D824–D827 *Nucleic Acids Research, 2013, Vol. 41, Database issue* *Published online 30 November 2012*
doi:10.1093/nar/gks1002

2P2Idb: a structural database dedicated to orthosteric modulation of protein–protein interactions

Marie Jeanne Basse, Stéphane Betzi, Raphaël Bourgeas, Sofia Bouzidi, Bernard Chetrit, Véronique Hamon, Xavier Morelli* and Philippe Roche*

Laboratory of Integrative Structural and Chemical Biology (iSCB), Centre de Recherche en Cancérologie de Marseille (CRCM), CNRS UMR 7258, INSERM U 1068, Institut Paoli-Calmettes & Aix-Marseille Universités, 13009 Marseille, France

Received August 3, 2012; Revised September 18, 2012; Accepted October 1, 2012

ABSTRACT

Protein–protein interactions are considered as one of the next generation of therapeutic targets. Specific tools thus need to be developed to tackle this challenging chemical space. In an effort to derive some common principles from recent successes, we have built 2P2Idb (freely accessible at http://2p2idb.cnrs-mrs.fr), a hand-curated structural database dedicated to protein–protein interactions with known orthosteric modulators. It includes all interactions for which both the protein–protein and protein–ligand complexes have been structurally characterized. A web server provides links to related sites of interest, binding affinity data, pre-calculated structural information about protein–protein interfaces and 3D interactive views through java applets. Comparison of interfaces in 2P2Idb to those of representative datasets of heterodimeric complexes has led to the identification of geometrical parameters and residue properties to assess the druggability of protein–protein complexes. A tool is proposed to calculate a series of biophysical and geometrical parameters that characterize protein–protein interfaces. A large range of descriptors are computed including, buried accessible surface area, gap volume, non-bonded contacts, hydrogen-bonds, atom and residue composition, number of segments and secondary structure contribution. All together the 2P2I database represents a structural source of information for scientists from academic institutions or pharmaceutical industries.

INTRODUCTION

Protein–protein interactions (PPIs) represent a promising new class of attractive therapeutic targets, and the advancement in drug discovery efforts against PPIs has been recently referred as 'the unmined biology gold reserve' (1). However, PPIs are still considered as extremely difficult for targeting by small-molecules due to the structural characteristics of the interface, and specific strategies need to be undertaken to tackle this particularly challenging class of drug targets [for reviews see (2–5)]. Successes in drug discovery developments against PPI targets face two major issues, i.e. druggability assessment and adequacy of the chemical libraries used for screening. Over the last decade more and more orthosteric PPI modulators have been reported, and hundreds of small molecule inhibitors have now been developed for more than 40 PPI targets (4). Our goal is to use the structural knowledge from these success stories to derive some common principles to help future target selection and to accelerate the process of drug discovery in this field.

There are many structural databases dedicated to protein–protein complexes (6–14), to protein–ligand (15,16) or to small molecule inhibitors of PPIs (17–19). We have recently developed a hand-curated structural database (2P2Idb) by collecting information about protein–protein interfaces for which both the protein–protein and protein–inhibitor complexes have been structurally characterized, and we identified key descriptors of PPIs with known inhibitors (20). To our knowledge, 2P2Idb is the only structural database dedicated to orthosteric PPI modulators with structural information for protein–protein and protein–ligand complexes as well as for small molecule compounds. Although this database is relatively small at the moment, the hope is that, as it

*To whom correspondence should be addressed. Tel: +33 491164506; Fax: +33 491164540; Email: proche@imm.cnrs.fr
Correspondence may also be addressed to Morelli Xavier. Tel: +33 486977331; Fax: +33 486977499; Email: xavier.morelli@inserm.fr

grows, patterns will emerge for both protein–protein interfaces and small molecule inhibitors.

RESULTS

Presentation of 2P2Idb

2P2Idb is a relational database that was built through data mining from literature and by exhaustive search of the Protein Data Bank (20). To focus on orthosteric inhibitors, we have selected the cases for which both the protein–protein and protein–ligand complexes had been 3D-characterized (by X-ray or nuclear magnetic resonance) and for which the inhibitor is clearly competing at the interface. As of today, it contains 14 protein–protein complexes, 60 protein–inhibitor complexes, 16 free proteins and 55 small molecule modulators. The protein–protein complexes were subdivided into two classes corresponding to protein–peptide (cluster 1) and to globular protein–protein (cluster 2) complexes based on the number of segments at the interface. An interface segment is defined as a stretch of residues that starts and ends with interface residues and may contain intervening non-interface residues, but only in stretches of not more than four (21). The general interface properties are summarized for the two clusters in Table 1 showing that they differ notably. In particular, complexes from Cluster 1 can be disrupted with modified peptides such as staple peptides or with peptide mimetics whereas complexes that belong to Cluster 2 cannot. Furthermore, protein–protein complexes from Cluster 1 usually correspond to lower affinity complexes whereas those from Cluster 2 correspond to higher affinity complexes, on average. We have compared the general biophysical, biochemical and structural properties of the interfaces found in 2P2Idb with those of representative datasets of hetero and homodimers to establish a characteristic profile for 'druggable' protein–protein complexes (20 and Table 1). A web interface has been developed to facilitate access to pre-calculated data and to related websites.

Description of 2P2Idb web interface

Since its first release in 2010, the 2P2Idb website has been completely revisited by including a user friendly interface and more features. For each PPI family, clickable information can be found about protein–protein, protein–ligand complexes and free proteins as well as small molecule orthosteric modulators. Several links to relevant databases are provided such as published abstracts (PubMed), protein information (UniProt), 3D structures (PDBsum, PDBe), ligand properties (ChemSpider), protein–protein and protein–ligand binding affinities (PDBBind, BindingDB, ChEMBL or MOAD). A large number of pre-calculated interface parameters are accessible for each protein–protein complex. These interface descriptors include, total interface area, gap volume, percentage of charged residues, segments, non-bonded contacts, hydrogen bonds, salt bridges, disulfide bonds, secondary structure as well as atom and residue properties for each chain. The detailed list of non-bonded contacts, hydrogen bonds and salt bridges can be accessed through popup windows. Protein–protein and protein–ligand complexes can be interactively visualized through Jmol applets with customized menus and predefined representations. Furthermore, all protein structures in 2P2I database can be easily downloaded from our website (http://2p2idb.cnrs-mrs.fr/download.html) and analysed with external molecular visualization program viewers. In the downloaded files, 3D structures from the same family of complexes have been superimposed to the unbound form to facilitate user analysis and comparison. PDB structures can be downloaded by protein family or by complex type (protein–protein or protein–ligand).

2P2Iinspector: a protein–protein interface analysis tool

Several tools are available to analyse protein–protein interfaces. However, most of them are dedicated to the prediction of hotspots residues or binding pockets (22). Other servers provide structural and chemical information on protein–protein associations. Protein Interactions Calculator (PIC) is a server which computes contact information but does not calculate topological parameters such as gap volume and surface area (23). PISA is a tool for exploring macromolecular interfaces and surfaces (24). However, it is more dedicated to the prediction of probable quaternary structures from crystal contacts. In the first release of the 2P2I database, most interface parameters had been calculated through the ProtorP web

Table 1. The table provides 'means' and 'standard deviations' of several interface parameters calculated for the two classes of druggable complexes in 2P2Idb

Interface properties	$2P2I_{DB}$		Heterodimers		Homodimers	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
No. of complexes	7	7	189	336	331	1442
BASA (Å^2)	1384.7 ± 516.1	1793.3 ± 591.6	2149.2 ± 1017.6	2769.3 ± 1411.4	2307.2 ± 1503.1	3042.1 ± 1823.1
Gap volume (Å^3)	2282.8 ± 1351.5	5085.2 ± 2199.7	3906.7 ± 1745.2	6670.3 ± 3128.1	3780.8 ± 1775.1	6969.9 ± 3716.9
Non-bonded contacts	74.4 ± 27.5	94.4 ± 39.9	114.5 ± 57.2	151.5 ± 84.1	114.2 ± 80.4	164.6 ± 114.1
Total no. of segments	4.1 ± 1.1	8.1 ± 1.8	6.1 ± 2.7	10.8 ± 3.8	3.9 ± 1.1	10.7 ± 4.6
No. of hydrogen bonds	2.4 ± 1.3	3.3 ± 2.5	4.8 ± 3.7	6.7 ± 5.0	4.6 ± 5.4	7.2 ± 6.3
No. of salt bridges	0.6 ± 0.8	0.6 ± 0.8	1.8 ± 1.8	2.0 ± 1.8	1.4 ± 1.8	2.0 ± 2.5
No. of disulfide bonds	0.0 ± 0.0	0.0 ± 0.0	0.03 ± 0.2	0.03 ± 0.2	0.01 ± 0.1	0.01 ± 0.1
% Charged residues	20.9 ± 8.8	28.9 ± 11.5	28.7 ± 13.2	26.6 ± 11.7	26.6 ± 12.8	25.6 ± 11.5

Complexes from Cluster 1 correspond to protein–peptide complexes and can be disrupted with modified peptide or peptide mimetics. Complexes from Cluster 2 correspond to higher affinity complexes. Values for nonredundant representative datasets of hetero- and homo-dimeric complexes collected through the Dockground server are indicated as comparison for both classes.

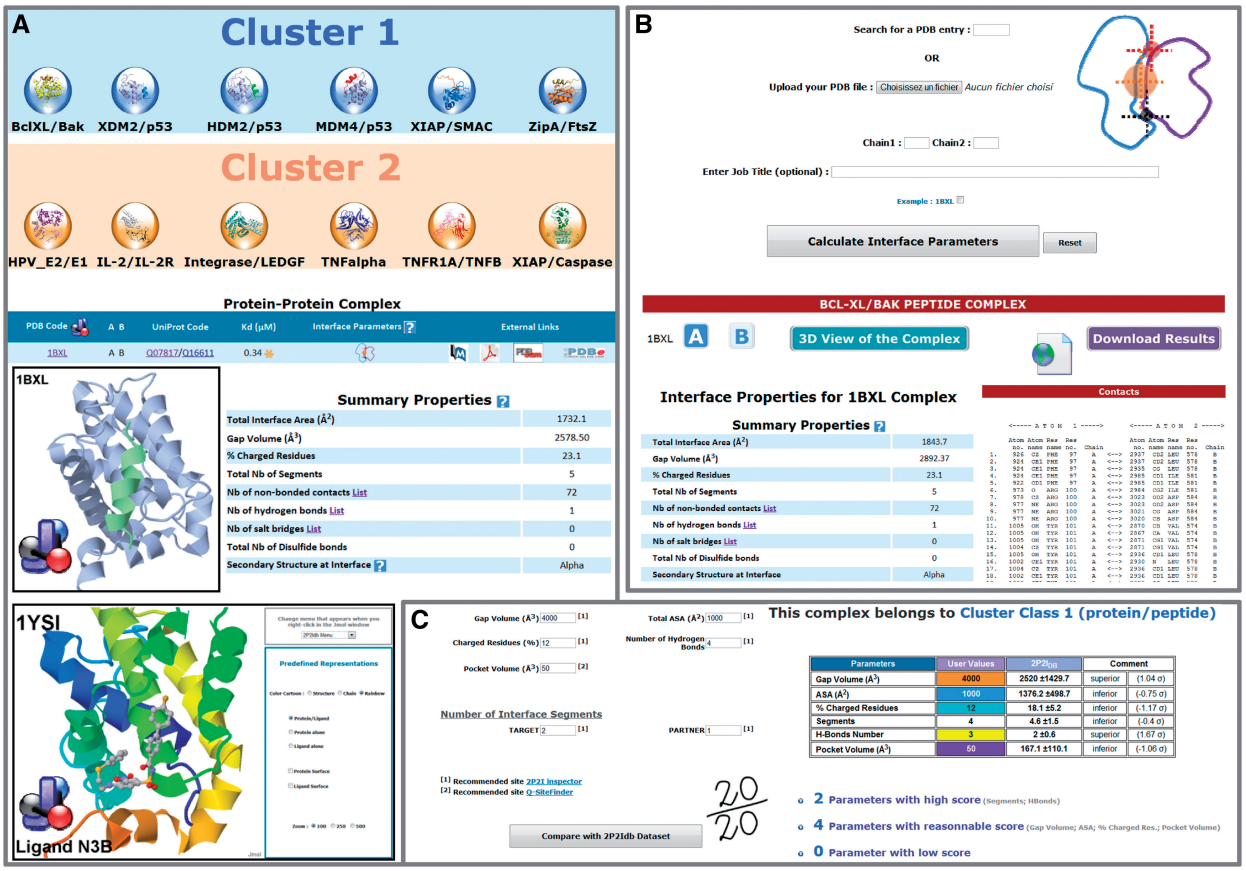

Figure 1. The 2P2I website and its main features. (**A**) 2P2Idb is a hand-curated database dedicated to the inhibition of protein–protein complexes with orthosteric modulators. It displays structural information about protein–protein, protein–ligand complexes and small molecule inhibitors. For each of the 14 families, sub-divided into two classes (protein–peptide and protein–protein), clickable html pages are provided with pre-calculated interface parameters, binding affinity data and links to related sites of interest (UniProt, PubMed, PDBsum, PDBe and ChemSpider). Protein–protein and protein–ligand complexes can be interactively visualized using Jmol applets and user-friendly menus. (**B**) 2P2Iinspector is a tool to analyse protein–protein interfaces in terms of geometric and physico-chemical descriptors. A total of 60 descriptors are computed including, buried accessible surface area, gap volume, non-bonded contacts, hydrogen-bonds, atom and residue composition, number of segments and secondary structure contribution. Users can analyze protein complexes from the PDB using standard four letter accession codes or upload their own files. (**C**) 2P2Iscore is a tool to assess the druggability of protein–protein interfaces. Comparison of protein–protein interfaces in 2P2Idb with standard heterodimers has allowed us to define six interface parameters to characterize protein–protein interfaces with a known modulator. Users are invited to compute five parameters using the 2P2Iinspector tool. The interfacial pocket volume should be calculated with Q-SiteFinder (http://www.modelling.leeds.ac.uk/qsitefinder). A color-coded table is provided to compare user defined parameters to those in 2P2Idb. A qualitative score is given for the six key parameters to assess the druggability of the interface. Detailed help documentation is available as PDF files for the different features.

server which has been discontinued and is no longer available (25). We have therefore developed our own, and enhanced, version of this tool by computing more interface parameters. 2P2Iinspector is a complete new tool that computes interaction properties from the 3D structure of protein–protein complexes. A total of 58 descriptors are now computed using in-house tcl scripts implemented in VMD (26) and SURFNET (27). These physical and chemical parameters include a large range of descriptors such as number of segments, buried accessible surface area, gap volume, non-bonded contacts, hydrogen-bonds, secondary structure contribution, atom and residue properties, and atomic composition. This new open access tool can be used to calculate interface parameters of protein complexes either from the PDB (using valid four-letter code) or by uploading a PDB file. The computed parameters can be accessed through the web interface for both chains of the protein–protein complex. Users can easily switch from the results of one chain to the other. Popup windows give easy access to the lists of non-bonded contacts, hydrogen bonds and salt bridges. The protein–protein complexes can be visualized interactively with a Jmol applet with the same functionalities described for the 2P2I database. Files are stored for 48 h before being deleted and during that period users can access their data via a direct unique link. Finally, results can be downloaded and then easily accessed locally as html files.

2P2Iscore: assessing the druggability of protein–protein interfaces

The difficulty of targeting PPIs emphasizes the importance of target selection. From a previous study, we defined six key interface parameters to characterize protein–protein complexes with a known modulator (20). We have used these descriptors to assess the druggability of protein–protein interfaces after the target has been assigned to a cluster type (protein–peptide or protein–protein) based on

the number of segments at the interface. A qualitative prediction is proposed, which is based on the standard deviation of each of the six interface parameters to the mean of the same parameter in the equivalent 2P2I cluster.

Users can compare parameters from their own protein–protein interface to a distribution of parameters from the 2P2Idb dataset. Five parameters can be easily computed on our website using 2P2Iinspector tool. We recommend using Q-SiteFinder to calculate the remaining descriptor, i.e. the interfacial pocket volume, because this server was used to estimate the size of pockets at the interface in 2P2Idb (20,28). A qualitative score is given for each parameter and a color coded table provides the deviation of each parameter compared to the mean of the same parameter in the equivalent 2P2I cluster.

CONCLUSIONS AND FUTURE DEVELOPMENTS

The 2P2I website (Figure 1) provides structural information about the modulation of PPIs with orthosteric inhibitors. A number of features give access to pre-calculated interface parameters and to related websites. A scoring function is available to qualitatively assess the druggability of protein–protein interfaces with prior 3D knowledge (2P2Iscore). A tool has been specifically developed to analyse protein–protein interfaces in terms of physico-chemical, topological or geometric features (2P2Iinspector). Future releases of the database will include new complexes and PPI modulators as they appear in the Protein Data Bank (29), new interface parameters (more particularly interfacial pockets) for the 2P2Iinspector tool and an automated version of 2P2Iscore with a quantitative scoring function.

We expect that this new version of the 2P2I database provides a useful source of information to characterize protein–protein interfaces and to design modulators of PPIs and is therefore of major interest for the scientific community.

ACKNOWLEDGEMENTS

The authors thank colleagues at the Cancer Research Center of Marseille for testing the webserver during construction.

FUNDING

Funding for open access charge: Agence Nationale de la Recherche [ANR-11-BS07-019-02].

Conflict of interest statement. None declared.

REFERENCES

1. Mullard,A. (2012) Protein–protein interaction inhibitors get into the groove. *Nat. Rev. Drug Discov.*, **11**, 173–175.
2. Wells,J.A. and McClendon,C.L. (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*, **450**, 1001–1009.
3. Roche,P. and Morelli,X. (2010) In Miteva M (ed). In silico lead discovery. *Bentham Science Publishers, eBook*, pp.118–143.
4. Morelli,X., Bourgeas,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.
5. Bienstock,R.J. (2012) Computational drug design targeting protein-protein interactions. *Curr. Pharm. Des.*, **18**, 1240–1254.
6. Davis,F.P. and Sali,A. (2005) PIBASE: a comprehensive database of structurally defined protein interfaces. *Bioinformatics*, **21**, 1901–1907.
7. Gong,S., Park,C., Choi,H., Ko,J., Jang,I., Lee,J., Bolser,D.M., Oh,D., Kim,D.-S. and Bhak,J. (2005) A protein domain interaction interface database: InterPare. *BMC Bioinformatics*, **6**, 207.
8. Teyra,J., Doms,A., Schroeder,M. and Pisabarro,M.T. (2006) SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces. *BMC Bioinformatics*, **7**, 104.
9. Winter,C., Henschel,A., Kim,W.K. and Schroeder,M. (2006) SCOPPI: a structural classification of protein-protein interfaces. *Nucleic Acids Res.*, **34**, D310–D314.
10. Jefferson,E.R., Walsh,T.P., Roberts,T.J. and Barton,G.J. (2007) SNAPPI-DB: a database and API of Structures, iNterfaces and Alignments for Protein-Protein Interactions. *Nucleic Acids Res.*, **35**, D580–D589.
11. Kundrotas,P.J. and Alexov,E. (2007) PROTCOM: searchable database of protein complexes enhanced with domain-domain structures. *Nucleic Acids Res.*, **35**, D575–D579.
12. Günther,S., von Eichborn,J., May,P. and Preissner,R. (2009) JAIL: a structure-based interface library for macromolecules. *Nucleic Acids Res.*, **37**, D338–D341.
13. Stein,A., Panjkovich,A. and Aloy,P. (2009) 3did Update: domain-domain and peptide-mediated interactions of known 3D structure. *Nucleic Acids Res.*, **37**, D300–D304.
14. Bickerton,G.R., Higueruelo,A.P. and Blundell,T.L. (2011) Comprehensive, atomic-level characterization of structurally characterized protein-protein interactions: the PICCOLO database. *BMC Bioinformatics*, **12**, 313.
15. Schüttelkopf,A.W. and van Aalten,D.M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 1355–1363.
16. Schreyer,A. and Blundell,T. (2009) CREDO: a protein-ligand interaction database for drug discovery. *Chem. Biol. Drug. Des.*, **73**, 157–167.
17. Higueruelo,A.P., Schreyer,A., Bickerton,G.R.J., 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.
18. Reynes,C., Host,H., Camproux,A.C., Laconde,G., Leroux,F., Mazars,A., Deprez,B., Fahraeus,R., Villoutreix,B.O. and Sperandio,O. (2010) Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods. *PLoS Comput. Biol.*, **6**, e1000695.
19. Sperandio,O., Reynès,C., Camproux,A. and Villoutreix,B. (2010) Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov. Today*, **15**, 220–229.
20. Bourgeas,R., Basse,M.-J., Morelli,X. and Roche,P. (2010) Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database. *PLoS One*, **5**, e9598.
21. Pal,A., Chakrabarti,P., Bahadur,R., Rodier,F. and Janin,J. (2007) Peptide segments in protein-protein interfaces. *J. Biosci.*, **32**, 101–111.
22. Marabotti,A. and Milanesi,L. (2012) Finding inhibitors of protein-protein interactions (i-ppis): a support from bioinformatics. *World Res. J. Peptide Protein*, **1**, 9–20.
23. Tina,K.G., Bhadra,R. and Srinivasan,N. (2007) PIC: protein interactions calculator. *Nucleic Acids Res.*, **35**, W473–W476.
24. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
25. Reynolds,C., Damerell,D. and Jones,S. (2009) ProtorP: a protein-protein interaction analysis server. *Bioinformatics*, **25**, 413–414.
26. Hsin,J., Arkhipov,A., Yin,Y., Stone,J.E. and Schulten,K. (2008) Using VMD: an introductory tutorial. *Curr. Protoc. Bioinformatics*, Chapter 5, Unit 5.7.
27. Laskowski,R. (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph*, **13**, 323–330, 307–328.
28. Laurie,A. and Jackson,R. (2005) Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics*, **21**, 1908–1916.
29. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.