IChem: An Open-Source Toolkit for Comparing Ligand Binding Modes and Protein Cavities

CS3 workshop 2 - 2026

Célien Jacquemard, Guillaume Bret, Didier Rognan and Esther Kellenberger

UMR 7200 – CNRS et Université de Strasbourg, Faculté de Pharmacie, 67400 Illkirch, France.

Structure-based drug discovery relies heavily on molecular docking to predict how candidate ligands engage a protein target. Despite decades of methodological progress, docking scoring functions remain unreliable: they correlate only loosely with experimental affinities, and even when a near-native pose is sampled, it is not systematically ranked first. The bottleneck of docking is therefore pose ranking, not pose sampling [1].

A complementary strategy exploits the rich content of the Protein Data Bank: rather than relying solely on a physics-based score, one can ask whether a docking pose reproduces an interaction pattern already observed in a known crystallographic complex. The underlying assumption is that ligands binding the same target with similar affinity tend to engage the same key residues through the same non-covalent interactions. This approach has been formalized through the concept of *interaction fingerprints* and *interaction graphs*, which encode a protein–ligand complex into compact, comparable representations [1, 2].

In this tutorial we present **IChem**, an open-source toolkit dedicated to the detection, comparison, and exploitation of protein–ligand interaction patterns [3]. Through two hands-on exercises built around a single biological case study (the μ-opioid receptor and its agonists), participants will learn how interaction-pattern similarity can recover near-native docking poses, and how it can be used to enrich a virtual screening campaign in true active compounds.

About IChem

IChem has been developed for more than fifteen years at the Laboratory of Therapeutic Innovation (LIT, UMR 7200 CNRS / University of Strasbourg) and applied to numerous academic and industrial drug discovery projects. It has recently been released under an open-source license, removing the main barrier to broader adoption and motivating the present tutorial.

The toolkit detects non-covalent interactions (hydrogen bonds, ionic bonds, aromatic stacking, hydrophobic contacts) between a protein and a ligand using geometric and chemical rules, and encodes them through two complementary representations (Figure 1):

- The **interaction fingerprint (IFP)** is a one-dimensional bit string in which residue is allocated a fixed block of bits encoding the interactions it engages with the ligand. Two complexes are compared via the Tanimoto coefficient (T_c) of their concatenated fingerprints. The IFP is fast, per-residue interpretable, but limited to comparisons across complexes sharing the same set of cavity residues.
- The **interaction graph** represents each interaction as a three-dimensional pseudo-atom labeled by pharmacophoric property. Two complexes are compared by graph matching (the **GRIM** algorithm), which identifies the largest common subgraph of consistent interactions. This representation is independent of cavity composition and can be applied across unrelated proteins.

In practice the two views are used jointly: the fingerprint captures *which residues* are engaged, while the graph captures *how* the interactions are spatially arranged.

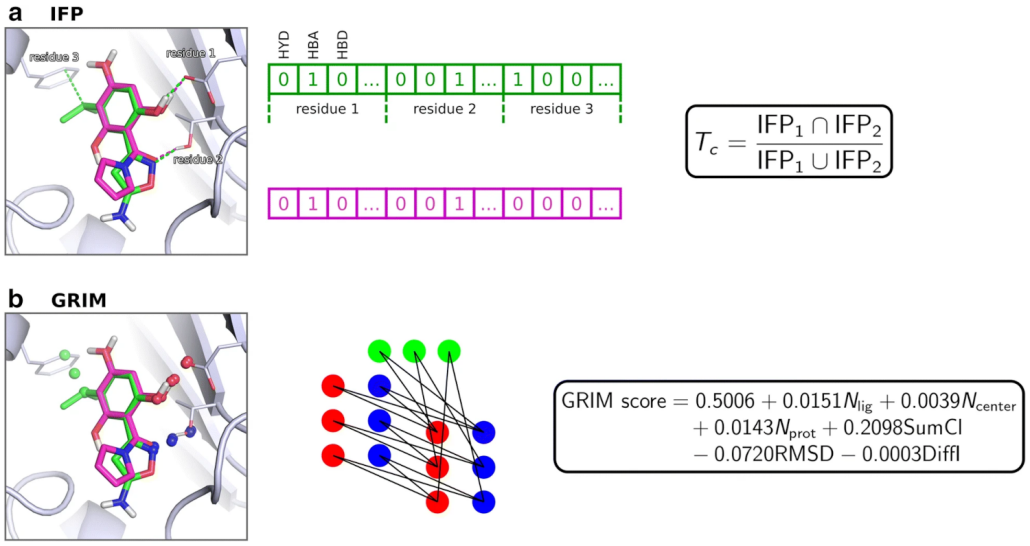

Figure 1: **Figure 1.** The two complementary representations of protein–ligand interactions used by IChem, compared between two complexes (shown in green and magenta on the structural panels). (**a**) The interaction fingerprint (IFP) encodes per-residue interactions as a 1D bit string, compared via the Tanimoto coefficient (T_c). (**b**) The interaction graph encodes each interaction as a 3D pseudo-atom; two complexes are compared by graph matching (GRIM score).

How to

CoCalc. CoCalc is a hosted scientific computing platform that exposes a Linux shell and a file browser from any web browser. A free account is sufficient for the present tutorial. By using CoCalc, no local installation is required regardless of the operating system. The workshop archive is provided directly inside the CoCalc project.

Local installation. Participants comfortable with a local terminal can run the tutorial offline on a Linux machine (Ubuntu, Debian, or Red Hat Enterprise Linux; x86_64 architecture). The workshop archive contains all required input data, pre-computed docking poses, and a self-contained Python environment.

Visualization. Detected interactions are visualized in the browser through standalone HTML viewers bundled with the workshop archive. No additional software is required.

IChem source code and documentation. IChem can be installed locally or rebuilt from the public repository at https://github.com/LIT-CCM-lab/IChem.

Overview of the tutorial

The tutorial is built around the **μ-opioid receptor** (μOR), a major pharmacological target with several recent crystallographic complexes covering chemically diverse agonists. It is organized in two hands-on parts of increasing complexity.

Part 1: Pose prediction on μ-opioid agonists

Eight crystallographic complexes of the activated μ-opioid receptor are used as references. Participants first compare the binding modes of these references pairwise, using both IFP/Tanimoto and IPA/GRIM scores, in order to assess the diversity and the conserved features of opioid binding. They

then assess whether IFP- and GRIM-based rescoring can recover near-native poses among 30 candidate docking solutions, both in a self-docking scenario (the ligand is re-docked into its own receptor) and in a more challenging cross-docking scenario (the ligand is docked into a non-cognate receptor). The expected take-home observation is that interaction-pattern similarity often selects a near-native pose more reliably than the native docking score, provided that an appropriate reference complex is available.

Part 2: Retrospective virtual screening

A small library of 100 compounds (10 known μ-opioid agonists retrieved from ChEMBL and 90 property-matched decoys generated with LUDE) has been docked in advance into six μOR receptor structures, producing 18,000 poses. Participants focus on the analysis. They first compare the native docking score (ChemPLP) and the standard TcIFP on classical virtual screening metrics (ROC AUC, enrichment factors at 1%, 5%, 10%, precision and recall at top-10). A perhaps unexpected pattern emerges: TcIFP outperforms ChemPLP on early enrichment but underperforms it on global AUC. Inspection of the top-ranked false positives reveals that the offending decoys lack a positively ionizable group, whereas all known μ-opioid agonists form a conserved salt bridge with **Asp147$^{3.32}$** [4, 5]. A refined scoring scheme (TcIFP_v2) is then introduced, combining a hard pharmacophore filter on Asp147 with a balanced contribution of hydrophobic and polar similarities, and the analysis is repeated. The exercise illustrates how an interpretable scoring scheme can be diagnosed and refined iteratively — a practical advantage of classical interaction fingerprints over opaque scoring functions.

References

[1] Marcou G., Rognan D. *Journal of Chemical Information and Modeling* **2007**, 47, 195–207 [doi: 10.1021/ci600342e]

[2] Desaphy J., Raimbaud E., Ducrot P., Rognan D. *Journal of Chemical Information and Modeling* **2013**, 53, 623–637 [doi: 10.1021/ci300566n]

[3] Da Silva F., Desaphy J., Rognan D. *ChemMedChem* **2018**, 13, 507–510 [doi: 10.1002/cmdc.201700505]

[4] Manglik A. et al. *Nature* **2012**, 485, 321–326 [doi: 10.1038/nature10954]

[5] Huang W. et al. *Nature* **2015**, 524, 315–321 [doi: 10.1038/nature14886]