In Situ Overlap and Sequence Synthesis During DNA Assembly

Bernhard Paetzold,*[,†,‡] Carlo Carolis,[‡,§] Tony Ferrar,[†,‡] Luis Serrano,[†,‡,‖] and Maria Lluch-Senar*[,†,‡]

[†]EMBL-CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain
[‡]Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, 08003 Barcelona, Spain
[§]Biomolecular Screening and Protein Technologies Unit, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain
[‖]Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain

Ⓢ *Supporting Information*

ABSTRACT: Modern cloning methods are independent from restriction enzyme recognition sites. However, nearly all current cloning methods still require the introduction of overlaps by PCR, which can introduce undesired mutations. Here, we investigated whether overlaps needed for DNA assembly can be synthesized *in situ* and we tested if the *de novo* synthesis of sequences can be simultaneously combined with

the assembly of larger double-stranded DNA fragments. We showed in a set of 44 cloning experiments that overlaps of 20 bp needed for DNA assembly can be synthesized *in situ* from single-stranded oligonucleotides. Short sequences of 30−255 bp can be synthesized from single-stranded oligonucleotides concurrently with DNA assembly, and both techniques can be combined. The assembly of similar constructs by state-of-the-art techniques would have required multiple rounds of cloning or tedious sample preparations, whereas our approach is a one-step reaction.

KEYWORDS: *de novo synthesis, in situ DNA assembly, cloning, oligonucleotides, overlaps*
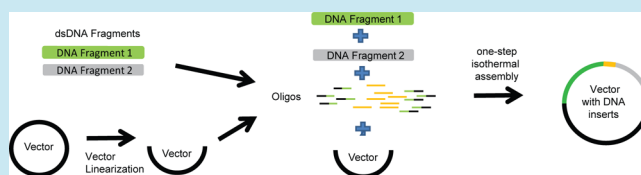
S ince the first restriction-enzyme-based cloning was performed in the early 1980's,[1] the techniques used for molecular cloning have made significant advancements. Restriction enzymes were the method of choice for molecular cloning for decades. However, restriction cloning requires the presence or absence of enzyme recognition sites, thereby complicating the design of DNA constructs and limiting high-throughput assembly. These problems were solved with the invention of sequence- and ligation-independent cloning (SLIC).[2] The SLIC method was the first sequence-independent cloning method reported. SLIC allows for the easy generation of scarless constructs and the assembly of DNA fragments independent of restriction sites. It is based on the introduction of overlaps between the vector and insert by PCR. These overlaps are processed to single-stranded overhangs by the 3′ to 5′ exonuclease activity of T4 polymerase and are annealed with their homologue partner strand. The annealed construct is transformed into bacteria, which subsequently ligate the construct. One drawback of this technique is that the exonuclease activity is difficult to control and tends to produce longer single-stranded regions than required. As result, single-stranded gaps, which are inefficiently repaired by the host, are introduced in the final construct, thus impairing cloning efficiency.

Gibson et al. improved the SLIC method further and created a very popular cloning method.[3] The main difference in the new method is the use of an enzyme mix containing an exonuclease to produce single-stranded ends, a high-fidelity polymerase to fill in single-stranded gaps, and a heat-stable ligase to ligate the final constructs. This method, called one-step isothermal assembly, can be used to clone double-stranded DNA (dsDNA) fragments seamlessly, even when they are hundreds of kilobases long,[3] or to synthesize short sequences (200−450 bp) from single-stranded oligonucleotides.[4] We hypothesized that it must be possible to combine both applications of the one-step isothermal assembly and thereby simplify existing cloning workflows even further.

Currently, all sequence-independent cloning methods available for *Escherichia coli* need homologous overlaps for the assembly of the fragments, which are usually introduced by primers in a PCR reaction. This can be challenging in cases where PCR amplification and subsequent sequencing is problematic. In addition, it requires a new PCR reaction for each overlap needed during subcloning, increasing the number of steps involved in sample preparation. Herein, we show that it is possible to synthesize these overlaps *in situ* from oligonucleotides during the cloning assembly. The *in situ* synthesis of overlaps suppresses the need for redundant PCRs each time a DNA fragment is subcloned into a new vector. We further show that it is simultaneously possible to assemble *de novo* sequences from oligonucleotides and to clone long DNA fragments. This approach allows the effortless and flexible introduction of modified DNA stretches from 30−255 bp. These modifications can be used to enhance protein expression or to characterize a protein of interest by the combination of

Table 1. Detailed Statistics of Nine Constructs from Assembling One dsDNA Insert and De Novo Synthesis of a Sequence Between 30−255 bp[a]

construct	hit rate colony PCR (%)	sequencing result mutated/failed/ correct	correct clones (%)	DNA synthesis length	errors total	error per bp	insertions/deletions/ mutations
13	58.33	3/0/1	25.00	87	5	0.057	4/1/0
15	41.67	2/0/2	50.00	114	2	0.017	2/0/0
22	83.33	1/0/3	75.00	96	2	0.020	2/0/0
19	66.67	1/1/2	50.00	69	1	0.014	1/0/0
30	66.67	2/0/2	50.00	96	3	0.031	1/0/2
31	83.33	3/0/1	25.00	96	5	0.052	3/0/2
36	41.67	0/0/4	100.00	46	0	0	0/0/0
46	75.00	1/0/3	75.00	73	1	0.013	0/0/1
41	58.33	2/1/1	25.00	52	9	0.173	1/8/0
total	63	15/2/19	52	729	21	0.03	14/2/5

[a]All nine constructs were obtained in the last round of transformation. The identifiers of the constructs are given in the first column (refer to Supporting Information Table 1). The percentage of positive clones from 12 colonies when screened by PCR is given in column two. The result from four positive clones sent to sequencing is given in the third column. Where indicated, failed corresponds to no sequencing result. The DNA synthesis length column specifies the length of sequence synthesized by oligonucleotides in the corresponding construct. The errors total column indicates the number of all errors found per construct, including multiple errors in one construct. The next column reports the error rate per synthesized base pair: the dsDNA parts were not counted towards this error rate. The last column reports the type of errors that were found. The bottom row reports the average for columns 2, 4, and 7 as well as the sum for all other columns.

customized ribosomal binding sites (RBS), affinity tags, antibody tags, secretion signals, localization signals, and promoters.

■ RESULTS AND DISCUSSION

Simultaneous Assembling of One dsDNA Fragment and the *De Novo* Synthesis of an Additional Sequence. Our initial task was to assemble a set of constructs in which a gene of interest was fused to a target sequence at the beginning of the gene. However, the size of the fusion (3−90 bp) made it difficult to introduce these sequences in a primer for PCR amplification. Therefore, we tested whether the simultaneous *de novo* synthesis of a sequence and assembly of the gene of interest is possible with one-step isothermal assembly.[3,4] In the state of the art, these methods can be used either to assemble multiple overlapping dsDNA molecules or, alternatively, to synthesize *de novo* DNA molecules from overlapping oligonucleotides. The combination of both in a one-step reaction has not yet been described.

We tested this approach by cloning 47 constructs combining six different promoters, four different N-terminal fusions, and five genes (Supporting Information Figure 1). Of the 120 theoretically possible constructs, we chose 47 for cloning. The library of 47 constructs was assembled in a miniTn4001-Puro-1 backbone (GenBank accession no. KC816623). The constructs consisted of a PCR insert (dsDNA between 500 and 2700 bp) and a *de novo* sequence (between 30 and 255 bp). These *de novo* sequences were built from combinations of two to eight oligonucleotides of around 60 bp with an overlap of 20 bp. For the complete experiment, 77 oligonucleotides were used (Supporting Information Table 2). The *de novo* sequence contained the promoter and a short sequence that should be fused to the PCR insert. In the first round of cloning, we obtained 29 out of 47 constructs with the correct sequence. After two more rounds of transformation and screening, we obtained 42 out of 47 constructs (Supporting Information Table 1). We reused the initial assembly reaction for the second and third rounds of transformation and colony screening. Two out of four constructs with seven oligonucleotides were not obtained with the correct sequence, even though both constructs with eight oligonucleotides were obtained with the

correct sequence. We observed background in the transformation, partly because of recircularization of the vector (32% of all colonies) as well as the generation of an unknown product during the assembly process (29% of all colonies). The unknown product is a consequence of the instability of the vector backbone (minitransposon vector) and was also detected during conventional cloning using this vector. Nonetheless, this strategy enabled a convenient one-pot assembly of constructs that would be difficult to prepare in any other way. We evaluated colony PCR hits and error rates in detail for the constructs obtained in the third round of the transformation (Table 1). To increase transformation efficiencies for these constructs, the assembly reaction mixtures were purified by MinElute columns before transformation. Twelve colonies for each construct were screened, with a hit rate of 63% in the colony PCR. However, the overall hit rate was only 40.5%, partially because of the problematic backbone. For each construct, we selected four colonies that showed a hit in the colony PCR, and we sequenced the construct. On average, we obtained a correct sequence for 50% of the clones. However, the individual rates for each construct varied from 25 to 100% of correct sequences. As a general trend, longer synthesized stretches contained more errors. This is not surprising because the synthesis of oligonucleotides is error-prone and the errors accumulate with an increasing number of oligonucleotides used for a construct. The errors observed were 14 insertions, five mutations, and two deletions; all were located in the part synthesized by oligonucleotides. In the screening of the complete library, only one truncation and one mutation were detected in the parts introduced as dsDNA PCR products.

***In Situ* Generation of Overlaps from Oligonucleotides.** After observing that sequence synthesis from oligonucleotides and DNA assembly can be done simultaneously using a one-step isothermal assembly master mix, we investigated whether the overlaps needed for assembly could be added simultaneously *in situ*.

To synthesize the overlaps necessary for the enzymatic assembly *in situ*, we added four different oligonucleotides to the enzymatic assembly mix. Each oligonucleotide overlaps 20 bp with the insert (the *bla* gene including the promoter, PCR amplified from pJET 1.2, GenBank accession no. EF694056.1,
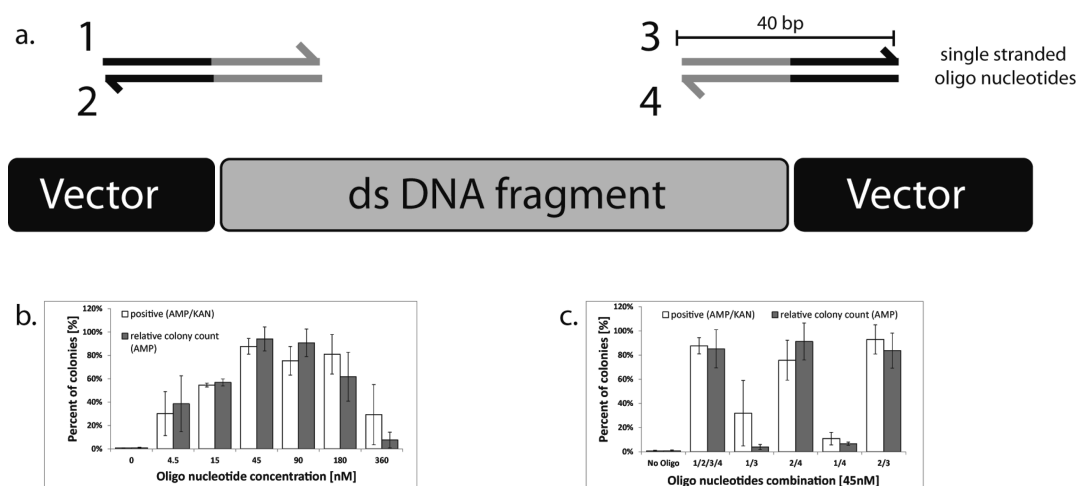
Figure 1. (a) Schematic representation of the oligonucleotides used for the *in situ* generation of overlaps. The oligonucleotides are shown with respect to the assembled dsDNA fragments. The bar indicates the size of the oligonucleotides (they are not to scale with the overlapping regions). (b) Effect of oligonucleotide concentration on the number of colonies obtained after transformation as well as the percentage of positive colonies obtained (c) Number of colonies obtained depending on different combinations of oligonucleotides as well as the percentage of positive colonies. The percentage of positive colonies was determined by dividing the colony count on a plate with ampicillin (AMP) by the colony count on a plate with kanamycin (KAN). The oligonucleotide numbers correspond to those in panel a.
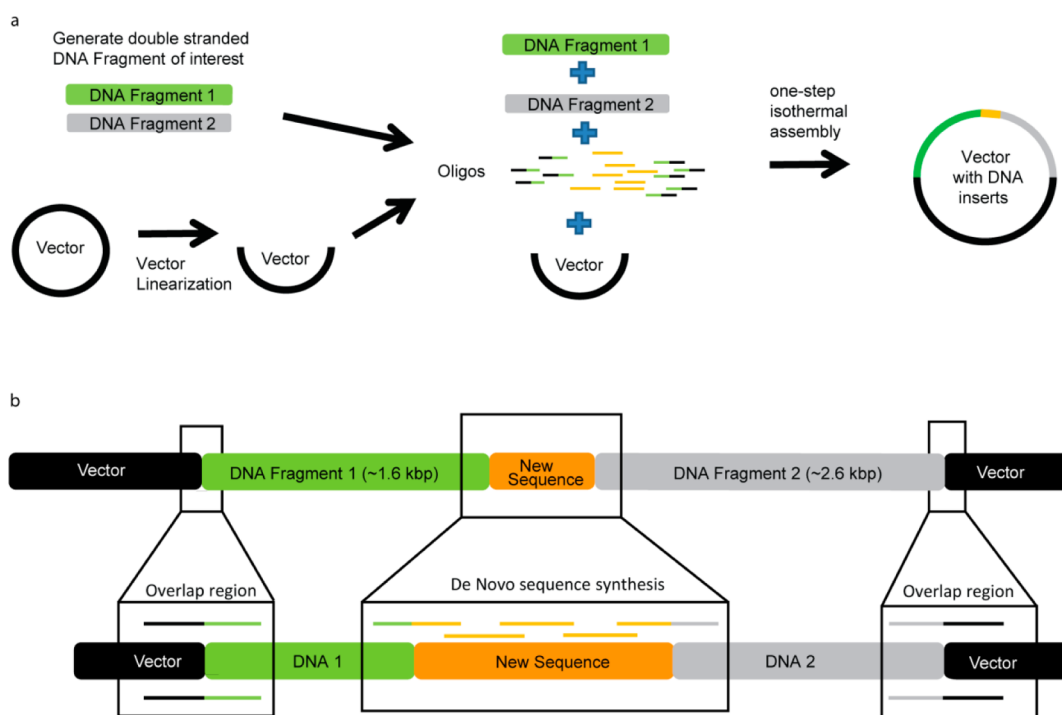
Figure 2. Schematic workflow and oligonucleotide design. (a) General workflow. dsDNA fragments and a linearized vector are obtained. Subsequently, the fragments, the vector, and the oligonucleotides are added to the one-step isothermal assembly master mix. (b) Distribution of the oligonucleotides on the construct for the assembly of importin-α (1.6 kb) and importin-β (2.6 kb) genes into the pCDF-Duet vector. A 156 bp spacer was synthesized in between them, which was composed of a T7 promoter and a modified RBS. The overlap between the individual oligonucleotides and the vector and DNA fragments is 20 bp.

was used as template) and 20 bp with the linearized vector pETM14[5] (Figure 1a) (GenBank accession no. KC816624). After transforming Top 10 *E. coli* competent cells with 1 μL of the assembly mix, 748 positive colonies were obtained on average.

We plated transformed cells on agar plates containing either ampicillin or kanamycin as the selective agent. This allowed us to distinguish between background colonies and colonies containing our insert on a global scale without the need to use colony PCR. We observed that two oligonucleotides can be sufficient for the assembly (Figure 1c), and it appears that oligonucleotide 2 is essential for assembly, whereas oligonucleotides 3 and 4 are exchangeable. We have no explanation for this, and we observed different dependencies on other occasions (data not shown). We hypothesize that properties of the individual oligonucleotide sequence, synthesis, or preparations are the determining factor. Therefore, we recommend using as a standard procedure all four stitching

oligonucleotides because this provides the most robust condition for the successful assembly of the construct. The optimal concentration of oligonucleotides in the final master mix was around 45 nM (Figure 1b). Very high concentrations of oligonucleotides (>150 nM) inhibited the DNA assembly step as well as reduced the efficiency of the transformation. Overall, we demonstrated that it is possible to *in situ* synthesize overlaps in a one-step isothermal assembly reaction, and we found robust parameters for this assembly. This facilitates the high-throughput cloning of the same fragment into many nonstandard vectors by exchanging only the stitching oligonucleotides.

We observed that isothermal assembly mixtures containing only the linearized vector yielded significant amounts of colonies, whereas the linearized vector without isothermal assembly or in the absence of Taq ligase yielded no colonies. We obtained the same results with a commercial version of the isothermal assembly mixture obtained from NEB in combination with a commercial linearized vector containing a death gene (pJET 1.2). The recircularized assembly products are truncated versions of the original vector and not a simple religation of the linearized vector. This is known to many in the field. It poses no problem in standard cloning applications because the rate of positive clones in a normal assembly is close to 90%. The background is observed only when the assembly fails or in the negative control of the vector. This is a drawback when assembly conditions have to be optimized against a negative control. Therefore, we recommend optimizing the assembly conditions using an antibiotic resistance gene as the insert or planning a colony-screening scheme with sufficient throughput.

Assembly of Two Inserts by *in Situ* Generation of Overlaps and *De Novo* Assembly of a Promoter and RBS (156 bp). After demonstrating that the *in situ* synthesis of overlaps from oligonucleotides is possible and that it can be combined with the assembly of dsDNA fragments, we tested whether it is possible to combine the two methods and clone multiple fragments (Figure 2). For this, we chose to clone the importin-α (1.6 kb) and importin-β (2.6 kb) genes in the pCDF-Duet vector while building a 156 bp spacer in between them composed of a T7 promoter and modified RBS (GenBank accession no. KC816625). To achieve this, three different oligonucleotide concentrations were tested in the final master mix: 550, 55, and 5.5 nM. As negative controls, we used a one-step isothermal assembly master mix without Taq ligase and one complete master mix missing the oligonucleotides (Supporting Information Table 4). We screened 12 colonies of the assembly mix with 55 nM oligonucleotides for the presence of both inserts, and we found eight colonies with a PCR signal for importin-α and four colonies for importin-β. Three of the colonies were positive for both inserts. The plasmid DNAs of the double-positive clones were purified, and no errors were detected after sequencing.

We obtained 26 colonies in the assembly of the negative control without oligonucleotides and 10 in the negative control without ligase. Sequencing results showed that the background is the result of a misprimed product of the PCR linearization of the backbone. This product produced two homologous ends that were able to combine during the reaction. Such byproducts can be easily eliminated by PCR screening and do not affect the overall convenience of this method.

We have shown that the synthesis of sequences from single-stranded oligonucleotides can be combined with the assembly

of DNA fragments into a vector. We also demonstrated that the overlaps needed for DNA assembly can be synthesized *in situ* during the assembly. By combining these two approaches, we could assemble, in a one-step reaction, two genes flanking a *de novo* built-up sequence.

This technique is also advantageous for individual constructs when *de novo* sequences have to be constructed that are too large to be introduced in a primer. For example, with our technique, a secretion signal of 30 amino acids or more can be easily synthesized in a one-step cloning reaction. By exchanging only individual oligonucleotides, mutations can be introduced in the amino acid sequence to investigate their effect on the phenotype.

The *in situ* synthesis of overlaps has multiple applications. First, it can be used to subclone dsDNA fragments with the versatility of one-step isothermal assembly cloning while omitting the need for any PCR reaction. Another application is in the generation of libraries for expression screening in which two genes need to be cloned in many vectors while varying the order of the genes, the promoters, and optimized RBS sites. The order of the genes can be easily changed in the *in situ* assembly of the overlaps by exchanging only the overlapping oligonucleotides. In the same way, a set of RBS sites can be screened or new promoters can be tested. Compared to a standard cloning, the advantages of this technique become greater with increasing sample numbers because oligonucleotides can be reused between constructs. The construction of combinatorial libraries using the isothermal assembly mixture have been described earlier.[6] However, this and other approaches depend on common linker sequences between the assembled parts, which are unnecessary in our approach. Unfortunately, we were not able to test our approach for the simultaneous assembly of a library in a single reaction mixture similar to that described by Ramon and Smith.[6]

We evaluated the error rates and location of errors for a small subset of the cloned library. The error rate per base pair for the synthesized stretches is about 1 order of magnitude larger than the error rate reported for the underlying technique by Gibson et al.[4] Although we cannot directly explain this, it is worth noting that the oligonucleotides used came from a different vendor then in the original study (Sigma in our case and IDT in the case of Gibson et al.[4]).

We have generated a library of 42 fusion genes in single-step reactions using a new approach on the basis of the one-step isothermal assembly cloning technique. This avoids a large number of PCR amplifications and multiple intermediate cloning steps. We show that sequences between 30 and 255 bp can be synthesized during the assembly of a vector with a PCR fragment of the size range between 500 and 2700 bp. The design and assembly of the constructs was very easy compared to other state-of-the-art methods and is compatible with automation. Therefore, we anticipate that this simple approach, when combined with rigorous screening methods, can be easily implemented in many applications and could be widely used in both molecular and synthetic biology.

■ METHODS

DNA Fragment Preparation. The vectors were linearized by inverted PCR or with a restriction enzyme (EcoRV-HF, New England Biolabs). PCR fragments were amplified using either Phusion (Finnzymes) or Kod (Merck Millipore) high-fidelity polymerases. A list of all primers used to generate PCR fragments and the amplified sequences can be found in the

Supporting Information. Vector and inserts were purified using the Qiagen MinElute Kit or Gel Purification Kit following the instructions of the manufacturer.

All oligonucleotide sequences and their combinations as well as the vectors used in this study are summarized in Supporting Information Tables 1, 2, and 3. The oligonucleotides were ordered from Sigma as desalted (desalt) or reverse-phase purified (RP). The pJET 1.2 vector was obtained as part of the CloneJET PCR Cloning Kit (Thermo Scientific no. K1231)

Oligonucleotide Design for De Novo Synthesis. All oligonucleotides were designed to be around 60 bp, with a 20 bp overlap to the next oligonucleotide, a dsDNA fragment, or the vector. Properties like melting temperature or secondary structure were ignored for the design process, as described earlier.[4] If the length of the synthesized stretch made it necessary, then individual oligonucleotides were designed shorter or longer than 60bp, and the overlapping region was extended. A list of all oligonucleotides can be found in Supporting Information Table 2.

Assemblies. The isothermal assembly mixture was either prepared in house as previously described[3] or purchased from New England Biolabs (no. E2611S). The assemblies were performed using a modified version of the protocols previously described or using the manual for the Gibson Assembly Master Mix available at neb.com.[3,4] Briefly, oligonucleotides were mixed in water in equimolar ratios from a 100 μM stock. Then, the mixture was heated at 95 °C for 5 min and slowly cooled to 4 °C with a ramp of 0.1 °C/s. The oligonucleotide mixture was diluted to a concentration of 900 nM, and 1 μL was added to a master mix for isothermal assembly, resulting in a final concentration of 45 nM of each oligonucleotide.

The PCR fragments were added either in equimolar amounts or as a 2-fold excess of the vector (50−100 ng). Unless otherwise stated, stitching oligonucleotides were also added to a concentration of 45 nM, in a final volume of 20 μL, as previously described.[3,4] The stitching oligonucleotides were not annealed before assembly. In contrast to the original protocol, the isothermal assembling was performed at 45 °C for 2 h instead of 50 °C and 1 h. This modification is based on previous observations reported that an incubation at 45 °C for 1 h increases the efficiency of the assembly compared to 50 °C and that an incubation at for 2 h at 50 °C also yields more clones than that at 1 h and 50 °C.[4] We confirmed the assumption that 2 h at 45 °C are superior to 1 h or 2 h at 50 °C by a pilot experiment (data not shown). The combinations of oligonucleotides and PCR fragments for each construct are listed in Supporting Information Table 1. The assemblies for the *in situ* generation of overlaps from oligonucleotides were performed in triplicate.

Bacterial Transformation. Top 10 (Life Technologies) or DH5α *E. coli* cells made competent with Z-Competent Kit (Zymo Research) were used for the transformations following the instructions of the manufacturer. The first transformation round for assembling one double-stranded insert and the de novo synthesis of a sequence between 30−255 bp was done with DH5 cells in combination with the Z-Competent Kit. The manufacturer gives a transformation efficiency for these cells of 10^8−10^9 transformants/μg of plasmid DNA. The second round was performed with Top 10 cells that have a transformation efficiency of 10^9 transformants/μg of plasmid DNA. In the third round, Top 10 cells were also used, but the assembly reaction mixture was purified by a MinElute column (Qiagen) before transformation, and 5 μl of the elution was used for

transformation. The removal of PEG8000 and various proteins increases the yields of colonies significantly.

The transformations for all other described experiments were done by adding 1 μL of the assembly master mix to Top 10 cells following the instruction of the manufacturer. The transformation for the *in situ* generation of overhangs experiments were split in half after recovery in SOC medium (half plated on ampicillin- and the other half on kanamycin-containing plates). Otherwise, selection on LB plates was performed with the antibiotic required by the vector.

Screening for Full-Length Inserts Assembled from Oligonucleotides and Nonoverlapping dsDNA Fragments. To screen for the successful assembly of our constructs, we performed colony PCR. In the case of the fusion library, we picked eight colonies from each transformation and screened using the primers Fwd Screen Puro and Rev Screen Puro. For each positive construct, four colonies were picked and prepared using the Millipore Montage Plasmid Miniprep HTS 96 Kit. If fewer than four colonies were positive in colony PCR, new unscreened colonies were selected, and the plasmid was purified. All purified plasmids were sequenced by Sanger sequencing (GATC, Germany).

The assembly in the pCDF_Duet Vector was screened by PCR with the primers used to generate the inserts. Colonies giving a positive signal for the inserts were purified using a Qiagen Miniprep Kit and sent for sequencing. All primer sequences used for screening are listed in Supporting Information Table 1.

***In Silico* Construct Design.** The constructs were designed using the Gene Designer software and the CLC Main workbench.[7]

■ ASSOCIATED CONTENT

ⓢ Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*Tel.: +34933160259; Fax: 34933160099; E-mail: bernhard. paetzold@crg.eu (B.P.).

*E-mail: maria.lluch@crg.eu (M.L.-S.).

Author Contributions

B.P. wrote the manuscript and designed the experiments. B.P, C.C., and T.F. contributed experimental data. M.L.S. and L.S. provided supervision and feedback; the work was performed in the lab of L.S.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS

SLIC, sequence- and ligation-independent cloning; RBS, ribosomal binding site; PCR, polymerase chain reaction; *E. coli*, *Escherichia coli*; dsDNA, double-stranded DNA; bp, base pair

■ REFERENCES

(1) Cohen, S. N., Chang, A. C. Y., Boyer, H. W., and Helling, R. B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U.S.A. 70*, 3240−3244.

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

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

(4) Gibson, D. G., Smith, H. O., Hutchison, C. A., Venter, J. C., and Merryman, C. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods 7*, 901−903.

(5) Dümmler, A., Lawrence, A.-M., and de Marco, A. (2005) Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. *Microb. Cell Fact. 4*, 34-1−34-10.

(6) 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.

(7) Villalobos, A., Ness, J., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006) Gene Designer: A synthetic biology tool for constructing artificial DNA segments. *BMC Bioinf. 7*, 285-1−285-8.