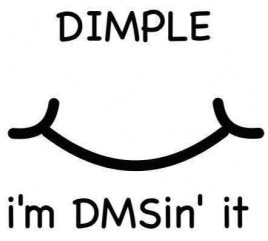
Apr 25, 2025 Version 6

DIMPLE library generation and assembly protocol V.6



DOI
https://dx.doi.org/10.17504/protocols.io.rm7vzy7k8lx1/v6

Christian Macdonald[1], David Nedrud[2], Patrick Rockefeller Grimes[1], Donovan Trinidad[1], Catherine Shin[1], Noah Greenwald[1], Matthew Howard[1], James Fraser[1], Willow Coyote-Maestas[1]

[1]University of California, San Francisco; [2]Bio-Techne

Christian Macdonald
University of California, San Francisco

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN ACCESS

DOI: https://dx.doi.org/10.17504/protocols.io.rm7vzy7k8lx1/v6

Protocol Citation: Christian Macdonald, David Nedrud, Patrick Rockefeller Grimes, Donovan Trinidad, Catherine Shin, Noah Greenwald, Matthew Howard, James Fraser, Willow Coyote-Maestas 2025. DIMPLE library generation and assembly protocol . **protocols.io** https://dx.doi.org/10.17504/protocols.io.rm7vzy7k8lx1/v6Version created by Noah Greenwald

Manuscript citation:
Macdonald, C.B., Nedrud, D., Grimes, P.R. *et al.* DIMPLE: deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. *Genome Biol* **24**, 36 (2023). https://doi.org/10.1186/s13059-023-02880-6

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working
We use this protocol and it's working

Created: April 17, 2025

Last Modified: April 25, 2025

Protocol Integer ID: 130951

Keywords: qcing mutagenic library, dimple protocol, dimple library generation, assembly protocol, mutagenic library, descriptions of qc, assembly, protocol, qc

Disclaimer

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to **protocols.io** is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with **protocols.io**, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

Abstract

This is a protocol for generating and QCing mutagenic libraries using the DIMPLE protocol. This version is updated to include expanded descriptions of QC and to clarify certain steps.

Materials

Enzymes and cells

PrimeSTAR GXL DNA Polymerase **Catalog #**R050A

MegaX DH10B T1R Electrocomptent Cells **Life Technologies Catalog #**C6400-03

NEB Golden Gate Assembly Mix **New England Biolabs Catalog #** E1601S (BsaI)

Kits (NEB kits given only as examples - any comparable kit should work)

Monarch PCR and DNA Cleanup Kit - 50 preps **New England Biolabs Catalog #**T1030S

Monarch® DNA Gel Extraction Kit **New England Biolabs Catalog #**T1020

Monarch Plasmid Miniprep Kit - 250 preps **New England Biolabs Catalog #**T1010L

Media and chemicals

SOC Outgrowth Medium - 100 ml **New England Biolabs Catalog #**B9020S

PCR Water (nuclease free) **Catalog #**PCPW

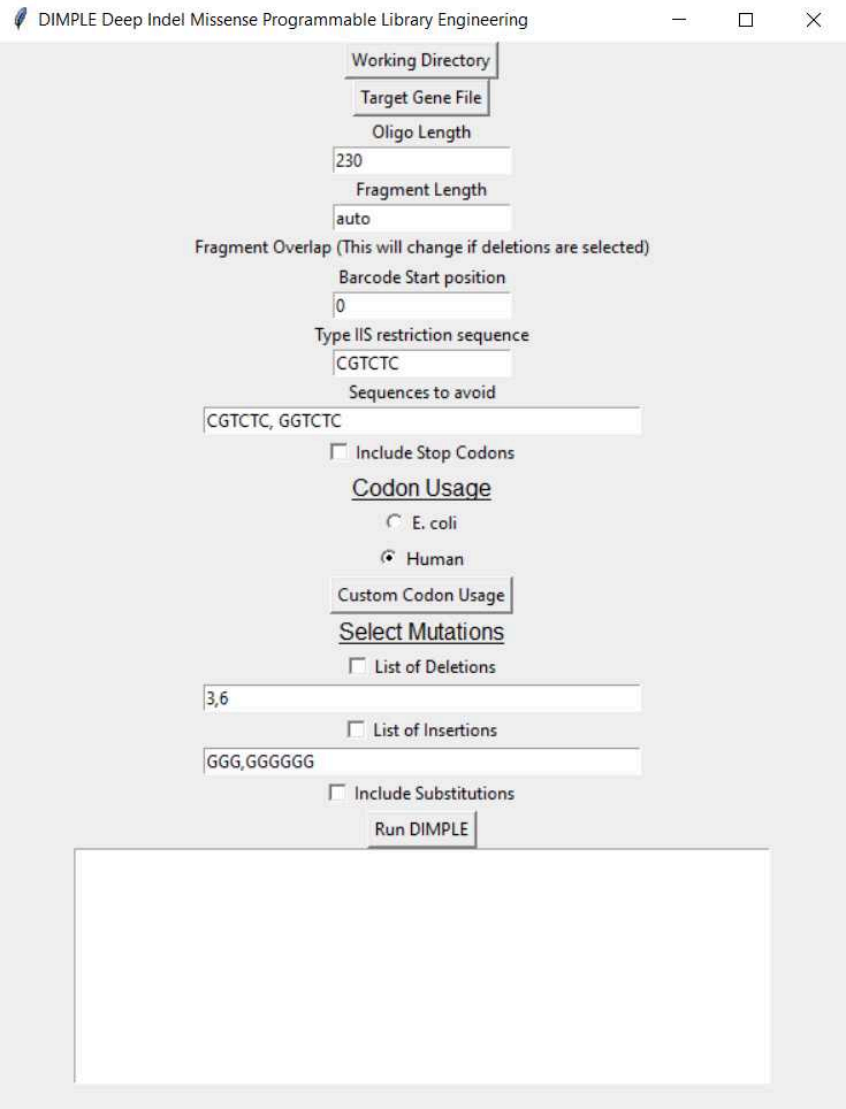
Equipment and consumables

- Thermocycler
- Electroporator
- Shaker
- Horizontal electrophoresis system
- Benchtop centrifuge
- OD meter

- Electroporation cuvettes (0.1 cm)
- Cuvettes for OD measurement
- Selection agar plates
- Agarose
- Optional: large BioAssay plates for plasmid purification

Preparation

1 **Use DIMPLE to generate mutagenic oligos and primers.**



Snapshot of the default DIMPLE GUI

2 **Important notes:** DIMPLE breaks a gene up into sub-library fragments and generates mutagenic insert oligo pools, where each oligo contains barcodes, Type IIS restriction cutsites, and a sub-region of the gene. Be sure to review your library generation vector and gene sequences and look for pre-existing Type IIS restriction sites. Use site-directed mutagenesis to remove unwanted off-target sites. Also note that, by default, DIMPLE

expects restriction enzymes with 6 bp recognition sites and 4 bp overhang lengths (as with BsaI and BsmBI).

3 **Input your wildtype gene sequence.**

Notes:
DIMPLE takes a reference sequence and provides a set of primers and oligos to mutagenize it according to the user's specifications. Specifying the mutagenic region properly is essential.

The reference gene sequence should contain the gene in the vector you are planning to construct the library in to allow primers for the fragments at either end to be properly created.

Example:

Snapgene view of target vector with region to be mutagenized (OCT1, red feature) selected. The sequence is displayed in a vector that will be used to assemble the library. Note that the displayed selection boundaries are **27** and **1688**.

Input wildtype gene sequence: Provide a fasta file of your gene sequence, including backbone as above. If multiple genes will be mutagenized in the same oligo pool, combine them all into one fasta file.

Optionally (but recommended), you can define the starting and ending nucleotide positions of your ORFs in each fasta sequence. Note that this uses **1 indexing**: for example, in the above sequence, we would provide the following fasta file with identical start and ending positions:

oct1_reference.fasta

```
>OCT1_lib_target_twist start:27 end:1688
GTGACGTCAGATCCGGCCGTCTCACCATGCCCACCGTGGATGACATTCTGGAGCAGGTTGGGGAGTCTGGCTGGTTCCAGAAG
CAAGCCTTCCTCATCTTATGCCTGCTGTCGGCTGCCTTTGCGCCCATCTGTGTGGGCATCGTCTTCCTGGGTTTCACACCTGA
CCACCACTGCCAGAGTCCTGGGGTGGCTGAGCTGAGCCAGCGCTGTGGCTGGAGCCCTGCGGAGGAGCTGAACTATACAGTGC
CAGGCCTGGGGCCCGCGGGCGAGGCCTTCCTTGGCCAGTGCAGGCGCTATGAAGTGGACTGGAACCAGAGCGCCCTCAGCTGT
GTAGACCCCCTGGCTAGCCTGGCCACCAACAGGAGCCACCTGCCGCTGGGTCCCTGCCAGGATGGCTGGGTGTATGACACGCC
CGGCTCTTCCATCGTCACTGAGTTCAACCTGGTGTGTGCTGACTCCTGGAAGCTGGACCTCTTTCAGTCCTGTTTGAATGCGG
GCTTCTTGTTTGGCTCTCTCGGTGTTGGCTACTTTGCAGACAGGTTTGGCCGTAAGCTGTGTCTCCTGGGAACTGTGCTGGTC
AACGCGGTGTCGGGCGTGCTCATGGCCTTCTCGCCCAACTACATGTCCATGCTGCTCTTCCGCCTGCTGCAGGGCCTGGTCAG
CAAGGGCAACTGGATGGCTGGCTACACCCTAATCACAGAATTTGTTGGCTCGGGCTCCAGAAGAACGGTGGCGATCATGTACC
AGATGGCCTTCACGGTGGGGCTGGTGGCGCTTACCGGGCTGGCCTACGCCCTGCCTCACTGGCGCTGGCTGCAGCTGGCAGTC
AGCCTGCCCACCTTCCTCTTCCTGCTCTACTACTGGTGTGTGCCGGAGTCCCCTCGGTGGCTGTTATCACAAAAAAGAAACAC
TGAAGCAATAAAGATAATGGACCACATCGCTCAAAAGAATGGGAAGTTGCCTCCTGCTGATTTAAAGATGCTTTCCCTCGAAG
AGGATGTCACCGAAAAGCTGAGCCCTTCATTTGCAGACCTGTTCCGCACGCCGCGCCTGAGGAAGCGCACCTTCATCCTGATG
```

*If you do not provide start & end position, the software will analyze different ORFs and ask you to select which one contains your gene of interest.

```
Analyzing Gene:pMV306_mOP_EccD3
ORF#1 MSENTVMPIVRVAVLAAGDDGGRLT...VGLFSLVLDR - length 475, strand 1, frame 2
ORF#2 MGIAEDSAQREDPSLIRQPAGRCEV...RGAPHLPGAR - length 102, strand 1, frame 2
ORF#3 MPRSEESHKSCRPLERKMARGRKHP...PESCTGVATR - length 89, strand 1, frame 3
ORF#4 MDAEAWLAGEKRLIEMWTPPQDRAK...EAMSKLAKTS - length 330, strand 1, frame 3
ORF#5 MSHIQRETSCSRPRLNSNMDADLYG...QFHLMLDEFF - length 271, strand 1, frame 3
ORF#6 MLLAAAGLVPTHRAAVLALAARRGS...WRHLCDRHFT - length 311, strand -1, frame 1
ORF#7 MVGRGINSVSQFSLTISSVTSLATL...SGAWGFPYKR - length 43, strand -1, frame 2
ORF#8 MVVVMFKSPVEHEAEQADQQRHHRD...LLNSAAGSEL - length 497, strand -1, frame 2
ORF#9 MPRGPCSSSYGPSCALGSSPRREAE...THAPELASSL - length 166, strand -1, frame 3
ORF#10 MYQAHGEAFAELTRATEPATPNSTP...VFSDILKLDS - length 367, strand -1, frame 3
Which ORF are you targeting? (number):1
MSENTVMPIV
Is this the beginning of your gene?(position 883) (y/n):y
VGLFSLVLDR
Is the size of your gene 1425bp? (y/n):n
Enter nucleotide length of your gene:1428
GLFSLVLDR*
Is this end correct? (y/n):y
```

It is important to include the plasmid backbone so that the software can avoid making primers that nonspecifically recognize a region outside of your gene.

```
Non specific Fragment:8
[162, 159, 159, 159, 156, 156, 159, 159, 159]
Creating Gene:pMV306_mOP_EccD3 --- Fragment:2-54
no thermodynamic data for neighbors 'AT/GT' available. Please check position manually:352 reverse
Primer:ATAGGTCTCcgccgatccggccgc
Match: ggcgcgcagcgccgcaccggccgc
Found non-specfic match at 790bp:
 match:cgaccacgcggcgacaccggccgc
primer:ATAGGTCTCcgccgatccggccgc Tm:35.03620585066153
------------------- Fragment size swapped due to non-specific primers -------------------
```

4 Using the DIMPLE GUI

4.1 *Working Directory:* When you open the DIMPLE GUI, first designate your working directory. This is the folder you wish your mutagenic oligo & primer outputs to be saved in.

4.2 *Target Gene File:* Upload your gene fasta text file with the Target Gene File button.

4.3 *Oligo length:* Designate the length of the oligos for the final pool. This will include the barcodes, Type IIS restriction cut site, and a region of your target gene that will be mutagenized. So, roughly, a given oligo length will allow fragments containing (oligo length - 50) bp of the gene.

4.4 *Fragment length:* You can adjust the length of your fragments, we recommend leaving the fragment length set to auto.

- DIMPLE will automatically break up your gene into roughly the same fragment sizes, and will determine which lengths work best to avoid matching overhangs.

4.5 *Fragment overlap:* The number of base pairs shared between sub-regions. We have seen that setting overlap to 0 base pairs leads to errors. The overlap is set to 4 base pairs by default.

4.6 *Barcode start position:* The software selects from a set pool of barcodes when designing oligos. You can define which number barcode in the list this starts from. Unless you have a very good reason to, this can safely be left at 0.

4.7 *Type IIS restriction sequence:* You can select which Type IIS restriction cutsite you'd like to append to your inserts and backbone. You can choose between the BsmBI (CGTCTC) or BsaI (GGTCTC) sequences.

4.8 *Sequences to avoid:* This will allow the code to design fragments without the cutsite you selected in the *Type IIS restriction sequence* section. If there is an error when running, this implies you have a cutsite in your gene or vector that needs to be removed.

```
Exception in Tkinter callback
Traceback (most recent call last):
  File "C:\Program Files\WindowsApps\PythonSoftwareFoundation.Pyth
    return self.func(*args)
  File "C:\Users\ASUS\Desktop\Fraser_CoyoteMaestas_Lab\DMS\DIMPLE-
    OLS = addgene(app.geneFile)
  File "C:\Users\ASUS\Desktop\Fraser_CoyoteMaestas_Lab\DMS\DIMPLE-
    tmpOLS.append(DIMPLE(gene, start, end))
  File "C:\Users\ASUS\Desktop\Fraser_CoyoteMaestas_Lab\DMS\DIMPLE-
    raise ValueError('Unwanted Restriction cut sites found. Please
ValueError: Unwanted Restriction cut sites found. Please input pla
```

4.9 *Codon usage:* Your oligos can be codon optimized to your organism of choice. The default options are *E. coli* or *Human*, but you can also upload a codon usage table for any organism by pressing "Custom codon usage."

4.10 *Select mutations:*

Deletions: If you'd like to generate a library with deletions across your gene, select "List of deletions". Enter how many **base pairs** you'd like to be deleted across your library.

- **For example,** if you'd like deletions that are one codon long, enter "3" for 3 nucleotides.
- **For example,** If you'd like to include both deletions that are one codon long AND two codons long in the same pool, enter "3,6" for both 3, and 6 nucleotide long deletions.

Insertions: If you'd like to generate a library with different sequences inserted across your gene, select "List of Insertions" and provide the nucleotides you wish to be inserted **(NOT amino acids!)**.

- **For example,** If you'd like to insert single glycine residues between every codon of your gene, include "GGG" (or "GGA", etc) in the box.
- **For example,** If you'd like to insert glycine-serine pairs between every codon of your gene, include "GGGTGC" in the box.
- **For example,** if you'd like to insert glycine AND glycine-serine residues between every codon of your gene, include "GGG,GGGTGC" in the box.

Substitutions: If you'd like to generate a library where an amino acid at each position is swapped with every other possible amino acid, select "Include Substitutions."

4.11 *Run DIMPLE:* Hit "Run DIMPLE" when you are ready to generate your pool.

5 **Examples of what the running code and outputs look like:**

The code will first iterate the ideal fragment length sizes across each gene of interest.

```
-------------------------------- Analyzing Gene:pMV306_mOP_EccD3 --------------------
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:1
[162, 156, 159, 159, 159, 159, 159, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:2
[162, 159, 156, 159, 159, 159, 159, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:3
[162, 159, 159, 156, 159, 159, 159, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:4
[162, 159, 159, 159, 156, 159, 159, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:6
[162, 159, 159, 159, 156, 156, 162, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:7
[162, 159, 159, 159, 156, 156, 159, 162, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:8
[162, 159, 159, 159, 156, 156, 159, 159, 159]
```

5.1 Next, barcodes are assigned to each fragment. The code will update how many barcodes it attempted to use to generate each fragment pool, and will remove these from the remaining barcode pool.

This will continue for each gene included in your fasta file.

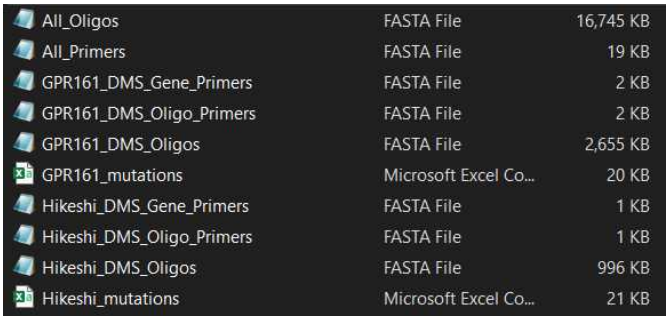
```
Creating Gene:Hikeshi --- Fragment:43-81
Barcodes used:9
Barcodes Remaining:1066
Creating Gene:Hikeshi --- Fragment:83-119
Barcodes used:3
Barcodes Remaining:1063
Creating Gene:Hikeshi --- Fragment:121-158
Barcodes used:6
Barcodes Remaining:1057
Creating Gene:Hikeshi --- Fragment:160-198
no thermodynamic data for neighbors 'GT/AG' available. Please check position manually:258 forward
Primer:ATAGGTCTCAGACAGCAAATGATGAAGCAAAATTGTAG
Match: CACGAATGGGAAGCCAAGTGCCATCTTCAAAATTTCAG
Barcodes used:9
Barcodes Remaining:1048
```

5.2 There is a final QC check for each primer set generated.

```
Running QC for barcode primer specificity
Checking primer set:TSHR_oligoP_DMS-1
Checking primer set:TSHR_oligoP_DMS-2
Checking primer set:TSHR_oligoP_DMS-3
Checking primer set:TSHR_oligoP_DMS-4
Checking primer set:TSHR_oligoP_DMS-5
Checking primer set:TSHR_oligoP_DMS-6
Checking primer set:TSHR_oligoP_DMS-7
Checking primer set:TSHR_oligoP_DMS-8
Checking primer set:TSHR_oligoP_DMS-9
Checking primer set:TSHR_oligoP_DMS-10
Checking primer set:TSHR_oligoP_DMS-11
Checking primer set:TSHR_oligoP_DMS-12
Checking primer set:TSHR_oligoP_DMS-13
Checking primer set:TSHR_oligoP_DMS-14
Checking primer set:TSHR_oligoP_DMS-15
Checking primer set:TSHR_oligoP_DMS-16
Checking primer set:ABCG2_oligoP_DMS-1
Checking primer set:ABCG2_oligoP_DMS-2
Checking primer set:ABCG2_oligoP_DMS-3
Checking primer set:ABCG2_oligoP_DMS-4
Checking primer set:ABCG2_oligoP_DMS-5
Checking primer set:ABCG2_oligoP_DMS-6
Checking primer set:ABCG2_oligoP_DMS-7
Checking primer set:ABCG2_oligoP_DMS-8
Checking primer set:ABCG2_oligoP_DMS-9
Checking primer set:ABCG2_oligoP_DMS-10
Checking primer set:ABCG2_oligoP_DMS-11
```

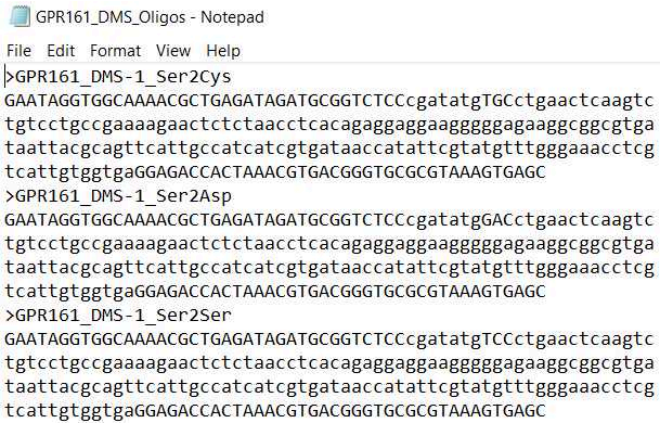
5.3 All outputs are saved in your working directory.

If you are using DIMPLE to mutagenize several genes, there will be separate files for the primers and mutagenic inserts for each gene, as well as a master list of **all** oligos and primers.



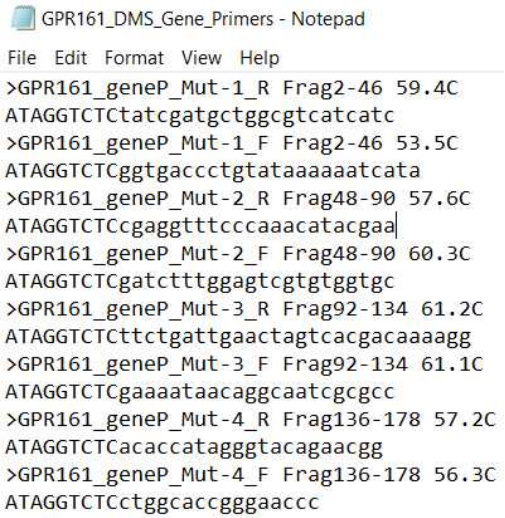
For each gene, DIMPLE generates a list of:

- **EXAMPLE_DMS_Oligos.** This is the list of sub-library oligo pools, where each oligo contains barcodes, Type IIS restriction cutsites, and a sub-region of the gene.

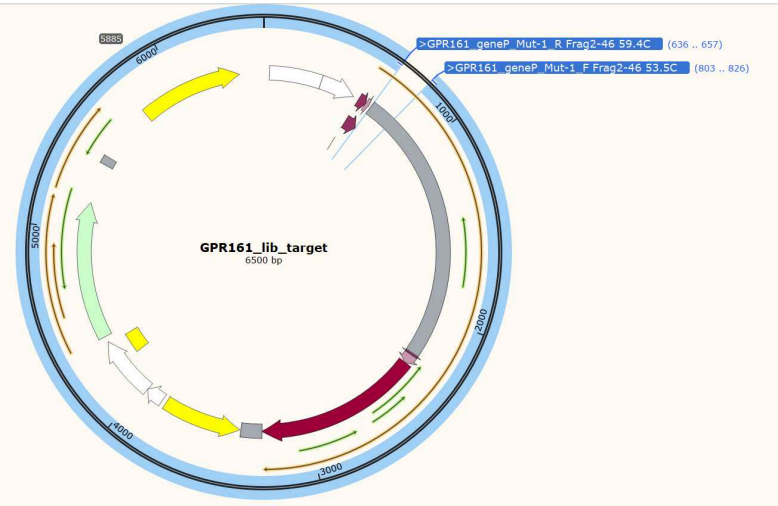


Screenshot of mutagenic inserts generated by DIMPLE

- **EXAMPLE_DMS_Gene_Primers.** These are primers used to add cutsites to and amplify the backbone each mutagenic insert will be ligated into.



Screenshot of backbone primers generated by DIMPLE. Each primer name also lists its melting temperature.



Each primer pair linearizes the backbone

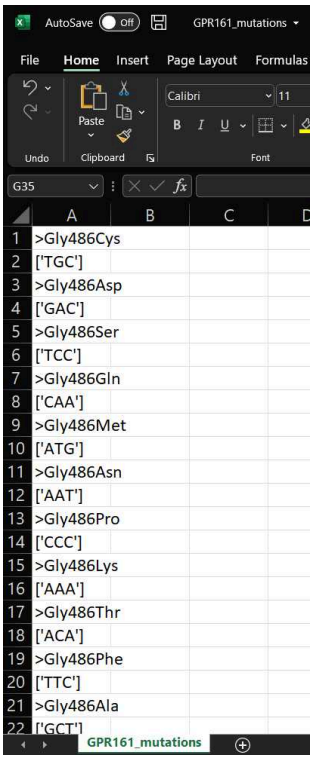
- **EXAMPLE_DMS_Oligo_Primers.** These are primers used to amplify the mutagenic inserts from your oligo pool.

GPR161_DMS_Oligo_Primers - Notepad

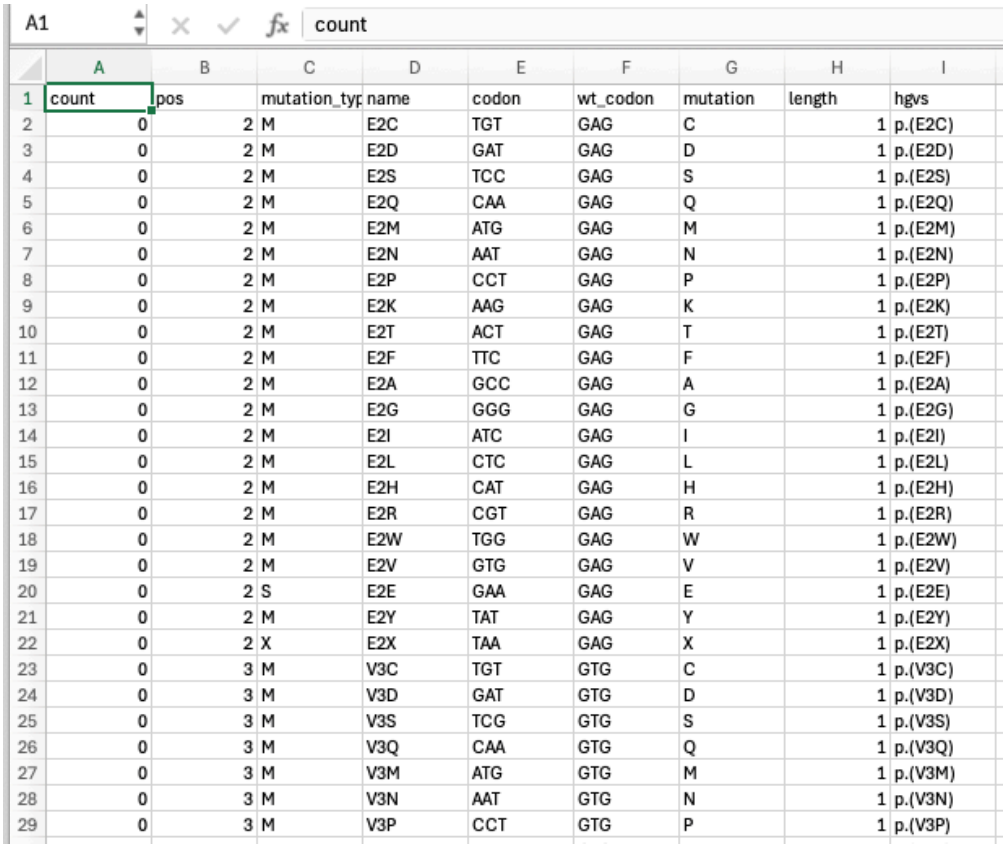
File Edit Format View Help

```
>GPR161_oligoP_DMS-1_F Frag2-46_57.9C
GAATAGGTGGCAAAACGCTGA
>GPR161_oligoP_DMS-1_R Frag2-46_58.3C
GCTCACTTTACGCGCACC
>GPR161_oligoP_DMS-2_F Frag48-90_58.6C
AGTTTGGGGATCTCGTACACG
>GPR161_oligoP_DMS-2_R Frag48-90_57.8C
TGAGGACTTTTGCGTTCGAC
>GPR161_oligoP_DMS-3_F Frag92-134_55.9C
ATCCATAGCGTCCCTATTGC
>GPR161_oligoP_DMS-3_R Frag92-134_58.7C
TTCCCCGTCATGCTTAGGGA
>GPR161_oligoP_DMS-4_F Frag136-178_58.6C
CTCTAACGGCGTGACACTTAG
>GPR161_oligoP_DMS-4_R Frag136-178_58.5C
TTGGCTAATAGTGGGAGGCG
```

- **EXAMPLE_mutations.** This is the list of mutations that will be present in your final library.



- **EXAMPLE_designed_variants.** This is a more detailed list of the variants expected to be present in your final library. This file can be directly used to analyze a DMS experiment using **Dumpling**, a pipeline we created to perform QC and generate scores with DIMPLE libraries.



A1 count

	A	B	C	D	E	F	G	H	I
1	count	pos	mutation_typ	name	codon	wt_codon	mutation	length	hgvs
2	0	2	M	E2C	TGT	GAG	C	1	p.(E2C)
3	0	2	M	E2D	GAT	GAG	D	1	p.(E2D)
4	0	2	M	E2S	TCC	GAG	S	1	p.(E2S)
5	0	2	M	E2Q	CAA	GAG	Q	1	p.(E2Q)
6	0	2	M	E2M	ATG	GAG	M	1	p.(E2M)
7	0	2	M	E2N	AAT	GAG	N	1	p.(E2N)
8	0	2	M	E2P	CCT	GAG	P	1	p.(E2P)
9	0	2	M	E2K	AAG	GAG	K	1	p.(E2K)
10	0	2	M	E2T	ACT	GAG	T	1	p.(E2T)
11	0	2	M	E2F	TTC	GAG	F	1	p.(E2F)
12	0	2	M	E2A	GCC	GAG	A	1	p.(E2A)
13	0	2	M	E2G	GGG	GAG	G	1	p.(E2G)
14	0	2	M	E2I	ATC	GAG	I	1	p.(E2I)
15	0	2	M	E2L	CTC	GAG	L	1	p.(E2L)
16	0	2	M	E2H	CAT	GAG	H	1	p.(E2H)
17	0	2	M	E2R	CGT	GAG	R	1	p.(E2R)
18	0	2	M	E2W	TGG	GAG	W	1	p.(E2W)
19	0	2	M	E2V	GTG	GAG	V	1	p.(E2V)
20	0	2	S	E2E	GAA	GAG	E	1	p.(E2E)
21	0	2	M	E2Y	TAT	GAG	Y	1	p.(E2Y)
22	0	2	X	E2X	TAA	GAG	X	1	p.(E2X)
23	0	3	M	V3C	TGT	GTG	C	1	p.(V3C)
24	0	3	M	V3D	GAT	GTG	D	1	p.(V3D)
25	0	3	M	V3S	TCG	GTG	S	1	p.(V3S)
26	0	3	M	V3Q	CAA	GTG	Q	1	p.(V3Q)
27	0	3	M	V3M	ATG	GTG	M	1	p.(V3M)
28	0	3	M	V3N	AAT	GTG	N	1	p.(V3N)
29	0	3	M	V3P	CCT	GTG	P	1	p.(V3P)

Library assembly

3d

6 **Prepare oligo pool stock**

Follow any recommendations for oligo pool resuspension: typically, this results in a 10 ng/µL solution, which the following steps assume. Lower or higher concentrations may require alterations.

7 **PCR amplification of oligos, backbone**

Thaw all components beforehand and follow general directions provided by manufacturer.

Amplification of mutagenic inserts and backbones should ideally be performed in parallel, assuming one has two thermocyclers. Alternatively, the backbone amplification can be performed first, then the the insert amplification can be done during the gel purification of the backbone.

8 **PCR amplification of backbone.**

Prepare a master mix with PrimeSTAR GXL polymerase:

	A	B	C	D	E
	Component	**Total amount in master mix (µL)**	Amount/reaction (µL)	Comment	Number of regions
	dNTP	**20**	4		5
	5X buffer	**50**	10		
	Template (backbone)	**5**	1	~10 ng	
	Enzyme	**5**	1		
	Nuclease-free H2O	**165**	33		
	Primers		1	Single mix of fwd & reverse at 10 µM each	
	Total	**245**	50		

Mix by vortexing and spin down each reagent (except for enzyme) prior to adding to master mix.

8.1 Transfer 49 µL of master mix into separate tubes for each reaction.

8.2 Add reaction-specific primers to each tube: 1 µL (combined at 10 µM each). Mix and spin down.

8.3 Place on thermocycler and amplify:

	A	B	C
	Step	T (° C)	Time (sec)
	1	98	10

	A	B	C
	2	55	15
	3	68	60 per kb
	4	Repeat 1-3 15-24 times	See note
	5	10	Hold

Note: the number of cycles should be optimized to minimize PCR cycles. Increasing cycles introduces the possibility of error. An initial comparison of 16, 18, and 22 total cycles can be used to find a minimum number of cycles which yields sufficient DNA for the assembly reaction (see **step 11** below) and no more. Increasing the template concentration (to 50 or 100 ng) could be preferable to increasing PCR cycles if large numbers are required for amplification. We suggest testing this on 1-2 subpools before scaling up to all subpools in a gene.

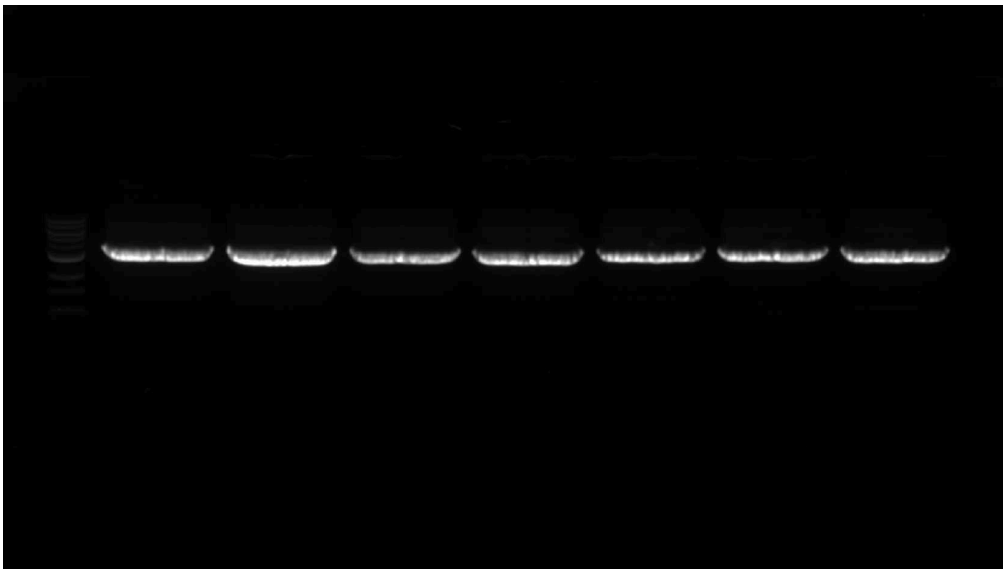
8.4 **Purify amplified product by gel extraction.**

Prepare an agarose gel for gel purification (0.5% - 0.75%). Ideally, it should be of sufficient size to load all samples.

8.5 Load and run backbone products.

8.6 Use a scalpel or razor blade to (carefully!) cut out each product. Using a gel extraction kit, purify the product and elute in 10 µL elution buffer .

8.7 **Important QC step** - evaluate the quality of the amplification with the gel. Each lane should have a single, crisp band at the expected size. Multiple bands, or no bands, might indicate improper annealing and may require Tm or extension time optimization. Blurry and smeary bands might indicate problems with the electrophoresis or amplification conditions. It is essential to get specific, clean amplification of the backbone, and libraries should not be generated with messy reactions.



Example visualization of expected backbone amplification reactions on SYBR Safe-stained agarose gel. Each lane has a single, crisp band at the expected position that is well-separated from other lanes.

9 **PCR amplification of mutagenic inserts.**

Prepare a master mix with PrimeSTAR GXL polymerase:

	A	B	C	D	E
	Component	**Total amount in master mix (µL)**	Amount/reaction (µL)	Comment	Regions
	5X buffer	**50**	10		5
	dNTP	**20**	4		
	Oligo pool	**5**	1	10 ng pool per reaction, assuming a stock at 10 ng/µL	
	Enzyme (PrimeSTAR GXL)	**5**	1		
	Primers		1	Single mix of fwd & reverse at 10 µM each	
	Nuclease-free H2O	**165**	33		
	Total	**245**	50		

	A	B	C	D	E

Mix by vortexing.

9.1 Transfer 49 µL insert master mix into separate tubes for each reaction.

9.2 Add reaction-specific primers to each tube: 1 µL (combined at 10 µM each). Mix and spin down.

9.3 Place on thermocycler and amplify:

	A	B	C
	Step	T (° C)	Time (sec)
	1	98	10
	2	55	15
	3	68	30
	4	Repeat 1-3 15-24 times	See note
	5	10	Hold

Note: the number of cycles should be optimized to minimize PCR cycles. Increasing cycles introduces both PCR bias and error. An initial comparison of 16, 18, and 22 total cycles can be used to find a minimum number of cycles which yields sufficient DNA for assembly step and no more. For the example libraries here, this is ~100 ng per oligo pool. We suggest testing this on 1-2 subpools before scaling up to all subpools in a gene.

9.4 Use a PCR cleanup kit to purify each product. Elute in 10 µL elution buffer.

10 **Important QC step** - run a gel with each PCR product and visualize it. Ideally, each reaction should give a single strong band at the expected size. Failure to do so may require changing PCR conditions.

11 **Golden gate assembly**

Prepare a master mix for the assembly. Use the table below to calculate: adjust number of regions and amount of necessary backbone and insert for desired amounts per

reaction. We have found 3 µL backbone and 1 µL oligo (insert) are usually good.

	A	B	C	D	E
	Component	**Total amount (µL)**	Amount/reaction (µL)	Notes	Number of regions
	10X buffer	**20**	4		**5**
	Enzyme	**10**	2		
	Backbone		3	300 ng	
	Insert (oligos)		1	2:1 molar ratio: for example libraries here, ~30-40 ng	
	Nuclease-free H2O	**150**	30		
	Total	**200**	40		

Vortex to mix.

11.1 Transfer appropriate master mix into separate tubes for each reaction.

Example: if each reaction has 3 µL backbone and 1 µL insert, then transfer **36 µL** to each tube.

11.2 To each tube, add the appropriate vector/insert pair:

For this example:

3 µL backbone (purified)

1 µL insert (purified)

Mix well and spin down.

11.3 Place on thermocycler and run the following program:

	A	B	C
	Step	T (° C)	Time (min)
	1	37	5

	A	B	C
	2	16	10
	3	Repeat 1-2 34 times (35 total cycles)	
	4	60	5
	5	10	Hold

Important: This assumes the assembly is performed with BsaI. If using a different enzyme (such as BsmBI), check the recommended temperatures from the vendor. If needed, a larger number of cycles can be used.

11.4 Use a PCR cleanup kit to purify each product. **Elute in** 10 µL nuclease-free H2O .

Important: using elution buffer or TE may cause electroporation to fail!

11.5 **Important QC step**: before electroporating libraries, transform 1 µL of each library into chemically competent cells and plate. If no or very few colonies are observed the next day, this could suggest a failure during assembly. Further check the assembly by running a small portion on a gel.

Transformation and recovery of sublibraries

2d

12 Using a high-efficiency electrocompetent strain of *E. coli*, such as MegaX DH10B, prepare a transformation for each assembly: 60 ng assembly product with 20 µL electrocompetent cells in a 0.1 cm cuvette. Follow the specific instructions for voltage and outgrowth media corresponding to your cells and electroporator. Outgrow cells for 1 hour before proceeding. We've found that 25uL of electrocompetent cells can speed the subpool growth process, but this increases the cost proportionally.

12.1 During the 1 hour of outgrowth, prepare tubes for the serial dilutions, as well as flasks to grow the cells.

12.2 For the serial dilutions, add 90 µL H_2O to six 200 uL microcentrifuge tubes for each subpool, and reserve for step 13.

12.3 For the bacterial outgrowth, aliquot 25 mL LB supplemented with appropriate antibiotic in either a Falcon tube or glass culture tube, and reserve for step 14.

12.4 Following 1-hour of growth, perform just the first part of step 13 (adding 10uL of outgrowth to the first tube of each dilution), and then take the remainder of the culture and begin growing it in the larger flasks (step 14).

13 **Essential QC step:** Count total number of transformants to ensure adequate coverage.

Take the strip tubes from step 12.2 and add 10 µL of the outgrowth mixture to the first tube, pipetting extremely slowly as this mixture is very viscous. Mix well by pipetting up and down, and add 10 µL of this to the second tube. Repeat for the rest of the tubes.

Note: Following this, proceed to step 14 (outgrowth) to start growing up subpools. Once the subpools are growing, continue the rest of the QC steps below.

13.1 Plate 90 µL of the last three dilutions (1:10,000, 1:100,000, 1:1,000,000) on LB-agar with appropriate antibiotic. Grow overnight and count colonies the next day. Calculate the total number of transformants and determine the variant coverage per reaction. At least 100-fold coverage is **essential**. Note: the relevant measurement here is not transformation *efficiency* (i.e., cfu/ng), but simply *total transformants!*

13.2 **Example calculations**

Let's say you are working with oligos with fragments of 50 codons from the protein of interest. For each residue, you generated 20 substitutions (including 1 synonymous), 3 insertions, and 3 deletions, for a total of 26 different variants at each position. Then this subpool will have a total of **1300** possible variants.

In order to have 100-fold coverage of these subpool, you would need to obtain $100 \times 1300 = 130,000$ total transformed bacteria. In order to get 200-fold coverage, you would need $200 \times 1300 = 260,000$ total transformed bacteria.

Given that we are calculating the total number of transformed bacteria, it is important to note that the relative dilution of a given plate (1:10, 1:100, etc) isn't the only thing that matters. The other key factor is the fraction of all transformed bacteria that went into the initial dilutions. Thus, to determine how the number of colonies on a given plate corresponds to the total number of transformed bacteria, we use the following formula:

$$N = C_n \times 10^{n-1} \times 100$$

where C_n is the total number of colonies on a particular plate, 10^{n-1} accounts for the serial dilutions, and 100 accounts for the fact that only 10 µL out of the 1 mL outgrowth media is being sampled.

For example, the number of colonies seen on the first plate will need to be multiplied by 100 to get the total number of transformed bacteria:
$C_1 \times 10^{1-1} \times 100 = C_1 \times 100$
That's because we took 10uL from 1mL of total culture, and the fact that we dilute that 10uL into 90uL doesn't impact the total number of bacteria that are plated.

The number of colonies seen on the second plate will need to be multiplied by 1000 to get the total number of transformed bacteria:
$C_2 \times 10^{2-1} \times 100 = C_2 \times 1000$
That's because we took 10uL from 1mL of total culture, dilute that 10uL into 90uL of water, and take 10uL from that to the second plate.

Thus, for a subpool with 1300 variants, we would need to see at least the following number of colonies in each plate to ensure 100X coverage, which corresponds to 130,000 total transformants:

	A	B
	Plate number	Minimum number of colonies
	1	1,300
	2	130
	3	13
	4	1.3
	5	0.13
	6	0.013

We usually see a substantial excess of colonies over the required coverage, for example multiple colonies on the 4th and 5th plates. We only plate the last three dilutions to avoid counting extremely dense clusters of colonies.

To determine if a given subpool has sufficient coverage, calculate the number of implied transformants from the plated dilutions and average them together. For example, let's assume we observed the following number of plated colonies from two subpools, each with 1300 variants

	A	B	C
	Plate number	Number of observed colonies in subpool 1	Number of observed colonies in subpool 2
	4	Many	6
	5	20	1
	6	3	0

This would imply the following total transformants for each of the dilutions by plugging into the equation described above:

	A	B	C
	Plate number	Subpool 1 implied total	Subpool 2 implied total
	4	NA	600,000
	5	20,000,000	1,000,000
	6	30,000,000	NA

Note that even though subpool 2 has significantly fewer implied transformants than subpool 1, it is still over the minimum (130k) needed for 100x coverage.

Note: It is challenging to estimate coverage if you observe plates without any colonies where the expected number of colonies is also less than 1, since measuring fractions of colonies is not possible. Therefore, it is important to adjust which of the serial dilutions you plate to ensure that you are consistently observing a measurable number of colonies (ideally 1-50) on each plate.

13.3 **Important QC step:** pick single colonies from the QC plates and use clonal sequencing to check for mutations (e.g., Sanger sequencing or whole-plasmid sequencing). We typically pick 5 colonies per subpool and expect the majority to contain a desired mutation. Too much WT could indicate problems during the backbone amplification, and too many off-target mutations could suggest a failure to amplify or purify properly. In both cases, these would be extremely deleterious to the quality of the final library.

14 Add the remaining outgrowth culture to 25 mL LB supplemented with appropriate antibiotic for further outgrowth in either a Falcon tube or glass culture tube. Shake until the culture reaches OD 0.6-0.7.

Note: it is important to harvest the cells in the exponential phase. This takes 4-6 hours in our experience, but may vary (including between samples in the same day).

15 Harvest DNA from cultures using a miniprep kit, splitting each reaction across 3 columns. Elute each column in 40 µL elution buffer and combine elutions corresponding to a single sublibrary.

Sublibrary pooling and subcloning

2d

16 Measure each sublibrary concentration with a fluorometric method, such as Qubit.

16.1 First, dilute each sublibrary 1:5 in H20. Use this as your working stock.

16.2 From the working stock of each sublibrary, generate two independent replicates of the sublibrary concentration.

16.3 Average the two estimates of sublibrary concentration, and then use these concentration estimates to combine the working stock subpools together into a single pooled library, ensuring that the same total amount of DNA is included from each working stock subpool.

17 Subclone the pooled library to transfer from library vector to selection vector. Depending on construct design, this may require digestion/ligation or a Golden Gate reaction.

18 Use PCR cleanup kit to purify reaction, and elute in 10 µL H_2O.

19 Using a high-efficiency electrocompetent strain of *E. coli*, such as MegaX DH10B, transform 3 µL subcloned product with 20 µL electrocompetent cells cells in a 0.1 cm cuvette. Follow the specific instructions for voltage and outgrowth media corresponding to your cells and electroporator. Outgrow cells for 1 hour before proceeding.

20 Repeat transformation and isolation following steps for subpool cloning and quantification as above, except use variant counts for the total library. Again, at least 100-fold coverage is **essential** for maintaining unbiased libraries. Pick colonies and submit for clonal sequencing as well: we typically submit ~10 samples for our final pooled library.