Highly Efficient Single-Pot Scarless Golden Gate Assembly

Mohammad HamediRad,†‡∥, Scott Weisberg,‡∥, Ran Chao,†‡⊥ and Huimin Zhao*†‡∥§∥∥

†Carl R. Woese Institute for Genomic Biology, ‡Department of Chemical and Biomolecular Engineering, §Department of Biochemistry, ∥College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China
∥Departments of Chemistry and Bioengineering, University of Illinois at Urbana—Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Golden Gate assembly is one of the most widely used DNA assembly methods due to its robustness and modularity. However, despite its popularity, the need for BsaI-free parts, the introduction of scars between junctions, as well as the lack of a comprehensive study on the linkers hinders its more widespread use. Here, we first developed a novel sequencing scheme to test the efficiency and specificity of 96 linkers of 4-bp length and experimentally verified these linkers and their effects on Golden Gate assembly efficiency and specificity. We then used this sequencing data to generate 200 distinct linker sets that can be used by the community to perform efficient Golden Gate assemblies of different sizes and complexity. We also present a single-pot scarless Golden Gate assembly and BsaI removal scheme and its accompanying assembly design software to perform point mutations and Golden Gate assembly. This assembly scheme enables scarless assembly without compromising efficiency by choosing optimized linkers near assembly junctions.

KEYWORDS: DNA assembly, Golden Gate assembly, synthetic biology

DNA assembly is one of the most fundamental techniques and often one of the very first steps in synthetic biology and metabolic engineering. The variety of techniques ranging from recombinant DNA technology in the 1970s to modern DNA assembly technologies have enabled scientists to make genetic modifications in all kingdoms of life for a wide variety of applications. A myriad of different ways to assemble DNA have been developed and presented to the scientific community. These methods assemble multiple DNA molecules together in a specific order using different approaches like restriction digestion-ligation, homology-based, and bridging oligo-based assembly.

One of the most widely used DNA assembly techniques is the Golden Gate method and its variations, thanks to their robustness and modular nature. In this method, type IIIS endonucleases are used, which cut outside of their recognition sequence and can generate any possible 4-bp sticky ends. These sticky ends can then anneal and ligate to their corresponding sticky ends, resulting in assembled DNA in the desired order. The flexibility of type IIIS restriction enzymes, most commonly BsaI, provides a high degree of modularity and robustness to Golden Gate assembly, making it one of the most widely used DNA cloning methods.

Despite these advantages, Golden Gate assembly has some limitations. All the parts used in the assembly should be free of the type IIIS restriction enzyme recognition site to avoid digesting the parts in the middle and considerably reducing the assembly efficiency. Given the relatively short recognition length of most type IIIS restriction enzymes, most commonly BsaI, the recognition site is relatively common in biological parts and must be removed by silent mutation. Moreover, Golden Gate assembly introduces a 4-bp scar between the DNA parts that can affect functionality, especially in areas upstream of the start codon where addition of multiple nucleotides can considerably change the expression level. The 4-bp overhangs or sticky ends generated after the digestion step that are used to assemble the parts in the correct order (linkers) are one of the key factors in designing the Golden Gate reaction. The choice of 4-bp linkers used in Golden Gate assembly has been shown to be crucial in the success and efficiency of the assembly.

There have been a few studies on the optimization of linkers based on computational and experimental data. Some general guidelines for choosing efficient linkers have been suggested and a piece of software has been developed to generate a number of efficient linkers following these guidelines. All the possible 4-bp linkers were also evaluated elsewhere, and a few efficient linker sets were identified and presented. However, none of these studies provided an extensive list of experimentally verified highly efficient 4-bp linkers. Due to the different objectives of...
these studies, no solution for scarless Golden Gate assembly and BsaI removal and no tools for designing assemblies and adaptive selection of linkers for the assemblies were proposed by these studies.

To address these limitations, we first developed a next-generation sequencing (NGS)-based method to evaluate the annealing frequency between all possible pairs of 96 linkers and identified the affinity of the linkers to anneal to themselves and their ligation specificity, i.e., the affinity of the linkers to anneal to others. Using this data, we identified more than 200 sets of optimized linkers to be used in Golden Gate assembly. We then designed a Golden Gate assembly scheme, where these optimized linkers can be used to perform scarless Golden Gate assembly, BsaI removal, or any point mutation in the assembly parts. In addition, we developed a web application, iBioCAD GGA, to facilitate a high-throughput design of efficient and scarless Golden Gate assembly and BsaI removal using optimized linkers and the assembly scheme presented here. Using this design tool and optimized linkers, as a proof of concept, we performed a 3-piece scarless assembly and BsaI removal with a total size of 3 kb of the ampicillin gene with 100% efficiency.

**RESULTS AND DISCUSSION**

**NGS-Based Linker Optimization.** Sequences of the 4-bp linkers have been shown to play an important role in the success of Golden Gate assembly. AT-rich linkers have been shown to have low affinity to each other and cause lower assembly efficiency. Linkers with similar sequences can misanneal to each other and drastically reduce the specificity of the assembly. Given the inherently high efficiency of Golden Gate assembly, these problems usually occur when there are more than 2 pieces of DNA fragments being assembled. With this constraint, a two-piece assembly to assess the specificity and affinity of Golden Gate reactions is not very effective since most of the products would likely turn out to be correct. Additionally, a multipiece assembly may not be very effective either, because it would be difficult to point the inefficient assembly to one linker versus the other where more than one or two are involved.

To address this problem, a method was designed where a one-piece assembly reaction was performed by adding the BsaI recognition site and appropriate linkers to the end of a linearized plasmid produced by PCR amplification and then performing Golden Gate assembly on the PCR product to produce a circular plasmid (Figure 1A). By generating PCR products containing all the linker combinations on each side, this method significantly increased the likelihood of a misligation being observed and low specificity linkers being detected. This intramolecular ligation design enables us to increase the chances of observing misalignments and effectively allowing us to compare linkers with a wide range of ligation efficiency with relatively low sequencing depth. To generate this intramolecular ligation library, we first created a forward and reverse library of primers, each primer containing the BsaI recognition site, linker, and an 8
amplify the linearized backbone as shown in Figure 1A. S1) were equimolarly mixed and used in a PCR reaction to all the primers are almost identical. All the 192 primers (Table PCR amplification, the 8 bp barcodes were chosen from the first 8 bp of a set of 240 000 orthogonal barcodes25 so that the ΔG are more likely to have higher affinity and perform better, all the possible 120 linkers were ranked by their ΔG and the 96 linkers with the lowest ΔG, calculated using the nearest neighbor method,24 were chosen. To avoid biases in PCR amplification, the 8 bp barcodes were chosen from the first 8 bp of a set of 240 000 orthogonal barcodes25 so that the Tm for all the primers are almost identical. All the 192 primers (Table S1) were equimolarly mixed and used in a PCR reaction to amplify the linearized backbone as shown in Figure 1A.

Golden Gate assembly was then performed on the PCR product, and the expected assembly product is shown in Figure 1B. The PCR products were expected to be digested and the linkers to be exposed. The linkers on both sides of the product are expected to be different for most of the population and only a small portion of the population is expected to have the same linker on both sides. After the ligation step, the linkers with higher affinity are expected to ligate to each other at a higher frequency than others. After 30 rounds of digestion and ligation, the Golden Gate product was purified and amplified in E. coli, and the region containing barcodes and the assembled linkers was then PCR amplified and sequenced. By analyzing the NGS data obtained from the sequencing results and comparing the barcodes on both sides of the linker, the number of matches (one linker ligated to the same linker) and misligation (one linker ligated to another linker) were identified.

We hypothesized that the number of matches has a positive correlation with the affinity of that linker because all the linkers have the same likelihood to be found in the PCR product and the higher frequency in the Golden Gate product can be attributed to the higher affinity of those linkers to themselves. Similarly, we hypothesized that the number of misligation between each linker to another shows their affinity to each other. Consequently, if the frequency of misligation between two linkers is high, they should not be used together in a Golden Gate reaction as it will reduce the specificity of the assembly. Moreover, if the number of matches between one linker to itself is high, the linker is desirable, and it will increase the efficiency of the Golden Gate reaction. The heatmap for the number of matches and misligation for all the linkers is shown in the matrix in Figure 2A. As expected, the diagonal cells on the matrix have a higher number of matches than the nondiagonal cells because it is easier for a linker to anneal to its perfect complement than to any other. The linkers in this matrix were sorted based on binding ΔG to themselves. Since the number of matches on average were considerably higher, the same data in a log-scale was also shown in Figure 2B for comparison. It is also noteworthy that by design and because of this intramolecular ligation reaction, this matrix is biased toward having more frequent misligations than correct ligations. This enables us to assess the efficiency of ligations and observe misligations without having to considerably increase the sequencing depth.

Analyzing the Trends in NGS Sequencing Data. First, we tried to validate the general guidelines that were suggested and used in many other previous studies,2,22 namely, GC content 25% to 75% and two allowable homologies. It was observed that the trend generally follows these rules but there are notable exceptions as shown in Figures 3A and 3B. Therefore, as expected, these are good rules of thumb but there are exceptions to these rules that can result in lower than optimal efficiency. For example, we found that there is a general trend in the data where the higher the GC content, the higher the number of matching ligations. This was expected given that GC bonds are stronger than AT bonds, but there are many exceptions to this general rule too (Figure 3A). We also observed the general trend that the less similar the 4-bp linkers are, the less likely they are to ligate to each other (Figure 3B). This observation is rather obvious and very much expected but demonstrates that the data generally matches the expectations. We then analyzed the number of misligations between 4-bp linkers with two noncomplementary base pairs, and categorized them to two categories of consecutive and nonconsecutive misligations and found that consecutive misligations are more likely to result in misligation than nonconsecutive ones (Figure 3C). We also analyzed the

Figure 2. Heatmap showing the number of matches between each overhang pair in (A) linear and (B) logarithmic scales.
correlation between the position of misligation, ends of the 4-bp linker or the middle of the linker, and misligation types (A−C, A−G, etc.) and found that the position of misligation had no impact on the number of misligations (Figure 3D). However, the misligation type did have an impact. (Figure 3E). Interestingly, we observed that transitions, purine−purine or pyrimidine−pyrimidine substitutions, have a significantly lower frequency of misligation than do transversions, purine−pyrimidine or pyrimidine−purine substitutions (p < 0.05).

**Experimentally Assessing the Predictive Power of the Matrix.** After acquiring this data on the affinity and specificity of the linkers, we set to evaluate the predictive power of this matrix.

**Figure 3.** Analysis of patterns observed in the affinity matrix. The interior of each violin plot includes a box plot showing the interquartile range (IQR), a white dot showing the median value, and whiskers reaching from the top and bottom of the box to 1.5 times the IQR. The exterior of each violin plot shows a kernel density estimation of the underlying distribution of data. The density of each violin is scaled to the same width. Groups within panels A−C are significantly different as calculated by single-factor ANOVA, p < 0.05. In panels A−E, N refers to the number of unique linker pairs observed in each subset. (A) GC content of linkers has mild effect on their affinity of self-binding. N = 15, 43, 36 for 25%, 50%, 75%, and 100% GC content, respectively. No linkers with 0% GC are shown as the 96 tested linkers were chosen from those with the lowest self-binding ΔG. (B) Similarity of linkers and fidelity of ligation are highly correlated but have many exceptions. N = 96, 720, 2438, 3744, 2218 for 0, 1, 2, 3, and 4 base difference, respectively. (C) Fidelity of linker pairs with a 2-base difference, plotted as “Nonconsecutive” (e.g., GGGG and GGGC) (n = 764) or “Consecutive” (e.g., GGGG and GGCC) (n = 1674). (D) Fidelity of linker pairs with a 1-base difference, plotted as “End” (e.g., GGGG and CGGG) (n = 294) or “Middle” (e.g., GGGG and GCGG) (n = 426). There is no significant difference between these groups (p > 0.05, calculated by single-factor ANOVA). (E) Fidelity of linker pairs with a 1-base difference, plotted as the identity of each mismatch (e.g., AGGG and CGGG as an “AC” mismatch). N = 33, 33, 33, 90, 70, 101 for “AC”, “AG”, “AT”, “CG”, “CT”, “GT” mismatches, respectively. Transition mismatches (purine−purine or pyrimidine−pyrimidine) have a significantly lower frequency of misligation than transversion mismatches (purine−pyrimidine or vice versa) with p < 0.05, calculated by single-factor ANOVA.
We first looked at the linkers that have been used in previously published Golden Gate assemblies, most notably MoClo\textsuperscript{21} and FairyTALE.\textsuperscript{21} We first observed that some of the linkers used in MoClo assembly are very AT-rich and were not among the 96 linkers with the lowest \(\Delta G\). This may have caused low efficiency in the MoClo assembly. We also observed that according to our data, some of the linker pairs used in these publications tend to have many misligations, which suggests the possibility of low assembly specificity. All the linker pairs with high observed misligation were different by only one base from the other, further validating the predictive power of our data as shown in Table 1. It is noteworthy that these sets have high specificity for the most part and these nonspecific pairs were picked from hundreds of pairs of linkers. However, they could still be improved by some modifications as suggested by the data.

To evaluate if the linkers have the expected performance, a multipiece Golden Gate reaction was performed. We chose a \(~10\) kb, 6-part zeaxanthin pathway assembly because of the number of parts, size of the assembly, and relative ease of evaluating efficiency and specificity with colorimetric assays. We generated 4 different sets of linkers with high and low specificity and affinity. These linker sets and the number of matches and misalignments in them are shown in Figures S3–S6. Everything else being equal, we expected to see more colonies on the plate with high-affinity linkers and fewer colonies for low-affinity linkers. We also expect to see a lower percentage of yellow colonies (with a complete zeaxanthin pathway) on the low-specificity plates because the occurrence of misassembly is more likely.

As expected, the number of colonies were significantly higher on the high-affinity plates (Table 2 and Figure S1). The percentage of correct colonies were also very high for sets with high specificity, and low to modest for sets with low specificity. Interestingly, one of the colonies in a low-specificity set turned out to be red, which likely indicates partial assembly of the zeaxanthin pathway only up until the lycopene pathway, showing the distinct red color of lycopene (Figure S2). Six of these yellow colonies were randomly picked and the junction between the parts were sequenced (36 reactions) and all of them showed correct sequences.

**Efficient Single-Pot Scarless Golden Gate Assembly Scheme.** Given the nature of type IIS restriction enzymes, Golden Gate reactions are easily and scarlessly performed due to the flexibility of choosing the Golden Gate linkers. However, it has been reported before and shown here that the linkers play an important role in the efficiency and specificity of the assembly.\textsuperscript{21} For example, sometimes the linkers adjacent to one junction may be very similar to another or they can be AT-rich which can adversely affect the assembly. However, given the data presented here on the specificity and affinity of the pairs of linkers, we have a lot of latitude in choosing the efficient linkers for each assembly.

To assemble two pieces of DNA, one efficient linker must be found near the junction for assembly as shown in Figure 4A. Then, primers will be designed such that the efficient linker and BsaI recognition site is added to the end of each PCR product. The search perimeter should ideally be the region from 30 bp downstream to 30 bp upstream of the junction such that the length of the designed primer does not exceed 60 bp suggested by most DNA synthesis companies. Unless the ideal linker is right at the junction of two parts, a small portion of one of the parts will be added to the other by PCR amplification. PCR products can then be digested and ligated in a Golden Gate reaction and assembled scarlessly and efficiently.

This strategy can be extended to assemblies with more pieces as well as performing mutations on the parts of interest before assembly. In the case of assemblies with more than one junction, as shown in Figure 4B, more than one linker must be found and other than the affinity, the specificity of the linkers and their compatibility with each other must be checked. Point mutations can also be performed by just treating the target point of mutation as a junction, finding the appropriate linkers around that point, and designing primers with the intended mutation. This point mutation is especially useful in Golden Gate assembly since all the parts used must be free of the type IIS restriction site and given the 6-bp length of the BsaI recognition sequence as an example, it occurs with relatively high frequency (~39% chance in a 2 kb gene, Supplementary Discussion A) in natural DNA sequences.

**Finding the Optimum Set of Linkers for DNA Assembly.** As the number of parts and consequently the number of junctions grow, the task of finding the right linker set gets increasingly difficult. Given the interdependency of the linkers and inherent sequence dependency of each assembly, one set of efficient linkers is not enough for all assemblies. The linkers in the set used in the DNA assembly should (1) exist on the 30 bp vicinity of the junction, (2) have high affinity with itself, and (3) have low affinity with all others. Meeting all three of these criteria is difficult, and manually checking all the possibilities is almost impossible. For example, we need to have a set of 14 linkers to expect to find only one subset of 5 efficient linkers to perform a 5-piece assembly. This number exponentially increases with the number of pieces in an assembly (Supplementary Discussion B).

To solve this problem, we first designed 200 sets of optimized linkers, 5 sets of each size from 10 linkers to 50, with high affinity and specificity that can be used for efficient Golden Gate assembly.
reactions (Table S7). Since fewer linkers were needed for the smaller sets, criteria for choosing them were more rigorous and they are expected to perform better than the larger sets. However, the larger sets are equally important for the cases where a larger number of parts are to be assembled or a larger set is required to find the appropriate linkers for the assembly. The sets were designed to be as different from each other as possible without compromising the efficiency as to increase the likelihood of finding a set of linkers for different assemblies.

We then developed a web-based application, iBioCAD GGA (http://ibiocad.igb.illinois.edu/) enabling the design of scarless and regular Golden Gate assemblies using the optimized linkers. For the case of scarless Golden Gate assembly, the DNA parts to be assembled and the backbone can be given to iBioCAD GGA and it searches the area around the junctions of the parts and finds the appropriate linkers to use for the assembly. To find the best linker set possible for an N-piece assembly, iBioCAD GGA first searches within the sets with N linkers to find the appropriate linkers. It will only try the larger sets if it cannot find the appropriate linkers. After the appropriate linkers are found and the assembly is designed, iBioCAD GGA designs the primers and assembly scheme for efficient, scarless single-pot assembly with instructions for performing the Golden Gate reaction.

For the cases where scarless assembly is not crucial, iBioCAD GGA can design regular Golden Gate assemblies with optimized linkers. The flexibility of adding a scar to the assembly enables the software to choose more efficient linkers and potentially increase the efficiency and specificity of the reaction. iBioCAD GGA can also perform high-throughput combinatorial assembly by letting the user upload multiple DNA parts for each location on the plasmid and designs multiple plasmids with only a few clicks.

**Scarless Golden Gate Assembly.** To test the software and this assembly scheme and as a quick proof of concept study, we used the ampicillin resistance gene and cloned it in a pET26b vector following the design in iBioCAD GGA. Both parts were given to iBioCAD GGA and it designed 6 primers to generate the parts needed for assembly. The length and $T_m$ of these primers were optimized to ensure the success of PCR amplification where possible. The ampicillin gene was divided into two parts, the BsaI recognition site was removed and the scarless Golden Gate assembly and BsaI removal were performed in one reaction. Twenty clones were tested, and all of them showed growth in LB media supplemented with ampicillin suggesting a 100% efficiency.

**CONCLUSION**

DNA assembly is one of the cornerstones of synthetic biology and is one of the first steps in most engineering efforts in this area. Golden Gate assembly is one of the most efficient and versatile DNA assembly methods to date and is used in a variety of applications from TALEN synthesis to biosynthetic pathway construction. The importance of linkers in efficiency and specificity of assembly has been demonstrated before, and some rough guidelines have been introduced for designing linkers used in Golden Gate assembly. A piece of software has also been developed to generate compatible linkers following these guidelines. Separately, an extended study was performed to analyze the ligation efficiency of 4-bp linkers with T4 and T7 DNA ligase, and a few efficient sets have been introduced.

In this work, a method in which next generation sequencing was used to evaluate the linkers in Golden Gate assembly to find their affinity and specificity was designed and implemented. This method resulted in design of 200 sets of optimized linkers that have been directly generated from the experimental data, demonstrating a significant improvement over a few sets designed elsewhere as well as the sets designed computationally using some general guidelines. The data presented here can be used to find the compatibility of any 4-bp linkers for assembly, generated by type IIS, type II, or even Artificial Restriction Enzymes (AREs) where any loci can be targeted and custom linkers can be generated. The library creation and sequencing method can also be expanded and used for characterizing longer linkers, although the cost and complexity of primer design for longer linkers increases exponentially due to the increase in the total number of possible linkers.

The predictive power of this method was then tested by crosschecking the linkers used in other publications to estimate their specificity. It was found that linkers predicted to be inefficient by the data presented here tend to be very similar to each other and in most cases are different by only one base. Four sets of linkers with high and low specificity were designed, and it was found that the efficiency and specificity matched the predictions.

Finally, two single-pot scarless Golden Gate assembly reactions were designed, and a web application, iBioCAD GGA, was developed to automate this assembly process. This assembly scheme eliminates some of the major problems with Golden Gate assembly, namely, the trade-off between scarless and efficient assemblies as well as the need to remove BsaI recognition sites before starting the assembly. We were able to efficiently perform scarless assembly and remove BsaI recognition sites from the sequences in a one-step and single-pot assembly reaction. The approach used to remove BsaI
recognition sites can be used to make any desired point mutation in the parts of interest for various applications from protein engineering to promoter tuning and mutagenesis.

**MATERIALS AND METHODS**

**Strains, Media, and Cultivation Conditions.** *E. coli* DH5α (New England Biolabs, Ipswich, MA) cells were used for making chemically competent cells using Mix & Go *E. coli* Transformation Kit (Zymo Research, Irvine, CA) for plasmid amplification. *E. coli* cells were grown in Luria Broth (LB) medium (Fisher Scientific, Pittsburgh, PA) supplemented with 25 μg/mL kanamycin (Kan) or 50 μg/mL of spectinomycin (Spec) or 50 μg/mL of ampicillin to maintain the plasmid. Antibiotics were purchased from Gold Biotechnology (St. Louis, MO). *E. coli* DH5α cells, starter cell cultures, were grown at 37 °C.

**DNA Manipulation and Library Construction.** All 192 primers were designed using Python programming language (Table S1) and ordered through Integrated DNA Technologies (Coralville, IA). The primers were then normalized to 100 pmol/μL and 10 pmol/μL using EPMotion robotic system from Eppendorf (Hamburg, Germany). 50 pmol of each primer was aspirated and mixed to create two forward and reverse mixed libraries with the final concentration of 10 pmol/μL. The pSPE-UPB1-EcoRV-UPB2 backbone was constructed by digesting pSPE with *AfII* and *XbaI*, and two primers P01 and P02 were annealed and ligated to the digestion product.

The primer library was then used to amplify the pSPE-UPB1-EcoRV-UPB2 backbone, and the PCR product was gel purified. 500 ng of the purified library was used in a 20 μL Golden Gate reaction consisting of 16 μL of the DNA, 1 μL *BsaI* restriction enzyme, 0.25 μL T4 DNA ligase, 2 μL of CutSmart buffer, 0.75 μL of ATP (25 mM), and 4 μL of molecular biology grade water. After the Golden Gate reaction, 5 μL of plasmid safe nuclease (Illumina, San Diego, CA) and *BsaI* master mix (0.25 μL *BsaI* (10 units/μL), 0.25 μL plasmid safe nuclease, 0.5 μL CutSmart buffer, 1 μL ATP (25 mM) and 3 μL water) was added to the reaction to linearize undigested backbone and remove all the linear parts from the mixture. The Golden Gate and plasmid safe master mix protocol were adopted from our previous work with some modifications, but the thermocycling protocol remained unchanged.

The inserts for the zeaxanthin pathway assembly in *E. coli* with high and low specificity and affinity linkers were obtained by amplifying the *CrtEBIYZ* genes with primers P05 to P50 where appropriate and blunt ligating them with PCR Cloning kit (New England BioLabs, Ipswich, MA). The pET26b backbone for these was also prepared by annealing primers P03, P04, P15,16, P27,28, and P39,40 and ligating them to pET26b plasmid digested with *XhoI* and *SphI* restriction enzymes.

QIAGEN Plasmid Mini Kit (QIAGEN, Valencia, CA) was used to isolate plasmids from *E. coli* cells, and Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA) was used for gel purification. All restriction enzymes, Q5 polymerase, and the *E. coli* shuttle vectors were purchased from New England Biolabs (Ipswich, MA), and all chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise specified.

**Next Generation Sequencing Library Construction and Analysis.** To attach NGS adaptors, a first step PCR was performed using 2 × Kapa HiFi HotStart Ready Mix (Kapa Biosystems, Wilmington, MA) and 10 ng extracted plasmid as template. The cycling condition is 95 °C for 3 min, (95 °C for 30 s, 46 °C for 30 s, 72 °C for 30 s) × 8 cycles, 72 °C for 5 min, and held at 4 °C. The PCR product was gel purified and 10 ng PCR product from the first step was used in a second step PCR to attach Nextera indexes using the Nextera Index kit (Illumina, San Diego, CA). The cycling condition is 95 °C for 3 min, (95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s) × 8 cycles, 72 °C for 5 min, and held at 4 °C. The second step PCR products were gel purified and quantitated with NanoDrop (ThermoFisher Scientific, Waltham, MA) and 40 ng of the library was then sequenced on one lane for 161 cycles from one end of the fragments on a HiSeq 2500 using a HiSeq SBS sequencing kit version 4 (Illumina, San Diego, CA).

Fastq files were generated and demultiplexed with the bc2fastq v2.17.1.14 conversion software (Illumina, San Diego, CA) and uploaded to the Galaxy Web site (https://usegalaxy.org). The fastq reads were trimmed to only include barcode1-linker-barcode2 using “fastqtrimmer” command. However, since the sequence of the linker was not obvious in the misligated assemblies, the linker sequence had to be removed and only barcode1-barcode2 was analyzed. To do this, “split” command was used to divide the sequence in two from the middle, and “fastqtrimmer” was used to remove the ends of each sequence, and “join” was used to reconnect these sequences resulting in barcode1-barcode2 sequence. A bowtie index (Table S3) was prepared using Python programming language from each two-base combination (96 × 96 = 9216). Extracted barcode sequences were mapped to the bowtie index using Map with Bowtie for Illumina (version 1.1.2) command in Galaxy with commonly used settings. Unmapped reads were removed and reads mapped to each unique barcode sequence were counted.

**Statistical Analysis.** Statistics calculations were performed using the single-factor analysis of variance (one-way ANOVA) method with the Analysis ToolPak extension for Microsoft Excel. To do this, the 96 by 96 ligase fidelity matrix (Table S8) was reformatted as a pairwise list, which was then sorted by the “base difference” between each pair of linkers. Different ranges of data, such as “mismatch type” or “GC content”, were extracted from this list and compared together using one-way ANOVA. A p-value of less than 0.05 was considered to denote statistical significance.

**ASSOCIATED CONTENT**

**Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00480.

Supporting discussions, Figures S1–S2, Tables S1–S6 (PDF)

Table S7 (XLSX)
Table S8 (XLSX)

**AUTHOR INFORMATION**

**Corresponding Author**

*E-mail: zhaos@illinois.edu.*

**ORCID**

Jiazhang Lian: 0000-0001-9784-9876
Huimin Zhao: 0000-0002-9069-6739

**Present Address**

**LifeFoundry, Inc., 60 Hazelwood Dr., Champaign, Illinois 61820, United States.**

**Author Contributions**

*M.H. and S.W. have contributed equally. M.H., R.C., and H.Z. designed the study, M.H., J.L., and S.W. performed the
experiments and analyzed the data, S.W. and M.H. developed the iBioCAD and M.H. and H.Z. drafted the manuscript. R.C. designed the initial iBioCAD GGA design software and the sequencing scheme. All the authors have read and approved the final manuscript.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS
This work was supported by the U.S. Department of Energy (DE-SC0018420). Biotechnology Center at UIUC is acknowledged for their help in NGS sequencing. Christopher Rao, Payman Toudiffar, and Adam Cornell are acknowledged for initial discussions on iBioCAD software design.

■ REFERENCES