

Proposed Data Model for the Next Version of the Synthetic Biology Open Language (SBOL)

Nicholas Roehner^{*}, Ernst Oberortner[†], Matthew Pocock[°], Jacob Beal[‡],
Kevin Clancy[§], Curtis Madsen[°], Goksel Misirli[°], Anil Wipat[°],
Herbert Sauro^{||}, and Chris J. Myers^{*}

^{*}University of Utah, [†]Boston University, [°]Newcastle University, [‡]Raytheon
BBN Technologies, [§]Life Technologies, and ^{||}University of Washington

July 15, 2014

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



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 sequence⁸. 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 oscillators⁹. Many more such examples are likely to emerge.

Because full sequence disclosure is critical,

and welcome contributions from the greater community.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

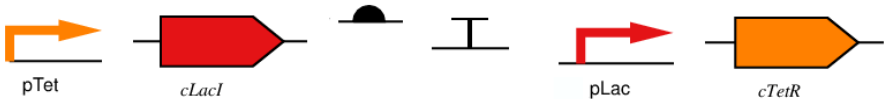
Jean Peccoud¹, J Christopher Anderson², Deepak Chandran³, Douglas Densmore⁴, Michal Galdzicki⁵, Matthew W Lux¹, Cesar A Rodriguez⁶, Guy-Bart Stan⁷ & Herbert M Sauro³

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.

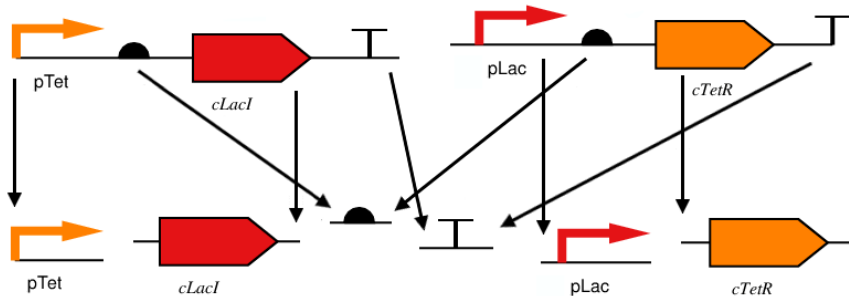
⁴Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts, USA. ⁵Biomedical and Health Informatics, University of Washington, Seattle, Washington, USA. ⁶BIOFAB, Emeryville, California, USA. ⁷Department of Bioengineering and Centre for Synthetic Biology and Innovation,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

Synthetic Biology Open Language (SBOL) Version 1.1



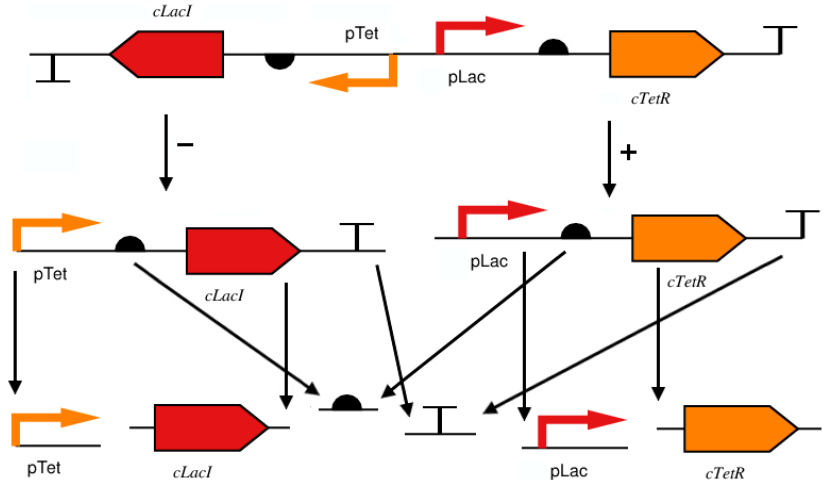
- Specification of DNA components.

Synthetic Biology Open Language (SBOL) Version 1.1



- Specification of DNA components.
- Hierarchical composition of DNA components.

Synthetic Biology Open Language (SBOL) Version 1.1



- Specification of DNA components.
- Hierarchical composition of DNA components.

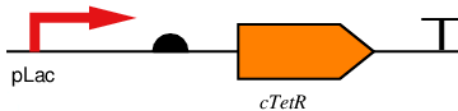
- Today's industry CAD session from 8:00-10:00 pm will feature a demo of SBOLv1.1-facilitated data exchange between multiple CAD tools.

Galdzicki, M. et al. The synthetic biology open language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat. Biotechnol.* 32 (2014), 545-550.

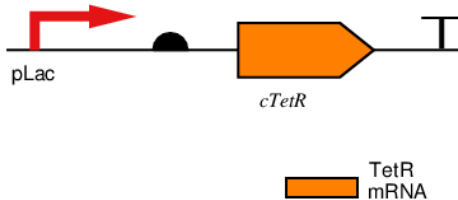
Goals for Next Version of SBOL

- Increase range of genetic structure and function that can be represented.
- Provide a basis for composing descriptions of genetic structure and function.

Increasing Structural Range

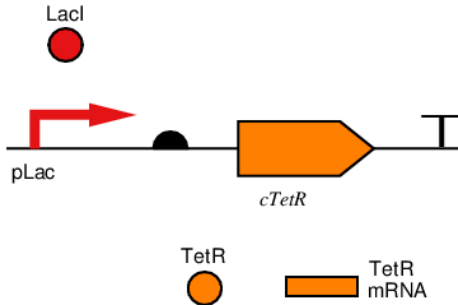


Increasing Structural Range



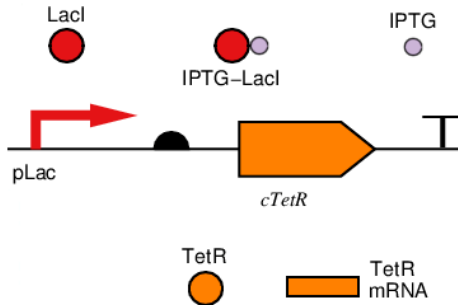
- RNA components (mRNA, tRNA, siRNA)

Increasing Structural Range



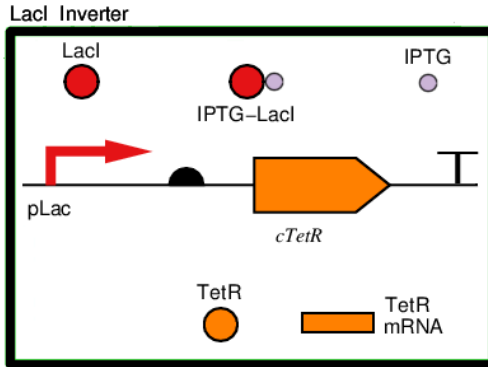
- RNA components
- Protein components (transcription factors, enzymes)

Increasing Structural Range



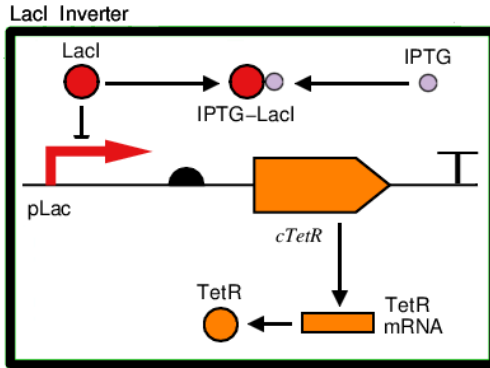
- RNA components
- Protein components
- Other Components (small molecules, molecular complexes, light)

Increasing Functional Range



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

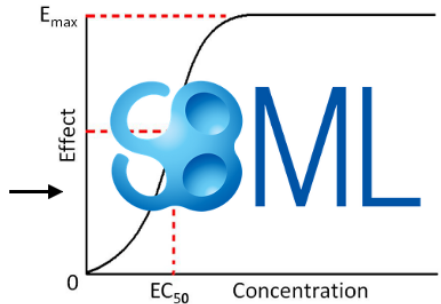
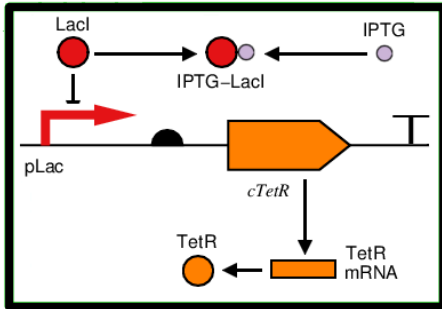
Increasing Functional Range



- Modules
- Interactions (activation, repression, complex formation, transcription, translation, phosphorylation)

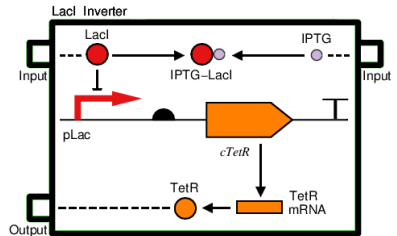
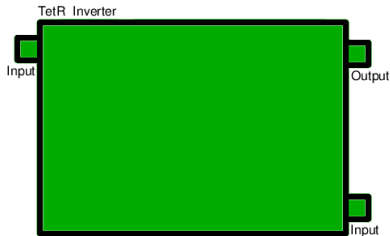
Increasing Functional Range

LacI Inverter



- Modules
- Interactions
- Models (SBML, CellML, MATLAB)

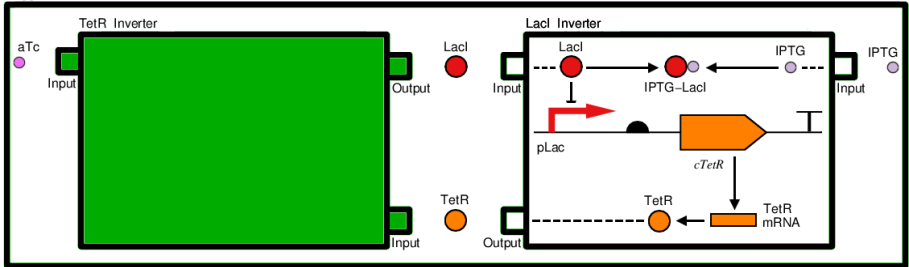
Basis for Composition of Structure and Function



- Ports

Basis for Composition of Structure and Function

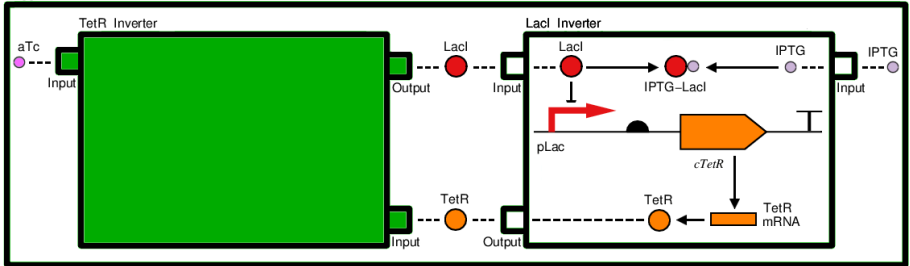
Toggle Switch



- Ports
- Instantiation and Mixed Hierarchy

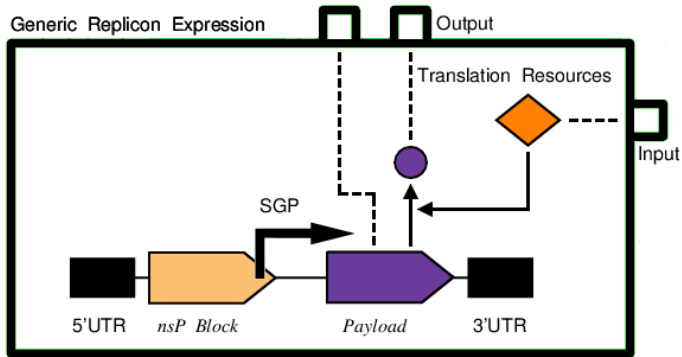
Basis for Composition of Structure and Function

Toggle Switch



- Ports
- Instantiation and Mixed Hierarchy
- Port Mapping

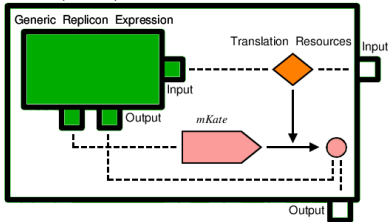
Example: RNA Replicons



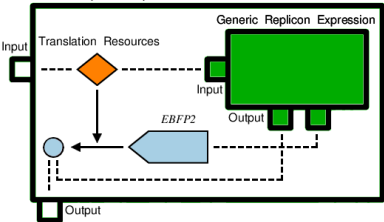
Beal, J., Wagner, T. E., Kitada, T., Azizgolshani, O., Parker, J. M., Densmore, D., and Weiss, R. Model-driven engineering of gene expression from RNA replicons. *ACS Synth. Biol.*, DOI: 10.1021/sb500173f

Example: RNA Replicons

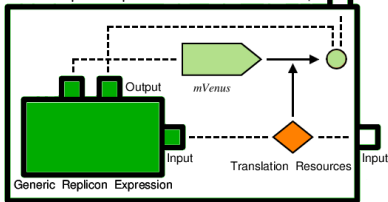
mKate Replicon Expression



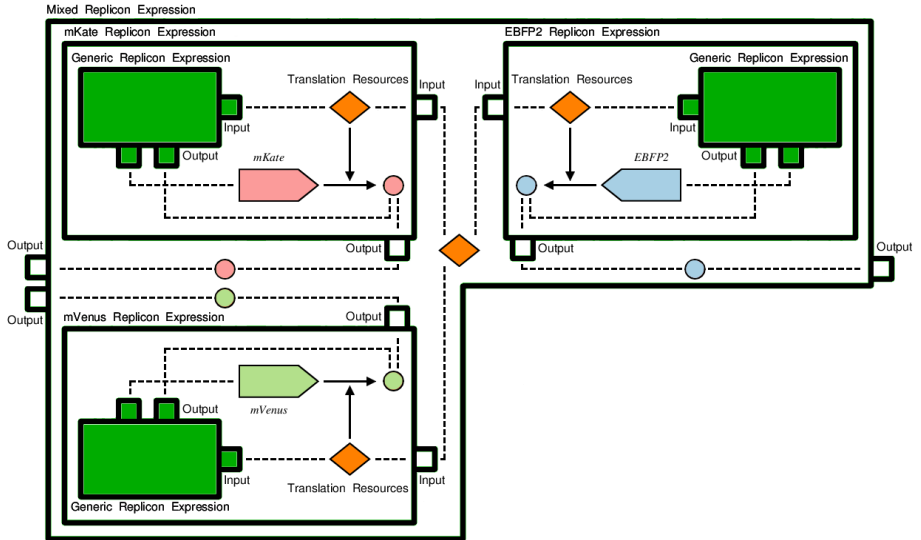
EBFP2 Replicon Expression



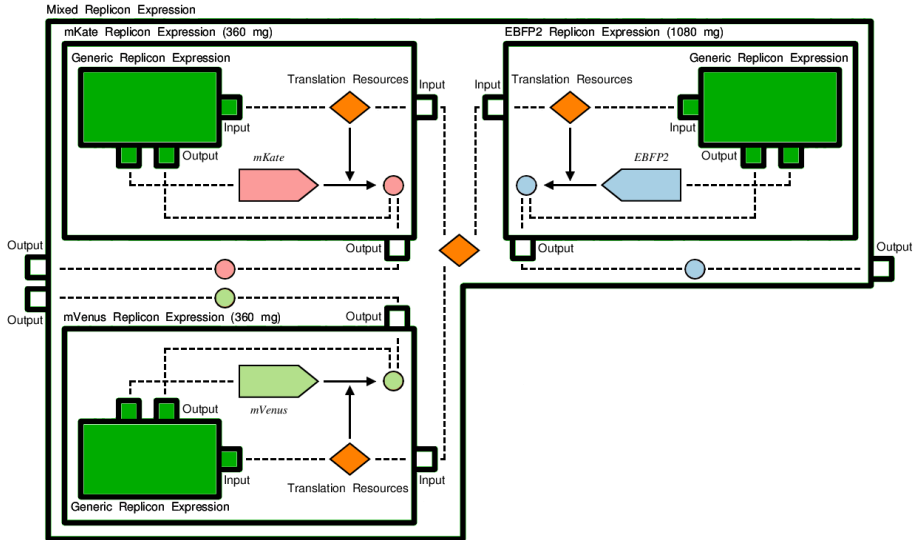
mVenus Replicon Expression



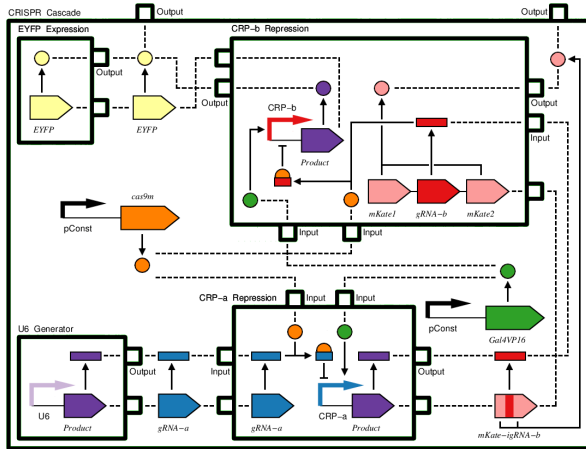
Example: RNA Replicons



Example: RNA Replicons

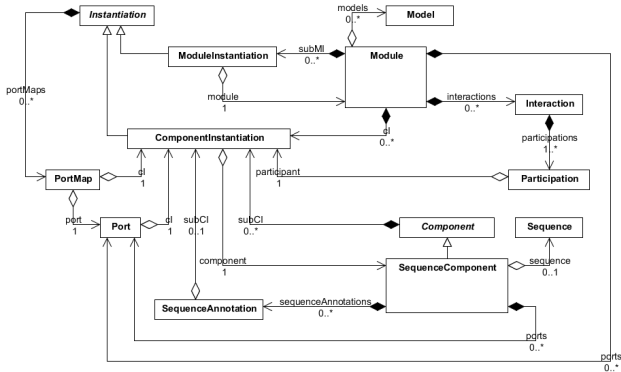


Example: CRISPR Cascade



Kiani, S., Beal, J., Ebrahimkhani, M. R., Huh, J., Hall, R. N., Xie, Z., Li, Y., and Weiss, R. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* 11 (2014), 723-726.

Proposed Data Model for the Next Version of SBOL



Roehner, N., Oberortner, E., Pocock, M., Beal, J., Clancy, K., Madsen, C., Misirli, G., Wipat, A., Sauro, H., and Myers, C. J. Proposed data model for the next version of the Synthetic Biology Open Language. *ACS Synth. Biol.*, DOI: 10.1021/sb500176h

Summary

- Proposal provides means to express and compose genetic designs exhibiting wide range of structure and function.
- Strikes balance between expressiveness and complexity.
- Avoids duplication of other standardization efforts where possible.

The Future

- As this data model is a proposal, we welcome community feedback.
- Construction of prototype for next version of `libSBOLj` has just begun.
- Details on joining SBOL effort at <http://www.sbolstandard.org/>.
- SBOL will still need to be extended to include data on environmental/host context, DNA synthesis, and characterization.

Acknowledgments



Agilent Technologies

CLARK  PARSIA 

Stardog

Raytheon

BBN Technologies



AUTODESK

life
technologies™



National Human
Genome Research
Institute

