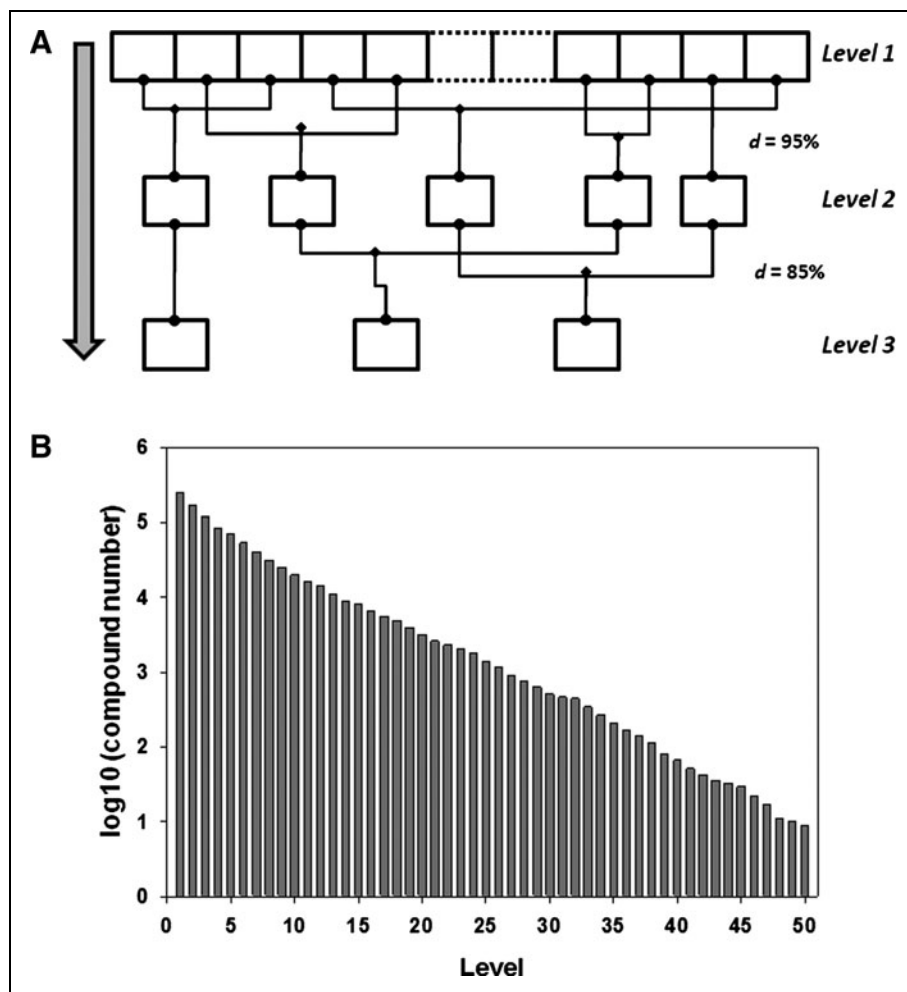


## SUPPLEMENTARY DATA



**Supplementary Fig. S1.** The compound clustering tree in the VLS process. **(A)** A tree construction algorithm. The arrow indicates the clustering direction. The squares indicate the individual compounds.  $d$ , chemical similarity threshold. **(B)** Distribution of the compounds at the individual levels of the clustering tree in which level 0 corresponds to the  $\sim 275,000$ -compound NCI ligand database.