

A general user interface for prediction servers of proteins' post-translational modification sites

Fengfeng Zhou^{1,4}, Yu Xue^{2,4}, Xuebiao Yao^{2,3} & Ying Xu¹

¹Computational Systems Biology Laboratory, Department of Biochemical and Molecular Biology, and Institute of Bioinformatics, University of Georgia, Athens, Georgia 30602, USA. ²Laboratory of Cellular Dynamics, Hefei National Laboratory for Physical Sciences, and the University of Science and Technology of China, Hefei, China 230027. ³Department of Physiology and Cancer Research Program, Morehouse School of Medicine, Atlanta, Georgia 30310, USA. ⁴These authors contributed equally to this work. Correspondence should be addressed to Y.X. (xyn@bmb.uga.edu) or X.Y. (yaoxb@ustc.edu.cn).

Published online 26 October 2006; doi:10.1038/nprot.2006.209

Post-translational modifications (PTMs) of proteins play essential roles in governing the functions and dynamics of proteins and are implicated in many cellular processes. Several types of PTMs have been investigated through computational approaches, including phosphorylation, sumoylation, palmitoylation, and lysine and arginine methylation, among others. Because the large diversity in the user interfaces (UIs) of different prediction servers for PTMs could possibly hinder experimental biologists in using these servers, we propose to develop a protocol for a unified UI for PTM prediction servers, based on our own work and that of other groups on PTM site prediction. By following this protocol, tool developers can provide a uniform UI regardless of the PTM types and the underlying computational algorithms. With such uniformity in the UI, experimental biologists would be able to use any PTM prediction server compliant with this protocol once they had learned to use one of them. It takes a typical PTM prediction server compliant with this unified UI several minutes to calculate the prediction results for a protein 1,000 amino acids in length.

INTRODUCTION

By covalently attaching to individual amino acids various functional molecules such as phosphates, lipids or proteins, PTMs alter proteins' biochemical natures significantly and play key roles in a wide variety of cellular processes. Although more than 350 types of PTMs have been discovered (see the database of protein translational modifications at <http://abrf.org/index.cfm/dm.home>), only a few of them have been well characterized^{1–10} because of the lack of sufficient data for analyses. Experimental identification of proteins PTM sites is labor-intensive and usually limited by the availability of detailed data on their enzymatic reactions. *In silico* prediction could be a promising strategy to conduct preliminary analyses and could greatly reduce the number of potential targets that need further *in vivo* or *in vitro* confirmation.

To the best of our knowledge, about 32 PTM site prediction servers have been developed and made publicly available through the internet. Although these prediction servers in general require the same input, i.e., protein sequences, and output similar information about different types of PTMs, such as the sequential location of a PTM, a flanking peptide sequence of the PTM site and an estimated prediction reliability score or a probability score, etc., they have all adopted different input-output interfaces, based on individual developers' preference. This has made it difficult for experimental biologists to learn how to use the many different interfaces with these prediction servers. With more and more experimental data becoming available, we believe that many more prediction servers will emerge for the prediction of different types of PTMs to meet the needs of proteomic studies. Hence, now seems like a good time to begin discussing and developing prototype systems for a unified UI for such prediction servers, so as to better serve the general user community.

In this paper, we describe a proposed protocol for setting up and using a unified UI that contains features widely used by most currently available PTM site prediction servers, as well as how to use and interpret the output results from such servers. Computa-

tional tool developers may follow this protocol to provide a uniform UI for their new prediction servers, regardless of the PTM types and the underlying computational algorithms. And experimental biologists would be able to use any PTM prediction servers compliant with this protocol after becoming familiar with one of them. To facilitate the adoption of our proposed protocol by other developers of PTM prediction servers, we have created a suite of tools useful for adapting to our proposed UI and supporting displaying predicted information (by other servers) in a richer context. These tools, along with a list of available PTM prediction servers, can be accessed at our Post-Translational Modification Site Prediction–User Interface page (<http://csbl.bmb.uga.edu/~ffzhou/PTMP-UI/>) and <http://bioinformatics.lcd-ustc.org/PTMP-UI/>). Although the current tool set is small, we plan to continually put out more powerful and attractive tools for displaying the predicted PTM information, including (i) display of PTMs in both protein sequence and three-dimensional (3D) structure if such information is available and (ii) a database for keeping track of predicted PTMs by available servers that will allow individual tool developers to cross-validate their predictions against PTMs predicted by other prediction servers. We believe that such a unified UI protocol, along with the associated toolkit, will be attractive enough to persuade other developers to either join forces with us to further develop this unified protocol or simply follow our proposed protocol in their development, which will ultimately benefit the users of PTM prediction tools.

Rationale for a unified UI

We have assessed 32 PTM prediction servers that are accessible through the internet. We found that the following features are the most commonly adopted, as indicated by the percentages of servers adopting them, and have included them in our proposed UI.

1. Input: protein primary sequences (31/32 = 96.88%)
2. Output: predicted PTM location (30/32 = 93.75%)

TABLE 1 | PTM site prediction servers.

PTM type	Web server and URL	Ref(s).
Phosphorylation	GPS: http://973-proteinweb.ustc.edu.cn/gps/gps_web/	1,2
	ScanSite: http://scansite.mit.edu/	3
	PPSP: http://bioinformatics.lcd-ustc.org/PPSP	4
	NetPhosK: http://www.cbs.dtu.dk/services/NetPhosK/	5
	PredPhospho: http://www.ngri.re.kr/proteo/PredPhospho.htm	6
	DISPHOS: http://core.ist.temple.edu/pred/pred.html	7
SUMOylation	SUM0sp: http://bioinformatics.lcd-ustc.org/sumosp/	8
	SUM0plot: http://www.abgent.com/doc/sumoplot	
Palmitoylation	CSS-Palm: http://bioinformatics.lcd-ustc.org/css_palm/	9
Methylation	MeMo: http://www.bioinfo.tsinghua.edu.cn/~tigerchen/memo.html	10

3. Output: flanking peptide of the PTM site (20/32 = 62.5%)

4. Output: evaluation score or probability of predicted PTM site (28/32 = 87.5%)

In addition to these commonly used features, we have also included in our unified UI several other easy-to-implement features that should make the display scheme richer in terms of its information content.

To convince and facilitate other developers to adopt our proposed UI, we have developed a number of tools that should make the adoption of our display protocol more attractive for other developers. Initially, these tools include two Perl scripts to help display the predicted PTM information on the 3D structure of a target protein, visualized using RasMOL¹¹, if such information is available. For the long-term development of this proposed UI, we plan to continually add to the toolkit other capabilities to make its display capabilities much richer. In the very near future, we expect to provide further support for more 3D structure visualization capabilities, e.g., PyMOL (DeLano, W.L., The PyMOL Molecular Graphics System (2002; <http://www.pymol.org>) and to develop comparative analysis tools in support of cross-validation. All of these supporting tools will be deposited into a database at <http://csbl.bmb.uga.edu/>

ffzhou/PTMP-UI/ and <http://bioinformatics.lcd-ustc.org/PTMP-UI/>, along with a list of available PTM prediction servers accessible on the Internet. We will ask developers to add an additional display mode, using our proposed protocol, in order to freely obtain and use our developed tool kit, which should greatly enrich the information displayed on their servers.

Using the procedure

Option A describes a procedure for a user to follow. We use the CSS-Palm server⁹ as an example to describe how to use a PTM prediction server and interpret the computing results. Users can access the CSS-Palm server at http://bioinformatics.lcd-ustc.org/css_palm/. It should be noted that any prediction server can serve the same purpose here. To help a user to identify a relevant PTM server, we provide a partial list of the available PTM prediction servers in Table 1. A complete list of such servers is given at our Post-Translational Modification Site Prediction–User Interface page (<http://csbl.bmb.uga.edu/~ffzhou/PTMP-UI/> and <http://bioinformatics.lcd-ustc.org/PTMP-UI/>), which is updated on regular basis. Option B describes how to design a PTM prediction server compliant with this protocol.

MATERIALS EQUIPMENT

Any graphical user interface–based web browser, running on a computer with Internet connection, should suffice.

PROCEDURE

1| Option A describes a procedure for a user to follow. Option B describes a procedure for a developer to follow, focusing on the UI of a PTM prediction server. The proposed features are organized in an order similar to that of the procedures for server users (Option A).

(A) Procedure for users

- (i) Enter query sequences. Prepare one or several protein sequences for PTM site prediction. Most of the prediction servers accept protein sequences in raw data or FASTA format (both of which are illustrated in Fig. 1).
- (ii) Use prediction server. To use CSS-Palm server (or any selected sever), access the prediction page at http://bioinformatics.lcd-ustc.org/css_palm/prediction.php. Enter the protein sequences (the input to the server)

```
MPAGTAARAWLVLALWGAVAGQNI TARIGEPLVLSCKGAPKPPQQLWKLNTGRTEA
WKVLSPOGGPWSVARILPNGSLLLPATGIVDEGTRFRCRATNRRGKEVKSNYRVRVYQIP
GKPEIVDPASELTASVFNKVGTCVSEGSYPAGTLSWHLDGKLLIPDGKETLVKEETRRHP
ETGLFTRSELTVIPTQGGTHPTFSCFSFSLGPRRRPLNTAPIQLRVREPDPPEGIQLLV
EPEGGIVAPGGTVLTCASIAQPPQVHWIKDGAAPLPLAPSPVLLPEVGHDEGTYSCV
ATHPSHGREQSPVSI RVTETGDEGPAEGSVGSEGLGLTALALGILGGLGVALLVGAITL
WRKRQPRREERKAPESQDEEERAE LNQSEEAEMPENAGGGP

Raw data of protein O35444

>tr|O35444|O35444_MOUSE RAGE (Advanced glycosylation end product-specific receptor).
MPAGTAARAWLVLALWGAVAGQNI TARIGEPLVLSCKGAPKPPQQLWKLNTGRTEA
WKVLSPOGGPWSVARILPNGSLLLPATGIVDEGTRFRCRATNRRGKEVKSNYRVRVYQIP
GKPEIVDPASELTASVFNKVGTCVSEGSYPAGTLSWHLDGKLLIPDGKETLVKEETRRHP
ETGLFTRSELTVIPTQGGTHPTFSCFSFSLGPRRRPLNTAPIQLRVREPDPPEGIQLLV
EPEGGIVAPGGTVLTCASIAQPPQVHWIKDGAAPLPLAPSPVLLPEVGHDEGTYSCV
ATHPSHGREQSPVSI RVTETGDEGPAEGSVGSEGLGLTALALGILGGLGVALLVGAITL
WRKRQPRREERKAPESQDEEERAE LNQSEEAEMPENAGGGP

>sp|Q62151|RAGE_MOUSE (Receptor for advanced glycosylation end products).
MPAGTAARAWLVLALWGAVAGQNI TARIGEPLVLSCKGAPKPPQQLWKLNTGRTEA
WKVLSPOGGPWSVAQILPNGSLLLPATGIVDEGTRFRCRATNRRGKEVKSNYRVRVYQIP
GKPEIVDPASELTASVFNKVGTCVSEGSYPAGTLSWHLDGKLLIPDGKETLVKEETRRHP
ETGLFTRSELTVIPTQGGTHPTFSCFSFSLGPRRRPLNTAPIQLRVREPDPPEGIQLLV
VEPEGGIVAPGGTVLTCASIAQPPQVHWIKDGAAPLPLAPSPVLLPEVGHDEGTYSCV
VATHPSHGREQSPVSI RVTETGDEGPAEGSVGSEGLGLTALALGILGGLGVALLVGAITL
LWRKRQPRREERKAPESQDEEERAE LNQSEEAEMPENAGGGP

Sequences of O35444 and Q62151 in FASTA format
```

Figure 1 | Examples of raw data and FASTA format of protein sequences.



into the 'input-sequence' box. Alternatively, press the button 'Example (CENPE_HUMAN)' in the prediction page (Fig. 2) to load an example sequence into the text box. The CSS-Palm server has only one adjustable parameter, the cutoff value, which has a default value set by the server developer. You can either choose your own values for the parameters through the interface or decide to use the default parameter values. CSS-Palm calculates a score for each cysteine residue in the input protein, based on the flanking residues of the residue, that indicates the possibility of the residue being a palmitoylation site. A cutoff threshold (a parameter) is used to make the final prediction. Usually the parameter-tuning step can be bypassed, in which case you will start the prediction procedure by pushing the 'Submit' button. The prediction results of an example protein, CENPE_HUMAN, are shown in Figure 3.

? TROUBLESHOOTING

- (iii) Interpret prediction results. CSS-Palm, like most of the PTM prediction servers, gives a table of the predicted PTM sites along with their scores. Usually, the higher the score, the higher the reliability of the prediction.

? TROUBLESHOOTING

(B) Procedure for developers

- (i) User interface. The web pages of a prediction server must be compliant with the W3C recommendation formats of HTML and XHTML, and there must be no errors when these web pages are displayed in various popular web browsers, such as Mozilla, Mozilla Firefox, Netscape Browser and Internet Explorer.

? TROUBLESHOOTING

- (ii) The PTM prediction server should have an introduction page to provide basic information about the definition of a particular PTM type; the computational algorithm used for PTM prediction; the parameters used by the algorithm, along with an intuitive explanation of each parameter and a suggested range of possible values; an intuitive explanation of how the predicted results should be interpreted; and links to other servers of the same PTM type.
- (iii) Input formats. 1. A PTM prediction server should accept both of raw sequence and FASTA formats of target protein sequences. We found that 31 out of 32 existing PTM prediction servers are based on protein sequences. But the requirements of the input formats are diverse: e.g., with ScanSite 3, it is necessary to assign a protein name³. 2. Each server should provide an 'Example' button. Clicking this button should cause an example sequence to be loaded into the sequence text box, displayed through its front prediction page. 3. There should also be a default value for each parameter that can be altered, for which the PTM prediction server should function correctly. This is because most experimentalists are not very familiar with underlying mathematical models.
- (iv) Output formats. Each prediction server should output at least the following information for each of its prediction: 1. the location in the protein sequence of a predicted PTM site, 2. the flanking sequences of a predicted PTM site on both the left and the right sides, 3. a reliability or probability score for each prediction and 4. a qualitative indicator about the confidence of the prediction (highly confident, marginally confident, not confident) that is consistent with the numerical values in 3.

In addition, each prediction server should provide its prediction results in a computer-readable format, specifically in a tab-delimited data file, which can be used by other computer programs to read and process.

● TIMING

For users, it takes a typical PTM prediction server compliant with this unified UI several minutes to calculate the prediction results for a protein with length 1,000 amino acids. For designers of PTM prediction servers, adaptation of this display protocol can be done easily and should take no more than a few minutes.

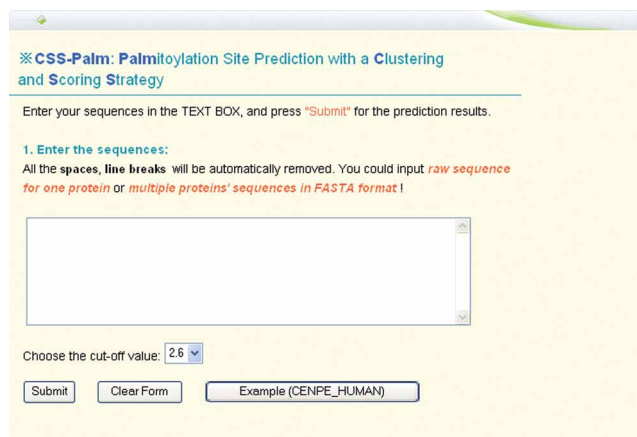


Figure 2 | Prediction page of the CSS-Palm web server.

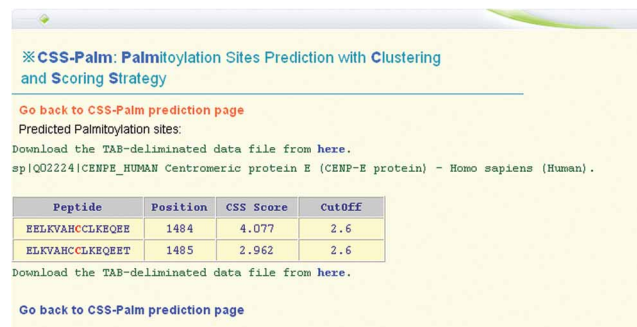


Figure 3 | Prediction result page of the example protein CENPE_HUMAN.

? TROUBLESHOOTING

For troubleshooting guidance see **Table 2**.

TABLE 2 | Troubleshooting table.

STEP AND PROBLEM	SOLUTION
Step 1A(ii): Why do I get an error for my raw/FASTA sequences?	The input formats of the PTM prediction servers are very diverse, but most accept at least one of the two formats: raw sequence of one protein or sequences of multiple proteins in FASTA format. Therefore, if you get an error with your sequences, please try another format.
Step 1A(iii): Why can't I get any predictions with my sequence?	Your sequence format may not be acceptable by the prediction server. Please refer to the entry above concerning Step 1A(ii) above. Your sequence may be predicted to contain no PTM sites with the provided parameter values, such as threshold cutoff value. You can lower the threshold to get more prediction results, and conduct a comparative analysis to improve the prediction. As in CSS-Palm, you could change the cutoff value from the default 2.6 to 1.5.
Step 1B(i): Where and how can I make my PTM prediction server compliant with the standards of HTML and XHTML?	Check whether your server is compliant with the standards of the validators provided at http://www.w3.org . Sometimes it is too strict to be fully compliant with these standards. For example, even PubMed cannot fully pass the above validators. Therefore, you may just want to validate your server in several popular web browsers: e.g., Mozilla, Mozilla Firefox, Netscape Browser and Internet Explorer.

ANTICIPATED RESULTS

A unified UI will make the lives of experimentalists easier when they try to use PTM prediction servers. Clearly our proposed unified UI, as simple as it is, is just a start toward getting the tool developer community to move in this direction. We plan to continuously push this forward and to get more PTM server developers to join in our effort. We also plan to continue to develop many more post-processing tools to help enrich information display for PTM prediction servers, which should be attractive to other developers. For our current effort, we have developed a number of tools that we offer for use by other prediction server developers as an incentive for them to take part (see INTRODUCTION above).

After this protocol is published, we intend to organize a workshop affiliated with one of the top bioinformatics conferences, such as RECOMB, ISMB, CSB, etc., to discuss with other PTM server developers the further development of a unified UI for all PTM prediction servers, possibly using our proposed UI as the basis for future development. This should help to push the adaptation of a unified UI for all PTM servers through a community-wide effort.

ACKNOWLEDGMENTS F. Zhou's and Y. Xu's work is supported in part by the National Science Foundation (NSF/DBI-0354771, NSF/ITR-IIS-0407204, NSF/DBI-0542119, NSF/CCF0621700) and a Distinguished Scholar award from the Georgia Cancer Coalition. Y. Xue's and X. Yao's work is supported by Chinese Natural Science Foundation (39925018, 30270654 and 30270293), Chinese Academy of Science (KSCX2-2-01), Chinese 973 project (2002CB713700), Chinese Minister of Education (20020358051), American Cancer Society (RPG-99-173-01), National Institutes of Health (DK56292; CA92080) and a Distinguished Scholar award from the Georgia Cancer Coalition.

COMPETING FINANCIAL INTERESTS The authors declare that they have no competing financial interests.

Published online at <http://www.natureprotocols.com/>
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Zhou, F., Xue, Y., Chen, G. & Yao, X. GPS: a novel group-based phosphorylation prediction and scoring method. *Biochem. Biophys. Res. Commun.* **325**, 1443–1448 (2004).
- Xue, Y., Zhou, F., Zhu, M., Ahmed, K., Chen, G. & Yao, X. GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res.* **33**, W184–W187 (2005).

- Obenauer, J.C., Cantley, L.C. & Yaffe, M.B. Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* **31**, 3635–3641 (2003).
- Xue, Y., Li, A., Wang, L., Feng, H. & Yao, X. PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinformatics* **7**, 163 (2006).
- Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633–1649 (2004).
- Kim, J.H., Lee, J., Oh, B., Kimm, K. & Koh, I. Prediction of phosphorylation sites using SVMs. *Bioinformatics* **20**, 3179–3184 (2004).
- Iakoucheva, L.M. *et al.* The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049 (2004).
- Xue, Y., Zhou, F., Fu, C., Xu, Y. & Yao, X. SUMOsp: a web server for SUMOylation site prediction. *Nucleic Acids Res.* **34**, W254–W257 (2006).
- Zhou, F., Xue, Y., Yao, X. & Xu, Y. CSS-Palm: palmitoylation site prediction with a clustering and scoring strategy (CSS). *Bioinformatics* **22**, 894–896 (2006).
- Chen, H., Xue, Y., Huang, N., Yao, X. & Sun, Z. MeMo: a web tool for prediction of protein methylation modifications. *Nucleic Acids Res.* **34**, W249–W253 (2006).
- Sayle, R.A. & Milner-White, E.J. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374–376 (1995).

