BIOINFORMATICS APPLICATIONS NOTE Vol. 22 no. 7 2006, pages 894–896
doi:10.1093/bioinformatics/btl013

Sequence analysis

CSS-Palm: palmitoylation site prediction with a clustering and scoring strategy (CSS)

Fengfeng Zhou[1,†], Yu Xue[2,†], Xuebiao Yao[2,3,*] and Ying Xu[1,*]

[1]Computational Systems Biology Laboratory, Department of Biochemical and Molecular Biology and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA, [2]Laboratory of Cellular Dynamics, Hefei National Laboratory for Physical Sciences, and the University of Science and Technology of China, Hefei, China 230027, China and [3]Department of Physiology and Cancer Research Program, Morehouse School of Medicine, Atlanta, GA 30310, USA

Received on December 20, 2005; revised on January 19, 2006; accepted on January 20, 2006
Advance Access publication January 24, 2006
Associate Editor: Martin Bishop

ABSTRACT

Summary: Palmitoylation is an important post-translational lipid modification of proteins. Unlike prenylation and myristoylation, palmitoylation is a reversible covalent modification, allowing for dynamic regulation of multiple complex cellular systems. However, *in vivo* or *in vitro* identification of palmitoylation sites is usually time-consuming and labor-intensive. So *in silico* predictions could help to narrow down the possible palmitoylation sites, which can be used to guide further experimental design. Previous studies suggested that there is no unique canonical motif for palmitoylation sites, so we hypothesize that the bona fide pattern might be compromised by heterogeneity of multiple structural determinants with different features. Based on this hypothesis, we partition the known palmitoylation sites into three clusters and score the similarity between the query peptide and the training ones based on BLOSUM62 matrix. We have implemented a computer program for palmitoylation site prediction, Clustering and Scoring Strategy for Palmitoylation Sites Prediction (CSS-Palm) system, and found that the program's prediction performance is encouraging with highly positive Jack-Knife validation results (sensitivity 82.16% and specificity 83.17% for cut-off score 2.6). Our analyses indicate that CSS-Palm could provide a powerful and effective tool to studies of palmitoylation sites.

Availability: CSS-Palm is implemented in PHP/PERL+MySQL and can be freely accessed at http://bioinformatics.lcd-ustc.org/css_palm/

Contact: yaoxb@ustc.edu.cn; xuyn@bmb.uga.edu

Supplementary information: Supplementary data are available at Bionformatics online.

INTRODUCTION

Many proteins are post-translationally modified by the addition of a palmitate molecule to a cysteine by thioesterification. Biochemically, palmitoylation enhances the surface hydrophobicity of protein substrates and promotes their interactions with membranes (Kleuss and Krause, 2003). Thus, some hydrophilic proteins like Ras and G proteins make use of this modification to attach themselves to membranes, often in combination with prenyl groups. Palmitoylation can also regulate intracellular trafficking (Kang *et al*., 2004), sorting (Schneider *et al*., 2005), subcellular localization (Van Itallie *et al*., 2005) and functional activities of the proteins (Sudo *et al*., 1992). Although palmitoylation is increasingly recognized as a frequent and important modification of eukaryotic signaling proteins, the molecular mechanism underlying protein palmitoylation remains elusive.

Identification of palmitoylation sites could provide an effective approach to understand the molecular mechanism of palmitoylation. To date, only a few palmitoylation sites have been identified experimentally, mainly through mutagenesis studies of candidate cysteine residues using conventional biochemical methods. The distinguishing features of palmitoylation sites had not been well-characterized, and most previous studies suggested that there is no canonical consensus sequence/motif for the palmitoylation sites (Bijlmakers and Marsh, 2003; el-Husseini Ael and Bredt, 2002; Linder and Deschenes, 2003; Smotrys and Linder, 2004; ten Brinke *et al*., 2002). Just as in the phosphorylation site prediction (Zhou *et al*., 2004), we propose that such a pattern may consist of multiple consensus motifs.

METHODS

Based on the above hypothesis, we present a novel computational method/software CSS-Palm—Palmitoylation Site Prediction using a Clustering and Scoring Strategy. We have collected 210 experimentally verified palmitoylation sites from 83 distinct proteins as training set, and grouped them into several subsets based on their sequence similarity, and the common characteristics of each subset are described by the high sequence similarity between the palmitoylation sites. For each given cysteine residue, its final score as a potential palmitoylation site is defined as the highest similarity score among all similarity scores against the above partitioned subsets. A detailed description of the algorithm can be found in the Supplementary Material. The sequence logo (WebLogo, Crooks *et al*., 2004)

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

(b)

Fig. 1. (**a**) Prediction page of CSS-Palm web server. (**b**) Prediction results of human (Q02224) and mouse CENP-E (Q6RT24). The default cut-off score is 2.6 (Sn 82.16% and Sp 83.17%).

of each of the subsets (Supplementary Fig. S1, S2 and S3) indicates that such strategy achieves better specificity over the whole set of collected palmitoylation sites (Supplementary Fig. S4). Its prediction performance on the curated dataset is highly encouraging with Jack-Knife sensitivity of 82.16% and specificity of 83.17%, respectively. To facilitate its application, we have developed an easy-to-use web server that is freely accessible from http://bioinformatics.lcd-ustc.org/css_palm/

APPLICATION

To explain how to use our CSS-Palm web server, we choose human Centromeric protein E (CENP-E, Q02224), a mitotic kinesin, as an example. Human CENP-E protein, as an important component of Kinetochore protein complexes, is localized in outerplate kinetochores during mitosis, attaches spindle microtubules to centromeres and orchestrates the fidelity of chromosomal segregation (Yao *et al*., 1997, 2000).

We first retrieved the primary sequences of both human and mouse CENP-E from ExPASy and pasted both the sequences in FASTA format into the textbox of CSS-Palm (http://bioinformatics.lcd-ustc.org/css_palm/prediction.php) and pressed the ‘Submit’ button to get the predicted palmitoylation sites by CSS-Palm (Fig.1a). As in Figure 1b, the prediction result suggests that both of the CENP-E proteins may be palmitoylated at a conserved pair of peptides in human (1477-EELKVAH**CC**LKEQEET-1492) and mouse (1378-EELNLAR**CC**LKEQENK-1393), which needs further experimental verification.

By pressing on 'here' in the button 'Download the TAB-deliminated data file from here' (as shown in Fig.1b), a user can get the TAB-deliminated result file that is convenient for further automatic processing.

ACKNOWLEDGEMENTS

The work of X. Yao and Y. Xue is supported by grants from 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) and National Institutes of Health (DK56292; CA92080). X. Yao. is a Georgia Cancer Coalition Eminent Research Scholar. The work of F. Zhou and Y Xu. is supported by the Georgia Cancer Coalition, National Science Foundation (NSF/DBI-0354771, NSF/ITR-IIS-0407204) and Department of Energy's Genomes to Life Program (http://doegenomestolife.org/) under the project 'Carbon Sequestration in *Synechochoccus* sp.: From Molecular Machines to Hierarchical Modeling'. Funding to pay the Open Access publication charges was provided by Chinese Natural Science Foundation and National Institutes of Health.

Conflict of Interest: none declared.

REFERENCES

Bijlmakers,M.J. and Marsh,M. (2003) The on-off story of protein palmitoylation. *Trends Cell Biol*., **13**, 32–42.

Crooks,G.E. *et al*. (2004) WebLogo: a sequence logo generator. *Genome Res*., **14**, 1188–1190.

el-Husseini Ael,D. and Bredt,D.S. (2002) Protein palmitoylation: a regulator of neuronal development and function. *Nat. Rev. Neurosci*., **3**, 791–802.

Kang,R. *et al*. (2004) Presynaptic trafficking of synaptotagmin I is regulated by protein palmitoylation. *J. Biol. Chem*., **279**, 50524–50536.

Kleuss,C. and Krause,E. (2003) Galpha(s) is palmitoylated at the N-terminal glycine. *EMBO J*., **22**, 826–832.

Linder,M.E. and Deschenes,R.J. (2003) New insights into the mechanisms of protein palmitoylation. *Biochemistry*, **42**, 4311–4320.

Schneider,A. *et al*. (2005) Palmitoylation is a sorting determinant for transport to the myelin membrane. *J. Cell. Sci*., **118**, 2415–2423.

Smotrys,J.E. and Linder,M.E. (2004) Palmitoylation of intracellular signaling proteins: regulation and function. *Annu. Rev. Biochem*., **73**, 559–587.

Sudo,Y. *et al*. (1992) Palmitoylation alters protein activity: blockade of G(o) stimulation by GAP-43. *EMBO J*., **11**, 2095–2102.

ten Brinke,A. *et al*. (2002) Palmitoylation and processing of the lipopeptide surfactant protein C. *Biochim. Biophys. Acta*, **1583**, 253–265.

Van Itallie,C.M. *et al*. (2005) Palmitoylation of claudins is required for efficient tight-junction localization. *J. Cell. Sci*., **118**, 1427–1436.

Yao,X. *et al*. (1997) The microtubule-dependent motor centromere-associated protein E (CENP-E) is an integral component of kinetochore corona fibers that link centromeres to spindle microtubules. *J. Cell. Biol*., **139**, 435–447.

Yao,X. *et al*. (2000) CENP-E forms a link between attachment of spindle microtubules to kinetochores and the mitotic checkpoint. *Nat. Cell Biol*., **2**, 484–491.

Zhou,F.F. *et al*. (2004) GPS: a novel group-based phosphorylation predicting and scoring method. *Biochem. Biophys. Res. Commun*., **325**, 1443–1448.