Synthetic Biology Open Language (SBOL)

Chris J. Myers and the SBOL Community

IWBDA

August 19, 2015

Synthetic Biology

- *Synthetic biology* extends *genetic engineering* by adding:
  - *Standards*
  - *Abstraction*
  - *Decoupling*

(source: Drew Endy)

Synthetic Biology

- *Synthetic biology* extends *genetic engineering* by adding:
  - *Standards*
  - *Abstraction*
  - *Decoupling*

(source: Drew Endy)

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).

© 2011 Nature America, Inc. All rights reserved.

Synthetic Biology Open Language (SBOL)

- In 2008, a small group of researchers proposed the development of the *synthetic biology open language* (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 (to appear)
- This month SBOL Version 2.0 was officially released.
  See Nicholas Roehner's talk on Wednesday.

SBOL Community

- 100+ people from all around the world.
- 30 universities, 14 companies, 8 other types of organizations.

Organizations Supporting SBOL

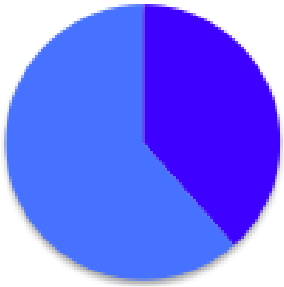
Current support for the development of SBOL provided by National Science Foundation Grants DBI-1356041 and DBI-1355909, and the Engineering and Physical Sciences Research Council under Grant Number EP/J02175X/1.

SBOL Governance

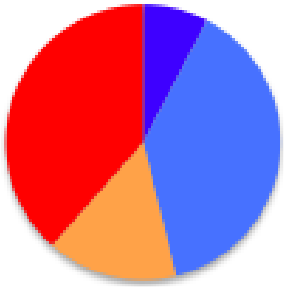
- SBOL Chair - Herbert Sauro (Washington)
- SBOL Editors - Bryan Bartley (Washington), Jacob Beal (Raytheon/BBN), Kevin Clancy (Life Technologies), Raik Grunberg (Montreal), and Goksel Misirli (Newcastle).
- Past SBOL Editors - Michal Galdzicki (Washington), Ernst Oberortner (DOE JGI), Matthew Pocock (Newcastle), Jacqueline Quinn (Google), Cesar Rodriguez (Autodesk), Nicholas Roehner (Boston), and Mandy Wilson (VBI).

SBOL/IWBDA Feedback

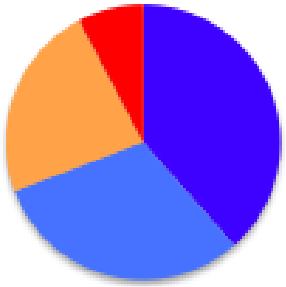
SBOL 2.0



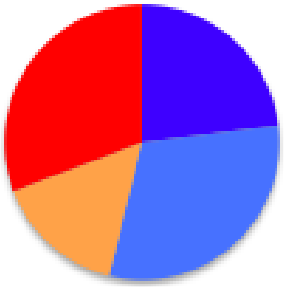
SBOL 1.0



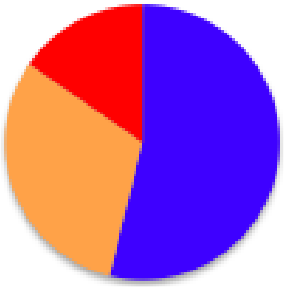
libSBOL



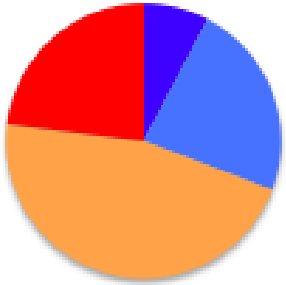
Hands-on



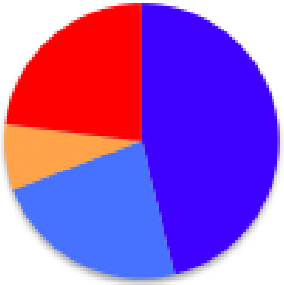
SBOLv



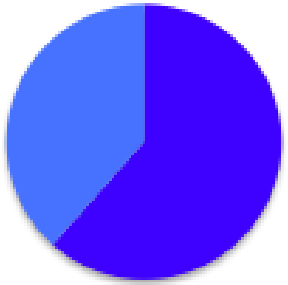
Upgrading



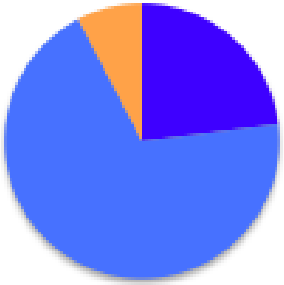
Community



Bench input



SBOL tools

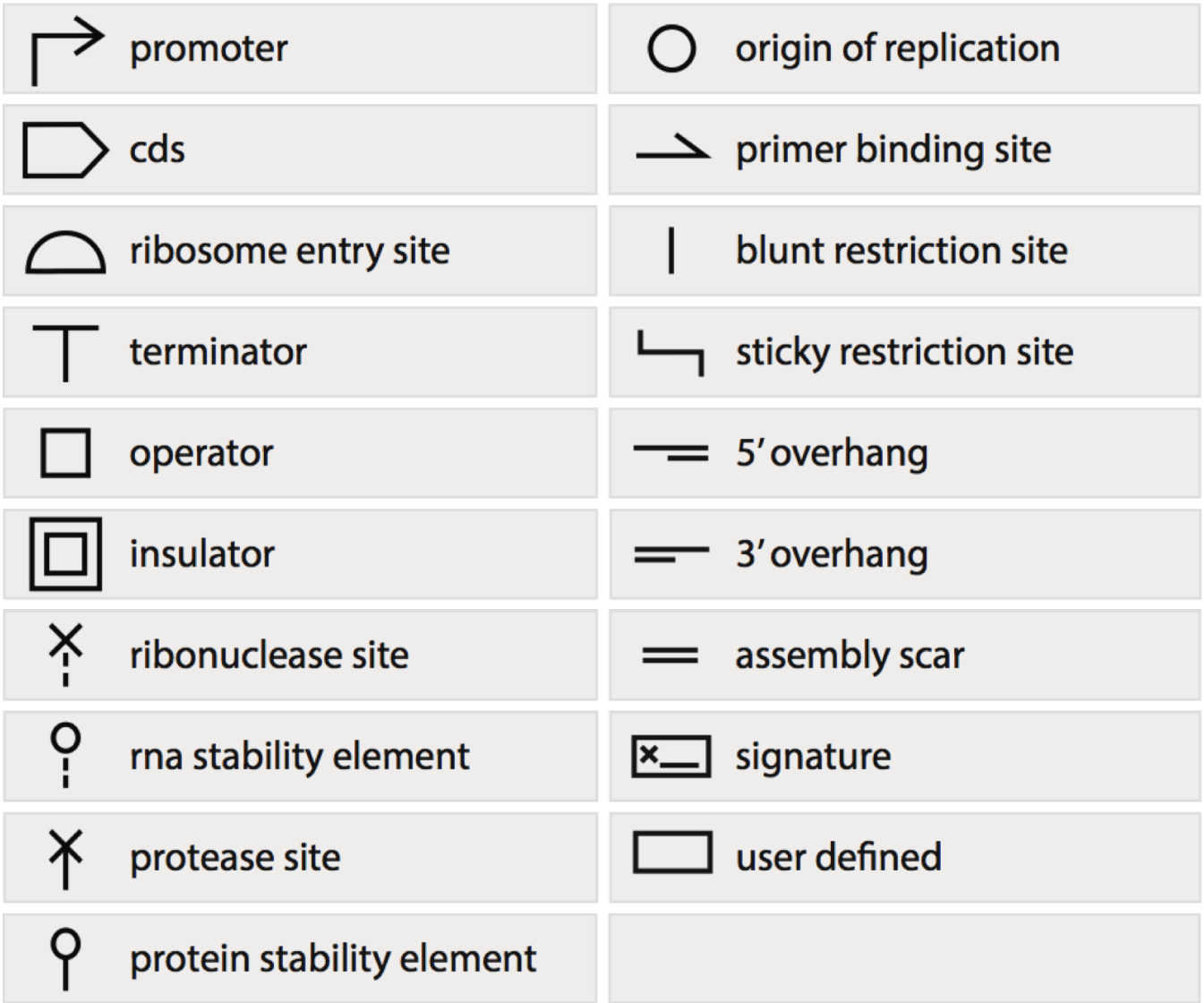


8/5, n=13

Schedule

- SBOL 2.0 - this talk
- SBOL tools - upcoming demonstrations
- Bench input - afternoon breakouts
- Community - discussion this evening
- libSBOL - informal meetings throughout IWBDA with libSBOL-team:
  - libSBOLc/pySBOL - Bryan Bartley, Jacob Beal, Herbert Sauro
  - libSBOLj - Tramy Nguyen, Nicholas Roehner, Chris Myers

SBOL Visual (Version 1.0)

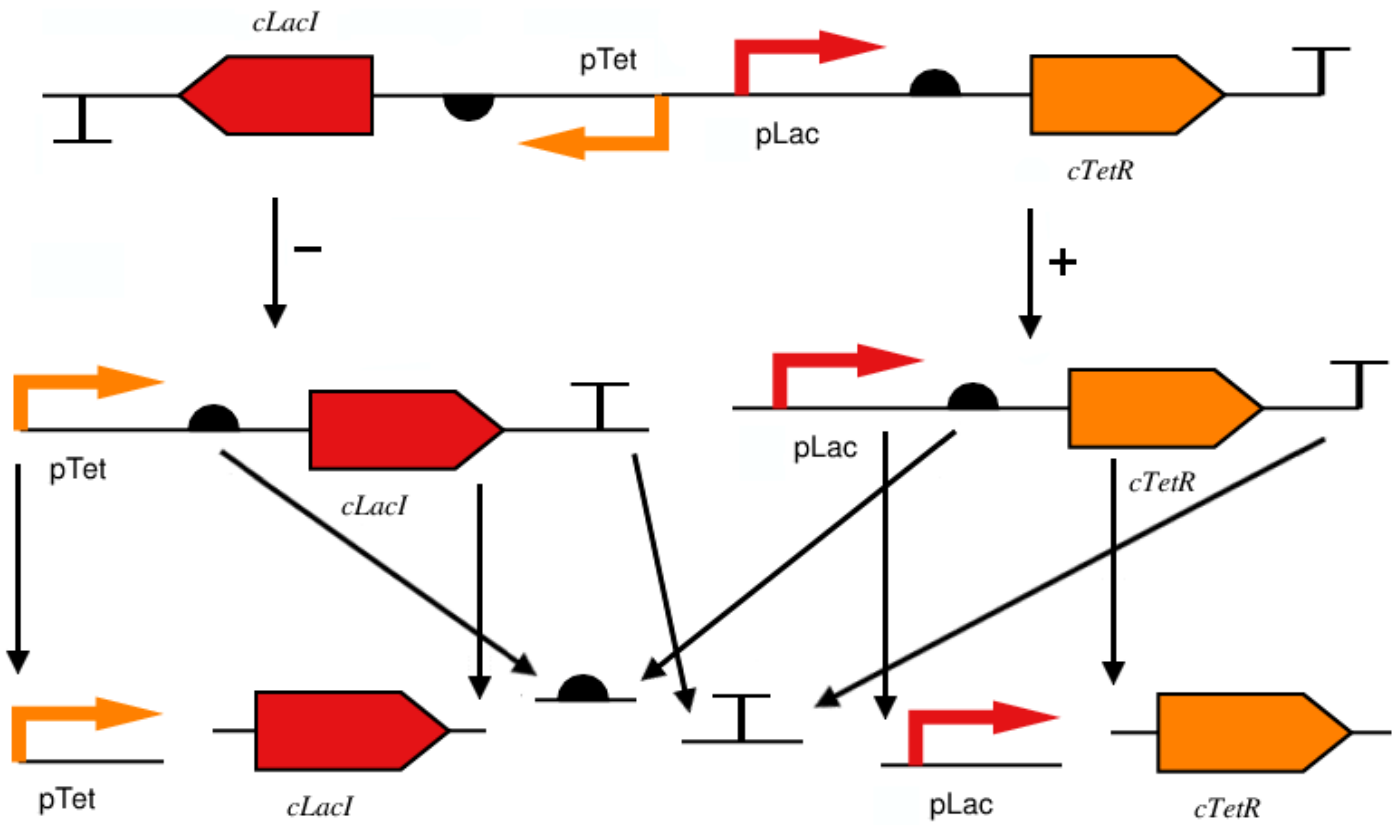


Quinn et al., PLOS Biology (to appear)

New symbols added on community consensus.

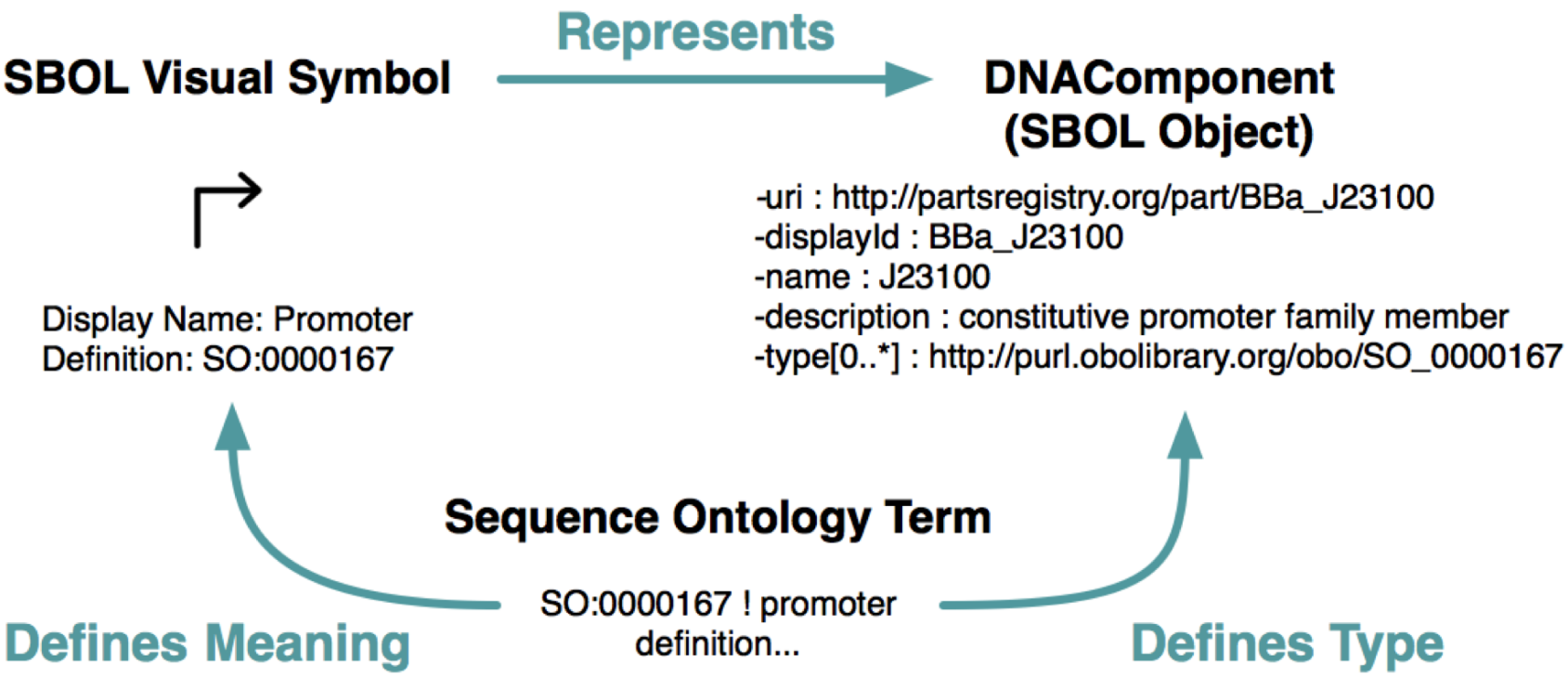
SBOL Data Model (Version 1.1)

- Version 1.1 specifies the hierarchical composition of *DNA components*.

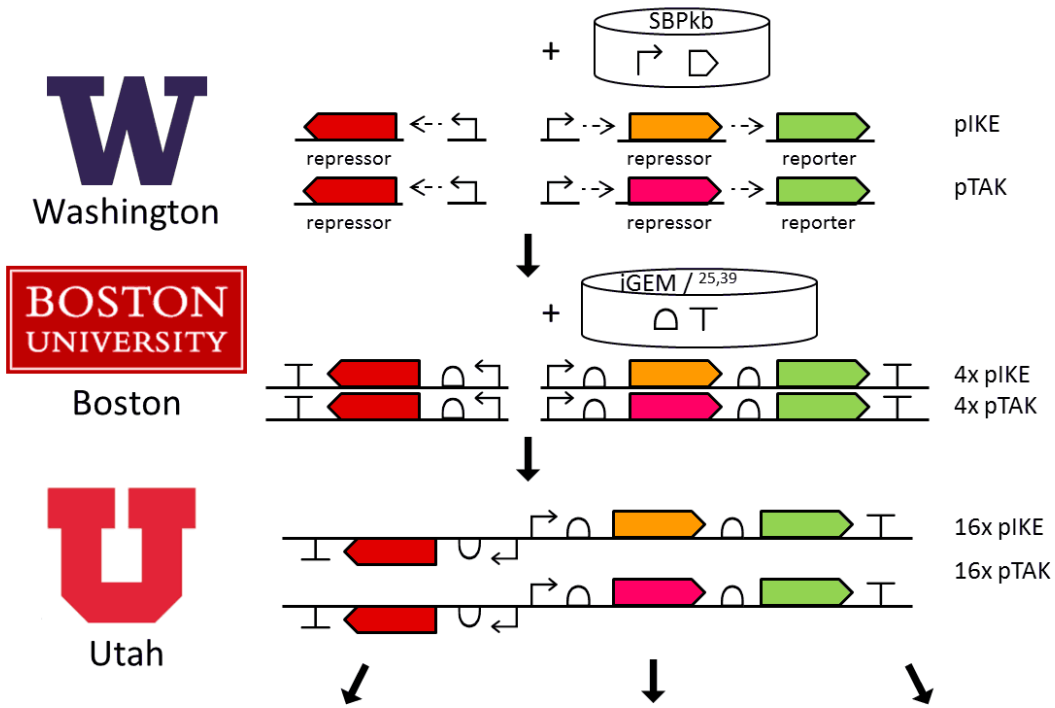


Galdzicki et al., Nature Biotechnology (2014).

Connection Between Visual and Data Model



SBOL Demonstration



Washington

Designed four expression cassettes, leaving RBS and terminator components unspecified.

Boston

Generated four variant designs of each cassette using six RBS and six terminator components.

Utah

Assembled and modeled sixteen variant designs of the toggle switch.

Life Technologies

Received design file for codon optimization and gene synthesis.

Newcastle

Stored design and model files in repository for dissemination.

Joint BioEnergy Institute

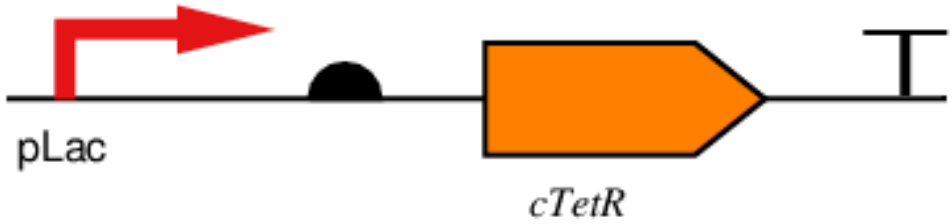
Stored design files in repository for dissemination.

Galdzicki et al., Nature Biotechnology (2014)

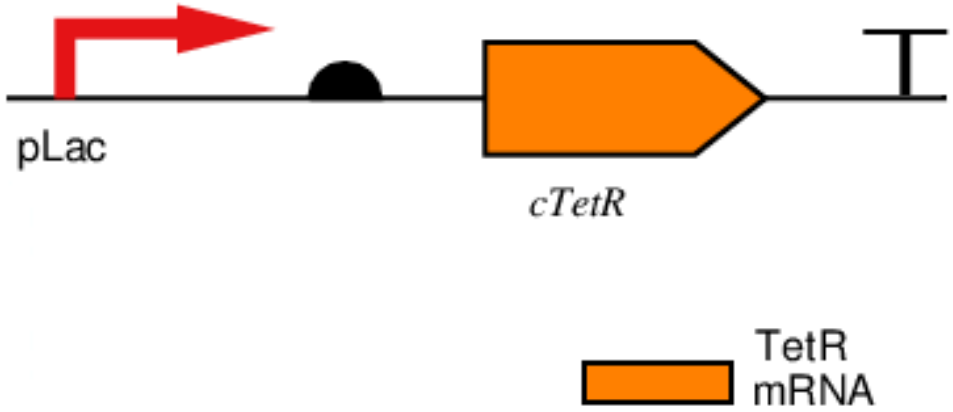
SBOL Data Model (Version 2.0)

- Proposed in Roehner et al., *ACS Synthetic Biology* (2014).
- Final specification released earlier this month.
  Available at `http://www.sbolstandard.org/`
- Provides the following new features:
  - *Generalized components* (DNA, RNA, protein, small molecule, etc.).
  - *Interactions* - qualitative relationships between components.
  - *Modules* - hierarchical groupings of components/interactions.
  - *Models* - means to connect modules to external models.
  - *Annotations* - means to encode special purpose information.

Generalized Components

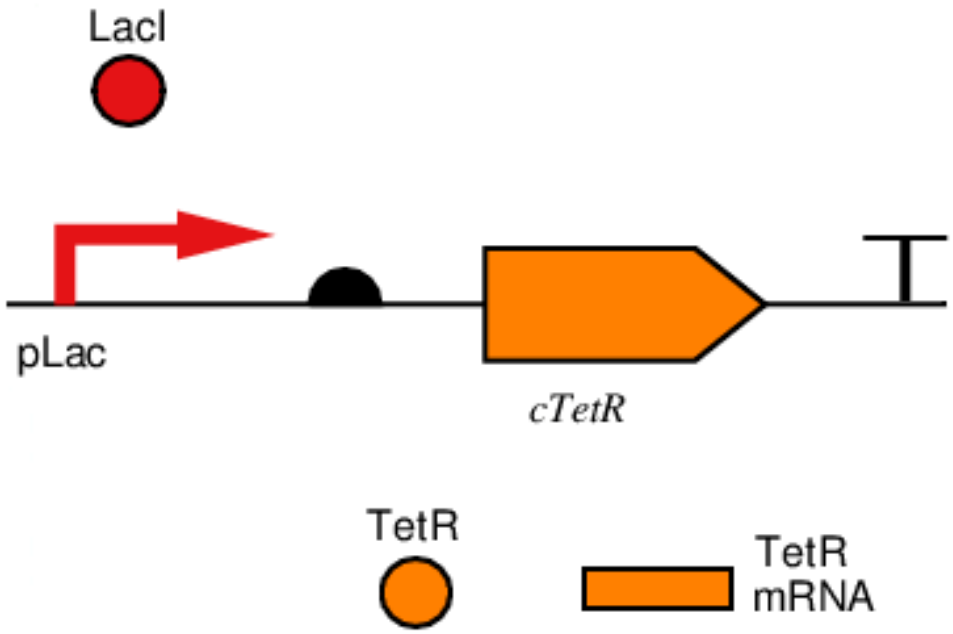


Generalized Components



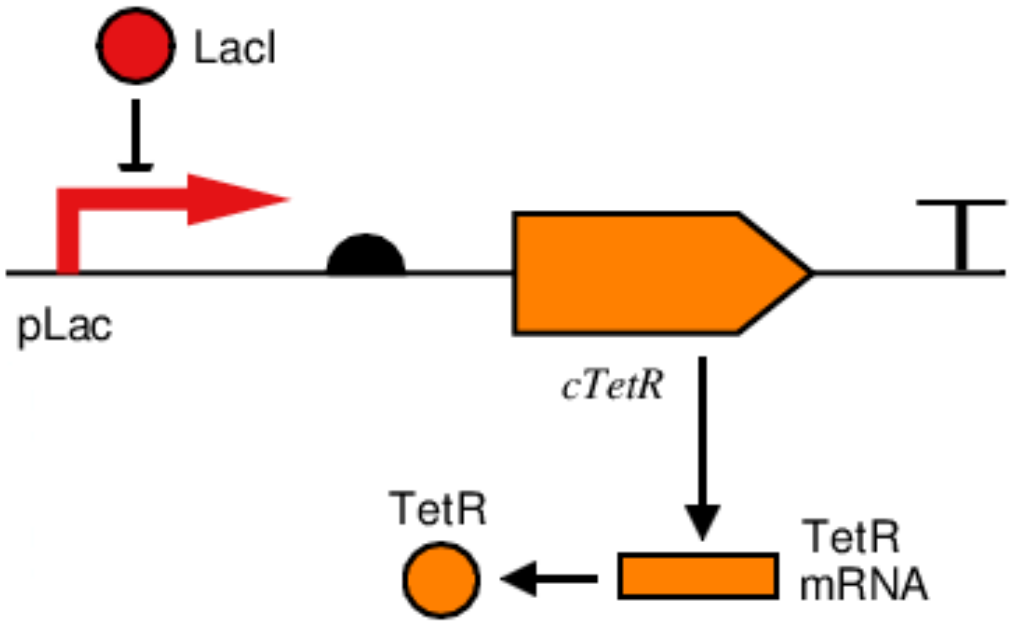
- RNA components (mRNA, tRNA, siRNA)

Generalized Components



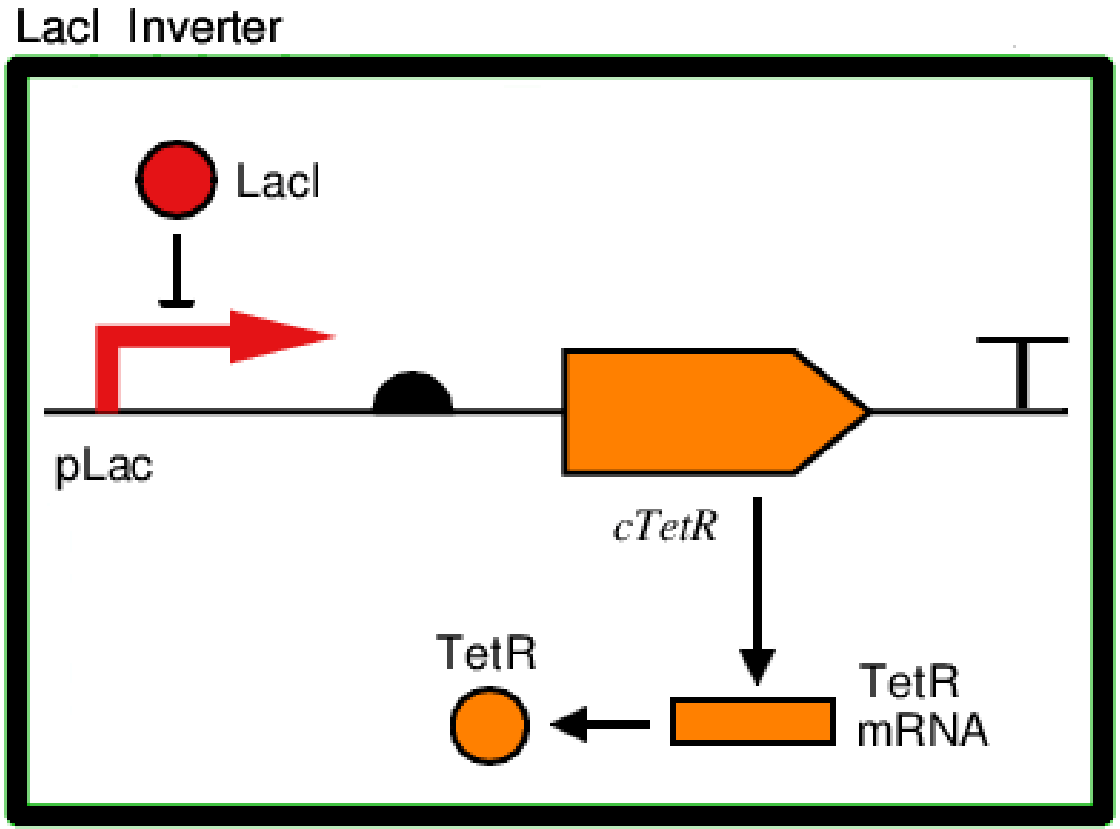
- RNA components
- Protein components (TFs, enzymes)

Interactions



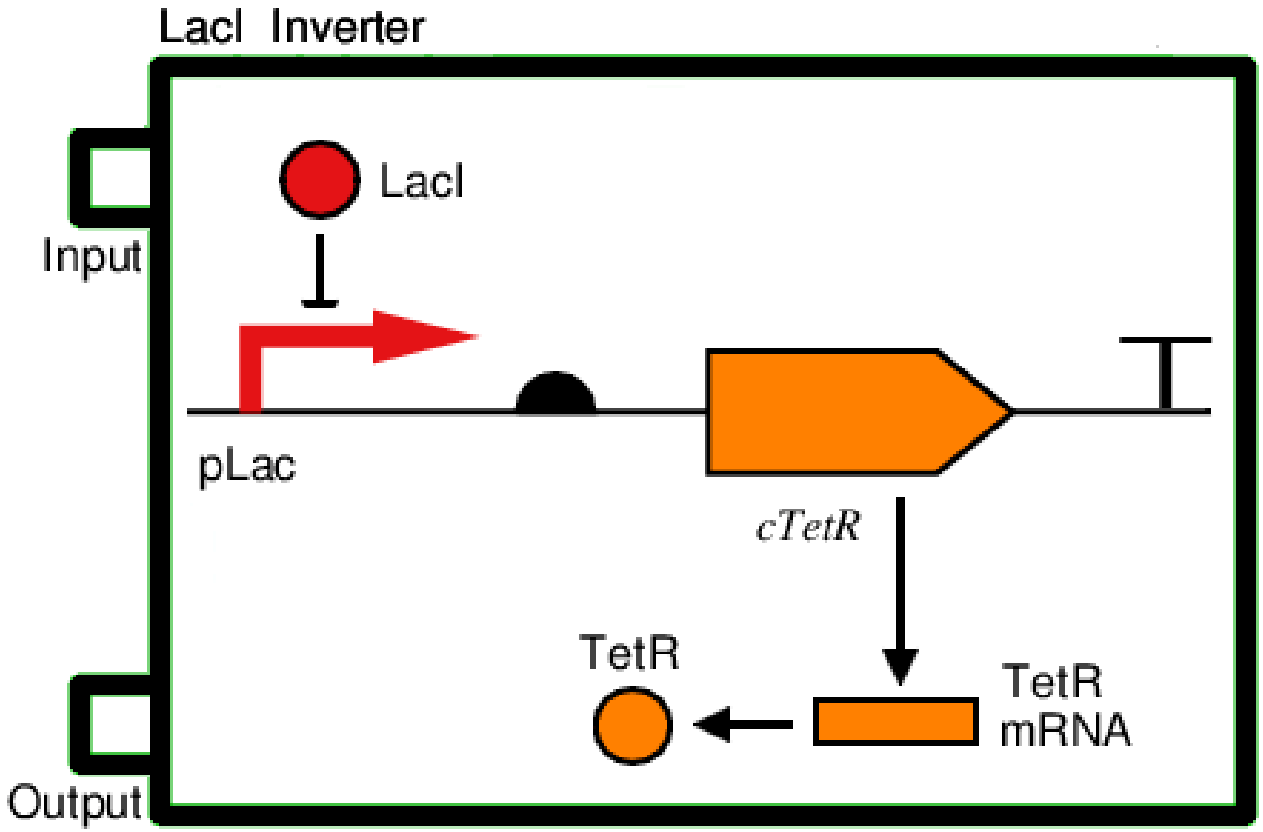
- Interactions (activation, repression, complexation, transcription, translation, phosphorylation)

Modules



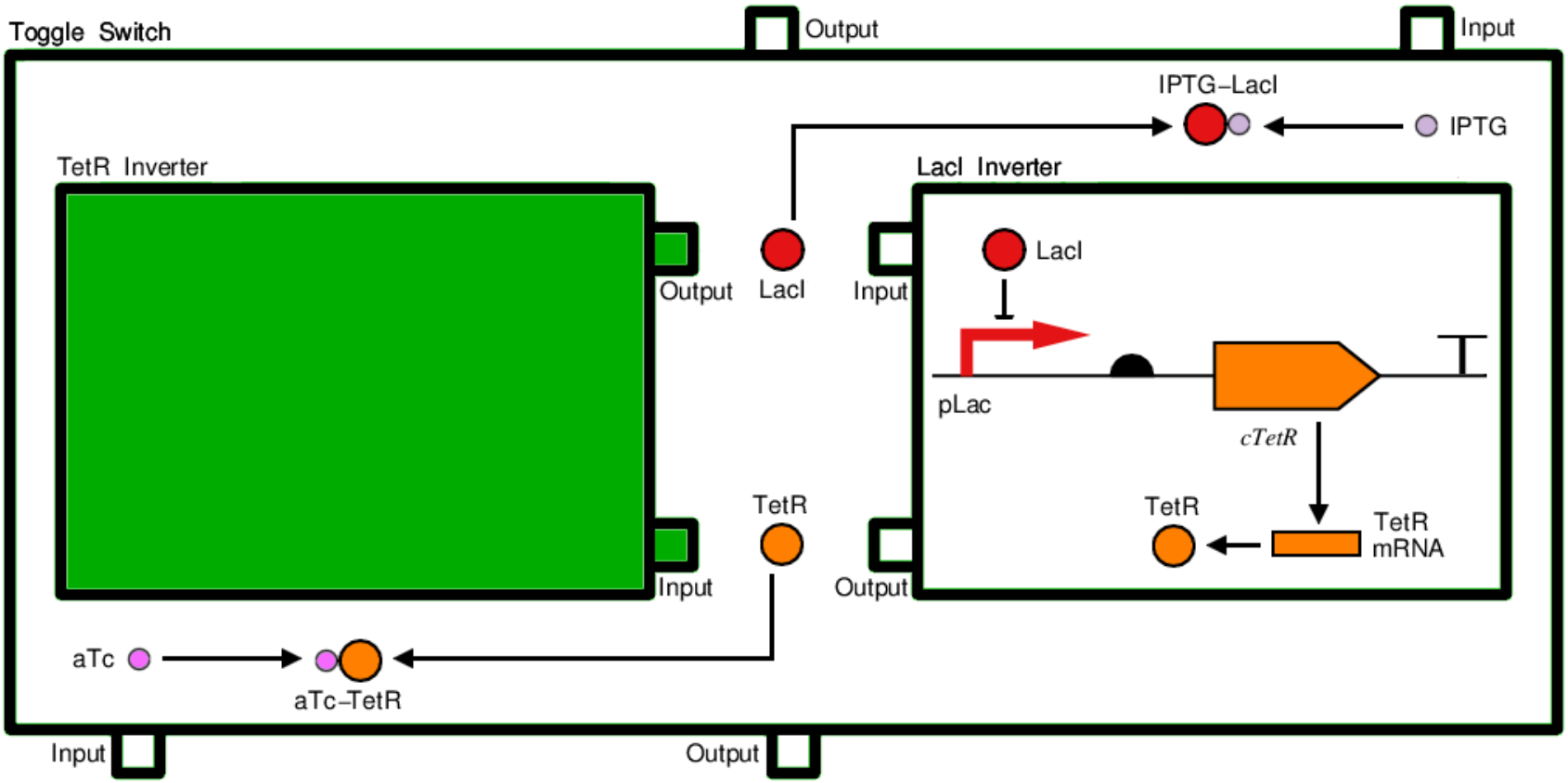
- Modules (logic gates, latches, oscillators, sensors, transducers, pathways, cascades)

Modules



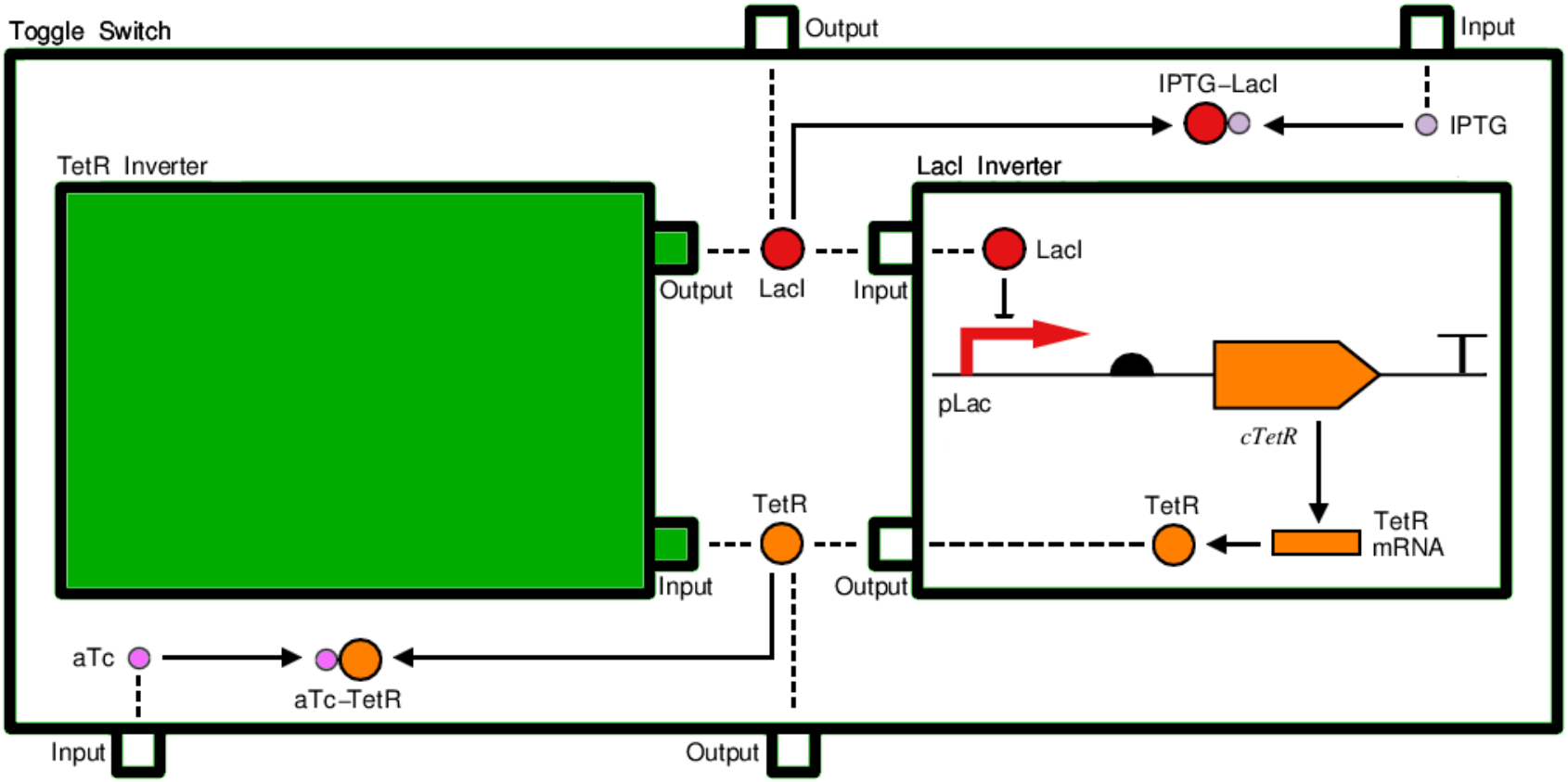
- Modules
- Ports

Modules



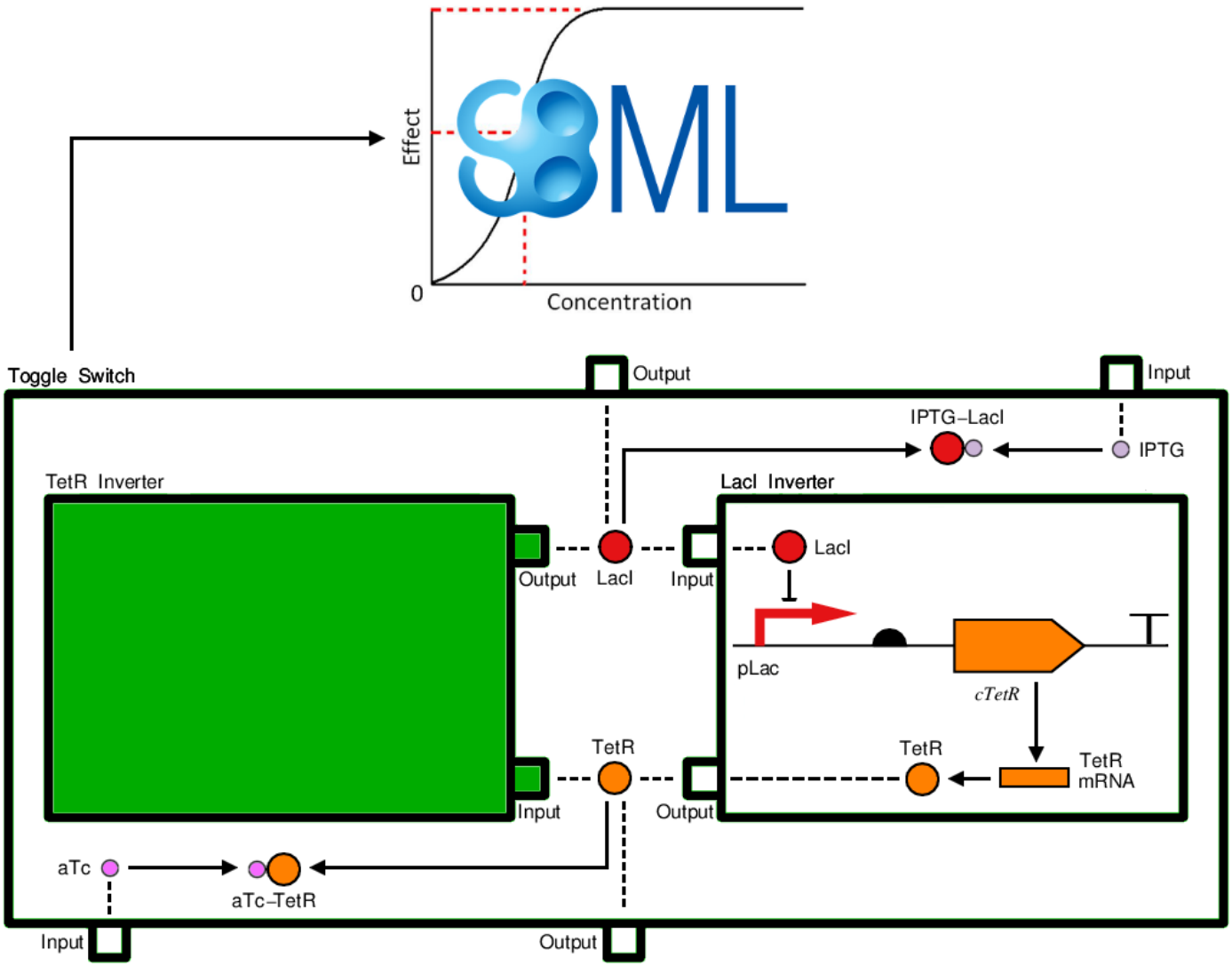
- Modules
- Ports
- Instantiation

Modules



- Modules
- Ports
- Instantiation
- Port Mapping

Models


Effect
0
Concentration
SBML
Toggle Switch
Output
Input
IPTG–LacI
IPTG
TetR Inverter
LacI Inverter
Output
LacI
Input
LacI
pLac
cTetR
TetR
TetR
mRNA
Input
TetR
Output
aTc
aTc–TetR
Input
Output

Annotations

```
<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:myerslab="http:www.async.ece.utah.edu"
  xmlns:sbol="http://sbols.org/v2#">
  <sbol:Collection rdf:about="http://www.async.ece.utah.edu/My_Parts/1.0">
    <dc:title>My_Parts</dc:title>
    <dc:description>These_are_my_parts.</dc:description>
    <sbol:member rdf:resource="http://www.async.ece.utah.edu/pLac/1.0"/>
    ...
  </sbol:Collection>
  <sbol:ModuleDefinition rdf:about="http://www.async.ece.utah.edu/LacI_Inverter/1.0">
    <sbol:role rdf:resource="http://www.async.ece.utah.edu/inverter"/>
    <sbol:component>
      <sbol:FunctionalComponent rdf:about="http://www.async.ece.utah.edu/LacI_Inverter/LacI/1.0">
        <sbol:definition rdf:resource="http://www.async.ece.utah.edu/LacI/1.0"/>
        <sbol:access rdf:resource="http://sbols.org/v2#public"/>
        <sbol:direction rdf:resource="http://sbols.org/v2#input"/>
      </sbol:FunctionalComponent>
    </sbol:component>
    ...
```

Annotations (cont)

```
    <sbol:interaction>
      <sbol:Interaction
           rdf:about="http://www.async.ece.utah.edu/LacI_Inverter/pLac_Repression/1.0">
        <sbol:type rdf:resource="http://identifiers.org/sbo/SBO:0000169"/>
        <myerslab:parameter rdf:resource="http://www.async.ece.utah.edu/Kr_LacI">
        <sbol:participation>
          <sbol:Participation
           rdf:about="http://www.async.ece.utah.edu/LacI_Inverter/pLac_Repression/p1/1.0">
            <sbol:role rdf:resource="http://www.async.ece.utah.edu/repressed"/>
            <sbol:participant
               rdf:resource="http://www.async.ece.utah.edu/LacI_Inverter/pLac/1.0"/>
          </sbol:Participation>
        </sbol:participation>
        <sbol:participation>
          <sbol:Participation
           rdf:about="http://www.async.ece.utah.edu/LacI_Inverter/pLac_Repression/p2/1.0">
            <sbol:role rdf:resource="http://identifiers.org/sbo/SBO:0000020"/>
            <sbol:participant
               rdf:resource="http://www.async.ece.utah.edu/LacI_Inverter/LacI/1.0"/>
          </sbol:Participation>
        </sbol:participation>
      </sbol:Interaction>
    </sbol:interaction>
    ...
  </sbol:ModuleDefinition>
  <myerslab:Parameter rdf:about="http://www.async.ece.utah.edu/Kr_LacI">
    <myerslab:type rdf:resource="http://identifiers.org/sbo/SBO:0000281"/>
    <myerslab:value>0.5</myerslab:value>
  </myerslab:Parameter>
  ...
</rdf:RDF>
```

libSBOL - Libraries to Support SBOL

- Crucial to the success of a standard is software infrastructure to support developers' integration of the standard within their tools.
- `libSBOL` libraries are 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, current release supports SBOL 2.0.
  - libSBOLc - C/C++ library, current release supports SBOL 1.1. SBOL 2.0 support coming soon.
  - pySBOL - Python library, current release supports SBOL 1.1. SBOL 2.0 support coming soon.
- Library distributions include detailed documentation for the class definitions and the methods provided by the API.
- Leveraging `libSBOj` many software tools now support SBOL 1.1.

Partial List of SBOL Compliant Software Tools

Software	Type	Visual	Data Model
ICE	Repository	Yes	Yes
SBOL Hub	Repository	No	Yes
GenBank Converter	Helper	No	Yes
MoSeC	Helper	No	Yes
DNAplotlib	Visualization	Yes	Yes
Pigeon	Visualization	Yes	No
VisBOL	Visualization	Yes	Yes
Benchling	DNA Design	No	Yes
DeviceEditor	DNA Design	Yes	Yes
Eugene	DNA Design	Yes	Yes
GeneGenie	DNA Design	No	Yes
GenoCAD	DNA Design	Yes	Yes
SynBad	DNA Design	Yes	Yes
TeselaGen	DNA Design	Yes	Yes
VectorEditor	DNA Design	No	Yes
Cello	Circuit Design	Yes	Yes
iBioSim	Circuit Design	No	Yes
RetroPath	Circuit Design	No	Yes
SBROME	Circuit Design	No	Yes
TinkerCell	Circuit Design	Yes	Yes
BioCompiler	Circuit Design	Yes	Yes
Clotho 3.0	Assembly	Yes	Yes
Raven	Assembly	Yes	No

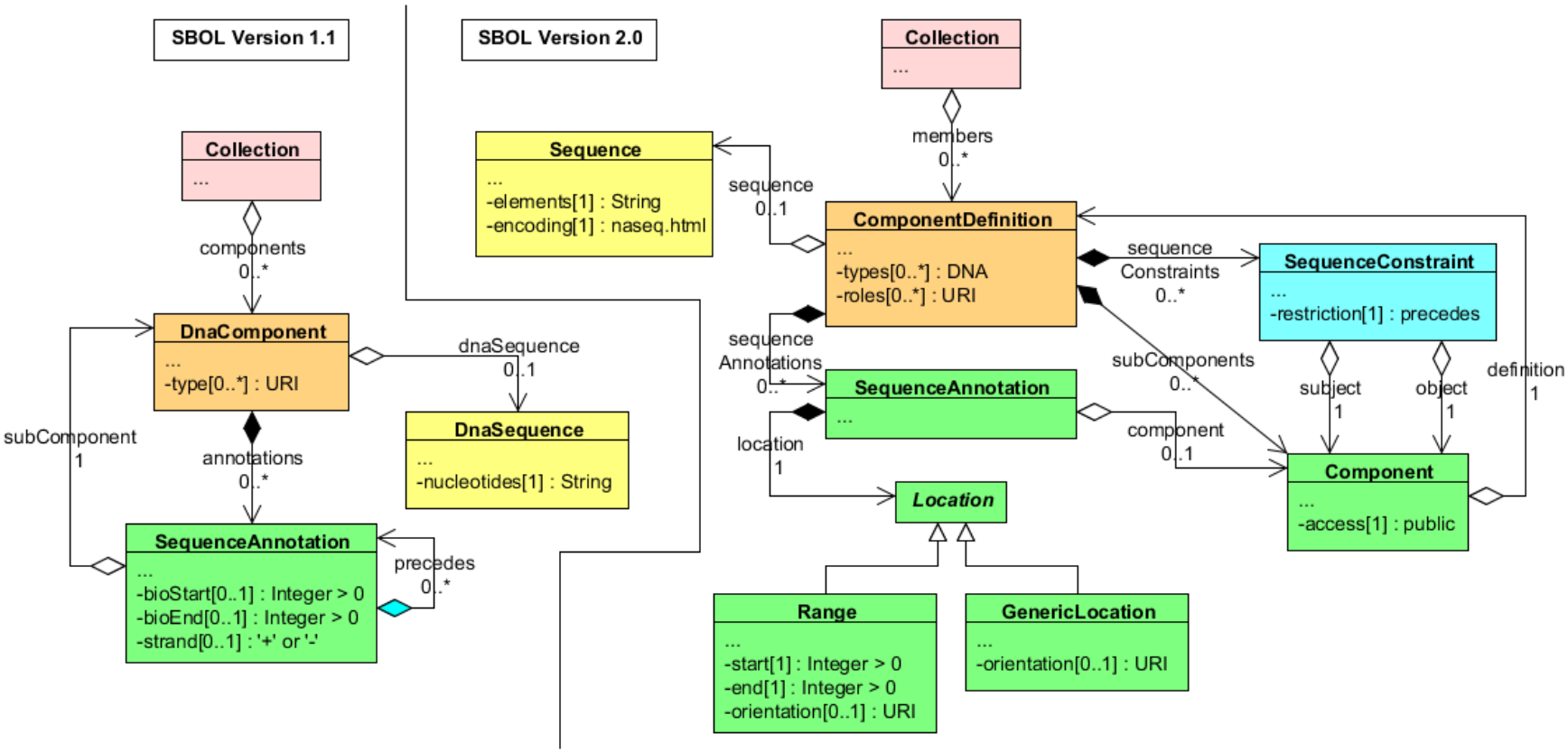
WARNING:
Changing rapidly, so likely not completely accurate.

libSBOLj 2.0

- `libSBOLj 2.0` easing the adoption of SBOL 2.0 by tool developers.
- Organizes all SBOL data within an *SBOL document*.
- Includes a list of each type of *top level* object: *collections*, *modules*, *components*, *sequences*, *models*, and *generic top level* objects.
- These lists are organized as hash maps to allow for easy search by their *unique reference identifiers* (URIs) and validation that they are distinct.
- Library includes methods for creating, updating, accessing, and removing these data objects, as well as, their child objects.
- Supports reading/writing data using RDF/XML, Turtle, and JSON formats.
- Software tools that need to store data that is not currently encoded within SBOL can do so using generic top level objects and custom annotations.

Support for SBOL 1.1

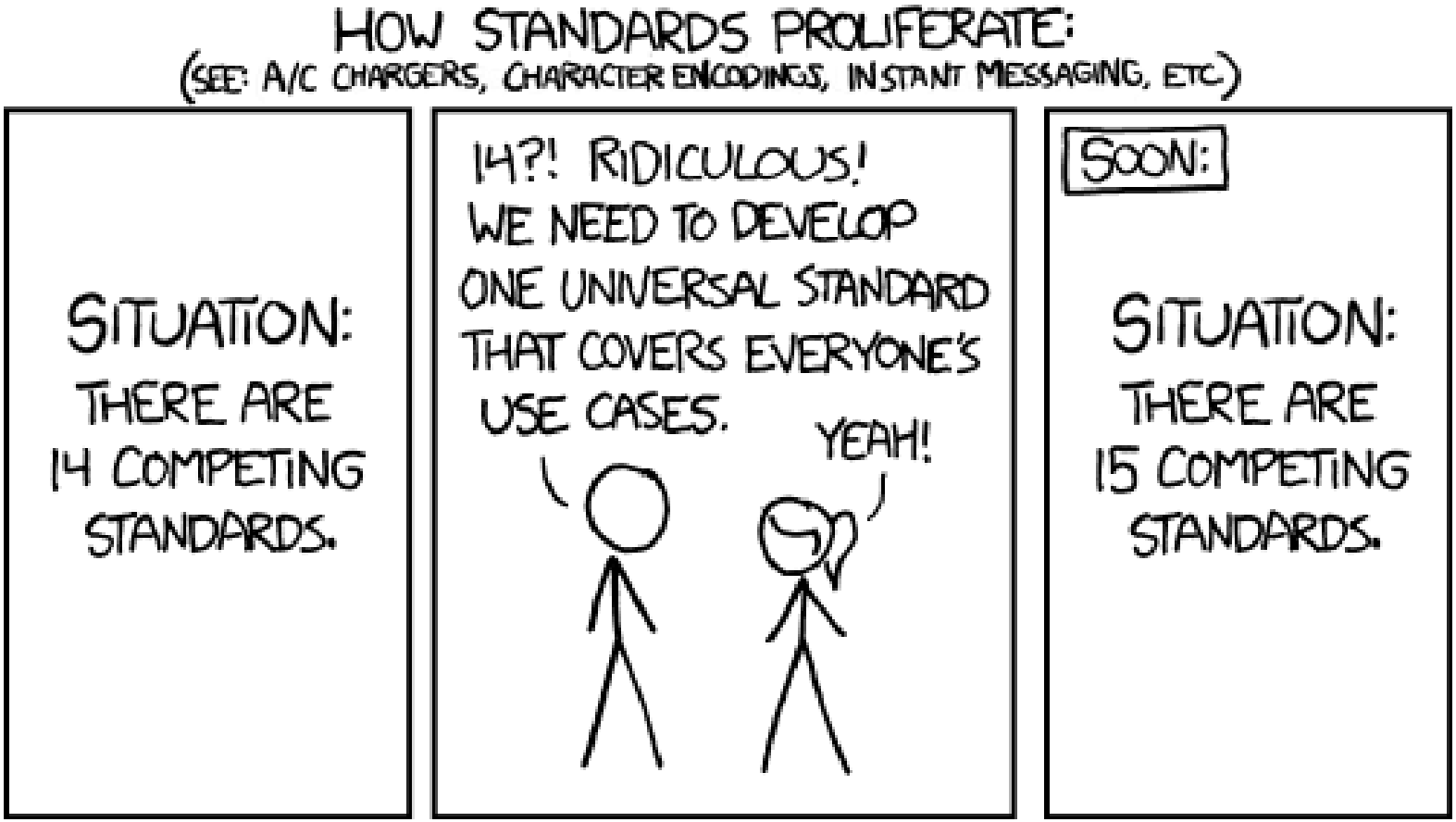
- `libSBOLj 2.0` eases transition for developers who adopted SBOL 1.1.
  1. `libSBOLj 1.1` is included unmodified to manage SBOL 1.1 data objects.
  2. New file reader automatically detects the SBOL version, and if SBOL 1.1 data is detected, it is automatically converted to an SBOL 2.0 data object.



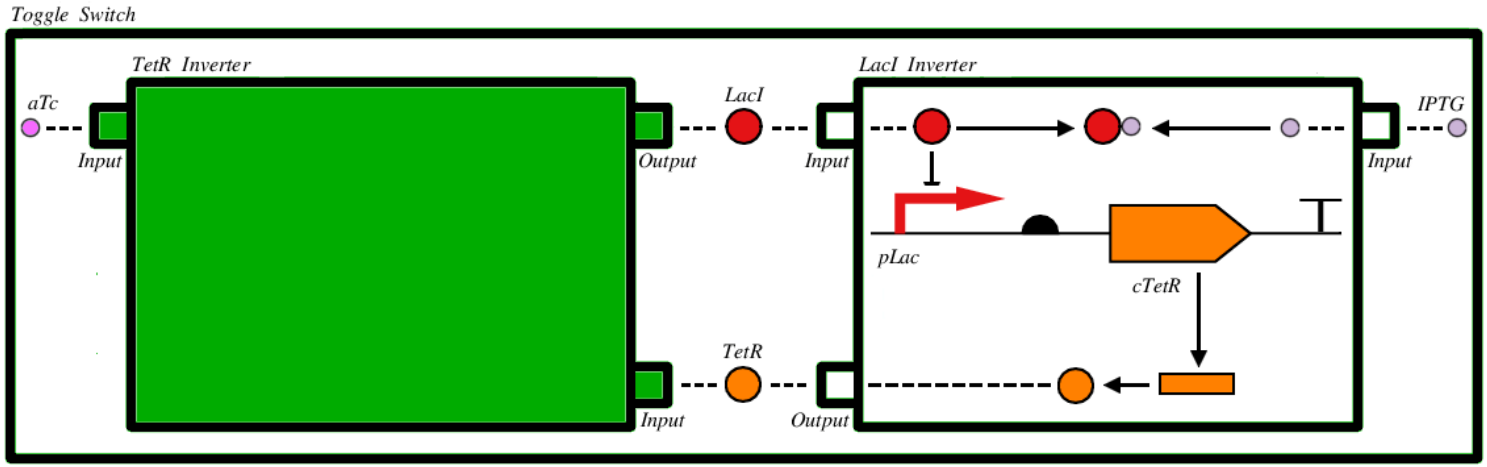
SBOL and COMBINE

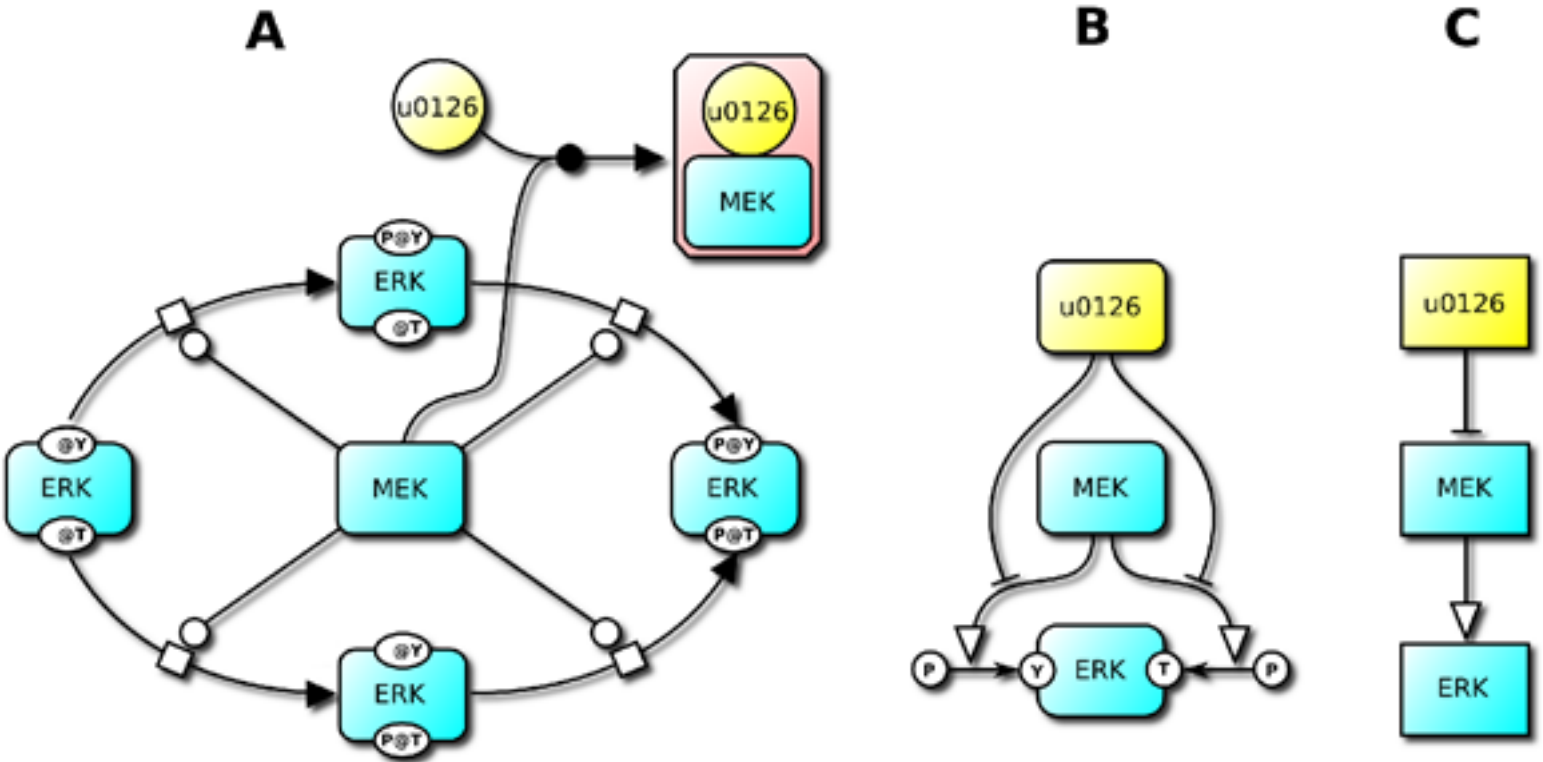
- *SBOL Developers Group* meet twice a year to discuss next steps.
- SBOL recently joined the COMBINE community of standards.
- Last meeting was in Wittenberg, Germany in April.
- Next meeting is in Salt Lake City, UT on October 12th to 16th.
- More information: `http://co.mbine.org/events/COMBINE_2015`

Why did SBOL join COMBINE?

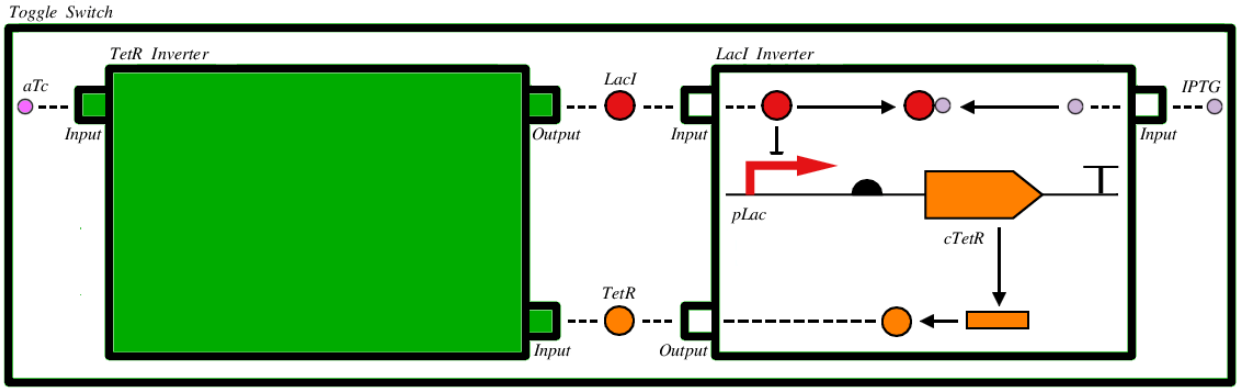


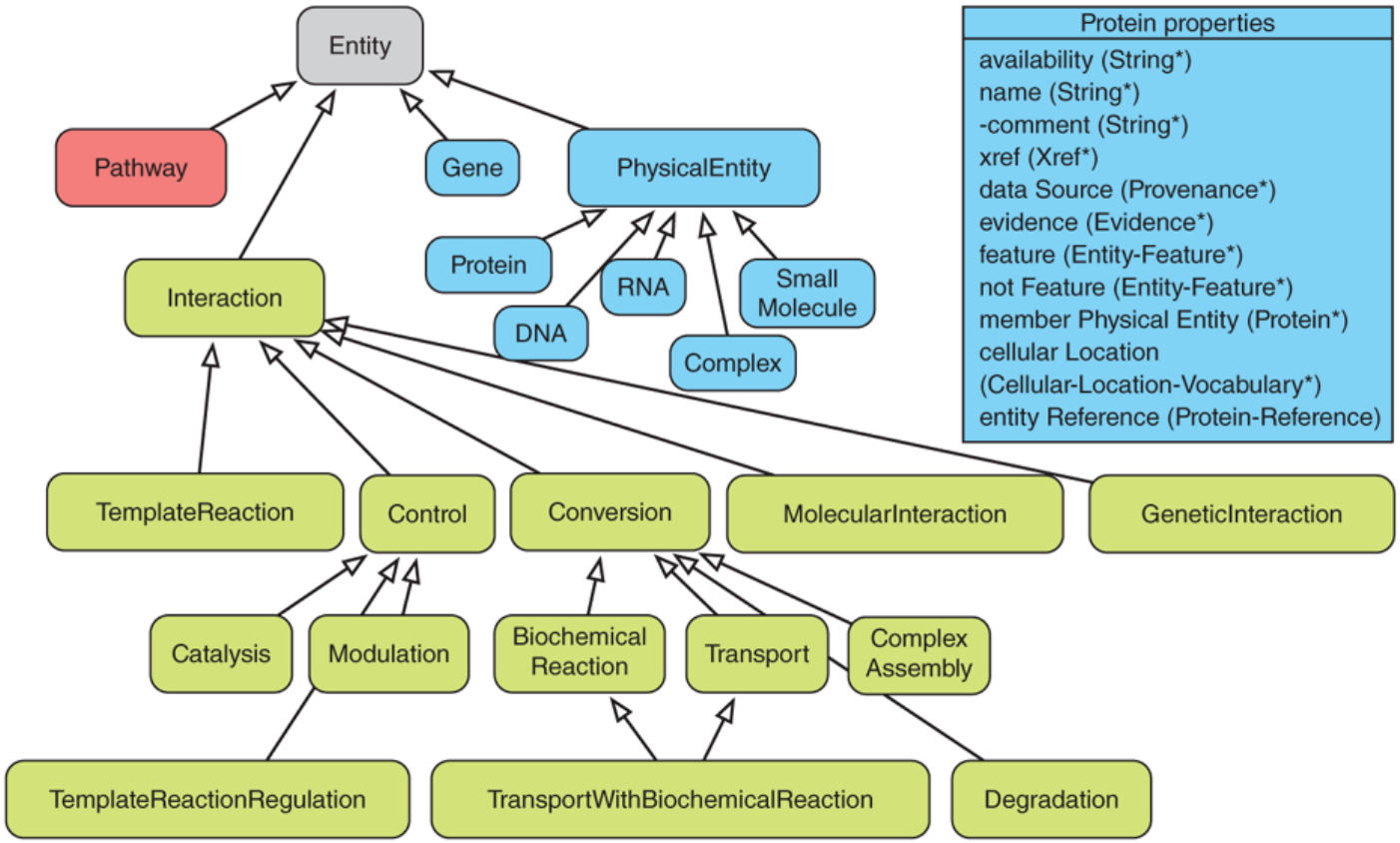
SBOL Visual/SBGN



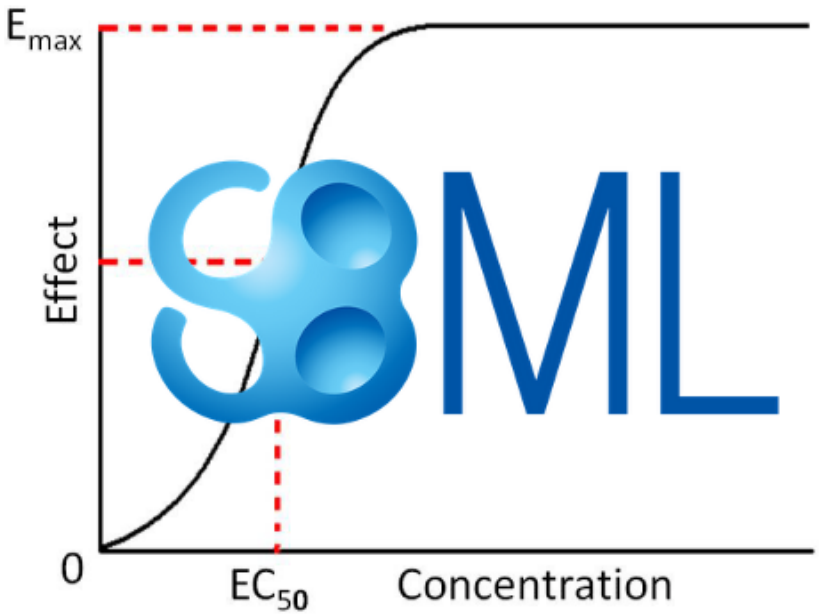


SBOL/BioPAX


Toggle Switch
TetR Inverter
LacI Inverter
aTc
Input
Output
LacI
IPTG
pLac
cTetR
TetR

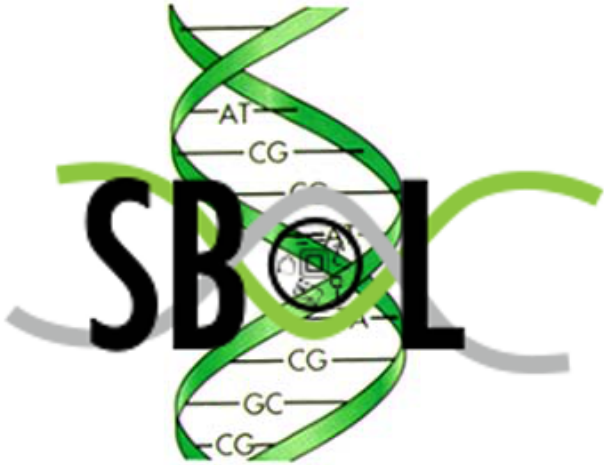

Entity
Pathway
Gene
PhysicalEntity
Protein
RNA
Small Molecule
DNA
Complex
Interaction
TemplateReaction
Control
Conversion
MolecularInteraction
GeneticInteraction
Catalysis
Modulation
Biochemical Reaction
Transport
Complex Assembly
TemplateReactionRegulation
TransportWithBiochemicalReaction
Degradation
Protein properties
availability (String*)
name (String*)
-comment (String*)
xref (Xref*)
data Source (Provenance*)
evidence (Evidence*)
feature (Entity-Feature*)
not Feature (Entity-Feature*)
member Physical Entity (Protein*)
cellular Location
(Cellular-Location-Vocabulary*)
entity Reference (Protein-Reference)

Connecting SBML to SBOL



SBML

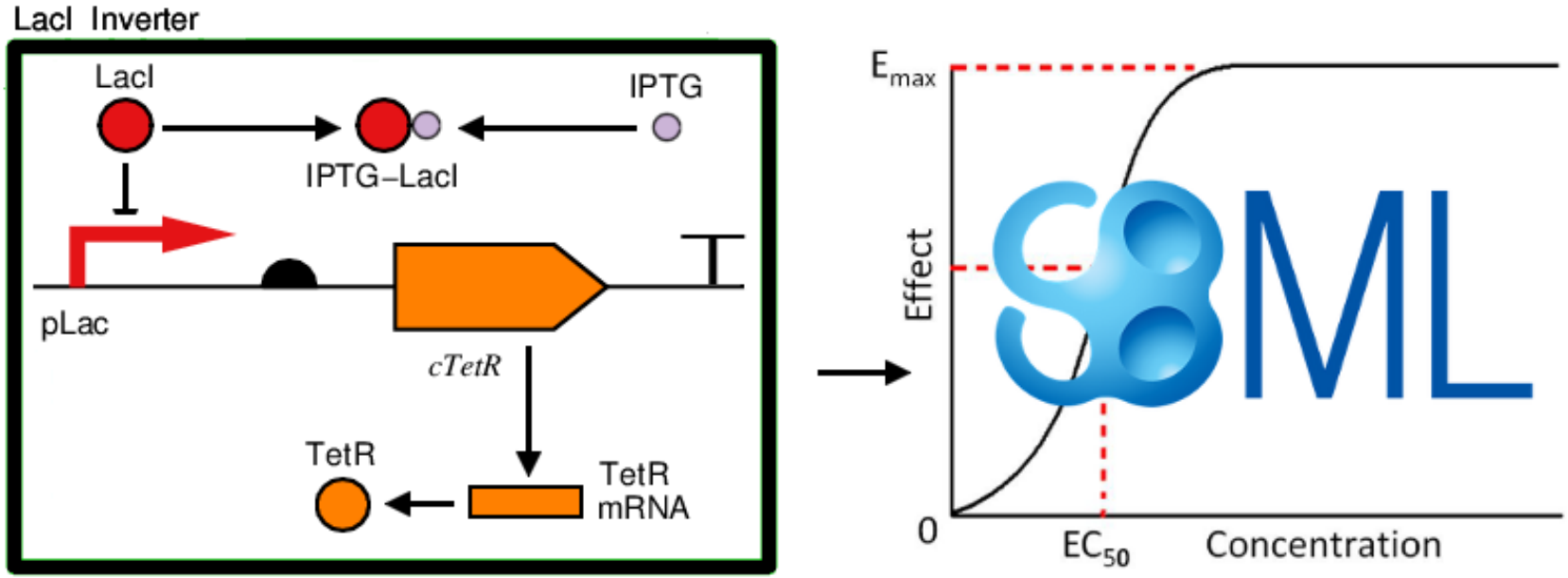
RDF/XML
Annotation



SBOL

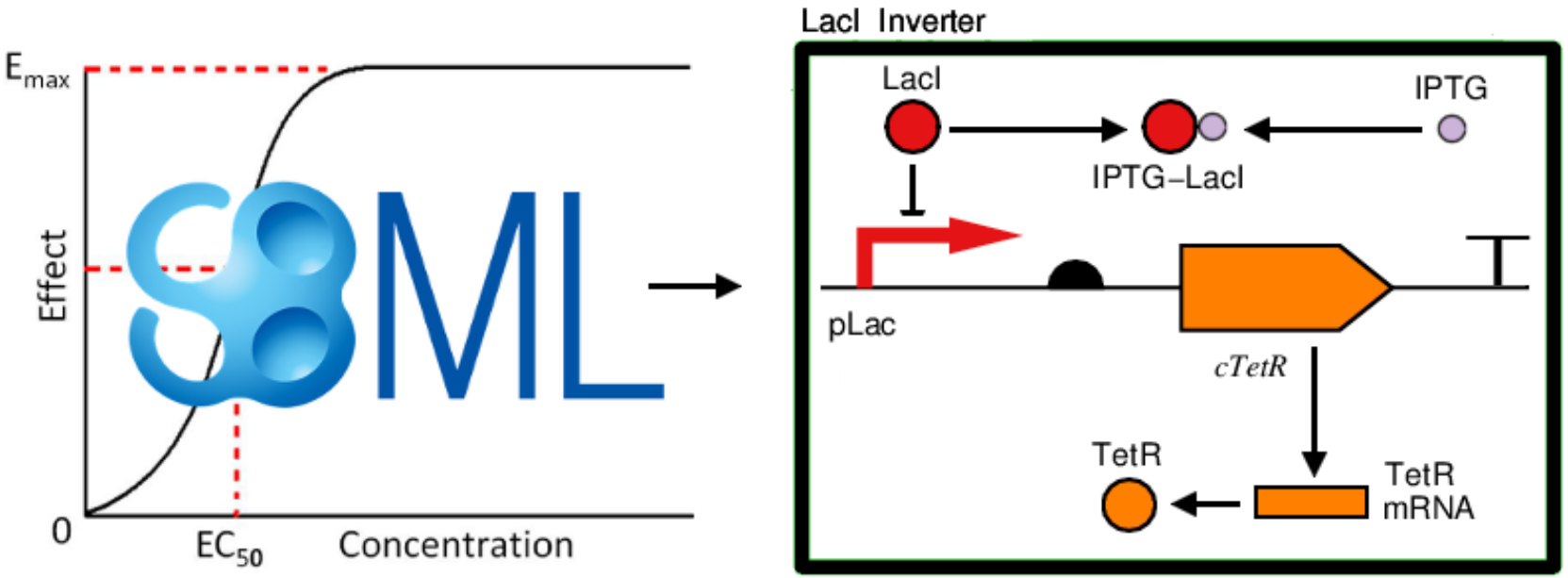
Roehner et al., ACS Synthetic Biology (2013).

Converting SBOL to SBML



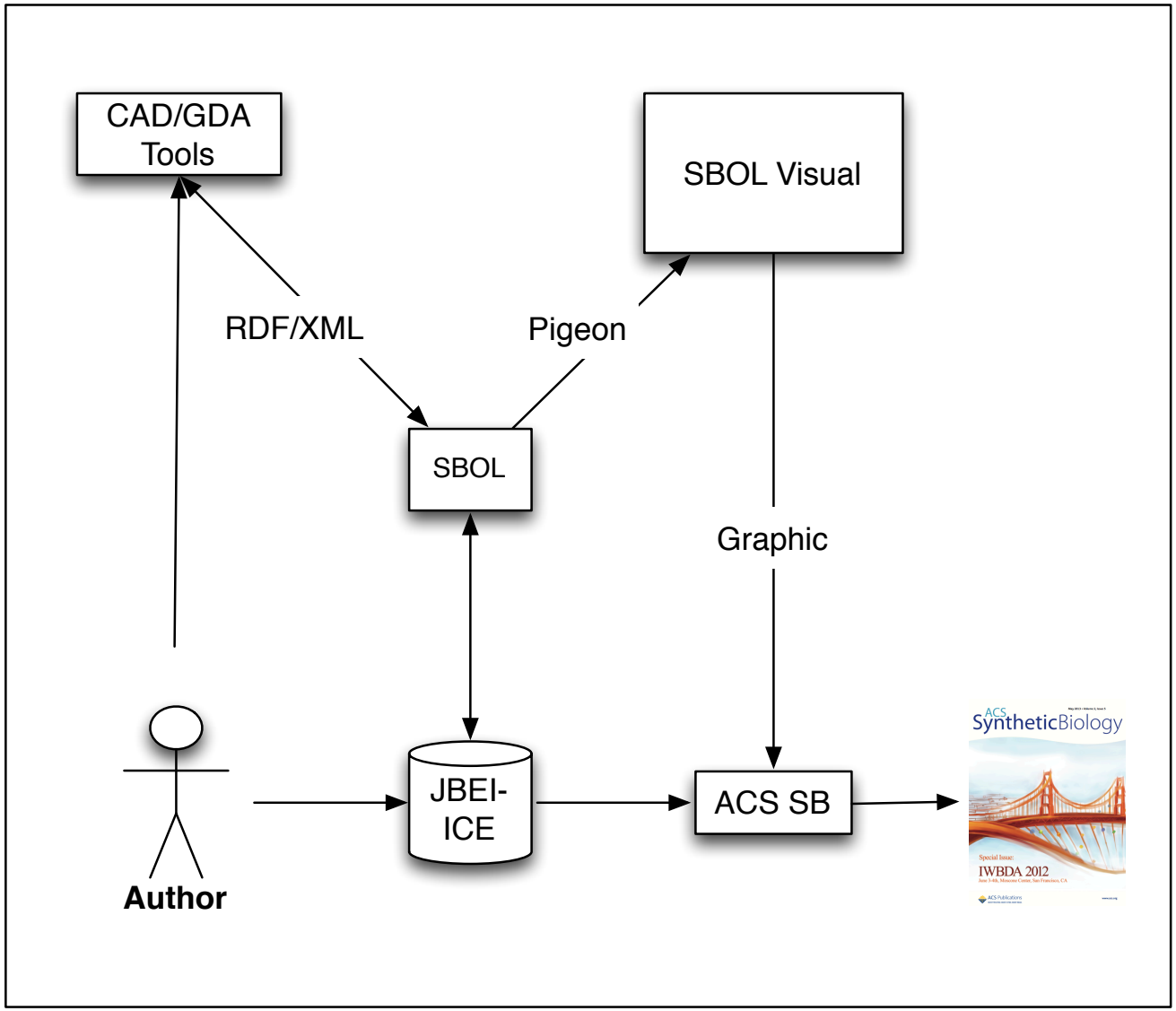
Roehner et al., ACS Synthetic Biology (2014)

Converting SBML to SBOL



Nguyen/Myers, IWBDA (2015)

SBOL and ACS Synthetic Biology



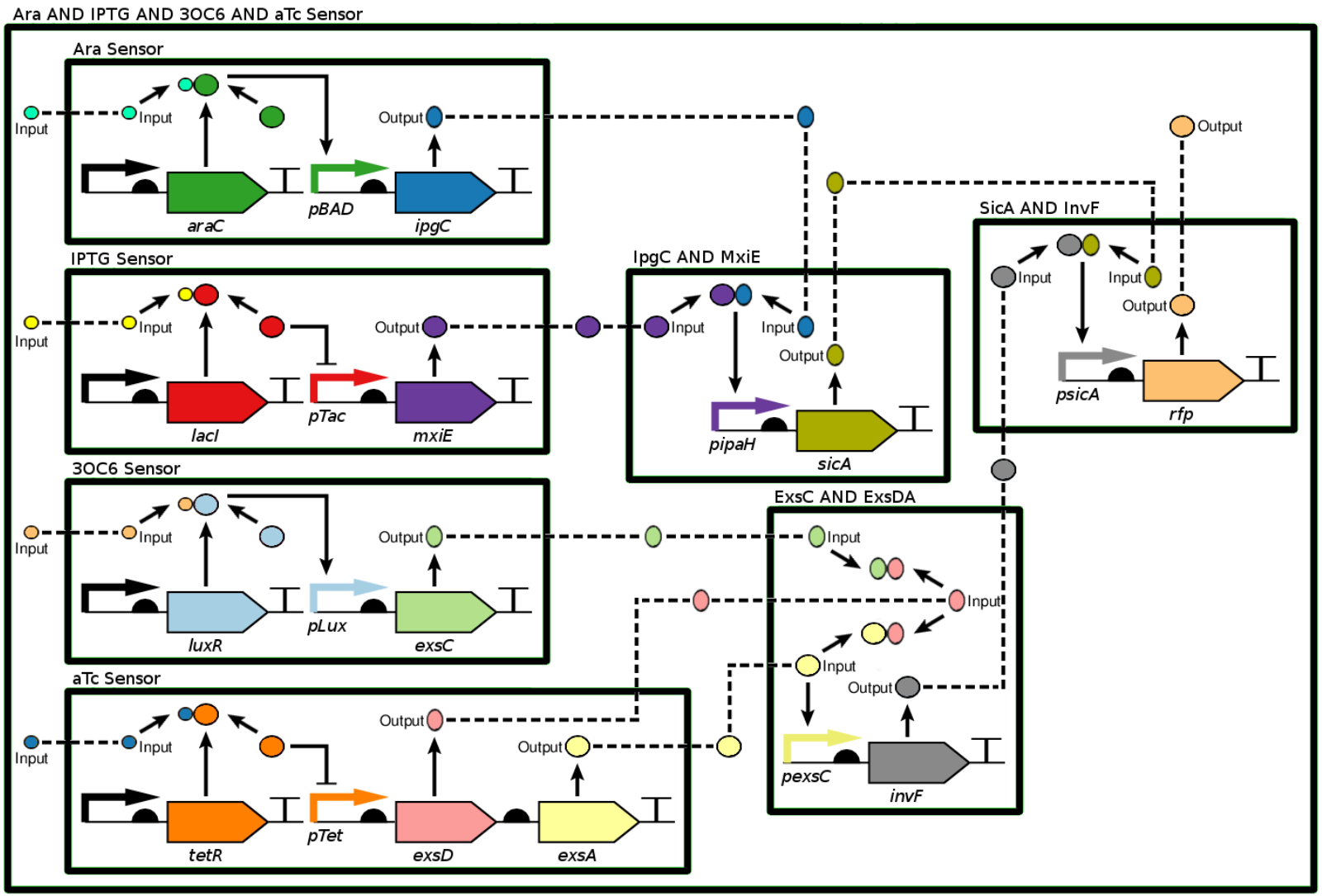
Conclusion

- Standards are an important enabler for progress in synthetic biology.
- SBOL 1.1 is an important step forward, but it remains limited to only describing components constructed from DNA.
- SBOL 2.0 introduces more generalized components, interactions between them, and hierarchical modules to group components implementing a desired function.
- Provides rationale for the design, making it possible to directly infer information about the system and its operation.
- Supported by `libSBOLj 2.0`, which provides an API, documentation, and utilities for managing the encoding and exchange of designs.
- Collaborations with the COMBINE community of standards will be essential to our standardization efforts.

More Information

- More information about SBOL: `http://www.sbolstandard.org/`.
- More information about COMBINE: `http://www.co.mbine.org/`.
- `libSBOLj` is open source under the Apache 2.0 License, and it is available from the SBOL website.
- Questions about the library can be sent to: `libsbol-team@googlegroups.com`.
- Join the SBOL Developers Group by emailing: `sbol-editors@googlegroups.org`.

SBOL Demonstration Example



Moon et al., Nature (2012)

SBOL Demonstration

- Proto BioCompiler - Jacob Beal (BBN)
- ICE - Oge Nnadi (JBEI)
- DNAPlotLib - Bryan Bartley (Washington)
- Cello - Nicholas Roehner (Boston)
- iBioSim - Chris Myers (Utah)
- SBOL Stack / SBOL Hub - Curtis Madsen (Newcastle)