Unmet Data / Infrastructure /Computational Needs

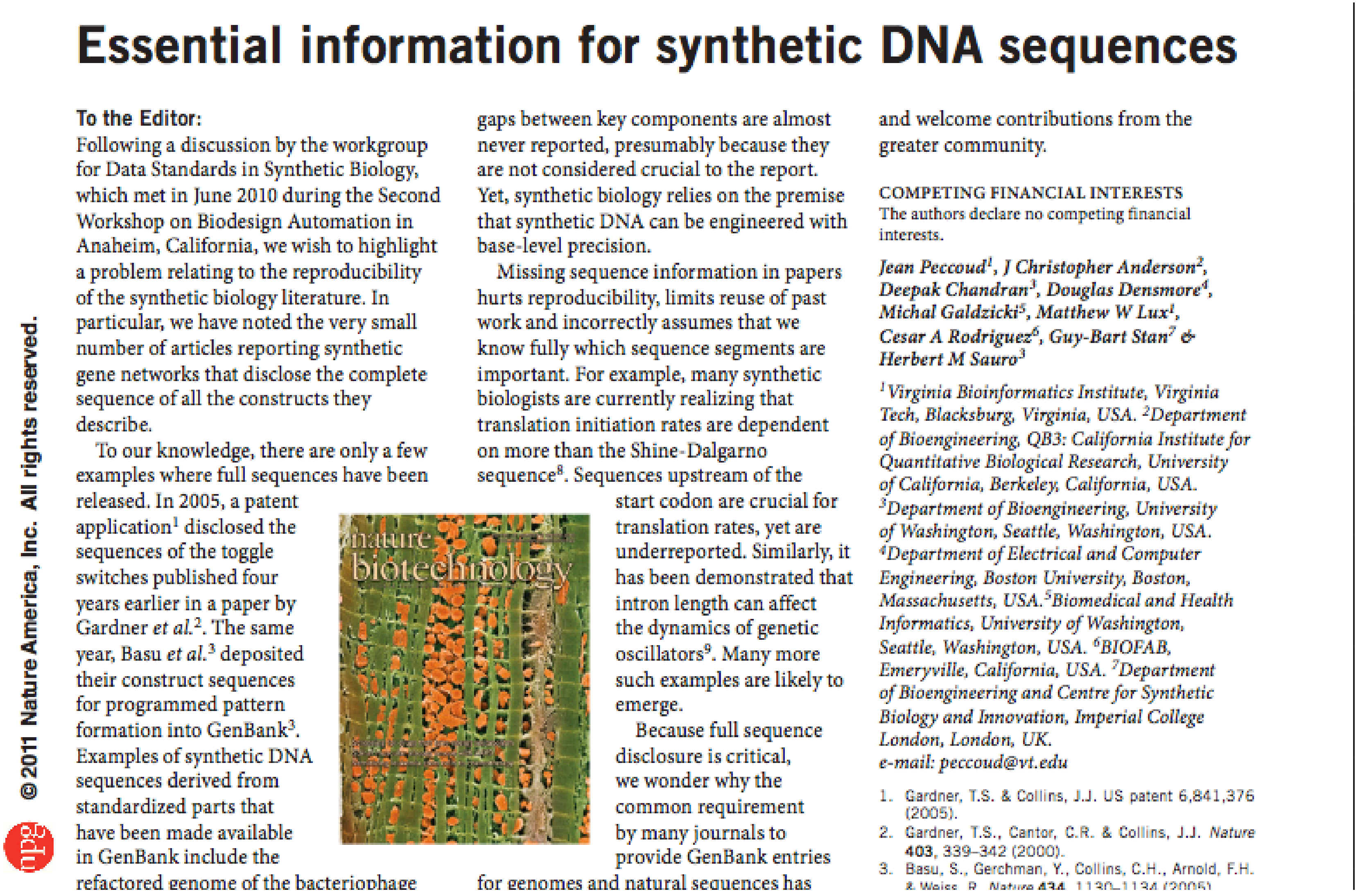
Doug Densmore
Nathan Hillson
Eric Klavins
Chris Myers

Reproducibility Crisis

Professor David Donoho
Stanford University

An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete ... set of instructions [and data] which generated the figures.

Reproducibility Crisis in SynBio



Essential information for synthetic DNA sequences

To the Editor:
Following a discussion by the workgroup for Data Standards in Synthetic Biology, which met in June 2010 during the Second Workshop on Biodesign Automation in Anaheim, California, we wish to highlight a problem relating to the reproducibility of the synthetic biology literature. In particular, we have noted the very small number of articles reporting synthetic gene networks that disclose the complete sequence of all the constructs they describe.

To our knowledge, there are only a few examples where full sequences have been released. In 2005, a patent application[1] disclosed the sequences of the toggle switches published four years earlier in a paper by Gardner *et al.*[2]. The same year, Basu *et al.*[3] deposited their construct sequences for programmed pattern formation into GenBank[3]. Examples of synthetic DNA sequences derived from standardized parts that have been made available in GenBank include the refactored genome of the bacteriophage

gaps between key components are almost never reported, presumably because they are not considered crucial to the report. Yet, synthetic biology relies on the premise that synthetic DNA can be engineered with base-level precision.

Missing sequence information in papers hurts reproducibility, limits reuse of past work and incorrectly assumes that we know fully which sequence segments are important. For example, many synthetic biologists are currently realizing that translation initiation rates are dependent on more than the Shine-Dalgarno sequence[8]. Sequences upstream of the start codon are crucial for translation rates, yet are underreported. Similarly, it has been demonstrated that intron length can affect the dynamics of genetic oscillators[9]. Many more such examples are likely to emerge.

Because full sequence disclosure is critical, we wonder why the common requirement by many journals to provide GenBank entries for genomes and natural sequences has

and welcome contributions from the greater community.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Jean Peccoud[1], J Christopher Anderson[2], Deepak Chandran[3], Douglas Densmore[4], Michal Galdzicki[5], Matthew W Lux[1], Cesar A Rodriguez[6], Guy-Bart Stan[7] & Herbert M Sauro[3]

[1]Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia, USA. [2]Department of Bioengineering, QB3: California Institute for Quantitative Biological Research, University of California, Berkeley, California, USA. [3]Department of Bioengineering, University of Washington, Seattle, Washington, USA. [4]Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts, USA. [5]Biomedical and Health Informatics, University of Washington, Seattle, Washington, USA. [6]BIOFAB, Emeryville, California, USA. [7]Department of Bioengineering and Centre for Synthetic Biology and Innovation, Imperial College London, London, UK.
e-mail: peccoud@vt.edu

1. Gardner, T.S. & Collins, J.J. US patent 6,841,376 (2005).
2. Gardner, T.S., Cantor, C.R. & Collins, J.J. *Nature* **403**, 339–342 (2000).
3. Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H. & Weiss, R. *Nature* **434**, 1130–1134 (2005).

Data Sharing

Lacking consistency in representation and annotation.

Limited characterization data, and usually without units.

EMBL-EBI

reactome

NCBI

addgene

The nonprofit plasmid repository

iGEM

Data Standards: SBOL

nature biotechnology

Home | Current issue | News & comment | Research | Archive ▼ | Authors & referees ▼ | About the journal

home ▸ archive ▸ issue ▸ computational biology ▸ perspective ▸ full text

NATURE BIOTECHNOLOGY | COMPUTATIONAL BIOLOGY | PERSPECTIVE

日本語要約

The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology

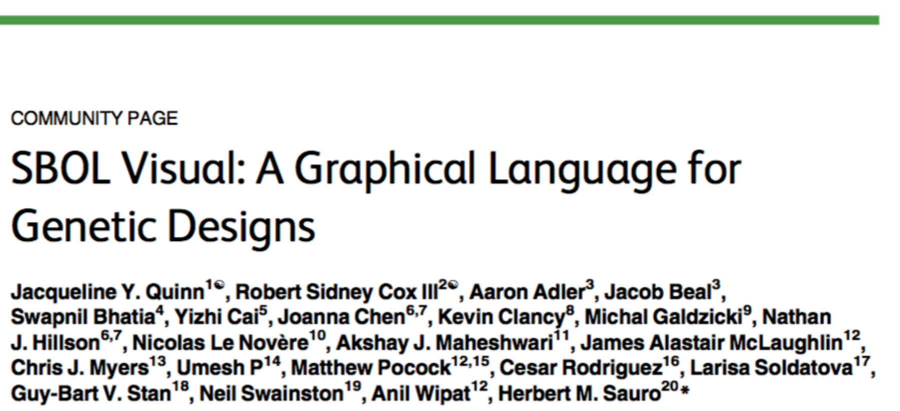
Michal Galdzicki, Kevin P Clancy, Ernst Oberortner, Matthew Pocock, Jacqueline Y Quinn, Cesar A Rodriguez, Nicholas Roehner, Mandy L Wilson, Laura Adam, J Christopher Anderson, Bryan A Bartley, Jacob Beal, Deepak Chandran, Joanna Chen, Douglas Densmore, Drew Endy, Raik Grünberg, Jennifer Hallinan, Nathan J Hillson, Jeffrey D Johnson, Allan Kuchinsky, Matthew Lux, Goksel Misirli, Jean Peccoud, Hector A Plahar, Evren Sirin, Guy-Bart Stan, Alan Villalobos, Anil Wipat, John H Gennari, Chris J Myers & Herbert M Sauro

⊟ Show fewer authors

Affiliations | Contributions | Corresponding author

Nature Biotechnology **32**, 545–550 (2014) | doi:10.1038/nbt.2891

Received 09 November 2013 | Accepted 20 December 2013 | Published online 06 June 2014

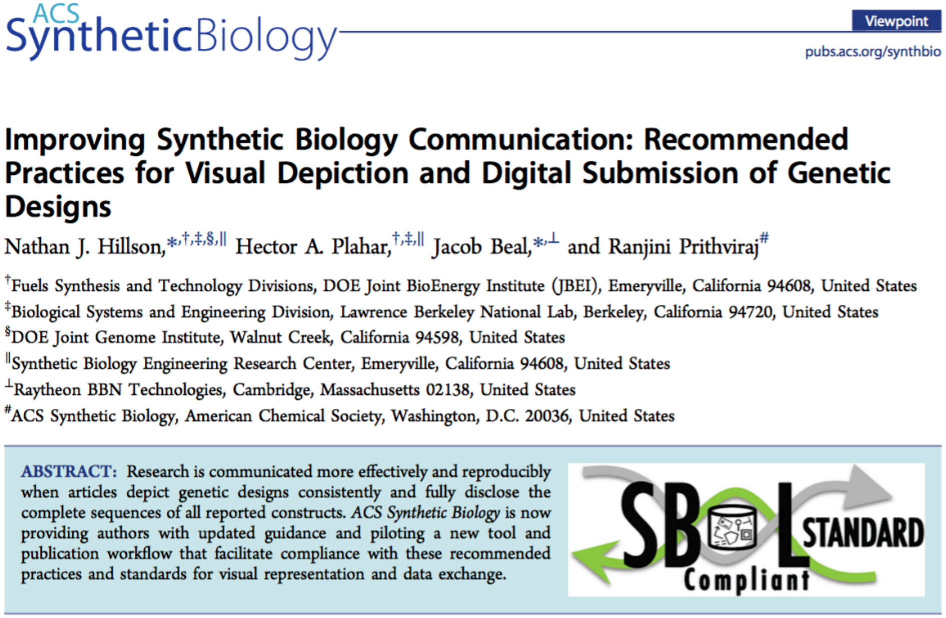


PLOS | BIOLOGY

COMMUNITY PAGE

SBOL Visual: A Graphical Language for Genetic Designs

Jacqueline Y. Quinn[1]◎, Robert Sidney Cox III[2]◎, Aaron Adler[3], Jacob Beal[3], Swapnil Bhatia[4], Yizhi Cai[5], Joanna Chen[6,7], Kevin Clancy[8], Michal Galdzicki[9], Nathan J. Hillson[6,7], Nicolas Le Novère[10], Akshay J. Maheshwari[11], James Alastair McLaughlin[12], Chris J. Myers[13], Umesh P[14], Matthew Pocock[12,15], Cesar Rodriguez[16], Larisa Soldatova[17], Guy-Bart V. Stan[18], Neil Swainston[19], Anil Wipat[12], Herbert M. Sauro[20]*



ACS SyntheticBiology

Viewpoint

pubs.acs.org/synthbio

Improving Synthetic Biology Communication: Recommended Practices for Visual Depiction and Digital Submission of Genetic Designs

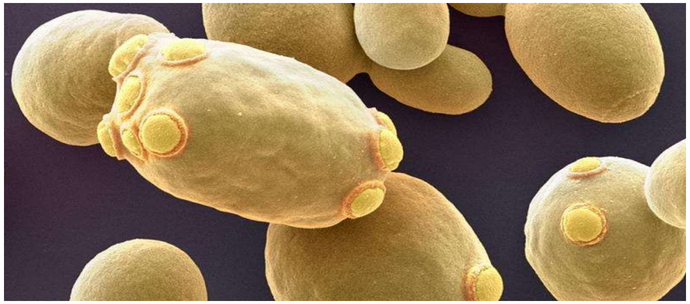
Nathan J. Hillson,[*,†,‡,§,∥] Hector A. Plahar,[†,‡,∥] Jacob Beal,[*,⊥] and Ranjini Prithviraj[#]

[†] Fuels Synthesis and Technology Divisions, DOE Joint BioEnergy Institute (JBEI), Emeryville, California 94608, United States
[‡] Biological Systems and Engineering Division, Lawrence Berkeley National Lab, Berkeley, California 94720, United States
[§] DOE Joint Genome Institute, Walnut Creek, California 94598, United States
[∥] Synthetic Biology Engineering Research Center, Emeryville, California 94608, United States
[⊥] Raytheon BBN Technologies, Cambridge, Massachusetts 02138, United States
[#] ACS Synthetic Biology, American Chemical Society, Washington, D.C. 20036, United States

ABSTRACT: Research is communicated more effectively and reproducibly when articles depict genetic designs consistently and fully disclose the complete sequences of all reported constructs. *ACS Synthetic Biology* is now providing authors with updated guidance and piloting a new tool and publication workflow that facilitate compliance with these recommended practices and standards for visual representation and data exchange.

SBOL STANDARD Compliant

Data: Novel Chassis

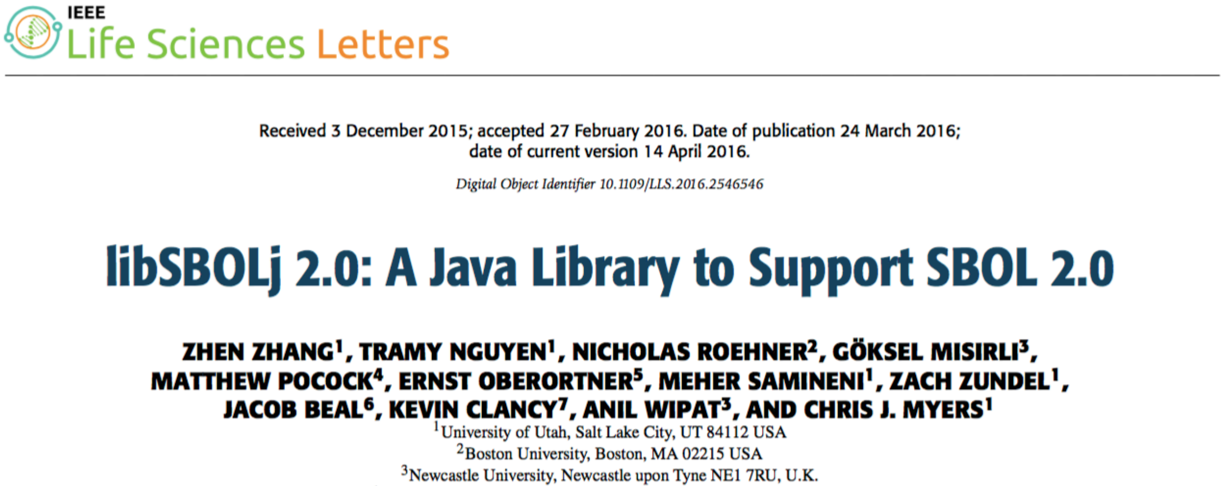


- Many parts used in synthetic biology come from or are initially tested in *E. Coli.*
- However, many applications require different bacteria or higher level organisms (i.e., yeast and other eukaryotic cells).
- Researchers use trial-and-error, since they cannot find reliable information about prior attempts.
- To scale, a wide range of data must be harnessed to assess confidence of success.

Data Opportunities

- Continue support of international standardization efforts, such as COMBINE and BioRoboost.
- Journals/funding agencies need to sufficiently incentivize data sharing (carrot and stick).
    - Repeating experiments, especially negative ones, results in substantial waste of resources.
    - Reproducibility and formalization of data need to demonstrate value and become part of the culture.
    - Reviewers are key to achieve this goal.
- Promote the curation of well characterized part libraries in a variety of host contexts.

Infrastructure: Software Libraries



Received 3 December 2015; accepted 27 February 2016. Date of publication 24 March 2016;
date of current version 14 April 2016.

Digital Object Identifier 10.1109/LLS.2016.2546546

libSBOLj 2.0: A Java Library to Support SBOL 2.0

ZHEN ZHANG[1], TRAMY NGUYEN[1], NICHOLAS ROEHNER[2], GÖKSEL MISIRLI[3],
MATTHEW POCOCK[4], ERNST OBERORTNER[5], MEHER SAMINENI[1], ZACH ZUNDEL[1],
JACOB BEAL[6], KEVIN CLANCY[7], ANIL WIPAT[3], AND CHRIS J. MYERS[1]

[1]University of Utah, Salt Lake City, UT 84112 USA
[2]Boston University, Boston, MA 02215 USA
[3]Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.

Technical Note

pySBOL: A Python Package for Genetic Desig
Standardization

Bryan A. Bartley*†, Kiri Choi‡, Meher Samineni§, Zach Zundel§, Tramy Nguyen§, Chris
Herbert M. Sauro‡

† Raytheon BBN Technologies, Cambridge, Massachusetts 02138, United States
‡ Department of Bioengineering, University of Washington, Seattle, Washington 98195, United States
§ Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, Utah 84112,

ACS Synth. Biol., Article ASAP
DOI: 10.1021/acssynbio.8b00336
Publication Date (Web): November 13, 2018
Copyright © 2018 American Chemical Society

Cite this: *ACS Synth. Biol.* XXXX, X

*E-mail: bryan.a.bartley@raytheon.com.
This article is part of the IWBDA 2018 special issue.

Technical Note

sboljs: Bringing the Synthetic Biology Open Language to the Web Browser

James Alastair McLaughlin†, Chris J. Myers‡, Zach Zundel‡, Nathan Wilkinson‡, Christian Atallah†, and Anil Wipat*†

† School of Computing, Newcastle University, Newcastle upon Tyne NE4 5TG, U.K.
‡ Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, Utah 84112, United States

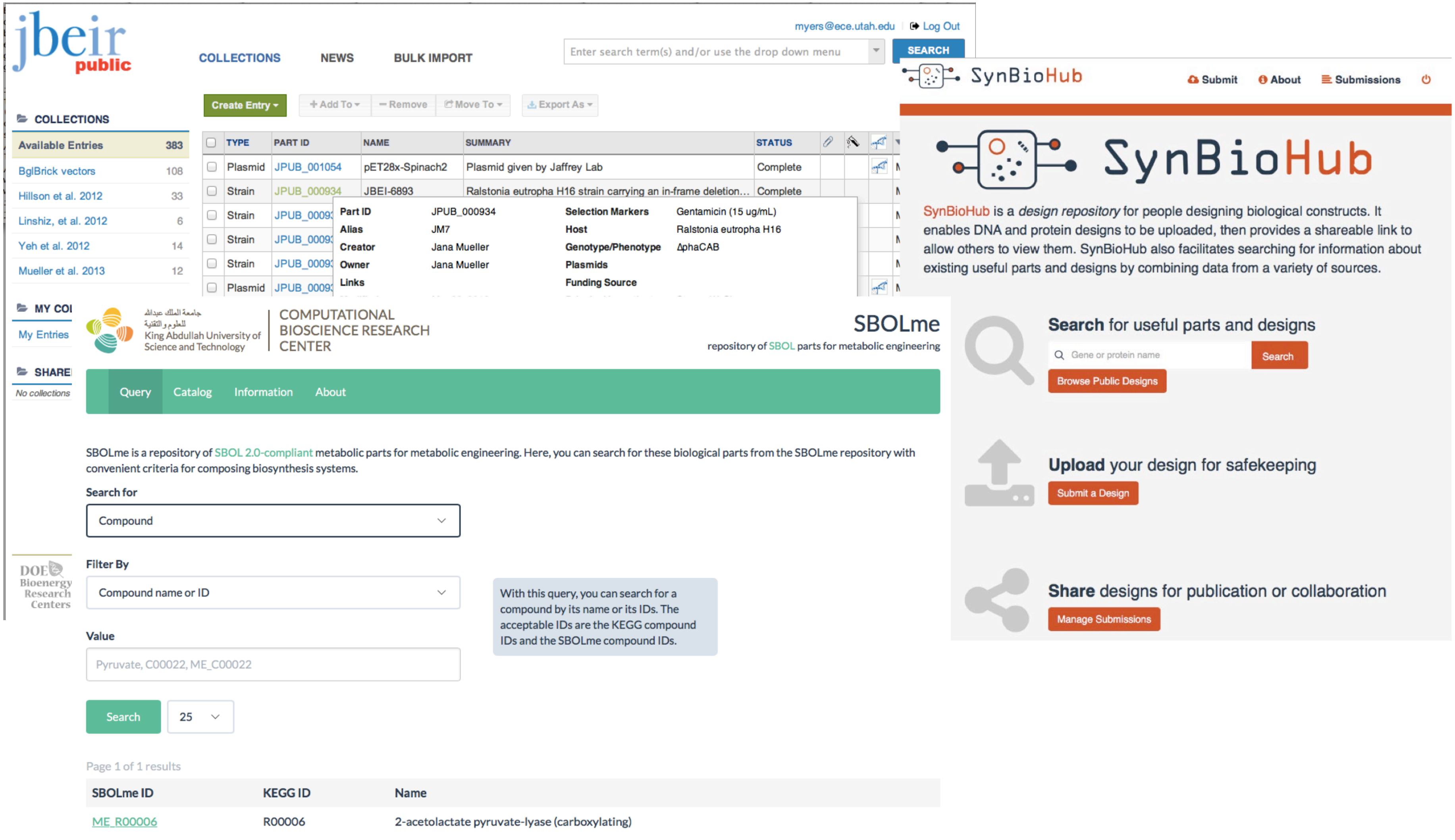
ACS Synth. Biol., **2019**, *8* (1), pp 191–193
DOI: 10.1021/acssynbio.8b00338
Publication Date (Web): November 7, 2018
Copyright © 2018 American Chemical Society

Cite this: *ACS Synth. Biol.* 2019, 8, 1, 191-193

*E-mail: anil.wipat@ncl.ac.uk.

Supported by NSF Grants:
DBI-1356041 and DBI-1355909

Infrastructure: Repositories



Collection	Entries
Available Entries	383
BglBrick vectors	108
Hillson et al. 2012	33
Linshiz, et al. 2012	6
Yeh et al. 2012	14
Mueller et al. 2013	12

☐	TYPE	PART ID	NAME	SUMMARY	STATUS
☐	Plasmid	JPUB_001054	pET28x-Spinach2	Plasmid given by Jaffrey Lab	Complete
☐	Strain	JPUB_000934	JBEI-6893	Ralstonia eutropha H16 strain carrying an in-frame deletion…	Complete
☐	Strain	JPUB_0009			
☐	Strain	JPUB_0009			
☐	Strain	JPUB_0009			
☐	Plasmid	JPUB_0009			

Part ID	JPUB_000934	**Selection Markers**	Gentamicin (15 ug/mL)
Alias	JM7	**Host**	Ralstonia eutropha H16
Creator	Jana Mueller	**Genotype/Phenotype**	ΔphaCAB
Owner	Jana Mueller	**Plasmids**	
Links		**Funding Source**	

SBOLme is a repository of SBOL 2.0-compliant metabolic parts for metabolic engineering. Here, you can search for these biological parts from the SBOLme repository with convenient criteria for composing biosynthesis systems.

Search for: Compound

Filter By: Compound name or ID

Value: Pyruvate, C00022, ME_C00022

With this query, you can search for a compound by its name or its IDs. The acceptable IDs are the KEGG compound IDs and the SBOLme compound IDs.

Page 1 of 1 results

SBOLme ID	KEGG ID	Name
ME_R00006	R00006	2-acetolactate pyruvate-lyase (carboxylating)

SynBioHub is a *design repository* for people designing biological constructs. It enables DNA and protein designs to be uploaded, then provides a shareable link to allow others to view them. SynBioHub also facilitates searching for information about existing useful parts and designs by combining data from a variety of sources.

Search for useful parts and designs

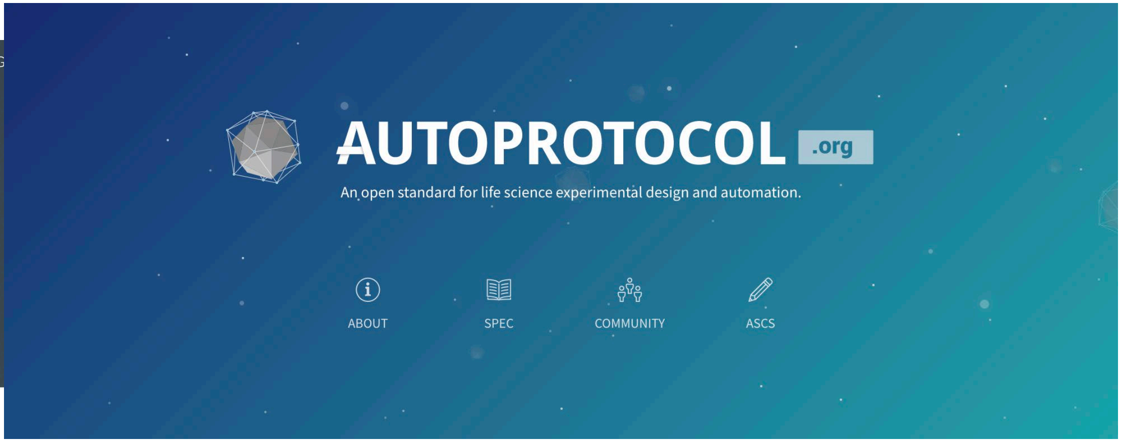
Upload your design for safekeeping

Share designs for publication or collaboration

Infrastructure: Automation Software

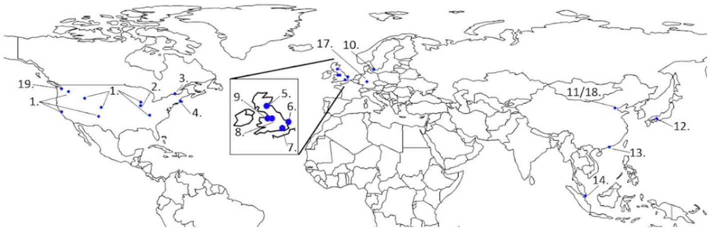
Global Biofoundries Alliance

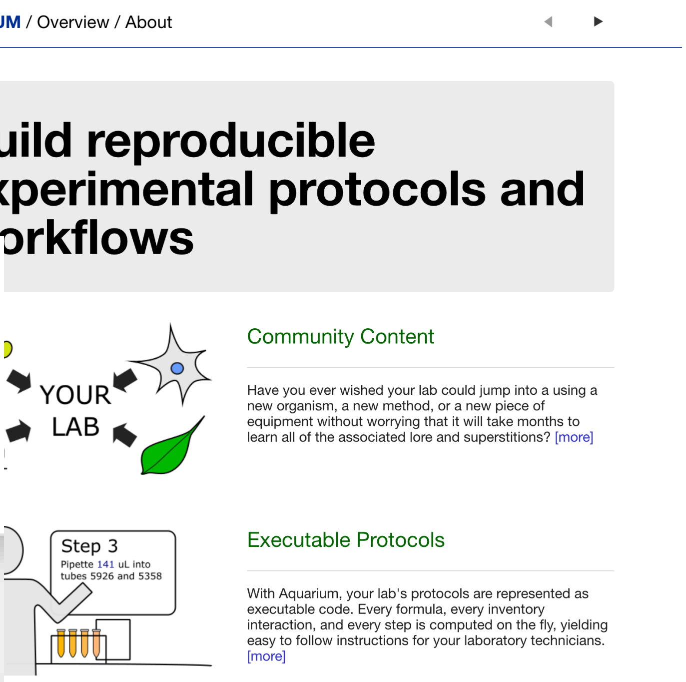
Formal Launch: 9AM (Kobe, Japan local time), 9th of May, 2019



AUTOPROTOCOL .org

An open standard for life science experimental design and automation.





AQUARIUM / Overview / About

The Laboratory Operating System

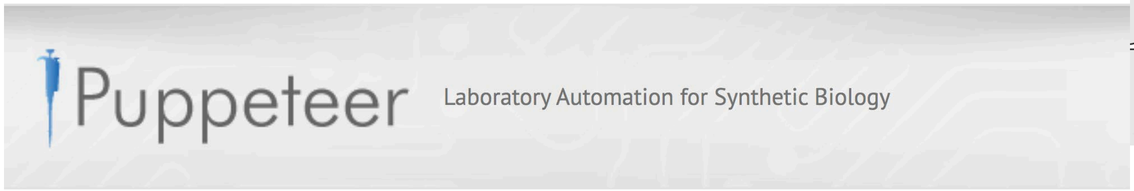
Build reproducible experimental protocols and workflows

Community Content

Have you ever wished your lab could jump into a using a new organism, a new method, or a new piece of equipment without worrying that it will take months to learn all of the associated lore and superstitions? [more]

Executable Protocols

With Aquarium, your lab's protocols are represented as executable code. Every formula, every inventory interaction, and every step is computed on the fly, yielding easy to follow instructions for your laboratory technicians. [more]

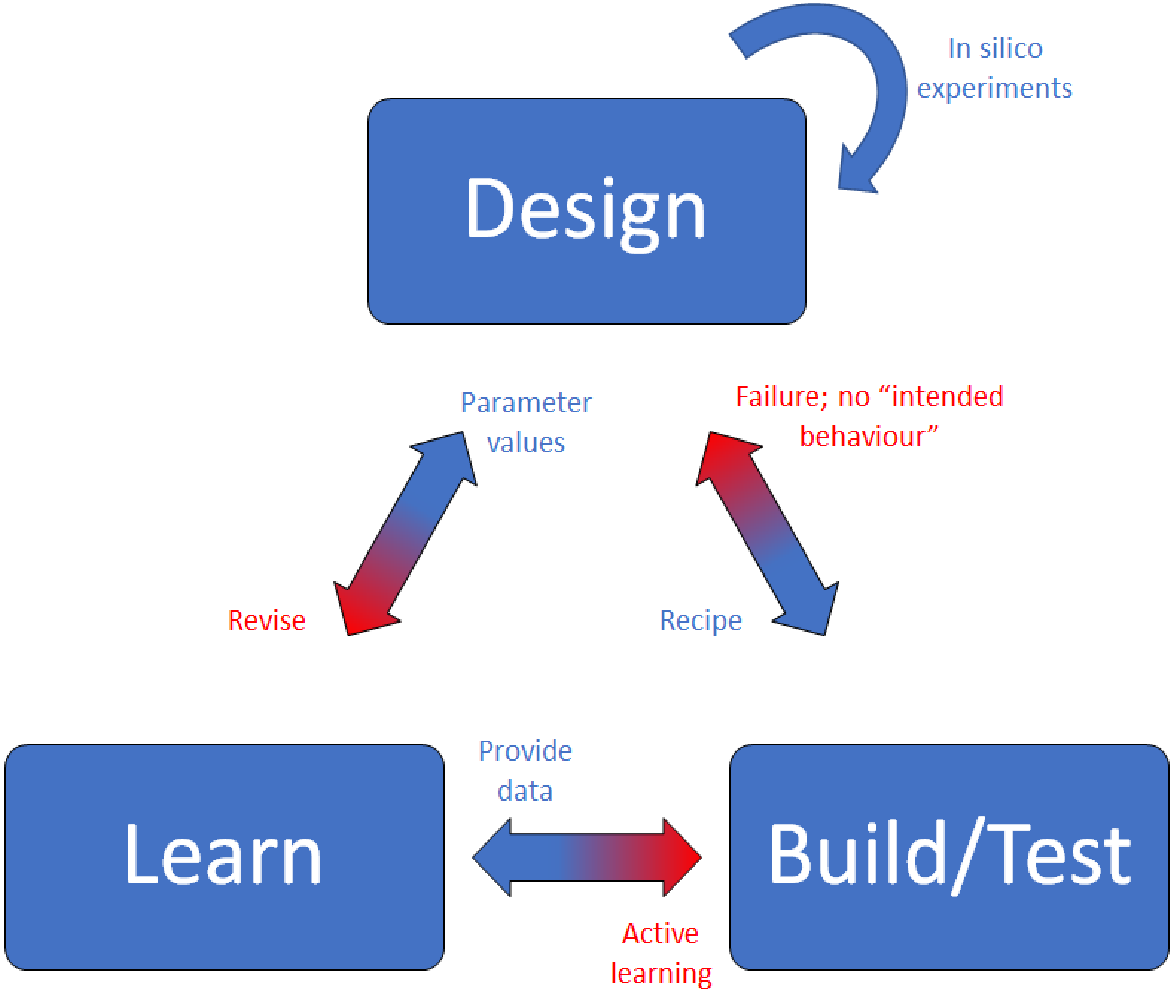


Laboratory Automation for Synthetic Biology

Infrastructure Opportunities

- Library, repository, automation infrastructure used by many projects, but often limited or no direct support of this infrastructure.
- While open standard development may only require meeting support, more substantial investment is needed for infrastructure.

Computational Support for Design/Build/Test/Learn



Design

In silico experiments

Parameter values

Failure; no “intended behaviour”

Revise

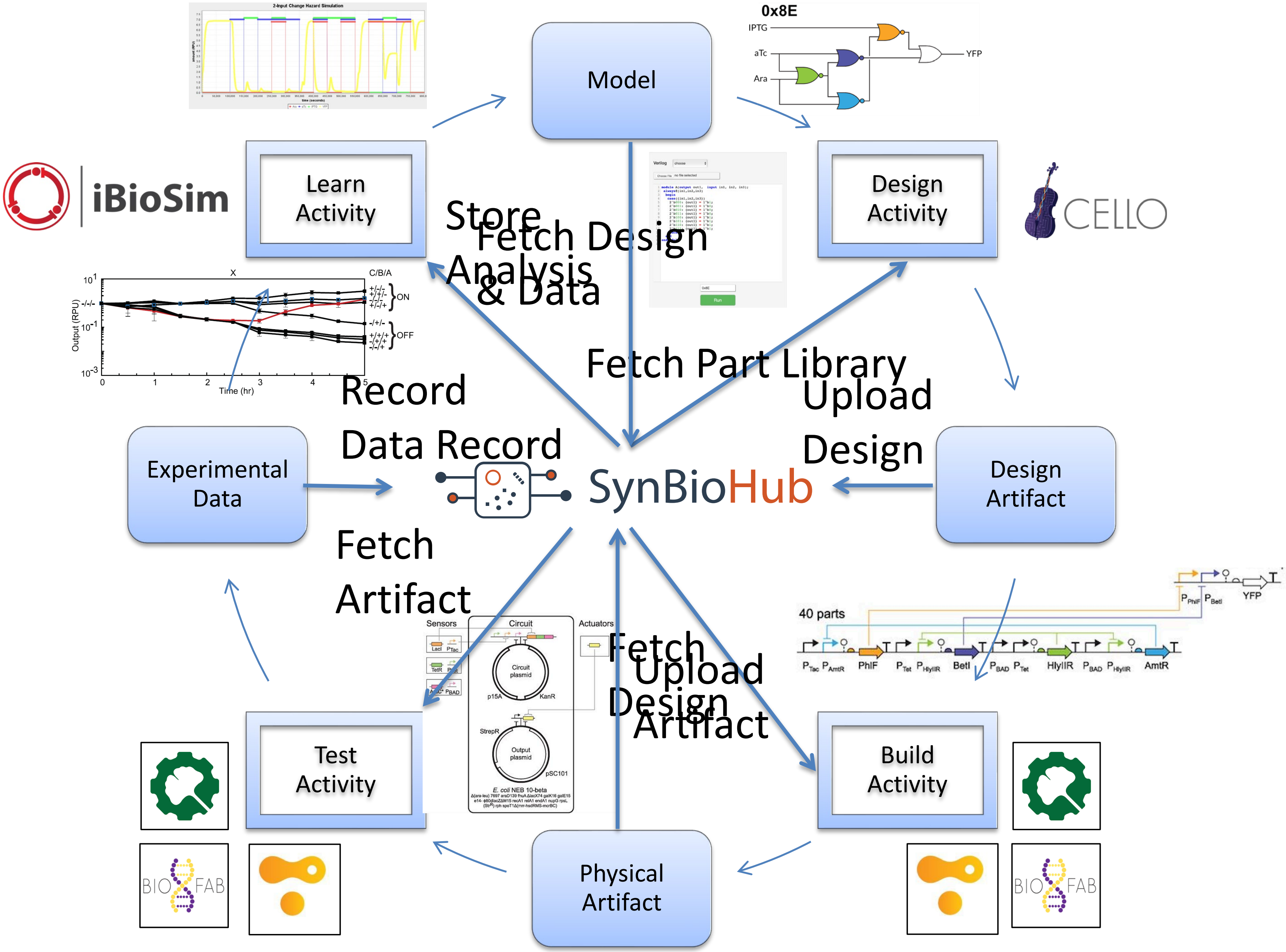
Recipe

Provide data

Learn

Build/Test

Active learning



Model

Design Activity

Design Artifact

Build Activity

Physical Artifact

Test Activity

Experimental Data

Learn Activity

SynBioHub

Store Analysis & Data

Fetch Design

Fetch Part Library

Upload Design

Record Data Record

Fetch Artifact

Fetch Design

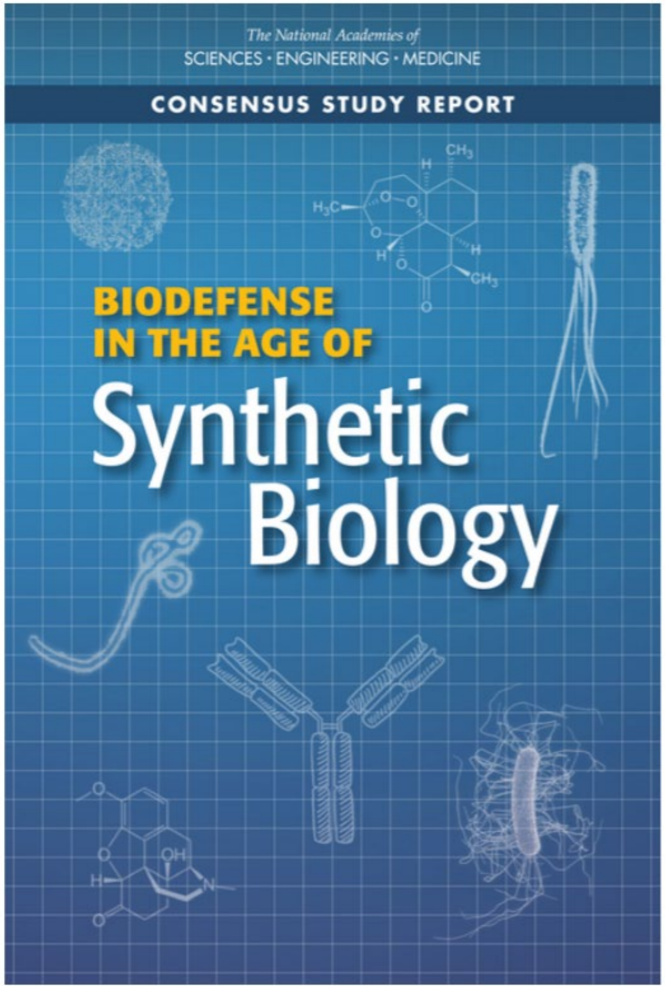
Upload Design Artifact

Computational Opportunities

- While data standards have evolved to capture the entire design/build/test/learn workflow, computational tools have not kept pace.
- Sequence editors are well supported by industry:
  - Benchling/SnapGene/Geneious
- Support is needed to promote the development of:
  - Effective languages for specifying design requirements/constraints.
  - Accurate multi-scale models that support automated abstraction.
  - Functional design, modeling, analysis, and visualization tools that exploit the inherent stochastic behavior of biological circuits.
  - Efficient methods for design space exploration using machine learning and other techniques.
  - Interfaces that can easily fit into existing experimental workflows.
- Need to change focus in the bioengineering curriculum from trial-and-error experimentation to in-silico design and automation.

Bio-security: Potential Opportunities for Improving Deterrence & Prevention Capabilities

(p. 130-131 of report)



The National Academies of
SCIENCES · ENGINEERING · MEDICINE

CONSENSUS STUDY REPORT

BIODEFENSE
IN THE AGE OF
Synthetic
Biology

- ***LEARN*** - Screening of activities with machine learning
- ***CONSTRAIN*** - Systems to constrain design capabilities
- ***REGISTER*** - Maintaining registries of known expertise and materials
- ***SCREEN*** - Maintaining registries of known biological threats
- ***TRACK*** - Tracking digital “signatures” in genetic designs