



# Lysine deserts prevent adventitious ubiquitylation of ubiquitin-proteasome components

Caroline Kampmeyer<sup>1</sup> · Martin Grønbaek-Thygesen<sup>1</sup> · Nicole Oelerich<sup>2</sup> · Michael H. Tatham<sup>3</sup> · Matteo Cagiada<sup>1</sup> · Kresten Lindorff-Larsen<sup>1</sup> · Wouter Boomsma<sup>4</sup> · Kay Hofmann<sup>2</sup> · Rasmus Hartmann-Petersen<sup>1</sup>

Received: 21 December 2022 / Revised: 15 March 2023 / Accepted: 17 April 2023 / Published online: 9 May 2023  
© The Author(s) 2023

## Abstract

In terms of its relative frequency, lysine is a common amino acid in the human proteome. However, by bioinformatics we find hundreds of proteins that contain long and evolutionarily conserved stretches completely devoid of lysine residues. These so-called lysine deserts show a high prevalence in intrinsically disordered proteins with known or predicted functions within the ubiquitin-proteasome system (UPS), including many E3 ubiquitin-protein ligases and UBL domain proteasome substrate shuttles, such as BAG6, RAD23A, UBQLN1 and UBQLN2. We show that introduction of lysine residues into the deserts leads to a striking increase in ubiquitylation of some of these proteins. In case of BAG6, we show that ubiquitylation is catalyzed by the E3 RNF126, while RAD23A is ubiquitylated by E6AP. Despite the elevated ubiquitylation, mutant RAD23A appears stable, but displays a partial loss of function phenotype in fission yeast. In case of UBQLN1 and BAG6, introducing lysine leads to a reduced abundance due to proteasomal degradation of the proteins. For UBQLN1 we show that arginine residues within the lysine depleted region are critical for its ability to form cytosolic speckles/inclusions. We propose that selective pressure to avoid lysine residues may be a common evolutionary mechanism to prevent unwarranted ubiquitylation and/or perhaps other lysine post-translational modifications. This may be particularly relevant for UPS components as they closely and frequently encounter the ubiquitylation machinery and are thus more susceptible to nonspecific ubiquitylation.

**Keywords** Lysine · Ubiquitin · Proteasome · Degradation · PTM · Intrinsically disordered protein

---

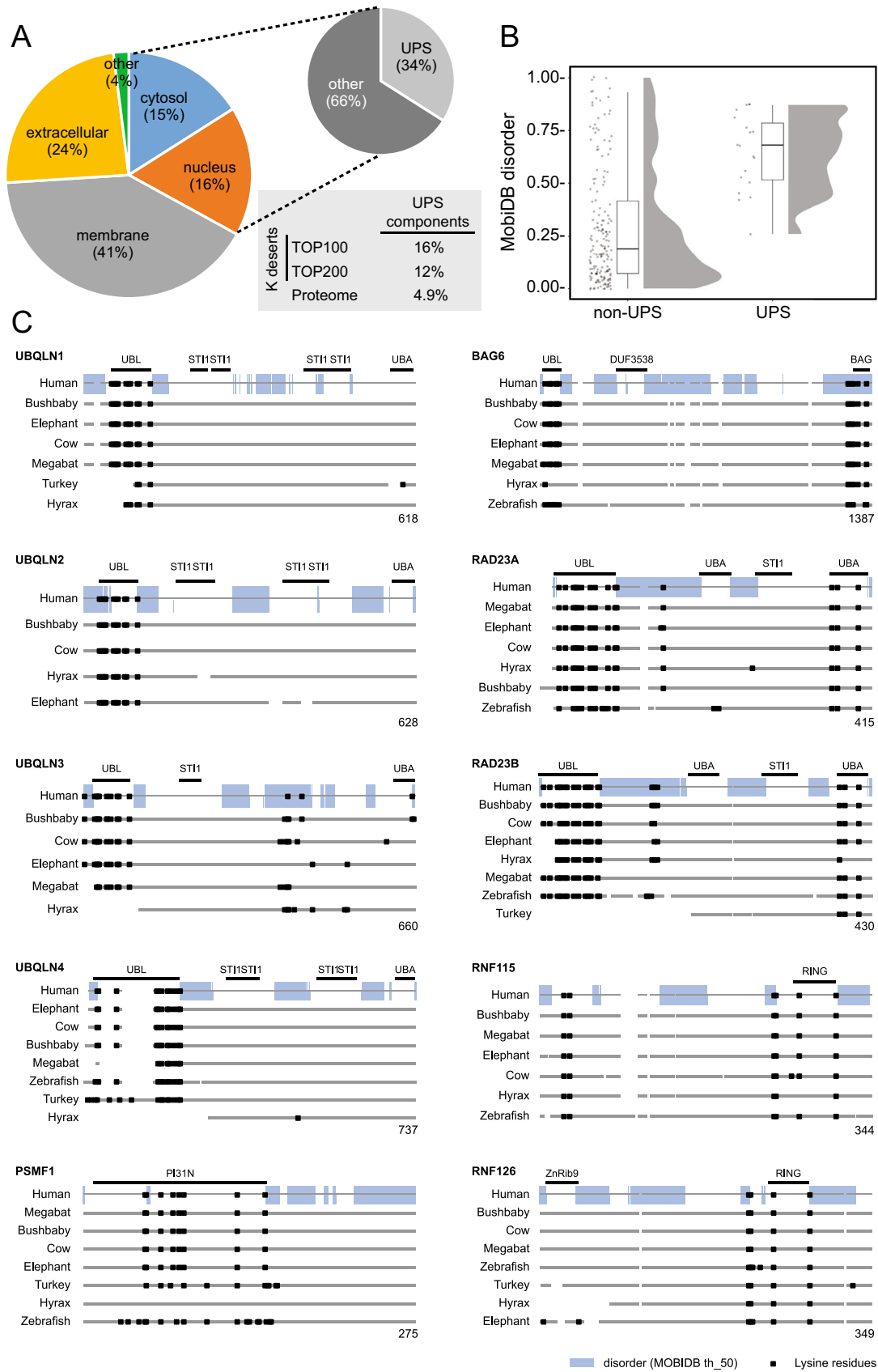
Caroline Kampmeyer, Martin Grønbaek-Thygesen have contributed equally to this work.

- 
- ✉ Wouter Boomsma  
wb@di.ku.dk
  - ✉ Kay Hofmann  
kay.hofmann@uni-koeln.de
  - ✉ Rasmus Hartmann-Petersen  
rhpetersen@bio.ku.dk

- <sup>1</sup> Department of Biology, The Linderstrøm-Lang Centre for Protein Science, University of Copenhagen, Copenhagen, Denmark
- <sup>2</sup> Institute for Genetics, University of Cologne, Cologne, Germany
- <sup>3</sup> Centre for Gene Regulation and Expression, Sir James Black Centre, School of Life Sciences, University of Dundee, Dundee, UK
- <sup>4</sup> Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

## Introduction

Of the 20 different amino acids that make up proteins, lysine represents a special case because its side-chain uniquely contains a primary amino group. This amino group is subject to various post-translation modifications (PTMs) and thus forms covalent bonds to other molecules and macromolecules, including other proteins. Lysine PTMs include the well-described methylation, acetylation, ubiquitylation and sumoylation, but more recently a range of other modifications have been observed [1]. In the cytosolic and nuclear compartments of eukaryotic cells, ubiquitylation represents a common PTM of lysine residues. Here, the small and highly conserved ubiquitin protein is covalently linked to a target protein, typically via an isopeptide bond between the C-terminal carboxyl group of ubiquitin and the  $\epsilon$ -amino group of a lysine residue within the target [2]. The process is catalyzed by three distinct enzymatic activities. First, a thioester bond is created between the C-terminal carboxyl of ubiquitin and a cysteine residue in the ubiquitin-activating



**Fig. 1** A bioinformatics screen for lysine desert proteins. **A** Pie diagram showing the distribution of the subcellular localization of the top 200 identified human lysine desert proteins after Shannon entropy filtering. Of the cytosolic and nuclear proteins, gene ontology (GO) classifications link 34% of the proteins to functions within the ubiquitin-proteasome system (UPS). The insert shows that the number of proteins with known or predicted functions in the UPS is also enriched for the top 100 and top 200 lysine desert proteins as compared to the entire human proteome. **B** The intrinsic disorder according to the MobiDB database for the lysine desert regions in the top 200 lysine desert proteins. The same data are presented as individual points (each representing a lysine desert region), box plots and rain cloud plots. The mean  $\pm$  standard error of the mean of the two sets are  $0.29 \pm 0.02$  (non-UPS) and  $0.65 \pm 0.04$  (UPS). **C** Conservation of the listed lysine desert proteins in the indicated species. Lysine residues are marked as black squares. The intrinsically disordered regions based on MobiDB in the human proteins are shown as a blue bar. The domain organization according to the SMART database is marked

enzyme (E1), driven by ATP hydrolysis. Next, the activated ubiquitin is transferred to an ubiquitin-conjugating enzyme (E2). Finally, a substrate-specific ubiquitin ligase (E3) interacts with the ubiquitin-loaded E2 and the target protein and catalyzes the ligation of ubiquitin to a lysine residue in the target protein. Repeated cycles of this ubiquitylation cascade lead to the formation of poly-ubiquitin chains, which target proteins for degradation by the 26S proteasome [3] or work as a signal in other processes, including DNA repair and endocytosis [4]. For degradation, ubiquitylated target proteins are either directly recognized and bound by the 26S proteasome via its ubiquitin-binding subunits or, alternatively, bound by UBL-UBA shuttle proteins. Through C-terminal ubiquitin-associated (UBA) domains [5], the UBL-UBA proteins interact with the ubiquitylated targets, while the N-terminal ubiquitin-like (UBL) domains interact with the proteasome [6–8]. The most prominent proteasomal shuttle proteins are yeast Rad23 and Dsk2, and their human orthologues RAD23AB and UBQLN1–4 [9], respectively. Another group of substrate shuttles is the UBL-BAG domain proteins, which include the human BAG1 and BAG6 [10] proteins. Unlike the UBL-UBA shuttles, these do not bind directly to ubiquitin. Instead, they contain a C-terminal Bcl-2 associated athanogene (BAG) domain that links to chaperones, thus bringing the chaperone and its bound substrate within proximity of the proteasome for degradation [11–13].

Although lysine is a common amino acid (5.6% occurrence in the eukaryotic proteome [14]), a few proteins wholly or largely without lysine residues have been reported, including certain viral proteins [15] and toxins [16]. In addition, the yeast E3 ubiquitin-protein ligases San1 and Slx5/8 both contain long stretches devoid of lysine, so-called lysine deserts, by which they avoid auto-ubiquitylation [17, 18]. By bioinformatics, long lysine depleted regions have also been observed in other E3s [19]. However, it is unknown whether there is a general functional significance of lysine depletion.

It has been speculated [15] that the viral proteins and toxins have also evolved lysine deserts to avoid ubiquitylation and degradation. In case of viral proteins, this would reduce MHC-I-mediated antigen presentation and thus limit their detection by the immune system. In addition, the viral protein lysine deserts might also protect them from conjugation to other lysine-targeting modifications such as ISG15, which inhibits viral infection [20]. For toxins that enter the cytosol through retrograde transport [21], the lysine deserts may allow them to escape ER-associated degradation (ERAD) as shown for ricin [22], cholera [23] and pertussis [24] toxin, thus resulting in increased potency. Lastly, lysine desert proteins are enriched in actinobacteria and the phages that infect them [25]. This enrichment correlates with the pupylation system found only in this phylum, which, similar to the UPS, involves conjugation to lysine residues as a signal to facilitate degradation [26].

Using bioinformatics, we identify a wide range of proteins across different species that contain long stretches completely devoid of lysines. Moreover, we analyze lysine deserts in the human proteome and find that they are conserved and pervasive for several proteins with known or predicted roles in the ubiquitin-proteasome system (UPS). Of these, we tested the consequences of artificial lysine introduction for a selected group of proteins (RAD23A, UBQLN1, UBQLN2, BAG6, RNF115 and PSMF1) and found increased ubiquitylation for most of the mutant proteins as compared to their wild-type counterparts. In case of BAG6, we show that ubiquitylation is catalyzed by the E3 RNF126, while RAD23A is ubiquitylated by E6AP. Despite the elevated ubiquitylation, mutant RAD23A appears stable, but displays a partial loss-of-function phenotype in fission yeast. In case of UBQLN1 and BAG6, introducing lysine leads to a reduced abundance due to proteasomal degradation of the proteins.

## Results

### A bioinformatics screen for lysine desert proteins

Although lysine desert proteins have been observed before [17–19] it remains unknown how widespread they are throughout proteomes. To this end, we used a bioinformatics approach to search the entire human proteome for proteins containing stretches without lysine residues and then sorted the proteins based on the length of the lysine-depleted region (the lysine desert) (Supplemental file 2, sheet 1). Since many of the highest scoring proteins would likely include proteins with simple repeated sequences lacking lysine, we next filtered for sequence complexity using the Shannon entropy of the amino acid distribution in the sequences [27] for the lysine desert regions (Supplemental file 2, sheet 2). Then,

we annotated the top 200 hits (based on the desert length in number of amino acids) for their subcellular localization as provided in UniProt (Supplemental file 2, sheet 3). Most of the lysine desert proteins were known or predicted to be localized to the plasma membrane (41%) or the extracellular matrix (24%), while the cytosol and nucleus accounted for only 15% and 16%, respectively (Fig. 1A). Remarkably, out of the cytosolic and nuclear lysine desert proteins, about one third of the lysine deserts were found in proteins with gene ontology (GO) annotations for biological process and molecular function connected with the ubiquitin-proteasome (UPS) system (Fig. 1A and Supplemental file 2). The abundance of UPS components among the lysine desert proteins is also evident without preselecting the cytosolic and nuclear proteins. Among the top 100 and the top 200 lysine desert proteins, 16% and 12%, respectively, are annotated as UPS components, corresponding to a 3.3- and 2.4-fold enrichment relative to the 4.9% of the human proteome (Fig. 1A and Supplemental file 2). This selection for UPS components was even more pronounced when we repeated our analyses for the proteomes of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, *Caenorhabditis elegans* and *Drosophila melanogaster* (Supplemental file 3). Thus, of the proteins with lysine deserts longer than 200 residues (after Shannon filtering for low-complexity regions) in the yeasts *S. cerevisiae* and *S. pombe*, we found by manual curation that 7 out of 14 (50%) and 17 out of 35 (48%), respectively, are UPS components (Supplemental file 3). The previously described budding yeast lysine desert proteins Slx5 and San1 [17, 18] were the highest scoring hits in *S. cerevisiae* (Supplemental file 3) [28, 29].

To be functionally relevant, a lysine desert likely must be exposed. Accordingly, the lysine deserts in the San1 and Slx5 E3s are localized in intrinsically disordered regions (IDRs) [17, 18]. Therefore, we next analyzed the top 200 human lysine desert proteins for intrinsic disorder within the lysine deserts, noting that lysines in general are modestly enriched in IDRs [28, 29]. Using the MobiDB database [30] annotation of intrinsic disorder, the average disorder for proteins in the human proteome was 26%, or  $0.26 \pm 0.23$  (mean  $\pm$  standard deviation). For the lysine desert regions in the non-UPS components in the top 200, this was not substantially different ( $0.29 \pm 0.28$ ). However, for the UPS components in the top 200 list, the average disorder of the desert regions was higher ( $0.64 \pm 0.18$ ) (Fig. 1B, C). Thus, while the individual non-UPS proteins may be more disordered than UPS proteins, the UPS proteins as a group on average show substantially more disorder. This shows that the lysine deserts in UPS components often, but not exclusively, overlap with regions of intrinsic disorder.

In agreement with our previous bioinformatics analyses [19], there are, among the UPS components in the top 200 lysine desert proteins (Table 1), several known or

**Table 1** Highest scoring human lysine desert UPS components

Name	UniProt ID	Desert length (aa)	Desert length (%)	Localiza-tion	Function
BAG6	P46379	950	83.9	Nucleus	Shuttle
AMBRA1	Q9C0C7	693	53.4	Other	E3 co-factor
RNF111	Q6ZNA4	663	66.7	Nucleus	E3
RNF6	Q9Y252	527	76.9	Cytosol	E3
UBQLN2	Q9UHD9	521	83.5	Cytosol	Shuttle
UBQLN4	Q9NRR5	514	85.5	Cytosol	Shuttle
UBQLN1	Q9UMX0	482	81.8	Cytosol	Shuttle
RNF12	Q9NVW2	477	76.4	Nucleus	E3
HUWE1	Q7Z6Z7	439	10.0	Cytosol	E3
HECW2	Q9P2P5	393	25.0	Cytosol	E3
RNF38	Q9H0F5	371	72.0	Nucleus	E3
UBR5	Q95071	342	12.2	Nucleus	E3
MARCHF7	Q9H992	329	46.7	Cytosol	E3
RNF43	Q68DV7	324	41.4	Nucleus	E3
HDAC6	Q9UBN7	317	26.1	Cytosol	Deacetylase
UBQLN3	Q9H347	311	47.5	Cytosol	Shuttle
ZSWIM8	A7E2V4	296	16.1	Cytosol	E3 co-factor
ANKIB1	Q9P2G1	284	26.1	Cytosol	E3
UBR4	Q5T4S7	275	5.3	Cytosol	E3
HERP1	Q15011	274	70.1	Mem-brane	UBL protein
RNF26	Q9BY78	270	62.4	Cytosol	E3
RNF165	Q6ZSG1	268	77.5	Nucleus	E3
RFWD3	Q6PCD5	258	33.3	Nucleus	E3
HECD1	Q9ULT8	253	9.7	Cytosol	E3

predicted E3s, including AMBRA1, RNF111 and RNF6. We also noted several non-E3 UPS components, including UBQLN1-4 and BAG6. Importantly, since the employed bioinformatics screen selects purely on the length of the lysine desert, deserts located in short proteins are overlooked. Conversely, long lysine deserts in large proteins will be detected, but may only cover a smaller fraction of the protein sequence. For instance, the HECD1 E3 contains a 253-residue stretch devoid of lysine residues. However, because HECD1 is 2610 residues long, the desert only accounts for about 10% of the entire protein, but still covers most of the unstructured part of the protein. The supplemental material also contains information on the lysine desert coverage as a fraction of sequence length (Supplemental File 2, sheet 2, column F).

In the following, we analyzed selected UPS lysine desert proteins (RAD23A, UBQLN1, UBQLN2, BAG6, RNF115 and PSMF1) identified broadly within the top

500, but all with lysine deserts covering at least 45% of the entire protein.

### The selected lysine deserts are conserved between species

It is possible that regions with low lysine counts occur through random chance rather than as a result of evolutionary selection. To investigate further whether the regions with lysine depletion are likely to be biologically relevant, we compared the lysine content for selected proteins across a diverse range of orthologues (Fig. 1C). For these we observed that generally, the lysine deserts are conserved across species, suggesting there was selective pressure for this trait.

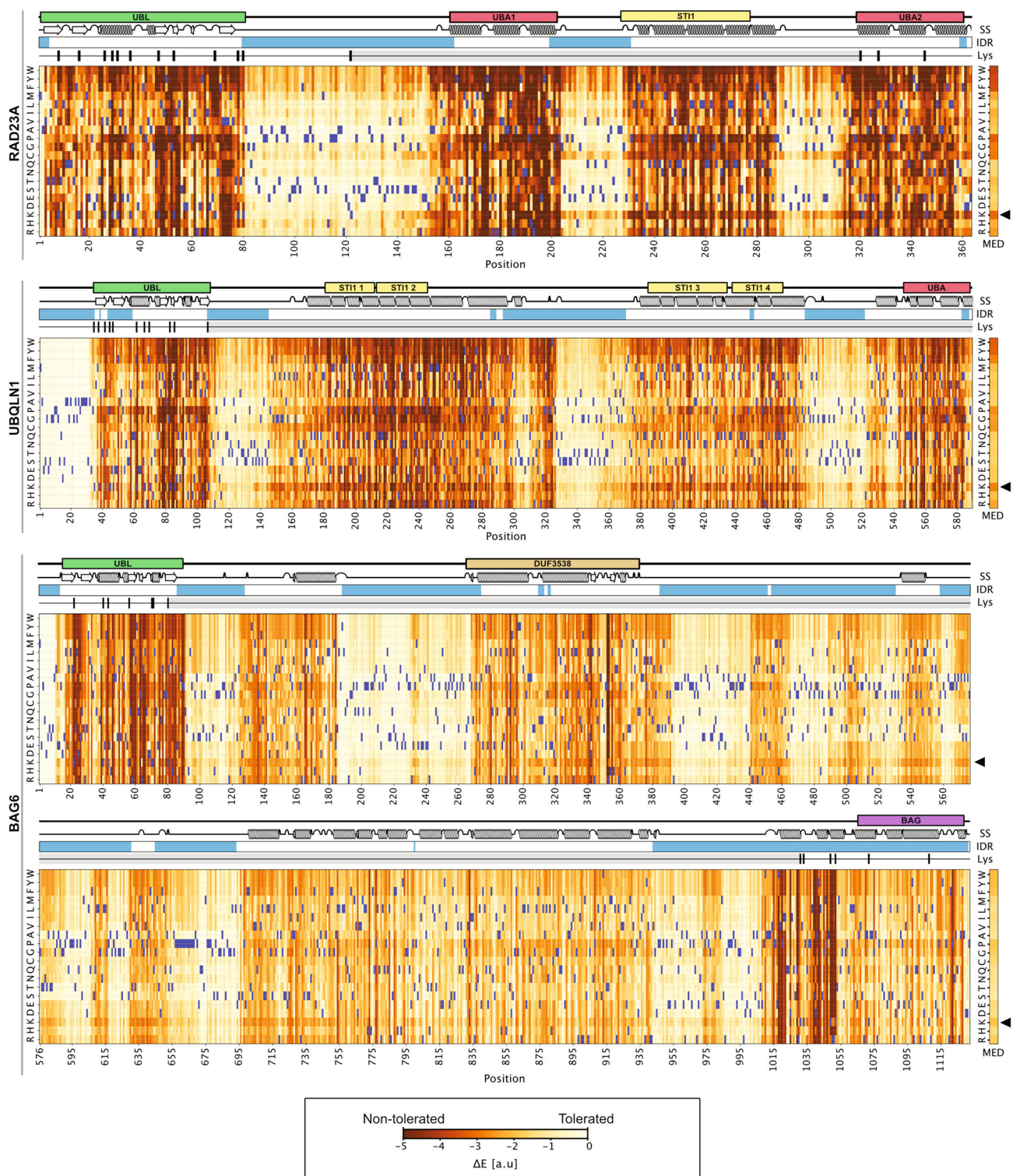
To evaluate a potential selection against introducing lysines into lysine deserts, we performed sequence analyses to predict the mutational landscapes of lysine desert proteins by modeling the evolutionary history of their sequences. Specifically, we used GEMME [31] (Global Epistatic Model for predicting Mutational Effects) and, as an input, multiple sequence alignments of RAD23A, UBQLN1 and BAG6 (with 1040, 623, and 169 entries, respectively) to calculate a score (available in supplemental file 4) for substituting the wild-type residue to all 19 other residue types. A score around zero suggests that a substitution would be acceptable (because it has been observed in this position and context during evolution), whereas more negative scores represent substitutions that the evolutionary model predicts to be detrimental. The evolutionary pressure to keep these regions without lysine was also evident from these analyses (Fig. 2). Thus, in general, these predictions indicate that substitutions to lysine are less tolerated in these proteins (Fig. 2). For RAD23A, UBQLN1 and BAG6, the GEMME scores indicated that substitutions to lysine were more unfavorable than substitutions to arginine, as evidenced by the fact that the GEMME scores for inserting lysines are on average more negative than when inserting an arginine at the same position (Supplemental file 1, Fig. S1). Similarly, we used GEMME to assess the effects of inserting lysines and arginines in the three desert proteins as compared to a set of ten non-desert proteins (proteins with short lysine desert regions). This revealed a small shift towards conservation in the mean value for substitution from any amino acid to lysine within lysine desert proteins when compared with nondesert proteins (Fig. S2). For comparison, an example heatmap of a nondesert UPS protein (SPOP) did not reveal substitutions to lysine as particularly detrimental (Fig. S3). Importantly, this shows that substitutions to lysine are not generally disfavored in eukaryotic proteins, whereas the observed selection against substitutions to bulky hydrophobic residues (W, V and Y) and proline (Figs. 2 and S3) is a common feature for most structured proteins [32]. However, we also note,

in particular in RAD23A and in UBQLN1, that cysteine is unfavorable at most positions (Fig. 2). Indeed, cysteine is rarely found in RAD23A and UBQLN1 orthologues, which we speculate is because cysteine is generally rare in IDRs [33]. Conversely, we note that lysine is common in IDRs [28, 29].

The human UBL-UBA shuttle RAD23A carries a 198-residue uninterrupted central lysine desert spanning the protein from the N-terminal UBL domain to the C-terminal UBA domain. UBQLN1 and UBQLN2 are even more dramatic, with their entire sequences downstream of the N-terminal UBL domain, corresponding to 482 and 521 residues, respectively, without any lysine. BAG6 is a massive protein of 1132 residues with a central uninterrupted desert of 950 residues without lysine. RNF115 is a 304-residue long RING-type E3 carrying a disordered N-terminal region with a desert of 173 residues. Finally, PSMF1, also known as PI31, is a 20S proteasome inhibitor [34] in which the entire disordered C-terminal region of 123 residues is a lysine desert. Importantly, in these cases, the regions are also depleted for lysine in their human paralogues (RAD23B, UBQLN3 and UBQLN4), despite an otherwise high sequence variability in the disordered regions. Finally, because these lysine deserts are not depleted of arginine (Fig. 3A), the lack of lysine is unlikely to be a consequence of an evolutionary pressure to maintain a specific overall net charge. This implies that the lysine deserts are rather a consequence of avoiding a functionality specific to the lysine side chain, which is undesirable in these regions.

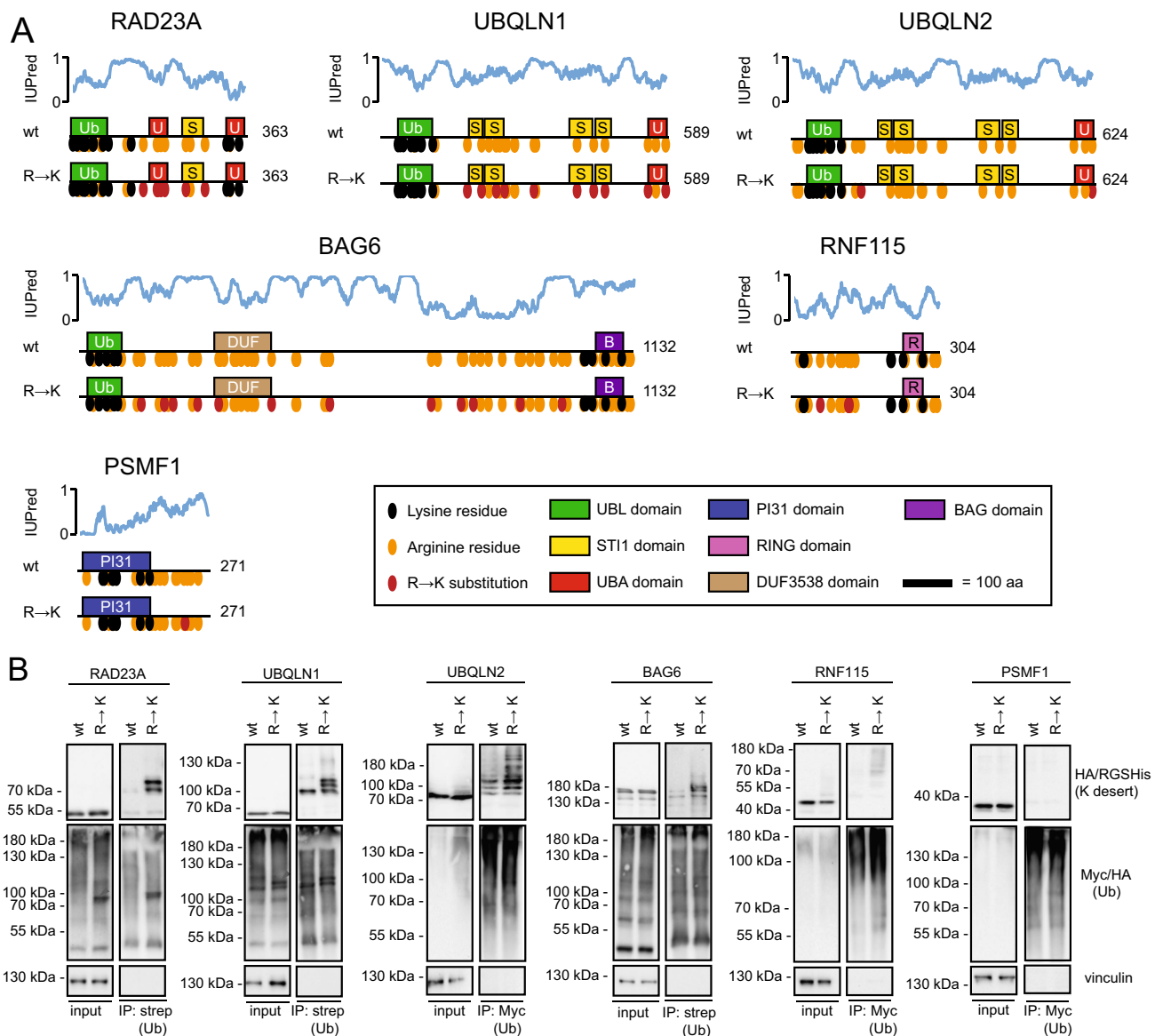
### Introduction of lysine residues into lysine deserts leads to increased ubiquitylation

One function of lysine residues is to accept covalent ubiquitin modifications. To test if the introduction of lysine residues into lysine deserts leads to altered ubiquitylation of the corresponding mutant protein, we transiently co-transfected human U2OS cells to express tagged ubiquitin (HA-strep- or myc-tagged) and one of six selected proteins, either wild type or a version where one or multiple arginine residues had been substituted for lysine (R → K) (Fig. 3A). The proteasome inhibitor bortezomib (BZ) was added to the cells 16 h prior to harvest to ensure that ubiquitylated proteins were not degraded. To purify only proteins covalently conjugated to ubiquitin, and not ubiquitin-interacting proteins, the cell lysates were first denatured by incubation at 100 °C in SDS. Then, SDS was diluted with Triton X-100 prior to precipitation of the tagged ubiquitin. We detected an increased ubiquitylation of the R → K variants of RAD23A, UBQLN1, UBQLN2, BAG6 and RNF115 as compared to their wild-type counterparts, but not of PSMF1 (Fig. 3B). Furthermore, in the precipitates of the K-variants, the ubiquitylated proteins all had



**Fig. 2** Evolutionary conservation of the lysine deserts. Evolutionary conservation analysis using multiple sequence alignments of human RAD23A, UBQLN1 and BAG6 presented as heat maps. Scores ( $\Delta E$ ) close to zero (white and light yellow colors) indicate that a given amino acid substitution is compatible with the alignments, while negative scores (red and dark yellow colors) indicate that the substitution is incompatible with the alignments and therefore likely detrimental to the protein structure and/or function. The wild-type amino acid residue at each position is

shown in blue. The domain organization, intrinsically disordered regions (IDR) (blue) and structural elements are shown above. The positions of lysine residues (black bars) and the extent of the lysine deserts (grey bars) are indicated. The median score (MED) for substitutions to the indicated amino acid residues across the entire protein is shown to the right. Note that substitutions to lysine (arrowhead) in general appear detrimental



**Fig. 3** The introduction of lysine residues leads to increased ubiquitylation of UPS components. **A** Schematic illustration of the indicated constructs, showing the positions of domains to scale, endogenous arginine (orange) and lysine (black) residues and introduced R→K substitutions (red). The prediction of intrinsically unstructured regions was performed with the IUPred web server (<https://iupred.elte.hu/>) [78]. **B** U2OS cells were transiently co-transfected

with either HA-strep- or Myc-tagged ubiquitin and the indicated constructs. After 24 h, cells were treated with 10  $\mu$ M BZ for 16 h and then used for denaturing immunoprecipitation (IP) of either strep or Myc. Protein was visualized by western blotting and vinculin was used as a loading control. Note that the ubiquitylation of RAD23A and UBQLN1 R→K variants was evident already in the input samples

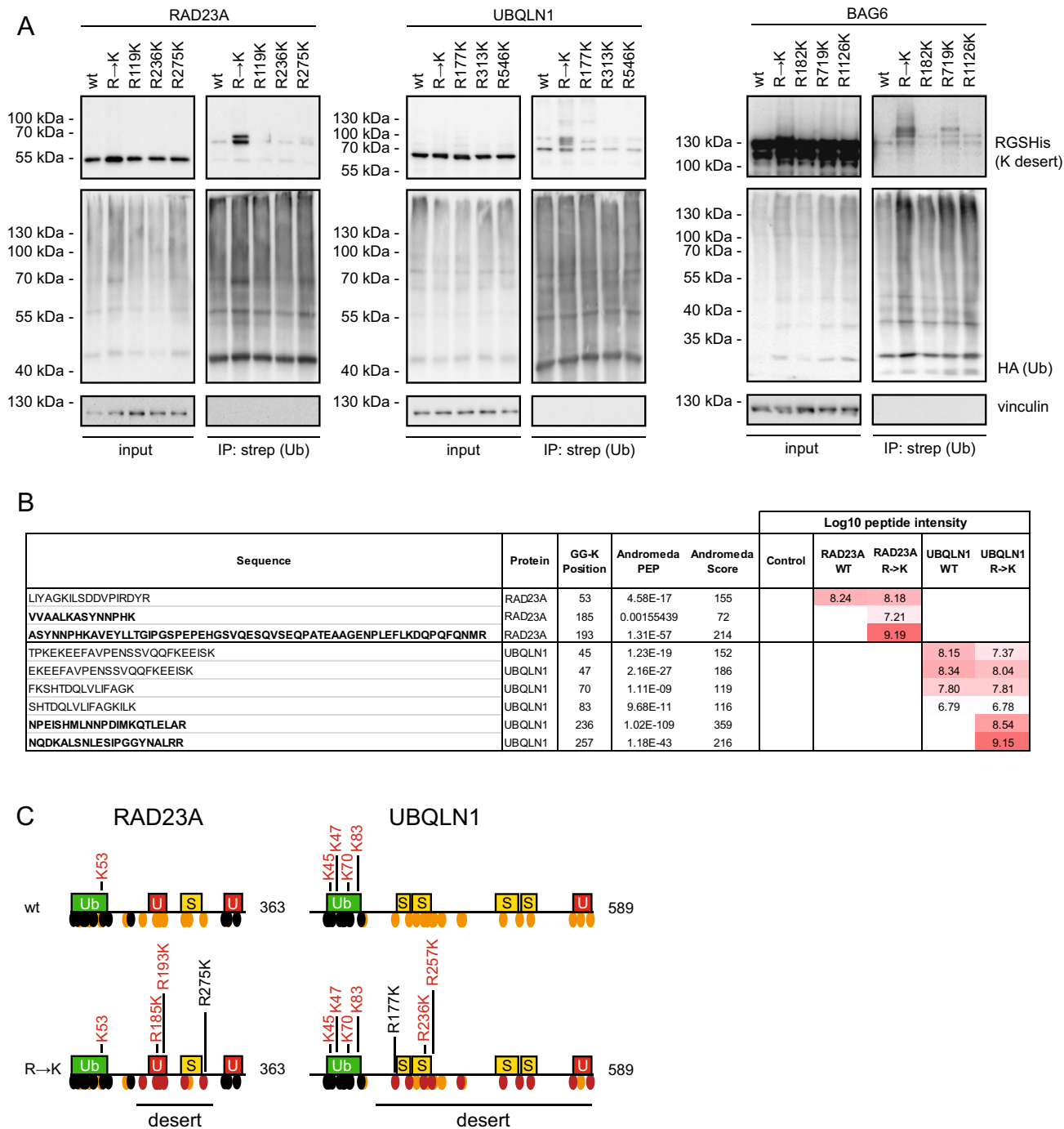
a higher molecular weight, supporting that the bands represent covalent conjugates. These results were specifically dependent on the introduction of lysine, as control experiments, where the arginine residues were replaced with glutamine (R→Q variants), appeared similar to wild type (Fig. S4). This indicates that the increased ubiquitylation of the K-variants is not triggered by misfolding induced by removing arginine, but is rather a consequence of specific ubiquitylation at the artificially introduced lysine residues.

### Ubiquitylation of introduced lysine residues is dependent on the lysine position

To test if the increased ubiquitylation is dependent on the position of the introduced lysines, we repeated the experiments with RAD23A, UBQLN1 and BAG6 variants carrying only single R→K substitutions. As expected, the effects were less dramatic. Some of the single mutant variants did not seem to affect ubiquitylation as compared to

the wild-type protein. However, at other positions, e.g. RAD23A R275K, UBQLN1 R177K and BAG6 R719K, increased ubiquitylation was observed (Fig. 4A). In parallel,

we analyzed both the wild type and the variants carrying multiple R → K substitutions in RAD23A and UBQLN1 by mass spectrometry. We observed that one and four of



**Fig. 4** Ubiquitylation is dependent on the position of the introduced lysine. **A** U2OS cells were transiently co-transfected with HA-strep-ubiquitin and the indicated constructs. After 24 h, cells were treated with 10 μM BZ for 16 h and then used for denaturing immunoprecipitation using StrepTactin beads, followed by Western blotting. Vinculin served as a loading control. **B** The wild-type and R → K versions of Myc-tagged RAD23A and UBQLN1 were immunoprecipitated

with Myc-trap beads, and peptides were analyzed by mass spectrometry for ubiquitylation. The Log<sub>10</sub> peptide intensities of the identified RAD23A and UBQLN1 peptides are listed. **C** Schematic illustration of RAD23A and UBQLN1, indicating the positions of modified lysine residues identified by mass spectrometry (red lettering) and by Western blotting (black lettering). Domain organizations as in Fig. 3A



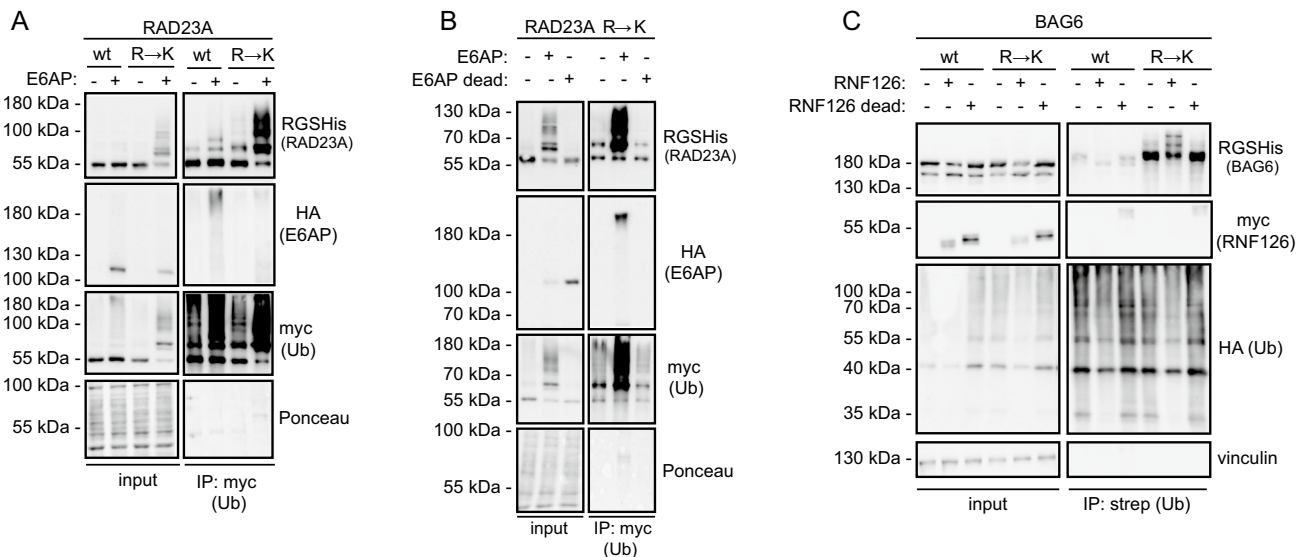
the naturally occurring lysine residues for RAD23A and UBQLN1, respectively, were ubiquitylated. In all cases, these are located within the UBL domain of both proteins (Fig. 4B, C) (Fig. S5). In the case of the R → K variants, we identified two additional ubiquitylation sites at the introduced lysine residues. For RAD23A, these were on K185 and K193, both located in the first UBA domain. UBQLN1 was ubiquitylated on K236 and K257, which are located in disordered regions between the structured domains (Fig. 4B, C). Together with the precipitation and Western blotting results (Fig. 4A), these findings suggest that some, but importantly not all, of the artificially introduced lysine residues are susceptible to ubiquitylation. Therefore, it appears that both the position and the number of substitutions to lysine are important for the extent of ubiquitylation.

### The E6AP and RNF126 E3s mediate ubiquitylation of lysines introduced in lysine deserts

Next, we sought to identify E3 enzymes that are involved in catalyzing the ubiquitylation of the UBL domain substrate shuttles. Since human RAD23A was previously shown to be

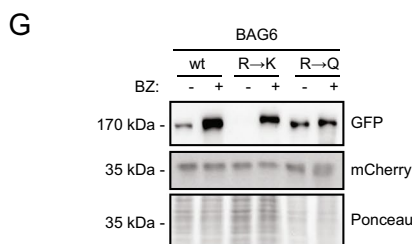
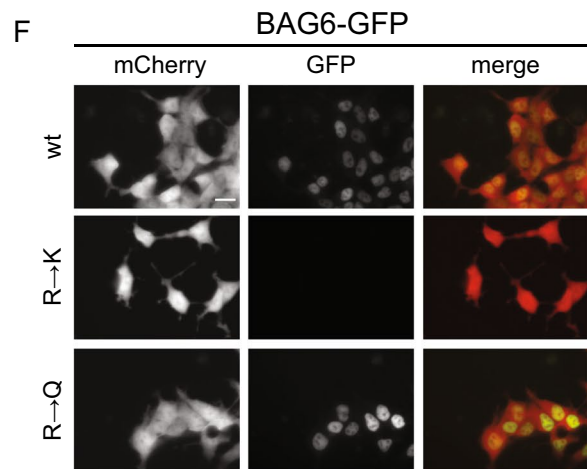
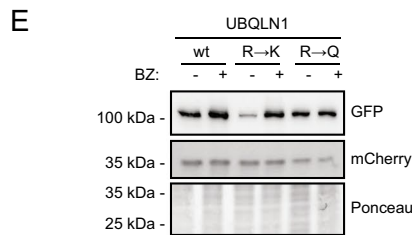
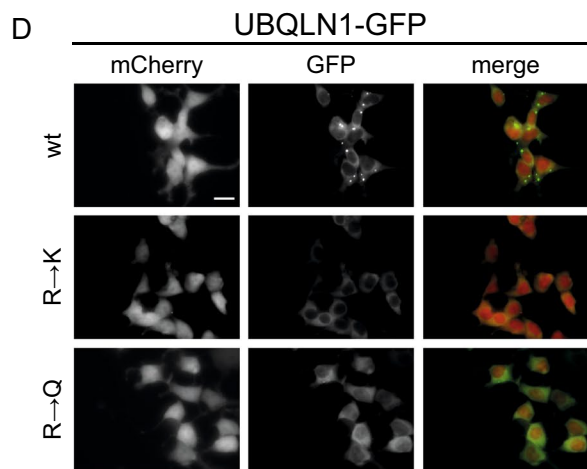
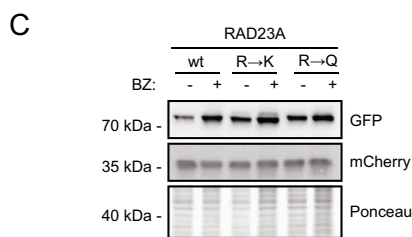
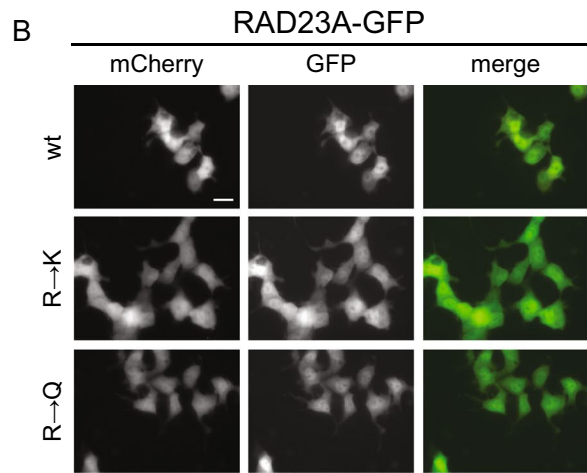
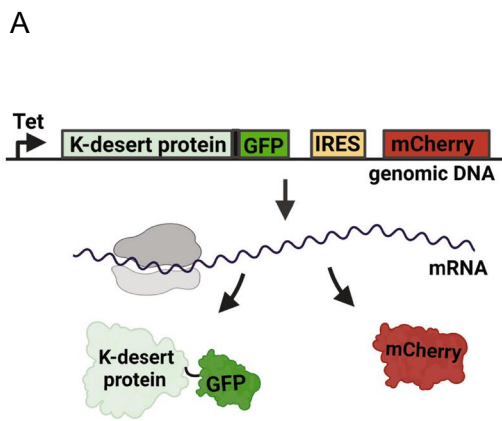
ubiquitylated by the E3 E6AP [35] (UBE3A), we tested if ubiquitylation of the lysine version of RAD23A was affected upon overexpression of E6AP. Indeed, RAD23A ubiquitylation was increased when E6AP was overexpressed, which was especially pronounced for the lysine version (Fig. 5A). This effect was specific for active E6AP since we did not detect any effect on RAD23A ubiquitylation when a catalytically inactive variant (C843A) of E6AP was used (Fig. 5B). This suggests that the lysine desert in RAD23A may have evolved to protect the protein from spurious ubiquitylation by proximal E6AP and possibly other E3s.

Because human BAG6 has been reported to interact with the E3 enzyme RNF126 via its UBL domain [36], we tested if RNF126 overexpression would affect BAG6 ubiquitylation. As a control, we included a catalytically inactive variant (C229/232A) of RNF126. The overexpression of wild-type RNF126, but not catalytically inactive RNF126, led to increased ubiquitylation of specifically the lysine version of BAG6 (Fig. 5C). We conclude that RNF126 is capable of catalyzing ubiquitylation of BAG6, and BAG6 may therefore have evolved the lysine desert to avoid RNF126 catalyzed ubiquitylation.



**Fig. 5** Distinct E3s catalyze the ubiquitylation of RAD23A and BAG6. **A** U2OS cells were transiently co-transfected with strep-Myc-ubiquitin, either wild-type RAD23A or the R → K variant and wild-type E6AP or empty vector. After 2 days, cells were harvested and used for denaturing immunoprecipitation with Myc-trap beads, followed by Western blotting. Staining of the membrane with Ponceau S served as a loading control. The cells were not incubated with BZ. **B** U2OS cells were transiently co-transfected with strep-Myc-ubiquitin, the RAD23A R → K variant and either wild-type E6AP, a catalytically inactive (C843A) E6AP mutant or empty vector. After

2 days, cells were harvested and used for denaturing immunoprecipitation with Myc-trap beads, followed by Western blotting. Staining of the membrane with Ponceau S served as a loading control. The cells were not incubated with BZ. **C** U2OS cells were transiently co-transfected with HA-strep-ubiquitin, either wild-type BAG6 or the R → K variant and with either wild-type RNF126, a catalytically inactive (C229/232A) RNF126 mutant or the empty vector. After 24 h, cells were treated with 10 μM BZ for 16 h and then used for denaturing immunoprecipitation with StrepTactin resin, followed by Western blotting. Vinculin served as a loading control



**Fig. 6** The UBQLN1 and BAG6 lysine deserts protect the proteins from proteasomal degradation. **A** The employed expression system relies on Bxb1-catalyzed site-specific plasmid integration at a “landing pad” in the genome of HEK293T cells [38]. Once the plasmid is integrated, the lysine desert protein (K-desert protein) is produced from a doxycyclin-regulated promoter (Tet) as a GFP fusion. To identify transfected cells and to correct for cell-to-cell variations in expression, the constructs also produced mCherry from an internal ribosomal entry site (IRES) located downstream of the GFP-fusions. This panel was generated using BioRender. **B** Representative fluorescence micrographs of human HEK293T cells after integration of the indicated expression constructs for RAD23A wild-type (wt), R→K, and R→Q fused to GFP. Scale bar, 20 μm. **C** Western blotting of whole cells lysates of HEK293T cells stably transfected to express the indicated expression constructs for RAD23A wild-type (wt), R→K, and R→Q fused to GFP. The cells were treated with 10 μM BZ, or DMSO as a control, 16 h prior to their harvest. The blots were probed with antibodies to GFP (RAD23A) and as a control with antibodies to mCherry. Ponceau S staining of the membrane served as a loading control. **D** As panel (B), but with UBQLN1 expression. **E** As panel (C), but with UBQLN1 expression. **F** As panel (B), but with BAG6 expression. **G** As panel (C), but with BAG6 expression

### Introduction of lysine residues into lysine deserts leads to an increased proteasomal degradation of UBQLN1 and BAG6, but not of RAD23A

The best-characterized role of ubiquitin is to target proteins for degradation by the proteasome. Because we noted that the introduction of lysine residues into the lysine deserts of UPS components leads to increased ubiquitylation, it is possible that this in turn leads to degradation. For more accurate determination of steady-state protein levels, we used site-specific genomic integration [37, 38] to generate stable HEK293T cell lines expressing C-terminally GFP-tagged wild-type or mutant RAD23A, UBQLN1 or BAG6. To correct for cell-to-cell variations in expression, the expression constructs also expressed mCherry from an internal ribosomal entry site (IRES) placed downstream of the GFP fusions (Fig. 6A).

By fluorescence microscopy (Fig. 6B) and western blotting (Fig. 6C), the levels of RAD23A appeared largely unaffected by substituting arginine residues to either lysine (R→K) or glutamine (R→Q). This indicates that in case of RAD23A, ubiquitylation of the lysine residues in the lysine desert does not lead to proteasomal degradation. However, for UBQLN1 (Fig. 6D, E) and BAG6 (Fig. 6F, G), the R→K variants, but not the R→Q variants, were less abundant and stabilized by the proteasome inhibitor bortezomib (BZ). Hence, for these proteins, ubiquitylation of lysine residues inserted in the lysine deserts leads to proteasomal degradation of the proteins. For BAG6, we noted that the level of wild-type protein was also increased upon treating with bortezomib (Fig. 6G), indicating that BAG6 is naturally unstable but further destabilized by the introduction of lysine residues. Quantification of the fluorescent signals by flow cytometry (Fig. S6) confirmed the results based on the

blotting and microscopy (Fig. 6), but also revealed slightly increased levels of the RAD23A variants as compared to wild type, as well as slightly increased levels of the R→Q variants of UBQLN1 and BAG6 as compared to wild type (Fig. S6).

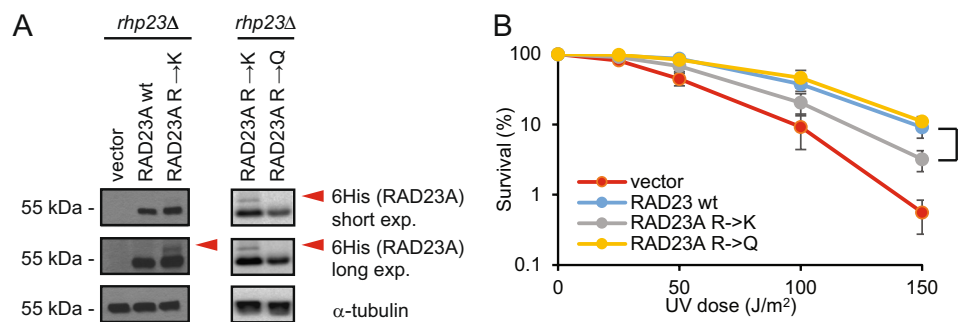
By fluorescence microscopy, the subcellular localization of RAD23A appeared unaffected by substituting arginines for lysine or glutamine (Fig. 6B). However, the UBQLN1 R→Q variant, although stable (Fig. 6E), appeared evenly distributed in the cytosol (Fig. 6D), whereas wild-type UBQLN1 formed cytosolic speckles (Fig. 6D), which may indicate condensate formation [39] similar to that reported for UBQLN2 [40]. BAG6 localized primarily to the nucleus, and its localization was not affected by the R→Q substitutions (Fig. 6F).

In conclusion, these data suggest that the lysine deserts protect UBQLN1 and BAG6 from proteasomal degradation, while the lysine desert in RAD23A does not appear to share this function. Moreover, the arginines in the UBQLN1 lysine desert are critical for UBQLN1 localization in cytosolic speckles, in line with the observation that arginines may help drive formation of cellular condensates [41].

### The RAD23A lysine desert is required for optimal function in *S. pombe*

Because the lysine desert in RAD23A seems to protect the protein from ubiquitylation but does not affect its cellular stability, we reasoned that the ubiquitylation of the lysine desert might lead to impaired protein function. To test this prediction, we used the fission yeast *Schizosaccharomyces pombe*, where the deletion of the *RAD23A* orthologue, *rhp23*, results in a UV-sensitive phenotype [42–47].

RAD23A and Rhp23 are conserved and both possess an extended lysine desert stretching between the UBL domain and the C-terminal UBA domain (Fig. S7A). When overexpressing RAD23A or its K and Q variants in *rhp23Δ* cells, we observed that the K variant, specifically, appeared to be modified in western blots of whole-cell lysates (Fig. 7A). As expected, the *rhp23Δ* strain exhibits an enhanced sensitivity towards UV radiation, which was complemented by both the RAD23A wild type and the R→Q variant (Fig. 7B). In comparison, the RAD23A R→K cells appeared more sensitive to UV irradiation than the cells carrying the wild type or R→Q variant, suggesting that the RAD23A R→K variant is partially compromised in function. However, we cannot exclude that this partial complementation is not due to indirect effects, e.g. on protein folding, structure or subcellular localization. In conclusion, the RAD23A lysine desert appears to be required to avoid ubiquitylation and for optimal RAD23A function in DNA repair, but is not required to protect from proteasomal degradation.



**Fig. 7** The RAD23A lysine desert is required for optimal function. **A** Western blot of RAD23A protein levels in an *rhp23Δ* *S. pombe* strain transformed with the indicated RAD23A fusion constructs. Blotting for  $\alpha$ -tubulin served as a loading control. Note that the RAD23A

R → K variant is modified (red arrows). **B** The survival after UV radiation of exponential phase *rhp23Δ* yeast cells transformed with the indicated constructs. The error bars indicate the standard deviation ( $n=3$ ). \* $p < 0.05$  (ANOVA)

## Discussion

Lysine deserts constitute regions in many hundreds of human proteins and may have evolved for various functions, including the evasion of lysine-linked PTMs or to maintain a specific protein structure. In this study, however, we describe the prevalence of lysine deserts as a conserved evolutionary mechanism to evade adventitious ubiquitylation. We focus on UPS components, reasoning that they are likely to become inadvertently modified, which might be particularly relevant for IDRs, where a lysine residue would be easily accessible. Accordingly, we note that it was recently reported that lysine deserts are also widespread in prokaryotes that harbor the ubiquitin-related pupylation system [25].

Even though lysine exhibits a relatively high propensity for  $\alpha$ -helices [48, 49], it is rated as a disorder-promoting amino acid, and IDRs have been shown to be enriched for lysines [29]. The fact that lysine is otherwise frequent in IDRs makes the existence of lysine deserts in IDRs even more remarkable and may represent a favorable trade-off, especially in the case of certain E3 enzymes. As shown for San1 [18], the IDRs in this enzyme allows for high conformational flexibility and direct substrate interaction, while the lysine deserts in the IDRs prevent its auto-ubiquitylation and degradation. We observed increased ubiquitylation when lysine was introduced into lysine deserts, but in the case of RAD23A, this did not lead to enhanced degradation. A possible reason, in agreement with the previous observations on this protein [50, 51], might be that RAD23A can escape proteasomal degradation since it lacks accessible disordered regions at its termini which are needed for the proteasome to engage its substrates. Hence, the lysine desert in RAD23A appears not to be essential to protect from degradation, but is still necessary for function. One possible explanation may be that ubiquitylation of lysine residues in the UBL domain of RAD23A seems to be required for function, but without affecting its degradation [52, 53]. Modification at other

lysine residues on the other hand may interfere with downstream signaling or act as a competitive acceptor for incoming ubiquitin. Thus, in some cases, perhaps lysine deserts have evolved to ensure focused ubiquitylation of a target protein at specific positions.

Although the lysine desert proteins contain long stretches without lysine, the position of an introduced lysine residue still appears to be critical. This dependence on position is likely a consequence of how exposed the particular position is and its proximity to the ubiquitylation machinery. Although we find that the lysine deserts tend to overlap with disordered regions, importantly even such areas of a protein may still form stable structures when interacting with binding partners. Finally, it is also possible that ubiquitylation occurs on any introduced lysine residue, albeit for some positions at a level below the detection limit.

We found that the overexpression of the E3 E6AP leads to increased ubiquitylation of RAD23A, especially when lysines were introduced into the lysine desert. Although no E6AP orthologue is found in yeast, we found that RAD23A is still modified when expressed in *S. pombe*, showing that at least one additional E3 may target the RAD23A desert. As the budding yeast homologs of the two UBL-UBA shuttles, RAD23A and UBQLN1 (Rad23 and Dsk2) were shown to bind via their UBL domains to the E3 ubiquitin ligase Ufd2 [54], we speculate that perhaps Ufd2 and its human orthologue UBE4B, also contribute to the observed ubiquitylation. For BAG6 we identify its binding partner, RNF126, as an E3 which can ubiquitylate lysines introduced into its lysine desert. This indicates that RNF126 must be relatively broad in its substrate selection, and accordingly, we note that RNF126 itself contains a lysine desert in its disordered N-terminal region. However, also in case of BAG6, we cannot exclude that other cellular E3s contribute to its ubiquitylation.

Aside from PTMs and despite their similar overall charge, arginine and lysine have distinct properties [55], and arginine

and lysine residues have for example been reported to display dramatically different contributions to phase separation of disordered proteins [56]. Hence, although both lysine-rich and arginine-rich sequences can form condensates, arginine to lysine substitutions have been reported to result in a decreased propensity to phase separate [41, 56–58] and arginine-rich condensates are more viscous [59]. UBQLN2, and likely its paralogues, form condensates of functional importance [40, 60, 61]. Accordingly, we observed that overexpressed GFP-tagged UBQLN1 localizes in cytosolic speckles. The arginine residues in the UBQLN1 lysine desert appear to be critical for speckle formation since this ability is lost upon substitutions to glutamine. Although we did not observe speckles for the UBQLN1 lysine variant, this difference in localization may be due to the increased degradation, leading to the UBQLN1 abundance being below a critical threshold for speckle formation. In addition, the lysine residues may alter the viscoelastic properties of the condensates, impacting the kinetics of their formation.

A limitation of our approach is that we cannot exclude that arginine to lysine substitutions do not affect the folding of lysine desert proteins, thereby making them protein quality control targets, which in turn may lead to their degradation. However, we note that the variants with substitutions to glutamine, which we predict in context of a folded protein should be more severe, are stable. Moreover, by microscopy, we never observed aggregation which could suggest that the proteins are misfolded.

Although we focus on the prominent group of UPS lysine deserts, we noticed that, in general, most proteins harboring pronounced lysine deserts were extracellular or plasma membrane proteins. Since the UPS is exclusively cytosolic and nuclear, there is likely a different underlying mechanism explaining why nature selected against lysine in these extracellular proteins. However, even for this group, it is still tempting to speculate that the avoidance of lysine-linked modifications is a major driving force. For instance, among these proteins, we noticed the extracellular lysyl oxidase LOXL1, which catalyzes the crosslinking of extracellular matrix proteins via substrate lysine residues [62]. LOXL1 possesses an extended lysine desert stretching the first two-thirds of the protein, which also overlaps with a predicted IDR. It is possible that the LOXL1 lysine desert has evolved to prevent itself from crosslinking nonspecifically to extracellular matrix (ECM) components, which would likely limit its diffusion in the ECM.

The current work has generated a long list of lysine desert proteins, whose functional roles remain to be deciphered. The list should be considered an exploratory tool, which presents the longest lysine deserts, and annotates them with additional information we believe might be useful for further downstream analysis. Some proteins will have long lysine-free stretches by chance, and any hypotheses must therefore

be carefully verified experimentally before conclusions can be drawn. That said, certain proteins were found to have strikingly long lysine-free stretches. We believe that these warrant further investigations and we hope the list will aid future studies to reveal more about the functions of the lysine-depleted regions. Along this line of thought, it would be interesting to investigate if also other types of deserts, e.g. serine or threonine deserts exist, and if so, if these have evolved to avoid undue modifications targeting these residues. Moreover, we note that several of the UPS lysine deserts, including RAD23A and UBQLN1, are also depleted for cysteine (Fig. S7AB). As cysteine is generally rare in IDRs [33], this is likely connected with the disordered nature of the lysine deserts. Next to the observations that some proteins are targeted for ubiquitin-independent proteasomal degradation [63] there are, however, several reports showing that ubiquitylation can also occur on the N-terminus and on serine, threonine or cysteine residues [64–72]. Recently, Szulc et al. reported nonlysine ubiquitylation of the lysine desert E3s VHL and SOCS [25], suggesting that some lysine deserts may have evolved as a means to ensure nonlysine ubiquitylation. The concomitant depletion of cysteine residues might potentially provide an evolutionary mechanism to avoid ubiquitylation of cysteine residues or other types of cysteine modifications, including oxidation.

## Materials and methods

### Bioinformatics

Lysine deserts were identified using a simple regular expression ([^K]+) on proteome files downloaded from UniProt [73] (human: UP000005640, retrieved Jan 2021; *Saccharomyces cerevisiae*: UP000002311, retrieved Jan 2022; *Schizosaccharomyces pombe*: UP000002485, retrieved Jan 2020; *Caenorhabditis elegans*: UP000001940, retrieved Jan 2020; *Drosophila melanogaster*: UP000000803, retrieved Jan 2020; *Arabidopsis thaliana*: UP000006548, retrieved Jan 2020). The Shannon entropy was used to filter out low-complexity sequences (containing long repeats). It was calculated by considering the residues in a sequence as independent observations from an underlying categorical distribution (20 outcomes) and reported in bits. The filtering was based on a Shannon entropy cutoff of 3.7, which was selected as a reasonable trade-off that removed most repeat proteins without affecting the remaining results notably. Disorder annotations for the proteomes were extracted from the MobiDB database [30] (retrieved at 2021/08/26 and 2022/01/04 for human and yeast, respectively). We used the “prediction-disorder-th\_50” annotation, which marks residues as disordered if at least 50% of the underlying annotations are disordered for that position. Orthologues were

extracted using Ensembl [74], filtering for a specific set of diverse species, and aligned using the muscle multiple sequence alignment algorithm [75]. Domain annotations were retrieved from the SMART database [76]. Gene ontology values were retrieved from UniProt using the Python BioServices module (v 1.7.11) [77]. UPS components were selected based on the following GO terms: Gene ontology (biological process) GO:0010498 (proteasomal protein catabolic process), GO:0006511 (ubiquitin-dependent protein catabolic process), GO:0016567 (protein ubiquitination) and/or gene ontology (molecular function) GO:0031593 (polyubiquitin modification-dependent protein binding), GO:0070628 (proteasome binding), GO:0031625 (ubiquitin protein ligase binding), GO:1990381 (ubiquitin-specific protease binding), GO:0061630 (ubiquitin protein ligase activity), GO:0004842 (ubiquitin-protein transferase activity), GO:0034450 (ubiquitin-ubiquitin ligase activity), GO:0043130 (ubiquitin binding), GO:0031624 (ubiquitin conjugating enzyme binding), GO:1990756 (ubiquitin ligase-substrate adaptor activity). The disorder predictions presented in Fig. 3 were from IUPred2A [78].

Evaluation of evolutionary conservation levels were performed measuring evolutionary distance of single site variants from the wild-type sequences of RAD23A (UniProt ID: P54725), UBQLN1 (UniProt ID Q9UMX0) and BAG6 (UniProt ID P46379). We started with the generation of multiple sequence alignments (MSAs) with the HHblits suite [79, 80], using standard parameters and an E-value threshold of  $10^{-20}$ . The obtained MSAs were comprised of 1059 sequences for RAD23A, 1270 sequences for UBQLN1, and 470 sequences for BAG6. Two additional filters were applied to the MSAs: First, all the amino acid columns in the MSA not included in the query wild-type sequences were removed. Then all the sequences with more than 50% of gaps were removed. After these filters, 1040 (RAD23A), 623 (UBQLN1), and 169 (BAG6) sequences were included in the alignments. To generate conservation scores for all the 19 possible substitutions at each position, we used the GEMME package [31] with standard parameters. The same procedure was applied to evaluate the conservation level of ten non-desert proteins reported in Supplemental file 2. For each protein, we used GEMME to predict the effect of inserting a lysine and an arginine in the whole protein. GEMME uses an evolutionary model to assess the functional effects of a given substitution; the score is similar to a standard conservation score, but also takes phylogeny and co-evolution across amino acid pairs into account. We calculated GEMME scores at each amino acid position both for substitutions to lysine and to arginine, excluding from the average wild-type lysine and arginine positions and their first neighbors. Lastly, we focus our analysis on the disorder regions of the target proteins using MobiDB prediction (selecting the “prediction-disorder-th\_50” score) to extract the selected

regions. The non-desert human proteins used were: HMGB3 (UniProt ID O15347), H1-O (UniProt ID: P07305), H1-2 (UniProt ID: P16403), SUB1 (UniProt ID: P53999), BASP1 (UniProt ID: P80723), C17orf64 (UniProt ID: Q86WR6), ERMN (UniProt ID: Q8TAM6), NHLH2 (UniProt ID: Q02577), YF016 (UniProt ID: A6NL46), KNOP1 (UniProt ID: Q1ED39). These ten proteins were selected from the bottom of Supplementary file 2 with the requirements to be at least 50 residues long and have an alignment with at least 100 homologs. Distribution plots of those two groups were generated using raincloud plots library [81] for Python3.

## Plasmids

The expression plasmids used are listed in the supplemental material (Supplemental file 1, Table S1).

## Cell culture and transfection

U2OS cells (ATCC) and HEK293T landing pad cells [38] were propagated in Dulbecco's Modified Eagle medium (DMEM), supplemented with 10% bovine serum (Sigma), 5000 IU/mL penicillin, 5 mg/mL streptomycin, 2 mM glutamine and, in the case of HEK293T landing pad cells, 2 µg/mL doxycycline (Dox) (Sigma-Aldrich, D9891), in a humidified atmosphere containing 5% CO<sub>2</sub> at 37 °C. Transient transfections were performed with FugeneHD (Promega) in reduced serum medium OptiMEM (Gibco). To generate stable transfectants in the HEK293T landing pad cells, 10<sup>6</sup> cells in 1 mL DMEM without doxycycline were seeded into 12-well plates. On the next day, the cells were transfected using 0.1 µg of the integrase vector pNLS-Bxb1-recombinase, mixed with 0.4 µg of the pVAMP recombination plasmid, 40 µL OptiMEM and 1.6 µL FugeneHD. Two days after transfection, the cells were treated with 10 nM of AP1903 (MedChemExpress) and 2 µg/mL doxycycline for two days to select for recombinant cells. After another two days of culturing in media with doxycyclin, but without AP1903, the cells were used directly for western blotting, microscopy and flow cytometry.

## Denaturing immunoprecipitation

A minimum of 10<sup>7</sup> U2OS cells were seeded into 10-cm dishes and transiently transfected on the following day using 4 µg of either HA-, HA-Strep- or Strep-Myc-ubiquitin along with 4 µg of the different target constructs, and 2 µg or 0.8 µg, respectively, for co-expression of E6AP or RNF126. One day after transfection, the cells were treated with 10 µM Bortezomib (LC Laboratories) in serum-free DMEM for 16 h. The cells were washed once with PBS and harvested in 300 µL lysis buffer A (30 mM Tris/HCl pH 8.1, 100 mM NaCl, 5 mM EDTA and freshly added 0.2 mM PMSF). On

ice, the samples were sonicated three times for 10 s. Subsequently, 75  $\mu$ L 8% SDS was added and the samples were boiled for 10 min with intermittent vortexing. Then 1125  $\mu$ L of lysis buffer A, supplemented with 2.5% Triton X-100, was added and samples incubated on ice for 30 min. Then, the samples were centrifuged at 16,000 g for 60 min at 4 °C. To prepare the agarose beads, 20  $\mu$ L of 50% slurry of either anti-HA resin (Sigma), StrepTactin (Qiagen) or Myc-Trap (ChromoTek) beads were washed twice in lysis buffer A. After centrifugation, 30  $\mu$ L cell extract was mixed with 12.5  $\mu$ L SDS sample buffer (94 mM Tris/HCl pH 6.8, 3% SDS, 40% glycerol, 0.0075% bromophenol blue, 0.0075% pyronin G) for input. The remaining extract was transferred to the tubes containing the washed beads. The samples were tumbled overnight at 4 °C and washed four times in 500  $\mu$ L lysis buffer A supplemented with 1% Triton X-100 and once with lysis buffer A (without detergent). After this last washing step, as much liquid as possible was removed and 40  $\mu$ L SDS sample buffer finally added to the beads. Both input and IP samples were boiled, and analyzed by SDS-PAGE and Western blotting.

### Electrophoresis and blotting

SDS-PAGE was performed with either 8% or 12.5% acrylamide gels. For the subsequent Western blotting, 0.2  $\mu$ m nitrocellulose membranes (Advantec, Toyo Roshi Kaisha Ltd.) were used. The membranes were blocked in PBS containing 5% fat-free milk powder, 0.1% Tween-20 and 5 mM Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> for at least 30 min. Incubation with the primary antibodies, all diluted in PBS containing 5% BSA and 0.1% Tween-20, was performed overnight at 4 °C. Primary antibodies used for this study were: anti-HA (Roche, 1:3000, 11867423001), anti-RGS-His (Qiagen, 1:2000, 34610), anti-Myc (ChromoTek, 1:1000, AB\_2631398), anti-Vinculin (Sigma, 1:2000, V9264), anti- $\alpha$ -tubulin (Merck, 1:1000, TAT-1 00020911), anti-GFP (ChromoTek, 1:1000, 3H9), anti-RFP (ChromoTek, 1:1000, 6G6). The secondary antibodies were: HRP-antimouse IgG (Dako, 1:5000, P0260), HRP-antirat IgG (Thermo Fisher Scientific, 1:5000, 31470). The blots were developed using Amersham ECL detection reagent (GE Healthcare) and a ChemiDoc Imaging System (BioRad).

### Identification of ubiquitylation sites in RAD23A and UBQLN1 by mass spectrometry

Both wild-type and the lysine variants of Myc-RGS6x-His-tagged RAD23A and UBQLN1 were purified from three confluent 10-cm dishes of transfected U2OS cells following the above protocol for denaturing immunoprecipitation using Myc-trap beads (ChromoTek). A control

purification from cells transfected with empty vector alone was performed in parallel. The protein was eluted from the beads with 20  $\mu$ L NuPAGE LDS sample buffer (Invitrogen) and separated by SDS-PAGE using Novex 4–20% Tris–Glycine gels (Invitrogen) and stained with Coomassie Brilliant Blue (CBB) (Sigma) (Fig. S8). Proteins from the regions of the gel relating to modified and unmodified forms were in-gel trypsin digested [82], alkylated with chloroacetamide and resultant peptides were resuspended in 0.1% TFA 0.5% acetic acid before analysis by LC–MS/MS. This was performed using a Q Exactive mass spectrometer (Thermo Scientific) coupled to an EASY-nLC 1000 liquid chromatography system (Thermo Scientific), using an EASY-Spray ion source (Thermo Scientific) running a 75  $\mu$ m  $\times$  500 mm EASY-Spray column at 45 °C. Two MS runs (of 60 and 150 min) were prepared using approximately 15% total peptide sample each. A top 3 data-dependent method was applied employing a full scan ( $m/z$  300–1800) with resolution  $R = 70,000$  at  $m/z$  200 (after accumulation to a target value of 1,000,000 ions with maximum injection time of 20 ms). The 3 most intense ions were fragmented by HCD and measured with a resolution of  $R = 70,000$  (60 min run) or 35,000 (150 min run) at  $m/z$  200 (target value of 1,000,000 ions and maximum injection time of 500 ms) and intensity threshold of  $2.1 \times 10^4$ . Peptide match was set to ‘preferred’. Ions were ignored if they had unassigned charge state 1, 8 or  $> 8$  and a 10 s (60 min run) or 25 s (150 min run) dynamic exclusion list was applied.

Data analysis used MaxQuant version 1.6.1.0 [83]. Default settings were used with a few exceptions. A database of the 4 transfected proteins was used as main search space (see below) with a first search using the whole human proteome (Uniprot 73,920 entries – April 2019). Digestion was set to Trypsin/P (ignoring lysines and arginines N-terminal to prolines) with a maximum of 3 missed cleavages. Match between runs was not enabled. Oxidation (M), Acetyl (Protein N-term) and GlyGly (K) were included as variable modifications, with a maximum of 4 per peptide allowed. Carbamidomethyl (C) was included as a fixed modification. Only peptides of maximum mass 8000 Da were considered. Protein and peptide level FDR was set to 1% but no FDR filtering was applied to identified sites. Manual MS/MS sequence validation was used to verify GlyGly (K) peptide identifications. Peptide intensity data were reported for each sample loaded on the original gel by pooling data from the two MS runs for only the upper (modified protein) gel slices of each lane. The protein sequences are listed in the supplemental material. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [84] partner repository (Project accession: PXD036108).

## Fluorescence microscopy

Stable transfected HEK293T landing pad cells, simultaneously expressing C-terminally tagged GFP-constructs and mCherry from an IRES (Fig. 6A), were grown in 6-well dishes. Images of the cells were recorded directly from the 6-well plates and with a Zeiss AxioVert.A1 microscope equipped with an AxioCam ICm 1 Rev. 1 FireWire (D) camera. EGFP was excited at 475 nm and mCherry at 590 nm. Image processing was performed using ImageJ (version 1.51j8).

## Flow cytometry

Immediately prior to flow cytometry, cells were washed by centrifugation in PBS and resuspended in 2% (v/v) bovine calf serum in PBS. Then the cells were filtered through 50 µm mesh filters into 5 mL tubes. A BD FACSJazz (BD Biosciences) instrument was used for flow cytometry. Data collection and analyses were performed using FlowJo (v10.8.1, BD), using the following gates: Live cells, singlet cells, BFP negative and mCherry positive.

## Yeast strains and techniques

The fission yeast *rhp23Δ* strain [85] was grown in standard rich media (YES media: 30 g/L glucose, 5 g/L yeast extract, 0.2 g/L adenine, 0.2 g/L uracil, 0.2 g/L leucine) at 30 °C. The cells were transformed with pREP1-based vectors using lithium acetate [86]. Transformants were cultured in Edinburgh Minimal Media 2 (EMM2) (Sunrise Science) without leucine for plasmid selection. For the survival assays, yeast cultures were grown at 30 °C to exponential phase and spread on solid media plates and irradiated with the indicated UV dosages under a calibrated UV lamp. Then, the plates were incubated for four days at 30 °C and colonies were counted. For western blotting, proteins were extracted as whole cell lysate using trichloroacetic acid (Sigma) and glass beads as described before [87].

## Statistical analyses

The mass spectrometry was performed once. All other presented experiments have been performed at least three times. For the cell survival assays, the error bars indicate the standard deviation ( $n=3$ ). ANOVA was performed using MS Excel.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00018-023-04782-z>.

**Acknowledgements** The authors thank Sofie V. Nielsen and Birthe B. Kragelund for the helpful discussions and comments on the manuscript,

and the REPIN consortium for discussions on IDPs. We thank Anne-Marie Lauridsen and Søren Lindemose for excellent technical assistance. We thank Vibe H. Østergaard and Michael Lisby for assistance with the flow cytometry. We thank Mads Gyrd-Hansen for the HA-strep ubiquitin expression construct. We acknowledge access to computing resources from the Biocomputing Core Facility at the Department of Biology, University of Copenhagen.

**Author contributions** CK, MGT, WB, NO, and MHT conducted the experiments. Data analyses by CK, MGT, WB, MHT, KH, MC and RHP. Experimental design by KL-L, KH, WB and RHP. KH and RHP conceived the study. CK, MGT and RHP wrote the paper

**Funding** Open access funding provided by Royal Danish Library. This work was supported by a Villum Fonden (<https://veluxfoundations.dk/>) research grant 40526 (to R.H.P.), the Novo Nordisk Foundation (<https://novonordiskfonden.dk>) challenge programmes PRISM (NNF18OC0033950; to R.H.P. and K.L.L.) and REPIN (NNF18OC0033926; to R.H.P.), NNF18OC0052441 (to R.H.P.), and the Novo Nordisk Foundation collaborative Data Science programme: Basic Machine Learning Research in Life Science, NNF20OC0062606 (to W.B.), and Danish Council for Independent Research (Natur og Univers, Det Frie Forskningsråd) (<https://dff.dk/>) <https://doi.org/10.46540/2032-00007B> (to R.H.P.). M.H.T was supported by Cancer Research UK (CRUK C434/A21747) and Wellcome (217196/Z/19/Z). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Availability of data and material** The mass spectrometry datasets generated in the current study are available from the ProteomeXchange Consortium: Project accession: PXD036108. Other data and material for this study are available upon request to the corresponding authors.

## Declarations

**Conflict of interest** The authors have no relevant financial or nonfinancial interests to disclose.

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Wang ZA, Cole PA (2020) The chemical biology of reversible lysine post-translational modifications. *Cell Chem Biol* 27:953–969
2. Swatek KN, Komander D (2016) Ubiquitin modifications. *Cell Res* 26:399–422



3. Collins GA, Goldberg AL (2017) The logic of the 26S proteasome. *Cell* 169:792–806
4. Kulathu Y, Komander D (2012) Atypical ubiquitylation—the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nat Rev Mol Cell Biol* 13:508–523
5. Hofmann K, Bucher P (1996) The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem Sci* 21:172–173
6. Hiyama H et al (1999) Interaction of hHR23 with S5a The ubiquitin-like domain of hHR23 mediates interaction with S5a subunit of 26 S proteasome. *J Biol Chem* 274:28019–28025
7. Wilkinson CRM et al (2001) Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nat Cell Biol* 3:939–943
8. Elsasser S et al (2002) Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nat Cell Biol* 4:725–730
9. Su V, Lau AF (2009) Ubiquitin-like and ubiquitin-associated domain proteins: Significance in proteasomal degradation. *Cell Mol Life Sci* 66:2819–2833
10. Corduan A, Lecomte S, Martin C, Michel D, Desmots F (2009) Sequential interplay between BAG6 and HSP70 upon heat shock. *Cell Mol Life Sci* 66:1998–2004
11. Höhfeld J, Jentsch S (1997) GrpE-like regulation of the Hsc70 chaperone by the anti-apoptotic protein BAG-1. *EMBO J* 16:6209–6216
12. Kabbage M, Dickman MB (2008) The BAG proteins: a ubiquitous family of chaperone regulators. *Cell Mol Life Sci* 65:1390–1402
13. Takayama S et al (1997) BAG-1 modulates the chaperone activity of Hsp70/Hsc70. *EMBO J* 16:4887–4896
14. Kozłowski LP (2017) Proteome-pI: Proteome isoelectric point database. *Nucleic Acids Res* 45:D1112–D1116
15. Randow F, Lehner PJ (2009) Viral avoidance and exploitation of the ubiquitin system. *Nat Cell Biol* 11:527–534
16. London E, Luongo CL (1989) Domain-specific bias in arginine/lysine usage by protein toxins. *Biochem Biophys Res Commun* 160:333–339
17. Sharma P et al (2017) A lysine desert protects a novel domain in the slx5–slx8 sumo targeted ub ligase to maintain sumoylation levels in *Saccharomyces cerevisiae*. *Genetics* 206:1807–1821
18. Fredrickson EK, Candadaí SVC, Tam CH, Gardner RG (2013) Means of self-preservation: How an intrinsically disordered ubiquitin-protein ligase averts self-destruction. *Mol Biol Cell* 24:1041–1052
19. Boomsma W, Nielsen SV, Lindorff-Larsen K, Hartmann-Petersen R, Ellgaard L (2016) Bioinformatics analysis identifies several intrinsically disordered human E3 ubiquitin-protein ligases. *PeerJ* 4:e1725
20. Durfee LA, Lyon N, Seo K, Huibregtse JM (2010) The ISG15 conjugation system broadly targets newly synthesized proteins: implications for the antiviral function of ISG15. *Mol Cell* 38:722–732
21. Nowakowska-Gołacka J, Sominka H, Sowa-Rogozińska N, Słomińska-Wojewódzka M (2019) Toxins utilize the endoplasmic reticulum-associated protein degradation pathway in their intoxication process. *Int J Mol Sci* 20:1307
22. Deeks ED et al (2002) The low lysine content of ricin a chain reduces the risk of proteolytic degradation after translocation from the endoplasmic reticulum to the cytosol. *Biochemistry* 41:3405–3413
23. Rodighiero C, Tsai B, Rapoport TA, Lencer WI (2002) Role of ubiquitination in retro-translocation of cholera toxin and escape of cytosolic degradation. *EMBO Rep* 3:1222–1227
24. Worthington ZEV, Carbonetti NH (2007) Evading the proteasome: absence of lysine residues contributes to pertussis toxin activity by evasion of proteasome degradation. *Infect Immun* 75:2946–2953
25. Szulc NA, Piechota M, Thapa P, Pokrzywa W (2023) Lysine-deficient proteome can be regulated through non-canonical ubiquitination and ubiquitin-independent proteasomal degradation. <https://doi.org/10.1101/2023.01.18.524605>.
26. Striebel F, Imkamp F, Özcelik D, Weber-Ban E (2014) Pupylation as a signal for proteasomal degradation in bacteria. *Biochim Biophys Acta Mol Cell Res* 1843:103–113
27. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
28. Roesgaard MA et al (2022) Deciphering the alphabet of disorder—Glu and Asp act differently on local but not global properties. *Biomolecules* 12:1426
29. Romero P et al (2001) Sequence complexity of disordered protein. *Proteins Struct Funct Genet* 42:38–48
30. Piovesan D et al (2021) MobiDB: intrinsically disordered proteins in 2021. *Nucleic Acids Res* 49:D361–D367
31. Laine E, Karami Y, Carbone A (2019) GEMME: A simple and fast global epistatic model predicting mutational effects. *Mol Biol Evol* 36:2604–2619
32. Høie MH, Cagiada M, Beck Frederiksen AH, Stein A, Lindorff-Larsen K (2022) Predicting and interpreting large-scale mutagenesis data using analyses of protein stability and conservation. *Cell Rep* 38:110207
33. Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527–533
34. Li X, Thompson D, Kumar B, DeMartino GN (2014) Molecular and cellular roles of PI31 (PSMF1) protein in regulation of proteasome function. *J Biol Chem* 289:17392–17405
35. Kumar S, Talis AL, Howley PM (1999) Identification of HHR23A as a substrate for E6-associated protein-mediated ubiquitination. *J Biol Chem* 274:18785–18792
36. Rodrigo-Brenni MC, Gutierrez E, Hegde RS (2014) Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol Cell* 55:227–237
37. Matreyek KA et al (2018) Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat Genet* 50:874–882
38. Matreyek KA, Stephany JJ, Chiasson MA, Hasle N, Fowler DM (2020) An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res* 48:1–12
39. Gerson JE et al (2021) Shared and divergent phase separation and aggregation properties of brain-expressed ubiquilins. *Sci Rep* 11:1–13
40. Dao TP et al (2018) Ubiquitin Modulates Liquid–Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Mol Cell* 69:965–978.e6
41. Wang J et al (2018) A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* 174:688–699.e16
42. Lombaerts M, Goeloe JI, den Dulk H, Brandsma JA, Brouwer J (2000) Identification and characterization of the rhp23+ DNA repair gene in *Schizosaccharomyces pombe*. *Biochem Biophys Res Commun* 268:210–215
43. Han TX, Xu XY, Zhang MJ, Peng X, Du LL (2010) Global fitness profiling of fission yeast deletion strains by barcode sequencing. *Genome Biol* 11:R60
44. Schaubert C et al (1998) Rad23 links DNA repair to the ubiquitin/proteasome pathway. *Nature* 391:715–718
45. Elder RT et al (2002) Involvement of rhp23, a *Schizosaccharomyces pombe* homolog of the human HHR23A and *Saccharomyces cerevisiae* RAD23 nucleotide excision repair genes, in cell cycle control and protein ubiquitination. *Nucleic Acids Res* 30:581–591
46. Miller RD, Prakash L, Prakash S (1982) Defective excision of pyrimidine dimers and interstrand DNA crosslinks in rad7 and rad23 mutants of *Saccharomyces cerevisiae*. *MGG Mol Gener Genet* 188:235–239

47. Masutani C et al (1994) Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *EMBO J* 13:1831–1843
48. Costantini S, Colonna G, Facchiano AM (2006) Amino acid propensities for secondary structures are influenced by the protein structural class. *Biochem Biophys Res Commun* 342:441–451
49. Nakashima H, Nishikawa K, Ooi T (1986) The folding type of a protein is relevant to the amino acid composition. *J Biochem* 99:153–162
50. Fishbain S, Prakash S, Herrig A, Elsasser S, Matouschek A (2011) Rad23 escapes degradation because it lacks a proteasome initiation region. *Nat Commun* 2:192
51. Yu H, Kago G, Yellman CM, Matouschek A (2016) Ubiquitin-like domains can target to the proteasome but proteolysis requires a disordered region. *EMBO J* 35:1522–1536
52. Gödderz D, Giovannucci TA, Laláková J, Menéndez-Benito V, Dantuma NP (2017) The deubiquitylating enzyme Ubp12 regulates Rad23-dependent proteasomal degradation. *J Cell Sci* 130:3336–3346
53. Sekiguchi T et al (2011) Ubiquitin chains in the Dsk2 UBL domain mediate Dsk2 stability and protein degradation in yeast. *Biochem Biophys Res Commun* 411:555–561
54. Hänzelmann P, Stingle J, Hofmann K, Schindelin H, Raasi S (2010) The yeast E4 ubiquitin ligase Ufd2 interacts with the ubiquitin-like domains of Rad23 and Dsk2 via a novel and distinct ubiquitin-like binding domain. *J Biol Chem* 285:20390–20398
55. Tesei G et al (2017) Self-association of a highly charged arginine-rich cell-penetrating peptide. *Proc Natl Acad Sci U S A* 114:11428–11433
56. Bremer A et al (2022) Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat Chem* 14:196–207
57. Nott TJ et al (2015) Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell* 57:936–947
58. Tsang B et al (2019) Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proc Natl Acad Sci U S A* 116:4218–4227
59. Fisher RS, Elbaum-Garfinkle S (2020) Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nat Commun* 11
60. Dao TP, Castañeda CA (2020) Ubiquitin-modulated phase separation of shuttle proteins: does condensate formation promote protein Degradation? *BioEssays* 42:2000036
61. Dao TP et al (2019) ALS-linked mutations affect UBQLN2 oligomerization and phase separation in a position- and amino acid-dependent manner. *Structure* 27:937–951.e5
62. Vallet SD, Ricard-Blum S (2019) Lysyl oxidases: From enzyme activity to extracellular matrix cross-links. *Essays Biochem* 63:349–364
63. Kravtsova-Ivantsiv Y, Ciechanover A (2012) Non-canonical ubiquitin-based signals for proteasomal degradation. *J Cell Sci* 125:539–548
64. Ben-Saadon R et al (2004) The tumor suppressor protein p16INK4a and the human papillomavirus oncoprotein-58 E7 are naturally occurring lysine-less proteins that are degraded by the ubiquitin system: direct evidence for ubiquitination at the N-terminal residue. *J Biol Chem* 279:41414–41421
65. Sánchez-Lanzas R, Castaño JG (2017) Lysine-less variants of spinal muscular atrophy SMN and SMNΔ7 proteins are degraded by the proteasome pathway. *Int J Mol Sci* 18:1–14
66. McClellan AJ, Laugesen SH, Ellgaard L (2019) Cellular functions and molecular mechanisms of non-lysine ubiquitination. *Open Biol* 9
67. Reinstein E, Scheffner M, Oren M, Ciechanover A, Schwartz A (2000) Degradation of the E7 human papillomavirus oncoprotein by the ubiquitin-proteasome system: Targeting via ubiquitination of the N-terminal residue. *Oncogene* 19:5944–5950
68. Skieterska K, Rondou P, van Craenenbroeck K (2016) Dopamine D 4 receptor ubiquitination. *Biochem Soc Trans* 44:601–605
69. Ciechanover A N-terminal ubiquitination. In: *Ubiquitin-Proteasome Protocols* vol. 301. Humana Press, pp 255–270
70. Ciechanover A, Ben-Saadon R (2004) N-terminal ubiquitination: More protein substrates join in. *Trends Cell Biol* 14:103–106. <https://doi.org/10.1016/j.tcb.2004.01.004>
71. Boban M, Ljungdahl PO, Foisner R (2015) Atypical ubiquitylation in yeast targets lysine-less Asi2 for proteasomal degradation. *J Biol Chem* 290:2489–2495
72. Kuo ML, den Besten W, Sherr CJ (2004) N-terminal polyubiquitination of the ARF tumor suppressor, a natural lysine-less protein. *Cell Cycle* 3:1367–1369
73. Bateman A et al (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480–D489
74. Howe KL et al (2021) Ensembl 2021. *Nucleic Acids Res* 49:D884–D891
75. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
76. Letunic I, Khedkar S, Bork P (2021) SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Res* 49:D458–D460
77. Cokelaer T et al (2013) BioServices: a common Python package to access biological Web Services programmatically. *Bioinformatics* 29:3241–3242
78. Mészáros B, Erdős G, Dosztányi Z (2018) IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* 46:W329–W337
79. Steinegger M et al (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 20:1–15
80. Remmert M, Biegert A, Hauser A, Söding J (2012) HHblits: Lightning-fast iterative protein sequence searching by HMM–HMM alignment. *Nat Methods* 9:173–175
81. Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA (2019) Raincloud plots: A multi-platform tool for robust data visualization [version 1; peer review: 2 approved]. *Wellcome Open Res* 4:1–40
82. Shevchenko A, Tomas H, Havliš J, Olsen JV, Mann M (2007) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1:2856–2860
83. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26:1367–1372
84. Perez-Riverol Y et al (2022) The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 50:D543–D552
85. Paraskevopoulos K et al (2014) Dss1 Is a 26S Proteasome Ubiquitin Receptor. *Mol Cell* 56:453–461
86. Suga M, Hatakeyama T (2005) A rapid and simple procedure for high-efficiency lithium acetate transformation of cryopreserved *Schizosaccharomyces pombe* cells. *Yeast* 22:799–804
87. Kampmeyer C et al (2017) The exocyst subunit Sec3 is regulated by a protein quality control pathway. *J Biol Chem* 292:15240–15253

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.