Nucleic Acids Research Advance Access published October 20, 2014

doi: 10.1093/nar/gku1009

The human DEPhOsphorylation database DEPOD: a 2015 update

Guangyou Duan†, Xun Li† and Maja Köhn*

European Molecular Biology Laboratory, Genome Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Received August 15, 2014; Revised October 08, 2014; Accepted October 8, 2014

ABSTRACT

Phosphatases are crucial enzymes in health and disease, but the knowledge of their biological roles is still limited. Identifying substrates continues to be a great challenge. To support the research on phosphatase–kinase–substrate networks we present here an update on the human DEPhOsphorylation Database: DEPOD (http://www.depod.org or http://www.koehn.embl.de/depod). DEPOD is a manually curated open access database providing human phosphatases, their protein and non-protein substrates, dephosphorylation sites, pathway involvements and external links to kinases and small molecule modulators. All internal data are fully searchable including a BLAST application. Since the first release, more human phosphatases and substrates, their associated signaling pathways (also from new sources), and interacting proteins for all phosphatases and protein substrates have been added into DEPOD. The user interface has been further optimized; for example, the interactive human phosphatase–substrate network contains now a 'highlight node' function for phosphatases, which includes the visualization of neighbors in the network.

INTRODUCTION

Reversible phosphorylation plays important roles across a wide range of cellular processes. Therein phosphatases are one important protein class that antagonizes the roles of kinases to maintain the phosphorylation homeostasis by hydrolyzing the phosphoryl group from protein or non-protein compounds. Compared to the available data resources for kinases (1), there are still very few public data resources about phosphatases, especially regarding the phosphatase–substrate relations. So far several data resources for phosphatases have been developed, and most of them focus on specific aspects. For example, PTP-central provides a comprehensive resource of one of the 18 phosphatase families (2), that is the protein tyrosine phosphatase family (PTPs), in eukaryotic species based on a genomic scale PTP prediction tool termed Y-Phosphatomer (3). Another comprehensive phosphatase resource is the Eukaryotic protein Kinase and protein Phosphatase Database (EKPD) which is a hierarchical database of eukaryotic protein kinases and protein phosphatases (4). Both, the EKPD and PTP-central do not provide the phosphatase–substrate relationship information. The human phosphatase portal (HuPho) includes phosphatase–substrate relations by integrating diverse annotation information from several public web resources or scientific literature using text-mining techniques (5). This database is however only searchable for proteins, and does not contain an interactive network. Several other older phosphatase data resources like the PTP Database and PhosphoregDB are either out of service or not continually updated anymore. While all these phosphatase-containing databases are very useful, we developed DEPOD to provide a comprehensive, high quality, manually curated, fully searchable and interactive human phosphatase resource to the signaling research community (2). DEPOD is continuously updated and extended. In the following, the previously only briefly described construction process and functionalities of DEPOD will be explained, including new features introduced after the first release (2).

THE DEPOD RESOURCE

Construction of DEPOD

The construction of DEPOD is shown in Figure 1. Human phosphatases with enzymatic activities were retrieved from the *Homo sapiens* gene data set in the Ensembl database (release 75) (2,6). In addition to the active phosphatases, we have now also included inactive ones, that is the ones without measured enzymatic activities, to provide a comprehensive human phosphatase database. The Gene Ontology (GO) molecular function term (7) and phosphatase-related Pfam protein family entries (8) were used to screen candidate phosphatase-encoding genes as shown in Figure 1. The human phosphatase substrate information is obtained by the combination of the 'dephosphorylation' post-translational modification data in the Human Protein Ref-

*To whom correspondence should be addressed. Tel: +49 6221 387 8544; Fax: +49 6221 387 518; Email: koehn@embl.de
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

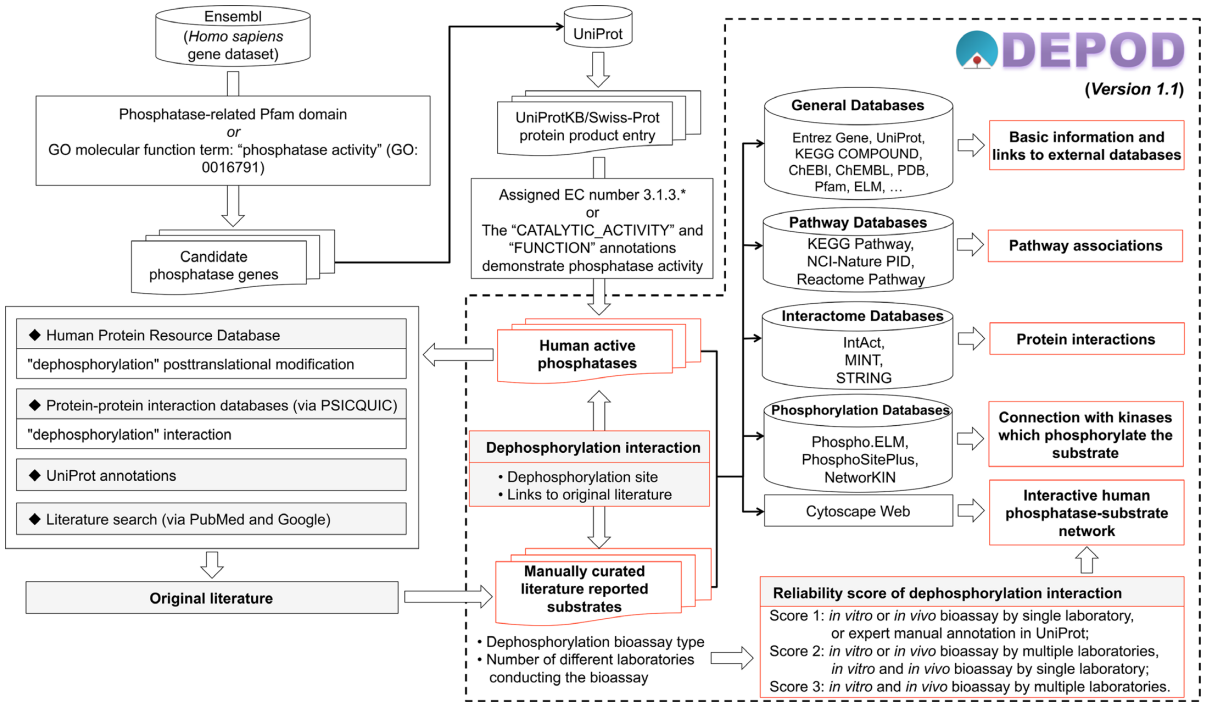

Figure 1. Flow chart of construction process of the DEPOD (version 1.1). The DEPOD core data are shown in red. EC, Enzyme Commission.

erence Database (9), 'dephosphorylation' interaction data retrieved from the PSICQUIC service (10), substrate information from UniProt annotation (11) and the scientific literature using PubMed and Google. After manual curation of literature reported substrates, the reliability score of the desphosphorylation interaction is assigned according to the dephosphorylation bioassay type and number of different laboratories conducting the corresponding bioassay (that is the number of unrelated corresponding authors of different publications, see also (2) and Figure 1). The corresponding literature to the phosphatase-substrate interaction is referenced via links to PubMed (www.ncbi.nlm.nih.gov/pubmed) and Europe PubMed Central (europepmc.org). The visualization of phosphatase–substrate relations is supported using Cytoscape Web (12). Furthermore, links to phosphorylation databases like Phospho.ELM (13), PhosphoSitePlus (14) and NetworKIN (15) provide the kinase information related to the substrates. Basic information and links to other databases integrated for phosphatases or substrates include but are not limited to HGNC (16), Entrez Gene (17), UniProt (11), KEGG COMPOUND (18), chEBI (19), chEMBL (20), PDB (21) and Pfam (8). Phosphatase or protein substrate related protein interaction motifs from ELM (22) are also added as a new feature for this release in the basic information page. The pathway information for phosphatases or substrates is extracted from different sources: KEGG Pathway (18), NCI-Nature PID Pathway (23) and, with this update, Reactome Pathway (24). Beyond the pathway information also as a new feature, the protein interaction information for phosphatase or substrate extracted from IntAct (25), MINT (26) and STRING (27) is provided as an integrated resource. The brief methodology of DEPOD can also be found by clicking on the 'About DEPOD' option on each page of DEPOD or on the bottom of the home page.

Statistics of DEPOD

Currently DEPOD contains 228 human active phosphatases of which 194 have 387 substrates including 298 protein substrates and 89 non-protein substrates, forming 1096 dephosphorylation interactions in total. We now added 11 inactive human phosphatases, thus in total there are 239 human phosphatases in the database. Furthermore, there are 213 pathways from the KEGG database (18), 206 pathways from the NCI-Nature Pathway Interaction Database (PID) (23) and 560 Reactome pathways (24) mapped onto phosphatases and substrates. As a new feature, there are 13984 protein interactions from IntAct, 6884 protein interactions from MINT and 6782 protein interactions from STRING including phosphatases or substrates.

Usage of DEPOD

Detailed documentation for the usage of DEPOD is provided by clicking on the 'User Manual' option on the left-hand navigation for each page of DEPOD or on the bottom of the home page. A brief usage introduction of DEPOD is shown below.

Search of DEPOD. Users of DEPOD can use the left-hand navigation on each page of DEPOD to browse the phosphatases ('Human Phosphatases'), protein substrates ('Protein Substrates'), non-protein substrates ('Non-Protein Substrates') and pathways ('Pathway Mapping') which provides an overview of all the related data in DEPOD. Another straightforward way is to use the keyword-

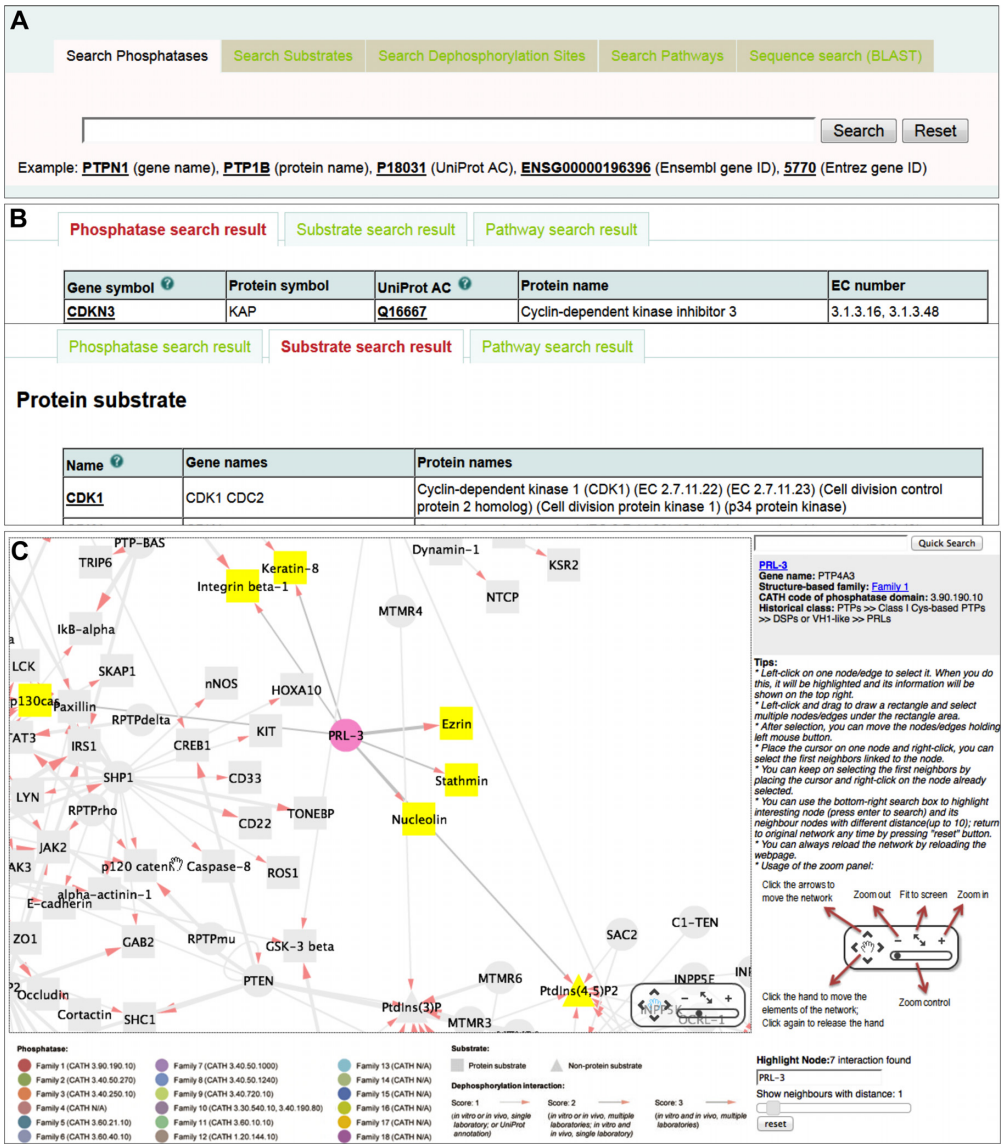

Figure 2. Usage screenshots of DEPOD. (**A**) Search of phosphatases/substrates/dephosphorylation site/pathways and sequence (BLAST); (**B**) Quick search of DEPOD; (**C**) Human phosphatase–substrate network.

based searches on phosphatases, substrates, pathways and, previously not available (2), dephosphorylation sites as shown in Figure 2A. Users can also search DEPOD using the quick search option on the top right corner of each page ('Quick Search'), which gives the results in the categories 'phosphatases', 'substrates' and 'pathways' as tabs (Figure 2B). A sequence homology search functionality is provided that users can use to search homologous phosphatases or protein substrates in DEPOD using their own interesting sequences from any species.

Visualization of the dephosphorylation network. As an important data component of DEPOD, manually curated phosphatase–substrate relations can be explored using the interactive visualization functionality based on Cytoscape Web (10) (with the flash installed in the web browser). Phosphatases and their substrates are connected via signal arrow links, for which the widths indicate the reliability scores of the corresponding interactions. The details of each node will be shown on the top right area by clicking on the corresponding node and its first neighbor nodes can be selected by right click on the node. Users can search their phosphatase or substrate of interest using the search box on the bottom right corner ('Highlight Node') and select any neighboring nodes with different path lengths as shown in Figure 2C.

Download of DEPOD data. On the download page, human phosphatases and related information included in DEPOD are downloadable in xls format for interested users. The human phosphatase-substrate interaction and dephosphorylation sites are now downloadable in TXT and PSI-MI Tab 2.5 formats (28). Beyond this, pathways from KEGG, NCI Nature PID and Reactome pathways mapped on phosphatases and their substrates are available for download.

Feedback and contact. As a human phosphatases data portal, users can also deposit substrate candidates of human active phosphatases to DEPOD using 'Deposit Data' on the left-hand navigation (see also the 'About DEPOD' page in the database). More general suggestions or comments for DEPOD can be sent to us by the 'Feedback' page that can be found on the left-hand navigation, or by sending us an email by clicking on the 'Contact' option on the left-hand navigation on each page. The 'Feedback' or 'Contact' option can also be accessed on the bottom of the home page.

Update on the structure-sequence-based phosphatase classification

Since we have now included also the inactive human phosphatases, we performed another re-classification of human phosphatases (for methodological details see (2)) that includes the inactive phosphatases as well as few more (potentially) active phosphatases that were previously not included. These were found by manual comparison of DEPOD data with Ensembl (STS-1, NT5DC1, NT5DC2, NT5DC4), HuPho (CDC14C) and EKPD (STS-1) data. Supplementary Figure S1 shows the phosphatase families where the new entries have been added. Of these newly added phosphatases, STS-1 represents a very interesting entry due to its novel classification as family 8 member. It was previously not classified into a historical phosphatase class (2), and thus this reveals a new family relationship. It is the only phosphatase in family 8 with tyrosine phosphatase activity.

CONCLUSION AND FUTURE PERSPECTIVES

DEPOD provides a web interface to access manually curated data on human active phosphatases and their substrates. These data can be used to develop tools for predicting dephosphorylation sites for phosphatases or to aid in substrate-based drug design (29). Pathway mapping provides clues to the potential involvement of phosphatases and substrates in novel pathways. External links to kinases conducting phosphorylation makes it possible to connect phosphatases and kinases through their common substrates, and could help to understand which pathways and phosphatases could be effected by kinase inhibition in drug discovery. The online interactive visualization tool enables a clear overview of known phosphatase interaction, and can be used to connect the phosphatase or substrate of interest with their nearest or further neighbors. DEPOD has become a valuable resource for the phosphorylation signaling community (30) and will continue to be an essential resource for future phospho-signaling studies.

DEPOD will be continuously updated according to the research progress of the phosphorylation signaling community by adding new human phosphatases, substrates of phosphatases, dephosphorylation sites, further protein–protein interaction data (31) and so on. With the advancement in our knowledge of the substrate specificity for phosphatases, more related information is planned to be added into DEPOD such as physicochemical properties (32) and protein interaction motifs (22).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

The authors would like to thank the users of DEPOD for their support as well as colleagues and users for their feedback to the DEPOD database.

FUNDING

X.L. thanks EMBL and Marie Curie Actions for the EMBL Interdisciplinary Postdoc (EIPOD) fellowship. M.K. thanks the German Science Foundation (DFG) for support within the Emmy-Noether program (KO 4013/1-1). Funding for open access charge: EMBL.
Conflict of interest statement. None declared.

REFERENCES

1. Xue,Y., Gao,X., Cao,J., Liu,Z., Jin,C., Wen,L., Yao,X. and Ren,J. (2010) A summary of computational resources for protein phosphorylation. *Curr. Protein Pept. Sci.*, **11**, 485–496.
2. Li,X., Wilmanns,M., Thornton,J. and Köhn,M. (2013) Elucidating human phosphatase–substrate networks. *Sci. Signal.*, **6**, rs10.
3. Hatzihristidis,T., Liu,S., Pryszcz,L., Hutchins,A.P., Gabaldón,T., Tremblay,M.L. and Miranda-Saavedra,D. (2014) PTP-central: a comprehensive resource of protein tyrosine phosphatases in eukaryotic genomes. *Methods*, **65**, 156–164.
4. Wang,Y., Liu,Z., Cheng,H., Gao,T., Pan,Z., Yang,Q., Guo,A. and Xue,Y. (2013) EKPD: a hierarchical database of eukaryotic protein kinases and protein phosphatases. *Nucleic Acids Res.*, **42**, D496–D502.
5. Liberti,S., Sacco,F., Calderone,A., Perfetto,L., Iannuccelli,M., Panni,S., Santonico,E., Palma,A., Nardozza,A.P., Castagnoli,L. *et al.* (2013) HuPho: the human phosphatase portal. *FEBS J.*, **280**, 379–387.
6. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S. *et al.* (2014) Ensembl 2014. *Nucleic Acids Res.*, **42**, D749–D755.
7. The Gene Ontology Consortium. (2012) Gene ontology annotations and resources. *Nucleic Acids Res.*, **41**, D530–D535.
8. Finn,R.D., Bateman,A., Clements,J., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Heger,A., Hetherington,K., Holm,L., Mistry,J. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.*, **42**, D222–D230.
9. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human protein reference database—2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
10. Aranda,B., Blankenburg,H., Kerrien,S., Brinkman,F.S.L., Ceol,A., Chautard,E., Dana,J.M., De Las Rivas,J., Dumousseau,M., Galeota,E. *et al.* (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods*, **8**, 528–529.
11. The UniProt Consortium. (2014) Activities at the universal protein resource (UniProt). *Nucleic Acids Res.*, **42**, D191–D198.
12. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
13. Dinkel,H., Chica,C., Via,A., Gould,C.M., Jensen,L.J., Gibson,T.J. and Diella,F. (2011) Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.*, **39**, D261–D267.
14. Hornbeck,P.V, Kornhauser,J.M., Tkachev,S., Zhang,B., Skrzypek,E., Murray,B., Latham,V. and Sullivan,M. (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.*, **40**, D261–D270.
15. Linding,R., Jensen,L.J., Pasculescu,A., Olhovsky,M., Colwill,K., Bork,P., Yaffe,M.B. and Pawson,T. (2008) NetworKIN: a resource for

exploring cellular phosphorylation networks. *Nucleic Acids Res.*, **36**, D695–D699.
16. Gray,K.A., Daugherty,L.C., Gordon,S.M., Seal,R.L., Wright,M.W. and Bruford,E.A. (2013) Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res.*, **41**, D545–D552.
17. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2011) Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **39**, D52–D57.
18. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.
19. Hastings,J., de Matos,P., Dekker,A., Ennis,M., Harsha,B., Kale,N., Muthukrishnan,V., Owen,G., Turner,S., Williams,M. *et al.* (2012) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.*, **41**, D456–D463.
20. Gaulton,A., Bellis,L.J., Bento,A.P., Chambers,J., Davies,M., Hersey,A., Light,Y., McGlinchey,S., Michalovich,D., Al-Lazikani,B. *et al.* (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.*, **40**, D1100–D1107.
21. Berman,H.M., Kleywegt,G.J., Nakamura,H. and Markley,J.L. (2012) The protein data bank at 40: reflecting on the past to prepare for the future. *Structure*, **20**, 391–396.
22. Dinkel,H., Van Roey,K., Michael,S., Davey,N.E., Weatheritt,R.J., Born,D., Speck,T., Krüger,D., Grebnev,G., Kuban,M. *et al.* (2014) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.*, **42**, D259–D266.
23. Schaefer,C.F., Anthony,K., Krupa,S., Buchoff,J., Day,M., Hannay,T. and Buetow,K.H. (2009) PID: the pathway interaction database. *Nucleic Acids Res.*, **37**, D674–D679.
24. Croft,D., Mundo,A.F., Haw,R., Milacic,M., Weiser,J., Wu,G., Caudy,M., Garapati,P., Gillespie,M., Kamdar,M.R. *et al.* (2014) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **42**, D472–D477.
25. Kerrien,S., Aranda,B., Breuza,L., Bridge,A., Broackes-Carter,F., Chen,C., Duesbury,M., Dumousseau,M., Feuermann,M., Hinz,U. *et al.* (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.*, **40**, D841–D846.
26. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardozza,A.P., Santonico,E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
27. Franceschini,A., Szklarczyk,D., Frankild,S., Kuhn,M., Simonovic,M., Roth,A., Lin,J., Minguez,P., Bork,P., von Mering,C. *et al.* (2013) STRING v9.1: protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, **41**, D808–D815.
28. Van Roey,K., Orchard,S., Kerrien,S., Dumousseau,M., Ricard-Blum,S., Hermjakob,H. and Gibson,T.J. (2013) Capturing cooperative interactions with the PSI-MI format. *Database*, **2013**, bat066.
29. Martin,K.R., Narang,P., Medina-Franco,J.L., Meurice,N. and MacKeigan,J.P. (2014) Integrating virtual and biochemical screening for protein tyrosine phosphatase inhibitor discovery. *Methods*, **65**, 219–228.
30. Goldman,A., Roy,J., Bodenmiller,B., Wanka,S., Landry,C.R., Aebersold,R. and Cyert,M.S. (2014) The calcineurin signaling network evolves via conserved kinase-phosphatase modules that transcend substrate identity. *Mol. Cell*, **55**, 422–435.
31. The International Molecular Exchange Consortium. (2012) Protein interaction data curation: the international molecular exchange (IMEx) consortium. *Nat. Methods*, **9**, 345–350.
32. Selner,N.G., Luechapanichkul,R., Chen,X., Neel,B.G., Zhang,Z.-Y., Knapp,S., Bell,C.E. and Pei,D. (2014) Diverse levels of sequence selectivity and catalytic efficiency of protein-tyrosine phosphatases. *Biochemistry*, **53**, 397–412.