

FIG. W4. The root mean square error (RMSE) and Pearson correlation coefficient (CC) curves used to determine the number of sites to select from the H3N2 data (1968 to 2007 viruses). The triangle curve plots the trend of RMSE against the number of sites selected, and the dot curve plots the trend of CC against the number of sites.

