Generation and Analysis of *malE* Mutations

The goals of this lab are for students to research the structure/function of a known protein, design a missense mutation they think will disrupt that protein's function and analyze the consequence of specific missense mutations. Due to the constraints imposed upon us as a result of the coronavirus pandemic, we won't actually be making new mutations this semester, but you will still engage in the processes of literature research, experimental design and data analysis.

The approach you will use to design your mutation is a site-directed mutagenesis approach. Site-directed mutagenesis allows researchers to create specific mutations in a sequence of interest and analyze its effect. This approach is referred to as "reverse genetics" because the researcher creates and assesses the effect of a particular mutation on gene function. If we were able to carry out all of the experimental procedures to introduce the mutation and analyze its consequences, you would employ a variety of molecular techniques including bacterial transformation, plasmid isolation, restriction digestion, electrophoresis and protein expression. Fortunately, you had the opportunity to implement most of these methodologies earlier in the semester via the Hox lab, so you already have experience with them.

The gene you will "mutate" for this exercise is a plasmid-borne copy of *malE*, which encodes a maltose binding protein, *MBP2**. The plasmid containing this gene is the expression vector pMAL-c2X (Figure 1); the *malE* gene encodes the protein tag in this expression vector and is located upstream of a multiple cloning site (polylinker). Induction of the strong bacterial promoter P_{tac} results in expression of the *malE* gene and any sequences cloned in frame into the polylinker. Resulting fusion proteins can be easily purified by passing the cell extract over an amylose resin column; fusion proteins would be retained in the column due to the affinity of *MBP2** for the amylose resin, while other cellular proteins would pass through the column.

The goal of this lab is to introduce a specific mutation into the *malE* gene and assess its effect on the maltose binding activity of *MBP2**. For this exercise, we will **NOT** be inserting additional sequences into the pMAL-c2x vector to generate a fusion gene/protein; rather, we will take advantage of the fact that *MBP2** can be expressed from this vector in order to study the function of its wild-type and mutant forms. The first step in this process is to determine what kind of mutation you want to introduce into *malE*. Each pair of students will design a missense mutation in the coding sequence of *malE*. The sequence of the *malE* gene is provided in Figure 2, with the start codon underlined and in bold, and the genetic code is provided in Figure 3. Details on how to design a mutagenic oligonucleotide will be provided in class and additional information can be found in the document "guidelines for mutagenic oligo design."

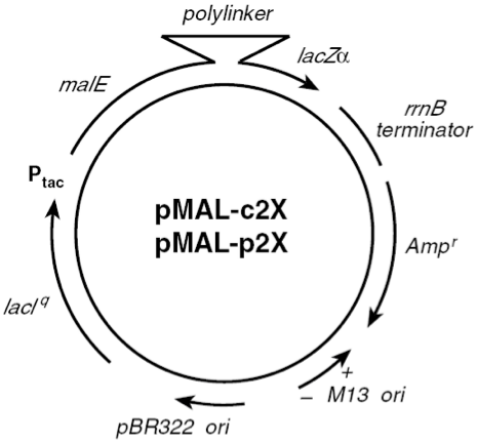

Figure 1: Schematic of the pMAL-c2X plasmid (www.neb.com).

sequence position	sequence feature
1406-1433	P_{tac} promoter
1528	*malE* start codon (ATG)
2587-2610	Sequence of *malE* primer
2629	Start polylinker, factor Xa site
2735	Start *lacZ*-α sequence (CACTGG)

pMAL™-c2X, -p2X polylinker:

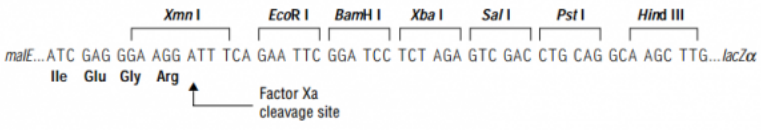

Figure 2. Sequence and translation of *malE* region of pMAL-c2X plasmid. The nucleotide sequence is provided, grouped into triplets, with the single letter amino acid indicated above each codon. Numbers on the left represent the position of the first nucleotide on each line.

```
       L  K  -  A  V  D  N  -  S  S  A  R  I  M  C  G  I  V  S  G
1393   CTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGA

       -  Q  F  H  T  G  N  S  Q  S  V  -  V  F  S  R  A  L  H  Q
1453   TAACAATTTCACACAGGAAACAGCCAGTCCGTTTAGGTGTTTTCACGAGCACTTCACCAA

       Q  G  P  -  H  M  K  I  E  E  G  K  L  V  I  W  I  N  G  D
1513   CAAGGACCATAGCATATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGAT

       K  G  Y  N  G  L  A  E  V  G  K  K  F  E  K  D  T  G  I  K
1573   AAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAA

       V  T  V  E  H  P  D  K  L  E  E  K  F  P  Q  V  A  A  T  G
1633   GTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGC

       D  G  P  D  I  I  F  W  A  H  D  R  F  G  G  Y  A  Q  S  G
1693   GATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGC

       L  L  A  E  I  T  P  D  K  A  F  Q  D  K  L  Y  P  F  T  W
1753   CTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGG

       D  A  V  R  Y  N  G  K  L  I  A  Y  P  I  A  V  E  A  L  S
1813   GATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCG

       L  I  Y  N  K  D  L  L  P  N  P  P  K  T  W  E  E  I  P  A
1873   CTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCG

       L  D  K  E  L  K  A  K  G  K  S  A  L  M  F  N  L  Q  E  P
1933   CTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCG

       Y  F  T  W  P  L  I  A  A  D  G  G  Y  A  F  K  Y  E  N  G
1993   TACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGC

       K  Y  D  I  K  D  V  G  V  D  N  A  G  A  K  A  G  L  T  F
2053   AAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTC

       L  V  D  L  I  K  N  K  H  M  N  A  D  T  D  Y  S  I  A  E
2113   CTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGAA

       A  A  F  N  K  G  E  T  A  M  T  I  N  G  P  W  A  W  S  N
2173   GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAAC

       I  D  T  S  K  V  N  Y  G  V  T  V  L  P  T  F  K  G  Q  P
2233   ATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCA

       S  K  P  F  V  G  V  L  S  A  G  I  N  A  A  S  P  N  K  E
2293   TCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAG
```

```
       L  A  K  E  F  L  E  N  Y  L  L  T  D  E  G  L  E  A  V  N
2353  CTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAAT

       K  D  K  P  L  G  A  V  A  L  K  S  Y  E  E  E  L  A  K  D
2413  AAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGAT

       P  R  I  A  A  T  M  E  N  A  Q  K  G  E  I  M  P  N  I  P
2473  CCACGTATTGCCGCCACTATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCG

       Q  M  S  A  F  W  Y  A  V  R  T  A  V  I  N  A  A  S  G  R
2533  CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGT

       Q  T  V  D  E  A  L  K  D  A  Q  T  N  S  S  S  N  N  N  N
2593  CAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAAC

       N  N  N  N  N  N  L  G  I  E  G  R  I  S  E  F  G  S  S  R
2653  AATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTTCAGAATTCGGATCCTCTAGA

       V  D  L  Q  A  S  L  A  L  A  V  V
2713  GTCGACCTGCAGGCAAGCTTGGCACTGGCCGTCGTTTT
```

Figure 3: Genetic code.

```
TTT phe F   TCT ser S   TAT tyr Y   TGT cys C
TTC phe F   TCC ser S   TAC tyr Y   TGC cys C
TTA leu L   TCA ser S   TAA OCH Z   TGA OPA Z
TTG leu L   TCG ser S   TAG AMB Z   TGG trp W

CTT leu L   CCT pro P   CAT his H   CGT arg R
CTC leu L   CCC pro P   CAC his H   CGC arg R
CTA leu L   CCA pro P   CAA gln Q   CGA arg R
CTG leu L   CCG pro P   CAG gln Q   CGG arg R

ATT ile I   ACT thr T   AAT asn N   AGT ser S
ATC ile I   ACC thr T   AAC asn N   AGC ser S
ATA ile I   ACA thr T   AAA lys K   AGA arg R
ATG met M   ACG thr T   AAG lys K   AGG arg R

GTT val V   GCT ala A   GAT asp D   GGT gly G
GTC val V   GCC ala A   GAC asp D   GGC gly G
GTA val V   GCA ala A   GAA glu E   GGA gly G
GTG val V   GCG ala A   GAG glu E   GGG gly G
```