第 11 回 2013 年 1 月 18 日（金）14:50 ~ 15:40

An exhaustive similarity search of known and putative ligand binding sites in proteins

タンパク質の既知及び推定基質結合部位の網羅的類似探索

Kentaro Tomii / 富井　健太郎
Computational Biology Research Center (CBRC)
National Institute of Advanced Industrial Science and Technology (AIST)
産業技術総合研究所　生命情報工学研究センター

We proposed an ultrafast alignment-free method for comparing huge number of protein-ligand binding sites, in which binding sites are mapped as vectors onto a high-dimensional feature space based on their physicochemical and geometrical properties. Once binding sites are represented as bit strings, called structural sketches, which is obtained by random projections of the vectors, a multiple sorting method is applied to the enumeration of all similar pairs in terms of the Hamming distance. With 3.4 million known and potential ligand binding sites, the proposed method found 24 million similar binding site pairs. Performing such a comprehensive comparison has been demonstrated to be useful for annotation of protein functions.

Keywords：protein-ligand interactions, similarity search, functional sites, database