

The evaluation of protein-ligand binding site similarity is crucial in many fields, from drug repurposing trials to evolutionary studies. Current state-of-the-art methods achieve good results on the benchmarking datasets. However, the current approaches operate over pairs of binding sites and are not applicable for searching databases of unprocessed protein structures. In cases when the binding sites are unknown, they have to be firstly located by using binding site prediction algorithms. That significantly increases the upfront costs of creating large databases of similar binding sites. This work covers the current methods for assessing binding site similarity and explores the possibility of fast searching of large databases of related structures by presenting a simple method that allows faster than linear search without the need for identification of the precise locations of the putative binding sites. The proposed approach shows promising preliminary results that merit further investigation, although more insight is still needed.