Supplementary figure 1 The percentage of two kinds of proteins in training and testing dataset