

In [1]:

```
#https://biopython.org/docs/1.75/api/Bio.pairwise2.html
#https://biopython.org/DIST/docs/tutorial/Tutorial.html#sec99
#https://www.ncbi.nlm.nih.gov/nuccore/NM_000546.6
#https://www.ncbi.nlm.nih.gov/nuccore/XM_003416902.3

#read a DNA interaction domain from fasta
from Bio import pairwise2
from Bio import SeqIO
elephantseq1 = SeqIO.read("loxodontap53_nucleotide_DNAbinding.fasta", "fasta")
humanseq2 = SeqIO.read("humanp53_nucleotide_DNAbinding.fasta", "fasta")
alignments = pairwise2.align.globalxx(elephantseq1.seq, humanseq2.seq)

print(pairwise2.format_alignment(*alignments[0]))
```

```
CACCA-TG---G--CCATC-TACAA--GAAGTCAGAGC-
|   ||   |   ||   |   |   |   ||   ||
---C-GTGTGGGCC-T-GT-C--CTG--G---GAG-A
Score=14
```

In [2]:

```
#https://logomaker.readthedocs.io/en/latest/
```

In [3]:

```
pip install logomaker
```

```
Requirement already satisfied: logomaker in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (0.8)
Requirement already satisfied: pandas in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from logomaker) (1.3.4)
Requirement already satisfied: numpy in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from logomaker) (1.20.3)
Requirement already satisfied: matplotlib in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from logomaker) (3.4.3)
Requirement already satisfied: cycler>=0.10 in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from matplotlib->logomaker) (0.10.0)
Requirement already satisfied: pillow>=6.2.0 in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from matplotlib->logomaker) (8.4.0)
Requirement already satisfied: python-dateutil>=2.7 in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from matplotlib->logomaker) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from matplotlib->logomaker) (1.3.1)
Requirement already satisfied: pyparsing>=2.2.1 in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from matplotlib->logomaker) (3.0.4)
Requirement already satisfied: six in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from cycler>=0.10->matplotlib->logomaker) (1.16.0)
Requirement already satisfied: pytz>=2017.3 in /Users/zachturman/opt/anaconda3/lib/python3.9/site-packages (from pandas->logomaker) (2021.3)
Note: you may need to restart the kernel to use updated packages.
```

I can't figure out how to read AA sequences into jupyter notebook.

In [6]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
plt.ion()

import logomaker
```

In [8]:

```
#add more sequences for sequence logo, only look at DNA binding domains.
#result is from C to N terminus
proteins = ["FRLGLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVRRCPHHERCSDSGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNCNSSCMGMNRRPILTIITLEDSSG"]
proteindf = pd.DataFrame(proteins)
print(proteindf)
proteindf.columns = ['Peptide Sequences']
proteindf.index = ['Human', 'Elephant', 'Mouse']
proteindf.head()

crp_sites_list = proteindf['Peptide Sequences'].values
crp_sites_list[:3]

crp_counts_df2 = logomaker.alignment_to_matrix(sequences=crp_sites_list, to_type='counts')
#logomaker.Logo(crp_counts_df2)

logo = logomaker.Logo(df=crp_counts_df2,
                      font_name='Arial',
                      fade_below=0.5,
                      shade_below=0.5,
                      figsize=(800,20))
```

In [9]:

```
print(len(proteins[0]))
print(len(proteins[1]))
print(len(proteins[2]))
```

```
180
180
180
```

In [10]:

```
#https://github.com/nglviewer/nglview
#import nglview
#view = nglview.show_file("entire_elephantp53_CITASSER.pdb")
#view.add_representation('cartoon', selection='chain A')
#view
```

In [11]:

```
#https://github.com/nglviewer/nglview
#http://nglviewer.org/nglview/release/v0.5.1/#usage
```

```
import nglview
view = nglview.show_file("humanp53_DNA_binding_structure.pdb")
view.add_representation('cartoon')
#view.camera = 'orthographic'
#view.frame = 100
view.add_licorice('ALA, GLU')
view
```

In [12]:

```
import nglview
view2 = nglview.show_file("loxodontap53_DNAbinding_structure.pdb")
view.add_representation('cartoon')
view2
```

In [13]:

```
from ipywidgets.widgets import Box
w1 = nglview.show_file("humanp53_DNA_binding_structure.pdb")
w2 = nglview.show_file("loxodontap53_DNAbinding_structure.pdb")
Box(children=(w1,w2))
```

In [14]:

```
#still need to compare the functional domains of each protein, will do this in final draft
```