3D High-throughput comparison of protein binding sites and protein-ligand complexes requires a uniform and generic data representation. We herewith present IChem, a novel toolkit dedicated to the description, encoding and comparison of protein-ligand binding sites and protein-ligand complexes.

When applied to apo-proteins, it enables to automatically detect cavities and predict their ligandability using an embedded support vector machine trained on cavity physicochemical descriptors. Moreover, it allows a high-throughput alignment-dependent comparison of protein-binding sites and the prediction of potential off-targets.

IChem can also convert protein–ligand coordinates into a simple fingerprint (TIFP) of 210 integers registering the corresponding molecular interaction pattern. TIFP fingerprints have been calculated for ca. 10 000 druggable protein–ligand complexes therefore enabling a wide comparison of relationships between interaction pattern similarity and ligand or binding site pairwise similarity. In addition we developed two tools (Ishape, Grim) to align protein–ligand complexes from their interaction patterns. Ishape is based on the overlap of interaction pseudoatoms using a smooth Gaussian function, whereas Grim utilizes a standard clique detection algorithm to match interaction pattern graphs. Both tools are complementary and enable protein–ligand complex alignments capitalizing on both global and local pattern similarities. The new fingerprint and companion alignment tools have been successfully used in three scenarios: (i) interaction-biased alignment of protein–ligand complexes, (ii) postprocessing docking poses according to known interaction patterns for a particular target, and (iii) virtual screening for bioisosteric scaffolds sharing similar interaction patterns.