

A Standard Enabled Workflow for Synthetic Biology

Chris J. Myers

2016 COMBINE Forum

September 19, 2016

Reproducibility in Synthetic Biology

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 application¹ 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 GenBank³. 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 sequence⁴. 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 oscillators⁵. 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 Peccoud¹, J Christopher Anderson², Deepak Chandran³, Douglas Densmore⁴, Michal Galczynski⁵, Matthew W Lux¹, Cesar A Rodriguez⁶, Guy-Bart Stan⁷ & Herbert M Sauro³

¹Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia, USA. ²Department of Bioengineering, QB3: California Institute for Quantitative Biological Research, University of California, Berkeley, California, USA.

³Department of Bioengineering, University of Washington, Seattle, Washington, USA.

⁴Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts, USA. ⁵Biomedical and Health Informatics, University of Washington, Seattle, Washington, USA. ⁶BIOFAB, Emeryville, California, USA. ⁷Department of Bioengineering and Centre for Synthetic Biology and Innovation, Imperial College London, London, UK.

e-mail: peccoud@vt.edu

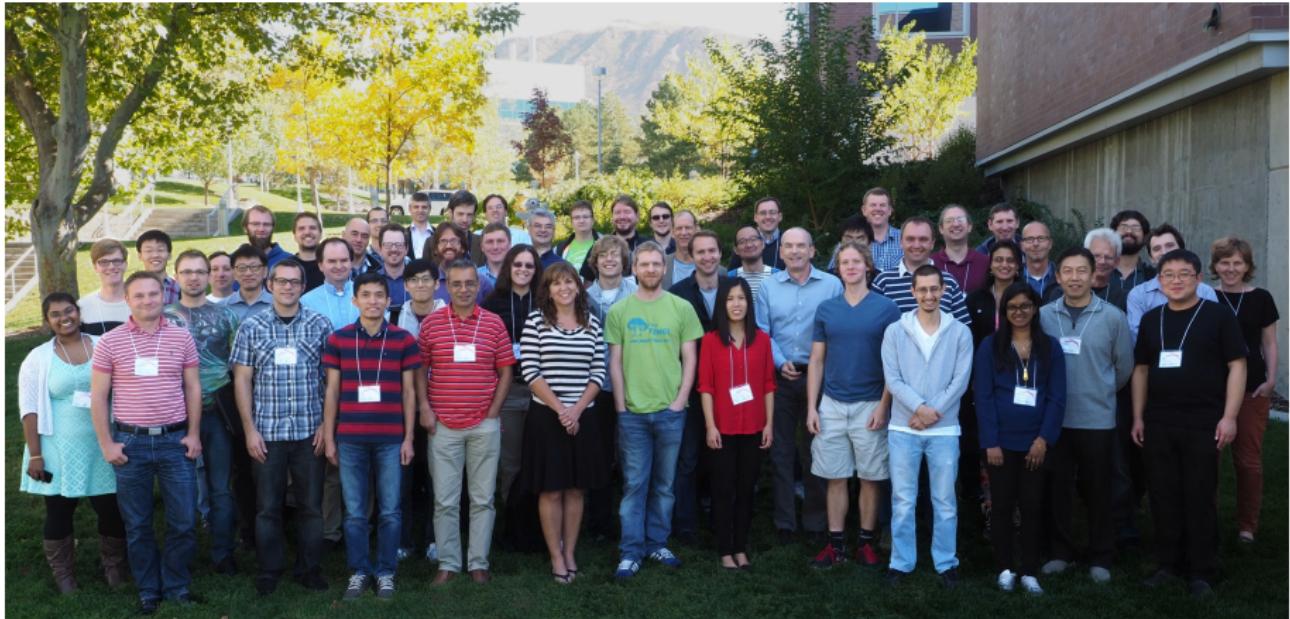
1. Gardner, T.S. & Collins, J.J. US patent 6,841,376 (2005).
2. Gardner, T.S., Cantor, C.R. & Collins, J.J. *Nature* **403**, 339–342 (2000).
3. Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H. & Voigt, C.A. *Nature* **431**, 1120–1124 (2004).

COMBINE Standards



- COMBINE Standards:
 - SBML, CellML, SED-ML, BioPAX, SBGN, SBOL, and NeuroML.
- Associated standardization efforts:
 - COMBINE Archive, identifiers.org, SBO, KISAO, and BioModels.net qualifiers.
- Related standardization efforts:
 - CNO, FieldML, GPML, MAMO, NineML, NuML, pharmML, PSI-MI, SpineML, and Teddy.

COMBINE Community



Attendees of 2015 COMBINE Forum in Salt Lake City, UT USA
66 participants from 11 countries and 6 continents

Synthetic Biology Open Language (SBOL) Community



At least 16 attendees at the forum represented the SBOL Community

Synthetic Biology Open Language (SBOL) Community



118 people from around the world

40 universities, 22 companies, 6 other types of organizations

Organizations Supporting SBOL



Current support for the development of SBOL provided by National Science Foundation Grants DBI-1356041 and DBI-1355909, and the Engineering and Physical Sciences Research Council under Grant Number EP/J02175X/1.

Synthetic Biology Open Language (SBOL)

- In 2008, a small group of researchers proposed the development of the SBOL, an open-source standard for the exchange of genetic designs.
- In 2011, the first version of the SBOL core data model was released.
 - Galdzicki et al., *Nature Biotechnology* (2014)
- In 2013, the first version of the SBOL Visual standard was released.
 - Quinn et al., *PLoS Biology* (2015)
- In 2015, SBOL Version 2.0 was officially released.
 - Roehner et al., *ACS Synthetic Biology* (2014)
 - Bartley et al., *Journal of Integrative Biology* (2015)
 - Roehner et al., *ACS Synthetic Biology* (2016)
- In 2016, ACS Synthetic Biology Journal recommended SBOL workflow.
 - Hillson et al., *ACS Synthetic Biology* (2016).

SBOL Governance

- **SBOL Chair** - Anil Wipat (Newcastle)
- **Past SBOL Chair** - Herbert Sauro (Washington)
- **SBOL Editors** - Jacob Beal (Raytheon/BBN), Robert Sidney Cox (Kobe), Raik Grunberg (KAUST), James McLaughlin (Newcastle), and Tramy Nguyen (Utah).
- **Past SBOL Editors** - Bryan Bartley (Washington), Kevin Clancy (ThermoFisher), Michal Galdzicki (Washington), Goksel Misirli (Newcastle), Ernst Oberortner (DOE JGI), Matthew Pocock (Newcastle), Jacqueline Quinn (Google), Cesar Rodriguez (Autodesk), Nicholas Roehner (Boston), and Mandy Wilson (VBI).
- **Steering Committee** - Jacob Beal (Raytheon/BBN), Kevin Clancy (ThermoFischer), Douglas Densmore (Boston), John Gennari (Washington), Nathan Hillson (JBEI), Chris Myers (Utah), and Herbert Sauro (Washington).
- **Advisory Panel** - Matthew Chang (Singapore), Victor de Lorenzo (CSIC), Traci Haddock (iGEM), Richard Kitney (Imperial), Sarah Munro (NIST), and Chris Voigt (MIT).
- **COMBINE Coordinator** - Chris Myers (Utah)

SBOL Visual (Version 1.0)

 promoter	 origin of replication
 cds	 primer binding site
 ribosome entry site	 blunt restriction site
 terminator	 sticky restriction site
 operator	 5' overhang
 insulator	 3' overhang
 ribonuclease site	 assembly scar
 rna stability element	 signature
 protease site	 user defined
 protein stability element	

New symbols
added by
community
consensus.

Quinn et al., PLoS Biology (2015)

SBOL Data Model (Version 2.0)

FASTA

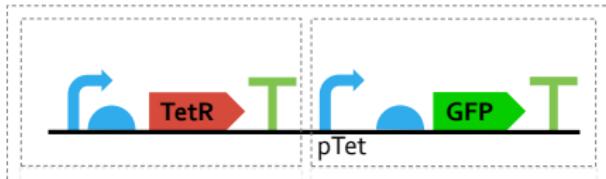
ACTGTGCCGTTAACGTGATTAAATCCGTACTGATAT...



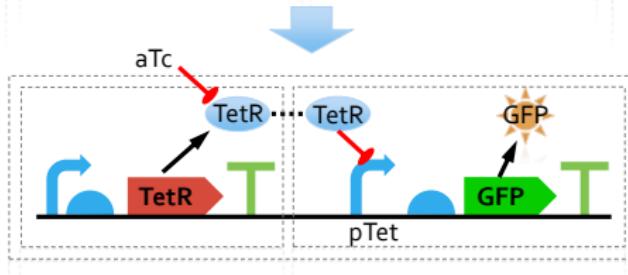
GenBank



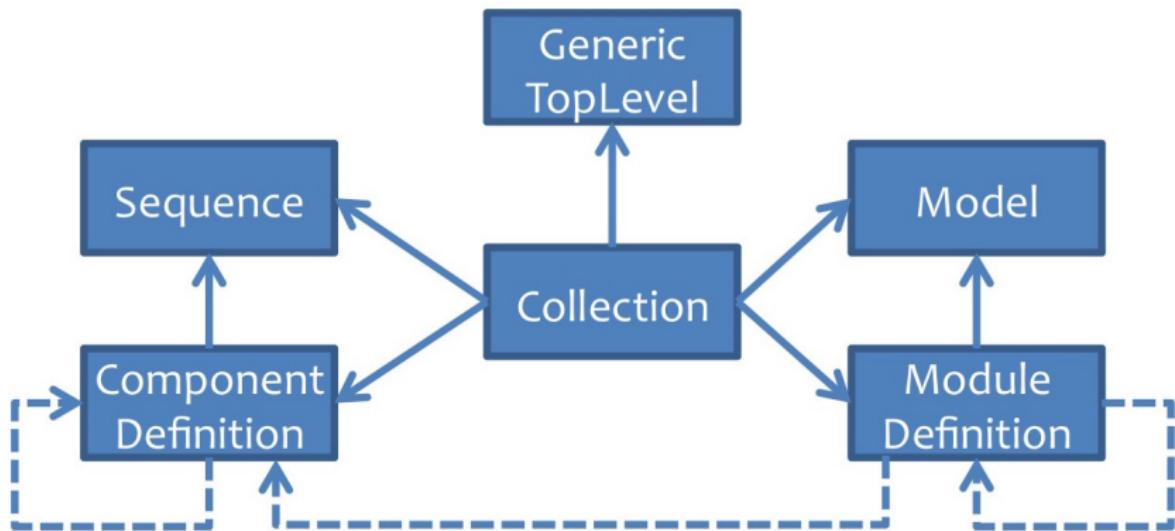
SBOL 1.1



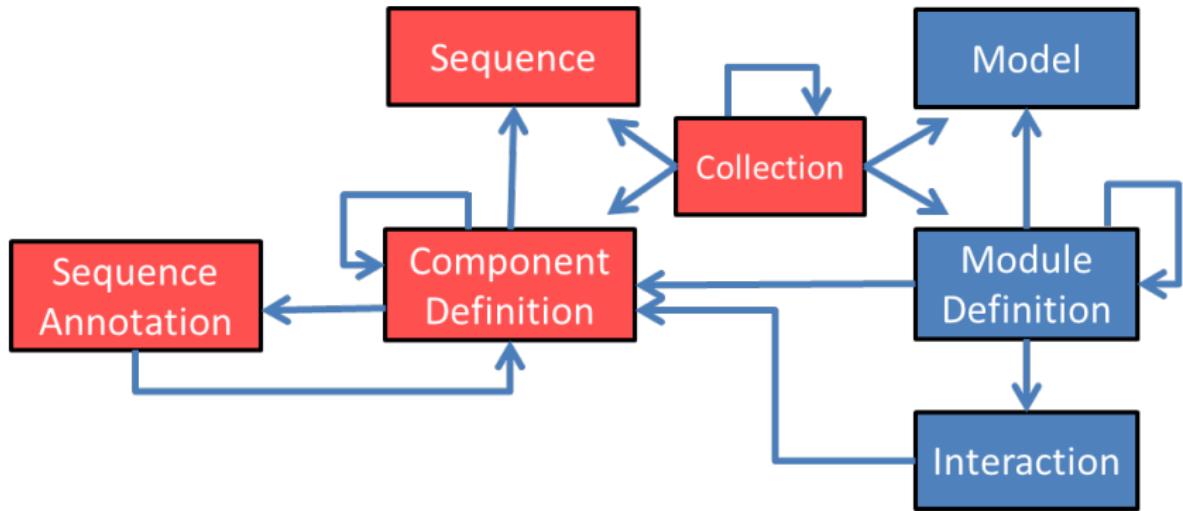
SBOL 2.0



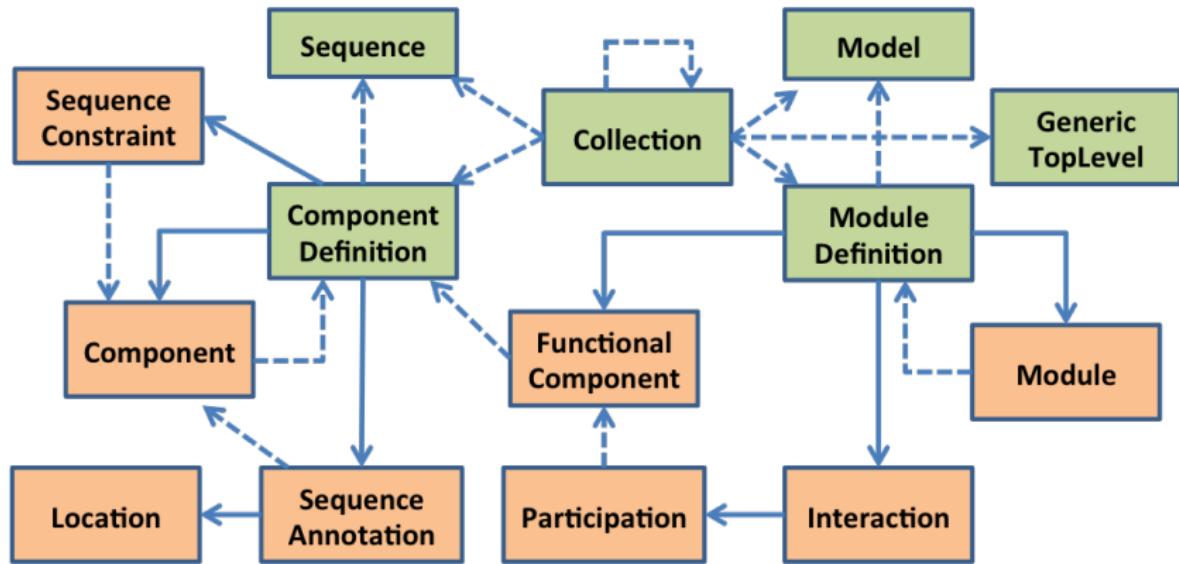
SBOL 2.0 Data Model



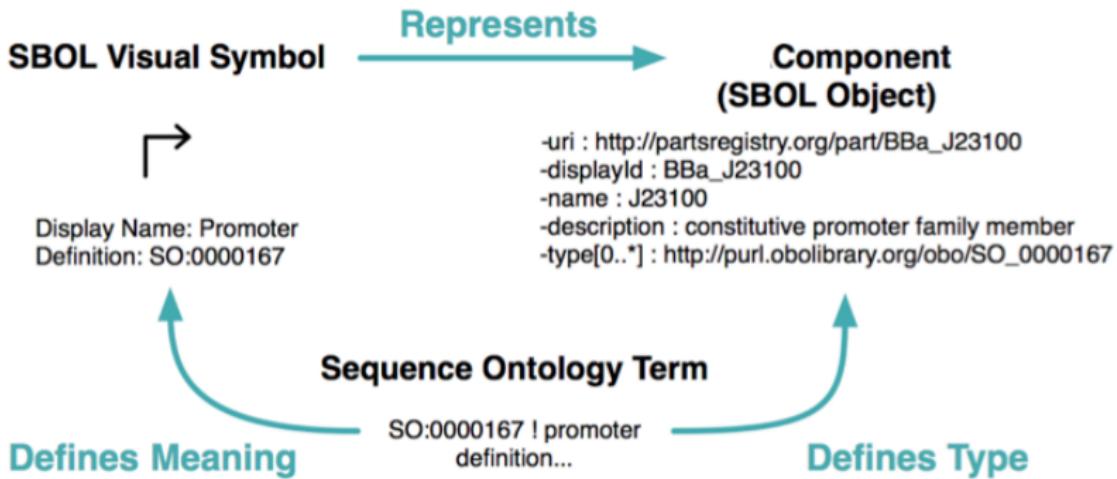
SBOL 2.0 Data Model



SBOL 2.0 Data Model



Connection Between Visual and Data Model



Library Support for SBOL 2.0

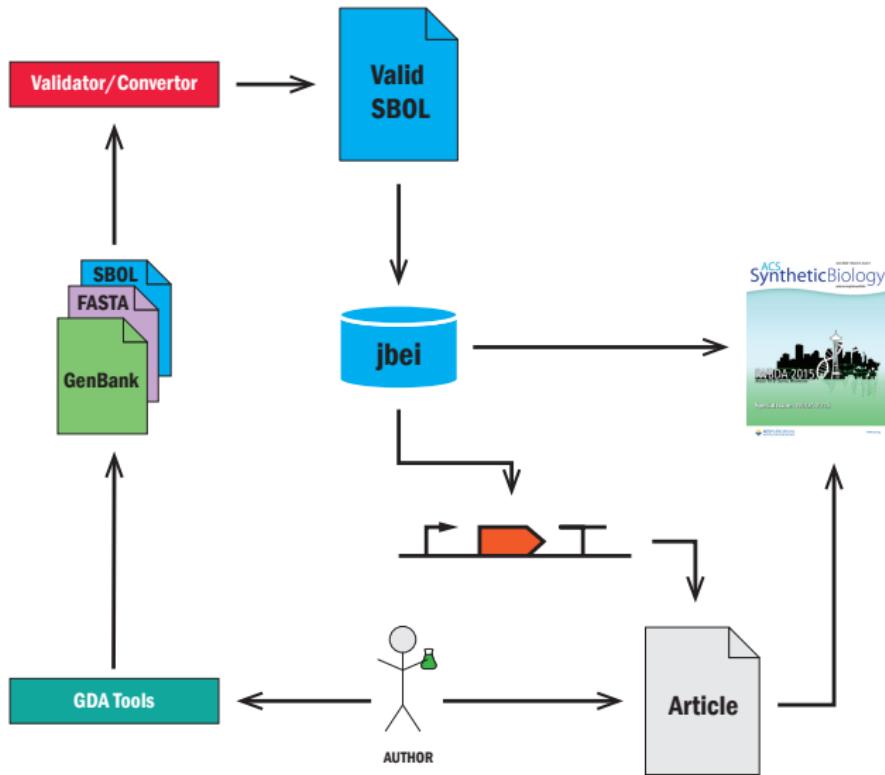
- Crucial to the success of a standard is software infrastructure to support developers' integration of the standard within their tools.
- There are several library implementations of the SBOL data structure, which provide an *application programmers interface* (API) for tool developers to interact with SBOL data objects.
 - libSBOLj - native Java library
 - libSBOL - C/C++ library
 - pySBOL - Python library
 - sboljs - Javascript library
- Library distributions include detailed documentation for the class definitions and the methods provided by the API.
- libSBOLj supports validation and conversion to/from FASTA, GenBank, and SBOL1.1.
- An online validator/converter is available from the SBOL website, and it also provides a webservice to be used by non-java applications.

Partial List of SBOL Compliant Software Tools

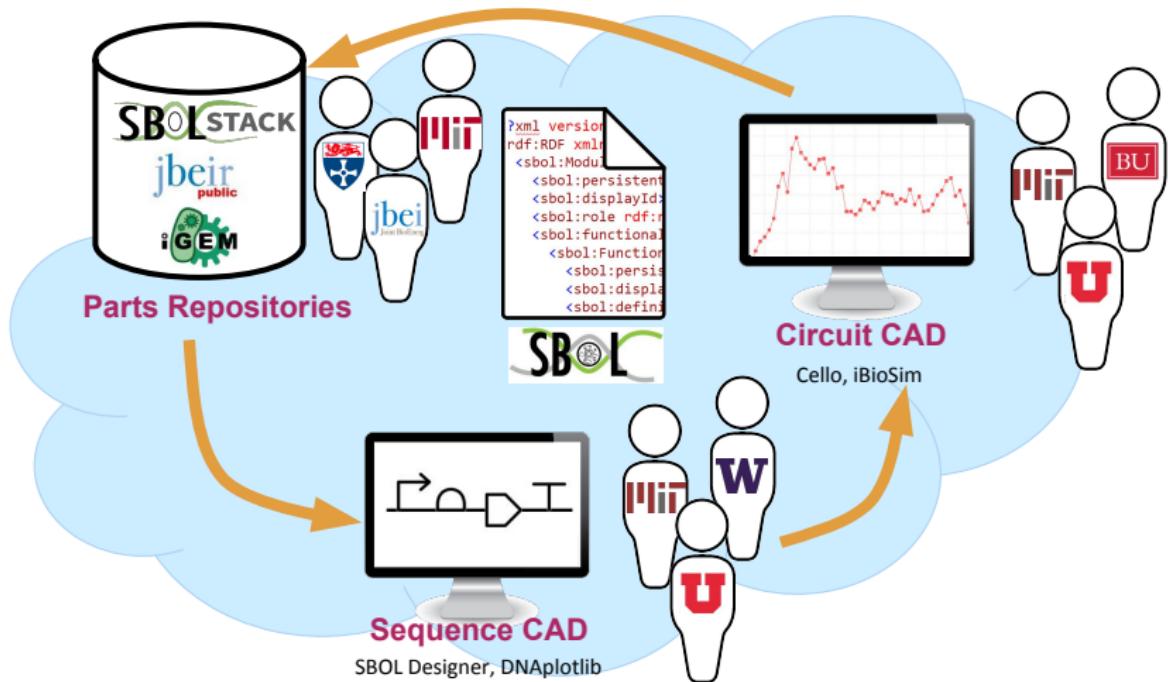
Name	Repository	Sequence	Circuit	Visual	Import	Export	Lossless
ICE	x			x	2.0	2.0	x
SBOL Stack	x				2.0	2.0	x
VirtualParts	x				2.0	2.0	x
Benchling		x				1.1	
DeviceEditor		x		x	1.1	1.1	
DNAPlotLib		x		x	1.1	1.1	
Eugene		x			1.1	1.1	
GeneGenie		x				2.0	
GenoCAD		x		x		1.1	
GraphViz		x		x			
GSL/Thumper		x			2.0	2.0	
j5		x			1.1	1.1	
Kera		x		x			
Pigeon		x		x			
Raven		x		x			
SBOLDesigner		x		x	2.0	2.0	x
TeselaGen		x		x	1.1	1.1	
Vector Editor		x			1.1	1.1	
VisBOL		x		x	2.0		
Cello			x	x		2.0	
iBioSim			x		2.0	2.0	x
MoSeC			x			1.1	
Proto BioCompiler			x	x		1.1	
SBROME			x		1.1	1.1	
TinkerCell			x	x	1.1	1.1	

WARNING:
Changing rapidly,
so likely not
completely
accurate.

ACS Synthetic Biology SBOL Workflow



Synthetic Biology Workflow Using SBOL



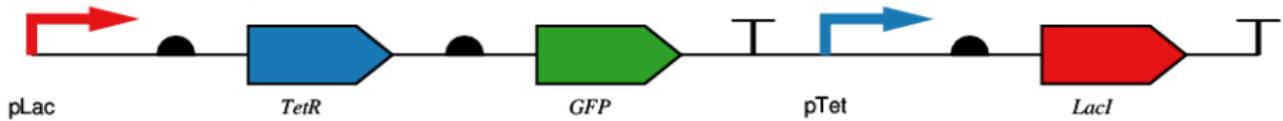
Data Repositories

- SBOL Stack (Newcastle University)
 - RDF database designed for storing SBOL data.
 - Provides RESTful API that can be queried with SPARQL queries.
 - Includes the full iGEM registry converted to SBOL format.
 - C. Madsen et al., ACS Synthetic Biology (2016) and COMBINE 2014.
- ICE (Joint BioEnergy Institute)
 - Web-based database for DNA parts.
 - While data model is GenBank-centric, provides SBOL import/export.
 - T. Ham et al., Nucleic Acid Research (2012).
- Learn more in Session 6, particularly Kuwahara's talk about SBOLme.
- Data should be paired with context information, discussion about extending SBOL to support this during Session 7.

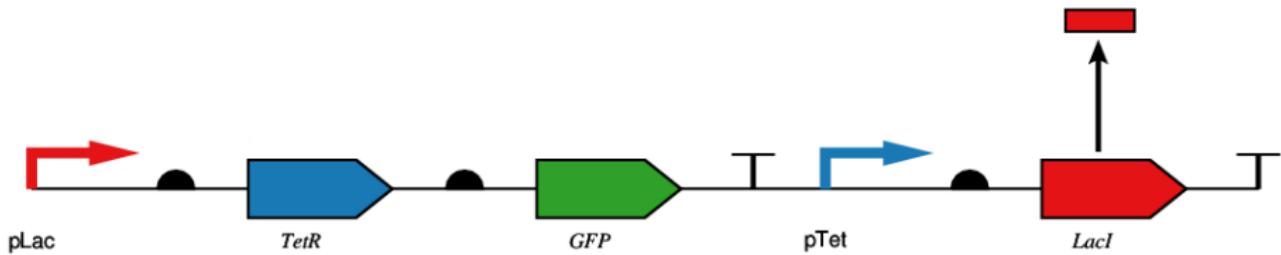
Sequence Editors

- Commercial sequence editors:
 - Benchling, Geneious, SnapGene, TeselaGen, Thumper, etc.
 - Either support SBOL 1.1/2.0 or can be used via GenBank conversion.
- Non-commercial sequence editors that support SBOL 1.1:
 - DeviceEditor, J5, VectorEditor (JBEI), DNAPlotLib (MIT), Eugene (Boston), GenoCAD (VBI), etc.
- Non-commercial sequence editors that support SBOL 2.0:
 - SBOLDesigner (Utah)

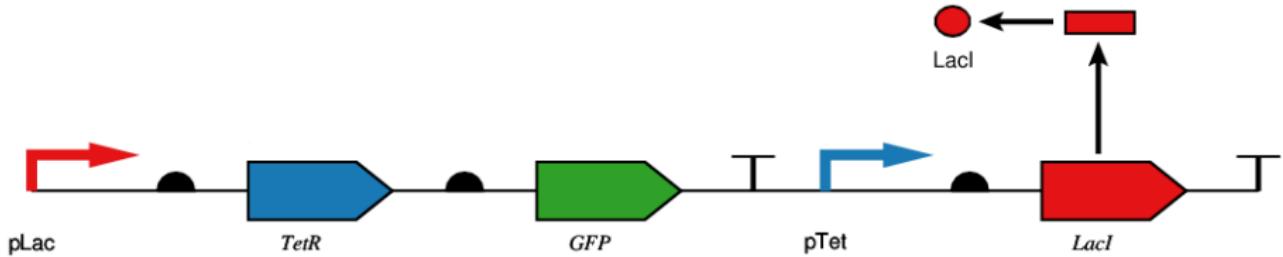
Genetic Toggle Switch (Gardner et al. 2000)



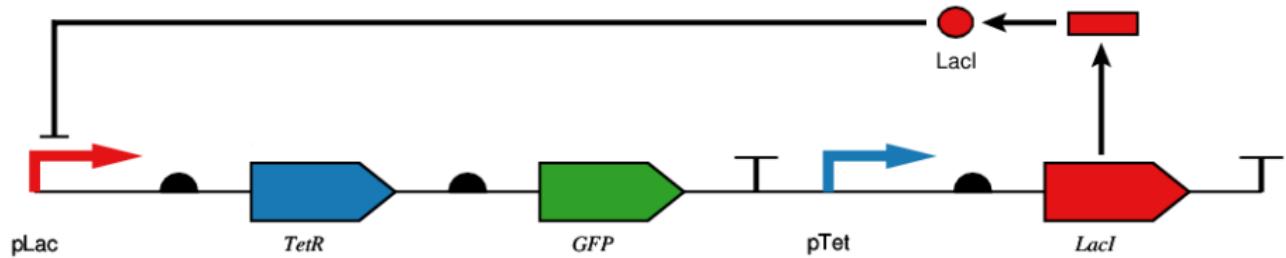
Genetic Toggle Switch (Gardner et al. 2000)



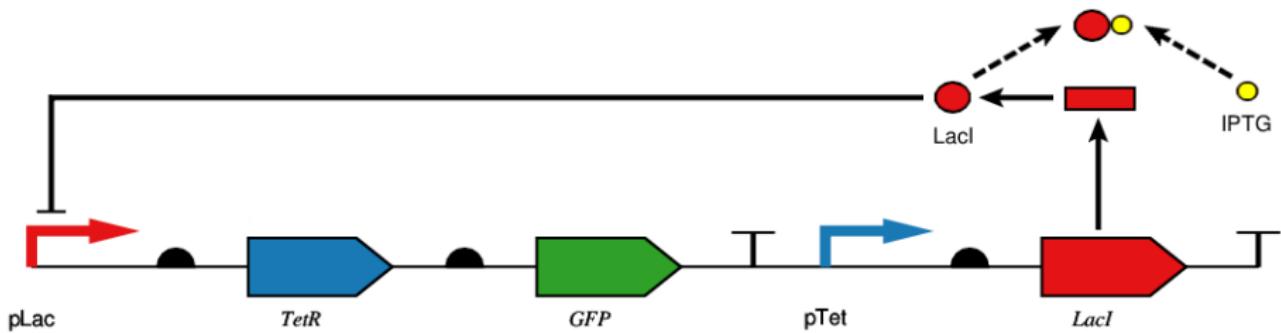
Genetic Toggle Switch (Gardner et al. 2000)



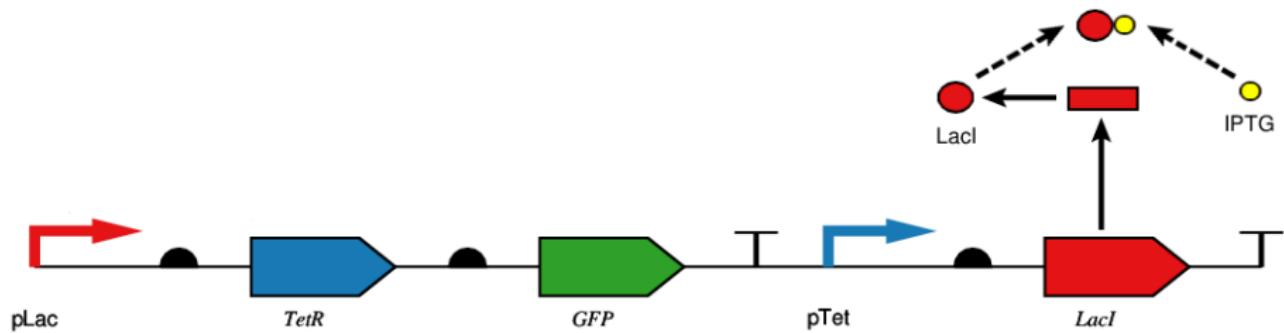
Genetic Toggle Switch (Gardner et al. 2000)



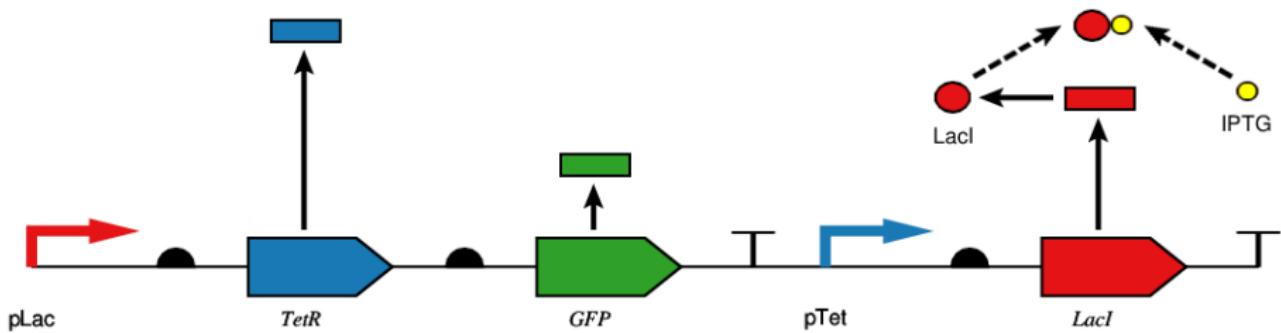
Genetic Toggle Switch (Gardner et al. 2000)



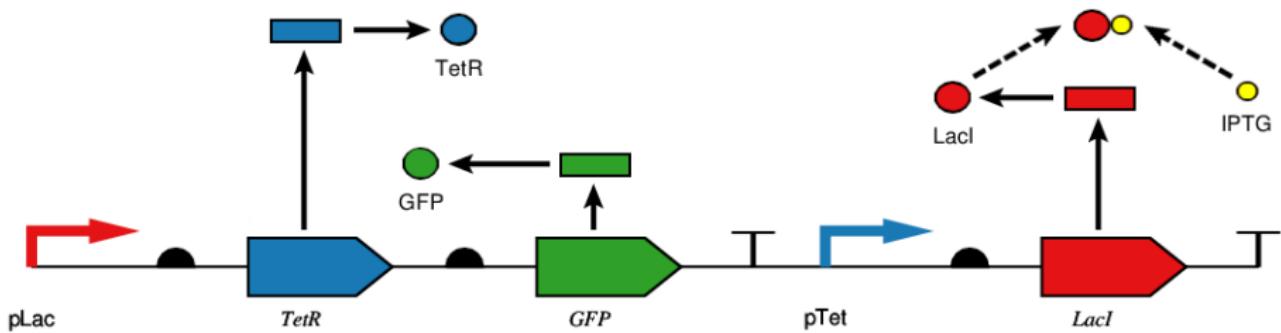
Genetic Toggle Switch (Gardner et al. 2000)



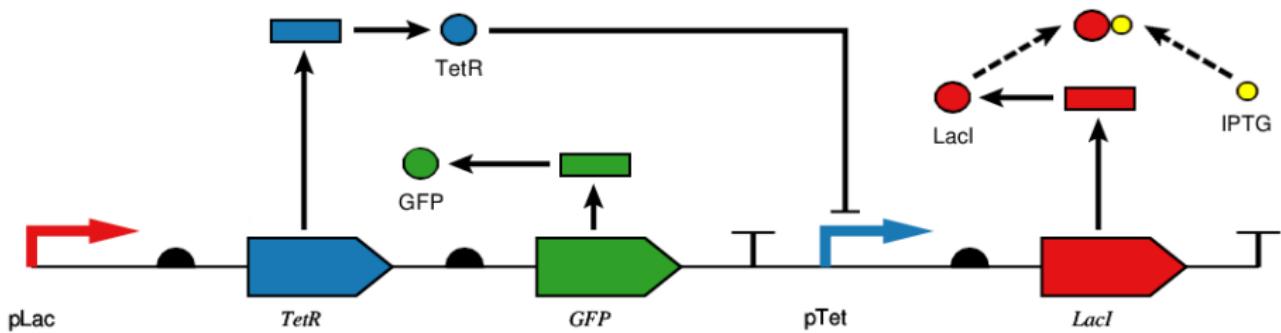
Genetic Toggle Switch (Gardner et al. 2000)



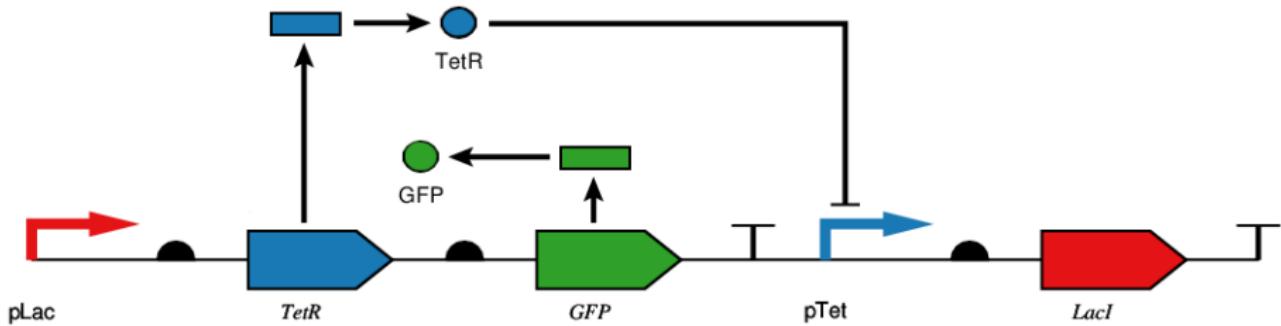
Genetic Toggle Switch (Gardner et al. 2000)



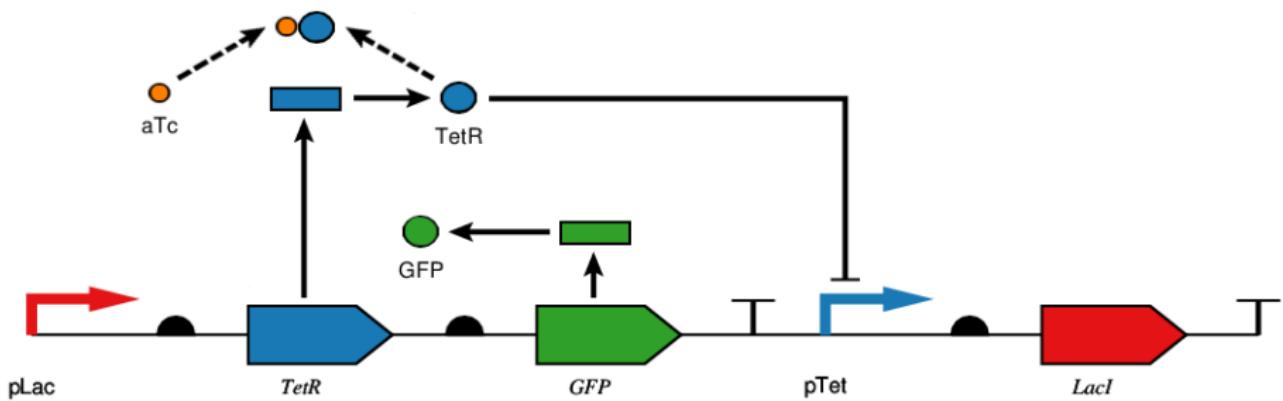
Genetic Toggle Switch (Gardner et al. 2000)



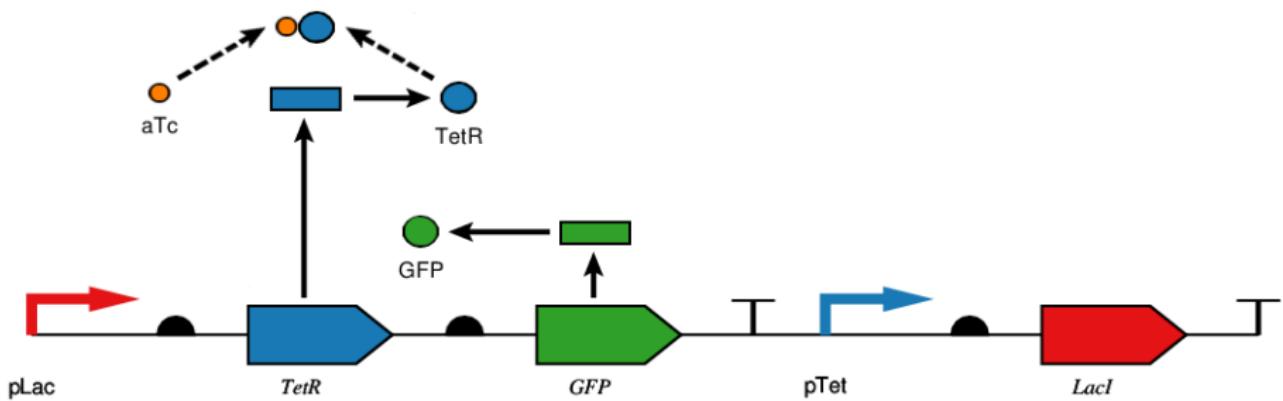
Genetic Toggle Switch (Gardner et al. 2000)



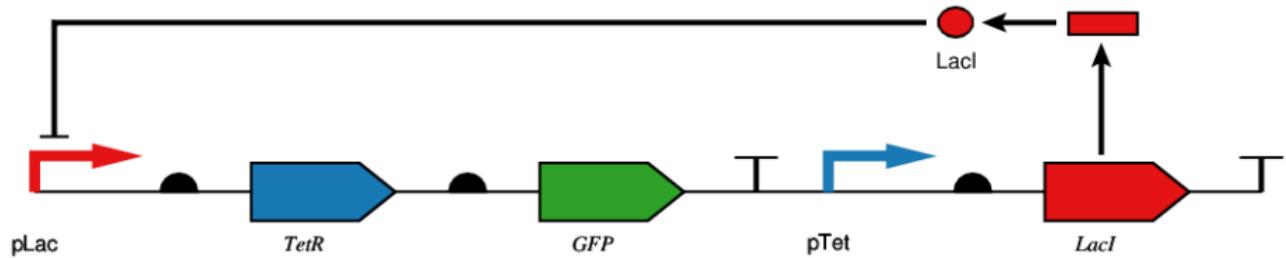
Genetic Toggle Switch (Gardner et al. 2000)



Genetic Toggle Switch (Gardner et al. 2000)



Genetic Toggle Switch (Gardner et al. 2000)

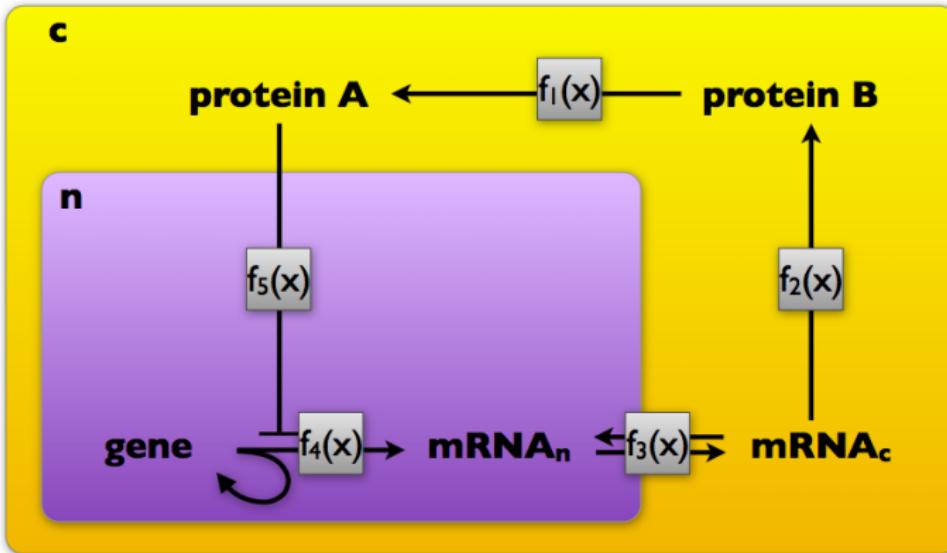


SBOL Stack / SBOLDDesigner Demo

Circuit Design Tools Supporting SBOL 2.0

- Cello (MIT) - Nielsen et al., Science (2016)
 - Compiles combinational designs described in Verilog into genetic circuit.
 - Uses a gate library of 2-input NOR gates and inverters.
 - Ensures good signal matching characteristics.
 - Designs were validated in the laboratory.
 - Supports export of designs using SBOL 2.0.
- iBioSim (Utah) - Madsen et al., IEEE Design & Test (2012)
 - Supports modeling, analysis, and design of genetic circuits.
 - Modeling uses SBML, analysis uses SED-ML, and design uses SBOL.
 - Includes SBOLDesigner as plugin providing a connection to SBOL Stack.
 - Supports import/export for SBML L3V1, SED-ML L1V2, and SBOL 2.0.
 - Provides converters between SBML and SBOL.

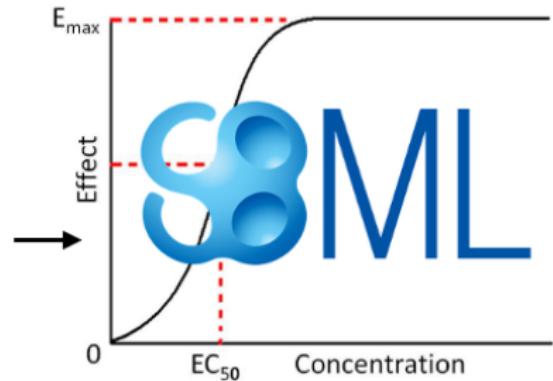
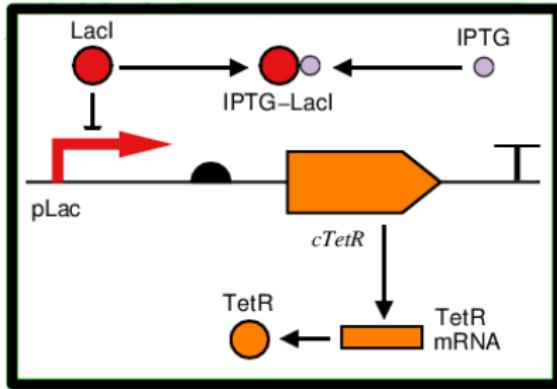
Systems Biology Markup Language (SBML)



- SBML breakouts for L3V2 (Session 8), extended math (Session 11), agent-based modeling (Session 12), qualitative models (Session 15), and multi-state models (Session 16).

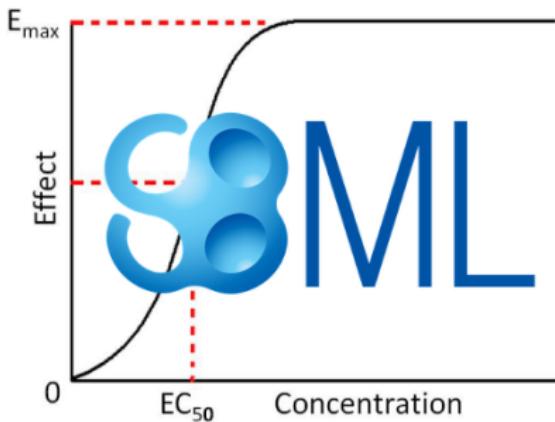
Converting SBOL to SBML

LacI Inverter

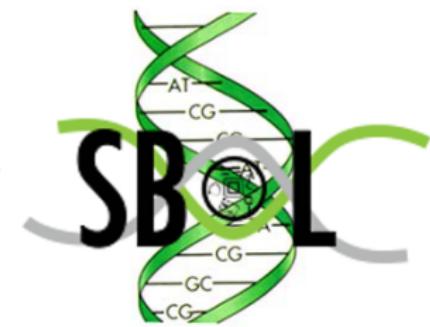


Roehner et al., ACS Synthetic Biology (2014)

Connecting SBML to SBOL



RDF/XML
Annotation



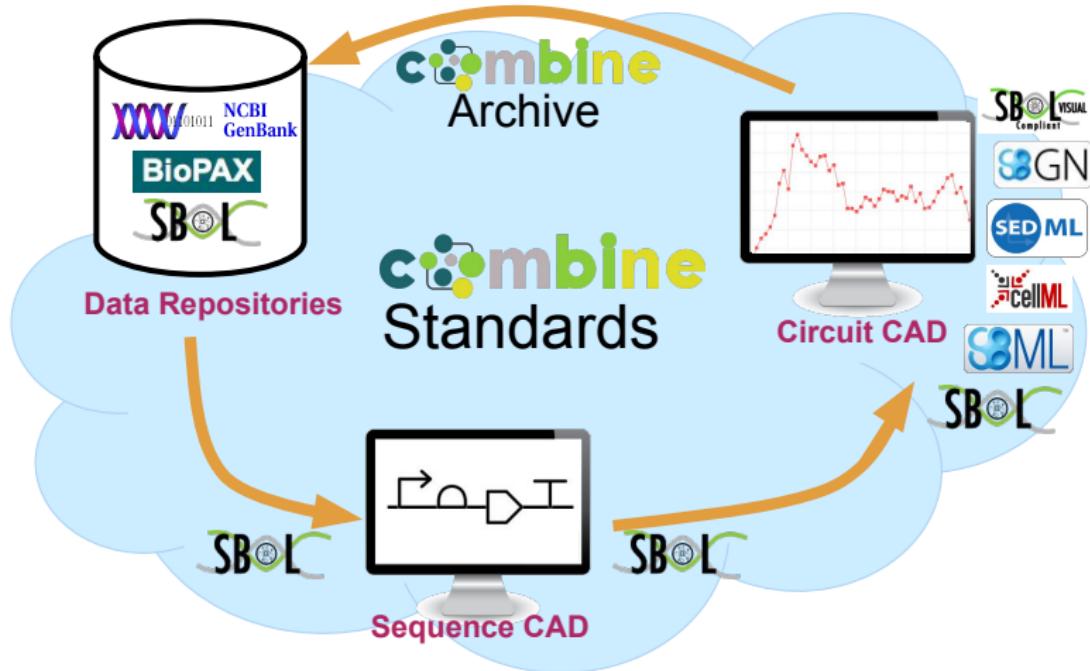
Roehner et al., ACS Synthetic Biology (2013).

Ontologies and Annotations

- Critical to conversion is consistent use of ontologies and annotations.
- Ontologies used by SBOL:
 - BioPAX Ontology for ComponentDefinition types.
 - Sequence Ontology (SO) for DNA ComponentDefinition roles.
 - Systems Biology Ontology (SBO) for Interaction types, Participation roles, and Model frameworks.
 - EMBRACE Data and Methods (EDAM) Ontology for Model languages.
- URIs are taken from identifiers.org, whenever possible.
- SBOL supports custom annotations for arbitrary RDF data that is not supported by existing fields within the data model.
- Session 12's SBOL breakout to discuss ontologies/URI best practices.
- Session 16's SBOL breakout to discuss use of SBO for interactions.
- Sessions 19 and 20 will discuss COMBINEd annotations.

iBioSim / Validator / ICE Demos

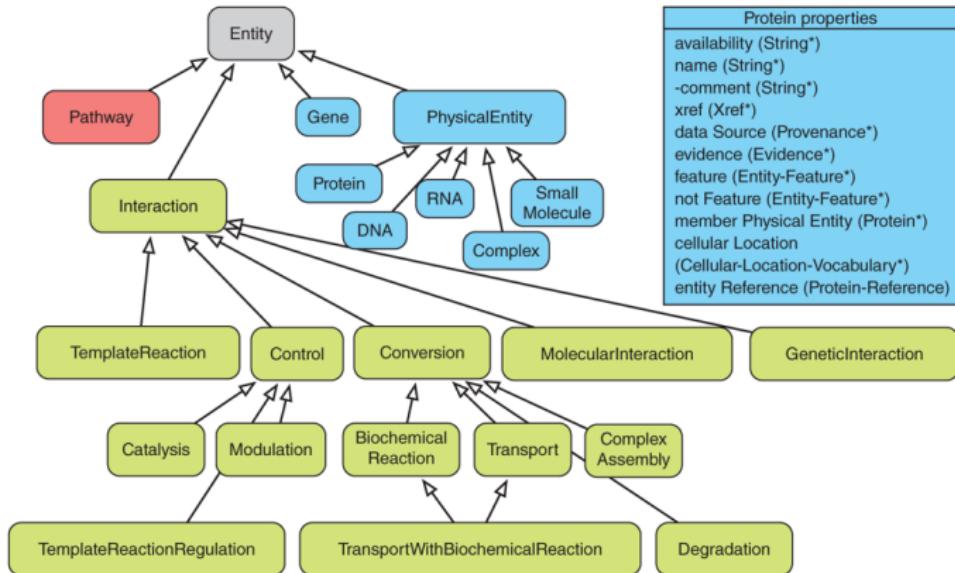
A Standard Enabled Workflow for Synthetic Biology



Fundamental Elements of a Standard Enabled Workflow

- Rich data repositories
- Consistent annotations
- Lossless data conversions
- Seemless connection of tools
- Intuitive visualizations

Data from BioPAX Repositories



- See Luna's talk in Session 17 for more about BioPAX.
- See Nguyen's talk in Session 17 to learn about a new BioPAX/SBML converter.

BioPAX Data from EcoCyc



BioCyc to Adopt
Subscription Model

Welcome, Chris

Enter a gene, protein, metabolite or pathway...

Searching *Escherichia coli* K-12 substr. MG1655 (EcoCyc)

Sites ▾ Search ▾ Genome ▾ Metabolism ▾ Analysis ▾ SmartTables ▾ Help ▾

gene polypeptide
lacI LacI DNA-binding transcriptional repressor

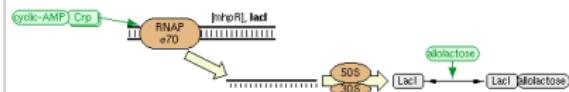
Add to SmartTable Advanced Query Results

Add

Synonym	transcriptional repressor of the lac operon
Accession IDs	EG10525 (EcoCyc) b0345 ECK0342 P03023 (UniProt)
Length	1083 bp / 360 aa
Map Position	[366,428 < 367,510] (7.89 centisomes, 28°)
Location	cytosol
Reaction	LacI + allolactose ↔ LacI-allolactose
Evidence	Inferred from experiment [2]

Summary GO Terms (8) Essentiality Reactions (1) Regulon Protein Features Operons References Show All

Regulation Summary Diagram



Summary

The Lactose Inhibitor, LacI, is a DNA-binding transcription factor that represses transcription of the operon involved in transport and catabolism of lactose [Hudson90, Vossen96, Lewis05]. In the absence of allolactose, LacI represses the lac operon by preventing open promoter complex formation for transcription [Hudson90, Sanchez11]. In this repression system, LacI binds to two operators, and formation of the repressor loop is critical [Fried96, Perros96, Balaef04, Hudson90]. This repressor binds in tandem to inverted repeat sequences that are 21 nucleotides long and possess conserved motifs [Hudson90, Fried96, Perros96].

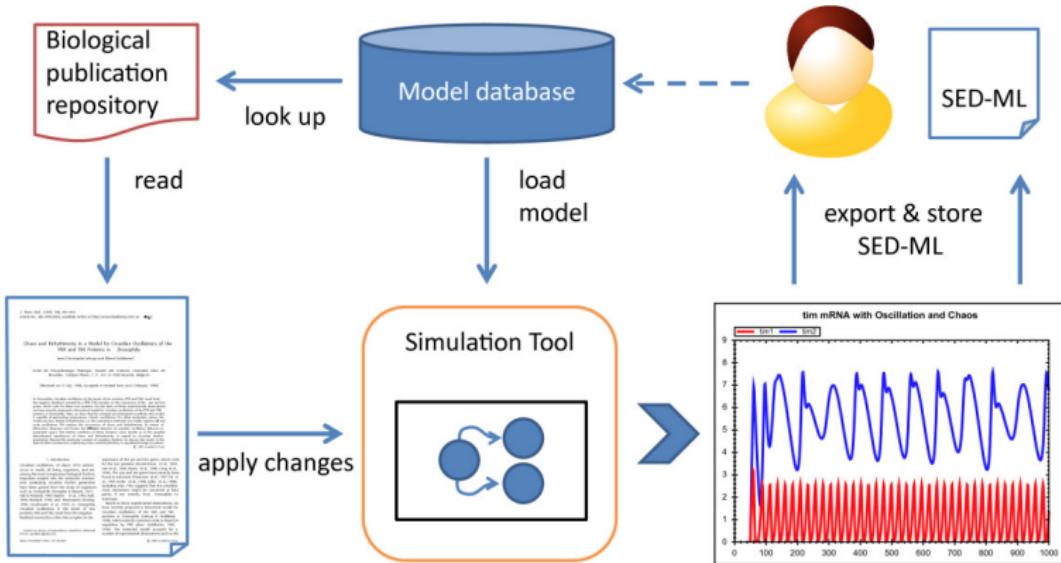
Laci is negatively autoregulated when it binds to two DNA-binding sites, one located downstream of the lac gene and the other one located in the coding sequence for the C terminus of LacI. The protein when bound to these sites forms a loop that inhibits the transcription elongation, thus producing truncated proteins that are targeted for degradation by the small peptide SsrA [Abo00, Semsey13].

Induction occurs when the physiological inducer, allolactose, binds to the lac repressor, preventing it from binding to the operator [Hudson90, Lewis05]. Nonphysiological analogs, such as thiogalactosides, can function as inducers too. A single mutation, W220F, in the inducer-binding site of the LacI repressor

Unification Links

ASAP	ABE-0001189
CGSC	578
DIP	DIP-10079N
DisProt	DP00433
EchoBASE	EB0520
EcoGene	EG10525
EcollWiki	b0345
Mint	MINT-6478062
ModBase	P03023
OU	b0345

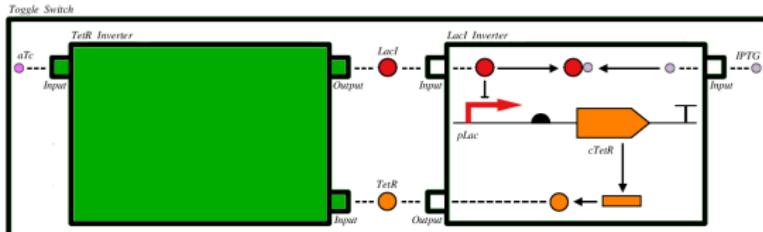
Simulation and SED-ML



- Simulation talks in Sessions 5, 10, 13, and 14.
- SED-ML breakouts to discuss L1V3 (Session 7), qualitative analysis (Session 15), and next steps (Session 19).

SED-ML Demo

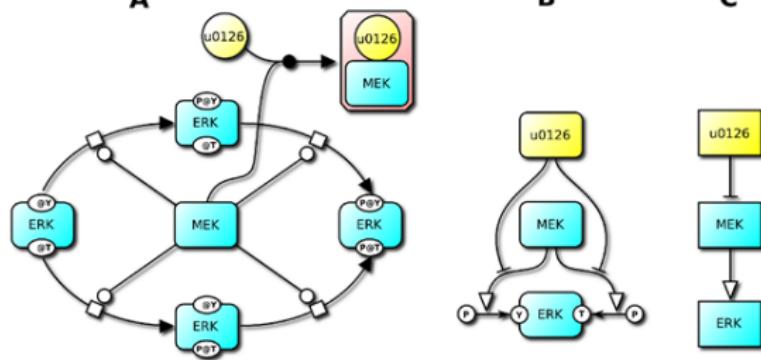
Visualization Standards: SBOL Visual and SBGN



A

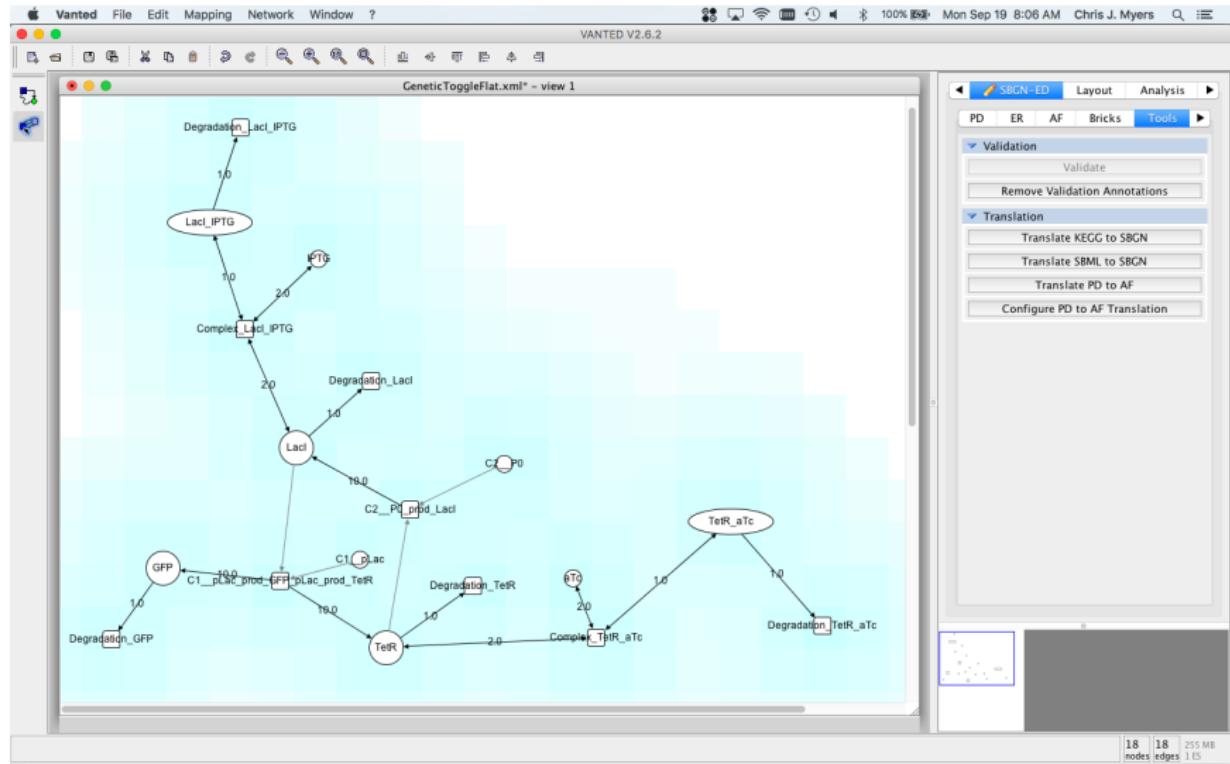
B

C

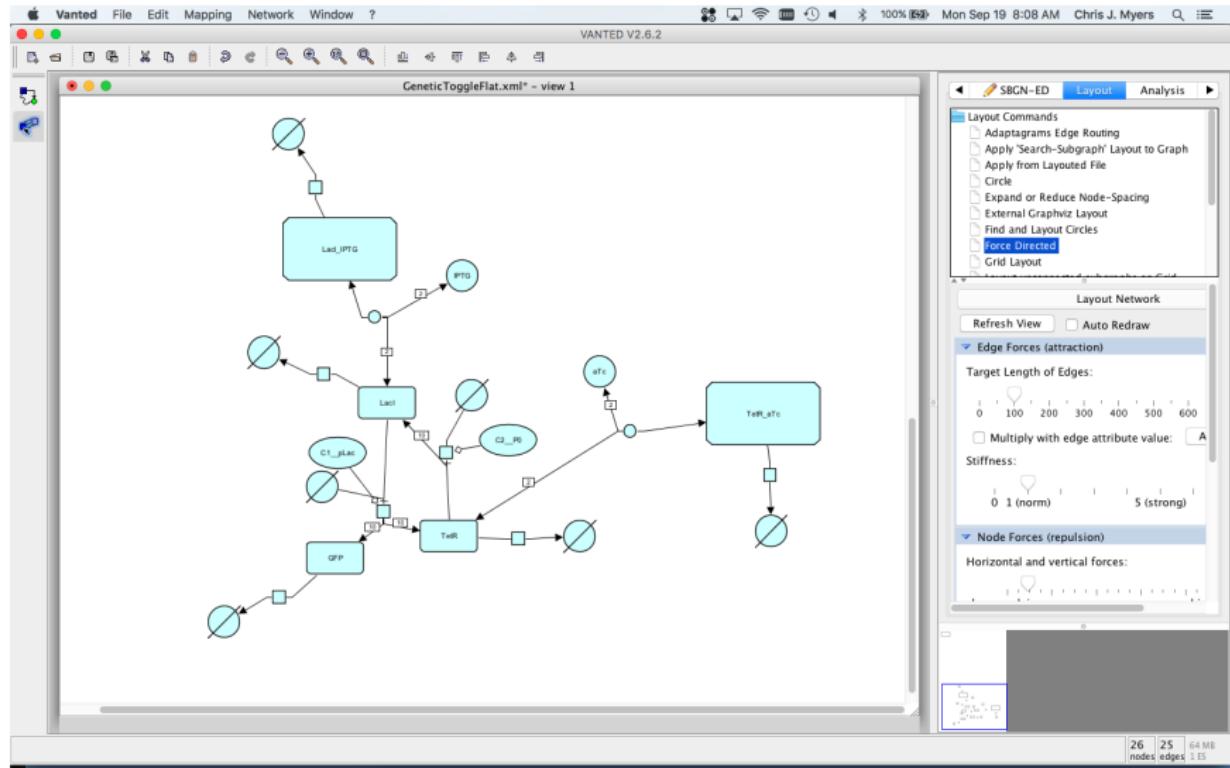


- Talks on visualization in Session 18.
- Session 11 breakout to discuss relationship b/w SBOLv 2.0 and SBGN.
- SBGN Workshop on Saturday, still space, please register, its free.

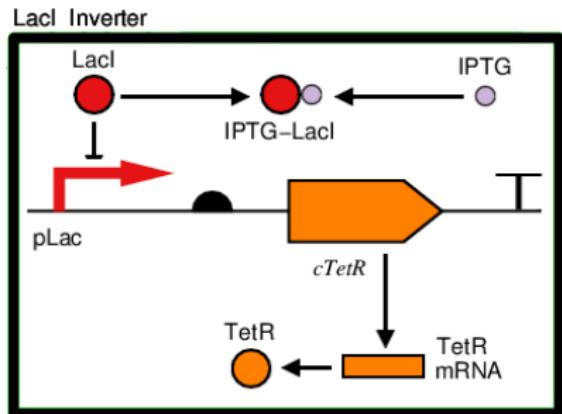
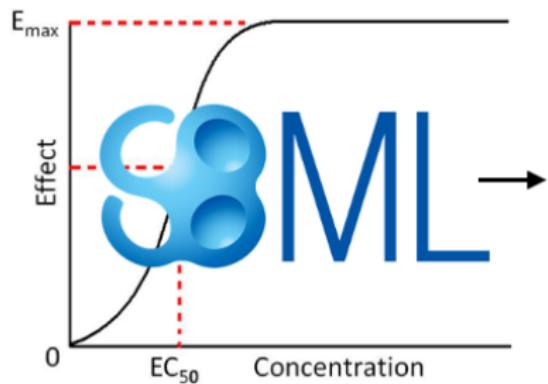
SBGN Visualization Using SBGN-ED



SBGN Visualization Using SBGN-ED

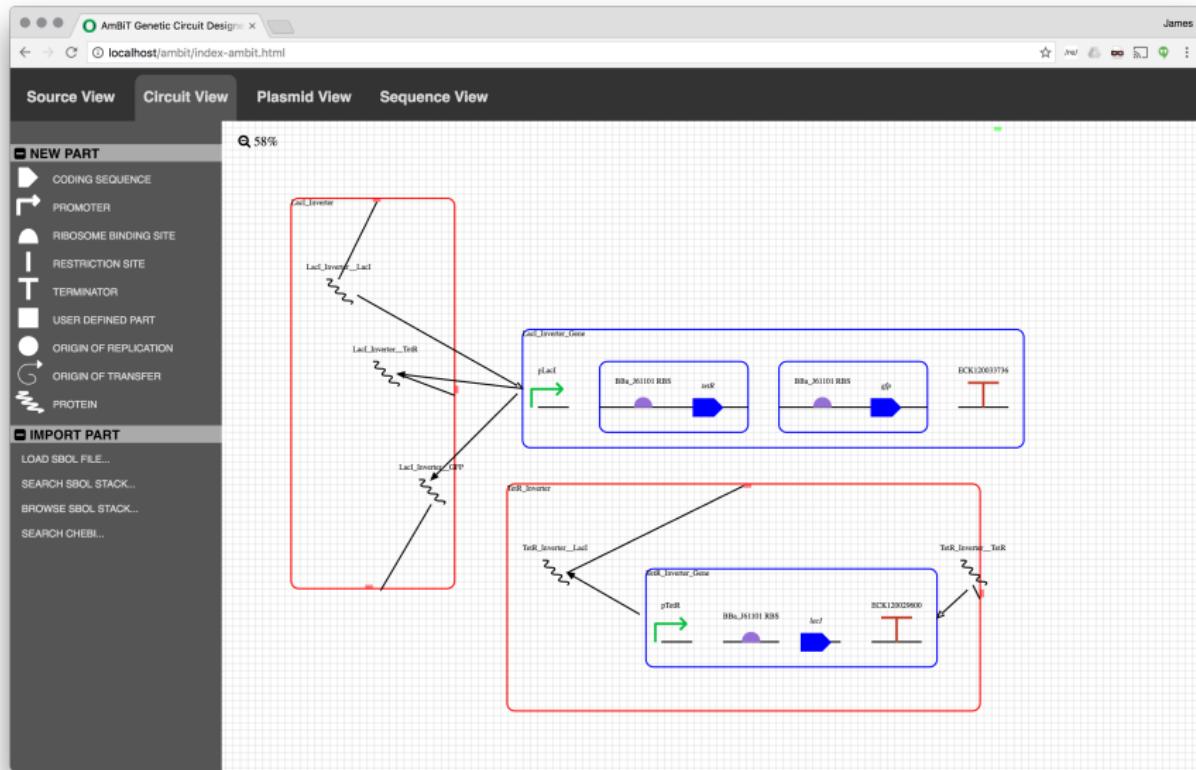


Converting SBML to SBOL



Nguyen et al., ACS Synthetic Biology (2016)

SBOLv Visualization Using AmBiT



Conclusion

- Standards are an important enabler for data sharing and reproducibility in synthetic biology.
- Collaborations within the COMBINE community are essential.
- Ultimate goal should be a complete standard-enabled workflow for synthetic biology.
- More information about SBOL: <http://www.sbolstandard.org/>.