DNA Shuffling using NEBridge® Golden Gate Assembly for Protein Engineering

Sean Lund, Ph.D., Annika Goldstein, Sean Johnson, Ph.D. and Matthew Norton, Ph.D., New England Biolabs

INTRODUCTION

DNA shuffling is a widely adopted method for recombining sequences to generate chimeric proteins with new or improved properties, such as thermotolerance, tolerance to harsh chemicals, or higher catalytic efficiency. Homologous recombination is typically used with highly similar sequences to scramble genes to make new proteins. These methods have been used to engineer beta-lactamases, P450s, and many more enzymes (1,2).

Random recombination of highly similar sequences is accomplished through random fragmentation of templates by DNase I followed by reconstruction of full-length genes using overlap extension PCR (3). While robust, these DNA shuffling methods require sequences to be highly homologous so that fragments of different genes can anneal to one another during cycles of PCR. The highly similar sequences are randomly recombined using overlapping homology. While this expands the accessible search area within the protein fitness landscape, it also generates libraries that may be difficult to sample effectively.

Modern protein engineering has pivoted from screening large libraries to constructing more thoughtful and purposely designed libraries. One method of constructing more targeted libraries is to use structure-guided designs in which proteins are broken down into functional units or domains that can then be targeted for protein recombination (4).

NEBridge Golden Gate Assembly allows for one-pot, scarless DNA assembly of multiple fragments in an ordered, modular fashion. While the library design determines the number of fragments, complex assemblies between 2–50+ fragments are supported with high efficiency and accuracy (5). Golden Gate Assembly allows for shuffling DNA with very low homology (6) using as little as 3 bp overlap. And, unlike random shuffling through DNase I treatment, users can specifically define fusion sites between different parts of a coding region. For each segment of a construct, multiple variant sequences can be added to the assembly to effectively scramble a region of a protein (Figure 1). For larger combinatorial libraries, all segments can be replaced with variant sequences allowing for the construction of larger, yet targeted libraries. Libraries can be constructed either as large pools of variants or, alternatively, parts can be selectively chosen by users or machine learning methods to purposefully construct desired mutants.

METHODS

Nine chromoprotein amino acid sequences (aeBlue, amajLime, amilCP, amilCP Orange, amilCP Pink, amilGFP, asPink, mRFP, and tsPurple) were all aligned using MUSCLE (Geneious). Stretches of two consecutive amino acids identical in the consensus were identified and five of these were chosen as break points for recombination. Amino acids were reverse translated, and DNA sequences for each fusion site were scanned for potential 4 bp overhangs by cross-referencing their compatibility using NEBridge Ligase Fidelity Viewer® (ligasefidelity.neb.com). Overhang sets were verified for 100% fidelity.

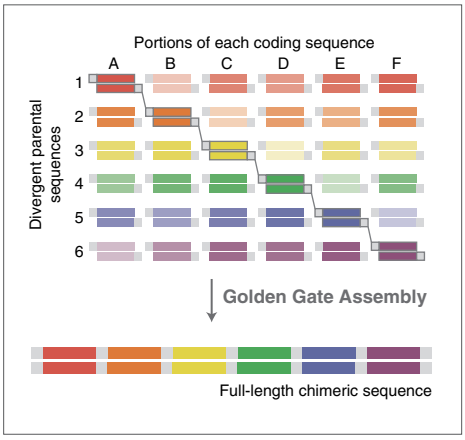
BsaI recognition sites, and Golden Gate overhangs where required, were appended to the six DNA fragment sequence parts to be properly constructed in a BsaI-domesticated pET28a backbone, for a seven-part total assembly. To ensure all fragments were 300 bp long, additional randomized sequences free of BsaI sites were then appended to each, and the 300 bp fragments were then ordered as eBlocks® (Integrated DNA Technologies®).

For control assemblies, parts for the parental (wild-type) sequences were assembled (20 ng each; 300 bp) in a domesticated pET28a vector (47 ng; 5,871 bp) using NEBridge Golden Gate Assembly Kit (BsaI-HFv2) (NEB #E1601). A 20 µL reaction was performed using 30 cycles of 37°C for 1 minute and 16°C for 1 minute before a final hold of 5 minutes at 60°C. 2 µl of each assembly was then transformed into T7 Express Competent *E. coli* (NEB #C2566). Cells were then plated on LB agar with kanamycin and grown at 37°C overnight. Once colonies were observed, plates were removed from the incubator and allowed to develop color on the benchtop.

MATERIALS

- NEBridge Golden Gate Assembly Kit (BsaI-HF®v2) (NEB #E1601)
- T7 Express Competent *E. coli* (High Efficiency) (NEB #C2566)
- Q5® Hot Start High-Fidelity 2X Master Mix (NEB #M0494)
- phi29-XT RCA Kit (NEB #E1603)
- NEBExpress® *E. coli* Lysis Reagent (NEB #P8116)

FIGURE 1: **Protein recombination using NEBridge Golden Gate Assembly DNA shuffling.**



Golden Gate Assembly can be used in a pooled fashion to generate novel combinations of different DNA parts by designing parts with compatible overhangs. Overhangs assign a sequence a position in an assembly allowing for multiple combinations to be accessed.

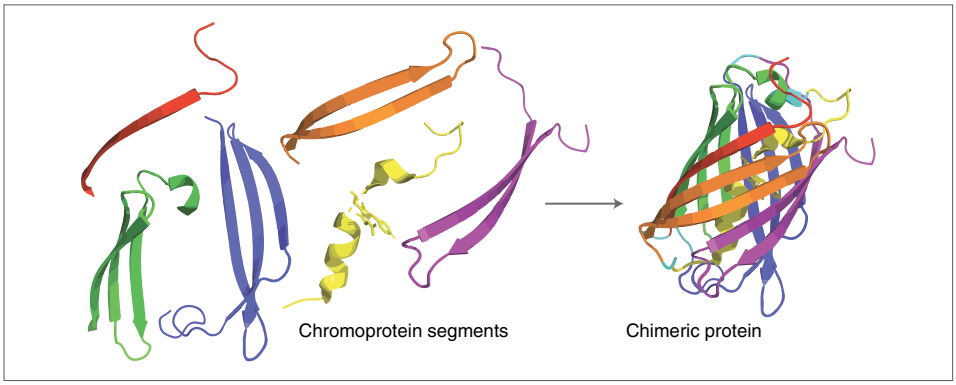
For the pooled assembly, 10 ng of each unique part (43 total), were mixed and 430 ng was added to a 60 µl NEBridge Golden Gate Assembly Kit (BsaI-HFv2) reaction with predigested vector (208 ng; 5,871 bp). Cycling conditions were identical to those used for the control assemblies. 2 µL of assembly was then transformed into T7 Express Competent *E. coli* and serial dilutions were plated onto LB agar plates with kanamycin and grown overnight at 37°C.

Pigmented colonies from the pooled assembly were picked and amplified using the phi29-XT RCA Kit (NEB #E1603). The resulting amplicons were then diluted 20-fold and Sanger sequenced. Parts were then mapped to the Sanger reads to identify which parts were contained in the assembly.

Once known, the new sequences were reconstructed by Golden Gate Assembly, as above, and again transformed. Colonies were then scraped using a pipette tip and lysed in 300 µl NEBExpress *E. coli* Lysis Reagent (NEB #P8116), spun down, and the supernatant was examined by UV-Vis spectroscopy. The resulting absorbance spectra were analyzed to determine if they represented a novel chromoprotein color.

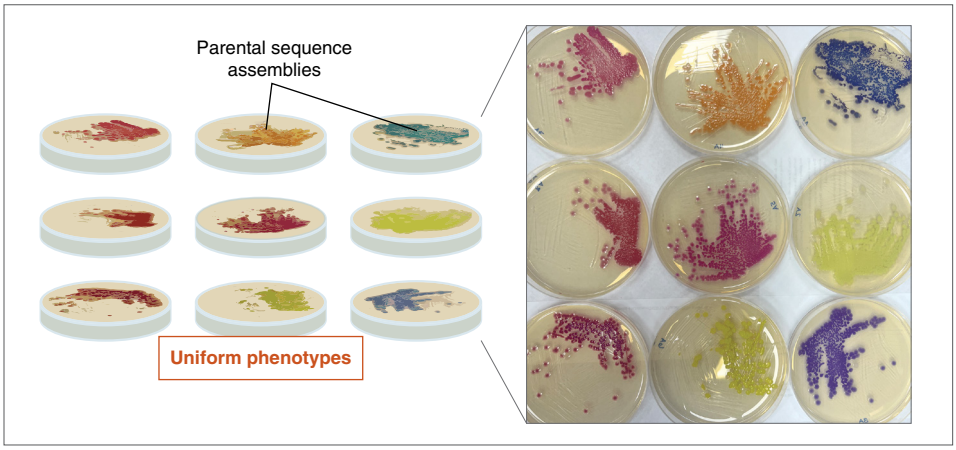
The combined fragments' sequences and resulting absorbance spectra were used to develop a computational model to predict further combinations of fragments that would result in additional novel colors. Gene fragments were reordered with adapters (Twist Bioscience®, South San Francisco, CA) for amplification and PCR amplified in 50 µl reactions using Q5® Hot Start High-Fidelity 2X Master Mix (NEB #M0494), cleaned using SPRI® beads, and eluted in 100 µl water. Using an Opentrons OT-2, a master mix containing destination vector and components for 15 µl NEBridge Golden Gate Assembly Kit (BsaI-HFv2) reactions were assembled on a 4°C temperature module then mixed off the deck by vortexing. The liquid handler then distributed the master mix across a 96-well plate without temperature control. Using the OT-2, 6 parts per assembly were then pipetted over the course of more than 3 hours (576 parts total). The plate was then sealed and subjected to 30 cycles of 37°C for 1 minute and 16°C for 1 minute before a final hold of 5 minutes at 60°C. 2 µl was then transformed into 20 µl of T7 Express Competent *E. coli* following the standard high-efficiency transformation protocol. 5 µL of diluted or concentrated transformation were then plated onto LB Kan and grown overnight at 37°C. After colonies grew, they were removed from the incubator and allowed to develop color on the benchtop overnight and then in a 4°C fridge.

FIGURE 2: **Structural chimeragenesis of chromoproteins**



Amino acid sequence alignments were performed for nine chromoproteins to determine structural break points. Elements from different proteins can then be recombined to generate new chimeras with different properties. Six structural elements from six different chromoproteins are shown here as an example.

FIGURE 3: **Constructing parental sequences from six separate parts**



Colonies from parental sequence assemblies show uniform phenotypes verifying reliable and accurate sequence construction. Greater than 95% of colonies showed the corresponding-colored phenotype.

RESULTS

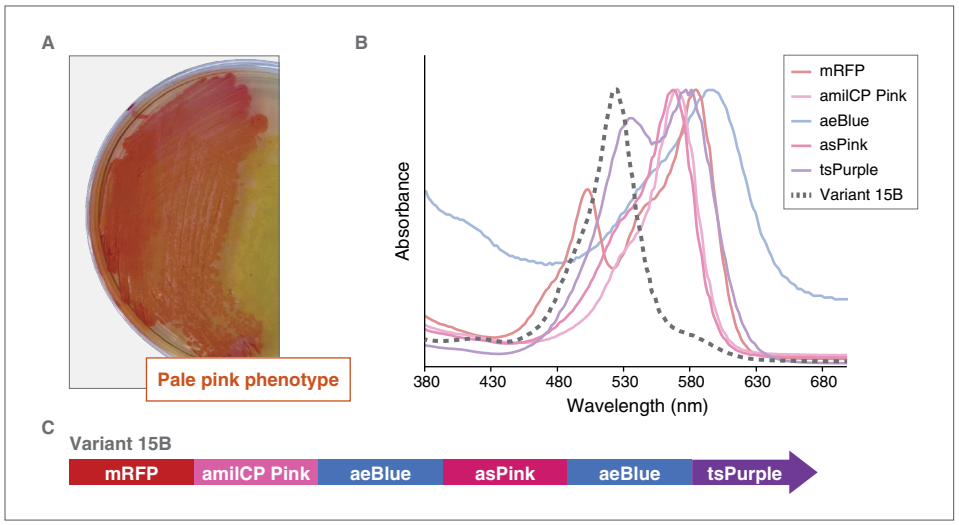
As a demonstration of the ability to shuffle sequences using Golden Gate Assembly, we attempted to make functional chimeric proteins from known chromoproteins (Figure 2). Chromoprotein sequences with as little as 34.4% amino acid identity were aligned and split into 6 separate parts where two amino acids were found to be conserved among the nine sequences. These low identity sequences would not be amenable to using traditional DNase I-based shuffling approaches. An iterative process was used in which 5 sets of two conserved amino acids between chromoprotein sequences were reverse translated into codons and NEBridge Ligase Fidelity Viewer was consulted to find optimal overhangs for an assembly. Fragments were then padded with random sequence to meet the minimum sequence length requirements for ordering.

Parental sequences were first constructed as a control to verify efficiency and accuracy of the 6-part DNA fragment assemblies. Plating of the transformations showed high assembly fidelity in that almost every colony from an assembly had the same phenotype (Figure 3). This gave high confidence that the vast majority of transformants from the pooled assembly would have all 6-parts needed to properly assemble a functional chromophore.

Next, we performed a pooled assembly of a library of variants and plated the reaction transformants. The theoretical library size was 129,654, so the odds of any given colony containing a parental sequence was negligible. Less than 1% of colonies generated had pigmentation and of the 39 colored colonies sequenced, none were wild-type sequences.

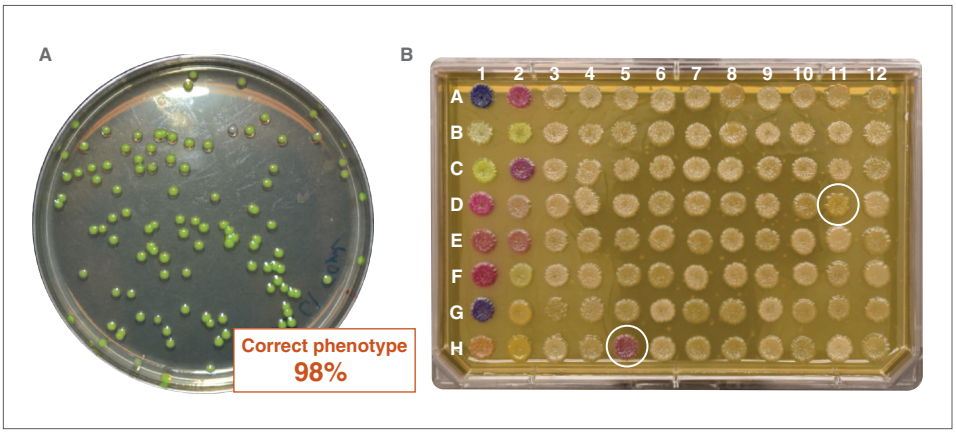
Once the parts making up the sequences were known, the sequences were reconstructed in defined Golden Gate Assemblies and plated to confirm the novel color phenotype (Figure 4A). The chromoproteins were then extracted alongside the parental proteins and assayed by UV-Vis spectroscopy to show that the new proteins in fact had novel absorbance properties relative to the parts from which they were made (Figure 4B and 4C).

FIGURE 4: **Novel variant resulting from protein recombination shows distinct absorbance properties relative to the component parts**



A) Variant 15B shows a pale pink phenotype. B) Variant 15B (dashes) shows a different absorbance profile relative to the parental chromoproteins that make up its parts. The absorbance maximum is shifted and shoulders are absent. C) Sequencing confirms that variant 15B is a chimeric protein comprised of 6 shuffled parts from 5 different proteins.

FIGURE 5: **Automated construction of 7-part Golden Gate Assemblies**



A) Automated construction of control chromoproteins from dsDNA fragments resulted in 98% of colonies displaying the correct phenotype. B) Construction of 96 different chromoprotein variants (9 parental sequences, 7 novel sequences shown previously to give color in the pooled library experiment, and 80 novel combinations not previously tested) showed homogeneity between spots giving confidence to the assembly method. Two new chromoproteins were found, H5 (purple) and D11 (orange).

Following the success of a pooled library construction, we reasoned that the sequences producing novel pigments could be used to generate a predictive model for new chromoproteins. A simple statistical model considering frequencies and co-occurrence of parts observed in successful clones from the previous experiment was made and 80 new designs were generated and constructed. Since a total of 96 seven-part assemblies were being constructed, we employed the use of a liquid handling robot to assemble a reaction mix and select the parts required for the construction. Reactions were set up at room temperature over the course of 3 hours before being transferred to a thermocycler to conduct the Golden Gate Assembly reaction. A portion of one control assembly was plated for individual colonies to verify that under these conditions, a high proportion of correctly assembled constructs could be obtained. We found greater than 98% of colonies displayed the desired pigment and obtained 200 CFU/uL of assembly (Figure 5A). Of the 80 new designs tested, 2 of the designs gave new pigments (Figure 5B).

CONCLUSION

NEBridge Golden Gate Assembly provides an accessible way to shuffle DNA to recombine proteins and find new properties. NEBridge Ligase Fidelity Viewer is an easy-to-use tool that allows users to identify compatible overhangs enabling reliable assembly of multiple parts. Using these tools, we showed that novel chimeric chromoproteins could be constructed from parts incompatible with traditional DNA shuffling approaches. The method demonstrated here can be applied to any enzyme of interest, as long as a suitable assay is available. Automation could be employed to set up assembly reactions using a liquid handler resulting in higher throughput, larger libraries, and investigation of additional space in a protein fitness landscape.

While the method demonstrated here is just one cycle in a design-build-test workflow, researchers could further iterate using NEBridge Golden Gate Assembly for additional rounds of DNA shuffling, NEBuilder® HiFi DNA Assembly for multi-site or combinatorial mutagenesis, or Q5 Site-Directed Mutagenesis for additional site-specific mutagenesis.

To learn more, visit www.neb.com/goldengate.

References

1. Meyer, M., et al. (2006) *Protein Eng Des Sel.* 19(12): 563–570.
2. Fasan, R., et al. (2007) *Angew Chem Int Ed Engl.,* 46(44): 8414–8418.
3. Zhao, H. and Arnold, F. H. (1997) *Nucleic Acids Res.* 25(6): 1307–1308.
4. Silberg, J. J., et al. (2004) *Methods Enzymol* 388: 35–42.
5. Pryor, J. M., et al. (2022) *ACS Synth Biol.* 11(6): 2036–2042.
6. Engler, C., et al. (2009) *PLoS One.* 4 (5), e5553.