Computational Methods to Assess the Production Potential of Bio-Based Chemicals

Elevated costs and long implementation times of bio-based processes for producing chemicals represent a bottleneck for moving to a bio-based economy. A prospective analysis able to elucidate economically and technically feasible product targets at early research phases is mandatory. Computational tools can be implemented to explore the biological and technical spectrum of feasibility, while constraining the operational space for desired chemicals. In this chapter, two different computational tools for assessing potential for bio-based production of chemicals from different perspectives are described in detail. The first tool is GEM-Path: an algorithm to compute all structurally possible pathways from one target molecule to the host metabolome. The second tool is a framework for Modeling Sustainable Industrial Chemicals production (MuSIC), which integrates modeling approaches for cellular metabolism, bioreactor design, upstream/downstream processes, and economic impact assessment. Integrating GEM-Path and MuSIC will play a vital role in supporting early phases of research efforts and guide the policy makers with decisions, as we progress toward planning a sustainable chemical industry.
Underground metabolism: network-level perspective and biotechnological potential
A key challenge in molecular systems biology is understanding how new pathways arise during evolution and how to exploit them for biotechnological applications. New pathways in metabolic networks often evolve by recruiting weak promiscuous activities of pre-existing enzymes. Here we describe recent systems biology advances to map such ‘underground’ activities and to predict and analyze their contribution to new metabolic functions. Underground activities are prevalent in cellular metabolism and can form novel pathways that either enable evolutionary adaptation to new environments or provide bypass to genetic lesions. We also illustrate the potential of integrating computational models of underground metabolism and experimental approaches to study the evolution of novel metabolic phenotypes and advance the field of biotechnology.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, ALE Technology & Software Development, Network Reconstruction in Silico Biology, Wageningen University, Hungarian Academy of Sciences
Authors: Notebaart, R. A. (Ekstern), Kintses, B. (Ekstern), Feist, A. (Intern), Papp, B. (Ekstern)
Pages: 108-114
Publication date: 2018
Main Research Area: Technical/natural sciences

Publication information
Journal: Current Opinion in Biotechnology
Volume: 49
ISSN (Print): 0958-1669
Ratings:
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
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 3.508 SNIP 2.327 CiteScore 7.93
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 3.313 SNIP 2.089 CiteScore 7.76
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 3.56 SNIP 2.223
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 3.772 SNIP 2.085
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 3.324 SNIP 2.099
Web of Science (2007): Indexed yes
Scopus rating (2007): SJR 3.058 SNIP 1.959
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 2.742 SNIP 2.235
Bacterial cells with improved tolerance to isobutyric acid
Bacterial cells genetically modified to improve their tolerance to certain commodity chemicals, such as isobutyric acid and related compounds, and methods of preparing and using such bacterial cells for production of isobutyric acid and related compounds.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Research Groups, iLoop, Bacterial Cell Factory Optimization, Global Econometric Modeling, Department of Biotechnology and Biomedicine, Bacterial Synthetic Biology, ALE Technology & Software Development
Authors: Lennen, R. (Intern), Nielsen, A. T. (Intern), Herrgård, M. (Intern), Sommer, M. O. A. (Intern), Feist, A. (Intern), Mohamed, E. T. T. (Intern)
Publication date: 16 Nov 2017

Publication information
IPC: C12P 7/52 A I
Patent number: WO2017194696
Date: 16/11/2017
Priority date: 09/06/2016
Priority number: EP20160173673
Original language: English
Main Research Area: Technical/natural sciences
Source: espacenet
Source-ID: WO2017194696
Publication: Research › Patent – Annual report year: 2017

Bacterial cells with improved tolerance to polyamines
Provided are bacterial cells genetically modified to improve their tolerance to certain commodity chemicals, such as polyamines, and methods of preparing and using such bacterial cells for production of polyamines and other compounds.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Research Groups, iLoop, Bacterial Cell Factory Optimization, Global Econometric Modeling, Department of Biotechnology and Biomedicine, Bacterial Synthetic Biology, ALE Technology & Software Development
Authors: Lennen, R. (Intern), Nielsen, A. T. (Intern), Herrgaard, M. (Intern), Sommer, M. O. A. (Intern), Feist, A. (Intern), Tharwat Tolba Mohamed, E. (Intern)
Publication date: 15 Jun 2017

Publication information
IPC: C12P 13/00 A I
Patent number: WO2017097828
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.
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.
Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution

There is a need to replace petroleum-derived with sustainable feedstocks for chemical production. Certain biomass feedstocks can meet this need as abundant, diverse, and renewable resources. Specific ionic liquids (ILs) can play a role in this process as promising candidates for chemical pretreatment and deconstruction of plant-based biomass feedstocks as they efficiently release carbohydrates which can be fermented. However, the most efficient pretreatment ILs are highly toxic to biological systems, such as microbial fermentations, and hinder subsequent bioprocessing of fermentative sugars obtained from IL-treated biomass. To generate strains capable of tolerating residual ILs present in treated feedstocks, a tolerance adaptive laboratory evolution (TALE) approach was developed and utilized to improve growth of two different Escherichia coli strains, DH1 and K-12 MG1655, in the presence of two different ionic liquids, 1-ethyl-3-methylimidazolium acetate ([C2C1Im][OAc]) and 1-butyl-3-methylimidazolium chloride ([C4C1Im]Cl). For multiple parallel replicate populations of E. coli, cells were repeatedly passed to select for improved fitness over the course of approximately 40 days. Clonal isolates were screened and the best performing isolates were subjected to whole genome sequencing. The most prevalent mutations in tolerant clones occurred in transport processes related to the functions of mdtJI, a multidrug efflux pump, and yhdP, an uncharacterized transporter. Additional mutations were enriched in processes such as transcriptional regulation and nucleotide biosynthesis. Finally, the best-performing strains were compared to previously characterized tolerant strains and showed superior performance in tolerance of different IL and media combinations (i.e., cross tolerance) with robust growth at 8.5% (w/v) and detectable growth up to 11.9% (w/v) [C2C1Im][OAc]. The generated strains thus represent the best performing platform strains available for bioproduction utilizing IL-treated renewable substrates, and the TALE method was highly successful in overcoming the general issue of substrate toxicity and has great promise for use in tolerance engineering.

General information

State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, iLoop, ALE Technology & Software Development, Research Groups, Global Econometric Modeling, Network Reconstruction in Silico Biology, Joint Bioenergy Institute
Authors: Mohamed, E. T. (Intern), Wang, S. (Ekstern), Lennen, R. M. (Intern), Herrgård, M. J. (Intern), Simmons, B. A. (Ekstern), Singer, S. W. (Ekstern), Feist, A. M. (Intern)
Number of pages: 15
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Microbial Cell Factories
Volume: 16
Issue number: 1
Article number: 204
ISSN (Print): 1475-2859
Ratings:
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.92 SJR 1.446 SNIP 1.228
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.501 SNIP 1.24 CiteScore 4.08
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.672 SNIP 1.471 CiteScore 4.25
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.

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
Authors: Monk, J. M. (Ekstern), Lloyd, C. J. (Ekstern), Brunk, E. (Ekstern), Mih, N. (Ekstern), Sastry, A. (Ekstern), King, Z. (Ekstern), Takeuchi, R. (Ekstern), Nomura, W. (Ekstern), Zhang, Z. (Ekstern), Mori, H. (Ekstern), Feist, A. M. (Intern),
Increased production of L-serine in Escherichia coli through Adaptive Laboratory Evolution

L-serine is a promising building block biochemical with a high theoretical production yield from glucose. Toxicity of L-serine is however prohibitive for high-titer production in E. coli. Here, E. coli lacking L-serine degradation pathways was evolved for improved tolerance by gradually increasing L-serine concentration from 3 to 100 g/L using adaptive laboratory evolution (ALE). Genome sequencing of isolated clones revealed multiplication of genetic regions, as well as mutations in thrA, thereby showing a potential mechanism of serine inhibition. Other mutations were evaluated by MAGE combined with amplicon sequencing, revealing role of rho, lrp, pykF, eno, and rpoB on tolerance and fitness in minimal medium. Production using the tolerant strains resulted in 37 g/L of L-serine with a 24% mass yield. The resulting titer is similar to the highest production reported for any organism thereby highlighting the potential of ALE for industrial biotechnology.

General Information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Research Groups, Applied Metabolic Engineering, Bacterial Cell Factory Optimization, iLoop, Network Reconstruction in Silico Biology, ALE Technology & Software Development
Number of pages: 24
Pages: 141–150
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Metabolic Engineering
Volume: 39
ISSN (Print): 1096-7176
Ratings:
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
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.
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.
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.
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.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, University of California, Columbia University, Universidad Nacional Autonoma de Mexico
Number of pages: 12
Pages: 260-271
Publication date: 2016
Main Research Area: Technical/natural sciences

Publication information
Journal: Cell
Volume: 2
ISSN (Print): 0092-8674

**Electronic versions:**

10.1016/j.cels.2016.04.003

**DOI:**

Publication: Research - peer-review › Journal article – Annual report year: 2016
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): 0974-7419
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
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

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
Authors: McCloskey, D. (Ekstern), Gangoiti, J. A. (Ekstern), Palsson, B. O. (Intern), Feist, A. M. (Intern)
Number of pages: 13
Pages: 1338-1350
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Metabolomics
Volume: 11
Issue number: 5
ISSN (Print): 1573-3882
Ratings:
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Development Of An Efficient Glycerol Utilizing Saccharomyces Cerevisiae Strain Via Adaptive Laboratory Evolution

With increasing interest in biosustainable technologies, the need for converting available non-saccharide carbon sources most efficiently is emerging. Highly abundant crude glycerol, a major waste residue in biodiesel production, has attracted attention as a cheap carbon source for microbial fermentation processes. The most commonly known microbial cell factory, the yeast Saccharomyces cerevisiae, has been extensively applied for the production of a wide range of scientifically and industrially relevant products using saccharides (mainly glucose) as carbon source. However, it was shown that popular wild-type laboratory yeast strains, commonly applied in metabolic engineering studies, did not grow or grew very slowly in glycerol medium. In this work, an adaptive laboratory evolution approach to obtain S. cerevisiae strains with an improved ability to grow on glycerol was applied. A broad array of evolved strains, which exhibited a significant increase in the specific growth rate and a higher glycerol consumption rate, were isolated. The best performing strains were further analyzed by classical genetics and whole genome re-sequencing in order to understand the molecular basis of glycerol catabolism in yeast. The knowledge acquired in this study may be further applied for rational S. cerevisiae strain improvement for using glycerol as a carbon source in industrial biotechnology processes. This work is a part of the DeYeastLibrary consortium financed by ERA-IB DeYeastLibrary - Designer yeast strain library optimized for metabolic engineering applications http://www.era-ib.net/deyeast-library

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Research Groups, Applied Metabolic Engineering, iLoop, Network Reconstruction in Silico Biology, EMBL Heidelberg
Development of an efficient glycerol utilizing Saccharomyces cerevisiae platform strain via adaptive laboratory evolution

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Research Groups, iLoop, Applied Metabolic Engineering
Authors: Strucko, T. (Intern), Tharwat Tolba Mohamed, E. (Intern), Feist, A. (Intern), Förster, J. (Intern)
Number of pages: 1
Pages: S242-S242
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Yeast
Volume: 32
Issue number: S1
Article number: PS14-29
ISSN (Print): 0749-503X
Ratings:
BFI (2017): BFI-level 1
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): SJR 0.816 SNIP 0.811 CiteScore 1.87
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 0.962 SNIP 0.745 CiteScore 2.01
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 0.875 SNIP 0.792 CiteScore 1.67
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.232 SNIP 0.72 CiteScore 2.09
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.197 SNIP 0.762 CiteScore 2.05
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.063 SNIP 0.701 CiteScore 1.77
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.049 SNIP 0.835
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 1.584 SNIP 0.81
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Fast Swinnex Filtration (FSF): A fast and robust sampling and extraction method suitable for metabolomics analysis of cultures grown in complex media

Liquid chromatography tandem mass spectrometry (LC–MS/MS) provides a powerful means to analyze intracellular metabolism. A prerequisite to accurate metabolomics analysis using LC–MS/MS is a robust sampling and extraction protocol. One unaddressed area in sampling is a detailed examination of a suitable method for anaerobic cultures grown in complex media. Given that a vast majority of bacteria are facultative or obligate anaerobes that grow to low biomass density and need to be cultured in complex media, a suitable sampling and extraction strategy for anaerobic cultures is needed. In this work, we develop a fast-filtration method using pressuredriven Swinnex filters. We show that the method is fast enough to provide an accurate snapshot of intracellular metabolism, reduces matrix interference from the media to improve the number of compounds that can be detected, and is applicable to anaerobic and aerobic liquid cultures grown in a variety of culturing systems. Furthermore, we apply the fast filtration method to investigate differences in the absolute intracellular metabolite levels of anaerobic cultures grown in minimal and complex media.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, iLoop, Network Reconstruction in Silico Biology, University of California
Authors: McCloskey, D. (Ekstern), Utrilla, J. (Ekstern), Naviaux, R. K. (Ekstern), Palsson, B. O. (Ekstern), Feist, A. (Intern)
Number of pages: 12
Pages: 198-209
Publication date: 2015
Main Research Area: Technical/natural sciences

Publication information
Journal: Metabolomics
Volume: 11
Issue number: 1
ISSN (Print): 1573-3882
Ratings:
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
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.
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. And 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.
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
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.
Obtaining optimal cofactor balance to drive production is a challenge metabolically engineered microbial strains. To facilitate identification of heterologous enzymes with desirable altered cofactor requirements from native content, we have developed Cofactory, a method for prediction of enzyme cofactor specificity using only primary amino acid sequence information. The algorithm identifies potential cofactor binding Rossmann folds and predicts the specificity for the cofactors FAD(H2), NAD(H), and NADP(H). The Rossmann fold sequence search is carried out using hidden Markov models whereas artificial neural networks are used for specificity prediction. Training was carried out using experimental data from protein cofactor structure complexes. The overall performance was benchmarked against an independent evaluation set obtaining Matthews correlation coefficients of 0.94, 0.79, and 0.65 for FAD(112), NAD(H), and NADP(H), respectively. The Cofactory method is made publicly available at http://www.cbs.dtu.dk/services/Cofactory.

Cofactory: Sequence-based prediction of cofactor specificity of Rossmann folds

Obtaining optimal cofactor balance to drive production is a challenge metabolically engineered microbial strains. To facilitate identification of heterologous enzymes with desirable altered cofactor requirements from native content, we have developed Cofactory, a method for prediction of enzyme cofactor specificity using only primary amino acid sequence information. The algorithm identifies potential cofactor binding Rossmann folds and predicts the specificity for the cofactors FAD(H2), NAD(H), and NADP(H). The Rossmann fold sequence search is carried out using hidden Markov models whereas artificial neural networks are used for specificity prediction. Training was carried out using experimental data from protein cofactor structure complexes. The overall performance was benchmarked against an independent evaluation set obtaining Matthews correlation coefficients of 0.94, 0.79, and 0.65 for FAD(112), NAD(H), and NADP(H), respectively. The Cofactory method is made publicly available at http://www.cbs.dtu.dk/services/Cofactory.

General information

State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Metagenomics, Novo Nordisk Foundation Center for Biosustainability, iLoop, Network Reconstruction in Silico Biology, Functional Human Variation, Integrative Systems Biology, Novozymes A/S
Authors: Geertz-Hansen, H. M. (Intern), Blom, N. (Intern), Feist, A. (Intern), Brunak, S. (Intern), Petersen, T. N. (Intern)
Pages: 1819-1828
Publication date: 2014
Main Research Area: Technical/natural sciences
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.

**General information**

State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Bacterial Cell Factories, Research Groups, iLoop, Big Data 2 Knowledge, Network Reconstruction in Silico Biology, Department of Systems Biology, Drug Resistance and Community Dynamics, University of California, San Diego, University of California
Pages: 2647-2662
Publication date: 2014
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Molecular Biology and Evolution
Volume: 31
Issue number: 10
ISSN (Print): 0737-4038
Ratings:
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): SJR 8.724 SNIP 7.289 CiteScore 13.93
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 8.142 SNIP 4.662 CiteScore 11.28
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 8.995 SNIP 6.947 CiteScore 14.08
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 6.804 SNIP 5.514 CiteScore 12.36
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 4.901 SNIP 3.176 CiteScore 8.5
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 4.23 SNIP 1.865 CiteScore 6.05
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 6.011 SNIP 3.774
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 5.106 SNIP 2.666
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 4.861 SNIP 1.898
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.
In Escherichia coli, the oxidative branch of the pentose phosphate pathway (oxPPP) is one of the major sources of NADPH when glucose is the sole carbon nutrient. However, unbalanced NADPH production causes growth impairment as observed in a strain lacking phosphoglucoisomerase (Δpgi). In this work, we studied the metabolic response of this bacterium to the replacement of its glucose-6-phosphate dehydrogenase (G6PDH) by an NADH-producing variant. The homologous enzyme from Leuconostoc mesenteroides was studied by molecular dynamics and site-directed mutagenesis to obtain the NAD-prefering LmG6PDH(R46E,Q47E). Through homologous recombination, the zwf loci (encoding G6PDH) in the chromosomes of WT and Δpgi E. coli strains were replaced by DNA encoding LmG6PDH(R46E,Q47E).

Contrary to some predictions performed with flux balance analysis, the replacements caused a substantial effect on the growth rates, increasing 59% in the Δpgi strain, while falling 44% in the WT. Quantitative PCR (qPCR) analysis of the zwf locus showed that the expression level of the mutant enzyme was similar to the native enzyme and the expression of genes encoding key enzymes of the central pathways also showed moderate changes among the studied strains. The phenotypic and qPCR data were integrated into in silico modelling, showing an operative G6PDH flux contributing to the NADH pool. Our results indicated that, in vivo, the generation of NADH by G6PDH is beneficial or disadvantageous for growth depending on the operation of the upper Embden-Meyerhof pathway. Interestingly, a genomic database search suggested that in bacteria lacking phosphofructokinase, the G6PDHs tend to have similar preferences for NAD and NADP. The importance of the generation of NADPH in a pathway such as the oxPPP is discussed.
Optimal cofactor swapping can increase the theoretical yield for chemical production in Escherichia coli and Saccharomyces cerevisiae

Maintaining cofactor balance is a critical function in microorganisms, but often the native cofactor balance does not match the needs of an engineered metabolic flux state. Here, an optimization procedure is utilized to identify optimal cofactor-specificity "swaps" for oxidoreductase enzymes utilizing NAD(H) or NADP(H) in the genome-scale metabolic models of Escherichia coli and Saccharomyces cerevisiae. The theoretical yields of all native carbon-containing molecules are considered, as well as theoretical yields of twelve heterologous production pathways in E. coli. Swapping the cofactor specificity of central metabolic enzymes (especially GAPD and ALCD2x) is shown to increase NADPH production and increase theoretical yields for native products in E. coli and yeast-including l-aspartate, l-lysine, l-isoleucine, l-proline, l-serine, and putrescine-and non-native products in E. coli-including 1,3-propanediol, 3-hydroxybutyrate, 3-hydroxypropanoate, 3-hydroxyvalerate, and styrene. © 2014 International Metabolic Engineering Society.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, iLoop, Network Reconstruction in Silico Biology, University of California
Authors: King, Z. A. (Ekstern), Feist, A. (Intern)
Number of pages: 12
Pages: 117-128
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: Metabolic Engineering
Volume: 24
ISSN (Print): 1096-7176
Ratings:
BFI (2017): BFI-level 2
Web of Science (2017): 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.
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.

General information
State: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Network Reconstruction in Silico Biology, CFB - Core Flow, University of California
Authors: McCloskey, D. (Ekstern), Palsson, B. (Intern), Feist, A. (Intern)
Pages: 661
Publication date: 2013
Main Research Area: Technical/natural sciences

Publication information
Journal: Molecular Systems Biology
Volume: 9
ISSN (Print): 1744-4292
Ratings:
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
Characterization and modelling of interspecies electron transfer mechanisms and microbial community dynamics of a syntrophic association.

Syntrophic associations are central to microbial communities and thus have a fundamental role in the global carbon cycle. Despite biochemical approaches describing the physiological activity of these communities, there has been a lack of a mechanistic understanding of the relationship between complex nutritional and energetic dependencies and their functioning. Here we apply a multi-omic modelling workflow that combines genomic, transcriptomic and physiological data with genome-scale models to investigate dynamics and electron flow mechanisms in the syntrophic association of Geobacter metallireducens and Geobacter sulfurreducens. Genome-scale modelling of direct interspecies electron transfer reveals insights into the energetics of electron transfer mechanisms. While G. sulfurreducens adapts to rapid syntrophic growth by changes at the genomic and transcriptomic level, G. metallireducens responds only at the transcriptomic level. This multi-omic approach enhances our understanding of adaptive responses and factors that shape the evolution of syntrophic communities.
Methanosarcina barkeri is an Archaeon that produces methane anaerobically as the primary byproduct of its metabolism. M. barkeri can utilize several substrates for ATP and biomass production including methanol, acetate, methyl amines, and a combination of hydrogen and carbon dioxide. In 2006, a metabolic reconstruction of M. barkeri, iAF692, was generated based on a draft genome annotation. The iAF692 reconstruction enabled the first genome-Scale simulations for Archaea.

Since the publication of the first metabolic reconstruction of M. barkeri, additional genomic, biochemical, and phenotypic data have clarified several metabolic pathways. We have used this newly available data to improve the M. barkeri metabolic reconstruction. Modeling simulations using the updated model, iMG746, have led to increased accuracy in predicting gene knockout phenotypes and simulations of batch growth behavior. We used the model to examine knockout lethality data and make predictions about metabolic regulation under different growth conditions. Thus, the updated metabolic reconstruction of M. barkeri metabolism is a useful tool for predicting cellular behavior, studying the methanogenic lifestyle, guiding experimental studies, and making predictions relevant to metabolic engineering applications.

Copyright © 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

General information
State: Published
Organisations: University of Illinois at Urbana-Champaign, University of California
Authors: Gonnerman, M. C. (Ekstern), Benedict, M. N. (Ekstern), Feist, A. (Intern), Metcalf, W. W. (Ekstern), Price, N. D. (Ekstern)
Pages: 1070-1079
Publication date: 2013
Main Research Area: Technical/natural sciences

Publication information
Journal: Biotechnology Journal
Volume: 8
Issue number: 9
ISSN (Print): 1860-6768
Ratings:
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
Central oxidoreductase enzymes (e.g., dehydrogenases, reductases) in microbial metabolism often have preferential binding specificity for one of the two major currency metabolites NAD(H) and NADP(H). These enzyme specificities result in a division of the metabolic functionality of the currency metabolites: enzymes reducing NAD+ to NADH drive oxidative phosphorylation, and enzymes reducing NADP+ to NADPH drive anabolic reactions. In this work, we introduce the computational method OptSwap, which predicts bioprocessing strain designs by identifying optimal modifications of the cofactor binding specificities of oxidoreductase enzyme and complementary reaction knockouts. Using the Escherichia coli genome-scale metabolic model iJO1366, OptSwap predicted eight growth-coupled production designs with significantly greater product yields or substrate-specific productivities than designs predicted with gene knockouts alone. These designs were identified for the production of L-alanine, succinate, acetate, and D-lactate under modeled conditions. Simulations predicted that production of L-alanine and D-lactate can be strongly coupled to growth by knocking out three reactions and swapping the cofactor specificity of one oxidoreductase reaction, while growth coupling was not predicted with four or fewer reaction knockouts under identical conditions. A succinate production design and an acetate production design were predicted to have higher maximum growth rates and higher substrate-specific productivities than designs predicted solely with reaction knockouts. The OptSwap formulation can be readily extended to additional organisms, and the constraints enforcing oxidoreductase specificity swaps can be extended to target other specificity sets of interest.
Sulfide-Driven Microbial Electrosynthesis

Microbial electrosynthesis, the conversion of carbon dioxide to organic molecules using electricity, has recently been demonstrated for acetogenic microorganisms, such as Sporomusa ovata. The energy for reduction of carbon dioxide originates from the hydrolysis of water on the anode, requiring a sufficiently low potential. Here we evaluate the use of sulfide as an electron source for microbial electrosynthesis. Abiotically oxidation of sulfide on the anode yields two electrons. The oxidation product, elemental sulfur, can be further oxidized to sulfate by Desulfobulbus propionicus, generating six additional electrons in the process. The eight electrons generated from the combined abiotic and biotic steps were used to reduce carbon dioxide to acetate on a graphite cathode by Sporomusa ovata at a rate of 24.8 mmol/day·m². Using a strain of Desulfuromonas as biocatalyst on the anode resulted in an acetate production rate of 49.9 mmol/day·m², with a Coulombic efficiency of over 90%. These results demonstrate that sulfide can serve effectively as an alternative electron donor for microbial electrosynthesis. © 2012 American Chemical Society.
Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli

Integrated approaches utilizing in silico analyses will be necessary to successfully advance the field of metabolic engineering. Here, we present an integrated approach through a systematic model-driven evaluation of the production potential for the bacterial production organism Escherichia coli to produce multiple native products from different representative feedstocks through coupling metabolite production to growth rate. Designs were examined for 11 unique central metabolism and amino acid targets from three different substrates under aerobic and anaerobic conditions. Optimal strain designs were reported for designs which possess maximum yield, substrate-specific productivity, and strength of growth-coupling for up to 10 reaction eliminations (knockouts). In total, growth-coupled designs could be identified for 36 out of the total 54 conditions tested, corresponding to eight out of the 11 targets. There were 17 different substrate/target pairs for which over 80% of the theoretical maximum potential could be achieved. The developed method introduces a new concept of objective function tilting for strain design. This study provides specific metabolic interventions (strain designs) for production strains that can be experimentally implemented, characterizes the potential for E. coli to produce native compounds, and outlines a strain design pipeline that can be utilized to design production strains for additional organisms.

General information
State: Published
Organisations: University of California
Authors: Feist, A. (Intern), Zielinski, D. C. (Ekstern), Orth, J. D. (Ekstern), Schellenberger, J. (Ekstern), Herrgard, M. (Ekstern), Palsson, B. (Ekstern)
Pages: 173-186
Publication date: 2010
Main Research Area: Technical/natural sciences

Publication information
Journal: Metabolic Engineering
Volume: 12
Issue number: 3
ISSN (Print): 1096-7176
Ratings:
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
Reconstruction of biochemical networks in microorganisms.

Systems analysis of metabolic and growth functions in microbial organisms is rapidly developing and maturing. Such studies are enabled by reconstruction, at the genomic scale, of the biochemical reaction networks that underlie cellular processes. The network reconstruction process is organism specific and is based on an annotated genome sequence, high-throughput network-wide data sets and bibliomic data on the detailed properties of individual network components. Here we describe the process that is currently used to achieve comprehensive network reconstructions and discuss how these reconstructions are curated and validated. This review should aid the growing number of researchers who are carrying out reconstructions for particular target organisms.

General information

State: Published
Organisations: University of California, University of Wisconsin-Madison, University of California, San Diego
Authors: Feist, A. (Intern), Herrgard, M. (Intern), Thiele, I. (Ekstern), Reed, J. L. (Ekstern), Palsson, B. Ø. (Ekstern)
Pages: 129-143
Publication date: 2009
Main Research Area: Technical/natural sciences

Publication information

Journal: Nature Reviews. Microbiology
Volume: 7
Issue number: 2
ISSN (Print): 1740-1526
Ratings:
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
The manner in which microorganisms utilize their metabolic processes can be predicted using constraint-based analysis of genome-scale metabolic networks. Herein, we present the constraint-based reconstruction and analysis toolbox, a software package running in the Matlab environment, which allows for quantitative prediction of cellular behavior using a constraint-based approach. Specifically, this software allows predictive computations of both steady-state and dynamic optimal growth behavior, the effects of gene deletions, comprehensive robustness analyses, sampling the range of possible cellular metabolic states and the determination of network modules. Functions enabling these calculations are included in the toolbox, allowing a user to input a genome-scale metabolic model distributed in Systems Biology Markup Language format and perform these calculations with just a few lines of code. The results are predictions of cellular behavior that have been verified as accurate in a growing body of research. After software installation, calculation time is minimal, allowing the user to focus on the interpretation of the computational results.

**Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox**

**General information**
- State: Published
- Organisations: University of California, University of California, San Diego
- Authors: Beck, S. A. (Forskerdatabase), Feist, A. (Intern), Mo, M. L. (Ekstern), Hannum, G. (Ekstern), Palsson, B. O. (Ekstern), Herrgard, M. (Intern)
- Pages: 727-738
- Publication date: 2007
- Main Research Area: Technical/natural sciences

**Publication information**
- Journal: Nature Protocols (Online)
- Volume: 2
- Issue number: 3
- ISSN (Print): 1750-2799
Projects:

**Generation of Raw Substrate Utilizing Platform Strains**

Nobo Nordisk Foundation Center for Biosustainability

Research Groups

iLoop

Network Reconstruction in Silico Biology

Period: 01/06/2015 → 30/04/2019
Number of participants: 4
PhD Student:
Tharwat Tolba Mohamed, Elsayed (Intern)
Supervisor:
Lennen, Rebecca (Intern)
Feist, Adam (Intern)
Main Supervisor:
Herrgard, Markus (Intern)

**Generation of Raw Substrate Utilizing Platform Strains**

Technical University of Denmark

Period: 01/06/2015 → 31/05/2018
Number of participants: 4
PhD Student:
Tharwat Tolba Mohamed, Elsayed (Intern)
Supervisor:
Feist, Adam (Intern)
Lennen, Rebecca (Intern)
Main Supervisor:
Herrgard, Markus (Intern)

Financing sources

Source: Internal funding (public)
Name of research programme: Samfinansieret - Andet
Project: PhD