Journal of Experimental Microbiology and Immunology (JEMI)

Increasing Overhang GC-Content Increases Sticky-End Ligation Efficiency

Tina Gao, Saki Konomura, Chris May, Charles Nieh
Department of Microbiology and Immunology, University of British Columbia

The ligation efficiency of sticky-ends differs between overhangs generated by different restriction enzymes. Differences between the composition of these overhangs include length and GC-content, thus indicating that these properties may play a role in ligation efficiency. Although ligases are a common and widely used tool in molecular biology, a method to accurately quantify ligation efficiency has yet to be fully developed and standardized. In this study, ligation efficiency of synthetic oligonucleotides with four-nucleotide overhangs containing varying GC-content was quantitatively investigated using qPCR. 2 sets of synthetic, double-stranded oligonucleotides with complementary overhangs were created. The complementary oligonucleotides were mixed with ligase, and the efficiency of ligation was quantified by measuring an increase in fluorescence intensity due to SYBR Green intercalation between amplified polymerase chain reaction products, where ligated oligonucleotide served as template. Using a control oligonucleotide, semi-quantitative qPCR analysis of ligations involving oligonucleotides with 0%, 50%, and 100% GC-content in their overhangs was developed. The ligation of oligonucleotides with overhangs containing 100% GC-content resulted in ~25% more ligated product than oligonucleotides with overhangs containing 0% GC-content. These findings suggest that oligonucleotides possessing overhangs with higher GC-content are more efficiently ligated.

Restriction enzyme digests and ligases are widely used in research and diagnostic tools in molecular biology (1). Restriction enzymes produce a double-stranded cut in DNA by recognizing specific nucleotide sequences (2). Depending on the restriction enzyme, the resulting termini of the cleaved DNA molecules are described as "blunt-ended" or "sticky-ended" (2). Sticky-ends have overhangs which are unpaired nucleotides that extend from the end of the double-stranded DNA molecule (2). Blunt-ended DNA molecules have no unpaired nucleotides. T4 DNA ligase, the most commonly used DNA ligase, catalyzes sticky-end ligations of juxtaposed DNA strands more efficiently than blunt-end ligations (1, 3). Different restriction enzymes generate sticky-ends with varying length and base composition in the overhangs, which leads to differences in ligation efficiency in downstream applications (4, 5). This indicates that overhang length and/or base composition may influence ligation efficiency.

In this study, we sought to develop a method to measure DNA ligation efficiency. Gel electrophoresis with DNA-specific staining or radioactive labeling is often used to determine DNA ligase activity by visualization of the band intensity of ligated products (7, 8). However, it is difficult to quantify the intensity of bands, which makes gel electrophoresis an insensitive method for determining ligation efficiency (4, 6). Qiu *et al.* demonstrated that real-time polymerase chain reaction (qPCR) using non-specific fluorescent dyes could be used to quantitatively measure ligation efficiency (5). In Qiu *et al.*'s method (5), dyes that are present in the qPCR reaction, such as SYBR Green, are able to intercalate between double-stranded DNA. These dyes only fluoresce once they incorporate into DNA duplexes, and the level of intercalation increases as DNA products accumulate, thus leading to an increase in fluorescence that can be detected by the qPCR machine in real time. By using primers that amplify the ligated region, this method can determine how much of the DNA in a ligation reaction has been successfully ligated, because unligated products will not amplify. Higher ligation efficiency results in more ligated product, and leads to a more rapid amplification and increase in fluorescence intensity (5).

There exist limitations to PCR-dependent cloning such as primer design and the introduction of random mutations (3). In addition, there are challenges with amplifying large or high GC-content DNA fragments (3). These challenges exist primarily due to intrinsic properties of DNA polymerase and DNA itself, namely that polymerases are error-prone and high-GC DNA is much more stable than low-GC DNA which makes it difficult to dissociate to support polymerase-dependent amplification. Thus, the classical digestion-ligation method is still widely used (3). Improving ligation efficiency will reduce the amount of ligase required as well as the time needed per ligation, thus saving reagent costs and time.

Higher GC-content leads to increased stability of DNA duplexes, however the reason for this has not been experimentally determined. It is thought that this is because G:C base pairs form three hydrogen bonds while A:T base pairs only form two. Therefore, DNA with higher GC-content is more stable and has a higher melting temperature (11). The stronger base pairing between guanine and cytosine may result in a lower rate of dissociation between sticky-ends with higher GC-content compared to sticky-ends with lower GC-content during ligation. DNA fragments with higher overhang GC-content would thus remain associated more often, leading to a higher chance of ligation by T4 ligase.

In this study, we further develop the qPCR method to quantify ligation efficiency established by Qiu *et al.* (5). Additionally, we examine the effects of differing GC-content on ligation efficiency, via the use of synthetic

oligonucleotides with 100% (100-GCs), 50% (50-GCs), and 0% (0-GCs) overhang GC-content. We hypothesize that ligation efficiency, will increase when overhangs possess higher GC-content. The results presented herein are consistent with this model, showing that ligation efficiency increases when overhangs of the same length have higher GC-content.

MATERIALS AND METHODS

Synthetic oligonucleotides and primers. Oligonucleotides and primers were obtained from Qiu *et. al* (5). Additional oligonucleotides with varying GC-content in their overhangs, and InternalF & InternalR primers (see Fig. 2 for primer annealing sites on synthetic oligonucleotides) were purchased from Integrated DNA Technologies. Refer to Table 1 for the sequences used. It should be noted that 0-GCs have 100% AT-content, and 50-GCs have 50% AT-content (Table 1). A brief flow-through of our methods is depicted in Figure 1.

Qiu *et al.* (2014)	Gao *et al.* (2015)	DNA Sequence
ETA2	O100F	5'ATCGCATGCACTACTGTCTCTC ACTGATCTGTCCTCGTAACTGAA TAT**CCGC** 3'
ETA6	O50F	5'ATCGCATGCACTACTGTCTCTC ACTGATCTGTCCTCGTAACTGAA TAT**AGT*** 3'
ETA4	O0F	5'ATCGCATGCACTACTGTCTCTC ACTGATCTGTCCTCGTAACTGAA TAT**AATA** 3'
ETA1	AnnealF	5'ATATTCAGTTACGAGGACAGA TCAGTGAGAGACAGTAGTGCAT GCGAT 3'
-	O100R	5'CCAGTTACGAGTATATCCACG CAACAGCTTCTCATCCGTCGAG GCCAT**GCGG** 3'
-	O50R	5'CCAGTTACGAGTATATCCACG CAACAGCTTCTCATCCGTCGAG GCCAT**GACT** 3'
O52	O0R	5'CCAGTTACGAGTATATCCACG CAACAGCTTCTCATCCGTCGAG GCCAT**TATT** 3'
O48	AnnealR	5'ATGGCCTCGACGGATGAGAAG CTGTTGCGTGGATATACTCGTAA CTGG 3'
Main Primer	PrimerF	5'ATCGCATGCACTACTGTCTC 3'
P2	PrimerR	5'CCAGTTACGAGTATATCCACG C 3'
-	InternalF	5'ATATTCAGTTACGAGGACAGA TCA 3'
-	InternalR	5'ATGGCCTCGACGGATGAG 3'

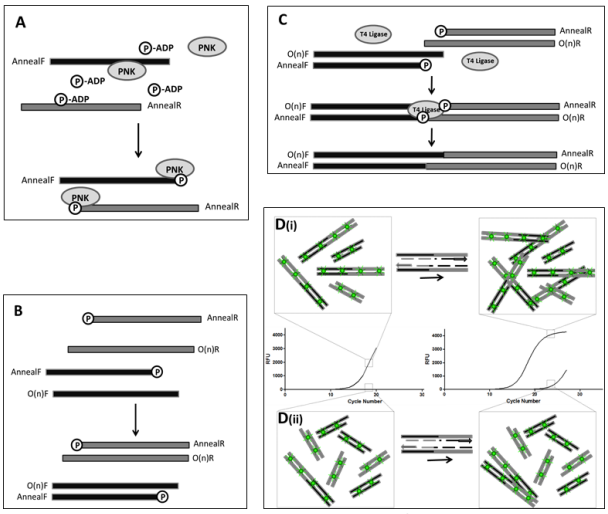

Fig 1 Diagram representing the method of determining ligation efficiency via qPCR. (a) 48-mer single-stranded oligonucleotides are ordered (IDT) and 5' ends are phosphorylated; PNK = Polynucleotide Kinase; P = Phosphate; ADP = Adenosine diphosphate. (b) Phosphorylated 48-mers are annealed to their respective 52-mer single-stranded oligonucleotide partners, forming double-stranded DNA oligonucleotides with 4-base overhangs and 5'-phosphorylated sticky-ends; O(n)R/O(n)F represents single-stranded oligonucleotides with overhangs of differing GC-content. (c) Double-stranded oligonucleotides with cognate overhangs are ligated under pre-set conditions. (d) ligation reactions are mixed with SYBR Green ("glowing" dots) which intercalates into double-stranded DNA and fluoresces. DNA is then amplified by qPCR, and fluorescence is read after each amplification step where (i) represents a ligation reaction where 75% of DNA is successfully ligated product, leading to a more rapid amplification and increase in fluorescence than (ii) which represents a ligation reaction where only 25% of DNA is successfully ligated product; RFU = Relative Fluorescence Units, units used to measure the amount of fluorescence in a sample.

Phosphorylation of synthetic oligonucleotides. T4 polynucleotide kinase (New England Biolabs, Cat no. M0201S) was used to phosphorylate the 5' end of oligonucleotides (AnnealF and AnnealR). Each 50µl reaction consisted of 1.5µl of 200µM DNA template, 1µl of kinase, 10µl of 5X T4 ligase buffer, and 37.5µl of dH_2O. The reaction was incubated at 37°C for 30 min, followed by kinase inactivation at 65°C for 20 min.

Annealing of synthetic oligonucleotides. Oligonucleotides were annealed as described in (9). 1.5µl of 20µM forward or reverse DNA oligonucleotides as shown in Table 1 (O100F, O50F, O0F / O100R, O50R, O0R) was annealed to 5µl of 6µM of phosphorylated AnnealF or AnnealR DNA oligonucleotide, respectively, in a 20µl reaction with 1.5µl of 10X annealing buffer (10mM Tris, 1mM EDTA, 50mM NaCl) and 12 µl dH_2O. Annealing was performed at 95°C for 4.5 min, followed by cooling at room temperature for 10 min.

Ligation of synthetic control oligonucleotide. Annealed forward and reverse oligonucleotides with 0% GC-content overhangs (O0F/AnnealF and O0R/AnnealR) were ligated in a 40µl reaction using 2µl T4 DNA ligase (Life Technologies, Cat no. 15224-025), 4µl 5X T4 DNA ligase buffer (Life Technologies, Cat no. 15224-025), 10µl of each annealed oligonucleotide, and 14µl of dH_2O. The ligation was performed at room temperature for 10 min, then the ligase was inactivated at 65°C for 10 min.

Agarose gel electrophoresis. Ligated products and qPCR products were run on a 2% agarose gel. Ultrapure™ agarose (Life Technologies, Cat no. 10821-015) was dissolved in a solution of 1.5X TBE buffer (89 mM Tris base, 89 mM boric acid, and 2mM

EDTA) and SYBR® Safe DNA Gel Stain in 0.5X TBE (Life Technologies, Cat no. S33100) in a 1:1 ratio. Samples were prepared with 6X DNA loading dye (Thermo Scientific, Cat no. R0611). 100 bp DNA ladder (Life Technologies, Cat no. 15628-050) was loaded with samples. 1X TBE was used as running buffer. Gels were run at 120V for 60 min, and imaged using an Alpha Innotech AlphaImager.

Gel extraction. Ligated oligonucleotides run on a 2% agarose gel were extracted and purified using PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Life Technologies, Cat no. K2200-01) as per the manufacturer's instructions.

qPCR for generating standard curve. Control oligonucleotides were run in qPCR reactions using CFX Connect™ Real-Time PCR Detection System. Each 20µl reaction was set up using 10µl SsoAdvanced Universal SYBR Green Supermix (BioRad, Cat. no. 1725265), 6µl dH_2O, 2µl of control oligonucleotide, and 1µl each of PrimerF and PrimerR (see Table 1). qPCR conditions were as follows: 3 min at 95°C, 35 cycles of 95°C for 1 minute and 45 °C for 7 seconds. Additionally, melt curve analysis was performed from 65°C to 95°C, with 0.5°C increments every 5 seconds.

Ligation of oligonucleotides with varying GC-content. Forward and reverse oligonucleotides with varying GC-content in overhangs were annealed and ligated with their complementary partner (Fig. 1B, 1C) in a 20µl reaction using 0.25µl T4 DNA ligase (Life Technologies, Cat no. 15224-025), 4µl 5X T4 DNA ligase buffer (Life Technologies, Cat no. 15224-025), 5µl of each annealed oligonucleotide, and 5.75µl of dH_2O. The ligation was performed at room temperature for 5 min or 20 hours, then the ligase was inactivated at 65°C for 10 min.

qPCR for oligonucleotides with varying GC-content. Oligonucleotides with varying GC overhang content were run in qPCR reactions using CFX Connect™ Real-Time PCR Detection System. Each 20µl reaction was set up using 10µl SsoAdvanced Universal SYBR Green Supermix (BioRad, Cat. no. 1725265), 6µl dH_2O, 2µl of template oligonucleotides, and 1µl each of PrimerF and PrimerR. Reactions with unligated oligonucleotides (serving as internal controls to ligated oligonucleotides) were set up under the same conditions, however the primers used were PrimerF/InternalF and PrimerR/InternalR. qPCR conditions were as follows: 3 min at 95°C, then 35 cycles of 95°C for 1 minute and 45 °C for 7 seconds. Additionally, melt curve analysis was performed from 65°C to 95°C, with 0.5°C increments every 5 seconds.

Analysis of qPCR results. qPCR analysis was completed using Bio-Rad CFX Manager™ Software on automatic baseline settings. Cq determination mode was set to single threshold.

(i). Generating standard curve. The standard curve of the control oligonucleotide was constructed using GraphPad Prism 5, and the equation of the best-fit line was found to be $y=-4.714log_2(x) + 14.82$, where y is the Cq value and x is amount of DNA in pg/reaction. This curve was used in later analyses to determine the relative amount of DNA in qPCR samples using their respective Cq values.

(ii). Normalizing DNA Concentrations. Ligated DNA concentrations had to be normalized due to inconsistencies in single-stranded DNA oligonucleotide concentrations before annealing, which could not be controlled for otherwise. Cq values obtained from the amplification of 100- and 0-GC ligations using primers InternalR and PrimerR were normalized to the Cq value of the 5-minute 100-GC ligation to obtain values for 5-minute 100-GC, 5-minute 0-GC, 20-hour 100-GC, and 20-hour 0-GC ligations. The Cq values obtained from the PrimerF/PrimerR (Fig. 2, Table 1) amplification of the same ligations were then multiplied by the normalized Cq/Cq values of their respective InternalR/PrimerR reactions to obtain new Cq values that were corrected for DNA concentration. The relative concentration of ligated product in these reactions was determined using the equation of the line generated from the standard curve ($y=-4.714log_2(x) + 14.82$) where y is the Cq value and x is the corresponding pg of DNA in the reaction. The resulting concentrations were then expressed as percentages of the 100-GC 5-minute ligation, as this is the ligation that appeared to have the higher concentration of ligated product (Fig. 5).

(iii). Statistical analysis. All p-values were obtained using unpaired, two-tailed Student's *t*-test.

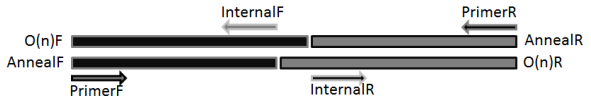

Fig 2 Location of primer annealing on ligated double-stranded synthetic oligonucleotides. (n) represents the different overhang GC-content of the oligonucleotides (100, 50, or 0). See Table 1 for primer sequences.

RESULTS

Phosphorylation of oligonucleotides prior to annealing yields desired product upon ligation of complementary overhangs. To examine the effect of GC-content in overhangs of the synethetic oligonucleotides on ligation efficiency, pairs of 48bp double-stranded oligonucleotides were synthesized with single-stranded overhangs of 4 nucleotides in length with differing GC-content. Double-stranded oligonucleotides were created by allowing 48-base single-stranded DNA fragments to anneal with 52-base fragments (9), with the additional 4 bases representing the resulting oligonucleotide overhangs. Following annealing, the resulting oligonucleotides were 5'-phosphorylated to facilitate ligation by T4 DNA ligase. However, agarose gel electrophoresis of ligated oligonucleotides that were annealed then phosphorylated showed not only the desired 100 bp band, but also additional bands in 50 bp increments (Fig. S1). This suggests blunt-end ligations were occurring in addition to the desired sticky-end ligations due to phosphorylation of both the blunt and sticky ends of each oligonucleotide. In addition, ligation of annealed and phosphorylated oligonucleotides with non-complementary sticky ends resulted in 100 bp products (Fig. S2), confirming the occurrence of blunt-end ligations. Sticky-end ligations were not possible due to the lack of complementary overhangs. Taken together these data show blunt-end ligations were possible when the 5' ends were phosphorylated. To address this issue, only oligonucleotides AnnealF and AnnealR were phosphorylated prior to annealing with complementary oligonucleotides, which ensured that only the 5'-end of the double-stranded oligonucleotide which is involved in sticky-end ligation would be phosphorylated. The ligation of these annealed oligonucleotides yielded 100bp products; the lack of a laddered banding pattern indicated blunt-end ligations no longer occurred (Fig. 3). Faint 50bp bands indicated the presence of unligated oligonucleotides (Fig. 3).

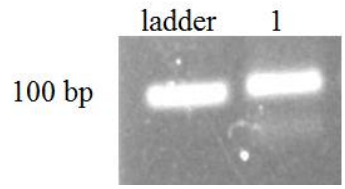

Fig 3 Phosphorylating AnnealF and AnnealR prior to annealing yields ligations with desired 100bp product. Ligated oligonucleotides were run on a 2% agarose gel in 1X TBE at 120V for 60 minutes. Lane 1 contains ligated oligonucleotides. 100 bp ladder was used.

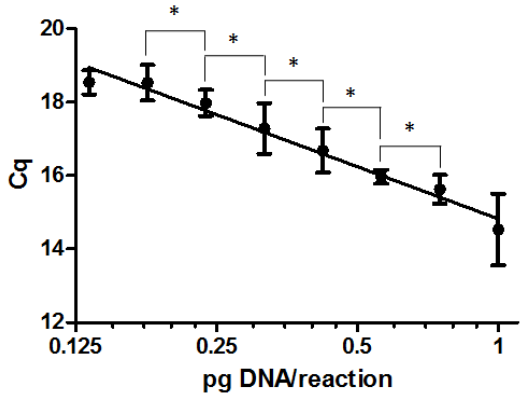

Fig 4 Standard curve for amplification of control oligonucleotide using qPCR. Control oligonucleotide starting at 1pg/reaction was serially diluted 8 times, with each dilution have 25% less DNA than the preceding dilution. Each sample was run in replicates of 8. Cq value represents the cycle number at which the fluorescence signal of the amplified products crosses the set threshold. Equation of the line: y=-4.714$\log_2(x)$ + 14.82. Unpaired, two-tailed Student's t-test was used for stastical analysis ($p<0.05$).

qPCR of control oligonucleotide shows a dose-dependent relationship, and can distinguish 1.33-fold differences in template concentration. Development of a standard curve is necessary when using qPCR to quantify the number of DNA copies in a sample; in this study, DNA copies refers to the concentration of ligated product. Thus, this standard curve compares Cq value and pg of control oligonucleotide. It allows for Cq values obtained from qPCR analysis of ligations to be compared and converted into relative concentrations of ligated product (e.g. ligation X resulted in 2-fold more product than ligation Y) but not absolute concentrations (e.g. ligation X resulted in 1 pg/ 100% ligated product). The standard curve was created by amplification of serial 1.33-fold dilutions of a control oligonucleotide (Fig. 4), which was identical to the 0-GC ligated product. An inverse, log-linear relationship was observed via qPCR between control oligonucleotide dilution and Cq value, for samples between 0.13pg and 1pg per 20µl reaction (Fig. 4), indicating a dose-dependent relationship. In addition, all dilutions on the standard curve in the range of 0.17-0.75 pg/reaction have significantly lower Cq values from thepreceeding dilution (Fig. 4). This indicates that the qPCR method can distinguish between differences in DNA template concentration of as little as 25%.

Oligonucleotides with higher overhang GC-content show higher ligation efficiency. Normalized Cq values from a qPCR of oligonucleotides with varying GC-content in their overhangs were used to determine the relative ligation efficiency of the different oligonucleotides. As shown in Figure 5, 100-GCs resulted in an increase in the amount of ligated product over 0-GCs in a 5-minute ligation. A similar trend was observed in a replicate experiment (Fig. S4). Taken together these data show that ligation of overhangs with higher GC-content is more efficient, and this can only be observed at early time points.

The results from the ligation of 50-GCs were not considered in the analysis of ligation efficiencies of oligonucleotides with varying GC-content. qPCR amplification of this sample was outside the expected range, and the Cq value was similar to the "No Template" control, which amplified due to background contamination (data not shown). These observations indicate a lack of successful ligation of 50-GCs.

The ligation efficiency of AT containing overhangs is greater than that of overhangs containing only GC nucleotides at longer incubation times. There was no difference in the amount of ligated products between 20-hour ligations of 100-GC and 0-GC oligonucleotides ($p=0.3$, Fig. 5). In comparison, the 20-hour ligation for 0-GC resulted in ~25% more ligated product compared to the 5-minute ligation of this oligonucleotide (Fig. 5). Together, these observations indicate that 0-GC ligation is able to produce the same amount of ligated product as the 100-GC ligation after a 20 hours.

It was also observed that the 5-minute ligation of the 100-GC resulted in a greater amount of ligated product than the 20-hour ligation of 100-GCs ($p=0.03$, Fig. 5). This was not expected, as the 20-hour ligation should have resulted in at least as much ligated product as the 5-minute ligation, due to the increase in reaction time. Furthermore, while there was no statistical difference in the amount of product between the two 20-hour ligations, there was also no difference in the amount of product between the 100-GC 5-minute ligation and the 0-GC 20-hour ligation ($p=0.25$, Fig. 5), indicating inconsistency in the results.

DISCUSSION

Sticky-end ligation is a popular tool used for molecular biology, and this popularity necessitates a thorough analysis of factors that may affect how it is applied. How the composition of sticky-ends influences ligation efficiency is not well understood, and there are few techniques that can provide a precise quantitation of these effects (4, 6). A method using qPCR has been developed to measure ligation efficiency (5). We further develop and validate this method by using synthetic oligonucleotides to determine the effect of GC-content of the overhangs on ligation efficiency. By employing this method, we compared the ligation efficiency of oligonucleotides containing 4-base overhangs with G+C content of 100% (100-GC), 50% (50-GC), and 0% (0-GC).

Results from our qPCR analysis of the ligation of oligonucleotides with differing GC-content in their overhangs indicate that higher GC-content increases the efficiency of ligation, which supports our hypothesis. We speculated that 3 hydrogen bonds between G and C dissociate at a lower rate than 2 hydrogen bonds between A and T within the given ligation time. 100-GC ligations resulted in 25% more product compared to the 0-GC , in 5 minute ligations with 25% of the recommended ligase concentration (10). However, the range of conditions in which this discrepancy can be observed is not known; it is possible that increasing the ligation time to 10 minutes or increasing the amount of ligase to the recommend concentration would eliminate any observed differences in ligation efficiency. Further experiments are required to determine the range of conditions in which increased GC-content in oligonucleotide overhangs leads to increased ligation
efficiency. This study also only examines overhangs composed of 4 bases; the numbers of bases may affect ligation efficiency, and altering the length of overhangs may eliminate any difference in ligation efficiency between overhangs of differing GC-content due to the increase in the absolute number of base pairs.

The results obtained from qPCR amplification of the 50-GC ligation (data not shown) are not reported because the original oligonucleotides obtained from Qiu *et. al* lacks a single base in the overhang region that is required for successful ligation (5) (Table 1). Consequently, Cq values obtained were only moderately lower than those obtained from the negative control lacking template DNA.

The observation that 20-hour ligation of 0-GCs results in a 25% increase in product formation compared to a 5-minute ligation of these oligonucleotides shows that 5 minutes is insufficient for this ligation reaction to reach completion, where all non-ligated DNA will be ligated. The observation that 20-hour ligation of 100-GCs did not result in more ligated product formation over 5-minute ligation of these oligonucleotides indicates that a 5 minute reaction results in equally as much ligated product as a 20-hour ligation, which highlights the efficiency of 100-GC ligation reactions. The implications of this result are significant; if a restriction enzyme is chosen which yields 4-bp overhangs consisting of 100% GC-content, downstream ligation of these overhangs can be achieved in 5 minutes or less, while only using 25% the suggested amount of ligase (10). This would result in a significant reduction in the time necessary for ligations, in addition to a significant reduction in the cost of individual ligation reactions. The validity of this result, however, has not been tested in ligation reactions consisting of vectors and inserts rather than 50bp synthetic oligonucleotides.

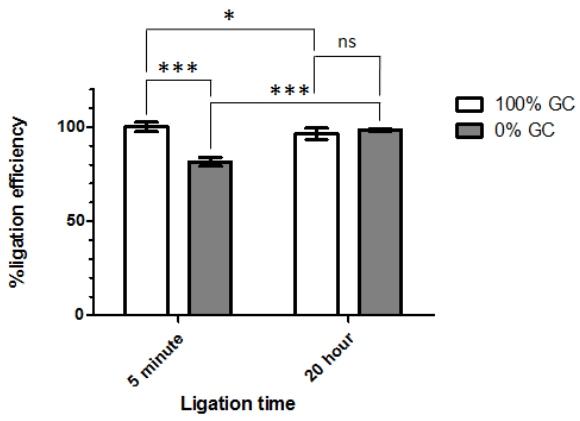

Fig 5 Ligation efficiencies of oligonucleotides with varying overhang GC-content. % ligation efficiency relative to 5-minute ligation of 100-GCs. Efficiencies were normalized to internal primer controls to account for differing pre-ligation DNA concentrations. * denotes $p<0.05$, *** denotes $p<0.001$. Unpaired, two-tailed Student's *t*-test was used for stastical analysis.

Considering the fact that different types of DNA would presumably result in different reaction kinetics (13), and also that ligations of inserts into vectors consist of two ligation sites per successful ligation rather than a single site as in this study, the results of the current study would require validation in these more experimentally applicable conditions.

The purpose of creating a standard curve using dilutions of control oligonucleotide that resulted in 25% differences in concentration between dilutions was twofold: it allowed Cq values obtained from experimental ligations (100-, 50-, and 0-GCs) to be converted into relative input DNA concentrations, and it partially validated the use of qPCR to determine ligation efficiency by demonstrating that significance could be achieved with input DNA concentration differences of as little as 25% (Fig. 4). The latter purpose was key in determining if it would be possible to identify a difference in ligation efficiency if one existed. We suspected that if a difference in ligation efficiency did exist between oligonucleotides with different overhang compositions, this difference would likely not be larger than 25%. qPCR is typically used to determine the copy number of a target gene in a cell sample compared to a reference gene, which typically exists as a single copy in each cell. This means that the copy number of the target gene will either be the same or at least twice as much as the reference, as copy number can only be expressed in whole numbers. As such, it has been shown that the limit of detection for qPCR is a 2-fold difference in copy number (15). A 2-fold difference represents a 50% reduction in DNA concentration/copy number, and thus we needed to determine if a 25% reduction (i.e. a 1.33-fold difference) in DNA concentration was detectable.

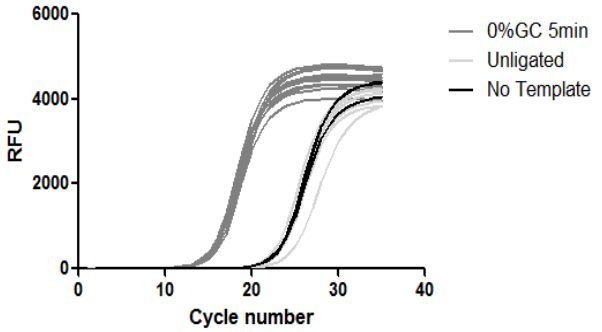

Fig 6 qPCR amplification curves of 0-GC ligation and negative controls. Dotted line represents the fluorescence threshold for Cq values. RFU is Relative Fluorescence Units, indicates the amount of fluorescence observed in a sample.

Indeed, we determined that 25% differences in DNA concentration could be significantly distinguished ($p<0.05$, Fig. 4), thus validating the use of qPCR to resolve slight differences in ligation efficiency. In reference to the former purpose for constructing a standard curve, the equation of the line obtained from the curve was later used to determine the amount of ligated product that resulted from the ligations of the various oligonucleotides. However, we were only able to determine the relative concentrations of ligated product, rather than the absolute concentrations. This is because we lacked the necessary control samples to compensate for inconsistencies in DNA concentration between samples. This is best understood in the following discussion on how differences in DNA concentration were adjusted for when comparing ligation efficiencies between the 100-, 50-, and 0-GC oligonucleotide ligations.

For this study, it is key to ensure that the concentration of DNA is consistent between ligation reactions; if more DNA is present in the ligation of 100-GCs than 0-GCs, for example, this will result in an increase in the amount of ligated product for the 100-GC ligation even if there is no difference in ligation efficiency between the oligonucleotides. Since it was not possible to quantify the amount of DNA in a sample using a spectrophotometer either before or after ligation due to agents in the lyophilized oligonucleotides and in the ligation buffer that interfered with accurate quantification, a method was devised to compensate for potential differences in DNA concentration during the qPCR reaction itself. This was done by creating internal primers (InternalF and InternalR, Fig. 2) that, when paired with PrimerF and PrimerR, respectively, would amplify the individual oligonucleotides whether they had been ligated or not. Each set of primers would amplify either the forward or reverse oligonucleotide, and the Cq values of these samples could be compared between the 100-, 50-, and 0-GC ligations. Since no standard curve was created for these internal primers, a direct correlation between Cq value and DNA concentration could not be determined, but a relative concentration could be determined. The amplification of the reverse oligonucleotide had a larger Cq value than that of the forward oligonucleotide for all samples (data not shown), therefore we concluded that the stock solution for the reverse oligonucleotide likely had a lower concentration of DNA. Accordingly, the Cq values obtained using PrimerF and PrimerR (Table 1, Fig. 2) for the 100-, 50-, and 0-GC ligation reactions were normalized to the Cq values obtained from their respective reactions using InternalR and PrimerR, as this reverse oligonucleotide was the limiting reagent (i.e. the maximum concentration of ligated products could not exceed the concentration of reverse oligonucleotide, due to a lack of ligation partner for the forward oligonucleotide). The normalized values were then analyzed for statistical significance, and relative percentage ligation efficiencies were determined (Fig. 5).

All qPCR reactions contained slight contaminants, as amplification in late cycles was observed in control samples containing primers but not template. If this contamination was unevenly dispersed between samples (e.g. aerosol), then we would not expect replicates of negative controls to have consistent Cq values, as they are unlikely to have received an identical "dose" of contamination. It is more likely that the contamination was present in a reagent that was used for all samples, thus providing a consistent background that should not impact results. The samples lacking template had an average Cq of ~24, which is >7 cycles later than the experimental samples (Fig. 6), indicating >100x decrease in DNA concentration (Fig. 6). This amount of contamination is sufficiently low to not interfere with experimental results and will be considered negligible for our experiment. qPCR amplification of samples containing unligated, complementary oligonucleotides was consistent with amplification of "No Template" controls, suggesting that only oligonucleotides which have successfully ligated are viable templates for amplification of desired 100 bp product (Fig. 6).

In this study, we validated the use of qPCR as a method to compare the ligation efficiencies of different oligonucleotides. Using this method, we determined that the ligation of oligonucleotides with 4-bp overhangs containing 100% GC-content is more efficient than those with 0% GC-content, thus indicating that sticky-end ligation efficiency is influenced by overhang GC-content. Our data also suggests that the ligation of 100% GC-content in 4-bp overhangs is able to reach completion within 5 minutes using 25% of the amount of ligase suggested by the manufacturer, which translates into downstream economic efficiency for the researcher.

FUTURE DIRECTIONS

We showed that qPCR is an effective method to identify differences in ligation efficiency. This method can be further applied to determine absolute ligation efficiency. Using internal primers such as the ones we have designed, two standard curves could be created alongside the standard curve we have generated. Together, these curves would allow values obtained for both the amplification of the ligated oligonucleotides and the un-ligated "internal" nucleotides to be normalized for differences in DNA concentration. This would lead to absolute quantification of ligation efficiency as opposed to the relative quantification observed in the current study.

Additionally, it would be beneficial to determine if ligation efficiency is discretely increased as GC-content is increased, which could be studied by comparing oligonucleotides with 4-bp overhangs consisting of 0%, 25%, 50%, 75%, and 100% GC-content using the method we have described. The differences in efficiency of these overhang types are likely to be minor and possibly undetectable under the conditions that we have employed. However, it is possible that the maximum difference in ligation efficiency between overhang types is not observed at 5 minutes with 0.25U of ligase; thus, titrations of ligation time and ligase concentration should be performed to further optimize this method.

It would be worthwhile to repeat these experiments using a vector and insert ligation approach, as this method is one of the most widely used tools in molecular cloning. Our study was aimed at understanding how GC-content of overhangs affects ligation efficiency when using this tool, it does not directly mirror the conditions that would be present in a typical research setting, despite the increasing use of small oligonucleotides in synthetic biology, such as with Gibson cloning.

Finally, GC-content is not the only factor that plays a role in ligation efficiency, as overhang length and possibly even stacking interactions may also impact efficiency. This method could also be used to quantify the effect of these parameters and others, thus opening the door to a plethora of possibilities in this area of research.

ACKNOWLEDGEMENTS

This work was financially supported by the Department of Microbiology and Immunology at the University of British Columbia (Vancouver, Canada). We would like to thank Dr. David Oliver, Jia Wang, Céline Michiels for their knowledge and mentorship during the course of this project. We would also like to thank the media room staff at Wesbrook Building for supplying the equipment required to perform our experiments.

REFERENCES

1. **Conze T, Shetye A, Tanaka Y, Gu J, Larsson C, Goransson J, Tavoosidana G, Soderberg O, Nilsson M, Landegren U.** 2009. Analysis of Genes, Transcripts, and Proteins via DNA Ligation. Annu. Rev. Anal. Chem. **2**:215-39.
2. **Pray L.** 2008. Restriction Enzymes. Nature Education. **1**:38.
3. **Gao S, Li Y, Zhang J, Chen H, Ren D, Zhang L, An Y.** 2014. A modified version of the digestion-ligation cloning method for more efficiency molecular cloning. Anal. Biochem. **453**:55-57.
4. **Bola G.** 2005. Evaluating the role of G,C-nucleotides and length of overhangs in T4 DNA ligase efficiency. J. Exp. Microbiol. Immunol. **8**:1-7.
5. **Qiu J, Chen P, Lin S**. 2014. Development of a Real-Time Polymerase Chain Reaction Method to Measure Ligation Efficiency. J. Exp. Microbiol. Immunol. (in press)
6. **Chang E, Ge B, Lee M, So M, Wang W**. 2005. Investigation of the ligation efficiency of NdeI digested fragments. J. Exp. Microbiol. Immunol. **7**:68-72.
7. **Odell M, Shuman S.** 1999. Footprinting of Chlorella virus DNA ligase bound at nick in duplex DNA. J. Biol. Chem. **274**:14032-9.
8. **Doherty A, Dafforn R**. 2000. Nick recognition by DNA ligases. J. Mol. Biol. **296**:43-56.
9. **Life Technologies**. 2015. Generating the Double Stranded Oligo. Thermo Fisher Scientific Inc. CA.
10. **Life Technologies**. 2015. T4 DNA Ligase. Thermo Fisher Scientific Inc. CA.
11. **New England Biolabs**. 2015. 5'-Phosphorylation of Oligonucleotides prior to using Phusion Site-Directed Mutagenesis Kit (F-541). New England Biolabs, Ipswich, MA.
12. **Yakovchuk P, Protozanova E, Frank-Kamenetskii M**. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res. **34**:564-574.
13. **Revie D, Smith D, Yee T.** 1988. Kinetic analysis for optimization of DNA ligation reactions. Nucleic Acids Res. **16**:10301-10321.
14. **D'haene B, Vandesompele J, Hellemans J.** 2010. Accurate and objective copy number profiling using real-time quantitative PCR. Methods. **50**:262-270.
15. **Bubner B, Gase K, Baldwin IT.** 2004. Two-fold differences are the detection limit for determining transgene copy numbers in plants by real-time PCR. BMC Biotechnology. **4**:14.