli. Biotechnol.
Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in Escherichia coli

The ferric uptake regulator (Fur) plays a critical role in the transcriptional regulation of iron metabolism. However, the full regulatory potential of Fur remains undefined. Here we comprehensively reconstruct the Fur transcriptional regulatory network in Escherichia coli K-12 MG1655 in response to iron availability using genome-wide measurements. Integrative data analysis reveals that a total of 81 genes in 42 transcription units are directly regulated by three different modes of Fur regulation, including apo- and holo-Fur activation and holo-Fur repression. We show that Fur connects iron transport and utilization enzymes with negative-feedback loop pairs for iron homeostasis. In addition, direct involvement of Fur in the regulation of DNA synthesis, energy metabolism and biofilm development is found. These results show how Fur exhibits a comprehensive regulatory role affecting many fundamental cellular processes linked to iron metabolism in order to coordinate the overall response of E. coli to iron availability.
Determining the Control Circuitry of Redox Metabolism at the Genome-Scale

Determining how facultative anaerobic organisms sense and direct cellular responses to electron acceptor availability has been a subject of intense study. However, even in the model organism Escherichia coli, established mechanisms only explain a small fraction of the hundreds of genes that are regulated during electron acceptor shifts. Here we propose a qualitative model that accounts for the full breadth of regulated genes by detailing how two global transcription factors (TFs), ArcA and Fnr of E. coli, sense key metabolic redox ratios and act on a genome-wide basis to regulate anabolic, catabolic, and energy generation pathways. We first fill gaps in our knowledge of this transcriptional regulatory network by carrying out ChIP-chip and gene expression experiments to identify 463 regulatory events. We then interfaced this reconstructed regulatory network with a highly curated genome-scale metabolic model to show that ArcA and Fnr regulate >80% of total metabolic flux and 96% of differential gene expression across fermentative and nitrate respiratory conditions. Based on the data, we propose a feedforward with feedback trim regulatory scheme, given the extensive repression of catabolic genes by ArcA and extensive activation of chemiosmotic genes by Fnr. We further corroborated this regulatory scheme by showing a 0.71 r(2) (p
Engineering synergy in biotechnology

In this article, the author focuses on approaches in metabolic engineering and synthetic biology for the creation of efficient cell factories, which can be used to convert biomass and other feedstocks for the generation of chemicals. Topics discussed include development of restriction enzymes, engineering plasmids and recyclable markers, production of 1,3-propanediol using a metabolically engineered Escherichia coli and production of isobutanol by using metabolically engineered yeast.

General information

State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Fungal Cell Factories, New Bioactive Compounds, Network Reconstruction in Silico Biology, Swiss Federal Institute of Technology, Technical University of Denmark, University of California, Massachusetts Institute of Technology
Authors: Nielsen, J. (Intern), Fussenegger, M. (Ekstern), Keasling, J. (Ekstern), Yup Lee, S. (Intern), Liao, J. C. (Ekstern), Prather, K. (Ekstern), Palsson, B. (Intern)
Number of pages: 4
Pages: 319-322
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: Nature Chemical Biology
Volume: 10
Issue number: 5
ISSN (Print): 1552-4450
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 9.41 SJR 8.778 SNIP 3.056
Evolution of Escherichia coli to 42 °C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations.

Adaptive laboratory evolution (ALE) has emerged as a valuable method by which to investigate microbial adaptation to a desired environment. Here, we performed ALE to 42 °C of ten parallel populations of Escherichia coli K-12 MG1655 grown in glucose minimal media. Tightly controlled experimental conditions allowed selection based on exponential-phase growth rate, yielding strains that uniformly converged toward a similar phenotype along distinct genetic paths. Adapted strains possessed as few as 6 and as many as 55 mutations, and of the 144 genes that mutated in total, 14 arose independently across two or more strains. This mutational recurrence pointed to the key genetic targets underlying the evolved fitness increase. Genome engineering was used to introduce the novel ALE-acquired alleles in random combinations into the ancestral strain, and competition between these engineered strains reaffirmed the impact of the key mutations on the growth rate at 42 °C. Interestingly, most of the identified key gene targets differed significantly from those found in similar temperature adaptation studies, highlighting the sensitivity of genetic evolution to experimental conditions and ancestral genotype. Additionally, transcriptomic analysis of the ancestral and evolved strains revealed a general trend for restoration of the global expression state back toward preheat stressed levels. This restorative effect was previously documented following evolution to metabolic perturbations, and thus may represent a general feature of ALE experiments. The widespread evolved expression shifts were enabled by a comparatively scant number of regulatory mutations, providing a net fitness benefit but causing suboptimal expression levels for certain genes, such as those governing flagellar formation, which then became targets for additional ameliorating mutations. Overall, the results of this study provide insight into the adaptation process and yield lessons important for the future implementation of ALE as a tool for scientific research and engineering.
Generation of an atlas for commodity chemical production in Escherichia coli and a novel pathway prediction algorithm, GEM-Path

The production of 75% of the current drug molecules and 35% of all chemicals could be achieved through bioprocessing (Arundel and Sawava, 2009). To accelerate the transition from a petroleum based chemical industry to a sustainable bio-based industry, systems metabolic engineering has emerged to computationally design metabolic pathways for chemical production. Although algorithms able to provide specific metabolic interventions and heterologous production pathways are available, a systematic analysis for all possible production routes to commodity chemicals in Escherichia coli is lacking. Furthermore, a pathway prediction algorithm that combines direct integration of genome-scale models at each step of the search to reduce the search space does not exist. Previous work (Feist. el. al., 2010) performed a model driven evaluation of the growth coupled production potential for E. coli to produce multiple native compounds from different feedstocks. In this study, we extended this analysis for non-native compounds by using an integrated approach through heterologous pathway integration and growth coupled metabolite production design. In addition to integration with genome-scale model integration, the GEM Path algorithm developed in this work also contains a novel approach to address reaction promiscuity. In total, 245 unique synthetic pathways for 20 large volume compounds were predicted. Host metabolism with these synthetic pathways was then analyzed for feasible growth-coupled production and designs could be identified for 1271 of the 6615 conditions evaluated. This study characterizes the potential for E coli to produce commodity chemicals, and outlines a generic strain design workflow to design production strains. (C) 2014 international Metabolic Engineering Society. Published by Elsevier Inc. All rights reserved.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, iLoop, University of California, University of Chile
Authors: Campodonico, M. A. (Ekstern), Andrews, B. A. (Ekstern), Asenjo, J. A. (Ekstern), Palsson, B. (Intern), Feist, A. (Intern)
Number of pages: 19
Pages: 140-158
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: Metabolic Engineering
Volume: 25
ISSN (Print): 1096-7176
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 8.33 SJR 3.54 SNIP 1.864
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 3.611 SNIP 1.822 CiteScore 8.2
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 3.381 SNIP 2.034 CiteScore 7.23
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 4.004 SNIP 2.185 CiteScore 8.43
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 3.032 SNIP 1.858 CiteScore 6.72
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 3.124 SNIP 2.144 CiteScore 6.75
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 2.373 SNIP 1.802
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 2.575 SNIP 1.421
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 1.757 SNIP 1.028
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 1.504 SNIP 1.184
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 1.269 SNIP 0.892
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 1.056 SNIP 1.065
Scopus rating (2004): SJR 1.657 SNIP 1.261
Scopus rating (2003): SJR 1.168 SNIP 0.884
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 1.182 SNIP 1.028
Web of Science (2002): Indexed yes
Scopus rating (2001): SJR 0.921 SNIP 0.766
Scopus rating (2000): SJR 0.724 SNIP 0.896

Original language: English
atlas, Facultatively Anaerobic Gram-Negative Rods Eubacteria Bacteria Microorganisms (Bacteria, Eubacteria, Microorganisms) - Enterobacteriaceae [06702] Escherichia coli species, 04500, Mathematical biology and statistical methods, 10515, Biophysics - Biocybernetics, 31000, Physiology and biochemistry of bacteria, 39008, Food microbiology - General and miscellaneous, Computational Biology, GEM-Path algorithm mathematical and computer techniques, metabolic engineering laboratory techniques, Bioprocess Engineering, Models and Simulations, BIOTECHNOLOGY, COMPLEX METABOLIC NETWORKS, CONSTRAINT-BASED MODELS, MICROBIAL-PRODUCTION, 2,3-BUTANEDIOL PRODUCTION, BIODIESEL PRODUCTION, BIOFUELS PRODUCTION, ADAPTIVE EVOLUTION, FLUX ANALYSIS, K-12 MG1655, SYSTEMS, Systems biology, Pathway predictions, Escherichia coli, Strain design
DOIs:
10.1016/j.ymben.2014.07.009
Source: Findit
Source-ID: 269548041
Publication: Research - peer-review › Journal article – Annual report year: 2014

Genome-scale reconstruction of the sigma factor network in Escherichia coli: topology and functional states
Background: At the beginning of the transcription process, the RNA polymerase (RNAP) core enzyme requires a sigma-factor to recognize the genomic location at which the process initiates. Although the crucial role of sigma-factors has long been appreciated and characterized for many individual promoters, we do not yet have a genome-scale assessment of their function. Results: Using multiple genome-scale measurements, we elucidated the network of s-factor and promoter interactions in Escherichia coli. The reconstructed network includes 4,724 sigma-factor-specific promoters corresponding to transcription units (TUs), representing an increase of more than 300% over what has been previously reported. The reconstructed network was used to investigate competition between alternative sigma-factors (the sigma(70) and sigma(38) regulons), confirming the competition model of sigma substitution and negative regulation by alternative s-factors. Comparison with sigma-factor binding in Klebsiella pneumoniae showed that transcriptional regulation of conserved genes in closely related species is unexpectedly divergent. Conclusions: The reconstructed network reveals the regulatory complexity of the promoter architecture in prokaryotic genomes, and opens a path to the direct determination of the systems biology of their transcriptional regulatory networks.
Minimal metabolic pathway structure is consistent with associated biomolecular interactions

Pathways are a universal paradigm for functionally describing cellular processes. Even though advances in high-throughput data generation have transformed biology, the core of our biological understanding, and hence data interpretation, is still predicated on human-defined pathways. Here, we introduce an unbiased, pathway structure for genome-scale metabolic networks defined based on principles of parsimony that do not mimic canonical human-defined textbook pathways. Instead, these minimal pathways better describe multiple independent pathway-associated biomolecular interaction datasets suggesting a functional organization for metabolism based on parsimonious use of cellular components. We use the inherent predictive capability of these pathways to experimentally discover novel transcriptional regulatory interactions in Escherichia coli metabolism for three transcription factors, effectively doubling the known regulatory roles for Nac and MntR. This study suggests an underlying and fundamental principle in the evolutionary selection of pathway structures; namely, that pathways may be minimal, independent, and segregated.

General information

State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, University of California, San Diego
Authors: Bordbar, A. (Ekstern), Nagarajan, H. (Ekstern), Lewis, N. E. (Ekstern), Latif, H. (Ekstern), Ebrahim, A. (Ekstern), Federowicz, S. (Ekstern), Schellenberger, J. (Ekstern), Palsson, B. (Intern)
Number of pages: 16
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: Molecular Systems Biology
Volume: 10
Issue number: 7
ISSN (Print): 1744-4292
Ratings:
BFI (2018): BFI-level 2
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 8.23 SJR 8.366 SNIP 2.15
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 8.731 SNIP 2.395 CiteScore 9.76
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 10.072 SNIP 3.505 CiteScore 11.8
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 9.637 SNIP 2.875 CiteScore 11.84
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 7.904 SNIP 2.417 CiteScore 10.13
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 7.481 SNIP 2.306 CiteScore 8.78
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 6.293 SNIP 2.45
Web of Science (2010): Indexed yes
Reconstruction and modeling protein translocation and compartmentalization in Escherichia coli at the genome-scale

Background: Membranes play a crucial role in cellular functions. Membranes provide a physical barrier, control the trafficking of substances entering and leaving the cell, and are a major determinant of cellular ultra-structure. In addition, components embedded within the membrane participate in cell signaling, energy transduction, and other critical cellular functions. All these processes must share the limited space in the membrane; thus it represents a notable constraint on cellular functions. Membrane- and location-based processes have not yet been reconstructed and explicitly integrated into genome-scale models.

Results: The recent genome-scale model of metabolism and protein expression in Escherichia coli (called a ME-model) computes the complete composition of the proteome required to perform whole cell functions. Here we expand the ME-model to include (1) a reconstruction of protein translocation pathways, (2) assignment of all cellular proteins to one of four compartments (cytoplasm, inner membrane, periplasm, and outer membrane) and a translocation pathway, (3) experimentally determined translocase catalytic and porin diffusion rates, and (4) a novel membrane constraint that reflects cell morphology. Comparison of computations performed with this expanded ME-model, named iJL1678-ME, against available experimental data reveals that the model accurately describes translocation pathway expression and the functional proteome by compartmentalized mass.

Conclusion: iJL1678-ME enables the computation of cellular phenotypes through an integrated computation of proteome composition, abundance, and activity in four cellular compartments (cytoplasm, periplasm, inner and outer membrane). Reconstruction and validation of the model has demonstrated that the iJL1678-ME is capable of capturing the functional content of membranes, cellular compartment-specific composition, and that it can be utilized to examine the effect of perturbing an expanded set of network components. iJL1678-ME takes a notable step towards the inclusion of cellular ultra-structure in genome-scale models.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Research Groups, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, iLoop, University of California
Authors: Liu, J. K. (Ekstern), O’Brien, E. J. (Ekstern), Lerman, J. A. (Ekstern), Zengler, K. (Intern), Palsson, B. (Intern), Feist, A. (Intern)
Number of pages: 15
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: BMC Systems Biology
Volume: 8
Issue number: 1
ISSN (Print): 1752-0509
Ratings:
BFI (2018): BFI-level 1
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Streptomyces species continue to attract attention as a source of novel medicinal compounds. Despite a long history of studies on these microorganisms, they still have many biochemical mysteries to be elucidated. Investigations of novel secondary metabolites and their biosynthetic gene clusters have been more systematized with high-throughput techniques through inspections of correlations among components of the primary and secondary metabolisms at the genome scale. Moreover, up-to-date information on the genome of Streptomyces species with emphasis on their secondary metabolism has been collected in the form of databases and knowledgebases, providing predictive information and enabling one to explore experimentally unrecognized biological spaces of secondary metabolism. Herein, we review recent trends in the systems biology and biotechnology of Streptomyces species.
Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli.

The genome-scale model (GEM) of metabolism in the bacterium Escherichia coli K-12 has been in development for over a decade and is now in wide use. GEM-enabled studies of E. coli have been primarily focused on six applications: (1) metabolic engineering, (2) model-driven discovery, (3) prediction of cellular phenotypes, (4) analysis of biological network properties, (5) studies of evolutionary processes, and (6) models of interspecies interactions. In this review, we provide an overview of these applications along with a critical assessment of their successes and limitations, and a perspective on likely future developments in the field. Taken together, the studies performed over the past decade have established a genome-scale mechanistic understanding of genotype-phenotype relationships in E. coli metabolism that forms the basis for similar efforts for other microbial species. Future challenges include the expansion of GEMs by integrating additional cellular processes beyond metabolism, the identification of key constraints based on emerging data types, and the development of computational methods able to handle such large-scale network models with sufficient accuracy.
Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome.

Chinese hamster ovary (CHO) cells, first isolated in 1957, are the preferred production host for many therapeutic proteins. Although genetic heterogeneity among CHO cell lines has been well documented, a systematic, nucleotide-resolution characterization of their genotypic differences has been stymied by the lack of a unifying genomic resource for CHO cells. Here we report a 2.4-Gb draft genome sequence of a female Chinese hamster, Cricetulus griseus, harboring 24,044 genes. We also resequenced and analyzed the genomes of six CHO cell lines from the CHO-K1, DG44 and CHO-S lineages. This analysis identified hamster genes missing in different CHO cell lines, and detected >3.7 million single-nucleotide polymorphisms (SNPs), 551,240 indels and 7,063 copy number variations. Many mutations are located in genes with functions relevant to bioprocessing, such as apoptosis. The details of this genetic diversity highlight the value of the hamster genome as the reference upon which CHO cells can be studied and engineered for protein production.
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 13.849 SNIP 5.416 CiteScore 10.45
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 10.76 SNIP 4.96 CiteScore 8.44
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 11.627 SNIP 6.248 CiteScore 8.21
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 7.763 SNIP 5.607
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 6.046 SNIP 5.07
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 5.039 SNIP 4.588
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 5.74 SNIP 4.596
Scopus rating (2005): SJR 5.151 SNIP 3.832
Scopus rating (2004): SJR 4.673 SNIP 3.635
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 3.804 SNIP 2.947
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 3.061 SNIP 2.955
Web of Science (2002): Indexed yes
Scopus rating (2001): SJR 2.736 SNIP 2.747
Scopus rating (2000): SJR 2.609 SNIP 2.269
Web of Science (2000): Indexed yes
Scopus rating (1999): SJR 2.844 SNIP 2.115
Original language: English
Electronic versions:
nbt.2624.pdf
DOIs:
10.1038/nbt.2624
Source: dtu
Source-ID: n::oai:DTIC-ART:npg/389835167::33714
Publication: Research - peer-review › Journal article – Annual report year: 2013

Structural Systems Biology Evaluation of Metabolic Thermotolerance in Escherichia coli
Improve the System A "systems biology" approach may clarify, for example, how particular proteins determine sensitivity of bacteria to extremes of temperature. Chang et al. (p. 1220) integrated information on protein structure with a model of metabolism, thus associating the protein structure of enzymes with their catalyzed metabolic reactions. The effects of temperature on susceptible proteins could be predicted and the key reactions that were likely to mediate sensitivity of bacteria to extremes of temperature were identified. Indeed, engineered thermotolerant proteins could be substituted for sensitive ones to improve the growth of thermosensitive strains of bacteria. Such control could come in handy when engineering strains of bacteria to produce compounds of industrial or therapeutic value.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California, San Diego, University of California, Sanford-Burnham Medical Research Institute
Authors: Chang, R. L. (Ekstern), Andrews, K. (Ekstern), Kim, D. (Ekstern), Li, Z. (Ekstern), Godzik, A. (Ekstern), Palsson, B. (Intern)
Proteomic Analysis of Chinese Hamster Ovary Cells

To complement the recent genomic sequencing of Chinese hamster ovary (CHO) cells, proteomic analysis was performed on CHO cells including the cellular proteome, secretome, and glycoproteome using tandem mass spectrometry (MS/MS) of multiple fractions obtained from gel electrophoresis, multidimensional liquid chromatography, and solid phase extraction of glycopeptides (SPEG). From the 120 different mass spectrometry analyses generating 682 097 MS/MS spectra, 93 548 unique peptide sequences were identified with at most 0.02 false discovery rate (FDR). A total of 6164 grouped proteins were identified from both glycoproteome and proteome analysis, representing an 8-fold increase in the number of proteins currently identified in the CHO proteome. Furthermore, this is the first proteomic study done using the CHO genome exclusively, which provides for more accurate identification of proteins. From this analysis, the CHO codon frequency was determined and found to be distinct from humans, which will facilitate expression of human proteins in CHO cells. Analysis of the combined proteomic and mRNA data sets indicated the enrichment of a number of pathways including protein processing and apoptosis but depletion of proteins involved in steroid hormone and glycosphingolipid metabolism. Five-hundred four of the detected proteins included N-acetylation modifications, and 1292 different proteins were observed to be N-glycosylated. This first large-scale proteomic analysis will enhance the knowledge base about CHO capabilities for recombinant expression and provide information useful in cell engineering efforts aimed at modifying CHO cellular functions.

General information

State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, Department of Systems Biology, Johns Hopkins University, Vanderbilt University, University of California
Pages: 5265-5276
Publication date: 2012
Main Research Area: Technical/natural sciences

Publication information

Journal: Journal of Proteome Research
Volume: 11
Issue number: 11
ISSN (Print): 1535-3893
Ratings:
- BFI (2018): BFI-level 2
- BFI (2017): BFI-level 2
- Web of Science (2017): Indexed yes
- BFI (2016): BFI-level 2
- Scopus rating (2016): CiteScore 4.34 SJR 1.705 SNIP 1.002
- Web of Science (2016): Indexed yes
- BFI (2015): BFI-level 2
- Scopus rating (2015): SJR 1.934 SNIP 1.092 CiteScore 4.45
- Web of Science (2015): Indexed yes
- BFI (2014): BFI-level 2
- Scopus rating (2014): SJR 1.945 SNIP 1.185 CiteScore 4.64
- Web of Science (2014): Indexed yes
- BFI (2013): BFI-level 2
The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line

Chinese hamster ovary (CHO)-derived cell lines are the preferred host cells for the production of therapeutic proteins. Here we present a draft genomic sequence of the CHO-K1 ancestral cell line. The assembly comprises 2.45 Gb of genomic sequence, with 24,383 predicted genes. We associate most of the assembled scaffolds with 21 chromosomes isolated by microfluidics to identify chromosomal locations of genes. Furthermore, we investigate genes involved in glycosylation, which affect therapeutic protein quality, and viral susceptibility genes, which are relevant to cell engineering and regulatory concerns. Homologs of most human glycosylation-associated genes are present in the CHO-K1 genome, although 141 of these homologs are not expressed under exponential growth conditions. Many important viral entry genes are also present in the genome but not expressed, which may explain the unusual viral resistance property of CHO cell lines. We discuss how the availability of this genome sequence may facilitate genome-scale science for the optimization of biopharmaceutical protein production.

General information
State: Published
Organisations: Center for Microbial Biotechnology, Department of Systems Biology, Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, BGI-Shenzhen, GT Life Sciences, Peking University Shenzhen Hospital, University of Delaware, Stanford University
Pages: 735-741
Publication date: 2011
Main Research Area: Technical/natural sciences
Projects:

Integration of Informatics and Metabolic Engineering for the discovery of Novel Antibiotics
Novo Nordisk Foundation Center for Biosustainability
New Bioactive Compounds
Network Reconstruction in Silico Biology

Research Groups
Bacterial Cell Factory Optimization

Fundación MEDINA
Korea Advanced Institute of Science and Technology (KAIST)
Period: 01/03/2017 → 31/03/2023
Number of participants: 12
Acronym: iimena
Project participant:
Weber, Tilmann (Intern)
Palsson, Bernhard (Intern)
Charusanti, Pep (Intern)
Jiang, Xinglin (Intern)
Damborg, Mie (Intern)
Durczak, Oliwia (Intern)
Kontou, Efthychia Eva (Intern)
Lizak, Dawid Mariusz (Intern)
Beck, Charlotte (Intern)
Kjiproski, Darko (Intern)
Rasmussen, Birte Kastrup (Intern)
Project Manager, organisational:
Lohmann, Ricarda (Intern)

Financing sources
Source: Forsk. Private danske - Fonde
Name of research programme: Novo Nordisk Foundation Challenge Program
Web address: http://www.novonordiskfonden.dk
Amount: 58,832,942.00 Danish Kroner
Year of approval: 2017

Relations
Activities:
Lectures on antibiotics biosynthesis: polyketides, aminoglycosides, RiPPs and others
Generation of click-able kirromycin derivatives by exploiting the substrate promiscuity of the discrete acyl transferase KirCII
In silico and experimental approaches to understand and engineer the biosynthesis of antibiotics
In silico and experimental approaches to understand and engineer the biosynthesis of antibiotics
Publications:
Dissemination of antibiotic resistance genes from antibiotic producers to pathogens
Press / Media items:
Video and Blog-post / interview at sciencenews.dk on iimena project (NNF Challenge Grant)
Millions for research into antibiotic resistance and better drugs
Scientists solve 30-year old mystery on how resistance genes spread
Research program on new antibiotics receives 58 M DKK

Press clippings:

**Millions for research into antibiotic resistance and better drugs**
Tilmann Weber, Pep Charusanti, Sang Yup Lee & Bernhard Palsson
26/01/2017 → 26/12/2017

**Description**
DTU Press release on NNF Challenge grants, including IIMENA
Network Reconstruction in Silico Biology, Novo Nordisk Foundation Center for Biosustainability, Big Data 2 Knowledge, New Bioactive Compounds

**Media coverage (2)**

- **Novo-millioner skal gøre lægemidler bedre**
  26/12/2017
  medwatch.dk (National), Denmark, Web
  LONNI PARK LYNGE
  http://medwatch.dk/secure/Medicinal___Biotek/article9319828.ece
  Tilmann Weber, Pep Charusanti, Sang Yup Lee & Bernhard Palsson
  Novo Nordisk Foundation Center for Biosustainability, New Bioactive Compounds, Big Data 2 Knowledge, Network Reconstruction in Silico Biology

- **GODT NYT I KAMPEN MOD ANTIBIOTIKARESISTENS**
  30/01/2017
  Dansk Kemi (National), Denmark, Web
  http://www.kemifokus.dk/godt-nyt-i-kampen-mod-antibiotikaresistens/
  Tilmann Weber, Sang Yup Lee, Pep Charusanti & Bernhard Palsson
  Novo Nordisk Foundation Center for Biosustainability, New Bioactive Compounds, Big Data 2 Knowledge, Network Reconstruction in Silico Biology

**Media contributions (2)**

- **Millions for research into antibiotic resistance and better drugs**
  26/01/2017
  DTU Homepage (International), Denmark, Web
  Viebeke Hempler
  http://www.dtu.dk/english/news/nyhed?id=39201475-593e-41c9-b63e-08c790731768
  Tilmann Weber, Sang Yup Lee, Pep Charusanti & Bernhard Palsson
  Novo Nordisk Foundation Center for Biosustainability, New Bioactive Compounds, Big Data 2 Knowledge, Network Reconstruction in Silico Biology

- **SPILDEVAND FRA 100 LANDE SKAL BIDRAGE TIL BEGRÆNSE ANTIBIOTIKA-RESISTENS**
  26/01/2017
  NNF Homepage (International), Denmark, Web
  http://novonordiskfonden.dk/da/content/spildevand-fra-100-lande-skal-bidrage-til-begraense-antibiotika-resistens
  Press release on 2017 NNF Challenge Grants by Novo Nordisk Foundation
  Tilmann Weber, Sang Yup Lee, Pep Charusanti & Bernhard Palsson
  Novo Nordisk Foundation Center for Biosustainability, New Bioactive Compounds, Big Data 2 Knowledge, Network Reconstruction in Silico Biology

**Relations**

Projects:
Integration of Informatics and Metabolic Engineering for the discovery of Novel Antibiotics

Press / Media