ACS Synthetic Biology

Technical Note

pubs.acs.org/synthbio

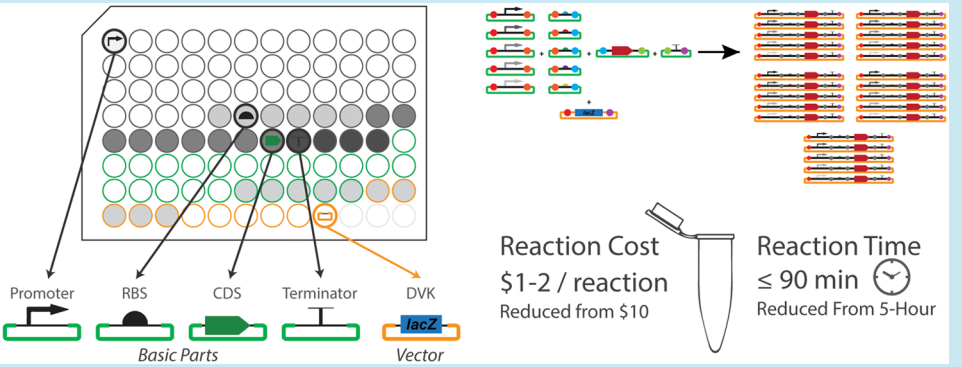
CIDAR MoClo: Improved MoClo Assembly Standard and New *E. coli* Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology

Sonya V. Iverson,[†,‡] Traci L. Haddock,[‡] Jacob Beal,[§] and Douglas M. Densmore*[,‡,∥]

[†]Molecular and Cellular Biology and Biochemistry, [‡]Biological Design Center (BDC), [∥]Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts 02215, United States

[§]Raytheon BBN Technologies, Cambridge, Massachusetts 02138, United States

Supporting Information



ABSTRACT: Multipart and modular DNA part libraries and assembly standards have become common tools in synthetic biology since the publication of the Gibson and Golden Gate assembly methods, yet no multipart modular library exists for use in bacterial systems. Building upon the existing MoClo assembly framework, we have developed a publicly available collection of modular DNA parts and enhanced MoClo protocols to enable rapid one-pot, multipart assembly, combinatorial design, and expression tuning in *Escherichia coli*. The Cross-disciplinary Integration of Design Automation Research lab (CIDAR) MoClo Library is openly available and contains promoters, ribosomal binding sites, coding sequence, terminators, vectors, and a set of fluorescent control plasmids. Optimized protocols reduce reaction time and cost by >80% from that of previously published protocols.

KEYWORDS: *assembly, modular, part library, Type IIS, multiplex*

Despite improvements in DNA assembly technologies, high-throughput iterative designs and combinatorial methods are still cost-prohibitive in synthesis-based assembly due to the lack of reusable parts. Multipart modular assembly part libraries have been developed for use in yeast[4] and mammalian systems;[1] however, no such library exists for *Escherichia coli*.

Modular cloning, or MoClo,[1] is a one-pot digestion and ligation multipart assembly method, derived from Golden Gate,[2] in which user-defined overhangs specific to each part type, such as a promoter or a coding sequence (CDS), create interchangeable DNA modules in the form of plasmids. This format allows for simple library propagation and combinatorial assembly from reusable parts with reliable ligation of up to six DNA fragments in a one-pot reaction. Three MoClo part libraries are currently available, providing reusable parts and vectors for plant transformation constructs[3] (Addgene, no. 1000000047), general eukaryotic multigene construct assembly[1] (Addgene, no. 1000000044), and yeast[4] (Addgene, no. 1000000061).

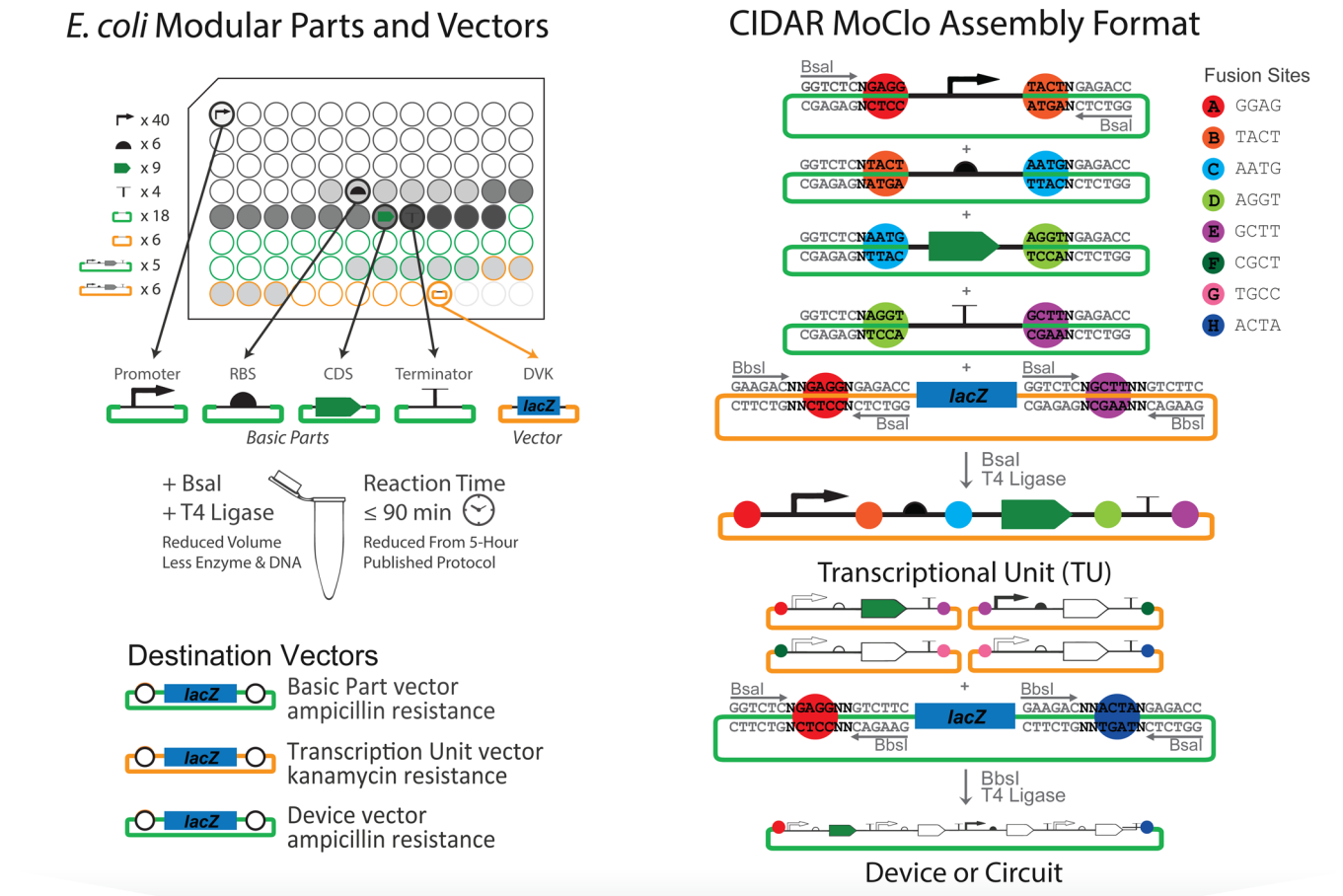
Recent publications have also detailed the development of yeast Golden Gate (yGG)[5] and BASIC assembly[6] methods, both of which employ a similar digestion and ligation reaction, although neither offers a part library to accompany the methods.

To address this lack of publically available standardize parts for bacterial systems, we have constructed and characterized a library of commonly used genetic parts and the necessary vectors in MoClo format for use in *E. coli* (Addgene, no. 1000000059).

MoClo relies upon Type IIS restriction enzymes (BbsI and BsaI) that each recognize a 6 bp nonpalindromic sequence and cut at a specified distance from that recognition sequence, resulting in modular overhanging 4 bp fusion sites (Figure 1).

Received: July 7, 2015

DOI: 10.1021/acssynbio.5b00124
ACS Synth. Biol. XXXX, XXX, XXX−XXX



CIDAR Workflow

Specification	Design	Build	Test	Share
Programming language for describing synthetic biology devices	Software tool for creating DNA assembly plans in common assembly formats	Collection of DNA part libraries based on CIDAR MoClo assembly.	Software tool to convert arbitrary fluorescent units to Methyl Equivilant Fluorescein (MEFL) physical units.	Inventory of Composable Elements (ICE), online database of sequences and characterization data.
Eugene	Raven	Hummingbird	TASBE Tools	CIDAR
www.eugenecad.org	www.ravencad.org	www.cidarlab.org/moclo	synbiotools.bbn.com	www.cidar-ice.org

Figure 1. CIDAR MoClo overview: basic assembly strategy. Left: The CIDAR MoClo Library is provided on a 96-well plate (Addgene, no. 1000000059). A standard CIDAR MoClo assembly is based on four-part transcription units composed of a promoter, ribosome binding site, coding sequence, and terminator assembled into a DVK (kanamycin-resistance destination vector). Destination vectors alternate antibiotic resistance at each assembly level and use *lacZα* blue-white selection. Right: Assembly strategy for MoClo Type IIS enzyme one-pot digestion ligation reactions. Basic parts shown with green backbones are prepared in DVAs with part-specific fusion sites, indicated by a single capital letter code. The 3′ fusion site of the upstream part must match the 5′ fusion site of the following part in order to be correctly assembled. Digesting parts and a DVK with the BsaI and simultaneous ligation with T4 DNA ligase results in a transcriptional unit (TU) shown here with the orange backbone. Multiple TUs can be combined into a complex device using BbsI in place of BsaI with the appropriate DVA. The library is described in more detail in Supporting Information Table S2.

For example, coding sequence (CDS) parts are flanked with -AATG- (abbreviated C) at the 5′ end and -ATTG- (abbreviated D) on the 3′ end, making all CDS parts interchangeable. Plasmid names include the fusion sites designated by single-letter abbreviations (ex. E0030m_CD) as in Figure 1. The restriction recognition sites are placed and oriented such that the digest product ends in these specific 4 bp overhangs and no longer contains the restriction enzyme recognition sequence. Once it is ligated, it cannot be recut, allowing for hierarchical multipart assembly (Figure 1).

The original MoClo protocol allowed for the reliable assembly of up to six parts with a 5 h digestion−ligation

DOI: 10.1021/acssynbio.5b00124
ACS Synth. Biol. XXXX, XXX, XXX−XXX

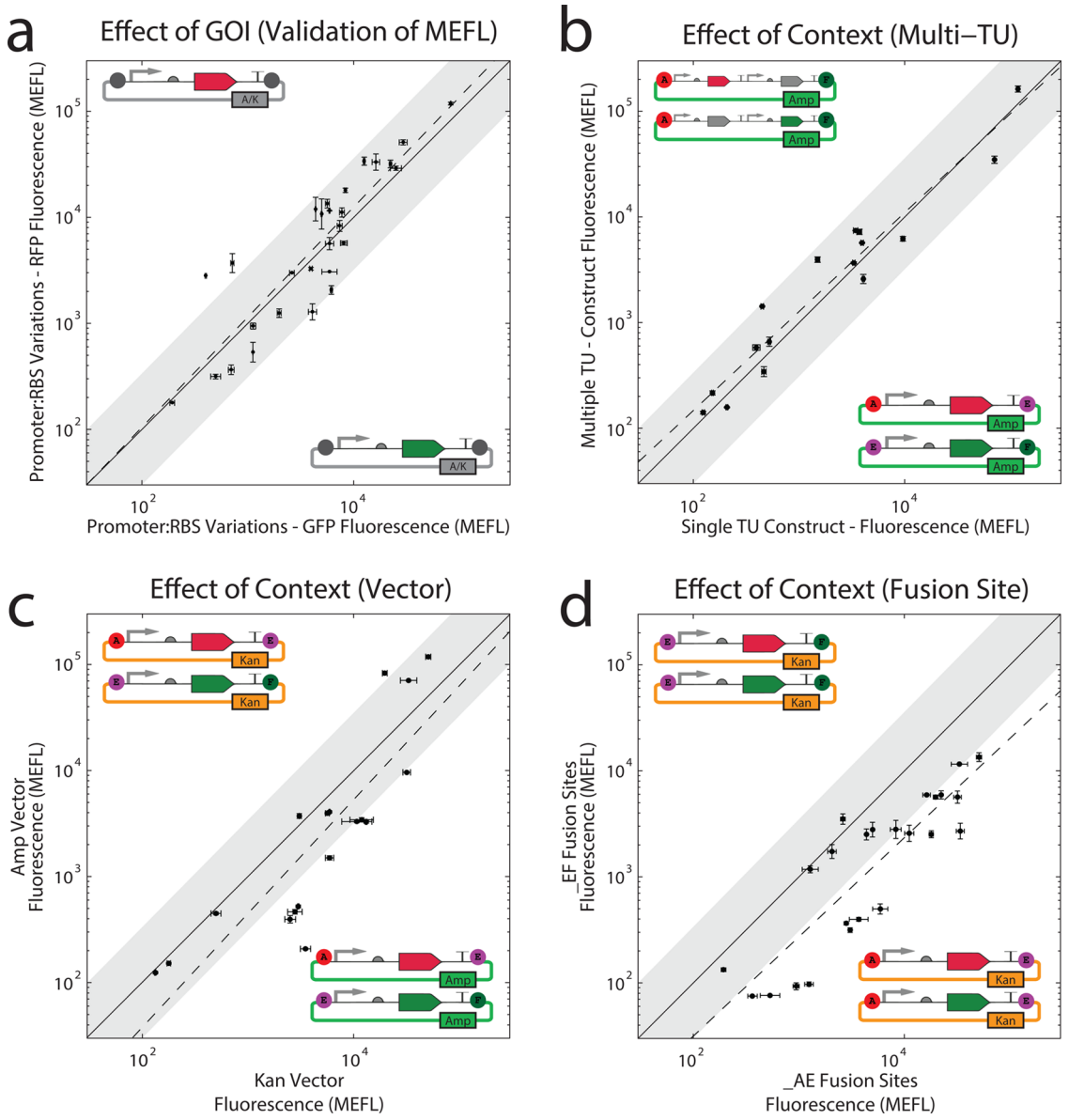


Figure 2. Molecules of equivalent fluorescein (MEFL) conversion and predictable expression. (a) SpheroTech RCP-30-5-A beads were used to provide a MEFL normalization of red and green fluorescence data in order to directly relate RFP to GFP in a rational unit. This normalization was validated by comparing 28 pairs of vectors in which each pair differed only in the choice of coding sequence, either GFP or RFP, in a transcriptional unit with a given promoters and RBS part. (b) Single TU expression compared to the same TU when expressed in the same plasmid with another TU. Expression of a single TU is consistent when it is assembled into a larger device. (c) Changing vector from DVA to DVK shows a high level of variability, suggesting that all expression tuning should be done in the same vector. (d) Changing the 4 bp fusion sites flanking a given TU influences expression likely due to the proximity of the 5′ fusion site to the minimal 35 bp promoter used in this study. The colored circles containing capital letters indicate the relevant 4 bp fusion sites, the sequences of which are noted in Figure 1.

reaction with large reagent volumes and rotating use of three antibiotics and two color selection modules.[6] Although this provided a substantial improvement in modularity compared to that of Golden Gate, the reaction conditions were still time- and cost-prohibitive for most applications. Here, we introduce a modified protocol (Supporting Information Table S1) that, while maintaining the same capacity for assembling multiple modules and a > 95% cloning efficiency, reduces reaction time from 5 h to 90 min and lowers reaction costs by 85% while simplifying the hierarchical assembly standard.

CIDAR MoClo uses an alternating series of vectors (Figure 1). Basic parts are provided in a destination vector-ampicillin (DVA) derived from BBa_pSB1A2, with BsaI sites flanking the part and fusion sites. These basic parts combine to create a simple transcriptional unit (TU) consisting of promoter:ribosome binding site (RBS):CDS:terminator in a kanamycin vector (DVK) derived from BBa_pSB1K3 with BbsI sites flanking the newly constructed TU. These units can be further combined into multi-TU devices in the same manner (Figure 1).

We created a standard four-part structure that allows users to easily assemble a genetic device consisting of a promoter, RBS, CDS, and terminator. This basic four-part format provides a solid foundation for high-throughput assembly while remaining adaptable to addition of new parts (i.e., new fusion site combinations) and part types.

DOI: 10.1021/acssynbio.5b00124
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Most of the parts contained in the CIDAR *E. coli* MoClo Library are derived from the BioBricks Registry (http://parts.igem.org/) and were selected for their functional reliability and utility in synthetic biological designs (see Supporting Information Table S2 for a full list of parts). To further enable rational design, three of the basic parts were selected from the BIOFAB collection of bicistronic design (BCD) translational elements that have been shown to enable more rational design in terms of protein expression.[7] These BCD parts contain a leader peptide followed by a secondary ribosome binding site (RBS) in order to physically separate transcriptional and translational regulation.

Destination vectors are included in the library to allow for simple cloning of new parts with any of the standard fusion site pairs. Additionally, a set of fluorescent expression plasmids is also included to be used as standards with TASBE flow cytometry analysis tools[8] (Supporting Information Table S2). Sequence and part information for all plasmids is available in the CIDAR Inventory of Composible Elements (ICE) registry[9] (www.cidar-ice.org). Other plasmids for use in *E. coli* not included in this library are available through the CIDAR ICE registry upon request.

To demonstrate the utility of this system, we used the library and protocols to perform a series of comparisons with red and green fluorescent proteins in an assay designed to provide a basis for expression tuning of complex devices. By normalizing flow cytometry fluorescence data to SpheroTech RCP-30-5-A beads[10] as a fluorescein standard, we created a color model with which to correlate RFP and GFP expression as molecules of equivalent fluorescein (MEFL) counts rather than arbitrary units.

This color model was validated by comparing MEFL counts of 28 pairs of plasmids with various promoter, RBS, and fusion site combinations (mean square error = 1.80-fold) in which each pair differed only in the coding sequence (E0040m_CD GFP or E1010m_CD RFP) (Figure 2a).

An array of GFP expressing TUs was made to further characterize a set of constitutive promoters to determine the extent of rational design capabilities. Using 16 promoters from the J23100 Anderson series and 6 RBS parts (3 Weiss RBS, 3 BCD), 96 iterations of a simple GFP expression plasmid in a DVK_AE vector were analyzed using flow cytometry (Supporting Information Figure S1). Two-way ANOVA analysis identified 36.3% of expression variation as being due to the promoter, 43.9% due to the RBS part, and 19.3% due to the interaction of the two factors. A subset of these promoters is included in the library and was selected to provide a full coverage of expression levels, as defined by GFP expression ranging from 10^2 to 10^5 MEFLs.

To further assess the capability for rational design under commonly varied genetic contexts, further pairwise comparisons were performed to evaluate the effects of gene order, variation in vector, and use of specific fusion sites flanking the transcriptional unit. Expression of a single TU was shown to remain constant when expressed in a plasmid with a second TU up- or downstream (Figure 2b) (mean sq. error = 1.53-fold). Changing the vector from DVA to DVK while maintaining the same transcriptional unit showed a higher variability yet retained a nearly linear relationship (mean sq. error = 2.61-fold) (Figure 2c).

Modifying the four base fusion sites did have an effect on expression ($p < 0.001$) in a paired analysis in the _EF transcriptional unit when compared to that of the _AE paired clone (mean sq. error = 2.02-fold) (Figure 2d). This difference may be due to the proximity of the 5′ fusion site to the simple promoter. Including an insulator upstream of the promoter sequences may mitigate this effect.

As a MoClo reaction is dependent upon equimolar ratios of each *part type*, a logical advance on the methodology is to multiplex reactions by adding various plasmids of the same part type at $1/n$ the concentration of each other part type, where n equals the total number of iterations. Multiplex MoClo (MMC) allows for an expanded utility including simultaneously screening a large number of iterations in parallel while retaining >95% cloning efficiency. Examples of this methodology include screening of variant sequences or modulating expression levels of a single transcriptional unit by multiplexing the promoter and/or RBS part (Supporting Information Figure S2).

This library, especially when combined with a related design tool,[11] allows for rapid assembly of synthetic gene networks and cost-efficient combinatorial assembly and is ideal for academic research and educational uses in teaching laboratory settings. Although it is formatted for cloning in *E. coli*, iterative design of eukaryotic circuits is also possible through the introduction of species-specific parts. Biodesign automation activities are further promoted with the included Eugene design files that capture not only the parts in the library but also initial design guidelines and data (Supporting Information Data). The CIDAR MoClo Library can be requested whole through Addgene (www.addgene.org/cloning/moclo/densmore), and additional individual plasmids are available for distribution through the CIDAR ICE Public Registry (www.cidar-ice.org) along with functional information and DNA sequence data for all parts described above.

METHODS

MoClo Cloning Protocols. In developing optimal protocols, various reaction conditions were tested. The following components were added to a 0.2 mL tube: 10−60 fmol of each DNA component with up to six components total, equimolar (PCR product or previously made MoClo DNA parts and the appropriate destination vector), 10−50 U of BsaI or BbsI (NEB), 5−50 U of T4 DNA ligase (cat. no. M1794, Promega, Madison, WI, USA, or no. M0202 NEB), 1× T4 DNA ligase buffer (Promega or NEB), and sterile, deionized water to at total volume of 10−60 μL. Reactions were performed using the following parameters: 15−40 cycles (37 °C 1.5−3 min, 16 °C 3−5 min), followed by 50 °C for 5 min and 80 °C for 10 min and were then held at 4 or −20 °C until they were transformed.

Optimized protocols are found in Supporting Information Table S1 using 10 fmol of each DNA component, 10 U of BsaI or BbsI, 20 U of T4 Ligase, 1× T4 DNA ligase buffer (Promega), final volume of 10−20 μL. Of this, 2−5 μL is used to transform cells as described above.

Multiplex MoClo Cloning Protocol. MoClo reactions were prepared as above with the following differences: The multiplex part type(s) was added such that the total concentration of that type was equimolar to the other part types. When multiplexing less than 6 of any one part type, samples were pipetted individually. To accommodate accurate measurements, multiplex reactions were performed in 20 μL volumes with 60 fmol of each part type. For example, in a reaction using 6 different promoters, 10 fmol each of these promoters was used along with 60 fmol of each other part. For

DOI: 10.1021/acssynbio.5b00124
ACS Synth. Biol. XXXX, XXX, XXX−XXX

larger multiplex examples all iterations of a given part type are mixed in equimolar ratio prior and added as one mixed part.

Statistical Analysis Methods. Flow cytometry data was converted from arbitrary units to compensated MEFL (molecules of equivalent fluorescein) using the TASBE characterization method.[13] An affine compensation matrix is computed from single color and blank controls: RFP alone (J23102:BCD2:E1010m:B0015, abbreviated as pJ02B2Rm, in _AE and EF DVAs), GFP alone (J23102:BCD2:E0040m:B0015 abbreviated as pJ02B2Gm, in _AE and EF DVAs), RFP:GFP (pJ02B2Rm:J02B2Gm in _AF DVAs) as well as the reciprocal GFP:RFP (pJ02B2Gm:J02B2Rm in _AF DVAs) together and untransformed DH5 Alpha Select *E. coli* cells (Bioline), respectively. FITC measurements (for GFP) are calibrated to MEFL using SpheroTech RCP-30-5-A beads.[10] An estimated mapping from RFP measured in the PE-Texas Red channel to equivalent FITC is computed from transformation of constitutive coexpression of RFP and GFP expressed together (RFP:GFP as pJ02B2Rm:J02B2Gm as _AF in DVA, GFP:RFP as pJ02B2Gm:J02B2Rm as _AF in DVA); RFP measurements are translated to MEFL by first mapping to estimated equivalent FITC. Geometric statistics are then computed over data in MEFL units.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.5b00124.

Additional methods, figures, and tables (PDF)

MEFL normalized data and plasmid descriptions for all samples represented in Figure 2 (XLSX)

CIDAR_MoClo_Eugene_Final.eug (a design file using the Eugene programming language) (TXT)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dougd@bu.edu.

Author Contributions

S.V.I., T.L.H., and D.M.D. conceived of the project. S.V.I., T.L.H., and J.B. designed the experiments; S.V.I. built the library, performed the experiments, analyzed data, and wrote the manuscript. J.B. performed computational analysis of flow cytometry data. D.M.D. and T.L.H. created the Eugene code. All authors edited the manuscript.

Notes

The authors declare the following competing financial interest(s): Douglas Densmore is the co-founder of Lattice Automation, Inc. Lattice creates bio-design automation software solutions.

■ ACKNOWLEDGMENTS

Thanks are extended to Dr. Mike Smanski for assistance with designing the MoClo library, the 2012 and 2013 iGEM teams for constructing various parts, Evan Appleton and Swati Carr for assistance in the lab, Dr. Swapnil Bhatia for continued support of Pigeon[12] (www.pigeoncad.org) used to make figures, Prashant Vaidyanathan for the configuration of the ICE repository, and the laboratories of Drs. James Collins, Wilson Wong, and Ahmad Khalil for use of miscellaneous equipment. This work was supported by NSF grant no. 1253856.

■ REFERENCES

(1) Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS One 6*, e16765.

(2) Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One 4*, e5553.

(3) Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D., Patron, N. J., and Marillonnet, S. (2014) A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synth Biol. 3*, 839−843.

(4) Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth Biol. 4*, 975−986.

(5) Agmon, N., Mitchell, L. A., Cai, Y., Ikushima, S., Chuang, J., Zheng, A., Choi, W. J., Martin, J. A., Caravelli, K., Stracquadanio, G., and Boeke, J. D. (2015) Yeast Golden Gate (yGG) for efficient assembly of S. cerevisiae transcription units. *ACS Synth Biol. 4*, 853−859.

(6) Storch, M., Casini, A., Mackrow, B., Fleming, T., Trewhitt, H., Ellis, T., and Baldwin, G. S. (2015) BASIC: a new Biopart Assembly Standard for Idempotent Cloning provides accurate, single-tier DNA assembly for synthetic biology. *ACS Synth Biol. 4*, 781−787.

(7) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods 10*, 354−360.

(8) Beal, J., Weiss, R., Densmore, D., Adler, A., Appleton, E., Babb, J., Bhatia, S., Davidsohn, N., Haddock, T., and Loyall, J. (2012) An End-to-End Workflow for Engineering of Biological Networks from High-Level Specifications. *ACS Synthetic Biology. 1*, 317−331.

(9) Ham, T. S., Dmytriv, Z., Plahar, H., Chen, J., Hillson, N. J., and Keasling, J. D. (2012) Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res. 40*, e141.

(10) (2001) Measuring molecules of equivalent florochrome (MEF) using SpheroTM rainbow and ultra rainbow calibration particles, *SPHERO™ Technical Note*, Spherotech, Inc., Lake Forest, IL.

(11) Appleton, E., Tao, J., Haddock, T., and Densmore, D. (2014) Interactive assembly algorithms for molecular cloning. *Nat. Methods 11*, 657−662.

(12) Bhatia, S., and Densmore, D. (2013) Pigeon: A Design Visualizer for Synthetic Biology. *ACS Synth Biol. 2*, 348−350.

(13) Beal, J., Weiss, R., Densmore, D., Adler, A., Appleton, E., Babb, J., Bhatia, S., Davidsohn, N., Haddock, T., Loyall, J., Schantz, R., Vasilev, V., and Yaman, F. (2012) An end-to-end workflow for engineering of biological networks from high-level specifications. *ACS Synth. Biol. 1*, 317−331.

DOI: 10.1021/acssynbio.5b00124
ACS Synth. Biol. XXXX, XXX, XXX−XXX