Exploring Protein-Protein Interactions as Drug Targets for Anti-cancer Therapy with In Silico Workflows

Alexander Goncearenco, **Minghui Li**, **Franco L. Simonetti**, **Benjamin A. Shoemaker**, and **Anna R. Panchenko**

Abstract

We describe a computational protocol to aid the design of small molecule and peptide drugs that target protein-protein interactions, particularly for anti-cancer therapy. To achieve this goal, we explore multiple strategies, including finding binding hot spots, incorporating chemical similarity and bioactivity data, and sampling similar binding sites from homologous protein complexes. We demonstrate how to combine existing interdisciplinary resources with examples of semi-automated workflows. Finally, we discuss several major problems, including the occurrence of drug-resistant mutations, drug promiscuity, and the design of dual-effect inhibitors.

Keywords

Protein-protein interactions; Cancer; Protein interaction inhibitors; Drug resistance mutations

1 Introduction

Protein-protein interactions (PPI) are remarkably complex and diverse. They form intricate interaction networks where some proteins (so-called hubs) make connections to many other proteins. Despite the diversity and complexity, there are recurrent motifs in the PPI networks and protein-protein binding sites [1, 2]. PPI interaction networks can be perturbed by differential gene expression and disease mutations [3–6] and a lot of efforts have been undertaken to study how signal flow in the networks is altered by these factors and how drugs can reestablish the intricate balance. Targeting of PPIs for therapeutic intervention is very challenging because of the size and shape of their binding interfaces, e.g., planar interfaces lacking binding pockets, the difficulty to construct a functional assay to screen out affected interactions. Nevertheless, for the last several years there has been a considerable interest in PPI drug targets. Furthermore, many existing drugs have promiscuous effects on interaction pathways affecting highly connected protein hubs, while PPI inhibitors can be designed to achieve highly specific drug binding to a target protein complex. The success stories include the design of inhibitors of bromo-domains [7], Bax-BclXL [8], p53-MDM2 [9], and VEGF receptor [10]. The up-to-date database 2P2Idb v.2 contains 27 structurally characterized protein-protein complexes and 274 protein-inhibitor complexes with 242 unique small molecule inhibitors [11].

Predicting the phenotypic effects of mutations and molecular mechanisms of drugs cannot be achieved without the detailed knowledge of cellular pathways, interaction networks, and PPI binding interfaces. Studies on PPI binding site properties go back several decades, and

some of them reported an extensive similarity between protein interfaces even in the absence of sequence or structure similarity between proteins [12–15]. Since the current coverage of the human proteome with protein structural complexes remains limited, and the number of novel PPI binding sites in the PDB database grows slower than the number of protein complexes [16, 17], different computational methods have been proposed to close this gap. It has been shown that protein interactions and binding sites for human complexes can be predicted from homologous protein structural complexes of another species [18–20]. Such an inference approach may be supported by a recent work that examined the growth dynamics and evolutionary roots of PPI binding sites and found that a majority of binding sites could be traced back to the universal common ancestor of all cellular organisms [16].

There are several challenges in the structural design of PPI inhibitors. First, the inhibitor should bind relatively strongly to the protein target. This is achieved by mimicking the interactions between two proteins, especially those interactions between residues that contribute most to the binding energy, so-called binding hot spots. Second, many anti-cancer therapies are prone to acquired drug resistance which is a major challenge in cancer treatment. Therefore, the designed inhibitor (peptide or small molecule) should retain its properties, even if two interacting target proteins undergo extensive selection in the tumor to eliminate binding to these inhibitors while retaining binding between the two proteins. The latter task can be accomplished by inhibitor design along with in-silico mutagenesis. Finally, the inhibitors should be active with respect to not only the proposed PPI targets but also with respect to their paralogs. We illustrate the PPI inhibitor design principles using the p53-MDM2 interaction as an example.

2 Materials

2.1 Online Resources

1. IBIS https://www.ncbi.nlm.nih.gov/Structure/ibis/ibis.cgi [21, 22]; IBIS API https://www.ncbi.nlm.nih.gov/research/ibis-api/.

2. PubChem BioAssay (https://pubchem.ncbi.nlm.nih.gov) [23, 24].

3. MutaBind (https://www.ncbi.nlm.nih.gov/research/mutabind/) [25, 26].

4. NCI drug dictionary (http://www.cancer.gov/publications/dictionaries/cancer-drug) [27].

5. Drug sensitivity in cancer (http://www.cancerrxgene.org/translation/Drug).

6. 2P2Idb database v2 (http://2p2idb.cnrs-mrs.fr/) [11].

7. Binding DB (https://www.bindingdb.org/bind/) [28].

8. SAAMBE and HotRegion (http://prism.ccbb.ku.edu.tr/hotregion/) [29, 30].

2.2 Software

1. KNIME (https://www.knime.org).

3 Methods

3.1 Practical Challenges in Cancer Drug Design: p53-MDM2 Inhibitors

Cytotoxicity of cancer drugs may lead to genomic instability and consequently to the activation of the p53 tumor suppressor. E3 ubiquitin-protein ligase (MDM2) endogenously inhibits p53, and the disruption of an interaction between the p53 transactivation domain and MDM2 leads to the activation of p53. MDM2 negatively regulates the p53 pathways and is often overexpressed in tumor cells. Different small molecule and stapled peptide-based drugs have been designed to inhibit the p53-MDM2 interaction [31, 32]. One of the very well-studied small molecule p53-MDM2 inhibitors is Nutlin-3 that exhibits anti-cancer effects even in those cells that do not express functional p53 via mechanisms involving p73 and E2F1 activation [33]. Other p53-MDM2 inhibitors are currently under clinical trials. Stapled peptides that inhibit protein-protein interactions are an emerging class of drugs since they can better resist the cancer clonal selection leading to drug resistance. Their effectiveness can be explained by a more extensive interface between the peptides and proteins compared to small molecules, so that a larger number of acquired drug-resistant mutations on the p53-MDM2 interface can be targeted.

3.2 Exploring Binding Interfaces of Homologous Protein Complexes

To find protein complexes that can potentially be targeted by small molecules, one can examine a superposition of protein-protein and protein-small molecule binding interfaces in the existing structural complexes and find if their binding modes overlap. This procedure can help detect druggable PPIs. Several methods have been proposed that explore different types of structurally homologous complexes to detect so-called multibinding interfaces [34, 35]. In the current study, we use the Inferred Biomolecular Interaction Server (IBIS) method [21, 22] that clusters similar binding sites found in homologous proteins based on their sequence and structure conservation, and validates these using various approaches. Analysis of IBIS binding site clusters offers an opportunity to directly compare PPI interfaces with protein-small molecule or with protein-peptide interfaces to identify those interfaces that can be potentially targeted by small molecules or peptides.

According to IBIS, there are three conserved binding site clusters between MDM2 and small molecules, and one conserved binding site cluster between MDM2 and p53 (PDB 1YCR used as a query in IBIS). IBIS allows one to compare binding sites of MDM2-p53 interactions with the binding sites of MDM2-small molecule complexes. As can be seen in Fig. 1, eight sites (sites 38, 41, 42, 45, 46, 77, 80, and 84) from the second protein-small molecule binding site cluster extensively overlap with the MDM2-p53 binding site cluster. It is also clear that the third small molecule binding site cluster overlaps with MDM2-p53 only partially (by one to two residues) and even without activity assays, one might suggest that the small molecules from this cluster might not be good inhibitors of this interaction and therefore not relevant drugs.

3.3 Identifying Binding Hot Spot Sites

It is challenging to develop modulators of protein–protein interactions because protein–protein interfaces are flat and lack binding pockets [37]. However, binding energy is usually

determined by a small number of binding "hot spots" [38]. Targeting binding hot spots in drug design is a simpler and more straightforward strategy. Resistance developed within the tumor cell subpopulations by utilizing preexisting or acquired mutations (drug resistance mutations) may also be predicted from examining binding hot spots.

Binding hot spots can be identified in silico, by calculating the evolutionary conservation of binding sites or by scanning of PPI interfaces and calculating changes in binding affinity upon amino acid substitutions (for example alanine scanning). This task can be achieved by recently developed methods MutaBind and Hot-Region. MutaBind uses molecular mechanics force fields, statistical potentials, and fast side-chain optimization algorithms [25]. The MutaBind server maps mutations on a structural protein complex, calculates the associated changes in binding affinity, determines the deleterious effect of a mutation, estimates the confidence of this prediction, and produces a mutant structural model for download. The HotRegion approach uses residue network topology and a statistical pairwise contact energy function [30].

Below is a step-by-step protocol to predict the impact of mutations on binding affinity of MDM2-p53 using MutaBind:

1.   Open MutaBind web page (*see* Subheading 2) and use the PDB code 1YCR as an input.

2.   Drag and drop "chain A" (MDM2) into the "Partner 1" box and "chain B" (p53) into the "Partner 2" box.

3.   Select "Chain A" and binding site residues annotated by IBIS (Fig. 1b) to mutate; mutate them into alanine. Each mutation will be treated independently. Please note the residue numbering difference between PDB (MutaBind uses PDB numbering) and IBIS (IBIS uses the full protein sequence numbering).

Table 1 shows that 12 alanine-scanning mutations are predicted to have deleterious destabilizing effects on the MDM2-p53 interaction (potential binding hot spots). Among these potential binding hot spots, eight are responsible for the interaction with Nutlin-3a inhibitor, whereas nine alanine-scanning mutations are predicted to be deleterious for the paralogous MDMX-p53 interaction (see the next section). Overall, five evolutionarily conserved binding hot spots overlap between MDM2-p53, MDM2-Nutlin-3a, and MDMX-p53 complexes are highlighted in bold fonts in Table 1.

3.4 Polypharmacology of Dual PPI Inhibitors

Drugs, including PPI inhibitors, may interact with multiple targets, inhibiting multiple PPIs and affecting various pathways. While cytotoxic in some cases, such polypharmacologic effects are desirable in other cases. For example, while targeting the p53-MDM2 interaction, the potency for paralogous interaction with MDMX should also be achieved. MDMX is structurally similar to MDM2 but lacks its p53 ubiquitin-mediated degradation activity. Nevertheless, MDMX can repress p53 and the effectiveness of p53-MDM2 inhibitors would be compromised in cells overexpressing MDMX [39]. Currently, several small-molecule dual p53-MDMX/MDM2 inhibitors are under development [40].

A solution to this problem of finding dual inhibitors requires extensive sampling of protein–small molecule binding modes in the space of similar binding interfaces of homologous complexes. For example, by querying IBIS with the MDMX-p53 structural complex (PDB 3DAB), one can see that MDM2-p53 and MDMX-p53 binding interfaces are similar (explore the first binding site cluster for "protein-peptide" interactions in IBIS, since disordered p53 is classified as peptide). Figure 2 illustrates a structural superposition of MDM2, p53 peptide, and Nutlin-3a small molecule. According to IBIS and MutaBind, the major binding hot spot residues (in bold in Table 1) are retained in both paralogous proteins and can be targeted in designing inhibitors that disrupt both p53-MDM2 and p53-MDMX interactions (Table 1).

In addition to the IBIS webserver, we provide an application program interface (API), discussed in the sections below, to identify binding sites from homologous structures and to find other compounds that would bind homologous complexes (Table 2, Fig. 3). We illustrate a combination of IBIS APIs with PubChem APIs in an example workflow in Figs. 4 and 5.

3.5 Incorporating Chemical Similarity and Biological Activity Data

While the biological relevance of binding interfaces via protein similarities can be assessed using the IBIS method, putative drug target alternatives can also be explored using the PubChem database [23]. PubChem is an open archive of chemical biology information including an expansive set of putative drug targets and bioactivity test results. PubChem currently contains over 92M unique chemical compounds and 3.6M substances evaluated for biological activity involving over 10K protein and 20K gene targets. Information can be obtained using the web interface, downloading from the FTP site or using programmatic tools as utilized in this protocol.

Similar compounds can have similar or optimized biological actions, such as the disruption of a protein-protein interaction binding interface. Given a query small molecule, a set of similar compounds can be retrieved via the proposed IBIS API, which, in turn, makes use of the PubChem REST API including both 2D and 3D chemical similarity searches. These two types of similarity searches complement each other. For 2D compound similarity, precomputed sets of binary fingerprints representing the constituent chemical substructures are compared to ensure that the key components of a query molecule are preserved. Using a 3D similarity search, shape features of the molecules can be compared to minimize steric hindrance in binding pockets.

In some cases, as for example in PDB 4HG7, a co-crystallized molecule corresponds to the compound of interest, nutlin-3a. In other cases (e.g., 2NOU, 1T4E, 3VGB, 3U15), the bound small molecules may be different enough to ensure that a compound receives a unique identifier, but their biological activity may be similar. In order to address such cases, an "expansion" of query compounds using similar compounds may be very useful for identifying all relevant hits. We illustrate this approach in an example workflow in Figs. 4 and 5, where an IBIS API (*see* also Table 2) is used to find similar compounds, and then for each compound a list of proteins is found that bind this compound. Finally, all protein-small molecule binding sites are identified for each structure in the list.

In the search for putative drug targets, PubChem offers a wealth of pertinent information to help choose one alternative over another. In combination with the chemical structure search, the large collection of bioactivity test results in PubChem [23, 24] provide the details required to narrow down a list of candidates. In a first pass, molecules designated "active," as defined by the original experimentalist, can be compared using filters on bioassay types like toxicity studies. If required, detailed analysis can be performed on individual assay experiments which contain large amounts of supporting data and explanations.

Finally, PubChem has collected extensive annotations for many chemical compounds. These include, but are not limited to, safety and hazard, pharmacological, toxicity, patent, vendor, interaction and pathway, classification and synonym annotations. Such annotations impact clinical and practical considerations like adverse reactions and purchasing. They also assist in making connections among existing literature references and categorizations.

3.6 Building Semiautomatic Workflows Using REST Services

Over the years, computational biology and computational chemistry communities have built a large number of services and resources. Some resources are integral parts of infrastructures provided by the large players such as EBI, NCBI, or PDB. However, integrating these resources in data analysis pipelines remains a tedious task that takes up significant amounts of project time and resources.

Some databases and web services can be combined in a modular way in workflows, thereby empowering users to perform complex interdisciplinary analytical processes. A workflow integrates a range of services to perform data-processing tasks. A service can be local or remote, it receives data and parameters as input, processes the request and returns the results. Although each service has a simple interface, complexity of workflows arises from the combination of services. The structure of a workflow is defined by the user and specific tasks. Automated workflows provide a way toward reproducible research. However, reproducibility becomes a complicated issue with distributed services as part of the workflow. Web services typically have a version prefix to make their interfaces and behavior backward-compatible. However, the databases are perpetually being updated and the API calls may return different data the next time a workflow is executed.

Pipelines and automated workflows have a long history in drug discovery, particularly with the advent of high-throughput screening. As robots speed up experiments at the bench, in-silico drug-screening is accelerated by computational pipelines [41]. For instance, the Chemistry Development Toolkit (CDK) has plugins to both KNIME and Taverna [42, 43]. A popular commercial small-molecule drug discovery and protein modeling suite by Schrödinger is automated with KNIME. KNIME may be more suited as a data mining tool [44] and may therefore be more appropriate for drug discovery applications involving data processing in addition to automated execution.

Several primers on connecting proteins with compounds using workflows and distributed web services exist, for example using BindingDB and KNIME [28, 45]. One of our goals here is to provide a primer on data mining in drug discovery for modulating and inhibiting protein-protein interactions. We exemplify an approach implemented with KNIME. Taverna

Workflow Management System [46] is an open platform under the Apache project, guaranteeing a long-term development commitment. Taverna offers prebuilt packages with Bioinformatics services and a comprehensive online repository for open exchange of workflows between users—MyExperiment [47]. The latter is a social network for the scientific community facilitating sharing of workflows and snippets. Creators and maintainers of web services may choose to list them in online catalogs, making it easier for the users to find, annotate, and monitor distributed services. Biocatalogue [48] is the largest online registry of SOAP and REST Web services, and all its services can be loaded into Taverna Workbench or into KNIME.

Web services operate on the top of the HTTP protocol, which is used as transport for the web. XML-based web-services, such as SOAP and WSDL specifications, became widespread in computational biology and cheminformatics, and are still well maintained. However, standardization of web browsers and JavaScript—frontend and backend frameworks stimulated a transition to RESTful services using a lightweight JSON-based (JavaScript Object Notation) data exchange. While XML-derived data formats remain heavily used, many services now support JSON in addition to XML. With REST, services behave more like documents or resources that are uniquely identified with a URI address. Not only does this approach simplify the development and scalability of servers but also allows one to create clients in a simple way in any modern scripting language: Perl, Python, or JavaScript.

Table 2 describes a list of IBIS REST resources that we created to facilitate drug discovery workflows aimed at protein-protein interaction inhibitors. All the following resources use HTTPS "GET" to retrieve single entries and "POST" to submit multiple entries. Examples of JSON formatted results for two IBIS API calls (a) requesting inferred protein-protein binding sites for a given protein structure and (b) requesting similar compounds given a PubChem compound ID are shown in Fig. 3. These API calls may be arranged in workflows accepting protein and gene names, protein structure, or small molecule identifier as input. A listing of a Python script (Fig. 4) illustrates the workflow where output of one API call is passed as input to another API call, thus creating a rather complex analysis pipeline. The script in Fig. 4 retrieves a list of small molecule compounds similar to Nutlin-3a, identifies the PDB structures where any of these compounds is cocrystallized with a protein, and uses IBIS to retrieve description of protein-small molecule binding sites. Due to the simplicity of implementation of HTTP clients, the script implementing this pipeline remains relatively short and easy to read. In addition to programmatic access to IBIS REST API, a graphical representation of workflows can be created, for example using KNIME, as we illustrate in Fig. 5. The diagram shown in Fig. 5 implements the same workflow as in the Python script shown in Fig. 4.

Acknowledgments

References

1. Shoemaker BA, Panchenko AR, Bryant SH. Finding biologically relevant protein domain interactions: conserved binding mode analysis. Protein Sci. 2006; 15(2):352–361. [PubMed: 16385001]

2. Aloy P, Russell RB. Ten thousand interactions for the molecular biologist. Nat Biotechnol. 2004; 22(10):1317–1321. [PubMed: 15470473]

3. Rolland T, Tasan M, Charloteaux B, et al. A proteome-scale map of the human interactome network. Cell. 2014; 159(5):1212–1226. [PubMed: 25416956]

4. Yates CM, Sternberg MJ. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. J Mol Biol. 2013; 425(21):3949–3963. [PubMed: 23867278]

5. Teng S, Madej T, Panchenko A, Alexov E. Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. Biophys J. 2009; 96(6):2178–2188. [PubMed: 19289044]

6. Li M, Goncearenco A, Panchenko AR. Annotating mutational effects on proteins and protein interactions: designing novel and revisiting existing protocols. Methods Mol Biol. 2017; 1550:235–260. [PubMed: 28188534]

7. Filippakopoulos P, Qi J, Picaud S, et al. Selective inhibition of BET bromodomains. Nature. 2010; 468(7327):1067–1073. [PubMed: 20871596]

8. Smith BJ, Lee EF, Checco JW, et al. Structure-guided rational design of alpha/beta-peptide foldamers with high affinity for BCL-2 family prosurvival proteins. Chembiochem. 2013; 14(13):1564–1572. [PubMed: 23929624]

9. Zhao Y, Aguilar A, Bernard D, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction (MDM2 Inhibitors) in clinical trials for cancer treatment. J Med Chem. 2015; 58(3):1038–1052. [PubMed: 25396320]

10. Haase HS, Peterson-Kaufman KJ, Lan Leven-good SK, et al. Extending foldamer design beyond alpha-helix mimicry: alpha/beta-peptide inhibitors of vascular endothelial growth factor signaling. J Am Chem Soc. 2012; 134(18):7652–7655. [PubMed: 22548447]

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

12. Aloy P, Ceulemans H, Stark A, Russell RB. The relationship between sequence and interaction divergence in proteins. J Mol Biol. 2003; 332(5):989–998. [PubMed: 14499603]

13. Ma B, Elkayam T, Wolfson H, Nussinov R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proc Natl Acad Sci U S A. 2003; 100(10):5772–5777. [PubMed: 12730379]

14. Valdar WS, Thornton JM. Protein-protein interfaces: analysis of amino acid conservation in homodimers. Proteins. 2001; 42(1):108–124. [PubMed: 11093265]

15. Guharoy M, Chakrabarti P. Conservation and relative importance of residues across protein-protein interfaces. Proc Natl Acad Sci U S A. 2005; 102(43):15447–15452. [PubMed: 16221766]

16. Goncearenco A, Shaytan AK, Shoemaker BA, Panchenko AR. Structural perspectives on the evolutionary expansion of unique protein-protein binding sites. Biophys J. 2015; 109(6):1295–1306. [PubMed: 26213149]

17. Goncearenco A, Shoemaker BA, Zhang D, et al. Coverage of protein domain families with structural protein-protein interactions: current progress and future trends. Prog Biophys Mol Biol. 2014; 116(2–3):187–193. [PubMed: 24931138]

18. Petrey D, Honig B. Structural bioinformatics of the interactome. Annu Rev Biophys. 2014; 43:193–210. [PubMed: 24895853]

19. Mosca R, Ceol A, Aloy P. Interactome3D: adding structural details to protein networks. Nat Methods. 2013; 10(1):47–53. [PubMed: 23399932]

20. Tyagi M, Hashimoto K, Shoemaker BA, et al. Large-scale mapping of human protein interactome using structural complexes. EMBO Rep. 2012; 13(3):266–271. [PubMed: 22261719]

21. Shoemaker BA, Zhang D, Thangudu RR, et al. Inferred Biomolecular Interaction Server—a web server to analyze and predict protein interacting partners and binding sites. Nucleic Acids Res. 2010; 38:D518–D524. Database issue. [PubMed: 19843613]

22. Shoemaker BA, Zhang D, Tyagi M, et al. IBIS (Inferred Biomolecular Interaction Server) reports, predicts and integrates multiple types of conserved interactions for proteins. Nucleic Acids Res. 2012; 40:D834–D840. Database issue. [PubMed: 22102591]

23. Kim S, Thiessen PA, Bolton EE, et al. PubChem Substance and Compound databases. Nucleic Acids Res. 2016; 44(D1):D1202–D1213. [PubMed: 26400175]

24. Wang Y, Bryant SH, Cheng T. PubChem BioAssay: 2017 update. Nucleic Acids Res. 2017; 45(D1):D955–D963. [PubMed: 27899599]

25. Li M, Simonetti FL, Goncearenco A, Panchenko AR. MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions. Nucleic Acids Res. 2016; 44(W1):W494–W501. [PubMed: 27150810]

26. Li M, Petukh M, Alexov E, Panchenko AR. Predicting the impact of missense mutations on protein? Protein binding affinity. J Chem Theor Comput. 2014; 10(4):1770–1780.

27. Yang W, Soares J, Greninger P. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013; 41:D955–D961. Database issue. [PubMed: 23180760]

28. Gilson MK, Liu T, Baitaluk M, Nicola G, et al. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res. 2016; 44(D1):D1045–D1053. [PubMed: 26481362]

29. Petukh M, Dai L, Alexov E. SAAMBE: webserver to predict the charge of binding free energy caused by amino acids mutations. Int J Mol Sci. 2016; 17(4):547. [PubMed: 27077847]

30. Cukuroglu E, Gursoy A, Keskin O. HotRegion: a database of predicted hot spot clusters. Nucleic Acids Res. 2012; 40:D829–D833. Database issue. [PubMed: 22080558]

31. Estrada-Ortiz N, Neochoritis CG, Domling A. How to design a successful p53-MDM2/X interaction inhibitor: a thorough overview based on crystal structures. Chem-MedChem. 2016; 11(8):757–772.

32. Shangary S, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. Annu Rev Pharmacol Toxicol. 2009; 49:223–241. [PubMed: 18834305]

33. Cinatl J, Speidel D, Hardcastle I, Michaelis M. Resistance acquisition to MDM2 inhibitors. Biochem Soc Trans. 2014; 42(4):752–757. [PubMed: 25109953]

34. Thangudu RR, Bryant SH, Panchenko AR, Madej T. Modulating protein-protein interactions with small molecules: the importance of binding hotspots. J Mol Biol. 2012; 415(2):443–453. [PubMed: 22198293]

35. Davis FP, Sali A. The overlap of small molecule and protein binding sites within families of protein structures. PLoS Comput Biol. 2010; 6(2):e1000668. [PubMed: 20140189]

36. Marchler-Bauer A, Derbyshire MK, Gonzales NR, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015; 43(D1):D222–D226. [PubMed: 25414356]

37. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature. 2007; 450(7172):1001–1009. [PubMed: 18075579]

38. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol. 1998; 280(1):1–9. [PubMed: 9653027]

39. Graves B, Thompson T, Xia M, Janson C. Activation of the p53 pathway by small-molecule-induced MDM2 and MDMX dimerization. Proc Natl Acad Sci U S A. 2012; 109(29):11788–11793. [PubMed: 22745160]

40. Ribeiro CJ, Rodrigues CM, Moreira R, Santos MM. Chemical variations on the p53 reactivation theme. Pharmaceuticals (Basel). 2016; 9(2)

41. Perez-Sanchez H, Rezaei V, Mezhuyev V, et al. Developing science gateways for drug discovery in a grid environment. Spring. 2016; 5(1):1300.

42. Beisken S, Meinl T, Wiswedel B, et al. KNIME-CDK: workflow-driven cheminformatics. BMC Bioinformatics. 2013; 14:257. [PubMed: 24103053]

43. Truszkowski A, Jayaseelan KV, Neumann S, et al. New developments on the cheminformatics open workflow environment CDK-Taverna. J Cheminform. 2011; 3:54. [PubMed: 22166170]

44. Mazanetz MP, Marmon RJ, Reisser CB, Morao I. Drug discovery applications for KNIME: an open source data mining platform. Curr Top Med Chem. 2012; 12(18):1965–1979. [PubMed: 23110532]

45. Nicola G, Berthold MR, Hedrick MP, Gilson MK. Connecting proteins with drug-like compounds: open source drug discovery workflows with BindingDB and KNIME. Database (Oxford). 2015 2015.

46. Wolstencroft K, Haines R, Fellows D, et al. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. Nucleic Acids Res. 2013; 41:W557–W561. Web Server issue. [PubMed: 23640334]

47. Goble CA, Bhagat J, Aleksejevs S, et al. myExperiment: a repository and social network for the sharing of bioinformatics workflows. Nucleic Acids Res. 2010; 38:W677–W682. Web Server issue. [PubMed: 20501605]

48. Bhagat J, Tanoh F, Nzuobontane E, et al. BioCatalogue: a universal catalogue of web services for the life sciences. Nucleic Acids Res. 2010; 38:W689–W694. Web Server. [PubMed: 20484378]

(a) MDM2 Sequence

Query sequence
CDD domain

SWIB

(b) Protein-Peptide Interaction

	Interaction partner sequence	Homologous complex	Homolog	Interaction partner	%Identity to query	Binding site
						9 E, 10 T, 34 M, 35 K, 37 V, 38 L, 39 F, 41 L, 42 G, 43 Q, 45 I, 46 M, 51 Y, 56 Q, 57 H, 59 V, 66 F, 70 F, 75 V, 76 K, 77 H, 78 I, 80 Y, 83 I, 84 Y, 87 I, 88 Y
Query		-	-	-	-	E T M K - L - L G - I M - Q H V - - V K H - Y - Y
☑	SQETFSDLWKLLPEN	1YCR	A	B	100	E T M K - L - L G - I M - Q H V - - V K H - Y - Y
☐	PLSQETFSDLWKLLPEN	1YCQ	A	B	73	- - - - - I - L G - I M Y Q - - - - V K P L Y - -
☐	TSFAEYWNLLSP	4RXZ	A	C	59	- - - - - M - L G - I M Y Q H V - - V K - - X - -
☐	SQETFSDLWKLLPEN	2Z5S	M	P	58	- - - - - M - L G - I M Y Q H V - - V - - - Y - -

c) Protein-small Molecule Interaction

	Chemical	Homologous complex	Chain	%Identity to query	Binding site
					34 M, 35 K, 37 V, 38 L, 39 F, 41 L, 42 G, 43 Q, 45 I, 46 M, 51 Y, 55 Q, 56 Q, 57 H, 59 V, 66 L, 70 F, 75 F, 76 S, 77 V, 78 K, 80 H, 83 I, 84 Y, 87 I
Query		-	-	-	M K - V L F L G Q I M Y Q Q H V L F F S V K H I Y I
☑	CHEMBL361103	1T4E	A	100	- - - - L - L G - I M Y - Q - - - - F - V - H I Y -
☐	CHEMBL2402574	4JSC	A	78	- - - - L - L G - I M Y - - H - - - F - V K H I Y -
☐	4-[(4s,5r)-4-(3-Chlorophenyl)-5-(4-Chlorophenyl...	2NOU	A	61	- - - M H L G - I M Y - Q H V - - F S V - - - Y -

Fig. 1.
Analysis of conserved binding sites using IBIS method with PDB 1YCR as a query. (**a**) MDM2 sequence annotation using Conserved Domain Database [36]; (**b**) Conserved protein binding sites between MDM2 and p53; (**c**) Conserved protein binding sites between MDM2 and small molecules. Residues are numbered with respect to the full protein sequence. Colors of binding site residues correspond to their conservation among homologs: conserved residues are colored in *red*, less conserved in *blue* and nonconserved in *gray*
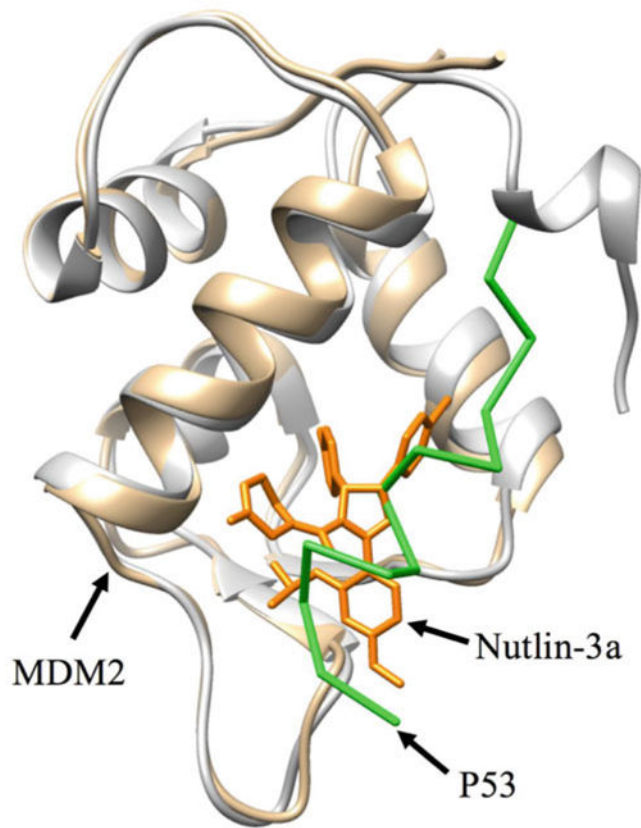
Fig. 2.
A superposition of MDM2-p53 (PDB code: 1YCR) and MDM2-Nutlin-3a complex (PDB code: 4HG7) structures

(a) /sites/inf/protein/pdb/4HG7

```
{
  "response": [
    {
      "CHAIN": "A",
      "INFERRED_PARTNER_CHAIN": "M",
      "INFERRED_PARTNER_PDB": "3DAC",
      "PDB": "4HG7",
      "SDI": 623460,
      "TEMPLATE_CHAIN": "C",
      "TEMPLATE_PARTNER_CHAIN": "",
      "TEMPLATE_PDB": "3DB2",
      "TEMPLATE_SDI": 280697,
      "residues": [
        {
          "id": 42,
          "resi": "53",
          "resn": "VAL"
        },
        {
          "id": 45,
          "resi": "56",
          "resn": "TYR"
        },
        {
          "id": 46,
          "resi": "57",
          "resn": "LEU"
        },
        ...
```

(b) /compounds/similar/compound/11433190

```
{
  "response": [
    24764418,
    67585029,
    87433224,
    68685834,
    10127376,
    118880883,
    11385877,
    118880884,
    90781721,
    118883353,
    118880885,
    118880886,
    53394469,
    ...
```

Fig. 3.
Examples of results of API calls: (**a**) to find inferred protein-protein binding sites for E3 Ubiquitin-protein Ligase MDM2 (PDB ID 4HG7); JSON structure shows a list of binding site residues from a nutlin-resistant structure of MDM2 (PDB ID 4UMN) with a stapled peptide; (**b**) A list of similar small molecule compounds for nutlin-3a (PubChem compound ID 11433190)

```python
import json
from urllib.request import urlopen
from urllib.error import HTTPError

prefix = "https://dev.ncbi.nlm.nih.gov/research/ibis-api/rest/v1"
API_similar_compound = "{}/compounds/similar/compound/{}"
API_structures_compound = "{}/structures/compound/{}"
API_sites = "{}/sites/obs/compound/pdb/{}"

combined_results = []

# Find similar compounds
query_cid = 11433190  # Nultin-3a
with urlopen(API_similar_compound.format(prefix, query_cid)) as f1:
    result_cids = json.loads(f1.read().decode('utf-8'))
    for cid in result_cids['response']:
        # Find structures for each compound
        try:
            with urlopen(API_structures_compound.format(prefix, cid)) as f2:
                result_struct = json.loads(f2.read().decode('utf-8'))
                for pdb in result_struct['response']:

                    # Find small molecules and binding sites for each structure
                    try:
                        with urlopen(API_sites.format(prefix, pdb)) as f3:
                            results_sites = json.loads(f3.read().decode('utf-8'))
                            combined_results.extend(results_sites['response'])
                    except HTTPError:
                        continue  # no small molecule binding sites found
        except HTTPError:
            continue  # no structures found for a compound

with open("results.json", 'w') as output:
    output.write(json.dumps(combined_results, indent=4, sort_keys=True))
```
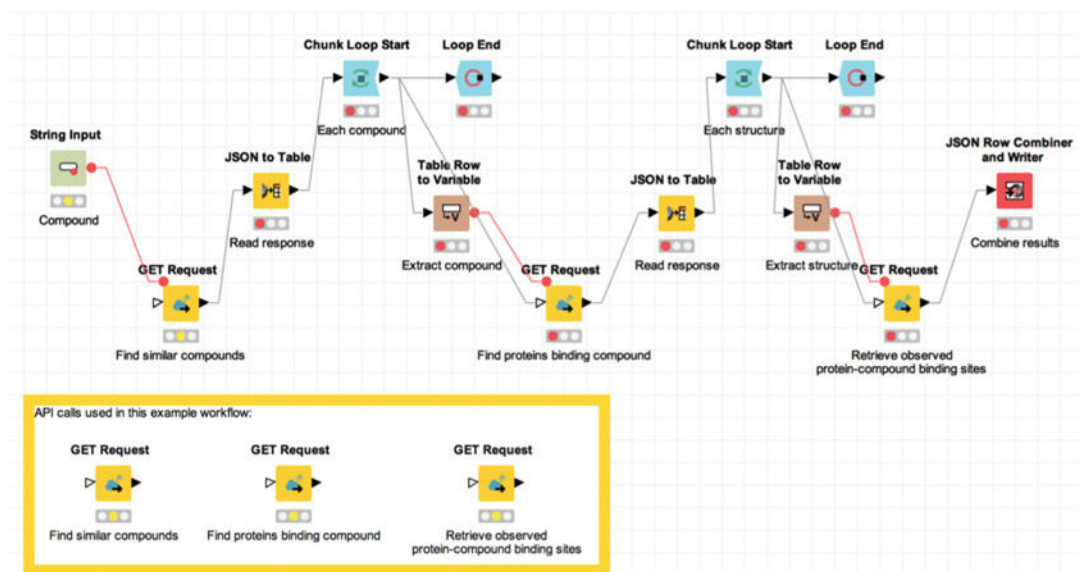
Fig. 4.

An example of a programmatic implementation (in Python 3) of a workflow for finding binding sites of similar small molecules in protein structures, given a PubChem Compound identifier. The workflow is using three IBIS API calls described in Table 2

Fig. 5.
An example of a workflow diagram built in KNIME using three REST API calls (shown in the figure legend). The workflow identifies binding sites of similar small molecules in protein structures, given a PubChem Compound identifier as an input. Nodes denote services, *while arrows* show direction of information flow in the workflow. *Red lines* show assignment of variables in the flow control structures

Table 1

Alanine scanning of binding sites annotated on MDM2 (PDB: 1YCR, chain A) or MDMX (PDB: 3DAB, chain A) by IBIS for MDM2-p53 (1YCR) and MDMX-p53 (3DAB) complexes using MutaBind webserver

MDM2/p53			MDMX/p53		
Mutation to Ala	G (kcal/mol)	Deleterious	Mutation to Ala	G	Deleterious
Glu25	0.56	No			
Thr26	0.44	No			
Met50	1.79	Yes	Val49	1.19	No
Lys51	1.01	No	Lys50	0.86	No
Val53	1.62	Yes	Met53	2.69	Yes
Leu54#	1.80	Yes			
Phe55	0.72	No			
Leu57	1.36	No	L56	1.91	Yes
Gly58#	**1.97**	**Yes**	**Gly57**	**2.00**	**Yes**
Gln59	0.58	No			
Ile61#	**2.12**	**Yes**	**Ile60**	**2.29**	**Yes**
Met62#	1.38	No	Met61	1.04	No
Tyr67	1.16	No	Tyr66	1.42	No
Gln72#	**1.67**	**Yes**	**Gln71**	**1.77**	**Yes**
His73#	1.69	Yes	His72	1.48	No
Val75#	**1.68**	**Yes**	**Val74**	**1.81**	**Yes**
Phe86	1.66	Yes			
Phe91	1.80	Yes			
Val93#	**2.61**	**Yes**	**Val92**	**2.22**	**Yes**
Lys94	0.65	No			
His96#	1.16	No			
Ile99#	2.05	Yes	Leu98	1.95	Yes
Tyr100#	1.12	No	Tyr99	2.68	Yes
Ile103	1.28	No			
Tyr104	0.14	No			

Observed binding sites in MDM2-Nutlin-3a complex (PDB: 4HG7) are labeled with "#." *G* refers to the change in binding affinity upon mutations, positive values correspond to the destabilizing effects. Overlapping binding hot spots (according to MutaBind and HotRegion) among three complexes, MDM2-p53, MDM2-Nutlin-3a, and MDMX-p53, are highlighted in bold fonts. Because of the difference in PDB (used in MutaBind) and full sequence (used in IBIS) numbering, there is a 16 residues offset and Glu25 in MutaBind or PDB corresponds to Glu9 in IBIS

Table 2

Applications program interface (API) to extract and process data on protein-protein and protein-small molecule interactions

URI	Description of retrieved results
/structures/protein/<PROTEIN>	PDB and chain identifiers given a human gene name or a UniProt protein name
/structures/compound/<CID>	PDB identifiers containing a given PubChem compound
/sites/obs/protein/pdb/<PDB>[/<CHAIN>]	Observed *protein-protein* binding site residues given a PDB identifier and an optional chain identifier. Each binding site specifies two interacting chains, the interacting domains and a list of residues
/sites/obs/compound/pdb/<PDB>[/<CHAIN>]	Observed *protein-small molecule* binding sites for each compound in a given protein structure
/sites/obs/peptide/pdb/<PDB>[/<CHAIN>]	Observed *protein-peptide* binding sites given a protein structure. The interacting peptide sequence is returned as well
/sites/obs/compound/compound/<CID>	Proteins and observed *protein-small molecule* binding sites given a compound of interest
/sites/inf/protein/pdb/<PDB>[/<CHAIN>]	Inferred *protein-protein* binding site residues given a protein structure
/sites/inf/peptide/pdb/<PDBID>[/<CHAIN>]	Inferred *protein-peptide* binding site residues given a PDB and an optional chain identifier
/sites/inf/compound/pdb/<PDB>[/<CHAIN>]	Inferred *protein-small molecule* binding site residues given a protein structure
/compounds/similar/compound/<CID>	Similar *small molecule compounds* given a small molecule. Performs both 2D fingerprint and 3D fast searches

All IBIS API resources have a prefix "https://www.ncbi.nlm.nih.gov/research/ibis-api/rest/v1," followed by more specific parts of the URI. Required parameters are specified in angle brackets, while optional parameters are specified in square brackets. The following parameters specified in angle brackets are used: <CID> PubChem compound identifier; <PDB> <CHAIN> PDB and chain identifiers. Additional instructions are available online at https://www.ncbi.nlm.nih.gov/research/ibis-api/