Standardizing cassette-based deep mutagenesis by Golden Gate assembly

Nicolas Daffern[1‡], Irene M. Francino-Urdaniz[1‡], Zachary T. Baumer[1‡], Timothy A. Whitehead[1,*]

[1]Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80305, USA

[‡] Authors contributed equally to this work

*Correspondence to:

Timothy A. Whitehead: timothy.whitehead@colorado.edu

JSC Biotechnology Building

3415 Colorado Avenue, Boulder, CO 80305

Phone: +1 (303)-735-2145

Abstract

Protocols for the construction of large, deeply mutagenized protein encoding libraries via Golden Gate assembly of synthetic DNA cassettes employ disparate, system specific methodology. Here we present a standardized Golden Gate method for building user-defined libraries. We demonstrate that a 25 µl reaction, using 40 fmol of input DNA, can generate a library on the order of $1x10^6$ members and that reaction volume or input DNA concentration can be scaled up with no losses in transformation efficiency. Such libraries can be constructed from dsDNA cassettes generated either by degenerate oligonucleotides or oligo pools. We demonstrate its real-world effectiveness by building custom, user-defined libraries on the order of 10^4 to 10^7 unique protein encoding variants for two orthogonal protein engineering systems. We include a detailed protocol and provide several general-use destination vectors.

Keywords: Golden Gate, Cassettes, Oligo Pools, Libraries, Protein Engineering, Deep Mutagenesis

Introduction

Cassette assembly has become a powerful way to create protein libraries thanks to modern synthesis technologies that can quickly and affordably produce custom DNA fragments[1–4]. Some of the most useful tools for assembling such fragments are Type IIs enzymes, which cut outside of their DNA recognition site and leave a user-defined four base pair overhang. While labs have employed these enzymes to manipulate DNA for the last several decades[5,6], a major development in their use came in 2008 when Engler *et al*. established a cloning method using the Type IIs enzyme BsaI, which they called Golden Gate assembly[7]. Since then, the method's popularity has grown[8–10], due in part to its ability to connect fragments in a specific order and without a restriction scar[11].

Although the assembly of synthetic DNA by Golden Gate has clear utility in building large protein libraries, a generalized procedure has yet to be established. While several labs have used Golden Gate to build libraries from cDNA and synthetic DNA fragments, their assembly protocols were not appropriate for general use by containing additional, system-specific steps such as sub-cloning into entry plasmids or multiple rounds of PCR to construct more complicated inserts. Additionally, these studies did not contain general instructions for insert design, and did not provide publicly available destination vectors to the public.[12–16] In 2019 Püllmann *et al.* developed a more general Golden Gate protocol for creating site-saturation libraries from oligonucleotides, which was well characterized and included a useful script for primer design[17]. However, the base version of this protocol was only shown to generate a single point mutation and a more complicated protocol involving subcloning was needed to make a library of 60 variants.

Here we provide a simple, broadly applicable Golden Gate procedure that can be used to build large ($>10^7$), site-specific, deeply mutagenized libraries. We present data on the procedure's efficiency under different use conditions and demonstrate its effectiveness in constructing libraries from mixed base-containing oligonucleotides and custom synthesized oligo pools. We also provide a detailed protocol with discussion of important design considerations and three general-use destination vectors deposited on Addgene.

Methods

Construction of destination vectors and cassettes

Destination vectors were created by combining a source vector and a GFP-encoding insert from pYTK047[10] (sfGFP) or pEDA5[18] (GFP mut3) using Gibson Assembly[19] or restriction enzyme cloning. Oligo pool and Ultramer derived double-stranded cassettes were obtained by performing PCR with single-stranded source DNA and a single reverse primer, followed by gel extraction. Sequences for primers, completed destination vectors, and wild-type versions of cassettes are listed in the Supporting Data (**Supplemental Table 2, Supplemental Data**).

Performance and assessment of Golden Gate reactions

All Golden Gate reactions were performed using a base protocol with some small number of changes which are listed below. The base protocol is as follows. 40 fmol of a destination vector and 40 fmol of each cassette are combined in a PCR tube along with 20 units of BsaI HF-V2, 400 units of T4 DNA ligase, 2.5 µl of 10x T4 ligase buffer, and nuclease-free H_2O up to 25 µl. The reaction mixture is then cycled between a 37 °C and 16 °C PCR step for a total of 60 cycles, followed by a final 5 minute, 37 °C step and a 10 minute, 65 °C step. The resulting DNA is then cleaned and concentrated using an NEB Monarch PCR DNA Cleanup Kit and eluted in 6 µl of nuclease-free H_2O. All 6 µl of concentrated DNA are then transformed by electroporation into 50 µl of TransforMax EPI300 *E. coli* cells. Additionally, 0.1 ng of pUC19 in 6 µl of nuclease-free water was transformed into TransforMax cells to assess cell competency. Transformations were carried out in 1 mm electroporation cuvettes (BTX, Cat# 45-0124) with an electroporator (Eppendorf, Cat# 4309000027) set to 1200 V, which generated time constants between 4-6 ms. Immediately after electroporation, cells were resuspended in 1 ml of SOC media (SOB from BD Cat# 244310 with 20mM dextrose), incubated at 37 °C for 1 hour, plated on LB agar containing the appropriate antibiotic, and were incubated overnight at 37 °C. Numbers of transformants and percentages of GFP negative colonies were determined by plating serial dilutions of the transformed cells ranging from 10^2-10^8 fold-diluted and counting the number of green and white colonies in the dilution that had the highest number of colonies totaling between 10 and 100. If no green colonies could be detected, the percentage incorporation was estimated as one over the total number of colonies in the lowest dilution, as extrapolated from the dilution containing all counted colonies. pUC19 transformation efficiencies were similarly determined by counting colonies from serial dilution plating, with the resulting numbers being scaled to reflect the number of transformants per 1 ng of transformed pUC19. For libraries, the remaining transformants not used for dilutions were plated onto bioassay plates and incubated at 37 °C. The next day colonies were scraped from the plates and mini prepped to obtain library DNA. For some experiments, the amount of input DNA or total reaction size was modified from the base protocol; these modifications are noted in the main text. Reaction size and input DNA along with the destination vector and cassettes used in each reaction are listed in **Supplemental Table 1**. The T7 RNAP library was generated with several small additional changes to our base protocol. Two 200 µl reactions (same PCR step reaction; 200 femtomoles of backbone and each insert in each reaction, 20 µl 10x T4 DNA Ligase Buffer, 5 µl BsaI-HFv2, 5 µl T4 DNA Ligase) were pooled, put through a PCR cleanup step, and used in a single electroporation transformation.

Characterization of Golden Gate assembled RBD constructs

Binding of GG and Wuhan-1 RBDs to ACE2/CC12.1 was assessed by yeast display as described by Francino-Urdaniz *et al*[20].

Deep sequencing preparation

Deep sequencing prep was performed as described in the "Method B" protocol from Kowalsky *et al*[21]. In brief, we did two rounds of PCR with an ExoI clean up in between. In the first round we amplify the amplicon using customized primers for each specific sequence. These primers also contain the TruSeq illumina adapters. The second PCR inserts the TruSeq barcodes for sequencing and is performed exactly as described in the "Method B" protocol. The primers used for the deep sequencing preparation are given in ***Supplementary Table 2***.

The first PCR thermocycler conditions for the RBD library are as follows: Initial Denaturation: 98 °C, 60 s. Amplification (18 cycles) Denaturation: 98 °C, 10 s; Annealing: 57 °C, 20 s; Extension: 72 °C, 30 sec. Final Extension: 72 °C, 10 min. Hold: 4 °C.

The first PCR thermocycler conditions for the T7 RNA polymerase library are as follows: Initial Denaturation: 98 °C, 60 s. Amplification (16 cycles): Denaturation: 98 °C, 10 s; Annealing: 69 °C, 20 s; Extension: 72 °C, 30 s. Final Extension: 72 °C, 10 min. Hold: 4 °C.

For deep sequencing of the RBD Ultramer 1 library assembled via Gibson Assembly, new primers were ordered and a single PCR was performed. The PCR reaction inputs were 0.5 uM of each primer, 500 ng of DNA, 500 μM of dNTPs, 10 μM of 5x HF Buffer, and 0.5 μL of Phusion polymerase. The PCR conditions were Initial Denaturation: 98 °C, 60 s. Amplification (20 cycles): Denaturation: 98 °C, 10 s; Annealing: 68 °C, 20 s; Extension: 72 °C, 45 s. Final Extension: 72 °C, 10 min. Hold: 4 °C.

DNA deep sequencing analysis

For the analysis of the RBD library deep sequencing data, sequences were merged using an in-house merging code, essentially as previously described by Haas *et al*[22] The occurrence of mutations at each designed location were counted and summed. From these values, the corresponding distributions were generated.

For analysis of the T7 RNAP library deep sequencing data, sequences were merged using FLASh with a maximum overhang of 100 bp[23]. Merged sequences were filtered to be 416 bp, start with amino acid sequence "EIL" in the appropriate reading frame, contain no "N" basecalls, and to be absent of any premature stop codons. Paired-end reads of length (412-417 bp) were used for comparison to correct lengths as these are the most likely errant lengths, and these showed up at elevated counts relative to merge lengths outside of this range. The occurrence of mutations at each designed location were counted and summed, including the occurrence of amino acids not in the original design space. From these values, the corresponding distributions were generated.

For analysis of the RBD Ultramer 1 library assembled via Gibson Assembly, sequences were merged using FLASh with a maximum overhang of 80 bp. Merged sequences were filtered to be 440 bp, start with amino acid sequence "GVS" in the appropriate reading frame, contain no "N"

basecalls, and to be absent of any premature stop codons. The occurrence of mutations at each designed location were counted and summed. From these values, the corresponding distributions were generated.

Gibson assembly mutagenesis

For Gibson Assembly, NEB DNA HiFi Assembly Master Mix was used with 40 fmol of each cassette as well as 40 fmol of insertion vector in a 25 µL total reaction volume. The double stranded cassette from the mixed base Ultramer was generated from a Gibson-specific template and reverse primer. The backbone for Gibson Assembly was prepared by gel extracting the BsaI digested destination vector, pIFU037. The remainder of the protocol (PCR cleanup, elution, electroporation) was identical to that of the Golden Gate protocol described above.

Results/Discussion

A standardized protocol for library generation

The general workflow for library generation using our protocol involves design and creation of a destination vector and mutagenic cassettes, assembly via a Golden Gate reaction and transformation into bacteria. The destination vector must be designed and created with a selection marker, BsaI sites, and specific overhangs such that cassette(s) introduction results in reconstitution of a full-length gene encoding sequence, and replacement of the marker which allows rapid assessment of incorporation efficiencies (**Figure 1A**). In parallel, cassettes are designed with BsaI sites and overhangs arranged for sequential insertion into the destination vector (**Figure 1A**) using subsets of the high-fidelity overhangs outlined by Potapov *et al.* to reduce inefficiencies from imperfect ligation[24]. Subsequently, the destination vector and cassettes are mixed with BsaI-HFv2 and T4 ligase and PCR cycled between 37 °C and 16 °C, during which time the selection marker and cassette ends are removed, the cassettes anneal with the vector, and the annealed DNA is ligated.

We developed a modified Golden Gate protocol that allowed us to rapidly assemble libraries while maintaining high numbers of transformants. For the base version of this protocol, we used a 25 µl reaction containing 40 fmol of the destination vector and each cassette. Additionally, our reactions were PCR cycled for 60 cycles with each step only lasting one minute, as described by Strawn *et al.*[25], allowing the Golden Gate reaction and bacterial transformation to be performed in under 8 hours (**Figure 1A**). We have included a protocol describing the design and creation of libraries using this technique in more detail (**Supporting Information**). Using this protocol, with properly designed cassettes and a destination vector in hand, a new library for any protein can be generated in a single day. To facilitate this, we built general use destination vectors for yeast surface display (pND003), MBP-tagged protein expression in *E. coli* (pND004), and yeast two-hybrid assays (pND005), each of which contain a GFP marker and BsaI sites with high fidelity overhangs. To confirm their functionality, a Golden Gate reaction was performed with pND003

and cassettes that encode for a monomeric variant of the plant abscisic acid receptor PYR1 (H60P, N90S)[26]. Comparison of plates from a transformation of pND003 alone and a transformation of the Golden Gate reaction confirmed cassette insertion and demonstrated that removal of the GFP marker allowed for assessment of incorporation efficiencies (**Figure 1B**). We used numbers of GFP negative colonies to calculate incorporation percentages for all Golden Gate reactions described in this paper with a median incorporation of 99.7% (range 81% to >99.9%, n=40) (**Figure 1C; Supplemental Table 1)**.

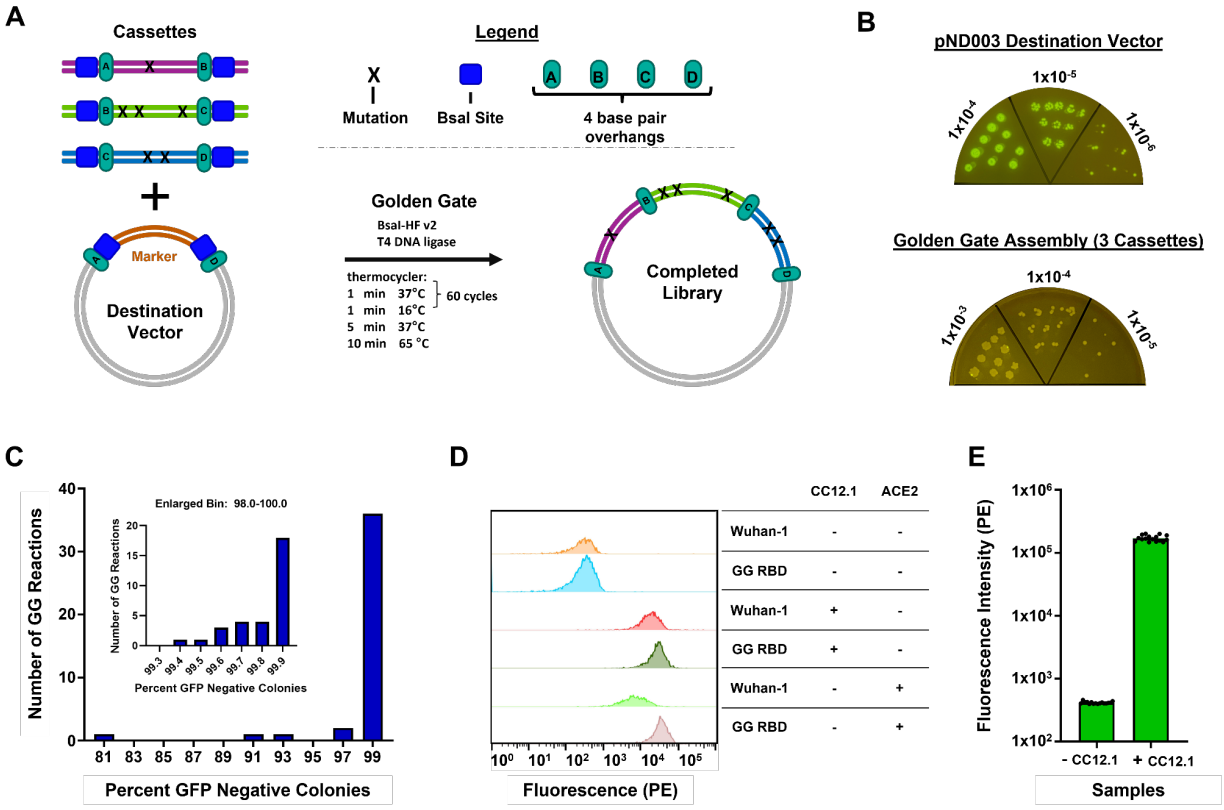


Figure 1. Establishing the functionality of a Golden Gate protocol for library generation. A. *Schematic showing library generation by assembly of mutagenic cassettes into a destination vector via Golden Gate.* ***B.*** *Dilution plating of E.coli transformations with 40 fmol of our destination vector pND003 (left) and a Golden Gate reaction performed with 40 fmol of both pND003 and three wild type PYR1 cassettes (right). Listed numbers represent fold dilution.* ***C.*** *Histogram showing the percent cassette incorporation for all performed Golden Gate reactions. Inset shows all reactions from the 98-100% bin redistributed in bins with 0.1% intervals. Incorporation percentages were calculated by comparing the number of green (GFP) and white (non-GFP) colonies on dilution plates after transformation.* ***D.*** *Functional comparison of GG RBD and Wuhan-1 RBD using yeast surface display. RBD displaying yeast were incubated in the absence (-) or presence (+) of saturating concentrations of PE-labeled CC12.1 (antibody) or ACE2, followed by assessment of binding by flow cytometry.* ***E.*** *Assessment of CC12.1 binding for 16 colonies displaying RBD from 4 different Golden Gate reactions. Individual values represent mean PE fluorescence intensity of RBD-displaying populations.*

To verify that our protocol would result in genes encoding full-length, functional protein, we assessed the ability of a Golden Gate-assembled SARS-CoV-2 Omicron chimeric S RBD to bind

ACE2 and the neutralizing monoclonal antibody CC12.1[20,27]. For these experiments we created a destination vector with the coding sequence for the Omicron chimeric RBD (SARS-CoV-2 S RBD (333-541) Wuhan-1 with mutations S477N, E484A, Q498R, N501Y and Y505H) and the corresponding dsDNA coding cassettes, assembled them into the RBD destination vector using our Golden Gate protocol, transformed the resulting plasmids into yeast, and expressed the isogenic Golden Gate-derived RBDs (GG RBD) on the surface of yeast. As a positive control, we also expressed the previously described Wuhan-1 S RBD N343Q (Wuhan-1)[20] on the surface of yeast. We then compared the ability of the GG RBD and Wuhan-1 to bind Fc-ACE2 and CC12.1 (which contains an Fc) using flow cytometry. When labeled with an anti-Fc PE, we observed specific PE fluorescence resulting from binding to ACE2 or CC12.1 for both GG RBD and Wuhan-1 (**Figure 1D**). We then tested GG RBD functionality in a similar manner for 16 different yeast colonies from four separate Golden Gate reactions and saw consistent levels of CC12.1 binding for all 16 variants, highlighting the ability of this protocol to reproducibly assemble full-length sequences (**Figure 1E**).

Benchmarking a scalable single day Golden Gate library generation protocol

Next, we benchmarked the number of transformants that can be generated using this protocol as a function of cassette number, reaction size, and input DNA. For all these experiments we used our pND003 vector and PYR1 cassettes. Our baseline for all three experiments was a 25 µl reaction containing 40 fmol of the destination vector and a single PYR1 cassette, which generally results in 7-8x10^5 transformants and >96% cassette incorporation (**Figure 2A-C, Supplemental Table 1**).

Since previous studies have noted decreasing Golden Gate efficiencies when increasing the number of 'parts' (number of cassettes plus destination vector)[13,28], we sought to quantitate the number of transformants as a function of the number of input cassettes. For this experiment, we performed four reactions with one to four PYR1 cassettes under conditions that were otherwise identical to the baseline reaction. We found that transformation efficiencies were modestly affected by increasing cassette number, with a four-cassette (five-part) assembly showing a minor reduction in transformants (3.7x10^5, n = 2) over a one-cassette (two-part) assembly (8.4x10^5, n = 2) (**Figure 2A, Supplemental Table 1**). Thus, increasing cassette numbers should not hinder most library designs.

We then assessed the ability of our protocol to be scaled by total reaction size by comparing our baseline 25 µl reaction with 100 µl and 200 µl reactions. We found that the number of transformants scaled at least linearly, with the 200 µl reaction (8x volume) generating an approx. 24-fold increase in transformants with no loss in incorporation efficiency (p-value 0.03**; Figure 2B, Supplemental Table 1**). We speculate that this trend results from decreased relative DNA loss while working with small constant volumes during the PCR cleanup and transformation steps. Since a 100 µl reaction can be performed in a single PCR tube, several orders of

magnitude higher numbers of transformants should easily be achieved by pooling multiple Golden Gate reactions in a single PCR cleanup column.

Subsequently we tested scaling of DNA concentration by comparing our baseline reaction with reactions in which the amount of destination vector and cassette DNA was increased to 80 and 160 fmol, without increasing reaction size. We again observed a linear increase in transformants, with 160 fmol (4x increase) of input DNA resulting in a 19-fold increase in transformants (p value 0.13 for whether the 160 fmol reaction gives more than a 4x increase in number of transformants) and no loss in incorporation efficiency (**Figure 2C, Supplemental Table 1**). Thus, reactions can be scaled by both volume and DNA concentration.

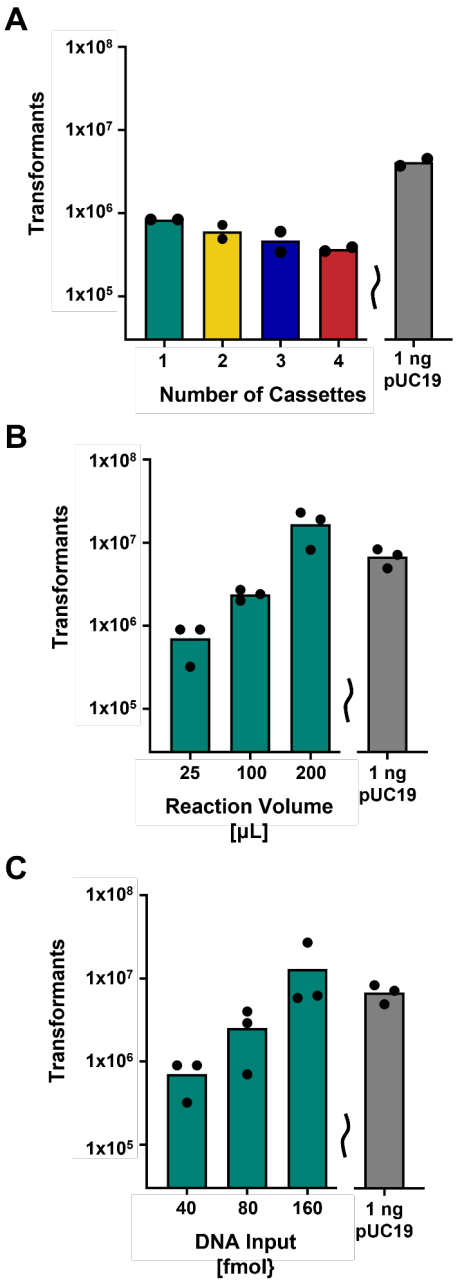


Figure 2. Standardizing a Golden Gate protocol for library generation A. *Transformation efficiencies for Golden Gate reactions performed with differing numbers of cassettes. 25 µl reactions were performed in duplicate with 40 fmol of both pND003 and one, two, three, or four wild type PYR1 cassettes. Numbers of transformants were calculated by plating serial dilutions of the recovered cells. pUC19 is shown as a control reaction for assessing efficiency of electrocompetent cells.* ***B.,C.:*** *Transformation efficiencies as a function of Golden Gate reaction size (**B**) or input DNA (**C**). A baseline 25 µl reaction using 40 fmol of both pND003 and one wild type PYR1 cassette was compared with reactions having increased size or input DNA. Reactions were performed in triplicate. Numbers of transformants were calculated by plating serial dilutions of the recovered cells. (**B.**) All reaction components were scaled 1:1, including DNA concentration, with increasing volume. (**C.**) The total amount of input DNA for the destination vector and cassette were increased while all other reaction conditions were held constant.*

Creation of a site-specific protein libraries using both DNA Ultramers and oligo pools

We next assessed the ability of our protocol to produce complete libraries from different types of synthetic DNA by building libraries for the Omicron chimeric RBD. For this we designed three site-specific combinatorial libraries, all covering the same 110 contiguous positions, assembled using a combination of three mutant cassettes (**Figure 3A, Supplemental Figure 1**).

We generated cassettes from synthetic dsDNA (eBlocks), mixed-base degenerate long oligonucleotides (Ultramers), and ssDNA sourced from custom oligo pools (**Figure 3B**). Generally, we used eBlocks to encode unmutated regions of protein, while Ultramers and oligo pools were used as mutagenic cassettes. dsDNA cassettes are generated from single-stranded Ultramers or oligo pool DNA using PCR with a reverse primer (**Figure 3B**). For these experiments, we assembled Libraries 1 and 2 from PCR-amplified mutant Ultramer cassettes and Library 3 from PCR-amplified mutant oligo pool cassettes (an example of this amplification for Library 1 is shown in **Figure 3C**). For comparison, we performed a fourth assembly with eBlock cassettes containing no sequence variation.

We performed PCR amplification of the oligo pools and Ultramers as well as the Golden Gate reactions using the base method described in our protocol, except for increasing input DNA from 40 to 200 fmol. We found similar transformation efficiencies using Ultramers and oligo pools, with both resulting in approximately $2.1x10^6$ transformants, and a slightly higher efficiency with eBlocks, which resulted in approximately $7.0x10^6$ transformants (**Figure 3D**). Thus, our protocol generates consistently high numbers of transformants with different types of input DNA, giving the user flexibility when designing their libraries.

To assess library quality, completed libraries were deep sequenced at a depth ranging from 5.9e5 to 2.9e6 reads (**Supplemental Table 3**). In contrast to other user-defined mutagenic protocols with high wild-type sequence carryover[18], no library contained more than 0.2% wild-type reads, and all libraries contained at least 96.7% of the desired variants (96.7, 99.9, 99.9%). This low wild-type percentage is comparable with another mutagenesis protocol[29]. The libraries ranged from 80.9-86.8% of on-target sequences. 9% of the oligo pool-derived library coded for chimera sequences that contained unintended combinations of designed mutations. These could have arisen during our deep sequencing preparation, as chimera formation is known to occur at low abundances during the associated PCR steps[20]. Alternatively, chimera formation could occur during the cassette generation step of our protocol; our analysis is unable to distinguish between these possibilities. We assessed library uniformity by comparing the theoretical frequency of different residues at each mutated position with the observed frequency seen in our deep sequencing data (**Figure 3E**). The relative frequencies observed varied between 0.34-1.52-fold as compared with expectations (n=62). Together, this data demonstrates that our Golden Gate protocol can generate user-defined combinatorial mutational libraries with almost complete coverage, little to no wild-type carryover, and near-uniform individual mutational distributions.

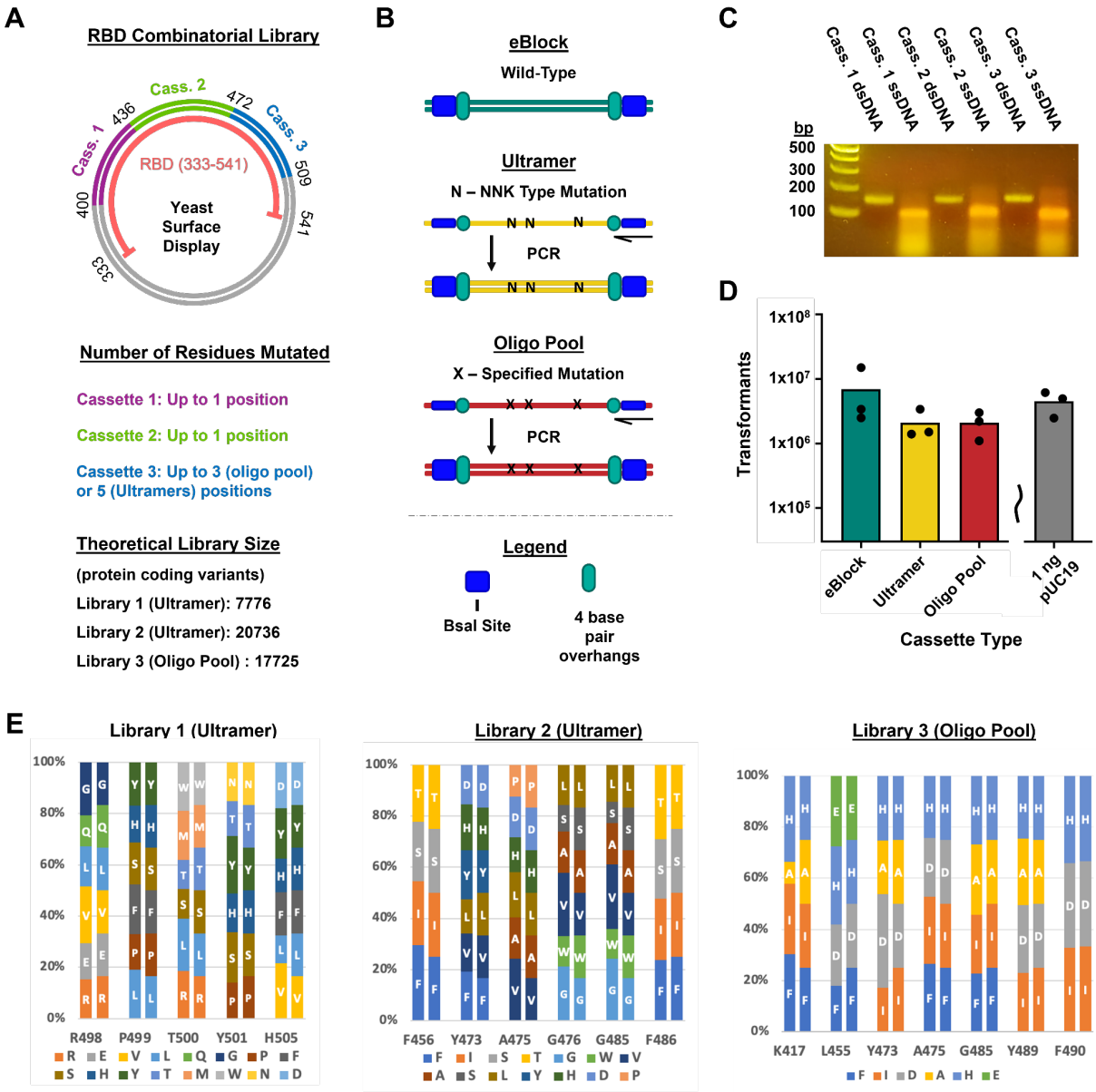
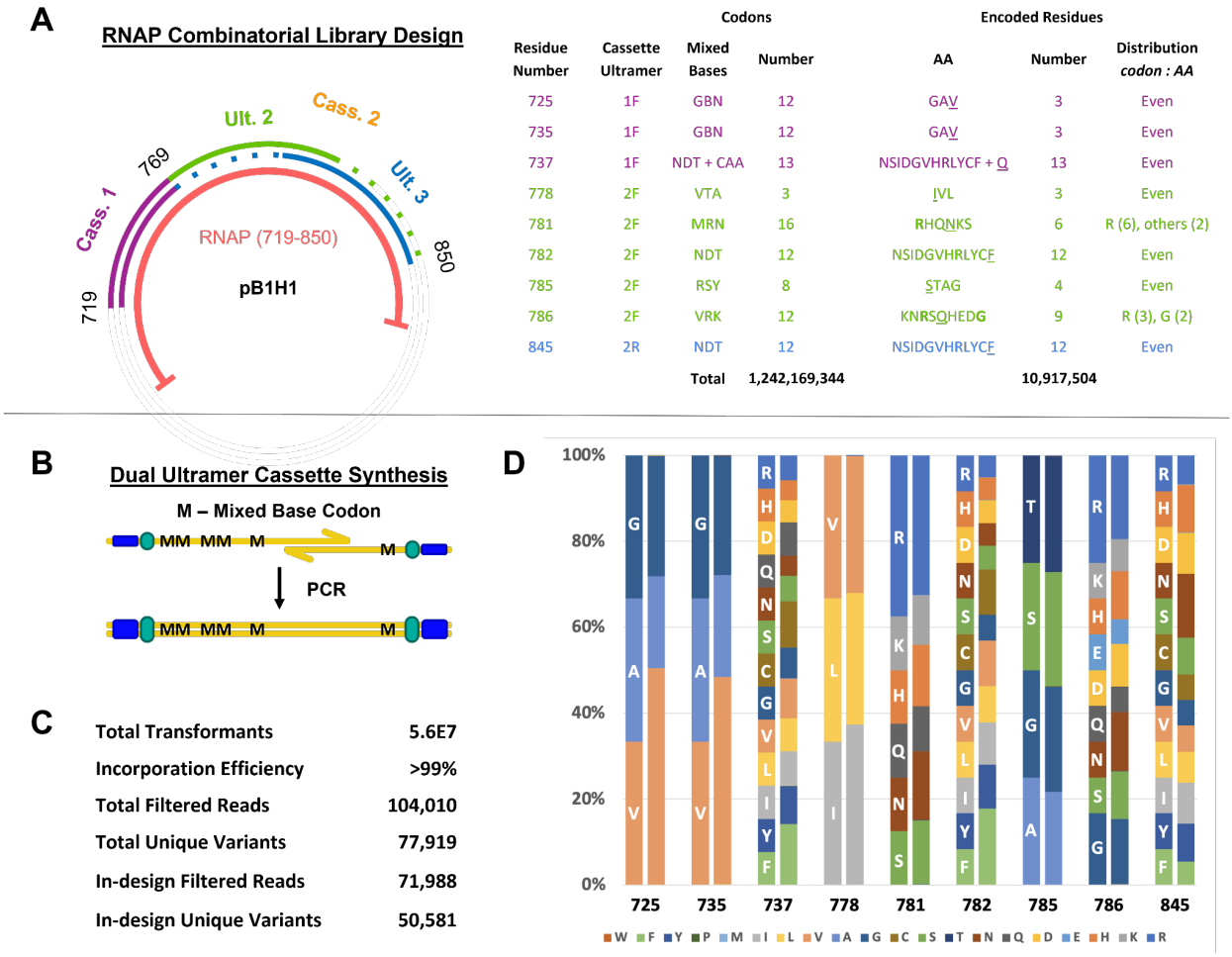


Figure 3. Golden Gate assembly of SARS-CoV-2 S RBD deep mutational libraries. A. *Schematic of the assembled plasmid for the S RBD combinatorial library. Unmutated S RBD residues 333-399 and 510-541 are encoded in the destination vector pIFU037 while residues 400-509 are encoded by three mutagenic cassettes. Cassettes can code for wild-type or mutant residues at each mutational site* ***B.*** *Different cassette DNA inputs for the Golden Gate assembly. eBlocks (dsDNA) are used as is, while Ultramers (mixed base-containing long oligonucleotides) and oligo pools are obtained as lyophilized ssDNA and require PCR synthesis using a reverse primer to generate dsDNA.* ***C.*** *Gel electrophoresis of three RBD library Ultramer cassettes before (ssDNA) and after PCR (dsDNA). Each cassette is 170 nts.* ***D.*** *Transformation efficiencies for Golden Gate reactions performed with pIFU037 and the three RBD library cassettes, each generated from different types of DNA. 25 µl reactions were performed in triplicate with 200 fmol of destination vector and cassettes. Numbers of transformants were calculated by plating serial dilutions of the recovered cells. pUC19 is shown as a control reaction for assessing efficiency of electrocompetent cells.* ***E.*** *Mutational distributions for RBD libraries. Expected (right bars) vs observed mutational frequency (left bars) at each mutated residue based on deep sequencing.*

Creation of a large site-specific mutational library

We next assessed if our protocol could be scaled to generate larger, high-quality libraries by designing a library for T7 RNA polymerase coding for $1.1x10^7$ theoretical protein variants using a reduced codon alphabet[30] (**Figure 4A**). We chose to build the library using Ultramer-based cassettes, where mutations spanned 363 nucleotide bases (V725-F845) and were grouped at three locations across two cassettes (**Figure 4A**). The second cassette, containing mutations at positions 781-786 and 845, was designed using two overlapping Ultramers (**Figure 4A, 4B**). To construct the library, we performed two separate 200 µl Golden Gate reactions that each contained 200 fmol of destination vector and each cassette, and pooled the DNA for use in a single transformation, generating $5.6x10^7$ transformants with an incorporation efficiency of >99% (**Figure 4C**).



Residue Number	Cassette Ultramer	Codons: Mixed Bases	Codons: Number	Encoded Residues: AA	Encoded Residues: Number	Encoded Residues: Distribution *codon : AA*
725	1F	GBN	12	GAV	3	Even
735	1F	GBN	12	GAV	3	Even
737	1F	NDT + CAA	13	NSIDGVHRLYCF + Q	13	Even
778	2F	VTA	3	IVL	3	Even
781	2F	MRN	16	RHQNKS	6	R (6), others (2)
782	2F	NDT	12	NSIDGVHRLYCF	12	Even
785	2F	RSY	8	STAG	4	Even
786	2F	VRK	12	KNRSQHEDG	9	R (3), G (2)
845	2R	NDT	12	NSIDGVHRLYCF	12	Even
		Total	1,242,169,344		10,917,504	

Figure 4. Golden Gate assembly of a large T7 RNAP library A. *Schematic of the assembly strategy and mutational designs for the T7 RNAP library. The assembly used two cassettes: one standard cassette generated from a single Ultramer, and one long cassette generated from two overlapping Ultramers (left). Different sets of mixed base codons were used to code for subsets of amino acids at each mutational site (right). Underlined single letter amino acids are the wild type residue at a given position and bold residues are encoded by a greater relative number of codons as a function of the mixed bases chosen.* ***B.*** *Schematic for using two overlapping Ultramers as forward and reverse primers to generate a longer cassette.* ***C.*** *Summary Library Statistics. Filter conditions included correct length of merged reads, the presence of "N" in the merged reads and the presence of stop codons. In-design reads*

exclude any mutations at protein level which are not in the design space. ***D.*** *Expected (left bars, with single letter amino acid code) vs observed mutational frequency (right bars) at each library residue. The sum of all amino acids not in the set of designed mutations was less than 0.5% at each position.*

Subsequently, we performed limited deep sequencing to assess the quality of our library. We obtained 104,010 reads after quality filtering and observed 77,919 total unique protein encoding variants, including 50,581 in-design protein encoding variants, with no wild-type sequences, further demonstrating the ability of our protocol to limit wild-type carryover. To assess library uniformity, we compared the theoretical frequency of different amino acids at each mutational position with the observed frequency seen in our deep sequencing data (**Figure 4D**). The relative frequencies observed varied between 0.6-2.1-fold compared with expectations (n=65), showing that large libraries can be assembled with minimal mutational bias. A conservative analysis (see **Supplemental Note 2)** results in an estimated 81% library coverage[31].

Yeast homologous recombination and Gibson assembly have also been used to generate large mutagenic libraries by segmenting protein encoding genes into several cassettes[19,32]. We compared the Golden Gate method with Gibson assembly using the chimeric Omicron RBD platform and the Library 1 ultramer (**Supplemental Figure 1A; Figure 3**). The three cassettes were redesigned to be comparable to the Golden Gate setup, except that each contains between 30-34 bp homology arms. The backbone for Gibson Assembly was prepared by gel extraction of a BsaI digested destination vector (pIFU037). Gibson assembly was performed side-by-side with Golden Gate using 40 fmol of each cassette and backbone, along with a pUC18 transformation control.

Both GG and Gibson had >99% white colonies and a similar number of transformants (Gibson: 1.3-1.9E5; GG 2.6E5) (**Supplemental Figure 1B**). The Gibson assembly library was at least 99.1% complete (7606/7776 unique library members) and of a similar library uniformity compared to GG (**Supplemental Figure 1C)**, as evaluated by NGS at a depth of 2.5E5 paired reads. Consistent with the existing body of literature, we conclude that mutational libraries of the same general quality of the GG protocol can also be prepared by Gibson assembly. The advantages of Golden Gate mutagenesis compared with Gibson are two-fold. First, as there is less overlap between cassettes (4 bp vs. 30 bp homology), oligo pools with higher diversity can be synthesized, and more contiguous mutational space can be mapped. Second, Gibson often requires an additional insert vector step either through PCR or by digestion and gel extraction. This can either add mutational biases by PCR or add to the hands-on time of the protocol.

Conclusions

Although the need for large, site-specific libraries for protein engineering workflows is almost universal, the methods for generating such libraries vary greatly. Modern DNA synthesis technology, which allows for the rapid creation of short DNA fragments containing user-defined mutations, has the potential to streamline these methods and make libraries more consistent and affordable. Here we provide a simple, broadly applicable protocol detailing how to build large

libraries from synthetic DNA cassettes using Golden Gate assembly. To provide an accurate assessment of this protocol's capabilities, we determined the obtainable transformation and incorporation efficiencies as a function of cassette number, reaction size, and input DNA. We then demonstrated its real-world applicability by creating libraries of different sizes in two orthogonal protein engineering systems. First, we generated libraries for the SARS-CoV-2 S RBD on the order of 10^4 members and demonstrated by deep sequencing that near-uniform, near-complete libraries can be made equally well using mixed base oligonucleotides or oligo pools. Subsequently, we built and characterized a larger library encoding T7 RNAP variants to show that this technique is broadly applicable, and that uniformity and coverage are maintained when our protocol is scaled up. We demonstrated that our protocol scales with assemblies of at least five fragments with only a limited decrease in the number of transformants. We speculate that this high efficiency results from higher quality enzymes engineered by NEB, the use of compatible overhangs enumerated by Potapov *et al*, and the specific reaction conditions optimized in this paper[24]. We expect our standardized method will facilitate implementation of cassette-based Golden Gate mutagenesis for the protein engineering and design community.

Author Information

Corresponding Author

Timothy A. Whitehead - *Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80305, USA; orcid.org/0000-0003-3177-1361*

Authors

Nicolas Daffern - *Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80305, USA;* orcid.org/0009-0006-6596-8612

Irene Francino-Urdaniz - *Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80305, USA;* orcid.org/0000-0001-8786-8002

Zachary T. Baumer - *Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80305, USA;* orcid.org/0000-0003-2691-4946

Author Contributions

Conceptualization: T.A.W., N.D., I.F.U., and Z.T.B. Experimental Design: T.A.W., N.D., I.F.U., and Z.T.B. Performed Experiments: N.D., I.F.U., and Z.T.B. Data Analysis: T.A.W., N.D., I.F.U., and Z.T.B. Wrote First Draft: N.D. Writing & Editing: T.A.W., N.D., I.F.U., and Z.T.B. Supervision: T.A.W. Funding Acquisition: T.A.W. and Z.T.B.

Notes

The authors declare no competing financial interest.

Acknowledgements

This work was supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health (Award Numbers 1R21AI174157-01 and 5R01AI141452-05 to T.A.W.), the Defense Advanced Research Projects Agency Advanced Plant Technologies (grant HR001118C0137 to T.A.W. - Approved for Public Release, Distribution Unlimited), the National Science Foundation (grant NSF CBET- 2030221 to T.A.W.), and the National Science Foundation Graduate Research Fellowship Program (Z.T.B. DGE Award Number 2040434, fellow ID: 2021324468).

Data Availability

The plasmids pNDD003, pND004, and pND005 and their corresponding sequences have been deposited at AddGene (AddGene accession numbers #200949, 200950, 20051, respectively). All raw deep sequencing reads have been deposited at the Sequencing Read Archive (SRA to be updated upon publication). Raw data used to prepare figures 1C, 2A-C, and 3D are present in **Supplemental Table 1**.

Supporting Information

Supplemental methods; Table of conditions used in all Golden Gate reactions; Table of all primer sequences used; Detailed protocol for performing Golden Gate cassette mutagenesis; Complete sequences of all plasmids and cassettes used

REFERENCES CITED

(1) Klein, J. C., Lajoie, M. J., Schwartz, J. J., Strauch, E.-M., Nelson, J., Baker, D., and Shendure, J. (2016) Multiplex pairwise assembly of array-derived DNA oligonucleotides. *Nucleic Acids Res 44*, e43–e43.

(2) Kuiper, B. P., Prins, R. C., and Billerbeck, S. (2022) Oligo Pools as an Affordable Source of Synthetic DNA for Cost-Effective Library Construction in Protein- and Metabolic Pathway Engineering. *ChemBioChem 23*.

(3) Medina-Cucurella, A. V., Steiner, P. J., Faber, M. S., Beltrán, J., Borelli, A. N., Kirby, M. B., Cutler, S. R., and Whitehead, T. A. (2019) User-defined single pot mutagenesis using unamplified oligo pools. *Protein Engineering, Design and Selection 32*, 41–45.

(4) Tresnak, D. T., and Hackel, B. J. (2023) Deep Antimicrobial Activity and Stability Analysis Inform Lysin Sequence–Function Mapping. *ACS Synth Biol 12*, 249–264.

(5) Hiraga, K., and Arnold, F. H. (2003) General Method for Sequence-independent Site-directed Chimeragenesis. *J Mol Biol 330*, 287–296.

(6) Szybalski, W., Kim, S. C., Hasan, N., and Podhajska, A. J. (1991) Class-IIS restriction enzymes — a review. *Gene 100*, 13–26.

(7) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One 3*, e3647.

(8) 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.

(9) Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS One 6*, e16765.

(10) Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth Biol 4*, 975–986.

(11) Bird, J. E., Marles-Wright, J., and Giachino, A. (2022) A User's Guide to Golden Gate Cloning Methods and Standards. *ACS Synth Biol 11*, 3551–3563.

(12) Sellmann, C., Pekar, L., Bauer, C., Ciesielski, E., Krah, S., Becker, S., Toleikis, L., Kügler, J., Frenzel, A., Valldorf, B., Hust, M., and Zielonka, S. (2020) A One-Step Process for the Construction of Phage Display scFv and VHH Libraries. *Mol Biotechnol 62*, 228–239.

(13) Chockalingam, K., Peng, Z., Vuong, C. N., Berghman, L. R., and Chen, Z. (2020) Golden Gate assembly with a bi-directional promoter (GBid): A simple, scalable method for phage display Fab library creation. *Sci Rep 10*, 2888.

(14) Gerstmans, H., Grimon, D., Gutiérrez, D., Lood, C., Rodríguez, A., van Noort, V., Lammertyn, J., Lavigne, R., and Briers, Y. (2020) A VersaTile-driven platform for rapid hit-to-lead development of engineered lysins. *Sci Adv 6*.

(15) Coussement, P., Bauwens, D., Maertens, J., and De Mey, M. (2017) Direct Combinatorial Pathway Optimization. *ACS Synth Biol 6*, 224–232.

(16) Macdonald, C. B., Nedrud, D., Grimes, P. R., Trinidad, D., Fraser, J. S., and Coyote-Maestas, W. (2023) DIMPLE: deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. *Genome Biol 24*, 36.

(17) Püllmann, P., Ulpinnis, C., Marillonnet, S., Gruetzner, R., Neumann, S., and Weissenborn, M. J. (2019) Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci Rep 9*, 10932.

(18) Wrenbeck, E. E., Klesmith, J. R., Stapleton, J. A., Adeniran, A., Tyo, K. E. J., and Whitehead, T. A. (2016) Plasmid-based one-pot saturation mutagenesis. *Nat Methods 13*, 928–930.

(19) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods 6*, 343–345.

(20) Francino-Urdaniz, I. M., Steiner, P. J., Kirby, M. B., Zhao, F., Haas, C. M., Barman, S., Rhodes, E. R., Leonard, A. C., Peng, L., Sprenger, K. G., Jardine, J. G., and Whitehead, T. A. (2021) One-shot identification of SARS-CoV-2 S RBD escape mutants using yeast screening. *Cell Rep 36*, 109627.

(21) Kowalsky, C. A., Faber, M. S., Nath, A., Dann, H. E., Kelly, V. W., Liu, L., Shanker, P., Wagner, E. K., Maynard, J. A., Chan, C., and Whitehead, T. A. (2015) Rapid fine conformational epitope mapping using comprehensive mutagenesis and deep sequencing. *Journal of Biological Chemistry 290*, 26457–26470.

(22) Haas, C. M., Francino-Urdaniz, I. M., Steiner, P. J., and Whitehead, T. A. (2021) Identification of SARS-CoV-2 S RBD escape mutants using yeast screening and deep mutational scanning. *STAR Protoc 2*, 100869.

(23) Magoč, T., and Salzberg, S. L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics 27*, 2957–2963.

(24) Potapov, V., Ong, J. L., Kucera, R. B., Langhorst, B. W., Bilotti, K., Pryor, J. M., Cantor, E. J., Canton, B., Knight, T. F., Evans, T. C., and Lohman, G. J. S. (2018) Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synth Biol 7*, 2665–2674.

(25) Strawn, I. K., Steiner, P. J., Newton, M. S., Baumer, Z. T., and Whitehead, T. A. (2023) A method for generating user-defined circular single-stranded DNA from plasmid DNA using Golden Gate intramolecular ligation. *Biotechnol Bioeng*.

(26) Beltrán, J., Steiner, P. J., Bedewitz, M., Wei, S., Peterson, F. C., Li, Z., Hughes, B. E., Hartley, Z., Robertson, N. R., Medina-Cucurella, A. V., Baumer, Z. T., Leonard, A. C., Park, S.-Y., Volkman, B. F., Nusinow, D. A., Zhong, W., Wheeldon, I., Cutler, S. R., and Whitehead, T. A. (2022) Rapid biosensor development using plant hormone receptors as reprogrammable scaffolds. *Nat Biotechnol 40*, 1855–1861.

(27) Rogers, T. F., Zhao, F., Huang, D., Beutler, N., Burns, A., He, W., Limbo, O., Smith, C., Song, G., Woehl, J., Yang, L., Abbott, R. K., Callaghan, S., Garcia, E., Hurtado, J., Parren, M., Peng, L., Ramirez, S., Ricketts, J., Ricciardi, M. J., Rawlings, S. A., Wu, N. C., Yuan, M., Smith, D. M., Nemazee, D., Teijaro, J. R., Voss, J. E., Wilson, I. A., Andrabi, R., Briney, B., Landais, E., Sok, D., Jardine, J. G., and Burton, D. R. (2020) Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science (1979) 369*, 956–963.

(28) Kucera, R., and Cantor, E. (2018) Breaking through the Limitations of Golden Gate Assembly– The Co-Evolution of Test Systems, Engineered Enzymes and Understanding Ligase Fidelity. *NEB*.

(29) Varadarajan, N., Cantor, J. R., Georgiou, G., and Iverson, B. L. (2009) Construction and flow cytometric screening of targeted enzyme libraries. *Nat Protoc 4*, 893–901.

(30) Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z.-G., Opperman, D. J., Reetz, M. T., and Acevedo, J. P. (2013) Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth Biol 2*, 83–92.

(31) Bosley, A. D., and Ostermeier, M. (2005) Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol Eng 22*, 57–61.

(32) Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M., and Wittrup, K. D. (2006) Isolating and engineering human antibodies using yeast surface display. *Nat Protoc 1*, 755–768.

Supporting Information to

Standardizing cassette-based deep mutagenesis by Golden Gate assembly

Nicolas Daffern[1‡], Irene Francino-Urdaniz[1‡], Zachary T. Baumer[1‡], Timothy A. Whitehead[1,*]

[1]Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80305, USA

[‡] Authors contributed equally to this work

This supplementary information contains:

Supplemental Note 1: Detailed Golden Gate-based cassette mutagenesis protocol

Supplemental Note 2: Analysis of large library size for T7 RNAP from limited sequencing depth

Supplementary Figure 1: DNA sequences involved in construction of the GG-RBD library 3

Supplementary Figure 2: Comparison of Golden Gate and Gibson Assembly protocols

Supplemental Table 1: Table of all Golden Gate reactions performed

Supplemental Table 2: Primer sequences

Supplemental Table 3: Library statistics for S RBD sequenced libraries

Supplemental Data: Plasmid and cassette sequences.

Supplemental Note 1: User-defined library generation by Golden Gate assembly

BACKGROUND & PROTOCOL OVERVIEW

Golden Gate assembly is a one-pot DNA assembly technique that uses temperature cycling with a type IIS restriction enzyme (BsaI-HFv2 reported here), T4 ligase, and input DNA to clone multiple DNA fragments sequentially into a vector. Generally, for this protocol, a PCR thermocycler cycles between a 37 °C step and a 16 °C step. During the 37 °C step BsaI generates four base pair overhangs downstream of its GGTCTC(X) binding motif and during the 16 °C step the reannealed DNA is ligated by T4 ligase. BsaI sites on the DNA fragments are designed such that overhangs will match complementary overhangs on upstream and downstream fragments or vector ends, allowing the fragments to be assembled in order. Additionally, the orientation of the BsaI sites on the vector and fragments are such that they are removed from the final product. Thus, cycling of these cutting and ligating steps continues to act on the input DNA and leaves the final product untouched, resulting in high incorporation efficiencies.

Here we detail a version of this protocol that can be applied to any vector and protein coding sequence to generate large, site-specific, combinatorial libraries by assembling synthetic mutagenic DNA fragments that we denote cassettes into a destination vector. Cassettes are linear dsDNA and can be purchased as synthetic dsDNA (e.g. gBlocks/eBlocks from IDT) or assembled from oligonucleotide or oligo pool ssDNA. Destination vectors have some, or all, of the protein coding sequence replaced with a selective marker, such that mutated cassettes coding for that sequence replace the marker during the Golden Gate reaction, resulting in functional plasmids. Users can follow our protocol to make their own destination vector or use one of our premade vectors; our protocol provides a detailed explanation for designing cassettes that work with pND003, pND004, and pND005. After the destination vector is obtained, cassette(s) (one to four demonstrated) are generated from ssDNA by PCR and co-incubated with the destination vector in a Golden Gate reaction to assemble the library, which is then transformed into *E. coli*. We describe a streamlined protocol such that, once the destination vector is in hand, new libraries can be built in a single day.

The protocol is listed in three sections: (1.) Design and construction of the destination vector; (2.) Design and construction of double stranded cassette DNA; (3.) Golden Gate assembly and transformation into *E. coli*. Additionally, we provide examples for different steps based on the creation of the SARS-CoV-2 S RBD (333-541) yeast display libraries (RBD libraries).

MATERIALS AND SUPPLIES

*All customized DNA fragments were purchased from IDT

Item	Supplier	Catalog number

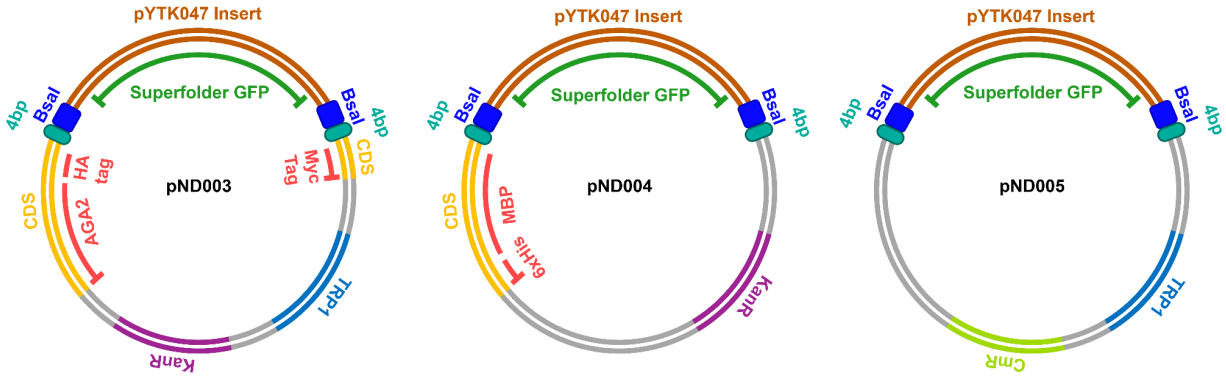
Section 1: Construction of destination vector		
pETconNK	Klesmith et al. 2017[1]	https://www.addgene.org/81169/
knock-out gene	Lee et al. 2015[2]	https://www.addgene.org/65154/
Section 2: Construction of dsDNA cassettes		
Cassette 1 ultramer/opool	This study	Supplementary Data
Cassette 2 ultramer/opool	This study	Supplementary Data
Cassette 3 ultramer/opool	This study	Supplementary Data
Forward primer	This study	Supplementary Data
Q5 High-Fidelity 2X Master Mix	NEB	M0492S
KAPA HiFi Ready Mix	Roche/Fisher	7958927001
Nuclease Free Water	IDT	11-05-01-14
Ultrapure Agarose	Invitrogen	16500-500
TAE Buffer (Tris-acetate-EDTA) (50X)	Thermo Scientific	B49
SYBR Safe DNA gel stain	Invitrogen	S33102
Gel loading dye, Purple	NEB	B7024
Nucleo-Spin® Gel and PCR Clean-up kit	Macherey-Nagel	740609.250
Section 3: Golden Gate assembly and transformation into *E. coli*		
Destination vector	This study	Step 1
dsDNA Cassette 1	This study	Step 2
dsDNA Cassette 2	This study	Step 2
dsDNA Cassette 3	This study	Step 2
BsaI-HF v2	NEB	R3733S

T4 DNA ligase	NEB	M0202S
T4 DNA ligase reaction buffer	NEB	B0202S
Nuclease Free Water	IDT	11-05-01-14
XL1-Blue Competent Cells	Agilent	200249
TransforMax EC100 Competent Cells	Biosearch Technologies	EC10010
Corning® Square Bioassay dish	Sigma	CLS431272
Glass Beads	Sigma	G8772
Kanamycin	GoldBio	K-120-25
ZymoPURE II Plasmid Midiprep kit	Zymo Research	D4201

PROTOCOL

Section 1: Design and construction of the destination vector

Background: We built three destination vectors (MBP *E. coli* expression, yeast surface display, yeast two-hybrid), each containing a GFP marker and BsaI recognition sequences next to high-fidelity overhangs **(Supplemental Note, Figure 1)**. If any of these destination vectors are suitable, section 1 can be skipped.



Supplemental Note, Figure 1. *Premade general-use destination vectors for Golden Gate-based cassette mutagenesis*

If these destination vectors are unsuitable, you can create your own destination vector. The following describes how to make a destination vector using Gibson Assembly, although a variety of cloning techniques could be used.

1: Pick a backbone

Pick a backbone for your destination vector which contains the protein of interest and the desired vector components (tags, promoters, selection markers, etc). Check to see if the vector already contains BsaI sites. If so, these should be removed by site-directed mutagenesis before proceeding.

2: Choose a selection marker

Our lab has used both GFP and RFP but other selection markers, such as LacZ[3,4], LacI[4], □-galactosidase[4] or sacB[5], have been demonstrated to be effective for GG assembly.

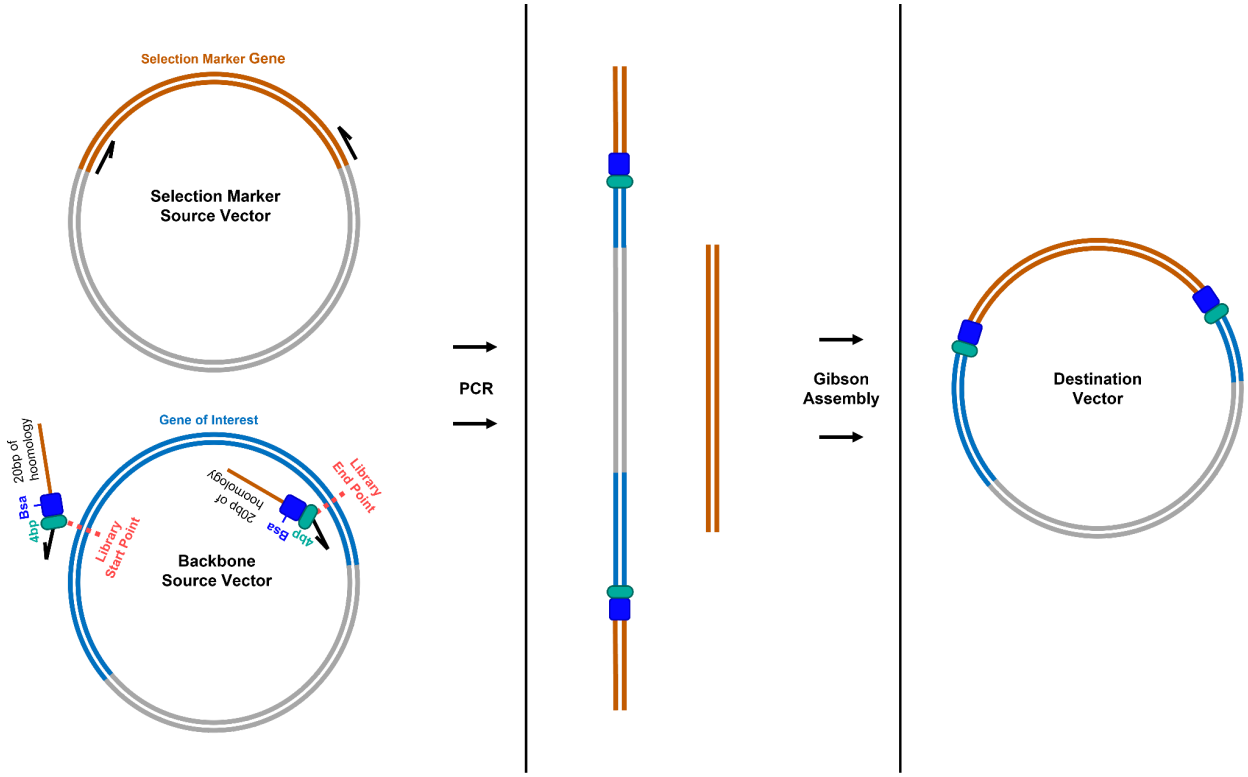
3: Choose the start and stop points for your cassettes

Before picking the start and end points for your library, you must choose the general region covered by your cassettes. Here you have two options. If you already have your protein in the desired vector, your cassettes can cover only the coding sequence containing your mutation sites. This allows for the fewest number of cassettes, is the simplest to design, and is ideal for large protein-coding sequences but requires that a new destination vector be created if you mutate a different region of your protein. Alternatively, you can have your cassettes begin and end outside the protein coding region, such that any library can be made for any protein without building a new destination vector (as with our premade destination vectors).

Once you have chosen the region covered by your cassettes, pick a set of high-fidelity 4 base pair overhangs from Table 1 in Potapov *et al*[4]. Then find a four base pair sequence near the beginning and end of that region from that set. These will be the start and end points for your cassettes.

4: Design primers for creation of the destination vector

Design primers to clone the desired selective marker and BsaI sites into the destination vector, in place of the backbone sequence covered by the cassettes. If using Gibson Assembly, this includes two primers to amplify the selection marker from its source as well as two primers to amplify the backbone vector. The primer set for amplifying the backbone should add BsaI sites directly adjacent to the first and last overhangs and 20 base pairs of homology with the selection marker from each end of the PCR fragment. Importantly, the BsaI sites should be in between the overhang and selection marker sequence (**Supplemental Note, Figure 2**).



Supplemental Note, Figure 2. *Building a destination vector using Gibson Assembly*

4. Create the destination vector

If using Gibson Assembly, the template vector and selection marker insert should be amplified using a standard PCR reaction. The resulting PCR products are then gel purified and combined by Gibson Assembly[6].

<u>Section 2: Design and construction of cassettes</u>

Background: After designing a destination vector or choosing to use one of our premade vectors, you must design cassettes that contain your mutated protein-coding sequence, high-fidelity overhangs, and a BsaI recognition sequence. Additional sequences must be included when designing cassettes for use with pND003, pND004, or pND005 (described below). These cassettes are ordered as synthetic ssDNA fragments and are converted to dsDNA using a reverse primer.

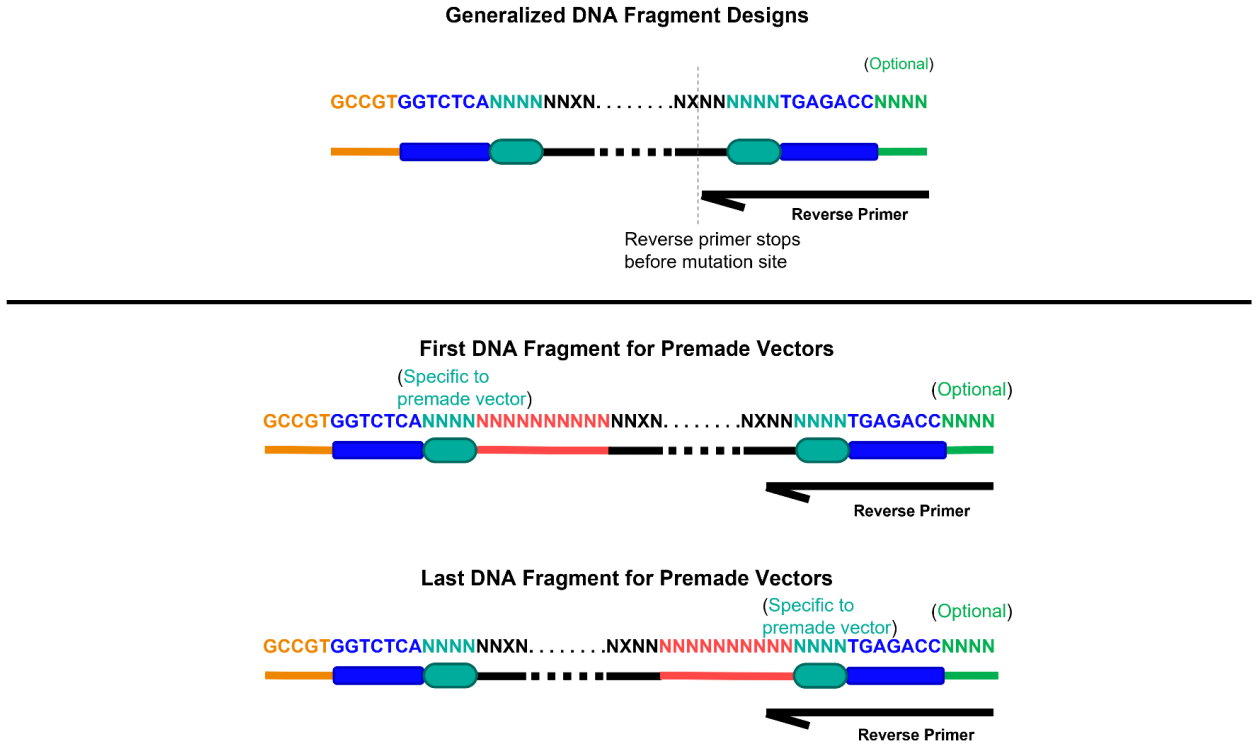
1: Cassette designs

The first step in designing cassettes is choosing your library's start and end points with respect to your protein coding sequence. If you created your own destination vector, as per section 1, these have already been chosen. If you are using our premade vectors, these can be chosen

arbitrarily. Next, you must decide the number and length of your cassettes, which is usually based on the maximum length of the synthetic DNA fragments you are using (It is important to remember that the length of each fragment must include at least 7 base pairs on either side of the coding sequencing for the BsaI site). This will then determine the general areas in your coding sequence where each cassette will begin and end.

Within these areas, look for four base pair sequences that match high-fidelity overhangs from the set you used for the destination vector or that are listed for our premade vectors at the bottom of **Supplemental Note, Table 1**. Note that all overhangs must be from the same set, cannot be used more than once and that residues coded for by base pairs in the overhangs cannot be mutated. These overhang sequences will be the starting and ending points of each cassette. If designed correctly, the first and last four protein-coding base pairs of each cassette will match the first or last four protein-coding base pairs of the adjacent cassettes (see Example 1). Next, you will add BsaI recognition sequences directly adjacent to the overhangs **(Supplemental Note, Figure 3)**. Note, it is a good idea to check that any mutations you make do not introduce a BsaI site into your coding sequence. We also added short sequences outside of the BsaI site, if no other sequence was present, to ensure BsaI binding; we have not tested whether this is necessary to achieve high transformation efficiencies.

Short sequence we included to support binding of BsaI to its recognition sequence. Removing it may not affect efficiencies but has not tested by us.	GCCGT	Only needed when working with our premade destination vectors or designing destination vectors with cassettes that extend beyond the protein coding sequence. Constant sequences that code for promoters, stop codons, terminators, and tags in source vectors. Exact sequences that match our premade vectors are found in **Supplemental Table 3** (colored in red)**.**	NNNNNNNNNN
BsaI recognition sequences (forward/ reverse).	GGTCTCA/TGAGACC	User's protein coding sequence. X represents the first and last mutation sites in the cassette	NNXN. . . . . .NXNN
4 base pair high-fidelity overhang sequence. When working with our premade destination vectors, the initial overhang in the first fragment and final overhang in the last fragment must contain a predetermined overhang that matches the vector. Exact sequences are found in **Supplemental Note, Table 1** (colored in cyan)**.**	NNNN	Optional sequence that can be added if there are not enough bases after the last mutation site to design a reverse primer with an appropriate Tm	NNNN



Supplemental Note, Figure 3. *Generalized design schematic of cassettes for Golden Gate-based cassette mutagenesis*

2: Cassettes designs for use with pND003, pND004, and pND005

Skip if not using our premade destination vectors.

Cassettes that will be used with pND003, pND004, or pND005 should be designed as described above, with two important exceptions:

First, cassettes should use overhang sequences as outlined in **Supplemental Note, Table 1**. All our destination vectors were made with high-fidelity overhangs from Set 1 from the main text's Table 1 in Potapov *et al*[4]. Once you have chosen a destination vector, the initial overhang from your first cassette and the final overhang from your last cassette should be the ones listed in the sequence for the chosen vector, as seen in **Supplemental Note, Table 1**. The rest of your overhangs should come from the remaining sequences in Set 1, found at the bottom of **Supplemental Note, Table 1**.

Second, your first and last cassettes should include additional sequences, specific to the chosen destination vector, that are shown in red in **Supplemental Note, Table 1.** These code for promoters, terminators, and continuations of tag sequences leading up to the overhang sequences. Note that stop codons do not need to be added to your sequence as these are already coded for in each vector.

	First Cassette	Last Cassette
pND003	GCCGTGGTCTCaCGGTAGCGGAGGCGG AGGGTCGGCTAGCCATNNNN……NNN NNNNNtGAGACCNNNN	GCCGTGGTCTCaNNNNNNNNN……NNNN CTCGAGGGGGGCGGATCCGAATGAGACCN NNNN
pND004	GCCGTGGTCTCaCGGTCGTCAGACTGTC GATGAAGCCCTGAAAGACGCGCAGACT GGCGGCGGGGGAGGTNNNN…….NN NNNNNNNtGAGACCNNNN	GCCGTGGTCTCaNNNNNNNNN……NNNNT GAGATCCGGCTGCTAACAAAGCCCGAAAG GAtGAGACCNNNN
pND005	GCCGTGGTCTCaTCCTGAAAGNNNN… ….NNNNNNNNNtGAGACCNNNN	GCCGTGGTCTCaNNNNNNNNN……NNNNT GAGTCGTATTACCTCAGCCAAGCTAATTCtG AGACCNNNN

High-fidelity overhangs that can be used in conjunction with pND003, pND004, and pND005	**TGCC, GCAA, ACTA, TTAC, CAGA, TGTG, GAGC, AGGA, ATTC, CGAA, ATAG, AAGG, AACT, AAAA, ACCG**

Supplemental Note, Table 1. *Sequence specific designs of cassettes for use with pND003, pND004, and pND005. Coloring of bases corresponds to schematics in Supplemental Note, Figure 3.*

Tips & tricks: A main area that this protocol will fail is incorrectly designing sequences and verifying the BsaI cut site. BsaI recognizes GGTCTC and cuts one base pair down, leaving the nucleotides 2-5 bases downstream as the 5' overhang on the adjacent strand. This is commonly denoted by the shorthand: GGTCTCN^NNNN_. It is always a good idea to create DNA fragments and run a "virtual" digest in a readily available online tool (e.g. Benchling or ApE - A plasmid Editor, and verify the resulting sequence leads to a proper open reading frame free of early stop codons or frameshifts.

3: Design primers for creation of double stranded cassette DNA

Next, a single reverse primer must be designed to create double stranded cassette DNA. The primer should not overlap with any mutation site. If mutation sites are located close to the overhang sequences, additional bases might need to be added to the end of a cassette to have a long enough complementary sequence for primer binding (**Supplemental Figure 3**). In this study, we used primers with melting temperatures (T_m) ranging from 56°C to 72°C with comparable results. For the RBD library, we used NEB Tm calculator to determine the primer melting temperature and the IDT Oligoanalyzer™ tool to analyze the possible hetero-dimers and the Gibbs free energy of the designed primers. We designed primers to avoid large negative ΔG of any possible mismatch (homodimer, unintended heterodimer, secondary structure) limiting this to > -22 kcal/mol. Additionally, we designed the ΔΔG of the intended heterodimer to any other structure to be <-20 kcal/mol. Inspection of the types of secondary structures may also be helpful to look for sequences which can amplify in an unintended manner from an annealed 3' end (e.g. a hairpin or dimer that anneals the 3' end with an extended 5' end can serve as a template for amplifying products like "primer dimers").

4. Generate cassette dsDNA from synthetic ssDNA.

Our experimental protocol below starts at this step, in which dsDNA is generated from single-stranded synthetic DNA fragments. We found that smearing sometimes occurred but that using at least 35 PCR cycles leads to sufficient amplification of the desired product over background. The PCR products are then gel purified and can be stored until the next step.

Day 1

1. Dilute the synthetic DNA fragments and the reverse primers to a final concentration of 10 µM in Nuclease Free Water. (See Example 2 for equations used to resuspend the synthetic DNA)

   Note: The IDT Resuspension Calculator tool can be used.

2. Generate dsDNA cassettes

   a. Assemble the PCR reactions as follows:

Cassette (10 µM)	1.25 µl

Primer reverse (10 µM)	1.25 µl
Q5 2X Master Mix	12.5 µl
Nuclease Free Water	10 µl
Reaction volume	**25 µl**

Note: We have used KAPA HiFi Ready Mix (Roche) and Phusion polymerase (NEB) with comparable results.

PCR cycling conditions			
Steps	**Temperature**	**Time**	**Cycles**
Initial Denaturation	98 °C	30 s	1
Denaturation	98 °C	10 s	35 cycles
Annealing	56°C*	30s	
Extension	72 °C	1:30 min	
Final extension	72 °C	2 min	1
Hold	4 °C	hold	

***Note:** We have demonstrated generation of dsDNA using primers with a Tm of 56°C as well as 72°C.
Tips & tricks: Set the annealing temperature at the optimal temperature for the designed primers. The NEB Tm calculator can be used.

b. Purify linear DNA by gel electrophoresis and gel extraction. One clear band is expected for each fragment. Purify all excised fragments using a Nucleo-Spin® Gel and PCR Clean-up kit (Macherey-Nagel). Elute in 20µl Nuclease free water. (See Example 3; if trouble getting the right size band see Troubleshooting 1)

   Tips & tricks: If the concentration obtained is not high enough for the Golden Gate reaction, reload the eluent in the column.

c. Quantify the linear DNA products by measuring A_{260} using a spectrophotometer. We often obtain concentrations of approx. 40 ng/µl.

Section 3: Golden Gate assembly and transformation into *E. Coli*

Background. Use a Golden Gate reaction to assemble the dsDNA fragments into your destination vector. This section takes about 5 hours, including bacterial transformation, and can be combined with section 2 for a single day library generation protocol

Day 1 (Cont.)

1. Assemble the pieces with a Golden Gate reaction.

    a. Per 40 fmol of input DNA, assemble the reaction as follows, adding the enzymes last. (See Example 4 for the calculations used)

dsDNA fragments	40 fmol each
Destination vector	40 fmol
10x T4 DNA ligase buffer	2.5 µl
BsaI-HF-v2	1 µl (20 units)
T4 DNA ligase	1 µl (400 units)
Nuclease Free Water	up to 25µl
Reaction volume	**25 µl**

Note: input DNA or total reaction size can be scaled up linearly, as shown in the main text.

Note: NEBridge Golden Gate Assembly kit (NEB) can also be used with comparable results following manufacturer's instructions.

Tips & tricks: The 10x T4 DNA ligase buffer should not have gone through multiple freeze thaw cycles. Making aliquots of the stock is recommended.

    b. Set the reaction on a thermal cycler using the conditions below:

Thermal cycler conditions		
Temperature	**Time**	**Cycles**
37 °C	1 min	60 cycles
16 °C	1 min	
37 °C	5 min	1
65 °C	10 min	1
4 °C	infinity and beyond	

Tips & tricks: The inclusion of an additional 5 minute at the 37 C cutting step significantly reduces the amount of incorrectly assembled sequences as any cut DNA will not transform efficiently.

    c. To concentrate the DNA, perform a PCR clean-up using a Monarch® PCR & DNA clean-up kit (NEB), eluting in 6 µl of nuclease free water.

2. Transform into *E. coli*.

a. Transform the 6 µl of eluent into electrocompetent cells. (See Example 5)

    i. Optionally, transform 0.1 ng of pUC18 (XL1 Blue) or pUC19 (TransforMax) into electrocompetent cells to assess competency.

b. Recover the cells in 1 ml SOC or TB media for 1 h at 37 °C with shaking at ~300 rpm.

   Tips & tricks: Use 2X 500µl of SOC or TB to resuspend the cells from the cuvette in order to better wash the cuvette and recover as many cells as possible.

c. Do 6 serial dilutions of the recovered cells and plate 10 µl of each dilution on an LB agar plate, containing the appropriate antibiotic, to assess transformation efficiencies. Plate the remainder of the cells on a LB agar square Bioassay dish, containing the appropriate antibiotic, using plating beads. Incubate the plates at 37 °C overnight.

   Tips & tricks: Running appropriate controls (e.g. pUC18 or pUC19 and undigested destination vector) will help identify whether the competency of the cells is low, or if the plasmid size is causing issues with reduced transformation efficiency.

Day 2

3. Check the transformation efficiency and if it is above the desired efficiency, scrape the Square Bioassay dish. Add ~3 ml of LB to the plate, using an L shaped cell scraper or a bended glass pasteur pipette, scrape the cells off the agar plate. Collect the cells in a corner and transfer them to a 50ml falcon tube by pipetting. Repeat this process five to seven times for a total 15-20 ml cell volume or until the plate is clear of cells. (See Example 6 to calculate the desired transformation efficiency)
   Note: Depending on the destination vector (especially dependent on plasmid copy number), it might take up to two days to see green colonies. A good control to compare fluorescence of the uncut destination vector is to transform the neat destination vector by itself, this can also reveal issues with low transformation efficiency as a result of large destination vectors.

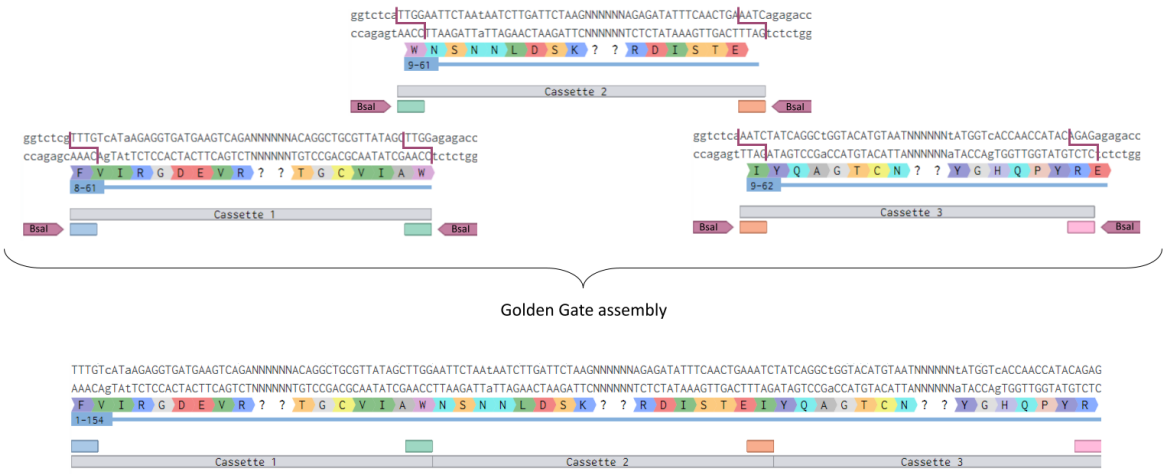
4. Make *E. Coli* glycerol stocks by spinning down 1.5 ml of the scraped cells and resuspending in 700 µl LB media and 300 µl 50 w/v% glycerol.
   Note: Since the desired transformation efficiency is 100 fold above the library size, 1.5ml of the scraped cells will have a good representation of the library.

5. Prepare 10 mL of the scraped cells using a ZymoPURE Midiprep kit (Zymo Research) to obtain the plasmid library.

EXAMPLES

Example 1:

In the case of the RBD library, positions 333-416 and 510-541 were not mutated, therefore, they are part of the destination vector. For the RBD we wanted to mutate separately the epitopes of class 1 and class 2 antibodies[7]. Therefore, we divided positions 400 to 509 into 3 cassettes: cassette 1 containing positions 400-436, cassette 2 containing positions 436-472 and cassette 3 containing positions 472-509. This means that positions 400, 436, 472 and 509 can not be mutated (**Figure S2**).



Supplemental Figure 4. *Schematic of cassette assembly and location of the BsaI overhangs. The BsaI overhang at the end of cassette 1 matches the beginning of cassette 2 (green). Since those 4 bp allow the assembly, the codon(s) containing the overhang nucleotides cannot be mutated. Notice that the BsaI sites are located on the outside of each cassette and are not part of the assembled product.*

Example 2:

If the ultramer ordered contains 4 nmol of DNA, resuspending in 200 µl will give a 20µM concentration.

$$4\ nmol \times \frac{1\ \mu mol}{10^3\ nmol} \times \frac{l}{20\ \mu mol} \times \frac{10^6 \mu l}{1\ l} = 200 \mu l\ \ The\ oligo\ pool\ ordered\ contained\ 10$$

pmol of DNA per individual oligo with 1644 oligos. Since we want the final concentration to be per pool:

$$10\ pmol \times 1644\ oligos \times \frac{1\ nmol}{10^3\ pmol} = \ 16.44\ nmol$$

Resuspending in 164.4 µl will give a 100µM concentration of the pool.

$$16.44\ nmol \times \frac{1\ \mu mol}{10^3\ nmol} \times \frac{l}{100\ \mu mol} \times \frac{10^6 \mu l}{1\ l} = 164.4\ \mu l$$

Example 3:

The fragments for the RBD are ~170 bp, so the 170bp band was excised and gel extracted.

Example 4:

*All dilutions are in nuclease free water

Cassette	Length (bp)	Concentration from PCR (ng/µl)
1	167	33.56
2	168	40.24
3	172	40.31

Cassette 1

$$167\ bp \times \frac{650\ Da}{1\ bp} \times \frac{1\ g}{1\ Da \cdot mol} \times 40 \cdot 10^{-15} mol \times \frac{10^9 ng}{1\ g} = 4.34 ng 4.34\ ng \times \frac{1\ \mu l}{33.56 ng} = 0.13\ \mu l$$

$\rightarrow$ 1.3 µl 1:10 dilution

Cassette 2

$$168\ bp \times \frac{650\ Da}{1\ bp} \times \frac{1\ g}{1\ Da \cdot mol} \times 40 \cdot 10^{-15} mol \times \frac{10^9 ng}{1\ g} = 4.37 ng 4.37\ ng \times \frac{1\ \mu l}{40.24 ng} = 0.11\ \mu l$$

$\rightarrow$ 1.1 µl 1:10 dilution

Cassette 3

$$172\ bp \times \frac{650\ Da}{1\ bp} \times \frac{1\ g}{1\ Da \cdot mol} \times 40 \cdot 10^{-15} mol \times \frac{10^9 ng}{1\ g} = 4.47 ng 4.47\ ng \times \frac{1\ \mu l}{40.31\ ng} = 0.11\ \mu l$$

$\rightarrow$ 1.1 µl 1:10 dilution

Destination vector

$$7461\ bp \times \frac{650\ Da}{1\ bp} \times \frac{1\ g}{1\ Da \cdot mol} \times 40 \cdot 10^{-15} mol \times \frac{10^9 ng}{1\ g} = 139.98\ ng 139.98\ ng \times \frac{1\ \mu l}{599\ ng} = 0.23\ \mu l$$

$\rightarrow$ 2.3 µl 1:10 dilution

Cassette 1	1.3 µl
Cassette 2	1.1 µl
Cassette 3	1.1 µl

Destination vector	2.3 µl
T4 DNA ligase buffer	2.5 µl
BsaI-HF-v2	1 µl
T4 DNA ligase	1 µl
Nuclease Free Water	14.7 µl
Reaction volume	**25 µl**

Example 5:

XL1-Blue Competent Cells (Agilent) or TransforMax EC100 Competent Cells (Biosearch Technologies) have been used with comparable results. In our hands when transforming 0.1ng of pUC18 into XL1-Blue we obtain $1.0e^5$ transformants ($1.0e^9$ transformants/µg DNA) and when transforming 0.1ng of pUC19 into TransforMax EC100 we obtain $6.2e^5$ ($6.2e^9$ transformants/µg DNA).

Example 6:

The RBD libraries have a theoretical library size on the order of 10k protein-encoding variants. To ensure that all the variants are transformed, we want a 100-fold coverage. Therefore, the minimum desired transformation efficiency is $1e^6$.

TROUBLESHOOTING

Troubleshooting 1:

If a single band is not observed, increase the extension time and/or increase the number of cycles. If Q5 1X Master Mix (NEB) does not give the desired results, the reaction can be set up with KAPA HiFi Ready Mix (Roche). Alternatively, using "touchdown PCR"[8] may also minimize smearing.

Supplemental Note 2: *Estimate of library coverage for T7 RNAP from limited sequencing depth*

This technical note provides an estimate of the percent of the library coverage for the T7 RNAP experiments presented in Figure 4. We start from the limiting assumption that the frequency of mutation k at any of the nine codons j (F_{jk}) is uncorrelated with that from another codon. This leads to an estimate of the frequency of any variant i in the population (F_i) as

$$F_i = \prod_j F_{jk} \tag{1}$$

We determine the expected frequency of each variant i by an estimate of the frequency of each codon variant jk observed from limited deep sequencing (Figure 4D).

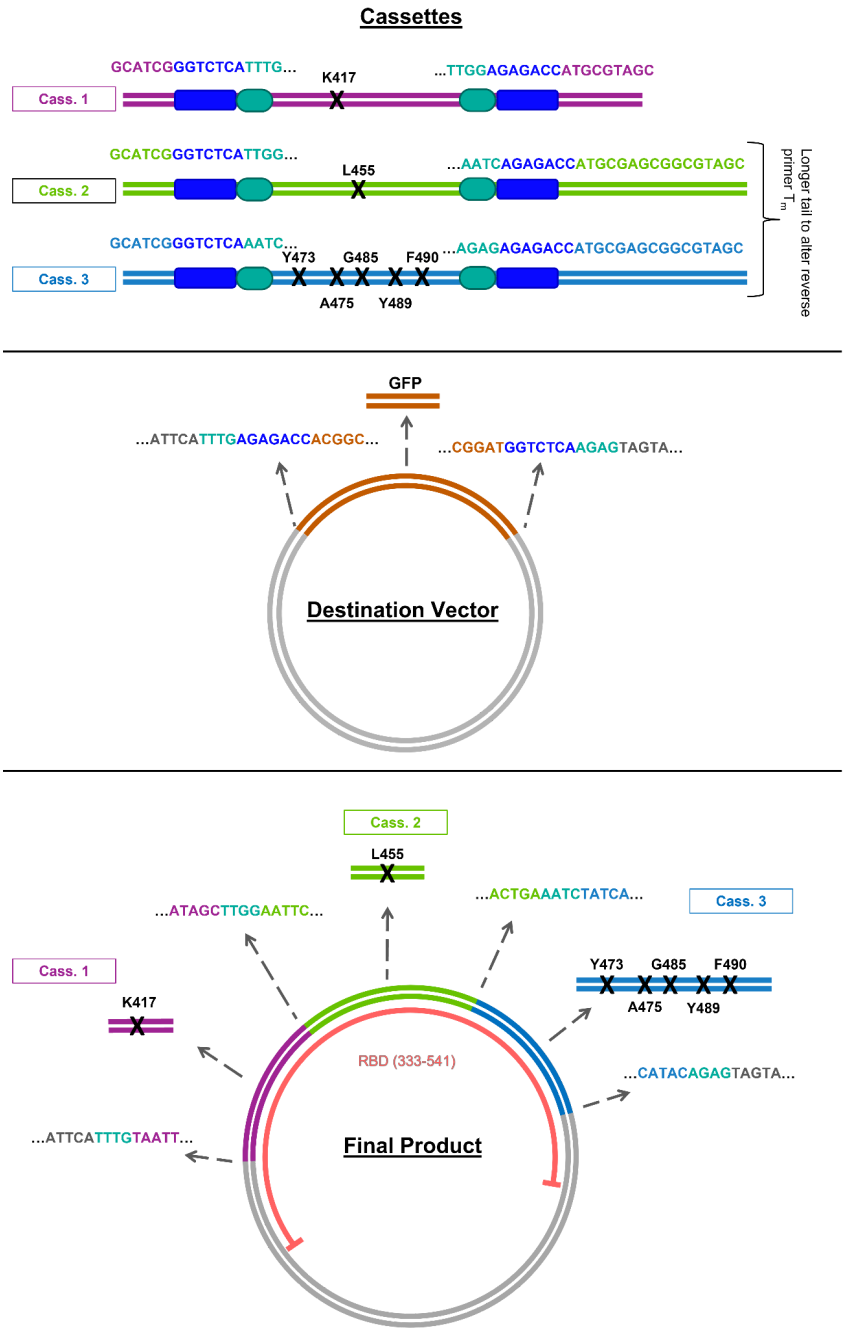
To estimate the probability (P_i) that any variant i is present in the library, we use a revision of equation 11 from Bosley & Ostermeier[12]:

$$P_i = 1 - (1 - F_i)^{0.692T} \tag{2}$$

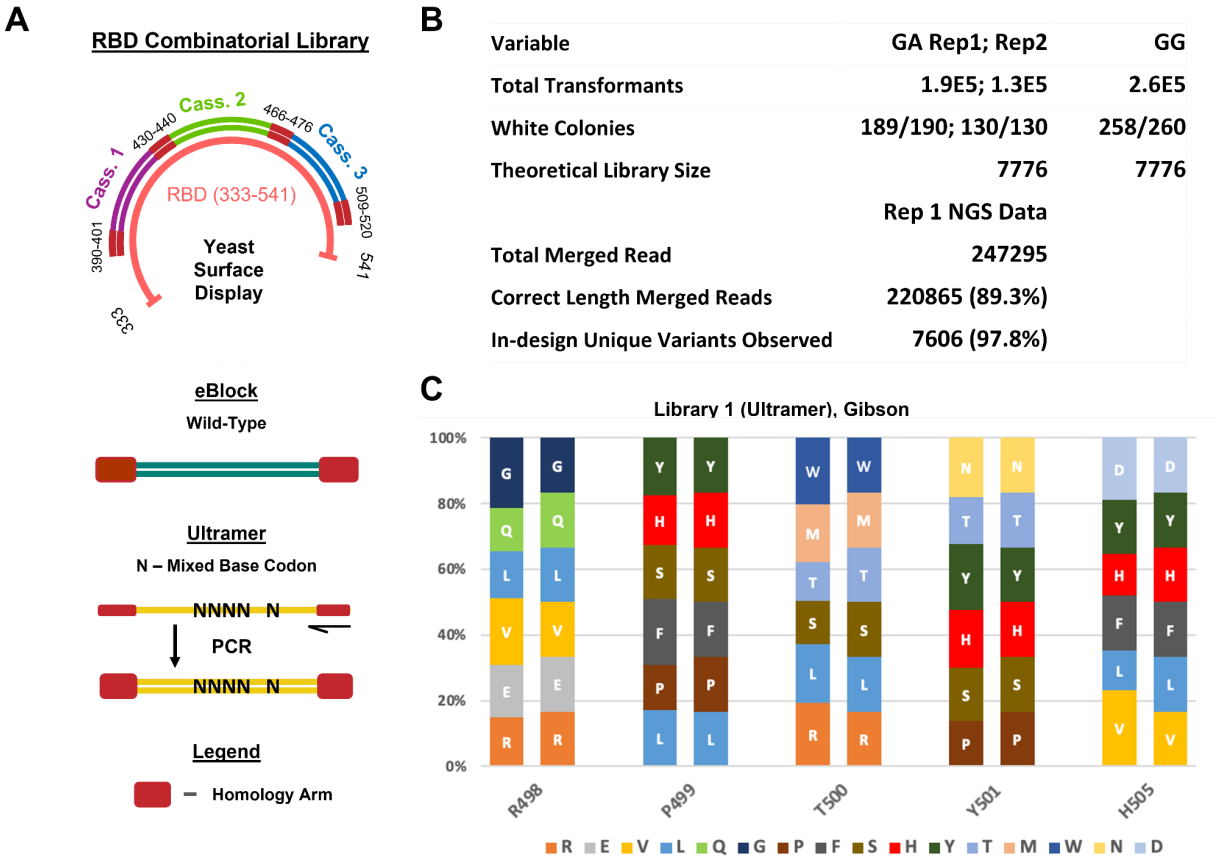
Where T is the number of transformants observed. Not all transformants result in on-target, full-length protein encoding sequences. Thus, the 0.692 in the equation is a conservative correction to the number of transformants observed, evaluated as "in-design reads", divided by "total filtered reads".

Finally, we estimate the percentage coverage by summing over the n variants expected in the library:

$$Estimated\ \%\ Coverage = \frac{1}{n}\sum_{i=1}^{n} P_i\, x\ 100\% \tag{3}$$



Supplementary Figure 1: **DNA sequences involved in construction of the GG-RBD library 3**
Schematic detailing the DNA sequences found in the cassettes, destination vector, and completed vector that were relevant for construction of the GG-RBD library 3 using oligo pools. Note that only three of the five mutations shown in cassette 3 were coded into any one oligo. Colors of the sequences correspond to their associated structure/function in the construct's architecture. Orange corresponds to parts of the GFP sequence, purple to RBD sequence on cassette 1, green to RBD sequence on cassette 2, light blue to RBD sequence on cassette 3, dark blue to the BsaI recognition sequences, cyan to RBD sequence used for the four base pair overhangs, and gray to RBD sequence found in the destination vector.



Supplementary Figure 2: **Comparison of Golden Gate and Gibson Assembly protocols for library generation** ***A.*** *Cartoon of the combinatorial library for the S RBD (333-541) constructed from Gibson assembly. Cassette 1 (purple) contains positions 390 – 440 with 30 (5') or 34 (3') bp homology arms on either end. Cassette 2 (green) contains positions 430-476 and has 30 bp homology arms on either end. Cassette 3 (blue) contains positions 466 – 520 with 30 (5') or 34 (3') bp homology arms on either end. DNA encoding positions 333-401 and 510-541 are part of the destination vector (grey) and contain homology arms to cassette 1 and 3, respectively. For comparison between Golden Gate and Gibson Assembly, Ultramer Library 1 was used which contains 7776 unique variants, all encoded on cassette 3.* ***B.*** *Tabulated results of the Gibson Assembly for comparison with Golden Gate Assembly* ***C.*** *Comparison of the observed (left column) vs expected (right column) diversity of mutations at each position. Out-of-design mutations accounted for less than 0.052% at each location, except for position 505. Position 505 had cysteine at 1.26% and glycine at 1.60%, but were expected to be 0% of the designed library.*

Supplementary Table 1

Experiment	Rep.	Destination Vector	Insert(s)	Input DNA (fmol)	Size (µl)	Colonies	% GFP Negative	pUC19 Colonies /µg DNA
Transformants / Cassette Number (Fig. 2A)	1	pND003	PYR1 Cass1+2+3+4	40	25	$8.4x10^5$	99.9	$3.7x10^9$ *(Rep. 1)* $4.5x10^9$ *(Rep. 2)*
	2		PYR1 Cass1+2+3+4			$8.4x10^5$	99.9	
	1		PYR1 Cass1+2+3, PYR1 Cass4			$4.9x10^5$	99.8	
	2		PYR1 Cass1+2+3, PYR1 Cass4			$7.2x10^5$	99.9	
	1		PYR1 Cass1+2, PYR1 Cass3, PYR1 Cass4			$3.4x10^5$	99.4	
	2		PYR1 Cass1+2, PYR1 Cass3, PYR1 Cass4			$6.0x10^5$	99.9	
	1		PYR1 Cass1, 2, 3, 4			$3.5x10^5$	99.7	
	2		PYR1 Cass1, 2, 3, 4			$3.9x10^5$	99.7	
GG RBD Function Test (Fig. 1D, 1E)	1	pIFU037	RBD WT eBlock (1+2+3)	40	20	$3x10^3$	80.0	NA
	2					$1x10^4$	91.6	
	3					$1.1x10^6$	96.0	
	4					$5.3x10^5$	98.0	
Transformants / Reaction Size (Fig. 2B)	1	pND003	PYR1 Cass1+2+3+4	40	25	$3.2x10^5$	96.9	$8.3x10^9$ *(Rep. 1)* $4.9x10^9$ *(Rep. 2)* $7.1x10^9$ *(Rep. 3)*
	2	pND003	PYR1 Cass1+2+3+4	40	25	$9x10^5$	98.9	
	3	pND003	PYR1 Cass1+2+3+4	40	25	$9x10^5$	98.9	
	1	pND003	PYR1 Cass1+2+3+4	160	100	$2x10^6$	99.5	
	2	pND003	PYR1 Cass1+2+3+4	160	100	$2.4x10^6$	99.6	
	3	pND003	PYR1 Cass1+2+3+4	160	100	$2.7x10^6$	99.6	
	1	pND003	PYR1 Cass1+2+3+4	320	200	$8.2x10^6$	99.9	
	2	pND003	PYR1 Cass1+2+3+4	320	200	$2.3x10^7$	99.9	
	3	pND003	PYR1 Cass1+2+3+4	320	200	$1.9x10^7$	99.9	
Transformants / Input DNA (Fig. 2C)	1	pND003	PYR1 Cass1+2+3+4	80	25	$7x10^5$	98.6	
	2	pND003	PYR1 Cass1+2+3+4	80	25	$2.9x10^6$	99.7	
	3	pND003	PYR1 Cass1+2+3+4	80	25	$4x10^6$	99.8	
	1	pND003	PYR1 Cass1+2+3+4	160	25	$5.8x10^6$	99.7	
	2	pND003	PYR1 Cass1+2+3+4	160	25	$2.7x10^7$	99.8	
	3	pND003	PYR1 Cass1+2+3+4	160	25	$6.2x10^6$	99.8	
RBD Library (Fig. 3A)	1	pIFU037	RBD ultramer 1+2+3	200	25	$5.6x10^6$	99.9	NA
	1		RBD ultramer 1+2+3			$2x10^6$	99.9	
	1		RBD oligo pool 1+2+3	400	50	$1.6x10^7$	98.0	
	2					$2.3x10^7$	92.0	
eBlock, Ultramer, oligo pool comparison (Fig. 3D)	1	pIFU037	WT RBD eBlock 1+2+3	200	25	$1.5x10^7$	99.9	$6.2x10^9$ *(Rep. 1)* $5x10^9$ *(Rep. 2)* $2.5x10^9$ *(Rep. 3)*
	2					$2.5x10^6$	99.9	
	3					$3.4x10^6$	99.9	
	1		WT RBD Ultramer 1+2+3			$3.4x10^6$	99.9	
	2					$1.5x10^6$	99.9	
	3					$1.4x10^6$	99.9	
	1		RBD oligo pool 1+2+3			$3.0x10^6$	99.9	
	2					$1.1x10^6$	99.9	
	3					$2.2x10^6$	99.9	
T7 Library (Fig. 4A)	1	pZB258	T7 Cass 1 + Cass 2	400	200	$5.6x10^7$	99.6	NA
RBD GA Library (Supp. Fig. S1)	1	pIFU037	RBD GA Ult 1+ Cass1 + Cass2	40	25	$1.9x10^5$	99.5	NA
	2	pIFU037	RBD GA Ult 1+ Cass1 + Cass2	40	25	$1.3x10^5$	99.9	NA
RBD Lib1 GG	1	pIFU037	RBD GG Ult 1+ Cass1 + Cass2	40	25	$2.6x10^5$	99.2	NA

Supplementary Table 2

Primer Name	Sequence
NDD-0126 pJS755 Backbone HA Rev	ACGTTTCACTTTCGGTCTCAACCGCTGGCGGCG
NDD-0125 pJS755 Backbone HA Fwd	TTCTGCGTTTATAGGTCTCAAGGAAGCTGAGTTGGCTGC
NDD-0124 pJS637 Backbone HA Rev	ACGTTTCACTTTCGGTCTCAACCGCCTCCACCAGAG
NDD-0123 pJS637 Backbone HA Fwd	TTCTGCGTTTATAGGTCTCACGAACAAAAGCTTATTTCTGAAGAGG
NDD-0122 pACL032 Backbone HA Rev	ACGTTTCACTTTCGGTCTCAAGGAGGCTTGCTTCAAGCT
NDD-0121 pACL032 Backbone HA Fwd	TTCTGCGTTTATAGGTCTCAATTCCGGGCGAATTTCTTATGATT
NDD-0120 GFP HA Rev	TGAGACCTATAAACGCAGAAAGG
NDD-0119 GFP HA Fwd	TGAGACCGAAAGTGAAACGT
NDD-0116 pJS755 BSa1 Removal Rev	CTTACCGACTTCAGCGAGTCCGTTATAGCCTTTATC
NDD-0115 pJS755 BSa1 Removal Fwd	GATAAAGGCTATAACGGACTCGCTGAAGTCGGTAAG
ZTB Cass1_for	GCATCGGGTCTCATTCGCAAGCGTTGCGCTGBNCATTGGGTAACTCCTGATGGTTTC CCTGBNTGGNDTGAATACAAGAAGCCTATTCAGACGCGCTTGAACCTGATGTTCCTC GGTCAGTTCCGCTTACAGCCTACCATTAACACCAACAAAGATAGCGAGAT
ZTB Cass1_Q737_for	GCATCGGGTCTCATTCGCAAGCGTTGCGCTGBNCATTGGGTAACTCCTGATGGTTTC CCTGBNTGGCAAGAATACAAGAAGCCTATTCAGACGCGCTTGAACCTGATGTTCCTC GGTCAGTTCCGCTTACAGCCTACCATTAACACCAACAAAGATAGCGAGAT
ZTB Cass1_rev	GCTACGGGTCTCAATCTCGCTATCTTTGTTGGTGTTAATGGTAGGCTGTAAGCGG
ZTB Cass2_for	GCATCGGGTCTCAAGATTGATGCACACAAACAGGAGTCTGGTVTAGCTCCTMRNND TGTACACRSYVRKGACGGTAGCCACCTTCGTAAGACTGTAGTGTGGGCACACGAGAA GTACGGAATCGAATCTTTTGCACTGATTCACGACTCCTTCGGTACCATTCCGGCTGAC GCTGCGAACCTGTTCAAAGCAGTGCGC
ZTB Cass2_rev	GCTACGGGTCTCAAGCGATCTGGTCGTAAHNATCAGCCAGTACATCACAAGACTCAT ATGTGTCAACCATAGTTTCGCGCACTGCTTTGAACAGGTTCGCAGCGTCAGCCGGAA TGGTACCGAAGGA
ZTB NGS_for	GTTCAGAGTTCTACAGTCCGACGATCGAGAGATTCTTCGCAAGCGTTGCG
ZTB NGS_rev	CCTTGGCACCCGAGAATTCCAGTGCAACTGGTCAGCGATCTGG
IFU-159 RBD cass1_rev	ACAGGCTGCGTTATAGCTTGGAGAGACCATAGATATGATGTAGCGTAGC
IFU-160 RBD cass2_rev	TTCGAGAGAGATATTTCAACTGAAATCAGAGACCATGCTAGAGCGTATGATGTAGCG TAGC
IFU-161 RBD cass3_rev	CAACCATACAGAGAGAGACCATGCTAGAGCGTATGATGTAGCGTAGC
IFU-162 RBD_fwd	GCATCGTTCCCTACTGCTCGTACATGC
IFU-163 RBD NGS_fwd	GTTCAGAGTTCTACAGTCCGACGATCTAGAGGTGATGAAGTCAGA
IFU-164 RBD NGS_rev	CCTTGGCACCCGAGAATTCCATACTACTACTCTGTATGGTTG
Forward Outer Primer	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATC
Reverse Outer Primer	GGCATACGAGATNNNNNNGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA ''NNNNNN'' is the barcode
ZB-p706 RBD GA Ultramer 1 Rev	ATGTAGAAGTTCAAAAGAAAGTACTACTACTCTGTATGGTTG;
ZB-p707 RBD NGS Lib Seq For	TCCCTACACGACGCTCTTCCGATCTGGTCCACAAACAGTTGCTGGTGC
ZB-p708 RBD NGS Lib Seq Rev	GTTCAGACGTGTGCTCTTCCGATCTGGAGTGTCTCCTACTAAATTAAATGATCTCTGC

Supplementary Table 3

	Library 1	**Library 2**	**Library 3**
DNA type	Ultramer	Ultramer	Oligo pool
Number of designed mutations	7776	20736	17725
Number of transformants obtained	$5.6x10^6$	$2x10^6$	$2x10^7$
Cassette incorporation efficiency	99.9%	99.9%	99.9%
Library coverage	100% (7776/7776)	96.7% (20043/20736)	99.9% (17701/17725)
% desired reads (designed variants - right mutation and number of expected mutations)	86.8% (824718/949555)	83.9% (594753/708766)	80.9% (2607572/3224158)
% chimera reads (variants with the correct mutations but with more than 3 mutations in cassette 3)	0% (0/949555)	0% (0/708766)	8.9% (289648/3224158)
% WT sequence reads	0.22% (1798/824718)	0.06% (347/594753)	0.02% (483/2607572)

Supplementary Data

Cassettes

PYR1 Cass1 TTCCCTACTGCTCGTACATCCGGTCTCACGGTAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCTTCG GAGTTAACACCAGAAGAACGATCGGAACTAAAAAACTCAATCGCCGAGTTCCACACATACCAACTCG ATCCAGGAAGCTGTTCATCACTCCACGCGCAACGAATCCACGCGCCTCCGGAACTCGTCTGTGAGAC CATAGCGGTTCTCACCCCTCAACACCTGCGCTGTCCGCACCGTTTGAAGTAAATTAGTTGAGGATTTA GCAGTGCTATCATGTGATCTCCAAATTAAAACATACCGTTCCATGAGGGCTAGAATTACTTACCGGCCT TCACCATGCCTGTACTATACGAACCCACTCTC
PYR1 Cass2 TTCCCTACTGCTCGTACATCCGGTCTCATCTGGTCAATCGTACGACGATTCGACAAACCACAAACATA CAAACCGTTCATCAAATCCTGCTCCGTCGAACAAAACTTCGAGATGCGCGTCGGATGCACGCGCGAC GTGATCGTCATCAGTGGATTACCGGCGTCTACATCAACGGAAAGACTCGATATACTCGACGACGAACG GAGAGTTACTGAGACCATAGCGGTTCTCACCCCTCAACACCTGCGCTGTCCGCACCGTTTGAAGTAA ATTAGTTGAGGATTTAGCAGTGCTATCATGTGATCTCCAAATTAAAACATACCGTTCCATGAGGGCTAG AATTACTTACCGGCCTTCACCATGCCTGTACTATACGAACCCACTCTC
PYR1 Cass3 TTCCCTACTGCTCGTACATCCGGTCTCATTACCGGATTCAGTATCATCGGAGGCGAACATAGGCTGACG AATTACAAATCCGTTACGACGGTGCATCGGTTCGAGAAAGAGAATCGGATCTGGACGGTGGTTTTGG AATCTTACGTCGTTGATATGCCTGAGACCATAGCGGTTCTCACCCCTCAACACCTGCGCTGTCCGCAC CGTTTGAAGTAAATTAGTTGAGGATTTAGCAGTGCTATCATGTGATCTCCAAATTAAAACATACCGTTC CATGAGGGCTAGAATTACTTACCGGCCTTCACCATGCCTGTACTATACGAACCCACTCTC
PYR1 Cass4 TTCCCTACTGCTCGTACATCCGGTCTCATGCCGGAAGGTAACTCGGAGGATGATACTCGTATGTTTGCT GATACGGTTGTGAAGCTTAATTTGCAGAAACTCGCGACGGTTGCTGAAGCTATGGCTCGTAACTCCGG TGACGGAAGTGGTTCTCAGGTGACGCTCGAGGGGGGCGGATCCGAATGAGACCATAGCGGTTCTCAC CCCTCAACACCTGCGCTGTCCGCACCGTTTGAAGTAAATTAGTTGAGGATTTAGCAGTGCTATCATGT GATCTCCAAATTAAAACATACCGTTCCATGAGGGCTAGAATTACTTACCGGCCTTCACCATGCCTGTAC TATACGAACCCACTCTC
PYR1 Cass1+2 TTCCCTACTGCTCGTACATCCGGTCTCACGGTAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCTTCG GAGTTAACACCAGAAGAACGATCGGAACTAAAAAACTCAATCGCCGAGTTCCACACATACCAACTCG ATCCAGGAAGCTGTTCATCACTCCACGCGCAACGAATCCACGCGCCTCCGGAACTCGTCTGGTCAAT CGTACGACGATTCGACAAACCACAAACATACAAACCGTTCATCAAATCCTGCTCCGTCGAACAAAAC TTCGAGATGCGCGTCGGATGCACGCGCGACGTGATCGTCATCAGTGGATTACCGGCGTCTACATCAAC GGAAAGACTCGATATACTCGACGACGAACGGAGAGTTACTGAGACCATAGCGGTTCTCACCCCTCAA
PYR1 Cass1+2+3 TTCCCTACTGCTCGTACATCCGGTCTCACGGTAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCTTCG GAGTTAACACCAGAAGAACGATCGGAACTAAAAAACTCAATCGCCGAGTTCCACACATACCAACTCG ATCCAGGAAGCTGTTCATCACTCCACGCGCAACGAATCCACGCGCCTCCGGAACTCGTCTGGTCAAT CGTACGACGATTCGACAAACCACAAACATACAAACCGTTCATCAAATCCTGCTCCGTCGAACAAAAC TTCGAGATGCGCGTCGGATGCACGCGCGACGTGATCGTCATCAGTGGATTACCGGCGTCTACATCAAC GGAAAGACTCGATATACTCGACGACGAACGGAGAGTTACCGGATTCAGTATCATCGGAGGCGAACAT AGGCTGACGAATTACAAATCCGTTACGACGGTGCATCGGTTCGAGAAAGAGAATCGGATCTGGACGG TGGTTTTGGAATCTTACGTCGTTGATATGCCTGAGACCATAGCGGTTCTCACCCCTCAA

PYR1 Cass1+2+3+4 TTCCCTACTGCTCGTACATCCGGTCTCACGGTAGCGGAGGCGGAGGGTCGGCTAGCCATATGCCTTCG GAGTTAACACCAGAAGAACGATCGGAACTAAAAAACTCAATCGCCGAGTTCCACACATACCAACTCG ATCCAGGAAGCTGTTCATCACTCCACGCGCAACGAATCCACGCGCCTCCGGAACTCGTCTGGTCAAT CGTACGACGATTCGACAAACCACAAACATACAAACCGTTCATCAAATCCTGCTCCGTCGAACAAAAC TTCGAGATGCGCGTCGGATGCACGCGCGACGTGATCGTCATCAGTGGATTACCGGCGTCTACATCAAC GGAAAGACTCGATATACTCGACGACGAACGGAGAGTTACCGGATTCAGTATCATCGGAGGCGAACAT AGGCTGACGAATTACAAATCCGTTACGACGGTGCATCGGTTCGAGAAAGAGAATCGGATCTGGACGG TGGTTTTGGAATCTTACGTCGTTGATATGCCGGAAGGTAACTCGGAGGATGATACTCGTATGTTTGCTG ATACGGTTGTGAAGCTTAATTTGCAGAAACTCGCGACGGTTGCTGAAGCTATGGCTCGTAACTCCGGT GACGGAAGTGGTTCTCAGGTGACGCTCGAGGGGGGCGGATCCGAATGAGACCATAGCGGTTCTCACC CCTCAA
RBD eblock Cass1 TTCCCTACTGCTCGTACATCCGGTCTCATTTGTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAG GGCAAACTGGAAAGATTGCTGATTATAATTATAAATTACCAGATGATTTTACAGGCTGCGTTATAGCTT GGAGAGACCATAGCGGTTCTCACCCCTCAA
RBD eblock Cass2 TTCCCTACTGCTCGTACATCCGGTCTCATTGGAATTCTAACAATCTTGATTCTAAGGTTGGTGGTAATTA TAATTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAGATATTTCAACTGAAAT CAGAGACCATAGCGGTTCTCACCCCTCAA
RBD eblock Cass3 TTCCCTACTGCTCGTACATCCGGTCTCAAATCTATCAGGCCGGTAACACACCTTGTAATGGTGTTGCAG GTTTTAATTGTTACTTTCCTTTACAATCATATGGTTTCCGACCCACTTATGGTGTTGGTCACCAACCATA CAGAGAGAGACCATAGCGGTTCTCACCCCTCAA
RBD Ultramer Cass1 GCATCGGGTCTCGTTTGTCATAAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAAACTGGAAAG ATTGCTGACTACAATTACAAGTTACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAGAGACCATAGA TATGATGTA
RBD Ultramer Cass2 GCATCGGGTCTCATTGGAATTCTAATAATCTTGATTCTAAGGTAGGTGGAAATTACAATTACCTGTATAG ATTGTTTAGGAAGTCTAATCTCAAACCTTTCGAGAGAGATATTTCAACTGAAATCAGAGACCATGCTA GAGCGTATGATGTA
RBD opool Cass1 GCATCGTTCCCTACTGCTCGTACATGCGGTCTCGTTTGTCATAAGAGGTGATGAAGTCAGACAAATCG CTCCAGGGCAAACTGGAAAGATTGCTGACTACAATTACAAGTTACCAGATGATTTTACAGGCTGCGTT ATAGCTTGGAGAGACCATAGATATGATGTA
RBD opool Cass2 GCATCGTTCCCTACTGCTCGTACATGCGGTCTCATTGGAATTCTAATAATCTTGATTCTAAGGTAGGTG GAAATTACAATTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTCGAGAGAGATATTTCAA CTGAAATCAGAGACCATGCTAGAGCGTATGATGTA
RBD opool Cass3 GCATCGTTCCCTACTGCTCGTACATGCGGTCTCAAATCTATCAGGCTGGTAACACACCTTGTAATGGTG TTGCAGGTTTTAATTGTTACTTTCCTTTACAATCATATGGCTTCCGACCCACTTATGGTGTAGGTCACCA ACCATACAGAGAGAGACCATGCTAGAGCGTATGATGTA

RBD GG Cassette 1 (Homology Arms in CAPITALS):

CTCTGCTTTACTAATGTCTATGCAGATTCAtttgtaattagaggtgatgaagtcagacaaatcgctccagggcaaactggaaagattgctgat
tataattataaattaccagatgattttacAGGCTGCGTTATAGCTTGGAATTCTAACAA

GG Cassette 2 (Homology Arms in CAPITALS):

AGGCTGCGTTATAGCTTGGAATTCTAACAAtcttgattctaaggttggtggtaattataattacctgtatagattgtttaggaagtctaatctcaa
acctttgagagAGATATTTCAACTGAAATCTATCAGGCtGG

GG Cassette 3 (Homology Arms in CAPITALS):

AGATATTTCAACTGAAATCTATCAGGCtGGtaacacaccttgtaatggtgttgcaggttttaattgttactttcctttacaatcatatggcttcsda
yhtwbghmtggtgttggtbwtcaaccatacagaGTAGTAGTACTTTCTTTTGAACTTCTACAT

NGS Amplicon (440 bp Amplicon insert (490 with Illumina Adapters), Homology Arms in CAPITALS, Illumina Adapters *italicized*, RBD sequence in **bold**):

*gttcagacgtgtgctcttccgatct***ggagtgtctcctactaaattaaatgatCTCTGCTTTACTAATGTCTATGCAGATTCAtttgtaa
ttagaggtgatgaagtcagacaaatcgctccagggcaaactggaaagattgctgattataattataaattaccagatgattttacAGGCTGCGTTA
TAGCTTGGAATTCTAACAAtcttgattctaaggttggtggtaattataattacctgtatagattgtttaggaagtctaatctcaaacctttgag
agAGATATTTCAACTGAAATCTATCAGGCtGGtaacacaccttgtaatggtgttgcaggttttaattgttactttcctttacaatcata
tggcttcsdayhtwbghmtggtgttggtbwtcaaccatacagaGTAGTAGTACTTTCTTTTGAACTTCTACATgcaccagca
actgtttgtggacc***agatcggaagagcgtcgtgtaggga***

T7 RNAP Cass1

GCATCGGGTCTCATTCGCAAGCGTTGCGCTGBNCATTGGGTAACTCCTGATGGTTTCCCTGBNTGGND
TGAATACAAGAAGCCTATTCAGACGCGCTTGAACCTGATGTTCCTCGGTCAGTTCCGCTTACAGCCTA
CCATTAACACCAACAAAGATAGCGAGATTGAGACCCGTAGC

T7 RNAP Cass2

GCATCGGGTCTCAAGATTGATGCACACAAACAGGAGTCTGGTVTAGCTCCTMRNNDTGTACACRSYV
RKGACGGTAGCCACCTTCGTAAGACTGTAGTGTGGGCACACGAGAAGTACGGAATCGAATCTTTTGC
ACTGATTCACGACTCCTTCGGTACCATTCCGGCTGACGCTGCGAACCTGTTCAAAGCAGTGCGCGAA
ACTATGGTTGACACATATGAGTCTTGTGATGTACTGGCTGATNDTTACGACCAGATCGCTTGAGACCC
GTAGC

Destination Vectors

pND003

TAATTATTTTTATAGCACGTGATGAAAAGGACCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACC
CCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCT
TCAATAATATTGAAAAAGGAAGAGTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGAT
TAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTG
CGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGC
GTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGAC
CATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGC
ATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCG

CCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGC
GCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTG
TTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGT
GATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTC
GGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTA
CAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGC
TCGATGAGTTTTTCTAACCTCAGCAACTGTCAACCAAGTTTACTCATATATACTTTAGATTGATTTAAAA
CTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAAC
GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTT
TTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGG
ATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTT
CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCT
GCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGAC
GATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG
AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGA
AGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGC
TTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGA
TTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGT
TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTG
AGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAAT
GCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTT
AGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGA
GCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCAC
TAAAGGGAACAAAAGCTGGTACCAATTCCTTGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGAC
GCAAAGAAGTTTAATAATCATATTACATGGCATTACCACCATATACATATCCATATCTAATCTTACTTATAT
GTTGTGGAAATGTAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACTTTCAGTAATA
CGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAAGCCGCCGAGCGGGTGACAGCCCTCCGAA
GGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCG
CCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGC
AGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATT
AGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATAT
AAATGCAAAAACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCA
AATGTAATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAAC
CCCGGATCGAATTCCCTACTTCATACATTTTCAATTAAGATGCAGTTACTTCGCTGTTTTTCAATATTTT
CTGTTATTGCTTCAGTTTTAGCACAGGAACTGACAACTATATGCGAGCAAATCCCCTCACCAACTTTAG
AATCGACGCCGTACTCTTTGTCAACGACTACTATTTTGGCCAACGGGAAGGCAATGCAAGGAGTTTTT
GAATATTACAAATCAGTAACGTTTGTCAGTAATTGCGGTTCTCACCCCTCAACAACTAGCAAAGGCAG
CCCCATAAACACACAGTATGTTTTTAAGGACAATAGCTCGACGATTGAAGGTAGATACCCATACGACG
TTCCAGACTACGCTCTGCAGGCTAGTGGTGGAGGAGGCTCTGGTGGAGGCGGTTGAGACCGAAAGT
GAAACGTGATTTCATGCGTCATTTTGAACATTTTGTAAATCTTATTTAATAATGTGTGCGGCAATTCACA
TTTAATTTATGAATGTTTTCTTAACATCGCGGCAACTCAAGAAACGGCAGGTTCGGATCTTAGCTACTA
GAGAAAGAGGAGAAATACTAGATGCGTAAAGGCGAAGAGCTGTTCACTGGTGTCGTCCCTATTCTGG
TGGAACTGGATGGTGATGTCAACGGTCATAAGTTTTCCGTGCGTGGCGAGGGTGAAGGTGACGCAAC
TAATGGTAAACTGACGCTGAAGTTCATCTGTACTACTGGTAAACTGCCGGTTCCTTGGCCGACTCTGG
TAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTCGTTATCCGGACCATATGAAGCAGCATGACTTCT
TCAAGTCCGCCATGCCGGAAGGCTATGTGCAGGAACGCACGATTTCCTTTAAGGATGACGGCACGTA
CAAAACGCGTGCGGAAGTGAAATTTGAAGGCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCATT
GACTTTAAAGAGGACGGCAATATCCTGGGCCATAAGCTGGAATACAATTTTAACAGCCACAATGTTTA
CATCACCGCCGATAAACAAAAAAATGGCATTAAAGCGAATTTTAAAATTCGCCACAACGTGGAGGAT
GGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACACTCCAATCGGTGATGGTCCTGTTCTGCTGC
CAGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAAAGATCCGAACGAGAAACGCGATCATAT
GGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGGATGAACTGTACAAATGACCA
GGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTG
AACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAGGTCTCACGA

ACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGAGATCTGATAACAACAGTGTAGATGTAACAAAAT
CGACTTTGTTCCCACTGTACTTTTAGCTCGTACAAAATACAATATACTTTTCATTTCTCCGTAAACAAC
ATGTTTTCCCATGTAATATCCTTTTCTATTTTTCGTTCCGTTACCAACTTTACACATACTTTATATAGCTAT
TCACTTCTATACACTAAAAAACTAAGACAATTTTAATTTTGCTGCCTGCCATATTTCAATTTGTTATAAA
TTCCTATAATTTATCCTATTAGTAGCTAAAAAAAGATGAATGTGAATCGAATCCTAAGAGAATTGAGCT
CCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAA
AACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGA
AGAGGCCCGCACCGATCGCCTTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTA
GCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCT
AGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCT
AAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATT
AGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCC
ACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTT
GATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAAC
GCGAATTTTAACAAAATATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGG
TATTTCACACCGCATAGATCGGCAAGTGCACAAACAATACTTAAATAAATACTACTCAGTAATAACCTA
TTTCTTAGCATTTTTGACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACCTCCG
CTTACATCAACACCAATAACGCCATTTAATCTAAGCGCATCACCAACATTTTCTGGCGTCAGTCCACCA
GCTAACATAAAATGTAAGCTTTCGGGGCTCTCTTGCCTTCCAACCCAGTCAGAAATCGAGTTCCAATC
CAAAAGTTCACCTGTCCCACCTGCTTCTGAATCAAACAAGGGAATAAACGAATGAGGTTTCTGTGAA
GCTGCACTGAGTAGTATGTTGCAGTCTTTTGGAAATACGAGTCTTTTAATAACTGGCAAACCGAGGAA
CTCTTGGTATTCTTGCCACGACTCATCTCCATGCAGTTGGACGATATCAATGCCGTAATCATTGACCAG
AGCCAAAACATCCTCCTTAGGTTGATTACGAAACACGCCAACCAAGTATTTCGGAGTGCCTGAACTAT
TTTTATATGCTTTTACAAGACTTGAAATTTTCCTTGCAATAACCGGGTCAATTGTTCTCTTTCTATTGGG
CACACATATAATACCCAGCAAGTCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCA
AGCCGCAAACTTTCACCAATGGACCAGAACTACCTGTGAAATTAATAACAGACATACTCCAAGCTGC
CTTTGTGTGCTTAATCACGTATACTCACGTGCTCAATAGTCACCAATGCCCTCCCTCTTGGCCCTCTCC
TTTTCTTTTTTCGACCGAATTAATTCTTAATCGGCAAAAAAAGAAAAGCTCCGGATCAAGATTGTACG
TAAGGTGACAAGCTATTTTTCAATAAAGAATATCTTCCACTACTGCCATCTGGCGTCATAACTGCAAAG
TACACATATATTACGATGCTGTTCTATTAAATGCTTCCTATATTATATATATAGTAATGTCGTGATCTATGG
TGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGC
TGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGG
AGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATAC
GCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGGATCGCTTGCCTGTAACTTACAC
GCGCCTCGTATCTTTTAATGATGGAATAATTTGGGAATTTACTCTGTGTTTATTTATTTTTATGTTTTGTAT
TTGGATTTTAGAAAGTAAATAAAGAAGGTAGAAGAGTTACGGAATGAAGAAAAAAAAAATAAACAAA
GGTTTAAAAAATTTCAACAAAAAGCGTACTTTACATATATATTTATTAGACAAGAAAAGCAGATTAAAT
AGATATACATTCGATTAACGATAAGTAAAATGTAAAATCACAGGATTTTCGTGTGTGGTCTTCTACACA
GACAAGATGAAACAATTCGGCATTAATACCTGAGAGCAGGAAGAGCAAGATAAAAGGTAGTATTTGT
TGGCGATCCCCCTAGAGTCTTTTACATCTTCGGAAAACAAAAACTATTTTTTCTTTAATTTCTTTTTTTA
CTTTCTATTTTTAATTTATATATTTATATTAAAAAATTTAAATTA

pND004

TGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACA
GATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGT
ACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGA
ATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCC
TGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAA
CGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAA
GAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAAC
CTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATA
CCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCA
AAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAA
GAATTAATTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCAC
ATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTT

AAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACC
GAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACG
TCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTT
TTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGA
CGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGG
CGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTACTACAGGGCGCGT
CCCATTCGCTGAGGGCCAATCCGGATATAGTTCCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCC
CCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTGAGACCTATAAA
CGCAGAAAGGCCCACCCGAAGGTGAGCCAGTGTGACTCTAGTAGAGAGCGTTCACCGACAAACAAC
AGATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTGATGCCTGGTCATTTGTACA
GTTCATCCATACCATGCGTGATGCCCGCTGCGGTTACGAACTCCAGCAGAACCATATGATCGCGTTTCT
CGTTCGGATCTTTAGACAGAACGCTTTGCGTGCTCAGATAGTGATTGTCTGGCAGCAGAACAGGACC
ATCACCGATTGGAGTGTTTTGCTGGTAGTGATCAGCCAGCTGCACGCTGCCATCCTCCACGTTGTGGC
GAATTTTAAAATTCGCTTTAATGCCATTTTTTTGTTTATCGGCGGTGATGTAAACATTGTGGCTGTTAAA
ATTGTATTCCAGCTTATGGCCCAGGATATTGCCGTCCTCTTTAAAGTCAATGCCTTTCAGCTCAATGCG
GTTTACCAGGGTATCGCCTTCAAATTTCACTTCCGCACGCGTTTTGTACGTGCCGTCATCCTTAAAGGA
AATCGTGCGTTCCTGCACATAGCCTTCCGGCATGGCGGACTTGAAGAAGTCATGCTGCTTCATATGGT
CCGGATAACGAGCAAAGCACTGAACACCATAAGTCAGCGTCGTTACCAGAGTCGGCCAAGGAACCG
GCAGTTTACCAGTAGTACAGATGAACTTCAGCGTCAGTTTACCATTAGTTGCGTCACCTTCACCCTCG
CCACGCACGGAAAACTTATGACCGTTGACATCACCATCCAGTTCCACCAGAATAGGGACGACACCAG
TGAACAGCTCTTCGCCTTTACGCATCTAGTATTTCTCCTCTTTCTCTAGTAGCTAAGATCCGAACCTGC
CGTTTCTTGAGTTGCCGCGATGTTAAGAAAACATTCATAAATTAAATGTGAATTGCCGCACACATTATT
AAATAAGATTTACAAAATGTTCAAAATGACGCATGAAATCACGTTTCACTTTCGGTCTCAACCGCTGG
CGGCGTTGATCACCGCAGTACGCACGGCATACCAGAAAGCGGACATCTGCGGGATGTTCGGCATGAT
TTCACCTTTCTGGGCGTTTTCCATAGTGGCGGCAATACGCGGATCTTTCACCAACTCTTCCTCGTAAGA
CTTCAGCGCTACGGCACCCAGCGGTTTGTCTTTATTAACCGCTTCCAGACCTTCATCAGTCAGCAGAT
AGTTTTCGAGGAACTCTTTTGCCAGCTCTTTGTTCGGACTGGCGGCGTTAATACCTGCGCTCAGCACG
CCAACGAACGGTTTGGATGGTTGACCCTTGAAGGTCGGCAGTACCGTTACACCATAATTCACTTTGCT
GGTGTCGATGTTGGACCATGCCCACGGGCCGTTGATGGTCATCGCTGTTTCGCCTTTATTAAAGGCAG
CTTCTGCGATGGAGTAATCGGTGTCTGCATTCATGTGTTTGTTTTTAATCAGGTCAACCAGGAAGGTC
AGACCCGCTTTCGCGCCAGCGTTATCCACGCCCACGTCTTTAATGTCGTACTTGCCGTTTTCATACTTG
AACGCATAACCCCCGTCAGCAGCAATCAGCGGCCAGGTGAAGTACGGTTCTTGCAGGTTGAACATCA
GCGCGCTCTTACCTTTCGCTTTCAGTTCTTTATCCAGCGCCGGGATCTCTTCCCAGGTTTTTGGCGGGT
TCGGCAGCAGATCTTTGTTATAAATCAGCGATAACGCTTCAACAGCGATCGGGTAAGCAATCAGCTTG
CCGTTGTAACGTACGGCATCCCAGGTAAACGGATACAGCTTGTCCTGGAACGCTTTGTCCGGGGTGAT
TTCAGCCAACAGGCCAGATTGAGCGTAGCCACCAAAGCGGTCGTGTGCCCAGAAGATAATGTCAGGG
CCATCGCCAGTTGCCGCAACCTGTGGGAATTTCTCTTCCAGTTTATCCGGATGCTCAACGGTGACTTT
AATTCCGGTATCTTTCTCGAATTTCTTACCGACTTCAGCGAGTCCGTTATAGCCTTTATCGCCGTTAATC
CAGATTACCAGTTTACCTTCTTCGATTTTCATGCCGCTGCTGTGATGATGATGATGATGGCTGCTATTCA
TGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTATCCGCTCACAATTCCC
CTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCGATCTCGATCCTCTACGCCGGACGCATCGTGG
CCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGA
TCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCC
GGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCA
ACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGAGATCCCGGACACCA
TCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGT
GAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCC
GCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCG
GAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCG
TTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGAT
CAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCG
GTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGC
CATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCAT
CAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTC
ACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGG

CATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTC
CGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAAC
GATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGACATCT
CGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAG
GATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGA
AGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAA
CCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAG
CGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACCGGGATCTCGACCGATG
CCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCAC
TTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCG
AGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCC
CTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCG
GCATGGCGGCCCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGG
TTGCCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAA
ACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGC
GGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGA
ACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATC
CATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATC
GTGAGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGGCA
TCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGC
TTCTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACC
ACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACAT
GCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAAACAAGCCCGTCAGGG
CGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGT
GTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATTGCGGTGTGAAATA
CCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT
GCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACA
GAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAA
AAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGC
TCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCC
TCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC
GTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGG
CTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTA
TGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTT
GGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAC
AAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATC
TCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGG
ATTTTGGTCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATAT
TCAACGGGAAACGTCTTGCTCTAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAA
ATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTA

pND005

AGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATA
TATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATG
TTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAA
CTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGG
CGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGT
ACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGT
TGCTACGCCTGAATAAGTGATAATAAGCGGATGAATGGCAGAAATTCGAAAGCAAATTCGACCCGGTC
GTCGGTTCAGGGCAGGGTCGTTAAATAGCCGCTTATGTCTATTGCTGGTTTACCGGTTTATTGACTACC
GGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGG
CTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGC
GCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTG
TAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGC

TGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAACGCATTTAAGCATAAACA
CGCACTATGCCGTTCTTCTCATGTATATATATATACAGGCAACACGCAGATATAGGTGCGACGTGAACA
GTGAGCTGTATGTGCGCAGCTCGCGTTGCATTTTCGGAAGCGCTCGTTTTCGGAAACGCTTTGAAGTT
CCTATTCCGAAGTTCCTATTCTCTAGCTAGAAAGTATAGGAACTTCAGAGCGCTTTTGAAAACCAAAA
GCGCTCTGAAGACGCACTTTCAAAAAACCAAAAACGCACCGGACTGTAACGAGCTACTAAAATATTG
CGAATACCGCTTCCACAAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATATCCCTATAT
AACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGACCTCTACATTTTTTATGTTTATC
TCTAGTATTACTCTTTAGACAAAAAAATTGTAGTAAGAACTATTCATAGAGTGAATCGAAAACAATACG
AAAATGTAAACATTTCCTATACGTAGTATATAGAGACAAAATAGAAGAAACCGTTCATAATTTTCTGAC
CAATGAAGAATCATCAACGCTATCACTTTCTGTTCACAAAGTATGCGCAATCCACATCGGTATAGAATA
TAATCGGGGATGCCTTTATCTTGAAAAAATGCACCCGCAGCTTCGCTAGTAATCAGTAAACGCGGGAA
GTGGAGTCAGGCTTTTTTTATGGAAGAGAAAATAGACACCAAAGTAGCCTTCTTCTAACCTTAACGGA
CCTACAGTGCAAAAAGTTATCAAGAGACTGCATTATAGAGCGCACAAAGGAGAAAAAAAGTAATCTA
AGATGCTTTGTTAGAAAAATAGCGCTCTCGGGATGCATTTTTGTAGAACAAAAAAGAAGTATAGATTC
TTTGTTGGTAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTT
GAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTTTACAAAAATGAAGCACAGATTCTTCGTTGGTAAA
ATAGCGCTTTCGCGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAATTAGCG
CTCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTTGCTTGCATGCAACTTCTTTT
CTTTTTTTTTCTTTTCTCTCTCCCCCGTTGTTGTCTCACCATATCCGCAATGACAAAAAAATGATGGAA
GACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTTGTTCCAGAGCTGATGA
GGGGTATCTCGAAGCACACGAAACTTTTTCCTTCCTTCATTCACGCACACTACTCTCTAATGAGCAAC
GGTATACGGCCTTCCTTCCAGTTACTTGAATTTGAAATAAAAAAAAGTTTGCTGTCTTGCTATCAAGTA
TAAATAGACCTGCAATTATTAATCTTTTGTTTCCTCGTCATTGTTCTCGTTCCCTTTCTTCCTTGTTTCTT
TTTCTGCACAATATTTCAAGCTATACCAAGCATACAATCAACTCCAAGCTTGAAGCAAGCCTCCTTGA
GACCGAAAGTGAAACGTGATTTCATGCGTCATTTTGAACATTTTGTAAATCTTATTTAATAATGTGTGC
GGCAATTCACATTTAATTTATGAATGTTTTCTTAACATCGCGGCAACTCAAGAAACGGCAGGTTCGGAT
CTTAGCTACTAGAGAAAGAGGAGAAATACTAGATGCGTAAAGGCGAAGAGCTGTTCACTGGTGTCGT
CCCTATTCTGGTGGAACTGGATGGTGATGTCAACGGTCATAAGTTTTCCGTGCGTGGCGAGGGTGAAG
GTGACGCAACTAATGGTAAACTGACGCTGAAGTTCATCTGTACTACTGGTAAACTGCCGGTTCCTTGG
CCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTCGTTATCCGGACCATATGAAGCA
GCATGACTTCTTCAAGTCCGCCATGCCGGAAGGCTATGTGCAGGAACGCACGATTTCCTTTAAGGATG
ACGGCACGTACAAAACGCGTGCGGAAGTGAAATTTGAAGGCGATACCCTGGTAAACCGCATTGAGCT
GAAAGGCATTGACTTTAAAGAGGACGGCAATATCCTGGGCCATAAGCTGGAATACAATTTTAACAGCC
ACAATGTTTACATCACCGCCGATAAACAAAAAAATGGCATTAAAGCGAATTTTAAAATTCGCCACAAC
GTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACACTCCAATCGGTGATGGTCCTG
TTCTGCTGCCAGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAAAGATCCGAACGAGAAACG
CGATCATATGGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGGATGAACTGTACA
AATGACCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTT
TGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAGG
TCTCAATTCCGGGCGAATTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGT
ATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTTATTCTTGAGTAACTCTTTCCTGTA
GGTCAGGTTGCTTTCTCAGGTATAGCATGAGGTCGCTCTTATTGACCACACCTCTACCGGCATGCCGG
CAAGTGCACAAACAATACTTAAATAAATACTACTCAGTAATAACCTATTTCTTAGCATTTTTGACGAAA
TTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACCTCCGCTTACATCAACACCAATAACGC
CATTTAATCTAAGCGCATCACCAACATTTTCTGGCGTCAGTCCACCAGCTAACATAAAATGTAAGCTTT
CGGGGCTCTCTTGCCTTCCAACCCAGTCAGAAATCGAGTTCCAATCCAAAAGTTCACCTGTCCCACCT
GCTTCTGAATCAAACAAGGGAATAAACGAATGAGGTTTCTGTGAAGCTGCACTGAGTAGTATGTTGC
AGTCTTTTGGAAATACGAGTCTTTTAATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGAC
TCATCTCCATGCAGTTGGACGATATCAATGCCGTAATCATTGACCAGAGCCAAAACATCCTCCTTAGGT
TGATTACGAAACACGCCAACCAAGTATTTCGGAGTGCCTGAACTATTTTTATATGCTTTTACAAGACTT
GAAATTTTCCTTGCAATAACCGGGTCAATTGTTCTCTTTCTATTGGGCACACATATAATACCCAGCAAG
TCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGCAAACTTTCACCAATGG
ACCAGAACTACCTGTGAAATTAATAACAGACATACTCCAAGCTGCCTTTGTGTGCTTAATCACGTATAC
TCACGTGCTCAATAGTCACCAATGCCCTCCCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAATT
CGTAATCATGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGC

CGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCT
CACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGGTTCTTTCCGCCTCAGGACTCTTCCTT
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCGCCCTGTAGCGGC
GCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGC
CCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATC
GGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGT
GATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTT
CTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTT
ATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGA
ATTTTAACAAAATATTAACGCTTACAATTTACGCGCCTAAATTATGCGGCCGCATATAACTGCATTAATG
AATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGAC
TCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATC
CACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACC
GTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCG
ACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGC
TCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCT
GGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGT
CCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAG
GTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTAT
TTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAA
CAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGAT
CTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGG
GATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTAT
CTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACG
GGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGATCCACGCTCACCGGCTCCAGATT
TATCAGCAATAAACCAGCCAGCCGCCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGT
CCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGA
GGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTGTCGAGATTTTCA
GGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCA
TCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGG
ATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTC
TTGCCCGCCTGATGAATGCTCATCCGGAATTACGTATGGCAATGAAAGACGGTGAGCTGGTGATATGG
GATAGTGTTCACCCTTGTTACACCGTTTTCCATG

pIFU037

ATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAA
CCCTCACTAAAGGGAACAAAAGCTGGTACCAATTCCTTGAATTTTCAAAAATTCTTACTTTTTTTTTGG
ATGGACGCAAAGAAGTTTAATAATCATATTACATGGCATTACCACCATATACATATCCATATCTAATCTTA
CTTATATGTTGTGGAAATGTAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACTTTC
AGTAATACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAAGCCGCCGAGCGGGTGACAGCC
CTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTG
CCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAAA
AATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAA
TGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACA
GATATATAAATGCAAAAACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTC
TTATTCAAATGTAATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGA
AAAAACCCCGGATCGAATTCCCTACTTCATACATTTTCAATTAAGATGCAGTTACTTCGCTGTTTTTCA
ATATTTTCTGTTATTGCTTCAGTTTTAGCACAGGAACTGACAACTATATGCGAGCAAATCCCCTCACCA
ACTTTAGAATCGACGCCGTACTCTTTGTCAACGACTACTATTTTGGCCAACGGGAAGGCAATGCAAGG
AGTTTTTGAATATTACAAATCAGTAACGTTTGTCAGTAATTGCGGTTCTCACCCCTCAACAACTAGCAA
AGGCAGCCCCATAAACACACAGTATGTTTTTAAGGACAATAGCTCGACGATTGAAGGTAGATACCCAT
ACGACGTTCCAGACTACGCTCTGCAGGCTAGTGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAG

GCGGAGGGTCGGCTAGCCATACAAACTTGTGCCCTTTTGGTGAAGTTTTTAACGCCACCAGATTTGCA
TCTGTTTATGCTTGGAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAATTCC
GCATCATTTTCCACTTTTAAGTGTTATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTTACTAATG
TCTATGCAGATTCATTTGAGAGACCACGGCCACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAG
GTCTATGAGTGGTTGCTGGATAACTTTACGGGCATGCATAAGGCTCGTAATATATATTCAGGGAGGCCA
CAACGGTTTCCCTCTACAAATAATTTTGTTTAACTTTCTAGAAATAATTTTGTTTAACTTTAAGAAGGA
GATATACATATGCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGT
GATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTA
CCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTT
ATGGTGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATG
CCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTG
AAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGAT
GGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAA
ACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTA
GCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCT
GTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGGGACCACATGGTCCTTCTTGAGTTT
GTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAACTAGAGTGAGATCCGGCTGCTAA
CAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGG
GCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATGGTCTCAAGAGTA
GTAGTACTTTCTTTTGAACTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTACTAATTTG
GTTAAAAACAAACTCGAGGGGGGCGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAGA
GATCTGATAACAACAGTGTAGATGTAACAAAATCGACTTTGTTCCCACTGTACTTTTAGCTCGTACAA
AATACAATATACTTTTCATTTCTCCGTAAACAACATGTTTTCCCATGTAATATCCTTTTCTATTTTTCGTT
CCGTTACCAACTTTACACATACTTTATATAGCTATTCACTTCTATACACTAAAAAACTAAGACAATTTTA
ATTTTGCTGCCTGCCATATTTCAATTTGTTATAAATTCCTATAATTTATCCTATTAGTAGCTAAAAAAAGA
TGAATGTGAATCGAATCCTAAGAGAATTGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCAC
TGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCA
CATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCTTTCCCAACAGTTGCG
CAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGC
GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTC
GCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGC
TTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGAT
AGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGA
ACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGT
TAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTCCT
GATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAGATCGGCAAGTGCACAAACA
ATACTTAAATAAATACTACTCAGTAATAACCTATTTCTTAGCATTTTTGACGAAATTTGCTATTTTGTTAG
AGTCTTTTACACCATTTGTCTCCACACCTCCGCTTACATCAACACCAATAACGCCATTTAATCTAAGCG
CATCACCAACATTTTCTGGCGTCAGTCCACCAGCTAACATAAAATGTAAGCTTTCGGGGCTCTCTTGC
CTTCCAACCCAGTCAGAAATCGAGTTCCAATCCAAAAGTTCACCTGTCCCACCTGCTTCTGAATCAAA
CAAGGGAATAAACGAATGAGGTTTCTGTGAAGCTGCACTGAGTAGTATGTTGCAGTCTTTTGGAAATA
CGAGTCTTTTAATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGACTCATCTCCATGCAGT
TGGACGATATCAATGCCGTAATCATTGACCAGAGCCAAAACATCCTCCTTAGGTTGATTACGAAACAC
GCCAACCAAGTATTTCGGAGTGCCTGAACTATTTTTATATGCTTTTACAAGACTTGAAATTTTCCTTGC
AATAACCGGGTCAATTGTTCTCTTTCTATTGGGCACACATATAATACCCAGCAAGTCAGCATCGGAATC
TAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGCAAACTTTCACCAATGGACCAGAACTACCT
GTGAAATTAATAACAGACATACTCCAAGCTGCCTTTGTGTGCTTAATCACGTATACTCACGTGCTCAAT
AGTCACCAATGCCCTCCCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAATTCTTAATCGGCAA
AAAAAGAAAAGCTCCGGATCAAGATTGTACGTAAGGTGACAAGCTATTTTTCAATAAAGAATATCTTC
CACTACTGCCATCTGGCGTCATAACTGCAAAGTACACATATATTACGATGCTGTTCTATTAAATGCTTCC
TATATTATATATATAGTAATGTCGTGATCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGT
TAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATC
CGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCG
AAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTT
TCTTAGACGGATCGCTTGCCTGTAACTTACACGCGCCTCGTATCTTTTAATGATGGAATAATTTGGGAA

TTTACTCTGTGTTTATTTATTTTTATGTTTTGTATTTGGATTTTAGAAAGTAAATAAAGAAGGTAGAAGA
GTTACGGAATGAAGAAAAAAAAATAAACAAAGGTTTAAAAAATTTCAACAAAAAGCGTACTTTACAT
ATATATTTATTAGACAAGAAAAGCAGATTAAATAGATATACATTCGATTAACGATAAGTAAAATGTAAAA
TCACAGGATTTTCGTGTGTGGTCTTCTACACAGACAAGATGAAACAATTCGGCATTAATACCTGAGAG
CAGGAAGAGCAAGATAAAAGGTAGTATTTGTTGGCGATCCCCCTAGAGTCTTTTACATCTTCGGAAAA
CAAAAACTATTTTTTCTTTAATTTCTTTTTTTACTTTCTATTTTTAATTTATATATTTATATTAAAAAATTTA
AATTATAATTATTTTTATAGCACGTGATGAAAAGGACCCAGGTGGCACTTTTCGGGGAAATGTGCGCG
GAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATA
AATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGC
CGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAAT
CAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAA
GGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCT
TCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAA
AACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTT
CCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCT
CAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTG
GCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTC
ATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGAC
GAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCT
TCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATT
TGATGCTCGATGAGTTTTTCTAACCTCAGCAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTG
ATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAAT
CCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGA
GATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTG
TTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCA
AATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATAC
CTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGA
CTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCC
CAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAC
GCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCA
CGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTT
GAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCT
TTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGT
GGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGC
GAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCG
ATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTA
ATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTG
GA

pZB558

ATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTC
CATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGA
CAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTT
CGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGG
GTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTAT
CCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGG
TAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTT
GGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAA
AACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCA
AGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCA
CCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTGGA
CGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTGGATA
ACCAGAAGCAATAAAAAATCAAATCGGATTTCACTATATAATCTCACTTTATCTAAGATGAATCCGATG
GAAGCATCCTGTTTTCTCTCAATTTTTTTATCTAAAACCCAGCGTTCGATGCTTCTTTGAGCGAACGAT
CAAAAATAAGTGCCTTCCCATCAAAAAAATATTGACAACATAAAAAACTTTGTGTTATACTTGTAACG

CTACATGGAGATTAACTCAATCTAGCTAGAGAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT
GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGGATTCACTGGAAC
TCTAGACCAAAGAGAGGACACAATGAACACGATTAACATCGCTAAGAACGACTTCTCTGACATCGAA
CTGGCTGCTATCCCGTTCAACACTCTGGCTGACCATTACGGTGAGCGTTTAGCTCGCGAACAGTTGGC
CCTTGAGCATGAGTCTTACGAGATGGGTGAAGCACGCTTCCGCAAGATGTTTGAGCGTCAACTTAAA
GCTGGTGAGGTTGCGGATAACGCTGCCGCCAAGCCTCTCATCACTACCCTACTCCCTAAGATGATTGC
ACGCATCAACGACTGGTTTGAGGAAGTGAAAGCTAAGCGCGGCAAGCGCCCGACAGCCTTCCAGTT
CCTGCAAGAAATCAAGCCGGAAGCCGTAGCGTACATCACCATTAAGACCACTCTGGCTTGCCTAACC
AGTGCTGACAATACAACCGTTCAGGCTGTAGCAAGCGCAATCGGTCGGGCCATTGAGGACGAGGCTC
GCTTCGGTCGTATCCGTGACCTTGAAGCTAAGCACTTCAAGAAAAACGTTGAGGAACAACTCAACAA
GCGCGTAGGGCACGTCTACAAGAAAGCATTTATGCAAGTTGTCGAGGCTGACATGCTCTCTAAGGGT
CTACTCGGTGGCGAGGCGTGGTCTTCGTGGCATAAGGAAGACTCTATTCATGTAGGAGTACGCTGCAT
CGAGATGCTCATTGAGTCAACCGGAATGGTTAGCTTACACCGCCAAAATGCTGGCGTAGTAGGTCAA
GACTCTGAGACTATCGAACTCGCACCTGAATACGCTGAGGCTATCGCAACCCGTGCAGGTGCGCTGG
CTGGCATCTCTCCGATGTTCCAACCTTGCGTAGTTCCTCCTAAGCCGTGGACTGGCATTACTGGTGGT
GGCTATTGGGCTAACGGTCGTCGTCCTCTGGCGCTGGTGCGTACTCACAGTAAGAAAGCACTGATGC
GCTACGAAGACGTTTACATGCCTGAGGTGTACAAAGCGATTAACATTGCGCAAAACACCGCATGGAA
AATCAACAAGAAAGTCCTAGCGGTCGCCAACGTAATCACCAAGTGGAAGCATTGTCCGGTCGAGGAC
ATCCCTGCGATTGAGCGTGAAGAACTCCCGATGAAACCGGAAGACATCGACATGAATCCTGAGGCTC
TCACCGCGTGGAGACGTGCTGCCGCTGCTGTGTACCGCAAGGACAAGGCTCGCAAGTCTCGCCGTAT
CAGCCTTGAGTTCATGCTTGAGCAAGCCAATAAGTTTGCTAACCATAAGGCCATCTGGTTCCCTTACA
ACATGGACTGGCGCGGTCGTGTTTACGCTGTGTCAATGTTCAACCCGCAAGGTAACGATATGACCAAA
GGACTGCTTACGCTGGCGAAAGGTAAACCAATCGGTAAGGAAGGTTACTACTGGCTGAAAATCCACG
GTGCAAACTGTGCGGGTGTCGATAAGGTTCCGTTCCCTGAGCGCATCAAGTTCATTGAGGAAAACCA
CGAGAACATCATGGCTTGCGCTAAGTCTCCACTGGAGAACACTTGGTGGGCTGAGCAAGATTCTCCG
TTCTGCTTCCTTGCGTTCTGCTTTGAGTACGCTGGGGTACAGCACCACGGCCTGAGCTATAACTGCTC
CCTTCCGCTGGCGTTTGACGGGTCTTGCTCTGGCATCCAGCACTTCTCCGCGATGCTCCGAGATGAGG
TAGGTGGTCGCGCGGTTAACTTGCTTCCTAGTGAAACCGTTCAGGACATCTACGGGATTGTTGCTAAG
AAAGTCAACGAGATTCTACAAGCAGACGCAATCAATGGGACCGATAACGAAGTAGTTACCGTGACCG
ATGAGAACACTGGTGAAATCTCTGAGAAAGTCAAGCTGGGCACTAAGGCACTGGCTGGTCAATGGCT
GGCTTACGGTGTTACTCGCAGTGTGACTAAGCGTTCAGTCATGACGCTGGCTTACGGGTCCAAAGAG
TTCGGCTTCCGTCAACAAGTGCTGGAAGATACCATTCAGCCAGCTATTGATTCCGGCAAGGGTCTGAT
GTTCACTCAGCCGAATCAGGCTGCTGGATACATGGCTAAGCTGATTTGGGAATCTGTGAGCGTGACGG
TGGTAGCTGCGGTTGAAGCAATGAACTGGCTTAAGTCTGCTGCTAAGCTGCTGGCTGCTGAGGTCAA
AGATAAGAAGACTGGAGAGATTCTTCGTGAGACCGAAAGTGAAACGTGATTTCATGCGTCATTTTGA
ACATTTTGTAAATCTTATTTAATAATGTGTGCGGCAATTCACATTTAATTTATGAATGTTTTCTTAACATC
GCGGCAACTCAAGAAACGGCAGGTTCGGATCTTAGCTACTAGAGAAAGAGGAGAAATACTAGATGCG
TAAAGGCGAAGAGCTGTTCACTGGTGTCGTCCCTATTCTGGTGGAACTGGATGGTGATGTCAACGGT
CATAAGTTTTCCGTGCGTGGCGAGGGTGAAGGTGACGCAACTAATGGTAAACTGACGCTGAAGTTCA
TCTGTACTACTGGTAAACTGCCGGTTCCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAG
TGCTTTGCTCGTTATCCGGACCATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCCGGAAGGCTAT
GTGCAGGAACGCACGATTTCCTTTAAGGATGACGGCACGTACAAAACGCGTGCGGAAGTGAAATTTG
AAGGCGATACCCTGGTAAACCGCATTGAGCTGAAAGGCATTGACTTTAAAGAGGACGGCAATATCCT
GGGCCATAAGCTGGAATACAATTTTAACAGCCACAATGTTTACATCACCGCCGATAAACAAAAAAATG
GCATTAAAGCGAATTTTAAAATTCGCCACAACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTA
CCAGCAAAACACTCCAATCGGTGATGGTCCTGTTCTGCTGCCAGACAATCACTATCTGAGCACGCAA
AGCGTTCTGTCTAAAGATCCGAACGAGAAACGCGATCATATGGTTCTGCTGGAGTTCGTAACCGCAGC
GGGCATCACGCATGGTATGGATGAACTGTACAAATGACCAGGCATCAAATAAAACGAAAGGCTCAGT
CGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGC
TCACCTTCGGGTGGGCCTTTCTGCGTTTATAGGTCTCACGCTGACCAGTTGCACGAGTCTCAATTGGA
CAAAATGCCAGCACTTCCGGCTAAAGGTAACTTGAACCTCCGTGACATCTTAGAGTCGGACTTCGCG
TTCGCGTAATAGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAAC
GAAGCGCTAACCGTTTTTATCAGGCTCTGGGAGGCAGAATAAATGATCATATCGTCAATTATTACCTCC
ACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTA
GTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGA

CCGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCA
GGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTG
TTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCA
GCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTG
TCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACA
TATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATA
TGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCA
TGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACG
GAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTAT
TTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAA
CTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGA
TTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTG
ATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGT
TGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTC
ACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGATGGTG
TTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGT
GCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCA
CTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCA
GAATATGTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCG
GCGAGCGGAAATGGCTTACGAACGGGGCGGAG

SUPPLEMENTARY REFERENCES CITED

(1) Klesmith, J. R., Bacik, J. P., Wrenbeck, E. E., Michalczyk, R., and Whitehead, T. A. (2017) Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proc Natl Acad Sci U S A 114*, 2265–2270.

(2) Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth Biol 4*, 975–986.

(3) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One 3*, e3647.

(4) Potapov, V., Ong, J. L., Kucera, R. B., Langhorst, B. W., Bilotti, K., Pryor, J. M., Cantor, E. J., Canton, B., Knight, T. F., Evans, T. C., and Lohman, G. J. S. (2018) Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synth Biol 7*, 2665–2674.

(5) Gerstmans, H., Grimon, D., Gutiérrez, D., Lood, C., Rodríguez, A., van Noort, V., Lammertyn, J., Lavigne, R., and Briers, Y. (2020) A VersaTile-driven platform for rapid hit-to-lead development of engineered lysins. *Sci Adv 6*.

(6) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods 6*, 343–345.

(7) Barnes, C. O., Jette, C. A., Abernathy, M. E., Dam, K.-M. A., Esswein, S. R., Gristick, H. B., Malyutin, A. G., Sharaf, N. G., Huey-Tubman, K. E., Lee, Y. E., Robbiani, D. F., Nussenzweig,

M. C., West, A. P., and Bjorkman, P. J. (2020) SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature 588*, 682–687.

(8) Green, M. R., and Sambrook, J. (2018) Touchdown Polymerase Chain Reaction (PCR). *Cold Spring Harb Protoc 2018*, pdb.prot095133.

(9) Kowalsky, C. A., Faber, M. S., Nath, A., Dann, H. E., Kelly, V. W., Liu, L., Shanker, P., Wagner, E. K., Maynard, J. A., Chan, C., and Whitehead, T. A. (2015) Rapid fine conformational epitope mapping using comprehensive mutagenesis and deep sequencing. *Journal of Biological Chemistry 290*, 26457–26470.

(10) Haas, C. M., Francino-Urdaniz, I. M., Steiner, P. J., and Whitehead, T. A. (2021) Identification of SARS-CoV-2 S RBD escape mutants using yeast screening and deep mutational scanning. *STAR Protoc 2*, 100869.

(11) Magoč, T., and Salzberg, S. L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics 27*, 2957–2963.

(12) Bosley, A. D.; Ostermeier, M. (2005) Mathematical Expressions Useful in the Construction, Description and Evaluation of Protein Libraries. *Biomolecular Engineering*, *22* (1–3), 57–61.