Variable region design and cloning protocol outline

Ethan Holleman

July 1, 2021

1 Insert design overview

The overall goal of this series of experiments is to transcribe carefully controlled DNA sequences, referred to as inserts, in order to systematically observe the effects of these sequences on R-loop formation *in vitro*. Inserts are composed of two types of DNA components, the variable region and flanking infrastructure regions. The variable region contains the sequence we are interested in observing R-loop dynamics over. Infrastructure sequences are additional nucleotide blocks that allow for the insertion of variable regions into specific plasmid backbones in a modular fashion.
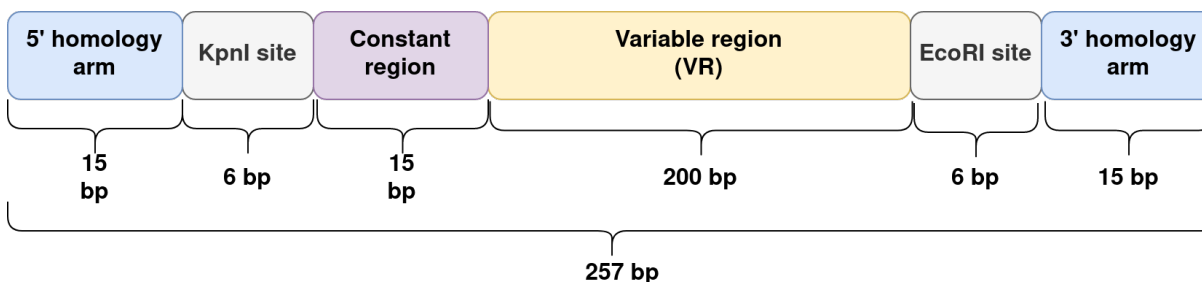

Figure 1: Diagram of a complement variable region insert showing the various regions each complete insert will contain, their relative order and lengths.

Each insert will be completely synthesized by an outside manufacturer and will not require assembly.

1.1 Variable regions

As is shown fig. 1 each insert will contain 1 variable region which will be designed to test the effects of different sequence properties on R-loop dynamics, namely GC / AT skew, GC content and G / C clustering. In total, 29 different variable regions with the properties described in table 1 will be generated using the variable region design workflow. The same pipeline will then be used to prepend or append infrastructure sequences to each variable region in order to produce a final fasta file containing the complete sequence of all inserts that can then be ordered for synthesis. The inserts will first be used to study the impact of sequence variation on R-loop initiation by insertion directly downstream of a promoter.

A subset of the variable regions will also be cloned further downstream of the same promoter species and oriented so the reverse complement of the sequence is transcribed in order to access their capacity for R-loop termination. The properties of the transcripts of these sequences are listed in table 2.

The cloning experiments required to produce these termination constructs will be undertaken after initiation experiments are completed. This will allow for the identification of a strong initiation sequence, and the determination of the average length of R-loops formed within plasmid constructs. This will determine the placement of termination regions relative to the promoter as it is critical that termination regions are far enough downstream as to not interfere with R-loop initiation but also not so distant as to prevent R-loops from actually terminating within the termination region.

Table 1: Properties of all syntheized inserts. The G clustering number refers to the number of G nucleotides in a given cluster with 0 being no clustering.

GC Skew	AT Skew	GC Content	G Clustering	Reverse Complement	Insert Number
0.2	0.0	0.4	0	1	0
0.1	0.0	0.3	0	0	1
0.6	0.0	0.6	0	0	2
0.4	0.0	0.3	0	1	3
0.2	0.0	0.3	0	1	4
0.0	0.0	0.3	0	1	5
0.0	0.0	0.6	0	0	6
0.6	0.0	0.5	0	0	7
0.1	0.0	0.6	0	0	8
0.4	0.0	0.6	0	0	9
0.2	0.0	0.6	0	0	10
0.0	0.0	0.5	0	1	11
0.6	0.0	0.4	0	0	12
0.0	0.0	0.4	0	1	13
0.1	0.0	0.5	0	0	14
0.6	0.0	0.3	0	0	15
0.4	0.0	0.5	0	1	16
0.2	0.0	0.5	0	1	17
0.1	0.0	0.4	0	0	18
0.4	0.0	0.4	0	1	19
0.4	0.0	0.6	2	1	20
0.4	0.2	0.6	2	1	21
0.4	0.4	0.6	2	1	22
0.4	0.0	0.6	3	1	23
0.4	0.2	0.6	3	1	24
0.4	0.4	0.6	3	1	25
0.4	0.0	0.6	4	1	26
0.4	0.2	0.6	4	1	27
0.4	0.4	0.6	4	1	28

Table 2

GC Skew	AT Skew	GC Content	C Clustering	Reverse Complement of Insert
-0.2	-0.0	0.4	0	0
-0.4	-0.0	0.3	0	3
-0.2	-0.0	0.3	0	4
-0.0	-0.0	0.3	0	5
-0.0	-0.0	0.5	0	11
-0.0	-0.0	0.4	0	13
-0.4	-0.0	0.5	0	16
-0.2	-0.0	0.5	0	17
-0.4	-0.0	0.4	0	19
-0.4	-0.0	0.6	2	20
-0.4	-0.2	0.6	2	21
-0.4	-0.4	0.6	2	22
-0.4	-0.0	0.6	3	23
-0.4	-0.2	0.6	3	24
-0.4	-0.4	0.6	3	25
-0.4	-0.0	0.6	4	26
-0.4	-0.2	0.6	4	27
-0.4	-0.4	0.6	4	28

Table 3

Total synthesized inserts	Total contructs
29	47

2 Assembly of DNA inserts
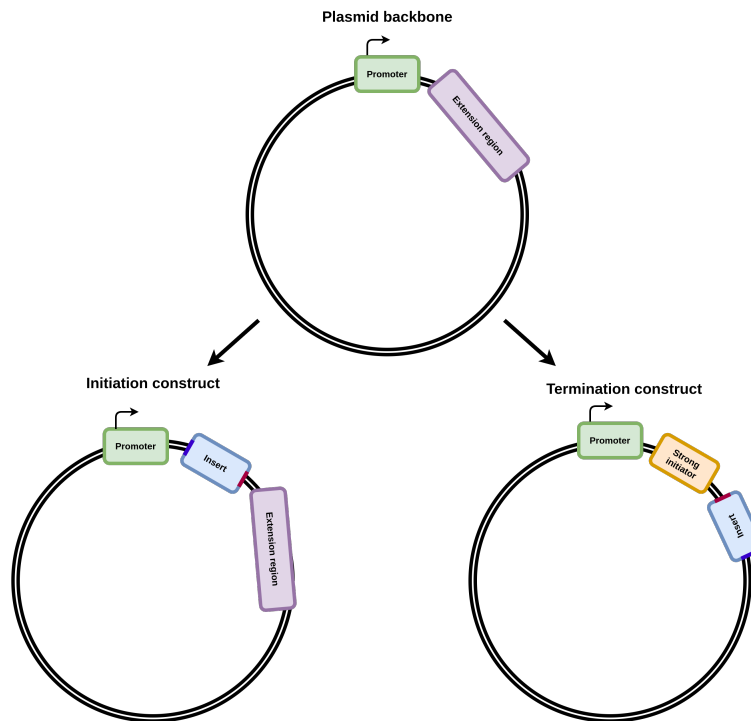
2.1 Basic Strategy

Figure 2: Diagram of basic cloning strategy. Inserts containing a variable region will be cloned into a backbone plasmid using either Gibson or restriction enzyme assembly to produce initiation or termination series constructs.

2.2 T7 initiation series constructs

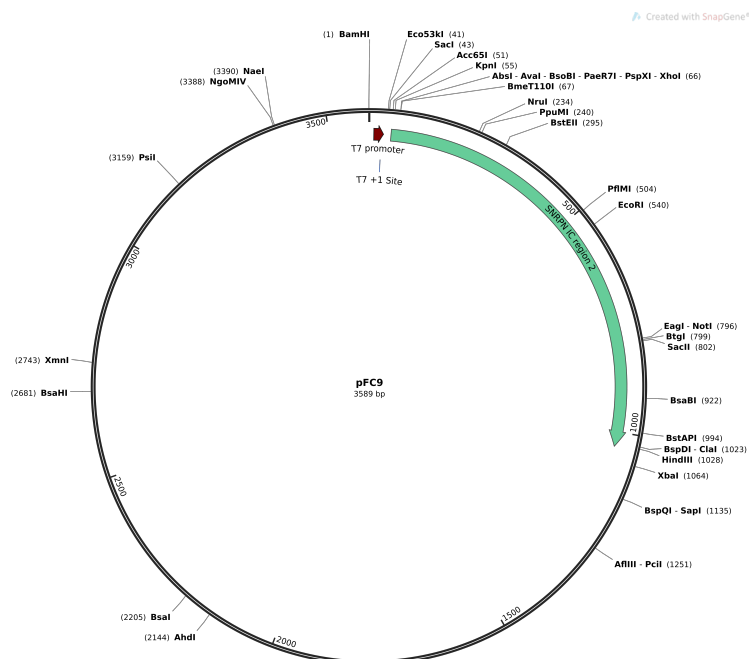The first series of constructs will be utilize pFC9 as the plasmid backbone.

Figure 3: Map of pFC9.

First, pFC9 will be cut using Eco53KI, producing blunt ends just downstream of the T7 promotor (fig 3). The complete ensemble of inserts will then be added in equal concentrations. Using the NEB Gibson assembly kit and protocol, the 5' and 3' homology arms will anneal to the T7 and extension region downstream of the Eco53KI cut site of pFC9 respectively producing a library of circular pFC9 plasmids with each insertion sequence in theoretically equivalent concentrations.
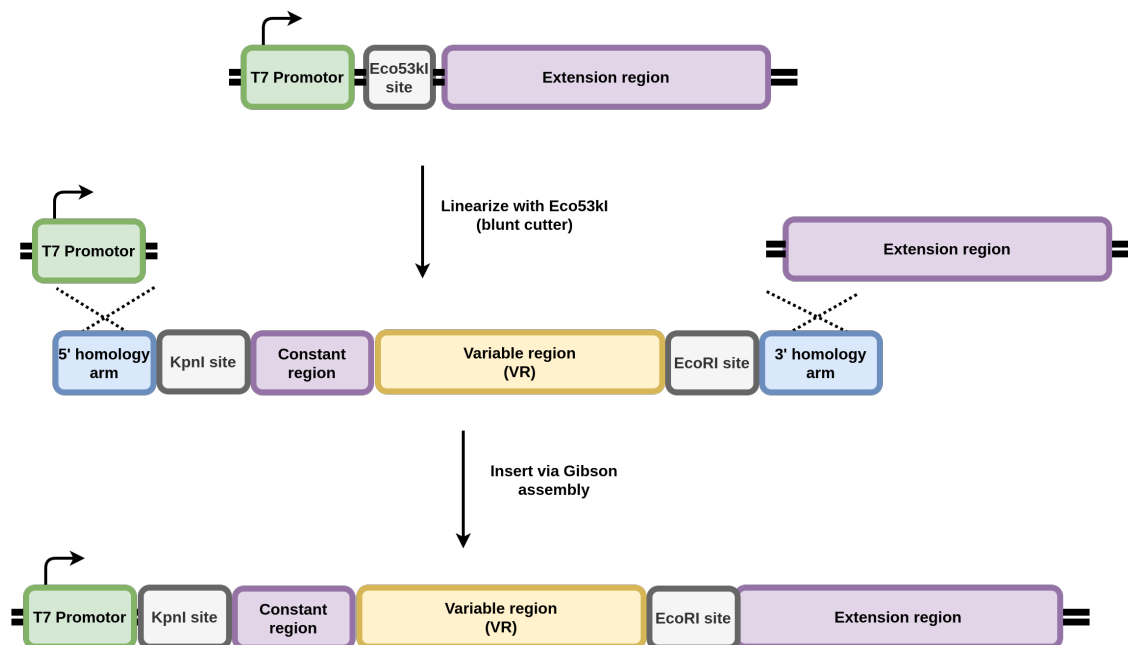

Figure 4: Diagram of pFC9 insertion series cloning strategy.

Additionally, primers will be designed for a subset of initiation regions and relative concentrations of each insert will be measured by qPCR. Having confirmed that inserts are present at relatively equal concentrations the library will be transcribed *in vitro* and prepared for single molecule R-loop footprinting using uniquely bar-coded PCR primers to facilitate the computational removal of PCR duplicates.

2.3 T7 termination series constructs

After the successful sequencing of the T7 initiation series, pFC8 will be utilized as the backbone for construction of the termination series library. First, the strongest and most consistent R-loop initiator identified from the T7 initiation series will be cloned into pFC9 without the presence of any other inserts using the methods described in section 2.2.
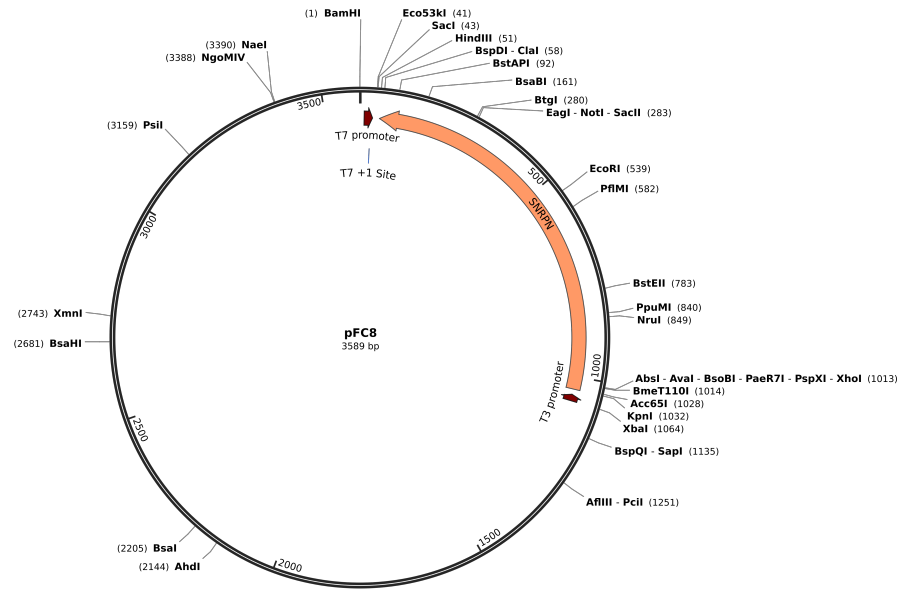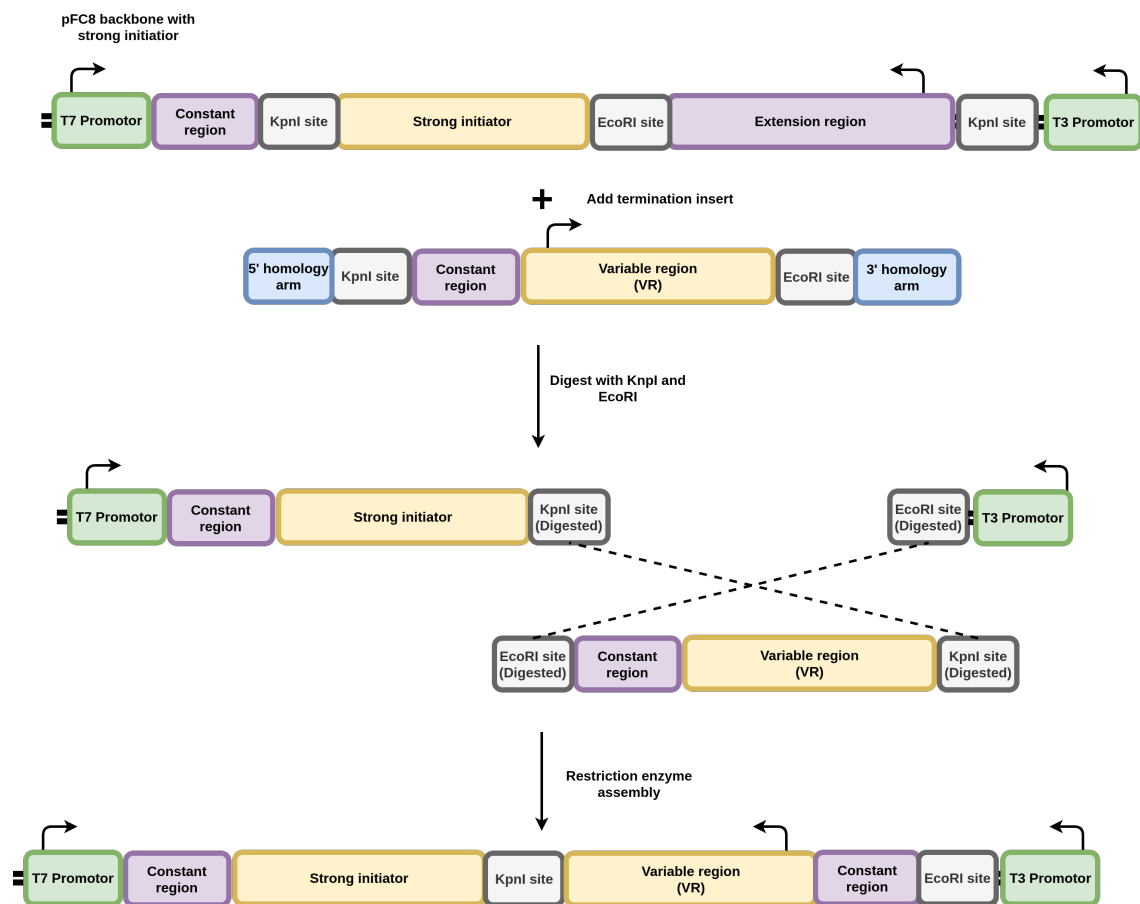
Figure 5: Map of pFC8.

Figure 6: Map of pFC9.

2.4 Tac initiation series constructs

BamHI (1)
BglII (37)
BbsI (64)

(3717) NaeI
(3715) NgoMIV

HindIII (379)
BspDI - ClaI (386)
PstI (416)
BstAPI (420)
BsaBI (489)

(3486) PsiI

T1T2 terminators

Btgl (608)
SacII (611)

(3070) XmnI

ApaLI

PflMI (910)

pFC8
3916 bp

SNRPN

Amp

BstEII (1111)

PpuMI (1168)
NruI (1177)

tac promoter

AbsI - AvaI - BsoBI - PaeR7I - PspXI - XhoI (1341)
BmeT110I (1342)
Acc65I (1356)
KpnI (1360)
XbaI (1391)

(2532) BsaI

(2471) AhdI

ApaLI

AflIII - PciI (1578)

Figure 7: Map of pFC8T$_1$T$_2$.
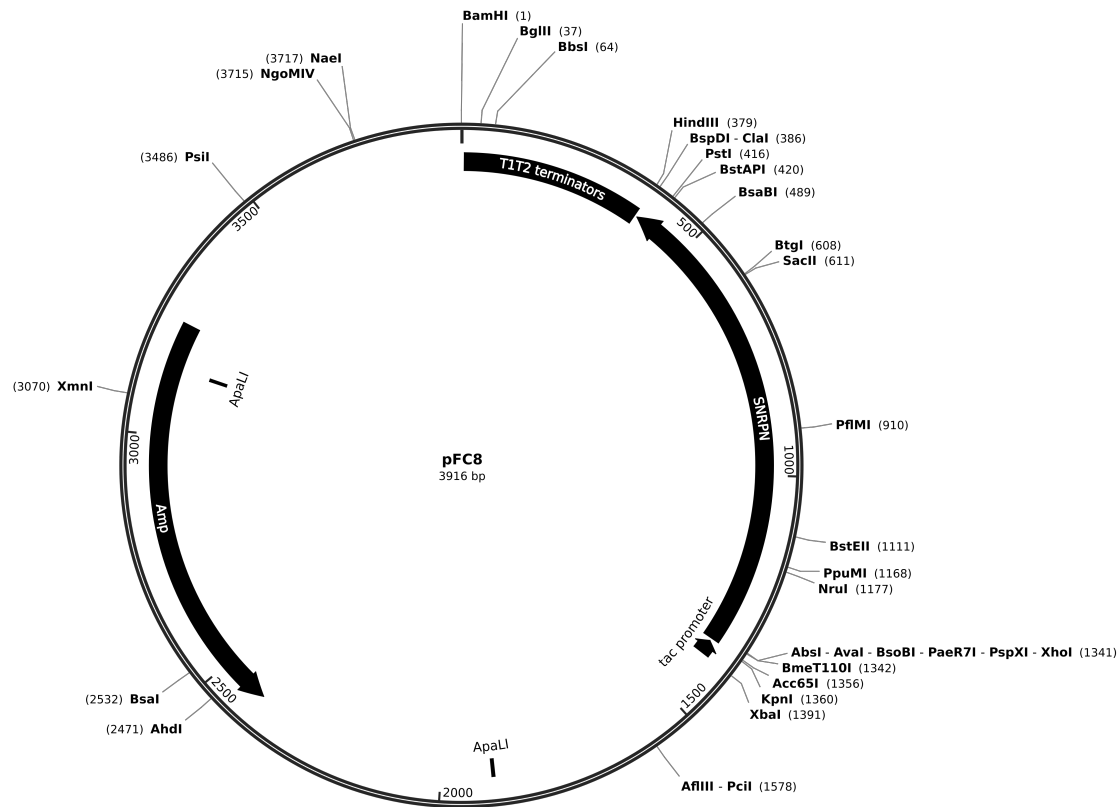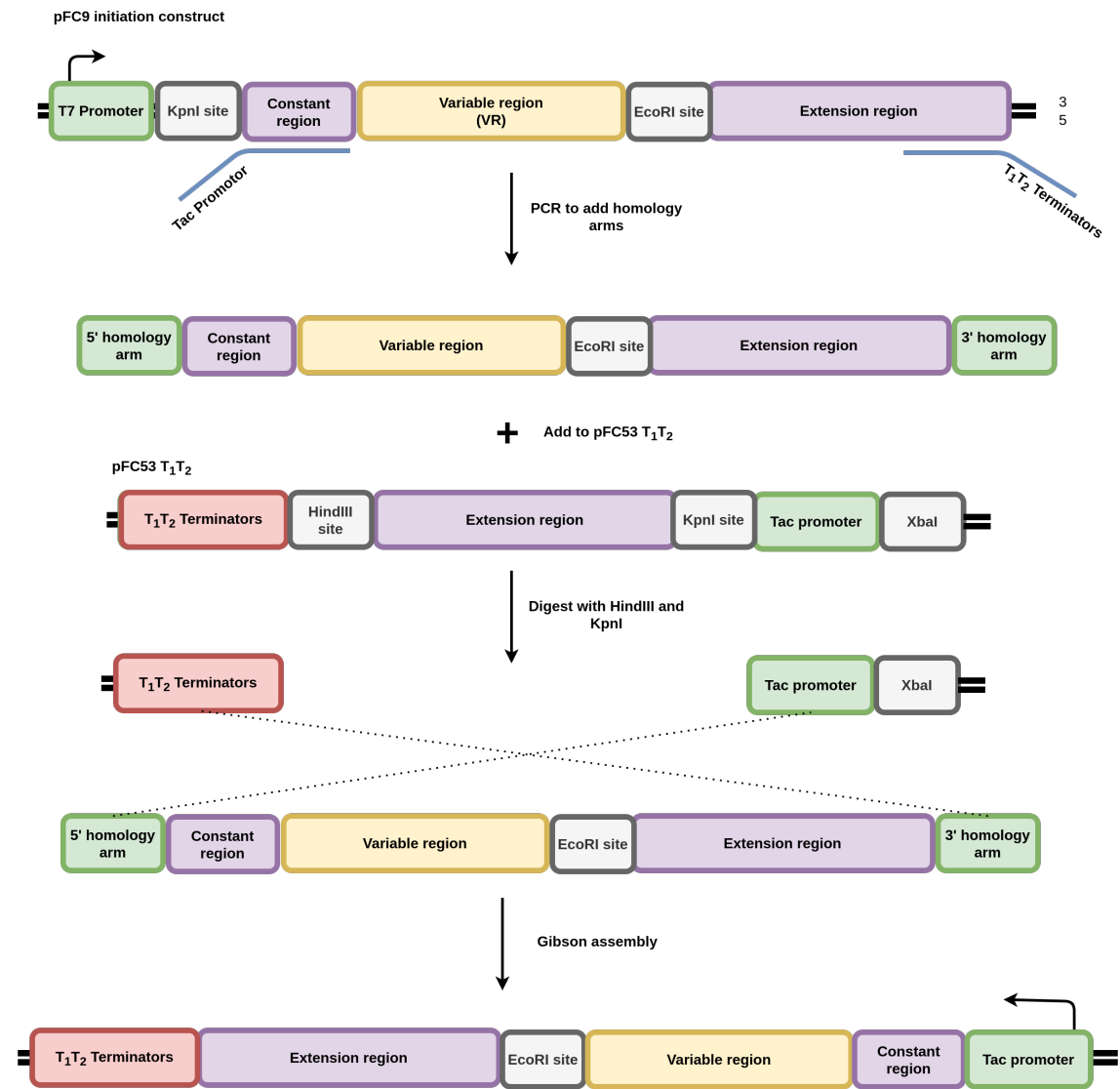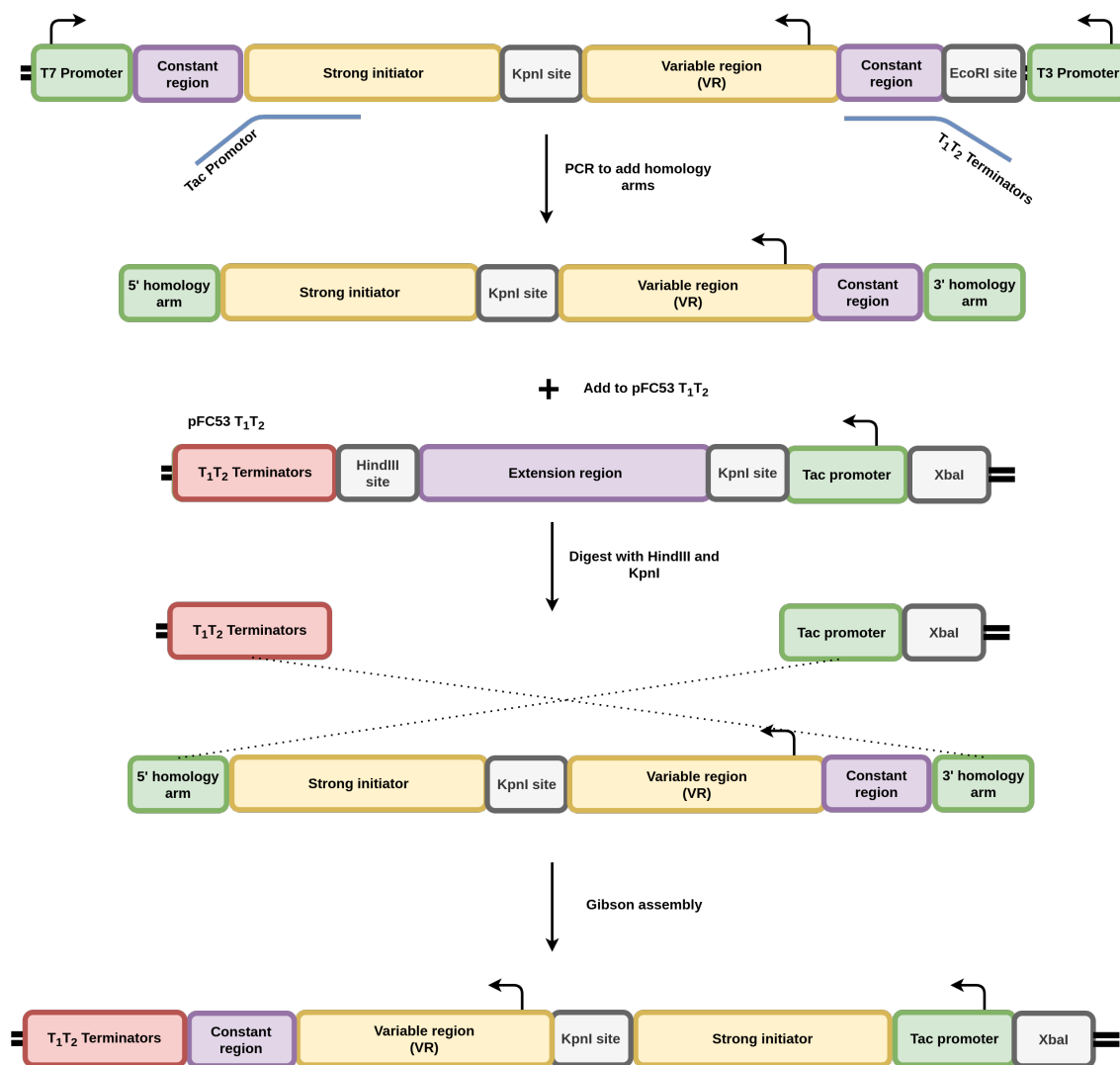
TExt here

Figure 8: Map of pFC9.

2.5 Tac termination series constructs

Figure 9: Map of pFC9.

References