

KSBI-BIML 2026

Bioinformatics & Machine Learning(BIML)
Workshop for Life Scientists

생명정보학 & 머신러닝 워크샵 (온라인)



Synthetic biology and AI

김현욱 _ KAIST



KSBI
KOREAN SOCIETY FOR
BIOINFORMATICS

한국생명정보학회



본 강의 자료는 한국생명정보학회가 주관하는 BIML 2026 워크샵을 목적으로 제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우 발생하는 **모든 법적 책임은 행위자 본인에게 있음**을 알립니다.

KSBI-BIML 2026

Bioinformatics & Machine Learning (BIML) Workshop for Life Scientists

한국생명정보학회가 주최하는 BIML-2026 동계 Bioinformatics & Machine Learning 교육 워크숍에 여러분을 초대합니다.

BIML 워크숍은 생명정보학 연구자들이 최신 AI바이오 분야의 인공지능 기반 분석 기술과 바이오 데이터 분석 기법을 이론과 실습을 통해 체계적으로 배울 수 있는 전문 교육 프로그램입니다. 2015년에 시작된 BIML 워크숍은 올해로 12년 차를 맞이하며, 국내 생명정보학 분야의 최초이자 최고 수준의 교육 프로그램으로 자리 잡았습니다. 이번 워크숍은 크게 인공지능바이오(AI바이오) 분야와 디지털바이오 분야, 두 분야로 구성됩니다.

AI바이오 분야에서는 생명정보 분석에 폭넓게 응용되고 있는 다양한 인공지능 기반 자료 모델링 기법을 다룰 예정입니다. 특히, 인공지능 심층학습을 활용한 단백질 구조 예측, 유전체 분석, 신약 개발에 대한 이론 및 실습 강의를 진행됩니다.

또한 디지털바이오 분야에서는 단일세포오믹스, 공간오믹스, 멀티오믹스, 메타오믹스에 대한 강의도 마련되어 있어, 연구자들의 분석 역량 강화에 실질적인 도움을 줄 것으로 기대됩니다.

또한 2024년부터 추가된 의료정보 자료 분석을 다루는 강의를 올해도 지속해서 운영하고자 합니다. 이는 최근 의료정보 자료 분석에 관한 연구 수요 증가를 반영한 것으로, 관련 연구를 수행하는 의과학자 및 의료정보 연구자들에게 유용한 지침을 제공할 것입니다.

또한, 올해도 생명정보학 기술의 다양화에 발맞춰 온라인 강좌를 대폭 확대했습니다. 올해는 무료 강좌 10개를 포함한 총 40개 이상의 강좌가 개설되며, 연구 주제에 맞는 강좌 추천과 강연료 할인 혜택도 제공합니다.

BIML-2026는 국내 주요 연구 중심 대학의 전임 교수 및 각 분야 최고 전문가들의 강의로 구성되어 있으며, 기초 이론부터 최신 연구 동향까지 아우르는 심도 있는 교육의 장이 될 것으로 확신합니다.

여러분의 많은 관심과 참여를 기대합니다!

2026년 2월

한국생명정보학회장 류 성 호

Synthetic biology and AI

합성생물학은 생물학, 공학, 컴퓨터 과학을 결합하여 효소, 생합성 경로, 또는 전체 생물학적 개체를 설계하고 구축하는 매우 다학제적인 분야이다. 그 목표는 새로운 생물학적 시스템을 창조하거나 기존 시스템을 유용한 목적으로 재설계하는 것이며, 이 과정에서 생물학적 도구를 설계할 때 모듈성, 표준화, 확장성이 중요하게 고려된다.

본 강의에서는 합성생물학의 배경을 먼저 설명하고, 합성생물학이 적용 가능한 다양한 분야 중 바이오제조와 대사공학을 중심으로 소개하고자 한다. 또한 이 과정에서 AI가 수행하는 핵심적인 역할을 설명한다. 특히 미생물 대사공학의 목표가 대사 네트워크와 생산 공정 전반을 통합적으로 분석하고 최적화하여 고수율·고생산성 균주를 개발하는 데 있음을 설명한다. 이를 위해 균주 설계부터 공정 최적화까지 전 과정을 포괄하는 10가지 전략을 체계적으로 소개한다. 더불어 이 과정에서 유용하게 활용될 수 있는 대사 네트워크 모델(genome-scale metabolic model)에 대해서도 이론과 실습을 제공한다. 본 강의에서 소개되는 개념과 방법론은 약물 표적 발굴 등 시스템 의학 분야에도 적용 가능하다.

강의는 다음의 내용을 포함한다:

- 합성생물학 기반 바이오제조
- 대사공학 전략
- 대사 네트워크 모델 (Genome-scale metabolic model)
- 합성생물학과 AI 실습

* 참고강의교재:

해당 논문은 수업자료를 통해서 공유 예정

* 교육생준비물:

노트북 (메모리 8GB 이상, 디스크 여유공간 30GB 이상)

* 강의 난이도: 초급/중급

* 강의: 김현욱 교수 (KAIST 생명화학공학과)

Curriculum Vitae

Speaker Name: Hyun Uk Kim, Ph.D.



► Personal Info

Name Hyun Uk Kim
Title Associate Professor
Affiliation KAIST

► Contact Information

Address 291 Daehak-ro, Yuseong-gu, Daejeon 34141
Email ehukim@kaist.ac.kr

Research Interest

Systems biology, Biological networks, Synthetic biology, Systems medicine

Educational Experience

2005 B.S., Biotechnology, Yonsei University, Seoul, Korea
2007 M.S., Chemical & Biomolecular Engineering, KAIST, Daejeon, Korea
2011 Ph.D., Chemical & Biomolecular Engineering, KAIST, Daejeon, Korea

Professional Experience

2023-Present Adjunct Professor, Graduate School of Engineering Biology, KAIST, Daejeon, Korea
2022-Present Adjunct Professor, BioProcess Engineering Research Center, KAIST, Daejeon, Korea
2018-Present Assistant and Associate Professor, Department of Chemical and Biomolecular Engineering, KAIST, Daejeon, Korea
2014-2016 Visiting Senior Researcher, Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark
2011-2018 Postdoctoral Researcher and Research Assistant Professor, Bioinformatics Research Center, KAIST, Daejeon, Korea

Selected Publications (3 maximum)

1. Lim J1, Jung HD1, Park SY, Jeon M, Kim DS, Cho R, Han D, Ryu HS, Kim Y* & Kim HU*. Genome-scale knockout simulation and clustering analysis of drug-resistant breast cancer cells reveal drug sensitization targets. Proceedings of the National Academy of Sciences U S A (PNAS) 122, e2425384122 (June 2025)
2. Kwon MS1, Lee J1 & Kim HU. A machine learning framework for extracting information from biological pathway images in the literature. Metabolic Engineering 86, 1-11 (November 2024)
3. Lee G1, Lee SM1, Lee S, Jeong CW, Song H, Lee SY, Yun H*, Koh Y*, Kim HU*. Prediction of metabolites associated with somatic mutations in cancers by using genome-scale metabolic models and mutation data. Genome Biology 25, 66 (March 2024)

Lecture 1

Synthetic biology and AI

-합성생물학 기반 바이오제조-

Kim, Hyun Uk

Department of Chemical and Biomolecular Engineering
Graduate School of Engineering Biology
Korea Advanced Institute of Science and Technology (KAIST)



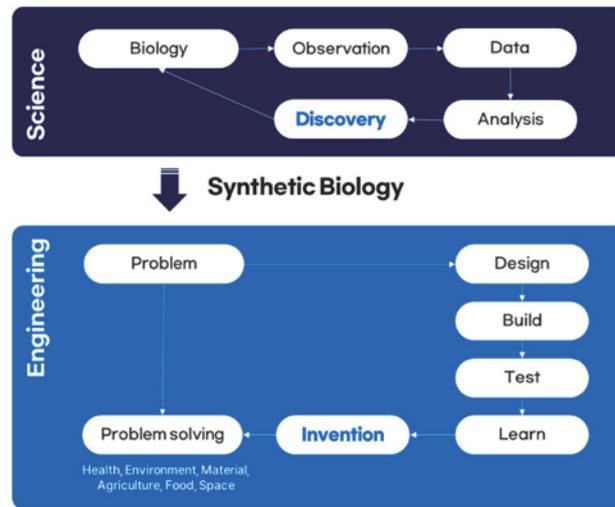
강의계획서

- 학습목표
 - 합성생물학은 생물학, 공학, 컴퓨터 과학을 결합하여 **효소, 생합성 경로, 또는 전체 생물학적 개체를 설계하고 구축하는** 매우 다학제적인 분야이다.
 - 그 목표는 새로운 생물학적 시스템을 창조하거나 기존 시스템을 유용한 목적으로 재설계하는 것이며, 이 과정에서 **생물학적 도구를 설계할 때 모듈성, 표준화, 확장성이 중요하게** 고려된다.

Emergence of synthetic biology as a tool for invention

- The integration of next-generation sequencing since 1996 has greatly enhanced the ability to analyze genomic and gene expression profiles, supported by bioinformatics tools. Synthetic biology adopts a bottom-up approach, using strategic DNA assembly to regulate gene expression for optimizing metabolic pathways. The iGEM competition, launched in 2004, has significantly promoted synthetic biology by engaging global participants in solving challenges and fostering innovation. It leverages the Registry of Standard Biological Parts, with over 20,000 documented parts, catalyzing the growth of startups and reinforcing collaboration between academia and industry.

Fig. 1 The paradigm shift in biological research from discovery to invention through the emergence of synthetic biology. The upper half of the diagram delineates the traditional approach to biological research, which is driven by observation leading to data collection. These data are then analyzed for discovery, signifying a key milestone in the scientific process. The lower half of the diagram depicts the engineering approach driven by synthetic biology. It begins with the identification of a specific problem and progresses through a cyclic process of design, build, test, and learn, resulting in innovative inventions. This shift highlights the evolution of biological research toward a more methodical, problem-solving framework characteristic of engineering disciplines within the scope of synthetic biology



강의계획서

- 학습목표
 - 본 강의에서는 **합성생물학의 배경**을 먼저 설명하고, 합성생물학이 적용 가능한 다양한 분야 중 **바이오제조와 대사공학**을 중심으로 소개하고자 한다.
 - 또한 이 과정에서 **AI**가 수행하는 핵심적인 역할을 설명한다.
 - 특히 미생물 대사공학의 목표가 대사 네트워크와 생산 공정 전반을 통합적으로 분석하고 최적화하여 **고수율·고생산성 균주**를 개발하는 데 있음을 설명한다. 이를 위해 균주 설계부터 공정 최적화까지 전 과정을 포괄하는 **10가지 전략**을 체계적으로 소개한다.
 - 더불어 이 과정에서 유용하게 활용될 수 있는 **대사 네트워크 모델(genome-scale metabolic model)**에 대해서도 이론과 실습을 제공한다.
 - 본 강의에서 소개되는 개념과 방법론은 **약물 표적 발굴 등 시스템 의학 분야**에도 적용 가능하다.

강의계획서

- 강의 1: 합성생물학 기반 바이오제조
- 강의 2: 대사공학 전략
- 강의 3: 대사 네트워크 모델 (Genome-scale metabolic model)
- 강의 4: 합성생물학과 AI 실습

정부, 과기정통부 12대 분야 '국가전략기술 공식 확정'

박진숙 기자
기사승인 2023. 12. 20. 12:00

12대 분야-50개 중점기술 선별
반도체·디스플레이, 이차전지 등
차세대 원자력·양자 등 미래혁신기술

<p>반도체</p> <p>고집적·저항기반 메모리 고성능·저전력 인공지능 반도체 전력반도체 반도체 첨단패키징 차세대 고성능 센서 프리폼 디스플레이 무기발광 디스플레이 반도체·디스플레이 소재·부품·장비 리튬이온전지 및 핵심소재 차세대 이차전지 소재·셀 이차전지 모듈·시스템 이차전지 재사용·재활용 자율주행시스템 도심항공교통(UAM) 전기·수소차 소형모듈형원자로(SMR) 선원원자력시스템·폐기물관리</p>	<p>첨단 바이오</p> <p>합성생물학 유전자·세포 치료 감염병 백신·치료 디지털 헬스데이터 분석·활용 대형 단단연소사이클 엔진 우주관측·센싱 달착륙·표면탐사 첨단 항공가스터빈 엔진·부품 해양자원탐사 수전해 수소생산 수소 저장·운송 수소연료전지 및 발전 데이터·AI 보안 디지털 취약점 분석·대응 네트워크·클라우드 보안 산업·가상융합 보안</p>	<p>인공지능</p> <p>효율적 학습 및 시인프라 고도화 첨단 AI 모델링-의사결정(인공지능)추론 안전·신뢰 AI 산업 활용·혁신 AI 5G 고도화(5G-Adv) 6G 오픈랜(Open-RAN) 5G·6G 고효율 통신부품 5G·6G 위성통신 로봇 정밀제어·구동 부품·SW 로봇 자율이동 고난도 자율조작 인간·로봇 상호작용 가상제조 양자컴퓨팅 양자통신 양자센싱</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<https://www.asiatoday.co.kr/view.php?key=20231220010012582>



보도자료

과학기술정보통신부
과학기술의
영광을 위하여

2023. 10. 30.(월)
15:30
2023. 10. 30.(월)
09:00
보도시점 (2023. 10. 31.(화) 초
간) 예보

바이오를 넘어 미래 산업과 인류의 삶을 바꿀 합성생물학 육성 및 확산 본격 추진

- 과기정통부 「합성생물학 핵심기술개발 및 확산전략」 현장발표회 개최
- 산업적으로 활용 가능한 인공세포 제작을 위한 6대 분야 핵심기술 중점 확보
 - ※ 세계 최고 대비 기술수준 (' 20) 75% → (' 30) 90%로 도약
 - 세포개발 및 대사최적화 기술 등 우리의 강점 기술은 세계 최고수준으로 고도화
- 의료분야 혁신, 오염물질 분해·대체, 고부가 소재 생산 등 합성생물학을 실제 산업에 적용하여 기존 한계를 뛰어넘는 9대 선도프로젝트 추진
- 2030년까지 합성생물학 기반 바이오 신물질* 100개 개발, 세계 최초 상용화 5개 달성 목표

* 합성제원료 생산 미생물, 혁신적인 유전자기위, 온실가스 분해 미생물 등

<https://www.msit.go.kr/bbs/view.do?mId=113&mPid=238&bbsSeqNo=94&nttSeqNo=3183610>

□ 9대 선도프로젝트 개요

합성생물학을 실제 산업에 적용하여 기존의 한계를 뛰어넘는

9대 선도프로젝트 추진

<p>1 미생물 활용 필수 의약소재 생산 및 치료기술 개발</p> <p>의료분야 혁신</p> <p>1. 미생물이 만드는 의약학 소재 후보물질 5개 개발(-28) ※ 항암제, 독성제 치료 2. 유용 광대 미생물이 약물을 생산 및 전달하는 신개념 치료기술 개발(-26)</p>	<p>2 혁신적 항체 설계 및 생산 플랫폼 개발</p> <p>1. 인공지능 항체설계 플랫폼 개발(-28) → 항체 개발시간 10배 단축 2. 혁신적항체 플랫폼 개발(-30) → 복합기능 항체 생산효율 10배 향상</p>	<p>3 RNA, 유전자기위 활용 혁신 치료기술 개발</p> <p>1. mRNA 기반 치료 핵심 요소 기술 국산화(-28) ※ Cap 유사제, RNA 전사제, Poly(A) 등 2. 고효율 유전자기위 활용 유전질환 치료기술 개발(-28) ※ 유전성 망막질환, 희귀 난치 유전질환 등</p>
<p>4 온실가스-고부가소재 생물학적 전환</p> <p>오염물질 분해, 대체</p> <p>1. 상용화 가능한 온실가스 생물학적 전환기술 확보(-28) ※ CO₂ 전환: 50% 향상, 20% 이상 전환, 유산율 1.2배 향상</p>	<p>5 미생물로 난분해 플라스틱 분해</p> <p>1. 난분해 플라스틱(PE, PVC, PS) 분해-미생물 탐색 바이오센서 개발(-27) 2. 신규 미생물 3종 발굴 및 개발(-32)</p>	<p>6 미생물로 화학농약, 필수비료 대체</p> <p>1. 유해선충 제어 미생물 10종 발굴(-28) → 대량 생산기술 확보(-30) 2. 필수비료 대체 가능 미생물 3종 발굴(-30) ※ 질소-인-인산 기반 미생물</p>
<p>7 대체식품용 천연 단백질 소재 개발</p> <p>고부가소재 생산</p> <p>1. 미생물을 활용하여 고기와 유사한 맛을 내는 바이오소재 10종 개발(-26) 2. 효소를 대체하는 동물 단백질 배양 혁신 플랫폼 생산기술 확보(-28)</p>	<p>8 바이오소재* 현장생산 무세포시스템 개발(-30)</p> <p>* 인공적, 합성·자연 등</p> <p>1. 현장전단 무세포 바이오센서 정확도, 민감도 향상(-30) ※ 유전자 증폭기형(PCR) 4중으로 향상 (기존 PCR 대비 0.01% 민감도)</p>	<p>9 유전자 편집을 통한 식물 경합성 효율 50% 향상(-30)</p> <p>※ 기존 4.6% - 목표: 7% 1. 고부가 바이오소재 생산 혁신적인 식물세포 플랫폼 개발(-28)</p>

THE WHITE HOUSE



SEPTEMBER 12, 2022

Executive Order on Advancing Biotechnology and Biomanufacturing Innovation for a Sustainable, Safe, and Secure American Bioeconomy

BRIEFING ROOM | PRESIDENTIAL ACTIONS

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. Policy. It is the policy of my Administration to coordinate a whole-of-government approach to advance biotechnology and biomanufacturing towards innovative solutions in health, climate change, energy, food security, agriculture, supply chain resilience, and national and economic security. Central to this policy and its outcomes are principles of equity, ethics, safety, and security that enable access to technologies, processes, and products in a manner that benefits all Americans and the global community and that maintains United States technological leadership and economic competitiveness.

<https://www.whitehouse.gov/briefing-room/presidential-actions/2022/09/12/executive-order-on-advancing-biotechnology-and-biomanufacturing-innovation-for-a-sustainable-safe-and-secure-american-bioeconomy/>

PRODUCT MONOGRAPH
INCLUDING PATIENT MEDICATION INFORMATION

WEGOVY®

semaglutide injection

Solution for Subcutaneous Injection in a pre-filled pen

Produced by recombinant DNA technology in *Saccharomyces cerevisiae*

Single-use pre-filled pen delivering doses of 0.25 mg, 0.5 mg, 1 mg, 1.7 mg or 2.4 mg

0.25 mg/pen (0.25 mg/0.5 mL)
0.5 mg/pen (0.5 mg/0.5 mL)
1 mg/pen (1 mg/0.5 mL)
1.7 mg/pen (1.7 mg/0.75 mL)
2.4 mg/pen (2.4 mg/0.75 mL)

and

Multi-use pre-filled pen (FlexTouch®) delivering doses of 0.25 mg, 0.5 mg, 1 mg, 1.7 mg or 2.4 mg

1 mg/pen (0.68 mg/mL)
2 mg/pen (0.68 mg/mL or 1.34 mg/mL)
4 mg/pen (1.34 mg/mL)
6.8 mg/pen (2.27 mg/mL)
9.6 mg/pen (3.2 mg/mL)

Glucagon-like Peptide-1 (GLP-1) Receptor Agonist

Novo Nordisk Canada Inc.
101-2476 Argenteia Road
Mississauga, Ontario
L5N 6M1

Date of Authorization:
8 April 2025

Submission Control No: 284023

Wegovy™ (semaglutide injection) Product Monograph

Page 1 of 73



https://www.novonordisk.ca/content/dam/nncorp/ca/en/products/Wegovy-product-monograph.pdf?utm_source=chatgpt.com

Keywords in this class

- Metabolic engineering
 - Definition
 - History
 - Target chemicals with examples
- Sustainability - Sustainable Development Goals (SDGs)
- Synthetic biology
 - Definition
 - History

Metabolic engineering vs Synthetic biology

- In essence, metabolic engineering can be seen as a specialized subset within the larger field of synthetic biology, focusing more on optimizing biological pathways for production purposes, while synthetic biology has a wider scope, aiming to create new and innovative biological systems.
- Metabolic engineering:
 - Metabolic engineering is the practice of optimizing genetic and regulatory processes within cells to increase the production of specific substances (e.g., biofuels, pharmaceuticals, or industrial chemicals). It involves the modification of existing metabolic pathways or the introduction of new ones, often through genetic engineering, to enhance the efficiency of biochemical reactions within a living organism.
- Synthetic biology:
 - Synthetic biology is a broader, interdisciplinary field that combines biology, engineering, and computer science to design and construct new biological entities, such as enzymes, genetic circuits, or even whole organisms. Its goal is to create new biological systems or redesign existing ones for useful purposes, often with a focus on modularity, standardization, and scalability in the creation of biological tools.

Metabolic engineering vs Synthetic biology

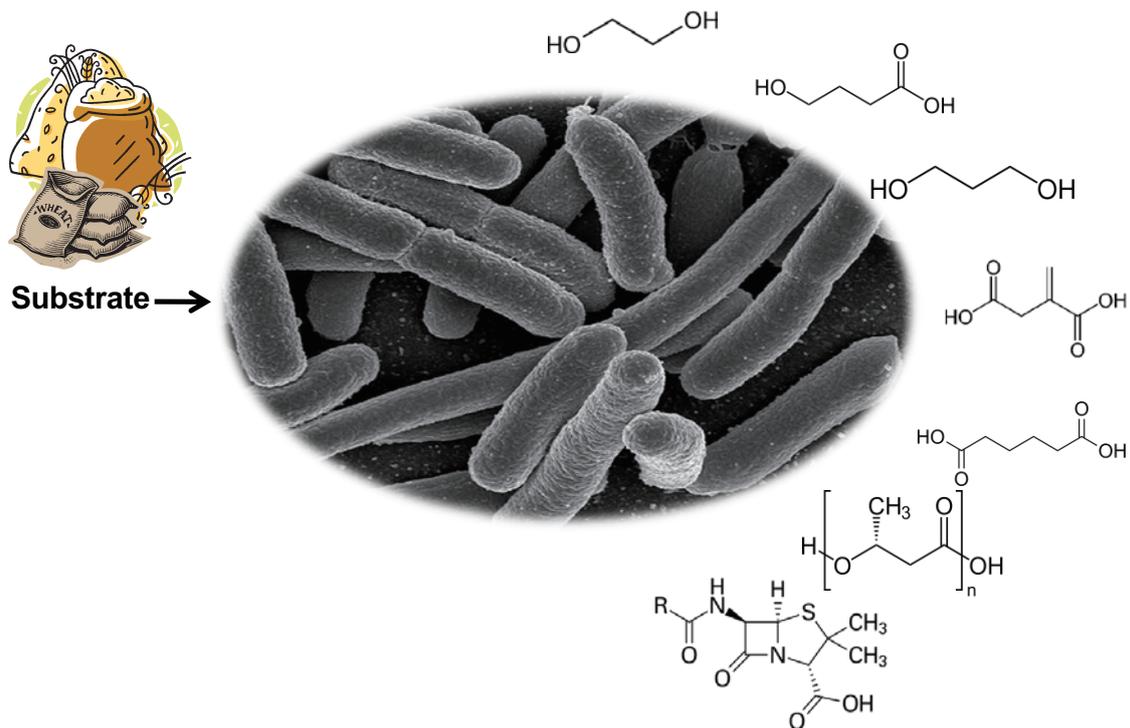
- Similarities:
 - Both metabolic engineering and synthetic biology involve the modification of organisms at the genetic level. They often use tools from molecular biology and biotechnology.
 - Both fields aim to solve practical problems, such as producing biofuels, pharmaceuticals, or environmentally friendly chemicals, using living organisms.
 - Both fields combine biology, engineering, and computational methods to achieve their goals.
 - Both fields contribute significantly to biotechnology, particularly in areas like biomanufacturing and green chemistry.

Metabolic engineering vs Synthetic biology

- Differences:
 - Scope:
 - Metabolic engineering focuses specifically on modifying metabolic pathways to improve or create specific product outputs within an organism.
 - Synthetic biology is broader and encompasses not just metabolic pathways but also the creation of entirely new biological systems or organisms.
 - Approach:
 - Metabolic engineering tends to **optimize existing pathways**, making small, targeted adjustments.
 - Synthetic biology often designs new systems or pathways from scratch, following engineering principles like **modularity and standardization**.
 - Focus:
 - Metabolic engineering is mainly focused on improving yields of specific chemicals or substances.
 - Synthetic biology has a broader aim, which includes building new biological parts, redesigning organisms, and solving complex problems like developing **biosensors or constructing artificial cells**.



To convert biomass feedstock into a valuable bioproduct



Brief history of biotechnology

- “Ancient” biotechnology (Pre-1800s): Biotechnology has its roots in ancient civilizations, where [fermentation processes](#) were used for making bread, beer, and wine. These processes, although not understood in terms of modern science, represented the early application of biotechnology.
- 19th century: The foundation for modern biotechnology was laid with [the discovery of microorganisms](#) by Louis Pasteur and the development of the germ theory of disease. This period also saw the development of vaccines and the understanding of fermentation at a microbial level.
- Early 20th century: Biotechnology advanced with the discovery of enzymes and their roles in metabolism. This period also saw the development of antibiotics, starting with [penicillin](#) in the 1920s, marking a significant milestone in medical biotechnology.
- 1970s: The advent of [recombinant DNA technology](#) marked the beginning of modern biotechnology. Herbert Boyer and Stanley Cohen’s successful genetic engineering experiments in 1973 opened the door to the manipulation of DNA and the development of genetically modified organisms (GMOs).
- 1980s to Present: The field has seen rapid advancements with the development of polymerase chain reaction (PCR) for amplifying DNA, [sequencing](#) of the human genome, and the emergence of [CRISPR-Cas9](#) as a tool for genome editing. Biotechnology now spans multiple sectors including healthcare, agriculture, industrial processes, and environmental remediation.

Brief history of metabolic engineering

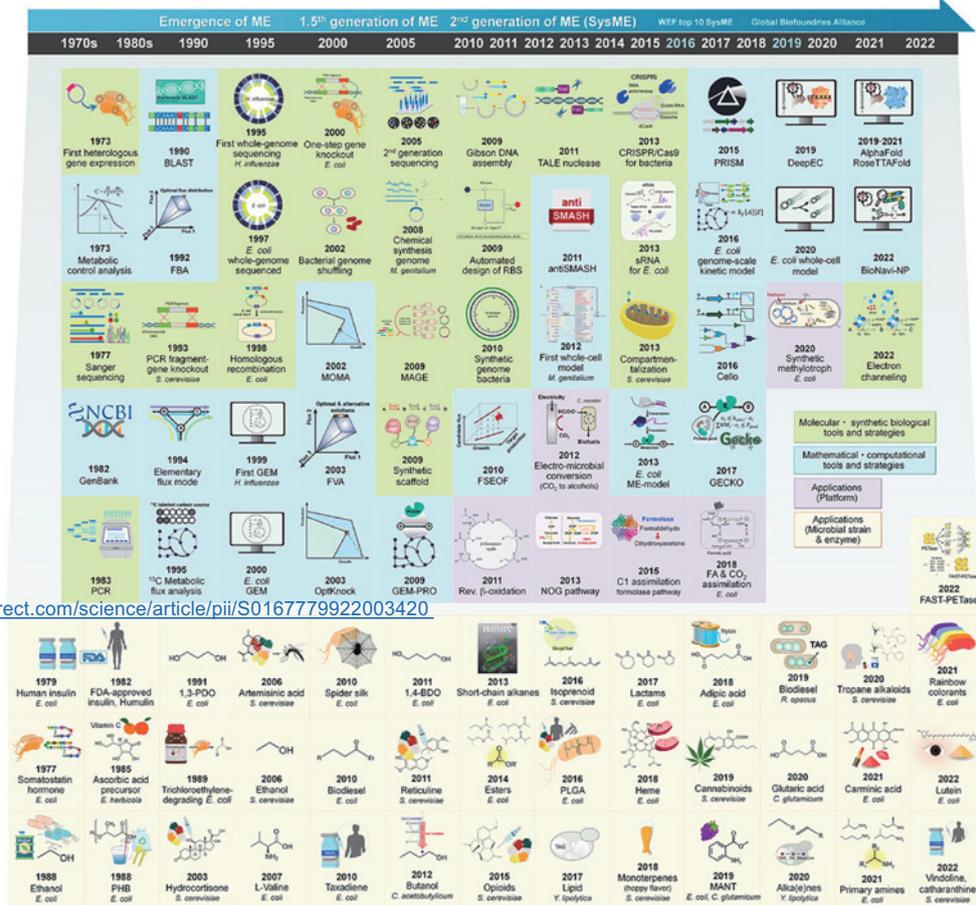
- Early 20th century: The concept of metabolic engineering, though not named as such, began with efforts to understand and manipulate metabolic pathways for improved fermentation processes.
- 1990s: The term “metabolic engineering” was officially coined, marking the field’s formal recognition. This period saw the integration of **genetic engineering** techniques to modify microbial metabolism for the production of desired compounds, such as pharmaceuticals, biofuels, and fine chemicals.
- 2000s to present: Metabolic engineering has evolved with the advent of **systems biology, synthetic biology, and high-throughput genetic engineering techniques**. These advancements have allowed for the precise manipulation of metabolic pathways and the development of microorganisms and plants with enhanced production capacities for various compounds. The integration of computational models has further refined the predictive capabilities in designing and optimizing metabolic pathways.

- **10,000-7,000 BC** Wine making developed in Eastern Mediterranean.
- **7,000-5,000 BC** Beer developed in Egypt and Babylon.
- **5,000 BC** Cheese making and some medicinal plants were developed.
- **4,000 BC** Vinegar was referenced in old testament.
- **500 AD** Algae was cultivated for food by Aztecs.
- **500 AD** Yogurt, sauces and fermented meats were developed.
- **1600 AD** The name fermentation was used.
- **1680 AD** Van Leeuwenhoek observed yeast cells in alcohol fermentation.
- **1781 AD** Pressed Baker’s yeast produced by Dutch (this was the first improvement process in handling Baker’s yeast).
- **1789 AD** Jenner demonstrated vaccination against smallpox infection (First immunization).
- **1837 AD** Cagniard-Latour, Schwann and Kutzing independently hypothesized that yeast is a living thing. (First knowledge on cell biology).
- **1847 AD** Blondeau studied fermentation of lactic acid, butyric acid, acetic acid and urea. He hypothesized that different fermentations carried out by different organisms (fungi?).
- **1857 AD** Pasteur demonstrated that living yeast cells ferment sugar into ethanol and carbon dioxide. Pasteur noted cylindrical organisms produce butyric acid only in absence of oxygen. (First knowledge on anaerobic fermentation).
- **1859 AD** Darwin published the Origin of Species.
- **1877 AD** Pasteur noted relationship between microbes / infectious disease. (First knowledge on pathogenic organisms).

Osama O. Ibrahim, "The history of biotechnology"

- **1916 AD** Germany produces baker's yeast grown on molasses as protein supplement. Also, produced glycerol by yeast fermentation.
- **1881 AD** Koch developed methods for handling microbial cultures (First knowledge on microbial technique and maintenance).
- **1881 AD** First commercial production of lactic acid by anaerobic fermentation.
- **1894 AD** Takamine patented a process to isolate diastase enzymes from molds that can break down starch. (First knowledge on enzymes it's applications).
- **1918 AD** Great Britain produced acetone and butanol by anaerobic fermentation process.
- **1923 AD** Commercial production of citric acid by surface cultures.
- **1929 AD** Fleming demonstrated that mold contaminant in a petri-dish causes bacterial death. (Fist discovery of microbial antibiotics)
- **1934 AD** Gautheret successfully cultured plant cells. (First knowledge on plant tissue culture).
- **1940 AD** Florey and Chain isolated penicillin, elucidated its structure and demonstrated its bacterial properties on G(+) bacteria.
- **1940s** Waxman discovered streptomycin's. and its properties on G(-) bacteria (first microbial screening method for new antibiotics discovery)
Waxman also discovered vitamin B12.
- **1950s** The production of Cortisone at the cost of \$200/g (now it's cost is \$16/g). The production of polio and peruses vaccines.
- **1960s** The production of Xanthus gum. The production of alkine protease for detergents industry.
- **1970s** The discovery of glucose isomerase and the production of high fructose corn syrup (HFCS) as a sweetener.

Osama O. Ibrahim, "The history of biotechnology"



Science 21 June 1991:
Vol. 252 no. 5013 pp. 1668-1675
DOI: 10.1126/science.2047876

< Prev | Table of Contents | Next >

Toward a science of metabolic engineering

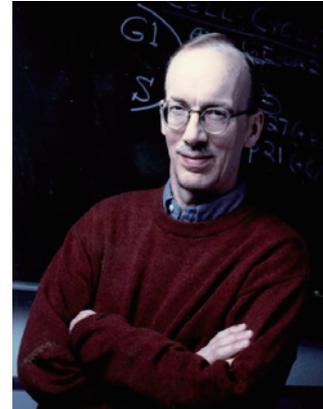
JE Bailey

 Author Affiliations

California Institute of Technology, Pasadena 91125.

ABSTRACT

Application of recombinant DNA methods to restructure metabolic networks can improve production of metabolite and protein products by altering pathway distributions and rates. Recruitment of heterologous proteins enables extension of existing pathways to obtain new chemical products, alter posttranslational protein processing, and degrade recalcitrant wastes. Although some of the experimental and mathematical tools required for rational metabolic engineering are available, complex cellular responses to genetic perturbations can complicate predictive design.



Jay Bailey (1944-2001)

<https://www.science.org/doi/10.1126/science.2047876>

Metabolic Engineering

“... **directed modification of cellular metabolism and properties** through the introduction, deletion, and/or modification of metabolic pathways by using recombinant DNA and other molecular biological techniques.”

“...considering reactions **in their entirety** rather than in isolation from one another.”

Lee and Papoutsakis. *The Challenges and promise of metabolic engineering*
in *Metabolic Engineering*. Marcel Dekker (1999)

Stephanopoulos et al. *The essence of metabolic engineering*
in *Metabolic engineering – Principles and methodologies*. Academic Press (1998)

Top 10 emerging technologies of 2016

SPECIAL REPORT

The Top 10 Emerging Technologies of 2016

Ten technologies that showcase the power of innovation to improve lives, transform industries and safeguard our planet

June 23, 2016 | Credit: World Economic Forum



Systems Metabolic Engineering Turns Microbes into Factories

Unlike fossil fuels, chemicals made from microorganisms are indefinitely renewable and emit relatively little greenhouse gas

June 23, 2016 — Lee Sang Yup

'Systems metabolic engineering' selected in 2016

<https://www.scientificamerican.com/report/the-top-10-emerging-technologies-of-2016/>



Definitions Not fully agreeable

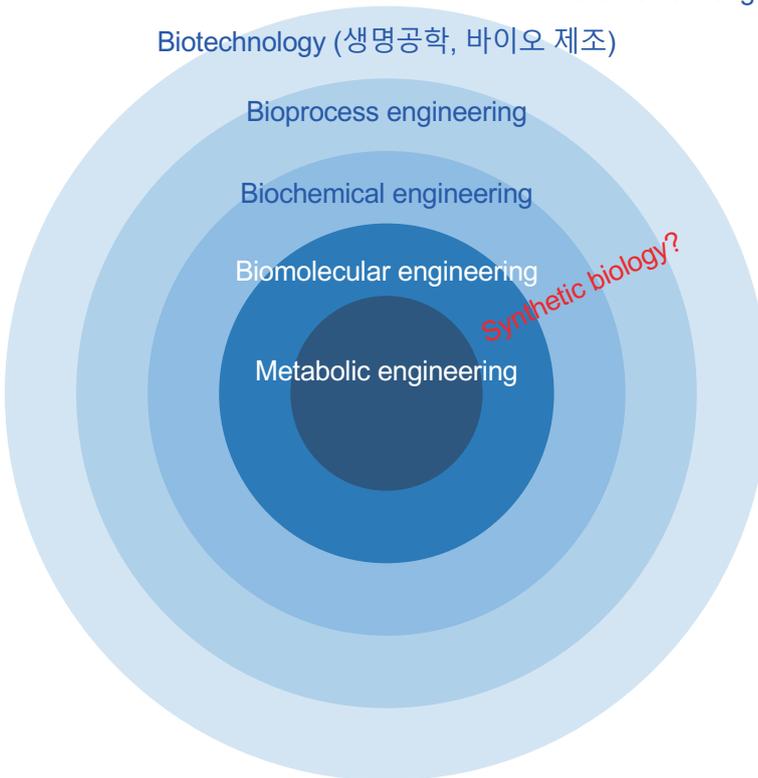
- **Biotechnology**: "... use or development of methods of direct genetic manipulation for a socially desirable goal. Such goals might be the production of a particular chemical, but they may also involve the production of better plants or seeds, or gene therapy, or the use of specially designed organisms to degrade wastes."
- **Bioengineering**: "... is a broad title and includes work on medical, bioprocess, agricultural and environmental systems; its practitioners include agricultural, electrical, mechanical, industrial, environmental and chemical engineers, and others."
- **Biological engineering**: "... is similar, but emphasizes applications to plants and animals."
- **Metabolic engineering**: "... the design of cells with genetically altered pathways to make small molecules that are often novel for that cell."
- **Biochemical engineering**: "... usually meant the extension of chemical engineering principles to systems using a biological catalyst to bring about desired chemical transformations. It is often subdivided into bioreaction engineering and bioseparations, particularly for production of biologics, chemicals, and fuels."
- **Biomedical engineering**: "... has been considered to be separate from biochemical engineering, although the boundary between the two is increasingly vague, particularly in the areas of cell surface receptors and animal cell culture. It focuses on the human body with an emphasis on application of engineering principles from a variety of disciplines to design medical devices, synthetic organs, novel methods for drug delivery and development of diagnostics and instrumentation."
- **Biomolecular engineering**: "... defined by the National Institutes of Health as '... research at the interface of biology and chemical engineering and is focused at the molecular level.'"
- **Bioprocess engineering**: "... includes the work of mechanical, electrical, environmental, and industrial engineers to apply the principles of their disciplines to processes based on using living cells or subcomponents of such cells."

From: Shuler & Kargi, 3rd Edition, Draft

When it comes to “chemical production”, I see it as...

- Bioengineering
- Biological engineering
- Biomedical engineering

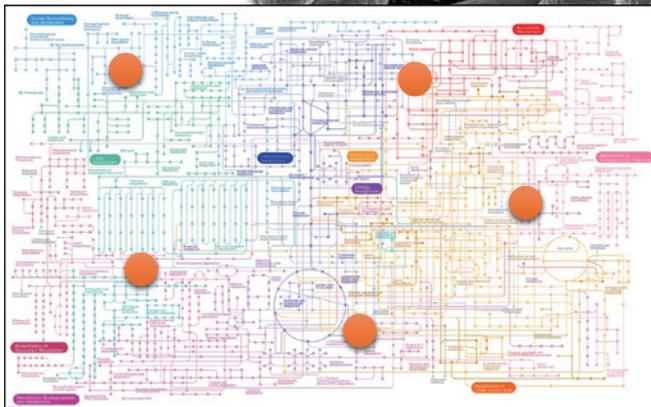
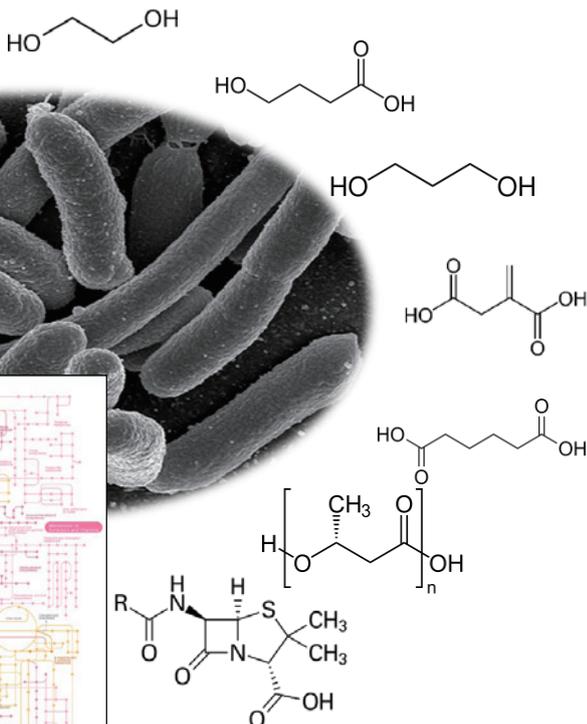
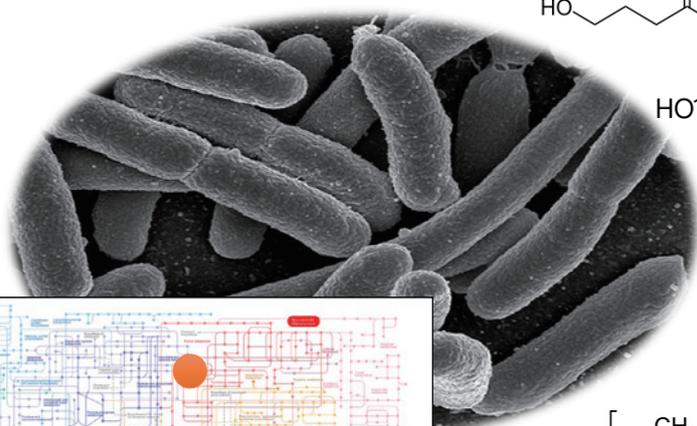
“*Bioprocess engineering*, also *biochemical engineering*, is a specialization of chemical engineering or biological engineering. It deals with the design and development of equipment and processes for the manufacturing of products such as agriculture, food, feed, pharmaceuticals, nutraceuticals, chemicals, and polymers and paper from biological materials & treatment of waste water.”
(Wikipedia)



A complex metabolic network is involved to convert biomass feedstock into a valuable bioproduct



Substrate →



Metabolic network

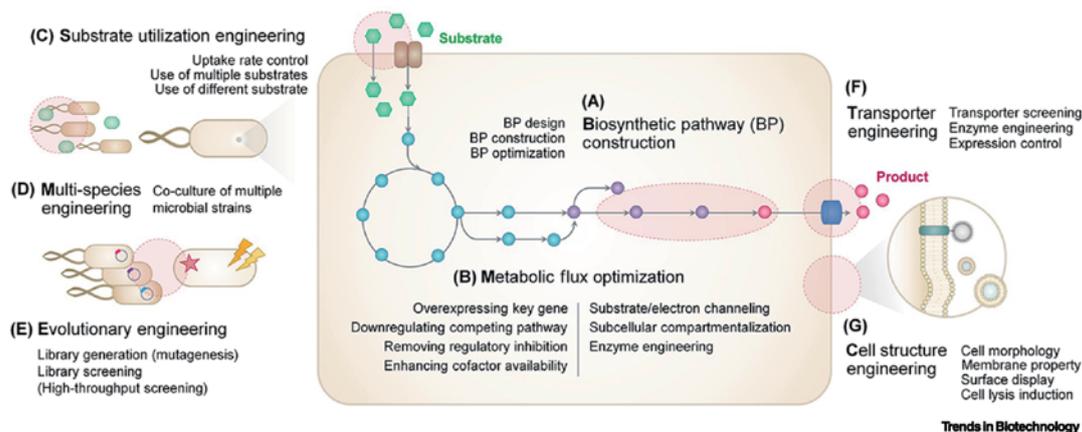
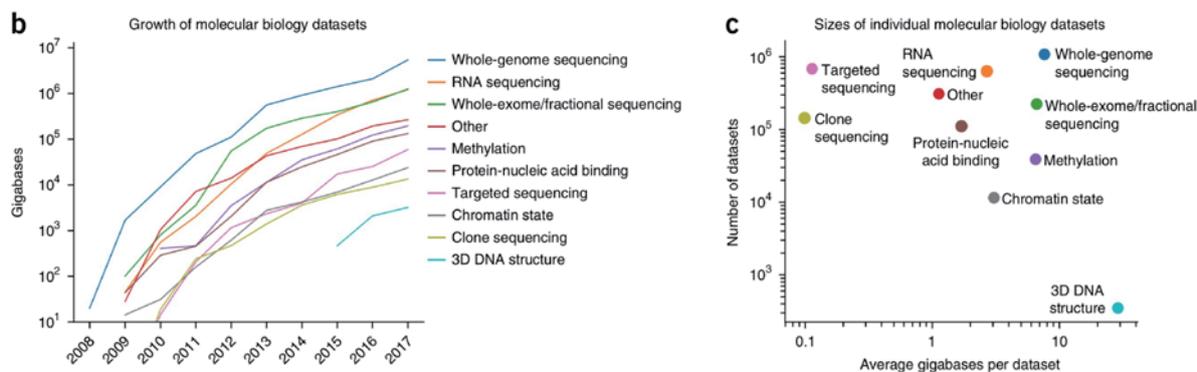


Figure 2. Representative metabolic engineering strategies that focus on microbial strain development. (A) Biosynthetic pathway construction includes biosynthetic pathway design, construction, and optimization processes. (B) Strategies for metabolic flux optimization includes overexpressing key genes, downregulating or removing competing pathways, removing regulatory inhibition, enhancing cofactor availability, substrate and electron channeling, subcellular compartmentalization, and enzyme engineering. (C) Strategies for substrate utilization engineering includes controlling substrate uptake rate and making cells to utilize multiple substrates or different substrates. (D) Multi-species engineering includes the co-culture of multiple microbial strains considering the multicellular interactions. (E) Evolutionary engineering involves adaptive laboratory evolution for cells and also library generation and screening processes for enzymes and pathways that can be accelerated by the use of high-throughput equipment. (F) Transporter engineering includes transporter screening, enzyme (transporter) engineering, and control of the expression levels of transporters. (G) Cell structure engineering includes modulation of cell morphology and membrane characteristics, and also cell surface display system and cell lysis induction system. The strategies are usually implemented in combination and repeatedly to optimize the cellular metabolic network for achieving the engineering goals such as efficient production of useful chemicals and degradation of toxic compounds.

<https://www.sciencedirect.com/science/article/pii/S016779922003420>

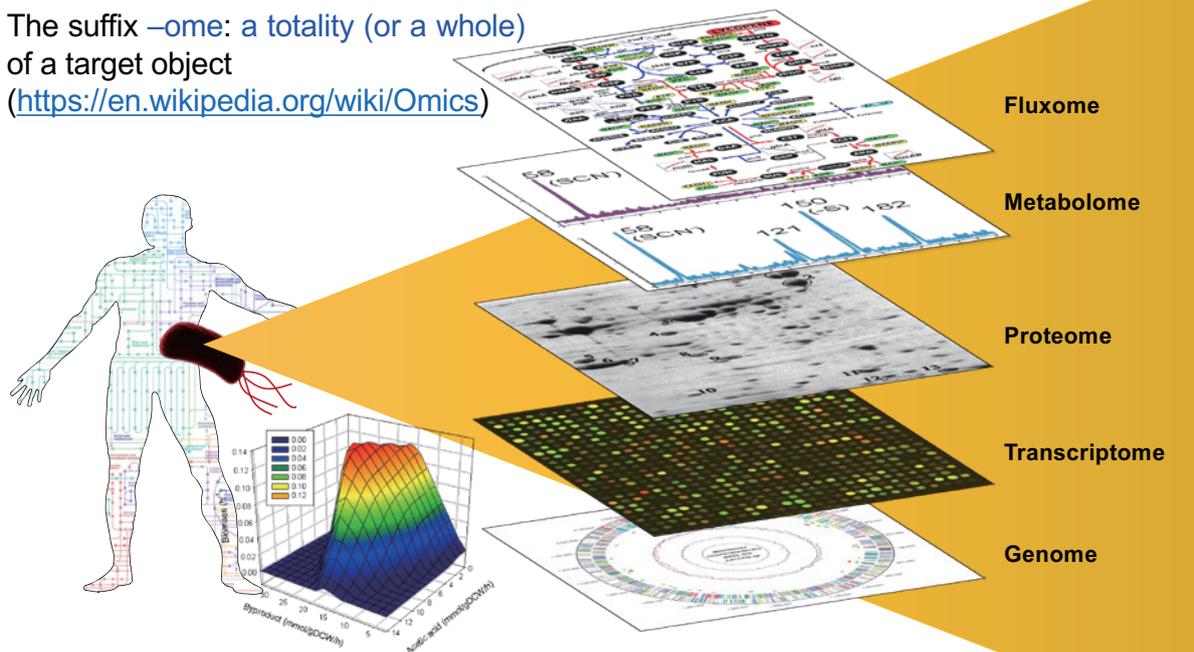


They are mostly omics data!

Wainberg et al. Nature Biotechnology (2018)
<https://doi.org/10.1038/nbt.4233>

Systems biology

The suffix **-ome**: a totality (or a whole) of a target object
<https://en.wikipedia.org/wiki/Omics>

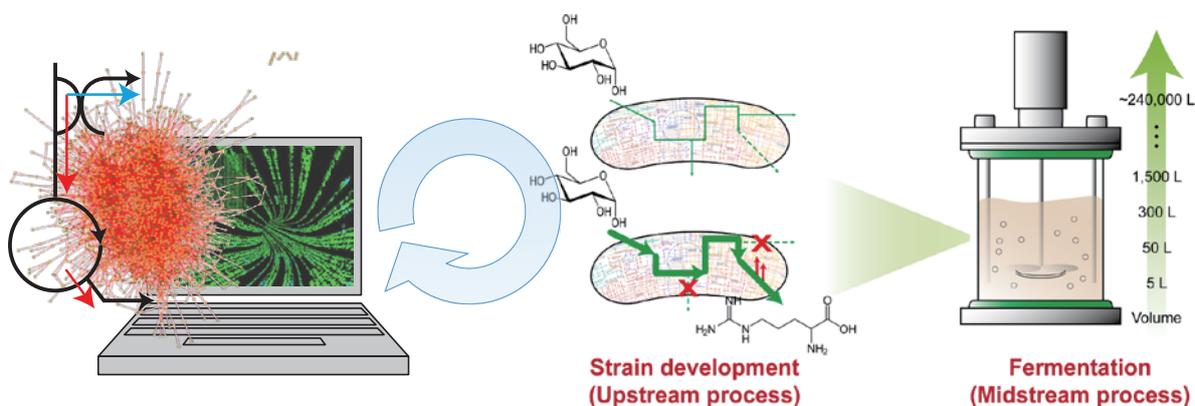


- Large-scale quantitative study of biological systems
- **Biological system**: Cells (a major target), tissues, body or population
- Computational and/or high-throughput experimental technologies
- Computational modeling for “Emergent properties”

Core Technology of the 4th IR



Systems metabolic engineering



Dry experiments

- Biological data analysis (e.g., omics)
- Metabolic modeling & simulations
- Designing strategies (e.g., gene targeting)
- Software development

Wet experiments

- Metabolic engineering
- Production of biochemicals
- Bioprocess engineering

 **GINKGO BIOWORKS** (<https://www.ginkgobioworks.com>)
THE ORGANISM COMPANY



ENGINEERING MADE SIMPLE

Integrated software and automation tools allow for scaling up our process while scaling down costs.



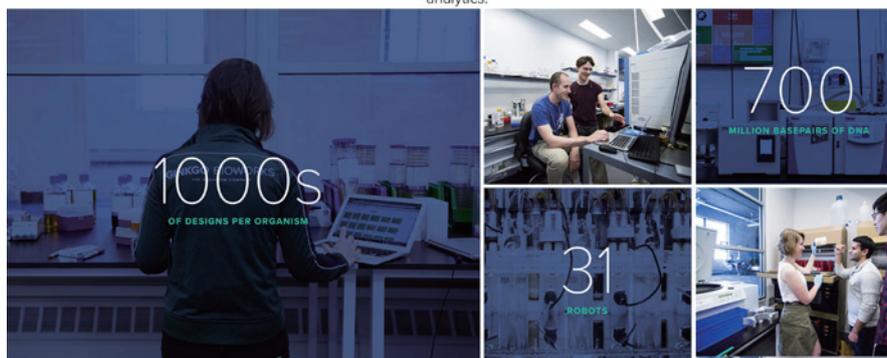
STATE OF THE ART TECHNOLOGY

The enabling tools for biotechnology are rapidly evolving. Our foundries leverage the best of DNA synthesis, laboratory automation, and high performance analytics.



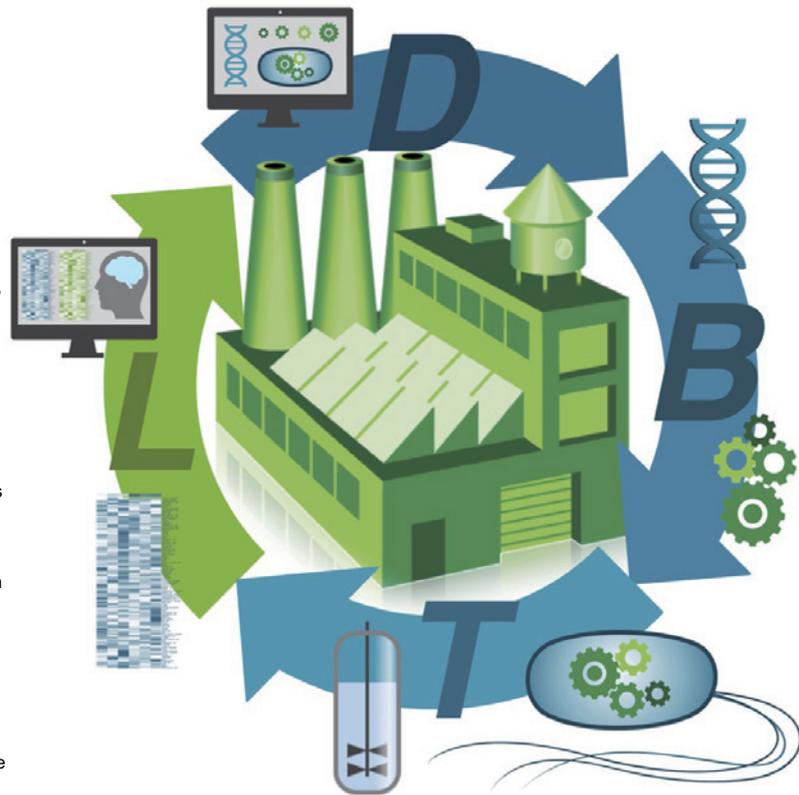
THOUSANDS OF BIOLOGICAL DESIGNS

The foundries allow organism designers to build and test many prototype enzyme pathways to hone in on the most effective design.



The Design-Build-Test-Learn (DBTL) biological engineering cycle.

In simple terms the DBTL framework aims to fulfill particular design criteria for a synthetic biology application, which might for example be the production of a specific product at an optimal titer or the detection of a specific clinical biomarker using an engineered gut microbiome. The cycle begins with **D (Design)**, which defines the desired target function/specifications and involves the computational design of genetic parts, circuits, regulatory and metabolic pathways to whole genomes; **B (Build)** involves the physical assembly of those designed genetic components; **T (Test)** involves the prototyping and testing of the assembled genetic designs in living cells (also called "chassis") at different scales, which also includes comprehensive analytical measurements ("omics") of specific cellular components. This can also include testing components in cell-free extract systems; **L (Learn)** is the application of modeling and computational learning tools, which uses the data obtained in T to inform the design process. Iterations of the DBTL cycle results in genetic designs that aim to fulfill the design specifications established in D. In the figure the DBTL cycle is depicted around an imagined biofactory or biorefinery where many products will be produced using more sustainable and circular economic processes forming the future infrastructure for a global bioeconomy. (Credit: Christopher Johnson, DOE Agile BioFoundry, Golden, CO, USA)



Hillson et al. Nature Communications (2019)

<https://www.nature.com/articles/s41467-019-10079-2>

DTU Biosustain

ABOUT | BIOFOUNDRY | **RESEARCH** | INNOVATION | INFORMATICS | EDUCATION | PRESS

Research Areas | Research Groups | Open science | Workflow | Scientific Tools | Publications

RESEARCH AREA

Sustainable chemicals

Explore the research area

RESEARCH AREA

Natural Products

Explore the research area

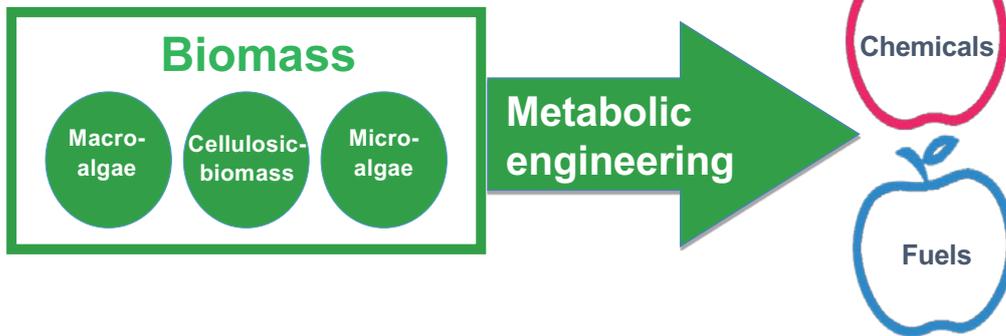
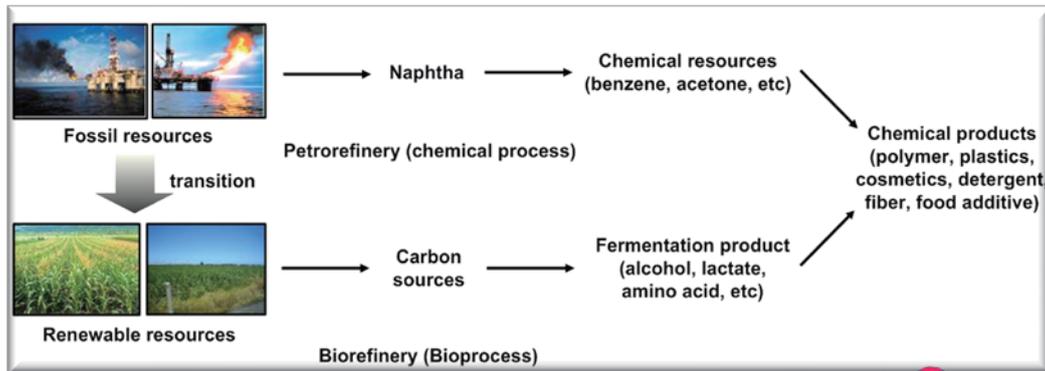
RESEARCH AREA

Microbial Foods

Explore the research area

<https://www.biosustain.dtu.dk/research/research-areas>

Transition towards biorefinery



nature biotechnology

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature biotechnology](#) > [news](#) > article

News | [Published: 15 February 2024](#)

Microbe-made jet fuel

[Emily Waltz](#)

Nature Biotechnology **42**, 163–166 (2024) | [Cite this article](#)

2857 Accesses | **27** Altmetric | [Metrics](#)

The aviation industry readies to embrace aviation fuel produced by fermentation.

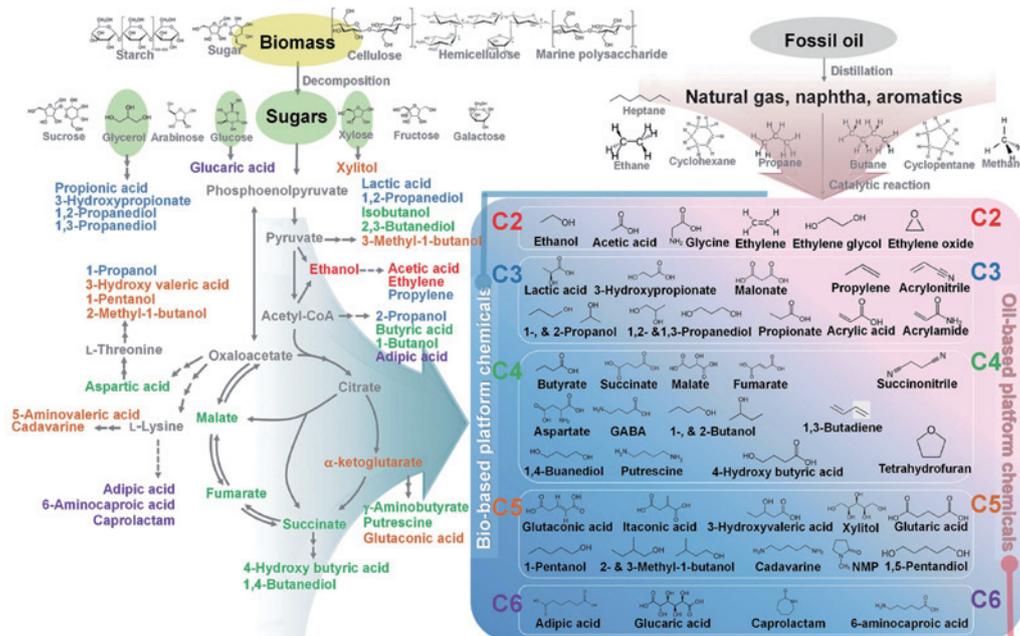
<https://www.nature.com/articles/s41587-024-02136-z>

Table 1 | Selected companies capturing or converting waste gas for fuel and chemical production using biotechnologies such as gas fermentation and enzymes

Company (location)	Gas feedstock	Capture or conversion technology	Products
LanzaTech (Chicago)	Carbon monoxide, carbon dioxide	Fermentation of <i>Clostridium autoethanogenum</i>	Ethanol for use in jet fuel, textiles, shoe soles, packaging, cleaning products, surfactants and detergents
Visolis (Hayward, California)	Gas, biomass and municipal solid waste	Fermentation of undisclosed microorganisms followed by chemical catalysis	Mevalonic acid as an intermediate for production of rocket fuel, aviation fuel, synthetic rubber and specialty chemicals
Krajete (Pasching, Austria)	Carbon dioxide	Fermentation of undisclosed archaea	Methane for use as natural gas
Electrochaeta (Planegg, Germany)	Carbon dioxide	Fermentation of the archaeon <i>Methanothermobacter thermoautotrophicus</i>	Methane for use as natural gas
Industrial Microbes (Alameda, California)	Methane	Fermentation of engineered <i>E. coli</i> to express an enzyme that catalyzes the direct oxidation of methane to methanol at standard temperature and pressure; pyrolysis and distillation are used to convert methanol to desired products such as acrylic acid, methacrylate and acrylonitrile	Acrylic acid, methacrylate and acrylonitrile
Phase Biolabs (Nottingham, UK)	Carbon dioxide	Fermentation of undisclosed microorganisms	Ethanol
Mango Materials (Redwood City, California)	Methane	Fermentation of undisclosed methanotrophs	PHAs for use in biodegradable fibers and goods
Newlight Technologies (Huntington Beach, California)	Carbon dioxide or methane	Fermentation of undisclosed, non-GMO microorganisms	Biodegradable polymer PHB
Circe (Boston)	Carbon dioxide	Fermentation of undisclosed engineered microorganisms	Biodegradable fatty acid polymers such as PHAs
Ucaneo (Berlin)	Carbon dioxide	Carbon capture using carbonic anhydrase enzymes	Captured carbon
Novonosis (Bagsværd, Denmark) and Saipem (Milan)	Carbon dioxide	Carbon capture using carbonic anhydrase enzymes	Captured carbon

PHA, polyhydroxyalkanoate; PHB, polyhydroxybutyrate. Source: SynBioBeta/Fiona Mischel

<https://www.nature.com/articles/s41587-024-02136-z>

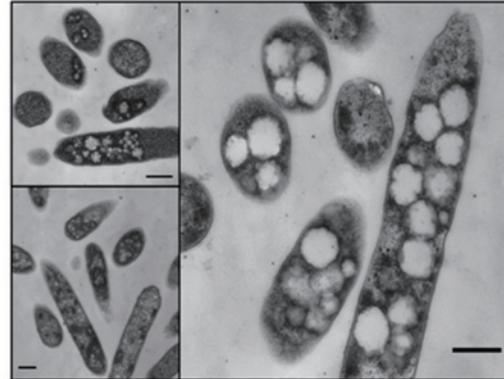
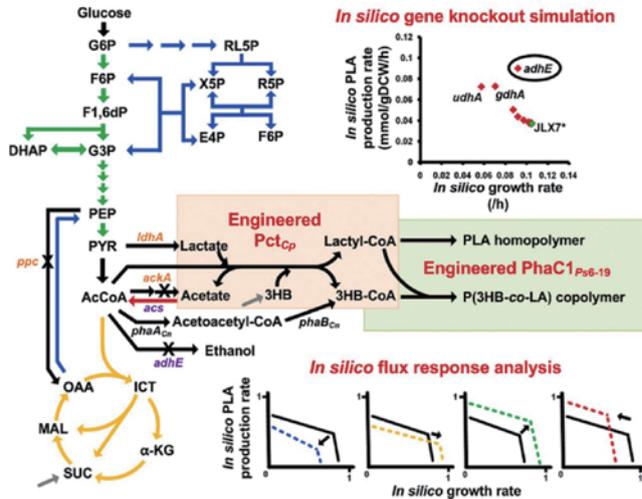


Jang et al., Biotechnol. Bioeng. (2012)

<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/bit.24599>

Biopolymers

Production of polylactic acid homopolymer and P(3-hydroxybutyrate-co-lactate) copolymers by metabolically engineered *E. coli* strains under various conditions



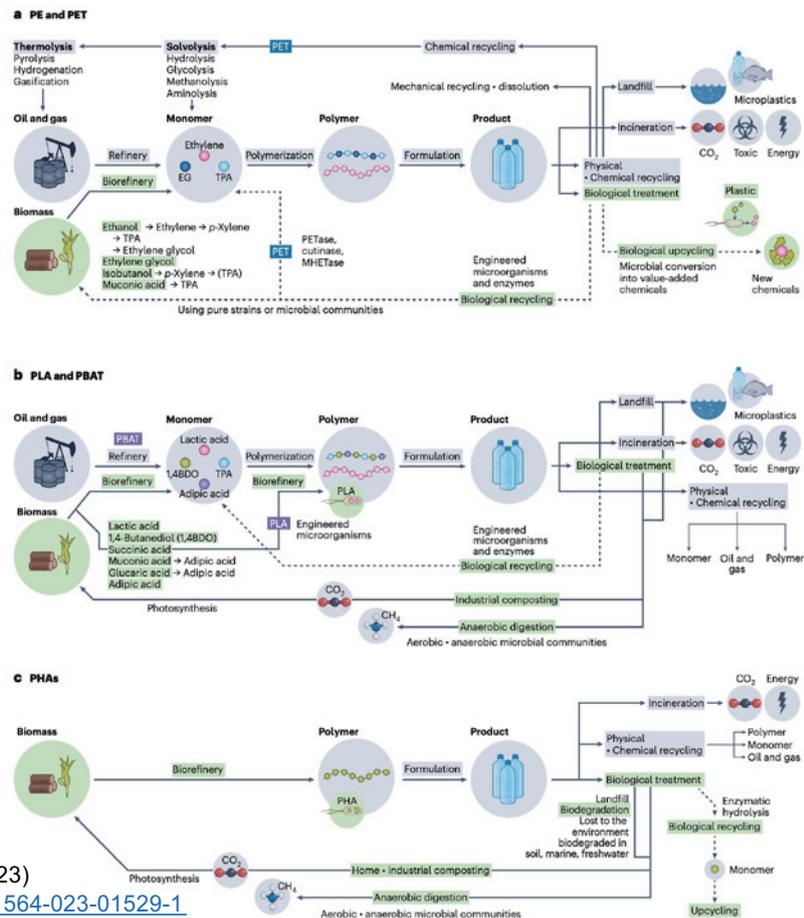
Transmission electron micrographs of intracellular PLA homopolymer and P(3HB-co-LA) copolymer granules

The overall metabolic network shown together with the introduced metabolic pathways for the production of PLA homopolymer and P(3HB-co-LA) copolymer in *E. coli* by combining enzyme engineering and metabolic engineering strategies.

Jung et al. Biotechnol. Bioeng. (2010)

<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/bit.22548>

- Life cycle of plastics is now considered beyond strain development.
- This paper also provides a comprehensive examples of microbial production of plastic precursors and polymers (Table 1).

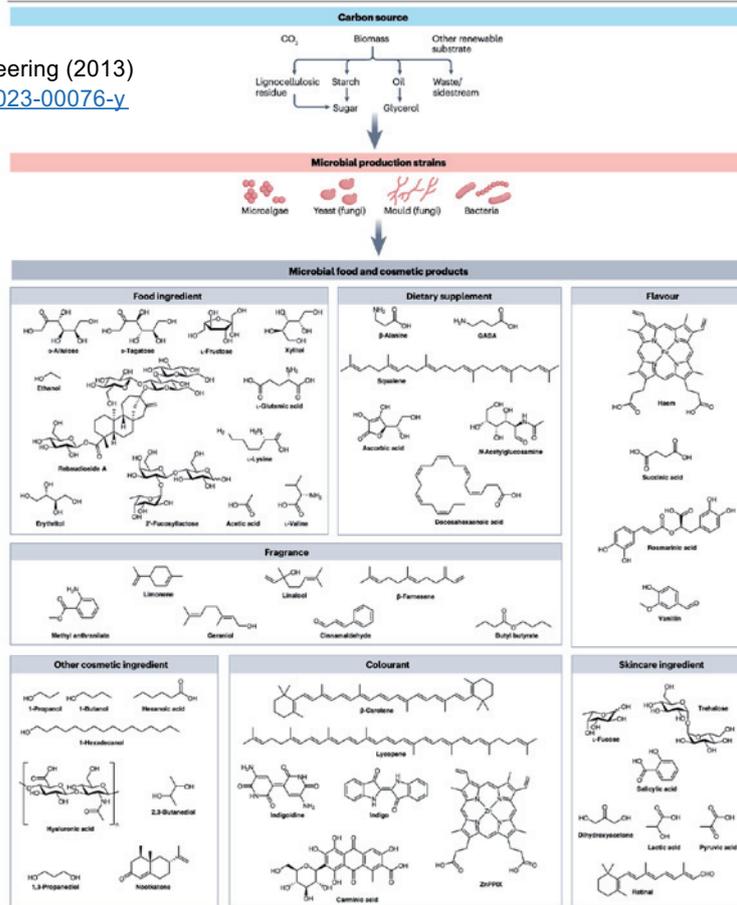


Choi et al. Nature Microbiology (2023)

<https://www.nature.com/articles/s41564-023-01529-1>

Choi and Lee. Nature Reviews Bioengineering (2013)
<https://www.nature.com/articles/s44222-023-00076-y>

- Various food and cosmetic products producible using microbes is displayed (right).
- This paper also provides a comprehensive examples of food and cosmetic compounds production using microorganisms (Table 1).



SUSTAINABLE DEVELOPMENT GOALS



<https://sdgs.un.org/goals>

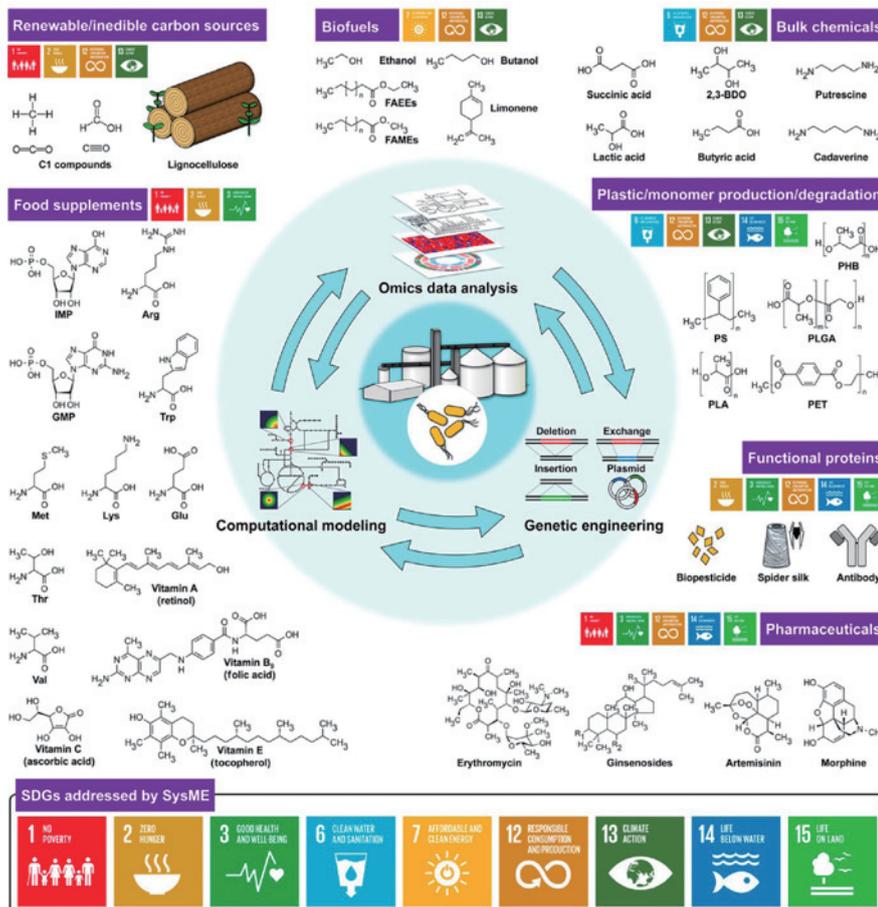


SUSTAINABLE DEVELOPMENT GOALS

17 GOALS TO TRANSFORM OUR WORLD

The 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015, provides a shared blueprint for peace and prosperity for people and the planet, now and into the future. At its heart are the 17 Sustainable Development Goals (SDGs), which are an urgent call for action by all countries - developed and developing - in a global partnership. They recognize that ending poverty and other deprivations must go hand-in-hand with strategies that improve health and education, reduce inequality, and spur economic growth – all while tackling climate change and working to preserve our oceans and forests.

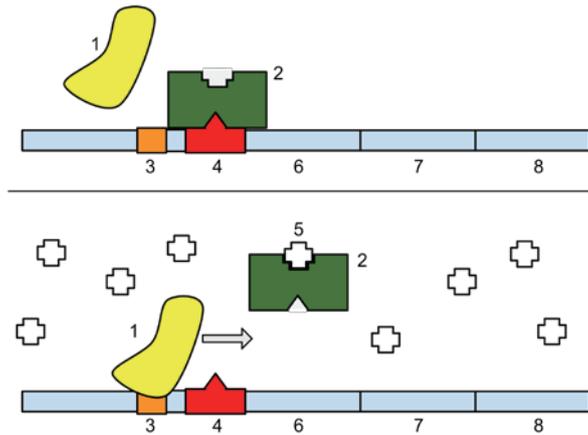
<https://sdgs.un.org/goals>



Yang et al.
Microbial Biotechnology
(2017)

<https://ami-journals.onlinelibrary.wiley.com/doi/10.1111/1751-7915.12766>

The roots of synthetic biology can be traced to the study of the *lac* operon in *E. coli*.

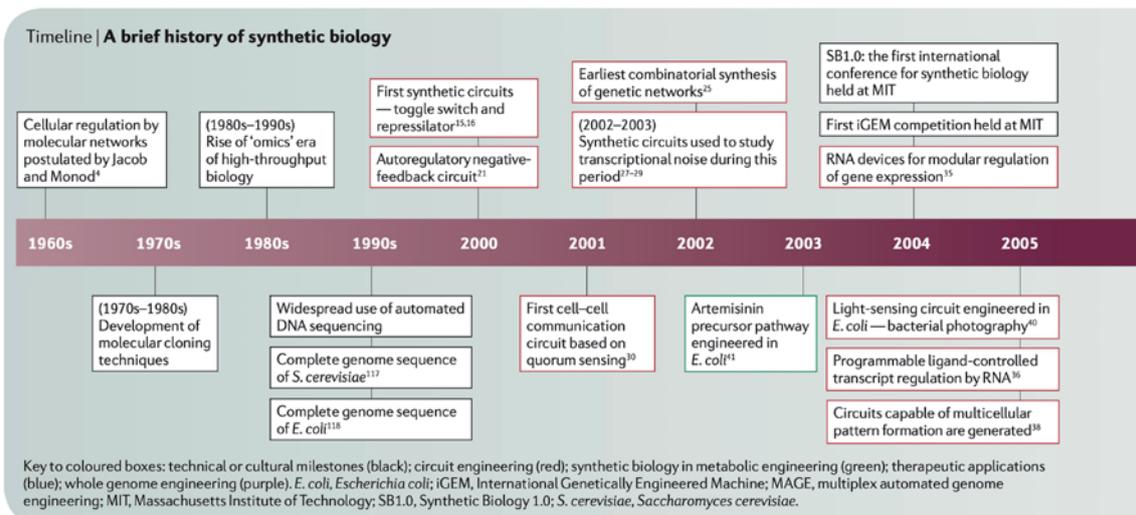


1: RNA Polymerase, 2: Repressor, 3: Promoter, 4: Operator, 5: Lactose, 6: lacZ, 7: lacY, 8: lacA.

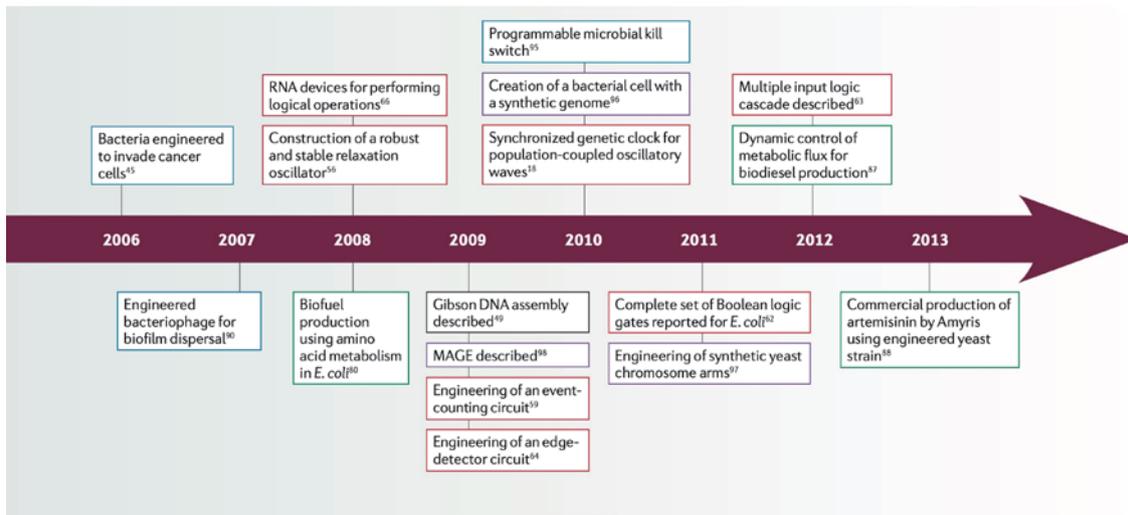
Top: The gene is essentially turned off. There is no allolactose to inhibit the *lac* repressor, so the repressor binds tightly to the operator, which obstructs the RNA polymerase from binding to the promoter, resulting in no *lacZYA* mRNA transcripts.

Bottom: The gene is turned on. Allolactose inhibits the repressor, allowing the RNA polymerase to bind to the promoter and express the genes, resulting in production of LacZYA. Eventually, the enzymes will digest all of the lactose, until there is no allolactose that can bind to the repressor. The repressor will then bind to the operator, stopping the transcription of the LacZYA genes.

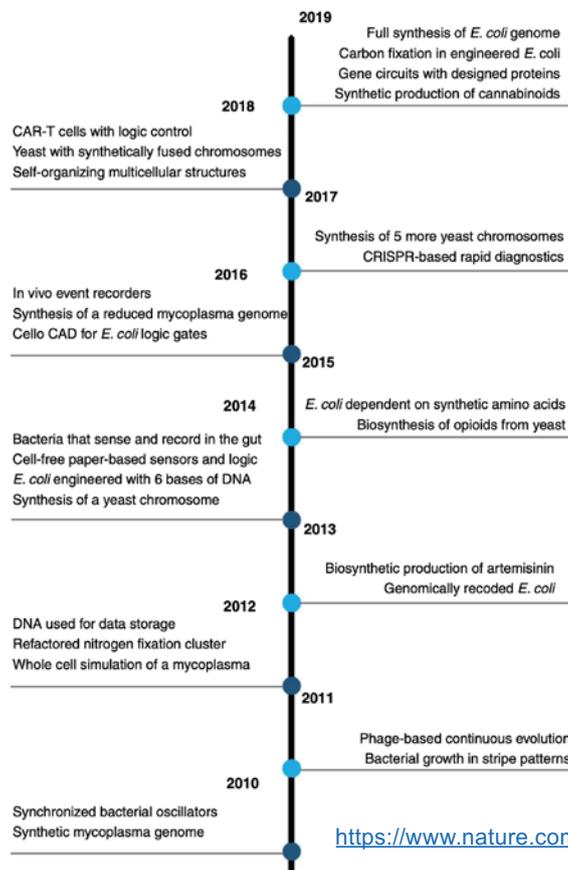
https://en.wikipedia.org/wiki/Lac_operon



<https://www.nature.com/articles/nrmicro3239>



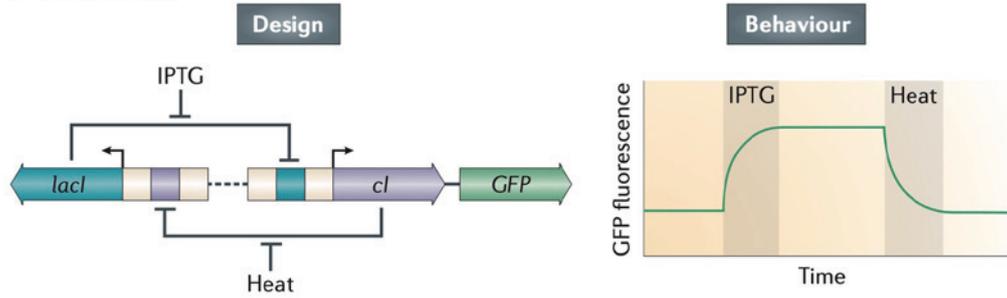
<https://www.nature.com/articles/nrmicro3239>



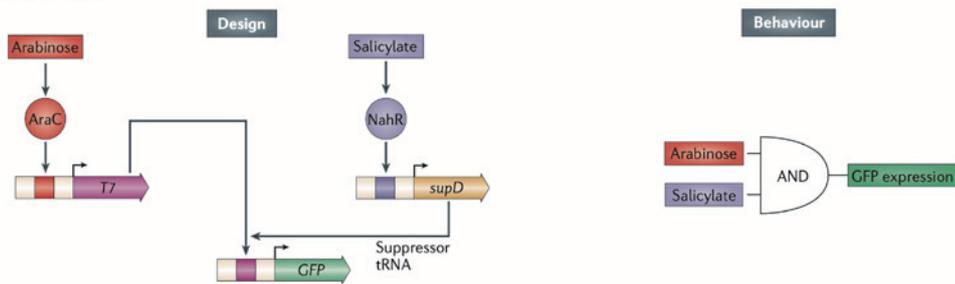
<https://www.nature.com/articles/s41467-020-19092-2>

Fig. 1 Landmark research achievements of synthetic biology from 2010 to 2020. A timeline is shown for the decade with brief summaries of some of the key research milestones published for each year.

a Toggle switch



b Two-input AND gate



<https://www.nature.com/articles/nrmicro3239>

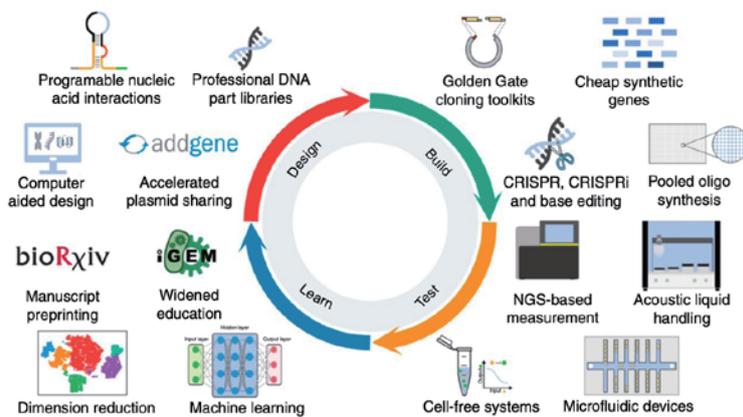


Fig. 2 New enabling technologies and ways of working that have accelerated the design-build-test-learn cycle of synthetic biology in the last decade. Diagram shows the engineering cycle used in synthetic biology (centre) with cartoons illustrating some of the key technologies and ways of working that now help accelerate each phase of the cycle.

<https://www.nature.com/articles/s41467-020-19092-2>

Lecture 2

Synthetic biology and AI

-대사공학 전략-

Kim, Hyun Uk

Department of Chemical and Biomolecular Engineering
Graduate School of Engineering Biology
Korea Advanced Institute of Science and Technology (KAIST)



강의계획서

- 학습목표
 - 합성생물학은 생물학, 공학, 컴퓨터 과학을 결합하여 **효소, 생합성 경로, 또는 전체 생물학적 개체를 설계하고 구축하는 매우 다학제적인 분야이다.**
 - 그 목표는 새로운 생물학적 시스템을 창조하거나 기존 시스템을 유용한 목적으로 재설계하는 것이며, 이 과정에서 **생물학적 도구를 설계할 때 모듈성, 표준화, 확장성이 중요하게 고려된다.**

강의계획서

- 학습목표
 - 본 강의에서는 **합성생물학의 배경**을 먼저 설명하고, 합성생물학이 적용 가능한 다양한 분야 중 **바이오제조와 대사공학**을 중심으로 소개하고자 한다.
 - 또한 이 과정에서 **AI**가 수행하는 핵심적인 역할을 설명한다.
 - 특히 미생물 대사공학의 목표가 대사 네트워크와 생산 공정 전반을 통합적으로 분석하고 최적화하여 **고수율·고생산성 균주**를 개발하는 데 있음을 설명한다. 이를 위해 **균주 설계부터 공정 최적화까지 전 과정**을 포괄하는 **10가지 전략**을 체계적으로 소개한다.
 - 더불어 이 과정에서 유용하게 활용될 수 있는 **대사 네트워크 모델(genome-scale metabolic model)**에 대해서도 이론과 실습을 제공한다.
 - 본 강의에서 소개되는 개념과 방법론은 **약물 표적 발굴 등 시스템 의학 분야**에도 적용 가능하다.

강의계획서

- 강의 1: 합성생물학 기반 바이오제조
- 강의 2: 대사공학 전략
- 강의 3: 대사 네트워크 모델 (Genome-scale metabolic model)
- 강의 4: 합성생물학과 AI 실습

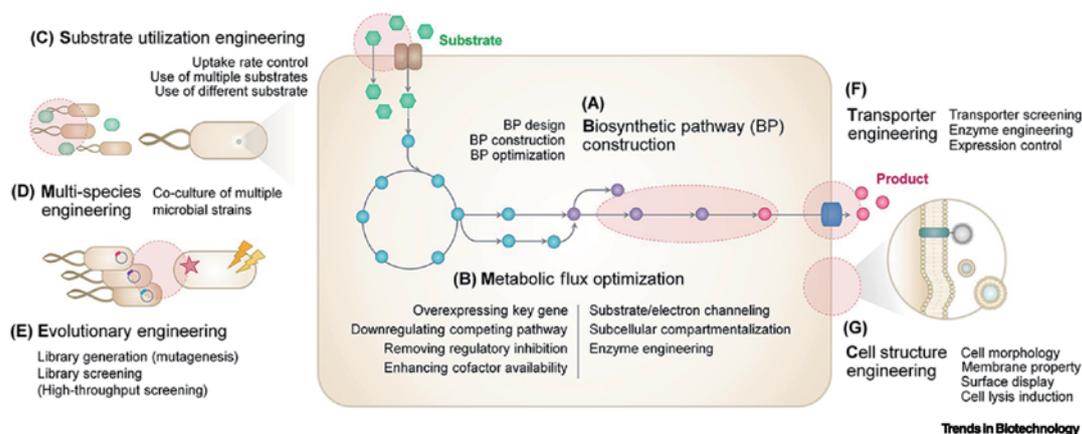


Figure 2. Representative metabolic engineering strategies that focus on microbial strain development. (A) Biosynthetic pathway construction includes biosynthetic pathway design, construction, and optimization processes. (B) Strategies for metabolic flux optimization includes overexpressing key genes, downregulating or removing competing pathways, removing regulatory inhibition, enhancing cofactor availability, substrate and electron channeling, subcellular compartmentalization, and enzyme engineering. (C) Strategies for substrate utilization engineering includes controlling substrate uptake rate and making cells to utilize multiple substrates or different substrates. (D) Multi-species engineering includes the co-culture of multiple microbial strains considering the multicellular interactions. (E) Evolutionary engineering involves adaptive laboratory evolution for cells and also library generation and screening processes for enzymes and pathways that can be accelerated by the use of high-throughput equipment. (F) Transporter engineering includes transporter screening, enzyme (transporter) engineering, and control of the expression levels of transporters. (G) Cell structure engineering includes modulation of cell morphology and membrane characteristics, and also cell surface display system and cell lysis induction system. The strategies are usually implemented in combination and repeatedly to optimize the cellular metabolic network for achieving the engineering goals such as efficient production of useful chemicals and degradation of toxic compounds.

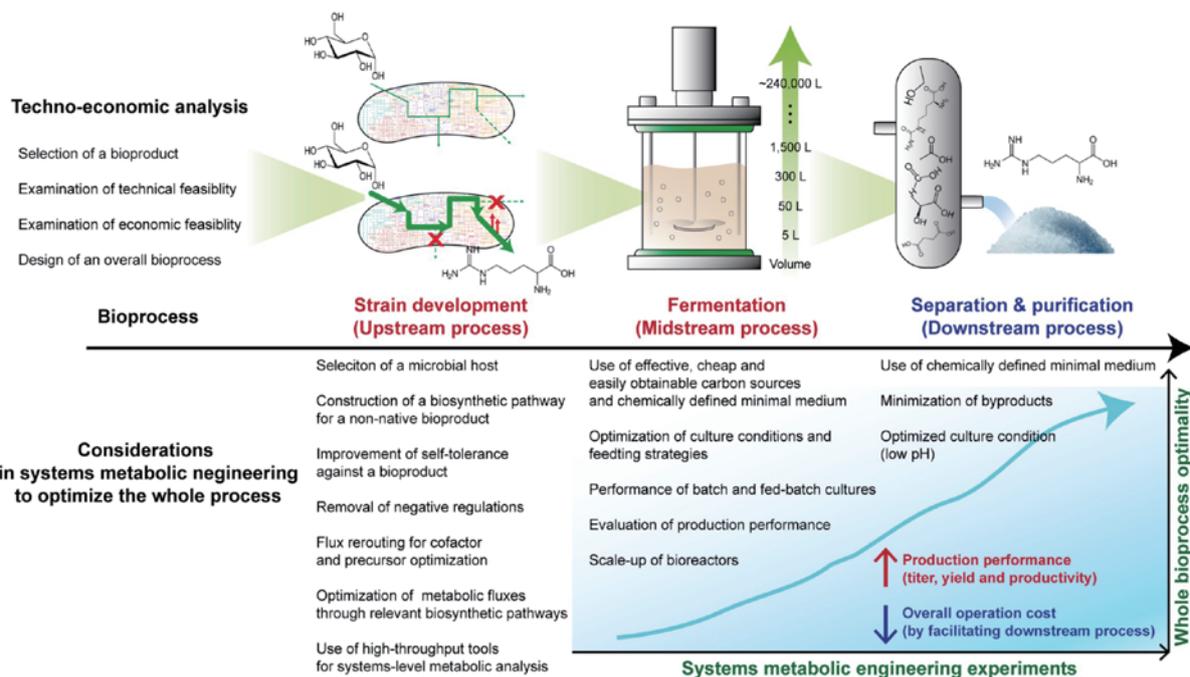
<https://www.sciencedirect.com/science/article/pii/S0167779922003420>

Systems metabolic engineering

- Strategy 1: Project design
- Strategy 2: Selection of host strain
- Strategy 3: Metabolic pathway reconstruction
- Strategy 4: Increasing tolerance to product
- Strategy 5: Removing negative regulatory circuits limiting overproduction
- Strategy 6: Rerouting fluxes to optimize cofactor and/or precursor availability
- Strategy 7: Diagnosing and optimizing metabolic fluxes toward product formation
- Strategy 8: Diagnosis and optimization of microbial culture conditions
- Strategy 9: System-wide gene manipulation to further increase production
- Strategy 10: Scale-up fermentation of the developed strain and diagnosis

- ✓ Interconnected
- ✓ Not always applied sequentially
- ✓ Often iterative

Lee & Kim. Nature Biotechnology (2015)
<https://www.nature.com/articles/nbt.3365>



Lee & Kim. Nature Biotechnology (2015)
<https://www.nature.com/articles/nbt.3365>

Strategy 1: Project design

- Selection of a target product based on various technical, economical, legal and regulatory factors
 - Often neglected in the academic sector
- Cost-effective carbon sources
 - Biomass readily available and production location?
- Three key performance indices; product concentration (titer), yield, and productivity
 - Bulk chemical vs fine chemical
 - Production host?
 - Fermentation mode: aerobic versus anaerobic fermentation, batch, fed-batch or continuous culture?
- Legal and regulatory considerations
 - Intellectual property
 - Genetically modified organisms (GMOs)
 - Generally recognized as safe (GRAS)

Strategy 1: Project design

Box 1 Metabolic engineering of a *C. glutamicum* strain overproducing L-arginine¹⁵

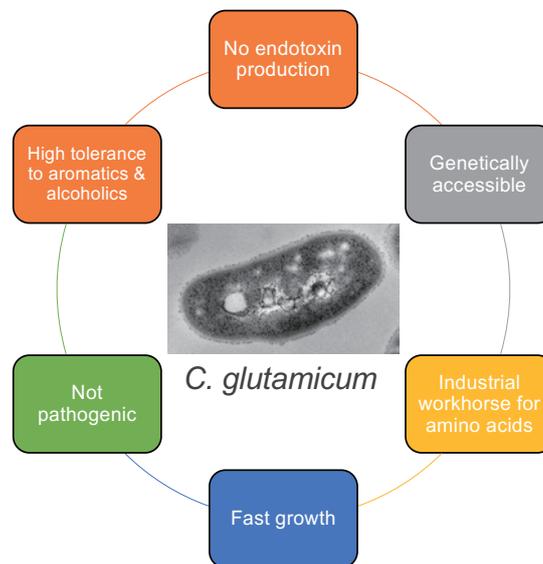
A challenge in the development of L-arginine-overproducing *C. glutamicum* was to achieve a production performance better than previous studies at laboratory scale and reproducibly at pilot-scale (**Fig. 2**). The final, high-performance AR6 strain was developed in about 9 person-years by the nine strategies described below (strategy 3 was skipped).

- Strategy 1. The end market for L-arginine was identified as for non-food, non-animal feed applications, such as for use as a plant growth stimulant⁶⁸, thus the use of a genetically modified *C. glutamicum* strain was considered acceptable. The preferred carbon substrates were derived from glucose and/or sucrose from industrially relevant feedstocks, corn starch hydrolysate and/or decomposed raw sugar from sugarcane. The initial target concentration, yield and productivity in the available 1,500-liter demo-plant scale fermenter were set at >80 g/L, 0.3~0.4 g/g and ~1.0 g/L/h, respectively.

Strategy 2: Selection of host strain

- *E. coli* and *S. cerevisiae* often considered as workhorses due to the availability of a wide range of genetic engineering tools
- Other microorganisms may be more suitable
 - e.g., *C. glutamicum* due to strong fluxes toward amino acids
 - Can a candidate organism utilize the desired carbon feedstock?
- Increasing availability of genetic engineering tools for a novel production host
 - e.g., the native succinic acid producer bacterium *M. succiniciproducens*
- A choice to make
 - Developing genetic engineering tools for a novel production host
 - Choosing a microorganism that can be more easily engineered
- Other considerations when selecting a host strain
 - e.g., carbon substrate utilization range, ease of fermentation in an inexpensive medium, ease of scale-up, requirement for anaerobic versus aerobic conditions, and the suitability of cost-effective downstream processes

***Corynebacterium glutamicum* as a production host**



Strategy 2: Selection of host strain

- Strategy 2. *C. glutamicum* ATCC 21831 (ARO) was selected for its ability to produce a substantial amount of L-arginine and to use the preferred carbon substrates, corn starch hydrolysate and/or decomposed raw sugar; the organism can also be metabolically engineered using reasonably well-established gene manipulation tools.

Strategy 3: Metabolic pathway reconstruction

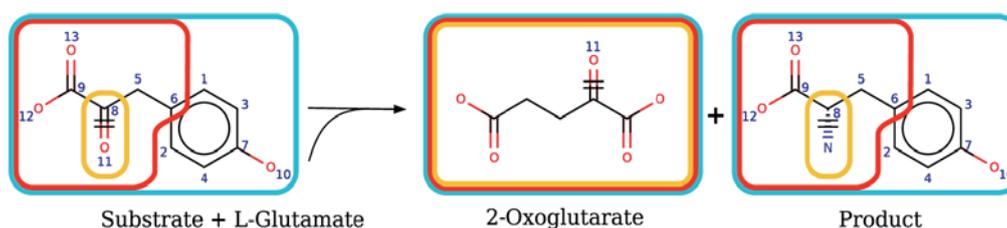
- Increasing importance due to the bio-based production of a diverse range of chemicals and fuels that are either non-natural or inefficiently produced in natural hosts
- Often based on intuitive biological knowledge
 - Unexpected inhibitory metabolic reactions caused by the promiscuous activities of introduced enzymes
- Use of genome mining and cheminformatic analyses
 - e.g., retrobiosynthesis
- Optimization of initially constructed, heterologous reactions via gene expression optimization and codon optimization
- Increasing availability of the combinatorial assembly of genes, operons and pathways, thanks to the relatively low cost of DNA synthesis and rapid advances in synthetic biology
- Often and still, various combinations of biosynthetic genes need to be tested

Retrobiosynthesis

A systematic design of a biosynthetic pathway

by using a set of biochemical reaction rules

that describe chemical transformation patterns between substrate and product molecules at an atomic level



Reaction rules:

D=∞ [#8:12]-[#6:9](=[#8:13])-[#6:8](=[#8:11])-[#6:5]-[#6:6]:1:[#6:1]:[#6:3]:[#6:7](-[#8:10]):[#6:4]:[#6:2]:1>>[#8]-[#6](=[#8])-[#6]-[#6]-[#6](=[#8:11])-[#6](-[#8])=[#8].[#7]-[#6@@H:8](-[#6:5]-[#6:6]:1:[#6:2]:[#6:4]:[#6:7](-[#8:10]):[#6:3]:[#6:1]:1)-[#6:9](-[#8:12])=[#8:13]

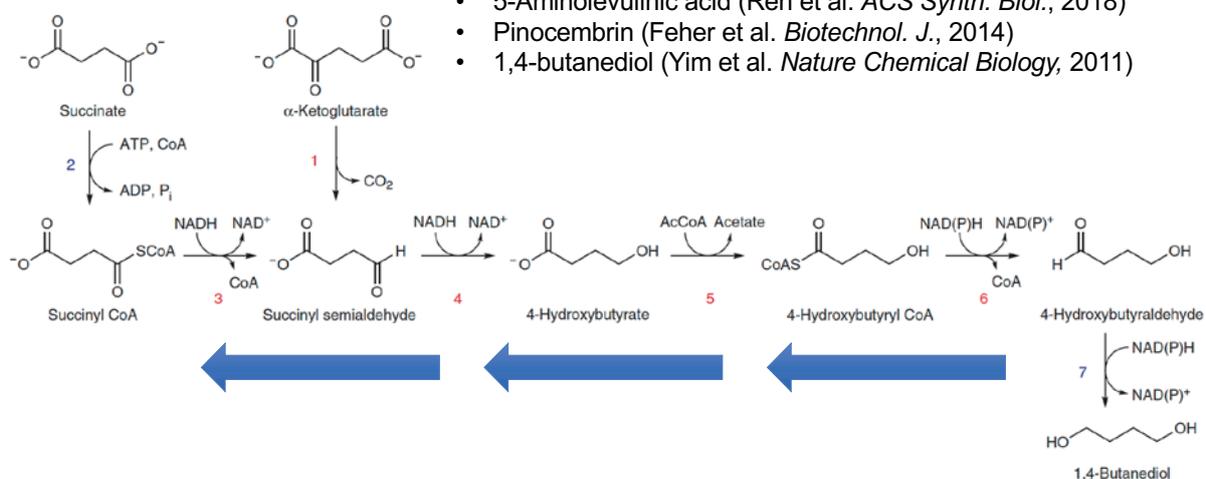
D=4 [#6:6]-[#6:5]-[#6:8](=[#8:11])-[#6:9](-[#8:12])=[#8:13]>>[#8]-[#6](=[#8])-[#6]-[#6]-[#6](=[#8:11])-[#6](-[#8])=[#8].[#6:6]-[#6:5]-[#6@@H:8](-[#7])-[#6:9](-[#8:12])=[#8:13]

D=0 [#6:8]=[#8:11]>>[#8]-[#6](=[#8])-[#6]-[#6]-[#6](=[#8:11])-[#6](-[#8])=[#8].[#6:8]-[#7]

Duigou et al. Nucleic Acids Research (2019)

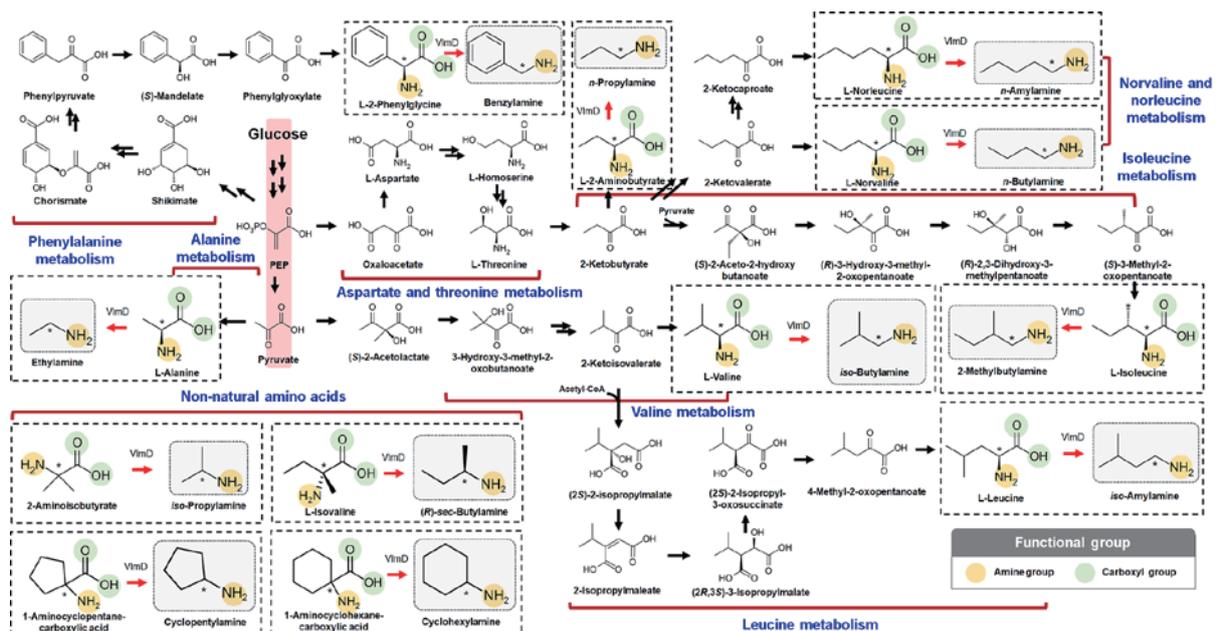
<https://academic.oup.com/nar/article/47/D1/D1229/5128930>

- Assimilation of formaldehyde (Yang et al. *Metab. Eng.*, 2019)
- 5-Aminolevulinic acid (Ren et al. *ACS Synth. Biol.*, 2018)
- Pinocembrin (Feher et al. *Biotechnol. J.*, 2014)
- 1,4-butanediol (Yim et al. *Nature Chemical Biology*, 2011)



Yim et al. *Nature Chemical Biology* (2011)
<https://www.nature.com/articles/nchembio.580>

Biosynthetic reactions constructed in *E. coli* for the in vivo production of 12 short-chain primary amines (SCPAs)



Kim et al. *Nature Communications* (2021)
<https://www.nature.com/articles/s41467-020-20423-6>

Strategy 4: Increasing tolerance to product

- Increasing product tolerance usually recommended upon construction of a base strain that generates a product at a requisite level
 - e.g., 50-80% of the level that starts to inhibit cell growth.
- Serial subculturing of cells with increasing concentrations of product or analogs with or without mutagen treatment, followed by identifying the cells that outgrow (i.e., grow faster)
 - Often repeated for the increased tolerance
- Engineering (often, overexpressing) efflux pumps to improve the product tolerance as a rational engineering approach

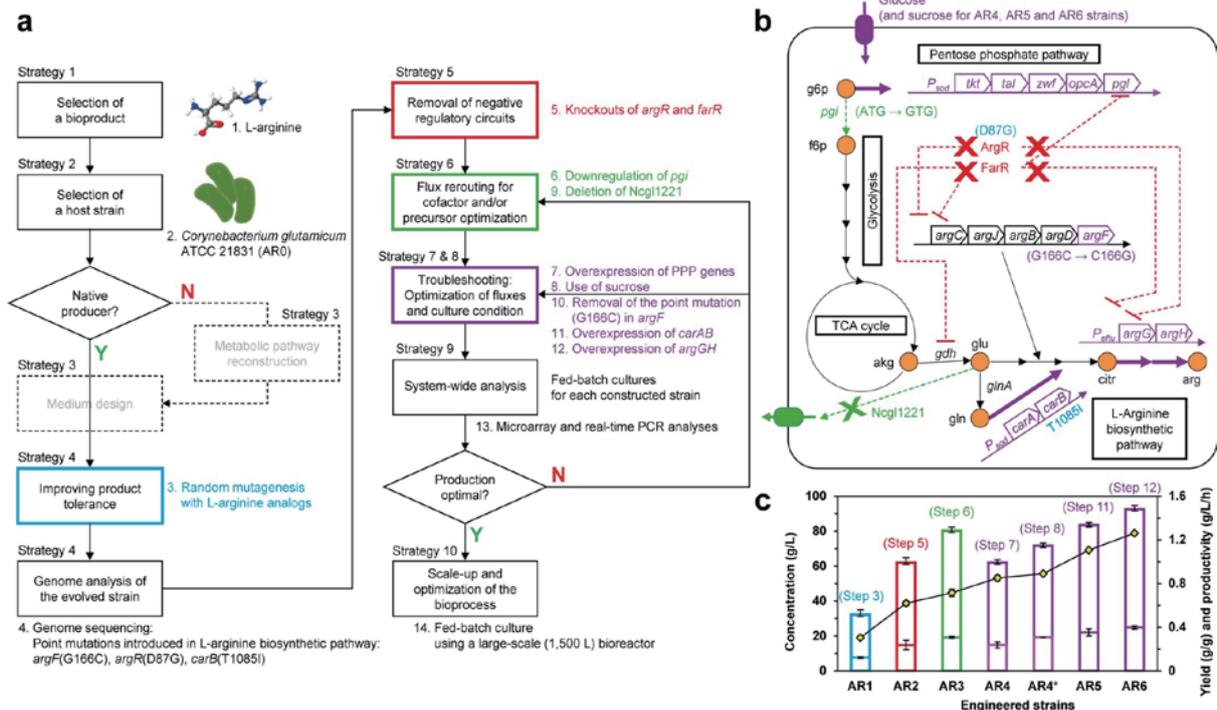
Strategy 4: Increasing tolerance to product

- Strategy 4. Random mutagenesis was performed to select an L-arginine-tolerant strain using L-arginine analogs, isolating the AR1 strain capable of producing 34.2 g/L of L-arginine in fed-batch culture (**Fig. 2c**), a more than twofold higher concentration than that obtained with the ARO strain. Subsequent genome sequencing and analysis of the AR1 strain revealed mutations in genes associated with the L-arginine biosynthetic pathway.

Strategy 5: Removing negative regulatory circuits limiting overproduction

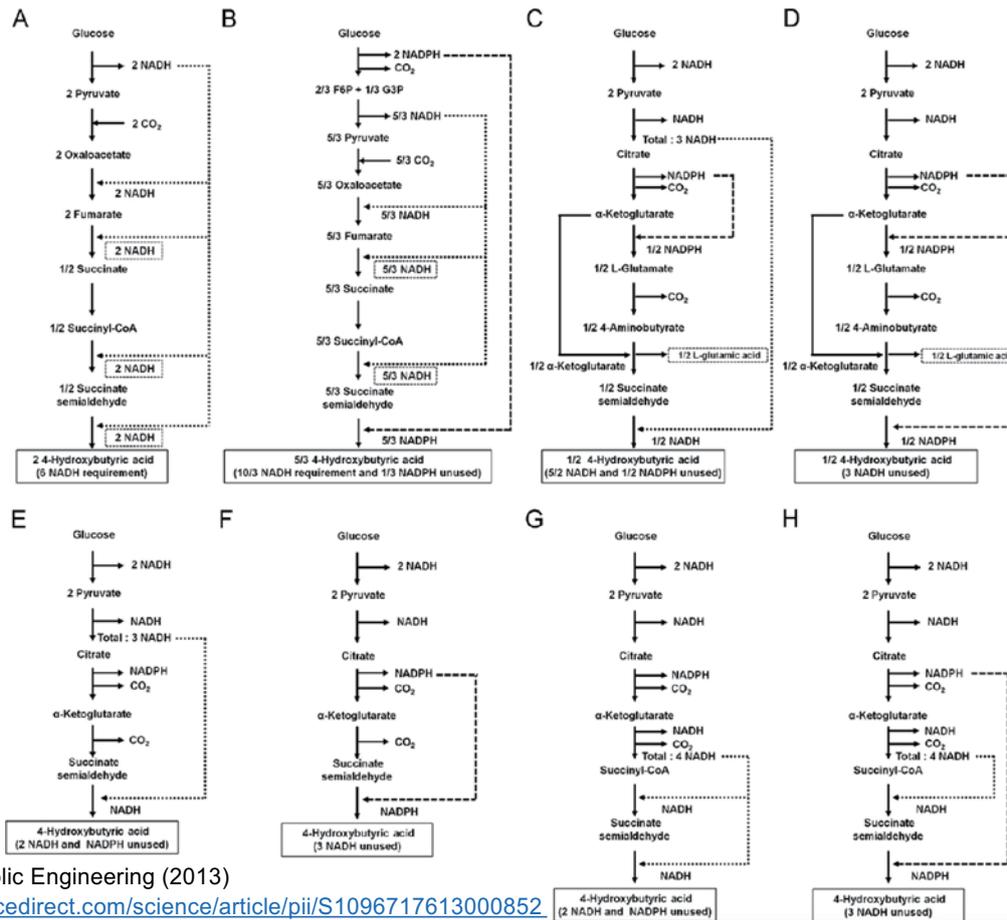
Strategy 6: Rerouting fluxes to optimize cofactor and/or precursor availability

- Transcriptional attenuation control and feedback inhibition, especially for amino acids biosynthesis, as typical examples of negative regulatory circuits
- Gene knockout as an obvious choice to reroute fluxes to increase the level of cofactors or precursors
 - e.g., removing competing pathways
 - Gene knockdown using synthetic small regulatory RNAs if target genes are essential
 - Construction of industrial strains via chromosomal integration rather than plasmid-based system
- In general, cofactor and precursor optimization usually requires more system-wide approaches than removal of pathway-specific negative regulatory circuits, because in the former case, global mass, energy and redox balances need to be considered



Lee & Kim. Nature Biotechnology (2015)
<https://www.nature.com/articles/nbt.3365>

Park et al. Nature Communications (2014)
<https://www.nature.com/articles/ncomms5618>



Strategy 5: Removing negative regulatory circuits limiting overproduction

Strategy 6: Rerouting fluxes to optimize cofactor and/or precursor availability

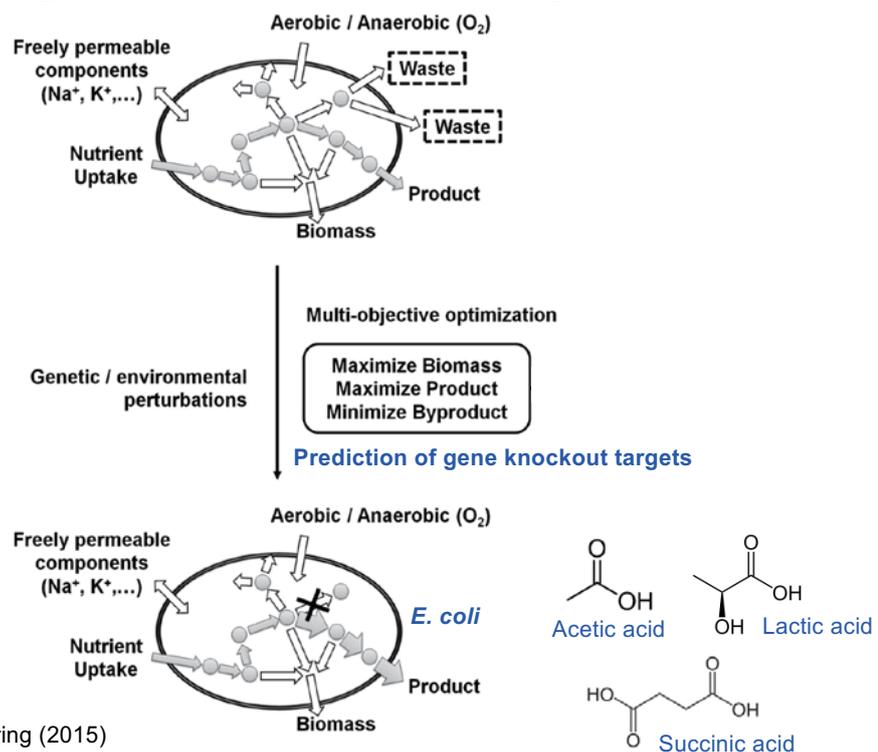
- Strategy 5. Negative feedback regulation in the AR1 strain was removed by inactivating two regulatory genes, *argR* and *farR*. The resulting AR2 strain was able to produce 61.9 g/L of L-arginine by fed-batch culture (Fig. 2c).
- Strategy 6. Downregulation of the *pgi* gene was achieved by replacing its start codon ATG with GTG to reroute fluxes toward the pentose phosphate pathway (PPP) for more NADPH, while still permitting flux through glycolysis for good enough cell growth. As a result, the AR3 strain capable of producing 80.2 g/L of L-arginine by fed-batch culture was developed.

Strategy 7: Diagnosing and optimizing metabolic fluxes toward product formation

Strategy 8: Diagnosis and optimization of microbial culture conditions

- Systematic diagnosis and optimization of intermediate strain under conditions that are as similar as possible to the final industrial fermentation conditions
 - Fed-batch fermentations necessary, at least under laboratory-scale conditions, during strain development
 - Diagnosis based on performance indices, such as productivity, yield and titer
 - Additional identification of potential metabolic bottlenecks and byproduct formation in fed-batch conditions
- Consideration of alternative substrates at this stage in light of the availability of feedstock and its economics
 - For industrial-scale fermentations, medium cost often considered more important than the nature of culture media being either defined, semi-defined or complex
 - Important to develop an industrial strain capable of efficiently producing a desired product using impure low-price substrates

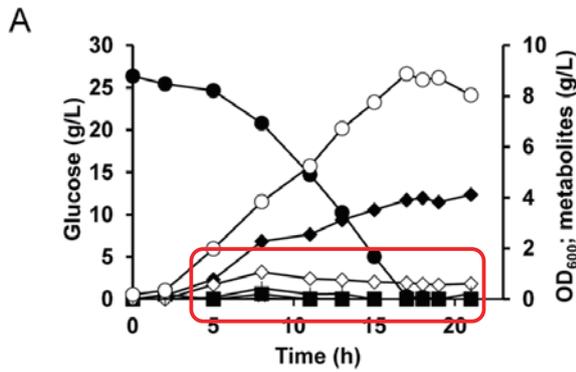
Use of multi-objective optimization for the homo-organic acid-producing strains



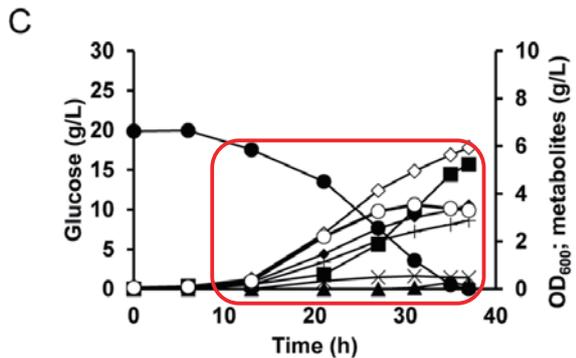
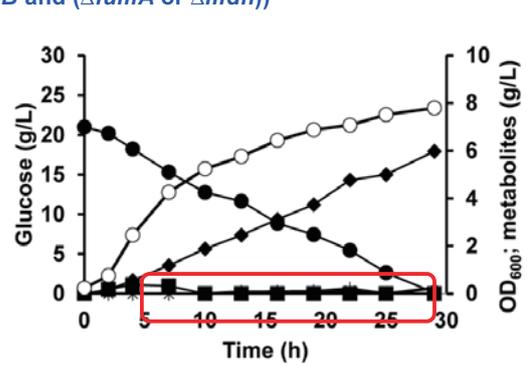
Kim et al. Metabolic Engineering (2015)

<https://www.sciencedirect.com/science/article/pii/S1096717614001621?via%3Dihub>

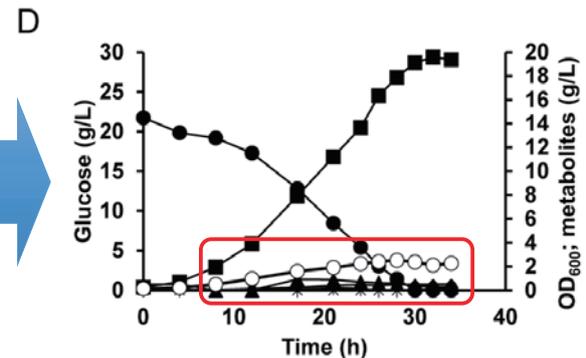
E. coli W3110 at 20% DO (microaerobic)



Homo-acetic acid producing strain ($\Delta pflB$ and $\Delta fumA$ or Δmdh)



E. coli W3110 under anaerobic



Homo-lactic acid producing strain ($\Delta pflB$ and $\Delta aceF$)

OD₆₀₀ (○), glucose (●), acetic acid (◆), ethanol (+), formic acid (◇), lactic acid (■), pyruvic acid (▲), succinic acid (x)

Kim et al. Metabolic Engineering (2015)

Strategy 7: Diagnosing and optimizing metabolic fluxes toward product formation

Strategy 8: Diagnosis and optimization of microbial culture conditions

- Strategy 7. After fed-batch cultures of AR3 revealed decreased consumption of glucose substrate and consequently longer fermentation times than AR2, it was found that the flux through the PPP was not streamlined after rerouting the flux from the glycolytic pathway. Thus, an operon associated with PPP flux was amplified using a stronger *sod* promoter. The resulting AR4 strain gave the highest L-arginine productivity yet of 0.85 g/L/h (Fig. 2c); however, L-arginine titer and yield were lower than those obtained with the AR3 strain, suggesting the need for further optimization.
- Strategy 8. Fed-batch culture of the AR4 strain using sucrose and glucose as carbon sources resulted in the production of 71.7 g/L of L-arginine with productivity and yield of 0.89 g/L/h and 0.31 g/g, respectively; these values are all higher than those obtained by the fed-batch culture using glucose as a single carbon source.

Strategy 9: System-wide gene manipulation to further increase production

- A final round of engineering to identify potential gene manipulation targets for further enhanced production capability
- Systems and synthetic biology approaches
 - Cultivation profile-based system-wide analyses, high-throughput genome-scale engineering, omics-based approaches and/or *in silico* metabolic simulations (e.g., genome-scale metabolic modeling and simulation)
 - Use of an increasing array of high-throughput genome-scale engineering tools, such as (co-selection) multiplex automated genome engineering, trackable multiplex recombineering and synthetic small regulatory RNAs

Strategy 9: System-wide gene manipulation to further increase production

- Strategy 9. The AR4 strain was subjected to final rounds of engineering by analyzing its fed-batch fermentation profiles in connection with metabolic characteristics. Because fed-batch culture of the AR4 strain consistently showed the excretion of L-glutamate, which can be further converted to L-arginine for an improved L-arginine yield and titer, the L-glutamate exporter was deleted. Analysis of genes associated with L-arginine biosynthesis using the sequenced genome data of the ARO and AR1 strains also identified two additional gene manipulation targets: removal of a negative mutation in the *argF* gene and overexpression of the *carAB* genes. The AR5 strain constructed this way produced 82 g/L of L-arginine, with a yield of 0.35 g/g by fed-batch culture. Finally, to solve the problem of the excretion of about 5 g/L of citrulline, two rate-controlling enzymes encoded by the *argGH* genes were overexpressed using a strong promoter, resulting in the AR6 strain. Fed-batch culture of the AR6 strain allowed production of 92.5 g/L of L-arginine with a high yield of 0.40 g/g and productivity of 1.26 g/L/h. Intended changes in the AR6 strain were confirmed with transcriptome analysis, enzyme assay and real-time PCR.

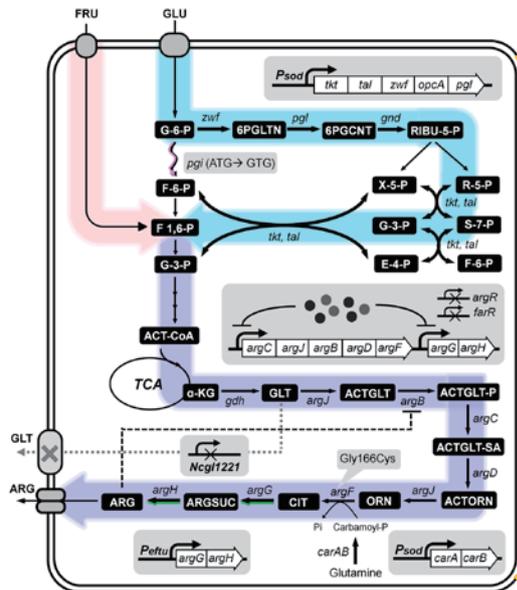
Strategy 10: Scale-up fermentation of the developed strain and diagnosis

- Validation of the developed microbial strain in a pilot plant-scale or demo plant-scale fermenter
 - Possible presence of discrepancies from one strain (or scale) to another
 - Mainly due to different mixing and aeration in laboratory-scale and pilot-scale fermenters (very difficult to predict)
- Possible reversion of a high-performance strain back to a low-performance strain through the loss of production capacity and phenotype alteration
 - Caused by genetic instability and/or fermentation conditions having different mass transfer rates of nutrients and oxygen, and substrate and product concentration profiles in the industrial-sized fermenter
 - Possible solutions (but, not guaranteed): permanent chromosomal manipulation (instead of plasmid-mediated engineering) and gene-level engineering (e.g., knockout of an entire gene) rather than changes of a few base pairs (e.g., single-point mutations)
 - In industrial fermentation, companies regularly isolate pure colonies of high performance at the end of the fermentation
- Contamination control in industrial-scale fermentation

Strategy 10: Scale-up fermentation of the developed strain and diagnosis

- Strategy 10. When the AR6 strain was cultured in a fed-batch mode in a 1,500-liter bioreactor, 81.2 g/L of L-arginine was produced with a yield and productivity of 0.353 g/g and 0.91 g/L/h, respectively. Even though these values are somewhat lower than those obtained by the lab-scale fed-batch culture, they met the project objectives defined in strategy 1 (high enough for initial industrial-scale fermentation). The AR6 strain's production performance was found to be highly reproducible even under scale-up fermentation.

- 92.5 g/L
- 0.40 g L-arginine / g total C source
- 81.2 g/L
- 0.35 g L-arginine / g total C source



Lab-scale 5 L bioreactor



Pilot-scale 1,500 L bioreactor

KSBi-BIML 2026

Lecture 3

Synthetic biology and AI

-대사 네트워크 모델 (Genome-scale metabolic model)-

Kim, Hyun Uk

Department of Chemical and Biomolecular Engineering
 Graduate School of Engineering Biology
 Korea Advanced Institute of Science and Technology (KAIST)



강의계획서

- 학습목표
 - 합성생물학은 생물학, 공학, 컴퓨터 과학을 결합하여 효소, 생합성 경로, 또는 전체 생물학적 개체를 설계하고 구축하는 매우 다학제적인 분야이다.
 - 그 목표는 새로운 생물학적 시스템을 창조하거나 기존 시스템을 유용한 목적으로 재설계하는 것이며, 이 과정에서 생물학적 도구를 설계할 때 모듈성, 표준화, 확장성이 중요하게 고려된다.

강의계획서

- 학습목표
 - 본 강의에서는 합성생물학의 배경을 먼저 설명하고, 합성생물학이 적용 가능한 다양한 분야 중 바이오제조와 대사공학을 중심으로 소개하고자 한다.
 - 또한 이 과정에서 AI가 수행하는 핵심적인 역할을 설명한다.
 - 특히 미생물 대사공학의 목표가 대사 네트워크와 생산 공정 전반을 통합적으로 분석하고 최적화하여 고수율·고생산성 균주를 개발하는 데 있음을 설명한다. 이를 위해 균주 설계부터 공정 최적화까지 전 과정을 포괄하는 10가지 전략을 체계적으로 소개한다.
 - 더불어 이 과정에서 유용하게 활용될 수 있는 대사 네트워크 모델(genome-scale metabolic model)에 대해서도 이론과 실습을 제공한다.
 - 본 강의에서 소개되는 개념과 방법론은 약물 표적 발굴 등 시스템 의학 분야에도 적용 가능하다.

강의계획서

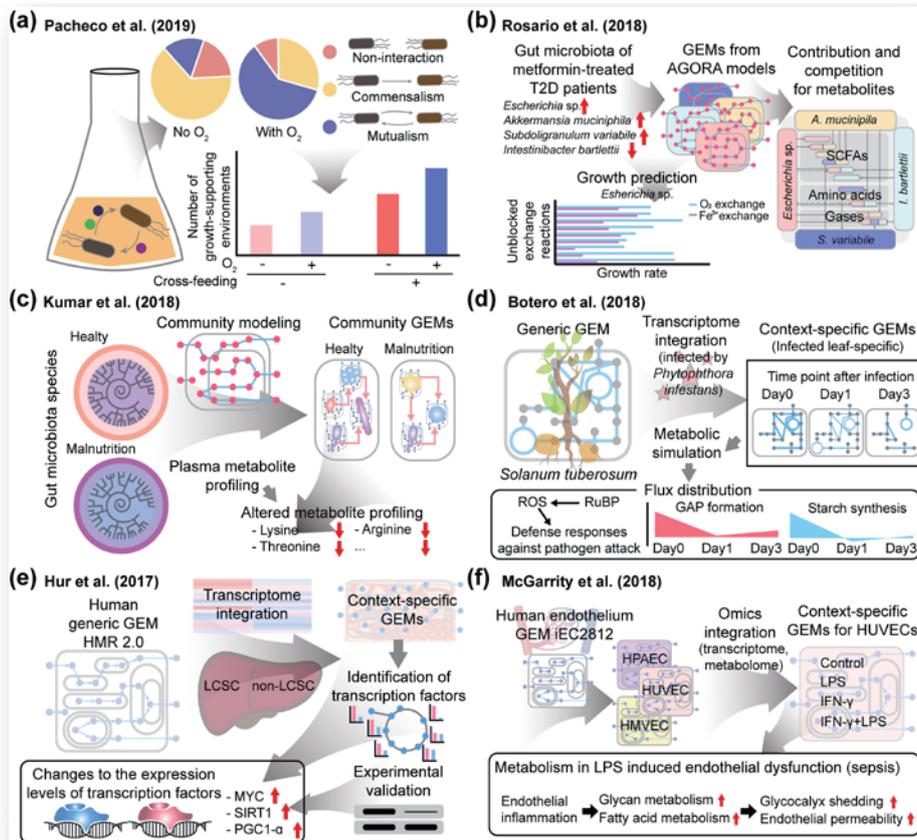
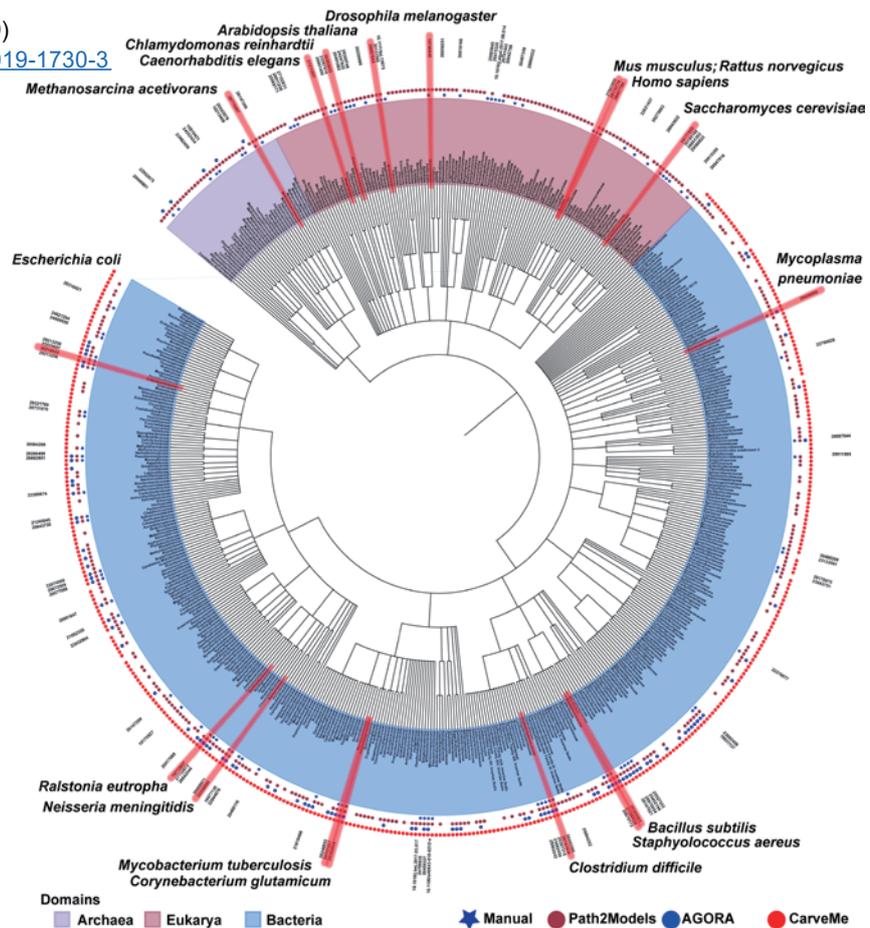
- 강의 1: 합성생물학 기반 바이오제조
- 강의 2: 대사공학 전략
- 강의 3: 대사 네트워크 모델 (Genome-scale metabolic model)
- 강의 4: 합성생물학과 AI 실습

Keywords in this class

- Simulation vs computational model
- Genome-scale metabolic models (GEMs)
 - Industrial and medical applications
 - Reconstruction procedure
 - Model repositories
 - Systems Biology Markup Language (SBML)
 - Boolean logic of gene-protein-reaction (GPR) associations
- Flux balance analysis
 - Prediction of metabolic flux distributions under a specific genotypic and environmental condition
 - Linear programming
 - Prediction of gene manipulation targets
 - Multi-objective optimization

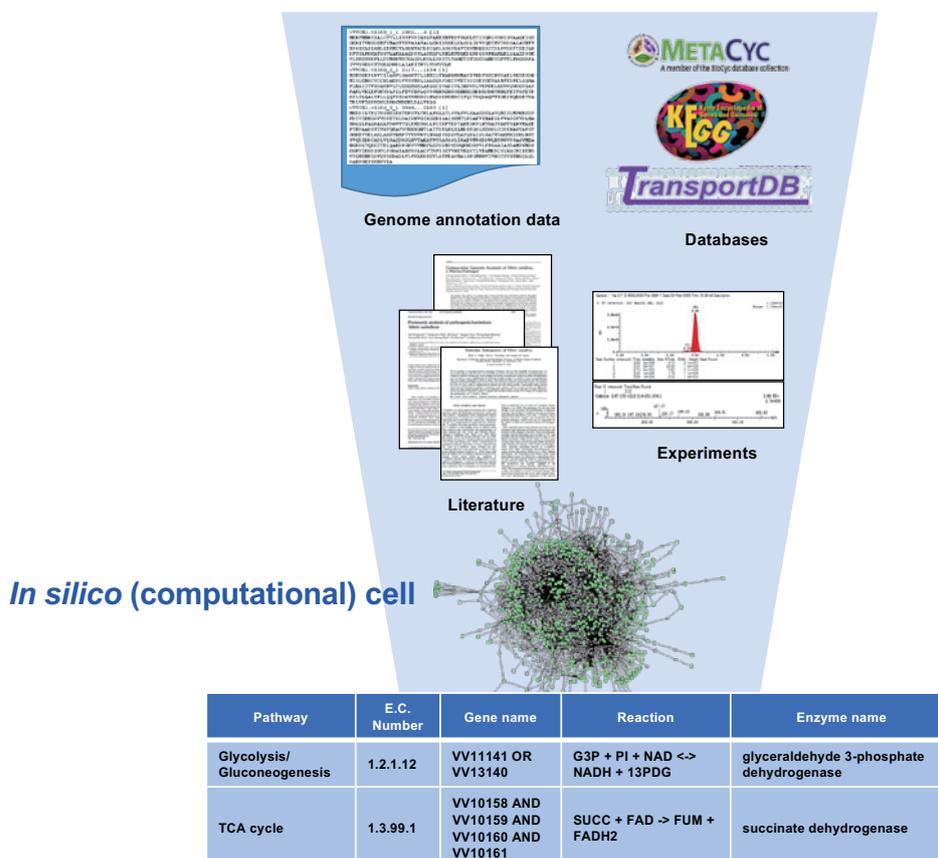
As of February 2019:

- GEMs have been reconstructed for **6239 organisms**
- **5897 bacteria**, 127 archaea & 215 eukaryotes
- Either manually or by using **automatic** GEM reconstruction tools

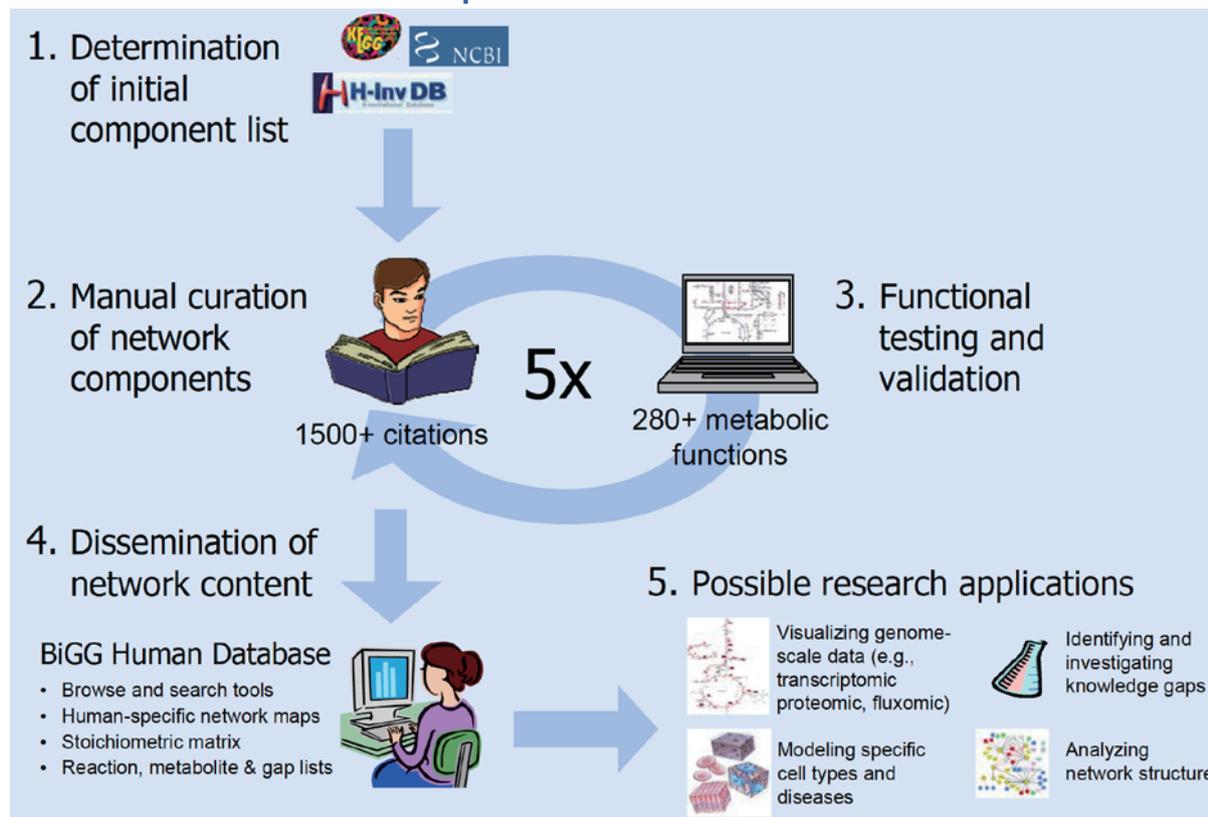


Reconstruction of a genome-scale metabolic model

- Manual reconstruction of a genome-scale metabolic model (GEM) was a very time-consuming process, as seen in Thiele and Palsson (Nature Protocols, 2010). Much effort was exerted on extensive literature survey and experimental validation of the model.
- Many reconstruction steps have now been automated (Gu et al. Genome Biology, 2019). This has now allowed the reconstruction of a large number of related organisms at the same time. Examples include all the *E. coli* strains, or multiple tumor cells.
- Reconstructed GEMs can be found at:
 - BiGG Models (<http://bigg.ucsd.edu>)
 - Metabolic Atlas (<https://metabolicatlas.org/gems/repository>)
 - Supplementary Data of individual papers and code repositories
- You don't need to build a new GEM unless you are working on a model organism. GEMs for most of the model organisms are now available online.



Reconstruction procedure of a human GEM 'Recon 1'



Duarte et al. PNAS (2007)

<https://doi.org/10.1073/pnas.0610772104>

PROTOCOL

A protocol for generating a high-quality genome-scale metabolic reconstruction

Ines Thiele^{1,2} & Bernhard Ø Palsson¹

¹Department of Bioengineering, University of California, San Diego, La Jolla, California, USA. ²Current address: Center for Systems Biology, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavik, Iceland. Correspondence should be addressed to B.Ø.P. (palsson@ucsd.edu).

Published online 7 January 2010; doi:10.1038/nprot.2009.203

Network reconstructions are a common denominator in systems biology. Bottom-up metabolic network reconstructions have been developed over the last 10 years. These reconstructions represent structured knowledge bases that abstract pertinent information on the biochemical transformations taking place within specific target organisms. The conversion of a reconstruction into a mathematical format facilitates a myriad of computational biological studies, including evaluation of network content, hypothesis testing and generation, analysis of phenotypic characteristics and metabolic engineering. To date, genome-scale metabolic reconstructions for more than 30 organisms have been published and this number is expected to increase rapidly. However, these reconstructions differ in quality and coverage that may minimize their predictive potential and use as knowledge bases. Here we present a comprehensive protocol describing each step necessary to build a high-quality genome-scale metabolic reconstruction, as well as the common trials and tribulations. Therefore, this protocol provides a helpful manual for all stages of the reconstruction process.

Thiele and Palsson. *Nature Protocols* (2010)

<https://doi.org/10.1038/nprot.2009.203>

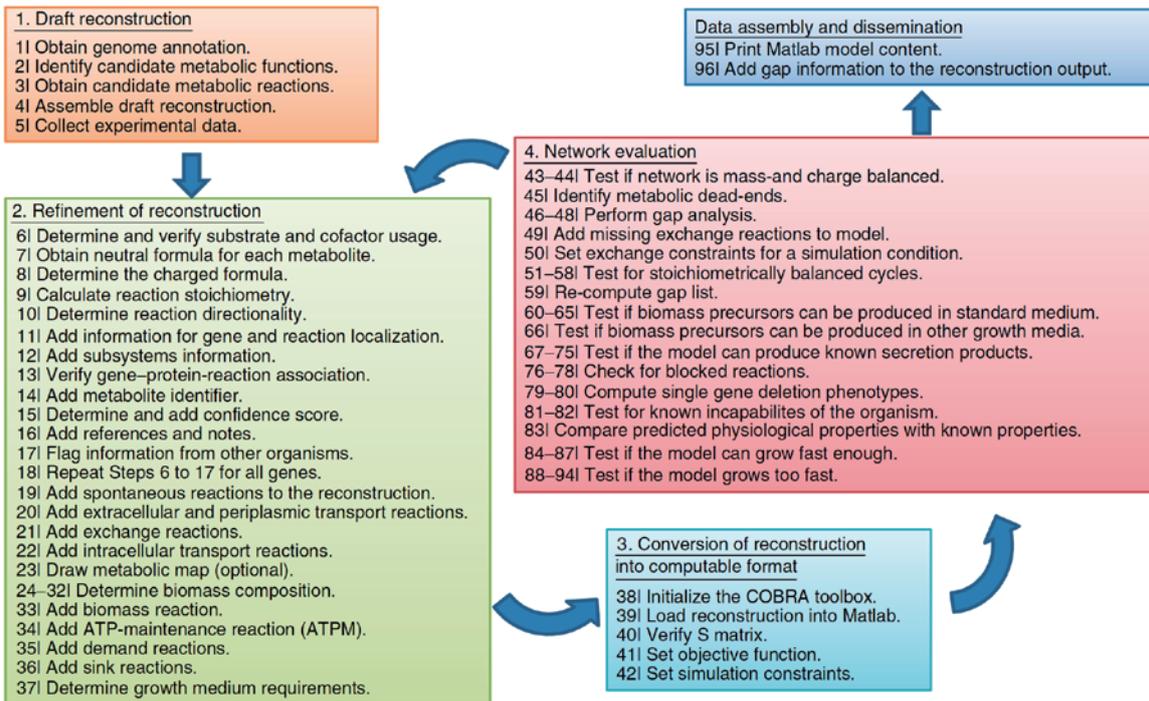
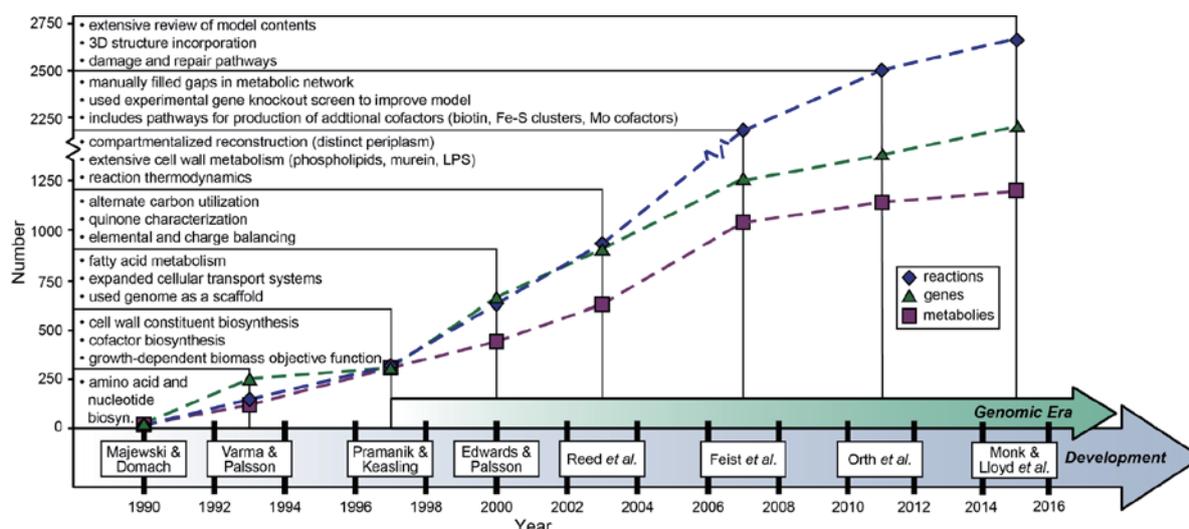


Figure 1 | Overview of the procedure to iteratively reconstruct metabolic networks. In particular, Stages 2–4 are continuously iterated until model predictions are similar to the phenotypic characteristics of the target organism and/or all experimental data for comparison are exhausted.

Thiele and Palsson. *Nature Protocols* (2010)
<https://doi.org/10.1038/nprot.2009.203>

Growth of reconstruction coverage and content in the series of published reconstruction of *E. coli* metabolism



Monk *et al.* *Nature Biotechnology* (2017)
<https://doi.org/10.1038/nbt.3956>

Model validation

(a) The Colony-live platform was used to measure **growth capabilities of 3,869 single-knockout mutant *E. coli* strains on minimal media with 16 different carbon sources**, forming a total of 62,272 measured phenotypes. Colony-live provides specific values for lag-time (LTG), maximum growth rate (MGR), and growth saturation point (GSP) for each gene knockout and condition.

(b) Subset of knockout data highlighting **growth rates for gene knockouts in the tricarboxylic acid (citric acid) cycle**.

(c) The iML1515 reconstruction is **93.4% accurate** in predicting the effect of gene knockouts, an increase in accuracy of 3.7% over the 89.8% accuracy of the iJO1366 *E. coli* metabolic reconstruction.

Monk et al. *Nature Biotechnology* (2017)
<https://doi.org/10.1038/nbt.3956>

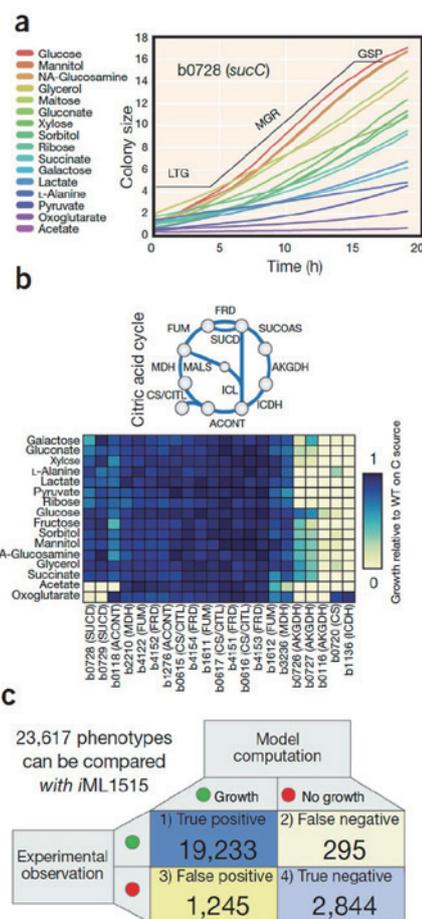
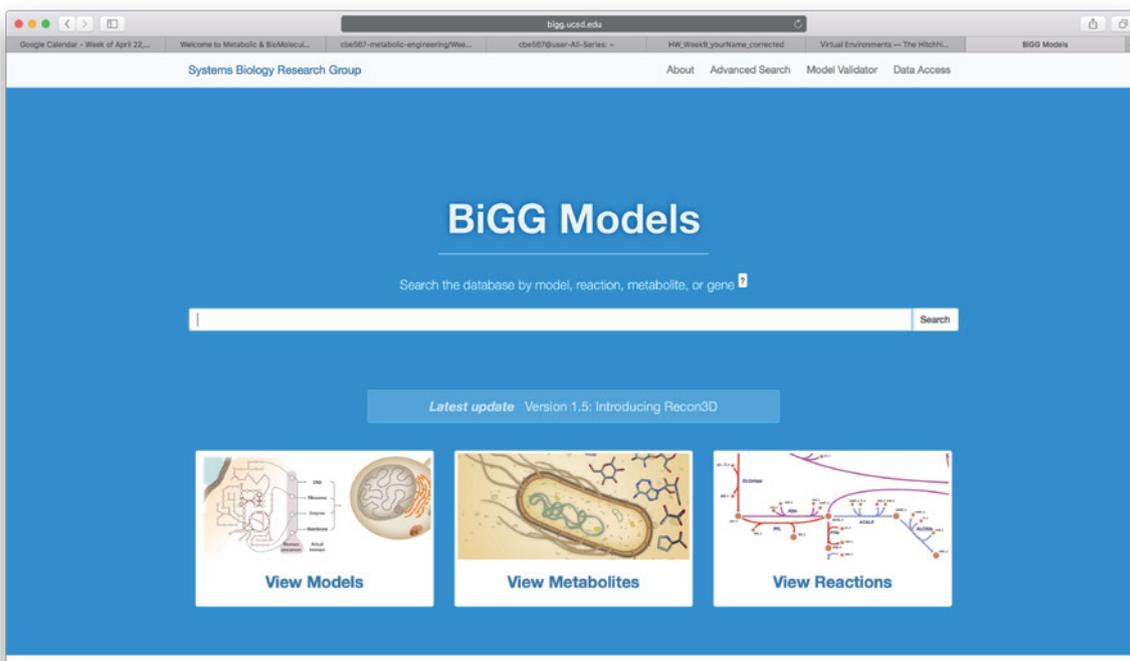


Table 1 Software programs for the reconstruction of GEMs

Tool	Language	Graphical user interface (GUI) available?	Source database for metabolic reactions	Use as a reference model?	Gap-filling	Eukaryote modeling	Simulation ready	Reference
AuReMe	Python	No	KEGG, BiGG Models, MetaCyc	No	Yes	Yes	Yes	[87]
AutoKEGGRec	Matlab	No	KEGG	No	No	No	No	[88]
CarveMe	Python	No	BiGG Models	Yes (universal model)	Yes	No	Yes	[15]
CoReCo	Python	No	KEGG	No	Yes	Yes	Yes	[89]
FAME	Python	Yes	KEGG	Yes	No	No	Yes	[90]
merlin	Java	Yes	KEGG, MetaCyc, UniProtKB, TCDB	No	No	Yes	No	[91]
MetaDraft	Python	Yes	BiGG Models	Yes	No	No	Yes	[92]
Model SEED	Web	Yes	In-house reaction database	Yes	Yes	Yes (only plants)	Yes	[93]
Pathway Tools	Python, Lisp	Yes	Pathway/Genome Database (PGDB), MetaCyc	No	Yes	Yes	Yes	[94]
RAVEN 2.0	Matlab	No	KEGG, MetaCyc	Yes	Yes	Yes	Yes	[95]
SubliMinal Toolbox	Java	No	KEGG, MetaCyc	No	No	No	Yes	[96]

Gu et al. *Genome Biology* (2019)
<https://doi.org/10.1186/s13059-019-1730-3>

Standardized genome-scale metabolic models (GEMs) are available at BiGG Models



<http://bigg.ucsd.edu>

BiGG Models

Home Advanced Search Data Access Memote Validator

Model: iML1515

Organism:
Escherichia coli str. K-12 substr. MG1655

Genome:
NC_000913.3 <http://bigg.ucsd.edu/models/iML1515>

Model metrics:

Component	Count
Metabolites	1877
Reactions	2712
Genes	1516

Download COBRA model from the BiGG Database:

SBML [?]: [iML1515.xml \(.xml.gz, compressed\)](#)
 JSON [?]: [iML1515.json \(.json.gz, compressed\)](#)
 MAT [?]: [iML1515.mat \(.mat.gz, compressed\)](#)

Downloads last updated Oct 31, 2019 | [BiGG License](#)

Systems Biology Markup Language (SBML) level 3 flux balance constraints (SBML3FBC) package as the primary description and exchange format

```

~/Downloads/IML1515.xml (no function selected)
1 <?xml version='1.0' encoding='UTF-8'?>
2 <sbml xmlns:fbc="http://www.sbml.org/sbml/level3/version1/fbc/version2" xmlns="http://www.sbml.org/sbml/level3/version1/core" level="3" version="1"
3 <model fbc:strict="true" id="IML1515">
4 <listOfUnitDefinitions>
5 <unitDefinition id="mmol_per_gDW_per_hr">
6 <listOfUnits>
7 <unit kind="mole" scale="-3" multiplier="1" exponent="1"/>
8 <unit kind="gram" scale="0" multiplier="1" exponent="-1"/>
9 <unit kind="second" scale="0" multiplier="3600" exponent="-1"/>
10 </listOfUnits>
11 </unitDefinition>
12 </listOfUnitDefinitions>
13 <fbc:listOfObjectives fbc:activeObjective="obj">
14 <fbc:objective fbc:id="obj" fbc:type="maximize">
15 <fbc:listOfFluxObjectives>
16 <fbc:fluxObjective fbc:reaction="R_BIOMASS_Ec_IML1515_core_75p37M" fbc:coefficient="1"/>
17 </fbc:listOfFluxObjectives>
18 </fbc:objective>
19 </fbc:listOfObjectives>
20 <listOfParameters>
21 <parameter value="-1000" id="cobra_default_lb" sboTerm="SBO:0000626" constant="true" units="mmol_per_gDW_per_hr"/>
22 <parameter value="1000" id="cobra_default_ub" sboTerm="SBO:0000626" constant="true" units="mmol_per_gDW_per_hr"/>
23 <parameter value="0" id="cobra_0_bound" sboTerm="SBO:0000626" constant="true" units="mmol_per_gDW_per_hr"/>
24 <parameter id="R_EX_glc_D_e_lower_bound" value="-10" sboTerm="SBO:0000625" constant="true" units="mmol_per_gDW_per_hr"/>
25 <parameter id="R_ATPM_lower_bound" value="6.86" sboTerm="SBO:0000625" constant="true" units="mmol_per_gDW_per_hr"/>
26 </listOfParameters>
27 <listOfCompartments>
28 <compartment id="c" name="cytosol" constant="true"/>
29 <compartment id="e" name="extracellular space" constant="true"/>
30 <compartment id="p" name="periplasm" constant="true"/>
31 </listOfCompartments>
32 <listOfSpecies>
33 <species id="M_octapb_c" constant="false" boundaryCondition="false" hasOnlySubstanceUnits="false" name="Octanoate (protein bound)" metaid="M
34 <sbml:annotation xmlns:sbml="http://www.sbml.org/sbml/level3/version1/core">
35 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
36 <rdf:Description rdf:about="#M_octapb_c">
37 <bqbiol:is xmlns:bqbiol="http://biomodels.net/biology-qualifiers"/>

```

- Reaction IDs
- Metabolite IDs
- Gene-protein-reaction (GPR) associations
- Enzyme name
- Reaction compartment
- Flux constraints
- Objective function

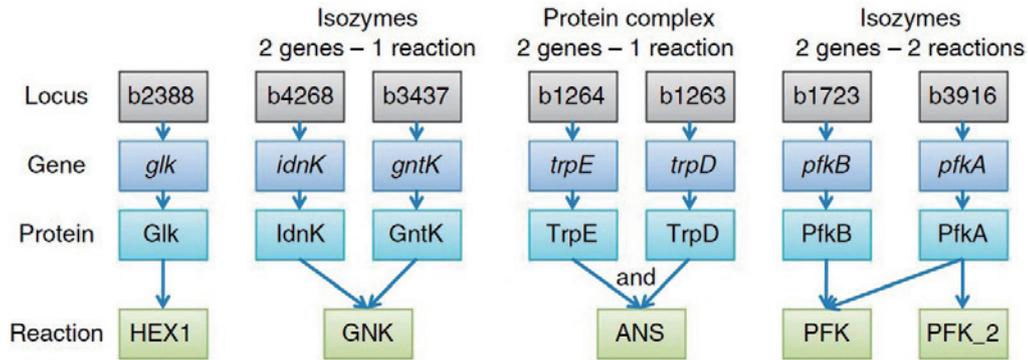
IML1515:
Genome-scale metabolic model of
E. coli

Monk et al. *Nature Biotechnology* (2017)
<https://doi.org/10.1038/nbt.3956>

```

168431 <reaction id="R_BIOMASS_Ec_IML1515_core_75p37M" fast="false" reversible="false"
168432 name="E. coli biomass objective function (IML1515) - core - with 75.37 GAM estimate"
168433 metaid="R_BIOMASS_Ec_IML1515_core_75p37M" sboTerm="SBO:0000629"
168434 fbc:upperFluxBound="cobra_default_ub" fbc:lowerFluxBound="cobra_0_bound">
168435 <sbml:annotation xmlns:sbml="http://www.sbml.org/sbml/level3/version1/core">
168436 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
168437 <rdf:Description rdf:about="#R_BIOMASS_Ec_IML1515_core_75p37M">
168438 <bqbiol:is xmlns:bqbiol="http://biomodels.net/biology-qualifiers"/>
168439 <rdf:Bag>
168440 <rdf:li
168441 rdf:resource="http://identifiers.org/bigg_reaction/BIOASS_Ec_IML1515_core_75p37M"/>
168442 </rdf:li>
168443 </rdf:Bag>
168444 </rdf:Description>
168445 </rdf:RDF>
168446 </sbml:annotation>
168447 <listOfReactants>
168448 <speciesReference species="M_10fthf_c" stoichiometry="0.000223"
168449 constant="true"/>
168450 <speciesReference species="M_2fe2s_c" stoichiometry="2.6e-05" constant="true"/>
168451 <speciesReference species="M_2ohph_c" stoichiometry="0.000223" constant="true"/>
168452 <speciesReference species="M_4fe4s_c" stoichiometry="0.00026" constant="true"/>
168453 <speciesReference species="M_ala_l_c" stoichiometry="0.513689"
168454 constant="true"/>
168455 <speciesReference species="M_amet_c" stoichiometry="0.000223" constant="true"/>
168456 <speciesReference species="M_arg_l_c" stoichiometry="0.295792"
168457 constant="true"/>
168458 <speciesReference species="M_asn_l_c" stoichiometry="0.241055"
168459 constant="true"/>
168460 <speciesReference species="M_asp_l_c" stoichiometry="0.241055"
168461 constant="true"/>
168462 <speciesReference species="M_atp_c" stoichiometry="75.55223" constant="true"/>
168463 <speciesReference species="M_bta_c" stoichiometry="2e-06" constant="true"/>
168464 <speciesReference species="M_ca2_c" stoichiometry="0.005205" constant="true"/>
168465 <speciesReference species="M_cl_c" stoichiometry="0.005205" constant="true"/>
168466 <speciesReference species="M_coa_c" stoichiometry="0.000576" constant="true"/>
168467 <speciesReference species="M_cobalt2_c" stoichiometry="2.5e-05"
168468 constant="true"/>
168469 <speciesReference species="M_ctp_c" stoichiometry="0.133508" constant="true"/>
168470 <speciesReference species="M_cu2_c" stoichiometry="0.000709" constant="true"/>
168471 <speciesReference species="M_cys_l_c" stoichiometry="0.09158" constant="true"/>
168472 <speciesReference species="M_datp_c" stoichiometry="0.026166" constant="true"/>
168473 <speciesReference species="M_dctp_c" stoichiometry="0.027017" constant="true"/>
168474 <speciesReference species="M_dgtp_c" stoichiometry="0.027017" constant="true"/>
168475 <speciesReference species="M_dntp_c" stoichiometry="0.026166" constant="true"/>
168476 <speciesReference species="M_fad_c" stoichiometry="0.000223" constant="true"/>
168477 <speciesReference species="M_fe2_c" stoichiometry="0.006715" constant="true"/>
168478 <speciesReference species="M_fe3_c" stoichiometry="0.007808" constant="true"/>
168479 <speciesReference species="M_gln_l_c" stoichiometry="0.26316" constant="true"/>
168480 <speciesReference species="M_glu_l_c" stoichiometry="0.26316" constant="true"/>
168481 <speciesReference species="M_gly_c" stoichiometry="0.612638" constant="true"/>
168482 <speciesReference species="M_gtp_c" stoichiometry="0.215096" constant="true"/>
168483 <speciesReference species="M_h2o_c" stoichiometry="70.028756" constant="true"/>
168484 <speciesReference species="M_his_l_c" stoichiometry="0.094738"
168485 constant="true"/>
168486 <speciesReference species="M_ile_l_c" stoichiometry="0.290529"
168487 constant="true"/>
168488 <speciesReference species="M_k_c" stoichiometry="0.195193" constant="true"/>
168489 <speciesReference species="M_kdo2lipid4_e" stoichiometry="0.019456"
168490 constant="true"/>
168491 <speciesReference species="M_leu_l_c" stoichiometry="0.450531"
168492 constant="true"/>
168493 <speciesReference species="M_lys_l_c" stoichiometry="0.343161"
168494 constant="true"/>

```



Reaction abbreviation	Reaction name	E. C.number	GPR
HEX1	Hexokinase (D-glucose:ATP)	2.7.1.1	(b2388)
GNK	Gluconokinase	2.7.1.12	(b3437) or (b4268)
ANS	Anthranilate synthase	4.1.3.27	(b1264) and (b1263)
PFK	Phosphofruktokinase	2.7.1.11	(b1723) or (b3916)
PFK_2	Phosphofruktokinase (2)	2.7.1.11	(b3916)

Figure 5 | Gene–protein–reaction (GPR) associations. Examples of GPR associations and their representation in Boolean format are shown for *Escherichia coli*.

Thiele and Palsson. Nature Protocols (2010)
<https://doi.org/10.1038/nprot.2009.203>

```

129891 <reaction id="R_GNK" fast="false" reversible="false" name="Gluconokinase" metaid="R_
129892 <sbml:annotation xmlns:sbml="http://www.sbml.org/sbml/level3/version1/core">
129893 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
129894 <rdf:Description rdf:about="#R_GNK">
129895 <bqbiol:is xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
129896 <rdf:Bag>
129897 <rdf:li rdf:resource="http://identifiers.org/biggs.reaction/GNK"/>
129898 <rdf:li rdf:resource="http://identifiers.org/biocyc/META:GLUCONOKIN-RXN"/>
129899 <rdf:li rdf:resource="http://identifiers.org/ec-code/2.7.1.12"/>
129900 <rdf:li rdf:resource="http://identifiers.org/kegg.reaction/R01737"/>
129901 <rdf:li rdf:resource="http://identifiers.org/metanetx.reaction/MNXR10039"/>
129902 <rdf:li rdf:resource="http://identifiers.org/rhea/19434"/>
129903 <rdf:li rdf:resource="http://identifiers.org/rhea/19436"/>
129904 <rdf:li rdf:resource="http://identifiers.org/rhea/19433"/>
129905 <rdf:li rdf:resource="http://identifiers.org/rhea/19435"/>
129906 <rdf:li rdf:resource="http://identifiers.org/sabiork/1845"/>
129907 <rdf:li rdf:resource="http://identifiers.org/seed.reaction/rxn01275"/>
129908 </rdf:Bag>
129909 </bqbiol:is>
129910 </rdf:Description>
129911 </rdf:RDF>
129912 </sbml:annotation>
129913 <listOfReactants>
129914 <speciesReference species="M_atp_c" stoichiometry="1" constant="true"/>
129915 <speciesReference species="M_glc_n_c" stoichiometry="1" constant="true"/>
129916 </listOfReactants>
129917 <listOfProducts>
129918 <speciesReference species="M_6pgc_c" stoichiometry="1" constant="true"/>
129919 <speciesReference species="M_adp_c" stoichiometry="1" constant="true"/>
129920 <speciesReference species="M_h_c" stoichiometry="1" constant="true"/>
129921 </listOfProducts>
129922 <fbc:geneProductAssociation>
129923 <fbc:or>
129924 <fbc:geneProductRef fbc:geneProduct="G_b3437"/>
129925 <fbc:geneProductRef fbc:geneProduct="G_b4268"/>
129926 </fbc:or>
129927 </fbc:geneProductAssociation>
129928 </reaction>

```

```

84175 <reaction id="R_ANS" fast="false" reversible="false" name="Anthranilate synthase" me
84176 <sbml:annotation xmlns:sbml="http://www.sbml.org/sbml/level3/version1/core">
84177 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
84178 <rdf:Description rdf:about="#R_ANS">
84179 <bqbiol:is xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
84180 <rdf:Bag>
84181 <rdf:li rdf:resource="http://identifiers.org/biggs.reaction/ANS"/>
84182 <rdf:li rdf:resource="http://identifiers.org/biocyc/META:ANTHRANSYN-RXN"
84183 <rdf:li rdf:resource="http://identifiers.org/ec-code/4.1.3.27"/>
84184 <rdf:li rdf:resource="http://identifiers.org/kegg.reaction/R00986"/>
84185 <rdf:li rdf:resource="http://identifiers.org/metanetx.reaction/MNXR95843
84186 <rdf:li rdf:resource="http://identifiers.org/rhea/21734"/>
84187 <rdf:li rdf:resource="http://identifiers.org/rhea/21733"/>
84188 <rdf:li rdf:resource="http://identifiers.org/rhea/21735"/>
84189 <rdf:li rdf:resource="http://identifiers.org/rhea/21732"/>
84190 <rdf:li rdf:resource="http://identifiers.org/sabiork/1280"/>
84191 <rdf:li rdf:resource="http://identifiers.org/seed.reaction/rxn00727"/>
84192 </rdf:Bag>
84193 </bqbiol:is>
84194 </rdf:Description>
84195 </rdf:RDF>
84196 </sbml:annotation>
84197 <listOfReactants>
84198 <speciesReference species="M_chor_c" stoichiometry="1" constant="true"/>
84199 <speciesReference species="M_gln_L_c" stoichiometry="1" constant="true"/>
84200 </listOfReactants>
84201 <listOfProducts>
84202 <speciesReference species="M_anth_c" stoichiometry="1" constant="true"/>
84203 <speciesReference species="M_glu_L_c" stoichiometry="1" constant="true"/>
84204 <speciesReference species="M_h_c" stoichiometry="1" constant="true"/>
84205 <speciesReference species="M_pyr_c" stoichiometry="1" constant="true"/>
84206 </listOfProducts>
84207 <fbc:geneProductAssociation>
84208 <fbc:and>
84209 <fbc:geneProductRef fbc:geneProduct="G_b1263"/>
84210 <fbc:geneProductRef fbc:geneProduct="G_b1264"/>
84211 </fbc:and>
84212 </fbc:geneProductAssociation>
84213 </reaction>

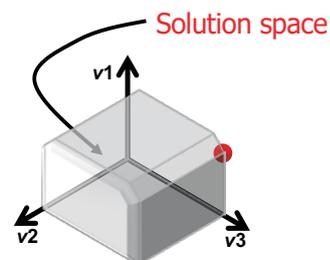
```

Flux balance analysis: simulation of the stoichiometric model with optimization technique (e.g. linear programming)

Maximize $Z = \sum_{j \in J} c_j v_j$ (Biomass formation equation)

Subject to $S \cdot v = 0$

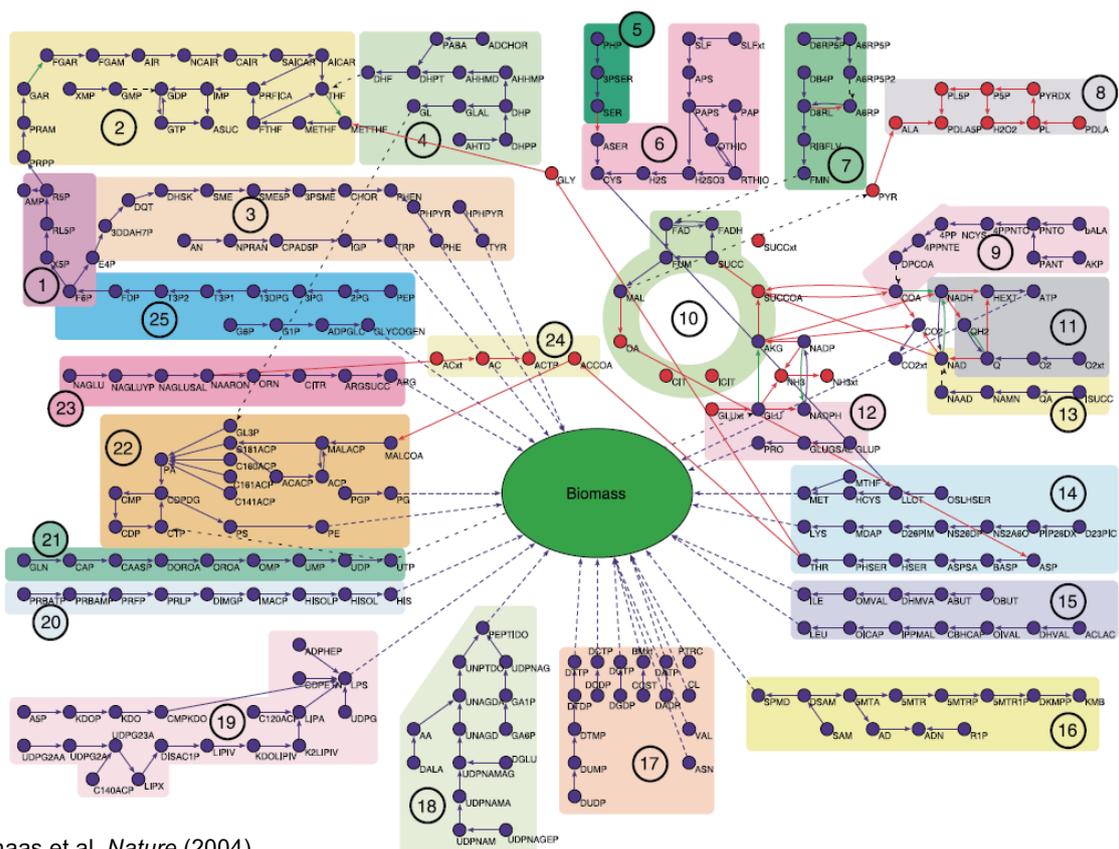
$$\alpha_j \leq v_j \leq \beta_j, \quad \forall j \in J$$



● Optimal flux distributions from constraint-based flux analysis

Options for objective function

- Maximizing **cell growth rate**
- Maximizing **particular metabolite (biochemical) production rate**
- Minimizing **energy consumption**
- ⋮



Almaas et al. *Nature* (2004)
<https://doi.org/10.1038/nature02289>

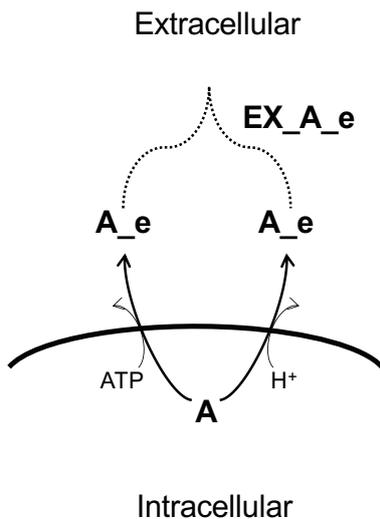
Biomass equation

Precursor metabolite	Building blocks	umol/gDW
Glucose-6-phosphate	UDP-glucose, UDP-galactose	205
Fructose-6-phosphate	UDP-N-acetylglucoseamine	71
Ribose-5-phosphate	His, Trp, nucleotides	898
Erythrose-4-phosphate	Phe, Trp, Tyr	361
Glyceraldehyde-3-P	Backbone of phospholipids	129
3-phosphoglycerate	Cys, Gly, Ser, choline, nucleotides	1496
Phosphoenolpyruvate	Phe, Trp, Tyr	519
Pyruvate	Alanine, isoleucine, valine	2833
Acetyl-CoA	Lipids	3747
2-oxoglutarate	Arg, Glu, Gln, Pro	1079
Succinyl-CoA	Hemes	
Oxaloacetate	Asp, Asn, Ile, Met, Thr, Lys, nucleotides	1787

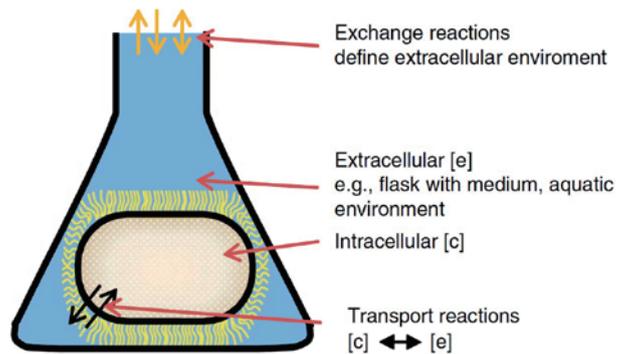
$- 0.20 \text{ G6P} - 0.07 \text{ F6P} - 0.89 \text{ R5P} - 0.36 \text{ E4P} - 0.12 \text{ T3P1} - 1.49 \text{ 3PG}$
 $- 0.51 \text{ PEP} - 2.83 \text{ PYR} - 1.78 \text{ OA} - 1.07 \text{ AKG} - 3.74 \text{ AcCoA}$
 $- 41.25 \text{ ATP} - 3.54 \text{ NAD} - 18.22 \text{ NADPH}$

$+ \text{"1 g DW Biomass"} + 3.74 \text{ CoA}$
 $+ 41.25 \text{ ADP} + 41.25 \text{ PI} + 3.54 \text{ NADH} + 18.22 \text{ NADP} = 0$

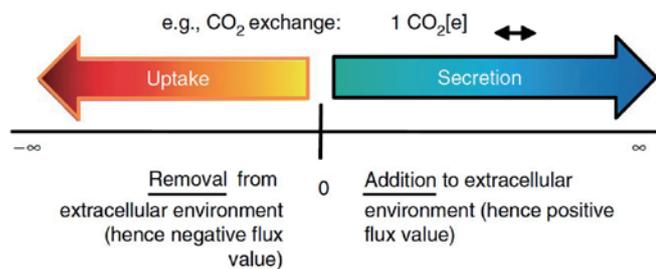
Exchange reactions define medium and environment



2. Definition of systems boundaries



Exchange reactions are defined as follows:



Thiele and Palsson. *Nature Protocols* (2010)
<https://doi.org/10.1038/nprot.2009.203>

Predicting the effects of nutrients

- Modify constraints for exchange reactions.
- Exchange reactions are identified by IDs that have the prefix 'EX_' and the suffix '_e'.

Reaction identifier	EX_co2_e
Name	CO2 exchange
Memory address	0x07fef04703d90
	co2_e <=>
Stoichiometry	CO2 <=>
GPR	
Lower bound	-1000.0
Upper bound	1000.0

- If you want to examine the effects of using succinate instead of glucose as a main carbon source, set lower constraint of 'EX_glc__D_e' as '0', and lower constraint of 'EX_succ_e' with some negative value (e.g., '-10').

Table 1
The frameworks of algorithms and description of objective functions.*

Algorithm	Objective function	Constraints	Solver	Description
FBA	$\max/\min v_j$	$\sum_{j=1}^M S_{ij} v_j = 0, \forall i \in N, \forall j \in M$ $v_{\min} \leq v \leq v_{\max}$	LP	Usually maximizing the growth rate
MOMA	$\min \sqrt{\sum_{j=1}^M (w_j - x_j)^2}$	$\sum_{j=1}^M S_{ij} v_j = 0, \forall i \in N, \forall j \in M$ $v_{\min} \leq v \leq v_{\max}$ $v_j = 0, \forall i \in A$	QP	Minimizing the Euclidian distance from a wild type flux distribution under knockout condition
ROOM	$\min y_j$	$\sum_{j=1}^M S_{ij} v_j = 0, \forall i \in N, \forall j \in M$ $v_{\min} \leq v \leq v_{\max}$ $v_j = 0, \forall i \in A$ $y_j - y_j (v_{\max,j} - w_j^*) \leq w_j^*$ $y_j - y_j (v_{\min,j} - w_j^*) \geq w_j^*$ $w_j^* = w_j + \delta w_j + \epsilon$ $w_j^* = w_j - \delta w_j - \epsilon$ $y_j \in \{0, 1\}$	MILP	Minimizing the number of significant flux changes from a wild type flux distribution under knockout condition
OptKnock	$\max v_{\text{biochemical}}$	$\max v_{\text{biochemical}} \sum_{j=1}^M S_{ij} v_j = 0, \forall i \in N, \forall j \in M$ $v_{\min} \leq v \leq v_{\max}$ $v_{\text{knockout}} > v_{\text{biochemical}}$ $v_{\text{knockout},j} \leq v_j \leq v_{\text{knockout},j}$ $\sum_{j \in K} (1 - y_j) \leq K, y_j \in \{0, 1\}$	MILP	Bilevel optimization framework: biomass, biochemical
OptReg	$\max v_{\text{biochemical}}$	$\max v_{\text{biochemical}} - \gamma \sum_j y_j$ $\sum_{j=1}^M S_{ij} v_j = 0, \forall i \in N, \forall j \in M$ $v_{\min} \leq v \leq v_{\max}$ Downregulation $v_{\text{knockout},j} \leq y_j \leq \left(\frac{v_{\text{knockout},j}}{v_{\text{biochemical},j}}\right) \cdot (1 - \gamma) + v_{\text{knockout},j}$ Upregulation $\left[\left(\frac{v_{\text{knockout},j}}{v_{\text{biochemical},j}}\right) \cdot (1 - \gamma) + v_{\text{knockout},j}\right] \cdot (1 - \gamma) + v_{\text{knockout},j} \leq v_j \leq v_{\text{knockout},j}$ Knockout $v_{\text{knockout},j} \leq y_j \leq v_{\text{knockout},j}$ $y_j \in \{0, 1\}$	MILP	Determining the activation/inhibition and elimination reaction set for biochemical production
OptStrain	Step1: $\max MW, \sum_{j=1}^M S_{ij} v_j, i = P$ Step2: $\min \sum_{j \in M_{\text{non-native}}} y_j$	Step1: $\sum_{j=1}^M S_{ij} v_j > 0, \forall i \in N, \forall i \in R, \forall j \in M$ $v_{\min} \leq v \leq v_{\max}$ $\sum_{i \in I} \left(\sum_{j=1}^M S_{ij} v_j \right) = -1$ Step2: $\sum_{j=1}^M S_{ij} v_j > 0, \forall i \in N, \forall i \in R, \forall j \in M$ $v_{\min} \leq v \leq v_{\max}$ $\sum_{j=1}^M S_{ij} v_j \geq \text{Yield}^{\text{opt}}, i = P$ $v_{\text{knockout},j} \leq v_j \leq v_{\text{knockout},j}, \forall j \in M_{\text{non-native}}$ $y_j \in \{0, 1\}, \forall j \in M_{\text{non-native}}$	LP MILP	Determining the maximum yield of the desired biochemical and minimizing the number of non-native reactions needed to meet the maximum yield of desired biochemical production
OMNI	$\min \sum_j w_j v_j^{\text{opt}} - v_j^{\text{exp}} $	$v_j^{\text{opt}} = \max v_{\text{biochemical}}$ $\sum_{j=1}^M S_{ij} v_j = 0$ $v_{\min} \leq v \leq v_{\max}, \forall i \in N, \forall j \in M$ $v_{\text{knockout},j} \leq v_j \leq v_{\text{knockout},j}, \forall j \in M_{\text{non-native}}$ $v_j = v_j^{\text{exp}}, i \in E$ $v_{\text{knockout},j} > v_{\text{biochemical},j}$ $\sum_{j \in K} (1 - y_j) \leq K, y_j \in \{0, 1\}$	MILP	Identifying the reaction set that leads the best agreement between prediction and experiment



In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data

Jeremy S. Edwards^{1,2}, Rafael U. Ibarra¹, and Bernhard O. Palsson^{1*}

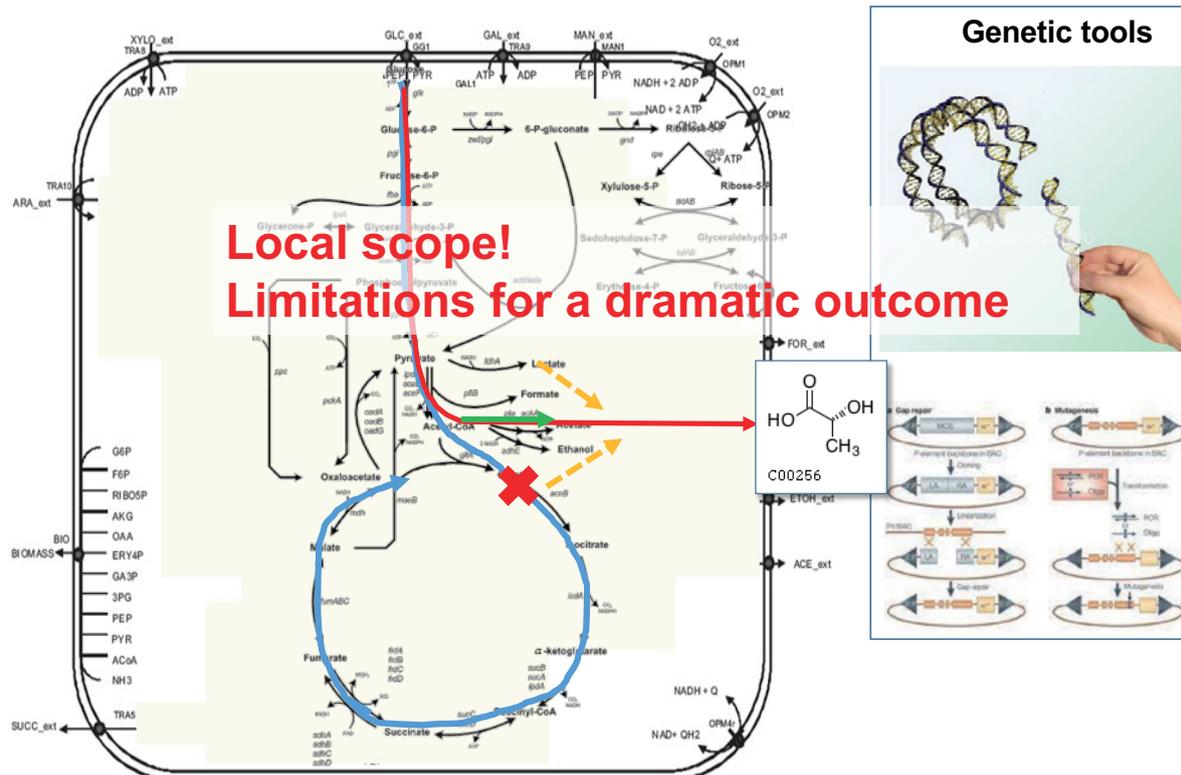
¹Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412. ²Current address: Department of Chemical Engineering, University of Delaware, Newark, DE 19716. *Corresponding author (palsson@ucsd.edu).

Received 19 September 2000; accepted 9 November 2000

A significant goal in the post-genome era is to relate the annotated genome sequence to the physiological functions of a cell. Working from the annotated genome sequence, as well as biochemical and physiological information, it is possible to reconstruct complete metabolic networks. Furthermore, computational methods have been developed to interpret and predict the optimal performance of a metabolic network under a range of growth conditions. We have tested the hypothesis that *Escherichia coli* uses its metabolism to grow at a maximal rate using the *E. coli* MG1655 metabolic reconstruction. Based on this hypothesis, we formulated experiments that describe the quantitative relationship between a primary carbon source (acetate or succinate) uptake rate, oxygen uptake rate, and maximal cellular growth rate. We found that the experimental data were consistent with the stated hypothesis, namely that the *E. coli* metabolic network is optimized to maximize growth under the experimental conditions considered. This study thus demonstrates how the combination of *in silico* and experimental biology can be used to obtain a quantitative genotype–phenotype relationship for metabolism in bacterial cells.

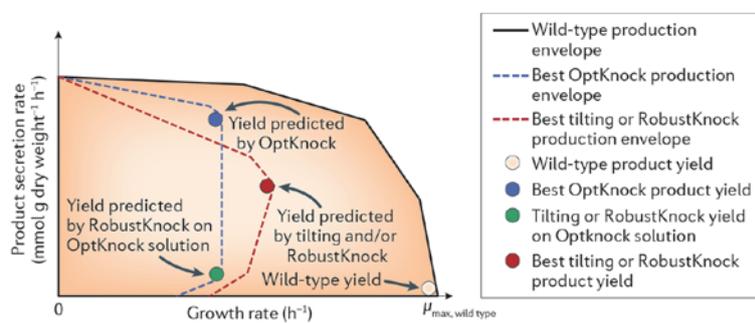
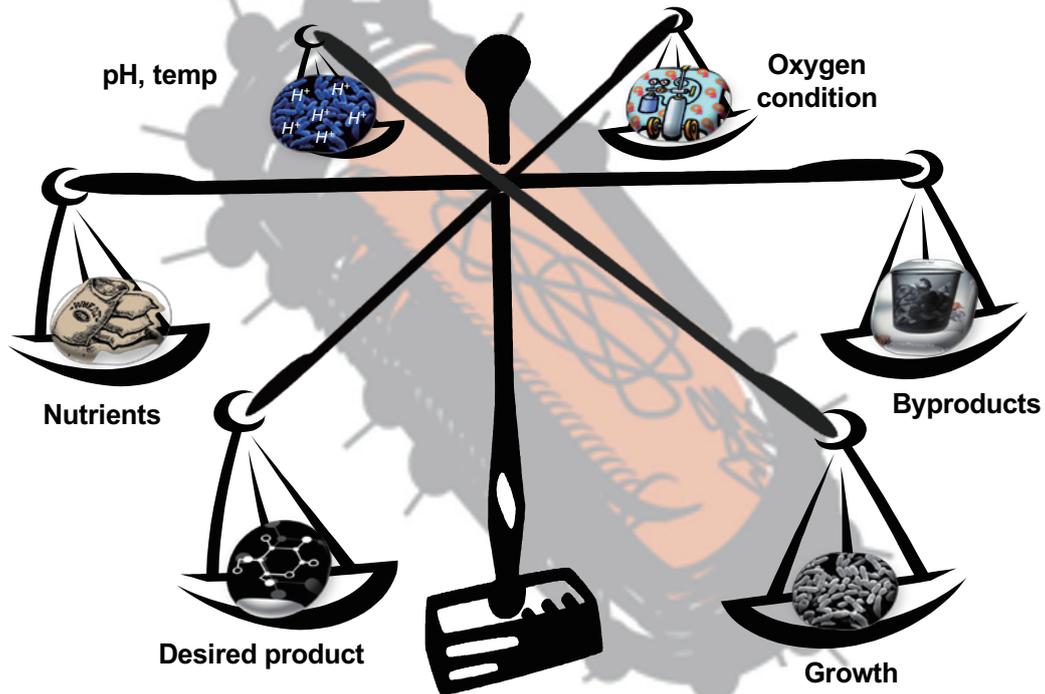
Keywords: *Escherichia coli*, genome analysis, metabolic reconstruction, computer simulation

Gene modification strategies for metabolic engineering



Factors affecting an organism's metabolism

Cells do not overproduce what we want!



Lewis et al. Nature Reviews Microbiology (2012)
<https://www.nature.com/articles/nrmicro2737>

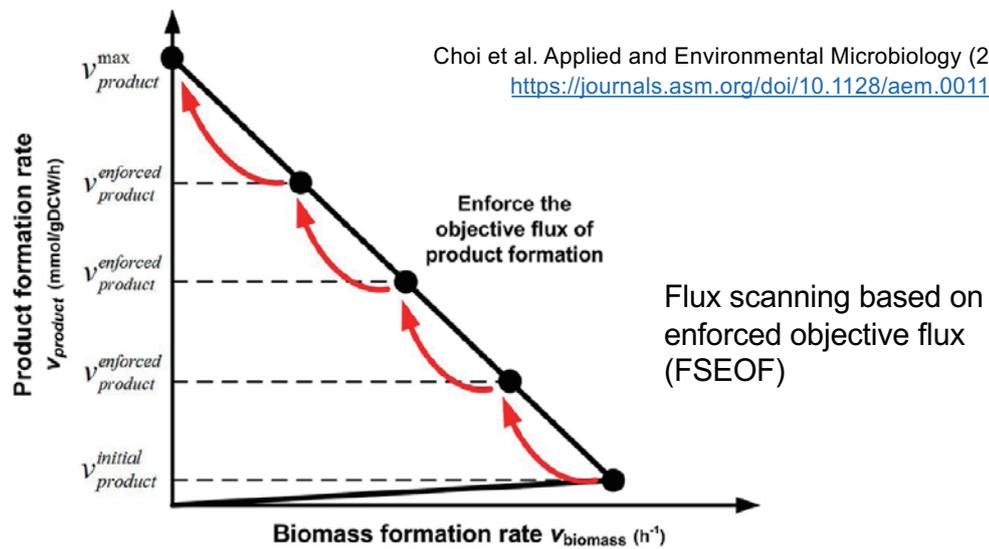
maximize bioengineering objective
 (through gene knockouts)
subject to **maximize** cellular objective
 (over fluxes)
subject to

- fixed substrate uptake
- network stoichiometry
- blocked reactions identified by outer problem

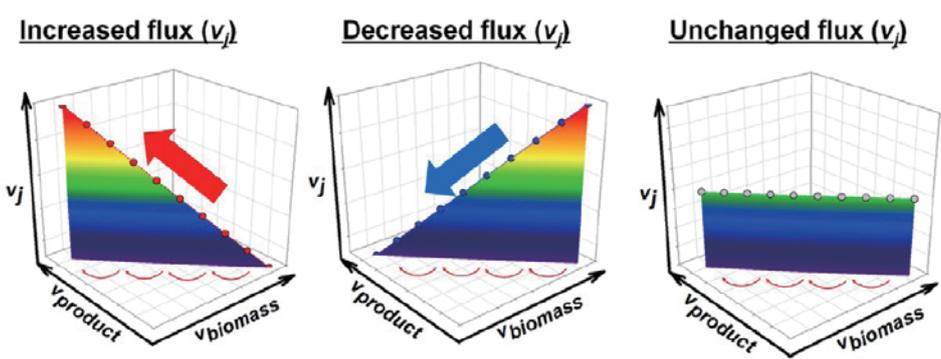
 number of knockouts \leq limit

maximize $v_{chemical}$ (OptKnock)
 y_j
 subject to maximize v_j (Primal)
 $v_{biomass}$
 subject to $\sum_{j=1}^M S_{ij} v_j = 0,$
 $v_{pts} + v_{glk} = v_{glc_uptake}$
 $v_{atp} \geq v_{atp_main}$
 $v_{biomass} \geq v_{biomass}^{target}$
 $v_j^{min} \cdot y_j \leq v_j \leq v_j^{max} \cdot y_j, \forall j \in \mathcal{M}$
 $y_j = \{0, 1\}, \forall j \in \mathcal{M}$
 $\sum_{j \in \mathcal{M}} (1 - y_j) \leq K$

Burgard et al. Biotechnology and Bioengineering (2003)
<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/bit.10803>



Choi et al. Applied and Environmental Microbiology (2010)
<https://journals.asm.org/doi/10.1128/aem.00115-10>



Systems metabolic engineering of *Streptomyces venezuelae* for the enhanced production of pikromycin

Min Kyung Cho¹ | Byung Tae Lee² | Hyun Uk Kim²  | Min-Kyu Oh¹ 

¹Department of Chemical and Biological Engineering, Korea University, Seoul, Republic of Korea

²Department of Chemical and Biomolecular Engineering (BK21 four), Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

Correspondence

Hyun Uk Kim, Department of Chemical and Biomolecular Engineering (BK21 four), Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea. Email: ehukim@kaist.ac.kr

Min-Kyu Oh, Department of Chemical and Biological Engineering, Korea University, Seoul 02841, Republic of Korea. Email: mkoh@korea.ac.kr

Funding information

Rural Development Administration; National Research Foundation of Korea

Abstract

Pikromycin is an important precursor of drugs, for example, erythromycin. Hence, systems metabolic engineering for the enhanced pikromycin production can contribute to the development of pikromycin-related drugs. In this study, metabolic genes in *Streptomyces venezuelae* were systematically engineered for enhanced pikromycin production. For this, a genome-scale metabolic model of *S. venezuelae* was reconstructed and simulated, which led to the selection of 11 metabolic gene targets. These metabolic genes, including four overexpression targets and seven knockdown targets, were individually engineered first. Next, two overexpression targets and two knockdown targets were selected based on the 11 strains' production performances to engineer two to four of these genes together for the potential synergistic effects on the pikromycin production. As a result, the NM1 strain with AQF52_RS24510 (methenyltetrahydrofolate cyclohydrolase/methylenetetrahydrofolate dehydrogenase) overexpression and AQF52_RS30320 (sulfite reductase) knockdown showed the best production performance among all the 22 strains constructed in this study. Fed-batch fermentation of the NM1 strain produced 295.25 mg/L of pikromycin, by far the best production titer using the native producer *S. venezuelae*, to the best of our knowledge. The systems metabolic engineering strategy demonstrated herein can also be applied to the overproduction of other secondary metabolites using *S. venezuelae*.

KEYWORDS

gene manipulation targets, genome-scale metabolic model, pikromycin, *Streptomyces venezuelae*, systems metabolic engineering

<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/bit.28114>

2 | MATERIALS AND METHODS

2.1 | Generation of a GEM of *S. venezuelae*

A GEM of *S. venezuelae* was reconstructed using a Python-based GEM reconstruction tool as previously described (Jang et al., 2019). This reconstruction tool processes entire protein sequences and their corresponding enzyme commission (EC) numbers for entire metabolic genes in an organism of interest to generate a draft GEM. Protein sequences for metabolic genes of *S. venezuelae* ATCC 15439 were obtained from the genome data (accession number NZ_CP013129.1) (Song et al., 2016) available at RefSeq database. EC numbers relevant to the protein sequences were predicted using DeepEC (Ryu et al., 2019). The Python-based GEM reconstruction tool also requires high-quality GEM and protein sequences of a biologically close organism to generate the draft GEM of a target organism. In this study, a high-quality *Streptomyces coelicolor* GEM iKS1317 was used as a template GEM (Kumelj et al., 2019). Finally, a pikromycin biosynthetic reaction (reaction ID: "PIKROMYCIN_SYN") was newly inserted to the GEM. The Systems Biology Markup Language (SBML) file of the resulting *S. venezuelae* GEM, SvenSBML1450, is available in Supporting Information: Data S1.

2.2 | Prediction of gene manipulation targets

A constraint-based flux analysis method, called flux scanning based on enforced objective flux (FSEOF), was applied to SvenSBML1450 to identify effective gene manipulation targets for the enhanced pikromycin production (Choi et al., 2010). FSEOF examines biochemical reactions whether they show flux values positively or negatively correlated with the flux of a target product formation. Reactions with positive correlations were considered as overexpression targets, and those with negative correlations as knockdown targets; genes associated with the positively or negatively correlated reactions were selected as gene manipulation targets. Glucose uptake rate was set to 10 mmol/g DCW/h, while exchange reaction fluxes of other nutrients were constrained with a lower bound of -1000 mmol/g DCW/h when implementing FSEOF was implemented by using Cameo (Cardoso et al., 2016).

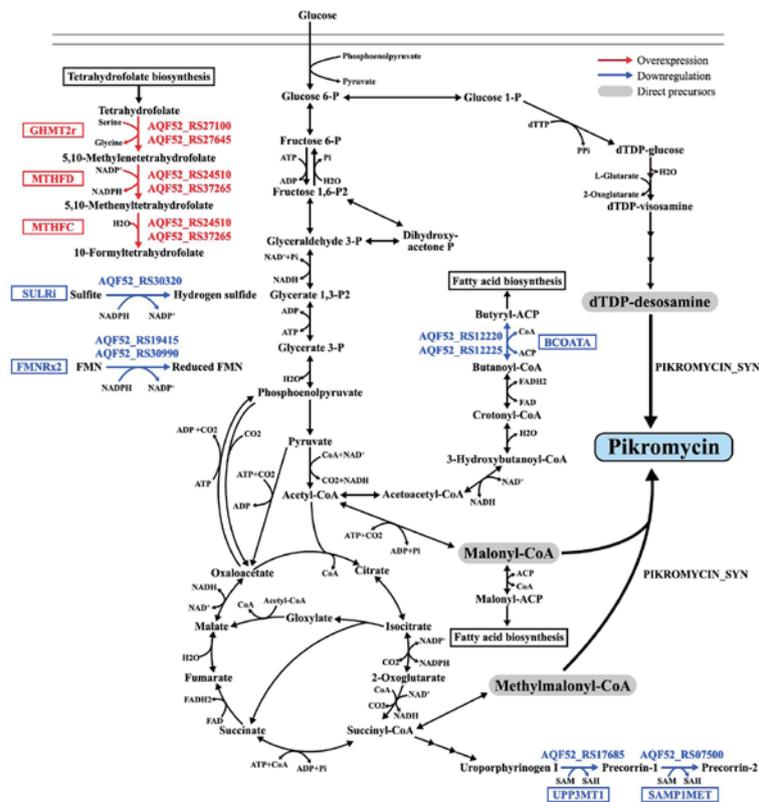
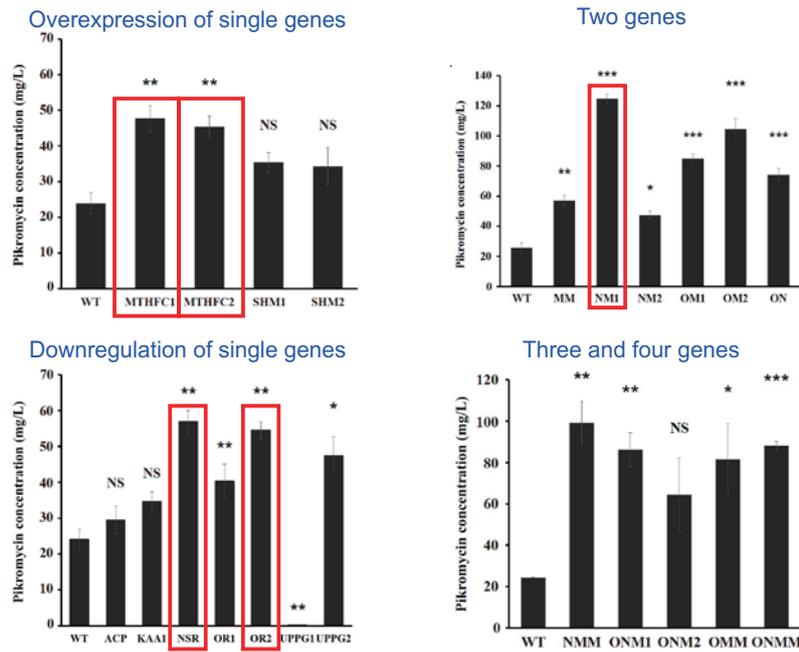
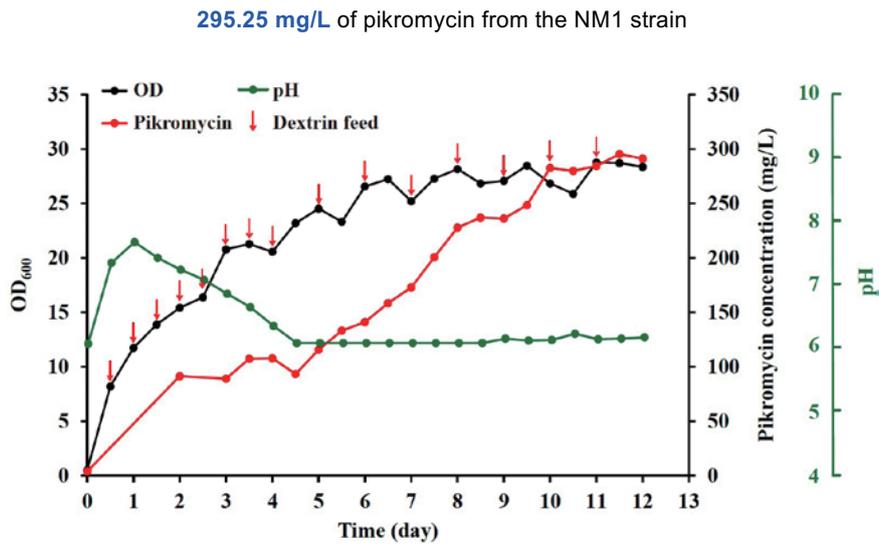


FIGURE 1 Metabolic pathways of *S. venezuelae* that present gene manipulation targets predicted using the genome-scale metabolic model SvenSBML1450. Locus tags in red and blue indicate overexpression and knockdown targets, respectively. Reaction IDs are also presented in boxes for each gene manipulation target. Enzyme names for the presented reaction IDs are: BCOATA, butyryl-CoA-ACP transacylase; FMNR2, FMN reductase; GHMT2r, serine hydroxymethyltransferase; MTHFC, methenyltetrahydrofolate cyclohydrolase; MTHFD, methylenetetrahydrofolate dehydrogenase (NADP⁺); SAMP1MET, precorrin-1 methyltransferase; SULR1, sulfite reductase; and UPP3MT1, uroporphyrin-I methyltransferase.

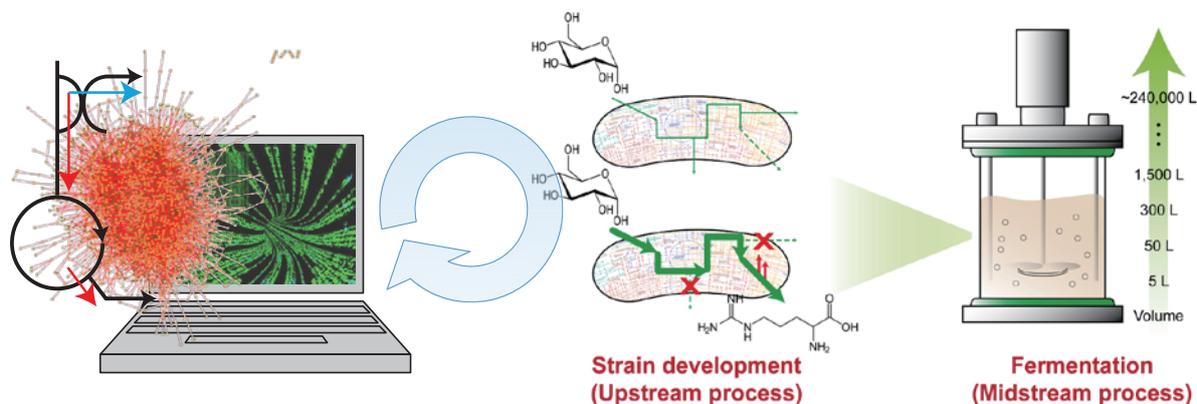


Cho et al. *Biotechnology and Bioengineering* (2022)
<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/bit.28114>



Cho et al. *Biotechnology and Bioengineering* (2022)
<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/bit.28114>

Systems metabolic engineering



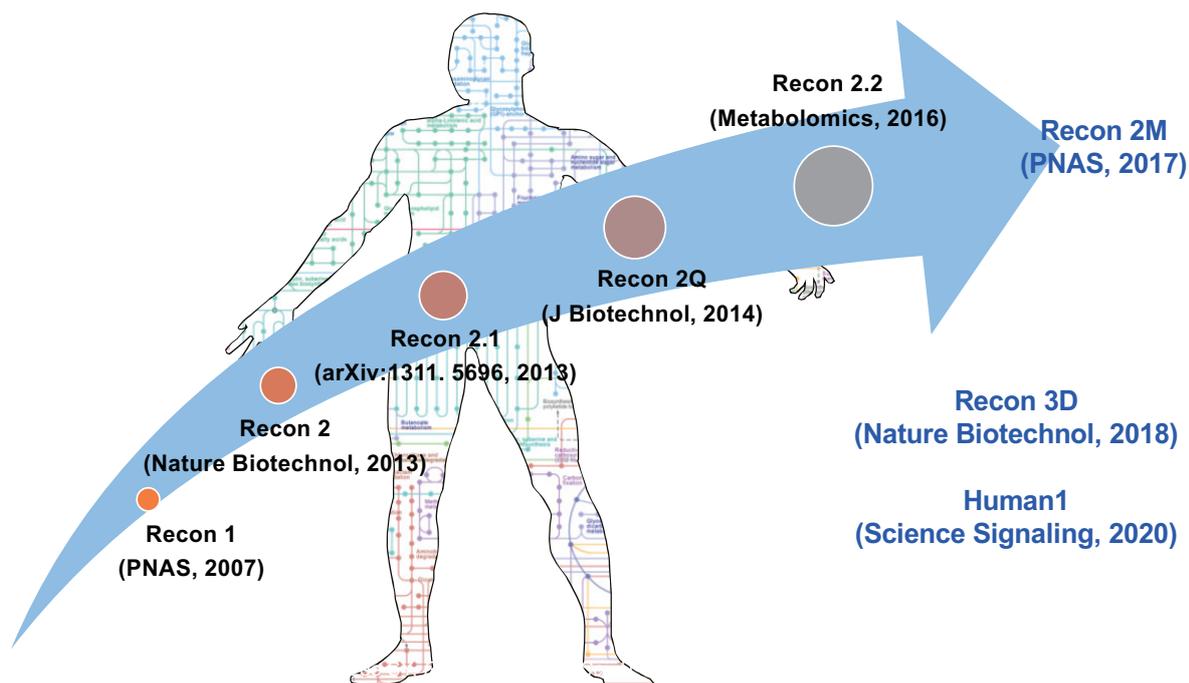
Dry experiments

- Biological data analysis (e.g., omics)
- Metabolic modeling & simulations
- Designing strategies (e.g., gene targeting)
- Software development

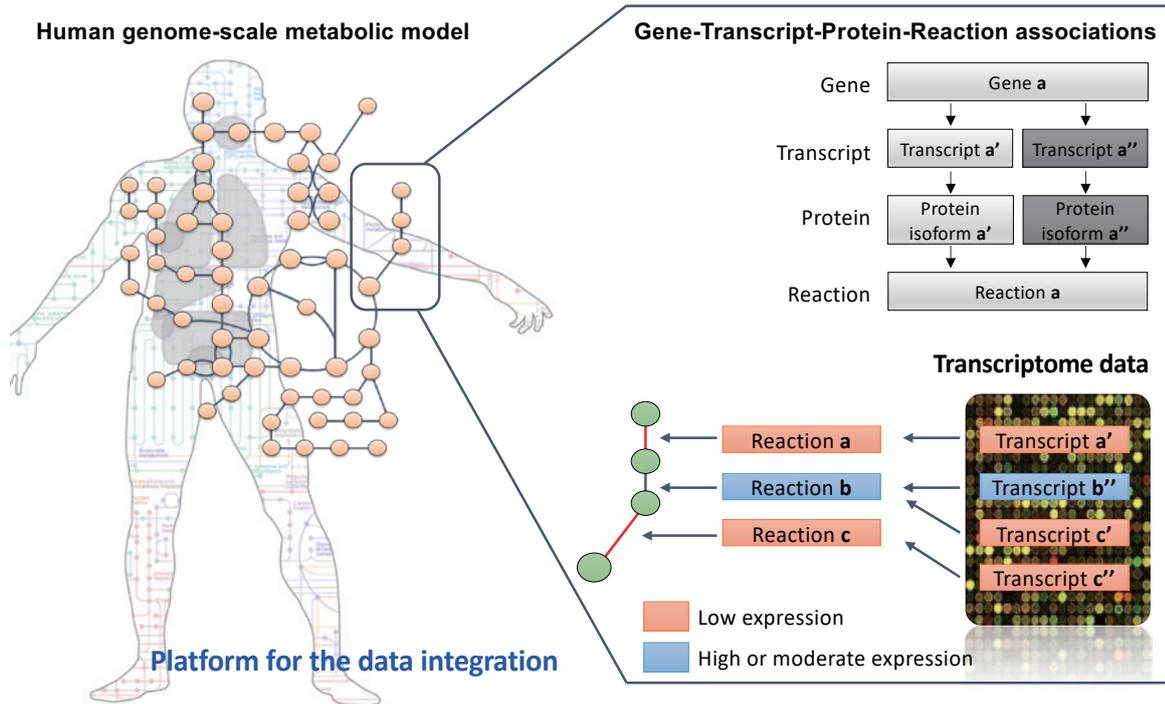
Wet experiments

- Metabolic engineering
- Production of biochemicals
- Bioprocess engineering

History of the development of human genome-scale metabolic models



Platform for the data integration – cell-specific model



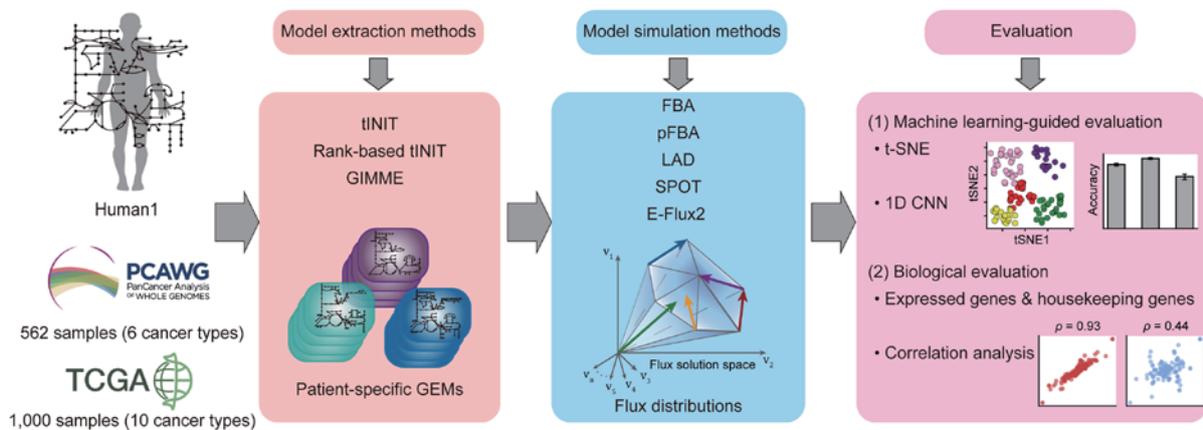
Source code: <https://bitbucket.org/kaistmbel/recon-manager>

Ryu et al. PNAS (2017)

Lee et al. Computational and Structural Biotechnology Journal (2022)

Machine learning-guided evaluation of extraction and simulation methods for cancer patient-specific metabolic models

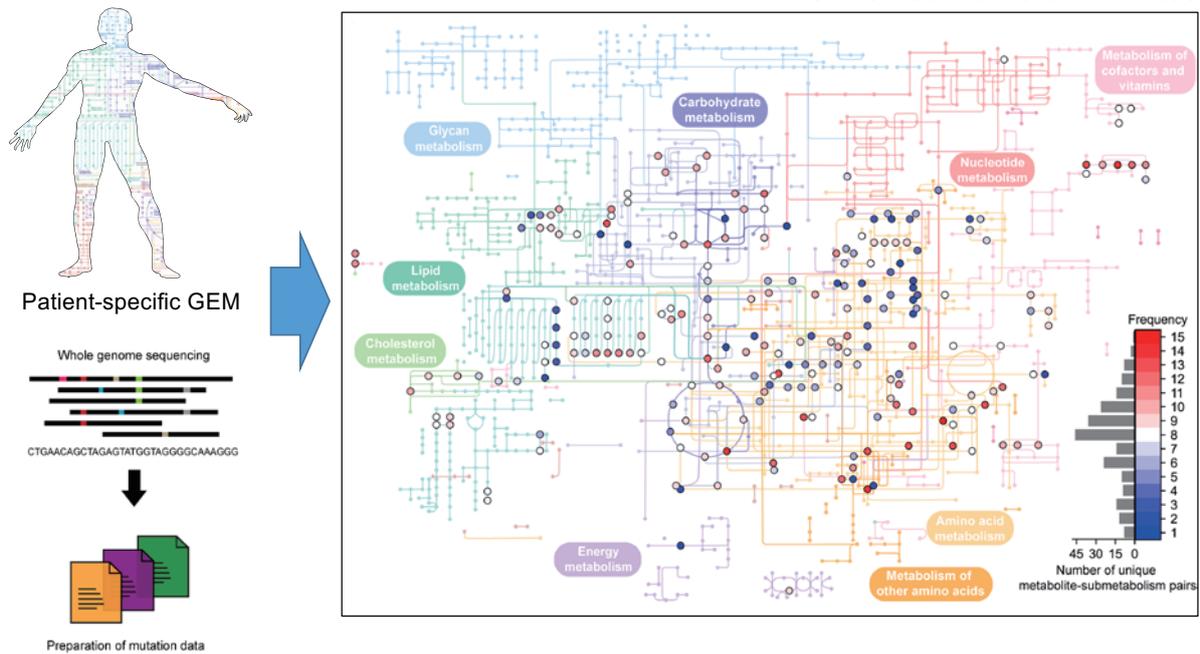
Best combination of algorithms?



Lee et al. Computational and Structural Biotechnology Journal (2022)

<https://www.sciencedirect.com/science/article/pii/S2001037022002434?via%3Dihub>

Oncometabolite candidates



Lee et al. Genome Biology (2024)

<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-024-03208-8>

Distribution of mutation-associated metabolites for the 18 cancer types across the genome-scale human metabolic pathways

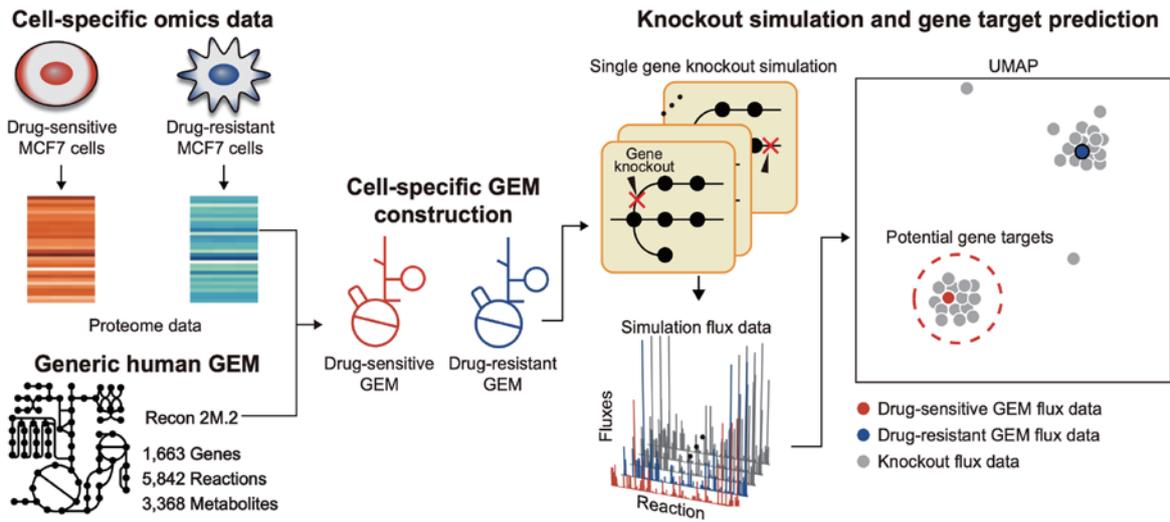
Supplementary Data 5. A list of 4,335 metabolite-gene-pathway sets (MGPs) predicted for the PCAWG and TCGA samples that represent 18 cancer types.

Cancer type	Target metabolite	Target metabolite MetaNetX	Target gene	Target pathway
Ovary-AdenoCA	geranyl diphosphate	MNXM100	GREB1	Cholesterol metabolism
Ovary-AdenoCA	O-acetyl-(R)-carnitine	MNXM1028	RELN	Fatty acid oxidation
Ovary-AdenoCA	3-hydroxy-2-methylpropanoyl-CoA	MNXM1034	ADAMTS7	Valine, leucine, and isoleucine metabolism
Ovary-AdenoCA	(3R,5S)-1-pyrroline-3-hydroxy-5-carboxylate	MNXM114091	COL3A1	Arginine and Proline Metabolism
Ovary-AdenoCA	(3R,5S)-1-pyrroline-3-hydroxy-5-carboxylate	MNXM114091	STAB1	Arginine and Proline Metabolism
Ovary-AdenoCA	Electron transfer flavoprotein oxidized	MNXM11479	AHNAK2	Fatty acid oxidation
Ovary-AdenoCA	Electron transfer flavoprotein reduced	MNXM11480	AHNAK2	Fatty acid oxidation
Ovary-AdenoCA	L-proline	MNXM114	MGAM	Arginine and Proline Metabolism
Ovary-AdenoCA	L-proline	MNXM114	WDR87	Arginine and Proline Metabolism
Ovary-AdenoCA	L-proline	MNXM114	LRR8C	Arginine and Proline Metabolism
Ovary-AdenoCA	(R)-pantetheine	MNXM1154	SMG1	CoA metabolism
Ovary-AdenoCA	urea	MNXM117	NRDE2	Urea cycle
Ovary-AdenoCA	urea	MNXM117	DST	Arginine and Proline Metabolism
Ovary-AdenoCA	urea	MNXM117	NEB	Urea cycle
Ovary-AdenoCA	putrescine	MNXM118	EPHA2	Urea cycle
Ovary-AdenoCA	putrescine	MNXM118	FNDC1	Arginine and Proline Metabolism
Ovary-AdenoCA	putrescine	MNXM118	ZFH4	Arginine and Proline Metabolism
Ovary-AdenoCA	cysteamine	MNXM1226	SMG1	CoA metabolism
Ovary-AdenoCA	(R)-S-lactoylglutathione	MNXM1253	IQUB	Pyruvate metabolism
Ovary-AdenoCA	3-phospho-D-glycerate	MNXM126	COL3A1	Glycolysis/gluconeogenesis
Ovary-AdenoCA	3-phospho-D-glycerate	MNXM126	MEGF8	Glycolysis/gluconeogenesis
Ovary-AdenoCA	N-formyl-L-kynurenine	MNXM1288	TEX13B	Tryptophan metabolism
Ovary-AdenoCA	N-formyl-L-kynurenine	MNXM1288	WDR87	Tryptophan metabolism
Ovary-AdenoCA	dimethylallyl diphosphate	MNXM132	GREB1	Cholesterol metabolism
Ovary-AdenoCA	acetoacetyl-CoA	MNXM133	NEB	Propanoate metabolism
Ovary-AdenoCA	L-histidine	MNXM134	IQUB	Transport, extracellular
Ovary-AdenoCA	L-histidine	MNXM134	LRP2	Transport, extracellular
Ovary-AdenoCA	L-leucine	MNXM140	CDK12	Transport, extracellular

Lee et al. Genome Biology (2024)

<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-024-03208-8>

The overall scheme of predicting drug-sensitizing gene targets



Lim et al. PNAS (2025)

<https://doi.org/10.1073/pnas.2425384122> a

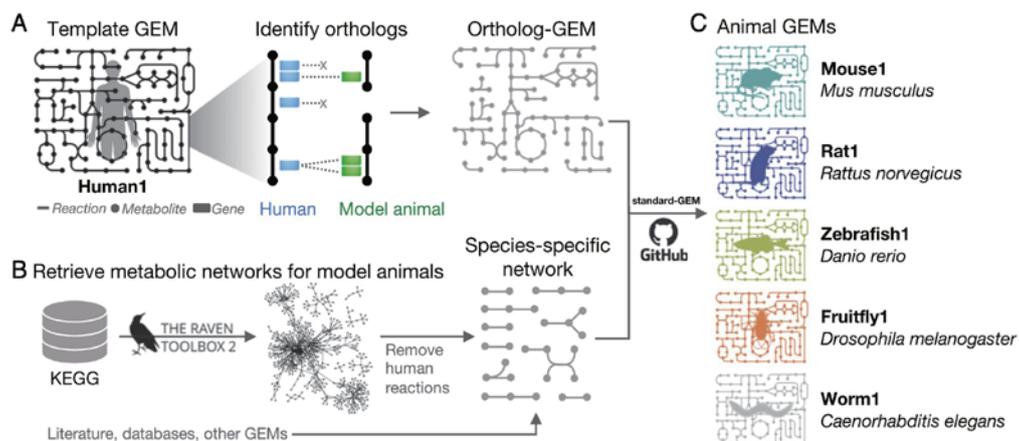


Fig. 1. Genome-scale metabolic modeling for model animals. A reconstruction approach of combining (A) ortholog-GEMs derived from the Human1 template and (B) species-specific metabolic networks extracted from the KEGG database by the RAVEN package was used to obtain (C) the model animal GEMs that were deposited on GitHub according to the standard-GEM scheme (28).

Wang et al. PNAS (2024)

<https://www.pnas.org/doi/full/10.1073/pnas.2102344118>

Lecture 4

Synthetic biology and AI

-합성생물학과 AI 실습-

Kim, Hyun Uk

Department of Chemical and Biomolecular Engineering
Graduate School of Engineering Biology
Korea Advanced Institute of Science and Technology (KAIST)



강의계획서

- 학습목표
 - 합성생물학은 생물학, 공학, 컴퓨터 과학을 결합하여 **효소, 생합성 경로, 또는 전체 생물학적 개체를 설계하고 구축하는** 매우 다학제적인 분야이다.
 - 그 목표는 새로운 생물학적 시스템을 창조하거나 기존 시스템을 유용한 목적으로 재설계하는 것이며, 이 과정에서 **생물학적 도구를 설계할 때 모듈성, 표준화, 확장성이 중요하게** 고려된다.

강의계획서

- 학습목표
 - 본 강의에서는 **합성생물학의 배경**을 먼저 설명하고, 합성생물학이 적용 가능한 다양한 분야 중 **바이오제조와 대사공학**을 중심으로 소개하고자 한다.
 - 또한 이 과정에서 **AI**가 수행하는 핵심적인 역할을 설명한다.
 - 특히 미생물 대사공학의 목표가 대사 네트워크와 생산 공정 전반을 통합적으로 분석하고 최적화하여 **고수율·고생산성 균주**를 개발하는 데 있음을 설명한다. 이를 위해 **균주 설계부터 공정 최적화까지 전 과정**을 포괄하는 **10가지 전략**을 체계적으로 소개한다.
 - 더불어 이 과정에서 유용하게 활용될 수 있는 **대사 네트워크 모델(genome-scale metabolic model)**에 대해서도 이론과 실습을 제공한다.
 - 본 강의에서 소개되는 개념과 방법론은 **약물 표적 발굴 등 시스템 의학 분야**에도 적용 가능하다.

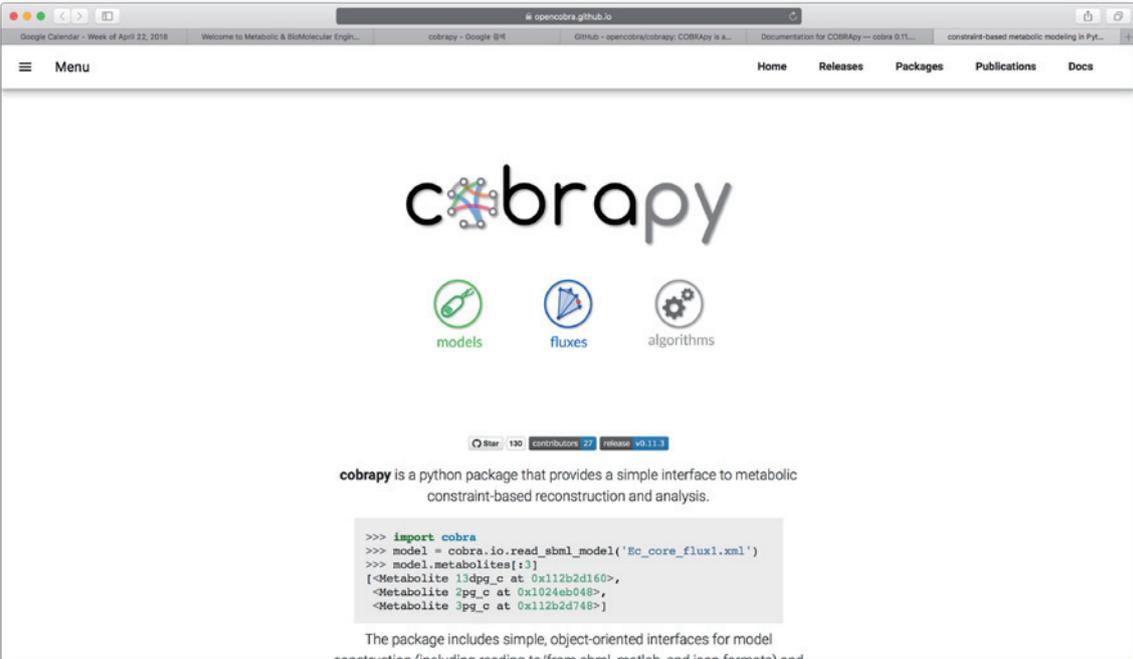
강의계획서

- 강의 1: 합성생물학 기반 바이오제조
- 강의 2: 대사공학 전략
- 강의 3: 대사 네트워크 모델 (Genome-scale metabolic model)
- 강의 4: 합성생물학과 AI 실습

Keywords for biofoundry

- Software programs for genome-scale metabolic models (GEMs)
 - COBRAPy
 - CAMEO (Computer Aided Metabolic Engineering and Optimization)
- Biofoundry
 - Automation
 - DBTL cycle
 - Workflows
 - Hardware
 - Software
 - Global Biofoundry Alliance
- Examples of high-throughput tools and AI tools

COBRA: CONstraint-Based Reconstruction and Analysis



The screenshot shows the COBRAPy website homepage. At the top, there is a navigation bar with links for Home, Releases, Packages, Publications, and Docs. The main content area features the COBRAPy logo, which consists of the word 'cobrapy' in a lowercase, sans-serif font with a colorful molecular structure icon above the 'o'. Below the logo are three circular icons representing 'models', 'fluxes', and 'algorithms'. A GitHub repository statistics bar shows 130 stars, 22 contributors, and the latest release v0.11.3. The text below describes COBRAPy as a Python package for metabolic constraint-based reconstruction and analysis. A code block shows a Python snippet for reading a model and listing metabolites. The snippet is as follows:

```
>>> import cobra
>>> model = cobra.io.read_sbml_model('Ec_core_flux1.xml')
>>> model.metabolites[:3]
[<Metabolite 13dpg_c at 0x112b2d160>,
 <Metabolite 2pg_c at 0x1024eb048>,
 <Metabolite 3pg_c at 0x112b2d748>]
```

The package includes simple, object-oriented interfaces for model construction (including reading to/from sbml, matlab, and json formats) and

Documentation: <https://cobrapy.readthedocs.io/en/latest/>

Source: <https://github.com/opencobra/cobrapy>

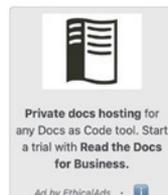
Home: <http://opencobra.github.io/cobrapy/>

cobra

Navigation

1. Getting Started
2. Global Configuration
3. Building a Model
4. Reading and Writing Models
5. Simulating with FBA
6. Simulating Deletions
7. Production envelopes
8. Flux sampling
9. Loopless FBA
10. Consistency testing
11. Gapfilling
12. Growth media
13. Solvers
14. Tailored constraints, variables and objectives
15. Dynamic Flux Balance Analysis (dFBA) in COBRAPy
16. FAQ
17. API

Quick search

Documentation for COBRAPy

For installation instructions, please see [INSTALL.rst](#).

Many of the examples below are viewable as IPython notebooks, which can be viewed at [nbviewer](#).

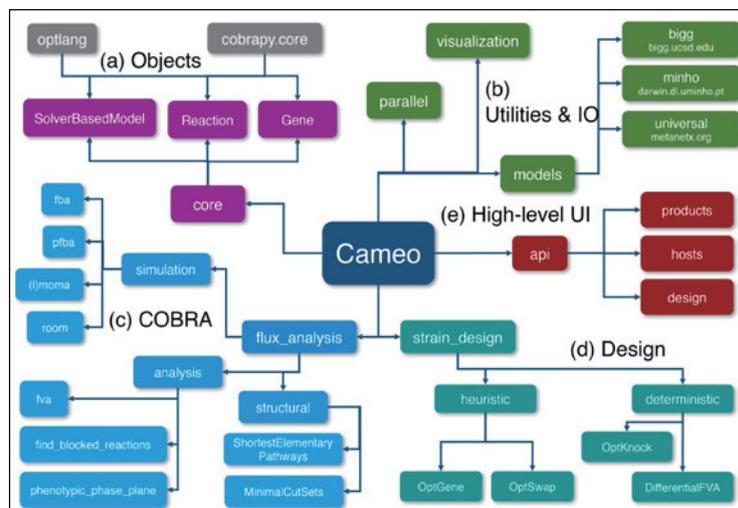
- **1. Getting Started**
 - [1.1. Loading a model and inspecting it](#)
 - [1.2. Reactions](#)
 - [1.3. Metabolites](#)
 - [1.4. Genes](#)
- **2. Global Configuration**
 - [2.1. The configuration object](#)
 - [2.2. Reaction bounds](#)
 - [2.3. Solver](#)
- **3. Building a Model**
 - [3.1. Model, Reactions and Metabolites](#)
 - [3.2. Objective](#)
 - [3.3. Model Validation](#)
 - [3.4. Exchanges, Sinks and Demands](#)
- **4. Reading and Writing Models**
 - [4.1. SBML](#)
 - [4.2. JSON](#)
 - [4.3. YAML](#)
 - [4.4. MATLAB](#)
 - [4.5. Pickle](#)
- **5. Simulating with FBA**
 - [5.1. Running FBA](#)
 - [5.2. Changing the Objectives](#)
 - [5.3. Running FVA](#)
 - [5.4. Running pFBA](#)
 - [5.5. Running geometric FBA](#)
- **6. Simulating Deletions**
 - [6.1. Knocking out single genes and reactions](#)
 - [6.2. Single Deletions](#)
 - [6.3. Double Deletions](#)
 - [6.4. Accessing individual deletion results](#)

Documentation: <https://cobrapy.readthedocs.io/en/latest/>

Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories

João G. R. Cardoso, Kristian Jensen, Christian Lieven, Anne Sofie Lærke Hansen, Svetlana Galkina, Moritz Beber, Emre Özdemir, Markus J. Herrgård, Henning Redestig, and Nikolaus Sonnenschein*

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark



Documentation: <https://cameo.readthedocs.io/en/latest/>

Source: <https://github.com/biosustain/cameo>

Example sources: <https://github.com/biosustain/cameo-notebooks>

Package organization and functionality overview

cameo Tutorials

Watch 27

codecov 63%

Navigation

Installation

Tutorials

- Import models
- Simulate models
- Analyzing models
- Predict gene knockout strategies
- Predict expression modulation targets
- Predict heterologous pathways
- Easy strain design using a high-level interface

Contributing

Development

API Docs

Quick search

Go

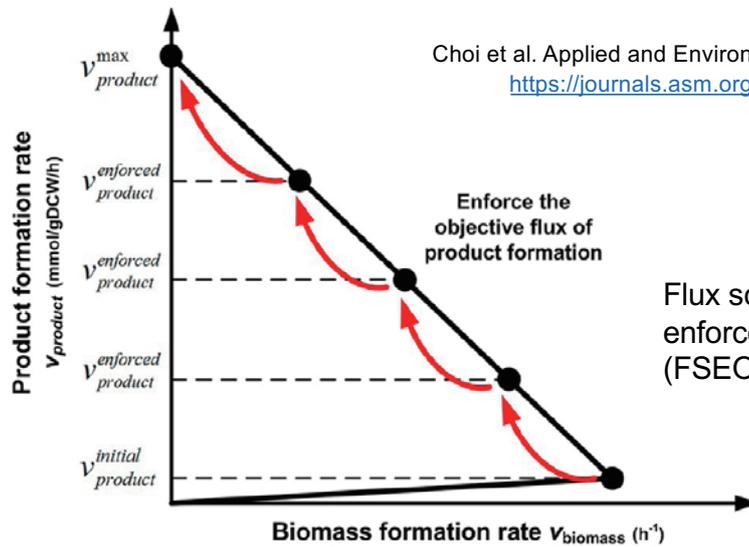
The following tutorials are based on Jupyter notebooks that are also available as live code at try.cameo.bio. Furthermore, course materials are available for a 2-day course in cell factory engineering.

- Import models
 - Import models from files
 - Import models from the internet
- Simulate models
 - Primer: Constraint-Based Modeling
 - Flux Balance Analysis
 - Parsimonious Flux Balance Analysis
 - Setp 2: Simulate knockouts phenotypes
- Analyzing models
 - Flux Variability Analysis
 - Phenotypic Phase Plane
 - Flux Balance Impact Degree
- Predict gene knockout strategies
 - OptGene
 - OptKnock
 - References
- Predict expression modulation targets
 - Flux Scanning based on Enforced Objective Flux
 - Differential flux variability analysis
- Predict heterologous pathways
- Easy strain design using a high-level interface
 - Searching for products
 - Host organisms
 - Computing strain engineering strategies

Documentation: <https://cameo.readthedocs.io/en/latest/>

Source: <https://github.com/biosustain/cameo>

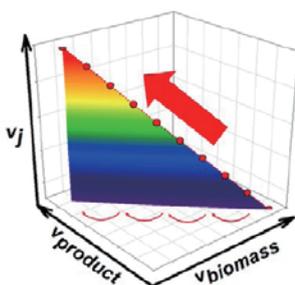
Example sources: <https://github.com/biosustain/cameo-notebooks>



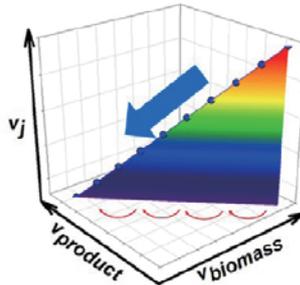
Choi et al. Applied and Environmental Microbiology (2010)
<https://journals.asm.org/doi/10.1128/aem.00115-10>

Flux scanning based on enforced objective flux (FSEOF)

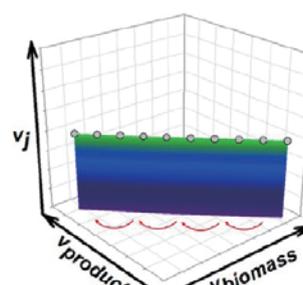
Increased flux (v_j)



Decreased flux (v_j)



Unchanged flux (v_j)



Tutorials

- Import models
- Simulate models
- Analyzing models
- Predict gene knockout strategies
- Predict expression modulation targets
- Predict heterologous pathways
- Easy strain design using a high-level interface

Contributing

Development

API Docs

Quick search


```
plotter = PlotlyPlotter()
```

<https://cameo.readthedocs.io/en/latest/06-predict-gene-modulation-targets.html#flux-scanning-based-on-enforced-objective-flux>

Flux Scanning based on Enforced Objective Flux

```
from cameo.strain_design.deterministic.flux_variability_based import FSEOF
```

```
fseof = FSEOF(model)
```

```
fseof.run(target=model.reactions.EX_succ_e)
```

Model: e_coli_core

Enforced objective: EX_succ_e

Primary objective: Maximize 1.0*BIOMASS_Ecoli_core_w_GAM - 1.0*BIOMASS_Ecoli_core_w_GAM_reverse_712e5

Reaction fluxes

	1	2	3	4	5	6	7	8	9	10
PFK	7.756941	8.036499	8.242515	8.361048	8.479582	8.598115	8.747793	9.393102	9.830787	9.910940
PGI	5.485630	6.110399	6.513173	6.651876	6.790579	6.929282	7.161108	8.874882	9.965126	9.981645
PGK	-16.431132	-16.838737	-17.173618	-17.421989	-17.670359	-17.918730	-18.198431	-18.976723	-19.547784	-19.761991
PGM	-15.238892	-15.761645	-16.212407	-16.577535	-16.942662	-17.307790	-17.704415	-18.602292	-19.293292	-19.628048
PPC	3.758320	5.012330	5.878875	6.284516	6.690157	7.095799	7.532996	8.472367	9.615323	11.108854
PPS	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.410338	1.527283
ADK1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.410338	1.527283
ENO	15.238892	15.761645	16.212407	16.577535	16.942662	17.307790	17.704415	18.602292	19.293292	19.628048
SUCc3	1.474575	2.949150	4.423725	5.898300	7.372875	8.847450	10.322025	11.796600	13.271175	14.745750
THD2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.119674	2.143848	2.916807	1.535161
TPI	7.756941	8.036499	8.242515	8.361048	8.479582	8.598115	8.747793	9.393102	9.830787	9.910940
FBA	7.756941	8.036499	8.242515	8.361048	8.479582	8.598115	8.747793	9.393102	9.830787	9.910940
FRD7	0.000000	0.000000	0.000000	1.453235	3.175302	4.897369	6.586703	7.755141	9.127842	10.852285
GAPD	16.431132	16.838737	17.173618	17.421989	17.670359	17.918730	18.198431	18.976723	19.547784	19.761991
ICL	0.000000	0.000000	0.386057	1.231342	2.076627	2.921912	3.735322	4.041459	4.143333	3.893465
MALS	0.000000	0.000000	0.386057	1.231342	2.076627	2.921912	3.735322	4.041459	4.143333	3.893465
PDH	8.809270	8.336008	8.179868	8.400957	8.622045	8.843134	9.064539	9.290984	9.107760	8.219084
EX_succ_e	1.474575	2.949150	4.423725	5.898300	7.372875	8.847450	10.322025	11.796600	13.271175	14.745750

REVIEW PAPER

Biofoundries: Bridging Automation and Biomanufacturing in Synthetic Biology

Dae-Hee Lee, Haseong Kim, Bong-Hyun Sung, Byung Kwan Cho, and Seung-Goo Lee

Received: 10 September 2023 / Revised: 27 October 2023 / Accepted: 31 October 2023
© The Korean Society for Biotechnology and Bioengineering and Springer 2023

Abstract Biofoundries represent advanced automation facilities pivotal for streamlining the Design-Build-Test-Learn (DBTL) paradigm in biomanufacturing and synthetic biology, suitable for both academic research and industrial applications. Nonetheless, establishing such a platform demands significant financial and temporal resources while maintaining a forward-looking perspective on automation, equipment compatibility, and operational efficiency. Despite its challenges, international collaborations between global biofoundries may offer solutions. The automated DBTL framework in biofoundries has transformed the production of bioproducts using engineered microbes. As the field advances, biofoundries are essential for streamlining and standardizing biotechnological processes, addressing efficiency, cost, and consistency challenges.

Keywords: biofoundry, DBTL cycle, synthetic biology, biomanufacturing, automation

<https://doi.org/10.1007/s12257-023-0226-x>

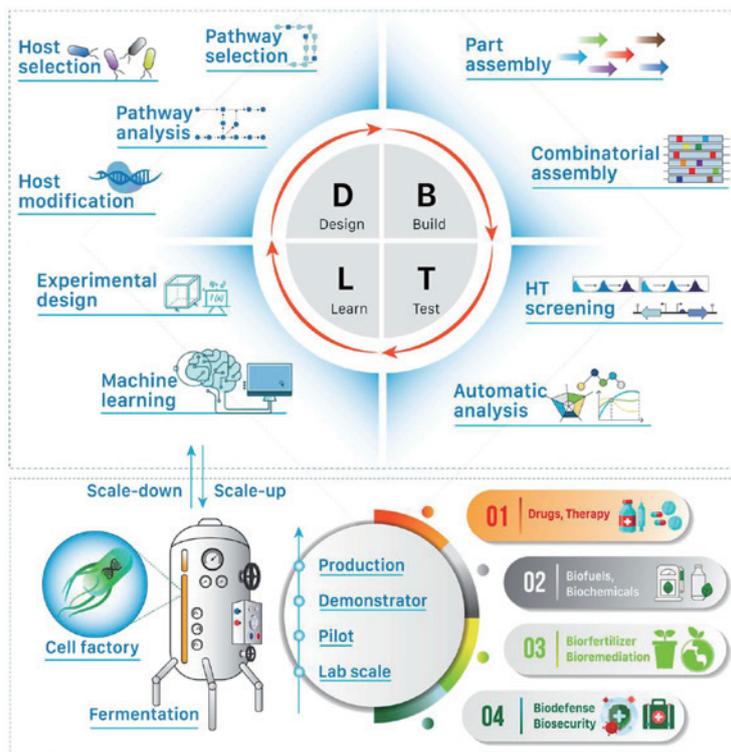


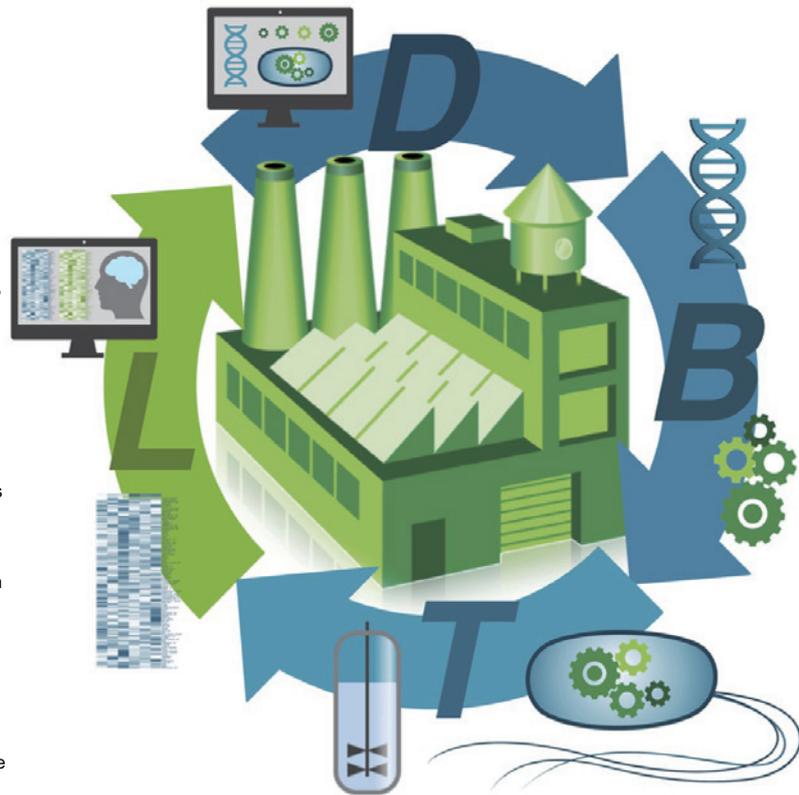
Fig. 1. The core of the biofoundry is established by the repetitive DBTL cycle. Through computational techniques, biofoundries create DNA segments and combine them, trying out and evaluating their function in live cells. Machine learning tools are then utilized to guide the design phase. By repeating the DBTL cycle, the goal is to produce genetic designs that meet the set criteria. After several rounds of this cycle, successful models progress from lab fermentation to commercial production through process development. DBTL: Design-Build-Test-Learn, HT: High Throughput.

Biofoundries are facilities that use advanced automation, synthetic biology tools, and computational tools to accelerate the design, construction, and testing of new biological systems [12,13]. These facilities are similar to traditional foundries used in metalworking, but instead of molds and furnaces, biofoundries use robotic systems and custom software to design and assemble biological components. Biofoundries provide a platform for researchers to rapidly prototype and test their synthetic biology designs, accelerating

the development process and enabling researchers to quickly iterate and optimize their designs [11]. By automating many of the time-consuming and error-prone aspects of biological research, biofoundries can dramatically increase the speed and efficiency of the Design-Build-Test-Learn (DBTL) cycle, ultimately leading to the faster development of new biotechnologies [9]. By using standardized protocols and automated systems in biofoundries, researchers can ensure that their experiments are consistent and reliable, which is essential for making meaningful comparisons between different systems [14]. This reproducibility also enables researchers to share their work more easily with others, which can accelerate the pace of scientific discovery and innovation [15]. Biofoundries are also highly flexible. Because they are designed to be modular and adaptable, they can be used for a wide variety of applications, from basic research to biomanufacturing. For example, biofoundry might be used to design and test new biological sensors for environmental monitoring, or it might be used to optimize the on-site and on-demand production of therapeutic proteins [16]. Finally, biofoundries are highly interdisciplinary [17]. Because they require expertise in biology, engineering, computer science, artificial intelligence, and other fields, they provide an excellent opportunity for researchers from different disciplines to collaborate and share knowledge. This interdisciplinary approach is essential for solving many of the complex challenges facing synthetic biology today, such as developing more sustainable and efficient biomanufacturing processes or designing new biosensors for detecting environmental contaminants and for disease diagnosis [10,18,19].

The Design-Build-Test-Learn (DBTL) biological engineering cycle.

In simple terms the DBTL framework aims to fulfill particular design criteria for a synthetic biology application, which might for example be the production of a specific product at an optimal titer or the detection of a specific clinical biomarker using an engineered gut microbiome. The cycle begins with **D (Design)**, which defines the desired target function/specifications and involves the computational design of genetic parts, circuits, regulatory and metabolic pathways to whole genomes; **B (Build)** involves the physical assembly of those designed genetic components; **T (Test)** involves the prototyping and testing of the assembled genetic designs in living cells (also called "chassis") at different scales, which also includes comprehensive analytical measurements ("omics") of specific cellular components. This can also include testing components in cell-free extract systems; **L (Learn)** is the application of modeling and computational learning tools, which uses the data obtained in T to inform the design process. Iterations of the DBTL cycle results in genetic designs that aim to fulfill the design specifications established in D. In the figure the DBTL cycle is depicted around an imagined biofactory or biorefinery where many products will be produced using more sustainable and circular economic processes forming the future infrastructure for a global bioeconomy. (Credit: Christopher Johnson, DOE Agile BioFoundry, Golden, CO, USA)



Hillson et al. Nature Communications (2019)

<https://www.nature.com/articles/s41467-019-10079-2>

Building a Biofoundry

- **Synthetic biology**, combining engineering and biology, enables the precise design of biological systems for applications in medicine, agriculture, and sustainable production.
- **Biofoundries, specialized facilities using automation and computational tools, accelerate the Design-Build-Test-Learn (DBTL) cycle, boosting research speed and reproducibility.** These adaptable, interdisciplinary hubs support innovations like environmental biosensors and on-demand therapeutic production. The collaborative nature of biofoundries leverages expertise across biology, engineering, and AI, pushing biotechnological progress and fostering global scientific advancement in synthetic biology and biomanufacturing.
- **Building a biofoundry involves integrating technical expertise and infrastructure** to support advanced synthetic biology research. Key components include:
 - **Robust infrastructure:** Equipped with specialized tools like DNA synthesis machines, high-throughput screening (HTS) instruments, and automated robotics.
 - **Expertise:** Teams skilled in synthetic biology and data analysis.
 - **Automated workflows:** For precision and efficiency in DNA synthesis and screening.
 - **Data management:** Robust systems for handling and analyzing large data volumes.
 - **Collaborative networks:** Partnerships to share resources, stay updated, and translate discoveries into practical applications.

Synthetic Biology for Building Biofoundries

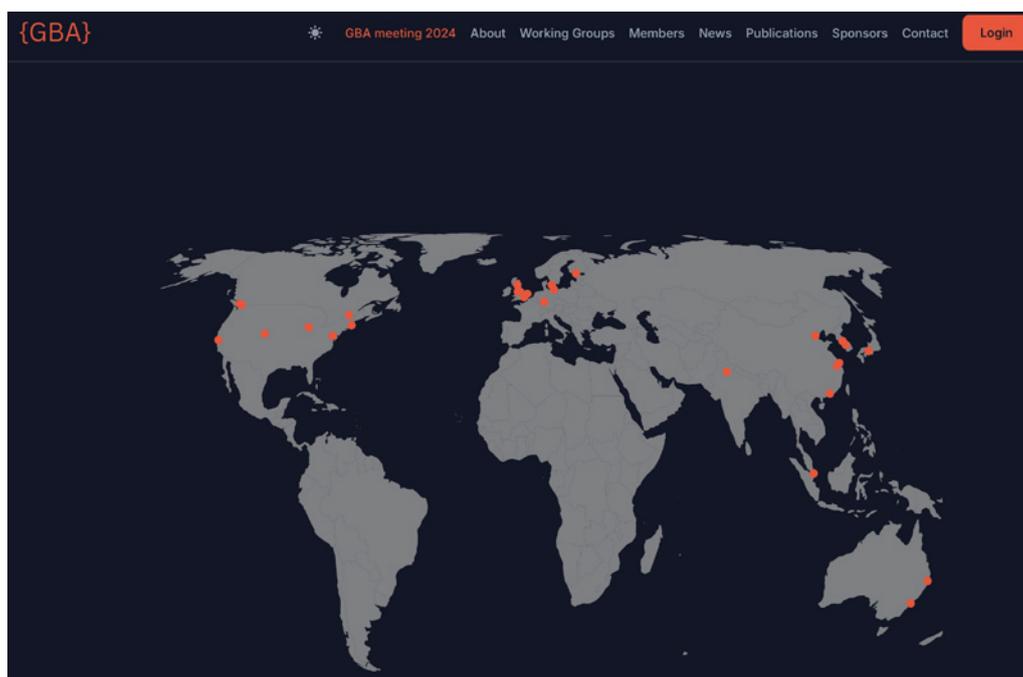
- The [DBTL \(Design-Build-Test-Learn\) cycle](#) is used in biofoundries to streamline the creation and optimization of biological systems:
 - [Design](#): Utilizes software like Cello for genetic circuit design, creating blueprints for synthetic constructs. KEGG and BRENDA databases help in selecting metabolic pathways and enzymes. For example, EnzymeMiner identifies uncharacterized enzymes for pathway optimization.
 - [Build](#): Uses methods like Gibson Assembly and Golden Gate Assembly for DNA synthesis. High-throughput platforms automate the creation of extensive DNA libraries.
 - [Test](#): Multi-omics, including next-generation sequencing (NGS) and Nanopore sequencing, assesses gene expression, metabolite levels, and protein function. For example, high-throughput screening identifies strains with maximum output.
 - [Learn](#): Employs machine learning for refining designs based on test results, allowing predictive adjustments for improved yields.
- Through these iterative cycles, biofoundries advance synthetic biology, optimizing processes for drugs, biofuels, and sustainable materials.

Global Biofoundry Alliance

- The [Global Biofoundry Alliance \(GBA\)](#) is an international network of public and private biofoundries aimed at promoting synthetic biology through shared expertise and resources. Established in 2018, it now includes 35 biofoundries worldwide.
- The GBA standardizes [data management, quality control, and experimental protocols to enhance reproducibility and collaboration](#). It has been pivotal in developing new technologies, such as COVID-19 diagnostics, and provides educational opportunities. The GBA recently restructured its working groups to focus on software/data, workflows/protocols, and sustainability.

Table 1. Current members of the GBA as of September 2023

No.	Name	Location	Website
1	Agile BioFoundry	USA, Emeryville	https://agilebiofoundry.org
2	Australian Genome Foundry	Australia, Sydney	https://www.mq.edu.au/research/research-centres-groups-and-facilities/innovative-technologies/facilities/australian-genome-foundry
3	Australian Institute for Bioengineering and Nanotechnology	Australia, Brisbane	https://aibn.uq.edu.au/welcome-biofoundry
4	BIOFAB, University of Washington	USA, Seattle	https://www.nano.uw.edu/uw-biofab-a-force-for-reproducible-science/
5	Biofactorial	Canada, Vancouver	https://biofactorial.microbiology.ubc.ca
6	The Biofoundry at UBC	Canada, Vancouver	https://biofoundry.sites.olt.ubc.ca
7	Biofoundry India	India, New Delhi	https://www.biofoundryindia.org
8	Colorado Cyberbiofoundry	USA, Colorado	http://cyberbiofoundry.org
9	CompuGene, TU Darmstadt	Germany, Darmstadt	https://www.compugene.tu-darmstadt.de/compugene/index.de.jsp
10	Concordia Genome Foundry	Canada, Montreal	https://genomefoundry.ca
11	CSIRO Biofoundry	Australia, Brisbane	https://www.csiro.au/biofoundry
12	DAMP lab, Boston University	USA, Boston	https://www.damplab.org
13	DTU Biosustain Biofoundry	Denmark, Lyngby	https://www.biosustain.dtu.dk/biofoundry
14	Earlham DNA Biofoundry	UK, Norwich	https://www.earlham.ac.uk/earlham-biofoundry
15	Edinburgh Genome Foundry	UK, Edinburgh	https://www.ed.ac.uk/biology/research/facilities/edinburgh-genome-foundry
16	GeneMill	UK, Liverpool	https://www.liverpool.ac.uk/health-and-life-sciences/research/liverpool-shared-research-facilities/multi-omics/genemill/
17	IBVT	Germany, Stuttgart	https://www.ibvt.uni-stuttgart.de/
18	iBioFAB-Illinois Biological Foundry for Advanced Biomanufacturing	USA, Illinois	https://www.igb.illinois.edu
19	K-Biofoundry (KAIST)	Korea, Daejeon	https://kbiofoundry.org
20	K-Biofoundry (KRIIBB)	Korea, Daejeon	https://kbiofoundry.org
21	Kobe Biofoundry	Japan, Kobe	https://www.kobe-u.ac.jp/research_at_kobe_en/index.html
22	LARA, Laboratory Automation Robotic Assistant Biochemistry Greifswald	Germany, Greifswald	https://lara.uni-greifswald.de
23	London Biofoundry, Imperial College London	UK, London	https://www.londonbiofoundry.org
24	Living Measurement Systems Foundry, NIST	USA, Maryland	https://www.nist.gov/programs-projects/nist-living-measurement-systems-foundry
25	Nucleo de Innovación de Sistemas Biológicos, NISB	Mexico, Monterrey	https://nisb.mx/index.html



<https://www.biofoundries.org>

Working Groups

To facilitate the growth of BioFoundries around the globe, we have established the following working groups.



Software

Working Group Co-chairs

- Nathan Hillson (Americas)
- Chueh Loo Poh (Asia-Australia)
- Lars Nielsen (Europe)



Business Model, Sustainability and Value Proposition

Working Group Co-chairs

- Benjamin Scott (Americas)
- Allison Rice (Asia-Australia)



Outreach, Industry, Policy, Security and Public

Working Group Chair

- To be decided



Metrology, Reproducibility and Data Quality

Working Group Chair

- To be decided



Grand Challenge and Demonstration Projects

Working Group Chair

- To be decided



K-Biofoundry (KAIST)

South Korea, Daejeon



K-Biofoundry (KRIBB)

South Korea, Daejeon



Kobe Biofoundry

Japan, Kobe



Lara, Laboratory Automation Robotic Assistant Biochemistry Greifswald

Germany, Greifswald



London Biofoundry, Imperial College London

UK, London



Living Measurement Systems Foundry, NIST

USA, Maryland



Nucleo de Innovación de Sistemas Biológicos, NISB

Mexico, Monterrey



SIAT Biofoundry

China, Shenzhen



SJTU Synbio Biofoundry

China, Shanghai



SKy Biofoundry, Sungkyunkwan University

South Korea, Seoul



Global Biofoundry Alliance Meeting 2024 (GBA2024)

October 10(Thu) - 12(Sat), 2024
Daejeon Convention Center(DCC), Republic of Korea

<https://www.gba2024.org>

GBA2024 Program Day 1 (10 October 2024)	
08:00 – 09:00 (60 min)	Registration and Refreshments
09:00 – 09:20 (20 min)	Welcoming Remarks Chair: Byung-Kwan Cho (K-biofoundry, Korea) Sang Yub Lee (K-biofoundry, Korea) Paul Freemont (London Biofoundry, UK) Lars Nielsen (DTU Biosustain Biofoundry, Denmark) Nathan Hillson (Aglie Biofoundry, USA) Chueh Loo Poh (NUS Biofoundry, Singapore)
09:20 – 10:20 (30 min each 60 min total)	Session 1 Invited Talks (2 slots) Session Chair: Byung-Kwan Cho (K-biofoundry, Korea) (09:20 – 09:50) Invited Talk 1: Michael Kopke (LanzaTech, USA) (09:50 – 10:20) Invited Talk 2: Joon Young Jung (CJ, Korea)
10:20 – 10:40 (20 min)	Coffee Break for Q&A Networking
10:40 – 13:00 (20 min each 140 min total)	Session 2 GBA Members I (7 slots) Session Chair: Dong-Myung Kim (K-biofoundry, Korea) (10:40 – 11:00) London Biofoundry (Paul Freemont, UK) (11:00 – 11:20) DAMP Lab (Courtney Tretheway, USA) (11:20 – 11:40) K-biofoundry (Seung-Goo Lee, Korea) (11:40 – 12:00) RISE High-throughput Centre (Peyam Ghiasi, Sweden) (12:00 – 12:20) Tianjin Biofoundry (Meng Wang, China) (12:20 – 12:40) IDEA Bio (Helen Wong, Australia) (12:40 – 13:00) NUS Biofoundry (Chueh Loo Poh, Singapore)
13:00 – 14:00 (60 min)	Lunch for Discussion & Networking
14:00 – 16:00 (20 min each 120 min total)	Session 3 GBA Members II (6 slots) Session Chair: Sung Sun Yim (K-biofoundry, Korea) (14:00 – 14:20) Australian Genome Foundry (Darren Plett, Australia) (14:20 – 14:40) Shenzhen Biofoundry (Tong Si, China) (14:40 – 15:00) Synthetic DNA Accelerator Lab (Konrad Herbst, Germany) (15:00 – 15:20) Aglie Biofoundry (Nathan Hillson, USA) (15:20 – 15:40) UC San Diego (Adam Feist, USA) (15:40 – 16:00) Kobe Biofoundry (Tomohisa Hasunuma, Japan)
16:00 – 16:20 (20 min)	Coffee Break & Group Photo
16:20 – 18:00 (25 min each 100 min total)	Session 4 Special Talks for Biofoundry Ecosystem (4 slots) Session Chair: Minkyu Oh (K-biofoundry, Korea) (16:20 – 16:45) GenoFAB software (Jean Peccoud, USA) (16:45 – 17:10) BioNet (Marc Sallit, USA) (17:10 – 17:35) Nucleic acid synthesis screening (Piers Millett, IBBIS) (17:35 – 18:00) Standardization (Paul Freemont, UK)
18:00 – 18:30 (30 min)	Refreshment for Discussion
18:30 – 21:00 (150 min)	Banquet on the Terrace supported by KSBA (Invitation only)

GBA2024 Program Day 2 (11 October 2024)	
08:00 – 08:45 (45 min)	Registration and Refreshments
08:45 – 10:15 (30 min each 90 min total)	Session 5 Invited Talks (3 slots) Session Chair: Bong Hyun Sung (K-biofoundry, Korea) (08:45 – 09:15) Invited Talk 3: Emily LeProust (Twist Biosciences, USA) (09:15 – 09:45) Invited Talk 4: Akihiko Komdo (Bacchus Bio Innovation, Japan) (09:45 – 10:15) Invited Talk 5: Massimo Merighi (Lesaffre, USA)
10:15 – 10:40 (25 min)	Coffee Break for Q&A Networking
10:40 – 13:00 (20 min each 140 min total)	Session 6 GBA Members III (7 slots) Session Chair: Daehee Lee (K-biofoundry, Korea) (10:40 – 11:00) Concordia Genome Foundry (Angela Quach, Canada) (11:00 – 11:20) DTU Biosustain Biofoundry (Jochen Forster, Denmark) (11:20 – 11:40) Estonian Biofoundry (Vallo Varik, Estonia) (11:40 – 12:00) Sparrow Biofoundry (Chen Xixian, Singapore) (12:00 – 12:20) VTT Biofoundry (Jussi Jantti, Finland) (12:20 – 12:40) Paris Biofoundry (Stephane Lemaire, France) (12:40 – 13:00) IBVT Biofoundry (Lorena Hagele, Germany)
13:00 – 14:00 (60 min)	Lunch for Discussion & Networking Luncheon Workshop: Screening Nucleic Acids (IBBIS)
14:00 – 15:30 (15 min each 90 min total)	Session 7 New Equipment and Applications (6 slots) Session Chair: Haseong Kim (K-biofoundry, Korea) (14:00 – 14:15) Twist Bioscience (Mark Dupal) (14:15 – 14:30) Bioneer (Namshin Kim) (14:30 – 14:45) Beckman Coulter Life Sciences (Maria Savino) (14:45 – 15:00) Thermo Fisher Scientific (Hildegard Tekotte) (15:00 – 15:15) Revvity (Jonathan Cechetto) (15:15 – 15:30) Tecan (Matthias Dzung)
15:30 – 16:00 (30 min)	Coffee Break for Q&A Networking (sponsored by MITRE)
16:00 – 17:40 (20 min each 100 min total)	Session 8 GBA Members IV (5 slots) Session Chair: Seung Goo Lee (K-biofoundry, Korea) (16:00 – 16:20) Sky Biofoundry (Han Min Woo, Korea) (16:20 – 16:40) CNPEM (Eduardo de Couto e Silva, Brazil) (16:40 – 17:00) UIUC Biofoundry (Yong-Su Jin, USA) (17:00 – 17:20) LARA Greifswald (Mark Doer, Germany) (17:20 – 17:40) TII Biofoundry (Nidia Hernandez, UAE)
17:40 – 18:00 (20 min)	Wrap-up, GBA announcement & Closing Remarks Lars Nielsen (GBA President)
18:00 – 19:00 (60 min)	Closed Session GBA Working Group Discussion (Directors only) Session Chair: Nathan Hillson, Chueh Loo Poh, Lars Nielsen
19:00 – 21:00 (120 min)	GBA director's dinner supported by KAIST, KRIBB, CNU (Invitation only)



Advancing biofoundry development: strategies and challenges

Aporva Gupta^{1,2} · Seung-Goo Lee^{1,2,3} · Bong-Hyun Sung^{1,2,3} · Dae-Hee Lee^{1,2,3} · Byung-Kwan Cho^{3,4} · Dong-Myung Kim⁵ · Haseong Kim^{1,2,3}

Received: 20 March 2024 / Revised: 12 June 2024 / Accepted: 17 June 2024 / Published online: 2 July 2024
© The Author(s), under exclusive licence to The Korean Society for Biotechnology and Bioengineering and Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

This study explores the development of biofoundries, emphasizing the integration of synthetic biology with artificial intelligence (AI) and robotics. It outlines critical challenges such as the necessity for interdisciplinary collaboration and the development of hardware, software, and AI specific to biofoundry operations. To address these challenges, we recommend strategies like rapid prototyping, soft integration, and the strategic implementation of AI. We underscore the vital role of synthetic biology researchers in advancing biofoundry capabilities and advocate for collaborative, multidisciplinary efforts to optimize the development and functionality of biofoundries.

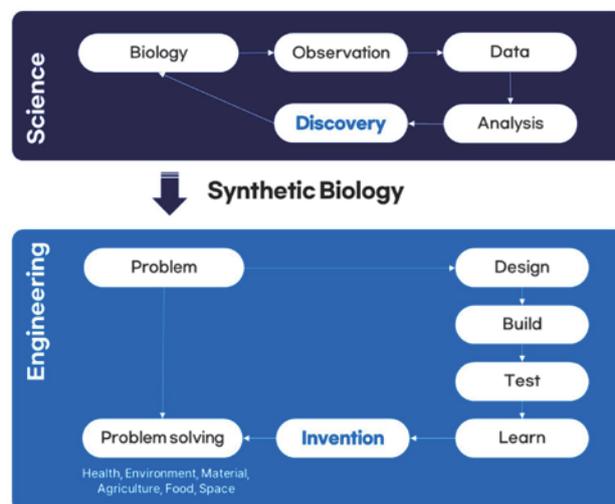
Keywords Biofoundry · Synthetic biology · Automation · Software

<https://doi.org/10.1007/s12257-024-00121-6>

Emergence of synthetic biology as a tool for invention

- The integration of next-generation sequencing since 1996 has greatly enhanced the ability to analyze genomic and gene expression profiles, supported by bioinformatics tools. Synthetic biology adopts a bottom-up approach, using strategic DNA assembly to regulate gene expression for optimizing metabolic pathways. The iGEM competition, launched in 2004, has significantly promoted synthetic biology by engaging global participants in solving challenges and fostering innovation. It leverages the Registry of Standard Biological Parts, with over 20,000 documented parts, catalyzing the growth of startups and reinforcing collaboration between academia and industry.

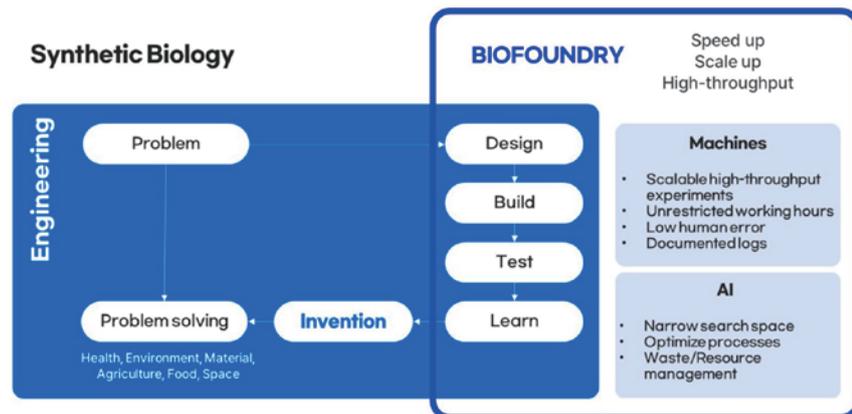
Fig. 1 The paradigm shift in biological research from discovery to invention through the emergence of synthetic biology. The upper half of the diagram delineates the traditional approach to biological research, which is driven by observation leading to data collection. These data are then analyzed for discovery, signifying a key milestone in the scientific process. The lower half of the diagram depicts to the engineering approach driven by synthetic biology. It begins with the identification of a specific problem and progresses through a cyclic process of design, build, test, and learn, resulting in innovative inventions. This shift highlights the evolution of biological research toward a more methodical, problem-solving framework characteristic of engineering disciplines within the scope of synthetic biology



Acceleration of synthetic biology research using biofoundry

- The [Design-Build-Test-Learn \(DBTL\) cycle](#) in synthetic biology enhances research reproducibility and efficiency, though the Build and Test phases can be labor-intensive. Automation has reduced these challenges, exemplified by innovations like robotic scientists Adam and Eve and biofoundries such as [Ginkgo Bioworks](#) and [Amyris](#), which have significantly boosted throughput and productivity. [Lanzatech](#) showcased industrial breakthroughs using automated systems to produce ethanol and improve yields for compounds like 1-hexanol. [AI integration further optimizes the Design and Learn phases](#), reducing costs and narrowing design space for faster market delivery.

Fig. 2 Biofoundry acceleration of the Design-Build-Test-Learn (DBTL) cycle in synthetic biology. It highlights the role of biofoundries in enhancing the efficiency of the synthetic biology DBTL cycle. Biofoundries primarily enable faster, larger-scale, and more efficient throughput in synthetic biology research, thereby accelerating the transition from problem-solving to invention. AI: artificial intelligence



Considerations required for Biofoundry construction

- **Hardware:**
 - Biofoundries are associated with automation, using robotic systems to advance the DBTL cycle in synthetic biology. However, [fully automated biofoundries with advanced machinery like robotic arms and automated devices come with high costs and planning challenges](#). Semi-automated DBTL strategies offer more cost-effective flexibility. The limited availability of such equipment highlights the need for continuous investment and policy support.
 - Importantly, [data management and analysis must keep pace with hardware](#) to avoid underutilization, emphasizing that effective workflows and software integration are crucial for maximizing biofoundry performance.
- **Workflows:**
 - In biofoundry development, two key terms are essential: [unit process \(the smallest automated protocol\)](#) and [workflow \(a series of unit processes aimed at a specific outcome\)](#). Workflows, derived from optimized manual protocols, must detail reagent handling and positioning for automated operations.
 - [Efficiency can be enhanced through sample consolidation and integrated scheduling of varied workflows](#), crucial for high-throughput demands. Advanced scheduling software supports this integration. Biofoundries reduce waste and errors, addressing a major issue where over 50% of research remains unpublished, partly due to preventable flaws.

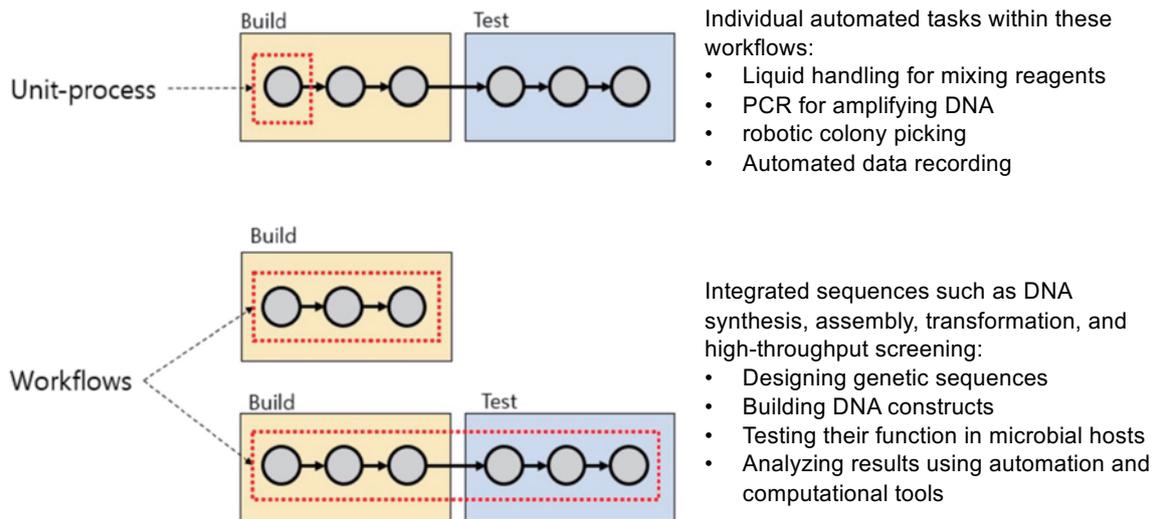


Fig. 3 Workflow and unit-process in biofoundry operations. It shows the concept of 'unit-processes' and 'workflows' within the context of biofoundry operations. The gray circles represent discrete unit-processes, which are the fundamental operational steps within the biofoundry. A workflow is depicted as an assembly of multiple unit-processes, which may occur within a single phase of the DBTL cycle or span across the entire cycle. DBTL: Design-Build-Test-Learn

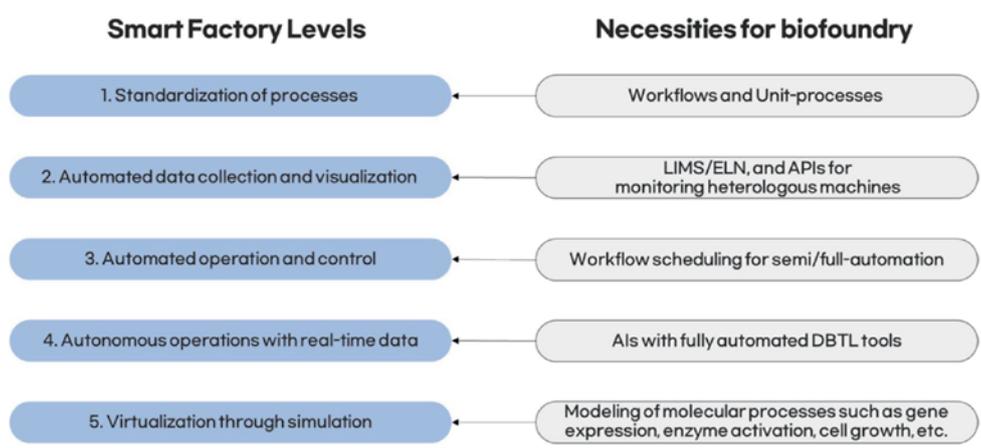


Fig. 5 Smart factory levels and related necessities for biofoundry. The comparison of the developmental stages of a smart factory to corresponding features within a biofoundry. The left column lists the five levels of a smart factory while the right column aligns these levels with analogous biofoundry components. This comparison underscores how biofoundries embody the principles of smart manufacturing to streamline and enhance biological research and development. *LIMS* Laboratory Information Management Systems, *ELN* Electronic Laboratory Notebooks, *DBTL* Design-Build-Test-Learn

Considerations required for Biofoundry construction

- Software & AI:
 - Developing robust software for biofoundry operations is critical but challenging due to unique requirements and the need for extensive integration. Current solutions, like Electronic Laboratory Notebooks, often fall short in handling the complex and large-scale operations of biofoundries. The 'rapid prototyping and soft integration' approach promotes quick creation and merging of essential functionalities using frameworks like Python's Streamlit or R's Shiny, with IDEs such as Visual Studio Code and RStudio for cycle management. Effective software must handle high-throughput data, monitor equipment and materials, and coordinate various automated devices via APIs like JSON for seamless communication. This enhances operational adaptability and interoperability, vital for the scalability and efficiency of biofoundry systems.
 - Biofoundries streamline the Design-Build-Test-Learn (DBTL) cycle in synthetic biology, but handling the vast complexity of biological systems requires precise library design supported by AI. Pre-trained models like ProtBERT and RFdiffusion help in transforming sequences into biologically relevant vectors and creating targeted libraries, significantly reducing experimental search space. Self-driving laboratories, such as the one demonstrated at the University of Wisconsin–Madison for protein engineering, use AI-guided systems to expedite DBTL cycles, enhancing thermostability of enzymes faster than traditional methods. Predictive maintenance using real-time data and machine logs reduces downtime, supporting efficient, reliable biofoundry operations. Large Language Models (LLMs) also contribute by assisting in training, optimizing protocols, and gathering preliminary data, integrating human expertise with advanced automation for enhanced productivity.

Examples of high-throughput tools and AI tools

- Recent development of methods for measuring/predicting enzyme kinetics
- DeepEC for predicting EC numbers
- AlphaFold for predicting highly accurate protein structures

RESEARCH ARTICLE

ENZYMOLOGY

Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics

C. J. Markin^{1†}, D. A. Mokhtari^{1†}, F. Sunden¹, M. J. Appel¹, E. Akiva², S. A. Longwell³, C. Sabatti^{4,5}, D. Herschlag^{1,6,7*}, P. M. Fordyce^{3,7,8,9*}

Systematic and extensive investigation of enzymes is needed to understand their extraordinary efficiency and meet current challenges in medicine and engineering. We present HT-MEK (High-Throughput Microfluidic Enzyme Kinetics), a microfluidic platform for high-throughput expression, purification, and characterization of more than 1500 enzyme variants per experiment. For 1036 mutants of the alkaline phosphatase PafA (phosphate-irrepressible alkaline phosphatase of *Flavobacterium*), we performed more than 670,000 reactions and determined more than 5000 kinetic and physical constants for multiple substrates and inhibitors. We uncovered extensive kinetic partitioning to a misfolded state and isolated catalytic effects, revealing spatially contiguous regions of residues linked to particular aspects of function. Regions included active-site proximal residues but extended to the enzyme surface, providing a map of underlying architecture not possible to derive from existing approaches. HT-MEK has applications that range from understanding molecular mechanisms to medicine, engineering, and design.

<https://doi.org/10.1126/science.abf8761>

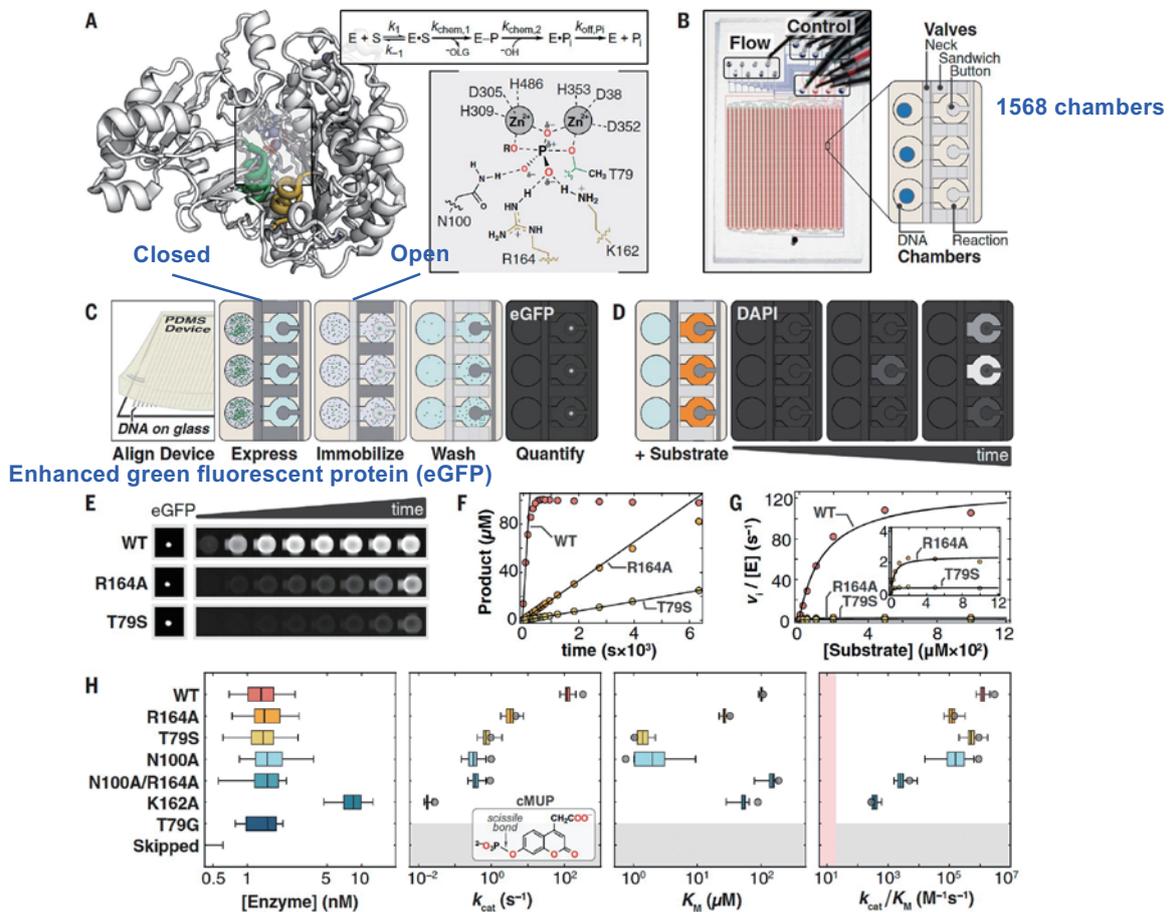
Enzyme kinetics

- Unit
 - The amount of enzyme that gives a predetermined amount of catalytic activity under specific conditions.
 - 1 unit: 1 μmol product per minute at a specified pH and temperature with a substrate concentration much greater than the value of K_m .
- Specific activity
 - The number of units of activity per amount of total protein.
- K_m (mol l^{-1}):
 - The substrate concentration value in which the substrate concentration is reaching halfway of the maximum reaction velocity.
 - A low K_m indicates a higher affinity for the substrate.
- k_{cat} (s^{-1}): turnover number
 - The maximum number of enzymatic reactions catalyzed per second.
- V_{max}
 - The enzyme's maximum catalytic rate.

BRENDA is the main collection of enzyme functional data available to the scientific community.

The screenshot shows the BRENDA website interface. At the top, it says "Leibniz Institute DSMZ" and "DSMZ Digital Diversity". The main header features the BRENDA logo and a search bar labeled "BRENDA Quicksearch". Below the header, there are navigation links: "Classic view", "All enzymes", "Enzyme history", and "BRENDA support". A search form is present with a "Search term" input field and a "Search BRENDA" button. Below the search form, there are dropdown menus for "Enzyme, Ligand" and "contains", along with "add search field" and "delete search field" buttons. The main content area is divided into several sections: "Text-based queries" (Full-text Search, Advanced Search, DRENDA, MeSH Ontology), "Structure-based queries" (Ligand Structure Search, Metabolic Pathways, Enzyme Structures), "Explorer" (Enzyme Classification, TaxTree, Protein folding: CATH, SCOPe, Ontologies), "Visualization" (Enzyme Word Maps, Organism Word Maps, Genomes, Functional Parameter Statistics, Metabolic Pathways), "Prediction" (Membrane Helices, Localization Prediction, EnzymeDetector), "Supporting & External" (BRENDA, Tissue Ontology, Biochemical Reactions, MetaboMAPS), and "News" (NEW Release online! - February 1, 2023, Release 2023.1 is online including 68 new and 479 updated enzyme classes).

<https://www.brenda-enzymes.org>



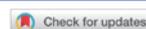
Overview of HT-MEK and PafA

- HT-MEK (High-Throughput Microfluidic Enzyme Kinetics) is a platform capable of simultaneously expressing, purifying, and characterizing more than 1500 rationally chosen enzyme mutants in parallel with the depth and precision of traditional site-directed mutagenesis.
- HT-MEK was applied to PafA (phosphate-irrepressible alkaline phosphatase of *Flavobacterium*). PafA and related phosphomonoesterases show the rate enhancements of up to $\sim 10^{27}$ -fold, providing a large dynamic range to explore. PafA, a secreted enzyme, was also considered to be highly stable, potentially allowing deeper analysis of the probe catalysis without obfuscation from global unfolding.
- To obtain catalytic rate parameters, HT-MEK quantifies (i) the concentration of immobilized enzyme in each chamber, using an eGFP calibration curve, and (ii) the amount of product formed as a function of reaction time, using a chamber-specific product calibration curve. Reaction progress curves are then fit in each chamber to obtain initial rates (v_i) for each substrate concentration using a custom image-processing pipeline, and observed rates (v_i) are converted to enzyme-normalized rate constants according to the eGFP intensity in each chamber and calibration curve (Fig. 1, E and F). This process-repeated on a single device for multiple substrate concentrations, multiple substrates, and multiple inhibitors-provides the data necessary to obtain Michaelis-Menten parameters and other kinetic and thermodynamic constants.
- Further information available in the section “HT-MEK device and experimental pipeline” of the paper.

ARTICLES

<https://doi.org/10.1038/s41929-022-00798-z>

nature
catalysis



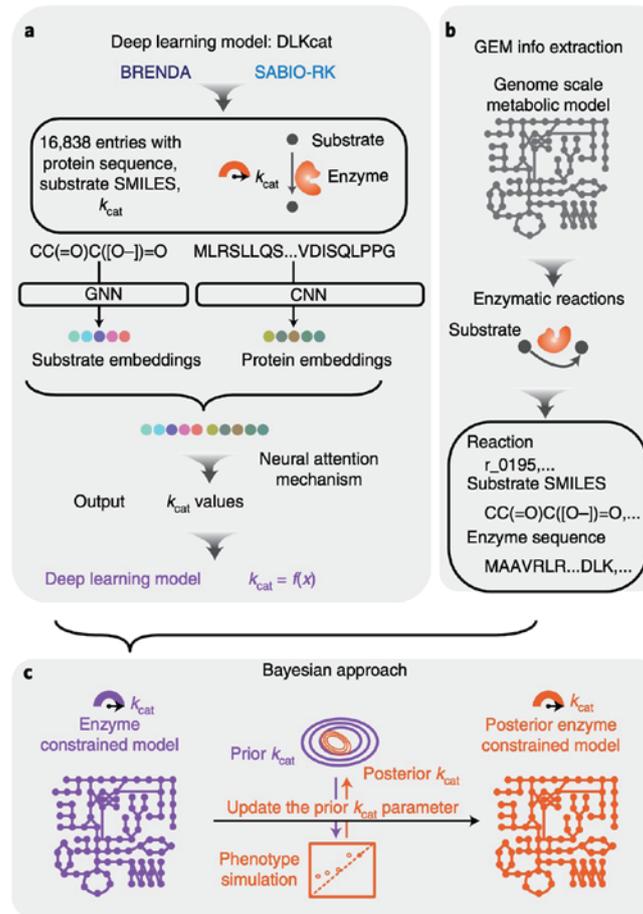
OPEN

Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction

Feiran Li^{1,4}, Le Yuan^{1,2,4}, Hongzhong Lu¹, Gang Li¹, Yu Chen¹, Martin K. M. Engqvist¹,
Eduard J. Kerkhoven^{1,2} and Jens Nielsen^{1,3}

Enzyme turnover numbers (k_{cat}) are key to understanding cellular metabolism, proteome allocation and physiological diversity, but experimentally measured k_{cat} data are sparse and noisy. Here we provide a deep learning approach (DLKcat) for high-throughput k_{cat} prediction for metabolic enzymes from any organism merely from substrate structures and protein sequences. DLKcat can capture k_{cat} changes for mutated enzymes and identify amino acid residues with a strong impact on k_{cat} values. We applied this approach to predict genome-scale k_{cat} values for more than 300 yeast species. Additionally, we designed a Bayesian pipeline to parameterize enzyme-constrained genome-scale metabolic models from predicted k_{cat} values. The resulting models outperformed the corresponding original enzyme-constrained genome-scale metabolic models from previous pipelines in predicting phenotypes and proteomes, and enabled us to explain phenotypic differences. DLKcat and the enzyme-constrained genome-scale metabolic model construction pipeline are valuable tools to uncover global trends of enzyme kinetics and physiological diversity, and to further elucidate cellular metabolism on a large scale.

<https://doi.org/10.1038/s41929-022-00798-z>



Development of DLKcat

- DLKcat is a deep learning approach that uses **substrate structures and protein sequences as inputs**, and demonstrated its capability for the large-scale **prediction of k_{cat} values** for various organisms, as well as for identifying key amino acid residues that affect these predictions.
- Even for those well-studied organisms, the k_{cat} coverage is far from complete. In a *S. cerevisiae* ecGEM (don't worry about this term), **only 5% of all enzymatic reactions** have fully matched k_{cat} values in BRENDA.
- The deep learning approach DLKcat was developed by combining a graph neural network (GNN) for substrates and a convolutional neural network (CNN) for proteins.
- A comprehensive dataset was generated from the BRENDA and SABIO-RK databases to train the neural network. Incomplete database entries with missing information and redundant entries were filtered out to ensure a dataset of unique entries with substrate name, substrate SMILES information, EC number, protein sequence, organism name and k_{cat} value. The final dataset contained 16,838 unique entries catalysed by 7,822 unique protein sequences from 851 organisms and converting 2,672 unique substrates.

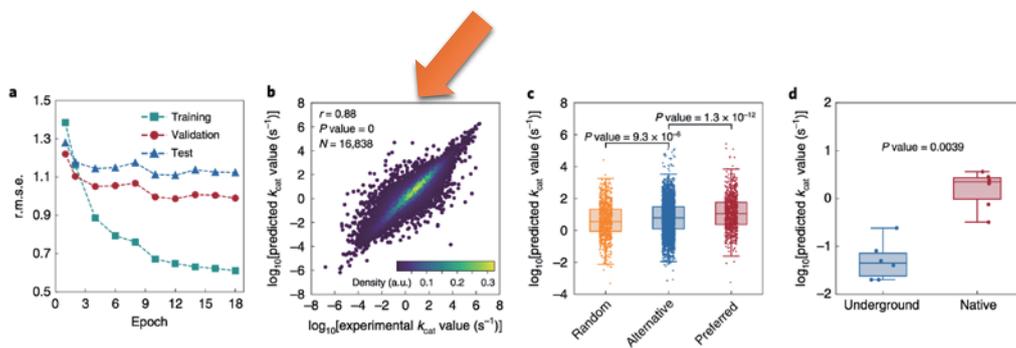
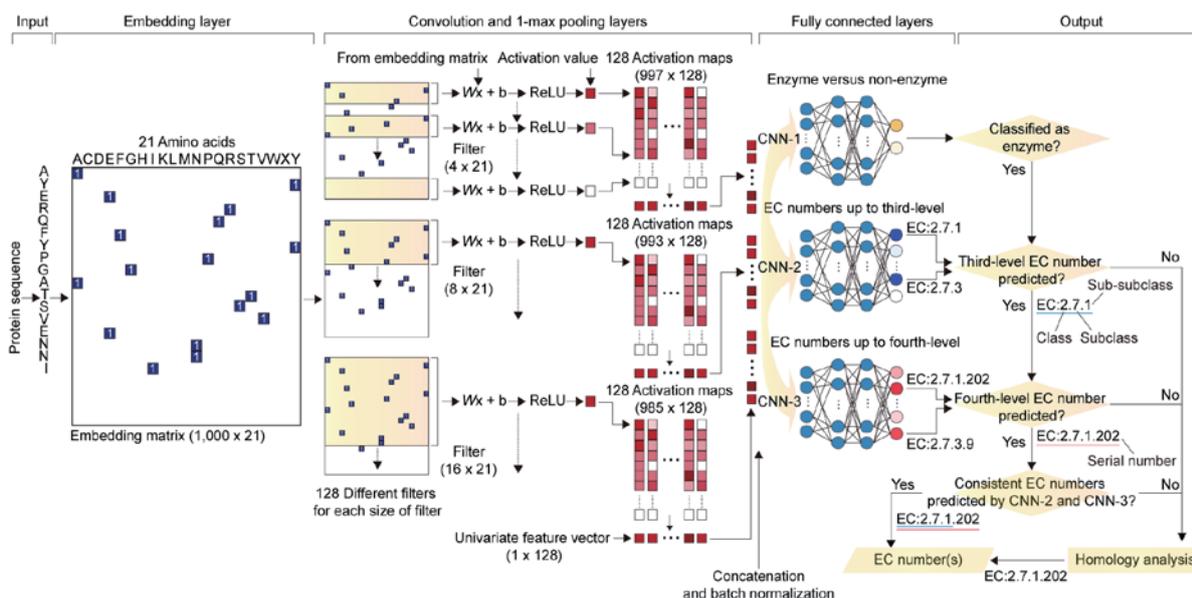


Fig. 2 | Deep learning model performance for k_{cat} prediction. **a**, The r.m.s.e. of k_{cat} prediction during the training process. **b**, Performance of the final deep learning model. The correlation between predicted k_{cat} values and those present in the whole dataset (training, validation and test datasets) was evaluated. The brightness of colour represents the density of data points. Student's *t*-test was used to calculate the *P* value for Pearson's correlation. **c**, Enzyme promiscuity analysis on the whole dataset. For enzymes with multiple substrates, we divided the substrates into preferred and alternative by their experimentally measured k_{cat} value, and then used the predicted k_{cat} values for this box plot. Random substrates were randomly chosen from the compound dataset in our training data, except for the documented substrates and products for the tested enzyme. We evaluated 945 promiscuous enzymes in the whole dataset ($n = 945$ for preferred substrates, $n = 4,238$ for alternative substrates, $n = 945$ for random substrates). **d**, Comparison of the predicted k_{cat} values for the native substrates and the underground substrates with the human aldo-keto reductase enzyme as a case study. Here, we defined those substrates with the top 10% catalytic ability (experimental k_{cat} value) as the native substrates ($n = 6$), while those with the last 10% catalytic ability (experimental k_{cat} value) were considered as the underground substrates as defined in the reference ($n = 6$)³¹. In each box plot (**c** and **d**), the central band represents the median value, the box represents the upper and lower quartiles and the whiskers extend up to 1.5 times the interquartile range beyond the box range. A two-sided Wilcoxon rank sum test was used to calculate the *P* values in **c** and **d**.

DeepEC: Use of three convolutional neural networks and homology analysis to predict EC numbers for a protein sequence



Source code: <https://bitbucket.org/kaistsystemsbiology/deepec>
<https://doi.org/10.1073/pnas.1821905116>

```

gene      393498..394961
         /locus_tag="SC00379"
         /gene_synonym="katA"
         /db_xref="GeneID:1095802"
CDS       393498..394961
         /locus_tag="SC00379"
         /gene_synonym="katA"
         /EC_number="1.11.1.6"
         /note="SCF62.05, katA, catalase (EC 1.11.1.6) len: 487 aa.
         Has been previously sequenced and characterised from
         Streptomyces coelicolor strain ATCC10147 TR:P77948 (EMBL;
         X96981) catalase (EC 1.11.1.6) (488 aa), fasta scores opt:
         3300 z-score: 3768.6 E():0 98.2% identity in 488 aa
         overlap. Also similar to another proposed Streptomyces
         coelicolor catalase (EC 1.11.1.6) TR:Q9Z598 (EMBL;
         AL035478) SC2G5.25C (487 aa), fasta scores opt: 1704
         z-score: 1896.1 E():0 53.8% identity in 483 aa overlap.
         Contains Prosite hits to PS00437 Catalase proximal
         heme-ligand signature and to PS00438 Catalase proximal
         active site signature. Also contains a Pfam match to entry
         PF00199 catalase, Catalase."
         /codon_start=1
         /transl_table=11
         /product="catalase"
         /protein_id="NP_624701.1"
         /db_xref="GI:21218922"
         /db_xref="GeneID:1095802"
         /translation="MPENNQKPLTTVAGAPVPDQNSLTSGPRGPMLLQDWWFLEKLA
         HFDREVIPERRMHAHGSGAGFTFTVTHDITRYTSAKIFSEIGKKTPLFTRFSTVAGER
         GAADAERDIRGFVAVKFYTDEGNWDLVGNNTPVFFFRDPLKFPDLNHAVKRDRPRTLNRN
         AENNWDFTWNLPEALHQVTIVMSDRGIPASYRHHMGGFSHTYSLINAAGERFWVKFHH
         RTQQGIGNLTDAAEALVGKDRESHQRDLFDAIEDGDFPKWKLFIQVMPEADAENYRF
         HPFDLTKVWSKKDYPLIEVGEWELNRPNDNYFADVEQAAFSPANVVPGISFSPDRMLQ
         GRLFSGYDAQRYRLGVNHHQIPVNAKPNVNSYHRDGMARVDGNQATPGVPEPNSYGR
         WQEQPAYRDPAAQAVGAVADRFNyreDDNYFEQPGNLFQMSPEQQQVLFENTARAID
         GASAQTIERHIGNCTQADPAYGAGVRKAIEALAAAGL"

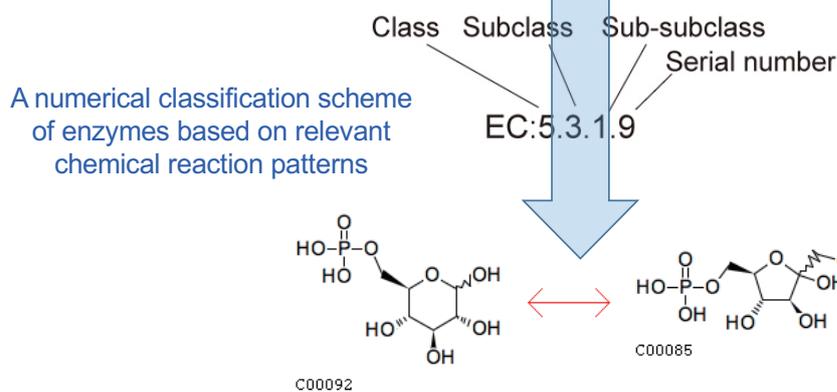
```

Enzyme Commission (EC) number

```

A5A9C3.1 14 5IKLWPPSQTIRLLVERMANNLST..PSIFSRK..YGLSKEEAARENAKQIEEVACSTANQ.....HYEKEPDGDDGSSAVQLYAKECSKLILEVELK 101
B4F917.1 13 5IKLWPPSESTRIMLDRMTNNLST..ESIFSRK..YRLGKDEAHENAKTIEELCFALADE.....HFREEPDGDGSSAVQLYAKETSKMMLLEVELK 100
A9S1V2.1 23 VFKLWPPSQGTREAVRQKMKALKSS..ACFESQS..FARIELADAQEHARAIIEEVAFGAQE.....ADSGGDKTGSVAVVMYAKHASKLMLLETLR 109
B9G5N7.1 13 5VKLWPPGQSTRMLVERMTKNFIT..PSFISRK..YGLLSKEEAEDAKKIEEVAFANAQ.....HYEKQPDGDDGSSAVQLYAKESRRLMLEVLK 100
Q8H056.1 30 5FSIWPPPTQRTDRAVVRRLVDTLGG..DTILSKR..YGAVPAADAEPAARGIEAEAFDAAA..SGEAAATASVEEGIKALQLYSKEVSRRLDFVK 120
Q0D4Z3.2 44 5LSIWPPSQRTDRAVVRRLVAVL..PSILSKR..YGAVPEAEAGRAAAVEAEAYAVTES..SSAAAAPASVEDGIEVLQAYSKEVSRRLLELAK 135
B9MVM8.1 56 5FSIWPPPTQRTDRAIISRLIETLST..TSVLSKR..YGTIPKEEAESARRIEEAFSGAST.....VASSEKDGLEVLQLYSKEISKRMLLETVK 141
Q0IYC5.1 29 5FAVWPPTRRTDRAVVRRLVAVL..SGDITLALRKYRYGAVPAADAEARAARAVEAQAFAAASA.....SSSSSSVEDGIEVLQYSREYSNRLAFVR 121
A9NW46.1 13 5IKLWPPSESTRMLVERMTDNLSS..VSFFSRK..YGLLSKEEAENAKRIEETAFLAAND.....HEAKEPNLDDSSVQFYAREASKLMLLEALK 100
Q9C500.1 57 5LRIWPPQKTRDRAVNLRIETLST..ESILSKR..YGTLSKDDATTVAKLIEEAFVAVSN.....AVSSDDGKILELYSKEISKRMLLESVK 142
Q2HR17.1 25 5YSIWPPKQRTDRAVKNRIETLST..PSVLSKR..YGTMSADEASAARIQIEEAFSVANA.....SSSTSNQNVITILEVYSKEISKRMIETVK 110
Q9M7N3.1 28 5FKIWPPQRTREAVRRLVETLST..QSVLSKR..YGVIPEDATSAARIIEEAFSVAVS..ASAASTGGRPEDEWIEVLHIYSQEIQRVVESAK 119
Q9M7N6.1 25 5FSIWPPPTQRTDRAVINRLIESLST..PSILSKR..YGTLPQDEASETARLIEEAFAAAQS.....TASDADGGIEILQVYSKEISKRMIETVK 110
Q9LE82.1 14 5VKMWPSSKSTRMLVERMTKNITT..PSIFSRK..YGLLSVEEAQDQAKRIEDLAFATANK.....HFQNEPDGDDGSAVHVYAKESKLMLDVTK 101
Q9M651.2 13 5IKLWPPSLPTRKALIERITNFFSS..KTIFTEK..YGLSKDQATENAKRIEDIAFSTANQ.....QFEREPDGDGSSAVQLYAKECSKLILEVELK 100
B9R748.1 48 5LSIWPPPTQRTDRAVITRLIETLSS..PSVLSKR..YGTISHDEAESARRIEEAFGVANT.....ATSADDDGLEILQLYSKEISRRLMDTVK 133

```



Enzyme: Enzyme Commission (EC) number

- A numerical classification scheme of enzymes based on relevant chemical reaction patterns.
- EC number serves to associate a protein sequence with relevant chemical reactions. EC number consists of four level numbers, each number separated by a period (i.e., 'a.b.c.d').
- An EC number having all four level numbers for a protein sequence is the most specific annotation, which allows associating the protein sequence with specific chemical reaction(s)

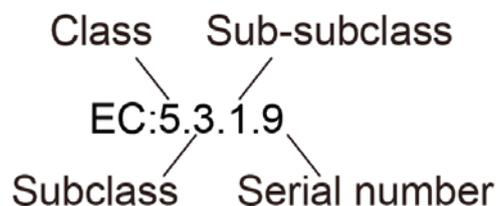


Table 1. Features of 6 different EC number prediction tools that are locally installable

EC number prediction tool	Disk space required (GB)	Number of predictable EC numbers	Last update (year)
DeepEC	0.045	4,669	2019
CatFam	2.072	1,653	2009
DETECT v2	0.854	786	2018
ECPred	9.427	858	2018
EFICAz2.5	24.682	2,757	2012
PRIAM	3.547	4,560	2018

- **Lightest**
- **Most comprehensive**
- **Most accurate**
- **Fastest**

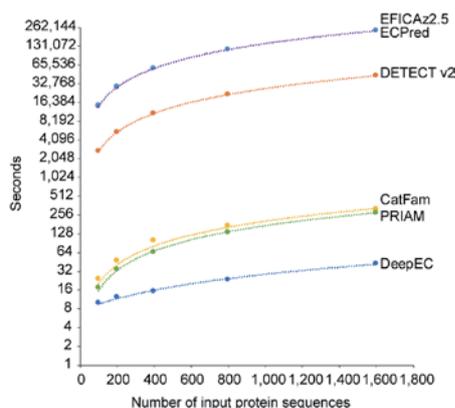


Table 2. Prediction performances of 6 different, locally installable EC number prediction tools using 201 enzyme protein sequences as inputs, which were never considered for the development of all these tools (from the Swiss-Prot database released August 2018)

EC number prediction tool	Precision	Recall	Run time (s)
DeepEC	0.920	0.455	13
CatFam	0.880	0.327	47
DETECT v2	0.804	0.203	5,480
ECPred	0.817	0.243	28,840
EFICAz2.5	0.737	0.416	29,093
PRIAM	0.809	0.356	51

Source code: <https://bitbucket.org/kaistsystemsbiology/deepec>
 Ryu et al. *PNAS* (2019)

EH

What is the greatest AI achievement?



As an AI language model, I am not programmed to have personal opinions, but I can provide some examples of what some experts and researchers consider to be some of the greatest AI achievements:



1. AlphaGo defeating the world champion in the game of Go, as it demonstrated the ability of AI to learn and excel at complex strategy games that require intuition and creativity.
2. The development of deep learning techniques, particularly convolutional neural networks, which have enabled significant advances in image recognition, natural language processing, and speech recognition.
3. The creation of GPT-3, a language model that can generate human-like text and has the ability to complete tasks such as writing stories, composing poetry, and even programming.
4. The development of self-driving cars, which use AI to perceive the environment, make decisions, and navigate roads safely and efficiently.
5. The breakthrough in protein folding prediction achieved by AlphaFold, which has the potential to revolutionize drug discovery and other areas of medicine.



These are just a few examples of significant AI achievements, but there are many more, and the field of AI continues to evolve rapidly.