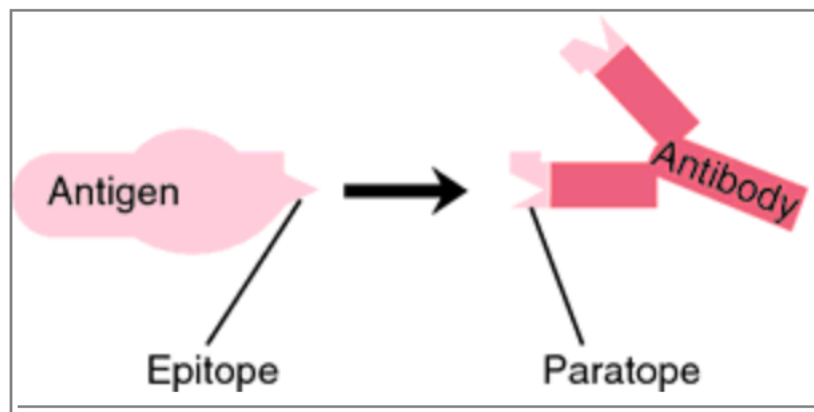# Estimate epitope residues in protein sequence.

Summary:

 - Sequence

   - LSTM or CNN

- Use BioInfoTools

Epitope:

The small site on an antigen to which a complementary antibody may specifically bind is called an epitope or antigenic determinant.

Antibodies have become an indispensable tool for many biotechnological and clinical applications. They bind their molecular target (antigen) by rec- ognizing a portion of its structure (epitope) in a highly specific manner. The ability to predict epi- topes from antigen sequences alone is a complex task. D

Input:

Protein Sequence - $a_1, a_2, \ldots, a_N$. a_i - Amino acid #i in chain

Output:

$$p(a_1), p(a_2), \ldots, p(a_N), \ p(a_k) - \text{probability of the amino acid } k \text{ to be in the}$$
epitope

Process (suggested):

- Process for each residue
  - Extract window (say 9 aa's)
  - Calculate :
    - computed volume
    - hydrophobicity (17)
    - polarity (18)
    - relative surface accessibility (RSA)
    - secondary structure (SS)
    - Type
- Run Convolution Neural Network on 9x6 Matrix to Estimate the probability of being in Epitope

---

Another Option - Use LSTM Network.

Same as BepiPred2 (see the paper) but use Convolution Neural Network (or other Deep Learning Method)

Database:

- Use linear epitope database for train and test http://www.cbs.dtu.dk/services/BepiPred/download.php

References:

PDB format - https://www.rcsb.org/pdb/static.do?p=file_formats/pdb/index.html
Chimera Tutorials - https://www.cgl.ucsf.edu/chimera/tutorials.html
BioPython - https://biopython.org

BepiPred-2.0:
- http://www.cbs.dtu.dk/services/BepiPred/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5570230/