

Introduction to Python

Day Four Exercises

Stephanie Spielman

Email: stephanie.spielman@gmail.com

1 Manipulating Sequence Data in BioPython

1. The first exercise will focus on manipulating sequences in the file "dopamine_sequences.fasta". Perform the following tasks:
 - (a) Use the SeqIO module to read in the file "dopamine_sequences.fasta".
 - (b) Determine the length for each sequence, and save this information to a dictionary. The keys should be the sequence ID, and the values should be the length.
 - (c) Translate each sequence to amino-acids, and then determine the number of tryptophans (coded as "W") in each sequence, and again save this information in a dictionary. (Hint: BioPython has implemented certain useful string methods, like `.count()` for Seq object sequences.)
 - (d) Re-do parts (b) and (c) by writing *functions* to count the sequence length and determine the number of tryptophans. Each function should take a *single* SeqRecord object as the argument.
 - (e) Create a new file **with format XXX**.

References

SJ Spielman, K Kumar, and CO Wilke. Comprehensive, structurally-informed alignment and phylogeny of vertebrate biogenic amine receptors. *PeerJ*, 3:e773, 2015. doi: 10.7717/peerj.773.