2. Primer Design

2.1 Multiple Cloning Sites

All commonly-used expression vectors used in the Jia Lab contain the following multiple cloning site:

```
                  BamHI     EcoRI   SmaI   SalI  XhoI   NotI
        XXX XXX GGA TCC CCG AAT TCC CGG GTC GAC TCG AGC GGC CGC XXX XXX
        XXX XXX CCT AGG GGC TTA AGG GCC CAG CTG AGC TCG CCG GCG XXX XXX
        XXX XXX Gly Ser Pro Asn Ser Arg Val Asp Ser Ser Gly Arg XXX XXX
```

When designing your primers, you should add one of these restriction sites to the end of your primer in such as was as to *preserve the reading frame shown here at both the 5' and the 3' end of the gene!!!*. This is really, really important for insuring that you express a properly tagged protein.

Here are the actual multiple cloning sites of our expression plasmids, including tags, stop codons, and other 'business':

P1 (pGEX-4T-3):
```
                 BamHI     EcoRI  SmaI   SalI  XhoI   NotI              Stop Codons
CTGGTTCCGCGT GGA TCC CCG AAT TCC CGG GTC GAC TCG AGC GGC CGC ATC GTG ACT GAC TGA
GACCAAGGCGCA CCT AGG GGC TTA AGG GCC CAG CTG AGC TCG CCG GCG TAG CAC TGA CTG ACT
LeuValProArg*Gly Ser Pro Asn Ser Arg Val Asp Ser Ser Gly Arg Ile Val Thr Asp ***
 Thrombin Cleavage (removes N-GST tag)
```

P13 (pET16bMCS):
```
ATG GGC CATCATCATCATCATCATCATCATCATCAC AGC AGC GGC CAT ATC GAA GGT CGT CAT
TAC CCG GTAGTAGTAGTAGTAGTAGTAGTAGTAGTG TCG TCG CCG GTA TAG CTT CCA GCA GTA
MET Gly HisHisHisHisHisHisHisHisHisHis Ser Ser Gly His Ile Glu Gly Arg*His
            10xHis-Tag                                      Factor Xa

     BamHI     EcoRI  SmaI   SalI  XhoI   NotI
ATA GGA TCC CCG AAT TCC CGG GTC GAC TCG AGC GGC CGC ATC GTG ACT GAC TGA
TAT CCT AGG GGC TTA AGG GCC CAG CTG AGC TCG CCG GCG TAG CAC TGA CTG ACT
Ile Gly Ser Pro Asn Ser Arg Val Asp Ser Ser Gly Arg Ile Val Thr Asp ***
```

P22 (pET21bMCS):
```
      BamHI     EcoRI  SmaI   SalI  XhoI   NotI
CATATA GGA TCC CCG AAT TCC CGG GTC GAC TCG AGC GGC CGC GTC GAGCACCACCACCACCACCACTGA
GTATAT CCT AGG GGC TTA AGG GCC CAG CTG AGC TCG CCG GCG CAG CTCGTGGTGGTGGTGGTGGTGACT
       Gly Ser Pro Asn Ser Arg Val Asp Ser Ser Gly Arg Val GluHisHisHisHisHisHis***
                                                           6xHis-Tag
```

P259 (pGEXHis):
```
          BamHI   EcoRI.SmaI.SalI.XhoI.NotI                       NdeI
CTGGTTCCGCGT GGATCCCCGAATTCCCGGGTCGACTCGAGCGGCCGCATCGAAGGTCGT CATATGCACCACCACCACCACCACTGA
GACCAAGGCGCA CCTAGGGGGCTTAAGGGCCCAGCTGAGCTCGCCGGCGTAGCTTCCAGCA GTATACGTGGTGGTGGTGGTGGTGACT
LeuValProArg*GlySerProAsnSerArgValAspSerSerGlyArgIleGluGlyArg*HisMETHisHisHisHisHisHis***
 Thrombin Cleavage (removes N-GST tag)         Factor Xa              6xHis-Tag
```

P270 (pPIC6L4zeo):
```
    BamHI     EcoRI   SmaI    SalI  XhoI_  NotI
TCG GGA TCC CCG AAT TCC CGG GTC GAC TCG AGC GGC CGC CTC GAC CATCATCATCATCATCATTGA
AGC CCT AGG GGC TTA AGG GCC CAG CTG AGC TCG CCG GCG GAG CTG GTAGTAGTAGTAGTAGTAACT
    Gly Ser Pro Asn Ser Arg Val Asp Ser Ser Gly Arg Leu Asp HisHisHisHisHisHis***
                                                             6xHis-Tag
```

2.2 Primer Choice Considerations

- **Select the first (or last) 21 bases of your gene of interest as the basis of your primer.** This is the complementary region that will be used to perform PCR to amplify your gene of interest. Keep in mind that DNA is synthesized 5'→3', and so you're 5' forward primer should be complementary to the bottom strand of you gene of interest, and the 3' reverse primer should be complementary to the top strand of your gene of interest. *See the example at the end of this section.*

- **The 5' (forward) primer must contain a start codon (ATG),** and if you think you might want to express the protein-of-interest in yeast (*Pichia pastoris*), make sure this start codon is part of a Pichia Kozak consensus (which is (G/A)NN**ATG**G).

- **Consider the amino acids you will be adding to the protein of interest with this primer.** In order to shift the reading frame appropriately, you will often have to add a few bases to the restriction sites in your primer. Also, adding the Kozak consensus above will often require You should figure out which amino acids these extra sequences will encode in the protein-of-interest—if possible avoid introducing 'weird' amino acids with large functional groups (tryptophan, phenylalanine, etc) or unique chemical/structural properties (cysteine). I usually look at the multiple cloning site and pick an amino acid residue encoded within it as a guide to relatively structurally neutral amino acids.

- **Avoid chosing SmaI as a restriction site for cloning.** The Pichia expression vector (p270, or pPIC6L4zeo) contains two SmaI sites (one within the multiple cloning site, plus another within the Zeocin resistance cassette). *For you, this means that SmaI can not be used to clone your gene of interest* in to the Pichia vector.

- **Choose a restriction site that is not present in the gene you are cloning.** For obvious reasons….okay, it's so that you don't cut up your gene while doing the cloning.

- **If possible, choose BamHI (5' forward primer) and NotI (3' reverse primer) to include in your primer design**

- **Make sure your primer choice will lead to the gene of interest being in the correct reading frame after subcloning.** I know I wrote this already, but it's really crucial and easy to screw up.

- **Double and triple check the sequence of the primer you choose.** Follow the measure-twice-cut-once maxim, because you don't want to waste weeks and weeks of work on an incorrect or frame-shifted primer.

2.1 Using DNAMan

2.1.1 Finding the Open Reading Frame

We often obtain cDNA clones of our genes, and wish to subclone them via PCR in to expression vectors. These cDNA clones are very often the product of large-scale cDNA library preparations of mRNA, and therefore contain non-coding sequences that were originally present in the mRNA

1. Load DNAMan and open a text file that contains the cDNA sequence of your gene of interst. Alternative, copy the sequence of your gene of interest from GenBank, and then (in DNAMan) go to Edit/Enter sequence and Paste it in.

2. Highlight the sequence, then go to Sequence/Load Sequence/From Selection. This loads the selected sequence in to Channel 1 (NOTE: the number down the left side of the DNAMan screen are individual channels. Each of these can hold separate protein/DNA sequences. Click on the channel of interest to work with the sequence it contains).

3. Go to Sequence/Search For/Open Reading Frame. Accept all defaults and click 'Okay'.

4. You will be presented with a list of open reading frames, including the location of both start and stop codons. Evey gene will have many—in general, you will want the one that codes for the longest protein, but there are exceptions to this rule. You'll undoubtedly have some extra information about your protein (ie, length, etc) to allow you to pick the correct reading frame.

2.1.2 Restriction Analysis

In order to find the restriction sites present in your gene of interest, use DNAMan as follows:

1. Load DNAMan and open a text file that contains the cDNA sequence of your gene of interst. Alternative, copy the sequence of your gene of interest from GenBank, and then (in DNAMan) go to Edit/Enter sequence and Paste it in.

2. Highlight the sequence, then go to Sequence/Load Sequence/From Selection. This loads the selected sequence in to Channel 1 (NOTE: the number down the left side of the DNAMan screen are individual channels. Each of these can hold separate protein/DNA sequences. Click on the channel of interest to work with the sequence it contains).

3. Go to Restriction/Restriction Analysis. Here, you will be walked through a series of dialogues asking a variety of question about the source DNA and the restriction site you want to search for. I usually select only the six restriction sites of the multiple cloning site to avoid a confusingly large list of sites. This part of the analysis should be self explanatory.

4. In the end, you will be presented with a list of restriction sites present in the sequence, with a list of non-cutting enzymes at the bottom. Among the non-cutting enzymes, you should find two that are present in the multiple cloning site that you can use to subclone your gene of interest.

2.3 Example: BIN1

Here is the ORF sequence for a gene called *bin1*:

```
Coding:        5'-ATGGCAGAGATGGGCAGTAAAGGGGTGACG .other stuff. CCCGAGAACTTCACTGAGAGGGTCCCATGA-3'
Complementary: 3'-TACCGTCTCTACCCGTCATTTCCCCACTGC .other stuff. GGGCTCTTGAAGTGACTCTCCCAGGGTACT-5'
```

Note the start (ATG) and stop (TGA) codons, the coding strand (shown in the 5' to 3' orientation) and the complementary non-coding strand (shown in 3' to 5' orientation). The first thing I did was pump the sequence through a restriction site analysis using DNAMan, and found that this sequence contains neither *BamHI* nor *NotI*. I will therefore choose these two sites to put in my primers for subcloning.

To design the forward primer, I select the first 21 bases or the above sequence:

```
5'- ATG GCA GAG ATG GGC AGT AAA
```

Luckily, I don't need to add a start codon, since I'm cloning this gene from the very beginning and it already has one. If you were cloning a gene from the middle someplace, then you would have to add an ATG here. I then add the BamHI site (GGA TCC) to the 5' end of this sequence:

```
5'- gga tcc ATG GCA GAG ATG GGC AGT AAA
```

Note that, to keep this primer straight in my head, I've used lowercase letters to designate sequence that I've added/changed to the complementary sequence. This primer seems pretty good, but then I remember that I may want to express this protein in *Pichia*, so I make some additions and changes to give my final product a Kozak consensus (G/A)NN<u>ATG</u>G:

```
5'- gga tcc gtc ATG GCA GAG ATG GGC AGT AAA
```

This requires the addition of a 'GTC' between the BamHI site and the start codon. Sometimes, you aren't lucky enough to have 'G' immediately downstream of the ATG start codon, and so you have to introduce a mutation here to ensure you have that ATGG start sequence. Finally, I double check which amino acid residues I'll be adding/changes with this primer:

```
5'- gga tcc gtc ATG GCA GAG ATG GGC AGT AAA
5'- Gly Ser Val Met etc....................
```

Nothing weird introduced, so this primer is good to go. I'll do one last thing, though—I like to add a couple of bases to the 5' end of the primer to preserve my restriction site during PCR and subsequent TA cloning (errors can be introduced right at the end during PCR, and I want to avoid the damage this can do to my restriction sites). So the final forward primer looks like this:

```
BIN1-F:  5'- gc gga tcc gtc ATG GCA GAG ATG GGC AGT AAA
```

For the reverse primer, I start in a similar way by selecting the last 21 bases in this sequence. NOTE that I am EXCLUDING the stop codon—this is really, really important. You will also see that I'm picking a sequence that is complementary to the coding strand (which is the opposite of the forward primer), and that I am writing it in the 5' to 3' orientation (required when we're ordering the primer).

```
5'- TGG GAC CCT CTC AGT GAA GTT
```

As above, I will add a NotI (GCG GCC GC) restriction site. Notice that when I write out these sequences, I write them in the triplets that correspond to the proper reading frame (from the multiple cloning sites described above).

```
5'- gcg gcc gc TGG GAC CCT CTC AGT GAA GTT
```

Right away, I see that if I use this as my primer, I will shift the reading frame by one base (between the restriction site and the complementary sequence). So I will add a base to shift the reading frame back in to place:

```
5'- gcg gcc gct TGG GAC CCT CTC AGT GAA GTT
```

And then I check the amino acids that will be introduced/modified by this primer. (To do this, keep in mind that this primer corresponds to the non-coding strand, so you have to make your brain read the *complement* of this primer from *right-to-left* to correctly guess the amino acids that are encoded:

```
5'- gcg gcc gct TGG GAC CCT CTC AGT GAA GTT
    cgc cgg cga acc ctg ........ect. (READ FROM RIGHT TO LEFT)
    Arg Gly Ser Pro Val ........ect.
```

Everything looks good. The final primer (with an added couple of bases to the 5' end) looks like this:

BIN1-R: `5'- ga gcg gcc gct TGG GAC CCT CTC AGT GAA GTT`

So now I am ready to double-check these primers and copy/paste these sequences in to a primer order form:

BIN1-F: `5'- gc gga tcc gtc ATG GCA GAG ATG GGC AGT AAA`

BIN1-R: `5'- ga gcg gcc gct TGG GAC CCT CTC AGT GAA GTT`

3. PCR

3.1.1 Polymerases

- **For standard PCR,** we use Taq (from New England Biolabs, available in the 3rd floor BioBar from Julie Clarke) as our polymerase. This enzyme comes with it's own 10X buffer.

- **For cloning by PCR**, we use a high-fidelity polymerase called Expand HF (Roche, available in the Roche freezer in the Microbiology Store Room on the 8th floor). This enzyme (a mixture of Taq and Pwo polymerase) also come with its own buffer. This is a proof-reading PCR mix, which is nice when you don't want mistakes in your PCR product. Unlike many proofreading mixes, Expan is very 'Taq-like' in it's properties—for example, it leaves A-overhangs which we can use for TA cloning.

3.1.2 Diluting primers

Most of the primers used by our are ordered using the 25 nmole process from Integrated DNA technologies. When they arrive, they are labeled with the amount of oligonucleotide (which is somewhere around 25 nmole). I dilute these to **100 μM** in ddH$_2$O:

$$
\begin{aligned}
\text{Amount of ddH}_2\text{O to add} &= \frac{\text{amount of primer (nmoles)}}{100 \frac{\text{micromole}}{\text{L}}} \\
&= \frac{25.6 \text{ nmole}}{100 \frac{\text{micromole}}{\text{L}}} \\
&= \frac{25.6 \text{ nmole}}{100 \frac{\text{nmole}}{\text{mL}}} \\
&= 0.256 \text{ mL}
\end{aligned}
$$

In this example, I would add 0.256 ml of sterile ddH$_2$O to the lyophilized oligo tube, and then immediate make a 1:10 dilution of it (10 microliters of primer in 90 microliters of ddH$_2$O) in a microtube so that I have a **10 μM** primer solution

3.1.3 Before Starting

Here is the grocery list of things you'll need to do your PCR:
- **Taq or Expand HF polymerase** (keep in the freezer until the last minute)
- **10X polymerase buffer with MgCl$_2$** (supplied with Taq and Expand HF)
- **10 mM dNTP mix**. 100 mM solutions of dTTP, dCTP, dATP, and dGTP are available from the New England Biolabs Biobar. Mix and aliquot a solution that contains each of these. (ie, 10 microliters of each 100 mM solution in 60 microliters of sterile ddH$_2$O. Avoid frequent freeze/thaw cycles of this mix—if your PCR isn't working, its often because this solution has gone 'off'.
- **10 μM forward primer**
- **10 μM reverse primer**
- **Sterile ddH$_2$O** (it's a good idea to keep a separate PCR-only bottle of ddH$_2$O, to avoid contamination).
- **Template DNA.** If the template is a plasmid, a 1:10 dilution of a standard miniprep works well. If the template is genomic DNA, try a 250 ng/microliter solution of DNA. PCR can be finicky, and you may have to play with these amounts depending on the situation.

3.2 Programming Eppendorf Mastercycler

For now, talk to Jim and he'll show you how to work this machine

3.3   PCR Reaction Mix

Here is the standard 1X PCR reaction mix:

Amount	Reagent	Final concentration
5 μL	10x polymerase buffer with MgCl$_2$	
1 μL	10 mM dNTPs	200 μM
1.5 μL	10 μM Forward primer	300 nM
1.5 uL	10 μM reverse primer	300 nM
1 to 5 μL	Template DNA (this amount can vary) • 1 μL of a 1:10 diluted plasmid • 1 μL of a 250 ng/μL genomic DNA solution • 5 μL of an E. coli colony resuspended in 10 μL ddH$_2$O • Etc…	250 ng genomic DNA 15 ng plasmid DNA
0.5 μL	Polymerase (Taq or Expand HF)	1.75 - 2.5 Units (depends on polymerase)
35.5 – 39.5 μL	ddH$_2$O	
50 μL	**TOTAL**	

You can scale this recipe up any way you like for doing multiple reactions. Just multiply all the amounts by the number of reactions you want to do plus one. You can also leave out things like the template and primers, then make 46 μL aliquots of a 'Master Mix' to which you add template and primer pairs (ie, for doing multiple PCR reactions of different genes the same time).

The instruction sheets that come with the polymerases contain many, many more details. Have a look at these when you get a chance, especially if you're trying to troubleshoot a PCR reaction that just isn't working.

Notes:

- Keep everything on ice when setting up your reactions

- Be very, very careful about cross-contaminating primers, templates, etc. A mistake here can give you really odd results.

- We use thin-walled 0.2 mL PCR tubes

- Ideally, include three controls: one where you leave out the forward primer, one where you leave out reverse primer other, and a third where you leave out only the template. None of these should show a PCR product, and if they do, you'll know that there is some kind of contamination or mispriming occurring.

- Some people like to keep a separate set of tips and reagents for PCR only. This is to avoid cross-contamination, and it's probably a good idea (although I don't do it myself!)

3.4   Conducting the PCR

Try the following times/temperature. There many variables that can adjusted here, but this standard cycle usually works:

1. 94°C for 3 minutes
2. 94°C for 15 seconds (denaturing)
3. 52°C for 30 seconds (annealing)
4. 72°C for 1 minute (extension) – Actually, you should program the machine to extend for 1 minute per kilobase. Ie, if the PCR product is about 1500bp, extend for 1 minute 30 seconds. Also, **when using Expand HF, extend at 68°C instead of 72°C.**
5.   Repeat from step 2 for 30 cycles
6. 72°C for 7 minutes
7.   Hold at 4^{0}

Notes:
- After conducting the PCR reaction, take 10 µL out of the reaction mix, add loading dye, and run it on an agarose gel to check the success of your reaction.

- We have a gradient thermocycler that allows you to set up multiple annealing temperatures within the same PCR cycle. This is useful for troubleshooting your PCR reaction in cases where you get multiple anomalous products.

- Both Taq and Expand HF polymerase leaves 3' A-overhangs at the very end of the PCR product. These are used for TA cloning (see next section), but they can be lost/degraded with time so be sure to do your TA cloning quickly. If you have a PCR product that has been hanging around in the freezer for a long time, add 0.5 µL Taq and put the mix back in to the thermocycler for 10 minutes at 72°C so that the A-overhangs are fresh for TA cloning.

TA Cloning

Vectors and Kits

Following a successful PCR reaction for the purpose of cloning (using Expand HF polymerase), we clone the PCR product in to a special cloning plasmid. By transferring the PCR product in to this plasmid, we are able to grow up as much of the clone as possible before transferring it in to the various expression vectors. *This makes your life easier, particularly if you are trying to clone your gene in to several different vectors at once.*

The method we use takes advantage of the fact that Taq (and 'Taq-like') polymerases leave 3' A-overhangs at the very ends of the PCR product. These miniature 'sticky ends' can be efficiently ligated in to a special vector that has been engineered to have 3'-T-overhangs. This is called 'TA cloning'.

For TA cloning, we use the Invitrogen TOPO-TA cloning kit (catalogue #K4550-01) in a version that includes the pCR2.1 vector and TOP10F' chemically competent cells. This kit refines the TA cloning method by covalently linking an enzyme (Topoisomerase I) to the T-overhangs of the pCR2.1 vector. The enzyme catalyzed the ligation of the PCR product in to the vector rapidly and efficiently. (See the TOPO-TA cloning manual for details).

Protocol

Follow the protocol as described in the instruction that come with this kit. I make the following modification to these instructions:

- All 6 µL of the TA cloning reaction mix is used to transform the TOP10F' cells.
- After performing the recovery after the heat shock transformation, I plate a full 100 µL of cells.
- I routinely use Kanamycin-containing plates for selection after transformation.

Blue/White Screening

You will notice after reading the TOPO-TA kit instructions that the pCR2.1 vector is capable of blue/white screening. Without getting in to the particulars of how it works, suffice it to say that this vector allows you to select positive clones simply by looking for white (As oppose to blue) colony on the plate after transformation and selection.

To use blue/white screening, take your Kanamycin containing plates and spread the following on them:

> **80 µL of X-Gal** (20 mg/ml solution in dimethylformamide (DMF))
> **40 µL of IPTG** (100 mM solution in water, then sterile filtered)

Then, warm/dry the plates to 37^oC as you normally would for a plating. After plating the transformation, colonies containing the recombinant clone will be white, whereas empty plasmid clones will be blue. (Note that putting the plate in the fridge for a few hours the day after plating intensifies the sometimes faint blue colour).