Journal of Cheminformatics

SOFTWARE

Open Access

ProLIF: a library to encode molecular interactions as fingerprints

Cédric Bouysset* and Sébastien Fiorucci*

Abstract

Interaction fingerprints are vector representations that summarize the three-dimensional nature of interactions in molecular complexes, typically formed between a protein and a ligand. This kind of encoding has found many applications in drug-discovery projects, from structure-based virtual-screening to machine-learning. Here, we present ProLIF, a Python library designed to generate interaction fingerprints for molecular complexes extracted from molecular dynamics trajectories, experimental structures, and docking simulations. It can handle complexes formed of any combination of ligand, protein, DNA, or RNA molecules. The available interaction types can be fully reparametrized or extended by user-defined ones. Several tutorials that cover typical use-case scenarios are available, and the documentation is accompanied with code snippets showcasing the integration with other data-analysis libraries for a more seamless user-experience. The library can be freely installed from our GitHub repository (https://github.com/chemosim-lab/ProLIF).

Keywords: Interaction fingerprint, Structural biology, Molecular dynamics, Docking, Virtual screening, Python

Introduction

Interactions between and within molecular structures are the driving force behind biological processes, from protein folding to molecular recognition. The decomposition of interactions by residues in biomolecular complexes can provide insights into structure–function relationships, and characterizing the nature of each of these interactions can guide medicinal chemists in structure-based drug discovery projects [1]. Approaches to encode the interactions observed in 3D structural data in the form of a binary fingerprint have been developed in the past [2–6] and applied successfully to a variety of projects. For example, de Graaf et al. [7] used the Tanimoto similarity between the interaction fingerprint (IFP) of a crystallographic reference and the IFP of docking poses to rescore virtual screening results on a G protein-coupled receptor (GPCR). Rodríguez-Pérez et al. [8] showed that IFPs can achieve superior predictive performance than ligand fingerprints (ECFP4) for the classification of kinase inhibitor binding modes with machine-learning models. Finally, Mpamhanga et al. [9] showed that one can use the IFP for clustering, and then shortlist a reasonable number of binding modes prior to visual inspection. More recently, the approach was also implemented for molecular dynamics (MD) simulations to study ligand unbinding [10]. While the typical IFP usually encodes pre-established interactions (hydrogen bond, π-stacking…etc.) on a per-residue basis, other implementations exist. Sato et al. [11] developed a pharmacophore-based IFP which relies on the pharmacophoric features of the ligand atoms in contact with the protein and the distance between each of these pharmacophores to generate a bitvector. Da et al. [12] developed an IFP that relies on the atomic environment of both the protein and ligand interacting atoms to set the positions of a bit in the fingerprint, rather than relying on protein residues and predefined interactions, which has the advantage of implicitly encoding every possible type of interaction. This protocol was later reimplemented in

*Correspondence: cedric.bouysset@univ-cotedazur.fr; sebastien.fiorucci@univ-cotedazur.fr
Institut de Chimie de Nice UMR7272, Université Côte d'Azur, CNRS, Nice, France

Python by Wójcikowski et al. [13], but other more classical Python-based IFP implementations exist [14–19]. In this paper, we introduce a new Python library, ProLIF, that overcomes several limitations encountered by these programs, namely working exclusively with the output of specific docking programs, not being compatible with the analysis of MD trajectories, being restricted to a specific kind of complex (usually protein–ligand complexes), depending on residue or atom type naming conventions, or not being extensible or configurable regarding interactions (Table 1).

Implementation

ProLIF can deal with RDKit [21] molecules or MDAnalysis [22] Universe objects as input, which allows supporting most 3D molecular formats, from docking to MD simulations. While most MD topology files do not keep explicit information about bond orders and formal charges, MDAnalysis is able to infer this information if all hydrogen atoms are explicit in the structure while converting the structure to an RDKit molecule. The RDKit parent molecule is then automatically fragmented in child residue molecules based on residues name, number, and chain to make it easier to work on a per-residue basis when encoding the interactions.

When calculating an interaction fingerprint, each interaction is typically defined as two groups of atoms that satisfy geometrical constraints based on distances and/or angles (Table 2). Here the selection of atoms is made using SMARTS queries (Table 3), which is more precise than relying on elements or atomic weights and is also more universal than relying on force-field-specific atom types.

The library is designed so that users can easily modify existing interactions, as there is usually no consensus on the empirical thresholds (distance, angles) that should be used. For example, the hydrogen bond DH…A can be defined as a distance between H and A lower or equal to 3.0 Å [9] or as a distance between D and A lower or equal to 3.5 [4, 14, 20] or 4.1 Å [15, 23], and the angles constraints can also vary. ProLIF is also designed to let users define custom interactions.

Each interaction is written as a Python class that implements a "detect" method which takes two RDKit molecules as input, typically a ligand and a protein residue, and outputs a Boolean (True if the interaction is present, else False) as well as the indices of atoms responsible for the interaction. All interaction classes are then gathered inside a "Fingerprint" class that can generate a bitvector from two RDKit molecules, and optionally return the atom indices. By default, the Fingerprint class is configured to generate a bitvector with the following interactions: hydrophobic, π-stacking, π-cation and cation-π, anionic and cationic, and H-bond donor and acceptor, although more specific interactions are available (see Table 2). This Fingerprint class is designed with two scenarios in mind, post-processing MD trajectories or docking results, thus it provides user-friendly functions to generate the complete array of interactions for each pair of interacting residues.

Finally, the interaction is stored inside the Fingerprint class as a mapping between a pair of "ligand" and "protein" residues, and the corresponding interaction bitvector. For easier post-processing, the interaction fingerprint can then be converted to a pandas DataFrame object [24], which facilitates the search for specific interactions and the aggregation of results.

Table 1 Comparative table of features available in non-commercial IFP software

Software	Complexes	Input format	MD	Configurable	Extensible	Web	CLI	Ref.
ProLIF	All	MDAnalysis and RDKit[c]	✓	✓	✓			
IChem	LP	MOL2		✓			✓	[4]
PLIP	All[a]	PDB		✓		✓	✓	[15, 18]
Arpeggio	All[a]	PDB		✓		✓	✓	[16]
PyPLIF HIPPOS	LP[a]	PDBQT, MOL2		✓			✓	[17]
getContacts	All[a,b]	VMD[c]		✓			✓	[19]
MD-IFP	LP[a]	MDAnalysis[c]	✓					[10]
ODDT	LP	OpenBabel and RDKit[c]		✓				[20]

MD native support of MD trajectories; *Web* available through a webserver; *CLI* command-line interface; *LP* Ligand–Protein; *All* all combinations between ligand, protein, DNA, and RNA molecules

[a] Relies on residue and atom naming convention to assign charges and/or aromaticity

[b] Only supports hydrogen-bond, hydrophobic and van der Waals contacts for LP complexes

[c] Compatible with the input formats supported by the aforementioned libraries

Table 2 Interactions currently available in ProLIF

Interaction	Ligand[a]	Protein[a]	Distance (Å)	Angle (deg)
Anionic	Anion	Cation	≤ 4.5	
Cationic	Cation	Anion		
CationPi	Cation	Aromatic	(+)-ctd $\leq$ 4.5	$\langle \vec{n}, \overrightarrow{ctd \cdots (+)} \rangle \in [0, 30]$
PiCation	Aromatic	Cation		
PiStacking	Aromatic	Aromatic	ctd–ctd $\leq$ 6.0 min $\leq$ 3.8	$\langle \vec{n}, \vec{n} \rangle \in [0, 90]$
EdgeToFace	Aromatic	Aromatic	ctd–ctd $\leq$ 6.0 Min $\leq$ 3.8	$\langle \vec{n}, \vec{n} \rangle \in [50, 90]$
FaceToFace	Aromatic	Aromatic	ctd–ctd $\leq$ 4.5 min $\leq$ 3.8	$\langle \vec{n}, \vec{n} \rangle \in [0, 40]$
HBAcceptor	HBAcceptor	HBDonor	D–A $\leq$ 3.5	$\langle \overrightarrow{HD}, \overrightarrow{HA} \rangle \in [130, 180]$
HBDonor	HBDonor	HBAcceptor		
XBAcceptor	XBAcceptor	XBDonor	X–A $\leq$ 3.5	$\langle \overrightarrow{XD}, \overrightarrow{XA} \rangle \in [130, 180]$
XBDonor	XBDonor	XBAcceptor		$\langle \overrightarrow{AX}, \overrightarrow{AR} \rangle \in [80, 140]$
MetalAcceptor	Ligand	Metal	≤ 2.8	
MetalDonor	Metal	Ligand		
Hydrophobic	Hydrophobic	Hydrophobic	≤ 4.5	

(−) anion; *(+)* cation; *ctd* centroid of the aromatic ring; *min* minimum value in the distance matrix between both aromatic rings; *n* normal to the aromatic ring plane; *D* hydrogen/halogen bond donor; *A* hydrogen/halogen bond acceptor; *H* hydrogen atom; *X* halogen atom; *R* atom linked to a halogen bond acceptor

[a] Although "ligand" and "protein" are used here, all the listed interactions can be applied to any molecular complex (protein–protein, DNA–protein…etc.)

Table 3 SMARTS patterns used in the definition of interactions

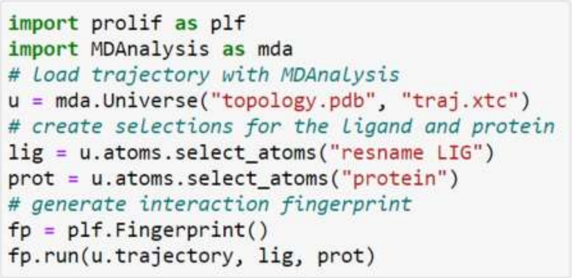
Name	SMARTS pattern(s)
Anion	[−{1−}]
Cation	[+{1−}]
Aromatic	a1:a:a:a:a:a:1 a1:a:a:a:a:1
HBAcceptor	[N,O,F,−{1−};!+{1−}]
HBDonor	[#7,#8,#16][H]
XBAcceptor	[#7,#8,P,S,Se,Te,a;!+{1−}][*]
XBDonor	[#6,#7,Si,F,Cl,Br,I]-[Cl,Br,I,At]
Metal	[Ca,Cd,Co,Cu,Fe,Mg,Mn,Ni,Zn]
Ligand	[O,N,−{1−};!+{1−}]
Hydrophobic	[#6,#16,F,Cl,Br,I,At;+0]

Results and discussion

By relying on the interoperability with popular open-source libraries (MDAnalysis and RDKit), it can support a wide range of molecular formats typically found in docking experiments and MD simulations. Because it directly relies on SMARTS patterns to define the chemical moieties that partake in interactions, it is also compatible with any kind of molecular complex, including complexes made of ligands, proteins, DNA or RNA molecules. Interoperability also allows for data analysis to be substantially easier: as mentioned in the Implementation section the IFP can be directly exported to a pandas DataFrame (one of Python's most popular data analysis library), and the documentation contains tutorials on how to visualize the interactions as graphs or how to display them on the 3D structure of the complex.

Analysis of an MD trajectory of a GPCR in complex with a ligand

The code to run ProLIF on an MD trajectory can be as simple as follow:



```python
import prolif as plf
import MDAnalysis as mda
# load trajectory with MDAnalysis
u = mda.Universe("topology.pdb", "traj.xtc")
# create selections for the ligand and protein
lig = u.atoms.select_atoms("resname LIG")
prot = u.atoms.select_atoms("protein")
# generate interaction fingerprint
fp = plf.Fingerprint()
fp.run(u.trajectory, lig, prot)
```

Here, we showcase an analysis based on the fingerprint obtained from a 500 ns MD simulation of the 5-HT1B receptor (class A aminergic GPCR) in complex with ergotamine retrieved from the GPCRmd webserver (id 90) [25]. In class A GPCRs, each position is annotated in superscript notation according to the Ballesteros-Weinstein numbering scheme [26], a generic residue numbering denoting both the helix and position relative to the most conserved residue labelled as number 50.

Exporting the fingerprint to a DataFrame allows to easily address common questions like which residues are involved in a specific type of interaction, which interactions does a specific residue do, which are the most frequent types of interactions, or which are the residues most frequently interacting with the ligand. In this MD trajectory, there is constantly at least one hydrophobic, H-bond donor and cationic interaction, while H-bond acceptor and π-stacking interactions occur respectively in 92% and 85% of the analyzed frames (see analysis notebook in supplementary data). F331$^{6.52}$ is responsible for half of the π-stacking interactions occurring during the simulation, and the ten residues that interact with the ligand the most frequently are (in descending order): D129$^{3.32}$, I130$^{3.33}$, F330$^{6.51}$, V201$^{ECL2.52}$, F331$^{6.52}$, S212$^{5.42}$, W327$^{6.48}$, V200$^{ECL2.51}$, C133$^{3.36}$ and F351$^{7.35}$ which are all in contact with the ligand in at least 97% of frames. This is in agreement with the known interactions available from experimental structures as listed on the GPCRdb webpage [27] for the human 5-HT1B receptor, except for S212$^{5.42}$ which isn't reported to make H-bond interactions with ligands. The difference is likely due to the fact that this analysis is based on an MD trajectory while GPCRdb gathers interactions from experimental structures. However, GPCRdb also lists mutational data for S212$^{5.42}$, and mutating this position to an alanine does not affect the binding affinity to ergotamine [28] which could coincide with the MD simulation since the ligand makes a hydrogen bond with the backbone and not the sidechain. Mutating S212$^{5.43}$ to a bulkier residue could potentially affect this interaction and decrease the binding affinity.

Because ProLIF keeps track of the atom indices responsible for interactions, it is possible to display detailed 2D or 3D interaction plots. Examples of scripts to generate such plots are given in the documentation. An exception is made for the ligand interaction network diagram which has been directly included in the source code of ProLIF under the LigNetwork class. This LigNetwork diagram (Fig. 1) is interactive and allows repositioning the residues but also hiding specific residue types or interactions by clicking the legend. It can show the interaction diagram at a precise frame or aggregate the results and only display interactions that appear frequently, controlled by a frequency threshold. In the latter case, to keep the plot readable for each ligand–protein-interaction group only the most frequent ligand atom is shown, as it might differ between frames.

The fingerprint can also be converted to an RDKit bitvector to make use of the similarity/distance metric functions implemented. This allows to investigate the presence of different binding modes in the simulation. In Fig. 2, we show the Tanimoto similarity matrix between each interaction fingerprint during the MD simulation. Two clusters are visible (from frame 400 to 1400, and from frame 1400 to 2100) which reveals changes in the interactions between ergotamine and 5-HT1B. Indeed, in the second cluster the phenyl ring of ergotamine gets closer to the indole moiety, which disrupts hydrophobic contacts with W125$^{3.28}$, H-bonding with S212$^{5.42}$ and π-stacking with F351$^{7.35}$ to create new hydrophobic interactions with T203^{ECL2}, T209$^{5.39}$, S334$^{6.55}$ and D352$^{7.36}$.

Analyzing protein–protein interactions (PPI)

The analysis of intra- and inter-molecular interactions can also be applied to investigate protein dynamics and function with ProLIF. Because ProLIF requires explicit hydrogen atoms, we preprocess PDB files of X-ray structures in the current section with the PDB2PQR [29] webserver as follows: AMBER force-field and naming scheme, protonation states assignments with PROPKA at pH 7.0, H-bond network optimization and removal of water molecules.

In this first example, we focus on the activation mechanism of a class A GPCR and show how ProLIF can help pinpoint intramolecular structural modifications upon receptor activation. GPCRs are membrane-embedded receptors arranged in seven helical transmembrane domains (labelled TM1 to TM7) followed by a shorter helix (H8) that lies at the interface between the membrane and the cytosol. This family shares conserved key motifs in each TM domain, and some of the motifs are part of molecular switches that mediate ligand binding or receptor activation. Among them, the DRY motif in TM3 and the NPxxY motif in TM7 have been reported to be part of the allosteric mechanism [30]. Briefly, upon ligand binding, the signal propagates from the binding pocket to the ionic lock (comprised of the DRY motif) through a network of hydrophobic residues. The ionic lock maintaining the receptor in its inactive form is disrupted, leading to an increase of the inter-helix distances (notably TM3-TM6). At the same time, the hydrophobic barrier cannot prevent anymore the flooding of the intracellular part of the receptor thereby creating an intracellular crevice required for G protein coupling. R$^{3.50}$ of the DRY motif is known to stabilize the inactive form of the rhodopsin receptor through a salt-bridge with D$^{6.30}$

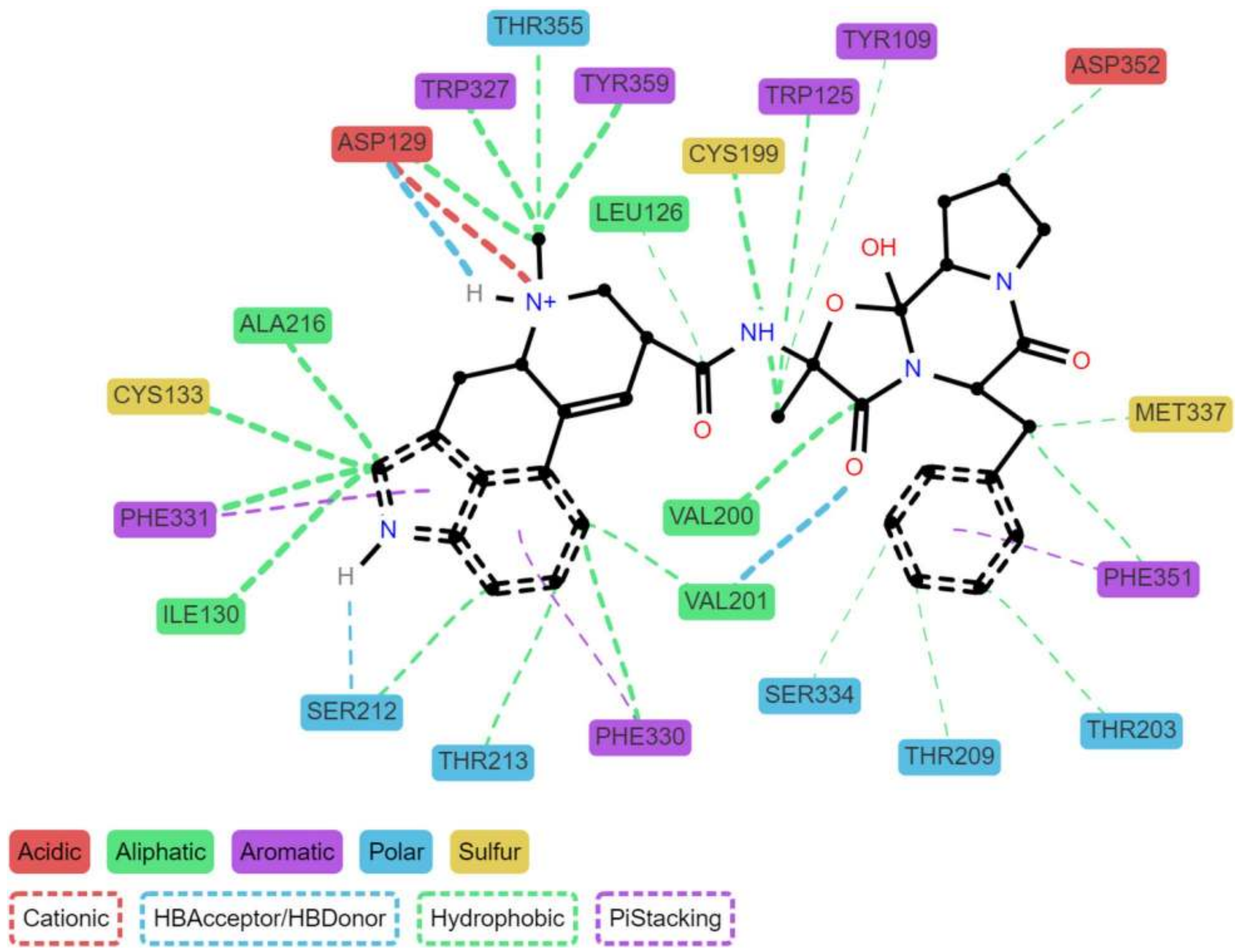


Fig. 1 Ligand interaction network for the ergotamine agonist bound to the 5-HT1B receptor. Each interaction is shown as a dashed line between the residue and the ligand, and the width of the line is linked to the frequency of the interaction in the simulation. Only interactions occurring in at least 30% of frames are shown here

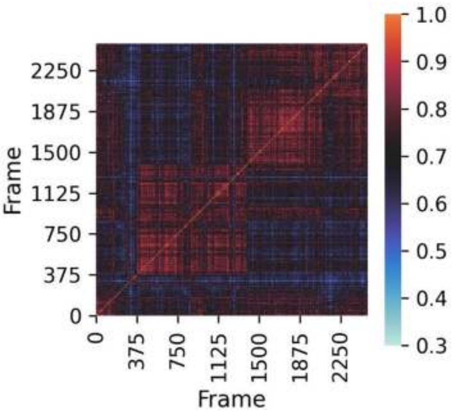


Fig. 2 Tanimoto similarity matrix of ligand–protein interactions between each frame of the MD trajectory

known as the “ionic-lock” [30]. This position can also interact with $Y^{5.58}$ through an H-bond, and is reported to be critical for the formation of the active state in the β2 adrenergic receptor [31]. For the NPxxY motif, the mutation of $Y^{7.53}$ disrupts interactions with $N^{2.40}$ in the β2 adrenergic receptor [32], and $Y^{7.53}$ is also reported to have an aromatic interaction with $F^{8.50}$ which stabilizes the inactive conformation of the rhodopsin receptor [33]. As an example, the residue interaction network of the bovine rhodopsin in both active (PDB 6FK6) and inactive (PDB 1U19) states is studied to reveal the structural changes involving these two motifs. As seen in Fig. 3, the ionic lock between $R135^{3.50}$ and $E247^{6.30}$ is only visible in the inactive form of the receptor, while the interaction between $R135^{3.50}$ and $Y223^{5.58}$ was only detected in the active form. $Y306^{7.53}$, which is part of the NPxxY motif in TM7, takes part in both key interactions that stabilize the inactive form of the receptor previously described: an

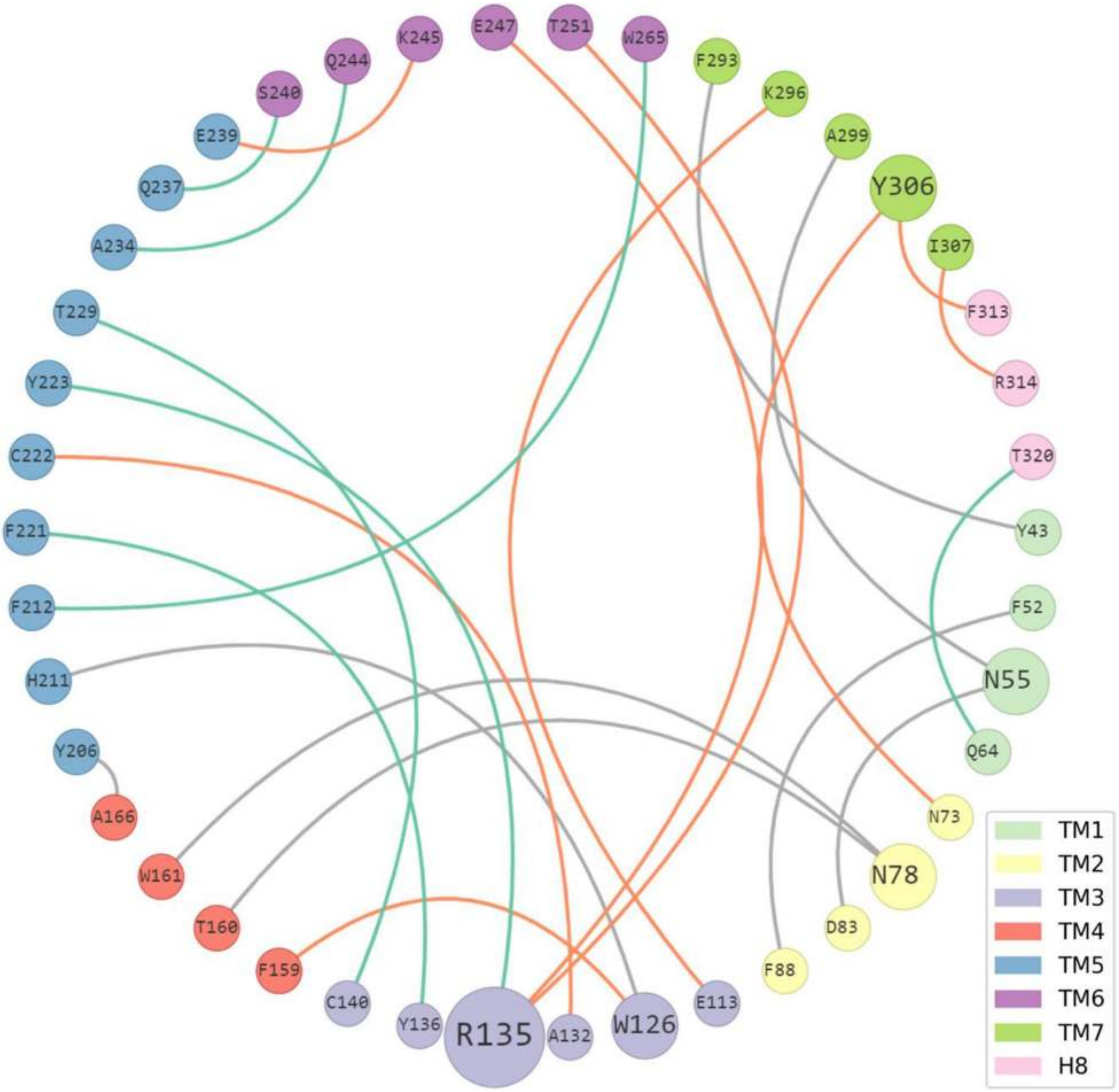


Fig. 3 Residue interaction network for the bovine rhodopsin. Residues are colored by transmembrane domain (TM). Interactions that only appear in the active (PDB 6FK6) or inactive (PDB 1U19) state of the receptor are respectively shown in green or orange, and the ones that appear in both are in grey. Each residue node is scaled based on its number of interactions. For clarity, interactions that occur within the same TM (as labelled by GPCRdb) and interactions between residues that are less than 3 residues apart are not shown, as well as hydrophobic interactions (as defined in the implementation) and residues that did not participate in any interaction

H-bond interaction with N73$^{2.40}$ and a π-stacking interaction with F313$^{8.50}$. Finally, in rhodopsin, the salt-bridge between K296$^{7.43}$ and E113$^{3.28}$ is known to be crucial in the activation cycle of the receptor and is only disrupted when K296$^{7.43}$ transiently bounds to retinal [34], which is in agreement with the interactions reported here.

The final step in GPCR signal transduction being an intermolecular process between the GPCR and a G-protein, ProLIF can also be used in this case to highlight positions that dictate the coupling specificity in a series of GPCR-G-protein complexes. Here, we reproduce the analysis of interactions between the β2 adrenoceptor and the Gαs/Gβ1 complex by Flock et al. [35] where the authors used a "van der Waals contact" interaction based on Venkatakrishnan et al. [36] which considers two residues as interacting if any interatomic distance is below or equal to their van der Waals interaction distance (the sum of their van der Waals radii plus a tolerance factor of 0.6 Å). We reimplemented this in ProLIF (see analysis notebook in supplementary data) and applied it to the same structure (PDB 3SN6) to obtain the PPI network shown in Fig. 4.

The interaction network remains mostly the same as with Figure S6 of the original study [35] and highlights the importance of positions I135$^{3.54}$, P138$^{34.50}$, F139$^{34.51}$, Q229$^{5.68}$, R239^{ICL3} and T274$^{6.36}$ for GPCR-G protein coupling. Using the default ProLIF implementation would help clarifying the types of interactions involved (H-bond, ionic…etc.) for a better understanding of coupling specificity when several GPCR-G protein complexes are investigated.

Conclusions

ProLIF is a new Python library that overcomes limitations encountered by other freely available IFP programs. One of the main differences is the support of MD trajectories, while still being compatible with other molecular structure files like docking and experimental structures. By design, it is also not restricted to a particular kind of molecular complex but supports any combination of ligand, protein, DNA, or RNA molecules, thanks to its absence of dependency to force-field specificities such as atom types or residue naming convention. It also has a user-friendly API, comes with several tutorials, and allows creating custom interactions or reconfiguring existing ones. Finally, it focuses on the integration with typical data-analysis packages and visualization tools for a seamless user experience within the Python ecosystem. Possible improvements include the addition of more interactions types, but also more types of fingerprints such as the pharmacophoric [11] or circular [12,

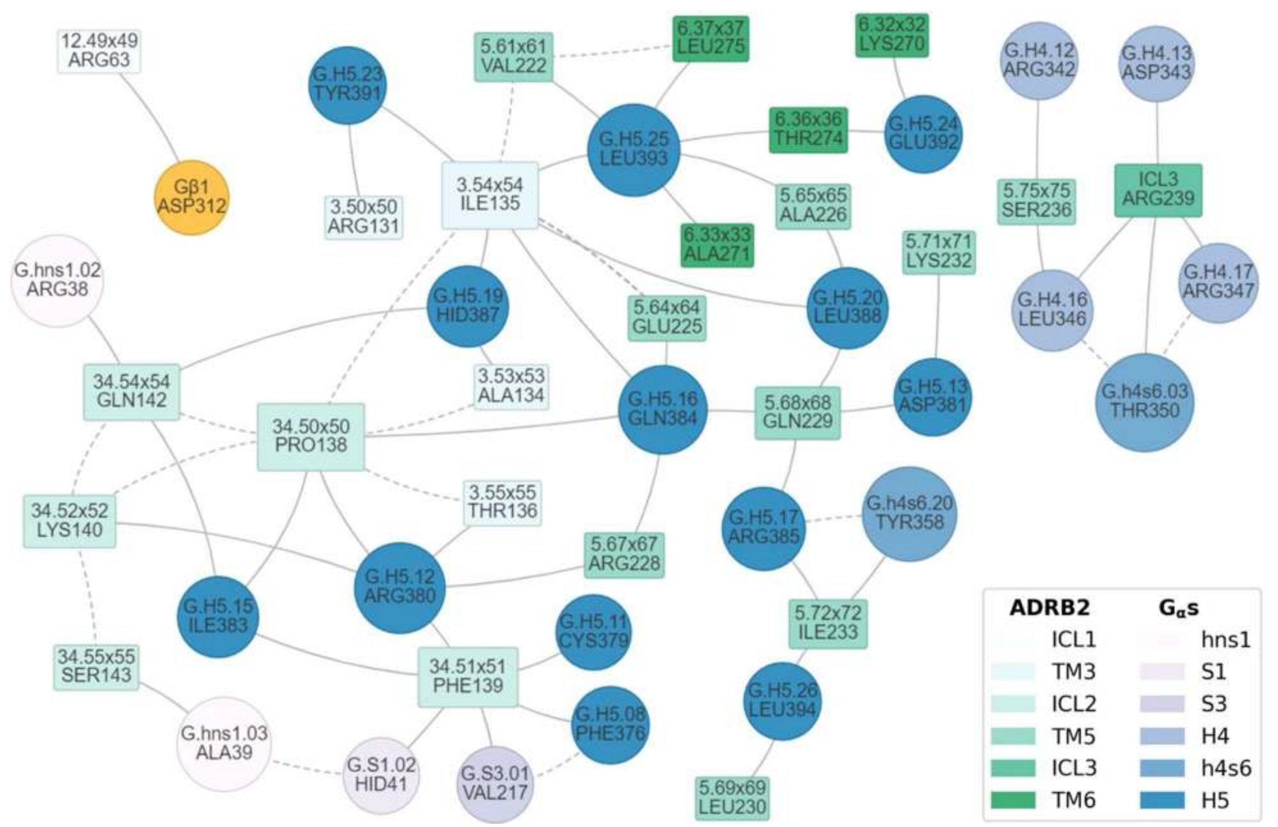


Fig. 4 Interaction network between the β2 adrenoceptor (ADRB2) and G protein complex (Gαs and Gβ1). ADRB2 residues are shown as rectangles in shades of green, and G protein residues are shown as ellipses in shades of blue for Gαs and in yellow for Gβ1. For ADRB2, ICL denotes the intracellular loops while TM corresponds to the transmembrane domains. For Gαs, the common Gα numbering (CGN) system is used [37]. Each node is scaled by its number of interactions. Inter and intra protein interactions are respectively shown as plain and dashed lines. Residues that do not participate in GPCR-G protein interactions are not shown, and interactions between covalently bonded residues or residues of the same helix (as labelled by GPCRdb) are hidden

13] fingerprints. Adding a command-line interface would also extend the userbase to researchers inexperienced in Python. Another point of interest could be the extension to other popular visualization libraries for a more streamlined data analysis experience for users.

Availability and requirements

Project name: ProLIF.

Project home page: https://github.com/chemosim-lab/ProLIF

Operating system(s): platform independent.

Programming language: Python.

Other requirements: Python 3.6 or higher, and several open-source Python packages listed in the project's documentation.

License: Apache License 2.0

Any restrictions to use by non-academics: none.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-021-00548-6.

Additional file 1. Jupyter notebook (html export) containing the analysis detailed in the manuscript. The code and the dataset are also available in the in the GitHub repository, https://github.com/chemosim-lab/ProLIF-paper, or through the Zenodo archive, https://doi.org/10.5281/zenodo.4945869.

Acknowledgements

The authors thank Jody Pacalon and Matej Hladiš for their helpful comments on the manuscript. The authors would also like to thank users who reported issues with the code and documentation which ultimately led to improvements

Authors' contributions

CB designed the software and analyzed the data. SF managed the study. CB and SF interpreted the data and wrote the paper. All authors read and approved the final manuscript.

Authors' information

CB received a stipend from Google for contributing code to MDAnalysis as part of a Google Summer of Code program. The contributed code, which allows for interoperability between MDAnalysis and RDKit, is part of the dependencies that the software described in this manuscript relies on.

Funding

This work was funded by the French Ministry of Higher Education and Research (PhD Fellowship to CB), by GIRACT (Geneva, Switzerland) (9th European PhD in Flavor Research Bursaries for first year students to CB) and the Gen Foundation (Registered UK Charity No. 1071026) (a charitable trust which principally provides grants to students/researchers in natural sciences, in particular food sciences/technology to CB).

Availability of data and materials

The datasets and code supporting the conclusions of this article are available in the supplementary materials and in the GitHub repository, https://github.com/chemosim-lab/ProLIF-paper, or through the Zenodo archive, https://doi.org/10.5281/zenodo.4945869.

Declarations

Competing interests

The authors declare no competing financial interests.

Received: 15 June 2021 Accepted: 30 August 2021
Published online: 25 September 2021

References

1. Fischer A, Smieško M, Sellner M, Lill MA (2021) Decision making in structure-based drug discovery: visual inspection of docking results. J Med Chem 64:2489–2500. https://doi.org/10.1021/acs.jmedchem.0c02227
2. Deng Z, Chuaqui C, Singh J (2004) Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. J Med Chem 47:337–344. https://doi.org/10.1021/jm030331x
3. Kelly MD, Mancera RL (2004) Expanded interaction fingerprint method for analyzing ligand binding modes in docking and structure-based drug design. J Chem Inf Comput Sci 44:1942–1951. https://doi.org/10.1021/ci049870g
4. Marcou G, Rognan D (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model 47:195–207. https://doi.org/10.1021/ci600342e
5. Perez-Nueno VI, Rabal O, Borrell JI, Teixido J (2009) APIF: a new interaction fingerprint based on atom pairs and its application to virtual screening. J Chem Inf Model 49:1245–1260. https://doi.org/10.1021/ci900043r
6. Jasper JB, Humbeck L, Brinkjost T, Koch O (2018) A novel interaction fingerprint derived from per atom score contributions: exhaustive evaluation of interaction fingerprint performance in docking based virtual screening. J Cheminform 10:1–13. https://doi.org/10.1186/s13321-018-0264-0
7. de Graaf C, Kooistra AJ, Vischer HF et al (2011) Crystal structure-based virtual screening for fragment-like ligands of the human histamine H_1 receptor. J Med Chem 54:8195–8206. https://doi.org/10.1021/jm2011589
8. Rodríguez-Pérez R, Miljković F, Bajorath J (2020) Assessing the information content of structural and protein–ligand interaction representations for the classification of kinase inhibitor binding modes via machine learning and active learning. J Cheminform 12:36. https://doi.org/10.1186/s13321-020-00434-7
9. Mpamhanga CP, Chen B, McLay IM, Willett P (2006) Knowledge-based interaction fingerprint scoring: a simple method for improving the effectiveness of fast scoring functions. J Chem Inf Model 46:686–698. https://doi.org/10.1021/ci050420d
10. Kokh DB, Doser B, Richter S et al (2020) A workflow for exploring ligand dissociation from a macromolecule: efficient random acceleration molecular dynamics simulation and interaction fingerprint analysis of ligand trajectories. J Chem Phys. https://doi.org/10.1063/5.0019088
11. Sato T, Honma T, Yokoyama S (2010) Combining machine learning and pharmacophore-based interaction fingerprint for in silico screening. J Chem Inf Model 50:170–185. https://doi.org/10.1021/ci900382e
12. Da C, Kireev D (2014) Structural protein-ligand interaction fingerprints (SPLIF) for structure-based virtual screening: method and benchmark study. J Chem Inf Model 54:2555–2561. https://doi.org/10.1021/ci500319f
13. Wójcikowski M, Kukiełka M, Stepniewska-Dziubinska MM, Siedlecki P (2019) Development of a protein–ligand extended connectivity (PLEC) fingerprint and its application for binding affinity predictions. Bioinformatics 35:1334–1341. https://doi.org/10.1093/bioinformatics/bty757
14. Radifar M, Yuniarti N, Istyastono EP (2013) PyPLIF: python-based protein-ligand interaction fingerprinting. Bioinformation 9:325–328. https://doi.org/10.6026/97320630009325
15. Salentin S, Schreiber S, Haupt VJ et al (2015) PLIP: fully automated protein–ligand interaction profiler. Nucleic Acids Res 43:W443–W447. https://doi.org/10.1093/nar/gkv315
16. Jubb HC, Higueruelo AP, Ochoa-Montaño B et al (2017) Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. J Mol Biol 429:365–371. https://doi.org/10.1016/j.jmb.2016.12.004

17. Istyastono EP, Radifar M, Yuniarti N et al (2020) PyPLIF HIPPOS: a molecular interaction fingerprinting tool for docking results of AutoDock Vina and PLANTS. J Chem Inf Model 60:3697–3702. https://doi.org/10.1021/acs.jcim.0c00305
18. Adasme MF, Linnemann KL, Bolz SN et al (2021) PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. Nucleic Acids Res gkab294. https://doi.org/10.1093/nar/gkab294
19. Venkatakrishnan AJ, Fonseca R, Ma AK et al (2019) Uncovering patterns of atomic interactions in static and dynamic structures of proteins. bioRxiv. https://doi.org/10.1101/840694
20. Wójcikowski M, Zielenkiewicz P, Siedlecki P (2015) Open drug discovery toolkit (ODDT): a new open-source player in the drug discovery field. J Cheminform 7:26. https://doi.org/10.1186/s13321-015-0078-2
21. G Landrum P Tosco B Kelley et al 2021 rdkit/rdkit: 2021_03_2 (Q1 2021) Release Zenodo Switzerland. https://doi.org/10.5281/zenodo.4750957
22. Gowers RJ, Linke M, Barnoud J et al (2016) MDAnalysis: a python package for the rapid analysis of molecular dynamics simulations. In Benthall S Rostrup S (eds) Proceedings of the 15th Python in Science Conference, SciPy, Austin, TX, 2016, pp 98–105. https://doi.org/10.25080/majora-629e541a-00e
23. Hajiebrahimi A, Ghasemi Y, Sakhteman A (2017) FLIP: an assisting software in structure based drug design using fingerprint of protein-ligand interaction profiles. J Mol Graph Model 78:234–244. https://doi.org/10.1016/j.jmgm.2017.10.021
24. J Reback WJ McKinney et al 2021 pandas-dev/pandas: Pandas 1.2.4 Zenodo Switzerland. https://doi.org/10.5281/zenodo.4681666
25. Rodríguez-Espigares I, Torrens-Fontanals M, Tiemann JKS et al (2020) GPCRmd uncovers the dynamics of the 3D-GPCRome. Nat Methods 17:777–787. https://doi.org/10.1038/s41592-020-0884-y
26. Ballesteros JA, Weinstein H (1995) Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Methods Neurosci 25:366–428. https://doi.org/10.1016/S1043-9471(05)80049-7
27. Kooistra AJ, Mordalski S, Pándy-Szekeres G et al (2021) GPCRdb in 2021: integrating GPCR sequence, structure and function. Nucleic Acids Res 49:D335–D343. https://doi.org/10.1093/nar/gkaa1080
28. Wang C, Jiang Y, Ma J et al (2013) Structural basis for molecular recognition at serotonin receptors. Science 340:610–614. https://doi.org/10.1126/science.1232807
29. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res 32:W665–W667. https://doi.org/10.1093/nar/gkh381
30. Katritch V, Cherezov V, Stevens RC (2013) Structure-function of the G protein-coupled receptor superfamily. Annu Rev Pharmacol Toxicol 53:531–556. https://doi.org/10.1146/annurev-pharmtox-032112-135923
31. Weis WI, Kobilka BK (2018) The molecular basis of G protein-coupled receptor activation. Annu Rev Biochem 87:897–919. https://doi.org/10.1146/annurev-biochem-060614-033910
32. Han DS, Wang SX, Weinstein H (2008) Active state-like conformational elements in the β_2-AR and a photoactivated intermediate of rhodopsin identified by dynamic properties of GPCRs. Biochemistry 47:7317–7321. https://doi.org/10.1021/bi800442g
33. Fritze O, Filipek S, Kuksa V et al (2003) Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. Proc Natl Acad Sci 100:2290–2295. https://doi.org/10.1073/pnas.0435715100
34. Robinson PR, Cohen GB, Zhukovsky EA, Oprian DD (1992) Constitutively active mutants of rhodopsin. Neuron 9:719–725. https://doi.org/10.1016/0896-6273(92)90034-B
35. Flock T, Hauser AS, Lund N et al (2017) Selectivity determinants of GPCR–G-protein binding. Nature 545:317–322. https://doi.org/10.1038/nature22070
36. Venkatakrishnan AJ, Deupi X, Lebon G et al (2013) Molecular signatures of G-protein-coupled receptors. Nature 494:185–194. https://doi.org/10.1038/nature11896
37. Flock T, Ravarani CNJ, Sun D et al (2015) Universal allosteric mechanism for Gα activation by GPCRs. Nature 524:173–179. https://doi.org/10.1038/nature14663

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.