

# LabGenius: Software engineer tech test

At LabGenius part of the job of the software team is to build tools to help our wet lab scientists with their experiments. This can involve automating repetitive, error prone tasks or using algorithms to unlock new solutions to problems they may face.

One of the common methods used in molecular biology labs is the Polymerase Chain Reaction (PCR). At a high level it involves designing short fragments of DNA (primers) to allow for the amplification of a specific section of DNA that interests us. Don't worry, you don't need to know how this works in detail to complete this task!

The goal of this tech test is to extend an existing web app to help our scientists design DNA primers to amplify DNA regions of interest.

## DNA

---

DNA typically exists as a double stranded sequence of 4 types of building blocks (nucleotides) arranged in the famous double helix structure. It can be represented as a linear sequence of 4 characters; A, C, G, T. For example a double strand of DNA could look like:

```
GCGCATGAACGTTGTA
CGCGTACTTGCAACAT
```

Although the DNA is double stranded the nucleotide sequence on one strand determines the sequence on the other strand through the following mapping:

```
A -> T
T -> A
C -> G
G -> C
```

Therefore DNA sequences are typically represented as a single sequence of nucleotides as the opposing strand can be generated from the simple rules introduced above. And so the DNA sequence above would be represented as:

```
GCGCATGAACGTTGTA
```

## Primers

---

Primers are short fragments of single stranded DNA which allow for a region of the target DNA to be

amplified. In order to amplify a region of DNA we need to design two primers; one at the start of the region of interest and one at the end. This primer pair is completely determined by the target DNA sequence. Consider the following DNA sequence (only one strand is shown), where we have introduced gaps to highlight the central region which we wish to amplify:

```
ACGGGA CACACACGGTTGACCAGTTA ATAGGCTACA
```

Possible primer pairs could be:

```
(CACA, GTTA),  
(CACAC, CCAGTTA) etc
```

Note that primers can have different lengths and in particular can have different lengths relative to each other. The first primer must begin at the start of the region to amplify and the second primer must end at the end of the region to amplify.

In practice when designing primers, the second primer in each pair should be transformed through the DNA mapping rule introduced above and then reversed, i.e. `GTTA -> ATTG -> TAAC`. So the possible primer pairs are represented as:

```
(CACA, TAAC),  
(CACAC, TAACTGG) etc
```

For a primer to work, the temperature at which the PCR experiment is performed must be tuned. The temperature at which a primer will no longer work is called its melting temperature, in general this depends on the length of the primer and the bonds it forms with the target DNA. It can be calculated as follows:

For primers of length  $< 14$ :

$$T_{\text{melting}} = 2 \times (N_A + N_T) + 4 \times (N_G + N_C)$$

For primers of length  $\geq 14$ :

$$T_{\text{melting}} = 64.9 + 41 \times (N_G + N_C - 16.4) / (N_A + N_T + N_G + N_C)$$

Where  $N_A$ ,  $N_C$ ,  $N_G$ ,  $N_T$  are the numbers of A, C, G and T nucleotides in the primer respectively.

For example `AACC` has  $T_{\text{melting}} = 12^\circ\text{C}$  and `ATAGGCTACATTGCA` has  $T_{\text{melting}} \approx 36.5^\circ\text{C}$

Don't worry too much about understanding the biology, just try to extract the rules required to design these primers. This should be all you need to know to be able to complete the following task!

## Set up

---

This task requires `python3.7` to run the python server. If you use new packages please add them to the requirements file. - to start the flask server: `python -m app.server` - to run tests: `pytest tests`  
- to run linting: `flake8 tests app`

To build the front end first navigate into the `frontend` directory and run `npm install` :- to build and watch for changes: `npm run start` - to run linting: `npm run lint` - to run Prettier: `npm run prettier`

## Task

---

The first stages of the task can be broken down as follows:

- Make sure you can run the flask server and get the front end building correctly.
- There is a failing test in the backend which needs to be resolved appropriately.

Once this is complete you can move on to the main task, which can be completed in any way you see fit.

- Create a tool to allow a wet lab scientist to submit a DNA sequence, indicating the region they wish to amplify and a desired primer length. Display the resultant primers as well as the melting temperatures in an appropriate way.
  - With the example DNA sequence `ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGG` , region to amplify `GGCGAGGAGCTG` and primer length 4 the tool should return the primers `GGCG` , `CAGC` with melting temperatures 16°C and 14°C respectively.
- Build a separate tool to allow the wet lab scientists to submit a DNA sequence, indicating the region to amplify, as well as a temperature range, e.g 50°C - 60°C. The tool should return a ranked list of primer pairs subject to the following constraints:
  - The primers should be between 6 and 20 nucleotides in length (note that the primers in each pair can have different lengths from each other).
  - The primer melting temperatures should both be between the submitted temperature range.
  - The best primer pair should be the pair which have the most similar melting temperatures.

Below is an example of a longer, real, input DNA sequence you can use for this task:

```
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA
GTTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGC
TGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAAGTCTCAGCCGCTACCCCGACCACATGAAG
CAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC
GGCATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC
CCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACG
AGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG
```

Ensure that you write tests for your work where appropriate and think about readability.

Don't worry if you don't complete the task, just get as far as you can!

After you are happy with your solution, create a zip of this project and send it back to us at [tech-test@labgeni.us](mailto:tech-test@labgeni.us)

Good luck!