Phosphorylation is the most common protein post-translational modification. Phosphorylated residues (serine, threonine and tyrosine) play critical roles in the regulation of many cellular processes. Since the amount of data produced by screening assays is growing continuously, the development of computational tools for collecting and analysing experimental data has become a pivotal task for unravelling the complex network of interactions regulating eukaryotic cell life. Here we present Phospho3D, a database of 3D structures of phosphorylation sites, which stores information retrieved from the phospho.ELM database and is enriched with structural information and annotations at the residue level. The database also collects the results of a large-scale structural comparison procedure providing clues for the identification of new putative phosphorylation sites.

DATABASE CONSTRUCTION AND CONTENT

The Phospho3D database was constructed by collecting data from the phospho.ELM database which gathers experimentally verified phosphorylation sites manually extracted from the literature. The phospho.ELM dataset used in this work (version 4.0) contains 5314 phosphorylation sites, or instances, belonging to 1805 different sequences. Although the amount of data produced in various screening assays is steadily growing (3–6), experimental identification of phosphoproteins and the determination of individual phosphorylation sites remains a difficult and time-consuming task. Hence, the implementation of computational tools proves to be very useful for collecting and analysing experimental data.
annotated links between SwissProt-TrEMBL and PDB sequences. Links are based on sequence alignment using pre-established highly reliable thresholds. From a list of 4530 sequence–structure links (for further details see website documentation), only the ones having the phosphorylatable residue in the alignment region were retained, this resulting in 2726 instances (166 unique phospho.ELM instances on 1219 protein chains).

The basic information stored in Phospho3D consists of the instance, its flanking sequence (10 residues) and any residue whose distance from the instance does not exceed 12 Å thus defining a 3D neighbourhood which we define as zone. For each zone, annotation at the residue level is provided, namely solvent accessibility supplied by the NACCESS program (17), secondary structure assignment given by the DSSP program (18) and residue conservation as from the Consurf-HSSP database (19).

Users can also retrieve information extracted from the phospho.ELM dataset; for instance, the Medline reference PMID and, when available, the kinase(s) that phosphorylate(s) the given site.

In addition, for each zone the results of a large-scale local structural comparison versus a representative dataset of PDB (20) protein chains from eukaryotic organisms are also given. The comparison was carried out using the Query3D sequence/fold independent algorithm (21). Structural matches are assessed by two criteria: structural similarity and biochemical similarity. The structural similarity demands that matching residues have a root mean square deviation (r.m.s.d.) lower than a given threshold, whereas the biochemical similarity is evaluated using a Dayhoff substitution matrix (22). The score of the match is the number of matching residues which fulfil the similarity criteria. The significance of the score is evaluated by calculating the Z-score over the score distribution of the query zone comparison to the whole dataset.

**THE WEB INTERFACE**

The Phospho3D database can be searched by kinase name, by PDB identification code or keyword. A browsing function has been also implemented.

The information returned to the user consists of a brief description of the PDB structure(s) which fulfil the search criterion and of a list of instances presented along with...
CONCLUSION AND FUTURE PERSPECTIVES

The Phospho3D database is a useful tool for the analysis of the structural features of experimentally verified phosphorylation sites. Moreover, it provides the results of a large-scale local structural comparison between the *zones* and a representative set of eukaryotic protein chains. The results of such a comparison identify new putative phosphorylation sites and suggest the kinase(s) responsible for phosphorylation.

Phospho3D will be regularly updated as soon as the new Phospho.ELM datasets are released. The annotations will be integrated as a feature in the pdbFun server (23). We are also planning to identify and annotate those sites which are recognized by protein phosphatases and phosphoresidues-binding modules (24–26).

The Phospho3D dataset (annotations at the residue level and structural comparison results) is available upon request.

ACKNOWLEDGEMENTS

The authors would like to thank Francesca Diella for support and suggestions and for providing the Phospho.ELM dataset. We acknowledge the support of AIRC, Telethon (GGP40273), a PNR 2001–2003 (FIRB art.8) and a PNR 2003–2007 (FIRB art.8). Funding to pay the Open Access publication charges for this article was provided by AIRC.

Conflict of interest statement. None declared.

REFERENCES

1. Cohen,P. (2002) The origins of protein phosphorylation. *Nature Cell. Biol.*, 4, E127–E130.
2. Manning,G., Ploiman,G.D., Hunter,T. and Sudarsanam,S. (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.*, 27, 514–520.
3. Salomon,A.R., Picarro,S.B., Brill,L.M., Brinker,A., Phang,Q.T., Ericson,C., Sauer,K., Brock,A., Horn,D.M., Schultz,P.G. et al. (2003) Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc. Natl Acad. Sci. USA*, 100, 443–448.
4. Shu,H., Chen,S., Bi,Q., Mumbly,M. and Brekken,D.L. (2004) Identification of phosphoproteins and their phosphorylation sites in the WEHI-231 B lymphoma cell line. *Mol. Cell. Proteomics*, 3, 279–286.
5. Brill,L.M., Salomon,A.R., Picarro,S.B., Mukherji,M., Stettler-Gill,M. and Peters,E.C. (2004) Robust phosphopeptide profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal. Chem.*, 76, 2763–2772.
6. Beausoleil,S.A., Jedrychowski,M., Schwartz,D., Elias,J.E., Villen,J., Li,J., Cohn,M.A., Cantley,L.C. and Gygi,S.P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl Acad. Sci. USA*, 101, 12130–12135.
7. Puntervoll,P., Linding,R., Gemund,C., Chabanis-Davidson,S., Mattingsdal,M., Cameron,S., Martin,D.M., Ausiello,G., Brannetti,B., Costantini,A. et al. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.*, 31, 3625–3630.
8. Oenbauer,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.
9. Blohm,N., Gammeltoft,S. and Brunak,S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.*, 294, 1351–1362.
10. Iakovouchea,L.M., Radiovicaj,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. Plewczynski,D., Tkacz,A., Godzik,A. and Rychlewski,L. (2005) A support vector machine approach to the identification of phosphorylation sites. *Cell. Mol. Biol. Lett.*, 10, 73–89.
12. Kim,J.H., Lee,J., Oh,B., Kim,M. and Koh,I. (2004) Prediction of phosphorylation sites using SVMs. *Bioinformatics*, 20, 3179–3184.
13. Huang,H.D., Lee,T.Y., Tseng,S.W., Wu,L.C., Hong,J.T., Tsou,A.P. and Huang,K.T. (2005) Incorporating hidden Markov models for identifying protein kinase-specific phosphorylation sites. *J. Comput. Chem.*, 26, 1032–1041.
14. Schwartz,D. and Gygi,S.P. (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.*, 23, 1239–1240.
15. Diella,F., Cameron,S., Gemund,C., Linding,R., Via,A., Kuster,B., Sicheritz-Ponten,T., Blom,N. and Gibson,T.J. (2005) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics*, 5, 79.
16. Via,A., Zanoni,A. and Helmer-Citterich,M. (2005) Seq2Struct: a resource for establishing sequence-structure links. *Bioinformatics*, 21, 551–553.
17. Hubbard,S. and Thornton,J.M. (1993) NACCESS Computer Program. Department of Biochemistry and Molecular Biology, University College, London.
18. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22, 2577–2637.
19. Glaser,F. and Rosenberg,Y. (1999) The SPOT technique as a tool for studying protein tyrosine phosphatase substrate specificities. *Cell. Mol. Biol. Lett.*, 5, 305–313.
20. Walchli,S., Espanel,X., Harrenga,A., Rossi,M., Cesareni,G. and Via,A. (2005) Query3d: a new method for high-throughput analysis of functional residues in protein structures. *BMC Bioinformatics*, 4, 55.
21. Vassy,M.O., Schwartz,R.M. and Orcutt,B.C. (1978) A model of evolutionary change in proteins. *J. Mol. Biol.*, 120, 443–448.
22. Beasoleil,S.A., Jedrychowski,M., Schwartz,D., Elias,J.E., Villen,J., Li,J., Cohn,M.A., Cantley,L.C. and Gygi,S.P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl Acad. Sci. USA*, 101, 12130–12135.
23. Manning,G., Plowman,G.D., Hunter,T. and Sudarsanam,S. (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.*, 27, 514–520.
24. Salomon,A.R., Picarro,S.B., Brill,L.M., Brinker,A., Phang,Q.T., Ericson,C., Sauer,K., Brock,A., Horn,D.M., Schultz,P.G. et al. (2003) Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc. Natl Acad. Sci. USA*, 100, 443–448.
25. Shu,H., Chen,S., Bi,Q., Mumbly,M. and Brekken,D.L. (2004) Identification of phosphoproteins and their phosphorylation sites in the WEHI-231 B lymphoma cell line. *Mol. Cell. Proteomics*, 3, 279–286.
26. Brill,L.M., Salomon,A.R., Picarro,S.B., Mukherji,M., Stettler-Gill,M. and Peters,E.C. (2004) Robust phosphopeptide profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal. Chem.*, 76, 2763–2772.
27. Beausoleil,S.A., Jedrychowski,M., Schwartz,D., Elias,J.E., Villen,J., Li,J., Cohn,M.A., Cantley,L.C. and Gygi,S.P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl Acad. Sci. USA*, 101, 12130–12135.