Engineering 5 (2019) 803–810

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Synthetic Biology—Article

Developing a Transformation-Independent and Unbiased qPCR Assay to Rapidly Evaluate the Determinants of DNA Assembly Efficiency

Xiaoyan Ma [a,#], Xinxin Liang [a,#], Yi-Xin Huo [a,b,*]

[a] Key Laboratory of Molecular Medicine and Biotherapy, School of Life Science, Beijing Institute of Technology, Beijing 100081, China
[b] UCLA Institute for Technology Advancement (Suzhou), Suzhou 215123, China

ARTICLE INFO

Article history:
Received 1 February 2018
Revised 29 July 2018
Accepted 19 March 2019
Available online 26 June 2019

Keywords:
Assembly efficiency
DNA assembly
qPCR
Secondary structure
Transformation

ABSTRACT

Synthetic biology is moving in the direction of larger and more sophisticated design, which depends heavily on the efficient assembly of genetic modules. Conventional evaluation of the DNA assembly efficiency (AE) requires transformation, and the whole process requires up to 10 h and is susceptible to various interferences. To achieve rapid and reliable determination of the AE, an alternative transformation-independent method was established using a modified quantitative polymerase chain reaction (qPCR) assay. The AE is represented by the proportion of the ligated fragment, which can be determined within 3 h. This qPCR-based measurement was tested by the commonly used restriction ligation, Golden Gate assembly, and Gibson assembly for the assembly of two or more DNA pieces; the results correlated significantly with the AEs represented by the counting of the colony-forming units (CFUs). This method outperformed the CFU-based measurement by reducing the measuring bias and the random deviations that stem from the transformation process. The method was then employed to investigate the effects of terminal secondary structures on DNA assembly. The results revealed the major effects of the overall properties of the overlap sequence and the negative effects of hairpin structures on the AE, which are relevant for all assembly techniques that rely on homologous annealing of the terminal sequences. The qPCR-based approach presented here should facilitate the development of DNA assembly techniques and the diagnosis of inefficient assemblies.

1. Introduction

The goal of synthetic biology is to modify the behavior of organisms or create new life forms to perform novel tasks [1]. This discipline builds on biological molecules such as DNA and protein, and follows the hierarchy to constitute biochemical reactions, metabolic pathways, cells, and finally populations [1,2]. However, development toward more sophisticated biological systems is greatly hindered by two challenges. The first of these is to increase the diversity of artificial biological constructs [3], as the current size of constructed libraries for genetic parts is usually orders of magnitude (10^3–10^5) below the desired level [4,5]. The second challenge is the integration of multiple genetic modules [6], especially when it comes to more than 10 biological parts [7]. Overcoming these challenges will rely on DNA assembly techniques that offer easy and fast operation, superior compatibility, and high efficiency [8].

The assembly efficiency (AE) reflects the probability of the ligation event between DNA fragments, and is the key parameter indicating the potential of an assembly method. It is generally determined by the number of colonies formed after transforming the assembly products [9,10]. Measurement of the AE begins with the transformation process, which takes about 1–1.5 h, followed by overnight incubation for at least 8 h to allow colony formation. Finally, there is the laborious step of colony counting and verification. These steps require up to 10 h, in addition to the time-consuming preparation of the competent cells. Thus, rapid determination of the AE would greatly speed up the cycle of testing and optimization in the development of assembly techniques.

The assembly of DNA molecules has evolved from site-dependent ligations of two DNA pieces to sequence-independent combinations of multiple fragments with desired orders [11]. Existing methods such as *in vivo* assembly (IVA) [12], twin-primer assembly (TPA) [13], sequence and ligation independent

* Corresponding author.
E-mail address: huoyixin@bit.edu.cn (Y.-X. Huo).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.eng.2019.06.002

cloning (SLIC) [10], and circular polymerase extension cloning (CPEC) [9] have the highest AEs, which range from 400 to 56 000 colony-forming units (CFUs) per unit of DNA. However, these values can be misleading. On the one hand, calculation of the AE varies among studies. Some studies use the amount of total DNA for normalization and express the efficiency as CFUs per nanogram DNA (CFUs·ng^{-1}) [14], while others bring the amount of either the vector or the insert into the calculation [9,15]. Expressions such as CFUs/plate [12,16] and the proportion of positive CFUs [17] have also been used. Therefore, the efficiencies reported by different assembly techniques may not be comparable. On the other hand, the AE represented by the CFU count can be affected by the transformation process. Any factor that affects the transformation efficiency, such as the genotype of the competent cell, the approach of transformation, or the size of the assembly product [18,19], may lead to dramatic variations in the colony count. As a result, the CFU-based measurement may not represent the actual AE, making comparisons among different techniques less reliable even if the same calculation method is employed. Thus, there is an urgent need to develop a transformation-independent method for the standardized and unbiased evaluation of the DNA AE.

Most assembly technologies, such as SLIC, IVA, In-Fusion [20], and Gibson assembly [21], require overlap sequences at the ends of the adjacent fragments. Assembly of the fragments relies on easy and stable annealing of the complementary overhangs once the overlap regions are recessed. Therefore, design of the overlap sequences is crucial in achieving high AE. The influence of secondary structures in the overlap region on the AE has received much less attention in comparison with factors such as the overlap length, which has been optimized intensively for most assembly techniques [10,12,14]. Since the overlap region must be processed into a single strand before assembly, secondary structures formed within or between the overhangs might prevent successful annealing of the complementary terminal sequences. This concern has been addressed by several online protocols (e.g., Gibson Assembly Cloning from Addgene; gBlocks Gene Fragments Protocol from the Integrated DNA Technologies, Inc.), yet experimental evidence is still lacking.

In this study, a transformation-independent measurement of the AE was established using a modified quantitative polymerase chain reaction (qPCR) assay (Figs. 1(a–c)). To summarize, the DNA fragments were first assembled, producing a mixture of ligated circular DNA and unligated linear fragments. After treatment by T5 exonuclease, only the assembled circular DNA remained intact, while the linear insert and vector were digested. The relative quantity of the ligated fragment to its initial amount was measured by qPCR assay, and was used to indicate the AE. This approach was applied to three widely used assembly techniques—namely, restriction ligation, Golden Gate assembly, and Gibson assembly—and the results were compared with those of CFU-based measurement. To demonstrate the application of the qPCR-based measurement, the effects of secondary structures in the overlap regions on the efficiency of Gibson assembly were investigated. The results of this study suggest that the qPCR-based method can generate a less biased evaluation of the AE in comparison with the conventional CFU-based approach, and in a much faster way; therefore, it can be employed for the development and optimization of DNA assembly techniques.

2. Materials and methods

2.1. Restriction ligation

Seven sets of restriction ligation (RL1–RL7) were tested. The enzymes BamHI and SalI, or PacI and FseI were used for double digestion. When necessary, the restriction sites were introduced into the ends of the DNA fragments by primers via polymerase chain reaction (PCR) amplification, and the remaining template plasmids were eliminated by DpnI. For double digestion, 1 μg of insert fragments and 1 μg of vectors were each cut by 1 U (one unit is defined here as the amount of restriction enzyme required to digest 1 μg of λ DNA in 1 h at 37 °C in a total reaction volume of 50 μL) of both enzymes in a total reaction volume of 50 μL. The reaction mix was incubated at 37 °C for 2 h and the targeted fragments were gel purified using a GeneJET Gel Extraction Kit (Thermo Fisher Scientific Inc.). A total of 60–350 ng DNA fragments

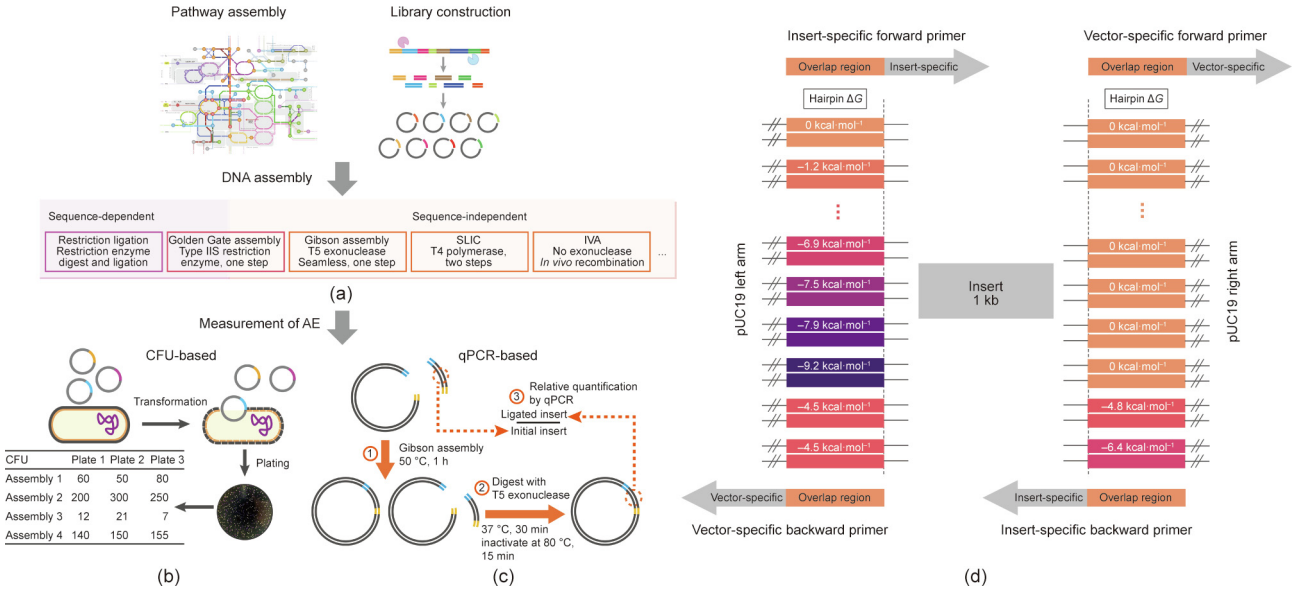

Fig. 1. (a) Constructions of complex metabolic pathways and genomic libraries rely on the efficient assembly of DNA fragments, a method that has been evolved from sequence-dependent to sequence-independent approaches. (b) The AE is usually determined by transformation and subsequent colony counting. (c) An alternative measurement of the AE was established by harnessing the qPCR assay; the ratio of the amount of the ligated fragment to its initial amount was used to indicate the AE. (d) A series of overlap sequences with different secondary structures were introduced to the ends of a linearized pUC19 and a 1 kilobase (kb) fragment in order to investigate the effects of terminal secondary structures on Gibson assembly. ΔG: free energy difference. 1 kcal = 4186 J.

with an insert-to-vector molar ratio of 1:1 or 5:1 were ligated by 400 cohesive end ligation units (CEU) of T4 DNA ligase in a total volume of 20 μL at 22 °C for 1 h. All enzymes were from New England Biolabs, Inc.

2.2. Golden Gate assembly

Eight sets of Golden Gate assembly (GG1–GG8) were tested. The BbsI restriction sites flanked by a four-base complementary sequence that directs the assembly of the ligation were added to the ends of the DNA fragments by PCR amplification. A total of 200–500 ng DNA with an insert-to-vector molar ratio of 1:1, 3:1, or 5:1, 2 μL of 10× T4 ligase buffer, 200 CEU of T4 DNA ligase, and 5 U of BbsI were mixed to a total volume of 20 μL. The reaction was performed for 30 cycles, for 5 min at 37 °C and for 5 min at 22 °C per cycle. This was followed by inactivation of the restriction enzyme at 55 °C for 15 min and inactivation of the T4 ligase at 85 °C for 15 min. All enzymes and buffers used were from New England Biolabs, Inc.

2.3. Gibson assembly

The fragments to be ligated were PCR-amplified to add the terminal overlaps. Overlaps of 20 base pairs (bp) were usually introduced for the assembly of two DNA fragments, and the overlap regions were extended to 40 bp for multi-part DNA assembly. Fifty femtomoles of each of the vector and insert were added to 11.25 μL lab-made assembly master containing 0.06 U of T5 exonuclease, 0.375 U of Phusion® High-Fidelity DNA Polymerase, 60 CEU of Taq DNA ligase, 13.3 mmol·L^{-1} of $MgCl_2$, 13.3 mmol·L^{-1} of 1,4-dithiothreitol (DTT), and 1.33× isothermal reaction buffer (0.1 mol·L^{-1} tris(hydroxymethyl)aminomethane-hydrochloride (Tris-HCl), pH 7.5, 0.2 mmol·L^{-1} each deoxy-ribonucleoside triphosphate (dNTP), 1 mmol·L^{-1} nicotinamide adenine dinucleotide (NAD), and 150 mg polyethylene glycol (PEG) 8000 for 1× buffer). A reaction mix of 15 μL was incubated at 50 °C for 1 h (two-part assembly) or 2.5 h (multi-part assembly). The enzymes used were from New England Biolabs, Inc. Six sets of Gibson assembly (GA1–GA6) were tested.

2.4. Transformation

Lab-made *Escherichia coli* strain XL10-Gold (Integrated Science & Technology, Inc.) with a transformation efficiency of 7×10^6 CFUs per microgram pUC19 DNA (CFUs·μg^{-1}) was used for chemical transformation. A volume of 1–6 μL of the assembly products was added to 50 μL of competent cells, recovered in 950 μL of SOC medium for 1 h and plated. The purple colonies were considered correct assemblies if the gene (iGEM Part: BBa_K1033906) encoding the purple chromogenic protein was used for assemblies. Assemblies that did not contain marker genes were verified by colony PCR using 10 randomly picked colonies. The AE was represented by the number of correct CFUs per microgram DNA.

2.5. qPCR

Before quantification, the residual linear fragments in the assembly mix were digested by 0.5 μL (5 U) T5 exonuclease at 37 °C for 30 min, followed by incubation at 85 °C for 15 min to inactivate the exonuclease. The insert or vector fragment in 1.5 μL assembly product was quantified in a 20 μL reaction system (Supplementary data, Table S1) prepared by SYBR® Premix Dimer-Eraser™ (TaKaRa). For the reference, a parallel reaction containing an identical amount of every DNA fragment used for assembly except the enzyme was prepared. A mixture of T5 digested linear vector and insert was used as the template in the control assay. Quantification was performed using a Roche LightCycler® 96 system with SYBR Green I detection. The reaction started with a preincubation at 95 °C for 30 s, followed by 40 cycles of 5 s at 95 °C, 30 s at 56 °C, and 30 s at 72 °C per cycle. Melting curve analysis was performed by raising the temperature from 60 to 95 °C at a rate of 0.1 °C·s^{-1} with five signal acquisitions per degree centigrade. Each sample was measured in duplicate.

Calculation of the AE was derived from the ΔC_t (i.e., difference in the threshold numbers for the target and the reference) method [22]. The threshold number of DNA molecules for the target (X_t) and reference (R_t) can be described as follows:

$$X_t = X_0 \times E_X^{C_{t,X}} \quad (1)$$

$$R_t = R_0 \times E_R^{C_{t,R}} \quad (2)$$

where X_0 and R_0 represent the amount of the target and the reference DNA, respectively. E_X is the amplification efficiency for the target sequence and E_R refers to that of the reference. $C_{t,X}$ is the threshold cycle for the target and $C_{t,R}$ is the threshold cycle for the reference. Accordingly, the ratio of X_0 to R_0 can be expressed as follows:

$$\frac{X_0}{R_0} = \frac{X_t}{R_t} \times \frac{E_R^{C_{t,R}}}{E_X^{C_{t,X}}} = R \quad (3)$$

In this case, quantification targets the same sequence for both the ligated fragment and its linear reference, and measurements are performed in the same PCR run. Thus, the ratio of X_t to R_t can be assumed to be a constant that approximates to one. In addition, the ΔC_t approach assumes equal amplification efficiency for both the target and the reference. In this case, the assumption is more valid because the target and the reference are identical. Ideally, the PCR product replicates once in every cycle, which means an amplification efficiency of 2. Therefore, Eq. (3) can be transformed into the following:

$$R = 2^{-\Delta C_t} \quad (4)$$

where R represents the ratio of the amount of the ligated fragment to its initial amount used for assembly, while ΔC_t is the difference in the threshold cycles between the target and the reference ($C_{t,X} - C_{t,R}$)—in this case, the difference in C_t values between the ligated and the initial linear fragments.

2.6. Data analysis

Pearson's two-tailed correlation was used to investigate the relationship between the AEs derived by the CFU- and qPCR-based measurements. When necessary, the differences between the AEs were analyzed by the two-tailed Student's *t*-test. The relationships between the efficiency of the Gibson assembly and the properties of the overlap sequence were modeled using boosted trees analysis [23] in R (2.7.2) with the gbmplus (1.5-17) package.

3. Results

3.1. Efficiencies of restriction ligation generated by different measurements

As shown by the CFU-based measurement, the seven sets of restriction cloning had different efficiencies ranging from 5×10^2 to 1.6×10^4 CFUs·μg^{-1} (Fig. 2(a)). RL6 and RL7, each of which generated a 2.1 kilobase (kb) DNA molecule, had an average efficiency that was 11 times higher than those of the other cloning

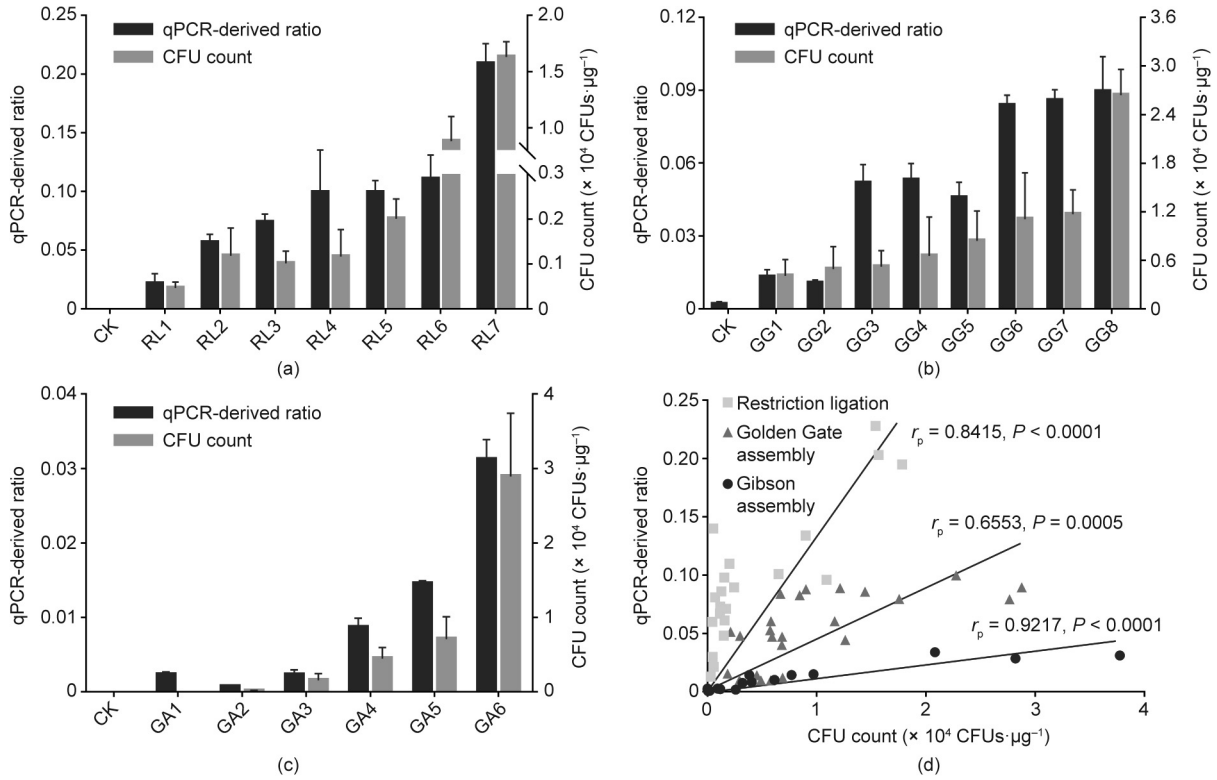

Fig. 2. The AEs of (a) restriction ligation, (b) Golden Gate assembly, and (c) Gibson assembly represented by the correct CFU count and the qPCR-derived ratio; (d) the correlations between these two measurements. CK indicates the level of residual linear fragments after digestion by T5 exonuclease. The qPCR-derived ratio and CFU count results of each sample are the mean values, and the error bars represent the standard deviations (n = 3). r_p: Pearson correlation coefficient.

sets, of which the sizes of the ligation products were between 5.5 to 8.7 kb. Similar differences in the AEs were revealed by the qPCR-based measurement, for which the qPCR-derived ratios were between 0.021 and 0.209 (Fig. 2(a)). The difference between the highest (RL7) and lowest (RL1) efficiencies decreased from 35-fold, as indicated by the CFU count, to 10-fold, as reflected by the qPCR-derived ratio. The larger difference for the CFU-based measurement might be attributed to a compromised transformation efficiency, as the size of the DNA molecule increased up to four-fold from RL7 to RL1, which could reduce the transformant yield. Thus, the AEs represented by the CFU count for large molecules such as RL1 were likely to be lower than the actual level of successfully ligated fragments as indicated by the qPCR-derived ratio. Nevertheless, a positive relationship was found between the efficiencies generated by the two measurements (Fig. 2(d)).

3.2. Efficiencies of Golden Gate assembly generated by different measurements

Both the CFU- and qPCR-based measurements detected an increase in the AE for groups from GG1 to GG8 (Fig. 2(b)). The CFU count increased 6.5-fold from 4×10^3 to 2.6×10^4 CFUs·μg^{-1}, and the qPCR-derived ratio increased 5.5-fold from 0.013 to 0.072. As shown by the qPCR-based measurement, the AE increased significantly from GG3 to GG5 by an average of four-fold compared with that of GG1 or GG2. The efficiency continued to increase sharply for GG6 and remained stable for GG7 and GG8. In comparison, a significant increase in the AE was only detected for GG7 and GG8 by the CFU-based method, while an increase in the AE for GG1 to GG6 were not particularly obvious. The non-significant results could be ascribed to the large deviation in the CFU-based measurement, which stemmed from the transformation process, even though the DNA molecules were of equal size and the transformation conditions were strictly controlled. Despite these discrepancies, the results derived from the qPCR-based measurement correlated positively with those generated by the CFU-based method (Fig. 2(d)).

3.3. Efficiencies of Gibson assembly derived from different measurements

According to the CFU-based measurement, the six sets (GA1–GA6) of DNA fragments had different AEs ranging from 0 to 2.9×10^4 CFUs·μg^{-1} (Fig. 2(c)). The efficiency increased as the size of the assembly product decreased from 9.7 kb (GA1) to 3.7 kb (GA6). The qPCR-derived ratios for the six assemblies followed the same order as that shown by the number of CFUs, and a positive relationship was found between these two measurements (Fig. 2(d)). It was notable that no colony was formed for GA1 that would produce the largest DNA molecule among all the assembly groups, whereas a value of 0.002, which was significantly higher than that of the T5 digested control, was detected for GA1 by the qPCR-based method. This result could also be attributed to the size effect of the DNA molecule on the transformation efficiency.

This method was also applied to measure the efficiency of multi-part DNA assembly. As expected, the efficiency determined by qPCR dropped significantly by 48% from 0.028 to 0.015 when the number of assembled fragments increased from two (M2) to three (M3) (Fig. 3). It continued to decrease by 55% from 0.015 to 0.007 when four fragments (M4) were assembled. Changes in the CFU count agreed with the qPCR-based measurement. Compared with the assembly of two fragments, a 54% decrease in the CFU count was detected when three fragments were assembled. The CFU count continued to decrease by 84% from 1.6×10^4 to 3×10^3 CFUs·μg^{-1} when the number of fragments increased to

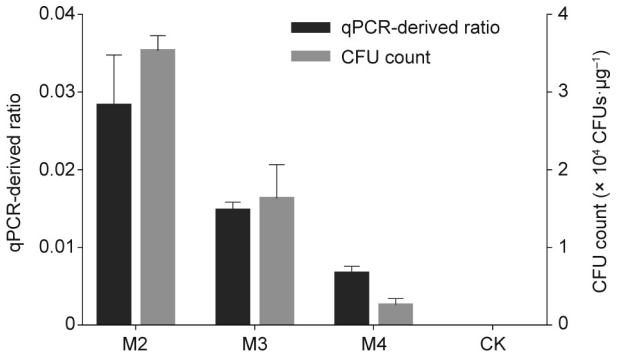

Fig. 3. The AEs for two to four (M2 to M4) DNA fragments determined by both the CFU- and the qPCR-based measurements. The qPCR-derived ratio and CFU count results of each sample are the mean values, and the error bars represent the standard deviations (n = 3).

four. In all, both measurements evidenced a significant decrease in the AE when multiple DNA fragments were assembled.

3.4. Influence of secondary structures in the overlap region on Gibson assembly

To study the effects of secondary structures in the overlap region on Gibson assembly, 11 short sequences (20 bp) capable of forming secondary structures were designed. The stability of the secondary structures was quantified by the free energy difference (ΔG) [24,25], where a lower ΔG indicates higher stability for a duplexed structure [26]. Of the 11 sequences, the hairpin ΔG ranged from −1.2 to −9.2 kcal·mol^{-1} (1 kcal = 4186 J), while the self-dimer ΔG ranged from −3.3 to −11.6 kcal·mol^{-1}. Another two sequences containing no secondary structures (ΔG = 0) were also designed. These short sequences were introduced into the ends of a linearized pUC19 vector and a 1 kb fragment encoding the purple protein (Supplementary data, Table S2) as shown in Fig. 1(d), generating a total of 12 fragment pairs. One pair of fragments did not contain secondary structures in both the overlap regions (OL1, as control), nine pairs had secondary structures embedded in only one overlap region (OL2–OL10), and the last two pairs could form secondary structures in both the overlap regions (OL11 and OL12). The assembly mixture from the same reaction tube was used for the measurement of AE by both the CFU- and the qPCR-based methods.

As shown by both the qPCR-derived ratio and the CFU count, the AE basically increased in the samples from OL1 to OL6, but decreased sharply in the samples from OL7 to OL12 (Fig. 4). The CFU count increased from 4 × 10^3 CFUs·μg^{-1} in OL1 with no terminal secondary structures, to the highest count of 1.1 × 10^4 CFUs·μg^{-1} in OL6 with a hairpin ΔG of −5.3 kcal·mol^{-1}. A dramatic decrease of 89% in the CFU count occurred when the hairpin ΔG of the overlap region decreased to −6.9 kcal·mol^{-1} (OL7). The fragment pair (OL10) with the lowest hairpin ΔG in the overlap region had the lowest CFU count of 2 × 10^2 CFUs·μg^{-1}, which was only 5% of that of OL1. Compared with OL1, a higher AE was also discovered in OL6 by the qPCR-based measurement. However, according to the qPCR-derived ratio, the highest AE (0.156) was achieved by OL3. The ratio decreased dramatically by 62% from 0.136 (OL6) to 0.052 (OL7) as the secondary structures became more stable for the samples from OL6 to OL7, and reached a lower level of 0.037 in OL10. Compared with OL1 (0.032), the formation of secondary structures in both overlap regions (OL11 and OL12) led to a significant decrease in the AE, although the ΔG (−4.5 to −6.4 kcal·mol^{-1}) indicated only medium structural stabilities among all the overlaps.

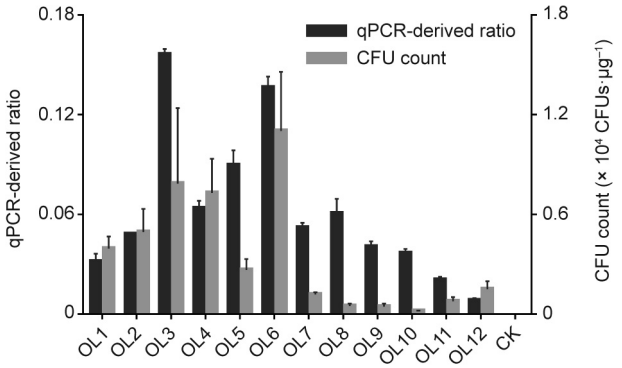

Fig. 4. The AEs for DNA fragments containing secondary structures in the overlap regions determined by both the CFU- and the qPCR-based measurements. The qPCR-derived ratio and CFU count results of each sample are the mean values, and the error bars represent the standard deviations (n = 3).

The terminal secondary structure characterized by ΔG is not the only factor that affects Gibson assembly. In fact, the overall properties of the overlap sequences can have stronger effects on the AE than the stability of the secondary structures (Fig. 5). According to the model generated from the results of the 12 sets of Gibson assembly, the primer melting temperature (T_m) and the guanine and cytosine (GC) content of the overlap sequence were the primary determinants that accounted for a total of 56% of the differences in the AE. The efficiency increased sharply when the overall T_m went above 60 °C (Fig. 5(b)), and increased with an increase of GC content ranging from 50% to 70% (Fig. 5(c)). The hairpin structures that formed within the overlap region showed stronger effects than those of the dimer structures on the AE. To be specific, the hairpin T_m, hairpin ΔG, and GC content of the hairpin structure explained 31% of the total variance in the AE, while the total relative influence of the T_m, ΔG, and GC content of the dimer structure was only 13%. It is likely that the AE would decrease dramatically once the hairpin ΔG decreased to below −4 kcal·mol^{-1} (Fig. 5(d)), and a negative relationship was discovered between the AE and the hairpin T_m, which ranged from 22 to 30 °C (Fig. 5(e)).

4. Discussion

Using a modified qPCR assay, this study developed a transformation-independent method for rapid determination of the DNA AE. As demonstrated by commonly used DNA assembly techniques including restriction ligation, Golden Gate assembly, and Gibson assembly, the qPCR-based measurement produced an estimation of the AE that was comparable to that of the conventional CFU-based method, and thus facilitated investigation of the determinants of DNA assembly.

The qPCR-based measurement ruled out the interferences from transformation on the AE, and thus is more reliable than the CFU-based measurement. In this study, the mean standard deviations of the AEs represented by the qPCR-derived ratios were 14% for restriction ligation, 13% for Golden Gate assembly, and 17% for Gibson assembly (Fig. 2). In comparison, even though the same batch of competent cells were used for the measurement of the AE and strict quality controls were established during transformation, the mean standard deviations for the AEs estimated by the CFU count were as high as 31%–48%, which were 2.2–3.7 times higher than the corresponding qPCR-derived ratios. The low reproducibility of the CFU-based measurement might hide the difference between samples such as GG6 and GG5, for which a sharp increase in the AEs of GG6 to that of GG5 should be detected, as shown by the qPCR-derived ratio. Another type of deviation that could be avoided

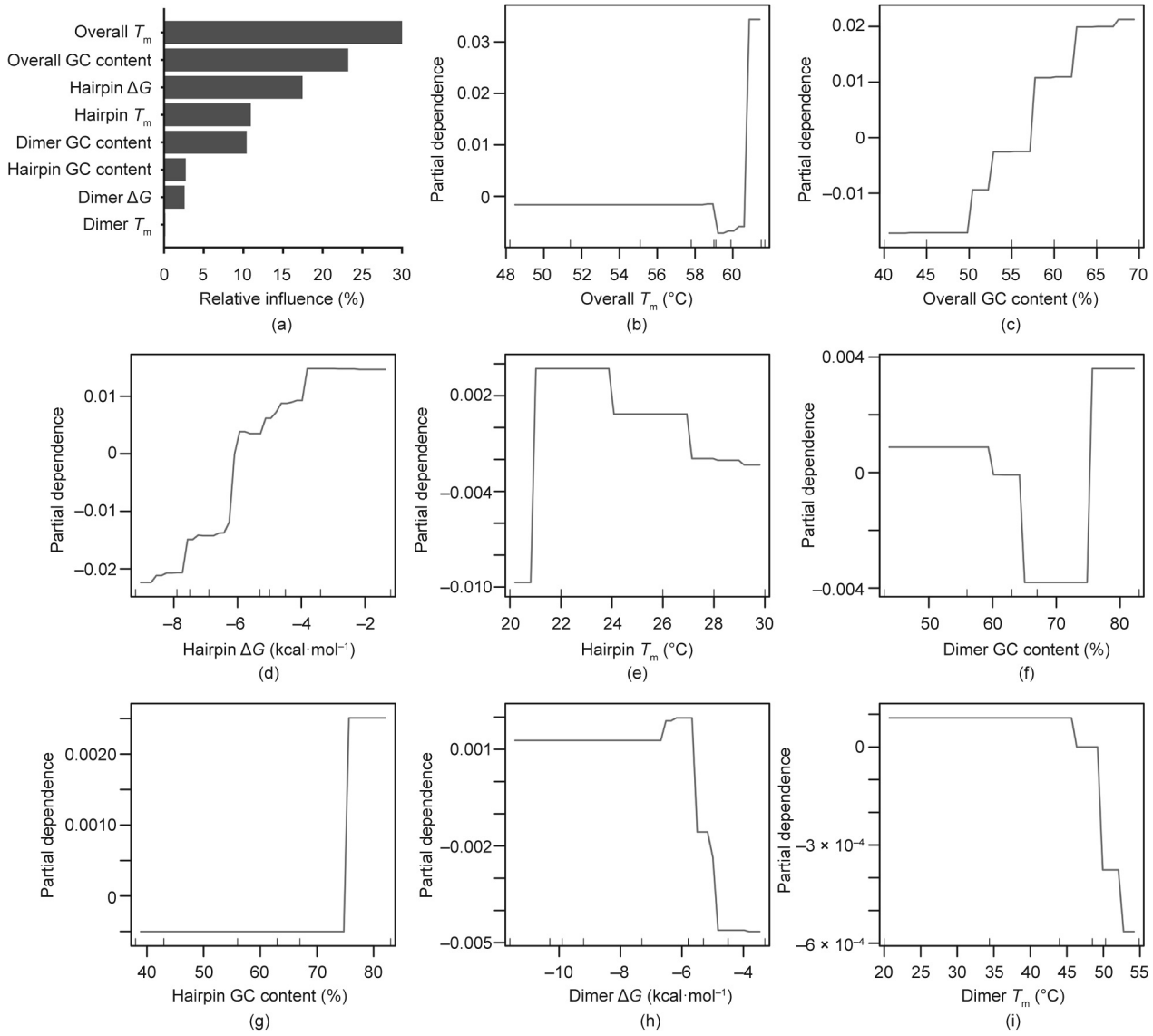

Fig. 5. (a) The relative influences of the terminal secondary structures and the overall properties of the overlap sequence on the AE and the partial dependences of the AE on (b) the overall T_m, (c) the overall GC content, (d) the hairpin ΔG, (e) the hairpin T_m, (f) the dimer GC content, (g) the hairpin GC content, (h) the dimer ΔG, and (i) the dimer T_m.

by the qPCR-based assay stems from the size preference. The transformation of a DNA molecule into cells becomes more difficult when its size increases [27,28], leading to decreased transformation efficiency and therefore an underestimated AE denoted by the CFU count. This could explain the contradiction between the failed and successful assembly of GA1, as indicated by the CFU- and qPCR-based measurements, respectively (Fig. 2(c)). Since the 10 kb DNA molecule produced by GA1 might be too large for cells to recover those rarely assembled DNA fragments via transformation, the AE could only be detected by the transformation-independent measurement. The reduced difference between RL7 and RL1 in the AE derived from the qPCR assay compared with that represented by the CFU count can also be partly attributed to the size bias. Thus, the qPCR-based assay offers higher sensitivity and reliability than the conventional CFU-based measurement.

An impartial comparison between different sets of assembly should minimize the random deviations that stem from the measurement of the AE. This could be laborious for CFU-based measurement, which depends on transformation. First, the competent cells must be of the same genotype among all measurements in order to exclude variations in DNA propagation and stability. More importantly, the transformation efficiency must be kept constant. To ensure this, the competent cells should come from the same batch, and transformation must follow identical steps—such as the exact same time spent in thawing the competent cells—as variations in these steps can result in dramatic differences in the transformation efficiency of up to 10 000-fold [29]. Even with strict controls, the AE can still be affected by *in vivo* manipulation of the assembly products, such as the recombination of the DNA molecules containing repetitive regions [30,31]. In comparison, the qPCR-based assay allows easy control of the measuring conditions by maintaining the basic rules for qPCR. It even simplifies the workflow by using the identical sequence for the target and the reference. Thus, only a single pair of primers are required for quantification, and the difference in the amplification efficiencies between the target and the reference is minimized. Note that the same amount of fragments should be used for the quantification reference as for the assembly. In addition, the remaining linear fragments in the assembly mix must be digested completely, and the exonuclease should be inactivated to avoid digestion of the qPCR primers in the subsequent quantification assay. When following these rules, qPCR-based measurement enables rapid

measurement of the AE. It usually takes 30 min for the digestion of the residue fragments, 15 min to inactivate the T5 exonuclease, and 2 h to prepare and run the qPCR. In total, this is a quick assay of less than 3 h, which takes only a quarter of the time required for a CFU-based measurement.

The development of qPCR-based measurement is not intended to replace transformation, but to provide a reliable and rapid alternative for determining the AE. Therefore, this method is particularly useful for the development of DNA assembly techniques and for the debugging of inefficient assembly, when there is a high demand for measurement throughput and data reliability. The effects of various factors (e.g., the length of the overlap region, the ratio between fragments, the amount of DNA molecules, and the assembly time) on the AE can be tested simultaneously in a single qPCR assay within 3 h, allowing the optimal protocol for this particular assembly to be determined. The *in vitro* measurement solely targets the AE, rather than focusing on a combined efficiency of both transformation and assembly, which enables the dissection of the multiple causes of low transformant yield. For example, if the qPCR-derived ratio does not suggest a low AE as indicated by the CFU count, then there could be something wrong with the transformation process. If an improved transformation efficiency does not yield more transformants, it might suggest product toxicities of the inserted genes [32,33], which could be solved by switching to a more tightly controlled vector that prohibits leaky expressions of the toxic genes. Moreover, the fidelity of DNA assembly can be determined by a modified quantification strategy. Instead of targeting the sequence within the ligated fragments, the region that spans the joint of the adjacent fragments is targeted for quantification, which should be detectable only when all the fragments are ligated in the desired order. Under this assumption, the ratio of the amount of the joint sequence to the amount of the ligated fragment is used to indicate the fidelity of the assembly (Fig. S1(a)). As exemplified by five sets of Gibson assembly with different correct rates, differences in the assembly fidelity can be identified by this alternative qPCR assay (Supplementary data, Fig. S1(b)).

The qPCR-based measurement uncovered the influence of terminal secondary structures on the efficiency of Gibson assembly. In general, the AE decreased significantly when stable hairpin structures were formed within the overlap regions. However, it should be noted that the overall properties of the entire overlap sequence, such as the base composition and the thermal stability, might have stronger effects than the local secondary structures in mediating DNA assembly. To be specific, a GC-rich overlap sequence with high T_m is likely to increase the chance of homologous ends annealing, resulting in a high AE. In this study, the T_m values of the overlap regions ranged from 48 to 62 °C and the overall GC contents ranged from 40% to 70%, which covered the general conditions used in DNA assembly [9,12,13,15]. Therefore, in order to design an efficient Gibson assembly, more GC should be included in the overlap sequence, as long as the GC content and the T_m are within the normal range. The presence of weak secondary structures within the overlap sequence is tolerable, but stable hairpin structures such as the GC-rich palindromic restriction site should be avoided. These rules are relevant for assembly techniques such as SLIC, TPA, IVA, and others that depend on homologous annealing of the terminal sequences.

5. Conclusions

In sum, this study established an alternative method for the determination of the DNA AE using a modified qPCR assay. The results were comparable to the commonly used CFU count in quantifying the AE. The proposed method outperformed the CFU-based measurement by reducing the measurement bias and the random deviations that stem from the transformation process. This rapid workflow can facilitate the development of DNA assembly techniques and the diagnosis of inefficient assemblies.

Acknowledgements

This work was supported by the National Key R&D Program of China (2017YFD0201400), the National Natural Science Foundation of China (21676026), and the Fundamental Research Funds for the Central Universities.

Compliance with ethics guidelines

Xiaoyan Ma, Xinxin Liang, and Yi-Xin Huo declare that they have no conflicts of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2019.06.002.

References

[1] Andrianantoandro E, Basu S, Karig DK, Weiss R. Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2006;2(1):2006.0028.
[2] Choffnes ER, Relman DA, Pray L. The science and applications of synthetic and systems biology: workshop summary. Report. Washington, DC: National Academies Press; 2011.
[3] Ramon A, Smith HO. Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. Biotechnol Lett 2011;33(3):549–55.
[4] Wingler LM, Cornish VW. Reiterative recombination for the *in vivo* assembly of libraries of multigene pathways. Proc Natl Acad Sci USA 2011;108(37):15135–40.
[5] Dietrich JA, McKee AE, Keasling JD. High-throughput metabolic engineering: advances in small-molecule screening and selection. Annu Rev Biochem 2010;79(1):563–90.
[6] Purnick PEM, Weiss R. The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 2009;10(6):410–22.
[7] Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, et al. Genetic circuit design automation. Science 2016;352(6281):aac7341.
[8] Chao R, Yuan Y, Zhao H. Recent advances in DNA assembly technologies. FEMS Yeast Res 2015;15(1):1–9.
[9] Quan J, Tian J. Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 2009;4(7):e6441.
[10] Li MZ, Elledge SJ. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. Nat Methods 2007;4(3):251–6.
[11] Cobb RE, Ning JC, Zhao H. DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. J Ind Microbiol Biotechnol 2014;41(2):469–77.
[12] García-Nafría J, Watson JF, Greger IH. IVA cloning: a single-tube universal cloning system exploiting bacterial *in vivo* assembly. Sci Rep 2016;6(1):27459.
[13] Liang J, Liu Z, Low XZ, Ang EL, Zhao H. Twin-primer non-enzymatic DNA assembly: an efficient and accurate multi-part DNA assembly method. Nucleic Acids Res 2017;45(11):e94.
[14] Jin P, Ding W, Du G, Chen J, Kang Z. DATEL: a scarless and sequence-independent DNA assembly method using thermostable exonucleases and ligase. ACS Synth Biol 2016;5(9):1028–32.
[15] Zhang Y, Werling U, Edelmann W. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res 2012;40(8):e55.
[16] Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. PLoS One 2008;3(11):e3647.
[17] Fu C, Donovan WP, Shikapwashya-Hasser O, Ye X, Cole RH. Hot fusion: an efficient method to clone multiple DNA fragments as well as inverted repeats without ligase. PLoS One 2014;9(12):e115318.
[18] Yoshida N, Sato M. Plasmid uptake by bacteria: a comparison of methods and efficiencies. Appl Microbiol Biotechnol 2009;83(5):791–8.
[19] Aune TEV, Aachmann FL. Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. Appl Microbiol Biotechnol 2010;85(5):1301–13.
[20] Sleight SC, Bartley BA, Lieviant JA, Sauro HM. In-Fusion BioBrick assembly and re-engineering. Nucleic Acids Res 2010;38(8):2624–36.
[21] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6(5):343–5.

[22] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 2001;25(4):402–8.
[23] De'ath G. Boosted trees for ecological modeling and prediction. Ecology 2007;88(1):243–51.
[24] SantaLucia J Jr, Hicks D. The thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct 2004;33(1):415–40.
[25] Lomzov AA, Vorobjev YN, Pyshnyi DV. Evaluation of the Gibbs free energy changes and melting temperatures of DNA/DNA duplexes using hybridization enthalpy calculated by molecular dynamics simulation. J Phys Chem B 2015;119(49):15221–34.
[26] Gao Y, Wolf LK, Georgiadis RM. Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. Nucleic Acids Res 2006;34(11):3370–7.
[27] Lu YP, Zhang C, Lv FX, Bie XM, Lu ZX. Study on the electro-transformation conditions of improving transformation efficiency for *Bacillus subtilis*. Lett Appl Microbiol 2012;55(1):9–14.
[28] Tee KL, Grinham J, Othusitse AM, González-Villanueva M, Johnson AO, Wong TS. An efficient transformation method for the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. Biotechnol J 2017;12(11):1700081.
[29] Russo R, Panangala VS, Wood RR, Klesius PH. Chemical and electroporated transformation of *Edwardsiella ictaluri* using three different plasmids. FEMS Microbiol Lett 2009;298(1):105–10.
[30] Bzymek M, Lovett ST. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc Natl Acad Sci USA 2001;98(15):8319–25.
[31] Lovett ST. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. Mol Microbiol 2004;52(5):1243–53.
[32] Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. Biotechnol J 2012;7(5):620–34.
[33] Pelicic V, Reyrat JM, Gicquel B. Expression of the *Bacillus subtilis sacB* gene confers sucrose sensitivity on mycobacteria. J Bacteriol 1996;178(4):1197–9.