

## RESEARCH ARTICLE

# Extending chemical perturbations of the ubiquitin fitness landscape in a classroom setting reveals new constraints on sequence tolerance

David Mavor<sup>1</sup>, Kyle A. Barlow<sup>2</sup>, Daniel Asarnow<sup>1</sup>, Yuliya Birman<sup>1</sup>, Derek Britain<sup>1</sup>, Weilin Chen<sup>2</sup>, Evan M. Green<sup>1</sup>, Lillian R. Kenner<sup>1</sup>, Bruk Mensa<sup>3</sup>, Leanna S. Morinishi<sup>2</sup>, Charlotte A. Nelson<sup>2</sup>, Erin M. Poss<sup>3</sup>, Pooja Suresh<sup>1</sup>, Ruilin Tian<sup>1</sup>, Taylor Arhar<sup>3</sup>, Beatrice E. Ary<sup>3</sup>, David P. Bauer<sup>1</sup>, Ian D. Bergman<sup>3</sup>, Rachel M. Brunetti<sup>1</sup>, Cynthia M. Chio<sup>3</sup>, Shizhong A. Dai<sup>3</sup>, Miles S. Dickinson<sup>3</sup>, Susanna K. Elledge<sup>3</sup>, Cole V. M. Helsell<sup>1</sup>, Nathan L. Hendel<sup>2</sup>, Emily Kang<sup>3</sup>, Nadja Kern<sup>1</sup>, Matvei S. Khoroshkin<sup>2</sup>, Lisa L. Kirkemo<sup>3</sup>, Greyson R. Lewis<sup>1</sup>, Kevin Lou<sup>3</sup>, Wesley M. Marin<sup>2</sup>, Alison M. Maxwell<sup>3</sup>, Peter F. McTigue<sup>3</sup>, Douglas Myers-Turnbull<sup>2</sup>, Tamas L. Nagy<sup>2</sup>, Andrew M. Natale<sup>1</sup>, Keely Olton<sup>3</sup>, Sergei Pourmal<sup>3</sup>, Gabriel K. Reder<sup>1</sup>, Nicholas J. Rettko<sup>3</sup>, Peter J. Rohweder<sup>3</sup>, Daniel M. C. Schwarz<sup>3</sup>, Sophia K. Tan<sup>1</sup>, Paul V. Thomas<sup>1</sup>, Ryan W. Tibble<sup>3</sup>, Jason P. Town<sup>2</sup>, Mary K. Tsai<sup>3</sup>, Fatima S. Ugur<sup>3</sup>, Douglas R. Wassarman<sup>3</sup>, Alexander M. Wolff<sup>1</sup>, Taia S. Wu<sup>3</sup>, Derek Bogdanoff<sup>4</sup>, Jennifer Li<sup>5</sup>, Kurt S. Thorn<sup>4</sup>, Shane O'Conchúir<sup>6</sup>, Danielle L. Swaney<sup>7</sup>, Eric D. Chow<sup>4</sup>, Hiten D. Madhani<sup>4</sup>, Sy Redding<sup>4</sup>, Daniel N. Bolon<sup>8</sup>, Tanja Kortemme<sup>6</sup>, Joseph L. DeRisi<sup>4</sup>, Martin Kampmann<sup>4,9,\*</sup> and James S. Fraser<sup>6,\*</sup>

## ABSTRACT

Although the primary protein sequence of ubiquitin (Ub) is extremely stable over evolutionary time, it is highly tolerant to mutation during selection experiments performed in the laboratory. We have proposed that this discrepancy results from the difference between fitness under laboratory culture conditions and the selective pressures in changing environments over evolutionary timescales. Building on our previous work (Mavor et al., 2016), we used deep mutational scanning to determine how twelve new chemicals (3-Amino-1,2,4-triazole, 5-fluorocytosine, Amphotericin B, CaCl<sub>2</sub>, Cerulenin, Cobalt Acetate, Menadione, Nickel Chloride, p-Fluorophenylalanine, Rapamycin, Tamoxifen, and Tunicamycin) reveal novel mutational sensitivities of ubiquitin residues. Collectively, our experiments have identified eight new sensitizing conditions for Lys63 and uncovered a sensitizing condition for every position in Ub except Ser57 and Gln62. By determining the ubiquitin fitness landscape under different chemical constraints, our work helps to resolve the inconsistencies between deep mutational scanning experiments and sequence conservation over evolutionary timescales.

**KEY WORDS:** Deep mutational scanning, Evolution, Ubiquitin

## INTRODUCTION

The increased capabilities of deep sequencing technologies have transformed our ability to interrogate pooled libraries of variants under selection or screening conditions (Fowler and Fields, 2014). In particular, protein sequence-structure-function studies are benefiting from the ability to comprehensively survey the functional effects of all possible single point mutants in experiments that have come to be called 'deep mutational scans' (Araya and Fowler, 2011; McLaughlin et al., 2012). Deep mutational scanning experiments are revealing new dimensions of protein stability (Araya et al., 2012), substrate specificity (Shah et al., 2018; Wrenbeck et al., 2017) and regulation (Bandaru et al., 2017). These experiments also provide insight into the evolutionary significance of the spectrum of mutational effects on fitness. For example, recent studies have also used deep mutational scanning to probe how mutations are tolerated in different sequence backgrounds, reflecting the local sequence space tolerated over long evolutionary timescales (Starr et al., 2017, 2018). These experiments rely on the connection between the character of the laboratory selection (or screen) and the pressures experienced by populations of organisms in the natural environment. Indeed, comparing phylogenetic analysis of naturally occurring variation between homologs with deep mutational scanning data of a single protein can reveal sites that are experiencing different selective pressures in nature versus the laboratory (Hilton et al., 2017). Across many studies, a general trend has emerged with the expected general correlation between sites that are poorly conserved in evolution tolerating more substitutions more readily in deep mutational scanning experiments and with highly conserved sites being less tolerant to substitutions.

One interesting contrast to the general trends between evolutionary sequence conservation and deep mutational scanning tolerance to substitution is the protein ubiquitin (Ub), an essential eukaryotic protein that acts as post-translational modification to mediate the degradation of ~80% of the proteome (Yau and Rape, 2016) and is also one of the first proteins subjected to a yeast-based

<sup>1</sup>Biophysics Graduate Group, University of California, San Francisco 94158, USA. <sup>2</sup>Bioinformatics Graduate Group, University of California, San Francisco 94158, USA. <sup>3</sup>Chemistry and Chemical Biology Graduate Program, University of California, San Francisco 94158, USA. <sup>4</sup>Department of Biochemistry and Biophysics, University of California, San Francisco 94158, USA. <sup>5</sup>Department of Chemistry Undergraduate Program, University of California, Davis 95616, USA. <sup>6</sup>Department of Bioengineering and Therapeutic Sciences, California Institute for Quantitative Biology (QBI), San Francisco 94158, USA. <sup>7</sup>Department of Cellular and Molecular Pharmacology, California Institute for Quantitative Biology (QBI), San Francisco 94158, USA. <sup>8</sup>Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester 01655, USA. <sup>9</sup>Institute for Neurodegenerative Diseases, University of California, San Francisco 94158, USA.

\*Authors for correspondence (martin.kampmann@ucsf.edu; jfraser@fraserlab.com)

 J.S.F., 0000-0002-5080-2859

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

deep mutational scanning experiment (Roscoe et al., 2013). The amino acid sequence of ubiquitin has been strikingly stable throughout evolutionary time: between yeast and humans, there are only three amino acid changes (96% sequence identity) (Finley et al., 2012). However, deep mutational scanning experiments in yeast have revealed that Ub is surprisingly robust to sequence changes, with 19 positions freely mutating to almost any other amino acid without a loss of fitness (Roscoe et al., 2013). Some of this pattern could be rationalized structurally: a sensitive structural surface is the known interface for many binding partners, whereas the positions that are tolerant to mutation make few contacts with structurally characterized binding partners (Roscoe et al., 2013). However, the outstanding disconnect between strong constraints during natural evolution and tolerance during laboratory selection remained unaddressed by the original study.

To interrogate the dichotomy between the strong sequence conservation and the mutational robustness of Ub, we initially hypothesized that sensitivities to mutations at new positions could be revealed by growing yeast under different selective pressures. To test this idea we previously determined, in a classroom setting, the fitness landscape of ubiquitin in four different chemical perturbations [DTT, caffeine, hydroxyurea (HU), and MG132] (Mavor et al., 2016). We showed that three of the perturbations (DTT, caffeine and HU) sensitize a shared set of positions to mutation, including several positions that were not sensitive to mutation under the standard growth conditions employed previously. Conversely, we showed that the proteasome inhibitor MG132 increases the mutational robustness of the ubiquitin sequence landscape. Inhibiting the proteasome reduces protein turnover through the same pathway as mutations in ubiquitin, leading to an alleviating interaction between MG132 and many of the mutant alleles. A major conclusion from this study was that the fitness defects, relevant for rationalizing evolutionary patterns, were buffered and undetectable in standard laboratory growth, but that these defects could be unmasked by simple chemical stresses. However, 12 of the 19 residues, the residues classified as ‘tolerant’ [almost all mutations at that position have near wild-type (WT) fitness] in standard growth conditions, were still tolerant under all chemical stresses.

To identify potential environmental perturbations that could help to rationalize the constraint on the ‘tolerant’ residues over evolutionary time, we again involved the first-year graduate students in UCSF’s iPQB and CCB programs to determine the fitness landscape of ubiquitin in distinct environments. We chose twelve new chemical perturbations [3-Amino-1,2,4-triazole (3-AT), 5-fluorocytosine (5-FC), Amphotericin B (AmpB), CaCl<sub>2</sub>, Cerulenin, Cobalt Acetate (Cobalt), Menadione, Nickel Chloride (Nickel), p-Fluorophenylalanine (p-FP), Rapamycin, Tamoxifen, and Tunicamycin], which were expected to impose a wide range of stresses upon the cell, including osmotic shock, protein folding stress and DNA damage. By using these additional stresses, we can now identify laboratory conditions that place strong constraints on the sequence preferences of all but two residues in ubiquitin. Our results represent an important next step towards how deep mutational scanning can be used to explain the evolutionary constraints on sequence conservation patterns.

## RESULTS

### Distinct chemical treatments can sensitize or increase robustness of Ub to mutation

As in our previous work, we performed deep mutational scanning experiments of a barcoded Ub library in the presence of distinct

chemical perturbations at concentrations that inhibited the growth of a strain expressing wild-type ubiquitin by 25% (Mavor et al., 2016). To quantify the effect of the chemical on growth of specific mutations, we subtracted the fitness values of our control (DMSO) dataset and generated difference fitness maps (Fig. 1). The most obvious global trends are the increased sensitization of many residues when treated with AmpB (Fig. 1L) and increased robustness of many residues when treated with Tamoxifen (Fig. 1K). In contrast, treatment with Menadione leads to a very similar pattern of fitness effects to control treatments (Fig. 1I).

To place these results in context of our previous chemical treatments (Mavor et al., 2016), we employed hierarchical clustering based on the pattern of fitness effects across each chemical treatment (Fig. 2). The treatments with Cobalt and p-FP form a new cluster near the previously described ‘sensitizing’ treatments DTT, Caffeine and Hydroxyurea (Fig. 2). These treatments share a sensitizing effect at positions near hydrophobic patch residues (8, 44, 70) and the C-terminus (Fig. 1). In contrast, treatment with Cerulenin, Menadione, Tunicamycin, and 5-FC clustered near to treatment with DMSO and MG132 (Fig. 2). These treatments are mild with many positions displaying mildly increased robustness to mutation and a few distinct mutations for each condition displaying stronger sensitization (Fig. 1).

Treatment with Nickel, 3-AT, or CaCl<sub>2</sub> form a second novel cluster (Fig. 2). These treatments share a dominant pattern of sensitizing positions 35, 46 and 63 coupled with increased robustness at position 58 (Fig. 1). Three treatments demonstrated more idiosyncratic responses (AmpB, Rapamycin, and Tamoxifen) and do not cluster with other treatments (Fig. 1). These treatments induced either extreme sensitization of residues (AmpB), extreme increased robustness of residues (Tamoxifen), or positional dependent, but strong, sensitization and robustness (Rapamycin) (Fig. 2). Collectively, these results demonstrate that many of our new perturbations tap into similar constraints on tolerated sequence space for ubiquitin as our previous study, but also that many of our new perturbations likely unmask distinct constraints.

### Deep mutational scanning in different chemical environments reveals constraints on most residues

To examine whether the new perturbations could help explain the high sequence conservation of Ub, we calculated the average fitness at each position for each condition. At each position, we used the fitness value from the condition with the lowest average fitness value and classified these minimum values based on the previous schemes (Mavor et al., 2016; Roscoe et al., 2013) as either sensitive ( $\leq -0.35$ ), intermediate ( $-0.35$  to  $-0.075$ ) and tolerant ( $\geq -0.075$ ) (Fig. 3). Previously we showed that twelve positions in Ub remained tolerant under the four different chemical stresses (Mavor et al., 2016). By expanding the number of perturbations, we now find that all but two positions, Ser57 and Gln62, are sensitive or intermediate in at least one condition. Although there is significant overlap in response of the Ub fitness landscape to these different perturbations, this result suggests that further exploration of chemical space might unmask constraints on the two residues for which a sensitizing condition has not yet been identified.

### Principal component analysis of deep mutational scanning data across chemical perturbations

To explore whether correlated patterns of fitness values across treatment conditions could provide mechanistic insight into the Ub sequence-structure-function relationship, we performed principal component analysis on the difference fitness data (Fig. 4).





3

**Fig. 1. The difference in fitness between DMSO and a perturbation for each Ub allele.** Chemical names are colored based on the hierarchical clustering presented in Fig. 1: (A) Cobalt, (B) p-FP, (C) Nickel, (D) 3-AT, (E)  $\text{CaCl}_2$ , (F) 5-FC, (G) Tunicamycin, (H) Cerulenin, (I) Menadione, (J) Rapamycin, (K) Tamoxifen, (L) AmpB. Difference in fitness is represented from 0.25 (Blue) to  $-0.25$  (Red) with white representing no change from DMSO. Wild-type amino acids are shown in green and mutations without fitness values (due to lack of barcode or competition sequencing reads) are shown in grey.

We focused our analysis on the first three principal components, which collectively explain 60 percent of the variance (Fig. S1). Projecting the treatments onto the first two principal components reveals two main clusters in this space that parallel many aspects of the hierarchical clustering performed earlier (Fig. 2). Consistent with the hierarchical clustering, treatment with AmpB, Rapamycin, or Tamoxifen appear as outliers (Fig. 4A).

Next, we mapped the contribution of each mutation to each of the first three principal components (PCs) by primary sequence (Fig. 4B) and three dimensional structure (Fig. 4C). PC1 is dominated by mild positive contributions for most mutations, with the strongest positive signals appearing at residues 11, 27, 40 and 41. Interestingly, the strongest negative contributions appear at Phe45, a large core residue that may be involved in long-range correlated motions that are important for recognition of Ub by interacting proteins (Fenwick et al., 2011). PC2 is most similar to the initial description of the fitness landscape in rich media (Roscoe et al., 2013): both are dominated by the contrast between positive contributions from the tolerant face of Ub and the negative contributions from the sensitive face, which includes the ‘hydrophobic patch’ that forms the interface for most interacting proteins.

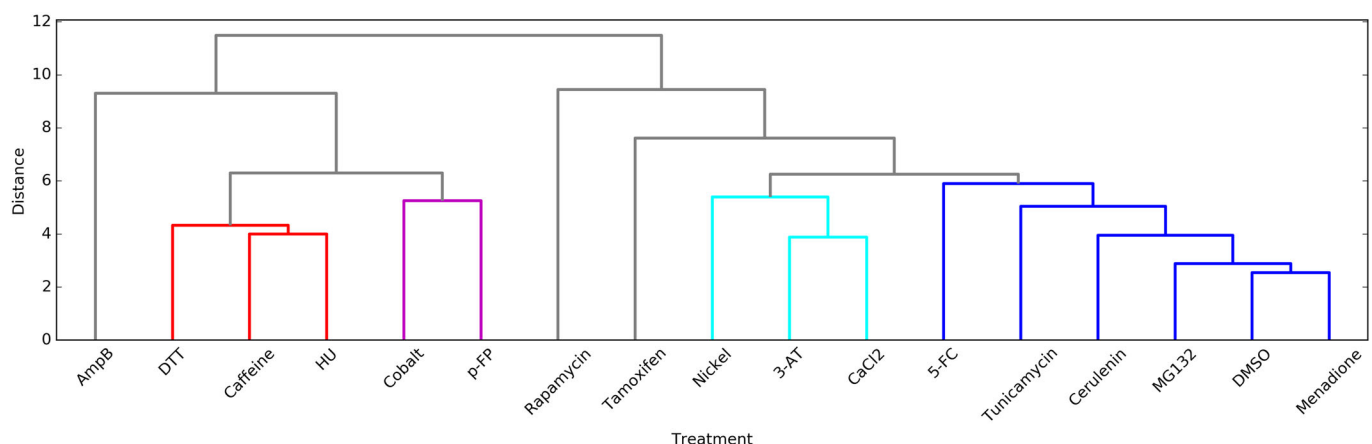
The most notable feature of PC3 is the response to mutation at Lys63, a key poly-Ub linkage site (Fig. 3B). In yeast, Lys63-linked poly-Ub is an important regulator of the DNA damage response and efficient intracellular cargo trafficking (Erpapazoglou et al., 2014). Since the other mutations with strong signals exposed by PC3 were not near K63 structurally (Fig. 3C), we investigated whether the pattern could be rationalized by examining structural complexes important for K63 linkages. We used the molecular modeling program Rosetta (Alford et al., 2017) to calculate the expected change in free energy of each mutation in various complexes involved in Lys63 linked poly-Ub assembly: the closed and open forms of Lys63 linked di-Ub (PDB ID: 2N2K (Liu et al., 2015) and

3H7P (Weeks et al., 2009)) and the donor and acceptor ubiquitin poses on the MMS/Ubc13 complex (PDB ID: 2GMI (Eddins et al., 2006)) (Fig. 3D; Fig. S2). Of these, only the MMS/Ubc13 donor Ub complex revealed any pattern correlated with PC3 (Fig. S2). Several positions predicted to destabilize this interface (including mutations at Lys11 and Pro37) have positive contributions to PC3 (Fig. 3D). This result suggests that conditions that are sensitized when K63-linked poly-Ub chains are compromised might have an increased relative fitness when positions that destabilize the donor ubiquitin pose are mutated. Consistent with this hypothesis is a mutational pattern observed for Lys11, which participates in a salt bridge with Glu65 of Ubc13. The Lys11Arg mutation, which is predicted to maintain the salt bridge and is the only mutation at position 11 that is predicted to stabilize the interface, has a negative contribution to PC3. Collectively, these results suggest that deep mutational scans from multiple chemical perturbations might reveal correlated responses that are difficult to uncover when analyzing only a single condition.

## DISCUSSION

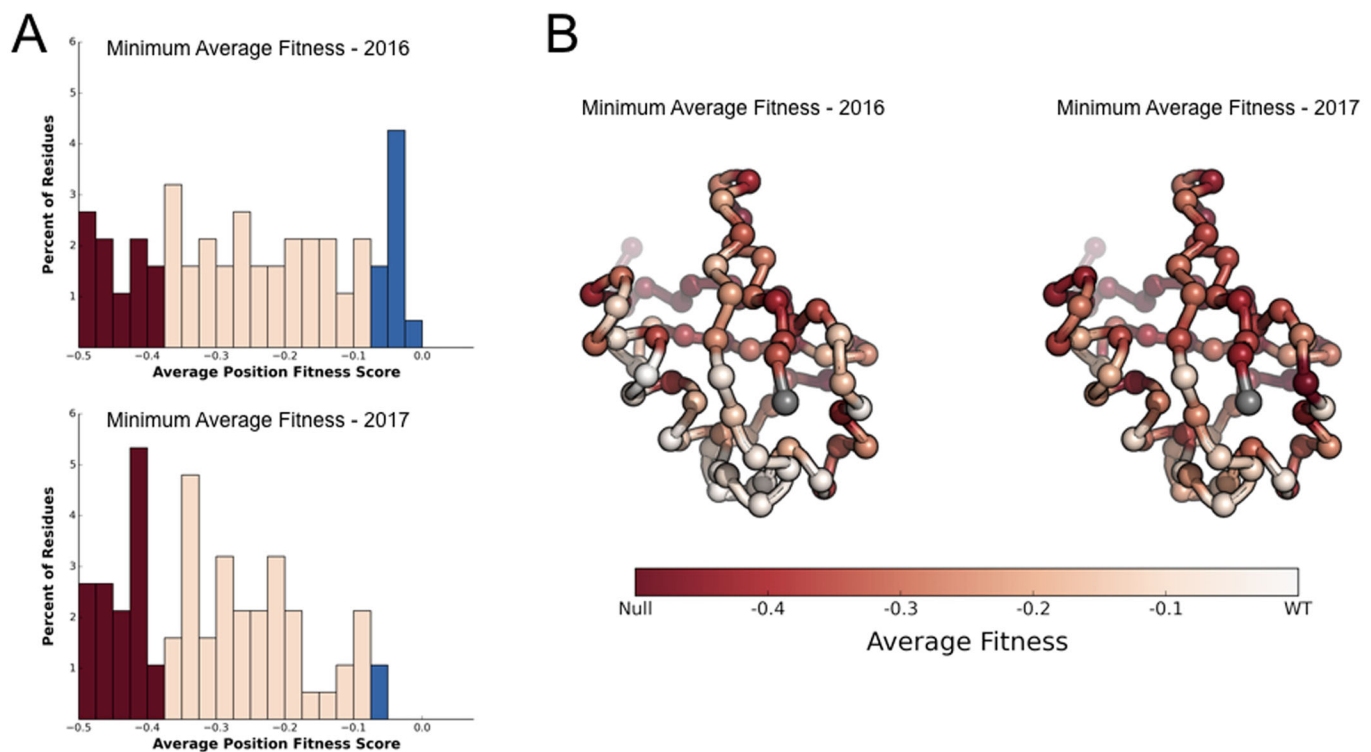
No single perturbation in the laboratory can easily replicate the diverse pressures that natively constrain protein evolution. However, in the case of ubiquitin, we can now rationalize the extreme sequence conservation of Ub after examining the fitness landscape under a large variety of conditions that included redox stress, osmotic stress, protein folding stress, DNA damage, ER stress, and anti-fungals. Notable exceptions are residues Ser57 and Gln62, which are not sensitive to mutation under any condition yet tested.

Of the newly revealed sensitivities, perhaps the most interesting is the sensitization of Lys63. This sensitivity to mutation dominated the third principal component (PC3) and is present in eight conditions. Traditionally, Lys63-linked poly-Ub is thought to participate in the response to DNA damage, where Lys63-linked poly-Ub chains form on PCNA to induce error-free postreplication repair (Zhang et al., 2011), and in endocytosis, where efficient endocytosis in cargo sorting to the vacuole requires Lys63-linked poly-Ub chains (Erpapazoglou et al., 2014). More recent studies (Kwon and Ciechanover, 2017; Silva et al., 2015) have shown that Lys63 chains are involved in the yeast response to oxidative stress and autophagy in metazoans, suggesting that the role of Lys63-linked poly-Ub chains may be more extensive than its previously recognized role in DNA damage and endocytosis.



**Fig. 2. Hierarchical clustering of the fitnesses reveals four distinct clusters.** Treatment with Cobalt and p-FP (magenta) cluster together and close to the previously described ‘sensitizing treatments’ (Mavor et al., 2016), DTT, Caffeine and HU (red). Treatment with Menadione, Cerulenin, Tunicamycin and 5-FC cluster with DMSO and the previously described ‘alleviating treatment’ MG132 (blue).





**Fig. 3. New perturbations reveal constraints on all but two Ub positions.** (A) The minimum average fitness of each position was calculated in: (top) DMSO, Caffeine, DTT, HU and MG132 and (bottom) in all conditions. Minimum average fitness was determined by calculating the average fitness of each position in each condition and taking the minimum value. Positions were binned into tolerant ( $\geq -0.075$  - Blue), intermediate ( $< -0.075$  to  $> -0.35$  - Pink) and sensitive ( $\leq -0.35$  - Red) and the distributions plotted. Calculating the minimum average fitness reveals how the new perturbations reveal additional constraints on the Ub fitness landscape. (B) Minimum average fitness score in: (left) DMSO, Caffeine, DTT, HU, and MG132 and (right) in all conditions mapped onto the Ub structure. C-alpha atoms are shown in spheres and the residues are colored according to average fitness. Met1 is colored grey. Treatment with Nickel, 3-AT and  $\text{CaCl}_2$  cluster together (cyan) and close to the 'alleviating treatment' cluster. Treatment with AmpB, Rapamycin and Tamoxifen appear as outliers in this clustering (grey). The clustering was performed using euclidean distance between the vectors and used Ward's method to join the clusters. Clusters are colored based on the treatments being within 6 distance of each other.

In contrast, we previously observed an increase in mutational robustness at Lys63 in DTT treatment, a reducing agent that interferes with ER protein folding. Interestingly, we also observed increased robustness under Tunicamycin treatment, a compound that interferes with ER protein folding via a distinct mechanism (Chawla et al., 2011). This result suggests an epistatic interaction between Lys63 signaling and the unfolded protein response, which may complement the suggested role of Lys11 under high (30 mM) DTT treatment (Xu et al., 2009). The Lys11Arg mutant is specifically sensitized in Tunicamycin suggesting that the origin of this effect may be structural, rather than due to a requirement for Lys11-linked poly-Ub.

In addition to the increased robustness at Lys63, Tunicamycin treatment leads to a unique increase in mutational robustness at several other positions, including Lys6, Lys11, and Lys33. These results address a major challenge in Ub biology: understanding the biological role of distinct poly-Ub species. While the mutational tolerance pattern at Lys6 and Lys11 appear to be due to disrupting a salt bridge, the increased robustness at Lys33 suggests a connection between Tunicamycin and Lys33 linked poly-Ub. We observed, further, but less conclusive, Lysine-specific effects for Lys27, Lys29, and Lys33 under treatment with AmpB, Cobalt, or Nickel.

Finally, these experiments continue to highlight the success of project-based courses. Building on our first effort (Mavor et al., 2016), we improved on our model: over the course of 6 weeks, first year graduate students in UCSF's CCB and iQCB programs generated and analyzed these data using their own computational pipelines. We believe that yeast-based deep mutational scanning

experiments present ideal systems for such project-based courses due to the low cost and wide range of stress responses accessible by readily purchasable and common chemicals. It is our hope that other graduate programs can offer similar project based classes in the future and we have made our reagents, code, and course material available to further that goal.

## MATERIALS AND METHODS

### Additional material is available

PUBS website ([www.fraserlab.com/pubs](http://www.fraserlab.com/pubs)).

GitHub (<https://github.com/fraser-lab/PUBS>).

Raw Sequencing reads are available via SRA (SRA Accession Number: SRP070953).

### Updated methods from Mavor et al. (2016)

For each compound, we determined the chemical concentrations that inhibited SUB328 (WT Ub) growth by 25% (3-Amino-1,2,4-triazole, 50 mM; 5-fluorocytosine, 1.25  $\mu\text{g/ml}$ ; Amphotericin B, 400 nM;  $\text{CaCl}_2$ , 500 mM; Cerulenin, 4.5  $\mu\text{M}$ ; Cobalt Acetate, 600  $\mu\text{M}$ ; Menadione, 500  $\mu\text{M}$ ; Nickel Chloride, 400  $\mu\text{M}$ ; p-Fluorophenylalanine, 800  $\mu\text{g/ml}$ ; Rapamycin, 200 nM; Tamoxifen, 25  $\mu\text{M}$ ; and Tunicamycin, 1 mg/ml). Other growth, sequencing and data processing methods are unchanged. All datasets, excepting Rapamycin, were collected in duplicate and the average fitnesses are presented.

### Hierarchical clustering

Clustering was performed using scipy (version 0.17.0) in Python with the following parameters:

```
scipy.cluster.hierarchy.linkage(method='ward')
```



**Fig. 4. Principal component analysis reveals specific signals related to K63 incorporation.**

(A) The first two principal components reveal a sensitizing cluster with negative values in each PC and an alleviating cluster in the center of the plot. Treatment with 3-AT, 5-FC or CaCl<sub>2</sub> appear between clusters with positive values in PC1 and negative values in PC2. The points are colored based on the hierarchical clustering shown in Fig. 1. (B) The contribution of each mutation to each principal component was visualized as a heat map. The percentage of the maximum contribution to that principal component is represented from 75% (Blue) to -75% (Red). PC1 is related to the general sensitivity of Ub mutants to perturbation. Regions with large positive contributions to PC1 correspond to the regions with increased mutational sensitivity in the sensitizing treatments. Strikingly this is coupled to negative contributions to PC1 for some mutations at the core residue Phe45 (top). PC2 differentiates 'Sensitive Face' residues (positive contributions) from 'Tolerant Face' residues (Negative Contributions) (middle). PC3 reveals mutations that are correlated with sensitization of Lys63 (bottom). (C) The average contribution of each mutation to a PC at each position was plotted from 75% (Blue) to -75% (Red) on the Ub monomer structure for PC1 (A), PC2 (B), and PC3 (C). (D) Rosetta  $\Delta\Delta G$  calculations revealed that mutations that strongly destabilize the donor Ub pose on the MMS/Ubc13 (2GMI) heterodimer are localized to Lys11 and Pro37 (shown in sticks). In the case of Lys11, all mutations other than that to Arg destabilize the interface suggesting a salt bridge between Lys11 and Glu65 of Ubc13 (shown in green sticks) ubiquitin residues are colored by the contribution to PC3 as in panel C.

**ROSETTA ddG predictions**

Interface ddG predictions were generated using the Rosetta macromolecular modeling suite, which is freely available for academic use. The git version used was 12e38402d9. For each amino acid position in the MMS/Ubc13 heterodimer, the interface ddG protocol was run as follows: (1) minimize (with constraints to the starting coordinates) the starting wild-type structure (PDB ID: 2GMI). (2) Generate an ensemble of 50 conformational states using Rosetta's backrub application (10,000 trials, temperature 1.2), using residues in an 8 Å radius of the specified amino acid position as backrub pivot residues. (3) Repack, or repack and mutate the side chains of the specified amino acid and the pivot residues from step 2. (4) Minimize (with constraints) the wild-type and mutant structures generated in step 3. (5) For each structure  $i$  (of 50), we calculate the ddG score as follows:

$$\Delta\Delta G_{bind} = (\Delta G_{MUT\ complex\ i} - \Delta G_{MUT\ partner\ Ai} - \Delta G_{MUT\ partner\ Bi}) - (\Delta G_{WT\ complex\ i} - \Delta G_{WT\ partner\ Ai} - \Delta G_{WT\ partner\ Bi}).$$

We then average all 50  $\Delta\Delta G_{bind}$  scores to obtain the final predicted value.

**Acknowledgements**

We acknowledge: administrative support from Rebecca Brown, Julia Molla, and Nicole Flowers; technical support from Jennifer Mann and Manny De Vera; gifts from David Botstein, and Illumina; and helpful discussions with Nevan Krogan and Ron Vale.

**Competing interests**

The authors declare no competing or financial interests.

**Author contributions**

Conceptualization: D.M., M.K., J.S.F., D.N.B., S.R., D.L.S., K.A.B.; Methodology: D.B., J.L., K.S.T., S.O., D.L.S., E.D.C., H.D.M., S.R.; Validation: D.B., J.L., K.S.T., S.O., D.L.S., E.D.C., H.D.M., S.R.; Formal analysis: D.B., J.L., K.S.T., S.O., D.L.S., E.D.C., H.D.M., S.R.; Investigation: D.B., J.L., K.S.T., S.O., D.L.S., E.D.C., H.D.M., S.R.; Resources: M.K., J.S.F., D.M., K.A.B., K.S.T., S.O., D.L.S., E.D.C., H.D.M., S.R., D.N.B., T.K., J.L.D.; Data curation: D.M., K.A.B., S.O.; Writing - original draft: D.M., M.K., J.S.F.; Writing - review & editing: D.B., J.L., K.S.T., S.O., D.L.S., E.D.C., H.D.M., S.R., J.L.D.; Visualization: D.B., J.L., K.S.T., S.O., D.L.S., E.D.C., H.D.M., S.R.; Supervision: D.M., M.K., J.S.F., S.R., H.D.M., D.L.S., K.S.T.; Project administration: M.K., J.S.F., D.M., J.L.D.; Funding acquisition: J.L.D., J.S.F.; Methodology, Software, Validation, Formal analysis, Investigation, Writing - review & editing, Visualization: D.M., K.A.B., D.A., Y.B., D.B., W.C., E.M.G., L.R.K., B.M., L.S.M., C.A.N., E.M.P., P.S., R.T., T.A., B.E.A., D.P.B., I.D.B., R.M.B., C.M.C., S.A.D., M.S.D., S.E., C.V.M.H., N.L.H., E.K., N.K., M.S.K., L.L.K., G.R.L., K.L., W.M.M., A.M.M., P.F.M., D.M.-T., T.L.N., A.M.N., K.O., S.P., G.K.R., N.J.R., P.J.R., D.M.C.S., S.K.T., P.V.T., R.W.T., J.P.T., M.K.T., F.S.U., D.R.W., A.M.W., T.S.W., D.N.B., T.K., M.K., J.S.F.

**Funding**

The Project Lab component of this work is specifically supported by a National Institute of Biomedical Imaging and Bioengineering T32 Training Grant, 'Integrative Program in Complex Biological Systems' [T32-EB009383]. UCSF iQPB and CCB Graduate programs are supported by US National Institutes of Health grants [EB009383, GM067547, GM064337, and GM008284], Howard Hughes Medical Institute/ National Institute of Biomedical Imaging and Bioengineering [56005676], School of Medicine, University of California, San Francisco, School of Pharmacy, University of California, San Francisco, University of California, San Francisco Graduate Division, University of California, San Francisco Chancellor's Office, and Discovery Funds. W.C., E.M.G., L.R.K., L.S.M., P.S., S.E.K., T.L.N., N.J.R. and F.S.U. are supported by National Science Foundation Graduate Research Fellowships. D.N.B. is supported by National Institutes of Health [GM112844]. T.K. is supported by National Institutes of Health [R01 GM117189, R01 GM110089] and National Science Foundation [MCB-1615990]. H.D.M. is supported by National Institutes of Health [R01 GM071801, R01 AI100272, R01 AI120464, R56 AI126726] and the Chan Zuckerberg Biohub. M.K. is supported by an National Institutes of Health Director's New Innovator Award National Institute of Health/National Institute of General Medical Sciences [DP2 GM119139], an Allen Distinguished Investigator Award (Paul G. Allen Family Foundation), a Stand Up to Cancer Innovative Research Grant, National Institutes of Health [K99/R00 CA181494], the Tau Center Without Walls National Institutes of Health/National Institute of Neurological Disorders and Stroke [U54 NS100717], the Chan Zuckerberg Biohub and the Paul F. Glenn Center for Aging Research. J.S.F. is a Searle Scholar, Pew Scholar, and Packard Fellow, and is supported by National Institutes of Health [OD009180].

**Supplementary information**

Supplementary information available online at

<http://bio.biologists.org/lookup/doi/10.1242/bio.036103.supplemental>

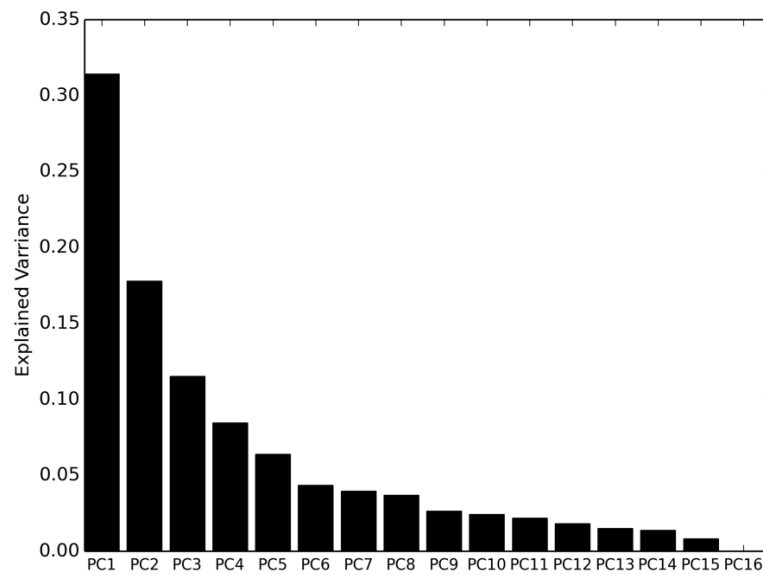
**References**

- Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K. et al. (2017). The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031-3048.
- Araya, C. L. and Fowler, D. M. (2011). Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* **29**, 435-442.
- Araya, C. L., Fowler, D. M., Chen, W., Muniez, I., Kelly, J. W. and Fields, S. (2012). A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc. Natl. Acad. Sci. USA* **109**, 16858-16863.
- Bandaru, P., Shah, N. H., Bhattacharyya, M., Barton, J. P., Kondo, Y., Cofsky, J. C., Gee, C. L., Chakraborty, A. K., Kortemme, T., Ranganathan, R. et al. (2017). Deconstruction of the Ras switching cycle through saturation mutagenesis. *eLife* **6**, e27810.
- Chawla, A., Chakraborty, S., Ghosh, G. and Niwa, M. (2011). Attenuation of yeast UPR is essential for survival and is mediated by IRE1 kinase. *J. Cell Biol.* **193**, 41-50.
- Eddins, M. J., Carlile, C. M., Gomez, K. M., Pickart, C. M. and Wolberger, C. (2006). Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. *Nat. Struct. Mol. Biol.* **13**, 915-920.
- Erpapazoglou, Z., Walker, O. and Haguenauer-Tsapis, R. (2014). Versatile roles of k63-linked ubiquitin chains in trafficking. *Cells* **3**, 1027-1088.
- Fenwick, R. B., Esteban-Martín, S., Richter, B., Lee, D., Walter, K. F. A., Milovanovic, D., Becker, S., Lakomek, N. A., Griesinger, C. and Salvatella, X. (2011). Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. *J. Am. Chem. Soc.* **133**, 10336-10339.
- Finley, D., Ulrich, H. D., Sommer, T. and Kaiser, P. (2012). The ubiquitin-proteasome system of *Saccharomyces cerevisiae*. *Genetics* **192**, 319-360.
- Fowler, D. M. and Fields, S. (2014). Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801-807.
- Hilton, S. K., Doud, M. B. and Bloom, J. D. (2017). phydm: software for phylogenetic analyses informed by deep mutational scanning. *PeerJ* **5**, e3657.
- Kwon, Y. T. and Ciechanover, A. (2017). The ubiquitin code in the ubiquitin-proteasome system and autophagy. *Trends Biochem. Sci.* **42**, 873-886.
- Liu, Z., Gong, Z., Jiang, W.-X., Yang, J., Zhu, W.-K., Guo, D.-C., Zhang, W.-P., Liu, M.-L. and Tang, C. (2015). Lys63-linked ubiquitin chain adopts multiple conformational states for specific target recognition. *eLife* **4**, e05767.
- Mavor, D., Barlow, K., Thompson, S., Barad, B. A., Bonny, A. R., Cario, C. L., Gaskins, G., Liu, Z., Deming, L., Axen, S. D. et al. (2016). Determination of ubiquitin fitness landscapes under different chemical stresses in a classroom setting. *eLife* **5**, e15802.
- McLaughlin, R. N., Jr, Poelwijk, F. J., Raman, A., Gosal, W. S. and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. *Nature* **491**, 138-142.

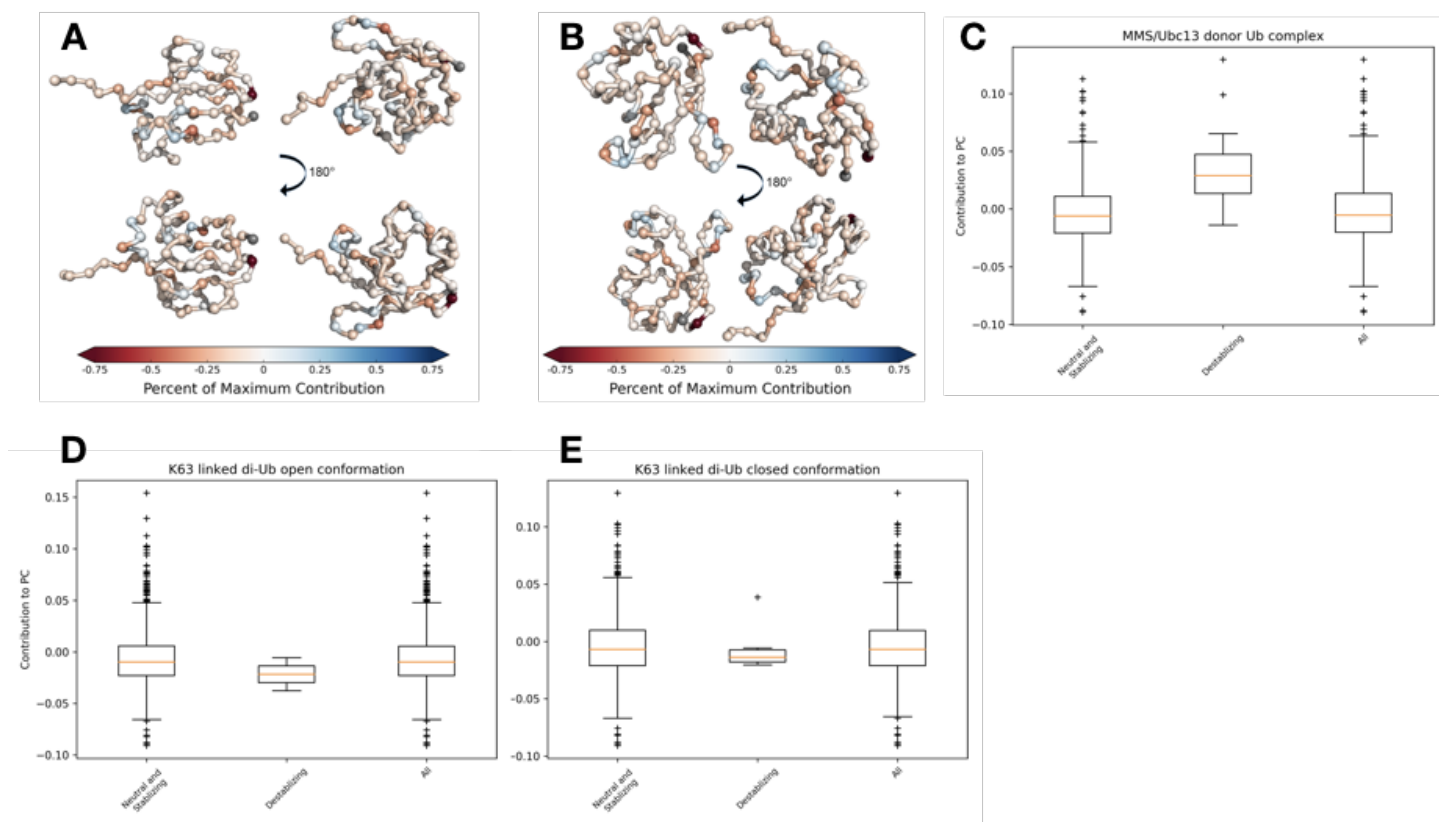


- Roscoe, B. P., Thayer, K. M., Zeldovich, K. B., Fushman, D. and Bolon, D. N. A. (2013). Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J. Mol. Biol.* **425**, 1363-1377.
- Shah, N. H., Löbel, M., Weiss, A. and Kuriyan, J. (2018). Fine-tuning of substrate preferences of the Src-family kinase Lck revealed through a high-throughput specificity screen. *eLife* **7**, e35190.
- Silva, G. M., Finley, D. and Vogel, C. (2015). K63 polyubiquitination is a new modulator of the oxidative stress response. *Nat. Struct. Mol. Biol.* **22**, 116-123.
- Starr, T. N., Picton, L. K. and Thornton, J. W. (2017). Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* **549**, 409-413.
- Starr, T. N., Flynn, J. M., Mishra, P., Bolon, D. N. A. and Thornton, J. W. (2018). Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proc. Natl. Acad. Sci. USA* **115**, 4453-4458.
- Weeks, S. D., Grasty, K. C., Hernandez-Cuevas, L. and Loll, P. J. (2009). Crystal structures of Lys-63-linked tri- and di-ubiquitin reveal a highly extended chain architecture. *Proteins* **77**, 753-759.
- Wrenbeck, E. E., Azouz, L. R. and Whitehead, T. A. (2017). Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* **8**, 15695.
- Xu, P., Duong, D. M., Seyfried, N. T., Cheng, D., Xie, Y., Robert, J., Rush, J., Hochstrasser, M., Finley, D. and Peng, J. (2009). Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137**, 133-145.
- Yau, R. and Rape, M. (2016). The increasing complexity of the ubiquitin code. *Nat. Cell Biol.* **18**, 579-586.
- Zhang, W., Qin, Z., Zhang, X. and Xiao, W. (2011). Roles of sequential ubiquitination of PCNA in DNA-damage tolerance. *FEBS Lett.* **585**, 2786-2794.





**Supplemental Figure 1.** Variance explained by the each of first 16 PCs of the difference fitness data.



**Supplemental Figure 2.** Association between PC3 and K63-related structures. (A) K63 linked di-Ub open conformation (3H7P) colored by average contribution to PC3. (B) K63 linked di-Ub closed conformation (2N2K) colored by average contribution to PC3. (C) Rosetta calculations reveal a potential association between destabilizing the MMS/Ubc13 donor complex (2GMI) and PC3. Binning mutations as either neutral and stabilizing mutations (Rosetta Energy Units (REU) < 0.75) or destabilizing (REU ≥ 0.75) reveals a significant difference in PC3 contribution magnitude (Student's t-test,  $p = 0.000215$ ) for this complex. (D) In contrast, Rosetta calculations reveal no significant difference for the K63 linked di-Ub open conformation (3H7P, Student's t-test,  $p = 0.887$ ) or (E) for the K63 linked di-Ub closed conformation (2N2K, Student's t-test,  $p = 0.887$ )