Software Infrastructure for Synthetic Biology

Zhen Zhang, Tramy Nguyen, Chris Myers
Department of Electrical and Computer Engineering
University of Utah

Introduction and Motivation

- *Synthetic biology* is an engineering discipline where biological components are assembled to form devices or systems with more complex functions [1].

- It includes the following three main aspects:
  - Standards define a data representation format for encoding synthetic biology designs.
  - Abstraction allows users to focus on high-level designs without worrying about low-level details.
  - Decoupling separates the design using software tools from building biological components.

- Missing sequence information in the synthetic biology literature precludes reproducibility and reuse of the published work [3].

- Encoding designs in a restricted standardized format allows different software tools to exchange models and store them for reproducibility and reuse of the past work.

- The infrastructure provided by `libSBOLj` enables integration of SBOL standards in software tools.

Synthetic Biology Open Language (SBOL)

- SBOL is a standard electronic data format for exchanging synthetic biology designs.

- Since 2008, SBOL has been developed under an international effort from both experimental and computational synthetic biologists.

- The SBOL 1.1 core data model enables specifications of DNA-level designs [1].

- SBOL 1.1 represents biological building blocks as DNA components.

- *DNA components* describe hierarchically annotated DNA sequences.
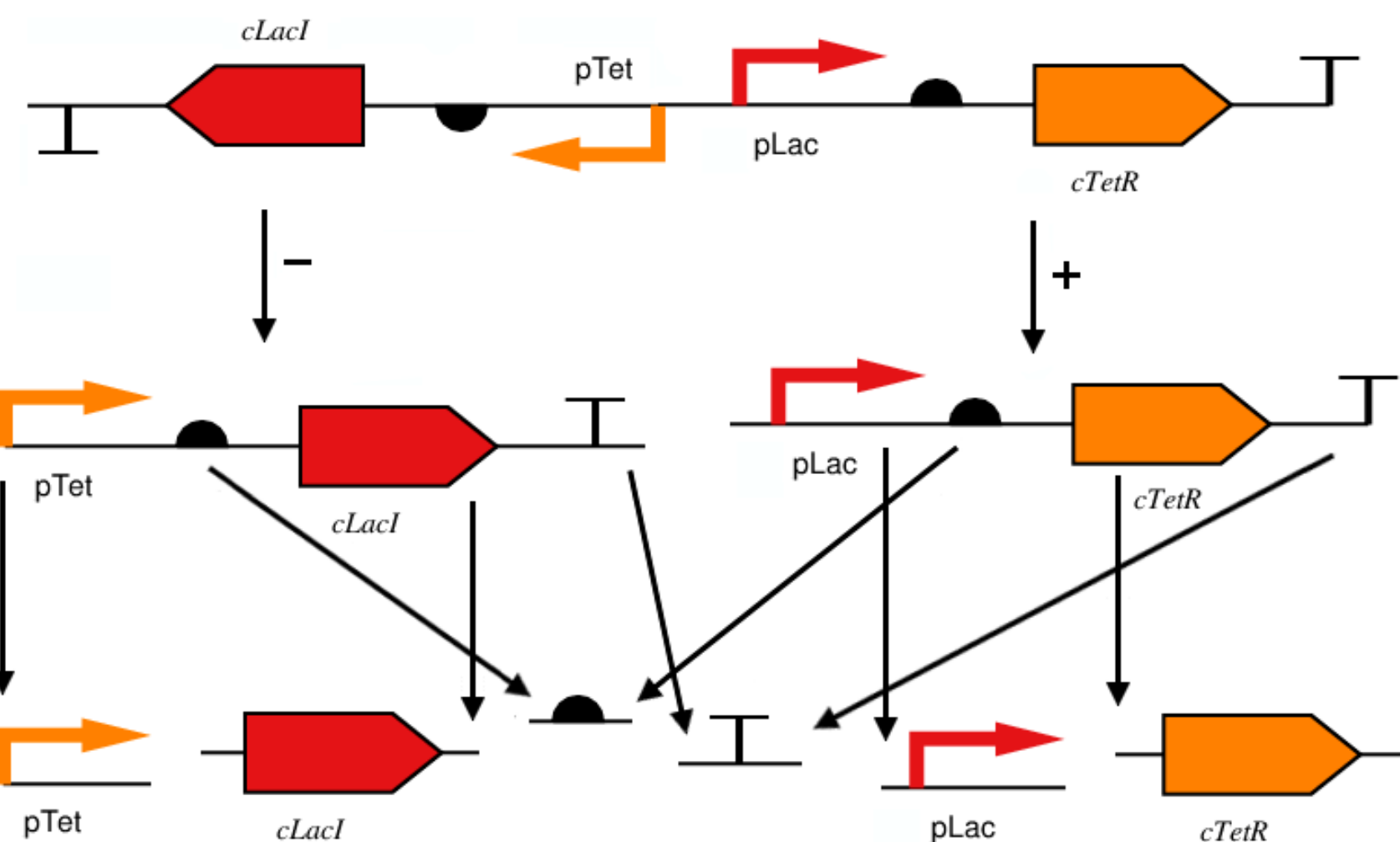
Hierarchical composition of the DNA component for a genetic toggle switch in SBOL Version 1.1.

- Encoding a genetic toggle switch [2] in SBOL 1.1 allows the design to be assembled, optimized, tested, and stored by software tools developed at different organizations.
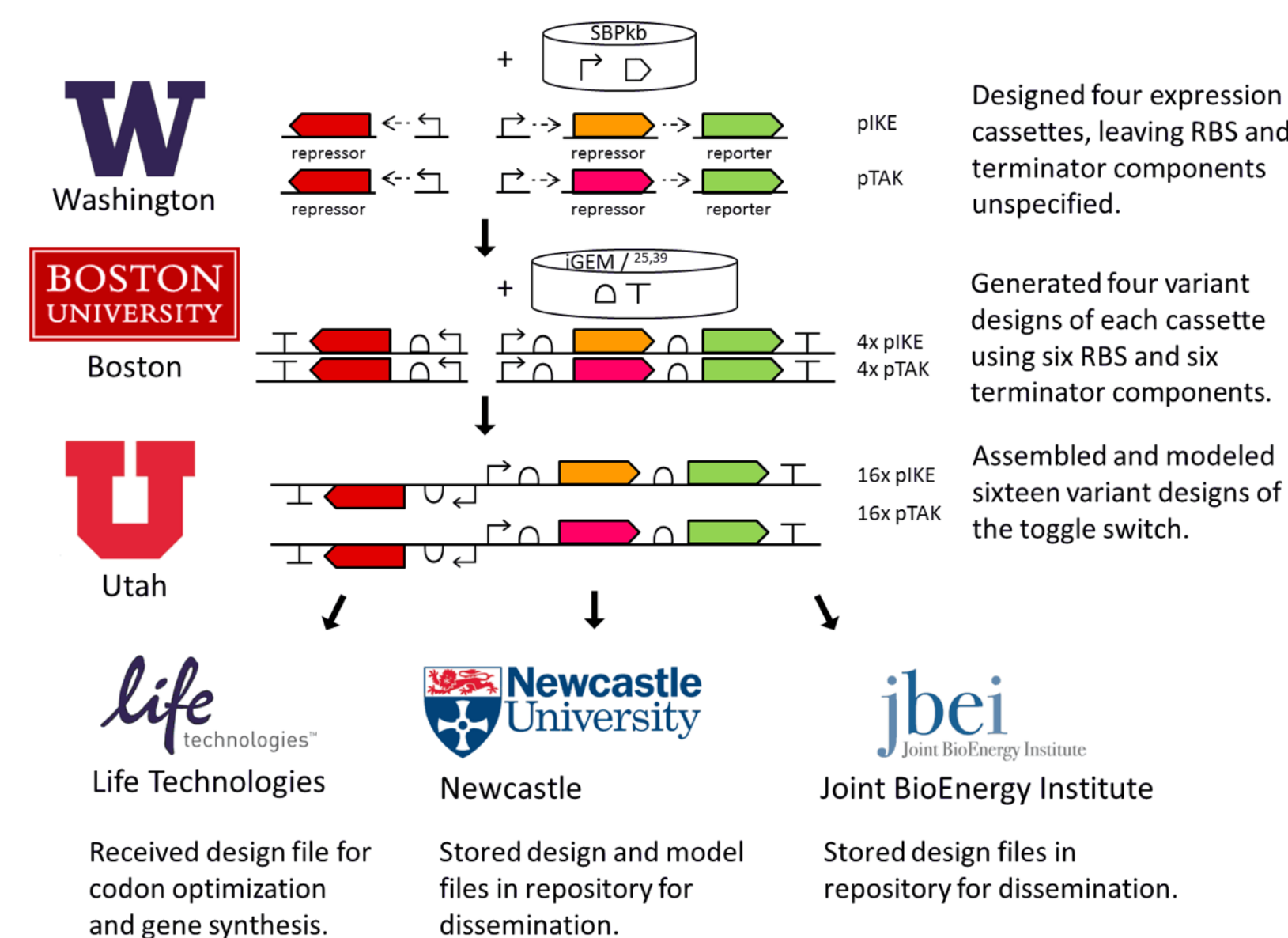

Demonstration of a collaboration to design a genetic toggle switch.

- It demonstrates the potential for collaborative work on complex synthetic biology designs.

SBOL 2.0

- The SBOL 2.0 proposal [4] allows specifications of more types of *components*, such as DNA, RNA, proteins, small molecules, etc.

- *Interactions* represent relationships between components, such as *activation*, *repression*, *complex formation*, etc.

- A *module* groups several interactions to represent a useful functional block.

- A module can be connected to an informative external behavioral *model* represented in a different language (i.e., SBML, CellML, Matlab, etc.).

- SBOL 2.0 includes constructs to hierarchical module definitions.
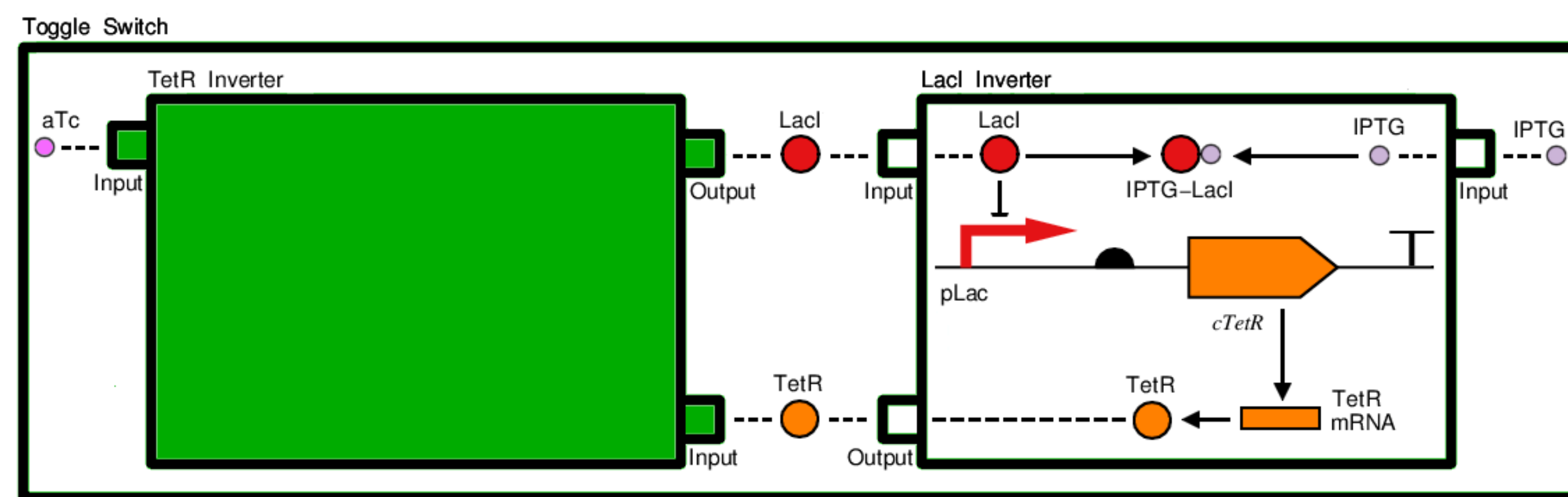
Hierarchical toggle switch module in SBOL 2.0.

libSBOLj 2.0

- `libSBOLj 2.0` is a native Java implementation of SBOL 2.0 data structure.

- It provides an API for tool developers to interact with SBOL data objects.

- Serialization of libSBOLj 2.0 has also been implemented.
  - Users are able to read/write SBOL 2.0 files.
  - SBOL 1.1 can be converted to SBOL 2.0 files.
  - Serialization can support RDF/XML, Turtle, and JSON formats.

- Methods are available for data operations (i.e., creating, storing, accessing, nesting, etc.).

- It includes detailed documentation for class definitions and methods in the API.

- The goal for the library development is to ease the adoption of SBOL 2.0 by tool developers.

Hierarchical UML model of a toggle switch in SBOL 2.0.

```xml
<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:sbol="http://sbols.org/v2#">
  <sbol:Collection rdf:about="http://www.async.ece.utah.edu/myParts/1/0">
    <sbol:timeStamp>2015-01-27 12:43:24.697</sbol:timeStamp>
    <sbol:name>myParts</sbol:name>
    <sbol:description>myParts</sbol:description>
    <sbol:member rdf:resource="http://www.async.ece.utah.edu/LacI_Inv/1/0"/>
    ...
  </sbol:Collection>
  <sbol:ModuleDefinition rdf:about="http://www.async.ece.utah.edu/LacI_Inv/1/0">
    <sbol:name>LacI_Inv</sbol:name>
    <sbol:role rdf:resource="http://some.ontology.org/Inverter"/>
    <sbol:component>
      <sbol:FunctionalComponent rdf:about="http://www.async.ece.utah.edu/LacI_Inv/LacIIn/1/0">
        <sbol:name>LacIIn</sbol:name>
        <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>
    ...
    <sbol:interaction>
      <sbol:Interaction rdf:about="http://www.async.ece.utah.edu/LacI_Inv/interact1/1/0">
        <sbol:type rdf:resource="http://some.ontology.org/repression"/>
        <sbol:participation>
          <sbol:Participation rdf:about="http://www.async.ece.utah.edu/LacI_Inv/interact1/p2a/1/0">
            <sbol:role rdf:resource="http://some.ontology.org/repressed"/>
            <sbol:participant rdf:resource="http://www.async.ece.utah.edu/LacI_Inv/LacIInv/1/0"/>
          </sbol:Participation>
        </sbol:participation>
        <sbol:participation>
          <sbol:Participation rdf:about="http://www.async.ece.utah.edu/LacI_Inv/interact1/p1a/1/0">
            <sbol:role rdf:resource="http://some.ontology.org/repressor"/>
            <sbol:participant rdf:resource="http://www.async.ece.utah.edu/LacI_Inv/LacIIn/1/0"/>
          </sbol:Participation>
        </sbol:participation>
      </sbol:Interaction>
    </sbol:interaction>
    ...
  </sbol:ModuleDefinition>
  ...
</rdf:RDF>
```
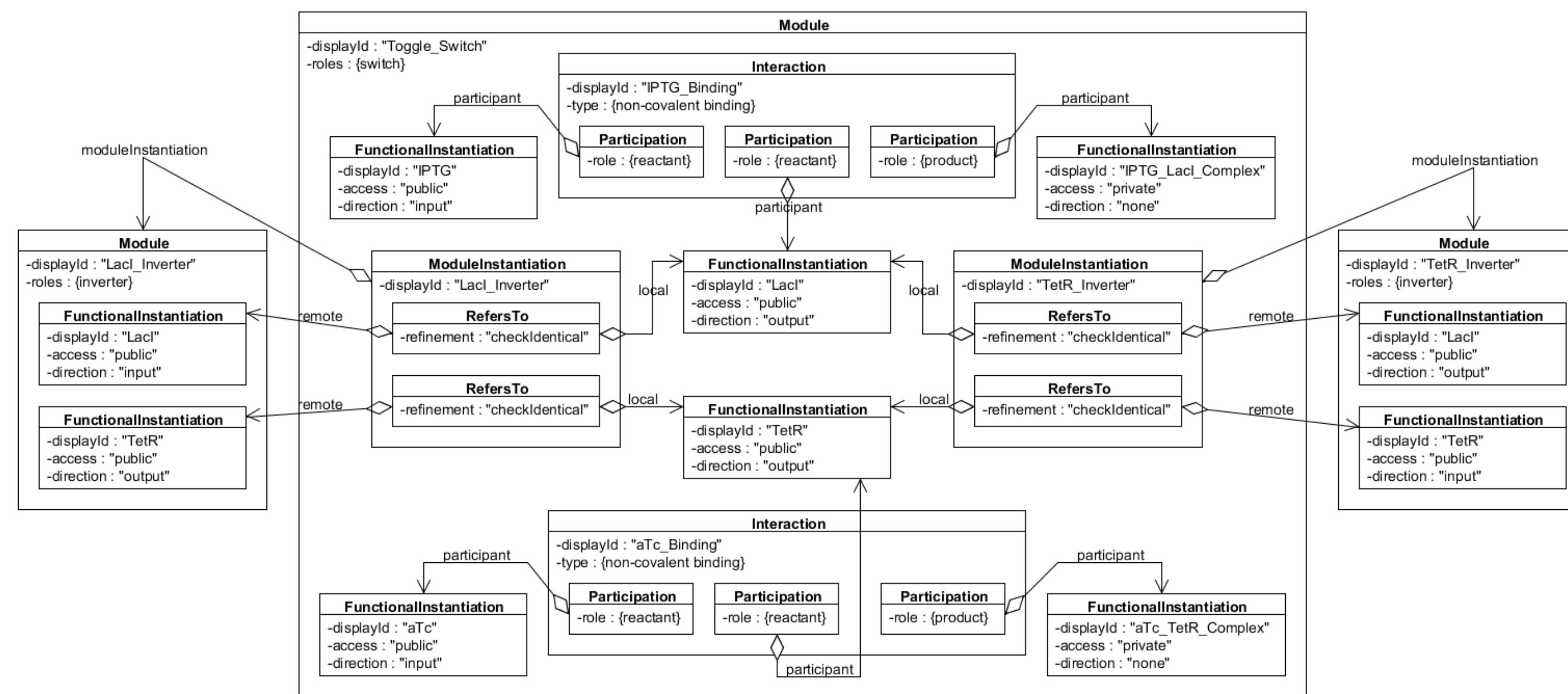
Discussion

- Standards are important in the synthetic biology field.

- SBOL 1.1 is limited to only describing components constructed from DNA.

- SBOL 2.0 introduces more generalized components, the interactions between them, and hierarchical module that group components implementing a desired function.

- A java based library of SBOL 2.0 has been implemented (libSBOLj 2.0).

- Additional functions and error checking routines continue to be added to ease adoption of SBOL 2.0.

- Please see: `http://www.sbolstandard.org/`

References

[1] M. Galdzicki, K. P. Clancy, E. Oberortner, M. Pocock, J. Y. Quinn, C. A. Rodriguez, N. Roehner, M. L. Wilson, L. Adam, J. C. Anderson, B. A. Bartley, J. Beal, D. Chandran, J. Chen, D. Densmore, D. Endy, R. Gruenberg, J. Hallinan, N. J. Hillson, J. D. Johnson, A. Kuchinsky, M. Lux, G. Misirli, J. Peccoud, H. A. Plahar, E. Sirin, G-B. Stan, A. Villalobos, A. Wipat, J. H. Gennari, C. J. Myers, and H. M. Sauro. The synthetic biology open language (sbol) provides a community standard for communicating designs in synthetic biology. *NATURE BIOTECHNOLOGY*, 32:545–550, 2014.

[2] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in *escherichia coli*. *Nature*, 403:339–342, 2000.

[3] J. Peccoud, J. C. Anderson, D. Chandran, D. Densmore, M. Galdzicki, M. W. Lux, C. A. Rodriguez, G-B. Stan, and H. M. Sauro. Essential information for synthetic DNA sequences. *Nature Biotechnology*, 29(1):22–22, Jan 2011.

[4] N. Roehner, E. Oberortner, M. Pocock, J. Beal, K. Clancy, C. Madsen, G. Misirli, A. Wipat, H. Sauro, and C. J. Myers. Proposed data model for the next version of the synthetic biology open language. *ACS Synthetic Biology*, 4(1):57–71, 2015. PMID: 24896221.