1 Primer Melting Point Prediction

1.1 Approach

My approach for finding melting point features was to search for different properties of the primer that might affect the melting point of the primer. The properties I settled on was:

1. The length of the primer
2. The count of each nucleotide
3. The first nucleotide
4. The last nucleotide

1.2 Findings

This gave me a $.96$ R^2 value, meaning that it worked quite well. The strongest indicator that I found with later testing was the length, but the count of each nucleotide mattered too. Lastly, including the first and last nucleotide of the primer increased the R^2 from $.94$ to $.96$.

2 Predict PCR Products

2.1 Approach

My approach to this was to sequentially ensure each constraint was satisfied and then run the sequence alignment script to find whether binding would succeed or not. The following was the process:

1. Check length of primers ($18 \leq$ length ≤ 35)
2. Check melting points ($58 \leq$ melting point ≤ 62)
3. Check alignment match ($.8 \leq$ alignment score)

If all of these requirements are satisfied we can find the predicted product. Note that we ensure the product length is less than 1000 bases downstream.

2.2 Findings

I found that this worked as intended. The test cases succeeded and my code returned the same results and my peers.

3 Generate Primers for PCR

3.1 Approach

To find primers that worked for multiple rRNA strands, I applied a brute force approach. That is, I attempted to try every possible valid primer pair. However, there are quite a few of these, so I lowered the search space in a few ways:

1. Only search through perfect matches on one of the strands
2. Reject primers immediately if their melting point is not in range
3. If a match is found in the strand, try all other strands to look for pairings or triples

The first point greatly decreases the number of strings we try, but it also cuts out some of the strings that might be good matches.

The second point helps avoid doing extra computation when the first primer chosen would never work.

The last point gives us a higher probability to find a three or more way match since we are search over more extra strands, instead of just an arbitrary three strands.

3.2 Findings

This procedure worked well, but was quite slow. I found many primers that worked, but I had to run the program for multiple hours. To fix this, I could work on parallelizing this to make my computer use every core. Additionally, I could run this via cloud computing to speed up the process as well.

4 Generate Primers for PCR Differentiation

4.1 Approach

The approach for this section is identical to the previous one, but this time we will filter the results for primers that create different lengths for each or some of the strands.

4.2 Findings

Finding a primer pair that created three different lengths for three different strands was extremely difficult, since they seem quite rare. Instead, our team found two pairs that would differentiate between many of the strands by using the combination of the products (see section 5). Running the code for this section was also slow, so we did not search the entire space.

5 Primer Binding Locations and Experimental Design

5.1 Primer Binding Locations

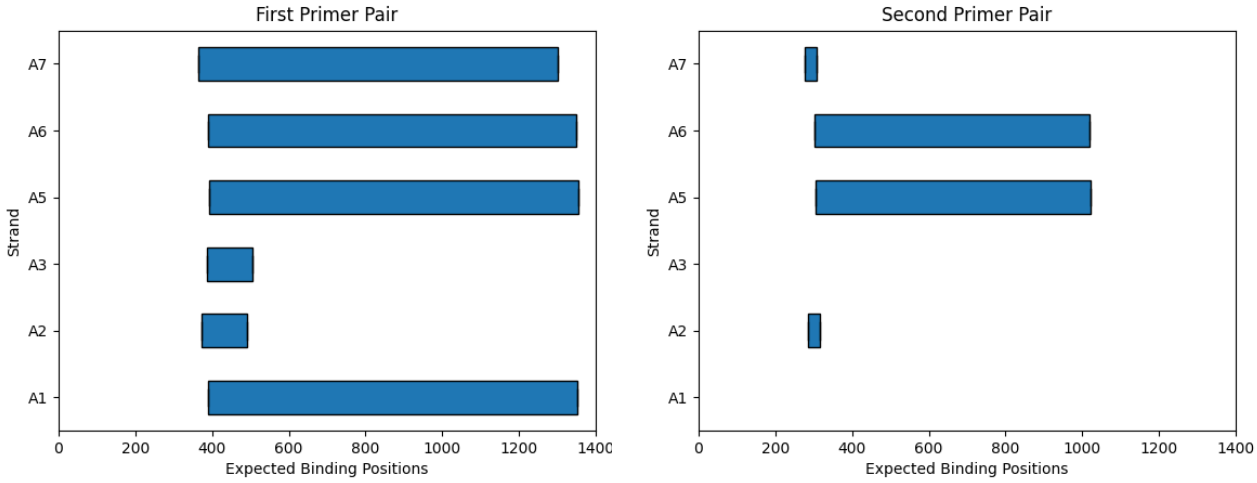

Figure 1: Binding locations of primer pairs

We selected two sets of primers, with expected binding locations shown above.

5.2 Experimental Design

First, to test if the primers are working as intended, I would use each of the primers on the pure bacteria DNA to ensure the products have the correct lengths. The primers were chosen such that the first primer pair is should be guaranteed to make a product. If it does not, we know that something is wrong. I would also test the second primer pair on the DNA to ensure that the correct product is made for A2, A5, A6, and A7, and that no product is made for A1 and A3.

Now that we have comfirmed the functionality of the primer, we will observe the relationship between the products and their lengths:

Length (bases)	Possible DNA Source	Primer Source
0	A3 or A1	2
≈ 50	A2 or A7	2
≈ 140	A2 or A3	1
≈ 740	A5 or A6	2
≈ 980	A1 or A5 or A6 or A7	1

Table 1: Length vs possible DNA and primer source

With this in mind, we should run PCR on our unknown sample. We can use both primer pairs at once, or we could run the PCR separately. This will not matter since the outputs are separable by the product lengths. First, lets assume we only have one of the known DNA in the sample:

DNA Source	Expected Products Lengths
A1	980
A2	50 and 140
A3	140
A5,A6	740 and 980
A7	50 and 980

Table 2: DNA source vs expected products

Note that A5 and A6 are inseparable in this experiment, but all other DNA sources have a unique set of product lengths. If we find any of these sets of product lengths, we can conclude that the corresponding DNA is in the

sample.

If we do not recieve a set of lengths similar to above or believe multiple types of the known DNA are present in our sample, we can refer to the complete table below to see what combinations of DNA might be possible. Product lengths from primer 1 and primer 2 have been separated, but they don't necessarily have to be separated when running the experiment.

Lengths from primer 1	Lengths from primer 2	Possible source sets
140		A3
140	50	A2 and (possibly A3)
980		A1
980	50	(possibly A1) and A7
980	740	(possibly A1) and (A5 or A6)
980	50 and 740	(possibly A1) and (A5 or A6) and A7
140 and 980		A1 and A3
140 and 980	50	(possibly A1) and A2 and (possibly A3) and A7
140 and 980	740	(possibly A1) and A3 and (A5 or A6)
140 and 980	50 and 740	(ALO: A1, A5, A6, A7) and (ALO: A2, A3) and (ALO: A2, A7) and (ALO: A5, A6)

Table 3: Lengths vs possible DNA source, ALO stands for at least one of the DNA following it. All combinations that are not mentioned are impossible.

In summary, our experiment is to run PCR on our control DNA sequences to ensure the primers work correctly (every primer pair, every DNA sequence). Next run PCR with both primers on the unknown DNA. Finally, measure the lengths of all PCR reactions and consult the tables above for conclusions.