

SCIENTIFIC REPORTS

OPEN

E2 superfamily of ubiquitin-conjugating enzymes: constitutively active or activated through phosphorylation in the catalytic cleft

Ilaria Valimberti¹, Matteo Tiberti¹, Matteo Lambrughì¹, Boris Sarcevic² & Elena Papaleo^{1,†}

Protein phosphorylation is a modification that offers a dynamic and reversible mechanism to regulate the majority of cellular processes. Numerous diseases are associated with aberrant regulation of phosphorylation-induced switches. Phosphorylation is emerging as a mechanism to modulate ubiquitination by regulating key enzymes in this pathway. The molecular mechanisms underpinning how phosphorylation regulates ubiquitinating enzymes, however, are elusive. Here, we show the high conservation of a functional site in E2 ubiquitin-conjugating enzymes. In catalytically active E2s, this site contains aspartate or a phosphorylatable serine and we refer to it as the conserved E2 serine/aspartate (CES/D) site. Molecular simulations of substrate-bound and -unbound forms of wild type, mutant and phosphorylated E2s, provide atomistic insight into the role of the CES/D residue for optimal E2 activity. Both the size and charge of the side group at the site play a central role in aligning the substrate lysine toward E2 catalytic cysteine to control ubiquitination efficiency. The CES/D site contributes to the fingerprint of the E2 superfamily. We propose that E2 enzymes can be divided into constitutively active or regulated families. E2s characterized by an aspartate at the CES/D site signify constitutively active E2s, whereas those containing a serine can be regulated by phosphorylation.

Phosphorylation is a major mechanism of post-translational regulatory modification, influencing almost all cellular functions^{1–3}. It has been estimated that more than 40% of proteins are phosphorylated *in vivo*⁴. Phosphorylation offers a dynamic mechanism to regulate protein function due to its fast kinetics and reversibility^{5–7}. Addition of a phosphate group to serine, threonine or tyrosine residues can induce large conformational modifications in proteins^{8–18}. This is driven by electrostatic perturbation or steric hindrance, which can affect the protein energy landscape, intermolecular protein-ligand interactions and enzymatic activity^{18–22}. In addition to the critical role of phosphorylation in normal cellular function, aberrant regulation of phosphorylation-induced electrostatic switches is involved in numerous pathologies, such as cancers^{23–25}. High-throughput mass spectrometric studies have identified thousands of post-translational phosphorylation sites on different proteins^{26,27}. However, precisely defining the role

¹Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126, Milan (Italy). ²Cell Cycle and Cancer Unit, St. Vincent's Institute of Medical Research and The Department of Medicine, St. Vincent's Hospital, The University of Melbourne, Fitzroy, Melbourne, Victoria 3065, Australia. [†]Present address: Unit of Statistics, Bioinformatics and Registry, Danish Cancer Society Research Center, Strandboulevarden 49, 2100, Copenhagen, Denmark. Correspondence and requests for materials should be addressed to E.P. (email: elena.papaleo@unimib.it or elena.papaleo.78@gmail.com)

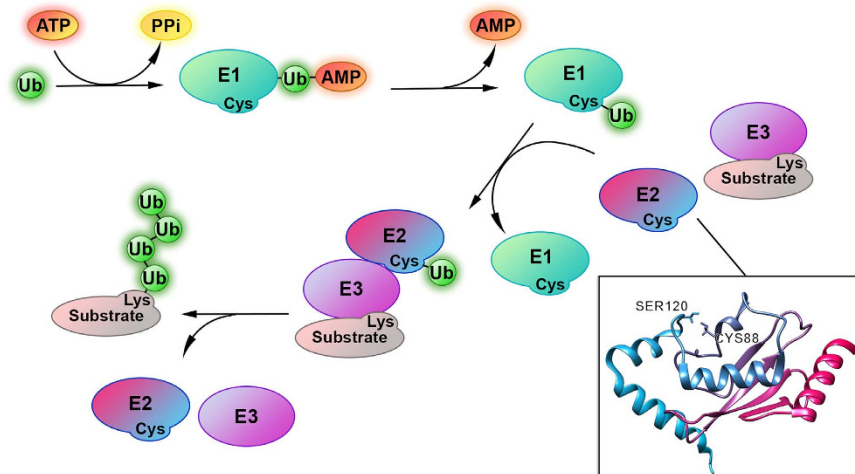


Figure 1. The process of ubiquitination consists of the sequential and cooperative actions of three classes of enzymes that catalyze the attachment of the protein Ubiquitin (Ub, green spheres) to substrate proteins. Ub is first activated by ubiquitin-activating enzyme (E1, green ovals) by an ATP-dependent reaction. The Ub-AMP product remains bound to the E1. Then a cysteine residue attacks the Ub, leading to an E1~Ub thioester intermediate. Subsequently the Ub is transferred to a conserved cysteine in the active site of an ubiquitin-conjugating enzyme (E2, hot pink-azure ovals). The E2s catalyze the transfer of Ub to a lysine on the substrate (pink-grey ovals) to form an isopeptide bond, together with an ubiquitin ligase (E3, purple ovals). Alternatively, E2s can transfer Ub to a catalytic cysteine on HECT family E3s, which then transfer Ub onto substrate to form an isopeptide bond. E3s are important for recognition and selection of the substrate. Certain E2/E3 pairs catalyze several rounds of Ub attachment to generate substrates with poly-Ub chains. Substrate can be mono or poly-ubiquitinated and these modifications mark the substrate for different destiny and functions. In the bottom-right panel the structure of Human Ubiquitin-conjugating enzyme E2 B (PDB ID 1JAS, hHR6B) is shown as cartoon. The catalytic cysteine (Cys88) present in the Ubiquitin Conjugation (UBC) domain of E2, that forms thioester bond with Ub, and the serine (Ser120) in the conserved E2 Ser/Asp site (CES/D), whose phosphorylation enhances E2 catalytic activity, are highlighted as sticks.

of these modifications on protein structure and dynamics at a molecular and atomic-level is still experimentally challenging.

Cyclin-dependent kinases (Cdks) are key enzymes involved in promoting cell cycle progression in eukaryotic cells²⁸. Over the last decade, several studies demonstrated that Cdks can regulate enzymes of the ubiquitination cascade^{29–39}. The ubiquitination pathway consists of three classes of enzymes, which catalyze the attachment of the protein Ubiquitin (Ub) to various substrates (Fig. 1)⁴⁰. In this enzymatic pathway the Ub C-terminus is first activated by an Ub-activating enzyme (E1) in an ATP-dependent reaction. The activated Ub is subsequently transferred to a conserved cysteine to one of the distinct Ub-conjugating enzymes (E2s), to generate a thioester bond. The E2s then catalyze the transfer of Ub to a substrate lysine to form an isopeptide bond, in conjunction with an Ub-protein ligase (E3), which is also important for substrate selection⁴¹. Since Ub contains seven lysines, in many cases certain E2/E3 pairs catalyze further rounds of Ub attachment to generate substrates with poly-Ub chains^{42–44}.

E2 Ub-conjugating enzymes are critical in this cascade, as they regulate both the topology of the poly-Ub chains and the processivity of the polyubiquitination reaction^{45–47}. The human genome encodes more than 40 E2s^{45–47}, which catalyze attachment of Ub to different lysines on protein substrate or Ub, during mono- or poly-ubiquitination. This versatility allows for a multitude of distinct ubiquitination events and the generation of diverse protein-Ub structures^{42–44}. The E2 superfamily has been divided in 17 families by comprehensive phylogenetic analysis⁴⁸. E2s share a ~150-amino acid conserved catalytic core domain, which is the minimal sufficient unit for enzymatic activity (Fig. 1). This catalytic core domain is known as the Ubiquitin Conjugation (UBC) domain and contains the catalytic cysteine, which forms a thioester bond with Ub^{45–47}. E2s recently gained attention as targets to design inhibitory molecules^{49–53} since their aberrant regulation is associated with cancer or neurodegenerative diseases^{54–56}.

Although several high-resolution three-dimensional (3D) structures of E2 and E3 enzymes have yielded many insights into the mechanisms of Ub or Ubiquitin-like (Ubl) attachment to substrates⁴¹, many of the fundamental mechanisms of Ub attachment remain to be elucidated. Studies with the small Ubl modifier SUMO and its human cognate E2, Ube2I, identified a triad of conserved residues in Ube2I, which are crucial in activating the substrate acceptor lysine to accept SUMO^{57,58}. Therefore, Asn85 of the His-Pro-Asn (HPN) motif, Tyr87 and Asp127 play important roles in lowering the substrate lysine pKa to allow its nucleophilic attack of the Ube2I~SUMO thioester bond. Several studies have demonstrated

the importance of these residues for the catalytic function of different E2s, especially the HPN motif^{59–64}. Asp127 plays an important role in the catalytic cleft of Ube2I⁵⁷, a role which is likely to be generally important in E2s. For example, Ser139 in the yeast E2, Cdc34, represents the residue homologous to Ube2I Asp127. Mutation of Cdc34 Ser139 to Asp (Cdc34 S139D), abrogated the ability of Cdc34 to generate lysine 48-linked poly-Ub chains but did not affect the ability of this enzyme to attach Ub to protein substrate⁶². Therefore, mutation of serine 139 from serine to aspartate, converted Cdc34 from a poly-ubiquitinating to a mono-ubiquitinating enzyme⁶². In other E2s, such as *Saccharomyces cerevisiae* Rad6 and its human ortholog hHR6A (Ube2A), the site homologous to Ube2I Asp127 is a serine (Ser120, Fig. 1), whose phosphorylation enhances E2 catalytic activity^{29,34}. The important regulatory nature of this site is further exemplified by mutational studies, demonstrating that the activity of hHR6A can be upregulated or downregulated, depending on the amino acid in this position^{29,34}. Hence mutations of hHR6A Ser120 to alanine (S120A) or threonine (S120T) abolished hHR6A Ub-conjugating activity, whereas S120E mutant variant displayed 35% of the activity of the wild type hHR6A. Conversely, a S120D mutation increased the activity of this enzyme three-fold, compared to the wild-type hHR6A, similar to the four-fold increase observed when hHR6A Ser120 is phosphorylated (pS120). Hereafter, we refer to this site as the conserved E2 Ser/Asp site (CES/D).

In this study we investigate this site in the whole E2 superfamily by phylogenetic analysis and more than 60 atomistic MD simulations, which are a powerful approach to investigate the role of phosphorylation in atomic detail^{9,12,15–18,20,65–72}. We investigate how mutation or phosphorylation of the site affects the dynamics, structure and function of free E2s and E2s in complex with the target substrate. Our studies show that E2 families that are catalytically active conserve a negatively charged residue (Asp) or serine (Ser). Phosphorylation of E2s with a serine at this site, by kinases such Cdks, increases E2 activity^{29,34}. Our data also indicates that a delicate balance between the size and charge of the CES/D site regulates alignment of the substrate lysine toward the E2 catalytic cysteine to control E2 catalytic activity.

Results and Discussion

A conserved E2 Ser/Asp (CES/D) site in catalytically active E2 members is characterized by a constitutive or regulated negatively charged residue. In the E2 superfamily, the site, which corresponds to hHR6A Ser120 (CES/D), generally contains an acidic residue (mostly Asp) or a ‘phosphorylatable’ serine residue (Fig. 2, Supplementary Fig. S1 and Table S1). Conversely, catalytically inactive members of the E2 superfamily, which retain the conserved E2 fold but lack a catalytic cysteine and are not involved in conjugation of Ub or Ub-like (Ubl) modifiers, such as the UEVs, do not show a strictly conserved residue at the same site (Supplementary Fig. S2). The only exceptions that we observed are families 5 and 13, which we included in the phylogenetic analysis since most of the members show a conserved cysteine residue at the catalytic site. Family 5 E2s are named Noncanonical Ubiquitin Conjugating Enzymes (NCUBE) as they lack the conserved HPN motif and have structural reorganizations in the region surrounding the catalytic cleft⁷³ making difficult a structural alignment with the canonical E2s. Family 13 E2 enzymes do not show a conserved HPN motif in all the members. Moreover a non-canonical Ubl-conjugating activity has been reported for the human member of family 13, i.e., UBE2W^{74–76}, so that we labeled this group as NCUBE E2s too. In addition, we should mention that Ube2D4 and Ube2U of family 4 do not feature a canonical Ubl-conjugating activity⁴⁴. Ube2U does not conserve an acid or phosphorylatable residue at the CES/D site, whereas most of the members of family 4 feature an Asp residue at the site. In families 14 and 15 the site is occupied by another negatively charged residue, i.e. a glutamate, which in members of family 15 that are Ube2L3-like are in position –1 (labeled as L3 E₋₁ in Fig. 2) with respect to the canonical CES/D site. To our knowledge, these two families are not extensively characterized, and thus additional studies will be needed to clarify the role of the glutamate at the CES/D site. Our analysis strongly supports the notion that the physical-chemical properties of this E2 site is very important for Ub and Ubl-conjugation activity and that the selective pressure for maintenance of a serine or aspartate at this position is missing in the UEVs or in other proteins, which are deprived of Ubl-conjugation activity but still conserve an E2-like fold.

Interestingly, the conserved site is almost invariably accompanied by a proximal C-terminal proline (Fig. 2, Supplementary Fig. S1). Therefore, E2s with a phosphorylatable serine at the conserved site conform to the consensus phosphorylation motif for Cdks⁷⁷ (Fig. 2, Supplementary Fig. S1). S120 is known to be phosphorylated by Cdk kinases in Rad6³⁴ and hHR6A from family 2²⁹ and our analysis suggests that it might be phosphorylated in other E2 enzymes. Mutational studies of the CES/D site of hHR6A indicate that a negatively charged residue at this site is important for optimal E2 conjugation activity²⁹. An aspartate residue at the CES/D site in Ube2I of family 7 is necessary for the ubiquitination of the target substrate in cooperation with the invariant asparagine of the HPN motif and a tyrosine residue⁵⁷.

In summary, we propose that a negatively charged residue at this site, which can be either constitutive or modulated by phosphorylation, is important for optimal E2 catalytic activity. We here defined an additional conserved site for the E2 superfamily, that together with HPN, the catalytic cysteine and the invariant tryptophan⁴⁸ contribute to define the signature of these enzymes.

A negatively charged residue of a specific size at the E2 CES/D site allows conserved principal motions of the free enzyme in solution. The conserved patterns observed in our phylogenetic analysis prompted us to compare the 3D structures and dynamics of representative E2 enzymes, which

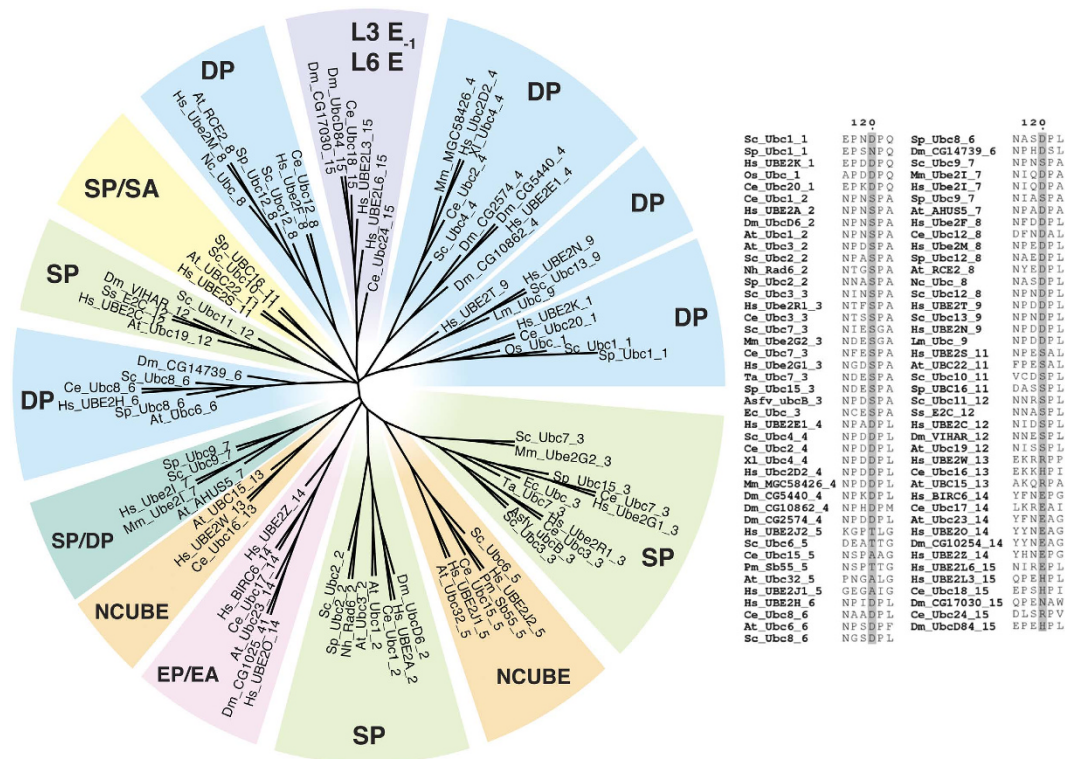


Figure 2. Phylogenetic tree of E2 families with the consensus patterns for the positions corresponding to the CES/D site. In the panel on the right, we reported a portion of the sequence alignment reported in Supplementary Fig. S1 to show the residues in the immediate proximity of the CES/D site. Each sequence is labelled as ‘Organism_E2_Family’ so for example *S. cerevisiae* Ubc2 (Rad6) from family 2 is labelled as ‘Sc_Ubc2_2’. The phylogenetic tree is coloured according to the conserved motif at the CES/D site within each family using light blue, light green, maroon, yellow, pink and violet colours for DP, SP, SP/DP, SP/SA, EP/EA and E sequence motifs, respectively. The NCUBE families are highlighted in orange.

feature either an aspartate or a serine at the CES/D site, by multi-replicate atomistic MD simulations in explicit solvent (Supplementary Table S2). In particular, we selected two E2 enzymes containing a constitutive aspartic acid residue (yeast Ubc1 and human Ube2I) and three enzymes with serine (yeast Rad6, human hHR6A and hHR6B) at the site. Ube2I is a SUMO-conjugating enzyme and thus we included in our study the Ub-conjugating enzyme Ubc1, which also features an aspartate at the CES/D site for a better coverage of the E2 members. Moreover, several Rad6 and hHR6A serine 120 point mutant variants have been experimentally investigated and provide a rich source of information on the role of the CES/D site in E2 catalytic activity^{29,34}. We also carried out MD simulations of all these mutants and phospho-variants. The information on the simulated systems and the MD replicates are summarized in the Supplementary Table S2.

To compare the dynamic properties and the structural effects induced by different amino acids or phosphorylation at the CES/D site, we carried out a principal component analysis (PCA) of a C α concatenated trajectory of all the MD simulations⁷⁸ using the same reference subspace and calculating the covariance matrix of corresponding C α atoms according to the multiple sequence alignment reported in Supplementary Fig. S3. PCA is a technique often employed to compare dynamic properties of different protein variants in an ‘essential’ subspace that retains only the largest amplitude fluctuations^{79–82}. It would be tempting to use the first three eigenvectors (Principal Components, PCs) to provide a 3D representation of the motions described by the different variants. This approach might encounter the risk to emphasize non-genuine differences. Indeed, for different single mutant variants of the same protein or even for proteins with a conserved fold, such as those that we are comparing here, the differences are more likely to be ascribed to changes in the eigenvalues rather than direction of motions. The usage of only the first or the first three PCs can thus encounter the risk of observing differences due to a ‘crossing-over’ of corresponding eigenvalues between difference PCs of two different systems. This is indeed the case in our simulations (Fig. 3A). We estimated the overlap between eigenvalues of the first PCs of the wild type variants and each of the first 10 PCs of the mutant variants, here exemplified in Fig. 3A for the eigenvalues of the first PC of wild type hHR6A which overlaps with higher index PCs (up to the fifth PC) in all the mutants, including either active and inactive variants.

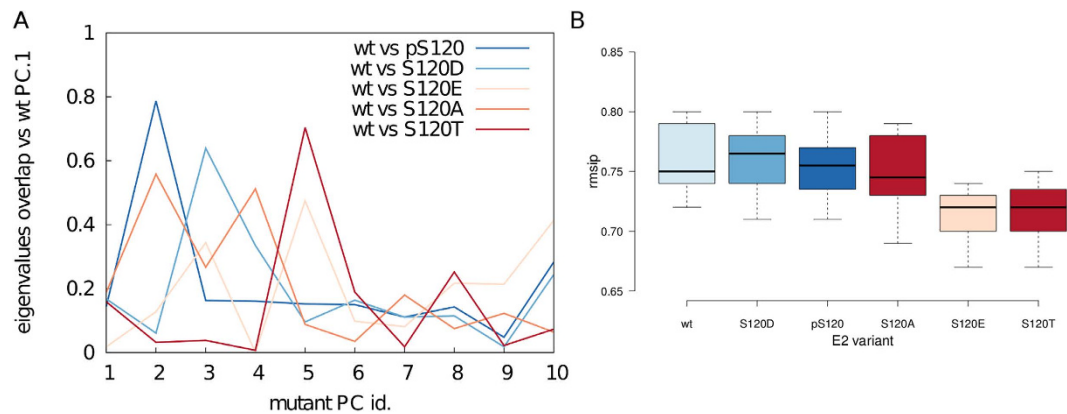


Figure 3. PCA analysis of MD trajectories of wild type, mutant or phosphorylated E2 variants.

(A) We evaluate if there is a crossing-over of eigenvalues between difference PCs of wild type and mutant E2 variants. The example of the overlap between the eigenvalues of the PC.1 of wild type hHR6A and the eigenvalues of each of the first 10 PCs of hHR6A mutants is shown to illustrate this result. We noticed that an interchange of PCs between wild type and mutant variants and this occurs independently of the effect that the mutation has on E2 Ub-activity. Indeed, it is equally true for active and inactive mutant variants. (B) To properly compare our different simulations we estimate the RMSIP between the essential subspaces described by the first 20 PCs of wild-type variants compared each other or compared to S120D, pS120, S120A, S120T and S120E mutant variants respectively. A boxplot analysis of the data has been carried out with *BoxPlotR*¹⁰¹. Each dataset is coloured according to the effects on Ub-conjugating activity (different shade of blue for active variants and different shade of orange for partially active or inactive variants). Center lines show the medians, box limit indicates the 25th and 75th percentiles and whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles.

We thus quantified the similarities in protein dynamics of the different variants using the first 20 PCs, which account for more than 80% of the total variance of the system. We calculated the root mean square inner product (RMSIP) as a measure of overlap. In particular, we carried out a boxplot analysis (Fig. 3B) comparing the RMSIP values obtained between different classes. Indeed, we compared the RMSIP values obtained when only wild-type E2 enzymes (hHR6A, hHR6B, UBC1 and UBE2I) are compared to each other with the RMSIP values of wild-type E2s vs S120D, pS120, S120A, S120E and S120T variants, respectively. We can see a generally conserved dynamics with high RMSIP values, i.e. quite overlapping essential subspaces. There are no radical changes, but we can observe a subtle effect due to fully inactive S120T and partially active S120E mutations, which feature slightly lower RMSIP values when compared to the wild type E2s, a signal of potential perturbation induced on native dynamics upon these mutations. S120A, which is also an inactive mutation for Ub-conjugating activity is characterized by a native-like behaviour in the free E2, suggesting that S120A does not affect the dynamics and stability of the free protein and it is more likely to exert an effect directly with respect to the substrate binding.

In summary, a PCA analysis of our MD trajectories shows that mutations at the CES/D site that abolish or impair (S120E and S120T) the Ub-conjugation activity induce perturbations in E2 native dynamics, whereas the other variants do not feature any remarkable effects.

Phosphorylation of CES/D site increases the solvent accessibility of the catalytic cysteine. To further unveil the effects induced upon phosphorylation or mutation in the active site, we calculated the solvent accessible surface (SAS) of the catalytic cysteine in our simulations. The phospho-hHR6A and -Rad6 variants present a higher average solvent accessibility of the cysteine than the non-phosphorylated or S120D counterparts (hHR6A and Rad6) (Fig. 4A). The solvent accessibility of the phospho-variants is comparable to accessibility of Ubc1, in which an aspartate residue occupies the CES/D site (Fig. 4A). Ube2I contains an insertion in proximity of the active site, which corresponds structurally to a loop that protrudes from the catalytic cleft and resembles the acidic loop of Cdc34-like E2 enzymes⁸³. Opening/closing motions of this loop may partially occlude the catalytic cysteine, as observed for Cdc34⁸³. Thus we did not include Ube2I in the SAS analysis since we are interested in assessing if mutations within the catalytic cleft correlate with accessibility of the catalytic cysteine and E2 activity. Mutations to alanine or threonine do not affect the accessibility of the catalytic cleft overall. Conversely, the glutamate mutation introduces a longer side chain in the catalytic cleft compared to either pS120 or the S120D mutation, partially occluding the catalytic cysteine.

A subtle balance between charge and size at the CES/D site is required for optimal Ub-conjugation activity. Mutations of the conserved CES/D site of hHR6A, Rad6 or Ube2I do not remarkably affect the Ub/Ubl-charging step of the E2, i.e., the step in which the E1 enzyme transfers the

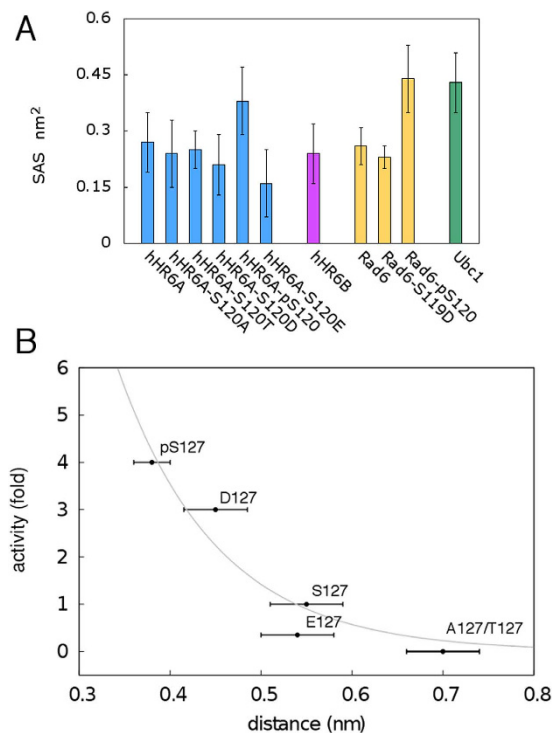


Figure 4. (A) Average Solvent Accessible Surface (SAS) of the side chain of the E2 catalytic cysteine in the simulations of the wild type, mutant or phosphorylated E2 variants. The SAS (nm²) of the side-chain atoms of the catalytic cysteine is reported as an average value over four replicates of each variants and the corresponding bars indicate the associated standard deviations. hHR6A, hHR6B, Rad6 and Ubc1 variants are shown in blue, magenta, yellow and dark green, respectively. (B) Correlation between the Ub-conjugating activity and the distance between the catalytic cysteine of the E2 donor and the side chain of the substrate acceptor lysine in wild type, mutant and phosphorylated Ube2I-RanGAP1 complexes. The experimental data are taken from²⁹. Each point represents the average distance calculated over four replicates of each variant and the corresponding bars indicate the associated standard deviations. The plot has been derived upon exponential fitting of the data according to $f(x) = e^{-(x^c)} * b$ with fit parameters b and c set equal to 136.28 and 9.12, which provide a correlation of 0.98.

Ub/Ubl to the cognate E2^{29,57}. Rather, the mutations are mainly impacting on the E2 Ub/Ubl-conjugation activity^{29,34,57}, i.e. the step in which the Ub/Ubl-loaded E2 enzyme, alone or together with the cognate E3, transfers the Ub/Ubl to the lysine of the target substrate. Thus, we investigate if different side chains at the CES/D site influence the orientation of the substrate lysine side chain by MD simulations of Ube2I-RanGAP1 wild type (Asp127) and mutant complexes (D127S, D127E, D127T, D127A and 127phospho-S). We selected this complex for our simulations since experimental structures of hHR6A in complex with its target substrate are not available and cannot be modelled by homology. MD simulations are known to encounter the risk of artefacts when they are initialized with a modelled structure rather than an experimental one⁸⁴, especially if sequence identity between targets and templates is low. Moreover, we limited our MD study to describe the local changes in the residues that are in the proximity of either the catalytic cysteine, the CES/D site or the acceptor lysine of the substrate.

We first monitored the distances between the substrate lysine and either CES/D side chain or the catalytic cysteine. We observed that the catalytically active or partially active variants (D127, D127S, pS127 and D127E) allow for interactions between the residue at the CES/D site and the substrate lysine at distances below 0.48 nm, with the phospho-variant exhibiting the minimal distance of ~0.38 nm. The interaction is not observed when alanine or threonine occupies the CES/D (distances higher than 0.7 nm). A similar trend is observed when we calculated the distances between the side chains of the substrate lysine and the E2 catalytic cysteine. The distances between the catalytic cysteine and target lysine residues in our simulated E2-RanGAP1 complex, correlate with an exponential behaviour to the experimentally observed Ub-conjugation activity of hHR6A on its substrate histone H2A²⁹ (Fig. 4B). The phospho-variant has the minimal distance between the two residues at values lower than 0.35, thus very close to the optimal distance for Ub/Ubl transfer to the target lysine⁸⁵. Conversely, the interaction between the lysine and the catalytic cysteine in the inactive variants (D127A and D127T) is sub-optimal, with the distance increasing up than 0.7 nm (Fig. 4B). The optimal distance for Ub/Ubl-transfer between the catalytic cysteine and the acceptor lysine of the substrate is expected to be around 0.3–0.35 nm⁸⁵.

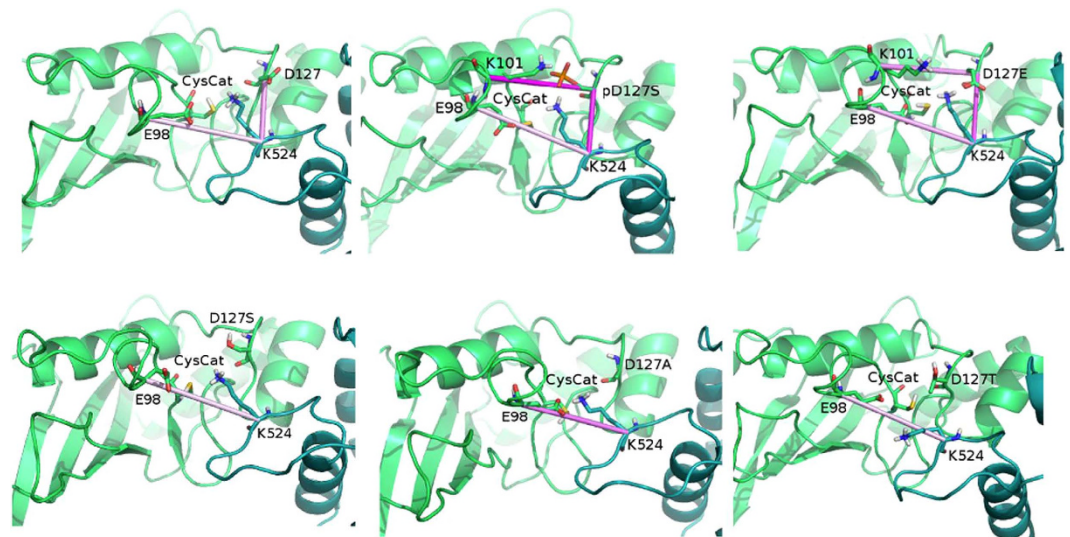


Figure 5. Network of electrostatic interactions between the substrate lysine and the residues surrounding the catalytic cleft of the E2 enzyme in MD simulations of wild type, mutant and phosphorylated Ube2I-RanGAP1 complexes. RanGAP1 and the E2 enzyme are shown as green and blue marine cartoons, respectively. Pink lines represent salt bridges with different shades of colors according to their persistence in the MD simulations.

Therefore, even minor perturbations can cause detrimental effects on E2 Ub/Ubl-conjugation activity, where even small changes in the distance between the catalytic cysteine and the target lysine are correlated to a marked drop in activity for the D127A and D127T mutants.

We thus speculate that the Ub transfer from the E2 catalytic cysteine to the target lysine is facilitated by the closer interaction between the E2~Ub thioester and the acceptor lysine. Moreover, the negatively charged group of the phospho-Ser can efficiently maintain the target lysine side chain in a conformation suitable for interaction with the catalytic cysteine, due to the electrostatic nature of the interaction between the residue at the E2 CES/D site and the target lysine of the substrate. Indeed, an ion pair and hydrogen bonds are observed between the side chains of the two residues in all the simulation frames of Ube2I-RanGAP1_{p127S} simulations (Supplementary Table S3, Fig. 5). The variant with a serine at this site can still interact with the lysine by hydrogen bonds but the interaction is weaker and observed in 10% of the Ube2I-RanGAP1_{D127S} simulations (Supplementary Table S3, Fig. 5).

The mutations to alanine or threonine cannot provide an interaction with the target lysine and causes a departure of the lysine side chain from the E2 catalytic cleft (Supplementary Table S3, Fig. 5 and Supplementary Fig. S4). We could have expected that the threonine, being a polar residue, could still interact with the target lysine. On the contrary, we observed that the -CH₃ group of threonine side chain is placed in a way that does not allow for interaction with the substrate lysine (Supplementary Fig. S4).

Moreover, we observed that the residue at the CES/D site is involved in a larger network of electrostatic interactions that contribute to the positioning of the substrate lysine within the E2 catalytic cleft (Fig. 5) in Ube2I-RanGAP1, Ube2I-RanGAP1_{pS127} and Ube2I-RanGAP1_{D127E}. The network includes Glu98 in all the three systems and also Lys101 in RanGAP1_{pS127} and Ube2I-RanGAP1_{D127E}. The substrate lysine is indeed bridging both the negatively charged residue at position 127 and the Glu98. In the inactive mutants the interaction of the lysine with the CES/D site is lost and the interaction with Glu98 is not sufficient to provide a proper orientation of the substrate lysine within the E2 catalytic cleft (Fig. 5 and S4). Indeed, the Glu98 attracts and positions the side chain of the substrate lysine outside the catalytic pocket.

The presence of a glutamate residue at position 127 allows for a tight interaction with the lysine of the substrate (Supplementary Table S3) but it impacts the lysine-cysteine distance, which is higher than 0.54 nm. The longer side chain of the glutamate residue with respect to an aspartate induces, in some of the simulated frames, a conformation of the lysine not optimal for interaction with the E2 catalytic cysteine, by attracting the lysine side chain away from the cysteine and outside the E2 catalytic cleft (Fig. 5 and S4). We observe that the D127E mutation results in a simulated conformational ensemble of the E2-substrate complex in which only a minor population of the substrate lysine is in a conformation competent for Ub/Ubl-conjugation. The majority of the conformations are not optimal for the activity of the enzyme, explaining the 75% drop in Ub-conjugating activity observed *in vitro*²⁹. The active variants, especially pS127, provide a suitable balance between side chain charge and size to maintain the substrate

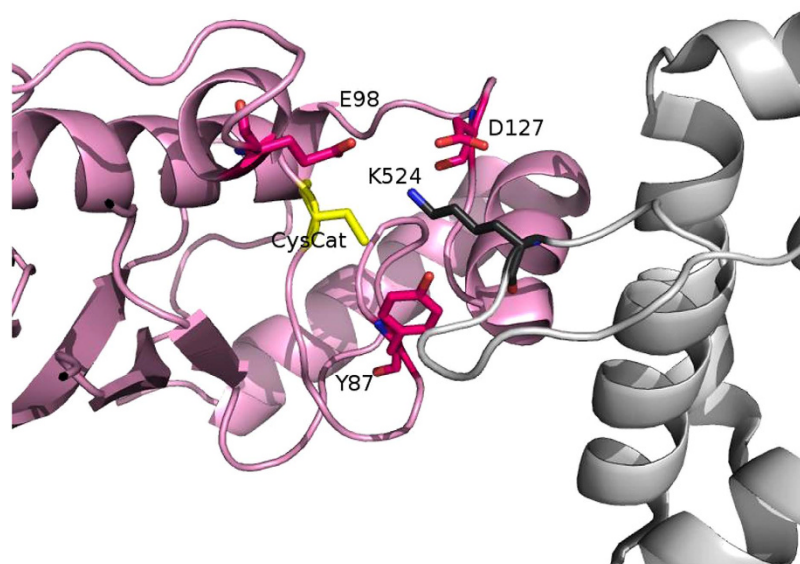


Figure 6. Tyr87, Glu98 and the CES/D residue cooperate in providing the optimal coordination for the acceptor lysine side chain within the catalytic cleft of the E2 enzyme. RanGAP1 and Ube2I are shown as grey and pink cartoons, respectively.

lysine closer to the catalytic site with the lysine side chain bridging the negatively charged residue at position 127 and the Glu98 (Fig. 5).

Discussion

Studies on Ube2I, which is the E2 enzyme specific for SUMO Ub-like modifier, showed that mutations at Asn85, Tyr87, and Asp127 result in defects in SUMO-conjugating activity to the substrate lysine, as severe as those achieved upon mutation of the catalytic cysteine⁵⁷. Asp127 corresponds to Ser120 of hHR6A, which is phosphorylated by CDKs²⁹. Phosphorylation or the mutation of Ser120 to aspartate can substantially increase the ubiquitin-conjugating activity of this enzyme²⁹. The catalytic defects observed for the Ube2I mutants have been ascribed to the inability of the mutated residues to coordinate the substrate lysine side chain within the E2 active site. This has been suggested by the observation that the lysine side chain is poorly ordered in the crystallographic structures of the Ube2I mutant variants in complex with the substrate. In addition, these residues decrease the lysine pKa in a range where catalysis can occur under physiological pH conditions, uncovering a dual role for this triad of residues⁵⁷.

Despite the central role that site 120 has in E2 enzymes, we have very limited knowledge of its role in E2 structure, dynamics and function, especially at the atomic level.

Progresses in genome sequencing and annotation of E2 enzymes, allowed the classification of E2 enzymes in 17 families of orthologs⁴⁸. We added new E2 enzymes from diverse organisms to the original pool of E2 sequences and focused on members of E2 families for which a catalytic Ub/Ubl-conjugating activity has been proved. Our analysis shows that the position corresponding to Ser120 of hHR6A is highly conserved in all the catalytically active E2 enzymes, allowing us to classify this position as a new CES/D functional site in E2s. At this position, generally a serine or negatively charged aspartic residue has been selected by evolution. The serine can be phosphorylated to generate a negatively charged side chain. Our studies demonstrate that not only is the charge of this residue important but also its size. Therefore, substitution with the longer glutamic acid side chain results in a less efficient enzyme²⁹. Our study suggests this is due to compromised native dynamics and the inability to efficiently coordinate the target lysine of the substrate in the catalytic site. Specifically, we show that when a residue distinct from Asp or the phospho-Ser is located at the CES/D site, the substrate acceptor lysine side chain is more mobile and results in orientations that fail to interact efficiently with the catalytic cysteine. We also observed a correlation between the Ub/Ubl-conjugating activity and the distance between the acceptor lysine side group and the E2 catalytic cysteine. It has been estimated that the optimal distance for Ub/Ubl transfer from the E2 ~ Ub thioester to the substrate acceptor lysine is lower than 0.35 nm⁸⁵. We observed the closest distances with E2s containing a phospho-Ser or an Asp at the CES/D site. Remarkably, even minor changes in the distance between the lysine and the catalytic cysteine cause rather drastic effects in Ub-conjugation activity.

In addition to the importance of the conserved serine/aspartate site, other key residues in the E2 catalytic site can be identified. In particular, Glu98 contributes to a tridentate network of electrostatic interactions, Glu98-Lys_{substrate}-CES/D127. Tyr87 is another crucial site for the coordination of the substrate lysine acceptor side chain and cooperates with the CES/D site⁵⁷. The aromatic side chain of this

residue can retain the lysine within the E2 catalytic cleft by steric effects (Fig. 6). Indeed, the mutation Y87A is known to reduce the Ubl-conjugating activity in Ube2I, similar to that observed for the D127A mutation⁵⁷. Conversely, a conservative Y87F mutation that retains the properties of the tyrosine side chain does not alter the activity of the enzyme⁵⁷. In agreement with these data, our multiple alignment shows that tyrosine or phenylalanine residues are conserved at the positions homologous to Ube2I Y87 site (Supplementary Fig. S1).

Residues at three sites, Tyr87, Asp or Ser at CES/D site and to a minor extent, Glu98 (Ube 2I numbering) thus cooperate for the optimal coordination of the substrate lysine within the E2 catalytic pocket (Fig. 6). Furthermore, phosphorylation of Ser at the CES/D site can generate a negative charge in this position to increase activity and therefore this site can regulate E2 activity through post-translational modification. The ability to dynamically and reversibly modulate the CES/D site through phosphorylation is likely to have crucial importance for E2 enzymes which need to be more active at a specific times in the cell, such as during cell cycle progression or particular signalling events. Our MD investigation provides insight on the role of the CES/D site at an atomic level. This residue is able to efficiently coordinate the substrate lysine by a subtle balance between its charge and size. The conservation of a phosphorylatable serine or an aspartate in almost all catalytically active E2s suggests that this is a general mechanism for the E2 Ub/Ubl-conjugating enzymes. The CES/D motif in the E2 superfamily characterized in this study, together with the catalytic cysteine, the HPN motif and the invariant Trp, defines a fingerprint of this superfamily of enzymes at the heart of the ubiquitination cascade.

Methods

Multiple sequence alignment and phylogenetic analysis. We obtained the unrooted phylogenetic tree using the PHYLIP package⁸⁶ on the basis of a multiple sequence alignment that has been carried out with ClustalW2⁸⁷. The multiple sequence alignment was manually corrected according to information on functional or structural conserved residues and secondary structures, along with a comparison with a structural alignment obtained by DALI and it is reported in Supplementary Fig. S1⁸⁸. The multiple sequence alignment includes the sequences of the UBC domains of several E2 enzymes belonging to each of the 17 E2 families⁴⁸. We discarded E2 enzymes belonging to families 10 and 16 from the analysis since they do not feature a catalytic cysteine and are known to be catalytically inactive^{48,73}. We also discarded E2 enzymes from family 17 since they do not present a canonical fold of the UBC domain⁴⁸. The sequences corresponding to the UBC domain of each E2 enzymes have been determined by intra-family sequence alignments, since for each family at least one member with known 3D structure is available. Initially, we collected more than 250 E2 sequences and redundant sequences were discarded by DivergentSet program⁸⁹, leading to a selection of 85 sequences in the final alignment (Supplementary Table S1 and Fig. S1). The phylogenetic analysis was carried out using bootstrap with 100 replicates and 12 trials. We employed the Maximum Likelihood method using the Proml program and a Jones-Taylor-Thornton substitution matrix. FigTree software (<http://tree.bio.ed.ac.uk/software/figtree>) was used for the tree illustration.

Molecular dynamics simulations. A summary of the starting structures, number of replicates and simulation length used for simulations of wild-type hHR6A, hHR6B, Rad6, Ube2I and Ubc1, along with hHR6A_{S120A}, hHR6A_{S120T}, hHR6A_{S120E}, hHR6A_{S120D}, hHR6A_{PS120}, Rad6_{S120A}, Rad6_{S120D}, Rad6_{PS120} is provided in Supplementary Table S1. hHR6A variants were modelled using as template hHR6B. They share 96% sequence identity and do not feature any insertion or deletion in the sequence. The modelling procedure was carried out with Modeller v.9.3⁹⁰ and the final model evaluated using the AIDE software⁹¹, which classifies them as excellent quality models with predicted root mean square deviation (RMSD) from the native structure, TM-score and LG-score of 0.14 nm, 0.78 and 0.25, respectively. The mutant variants have been selected according to the mutants previously investigated experimentally^{29,34}. Simulations of Ube2I-RanGAP1 complex have been also carried out for both the wild type (Asp127) and all the mutant variants studied on the free proteins (Ube2I_{D127S}-RanGAP1, Ube2I_{D127A}-RanGAP1, Ube2I_{D127E}-RanGAP1, Ube2I_{D127T}-RanGAP1 and Ube2I_{PS127}-RanGAP1).

MD simulations were performed with Gromacs 4.5⁹² and a modified version of the Gromos 54a7 force field⁹³ to account for phospho-residues, as provided by VIENNA-PTM server^{94,95}. The systems were solvated in a dodecahedral box (minimum distance between protein and box edges: 0.9 nm) of SPC (Simple Point Charge) water molecules, applying periodic boundary conditions. The ionization state of residues was set to be consistent with neutral pH and the tautomeric form of histidine residues was derived using GROMACS tools and confirmed by visual inspection. The system was equilibrated according a protocol previously applied to other E2 enzymes^{12,96}. We collected overall 64 MD simulations with two or four replicates for each system, collecting in total more than nine μ s of MD sampling (Supplementary Table S2). The use of multiple MD replicates of the same system can help to identify recurring features and to avoid simulations artifacts⁹⁷. Simulations were carried out in the isothermal-isobaric ensemble at 300 K at 1 atm using an external Berendsen bath with thermal and pressure coupling of 0.1 and 1 ps, respectively. The LINCS algorithm was used to constraint the heavy atom bonds to use a two fs time-step. Long-range electrostatic interactions were calculated using the Particle-mesh Ewald (PME) summation scheme. Van der Waals and Coulomb interactions were truncated at 0.9 nm.

We discarded the first ten ns of MD simulations from analysis, since they are required to reach stable main-chain root mean square values in all the MD replicates (Supplementary Figs S5–S7). For each system, the equilibrated portions of each replicate were joined together in a macro-trajectory, which is representative of the different direction of sampling around the starting structure of that system.

Principal Component Analysis (PCA). PCA of MD trajectories reveals large amplitude concerted motions in the MD ensemble through the eigenvectors of the mass-weighted covariance matrix of the atomic positional fluctuations⁷⁸. We thus carried out PCA using either C α or all protein atoms on the macro-trajectories of the wild type and mutant variants of E2 enzymes collected in our study using as a guide the alignment reported in Supplementary Fig. S3. We indeed used the same reference subspace to better compare them. The similarity between two eigenvectors, each calculated from a different MD simulation, was evaluated by computing the inner product between them. We have used this method to systematically compare the most relevant eigenvectors of different protein variants to verify where a interchange between eigenvalues of different principal components (PCs) was occurring. We could observe a crossing-over of eigenvalues in the first 10 PCs of different systems. We thus compared pairwise the first 20 PCs of each system using as a measure of overlap the Root Mean Square Inner Product (RMSIP)^{98,99}.

The cosine content of the principal components (eigenvectors) of the covariance matrix of atomic positional fluctuations is an index that can range between 0 (no cosine) and 1 (perfect cosine). It has been demonstrated that insufficient sampling lead to high cosine content values in MD simulations, representative of random motions¹⁰⁰. We thus calculated the cosine content of the first 20 PCs of each MD replicate, achieving values always lower than 0.30.

References

- Deribe, Y. L., Pawson, T. & Dikic, I. Post-translational modifications in signal integration. *Nat. Struct. Mol. Biol.* **17**, 666–72 (2010).
- Cohen, P. The origins of protein phosphorylation. *Nat. Cell Biol.* **4**, E127–30 (2002).
- Pawson, T. & Scott, J. D. Protein phosphorylation in signaling—50 years and counting. *Trends Biochem. Sci.* **30**, 286–90 (2005).
- Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–20 (2015).
- Nishi, H., Shaytan, A. & Panchenko, A. R. Physicochemical mechanisms of protein regulation by phosphorylation. *Front. Genet.* **5**, 270 (2014).
- Olsen, J. V. *et al.* Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–48 (2006).
- Hunter, T. Why nature chose phosphate to modify proteins. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 2513–6 (2012).
- Gallagher, E., Gao, M., Liu, Y.-C. & Karin, M. Activation of the E3 ubiquitin ligase Itch through a phosphorylation-induced conformational change. *Proc. Natl. Acad. Sci. USA* **103**, 1717–22 (2006).
- Dago, A. E. *et al.* Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis. *Proc. Natl. Acad. Sci. USA* **109**, E1733–42 (2012).
- Kim, S. *et al.* Phosphorylation of the spindle checkpoint protein Mad2 regulates its conformational transition. *Proc. Natl. Acad. Sci. USA* **107**, 19772–19777 (2010).
- Burke, J. R., Deshong, A. J., Pelton, J. G. & Rubin, S. M. Phosphorylation-induced conformational changes in the retinoblastoma protein inhibit E2F transactivation domain binding. *J. Biol. Chem.* **285**, 16286–93 (2010).
- Papaleo, E. *et al.* An acidic loop and cognate phosphorylation sites define a molecular switch that modulates ubiquitin charging activity in Cdc34-like enzymes. *PLoS Comput. Biol.* **7**, e1002056 (2011).
- Fraser, J. A., Vojtesek, B. & Hupp, T. R. A novel p53 phosphorylation site within the MDM2 ubiquitination signal: I. phosphorylation at SER269 *in vivo* is linked to inactivation of p53 function. *J. Biol. Chem.* **285**, 37762–72 (2010).
- Kim, S., Ma, B., Tsai, C.-J. & Csermely, P. Allosteric conformational barcodes direct signaling in the cell. *Structure* **21**, 1509–21 (2013).
- Autore, F., Pagano, B., Fornili, A., Rittinger, K. & Fraternali, F. In silico phosphorylation of the autoinhibited form of p47(phox): insights into the mechanism of activation. *Biophys. J.* **99**, 3716–25 (2010).
- De Simone, A. *et al.* Structures of the excited states of phospholamban and shifts in their populations upon phosphorylation. *Biochemistry* **52**, 6684–94 (2013).
- Stanley, N., Esteban-Martín, S. & De Fabritiis, G. Kinetic modulation of a disordered protein domain by phosphorylation. *Nat. Commun.* **5**, 5272 (2014).
- Papaleo, E., Casiraghi, N., Arrigoni, A. & Ranzani, V. Atomistic Insights Into the Regulatory Mechanisms Mediated by Post-Translational Modifications: Molecular Dynamics Investigations. *Curr. Phys. Chem.* **2**, 344–362 (2012).
- Groban, E. S., Narayanan, A. & Jacobson, M. P. Conformational changes in protein loops and helices induced by post-translational phosphorylation. *PLoS Comput. Biol.* **2**, e32 (2006).
- Narayanan, A. & Jacobson, M. P. Computational studies of protein regulation by post-translational phosphorylation. *Curr. Opin. Struct. Biol.* **19**, 156–63 (2009).
- Roca, M., Aranda, J., Moliner, V. & Tuñón, I. Modeling methods for studying post-translational and transcriptional modifying enzymes. *Curr. Opin. Chem. Biol.* **16**, 465–71 (2012).
- Lätzer, J., Shen, T. & Wolynes, P. G. Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles. *Biochemistry* **47**, 2110–22 (2008).
- Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355–65 (2001).
- Reimand, J., Wagih, O. & Bader, G. D. The mutational landscape of phosphorylation signaling in cancer. *Sci. Rep.* **3**, 2651 (2013).
- Reimand, J., Wagih, O. & Bader, G. D. Evolutionary Constraint and Disease Associations of Post-Translational Modification Sites in Human Genomes. *PLoS Genet.* **11**, e1004919 (2015).
- Derouiche, A., Cousin, C. & Mijakovic, I. Protein phosphorylation from the perspective of systems biology. *Curr. Opin. Biotechnol.* **23**, 585–90 (2012).
- Ptacek, J. & Snyder, M. Charging it up: global analysis of protein phosphorylation. *Trends Genet.* **22**, 545–54 (2006).
- Harper, J. W. & Adams, P. D. Cyclin-dependent kinases. *Chem. Rev.* **101**, 2511–26 (2001).
- Sarcevic, B., Mawson, A., Baker, R. T. & Sutherland, R. L. Regulation of the ubiquitin-conjugating enzyme hHR6A by CDK-mediated phosphorylation. *EMBO J.* **21**, 2009–18 (2002).

30. Rizzardi, L. F. *et al.* CDK1-dependent inhibition of the E3 ubiquitin ligase CRL4CDT2 ensures robust transition from S Phase to Mitosis. *J. Biol. Chem.* **290**, 556–67 (2015).
31. Barbash, O., Lee, E. K. & Diehl, J. A. Phosphorylation-dependent regulation of SCF(Fbx4) dimerization and activity involves a novel component, 14-3-3 ϵ . *Oncogene* **30**, 1995–2002 (2011).
32. Shchebet, A., Karpiuk, O., Kremmer, E., Eick, D. & Johnsen, S. A. Phosphorylation by cyclin-dependent kinase-9 controls ubiquitin-conjugating enzyme-2A function. *Cell Cycle* **11**, 2122–7 (2012).
33. Su, Y.-F., Yang, T., Huang, H., Liu, L. F. & Hwang, J. Phosphorylation of Ubc9 by Cdk1 enhances SUMOylation activity. *PLoS One* **7**, e34250 (2012).
34. Wood, A., Schneider, J., Dover, J., Johnston, M. & Shilatifard, A. The Bur1/Bur2 complex is required for histone H2B monoubiquitination by Rad6/Bre1 and histone methylation by COMPASS. *Mol. Cell* **20**, 589–99 (2005).
35. Oh, R. S., Bai, X. & Rommens, J. M. Human homologs of Ubc6p ubiquitin-conjugating enzyme and phosphorylation of HsUbc6e in response to endoplasmic reticulum stress. *J. Biol. Chem.* **281**, 21480–90 (2006).
36. Park, H.-J. *et al.* Identification of phosphorylation sites of TOPORS and a role for serine 98 in the regulation of ubiquitin but not SUMO E3 ligase activity. *Biochemistry* **47**, 13887–96 (2008).
37. Smith, H. *et al.* Identification of the phosphorylation sites on the E3 ubiquitin ligase Pellino that are critical for activation by IRAK1 and IRAK4. *Proc. Natl. Acad. Sci. USA* **106**, 4584–90 (2009).
38. Kumar, B., Lecompte, K. G., Klein, J. M. & Haas, A. L. Ser(120) of Ubc2/Rad6 regulates ubiquitin-dependent N-end rule targeting by E3[alpha]/Ubr1. *J. Biol. Chem.* **285**, 41300–9 (2010).
39. Lewandowski, K. T. & Piwnicka-Worms, H. Phosphorylation of the E3 ubiquitin ligase RNF41 by the kinase Par-1b is required for epithelial cell polarity. *J. Cell Sci.* **127**, 315–27 (2014).
40. Streich, F. C. & Lima, C. D. Structural and functional insights to ubiquitin-like protein conjugation. *Annu. Rev. Biophys.* **43**, 357–79 (2014).
41. Dye, B. T. & Schulman, B. A. Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 131–50 (2007).
42. Trempe, J.-F. Reading the ubiquitin postal code. *Curr. Opin. Struct. Biol.* **21**, 792–801 (2011).
43. Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–29 (2012).
44. Van Wijk, S. J. L. *et al.* A comprehensive framework of E2-RING E3 interactions of the human ubiquitin-proteasome system. *Mol. Syst. Biol.* **5**, 295 (2009).
45. Van Wijk, S. J. L. & Timmers, H. T. M. The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins. *FASEB J.* **24**, 981–93 (2010).
46. Ye, Y. & Rape, M. Building ubiquitin chains: E2 enzymes at work. *Nat. Rev. Mol. Cell Biol.* **10**, 755–64 (2009).
47. Wenzel, D. M., Stoll, K. E. & Klevit, R. E. E2s: structurally economical and functionally replete. *Biochem. J.* **433**, 31–42 (2011).
48. Michelle, C., Vourc'h, P., Mignon, L. & Andres, C. R. What was the set of ubiquitin and ubiquitin-like conjugating enzymes in the eukaryote common ancestor? *J. Mol. Evol.* **68**, 616–28 (2009).
49. Ceccarelli, D. F. *et al.* An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. *Cell* **145**, 1075–87 (2011).
50. Sanders, M. A. *et al.* Novel inhibitors of Rad6 Ubiquitin Conjugating Enzyme: Design, synthesis, identification and functional characterization. *Mol. Cancer Ther.* **12**, 373–83 (2013), doi: 10.1158/1535-7163.MCT-12-0793.
51. Arrigoni, A., Bertini, L., De Gioia, L. & Papaleo, E. Inhibitors of the Cdc34 acidic loop: A computational investigation integrating molecular dynamics, virtual screening and docking approaches. *FEBS Open Bio* **4**, 473–84 (2014).
52. Pulvino, M. *et al.* Inhibition of proliferation and survival of diffuse large B-cell lymphoma cells by a small-molecule inhibitor of the ubiquitin-conjugating enzyme Ubc13-Uev1A. *Blood* **120**, 1668–77 (2012).
53. Huang, H. *et al.* E2 enzyme inhibition by stabilization of a low-affinity interface with ubiquitin. *Nat. Chem. Biol.* **10**, 156–63 (2014).
54. Hao, Z., Zhang, H. & Cowell, J. Ubiquitin-conjugating enzyme UBE2C: molecular biology, role in tumorigenesis, and potential as a biomarker. *Tumour Biol.* **33**, 723–30 (2012).
55. Voutsadakis, I. A. Ubiquitin- and ubiquitin-like proteins-conjugating enzymes (E2s) in breast cancer. *Mol. Biol. Rep.* **40**, 2019–34 (2013).
56. Edelmann, M. J., Nicholson, B. & Kessler, B. M. Pharmacological targets in the ubiquitin system offer new ways of treating cancer, neurodegenerative disorders and infectious diseases. *Expert Rev. Mol. Med.* **13**, e35 (2011).
57. Yunus, A. A. & Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nat. Struct. Mol. Biol.* **13**, 491–9 (2006).
58. Knipscheer, P. & Sixma, T. K. Divide and conquer: the E2 active site. *Nat. Struct. Mol. Biol.* **13**, 474–6 (2006).
59. Cook, B. W. & Shaw, G. S. Architecture of the catalytic HPN motif is conserved in all E2 conjugating enzymes. *Biochem. J.* **445**, 167–74 (2012).
60. Berndsen, C. E., Wiener, R., Yu, I. W., Ringel, A. E. & Wolberger, C. A conserved asparagine has a structural role in ubiquitin-conjugating enzymes. *Nat. Chem. Biol.* **9**, 154–56 (2013).
61. Wu, P.-Y. *et al.* A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J.* **22**, 5241–50 (2003).
62. Sadowski, M., Suryadinata, R., Lai, X., Heierhorst, J. & Sarcevic, B. Molecular basis for lysine specificity in the yeast ubiquitin-conjugating enzyme Cdc34. *Mol. Cell. Biol.* **30**, 2316–29 (2010).
63. Zhen, Y. *et al.* Exploring the RING-catalyzed ubiquitin transfer mechanism by MD and QM/MM calculations. *PLoS One* **9**, e101663 (2014).
64. Markin, C. J. *et al.* Catalytic proficiency of ubiquitin conjugation enzymes: balancing pK(a) suppression, entropy, and electrostatics. *J. Am. Chem. Soc.* **132**, 17775–86 (2010).
65. Meng, Y. & Roux, B. Locking the active conformation of c-Src kinase through the phosphorylation of the activation loop. *J. Mol. Biol.* **426**, 423–35 (2014).
66. Srinivasaraghavan, K., Nacro, K., Grüber, G. & Verma, C. S. Effect of Ser392 phosphorylation on the structure and dynamics of the polybasic domain of ADP ribosylation factor nucleotide site opener protein: a molecular simulation study. *Biochemistry* **52**, 7339–49 (2013).
67. Shen, T., Zong, C., Hamelberg, D., McCammon, J. A. & Wolynes, P. G. The folding energy landscape and phosphorylation: modeling the conformational switch of the NFAT regulatory domain. *FASEB J.* **19**, 1389–95 (2005).
68. Lindert, S., Cheng, Y., Kekenus-Huskey, P., Regnier, M. & McCammon, J. A. Effects of HCM cTnI Mutation R145G on Troponin Structure and Modulation by PKA Phosphorylation Elucidated by Molecular Dynamics Simulations. *Biophys. J.* **108**, 395–407 (2015).
69. Arrigoni, A., Grillo, B., Vitriolo, A., De Gioia, L. & Papaleo, E. C-Terminal acidic domain of ubiquitin-conjugating enzymes: a multi-functional conserved intrinsically disordered domain in family 3 of E2 enzymes. *J. Struct. Biol.* **178**, 245–59 (2012).
70. Caulfield, T. R. *et al.* Phosphorylation by PINK1 releases the UBL domain and initializes the conformational opening of the E3 ubiquitin ligase Parkin. *PLoS Comput. Biol.* **10**, e1003935 (2014).
71. Missimer, J. H., Steinmetz, M. O., van Gunsteren, W. F. & Dolenc, J. Influence of 63Ser phosphorylation and dephosphorylation on the structure of the stathmin helical nucleation sequence: a molecular dynamics study. *Biochemistry* **51**, 8455–63 (2012).

72. Velazquez, H. A. & Hamelberg, D. Conformation-directed catalysis and coupled enzyme-substrate dynamics in Pin1 phosphorylation-dependent cis-trans isomerase. *J. Phys. Chem. B* **117**, 11509–17 (2013).
73. Burroughs, A. M., Jaffee, M., Iyer, L. M. & Aravind, L. Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. *J. Struct. Biol.* **162**, 205–18 (2008).
74. Scaglione, K. M. *et al.* The ubiquitin-conjugating enzyme (E2) Ube2w ubiquitinates the N terminus of substrates. *J. Biol. Chem.* **288**, 18784–8 (2013).
75. Tatham, M. H., Plechanová, A., Jaffray, E. G., Salmen, H. & Hay, R. T. Ube2W conjugates ubiquitin to α -amino groups of protein N-termini. *Biochem. J.* **453**, 137–45 (2013).
76. Vittal, V. *et al.* Intrinsic disorder drives N-terminal ubiquitination by Ube2w. *Nat. Chem. Biol.* **11**, 83–9 (2015).
77. Takeda, D. Y., Wohlschlegel, J. A. & Dutta, A. A bipartite substrate recognition motif for cyclin-dependent kinases. *J. Biol. Chem.* **276**, 1993–7 (2001).
78. Amadei, A., Linssen, A. B. & Berendsen, H. J. Essential dynamics of proteins. *Proteins* **17**, 412–25 (1993).
79. Bhakat, S., Martin, A. J. M. & Soliman, M. E. S. An integrated molecular dynamics, principal component analysis and residue interaction network approach reveals the impact of M184V mutation on HIV reverse transcriptase resistance to lamivudine. *Mol. Biosyst.* **10**, 2215–28 (2014).
80. Invernizzi, G., Tiberti, M., Lambrugh, M., Lindorff-Larsen, K. & Papaleo, E. Communication routes in ARID domains between distal residues in helix 5 and the DNA-binding loops. *PLoS Comput. Biol.* **10**, e1003744 (2014).
81. Kamaraj, B. & Purohit, R. Computational screening of disease-associated mutations in OCA2 gene. *Cell Biochem. Biophys.* **68**, 97–109 (2014).
82. Masureadze, G. G., Zhou, R., Liwo, A., Xiao, Y. & Scheraga, H. A. Effects of mutation, truncation, and temperature on the folding kinetics of a WW domain. *J. Mol. Biol.* **420**, 350–65 (2012).
83. Papaleo, E. *et al.* An acidic loop and cognate phosphorylation sites define a molecular switch that modulates ubiquitin charging activity in Cdc34-like enzymes. *PLoS Comput. Biol.* **7**, e1002056 (2011).
84. Raval, A., Piana, S., Eastwood, M. P., Dror, R. O. & Shaw, D. E. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. *Proteins* **80**, 2071–9 (2012).
85. Verdecia, M. A. *et al.* Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. *Mol. Cell* **11**, 249–59 (2003).
86. Felsenstein, J. PHYLIP- Phylogeny inference package. *Cladistics* **5**, 164–166 (1989).
87. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–8 (2007).
88. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–9 (2010).
89. Widmann, J., Hamady, M. & Knight, R. DivergentSet, a tool for picking non-redundant sequences from large sequence collections. *Mol. Cell. Proteomics* **5**, 1520–32 (2006).
90. Eswar, N., Eramian, D., Webb, B., Shen, M.-Y. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol Biol* **426**, 145–159 (2008).
91. Mereghetti, P., Ganadu, M. L., Papaleo, E., Fantucci, P. & De Gioia, L. Validation of protein models by a neural network approach. *BMC Bioinformatics* **9**, 66 (2008).
92. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
93. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843–56 (2011).
94. Margreitter, C., Petrov, D. & Zagrovic, B. Vienna-PTM web server: a toolkit for MD simulations of protein post-translational modifications. *Nucleic Acids Res.* **41**, W422–6 (2013).
95. Petrov, D., Margreitter, C., Grandits, M., Oostenbrink, C. & Zagrovic, B. A systematic framework for molecular dynamics simulations of protein post-translational modifications. *PLoS Comput. Biol.* **9**, e1003154 (2013).
96. Papaleo, E. *et al.* Loop 7 of E2 enzymes: an ancestral conserved functional motif involved in the E2-mediated steps of the ubiquitination cascade. *PLoS One* **7**, e40786 (2012).
97. Caves, L. S., Evanseck, J. D. & Karplus, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* **7**, 649–66 (1998).
98. Hess, B. Convergence of sampling in protein simulations. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **65**, 031910 (2002).
99. Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R. & Gioia, L. De. Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case. *J. Mol. Graph. Model.* **27**, 889–899 (2009).
100. Hess, B. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E. Stat. Phys. Plasmas. Fluids. Relat. Interdiscip. Topics* **62**, 8438–48 (2000).
101. Spitzer, M., Wildenhain, J., Rappsilber, J. & Tyers, M. BoxPlotR: a web tool for generation of box plots. *Nat. Methods* **11**, 121–2 (2014).

Acknowledgements

The results of this research have been achieved using the IS CRA-CINECA (E2-Cons/HP10CQE11L, Dyn-E2/HP10COBO3E and PE2cat/HP10C4FKX6). Funding for work on ubiquitination to BS has been supported by Cancer Council Victoria, the Association for International Cancer Research, the Multiple Myeloma Research Foundation and the National Health and Medical Research Council.

Author Contributions

E.P. conceived and designed the research; E.P., I.V., M.T. and M.L. carried out the experiments; E.P. and I.V. analysed the data; E.P., I.V. and B.S. discussed the data; E.P. and B.S. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Valimberti, I. *et al.* E2 superfamily of ubiquitin-conjugating enzymes: constitutively active or activated through phosphorylation in the catalytic cleft. *Sci. Rep.* **5**, 14849; doi: 10.1038/srep14849 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>