Methods 65 (2014) 156–164

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

PTP-central: A comprehensive resource of protein tyrosine phosphatases in eukaryotic genomes

Teri Hatzihristidis [a,b,1], Shaq Liu [c,1], Leszek Pryszcz [d], Andrew P. Hutchins [c], Toni Gabaldón [d], Michel L. Tremblay [a,b,*], Diego Miranda-Saavedra [c,*]

[a] Goodman Cancer Research Center, McGill University, 1160 Pine Avenue, Montreal H3A 1A3, QC, Canada
[b] Department of Biochemistry, McGill University, Montreal, QC, Canada
[c] Bioinformatics and Genomics Laboratory, World Premier International (WPI), Immunology Frontier Research Center (IFReC), Osaka University, 3-1 Yamadaoka, Suita, 565-0871 Osaka, Japan
[d] Centre for Genomic Regulation (CRG), UPF, Dr. Aiguader, 88, 08003 Barcelona, Spain

ARTICLE INFO

Article history:
Available online 31 July 2013

Keywords:
Tyrosine phosphatase
Tyrosine phosphorylation
Database
HMM
Sequence analysis
Signal transduction

ABSTRACT

Reversible tyrosine phosphorylation is a fundamental signaling mechanism controlling a diversity of cellular processes. Whereas protein tyrosine kinases have long been implicated in many diseases, aberrant protein tyrosine phosphatase (PTP) activity is also increasingly being associated with a wide spectrum of conditions. PTPs are now regarded as key regulators of biochemical processes instead of simple "off" switches operating in tyrosine kinase signaling pathways. Despite the central importance that PTPs play in the cell's biochemistry, the tyrosine phosphatomes of most species remain uncharted. Here we present a highly sensitive and specific sequence-based method for the automatic classification of PTPs. As proof of principle we re-annotated the human tyrosine phosphatome, and discovered four new PTP genes that had not been reported before. Our method and the predicted tyrosine phosphatomes of 65 eukaryotic genomes are accessible online through the user-friendly PTP-central resource (http://www.PTP-central.org/), where users can also submit their own sequences for prediction. PTP-central is a comprehensive and continually developing resource that currently integrates the predicted tyrosine phosphatomes with structural data and genetic association disease studies, as well as homology relationships. PTP-central thus fills an important void for the systematic study of PTPs, both in model organisms and from an evolutionary perspective.

1. Introduction

The reversible phosphorylation of proteins as carried out by protein kinases and phosphatases is one of the most widespread mechanisms for controlling cellular functions [1]: cells can quickly respond to intracellular and extracellular cues by altering the phosphorylation status of target proteins with the effect of increasing or decreasing their biological activity, modifying their sub-cellular localisation, or affecting protein stability and protein–protein interactions [2]. Reversible protein phosphorylation is thus a simple and flexible regulatory system that has been positively selected for in evolution as a general mechanism of cellular control. The first serine/threonine protein kinase ('Phosphorylase Kinase') was reported by Fischer and Krebs in1955 [3,4]. It took another 25 years to realize that v-Src (encoded by the Rous sarcoma virus) was a protein kinase [5] that phosphorylates tyrosine residues (a PTK [6]). On the other hand, the first serine/threonine protein phosphatases were discovered during the late 1970s and early 1980s [7], and the first tyrosine-specific phosphatase (PTP1B) in 1988 [8].

Tyrosine phosphorylation in metazoans is a fundamental signaling mechanism controlling a plethora of processes ranging from development to cellular shape and motility, transcriptional regulation, and proliferation vs. differentiation decisions. Not surprisingly, the abnormal regulation of tyrosine phosphorylation on target proteins is responsible for a wide spectrum of human conditions, including diabetes, obesity, cancer and inflammatory diseases. Many diseases have been associated with PTKs as well as protein tyrosine phosphatase (PTP) over-expression and deficiencies [9]. Historically, research on PTKs has advanced at a faster rate than investigations into PTPs. Not only were PTKs identified nearly a decade earlier than PTPs, but also the intrinsic difficulties of investigating the "disappearance" of a phosphate moiety as opposed to the appearance of the radioactive phosphate represented

* Corresponding authors. Addresses: Goodman Cancer Research Center, McGill University, 1160 Pine Avenue, Montreal H3A 1A3, QC, Canada. Fax: +1 514 398 6769 (M.L. Tremblay), Bioinformatics and Genomics Laboratory, World Premier International (WPI), Immunology Frontier Research Center (IFReC), Osaka University, 3-1 Yamadaoka, Suita, 565-0871 Osaka, Japan. Fax: +81 6 6879 4272 (D. Miranda-Saavedra).

E-mail addresses: michel.tremblay@mcgill.ca (M.L. Tremblay), diego@ifrec.osaka-u.ac.jp (D. Miranda-Saavedra).

[1] These authors contributed equally to this work.

1046-2023/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2013.07.031

a major burden for the PTP field. Yet the use of both embryonic gene targeting models, and the knockdown technologies of siRNA and shRNA have validated the importance of PTP activities in a great number of signaling pathways. Moreover, major advances have been made with the development of substrate trapping techniques where specific mutations of the PTP catalytic domain allow for the purification, detection and identification of their physiological substrates.

Whereas genome-wide catalogs of protein kinases have been available since the dawn of the genomics era (e.g. the *S. cerevisiae* kinome was published in 1997 [10]), the systematic identification of tyrosine phosphatomes has lagged behind. Previous work by Andersen et al. on classical (tyrosine-specific) PTPs has provided fundamental insights into the structure of the tyrosine phosphatase catalytic domain, a ~280 amino acid region containing 10 discrete and highly conserved motifs [11]. This was followed by the characterization of the tyrosine-specific PTPs in humans [12], and a three-way comparison including human, fly and worm PTPs [13]. These efforts were later extended by a thorough description of the entire human phosphatome (including tyrosine-specific and non tyrosine-specific PTPs) [14], and a recent review of the functions of human PTPs alongside the development of the PTP field from its origins [15]. Publicly available resources for PTPs include the 'PTP Database' (http://ptp.cshl.edu/) and PhosphoregDB (http://phosphoreg.imb.uq.edu.au/home.shtml). The PTP Database provides access to tyrosine-specific PTP sequences, multiple sequence alignments, phylogenetic trees, structures and links to human diseases. Although a fundamental resource for the PTP field, the sequences stored in the PTP Database are derived from computational analyses done nearly 10 years ago, with the further limitation that it is neither a continually updated or interactive database, nor is it linked to modern databases, but instead consists of downloadable files mostly in the form of text files, PDFs and Excel tables [13]. PhosphoregDB, on the other hand, is limited to providing information on mouse and human protein kinases and phosphatases, including protein sequences and interacting partners, some pathway information, tissue-specific distribution (drawn from the mouse GNF expression atlas), and a systematic subcellular localization screen done in HeLa cells [16]. The Human Phosphatase (HuPho) web portal was recently launched with a focus on human phosphatases only, including serine, threonine and tyrosine phosphatases, as a well as hydrolases [17], and information is provided on expression profiles, substrates, interactions and structures. However, since HuPho focuses on more than just PTPs, their proper identification or classification is often sacrificed (for instance, EyA phosphatases are categorized only as haloacid dehalogenases and not PTPs). Finally, an important resource that focuses on human phosphatases is the 'human DEPhOsphorylation Database' (DEPOD), where the authors, among other important analyses, integrate substrate information with cellular localization, coexpression data and pathway information to uncover important enzyme-substrate relationships [18].

Despite the central importance that PTPs play in the regulation of cellular biochemistry and the recent advances in high-throughput sequencing, the tyrosine phosphatomes of most genomes remain uncharacterized. We present a highly sensitive and specific sequence analysis method for the automatic classification of PTPs. Our method ('Y-Phosphatomer') relies on a collection of publicly available protein domain models to represent the diversity of the PTP sequence universe. Upon evaluation Y-Phosphatomer gave perfect coverage and class-level classification rates on characterized PTPs, suggesting its broad utility for the genome-wide characterization of tyrosine phosphatomes. We scanned 65 distinct eukaryotic genomes and describe their PTP complements and the evolutionary distribution of the various PTP classes. Finally we present PTP-central, a comprehensive resource of tyrosine phosphatomes in multiple genomes, where users can: (i) access our PTP sequence predictions and perform multiple sequence and phylogenetic analyses online; (ii) analyze the orthology and paralogy relationships of PTP sequences across 65 eukaryotic genomes; (iii) access an up-to-date database of PTP structures as well as a curated set of genetic studies where PTPs have been implicated; and (iv) also submit their own sequences for prediction.

2. Material and methods

2.1. Current classification of protein tyrosine phosphatases (PTPs)

The PTP superfamily has been divided into 4 distinct classes that differ both in their catalytic mechanisms and phosphatase catalytic domain sequences [14]. Class I are cysteine-based PTPs including the classical tyrosine-specific phosphatases (both receptor and cytoplasmic), and the dual-specificity phosphatases (DSPs, or VH1-like). DSPs represent the most promiscuous group of PTPs in terms of substrate specificity, with some members dephosphorylating mRNAs while other enzymes dephosphorylate lipids. Class II PTPs are a small but evolutionarily conserved class of PTPs with only one member in human (ACP1); they are also found in bacteria and are structurally related to bacterial arsenate reductases. Class III PTPs, like the Class I and II members, are also cysteine-based enzymes displaying specificity towards phosphotyrosine and phosphothreonine residues. The human enzymes (CDC25A, CDC25B and CDC25C) control cell cycle progression by dephosphorylating cyclin-dependent kinases. Despite sharing a cysteine-based catalytic mechanism, Class I, II, and III PTPs are believed to have evolved independently. A fourth class of PTPs displays an aspartic acid–based catalytic mechanism with dependence on a cation, and is represented by the developmentally important EyA ('Eyes Absent') genes, of which only 1 member is found in the fruit fly versus 4 genes in mouse and human.

2.2. Systematic prediction of PTPs

We present a general method for the automatic prediction and class-level classification of eukaryotic PTPs called Y-Phosphatomer. Y-Phosphatomer relies on a specific combination of publicly available protein domain models (mostly in the form of profile hidden Markov models, or HMMs) diagnostic for the various PTP families. Briefly, we built Y-Phosphatomer by characterizing the specific combination of protein domain signatures from the smallest number of protein domain databases that allows the identification and classification of a curated set of human PTPs into their correct classes, without cross-hitting other classes (Fig. 1).

2.2.1. Working data sets and identification of protein domain models specific to PTPs

In 2004 Alonso et al. reported the PTP complement of the human genome following comprehensive searches in public databases for genes encoding PTPs and PTP-like genes [14]. A total of 107 human genes were found to encode PTPs belonging to 4 distinct classes: Class I (receptor PTPs, cytoplasmic PTPs and dual-specificity phosphatases), Class II (low molecular weight phosphatases), Class III (CDC25), and aspartic acid (Asp)-based phosphatases (EyA homologues). We mapped these 107 human genes onto the Ensembl database (release 67), leading to the removal of MTMR15 and PTPRVP (a pseudogene, originally annotated as PTPRV). We also expanded gene PTPN20 into 3 distinct genes (PTPN20A, PTPN20B and PTPN20C), and merged DUSP13A and DUSP13B into the DUSP13 gene. This brings the updated list of human PTPs to 105 protein-coding genes, including 20 receptor PTPs,