# Experimental and theoretical aspects of Golden Gate Assembly assays

**Patryk Strzelecki[1,2], Anastasiia Shpiruk[1], Elise Hoffmann[3], Grzegorz M. Cech[2] ✉, Anna Kloska[4], Pascal Hébraud[1], Nicolas Beyer[1], Florent Busi[5,6] ✉, and Wilfried Grange[5,1] ✉**

[1]Institut de Physique et Chimie des Matériaux de Strasbourg, Université de Strasbourg, CNRS, UMR7504, 23 rue du Loess, F-67034 Strasbourg Cedex
[2]Department of Bacterial Molecular Genetics, Faculty of Biology, University of Gdańsk, Wita Stwosza 59, 80-308 Gdańsk, Poland
[3]Institut de Biologie Moléculaire des Plantes, Université de Strasbourg, CNRS, UPR2357, F-67084 Strasbourg Cedex
[4]Department of Medical Biology and Genetics, Faculty of Biology, University of Gdańsk, Wita Stwosza 59, 80-308 Gdańsk, Poland
[5]Université Paris Cité, UFR Sciences du vivant, F-75006 Paris
[6]Université Paris Cité, CNRS, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris

**Molecular cloning is a routine yet essential technique. Its efficiency relies on ligation and can be greatly improved when using a procedure known as Golden Gate Assembly (GGA). Essential to GGA are type IIS enzymes that have the unique property to cleave downstream their recognition sequence and thus generate any non-palindromic overhangs. Today, GGA benefits from new engineered enzymes with enhanced activity. Concomitantly, high throughput GGA assays (involving the simultaneous study of all overhangs at a time) have proposed optimal GGA substrates with high efficiencies and fidelities. Surprisingly, those assays show either no or unexpected correlation between ligation efficiencies and overhang stabilities. To explain those observations, we present here experiments involving one or two substrates (overhangs) only. When performing GGA at a stable temperature of 37°C, we found that GGA efficiency strongly correlates with overhang stability. Combining those experimental results with a kinetic model, we were able to determine how relevant parameters (time, temperature, molarity, stoichiometry, stacking energy) influence GGA. This work provides a comprehensive view of low-assembly GGA, defines the required optimal conditions and substrates (allowing the fabrication of constructs for single-molecule experiments with unprecedented yields) and gives new insights into DNA ligation that are crucial in biology.**

Correspondence: *Wilfried Grange : wilfried.grange@u-paris.fr. Correspondence can be also addressed to Florent Busi : florent.busi@u-paris.fr or Grzegorz M. Cech : grzegorz.cech@ug.edu.pl.*

# INTRODUCTION

Type IIS restriction endonucleases recognize non-palindromic DNA sequences (of typical length 5 to 6 nucleotides) and cleave at a defined distance from them Szybalski et al. (1991); Pingoud and Jeltsch (2001). These enzymes have contributed to the development of fast and reliable DNA cloning techniques (such as the Golden Gate Assembly (GGA) method Engler et al. (2008)), which have been used to (*i*) insert large assemblies of 52 DNA fragments Pryor et al. (2022), (*ii*) generate libraries of engineered metabolic pathways Tong et al. (2021) and (*iii*) greatly enhance the yield of DNA constructs in single-molecule assays Bell and Molloy (2022). Those methods consist in using rather long primers (e.g. 40 nucleotides in length) to introduce the recognition sequence of type IIS enzymes at both ends of Polymerase Chain Reaction (PCR) products. As the recognition sequence is lost after cleavage, an appropriate choice of non-palindromic overhangs prevents self ligation and thus allows a high yield of target products in a one-pot reaction where the ligase, the restriction endonuclease and all amplified DNA fragments are mixed together in a single tube.

In the past few years, high-throughput sequencing based GGA assays have been developed to measure ligation frequencies of substrates digested by different type IIS enzymes HamediRad et al. (2019); Pryor et al. (2020). These studies report both on the efficiency (*i.e.* the capability of ligase to ligate certain overhangs) and the fidelity (*i.e.* the capability of ligase to ligate a mismatch-containing sequence). Such high-throughput assays generate a quantity of information that cannot be obtained by looking at individual sequences at a time. For 4 base pairs (bp) overhangs, this represents a total of 256 possible substrates and thousands possible mismatches. Because ligase recognizes nicked dsDNA Weiss et al. (1968); Lehman (1974), one would expect that strong (weak) overhangs would have the largest (lowest) ligation efficiencies. Interestingly, the latter were found to be poorly correlated with the GC content HamediRad et al. (2019); Pryor et al. (2020). In addition, experiments performed with far fewer substrates (13 or 35 fragments) with supposedly non-optimal efficiencies yield a large number of transformants Pryor et al. (2020).

Regarding the above-mentioned high-throughput experiments, it should be noted that the simultaneous presence of all 256 overhang sequences in the same tube leads to a competition between the annealing of perfect-matched and mismatched overhangs. As T4 DNA ligase can also catalyze the joining of nicked DNA substrates containing mismatched base pairs Wu and Wallace (1989), this results in two different dsDNA substrates populations available for ligation. The relative concentration of dsDNA substrates with mismatches depends on both GC-content and the number of competitive overhangs and so efficiency values (and possibly fidelity values) will be biased. It remains however difficult, if not impossible, to determine how individual overhang would be affected in a given set of overhangs. Undoubtedly, spurious effects have to be expected. Still, based on those high throughput assays HamediRad et al. (2019); Pryor et al. (2020), online tools have been developed and propose optimal substrates. Finally, we point out that having all substrates at a time may not be practically relevant as the number of fragments to assemble has to remain low in a typical GGA assay Eisenstein (2020). The reason is that the recognition site chosen for GGA cannot be internal to assembly fragments : considering the assembly of 10 fragments of 500 bp (each) and a 6 bp recognition site, there is a 30% chance not to find the chosen recognition site (see Supplementary Text).

In this article, we investigate GGA both from an experimental and theoretical perspective. To this end, we have performed time-dependent GGA experiments on one substrate at a time. Combining gel electrophoresis with a quantitative and an automatic analysis, we have studied GGA efficiency on a total of six specific substrates. Those are found to have either extremely high or extremely low ligation efficiencies by high throughput GGA assays. We determine how temperature and cycling affect ligation as, in traditional GGA assays, cycling is recommended and consists in cycling the temperature between 37°C, which is the optimal temperature for the restriction endonuclease and 16°C, which is the traditionally recommended sub-optimal temperature for T4 DNA ligase. When performing experiments at a stable temperature of 37°C, we observe a strong and statistically significant correlation between ligation efficiency and overhang stability. Interestingly, for optimal substrates (*i.e.* strong overhangs), we find

that cycling is dispensable and may even slightly decrease ligation efficiency. In addition, we show that a given non palindromic overhang (e.g. AGAA) and its complementary partner (e.g. TTCT) yield different results due to the presence of different stacking energies. All those experimental observations are in agreement with a kinetic model, which then defines optimal conditions such as the molarity and stoichiometry of DNA fragments. Based on those findings, we finally show how to easily and rapidly fabricate DNA tethers for single-molecule experiments with unprecedented yields.

# MATERIALS AND METHODS

All enzymes were obtained from New England Biolabs (NEB). Unless otherwise noted, chemicals were obtained from Merck.

### PCR reactions

The synthesis of DNA fragments was performed either with Q5 Hot Start High-Fidelity DNA Polymerase (M0493 for $\beta'\beta$ fragments) or Hot Start Taq polymerase (M0495 for $\alpha\alpha'$ fragments) following standard protocols. The recognition site of BsaI-HFv2 was introduced using primers of $\sim$40 nucleotides in length. The number of cycles was set to 25. For Magnetic Tweezers constructs, the fragment of interest was synthesized with Q5 DNA polymerase and handles were synthesized with Taq using either biotin-16-dUTP (11093070910, Roche) or digoxigenin-11-dUTP (11093088910, Roche) in a 1:4 dUTP to dTTP ratio. PCR products were purified using a Monarch PCR & DNA Cleanup Kit (5 $\mu$g) from NEB and then visualized on an 0.8 to 1.2 % percent agarose gel resolved in Tris-acetate-EDTA electrophoresis buffer. Detailed protocols are given in the Supplementary Text.

### Golden Gate Assembly-like assays

Experiments were performed with T4 DNA ligase (M0202) at a final concentration 12 U/$\mu$L and BsaI-HFv2 (R3733) at a final concentration of 0.6 U/$\mu$L in CutSmart or rCutSmart Buffer supplemented with 1 mM ATP (P0756, NEB) and 10 mM dithiothreitol (DTT, 43816) in a $\sim$10 $\mu$L volume. All experiments were performed with DNA molecules at molarities equal or less than 21 nM. For cycling experiments (37°C for 5 minutes and 16°C for 5 minutes, 5 minutes GGA-cycling), tubes were placed in a thermal cycler. For static experiments (37°C), a dry bath was used. While the time is well defined for static experiments, it takes about 10 to 20 seconds for the thermal cycler to reach and stabilize the temperature and so there might be an uncertainty in time of 5 to 10% for those experiments. Reactions were stopped with the addition of sodium dodecyl sulfate (SDS, 71725; final concentration of 0.2 %) and an incubation at 65°C for at least 10 minutes. We have kept reaction volumes as low as possible (about 10 $\mu$L) to ensure proper thermal exchange in the cycler. Detailed protocols are given in the Supplementary Text.

### Gel assays and analysis

1.2 % agarose gels (in Tris-acetate-EDTA buffer) were used to follow the kinetic of GGA experiments. Gels were post-stained in ethidium bromide (EtBr, 0.5 $\mu$g/ml) for 15 minutes and destained in water for 10 minutes (twice). The quantification of DNA bands (intensity) was performed automatically using a code written in R (`https://github.com/github.com/wilfriedgrange/gel_analysis`). Details are also given in Supplementary Figure S2. To quantify the stoichiometry of GGA experiments, gels contain two lanes with $\alpha\alpha'$ and $\beta'\beta$ fragments at time zero. From those, we calculate the ratio $S_\beta$ defined as $100 \times \beta'\beta/(\beta'\beta + \alpha\alpha')$, where $\beta'\beta$ and $\alpha\alpha'$ are intensities (determined from gel assays). For a stoichiometry of 3 ($[\alpha\alpha']/[\beta'\beta]$) and 2 fragments experiments (e.g. those reported in Figure 2), $S_\beta$ = 49.7 (see Supplementary Figures S2).

### Magnetic Tweezers

We use a home-built Magnetic Tweezers instrument capable of tracking a few tens of beads in real-time at 400 Hz. For illumination, we use a Superluminescent Diode (SLD-mCS-261-MP2-SM, Superlum) providing up to 2 mW at 670 nm ($\sim$0.43 mm (1/e2) beam, $\sim$2.5 mrad using a PAF-X-2-B fiberport (Thorlabs)). The force acting on the magnetic beads is controlled with a pair of magnets (W-05-N50-G, Supermagnete) placed in a vertical configuration (gap size of 1 mm) Lipfert et al. (2009). The position and rotation of the magnets are adjusted with two motors: M-126.PD1 (Physik Instrumente) and ST4118M1804-L (Nanotec), respectively. To image the magnetic beads (65001, Invitrogen) we use a UPLFL 100XOI objective (Olympus) mounted on a piezo-focusing system (P-726.1CD, Physik Instrumente) and a high-speed camera (HXC-13, Baumer) set at 400 Hz (800 x 800 pixels) controlled with a

microEnable IV AD4-CL frame grabber (Silicon Software). Using an EF 300mm F/4 L USM telephoto lens (Canon), this gives a pixel size of 84 nm. The position of the beads (Dynabeads MyOne Streptavidin C1; 65001, Thermo Fisher Scientific) is determined using a software based on the work of V. Croquette Gosse and Croquette (2002); Lionnet et al. (2012) and ported later on CUDA by S.J. Tabrizi Lansdorp et al. (2013). The flow cell consists of two $\sim$0.16mm thick coverslips and a 0.074 mm double-sided acrylic tape (9495MP, 3M). The inner bottom surface of the flow cell is coated with an hydrophobic solution (nitrocellulose or Sigmacote) to allow attachement of anti-Digoxigenin (11093274910, Roche) Lionnet et al. (2012). Measurements were performed in phosphate-buffered saline (pH 7.4), containing 0.5 mg/mL bovine serum albumin (BSA) and 0.1 % Tween 20.

## Kinetic model

We solved differential equations using the R package deSolve and the R function lsoda. Details are given in the Supplementary Text.

# RESULTS

## Two fragments reaction

We first show GGA experiments performed for different sequences having different overhang stabilities (as reported by SantaLucia (1998)) and different stacking energies, *i.e.* the energy determined from previous nicking experiments Protozanova et al. (2004). Here, we define a sequence as a succession of 6 nucleotides, where nucleotides in positions 2,3,4 and 5 form the non-palindromic overhang generated by the restriction endonuclease BsaI-HFv2. In contrast, nucleotides in positions 1 and 6 govern stacking energies (see Supplementary Tables S1 and S2 for energy values). These experiments were performed at a static temperature of 37°C or by cycling the temperature between 37°C and 16°C every 10 minutes (5 minutes at each temperature). The kinetics was determined by measuring the concentration of products at 10, 30, 70, 140 and 240 minutes.

Specifically, nucleotides in positions 2,3,4 and 5 have been chosen by looking at the results of GGA sequencing-based assays Pryor et al. (2020), which report on the ligation efficiency and fidelity when using four different type IIS enzymes. Based on this study, we selected 4 bp overhangs with the highest ligation efficiencies: `AGGA`, normalized ligation frequency of 739 (reported values are normalized to 100,000 ligation events, see also Supplementary Table S7) and `AGAA` and `TTCT`. These overhangs are considered to be equivalent and non discernable in Pryor et al. (2020) and yield an averaged frequency of 731. Similarly, we have selected the 2 lowest frequencies values (247 for `CCGC` and 234 for `CACC`, respectively). Here the 2 fragments $\alpha\alpha'$ (565 bp) and $\beta'\beta$ (1678 bp) synthesized by PCR (Figure1) were mixed in a 3:1 ratio (final molarity of 21 and 7 nM, respectively) and T4 DNA ligase and BsaI-HFv2 were added to the reaction tubes. Both $\alpha$ and $\beta$ do not contain the recognition site of BsaI-HFv2 (see Figure 1) so that the product $\alpha\beta$ remains stable after ligation. In contrast, all other ligations yield non-stable products ($\alpha\alpha'$, $\beta'\beta$ and $\beta'\alpha'$), which can be re-digested as they do contain the recognition site of BsaI-HFv2. Note that nucleotides in positions 1 and 6, adjacent to the overhang, have been carefully chosen to mimic what happens in a real GGA experiment (e.g. assembling a PCR fragment in a destination vector) where stacking energies are usually not conserved among the different possible products (see Figure 1 and Supplementary Figure S1). Except for `GCCGCC` (see Supplementary Table S1), all our experiments reproduce this behavior.

Figure 2 reports the percentage of product formed along the course of the experiment, which we define as $\Psi_\beta$ ($\Psi_\beta = 100 \times [\alpha\beta]/([\beta'\beta] + [\alpha\beta])$, where square brackets represent molar concentrations). To accurately determine the concentrations of those products, we perform agarose gel electrophoresis combined with an automated and quantitative analysis (Supplementary Figure S2). At a given time, we first observe that the percentage of product formed increases with the stability of the overhangs : `CCGC` or `CACC` overhangs (energy of -6.25 and -4.73 kcal/mol, Table S1) are far more efficient than `AGAA` (-3.58 kcal/mol) and allow to reach high $\Psi_\beta$ values within 10 minutes. This trend is observed at all time points (Supplementary Tables S3 and S4) for both the static and cycling assays. At 37°C, we observe significant (Supplementary Table S5) and high correlations (absolute values in excess of 0.8) between $\Psi_\beta$ and the overhang stability (or the total free energy that accounts for stacking energies). For cycling experiments, the correlation is lower but non negligible (absolute values of $\sim$0.6). Most probably, lowering the temperature (as done in a cycling assay) prevents us from observing temperature-dependent characteristics (such as the stability of overhangs). These are the expected results when performing a simple ligation at low temperature (e.g. 16°C or 4°C) as recommended in ligation protocols. For comparison, we performed a statistical analysis on results obtained by high throughput assays HamediRad et al. (2019); Pryor et al. (2020) (Supplementary Figure S3 and S4). There, experiments have always been performed with temperature cycling, as recommended by GGA protocols, and show either unexpected correlation or no correlation.

As seen in Figure 2, the difference in $\Psi_\beta$ is time-dependent and pronounced at short times (30 minutes and less). In contrast, for longer reaction times, all overhangs show comparable $\Psi_\beta$ (albeit still different and showing a significant correlation for static assays, Supplementary Tables S3 and S4). After 240 minutes, we obtain $\Psi_\beta$ values of 80% to 90% for all overhangs, which is consistent with time-dependent GGA experiments of Bell and Molloy (2022). This

behavior is expected as the last reaction is irreversible Lohman et al. (2011) and so, for any overhang, $\Psi_\beta$ values shall reach ∼100% when the reaction approaches completion. Surprisingly, ligation efficiencies values as reported by high throughput assays differ by up to 300% Pryor et al. (2020) or up to 400% (removing outliers) HamediRad et al. (2019) after 300 minutes.

For stable overhangs, Figure 2 shows that static assays should be favored over cycling assays. In other words, for those overhangs, the loss in T4 DNA ligase activity at 16°C (and the complete loss of restriction endonuclease activity) is far more prejudicial than the possible gain in the dissociation constant when decreasing temperature. Here, this observation has important implications and suggests that, for stable overhangs, ligation should be performed at 37°C (as expected from the temperature-dependence activity of T4 DNA ligase Weiss et al. (1968); Pohl et al. (1982)). Finally, we observe that stacking energies play a role: while `TTTCTT` and `TAGAAT` have exactly the same overhang stability (as `TTCT` and `AGAA` are complementary to each other and not discernible in sequencing-based assays), `TAGAAT` that has a much weaker stacking energy (-1.53 as compared to -2.22 kcal/mol, Supplementary Table S2) and is far more inefficient in a static assay. To a lesser extent (see time points at 10 and 30 minutes), `GCCGCC` with a stacking energy of -3.61 kcal/mol seems to be more efficient than `TCCGCT` with a stacking energy of -2.49 kcal/mol.

## Kinetic model

To strengthen the above findings, we present in this section a kinetic model that is able to qualitatively reproduce experimental trends (Figure 3.A). We assume that enzymatic reactions follow a Michaelis–Menten model involving a binding step between the enzyme and the substrate (represented by a centered dot) and a catalytic step. At $t$ = 0, only $\alpha\alpha'$ and $\beta'\beta$ have a non-null molarity. The formation of the final product $\alpha\beta$ involves the binding of $\alpha$ and $\beta$ (association rate $k_+$, dissociation rate $k_-$) and so the restriction endonuclease needs first to bind (rate constant $\kappa[E]$) and cleave (rate constant $k_\epsilon$) DNA (steps 1 and 2 for both $\alpha\alpha'$ and $\beta'\beta$). Ligase can then bind (rate constant $\kappa_{\alpha\cdot\beta}[L]$ for the nicked $\alpha\cdot\beta$ intermediate) and form appropriate phosphodiester bonds (rate constant $k_\lambda$). This results in joining $\alpha$ and $\beta$ (steps 3 and 4) and so the formation of $\alpha\beta$. Note that $\kappa$ values depend on DNA substrates (such as $\alpha\cdot\beta$ or $\alpha\cdot\alpha'$) as those involve different stacking energies, which possibly influence binding of ligase to the nicked substrate. The route involving steps 1,2,3 and 4 is the most direct route (highlighted in red in Figure 3.A) to form the target product but other routes exist that can significantly alter the overall kinetics: for instance, $\alpha$ and $\alpha'$ (as well as $\beta'$ and $\beta$) can bind and be joined by ligase to form initial reactants $\alpha\alpha'$ ($\beta'\beta$) (steps 5 and 6). As both these routes involve binding of $\alpha$ or $\beta$, they reduce the rate formation of $\alpha\beta$. In contrast, there is one route that indirectly favors the formation of $\alpha\beta$: $\beta'$ and $\alpha'$ can bind and be joined by ligase to form a product $\beta'\alpha'$ (steps 7 and 8), which will eventually be cleaved by the restriction endonuclease (steps 9 and 10). In this model, we assume (*i*) all reactions involving enzymes to be irreversible, (*ii*) identical dissociation and association rates for all possible partners (hybridization of 4 bp overhangs) and (*iii*) constant enzyme concentrations so that $\kappa[E]$ (binding of the restriction endonuclease $E$), $\kappa_{\alpha\cdot\beta}[L]$, $\kappa_{\alpha\cdot\alpha'}[L]$, $\kappa_{\beta'\cdot\beta}[L]$ and $\kappa_{\beta'\cdot\alpha'}[L]$ (binding of the ligase $L$) are considered as first-order reaction rates (Supplementary Text and Figure S5). We furthermore assume the activity of enzymes to be temperature dependent with a change in activity of 100 % (60 %) for the restriction endonuclease (ligase) at 16°C Weiss et al. (1968); Pohl et al. (1982). Finally, note that it would have been equivalent to define new first-order constants (e.g. $k_\lambda^\star$ defined as : $\left(k_\lambda^\star\right)^{-1} = \left(k_\lambda\right)^{-1} + \left(\kappa_{\alpha\cdot\beta}[L]\right)^{-1}$) as enzymatic reactions here involve two irreversible first-order reactions.

Figure 3.B shows the result of the numerical calculations with $\kappa[E]$, $\kappa_{\alpha\cdot\beta}[L]$, $\kappa_{\alpha\cdot\alpha'}[L]$, $\kappa_{\beta'\cdot\beta}[L]$ and $\kappa_{\beta'\cdot\alpha'}[L]$ equal to 0.1 s$^{-1}$, $k_\lambda$ and $k_\epsilon$ equal to 1 s$^{-1}$ (37°C) and $k_+$ equal to 10$^6$ M$^{-1}$s$^{-1}$. For a stable overhang, we set $k_-$ at 1 s$^{-1}$ (at both 37°C and 16°C); for a weaker overhang we set $k_-$ at 10 s$^{-1}$ (37°C) and $k_-$ at 2 s$^{-1}$ (16°C). Those values are found to best reproduce experimental trends and are in qualitative agreement with the literature. Most enzymes have turnovers in the range 1 to 100 s$^{-1}$ Bar-Even et al. (2011). Given the limit of diffusion of ∼10$^8$-10$^9$ M$^{-1}$s$^{-1}$ Alberty and Hammes (1958), it is reasonable to assume $\kappa$ values in the order of 10$^6$-10$^7$ M$^{-1}$s$^{-1}$ (which yields $\kappa[L]$ of ∼0.1 s$^{-1}$ for an enzyme in the sub micro-molar range). Finally, the values of the association $k_+$ and dissociation

rates $k_-$ are in qualitative agreement with recent ensemble and single-molecules experiments, which report $k_+$ of $\sim$0.2$\times$10$^5$ M$^{-1}$s$^{-1}$ and $k_-$ of $\sim$ 20 s$^{-1}$ for an immobilized 8-mer DNA molecules Yazawa and Furusawa (2018). As shown in Figure 3.B, the model fails to reproduce the exact $\Psi_\beta$ values but successfully reproduces all experimental trends: (*i*) cycling is prejudicial for stable overhangs, (*ii*) stable overhangs are more efficient than weaker ones and (*iii*) there is a marked difference in $\Psi_\beta$ for weak overhangs for static and cycling assays. Interestingly (and as expected) a change in only $k_-$ is sufficient to reproduce the experimental behavior of both strong and weak overhangs.

Numerical calculations allow to explore the influence of various parameters. For cycling experiments (when one has to work with weak overhangs), our model suggests that a 5 minutes GGA-cycling is certainly a good choice (Supplementary Figure S6). Longer cycling durations significantly decrease $\Psi_\beta$ and shorter cycling durations only slightly enhance $\Psi_\beta$ (knowing that it may be difficult to experimentally control fast changes in temperature). For practical reasons (e.g. amount of DNA produced per PCR reactions, sensitivity of agarose gels-based assays) as well as to optimize DNA molecules for single-molecule experiments (see below), we have chosen to work at $[\alpha\alpha']$ = 21 nM and $[\beta'\beta]$ = 7 nM. Obviously $\Psi_\beta$ strongly depends on both concentrations and stoichiometries. As shown in Figure 4, increasing both the $[\alpha\alpha']$ to $[\beta'\beta]$ ratio and $[\alpha\alpha']$ increase $\Psi_\beta$ and so, at any time point, one obtains the largest $\Psi_\beta$ values in the fourth quadrant of the graphs ($[\alpha\alpha'] \geq$ 16nM and $[\beta'\beta] \leq$ 7nM, which is what we have done). In contrast, working at concentrations less than 10 nM for both $\alpha'\alpha$ et $\beta\beta'$ is not recommended (see also Supplementary Figure S7 showing experiments and calculations performed at $[\alpha\alpha']$ = $[\beta'\beta]$=7 nM). At a given $\beta'\beta$ concentration, it is always possible to reach high $\Psi_\beta$ by increasing the $\alpha\alpha'$ concentration. However some experiments such as the assembly of several fragments for the construction of libraries Tong et al. (2021) are usually performed at equal molarities. There, it may be relevant to define an additional metric $\xi$ ($\xi = 100 \times [\alpha\beta]/([\beta'\beta] + [\alpha\alpha'] + [\alpha\beta])$) that introduces a penalty upon increasing the molarity of one fragment only. As in Ng and Sarkar (2012), $\xi$ may be referred as an efficiency. Supplementary Figure S8 shows however that it remains challenging to reach high $\xi$ values at equimolar concentrations (in a few hours) as this requires to work at molarities larger than 100 nM.

So far, we have falsely assumed identical $\kappa_{\alpha\cdot\beta}[L]$, $\kappa_{\alpha\cdot\alpha'}[L]$, $\kappa_{\beta'\cdot\beta}[L]$ and $\kappa_{\beta'\cdot\alpha'}[L]$ values. In other words we have considered that stacking energies are conserved among the different products. As seen in Supplementary Figure S1, this is usually not the case in real experiments as stacking energies may be constrained by the choice of the vector. In Figure 5, we show additional experiments performed on `CCGC` and `AGAA` overhangs. In contrast to experiments performed in Figure 2, nucleotides in positions 1 and 6 are conserved for all four ligation products ($\alpha\beta$, $\alpha\alpha'$, $\beta'\beta$ and $\beta'\alpha'$). For both experiments (that of Figures 5 and 2), $\kappa_{\alpha\cdot\beta}$ remains however identical as the final product ($\alpha\beta$) leads to an identical sequence (e.g. `TCCGCT`). Here, both experiments show that stacking energies of intermediate products ($\alpha\alpha'$, $\beta'\beta$ and $\beta'\alpha'$) do not alter the kinetics of GGA. This result is in agreement with calculations (Supplementary Figure S9) demonstrating that $\Psi_\beta$ does not show any significant dependence on $\kappa_{\alpha\cdot\alpha'}[L]$, $\kappa_{\beta'\cdot\beta}[L]$ and $\kappa_{\beta'\cdot\alpha'}[L]$ (over 2 orders of magnitude). In contrast, a large change in stacking energies involved in the formation of the main product $\alpha\beta$ dramatically alters $\Psi_\beta$ values.

## High yield molecules for single-molecule experiments

In the previous sections, we have identified sequences that allow to reach high $\Psi_\beta$ within minutes (e.g. `TCCGCT` and `TCACCT`). Here, we show how one can use those sequences to generate molecules for Magnetic Tweezers (MT) with unprecedented yields. Our approach differs from that of Bell and Molloy (2022) who have recently used GGA to fabricate either 10 kb-long DNA molecules or hairpin-like structures. Here, we will focus on short molecules (e.g. 1,500 to 2000 bp) that are best suited for high-resolution MT experiments Revyakin et al. (2005); Dulin et al. (2015). Figure 6.A shows the results of our static GGA assay using a $\sim$1700 bp DNA molecule (equivalent to a $\beta'\beta$ fragment) and two handles (equivalent to $\alpha\alpha'$ fragments). Both DNA handles are $\sim$550 bp DNA molecules containing either multiple biotin or digoxigenin labels (see Supplementary Text and Supplementary Figure S11) and so allow to create torsionally constrained substrates (Figure 6.B). As before, we have chosen to work at $[\alpha\alpha']$ = 21 nM and $[\beta'\beta]$ = 7 nM.

While the full length construct is barely visible on a gel using conventional ligation methods (see Supplementary Figure S1 of Papini et al. (2019)), it is here clearly visible after 10 minutes of reaction only (Figure 6.A). Because the two ligation events occurring at both ends of the $\sim$1700 bp fragment are independent events, we can easily estimate the percentage of molecules with either 0, 1 or 2 handles from Figure 2 and Supplementary Table S6. Defining $\Psi_{\beta}^{37,t}(s)$ as the percent of product formed for the sequence $s$ at time $t$ for a static assay, the expected percentage of fragments with 2 handles at time $t$ is $\Psi_{\beta}^{37,t}(s1) \times \Psi_{\beta}^{37,t}(s2)$, where $s_1$ ($s_2$) denotes the TCCGCT (TCACCT) sequence, respectively. A similar argument (independence) allows to determine the percentage of molecules with 0 and 1 handle. For reaction times of 70 minutes and above, we find (Supplementary Table S6) that almost 70 percent of molecules are full length fragments, *i.e.* a main fragment with a biotin and a dig handle. Interestingly, the presence of biotin and digoxigenin labels (at least for those overhangs) is not an obstacle to the formation of the full length fragment: in a control experiment using non-modified dUTPs for the PCR fabrication of DNA handles, we found similar results (see Supplementary Figure S10).

The typical high yield of our approach allows to use the reaction mixture as it is, without the need for an additional step of gel purification. In addition, this allows to determine the concentration of the full length fragment with a high precision and so control the DNA to bead ratio Revyakin et al. (2005). As a benchmark, we show in Figure 6.C typical MT-hat curves that represent the extension of the DNA fragment as a function of the degree of supercoiling. Those curves were recorded at 0.57 pN (the typical fraction of supercoilable molecules was found to be 0.5 or above). Here we have used DNA molecules after a 140 minutes ligation but shorter reaction times could be obviously used (e.g. 30 or 70 minutes). Still, our approach allows to fabricate molecules for MT in a reproducible way in a time-frame of about three hours (starting from plasmid DNA), which is half that of conventional methods (see Figure 6 of Papini et al. (2019)).

# DISCUSSION

High-throughput DNA sequencing has recently provided a comprehensive analysis of GGA. When tens of different overhangs are present, those assays give new and valuable information regarding both the efficiency and fidelity of GGA. For those assays, competitive effects are however likely to occur and may lead to unexpected results. For instance, those assays suggest that stable overhangs (as obtained from thermodynamical data) have lower ligation frequencies (and *vice versa*, Supplementary Figures S3 and S4). A more subtle yet important point is the following: given that that phosphodiester bond formation is irreversible Lohman et al. (2011), one would expect overhangs to show comparable efficiencies at completion. Surprisingly, reported ligation frequencies vary by 300% or more after 300 minutes. Either those experiments report on values that cannot represent end-time points or other effects (e.g. competitive inhibition-like effects) have to be taken into account.

DNA ligase (e.g. T4) is an enzyme that seals dsDNA and so unstacking events at nicked sites certainly play a major role in the ligation process Häse and Zacharias (2016); Zacharias (2020), *i.e* the formation of phosphodiester bonds between adjacent 3'-hydroxyl and 5'-phosphate termini. In contrast, it seems reasonable to assume that the binding of Watson-Crick pairs plays only a minor role in this process. Still, the natural substrate of ligase is nicked dsDNA and so any experimental assay involving the simultaneous determination of DNA overhang hybridization and DNA ligation in the presence of competitive partners (mismatched overhang substrates) would introduce biases. In this hypothesis, an overhang prone to mismatched base pairing (e.g. CG-rich overhang) will likely show a reduced ligation efficiency simply because, in a given time, some ligation events also involve mismatched substrates. This is evidenced by our analysis presented in Supplementary Tables S7 where we have, for each overhang, calculated the ratio $\rho$ of the ligation efficiency to the total number of counts per overhang (involving mismatched and non-mismatched events). There, for overhangs less prone to mismatched base pairing and which have been falsely proposed as efficient in DNA sequencing assays Pryor et al. (2020), $\rho$ approaches 1. Such competitive-like effects depend on the number of possible partners and it remains to determine if this also influences ligation fidelity values. In contrast, those competitive effects are unlikely to occur with far fewer fragments. One may then wonder if the results obtained by high throughput assays are transposable to experiments with few substrates only. For example, assemblies with a relative low number of fragment (e.g. 13) having non optimal ligation efficiencies (as reported by sequencing-based GGA assays) unexpectedly show a large number of transformants Pryor et al. (2020). It is worth mentioning that other recent high-throughput DNA sequencing assays implicitly suggest a possible bias of multiplexed experiments : when directly ligating two phosphorylated oligonucleotides with T4 DNA ligase, initial velocities strongly correlate with GC content (see Supplementary Figure S3 of Bilotti et al. (2022)). In contrast, multiplexed experiments involving the study of ligation by T4 DNA ligase show a moderate correlation between efficiency and GC content.

To answer the above question, we have performed GGA assays involving one substrate only, *i.e.* one overhang at a time. Cycling experiments (as recommended in GGA protocols) show a correlation of $\sim$0.6 (absolute value, Supplementary Tables S4 and S5). Even higher correlation coefficients are found when performing GGA at a static temperature of 37°C. There, a strong and statistically significant correlation (absolute values of $\sim$0.8 and above) is found for the 6 substrates under study. In particular, strong overhangs (`CCGC`), which were reported as inefficient by multiplexed GGA assays, show here remarkable efficiencies. In contrast, weak overhangs (`AGAA`), which were reported as efficient by multiplexed GGA assays, show here a low efficiency. As expected, those findings are stoichiometry independent (Supplementary Figure S7).

Our experiments suggest to use 4 bp overhangs with energies $\leq$ -4.7 kcal/mol (Figure 2). Here, it is interesting to mention that proposed sets of sequences (such as those of HamediRad et al. (2019)) based on efficiency and fidelity do not meet this criterion (Supplementary Table S8). When dealing with a low number of fragments (e.g. five), fidelity is certainly of less importance than efficiency. As such, online tools such as `https://github.com/scottweis1/iBioCAD` on the basis of HamediRad et al. (2019) and `https://goldengate.neb.com` on the basis ofPryor et al. (2020) may give inappropriate results. For our 3 fragments reaction (single-molecule DNA tether), the

proposed overhangs are GAGA and CCGA with overhang stabilities of -3.88 and -5.31 kcal/mol (https://github.com/scottweis1/iBioCAD) and TAGC and CGAA with overhang stabilities of -4.1 and -4.47 kcal/mol (https://goldengate.neb.com).

Our study not only reconciles thermodynamics data and GGA but also gives new insights to optimize GGA. For stable overhangs, we show that cycling the temperature (as recommended in typical GGA assays) is prejudicial (Figure 2). This may explain why large assembly (52 fragments) have failed when cycling and have been successful when performing the reaction at a given temperature of 37°C Pryor et al. (2020). For those stable overhangs, our experimental data ($[\alpha\alpha']$ = 21 nM and $[\beta'\beta]$ =7 nM) show that $\Psi_\beta$ values as high as 84% are reached within 30 minutes (Supplementary Table S3) and so allow to perform GGA in very short times. $\Psi_\beta$ values then slowly increase by a few percent within hours to reach about 90% (strong overhang, $\Psi_\beta^s$) and 80% (weak overhang $\Psi_\beta^w$). When assembling several $n$ fragments, a small difference however accounts for a large effect as the overall difference in efficiency scales as $(\Psi_\beta^s/\Psi_\beta^w)^n$. Obviously, the effect would be even more pronounced if one decides to perform the experiment in a short time (e.g. 30 minutes). Those findings allow to propose optimal substrates with applications in single-molecule research. As a benchmark, we have shown how to easily fabricate DNA assemblies for MT experiments with yields and speed by far superior to conventional methods. Our study also evidences the role played by stacking energies in GGA. For weak overhangs, we show that a given overhang and its complement give clearly different $\Psi_\beta$ values (Figure 2). Again, this role has been known for long (and has been accounted for in simple ligation assays Ng and Sarkar (2012)) but has not been investigated by high-throughput assays. Here, this may also play a role in fidelity and needs being addressed in the future.

All our experimental findings (role of the overhang stability and stacking energy, dependence on the temperature) have been reproduced with a kinetic model that gives experimentalists optimal choices (stoichiometry and molarity) to generate GGA assembly involving a low number of fragments. Overall, our experiments and calculations provide new and valuable informations that have important implications for high-throughput DNA sequencing assays studying ligation (e.g. Bilotti et al. (2022)). In turns, this shall stimulate both experimental assays and theoretical developments.

## SUPPLEMENTARY DATA

Supplementary Data is available.

# ACKNOWLEDGEMENTS

The authors thank Iryna Riabko and Nelson Gartiser (Unistra, Strasbourg) for assistance with Magnetic Tweezers experiments.

Author contributions: W.G. planned the experiments. P.S, E.H, A.K, G.M.C, F.B. and W.G. performed GGA assays and/or contributed to molecular biology assays. P.S, A.S. and W.G. performed single-molecule experiments and analyzed corresponding data. P.S and W.G analyzed GGA data. P.H and W.G derived the kinetic model and performed calculations. N.B. designed and built the setup. W.G. made all figures and R programs, performed statistical analysis and wrote the paper. All authors read the manuscript and contributed to comments or corrections.

## FUNDING

This research was funded by the National Science Centre, Poland (grant SONATINA 2017/24/C/NZ1/00456 to G.M.C) and was supported by a French Government Scholarship (Campus France to P.S.) and by the IdEx Université Paris Cité, ANR-18-IDEX-0001.


***Data availability statement.*** Softwares and source codes related to this work have been deposited in GitHub and are available at https://github.com/wilfriedgrange.

***Conflict of interest statement.*** None declared.

# Bibliography


Alberty, R. A. and Hammes, G. G. Application of the theory of diffusion-controlled reactions to enzyme kinetics. *The Journal of Physical Chemistry*, 62(2):154–159, 1958.

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–10, 2011. doi: 10.1021/bi2002289.

Bell, N. A. W. and Molloy, J. E. Efficient golden gate assembly of dna constructs for single molecule force spectroscopy and imaging. *Nucleic Acids Res*, 50(13):e77, 2022. doi: 10.1093/nar/gkac300.

Bilotti, K., Potapov, V., Pryor, J. M., Duckworth, A. T., Keck, J. L., and Lohman, G. J. S. Mismatch discrimination and sequence bias during end-joining by dna ligases. *Nucleic Acids Res*, 50(8):4647–4658, 2022. doi: 10.1093/nar/gkac241.

Dulin, D., Cui, T. J., Cnossen, J., Docter, M. W., Lipfert, J., and Dekker, N. H. High spatiotemporal-resolution magnetic tweezers: Calibration and applications for dna dynamics. *Biophys J*, 109(10):2113–25, 2015. doi: 10.1016/j.bpj.2015.10.018.

Eisenstein, M. How to build a genome. *Nature*, 578(7796):633–635, 2020. doi: 10.1038/d41586-020-00511-9.

Engler, C., Kandzia, R., and Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, 3(11):e3647, 2008. doi: 10.1371/journal.pone.0003647.

Gosse, C. and Croquette, V. Magnetic tweezers: micromanipulation and force measurement at the molecular level. *Biophys J*, 82(6):3314–29, 2002. doi: 10.1016/S0006-3495(02)75672-5.

HamediRad, M., Weisberg, S., Chao, R., Lian, J., and Zhao, H. Highly efficient single-pot scarless golden gate assembly. *ACS Synth Biol*, 8(5):1047–1054, 2019. doi: 10.1021/acssynbio.8b00480.

Häse, F. and Zacharias, M. Free energy analysis and mechanism of base pair stacking in nicked dna. *Nucleic Acids Res*, 44(15): 7100–8, 2016. doi: 10.1093/nar/gkw607.

Lansdorp, B. M., Tabrizi, S. J., Dittmore, A., and Saleh, O. A. A high-speed magnetic tweezer beyond 10,000 frames per second. *Rev Sci Instrum*, 84(4):044301, 2013. doi: 10.1063/1.4802678.

Lehman, I. R. Dna ligase: structure, mechanism, and function. *Science*, 186(4166):790–7, 1974. doi: 10.1126/science.186.4166.790.

Lionnet, T., Allemand, J.-F., Revyakin, A., Strick, T. R., Saleh, O. A., Bensimon, D., and Croquette, V. Single-molecule studies using magnetic traps. *Cold Spring Harb Protoc*, 2012(1):34–49, 2012. doi: 10.1101/pdb.top067488.

Lipfert, J., Hao, X., and Dekker, N. H. Quantitative modeling and optimization of magnetic tweezers. *Biophys J*, 96(12):5040–9, 2009. doi: 10.1016/j.bpj.2009.03.055.

Lohman, G. J. S., Chen, L., and Evans, T. C., Jr. Kinetic characterization of single strand break ligation in duplex dna by t4 dna ligase. *J Biol Chem*, 286(51):44187–44196, 2011. doi: 10.1074/jbc.M111.284992.

Ng, D. T. W. and Sarkar, C. A. Model-guided ligation strategy for optimal assembly of dna libraries. *Protein Eng Des Sel*, 25(10): 669–78, 2012. doi: 10.1093/protein/gzs019.

Papini, F. S., Seifert, M., and Dulin, D. High-yield fabrication of dna and rna constructs for single molecule force and torque spectroscopy experiments. *Nucleic Acids Res*, 47(22):e144, 2019. doi: 10.1093/nar/gkz851.

Pingoud, A. and Jeltsch, A. Structure and function of type ii restriction endonucleases. *Nucleic Acids Res*, 29(18):3705–27, 2001. doi: 10.1093/nar/29.18.3705.

Pohl, F. M., Thomae, R., and Karst, A. Temperature dependence of the activity of dna-modifying enzymes: endonucleases and dna ligase. *Eur J Biochem*, 123(1):141–52, 1982. doi: 10.1111/j.1432-1033.1982.tb06510.x.

Protozanova, E., Yakovchuk, P., and Frank-Kamenetskii, M. D. Stacked-unstacked equilibrium at the nick site of dna. *J Mol Biol*, 342(3):775–85, 2004. doi: 10.1016/j.jmb.2004.07.075.

Pryor, J. M., Potapov, V., Kucera, R. B., Bilotti, K., Cantor, E. J., and Lohman, G. J. S. Enabling one-pot golden gate assemblies of unprecedented complexity using data-optimized assembly design. *PLoS One*, 15(9):e0238592, 2020. doi: 10.1371/journal.pone.0238592.

Pryor, J. M., Potapov, V., Bilotti, K., Pokhrel, N., and Lohman, G. J. S. Rapid 40 kb genome construction from 52 parts through data-optimized assembly design. *ACS Synth Biol*, 11(6):2036–2042, 2022. doi: 10.1021/acssynbio.1c00525.

Revyakin, A., Ebright, R. H., and Strick, T. R. Single-molecule dna nanomanipulation: improved resolution through use of shorter dna fragments. *Nat Methods*, 2(2):127–38, 2005. doi: 10.1038/nmeth0205-127.

SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A*, 95(4):1460–5, 1998. doi: 10.1073/pnas.95.4.1460.

Szybalski, W., Kim, S. C., Hasan, N., and Podhajska, A. J. Class-iis restriction enzymes–a review. *Gene*, 100:13–26, 1991. doi: 10.1016/0378-1119(91)90345-c.

Tong, Y., Zhou, J., Zhang, L., and Xu, P. A golden-gate based cloning toolkit to build violacein pathway libraries in yarrowia lipolytica. *ACS Synth Biol*, 10(1):115–124, 2021. doi: 10.1021/acssynbio.0c00469.

Weiss, B., Jacquemin-Sablon, A., Live, T. R., Fareed, G. C., and Richardson, C. C. Enzymatic breakage and joining of

deoxyribonucleic acid. vi. further purification and properties of polynucleotide ligase from escherichia coli infected with bacteriophage t4. *J Biol Chem*, 243(17):4543–55, 1968.

Wu, D. Y. and Wallace, R. B. Specificity of the nick-closing activity of bacteriophage t4 dna ligase. *Gene*, 76(2):245–54, 1989. doi: 10.1016/0378-1119(89)90165-0.

Yazawa, K. and Furusawa, H. Probing multiple binding modes of dna hybridization: A comparison between single-molecule observations and ensemble measurements. *ACS Omega*, 3(2):2084–2092, 2018. doi: 10.1021/acsomega.8b00135.

Zacharias, M. Base-pairing and base-stacking contributions to double-stranded dna formation. *J Phys Chem B*, 124(46): 10345–10352, 2020. doi: 10.1021/acs.jpcb.0c07670.

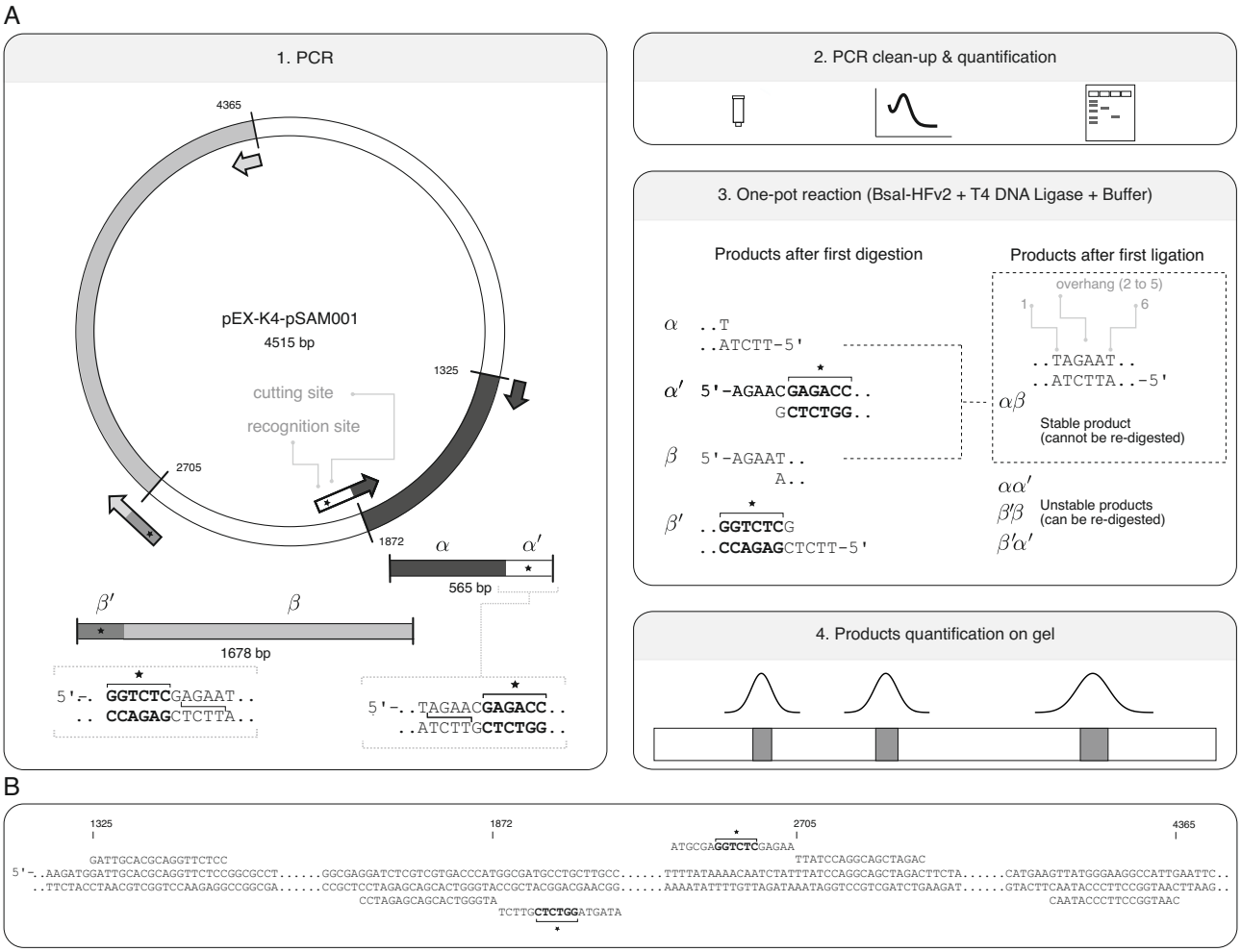


**Figure 1.** Experimental flow for the Golden Gate Assembly (GGA) assay. A. First, two PCR reactions are performed on the pEX-K4-pSAM001 plasmid (4515 bp) using primers that introduce both the recognition site of the type IIS restriction endonuclease BsaI-HFv2 (GGTGTC, asterisk) and specific sequences located downstream the recognition site (step 1). After purification of PCR products, determination of DNA concentration and visualization of DNA fragments by electrophoresis (step 2), a one-pot reaction is performed (step 3). Here, both PCR fragments ($\alpha\alpha'$, 565 bp and $\beta'\beta$, 1678 bp) are mixed at a given molarity and stoichiometry together with BsaI-HFv2 and T4 DNA ligase. Among the fragments produced by BsaI-HFv2, only those not containing BsaI-HFv2's recognition sites yield a stable product after ligation. Shown here are nucleotides for the TAGAAT sequence, which is a sequence of six nucleotides containing a non palindromic 4 bp overhang at position 2. Note that nucleotides at positions 1 and 6 may also play a role as they govern stacking energies. Here, and except for the GCCGCC sequence, nucleotides at positions 1 and 6 vary among possible ligation products $\alpha\beta$, $\alpha\alpha'$, $\beta'\beta$ and $\beta'\alpha'$. To determine the percentage of product formed $\Psi_\beta$ (defined as $100 \times [\alpha\beta]/([\beta'\beta] + [\alpha\beta])$), we perform a quantitative and robust analysis on agarose gels (see Supplementary Figure S2). B. DNA sequences showing the local binding of primers on the plasmid (TAGAAT sequence found in the $\alpha\beta$ product).

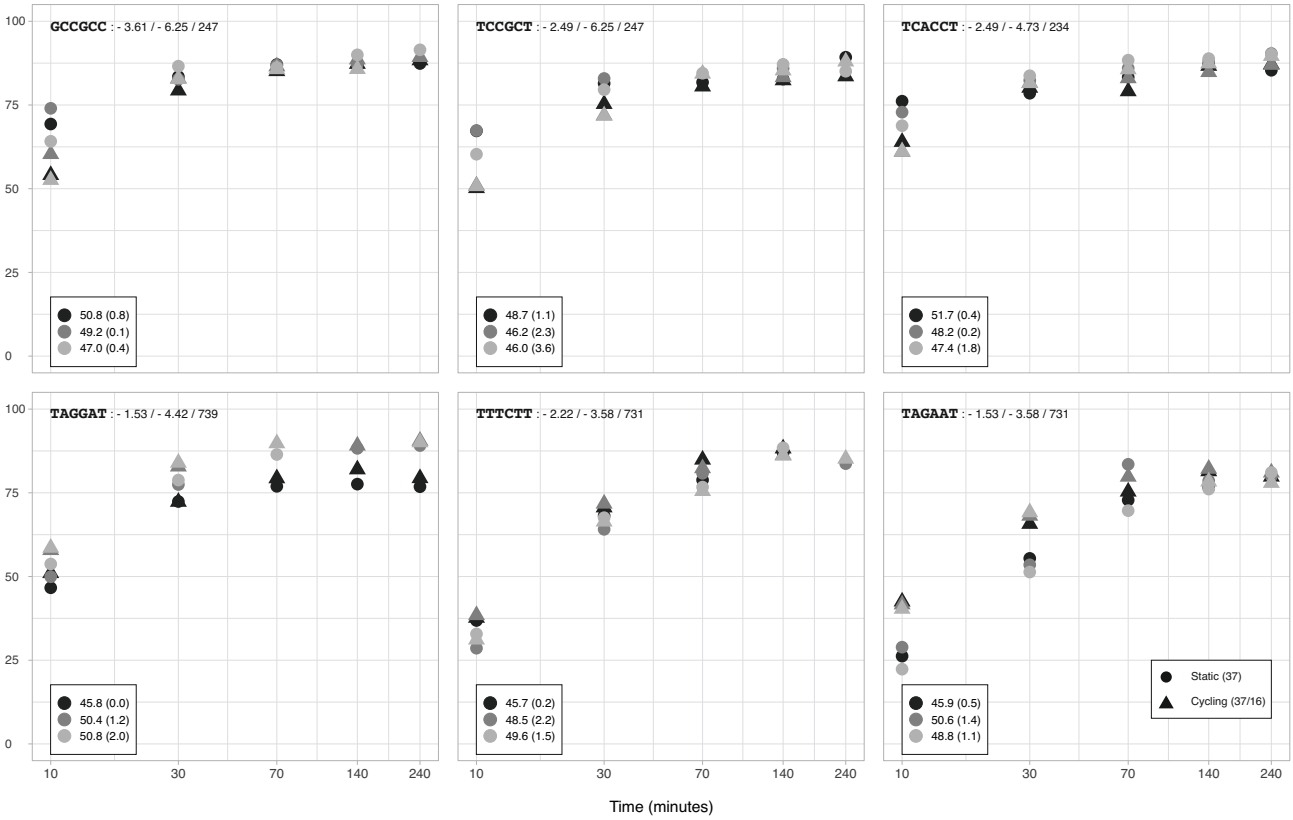


**Figure 2.** The percentage of product formed $\Psi_\beta$ deduced from gel assays ($[\alpha\alpha']$ = 21 nM, $[\beta'\beta]$ = 7 nM). Experiments are performed in triplicates for different sequences as well as different GGA assays conditions (static, circles and cycling, upper triangles at 10, 30, 70, 140 and 240 minutes). For every sequence, stacking energies and overhang stabilities (in kcal/mol) are given (e.g. -3.61 and -6.25 for GCCGCC, respectively). Also shown are ligation efficiencies (e.g. 247 for CCGC) deduced from sequencing-based assays (Pryor et al., 2020). In the insets are given $S_\beta$ values, which report on the stoichiometry (the value in bracket is the corresponding error). Here, the expected value of $S_\beta$ is 49.7 (see Materials and Methods).

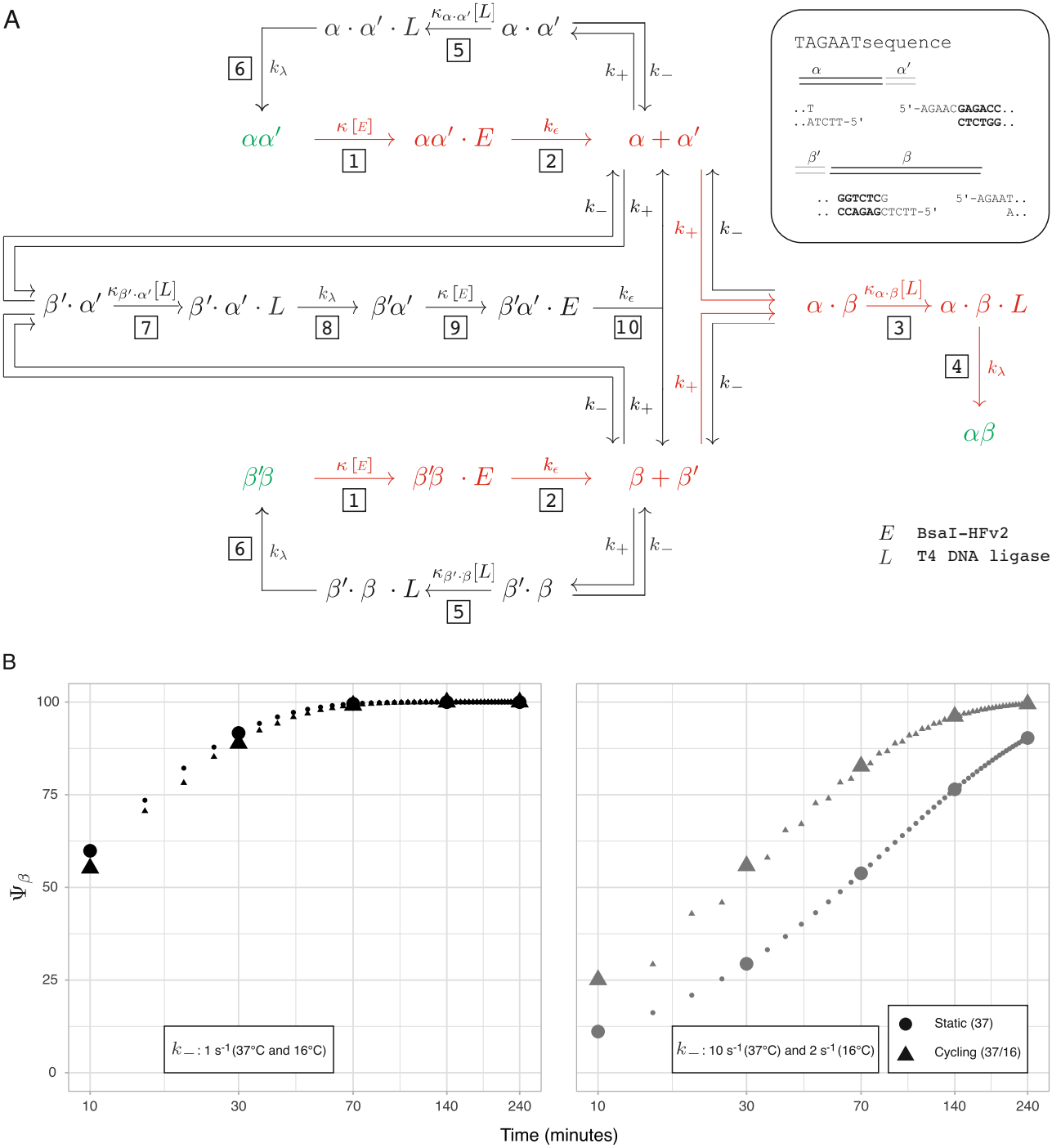


**Figure 3.** Kinetic model for our GGA assays. A. Initial reactants ($\alpha\alpha'$ and $\beta'\beta$) as well as the final (stable) product ($\alpha\beta$) are highlighted in green. Red arrows show the most direct route to form $\alpha\beta$. Enzymatic reactions are considered irreversible and are assumed to follow a two-steps Michaelis Menten process involving the binding of the enzyme to the substrate (represented by a centered dot) and a catalytic step ($k_\epsilon$, $k_\lambda$). To account for a possible difference in stacking energy among the different nicked DNA molecules ($\alpha \cdot \beta$, $\alpha \cdot \alpha'$, $\beta' \cdot \beta$ and $\beta' \cdot \alpha'$), different second order-reaction rates have been introduced (e.g. $\kappa_{\alpha \cdot \beta}$). The concentration of enzymes is supposed constant so that $\kappa[E]$, $\kappa_{\alpha \cdot \beta}[L]$, $\kappa_{\alpha \cdot \alpha'}[L]$, $\kappa_{\beta' \cdot \beta}[L]$ and $\kappa_{\beta' \cdot \alpha'}[L]$ can be considered as first-order rates. Boxes with text numbers are used to identify the different reaction steps, additional details are given in the main text. B. Corresponding calculations (time dependence of $\Psi_\beta$) for a static (37°C) and a cycling assays (37°C for 5 minutes and then 16°C for 5 minutes). Large symbols indicate times at which the concentration of products was experimentally determined (10, 30, 70, 140 and 240 minutes). For a stable overhang (e.g. `TCCGCT` sequence), we set $k_-$ at 1 s$^{-1}$ (at both 37°C and 16°C); for a weaker overhang (e.g. `TAGAAT` sequence), we set $k_-$ at 10 s$^{-1}$ (37°C) and $k_-$ at 2 s$^{-1}$ (16°C). All other parameters are identical : $\kappa[E]$, $\kappa_{\alpha \cdot \beta}[L]$, $\kappa_{\alpha \cdot \alpha'}[L]$, $\kappa_{\beta' \cdot \beta}[L]$ and $\kappa_{\beta' \cdot \alpha'}[L]$ equal to 0.1 s$^{-1}$, $k_\epsilon^{37}$ = 1 s$^{-1}$, $k_\epsilon^{16}$ = 0 s$^{-1}$, $k_\lambda^{37}$ = 1 s$^{-1}$, $k_\lambda^{16}$ = 0.4 s$^{-1}$ and $k_+$ = 10$^6$ M$^{-1}$s$^{-1}$.

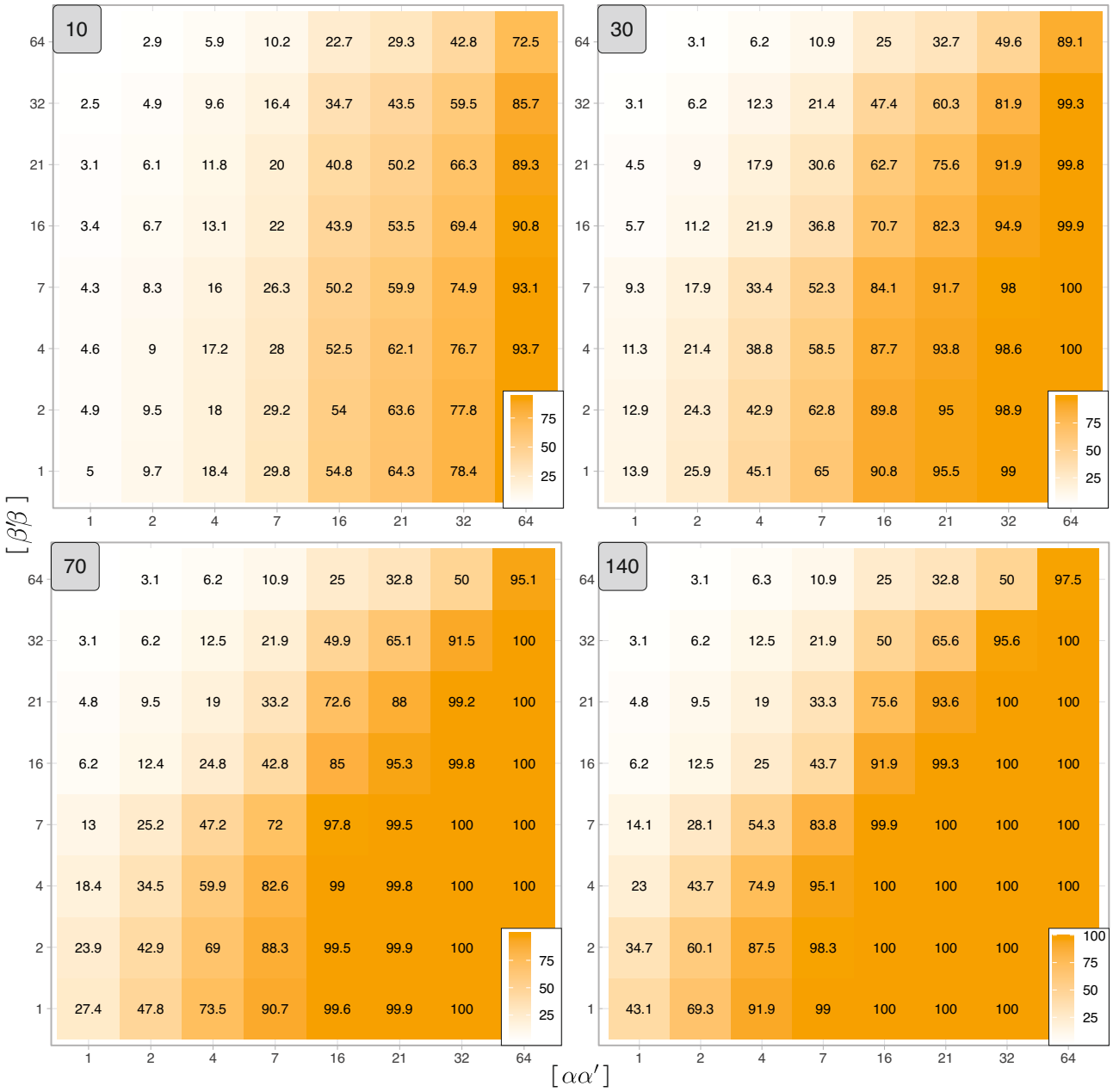


**Figure 4.** Calculations using our kinetic model for a static (37°C) assay and a strong overhang (*i.e.* $k_{-}$ = 1 s$^{-1}$). Values denote the percentage of product formed $\Psi_\beta$ at different times (10, 30, 70 and 140 minutes) and different concentrations of $\alpha\alpha'$ and $\beta'\beta$ (expressed in nM). Parameters identical to those of Figure 3.

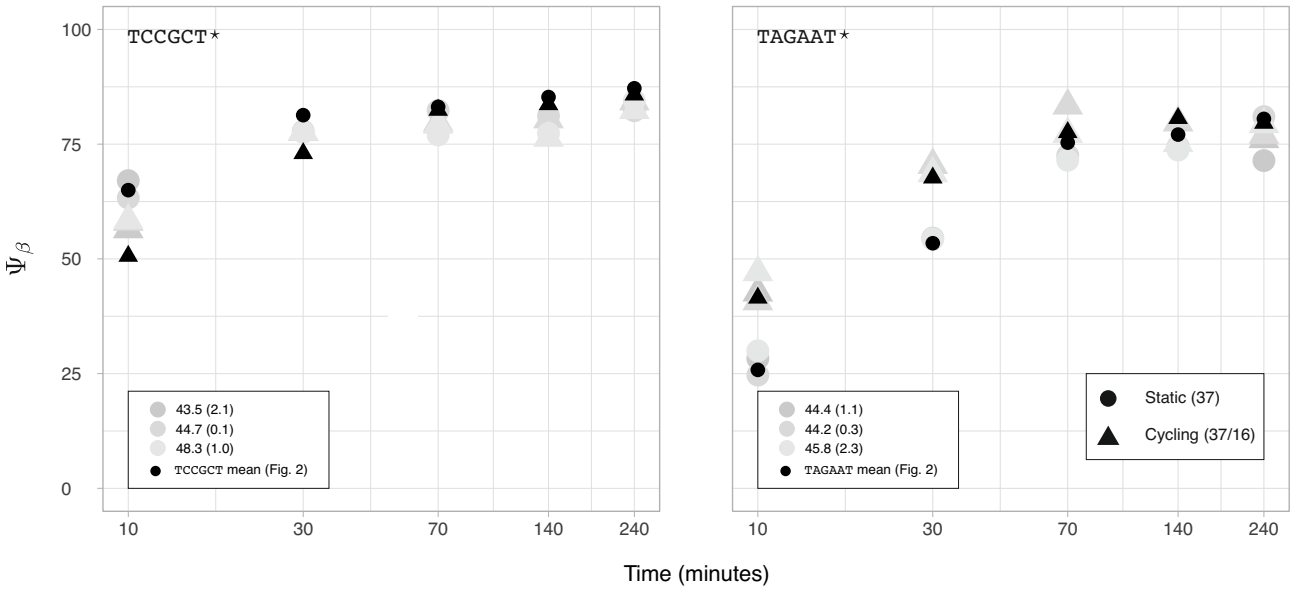


**Figure 5.** Golden-gate assembly of DNA fragments for Magnetic Tweezers. Experiments for TCCGCT$^\star$ and TAGAAT$^\star$ sequences, where nucleotides in positions 1 and 6 are conserved through possible intermediates ($\alpha \cdot \beta$, $\alpha \cdot \alpha'$, $\beta' \cdot \beta$ and $\beta' \cdot \alpha'$, see Figure 1). Experiments are performed in triplicates (grey circles: static assay, upper grey triangles : cycling assay). Black symbols are experimental averaged $\Psi_\beta$ values for TCCGCT and TAGAAT (Figure 2), *i.e.* sequences in which stacking energies are not conserved. Finally and as in Figure 2, the ratio $S_\beta$ is also given (the value in bracket is the corresponding error).

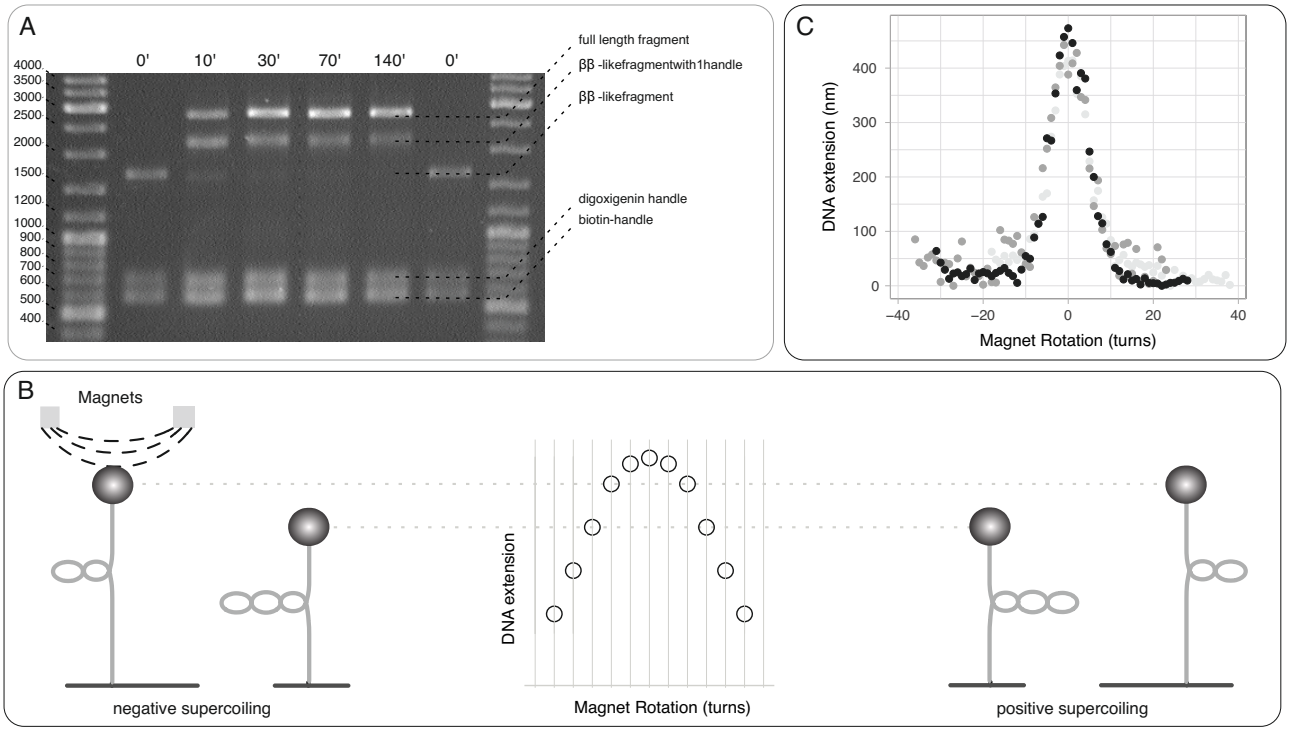


**Figure 6.** A. The results of GGA experiments using DNA handles ($\alpha\alpha'$-like fragments) labelled with biotin and digoxigenin (521 bp and 565 bp). The main fragment ($\beta'\beta$-like fragment, 1695 bp) is a PCR product (similar to that of Figure 1) where `TCCGCT` and `TCACCT` sequences have been introduced using appropriate primers (see Supplementary Text) and can specifically ligate to the biotin- and digoxigenin-handles (where compatible overhangs have been introduced). Raw gel images can be found in Supplementary Figure S9. B. Magnetic Tweezers experiments. Under torsional constraint (by imposing a torque and having attachment points on both strands of both handles), a DNA molecule will release stress by forming specific structures (plectoneme conformations) that result in a decrease in the the magnetic bead height. By varying the degree of supercoiling (the number of applied turns with the magnets), typical curves (so-called hat curves) can be obtained. C. Typical experimental hat curves measured on the final GGA fragment at 0.57 pN.

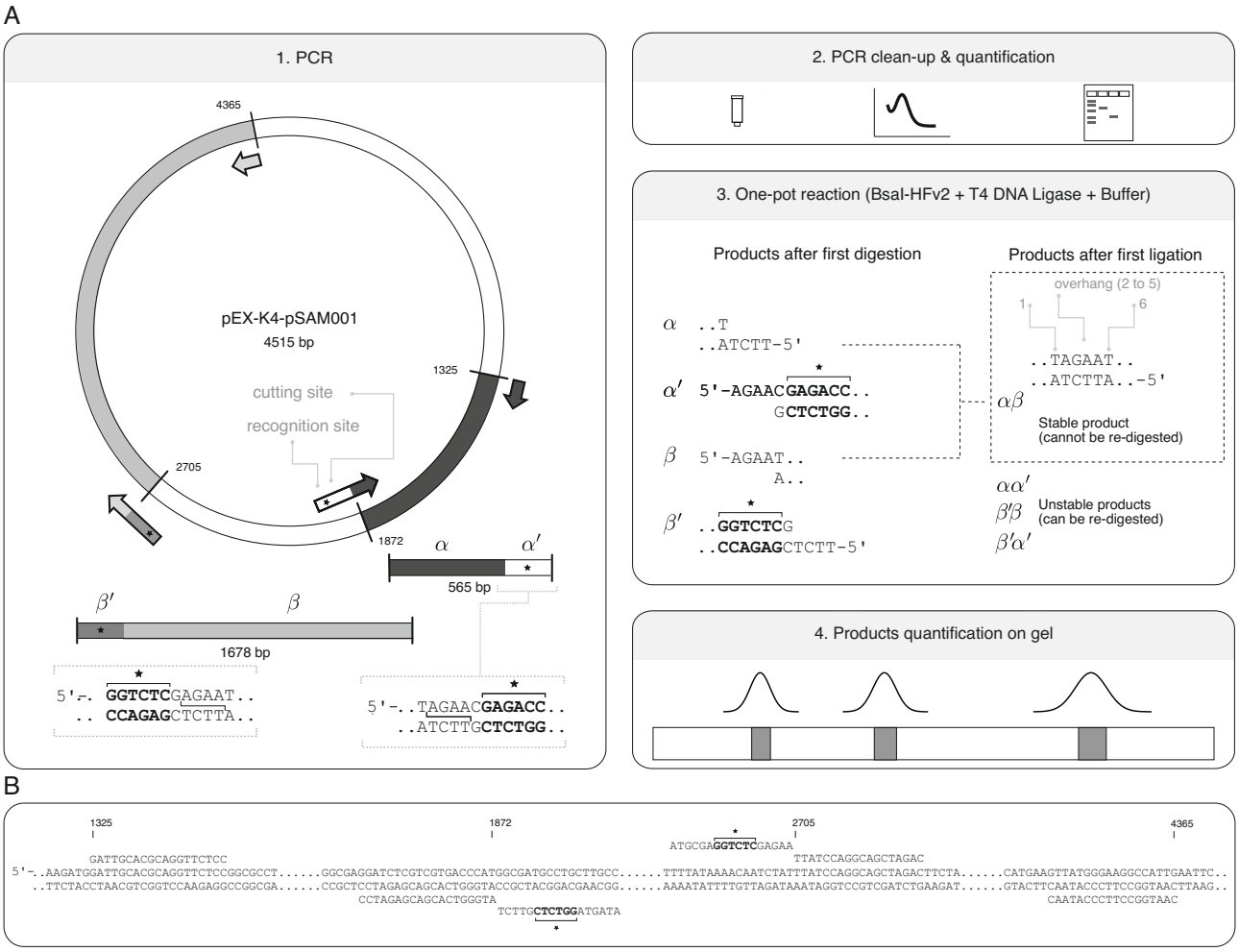


**Figure 1.** Experimental flow for the Golden Gate Assembly (GGA) assay. A. First, two PCR reactions are performed on the pEX-K4-pSAM001 plasmid (4515 bp) using primers that introduce both the recognition site of the type IIS restriction endonuclease BsaI-HFv2 (GGTGTC, asterisk) and specific sequences located downstream the recognition site (step 1). After purification of PCR products, determination of DNA concentration and visualization of DNA fragments by electrophoresis (step 2), a one-pot reaction is performed (step 3). Here, both PCR fragments ($\alpha\alpha'$, 565 bp and $\beta'\beta$, 1678 bp) are mixed at a given molarity and stoichiometry together with BsaI-HFv2 and T4 DNA ligase. Among the fragments produced by BsaI-HFv2, only those not containing BsaI-HFv2's recognition sites yield a stable product after ligation. Shown here are nucleotides for the TAGAAT sequence, which is a sequence of six nucleotides containing a non palindromic 4 bp overhang at position 2. Note that nucleotides at positions 1 and 6 may also play a role as they govern stacking energies. Here, and except for the GCCGCC sequence, nucleotides at positions 1 and 6 vary among possible ligation products $\alpha\beta$, $\alpha\alpha'$, $\beta'\beta$ and $\beta'\alpha'$. To determine the percentage of product formed $\Psi_\beta$ (defined as $100 \times [\alpha\beta]/([\beta'\beta] + [\alpha\beta])$), we perform a quantitative and robust analysis on agarose gels (see Supplementary Figure S2). B. DNA sequences showing the local binding of primers on the plasmid (TAGAAT sequence found in the $\alpha\beta$ product).

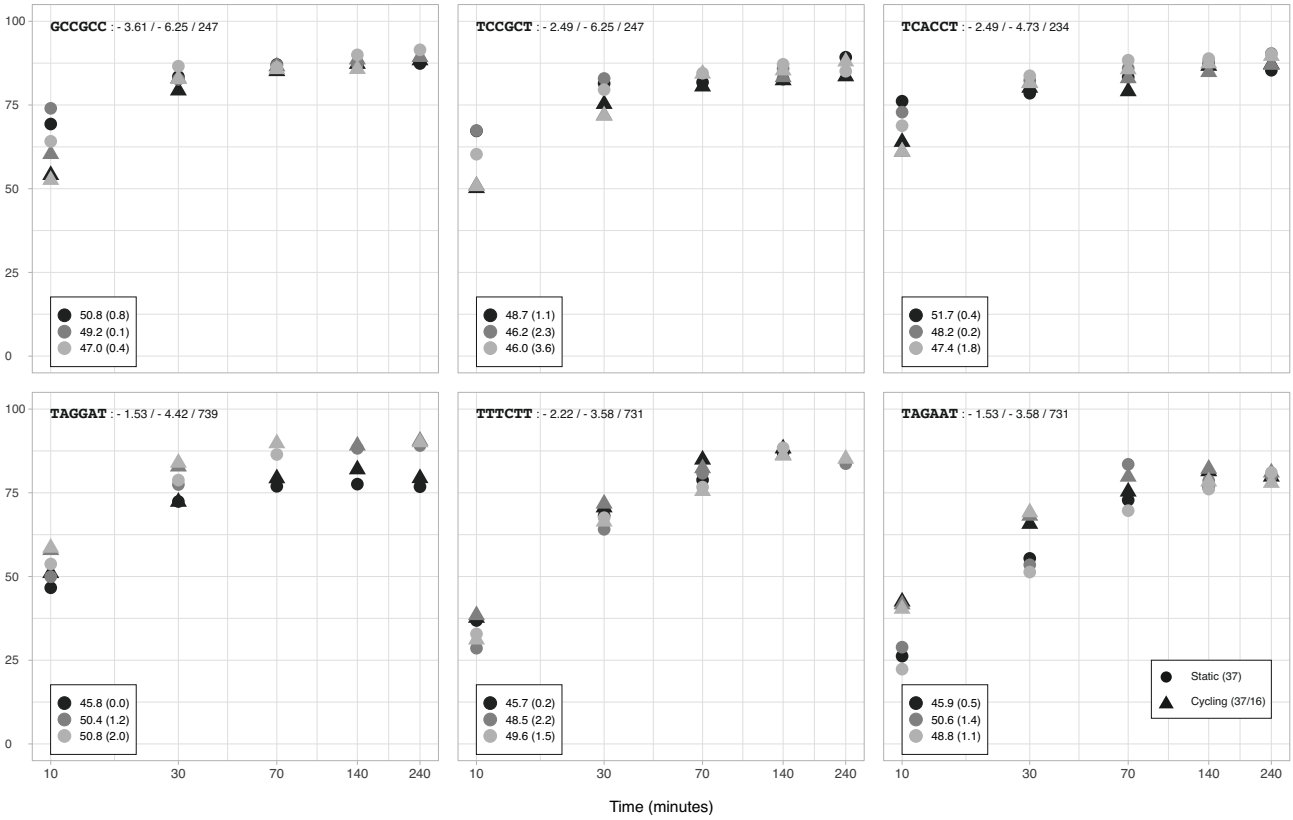


**Figure 2.** The percentage of product formed $\Psi_\beta$ deduced from gel assays ($[\alpha\alpha']$ = 21 nM, $[\beta'\beta]$ = 7 nM). Experiments are performed in triplicates for different sequences as well as different GGA assays conditions (static, circles and cycling, upper triangles at 10, 30, 70, 140 and 240 minutes). For every sequence, stacking energies and overhang stabilities (in kcal/mol) are given (e.g. -3.61 and -6.25 for GCCGCC, respectively). Also shown are ligation efficiencies (e.g. 247 for CCGC) deduced from sequencing-based assays (Pryor et al., 2020). In the insets are given $S_\beta$ values, which report on the stoichiometry (the value in bracket is the corresponding error). Here, the expected value of $S_\beta$ is 49.7 (see Materials and Methods).

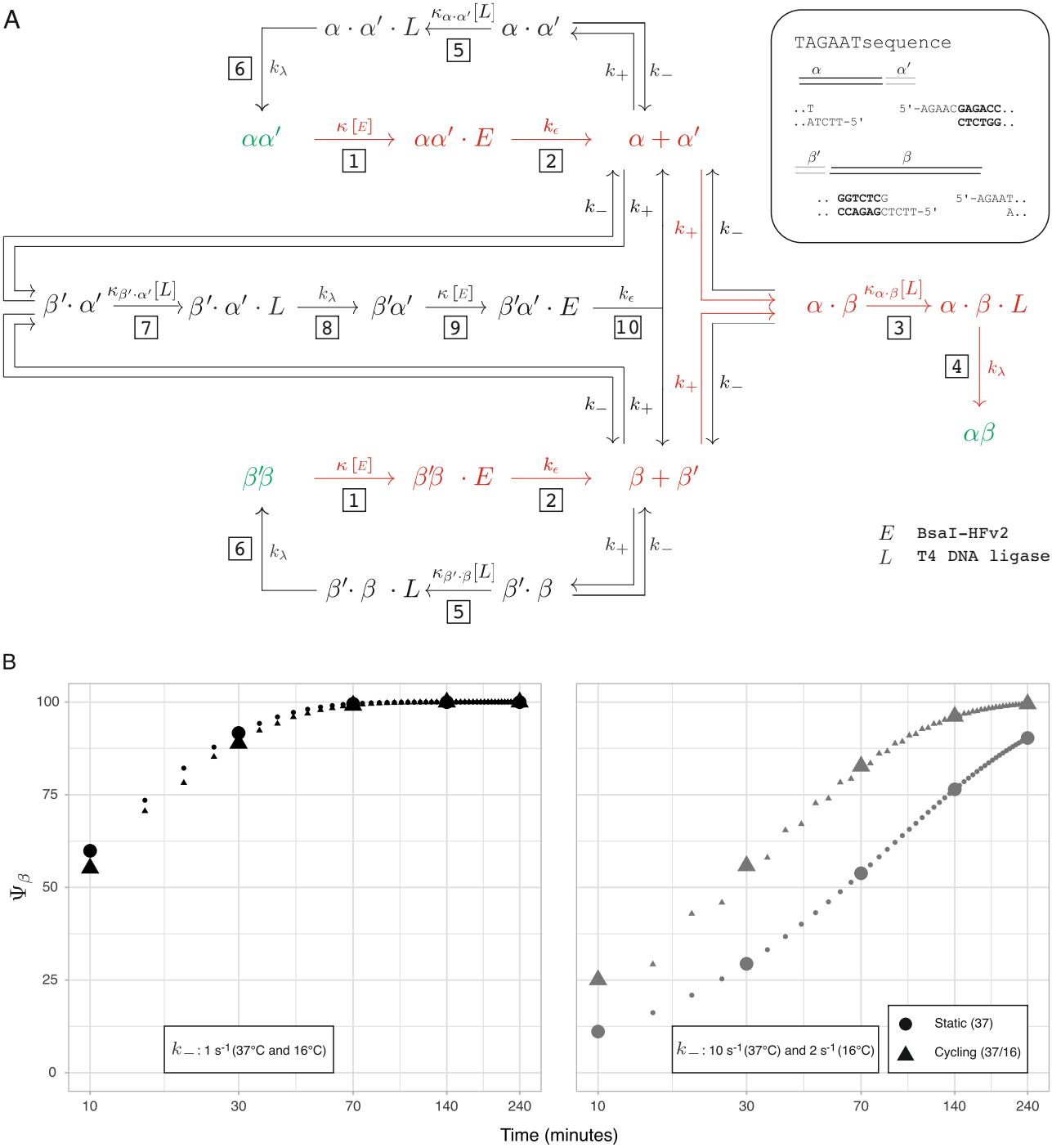


**Figure 3.** Kinetic model for our GGA assays. A. Initial reactants ($\alpha\alpha'$ and $\beta'\beta$) as well as the final (stable) product ($\alpha\beta$) are highlighted in green. Red arrows show the most direct route to form $\alpha\beta$. Enzymatic reactions are considered irreversible and are assumed to follow a two-steps Michaelis Menten process involving the binding of the enzyme to the substrate (represented by a centered dot) and a catalytic step ($k_\epsilon$, $k_\lambda$). To account for a possible difference in stacking energy among the different nicked DNA molecules ($\alpha \cdot \beta$, $\alpha \cdot \alpha'$, $\beta' \cdot \beta$ and $\beta' \cdot \alpha'$), different second order-reaction rates have been introduced (e.g. $\kappa_{\alpha \cdot \beta}$). The concentration of enzymes is supposed constant so that $\kappa[E]$, $\kappa_{\alpha \cdot \beta}[L]$, $\kappa_{\alpha \cdot \alpha'}[L]$, $\kappa_{\beta' \cdot \beta}[L]$ and $\kappa_{\beta' \cdot \alpha'}[L]$ can be considered as first-order rates. Boxes with text numbers are used to identify the different reaction steps, additional details are given in the main text. B. Corresponding calculations (time dependence of $\Psi_\beta$) for a static (37°C) and a cycling assays (37°C for 5 minutes and then 16°C for 5 minutes). Large symbols indicate times at which the concentration of products was experimentally determined (10, 30, 70, 140 and 240 minutes). For a stable overhang (e.g. `TCCGCT` sequence), we set $k_-$ at 1 s$^{-1}$ (at both 37°C and 16°C); for a weaker overhang (e.g. `TAGAAT` sequence), we set $k_-$ at 10 s$^{-1}$ (37°C) and $k_-$ at 2 s$^{-1}$ (16°C). All other parameters are identical : $\kappa[E]$, $\kappa_{\alpha \cdot \beta}[L]$, $\kappa_{\alpha \cdot \alpha'}[L]$, $\kappa_{\beta' \cdot \beta}[L]$ and $\kappa_{\beta' \cdot \alpha'}[L]$ equal to 0.1 s$^{-1}$, $k_\epsilon^{37}$ = 1 s$^{-1}$, $k_\epsilon^{16}$ = 0 s$^{-1}$, $k_\lambda^{37}$ = 1 s$^{-1}$, $k_\lambda^{16}$ = 0.4 s$^{-1}$ and $k_+$ = 10$^6$ M$^{-1}$s$^{-1}$.

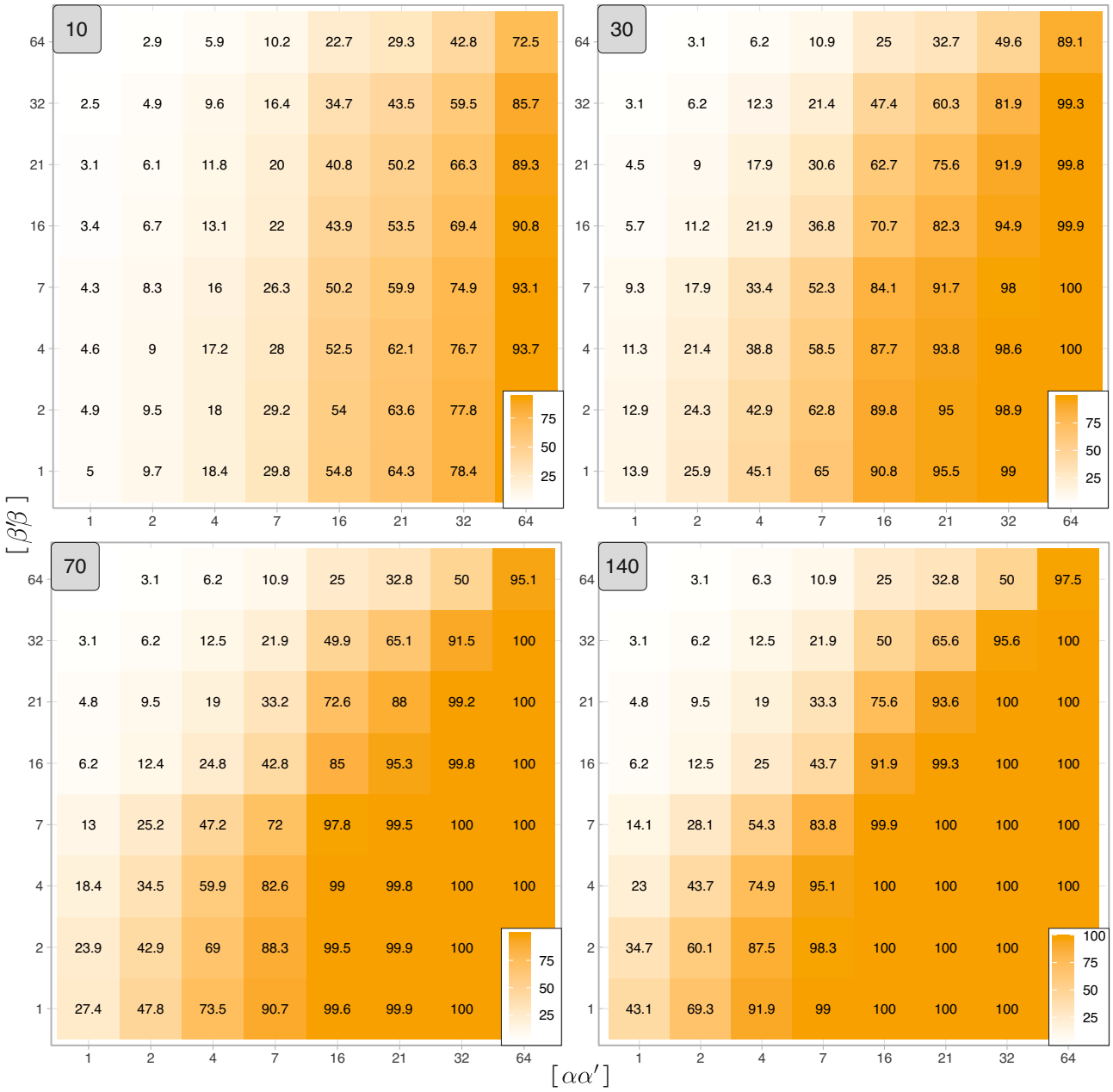


**Figure 4.** Calculations using our kinetic model for a static (37°C) assay and a strong overhang (*i.e.* $k_{-}$ = 1 s$^{-1}$). Values denote the percentage of product formed $\Psi_\beta$ at different times (10, 30, 70 and 140 minutes) and different concentrations of $\alpha\alpha'$ and $\beta'\beta$ (expressed in nM). Parameters identical to those of Figure 3.

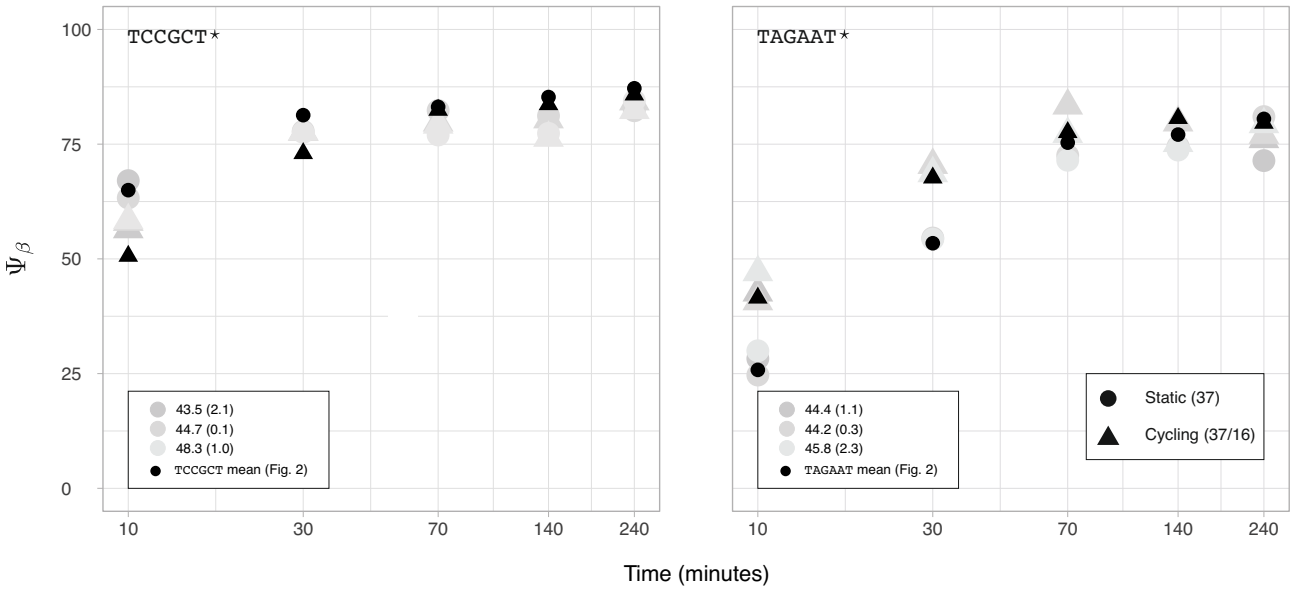


**Figure 5.** Golden-gate assembly of DNA fragments for Magnetic Tweezers. Experiments for TCCGCT$^\star$ and TAGAAT$^\star$ sequences, where nucleotides in positions 1 and 6 are conserved through possible intermediates ($\alpha \cdot \beta$, $\alpha \cdot \alpha'$, $\beta' \cdot \beta$ and $\beta' \cdot \alpha'$, see Figure 1). Experiments are performed in triplicates (grey circles: static assay, upper grey triangles : cycling assay). Black symbols are experimental averaged $\Psi_\beta$ values for TCCGCT and TAGAAT (Figure 2), *i.e.* sequences in which stacking energies are not conserved. Finally and as in Figure 2, the ratio $S_\beta$ is also given (the value in bracket is the corresponding error).

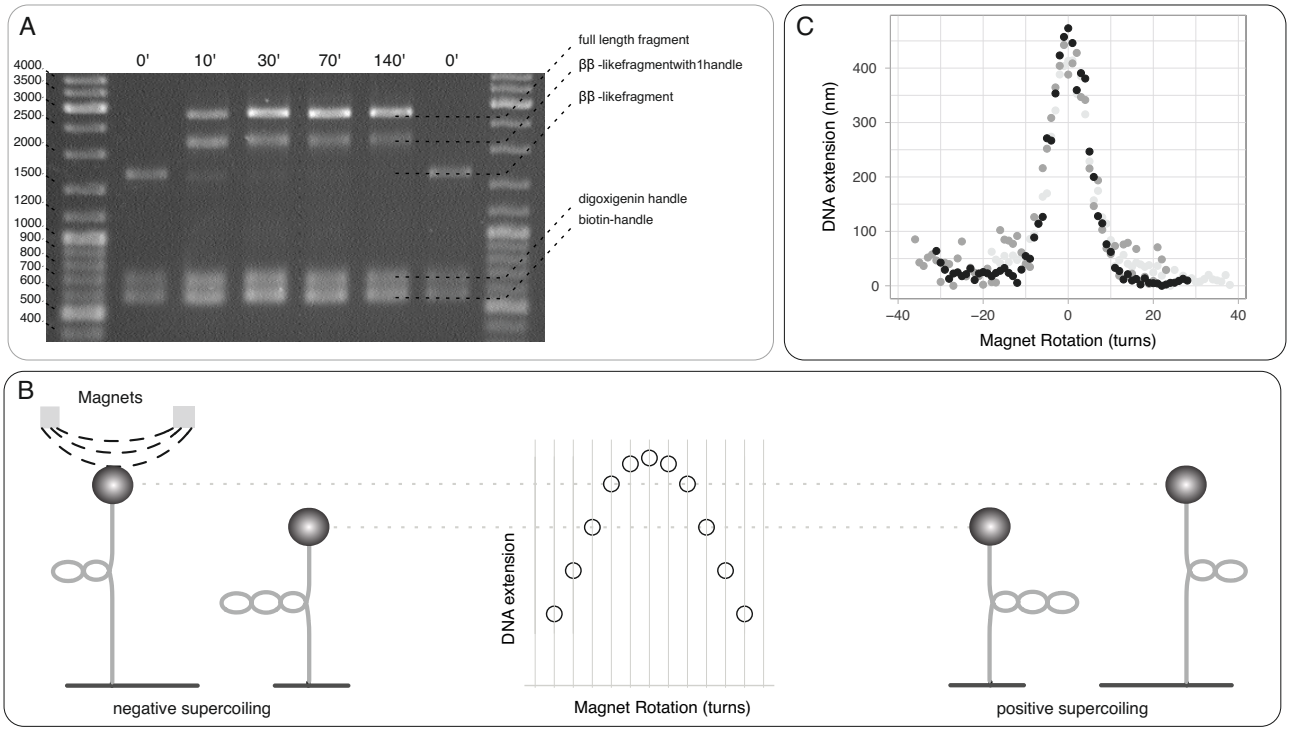


**Figure 6.** A. The results of GGA experiments using DNA handles ($\alpha\alpha'$-like fragments) labelled with biotin and digoxigenin (521 bp and 565 bp). The main fragment ($\beta'\beta$-like fragment, 1695 bp) is a PCR product (similar to that of Figure 1) where `TCCGCT` and `TCACCT` sequences have been introduced using appropriate primers (see Supplementary Text) and can specifically ligate to the biotin- and digoxigenin-handles (where compatible overhangs have been introduced). Raw gel images can be found in Supplementary Figure S9. B. Magnetic Tweezers experiments. Under torsional constraint (by imposing a torque and having attachment points on both strands of both handles), a DNA molecule will release stress by forming specific structures (plectoneme conformations) that result in a decrease in the the magnetic bead height. By varying the degree of supercoiling (the number of applied turns with the magnets), typical curves (so-called hat curves) can be obtained. C. Typical experimental hat curves measured on the final GGA fragment at 0.57 pN.