

# BIO122 Lab: Designing a Gene for Expression using Modular Components

**Aim:** To learn to design a gene for expression and testing a GFP construct

## Learning goals:

1. Practical understanding of the concept of a gene
2. To use sequence editing software
3. Understand the role of the regulatory steps in gene expression
4. Simulate the first step in recombinant protein expression
5. Design a test at the level of the construct to know if the construct is successful
6. Represent the result of the design

## Introduction:

The process of biological engineering has multiple components and the most basic step is genetic engineering. In 1979 the first human growth hormone (hGH) was produced as a recombinant protein in bacteria, as a part of the recombinant DNA (rDNA) revolution(1, 2). This was commercialized by the company Genentech, USA. This hGH is a therapeutic for dwarfism, bone fractures and burns recovery.

In order to develop more complex genetically engineered systems, to perform functions that go beyond treating them as 'cell factories', a Systems Biology level understanding is required- of the parts and their interconnections. How such systems biology knowledge can be transferred into a functional 'genetically engineered machine' (GEM) is determined by the following steps:

- 1) Problem definition
- 2) Abstraction
- 3) Design
- 4) Research into the genes
- 5) Synthesis



**Fig. 1.** The Part is inserted in a standard plasmid backbone. The part is labeled with a unique number, flanked before and after the sequence by a Prefix (Pref) and Suffix (Suff).

In order to achieve the synthesis in a manner that is reproducible and modular, we need some standard. These are defined by the BioBricks (BBa) standards. All **Parts** are inserted into a standard plasmid backbone (Fig. 1). The **Part** is defined by a **BBa\_** number, where the underscore is followed by a unique identifier (Fig. 1). Once you have such a standardized part in a standardized plasmid, you need to combine the parts to develop the device. Knowing which bit of the part to take, requires you to search for the marker sequences **Prefix** and **Suffix**. Prefix (Pref) is the part pre-ceding the PART of interest and suffix (Suff) the part following it as seen in **Table 1**.

**Table 1.** The prefix and suffix sequences to look for in the BioBricks<sup>1</sup>.

S.No.	Part name	Condition	
1	Pref	Prefix, if the part starts with	5' GAATTCGCGCCGCTTCTAG '3

<sup>1</sup> <http://parts.igem.org/Help:Prefix-Suffix>

		ATG	
2	Pref	Prefix for all other parts	5' GAATTCGCGGCCGCTTCTAGAG '3
3	Suff	Suffix for all parts	5' TACTAGTAGCGGCCGCTGCAG '3

In order to make a construct with no pre-existing sequence in the BioBricks repository, you will need to amplify the gene using PCR primers. Designing PCR primers is based on approximately 15 base pair complementary regions to the template DNA. Primer-BLAST<sup>2</sup> is an online program that can allow such design.

To use this program, you need to use a FASTA format for the sequence of interest. FASTA is a text-based format for representing either nucleotide sequences (5' to 3') or peptide sequence (N to C terminal). Here, nucleotides or amino acids are represented using single-letter codes. The format may be preceded by comments and names of the sequences.

Table 2. The nucleotide single letter codes taken from NCBI Learning Center<sup>3</sup>

Single Letter Codes for Nucleotides				
Nucleotide(s)	Code	Acceptable by		Treated as (in alignment)
		makeblastdb/ formatdb	BLAST	
A	A	+	+	A
C	C	+	+	C
G	G	+	+	G
T	T	+	+	T
U	U	+	+	T
A, C	M (amino)	-	+	N
A, G	R (purine)	-	+	N
A, T	W (weak)	-	+	N
C, G	S (strong)	-	+	N
C, T	Y (pyrimidine)	-	+	N
G, T	K (keto)	-	+	N
A,C,G	V (not T)	-	+	N
A,C,T	H (not G)	-	+	N
A,G,T	D (not C)	-	+	N
C,G,T	B (not A)	-	+	N
A,C,G,T	N (any nucleotide)	+	+	N

Proteins are written in single letter amino-acid notations as seen in Table 3.

Table 3. The single letter amino acid code.

1-letter	3-letter	description
A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartic acid
C	Cys	Cysteine
Q	Gln	Glutamine
E	Glu	Glutamic acid
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine

<sup>2</sup> <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>

<sup>3</sup> [https://www.ncbi.nlm.nih.gov/staff/tao/tools/tool\\_lettercode.html](https://www.ncbi.nlm.nih.gov/staff/tao/tools/tool_lettercode.html)

L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine
B	Asx	Aspartic acid or Asparagine
Z	Glx	Glutamine or Glutamic acid
X	Xaa	Any amino acid

The constructs you will use today for building a minimal gene-expression system consist of 3 plasmids:

- (1) a lactose promoter (pLac) derived from the Lac Operon cloned into a plasmid,
- (2) a cassette consisting of a ribosome binding site (RBS) followed by a coding sequence for GFP-mut3 and a double-terminator of transcription and
- (3) cloning backbone with an origin of replication, *bla* gene for ampicillin resistance and the E-X—S-P prefix and suffix.

The purpose of such modularity is to be able to clone genes from one set of plasmids into another, while maintaining the “standard” structure (prefix-geneticElement-suffix).

Here, we will be making one such expression construct in theory, that we can indeed perform in experiment, so as to give you an idea of how one might go about designing a genetic construct for expression.

## Materials

1. APE: a plasmid editor program. The program has been installed on your workstations.
2. 2155 bps linearized plasmid DNA pSB1A3 (Amp selection)  
<http://parts.igem.org/Part:pSB1A3>
3. 2270 bps plasmid pSB1lacP promoter sequence (BBa\_R0010) inside a pSB1C3 backbone [http://parts.igem.org/Part:BBa\\_R0010](http://parts.igem.org/Part:BBa_R0010)
4. 2945 bps plasmid containing rbs + GFPmut3 + ter1 + ter2 sequence allows us to include a 'functional' version of the sequence for translation (BBa\_I13504)  
[http://parts.igem.org/Part:BBa\\_I13504](http://parts.igem.org/Part:BBa_I13504)

## Procedure

1. You are provided the Promoter and RBS+ in the pSB1C3 backbone with a chloramphenicol resistance marker containing GFP and a lac promoter.
2. Note where the prefix, suffix and enzyme cutting sites are located based on the plasmid map provided for pSB1C3 from <http://parts.igem.org/Part:pSB1C3>
3. The rbs + GFPmut3 + ter1 + ter2 sequence allows us to include a 'functional' version of the sequence for translation (BBa\_I13504) [http://parts.igem.org/Part:BBa\\_I13504](http://parts.igem.org/Part:BBa_I13504)
4. The lacP promoter sequence (BBa\_R0010) has been put inside a pSB1C3 backbone [http://parts.igem.org/Part:BBa\\_R0010](http://parts.igem.org/Part:BBa_R0010)
5. Simulate the cutting of the sequences of the two plasmids by EX and SP by copy pasting the sequence corresponding to the restriction enzyme sites.
6. Copy paste these two sequences into a third file and save.
7. Copy paste this lac promoter + rbs + GFPmut3 + ter1 + ter2 sequence into the backbone. Use the backbone pSB1A3 (Amp selection)  
<http://parts.igem.org/Part:pSB1A3>

## RECORD KEEPING:

In your lab-notebooks note the following:

1. The size of the 3 plasmids
2. The schematic of the cloning strategy (i.e. draw the 3 plasmids you started with, the enzymes you digested them with virtually, the ligation reaction, planned transformation and selection.
3. Sketch a representative figure of the DNA digest (in APE "Enzyme Selection" --> "Digest")
4. Note the sequence of the transcriptional terminator in your construct.
5. How the transcriptional terminator function?

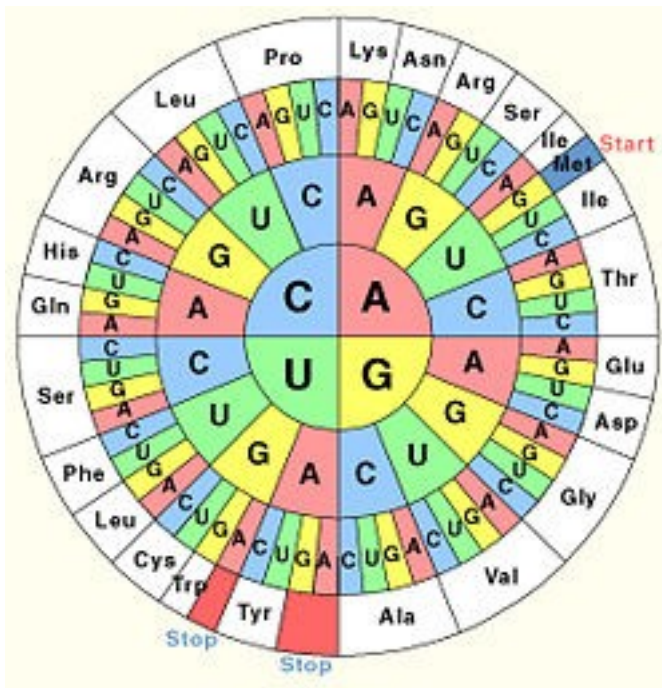


Fig. 2. The codon table

**Resources:**

1. Gene at PubMed, NIH USA <https://www.ncbi.nlm.nih.gov/gene/> [Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.]
2. Molecular Biology of the Gene by James Watson et al.
3. Molecular Biology of the Cell by Bruce Alberts et al.
4. BioBuilder: A textbook and resource for synthetic biology
5. Codon usage database <http://www.kazusa.or.jp/codon/>
6. Nucleotide sequence resource <https://www.ncbi.nlm.nih.gov/nucco>
7. Sequence details of GFP-mut3 <https://www.addgene.org/20410/>

**References:**

1. Seeburg PH, et al. (1978) Synthesis of growth hormone by bacteria. *Nature* 276(5690):795–798.
2. Goeddel D V, et al. (1979) Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone. *Nature* 281(5732):544–548.