Phosphatases play an essential role in the regulation of protein phosphorylation. Less abundant than kinases, many phosphatases are components of one or more macromolecular complexes with different substrate specificities and specific functionalities. The expert scientific curation of phosphatase complexes for the UniProt and Complex Portal databases supports the whole scientific community by collating and organising small- and large-scale experimental data from the scientific literature into context-specific central resources, where the data can be freely accessed and used to further academic and translational research. In this review, we discuss how the diverse biological functions of phosphatase complexes are presented in UniProt and the Complex Portal, and how understanding the biological significance of phosphatase complexes in Caenorhabditis elegans offers insight into the mechanisms of substrate diversity in a variety of cellular and molecular processes.

Introduction

The spatial and temporal phosphorylation of proteins and lipids by kinases is tightly regulated by their subsequent dephosphorylation by phosphatases and is necessary for the progression and/or termination of a multitude of signalling pathways. Phosphatases are dynamic hydrolases that catalyse the removal of phosphate groups from specific amino acid residues or lipid moieties in target protein or lipid substrates, respectively. Phosphatase activity is associated with various catalytic mechanisms and different protein folds which have been used for their classification [1,2].

Despite their relevance in signalling pathways and biological processes, the number of phosphatase genes is significantly less abundant than kinases in most genomes, and constitute only between 0.5% and 1.5% of proteome content across a broad taxonomic range [3]. Phosphatases therefore employ various mechanisms to act on a similar range of targets as kinases. They can act as single units displaying a broad spectrum of substrates, such as the mitogen-activated protein kinase phosphatases (MKPs) which dephosphorylate mitogen-activated protein kinases ERK1/2, p38 and JNK, to terminate MAP kinase-mediated signalling cascades.
They can also stably associate with other structural proteins and form holoenzymes, such as PPP2CA with the constant subunit PPP2R1A, which in turn can associate with various regulatory subunits to confer substrate specificity and regulate catalytic activity of the phosphatase [5,6]. Despite their reduced gene number in comparison to kinases, phosphatases play vital roles in a variety of biological processes. In humans, disruption in the balance between kinase and phosphatase activities results in diseases such as diabetes, a range of cancers and neurodegenerative diseases such as Alzheimer’s [7]. Drugs based on manipulating kinase activity have proved invaluable in the treatment of some of these diseases [8]. As modulators of phosphorylation, phosphatases are also appropriate candidates for drug development [1,9,10]. Indeed, manipulating or inhibiting phosphatase activity is a growing field, with many phosphatase inhibitors currently being considered as potential drug targets, but to date only few are available as licensed drugs [11]. For example, cyclosporine A and FK506, two inhibitors of protein serine/threonine phosphatase 2B (also known as calcineurin), have been used successfully as immunosuppressants for the treatment of patients following organ transplantation [12]. To support the progression of translational research and drug development, biological databases collate and organise data from 3D structure analyses and large-scale proteomics studies as well as from smaller-scale initiatives presented in the scientific literature. This facilitates the identification of specific phosphatases, and thus, such databases constitute invaluable research tools. Whilst some databases such as Phosphatome.net (phosphatome.net/3.0) focus specifically on phosphatases, databases such as UniProt (www.uniprot.org) provide a more general, but comprehensive overview on all proteins in a wide variety of taxa [13]. Other resources specialise in particular aspects of protein biology. One such example is the Complex Portal (www.ebi.ac.uk/complexportal) [14], which focuses on the arrangement of proteins in macromolecular complexes. In the phosphatase field, the collation of such data is of great relevance as a large number of human phosphatases are components of one or more macromolecular complexes with specific biological functions [1,15,16]. Specifically, UniProt and the Complex Portal are two resources that provide comprehensive functional and structural data on complexes of a large variety of species in different, but complementary ways. UniProt provides functional and structural data and sequences of proteins on a protein-by-protein basis for over 1 million species (UniProt release 2019_09). The Complex Portal also provides functional and structural data on proteins from a broad taxonomic range (including human and key model organisms such as Escherichia coli, Mus musculus, Drosophila melanogaster and Caenorhabditis elegans), but within the context of macromolecular complexes. In this review, we provide an analysis of the C. elegans phosphatome and phosphatase-containing complexes based on their curation and presentation in UniProt and the Complex Portal. We highlight the challenges in gathering data from the scientific literature and inferring biological function from orthologs, and discuss how the two databases display this data. Ultimately, we demonstrate how understanding the biological significance of phosphatase complexes in a model organism such as C. elegans allows for insight into how phosphatases are able to target such a diverse range of substrates in a variety of biological processes. This is particularly necessary for the identification of potential drug targets for therapies and treatments for a range of human diseases.

**C. elegans phosphatome and complexes**

The study of phosphatase evolution across species underwent an important step forward with the recent compilation of the phosphatomes for human and several model organisms including C. elegans [3]. However, whilst this study provides a list of phosphatases and a classification based on the fold of their catalytic domains, it does not provide information about the extent of their functions or their interactions. Phosphatase-specific databases such as DEPOD (www.depod.bios.uni-freiburg.de/) provide information about interacting partners together with their substrates and biological functions; however, their focus is the characterisation of human phosphatases [17]. To help fill the gap this focus creates and provide contextual understanding of the dynamics of these proteins in C. elegans, we comprehensively updated the C. elegans phosphatome in the UniProt database with functional data from the scientific literature (for a description of the curation process, see [18]). We chose to focus on this nematode species as it has been an invaluable tool in understanding protein biology [19]. More than 50% of human phosphatases have a counterpart in C. elegans and many are involved in the same biological processes [20]. Moreover, the relative ease in manipulating the C. elegans genome, combined with its fast life cycle, make it an ideal model to complement human and mouse studies and allows for the
investigation of protein function in the context of a whole organism [19].

The first step in the curation of the *C. elegans* phosphatome was to create a list of phosphatases and their corresponding UniProtKB accessions. As a starting point, we used the gene list compiled by Chen et al. for *C. elegans* [3]. Of the 244 genes in this list, six genes corresponded to pseudogenes in the nematode database Wormbase [21] and were discarded, and two genes had the same sequence and thus were considered as one. Secondly, to identify additional phosphatases, we also took advantage of the cross references to the protein family database Pfam [22]. These cross references are automatically added to UniProtKB entries. Using Pfam signatures for phosphatase families (e.g. PF00102 Y_phosphatase), we could not find any other phosphatases besides the one identified by Chen et al. Our final list contains 237 phosphatase genes belonging to the CC1 (Cys-based class I fold), CC2 (Cys-based class II fold), CC3 (Cys-based class III fold), PPM (metal-dependent protein phosphatase fold), PPPL (protein phosphatase-like fold), HAD (haloacid dehalogenase-like fold), HP (histidine acid phosphatase fold), PHP (protein histidine phosphatase fold), RTR1 (regulator of transcription fold) and AP (alkaline phosphatase fold) superfamilies (Table S1). As the focus of this study is protein phosphatases, we did not include members of the inositol-4-phosphate, inositol-5-phosphate, glucose/lipid phosphatase and adenosine/fructose/inositol-x-phosphatases superfamilies. We used experimentally validated information extracted from the scientific literature to update the annotation of over 25 phosphatases of the 237 UniProt entries. In total, there are now 71 phosphatases (representing 30% of the *C. elegans* phosphatome) whose functions have been experimentally characterised (known as reviewed entries in UniProt) (Fig. 1A). This effort provides an up-to-date view of how much is currently known about the *C. elegans* phosphatome. Overall, the level of experimental characterisation is around 25–30% for the three major superfamilies CC1 (Cys-based class I), PPPL (protein phosphatase-like) and HP (histidine acid phosphatase) (Fig. 1B). Unsurprisingly, most of the well-characterised phosphatases have human orthologs (Table 1).

In parallel with this functional update effort, we also added experimental characterisation data for phosphatase complexes to both UniProt and the Complex Portal (see below for examples and an overview of the process). Among the 237 *C. elegans* phosphatases, 12 have been experimentally proven to be part of 16 stable complexes (Table 2). Each phosphatase-containing complex has specific functional features. In particular, among the 12 characterised phosphatases, MTM-9 (UniProtKB Q965W9) and the *C. elegans*-specific EGG-3 (UniProtKB Q20402), EGG-4 (UniProtKB O01767) and EGG-5 (UniProtKB O61789) are pseudophosphatases which lack catalytic activity [23,24]. Interestingly, whilst the majority of phosphatase complexes usually contain one active phosphatase or a combination of an active and an inactive phosphatase, three of the inactive phosphatases, EGG-3, EGG-4 and EGG-5, are found in a single complex, the Egg-3/4/5/MBK-2 complex (Complex Portal: CPX-3381), which does not contain an active phosphatase [24]. All characterised active phosphatases are Ser/Thr phosphatases, except for MTM-6, which is a phosphoinositide phosphatase [25]. The *C. elegans* phosphatase complexes are remarkably well conserved. To date, apart from the Egg-3/4/5/MBK-2 complex, which appears to be specific to *C. elegans*, all the other experimentally characterised phosphatase complexes are present in a wide variety of species including mammals, fruit fly and *Saccharomyces cerevisiae*. We describe below how UniProt and the Complex Portal present phosphatase complex data and use specific examples to highlight the broad data types.

**UniProt and the Complex Portal databases**

The Complex Portal and UniProt present complementary data on complexes to provide biologically relevant functional information. Both databases define macromolecular complexes as stable sets of two or more interacting protein molecules that can be copurified and that exist as functional units in vivo. These complexes may also contain nonprotein molecules such as small molecules and nucleic acids. Interactions between enzymes and their substrates are usually not annotated as complexes as they are often transient. However, if the substrate remains stably associated with the enzyme and their association is essential for the function, then they are annotated as complexes.

Identifying complexes and their components is experimentally challenging. To take this into account, UniProt and the Complex Portal curators use stringent criteria to determine what constitutes a complex. Briefly, copurification by sequential column chromatography, reciprocal co-immunoprecipitation and/or GST pull-downs, and biophysical techniques such as cocrystallisation, electron microscopy or fluorescence resonance energy transfer methods associated with functional assays are suitable evidence for the interaction. However, large-scale pull-downs followed by mass spectrometry or yeast two-hybrid assays are
not considered unless the interactions are further confirmed with additional experiments. Also, we do not annotate complexes if they are only inferred from genetic interactions or are only detectable when amino acid residues are mutated to ‘stabilise’ the association. Nonetheless, if there is firm evidence in another species that the complex does exist, we do annotate the complex using orthology-based evidence (see below).

UniProt presents complex data on a protein-by-protein basis in the ‘Interaction’ section of each protein entry participating in the complex (Fig. 2A). If known, binding regions for each protein are provided in the subunit comment and/or in the sequence feature table and the ProtVista feature viewer (Fig. 2B, C) [26]. Proteins that are integral components of a macromolecular complex will have a supporting cross reference that provides a direct link to the relevant entry page in the Complex Portal (Fig. 3A). The Complex Portal presents functional and structural data on macromolecular complexes as a whole, which can be identified by a complex-specific accession number (Fig. 3A, E). Specific proteins within the complexes can be identified by UniProt accession numbers with a link to the corresponding UniProt records (Fig. 3D) [14]. The Complex Portal provides an interactive graphical view of the complex which includes not only protein participants, but also any small molecules, such as metal ions, that are an integral part of the protein function. For each key protein, domains, features and metal binding sites can be viewed (Fig. 3B) and where known, the specific amino acids involved in the interactions are specified (Fig. 3C).

There are multiple challenges associated with the annotation of macromolecular complexes in both databases, for example, (a) how we assign complex names, taking into account functional conservation among species, (b) how we annotate proteins that belong to multiple complexes and (c) how information about complex formation, stability and regulation is included. To illustrate how UniProt and Complex Portal curators address these challenges, we chose six phosphatases among the 12 phosphatases that have been experimentally shown to be part of complexes as they are well characterised. These are the three phosphatases LET-92 (UniProtKB G5EGK8), TAX-6 (UniProtKB Q0G819) and FEM-2 (UniProtKB P49594), and the three pseudophosphatases EGG-3, EGG-4 and EGG-5.

**LET-92: one phosphatase for 8 complexes**

Reversible phosphorylation is essential for a range of cellular processes. In *C. elegans*, there are 438 kinases and to date, over half of them have been experimentally characterised [18]. By contrast, there are only 237 phosphatases in the *C. elegans* phosphatome (Table S1). How do so few phosphatases counteract the phosphorylation activity of such a high number of kinases in diverse processes? In order to achieve functional diversity, one phosphatase may associate with different ‘regulatory’ proteins, which confer substrate and functional specificity. The participation of a phosphatase in multiple complexes raises several questions in terms of how databases can comprehensively and clearly present this.

---

**Fig. 1.** Curation of the *Caenorhabditis elegans* phosphatome in UniProt. (A) Percentage of phosphatases with (reviewed; yellow) and without (unreviewed; blue) experimental data available. (B) Percentage of phosphatases within the major phosphatase superfamilies with (reviewed; yellow) and without (unreviewed; blue) experimental data. CC1, Cys-based class I fold; CC2, Cys-based class II fold; CC3, Cys-based class III fold; PPM, metal-dependent protein phosphatase fold; PPPL, protein phosphatase-like fold; HAD, haloacid dehalogenase-like fold; HP, histidine acid phosphatase fold; PHP, protein histidine phosphatase fold; RTR1, regulator of transcription fold; AP, alkaline phosphatase fold.
### Table 1. Human homologues of the 71 reviewed Caenorhabditis elegans phosphatases. Ensembl Compara was used to identify homologous human genes (https://www.ensembl.org/info/genome/compara/index.html)

| Classification | C. elegans (reviewed entries) | Human homologues |
|----------------|-------------------------------|-------------------|
|                | Gene name | UniProtKB AC | Gene name | |
| CC1: DSP: CDC14 | cdc-14 | P81299 | CDC14A | CDC14B | CDC14C |
| CC1: DSP: CDC11 | F54C8.4 | P34442 | DUSP11 |
| CC1: DSP: CDC11 | pir-1 | Q22707 | DUSP11 |
| CC1: DSP: CDC14 | ZK757.2 | P34680 | DUSP14 |
| CC1: DSP: DSP6 | lip-1 | V6CIV8 | DUSP6 | DUSP7 | DUSP9 |
| CC1: DSP: DSP8 | vhp-1 | Q10038 | DUSP16 | DUSP8 |
| CC1: DSP: RNGTT | cel-1 | Q17607 | RNGTT |
| CC1: Myotubulin: MTMR1 | mmt-1 | Q9NL89 | MTM1 | MTM1R | MTMR2 |
| CC1: Myotubulin: MTMR3 | mmt-3 | Q22712 | MTMR3 | MTMR4 |
| CC1: Myotubulin: MTMR5 | mmt-5 | Q97X53 | SBF1 | SBF2 |
| CC1: Myotubulin: MTMR6 | mmt-6 | Q5ED8 | MTMR6 | MTMR7 | MTMR8 |
| CC1: Myotubulin: MTMR9 | mmt-9 | Q96W9 | MTMR9 |
| CC1: PTEN: PTEN | daf-18 | G5EE01 | PTEN |
| CC1: PT: CelePTP-s14 | B0280.11 | P42083 | None |
| CC1: PT: CelePTP-s14 | C15H7.3 | P34337 | None |
| CC1: PT: eak | eak-6 | G5EGU9 | None |
| CC1: PT: egg | egg-3 | Q20402 | None |
| CC1: PT: egg | egg-4 | O01767 | None |
| CC1: PT: egg | egg-5 | O61789 | None |
| CC1: PT: Ptp690 | crz-1 | H2KZM6 | PTPR |
| CC1: PT: PTPN3 | ptp-1 | P28191 | PTPN3 | PTPN4 |
| CC1: PT: PTPN6 | ptp-2 | G5EC24 | PTPN11 | PTPN6 |
| CC1: PT: PTPRB | dep-1 | G5EGJ9 | PTPR8 | PTPRH | PTPRJ |
| CC1: PT: PTPRD | ptp-3 | Q9BMN8 | PTPRD | PTPR | PTPR5 |
| CC1: PT: PTPRG | ptp-4 | P28192 | PTPRG | PTPRZ1 |
| CC1: PT: PTPRN | ida-1 | G5EG5J | PTPRN | PTPRN2 |
| CC1: Sac: Synaptotagmin | unc-26 | G5EC12 | SYNJ1 | SYNJ2 |
| CC2: SSU72: SSU72 | ssup-72 | Q22453 | SSU72 | SSU72P2 | SSU72P3 |
| CC3: CDC25: CDC25 | F16B4.8 | O44628 | CDC25A | CDC25B | CDC25C | SSU72P8 |
| CC3: CDC25: CDC25 | K06A5.7 | O44552 | CDC25A | CDC25B |CDC25C |
| CC3: CDC25: CDC25 | ZK637.11 | P30634 | CDC25A | CDC25B |CDC25C |
| HAD: HAD: EYA: EYA | eya-1 | S71670 | EYA1 | EYA2 | EYA3 |
| HAD: HAD: FCP: CTDSL2 | scp-3 | Q9NN4V | CTDSL2 | EYA4 |
| HAD: HAD: FCP: DULLARD | cne-1 | Q20432 | CTDSL2 | EYA4 |
| HAD: HAD: FCP: FCP1 | fcp-1 | Q9S5Q8 | CTDP1 |
| HAD: HAD: FCP: SCP | scp-1 | Q8T312 | CTDSP1 | CTDSP2 | CTDSL |
| HAD: HAD: FCP: TIM50 | scp-4 | Q22647 | TIM50 |
| HAD: HAD: NagD: PGP | K02D10.1 | P34492 | NIPSAP1 | NIPSAP2 | PDXP |
| HP: HP: PFGFB | pffk-1.2 | Q21122 | PFFFB1 | PFFFB2 | PFFFB3 | PGP |
| HP: HP: PGAM5 | pgam-5 | Q94222 | PGAM5 |
| HP: HP: ACP2 | acp-1 | Q23534 | None |
| HP: HP: ACP2 | F26C11.1 | Q90549 | None |
| HP: HP: ACP2 | pho-1 | Q19076 | None |
| HP: HP: ACP2 | pho-10 | Q90444 | None |
| HP: HP: ACP2 | pho-10 | Q90451 | None |
| HP: HP: ACP2 | pho-5 | Q10944 | ACP2 | ACP3 | ACP4 |
| HP: HP: PPIPK5 | F46F11.1 | P91309 | PPIPK1 | PPIPK2 |
| PPM: PPM: PPM: PHLPP | philip-2 | Q99564 | PHILPP1 | PHILPP2 |
| PPM: PPM: PPM: PPM1A | pmm-1 | Q19775 | PPM1A | PPM1B | PPM1N |
| PPM: PPM: PPM: PPM1E | tem-2 | F49594 | PPM1E | PPM1F |
| PPM: PPM: PPM: PPM1G | F42G9.1 | F49595 | PPM1G |
| PPM: PPM: PPM: PPM12 | ppm-2 | F49596 | None |
An interesting example is the PP2A family (Fig. 4). These proteins are among the most well-characterised serine/threonine phosphatases. They play essential roles in many key and diverse biological processes such as cell division and apoptosis [27,28]. Several PP2A-containing complexes have been experimentally characterised in mammals and functional conservation is apparent throughout the animal kingdom in species such as *D. melanogaster* and *S. cerevisiae* (Fig. 4A) [29–31]. In particular, a phylogenetic analysis of the catalytic subunit shows that PP2A is highly conserved across several species. In *C. elegans*, there is only one PP2A family member encoded by the *let-92* gene (Fig. 4A). Based on protein sequence similarities, *LET-92* most likely associates with the scaffolding protein *PAA-1* (UniProtKB Q09543) to form the common heterodimeric core enzyme that is the basis of multiple PP2A complexes (Fig. 4B). The association of this core with regulatory proteins confers distinct functional properties. Experimental data demonstrates that this core is involved in processes including cell division, centriole duplication, microtubule outgrowth and vulval development in *C. elegans* [5,6,32].

Whilst the interaction between a PP2A catalytic subunit and a constant scaffolding protein has been experimentally demonstrated in mammals, to date, no experimental data confirms the physical association between *C. elegans* LET-92 and PAA-1 subunits (Fig. 4B). However, the existence of the LET-92-PAA-1 heterodimer has been inferred from genetic interaction studies, a common tool in *C. elegans*, whereby the analysis of animals lacking *paa-1* and/or *let-92* has shown that the constant and catalytic subunits share similar functions and appear to be part of a common signalling pathway [6]. Despite the lack of direct evidence for the existence of the LET-92-PAA-1 heterodimer, biochemical studies provide evidence for an interaction between the regulatory subunit RSA-1 (UniProtKB O02217) and the core proteins LET-92 and PAA-1 (Fig. 4B) [6], and between the regulatory subunit SUR-6 (UniProtKB G5EDR3) and LET-92 [5]. Other potential regulatory subunits have been identified based on sequence similarity (*F43B10.1, F47B8.3, T22D1.5, Y71H2AM.20*) or genetic interaction studies (*pptr-1, pptr-2*) (Fig. 4A) [33,34]. Based on this information, LET-92 appears to be the common component of at least 8 distinct complexes (Table 2).

Table 1. (Continued).

| Classification | *C. elegans* (reviewed entries) | Human homologues |
|----------------|---------------------------------|-----------------|
| PPM: PPM: PPM: TAB1 | tap-1 | G5EE5M6 | TAB1 |
| PPL: PPL: PPL: C06A1.3 | C06A1.3 | P48459 | None |
| PPL: PPL: PPL: C30A5.4 | C30A5.4 | P34350 | None |
| PPL: PPL: PPL: C27B7.6 | C27B7.6 | P48460 | None |
| PPL: PPL: PPL: C23G10.1 | C23G10.1 | P48459 | None |
| PPL: PPL: PPL: O02658 | O02658 | None | None |
| PPL: PPL: PPL: P91420 | P91420 | None | None |
| PPL: PPL: PPL: F44B9.9 | F44B9.9 | P34430 | None |
| PPL: PPL: PPL: Q27497 | Q27497 | PFP1CB | PFP1CB |
| PPL: PPL: PPL: P34727 | P34727 | PFP1CA | PFP1CA |
| PPL: PPL: PPL: G5EGK8 | G5EGK8 | PFP2CA | PFP2CA |
| PPL: PPL: PPL: Q0G819 | Q0G819 | PFP3CA | PFP3CA |
| PPL: PPL: PPL: Q9KV79 | Q9KV79 | PFP4C | PFP4C |
| PPL: PPL: PPL: Q9XTT8 | Q9XTT8 | PFP4B | PFP4C |
| PPL: PPL: PPL: Q9NES8 | Q9NES8 | PFP5C | PFP5C |
| PPL: PPL: PPL: O09496 | O09496 | PFP6B | PFP6C |
| PPL: PPL: PPL: G5EBX9 | G5EBX9 | PPEF1 | PPEF1 |
| PPL: PPL: PPL: R08D7.2 | R08D7.2 | P30641 | P30641 |

The FEBS Journal 287 (2020) 2664–2684 © 2020 European Bioinformatics Institute. The FEBS Journal published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies.
regulatory subunits, and not all are experimentally confirmed, only the components of the common heterodimer core, LET-92 and PAA-1, are explicitly named (Fig. 4C). The interactions with specific subunits supported by experimental evidence are described separately. The ‘Function’ section of a UniProt protein entry provides more information about the individual role of each component in any complexes it may form. For example, during mitosis, the regulatory subunit RSA-1 recruits the heterodimeric core of LET-92 and PAA-1 to the centrosomes, and thereby regulates microtubule outgrowth from centrosomes and mitotic spindle assembly (Fig. 4D) [6]. By a different mechanism, the regulatory subunit SUR-6 associates with LET-92 to regulate centriole duplication during early embryonic cell divisions [5].

In UniProt, all the complexes that an individual protein is involved in are described together. In the Complex Portal, separate records are provided for each complex and so multiple complex entries may contain the same proteins, which is indeed the case for the LET-92-PAA-1 heterodimeric core. At the top of each Complex Portal entry page, the interactive graphical viewer shows the various components of a complex (Fig. 3B). When there is experimental evidence for the interaction between two components, a direct link is provided (Fig. 4B).

For LET-92- and PAA-1-containing complexes, there is no experimental confirmation for a direct physical interaction, and only genetic interaction studies are available. However, as the heterodimeric core is conserved between species, a direct connection between the catalytic and the constant subunits is inferred even though it is not supported by physical interaction data (Fig. 4B). Similarly, in six of the eight LET-92-containing complexes, there is no experimental evidence for the interaction between the various regulatory subunits and the heterodimeric core. However, as it is not always clear whether an orthologous complex with a functionally comparable regulatory subunit exists in other species, the subunit is included in the complex, but no binary interactions are presented between the subunit and core as without experimental data, it is not possible to predict which proteins in the core interact with the regulatory subunit. Complex names are assigned based on conservation between species.

### Table 2. Caenorhabditis elegans phosphatase complexes. For each complex member, the gene name and UniProt accession are provided.

| Phosphatase (catalytic subunit) | Regulatory subunits | Human homologues | Cross reference in Complex Portal | References |
|-------------------------------|---------------------|-------------------|----------------------------------|------------|
| LET-92 (G5EGK8)               | Constant            | Yes               | CPX-1357 RSA centrosome-targeting complex | [5,6]      |
|                               | PAA-1 (Q09543)      |                   | CPX-1361 PP2A-RSA-1 phosphatase complex   |            |
|                               |                     |                   | CPX-1366 PP2A-SUR-6 phosphatase complex   |            |
|                               | SUR-6 (G5EDR3)      |                   | CPX-1367 PP2A-PPTR-1 phosphatase complex |            |
|                               | RSA-1 (Q02217)      |                   | CPX-1368 PP2A-PPTR-2 phosphatase complex |            |
|                               | F43B10.1 (Q17744)   |                   | CPX-1373 PP2A-F43B10.1 phosphatase complex|            |
|                               | F47B8.3 (Q20519)    |                   | CPX-1375 PP2A-F47B8.3 phosphatase complex|            |
|                               | T22D1.5 (Q392H6)    |                   | CPX-1376 PP2A-T22D1.5 phosphatase complex|            |
|                               | Y71H2AM.20 (Q9BL35) |                   |                                   |            |
|                               | PPTR-1 (Q18178)     |                   |                                   |            |
|                               | PPTR-2 (A9UJN4)     |                   |                                   |            |
| TAX-6 (Q0G819)                | CNB-1 (G5EDN6)      | Yes               | CPX-1128 calcineurin–calmodulin complex | [38]       |
| PPH-5 (O9NE58)                | DAIF-21 (Q18868)    |                   | CPX-3983 Hsp90-cdc-37-pph-5 phosphatase complex | [69] |
| PPH-6 (Q09496)                | SAPS-1 (B2D6K9)     | Yes               | CPX-4025 pph-6-saps-1 phosphatase complex | [70]       |
| MTM-6 (G5ED68)                | MTM-9* (Q6EBBV8)    | Yes               | CPX-4024 MTM6-9* phosphatase complex | [23]       |
| FEM-2 (F49594)                | FEM-1 (P17221)      | Yes               | CPX-3385 FEM-2 phosphatase complex | [57,58] |
|                               | FEM-3 (P34691)      |                   | CPX-3383 CBC-fem-1 ubiquitin ligase complex |            |
| DAF-18 (G5EE01)               | ARR-1 (P51485)      | Yes               | CPX-3881 arr-1-mpz-1-daf-18 complex | [71]       |

| Pseudophosphatase | Human homologues | Cross reference in Complex Portal | References |
|-------------------|------------------|----------------------------------|------------|
| EGG-4* (O01767)   | No               | CPX-3381 Egg-3/4/5 MBK-2 complex | [61,24]    |
| EGG-3* (O20402)   |                  |                                  |            |
| EGG-5* (O61789)   |                  |                                  |            |

* Pseudophosphatase.
the PP2A complexes, typically both the common catalytic subunit and the specific regulatory subunit are included in the name. For example, the LET-92/PAA-1/RSA-1 complex is called PP2A-RSA-1 phosphatase complex (Complex Portal: CPX-1361). However, for complexes with multiple components, a functionally specific name may be assigned such as in the case of RSA centrosome-targeting complex (Complex Portal: CPX-1357) (Fig. 4B). The ‘Function’ and ‘Properties’ sections provide more information about the biological role and features of the complex as a whole. In contrast to UniProt, individual protein functions are only indicated when integral to the function of the complex as a whole. In the ‘Properties’ section of all the PP2A complexes, the same description of the heterodimeric core enzyme is given, and in the ‘Function’ section, a description of the roles of the regulatory subunits in terms of the overall function of the complex is provided. Therefore, in the Complex Portal, the number of individual complex records highlights the diverse functions of the regulatory subunits in the PP2A-containing complexes.

TAX-6: a calcium-dependent complex

As the previous LET-92 example shows, the regulatory subunits in phosphatase complexes exert multiple roles, which allow the complexes to have an involvement in a wide range of processes. In addition to the diverse roles indicated above, regulatory subunits can also regulate the activation of their catalytic binding partner. For instance, in response to external signals, regulatory subunits may undergo post-translational modifications or, in some cases, facilitate the recruitment of other components to the complex, which ultimately leads to the activation of the phosphatase. One example of this is the calcineurin complex.

The calcineurin–calmodulin complex consists of a catalytic subunit and a Ca^{2+}-binding regulatory subunit. This heterodimeric complex is conserved
throughout evolution. It is present in mammals and *D. melanogaster* as well as in *S. cerevisiae* [35–37]. Whilst in humans, the catalytic subunit is encoded by three genes (PPP3CA, PPP3CB and PPP3CC) and the regulatory subunit by two genes (PPP3R1 and PPP3R2), in *C. elegans* there is only one gene each for the catalytic (*tax-6*) subunit and the regulatory subunit (*cnb-1*; UniProtKB G5EDN6), which facilitates functional studies (Table 3). The catalytic subunit, also known as calcineurin A (CNA), is a Ca$^{2+}$-dependent, calmodulin-stimulated serine/threonine-protein phosphatase which plays an essential role in the transduction of intracellular Ca$^{2+}$-mediated signals [38]. In mammals, CNA is essential for immunity and for synaptic transmission in the brain [39–43]. However, studies in *C. elegans* show that CNA may have more widespread functions. Indeed, TAX-6 regulates a wide range of biological processes including egg laying, fertility, growth, lifespan, movement, sensory behaviours, and a specific set of endocytic processes such as coelomocyte endocytosis and synaptic vesicle recycling [44–51].

Figure 5A shows the current model for the assembly and regulation of the calcineurin–calmodulin complex based on experimental evidence from structural and biochemical studies across various species. In humans, the Ca$^{2+}$-induced complex is inferred based on experimental evidence [52]. In *C. elegans*, experimental evidence for the interaction is only available between TAX-6 and CNB-1, and for TAX-6 and calmodulin (CMD-1; UniProtKB O16305), and requires calcium [38,53]. As this complex is present in a range of
species, we assessed the evidence across species and inferred information to complete the characterisation. As in humans, the *C. elegans* catalytic subunit TAX-6 forms a stable association with the regulatory subunit CNB-1. CNB-1 contains 4 Ca\(^{2+}\)-binding sites, two with high affinity and two with low affinity. When

---

**Fig. 4.** Curation of the phosphatase complexes containing the LET-92 phosphatase, a member of the PP2A family. (A) Conservation of the PP2A complexes in *Saccharomyces cerevisiae*, *Homo sapiens* and *Caenorhabditis elegans*. The gene names for the catalytic, structural and regulatory subunits are shown. Association of the phosphatase (catalytic subunit) and constant scaffolding protein (structural subunit) with a variable regulatory subunit confers functional diversity. (B) Graphical view of the PP2A-RSA-1 phosphatase complex (Complex Portal: CPX-1361, left) and RSA centrosome-targeting complex (Complex Portal: CPX-1357, right) in the Complex Portal. The proteins are identified by circles and their corresponding UniProt entry name, for example ‘pp2a_caeel’, the gene name has been provided in brackets. The core of the complex (highlighted by a red square), common to the 2 complexes, is composed of the catalytic subunit LET-92 and the structural subunit PAA-1. Although no direct physical evidence confirms the interaction, the association is inferred based on orthology and genetic interaction studies both in UniProt and the Complex Portal. By contrast, direct association between the core and regulatory subunits has been demonstrated. In the Complex Portal direct links between complex components denote an association and this primarily based on direct physical evidence but also on genetic and/or orthology evidences. Complexes may be named according to their components (PP2A-RSA-1 phosphatase complex: CPX-1361) or based on function (RSA centrosome-targeting complex: CPX-1357). (C) In UniProt, the experimentally proven binary one-to-one interactions between the LET-92 phosphatase and the regulatory subunit (and vice versa) are described. (D) The ‘Function’ section of LET-92, RSA-1 and SUR-6 describes the individual roles of the protein components.
intracellular calcium levels are low, only the two high-affinity binding sites are occupied. Following an increase in intracellular Ca\(^{2+}\), the occupancy of the low-affinity sites on CNB-1 by Ca\(^{2+}\) causes a conformational change of the C-terminal regulatory domain of TAX-6, resulting in the exposure of the calmodulin-binding domain and in the partial activation of TAX-6 [38]. The subsequent binding of Ca\(^{2+}\)-bound CMD-1/calmodulin leads to the full displacement of the autoinhibitory domain from the active site and possibly of the autoinhibitory segment from the substrate binding site, which fully activates TAX-6.

In UniProt, the stable binary interactions between TAX-6 and CNB-1, and TAX-6 and CMD-1 are presented in the ‘Interaction’ section for the relevant protein entries (Fig. 5B). As this complex is only formed in response to an increase in intracellular Ca\(^{2+}\), the ‘Interaction’ section and ‘Activity regulation’ subsection of the TAX-6 UniProt protein entry contain a detailed description of how the induced complex is formed, how it is regulated by calcium, and how it leads to the activation of TAX-6. Interactions with small molecules such as calcium are indicated on a protein-by-protein basis (Fig. 5C). In the Complex Portal, all the stable and transient interactions between all proteins and all small molecules in the heterotrimeric complex are consolidated into a single calcineurin–calmodulin complex entry (Complex Portal: CPX-1128) (Fig. 5D). The assembly and the conditions and requirements for TAX-6 activation are described in the ‘Properties’ section (Fig. 5D). This demonstrates how the two databases display complementary information to provide a complete overview of the complex and its members.

**FEM-2: ferrying substrate to complexes**

The LET-92 and TAX-6 examples above show how single phosphatases can function in a range of processes by associating with different partners. In the next example, we describe the functional diversity achieved by the association of a phosphatase complex with another macromolecular complex, and explain how UniProt and the Complex Portal capture this information.

FEM-2 is a serine/threonine-protein phosphatase which belongs to the PP2C family [54–56]. In contrast to other members of this family, it contains a unique N-terminal domain bearing no similarity to any other protein domains, which functions as a scaffold to aid the interaction and assembly of macromolecular complexes [57]. At present, there are two experimentally verified complexes containing fem-2 (Fig. 6A). The first one, named the Fem-2 phosphatase complex (Complex Portal: CPX-3385), contains the phosphatase FEM-2 and two other proteins FEM-1 (UniProtKB P17221) and FEM-3 (UniProtKB P34691) (Fig. 6B) [57,58]. This complex associates with the CBC (cullin-2–elongin-B–elongin-C) E3 ubiquitin ligase complex to form a second bigger complex, the CBC-fem-1 ubiquitin ligase complex (Complex Portal: CPX-3383) (Fig. 6C) [58].

To date, the experimental data available show that FEM-2 associates with both FEM-1 and FEM-3 and that FEM-1, by interacting with ELC-1 (UniProtKB Q9BKS1), a component of the CBC E3 ubiquitin ligase complex, mediates the interaction between the two complexes (Fig. 6A) [57,58]. This macrocomplex is essential to control male sexual development in *C. elegans* [55,58]. Several models have been proposed to explain how the complex regulates this process. In one model, the FEM-2 phosphatase complex interacts with the zinc finger transcription factor TRA-1 (UniProtKB P34708), a sex determination regulator in *C. elegans* [58]. This association leads to the repression of the transcription of male-specific genes in somatic cells, and results in TRA-1 dephosphorylation. FEM-1 then associates with the CBC-fem-1 ubiquitin ligase complex and facilitates the ubiquitin-mediated degradation of TRA-1 (Fig. 6A).

UniProt presents the two experimentally proven FEM-2-containing complexes in the ‘Interaction’ sections of the individual protein entries for FEM-1, FEM-2 and FEM-3 (Fig. 6D). UniProt entries do not usually include substrates as a complex component due to their transient association. However, as TRA-1 appears to interact relatively stably with the complex prior to its degradation, it is included in the composition of the CBC-fem-1 ubiquitin ligase complex. By contrast, the Complex Portal is more stringent and does not present ligands as components of the complex unless they have been shown to be integral to the function of the complex. As TRA-1 acts as a substrate and is dispensable for the complex to function (other substrates may exist), it is not included in either of the Complex Portal entries for FEM-2-containing complexes (Fig. 6B). The Complex Portal does not standardly present more than one complex per entry. However, as the CBC-fem-1 ubiquitin ligase complex exerts its function on the TRA-1 substrate as a consequence of the activity of the Fem-2 phosphatase complex, this complex is also present on the entry page of the CBC-fem-1 ubiquitin ligase complex (Fig. 6B). Thus, by combining the information in UniProt and the Complex Portal, users have access to all binary
Table 3. Comparison of the phosphatase complex members in humans and Caenorhabditis elegans as of Complex Portal release 225. When known and/or available, orthologues/homologues for both species are provided. For the human homologues, their involvement in diseases is shown.

| Homo sapiens | UniProtKB AC | Gene name | Cross reference (Complex Portal) | Involvement in disease | UniProtKB AC | Gene name | Cross reference (Complex Portal) |
|--------------|--------------|-----------|----------------------------------|------------------------|--------------|-----------|----------------------------------|
| Catalytic    | Q08209       | PPP3CA    | CPX-1003; CPX-1048; CPX-1114; CPX-674 | Epileptic encephalopathy, infantile or early childhood, 1 (IECEE1) | Q0G819       | tax-6     | CPX-1128                          |
|              | P16298       | PPP3CB    | CPX-1002; CPX-1009; CPX-1116; CPX-998 |                         |              |           |                                  |
|              | P48454       | PPP3CC    | CPX-1001; CPX-1050; CPX-1112; CPX-1118 |                         |              |           |                                  |
| Regulatory   | P63098       | PPP3R1    | CPX-1001; CPX-1003; CPX-1009; CPX-1112; CPX-674; CPX-998 |                         | G5EDN6       | cnb-1     | CPX-1128                          |
|              | Q96LZ3       | PPP3R2    | CPX-1002; CPX-1048; CPX-1050; CPX-1114; CPX-1116; CPX-1118 |                         |              |           |                                  |
|              | P0DP23       | CALM1     | Ventricular tachycardia, catecholaminergic polymorphic, 4 (CPVT4), long QT syndrome 14 (LQT14) |                         | O16305       | cmd-1     | CPX-1128                          |
|              | P0DP24       | CALM2     | Long QT syndrome 15 (LQT15) |                         |              |           |                                  |
|              | P0DP25       | CALM3     |                         |                         |              |           |                                  |
| Catalytic    | P67775       | PPP2CA    |                         |                         | G5EGK8       | iet-92    | CPX-1357; CPX-1361; CPX-1366; CPX-1367; CPX-1368; CPX-1373; CPX-1375; CPX-1376 |
|              | P62714       | PPP2CB    |                         |                         | Q09543       | paa-1     | CPX-1357; CPX-1361; CPX-1366; CPX-1367; CPX-1368; CPX-1373; CPX-1375; CPX-1376 |
| Constant     | P30153       | PPP2R1A   | Mental retardation, autosomal dominant 36 (MRD36) |                         | Q00005       | rsa-1     | CPX-1357; CPX-1361 |
|              | P30154       | PPP2R1B   |                         |                         | Q9Y2T4       | F43B10.1  | CPX-1373                          |
|              | P63151       | PPP2R2A   | Spincerebellar ataxia 12 (SCA12) |                         | Q66LE6       | F47B8.3   | CPX-1375                          |
| Regulatory (B) | P00005     | PPP2R2B   |                         |                         | Q06190       | rsa-1     | CPX-1357; CPX-1361 |
|              | Q9Y2T4       | PPP2R2C   |                         |                         | Q996Q6       | T22D1.5   | CPX-1376                          |
| Regulatory (B') | Q00005    | PPP2R2D   |                         |                         | Q06190       | rsa-1     | CPX-1357; CPX-1361 |
|              | Q9Y5P8       | PPP2R3A   |                         |                         | Q17744       | F43B10.1  | CPX-1373                          |
|              | Q969Q6       | PPP2R3B   |                         |                         | Q20519       | F47B8.3   | CPX-1375                          |
|              | Q06190       | PPP2R3C   |                         |                         | Q9G2H6       | T22D1.5   | CPX-1376                          |
| Homo sapiens | | | | | C. elegans |
|---|---|---|---|---|---|
| SSU | UniProtKB AC | Gene name | Cross reference (Complex Portal) | Involvement in disease | UniProtKB AC | Gene name | Cross reference (Complex Portal) |
| Regulatory (B''') | Q15257 | PTPA (PPP2R4) | | Mental retardation, autosomal dominant 35 (MRD35) | O9BL35 | Y71H2AM.20 |
| Regulatory (B'') | Q15172 | PPP2R5A | | | O18178 | pptr-1 | CPFX-1367 |
| | Q15173 | PPP2R5B | | | A9UJN4 | pptr-2 | CPFX-1368 |
| | Q13362 | PPP2R5C | | | | |
| | Q14738 | PPP2R5D | | | | |
| | Q16537 | PPP2R5E | | | | |
| Catalytic | P60510 | PPP4C | CPX-156; CPX-1842; CPX-1843; CPX-1844 | | Q9XW79 | pph-4.1 |
| Regulatory | Q8TF05 | PPP4R1 | CPX-1842 | | Q9XT8 | pph-4.2 |
| | Q9NY27 | PPP4R2 | CPX-156; CPX-1843; CPX-1844 | | G5ECH5 | ppfr-1 |
| | Q6IN85 | PPP4R3A | CPX-1843 | | P91198 | ppfr-2 |
| | Q5MIZ7 | PPP4R3B | CPX-1844 | | | |
| | Q6ZMV5 | PPP4R3C | | | | |
| | P78318 | IGBP1 | Mental retardation, X-linked, syndromic, 28 (MRXS28) | | Q9N4E9 | ppfr-4 |
| Catalytic | Q8WY54 | PPM1E | | | P49594 | fem-2 | CPFX-3385 |
| | P49593 | PPM1F | | | | |
| Regulatory | Q9BSK4 | FEM1A | | | P17221 | fem-1 | CPFX-3385 |
| | Q96JP0 | FEM1C | | | | |
| | Q96M16 | PPM1M | | | | |
| Catalytic | Q13496 | MTM1 | | Myopathy, centronuclear, X-linked (CNMX) | P34691 | fem-3 | CPFX-3385 |
| | Q13613 | MTMR1 | | | Q9N589 | mtm-1 |
| | Q13614 | MTMR2 | | | | |
| | Q13615 | MTMR3 | | Charcot–Marie–Tooth disease 4B1 (CMT4B1) | Q22712 | mtm-3 |
| | Q9NY4A | MTMR4 | | | | |
| | Q9Y217 | MTMR6 | | | G5ED68 | mtm-6 | CPFX-4024 |
| | Q9Y216 | MTMR7 | | | | |
| | Q96EF0 | MTMR8 | | | | |
| | Q8NCE2 | MTMR14 | Myopathy, centronuclear, 1 (CNM1) | | | |
| Regulatory | Q95248 | SBF1 (MTMR5) | | Charcot–Marie–Tooth disease 4B3 (CMT4B3) | Q9TXP3 | mtm-5 |
| | Q96QG7 | MTMR9 | | | Q965W9 | mtm-9 | CPFX-4024 |
| | Q9NXD2 | MTMR10 | | | | |
| | A4FU01 | MTMR11 | | | | |
| | Q9C011 | MTMR12 | | | | |
| | Q86WG5 | SBF2 (MTMR13) | | Charcot–Marie–Tooth disease 4B2 (CMT4B2) | | | |
one-to-one interactions, the context for the interactions and substrate interaction data in the context of a complex.

**Egg-3/4/5/MBK-2 complex: a kinase trapping complex**

Phosphatases within complexes have dynamic roles. In the previous fem-2 example, we have seen that phosphatases are not only present in a complex to exert their catalytic activity, but they may also facilitate the delivery of a protein to another complex for degradation. In the next example, we use the Egg-3/4/5/MBK-2 complex to demonstrate how subcellular location dictates the function of a phosphatase complex (Fig. 7A).

The Egg-3/4/5/MBK-2 complex consists of three inactive phosphatases (or pseudophosphatases), EGG-3, EGG-4 and EGG-5, and the serine/threonine-protein kinase MBK-2 (UniProtKB Q9XTF3). Whilst MBK-2 is conserved throughout the animal kingdom in species such as yeast and human [59], the three pseudophosphatases appear to be specific to nematodes as no homologues exist in other species. Therefore, the Egg-3/4/5/MBK-2 complex is likely to be unique to *C. elegans* and other nematode species. A combination of yeast two-hybrid, coimmunoprecipitation and *in vitro* pull-down assays provide experimental evidence for the existence of the interactions between the participants [24,60,61]. Interestingly, the interaction of EGG-4 and EGG-5 with MBK-2 relies on EGG-4 and EGG-5 pseudophosphatases, which, despite the lack of catalytic activity [24], have the capacity to bind to phosphorylated proteins. This feature allows the pseudophosphatases to sequester MBK-2 and thus prevent MBK-2 from phosphorylating its substrates.

Some of the steps required for the assembly/disassembly of this complex have been elucidated based on the subcellular location of the different complex components in *C. elegans* lacking either egg-3, egg-4 or egg-5 (Fig. 7A). This complex is transiently formed in oocytes, and is required for the oocyte-to-zygote transition [62]. It localises to the cortex of the maturing oocyte and it is retained there until the end of meiosis I. Whilst EGG-4 and EGG-5 bind and inhibit MBK-2...
kinase activity, EGG-3 acts as a scaffold to tether MBK-2, EGG-4 and EGG-5 to the oocyte cortex, and thus prevents MBK-2 accessing its cytoplasmic substrates during meiosis I. During anaphase of meiosis I, EGG-3 relocates to the cytoplasm and is probably degraded by the anaphase-promoting complex/cyclosome (Complex Portal: CPX-3382), resulting in MBK-2 release from the complex. In UniProt, the information regarding the subcellular localisation of the individual protein components in this complex is described in the ‘Subcellular location’ section (Fig. 7B) and the role of each component is documented in the ‘Function’ section of each protein entry. In the Complex Portal, this localisation information is summarised in the ‘Function’ section as the subcellular localisation determines how the complex functions (Fig. 7C). Through the complementary data captured by each resource, users are provided with a summary of subcellular location at the level of the individual complex components as well as a description of the role of location of the complex in determining its activity.

Discussion
Phosphatases are largely signal-transducing molecules that control the transmission of intracellular signals by dephosphorylating molecules, reversing the action of cellular kinases. For most species, the phosphatome contains significantly fewer phosphatases than there are kinases in the kinome, and protein phosphatases...
have developed various mechanisms for ensuring the diversity of substrates they are capable of dephosphorylating. Using members of the *C. elegans* phosphatome as examples, we have shown how a phosphatase can specifically target a range of substrates depending on which regulatory subunit it associates with, how small molecules such as calcium can promote the formation of complexes, how a phosphatase complex functions together with a non-phosphatase-containing complex in order to achieve functional diversity, and how the subcellular location of a complex dictates its function. In other species, phosphatases that are part of complexes belong to various phosphatase superfamilies. In *C. elegans*, 13 out of the 16 complexes that have been characterised contain Ser/Thr phosphatases. Interestingly, even though 44% of phosphatases are tyrosine phosphatases, to date none have been experimentally proven to function within a complex.

The UniProt protein sequence database collates and organises protein-specific experimental and computationally predicted data. Expert curators analyse,
organise and categorise data to facilitate data retrieval, and are responsible for transferring key information from the scientific literature to the appropriate reviewed entries. It is incredibly rare, and arguably impossible, for a protein to function in sole isolation and without impacting another biological entity.
process or pathway. Thus, simply documenting the function of one biological entity in one record without considering interactors, protein networks and pathways is not feasible or indeed useful to scientific communities. UniProt acts as a central hub of knowledge, both describing interactions and pathways within entries, and also linking out to more detailed information in domain-specific resources such as model organism databases and molecular interaction resources such as the ImEx databases [63] for more details of binding partners. However, as described above, to fully understand the role of a phosphatase it is often necessary to comprehend how that protein operates in the context of a macromolecular complex of which it is a member. The Complex Portal supplies this additional level of detail, enabling a full picture of the way in which a phosphatase molecule may be regulated, may achieve functional diversity, and may dephosphorylate multiple substrates in a dynamic, but tightly controlled, series of events required for the fine control of cellular function.

Apart from the Egg-3/4/5/MBK-2 complex, all the complexes experimentally characterised in C. elegans are conserved in human (Table 3). Several of the phosphatases contained in these complexes are associated with diseases in human (Table 3). Therefore, mechanistic information on the regulation of phosphatase activity is becoming critical as drug development based on inhibiting/enhancing phosphatase activity is an actively growing field. Some phosphatase inhibitors are already available as licensed drugs, such as the immunosuppressants cyclosporin A and tacrolimus (FK506), which inhibit calcineurin activity [64]. The PP2A phosphatase complexes have been shown to be involved in several disease processes [65] and a range of drugs are already available, although not licensed yet, which inhibit some of its activity directly and indirectly by targeting its interactors and substrates [66]. Conversely, drugs with neuroprotective properties designed to prevent inhibition of PP2A in Alzheimer’s disease are also currently being developed [67,68].

We describe above a comprehensive review of the existing experimental literature pertaining to the C. elegans phosphatome by expert biocurators, and the capture of summary data into the UniProt and Complex Portal databases. Now that this proof-of-concept exercise has been completed for C. elegans, efforts need to switch to updating the human phosphatome and creating the human phosphatase complexosome. Both of these are much more challenging tasks due to the much higher number of catalytic molecules, regulatory subunits and also unanswerable questions as to the role of the many paralogous proteins which exist in the human proteome as a result of gene duplication. Ongoing efforts will also ensure that entries for both species are kept up to date for new experimental data as it appears in the literature, and also that key information is transferred to less well-studied, but closely related, organisms. For example, the annotation of C. elegans entries can be used to inform the study of parasitic worms which infect both humans and animals, causing weakness and disease of the host. The work of the expert biocurators is critical to support and enhance the work of laboratory researchers in the fields of both molecular mechanistic studies and, health and disease.

**Acknowledgements**

We thank Elena Speretta for the help in finding human homologue genes. UniProt has been prepared by Alex Bateman, Michele Magrane, Maria Martin, Sandra Orchard, Emily Bowler, Ramona Britto, Hema Bye-A-Jee, Penelope Garmiri, George Georghiou, Emma Hatton-Ellis, Yvonne Lussi, Alistair MacDougall, Elena Speretta, Nidhi Tyagi, Kate Warner, Roshana Zaru, Shadab Ahmed, Emanuele Alpi, Borisas Bursteinas, Leonardo Gonzales, Alexandr Ignatchenko, Giuseppe Insana, Rizwan Ishtiaq, Vishal Joshi, Dushyanth Jyothi, Jie Luo, Mahdi Mahmoudy, Andrew Nightingale, Joseph Onwubiko, Sangya Pundir, Guoying Qi, Daniel Rice, Rabie Saidi, Edward Turner, Preethi Vasudev, Vladimir Volynkin, Xavier Watkins and Hermann Zellner at the European Bioinformatics Institute; Alan Bridge, Lionel Breuza, Elisa Watkins and Hermann Zellner at the European Bioinformatics Institute; Hemanth Garge, George Georghiou, Leslie Arminski, Chuming Hulo, Nevila Hyka-Nouspikel, Florence Jungo, Philippe Le Mercier, Patrick Masson, Anne Morgat, Sandrine Pilbout, Catherine Rivoire, Christian Sigrist, Shyamala Sundaram, Parit Bansal, Delphine Baratin, Teresa Batista Neto, Jerven Bolleman, Beatrice Cuche, Edouard De Castro, Elisabeth Gasteiger, Sebastien Gehant, Arnaud Kerhornou, Thierry Lombardot and Monica Pozzo at the SIB Swiss Institute of Bioinformatics; and Cathy Wu, Cecilia Arighi, Hongzhan Huang, Peter McGarvey, Darren Natale, John S. Garavelli, Kati Laiho, Karen Ross, C. R. Vinayaka, Qinhua Wang, Lai-Su Yeh, Leslie Arminski, Chuming...
Chen, Yongxing Chen, Yuqi Wang and Jian Zhang at the Protein Information Resource. This work was supported by the National Eye Institute (NEI), National Human Genome Research Institute (NHGRI), National Heart, Lung, and Blood Institute (NHLBI), National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of General Medical Sciences (NIGMS) and National Institute of Mental Health (NIMH) of the National Institutes of Health under Award Number [U24HG007822]; Swiss Federal Government through the State Secretariat for Education, Research and Innovation; and European Molecular Biology Laboratory core funds.

Conflict of interest
The authors declare no conflict of interest.

Author contributions
HB and RZ performed biocuration and bioinformatic analyses and wrote the manuscript. SO and MM commented on the manuscript and supervised the work.

References
1 Sacco F, Perfetto L, Castagnoli L & Cesareni G (2012) The human phosphatase interactome: an intricate family portrait. FEBS Lett 586, 2732–2739.
2 Shi Y (2009) Serine/threonine phosphatases: mechanism through structure. Cell 139, 468–484.
3 Chen MJ, Dixon JE & Manning G (2017) Genomics and evolution of protein phosphatases. Sci Signal 10, eaag1796.
4 Patterson KJ, Brummer T, O’Brien PM & Daly RJ (2009) Dual-specificity phosphatases: critical regulators with diverse cellular targets. Biochem J 418, 475–489.
5 Song MH, Liu Y, Anderson DE, Jahng WJ & O’Connell KF (2011) Protein phosphatase 2A-SUR-6/B55 regulates centriole duplication in C. elegans by controlling the levels of centriole assembly factors. Dev Cell 20, 563–571.
6 Schlaitz A-L, Srayko M, Dammermann A, Quintin S, Wielisch N, MacLeod I, de Robillard Q, Zinke A, Yates JR, Müller-Reichert T et al. (2007) The C. elegans RSA complexlocalizes protein phosphatase 2A to centrosomes and regulates mitotic spindle assembly. Cell 128, 115–127.
7 den Hertog J (2003) Regulation of protein phosphatases in disease and behaviour. EMBO Rep 4, 1027–1032.
8 Bhullar KS, Lagarón NO, McGowan EM, Parmar I, Jha A, Hubbard BP & Rupasinghe HPV (2018) Kinase-targeted cancer therapies: progress, challenges and future directions. Mol Cancer 17, 48.
9 Lazo JS, McQueeny KE & Sharlow ER (2017) New approaches to difficult drug targets: the phosphatase story. SLAS Discov 22, 1071–1083.
10 Zhang M, Yogesh SD, Mayfield JE, Gill GN & Zhang Y (2013) Viewing serine/threonine protein phosphatases through the eyes of drug designers. FEBS J 280, 4739–4760.
11 Mullard A (2018) Phosphatases start shedding their stigma of undruggability. Nat Rev Drug Discov 17, 847–849.
12 McConnell JL & Wadzinski BE (2009) Targeting protein serine/threonine phosphatases for drug development. Mol Pharmacol 75, 1249–1261.
13 UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47, D506–D515.
14 Meldal BHM, Bye-A-Jee H, Gajdos L, Hammerová Z, Horácková A, Melicher F, Perfetto L, Pokorný D, Lopez MR, Túrková A et al. (2019) Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. Nucleic Acids Res 47, D550–D558.
15 Yadav L, Tamene F, Göös H, van Drogen A, Katalinen R, Aebersold R, Gstaiger M & Varjosalo M (2017) Systematic analysis of human protein phosphatase interactions and dynamics. Cell Syst 4, 430–444.e5.
16 Janssens V, Longin S & Goris J (2008) PP2A holoenzyme assembly: in cauda venenum (the sting is in the tail). Trends Biochem Sci 33, 113–121.
17 Duan G, Li X & Köhn M (2015) The human DEPhosphorylation database DEPOD: a 2015 update. Nucleic Acids Res 43, D531–D535.
18 Zarur M, Magrane M, O’Donovan C & UniProt Consortium (2017) From the research laboratory to the database: the Caenorhabditis elegans kinase in UniProtKB. Biochem J 474, 493–515.
19 Corsi AK, Wightman B & Chalfie M (2015) A transparent window into biology: a primer on Caenorhabditis elegans. WormBook 1–31.
20 Kim W, Underwood RS, Greenwald I & Shaye DD (2018) OrthoList 2: a new comparative genomic analysis of human and Caenorhabditis elegans genes. Genetics 210, 445–461.
21 Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, Davis P, Gao S, Grove CA, Kishore R et al. (2019) WormBase: a modern model organism information resource. Nucleic Acids Res 48, D762–D767.
22 El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A et al. (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47, D427–D432.
23 Dang H, Li Z, Skolnik EY & Fares H (2004) Disease-related myotubularins function in endocytic traffic in Caenorhabditis elegans. Mol Biol Cell 15, 189–196.
24 Cheng KC-C, Klancer R, Singson A & Seydoux G (2009) Regulation of MBK-2/DYRK by CDK-1 and the pseudophosphatases EGG-4 and EGG-5 during the oocyte-to-embryo transition. Cell 139, 560–572.

25 Xue Y, Fares H, Grant B, Li Z, Rose AM, Clark SG & Skolnik EY (2003) Genetic analysis of the myotubularin family of phosphatases in Caenorhabditis elegans. J Biol Chem 278, 34380–34386.

26 Watkins X, Garcia LJ, Pundir S, Martin MJ & UniProt Consortium (2017) ProtVista: visualization of protein sequence annotations. Bioinformatics 33, 2040–2041.

27 Reynhout S & Janssens V (1991) Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. Biochem J 353, 417–439.

28 Li Y-X & Dijkers PF (2015) Specific calcineurin Bγ determines substrate binding. Biochemistry 48, 1900–1910.

29 Bandyopadhyay J, Lee J, Lee J, Il Lee J, Yu J-R, Jee C, Cho J-H, Jung S, Lee MH, Zannoni S et al. (2002) Calcineurin, a calcium/calmodulin-dependent protein phosphatase, is involved in movement, fertility, egg laying, and growth in Caenorhabditis elegans. Mol Biol Cell 13, 3281–3293.

30 Lesch J, Garcia T, von Schwedler U, Lin Y, van der Bliek AM, Goldstein DB & Bao Y (2009) The proline-rich N-terminal sequence of calcineurin Aβ mediates the effect of calcineurin on life span in C. elegans. Autophagy 5, 604–607.
Hallem EA & Sternberg PW (2008) Acute carbon dioxide avoidance in Caenorhabditis elegans. Proc Natl Acad Sci USA 105, 8038–8043.

Kuhara A, Inada H, Katsura I & Mori I (2002) Negative regulation and gain control of sensory neurons by the C. elegans calcineurin TAX-6. Neuron 33, 751–763.

Li S-J, Wang J, Ma L, Lu C, Wang J, Wu J-W & Wang Z-X (2016) Cooperative autoinhibition and multi-level activation mechanisms of calcineurin. Cell Res 26, 336–349.

Shen X, Valencia CA, Gao W, Cotten SW, Dong B, Pilgrim D, McGregor A, J & Becker W (2012) Emerging role of DYRK family phosphatases that promote cell death during sensory neuron apoptosis. Cell Calcium 43, 444–456.

Huang B-C & Liu R (2008) Ca(2+)/calmodulin-binding phosphatases that promote cell death during sensory neuron apoptosis. Cell Calcium 43, 444–456.