The Repertoires of Ubiquitinating and Deubiquitinating Enzymes in Eukaryotic Genomes

Andrew Paul Hutchins,†,1 Shaq Liu,†,1 Diego Diez,1 and Diego Miranda-Saavedra†,*

†Bioinformatics and Genomics Laboratory, World Premier International (WPI) Immunology Frontier Research Center (IFReC), Osaka University, Osaka, Japan

†These authors contributed equally to this work.

*Corresponding author: E-mail: diego@ifrec.osaka-u.ac.jp.
Associate editor: Kenneth Wolfe

Abstract

Reversible protein ubiquitination regulates virtually all known cellular activities. Here, we present a quantitatively evaluated and broadly applicable method to predict eukaryotic ubiquitinating enzymes (UBE) and deubiquitinating enzymes (DUB) and its application to 50 distinct genomes belonging to four of the five major phylogenetic supergroups of eukaryotes: unikonts (including metazoans, fungi, choanozoa, and amoebozoan), excavates, chromalveolates, and plants. Our method relies on a collection of profile hidden Markov models, and we demonstrate its superior performance (coverage and classification accuracy >99%) by identifying approximately 25% and approximately 35% additional UBE and DUB genes in yeast and human, which had not been reported before. In yeast, we predict 85 UBE and 24 DUB genes, for 814 UBE and 107 DUB genes in the human genome. Most UBE and DUB families are present in all eukaryotic lineages, with plants and animals harboring massively enlarged repertoires of ubiquitin ligases. Unicellular organisms, on the other hand, typically harbor less than 300 UBEs and less than 40 DUBs per genome. Ninety-one UBE/DUB genes are orthologous across all four eukaryotic supergroups, and these likely represent a primordial core of enzymes of the ubiquitination system probably dating back to the first eukaryotes approximately 2 billion years ago. Our genome-wide predictions are available through the Database of Ubiquitinating and Deubiquitinating Enzymes (www.DUDE-db.org), where users can also perform advanced sequence and phylogenetic analyses and submit their own predictions.

Key words: ubiquitination, functional prediction, profile hidden Markov model.

Introduction

Protein ubiquitination is a reversible post-translational modification that regulates virtually all known cellular activities (Grabbe et al. 2011). In a hierarchical cascade, ubiquitinating enzymes (UBEs) catalyze the covalent addition of the 76 amino acid molecule ubiquitin to target proteins by the sequential action of a ubiquitin-activating enzyme (E1) and a ubiquitin-conjugating enzyme (E2) that selectively interacts with a ubiquitin ligase (E3) to control substrate specificity. Protein ubiquitination can be reversed by the action of deubiquitinating enzymes (DUBs). The ubiquitin signaling system is both functionally diverse and tightly regulated, and its dysregulation is responsible for a large number of conditions including many types of cancer, neurodegenerative, metabolic, and muscle wasting disorders (reviewed in Ciechanover and Schwartz 2004; Petroski 2008; Cohen and Tcherpakov 2010; Kirkin and Dikic 2011; Fulda et al. 2012). Moreover, viruses are known to employ a variety of strategies to manipulate the ubiquitin system to facilitate their replication and also encode their own viral UBEs and DUBs to evade the immune response (Randow and Lehner 2009).

The UBEs consist of three structurally and functionally different classes of enzymes (E1, E2, and E3): humans possess only two E1 genes and yeast only one. The E2s are a relatively small class of enzymes with 41 genes reported in humans and 14 in yeast and are characterized by the highly conserved ubiquitin-conjugating (UBC) domain (van Wijk and Timmers 2010). On the other hand, E3s are a large and structurally very diverse class of enzymes and have been classified into three main families (reviewed in Ardley and Robinson 2005): HECT (homologous to E6-associated protein C-terminus), RING (really interesting new gene), and U-box (a type of modified RING motif). The HECT are E3s with catalytic activity as the HECT domain contains a central cysteine residue that acts as an acceptor for ubiquitin. Neither RING nor U-box E3s have a direct role in catalysis and function as adaptor-like molecules to facilitate protein ubiquitination while needing the presence of E2s and other proteins. RING finger E3s are by far the largest family of E3s and are characterized by the presence of a RING finger domain (a cysteine/histidine-rich, zinc-chelating domain that promotes both protein–protein and protein–DNA interactions). Besides the simple E3s containing a RING finger domain (“RING domain”), two classes of complex RING E3s have been described: 1) the RBR (RING in between RING–RING), which contain an IBR domain (“In Between Ring fingers”) and two RING domains and are capable of E2 binding and...
substrate recognition. A recently proposed mechanism for RBR ubiquitin transfer suggests that RBR ligases combine features of both RING- and HECT-type ligases, with independent catalytic activity (Wenzel and Klevit 2012). 2) The CRL (Cullin–RING ligases), large multiprotein complexes comprising at least a cullin protein and a RING protein. The cullin protein provides the central scaffolding component for each of the complexes. The main CRL families are the F-box, the BTB (Broad complex, Tramtrak, Bric-a-brac), and the SOCS (Suppressor Of Cytokine Signaling). The third family of E3s is the U-box, characterized by their approximately 75 amino acid (U-box) domain. The U-box is structurally similar to the RING finger domain but lacks the full complement of zinc-binding ligands required for metal chelation. Initially, all U-box E3s were termed “E4” as they were thought to be proteins auxiliary to E2–E3 complexes and involved in ubiquitin chain elongation. However, it was later recognized that other U-box E3s actually interact with E2s and present E3 ligase activity independently of other E3s. Their function as E3s or E4s may depend not only on the nature of the substrate but also on the presence of other proteins in the multiprotein complex. Finally, there is a very small set of E3s that cannot be classified as HECT, RING finger, or U-box E3s, and these comprise the ZnF A20 and DDB1-like families.

The DUBs are a smaller enzyme set than the UBEs, and mechanistically all DUBs are either cysteine or metalloproteases. DUBs have been classified into five distinct families: the JAMM motif proteases (JAMM), the Machado–Joseph Disease protein domain proteases (MJD), the ovarian tumor proteases (OTU), the ubiquitin C-terminal hydrolases, and the ubiquitin-specific proteases (USP) (Nijman et al. 2005).

Despite the central importance that reversible protein ubiquitination plays in cellular signaling, the UBE and DUB repertoires of almost all genomes remain uncharted. Here, we describe a highly sensitive and specific sequence analysis method for the automatic classification of UBEs and DUBs into their corresponding families (N.B. our method does not attempt to predict large suites of cullin adapters, such as F-box-LRR or MATH-BTB). First, we describe the method in detail and a series of tests to evaluate its performance. Second, we also provide proof of concept of the general applicability of the method by scanning 50 completely sequenced eukaryotic genomes. General aspects of UBE/DUB genomic content and conservation of function are discussed across four eukaryotic supergroups. We chose a wide spectrum of species covering most of the sequenced eukaryotic phylogenetic diversity, including unikonts (metazoans, fungi, choanozoa, and amoebozoa), plants, excavates, and chromalveolates. Finally, we present the Database of Ubiquitinating and Deubiquitinating Enzymes (DUDE-db), a comprehensive repository of UBE and DUB. DUDE-db is publicly available at http://www.DUDE-db.org and features an annotated database of predicted UBEs and DUBs in 50 eukaryotic genomes, and a web server interface where users can scan a set of proteins for the presence of sequence domains diagnostic for any of the UBE and DUB families.

Results

A Novel Method for the Automatic Classification of UBEs and DUBs into Families

We have developed an automatic and generally applicable method for the systematic retrieval and classification of UBEs and DUBs into families (fig. 1). Briefly, our method is based on a collection of 33 specifically selected profile hidden Markov models (HMMs) from the Pfam (Punta et al. 2012), SMART (Letunic et al. 2012), and SUPERFAMILY (Wilson et al. 2009) databases. The selection of HMMs was modeled on the members of the various UBE and DUB families and subfamilies of the human genome. The evaluation for sensitivity was done on the manually annotated yeast UBE/DUB repertoire and on a phylogenetically diverse set of experimentally validated UBEs and DUBs from the UniProt database. These tests reported a coverage of more than 99% for the "UBE/DUB HMM Library." Additionally, the correct classification rate on the family level was 100% for both the yeast and UniProt data sets. Therefore, we expect the UBE/DUB HMM Library to retrieve more than 99% of proteins encoding UBEs and DUBs, which will likely be correctly assigned to specific UBE/DUB protein families in all cases. It must be noted that our method relies on the "trusted cut-offs" of the distinct HMMs as implemented in InterProScan (Zdobnov and Apweiler 2001). These cutoffs are those recommended by the individual InterPro member databases and therefore are believed to report relevant matches. Other than this, no test for false positives was implemented as no gold standard exists yet for the correct classification of UBEs and DUBs.

To show the wide applicability of the UBE/DUB HMM Library, we first scanned the yeast and human genomes and present an extended repertoire of UBE and DUB genes that have remained uncharacterized in these model organisms. We then extend our analyses to 50 eukaryotic genomes, which are made available through the DUDE-db.

The Enlarged Repertoire of UBE and DUB in Human and Yeast

We applied the UBE/DUB HMM Library to scan the yeast and human genomes for UBEs and DUBs, the only two species for which the genomic UBE/DUB complements have been reported. Analysis of the yeast genome (Ensembl 67) predicted 85 UBE genes (of which 72 had been characterized previously) and 24 DUB genes (of which 17 were already known) (table 1), thus increasing the yeast repertoire of UBEs and DUBs by approximately 25% (fig. 2A). Most importantly, we describe two and four new members of the OTU and JAMM families of DUBs that were not reported in the yeast catalog. At 109 genes, the yeast complement of UBEs and DUBs is one-ninth the size of the human complement by gene number, but more than two-thirds of the yeast enzymes have an identifiable human ortholog. Moreover, of the 22 yeast genes whose genetic deletion produces an inviable phenotype, 20 are conserved in human (supplementary table S1, Supplementary Material online). This suggests that yeast might be a very
useful model organism with which to investigate the functions of a number of human UBEs and DUBs.

Similarly, we predicted 2,593 UBEs and 392 DUBs in the human genome, encoded by 814 UBE and 107 DUB genes. This figure represents an approximately 35% increase in the number of UBE and DUB genes in the human genome in comparison to the study by Li et al. (2008) (fig. 2A) and brings the full human UBE/DUB complement to a new total of 921 genes (table 1). Although yeast has an approximately similar number of UBEs and protein kinases (Miranda-Saavedra et al. 2007), humans possess 65% more genes encoding UBEs than protein kinases. A comparison of the yeast and human UBE/DUB complements shows that enzymes harboring an RNF domain make up approximately 50% of the entire UBE/DUB repertoire and that genes of the families HECT, RNF domain, and F-box are present in approximately equal proportions in both organisms (fig. 2B). Finally, humans possess three families that are absent from yeast: SOCS, ZnF A20, and the DUB family MJD (fig. 2B).

We report seven genes in the human genome that have multiple distinctive domains and so fall into a new category
that we term “Mixed-family” (fig. 2C). These enzymes harbor various combinations of domains diagnostic for distinct UBE/DUB families, including OTU/ZnF A20, RNF/HECT, U-box/RNF, and Cullin/RNF (fig. 2C). Mixed-family enzymes are found in limited numbers in many of the species surveyed but are particularly abundant in metazoans and absent in yeasts and excavates. Even though mixed-family enzymes represent an almost negligible portion of any genome, they illustrate the flexibility of the ubiquitination system to produce protein machines that combine specific functions in new and creative ways.

The UBE/DUB Complements of 50 Eukaryotic Genomes

We applied our method on 50 completely sequenced eukaryotic genomes belonging to four eukaryotic supergroups as defined by Keeling et al. (2005): unikonts (including metazoans, Table 1.

| Species Class Family Genes Genes |
|----------------|----------------|----------------|
| Yeast E1 E1 | 1 | 3 |
| E2 E2 | 11 | 14 |
| E3 | HECT | 5 | 5 |
| | RING/RNF domain | 39 | 43 |
| | RING/BTB | 3 | 6 |
| | RING/SOCS box | 0 | 0 |
| | RING/F-box | 10 | 10 |
| | U-box | 2 | 2 |
| | Other/ZnF A20 | 0 | 0 |
| | Other/DDB1-like | 2 | 2 |
| | Mixed family | 0 | 0 |
| | JAMM | 0 | 4 |
| | MJD | 0 | 0 |
| | OUT | 0 | 2 |
| | UCH | 1 | 1 |
| | USP | 16 | 17 |
| Total | 90 | 109 |

| Human E1 E1 | 2 | 5* (12) |
| E2 E2 | 32 | 43 (178) |
| E3 | HECT | 26 | 28 (104) |
| | RING/RNF domain | 271 | 407 (1,319) |
| | RING/BTB | 162 | 199 (592) |
| | RING/SOCS box | 35 | 39 (96) |
| | RING/F-box | 56 | 73 (209) |
| | U-box | 7 | 5 (20) |
| | Other/ZnF A20 | 6 | 4 (35) |
| | Other/DDB1-like | 3 | 4 (12) |
| | Mixed family | 0 | 7 (18) |
| | JAMM | 12 | 12 (45) |
| | MJD | 4 | 4 (45) |
| | OUT | 13 | 13 (41) |
| | UCH | 4 | 4 (25) |
| | USP | 51 | 74 (236) |
| Total | 684 | 921 (2,985) |

**Note.**—UCH, ubiquitin C-terminal hydrolases. Our count of five and three E1s in human and yeast, respectively, does not reflect an increase in the number of ubiquitin-activating genes, rather it reflects the sequence identity shared with the three other ubiquitin-like activating enzymes, namely (in human) UBEA2 (SUMO pathway), UBEA3 (NEDD8 pathway), and UBEA7 (ISG15 pathway).

The UBE/DUB Complements of 50 Eukaryotic Genomes

We applied our method on 50 completely sequenced eukaryotic genomes belonging to four eukaryotic supergroups as defined by Keeling et al. (2005): unikonts (including metazoans,
choanozoa, fungi, and amoebozoa), excavates, chromalveolates, and plants. These genomes display a wide range of genome sizes and adaptations to their environments, including important human pathogens such as *Trypanosoma brucei* and *Leishmania major*. If we accept the classification as presented here, an enormous variation in the sizes of eukaryotic UBE/DUB complements exists, ranging from 37 UBEs and 7 DUBs for the intracellular microsporian parasite *Encephalitozoon cuniculi* to 2,593 UBEs and 392 DUBs in human (supplementary table S2, Supplementary Material online). These numbers, however, represent predicted proteins (including splice isoforms) and not necessarily individual genes. Therefore, the number of genes encoding UBEs and DUBs per genome is likely to be smaller, especially for metazoan species. Examination of the phylogenetic distribution of UBEs and DUBs (fig. 3A and supplementary table S2, Supplementary Material online) shows that all E1, E2, and E3 families, with the exception of SOCS and ZnF A20, are universally distributed across the four eukaryotic supergroups. SOCS family E3s are found in all metazoans, and their absence from other taxa, especially the choanoflagellate *Monosiga brevicollis* (the closest outgroup to metazoa), suggests that SOCSs are exclusively a metazoan innovation. Similarly, all DUB families are universally found in all phylogenetic lineages, except for the MJD proteases, which are absent from excavates.

A linear correlation exists between the number of proteins encoded in a genome and the size of its UBE/DUB complement (fig. 3C), with the proportion of genes encoding UBEs/DUBs ranging from 5.3% (*Arabidopsis thaliana*) to 1.2% (*Phytophthora sojae*). Only 1.7% of human genes encode UBE/DUB, a figure similar to that of protein kinases (Martin et al. 2009). Unicellular organisms typically have less than 300 UBEs and less than 40 DUBs, whereas the more complex (plant and animal) multicellular organisms possess massively enlarged repertoires of E3s and to some extent larger E2 and E1 families.
DUB complements too (fig. 4 and supplementary table S2, Supplementary Material online). Therefore, it appears that organismal complexity roughly correlates with an increase in the complexity of the ubiquitination system.

Because UBEs and DUBs are found in all eukaryotic lineages explored, we interrogated the OrthoMCL database (Chen et al. 2006) for orthologous genes across the four eukaryotic supergroups. OrthoMCL is a sensitive genome-scale algorithm for grouping orthologous protein sequences, with release 5 featuring approximately 1.4 million sequences from 150 genomes (including 57 unikonts, 13 excavates, 11 plants, 17 chromalveolates, and 52 bacteria). We found that 91 genes are orthologous across all four eukaryotic supergroups (i.e., orthologs are present in at least one genome from each supergroup) (fig. 5). These 91 genes represent an essential core of UBEs and DUBs that probably dates back to the first eu- karyotes approximately 2 billion years ago. However, some of these genes appear to have been lost secondarily in specific lineages, especially excavates and chromalveolates, two lineages that include many parasitic species. This suggests that their likely essential functions have been taken over by other enzymes or that the original enzymes have evolved beyond recognition. If we select only those genes that are present in at least 75% of the genomes of each eukaryotic supergroup, a

**Fig. 3.** Evolutionary conservation of the distinct UBE and DUB families in four eukaryotic supergroups. (A) All UBE families are present in all eukaryotic genomes surveyed, except for SOCS (metazoan specific) and ZnF A20 (absent in excavates and fungi). All DUB families are universally present in eukaryotes, except for the MJD DUBs, which appear to have been lost from excavates. (B) Evolutionary tree displaying the relationships among the various eukaryotic groups included in this study, with the number of species indicated in parentheses. The area of the circles is proportional to the total number of sequences identified for each group, which are also shown. (C) In eukaryotic genomes a linear relationship exists between the number of UBE/DUB and the total number of predicted proteins.
**Fig. 4.** Bar plot of the UBE and DUB complements of 50 eukaryotic genomes split into E1, E2, E3, and DUB.
Fig. 5. Circle plots showing the presence of the 91 orthologous UBE/DUB genes found in all four eukaryotic supergroups. A fully colored circle indicates that 100% of the species included in that eukaryotic supergroup possess an identifiable ortholog (see inset key). Four genes encoding cullins also appear to be largely conserved in all four eukaryotic supergroups.
subset of 18 genes is identified, including 12 UBE and 6 DUB genes. Figure 5 lists the human genes associated with these 91 orthologous groups (with the 18 “ultra conserved” groups underlined). Examination of the mouse orthologs in the Mouse Genome Database (Eppig et al. 2012) shows that the genetic deletion of the majority of these genes produces embryonic lethal phenotypes or leads to serious conditions later in life (e.g., abnormal function and development of the nervous system, inflammatory diseases, infertility, and major nuclear organization defects leading to abnormal chromosome condensation and segregation).

The DUDE-db

Our method and predictions are accessible via an online web resource (http://www.DUDE-db.org/). DUDE-db contains the precomputed predictions for 50 eukaryotic genomes, including 35,228 distinct UBEs and DUBs proteins, which have been complemented by including proteins encoding the important scaffolding proteins of the “Cullin” and “VHL” families in the same 50 genomes. Cullin and VHL proteins were identified by the presence of specific protein domains as provided by high-quality models from the Pfam, SMART, and SUPERFAMILY databases (Cullin: PF00888 [IPR001373] and SM00182/SSF75632 [IPR016158]; VHL: PF01847/SSF49468 [IPR022772]). This brings the total number of proteins in DUDE-db to 35,778. UBEs constitute approximately 87% of all database entries and are mainly composed of E3s (~81%), followed by E2s (~6%) and E1s (<1%). Among the E3s, the most abundant family is the RNF domain, making up approximately 50% of all E3s, followed by BTB enzymes (~23%) and F-box enzymes (~16%). DUBs represent approximately 11% of all entries, with USPs being the main DUB family (~57% of all DUBs). The scaffolding proteins of the Cullin and VHL families represent a mere 1.5% of all entries. Finally, enzymes belonging to more than one UBE/DUB family (the “Mixed-family” category) account for only 0.4% of the data set. The annotation of the proteins in DUDE-db was enriched with the information provided in the original genome releases, including not only the original protein ID but also the gene ID each protein maps to, the annotated gene names and gene descriptions, and the protein sequence.

DUDE-db provides a comprehensive interface for accessing the database: The “Search database” tab allows the user to browse the database by keyword (that will match sequence and gene IDs, gene names, and gene descriptions) or by any combination of UBE/DUB family and species (fig. 6). The results are tabulated in an output HTML page, detailing not only the original protein IDs and their family classification but also their associated gene IDs, gene names and gene descriptions, and a mini-plot of the protein domain architecture. By clicking on the mini-plot, a full-size image of the protein domain architecture will be displayed (fig. 6). The user is given the option to download the data sets, either in the “text-only” version (fully annotated protein sequences in FASTA format) or also including full-size protein domain architecture plots for each protein. The “Start Jalview” button allows the user to launch a Java applet of the multiple sequence analysis tool Jalview (Waterhouse et al. 2009) to visualize the query results. Using Jalview, the user can easily edit and color the sequences by conservation, protein secondary structural properties, or amino acid chemical characteristics and perform on-the-fly calculations of phylogenetic trees (Neighbor-Joining and average distance). The full Jalview application can be readily launched via the “File→View in Full Application” option. This provides access to various multiple sequence alignment algorithms (MAFFT, Muscle, ClustalW, T-Coffee, and Probcons), secondary structure prediction by JNet, and structure display with Jmol. Furthermore, multiple alignments in Jalview can now be exported to TOPAli v2 (Milne et al. 2009) via a synchronized interface where the user can access more sophisticated methods of sequence evolutionary analysis. Through the “Peptide scan” interface, users can submit their own sequences for prediction, either by “copying and pasting” the sequences in a text box or uploading them from a local file. The input sequences are first subjected to basic quality assurance checks before submission to a multi-node Linux cluster. In the meantime, the user can also check the job’s running status and even submit a second set of sequences for scanning, while the first job is still in progress. The user is asked to provide a job name and e-mail address where a hyperlink with the results will be sent upon job submission. The results page reports those proteins that are predicted to be UBEs, DUBs, or scaffolding proteins (Cullin/VHL), and their corresponding domain architecture plots.

Discussion

Protein ubiquitination is a fundamental mechanism controlling virtually all known cellular functions in complex ways that we are only beginning to understand (Grabbe et al. 2011). Despite the importance of ubiquitination in normal physiological and disease states, partial catalogs of these important enzymes only existed for a few select species. We report here a broadly applicable computational method to predict UBEs and DUBs on the basis of identifying protein domains diagnostic of the various UBE/DUB families. Our method relies on HMMs specifically selected from three distinct protein domain databases (Pfam, SMART, and SUPERFAMILY), and upon evaluation and manual examination of the results, we report a coverage and a family-level classification accuracy greater than 99%. This means that using the UBE/DUB HMM Library, we expect not only to retrieve more than 99% of these enzymes from any genome but also that these enzymes will be correctly assigned to a specific protein family in all cases. HMMs are known to be both especially sensitive for database searches and highly specific for the classification of protein sequences into specific groups, as we previously showed for the protein kinase superfamily (Miranda-Saavedra and Barton 2007). The application of our method on the previously reported UBE/DUB sets of yeast and human allowed us to identify proteins that most likely are annotation mistakes and most importantly to find new genes encoding UBEs and DUBs that had not been reported before. In human, we report an additional 35% genes encoding UBEs and DUBs, and in the well-studied Baker’s yeast an additional 25%
thus highlighting the power of our method for characterizing the complements of UBEs and DUBs genome wide. Our predictions on 50 eukaryotic genomes cover the major phylogenetic supergroups of eukaryotes and are now available through DUDE-db, a comprehensive database of UBEs and DUBs. DUDE-db also features an easy-to-use prediction server where users can scan their protein data sets. Therefore, our prediction method and its web interface, DUDE-db, represent a significant advance in the field and fill an essential void for researchers studying protein ubiquitination. DUDE-db will be updated periodically to incorporate additional predicted and manually annotated UBE/DUB complements, as well as predictions of proteins with ubiquitin-binding domains (Dikic et al. 2009). Recently, a database of UBE and DUB covering a limited portion of the eukaryotic tree was published (Gao et al. 2013). The authors used the sequences of
Protein ubiquitination and phosphorylation are similar in that both types of chemical modification can be reversed (by deubiquitinases and phosphatases, respectively) and specifically recognized (by phosphoprotein- and ubiquitin-binding domains), both can be autoregulated (by auto-ubiquitination and autophosphorylation), and both can be induced by various stimuli. Moreover, both systems are tightly interlinked because ubiquitination is often dependent on protein phosphorylation (e.g., to create binding sites for E3s or to modulate E3 activity), and many protein kinases are known to be ubiquitinated as exquisitely illustrated for the EGF-receptor signaling network (Woelk et al. 2007; Argenzio et al. 2011).

The ubiquitination system is larger by number of enzymes than the phosphorylation system, with the added complexity that all seven lysines of the ubiquitin molecule can be ubiquitinated, giving rise to poly-ubiquitinated and branched poly-ubiquitinated chains. All these chemical modifications introduce topologies that encode functionally distinct signals that are recognized by cells and affect not only the half-lives of target proteins but also their structure and activity, localization, and interaction partners (Woelk et al. 2007). The sheer size of ubiquitin chains has hindered their identification by quantitative proteomics, but improvements in novel controlled protein fragmentation techniques, together with more sensitive and accurate mass spectrometers, will eventually enable the site-specific identification of ubiquitin modifications with the same degree of detail of acetylation and phosphorylation, on a proteome-wide scale (Vertegaal 2011).

Coupling mass spectrometry to experiments in genetically modified cells will lead to understanding the involvement of specific E2/E3 pairs in regulating specific targets and cellular processes in the dynamic context of other post-translational modifications (Nagy and Dikic 2010) and how the subversion of the system contributes to disease. Abnormal ubiquitination has been implicated in dozens of diseases (Ciechanover and Schwartz 2004; Petrock 2008; Cohen and Tcherpakov 2010; Kirkin and Dikic 2011; Fulda et al. 2012), and UBEs and DUBs could become the next large set of drug targets for either discontinued genes (USP17-like, USP51, DUB-3, ENSG00000197767, IFP38, and ENSG00000198817) or pseudogenes (TL132, TL132-like, and PARP11). This produced a set of 86 genes encoding DUBs in human.

Protein kinases are such a successful group of targets because we have a thorough understanding of their druggability principles as most kinase inhibitors target the ATP-binding pocket. As a result, chemical libraries can be synthesized and tested directly against panels of kinases. On the other hand, ubiquitination is more complex and versatile than phosphorylation as a control mechanism, and as such, the potential for therapeutic intervention is enormous. The only commercially available inhibitor of the ubiquitin system (Bortezomib) targets the 26S proteasome, but recently some E3 inhibitors have entered clinical trials (reviewed in Cohen and Tcherpakov 2010). The development of general principles for identifying inhibitors against E3s, E2/E3 pairs, or DUBs will give birth to chemical libraries that can be tested against enzymes of the ubiquitin system. In this context, DUDE-db provides not only a data set of predicted UBEs and DUBs in humans, but also an essential toolkit for pharmacophylogenomic analysis (Searls 2003). Phylogenetic reconstructions including a large number of species make more reliable ortholog and paralog calls. These in turn can help identify functional shifts (orthologs) and broader issues of coevolution, pleiotropy, and functional redundancy, all of which are essential considerations when reasoning about species differences and to identify appropriate drug targets and disease models for drug development pipelines.

**Materials and Methods**

**Systematic Prediction of UBEs and DUBs**

The method presented here relies on distinguishing among the various eukaryotic UBE and DUB families by virtue of diagnostic protein domain signatures. We aimed to characterize the specific combination of protein domain signatures from the smallest number of protein domain databases that allows the identification and classification of characterized sets of UBEs and DUBs into their correct families, without cross-hitting other families (fig. 1).

**Working Data Sets**

A data set of 2 human E1 and 32 E2 genes was previously described by van Wijk and Timmers (2010). E1s and E2s are characterized by ubiquitin-activating and UBC domains, respectively. E3s constitute a much larger and diverse group of enzymes: In 2008, Li et al. reported a catalog of 616 distinct human E3s representing individual human genes and classified into the HECT, RING finger, U-box, ZnF A20, and DDB1-like families. Most of these sequences had been predicted by sequence similarity to other characterized ligases. The RefSeq protein IDs provided for these E3s were mapped to the human RefSeq database (release of July 2012) (Pruitt et al. 2009). The original proteins could be retrieved for 598/616 RefSeq IDs. This E3 data set was curated by reannotating duplicate entries (NP_055763 as “RING finger” and NP_694578 as “BTB”). Next, to obtain a final, nonredundant gene data set, RefSeq IDs were mapped to the human Ensembl gene set (release 67) (Flicek et al. 2012). Nonmatches and double annotations to human Ensembl genes were curated to produce a set of 2 E1, 32 E2, and 592 E3 genes in the human genome.

The human complement of DUBs was previously reported to consist of 95 genes (Nijman et al. 2005). Examination of this data set led to the removal of nine DUBs that were either discontinued genes (USP17-like, USP51, DUB-3, ENSG00000197767, IFP38, and ENSG00000198817) or pseudogenes (TL132, TL132-like, and PARP11). This produced a set of 86 genes encoding DUBs in human.

**Characterization of Protein Domain Signatures Specific to UBE and DUB**

The sets of human UBEs and DUBs described above were analyzed with a local installation of InterProScan.
Prediction of Ubiquitinating and Deubiquitinating Enzymes.

Automatically integrates trusted cutoffs for each protein individually with the added advantage that InterPro makes the most of their regions of the protein space. Thus, by combining various protein signature databases, InterPro makes the most of their individual strengths with the added advantage that InterPro automatically integrates trusted cutoffs for each protein domain model. We found that 26/592 (~4.5%) of the E3s reported by Li et al. (2008) do not harbor any protein domains characteristic of any E3 family nor experimental evidence for their catalytic activity is reported and were therefore disregarded. For these same reasons, 2/86 DUBs from the Nijman et al. data set (Nijman et al. 2005) were also removed, producing a final superset (“the working dataset”) of 600 UBEs (2 E1, 32 E2, and 566 E3 genes) and 84 DUB genes in the human genome (supplementary table S3, Supplementary Material online).

Analysis of the protein domain signatures of the working data set led to the identification of a minimal group of 33 distinct protein domain signatures that allows the automatic retrieval and correct family-level classification of all the enzymes in the working data set (supplementary table S4, Supplementary Material online). We call this collection of HMMs the “UBE/DUB HMM Library.” We chose to combine HMMs from Pfam (Punta et al. 2012), SMART (Letunic et al. 2012), and SUPERFAMILY (Wilson et al. 2009) as these databases differ in their contents and therefore their coverage. Pfam (release 26.0) is a large collection of more than 13,000 protein family alignments, whereas SMART (version 7) contains models for more than 1,000 protein domains. Although Pfam and SMART are built from manually curated alignments of multiple protein sequences, SUPERFAMILY is based on the sequences of domains with known three-dimensional structure as contained in the SCOP database (Andreeva et al. 2008). Therefore, the integration of HMMs from the three databases leads to an improvement in the predictive power of the method when compared with using any individual database in isolation. In InterPro, each HMM is usually integrated into an “InterPro entry” that typically includes protein and domain models not only from the Pfam, SMART, and SUPERFAMILY databases but also from other InterPro member databases. In fact, many of the 33 HMMs selected above are part of InterPro entries that feature additional models from other InterPro member databases, such as Gene3D, PANTHER, PRINTS, and PROSITE. We proved empirically that including any other protein domain model from other InterPro member databases in addition to the 33 HMMs of the UBE/DUB HMM Library (supplementary table S4, Supplementary Material online) did not improve the coverage on the annotated set of UniProt proteins associated with each InterPro entry. Therefore, the select combination of HMMs from the Pfam, SMART, and SUPERFAMILY databases that makes up the UBE/DUB HMM Library represents both a minimal and comprehensive collection of protein domain models that cover all the protein space associated with each one of the InterPro entries they are featured in.

Evaluation

To evaluate the coverage and accuracy of our method, we performed a series of tests on sequences that had been manually annotated as UBEs and DUBs, and for many of which, experimental evidence is available. In a first exercise, we attempted to classify the manually annotated sets of UBEs and DUBs from the model yeast *Saccharomyces cerevisiae*. The set of E1s and E2s was retrieved from the *S. cerevisiae* Ubiquitination Database (SCUD) (Lee et al. 2008). Yeast E3s had previously been characterized by Li et al. (2008), and the DUBs were those reported by Amerik et al. (2000). According to these annotated data sets, the yeast UBE/DUB complement consists of 109 genes, including E1 (1 gene), E2 (11), E3 (80), and DUB (17). Our method retrieved 89/109 proteins (81.6%), with a correct classification rate of 100% on the family level (table 2). A closer inspection of the proteins that we failed to identify unveiled the incorrect annotation of 19 proteins. These include the following eight RNF genes: ASI1 (YMR119W) and ASI3 (YNL008C) are inner nuclear membrane proteins, and although the authors suggested the presence of nucleoplasmically oriented C-terminal RING domains, they could not show auto- or transubiquitination activity (Zargari et al. 2007). Moreover, not a single protein domain could be identified in these proteins using InterProScan. Similarly, SLX5 (YDL013W) has no identifiable

| Species | Class | Family | Genes | Coverage (%) | Family-Level Classification Accuracy on Data Set Covered (%) |
|---------|-------|--------|-------|--------------|-------------------------------------------------------------|
| E1      | E1    |        | 1     | 1 (100)      | 1 (100)                                                   |
| E2      | E2    |        | 11    | 11 (100)     | 11 (100)                                                  |
| E3      |       |        |       |              |                                                            |
| Yeast   |       |        |       |              |                                                            |
| JAMM    | 0     | NA     | NA    | NA           | NA                                                         |
| MJD     | 0     | NA     | NA    | NA           | NA                                                         |
| OUT     | 0     | NA     | NA    | NA           | NA                                                         |
| UCH     | 1     | 1 (100)| 1 (100)|              |                                                            |
| USP     | 16    | 16 (100)| 16 (100)|             |                                                            |
| Total   | 109   | 89 (81.6)| 89 (100)|             |                                                            |

Note.—UCH, ubiquitin C-terminal hydrolases; NA, not applicable.
protein domains, and although it was shown that the Slx5p-Slx8p dimer has robust substrate-specific E3 ligase activity, such activity is ascribed to the SLX8 gene, with SLXS working to enhance the catalytic process only (Xie et al. 2007). The genes SNT2 (YGL131C) and PEP3 (YLR148W) are not reported as having E3 activity nor encode any protein domains that may suggest so. The gene STE5 (YDR103W) is not an E3 but a MAPK scaffold protein that controls the mating decision, and the ubiquitination of Ste5p appears to be controlled by the Cdc4p E3 (an F-box) and requires Cdk1p-mediated phosphorylation (Garrenton et al. 2009). PEX2 (YIL210W) does not encode an identifiable RNF domain, and neither does PEX12, unlike PEX10, which may be the E3 responsible for the ubiquitination reactions reported for this peroxisomal protein import complex (Platta et al. 2009). The Ssl1 gene (YLR005W), part of the TFIH complex, has been reported to display E3 activity (Takagi et al. 2005). Analysis of its domain structure showed that the C-terminal domain is a C2H2 zinc finger domain (a very common DNA-binding domain found in many transcription factors) and not a C3HC4 zinc finger domain (which is diagnostic for RNF domain E3s). As the authors noted, Ssl1p might depend upon another subunit for its activity. Among the proteins annotated as F-box proteins, the following lack an identifiable F-box domain, and no specific E3 activity has been conclusively ascribed: COS111 (YBR203W), SAF1 (YBR280C), RCY1 (YIL204C), HRT3 (YLR097C), YLR224W, YLR352W, AMN1 (YBR158W), CTF13/CBF3 (YMR094W), RAD7 (YJR052W), YDR306C, and ROY1 (YMR258C). Discounting these 19 false positives from the original annotation brings the yeast complement to 73 UBEs and 17 DUBs. Therefore, because our method identified 89 of these 90 enzymes, we estimate a coverage of approximately 99% (89/90) and a family-level classification accuracy of 100% on this curated data set. The only gene that we failed to identify is Elongin A (E4A1 or YNL230C), an E3 containing an F-box domain that displays marginal sequence similarity to canonical F boxes. This E4A1 F-box domain can only be identified with a PROSITE matrix (and with a very poor score).

In a second test on a more phylogenetically diverse data set, we examined the coverage and classification accuracy of our method on a select set of proteins from the UniProt database (Uniprot Consortium 2010), the largest public repository of protein sequences. First, we examined the Gene Ontology (GO) resource (Gene Ontology Consortium 2000) for the GO terms under the “Molecular Function” category that would help categorizing UBEs and DUBs. These GO terms ( supplementary table S5, Supplementary Material online) were used to select a set of protein sequences (UniProt release of 16 May 2012; http://www.ebi.ac.uk/uniprot) as our benchmark. Moreover, these proteins should also have been annotated under the experimental evidence codes IDA (“Inferred from Direct Assay”), IPI (“Inferred from Physical Interaction”), IMP (“Inferred from Mutant Phenotype”), IGI (“Inferred from Genetic Interaction”), and EXP (“Inferred from experiment”). This selection resulted in a set of 574 UniProt proteins from a variety of phylogenetic lineages including chordates (human, mouse, rat, chicken, Xenopus laevis, and zebrafish), nonchordate animals (Drosophila melanogaster and Caenorhabditis elegans), plants (A. thaliana, Oryza sativa, and Capsicum annuum), fungi (Schizosaccharomyces pombe, S. cerevisiae, Candida albicans, and Emericella nidulans), a number of (mainly Gram-negative) bacteria (Chlamydia trachomatis, Legionella pneumophila, Mycobacterium tuberculosis, Salmonella typhimurium, and Shigella flexneri), and a virus (Human herpesvirus). The UBE/DUB HMM Library retrieved 514/572 proteins (~89.9%). Manual inspection of the set of 58 proteins that our method failed to identify showed that 16 of these were fragments instead of full proteins. Moreover, 24 of the proteins were misannotated as UBEs or DUBs, 10 proteins did not harbor any identifiable protein domains at all, and 7 proteins were of bacterial origin. Bacterial enzymes of the ubiquitination pathway differ substantially from the eukaryotic enzymes sequence wise, and the UBE/DUB HMM Library was not expected to find them, although we successfully identified and classified a bacterial E3 ligase from Leg. pneumophila (LubX, Uniprot identifier LUBX_LEGPH). We also failed to identify human ZFP91 (UniProt: ZFP91_HUMAN), an atypical E3 reported to mediate the activatory Lys(63)-linked ubiquitination of MAP3K14/NIK (jin et al. 2010). ZFP91 cannot possibly be identified as an E3 on the sequence level as it harbors C2H2 zinc finger domains (which typically bind DNA), and as the name implies, it is “atypical.” Therefore, if we take into account the presence of these false positives in the original UniProt annotation, we report a coverage of 514/515 (99.8%) on this data set. Among the 514 proteins that were retrieved are representatives of all UBE and DUB families except DDB1-like, and the correct classification rate of all identified proteins was 100% on the family level ( table 3). Collectively, the results from these two tests indicate a coverage more than 99% for the retrieval of UBEs and DUBs from eukaryotic genomes, with a correct classification rate of 100% on the family level.

The DUDE-db: Contents and Sequence Data Sources

The UBE/DUB HMM Library was used to scan the predicted protein data sets of 50 completely sequenced and published eukaryotic genomes. The genomes analyzed belong to four of the eukaryotic supergroups, as defined by Keeling et al. (2005). These include the unikonts: Anopheles gambiae (Holt et al. 2002), Branchiostoma floridae (Putnam et al. 2008), C. elegans (C. elegans Sequencing Consortium 1998), Can. glabrata (Dujon et al. 2004), Ciona intestinalis (Dehal et al. 2002), Cryptococcus neoformans (Loftus et al. 2005), Danio rerio (Flicek et al. 2012), Dictostylium discoideum (Eichinger et al. 2005), D. melanogaster (Adams et al. 2000), E. cuniculi (Katinka et al. 2001), Gallus gallus (International Chicken Genome Sequencing Consortium 2004), Homo sapiens (Lander et al. 2001), Klyveromyces lactis (Dujon et al. 2004), Macaca mulatta (Gibbs et al. 2007), Monodelphis domestica (Mikkelsen et al. 2007), Monosiga brevicollis (King et al. 2008), Mus musculus (Waterston et al. 2002), Nematostella vectensis (Putnam et al. 2007), Neurospora crassa (Galagan et al. 2003), Ornithorhynchus anatinus.
DUDE-db was designed to store the UBE/DUB complement of an unlimited number of genomes and is freely available at http://www.DUDE-db.org/. DUDE-db is stored as a MySQL relational database (http://www.mysql.com/). The server is implemented as a set of Python and Perl CGI scripts running under Apache (http://www.apache.org/).

Supplementary Material
Supplementary tables S1–S5 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
The authors thank Dr. David Martin (Dundee) for help with Jalview, Dr. Shandar Ahmad (Osaka) for reading the manuscript, and Ms. Mineko Tanimoto for secretarial support. Japan Society for the Promotion of Science (JSPS) through the WPI-IFREC Research Program and a Kakenhi grant; the Kishimoto Foundation; and the ETHZ-JST Japanese-Swiss Cooperative Program.

References
Adams MD, Celniker SE, Holt RA, et al. (195 co-authors). 2000. The genome sequence of Drosophila melanogaster. Science 287: 2185–2195.
Amerik AY, Li SJ, Hochstrasser M. 2000. Analysis of the deubiquitinating enzymes of the yeast Saccharomyces cerevisiae. Biol Chem. 381: 981–992.
Andreeva A, Howorth D, Chandion JM, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. 2008. Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res. 36: D419–D425.
Aparicio S, Chapman J, Stupka E, et al. (41 co-authors). 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 297:1301–1310.
Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815.
Ardley HC, Robinson PA. 2005. E3 ubiquitin ligases. Essays Biochem. 41: 15–30.
Argenzio E, Bange T, Oldrini B, Bianchi F, Peesari S, Mari S, Di Fiore PP, Mann M, Polo S. 2011. Proteomic snapshot of the EGF-induced ubiquitin network. Mol Syst Biol. 7:462.
Berriman M, Ghedin E, Hertz-Fowler C, et al. (102 co-authors). 2005. The genome of the African trypanosome Trypanosoma brucei. Science 309:416–422.
Bowler C, Allen AE, Badger JH, et al. (80 co-authors). 2008. The genome of the flowering plant Arabidopsis thaliana. Nature 452:239–244.
C. elegans Sequencing Consortium. 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282:1022–1028.
Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 34:D363–D368.
Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.
Ciechanover A, Schwartz AL. 2004. The ubiquitin system: pathogenesis of human diseases and drug targeting. Biochim Biophys Acta 1695: 3–17.
Cohen P. 2002. Protein kinases—the major drug targets of the twenty-first century? Nat Rev Drug Discov. 1:309–315.

Table 3. Family-Level Classification Accuracy of the UBE/DUB HMM Library on a Manually Annotated and Experimentally Verified Set of UBEs and DUBs from the UniProt Database.

| Class | Family | Proteins | Coverage (%) | Family-Level Classification Accuracy on the Data Set Covered (%) |
|-------|--------|----------|--------------|---------------------------------------------------------------|
|       | E1     | 3        | (100)        | 3 (100)                                                      |
|       | E2     | 79       | (100)        | 79 (100)                                                     |
| E3    | HECT   | 33       | (100)        | 33 (100)                                                     |
|       | RING/RNF domain | 244 | (100) | 244 (100) |
|       | RING/BTB | 2 | (100) | 2 (100) |
|       | RING/SOCS box | 1 | (100) | 1 (100) |
|       | RING/F-box | 14 | (100) | 14 (100) |
|       | U-box | 37 | (100) | 37 (100) |
|       | Other/ZnF A20 | 1 | (100) | 1 (100) |
|       | Other/DDB1-like | 0 | NA | NA |
|       | Mixed family | 6 | (100) | 6 (100) |
| DUB   | UCH    | 6        | (100)        | 6 (100)                                                      |
|       | USP    | 69       | (100)        | 69 (100)                                                     |
|       | MJD    | 4        | (100)        | 4 (100)                                                      |
|       | OUT    | 10       | (100)        | 10 (100)                                                     |
|       | JAMM   | 5        | (100)        | 5 (100)                                                      |
|       | Atypical | 1 | (0) | 0 (0) |
|       | Total  | 515      | (99.8)       | 514 (100)                                                    |

Note.—The UBE/DUB HMM Library recovered 514/515 (99.8%) of all bona fide UBE and DUB from the UniProt database. This near-perfect coverage was mirrored by a perfect family-level classification across all enzyme families. The ‘Mixed-family’ category includes proteins that harbor domains diagnostic for more than UBE or DUB family. UCH, ubiquitin C-terminal hydrolases; NA, not applicable.

(Warren et al. 2008), Pan troglodytes (Chimpanzee Sequencing and Analysis Consortium 2005), Phanerochaete chrysosporium (Martinez et al. 2004), Pongo pygmaeus (Locke et al. 2011), Rattus norvegicus (Gibbs et al. 2004), S. cerevisiae (Goffeau et al. 1996), Sch. pombe (Wood et al. 2002), Strongylocentrotus purpuratus (Sodergren et al. 2006), Takifugu rubripes (Aparicio et al. 2002), Trichoplax adhaerens (Srivastava et al. 2008), Ustilago maydis (Kamper et al. 2006), X. tropicalis (Hellsten et al. 2010), and Yarrowia lipolytica (Dujon et al. 2004); the excavates: L. infantum (Peacock et al. 2007), L. major (Ivens et al. 2005), T. brucei (Berriman et al. 2005) and T. cruzi (El-Sayed et al. 2005); the plants: A. thaliana (Arabidopsis Genome Initiative 2000), Chlamydomonas reinhardtii (Merchant et al. 2007), Cyanidioschyzon merolae (Matsuzaki et al. 2004), O. sativa (International Rice Genome Sequencing Project 2005), Ostreococcus lucimarinus (Palenik et al. 2007), Ost. tauri (Derele et al. 2006), Physcomitrella patens (Rensing et al. 2008), Populus trichocarpa (Tuskan et al. 2006), Volvox carteri (Prochnik et al. 2010), and Zea mays (Wei et al. 2009); and the chromalveolates: Phaeodactylum tricornutum (Bowler et al. 2008), P. ramorum (Tyler et al. 2006), P. sojae (Tyler et al. 2006), and Thalassiosira pseudonana (Armbrust et al. 2004).
Prediction of Ubiquitinating and Deubiquitinating Enzymes · doi:10.1093/molbev/mst022

Petroski MD. 2008. The ubiquitin system, disease, and drug discovery. BMC Biochem. 9(1 Suppl):S7.

Platta HW, El Magraoui F, Baumer BE, Schlee D, Girzalsky W, Erdmann R. 2009. Pex2 and pex12 function as protein-ubiquitin ligases in peroxisomal protein import. Mol Cell Biol. 29:5505–5516.

Prochnik SE, Umen J, Nedelcu AM, et al. (28 co-authors). 2010. Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. Science 329:223–226.

Pruitt KD, Tatusova T, Klimke W, Maglott DR. 2009. NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Res. 37:D32–D36.

Putnam NH, Butts T, Ferrier DE, et al. (37 co-authors). 2008. The amphiioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071.

Putnam NH, Srivastava M, Hellsten U, et al. (19 co-authors). 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317:86–94.

Randow F, Lehner PJ. 2009. Viral avoidance and exploitation of the ubiquitin system. Nat Cell Biol. 11:527–534.

Rensing SA, Lang D, Zimmer AD, et al. (71 co-authors). 2008. The Physcomitrella genome and the evolution of the chordate karyotype. Nature 319:64–69.

Rensing SA, Lang D, Zimmer AD, et al. (71 co-authors). 2008. The Physcomitrella genome and the nature of placozoans. Nature 453:175–183.

Rong Y, Ma C, Li Y, et al. (33 co-authors). 2007. The Paramecium genome reveals ancestral eumetazoan gene repertoire and genomic organization. PLoS Genet. 3: e1000715.

Wenzel DM, Klevit RE. 2012. Following Ariadne’s thread: a new perspective on RBR ubiquitin ligases. BMC Biol. 10:24.

Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, Chothia C, Gough J. 2009. SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res. 37:D380–D386.

Woelfl S, Sigismund S, Penengo L, Polo S. 2007. The ubiquitination code: a signalling problem. Cell Div. 2:11.

Wood V, Gwilliam R, Rajandream MA, et al. (134 co-authors). 2002. The genome sequence of Schizosaccharomyces pombe. Nature 415:871–880.

Xie Y, Kerscher O, Kroetz MB, McConchie HF, Sung P, Hochstrasser M. 2007. The yeast Hex3.Slx8 heterodimer is a ubiquitin ligase stimulated by substrate sumoylation. J Biol Chem. 282:34176–34184.

Zargari A, Boban M, Heessen S, Andréasson C, Thyberg J, Ljungdahl PO. 2007. Inner nuclear membrane proteins Asp1, Asp2, and Asp3 function in concert to maintain the latent properties of transcription factors Stp1 and Stp2. J Biol Chem. 282:5954–605.

Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847–848.