Database, Vol. 2013, Article ID bat039, doi:10.1093/database/bat039

Database update

TIMBAL v2: update of a database holding small molecules modulating protein–protein interactions

Alicia P. Higueruelo*, Harry Jubb and Tom L. Blundell

Department of Biochemistry, University of Cambridge, Cambridge CB1 2GA, UK

***Corresponding author:** Tel: +44 1223 766001; Fax: +44 1223 766002; Email: alicia@cryst.bioc.cam.ac.uk

Citation details: Higueruelo,A.P., Jubb,H. and Blundell,T.L. TIMBAL v2: update of a database holding small molecules modulating protein–protein interactions. *Database* (2013) Vol. 2013: article ID bat039; doi:10.1093/database/bat039

Submitted 10 April 2013; Revised 7 May 2013; Accepted 13 May 2013

TIMBAL is a database holding molecules of molecular weight <1200 Daltons that modulate protein–protein interactions. Since its first release, the database has been extended to cover 50 known protein–protein interactions drug targets, including protein complexes that can be stabilized by small molecules with therapeutic effect. The resource contains 14 890 data points for 6896 distinct small molecules. UniProt codes and Protein Data Bank entries are also included.

Database URL: http://www-cryst.bioc.cam.ac.uk/timbal

Introduction

The idea of modulating protein–protein interactions (PPI) with small molecules has been intentionally pursued for more than a decade. The concept is attractive, but there are many challenges still ahead. In the UK, a network was recently created to bring the PPI scientists closer and facilitate collaboration to overcome the many hurdles (http://ppi-net.org). A contribution to these efforts has been to create TIMBAL, a resource that holds known small molecules modulating protein–protein complexes. The first release of the TIMBAL database in 2009 (1) included an analysis of 104 small molecules, 27 of which were structurally characterized with their targets in the Protein Data Bank (PDB) (2). A year later, Bourgeas *et al*. (3) released the 2P2I database, a hand-curated database of the structures of protein–protein complexes with known inhibitors. Several updates (4, 5) have refined the 2P2I to a structural database dedicated to orthosteric modulation of PPI containing 14 protein–protein complexes, 60 protein–inhibitor complexes, 16 free proteins and 55 small molecule modulators.

To our knowledge, there are no other resources for PPI modulators. The growth of data in the past years makes hand-curated databases a phenomenally time-consuming task. The maintenance of TIMBAL is achieved now through automated searches of the ChEMBL database (6) (currently using ChEMBL_15), and this report is a brief description of the update and its current contents.

Methods

ChEMBL database (6) (https://www.ebi.ac.uk/chembldb) holds bioactivity data for molecules manually extracted from a selection of peer-reviewed journals relevant to drug discovery. Chemical structures are checked and standardized to ensure consistency across the resource before deposition in the database. Assays are classified as 'binding' when there is direct interaction between the compound and the target, 'functional' when the interaction is indirect or against the whole organism or cell and 'ADMET' when there are pharmacokinetic data. Target assignment is checked by curators and a confidence score flagged. A further sub-classification depends on whether the assay is against an isolated *in vitro* target, a multi-protein complex (or nucleic acids), or not assigned because the assay is cell or tissue based. The database also contains a target dictionary

Database, Vol. 2013, Article ID bat039, doi:10.1093/database/bat039

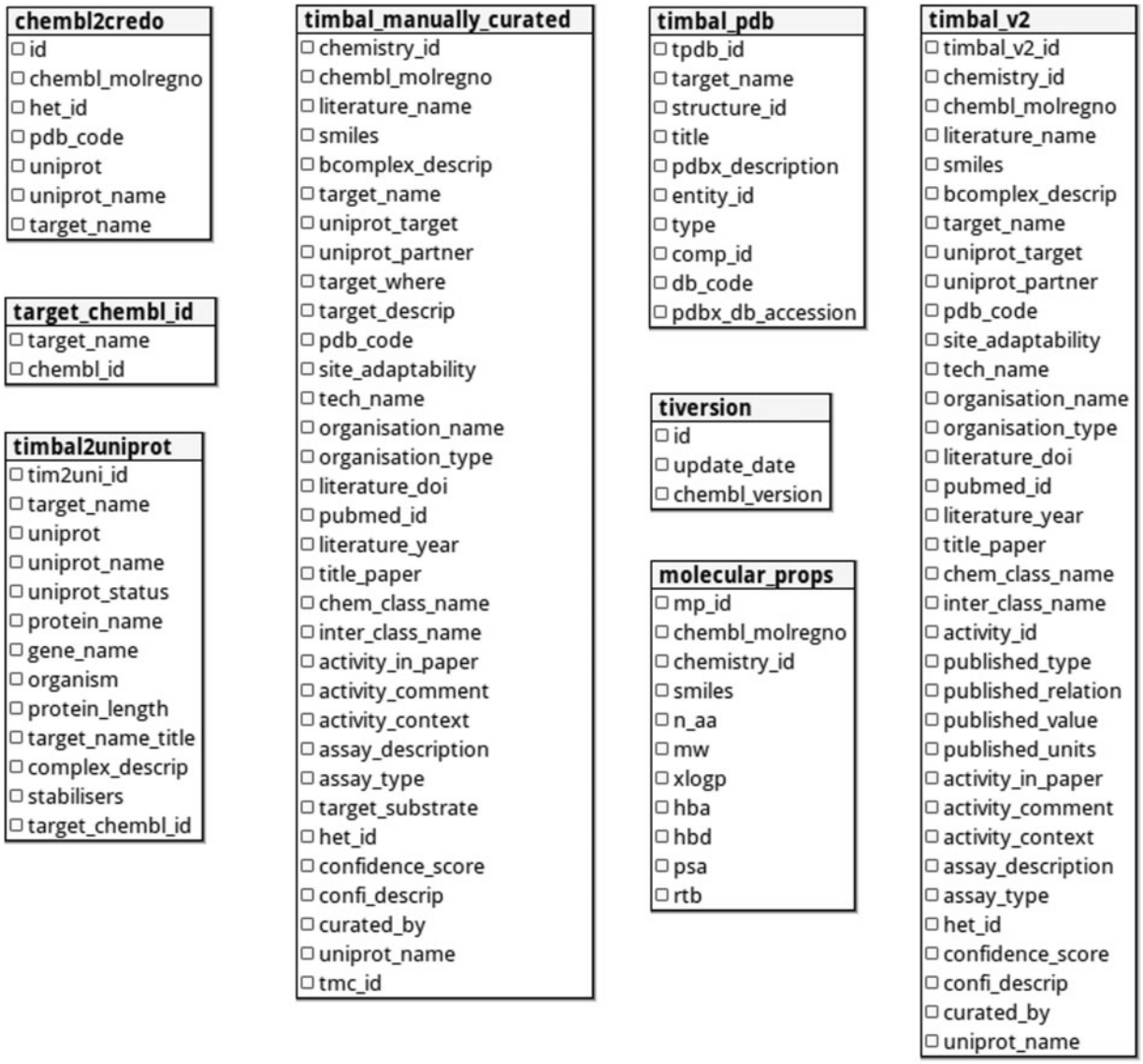

Figure 1. Schema of the database showing all tables and fields.

that allows users to browse target components by standard identifiers like UniProt accession code as well as NCBI taxonomy. In addition to a rich interactive web-based interface, ChEMBL is also conveniently downloadable in full in a variety of formats, which has allowed us to use a local copy to derive the TIMBAL update.

Target list

The initial list of 17 known PPI targets has been extended to 50 targets by PPI-Net members and TIMBAL users, and from conference talks and ChEMBL classification. For each target, we have generated a list of reviewed UniProt (7) codes for its orthologs. The codes are used in ChEMBL for searching small molecule data related to these proteins in binding assays where there is confidence that the assay is directly assigned to either a single protein or its homolog (e.g. binding affinity to Bcl-XL by isothermal titration calorimetric assay) or to a protein complex or its homologs (such as p53/MDM2 complex).

Automated update and manual curation

We maintain a small table for manually curated entries that are not available from ChEMBL, e.g. the newly described Mixed Lineage Leukemia (MLL) inhibitors (8) are reported in a journal not fully screened by the ChEMBL curators. A completely automated script updates the database merging the manually curated entries and the data extracted from the local copy of the ChEMBL. Searches against the

Table 1. Summary of the TIMBAL contents

Target name	Protein complex	N data points	N unique SM	N papers	N prot-sm PDB	N total PDB	N unique SM in v1
14-3-3[a]	14-3-3/PMA	3	3	2	3	8	
Adenylyl Cyclase[a]	Adenylyl Cyclase dimer C1-C2 domains	7 (2)	3 (1)	3	2	17	
Annexin A2	Annexin A2/S100-A10	164 (22)	54 (10)	1	0	9	
ARF1[a]	ARF1/SEC7	4	2	2	1	19	
AuxinIAA[a]	AuxinIAA-TIR1	1	1	1	1	8	
Bcl-XL and Bcl-2	Bcl-2 and Bcl-XL with BAX; BAK and BID	1256 (77)	645 (71)	65	16	78	26
Beta-catenin	BetaCatenin/Tcf4 and Tcf3	12 (7)	12 (7)	4	0	26	4
BIII	BIII/X11a	0	0	0	0	13	
BRD2	BRD2/Ack	93 (5)	44 (4)	7	12	21	
BRD4	BRD4/NUT	109 (2)	52 (2)	8	4	35	
BRDT	BRDT/H4	29 (2)	28 (2)	4	1	4	
CD154	CD40/CD154	1 (1)	1 (1)	1	0	8	
CD74	CD74/MIF	0	0	0	0	49	
CD80 (B7-1)	CD80/CD28 (or CTLA-4)	4	4	3	0	10	4
Clathrin	Clathrin/adaptor and accessory proteins	2	2	1	2	18	
c-Myc	c-Myc/Max	1	1	1	0	10	1
CRM1	CRM1/Rev	182 (144)	59 (51)	4	0	23	2
Cyclophilins	Cyclophilins	261 (37)	194 (33)	11	0	69	
E2	E1/E2	50 (1)	44 (1)	6	1	30	4
HIF-1a	HIF-1a/p300	274 (43)	182 (36)	20	0	12	
IL-2	IL-2/IL-2Ra	52 (2)	48 (2)	5	4	19	6
Immunophilin FKBP1A[a]	FKBP1A/FK506	571 (9)	540 (9)	30	10	44	
Integrins	Integrins	9730 (498)	3685 (307)	210	2	83	
K-Ras	K-Ras/SOS1	5	5	1	5	9	
Keap1	Nrf2/Keap1	0	0	0	0	31	
LMO2	LMO2/LDB1 or TAL1	0	0	0	0	5	
MDM2	p53/MDM2	320 (52)	236 (47)	23	8	34	16
MDMX	p53/MDMX	44 (16)	40 (16)	4	1	15	
Max[a]	Max dimer	0	0	0	0	8	
MLL	MLL/Menin	2	2	1	2	22	
Neuropilin-1	Neuropilin-1/VEGF-A	177 (11)	157 (11)	6	1	37	
PPAR-gamma	PPAR-gamma/NRCoA1	0	0	0	0	235	
Plk1(PBD)	Plk1(PBD)/PBD substrate	2	2	1	2	35	
Rac1	Rac1/GEFs	118 (11)	76 (11)	3	0	28	
Rad51	Rad51/BRCA2	34 (4)	10 (2)	2	8	33	
RGS4	RGS4/Galpha-o protein	1	1	1	0	3	1
RRTF1	RRTF1/CBFb	0	0	0	0	15	
S100B	S100B/p53	19	18	4	5	32	7
SOD1[a]	SOD1 dimer	28 (17)	16 (11)	5	2	109	
STAT3	STAT3 dimer	42 (7)	33 (6)	3	0	2	
STAT5	STAT5 dimer	19	5	2	0	1	
Sur-2	ESX/Sur-2 (DRIP130)	29 (8)	9 (4)	2	0	1	1

(Continued)

Database, Vol. 2013, Article ID bat039, doi:10.1093/database/bat039

Table 1. Continued.

Target name	Protein complex	N data points	N unique SM	N papers	N prot-sm PDB	N total PDB	N unique SM in v1
Tak1	Tak1/Tab1	1	1	1	0	7	
TNFa	TNFa trimer or TNFa/TNFR	8	7	3	1	13	2
Transthyretin[a]	Transthyretin tetramer	592 (71)	350 (69)	18	24	180	
ToxT	ToxT dimer	1	1	1	0	1	1
Tubulin[a]	Tubulin dimer	75 (36)	64 (36)	9	1	18	
UL42	UL30(Pol)/UL42 subunits of HSV type 1 DNA polymerase	4	4	1	0	1	3
XIAP	XIAP/Caspase9 or SMAC (BIR3 domanin)	538 (23)	312 (18)	30	8	38	5
ZipA	ZipA/FtsZ	24	23	6	4	8	21

N data points, number of data points for each target; N unique SM, number of distinct small molecules for each target; N papers, number of distinct publications per target; N prot-sm PDB, number of protein–small molecule complexes in the PDB for each target; N total PDB, number of PDB for each target, including protein–protein, protein–small molecule and apo protein structures; N unique SM in v1, For comparison, number of unique small molecules per target that were in previous version of the database.
Numbers in parentheses for data points and unique small molecules refer to inactive molecules.
[a]SM for the targets are stabilizers of PPI.

PDB bring the experimental structures for these targets, including protein–small molecule, protein–protein complexes and unbound proteins. Links to the CREDO database (9) allows the user to explore in detail the atomic interactions of these complexes. These links are matches to the chemical structure of the small molecule and the UniProt identifier of the appropriate target in the PDB entry.

The final step is a check of the contents of the database to ensure that the data reported are binding of small molecules to protein interfaces. Any discrepancy found is reported to the ChEMBL curators and removed from the TIMBAL database.

Thus, TIMBAL is no longer a manually curated database; there is a trade-off between automation and curation. Although every effort has been put in place to avoid noise in the data, it is clear that >9000 data points for the integrins cannot be fully curated. Researchers using TIMBAL are encouraged to report mistakes, comments or improvements.

Allosteric modulators that do not bind to interfacial residues have not been included, as their identification requires dedicated curation, and this is out of the scope of this update. Researchers interested in allosteric modulation are referred to AlloSteric Database (ASD) (10), a manually curated resource with announced updates every 6 months.

Owing to the characteristics of PPI targets, the small molecule term is a generic name to refer to synthetic molecules and small peptides that bind to these interfaces. For example, subnanomolar synthetic inhibitors for Bcl-2/Bcl-XL have been reported with molecular weight >1100 Daltons (11). The small peptides are also kept (up to 10 peptide bonds), as they might be useful for researchers as a tool compounds. In this way, TIMBAL molecules have molecular weight below 1200 Daltons and no more than 10 peptide bonds.

Web resource

Data extracted from ChEMBL and manually curated are stored into a PostgreSQL (http://www.postgresql.org) database. We use SQLAlchemy (http://www.sqlalchemy.org) to generate python objects from the database tables and Flask (http://flask.pocoo.org) to create web pages from these objects. User requests are handed on the fly using Flask generators and direct responses. Bootstrap (http://twitter.github.com/bootstrap) gives the Cascading Style Sheets framework and javascript functionality to create an efficient resource with minimal coding.

Results and Discussion

TIMBAL can be publicly accessed and downloaded at http://www-cryst.bioc.cam.ac.uk/timbal. The schema of the database is presented in Figure 1.

It contains >14 000 data points for ~7000 small molecules with 50 PPI targets. More than 9000 data entries are for integrins, the cell surface receptors that have been pursued as therapeutic targets for almost two decades (12).

Table 1 summarizes the contents of the database that also holds inactive molecules against PPI targets (7% of the total content), as ChEMBL stores all reported data, including non-active readings.

TIMBAL also holds small molecules that stabilize protein complexes with possible therapeutic effect (13), such as stabilizers of transthyretin oligomer that inhibit harmful amyloid fibril formation.

The resource will be updated with each new release of the ChEMBL database. Since its first release, TIMBAL has grown not only in terms of number of entries but also in terms of content, including now stabilizers and inactive molecules. It is our aim that this database helps in the quest of identifying small molecules binding to protein interfaces.

Acknowledgements

The authors thank Colin Groom for his encouragement to include stabilisers of protein complexes as well as their inhibitors; John Overington and Yvonne Light for their support with the ChEMBL database; Adrian Schreyer for maintenance of the local copy of ChEMBL; Will Pitt, Colin Groom, Adrian Schreyer, Bernardo Ochoa, Richard Bickerton, Marko Hyvonen, Mike Hann, Mark Searcey, Terry Rabbitts, Ian Hardcastle, Alexander Metz, John Karanicolas, Nikolay Todorov, David Bettinson, Hsin-Wei Wang, Alvaro Olivera-Nappa and Alessandro Contini for their contribution to the target list and testing the resource.

Funding

This work was supported by the University of Cambridge, BBSRC and UCB. Funding for open access charge: University of Cambridge.

Conflict of interest. None declared.

References

1. Higueruelo,A.P., Schreyer,A., Bickerton,G.R.J. *et al*. (2009) Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chem. Biol. Drug Des.*, **74**, 457–467.
2. Berman,H.M., Westbrook,J., Feng,Z. *et al*. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
3. Bourgeas,R.l., Basse,M.J., Morelli,X. *et al*. (2010) Atomic analysis of protein-protein interfaces with known inhibitors: The 2P2I database. *PLoS One*, **5**, e9598.
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. Basse,M.J., Betzi,S., Bourgeas,R. *et al*. (2013) 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res.*, **41**, D824–D827.
6. Gaulton,A., Bellis,L.J., Bento,A.P. *et al*. (2011) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.*, **40**, D1100–D1107.
7. The UniProt,C. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.*, **39**, D214–D219.
8. Shi,A., Murai,M.J., He,S. *et al*. (2012) Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia. *Blood*, **120**, 4461–4469.
9. Schreyer,A. and Blundell,T. (2009) CREDO: a protein-ligand interaction database for drug discovery. *Chem. Biol. Drug Des.*, **73**, 157–167.
10. Huang,Z., Zhu,L., Cao,Y. *et al*. (2011) ASD: a comprehensive database of allosteric proteins and modulators. *Nucleic Acids Res.*, **39**, D663–D669.
11. Zhou,H., Chen,J., Meagher,J.L. *et al*. (2012) Design of Bcl-2 and Bcl-xL inhibitors with subnanomolar binding affinities based upon a New Scaffold. *J. Med. Chem.*, **55**, 4664–4682.
12. Fry,D.C. (2006) Protein-protein interactions as targets for small molecule drug discovery. *Biopolymers*, **84**, 535–552.
13. Thiel,P., Kaiser,M. and Ottmann,C. (2012) Small-molecule stabilization of protein-protein interactions: an underestimated concept in drug discovery? *Angew. Chem. Int. Ed. Engl.*, **51**, 2012–2028.