**SCOPE**

The scope of this project encompasses the identification and analysis of anchor residues within transcription factor binding sites (TFBS) using UniProbe data and MEME Suite. By doing this, it will help us make progress towards the primary objective, which is to be able to design drugs that restrict binding by targeting anchor residues to inhibited gene expression in diseased states.

Constraints:  
  
The project will be constrained by the availability and quality of data from UniProbe.  
Resource limitations, including computational resources and time constraints, may impact the scale and complexity of the analysis.  
  
Deliverables:  
  
I'll be able to produce a report detailing the methodology, findings, and conclusions of the project. Additionally, I'll may be able to represent the findings as a visualization. Finally, my code may be implemented to any chosen dataset.

**PLAN OF WORK**

Through this project, I will focus on the following key aspects:  
  
(1) Data Procurement and Preparation:  
  
I'll be procuring relevant data from the UniProbe database for a human TF (eg: GATA4). I'll be identifying the dataset required and will work on preprocessing and formatting the data to ensure compatibility with MEME Suite.  
  
(2) Identification of Motifs:  
  
I'll first familiarise myself with the command line interface of the MEME Suite tool, which is available through a docker container. I'll be experimenting with the ideal number of motifs to be identified, interpreting PWM (Probability Weight Matrix).  
  
(3) Binning dataset to identify Anchor Residues:  
  
To understand what are the anchor residues, I'll have to find the motifs for each bin (to be experimentally set) acquired from the top-enrichment file. Then, I'll have to see which motifs are common across all the bins - the identified should be the anchor residues.