**Additional Data File 3**: The distributions of a 23-state structural alphabet on each kind of eight DSSP secondary structure codes, including the H, I, G, E, B, T, S, and other. In general, the H, I, and G are considered as $\alpha$-helixs; the E and B considered as $\beta$-strands, and other codes are considered as coils.