ACS Synthetic Biology
Research Article
Cite This: ACS Synth. Biol. 2018, 7, 2665−2674
pubs.acs.org/synthbio

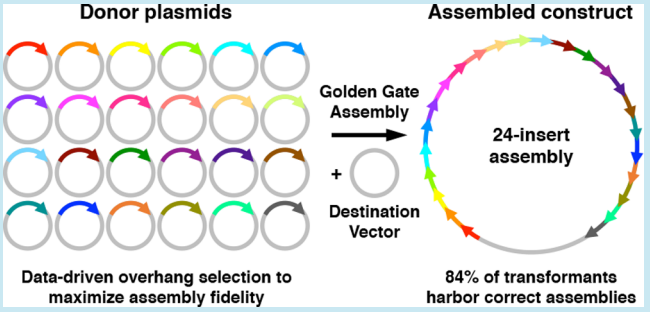
Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly

Vladimir Potapov,[†] Jennifer L. Ong,[†] Rebecca B. Kucera,[‡] Bradley W. Langhorst,[‡] Katharina Bilotti,[†] John M. Pryor,[†] Eric J. Cantor,[‡] Barry Canton,[§] Thomas F. Knight,[§] Thomas C. Evans, Jr.,[†] and Gregory J. S. Lohman*[,†]

[†]Research Department, New England Biolabs, Ipswich, Massachusetts 01938, United States
[‡]Applications and Product Development, New England Biolabs, Ipswich, Massachusetts 01938, United States
[§]Ginkgo Bioworks, Boston, Massachusetts 02210, United States

Supporting Information

ABSTRACT: Synthetic biology relies on the manufacture of large and complex DNA constructs from libraries of genetic parts. Golden Gate and other Type IIS restriction enzyme-dependent DNA assembly methods enable rapid construction of genes and operons through one-pot, multifragment assembly, with the ordering of parts determined by the ligation of Watson−Crick base-paired overhangs. However, ligation of mismatched overhangs leads to erroneous assembly, and low-efficiency Watson Crick pairings can lead to truncated assemblies. Using sets of empirically vetted, high-accuracy junction pairs avoids this issue but limits the number of parts that can be joined in a single reaction. Here, we report the use of comprehensive end-joining ligation fidelity and bias data to predict high accuracy junction sets for Golden Gate assembly. The ligation profile accurately predicted junction fidelity in ten-fragment Golden Gate assembly reactions and enabled accurate and efficient assembly of a *lac* cassette from up to 24-fragments in a single reaction.



KEYWORDS: *DNA ligases, DNA ligase fidelity, Golden Gate assembly, DNA assembly, single-molecule sequencing*

The one-pot assembly of large DNA constructs from smaller component parts is a key technology in modern synthetic biology, with common *in vitro* methods dependent on high-fidelity ligation steps to produce the desired constructs. In restriction enzyme-dependent assembly methods such as BioBricks and Golden Gate cloning, the assembly of large constructs is achieved by the joining of multiple DNA fragments linked by short overhangs.[1−6] To produce the desired final assembly, fragments must be joined only by Watson−Crick overhang pairs; if mismatched overhangs ligate, incorrect assemblies will result with large insertions or deletions of entire fragments or result in one or more fragments being inserted in the incorrect orientation.

T4 DNA ligase is commonly employed in these methods due to its high efficiency in end-joining reactions. However, this enzyme is well-known to join mismatches, gaps, and other imperfect structures with varying levels of efficiency.[7−11] In order to ensure high-fidelity assembly in Golden Gate and derived methods, several rules of thumb have been adopted to minimize the risk of ligating imperfectly base-paired partners during an assembly reaction. In addition to the need to avoid self-complementary palindromic overhangs, it is typically advised to have at least a two base-pair difference between overhangs and to ensure all overhangs have similar GC content in a given assembly.[3−5] Following these rules limits the number of four-base overhangs that can be used in a single pot and is particularly constraining when junction sequences are restricted (e.g., when assemblies must break within coding sequences). Several Golden Gate based assembly systems (e.g., MoClo, Golden Braid, Mobius Assembly, Loop Assembly, MIDAS) have further restricted the number of overhangs to standardized, reliable sets in an effort to improve efficiency and fidelity.[12−20] While very large DNA constructs can be produced from successive hierarchical assembly rounds in these methods, the number of fragments that can be assembled in a single pot is limited by the number of allowable overhang pairs (typically six to eight). While these sets have been vetted empirically, there has been a lack of informatics-driven efforts to choose overhang junctions from comprehensive ligase fidelity data. Thus, the flexibility of Golden Gate methods could be greatly expanded through the identification of large high-fidelity overhang sets, potentially allowing the assembly of many more fragments in a single reaction.

Here, we report the application of a single-molecule sequencing assay to probe the fidelity of DNA ligase end-

Received: August 3, 2018
Published: October 18, 2018

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665−2674

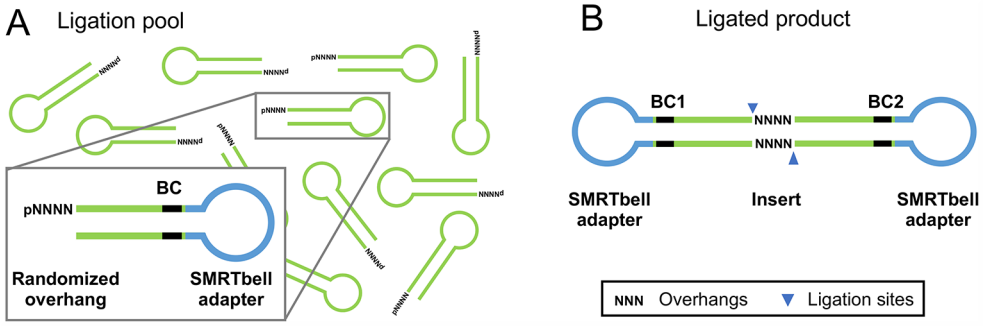


Figure 1. Schematic of multiplexed ligation fidelity and bias profiling assay. (A) Libraries containing randomized four-base overhangs were synthesized and ligated with DNA ligase under various conditions. The hairpin substrates contain the Pacific-Biosciences SMRTbell adapter sequence, an internal 6-base random barcode used to confirm strand identity and monitor the substrate sequence bias derived from oligonucleotide synthesis, and randomized four-base overhangs. (B) Ligated substrates form circular molecules, in which a double-stranded insert DNA is capped with SMRTbell adapters. These products were sequenced utilizing Pacific Biosciences SMRT sequencing, which produced long rolling-circle sequencing reads. Consensus sequences were built for the top and bottom strands independently, allowing extraction of the overhang identity and barcode sequence.

joining of four-base overhangs.[21] We have quantified the ligation efficiency of all four-base Watson−Crick pairs and the prevalence of all possible mismatched overhang combinations by T4 and T7 DNA ligases under typical reaction conditions. We have further applied this data to predict the accuracy of Golden Gate assembly, demonstrating the ability to predict overall assembly fidelity, specific assembly errors, and ligation pairs that exhibit relatively low ligation efficiency despite perfectly complementary overhangs. Finally, we apply the ligation data to guide the choice of ligation junctions in the design of 12- and 24-fragment assemblies of a *lac* cassette.[22] When predicted high-fidelity junction sets are used, efficient assembly of the correct gene construct was observed. Likewise, a designed deletion-prone 12-fragment assembly closely matched predictions. The current data set can thus be used to enumerate sets of junctions with little to no predicted mismatch ligation among any members of the set, allowing >20 fragment, one-pot assemblies with much greater flexibility in choosing junctions than allowed by traditional rules of thumb.

RESULTS

In the current study, we have extended our previously described single molecule sequencing method for profiling DNA ligase end-joining fidelity and bias to four-base overhangs (Figure 1).[21] The multiplexed ligation profile results for overnight ligation at 25 °C with T4 DNA ligase are shown in Figure 2. Four-base overhang ligation libraries showed increased yields with prolonged incubation at 25 °C, with 56 ± 6% ligation at 1 h increasing to 82 ± 3% yield at 18 h. However, fidelity for each overhang changed little from short to long incubation times, despite the increase in yield (Supporting Information, Figure S1), indicating that mismatch ligation was occurring proportionately throughout the ligation time course. Overall fidelity was dramatically improved for four-base overhangs at 37 °C (Supporting Information, Figure S2 and S3); however, the bias in efficiency between overhangs was much more pronounced at 37 °C than at 25 °C, with high AT overhangs notably underrepresented compared to high GC overhangs. This sequence-dependent bias was reduced after 18 h incubation. The increased bias was reflected in lower ligation yields at 37 °C, with 45 ± 2% product at 1 h increasing to 69 ± 3% at 18 h.

In discussing the ligation profile results, individual overhangs are written in the 5′ to 3′ direction with the phosphate omitted, and ligation products are written as overhang pairs with the top overhang written in the 5′ to 3′ direction and the bottom overhang in the 3′ to 5′ direction. For example, $\frac{ATTT}{TAAA}$ represents the fully Watson−Crick paired ligation product between a substrate with a 5′-pATTT overhang and a substrate with a 5′-pAAAT overhang.

Figure 2, Panel A shows a log-scale frequency heat map of all ligation events; Panel B shows the linear frequency of Watson−Crick and mismatch ligation events for each overhang. While a similar overall frequency was observed for the majority of Watson−Crick ligation (Figure 2B, blue bars), several overhangs, notably TNNA, appeared in reduced incidence, despite the ANNT overhangs appearing with similar efficiency to other overhangs. This result mirrored our previous observation with three-base overhangs that TNA overhangs ligate significantly slower than ANT overhangs,[21] though in this case several other high AT% overhangs, such as AAAA and TTTT, were also seen in reduced numbers.

The range of observed ligation fidelity as a function of overhang identity is quite broad, from overhangs with very few mismatch ligation events (e.g., AAGA, CCAA), to those where the majority of ligation partners had at least one base-pair mismatch (e.g., GGCG and GGCC). Overall, there was a weak trend toward lower fidelity with higher GC content. Additionally, GNNN and GGNN sequences were particularly overrepresented in the low-fidelity region. The position dependence of the specific mismatched base-pairings observed with 18 h incubation at 25 °C are visualized in Figure 3 (for additional ligation conditions, see Supporting Information, Figure S4). For both edge (N1:N4, Figure 3A) and middle (N2:N3, Figure 3B) positions, G:T mismatches were the most frequently observed ligation event, with a 5′G and a templating T being significantly more prevalent than a 5′T with a templating G. T:T mismatches were also common, as well as purine:purine mispairs. In the latter case, G:G mismatches dominated in the middle positions, while A:G and G:A mismatches were preferred at the edge position. For most overhangs, it should be noted that the bulk of the mismatch ligation events were derived from pairing a few specific ligation partners. This result suggested the possibility of even "low-fidelity" overhangs being used in high-fidelity ligation sets, as long as their most favorable mismatch ligation partners were not present.

We also tested the fidelity and bias of T7 DNA ligase using the multiplexed ligation profile assay. T7 shows a modest increase in

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665−2674

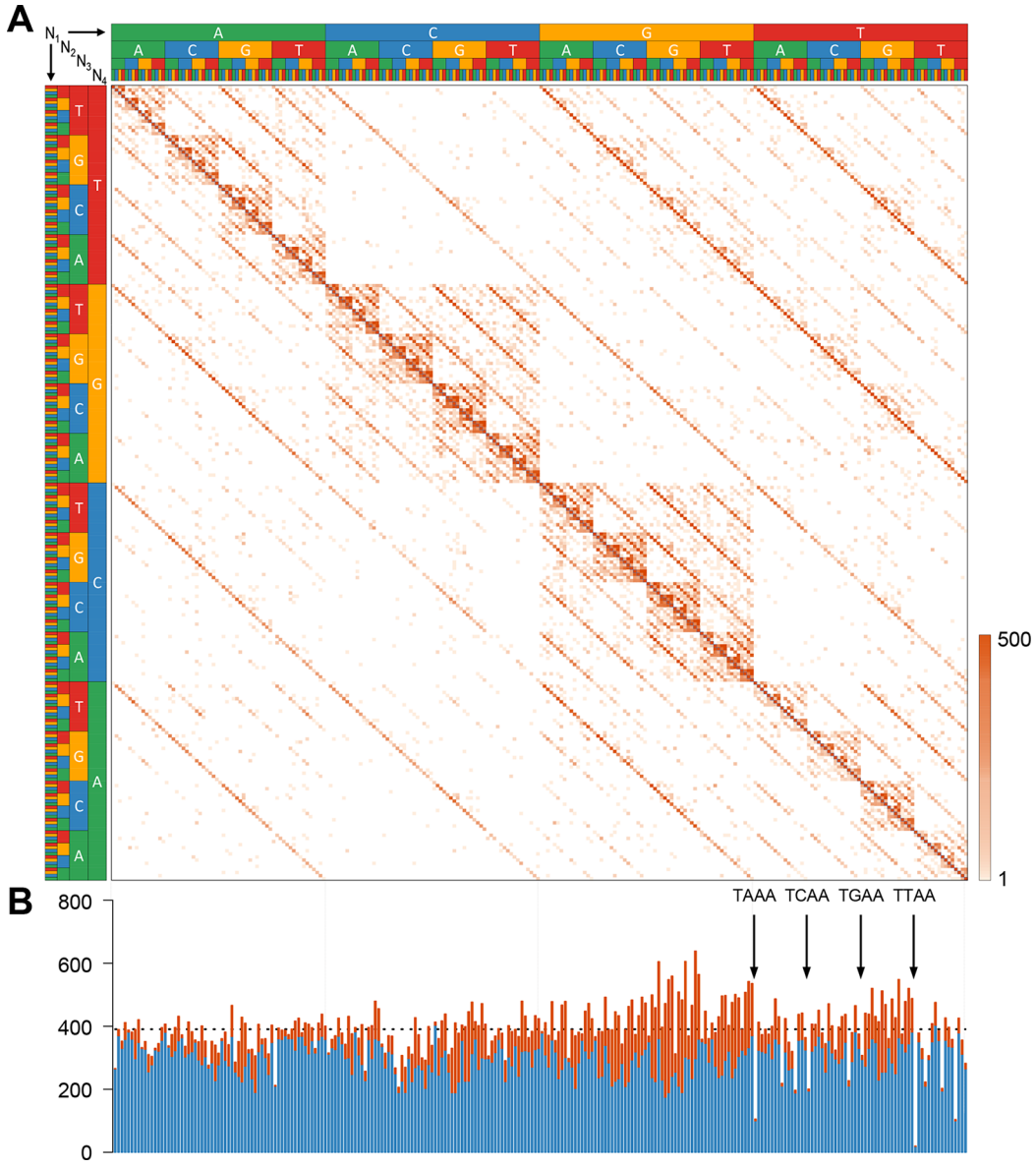


Figure 2. Assay results for the ligation of randomized four-base overhangs by T4 DNA ligase. SMRT sequencing results for ligating 100 nM of the multiplexed four-base overhang substrate 18 h at 25 °C, with 1.75 μM T4 DNA ligase in standard ligation buffer. Observations have been normalized to 100,000 ligation events (see Supporting Information data files for actual observation totals). (A) Frequency heat map of all ligation events (log-scaled). Overhangs are listed alphabetically left to right (AAAA, AAAC···TTTG, TTTT) and bottom to top such that the Watson−Crick pairings are shown on the diagonal. (B) Stacked bar plot showing the frequency of ligation products containing each overhang, corresponding to each column in the heat map in (A). Fully Watson−Crick paired ligation results are indicated in blue, and ligation products containing one or more mismatches are in orange. The dashed line indicates the expected level of ligation if all overhangs appeared in equal frequency.

fidelity compared to T4 DNA ligase, but at the cost of a dramatic increase in overhang sequence bias. (Supporting Information, Figure S5). Indeed, there are many overhangs which resulted in minimal or zero product. The increased bias of T7 also manifests as a loss of ligation efficiency compared to T4 DNA ligase, with a 42 ± 1% yield following a prolonged 18 h incubation at 25 °C. Consistent with trends observed for T4 DNA ligase, the sequence-dependent bias of T7 DNA ligase was even more pronounced at 37 °C and the 18 h ligation yield decreased further to 28 ± 2% (Supporting Information, Figure S6).

Multiplex Ligation Fidelity Data Predicts Multifragment DNA Assembly Accuracy. A key potential application of the ligation fidelity and bias data was to predict the results of one-pot, multifragment assembly reactions. The fidelity profiles were used to guide the selection of junction sets predicted to have low potential for mismatch ligation within the set. To test the predictive power of our ligation fidelity data, 10-fragment systems were designed for Golden Gate assembly, selecting junctions to give a range of predicted outcomes. Fragment inserts (A−J) consisted of 10 randomized 300 bp sequences flanked by specific four-base overhangs and BsaI restriction sites (Figure 4A, Supporting Information, Table S1). For all assemblies, inserts A and J were terminated with an overhang not predicted to pair or mispair with any other overhang present in the assembly reaction (AAAC). The junctions between fragment pairs (Junctions 1−9) were selected to either be 9 high-fidelity (HF) junctions or a low-fidelity (LF) set where 9 junction pairs were chosen such that many mismatch ligation

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665−2674

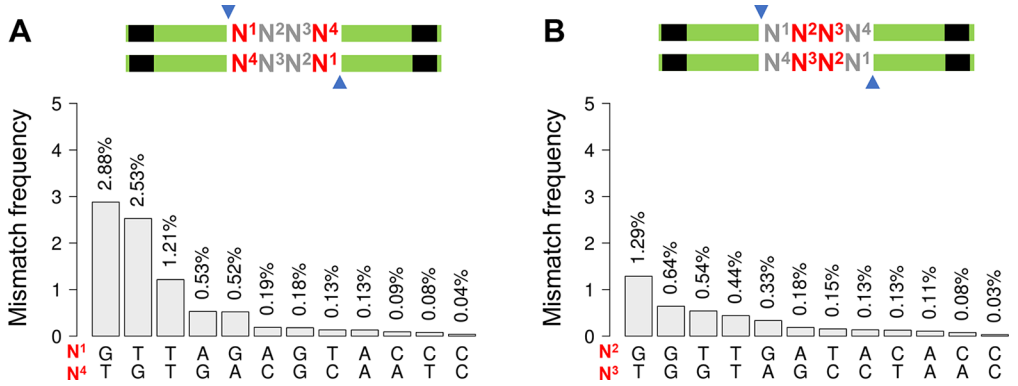


Figure 3. Frequency of specific base pair mismatches by position. Incidence of each possible mismatched base pair observed during ligation of four-base overhangs, with 100 nM of the multiplexed substrate, 1.75 μM T4 DNA ligase, and 18 h incubation at 25 °C in standard ligation buffer. This figure was generated from the same data as shown in Figure 1. (A) shows the results for the edge position (N1:N4); (B) for the middle position (N2:N3). The overhang positions (N^1, N^2, N^3, N^4) are numbered from 5′- to 3′-end for each strand. Each position in N^1:N^4 and N^2:N^3 refers to bases in opposite strands. Note that as strand designation is arbitrary, all ligation products were counted in both orientations (top-to-bottom and bottom-to-top strand).

events were predicted (Supporting Information, Table S2). Two additional sets were designed: a deletion-prone (DP) set, where junction 7 of the HF set was changed to $\frac{GCTG}{CGAC}$ such that deletion (and to a lesser extent, duplication) of insert G was predicted to result, and a failure-prone (FP) set, where junction 7 was replaced with the high fidelity but low efficiency $\frac{TAAA}{ATTT}$ pair.

For each set of inserts, Golden Gate assembly was tested using typical cycling conditions (5 min 37 °C, 5 min 16 °C, 30 cycles), and the product assemblies (all lengths) were sequenced by SMRT sequencing. This method allowed identification of the number, identity, and orientation of fragments in each assembly. The HF set indeed assembled with high fidelity: 99.9% of all observed assemblies were formed by only correct Watson–Crick pairings (Figure 4C). Among assemblies starting with Insert A and ending with Insert J (effectively, those that would be predicted to assemble into the destination plasmid), 99.9% contained all 10 fragments in the proper order and orientation under both reaction conditions. Shorter, incomplete assemblies were equally high fidelity, with most short fragments representing incompatible partial assemblies (e.g., ABCDEFG and GHIJ; see Supporting Data for a list of all fragments), indicating truncations primarily resulted from depletion of limiting fragments, not a failure to fully ligate compatible ends. The LF set showed dramatically increased numbers of erroneous junctions, with 74.1% of assemblies containing at least one erroneous junction (Figure 4D). For the DP set, we observed an increase in 9- and 11-fragment assemblies with an incorrect junction; analysis of the specific structures showed that these were dominated by deletion or duplication of fragment G, as predicted (Figure 4E). The FP (Figure 4F) set displayed a decrease in ligation events at junction 7 (~30% decrease relative to other junctions), and there were many fragments observed truncated at this junction (ABCDEFG and HIJ). While this connection did fail with increased frequency, it was not an impediment to seeing full length assemblies. Additionally, the HF, DP, and FP sets all showed an increased incidence of truncations at junction 6 (the 100% GC junction GCCG/CGGC) and a drop of roughly 23% of connections at junction 6 relative to the average for all junctions, despite this sequence not being a low efficiency junction in the ligation fidelity data set.

In addition to successfully predicting overall fidelity, the specific observed junctions were well predicted by the ligation fidelity profile (Figure 5). Low-frequency mismatch ligation events were not well predicted in any set (less than 10 observations per 100,000 ligation events), with some matching predicted mistakes and others not seen in the fidelity libraries at all. For higher frequency errors, the expected junctions were observed at a frequency very similar to that predicted by the fidelity data, though there was overall a moderately higher incidence of many mismatches than the fidelity library data gathered at 25 °C would predict. Nevertheless, in no case did we see a significant incidence of mismatches not predicted by the fidelity data, nor fail to see a mismatch that was predicted. The same was true of the predictions of the DP and FP sets (Figure S7). Assemblies were also performed at 37 °C; results can be found in the Supporting Data (Supporting Note and Supporting Information, Figure S8 and S9).

Use of Fidelity Predictions Enables 12- and 24-Fragment One Pot Assembly of *lac* Cassettes. We next sought to test the predictive power of our ligation fidelity data in a practical application to select DNA fragment breakpoints and overhang sequences. Thus, we designed 12- and 24-fragment Golden Gate assemblies of a cassette containing both the *lacI* and *lacZ* genes, as well as necessary regulatory elements to drive expression of β-galactosidase (β-gal) (see Supporting Information, Table S3 and S4 for all sequences). Easily scored blue colonies indicated a correctly assembled cassette; white colonies indicated an intact plasmid granting chloramphenicol resistance, but with an incorrectly assembled *lac* cassette. The fidelity data was applied to select junctions for 12- and 24-fragment test systems that were predicted to be high fidelity orthogonal sets, without modification of native sequence. Additionally, we designed a deletion-prone 12-fragment test system with overhangs predicted to frequently result in deletions within the *lac* cassette. The predicted fraction of correctly to incorrectly assembled lac cassette-containing circular plasmids was 99% and 91% for the high fidelity 12- and 24-fragment sets, respectively, and 33% for the low fidelity 12-fragment set.

Assembly of the high-fidelity 12-fragment test system using the typical cycling protocol resulted in large numbers of transformants, with ~99% harboring plasmids with accurate assemblies based on blue/white scoring (Figure 6A and Supporting Information, Table S5), closely agreeing with predictions. Exclusion of any of the 12 fragments resulted in plates with no observed blue colonies (Supporting Information, Table S6). Use of the predicted low-fidelity, deletion-prone test

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665−2674

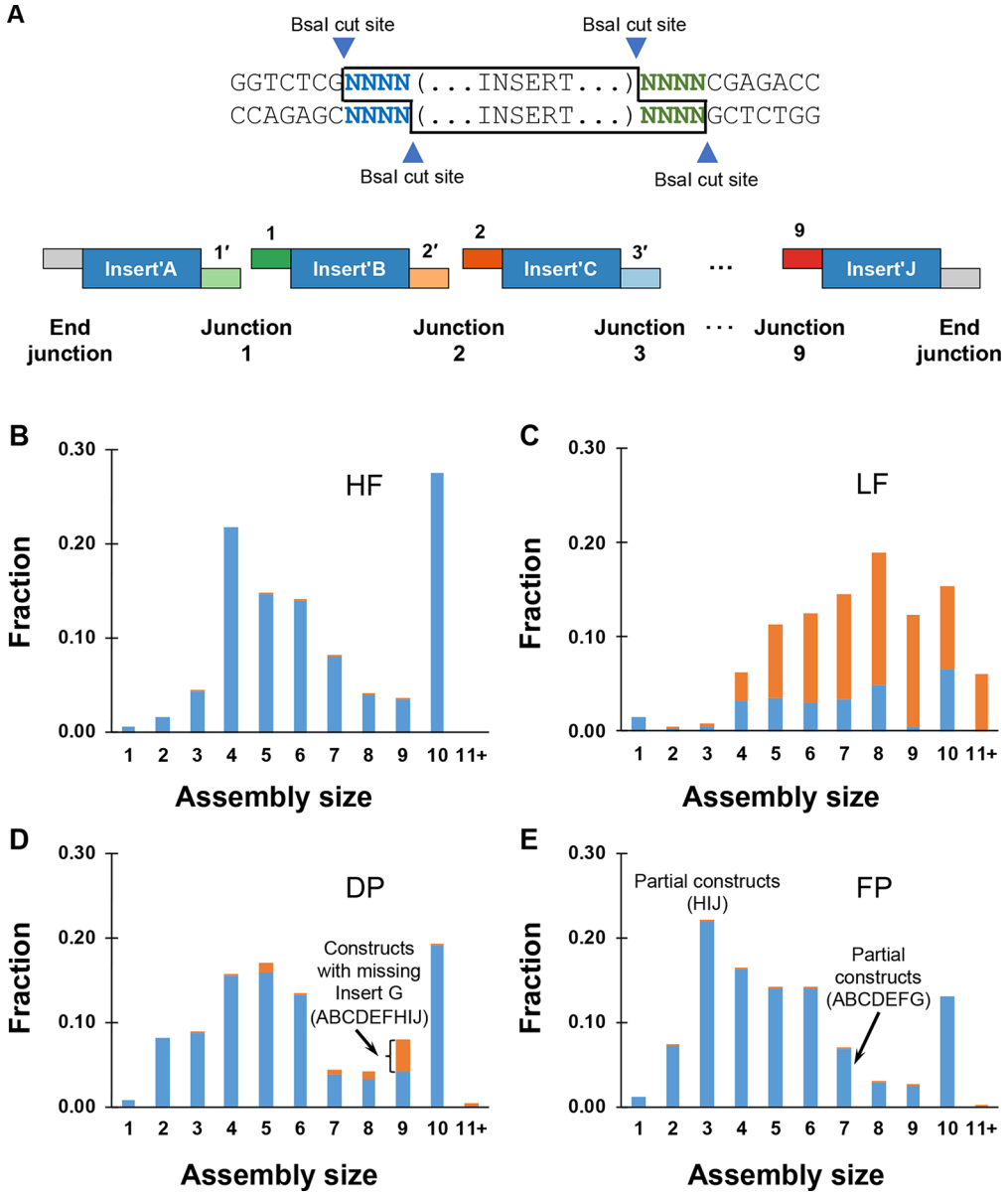


Figure 4. Overview of Golden Gate assembly design. (A) Ten fragments of arbitrary, randomized sequence (Supporting Information, Table S2) were designed, giving 9 junctions and an "end junction" designed with sequence AAAC, which was not predicted to have significant mismatch ligation potential with any overhang used for the junctions. The sequences chosen for the junction differ among the HF, LF, DP, and FP sets, as indicated in Supporting Information, Table S1. The order of assembly could be determined by SMRT sequencing of the products, with the unique insert sequences defining the order of assembly and, thus, which overhangs ligated to produce the connection. For assembly results (B–E), constructs are in blue; constructs containing at least one incorrect junction are shown in orange. (B) HF set results in correctly assembled constructs with the full-length product ABCDEFGHIJ being the most common. (C) LF set expectedly results in a significant fraction of incorrectly assembled constructs. (D) DP set leads to accumulation of a construct with missing Insert G and a slight uptick in 11-insert assemblies duplicating fragment G. (E) FP set has a ligating junction 7 ($\frac{TAAA}{ATTT}$) with predicted low efficiency; this junction did join, but with ∼33% reduced incidence as compared to the other junctions. Additionally, many product fragments truncated at this junction (ABCDEFG and HIJ) were observed (Supporting Data).

system resulted in 45% ± 5% of the transformants harboring correct assemblies, in good agreement with the predicted frequency of ∼33% (Figure 6B, Supporting Information, Table S5). The assembly of the predicted high-fidelity 24-fragment test system (Figure 6C) resulted in a lower count of transformants (∼1 transformant/μL of assembly mix, vs ∼ 100/μL for the twelve-fragment assembly), likely due to the increased number of junctions. However, the observed frequency of transformants expressing β-gal was still quite high, 84 ± 5%, only modestly lower than the predicted 91% of correct assemblies based on the fidelity data. As in the 12-fragment assembly, omitting any one of these fragments eliminated the blue phenotype on scored plates (Supporting Information, Table S7).

DISCUSSION

Herein, we report the comprehensive fidelity and bias profile of T4 and T7 DNA ligases in the joining of four-base overhangs and applied the results to accurately predict high-fidelity overhang sets for Golden Gate assembly. For T4 DNA ligase, similar trends were observed for four-base overhang ligation fidelity with previously reported data for three-base overhangs.[21] As in the case of three-base overhangs, G:T mismatches were

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665–2674

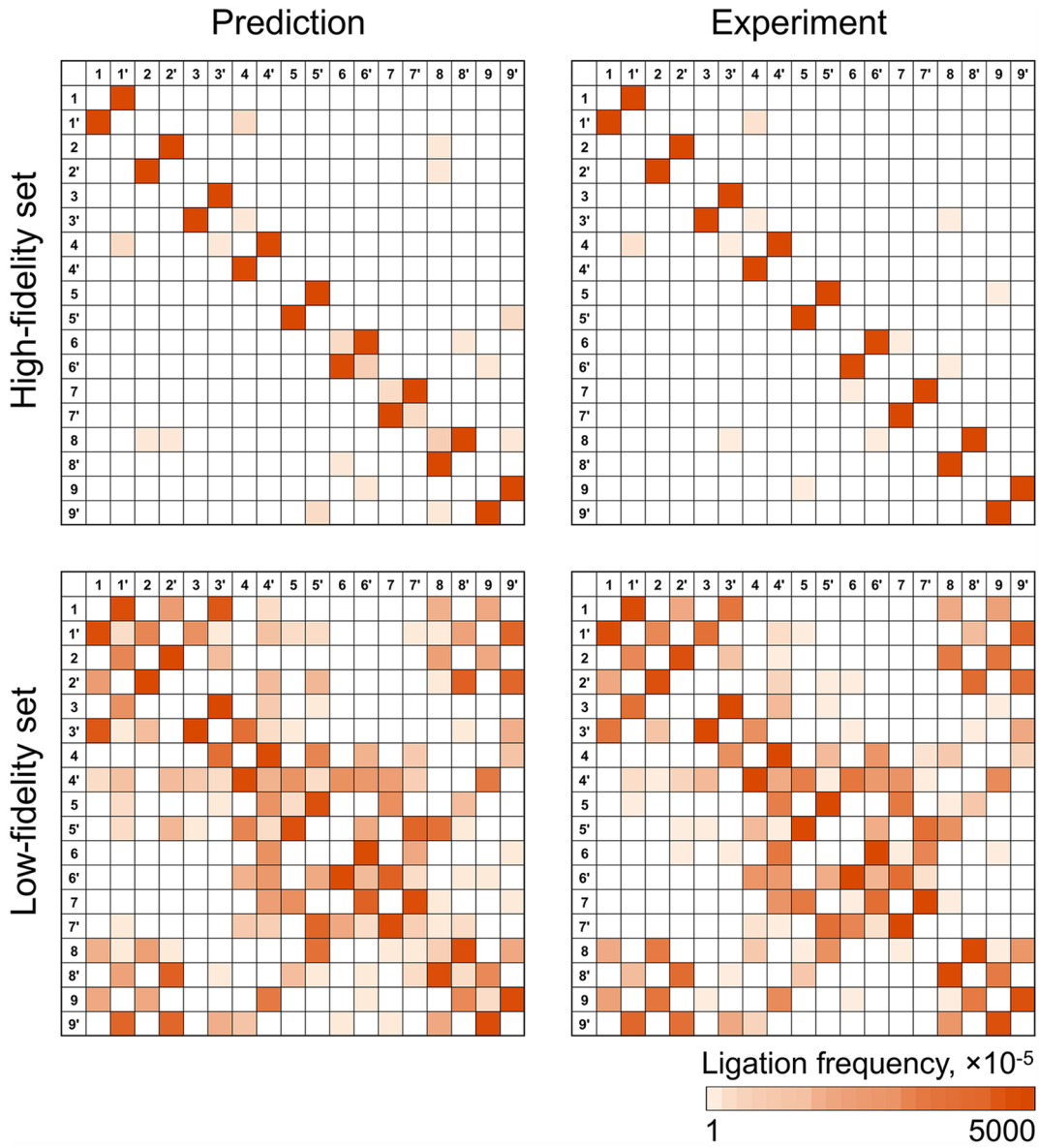


Figure 5. Predicted versus observed fragment linkages in Golden Gate assembly of the HF and LF 10-fragment assemblies. Junction overhangs can be found in Table 1. The intensity of the color corresponds to the number of instances of that junction observed in a Pacific Biosciences SMRT sequencing experiment, normalized to 100,000 total junctions. Predicted frequencies of junctions are based on the fidelity library data generated for the four-base overhang substrate ligated with T4 DNA ligase at 25 °C for 18 h. The experimental observations shown are for assembly of the 10-fragment HF and LF sets with Golden Gate Assembly mix, 37 °C 5 min/16 °C 5 min, 30 cycles.

highly favored, along with lesser amounts of T:T, purine:purine, and A:C mismatches. However, here, very similar profiles were seen at the edge and middle positions, in contrast to three-base overhangs in which mismatches were dramatically disfavored at the middle position, with only T:T mismatches prevalent. Four-base overhangs also lacked the dramatic asymmetry in preference for 5′-purines; while G:T with T in the template was modestly favored over G in the template, this preference was dwarfed by the 10-fold preference observed for three-base overhangs. This result suggests a stronger annealing influence on the mismatch preferences for four-base overhangs compared to three-base overhangs.

T7 DNA ligase was found to have higher overall fidelity than T4 DNA ligase at 25 °C, with the percentage of mismatched products formed by T7 DNA ligase at 25 °C to be comparable to T4 DNA ligase at 37 °C. However, two key drawbacks lead us to caution against the general use of this enzyme in Golden Gate-type assembly. First, the overall ligation efficiency of T7 was notably lower than T4 DNA ligase, showing only half the library yield after 18 h. Second, T7 DNA ligase is dramatically more biased, with many overhangs ligating with efficiency below the mean; its bias at 25 °C is significantly larger than the bias of T4 ligase at 37 °C. Thus, if high fidelity is desired, our data suggests a better approach would be to use T4 DNA ligase with prolonged incubation at 37 °C rather than using traditional cycling protocols and T7 DNA ligase. Nevertheless, it should be possible to achieve high fidelity and efficient assembly if the T7 DNA ligase fidelity data is used to guide the selection of junctions and to avoid any low efficiency Watson−Crick pairs, though the bias of T7 DNA ligase will limit junction choice and maximum assembly size relative to sets designed for use with T4 DNA ligase.

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665−2674

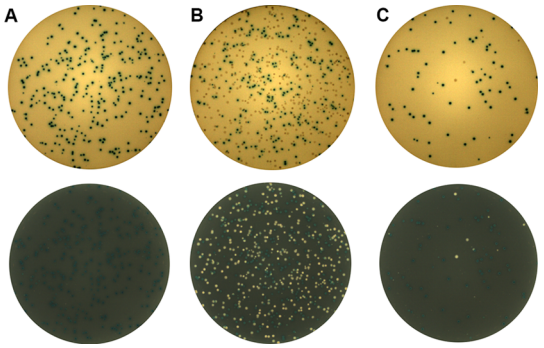


Figure 6. Twelve- and 24-fragment Golden Gate assembly of lac cassettes. Assemblies were twelve (A and B) or twenty-four fragments (C) in a single pot, with choice of junctions driven by the ligation fidelity and bias profile. Reactions were performed as described in the Materials and Methods section, plating 5 μL assembly reaction for 12-fragment assemblies, and 100 μL for 24-fragment assemblies. Plates shown are representative replicates, imaged and counted using the aCOLyte 3 automated colony counting system with a white filter (top) to show blue colonies expressing β-gal, and a black filter (bottom) to visualize white colonies containing antibiotic resistance but a nonfunctional *lac* cassette. Data for all replicates can be found in the Supporting Information, Table S6. (A) shows the results of a designed predicted high fidelity 12-fragment set, predicted 99% blue colonies, observed average over 8 replicates, 99.2 ± 0.6%. (B) shows results of the designed low fidelity, deletion-prone 12-fragment set; predicted 31% blue colonies, observed average of 8 replicates, 45 ± 5%. (C) shows the results of assembly of the designed 24-fragment high fidelity set, predicted 91% blue colonies, observed average over 10 replicates, 84 ± 5%.

While the trends in the ligation fidelity profile data presented here are informative as to the overall mismatch tolerance of DNA ligases, practical utility can also be extracted from the knowledge of precisely which mismatched four-base overhang pairs ligate and which do not. This data set can be used to enumerate sets of "orthogonal" Watson−Crick pairs of overhangs, predicted to result in minimal mismatch ligation. This allows choice of overhangs for use in Golden Gate assemblies with great flexibility. In short, it is not necessary to ensure all overhangs have at least two bases different from all other overhangs, as many pairs with only a single base difference these (e.g., $\frac{AACT}{TTCA}$, $\frac{GAAG}{CTAC}$) form very few if any mismatch ligation products with each other. The ligation profile of T4 DNA ligase can also be used to select a very large number of sets of 10, 12, or even 20+ Watson−Crick pairs with low levels of cross-talk. Further, the data allow identification of low-efficiency Watson−Crick pairs, allowing poorly ligating pairs to be excluded from assembly design. Using these design principles, we successfully produced 10-fragment test assemblies, as well as 12- and 24-fragment lac cassettes that accurately predicted the degree and identity of mismatched connections. Further, assembly of the high fidelity 12- and 24- fragment *lac* cassettes demonstrate that native coding sequences can be divided into >20 fragments without modifying native sequence, yet still assemble with high accuracy in one pot. As an additional possibility, the ability to accurately predict specific mismatch-prone junctions (as in the 10-fragment DP set) could allow design of Golden Gate assembly sets that allow for *in vitro* "alternative splicing" of constructs. As a final note on joining accuracy, overhang mispairing is not predicted to be an issue in traditional cloning methods dependent on Type IIP palindromic cutters. Only trace cross-talk is predicted between all possible palindromic overhangs (Supporting Information, Figure S10), though our data suggests experimenters should avoid the poor-reacting TTAA and TATA overhangs.

It should be noted that while tested assemblies matched predictions very closely, Golden Gate assembly involves a restriction enzyme digestion step plus melting of the overhangs; the multiplexed ligation profiling assay accounts only for annealing and ligation. Thus, slow-melting overhangs (e.g, 100% G/C overhangs) may not assemble with efficiency as high as predicted. Indeed, in the observed 10-fragment assembly results, the frequency of ligation events observed at junction 6 ($\frac{GCCG}{CGGC}$) lagged behind predictions, while other junctions lined up much better with predicted efficiencies. Thus, the reduced efficiency of high GC junctions should be considered along with those of predicted low efficiency (e.g., TNNA) to avoid a reduced yield in assemblies.

When junctions can be arbitrarily chosen, very large assembly sets with high fidelity should be possible. For use in cases where junctions are not restricted by coding sequence, we have enumerated several sets of overhangs predicted to have negligible mismatch-ligation cross talk (>98% correct ligation events). Set 1 (Table 1) contains 15 overhang pairs that include the MoClo standard overhangs; this set is preferred when at least some of the parts have already been designed with this standard in mind.[12,13] If users are not restricted by MoClo, Sets 2−4 contain 20, 25, and 30 overhang pairs that are predicted to achieve the highest fidelity possible for sets of that size. To ensure efficient assembly, these sets exclude all Watson−Crick pairs that ligate significantly below the mean. Overhangs with 100% GC content have additionally been avoided to remove concerns with inefficient melting of these sequences. Use of

Table 1. Predicted High Fidelity Four-Base Overhang Sets for Use with Golden Gate Assembly Methods[a]

Set	Number of overhangs	Estimated fidelity	Overhang sequences[b]
1	15	98.5%	TGCC, GCAA, ACTA, TTAC, CAGA, TGTG, GAGC, AGGA, ATTC, CGAA, ATAG, AAGG, AACT, AAAA, ACCG
2	20	98.1%	AGTG, CAGG, ACTC, AAAA, AGAC, CGAA, ATAG, AACC, TACA, TAGA, ATGC, GATA, CTCC, GTAA, CTGA, ACAA, AGGA, ATTA, ACCG, GCGA
3	25	95.8%	CCTC, CTAA, GACA, GCAC, AATC, GTAA, TGAA, ATTA, CCAG, AGGA, ACAA, TAGA, CGGA, CATA, CAGC, AACG, AAGT, CTCC, AGAT, ACCA, AGTG, GGTA, GCGA, AAAA, ATGA
4	30	91.7%	TACA, CTAA, GGAA, GCCA, CACG, ACTC, CTTC, TCAA, GATA, ACTG, AACT, AAGC, CATA, GACC, AGGA, ATCG, AGAG, ATTA, CGGA, TAGA, AGCA, TGAA, ACAT, CCAG, GTGA, ACGA, ATAC, AAAA, AAGG, CAAC

[a]Sets are provided for use with cycled assembly (16°C/37°C cycles). Set 1 is an extended MoClo set (TGCC, GCAA, ACTA, TTAC, CAGA, TGTG, GAGC) with additional 8 overhangs. All sets are predicted to assemble with a specified overall fidelity if every overhang and its complement is used; subsets of these sets are predicted to have even higher fidelity. Sets for use with static incubation at 37°C can be found in the Supporting Information, Table S8. [b]Only one member of each complementary overhang pair is shown.

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665−2674

these overhang sets, or any subsets thereof, should provide efficient, high-fidelity assembly in Golden Gate reactions.

The current method has proven effective in rapidly profiling the ligation fidelity of T4 and T7 DNA ligases in a single experiment. The data generated has allowed us to accurately predict the efficiency and fidelity of assembly reactions of up to 24 fragments. Further application of the method will allow for profiling the effect of other ligases, buffers, and protocols on ligation fidelity and bias. These data will allow for discovery of high fidelity, low bias ligation conditions that could extend the utility of Type IIS restriction-based assembly systems even further. Finally, modifications of the substrate to include the restriction cleavage and melting steps should increase the accuracy of predictions and allow coscreening of different Type IIS restriction enzymes and ligases in combination. Thus, by combining informatics to guide junction choice and high-throughput screening of conditions, the use of dozens of fragments in a single pot, resulting in highly efficient and highly accurate assembly, is within reach.

MATERIALS AND METHODS

All enzymes and buffers were obtained from New England Biolabs (NEB, Ipswich, MA) unless otherwise noted. T4 DNA ligase reaction buffer (1X) is: 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT. NEBuffer 2 (1X) is: 10 mM Tris-HCl (pH 7.9), 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT. CutSmart Buffer (1X) is: 20 mM Tris-acetate (pH 7.9), 50 mM Potassium Acetate, 10 mM Magnesium Acetate, 100 μg/ml BSA. Thermopol buffer is: 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton-X-100. Standard Taq polymerase buffer is: 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$. SOC outgrowth medium and competent *E. coli* strain T7 Express were from New England Biolabs. The T7 express cell line lacks a functional *lacZ* gene, full genotype: *fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^S*)2 [dcm] R(zgb-210::Tn10--*Tet^S*) endA1* Δ*(mcrC-mrr)114::IS10.* All column cleanup of oligonucleotides and ligated libraries was performed using Monarch PCR & DNA Cleanup Kit columns (NEB), following the published Oligonucleotide Cleanup Protocol (https://www.neb.com/protocols/2017/04/25/oligonucleotide-cleanup-using-monarch-pcr-dna-cleanup-kit-5-g-protocol-neb-t1030). Oligonucleotide purity and sizing was performed using an Agilent Bioanalyzer 2100, using a DNA 1000 assay, following the standard protocols. Synthetic oligonucleotides were obtained from Integrated DNA Technologies as lyophilized solid (Coralville, IA). Fragments for Golden Gate assembly assays were obtained from GenScript (Piscataway, NJ), as precloned inserts flanked by BsaI cut sites in a pUC57-mini plasmid with the native BsaI site in the amp^R gene removed through silent mutagenesis.

Preparation and Pacific Biosciences SMRT Sequencing of Ligation Fidelity Libraries. The substrate for the four-base overhang ligation fidelity assay was produced using the protocol previously published for three-base overhangs, with the following modifications.[21] Initial PAGE-purified substrate precursor oligonucleotide (IDT) contained a 5′-terminal region, a randomized four-base region, a BsaI binding site, a constant region, an internal 6-base randomized region as a control for synthesis bias, and a region corresponding to the SMRT-bell sequencing adapter for Pacific Biosciences SMRT sequencing (Supporting Information, Table S9). The oligonucleotide was designed with a short (7-base) complementary region such that they form a primer-template junction hairpin structure. The precursor oligonucleotide was extended as per the published method.[21] The extended DNA was purified (Monarch PCR & DNA Cleanup Kit), and the concentration of the purified DNA (typically 25−30 μM) was determined using an Agilent Bioanalyzer 2100, DNA 1000 kit.

The extended DNA was cut using BsaI to generate a four-base overhang. Typically, 1 μL of DNA from the extension reaction was combined with 900 U of BsaI in a 100 μL total volume of NEB CutSmart buffer and incubated for 2 h at 37 °C. Reactions were halted by addition of 1 μL Proteinase K followed by 20 min incubation at 37 °C and then purified using the Monarch PCR & DNA Cleanup Kit (NEB). Final concentration and extent of cutting was determined by Agilent Bioanalyzer (DNA 1000) and confirmed to be >95% cut. Remaining uncut starting material (∼5%) was not 5′ phosphorylated and thus should not interfere with subsequent cohesive-end joining reactions. For use in subsequent steps, DNA substrates were diluted to ∼500 nM in 1X TE buffer, with precise concentration determined by Bioanalyzer. The final substrate sequence can be found in the Supporting Information, Table S1.

In a typical ligation reaction, substrate (100 nM) was combined with 2.5 μL high concentration T4 DNA ligase or T7 DNA ligase (2000 U, 1.75 μM final concentration) in 1x T4 DNA ligase buffer in a 50 μL total reaction volume and incubated for 1 or 18 h at 25 or 37 °C. Reactions were quenched with 2.5 μL of 500 mM EDTA and purified using the Monarch PCR & DNA Cleanup Kit, oligonucleotide cleanup protocol. Each ligation was performed in a minimum of duplicates, and the ligation yield was determined by Agilent Bioanalyzer (DNA 1000) with error reported as one standard deviation. The ligated library was treated with Exonuclease III (50U) and Exonuclease VII (5 U) in a 50 μL volume in 1X Standard Taq Polymerase buffer for a 1 h incubation at 37 °C. The library was purified using a Monarch PCR & DNA Cleanup Kit, oligonucleotide cleanup protocol, including a second wash step, and then quantified by Agilent Bioanalyzer (DNA 1000). Typical concentrations of final library were between 0.5 and 2 ng/μL. Sequencing and analysis of sequencing data were performed as previously described,[21] with the scripts modified to use the expected insert sequence from the four-base overhang ligation reactions (Supporting Information, Table S9). Total observations for all experiments can be found in Supporting Information, Table S10. Full results from each experiment are supplied as .csv files in the Supporting Information.

Preparation and SMRT Sequencing Analysis of Ten-Fragment Golden Gate Assemblies. Insert sequences were designed by combining all possible 4-base sequences in random order to generate 1024nt sequences. Three such sequences were combined and then divided into ten 300nt fragments (discarding the remaining 72nt fragment). Homopolymer regions of length greater than 4 and BsaI restriction sites were excluded from all sequences. The final ten fragments (Supporting Information, Table S2) were obtained precloned into pUC57-mini plasmids, flanked with BsaI cut sites designed to allow arbitrary 4-base overhangs to be installed on the ends of each insert; the junction overhangs used in each assembly set are specified in Table 1.

Insert fragments were amplified using PCR primers designed to anneal to the plasmid region flanking the insert site (P1 GGGTTCCGCGCACATTTC; P2 TTTGCTGGCC-TTTTGCTCACAT). PCR reactions included 100 pg/μL plasmid, 0.5 μM each primer, 2 U Q5 High Fidelity DNA

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665−2674

polymerase, and 0.2 mM each dNTP in Q5 Reaction buffer in a 100 μL total reaction volume. Reactions were incubated 30 s at 98 °C, 16 cycles of 5 s at 98 °C, 10 s at 62 °C, and 20 s at 72 °C, with a final incubation of 5 min at 72 °C, with the exception of insert D for the low-fidelity set, which was amplified by EpiMark Hot Start *Taq* DNA polymerase using the plasmid, primer, dNTP concentration as above, and 2.5 U EpiMark Hot Start *Taq* DNA polymerase in 1X EpiMark Hot Start *Taq* Reaction Buffer (and cycled 30 s at 95 °C, 20 cycles of 15 s at 95 °C, 15 s at 55 °C, 3 0s at 68 °C, with a final incubation of 1 min at 68 °C). The amplified inserts were purified and size selected using AMPure XP beads with a first bead selection of 0.55X volume of beads and a second bead selection of 0.35X sample volume. Concentration and purity of the final fragments were assessed by Agilent Bioanalyzer (DNA 1000).

Ligation assemblies were prepared with each fragment at a 5 nM final concentration in 1X T4 DNA ligase buffer with 2.5 μL of NEB Golden Gate Assembly Enzyme Mix in a volume of 50 μL. Reactions were incubated for 1 or 18 h at 37 °C and then heat inactivated for 10 min at 65 °C. Alternatively, reactions were carried out with 30 cycles of 1 or 5 min at 37 °C and 1 or 5 min at 16 °C, followed by a 5 min 65 °C heat inactivation step. Ligation products were blunted using the NEB Quick Blunting kit, adding 10 μL of 10X Quick Blunting buffer, 10 μL of 1 mM dNTPs, and 4 μL of Quick Blunting Enzyme Mix (final volume 75 μL) incubating for 1 h at 25 °C, followed by cleanup using Monarch PCR & DNA Cleanup Kit columns. SMRT adapters were ligated by adding 2 μL of SMRTbell blunt adapter (20 μM, Pacific Biosciences) and 25 μL of Blunt/TA Ligase Master Mix in a final volume of 50 μL, incubating for 1 h at 25 °C, and then purification with Monarch PCR & DNA Cleanup Kit columns. The adapter-ligated library was treated with 1 μL of PreCR Repair Mix in 1X Thermopol buffer, 1 mM each dNTP and 0.5 mM NAD+ in a 50 μL total volume, incubated 20 min at 37 °C, then cleaned up with Monarch PCR & DNA Cleanup Kit columns. Finally, libraries were treated with 50 U Exonuclease III and 5 U Exonuclease VII in 1X Standard Taq Polymerase buffer in a 50 μL total volume and incubated for 1 h at 37 °C. Final libraries were purified twice using a 1X volume of Ampure PB beads (Pacific Biosciences). Average fragment size was estimated by Agilent Bioanalyzer DNA 12000 assay (typically 1800−2400 bp), and total DNA concentration (typically 2−4 ng/μL) was determined. Libraries were prepared for sequencing according to the Pacific Biosciences Binding Calculator Version 2.3.1.1 and the DNA/Polymerase Binding Kit P6 v2 using the Magbead OCPW protocol and no DNA control complex. Libraries were sequenced on a Pacific Biosciences RSII, 1 SMRT cells per library, with a 6 h data collection time.

Consensus sequences were built for fragment assembly libraries with the Arrow algorithm using the ccs program from SMRT Link software. Each consensus sequence represents a result of ligating multiple Golden Gate fragments into a single assembly such that the resulting consensus reads are comprised of long fragments separated by short regions corresponding to ligation junctions. Given the ten known 300nt fragments, their coordinates and mapping direction in each consensus read from assembly libraries were determined using BLAST software. This information was then used to tabulate the frequency of pairwise ligation events and overall composition of assemblies. A number of filtering steps were applied to ensure the integrity of the derived data. Any 300nt fragment was required to map entirely from the first to the last nucleotide in the consensus read. Additionally, only two types of ligation junctions were expected to be seen in consensus reads: junctions of length 4 corresponding to overhang ligation during assembly reaction and junction of length 8 corresponding to blunt ligation during SMRT library preparation workflow. A 1nt variation in length was permitted for each junction type to account for possible errors in the sequencing reads. If any of the above conditions were not met, the resulting consensus read was excluded from the analysis. When a blunt ligation junction was detected in the consensus read, the entire read was split apart at such junctions.

Golden Gate Cloning of 12 and 24 Fragment *lac* Cassettes. Golden Gate assembly reactions consisted of 75 ng pGGA destination plasmid, 75 ng of each precloned DNA fragment (See Supporting Information, Tables S3 and S4 for fragment and junction identities), 500 U T4 DNA ligase, and 15 U BsaI-HFv2 (BsaI isoschizomer) in a final volume of 20 μL (12 fragment assemblies) or 25 μL (24 fragment assemblies) unless otherwise noted in the text. Reactions were kept cold with the use of a prechilled aluminum cold block before transfer to a T100 thermal cycler (Bio-Rad). Assembly reactions were incubated with thirty cycles consisting of 5 min at 37 °C and 5 min at 16 °C, followed by a 5 min final incubation step at 55 °C and then a final 4 °C hold prior to transformation. Transformations were performed using 2 μL of each assembly reaction added to 50 μL competent T7 Express cells, incubation on ice for 30 min, and incubation at 42 °C for 10 s, with a final 5 min recovery period on ice. SOC outgrowth medium (950 μL) was added and the cells were incubated 1 h at 37 °C with rotation. The outgrowth was then placed on ice for 5 min before plating 5 μL (12 fragment assemblies) or 100 μL (24 fragment assemblies) using bead spreading on prewarmed agar plates (Luria−Bertani broth supplemented with 1 mg/mL dextrose, 1 mg/mL $MgCl_2$, 30 μg/mL Chloramphenicol, 200 μM IPTG and 80 μg/mL X-gal).[22] Plates were inverted and placed at 37 °C for 18 h and then stored at 4 °C for 8 h before scoring colony color phenotype. Plates were imaged and counted using the aCOLyte 3 automated colony counting system (Synbiosis) or by hand. For each assembly type, total transformants and percentage correct assemblies (blue colonies) are reported as the average result of at least three independent assembly reaction replicates, with the reported error one standard deviation from the mean.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00333.

Supporting Note, Supporting Figures S1−S10, and Supporting Tables S1−S10 (PDF)

Supporting Data (CSV formatted data tables for raw ligation product observation counts and 10-fragment Golden Gate assembly reactions) (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*Tel: 1-978-998-7916; Fax: 978-921-1350; E-mail: lohman@neb.com.

ORCID

Gregory J. S. Lohman: 0000-0002-0638-7555

Author Contributions

G.J.S.L, T.C.E, B.C., and T.F.K. conceived the project. G.J.S.L. with J.O., K.B., R.B.K., and J.P. performed the experiments. V.P. with B.W.L performed the bioinformatics analysis of the sequencing data and designed the data visualizations. V.P. and

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665−2674

G.J.S.L. wrote the manuscript with editing assistance from K.B., B.C., and T.C.E. G.J.S.L. and E.J.C supervised the project.

Notes

The authors declare the following competing financial interest(s): Vladimir Potapov, Jennifer L. Ong, Rebecca B. Kucera, Bradley W. Langhorst, Katharina Bilotti, John M. Pryor, Eric J. Cantor, Thomas C. Evans, Jr., and Gregory J. S. Lohman are employees of New England Biolabs, a manufacturer and vendor of molecular biology reagents, including DNA ligases and restriction endonucleases. This affiliation does not affect the authors' impartiality, adherence to journal standards and policies, or availability of data. Barry Canton and Thomas F. Knight are employees of Ginkgo Bioworks, Inc., a corporation that uses enzymes and reagents for gene synthesis in the course of developing engineered microbes. This affiliation does not affect the authors' impartiality, adherence to journal standards and policies, or availability of data.

Sequencing data pertaining to this study has been deposited into the Sequencing Read Archive under accession numbers SRP144368 (multiplexed ligase fidelity sequencing data) and SRP144386 (golden gate sequencing data). Custom software tools are available in the GitHub repository at: https://github.com/potapovneb/golden-gate.

■ ACKNOWLEDGMENTS

We would like to thank Laurence Ettwiller, Laurie Mazzola, Rick Morgan, Yvette Luyten (NEB), and Pacific Biosciences for assistance with sequencing reactions. We are grateful to Bill Jack, Andy Gardner, and Karen Lohman for critical feedback on this manuscript. This work was supported entirely by internal funding from NEB and Ginkgo Bioworks. Funding for open access charge: New England Biolabs.

■ REFERENCES

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

(2) Smolke, C. D. (2009) Building outside of the box: iGEM and the BioBricks Foundation. *Nat. Biotechnol. 27*, 1099−1102.

(3) 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.

(4) Engler, C., and Marillonnet, S. (2011) Generation of families of construct variants using golden gate shuffling. *Methods Mol. Biol. 729*, 167−181.

(5) Engler, C., and Marillonnet, S. (2014) Golden Gate cloning. *Methods Mol. Biol. 1116*, 119−131.

(6) Marillonnet, S., and Werner, S. (2015) Assembly of Multigene Constructs Using Golden Gate Cloning. *Methods Mol. Biol. 1321*, 269−284.

(7) Nilsson, S. V., and Magnusson, G. (1982) Sealing of gaps in duplex DNA by T4 DNA ligase. *Nucleic Acids Res. 10*, 1425−1437.

(8) Goffin, C., Bailly, V., and Verly, W. G. (1987) Nicks 3′ or 5′ to AP sites or to mispaired bases, and one-nucleotide gaps can be sealed by T4 DNA ligase. *Nucleic Acids Res. 15*, 8755−8771.

(9) Wu, D. Y., and Wallace, R. B. (1989) Specificity of the nick-closing activity of bacteriophage T4 DNA ligase. *Gene 76*, 245−254.

(10) Harada, K., and Orgel, L. E. (1993) Unexpected substrate specificity of T4 DNA ligase revealed by in vitro selection. *Nucleic Acids Res. 21*, 2287−2291.

(11) Showalter, A. K., Lamarche, B. J., Bakhtina, M., Su, M. I., Tang, K. H., and Tsai, M. D. (2006) Mechanistic comparison of high-fidelity and error-prone DNA polymerases and ligases involved in DNA repair. *Chem. Rev. 106*, 340−360.

(12) 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.

(13) Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012) Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng Bugs 3*, 38−43.

(14) 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.

(15) Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., and Orzaez, D. (2013) GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol. 162*, 1618−1631.

(16) Vazquez-Vilar, M., Sarrion-Perdigones, A., Ziarsolo, P., Blanca, J., Granell, A., and Orzaez, D. (2015) Software-assisted stacking of gene modules using GoldenBraid 2.0 DNA-assembly framework. *Methods Mol. Biol. 1284*, 399−420.

(17) Iverson, S. V., Haddock, T. L., Beal, J., and Densmore, D. M. (2016) CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol. 5*, 99−103.

(18) Andreou, A. I., and Nakayama, N. (2018) Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLoS One 13*, e0189892.

(19) van Dolleweerd, C. J., Kessans, S. A., Van de Bittner, K. C., Bustamante, L. Y., Bundela, R., Scott, B., Nicholson, M. J., and Parker, E. J. (2018) MIDAS: A Modular DNA Assembly System for Synthetic Biology, *ACS Synth. Biol.*. 7, 1018

(20) Pollak, B., Cerda, A., Delmans, M., Álamos, S., Moyano, T., West, A., Gutiérrez, R. A., Patron, N., Federici, F., and Haseloff, J. (2018) Loop Assembly: a simple and open system for recursive fabrication of DNA circuits. *BioRxiv*, DOI: 10.1101/247593

(21) Potapov, V., Ong, J. L., Langhorst, B. W., Bilotti, K., Cahoon, D., Canton, B., Knight, T. F., Evans, T. C., Jr., and Lohman, G. J. S. (2018) A single-molecule sequencing assay for the comprehensive profiling of T4 DNA ligase fidelity and bias during DNA end-joining. *Nucleic Acids Res.*, https://doi.org/10.1093/nar/gky303.

(22) Sambrook, J., and Russell, D. W. (2001) Screening Bacterial Colonies Using X-gal and IPTG: alpha-Complementation. In *Molecular Cloning - A Laboratory Manual* 3rd ed., p 127, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

DOI: 10.1021/acssynbio.8b00333
ACS Synth. Biol. 2018, 7, 2665−2674