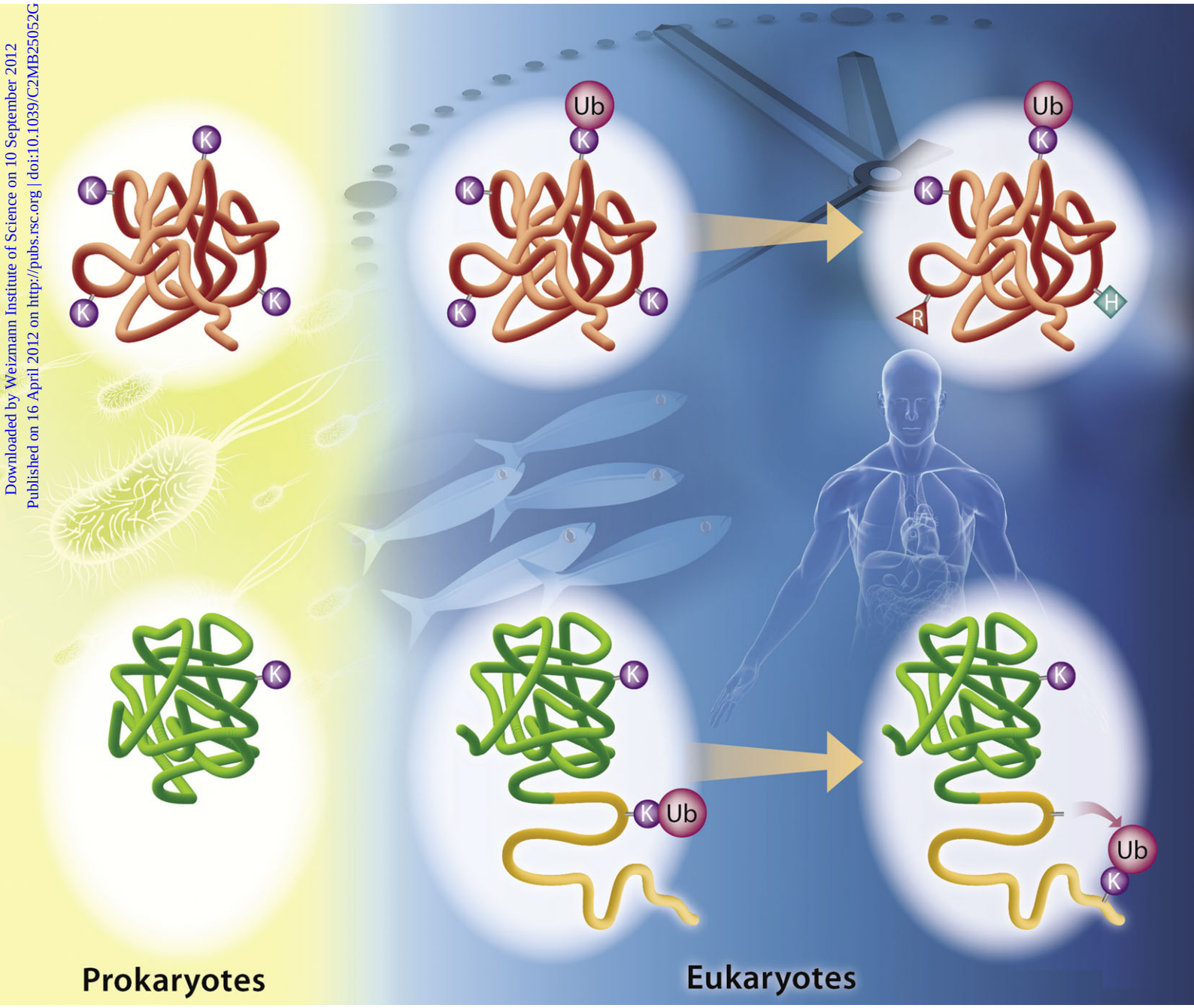
Molecular BioSystems

www.molecularbiosystems.org

Volume 8 | Number 7 | July 2012 | Pages 1843–2012



ISSN 1742-206X

RSC Publishing

PAPER

Yaakov Levy *et al.*

The origins and evolution of ubiquitination sites

1742-206X(2012)8:7;1-6

Molecular BioSystems

Dynamic Article Links
View Online

Cite this: *Mol. BioSyst.*, 2012, **8**, 1865–1877

www.rsc.org/molecularbiosystems

PAPER

The origins and evolution of ubiquitination sites†

Tzachi Hagai, Ágnes Tóth-Petróczy, Ariel Azia and Yaakov Levy

Received 14th February 2012, Accepted 13th April 2012
DOI: 10.1039/c2mb25052g

Protein ubiquitination is central to the regulation of various pathways in eukaryotes. The process of ubiquitination and its cellular outcome were investigated in hundreds of proteins to date. Despite this, the evolution of this regulatory mechanism has not yet been addressed comprehensively. Here, we quantify the rates of evolutionary changes of ubiquitination and SUMOylation (Small Ubiquitin-like MOdifier) sites. We estimate the time at which they first appeared, and compare them to acetylation and phosphorylation sites and to unmodified residues. We observe that the various modification sites studied exhibit similar rates. Mammalian ubiquitination sites are weakly more conserved than unmodified lysine residues, and a higher degree of relative conservation is observed when analyzing *bona fide* ubiquitination sites. Various reasons can be proposed for the limited level of excess conservation of ubiquitination, including shifts in locations of the sites, the presence of alternative sites, and changes in the regulatory pathways. We observe that disappearance of sites may be compensated by the presence of a lysine residue in close proximity, which is significant when compared to evolutionary patterns of unmodified lysine residues, especially in disordered regions. This emphasizes the importance of analyzing a window in the vicinity of functional residues, as well as the capability of the ubiquitination machinery to ubiquitinate residues in a certain region. Using prokaryotic orthologs of ubiquitinated proteins, we study how ubiquitination sites were formed, and observe that while sometimes sequence additions and rearrangements are involved, in many cases the ubiquitination machinery utilizes an already existing sequence without significantly changing it. Finally, we examine the evolution of ubiquitination, which is linked with other modifications, to infer how these complex regulatory modules have evolved. Our study gives initial insights into the formation of ubiquitination sites, their degree of conservation in various species, and their co-evolution with other posttranslational modifications.

Introduction

Ubiquitination has emerged as a major regulatory mechanism that controls protein homeostasis and affects important cellular pathways, such as DNA repair and protein trafficking.[1–7] A tightly regulated cascade of enzymes leads to the covalent attachment of a monoubiquitin or a polyubiquitin chain to a lysine residue of the modified protein, which is subsequently recognized by specific ubiquitin-binding proteins, affecting its function, turnover, localization, or interactions.

Ubiquitin-mediated signaling is thought to be universal in eukaryotes, and its utilization was shown to be prevalent in various eukaryotic species. However, how this system evolved from its prokaryotic antecedents and how it has developed in the course of eukaryotic evolution are less clear.[8–11] The presence of analogous ubiquitin-like protein conjugation systems has been characterized in several prokaryotes.[12–14] Sulfur activation and delivery systems, which resemble the ubiquitination machinery in sequence and in enzymatic mechanism, were shown to be widespread in bacterial genomes.[15] Therefore, the ubiquitin signaling system was likely to be present in the last eukaryotic common ancestor.[10]

While the origins of the ubiquitin system have been uncovered, the evolution of ubiquitination sites, the modified positions on the proteins' surface, has not yet been investigated. Hundreds of ubiquitination sites are now known experimentally, most of them from a few mammals and from baker's yeast (*Saccharomyces cerevisiae*) and many of them with characterized functions. However, the time at which these sites formed and the degree of their conservation in other uncharacterized species are rarely known. On account of their important role in regulation and the high specificity required in the ubiquitination process, ubiquitination sites are expected to be conserved. On the other hand, a lack of conservation of ubiquitination sites may be an important source of diversity between species

Department of Structural Biology, Weizmann Institute of Science, Rehovot, 76100, Israel. E-mail: koby.levy@weizmann.ac.il; Fax: +972-8-9344136; Tel: +972-8-934-4587
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c2mb25052g

reflecting altered regulation. In cases where the ubiquitin moiety interacts with the substrate to form a specific interface[16] or to affect its physical nature,[17–19] it is expected that the site location would be evolutionarily constrained, however in cases where no important interaction occurs,[20] the specific site's location may be less important and may therefore be less conserved.

In some substrates, only a few lysine residues can be ubiquitinated and their mutation leads to impairment of the cellular process mediated by ubiquitination.[21,22] Other substrates display a high degree of redundancy and have many lysine residues that can serve as alternative ubiquitination sites.[23,24] Cases where numerous residues may serve as ubiquitination sites may point to the ability of the ubiquitination machinery to recognize and conjugate ubiquitin to various regions, while maintaining a high level of specificity, perhaps by tight temporal and spatial regulation of E2 and E3 enzymes.[4] Redundancy in ubiquitination sites may decrease the evolutionary constraints on specific sites and increase their evolutionary rates.

Here, we study the evolutionary conservation and formation of ubiquitination sites in comparison with other posttranslational modifications, specifically SUMOylation (a ubiquitin-like modification which regulates various pathways), acetylation (an unrelated lysine modification), and phosphorylation (the evolution of the latter two modifications as well as of glycosylation was studied previously).[25–34]

We evaluate the time at which these sites appeared and their evolutionary rates, and compare them to analogous unmodified residues in the same protein as done previously.[28,30] In addition, in cases where ubiquitination sites seem not to be conserved at exactly the same position, we test the possibility of a compensating mechanism, in which lysine residues nearby may backup the disappearance of the site from its original location. In order to learn about the sequence rearrangements that occurred prior to the formation of ubiquitination sites, we study the characteristics of the region in which the ubiquitination sites reside in prokaryotic orthologs and so gain insights into how regulatory regions such as ubiquitination sites are formed within an existing protein. Finally, we investigate proteins whose ubiquitination is linked to additional modifications to better comprehend how these complex regulatory modules were created during evolution.

Materials and methods

Dataset assembly

Mammalian and yeast proteins with known ubiquitination and SUMOylation sites were extracted from the ubiprot,[35] uniprot[36] Release 2011_05, 3-5-2011, and phosphosite[37] databases and from a literature search. We manually annotated the proteins and removed any redundant protein and any protein whose sites were ambiguous or whose ubiquitination process was not characterized *in vivo*. Proteins whose ubiquitination process may have evolved in a manner different from that of common modified proteins (such as substrates that are ubiquitinated by viral proteins, and proteins from the ubiquitin family) were discarded. For each protein, we extracted annotated eukaryotic orthologs using the InParanoid database (version 7.0)[38] and in the yeast set we additionally used the Fungal Orthogroups Repository.[39] When more than one ortholog was assigned, the sequence with the highest similarity score was used. For further analysis, we used proteins that had at least 6 orthologs in their defined phylogeny of 19 vertebrates and 16 yeast species (see Fig. S1 for details, ESI†). Thus, the final dataset comprised two components, yeast (210 proteins, 16 orthologs) and vertebrate (416 mammalian proteins, 19 vertebrate orthologs), where the 210 yeast proteins and 416 mammalian proteins contain 753 known ubiquitination and 398 known SUMOylation sites (details in Tables S1 and S2, ESI†). In addition to ubiquitination and SUMOylation sites, the acetylation and phosphorylation sites of these proteins were extracted from the same databases and analyzed for comparison. Additionally, we extracted known cases of cross-talks between ubiquitination and other modifications and cases where primary and secondary sites have been identified. Finally, we divided the mammalian ubiquitinated proteins by their characterization method—high-throughput MS[40] or functional mutagenesis.

In order to extract prokaryote orthologs, we used the Inparanoid program,[38] with the BLOSUM62 matrix, a cutoff score of 35, and an overall sequence overlap cutoff of 0.4 with the complete sequenced genomes of 32 representative prokaryote species (see Table S4, ESI†).

Sequence analysis

Orthologs were aligned using MUSCLE,[41] and were later manually examined and corrected. A vertebrate phylogenetic tree was created with PhyML[3] with default parameters, using a concatenated alignment of proteins that had all 19 possible orthologs, whereas in the yeast phylogeny we used the tree constructed previously.[39] Evolutionary rates were computed using Rate4Site[42] with default parameters using the same species tree for all proteins (Fig. S1 and S2, ESI†). The evolutionary model used, JTT, was selected after assessing the best-fit model using ProtTest;[43] other models gave similar results. Ancestral reconstruction was achieved using PAML,[44] and the first appearance of a residue of a certain type was defined as the earliest appearance of that residue in the inferred ancestral sequence, which was conserved from that ancestral node.

Since the evolutionary rates of different structural regions differ,[45–48] we characterized the structural properties of the substrates (each residue's tendency to be disordered and its surface accessibility). The tendency of residues to be in disordered regions was predicted using IUPred,[49,50] and a site was considered to be disordered if the predictor returned an average value of ≥ 0.5 for an 11-residue long window surrounding it (we have repeated this analysis with two additional disorder predictors—PONDR-FIT[51] and foldIndex[52]—the results of the three predictors were similar and did not significantly affect the evolutionary analysis). Ordered regions were divided into core and surface by assessing surface accessibility (ASA) using the CSU method[53] in cases where structures were available (ASA was defined as the ratio of the solvent accessible surface for a given residue within the structured

protein *versus* in completely solvated residue). Residues were classified as core for ASA $\leq$ 0.2 and, where no structure was available, by prediction of Prof PHD.[54]

Statistical tests

The significance of the results was determined using the Fisher exact test, Kolmogorov–Smirnov test, and empirical sampling of subgroups. A P value of $\leq$0.05 was considered significant. All P-values that appear in the main text were determined using the Kolmogorov–Smirnov test.

In order to statistically test whether modification sites are relatively conserved, we composed two distributions: the first "modified" distribution was composed of all the sites belonging to a specific category (for example—ubiquitination sites in ordered regions in the mammalian set). The second "unmodified" distribution was created so that it will have the same number of points as in the modified distribution, and that the number of values from each protein would be identical in both distributions. To achieve this, for every modified site, a corresponding value was created in the unmodified distribution, which was calculated as the median of the respective unmodified residues in the same protein with similar structural properties (ordered or disordered regions). In this manner, we controlled for both protein's rate variation and distribution size.

In order to examine the presence of Lys residues in the vicinity of non-conserved ubiquitination sites, we used fish orthologs of the mammalian functional-mutagenesis set to determine which Lys residues are conserved (we used the fish sequences as they are the most distant orthologs in the vertebrate clade, which gives the highest number of non-conserved Lys residues and allows a rigorous statistical examination). We then separated the non-conserved Lys residues into four categories: ubiquitinated and unmodified Lys residues in ordered and disordered regions. In each category, we examined how many of the non-conserved Lys have a Lys residue in a window of a certain length surrounding the original Lys location ("compensated Lys"), and how many do not have it ("uncompensated Lys"). We calculated the normalized ratio between these two types of Lys as (compensated Lys – uncompensated Lys)/(compensated Lys + uncompensated Lys). A more positive ratio may reflect a stronger compensation. Next, we randomly shuffled the fish sequences (still keeping the ordered and disordered content and locations separated) 10 000 times. In each of these shuffling iterations, we determined the normalized ratio between the Lys types, and compared it to the ratio in the original non-shuffled fish sequence. We obtained empirical P-values by using the number of iterations in which the shuffled sequence had a higher ratio.

Results and discussion

The level of conservation of ubiquitination sites in comparison with other posttranslational modifications and with unmodified lysine residues

To explore the evolutionary history of ubiquitination sites, we composed a dataset of experimentally characterized ubiquitination and SUMOylation sites from 210 yeast and 416 mammalian proteins (Tables S1 and S2, ESI†). We extracted their orthologs from 16 yeast and 19 vertebrate genomes, respectively (Fig. S1, ESI†), with divergence estimated to have occurred about 420 or 450 Ma ago, respectively. Thus, we created a dataset with two components, yeast and vertebrate, where the latter contained sites identified in mammalian species. We estimated the evolutionary rates of each protein at the residue level using the Bayesian method implemented in the Rate4Site program,[42] and compared the evolutionary rates of modification sites with those of lysine residues that are not known to be modified from within the same proteins (see Methods and Fig. S2, ESI†). Since the evolutionary rate of core residues is different from that of surface residues[45,47] and assuming that modification sites are on the surface, we discarded any unmodified lysine that was predicted to be in the core (there was no modified lysine that was predicted to be in the core). In addition, we compared ordered and disordered residues separately (both for the modified and the unmodified sites), since disordered regions tend to evolve faster than ordered ones.[46] Using the same set of proteins, we repeated this analysis for acetylation and phosphorylation sites.

As expected, we observed that all modification sites tend to evolve faster in disordered regions than in ordered domains (Fig. 1 and 2 and Fig. S2, ESI†), in agreement with previous results on phosphosites.[28]

We first examined the evolutionary rates of the modified and unmodified residues at the level of each of the proteins. In Fig. 1A and B, we show examples of such an analysis by plotting the evolutionary rates of various types of ubiquitination sites *versus* the median rates of the unmodified lysine residues in the same protein (the proteins are ordered by increasing the median evolutionary rate for the unmodified lysine residues). Using this plot, we can specifically analyze in which proteins the ubiquitination sites seem to be relatively non-conserved and why. For example, in the case of mammalian ubiquitination sites located in disordered regions, 52.5% of the sites are more conserved than their unmodified counterparts (their rates are below the median rate for unmodified lysine residues), 32.8% of the sites are less conserved than their unmodified counterparts, and 14.7% of the sites are similarly conserved.

In order to test the significance of our results, we compared the distribution of the rates of the modified sites with the corresponding distribution of the medians of the rates of the respective unmodified residues in the same protein (so that we have two distributions with the same size, and with the same number of values from each protein). For simplicity, we display in Fig. 2 the average of each of the modified distributions, with the corresponding unmodified distribution.

Modified lysine residues in ordered regions have average evolutionary rates that are slower than the mean of the proteins in both yeast and mammals (for example, the average evolutionary rates of ubiquitination sites in ordered regions in mammals and in yeast are 0.72 and 0.67, respectively, which means that these sites evolve at a rate that is slower by $\sim$30% than the mean of the entire content of the proteins set). In both clades, all the lysine modification types—ubiquitination, SUMOylation, and acetylation—have comparable rates ranging between 0.57–0.72 on average in ordered regions. In disordered regions, these sites tend to evolve faster, with average rates ranging between 0.69–1.36. Phosphorylation sites seem to evolve at rates similar to those of lysine-modification sites, and their

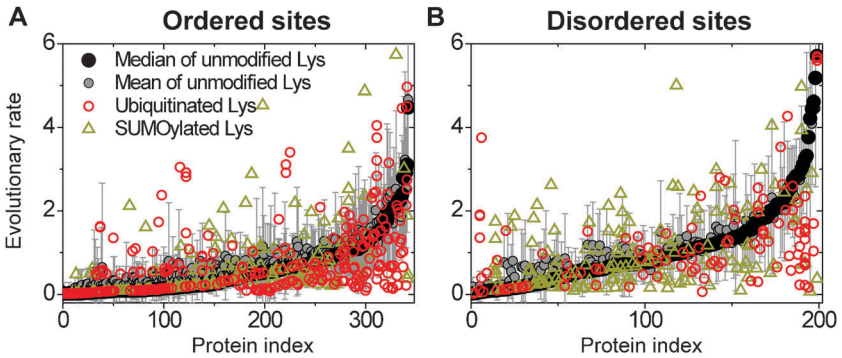


Fig. 1 Evolutionary rates of ubiquitination sites and SUMOylation sites and their corresponding unmodified lysine residues. (A and B) The individual evolutionary rates of ubiquitination and SUMOylation sites and the respective median of the unmodified lysine residues in (A) ordered and (B) disordered regions (proteins are presented in order of increasing median rate for the unmodified lysine residue).

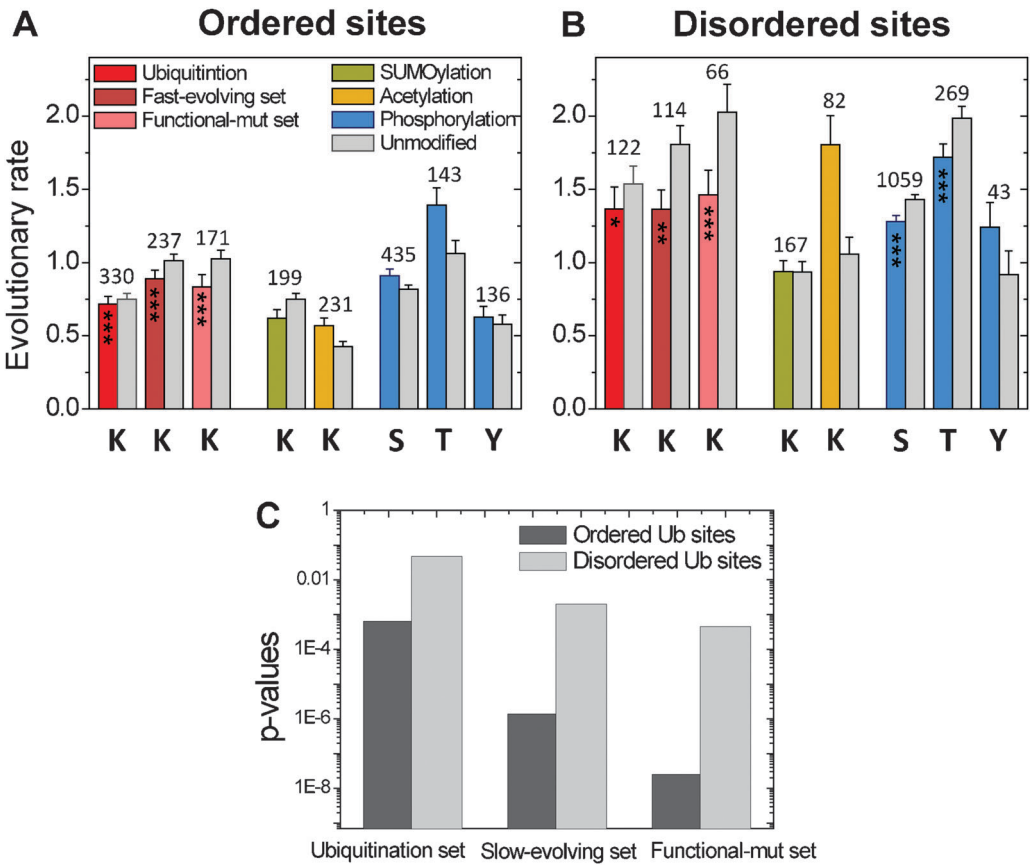


Fig. 2 Evolutionary rates of ubiquitination sites and other post-translational modifications in mammalian proteins. Means of evolutionary rate distributions of ubiquitination, SUMOylation, acetylation, and phosphorylation sites and their analogous unmodified residues (lysine, serine, threonine and tyrosine) in ordered (A) and disordered (B) regions of mammalian proteins. The numbers of sites of each type are indicated above each column. The distributions of the modified and unmodified set were compared using the Kolmogorov–Smirnov test. Ubiquitination sites are significantly more conserved than their unmodified counterparts in both ordered and disordered regions. These sites were further divided into two groups in two different analyses. (1) Fast and slow evolving proteins: in the fast-evolving set, sites are significantly more conserved than their unmodified residues. (2) Functional mutagenesis and MS-identified ubiquitination sites: in the functional-mutations set, ubiquitination sites are significantly more conserved. SUMOylation and acetylation sites were not found to be significantly more conserved, as most phosphorylation sites. See details in Table S3 (ESI†). (C) The significance of relative conservation of ubiquitination sites in ordered and disordered regions in the entire set of mammalian proteins, in the fast-evolving subset, and in the functional-mutagenesis set.

average rates are usually in good agreement with a previous study conducted on a different set.[28] One exception to this are the observed rates of Ser- and Thr-phosphosites in mammalian proteins, which in our dataset exhibited higher evolutionary rates compared with previous observations, possibly due to the different set used in that study.

We compared the distributions of the evolutionary rates of modified sites to those of the corresponding unmodified

residues in the same protein set to examine whether modified sites have higher evolutionary constraints than unmodified residues. In most cases, we observe that the rates of the unmodified residues are similar to those of the modified residues. In the mammalian set (Fig. 2), ubiquitination sites are slightly more conserved than unmodified lysine residues in both ordered and disordered regions (Fig. 2A and B). However slight, this excess conservation is significant ($P = 6.4 \times 10^{-4}$ for ordered regions, $P = 0.047$ for disordered regions, Kolmogorov–Smirnov test).

In the yeast set, ubiquitination sites are not more conserved than unmodified lysine residues (Fig. S2, ESI†). When comparing SUMOylation sites and acetylation sites to unmodified lysine residues and when comparing phosphosites to their unmodified counterparts (Ser, Thr or Tyr) we usually do not observe significant conservation in the modified set. In summary, with the exception of mammalian ubiquitination sites, modification sites in our set do not seem to be significantly more conserved than analogous unmodified residues, and even for mammalian proteins, the level of conservation is small.

It is probable that some of the unmodified lysine residues are highly conserved since they are part of active sites or important binding regions, and this is part of the reason for the mild difference between modified and unmodified sites. However, when looking at the absolute rate values in Fig. 1 (the *y*-axis), we note that a significant portion of our dataset has very low rates for both ubiquitinated and unmodified lysine residues, which corresponds to highly conserved proteins overall. In these cases, we cannot draw conclusions from our comparison of modified residues and unmodified residues, as we deal with proteins that tend to be conserved within the set of orthologs we analyzed.

Since the set of mammalian proteins seems to be enriched with proteins having very old modified and unmodified lysine residues, we examined whether this may bias our results. A recent study showed that proteins can be separated into several groups based on the distribution of their evolutionary rates.[47] In the group whose average rate is ≤ 0.5, most residues, even on the surface, tend to evolve very slowly. While it was reported that these slow-evolving proteins compose $\sim 2\%$ of the entire proteome, in our set, $\sim 22\%$ of the proteins have an average rate of ≤ 0.5. As this may affect our results, we divided our dataset into slow-evolving proteins (with an average evolutionary rate of 0.5 or below) and faster-evolving proteins (Fig. 2 and Fig. S5, ESI†). While in the slow-evolving set of mammalian proteins we observe that ubiquitination sites have higher rates than the respective unmodified lysine residues in the same protein ($P = 0.17$ and 0.008 for ordered and disordered regions, respectively, Kolmogorov–Smirnov test), in the faster-evolving set the opposite trend is seen and the ubiquitination sites are significantly more conserved than their unmodified counterparts ($P = 1.4 \times 10^{-6}$ and 0.002 for ordered and disordered regions, respectively, Kolmogorov–Smirnov test—see Fig. 2C). However, when repeating this analysis with SUMOylation and acetylation sites, we did not observe this significant separation between slow and fast evolving proteins and the relative evolution of their modification sites.

If a protein contains uncharacterized modification sites, they will necessarily be counted as unmodified sites because they are not known to be otherwise. We therefore wondered to what extent our analysis might be affected by the fidelity of the unmodified dataset. We separated the set of mammalian proteins into ubiquitinated proteins whose sites were characterized by high-throughput Mass-Spectrometry analysis (MS-based set) and proteins whose sites were characterized by point mutations (functional-mutations set). We chose the set of ubiquitinated mammalian proteins as it has a sufficient number of sites characterized by both Mass-Spectrometry and functional-mutagenesis to allow statistical analysis. The major difference between these sets is that in the functional-mutations set, the entire set of lysine residues that can serve as ubiquitination sites is often known, as well as the cellular pathway that this modification regulates. In the MS-based set, only the *wildtype* protein is tested, thus alternative ubiquitination sites that can serve as backup sites upon mutation of the original sites cannot always be traced. Therefore the MS-based set may contain only a subset of the possible ubiquitination sites. By comparing the conservation of modified and unmodified residues in both sets, we can see whether undetected modified residues (which are presumably more abundant in the MS-based set and are considered in our analysis as a subset of the unmodified residues) may significantly affect our results. Indeed, in the functional-mutations set, ubiquitination sites are significantly more conserved than their unmodified counterparts ($P = 2.5 \times 10^{-8}$ for the ordered set and $P = 4.5 \times 10^{-4}$ for the disordered set, Kolmogorov–Smirnov test, Fig. 2A and B), and the difference is much higher than in the entire set, which includes all types of ubiquitinated proteins (see Fig. 2C). This difference is diminished in the MS-based set, where ubiquitination sites are as conserved as the unmodified lysine set ($P > 0.05$, Fig. S5, ESI†), possibly because of the presence of uncharacterized backup sites being part of the unmodified set.

Why ubiquitination sites are not strongly conserved

From the above discussion, it seems that ubiquitination sites are weakly more conserved than unmodified lysine residues. However, while a distinct relative conservation is observed in some cases, it might be absent in others. A close inspection of our set revealed that part of this could be explained by the presence of highly conserved proteins with slow substitution rates (having a high degree of conservation throughout their surface) and by uncharacterized modification sites that, because they are unknown, are included in the unmodified set by default. Although these two factors may significantly bias our results and should be examined when additional substrates are discovered, they are unlikely to be strong enough to entirely explain them. Therefore, bearing in mind these technical limitations, it is reasonable to view the observed absence of strong conservation of ubiquitination sites as related to the processes through which ubiquitination sites are formed during evolution, to the mechanisms by which lysine was chosen by the ubiquitination enzymatic machinery and to changes in regulation or in site location during evolution. In addition, as recently suggested for phosphorylation sites,[28] it is possible that some of

the ubiquitination sites are non-functional, and therefore are expected to be less conserved. These important issues are not completely understood and we will endeavor to address them with respect to our findings and using specific known examples.

Several studies on the evolution of phosphorylation sites observed a varying degree of conservation, which was attributed to several factors. While most of the studies found evidence for a greater evolutionary constraint on phosphorylation sites relative to unmodified residues, they also identified a significant fraction of sites that are not conserved. This lack of conservation may stem from changes in regulatory pathways. Indeed, there is evidence of changes in phosphoregulation in several case studies,[55–57] that the rate of change in kinase-substrate networks is relatively high,[58] and that the changes in phosphorylation patterns may be an important source of phenotypic diversity.[26] The suggestion that some of the sites may be nonfunctional was also investigated,[28,59] and may explain the lack of conservation in some of these cases.

Another possibility for the apparent lack of conservation is the existence of alternative sites or a shift in the position of the modification site from its original location. In these cases, while the regulation is maintained, the exact location of the modification is not crucial, and may move over time. This possibility was explored in studies concerning phosphorylation sites,[27,29,34] where it was shown to be common in some cases, and in particular in disordered regions.

We used the mammalian functional-mutagenesis set, in which all the sites were experimentally validated and manually curated by us, to examine the possibility of utilization of a nearby lysine residue or shifts in site locations (a possible compensation mechanism in cases of lack of conservation at the specific location). We first quantified how many Lys residues from each group—ubiquitinated and unmodified residues, in ordered regions and disordered regions—are conserved in orthologs' sequences of fish species—the most distant group of species in the vertebrate clade. Since almost all of the dataset is composed of sites characterized in human, mouse and rat, using their most distant orthologs in the clade allows a rigorous statistical examination, as Lys residues are significantly less conserved in fish orthologs than in orthologs from other species which are closer to mammals (for example, Lys residues in the frog *X. tropicalis*, the next furthest species are more conserved by 8–12%).

Next, we determined how many of the residues that are not conserved have a Lys residue in a small window surrounding the original site (Fig. 3A). For example, in a window size of ±2, 12 non-conserved ubiquitination sites in ordered regions have a Lys residue in this window ("compensated sites"), while 27 sites lack any Lys residue in this window ("uncompensated sites"). We repeated this analysis with windows of varying lengths (3–11 residues in total). In Fig. 3B and C the percentages of compensated and uncompensated Lys residues appear for each of these windows from the total number of non-conserved Lys residues.

In order to test the significance of our findings, we shuffled the fish sequences (still separating the ordered and disordered regions as in the non-shuffled fish sequences), and evaluated the ratio between the compensated sites and the uncompensated sites within the examined window in the shuffled sequences.

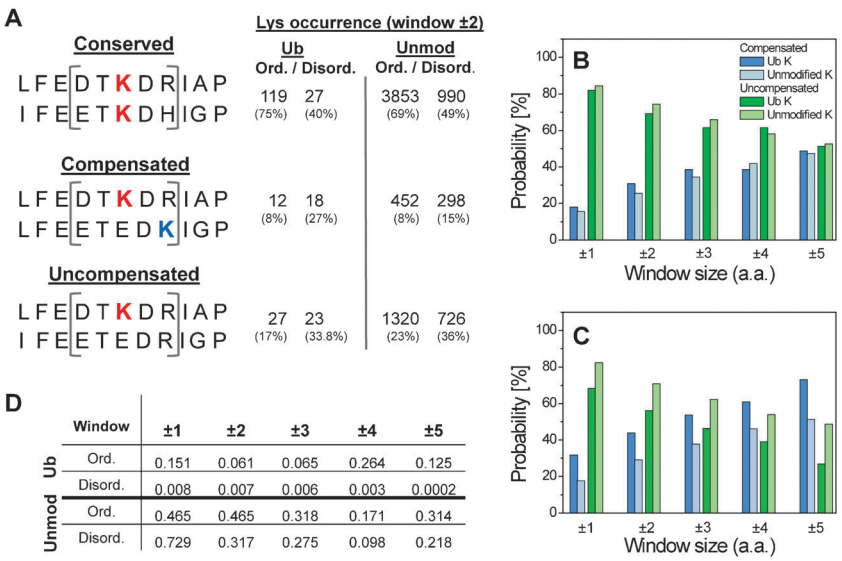


Fig. 3 Analysis of non-conserved ubiquitination sites. (A) We compared ubiquitination sites and unmodified Lys residues from the annotated mammalian functional-mutagenesis set to the corresponding sites in orthologous fish sequences. Each Lys can be either conserved (top), or non-conserved (middle and bottom). Non-conserved Lys residues can be either "compensated" by the presence of a Lys residue in a defined window surrounding the original Lys (middle), or "uncompensated" (bottom). The numbers of Lys residues (and their percentages) from each of these categories in a window of ±2 residues appear in the right (for ubiquitination and unmodified Lys in ordered and disordered regions). (B and C) The percentages of compensated and uncompensated ubiquitination sites and unmodified Lys residues in ordered regions (B) and disordered regions (C) from the total non-conserved Lys. The windows examined are of ±1–±5 residues, centered on the non-conserved Lys. (D) The *P*-values obtained by shuffling of fish sequences and comparing the normalized ratios between compensated and uncompensated Lys residues in the original and shuffled fish sequence. Note that the percentage in panel A was calculated relative to all types of Lys (conserved, compensated, and uncompensated) while the percentages in panels C and D are relative to the compensated and uncompensated Lys only.

The *P*-values were determined after repeating this procedure 10 000 times and comparing the ratio in the original non-shuffled sequence with the ratio in the shuffled sequence (see Methods for a detailed description of the procedure). We observed that the appearance of a "compensating" Lys residue in the vicinity is more common in ubiquitination sites, in comparison with this pattern in respective unmodified Lys residues. In disordered regions, the putative compensation appears to be stronger and is highly significant (see Fig. 3D for details on *P*-values). These findings were strong and consistent for sites in disordered regions for all window lengths that were examined, and were apparent in windows of ±2 and ±3 residues in the case of ordered sites, although to a lesser extent. The enrichment of Lys residues in the vicinity of ubiquitination sites was observed previously on different sets of proteins.[60,61] Our finding that ubiquitination sites in disordered regions may be compensated by the presence of nearby Lys is similar to findings on phosphosites, where it was suggested that a compensation mechanism may exist in several kinase systems, and that phosphosites location in disordered regions may rapidly change.[27]

In various ubiquitination substrates it was shown that site mutagenesis does not significantly impair ubiquitination, and that other lysine residues can be utilized as backup.[1] The fact that ubiquitination is thought to be achieved by increasing the effective concentrations of the region to be modified in the substrate at the active site of the E2 ubiquitin in RING E3s[62] lends support to the idea that the exact position of the lysine may not be crucial and may move along the sequence. In several ubiquitination substrates, the lysine position was engineered to be located further away from the original location.[63,64] In these cases it was shown that the engineered substrate may still be ubiquitinated; however the reaction kinetics may slow. Additionally, in several proteins the ubiquitin or SUMO modifications were fused to the substrates' termini and other regions,[65,66] yet they achieved the same function as the modified substrate in the original location.

PCNA is ubiquitinated and SUMOylated in response to certain DNA damage and can serve as a good example for examination of the above arguments. In these events, PCNA is alternatively ubiquitinated and SUMOylated on Lys164, and Lys127 serves as a backup site for SUMOylation.[67] While Lys164 is highly conserved across the eukaryotic kingdom, Lys127, which is part of the IDCL loop, is not as conserved, and changes in that region may be related to changes in interactions of PCNA with several of its partners[68] (see Fig. 4A and B). The fact that it only serves as an auxiliary position may reflect the decreased evolutionary constraints on it.

Yeast PCNA is also ubiquitinated on Lys107 in response to a deficiency in DNA ligase I.[69] While this residue is the sole acceptor of ubiquitin in this pathway, and this regulation was shown to be conserved in humans as well, the lysine residue at position 107 is mutated in mammals (Fig. 4C). However, it was suggested that Lys110 (which does not appear in yeast) can be used as the ubiquitin acceptor site, and that a shift in the location of the site occurred.[69] We observed a few similar cases where the experimentally known site disappears in several orthologs and another lysine residue emerges at the same time (Fig. S6, ESI† and examples in the following section).

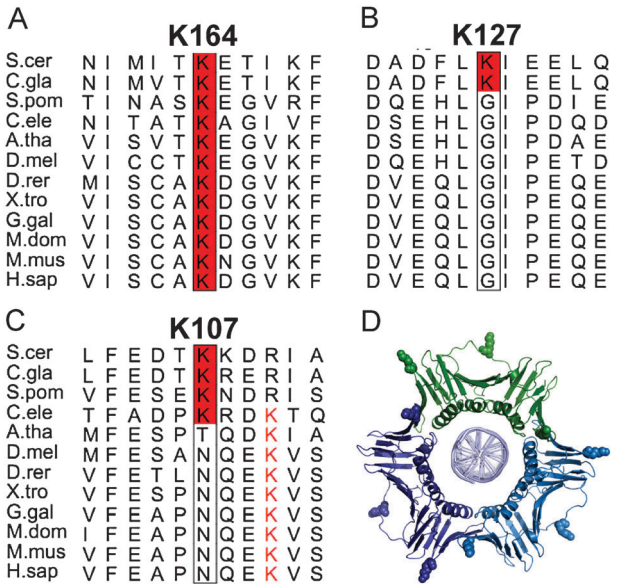


Fig. 4 The evolutionary history of PCNA ubiquitination and SUMOylation sites. The regions of PCNA ubiquitination and SUMOylation as characterized in yeast (*S. cerevisiae*) and in their corresponding regions in representative eukaryotes. Ubiquitination and SUMOylation sites are highlighted in the sequence in red, putative alternative ubiquitination site at K110 is marked with red letters. (A) K164 is mono- or polyubiquitinated and alternatively SUMOylated in response to certain types of DNA damage. (B) K127 is SUMOylated to a lesser extent and serves as a secondary site. (C) K107 was shown to be ubiquitinated in *S. cerevisiae* in DNA-ligase-I deficiency. While this regulatory pathway was shown to occur in humans, the exact site is not known, and it was suggested that it is shifted to K110. (D) PCNA structure and the ubiquitin and SUMO acceptor residues (which are shown with spheres).

These examples, in addition to the statistically significant bias we observe towards the appearance of lysine residues in close proximity to non-conserved sites, may imply that indeed a mechanism of compensation exists which can be utilized by the ubiquitination machinery to modify nearby lysine residues.

How ubiquitination sites are formed

We calculated the age of each of the lysine residues in our analysis, and the predicted time at which it first appeared, using ancestral reconstruction as implemented using the Maximum Likelihood method in PAML.[44] From this method, the most probable ancestral sequences of existing sequences may be inferred. By looking at the analogue position of a modified (or unmodified) residue in the ancestral sequence, we can derive the presumed identity of this residue (or, in our case, whether it is predicted to be a lysine or other residue). In this manner, we analyzed the predicted time of appearance of modified and unmodified lysine residues in our full set of ubiquitinated proteins (Fig. 5 and Fig. S4, ESI†). In the mammalian set, we observe that 76.9% and 63.5% of the residues that serve as ubiquitination sites in ordered and disordered regions, respectively, are predicated to appear before the split of fish and tetrapods ~450 Ma ago. Similar observations are seen in analyses of SUMOylation sites and acetylation sites in vertebrates. However, unmodified lysine

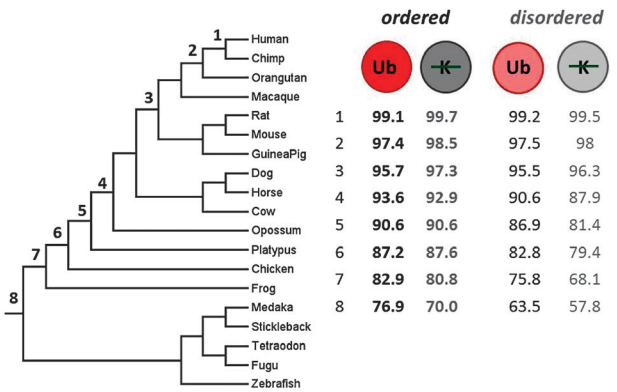


Fig. 5 Inferred age of ubiquitination sites and their analogous unmodified lysine residues in vertebrate proteins. The inferred ancestral nodes of ordered and disordered ubiquitination sites (in red) and of the analogous unmodified Lys residues (in gray). The node was inferred by ancestral reconstruction using the tree and the given multi-sequence alignment. Each lysine was assigned a node based on the most ancient node at which it was predicted to exist using the PAML program.[44] Numbers of nodes are indicated on the tree (to the left), and the percentages of lysine residues in each node are shown to the right.

residues appear to be relatively ancient as well, as 70% and 57.8% of the unmodified lysine residues in ordered and disordered regions, respectively, appeared at the same time (Fig. 5). In the yeast dataset, we again observe similar age distributions of ubiquitinated and unmodified lysine residues, although in this case both the modified and unmodified lysine residues seem to appear in later times than in the vertebrate phylogeny (Fig. S4, ESI†). For example, in ordered regions only 25% and 21% of the ubiquitinated and unmodified lysine residues, respectively, are predicted to appear before the split of *S. cerevisiae* and *S. pombe* ~420 Ma ago.

In order to gain insights into how new ubiquitination sites are formed, we analyzed the sequences of representative eukaryotic proteins from our combined yeast and mammalian datasets with their prokaryote orthologs. Since ubiquitination is predominantly used by eukaryotes,[10,15] such a comparison can estimate the extent to which the ubiquitination machinery scavenged already-existing surface lysine residues in a 'piggy-back' mechanism. We found 69 yeast and 67 mammalian proteins from our database that have at least two prokaryote orthologs in the prokaryotic genomes we analyzed (see methods and Table S4, ESI†), and examined the analogous location of the ubiquitination (and SUMOylation) sites in the prokaryotic species and the region surrounding it. We found that 54% of the sites in yeast proteins and 27% in mammalian proteins have analogous lysine residues in at least one of the prokaryotic species. This analysis gives a conservative rough estimation of the utilization of already-existing lysine residues by the ubiquitin machinery. In other cases, however, the site and its region are either very different or completely absent in prokaryote analogs.

In Fig. 6, we show examples illustrating how ubiquitination sites and their neighboring regions appear in prokaryote orthologs. In the first substrate, plasma membrane ATPase 1 (Fig. 6A and B), the ubiquitination site on Lys555 and its surrounding region are highly conserved in prokaryotic species, however a nearby site, Lys566, appears to significantly differ between eukaryotic and prokaryotic species. Similarly, in PCNA, Lys164, which serves as a major site for ubiquitination and

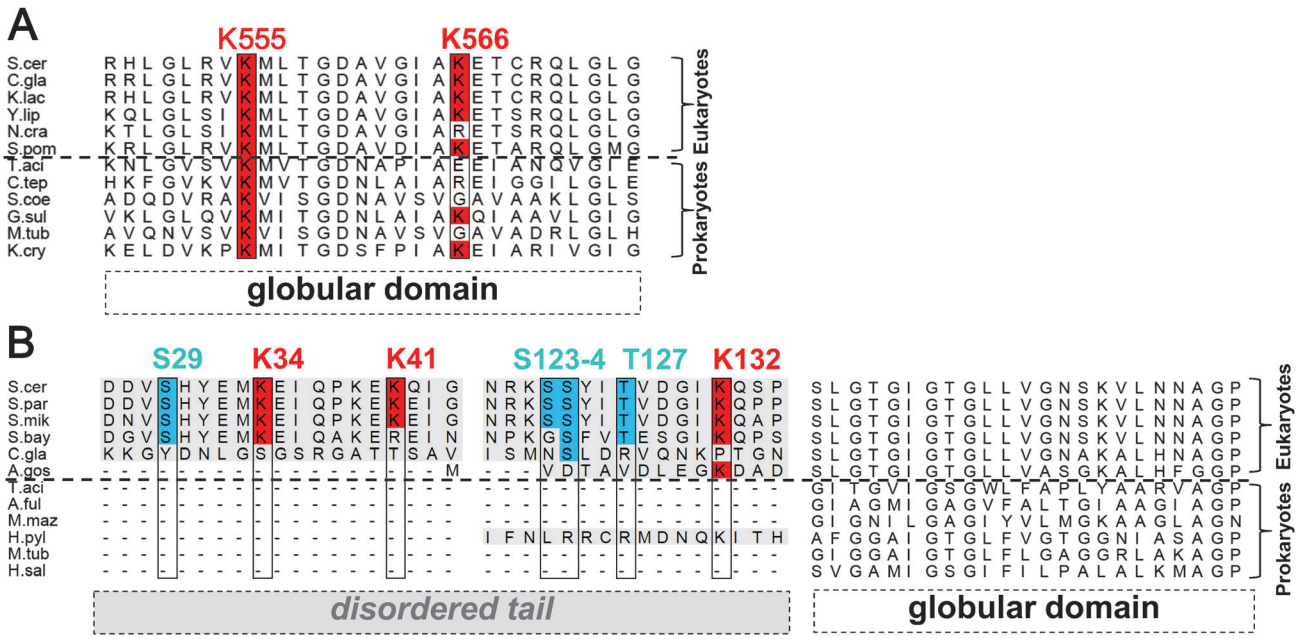


Fig. 6 Ubiquitination sites and their prokaryotic antecedents. Representative examples of proteins that are known to be ubiquitinated in *S. cerevisiae* and the corresponding regions in other yeast and prokaryotic orthologues. (A) Yeast plasma membrane ATPase 1 is ubiquitinated at K555 and K566. While K555 and its surrounding regions are highly conserved in both eukaryotes and prokaryotes, K566 is not as conserved. (B) High-affinity glutamine permease is ubiquitinated and phosphorylated at various sites on its disordered tail. These sites are not conserved in distant yeast species and the disordered tail is evolving relatively rapidly in yeast and is absent almost entirely in the prokaryotic species examined. However, a higher degree of conservation is observed in all species in the globular domain.

SUMOylation and which is almost entirely conserved in eukaryotes, is absent in prokaryotes and the region surrounding it undergoes many alterations. In several cases, the region surrounding the modification sites seems to be entirely missing in prokaryotic orthologs, such as in the high-affinity glutamine permease. In this protein ubiquitination (and phosphorylation) sites are found at the disordered N-terminal tail of the *S. cerevisiae* ortholog. While the disordered tail itself is absent in prokaryotic species, the ordered globular domain is relatively conserved in all examined species (Fig. 6B). In this case, it seems that the ordered parts of the protein are relatively conserved because of functional constraints, while the regulatory region resides in the disordered tail and evolves relatively rapidly as suggested previously.[70,71] This analysis suggests a wide spectrum of evolutionary origins for ubiquitination sites and their regions; although some of the sites and their regions appeared prior to their utilization by the ubiquitination machinery, others underwent various substitutions and insertions.

The evolution of posttranslational crosstalks

The evolution of complex regulation and of protein networks in general has been the focus of recent studies. In various case studies, substrate ubiquitination is linked with other modifications that either compete, promote, or inhibit it, in order to achieve tight and timely regulation.[1,4] The development of these complex crosstalks between various posttranslational modifications has received scant research attention. Here, we collected a number of ubiquitination substrates that are known to undergo additional modifications linked to the ubiquitination events with the aim of studying the origins and development of crosstalks between the various modifications. As suggested earlier,[72,73] many of these complex regulatory modules have ancient origins. For example, in the case of β-catenin, in which a cascade of phosphorylation events precedes ubiquitination, all the sites seem to be highly conserved from higher invertebrates to humans (see Fig. 7A). However, in other substrates, such as Iκbα, the modified sites are not as conserved (Fig. S7A, ESI†).

In other cases, however, this high level of conservation is not apparent and the sites of some of the modifications are altered. For example, in HIF1α, which undergoes hydroxylation prior to ubiquitination and degradation, the hydroxylated proline residues are highly conserved, and this degradation pathway was shown to be conserved experimentally from *Trichoplax adhaerens*, a basal metazoa to humans.[72] However, the lysine residues, which have been characterized in human to be ubiquitinated, are absent in many organisms (Fig. 7B). Since the regulatory pathway is conserved, a probable shift in the location of the lysine residues has occurred. As discussed previously, the precise location of the site may not be critical so long as the ubiquitination machinery is able to successfully ligate ubiquitin to the substrate. Indeed, in the case of HIF1α, it was shown that when lysine residues are shifted (in certain constructs, up to 15 residues away from the original site), ubiquitination of the substrate still occurs, albeit more slowly.[74]

A similar case may be the complex regulation of NEMO, which is first SUMOylated on Lys277 and Lys 309, then phosphorylated on Ser85 by the ATM kinase, which leads to the SUMO attachment being replaced with ubiquitin.[65] This complex cascade of events was characterized in humans and is conserved in many placental mammals. However, in orthologs

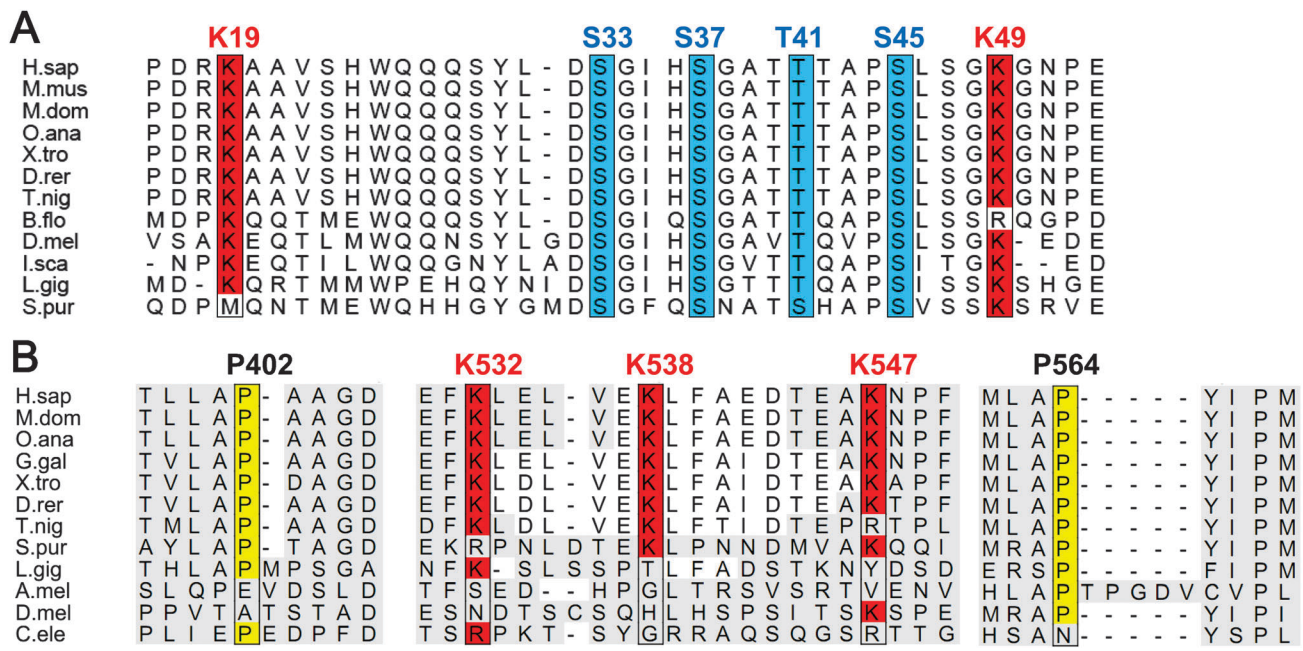


Fig. 7 The evolution of ubiquitination and its crosstalk with other post-translational modifications in β-catenin and HIF1α. Representative examples of proteins that are known to be regulated by several interlinked posttranslational modifications in humans and their corresponding regions in several representative orthologues: (A) β-catenin is phosphorylated at a number of sites (the blue-highlighted S33, S37, T41, S45 sites) before being polyubiquitinated (at the red-highlighted sites) and degraded. (B) HIF1α is hydroxylated at two (yellow-highlighted) proline residues prior to being ubiquitinated (at the red-highlighted sites) and degraded.

of other organisms, from placental mammals to insects, while the ubiquitination site is conserved, the phosphosite is absent in the original characterized location (Fig. S7b, ESI†). The disappearance of the consensus phosphorylation site (SQ) may point to a change in the regulation of NEMO or to a shift in its location. Without experimental evidence, we cannot rule out either of these two hypotheses. However, the existence of ATM in all these organisms as well as the appearance of SQ at least once in each of the examined orthologs in a different location (although this can occur at random) may suggest that, while the complex regulation in NEMO was conserved during the evolution from insects to mammals, the exact location of the different regulatory modules changed.

The CDK-inhibitor $p19^{4inkd}$ was shown in some cases to undergo two phosphorylation events preceding its degradation-dependent ubiquitination. The roles of these phosphorylation events were studied by mutating them to the phosphomimetic residue glutamic acid. While the phosphorylation of Ser76 is thought to act as a conformational switch perturbing the folded substrate to assist in proteasomal degradation, the phosphorylation of Ser66 was shown to increase the efficiency of the subsequent ubiquitination. When looking at its evolutionary history, an interesting picture emerges. The phosphorylation and ubiquitination sites are highly conserved in tetrapods (the phosphosite at Ser66 becomes asparagine in two species). In fish, however, there are prominent changes – the phosphosite at Ser76 becomes alanine, the phosphosite at Ser66 becomes the phosphomimetic aspartic acid, and Lys62, which serves as a ubiquitination site in tetrapods, is absent, while another lysine at position 63 emerges (Fig. 8).

As the fish and tetrapod proteins are almost identical in sequence, and since no other candidate for phosphorylation sites seems to appear in fish in the vicinity of that found in the tetrapods (and in this case, the physical proximity of the phosphosite to the ubiquitination site appears important), it seems that the phosphosites are not shifted in location. Instead, we putatively observe here a case of the development of the complex regulation of $p19^{4inkd}$ degradation during the evolution of vertebrates. While the ubiquitination site is presumed to exist in all vertebrates, the additional layer of control by phosphorylation seems to have been added on top of it only after the split between fish and tetrapods. Interestingly, while the phosphorylation site, which acts as a conformational switch, appears in fish as a neutral alanine residue (site 76), the second site (whose phosphorylation increases ubiquitination efficiency) appears as the phosphomimetic residue aspartic-acid (site 66). This raises the possibility that the second phosphorylation site in fish acts as a constitutive weak signal, as suggested recently in various phosphosites.[75] It may also be related to the shift in location observed in the ubiquitination site. In fish species, the lysine residue is located one residue closer to the 'phosphosite', perhaps because the aspartic-acid charge does not mimic the phospho-serine well enough. As the kinase and the ligase involved in this process are unknown, this hypothesis awaits experimental verification. However, a similar case, which was experimentally proven, was found in activation-induced cytidine deaminase, where phosphorylation at Ser38 in tetrapods is missing in fish, but an aspartic acid which is exclusively found nearby in the fish species was shown to act as its replacement that gives a constitutive signal[76] (see Fig. S8, ESI†).

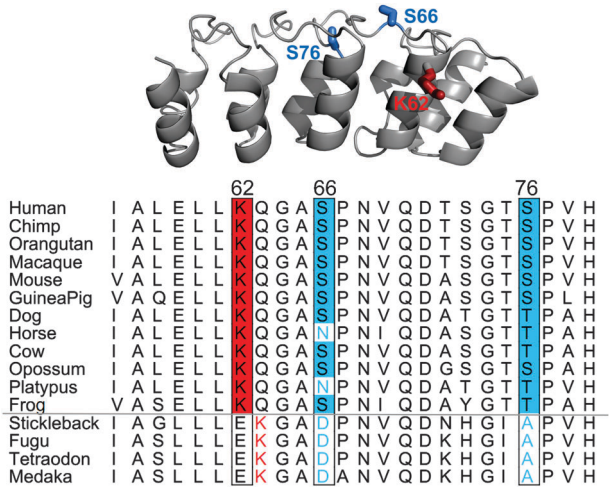


Fig. 8 The evolution of ubiquitination and related phosphorylation in $p19^{4inkd}$. (A) The structure (pdb: 1bd8) and (B) the multi-sequence alignment of the region in $p19^{4inkd}$ that is known to undergo phosphorylation and ubiquitination prior to its degradation in humans. In tetrapod organisms (mammals and amphibians) residues Ser66, Ser76 and Lys62, which were characterized in human cells as phosphorylation and ubiquitination sites, are almost entirely conserved. However, in fish (below the line), Ser76 is replaced by Ala, Ser66 is replaced by Asp, and Lys62 is replaced by Glu. Putatively, while the phosphorylation event at Ser76 is absent in fish, the phosphorylation at Ser66 is putatively replaced by a constitutive weak negative signal, and the ubiquitination site has shifted one residue closer to it (marked in red). The phosphorylation sites are marked in blue and the ubiquitination site is marked in red in both A and B.

Conclusions

In this work, we studied the formation and evolution of ubiquitination sites in comparison to other modifications. We showed that various lysine modifications—ubiquitination, SUMOylation and acetylation—exhibit similar evolutionary rates. Mammalian ubiquitination sites are significantly but weakly more conserved than unmodified lysine residues within the same proteins. In a non-negligible number of cases, a similar degree of conservation was observed for the modified and analogous unmodified lysine residues. The observed lack of additional evolutionary constraints on the modified lysine residues may stem from changes in the regulatory mechanism or simply from changes in the locations of the modification sites during evolution. In addition, non-functional ubiquitination sites which may be present in the dataset may also rapidly evolve. However, to some extent, these observations may stem from methodological factors, such as there being a high fraction of highly conserved proteins in the current set of ubiquitinated proteins and insufficient characterization of modification sites leading to modified lysine residues being erroneously treated as unmodified residues because they were not yet characterized. These issues may be resolved when the experimental characterization of ubiquitination sites would be extended to a larger set.

In cases where an interface is formed between the substrates and the ubiquitin-like moiety, such as in SUMOylated

uracil/thymine DNA glycosylase,[16] or in cases where the ubiquitin moiety may physically destabilize the protein prior to its degradation,[17] it is expected that the modification would be confined to a specific location. However, in other cases, where no important interactions occur, such as in modified PCNA,[20] other lysine residues may be utilized as modification sites, or the original location may shift. Indeed, we found a significant enrichment of Lys residues in the vicinity of non-conserved ubiquitination sites. The fact that a significant portion of the lysine residues that serve as ubiquitination sites was found in prokaryotic species implies that, at least in some cases, already-existing residues were 'scavenged' by the ubiquitination machinery, and may lend support to the notion that the location of the sites is not always important. Indeed, it was shown that engineered ubiquitinated-PCNA with lysine residues at different locations on the surface functions similarly to the ubiquitinated wildtype.[66] It is not clear how common these cases are, and the data on the capabilities of different E3 ligases to direct ubiquitin conjugation to different locations are relatively scarce. It would be interesting to compare between types of ubiquitination machineries and the use of different available residues or shifts that may have occurred in modification sites over the course of evolution when sufficient data become available. It is important to note, however, that in cases where various lysine residues can be utilized by the ubiquitination machinery, the tight regulation of this process must stem from other origins—such as the effective concentration of the enzymatic ubiquitination machinery components in the vicinity of the substrates. Additionally, our analysis of lysine residues' presence in the vicinity of non-conserved ubiquitination sites highlights the importance of looking at windows surrounding functional sites, when testing for their evolutionary conservation.

We also studied the development of complex posttranslational regulation as manifested in several modifications that affect one another. While some of these complex regulatory modules seem to be very ancient, and we could not trace their gradual evolution, in others cases we saw evidence for variations in orthologous sequences from the experimentally characterized proteins. In some of these cases, this variation may reflect a shift in the location of components of the regulation modules—as exemplified in the ubiquitination sites in HIF1α and, putatively, the ubiquitination-related phosphosite in NEMO. In other cases, these variations may reflect the course the complex regulation followed during its emergence—such as in the case of $p19^{4inkd}$, where phosphorylation sites that are thought to precede ubiquitination in mammals seem to be missing in fish, perhaps reflecting the recent addition of a regulatory layer on top of the basal ubiquitination process. Further studies on the evolution of complex posttranslational regulations are needed to shed light on how these intricate regulation processes have developed.

In summary, our study gives primary insights into the origins and evolution of ubiquitination sites. Characterization of additional ubiquitination sites by either comprehensive proteomic approaches or mutational studies in the future will allow researchers to address additional fundamental questions. These include exploring the relationship between the class of ubiquitination enzymes and the evolutionary conservation of the sites they modify, and investigating how changes in ubiquitination patterns are related to changes in regulatory pathways. In addition, it would be interesting to investigate evolutionary patterns of ubiquitination which are related to specific pathways, when sufficient data will become available.

Acknowledgements

We would like to thank Peter Hornbeck, Predrag Radivojac, Ilan Wapinski and Gabriel Östlund and Vladimir Uversky for supplying data and Itay Mayrose for advice regarding rate4site. This work was supported by the Kimmelman Center for Macromolecular Assemblies and the Israeli Science Foundation. Y.L. is the incumbent of the Lillian and George Lyttle Career Development Chair.

References

1 T. Ravid and M. Hochstrasser, Diversity of degradation signals in the ubiquitin-proteasome system, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 679–690.
2 A. Hershko and A. Ciechanover, The ubiquitin system, *Annu. Rev. Biochem.*, 1998, **67**, 425–479.
3 S. Guindon and O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, *Syst. Biol.*, 2003, **52**, 696–704.
4 C. Grabbe, K. Husnjak and I. Dikic, The spatial and temporal organization of ubiquitin networks, *Nat. Rev. Mol. Cell Biol.*, 2011, **12**, 295–307.
5 J. Spence, R. R. Gali, G. Dittmar, F. Sherman and M. Karin, *et al.*, Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain, *Cell*, 2000, **102**, 67–76.
6 A. Varshavsky, The N-end rule, *Cell*, 1992, **69**, 725–735.
7 A. Varshavsky, Regulated protein degradation, *Trends Biochem. Sci.*, 2005, **30**, 283–286.
8 M. Hochstrasser, Origin and function of ubiquitin-like proteins, *Nature*, 2009, **458**, 422–429.
9 A. Varshavsky, The N-end rule pathway and regulation by proteolysis, *Protein Sci.*, 2011, **20**, 1298–1345.
10 E. V. Koonin, The origin and early evolution of eukaryotes in the light of phylogenomics, *Genome Biol.*, 2010, **11**, 209.
11 D. Finley, B. Bartel and A. Varshavsky, The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis, *Nature*, 1989, **338**, 394–401.
12 M. J. Pearce, J. Mintseris, J. Ferreyra, S. P. Gygi and K. H. Darwin, Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis, *Science*, 2008, **322**, 1104–1107.
13 M. A. Humbard, H. V. Miranda, J. M. Lim, D. J. Krause and J. R. Pritz, *et al.*, Ubiquitin-like small archaeal modifier proteins (SAMPs) in Haloferax volcanii, *Nature*, 2010, **463**, 54–60.
14 T. Nunoura, Y. Takaki, J. Kakuta, S. Nishi and J. Sugahara, *et al.*, Insights into the evolution of archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group, *Nucleic Acids Res.*, 2011, **39**, 3204–3223.
15 L. M. Iyer, A. M. Burroughs and L. Aravind, The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains, *Genome Biol.*, 2006, **7**, R60.
16 D. Baba, N. Maita, J. G. Jee, Y. Uchimura and H. Saitoh, *et al.*, Crystal structure of SUMO-3-modified thymine-DNA glycosylase, *J. Mol. Biol.*, 2006, **359**, 137–147.
17 T. Hagai and Y. Levy, Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 2001–2006.
18 T. Hagai, A. Azia, A. Toth-Petroczy and Y. Levy, Intrinsic disorder in ubiquitination substrates, *J. Mol. Biol.*, 2011, **412**, 319–324.
19 J. Gsponer, Who initiates proteasomal degradation?, *J. Mol. Biol.*, 2011, **412**, 317–318.

20 B. D. Freudenthal, J. E. Brogie, L. Gakhar, C. M. Kondratick and M. T. Washington, Crystal structure of SUMO-modified proliferating cell nuclear antigen, *J. Mol. Biol.*, 2011, **406**, 9–17.
21 R. E. Amir, H. Haecker, M. Karin and A. Ciechanover, Mechanism of processing of the NF-kappa B2 p100 precursor: identification of the specific polyubiquitin chain-anchoring lysine residue and analysis of the role of NEDD8-modification on the SCF(beta-TrCP) ubiquitin ligase, *Oncogene*, 2004, **23**, 2540–2547.
22 M. D. Petroski and R. J. Deshaies, Context of multiubiquitin chain attachment influences the rate of Sic1 degradation, *Mol. Cell*, 2003, **11**, 1435–1444.
23 T. K. Fung, C. H. Yam and R. Y. Poon, The N-terminal regulatory domain of cyclin A contains redundant ubiquitination targeting sequences and acceptor sites, *Cell Cycle*, 2005, **4**, 1411–1420.
24 R. W. King, M. Glotzer and M. W. Kirschner, Mutagenic analysis of the destruction signal of mitotic cyclins and structural characterization of ubiquitinated intermediates, *Mol. Biol. Cell*, 1996, **7**, 1343–1357.
25 B. Macek, F. Gnad, B. Soufi, C. Kumar and J. V. Olsen, *et al.*, Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation, *Mol. Cell Proteomics*, 2008, **7**, 299–307.
26 P. Beltrao, J. C. Trinidad, D. Fiedler, A. Roguev and W. A. Lim, *et al.*, Evolution of phosphoregulation: comparison of phosphorylation patterns across yeast species, *PLoS Biol.*, 2009, **7**, e1000134.
27 L. J. Holt, B. B. Tuch, J. Villen, A. D. Johnson and S. P. Gygi, *et al.*, Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution, *Science*, 2009, **325**, 1682–1686.
28 C. R. Landry, E. D. Levy and S. W. Michnick, Weak functional constraints on phosphoproteomes, *Trends Genet.*, 2009, **25**, 193–197.
29 A. N. Ba and A. M. Moses, Evolution of characterized phosphorylation sites in budding yeast, *Mol. Biol. Evol.*, 2010, **27**, 2027–2037.
30 V. E. Gray and S. Kumar, Rampant purifying selection conserves positions with posttranslational modifications in human proteins, *Mol. Biol. Evol.*, 2011, **28**, 1565–1568.
31 Z. Wang, G. Ding, L. Geistlinger, H. Li and L. Liu, *et al.*, Evolution of protein phosphorylation for distinct functional modules in vertebrate genomes, *Mol. Biol. Evol.*, 2011, **28**, 1131–1140.
32 C. Park and J. Zhang, Genome-wide evolutionary conservation of N-glycosylation sites, *Mol. Biol. Evol.*, 2011, **28**, 2351–2357.
33 B. T. Weinert, S. A. Wagner, H. Horn, P. Henriksen and W. R. Liu, *et al.*, Proteome-wide mapping of the Drosophila acetylome demonstrates a high degree of conservation of lysine acetylation, *Sci. Signaling*, 2011, **4**, ra48.
34 L. Freschi, M. Courcelles, P. Thibault, S. W. Michnick and C. R. Landry, Phosphorylation network rewiring by gene duplication, *Mol. Syst. Biol.*, 2011, **7**, 504.
35 A. L. Chernorudskiy, A. Garcia, E. V. Eremin, A. S. Shorina and E. V. Kondratieva, *et al.*, UbiProt: a database of ubiquitylated proteins, *BMC Bioinf.*, 2007, **8**, 126.
36 E. Jain, A. Bairoch, S. Duvaud, I. Phan and N. Redaschi, *et al.*, Infrastructure for the life sciences: design and implementation of the UniProt website, *BMC Bioinf.*, 2009, **10**, 136.
37 P. V. Hornbeck, I. Chabra, J. M. Kornhauser, E. Skrzypek and B. Zhang, PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation, *Proteomics*, 2004, **4**, 1551–1561.
38 A. C. Berglund, E. Sjolund, G. Ostlund and E. L. Sonnhammer, InParanoid 6: eukaryotic ortholog clusters with inparalogs, *Nucleic Acids Res.*, 2008, **36**, D263–D266.
39 I. Wapinski, A. Pfeffer, N. Friedman and A. Regev, Natural history and evolutionary principles of gene duplication in fungi, *Nature*, 2007, **449**, 54–61.
40 J. Peng, D. Schwartz, J. E. Elias, C. C. Thoreen and D. Cheng, *et al.*, A proteomics approach to understanding protein ubiquitination, *Nat. Biotechnol.*, 2003, **21**, 921–926.
41 R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, 2004, **32**, 1792–1797.
42 T. Pupko, R. E. Bell, I. Mayrose, F. Glaser and N. Ben-Tal, Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues, *Bioinformatics*, 2002, **18**(Suppl 1), S71–S77.
43 F. Abascal, R. Zardoya and D. Posada, ProtTest: selection of best-fit models of protein evolution, *Bioinformatics*, 2005, **21**, 2104–2105.
44 Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, *Mol. Biol. Evol.*, 2007, **24**, 1586–1591.
45 N. Goldman, J. L. Thorne and D. T. Jones, Assessing the impact of secondary structure and solvent accessibility on protein evolution, *Genetics*, 1998, **149**, 445–458.
46 J. Nilsson, M. Grahn and A. P. Wright, Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins, *Genome Biol.*, 2011, **12**, R65.
47 A. Toth-Petroczy and D. S. Tawfik, Slow protein evolutionary rates are dictated by surface-core association, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11151–11156.
48 C. J. Brown, A. K. Johnson and G. W. Daughdrill, Comparing models of evolution for ordered and disordered proteins, *Mol. Biol. Evol.*, 2010, **27**, 609–621.
49 Z. Dosztanyi, V. Csizmok, P. Tompa and I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, *Bioinformatics*, 2005, **21**, 3433–3434.
50 *Structure and Function of Intrinsically Disordered Proteins*, ed. P. Tompa, Taylor & Francis Group, Boca Raton, 2009, 1 edn, 331 p.
51 B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker and V. N. Uversky, PONDR-FIT: a meta-predictor of intrinsically disordered amino acids, *Biochim. Biophys. Acta*, 2010, **1804**, 996–1010.
52 J. Prilusky, C. E. Felder, T. Zeev-Ben-Mordehai, E. H. Rydberg and O. Man, *et al.*, FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded, *Bioinformatics*, 2005, **21**, 3435–3438.
53 V. Sobolev, R. C. Wade, G. Vriend and M. Edelman, Molecular docking using surface complementarity, *Proteins*, 1996, **25**, 120–129.
54 B. Rost, G. Yachdav and J. Liu, The PredictProtein server, *Nucleic Acids Res.*, 2004, **32**, W321–W326.
55 L. J. Jensen, T. S. Jensen, U. de Lichtenberg, S. Brunak and P. Bork, Co-evolution of transcriptional and post-translational cell-cycle regulation, *Nature*, 2006, **443**, 594–597.
56 A. M. Moses, M. E. Liku, J. J. Li and R. Durbin, Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 17713–17718.
57 J. Boulais, M. Trost, C. R. Landry, R. Dieckmann and E. D. Levy, *et al.*, Molecular characterization of the evolution of phagosomes, *Mol. Syst. Biol.*, 2010, **6**, 423.
58 C. Shou, N. Bhardwaj, H. Y. Lam, K. K. Yan and P. M. Kim, *et al.*, Measuring the evolutionary rewiring of biological networks, *PLoS Comput. Biol.*, 2011, **7**, e1001050.
59 G. E. Lienhard, Non-functional phosphorylations?, *Trends Biochem. Sci.*, 2008, **33**, 351–352.
60 P. Radivojac, V. Vacic, C. Haynes, R. R. Cocklin and A. Mohan, *et al.*, Identification, analysis, and prediction of protein ubiquitination sites, *Proteins*, 2009, **78**, 365–380.
61 T. Y. Lee, S. A. Chen, H. Y. Hung and Y. Y. Ou, Incorporating distant sequence features and radial basis function networks to identify ubiquitin conjugation sites, *PLoS One*, 2011, **6**, e17331.
62 B. T. Dye and B. A. Schulman, Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins, *Annu. Rev. Biophys. Biomol. Struct.*, 2007, **36**, 131–150.
63 G. Wu, G. Xu, B. A. Schulman, P. D. Jeffrey and J. W. Harper, *et al.*, Structure of a beta-TrCP1-Skp1-beta-catenin complex: destruction motif binding and lysine specificity of the SCF(beta-TrCP1) ubiquitin ligase, *Mol. Cell*, 2003, **11**, 1445–1456.
64 S. Paltoglou and B. J. Roberts, HIF-1alpha and EPAS ubiquitination mediated by the VHL tumour suppressor involves flexibility in the ubiquitination mechanism, similar to other RING E3 ligases, *Oncogene*, 2007, **26**, 604–609.
65 T. T. Huang, S. M. Wuerzberger-Davis, Z. H. Wu and S. Miyamoto, Sequential modification of NEMO/IKKgamma by

SUMO-1 and ubiquitin mediates NF-kappaB activation by genotoxic stress, *Cell*, 2003, **115**, 565–576.
66 J. Chen, Y. Ai, J. Wang, L. Haracska and Z. Zhuang, Chemically ubiquitylated PCNA as a probe for eukaryotic translesion DNA synthesis, *Nat. Chem. Biol.*, 2010, **6**, 270–272.
67 S. Bergink and S. Jentsch, Principles of ubiquitin and SUMO modifications in DNA repair, *Nature*, 2009, **458**, 461–467.
68 Y. Fridman, N. Palgi, D. Dovrat, S. Ben-Aroya and P. Hieter, *et al.*, Subtle alterations in PCNA-partner interactions severely impair DNA replication and repair, *PLoS Biol.*, 2010, **8**, e1000507.
69 S. Das-Bradoo, H. D. Nguyen, J. L. Wood, R. M. Ricke and J. C. Haworth, *et al.*, Defects in DNA ligase I trigger PCNA ubiquitination at Lys 107, *Nat. Cell Biol.*, 2010, **12**, 74–79pp. 71–20.
70 A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner and C. J. Brown, Intrinsic protein disorder in complete genomes, *Genome Inform. Ser. Workshop Genome Inform.*, 2000, **11**, 161–171.
71 J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton and D. T. Jones, Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *J. Mol. Biol.*, 2004, **337**, 635–645.
72 C. Loenarz, M. L. Coleman, A. Boleininger, B. Schierwater and P. W. Holland, *et al.*, The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens, *EMBO Rep.*, 2011, **12**, 63–70.
73 K. T. Rytkonen and J. F. Storz, Evolutionary origins of oxygen sensing in animals, *EMBO Rep.*, 2011, **12**, 3–4.
74 S. M. Paltoglou and B. J. Roberts, Role of the von Hippel–Lindau tumour suppressor protein in the regulation of HIF-1alpha and its oxygen-regulated transactivation domains at high cell density, *Oncogene*, 2005, **24**, 3830–3835.
75 S. M. Pearlman, Z. Serber and J. E. Ferrell, Jr, A mechanism for the evolution of phosphorylation sites, *Cell*, 2011, **147**, 934–946.
76 U. Basu, Y. Wang and F. W. Alt, Evolution of phosphorylation-dependent regulation of activation-induced cytidine deaminase, *Mol. Cell*, 2008, **32**, 285–291.