

# Which Three-Dimensional Characteristics Make Efficient Inhibitors of Protein–Protein Interactions?

Mélaine A. Kuenemann,<sup>†,‡,○</sup> Laura M. L. Bourbon,<sup>†,‡,○</sup> Céline M. Labbé,<sup>†,‡,§,○</sup>  
Bruno O. Villoutreix,<sup>†,‡,§,○</sup> and Olivier Sperandio<sup>\*</sup>

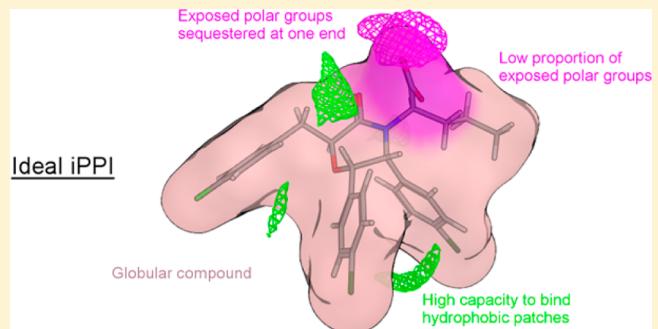
<sup>†</sup>Université Paris Diderot, Sorbonne Paris Cité, UMRS 973 Inserm, Paris 75013, France

<sup>‡</sup>Inserm, U973, Paris 75013, France

<sup>§</sup>Faculté de Pharmacie, CDithem, 1 rue du Prof Laguesse, 59000 Lille, France

## Supporting Information

**ABSTRACT:** The specific properties of protein–protein interactions (PPI) (flat, large and hydrophobic) make them harder to tackle with low-molecular-weight compounds. Learning from the properties of successful examples of PPI interface inhibitors (iPPI) at earlier stages of developments, has been pinpointed as a powerful strategy to circumvent this trend. To this end, we have computationally analyzed the bioactive conformations of iPPI and those of inhibitors of conventional targets (e.g enzymes) to highlight putative iPPI 3D characteristics. Most noticeably, the essential property revealed by this study illustrates how efficiently iPPI manages to bind to the hydrophobic patch often present at the core of protein interfaces. The newly identified properties were further confirmed as characteristics of iPPI using much larger data sets (e.g iPPI-DB, [www.ippidb.cdithem.fr](http://www.ippidb.cdithem.fr)). Interestingly, the absence of correlation of such properties with the hydrophobicity and the size of the compounds opens new ways to design potent iPPI with better pharmacokinetic features.



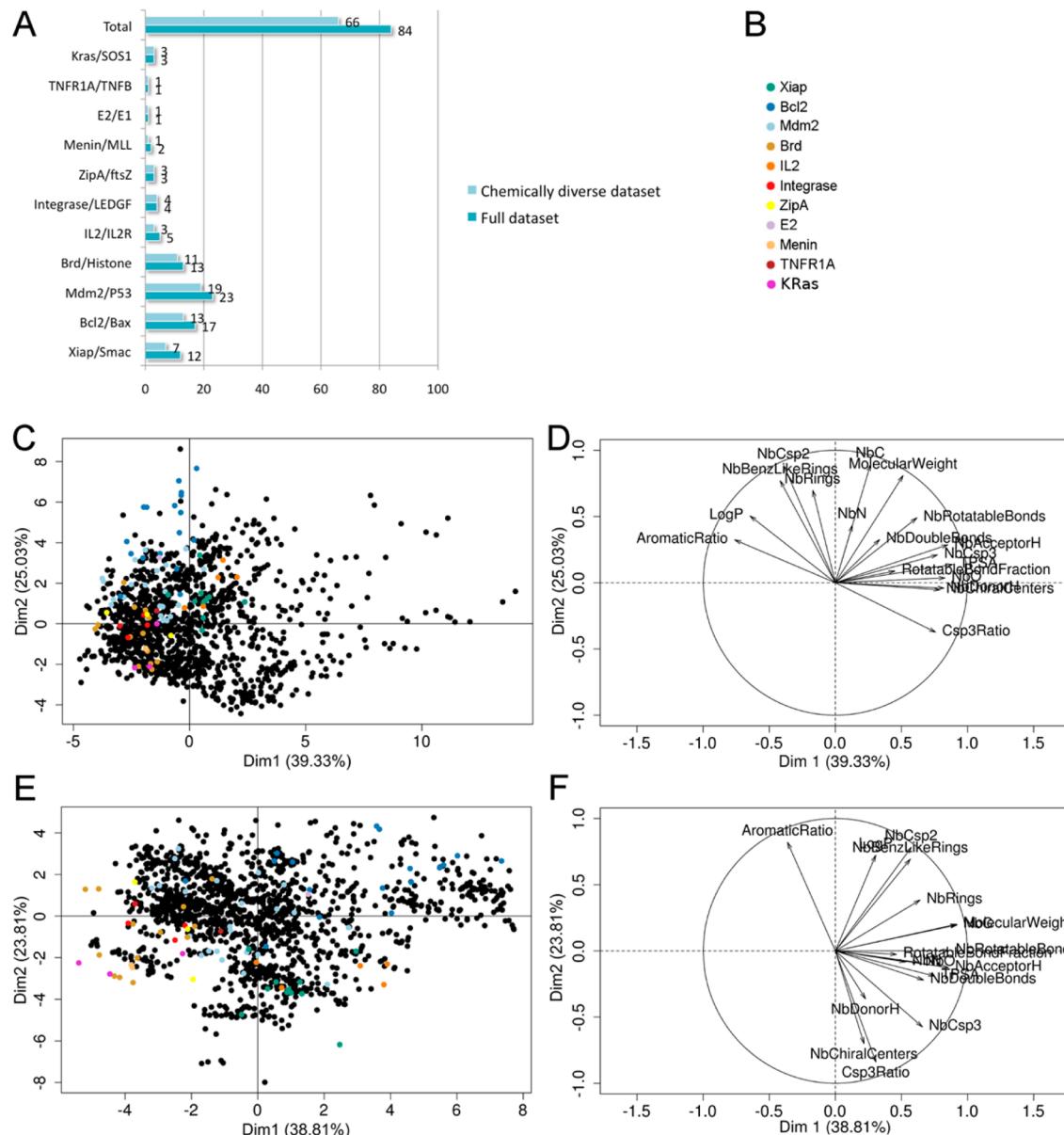
## INTRODUCTION

In the past 50 years, drug discovery was focused on a relatively low number of targets<sup>1–3</sup> including G-Protein Coupled Receptors (GPCR), nonkinase enzymes, ion channels, nuclear receptors, and more recently kinases. In 2006, these represented only about 324 drug targets.<sup>4</sup> This number stands in sharp contrast with the about 370 000 protein–protein interactions (PPI)<sup>5</sup> predicted in humans, not to mention the transorganism PPI. Within this vast interactome, there is a strong hope to see the emergence of more and more therapeutic targets associated with various diseases and cellular mechanisms, as illustrated by the existing PPI targets spanning through biological recognition, pathogenicity, enzyme regulation, apoptosis, immune system, etc.

PPI interfaces are usually described as larger, flatter, and more hydrophobic than more conventional targets (e.g GPCR and enzymes). These properties are thought to regulate the formation of their functional protein complexes, whose predictions still remain a major challenge using in silico protocols.<sup>6</sup> The PPI interface tends to lack a single clearly defined low-molecular-weight ligandable binding cavity<sup>7</sup> but instead presents several small pockets at or near the interface area.<sup>8</sup> Such features do not facilitate ligand binding and therefore impede the development of low-molecular-weight drugs. Indeed, high throughput screening (HTS) technologies (commonly used to identify the initial hits on a drug target)

using traditional screening compound collections most often lead to poor hit rates and/or high false positive rates in the case of PPI targets.<sup>9</sup> The direct consequence of applying such strategies is to end up with a very short list of high quality starting hits, if any. This usually impedes the following steps of development and significantly reduces the chance of success. Yet, in order to facilitate the emergence of drugs on PPI targets, it is first necessary to boost the identification of the initial hits. To this end, the chemical properties that promote the binding of low-molecular-weight compounds to relatively flat PPI interfaces must be elucidated. Noticeably, years of development and the accumulation of successful iPPI on various PPI targets have given the scientific community the opportunity to get some insights into the binding profile of such compounds. Several hundred iPPI, sometimes identified using innovative strategies,<sup>10</sup> are now available in the literature and can be found in several databases, iPPI-DB (<http://www.ippidb.cdithem.fr/>; mostly hit compounds but this also includes about 30 more advanced drug candidates),<sup>11</sup> 2P2Idb,<sup>12</sup> and Timbal.<sup>13</sup> All these databases can be used to help to rationalize key physicochemical properties, to highlight privileged chemotypes, and to understand the singular features that make them capable of binding to such intricate interfaces. Previous studies have

Received: August 6, 2014



**Figure 1.** Distribution of iPPI across PPI families and principal component analysis (PCA) on the data sets. (A) Number of compounds and diverse compounds per PPI complexes. (B) PPI family represented by the protein that is bound by the compound and associate with the color used for the PCA. (C–F) PCA calculations were run using 18 physicochemical properties (Table S2). (C) PCA on iPPI versus non-iPPI compounds of the X-ray data set: individual map of compounds, PPI inhibitors are represented as color dots, and non-PPI inhibitors are represented as black dots. (D) PCA on iPPI versus non-iPPI compounds of the X-ray data set: variable map. (E) PCA on the iPPI of the X-ray data set versus the iPPI-DB compounds: individual map of compounds, X-ray iPPI are represented as color dots, and iPPI-DB compounds are represented as black dots. (F) PCA on the iPPI of the X-ray data set versus the iPPI-DB compounds: variable map.

already shown that iPPI have unusual physicochemical properties<sup>14,15</sup> like a higher molecular weight, higher hydrophobicity and higher aromaticity to name a few. These studies provide a rationale to explain the low success rates of previously reported HTS campaigns. They also suggest that the chemical collections used this far on PPI targets were historically designed for conventional targets (GPCR, enzyme, etc.). Such chemical collections usually have a Lipinski's ROS<sup>16</sup> compliant profile and do certainly not comply with the electronic and the topological constraints of PPI interfaces. Thus, one can measure the leap in chemical space that needs to be undertaken to have the physicochemical properties of iPPI better represented in the commercial chemical screening libraries, including those used in HTS campaigns. But more importantly,

such analysis of the iPPI properties calls for the definition of new rules that assist the identification of novel inhibitors presenting an initial profile more compatible with compound optimizations moving away from the so-called molecular obesity.<sup>17</sup>

Among the properties that are known to determine the binding capacity of a compound to a target, 3D characteristics have been identified as critical.<sup>18</sup> Indeed, shape is a fundamental molecular feature that will often determine the fate of a compound in terms of molecular interactions with eventually privileged biological targets.<sup>19</sup> More specifically for PPI interfaces, shape has been continuously highlighted as a key feature for pharmacological modulation and therefore justifies a primary focus.<sup>20–22</sup>

The objective of this study is thus to attempt to detect which 3D characteristics the structures of iPPI may possess that maximize their binding capacity to PPI interfaces. Conversely, the most documented properties about iPPI, high molecular weight and high hydrophobicity, are known to be a potential liability for drug development. Within this analysis, we have therefore attempted to discover new characteristics that do not rely on those two properties. In the present study, iPPI were collected in their bioactive conformation and compared to inhibitors of conventional targets. Four 3D characteristics were highlighted. They describe either the shape of the compounds (globularity) or the 3D distributions of the hydrophobic and hydrophilic interacting regions of the compounds (VolSurf descriptors<sup>23</sup>). More specifically the most essential property revealed in the analysis (EDmin3) illustrates how iPPI manage to bind to the hydrophobic patches often present at the core of PPI targets. Our study shows that even iPPI with a modest hydrophobicity, like those acting on the Interleukin 2 (IL2)/Interleukin 2 alpha Receptor (IL2R $\alpha$ ) interaction or the X-linked inhibitor of apoptosis protein (Xiap)/second mitochondria-derived activator of caspases (smac) interaction, manage to display a critical capacity to interact with a PPI hydrophobic patch due to the proper layout of their hydrophobic interacting regions. These four properties were first analyzed on 3D bioactive conformations and then confirmed on larger data sets starting from 2D conformations including the data of iPPI-DB.<sup>11</sup> Interestingly, the absence of correlation of such properties with the hydrophobicity and the size of the compounds could allow the development of compounds maximizing such profile but with a better balance in terms of ligand efficiency and lipophilic efficiency.<sup>24</sup>

## RESULTS AND DISCUSSION

**Construction of the Dataset.** In order to investigate the 3D characteristics of iPPI, we have constructed on one hand a data set of cocrystallized iPPI (i.e. bioactive conformations) mostly from 2P2Idb<sup>12</sup> referred to as the iPPI X-ray data set and on the other hand cocrystallized inhibitors of conventional targets, named here the non-iPPI X-ray data set (non-iPPI from PDBind<sup>25</sup>). In order to remove any fragment hits (as opposed to full compound hits) we removed from the data sets all molecules with a molecular weight below 250 g/mol. We ended up with 84 iPPI and 1282 inhibitors of conventional targets (list of pdb codes: Table S1). Furthermore, to prevent any chemical bias and prevent the overrepresentation of any given chemical series in the selection of specific 3D properties, we have also used a fingerprint-based clustering approach (see Methods section Data Set Compilation) to build a chemically diverse subset that eventually contains 66 representative iPPI (out of 84) and 878 representative (out of 1282) non-iPPI (Figure 1A). Both data sets (full and diverse chemically) were jointly used in the analysis and in the identification of new 3D characteristics.

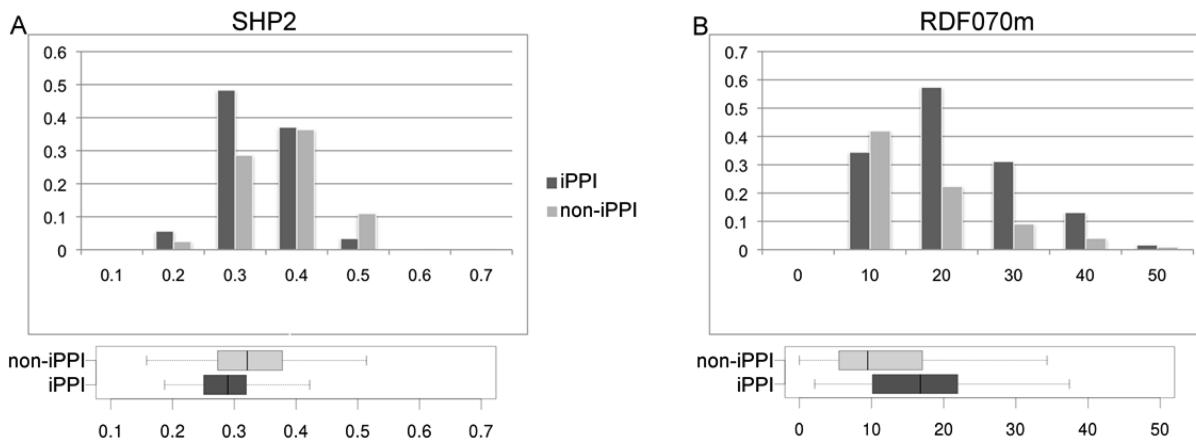
The profile of each individual PPI target was also considered through the properties of its corresponding compounds to determine which properties were essential regardless of the type of PPI targets, and which properties were target specific. The data set contains 11 different targets that are not equally populated (Figure 1A). Six targets contain more than 3 compounds (Xiap/Smac, apoptosis regulator (Bcl2)/apoptosis regulator (Bax), bromodomain-containing protein (Brd)2/3/4/T/(Histone), Minute double mutant 2 (Mdm2)/tumor suppressor (p53), Lens Epithelium-Derived Growth Factor

(LEDGF)/human (Integrase)]; and IL2/IL2R $\alpha$ ). Five PPI targets contain 3 or less compounds (Tumor Necrosis Factor (TNFR1A) receptor 1A/human Tumor Necrosis Factor beta (TNFB), Human papillomavirus (HPV) E1/E2 and HIV-1, Menin/Mixed Lineage Leukemia (MLL), Growth factor receptors activate (KRas)/factor son of sevenless (SOS1) and cell division protein (ZipA)/cell division protein (ftsZ)) and were therefore not considered in our statistical analysis (i.e except for visualization) for target specificity as suggested in ref 26.

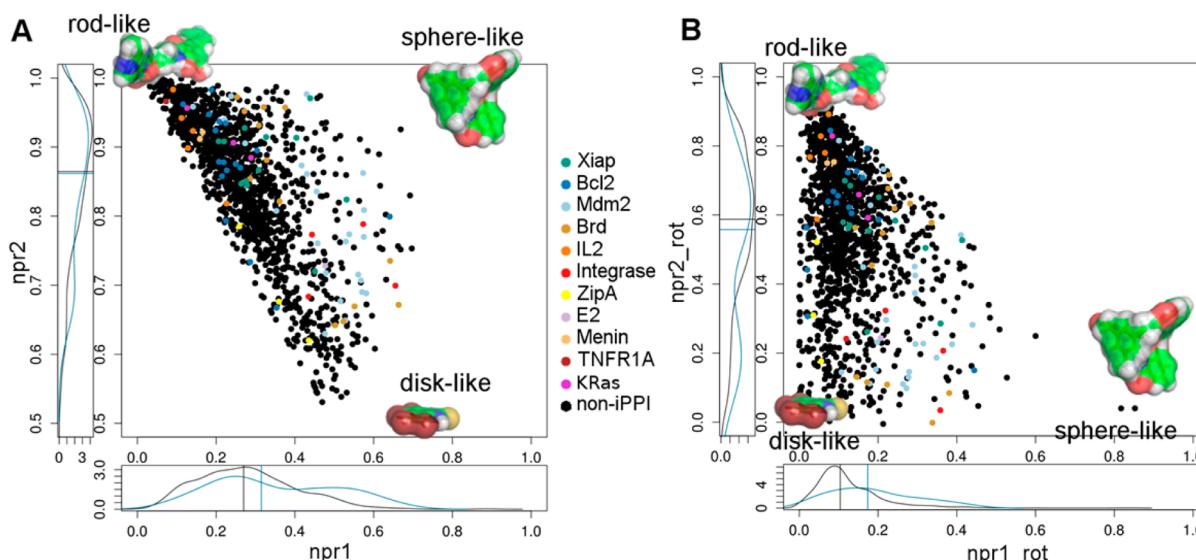
**Data Set Inspection.** We ran two PCAs using 18 commonly used molecular descriptors (Table S2) first between the iPPI and the non-iPPI of the X-ray data set and second between the iPPI of our X-ray data set and the iPPI compounds from iPPI-DB.<sup>11</sup>

The first PCA was calculated to visualize the chemical space of our full X-ray data set but also to confirm the already known differences between iPPI and non-iPPI. The two first axes of the PCA represent 64.36% of the total variance. Therefore, it provides a fair visualization of the corresponding chemical space (Figure 1C and D). The first axis is characterized by the polarity and the complexity of compounds (topological polar surface area TPSA, number of H-bond acceptors or donors, and number of sp<sup>3</sup> carbons). The second axis is represented by the size (number of carbon atoms) and the compounds' aromaticity (number of benzene-like rings and number of sp<sup>2</sup> carbons). The global position of the iPPI population is in the upper left part of the individual map. This confirms that, as a general trend, iPPI are more hydrophobic, more aromatic, and with a higher size than conventional inhibitors. This is in agreement with previously reported profiles.<sup>9,14,15</sup> The dissimilarity between these populations logically illustrates the properties of PPI interfaces, known to be in general flatter and more hydrophobic than the binding regions of conventional targets. But interestingly, this representation also highlights poorly documented observations such as the differences of compound profiles across PPI targets. Indeed, it is clear from Figure 1C that not all PPI targets are equivalent when it comes to the properties of their corresponding iPPI. They present a different profile with respect to the inhibitors of conventional targets without being characterized by a fully homogeneous profile of their own. For example, the Bcl2 compounds are clearly more aromatic (e.g. higher number of sp<sup>2</sup> carbon and benzene-like rings) and with a higher molecular size, while the compounds of Mdm2 are characterized by higher proportion of aromatic rings but not necessarily by higher sizes. Conversely, Xiap compounds are clearly more complex, more polar, and not as aromatic.

A second PCA was calculated between the iPPI of our X-ray data set and the iPPI-DB compounds. iPPI-DB v1 ([www.ippidb.cdithem.fr](http://www.ippidb.cdithem.fr)) is a database containing 1650 iPPI manually curated with an activity (e.g. IC<sub>50</sub>) under 30  $\mu$ M across 13 families of PPI targets<sup>11</sup> and constitutes a representative overview of the available iPPI in the literature. This PCA was run with the same 18 descriptors (Table S2) to confirm that our data set, although containing 84 iPPI covers a similar chemical space than compounds from iPPI-DB (Figure 1E and F). Combination of the first and second axis presents 62.62% of variance, which constitutes a fair representation of the corresponding chemical space. It is clear from this individual map that the iPPI of our X-ray data set are sparsely distributed and occupy similar regions of the chemical space than the iPPI-



**Figure 2.** Distribution of SHP2 and RDF070m for iPPI and non-iPPI. (A) Distribution of SHP2 in darkgrey for iPPI (histogram and boxplot) and gray for non-iPPI (histogram and boxplot). (B) Distribution of RDF070m in darkgrey for iPPI (histogram and boxplot) and gray for non-iPPI (histogram and boxplot).



**Figure 3.** Principal moments of inertia represented as a 2D-plot with non-iPPI and iPPI. (A) The normal plot. (B) The 45° rotated normalized plot. Each compound is represented by a dot through its normalized moments of inertia coordinates into a plot (in the form of a triangle) whose extreme accessible values correspond to extreme shapes (rod-, disk-, and sphere-like). iPPI compounds are colored by their corresponding target, and non-iPPI are in black. Density curves are near the X- and Y-axis highlight the discrepancies according to both axis between iPPI and non-iPPI as a whole along with a straight line within them to mark the median value of each population (blue for iPPI and black for non-iPPI).

DB compounds. This confirms that our X-ray data set seems to be a representative set of iPPI.

**Continuity with Previous Works on iPPI.** In 2007, Neugebauer et al.<sup>20</sup> identified SHP2 as important for iPPI. SHP2 is referred to as a shape descriptor and was introduced by Randic.<sup>27</sup> It is an average of various lower-level descriptors that are derived from the interatomic distances of the atoms at the periphery of a molecule, therefore susceptible to contribute to protein binding. In 2010, our group had shown that RDF070m is a specific shape descriptor for iPPI.<sup>15,21</sup> RDF070m is a radial distribution function (RDF) weighted by the atomic masses. It represents the probability of finding a compound atom on a sphere of radius 7 Å. Values of this descriptor dramatically increase in the case of branched compounds (e.g T-shaped or star-shaped compounds) whose structures manage to maximize such probability. But RDF070m also highlights the need for a critical size of the compound.<sup>15</sup> Such a requirement for critical size is also described in other studies.<sup>14</sup> Interestingly, historical

chemical providers such as Asinex Ltd. (<http://www.asinex.com>), Life Chemicals (<http://www.lifechemicals.com>), or Otava Chemicals (<http://www.otavachemicals.com>) now include such a property in the design of PPI-focused chemical libraries.

We therefore decided to evaluate these characteristics on our X-ray data set. Both SHP2 and RDF070m remain highly discriminative toward iPPI compounds as illustrated by the distributions and boxplots of Figure 2. Indeed, iPPI present a significantly lower value for SHP2 ( $\text{mean}_{\text{iPPI}} = 0.28 \pm 0.05$ ,  $\text{mean}_{\text{non-iPPI}} = 0.32 \pm 0.07$ , and  $P\text{-value} = 7.34 \times 10^{-6}$ ) (Figure 2A), while they are characterized by significantly higher values of RDF070m ( $\text{mean}_{\text{iPPI}} = 17.28 \pm 10.56$ ,  $\text{mean}_{\text{non-iPPI}} = 12.69 \pm 10.46$ , and  $P\text{-value} = 3.18 \times 10^{-5}$ ) (Figure 2B). We could hypothesize that the existence of more branched structures for iPPI can facilitate their binding to the combination of several protein subpockets often described at PPI interfaces.<sup>8</sup> The properties illustrated by those descriptors are confirmed as

essential for iPPI. Despite a correlation with the size of the compounds ( $\text{cor}_{\text{shp2-MW}} = -0.83$  and  $\text{cor}_{\text{RDF070m-MW}} = 0.85$ ), it was demonstrated that at equivalent molecular weight two compounds can have important variations for the values of RDF070m<sup>15,21</sup> and that higher values for this descriptor should be favored. More specifically, this descriptor brings important insights into the intrinsic structure of an ideal iPPI, even though a critical size of the compound may be required for the value to be in the desired threshold.

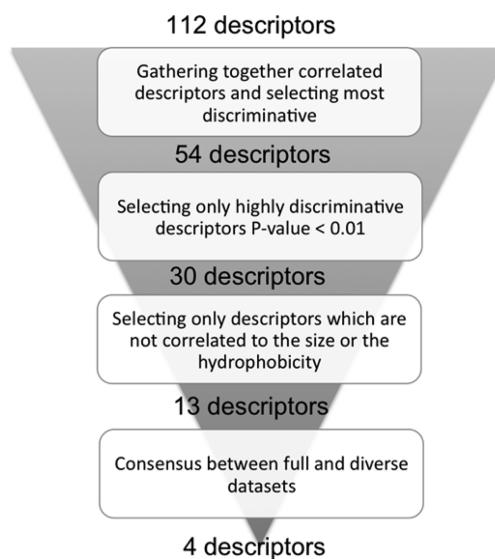
Other studies have also highlighted the better occupancy of the 3D space of iPPI using the normalized moment of inertia (npr1 and npr2).<sup>22</sup> Normalized moments of inertia were therefore calculated on our data set of bioactive conformations and plotted to distinguish the rod- (top left), disk- (bottom center), and sphere-like (top right) compounds (Figure 3A). According to this representation, there is a discriminative character on the first axis ( $P\text{-value}_{\text{npr1}} = 9.33 \times 10^{-5}$ ), with iPPI presenting a higher value of npr1 than inhibitors of conventional targets ( $\text{mean}_{\text{iPPI}} = 0.35 \pm 0.16$ ,  $\text{mean}_{\text{non-iPPI}} = 0.28 \pm 0.12$ ). A low value of npr1 represents an extended shape (rod-like) and a high value represents a sphere-like shape. But in the intermediate region of the X-axis (npr1), the plot does not help to distinguish disc- from sphere-like shapes without inspecting the Y-axis (npr2). So we applied a 45° rotation and a normalization (between 0 and 1) of the plot npr1-vs-npr2 (Figure 3B). These rotation and normalization were carried out to differentiate a flat shape (left part, rod- or disc-like) from a more tridimensional shape (right part) according to the new X-axis (npr1\_rot) alone. The new descriptor npr1\_rot better illustrates that the occupation of the 3D space of iPPI ( $\text{mean}_{\text{iPPI}} = 0.18 \pm 0.11$ ) is very significantly ( $P\text{-value} = 1.95 \times 10^{-8}$ ) superior to those of inhibitors of conventional targets ( $\text{mean}_{\text{non-iPPI}} = 0.12 \pm 0.08$ ). Besides this general trend, it is worth noting that depending on their targets, iPPI can have different profiles regarding this property as seen in Figure 3B. For example, Mdm2, Bromodomain, Xiap, E2, and Integrase inhibitors present higher npr1\_rot values of their iPPI, while IL2 compounds are more similar to conventional inhibitors.

Other studies have also used the normalized moments of inertia to characterize the 3D shape of drugs, or compounds from chemical libraries.<sup>18,28</sup> They showed that existing drugs or hits are more characterized by rod- or disc-like shapes and that more tridimensional compounds (toward the sphere-like region or higher globularity) are less inclined to become hits during experimental screenings. Analogously, we obtain a similar pattern in the case of the non-iPPI compounds, which present rod- or disc-like profiles as opposed to iPPI (more sphere-like). Noticeably, our non-iPPI subset is mainly composed of enzyme, receptor, or even kinase inhibitors. It is therefore logical that they possess a similar 3D profiles to existing drugs largely dominated by these targets. But, the results presented herein show that while globularity (more sphere-like shape) can be a liability and impede binding of compounds on regular targets, it seems to be an asset on PPI targets, at least for some of them.

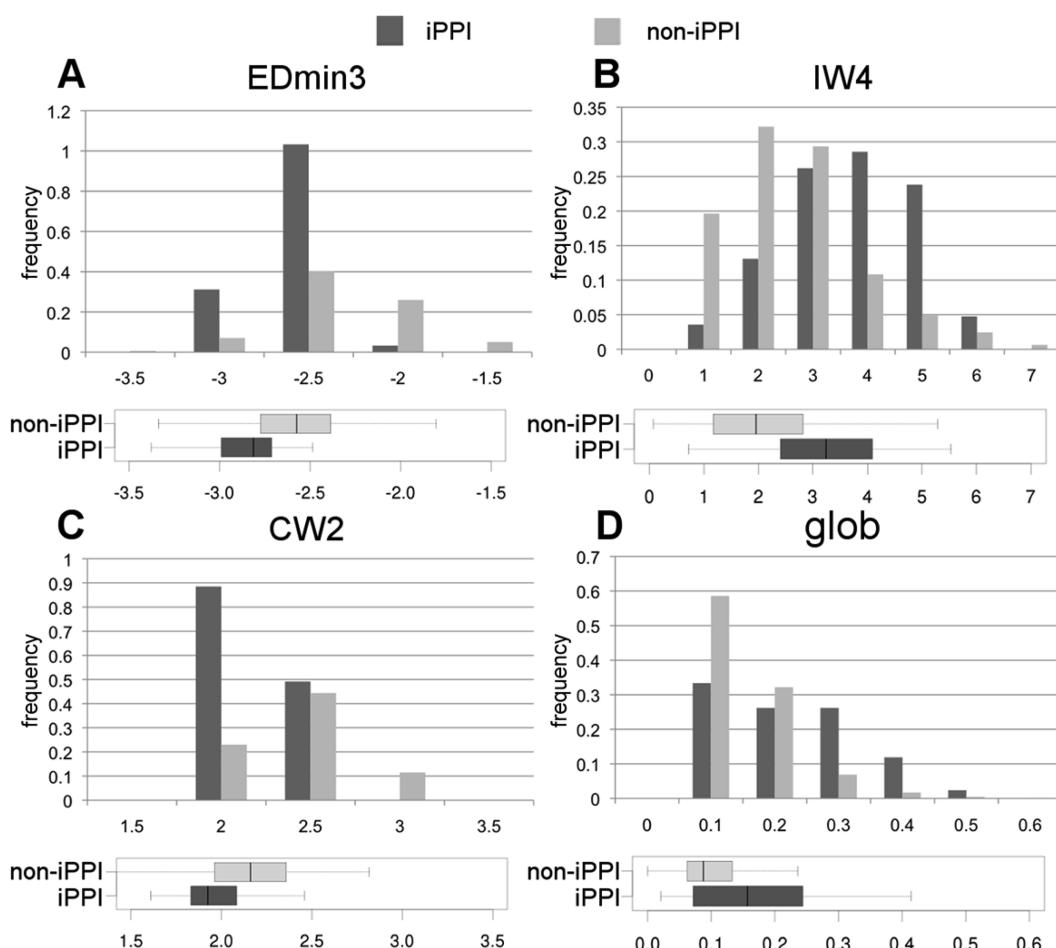
**Highlight New 3D Properties.** Previous studies have already shown that inhibitors of protein–protein interactions have specific properties like a higher size, higher hydrophobicity, or higher aromaticity.<sup>14,15</sup> This trend is confirmed again in our data set of bioactive conformations with size (molecular weight:  $\text{mean}_{\text{iPPI}} = 507.32 \pm 145.00$  g/mol;  $\text{mean}_{\text{non-iPPI}} = 435.94 \pm 141.45$  g/mol,  $P\text{-value} = 3.12 \times 10^{-5}$ ); hydrophobicity ( $\log P$ : $\text{mean}_{\text{iPPI}} = 4.21 \pm 2.27$ ;

$\text{mean}_{\text{non-iPPI}} = 1.40 \pm 3.42$ ,  $P\text{-value} = 3.35 \times 10^{-14}$ ); aromaticity (aromatic ratio: $\text{mean}_{\text{iPPI}} = 0.26 \pm 0.09$ ;  $\text{mean}_{\text{non-iPPI}} = 0.14 \pm 0.08$ ,  $P\text{-value} = 4.36 \times 10^{-3}$ ); or some specific 3D properties so far all correlated with the size of the compounds. The above-mentioned 3D properties (SHP2, RDF070m) describe in their own way the specific shape of iPPI. Therefore, these properties should be favored in the selection of privileged structures for PPI target inhibition. Yet, it would be highly desirable to characterize other iPPI features that could not only be independent of the size of the compounds but also of their hydrophobicity. This might further facilitate the design (or selection) of new iPPI with a better “drug-likeness” profile. Indeed, it is commonly admitted that molecular size, hydrophobicity<sup>29,30</sup> and also aromaticity<sup>31</sup> can become a liability for drug development even though there are still some debate about the latter.<sup>32</sup> Conversely from a methodological standpoint, the PCA discussed above (Figure 1C and D) applied on our combined data set of bioactive conformations (iPPI + non-iPPI) shows that the largest discrepancies between iPPI and non-iPPI are associated with the size, the aromaticity and the hydrophobicity of the compounds. Although it interestingly confirms already known facts about iPPI characteristics, this also illustrates that other specific properties regarding iPPI compounds may be concealed behind this overall. In order to identify new intrinsic characteristics of iPPI, we decided to focus our analysis on 3D properties that are independent of the size and hydrophobicity of the compounds. Our protocol to select such 3D descriptors can be separated into four steps as shown in Figure 4 and is described in the methods section Descriptors’ Selection.

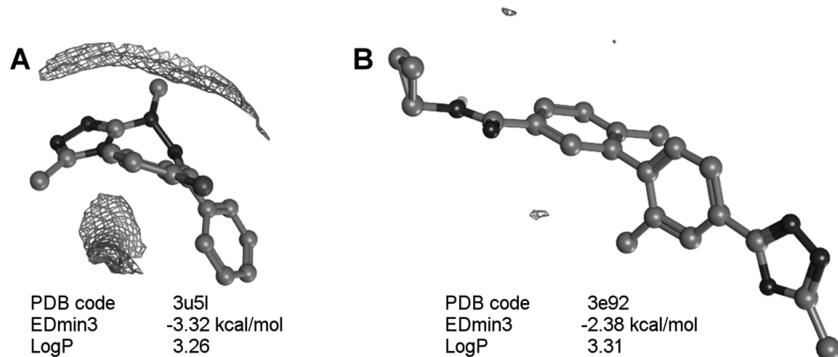
To prevent any chemical bias from an overrepresentation of some chemical series in the selection, the same protocol was



**Figure 4.** Protocol for descriptors’ selection. Step 1: gathering together correlated descriptors (Pearson’s correlation coefficient  $< -0.9$  or Pearson’s correlation coefficient  $> 0.9$ ) and selecting the most discriminative descriptor from each group according to the appropriate comparison test. Step 2: Selecting only highly discriminative descriptors ( $P\text{-value} < 0.01$ ). Step 3: Selecting only descriptors not correlated ( $-0.6 < \text{Pearson's correlation coefficient} < 0.6$ ) to size (molecular weight, number of non-H atoms), or the hydrophobicity ( $\log P$ , TPSA). Step 4: Selecting only descriptors obtained for the two data sets (full and diverse chemically).



**Figure 5.** Distribution of the four selected descriptors for iPPPI (darkgrey), non-iPPPI (gray). (A) Double histogram and boxplot of EDmin3. (B) Double histogram and boxplot of IW4. (C) Double histogram and boxplot of CW2. (D) Double histogram and boxplot of glob.



**Figure 6.** Illustration of EDmin3 on two molecules. EDmin3 is represented by calculating the hydrophobic interaction molecular fields (wire) and was calculated using Moe 2012.10 at the levels of energy equal to -2.4 kcal/mol using a dry probe. (A) Inhibitor of PPI (Bromodomain brd4) with a low value EDmin3. (B) Inhibitor of mitogen activated protein kinase 14 with a high value of EDmin3.

applied on the chemically diverse subset described above (66 iPPPI + 878 non-iPPPI) and our full data set (84 iPPPI + 1282 non-iPPPI). We selected the four discriminative descriptors that were identified in both cases as specific to iPPPI (Figure 5):

- Local interaction minimum (EDmin<sup>23</sup>)
- Hydrophilic INTeraction enerGY (integY) moment at energy level at -2.0 kcal/mol (IW4<sup>23</sup>)
- Capacity factor at the energy level equal to -0.5 kcal/mol (CW2<sup>23</sup>)
- Globularity (glob)

As specified in the selection procedure, neither of these descriptors is correlated to the molecular weight (MW), the octanol/water partition coefficient (logP), the topological polar surface area (TPSA), nor the number of heavy atoms (NbAtomNonH) (Table S3). Neither are they correlated to previously identified shape descriptors such as SHP2 or RDF070m (Figure S1). We suggest that these descriptors actually describe a set of so far undisclosed properties for iPPPI.

EDmin3, IW4, and CW2 are three VolSurf descriptors.<sup>23</sup> VolSurf derives its molecular descriptors from molecular

interaction fields using water or dry probes that aims at describing the general features of the molecule.

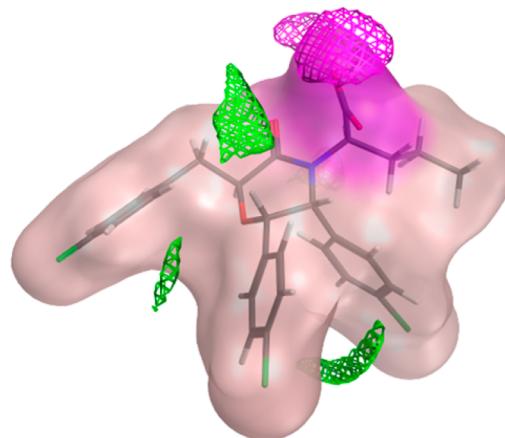
Local interaction minimum (EDmin3) is the third lowest local minimum of the interaction energy (in kilocalories per mole) of a dry probe. This property, derived from the ligand alone, measures therefore the potential interaction energy of the ligand with a hydrophobic object. The lower this energy is the more stable the interaction would be with a hydrophobic partner. When transposed to an interaction between a ligand and a PPI interface, it therefore represents the efficacy with which a compound can bind to a hydrophobic patch. EDmin3 is significantly lower for iPPI than non-iPPI ( $P\text{-value} = 1.24 \times 10^{-16}$ ). The mean of iPPI is equal to  $-2.85 \pm 0.21$  kcal/mol whereas the mean of non-iPPI is equal to  $-2.56 \pm 0.35$  kcal/mol (Figure 5A). Interestingly, this descriptor does not correlate with the hydrophobicity of the compound ( $\log P$ ,  $\text{cor}_{\log P-\text{EDmin3}} = -0.43$ ). A molecule with the same range of hydrophobicity ( $\log P$ ) can present a different EDmin3 as presented in Figure 6. This allows compounds with lower  $\log P$  to have appropriate values for EDmin3 and therefore make hydrophobic binding feasible with the PPI target. In order to make a hydrophobic interaction, a compound does need evidently some hydrophobic groups. But it seems that the spatial distribution (layout) of the hydrophobic functions within the compound is more determinant to promote ligand binding than simply the proportion of them.

The second descriptor shown as important for iPPI is the integy moment IW4. Integy moments express the unbalance between the center of mass of a molecule and the barycenter of its hydrophilic (IW) interacting regions (calculated using a water probe). The integy moment for hydrophilic regions is the vector pointing from the center of mass to the barycenter of the hydrophilic regions. A high integy moment is a clear concentration of hydrophilic interacting regions at one extremity of the compound. Small moments indicate that the hydrophilic interacting moieties are either close to the center of mass or that they are equally distributed around the molecule, so that their resulting barycenter is close to the center of the mass of the compound.<sup>33</sup> IW4 is measured at the energy level equal to  $-2.0$  kcal/mol. This level of energy represents more potential steric interaction between polar groups using dipole-dipole interactions rather than stronger interactions like hydrogen bonds or even charge-charge interactions. The average value of hydrophilic integy moment (IW4) is significantly higher for iPPI than non-iPPI ( $P\text{-value} = 5.82 \times 10^{-14}$ ). iPPI have a mean value equal to  $3.20 \pm 1.16$  while non-iPPI have an average equal to  $2.16 \pm 1.41$  (Figure 5B). So it means that on average an iPPI compound tend to have hydrophilic interacting regions more concentrated to one extremity of the molecule.

The third descriptor highlighted by our analysis is CW2. The capacity factor (CW2) represents the ratio between the surface of the hydrophilic regions calculated at  $-0.5$  kcal/mol and the total molecular surface. This descriptor is proportional to the concentration of hydrophilic regions (involved of weak potential polar interactions) compared to the total surface area. As seen from the distributions (Figure 5C), the value of CW2 of iPPI is significantly lower than inhibitors of classical targets ( $\text{mean}_{\text{iPPI}} = 1.94 \pm 0.17$ ;  $\text{mean}_{\text{non-iPPI}} = 2.17 \pm 0.27$  and  $P\text{-value} = 1.38 \times 10^{-13}$ ). This means that iPPI present less exposed hydrophilic regions than inhibitors of conventional targets. The fourth descriptor is the globularity (glob). It evaluates the resemblance of a compound with a sphere. A

globularity equal to 1 represents a perfect spherical molecule. Conversely, globularity equal to 0 represents a totally flat molecule. It is actually strongly correlated with the modified normalized moments of inertia we described above npr1\_rot ( $\text{cor}_{\text{npr1\_rot-glob}} = 0.91$ ). The average value of the globularity is significantly higher for iPPI than non-iPPI (Figure 5D). The mean of iPPI ( $\text{mean}_{\text{iPPI}} = 0.17 \pm 0.10$ ) is significantly different ( $P\text{-value} = 1.67 \times 10^{-7}$ ) than the mean of non-iPPI ( $\text{mean}_{\text{non-iPPI}} = 0.10 \pm 0.07$ ). This interestingly confirms the results from the previous (Continuity with Previous Works on iPPI) section that iPPI atoms have a better occupancy of the 3D space.

To summarize, the 3D characteristics highlighted in this analysis illustrate that an ideal iPPI could be described as more globular (higher glob), with a stronger capacity to bind hydrophobic patches at the core of PPI interfaces (lower EDmin3), with a smaller proportion of exposed hydrophilic regions (lower CW2) and conversely with a concentration of these hydrophilic regions to one extremity of the compound (higher IW4). This can be exemplified by compound 1MQ, which displays all these properties and is an inhibitor of the Mdm2/p53 interaction (Figure 7).



**Figure 7.** Bioactive conformation of compound 1MQ. The 1MQ compound as cocrystallized with Mdm2 (pdb code 4JVE). The compound is represented as transparent molecular surface and molecular sticks. Molecular surface is colored by lipophilicity (polar regions in pink). Values for the identified descriptors on this compound are EDmin3 =  $-3.18$  kcal/mol (illustrated by the green molecular field calculated using Moe 2012.10 at the levels of energy equal to  $-2.4$  kcal/mol using a dry probe), IW4 = 4.13 (illustrated by the pink molecular field calculated using Moe 2012.10 at the levels of energy equal to  $-5.5$  kcal/mol using a water probe), glob = 0.20 (illustrated by the molecular surface), and CW2 = 1.90 (illustrated by the proportion of pink surface over the full molecular surface).

**Taking into Account the PPI Target.** The results from previous paragraphs (Continuity with Previous Works on iPPI and Figure 3) have highlighted the variations that can be observed for some of the properties when considering the type of PPI targets. To see if the four identified 3D descriptors are discriminative for all PPI targets, we decided to consider each PPI target individually.

An analysis of variance (ANOVA; Kruskall-Wallis) experiment was used to evaluate the differences between them, and then a pairwise (Mann-Witney-Wilcoxon) test was performed (Figure S2).

		iPPI-DB	Bcl2	CD80	CTNNB1	E2	gp120	IL2	ITGAL	Max	Mdm2	Integrase	Xiap	Brd
EDmin3	eDrug	***	***	***	***	***	***	***	***	***	***	**	***	***
EDmin3	ion channel	***	***	***	***	*	**	***	***	***	***	**	***	
EDmin3	enzyme	***	***	***	***	*	*	***	***	*	***	*	***	
EDmin3	gpcr	***	***	***	***	*	**	***	***	*	***	*	***	
EDmin3	kinase	***	***	*	***			**	***		***			
EDmin3	nuclear	***	***	***	***	*	***	***	***	***	***	*	***	
IW4	eDrug	***	***						*		***	*		
IW4	ion channel	***	***						***		***	**		
IW4	enzyme		***						***		***	**		
IW4	gpcr		*								***	*		
IW4	kinase	***	***						***		***	**	**	
IW4	nuclear									***	***	*		
CW2	eDrug		*										***	
CW2	ion channel	***	***							*	***		***	***
CW2	enzyme	***	***							*	***		***	***
CW2	gpcr												**	
CW2	kinase	***	***				**			**	***	*	***	***
CW2	nuclear	***	***								***		***	***
glob	eDrug	***									***		***	***
glob	ion channel	***	***		***	***	***	***	***		***	*	***	***
glob	enzyme	***	***								***		***	***
glob	gpcr	***	**								***		***	***
glob	kinase	***	***		***	***	***	***	***		***	*	***	***
glob	nuclear	***									***		***	***

**Figure 8.** Statistical test results on iPPI-DB family versus inhibitors of other targets. Cells are flagged with different degrees of discrimination based on comparison test against non-iPPI. We consider the P-value, from not significant (no star) to modest  $0.01 < \text{P-value} < 0.05$  (\*), moderate  $0.001 < \text{P-value} < 0.01$  (\*\*), and very significant  $\text{P-value} < 0.001$  (\*\*\*)

The results obtained with the comparison test against non-iPPI (only for population with more than three compounds) show that for all PPI targets EDmin3 and CW2 are significantly lower. It seems that the property illustrated by EDmin3, i.e., the efficiency to bind a hydrophobic patch at PPI interfaces is essential. Even for PPI target such as Xiap and IL2 whose iPPI are known to be more polar,<sup>11</sup> this capacity seems to be critical to bind their target. Conversely, this seems to come with a lower proportion of exposed polar groups at the ligand's surface (low CW2) as well.

Yet, the two other descriptors (IW4 and glob) are significantly different for only four out the six PPI targets for which a quantitative comparison test was possible (more than three compounds). Brd and IL2 compounds are not characterized by the same iPPI profile for IW4 and Bcl2 and IL2 for globularity. Nevertheless, it seems that for all these three targets both CW2 values and more convincingly EDmin3 values are compensating for the other properties by being significantly different (CW2) or very significantly different (EDmin3). This confirms that EDmin3 is an interesting characteristics of iPPI regardless of the PPI target, and so is CW2 although to a lower extent in terms of significance.

**Application on 2D Databases.** In order to expand our analysis and validate the iPPI shape profile described herein, we needed to evaluate the aforementioned 3D characteristics on much larger data sets than the one used to identify them. Nevertheless, before attempting to establish such an evaluation, we first tried to find closely related properties described by 2D descriptors that could highly correlate with our 3D descriptors.

This would have the major benefit to prevent the generation of 3D conformations prior to the evaluation of those properties. However, none of the 4872 Dragon descriptors<sup>34</sup> we calculated, correlated with the identified descriptors confirming the distinctiveness of the above-described 3D characteristics. Such that it was necessary to find a protocol to properly generate, from 2D structures, the 3D de novo conformations from which our 3D descriptors could be calculated with a reasonable accuracy. So, after assessing different programs (Corina,<sup>35</sup> Concord,<sup>36</sup> Frog2,<sup>37</sup> and Moe 2012.10<sup>38</sup>) that generate 3D conformations and protocols to obtain the descriptors's values (Table S4), Moe 2012.10 was chosen to calculate them using the mean value of the 3D descriptors calculated on the 50 best (lowest internal energies) conformations for each compound (see Methods section Determination of 3D Conformations for 2D Databases; Table S5).

We thus evaluated the shape profile of three databases iPPI-DB,<sup>11</sup> eDrugs3D,<sup>39</sup> and a subset of bindingDB,<sup>40</sup> all from which we removed compounds with a molecular weight below 250 g/mol to be consistent with our data set of bioactive conformations (Table S6). eDrugs3D (June 2012) is a database of FDA (Food and Drug Administration) approved drugs from which we selected the 687 compounds that are orally bioavailable drugs. Herein, the bindingDB data set is a representative subset of the 100 most represented protein targets (classified by targets: GPCR (1025 compounds), nonkinase enzyme (965 compounds), kinase (971 compounds), ion channel (724 compounds), and nuclear receptors

(985 compounds)), which collectively represents for our bindingDB subset a total of 4670 compounds.

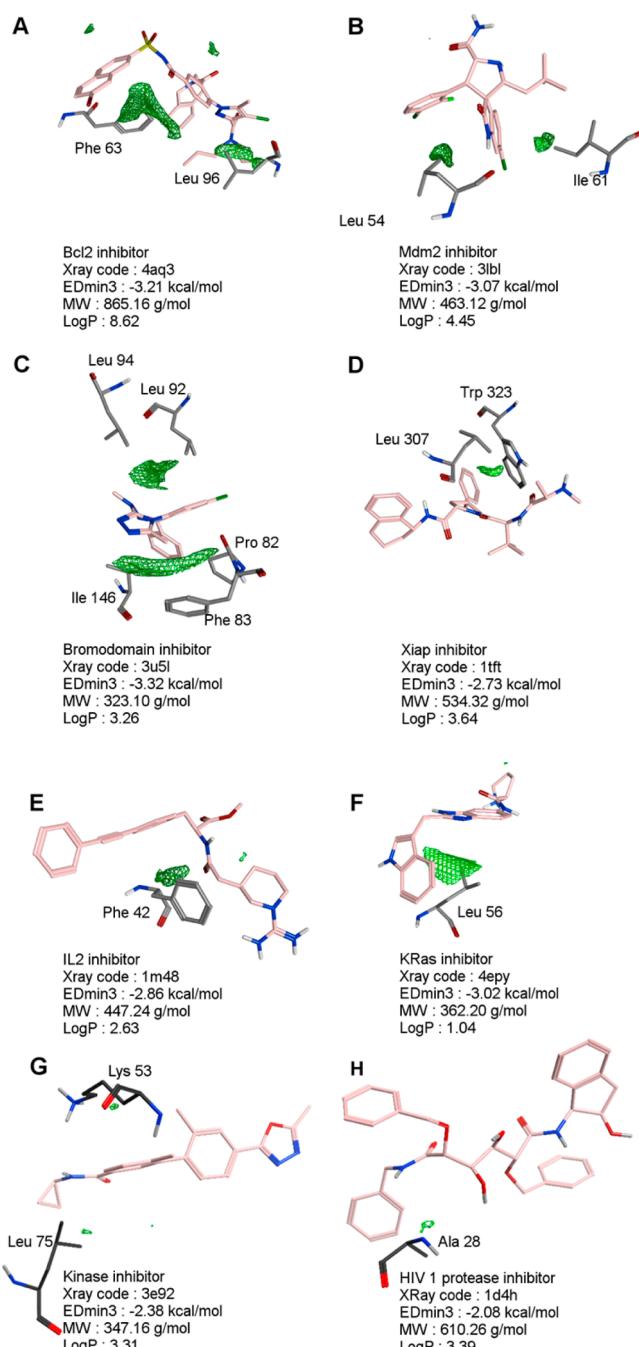
We first compared iPPI-DB as a whole data set with the other subsets of non-iPPI: eDrug and a subset of bindingDB (nonkinase enzymes, GPCR, kinases, ion channels, and nuclear receptors). For all four descriptors, the shape profile established from the data set of bioactive confirmations could be confirmed as a general trend for the global iPPi-DB population, especially for EDmin3 and globularity that remains significantly different regardless of the considered non-iPPI subset.

Similarly, to what was done on the bioactive conformation data set, we then assessed differences of shape profiles between PPI targets using the same approach combining ANOVA and pairwise comparison of populations (Figures S3 and 8).

From Figure 8, there is striking evidence that EDmin3 is the most essential property this analysis has revealed. This is also illustrated on Figure S3A where the differences of boxplots between PPI targets and non-PPI targets are even more pronounced than for the other descriptors. Indeed, a significant difference remains between all non-iPPI subsets and the iPPi-DB compounds of all PPI targets with the exception of Bromodomain compounds, which only display significant differences with respect to eDrug. But when one considers the most advanced compounds (either in preclinical or clinical phase) for the Bromodomain family such as I-BET762 and (+)-JQ1, values of EDmin3 are respectively equal to -3.31 and -3.33 kcal/mol which are lower than the iPPi-DB median values for this descriptor -2.84 kcal/mol (Table S7). Furthermore, every PPI target has a significantly lower value for EDmin3 with respect to the eDrug data set. This confirms the importance of this descriptor for all PPI families compared to existing orally bioavailable drugs.

The pairwise comparison across all targets for the three other descriptors illustrates a subtler pattern. Indeed, it is clear that, as opposed to EDmin3, those other properties do not characterize the iPPi of every single PPI target. It depends on the PPI target one considers. Interesting observations can be driven from these other properties as well and fit the profile of other PPI targets. Globularity is the property that can be found significantly different for most of the PPI targets at least with some non-iPPI subsets. The two other properties, IW4 and CW2 are clearly more specific to some PPI targets such as Bcl2, ITGAL (intercellular adhesion molecule (ICAM)-1/lymphocyte function-associated antigen (LFA)-1 interaction), Mdm2, Integrase, Xiap, or Bromodomain. Collectively, it is interesting to note in the very few cases where EDmin3 may not be significantly different with some non-iPPI subsets, that there are systematically one or several of these other newly identified property that compensate. For example, the Bromodomain inhibitors manage to display a significantly higher globularity than all non-iPPI subsets and a significantly lower CW2 than most of the non-iPPI subsets. Similar patterns can be seen for the rare other cases where EDmin3 is not significantly different.

Together, these results show the importance of EDmin3 as one prerequisite 3D characteristic to bind PPI interfaces and the importance of the other properties to complement this shape profile when necessary and to maximize the binding efficacy. Even in the case of modestly hydrophobic inhibitors like those of IL2 and Xiap, a tailored tridimensional layout of the few hydrophobic groups manages to promote the emergence of a hydrophobic interacting region (molecular fields) where it is needed to interact with hydrophobic residues within the protein partner (Figure 9). Even though, as all



**Figure 9.** Representation of EDmin3 on eight cocrystallized compounds and the corresponding interacting amino acids within the different binding pockets. EDmin3 is illustrated by calculating the hydrophobic interaction molecular fields (green wire) on six molecules at an energy level of -2.4 kcal/mol. The more extended is the molecular field the lower is the value of EDmin3. Protein residues localized close to the molecular fields are shown as well. The molecular fields derived from the ligands alone are well superimposed onto the hydrophobic residues of the protein partners. (A) Bcl2 inhibitor pdb code 4aq3. (B) Mdm2 inhibitor pdb code 3lbl. (C) Bromodomain inhibitor pdb code 3u5l. (D) Xiap inhibitor pdb code 1tft. (E) IL2 inhibitor pdb code 1m48. (F) KRas inhibitor pdb code 4epy. (G) Mitogen activated protein kinase 14 inhibitor pdb code 3e92. (H) HIV protease inhibitor pdb code 1d4h. Interestingly, in the case of Mdm2, Xiap, and IL2 not only the residues close to the hydrophobic interaction molecular fields are logically hydrophobic but there are known or predicted as PPI hot spots by DrugScore<sup>PPI</sup>.<sup>41</sup>

molecular descriptors in this analysis, EDmin3 is calculated from the ligand alone, the protein residues in the vicinity of the ligand hydrophobic interacting regions are always hydrophobic residues at the core of the interface. Furthermore, these residues are in some cases predicted (using DrugScore<sup>PPI</sup><sup>41</sup>) or known as hot spots, like Leu54 and Ile61 for Mdm2, Trp323 for Xiap, and Phe42 for IL-2 and are therefore essential for the PPI target inhibition. This illustrates how determinant EDmin3 is to profile compounds that have the best potential to hit the hydrophobic patch often present at the core of the PPI interfaces. Conversely, the lower values of EDmin3 for the non-iPPI compounds are clearly associated with different types of binding modes. Indeed, non-iPPI compounds do not rely as much on hydrophobic patch binding as illustrated by the size of the hydrophobic molecular fields (green wired) for the two examples in Figure 9: a kinase inhibitor (pdb code 3e92) and a HIV protease inhibitor (pdb code 1d4h).

**Advanced iPPI Drug Candidates.** We also looked at the profile of the few iPPI compounds within iPPI-DB that are presently in preclinical or clinical phases. In March 2014, iPPI-DB contained 29 of those compounds (Table S7). In order to evaluate the shape profile of those iPPI drug candidates we more specifically compared them to the drugs of the eDrug data set. For each of those 29 compounds we compared the values for molecular weight, logP, and the 4 shape descriptors to the thresholds for those descriptors that distinguish them from 80% of the eDrug compounds, to evaluate if those descriptors remain specific to iPPI. As a general trend, it can be noted that most of the iPPI drug candidates (20 out of 29) have a higher molecular weight than 80% of the conventional drugs. This confirms that also at this stage of development iPPI remain heavier than conventional drugs even though we removed the eDrug compounds with a molecular weight below 250 g/mol. But unexpectedly, not even half of the advanced iPPI (13 out of 29) have a higher logP value than 80% of the conventional drugs. Indeed, despite a higher hydrophobicity of iPPI in general (therefore when considering mostly hits), it seems that advanced iPPI are not characterized by this trend. It is now well documented that the higher hydrophobicity (high logP) of drug candidates can be a liability in some cases for further development and may explain this low proportion. While the binders of some targets are characterized by a higher logP (e.g. Bcl2, Mdm2), therefore matching the usually observed profile of iPPI, the property described by EDmin3 remains specific to iPPI regardless of the PPI target. As matter of fact, 24 out 29 of those iPPI drug candidates have a lower value for EDmin3 (i.e. better efficiency for hydrophobic binding) than 80% of the conventional drugs (threshold = -2.74 kcal/mol to distinguish them from 80% of drugs). And, this concerns all PPI targets. On the other hand, it is not as global for the other identified properties, IW4, glob, and CW2. But as specified in the previous paragraphs a clear pattern is still observable concerning the globularity of the inhibitors of some PPI targets such as Mdm2, PSIP1, Bromodomain, and Xiap for which this property remains important (16 out of 29). Conversely, CW2 and IW4 remains descriptive of advanced iPPI only for 5 iPPI out of 29 and not for the same PPI targets.

Interestingly, at latter stage of development only EDmin3 seem to be confirmed as characteristics to iPPI for all PPI targets as opposed to the three other identified properties. The fact that for all targets, except Bcl2, iPPI with a reasonable molecular weight (i.e. below 500 g/mol) can be associated with favorable values of EDmin3 provide good hopes to see the

emergence of privileged structures at latter stages of drug development balancing the size of the compounds with their efficiency to bind the hydrophobic core of PPI targets. More generally, for all the targets in iPPI-DB except for E2 and ZipA, one or several iPPI can be found with a molecular weight inferior to 500 g/mol and yet a low value for EDmin3. This confirms that such balance can be found for most PPI targets and should be therefore favored in the future.

If confirmed, this important finding might help the PPI community to identify at early stage of development compounds with the best drug potential, i.e., those balancing the critical need for efficient target binding with an acceptable hydrophobicity but also molecular size, similarly to the now commonly used concept of ligand efficiency and lipophilic efficiency.

**Toward the Profiling of PPI-Friendly Compounds.** The properties described herein, particularly EDmin3, could be used to better profile PPI-friendly molecules including among existing drugs. For example, the FDA approved drug Raloxifene is known to be a selective estrogen receptor modulator (SERM) commonly used as treatment for osteoporosis. Yet, a recent study has suggested that it could also be used as an inhibitor of the Interleukin 6 (IL-6)/GP130 interaction therefore as an advanced lead compound in targeting the IL-6/GP130/STAT3 cancer signaling pathway.<sup>42</sup> Interestingly, this compound has a value for EDmin3 equal to -2.76 kcal/mol, i.e., in the same range of the 29 iPPI-DB compounds in clinical trials and inferior to 80% of existing drugs (Figure S4). Therefore, it could be an interesting strategy to use such a property to profile existing drugs that have the most promising potential as iPPI in the context of a drug-repositioning program. But such a property could be even more widely be used for any small molecules and the purpose of building PPI focused libraries.

## METHODS

**Data Set Compilation.** A data set of 2952 compounds in their bioactive conformation was collected. A set of 2835 non-iPPI were selected from the PDDBbind database<sup>25</sup> ([www.pdbbind.org.cn](http://www.pdbbind.org.cn), version 2012). A set of 117 iPPI were selected from 2P2Idb ([www.2p2idb.cnrs-mrs.fr](http://www.2p2idb.cnrs-mrs.fr), version april 2014)<sup>12</sup> and Protein Database ([www.rcsb.org/pdb/home/home.do](http://www.rcsb.org/pdb/home/home.do), version 2012). Redundant compounds, amino acids, salts, compounds with less than 10 heavy atoms, carbocations, compounds with less than 250 g/mol, and compounds with phosphorus and nonorganic compounds were removed. All partial charges were calculated with the MMF94x force field using Moe software (version 2012.10).<sup>38</sup> All allosteric compounds for the iPPI data set were removed. A set of 1366 nonredundant compounds that is composed of 1282 non-iPPI and 84 iPPI was obtained (pdb codes: Table S1). A diversity criterion was applied on our data set using Functional-Class Fingerprints (FCFP4) with a dissimilarity threshold equal to 0.3 using PipelinePilot V8.5. We obtained 66 iPPI and 878 non-iPPI chemically diverse to use in combination with the full data set.

**Molecular Descriptors.** For the 3D descriptors selection, all internal 3D (i3D) molecular descriptors except for the potential energy descriptors and semi-empirical descriptors such as MNDO, PM3, and AM1 descriptors, were calculated using Moe software (version 2012.10).<sup>38</sup> The i3D descriptor data set use 3D coordinates' information about each molecule, but they are invariant to rotations and translations of the

conformation. Six 3D descriptors created in the laboratory by Dr. Petitjean were also calculated.<sup>43,44</sup> Two other 3D descriptors were added SHP2 and RDF070m and were calculated using the Dragon software<sup>34</sup> (version 6.0.32).

A set of molecular 2D descriptors was added to this analysis in order to compare them to the 3D descriptors. This set contains the logP, the molecular weight (MW), the topological polar surface area (TPSA), and the number of heavy atoms and was used to find descriptors having no correlation with the size and hydrophobicity. These descriptors were calculated with a java program using ChemAxon JChem library (version 5.10.1, <https://www.chemaxon.com/>). Finally, another set contains 4872 2D descriptors calculated using the Dragon software<sup>34</sup> (version 6.0.32). This set was calculated to compare them with our four selected descriptors and find a way to replace our four 3D descriptors.

**Molecular Descriptors Analysis.** Univariate analysis and multivariate analysis were performed in order to investigate the different physicochemical properties of the different compounds. All the analysis was done using script from the software R<sup>45</sup> (version 2.15.1).

**Analysis of the Discrimination between Two Populations.** Comparisons between two populations were done using Student's test if the two populations follow a normal distribution according to Shapiro test and have different variances according to the Fisher–Student test or using the nonparametric Mann–Witney–Wilcoxon test when the previous conditions were not fulfilled.

Mean group values were compared using nonparametric Kruskal–Wallis ANOVA. Post hoc comparisons were carried out using a pairwise Mann–Whitney–Wilcoxon test.<sup>46</sup>

**Descriptors' Selection.** A protocol described (Figure 4) herein was used to select the most discriminating descriptors regardless of their size and hydrophobicity. This protocol is divided into four steps. In the first step, descriptors were gathered according to the Pearson correlation coefficient above 0.9 (or below −0.9) and selected the most discriminating descriptors according to the P-value from comparison test between iPPPI and non-iPPPI. For the comparison test, if the two populations follow a normal distribution according to Shapiro test and have different variance according to the Fisher–Student test, we apply a Student's test between these two populations. If not, we apply a Mann–Witney–Wilcoxon test<sup>46</sup> between the two populations.

In the second step, only descriptors with a P-value under 0.01 for the comparison test between iPPPI and non-iPPPI were selected. In the third step, only descriptors with a Pearson's correlation coefficient with molecular weight, LogP, TPSA, and number of heavy atoms between −0.6 and 0.6 were selected.

In the last step, this protocol was calculated on the full bioactive data set and also the chemically diverse bioactive data set and only descriptors obtained for the two data sets (full and diverse data sets) were selected.

**Determination of 3D Conformations for 2D Databases.** A set of 50 conformations for each ligand were generated using Moe 2012.10 software. A stochastic search with rejection equal to 100, iteration equal to 5000, rms gradient equal to 0.005, and an MM iteration limit equal to 200 was done. We also fixed the RMSD limit at 0.75, the energy windows at 7 kcal/mol, and the conformation limit at 50. All partial charges were calculated with the MMF94x force field using Moe 2012.10 software.<sup>38</sup>

For each conformation, all the 3D descriptors were calculated and the mean of these values for each compound was used as the value of the descriptor for a given compound.

## CONCLUSIONS

Identifying low-molecular-weight iPPPI is known to be a difficult task. This has usually been translated into designing compounds with higher size, aromaticity, and hydrophobicity. Yet, lessons are being learnt from iPPPI bioactive conformations in an attempt to circumvent this trend and to bring new insights into their intrinsic capacity to bind PPI interfaces. During this analysis, we demonstrated that this capacity partially rely on the combination of several structural and electrostatic features including the globularity and the distribution of hydrophilic regions but most importantly of hydrophobic interacting regions. More distinctively, iPPPI seem to be characterized by a significantly higher efficiency to bind the hydrophobic patches often present at PPI interfaces. The perspective that this chemical property may facilitate the binding to such intricate surfaces, often described as hydrophobic and devoided from tractable binding pocket, could expand our pharmacological options to tackle poorly ligandable PPI targets. Furthermore, the absence of correlation of this type of property with the hydrophobicity and the size of the compounds could open new ways to design iPPPI with improved ligand and lipophilic efficiencies and may allow the scientific community to anticipate an era of more drug-like iPPPI.

## ASSOCIATED CONTENT

### Supporting Information

Figure S1: a pairwise correlation plot for the four highlighted descriptors (EDmin3, IW4, CW2, and glob) and already known descriptors characteristics to iPPPI (molecular weight, LogP, number of heavy atoms, topological surface area, SHP2, and RDF070m). Figure S2: boxplots of EDmin3, IW4, CW2, and glob for the different PPI families of our X-ray data set. Figure S3 shows boxplots of EDmin3, IW4, CW2, and glob for the different PPI families of iPPPI-DB, and also compounds from eDrug and bindingDB. Figure S4: structure and characteristics of raloxifene. Table S1: all the pdb codes of our iPPPI X-ray data sets. Table S2: list of all 2D descriptors used in our PCA. Table S3: Pearson correlation coefficient between the four highlighted descriptors (EDmin3, IW4, CW2, and glob) and 2D descriptors known to be characteristics to iPPPI (molecular weight, LogP, number of heavy atoms, and topological surface area). Table S4: results of the Pearson coefficient correlation for EDmin3, IW4, CW2, and glob between the X-ray conformation and the de novo conformation for the various assessed software. Table S5: mean and P-value for EDmin3, IW4, CW2, and glob between the de novo conformations generated by Moe and the X-ray data set. Table S6: number of compounds for the 2D databases used in our analysis (iPPPI-DB, eDrug, and bindingDB). Table S7: values of EDmin3, IW4, CW2, and glob for iPPPI-DB compounds in clinical trials. All X-ray data set (iPPPI and non-iPPPI) used in this study are provided in csv files (pdb code and smiles). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: olivier.sperandio@inserm.fr.

**Present Address**

<sup>○</sup>Inserm UMR-S973, Université Paris Diderot, 35 Rue Hélène Brion, 75205 Paris cedex 13, case courier 7113.

**Funding**

This work was funded by the French agency ANR within the project Piribio\_PPIFocusDb (ANR-09-PIRI-0017) Edition 2009 du Programme Physique et Chimie du Vivant.

**Notes**

The authors declare no competing financial interest.

**ABBREVIATIONS**

ANOVA, analysis of variance; Bcl2, apoptosis regulator (Bcl2)/apoptosis regulator (Bax); Brd, bromodomain-containing protein (Brd)2/3/4/T/(Histone); CD80, T-lymphocyte activation antigen CD80/T-cell-specific surface glycoprotein CD28; CTNNB1, beta catenin/TCF-4; CW2, capacity factor at the energy level equal to -0.5 kcal/mol; Edmin3, local interaction minimum; E2, human papillomavirus (HPV) E1/E2; FCFP4, Functional-Class Fingerprints 4; FDA, Food and Drug Administration; Glob, globularity; GPCR, G-Protein Coupled Receptors; Gp120, envelope glycoprotein GP120/CD4 receptor; HTS, high throughput screening; IL2, interleukin 2 (IL2)/interleukin 2 alpha receptor; integy, interaction energy; integrase, Lens epithelium-derived growth factor (LEDGF)/human (integrase); ITGAL, intercellular adhesion molecule (ICAM)-1/lymphocyte function-associated antigen (LFA)-1; iPPi, inhibitors of protein–protein interactions; IW4, hydrophilic interaction energy (integy) moment at energy level at -2.0 kcal/mol; KRas, growth factor receptors activate (KRas)/factor son of sevenless (SOS1); logP, octanol/water coefficient; Mdm2, minute double mutant 2 (Mdm2)/tumor suppressor (p53); Menin, menin/mixed lineage leukemia (MLL); MW, molecular weight; NbAtomNonH, the number of heavy atoms; non-iPPi, non inhibitors of protein–protein interactions; npr1/npr2, normalized moments of inertia 1/2; npr1/2, rotate normalized moments of inertia1/2; PPI, protein–protein interactions; RDF070m, radial distribution function 070m; ROS, Lipinski's rules of five; SERMs, selective estrogen receptor modulators; TNFR1A, tumor necrosis factor (TNFR1A) receptor 1A/human tumor necrosis factor beta (TNFB); TPSA, topological polar surface area; Xiap, X-linked inhibitor of apoptosis protein (XIAP)/second mitochondria-derived activator of caspases (smac) interaction; ZipA, ZipA cell division protein; ftsZ, cell division protein

**REFERENCES**

- (1) Chene, P. Drugs targeting protein-protein interactions. *ChemMedChem*. 2006, 1, 400–411.
- (2) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. rev. Drug discovery* 2002, 1, 727–730.
- (3) Teague, S. J. Learning lessons from drugs that have recently entered the market. *Drug discovery today* 2011, 16, 398–411.
- (4) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discovery* 2006, 5, 993–996.
- (5) Zhang, Q. C.; Petrey, D.; Garzon, J. I.; Deng, L.; Honig, B. PrePPI: a structure-informed database of protein-protein interactions. *Nucleic acids Research* 2013, 41, D828–833.
- (6) Lensink, M. F.; Moal, I. H.; Bates, P. A.; Kastritis, P. L.; Melquiond, A. S.; Karaca, E.; Schmitz, C.; van Dijk, M.; Bonvin, A. M.; Eisenstein, M.; Jimenez-Garcia, B.; Grosdidier, S.; Solernou, A.; Perez-Cano, L.; Pallara, C.; Fernandez-Recio, J.; Xu, J.; Muthu, P.; Praneeth Kilambi, K.; Gray, J. J.; Grudinin, S.; Derevyanko, G.; Mitchell, J. C.; Witing, J.; Kanamori, E.; Tsuchiya, Y.; Murakami, Y.; Sarmiento, J.; Standley, D. M.; Shirota, M.; Kinoshita, K.; Nakamura, H.; Chavent,

M.; Ritchie, D. W.; Park, H.; Ko, J.; Lee, H.; Seok, C.; Shen, Y.; Kozakov, D.; Vajda, S.; Kundrotas, P. J.; Vakser, I. A.; Pierce, B. G.; Hwang, H.; Vreven, T.; Weng, Z.; Buch, I.; Farkash, E.; Wolfson, H. J.; Zacharias, M.; Qin, S.; Zhou, H. X.; Huang, S. Y.; Zou, X.; Wojdyla, J. A.; Kleanthous, C.; Wodak, S. J. Blind prediction of interfacial water positions in CAPRI. *Proteins* 2014, 82, 620–632.

(7) Perot, S.; Sperandio, O.; Miteva, M. A.; Camproux, A. C.; Villoutreix, B. O. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug discovery today* 2010, 15, 656–667.

(8) Fuller, J. C.; Burgoyne, N. J.; Jackson, R. M. Predicting druggable binding sites at the protein-protein interface. *Drug discovery today* 2009, 14, 155–161.

(9) Fry, D. C. Drug-like inhibitors of protein-protein interactions: a structural examination of effective protein mimicry. *Curr. Protein Pept Sci.* 2008, 9, 240–247.

(10) Huang, Y.; Wolf, S.; Beck, B.; Kohler, L. M.; Khouri, K.; Popowicz, G. M.; Goda, S. K.; Subklewe, M.; Twarda, A.; Holak, T. A.; Domling, A. Discovery of highly potent p53-MDM2 antagonists and structural basis for anti-acute myeloid leukemia activities. *ACS Chem. Biol.* 2014, 9, 802–811.

(11) Labb  , C. M.; Laconde, G.; Kuenemann, M. A.; Villoutreix, B. O.; Sperandio, O. iPPi-DB: a manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions. *Drug discovery today* 2013, 18, 958–968.

(12) Basse, M. J.; Betzi, S.; Bourgeas, R.; Bouzidi, S.; Chetrit, B.; Hamon, V.; Morelli, X.; Roche, P. 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic acids research* 2013, 41, D824–827.

(13) Higuero, A. P.; Jubb, H.; Blundell, T. L. TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions. *Database: the journal of biological databases and curation* 2013, 2013, bat039.

(14) Morelli, X.; Bourgeas, R.; Roche, P. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr. Opin. Chem. Biol.* 2011, 15, 475–481.

(15) Sperandio, O.; Reynes, C. H.; Camproux, A. C.; Villoutreix, B. O. Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug discovery today* 2010, 15, 220–229.

(16) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews* 2001, 46, 3–26.

(17) Hann, M. M. Molecular obesity, potency and other addictions in drug discovery. *Med. Chem. Commun.* 2011, 2, 349–355.

(18) Akritopoulou-Zanze, I.; Metz, J. T.; Djuric, S. W. Topography-biased compound library design: the shape of things to come? *Drug discovery today* 2007, 12, 948–952.

(19) Kortagere, S.; Krasowski, M. D.; Ekins, S. The importance of discerning shape in molecular pharmacology. *Trends in pharmacological sciences* 2009, 30, 138–147.

(20) Neugebauer, A.; Hartmann, R. W.; Klein, C. D. Prediction of protein-protein interaction inhibitors by chemoinformatics and machine learning methods. *Journal of medicinal chemistry* 2007, 50, 4665–4668.

(21) Reynes, C.; Host, H.; Camproux, A. C.; Laconde, G.; Leroux, F.; Mazars, A.; Deprez, B.; Fahraeus, R.; Villoutreix, B. O.; Sperandio, O. Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods. *PLoS computational biology* 2010, 6, e1000695.

(22) Villoutreix, B. O.; Labb  , C. M.; Lagorce, D.; Laconde, G.; Sperandio, O. A leap into the chemical space of protein-protein interaction inhibitors. *Current pharmaceutical design* 2012, 18, 4648–4667.

(23) Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *European journal of pharmaceutical sciences: official journal of the European Federation for Pharmaceutical Sciences* 2000, 11 (Suppl 2), S29–39.

- (24) Hopkins, A. L.; Keseru, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nature reviews. Drug discovery* **2014**, *13*, 105–121.
- (25) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *Journal of medicinal chemistry* **2004**, *47*, 2977–2980.
- (26) Zhao, Y. D.; Rahardja, D.; Qu, Y. Sample size calculation for the Wilcoxon-Mann-Whitney test adjusting for ties. *Statistics in medicine* **2008**, *27*, 462–468.
- (27) Randic, M. Molecular shape profiles. *Journal of Chemical Information Computer Sciences* **1995**, *35*, 373–382.
- (28) Wirth, M. S.; W, H. B. Bioactive Molecules: Perfectly Shaped for Their Target? *Molecular Informatics* **2011**, *30*, 677–688.
- (29) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature reviews. Drug discovery* **2007**, *6*, 881–890.
- (30) Gleeson, M. P.; Hersey, A.; Montanari, D.; Overington, J. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nature reviews. Drug discovery* **2011**, *10*, 197–208.
- (31) Ritchie, T. J.; Macdonald, S. J. The impact of aromatic ring count on compound developability—are too many aromatic rings a liability in drug design? *Drug discovery today* **2009**, *14*, 1011–1020.
- (32) Ward, S. E.; Beswick, P. What does the aromatic ring number mean for drug design? *Expert opinion on drug discovery* **2014**, *1*–9.
- (33) Zamora, I.; Oprea, T.; Cruciani, G.; Pastor, M.; Ungell, A. L. Surface descriptors for protein-ligand affinity prediction. *Journal of medicinal chemistry* **2003**, *46*, 25–33.
- (34) Todeschini, R.; Consonni, V. *Molecular descriptors for chemoinformatics*. John Wiley & Sons: Weinheim, Germany, 2009; p 1257.
- (35) Sadowski, J.; Schwab, C. H.; Gasteiger, J. *Corina*, software for the generation of high-quality three-dimensional molecular models; Molecular Networks GmbH Computerchemie: Erlangen, Germany, 2004; [www.molecular-networks.com](http://www.molecular-networks.com).
- (36) Rusinko, A.; Sheridan, R. P.; Nilakantan, R.; Haraki, K.; Bauman, N.; Venkatraghavan, R. Using concord to construct a large Database of three-dimensional coordinates from connection tables. *Journal of chemical information and computer sciences* **1989**, *29*, 251–255.
- (37) Miteva, M. A.; Guyon, F.; Tuffery, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic acids research* **2010**, *38*, W622–627.
- (38) MOE, M.O.E., Chemical Group. Inc.: Montreal, 2012.
- (39) Pihan, E.; Colliandre, L.; Guichou, J. F.; Douguet, D. e-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics* **2012**, *28*, 1540–1541.
- (40) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic acids research* **2007**, *35*, D198–201.
- (41) Kruger, D. M.; Gohlke, H. DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. *Nucleic acids research* **2010**, *38*, W480–486.
- (42) Li, H.; Xiao, H.; Lin, L.; Jou, D.; Kumari, V.; Lin, J.; Li, C. Drug Design Targeting Protein-Protein Interactions (PPIs) Using Multiple Ligand Simultaneous Docking (MLSD) and Drug Repositioning: Discovery of Raloxifene and Bazedoxifene as Novel Inhibitors of IL-6/GP130 Interface. *Journal of medicinal chemistry* **2014**, *57*, 632–641.
- (43) Petitjean, M. On the Analytical Calculation of van der Waals Surfaces and Volumes: Some Numerical Aspects. *J. Comput. Chem.* **1994**, *15*, 507–523.
- (44) Petitjean, M. Applications of the radius-diameter diagram to the classification of topological and geometrical shapes of chemical compounds. *Journal of chemical information and computer sciences* **1992**, *32*, 331–337.
- (45) R Development Core team. *R: A language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2012.
- (46) Bauer, D. F. Constructing Confidence Sets Using Rank Statistics. *Journal of the American Statistical Association* **1972**, *67*, 687–690.