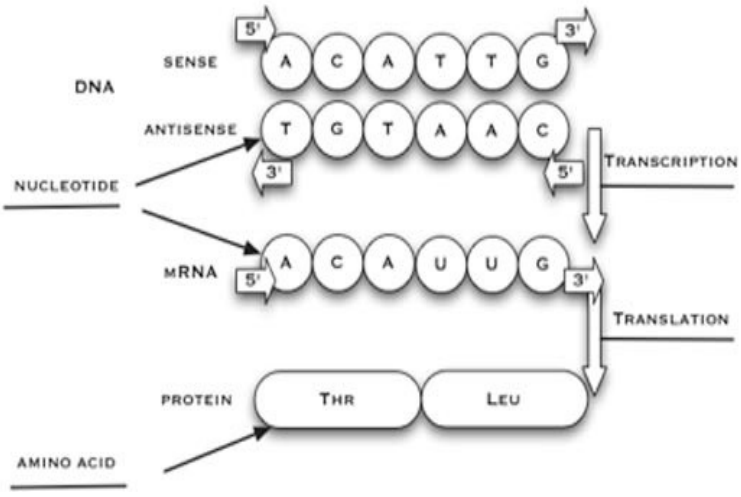
BIO722P Assessment: Background Information

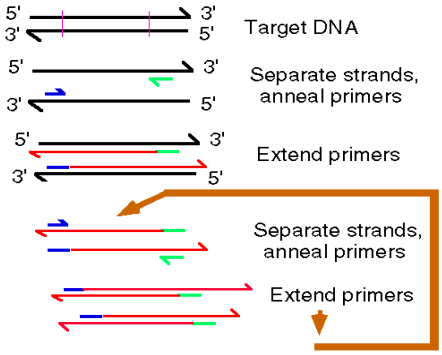
Question 1: PCR and primers

The Polymerase Chain Reaction (PCR) is a molecular biology technique used to amplify few copies of a particular DNA sequence into thousands of copies. PCR is used to amplify a specific region of a DNA strand (the DNA target). PCR is performed in a thermocycler, a machine that performs several cycles of heating and cooling. These thermal cycling steps are necessary first to physically separate the two strands in a DNA double helix at the higher temperature in a process called DNA melting. At the lower temperature, each strand is then used as the template in DNA synthesis by the DNA polymerase to selectively amplify the target DNA. The selectivity of PCR results from the use of *primers* that are complementary to the DNA region targeted for amplification.

The primers are short, chemically syntesized strands of nucleic acid that are about 20 nucleotides long and serve as a starting point for DNA synthesis. DNA-polymerase amplifies the target DNA starting from those sequences, that specifically bind the extremities of the sequence you want to clone.



A DNA sequence is always read from 5' end to 3' end and is usually written in this way: the sense sequence is written from 5' to 3', while the antisense sequence, which is the complementary sequence (because DNA is double stranded) and which is the one transcribed into mRNA, is written from 3' to 5' (antiparallel strands).



If you want to do a PCR in order to amplify a particular sequence of DNA you need to enhance both strands, so you need a primer for one strand, called the Reverse primer, that binds the 3' end of the sense strand, and a primer for the other strand, called the Forward primer, that binds the 3' end of the antisense strand. You thus have a primer for each extremity of the DNA piece you want to amplify, and you can amplify selectively the specific sequence for which the primers are designed. The diagram on the left illustrates the PCR process and the role of the primers.

The melting temperature (Tm) is the temperature at which a particular DNA duplex will dissociate and become single strand DNA. The stability of a primer-template DNA duplex can be measured by its Tm. Primers with melting temperatures in the range of 52-58°C generally produce better results than primers with lower melting temperatures. Primer length and sequence are of critical importance in designing the parameters of a successful amplification. The melting temperature of a nucleic acid duplex increases both with its length, and with increasing GC content. A simple empirical formula for calculation of Tm is:

$$\mathbf{Tm = 4(G + C) + 2(A + T)}$$

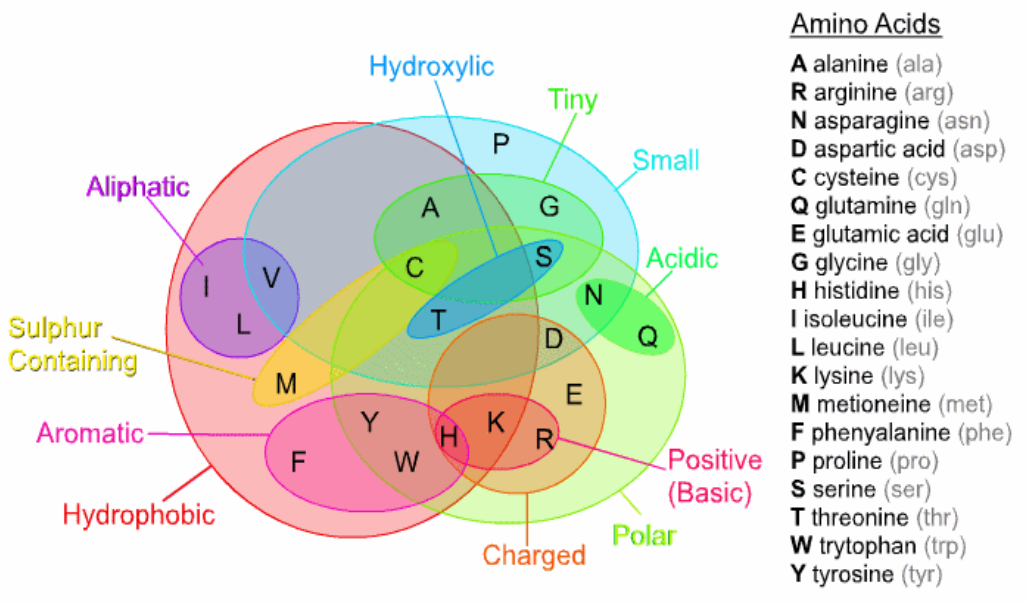
where A,C,G and T stand for the number of nucleotides of each type contained in the primer and Tm is obtained in degrees Celsius.

Question 2: Translation and reading frames

Reverse reading frames are usually taken from the reverse complement of the sequence. However, as a simplification, for the purpose of these exercises **do not complement the sequence**; just reverse it.

Question 3: Classification of amino acids

This question asks you to determine the aminoacid usage of a protein sequence in terms of three standard categories of amino acids (polar, small and hydrophobic). Refer to the Venn diagram below for a list of amino acids belonging to each category. Note that some amino acids belong to more than one category, so that the sum of the fractions of amino acids of each type contained in a sequence can be greater than 1.



Question 4: Clustering proteins based on AA usage

No supplementary information provided.