BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 21 2013, pages 2806–2807
doi:10.1093/bioinformatics/btt483

Databases and ontologies

Advance Access publication August 19, 2013

HippDB: a database of readily targeted helical protein–protein interactions

Christina M. Bergey[1,†], Andrew M. Watkins[2,†] and Paramjit S. Arora[2,*]

[1]Department of Anthropology and [2]Department of Chemistry, New York University, New York, NY 10003, USA

Associate Editor: Burkhard Rost

ABSTRACT

Summary: HippDB catalogs every protein–protein interaction whose structure is available in the Protein Data Bank and which exhibits one or more helices at the interface. The Web site accepts queries on variables such as helix length and sequence, and it provides computational alanine scanning and change in solvent-accessible surface area values for every interfacial residue. HippDB is intended to serve as a starting point for structure-based small molecule and peptidomimetic drug development.

Availability and implementation: HippDB is freely available on the web at http://www.nyu.edu/projects/arora/hippdb. The Web site is implemented in PHP, MySQL and Apache. Source code freely available for download at http://code.google.com/p/helidb, implemented in Perl and supported on Linux.

Contact: arora@nyu.edu

Received on May 1, 2013; revised on July 11, 2013; accepted on August 15, 2013

1 INTRODUCTION

Protein–protein interactions (PPIs) mediate fundamental signaling pathways and cellular processes. Although PPIs are highly promising pharmaceutical targets, they are not preferred targets in conventional drug development because of their extended flat interfaces. In particular, compound libraries for high-throughput screening that offer attractive lead compounds for enzymatic targets lack the topological and functional complexity necessary for PPI inhibition (Hajduk and Greer, 2007; Raj *et al.*, 2013; Wells and McClendon, 2007). One successful method to inhibit PPIs is the mimicry of secondary structure motifs that contribute to complex formation (Azzarito *et al.*, 2013; Boersma *et al.*, 2012; Jochim and Arora, 2010; Moellering *et al.*, 2009; Patgiri *et al.*, 2011).

Often a subset of the residues at a protein–protein interface can contribute significantly to the binding interaction (Clackson and Wells, 1995). Because solubility and specificity are eternal problems in drug design, it is advantageous to identify and prioritize most important residues, leaving less important positions free for fine-tuning (Bullock *et al.*, 2011; Jochim and Arora, 2010).

Conventional computational methods to predict important residues include alanine scanning (Jochim and Arora, 2010; Kortemme and Baker, 2002; Kortemme *et al.*, 2004) and solvent-accessible surface area (ΔSASA) analysis (Koes and Camacho, 2012). Alanine scanning provides the change in ΔG resulting from a contact residue being mutated to alanine, while a ΔSASA value describes how much of the residue is buried from solvent on binding. We have previously developed a scoring strategy to rank protein interfaces by their promise for synthetic inhibition (Bullock *et al.*, 2011; Jochim and Arora, 2010) and designed inhibitors of formerly 'undruggable' PPIs (Patgiri *et al.*, 2011).

To follow-up this work, we sought to derive a readily accessible resource for the chemical biology community. Research groups with potent small-molecule scaffolds might be interested in small interfaces with hotspot residues in two consecutive positions or in the *i* and *i+4* positions, whereas those developing peptoid or beta-peptide foldamers might be more interested in long interfaces with high total ΔΔG. HippDB—a database of helical interfaces in PPIs—lists all the helical PPIs in the Protein Data Bank (PDB) and catalogs computational alanine scanning results (ΔΔG in Rosetta energy units) and ΔSASA for each interfacial residue. We expect this dataset will be a useful resource for PPI inhibition.

Figure 1 depicts a typical workflow in HippDB. The user might first search for interface helices found in humans by constraining the organism name. Next, the user might trim the results for complexes with exactly three hotspot residues, then for helices <10 residues long, then for a ΔΔG average >2. By clicking on the PDB codes that result, the user can view any of the five complexes fitting these criteria in JMol, with their hotspot residues displayed in wireframe.

2 METHODS

Structures of multi-entity protein complexes' asymmetric units were obtained from the PDB (Berman *et al.*, 2000). We identified all interacting interface chains within each PDB file and created a new PDB file for each chain and each pair of interacting chains. If the original PDB file contained more than one model, only the lowest-scoring model was used (according to Rosetta's 'Relax' protocol).

Each qualifying pair of chains was analyzed using the RosettaScripts AlaScan filter, averaging 100 runs (Baker and Sali, 2001; Fleishman *et al.*, 2011). Following alanine scanning, we isolated all interface helices containing two or more hotspot residues (ΔΔG >1.0 Rosetta energy units, which approximately scale as 1 kcal/mol) and computed ΔSASA using NACCESS (Hubbard and Thornton, 1993). Interface helices were required to possess at least four consecutive residues, each assigned as helical by Dictionary of Secondary Structure Prediction acquired from the Center for Molecular and Biomolecular Informatics Web site (Kabsch and Sander, 1983). For each interface helix, parameters including average and total ΔΔG and ΔSASA, the percentage of the complex's ΔΔG and ΔSASA contributed by the helix, the helix length, hotspot

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

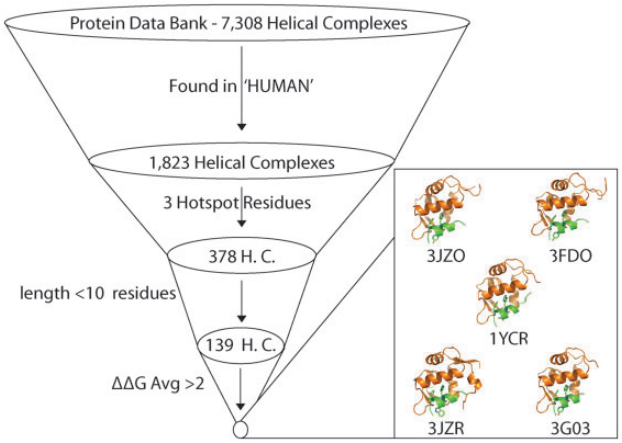

Fig. 1. A typical HippDB query. The five resulting complexes are depicted with the qualifying chain in green and the partner chain in orange. 1YCR is the native p53/mdm2 complex; 3FDO and 3JZO are complexes of p53-like synthetic peptides with mdm4 and 3G03 and 3JZR are complexes of synthetic peptides with mdm2

Table 1. A selection of the fields found in HippDB[a]

Field name	Description
Average ΔΔG, helix	Average ΔΔG contributed by a residue in the helix
Percentage ΔΔG, helix	Percentage of the chain's total ΔΔG due to the helix
Percentage ΔSASA, helix	Percentage of the chain's total ΔSASA due to the helix
Helix sequence	Sequence of the interface helix
Hotspot IDs	List of hotspot residues with residue type and ΔΔG
MimeticScore	Sum of the top three hotspot ΔΔG values

[a]HippDB includes standard search fields such as the PDB code, organism along with specific fields listed above. The fields are searchable and sortable.

distance and sequence and the organism of origin are recorded in the database.

The Web site interface uses original JavaScript for constructing queries, a standard AJAX protocol to execute the queries and a JQuery extension (DataTables) to format the query results (Table 1).

3 RESULTS

From 11 818 multiprotein entries in the PDB, 379 877 files of two protein chains were produced and subjected to alanine scanning. Of these interfaces, we found 7308 helices of four residues or longer with the two hotspots necessary to qualify for the database. A qualifying alpha helix is, on average, 13.2 residues long and contains 2.7 hotspots. The end-to-end distance separating these hotspots is 7.3 residues. On average, the three best hotspot residues sum to a ΔG of 3.9 in Rosetta energy units, and the helix overall contributes 48% of the chain's total ΔΔG and 37% of its ΔSASA.

The rational design of PPI inhibitors involves a systematic analysis of native interactions. By cataloging the results of this analysis for every known structure, and by describing second-order metrics to help prioritize design efforts, this database will eliminate often-reduplicated effort and greatly accelerate the design process. In this way, HippDB complements existing resources such as PocketQuery and HotSprint, the former of which catalogues regions of high solvent burial and the latter of which highlights evolutionary conservation to evaluate the role, functional or structural, of individual hotspots (Camacho and Koes, 2012; Guney *et al.*, 2008).

ACKNOWLEDGEMENTS

The authors thank Andrea Jochim for her work on the project that inspired HippDB.

Funding: This work was supported by the National Institutes of Health [R01 GM073943]. A.M.W. thanks the NYU Chemistry Department for a Kramer Fellowship. C.M.B. thanks the National Science Foundation for a Graduate Research Fellowship.

Conflict of Interest: none declared.

REFERENCES

Azzarito,V. *et al.* (2013) Inhibition of [alpha]-helix-mediated protein-protein interactions using designed molecules. *Nat. Chem.*, **5**, 161–173.
Baker,D. and Sali,A. (2001) Protein structure prediction and structural genomics. *Science*, **294**, 93–96.
Berman,H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
Boersma,M.D. *et al.* (2012) Evaluation of diverse alpha/beta-backbone patterns for fnctional alpha-helix mimicry: analogues of the Bim BH3 domain. *J. Am. Chem. Soc.*, **134**, 315–323.
Bullock,B.N. *et al.* (2011) Assessing helical protein interfaces for inhibitor design. *J. Am. Chem. Soc.*, **133**, 14220–14223.
Camacho,C.J. and Koes,D.R. (2012) PocketQuery: protein–protein interaction inhibitor starting points from protein–protein interaction structure. *Nucleic Acids Res.*, **40**, W387–W392.
Clackson,T. and Wells,J.A. (1995) A hot-spot of binding-energy in a hormone-receptor interface. *Science*, **267**, 383–386.
Fleishman,S.J. *et al.* (2011) RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One*, **6**, e20161.
Guney,E. *et al.* (2008) HotSprint: database of computational hot spots in protein interfaces. *Nucleic Acids Res.*, **36**, D662–D666.
Hajduk,P.J. and Greer,J. (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.*, **6**, 211–219.
Hubbard,S.J. and Thornton,J.M. (1993) NACCESS. (1993) *Department of Biochemistry and Molecular Biology*. University College, London.
Jochim,A.L. and Arora,P.S. (2010) Systematic analysis of helical protein interfaces reveals targets for synthetic inhibitors. *ACS Chem. Biol.*, **5**, 919–923.
Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure. *Biopolymers*, **22**, 2577–2637.
Koes,D.R. and Camacho,C.J. (2012) Small-molecule inhibitor starting points learned from protein-protein interaction inhibitor structure. *Bioinformatics*, **28**, 784–791.
Kortemme,T. and Baker,D. (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl Acad. Sci. USA*, **99**, 14116–14121.
Kortemme,T. *et al.* (2004) Computational alanine scanning of protein-protein interfaces. *Sci. STKE*, **2004**, pl2.
Moellering,R.E. *et al.* (2009) Direct inhibition of the NOTCH transcription factor complex. *Nature*, **462**, 182–188.
Patgiri,A. *et al.* (2011) An orthosteric inhibitor of the Ras-Sos interaction. *Nat. Chem. Biol.*, **7**, 585–587.
Raj,M. *et al.* (2013) Plucking the high hanging fruit: a systematic approach for targeting protein-protein interactions. *Bioorg. Med. Chem.*, **21**, 4051–4057.
Wells,J.A. and McClendon,C.L. (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*, **450**, 1001–1009.