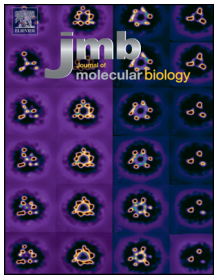

Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast

Benjamin P. Roscoe and Daniel N.A. Bolon

Department of Biochemistry and Molecular Pharmacology, *University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA*

Correspondence to Daniel N.A. Bolon: *dan.bolon@umassmed.edu.*
http://dx.doi.org/10.1016/j.jmb.2014.05.019
Edited by A. Keating

Abstract

The complexity of biological interaction networks poses a challenge to understanding the function of individual connections in the overall network. To address this challenge, we developed a high-throughput reverse engineering strategy to analyze how thousands of specific perturbations (encompassing all point mutations in a central gene) impact both a specific edge (interaction to a directly connected node) and an overall network function. We analyzed the effects of ubiquitin mutations on activation by the E1 enzyme and compared these to effects on yeast growth rate. Using this approach, we delineated ubiquitin mutations that selectively impacted the ubiquitin-E1 edge. We find that the elasticity function relating the efficiency of ubiquitin-E1 interaction to growth rate is non-linear and that a greater than 50-fold decrease in E1 activation efficiency is required to reduce growth rate by 2-fold. Despite the robustness of fitness to decreases in E1 activation efficiency, the effects of most ubiquitin mutations on E1 activation paralleled the effects on growth rate. Our observations indicate that most ubiquitin mutations that disrupt E1 activation also disrupt other functions. The structurally characterized ubiquitin-E1 interface encompasses the interfaces of ubiquitin with most other known binding partners, and we propose that this enables E1 in wild-type cells to selectively activate ubiquitin protein molecules capable of binding to other partners from the cytoplasmic pool of ubiquitin protein that will include molecules with chemical damage and/or errors from transcription and translation.

Introduction

Determining how genes function together as biological systems is a defining challenge of the genomic era. While genome sequences reveal the DNA blueprint of organisms, deciphering how this blueprint leads to biological function is challenging due in large part to the complexity of protein interaction networks [1,2]. For example, many phenotypes are mediated by multiple genes [3], and numerous genes exhibit pleiotropy [4]. Tremendous progress has been made in mapping the connections (also known as edges) between genes and gene products by both genetic [5–7] and biochemical approaches [8,9]. Epistatic analyses of gene knockout combinations have provided a broad understanding of the impacts of node deletions on network function [5]. In addition, approaches have been developed to analyze the effects of disrupting individual network edges by identifying mutations that eliminate a specific interaction [10–12]. However, for most complex biological networks, the elasticity function [13,14] relating network edge strength (e.g., the affinity of a specific protein–protein interaction) to overall network function (e.g., growth rate) is poorly understood. To address this challenge, we developed a high-throughput strategy to analyze how all point mutations in a central gene impact both an edge to a directly connected node in its network and an overall network function. Of note, we assess overall network function by quantifying yeast growth rate as a measure of experimental fitness under defined environmental conditions. In this work, we report experiments with ubiquitin and the E1 enzyme that provide fundamental insights into regulated protein degradation in eukaryotes.

Systematic investigations of the relationships among gene or protein sequence, function, and fitness provide new opportunities to bridge molecular, systems, and

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

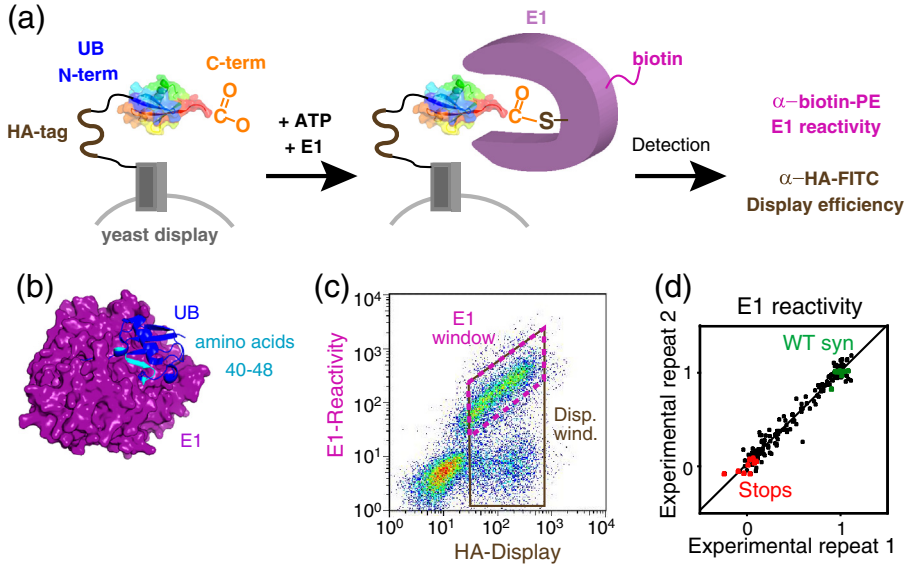

Fig. 1. E1 reactivity of ubiquitin mutants assessed using yeast display and FACS. (a) Experimental setup: the C-termini of displayed ubiquitin variants are free to react with E1 and the upstream HA tag enables normalization for display level. (b) Molecular representation of the adenylation domain of E1 shown in surface representation and colored purple complexed with ubiquitin shown in cartoon form with amino acids 40–48 that were chosen as an initial test shown in cyan. (c) FACS analysis of pools of ubiquitin point mutants in the test region. Cells in the sort windows were independently collected and analyzed by focused deep sequencing in order to estimate the E1 reactivity of each mutant. (d) E1 reactivity estimates are reproducible (R^2 = 0.96) in a full experimental repeat.

evolutionary biology [15–18]. While a wealth of studies demonstrate that the fitness effects of mutations are mediated by biochemical changes [19–25], most systematic studies of mutants have focused predominantly on either growth effects [14,26,27] or biochemical effects [11,28–30]. The relationships between mutant effects on biochemical properties and experimental fitness under defined conditions have been studied using traditional approaches for a handful of genes, almost all of which encode enzymes that catalyze a single critical chemical transformation. In many of these systems [24,31–33], the experimental fitness effects of a set of mutants can be accurately predicted based on both the proficiency of the mutant enzyme and physiological models of biochemical fluxes [19]. However, for the majority of genes (particularly those that perform multiple functions or whose functions are not fully appreciated), the relationships between a mutation's impact on biochemical properties and fitness remain unclear. In theory, each activity of a multi-functional protein may contribute independently to fitness and may be predicted based on flux models, or the contributions of each activity to function may be interdependent, likely depending on the molecular and evolutionary context of each particular gene product. Distinguishing these possibilities provides insights into network function and can be accomplished by systematically investigating the effects of mutations on both biochemical function and experimental fitness.

We determined the effects of all possible point mutants in ubiquitin on activation by the E1 enzyme using a biochemical assay and compared this new functional map to a corresponding map of experimental fitness effects in yeast that we had previously determined [27]. Through its ability to covalently link to other proteins, ubiquitin contributes to multiple important cellular processes including regulated protein degradation [34]. The covalent attachment of ubiquitin is mediated by a series of enzymes, with E1 activation serving as the first step in this process. E1 activates ubiquitin by first adenylating the C-terminus of ubiquitin and subsequent covalent attachment via a catalytic cysteine in E1 [35,36]. In this work, we developed a biochemical screen for the relative effects of ubiquitin mutations on E1 reactivity. We find that most ubiquitin variants that were deficient for E1 activation in this screen also failed to support robust yeast growth, consistent with the essential role of this reaction [37]. However, our results also demonstrate that activation of wild-type ubiquitin is far more efficient than required to support robust growth and that the relationship between the E1 reactivity of an ubiquitin mutant and yeast growth rate is non-linear. Despite this non-linear elasticity function [13,14], the effects of most ubiquitin mutants on E1 activation were similar to their effects on

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

yeast growth rate. These observations suggest that most ubiquitin mutations that lead to defects in E1 activation also lead to defects in other ubiquitin network edges (e.g., binding to the proteasome) and that the combined biochemical defects of these ubiquitin mutations are responsible for the observed fitness defect.

Results and Discussion

Investigating E1 reactivity

We developed a bulk competition approach to interrogate the effects of all possible ubiquitin point mutations on E1 reactivity (Fig. 1). Comprehensive site saturation libraries of ubiquitin point mutations in eight pools of 9–10 consecutive amino acid positions that can be efficiently interrogated with short read sequencing [27,38] were transferred to a yeast-display system. Importantly, this yeast-display setup (Fig. 1a) presents ubiquitin molecules with a free C-terminus, which is required for activation by E1. For initial method development, we focused on a region of ubiquitin encoding amino acids 40–48 that form a close contact with E1 in the co-crystal structure [36] (Fig. 1b). Display cells were reacted with a limiting concentration of yeast E1 (Uba1) (Supplementary Fig. S1), labeled with fluorescent antibodies targeted to E1 and an HA epitope used to assess display efficiency, and were separated by flow cytometry into pools of E1-reactive cells and HA-displaying cells (Fig. 1c). Plasmids encoding the mutated ubiquitin library were recovered and the mutated region was sequenced [38]. Differences in observed mutant frequency from E1-reactive cells and HA-displaying cells were used to assess the effects of each ubiquitin point mutation on E1 reactivity.

In order to achieve the throughput required to systematically scan all ubiquitin point mutants, we utilized a display system that enabled a fluorescence-activated cell sorting (FACS) and deep sequencing readout to report on E1 activation efficiency. This provides a reasonable approximation of the E1 activation process that occurs in cells, but reactions that occur on a cell surface are not perfect mimics of reactions in solution. E1 contains multiple domains and can bind two ubiquitin molecules at the same time: one that is adenylated (through strong non-covalent association between E1 and adenosine) and one that is covalently attached to the catalytic cysteine of E1. In our bulk competitions, E1 attachment to the displaying yeast cells requires adenylation, but not necessarily the transfer and attachment of displayed ubiquitin to E1's catalytic cysteine. In addition, the display of multiple copies of the same ubiquitin molecule on the surface of each display cell will constrain E1 to primarily react with two ubiquitin molecules of the same sequence (i.e., in the assay, E1 should preferentially bind two ubiquitins of the same sequence, one covalently and one non-covalently, for ubiquitin variants capable of both adenylation and transfer to the catalytic cysteine).

We developed the system to interrogate the kinetics of covalent activation of ubiquitin mutants with high sensitivity. To develop a sensitive assay for this rapid kinetic process, we utilized a limiting amount of E1 enzyme mixed into a suspension of display cells and quenched with free ubiquitin such that displayed ubiquitin variants were in competition with each other. Ubiquitin mutations that could be efficiently sequenced in the same reaction were assayed together resulting in eight pools of ubiquitin mutations that were competed with E1 separately. While quenching with free ubiquitin serves to provide a final stop to the reaction, the limiting concentration of E1 means that depletion of E1 during the competition phase can vary depending on the relative activation efficiency of library variants, leading to potential distinctions in the sensitivity of our assay for regions of ubiquitin analyzed in different pools. Region-to-region consistency could be achieved by using conditions of ultra-limiting E1 (e.g., by including an equivalent concentration of soluble ubiquitin to E1 in the competitions or by using rapid mixing techniques to quench reactions before E1 is depleted). Of note, experiments using ultra-limiting E1 conditions would be strongly influenced by mixing conditions, introducing additional sources of potential experimental variation.

Given the caveats of our experimental approach, we performed a number of control analyses to assess the quality of our data. The site saturation ubiquitin libraries include all 64 codons at each position and thus encode many wild-type synonyms and stop codons. Across all measurements, wild-type synonyms exhibit similarly robust E1 reactivity and stop codons exhibit baseline levels of E1 reactivity (Fig. 1d and Supplementary Table 1), consistent with selection on the ubiquitin amino acid sequence in our screen and the known requirement of the C-terminus of ubiquitin for E1 reactivity. This approach resulted in highly reproducible (R^2 = 0.96) measurements in a full experimental repeat (Fig. 1d) and was used to interrogate the E1 reactivity of mutants across all of ubiquitin (Fig. 2a and Supplementary Tables 1 and 2). We also developed an independent assay using purified proteins to measure the E1 reactivity of individual mutants relative to wild type (Supplementary Fig. S2). Analyses of a panel of mutants indicated that yeast-display E1 reactivity measurements of ubiquitin mutants in different regions correlate reasonably well (R^2 = 0.6) with measurements made with purified proteins (Fig. 2b). Multiple factors may contribute to distinctions in E1 activation observed in the yeast display and purified protein analyses including noise in each experiment (experimental variation in display and purified assays were both about 5%; see Fig. 1d and Supplementary Fig. S2e), variations in selection strength for different regions

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

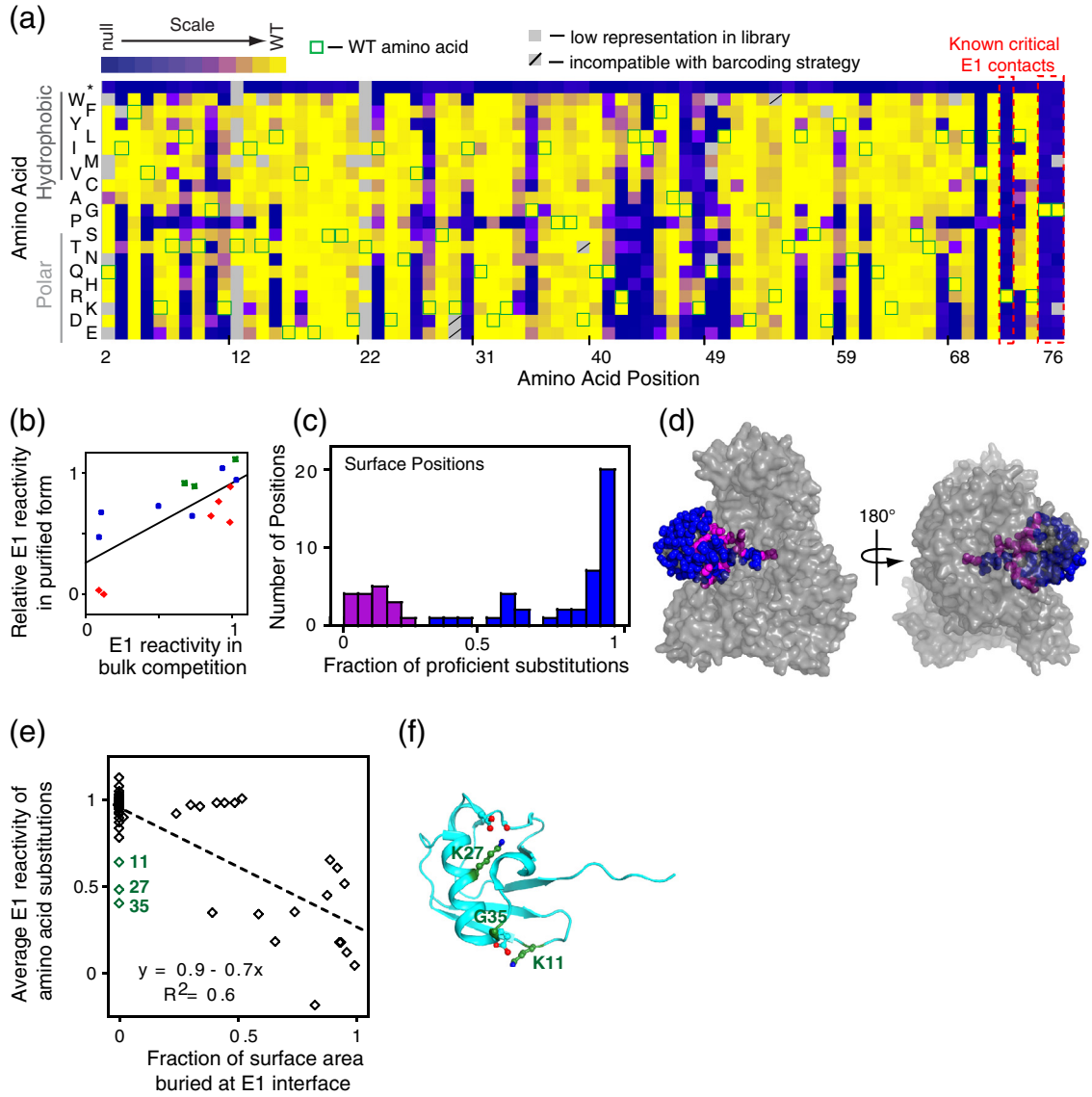

Fig. 2. Mapping the effects of ubiquitin mutants on E1 reactivity to structure. Heat map representation of E1 reactivity for ubiquitin mutants normalized to wild-type synonyms and stop codons. (a) Map of entire ubiquitin sequence except for the initiating methionine. Of note, we did not observe any ubiquitin mutants present in our plasmid libraries that failed to efficiently display the HA epitope. (b) Comparison of E1 reactivity estimates from bulk competitions with independent measurements made using purified proteins. Individual mutants from three regions analyzed in separate bulk competitions are distinguished by colors: green squares (K33A, E34G, G35N), blue circles (Q40A, I44M, I44V, K48R), and red diamonds (H68N, H68Q, H68S, L69S, R72S, G75D). Overall estimates from bulk competitions positively correlate (R^2 = 0.6) with those using purified proteins. (c) Sensitivity of ubiquitin surface positions for E1 reactivity assessed by measuring the number of amino acids compatible with proficient E1 activation within 20% of wild-type ubiquitin. Sensitive positions that tolerate less than 25% of analyzed amino acid substitutions were color coded purple and all other positions were color coded blue. (d) Mapping sensitive (purple) and tolerant (blue) positions on the ubiquitin surface onto the structurally [36] characterized complex with E1 (shown in transparent gray). (e) Correlation between the average impact of substitutions at each ubiquitin position on E1 activation and the fraction of wild-type side-chain surface area buried at the E1 interface. Positions 11, 27, and 35 [colored green in (d)] stand out as sensitive for E1 activation despite not burying side-chain surface at the binding interface. (f) The wild-type amino acids at these positions (K11, K27, and G35) all form intramolecular interactions that likely contribute to the ground state structure and/or dynamics of ubiquitin.

of ubiquitin in the yeast-display experiments, and biochemical distinctions due to yeast surface *versus* solution reaction conditions. These analyses indicate that the bulk yeast-display studies distinguish highly active from weakly active variants but that smaller distinctions in relative activation efficiency

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

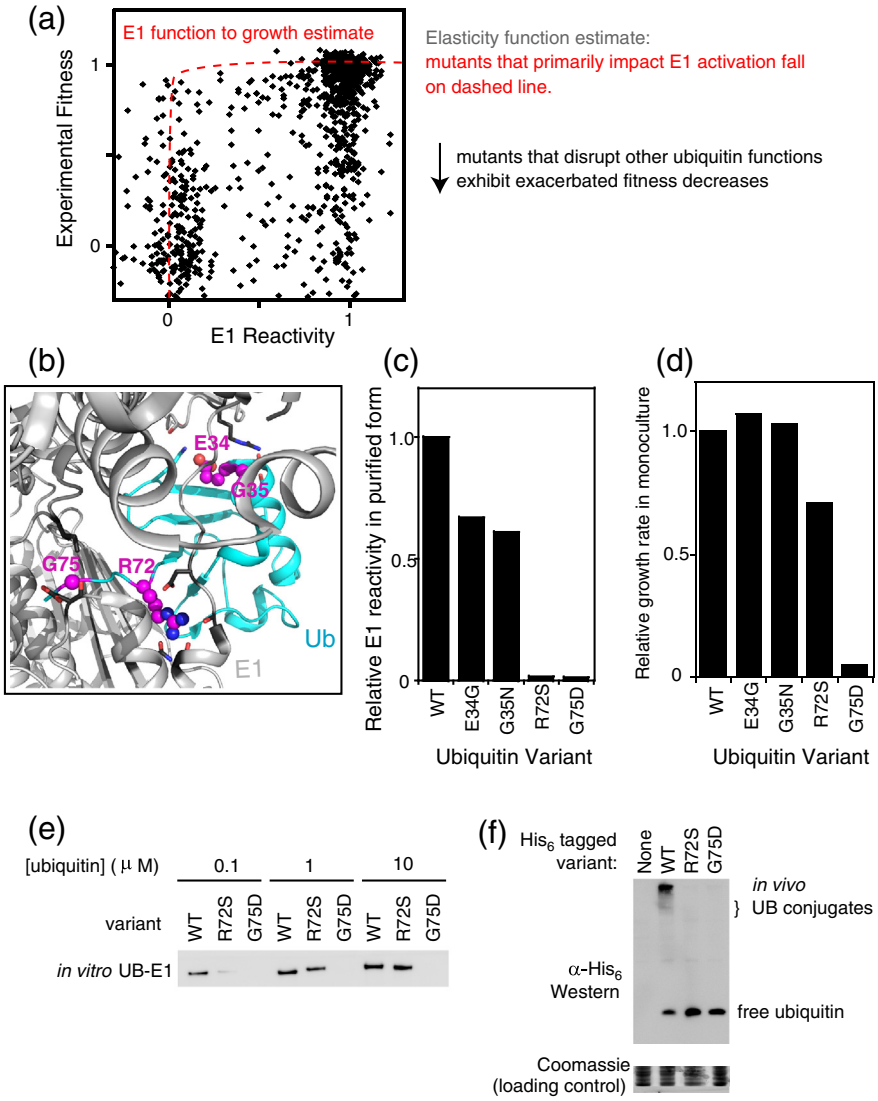

Fig. 3. Relating the effects of ubiquitin mutations on E1 reactivity to experimental fitness. (a) Comparison of the effects of ubiquitin mutations on E1 reactivity and yeast experimental fitness. (b–f) Ubiquitin mutations at four positions located at the E1 interface were analyzed further using purified components. (b) Illustration indicating the location of ubiquitin amino acids E34, G35, R72, and G75. E34 and G35 are located at the periphery of the interface between ubiquitin and E1. R72 is located in a deep cavity on the surface of E1 and forms multiple hydrogen bonds across the interface. G75 is located in a narrow cleft adjacent to the active site. (c) E1 reactivities for a panel of mutants at these positions were determined using purified proteins and binary competitions with wild-type ubiquitin (see also Supplementary Fig. S2). (d) The experimental fitness of this panel of mutants was analyzed by monitoring the growth rate of each variant in isolation. (e) The E1 activation potential of purified wild type (WT) and R72S and G75D ubiquitin variants analyzed without competition by Western blotting for high-molecular-weight ubiquitin. (f) The accumulation pattern of epitope-tagged ubiquitin variants expressed in yeast co-expressing endogenous untagged ubiquitin.

may not be determined with confidence from the bulk competitions.

Mapping mutant effects on E1 activation to structure

Structural mapping indicates a general correspondence between E1 reactivity and contact surfaces observed [36] between E1 and ubiquitin. To estimate the sensitivity of each ubiquitin position, we calculated the fraction of mutations at each position in ubiquitin that were proficient for E1 activation (within 20% of the average wild-type synonym). Of note, 90% of wild-type synonyms but no stop codons classify as E1 proficient under this definition. Most positions on the surface of ubiquitin either tolerated almost every amino acid substitution or were highly sensitive to mutation (Fig. 2c and Supplementary

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

Fig. S3a). Mapping the tolerant surface ubiquitin positions to the structure indicates that sensitive positions were located almost exclusively at the interface with E1 and tolerant positions remained predominantly solvent accessible (Fig. 2d). The fraction of ubiquitin side-chain surface area buried at the interface with E1 correlates with the observed variation in the average effect of ubiquitin mutations on E1 reactivity (Fig. 2e), indicating that surface area burial is a major determinant of mutational sensitivity. All ubiquitin positions that bury greater than 60% of their side-chain surface area at the E1 interface were strongly sensitive to mutation. Conversely, the majority of ubiquitin positions that do not bury any side-chain surface area at the E1 interface are almost completely tolerant to mutation. Three ubiquitin positions that do not bury side-chain surface area at the E1 interface exhibit mutational sensitivity that stands out (positions 11, 27, and 35 shown in Fig. 2e and f). The side chains at these three positions all make intramolecular contacts that may impact ubiquitin structure and dynamics: both K11 and K27 form salt bridges between different secondary structure elements, and G35 is part of a turn structure and has a main-chain conformation (positive phi angle) energetically disfavored for non-glycine amino acids (Fig. 2f). These structural analyses are consistent with the chemical intuition that the functional sensitivity of a position to mutation is primarily determined by direct binding interfaces [39], as well as structural integrity [40] and dynamics [41].

In the solvent-inaccessible core of ubiquitin, most positions exhibited a similar pattern of mutational tolerance for E1 reactivity (Supplementary Fig. S3b–d). Of the 16 core positions, 13 have aliphatic side chains in wild-type ubiquitin (six Leu, four Ile, and three Val) that form a hydrophobic cluster known to be a driving force for stabilizing native protein structures [42,43]. Consistent with observations that the protein folding stability of wild-type ubiquitin is far greater than required for yeast growth [27,44], we observed that modest substitutions to other aliphatic side chains were generally well tolerated for activation by E1 (Supplementary Fig. S3b). In contrast, substitutions to polar amino acids were poorly tolerated, suggesting that these substitutions likely disrupt the ground state structure or the dynamics of ubiquitin. Three positions that are exceptions to this rule are all located at the edge of the solvent-inaccessible core (Supplementary Fig. S3c and d) where long polar amino acid side chains may be able to access solvent without disrupting the structure.

Relationship between ubiquitin mutant effects on E1 activation and experimental fitness

The fitness effects of ubiquitin mutations integrate over impacts on the entire ubiquitin interaction network. For this reason, ubiquitin mutants with identical impacts on E1 activation can have different fitness effects (e.g., due to different effects of each mutation on binding to and recycling by the proteasome [41,44–46]). Comparing the upper bound of fitness effects to E1 reactivity (see red broken line in Fig. 3a, which presumes that negative measures of E1 activation are due to experimental noise) provides an estimate of the underlying elasticity relationship demarcated by ubiquitin mutations that primarily impact E1 activation. Of note, the red broken line in Fig. 3a represents our conceptualization of the behavior of mutations that impact E1 activation alone and is not the result of an explicit fit to the data. Our conceptualization is illustrated as a curve because this represents the general shape of elasticity relationships that have been observed for other proteins [14,19,32]. Our systematic scan of ubiquitin mutations indicates that the elasticity function relating E1 activation efficiency to growth rate is non-linear and that E1 reactivity can be reduced to levels below the threshold for measurement in our screen with minimal impacts on fitness.

To further assess the minimum level of E1 reactivity for a ubiquitin mutant required to support yeast growth, we investigated a panel of individual ubiquitin mutations (Fig. 3b–f). We chose non-conservative mutations located at the structurally determined interface with E1 (Fig. 3b) that exhibited E1 activation defects in display competitions. We independently determined the E1 reactivity of each mutant in our panel using purified proteins (Supplementary Fig. S4). Consistent with our bulk experiments, each mutation in this panel reduced E1 reactivity compared to wild type (Fig. 3c). Both the E34G and G35N ubiquitin mutations reduced E1 activation by roughly 40%, but both support yeast growth rates as the sole ubiquitin in cells that are indistinguishable from wild type (Fig. 3c and d). The R72S and G75D mutations caused severe defects (~50-fold) for E1 reactivity relative to wild type. Of these two severely E1-deficient mutations, R72S supported yeast growth albeit at a rate 30% slower than wild type and G75D exhibited null-like growth based on monoculture experiments (Fig. 3d and Supplementary Fig. S4).

To further investigate the E1 activation potential of R72S and G75D, we tested them individually in purified form with varying concentrations of E1 (Fig. 3e). At low E1 concentrations, both R72S and G75D ubiquitin reacted poorly with E1 compared to wild type providing an additional confirmation of the activation defects of these mutations. At higher concentrations, R72S was capable of reacting with E1. We did not observe reaction of G75D ubiquitin with E1 even at concentrations 100-fold greater than those where we observed reaction with wild type or at concentrations 10-fold greater than for the R72S variant. The G75D ubiquitin variant was recently recovered in a phage display selection for E1 reactivity [47], which may be due to the use of non-covalent and

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

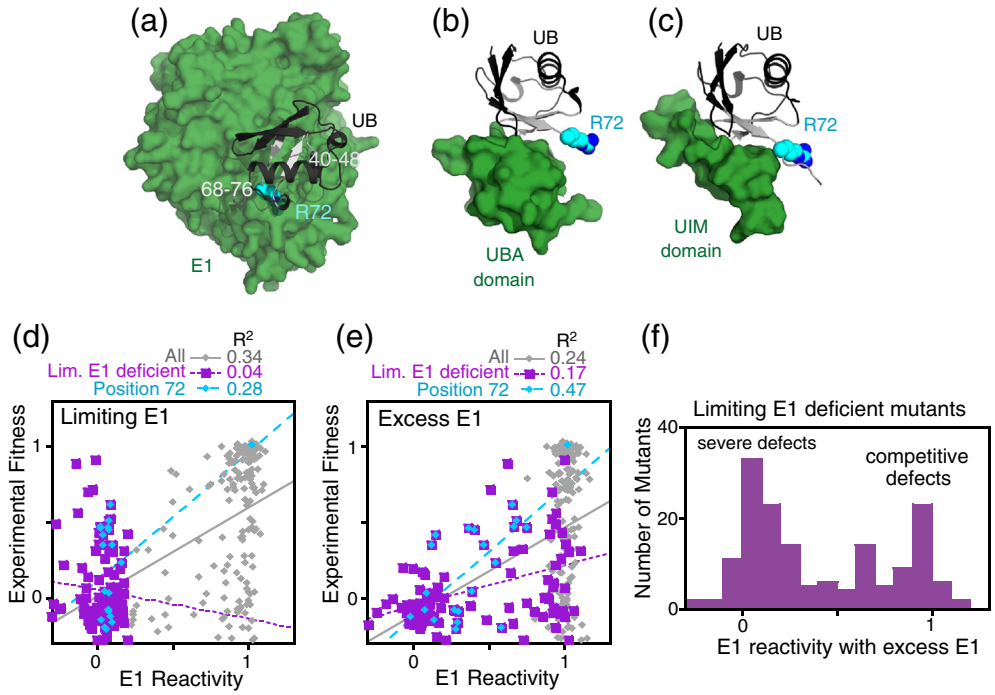

Fig. 4. Distinguishing the E1 reactivity of ubiquitin mutants near the threshold required to support yeast growth. Two regions of ubiquitin encompassing amino acids 40–48 and 68–76 were analyzed using bulk competitions performed with excess E1. (a) Molecular illustration highlighting the contacts between these two ubiquitin regions (shown in light gray) and E1 (shown in dark green) based on 3CMM.PDB [36]. (b and c) Illustrations of contacts between ubiquitin and two common ubiquitin binding domains (UBA and UIM) based on 1QOW.PDB [54] and 2OOB.PDB [55], respectively. (d and e) The impacts of ubiquitin mutations in these two regions on experimental fitness compared with effects on reactivity with either limiting E1 (d) or excess E1 (e). Ubiquitin mutations deficient for activation with limiting E1 are shown as purple squares and mutations at position 72 are highlighted with a light-blue diamond in (d) and (e). (f) For ubiquitin mutations deficient for reactivity with limiting E1, the distribution of E1 reactivity observed in experiments with excess E1.

unstable [48] Fos-Jun-mediated association between ubiquitin and phage particles or other distinctions between the experimental setups. Our observations with purified proteins show that purified G75D ubiquitin was severely defective for E1 reactivity.

The ability of R72S ubiquitin to support yeast growth, albeit with an approximately 30% defect relative to wild type, was unexpected as position 72 is the main determinant of activation specificity for ubiquitin-like proteins [49–51]. Consistent with the importance of R72 in E1 activation, our binary competitions with purified proteins (Fig. 3c) indicate that R72S ubiquitin is activated by E1 approximately 2% as efficiently as wild type. This represents an upper estimate on the E1 reactivity of a ubiquitin mutant required to support yeast viability, as we cannot rule out the possibility that the R72S mutation impacts other ubiquitin functions such as ubiquitin transfer to E2 or E3 [52].

To examine how the E1 reactivity that we observed *in vitro* extends to *in vivo* utilization of ubiquitin, we measured the accumulation pattern of the R72S and G75D ubiquitin variants in yeast cells. In cells, ubiquitin exists primarily in two pools: free ubiquitin monomers of low molecular weight or covalent conjugates of far greater molecular weight (depending on the mass of the targeted protein and the number of ubiquitin molecules attached). To examine how ubiquitin variants accumulate in these two pools, we co-expressed untagged wild-type ubiquitin with mutant versions tagged with an epitope tag that is compatible with *in vivo* function [53]. The separation of denatured cell lysates by gel electrophoresis followed by Western blotting for the epitope tag enabled estimation of the fraction of the tagged ubiquitin variant incorporated into conjugates while in competition with wild-type ubiquitin in cells. These experiments provide a valuable examination of ubiquitin and E1 in the complex cellular environment, but they do not distinguish E1 activation from contributions of other enzymes (e.g., E2s and E3s) in the conjugation process. While epitope-tagged wild-type ubiquitin readily accumulated as conjugated species, we did not observe appreciable accumulation of conjugates of either R72S or G75D (Fig. 3f), consistent with our observations that both of these mutations cause severe defects for E1 activation relative to wild type.

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

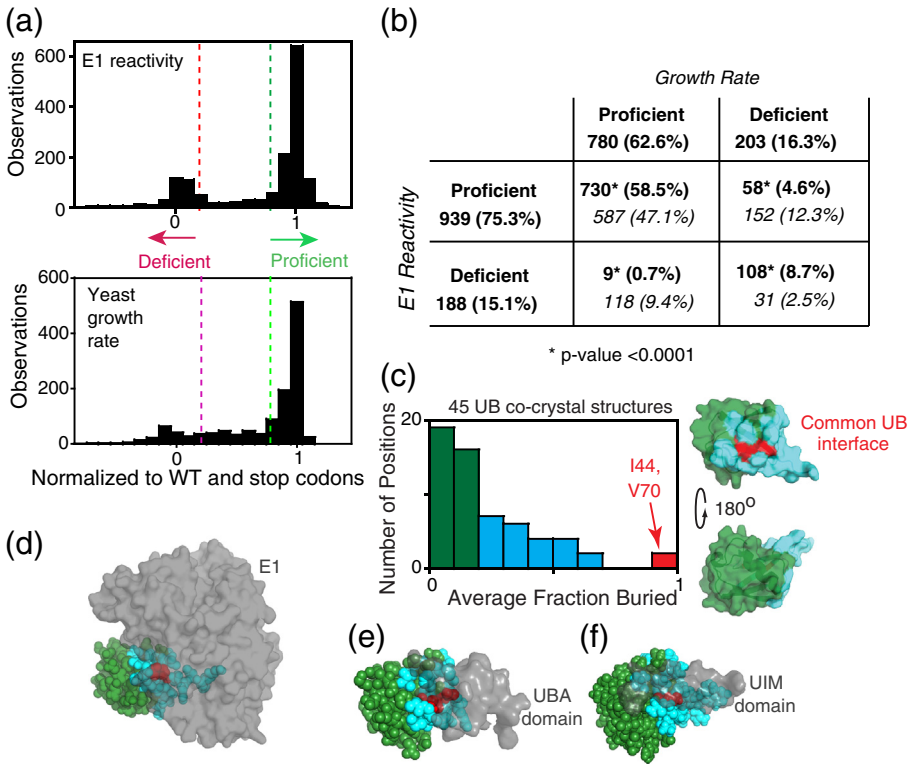

Fig. 5. Similar impacts of ubiquitin mutations on E1 reactivity and experimental fitness. (a) Distribution of the effects of ubiquitin mutations on E1 reactivity with limiting E1 (top) and previously determined [27] effects on yeast growth rate (bottom). Stop codons were not included in the E1 or fitness panels, and four severely depleted reactivity measurements were excluded in order to focus on the main features of the distribution. (b) Contingency table describing the observed overlap of the effects of ubiquitin mutations on E1 reactivity and yeast growth rate. All observed frequencies were statistically skewed compared to expectations from independent binomial distributions, which are shown in italics. (c) Analyses of the side-chain surface area buried between ubiquitin and many different binding partners in 45 high-resolution co-crystal structures indicate that I44 and V70 are almost always fully buried at the binding interface. Positions surrounding I44 and V70 are buried in a large fraction of interfaces depending on the orientation of the binding partner relative to ubiquitin. (d) The adenylation domain of E1 (shown in gray) almost completely encompasses ubiquitin surfaces that were structurally determined to contribute to binding interfaces with other proteins. (e and f) Domains that commonly mediate binding to ubiquitin bind to smaller surface regions of ubiquitin than E1.

Investigating activation potential of ubiquitin mutants with excess E1

To delineate the effects of ubiquitin mutations on E1 activity near the threshold required to support robust yeast growth rates, we performed display experiments under conditions of excess E1 for two ubiquitin regions encompassing amino acids 40–48 and 68–76 located at the E1 interface (Fig. 4a). Excess E1 in these experiments provides the opportunity for each displayed mutant to react with minimal competition from other variants and distinguishes ubiquitin mutants with severe E1 activation defects from those with competitive activation defects that may not compromise fitness on their own. The regions we chose to study in these experiments were located at structurally characterized interfaces with other ubiquitin binding partners [56] including ubiquitin-associated domains (UBA) and ubiquitin-interacting motifs (UIM) as illustrated in Fig. 4b and c.

The relationship between the effects of mutations on fitness and activation efficiency with limiting E1 in these two regions (Fig. 4d) is similar to the pattern observed across all positions in ubiquitin (Fig. 3a). In particular, these regions contain many mutations that cause deficient E1 activation with limiting E1, including some that exhibit growth rates approaching wild type (Fig. 4d). Of note, R72 forms extensive contacts with E1 but is largely exposed in complexes with UBA or UIM proteins such that mutations at this position may primarily impact E1 activation. The contact between arginine 72 and E1 has previously been demonstrated to be important for efficient ubiquitin activation [36,49,50]. As expected based on these previous observations, only the wild-type amino

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

acid at position 72 was compatible with proficient E1 activation under limiting conditions (Fig. 4d). In terms of fitness effects, all point mutations at position 72 were deleterious, though they ranged from roughly 40% growth defects to null in estimates from bulk competitions (Fig. 4d).

With excess E1, the relationship between the effects of ubiquitin mutations on activation and fitness shifted distinctly (Fig. 4e and Supplementary Tables 3 and 4). Excess E1 caused an increase in the E1 reactivity observed for many ubiquitin mutations. All ubiquitin mutations that caused severe activation defects with excess E1 also exhibited deficient growth in yeast, suggesting that E1 activation in this set of ubiquitin mutations is below the level required for even modest yeast growth. While this class of ubiquitin mutations likely has fitness limiting E1 activation defects, they may also have defects in other critical ubiquitin functions due to the structural location of many of these residues at contact sites with other ubiquitin binding domains (Fig. 4b and c).

We observed many ubiquitin mutations that caused activation defects at limiting E1 but were capable of activation with excess E1 (Fig. 4e and f). This class of ubiquitin mutation has competitive activation defects that were not severe enough on their own to prohibit yeast growth but that would competitively hinder activation in the presence of ubiquitin molecules that are more E1 reactive. Most ubiquitin mutations that caused competitive E1 activation defects exhibited severely impaired fitness, suggesting that these mutants also caused biochemical defects in other critical ubiquitin properties.

Amino acid substitutions at position 72 resulted in activation efficiencies with excess E1 that correlated positively with fitness effects (Fig. 4e). The location of R72 in a deeply buried cleft in the structure with E1 (Fig. 4a) and at the periphery of structures of ubiquitin with other binding domains (Fig. 4b and c) suggests that mutations at position 72 may primarily impact E1 activation within the ubiquitin interaction network. Consistent with this structural inference, the activation observed with excess E1 of ubiquitin mutants at position 72 disproportionately correlates with fitness effects compared to mutants at other positions.

Correspondence between ubiquitin mutant effects on E1 activation and fitness

The overall distribution of the effects of ubiquitin mutations on E1 reactivity under conditions of limiting E1 is similar to the distribution of experimental fitness effects (Fig. 5a). Both distributions are bi-modal with a main peak near wild type (defined as 1) and null (defined as 0). In principle, the similar profiles of mutant effects on fitness and this E1 function could be due to a strong or linear relationship where E1 activation is rate limiting for yeast growth. However, this idea is incompatible with our observation that purified R72S ubiquitin reduces E1 activation 50-fold while impairing growth rate less than 2-fold, as well as observations of many other mutations that in bulk experiments (Figs. 3a and 5d and e) exhibit either strong competitive E1 defects with robust growth or proficient E1 activation with poor growth. As an alternative hypothesis, we propose that the similar observed distributions of ubiquitin mutant effects may be due to parallel biochemical impacts of many ubiquitin mutants on E1 activation and other ubiquitin functions.

For both experimental fitness and E1 reactivity, we classified each mutant as proficient (within 20% of the average wild-type synonym), deficient (within 20% of the average stop codon), or intermediate (Fig. 5a and b). To assess the reasonableness of these cutoffs, we examined wild-type synonyms and stop codons as positive and negative controls. We observe that greater than 90% of wild-type synonyms classify as proficient and greater than 90% of stop codons classify as deficient. In addition, these control sets exhibited no full misclassifications (e.g., proficient misclassified as deficient). The majority of E1-deficient ubiquitin mutations were deficient for yeast growth, and the preponderance of E1-proficient mutations supported robust yeast growth (Fig. 5b and Supplementary Table 5). The large number of mutations that exhibit robust yeast growth and E1 reactivity indicates that most ubiquitin variants that react efficiently with E1 tend to also function well in all other essential ubiquitin activities. The strong correspondence between E1 reactivity and fitness is consistent with the structurally characterized interfaces between ubiquitin and many binding partners that previous analyses have shown often center on the isoleucine 44 of ubiquitin [56].

Common ubiquitin binding interface

Structural analyses indicate that ubiquitin almost universally contacts partner proteins via a common binding surface (Fig. 5c–f). Our own analyses of 45 high-resolution co-crystal structures of ubiquitin with a variety of binding partners indicate that the side chains of I44 and V70 in ubiquitin are almost always fully buried at the interface (Fig. 5c) and surrounding positions are buried in a fraction of these structures depending on the shape and orientation of the binding partner (Fig. 5d–f and Supplementary Table 5). The adenylation domain of E1 forms a very large contact surface with ubiquitin that encompasses nearly all of the structurally characterized contacts between ubiquitin and other common binding domains including ubiquitin-associated domains (UBA) and ubiquitin-interacting motifs (UIM) as illustrated in Fig. 5d–f.

One-third (3300 $Å^2$) of the total surface area of ubiquitin is buried by contact with E1 [36], including the hydrophobic patch formed by L8, I44, and V70 that is required for binding to proteasomal [57] and

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

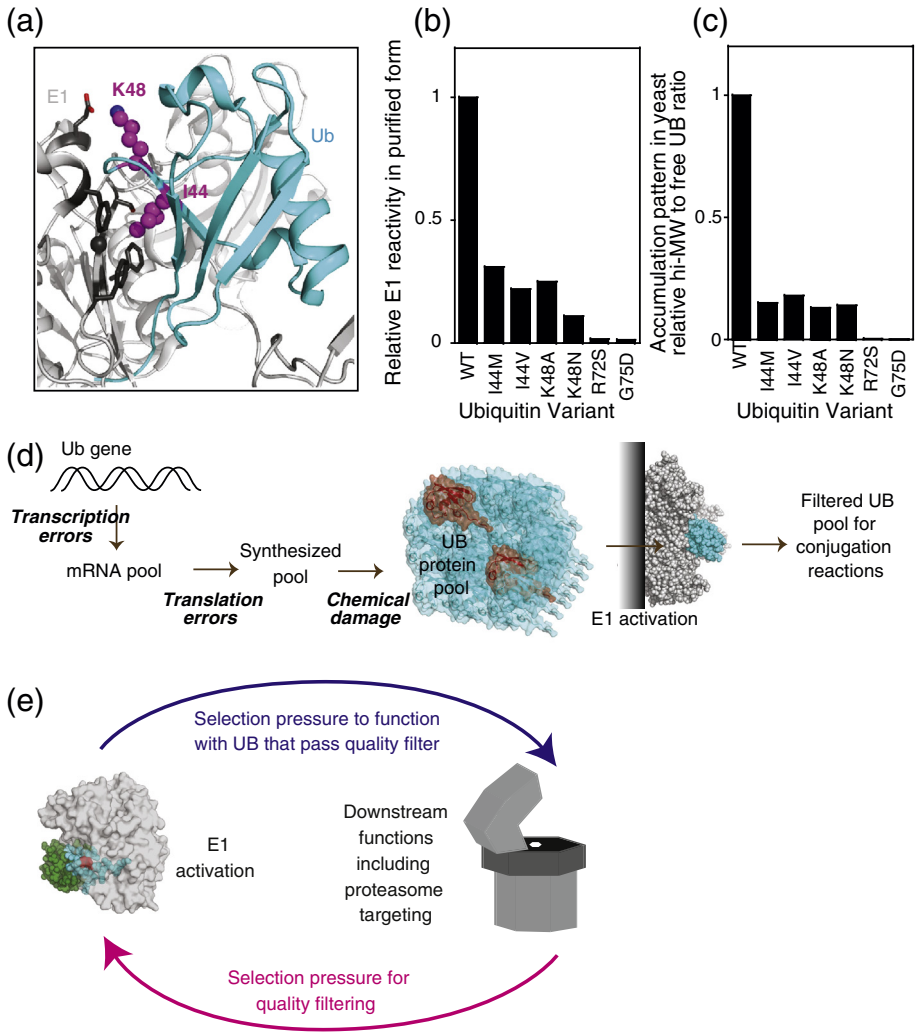

Fig. 6. E1 inefficiently activates ubiquitin variants with known biochemical defects in downstream pathways. (a) Molecular illustration of contacts between ubiquitin amino acids I44 and K48 with the adenylation domain of E1. Covalent linkage by the ubiquitin K48 side chain is a critical signal for proteasome-mediated degradation, and I44 forms direct binding contacts with almost all structurally characterized ubiquitin binding partners. Conservative substitutions to I44 and substitutions that removed the positive charge at K48 exhibited decreased E1 reactivity in purified form (b) and a reduced ability to conjugate to other proteins in cells co-expressing wild-type ubiquitin [(c) and Supplementary Fig. S7]. (d) The capability of E1 to preferentially activate ubiquitin protein molecules functional in downstream pathways provides the potential for post-translational quality control over the pool of ubiquitin protein in wild-type cells that will include ubiquitin protein molecules with errors from synthesis and/or chemical damage. (e) Selection model indicating that pressure for E1 to quality filter ubiquitin protein molecules could lead to parallel selection for downstream processes to function with the pool of ubiquitin variants efficiently activated by E1. This hypothesis is consistent with the correspondence we observe between ubiquitin mutant effects on competitive fitness and E1 activation, as well as the extensive binding interface between E1 and ubiquitin that encompasses the interfaces of ubiquitin with other binding partners.

many other ubiquitin receptors [56]. This large interface is not a chemical prerequisite for activation, as the chemistry of this reaction is localized to the C-terminal carboxyl group of ubiquitin. The same chemical mechanism is utilized to activate SUMO (a ubiquitin-like protein) despite a far smaller interface (1600 $Å^2$) [58]. The strong evolutionary conservation of E1, whose protein sequence is 50% identical between human and yeast [36], suggests that the large ubiquitin-E1 interface has been subject to stringent purifying selection in nature. The structural interface between ubiquitin and E1 is among the surfaces that exhibit the strongest evolutionary conservation in these proteins (Supplementary Fig. S5). Our results indicate that the large E1-ubiquitin interface enables E1 to preferentially activate ubiquitin variants that are functional across the majority of the ubiquitin interaction network.

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

While E1 poorly activates most ubiquitin variants that were growth deficient, we observed a small fraction of ubiquitin variants that were activated efficiently by E1 but that were incompatible with robust growth. Of the 939 ubiquitin point mutants that were E1 proficient (Fig. 5b), only 58 (6%) were growth deficient (Supplementary Fig. S6). These E1-proficient and growth-deficient ubiquitin point mutations were frequently located at the periphery of the interface between ubiquitin and the adenylation domain of E1 (Supplementary Fig. S6), consistent with the chemical intuition that peripheral contacts have smaller contributions to binding and reactivity than central contact points [59].

Discriminating activation by E1

We investigated the effects of ubiquitin mutations at two positions (I44 and K48) that are both critical for many downstream functions but whose mutational sensitivity for E1 reactivity was not predicted nor discussed in the description of the ubiquitin-E1 crystal structure [36]. I44 is at the center of a hydrophobic patch on the surface of ubiquitin that forms central contacts with most structurally characterized ubiquitin binding domains [56,57], and K48 is the site of covalent linkage to form ubiquitin polymers that target substrates for proteasome-mediated degradation [60]. In the structurally characterized complex with E1, I44 of ubiquitin forms hydrophobic contacts with multiple side chains of E1, while K48 forms a partially solvent accessible salt bridge with E892 from E1 (Fig. 6a). While hydrophobic contacts stabilize interfaces, they tend to tolerate slight changes to geometry and often permit conservative substitutions [26–29]. Using binary competitions with purified proteins, we observed that E1 reactivity was sensitive to even the most conservative substitutions of Ile to Val at position 44 (Fig. 6b and Supplementary Fig. S4). Similarly, partially solvent exposed salt bridges such as the one formed by K48 of ubiquitin with E1 often fail to stabilize protein structures and complexes due to the cost of displacing water molecules from unbound states [61]. However, the ubiquitin K48-mediated salt bridge to E1 is critical for efficient activation as mutations that remove the positive charge at position 48 reduce competitive reactivity with E1 (Fig. 6b and Supplementary Fig. S2).

While ubiquitin mutations at positions 44 and 48 almost universally caused strong activation defects with limiting E1, many were capable of activation with excess E1. With excess E1, eight substitutions at isoleucine 44 (Q, H, T, A, V, L, M, and W) and all substitutions except D and E at lysine 48 could be activated in experiments with excess E1 to an extent similar to wild-type ubiquitin and greater than R72S and many other mutations that were capable of supporting moderate yeast growth rates (Supplementary Table 4). These observations indicate that the strong fitness defects of most mutations at positions 44 and 48 were caused by biochemical defects other than E1 activation, consistent with the known biochemical function of K48 in forming critical polymers and I44 in binding to essential receptors [56,57].

We investigated how mutants at positions 44 and 48 accumulated *in vivo*. We observed that ubiquitin mutants at I44 or K48 exhibited decreased accumulation as conjugated species in cells co-expressing wild-type ubiquitin (Fig. 6c and Supplementary Fig. S7). Together with our observation that these mutations caused a competitive E1 defect *in vitro*, we infer that E1 likely contributes to the limited conjugation of these ubiquitin variants *in vivo*.

Post-translational quality filtering model

Based on our observations, we propose that E1 can discriminately activate ubiquitin protein molecules that are capable of binding to other partners from the pool of ubiquitin protein in cells that will include molecules with synthetic errors [62] and/or chemical damage (e.g., deamination of glutamine to glutamate) [63] (Fig. 6d). Of note, the E1 quality filtering that we propose occurs on the protein pool of ubiquitin generated from the wild-type ubiquitin gene. Estimates of the rate of transcription and translational errors [64] suggest that ~0.1% of ubiquitin protein molecules generated from the wild-type gene will contain an amino acid substitution error. The average observed fitness effect of an amino acid substitution in ubiquitin ($s_{avg} = -0.25$) leads to a rough estimate of the fitness benefit from E1 quality filtering of 0.025% ($s = 0.00025$) under the simplifying assumption that all amino acid substitutions are equally probable. In natural populations, selection coefficients above ≈0.0001% ($s \approx 10^{-6}$) would be subject to natural selection based on estimates of the effective population size of yeast [65] and the nearly neutral model [66]. Of note, the rates for most amino acid substitution errors (e.g., Ser to Thr) are largely unknown, making it challenging to infer potential selection at this level of resolution. Nonetheless, the available overall approximations of protein synthesis errors and fitness costs indicate that quality filtering by E1 could impact fitness by a magnitude sufficient for selection in natural populations of yeast.

The tendency for partner proteins to bind to a similar ubiquitin surface is consistent with our quality filtering hypothesis but does not rule out alternative explanations. For example, the ubiquitin surface that binds to E1 has biophysical properties (exposed hydrophobic side chains including I44) that favor macromolecular interactions and likely contribute to this surface serving as a common target of other binding partners. However, the observed biophysical diversity at protein–protein interfaces [67] makes it unlikely that biophysical preferences alone would lead to a near-universal ubiquitin binding surface that

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

is encompassed by the E1-ubiquitin interface, while the quality filtering model would.

In addition, the quality filtering model provides a rationale for the impacts of mutations to K48 on E1 activation efficiency. Our display studies and follow-up studies with individual ubiquitin mutations demonstrate that E1 can selectively filter ubiquitin protein molecules with substitutions at position 48 that cause known biochemical defects when activated and attached to substrates [60]. In future studies, it will be important to further test the quality filtering model. In particular, this model makes the potentially testable prediction that loss of quality filtering by E1 would lead to fitness defects.

The quality filtering model provides a possible rationale for the large contact area observed between E1 and ubiquitin: to enable E1 to extensively interrogate the properties of ubiquitin molecules and discriminate functional ubiquitin variants. From a theoretical perspective, large protein interfaces should provide a general opportunity to quality filter by preferentially binding to molecules without errors or damage that weaken binding affinity. Indeed, the large contact area previously observed between E1 and ubiquitin provides strong support for the proposed quality filtering model, though to our knowledge, this type of quality control mechanism has not been previously described in the literature. Our systematic analyses explicitly demonstrated the sensitivity of the ubiquitin interface with E1 to mutation eliminating the possibility that some sites at the interface may not have been sensitive to mutation. All positions at an interface can contribute to relative affinity (e.g., $\Delta\Delta G$ of binding compared to wild type) and this in turn contributes directly to competitive or relative reactivity for systems under equilibrium control [27,39,68]. Thus, all positions at interfaces have the potential to contribute strongly to competitive affinity and reactivity. Consistent with this idea, the interface between ubiquitin and E1 is conserved relative to other surfaces on these proteins (Supplementary Fig. S5). This type of interface-mediated quality filtering may reduce potential toxic consequences from flawed macromolecules in other systems with large contact surfaces (e.g., ribosome assembly). Of note, the expression of flawed proteins can impose a fitness cost even in the absence of aberrant function [69] that is of sufficient magnitude to be under selection in natural populations [66]. Quality filtering of ubiquitin pools by E1 may be particularly important because proteasome-mediated surveillance, an important quality control component for the majority of the proteome [70], may be unavailable for ubiquitin due to the inherent ubiquitin recycling function of the proteasome [71,72].

The selection we propose for quality filtering of ubiquitin protein pools by E1 predicts feedback selection such that ubiquitin mutations will often have similar biochemical effects across many ubiquitin functions (Fig. 6e). In this model, downstream ubiquitin functions impose selection pressure for quality filtering by E1, and quality filtering imposes feedback selection on downstream functions to be efficient with the set of ubiquitin variants that pass E1 quality filtering. Quality filtering and feedback selection provide a plausible evolutionary rationale for the structurally observed large ubiquitin-E1 interface that encompasses interfaces of ubiquitin with other binding partners.

Conclusions

Understanding the connections between function and fitness is a primary goal of many biological disciplines including systems biology and molecular evolution. While sound approaches have been developed to understand the connections between function and fitness for proteins that perform a single function [19,24,31–33], investigating potential interdependencies in multi-functional proteins had posed daunting technical challenges. This study demonstrates that systematic analyses of the effects of mutations on biochemical function and growth rates provide a powerful approach to investigate how edge-rich protein interaction networks contribute to overall biological function.

Materials and Methods

Libraries of ubiquitin point mutants were displayed on the surface of yeast as C-terminal fusions with Aga2-HA similar to previous descriptions [73,74]. Pools of yeast-displayed mutants were reacted with E1, labeled with fluorescent antibodies directed to either E1 or the HA tag. FACS was used to isolate E1-reactive cells (E1 and HA positive) and/or HA-displaying (HA positive) cells. Deep sequencing [38] was used to determine the enrichment or depletion of each mutation in E1-reactive cells compared to HA-displaying cells. The relative E1 reactivity of a panel of individual ubiquitin variants was independently determined relative to wild-type ubiquitin using purified proteins. The accumulation pattern of His_6-ubiquitin variants in yeast harboring untagged wild-type ubiquitin was monitored by inducing expression of the epitope-tagged variant followed by Western blotting.

Expression and purification of E1 (Uba1)

The yeast E1 (Uba1) open reading frame was cloned with a biotin ligase acceptor peptide [75,76] encoded at the far C-terminus into a pAC-T7, an expression vector with a T7 promoter and a chloramphenicol resistance marker. This expression plasmid was co-transformed into BLR(DE3) *Escherichia coli* together with pET24-birA to co-express biotin ligase. Cells were grown at 37 °C in 2× YT media to an OD_{600} of 0.8. Cells were then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside at 25 ° C for 6 h, harvested by centrifugation, and resuspended in IMAC binding buffer [20 mM potassium phosphate

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

(pH 7.2), 300 mM sodium chloride, and 10 mM imidazole]. Bacterial pellets were lysed with a combination of lysozyme, DNase I, and sonication in the presence of 1 mM phenylmethanesulfonyl fluoride to inhibit proteolysis. Biotinylated Uba1 was then purified by cobalt immobilized metal affinity chromatography followed by anion-exchange chromatography. Active E1 concentration was estimated by titration with purified wild-type ubiquitin and was routinely 20–40% of the E1 concentration estimated by absorbance at 280 nm. E1 concentrations based on absorbance at 280 nm were more precise and were used throughout the text and figures.

Yeast surface display of ubiquitin point mutants

Systematic libraries of ubiquitin point mutants were generated in the pCTCON2 yeast-display plasmid [73] with a galactose-dependent promoter driving a fusion of Aga2 with HA followed by a glycine-rich linker and ubiquitin with its native C-terminus. Libraries of ubiquitin point mutants were generated in eight pools. Each pool contained mutants in 9–10 consecutive amino acids as previously described [27]. Ubiquitin mutant libraries were transferred into pCTCON2 using sequence and ligation-independent cloning (SLIC) [77]. To facilitate transfer, we constructed a modified pCTCON2 destination plasmid with the sequence GCTAGCGATTCTAGAACTAGTAATATGCATGCTC GAGTCATGTAATTAGTTAGGATCC immediately following the HA tag and glycine-rich sequence in pCTCON2. This vector was prepared for SLIC by digestion with SphI and treatment with T4 DNA polymerase as previously described [77]. SLIC inserts were prepared by 8 cycles of PCR with previously described [27] ubiquitin libraries in p427GPD as template and forward (GATTCTAGAAC TAGTAATATG) and reverse (TAACTAATTACATGACTC GAG) primers that bind immediately upstream and downstream of the ubiquitin open reading frame in this template, followed by treatment with T4 DNA polymerase as previously described [77]. After annealing of prepared vector and inserts, we transformed samples into competent bacteria and prepared plasmid libraries in bulk as previously described [38]. The library generation procedures were developed to maximize the fraction of the library with relevant point mutations and minimize chances for secondary mutations, especially those outside of the regions directly sequenced and hence undetectable to our analyses. The starting ubiquitin libraries [27] were generated using a cassette ligation strategy such that all regions outside of those directly sequenced were copied entirely in bacteria where fidelity should virtually eliminate the probability of secondary mutations. In transferring the libraries to the display plasmid, we performed 8 cycles of PCR using Pfusion DNA polymerase (New England Biolabs). According to the manufacturer's estimated error rate for this polymerase (4.4×10^{-7}) and the amplification details, we estimate that less than 1 in 10,000 molecules would have a secondary mutation outside the region that we sequence.

Pooled mutant libraries of each region were transformed separately into the EBY100 yeast-display strain [77] as previously described [27]. Following plasmid transformation, we pelleted yeast cells and washed them three times in 1× Tris-buffered saline (TBS) to remove extracellular plasmid. Each pellet was then resuspended in 50 mL of synthetic dextrose (SD) media lacking tryptophan and uracil to select for transformed cells. Cells were grown for 48 h (to an OD_{600} of about 1) at 30 °C in a shaking incubator. Aliquots of approximately 10^8 cells were collected for each library and stored in 20% glycerol at −80 °C. Aliquots for each library region were thawed and used to inoculate 50 mL of casamino acid and dextrose (CAAD) media. These cultures were grown at 30 °C to near saturation for 24 h, and then diluted 50-fold into 50 mL of fresh CAAD. Yeast proliferation was then monitored by OD_{600} reading and kept in mid-log growth by dilution with fresh CAAD for 16 h. Cells in log phase were collected by centrifugation and washed 3 times with CAA-RG media (casamino acids media with 1% raffinose and 1% galactose), resuspended in CAA-RG media to an OD_{600} of 0.5, and grown at 30 °C for a further 16 h. As a control for non-displaying cells, cultures were also grown in CAAD to repress expression from the gal-inducible promoter.

Labeling and sorting of yeast-display cells

For each ubiquitin region, a sample of 10^7 display cells were collected in a microfuge tube, washed twice with TBS, and resuspended in 100 μL of TBS. A 2× E1 reaction mixture was made in a separate tube [50 mM Tris (pH 7.5), 5 mM magnesium chloride, 2.5 mM ATP, and 200 nM total E1 enzyme for limiting conditions or 2000 nM total E1 for excess conditions]. We mixed 100 μL of E1 reaction mixture with cells and incubated it at room temperature for 1 min, and the reaction was quenched with an excess of free ubiquitin. Following reaction, we washed yeast cells twice with 500 μL of TBS containing 0.1% bovine serum albumin (TBSB). Cells were resuspended in 100 μL of TBSB and incubated for 30 min on ice with a 1:100 dilution of both α-HA rabbit polyclonal (Abcam 13834-100) and mouse monoclonal α-biotin (Jackson Immuno Research 200-002-211) antibodies. Cells were then collected by centrifugation, washed twice with TBSB, and incubated in a 100-μL volume on ice with α-rabbit-IgG-FITC and α-mouse-IgG-phycoerytherin (Sigma F0382 and P9287). Labeled cells were diluted to 10^6 cells/mL and transferred to polystyrene FACS tubes.

Labeled samples were sorted for display efficiency and E1 reactivity on a BD FACSVantage DV-1 cell sorter by collecting all FITC-positive cells as one population (HA-display positive), followed by double-positive FITC + phycoerytherin cells (HA-display positive and E1 reactive). To ensure adequate library coverage, we sorted at least 150,000 cells of each population and collected them into sterile SD media. Sorted yeast cells were amplified in 50 mL of SD-U-W media (display off) for 24 h at 30 °C to an OD_{600} of approximately 1. These yeast samples were collected by centrifugation and washed with TBS, and cell pellets were stored at −80 °C.

Quantifying mutant responses to selection by sequencing

Plasmid DNA from yeast pellets was prepared for deep sequencing as previously described [27,38]. Briefly, plasmid DNA was isolated from yeast and the display ubiquitin open reading frame was amplified with primers specific to the pCTCON2 promoter and terminator regions. A second PCR step was used to focus on the randomized region of each library, including addition of an MmeI site adjacent to the mutated region. Three-base barcodes each

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

differing by at least two bases were ligated to MmeI-digested samples to differentiate between unsorted cells, HA-positive-displaying cells, and double-positive E1- and HA-displaying cells. FastQ files from deep sequencing were analyzed as previously described [38]. Raw counts of each mutation were normalized to the wild-type ubiquitin sequence count. The relative enrichment or depletion of each mutation in E1-reactive cells to HA-displaying cells was calculated in log scale. Because the last amino acid of ubiquitin is strictly required for E1 activation, stop codons at each position should be biochemically null. To normalize for small differences in observed raw enrichment and depletion values for different regions, we linearly scaled the apparent E1 reactivity's of mutations in each region such that the average stop codon was 0 and that the average wild-type synonym was 1. Of note, analyses of unsorted and HA-displaying cells indicated that all mutations (including stop codons and cysteine substitutions throughout ubiquitin) were displayed with similar efficiency. As previously described [27], mutations that were low in abundance in our libraries (mutant:WT ratio less than 2^{-8}) or that introduced an internal MmeI site were omitted from analysis. For wild-type amino acids where synonyms were not available or analyzed, E1 reactivity was set to the average of all wild-type synonyms in the region (1 by definition).

Monoculture growth rate of yeast with individual ubiquitin mutations

Growth rates of yeast supported by ubiquitin variants were determined as previously described [27,41]. Briefly, plasmid (p427GPD) encoded ubiquitin variants driven by a constitutive promoter were transformed into a ubiquitin shutoff strain (Sub328[53]). Growth rates at 30 °C were determined in SD media by following the change in OD_{600} after 12 h of pre-equilibration under shutoff conditions.

Quantification of ubiquitin activation by E1 using purified proteins

We developed a binary competition assay to determine the E1 reactivity of ubiquitin mutants relative to wild type using purified proteins. We generated and purified wild-type ubiquitin with a His_6 tag and a unique cysteine at the N-terminus (MGHHHHHHCGG). Purified protein was reacted with fluorescein iodoacetamide, and fluorescently labeled ubiquitin (FL-UB) further purified by size-exclusion chromatography using a Superdex-200 column. Competition experiments between fluorescently labeled wild-type ubiquitin and unlabeled competitors were setup with 100 nM total E1, 500 nM FL-UB, and a range of competitor concentrations. Reactions were performed at room temperature in E1 reaction buffer [25 mM Tris (pH 7.2), 50 mM sodium chloride, 5 mM magnesium chloride, and 5 mM ATP] supplemented with 0.1 mg/mL bovine serum albumin. After 1 min, reactions were halted by the addition of sodium dodecyl sulfate to 2%. Reaction products were separated on a non-reducing SDS-PAGE and imaged on a fluorescent imager. The intensity of the FL-UB-E1 band was quantified using the program Multigauge (Fuji) and plotted as a function of the concentration of unlabeled competitor. These plots were fit to a kinetic competition model (Supplementary Fig. S4) to estimate relative E1 reactivity. Relative E1 reactivity was log transformed and normalized (wild type set to 1 and the null mutant G75D set to 0) in order to facilitate comparison to reactivity estimates from display experiments.

Analyzing ubiquitin accumulation in yeast

To examine the accumulation profile of ubiquitin mutants in yeast, we generated inducible epitope-tagged ubiquitin constructs. Selected ubiquitin mutants with an N-terminal His_6-Myc epitope were cloned with a galactose-inducible promoter [78] into p427 plasmids. These constructs were transformed into W303 yeast cells that express wild-type ubiquitin from endogenous loci. Following transformation, we grew single colonies to saturation at 30 °C in SD media, and then grew them at 30 °C for 16 h in synthetic media with 2% raffinose to an OD_{600} of 1. At this point, a sample of control (uninduced) cells were collected and frozen at −80 °C. The remaining culture was grown in synthetic media with 1% galactose and 1% raffinose for 2 h at 30 °C. Samples were collected by centrifugation, washed once with TBS, and stored at −80 °C. Frozen samples were lysed by vortexing with glass beads and treatment with 2% SDS buffer with 1 mM phenylmethanesulfonyl fluoride at 95 °C for 5 min. After removing cell debris by centrifugation, we determined the protein concentration in each sample using a BCA assay (Pierce). Samples (20 μg of total protein from each lysate) were analyzed by Western blotting with anti-HisG antibody (Invitrogen 46-1008). Multigauge (Fuji) densitometry software was used to quantify both free and conjugated ubiquitin species.

Structural analyses

Structural analyses were performed with PyMOL (Schrödinger) or Chimera (UCSF), and these programs were also used to generate all molecular images. The average surface area buried at structurally characterized ubiquitin interfaces was calculated from previously published surface area measurements [27].

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.05.019.

Acknowledgements

This work was aided by discussions with R. Hietpas, P. Mishra, L. Jiang, C. R. Matthews, R. Gilmore, and P. Pryciak. We are grateful for assistance from the University of Massachusetts Medical School Core Flow Cytometry Laboratory and R. Konz with cell sorting experiments. This work was supported by the National Institutes of Health (grant number R01-GM083038 to D.N.B.).

Received 10 April 2014;
Received in revised form 13 May 2014;
Accepted 18 May 2014
Available online xxxx

Keywords:
systematic mutagenesis;
E1 activation;
quality filtering;
elasticity function

Abbreviations used:
FACS, fluorescence-activated cell sorting; SLIC, sequence and ligation-independent cloning; TBS, Tris-buffered saline; SD, synthetic dextrose; CAAD, casamino acid and dextrose.

References

[1] Sahni N, Yi S, Zhong Q, Jailkhani N, Charloteaux B, Cusick ME, et al. Edgotype: a fundamental link between genotype and phenotype. Curr Opin Genet Dev 2013;23:649–57.
[2] Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. Cell 2011;144:986–98.
[3] Ideker T, Sharan R. Protein networks in disease. Genome Res 2008;18:644–52.
[4] Stearns FW. One hundred years of pleiotropy: a retrospective. Genetics 2010;186:767–73.
[5] Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, et al. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell 2005;123:507–19.
[6] Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 2001;294:2364–8.
[7] Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, et al. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. Nature 2000;403:623–7.
[8] Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002;415:141–7.
[9] Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 2002;415:180–3.
[10] Vidal M, Brachmann RK, Fattaey A, Harlow E, Boeke JD. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein–protein and DNA–protein interactions. Proc Natl Acad Sci U S A 1996;93:10315–20.
[11] Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, et al. Understanding the function–structure and function–mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proc Natl Acad Sci U S A 2003;100:8424–9.
[12] Dreze M, Charloteaux B, Milstein S, Vidalain PO, Yildirim MA, Zhong Q, et al. "Edgetic" perturbation of a *C. elegans* BCL2 ortholog. Nat Methods 2009;6:843–9.
[13] Kacser H, Fell DA. The control of flux: 21 years on. Biochem Soc Trans 1995;23:341–66.
[14] Jiang L, Mishra P, Hietpas RT, Zeldovich KB, Bolon DN. Latent effects of Hsp90 mutants revealed at reduced expression levels. PLoS Genet 2013;9:e1003600.
[15] Dean AM, Thornton JW. Mechanistic approaches to the study of evolution: the functional synthesis. Nat Rev Genet 2007;8:675–88.
[16] Liberles DA, Teichmann SA, Bahar I, Bastolla U, Bloom J, Bornberg-Bauer E, et al. The interface of protein structure, protein biophysics, and molecular evolution. Protein Sci 2012;21:769–85.
[17] Harms MJ, Thornton JW. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat Rev Genet 2013;14:559–71.
[18] Fraser JS, Gross JD, Krogan NJ. From systems to structure: bridging networks and mechanism. Mol Cell 2013;49:222–31.
[19] Kacser H, Burns JA. The molecular basis of dominance. Genetics 1981;97:639–66.
[20] Zhu G, Golding GB, Dean AM. The selective cause of an ancient adaptation. Science 2005;307:1279–82.
[21] Bridgham JT, Carroll SM, Thornton JW. Evolution of hormone-receptor complexity by molecular exploitation. Science 2006;312:97–101.
[22] Linnen CR, Poh YP, Peterson BK, Barrett RD, Larson JG, Jensen JD, et al. Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 2013;339:1312–6.
[23] Rennell D, Bouvier SE, Hardy LW, Poteete AR. Systematic mutation of bacteriophage T4 lysozyme. J Mol Biol 1991;222:67–88.
[24] Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. Nature 2005;436:588–92.
[25] Eames M, Kortemme T. Cost-benefit tradeoffs in engineered lac operons. Science 2012;336:911–5.
[26] Hietpas RT, Jensen JD, Bolon DN. Experimental illumination of a fitness landscape. Proc Natl Acad Sci U S A 2011;108:7896–901.
[27] Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DN. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J Mol Biol 2013;425:1363–77.
[28] Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence–function relationships. Nat Methods 2010;7:741–6.
[29] McLaughlin RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. Nature 2012;491:138–42.
[30] Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW, Fields S. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. Proc Natl Acad Sci U S A 2012;109:16858–63.
[31] Weinreich DM, Delaney NF, Depristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 2006;312:111–4.
[32] Lunzer M, Miller SP, Felsheim R, Dean AM. The biochemical architecture of an ancient adaptive landscape. Science 2005;310:499–501.
[33] Dean AM, Dykhuizen DE, Hartl DL. Fitness as a function of beta-galactosidase activity in *Escherichia coli*. Genet Res 1986;48:1–8.
[34] Hershko A, Ciechanover A. The ubiquitin system. Annu Rev Biochem 1998;67:425–79.
[35] Haas AL, Warms JV, Hershko A, Rose IA. Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. J Biol Chem 1982;257:2543–8.
[36] Lee I, Schindelin H. Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. Cell 2008;134:268–78.

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

[37] McGrath JP, Jentsch S, Varshavsky A. UBA 1: an essential yeast gene encoding ubiquitin-activating enzyme. EMBO J 1991;10:227–36.
[38] Hietpas R, Roscoe B, Jiang L, Bolon DN. Fitness analyses of all possible point mutations for regions of genes in yeast. Nat Protoc 2012;7:1382–96.
[39] King JL, Jukes TH. Non-Darwinian evolution. Science 1969; 164:788–98.
[40] Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. Curr Opin Struct Biol 2009;19: 596–604.
[41] Haririnia A, Verma R, Purohit N, Twarog MZ, Deshaies RJ, Bolon D, et al. Mutations in the hydrophobic core of ubiquitin differentially affect its recognition by receptor proteins. J Mol Biol 2008;375:979–96.
[42] Gangadhara BN, Laine JM, Kathuria SV, Massi F, Matthews CR. Clusters of branched aliphatic side chains serve as cores of stability in the native state of the HisF TIM barrel protein. J Mol Biol 2013;425:1065–81.
[43] Dill KA. Dominant forces in protein folding. Biochemistry 1990;29:7133–55.
[44] Lee SY, Pullen L, Virgil DJ, Castaneda CA, Abeykoon D, Bolon DN, et al. Alanine scan of core positions in ubiquitin reveals links between dynamics, stability, and function. J Mol Biol 2014;426:1377–89.
[45] Matyskiela ME, Lander GC, Martin A. Conformational switching of the 26S proteasome enables substrate degradation. Nat Struct Mol Biol 2013;20:781–8.
[46] Phillips AH, Zhang Y, Cunningham CN, Zhou L, Forrest WF, Liu PS, et al. Conformational dynamics control ubiquitin-deubiquitinase interactions and influence *in vivo* signaling. Proc Natl Acad Sci U S A 2013;110: 11379–84.
[47] Zhao B, Bhuripanyo K, Schneider J, Zhang K, Schindelin H, Boone D, et al. Specificity of the E1-E2-E3 enzymatic cascade for ubiquitin C-terminal sequences identified by phage display. ACS Chem Biol 2012;7:2027–35.
[48] Patel LR, Curran T, Kerppola TK. Energy transfer analysis of Fos-Jun dimerization and DNA binding. Proc Natl Acad Sci U S A 1994;91:7360–4.
[49] Walden H, Podgorski MS, Schulman BA. Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. Nature 2003;422:330–4.
[50] Whitby FG, Xia G, Pickart CM, Hill CP. Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes. J Biol Chem 1998;273: 34983–91.
[51] Burch TJ, Haas AL. Site-directed mutagenesis of ubiquitin. Differential roles for arginine in the interaction with ubiquitin-activating enzyme. Biochemistry 1994;33: 7300–8.
[52] Pruneda JN, Littlefield PJ, Soss SE, Nordquist KA, Chazin WJ, Brzovic PS, et al. Structure of an E3:E2 ~ Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. Mol Cell 2012;47:933–42.
[53] Spence J, Sadis S, Haas AL, Finley D. A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. Mol Cell Biol 1995;15:1265–73.
[54] Swanson KA, Kang RS, Stamenova SD, Hicke L, Radhakrishnan I. Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. EMBO J 2003;22:4597–606.
[55] Peschard P, Kozlov G, Lin T, Mirza IA, Berghuis AM, Lipkowitz S, et al. Structural basis for ubiquitin-mediated dimerization and activation of the ubiquitin protein ligase Cbl-b. Mol Cell 2007;27:474–85.
[56] Hicke L, Schubert HL, Hill CP. Ubiquitin-binding domains. Nat Rev Mol Cell Biol 2005;6:610–21.
[57] Beal R, Deveraux Q, Xia G, Rechsteiner M, Pickart C. Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. Proc Natl Acad Sci U S A 1996;93:861–6.
[58] Lois LM, Lima CD. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. EMBO J 2005;24:439–51.
[59] Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. Science 1995;267:383–6.
[60] Thrower JS, Hoffman L, Rechsteiner M, Pickart CM. Recognition of the polyubiquitin proteolytic signal. EMBO J 2000;19:94–102.
[61] Hendsch ZS, Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci 1994;3: 211–26.
[62] Drummond DA, Wilke CO. The evolutionary consequences of erroneous protein synthesis. Nat Rev Genet 2009;10:715–24.
[63] Khoury GA, Baliban RC, Floudas CA. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep 2011;1.
[64] Kramer EB, Vallabhaneni H, Mayer LM, Farabaugh PJ. A comprehensive analysis of translational missense errors in the yeast *Saccharomyces cerevisiae*. RNA 2010;16: 1797–808.
[65] Tsai IJ, Bensasson D, Burt A, Koufopanou V. Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. Proc Natl Acad Sci U S A 2008;105:4957–62.
[66] Ohta T. Slightly deleterious mutant substitutions in evolution. Nature 1973;246:96–8.
[67] Jones S, Thornton JM. Analysis of protein–protein interaction sites using surface patches. J Mol Biol 1997;272: 121–32.
[68] Zuckerkandl E. Evolutionary processes and evolutionary noise at the molecular level. I. Functional density in proteins. J Mol Evol 1976;7:167–83.
[69] Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proc Natl Acad Sci U S A 2011;108:680–5.
[70] Joshi KK, Chen L, Torres N, Tournier V, Madura K. A proteasome assembly defect in rpn3 mutants is associated with Rpn11 instability and increased sensitivity to stress. J Mol Biol 2011;410:383–99.
[71] Yao T, Cohen RE. A cryptic protease couples deubiquitination and degradation by the proteasome. Nature 2002;419:403–7.
[72] Verma R, Aravind L, Oania R, McDonald WH, Yates JR, Koonin EV, et al. Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. Science 2002;298:611–5.
[73] Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. Nat Biotechnol 1997;15:553–7.
[74] Shusta EV, Holler PD, Kieke MC, Kranz DM, Wittrup KD. Directed evolution of a stable scaffold for T-cell receptor engineering. Nat Biotechnol 2000;18:754–9.

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019

[75] Duffy S, Tsao KL, Waugh DS. Site-specific, enzymatic biotinylation of recombinant proteins in *Spodoptera frugiperda* cells using biotin acceptor peptides. Anal Biochem 1998;262:122–8.

[76] Chen I, Howarth M, Lin W, Ting AY. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. Nat Methods 2005;2:99–104.

[77] Li MZ, Elledge SJ. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. Nat Methods 2007;4:251–6.

[78] Mumberg D, Muller R, Funk M. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res 1994;22:5767–8.

Please cite this article as: Roscoe Benjamin P., Bolon Daniel N.A., Systematic Exploration of Ubiquitin Sequence, E1 Activation Efficiency, and Experimental Fitness in Yeast, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.05.019