Creation of a Scaffold Hopping Database

Philippe Pinel[1,2], Stéphanie Labouille[1], Nicolas Drizard[1], Brice Hoffmann[1], Véronique Stoven[2]

1. Iktos SAS ; 2. MINES Paristech, CBIO

- An important issue in drug discovery is the ability to identify new molecules with optimized biological and chemical characteristics for a given protein target, potentially belonging to a different structural family than that of known hits.
- We created a Scaffold Hopping Database from PDB Bind in order to develop and evaluate drug discovery methods able to identify and solve scaffold hopping issues, in particular for targets with unknown 3D structures.

Interaction Database

Detection of protein-ligand interactions for each entry of the PDB Bind[1] database of protein-ligand complex structures.

PDBBind v2019
17,652 PDBs
→ Filtering
- Several PDBs for a protein
- Ligand Weight ≤ 900 g/mol
→ Interaction Database
12,826 PDBs
1,495 proteins

Considered Interactions

Protein-Ligand Interactions detected using an internal software inspired from PLIP package[2].

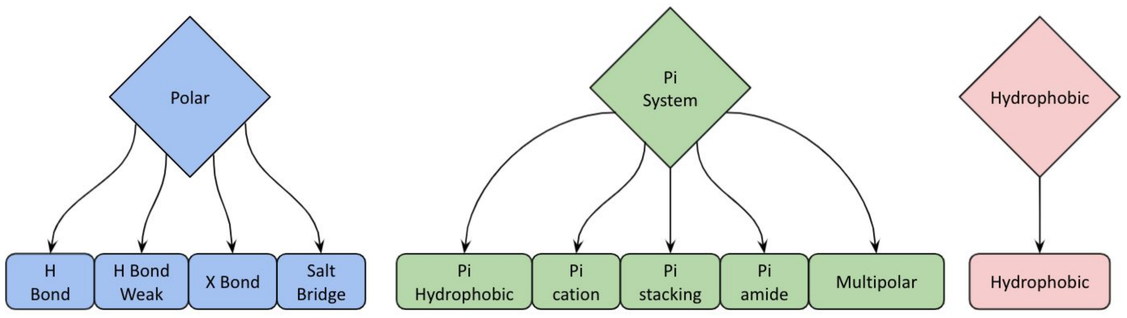
Interaction	Ligand	Protein
Hydrophobic	Hydrophobic atom	Hydrophobic atom
Pi-stacking	Aromatic ring	Aromatic ring
Pi-amide	Aromatic ring Amide	Amide Aromatic ring
Pi-cation	Aromatic ring Cation	Cation Aromatic ring
Pi-hydrophobic	Aromatic ring Hydrophobic atom	Hydrophobic atom Aromatic ring
Multipolar	Halogen Polar C	Polar C Halogen
H-bond	H-bond donor H-bond acceptor	H-bond acceptor H-bond donor
Weak H-bond	Weak H-bond donor Weak H-bond acceptor	Weak H-bond acceptor Weak H-bond donor
Salt bridge	Anion Cation	Cation Anion
Halogen bond	X-bond donor X-bond acceptor	X-bond acceptor X-bond donor

Interaction Fingerprints

Encode protein-ligand interactions:

- **Classical** IFP: 10 bits per residue (interactions) ;
- **Simple** IFP: 3 bits per residue (family of interactions) ;
- **Hierarchical** IFP: 13 bits per residue ;

Tanimoto similarity of interaction fingerprints computed to compare binding modes of 2 ligands of the same protein.

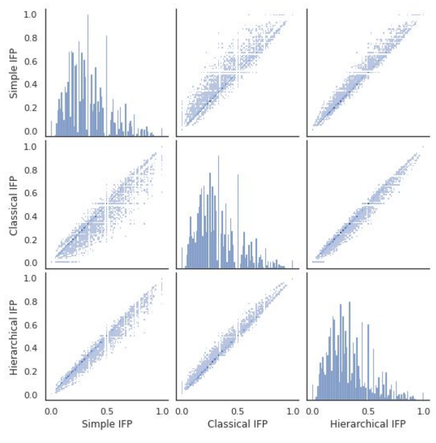


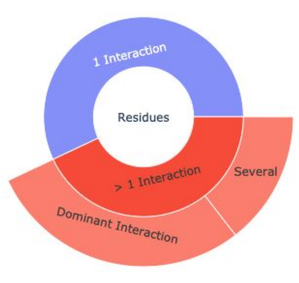
Correlation of Interaction Fingerprints

Interaction Tanimoto similarities between ligands according to previous IFPs are correlated. Encoding the specific interaction or its family is equivalent in this context.

Explained by the tendency of residues to create one type of interaction:

- **85%** of interacting residues have a **dominant interaction** (frequency > 0.5) ;
- **66%** of them only interact one way with the ligands ;





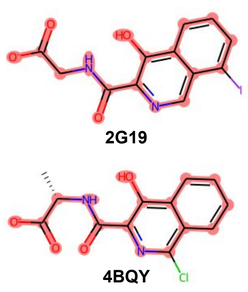
Scaffold Hopping

Corresponds to situations involving dissimilar ligands with similar binding modes within a specific target pocket. Identification:

- Low Morgan Similarity;
- High Classical IFP Similarity;

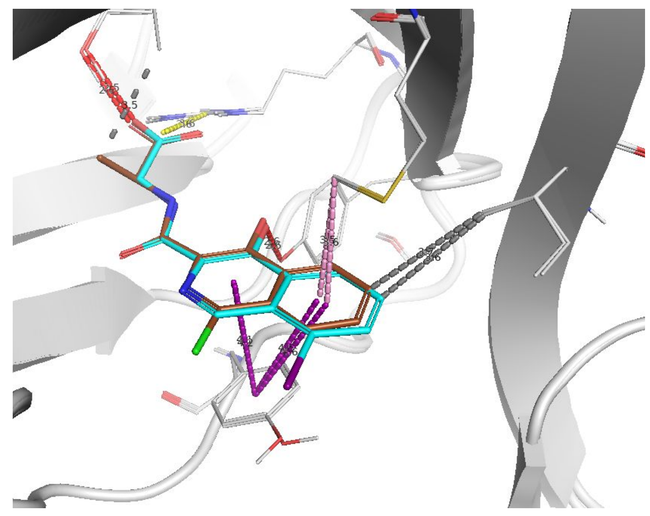
Issue: Binding interactions may involve a common substructure.

Example: Egl nine homolog 1



H bond
Salt Bridge
Pi Stacking
Pi Hydrophobic
Hydrophobic

Morgan Similarity	Classical IFP Similarity
0.19	0.85



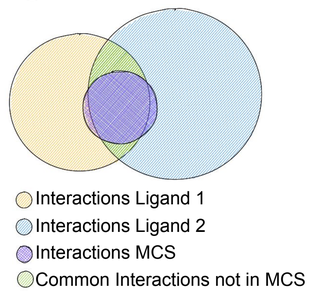
Maximum Common Substructure

Does the **MCS** between two ligands explain **similar binding modes** to avoid “false” scaffold hopping cases?

Comparison of ligands binding modes using MCS:

- Element exact match ;
- Bond type conserved ;
- Ring breaking allowed ;

$$Ratio_{MCS_{Interactions}} = \frac{|Interactions_{Common} \bigcap Interactions_{MCS}|}{|Interactions_{Common}|}$$



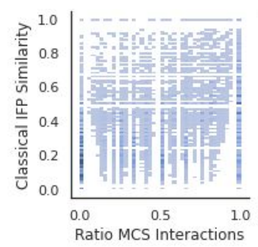
Scaffold Hopping Database

For each protein: pairwise comparison of its ligands interactions and detection of scaffold hopping situations.

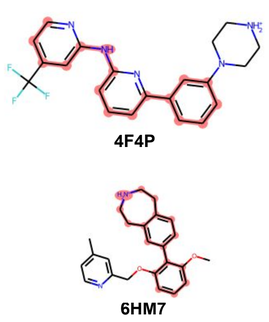
Interaction Database
→ Filtering
- Resolution ≤ 2.5 Å
- Interactions with 1 chain
- ≥ 6 Interactions
→ Ligands Comparison
91,867 ligand pairs
6,023 PDB
735 proteins
→ Scaffold Hopping Database
570 situations
59 proteins

Identification of Scaffold Hopping Situations:

- **Classical IFP Similarity** ≥ 0.7 ;
- **Ratio MCS Interactions** ≤ ⅓ ;
- **Morgan Similarity** ≤ 0.4 ;

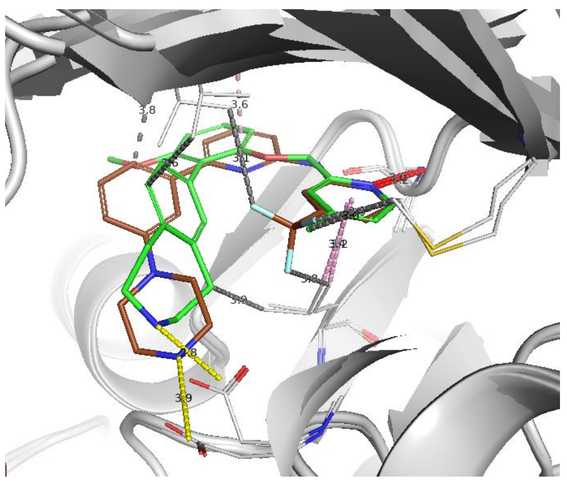


Example: Tyrosine-protein kinase SYK ; 5 scaffold hopping situations identified.



H bond
Salt Bridge
Pi Stacking
Pi Hydrophobic
Hydrophobic

Morgan Similarity	Classical IFP Similarity	Ratio MCS Interactions
0.11	0.75	0.0



References

[1] Wang R.; Fang X.; Lu Y.; Wang S. The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. J. Med. Chem. 2004, 47, 2977-2980

[2] Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: fully automated protein-ligand interaction profiler. Nucleic Acids Res. 2015 Jul 1;43(W1):W443-7

Contact

Philippe Pinel: philippe.pinel@iktos.com