Exhaustive Repertoire of Druggable Cavities at Protein–Protein Interfaces of Known Three-Dimensional Structure

Franck da Silva, Guillaume Bret, Leandro Teixeira, Claudio Gonzalez, Didier Rognan

Exhaustive repertoire of druggable cavities at protein-protein interfaces of known three-dimensional structure.

Franck Da Silva[†], Guillaume Bret[†], Leandro Teixeira[‡], Claudio F. Gonzalez[‡] and Didier Rognan[†*]

[†]Laboratoire d'Innovation Thérapeutique, UMR 7200 CNRS-Université de Strasbourg, 67400 Illkirch, France

[‡]Department of Microbiology and Cell Science, Genetics Institute, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL 32610-3610, United States of America.

*To whom correspondence should be addressed (phone: +33 3 68 85 42 35, fax: +33 3 68 85 43 10, email: rognan@unistra.fr)

ABSTRACT

Protein-protein interactions (PPI) offer the unique opportunity to tailor ligands aimed at specifically stabilizing or disrupting the corresponding interfaces and providing a safer alternative to conventional ligands targeting monomeric macromolecules. Selecting biologically relevant protein-protein interfaces for either stabilization or disruption by small molecules is usually biology-driven, on a case-by-case basis and does not follow a structural rationale that could be applied to an entire interactome. We herewith provide a first step to the latter goal by using a fully automated and structure-based workflow, applicable to any PPI of known three-dimensional (3D) structure, to identify and prioritize druggable cavities at and nearby PPIs of pharmacological interest. When applied to the entire Protein Data Bank, 164,514 druggable cavities were identified and classified in four groups (interfacial, rim, allosteric, orthosteric) according to their properties and spatial locations. Systematic comparison of PPI cavities with pockets deduced from druggable protein-ligand complexes shows almost no overlap in property space, suggesting that even the most druggable PPI cavities are unlikely to be addressed with conventional drug-like compound libraries. The archive is freely accessible at http://drugdesign.unistra.fr/ppiome.

INTRODUCTION

Until recently, mostly single macromolecules (proteins, nucleic acids) have been considered as potential drug targets. Out of the 30,000 proteins currently annotated in the human proteome, [1] about 450 targets have currently been addressed by low molecular-weight drugs.[2] Beside single targets, large-scale genomics and proteomics have identified complex networks of targets and pathways regulating physiopathological processes in a coordinated manner.[3] Protein-protein interactions (PPI) therefore describe a new biological space that attracts more and more attention, with already 36 PPI inhibitors under clinical development,[4] mostly in oncology.[5] Despite PPIs may adopt quite different sizes, shapes and electrostatics, identifying high-affinity PPI modulators (inhibitors, stabilizers) is a considerable challenge.[6] First and contrarily to conventional targets, a medicinal chemist cannot start inhibitor design from the structure of endogenous ligands. Second, PPIs often involve flat surfaces delocalized over multiple epitopes, and lack well-defined buried cavities typical of conventional targets. Last, high-throughput screening of traditional compound libraries often return no viable hits for the main reason that PPI modulator chemical space is quite different from that described by traditional drug-like compounds.[7] Nonetheless, coupling bioinformatics and proteomics-guided prioritization of therapeutically relevant protein-protein complexes with efficient screening strategies[8] yield more and more PPI modulators.[4,9]

Although the current human protein-protein interactome has been estimated at between 120,000[10] and 650,000 complexes,[11] a precise map of the human interactome is still missing because even the most recent and highest quality methods (e.g. yeast two-hybrid, affinity purification or co-fractionation) yields to almost non-overlapping maps.[12] Many software, metadatabases and internet resources[13] have been developed to mine protein-protein interactions at different detail levels (domain, residue, and atom) but none of them is directly usable for rational drug design purpose. Hence, at least three steps are necessary to post-process the vast amount of available structural data using heterogeneous tools and file formats. A first restriction, if one aims at using a structure-guided

approach, is to start from PPIs of known experimental structure, deposited in the Protein Data Bank (PDB).[14] Second, interfaces need to be prioritized in order to focus ligand design on pharmacologically relevant PPIs.[15] Third, once the relevant interfaces have been identified, the surface of the protein-protein complex need to be scanned to identify druggable cavities able to accommodate a high-affinity ligand.[16] It is only at this stage that a structure-based computational approach (e.g. molecular docking,[17] de novo design,[18] protein-based pharmacophore search)[19] can be applied to convert the structural description of a PPI into a set of putative ligands.

We herewith present a computational approach aimed at filling the above-described gap and fostering drug discovery from the structural knowledge of protein-protein interfaces. A fully automated flowchart, made of several steps utilizing in-house developed cheminformatics tools, has been designed to answer key issues in detecting biologically relevant PPIs and identifying druggable cavities at the interface or its close vicinity. When applied to the entire PDB, it reveals an almost unexplored space of 164,514 druggable pockets, specifically geared to design PPI modulators. Intriguingly, a systematic comparison of this PPI pocketome to that described by known druggable protein-ligand targets[20] shows almost no overlap.

RESULTS

We designed a computational flowchart (**Figure 1**) that reads the entire PDB content (PDB file format) and performs a succession of checks for (i) assigning the oligomeric state of the entry, (ii) detecting all possible two chain-interfaces, (iii) predicting their biological relevance, (iv) detecting all cavities at and nearby the interface, (v) predicting their structural druggability, and (vi) classifying predicted druggable cavities into four categories (interfacial, rim, allosteric, orthosteric) according to their properties and spatial location.

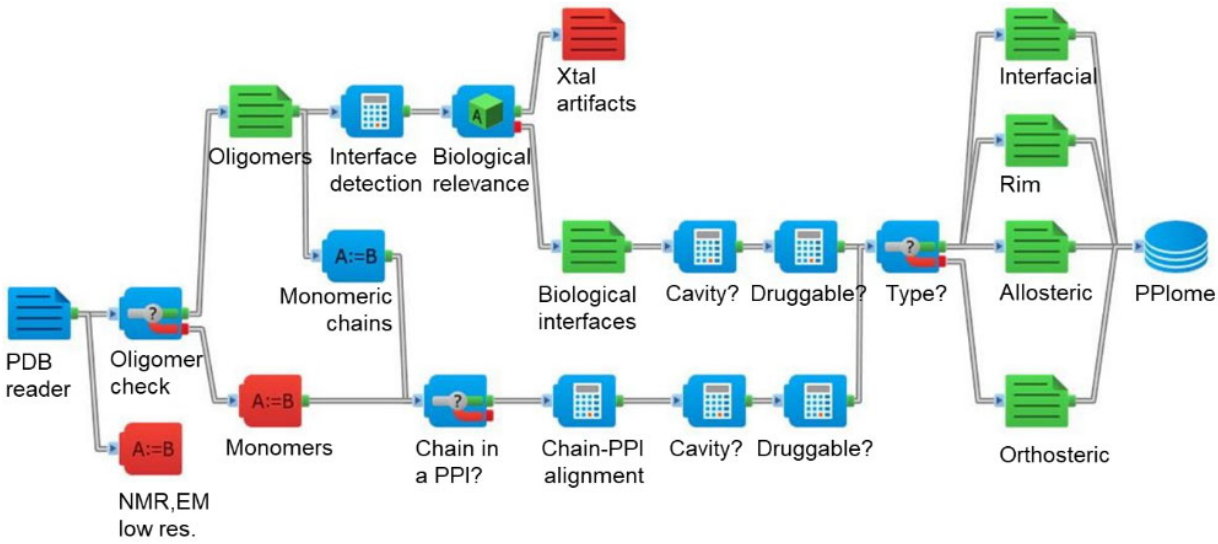

Figure 1. Computational flowchart to mine the Protein Data Bank. After eliminating low resolution (> 3 Å) and diverse entries not solved by X-ray diffraction, the oligomeric status of the entry is checked for distinguishing monomers from oligomers. For oligomers, all possible binary interfaces are checked for their biological relevance with a Random Forest model. Predicted biologically relevant interfaces are then submitted to a full cavity scan and the structural druggability of each cavity predicted with a support vector machine model. Druggable cavities are then classified in three categories (interfacial, rim, allosteric) according to their properties and spatial location with respect to the interface. Monomeric chains, discarded at the first step of the flowchart, are structurally aligned to the same chain present if part of a biologically relevant protein-protein interface. Cavity detection and ligandability estimates are performed as described above, and resulting orthosteric cavities added to the first pool of cavities.

Characterizing biologically relevant PPIs. All entries of the PDB were parsed to retain 66,621 oligomeric entries solved by X-ray diffraction with a resolution lower than 3.0 Å. To prevent spending most of the computing time on a tiny fraction of the database, entries presenting more than 10 unique protein chains (e.g. ribosomes) were discarded from the present analysis. Monomeric entries (n=30,584) passing the resolution threshold were conserved for later. A total number of 172,635 unique homo- and heterodimeric chain combinations could be detected for which 36% (n=62,278) were predicted biologically relevant using a previously-developed Random Forest model[21] trained to distinguish crystallographic artifacts from known biologically relevant interfaces For predicted biologically relevant interfaces, the number of molecular interactions between two chains varies from 7 to 853 with a mean value of 115 (**Figure 2a**). Analyzing the distribution of interaction types across all interfaces confirm previous observations[22-23] that PPIs are composed by a very large proportion of hydrophobic contacts (85%) whereas hydrogen bonds and ionic bonds are far less frequent (13.7 and 2.5%, respectively; **Figure 2b**). Comparing the distribution of molecular interactions in protein-protein (this work) and drug-like protein-ligand complexes[20] confirms that shape complementary geared by apolar contacts is really an hallmark of PPIs (**Figure 2b**), protein-ligand complexes exhibiting a much higher proportion of polar interactions (hydrogen bonds and ionic bonds). A large majority of biologically relevant PPIs concerned homodimeric entries (72%).

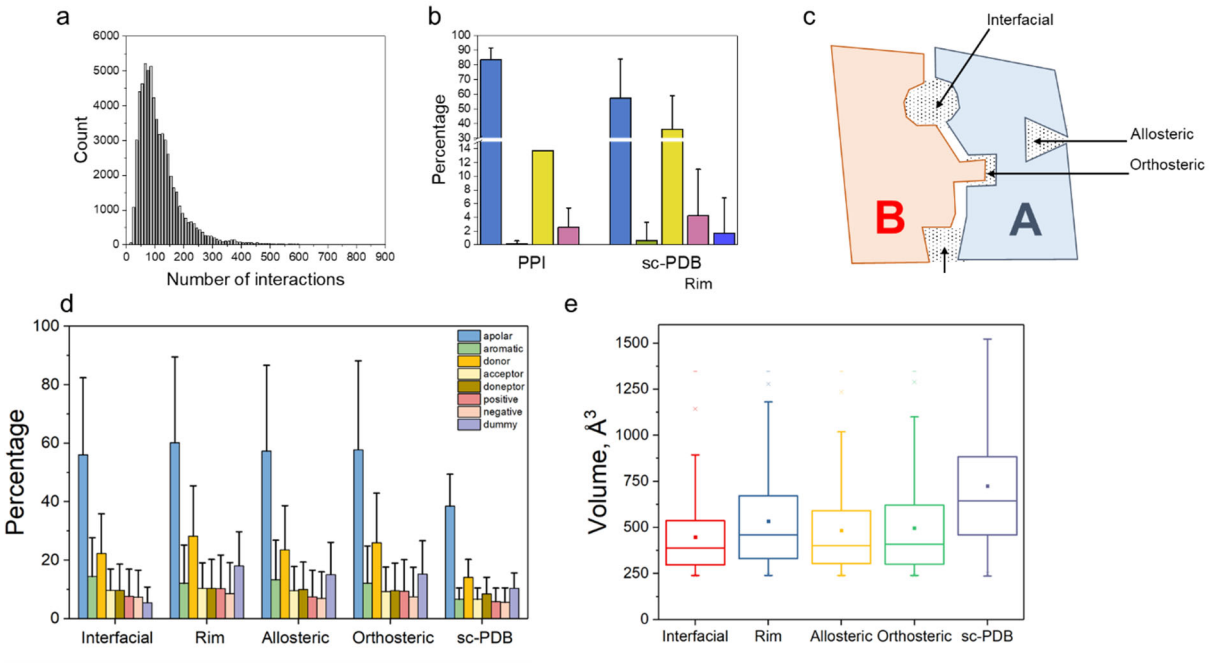

Figure 2. Biologically relevant protein-protein interfaces and their cavities. (**a**) Number of non-covalent interactions at biologically relevant protein-protein interfaces. (**b**) Distribution of molecular interactions (hydrophobic, blue; aromatic, green; hydrogen bond, yellow; ionic bond, rosy brown; metal chelation, violet) for biologically relevant protein-protein (PPI, n=62,278) and druggable protein-ligand complexes (sc-PDB, n=16,034) retrieved from the sc-PDB database.[20] (**c**) Schematic representation of four possible cavity types at or nearby an interface between two protein chains A and B. (**d**) Distribution of pharmacophoric properties (blue, apolar; green, aromatic; orange, hydrogen bond donor, yellow, hydrogen bond acceptor; sienna, hydrogen bond donor and acceptor; rosy brown, positive ionizable; tan, negative ionizable; violet, dummy) for PPI cavities (n=164,514) and drug-like ligand-bound cavities (sc-PDB, n=16,034). (**e**) Distribution of cavity volumes (in Å) for PPI cavities and drug-like ligand-bound cavities (sc-PDB). The boxes delimits the 25th and 75th percentiles, the whiskers delimit the 5th and 95th percentiles, respectively. The median and mean values are indicated by a horizontal line and a filled square in the box.

Identifying druggable cavities at and nearby biologically relevant PPIs. Having identified the protein-protein complexes of interest, we next scanned the surface of all corresponding dimeric complexes in

order to detect cavities and estimate their druggability with the previously reported VolSite method.[24] Briefly, the method first assigns pharmacophoric features (H-bond acceptor, H-bond donor, H-bond acceptor and donor, negative ionizable, positive ionizable, hydrophobic, aromatic) to grid points encompassing each detected cavity. Since pharmacophoric features are defined by complementarity to the nearest protein atoms, cavity grid points define a negative image of an ideal ligand covering the entire cavity. Then a support vector machine (SVM) model, trained on 73 VolSite descriptors (number of cavity points, proportion of each feature type, accessibility of every grid point) of 113 cavity structures (71 druggable, 42 undruggable) is used to predict in a binary manner (druggable, undruggable) the structural druggability of each detected cavity. Since its original description,[24] the SVM model has been updated several times, along with VolSite improvments (e.g. better definition of hydrophobic contacts)[25] while conserving the same accuracy of 89%.

In order to identify orthosteric cavities that are unmasked only at the surface of monomeric proteins contributing to the PPI, the protocol was divided in two different workflows (**Figure 1**). The first one was run on dimeric complexes selected in the previous section and aims at identifying any but orthosteric cavities. The second one was only performed on isolated monomeric chains at the condition that the corresponding protein chain (annotated by chain name and UniProtKB identifier) was also part of a predicted biologically relevant dimeric complex. To analyze both entries in the same coordinate space, the structure of the monomeric chain was aligned to that of the corresponding chain in the dimer.[26] Detected cavities at the surface of the aligned monomeric chain were then merged to the coordinate space of the corresponding dimer, enabling to locate the cavities with respect to the dimeric interface. According to our previous work on protein-ligand complexes,[20] lower and upper volume thresholds (230 $Å^3$, 1350 $Å^3$) were applied to remove pockets that would be either too small or too large to accommodate any high-affinity drug-like ligand. Out of the 919,599 cavities identified, 164,514 (17.9 %) were predicted druggable by our support vector machine (SVM) model[24] trained on

45 pockets descriptors, including size, physicochemical properties, buriedness and curvature (**Table 1**; **Figure 2c**). Druggable cavities were next classified in four categories (interfacial, rim, allosteric, orthosteric; **Figure 2d**) depending on their chain composition, average buriedness and distance to the PPI interface (**Table 1**). Cut-off values for the two descriptors (buriedness, distance to interface) were chosen from the manual inspection of known interfacial[9] and orthosteric[27] cavities. We could detect 6,332 interfacial cavities[9] which are by definition located at the interface and almost entirely buried by the two protein chains in contact. By contrast, a large number of solvent-exposed rim cavities (72,242) could be found as they originate from the assembly of three-dimensional objects[28]. By nature, they are also formed by the two chains in contact but remains largely accessible. 41,978 allosteric PPI cavities[29] are located at the surface of only one of the two chains in complex. To limit the relevance of such cavities, only pockets distant by less than 10 Å from the interface were considered here. Last, 43,962 orthosteric cavities were identified. These pockets are only present at the surface of one of the two interacting chains (in their free state) and are located exactly at a protein-protein interface when the latter chain is involved in a biologically relevant PPI.

Table 1. Classification of PPI cavities

Type	Chains[a]	Buriedness[b]	Distance to interface[c]	Number
Interfacial	A and B	≥ 75%	< 3 Å	6,332
Rim	A and B	< 75%	< 3 Å	72,242
Allosteric	A or B	≥ 60%	3-10 Å	41,978
Orthosteric	A or B	≥ 60%	< 4.5 Å	43,962
Total count				164,514

[a]Prototypical AB dimer. [b]Average buriedness over all voxels defining the cavity. [c]Smallest distance in Å between any cavity voxel and any interaction pseudoatom representing intermolecular interactions between the two protein chains A and B.

We should point that neither sequence nor structure redundancy has been applied here, as we wanted to define a full repertoire of druggable cavities that would be insensitive to local minor changes (e.g. rotameric states of cavity-lining residues) and enable full pairwise cavity comparisons.

Analyzing the pharmacophoric properties of all druggable cavities does not reveal major differences among the four cavity types (**Figure 2d**). They are dominated by hydrophobic features (ca. 60% of all features). The only significant difference, as expected, relates to the solvent exposure that is lower for interfacial cavities, as revealed by the lower proportion of dummy features (propertyless cavity features farther than 4.5 Å away from any protein heavy atom; **Figure 2d**). Interestingly, the overall distribution of pharmacophoric properties observed for PPI cavities deviates from that observed for conventional ligand-bound druggable cavities in the sc-PDB database.[20] PPI druggable cavities are clearly more hydrophobic and less buried (**Figure 2d**). Another noticeable difference lies in the distribution of cavity volumes that are significantly smaller for PPI pockets (**Figure 2e**).

Ligand occupancy. We next inspected the location of co-crystallized molecules (ligands, biochemical reagents, ions, etc…) with respect to the previously identified druggable PPI cavities by measuring for each entry, the closest distance between the ligand's center of mass and any cavity pharmacophoric feature. If the distance is below a user-defined threshold (1.06 Å for interfacial, rim and allosteric cavities; 2.12 Å for orthosteric cavities; see Experimental section), the ligand is annotated as being bound to the corresponding cavity. A vast proportion of druggable PPI cavities are indeed free of any ligand, however with noticeable differences according to the cavity type (**Figure 3a**). Rim cavities are significantly less occupied by bound ligands (9% occupancy on average) than the three other types for which a ligand is found in 14-19% of cases (**Figure 3a**). Using an previously developed functional annotation of PDB chemical components,[30] we could annotate pocket occupancy by ligand type (**Table 2**). For all cavity types, about 50% of unique bound ligands are represented by pharmacological tools devoid of drug-likeness character (e.g. S-adenosyl-L-homocysteine, retinal). True drug-like ligands

constitute the second largest category of bound molecules (ca. 30%), notably for orthosteric cavities for which we could identify 1,572 unique ligands (**Table 2**). Biochemical reagents (e.g. glycerol, ethanediol) and cofactors are present in all cavity types, with lower percentages (6-12%) considering the number of unique molecules (**Table 2**) but a very high frequency (**Table S1**, Supporting Information).

Table 2. Small molecules (unique HET groups) bound to PPI cavities

Ligand type[a]	Cavity			
	Interfacial	Rim	Allosteric	Orthosteric
Drug-like[b]	149	455	559	1,572
Pharmacological tool[c]	209	1,059	993	2,618
Reagents	40	221	205	293
Cofactor	1	13	12	27
Prosthetic group	0	2	3	3
Natural aminoacid/peptide	3	5	8	0
Modified aminoacid/peptide	9	30	19	68
Sugar	1	31	44	58
Organometallic	5	22	35	56
Nucleic acid/nucleotide/nucleoside	8	19	19	26
Metal	0	3	4	1
Ion	3	13	11	31
Water	0	1	0	1
Total	428	1,875	1,915	4,574

[a]Ligand annotation was realized with IChem, as previously described.[30] [b]Drug-likeness was defined by a set of topological fiters (**Table S2**, Supporting Information) implemented in OpeneEye's Filter program (OpenEye Scientific Software, Sante Fe, U.S.A.). [c]Ligand rejected by the drug-likeness filter.

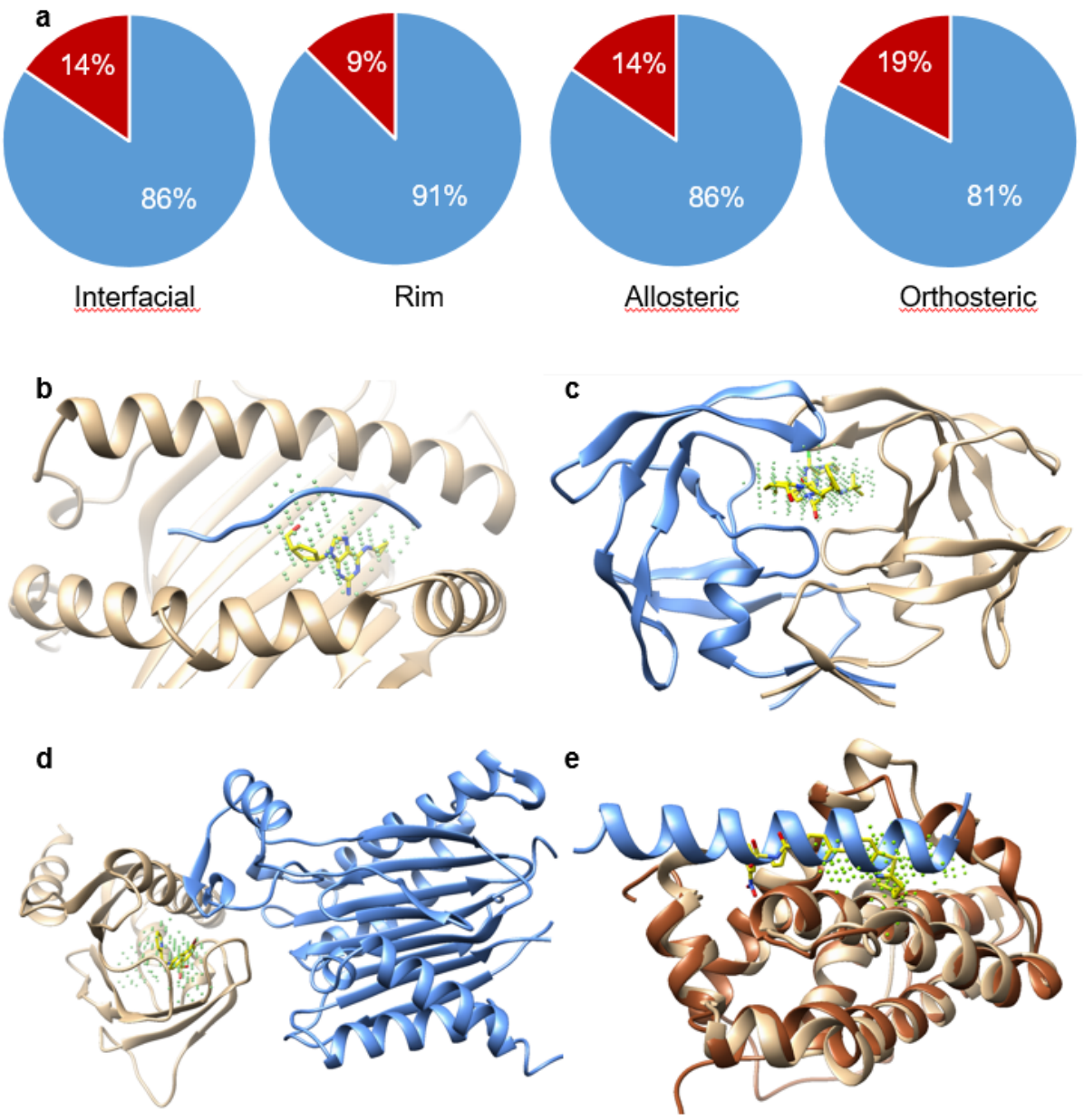

Figure 3. Occupation of PPI cavities by known ligands. **(a)** Percentages of ligand-free (blue) and ligand-bound (red) PPI cavities. **(b)** Example of interfacial cavity (green sphere) between the HLA-B*57:01 protein (tan) and a peptide antigen (blue) with bound ligand (abacavir, yellow sticks; PDB accession number 3VRI). **(c)** Example of rim cavity (green sphere) between the alpha chain (tan) and beta chain (blue) of human HIV-1 protease with bound inhibitor (3IN, yellow sticks; PDB accession number 2C6X). **(d)** Example of allosteric cavity (green sphere) between the PYL1 receptor (tan) and the type 2C protein phosphatase ABI1 (blue) with bound ligand (pyrabactin, yellow sticks; PDB accession number 3NMN). **(e)** Example of orthosteric cavity on Bcl-2 (brown) with bound inhibitor (43B, yellow sticks; PDB accession number 2O21). The inhibitor-bound structure is aligned to the structure of Bcl-2 (tan) in complex with the Bax BH3 peptide (blue; PDD accession number 2XA0).

Four prototypical examples of drug-like ligands occupying PPI druggable cavities are hihlighted in **Figure 3b-e**. For example, the antiviral compound abacavir binds to an interfacial cavity delimited by two chains of a class I human histocompatibility-peptide complex (PDB accession number 3VRI),[31] thereby altering the repertoire of natural peptide antigens presented to the HLA-B*57:01 allele and inducing severe immunological side effects.[32] A perfect illustration of a ligand bound to a rim pocket is given by HIV-1 protease inhibitors that bind to the catalytic pocket at the interface of the protease dimer (PDB accession number 1C6X). Allosteric PPI modulation is here examplified by pyrabactin that binds to a specific pocket of the PYL1 receptor and further locks the type 2C protein phosphatase ABI1 (PDB accession number 43NMN) in an inactive conformation.[33] Last, a classical example of orthosteric inhibition is provided by the anti-apoptotic Bcl-2 inhibitor 43B bound to the Bax BH3 binding pocket (PDB accession number 2O21).

Druggable PPI cavities are unique. Inferring potential ligands for PPI pockets may be obtained by aligning known ligand-bound pockets to PPI cavities and merging the ligand of the best-aligned pockets into the PPI cavity space.[16] To represent holo pockets, we extracted 16,034 druggable cavities from the sc-PDB database.[20] We next systematically compared all but allosteric PPI cavities to 16,034 sc-PDB pockets with the previously reported Shaper algorithm,[24] thereby defining a huge similarity matrix of 1.98 billion comparisons. In this exercise, allosteric PPI cavities were not considered as they are supposed to be already present in the sc-PDB archive. To estimate pair-wise cavity similarities, we used the in-house developed Shaper software[24] that uses a smooth Gaussian function to maximize the overlap of both shapes and properties of cavity-defining pharmacophoric features. To facilitate the analysis of the large matrix, only pairs exhibiting a similarity higher than 0.6 (expressed by a Tversky coefficient on pharmacophoric features overlap; See Experimental section) and a limited difference in size (less than 10%) were further considered (Filter 1, **Table 3**). Removing trivial matches (Filter 2, **Table 3**) between PPI and sc-PDB entries limit the number of possible matching pairs to very small numbers

(**Tables S3-S5**, Supporting Information) that were inspected in more details. We notably compared the predicted location of the sc-PDB-bound ligand in the PPI cavity according to two methods: (i) merging the sc-PDB ligand in the PPI cavity according to the rotation/translation matrix corresponding to the best-scored cavity match, (ii) docking the sc-PDB ligand in the PPI pocket. If both methods agreed to place the ligand in the PPI cavity within 2 Å root-mean square deviations (rmsd) on heavy atoms, the corresponding cavity match was considered as potentially interesting (Filter 3, **Table 3**). Out of the 1.98 billion comparisons, only 2 matches were retained, both concerning a predicted similarity between an interfacial cavity of the LJ0536 cinnamoylesterase homodimer (PDB accession number 3S2Z)[34] and the flurbiprofen (FLP)-bound cavities of cyclocygenase-1 (PDB accession number 1EQH)[35] and cyclooxygenase-2 (PDB accession number 3PGH).[36]

Table 3. Pairwise comparison of PPI to sc-PDB cavities

Cavity type	Number of Comparisons	Remaining pairs		
		Filter 1[a]	Filter 2[b]	Filter 3[c]
Interfacial	101 million	2,203	11	2
Rim	1.16 billion	3,121	23	0
Orthosteric	727 million	1,474	8	0

[a]Similarity > 0.60, less than 10% differences in cavity volumes. [b]Different UniProt identifiers and names, different gene names, different protein families. [c]Less than 2 Å rmsd between ligand positions obtained by either cavity alignment or molecular docking

Both cavities (**Figure 4a-b**) have a similar shape (3S2Z, 239 $Å^3$; 1EQH, 258 $Å^3$) and exhibit a strong hydrophobic character (ca 85% of cavity features are hydrophobic in both cases). When aligned to optimize the overlap of their pharmacophoric feature, the shape of the cavities matched almost perfectly. Interestingly, FLP automatically moved with its pocket upon alignment of cyclooxygenase-1 to cinnamoylesterase) is well centered in the cinnamoylesterase cavity in a position quite similar to

that predicted by docking (**Figure 4c**). FLP is predicted to interact only with apolar side chains lining the interfacial cavity of the esterase (Figure 4c).

The prediction was validated *in vit*ro by demonstrating a weak but significant dose-dependent inhibition by FLP of the catalytic activity of the LJ0536 cinnamoylesterase in converting 4-nitro phenylbutyrate into 4-nitro phenol (**Figure 4d**). In order to evaluate if inhibition over the esterase activity of the purified enzyme from Lactobacillus johnsonii N6.2 was not due the high concentrations of the inhibitor, we tested different concentrations of esterase in the presence or absence of a fixed concentration of FLP (400µM) in the conversion of 4-nitrophenyl butyrate into p-nitrophenol. The results obtained show that the esterase activity of the enzyme was significantly impaired in the presence of FLP in all concentrations of esterase tested (**Figure 4e**). Next, we evaluate if the inhibitory activity of FLP on the esterase activity of the purified enzyme was competitive or noncompetitive. To address this question, we ran a kinetic assay evaluating the breakthrough of 4-nitrophenyl butyrate in the presence or absence of flurbiprofen at a concentration of 400µM (**Figure 4f**). In the presence of flurbiprofen, the Vmax of the enzyme was significantly decreased (3.7µmol/min/mg) with respect to the control (4.094 µmol/min/mg), thereby demonstrating the noncompetive nature of flurbiprofen inhibition. Given that the cinnamoylesterase dimer exhibits only two catalytic sites (one per monomer) and one allosteric interfacial cavity, flurbiprofen is very likely to bind to the interfacial cavity, as predicted above. Of course, we cannot rule out the possibility that flurbiprofen unmasks a previously unknown binding pocket at the surface of the esterase. The rather weak inhibitory activity reported in enzymatic assays is however compatible with the predicted binding mode exhibiting a good shape complementarity between the ligand and the interfacial pocket but an important electrostatic mismatch between the carboxylic acid of the ligand and its apolar protein environment (**Figure 4c**).

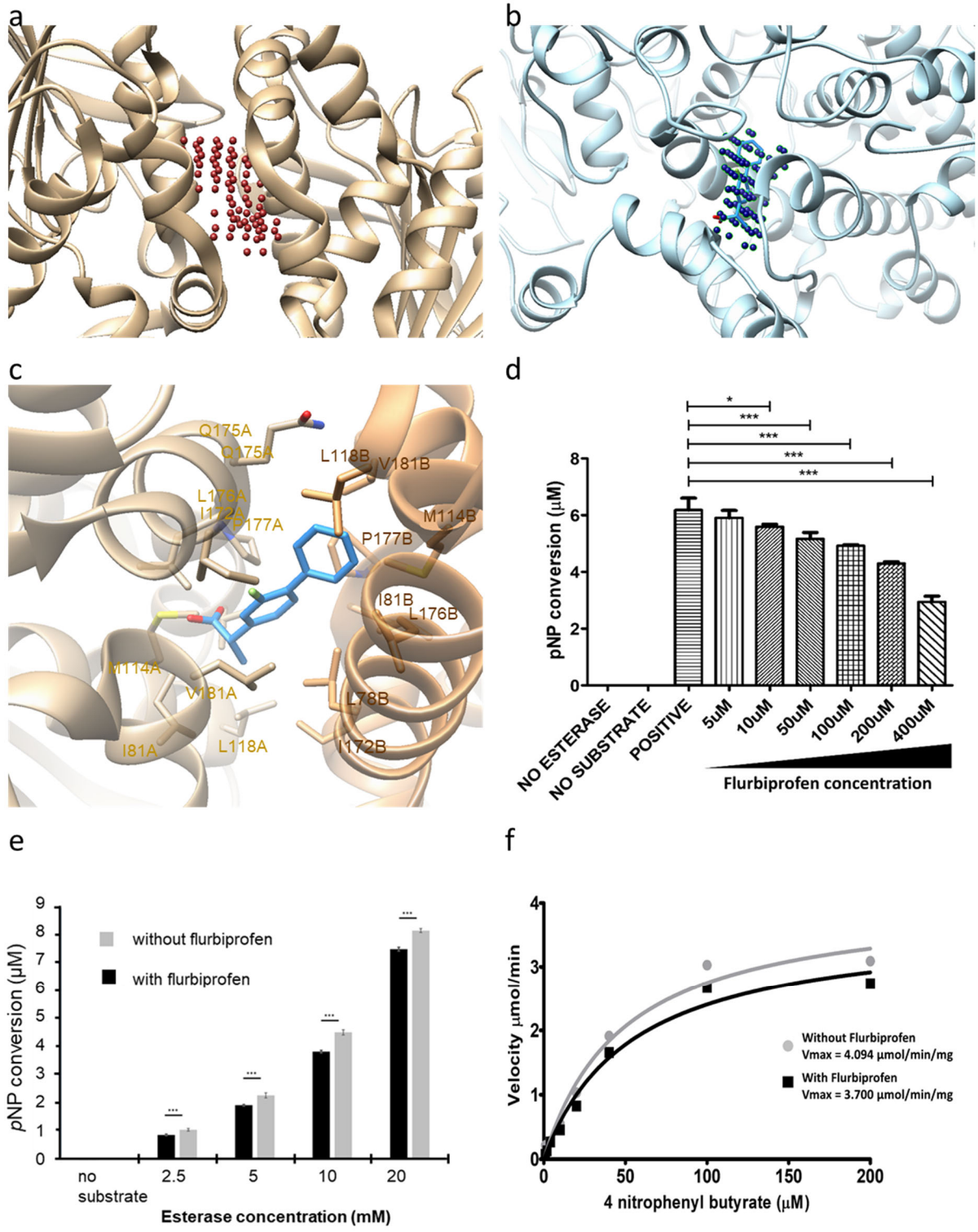

Figure 4. Similarity between cavities of LJ0536 cinnamoyl esterase and cyclooxygenase-1. **(a)** Interfacial cavity (brown dots) at the surface of the LJ0536 cinnamoylesterase dimer (tan ribbons, PDB accession number 3S2Z), **(b)** Flurbiprofen (dodger blue sticks)-binding cavity (dark blue dots) in cyclooxygenase-1 (cyan ribbons, PDB identifier 1EQH), **(c)** Shaper-based alignment of 3S2Z and 1EQH cavities. Predicted

location of flurbiprofen in the esterase cavity is inferred from the cavity-based alignment of both targets, **(d)** Concentration-dependent inhibition by flurbiprofen of the catalytic activity of LJ0536 cinnamoylesterase (600 µM), estimated by the in vitro hydrolysis of 4-nitro phenylbutyrate in 4-nitro phenol (pNP). Results are mean ± standard deviations of three experiments. * $P < 0.05$, *** $P < 0.001$ vs. no inhibitor (POSITIVE). **(e)** Conversion rate of 4-nitrophenyl butyrate into 4-nitro phenol (pNP) by different concentrations of purified esterase from Lactobacillus johnsonii N6.2 in the presence or absence of flurbiprofen at 400µM. Results are mean ± standard deviations of three experiments (*** $P < 0.001$), **(f)** Representative enzymatic kinetic curves of the activity of purified esterase from Lactobacillus johnsonii N6.2 as spectrophotometrically monitored by the appearance of pNP (absorbance at 412 nm) at the indicated substrate concentrations (4-nitrophenyl butyrate) in the presence or absence of flurbiprofen at 400µM.

Querying the database. A web interface to the full repertoire of PPIs and their associated druggable pockets and bound ligands is accessible at http://drugdesign.unistra.fr/ppiome. The database can be searched from multiple point of views (PDB structure, protein annotations, protein-protein interface, druggable cavity, bound ligand) therefore enabling complex queries. For example, selecting heterodimeric PPIs devoid of enzymatic activity with a structure solved at a resolution below 2 Å and exhibiting interfacial cavities larger than 500$Å^3$, distant by less than 2 Å from the interface, and presenting an average buriedness higher than 75% can be done with a few mouse clicks (**Figure S1A**, Supporting Information**),** and returns a total of 22 different interfaces in a results summary (**Figure S1B**, Supporting Information) that can be individually visualized by selecting the 'view' option. Choosing for example the 1R8Q interface between chains A and B (ppi_interface_id = 1R8QAE) permits to visualize both the PPI, related cavities and bound ligands in a molecular viewer along with corresponding molecular properties in expandable spreadsheets (**Figure S2**, Supporting Information).

DISCUSSION

Up to now, PPI modulation by low molecular weight compounds has been essentially focused on ligands binding to orthosteric pockets, directly competing with one of the two protein partners [4]. With noticeable exceptions,[37] current PPI modulators are usually high molecular weight ligands with rather poor pharmacokinetic properties.[4] Such compounds rarely belong to conventional screening libraries, therefore explaining why high-throughput screening for PPI modulators often return yield low hit rates.[38] Designing PPI-focused libraries from the current knowledge on PPI inhibitors has led to remarkable successes[39] but the overall applicability domain of these libraries remain obscure, as they are derived from known inhibitors covering only ca. 30 unique PPIs.[40] The current study provides a first answer to the latter issue since the herein described computational workflow is able to provide for the first time a clear picture of all druggable cavities present at the surface of biologically relevant PPIs of known X-ray structures. The proposed protocol exhibits three key advantages: (i) it is fully automated and fast enough to be applied at the entire PDB scale, (ii) it distinguishes biologically relevant interfaces from that suspected to be artifacts from crystallization conditions, (iii) it can detect all cavities located at or nearby the interface and predict their druggability (or ligandability).

For a matter of semantics, it is important to notice first that protein-protein interfaces have been here selected from a structural point of view. In other words, any homo/heterodimeric protein (enzymatic or not), exhibiting enough non-covalent interactions between the two chains in contacts, will be considered. Therefore, many PDB entries (e.g. HIV-1 protease) that are not considered as regular PPIs by the drug design community are taken into account in the present study. A first surprise in the application of the current computational workflow is the very large proportion of binary interfaces (64%) predicted as non-biologically relevant by our machine-learning model. Many of these cases apply to homodimeric structures of proteins known to be monomeric in solution but for which both the asymmetric unit and the biological assembly provided by the PDB describes a homodimer. A smaller

but significant fraction of interfaces that were removed from our analysis concerns very large protein-protein interfaces (> 3000 Å^2) that might be incorrectly classified by our model trained on interfaces of smaller sizes (700-2000 Å^2).[30] As already stated in previous studies,[41-42] we here confirm on a very large scale (>66,000 PDB entries) that molecular interactions contributing to biologically relevant protein-protein interfaces are largely dominated by hydrophobic contacts.

A second surprise of the current work is the huge repertoire of PPI pockets (164,514 cavities) presenting all physicochemical properties of truly druggable cavities. Please remind that no flat and featureless pockets have been considered here. Neither were selected proximal but distinct non-druggable sites that could be simultaneously occupied by a single high molecular weight inhibitor.[43] No PPI-specific druggability model was used in the current study because of the very unbalanced and scarce nature of currently available experimental data on truly druggable PPI cavities. Moreover, we believe that druggability (or ligandability or bindability) is a ligand-independent property solely determined by physicochemical and topological properties of the cavity of interest.

Beside interfacial and orthosteric cavities, whose druggable potential are well documented,[9, 4, 29] we identified 72,242 pockets formed at the rim of the interface that represent an almost uncharted pocket space of high druggability. A similar observation was previously reported on a much smaller scale but no particular focus on either biologically-relevant interfaces and potentially druggable cavities.[28] Most of PPI cavities (85% on average) are present in a ligand-free state. However, over 2,000 drug-like compounds were found in orthosteric, interfacial and rim pockets. Out of the 756 PPI inhibitors reported in the 2P2I database,[27] 270 compounds (36%) could be recovered using the current computational workflow. Reasons for failures are twofold: (1) the PPI is not predicted as biologically relevant by our Random Forest model because one of the two protein chains is a short peptide that generates a small-sized interface (e.g. bromodomain-H4 histone interfaces), (ii) no cavity at or nearby the PPI could be predicted as druggable by our support vector machine model (e.g. Bcl-2 cavity bound to ligand LIU, PDB identifier 2O22).

Out of the four categories of ligands targeting each a cavity type (interfacial, rim, allosteric, orthosteric), the drug design community has almost only followed the concept of orthosteric inhibition by designing usually high molecular-weight compounds disrupting the PPI by occupying a cavity at the surface of one of the two partners.[4] Such ligands remain the most straightforward approach to inhibit a PPI by a small molecule. Interfacial inhibitors (or interfacial stabilizers), [9,44-45] that stabilize a PPI through binding to interfacial cavities, have long been neglected but are getting more and more interest, inspired by natural products (e.g. fusicoccin, rapamycin, brefeldin A) that efficiently inhibit the function of a protein-protein complex by stabilizing the corresponding dimeric assembly.[9] Examples of rim and interfacial ligands modulating PPIs are still scarce, for the simple reason that high-throughput screening assays might not have been developed for that purpose. Paclitaxel remains a good example of allosteric PPI stabilization, by binding to a hydrophobic pocket of the tubulin β-subunit and promoting the stabilization microtubules.[46] Predicting the functional outcome of ligand binding to either allosteric or rim cavities is currently out of reach. We, however, do hope that the herein presented PPI-focused pocketome will foster PPI modulator design by addressing allosteric and rim cavities.

Noteworthy, most ligands (71% on average) do not exhibit a single violation to Lipinski's rule-of-five,[47] thereby attesting a clear potential for medicinal chemistry development. Analyzing standard molecular properties of cavity-bound drug-like and pharmacological ligands reveals that rim and allosteric and ligands tend to share similar properties. Interestingly, interfacial ligands tend to exhibit lower molecular weights, and are more rigid and hydrophobic (**Figure S3**, Supporting Information). We next computed the PBF (plane of best fit) score,[48] as general globularity descriptor for herein described PPI ligands, conventional drug-like compounds (sc-PDB) and true PPI inhibitors from the 2P2I database (**Figure S4**, Supporting Information). Ligands bound to rim, orthosteric, and allosteric pockets exhibit quite similar properties. Interestingly, interfacial ligands tend to be less globular (lower PBF score) and smaller (lower radius of gyration). As to be expected true PPI inhibitors from the 2P2IDB archive are significantly more globular. This difference is however biased by the fact that the 2P2IDB repository

does not take into account the druggability of ligands and pockets, thereby biasing the selection of ligands towards higher molecular weight compounds. The same observation applies to sc-PDB ligands for which the drug-like subset has a much lower PBF median score (0.71) than the non drug-like subset (PBF=0.87).

Despite their apparent drug-like character, the large repertoire of PPI pockets do not resemble cavities bound to drug-like ligands. Out of 1.98 billion pairwise comparisons, only two matches could be found, one of which (LJ0536 cinnamoylesterase dimer vs. cyclo-oxygenase-1) being validated in vitro. Due to the enormous size of the comparison matrix, we had to restrict experimental validations to a subset of similar-sized pocket pairs. We cannot therefore rule out the possibility of local similarity between overlapping parts of PPI pockets and ligand-bound cavities that could be occupied for example by fragments. Moreover, the herein inspected protein-protein interactome is only a subset of a much larger space. However, the current analysis provides a strong evidence that both pocketomes do not overlap in 3D space, and consequently a clear structural basis for the known difficulty of conventional drug-like ligands to target PPIs.

CONCLUSIONS

The herein characterized pocketome offers novel opportunities for PPI modulation by low molecular weight compounds that significantly differ from current strategies. First, it demonstrates that PPI stabilization or disruption by high molecular weight ligands addressing flat and featureless interfaces is not an absolute dogma. We notably characterized a huge repertoire of druggable cavities, present at the surface of PPIs, and still awaiting ligands to be discovered by *in silico* or experimental screening. We acknowledge that the functional effects of such ligands might be difficult to infer from the simple location of the targeted cavity. It is more than likely that ligands binding to orthosteric pockets will indeed disrupt the corresponding PPI and therefore be further developed as true PPI inhibitors.[4] Moreover, targeting an interfacial pocket will probably lead to interfacial inhibitors[9,44-45] stabilizing the

protein-protein interface and thereby enabling activating or inhibiting a specific biological function. The functional outcome of ligands targeting herein described rim and allosteric pockets is much more difficult to predict. We here confirm at a large scale than druggable pockets tend to be created nearby PPIs by the simple association of three-dimensional objects. The true potential of rim and allosteric inhibitors has already been demonstrated for homodimeric enzymes (e.g. HIV-1 protease, HIV_1 reverse transcriptase). Whether this paradigm might be applied to heterodimeric non-enzymatic assemblies still remain to be addressed and offers interesting possibilities to modulate PPIs not amenable to classical orthosteric inhibition.

To support our view, we recently identified extracellular allosteric modulators of receptor tyrosine kinases (RTK) preventing, with a high affinity and exquisite selectivity the binding of endogenous ligands (cytokines, growth factors) to the extracellular domain of targeted RTKs.[49-50]. Second, our study suggests that neither conventional drug-like libraries nor PPI-focused compound collections are likely to yield high affinity ligands to the herein described pocketome. One the one hand, current PPI-focused libraries are enriched in high molecular weight ligands designed to occupy multiple, small and shallow pockets, observed on a small number of unique PPIs.[40] On the other hand, standard drug-like ligand collections have been built to mimic the chemical space intersecting the most valuable pharmaceutical targets (kinases, nuclear hormone receptors, proteases, G-protein coupled receptors, ion channels) whose cavities are very different from that presented here. The physicochemical properties of this newly disclosed pocketome and the corresponding drug-like ligand repertoire suggests some hints about which chemotypes to consider for hit identification. First, ligands of limited volumes (400-500 $Å^3$) should be considered in order to fit the observed distribution of PPI cavity volumes. Applying this strategy is perfectly suited for biophysical fragment screening methods.[8, 4] Second, prioritizing rather rigid compounds is another advisable strategy[29] to efficiently disrupt or stabilize peculiar conformations of protein-protein interfaces. We acknowledge that our approach does not explicitly consider protein flexibility and is unable per se to identify cryptic pockets transiently unmasked at the surface of protein-protein complexes.[51] However, the workflow can easily be applied to molecular

dynamics trajectories of any PPI complex to address the important issue of transient pockets, but on a case-by-case basis. Moreover, we have only considered high-resolution X-ray structures up to now. The spectacular development of cryo-electron microscopy[52] to solve the structure of macromolecular assemblies will undoubtedly increase the applicability domain of our approach in a very near future.

EXPERIMENTAL SECTION

The overall post-processing of raw PDB data was done using the in-house developed IChem toolkit[30] and a series of standalone python scripts. Two tables (pdb_chain_uniprot.csv, uniprot_pdb.csv) were retrieved from the SIFTS resource[53] to map, for each each PDB entry, a chain name to a unique UniProtKB[54] identifier.

Detection of protein-protein interfaces. Raw PDB files were parsed to retain entries at the condition that three criteria were satisfied: (i) the experimental method is X-ray diffraction, (ii) the resolution is below 3 Å, (iii) the number of unique chain names described by at least 10 consecutive residues is between 2 and 9. For each possible interface between 2 different chains, the corresponding dimeric structure was saved, conserving in case of multiple occupancy values for a single atom, the atomic coordinates corresponding to the highest occupancy. Hydrogen atoms were then added using the Protoss algorithm[55] and the protonated dimer was saved in MOL2 file format. An interface (continuous or discontinuous) was kept if at least 20 atoms from at least 10 unique amino acids (involving both chains) were closer than 5 Å from each other. Precise intermolecular interactions (apolar, aromatic,

hydrogen bond, ionic bond) were then computed with IChem,[30] relying on a set of standard topological descriptors.[56] The corresponding interaction pattern was saved in MOL2 file format as a set of interaction pseudoatoms[24] placed at the geometric barycenter of each interacting atom pair, and annotated according to the type of interaction. The biological relevance of the protein-protein interaction was last assessed, from the value of 45 descriptors assigned to the above-described interaction pseudoatoms, with a Random Forest binary classification model, as previously described.[21]

Detection of ligand-binding druggable cavities. Dimeric structures predicted to exhibit a biologically relevant interface were then inspected for the existence of druggable cavities with the in-house developed VolSite method.[24] Each cavity was represented by a set of pharmacophoric features, complementary to that of the nearest protein atom, and centered on the corresponding cavity voxel. A support vector machine model, previously trained on 73 molecular descriptors of the cavity-dependent pharmacophoric features set,[24] was applied to estimate the druggability (ligandability) of each detected cavity. The druggability score (≥0 if druggable, < 0 if non-druggable), volume (in $Å^3$), pharmacophoric feature composition (% hydrophobic, % aromatic, %h-bond donor, %h-bond acceptor, %h-bond acceptor and donor, %positive ionizable, % negative ionizable) and average buriedness was saved for each cavity along with atomic coordinates (MOL2 format) of pharmacophoric features describing the cavity.

Detection of orthosteric cavities. For orthosteric cavities, a specific procedure was applied since these cavities require the inspection of monomeric chains. If a monomeric protein (characterized by a chain name and UniProtKB identifier) is contributing to a previously identified biologically relevant PPI interface, the structure of the monomeric chain was aligned to that of the same chain in the dimer with the Combinatorial Extension (CE v1.02) algorithm,[26] keeping the alignment with the lowest rmsd as final result. The cavity detection step was then performed on the aligned monomeric structure, as

previously described. Any predicted druggable cavity located with 4.5 Å of the corresponding protein-protein interface (shortest distance between any cavity point and any PPI interaction pseudoatom) was assigned the orthosteric type.

Pairwise cavity comparisons. All pairwise cavity comparisons were done with the Shaper method,[24] by comparing pharmacophore-annotated cavity shapes generated by VolSite. Shaper attempts to maximize the overlap of pharmacophoric features using a smooth Gaussian function taken from the ShapeTK toolkit (OpenEye Scientific Software Inc, Santa Fe, U.SA.), and outputs a similarity score $S_{A,B}$ between cavity A (reference) and B (fit) by a Tversky index as follows:

$$S_{A,B} = \frac{O_{A,B}}{0.95\, I_A + 0.05\, I_B + O_{A,B}}$$

where $O_{A,B}$ is the overlap between pharmacophoric features of cavities A and B, and I non-overlapped features of each entity A and B.

Calculation of ligand properties. Drug-likeness was predicted from a set of topological and physicochemical filters (**Table S2**, Supporting Information) implemented in OpenEye's Filter v.2.5.1.4 (OpenEye Scientific Sofware, Sante Fe, U.S.A). 3D MOL2 files of each ligand were obtained by converting SMILES strings, downloaded from the Protein Data Bank, with Corina v3.40 (Molecular Networks GmbH, Erlangen, Germany). Molecular properties of ligands were then computed with Pipeline Pilot v16.5.0.143 (Dassault Systèmes Biovia Corp., San Diego, U.S.A.)

Assigning ligands to cavities. Cavity occupancy was estimated, for each PDB entry, by measuring the shortest distance between any cavity point and any ligand's center of mass (ligand being here defined as any PDB HET identifier present in the PDB entry). If the distance is below a user-defined threshold

T dependent on the grid resolution R (by default equal to 1.5 Å), the ligand is considered located in the corresponding cavity.

For any but orthosteric cavity, $T = \frac{\sqrt{2\,R^2}}{2}$ = 1.06 Å

For orthosteric cavities, $T = \sqrt{2\,R^2}$ = 2.12 Å

The larger distance threshold used to map ligands to orthosteric cavities is explained by the prior alignment of ligand-bound monomeric chains to the corresponding PPI that might not be perfect and by the potential conformational rearrangement of the PPI pocket when comparing free and bound states. It was therefore set to twice the value of the grid resolution used to detect cavities.

Cloning, expression and purification of the LJ0536 esterase. The gene of interest was PCR amplified from genomic DNA isolated as previously reported.[57] The PCR fragments were cloned into pET-15b-TV according to methods described previously.[58] The expression of His6-tagged proteins was carried out in Escherichia coli (DE3) cells (Stratagene) by using IPTG (isopropyl-β-D-thiogalactopyranoside) (1 mM) as inducer. Escherichia coli BL21cells were collected by centrifugation, suspended in binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl [pH 7.9]), and disrupted using a French press. The His6-tagged proteins were purified by affinity chromatography as previously described.[59] The purified protein was dialyzed during 16 h at 4°C against a solution containing 50 mM Tris-HCl buffer (pH 8.00), 500 mM NaCl, and 1 mM dithiothreitol. After dialysis, the samples were flash-frozen and preserved at −80°C in small aliquots until use.

Enzymatic assays. Aliquots of purified LJ0536 esterase was dialyzed against 5mM BES [N,N-bis (2-hydroxyethyl)-2-aminoethanesulfonic acid] buffer (pH 7.2). The enzymatic activity was determined using a colorimetric method using the model substrate p-nitrophenyl butyrate. Flurbiprofen was

solubilized in 10% methanol. At this methanol concentration, the enzyme activity was not affected. A typical enzymatic reaction mixture contained 7.1% acetonitrile, 1mM p-nitrophenyl butyrate, 0.4 µg (600 µM) of the purified esterase, 10% methanol, and 4.5 mM BES (pH7.2). Flurbiprofen was tested at 5, 10, 50, 100, 200, and 400 µM. The enzymatic assays were performed with 200 µl in 96-well plates in a Synergy HT Biotek reader equipped with temperature control unit. The effect of flurbiprofen on the esterase activity was evaluated with and without pre-incubation with the enzyme for 2h at room temperature. The increase in absorbance was continuously monitored during 30 min at 412 nm, and the concentration of p-nitrophenol was estimated by using the extinction coefficient (ε = 16,300 M−1 cm−1). All assays and controls were performed in triplicate. Results are shown as means of 3 to 5 determinations, ± standard deviations. Statistical significance was determined by analysis of variance (ANOVA one-way). In all experiments in which a significant result was obtained, the F test was followed by Tukey's multiple comparison test.

ASSOCIATED CONTENT

Supporting Information

Most frequent ligands in PPI cavities; Set of rules implemented to filter PDB ligand by drug-likeness; Selection of best matching pairs between interfacial and sc-PDB cavities; Selection of best matching pairs between rim and sc-PDB cavities; Selection of best matching pairs between orthosteric and sc-PDB cavities; Querying the PPIome database (http://drugdesign.unistra.fr/ppiome); Visualizing the interface between the ADP-ribosylation factor 1 (Arf1, chain A) and the Arf exchange factor (Arno, chain E) bound to the PPI stabilizer brefeldin A; Distribution of molecular properties for ligands (drug-like, pharmacological tools) bound to druggable PPI cavities; Distribution of the ligand globularity PBF score for PPI modulators (Interfacial, Rim, Allosteric, Orthosteric) bound to druggable cavities (this work), conventional drug-like ligands (sc-PDB) and PPI inhibitors (2P2IDB).

AUTHOR INFORMATION

Corresponding author

E-mail: rognan@unistra.fr

ORCID

Didier Rognan : 0000-0002-0577-641X

ACKNOWLEDGMENTS

This work was funded by the French National Research Agency (ANR) through the Programme d'Investissement d'Avenir under contract ANR-10-LABX-0034. We acknowledge the Calculation center of IN2P3-CNRS (Villeurbanne, France) for allocation of computing time and excellent support

ABBREVIATIONS USED

3D, three-dimensional; PDB, Protein Data Bank; PPI, protein-protein interaction; RMSD, root-mean square deviation; SVM, support vector machine.

REFERENCES

1. Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K.; Muthusamy, B.; Leal-Rojas, P.; Kumar, P.; Sahasrabuddhe, N. A.; Balakrishnan, L.; Advani, J.; George, B.; Renuse, S.; Selvan, L. D.; Patil, A. H.; Nanjappa, V.; Radhakrishnan, A.; Prasad, S.; Subbannayya, T.; Raju, R.; Kumar, M.; Sreenivasamurthy, S. K.; Marimuthu, A.; Sathe, G. J.; Chavan, S.; Datta, K. K.; Subbannayya, Y.; Sahu, A.; Yelamanchi, S. D.; Jayaram, S.; Rajagopalan, P.; Sharma, J.; Murthy, K. R.; Syed, N.; Goel, R.; Khan, A. A.; Ahmad, S.; Dey, G.; Mudgal, K.; Chatterjee, A.; Huang, T. C.; Zhong, J.; Wu, X.; Shaw, P. G.; Freed, D.; Zahari, M. S.;

Mukherjee, K. K.; Shankar, S.; Mahadevan, A.; Lam, H.; Mitchell, C. J.; Shankar, S. K.; Satishchandra, P.; Schroeder, J. T.; Sirdeshmukh, R.; Maitra, A.; Leach, S. D.; Drake, C. G.; Halushka, M. K.; Prasad, T. S.; Hruban, R. H.; Kerr, C. L.; Bader, G. D.; Iacobuzio-Donahue, C. A.; Gowda, H.; Pandey, A., A draft map of the human proteome. *Nature,* **2014,** *509*, 575-581.

2. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrian-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magarinos, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R., The ChEMBL database in 2017. *Nucleic Acids Res,* **2017,** *45*, D945-D954.

3. Legrain, P.; Rain, J. C., Twenty years of protein interaction studies for biological function deciphering. *J Proteomics,* **2014,** *107*, 93-97.

4. Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J., Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nat Rev Drug Discov,* **2016,** *15*, 533-550.

5. Li, Z.; Ivanov, A. A.; Su, R.; Gonzalez-Pecchi, V.; Qi, Q.; Liu, S.; Webber, P.; McMillan, E.; Rusnak, L.; Pham, C.; Chen, X.; Mo, X.; Revennaugh, B.; Zhou, W.; Marcus, A.; Harati, S.; Chen, X.; Johns, M. A.; White, M. A.; Moreno, C.; Cooper, L. A.; Du, Y.; Khuri, F. R.; Fu, H., The OncoPPi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. *Nat Commun,* **2017,** *8*, 14356.

6. Arkin, M. R.; Wells, J. A., Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov,* **2004,** *3*, 301-317.

7. Bosc, N.; Kuenemann, M. A.; Becot, J.; Vavrusa, M.; Cerdan, A. H.; Sperandio, O., Privileged substructures to modulate protein-protein interactions. *J Chem Inf Model,* **2017,** *57*, 2448-2462.

8. Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H., Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov,* **2016,** *15*, 605-619.

9. Thiel, P.; Kaiser, M.; Ottmann, C., Small-molecule stabilization of protein-protein interactions: an underestimated concept in drug discovery? *Angew Chem Int Ed Engl,* **2012,** *51*, 2012-2018.

10. Venkatesan, K.; Rual, J. F.; Vazquez, A.; Stelzl, U.; Lemmens, I.; Hirozane-Kishikawa, T.; Hao, T.; Zenkner, M.; Xin, X.; Goh, K. I.; Yildirim, M. A.; Simonis, N.; Heinzmann, K.; Gebreab, F.; Sahalie, J. M.; Cevik, S.; Simon, C.; de Smet, A. S.; Dann, E.; Smolyar, A.; Vinayagam, A.; Yu, H.; Szeto, D.; Borick, H.; Dricot, A.; Klitgord, N.; Murray, R. R.; Lin, C.; Lalowski, M.; Timm, J.; Rau, K.; Boone, C.; Braun, P.; Cusick, M. E.; Roth, F. P.; Hill, D. E.; Tavernier, J.; Wanker, E. E.; Barabasi, A. L.; Vidal, M., An empirical framework for binary interactome mapping. *Nat Methods,* **2009,** *6*, 83-90.

11. Stumpf, M. P.; Thorne, T.; de Silva, E.; Stewart, R.; An, H. J.; Lappe, M.; Wiuf, C., Estimating the size of the human interactome. *Proc Natl Acad Sci U S A,* **2008,** *105*, 6959-6964.

12. Luck, K.; Sheynkman, G. M.; Zhang, I.; Vidal, M., Proteome-scale human interactomics. *Trends Biochem Sci,* **2017,** *42*, 342-354.

13. Alonso-Lopez, D.; Campos-Laborie, F. J.; Gutierrez, M. A.; Lambourne, L.; Calderwood, M. A.; Vidal, M.; De Las Rivas, J., APID database: redefining protein-protein interaction experimental evidences and binary interactomes. *Database-Oxford,* **2019**, doi: 10.1093/database/baz005.

14. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The protein data bank. *Nucleic Acids Res,* **2000,** *28*, 235-242.

15. Capitani, G.; Duarte, J. M.; Baskaran, K.; Bliven, S.; Somody, J. C., Understanding the fabric of protein crystals: computational classification of biological interfaces and crystal contacts. *Bioinformatics,* **2016,** *32*, 481-489.

16. Ehrt, C.; Brinkjost, T.; Koch, O., Impact of binding site comparisons on medicinal chemistry and rational molecular design. *J Med Chem,* **2016,** *59*, 4121-4151.

17. Pagadala, N. S.; Syed, K.; Tuszynski, J., Software for molecular docking: a review. *Biophys Rev,* **2017,** *9*, 91-102.

18. Schneider, P.; Schneider, G., De novo design at the edge of chaos. *J Med Chem,* **2016,** *59*, 4077-4086.

19. Löwer, M.; Proschak, E., Structure-based pharmacophores for virtual screening. *Mol Inform,* **2011,** *30*, 398-404.

20. Desaphy, J.; Bret, G.; Rognan, D.; Kellenberger, E., sc-PDB: a 3D-database of ligandable binding sites--10 years on. *Nucleic Acids Res,* **2015,** *43*, D399-404.

21. Da Silva, F.; Desaphy, J.; Bret, G.; Rognan, D., IChemPIC: A random forest classifier of biological and crystallographic protein-protein interfaces. *J Chem Inf Model,* **2015,** *55*, 2005-2014.

22. Bogan, A. A.; Thorn, K. S., Anatomy of hot spots in protein interfaces. *J Mol Biol,* **1998,** *280*, 1-9.

23. Chakrabarti, P.; Janin, J., Dissecting protein-protein recognition sites. *Proteins,* **2002,** *47*, 334-343.

24. Desaphy, J.; Azdimousa, K.; Kellenberger, E.; Rognan, D., Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. *J Chem Inf Model,* **2012,** *52*, 2287-2299.

25. Tran-Nguyen, V. K.; Da Silva, F.; Bret, G.; Rognan, D., All in one: cavity Detection, druggability estimate, cavity-based pharmacophore perception, and virtual screening. *J Chem Inf Model,* **2019,** *59*, 573-585.

26. Shindyalov, I. N.; Bourne, P. E., Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng,* **1998,** *11*, 739-747.

27. Basse, M. J.; Betzi, S.; Morelli, X.; Roche, P., 2P2Idb v2: update of a structural database dedicated to orthosteric modulation of protein-protein interactions. *Database (Oxford),* **2016,** *2016*.

28. Gao, M.; Skolnick, J., The distribution of ligand-binding pockets around protein-protein interfaces suggests a general mechanism for pocket formation. *Proc Natl Acad Sci U S A,* **2012,** *109*, 3784-3789.

29. Lawson, A. D. G.; MacCoss, M.; Heer, J. P., Importance of rigidity in designing small molecule drugs to tackle protein-protein interactions (PPIs) through stabilization of desired conformers. *J Med Chem,* **2018,** *61*, 4283-4289.

30. Da Silva, F.; Desaphy, J.; Rognan, D., IChem: A versatile toolkit for detecting, comparing, and predicting protein-ligand interactions. *ChemMedChem,* **2018,** *13*, 507-510.

31. Illing, P. T.; Vivian, J. P.; Dudek, N. L.; Kostenko, L.; Chen, Z.; Bharadwaj, M.; Miles, J. J.; Kjer-Nielsen, L.; Gras, S.; Williamson, N. A.; Burrows, S. R.; Purcell, A. W.; Rossjohn, J.; McCluskey, J., Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. *Nature,* **2012,** *486*, 554-558.

32. Hetherington, S.; McGuirk, S.; Powell, G.; Cutrell, A.; Naderer, O.; Spreen, B.; Lafon, S.; Pearce, G.; Steel, H., Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. *Clin Ther,* **2001,** *23*, 1603-1614.

33. Melcher, K.; Xu, Y.; Ng, L. M.; Zhou, X. E.; Soon, F. F.; Chinnusamy, V.; Suino-Powell, K. M.; Kovach, A.; Tham, F. S.; Cutler, S. R.; Li, J.; Yong, E. L.; Zhu, J. K.; Xu, H. E., Identification and mechanism of ABA receptor antagonism. *Nat Struct Mol Biol,* **2010,** *17*, 1102-1108.

34. Lai, K. K.; Stogios, P. J.; Vu, C.; Xu, X.; Cui, H.; Molloy, S.; Savchenko, A.; Yakunin, A.; Gonzalez, C. F., An inserted alpha/beta subdomain shapes the catalytic pocket of lactobacillus johnsonii cinnamoyl esterase. *PLoS One,* **2011,** *6*, e23269.

35. Selinsky, B. S.; Gupta, K.; Sharkey, C. T.; Loll, P. J., Structural analysis of NSAID binding by prostaglandin H2 synthase: time-dependent and time-independent inhibitors elicit identical enzyme conformations. *Biochemistry,* **2001,** *40*, 5172-5180.

36. Kurumbail, R. G.; Stevens, A. M.; Gierse, J. K.; McDonald, J. J.; Stegeman, R. A.; Pak, J. Y.; Gildehaus, D.; Miyashiro, J. M.; Penning, T. D.; Seibert, K.; Isakson, P. C.; Stallings, W. C., Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature,* **1996,** *384*, 644-648.

37. Perez-Salvia, M.; Esteller, M., Bromodomain inhibitors and cancer therapy: from structures to applications. *Epigenetics,* **2017,** *12*, 323-339.

38. Wendt, M. D., Protein-Protein Interactions as Drug Targets. In *Protein-Protein Interactions*, Wendt, M. D., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp 1-55.

39. Milhas, S.; Raux, B.; Betzi, S.; Derviaux, C.; Roche, P.; Restouin, A.; Basse, M. J.; Rebuffet, E.; Lugari, A.; Badol, M.; Kashyap, R.; Lissitzky, J. C.; Eydoux, C.; Hamon, V.; Gourdel, M. E.; Combes, S.; Zimmermann, P.; Aurrand-Lions, M.; Roux, T.; Rogers, C.; Muller, S.; Knapp, S.; Trinquet, E.; Collette,

Y.; Guillemot, J. C.; Morelli, X., Protein-protein interaction inhibition (2P2I)-oriented chemical library accelerates hit discovery. *ACS Chem Biol,* **2016,** *11*, 2140-2148.

40. Kuenemann, M. A.; Labbe, C. M.; Cerdan, A. H.; Sperandio, O., Imbalance in chemical space: how to facilitate the identification of protein-protein interaction inhibitors. *Sci Rep,* **2016,** *6*, 23815.

41. Bickerton, G. R.; Higueruelo, A. P.; Blundell, T. L., Comprehensive, atomic-level characterization of structurally characterized protein-protein interactions: the PICCOLO database. *BMC Bioinformatics,* **2011,** *12*, 313.

42. Mathew, O. K.; Sowdhamini, R., PIMADb: a database of protein-protein interactions in huge macromolecular assemblies. *Bioinform Biol Insights,* **2016,** *10*, 105-109.

43. Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C., Binding of small molecules to an adaptive protein-protein interface. *Proc Natl Acad Sci U S A,* **2003,** *100*, 1603-1608.

44. Fischer, E. S.; Park, E.; Eck, M. J.; Thoma, N. H., SPLINTS: small-molecule protein ligand interface stabilizers. *Curr Opin Struct Biol,* **2016,** *37*, 115-122.

45. Zarzycka, B.; Kuenemann, M. A.; Miteva, M. A.; Nicolaes, G. A. F.; Vriend, G.; Sperandio, O., Stabilization of protein-protein interaction complexes through small molecules. *Drug Discov Today,* **2016,** *21*, 48-57.

46. Nogales, E.; Wolf, S. G.; Khan, I. A.; Luduena, R. F.; Downing, K. H., Structure of tubulin at 6.5 A and location of the taxol-binding site. *Nature,* **1995,** *375*, 424-427.

47. Lipinski, C. A., Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol,* **2004,** *1*, 337-341.

48. Firth, N. C.; Brown, N.; Blagg, J., Plane of best fit: a novel method to characterize the three-dimensionality of molecules. *J Chem Inf Model,* **2012,** *52*, 2516-2525.

49. Cazorla, M.; Premont, J.; Mann, A.; Girard, N.; Kellendonk, C.; Rognan, D., Identification of a low-molecular weight TrkB antagonist with anxiolytic and antidepressant activity in mice. *J Clin Invest,* **2011,** *121*, 1846-1857.

50. Rivat, C.; Sar, C.; Mechaly, I.; Leyris, J. P.; Diouloufet, L.; Sonrier, C.; Philipson, Y.; Lucas, O.; Mallie, S.; Jouvenel, A.; Tassou, A.; Haton, H.; Venteo, S.; Pin, J. P.; Trinquet, E.; Charrier-Savournin, F.; Mezghrani, A.; Joly, W.; Mion, J.; Schmitt, M.; Pattyn, A.; Marmigere, F.; Sokoloff, P.; Carroll, P.; Rognan, D.; Valmier, J., Inhibition of neuronal FLT3 receptor tyrosine kinase alleviates peripheral neuropathic pain in mice. *Nat Commun,* **2018,** *9*, 1042.

51. Fischer, G.; Rossmann, M.; Hyvonen, M., Alternative modulation of protein-protein interactions by small molecules. *Curr Opin Biotechnol,* **2015,** *35*, 78-85.

52. Renaud, J. P.; Chari, A.; Ciferri, C.; Liu, W. T.; Remigy, H. W.; Stark, H.; Wiesmann, C., Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat Rev Drug Discov,* **2018,** *17*, 471-492.

53. Velankar, S.; Dana, J. M.; Jacobsen, J.; van Ginkel, G.; Gane, P. J.; Luo, J.; Oldfield, T. J.; O'Donovan, C.; Martin, M. J.; Kleywegt, G. J., SIFTS: structure integration with function, taxonomy and sequences resource. *Nucleic Acids Res,* **2013,** *41*, D483-489.

54. The UniProt Consortium, UniProt: the universal protein knowledgebase. *Nucleic Acids Res,* **2018,** *46*, 2699.

55. Bietz, S.; Urbaczek, S.; Schulz, B.; Rarey, M., Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. *J Cheminform,* **2014,** *6*, 12.

56. Marcou, G.; Rognan, D., Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model,* **2007,** *47*, 195-207.

57. Lai, K. K.; Lorca, G. L.; Gonzalez, C. F., Biochemical properties of two cinnamoyl esterases purified from a Lactobacillus johnsonii strain isolated from stool samples of diabetes-resistant rats. *Appl Environ Microbiol,* **2009,** *75*, 5018-5024.

58. Lorca, G. L.; Ezersky, A.; Lunin, V. V.; Walker, J. R.; Altamentova, S.; Evdokimova, E.; Vedadi, M.; Bochkarev, A.; Savchenko, A., Glyoxylate and pyruvate are antagonistic effectors of the Escherichia coli IclR transcriptional regulator. *J Biol Chem,* **2007,** *282*, 16476-16491.

59. Gonzalez, C. F.; Proudfoot, M.; Brown, G.; Korniyenko, Y.; Mori, H.; Savchenko, A. V.; Yakunin, A. F., Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from Escherichia coli, FrmB and YeiG. *J Biol Chem,* **2006,** *281*, 14514-14522.

FOR TABLE OF CONTENTS USE ONLY

Exhaustive Repertoire of Druggable Cavities at Protein-Protein Interfaces of Known Three-Dimensional Structure

Franck Da Silva, Guillaume Bret, Leandro Teixeira, Claudio F. Gonzalez and Didier Rognan