PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A common structural scaffold in CTD phosphatases that supports distinct catalytic mechanisms

Tirso Pons,[1]* Ida Paramonov,[2] César Boullosa,[1] Kristina Ibáñez,[1] Ana M. Rojas,[2,3] and Alfonso Valencia[1]*

[1] Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

[2] Institute of Predictive and Personalized Medicine of Cancer (IMPPC), Badalona, Spain

[3] Life Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain

ABSTRACT

The phosphorylation and dephosphorylation of the carboxyl-terminal domain (CTD) of the largest RNA polymerase II (RNAPII) subunit is a critical regulatory checkpoint for transcription and mRNA processing. This CTD is unique to eukaryotic organisms and it contains multiple tandem-repeats with the consensus sequence Tyr^1–Ser^2–Pro^3–Thr^4–Ser^5–Pro^6–Ser^7. Traditionally, CTD phosphatases that use metal-ion-independent (cysteine-based) and metal-ion-assisted (aspartate-based) catalytic mechanisms have been considered to belong to two independent groups. However, using structural comparisons we have identified a common structural scaffold in these two groups of CTD phosphatases. This common scaffold accommodates different catalytic processes with the same substrate specificity, in this case phospho-serine/threonine residues flanked by prolines. Furthermore, this scaffold provides a structural connection between two groups of protein tyrosine phosphatases (PTPs): Cys-based (classes I, II, and III) and Asp-based (class IV) PTPs. Redundancy in catalytic mechanisms is not infrequent and may arise in specific biological settings. To better understand the activity of the CTD phosphatases, we combined our structural analyses with data on CTD phosphatase expression in different human and mouse tissues. The results suggest that aspartate- and cysteine-based CTD-dephosphorylation acts in concert during cellular stress, when high levels of reactive oxygen species can inhibit the nucleophilic function of the catalytic cysteine, as occurs in mental and neurodegenerative disorders like schizophrenia, Alzheimer's and Parkinson's diseases. Moreover, these findings have significant implications for the study of the RNAPII-CTD dephosphorylation in eukaryotes.

Proteins 2014; 82:103–118.

Key words: structure comparison; RNA polymerase II carboxyl-terminal domain; protein phosphatases; catalytic mechanism; HAD superfamily; prolyl *cis*/*trans* isomerization; oxidative stress; neurodegeneration; mental disorders; Alzheimer's disease; Parkinson's disease; schizophrenia.

INTRODUCTION

The nucleus is a highly dynamic cellular compartment in which protein phosphorylation plays a dominant regulatory role.[1] Protein kinases act in counterpoint with phosphatases to control the phosphorylation states of proteins that regulate virtually every aspect of eukaryotic biology. While kinases are highly abundant, distributed among many different families that generally exhibiting high substrate specificity,[2,3] the variety of phosphatases is more limited and they are more promiscuous in terms of substrate specificity. Protein phosphatase activity has been implicated in a multitude of diverse nuclear processes, including DNA replication and repair, chromosome

Additional Supporting Information may be found in the online version of this article.

Abbreviations: RNAPII, RNA polymerase II; CTD, carboxyl-terminal domain of the largest subunit of RNAPII; PTPs, protein tyrosine phosphatases; DSPs, dual-specificity PTPs; LMWPTP, low-molecular-weight PTPs; HAD, haloacid dehalogenase phosphatases; ROS, reactive oxygen species.

Grant sponsor: EU FP7 ASSET project; Grant number: 259348 (to AV); Grant sponsor: Obra Social laCaixa (to KI); Grant sponsor: FPI; Grant number: BES-2008-006332 (to CB).

*Correspondence to: Tirso Pons and Alfonso Valencia, Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre (CNIO), C/ Melchor Fernández Almagro 3, 28029 Madrid, Spain. E-mail: tpons@cnio.es; avalencia@cnio.es

Received 9 April 2013; Revised 28 June 2013; Accepted 12 July 2013
Published online 31 July 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24376

condensation, ribosome biogenesis, and chromatin remodeling, as well as in multiple signal-transduction pathways.[4] Aberrant regulation of protein phosphorylation significantly perturbs the associated signaling pathways and it has been linked to several disorders, including neurodegeneration, cancer, diabetes and obesity, as well as cognitive ageing.[4,5] Structural genomics projects that have focused on human protein phosphatases and those from biomedically relevant pathogens have provided important insight into the role these enzymes in both normal and pathological processes.[6,7]

In this study, we focused specifically on protein phosphatases that dephosphorylate the carboxyl-terminal domain (CTD) of the largest RNA polymerase II (RNAPII) subunit. These CTD phosphatases are known to control the transcription by RNAPII of protein-coding and non-coding genes in eukaryotes, and the phosphorylation state of CTD also influences chromatin modifications,[8–12] In addition, CTD phosphorylation is implicated in a variety of processes not directly related to transcription, such as mRNA export and the stress response to DNA damage.[12]

Primary structure of the RNAPII-CTD domain

The CTD domain contains multiple heptapeptide repeats bearing the consensus or "canonical" $Y^1S^2P^3T^4S^5P^6S^7$ sequence, as well as several non-consensus repeats. Notably, the amino acids of the consensus and non-consensus repeats may all be modified (e.g., serine, threonine, and tyrosine residues can be phosphorylated *in vivo*). The non-consensus repeat sequences are associated with modifications such as glycosylation, acetylation, methylation, sumoylation, and ubiquitination, which further expand the complexity of the "CTD code".[11] The combination of these distinct CTD modifications with the *cis–trans* isomerization of the proline residues plays key roles in different stages of the transcription cycle.[10,11] Phosphorylation/dephosphorylation is an important regulator of proteins and global dephosphorylation of the CTD facilitates the release of RNAPII from DNA, after which it recycles and again associates with promoters for the next cycle of transcription. Thus, eukaryotic organisms depend on finely tuned mechanisms to control CTD dephosphorylation, as the impairment of this process prevents RNAPII from terminating efficiently and probably impedes its subsequent assembly into the pre-initiation complex.[13]

Current classification of CTD phosphatases

Unlike kinases, there are three different mechanisms used by protein phosphatases to catalyze phosphoryl-transfer reaction: (i) aspartate-based catalysis; (ii) cysteine-based catalysis; and (iii) di-metal ions based catalysis.[14–16] The first two mechanisms involve a two-step reaction with the formation of a phosphoenzyme intermediate,[15,17] and these are used by CTD phosphatases, preferentially dephosphorylating CTD at Ser2 and Ser5 of the heptad repeat.[11,12,18,19] Accordingly, the catalytic mechanism employed is the main criterion used to classify CTD phosphatases, taking into account the identity of the nucleophile residues and the sequence motif at the active site.[8] Phosphatases that employ aspartate-based catalysis, henceforth referred to as the "Asp-CTD class" have been included in the HAD (haloacid dehalogenase)-like phosphatase family, one of the largest and most ubiquitous of the phosphotransferase families characterized to date.[20–22] The "Asp-CTD class" contains several proteins, including TFIIF-associating CTD-phosphatase-1 (FCP1), small CTD phosphatases (SCPs), and plant CTD phosphatase-like proteins (CPLs),[23,24] all of which contain the active-site sequence signature **D**X**D**X(T/V).[8]

CTD phosphatases that use cysteine as the nucleophile, henceforth referred to as the "Cys-CTD class", include the SSU72 protein,[13,18,19] and the human cell cycle regulatory enzyme CDC14B,[25,26] both of which contain the active-site sequence signature **C**X_5**R** (P-loop signature) of protein tyrosine phosphatases (PTPs). Interestingly, SSU72 is the only CTD phosphatase described to date that recognizes the phospho-Ser-Pro (pSer-Pro) motif when proline adopts a *cis* conformation[13,18,19,27] with all other CTD-binding proteins being *trans*-Pro specific.[10–12,28] The CTD phosphatase activity of the Cys-based RTR1 (yeast regulator of transcription),[29,30] and RPAP2 (human RNA Pol II-associated protein 2)[31,32] has been challenged in a recent study where no evidence of this type of activity was found.[33] Accordingly, we have excluded these proteins from our analyses.

Alternative classification systems exists for proteins that bind to the RNAPII-CTD based on the conservation of the consensus or "canonical" sequence of CTD repeats[34]: proteins can be classified as class 1 ("core functions recruited to CTD") or class 2 ("Co-evolution with CTD"). Class 1 proteins are found in a broad range of eukaryotes that have conserved and divergent heptapeptide sequences in their CTD, while the class 2 proteins identified in eukaryotes that have conserved heptapeptide sequences in their CTD. Accordingly, in this classification system, the SCP1 Asp-CTDs would belong to class 1 and FCP1 to class 2. However, Cys-CTD family phosphatases are not included in this classification.

Sequence and structural relationships between CTD phosphatases and other protein phosphatases

Based on sequence and structural data, in previous studies evolutionary relationships have only been proposed for protein phosphatases that share the same catalytic mechanism. Examples include the classification of the FCP1/SCP Asp-CTD family within the HAD superfamily of the aspartate-based phosphatases,[6,20,22,24,35]

the structural relationships between Cys-CTD SSU72 and low-molecular-weight PTPs (LMWPTP),[36–39] and those between Cys-CTD CDC14B (B-domain, residues 213–379) and "dual-specificity" PTPs (DSPs).[40]

To date, however, no structural similarities have been proposed for phosphatases that mediate distinct catalytic reactions,[14] and attempts to demonstrate convergent evolution of active sites have revealed no similarities between Asp-CTD and Cys-CTD.[41] Furthermore, although structural similarities exist in the "core" domain of the Rossmann fold of SSU72 and LMWPTP,[39,42] these Cys-CTD and Cys-based PTP (class II) enzymes lack the active-site sequence signature **D**X**D**X(T/V) and they are not classified in the HAD superfamily (reviewed in Ref. 20).

Distinct mechanisms of RNAPII-CTD dephosphorylation

It is unclear why both Asp- and Cys-based mechanisms are recruited to dephosphorylate the CTD. One potential explanation is based on the strong acidity of the Asp nucleophile. While oxidation of the cysteine at the active site in classical Cys-based PTPs abrogates its biochemical activity, the Asp-based mechanism is not affected by oxidation events.[4,43]

Mechanisms to prevent irreversible oxidation of the catalytic cysteine have been described[43,44] and protein tyrosine phosphatase 1B (PTP1B) for example forms a reversible cyclic sulfonamide by bonding the catalytic cysteine residue with the amide nitrogen of the neighboring serine residue.[45] The formation of disulfide bonds with another cysteine residue in the same protein, has also been described in DSPs, LMWPTP and CDC25 groups.[43] Briefly, following oxidation of the nucleophilic cysteine, a disulfide bond is formed with the neighboring cysteine that protects the enzymes from the irreversible inactivation that results from exposure to more oxidized species. This S–S bond can be readily reduced, which ensures the transient nature of the modification and returns the enzymes to its active form.[4]

Oxidation of the nucleophilic cysteine residue can be triggered by hydrogen peroxide, a reactive oxygen species (ROS). Reversible oxidation of the catalytic cysteine (e.g., by PTPs[4] and cysteine proteases[46]) has emerged as a putative mechanism to regulate the proteins activity in response to physiological cell stimulation with growth factors, engagement of antigen receptors or exposure to adverse conditions (e.g., UV irradiation).[43]

Isomerism of the prolines adjacent to phosphorylated serines has also emerged as a key mechanism that regulates of CTD dephosphorylation and successful progression through the transcription cycle.[11,28] Around 75% of the prolines in proteins are in *trans* conformation, while 10–30% are in *cis*.[28]. The transition between *cis* and *trans* conformations of prolines at positions 3 and 6 is catalyzed by phosphorylation-specific peptidyl–prolyl *cis/trans* isomerase PIN1 (ESS1 in yeast),[19,28,47,48], to date the only protein known to directly affect CTD dephosphorylation.

In summary, there is no clear explanation for the existence of both Cys- and Asp-based catalytic mechanisms, coupled to a third process (prolyl isomerization), in the context of CTD dephosphorylation.[10–12] Our results identify a common structural framework that is flexible enough to accommodate diverse catalytic reactions of a consensus sequence (pSer-Pro or pThr-Pro). We propose that this redundancy serves to compensate for the existence of an alternative mechanism that indirectly affects the conformation of the same consensus sequences (*cis/trans*-Pro). Based on this proposal, we analyzed CTD phosphatase expression data in several tissues from humans and mice. Our preliminary results suggest specific gene expression trends in neurodegenerative and mental disorders.

MATERIAL AND METHODS

Dataset of eukaryotic genomes

We investigated 20 complete or near complete (draft format) genomes based on their correspondence to relevant moments in eukaryotic evolution. These key events were represented by the inclusion of sequences from: ancient eukaryotes (red algae *Cyanidioschyzon merolae*); parasitic protozoa from Apicomplexa (*Plasmodium falciparum*) and Kinetoplastida (*Leishmania braziliensis* and *Trypanosoma brucei*); a free-living ciliate (*Paramecium tetraurelia*); the placozoan *Trichoplax adhaerens*, the simplest known free-living animal possessing only four somatic cell types[49]; the anthozoan cnidarian *Nematostella vectensis*, which has a tissue grade of organization[50]; and different chordate lineages represented by the ascidian *Ciona intestinalis* and the lancelet *Branchiostoma floridae*. In addition, the *Amoebozoa* genomes of *Dictyostelium discoideum* and *Entamoeba histolytica* represent an ancient split in the Conosa lineage, which is subdivided into Mycetozoa and Archamoebae.[51] Recent proteome-based phylogeny analyses have confirmed that Amoebozoa diverged from the animal–fungal lineage after the plant–animal split.[52] The *D. discoideum* genome is noteworthy in that the proteins it encodes are commonly found in fungi, plants, and animals.[53]

Ten of the selected genomes (*Giardia lamblia*, *C. merolae*, *T. brucei*, *P. falciparum*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Homo sapiens*) have been analyzed previously,[34] while the unexplored genomes included that of *P. tetraurelia*, *L. braziliensis*, *E. histolytica*, *D. discoideum*, *T. adhaerens*, *N. vectensis*, *B. floridae*, *C. intestinalis*, *Xenopus tropicalis*, and *Mus musculus*.

While well-annotated genomes were available for the majority of the species included in the analysis, in some

cases the genome was poorly annotated (e.g., *P. falciparum*) or was only available in a draft format (e.g., *G. lamblia*, *T. adhaerens, C. intestinalis*, *L. braziliensis,* and *X. tropicalis*) according to the Genomes Online Database (GOLD) at http://www.genomesonline.org/cgi-bin/GOLD/index.cgi.[54]

Sequence database searches

PSI-BLAST searches[55] of the NCBI database were performed using the non-redundant database and default parameters (*E*-value inclusion threshold of 0.005). Additional BLAST searches at organism-specific sites were used to help identify potentially divergent orthologs reducing the size of target databases and optimizing search parameters for a given genome, including: *C. merolae* (http://merolae.biol.s.u-tokyo.ac.jp/blast/blast.html); *D. discoideum* and *E. histolytica* (http://amoebadb.org/amoeba/); *P. tetraurelia* (http://paramecium.cgm.cnrs-gif.fr/cgi/tool/blast); *C. intestinalis* (http://genome.jgi-psf.org/Cioin2/Cioin2.info.html); kinetoplastids (Leishmania and Trypanosoma; http://tritrypdb.org/tritrypdb/); and *P. falciparum* (http://www.broadinstitute.org/annotation/genome/plasmodium_falciparum_spp/Blast.html).

When an entire gene was absent from a particular species or a particular domain was reported, we executed translated searches against the nucleic acid sequences. Each unannotated sequence recovered was also queried against the PFAM database[56] to ensure that it matched the protein family as closely as possible. Both well-annotated and unannotated protein sequences should have the same significant PFAM-A matches. Orthology was verified by querying the repositories OrthoMCL-DB (http://www.orthomcl.org/)[57] and Roundup 2.0 (http://roundup.hms.harvard.edu - Divergence = 0.5; BLAST *E*-value = 1E-20; Distance Lower/Upper Limit = none).[58] These repositories complement each other, both in the species included as in the implemented methods. OrthoMCL-DB implements an all-against-all BLAST search of each species' proteome, followed by normalization of inter-species differences, and a Markov clustering strategy.[57] Roundup 2.0 compute gene orthologs using a reciprocal smallest distance (RSD) algorithm, which improves the sensitivity of reciprocal best blast hits by considering global alignment and maximum likelihood evolutionary distance between sequences.[58] OrthoMCL-DB includes 150 genomes mainly from eukaryotes, while Roundup 2.0 covers gene orthologs for over 1800 genomes including 226 Eukaryote.

Crystal structure dataset and structural comparison methods

The crystal structures were retrieved from the Protein Data Bank (http://www.rcsb.org/pdb/) for: human SCP1 (PDB ID: 2ght),[35] SCP2 (PDB ID: 2q5e), SCP3 (PDB ID: 2hhl),[6] SSU72 (PDB ID: 3o2q),[18] CDC14B (PDB ID: 1ohe)[40]; *Drosophila melanogaster* SSU72 (PDB ID: 3omx)[39]; and *Schizosaccharomyces pombe* FCP1 (PDB ID: 3ef1).[24] Pairwise and multiple superimposition of CTD phosphatases were carried out using the CE,[59] DALI,[60] and PDBeFold[61] methods.

Gene expression datasets and data analyses

The gene expression patterns for Asp-CTD and Cys-CTD were analyzed in a large collection of tissues from human and mouse, assessing expression versus non-expression, using the Barcode database (http://rafalab.jhsph.edu/barcode).[62] We complemented this analysis using the annotations extracted from COXPRESdb (http://coxpresdb.jp), which identifies co-expressed gene sets.[63] Barcode includes microarray data for normal (97) and tumoral (24) human tissue, as well as normal mouse cells (89), whereas COXPRESdb uses publicly available GeneChip data from the GEO database.[64]

Gene expression was assessed by downloading the normalized and preprocessed data for 84 human tissues from the BioGPS portal[65] (also available from the GEO database: accession number GSE1133[66]), and the normalized data for 91 mouse tissues from the GEO database with the accession number GSE10246.[67] A relative expression value was calculated for each gene in each sample using the percent rank, defined as the rank of a value in a dataset as a percentage of the dataset. Thus, non-expressed genes are assigned a value of 0%, while 100% reflects those with the strongest expression. In general, a value of $<40\%$ indicates that a gene is not expressed (background noise) and strongly expressed genes have a percent rank of $>80\%$.

To study neurodegenerative and mental disorders, the gene expression profiles from patients and control subjects were retrieved from the GEO database for Alzheimer's disease (GSE5281), Parkinson's disease (GSE20292, GSE8397, and GSE7621), and schizophrenia (GSE4036). We also used the schizophrenia datasets from the Stanley Medical Research Institute Online Genomics Database (SMRIDB): AltarA, AltarC, Bahn, Kato, Kemether, and Laeng) (https://www.stanleygenomics.org). The Alzheimer's gene expression dataset pertains to 6 different brain regions (hippocampus; medial-temporal, and superior-frontal gyrus; posterior-cingulate; and primary-visual and entorhinal cortex).[68] The Parkinson's disease profiles correspond to substantia nigra[69–71] and the schizophrenia dataset pertains to the cerebellum, cortex, hippocampus, and thalamus.[72]

All these datasets were analyzed using a variety of tools: (i) Affymetrix package for R (http://www.R-project.org/); (ii) the fRMA (frozen Robust Multiarray Analysis) method[73]; and (iii) the Limma program.[74] The *post hoc* analysis produced a list of overexpressed (Up) and underexpressed (Down) genes after multiple testing corrections (FDR < 0.05).

RESULTS

Evolutionary analysis of Asp-CTD and Cys-CTD

We investigated the protein domains present not only in Asp-CTD but also in Cys-CTD proteins across diverse taxa that represent all six major eukaryotic supergroups (Fig. 1). We are aware that this list of CTD phosphatases is necessarily incomplete as additional CTD–protein interactions continue to be identified. Nevertheless, we feel that this is a representative list to be able to perform evolutionary studies given the current state-of-the-art.

Of the Asp-CTD and Cys-CTD proteins analyzed, only the SCPs (Asp-CTD class) were detected universally in all 20 eukaryotic genomes investigated (Fig. 1). Human CDC14B orthologues (Cys-CTD class) were identified in all species analyzed with the exception of plants (reviewed in Ref. 75 and 14), *P. falciparum* and *E. histolytica*. As CDC14B is involved in both transcription and cell cycle regulation, and it is present in all other complete and draft genomes examined, we speculated that CDC14B could be included in the Class 1 "core functions recruited to CTD," as well as the Asp-CTD SCP group.[34]

Other Asp-CTD (FCP1) and Cys-CTD (SSU72) proteins were not detected in one or several highly divergent organisms (e.g., the cnidarian *N. vectensis* and the free-living ciliate *P. tetraurelia*) or in human parasites (*e.g.*, *E. histolytica*, *L. braziliensis*, *T. brucei*, *P. falciparum*, and *G. lamblia*) with a complete or draft genome (as indicated by dashes in Figure 1). With the exception of the cnidarian *N. vectensis*, all of the species mentioned above

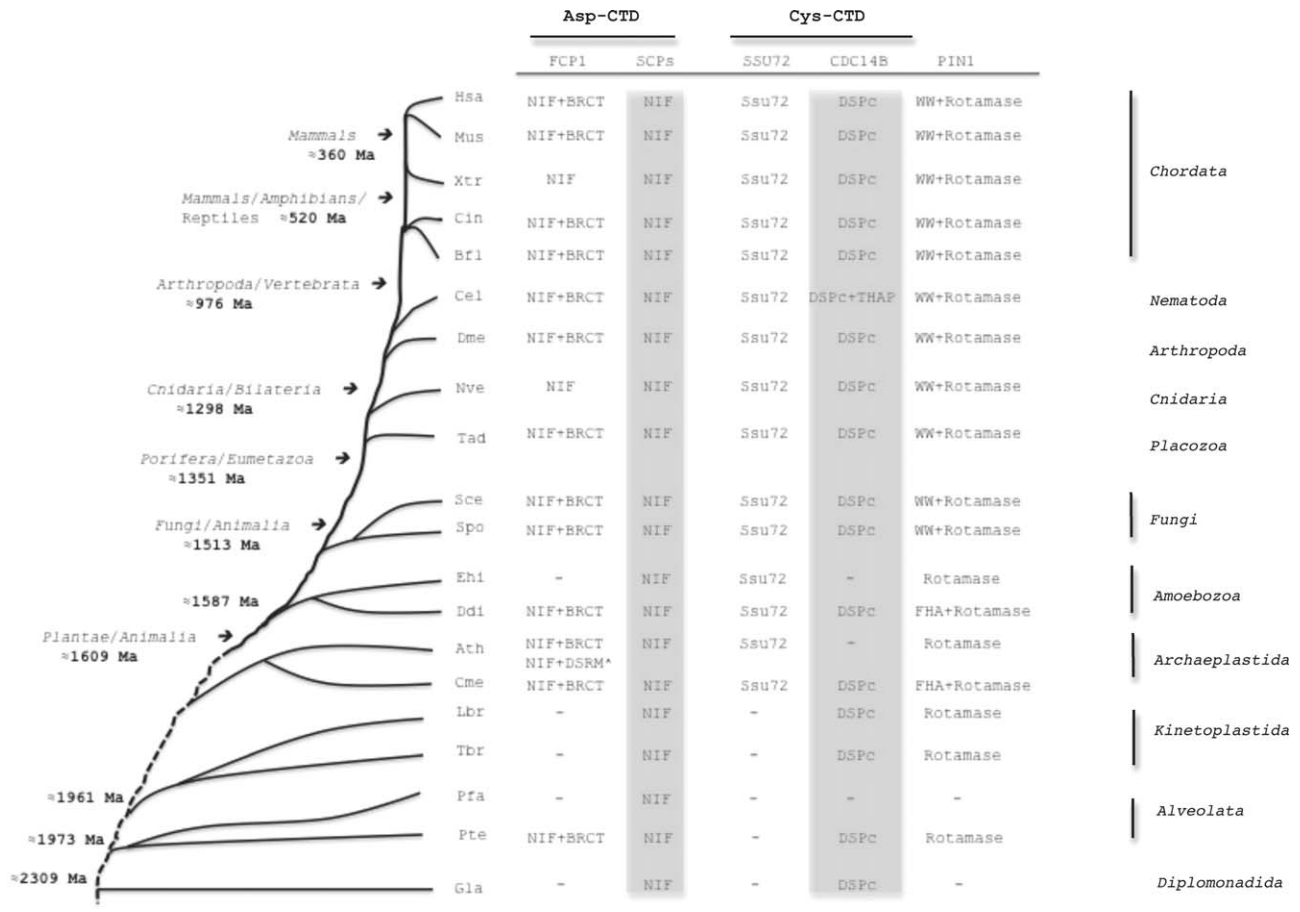


Figure 1

Orthologs of human Asp-CTD, Cys-CTD and peptidyl-prolyl *cis/trans* isomerase PIN1 in 20 proteomes. Columns represent the PFAM[56] domain composition of the orthologous proteins. Dashes indicate that no orthologous protein with an above-threshold score and similar domain architecture was retrieved in the database searches. The numbers in bold indicate millions of years according to a recently revised scale[114] and the time line is an approximate scale for the purposes of illustration. The arrows point to important splits that occurred in the course of evolution and the gray areas indicate "Class 1" CTD-related proteins according to Ref. 34. Gla: *Giardia lamblia*; Cme: *Cyanidioschyzon merolae*; Tbr: *Trypanosoma brucei*; Pfa: *Plasmodium falciparum*; Ath: *Arabidopsis thaliana*; Sce: *Saccharomyces cerevisiae*; Spo: *Schizosaccharomyces pombe*; Cel: *Caenorhabditis elegans*; Dme: *Drosophila melanogaster*; Hsa: *Homo sapiens*; Pte: *Paramecium tetraurelia*; Lbr: *Leishmania braziliensis*; Ehi: *Entamoeba histolytica*; Ddi: *Dictyostelium discoideum*; Tad: *Trichoplax adhaerens*; Nve: *Nematostella vectensis*; Bfl: *Brachiostoma floridae*; Cin: *Ciona intestinalis*; Xtr: *Xenopus tropicalis*; Mus: *Mus musculus*.

have a "non-conserved CTD" (Supporting Information Fig. S1). Thus, that in addition to FCP1,[34] we propose that the SSU72 protein may be included in Class 2 "Co-evolution with CTD."

Possible explanations for the absence of certain Asp-CTDs and/or Cys-CTDs from the group of complete and well-annotated genomes include a high level of sequence divergence from other eukaryotes or complete gene loss. Remarkably, SSU72 homologues are absent in Kinetoplastida (*L. braziliensis* and *T. brucei*), Alveolata (*P. falciparum* and *P. tetraurelia*), and Diplomonadida (*G. lamblia*; Fig. 1). All these organisms have a "non-conserved CTD" and the majority is parasitic protozoa. In our PSI-BLAST searches, we found no sequences with significant similarity to Class 2 proteins (FCP1 and SSU72) in the parasitic protozoan *G. lamblia*. Indeed, this human parasite is considered to be among the earliest-diverging eukaryotic lineages, displaying putative transitional stages in its transcriptional machinery that are interpreted as "prokaryotic properties."[76,77]

Perhaps unsurprisingly, some eukaryotic organisms have discarded one or two of the four CTD phosphatases identified to date, probably due to functional redundancy between Asp-CTD and Cys-CTD. A similar situation was described for many prokaryotic protein-serine/threonine/tyrosine phosphatases.[78]

The Asp-CTD and Cys-CTD domain architecture is related to their functional specificity

In both Asp-CTD and Cys-CTD we observed different loss and gain events indicative of functional diversification (Fig. 1). For example, Asp-CTD FCP1 in both *Cnidaria* (Nematostella) and *Chordata* (Xenopus) only possesses the catalytic NIF-domain (PFAM code: PF03031) instead of the NIF and BRCT (PFAM code: PF12738) combination. Similarly, the CPL1 and CPL2 proteins of the green plant *A. thaliana* that are implicated in jasmonic acid biosynthesis, as well as stress and auxin responses,[79,80] possess one NIF-domain and an additional double-stranded RNA binding motif (DSRM; PFAM code: PF00035). Both these proteins lack the BRCT domain that is commonly found in the FCP1 orthologs and that interacts with phosphorylated protein targets containing the pSer-X-X-Phe sequence, where X indicates any residue.[81] However, BCRT is not essential for FCP1 function *in vivo* or *in vitro*[24] and indeed the Class 1 Asp-CTD SCP1 does not contain a BRCT domain.

A third example is the Cys-CTD CDC14B from *C. elegans*. This *C. elegans* CDC14 protein was seen to have gained a putative DNA-binding domain (THAP, PFAM code: PF05485) in addition to containing the conserved DSPc catalytic domain (PFAM code: PF00782). Studies of CDC14 in different eukaryote species have revealed divergent functions and cellular locations (reviewed in Ref. 82). *C. elegans* CDC14 localizes to the spindle and centrosomes during mitosis and to the cytoplasm during interphase.[83,84] Unlike nematode CDC14, the human CDC14B ortholog is primarily nucleolar, as in yeast, although it is also detected in nuclear filaments and at the spindle.[85–87]

Some of the phosphatases discussed here evolve by means of domain rearrangements. These adaptation events were observed in both CTD phosphatases and in the prolyl-isomerase PIN1 (as described below), and they may indicate some correlation in the variation between RNAPII-CTD and the transcription machinery.[88,89]

Loss and gain events in prolyl-isomerase PIN1

Prolyl-isomerase PIN1 directly influences the mechanism of CTD dephosphorylation.[11,28] We observed different loss and gain events in the PIN1 family that are suggestive of functional adaptation similar to that observed in the Asp-CTD FCP1 and Cys-CTD CDC14B families (Fig. 1). The domain architecture of PIN1 orthologues from Fungi, Placozoa, Cnidaria, Arthropoda, Nematoda and Chordata consists of Rotamase (PFAM code: PF00639) and WW (PFAM code: PF00397) domains. However, PIN1 orthologues from Amoebozoa (*Dictyostelium*) and Archaeplastida (*Cyanidioschyzon*) contain a forkhead-associated domain (FHA domain; PFAM code: PF00498) instead of the WW domain. Remarkably, the WW domain binds proline-rich peptide motifs,[90–92] while the FHA domain recognizes a phosphopeptide motif found in a wide range of proteins (e.g., kinases, phosphatases, transcription factors, RNA-binding proteins, and metabolic enzymes).[93] The FHA domain displays specificity for phosphothreonine-containing motifs but it also recognizes phosphotyrosine with relatively high affinity.[94] To date, genes encoding FHA-containing proteins have been identified in eubacterial and eukaryotic genomes, as documented in the PFAM database.[56] Therefore, the domain architecture (FHA + Rotamase) observed here in PIN1 from the divergent eukaryotes *Dictyostelium* (Amoebozoa) and *Cyanidioschyzon* (Archaeplastida) is not unusual.

Another event that is suggestive of functional adaptation of PIN1 was observed in Archaeplastida (Plants), Amoebozoa (*Entamoeba*), Kinetoplastida (*Leishmania* and *Trypanosoma*), and the free-living ciliate *Paramecium* (Alveolata). This event involves the loss of the WW and FHA accessory domains in PIN1 homologues, although they retain the Rotamase domain. The PIN1 of *A. thaliana* is the best characterized of these species and it retains pSer/Thr prolyl *cis*/*trans* isomerization specificity.[95,96] One explanation for this putative functional adaptation is that the WW domain binds specific pSer/Thr-Pro sites in the substrate and it acts as a binding module that places the Rotamase domain close to its substrate, increasing the local concentration of this enzyme.[97] A similar explanation for the functional

implications of domains loss could be applicable to the other "Class 2" enzymes mentioned above (e.g., Asp-CTD FCP1, which lacks a BRCT domain).

Identification of a common structural scaffold for the Asp-CTD and Cys-CTD classes

To study the structural relationship between the Asp-CTD and Cys-CTD classes, we analyzed the crystal structures available for human SCP1 (PDB ID: 2ght),[35] SCP2 (PDB ID: 2q5e) and SCP3 (PDB ID: 2hhl),[6] SSU72 (PDB ID: 3o2q),[18] CDC14B (PDB ID: 1ohe),[40] *Drosophila melanogaster* SSU72 (PDB ID: 3omx),[39] and *Schizosaccharomyces pombe* FCP1 (PDB ID: 3ef1).[24]

The pairwise structure-superposition of human SCP1 (Asp-CTD), yeast FCP1 (Asp-CTD), and human and fly SSU72 (Cys-CTD) revealed distant but significant relationships. Human SCP1 and yeast FCP1 share 25% sequence identity, whereas human and fly SSU72 share 60% sequence identity. Searching for structural similarities with DALI,[60] we detected a low but significant match between SCP1 and SSU72 (Z-score = 3.3, rmsd = 3.3 Å over 97 aligned residues with 12% identity for human SSU72; Z-score = 3.5, rmsd = 3.4 Å over 96 aligned residues with 9% identity for fly SSU72), and between FCP1 and SSU72 (Z-score = 3.9, rmsd = 3.7Å over 104 aligned residues with 8% identity for human SSU72; Z-score = 3.9, rmsd = 3.3 Å over 105 aligned residues with 11% identity for fly SSU72). A DALI Z-score < 2 indicates a spurious match. Additional superimpositions were performed with the X-ray structure of human CDC14B to determine the significance of the resemblance between Asp-CTD and Cys-CTD phosphatases. While DALI superposition was less significant for CDC14B matched to SCP1 (Z-score = 0.3, rmsd = 3.8Å over 61 aligned residues with 5% identity) or FCP1 (Z-score = 1.3, rmsd = 2.6 Å over 51 aligned residues with 6% identity), a similar arrangement of core parallel β-stands remained. Similar results were obtained when applying the CE method (Supporting Information Table S2).

Based on these findings, we performed a detailed structural comparison between human SCP1, SCP2 and SCP3, yeast FCP1, human and fly SSU72, and human CDC14B. A multiple structural alignment of secondary structures with PDBeFold[61] revealed a common structural scaffold for these CTD phosphatases. This scaffold consists of four parallel and twisted β-stands flanked by α-helices (Fig. 2 and Supporting Information Fig. S2), which resembles the four-stranded core in the Rossmann fold of the HAD superfamily. Furthermore, the scaffold contains the conserved sequence motifs I, II, and IV described in the active site of the HAD superfamily. The overall low degree of sequence identity is offset by the striking concordance of catalytic aspartic acid and cysteine residues, and the presence of a central core of parallel β-stands.

The secondary structure elements are not always consecutive in the polypeptide chain as seen in human CDC14B, in which the first β-stand and α-helix are located in the C-terminus of the polypeptide chain (Fig. 2, panels B and C). This superposition places both nucleophile aspartate and cysteine residues in the same topographic location (panel A in Figs. 2 and 3). Moreover, the central axis of the first α-helix (α1 for SSU72 and α4B for CDC14B) in the Cys-CTD phosphatases is slightly angled as compared with the equivalent α-helix in Asp-CTD phosphatases. This small inclination positions the arginine residue (transition-state stabilizer) within structural proximity of the cysteine nucleophile.

Another important observation regarding the common structural scaffold of these CTD phosphatases is that other functional residues that have been identified by site-directed mutagenesis, residues known to be important for substrate binding or metal-ion coordination, also map to the C-terminal region of the four parallel and twisted β-stands of the core (Fig. 2, panel C). Moreover, the existence of a common structural scaffold in CTD phosphatases is in agreement with previous findings demonstrating that Asp-based catalysis of the FCP1/SCP family resembles the two-step reaction mechanism of Cys-based phosphatases, but using aspartic acid as a nucleophile.[17]

Despite the common structural scaffold, we observed differences among the phosphatases studied in the CTD recognition pocket. These differences arose from inserts (termed caps) at the end of β-strands in the common scaffold (Fig. 2, panel C; and Fig. 3 panel B) similar to those described in the core of the Rossmann fold of the HAD superfamily. Diversification of the cap module is a major determinant of evolution of this superfamily.[22]

Asp-CTD and Cys-CTD exhibit no apparent differences in genes expression

We investigated whether the distinct catalytic mechanisms are reflected by the pattern of expression of these genes, initially analyzing the overall expression profiles of the Asp-CTD and Cys-CTD genes in a large number of human and mouse tissues (Barcode), and subsequently analyzing their relative expression. The Barcode analyses revealed that Asp-CTD and Cys-CTD genes are expressed in almost all human and mouse tissues (Supporting Information Fig. S3), with no clear differences in the levels of expression in any given normal tissue type (Fig. 4 and Supporting Information Tables S3 and S4). The most striking difference observed between human and mouse was observed for FCP1, which was expressed to a variable degree in mouse tissues (Fig. 4 panel A) but was silenced or expressed very weakly (<50%) in nearly all human tissues, except for blood, testis and lung (Fig. 4,

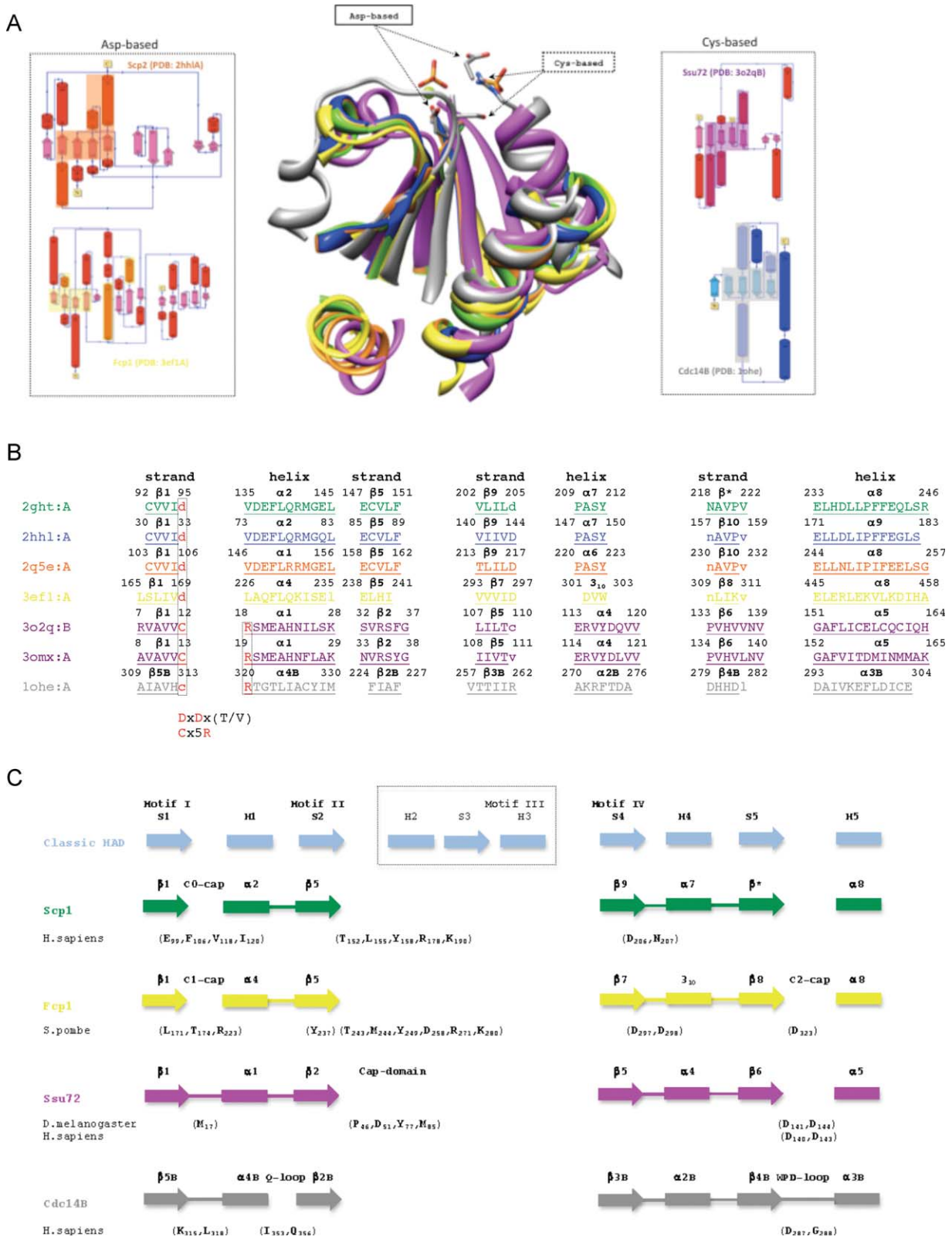

A
Asp-based
Scp2 (PDB: 2hhlA)
Fcp1 (PDB: 3ef1A)
Asp-based
Cys-based
Cys-based
Ssu72 (PDB: 3o2qB)
Cdc14B (PDB: 1ohe)
B
strand
helix
strand
strand
helix
strand
helix
2ght:A CVVId VDEFLQRMGEL ECVLF VLILd PASY NAVPV ELHDLLPFFEQLSR
2hhl:A CVVId VDEFLQRMGQL ECVLF VIIVD PASY nAVPv ELLDLIPFFEGLS
2q5e:A CVVId VDEFLRRMGEL ECVLF TLILD PASY nAVPv ELLNLIPIFEELSG
3ef1:A LSLIVd LAQFLQKISE1 ELHI VVVID DVW nLIKv ELERLEKVLKDIHA
3o2q:B RVAVVC RSMEAHNILSK SVRSFG LILTc ERVYDQVV PVHVVNV GAFLICELCQCIQH
3omx:A AVAVVC RSMEAHNFLAK NVRSYG IIVTv ERVYDLVV PVHVLNV GAFVITDMINMMAK
1ohe:A AIAVHc RTGTLIACYIM FIAF VTTIIR AKRFTDA DHHDl DAIVKEFLDICE
DxDx(T/V)
Cx5R
C
Motif I
Motif II
Motif III
Motif IV
Classic HAD
Scp1
H.sapiens
Fcp1
S.pombe
Ssu72
D.melanogaster
H.sapiens
Cap-domain
Cdc14B
H.sapiens
Q-loop
WPD-loop

Figure 2

panel B). Interestingly, knockdown of this gene is deleterious to cells,[98] while its overexpression provokes cell-cycle arrest,[99] suggesting it fulfills a role in fine tuning cell-cycle regulation, although no periodicity has been described for this gene.[100]

The expression of human phosphatases revealed no trends towards differential expression in normal *versus* tumoral tissues (Supporting Information Table S4), indicating that both catalytic mechanisms are represented equally in the samples. The strongest expression (percent rank >90%) was observed in blood, lung and thyroid tissues. The annotations in COXPRESdb, which identifies co-regulated gene sets, suggest that the human SCP1 and SCP2 Asp-CTDs are the only highly co-expressed gene pairs with a mutual rank (MR) value of 12.7 (significant values are <50[63,101]), consistent with our expression analyses (Supporting Information Table S4).

Considering all the limitations associated with using whole tissues, overall these results suggest that neither Asp-CTD nor Cys-CTD-based catalysis is subject to strong expression selection, as representative genes from both mechanisms were expressed similarly in almost all human and mouse tissues analyzed. These findings support the view that the two mechanisms co-exist and maintain robust transcription.

Dephosphorylation mechanisms are affected by oxidative stress, which may have pathological consequences in certain tissues

The human brain is responsible for about 20% of the oxygen consumption of the body and it is therefore exposed to high levels of ROS, which can result in oxidation of the catalytic cysteine in proteins.[102,103] We propose a biological model that integrates both observations with structural implications that affect key residues of the core heptapeptide in RNAPII-CTD, the redundancy of the dual Asp- and Cys-based catalytic mechanisms and the isomerization of adjacent prolines (Fig. 5). We believe that this hypothetical model is potentially important in the context of human neurodegenerative and mental diseases (e.g., schizophrenia or Alzheimer's and Parkinson's diseases). Indeed, the role of oxidative stress (or high ROS levels) in the pathogenesis of Alzheimer's and Parkinson's diseases has been well documented.[44,102,104]

According to the hypothetical model (Fig. 5), the RNAPII-CTD dephosphorylation of $pSer^2$-Pro^3 and $pSer^5$-Pro^6 requires the action of Asp-CTDs, Cys-CTDs (*trans*-acting), Cys-CTD SSU72 (*cis*-acting), and the proline isomerase PIN1. The role of PIN1 in RNAPII-CTD dephosphorylation by Asp-CTD SCP1 and Cys-CTD SSU72 has recently been described.[28] In normal conditions, the dephosphorylation of the Asp/Cys-CTDs *trans* conformations is reinforced by PIN1-mediated isomerization of *cis*-prolines, while Cys-CTD SSU72 simultaneously acts directly on *cis* conformations (Fig. 5, panel A).

In conditions in which high levels of ROS inhibit the Cys-based catalytic mechanism, only Asp-CTDs can dephosphorylate the RNAPII-CTD and the increase in PIN1 expression might decrease the pool of *trans*-proline, thereby favoring Asp-CTD-mediated dephosphorylation (Fig. 5, panel B). Interestingly, PIN1 is strongly expressed in healthy human brain tissue from different regions (~90% in whole brain) compared with other tissues (Supporting Information Table S4), suggesting an important role of this mechanism in ensuring proper phosphorylation/dephosphorylation and compensating for the possible hindrance to Cys-based catalytic mechanisms.

Our comparative analysis of the expression of CTD phosphatases in samples from Alzheimer's, Parkinson's and schizophrenia patients,[68] indicated that the expression of Cys-CTDs (SSU72 and CDC14B) appears to be silenced in these patients whereas Asp-CTD expression is upregulated (Fig. 5 panel D), suggesting oxidative damage to the Cys-CTDs catalytic mechanism. Moreover, we found that PIN1 expression was downregulated in these patients. These observations suggest complete deregulation of the RNAPII-CTD dephosphorylation mechanism, whereby: (i) the poor (or null) availability of Cys-CTD SSU72 produces an increase in phosphorylated serine with adjacent proline residues in the *cis* conformation; and (ii) the low levels of PIN1 expression affect the isomerization process (Fig. 5, panel C).

Taken together, these findings underscore the importance of a finely tuned mechanism for the dephosphorylation of the RNAPII-CTD. Such a mechanism could alleviate potential unscheduled phosphorylation events particularly in situations of oxidative stress, as occurs in human neurodegenerative and mental disorders.

Figure 2

Topology diagrams of CTD phosphatases and the structural scaffold resembling a four-stranded core of the Rossmann fold. Strands are shown as arrows with the arrowhead indicating the C-terminal end. *Panel A.* Structural representation of human SCP1 (green), SCP2 (blue), SCP3 (orange), SSU72 (magenta), CDC14B (gray), and *S. pombe* FCP1 (yellow). The catalytic residues (aspartate, cysteine, arginine), phosphate groups and magnesium ions are also represented (the figure was produced with Chimera software[115]). *Panel B.* Amino acid sequence of the equivalent secondary structure elements for human SCP1 (PDB ID: 2ght, chain A), SCP2 (PDB ID: 2hhl, chain A), SCP3 (PDB ID: 2q5e, chain A), SSU72 (PDB ID: 3o2q, chain B), CDC14B (PDB ID: 1ohe, chain A), and *S. pombe* FCP1 (PDB ID: 3ef1, chain A). Upper case and underlined letters indicate amino acid residues in the secondary structure elements, while the numbers correspond to the first and last positions of the structural elements. Active-site residues are boxed and highlighted in red. The nomenclature and atom number is in accordance with PDB annotations: β*, Beta bridge 2 (according to DSSP). *Panel C.* Representation of the functional residues in the common structural scaffold. The functional residues for substrate recognition and metal-ion coordination (in parenthesis), and the secondary structure elements (bar and arrows indicate helices and strands, respectively) are annotated. Suffix B was added to indicate the B-domain of CDC14B. For simplicity, the SCP2 and SCP3 proteins have been omitted.

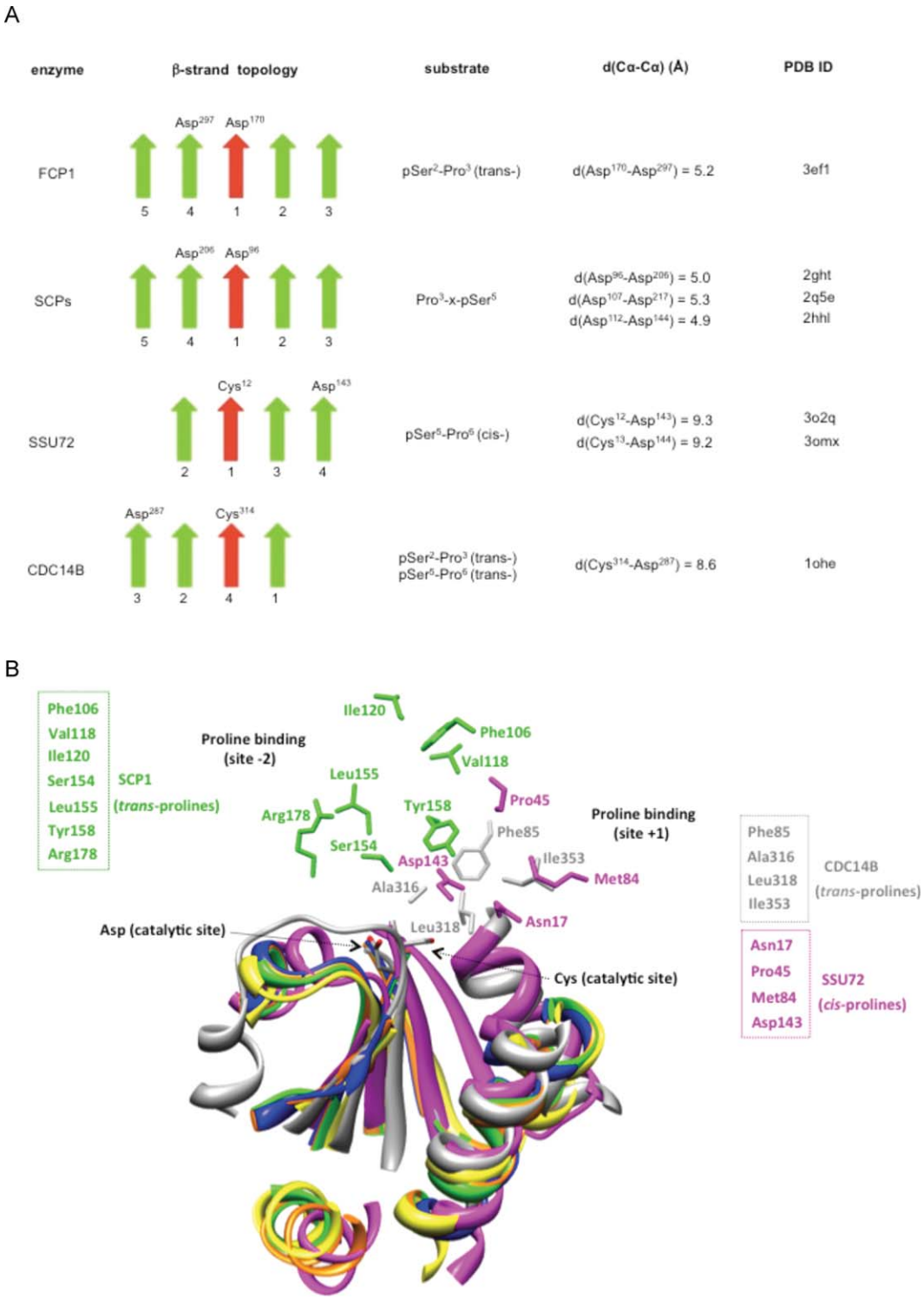


Figure 3

Summary of the β-strand topology, substrate specificity and spatial disposition of catalytic residues in Asp-CTD and Cys-CTD. *Panel A.* The β-strand order in the central parallel β-sheet of CTD phosphatases is indicated by numerals. The β-strand that accommodates either the aspartate or cysteine catalytic residues is shown in red. The alpha-carbon distances between catalytic residues, and the substrate specificity for Asp-CTD and Cys-CTD are also shown. *Panel B.* Representation of Pro-binding cavities in Asp-CTD and Cys-CTD. The following crystal structures for enzyme-substrate complexes were used: CDC14B (PDB ID: 1ohe), SSU72 (PDB ID: 4h3k), and SCP1 (PDB ID: 2ght, 2ghq). The side chain of residues in the vicinity of the proline substrates (heavy-atom inter-atomic distances $\leq$4 Å), and the side chains of the catalytic Asp and Cys residues are shown. Site −2 refers to Pro3 (two residues N-terminal to the phospho-serine site) in the single and dual phosphorylated CTD peptides (SCP1 complexes). Site +1: refers to Pro6 (a residue immediately adjacent to the phospho-serine site) in the phosphorylated 7 amino acid CTD peptide (SSU72 complex), and the unique proline residue in a phosphorylated 3 amino acid peptide ligand (CDC14B complex).

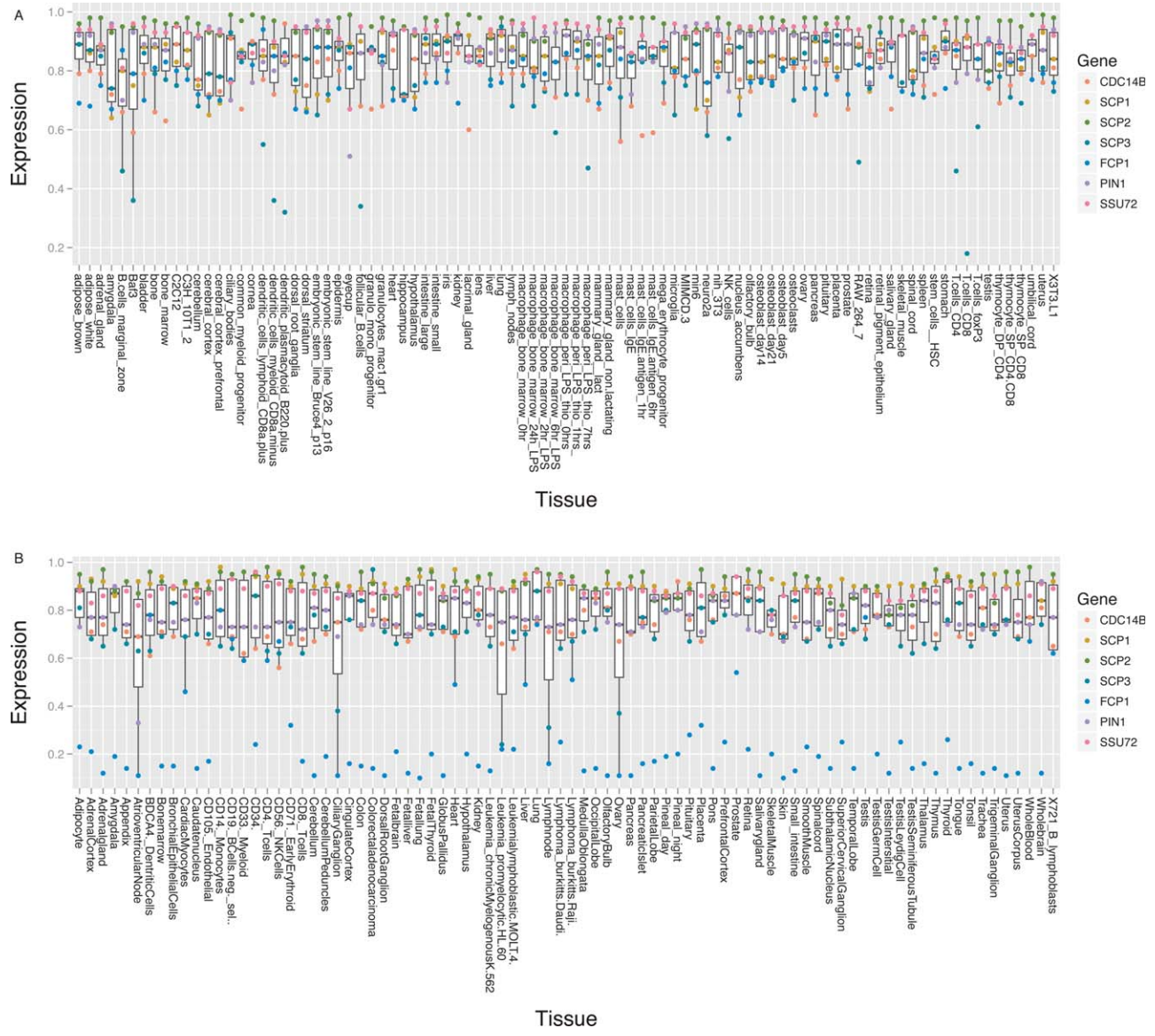


Figure 4

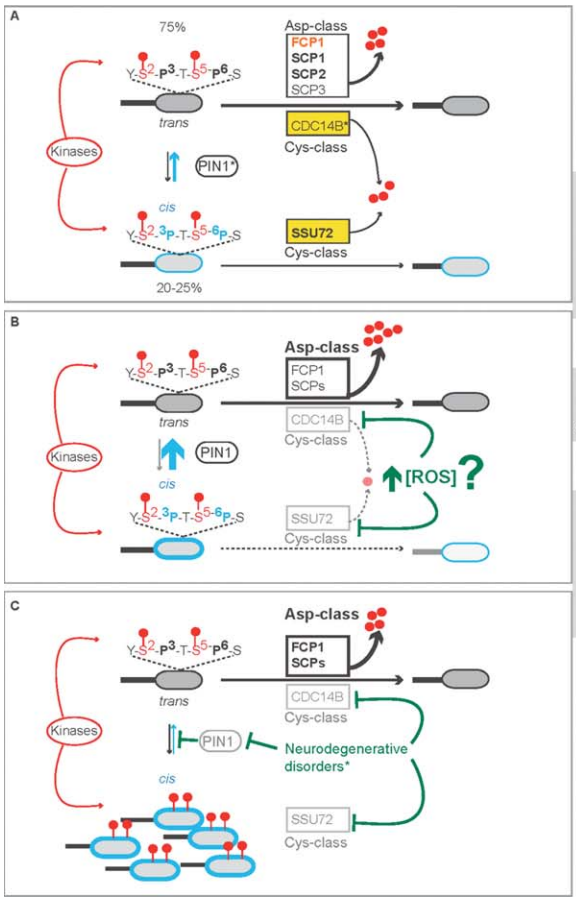
Comparison of Asp-CTD and Cys-CTD gene expression. *Panel A*, Mouse; *Panel B*, Human.

DISCUSSION

Traditionally, the presence of a characteristic set of conserved residues has been the cornerstone to identify enzymes through sequences analysis. However, in cases where structure was preserved but catalytic residues varied, the detection of evolutionary relationships was far more difficult.[105,106] Based on the findings presented here, together with a review of the literature, we are unable to draw hard conclusions about divergent/convergent evolutionary events for Asp- and Cys-CTD phosphatases, as occurs for other protein families (e.g., TIM-barrel fold).[107–109].

It is important to note that the evolution of enzymes does not merely involve changes in substrate specificity but also, those that involve the entire reaction space. Evolutionary changes can range from the use of similar catalytic residues on similar bonds to variations in catalytic residues, while retaining the same structural scaffold.[105,110] Examples of multi-catalytic folds include β-propellers,[111] the RMM-like fold,[105] and TIM-barrel[112] (for a recent review of the evolution of enzyme active sites[41]).

The borders between protein phosphatase classes are becoming more diffuse, posing significant challenges to



D

	Asp-CTD class				Cys-CTD class		
Alzheimer´s vs controls	FCP1	SCP1	SCP2	SCP3	CDC14B	SSU72	PIN1
hippocampus	down	-	-	up	-	down	down
medial_temporal_gyrus	-	up	up	up	-	down	down
superior_frontal_gyrus	-	up	up	up	down	-	-
posterior_cingulate	-	up	up	down	-	down	down
primary_visual_cortex	up	-	-	-	down	-	-
entorhinal_cortex	-	up	up	up	-	down	down
Parkinson´s vs Control	**FCP1**	**SCP1**	**SCP2**	**SCP3**	**CDC14B**	**SSU72**	**PIN1**
GSE7621 (substantia nigra)	-	-	-	up	up	-	-
GSE8397 (substantia nigra)	-	-	up	up	down	-	down
GSE20292 (substantia nigra)	-	-	-	down	down	-	-
schizophrenia vs control	**FCP1**	**SCP1**	**SCP2**	**SCP3**	**CDC14B**	**SSU72**	**PIN1**
AltarA (cortex)	-	-	-	-	-	-	-
AltarC (cortex)	up	up	-	up	-	-	-
Bahn (cortex)	-	-	-	-	down	-	-
Kato (cortex)	-	-	-	-	down	-	-
Kemether (thalamus)	up	up	-	up	up	down	down
Laeng (hippocampus)	-	-	-	up	down	down	-

Figure 5

Dephosphorylation mechanisms involving Asp- and Cys-CTD phosphatases in humans. (**A**) Normal conditions. Bold names indicate increased expression and orange FCP1 is silenced in most tissues: *indicates increased expression of PIN1 and CDC14B in the brain, a tissue affected strongly by ROS. (**B**) Potential effect of ROS as a consequence of oxidative stress. Inhibition of Cys-CTDs should increase PIN1 levels to transform the pool of *cis*-proline phosphorylated domains to *trans*-proline. (**C**) Model of unscheduled phosphorylation in neurodegenerative and mental disorders (* schizophrenia or Alzheimer's and Parkinson's diseases). The expression data indicates that Cys-CTDs and PIN1 are down regulated, which should lead to an increased pool of phosphorylated domains *cis*-proline domains that cannot be isomerized into *trans* domains, and the levels of Asp-CTDs are increased as compared with healthy samples. (**D**) Comparative analysis of CTD phosphatase expression in neurodegenerative and mental disorders. Up/Down, overexpressed or underexpressed genes in samples from patients with schizophrenia, Alzheimer's disease or Parkinson's disease compared with controls, determined as described in the Materials and Methods: (–), not significant; s.f.gyrus.normal, superior frontal gyrus; p.v.cortex, primary visual cortex; m.t.gyrus, medial temporal gyrus.

their classification.[113] Here, we have addressed an as-yet-unanswered question: is there an evolutionary connection between Asp-CTDs and Cys-CTDs? The common structural scaffold described in this study, provides an important basis for further studies of CTD phosphatases that may herald the reclassification of these enzymes, as well as other protein phosphatases that employ Asp- and Cys-based catalysis.

Another key question is why evolution has not simply provided a comprehensive set of Asp-CTDs, thereby circumventing the oxidation problem completely? It is known that transition between *cis*- and *trans*-Pro conformations adjacent to pSer-Pro motifs may affect the reaction rate of catalysis.[28] All Asp-CTDs are *trans*-Pro specific[28] and we speculate that a dephosphorylation mechanism that combines Asp-CTDs with peptidyl–prolyl *cis/trans* isomerase is not as efficient as one that uses a CTD-phosphatase for *cis*-Pro conformations. Indeed, Cys-CTD SSU72 is the only CTD-phosphatase identified to date that recognizes the pSer-Pro motif when proline adopts a *cis* conformation.[18,19] Therefore, we believe that the concerted action of Asp-CTD and Cys-CTD may help to preserve efficiency when the prolyl isomerization mechanism fails.

CONCLUSIONS

This study describes for the first time a common structural scaffold for the CTD phosphatases FCP1,

SCP1, SCP2, SCP3, SSU72, and CDC14B. This scaffold accommodates two distinct CTD phosphatase catalytic mechanisms, metal-ion-independent (Cys-CTD), and metal-ion-assisted (Asp-CTD) catalysis, in the same topographic region. Moreover, the scaffold resembles the four-stranded core of the Rossmann-fold, and it encompasses the active site sequence-motifs I, II and IV of the L-2-haloacid dehalogenase (HAD) superfamily, in which only Asp-CTDs have been included previously. The structural scaffold described here provides a connection between Cys-based PTPs (classes I, II, and III) and Asp-based PTPs (class IV).

We propose that additional players may participate in the fine-tuning of CTD-dephosphorylation in biological settings in which the catalytic mechanism is impaired, for example when additional enzymes act on residues positioned close to the targeted phosphorylated Ser/Thr residues, as occurs during proline isomerization in the heptapeptide sequence by the PIN1 isomerase. We investigated the functional implication of the putative relationship between Cys-based, Asp-based and Pro isomerization, screening functional gene expression data from a wide array of healthy and diseased tissues. We propose that the concerted action of Asp-CTD and Cys-CTD sustain robust transcription when the Cys-based mechanism is compromised and/or when alterations in proline isomerization affect the reaction rate of the catalysis. This hypothesis is partially supported by gene expression changes observed in human neurodegenerative and mental disorders, including schizophrenia and Alzheimer's or Parkinson's diseases.

ACKNOWLEDGMENTS

We thank Luis Sánchez-Pulido (University of Oxford), the members of the Structural Computational Biology Group at the CNIO, David de Juan, Victor de La Torre and Iakes Ezkurdia for their helpful discussions. We also thank Daniel Rico for providing us with a curated Barcode dataset.

REFERENCES

1. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 2006;127:635–648.
2. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science 2002; 298:1912–1934.
3. Manning BD. Challenges and opportunities in defining the essential cancer kinome. Sci Signal 2009;2:pe15.
4. Tonks NK. Protein tyrosine phosphatases: from genes, to function, to disease. Nat Rev Mol Cell Biol 2006;7:833–846.
5. MacKeigan JP, Murphy LO, Blenis J. Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. Nat Cell Biol 2005;7:591–600.
6. Almo SC, Bonanno JB, Sauder JM, Emtage S, Dilorenzo TP, Malashkevich V, Wasserman SR, Swaminathan S, Eswaramoorthy S, Agarwal R, Kumaran D, Madegowda M, Ragumani S, Patskovsky Y, Alvarado J, Ramagopal UA, Faber-Barata J, Chance MR, Sali A, Fiser A, Zhang Z-Y, Lawrence DS, Burley SK. Structural genomics of protein phosphatases. J Struct Funct Genom 2007;8:121–140.
7. Barr AJ, Ugochukwu E, Lee WH, King ONF, Filippakopoulos P, Alfano I, Savitsky P, Burgess-Brown NA, Müller S, Knapp S. Large-scale structural analysis of the classical human protein tyrosine phosphatome. Cell 2009;136:352–363.
8. Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. A structural perspective of CTD function. Genes Dev 2005;19:1401–1415.
9. Buratowski S. Progression through the RNA polymerase II CTD cycle. Mol Cell 2009;36:541–546.
10. Bataille AR, Jeronimo C, Jacques P-É, Laramée L, Fortin M-È, Forest A, Bergeron M, Hanes SD, Robert F. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Mol Cell 2012;45:158–170.
11. Egloff S, Dienstbier M, Murphy S. Updating the RNA polymerase CTD code: adding gene-specific layers. Trends Genet 2012;28:333–341.
12. Zhang DW, Rodríguez-Molina JB, Tietjen JR, Nemec CM, Ansari AZ. Emerging views on the CTD Code. Genet Res Int 2012;2012: 347214.
13. Zhang DW, Mosley AL, Ramisetty SR, Rodríguez-Molina JB, Washburn MP, Ansari AZ. Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. J Biol Chem 2012;287:8541–8551.
14. Moorhead GBG, De Wever V, Templeton G, Kerk D. Evolution of protein phosphatases in plants and animals. Biochem J 2009;417: 401–409.
15. Shi Y. Serine/threonine phosphatases: mechanism through structure. Cell 2009;139:468–484.
16. Alonso A, Sasin J, Bottini N, Friedberg I, Friedberg I, Osterman A, Godzik A, Hunter T, Dixon J, Mustelin T. Protein tyrosine phosphatases in the human genome. Cell 2004;117:699–711.
17. Allen KN, Dunaway-Mariano D. Phosphoryl group transfer: evolution of a catalytic scaffold. Trends Biochem Sci 2004;29:495–503.
18. Xiang K, Nagaike T, Xiang S, Kilic T, Beh MM, Manley JL, Tong L. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. Nature 2010;467:729–733.
19. Werner-Allen JW, Lee C-J, Liu P, Nicely NI, Wang S, Greenleaf AL, Zhou P. cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. J Biol Chem 2011;286:5717–5726.
20. Seifried A, Schultz J, Gohla A. Human HAD phosphatases: structure, mechanism, and roles in health and disease. FEBS J 2013;280: 549–571.
21. Allen KN, Dunaway-Mariano D. Markers of fitness in a successful enzyme superfamily. Curr Opin Struct Biol 2009;19:658–665.
22. Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. J Mol Biol 2006;361:1003–1034.
23. Chambers RS, Kane CM. Purification and characterization of an RNA polymerase II phosphatase from yeast. J Biol Chem 1996;271: 24498–24504.
24. Ghosh A, Shuman S, Lima CD. The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. Mol Cell 2008;32:478–490.
25. Guillamot M, Manchado E, Chiesa M, Gómez-López G, Pisano DG, Sacristán MP, Malumbres M. Cdc14b regulates mammalian RNA polymerase II and represses cell cycle transcription. Sci Rep 2011;1: 189.
26. Clemente-Blanco A, Sen N, Mayan-Santos M, Sacristán MP, Graham B, Jarmuz A, Giess A, Webb E, Game L, Eick D, Bueno A, Merkenschlager M, Aragón L. Cdc14 phosphatase promotes segregation of telomeres through repression of RNA polymerase II transcription. Nat Cell Biol 2011;13:1450–1456.

27. Xiang K, Manley JL, Tong L. An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72. Genes Dev 2012;26:2265–2270.
28. Zhang M, Wang XJ, Chen X, Bowman ME, Luo Y, Noel JP, Ellington AD, Etzkorn FA, Zhang Y. Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code. ACS Chem Biol 2012;7:1462–1470.
29. Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L, Workman JL, Washburn MP. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Mol Cell 2009;34:168–178.
30. Gibney PA, Fries T, Bailer SM, Morano KA. Rtr1 is the Saccharomyces cerevisiae homolog of a novel family of RNA polymerase II-binding proteins. Eukaryotic Cell 2008;7:938–948.
31. Jeronimo C, Forget D, Bouchard A, Li Q, Chua G, Poitras C, Thérien C, Bergeron D, Bourassa S, Greenblatt J, Chabot B, Poirier GG, Hughes TR, Blanchette M, Price DH, Coulombe B. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol Cell 2007;27:262–274.
32. Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. Mol Cell 2012;45:111–122.
33. Xiang K, Manley JL, Tong L. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. Nat Commun 2012;3:946.
34. Guo Z, Stiller JW. Comparative genomics and evolution of proteins associated with RNA polymerase II C-terminal domain. Mol Biol Evol 2005;22:2166–2178.
35. Zhang Y, Kim Y, Genoud N, Gao J, Kelly JW, Pfaff SL, Gill GN, Dixon JE, Noel JP. Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1. Mol Cell 2006;24:759–770.
36. Ganem C, Devaux F, Torchet C, Jacq C, Quevillon-Cheruel S, Labesse G, Facca C, Faye G. Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. EMBO J 2003;22:1588–1598.
37. Meinhart A, Silberzahn T, Cramer P. The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. J Biol Chem 2003;278:15917–15921.
38. Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. Ssu72 Is an RNA polymerase II CTD phosphatase. Mol Cell 2004; 14:387–394.
39. Zhang Y, Zhang M, Zhang Y. Crystal structure of Ssu72, an essential eukaryotic phosphatase specific for the C-terminal domain of RNA polymerase II, in complex with a transition state analogue. Biochem J 2011;434:435–444.
40. Gray CH, Good VM, Tonks NK, Barford D. The structure of the cell cycle protein Cdc14 reveals a proline-directed protein phosphatase. EMBO J 2003;22:3524–3535.
41. Gherardini PF, Wass MN, Helmer-Citterich M, Sternberg MJE. Convergent evolution of enzyme active sites is not a rare phenomenon. J Mol Biol 2007;372:817–845.
42. Wang S, Tabernero L, Zhang M, Harms E, Van Etten RL, Stauffacher CV. Crystal structures of a low-molecular weight protein tyrosine phosphatase from Saccharomyces cerevisiae and its complex with the substrate p-nitrophenyl phosphate. Biochemistry 2000;39: 1903–1914.
43. Barford D. The role of cysteine residues as redox-sensitive regulatory switches. Curr Opin Struct Biol 2004;14:679–686.
44. Finkel T. Oxidant signals and oxidative stress. Curr Opin Cell Biol 2003;15:247–254.
45. Salmeen A, Andersen JN, Myers MP, Meng T-C, Hinks JA, Tonks NK, Barford D. Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate. Nature 2003;423:769–773.
46. Kulathu Y, Garcia FJ, Mevissen TET, Busch M, Arnaudo N, Carroll KS, Barford D, Komander D. Regulation of A20 and other OTU deubiquitinases by reversible oxidation. Nat Commun 2013;4:1569.
47. Singh N, Ma Z, Gemmill T, Wu X, Defiglio H, Rossettini A, Rabeler C, Beane O, Morse RH, Palumbo MJ, Hanes SD. The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. Mol Cell 2009;36:255–266.
48. Krishnamurthy S, Ghazy MA, Moore C, Hampsey M. Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. Mol Cell Biol 2009;29:2925–2934.
49. Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B. Mitochondrial genome of Trichoplax adhaerens supports placozoa as the basal lower metazoan phylum. Proc Natl Acad Sci USA 2006;103:8751–8756.
50. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 2007;317:86–94.
51. Song J, Xu Q, Olsen R, Loomis WF, Shaulsky G, Kuspa A, Sucgang R. Comparing the Dictyostelium and Entamoeba genomes reveals an ancient split in the Conosa lineage. PLoS Comput Biol 2005;1: e71.
52. Eichinger L, Pachebat JA, Glöckner G, Rajandream M-A, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Madan Babu M, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabbinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A. The genome of the social amoeba Dictyostelium discoideum. Nature 2005;435:43–57.
53. Eichinger L, Noegel AA. Crawling into a new era-the Dictyostelium genome project. EMBO J. 2003;22:1941–1946.
54. Pagani I, Liolios K, Jansson J, Chen I-MA, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 2012;40:D571–D579.
55. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
56. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD. The Pfam protein families database. Nucleic Acids Res 2012;40:D290–D301.
57. Chen F, Mackey AJ, Stoeckert CJ, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res 2006;34:D363–D368.
58. DeLuca TF, Cui J, Jung J-Y, St Gabriel KC, Wall DP. Roundup 2.0: enabling comparative genomics for over 1800 genomes. Bioinformatics 2012;28:715–716.
59. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747.
60. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res 2010;38:W545–W549.
61. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 2004;60:2256–2268.

62. McCall MN, Uppal K, Jaffee HA, Zilliox MJ, Irizarry RA. The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. Nucleic Acids Res 2011;39:D1011–D1015.
63. Obayashi T, Kinoshita K. COXPRESdb: a database to compare gene coexpression in seven model animals. Nucleic Acids Res 2011;39: D1016–D1022.
64. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Holko M, Ayanbule O, Yefanov A, Soboleva A. NCBI GEO: archive for functional genomics data sets–10 years on. Nucleic Acids Res 2011;39:D1005–D1010.
65. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW, Su AI. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol 2009;10:R130.
66. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA 2004;101:6062–6067.
67. Lattin JE, Schroder K, Su AI, Walker JR, Zhang J, Wiltshire T, Saijo K, Glass CK, Hume DA, Kellie S, Sweet MJ. Expression analysis of G Protein-Coupled Receptors in mouse macrophages. Immunome Res 2008;4:5.
68. Liang WS, Reiman EM, Valla J, Dunckley T, Beach TG, Grover A, Niedzielko TL, Schneider LE, Mastroeni D, Caselli R, Kukull W, Morris JC, Hulette CM, Schmechel D, Rogers J, Stephan DA. Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. Proc Natl Acad Sci USA 2008;105:4441–4446.
69. Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, de Andrade M, Henley JR, Rocca WA, Ahlskog JE, Maraganore DM. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. PLoS Genet 2007;3:e98.
70. Moran LB, Duke DC, Deprez M, Dexter DT, Pearce RKB, Graeber MB. Whole genome expression profiling of the medial and lateral substantia nigra in Parkinson's disease. Neurogenetics 2006;7:1–11.
71. Zhang Y, James M, Middleton FA, Davis RL. Transcriptional analysis of multiple brain regions in Parkinson's disease supports the involvement of specific protein processing, energy metabolism, and signaling pathways, and suggests novel disease mechanisms. Am J Med Genet B Neuropsychiatr Genet 2005;137B:5–16.
72. Higgs BW, Elashoff M, Richman S, Barci B. An online database for brain disease research. BMC Genomics 2006;7:70.
73. McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). Biostatistics 2010;11:242–253.
74. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:Article 3.
75. Andreeva AV, Kutuzov MA. Protozoan protein tyrosine phosphatases. Int J Parasitol 2008;38:1279–1295.
76. Best AA, Morrison HG, McArthur AG, Sogin ML, Olsen GJ. Evolution of eukaryotic transcription: insights from the genome of Giardia lamblia. Genome Res 2004;14:1537–1547.
77. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-Nesselquist E, Manning G, Nigam A, Nixon JEJ, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svard SG, Sogin ML. Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science 2007;317:1921–1926.
78. Shi L, Potts M, Kennelly PJ. The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: a family portrait. FEMS Microbiol Rev 1998;22:229–253.
79. Matsuda O, Sakamoto H, Nakao Y, Oda K, Iba K. CTD phosphatases in the attenuation of wound-induced transcription of jasmonic acid biosynthetic genes in Arabidopsis. Plant J 2009;57:96–108.
80. Ueda A, Li P, Feng Y, Vikram M, Kim S, Kang CH, Kang JS, Bahk JD, Lee SY, Fukuhara T, Staswick PE, Pepper AE, Koiwa H. The Arabidopsis thaliana carboxyl-terminal domain phosphatase-like 2 regulates plant growth, stress and auxin responses. Plant Mol Biol 2008;67:683–697.
81. Williams RS, Lee MS, Hau DD, Glover JNM. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. Nat Struct Mol Biol 2004;11:519–525.
82. Mocciaro A, Schiebel E. Cdc14: a highly conserved family of phosphatases with non-conserved functions? J Cell Sci 2010;123:2867–2876.
83. Gruneberg U, Glotzer M, Gartner A, Nigg EA. The CeCDC-14 phosphatase is required for cytokinesis in the Caenorhabditis elegans embryo. J Cell Biol 2002;158:901–914.
84. Saito RM, Perreault A, Peach B, Satterlee JS, van den Heuvel S. The CDC-14 phosphatase controls developmental cell-cycle arrest in C. elegans. Nat Cell Biol 2004;6:777–783.
85. Mocciaro A, Berdougo E, Zeng K, Black E, Vagnarelli P, Earnshaw W, Gillespie D, Jallepalli P, Schiebel E. Vertebrate cells genetically deficient for Cdc14A or Cdc14B retain DNA damage checkpoint proficiency but are impaired in DNA repair. J Cell Biol 2010;189: 631–639.
86. Cho HP, Liu Y, Gomez M, Dunlap J, Tyers M, Wang Y. The dual-specificity phosphatase CDC14B bundles and stabilizes microtubules. Mol Cell Biol 2005;25:4541–4551.
87. Wu J, Cho HP, Rhee DB, Johnson DK, Dunlap J, Liu Y, Wang Y. Cdc14B depletion leads to centriole amplification, and its overexpression prevents unscheduled centriole duplication. J Cell Biol 2008;181:475–483.
88. Liu P, Kenney JM, Stiller JW, Greenleaf AL. Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. Mol Biol Evol 2010;27:2628–2641.
89. Kishore SP, Perkins SL, Templeton TJ, Deitsch KW. An unusual recent expansion of the C-terminal domain of RNA polymerase II in primate malaria parasites features a motif otherwise found only in mammalian polymerases. J Mol Evol 2009;68:706–714.
90. Hu H, Columbus J, Zhang Y, Wu D, Lian L, Yang S, Goodwin J, Luczak C, Carter M, Chen L, James M, Davis R, Sudol M, Rodwell J, Herrero JJ. A map of WW domain family interactions. Proteomics 2004;4:643–655.
91. Sudol M, Chen HI, Bougeret C, Einbond A, Bork P. Characterization of a novel protein-binding module–the WW domain. FEBS Lett 1995;369:67–71.
92. André B, Springael JY. WWP, a new amino acid motif present in single or multiple copies in various proteins including dystrophin and the SH3-binding Yes-associated protein YAP65. Biochem Biophys Res Commun 1994;205:1201–1205.
93. Hofmann K, Bucher P. The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. Trends Biochem Sci 1995;20:347–349.
94. Durocher D, Jackson SP. The FHA domain. FEBS Lett 2002;513: 58–66.
95. Landrieu I, De Veylder L, Fruchart JS, Odaert B, Casteels P, Portetelle D, Van Montagu M, Inzé D, Lippens G. The Arabidopsis thaliana PIN1At gene encodes a single-domain phosphorylation-dependent peptidyl prolyl cis/trans isomerase. J Biol Chem 2000; 275:10577–10581.
96. Landrieu I, Wieruszeski J-M, Wintjens R, Inzé D, Lippens G. Solution structure of the single-domain prolyl cis/trans isomerase PIN1At from Arabidopsis thaliana. J Mol Biol 2002;320: 321–332.
97. Zhou XZ, Lu PJ, Wulf G, Lu KP. Phosphorylation-dependent prolyl isomerization: a novel signaling regulatory mechanism. Cell Mol Life Sci 1999;56:788–806.
98. Kobor MS, Archambault J, Lester W, Holstege FC, Gileadi O, Jansma DB, Jennings EG, Kouyoumdjian F, Davidson AR, Young RA, Greenblatt J. An unusual eukaryotic protein phosphatase

required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae. Mol Cell 1999;4:55–62.
99. Niu W, Li Z, Zhan W, Iyer VR, Marcotte EM. Mechanisms of cell cycle control revealed by a systematic and quantitative overexpression screen in S. cerevisiae. PLoS Genet 2008;4:e1000120.
100. Gauthier NP, Jensen LJ, Wernersson R, Brunak S, Jensen TS. Cyclebase.org: version 2.0, an updated comprehensive, multi-species repository of cell cycle experiments and derived analysis results. Nucleic Acids Res 2010;38:D699–D702.
101. Obayashi T, Kinoshita K. Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. DNA Res 2009;16:249–260.
102. Pimentel C, Batista-Nascimento L, Rodrigues-Pousada C, Menezes RA. Oxidative stress in Alzheimer's and Parkinson's diseases: insights from the yeast Saccharomyces cerevisiae. Oxid Med Cell Longev 2012;2012:132146.
103. Ramalingam M, Kim S-J. Reactive oxygen/nitrogen species and their functional correlations in neurodegenerative diseases. J Neural Transm 2012;119:891–910.
104. Su B, Wang X, Nunomura A, Moreira PI, Lee H-G, Perry G, Smith MA, Zhu X. Oxidative stress signaling in Alzheimer's disease. Curr Alzheimer Res 2008;5:525–532.
105. Anantharaman V, Aravind L, Koonin EV. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. Curr Opin Chem Biol 2003;7:12–20.
106. Andreeva A, Murzin AG. Structural classification of proteins and structural genomics: new insights into protein folding and evolution. Acta Crystallogr Sect F Struct Biol Cryst Commun 2010;66:1190–1197.
107. Murzin AG. How far divergent evolution goes in proteins. Curr Opin Struct Biol 1998;8:380–387.
108. Grishin NV. Fold change in evolution of protein structures. J Struct Biol 2001;134:167–185.
109. Copley RR, Bork P. Homology among (betaalpha)(8) barrels: implications for the evolution of metabolic pathways. J Mol Biol 2000;303:627–641.
110. Bazan JF, de Sauvage FJ. Structural ties between cholesterol transport and morphogen signaling. Cell 2009;138:1055–1056.
111. Chen CK-M, Chan N-L, Wang AH-J. The many blades of the β-propeller proteins: conserved but versatile. Trends Biochem Sci 2011;36:553–561.
112. Nagano N, Orengo CA, Thornton JM. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J Mol Biol 2002;321:741–765.
113. Li X, Wilmanns M, Thornton J, Köhn M. Elucidating human phosphatase-substrate networks. Sci Signal 2013;6:rs10.
114. Hedges SB, Blair JE, Venturi ML, Shoe JL. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol 2004;4:2.
115. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera-A visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612.