



**Figure S2.** Length distribution histogram of sequence entries in the GH19ED database, with a bin size of 20 residues. The two main peaks are around 200 and 580 residues. Only few sequences are longer than 1100 residues (up to 6000 residues).