



Figure S3. Distribution of occurrence frequencies for 20 amino acids. This figure demonstrates the non-normal distribution of occurrence frequencies of each amino acid. The background group (nrCPDB-40; blue bars) and CP site groups (nrCPsite_{cpdb}-40, $p_{cut} \pm 3$; red bars) were the same as those used in Figure S2. In each diagram, the x axis represents the occurrence frequency of a specific amino acid in a polypeptide, and the y axis represents the proportion of polypeptides that possessing a certain occurrence frequency of the amino acid. For instance, The first blue bar in diagram (a) indicates that there were 20% (*i.e.*, 0.2) of all polypeptides in the background group that possessed 0% – 5% Ala residues. The distributions of the CP site groups are all skewed to the left and clearly non-normal. Although some amino acids, *e.g.*, Ala and Gly, have a visually more normal distribution than the CP site groups do, most amino acids still have very skewed distributions. To determine the significance of difference between samples with non-normal distribution(s), the traditional t-test is inadequate; instead, we performed the permutation test (see **Materials and Methods**).