ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

# j5 DNA Assembly Design Automation Software

Nathan J. Hillson,*[,†,‡] Rafael D. Rosengarten,[†,‡,∥] and Jay D. Keasling[†,‡,§]

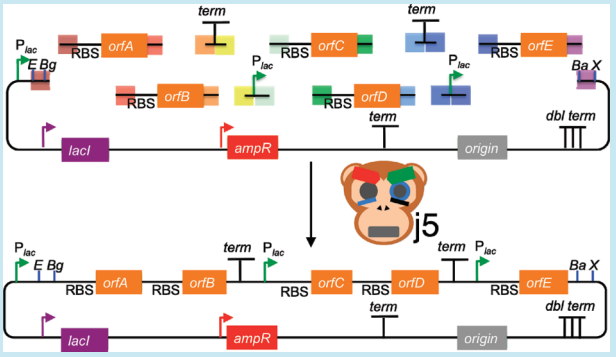
[†]Fuels Synthesis Division, Joint BioEnergy Institute, Emeryville, California 94608, United States

[‡]Physical Bioscience Division, Lawrence Berkeley National Lab, 1 Cyclotron Road Mail Stop 978R4121, Berkeley, California 94720, United States

[§]Department of Chemical & Biomolecular Engineering, Department of Bioengineering, University of California, Berkeley, California 94720, United States

Supporting Information

**ABSTRACT:** Recent advances in Synthetic Biology have yielded standardized and automatable DNA assembly protocols that enable a broad range of biotechnological research and development. Unfortunately, the experimental design required for modern scar-less multipart DNA assembly methods is frequently laborious, time-consuming, and error-prone. Here, we report the development and deployment of a web-based software tool, j5, which automates the design of scar-less multipart DNA assembly protocols including SLIC, Gibson, CPEC, and Golden Gate. The key innovations of the j5 design process include cost optimization, leveraging DNA synthesis when cost-effective to do so, the enforcement of design specification rules, hierarchical assembly strategies to mitigate likely assembly errors, and the instruction of manual or automated construction of scar-less combinatorial DNA libraries. Using a GFP expression testbed, we demonstrate that j5 designs can be executed with the SLIC, Gibson, or CPEC assembly methods, used to build combinatorial libraries with the Golden Gate assembly method, and applied to the preparation of linear gene deletion cassettes for *E. coli*. The DNA assembly design algorithms reported here are generally applicable to broad classes of DNA construction methodologies and could be implemented to supplement other DNA assembly design tools. Taken together, these innovations save researchers time and effort, reduce the frequency of user design errors and off-target assembly products, decrease research costs, and enable scar-less multipart and combinatorial DNA construction at scales unfeasible without computer-aided design.




**KEYWORDS:** *DNA assembly, design automation, BioCAD, combinatorial library*

Developing the ability to construct large and functionally complex DNA sequences, such as those encoding biosynthetic pathways, genetic circuits, partially synthetic chromosomes,[1] or synthetic genomes,[2] will be crucial for engineering microbes, plants and mammalian cells for vaccine, biofuel, and bio-based chemical production.[3] Recent advances in DNA assembly[4,5] have introduced protocols that offer substantial time- and cost-savings over traditional multiple cloning-site approaches, especially when constructing long DNA sequences that contain multiple genes. These new methods are automatable and standardized, that is, the same enzymatic reagents and conditions are used for every task.

Methods such as SLIC,[6] isothermal *in vitro* recombination (hereafter Gibson),[7,8] CPEC,[9,10] type IIs endonuclease mediated (hereafter Golden Gate),[11,12] USER,[13] and DNA Assembler[14,15] are scar-less, providing control over every DNA base pair, and enable more than two DNA fragments to be put together in a single step. These methods can offer benefits over BioBrick-style assembly,[16,17] for which 6 base pair scars result at every assembly junction and only two fragments can be assembled per step. In contrast with BioBrick assembly, however, designing optimized protocols for scar-less multipart DNA assembly methods is often tedious, laborious, and error-prone. Toward addressing this challenge, two recent methodological developments, MoClo[18] and GoldenBraid,[19] report consistent design patterns employing standardized subcloning steps for hierarchical Golden Gate-style assembly. While elegant, these techniques introduce predetermined sets of 4 bp assembly junction scars, may require elaborate plasmid libraries (MoClo employs approximately 35 intermediate vectors), and/or sacrifice full multipart assembly (GoldenBraid assembles only two fragments per step).

Here we report a new approach to the challenge of designing scar-less multipart DNA assembly protocols including the SLIC, Gibson, CPEC, and (combinatorial) Golden Gate methods. We have developed the web-based computer-aided design (CAD) software, "j5", to automate protocol design and optimization, while fully preserving scar-less and multipart assembly without prerequisite plasmid libraries.

**Received:** October 4, 2011
**Published:** December 7, 2011

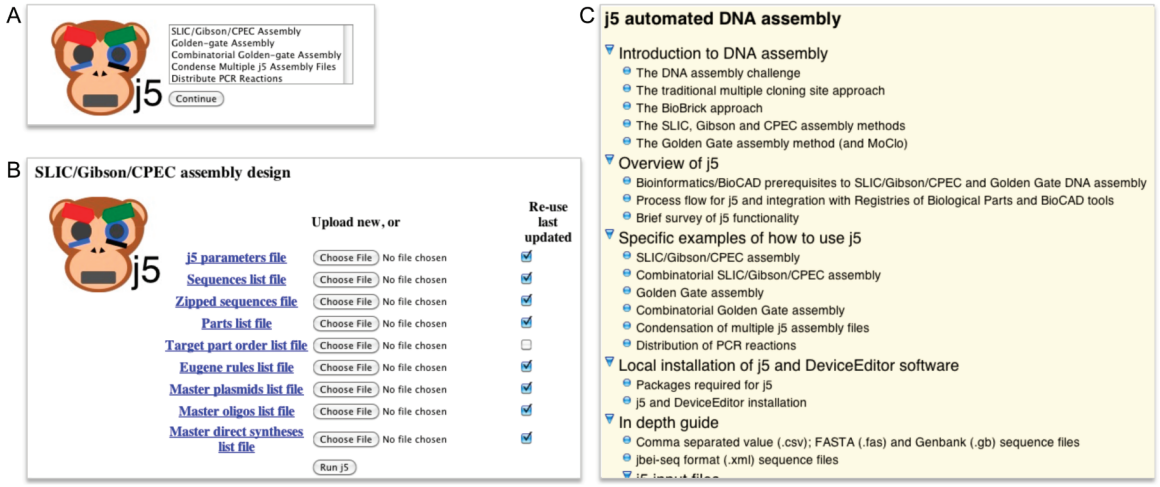



**Figure 1.** j5 web-based interface and user's manual. (A) Top level design task menu. (B) SLIC/Gibson/CPEC assembly design entry-form interface. A hyperlink to the user's manual provides a description of and a downloadable example for each input file type. For each input file, users may opt to upload a new file or to reuse the version they last updated on the server. (C) Online user's manual table of contents (truncated), providing a review of selected DNA assembly methodologies, an overview of j5 functionality, specific step-by-step examples of how to use j5, in-depth guides documenting input and output file specifications, etc.

Two factors, the decreasing price of DNA synthesis and the increasing demand for scalable construction of combinatorial DNA libraries, are now impacting the DNA assembly design process. As the price of DNA synthesis decreases, the cost of outsourcing end-to-end construction (or perhaps portions thereof) becomes comparable to that of in-house cloning. Researchers should now consider the cost-effectiveness and efficiency of each DNA construction task relative to commercial DNA synthesis. Even as inexpensive synthesis supplants single construct cloning, the exploration of numerous combinations of genes, protein domains, regulatory elements, etc. requires technology to enable the design and scar-less assembly processes. These combinatorial DNA libraries have become increasingly important, especially as a means of engineering fusion proteins and metabolic pathways toward the production of biofuels and biobased chemicals.[8,10,14]

As a way of understanding the utility of this approach, consider the example of constructing green fluorescent protein (GFP) with peptide tags specifying subcellular localization and degradation (Figure 3A). Selecting one of two N-terminal signal peptides, either a long or short linker sequence, and either a normal or enhanced C-terminal degradation tag, yields a total of 8 variant molecules (three variable components with two options each). With no *a priori* expectation of which variants might fold functionally, localize correctly, and degrade most efficiently, one must try them all. Leveraging a combinatorial assembly approach allows the researcher to reuse parts, such as a vector backbone, across multiple combinations instead of generating a custom version of each part for each distinct plasmid. Still, one must identify the optimal assembly junctions, design oligos to amplify the input components, and incorporate complementary overhangs and restriction sites. In a more ambitious example of a 10-gene pathway with 3 orthologs for each gene ($3^{10}$ or 59,000 variations), the design challenge is not only daunting but virtually impossible to tackle by hand.

To the best of our knowledge, j5 is the first DNA assembly design tool (for any assembly method including BioBricks) that recommends DNA synthesis when cost-effective to do so and has the capacity to direct the construction of scar-less combinatorial DNA libraries. j5 is also unique among scar-less DNA assembly design tools in its ability to perform cost-optimization, design combinatorial libraries or hierarchical assembly strategies to mitigate putative off-target assembly products, and to enforce design specification rules. For combinatorial DNA libraries of around 200 plasmids, the time-savings can be 3- to 8-fold over traditional approaches, and the cost-savings can be 10- to 20-fold over DNA synthesis services (see Supplementary Table S1).

## RESULTS AND DISCUSSION

**j5 DNA Assembly Design Automation Software.** j5 automates the cost-optimal design of scar-less multipart DNA assembly protocols including SLIC,[6] Gibson,[7] CPEC,[9] and Golden Gate.[12] j5 is web-based, available across computer platforms via a common web-browser interface (Figure 1A,B), and as such does not require the user to install or update the software. j5 also provides XML-encoded Remote Procedure Calling protocol over HTTP (XML-RPC) web-services, enabling alternative graphical user interfaces or third-party applications to exploit the full j5 feature set. An online user's manual (Figure 1C) provides a brief review of DNA assembly methodologies, an overview of j5 functionality, step-by-step how-to examples, in-depth descriptions of input and output files, detailed documentation of the j5 XML-RPC web-services API, error-message explanations, and experimental protocols for the aforementioned DNA construction techniques.

To begin the j5 DNA assembly design process, the user first selects the assembly methodology for j5 to design, namely, SLIC/Gibson/CPEC, Golden Gate, or combinatorial Golden Gate (Figure 1A). Next, the user defines the biological "parts" (here synonymous with DNA sequences) to be assembled. The input format is a comma separated value (CSV) file that can be manipulated by any spreadsheet (e.g., Excel, OpenOffice, etc.) or text editor software, as shown in Figure S1A (see Supporting Information online). Each part is defined by a start and an end base pair within a source sequence and by an orientation on the top or bottom strand. Since j5 designs for assembly methods that do not require predetermined flanking sequences, the defined parts do not need to be "packaged" in any particular


dx.doi.org/10.1021/sb2000116 | ACS Synth. Biol. 2012, 1, 14–21

format, in contrast to BioBrick assembly.[16,17] The source DNA sequence files, user-specified (as shown in Figure S1B) and incorporated into a single ZIP file, may be in community-standard FASTA and Genbank formats or alternatively the jbei-seq (XML) format of the JBEI-ICE repository platform. After defining the parts to be assembled, the user then sequentially orders and sets the direction (forward or reverse) of each of the parts in the final target construct(s) (as shown in Figure S1C for a single construct and in Figure S1D for a combinatorial library). The user may also dictate Eugene biological design specification rules.[20,21] These rules can limit the total number of times a given part may appear in a given construct, prevent any two parts from appearing together in the same construct, ensure that two given parts always appear together in the same construct (see Figure S1E), and are of particular utility when only a subset of all possible combinatorial variants is desired. For example, if prior research demonstrated that a long linker sequence must follow tag *sig1* for proper GFPuv localization (see Figure 3A), Eugene rules can be specified to ensure that *sig1* and the long linker are always constructed together, eliminating the two (of the eight) possible combinations that have the tag *sig1* followed by the short linker. j5 enforces these rules by designing assemblies only for those constructs that satisfy the specifications.

To determine the most cost-effective assembly strategy, j5 uses Algorithm S1, based on the user-adjustable cost analysis parameters shown in Figure S1F. Given an ordered list of parts to assemble, Algorithm S1 utilizes alternative cost comparison and iterative DNA synthesis fragment extension to determine for each part if direct DNA synthesis, PCR, or oligo embedding (i.e., direct incorporation of a small part into the primer that will be used to amplify the adjacent part) is the best approach from a monetary cost standpoint (see Figure S1G). Algorithm S1 precludes researchers from wasting several days upfront to the design of a cost-ineffective PCR-based assembly strategy. For the design shown in Figure 3A, Algorithm S1 eliminated 2591 alternate assembly strategies with more expensive combinations of DNA synthesis, PCR, and oligo embedding. The user alternatively may specify the strategy for generating a given part (e.g., the vector backbone will be restriction digested instead of PCR amplified; see Figures 2A and S1C).

To design flanking "homology sequences" for the SLIC, Gibson, and CPEC methods that direct assembly between neighboring parts with complementary ends, j5 uses Algorithms S2 and S3. For the Gibson and CPEC methods specifically, the homology sequences prime neighboring assembly pieces for polymerase extension. j5 utilizes Primer3[22] to optimize these homology sequences as primers for their neighboring assembly pieces while adhering to the user's specifications, such as the minimum homology sequence length (see Figure S1F). Primer3 does not return a sequence design if none of the sequences it considers meets its design specifications. When this happens, it is necessary to adjust the design constraint (e.g., self-complementarity) parameters until at least one sequence is acceptable. Algorithm S2 removes this burden from the user, ensuring that Primer3 generates an optimized design for each homology sequence by progressively relieving any design constraints (issuing warning messages as it does so) that lead to the rejection of all considered sequences. For the design shown in Figure 2A, Algorithm S2 progressively relieved Primer3 of 6 design constraints and eliminated 396 less optimal homology sequence pairs. The flanking sequence to append to each terminus of each part, then, is the portion of the optimized

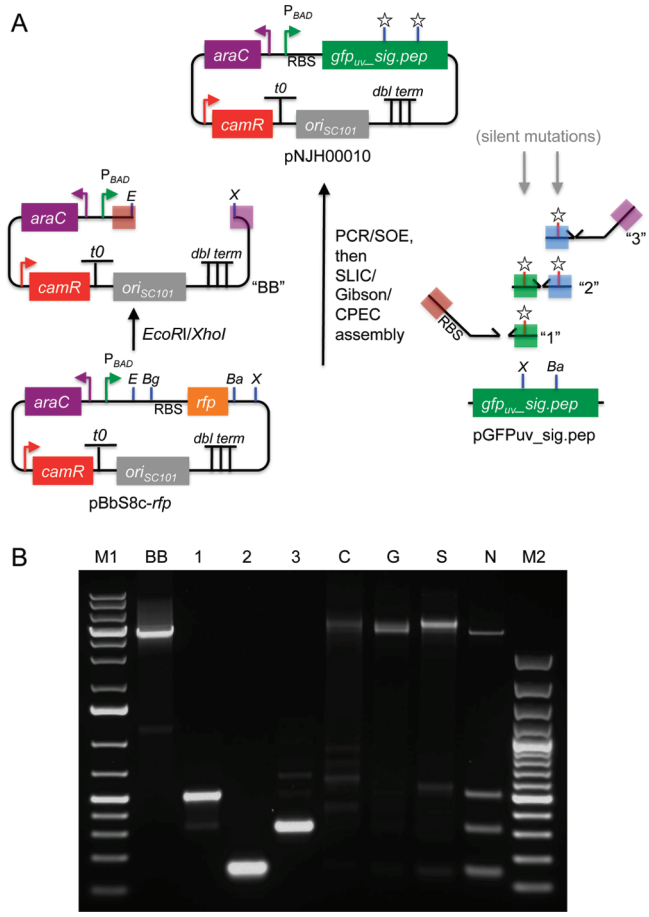



**Figure 2.** SLIC/Gibson/CPEC assembly design. (A) Schematic of the SLIC/Gibson/CPEC DNA assembly task. The vector pBbS8c-*rfp*[25] is double digested with *Eco*RI/*Xho*I and the vector backbone fragment "BB" is gel-purified away from the *rfp* insert. Two silent mutations (marked with stars, removing internal *Xho*I and *BamH*I sites) are introduced into $gfp_{uv}$*_sig.pep* via primer-templated mutagenesis of plasmid pGFPuv_sig.pep, splitting $gfp_{uv}$*_sig.pep* into three overlapping fragments, "1", "2", and "3" (sequence homology depicted in green and light blue). Sequence homologous to the 3′ end of the digested vector backbone (brown) and a RBS are introduced into the 5′ end of fragment "1" via the forward PCR primer. Similarly, sequence homologous to the 5′ end of the vector backbone (purple) is introduced into the 3′ end of fragment "3". The four DNA fragments are then assembled with SLIC, Gibson, or CPEC assembly to yield plasmid pNJH00010. (B) DNA electrophoresis of the four DNA fragments, and the resulting DNA assembly reactions. Lane 1: 1 kb+ DNA ladder "M1"; lane 2: digested vector backbone "BB"; lanes 3−5: PCR amplified fragments "1", "2", and "3"; lanes 6−8: CPEC "C", Gibson "G", and SLIC "S" assembly reactions; lane 9: negative "N" assembly reaction control; lane 10: 100 bp DNA ladder "M2".

homology sequence derived from the adjacent part(s). Once the flanking sequences have been designed, j5 utilizes BLAST[23] to check for assembly pieces that are incompatible with each other (i.e., unintentionally share highly similar flanking sequences), which can lead to undesired assembly products (see Figure S2A,B). If any BLAST-identified incompatible sequences exceed a user-specified $T_m$ threshold (see Figure S1F), Algorithm S3 identifies contiguous sets of compatible assembly pieces and then designs a hierarchical assembly process, mitigating the risk of assembling off-target products (see Figure S2C) and bolstering correct assembly efficiency.

To design cohesive single stranded overhangs to guide the Golden Gate method assembly process, j5 uses Algorithm S4. The type IIs endonuclease (e.g., BsaI) cleavage-generated overhang sequences at each end of each part should not be self-cohesive nor anneal to off-target overhangs (see Figure S2D). For each assembly junction, there may be multiple overhang sequences to choose from that would result in the same assembly product (see Figure S2E). "Neutral" positioned overhang sequences (i.e., centered on part boundaries) are preferable, since (at least for PCR-derived assembly pieces) non-neutral overhangs result in longer, and thus more expensive, DNA oligos. For scar-less combinatorial Golden Gate assembly, the set of putative overhang sequences is bounded at the 5′ and 3′ ends of each assembly junction by the first nonidentical base pairs among the combinatorial variants. Algorithm S4 searches through all combinations of putative overhangs and selects the set of overhang sequences that are compatible with themselves and each other, are as neutral as possible, and satisfy a user-determined maximum number of off-target overhang base-pair matches (see Figures S1F, S2D). Algorithm S4 uses a branched search strategy that is pruned to avoid redundant paths and paths that will not lead to compatible sets of overhangs. For the two-fragment (two overhang) combinatorial library design shown in Figure 3A, it was necessary to evaluate 25 overhang combinations before identifying the optimal compatible set of overhang sequences. We are currently pursuing a more complicated metabolic pathway combinatorial library design requiring 11 assembly junctions, for which it was necessary to evaluate over 2.4 *billion* overhang combinations before finding the selected set of sequences. Without the use of Algorithm S4, identifying the compatible set of overhang sequences for this metabolic pathway design would not be possible. With the overhang sequences selected, the flanking sequence to append to each terminus of each part is a type IIs endonuclease motif (including buffer sequence, a recognition site, and a spacer sequence; all user-specifiable, see Figure S1F) followed by any sequence from the overhang to the part boundary derived from adjacent part(s).

To optimize the design of template-priming sequences for those parts for which the most cost-effective approach is PCR, j5 utilizes Primer3 (applying Algorithm S2) while adhering to the user's specifications (such as the GC-clamp length, see Figure S5). For the design shown in Figure 3A, Algorithm S2 eliminated 720 less optimal, template-priming, DNA oligo pairs. The designed full-length DNA oligos are concatenations of flanking homology or overhang-generating sequences and template-priming sequences. For those parts for which the most cost-effective approach is direct DNA synthesis, the synthesis products are the specified assembly parts flanked with homology or overhang-generating sequences. j5 checks whether any of the designed DNA oligos or synthesis fragments can be reused from the user's existing collection (decreasing research costs) before automatically naming, iteratively numbering, and appending to the user's collection any new DNA oligos and direct synthesis fragments to be ordered (see Figure S3A,B). j5 then outputs the required PCR reactions (Figure S3C), details the set of DNA assembly pieces (for SLIC/Gibson/CPEC assembly, providing the length and $T_m$ of each flanking homology sequence, Figure S3D; for Golden Gate assembly, providing the sequence of each overhang, Figure S3E), and specifies the subset of pieces to be assembled into each combinatorial variant (Figure S3F). Finally, j5 appends the

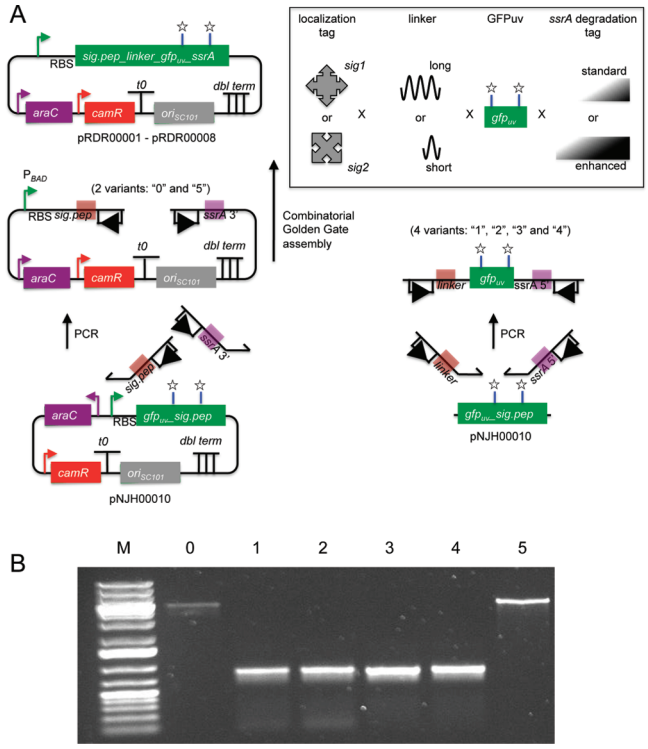



**Figure 3.** Combinatorial Golden Gate assembly design. (A) Schematic of a portion of the combinatorial Golden Gate DNA assembly task. The vector backbone of pNJH00010 is PCR amplified from just after the *gfp*$_{uv}$*_sig.pep* coding sequence to just before the *gfp*$_{uv}$*_sig.pep* coding sequence, with the forward primer introducing a BsaI recognition site (schematically depicted as a rectangle labeled with an arrowhead pointing to the 4-bp Golden Gate overhang sequence, here shown in purple) and the 3′ portion of the *ssrA* degradation tag, and the reverse primer introducing either the *sig1* or *sig2* localization tag and a BsaI recognition site (the Golden Gate overhang sequence shown here in brown), resulting in fragments "0" (*sig1*) and "5" (*sig2*), respectively. The *gfp*$_{uv}$ coding sequence of pNJH00010 is PCR amplified, with the forward primer introducing a BsaI recognition site (the Golden Gate overhang sequence shown here in brown) and either the long or short Gly/Ser linker, and the reverse primer introducing the 5′ portion of either the standard or enhanced (NYNY) *ssrA* degradation tag[27] and a BsaI recognition site (the Golden Gate overhang sequence shown here in purple), resulting in fragments "1" (long/standard), "2" (long/enhanced), "3" (short/standard), and "4" (long/enhanced), respectively. The two vector backbone fragments "0" and "5" are then combinatorially assembled with the insert fragments "1", "2", "3", and "4", resulting in the 8 plasmid variants pRDR00001−pRDR00008. (B) DNA electrophoresis of the combinatorial Golden Gate DNA fragments. Lane 1: 1kb+ DNA ladder "M"; lanes 2−7: PCR amplified fragments "0" through "5".

user's plasmid collection (see Figure S3G) and prepares a Genbank format sequence file for each of the designed assembly products. Annotations from the input source files are preserved in the output Genbank files, providing a rapid means to visual design validation (Figure S4). This is a tremendous time-saving and error-reducing feature, since the user must otherwise copy and paste the sequence fragments (including feature annotations) for each construct.

Toward enabling a single person or automation platform to execute an entire laboratory's worth of DNA assembly projects in parallel, j5 aggregates multiple independent designs into a single consolidated assembly file. Algorithm S5 makes it easy for the user to take advantage of thermocycler annealing


dx.doi.org/10.1021/sb2000116 | *ACS Synth. Biol.* 2012, 1, 14−21

temperature gradient features, which can positively affect PCR performance but are often underutilized due to the effort required to select the appropriate temperatures and place the PCR reactions accordingly. The user inputs a consolidated assembly file, the multiwell plate locations and volumes of the requisite DNA templates and primers (Figure S5A), and a set of user-adjustable parameters (Figure S5B). j5 then uses Monte Carlo simulated annealing Algorithm S5 to optimize the distribution of the PCR reactions required for the collective assembly process across multiwell plates (Figure S5C) within optimized thermocycler block annealing temperature gradients (Figure S5D), as schematically depicted in Figure S5E. j5 also generates the corresponding PCR setup control file for the NextGen (eXeTek) Expression workstation liquid-handling robotics platform (Figure S5F). Control files for other robotics platforms, such as the Tecan EvoLab, is an ongoing endeavor in our group.

In summary, we have implemented Algorithms S1−S5 and other features in our j5 DNA assembly design automation software that not only save researchers and their host institutions time, effort, and expense but also enable scar-less multipart and combinatorial DNA construction at scales feasible only with computer-aided design software.

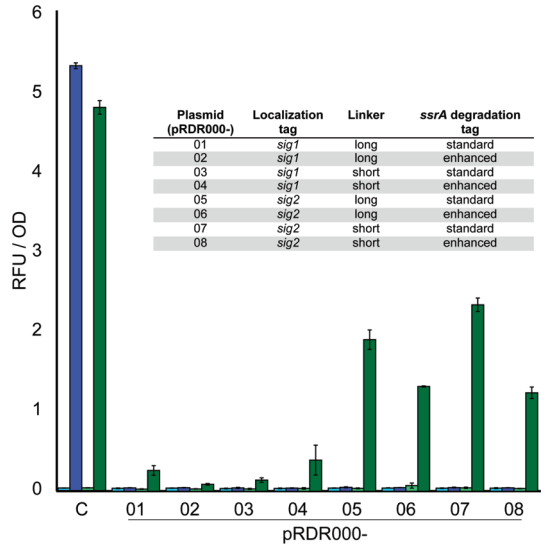
**SLIC/Gibson/CPEC Assembly Design with j5 and Plasmid Construction.** To show that j5 can design assembly protocols for the SLIC,[6] Gibson,[7] and CPEC[9] methods, plasmid pNJH00010 was designed as a four fragment assembly, introducing two silent mutations into $gfp_{uv}$_*sig.pep* and placing this modified gene under the control of the $P_{BAD}$ promoter (Figure 2A). For each of the three methods, DNA electrophoresis of the completed assembly reaction shows the depletion of the four j5-designed input fragments and the emergence of a larger assembly product, compared with the no-assembly reaction negative control (Figure 2B). Colony PCR screening of *E. coli* DH10b transformed with the assembly reaction products revealed that for each of the three methods, all (8/8) randomly screened colonies were consistent with the desired assembly product (Figure S7A−C). These results demonstrate that j5 can be used to design successful SLIC, Gibson, and CPEC protocols and that these three assembly methods can be used interchangeably if j5 design methods are used.

**Combinatorial Golden Gate Assembly Design with j5 and Plasmid Library Construction.** To evaluate the ability to use j5 to design combinatorial protocols for a variant of the Golden Gate[11,12] method, a library of eight plasmids (pRDR00001−pRDR00008) was designed, each consisting of two DNA fragments. These flanked $gfp_{uv}$ with sequences encoding one of two varieties of a localization tag (sig1 and sig2), glycine/serine linker (short and long) and *ssrA* degradation tag (regular and enhanced), and placing these modified gene combinations under the control of the $P_{BAD}$ promoter (Figure 3A). The fragments to be assembled were PCR-derived, contrasting with the previously reported Golden Gate approach[11,12] that utilizes plasmid-borne fragments. DNA electrophoresis of the six j5-designed, PCR amplified fragments to be assembled is shown in Figure 3B. Colony PCR screening of *E. coli* DH10b transformed with the assembly reaction products revealed that for each combination, all (4/4) randomly screened colonies contained the desired assembly product (Figure S7D). These results demonstrate that j5 can be used to design successful combinatorial Golden Gate variant protocols.

**Linear Gene Deletion Cassette Assembly Design with j5 and *clpX* Protease Markerless Deletion.** In preparation for assessing the ClpX protease dependence of the assembled $gfp_{uv}$ variant library, the construction of the linear *clpX* deletion cassette JPUB_000253 (Figure S8A−C) was designed with j5. Briefly, sequence fragments homologous to the *E. coli* DH10b *clpX* genomic locus were assembled with a portion of plasmid pSKI[24] containing markers for selection and counter-selection, and a homing endonuclease motif for marker excision, into a linear deletion cassette. Following a previously described strategy[24] schematically depicted in Figure S8, this deletion cassette was exploited to accomplish the markerless deletion of *clpX* (Figure S8D,E), demonstrating that j5-designed linear cassette assembly can be successfully applied to markerless gene deletion efforts.

**Experimental Characterization of GFPuv Variant Library.** To assess ClpX protease dependence, the control plasmid pNJH00010 (Figure 2) along with the assembled $gfp_{uv}$ library pRDR00001−pRDR00008 (Figure 3) were transformed into modified *E. coli* DH10b Δ*araFGH* Δ*araE* $P_{CP18}$::araE backgrounds, for which gene expression from the $P_{BAD}$ promoter is linear with arabinose concentration and uniform across the induced cellular population,[25] in the absence or presence of Δ*clpX*. The resulting strains were conditionally induced with arabinose, and the relative GFPuv fluorescence was measured for each plasmid variant for each genetic background for each induction condition (Figure 4). Consistent with previous reports, there was very little detectable GFPuv fluorescence without arabinose induction for any of the strains.[25] The fluorescence of the control GFPuv (lacking a *ssrA* degradation tag) was not dramatically affected by the deletion of *clpX*.[26] In the presence of functional *clpX*, little fluorescence was observed in any of the *ssrA*-tagged variants,[26] while in the Δ*clpX* background, GFPuv fluorescence of the *ssrA*-tagged variants was readily detected, albeit at lower levels than the control.[26] The GFPuv fluorescence of the enhanced *ssrA*-tagged variants was lower than their standard *ssrA*-tagged counterparts (with the exception of plasmids pRDR00003 and pRDR00004).[27] The GFPuv fluorescence of the *sig1*-tagged variants was consistently lower than their *sig2*-tagged counterparts.

**Summary and Conclusion.** While automated DNA construction design and optimization has been recently reported for BioBrick assembly,[28−30] designing optimized protocols for scar-less multipart DNA assembly methods has remained tedious, laborious, and error-prone. MoClo[18] and GoldenBraid[19] address this challenge through the use of consistent Golden Gate style designs that introduce predetermined sets of assembly junction scars, require elaborate plasmid libraries and/or sacrifice multipart assembly. To circumvent these limitations, j5 encompasses computer-aided design (via Algorithms S1−S5) that automate protocol design and process optimization as part of an integrated synthetic biology platform (Figures 5 and S9), while fully preserving scar-less and multipart assembly without prerequisite plasmid libraries. j5 can be used on its own or in conjunction with BioBrick, MoClo, GoldenBraid, and Reiterative Recombination,[31] where j5 is utilized to design the construction of the BioBricks, "Level 0 modules" (MoClo), "Level α entry-points" (GoldenBraid), or "Donor plasmids" (Reiterative Recombination). Although j5 does not currently design protocols for DNA Assembler,[14,15] USER,[13] or combinatorial assembly protocols for SLIC, Gibson[8] or CPEC,[10] the algorithms developed here




| Plasmid (pRDR000-) | Localization tag | Linker | *ssrA* degradation tag |
|---|---|---|---|
| 01 | *sig1* | long | standard |
| 02 | *sig1* | long | enhanced |
| 03 | *sig1* | short | standard |
| 04 | *sig1* | short | enhanced |
| 05 | *sig2* | long | standard |
| 06 | *sig2* | long | enhanced |
| 07 | *sig2* | short | standard |
| 08 | *sig2* | short | enhanced |

**Figure 4.** Experimental characterization of the assembled GFPuv variants. GFPuv variant plasmids (pNJH00010 control "C", pRDR00001−pRDR00008) were transformed into either "wild-type" (Keasling-1484; *E. coli* DH10b Δ*araFGH* Δ*araE* $P_{CP18}$::araE[25]) or Δ*clpX* (JBEI-3083; Keasling-1484 Δ*clpX*) backgrounds. The resulting strains (JBEI-2804; JBEI-2747 to 2749; JBEI-2751 to 2753; JBEI-2755; JBEI-2923; JBEI-3144; JBEI-3133 to JBEI-3140) were grown in quadruplicate in 2 mL 96-well plates on an orbital shaker at 37 °C at 900 rpm in 1 mL of LB media supplemented with 30 μg/mL chloramphenicol, conditionally induced in exponential phase at $OD_{600}$ = 0.2 with 10 mM arabinose, grown for an additional 6 h, pelleted, washed twice with M9 medium, and then assayed for GFPuv fluorescence signal using a SpectroMax Plus384 (Molecular Devices) plate-reader. The relative fluorescence units per $OD_{600}$ (RFU/OD) are shown for each GFPuv variant for each strain background (wild-type, blue; Δ*clpX*, green) matrixed with each growth condition (−arabinose, light blue/green; +arabinose, dark blue/green). Error bars show the standard error of four biological and two technical replicates. Inset table presents the localization tag, linker, and *ssrA* degradation tag combination for each plasmid.

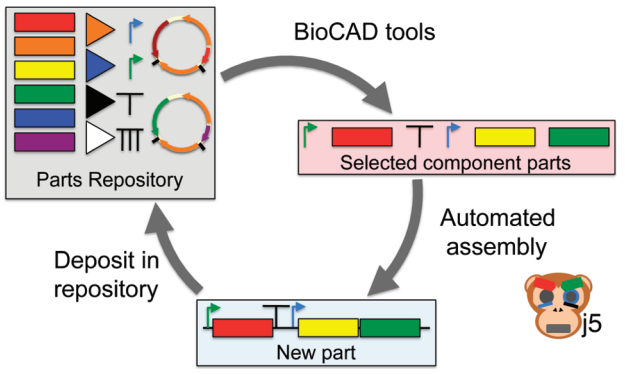



**Figure 5.** j5 DNA assembly design automation as part of an integrated Synthetic Biology design-implement-assay cycle.

are broadly applicable to, and under development to support, other classes of methods. These algorithms could also supplement other recently reported scar-less DNA assembly design tools (such as GeneDesign,[32] PHUSER,[33] Gibthon (http://gibthon.org), and DNA Oligo Designer (http://escience.invitrogen.com/oligoDesigner)). Exploring the entire combinatorial space of fusion proteins, overexpression schemes, genetic pathways, etc. has become a valuable tool for metabolic engineers.[34] Only a subset of the correctly assembled and sequence validated GFP localization tag variants (Figure 3A) constructed here displayed readily detectable levels of protease-dependent fluorescence (Figure 4). These data demonstrate the utility of employing a combinatorial approach to identify assemblies of genetic elements yielding a functional DNA device.

As prices fall, DNA synthesis is anticipated to play an increasingly large role in DNA construction. This makes j5's cost-optimal DNA assembly design process, which considers alternative strategies leveraging DNA synthesis (via Algorithm S1), an especially timely innovation. While DNA synthesis (at $\gtrsim$ US$0.35/bp) is not currently more cost-effective than the schemes depicted in Figures 2A and 3A, this strategic calculation might have a different outcome in cases requiring codon optimization or extensive sequence modifications or in the future as technological developments further reduce the cost of chemical synthesis. Future advances in DNA assembly methodology will also significantly impact the cost-optimal process calculus and drive the continual development of j5. In the near term, however, it is unlikely that the end-to-end synthesis of each individual DNA construct in a large combinatorial library (consisting of thousands to millions of variants) will be cost-effective. Instead, we speculate that DNA fragments (such as the 6 shown in Figure 3) will be individually synthesized, subsequently liberated (if necessary) from their cloning plasmids with a type IIs endonuclease, and then combinatorially assembled. Lacking an intervening PCR step, this envisioned process would not incur extensive sequence validation costs, although correct assembly junctions would still need to be confirmed. Thus, j5's combinatorial assembly protocol design (leveraging DNA synthesis) may emerge as the most valuable feature. Software tools like j5 may enable DNA synthesis companies to offer combinatorial DNA assembly services and may also make possible the integration of combinatorial DNA library construction, clonal transformation, and functional assay into an affordable benchtop device.

Finally, j5 specializes in DNA assembly protocol design and as such is not intended to facilitate the biological design of the DNA to be assembled. For example, j5 does not assist the selection of the genetic expression systems or metabolic enzymes to be assembled into functional biosynthetic pathways. However, j5's nonproprietary input and output text file formats facilitate interoperability with independent tools (e.g., the RBS Calculator[35] and GLAMM[36]) that do support biological function design. Furthermore, j5's open web-service interface enables full-feature plug-in development for all-in-one biological design platforms such as Clotho.[29] Thus, j5 is well-positioned to participate in emerging Synthetic Biology computer-aided design frameworks, whether all-in-one integrated environments or those favoring collections of independent specialist tools.

## METHODS

**j5 Software License and Availability.** j5 is available at no cost to noncommercial (e.g., academic, nonprofit, or government) users, under a Lawrence Berkeley National Lab end-user license agreement (http://j5.jbei.org/index.php/License). The software is available through the public j5 webserver (http://j5.jbei.org) and is also available for download upon request. Commercial use is available through the Technology Transfer Department of Lawrence Berkeley National Laboratory (ttd@lbl.gov).


dx.doi.org/10.1021/sb2000116 | *ACS Synth. Biol.* 2012, 1, 14–21

**j5 Software Implementation.** Mediawiki software (http://www.mediawiki.org) coupled with a PostgreSQL database (http://www.postgresql.org/) serves to automate the creation and maintenance of user accounts on the public j5 web-server. Perl-CGI web-form entry provides an interface to j5 (Figures 1A,B), although XML-RPC web-services and command-line interfaces are also available. j5 is written in the Perl programming language (http://www.perl.org/) and heavily draws upon the BioPerl[37] package as well as modules from the Comprehensive Perl Archive Network (CPAN, http://www.cpan.org) repository. j5 makes external calls to Primer3,[22] for primer and flanking homology sequence design, and to BLAST,[23] for identifying putative mis-priming and flanking homology sequence incompatibility events (see Results and Discussion). Circus Ponies Notebook software (http://www.circusponies.com/) was used to compose and generate the online j5 user's manual (Figure 1C).

**Strain and Sequence Availability.** *E. coli* strains (JBEI-2747 to 2749, JBEI-2751 to 2753, JBEI-2755, JBEI-2804, JBEI-2923, JBEI-2948, JBEI-3083, JBEI-3133 to JBEI-3140, and JBEI-3144) and DNA sequences (pNJH00010, pRDR00001–pRDR00008, and deletion cassette JPUB_000253), along with their associated information (annotated Genbank-format sequence files, j5 assembly design files including DNA oligo sequences, and sequencing trace files) have been deposited in the public instance of the JBEI Registry (https://public-registry.jbei.org) and are physically available from the authors and/or addgene (http://www.addgene.org) upon request. Additional details of plasmid and strain construction and functional characterization, beyond that described in the Results and Discussion section and in Figures 2–4 and S8, are provided Supporting Information.

## ■ ASSOCIATED CONTENT

### Supporting Information

Supporting tables, methods, algorithms, and figures. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

*Tel: +1 510 486 6754. Fax: +1 510 486 4252. E-mail: njhillson@lbl.gov.

### Present Address

[‖]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030.


### Author Contributions

N.J.H. designed and developed the software, N.J.H and R.D.R designed the experiments, R.D.R performed all experiments, N.J.H. wrote the software user's manual, and N.J.H, R.D.R, and J.D.K wrote the manuscript.


### Funding

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy (Contract No. DE-AC02-05CH11231); and the Berkeley Laboratory Directed Research and Development Program (to N.J.H.).


### Notes

The authors declare the following competing financial interest(s):The authors declare competing financial interests in the form of pending patent applications whose value may be affected by the publication of this article.

## ■ ACKNOWLEDGMENTS


The authors thank James N. Kinney for providing plasmid pGFPuv_sig.pep and the *sig1* and *sig2* localization tag sequences; Taek Soon Lee for providing plasmid pBbS8c-rfp; John W. Thorne for assistance constructing plasmid pNJH00010 and pSKI; Anna Chen for assistance constructing plasmids pRDR00001-pRDR00008; Chris Fields for incorporating proposed changes into the Primer3Redux BioPerl package; David Pletcher, Steve Lane, Zinovii Dmytriv, Ian Vaino, and William Morrell for providing information technology support; and Timothy Ham, James Carothers, and Vivek Mutalik for constructive comments on the manuscript.


## ■ REFERENCES


(1) Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L., Boeke, A. C., Gottschling, D. E., Chandrasegaran, S., Bader, J. S., and Boeke, J. D. (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature 477*, 471–476.

(2) Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z. Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A. 3rd, Smith, H. O., and Venter, J. C. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science 329*, 52–56.

(3) Nielsen, J., and Keasling, J. D. (2011) Synergies between synthetic biology and metabolic engineering. *Nat. Biotechnol. 29*, 693–695.

(4) Ellis, T., Adie, T., and Baldwin, G. S. (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol. 3*, 109–118.

(5) Hillson, N. J. (2011) DNA Assembly Method Standardization for Synthetic Biomolecular Circuits and Systems, in *Design and Analysis of Bio-molecular Circuits* (Koeppl, H., Densmore, D., di Bernardo, M., Setti, G., Eds.) 1st ed., pp 295–314, Springer-Verlag, Dordrecht.

(6) Li, M. Z., and Elledge, S. J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods 4*, 251–256.

(7) Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343–345.

(8) Ramon, A., and Smith, H. O. (2011) Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol. Lett. 33*, 549–555.

(9) Quan, J., and Tian, J. (2009) Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One 4*, e6441.

(10) Quan, J., and Tian, J. (2011) Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc. 6*, 242–251.

(11) 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.

(12) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One 3*, e3647.

(13) Bitinaite, J., Rubino, M., Varma, K. H., Schildkraut, I., Vaisvila, R., and Vaiskunaite, R. (2007) USER friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res. 35*, 1992–2002.

(14) Shao, Z., Luo, Y., and Zhao, H. (2011) Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol. Biosyst. 7*, 1056–1059.

(15) Shao, Z., and Zhao, H. (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res. 37*, e16.
(16) Anderson, J. C., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. A., Arkin, A. P., and Keasling, J. D. (2010) BglBricks: A flexible standard for biological part assembly. *J. Biol. Eng. 4*, 1.
(17) Shetty, R. P., Endy, D., and Knight, T. F. Jr. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng. 2*, 5.
(18) 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.
(19) Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juarez, P., Fernandez-Del-Carmen, A., Granell, A., and Orzaez, D. (2011) GoldenBraid: An iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One 6*, e21622.
(20) Bilitchenko, L., Liu, A., Cheung, S., Weeding, E., Xia, B., Leguia, M., Anderson, J. C., and Densmore, D. (2011) Eugene—a domain specific language for specifying and constraining synthetic biological parts, devices, and systems. *PLoS One 6*, e18882.
(21) Bilitchenko, L., Liu, A., and Densmore, D. (2011) The Eugene language for synthetic biology. *Methods Enzymol. 498*, 153—172.
(22) Rozen, S., and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol. 132*, 365—386.
(23) Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *J. Comput. Biol. 7*, 203—214.
(24) Yu, B. J., Kang, K. H., Lee, J. H., Sung, B. H., Kim, M. S., and Kim, S. C. (2008) Rapid and efficient construction of markerless deletions in the *Escherichia coli* genome. *Nucleic Acids Res. 36*, e84.
(25) Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L., and Keasling, J. D. (2001) Homogeneous expression of the P(BAD) promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology 147*, 3241—3247.
(26) Farrell, C. M., Grossman, A. D., and Sauer, R. T. (2005) Cytoplasmic degradation of ssrA-tagged proteins. *Mol. Microbiol. 57*, 1750—1761.
(27) Hersch, G. L., Baker, T. A., and Sauer, R. T. (2004) SspB delivery of substrates for ClpXP proteolysis probed by the design of improved degradation tags. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12136—12141.
(28) Densmore, D., Hsiau, T. H., Kittleson, J. T., DeLoache, W., Batten, C., and Anderson, J. C. (2010) Algorithms for automated DNA assembly. *Nucleic Acids Res. 38*, 2607—2616.
(29) Xia, B., Bhatia, S., Bubenheim, B., Dadgar, M., Densmore, D., and Anderson, J. C. (2011) Developer's and user's guide to Clotho v2.0 A software platform for the creation of synthetic biological systems. *Methods Enzymol. 498*, 97—135.
(30) Leguia, M., Brophy, J., Densmore, D., and Anderson, J. C. (2011) Automated assembly of standard biological parts. *Methods Enzymol. 498*, 363—397.
(31) Wingler, L. M., Cornish, V. W. (2011) Reiterative Recombination for the in vivo assembly of libraries of multigene pathways, *Proc. Natl. Acad. Sci. U.S.A.*
(32) Richardson, S. M., Nunley, P. W., Yarrington, R. M., Boeke, J. D., and Bader, J. S. (2010) GeneDesign 3.0 is an updated synthetic biology toolkit. *Nucleic Acids Res. 38*, 2603—2606.
(33) Olsen, L. R., Hansen, N. B., Bonde, M. T., Genee, H. J., Holm, D. K., Carlsen, S., Hansen, B. G., Patil, K. R., Mortensen, U. H., and Wernersson, R. (2011) PHUSER (Primer Help for USER): a novel tool for USER fusion primer design. *Nucleic Acids Res. 39*, W61—67.
(34) Santos, C. N., and Stephanopoulos, G. (2008) Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr. Opin. Chem. Biol. 12*, 168—176.
(35) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol. 27*, 946—950.
(36) Bates, J. T., Chivian, D., and Arkin, A. P. (2011) GLAMM: Genome-Linked Application for Metabolic Maps. *Nucleic Acids Res. 39*, W400—405.
(37) Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., Fuellen, G., Gilbert, J. G., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C. J., Osborne, B. I., Pocock, M. R., Schattner, P., Senger, M., Stein, L. D., Stupka, E., Wilkinson, M. D., and Birney, E. (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res. 12*, 1611—1618.