

iPPI-DB: an online database of modulators of protein–protein interactions

Céline M. Labbé^{1,2}, Méline A. Kuenemann^{1,2}, Barbara Zarzycka³, Gert Vriend⁴, Gerry A.F. Nicolaes³, David Lagorce^{1,2}, Maria A. Miteva^{1,2}, Bruno O. Villoutreix^{1,2} and Olivier Sperandio^{1,2,*}

¹Université Paris Diderot, Sorbonne Paris Cité, Molécules Thérapeutiques, In Silico, INSERM UMR-S 973, Paris, France, ²INSERM, U973, Paris, France, ³Department of Biochemistry, Cardiovascular Research Institute Maastricht (CARIM), Maastricht University, Maastricht, The Netherlands and ⁴Centre for Molecular and Biomolecular Informatics (CMBI), Radboudumc, PO Box 9101, 6500 HB Nijmegen, The Netherlands

Received August 02, 2015; Revised September 18, 2015; Accepted September 19, 2015

ABSTRACT

In order to boost the identification of low-molecular-weight drugs on protein–protein interactions (PPI), it is essential to properly collect and annotate experimental data about successful examples. This provides the scientific community with the necessary information to derive trends about privileged physicochemical properties and chemotypes that maximize the likelihood of promoting a given chemical probe to the most advanced stages of development. To this end we have developed iPPI-DB (freely accessible at <http://www.ippidb.cdithem.fr>), a database that contains the structure, some physicochemical characteristics, the pharmacological data and the profile of the PPI targets of several hundreds modulators of protein–protein interactions. iPPI-DB is accessible through a web application and can be queried according to two general approaches: using physicochemical/pharmacological criteria; or by chemical similarity to a user-defined structure input. In both cases the results are displayed as a sortable and exportable datasheet with links to external databases such as Uniprot, PubMed. Furthermore each compound in the table has a link to an individual ID card that contains its physicochemical and pharmacological profile derived from iPPI-DB data. This includes information about its binding data, ligand and lipophilic efficiencies, location in the PPI chemical space, and importantly similarity with known drugs, and links to external databases like PubChem, and ChEMBL.

INTRODUCTION

Drug discovery is a remarkably complicated process and among the many hurdles that drug hunters have to face is the paucity of targets. The focus of these past 50 years has thus been centered on certain large enzyme families, ion channels and/or receptors because they were deemed more amenable to modulation by low molecular weight (LMW) compounds (1–3). These observations stand in sharp contrast to the large number of mainly untapped protein–protein interactions (PPI). PPIs play an essential role in nearly all biological processes and their deregulation is often associated with disease states. For this reason, there is a growing interest to target them for therapeutic interventions using LMW compounds (<1000 g/mol). Still, targeting PPIs with LMW drugs remains one of the most difficult challenges in molecular medicine. As opposed to most traditional targets, PPIs have not evolved to bind small molecules (4). Indeed, the molecular topography of most known PPIs, often described as shallow, large and hydrophobic, makes them harder to tackle with small compounds and these features have often been translated in the design of larger and more hydrophobic modulators. In fact, such interfaces are now known to preferentially bind compounds that display some specific physicochemical characteristics and chemotypes (5–7). Yet, analyzing further successful LMW PPI modulators should be essential to rationalize what makes those molecules so special and capable to bind to such intricate surfaces and thus assist the design of future generations of PPI inhibitors. Two databases already propose to access the structural and pharmacological data of existing successful examples of PPI modulators. First, the TIMBAL (8) database proposes compounds that are automatically imported from ChEMBL (9) (<https://www.ebi.ac.uk/chembl/>) following a manual selection of the PPI target type, it contains the data of about 8900 compounds on several PPI targets. Most of the data come from

*To whom correspondence should be addressed. Tel: +33 1 57 27 83 90; Fax: +33 1 57 27 83 72; Email: olivier.sperandio@inserm.fr

a large pool of integrins for which the target is not always clearly identified. Second, the 2P2I-db (10) is a manually curated database from the PDB (11) (Protein Data Bank) that collects the crystallographic data of cocrystallized orthosteric PPI inhibitors. In the last version, it contains 242 compounds.

In order to help the scientific community to gain new knowledge about LMW modulators of this new target class, we propose a database, named iPPI-DB, together with a user-friendly web interface (<http://www.ippidb.cdithem.fr>). The database is actually the second release of a previous version of the database that was before password protected and for which only a representative fraction of the compounds were accessible. We have decided to now make the data fully available to everyone while also adding new functionalities (described below) (12), such as an embedded chemical similarity search, a query toward drug candidates and the possibility to export all results as a CSV file.

RESULTS

Presentation of iPPI-DB

iPPI-DB is a relational database containing the structure, the physicochemical characteristics, the pharmacological data (biochemical and/or cellular binding data) of compounds modulating known PPI targets as well as the profile of the corresponding targets. As those data are manually extracted from the literature and curated by experts, we set up a comprehensive protocol to decide whether or not a given compound should enter the database and thus such as to ensure assure the data quality. First, we consider only world patents and peer-reviewed articles from scientific journals with expertise in medicinal chemistry. Also considered, the PPI targets must have been discussed in several scientific publications with for instance links between biological studies previously published, therefore with some demonstrating the pertinence of the interactions and their contribution in a given disease state and more specifically in terms of functional mechanism. Moreover, only small non-peptide compounds are selected such that metal-based compounds, macrocycles and molecules containing atoms other than C, N, O, S, P and halogens are not currently included. Furthermore, to be confident about the actual compound activities, we have chosen to rule out assays containing only percentage of inhibition and select compounds and assays for which a dose-response study was carried out and led to any of the following measures of activity: K_d , K_i , IC_{50} or EC_{50} . Regardless of the assay type, we also impose a 30 μM threshold on that activity to prevent as much as possible adding non specific binders in the collection. A series of nine descriptors are calculated for each compound using the Chemaxon JChem library v6.1 (www.chemaxon.com). Those descriptors are commonly used to characterize the physicochemical profile of small molecules, namely molecular weight, A_{logP} (13), number of Hydrogen bond donors and acceptors, the topological polar surface area (14), the number of rotatable bonds, the number of aromatic rings according to Ritchie (15), the proportion of sp^3 carbon— F_{sp^3} , and the number of chiral centers— R/S . At this time, a total of 2461 binding data have been collected on 1650 compounds across 13 families of highly homologous

PPI targets. This represents a significantly higher number of accessible data than in the first version, which only provided access to a representative subset of 352 compounds. As described below, the interface now allows the user to perform a chemical similarity search on iPPI-DB, to restrict the search only to drug candidates, and to export all results, which was not available in the first version.

iPPI-DB webapps


In order for the PPI scientific community to fully make use of the database, we have designed a new user-friendly web application that offers tools and predictive models as well as two approaches to query iPPI-DB. The first possibility to search the database is to use physicochemical and pharmacological criteria on a given target. The second possibility is to import an input molecule as the query compound and search by chemical similarity the entire collection regardless of their target type.

Querying iPPI-DB using physicochemical or pharmacological criteria

In the first approach physicochemical and pharmacological criteria can be defined to search active compounds within iPPI-DB on a given PPI target. The user can refine the search by tuning several properties such as potency thresholds, physicochemical descriptors (e.g. molecular weight, hydrophobicity, F_{sp^3}). As an alternative, the user can apply one of the three commonly used physicochemical rules namely Lipinski's rule of five (16), Veber's (17) and Pfizer's 3/75 (18) in order to select only compounds that are predicted as orally bioavailable in humans, rats or as potentially less toxic, respectively. Finally, the user can choose to extract only drug candidates on the PPI targets that have some compounds presently tested in preclinical or clinical phases (annotated using the data MDDR march 2012).

Querying iPPI-DB using chemical similarity to a user-defined compound

The second approach to query iPPI-DB is to use a compound structure as an input and perform a chemical similarity search between this compound and all the modulators present in iPPI-DB. The user can either copy and paste a SMILES string or sketch directly a molecule within the Marvin JS editor (www.chemaxon.com) embedded in the interface. Once the input molecule has been properly imported, one of the two fingerprints ECFP4 or FCFP4 can be chosen. The difference with respect to the first approach to query iPPI-DB is that in this case the search is made on all compounds of iPPI-DB regardless of the corresponding PPI target. This can help the user to evaluate whether its input structure can contain a scaffold that match more than one PPI target and may therefore constitute either a possible privileged substructure or a non-specific chemical moiety. The chemical search provides the user with all the binding data of the 20 closest iPPI-DB compounds to the input structure according to the chosen type of fingerprint and their Tanimoto index. It also recalls the structure of the query compound along with its compliance with respect to the three chemistry rules mentioned above (Figure 1).

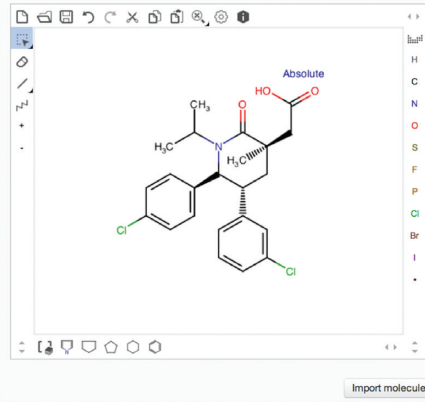
Search by chemical similarity (powered by ChemAxon v6.1) 

1- Import your molecule

Paste the SMILES of your molecule here:

OR

Sketch or upload (in .mol format and in 2D only) it here:



2- Choose your search criteria

Which fingerprints do you want to use?




- Extended-Connectivity Fingerprints 4 (ECFP4)
- Functional-Class Fingerprints 4 (FCFP4)

Search

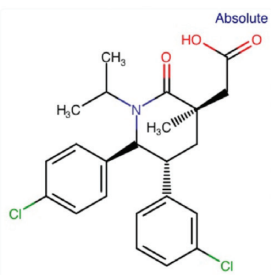
The process is retrieving the 20 closest iPPI-DB compounds. This may take some time, please wait...

Chemical similarity

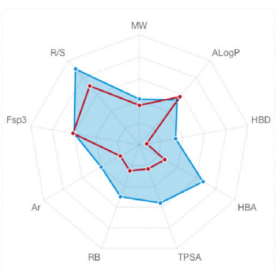
- Paste a SMILES or sketch a molecule
- Import the molecule
- Select a fingerprint (ECFP4 or FCFP4)
- Run the search
- Get the 20 most similar iPPI-DB compounds

Your input molecule   

Molecule's structure






Radar chart



Physicochemical filters

	Lipinski's RO5	Veber	3 / 75
MW	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
HBA	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HBD	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HBA + HBD	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
AlogP	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
TPSA	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
RB	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Similarity search using ECFP4: 47 binding affinities from 20 compound(s).   

Showing 1 to 47 of 47 entries










Tn	ID	Compound	RadarChart	Target	Assay	Type	Activity	MW	ALogP	HBD	HBA	TPSA	RB	Ar	Fsp3	R/S	LE	LLE	Biblio
0.34	682			MDM2 Q00987	FP	pKi	6.52	459.15	4.45	2	5	61.44	4	2	0.42	4	0.29	2.07	 [57]
0.34	682			MDM2 Q00987	proliferation assay *	pC50	5.88	459.15	4.45	2	5	61.44	4	2	0.42	4	0.25	1.23	 [57]
0.34	682			MDM2 Q00987	proliferation assay *	pC50	4.77	459.15	4.45	2	5	61.44	4	2	0.42	4	0.21	0.32	 [57]

Figure 1. Example of a query on iPPI-DB using the mode 'Chemical Similarity'. As the query compound, the user can either copy and paste a SMILES string into the interface, or directly sketch the compound within a Marvin JS editor. Here, we show an example with the structure with a recently identified piperidinone-based inhibitor of the MDM2/p53 interaction (25). Then, the compound can properly be imported, and the desired type of fingerprints for the chemical similarity search has to be chosen between ECFP4 and FCFP4. The search provides the user with all the binding data (here 47) of the 20 most similar iPPI-DB compounds in a table that can be sorted according to various criteria: molecular descriptors, activity, efficiencies or bibliographic IDs. The interface also provides a summary of the properties of the input compound using a radar chart based on nine physicochemical properties (molecular weight—*MW*, hydrophobicity—*AlogP*, the number of H-bond donors—*HBD* and acceptors—*HBA*, Topological Polar Surface Area—*TPSA*, the number of rotatable bonds—*RB*, the number of aromatic rings—*Ar*, the proportion of sp³ carbon atoms—*Fsp3* and the number of chiral centers—*R/S*) and its compliance to the three chemistry rules, Lipinski's RO5, Veber's and Pfizer's 3/75.

Compound 682

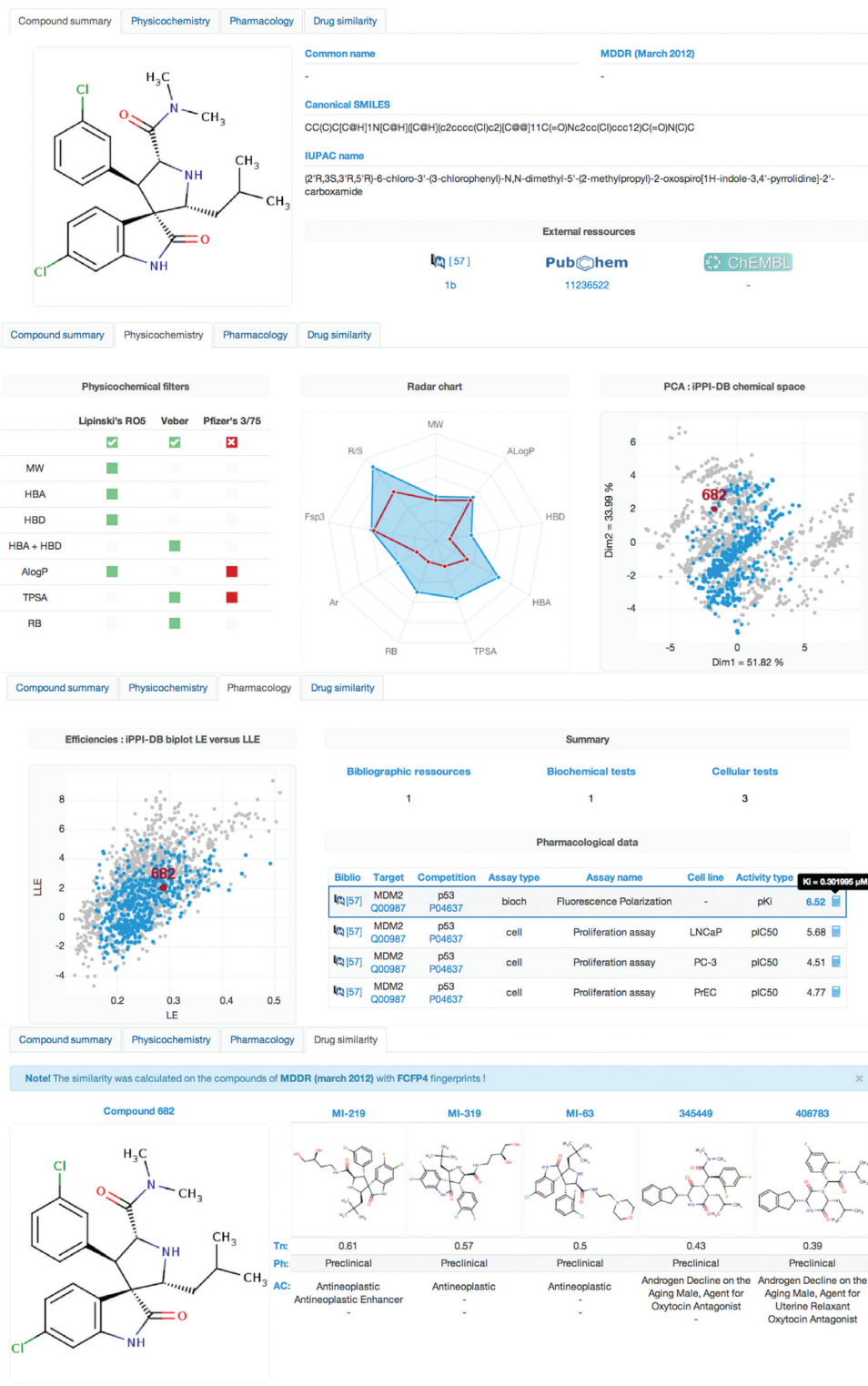


Figure 2. iPPI-DB ID card of compound 682. All data available of a given compound (here compound 682 as the first hit on the chemical similarity search from Figure 1) are provided through four different tabs: *compound summary*, *physicochemistry*, *pharmacology* and *drug similarity*.

Visualizing the results within a sortable and exportable datasheet

The results of any query are displayed as a sortable and exportable datasheet that contains all the binding data found in iPPI-DB concerning the compounds fitting the query criteria. Each line matches a compound's binding data and recalls the iPPI-DB compound ID. Also displayed are a 2D representation of the compound, a radar chart summarizing the values of nine computed physicochemical descriptors and the activity, the PPI target bound by the compound (with a clickable link to the Uniprot (19) web server (<http://www.uniprot.org>)), the name of the assay (with a mouse over to have its full name) and whether it is a cellular assay, the type of activity that was derived from the assay (pIC_{50} , pK_i , etc.), the potency of the compound to bind its target, the nine physicochemical descriptors, the ligand and the lipophilic efficiencies (20), and the iPPI-DB ID of the bibliographic source from which the data were retrieved. The icon indicates if it is a world patent or a research article. The link is clickable and redirects the user to either the Wipo (www.wipo.int) or the Pubmed (21) web page. All results can be exported as a CSV file that contains all the above-mentioned data along with a Chemaxon canonical SMILES of the compounds' structures.

Accessing all data of a given iPPI-DB compound through its ID card

If the user is interested in a particular modulator, a click on its iPPI-DB ID within the results' datasheet will open an individual compound ID card where all data about the compound are summarized through four different tabs: *compound summary*, *physicochemistry*, *pharmacology* and *drug similarity* (Figure 2).

In the *compound summary* tab, the chemical structure of the compound is recalled along with its canonical SMILES, its IUPAC (<http://www.iupac.org>) name, brand name if any, development phase if the compound is present in MDDR march 2012 version), external links to the compound within other databases such as PubChem (21) and ChEMBL if any, and a link to the Pubmed article or the Wipo patent with the name of the compound as defined in the bibliographic source to facilitate its identification within the document.

In the *physicochemistry* tab, the data provided give an estimation of the physicochemical profile of the compound. This includes the compliance of the modulator with respect to the three above mentioned physicochemistry rules: Lipinski's RO5, Veber's and Pfizer's 3/75. But the user can also consult a radar chart recalling the physicochemical profile of the compound with respect to a hypothetical reference compound having the following properties: MW = 500 $g\text{mol}^{-1}$, AlogP = 5, HBD = 5, HBA = 10, TPSA = 140, RB = 10, Ar = 4, Fsp3 = 0.4 and R/S = 1. Those correspond either to the limits of both the Lipinski's RO5 and Veber's rule, or to the rule of thumb of Ritchie for the number of aromatic rings (15)–Ar, and the mean values usually observed among drugs for Fsp3 and the number of chiral centers–R/S. Finally, a principal component analysis individual map is provided to locate the position (in red) in the iPPI chemical space of the selected compound with respect

to all iPPI-DB compounds (in gray) and all the other iPPI-DB compounds on the same target (in blue).

In the *pharmacology* tab, all binding data available on the selected compound are given. A biplot also represents the lipophilic- and ligand-efficiencies of the selected compound (in red) with respect to the one of all iPPI-DB compounds (in gray) and to the one of all the other modulators available on the same target (in blue).

Finally in the *drug similarity* tab, the chemical structure of the selected compounds is shown next to the five most similar drugs that are found in the MDDR database (march 2012 version). Along with their structures, the following data are provided about the MDDR drugs: the Tanimoto index (using FCFP4 fingerprint), the development phase and the activity class (e.g. antineoplastic).

CONCLUSION

Given the importance taken by protein–protein interactions in the last decade in the field of drug discovery, new tools and data collections are necessary to address the challenge raised by this intricate class of therapeutic targets. Indeed, one avenue to assist the identification of new chemical probes on PPI is to collect successful examples of such modulators and learn from these molecules. With the iPPI-DB initiative, we hope that the presence in one database of manually annotated data with freely accessible and intuitive tools to query them will boost drug discovery and chemical biology projects for this target class. Along the same line, big data analysis of such databases and of databases containing regular compounds should help to rationalize why some compounds are capable of modulating PPI interfaces (22). We plan to proceed to regular updates of the data as the number of inhibitors available is constantly rising. To this end, by the end of the year we anticipate to add around 300 new compounds on about 20 new PPI targets. The addition of new inhibitors and new targets is key to have a representative subset of both the PPI chemical and target spaces. In addition, we will also implement stabilizers of protein–protein interactions as they represent a new era of PPI modulation that also needs to be addressed. Hopefully, our iPPI-DB initiative together with, for instance, the free online physical chemistry-toxicophore-PAINS filtering tool (23) and structure-based virtual screening server (24), will help biologists, chemists and clinicians in their attempt to discover new drugs and chemical probes for this challenging target class.

FUNDING

Funding for open access charge: French Ministry of Research through INSERM institute.

Conflict of interest statement. None declared.

REFERENCES

- Jubb,H., Higuero, A.P., Winter,A. and Blundell,T.L. (2012) Structural biology and drug discovery for protein-protein interactions. *Trends Pharmacol. Sci.*, **33**, 241–248.
- Kuenemann,M.A., Sperandio,O., Labbe,C.M., Lagorce,D., Miteva,M.A. and Villoutreix,B.O. (2015) In silico design of low molecular weight protein-protein interaction inhibitors: Overall concept and recent advances. *Prog. Biophys. Mol. Biol.*, **119**, 20–32.

3. Villoutreix, B.O., Kuenemann, M.A., Poyet, J.L., Bruzzoni-Giovanelli, H., Labbe, C., Lagorce, D., Sperandio, O. and Miteva, M.A. (2014) Drug-Like Protein-Protein Interaction Modulators: Challenges and Opportunities for Drug Discovery and Chemical Biology. *Mol. Inform.*, **33**, 414–437.
4. Arkin, M.R., Tang, Y. and Wells, J.A. (2014) Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.*, **21**, 1102–1114.
5. Sperandio, O., Reynès, C.H., Camproux, A.-C. and Villoutreix, B.O. (2010) Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov. Today*, **15**, 220–229.
6. Higuero, A.P., Schreyer, A., Bickerton, G.R., Pitt, W.R., Groom, C.R. and Blundell, T.L. (2009) Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chem. Biol. Drug Des.*, **74**, 457–467.
7. Morelli, X., Bourgeois, R. and Roche, P. (2011) Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr. Opin. Chem. Biol.*, **15**, 475–481.
8. Higuero, A.P., Jubbs, H. and Blundell, T.L. (2013) TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions. *Database (Oxford)*, **2013**, bat039.
9. Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L. and Overington, J.P. (2015) ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.*, **43**, W612–W620.
10. Basse, M.-J., Betzi, S., Bourgeois, R., Bouzidi, S., Chetrit, B., Hamon, V., Morelli, X. and Roche, P. (2012) 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res.*, **41**, D824–D827.
11. Rose, P.W., Prlic, A., Bi, C., Bluhm, W.F., Christie, C.H., Dutta, S., Green, R.K., Goodsell, D.S., Westbrook, J.D., Woo, J. *et al.* (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **43**, D345–D356.
12. Labbe, C.M., Laconde, G., Kuenemann, M.A., Villoutreix, B.O. and Sperandio, O. (2013) iPPI-DB: a manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions. *Drug Discov. Today*, **18**, 958–968.
13. Ghose, A.K., Viswanadhan, V.N. and Wendoloski, J.J. (1998) Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A*, **102**, 3762–3772.
14. Ertl, P., Rohde, B. and Selzer, P. (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.*, **43**, 3714–3717.
15. Ritchie, T.J. and Macdonald, S.J. (2009) The impact of aromatic ring count on compound developability—are too many aromatic rings a liability in drug design? *Drug Discov. Today*, **14**, 1011–1020.
16. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **46**, 3–26.
17. Veber, D.F., Johnson, S.R., Cheng, H.-Y., Smith, B.R., Ward, K.W. and Kopple, K.D. (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.*, **45**, 2615–2623.
18. Hughes, J.D., Blagg, J., Price, D.A., Bailey, S., Decrescenzo, G.A., Devraj, R.V., Ellsworth, E., Fobian, Y.M., Gibbs, M.E., Gilles, R.W. *et al.* (2008) Physicochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.*, **18**, 4872–4875.
19. UniProt, C. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
20. Hopkins, A.L., Keserü, G.M., Leeson, P.D., Rees, D.C. and Reynolds, C.H. (2014) The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discov.*, **13**, 105–121.
21. Coordinators, N.R. (2015) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **43**, D6–D17.
22. Kuenemann, M.A., Bourbon, L.M., Labbe, C.M., Villoutreix, B.O. and Sperandio, O. (2014) Which three-dimensional characteristics make efficient inhibitors of protein-protein interactions? *J. Chem. Inf. Model.*, **54**, 3067–3079.
23. Lagorce, D., Sperandio, O., Baell, J.B., Miteva, M.A. and Villoutreix, B.O. (2015) FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res.*, **43**, W200–W207.
24. Labbe, C.M., Rey, J., Lagorce, D., Vavrusa, M., Becot, J., Sperandio, O., Villoutreix, B.O., Tuffery, P. and Miteva, M.A. (2015) MTiOpenScreen: a web server for structure-based virtual screening. *Nucleic Acids Res.*, **43**, W448–W454.
25. Yu, M., Wang, Y., Zhu, J., Bartberger, M.D., Canon, J., Chen, A., Chow, D., Eksterowicz, J., Fox, B., Fu, J. *et al.* (2014) Discovery of Potent and Simplified Piperidinone-Based Inhibitors of the MDM2-p53 Interaction. *ACS Med. Chem. Lett.*, **5**, 894–899.