**Functional Specification for pcr\_optimizer**

Lionel Tukei and Lily Torp

1. **Problem Addressed:**

Polymerase chain reaction (PCR) is a common molecular biology technique used to amplify DNA. Since its discovery in 1983 many different PCR enzymes, reagents, and protocols have been developed. Each enzyme has different optimal reaction conditions such as extension times and primer annealing temperatures depending on the properties of the target gene, the primers used to amplify it, and buffers present in the reaction. PCR enzymes are expensive reagents, so it is also important to consider cost before choosing a protocol. Additionally, PCRs can be ineffective due to user design errors. Primers must anneal at the correct sites on a gene and have a carefully calculated annealing temperature to ensure proper binding conditions. Checking for any errors before conducting PCR is also crucial to avoid wasting time and money due to a failed reaction. Considering all these variables can make designing an effective PCR experiment difficult.

This tool optimizes the PCR protocol decision making process. First the gene and primer sequences are checked for errors and evaluated to ensure the primers are binding in the correct location on the target sequence. Next, enzyme amount, cost per reaction, annealing temperature, annealing time, extension time, and total PCR reaction time for two common PCR enzymes (iProof High-Fidelity Polymerase and Taq Polymerase) are reported to the user. The user can also designate a factor (time or cost) to focus their optimization.

1. **User Profiles:**

The target user is a scientific researcher who performs polymerase chain reactions (PCR). This includes undergraduates, research technicians, grad students, post-docs, and even principal investigators. They will have a fundamental understanding of how PCR works on the molecular level and use PCR regularly. They will have an understanding of where primers anneal to DNA, the 5’-3’ universal structure of gene/primer sequences and have a specific goal in mind when it comes to optimizing PCR reactions (cost or time).

The user should have beginner to intermediate python skills. This includes knowing how to pip install in python using a virtual environment like Colab or Jupyter Notebook. They should also understand how strings are formatted and how to define a class.

1. **Use Cases:**

The user will input the gene or DNA sequence they are amplifying and the primers they will use to amplify it. As an example, the user might want to focus on optimizing time or cost.

* Optimize PCR protocol for time:
  1. Objective: The user wants to optimize their PCR reaction for time.
  2. The user will use the python environment of their choice, install pcr\_optimizer using pip, and define a new pcr object. In the pcr object they will input the target gene, forward primer, and reverse primer sequences as strings (bases 5’-3’) and specify the template type. Next, they will run the “check” function on the object to ensure the genes and primers are compatible with no unacceptable bases. If there are any errors at this point, the user will manually fix the inputs, redefine the pcr object, and re-run the “check” function. Next the user will run the “recommend” function. They will designate “time” as the factor in “recommend”, and the function will return a table detailing annealing time, extension time, and total PCR reaction time for both enzymes. The user will then compare relevant times between the iProof High-Fidelity Polymerase and Taq Polymerase enzymes and choose which one best suits their needs.
* Optimize PCR protocol for cost:
  1. The user wants to optimize their PCR reaction for cost.
  2. The user will use the python environment of their choice, install pcr\_optimizer using pip, and define a new pcr object. In the pcr object they will input the target gene, forward primer, and reverse primer sequences as strings (bases 5’-3’) and specify the template type. Next, they will run the “check” function on the object to ensure the genes and primers are compatible with no unacceptable bases. If there are any errors at this point, the user will manually fix the inputs, redefine the pcr object, and re-run the “check” function. Next the user will run the “recommend” function. They will designate “cost” as the factor in “recommend”, and the function will return a table detailing enzyme amount and cost per reaction for both enzymes. The user will then compare relevant costs between the iProof High-Fidelity Polymerase and Taq Polymerase enzymes and choose which one best suits their needs.