Project 3

PSSM

Compute PSSM

- create count matrix (each AA times number of AA)

- count the number of times each AA acid appears in the different sequences by column $n_{u,b}$

- compute AA frequency by column

$$f_{u,b} = \frac{n_{u,b}}{N_{seq}}$$

- compute probability of finding AA in a column (p_a obtained from swissprot)

$$q_{u,a} = \frac{\alpha f_{u,a} + \beta p_a}{\alpha + \beta}$$

- compute final score

$$m_{u,a} = \log \frac{q_{u,a}}{p_a}$$

$$\beta = \sqrt{N_{seq}}$$

$$\alpha = N'_{seq}$$

N'_{seq} number of sequences without gap in that position

Alignment

- adapt your alignment algorithm:
  - alignment matrix is now computed using
    - initialize: with gap penalties as before
    - matrix S(i,j)=max{S(i-1,j-1)+PSSM(seq(i), j),
                        S(i-1,j)+PSSM("_",j),
                        S(i,j-1)+PSSM("_",j-1)
                        0 }
- j is column of the PSSM
- PSSM("_",j) corresponds to the penalty you assigned for a gap in position j, this is a choice we made.