Primer Design Assignment

8/24

Eric Hao

BjergSun Co.

Primer Design Assignment

Goal:

Create a program that can expedite the process of creating and optimizing primers for PCR

Table of Contents

Background:

PCR (Polymerase Chain Reaction) is a method widely used by biologists to duplicate a segment of DNA by nearly one million-fold. This is very useful for many different applications, including analyzing the effect of gene therapy and creating a higher concentration of DNA for expression in target cells.

Typically, PCR relies upon a mix of either 25 or 50 uL of ingredients. These ingredients include:

1. Water
2. A type of polymerase (*Phusion is a widely used type for DNA amplification*)
3. A set of primers (*forward and reverse; primers are directional!*)
4. Template DNA
5. dNTP's (*dATP, dGTP, dTTP, dGTP; these are the nucleotides that help make the DNA longer*)
6. Some other stuff we don't have to worry about ☺

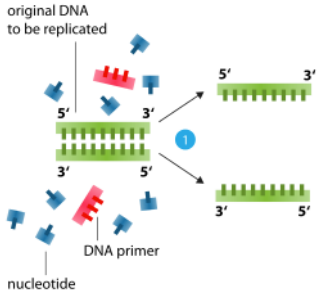
Some terms:

- A **polymerase** is an enzyme (a protein) that **synthesizes** long chains of nucleic acids (in our case, DNA).
- A **primer** is a short segment of **DNA**, around 40-60 base pairs long (a base pair refers to a single nucleotide base on one strand and its complement on another strand; for example, an A paired with a T).

Polymerase Chain Reaction:

PCR is comprised of several steps, characterized by temperature.

1. **Initialization**: 98°C to activate the polymerase
2. **Denaturation**: 98°C to cause double-stranded DNA to separate into single strands

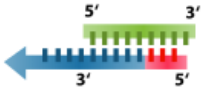


DNA denatures and separates into 2 strands of DNA.

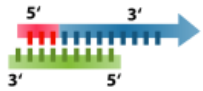
3. **Annealing**: 50-60°C to allow the primers to attach to single-stranded DNA. This temperature is **VERY important to primer design**! The primers are specific to a certain piece of DNA you want to amplify. More on this later.

*Primers (red) attach to the single-stranded DNA. Note that the primes attach at the **3′ end***

4. **Elongation**: 75-80°C to allow the polymerase to extend the sequence, starting at the primer and moving along as it adds more bases. ALSO a temperature **used in primer design**!

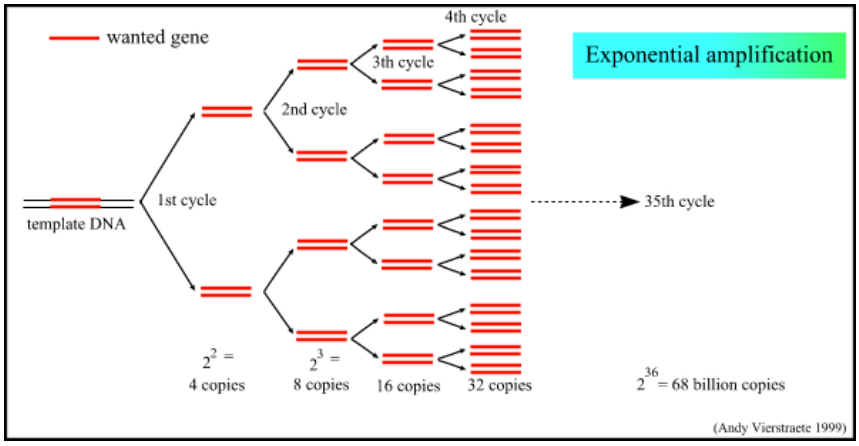


*Primers (red) elongate via the polymerase (not pictured). Note the **directionality**! Elongation continues along the 3′ end of the PRIMER (Along the 5′ end of DNA). Usually DNA is **WRITTEN from 5′ to 3′***



(Note: Steps 3-4 are often repeated for many cycles before step 5)

5. **Final elongation**: 72°C to allow the polymerase to finish up what it's doing
6. **Final hold**: 4°C to keep the DNA intact



This image shows the cycles of steps 3 & 4 repeated over many cycles to amplify the DNA

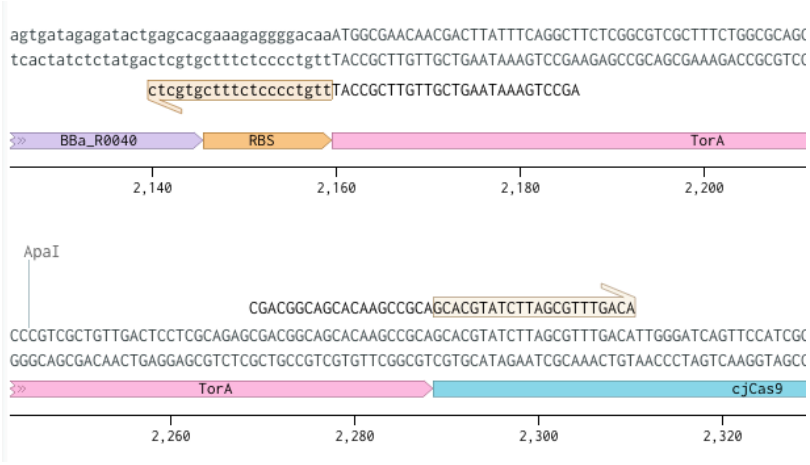
PRIMER DIRECTIONALITY:

This is very hard to grasp but also VERY important. It even gets its own section! If you don't understand this part, I will be explaining more in the "what to do" section later on, and hopefully you will be able to execute without understanding. Shrug.

As you saw above, primers have DIRECTIONALITY. Shown below is a visualization of what primers look like in a program I used called "Benchling." It is evil. I am trying to replace it with the program you will (hopefully be able to) make.

The primers are shown in ORANGE, attached to a **TAIL**: ctcgtgctttctcccctgttTACCGCTTGTTGCTGAATAAAGTCCGA

Let's call the orange arrowhead the **"HEAD"** and the non-colored part the **"TAIL"**.



SENSE and ANTISENSE strands

See how there are 2 stands of DNA?

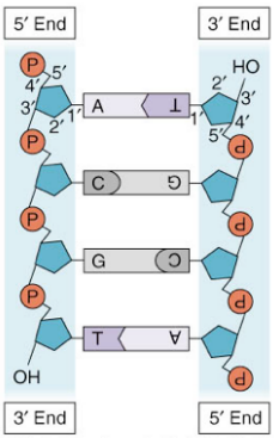
We will call the top strand the "sense" strand. This strand runs from 5' → 3'.

We will call the bottom strand of DNA the "antisense" strand. This strand runs from 3' → 5'. It is a complement of the top strand.

```
5' ...A T G G C C T G G A C T T C A... 3'   Sense strand of DNA
3' ...T A C C G G A C C T G A A G T... 5'   Antisense strand of DNA
```

Important!!! **A – T, G – C**.

This directionality is based on the chemistry of DNA. Wanna know more? Curiosity killed the cat:



(I don't mind if u don't understand this)

Forward primer

The arrow pointing to the **right** is called the **FORWARD** primer. This primer points in the direction of 3′:

5′ ← CGACGGCAGCACAAGCCGCAGCACGTATCTTAGCGTTTGACA → 3′

As you can see, the **FORWARD** primer is the SAME as the **top strand** of DNA:

CGACGGCAGCACAAGCCGCAGCACGTATCTTAGCGTTTGACA
CGACGGCAGCACAAGCCGCAGCACGTATCTTAGCGTTTGACAT

Reverse primer

The arrow pointing to the **left** is called the **REVERSE** primer. Does this primer point in the direction of 5′? NO!!!!! It points in the direction of 3′. This primer is the SAME as the **complementary (antisense) strand** of DNA (meaning, it is the **complement** of the **sense** strand):

actcgtgctttctcccctgttTACCGCTTGTTGCTGAATAAAGTCCGAA
ctcgtgctttctcccctgttTACCGCTTGTTGCTGAATAAAGTCCGA

BUT because it is a **CONVENTION** for us to read DNA from 5′ to 3′, when we display our **REVERSE** primers in **writing**, they must be written **IN REVERSE**.
A REVERSE primer shown as **3′ – c t c g t g – 5′** will instead, be written as **5′ – g t g c t c – 3′**.

Reverse complement

Another way people describe REVERSE primers is: **"REVERSE COMPLEMENT"**. Literally, you take the **reverse** of the **complement** of the **SENSE** strand. Follow?

END OF BACKGROUND:
Amazing! Now you have learned some of the main biological background necessary for this project. Some things that are important to keep in mind: **ANNEALING** temperature, primer **DIRECTIONALITY**, **FORWARD** primer, **REVERSE** primer, **REVERSE COMPLEMENT**.

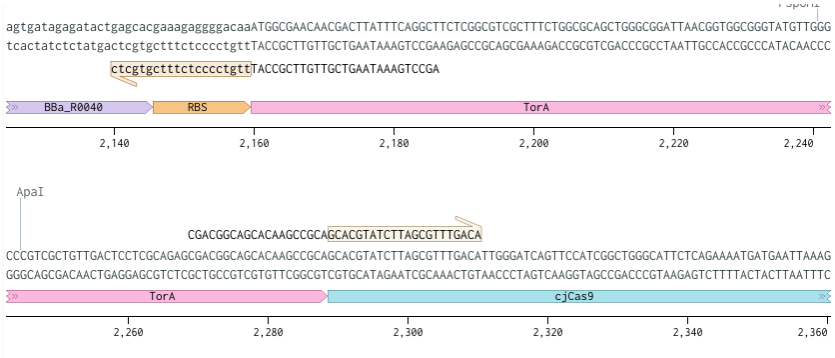
On to what you are supposed to do for me….

Primer Design Assignment

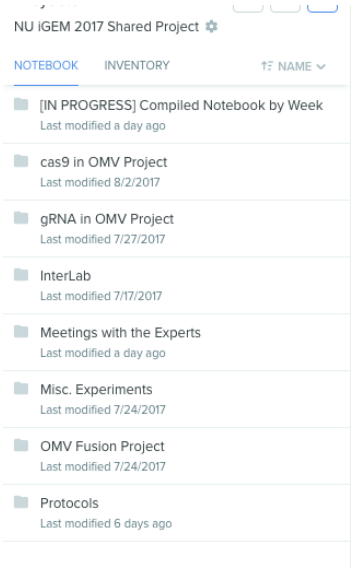
What do we have to use already?

Currently, there are two main programs used for primer design. They both work, but there is still a large amount of manual work required. One is Benchling, and the other is called ApE.

Benchling is very great because it has very clear annotations and can be very easy to organized with its filing system.



Annotations look nice!
Primers are easy to find!
There are even numbers telling you how many base pairs you are along the DNA!

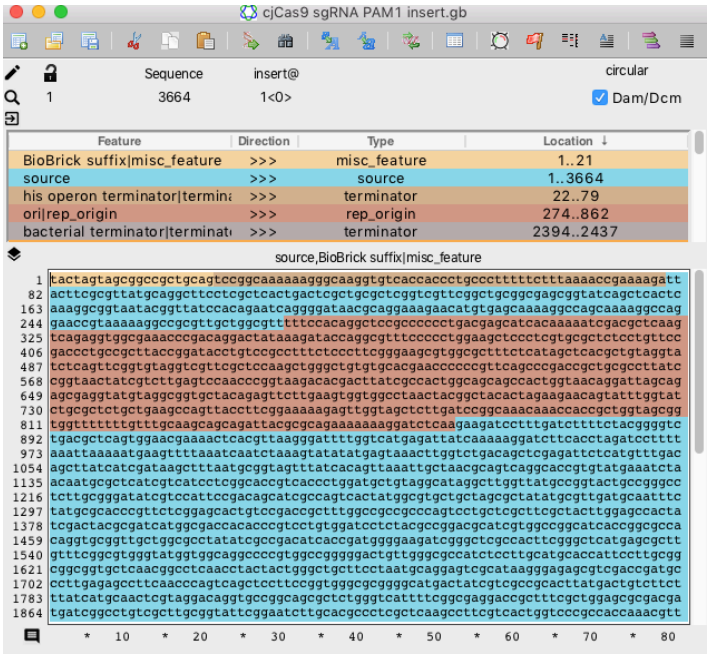


Folders are great! Very organized! Nice.

The greatest downfall is Benchling's primer design ability. The way it designs primers is very stupid and clumsy. Don't worry about how bad it is, but the fact that almost no real scientist uses the primer design functionality should say something about how bad it is.

I would not mind using your program to design primers and then adding them into my Benchling DNA map. That part is very easy. Front-end stuff? Don't worry about it.

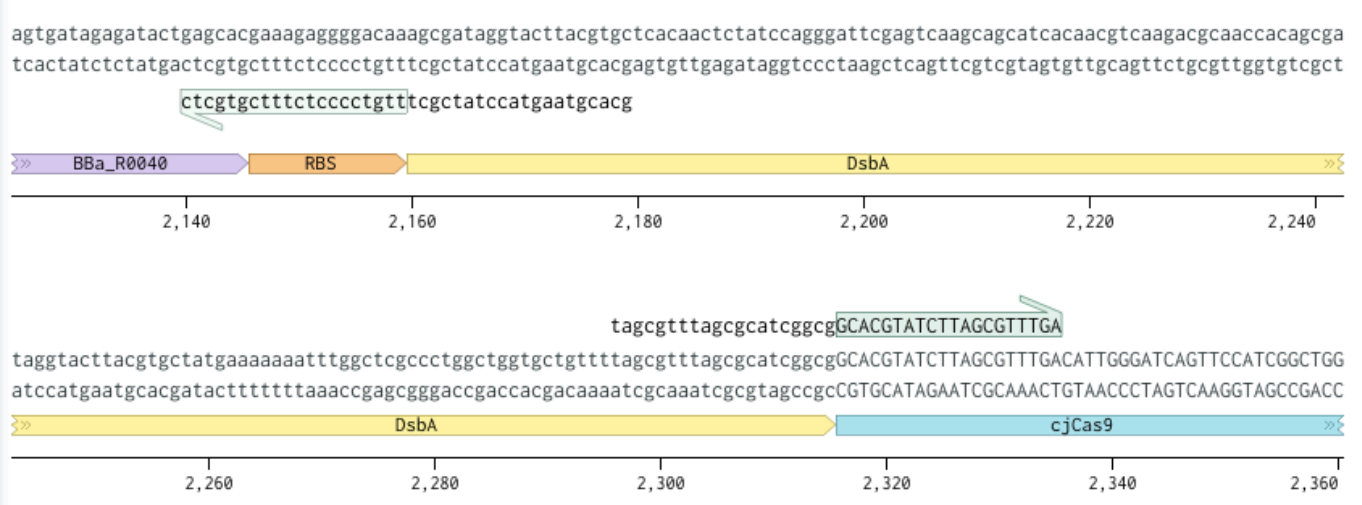
ApE is also nice, but I haven't used it that much. Design-wise, it is much less sophisticated than Benchling. The interface is literally a nightmare to look at. It also doesn't show both strands of DNA (sense and antisense), but Benchling does. I'm not too sure why people use this. Habit? It seems like the program can do a lot, but it is very anti-user. Primer design is very annoying on this.



My process of designing primers:

Part of the project needs a sequenced called DsbA to be amplified in PCR. This means I need a set of primers, one FWD and one REV that can attach to the sequence. Through PCR, only the segment between these two primers will become amplified.

Part of the project also needs the rest of the DNA that is NOT DsbA. As shown below, the FWD and REV primers point away from DsbA, and will amplify all the DNA around DsbA.



The reason why we need tails is…. Complicated. It's called "Gibson assembly" if you are interested, but how that works is of little relevance to the project I want you to do. Basically, we use these tails as "sticky ends", and if two pieces both have sticky ends that complement each other, they will attach and make a full circle of DNA!!! (remember, we're working with plasmids, or circular DNA).

Secret mission: If your program can allow me to make sticky ends that are NOT in the same plasmid (in the above pictures, the sticky ends are in the SAME plasmid), that would be REALLY FANTASTIC and I would love you forever.

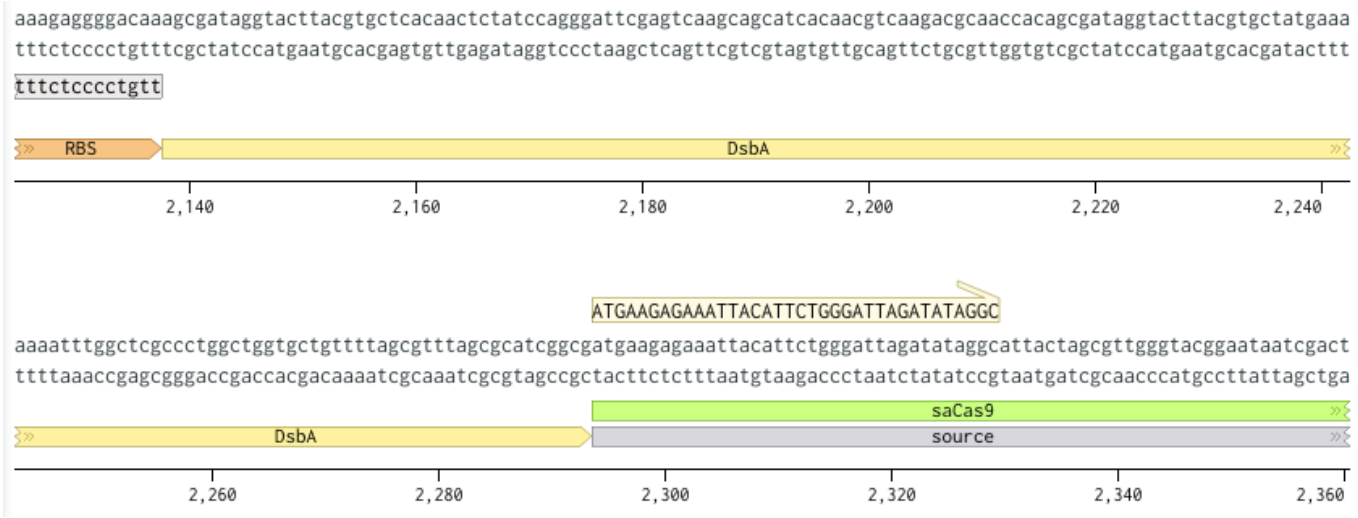
The steps of picking primers are:

1. Choose what part I want to amplify (an example will follow this section)
2. Create FWD and REV primers, starting with 20 bp HEAD & 20 bp TAIL (total of 40 bp).
3. Check the temperatures:
    a. Paste sequence of HEAD into idtdna.com/calc/analyzer : the temperature reported is going to be "T anneal"
    b. Paste sequence of HEAD + TAIL into idtdna.com/calc/analyzer : the temperature reported is going to be "T whole"
    c. *T anneal* of FWD must be 3°C around *T anneal* of REV; likewise, *T whole* of FWD must be 3°C around *T whole* of REV
    d. Temperatures MUST NOT exceed 72°C
    e. Generally, *T anneal* is around 50-65°C, and *T whole* is around 65-72°C, with exception.
4. If the *T anneal* temperatures greater than 3°C apart:
    a. Take the sequence with the lower *T anneal* and elongate the size of the HEAD (elongate on the end farthest from the tail, not the end adjacent to the tail)
    b. Elongate one base-pair at a time. We want the shortest primer possible within the acceptable temperature constraints.
    c. Total primer length **MUST NOT exceed 60 bp**
5. If the *T whole* temperatures are greater than 3°C apart:
    a. Take the sequence with the lower *T whole* and elongate the size if the TAIL (elongate on the end farthest from the head, not the end adjacent to the head)
    b. Elongate one base-pair at a time.
    c. Total primer length **MUST NOT exceed 60 bp**
6. Keep track of the "GC content" of the ENTIRE PIECE! Should be between **40%-60%!!**
    a. GC content is how many G's and C's are in this strand. Important, because G's and C's stick together rly strongly, so if there are too many G's and C's, the piece of DNA will want to stick to each other.

Remember, the goal of your project is to be able to automate as much of the process above as possible, outside of allowing me to choose what pieces I want to amplify.

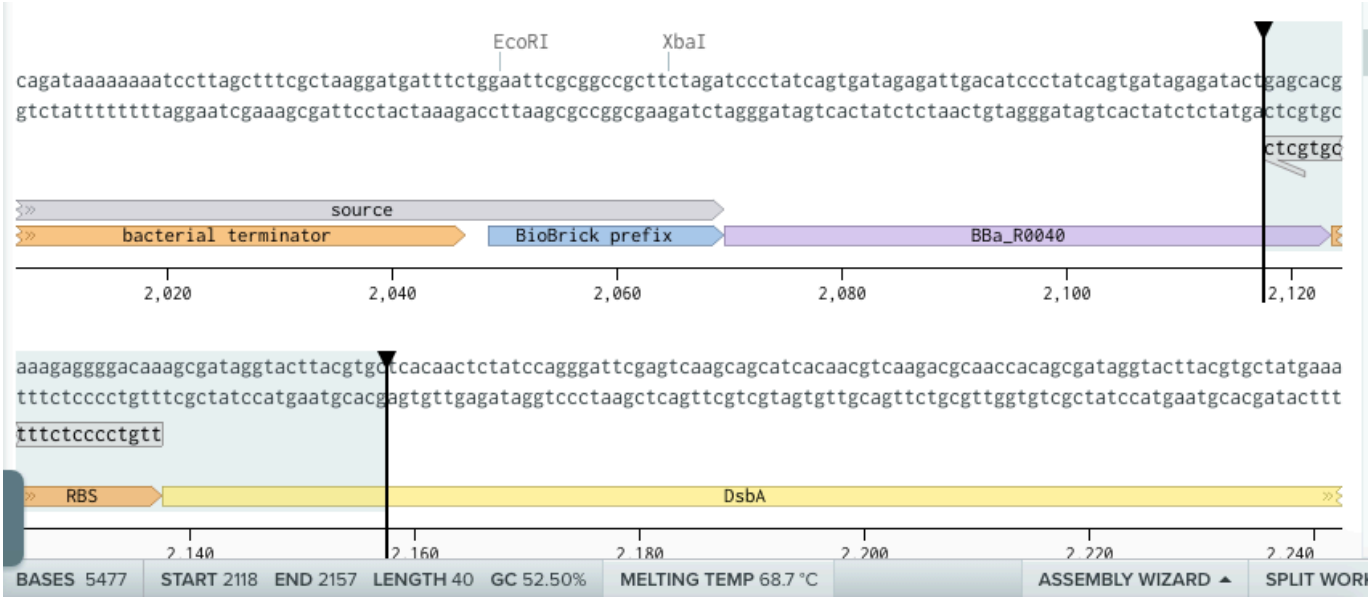
Primer Design Example:

I want to amplify DsbA (ignore the other primers). I call DsbA the "Insert":



To amplify the sequence, the FWD and REV primers must point inward toward DsbA. First, I choose primers that are exactly 40bp length (20 bp HEAD, 20bp TAIL).

On benchling, I select this, and I copy it onto an Excel sheet as my FWD primer. When I copy on benchling, it automatically copies the SENSE strand (the top strand or the 5'→3' strand). More on this later.

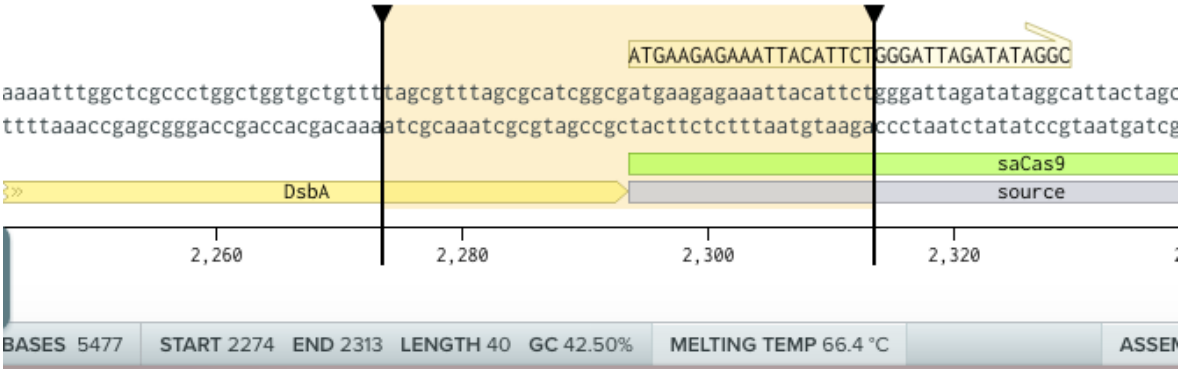


Notice that on the bar at the bottom, it tells me the position I'm at (Start: 2118, End: 2157), the length of what I'm selecting (Length: 40), the GC content (GC: 52.50%), and the melting temp (melting temp: 68.7C).

One thing your project will have to mimic is providing this kind of data for a given sequence.
Your project will improve upon the "melting temp" statistic: here, it only shows one temp (because it doesn't know head or tail. Your program should know, or give an option to mark which section is head

or tail). Also, the temp your program provides will be more accurate, from the IDT website (idtdna.com/calc/analyzer)

Now I want to create my initial REV strand. I highlight this:



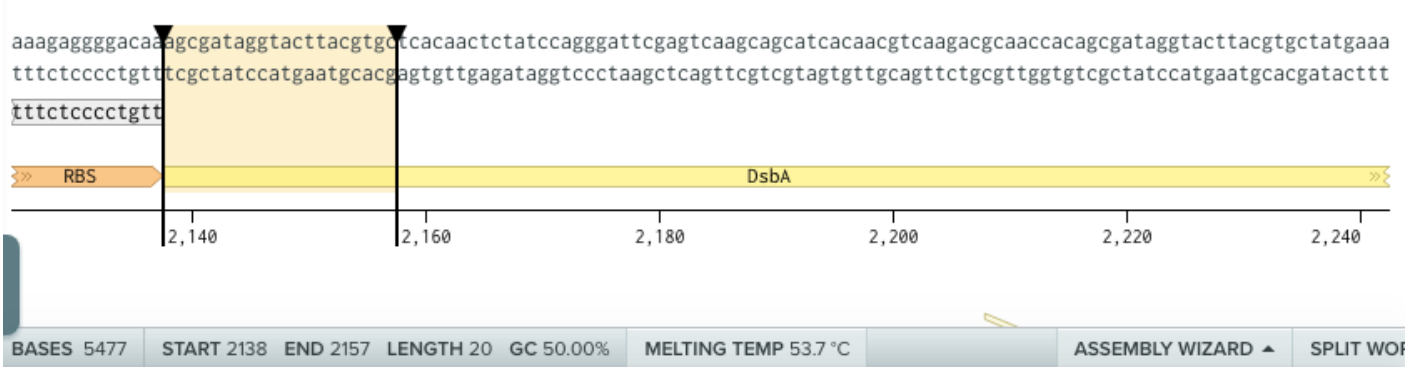
And I copy the REVERSE COMPLEMENT onto the excel sheet. Benchling gives me these options when I want to copy a sequence:

I select "Reverse Complement", which gives me the correct primer written 5'→3'. Whenever I want to check the temperature, I choose the "Reverse Complement" to check.

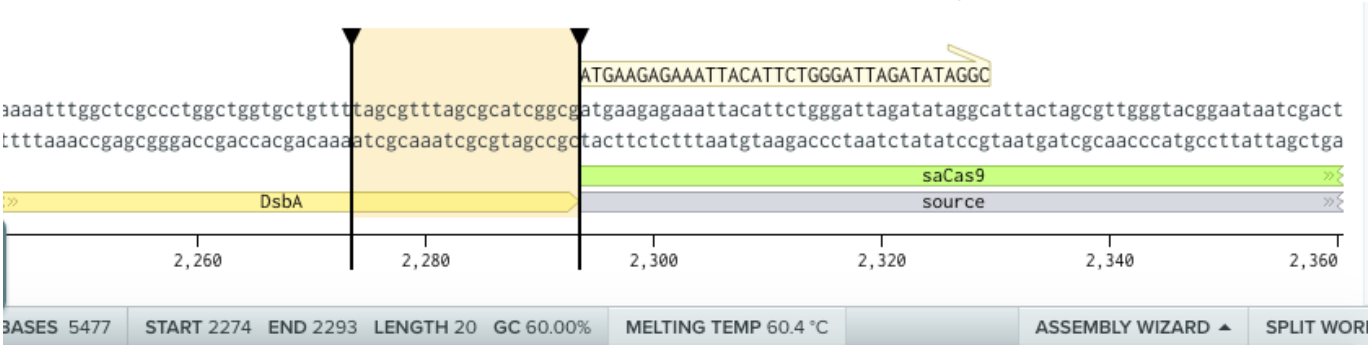
Your program should be able to do this as well: allow me to copy either the SENSE or the ANTISENSE strand (aka, the SEQUENCE or the REVERSE COMPLEMENT). There is never the case where I want ONLY the complement or ONLY the reverse.

Now, I get to take these 2 sequences, and check on IDT what the T anneal and T whole are. Remember, T anneal is based on the part of the primer that will attach to DsbA.

For the **FWD** primer, it is this selection: (5'- agcgataggtacttacgtgc – 3')



For the REV primer, it is this selection: (5'- cgccgatgcgctaaacgcta – 3'). Note that this is the REVERSE of the second strand (antisense), aka the REVERSE COMPLEMENT of the top strand (sense).



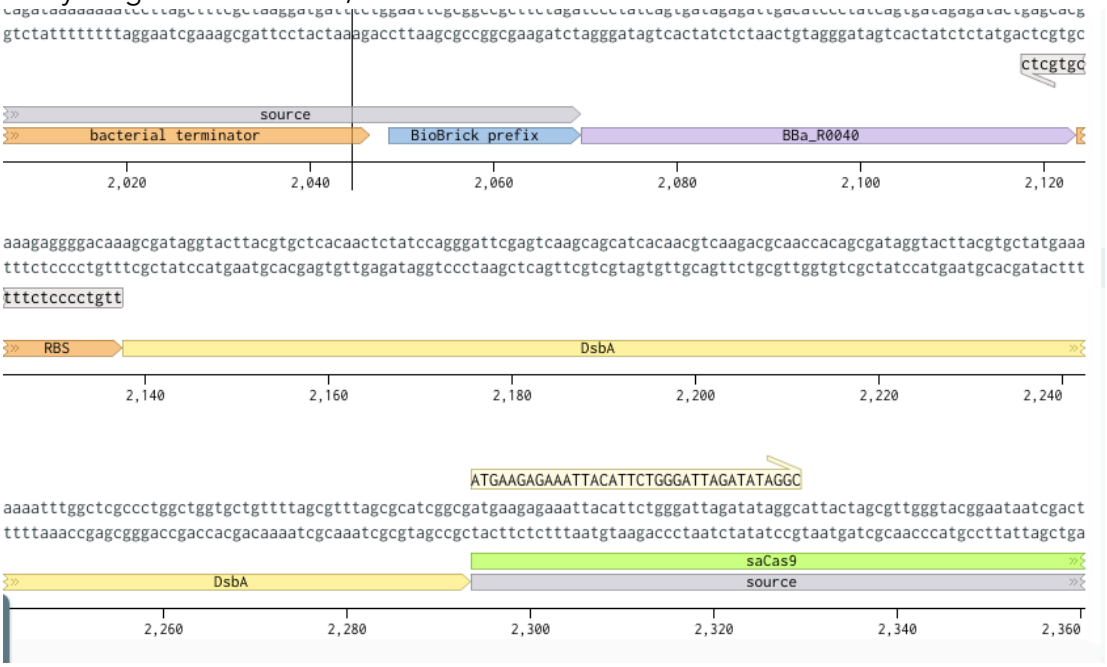
So I go to the IDT website idtdna.com/calc/analyzer and I compare the temperatures. Then I follow the same steps as listed above, if the temperatures don't match.

After the T anneal, T whole, and GC content check out, I will copy the primers over into Benchling.

I give my primers names, and everything is done.

Tangent:

If I wanted to amplify everything EXCEPT DsbA, the primers would be facing in the opposite direction. Everything else I call "bb", or "backbone".



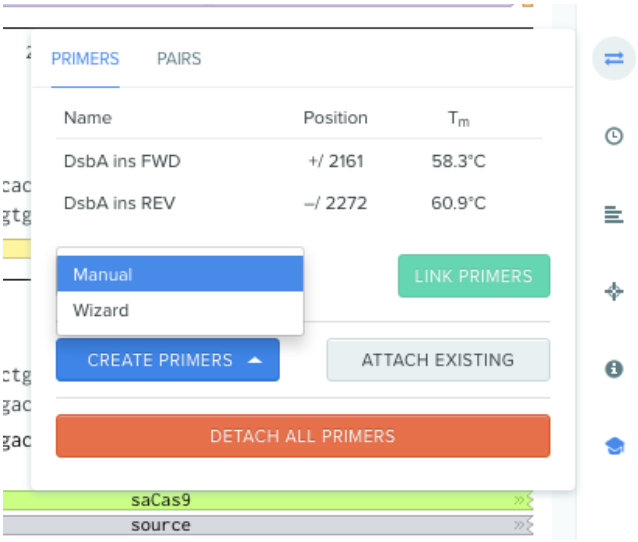
In the end, my excel sheet will look something like this:

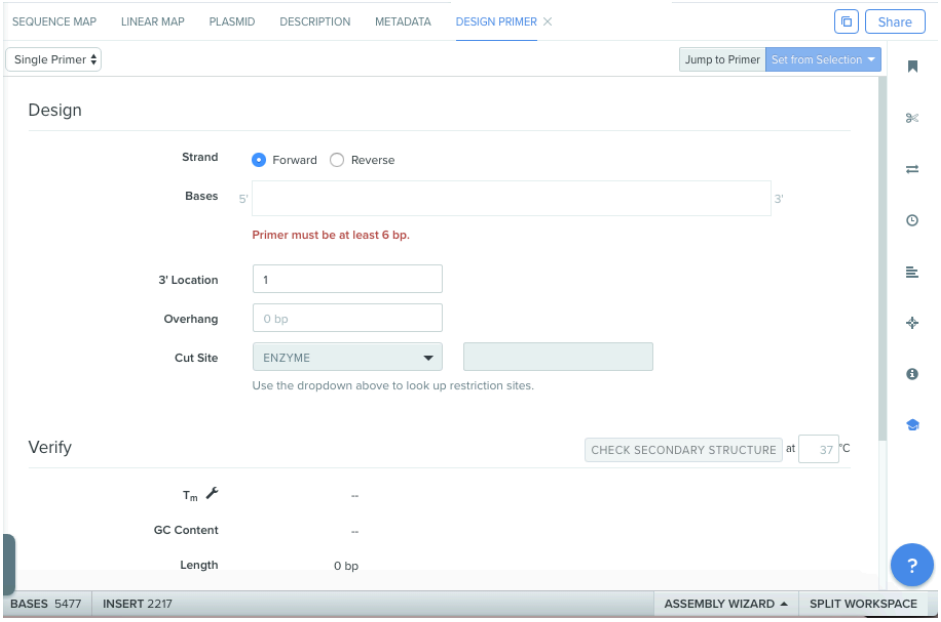
B	C	D	E	F	G	H	I	J	K	L
Part	**Sequence**	**Length**	**T anneal**	**T whole**	**GC content**		**Anneal seq**	**Length**	**Overhang**	**Length**
DsbA										
bb FWD	tagcgtttagcgcatcggcgGCACGTATCTTAGCGTTTGA	42	55.4	69.7	52.40%		GCACGTATCTTAGCGTTT(	22	tagcgtttagcgcatcggcg	20
bb REV	gcacgtaagtacctatcgctttgtcccctctttcgtgctc	40	57.2	67.1	52.50%		ttgtcccctctttcgtgctc	20	gcacgtaagtacctatcgct	20
ins FWD	gagcacgaaagaggggacaaagcgataggtacttacgtgctcac	44	58.2	67.9	52.30%		agcgataggtacttacgtgctcac	24	gagcacgaaagagggaca	20
ins REV	ccagaatgtaatttctcttcatcgccgatgcgctaaacgctaaa	44	60.6	65.8	43.20%		cgccgatgcgctaaacgctaaa	22	ccagaatgtaatttctcttcat	22

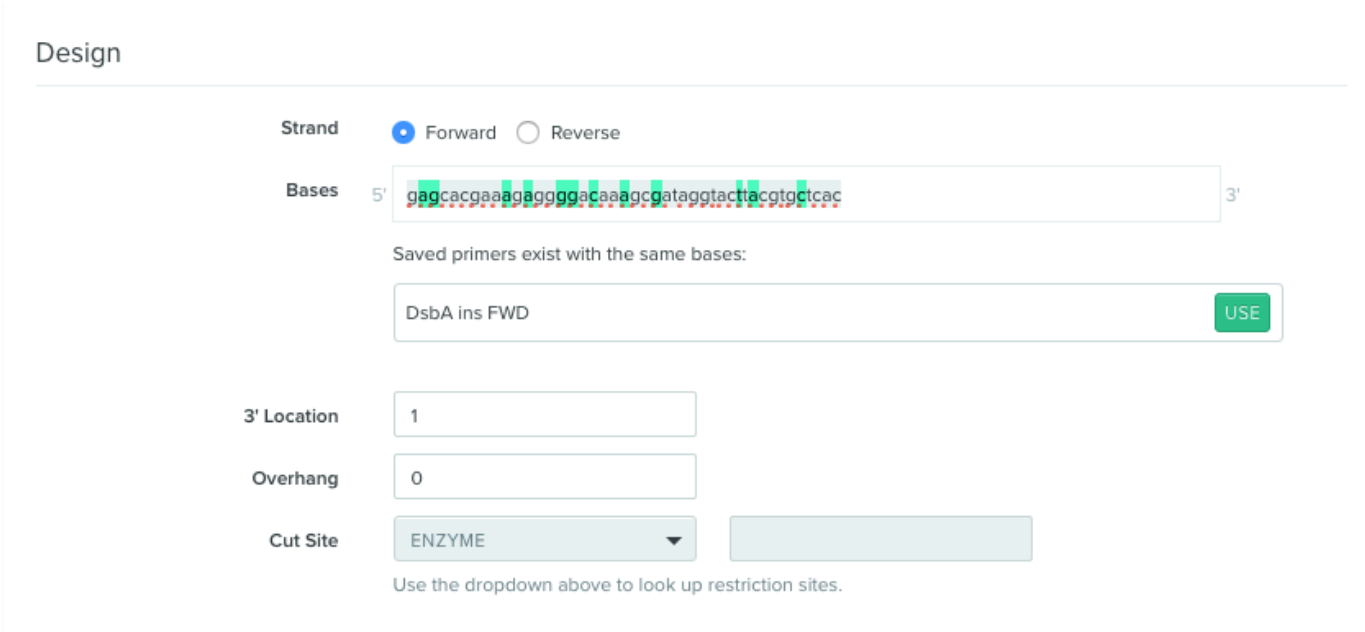
This excel sheet helps me keep track of every piece of information I need.

Your program should be able to give me a report like this as well. All sequences are reported 5′ to 3′. Do not worry about the visuals, I would work off this excel sheet to input (manually) all of the primers into Benchling.

If your program is very smart, it may also be allowed to use Benchling on my behalf.

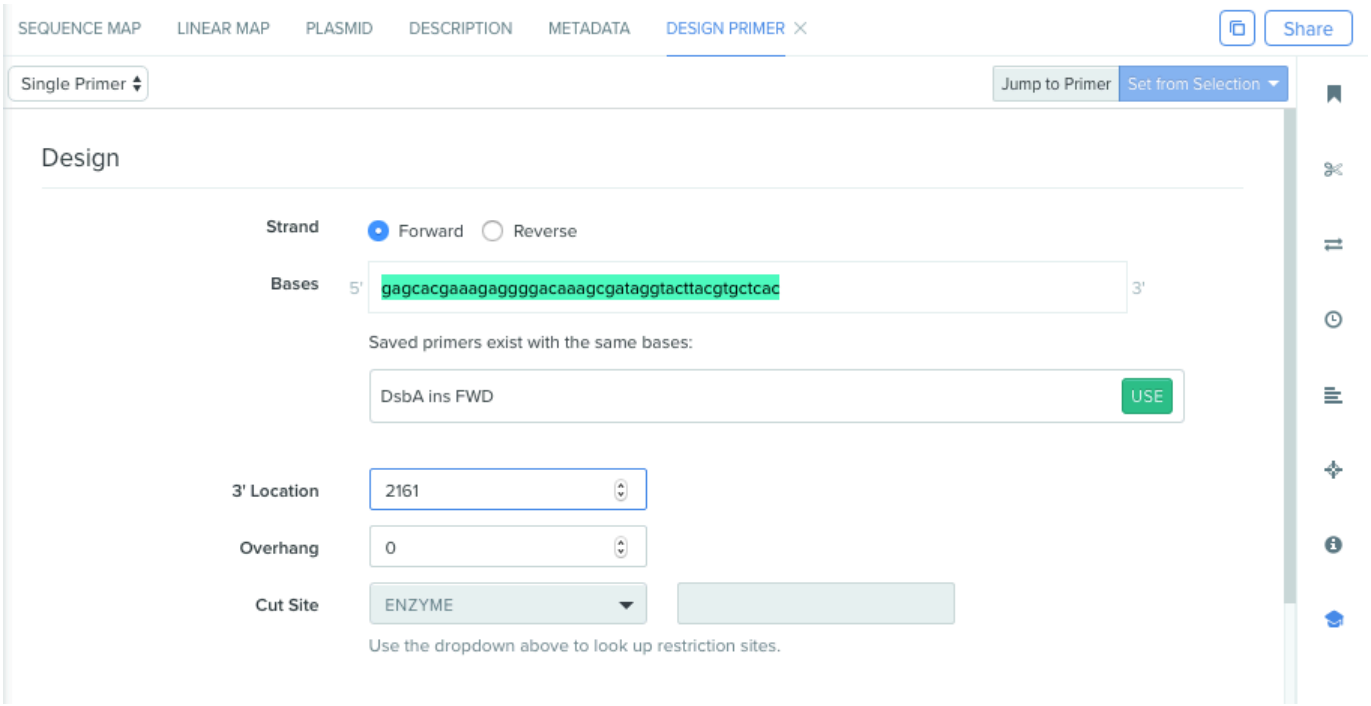




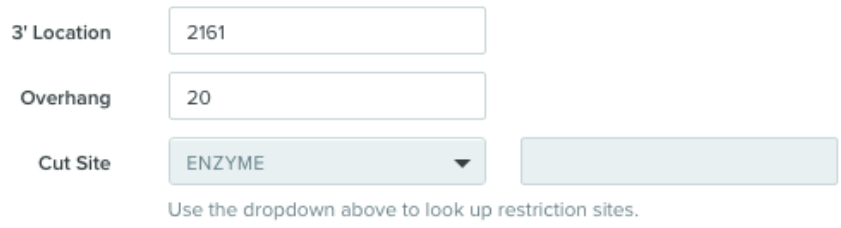


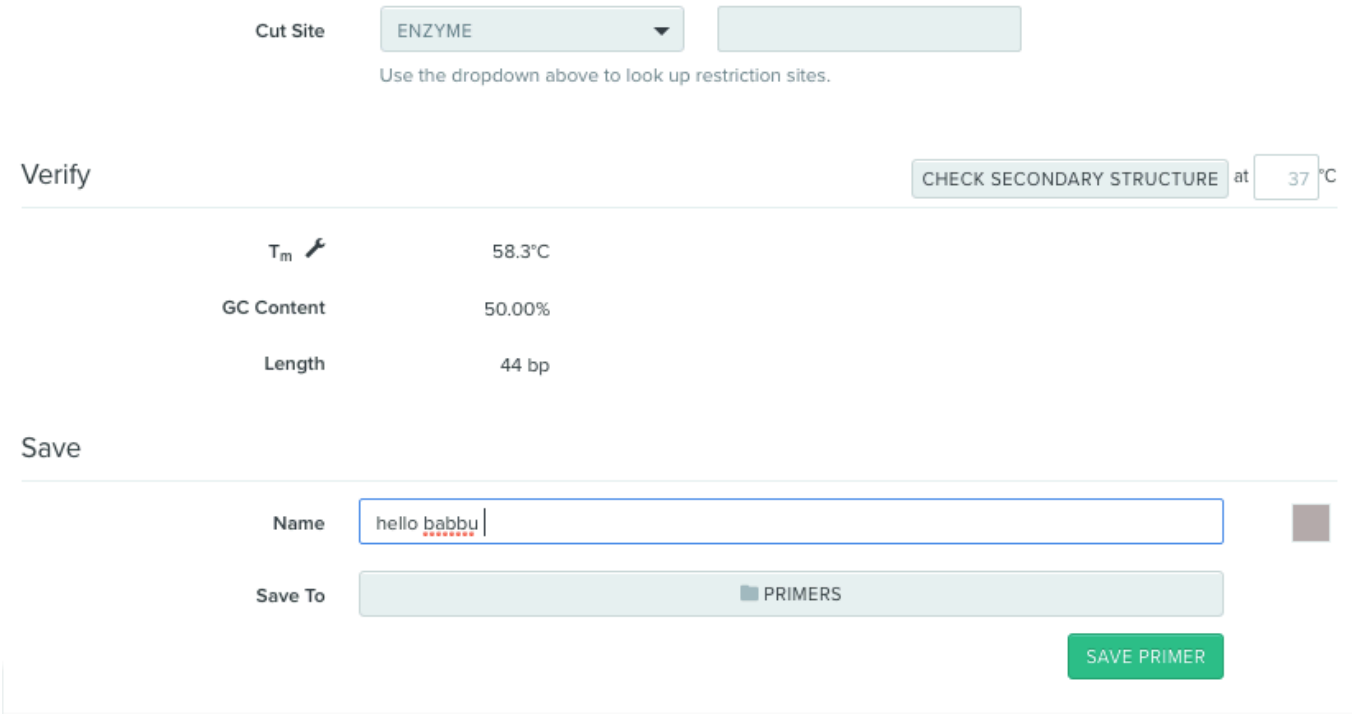
Where is the location? Search on **SEQUENCE MAP**

Put the cursor at the correct location: Insert ends at 2161



Report to the DESIGN PRIMER tab and set the location and overhang length.





Give it a cute name, and its done!

In summary:

Your program should:

1. Show SENSE and ANTISENSE strands
2. Copy either the SEQUENCE (sense) or the REVERE COMPLEMENT (reverse of antisense)
3. Mark names of sequences (DsbA, for example)
4. Know what segment is being amplified and what direction the FWD and REV primers should be in
5. Execute the steps of checking temperature, GC content, and primer length (HEAD & TAIL)
6. Keep track of primer data in an excel report
7. Use Benchling, if possible.