

# Online Tools for Characterization, Design, and Debugging

Aaron Adler  
Raytheon BBN Technologies  
10 Moulton Street  
Cambridge, MA, USA 02138  
aadler@bbn.com

Fusun Yaman  
Raytheon BBN Technologies  
10 Moulton Street  
Cambridge, MA, USA 02138  
fusun@bbn.com

Jacob Beal  
Raytheon BBN Technologies  
10 Moulton Street  
Cambridge, MA, USA 02138  
jakebeal@bbn.com

## 1. MOTIVATION

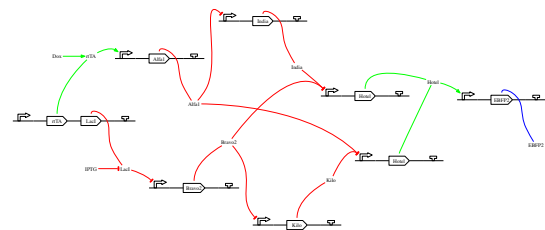
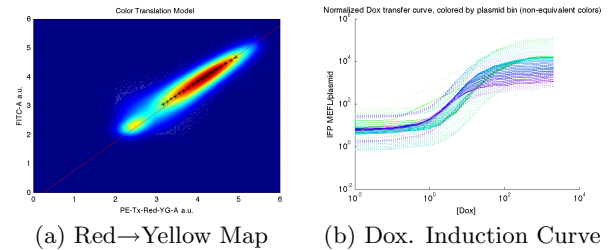
The engineering of biological systems can be greatly aided by better models, derived from high-quality characterization data, and by better means for designing and debugging new genetic circuits. Web-based tools and repositories have proven a successful approach to distributing such techniques, particularly because the centralization of infrastructure greatly decreases adoption cost for new users. Notable examples include the Parts Registry [8], the RBS calculator [10], GeneDesign [9], GenoCAD [4], BioFab [7], and JBEI ICE [6].

No prior web-based tools, however, have supported either analysis of characterization data or high-level design that can take advantage of such data. Previously, we have constructed a number of such tools during the course of building the TASBE end-to-end tool-chain for biological design [2]. We have now improved these tools to be more user friendly and broadly applicable, and placed them online as free web-based tools, embedded in a secure architecture to preserve data privacy.

## 2. TASBE WEB-BASED TOOLS

At the present time, the TASBE Tools website provides access to a suite of three tools: Color Models, Characterization Experiment, and BioCompiler. These tools are accessible online at <https://synbiotools.bbn.com>. Access is free, and may be done either anonymously or with a registered account that allows data to be kept private, as described below in Section 4.

The Color Models tool uses flow cytometry data from control samples to create a calibrated model of single-cell fluorescent expression. This tool implements the fluorescent calibration methodology described in [3], in which the arbitrary units produced by a flow cytometer are mapped into standardized FITC units. This requires four controls: two standard controls: (1) blank cells for estimating autofluorescence, and (2) single constitutive expression of each fluorescent color, and two non-standard controls: (3) the fluorescent beads typically used for calibration and maintenance of flow cytometers (used for mapping the cytometer's FITC channel to absolute units), and (4) co-expression controls with two or three colors independently expressed using identical promoters (used for determining equivalent fluorescent expression levels in context). Given these inputs, the tool produces a color model that can be used to translate

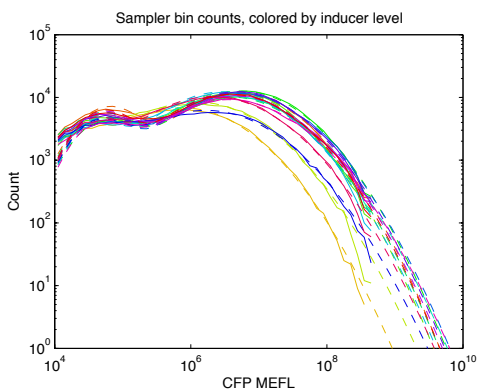


(c) Generated Exclusive-Or Circuit

**Figure 1: Example results from the three web services currently available on the TASBE tools site: (a) calibration of unit translation for a Color Model, (b) transfer curve (mean (solid lines) and standard deviations (dashed lines)) binned and normalized by constitutive expression from a Characterization Experiment, and (c) optimized transcriptional exclusive-or designed and visualized by BioCompiler.**

flow cytometer data into reproducible absolute units of measurement, as well as compensate for autofluorescence and spectral overlap. Figure 1(a) shows an example of the color model's collateral outputs: a translation model for mapping between red and yellow fluorescence.

The Characterization Experiment tool provides detailed analysis of single-variable flow cytometry experiments, again following the methodology presented in [3]. In particular, the tool computes the relationship between a controlled variable (e.g., time, inducer) and the statistical distributions of a constitutive marker and up to two other fluorescent proteins. Using a color model produced by the prior tool, the results are given in standardized MEFL units that are replicable between labs and experiments. This approach is particularly useful for transient transfections, where the number of



**Figure 2: TASBE tools reveal normally hidden problems in experimental data (solid lines; model fits shown as dashed lines), such as the three low-quality transfections (orange, light green, and dark blue) in this data set.**

plasmids per cell may vary by orders of magnitude, but has also been applied to tighter distributions, such as replicating plasmids or genomic integrations. Figure 1(b) shows an example of its outputs: an input/output curve for induction of rtTA with doxycycline, normalized by estimated plasmid count.

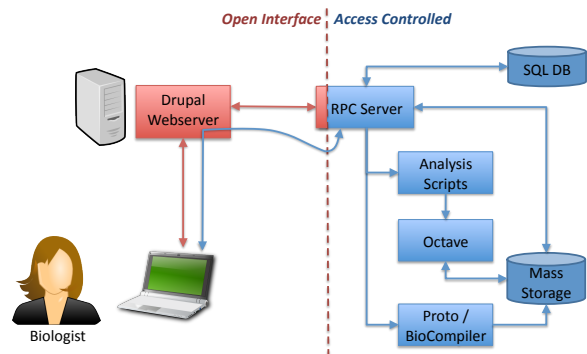
Finally, the Proto BioCompiler [1] takes high-level computational specifications and creates optimized genetic regulatory networks that implement those specifications. The newly designed network is then exported for the user in two standards-based formats—an SBOL [11] XML file specifying the design, and a visualization in GraphViz [5] using its built-in SBOLv symbols—plus Matlab files specifying an ODE-based simulation. Figure 1(c) shows an example of its outputs: a GraphViz visualization of a genetic regulatory network generated from a specification for a transcriptional exclusive-or circuit.

### 3. BIOLOGIST-FOCUSED DEBUGGING

A key challenge in biological experiments is differentiating between erroneous results that should be discarded and odd or unexpected results that bear further investigation. The TASBE tools encode knowledge to assist biologists in making this determination, by automatically examining color model and characterization experiment data for known classes of problems. These are then reported to the user in the form of warnings (e.g., of a more than 5% mismatch to the expected fluorescent bead distribution) or errors (e.g., such poor control fluorescence that color translation is not possible).

There are many ways that experiments can go wrong, and any human is likely to overlook some potential problems, especially subtle ones. A software tool reliably checks all its encoded failure cases, where a hurried experimenter may skip or overlook something. We provide an online biologist-focused manual that describes how the messages from the tools typically relate to issues in experimental protocols, and provides recommended strategies for addressing these issues.

The tools also take advantage of the internal cross-validation in high-resolution characterization data, presenting results in such a way as to reveal problems that are obscured by



**Figure 3: The architecture of the TASBE tools web service is designed to ensure data security and scalability.**

typical analyses. For example, Figure 2 shows constitutive distribution curves for a characterization experiment (experimental data shown as solid lines, model fits as dashed lines): three are clearly indicated by their relatively much lower curves as poor quality transfections that should be discarded, even though their data covers the approximately same overall range.

### 4. WEB SERVICE ARCHITECTURE

Two key requirements drove our architectural decisions for these web services. First and foremost is the necessity to ensure the biological data remains secure and private. Not only are there potential ethical, legal, and intellectual property considerations, but also biologists typically consider secrecy of data vital in order to avoid being scooped. Second is the need for scalability, which comes from the large requirements for data and computation when processing flow cytometry data; data sets are frequently more than a gigabyte in size and can take more than ten minutes to process.

We thus designed the architecture shown in Figure 3, executing on an SE Linux platform. Critically, no data is ever placed in a location accessible to the web server: instead, the web server is simply used for routing data on a secure and authenticated connection between the RPC server (which is not externally exposed) and the user’s browser. This means that no third party accessing the web server can obtain a user’s data without being able to authenticate as that user.

### 5. FUTURE DIRECTIONS

The TASBE tools are already being used by several laboratories, and we aim to continue improving the tools to better serve the needs of an expanding user community and according to priorities based on user feedback. Key expected improvements include batch file upload, laboratory and project data sharing, and incremental addition of new knowledge-driven debugging assistance. We also plan to continue adding more tools from the TASBE suite; near-term candidates include MatchMaker [12], control experiment planning, color selection tools, and multi-variable characterization.

### 6. ACKNOWLEDGMENTS

Example data produced by Noah Davidsohn and Ron Weiss.

## 7. REFERENCES

- [1] J. Beal, T. Lu, and R. Weiss. Automatic compilation from high-level biologically-oriented programming language to genetic regulatory networks. *PLoS ONE*, 6(8):e22490, August 2011.
- [2] J. Beal, R. Weiss, D. Densmore, A. Adler, E. Appleton, J. Babb, S. Bhatia, N. Davidsohn, T. Haddock, J. Loyall, et al. An end-to-end workflow for engineering of biological networks from high-level specifications. *ACS Synthetic Biology*, 1(8):317–331, 2012.
- [3] J. Beal, R. Weiss, F. Yaman, N. Davidsohn, and A. Adler. A method for fast, high-precision characterization of synthetic biology devices. Technical Report MIT-CSAIL-TR-2012-008, MIT, April 2012. Technical Report: MIT-CSAIL-TR-2012-008 <http://hdl.handle.net/1721.1/69973>.
- [4] M. J. Czar, Y. Cai, and J. Peccoud. Writing dna with genocad. *Nucleic Acids Research*, 37(Web Server issue):W40–W47, 2009.
- [5] J. Ellson, E. R. Gansner, E. Koutsofios, S. C. North, and G. Woodhull. Graphviz and dynagraph: static and dynamic graph drawing tools. In *GRAPH DRAWING SOFTWARE*, pages 127–148. Springer-Verlag, 2003.
- [6] T. S. Ham, Z. Dmytriv, H. Plahar, J. Chen, N. J. Hillson, and J. D. Keasling. Design, implementation and practice of jbei-ice: an open source biological part registry platform and tools. *Nucleic acids research*, 40(18):e141–e141, 2012.
- [7] V. K. Mutalik, J. C. Guimaraes, G. Cambray, C. Lam, M. J. Christoffersen, Q.-A. Mai, A. B. Tran, M. Paull, J. D. Keasling, A. P. Arkin, and D. Endy. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, 10:354–360, April 2013.
- [8] Registry of standard biological parts, Retrieved May, 2013.
- [9] S. Richardson, S. Wheelan, R. Yarrington, and J. Boeke. Genedesign: rapid, automated design of multikilobase synthetic genes. *Genome Res.*, 16(4):550–6, April 2006.
- [10] H. M. Salis, E. A. Mirsky, and C. A. Voigt. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27(10):946–950, 2009.
- [11] SBOL. Synthetic Biology Open Language. <http://www.sbolstandard.org>, January 2013.
- [12] F. Yaman, S. Bhatia, A. Adler, D. Densmore, and J. Beal. Automated selection of synthetic biology parts for genetic regulatory networks. *ACS Synthetic Biology*, 1(8):332–344, 2012.