

Entropy

The site **entropy** h_r for site r is defined as $h_r = - \sum_x \pi_{r,x} \log(\pi_{r,x})$ where $\pi_{r,x}$ is the preference for character (e.g., amino acid) x at site r , and the log is the natural logarithm.

You can use the `scipy` function to calculate entropy:

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.entropy.html>

You would just give the function a single sites preferences as a scipy array:

`Prefs = scipy.array([0.2, 0.3, 0.4])` (if there were only three amino acids)

Sites with low entropy usually have only a few highly preferred amino acids. We would say they are mutationally constrained.

Sites with a high entropy have many basically equally preferred amino acids. We would say they are mutationally tolerant.

You might expect the constrained (low entropy) sites to be on the interior of the protein

