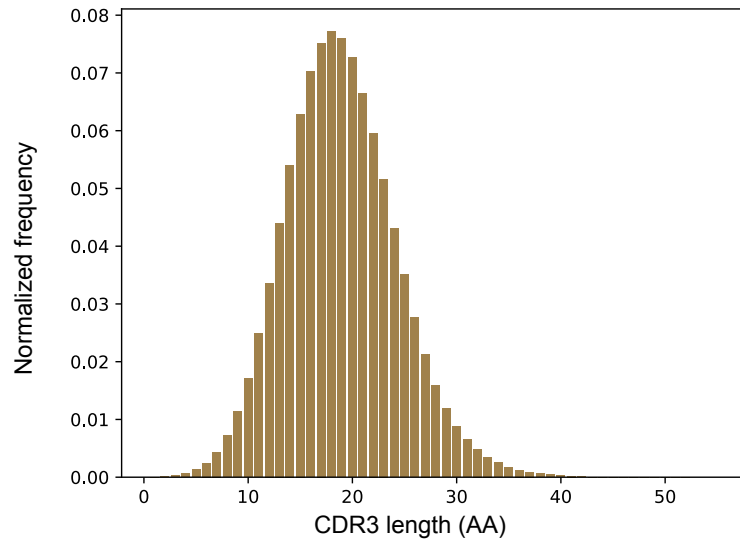


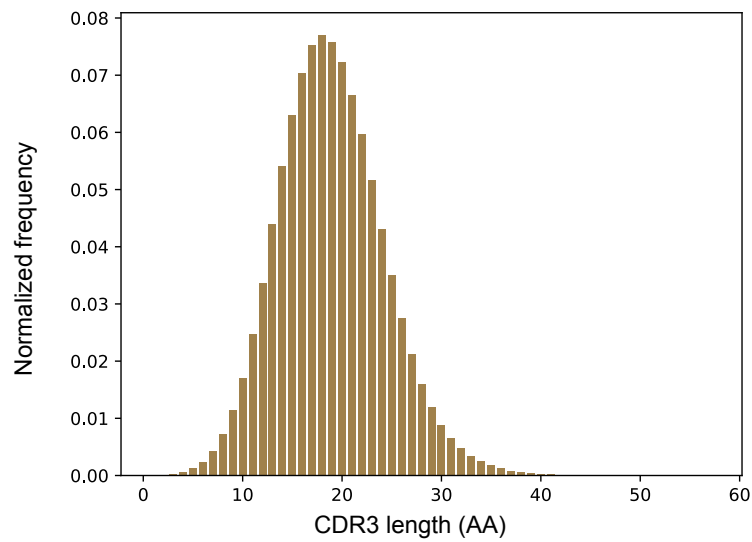
A

10 million sequences



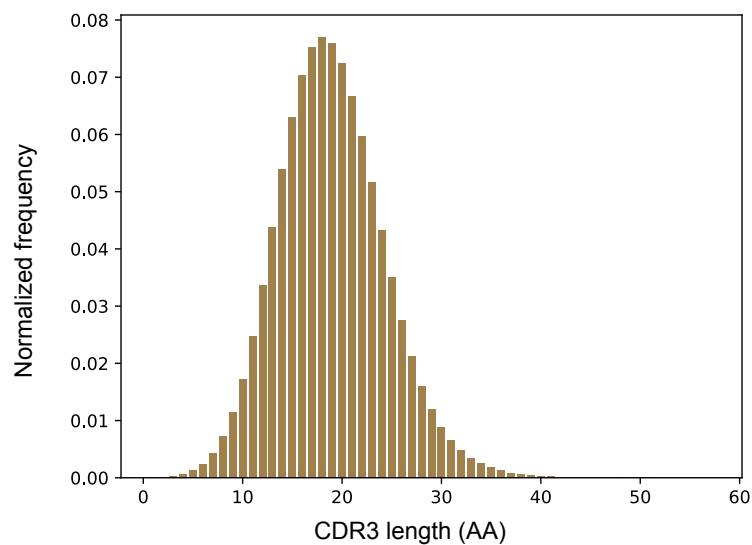
B

50 million sequences



C

100 million sequences



Supplementary Figure 1: **Generating CDR3 length distributions with PyIR's API.** All data used in generating histograms come from synthetic heavy chain sequence data used in the Briney *et al.*, study. All CDR3 length distributions were normalized by the total number of productive sequences after processing and filtering sequences with PyIR. (A) 10 million sequences (B) 50 million sequences and (C) 100 million sequences.