PhosphoSite®: A Knowledge Base of Post-Translational Modifications and Their Role in Cell Signaling.

David Merberg[1], Roland Carel[2], Jon Kornhauser[1], Juan Munoz[2], Beth Murray[1], Julie Nardone[1], Ousama Shamma[2], Elzbieta Skrzypek[1], Bin Zhang[1] and Peter Hornbeck[1] • [1]Cell Signaling Technology, Danvers, MA USA / [2]Third Millennium, Waltham, MA USA

Abstract

Protein modifications are important modulators of protein activity. Because modification events occur subsequent to transcription and translation, they frequently exert a more direct effect on cellular processes. Hence, it is important for models of cellular systems to incorporate information about the post-translational modifications of their protein components.

In particular, phosphorylation of tyrosine, threonine, and serine residues on substrate proteins by specific protein kinases frequently modulates pathways that regulate cellular proliferation and differentiation. These pathways are the subject of intense study because of their role in oncogenesis and other disease processes.

PhosphoSite® provides systems biologists with comprehensive source of information concerning protein modifications and their role in regulating cellular functions and processes. Currently, the database contains over 58,000 phosphorylation sites on over 17,000 proteins. Approximately 15% of the sites have been identified via an internal high-throughput site discovery program, and have not been published elsewhere.

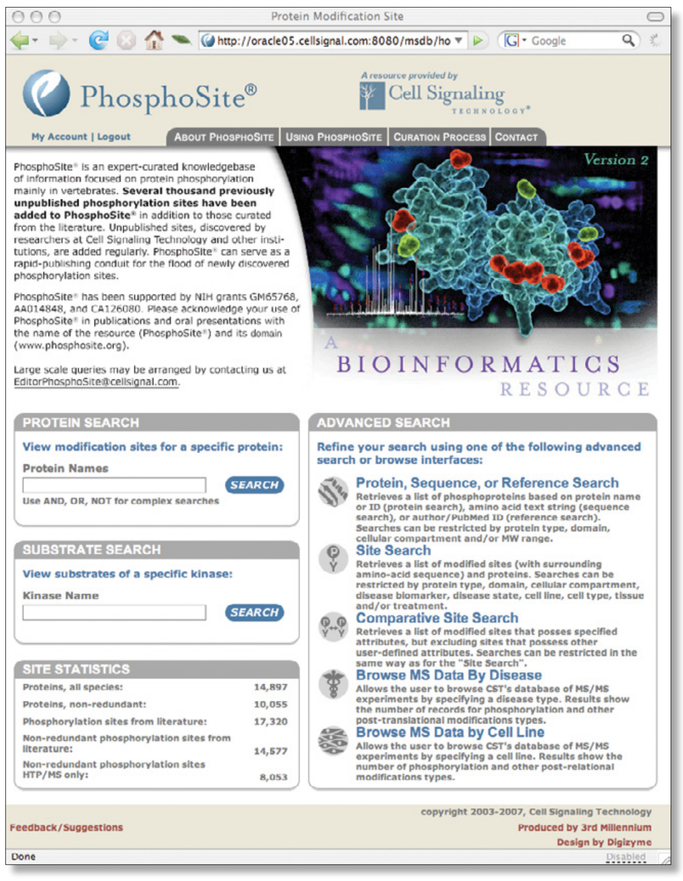
PhosphoSite® has been completely re-engineered to provide better access to the rapidly increasing volume of post-translational modification data. New user interfaces enable the user to formulate sophisticated queries based on treatments, diseases, cell lines, and tissues, and result in retrieval of bulk data sets. These bulk data sets can be downloaded for analysis and model-building. There are provisions for bulk queries that allow high-throughput labs to determine whether sites from their modification site discovery programs are novel. In addition, PhosphoSite® has been expanded to include acetylation and methylation sites, both from the literature and internal mass spectroscopy studies.

Finally, the application and database software have been rebuilt for improved performance and scalability. There have been numerous improvements to the underlying schema and all components have been developed using current generation enterprise technologies (Java 5, Tomcat, Oracle 10g).

PhosphoSite®

Future Enhancements:

- Export PhosphoSite® information in BioPAX, SBML, and other standard formats.
- View pathway diagrams showing post-translational modifications and upstream and downstream events.



PhosphoSite® version 2.0 – New Features *Available late summer 2007*

- Improved access to modification sites determined by mass spectrometry
- View all modification sites observed in one or more high throughput experiment(s).
- Utilize advanced searches to focus on proteins, treatments, tissues, sequences, substrate-kinase relationships, and references of particular interest.
- Download search results.
- View spectra for selected sites.
- Browse through mass spectrometry data, focusing on experiments of interest.
- View selected acetylation, methylation, and other modifications as determined by mass spectrometry.
- View and manipulate available protein structures with the AstexViewer™.
- Zoom and pan through a graphic representation of sites and domains on each protein.
- More intuitive navigation between protein isoforms.
- Improved reliability and scalability.

www.phosphosite.org

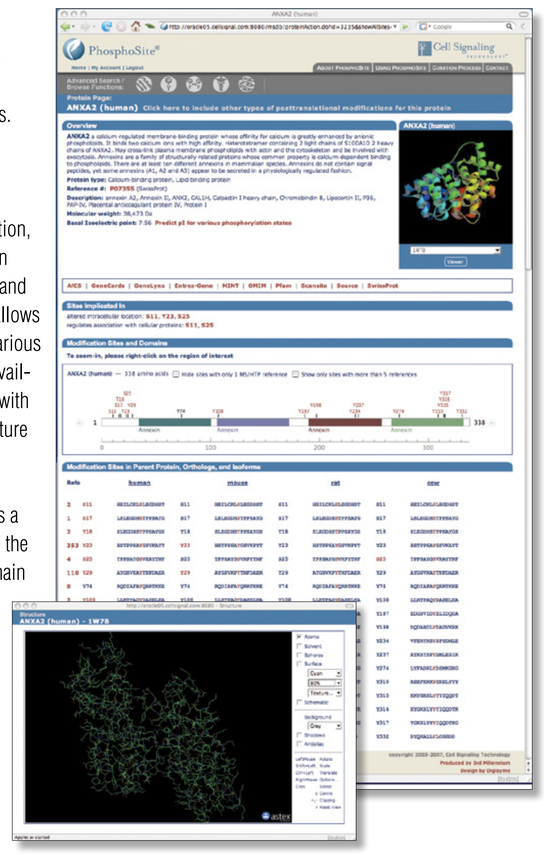
Protein Pages

The Protein Page summarizes information about each protein and its modification sites. There are links at the top of the page for navigating through the database, including to the Advanced Search and Browse functions. There is also a link to toggle between viewing all types of modifications or phosphorylations only.

Below the navigation panel is the Protein Overview section, which includes a brief description of the protein, protein type, the protein accession number, alternative names, and molecular weight. An interactive isoelectric calculator allows the user to retrieve the predicted isoelectric points of various phosphorylated isoforms of the parent protein. When available, a ribbon image of the protein structure is shown, with a link to activate the AstexViewer™ preloaded with structure coordinates.

The Domains section in the middle of the page contains a linear schematic of the protein, predicted domains, and the location of known modification sites relative to the domain structure. The diagram is interactive, so that the user can zoom into regions of interest that are crowded with many features.

The Sequence and Sites section at the bottom of the page presents the modified residues with surrounding amino acids (+/- 7) in multiple species. The number to the left of a sequence line is the number of references or MS experiments in the database for that site; it is hyperlinked to the full list of references/experiments. The modified residue is hyperlinked to the Site Page.



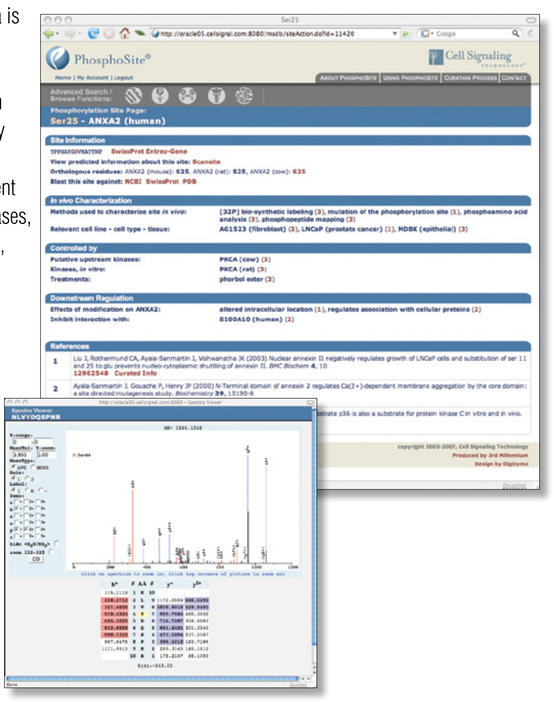
Modification Site Pages

Modification Site Pages serve detailed information about each modification site. At the top of the page is a Navigation Pane, similar to that found on the Protein Page.

Immediately below the Navigation Pane is a section that provides general information about the modification site: its sequence and orthologs. There are links to ScanSite (Obenauer, et al. 2003 Nucleic Acids Research) and Blast servers. When spectral data is available, there is a link to view them.

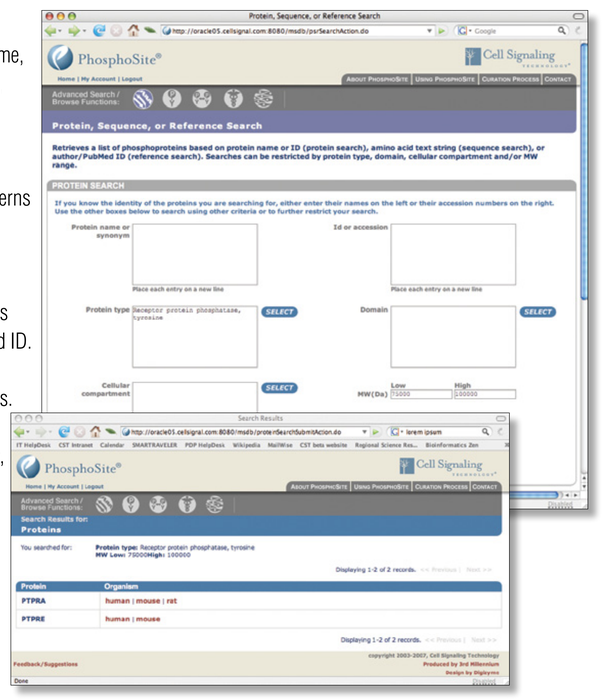
The next sections contain information about how the site was experimentally characterized (mass spectrometry, phospho-antibodies, etc.), the apparent regulation of phosphorylation by kinases, phosphatases, ligands, and receptors, and downstream regulation.

The bibliography contains a list of hyperlinked references and high-throughput experiments in which the modification site was observed.



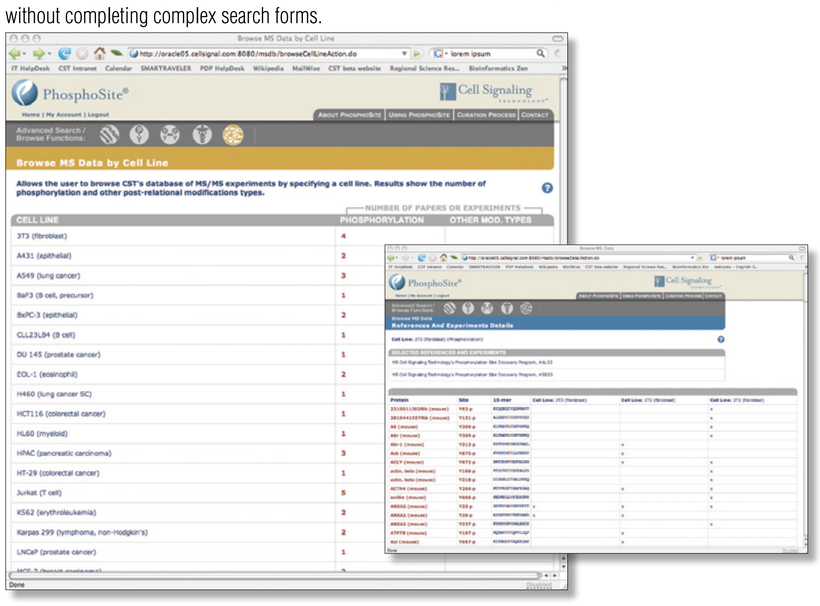
Advanced Searches

- Protein searches find modified proteins by name, accession, protein type, cellular compartment, domains or molecular weight range.
- Sequence searches find modified proteins by sequence, including ambiguous sequence patterns
- Reference searches return a list of references with links to the modification information that has been curated from them. Search parameters include protein name, author name, or PubMed ID.
- Site searches retrieve a list of modification sites. Possible search parameters include treatment, cell type, cell line, tissue, disease, protein type, domain, and cellular compartment.



Interfaces for browsing MS Data

Browse through the PhosphoSite® collection of modification sites reported in the literature and observed in Cell Signaling Technology's internal Cancer Biology program (Rush, et al. 2004 Nature Biotechnology). Sites are organized by cell line and disease. The browse interfaces enable you to see what data is available without completing complex search forms.



Cell Signaling TECHNOLOGY®

www.cellsignal.com