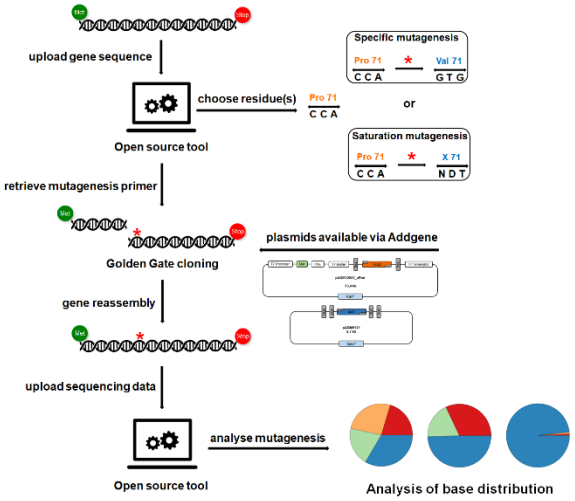
Golden Mutagenesis: An efficient multi-site saturation mutagenesis approach by Golden Gate cloning with automated primer design

Pascal Püllmann,[†,‡] Chris Ulpinnis,[†,‡] Sylvestre Marillonnet,[†] Ramona Gruetzner,[†] Steffen Neumann,[*,†] Martin J. Weissenborn[*,†,‡]

[†]Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle (Saale) and [‡] Institute of Chemistry, Martin-Luther-University Halle-Wittenberg, Kurt-Mothes-Str. 2, 06120 Halle (Saale), Germany

ABSTRACT: Site-directed methods for the generation of genetic diversity are essential tools in the field of directed enzyme evolution. The Golden Gate cloning technique has been proven to be an efficient tool for a variety of cloning setups. The utilization of restriction enzymes which cut outside of their recognition domain allows the assembly of multiple gene fragments obtained by PCR amplification without altering the open reading frame of the reconstituted gene. In this technical note, we developed a protocol termed Golden Mutagenesis for the rapid, easy, reliable and inexpensive construction of mutagenesis libraries. One to five positions within a coding sequence could be altered simultaneously using a protocol which can be performed within two days. To facilitate the implementation of this technique, a software library for automated primer design and for the graphical evaluation of the sequencing results has been developed, allowing an easy determination of the library quality.



Keywords: iterative saturation mutagenesis (ISM), primer design algorithm, sequencing analysis, quick quality control (QCC), site-directed mutagenesis

Directed evolution approaches require a highly efficient molecular cloning setup to simultaneously alter multiple residues in a rapid, reproducible and cost-effective manner. It is moreover desirable that utilized techniques avoid general bias in the created library and ideally create statistical nucleobase distributions. In recent years, targeted combinatorial approaches for directed evolution such as CASTing[1, 2] or iterative saturation mutagenesis (ISM)[3, 4] have evolved as tremendously successful and widely applied techniques for protein engineering using "smart" directed evolution. A general requirement within these approaches is the ability to saturate specific protein residues simultaneously.

Commonly employed techniques for single site-saturation mutagenesis include MOD-PCR (Mutagenic Oligonucleotide-Directed PCR Amplification)[5] Codon Cassette Mutagenesis,[6] Overlap Extension PCR,[7] Megaprimer PCR[8] and the commercial kit QuikChange®.[9] The successful generation of libraries of gene sequences containing multiple randomization sites, however, often requires more sophisticated methods to create high-quality libraries with a high success rate. Recent developments to address this issue include the introduction of Omnichange[10], Darwin Assembly[11], and ISOR.[12]

Golden Gate cloning has been introduced in 2008[13] and is based on the utilization of Type IIS restriction enzymes. This subclass of restriction enzymes is defined by their ability to cleave double-stranded DNA templates outside of their recognition site. A remarkable feature of the technique is that is allows to perform the restriction and ligation reaction in a time and cost saving one-pot reaction setup with an overall efficiency of correct assembly close to 100 %.[14, 15] An increasingly large number of Golden Gate compatible plasmids is furthermore accessible through the non-profit plasmid repository Addgene.

Arguably, the most valuable feature of the Golden Gate technique is that it allows highly specific assembly of several gene fragments mediated through the generation of 4 base pair overhangs that flank the fragments after restriction. 240 unique overhangs (4^4 – 16 palindromic sequences) can be employed for the joining of adjacent fragments, therefore enabling a broad applicability in the field of multiple site-directed mutagenesis. Correct assembly occurs in a scarless manner, i.e. the original open reading frame of the target gene is restored. Furthermore, the conducted blue/orange against white bacterial colony screening enables the distinction of negative events (no restriction/ligation) and hence saves screening effort, since it allows the exclusion of those events for subsequent testing. In contrast to other widespread mutagenesis techniques, amplification of the acceptor plasmid backbone via PCR is not necessary for Golden Gate cloning. Therefore the risk of introducing unwanted mutations within the plasmid backbone can be eliminated. However, thus far, only few examples of the use of Golden Gate cloning in the context of mutagenesis for directed evolution have been reported.[16]

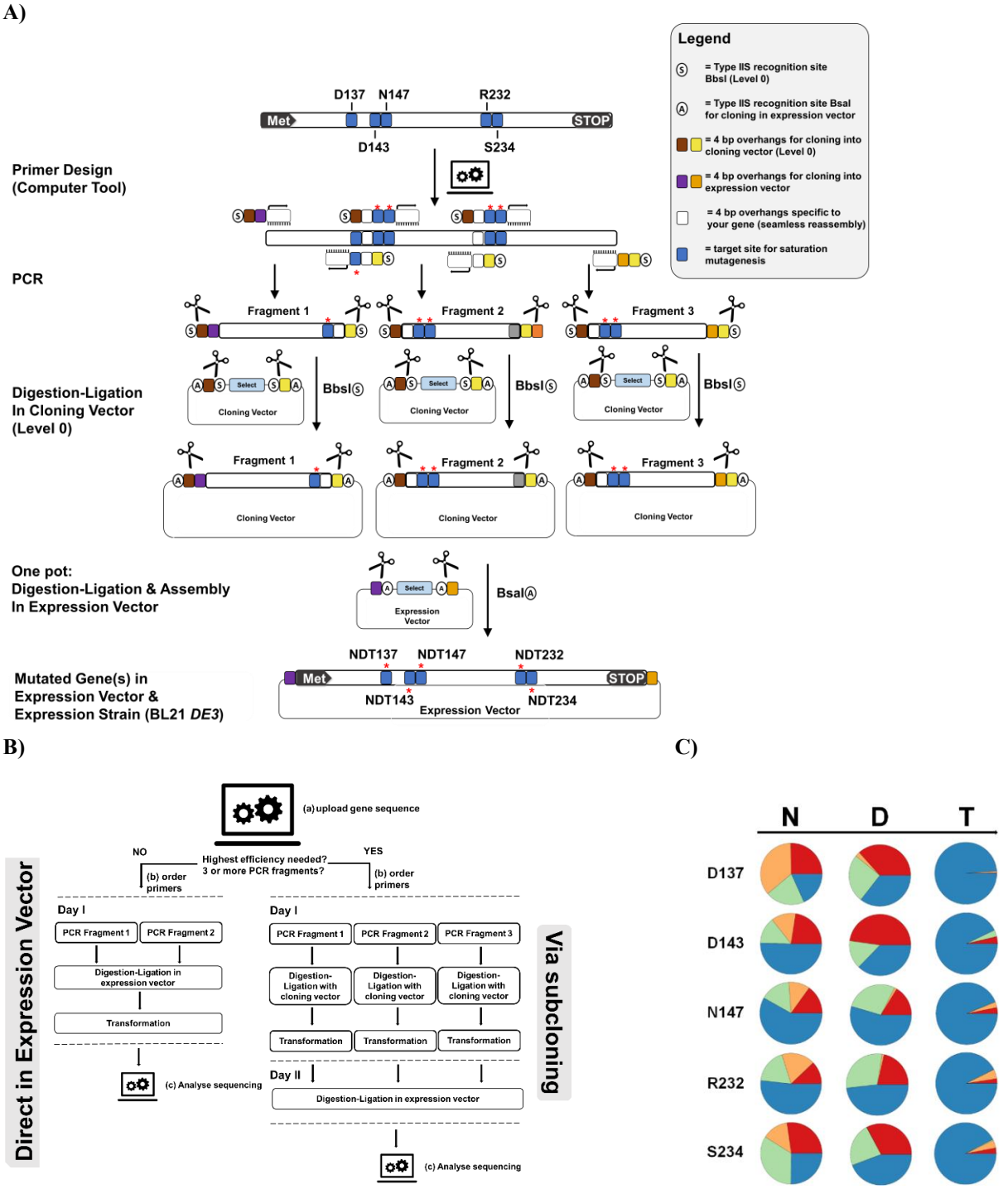


Figure 1. A) Conceptual background of Golden Mutagenesis. Primers are designed, which carry Type IIS recognition sites (circle with S = BbsI and circle with A = BsaI), 4 bp sequences for compatibility with the cloning vector (Level 0, brown [5'] and yellow [3']). Additionally, the primers at the 5' and 3' end include specific expression vector (Level 2) overhangs (purple and bordeaux-red) and the "inner" primers carry gene-specific 4 bp overhangs for a seamless reassembly of the fragments (white). Flanking the template binding part of the primer, randomization sites (blue with a red asterisk) are introduced. In a subsequent "one pot" Golden Gate digestion-ligation reaction BbsI binds to its recognition sequence and creates specific, pre-designed 4 bp overhangs (brown and beige) flanking the mutated PCR fragments, which are complementary to the exposed overhangs in the acceptor vector (cloning vector). The three created plasmids carry a vector located recognition site for the Type IIS restriction enzyme BsaI and through the previously introduced 4 bp overhangs for the expression vector (purple, bordeaux-red) as well as the gene-specific overhangs (white) a highly efficient one-pot cloning and reassembly of the separate gene fragments into the expression vector is achieved. B) Schematic overview of the work-flow: Upload of a gene of interest into the computer tool, decide whether to go for high-efficiency cloning or a quicker, single day procedure. The quick procedure designs primers for direct cloning into the expression/cloning vector. Particularly suitable for point mutations and less than three PCR fragments. To reach higher efficiencies an extra subcloning step as described in A) can be performed. The pooled sequencing results of the transformants can be analyzed using the software tool and a base distribution is given as in C) with the example of NDT saturations at 5 positions (simultaneous NDT (N= A,T,C,G; D=A,T,G) randomization of 5 amino acids residues of the enzyme YfeX)(color code: blue: Thymine; orange: Cytosine; green: Guanine and red: Adenin).

Table 1. Overview of Cloning results for NDT saturation mutagenesis.

Experiment	I site via Level 2	I site via Level 0	5 residues via Level 2	5 residues via Level 0
Efficiency[a]	96 %	100 %	100 %	100 %
Total amount white/orange colonies	862/0	4944/12	1364/55	1560/29
Ratio white to orange colonies	100	99:1	96:4	98:2

[a]Percentage of bacterial colonies containing plasmids of the expected size as determined by colony PCR (24 colonies tested in each approach)

Herein, we describe an optimized hands-on protocol for the implementation of Golden Gate Mutagenesis (coined Golden Mutagenesis) in any laboratory focusing on rational or random protein engineering. An open source software tool has been developed to facilitate primer design and analysis of the mutagenized library. This tool allows users to upload any gene sequence of interest, select a number of sites that shall be mutagenized and select a cloning/expression vector of choice. All required primers are automatically designed taking into account the desired codon degeneracies (e.g. NDT) and minimizing differences in primer melting temperatures. After physical construction of the library and sequencing, the generated randomizations can be assessed using the software since the distributions patterns of single nucleobases at specific positions are graphically illustrated in pie charts.

The concept of Golden Mutagenesis

Information about the conceptual basics of Golden Gate cloning is described in detail elsewhere.[14, 15, 17] Briefly, Type IIS restriction enzymes (BsaI or BbsI) are utilized, which cut outside of their recognition site. This characteristic feature allows the introduction of mutations within the gene of interest in a seamless/scarless manner. PCR products generated with primers designed to introduce mutations in the gene of interest are cloned in a target expression vector as a one-pot assembly reaction (Figure 1B). Alternatively, since the direct assembly of multiple PCR fragments may become inefficient for large numbers of PCR products, individual PCR fragments can be subcloned as an intermediate step in a cloning vector (Level 0). DNA preparations of the resulting fragment constructs, which contain randomization patterns at specified residues, are therefore prepared as libraries. The subcloned gene fragments are then reassembled into the final expression vector (Figure 1A). The availability of cloned gene fragments in separate vectors also offers the advantage of independently and sequentially screening individual mutational sites, by combining each mutated fragment with the corresponding wildtype gene regions. This is particularly relevant for the applications in CASTing and ISM protocols.[1-4]

The cloning step into the final expression vector—either directly from the PCR fragments or via intermediate subcloning of each fragment—could be done using a standard *E.coli* cloning strain for subsequent transformation, but should preferably be performed using an *E. coli* expression strain enabling the direct assessment of the phenotypes of the respective clones. The strain BL21 (*DE3*)pLysS was selected in this approach. Expression of the cloned gene requires induction of the T7 promoter using IPTG as an inductor. The screening of colonies containing recombinant constructs without inducing expression of the cloned genes requires a visual selection marker other than LacZ which requires IPTG induction. Therefore a novel expression vector, pAGM22082_CRed (Figure S2, available as AddGene plasmid #117225), was constructed. As a visual selection marker, this plasmid includes a canthaxanthin biosynthesis operon within the cloning site under control of a constitutively active bacterial promoter. This marker produces the orange carotenoid canthaxanthin independently of an exogenous inductor (and therefore independently of T7 promoter activation) and therefore allows orange/white selection in *E. coli* BL21(*DE3*)pLysS cells.

R Software Package: Primer design and quick quality control (QQC) of randomization

Since primer design in general and also for Golden Mutagenesis is relatively tedious and error-prone, a freely available application for primer design implemented on the programming language R has been developed (available at https://msbi.ipb-halle.de/GoldenMutagenesis). An implementation of an easy to handle online tool is currently under construction and will be accessible using the same URL.

The tool is implemented into the workflow as shown in Figure 1B). The user can specify the gene of interest and determine which residues shall be mutated. Removal of internal Type IIS restriction sites, which is necessary for Golden Gate cloning, can also be performed at the same time and is automatically done by the program; this process, called domestication, is advisable to be performed before generating a mutagenesis library.

Two different cloning options are available: The different vectors and their levels can be chosen among different options. a) Direct cloning from PCR fragments into a cloning vector (Level 0): pAGM9121 or expression plasmid (Level 2): pAGM22082_cRed; b) subcloning of individual gene fragments into Level 0 and then reassembly in Level 2 expression vector. The tool calculates and displays the full set of required mutagenic primers, which carry the distinct restriction sites, mutagenic sites as well as 4 bp overhangs for gene reassembly. Primer pairs are then calculated with minimal differences in melting temperature. A text protocol for helping the user perform all steps of the mutagenesis procedure is generated.
In addition, an analysis feature is available to easily assess the efficiency of the performed mutagenesis, a method that has previously been referred to as Quick Quality Control (QQC).[18] The sequencing results in ".ab1" format can be loaded by the software which is based on sangerseq R package.[19] The base distribution can then be visualized as pie-charts (Figure 1C).

Use Case I: Single site-directed mutagenesis and cloning into the Golden Gate plasmid

As a proof-of-concept, the widely used fluorescent protein mCherry[20] was used as a target, with the aim of introducing a single amino acid change (L69V) using one point mutation. Since

mCherry contains an internal BbsI site, which needs to be removed for Golden Gate cloning, a second nucleotide change is required. Therefore, the mCherry gene was split into three gene fragments using PCR. This required a total of six primers flanked by BbsI recognition sites (Table S1). All primers were designed by the software tool and the requirement to remove the internal restriction site was automatically detected within the process. New genetic information was also introduced at the terminal gene regions to provide compatibility with the acceptor plasmid pAGM9121 (Figure S3). The performed cloning led to 28688 colonies with a 99:1 ratio of white to blue colonies.

Use Case II: Multiple-Site Saturation Mutagenesis

Saturation mutagenesis was performed using the genetic information of the heme-protein YfeX[21, 22] as a template to introduce an NDT codon degeneracy at either one or five positions simultaneously. Two strategies were pursued: direct cloning and assembly of the mutagenic PCR fragments into the final expression vector, or alternatively, a subcloning step for each fragment into an intermediate cloning vector (Level 0), followed by a subsequent assembly in the final expression vector (Figure 1A, 1B). For the two approaches, the designed primers conceptually differ from the design of point mutational primers, since randomization sites are not suitable for specified reassembly and hence cannot be positioned within overhang regions. Therefore randomizations are rather introduced at the primer annealing sites flanking the 4 base pair reassembly overhangs (Figure 1A, blue with an asterisk).

The direct cloning into the expression vector for one NDT saturation site (two PCR fragments) led to 862 white colonies (no orange colonies detected) and a 96 % efficiency rate of reassembly (Table 1). In the case of five NDT saturation sites (three PCR fragments) 1364 colonies white and 55 orange were obtained. The assembly efficiency was ideal in the case of all 24 quick-tested colonies (100 %).

By using the subcloning step (via Level 0) for each PCR fragment followed by gene reassembly into the expression vector substantially higher colony numbers in case of the one position saturation were achieved. 4944 white and twelve orange colonies could be obtained (ratio: 99:1). The achieved efficiency of reassembly proved to be 100 % again. In the case of five saturation sites (three fragments) ideal efficiencies (100 %) as well as a high colony number (1589 in total) and an excellent white to orange colony ratio of 99:1 could be achieved. To access the base distribution within the randomization sites for this approach, a QQC analysis was performed and then analyzed using the software tool (Figure 1C). The expected NDT distribution pattern could be demonstrated at all 5 targeted positions, with a slight overrepresentation of Thymine, which is most likely caused by a bias within the ordered primer mixtures. As a further proof that the intended diversity of the library is not hampered due to the intermediate subcloning step, the individual cloning vectors (Level 0) have been plated as well leading to very high colony numbers and excellent white/blue ratios (up to 57948 colonies and 997:3 ratio; Table S2)

Conclusion

A protocol for the introduction of defined site-specific mutations within a gene of interest as well as for the generation of site-specific randomization libraries based on the Golden Gate cloning principle has been developed. Efficient cloning protocols have been developed including a newly build acceptor plasmid for direct T7 dependent expression in *E.coli*. The rather complicated process of primer design has been simplified by the construction of a freely available software tool, which also facilitates the analysis of the quality of the obtained library. The introduced Golden Mutagenesis technique is in particular suited for pursuing ISM protocols.

ASSOCIATED CONTENT

Supporting Information: Experimental details (File Supplemental info), Open Source Computer Tool examples for Multi-Site Directed Mutagenesis **(**File: Vignette Multi SD, Vignette Multi SD2, Vignette SD3), point mutagenesis and domestication (File: Vignette Point_Mutagenesis and Domestication), and Quick Quality Control (File: Vignette QQC).

AUTHOR INFORMATION

‡ These authors contributed equally.

Corresponding Author

E-mail: martin.weissenborn@ipb-halle.de ;
E-mail: sneumann@ipb-halle.de

Notes

The authors declare no competing financial interests.

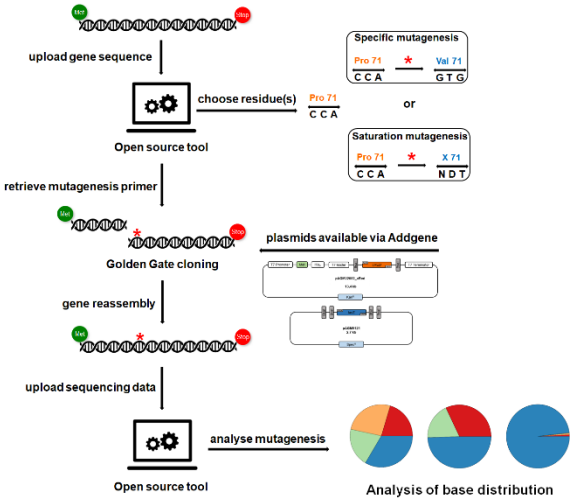
ACKNOWLEDGMENT

M.J.W thanks the BMBF („Biotechnologie 2020+ Strukturvorhaben: Leibniz Research Cluster“, 031A360B) for generous funding. P.P. thanks the Landesgraduiertenförderung Sachsen-Anhalt for a PhD scholarship. C.U thanks the Leibniz SAW funding. The authors are moreover grateful to Dr. Martin Dippe for providing the mCherry gene as well as Jonathon Hill who developed the PolyPeakParser Package and provided helpful tips for accessing the sequencing data via R.

REFERENCES

(1) Reetz, M. T., Wang, L. W., and Bocola, M. (2006) Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew Chem Int Ed Engl 45*, 1236-1241.
(2) Reetz, M. T., Carballeira, J. D., Peyralans, J., Hobenreich, H., Maichele, A., and Vogel, A. (2006) Expanding the substrate scope of enzymes: combining mutations obtained by CASTing. *Chemistry 12*, 6031-6038.
(3) Reetz, M. T., and Carballeira, J. D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature protocols 2*, 891-903.
(4) Acevedo-Rocha, C. G., Hoebenreich, S., and Reetz, M. T. (2014) Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. *Methods Mol Biol 1179*, 103-128.
(5) Chiang, L. W., Kovari, I., and Howe, M. M. (1993) Mutagenic oligonucleotide-directed PCR amplification (Mod-PCR): an efficient method for generating random base substitution mutations in a DNA sequence element. *PCR Methods Appl 2*, 210-217.
(6) Kegler-Ebo, D. M., Docktor, C. M., and DiMaio, D. (1994) Codon cassette mutagenesis: a general method to insert or replace individual codons by using universal mutagenic cassettes. *Nucleic Acids Res 22*, 1593-1599.
(7) Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene 77*, 51-59.

(8) Tyagi, R., Lai, R., and Duggleby, R. G. (2004) A new approach to 'megaprimer' polymerase chain reaction mutagenesis without an intermediate gel purification step. *BMC Biotechnol 4*, 2.
(9) Liu, H., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol 8*, 91.
(10) Dennig, A., Shivange, A. V., Marienhagen, J., and Schwaneberg, U. (2011) OmniChange: the sequence independent method for simultaneous site-saturation of five codons. *PLoS One 6*, e26222.
(11) Cozens, C., and Pinheiro, V. B. (2018) Darwin Assembly: fast, efficient, multi-site bespoke mutagenesis. *Nucleic Acids Res 46*, e51.
(12) Herman, A., and Tawfik, D. S. (2007) Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng Des Sel 20*, 219-226.
(13) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One 3*, e3647.
(14) Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009) Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLOS ONE 4*, e5553.
(15) Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012) Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioengineered bugs 3*, 38-43.
(16) Quaglia, D., Ebert, M. C., Mugford, P. F., and Pelletier, J. N. (2017) Enzyme engineering: A synthetic biology approach for more effective library generation and automated high-throughput screening. *PLoS One 12*, e0171741.
(17) Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D., Patron, N. J., and Marillonnet, S. (2014) A golden gate modular cloning toolbox for plants. *ACS Synth Biol 3*, 839-843.
(18) Acevedo-Rocha, C. G., Reetz, M. T., and Nov, Y. (2015) Economical analysis of saturation mutagenesis experiments. *Sci Rep 5*, 10654.
(19) Hill, J. T., Demarest, B. L., Bisgrove, B. W., Su, Y. C., Smith, M., and Yost, H. J. (2014) Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. *Dev Dyn 243*, 1632-1636.
(20) Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol 22*, 1567-1572.
(21) Weissenborn, M. J., Löw, S. A., Borlinghaus, N., Kuhn, M., Kummer, S., Rami, F., Plietker, B., and Hauer, B. (2016) Enzyme-Catalyzed Carbonyl Olefination by the E. coli Protein YfeX in the Absence of Phosphines. *ChemCatChem 8*, 1636-1640.
(22) Hock, K. J., Knorrscheidt, A., Hommelsheim, R., Ho, J., Weissenborn, M. J., and Koenigs, R. M. (2018) Iron-Catalyzed C—H Insertions: Organometallic and Enzymatic Carbene Transfer Reactions. *ChemRxiv*.

.

Insert Table of Contents artwork here