

# PepCyber:P~PEP: a database of human protein–protein interactions mediated by phosphoprotein-binding domains

Wuming Gong<sup>1</sup>, Dihan Zhou<sup>1</sup>, Yongliang Ren<sup>1</sup>, Yejun Wang<sup>1</sup>, Zhixiang Zuo<sup>1</sup>, Yanping Shen<sup>1</sup>, Feifei Xiao<sup>1</sup>, Qi Zhu<sup>2</sup>, Ailing Hong<sup>3</sup>, Xiaochuan Zhou<sup>4</sup>, Xiaolian Gao<sup>2</sup> and Tongbin Li<sup>1,\*</sup>

<sup>1</sup>Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455, <sup>2</sup>Department of Biology and Biochemistry, University of Houston, Houston, TX 77004, <sup>3</sup>Atactic Technologies and <sup>4</sup>LC Sciences, Houston, TX 77054, USA

Received August 16, 2007; Revised September 25, 2007; Accepted September 26, 2007

## ABSTRACT

**Phosphoprotein-binding domains (PPBDs) mediate many important cellular and molecular processes. Ten PPBDs have been known to exist in the human proteome, namely, 14-3-3, BRCT, C2, FHA, MH2, PBD, PTB, SH2, WD-40 and WW. PepCyber:P~PEP is a newly constructed database specialized in documenting human PPBD-containing proteins and PPBD-mediated interactions. Our motivation is to provide the research community with a rich information source emphasizing the reported, experimentally validated data for specific PPBD–PPEP interactions. This information is not only useful for designing, comparing and validating the relevant experiments, but it also serves as a knowledge-base for computationally constructing systems signaling pathways and networks. PepCyber:P~PEP is accessible through the URL, <http://www.pepcyber.org/PPEP/>. The current release of the database contains 7044 PPBD-mediated interactions involving 337 PPBD-containing proteins and 1123 substrate proteins.**

## INTRODUCTION

Protein phosphorylation-mediated signal transduction is an important post-translational modification (PTM)-based regulatory mechanism, and is implicated in a broad spectrum of key cellular molecular processes, including cell cycle, oncogenic transformation,

immunological responsiveness, apoptosis and development (1–5). In these activities, ‘phosphoprotein-binding domains’ (PPBDs, denoting domains that have specific binding affinity to phosphorylated sites in proteins) play a pivotal role in connecting the kinases and the effector molecules, forming multi-protein complexes, and inducing specific protein–protein interactions responsible for changes in these proteins’ subcellular localization, folding state, binding specificity or activity (6). PPBDs achieve their binding specificity to their substrate proteins primarily through the recognition of a phosphopeptide (PPEP) region, which are short peptide sequences (~6–15 residues) containing phosphorylated residues (i.e. pS, pT or pY, where S is serine, T is threonine and Y is tyrosine) (1,7). Other factors, such as the tertiary structures, subcellular localization of the substrate proteins, as well as domain competition, are also known to influence PPBD–phosphoprotein interactions *in vivo* (7,8). Phosphorylation sites are frequently found in intrinsically disordered or unstructured regions of the proteins (9–11), making these regions good candidate sites for PPBD binding. In the human proteome, 10 protein domains—14-3-3, BRCT, C2, FHA, MH2, PBD, PTB, SH2, WD-40 and WW—have been identified as PPBDs, i.e. they possess phosphoprotein or PPEP-binding activities (Table 1).

The interactions between PPBDs and their PPEP substrates have been studied extensively using a variety of techniques, including structural determination (using X-ray crystallography or NMR spectroscopy), peptide assay (using phage display, synthetic peptide library or oriented peptide library), combinatorial screening, mass spectrometry analysis, mutagenesis (usually followed

\*To whom correspondence should be addressed. Tel: 612 626 3481; Fax: 612 626 3257; Email: [tolib@biocompute.umn.edu](mailto:tolib@biocompute.umn.edu)  
Correspondence may also be addressed to Xiaolian Gao. Tel: 713 743 2805; Fax: 713 743 2636; Email: [xgao@uh.edu](mailto:xgao@uh.edu)

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

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Table 1.** Summary of the 10 human PPBD classes

PPBD (class name)	Reported substrate specificity <sup>a</sup>	Refs.
14-3-3	R-S-X-pS-X-P, R-X-X-X-pS-X-P	(1,20)
BRCT	pS-X-X-F	(21,22)
C2	(Y/F)-(S/A)-(V/I)-pY-(Q/R)-X-(Y/F)-X	(23)
FHA (Forkhead-associated)	pT-X-X-D	(24,25)
MH2 (MAD homology 2)	pS-X-pS	(26)
PBD (Polo-Box domain)	S-(pT/pS)-(P/X)	(27)
PTB (phosphotyrosine binding)	N-P-X-pY, N-P-X-Y	(28)
SH2 (Src homology 2)	pY-X-X-(I/P) (CRK); pY-(I/V/L)-X-(I/V/L) (PTPN11); pY-(M/I/L/V/E)-X-M (PIK3R1)	(29,30) <sup>b</sup>
WD-40	D-pS-G-Φ-X-pS (BTRC); (I/L/P)-(I/L/P)-pT-P (FBXW7)	(31,32)
WW	(A/P)-P-P-(A/P)-Y; pS-P; pT-P	(33)

<sup>a</sup>The listed information is representative of the vast amount of literature information.

<sup>b</sup>The information also includes our unpublished data on SH2-PPEP microarray data.

by GST pull-down or yeast two-hybrid assays) and computational sequence analysis (3). The rich information generated by these means is now partially captured by a few database resources, where the information about PPBDs, PPBD-containing proteins and their interactions with PPEP-containing substrate proteins can be obtained. These resources include general protein–protein interaction databases such as BIND (12), HPRD (13) and DOMINO (14), functional motif databases such as ELM (8) and Phospho.ELM (15), and a specialized prediction server—Scansite, which makes predictions about the PPBD–PPEP interactions based on results obtained from oriented peptide library experiments (16). Despite the availability of these existing resources, a database that offers integrated, comprehensive, detailed annotations regarding the proteomic interactions mediated by PPBDs is still lacking.

PepCyber:P~PEP was constructed with the intention of filling this gap. In PepCyber:P~PEP, the information about PPBD-mediated protein interactions was carefully compiled through curation of peer-reviewed publications, and deposited into a relational database. For each interaction, specific information about the PPBD, PPBD-containing protein, the specific PPEP substrate bound, the substrate protein, the evidence of the interaction and the citations were recorded. Moreover, information regarding the signaling pathways associated with the PPBD–PPEP-binding interactions (in particular, tumorigenesis-related signaling pathways and tumor types) is also documented and stored in the database. The data hosted in PepCyber:P~PEP meet the Human Proteome Organization (HUPO), Proteomics Standards Initiative (PSI) standard (17), which is supported by most major protein–protein interaction databases.

## UTILITY

### Data content

We term an occurrence of a PPBD in a specific protein a ‘PPBD instance’, and the collection of similar PPBD instances with a high level of sequence homology and structural similarity a ‘PPBD class’. For example, the SH2

domain located close to the N-terminus of the protein, PTPN11 (SHP2), is an ‘instance’ belonging to the SH2 ‘PPBD class’. Presently, there are 10 reported human PPBD classes (Table 1). The current PepCyber:P~PEP release (V.1.0, release 31 July 2007) includes 7044 PPBD-mediated interactions involving 337 PPBD-containing proteins and 1123 substrate proteins. This rich information was obtained through the curation of 2446 peer-reviewed research articles published between 1975 and 2007. The largest number of interactions involves the SH2 PPBD class (4290 interactions) that is followed by the WW PPBD class (1389 interactions) and 14-3-3 PPBD class (1086 interactions). These interactions were classified into three categories based on whether the interaction was known to be mediated by a concerned PPBD instance, and whether the substrate peptide had been identified: if the interaction was known to be mediated by a PPBD instance and the substrate peptide had been identified, the interaction was classified as ‘category A’; if the interaction was known to be mediated by a PPBD instance but the substrate peptide had not been identified, the interaction is classified as ‘category B’; if it was not known whether the interaction was mediated by a PPBD instance (though one of the interacting proteins was a PPBD protein), then the interaction was classified as ‘category C’. Among the 7044 interactions documented in the current release of PepCyber:P~PEP, 5376 (76%) are category A interactions.

The 14-3-3, PTB and WW PPBD classes are unique in that they are also capable of binding to non-PPEP substrates in certain instances. These non-PPEP interactions are also documented in PepCyber:P~PEP. There are a total of 1068 non-PPEP interactions, accounting for 15% of the total collection of the current PepCyber:P~PEP release. All non-PPEP interactions documented are category A interactions.

### Web interface

The web interface of the PepCyber:P~PEP database can be accessed through the URL <http://www.pepcyber.org/PPEP/>. Five tabs are located underneath the logo of

the web site, namely, 'PPBD Classes', 'PPBD Proteins', 'Interaction Search', 'Tutorial' and 'Glossary'. These five tabs are described subsequently.

The 'PPBD Classes' tab leads to the introduction pages of the 10 PPBD classes (Table 1), where information about the lengths, structures, representative instances and the reported binding specificity for each of the 10 PPBD classes is presented.

The 'PPBD Proteins' tab leads to the PPBD-containing proteins browsing page, where the user can select a PPBD-containing protein to view the details for the protein of interest, including the gene symbol, description, NCBI RefSeq and Swiss-Prot accessions and a graphical representation of all PPBD-mediated interactions involving this protein. The information regarding the interactions involving each PPBD instance of the protein is then listed separately. When the number of the known interactions involving a domain is sufficiently large ( $\geq 10$ ), the positional amino acid composition preference of the substrate sequences (17 amino acid long) is also available as a WebLogo image (18).

The 'Interaction Search' tab leads to the page where custom search functions can be executed. The user has multiple search options for the PPBD-PPEP interactions: by the PPBD class; by the PPBD instance; by the name, NCBI RefSeq accession or Swiss-Prot accession of either the PPBD protein or the substrate protein; by the substrate peptide sequence, or by the pathway involved. The search can also be conducted using any combinations of the above criteria. The search result is presented as a list of interactions, each with the names of the PPBD-containing protein and the substrate protein, the substrate sequence, index site, evidence type, category of the interaction and the number of records matching the interaction. The 'index site' refers to the locus of the phosphorylated residue on a PPEP substrate protein, or the locus of the central contact residue on a non-PPEP substrate protein. The 'evidence type' indicates the type of analysis conducted to support the presence of the interaction. Currently, four evidence types are defined: (i) structural determinations; (ii) peptide library experiments; (iii) mutagenesis and (iv) sequence analyses. By clicking the 'Details' link for each listed interaction, the user can view more detailed information about the interaction, including the names, NCBI RefSeq accessions, Swiss-Prot accessions of the PPBD-containing protein and the substrate protein, the sequence of the substrate protein, the evidence type and the references for the interactions reported. The user can choose to plot all interactions or a selected set of interactions as PPBD-mediated protein-protein interaction networks. The network graphs are rendered dynamically using the graph visualization software, GraphViz (19). In the network plot, each node represents one protein: PPBD-containing proteins are labeled in green, and other proteins are labeled in yellow. Each directed edge represents an interaction between a PPBD instance and substrate protein, with the index site displayed on the edge. The user can click on a node representing a PPBD-containing protein to reach the PPBD protein information page.

The 'Tutorial' tab leads to the page where the utility of the database is demonstrated in a graphical manner. The 'Glossary' tab leads to the glossary page where terms and abbreviations used in the web site are explained.

#### Data access

PepCyber:P~PEP is publicly accessible through the URL <http://www.pepcyber.org/PPEP/> and the data sets are available, free of charge, to researchers from academic and non-profit institutions. Additional requests can be made by emailing to [help@pepcyber.org](mailto:help@pepcyber.org).

#### Implementation

The PepCyber:P~PEP database is a relational database implemented with MySQL on a Fedora Core 2 Linux system. The front-end web interface is implemented as a PHP project running under Apache 2.0.

#### COMPARISON WITH DATABASES RELEVANT TO PPBPS AND/OR PPEPS

PepCyber:P~PEP is the first database specialized in documenting human PPBDs, PPBD-containing proteins and PPBD-mediated protein-protein interactions. However, the information about human PPBDs and PPBD-mediated interactions is also hosted in existing general protein-protein interaction databases such as BIND (12), HPRD (13) and DOMINO (14), and functional motif databases such as ELM (8) and Phospho.ELM (15). All these databases have different focuses, and as such the types of information stored vary among them (Table 2).

PepCyber:P~PEP hosts a substantially richer collection of data about PPBD-mediated interactions than any other database. Table 3 provides a quantitative comparison between PepCyber:P~PEP and the existing databases in the number of PPBD instances and PPBD-mediated interactions. In addition to this advantageous depth and breadth of information, the PepCyber:P~PEP data collection is also of notably high quality, attributed to the meticulous data curation procedure followed and the rigorous quality control (QC) process carried out before the data is deposited into the MySQL database. For example, special attention was given to allow synonymous protein symbols used in the search, allowing a user to obtain consistent results, as different original studies may use different symbols to represent the same gene or protein. During data curation, each gene/protein symbol used in the original articles was checked against three databases—NCBI GenBank, Swiss-Prot and NCBI Entrez Gene to ensure that different symbols (or synonyms) of the same gene/protein are represented by the same entity in the data set. During QC, the curated entries were checked against a local copy of the three public gene/protein databases. If any inconsistency was identified, the entry was returned to the curation process for re-checking. These procedures guarantee that only high-confidence data were deposited into the released PepCyber:P~PEP database. Problems with gene/protein

**Table 2.** A comparison between PepCyber:P~PEP and similar databases in the type of information provided for PPBD-mediated protein-protein interactions<sup>a</sup>

Database	URL	Institution/company	PPBD class	PPBD protein	PPBD-mediated networks	Substrate protein	Substrate sequence	Binding pattern	Index site	Evidence	Pathway
BIND	<a href="http://bind.ca">http://bind.ca</a>	Unleashed Informatics	No	Yes	No	Yes	Yes	No	No	Yes	No
HPRD	<a href="http://www.hprd.org/">http://www.hprd.org/</a>	Institute of Bioinformatics, Bangalore, India and Johns Hopkins University	Yes	Yes	Yes	No	No	Yes	No	No	No
DOMINO	<a href="http://mint.bio.uniroma2.it/domino/">http://mint.bio.uniroma2.it/domino/</a>	University of Rome Tor Vergata, Italy	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
ELM	<a href="http://elm.eu.org/">http://elm.eu.org/</a>	Eukaryotic Linear Motif (ELM) consortium	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes
Phospho.ELM	<a href="http://phospho.elm.eu.org/">http://phospho.elm.eu.org/</a>	Eukaryotic Linear Motif (ELM) consortium	Yes	No	No	Yes	Yes	No	Yes	No	No
<b>PepCyber:P~PEP</b>	<b><a href="http://www.pepcyber.org/PPEP/">http://www.pepcyber.org/PPEP/</a></b>	<b>University of Minnesota, University of Houston, LC Sciences</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>

<sup>a</sup>The comparison was made based only on PPBDs and PPBD-mediated interactions. BIND, HPRD and DOMINO are general protein-protein interaction databases and host information about interactions of much broader scopes than PepCyber:P~PEP. ELM is a general functional motif database and has much broader scope of coverage than the PPBD substrates that PepCyber:P~PEP focuses on. Phospho.ELM primarily focuses on documenting information related to phosphorylation sites. Bold indicates that PepCyber:P~PEP is the database we are presenting, and that it is superior to other databases in comparison.

symbols that occur from time to time in other databases were minimized. As an example of such problems, three symbols—SHP2, SHP-2 and SHPTP2—were used in different entries in Phospho.ELM, without indication that they are all synonyms of the same gene PTPN11.

## FUTURE DIRECTIONS

PepCyber:P~PEP is intended to provide comprehensive, up-to-date, dynamic information and tools to researchers who require information on PPBD-mediated protein-protein interactions as well as the sequence patterns and connecting maps of these interactions in the human proteome. The reported information herein represents an initial step toward our long-term goals. Our continuing effort will be in the following several areas: (i) Update and expand the content and functions of PepCyber:P~PEP as published studies on the PPBDs and PPBD-mediated interactions continue to accumulate; (ii) develop and implement novel analysis methods of proteins and peptides to mine the rich data compilation stored in PepCyber:P~PEP; (iii) develop strategies and methods to predict substrate specificity for one or more PPBD instances within a PPBD class and (iv) develop necessary tools for systems biology modeling using the PepCyber:P~PEP data. These developments will complement experimental efforts, lead to savings in time and cost in experiments, and accelerate our understanding of the key processes in cellular regulation mechanisms. We envision that such an information source will have significant value for not only proteomic research, but also for discovery and development of drug candidates, drug targets and biomarkers.

PepCyber:P~PEP is merely the first significant component of the overall PepCyber, a valuable information source for important peptide-related biological and biomedical subject areas. The PepCyber effort will eventually result in a suite of database resources and computational tools assisting the development of peptide microarray-based proteomics profiling analysis. Future developments of PepCyber will include database resources with expanded scopes, e.g. non-PPBD-mediated protein-protein interactions, as well as non-database components such as peptide microarray design and data analysis tools.

## ACKNOWLEDGEMENTS

We thank the Supercomputing Institute, University of Minnesota for computational resources and Dr K. Cassidy for reviewing the database and the manuscript. We also acknowledge the support of NIH (1R21CA126209 to X.G. and T.L.), Minnesota Medical Foundation (to T.L.), NIH/GM/AI (R43 GM076941 to X.G.) and the R. A. Welch Foundation (E-1270 to X.G.). Funding to pay the Open Access publication charges for this article was provided by NIH/NCI.

*Conflict of interest statement.* None declared.

**Table 3.** A quantitative comparison between PepCyber:P~PEP and similar databases in the numbers of PPBD instances, PPBD-mediated interactions for all PPBD classes as well as for the four most popular PPBD classes: 14-3-3, PTB, SH2 and WW

Database	#PPBD instances	#PPBD-mediated interactions	#14-3-3-mediated interactions	#PTB-mediated interactions	#SH2-mediated interactions	#WW-mediated interactions
BIND	81	149	4	19	82	4
HPRD	57	105	7	7	82	1
DOMINO	195	1020	234	85	468	157
ELM	14	68	20	0	47	1
Phospho.ELM	47	220	0	11	209	0
<b>PepCyber:P~PEP</b>	<b>337</b>	<b>7044</b>	<b>1086</b>	<b>158</b>	<b>4290</b>	<b>1389</b>

Bold indicates that PepCyber:P~PEP is the database we are presenting, and that it is superior to other databases in comparison.

## REFERENCES

1. Yaffe, M.B. (2002) Phosphotyrosine-binding domains in signal transduction. *Nat. Rev. Mol. Cell Biol.*, **3**, 177–186.
2. Zhou, M.M. (2000) Phosphothreonine recognition comes into focus. *Nat. Struct. Biol.*, **7**, 1085–1087.
3. Yaffe, M.B., Leparo, G.G., Lai, J., Obata, T., Volinia, S. and Cantley, L.C. (2001) A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat. Biotechnol.*, **19**, 348–353.
4. Pawson, T. and Scott, J.D. (2005) Protein phosphorylation in signaling—50 years and counting. *Trends Biochem. Sci.*, **30**, 286–290.
5. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912–1934.
6. Yaffe, M.B. (2004) Master of all things phosphorylated. *Biochem. J.*, **379**, e1–e2.
7. Pawson, T. and Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science*, **300**, 445–452.
8. Puntervoll, P., Linding, R., Gemund, C., Chabanis-Davidson, S., Mattingsdal, M., Cameron, S., Martin, D.M., Ausiello, G., Brannetti, B. et al. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.*, **31**, 3625–3630.
9. Fuxreiter, M., Tompa, P. and Simon, I. (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, **23**, 950–956.
10. Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z. and Dunker, A.K. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.*, **32**, 1037–1049.
11. Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M. and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry*, **41**, 6573–6582.
12. Bader, G.D., Betel, D. and Hogue, C.W. (2003) BIND: the biomolecular interaction network database. *Nucleic Acids Res.*, **31**, 248–250.
13. Mishra, G.R., Suresh, M., Kumaran, K., Kannabiran, N., Suresh, S., Bala, P., Shivakumar, K., Anuradha, N., Reddy, R. et al. (2006) Human protein reference database—2006 update. *Nucleic Acids Res.*, **34**, D411–D414.
14. Ceol, A., Chatranyamontri, A., Santonico, E., Sacco, R., Castagnoli, L. and Cesareni, G. (2007) DOMINO: a database of domain-peptide interactions. *Nucleic Acids Res.*, **35**, D557–D560.
15. Diella, F., Cameron, S., Gemund, C., Linding, R., Via, A., Kuster, B., Sicheritz-Ponten, T., Blom, N. and Gibson, T.J. (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics*, **5**, 79.
16. Obenaus, J.C., Cantley, L.C. and Yaffe, M.B. (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.*, **31**, 3635–3641.
17. Hermjakob, H., Montecchi-Palazzi, L., Bader, G., Wojcik, J., Salwinski, L., Ceol, A., Moore, S., Orchard, S., Sarkans, U. et al. (2004) The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.*, **22**, 177–183.
18. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
19. Gansner, E.R. and North, S.C. (2000) An open graph visualization system and its applications to software engineering. *Softw. Pract. Exp.*, **30**, 1203–1233.
20. Isobe, T., Ichimura, T., Sunaya, T., Okuyama, T., Takahashi, N., Kuwano, R. and Takahashi, Y. (1991) Distinct forms of the protein kinase-dependent activator of tyrosine and tryptophan hydroxylases. *J. Mol. Biol.*, **217**, 125–132.
21. Bork, P., Hofmann, K., Bucher, P., Neuwald, A.F., Altschul, S.F. and Koonin, E.V. (1997) A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J.*, **11**, 68–76.
22. Yu, X., Chini, C.C., He, M., Mer, G. and Chen, J. (2003) The BRCT domain is a phospho-protein binding domain. *Science*, **302**, 639–642.
23. Benes, C.H., Wu, N., Elia, A.E., Dharia, T., Cantley, L.C. and Soltoff, S.P. (2005) The C2 domain of PKCdelta is a phosphotyrosine binding domain. *Cell*, **121**, 271–280.
24. Hofmann, K. and Bucher, P. (1995) The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem. Sci.*, **20**, 347–349.
25. Durocher, D., Taylor, I.A., Sarbassova, D., Haire, L.F., Westcott, S.L., Jackson, S.P., Smerdon, S.J. and Yaffe, M.B. (2000) The molecular basis of FHA domain: phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol. Cell*, **6**, 1169–1182.
26. Wu, J.W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D.J., Kyin, S., Muir, T.W. et al. (2001) Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. *Mol. Cell*, **8**, 1277–1289.
27. Elia, A.E., Cantley, L.C. and Yaffe, M.B. (2003) Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates. *Science*, **299**, 1228–1231.
28. Zhou, M.M., Ravichandran, K.S., Olejniczak, E.F., Petros, A.M., Meadows, R.P., Sattler, M., Harlan, J.E., Wade, W.S., Burakoff, S.J. et al. (1995) Structure and ligand recognition of the phosphotyrosine binding domain of Shc. *Nature*, **378**, 584–592.
29. Russell, R.B., Breed, J. and Barton, G.J. (1992) Conservation analysis and structure prediction of the SH2 family of phosphotyrosine binding domains. *FEBS Lett.*, **304**, 15–20.
30. Sadowski, I., I. Stone, J.C. and Pawson, T. (1986) A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol. Cell. Biol.*, **6**, 4396–4408.
31. Li, D. and Roberts, R. (2001) WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cell. Mol. Life Sci.*, **58**, 2085–2097.
32. van der Voorn, L. and Ploegh, H.L. (1992) The WD-40 repeat. *FEBS Lett.*, **307**, 131–134.
33. Bork, P. and Sudol, M. (1994) The WW domain: a signalling site in dystrophin? *Trends Biochem. Sci.*, **19**, 531–533.