

# eScholarship@UMassChan

## Determination of Ubiquitin Fitness Landscapes Under Different Chemical Stresses in a Classroom Setting [preprint]

Item Type	Preprint
Authors	Mavor, David;Roscoe, Benjamin P.;Bolon, Daniel N A;Fraser, James S.
Citation	<p>bioRxiv 025452; doi: <a href="https://doi.org/10.1101/025452">https://doi.org/10.1101/025452</a> . <a href="https://doi.org/10.1101/025452" target="_blank">Link to preprint on bioRxiv service</a>.</p>
DOI	<a href="https://doi.org/10.1101/025452">10.1101/025452</a>
Rights	The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.
Download date	2026-03-14 17:20:48
Item License	<a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>
Link to Item	<a href="https://hdl.handle.net/20.500.14038/29346">https://hdl.handle.net/20.500.14038/29346</a>

## Determination of Ubiquitin Fitness Landscapes Under Different Chemical Stresses in a Classroom Setting

David Mavor<sup>1</sup>, Kyle A. Barlow<sup>2</sup>, Samuel Thompson<sup>1</sup>, Benjamin A. Barad<sup>1</sup>, Alain R. Bonny<sup>1</sup>, Clinton L. Cario<sup>2</sup>, Garrett Gaskins<sup>2</sup>, Zairan Liu<sup>1</sup>, Laura Deming<sup>9</sup>, Seth D. Axen<sup>2</sup>, Elena Caceres<sup>2</sup>, Weilin Chen<sup>2</sup>, Adolfo Cuesta<sup>3</sup>, Rachel Gate<sup>2</sup>, Evan M. Green<sup>1</sup>, Kaitlin R. Hulce<sup>3</sup>, Weiyue Ji<sup>1</sup>, Lillian R. Kenner<sup>1</sup>, Bruk Mensa<sup>3</sup>, Leanna S. Morinishi<sup>2</sup>, Steven M. Moss<sup>3</sup>, Marco Mravic<sup>1</sup>, Ryan K. Muir<sup>3</sup>, Stefan Niekamp<sup>1</sup>, Chimno I. Nnadi<sup>3</sup>, Eugene Palovcak<sup>1</sup>, Erin M. Poss<sup>3</sup>, Tyler D. Ross<sup>1</sup>, Eugenia Salcedo<sup>3</sup>, Stephanie See<sup>3</sup>, Meena Subramaniam<sup>2</sup>, Allison W. Wong<sup>3</sup>, Jennifer Li<sup>4</sup>, Kurt S. Thorn<sup>5</sup>, Shane Ó. Conchúir<sup>6</sup>, Benjamin P. Roscoe<sup>7</sup>, Eric D. Chow<sup>5,8</sup>, Joseph L. DeRisi<sup>5,9</sup>, Tanja Kortemme<sup>6</sup>, Daniel N. Bolon<sup>7</sup>, James S. Fraser<sup>6,\*</sup>

1 - Biophysics Graduate Group, University of California, San Francisco, California 94158, United States

2 - Bioinformatics Graduate Group, University of California, San Francisco, California 94158, United States

3 - Chemistry and Chemical Biology Graduate Program, University of California, San Francisco, San Francisco, California 94158, United States

4 - UCSF Science and Health Education Partnership, University of California, San Francisco, San Francisco, California 94158, United States; Lowell High School, San Francisco, California 94132, United States

5 - Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, United States

6 - Department of Bioengineering and Therapeutic Science and California Institute for Quantitative Biology, University of California, San Francisco, California 94158, United States

7 - Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Massachusetts 01605, United States

8 - Center for Advanced Technology, University of California, San Francisco, California 94158, United States

9 - Howard Hughes Medical Institute, UCSF, San Francisco, 94158 California, United States

\* - [james.fraser@ucsf.edu](mailto:james.fraser@ucsf.edu)

### ABSTRACT

Ubiquitination is an essential post-translational regulatory process that can control protein stability, localization, and activity. Ubiquitin is essential for eukaryotic life and is highly conserved, varying in only 3 amino acid positions between yeast and humans. However, recent deep sequencing studies in *S. cerevisiae* indicate that ubiquitin is highly tolerant to single amino acid mutations. To resolve this paradox, we hypothesized that the set of tolerated substitutions would be reduced when the cultures are not grown in rich media conditions and that chemically induced physiologic perturbations might unmask constraints on the ubiquitin sequence. To test this hypothesis, a class of first year UCSF graduate students employed a deep mutational scanning procedure to determine the fitness landscape of a library of all possible single amino acid mutations of ubiquitin in the presence of one of five small molecule perturbations: MG132, Dithiothreitol (DTT), Hydroxyurea (HU), Caffeine, and DMSO. Our data reveal that the number of tolerated substitutions is greatly reduced by DTT, HU, or Caffeine, and that these perturbations uncover “shared sensitized positions” localized to areas around the hydrophobic patch and to the C-terminus. We also show perturbation specific effects including the sensitization of His68 in HU and tolerance to mutation at Lys63 in DTT. Taken together, our data suggest that chemical stress reduces buffering effects in the ubiquitin proteasome system, revealing previously hidden fitness defects. By expanding the set of chemical perturbations assayed, potentially by other classroom-based experiences, we will be able to further address the apparent dichotomy between the extreme sequence conservation and the experimentally observed mutational tolerance of ubiquitin. Finally, this study demonstrates the realized potential of a project lab-based interdisciplinary graduate curriculum.

51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94

## INTRODUCTION

Protein homeostasis enables cells to engage in dynamic processes and respond to fluctuating environmental conditions (Powers et al., 2009). Misregulation of proteostasis leads to disease, including many cancers and neurodegenerative diseases (Balch et al., 2008; Lindquist and Kelly, 2011). Protein degradation is an important aspect of this regulation. In eukaryotes ~80% of the proteome is degraded by the highly conserved ubiquitin proteasome system (UPS) (Zolk et al., 2006). The high conservation of the UPS is epitomized by ubiquitin (Ub), a 76 amino acid protein post-translational modification that is ligated to substrate amine groups, including on Ub itself in poly-Ub linkages, via a three enzyme cascade (Finley et al., 2012).

Perhaps due to its central role in regulation, the sequence of ubiquitin has been extremely stable throughout evolution. Only three residues vary between yeast and human (96% sequence identity). This remarkable conservation implies that the UPS does not acquire new functions through mutations in the central player, Ub. Instead the evolution of proteins that add Ub to substrate proteins (E2/E3 enzymes), remove Ub (deubiquitinating enzymes, DUBs), or recognize Ub (adaptor proteins) combine to create new functions, many of which rely on various poly-Ub topologies (Sharp and Li, 1987; Zuin et al., 2014). The role of Lys48 linked poly-Ub in protein degradation (Thrower et al., 2000) appears to be universally conserved, but the functions of other linkages are more plastic. Although mass spectroscopy of cell lysates has shown that every possible poly-Ub lysine linkage exists within yeast cells (Peng et al., 2003), only the roles of Lys11 linked poly-Ub in ERAD (Xu et al., 2009) and Lys63 linked poly-Ub in DNA damage (Zhang et al., 2011) and endocytosis (Erpapazoglou et al., 2014) are well characterized in yeast. Both of these linkages are central to stress responses, mirroring some of the established roles for non-Lys48 linkages in other organisms (Komander and Rape, 2012).

Given this central role in coordinating a diverse set of stress responses, perhaps the high sequence conservation of ubiquitin is not surprising. However, classic Alanine-scanning studies showed that ubiquitin is quite tolerant of mutation under normal growth conditions (Sloper-Mould et al., 2001). The mutational tolerance of Ub was further confirmed using EMPIRIC (“extremely methodical and parallel investigation of randomized individual codons”), where growth rates of yeast strains harboring a nearly comprehensive library of all ubiquitin point mutations were assessed in bulk by deep sequencing (Roscoe et al., 2013). Subsequent studies revealed that many of the constraints on the Ub sequence are enforced directly by the E1-Ub interaction (Roscoe and Bolon, 2014); however, the surprisingly high number of tolerant positions remained unexplained. Previous EMPIRIC experiments on HSP90 suggested that reducing protein expression could reveal fitness defects that are otherwise buffered (Jiang et al., 2013). Similarly, we hypothesized that a buffer might be removed by subjecting cells to chemical stresses. Moreover, this chemical genetic approach might also allow us to relate specific residues to the stress response induced by a specific chemical.

To address the paradox of the high sequence conservation and mutational tolerance of ubiquitin, we posed the problem to the first year students in UCSF’s iPQB (Integrative Program in Quantitative Biology and CCB (Chemistry & Chemical Biology) graduate programs. The students performed the bulk competition experiments, deep sequencing and data analysis as part of an 8-

95 week long research class held in purpose-built Teaching Lab. In small teams of 4-5 students  
96 working together for 3 afternoons each week, they each examined a chemical stressor: Caffeine,  
97 which inhibits TOR and consequently the cell cycle (Reinke et al., 2006; Wanke et al., 2008);  
98 Dithiothreitol (DTT), which reduces disulfides and induces the unfolded protein response (Frandsen  
99 and Kaiser, 1998) and the ER associated decay (ERAD) pathway (Friedlander et al., 2000);  
100 Hydroxyurea (HU), which causes pausing during DNA replication and induces DNA damage (Koc  
101 et al., 2004; Petermann et al., 2010); or MG132, which inhibits the protease activity of the  
102 proteasome (Jensen et al., 1995; Rock et al., 1994). We expected MG132 to desensitize the yeast  
103 to deleterious mutations in Ub, as the inhibition acts on the final degradation of UPS substrates.  
104 For the other three chemicals we expected that specific sites on Ub would become sensitized to  
105 mutation. These sites could represent important Ub/protein binding interfaces that are required for  
106 Ub to bind to adaptor proteins and ligation machinery required to respond to a specific stress.  
107 Furthermore, we expected that Caffeine induced stress would be mediated through Lys48 linked  
108 poly-Ub (cell cycle), DTT induced stress would be mediated through Lys11 linked poly-Ub (ERAD),  
109 and HU induced stress would be mediated through Lys63 linked poly-Ub (DNA damage response).

110 Our data collectively show that stress reduces a general buffering effect and unmasks a  
111 shared set of residues that become less tolerant to mutation. Additionally, we have identified a  
112 small set of mutations that are specifically aggravated or alleviated by each chemical. We suggest  
113 that expanding the set of environmental stresses might be able to explain the high sequence  
114 conservation of ubiquitin, as different positions in the protein are important for interactions  
115 mediating the specific responses to a wide variety of perturbations.

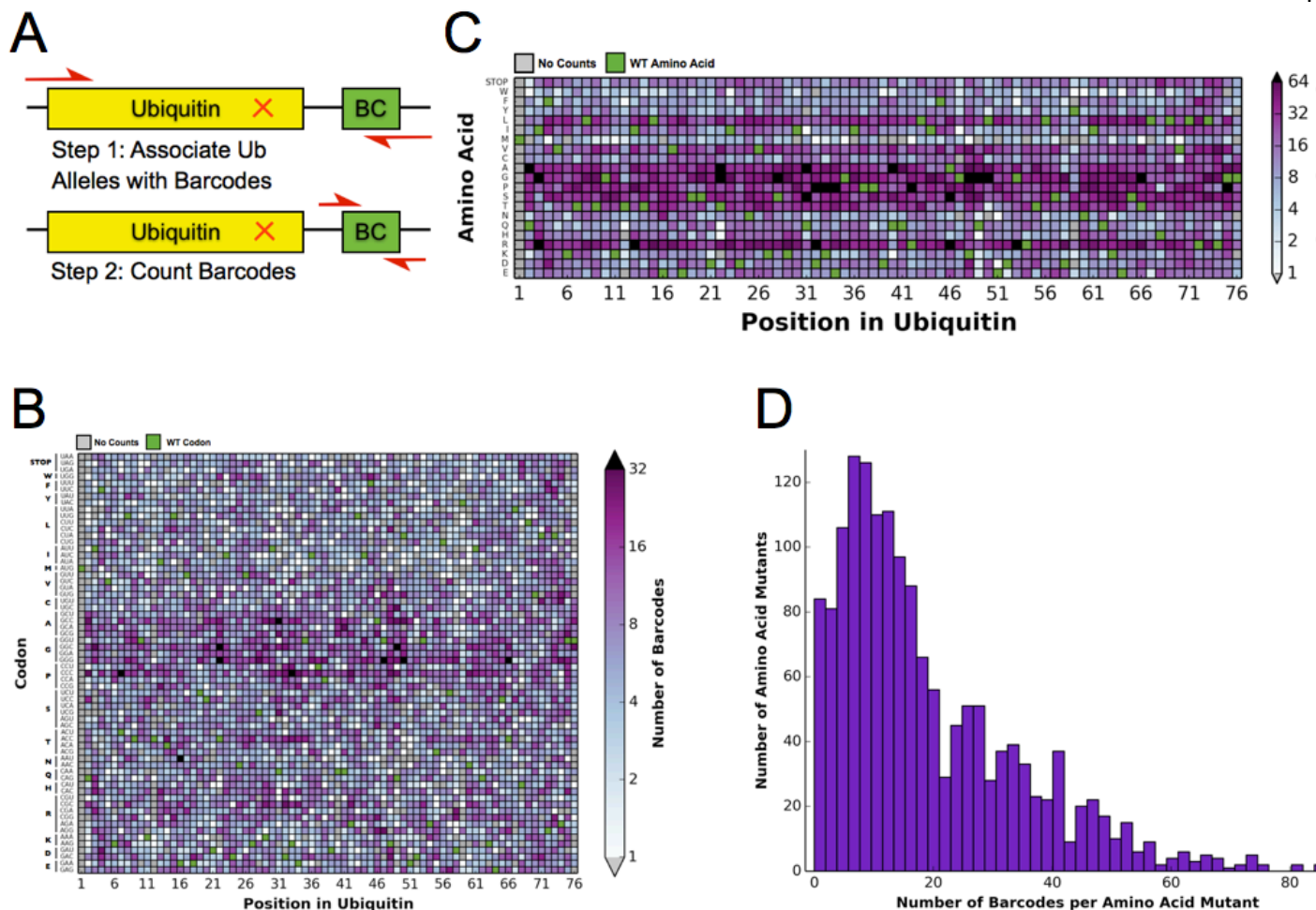
116

## 117 **RESULTS**

118

### 119 **Library Construction:**

120 Previously, the fitness landscape of Ub in yeast was determined using eight competition  
121 experiments using the EMPIRIC strategy of deep sequencing short regions of all possible single  
122 amino acid substitutions during a growth competition experiment in rich media (Roscoe et al.,  
123 2013). These experiments measured all point mutants contained in short 30 base pair (bp)/10  
124 amino acid residue stretches of the Ub open reading frame (ORF), which necessitated 8 separate  
125 competition experiments. To increase the throughput and reduce the cost of the experiment, we  
126 designed a barcoding strategy (Fowler et al., 2014), that allowed us to determine allele fitness in a  
127 single experiment using EMPIRIC with barcodes (EMPRIC-BC). We synthesized eighteen bp  
128 random barcodes (N18 BCs), which were ligated upstream of the Illumina sequencing primer  
129 binding site. The specific association of each unique N18 BC with a given mutant Ub allele was  
130 then established through paired end sequencing of the Ub ORF and the N18 BC (**Figure 1A**) The  
131 resulting lookup table of BCs and alleles was then employed in our competition experiments to  
132 count alleles by directly sequencing the N18 BCs. In addition to simplifying the experiment, this  
133 strategy enabled us to count the alleles with a short, single end sequencing run, substantially  
134 reducing cost. The library is nearly complete at the amino acid level. We observed a slight GC bias  
135 in the codon coverage (**Figure 1 B-C**), which is likely due to the cloning method that initially  
136 generated the Ub mutants (Hietpas et al., 2012). Most substitutions are associated with many N18  
137 BCs, with a median of fifteen unique barcodes representing a specific amino acid substitution  
138 (**Figure 1D**).



139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

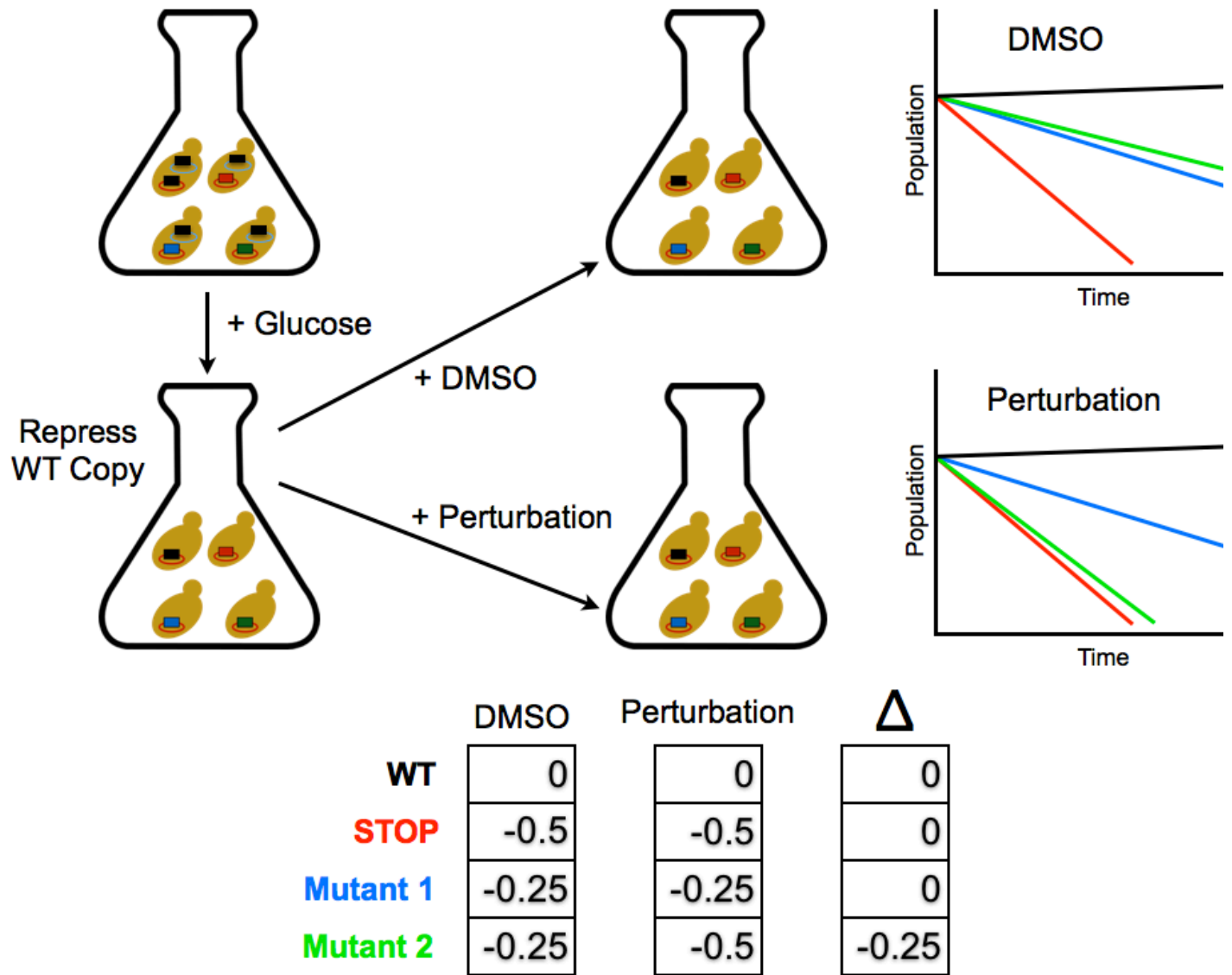
157

158

**Figure 1) Barcoding enables a bulk competition experiment of ~1500 Ubiquitin variants. A)** Prior to the competition experiment, ubiquitin alleles were specifically associated with unique barcodes through a paired end sequencing. To monitor the frequency of different alleles during the competition experiments, we directly sequenced the barcodes in a short single end read. **(B)** The library contains most codon substitutions and almost all are associated with multiple barcodes. A slight GC bias is seen in the cloning. WT codons are shown in green and missing alleles are shown in grey. **(C)** The amino acid coverage of the library is almost complete. WT residues are shown in green and missing alleles are shown in grey. **(D)** Examining the number of barcodes per amino acid substitution shows that 2.5% of the library is missing and the median number of barcodes per substitution is 15.

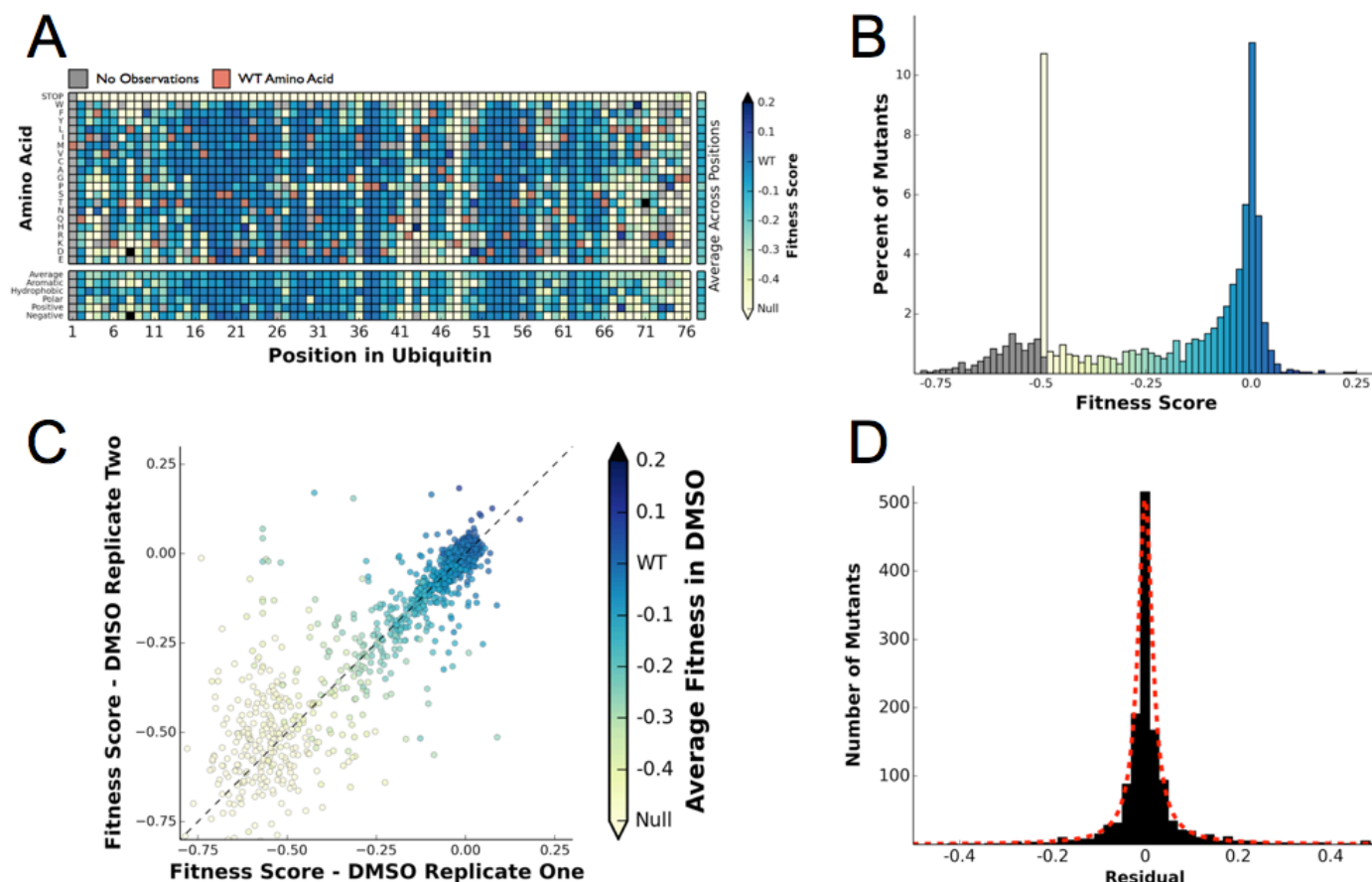
### Determining the Ub Fitness Landscape in DMSO

To determine the differential fitness landscape of Ub under different chemical stresses, we first conducted an EMPIRIC-BC experiment under 0.5% DMSO to serve as a control (**Figure 2**). The resulting fitness landscape is quite similar to the previously published dataset, which was collected under no chemical stress (Roscoe et al., 2013) (**Figure 3A**). The lowest fitness scores occurred at premature stop codons and residues that are critical to build Lys48 poly-Ub linkages (Lys48, Ile44, Gly75, Gly76). As previously observed, much of the protein surface is tolerant to mutation. Based on the average value of the stop codon substitutions, we set a minimum fitness score of -0.5 (**Figure 3B**). Comparisons of biological replicates indicated that the data were reproducible and well fit by a Lorentzian function centered at 0 (**Figure 3C,D**).



159  
160  
161  
162  
163  
164

**Figure 2) Competition experiment based on a galactose inducible Ub.** The fitness of all ubiquitin mutants was measured in a single culture by shutting off the galactose-driven wild type copy. This allows a constitutively expressed mutant to be the sole source of ubiquitin for the cell. Upon repression of the wild type copy, chemical perturbations were added and the yeast were grown for multiple generations. Fitness scores were calculated for each mutant based on the relative frequencies of mutant and wild type alleles over multiple generations.



165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

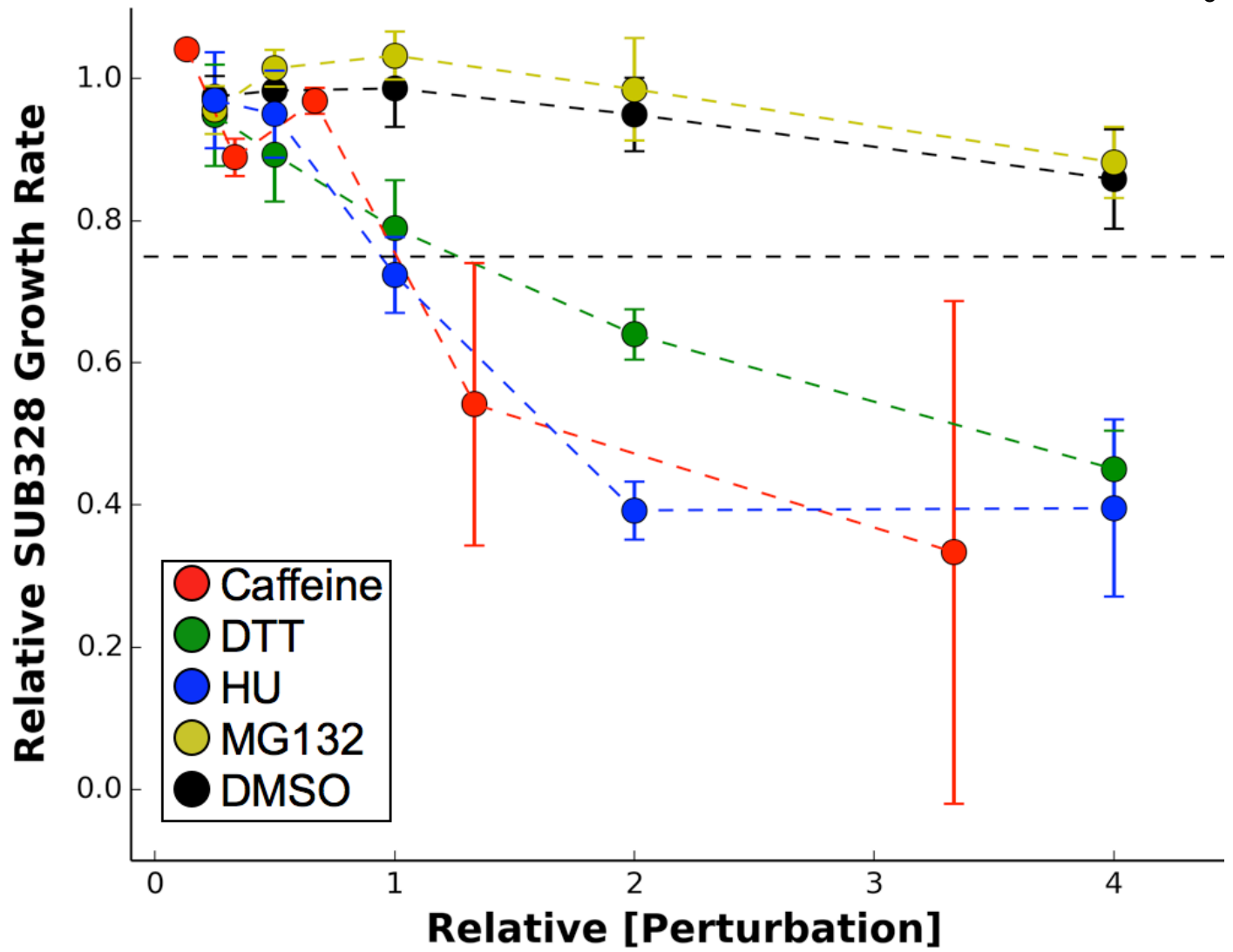
**Figure 3) Ubiquitin fitness scores determined in DMSO recapitulate previously determined unperturbed fitness maps.** (A) Heatmap showing the fitness of observed ubiquitin alleles. Scores presented are the average of three biological replicates. Wild type amino acids are shown in red and mutations without fitness values (due to lack of barcode or competition sequencing reads) are shown in grey. The average fitness score of each position and the averages of substitutions binned by amino acid characteristics are shown below. The single column on the far right shows the average of each amino acid substitution across all positions. (B) The distribution of fitness values is shown and colored based on fitness score. Grey bins reflect fitness scores that were reset to -0.5. (C) Biological replicates of the competition experiment in DMSO are well correlated ( $R^2 = 0.79$ ). Each point represents a single mutant and the color of the points corresponds to the fitness score determined by averaging 3 biological replicates (D) The distribution of the residuals to the identity line between two DMSO replicates is symmetric and well modeled by a Lorentzian ( $X_0 = 0$ ,  $\Gamma = 0.0035$ , scaled by 1600).

### Chemical Perturbations Sensitize Many Positions in Ub

Because our competition experiments require cells growing for multiple generations in log phase, we conducted our experiments at chemical concentrations that inhibit yeast growth by 25%. These chemical concentrations are not as high as used in previous transcriptional studies of yeast chemical stress responses (Gasch et al., 2000). For Caffeine (7.5mM), and DTT (1mM) we determined the IC25 for each drug by following growth via optical density (Figure 4 - Figure Supplement 1). Since HU (25mM) induced a lag phase followed by near wild type like growth, we determined the IC25 concentration by monitoring yeast growth from two to five hours post treatment. DMSO (0.5%) and MG132 (50uM) did not inhibit growth.

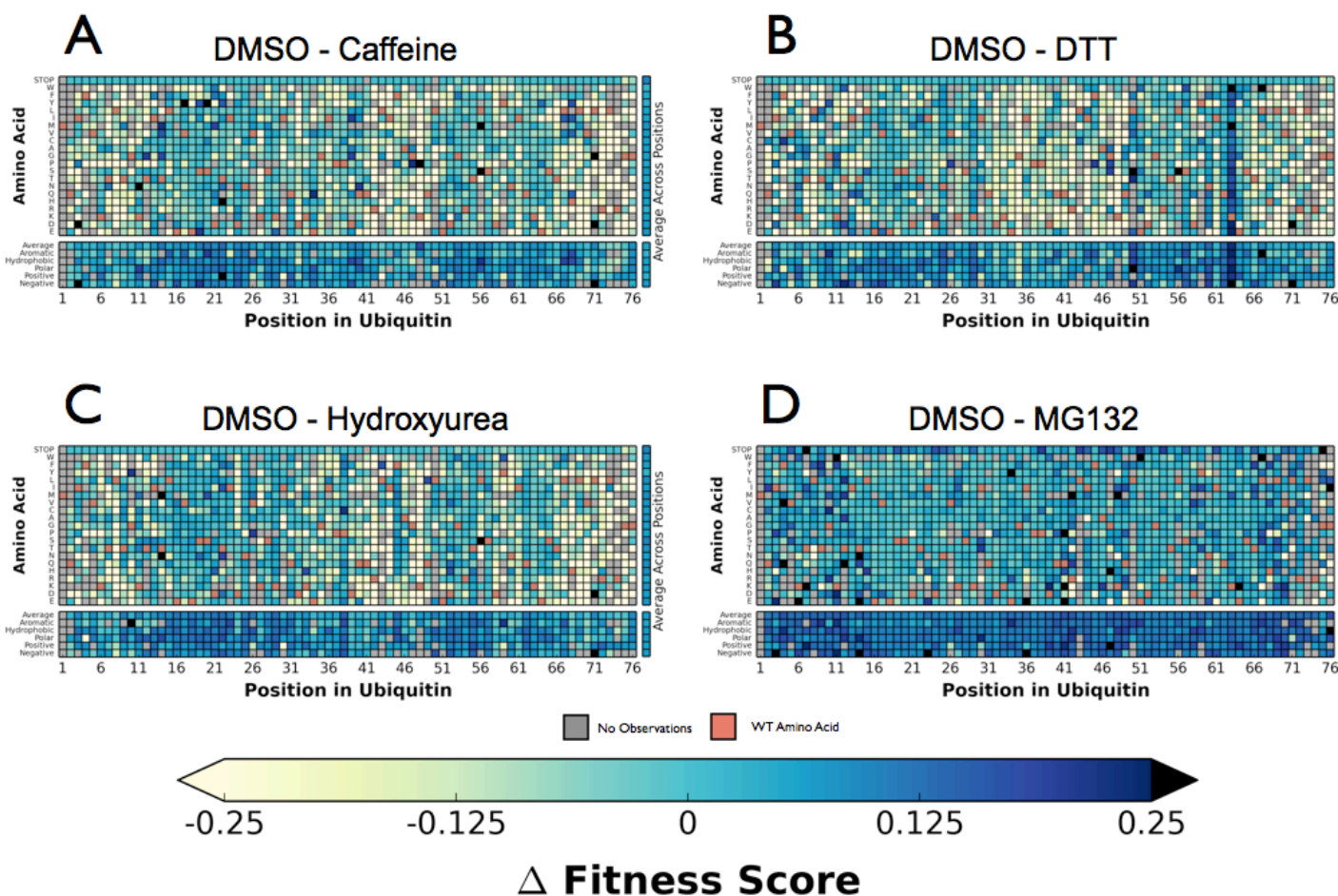
Next, we performed the EMPIRC-BC experiment with each chemical perturbation (Figure 4B). In Caffeine, DTT, and HU (Figure 4A-D, Figure 4 - Figure Supplement 2) many mutations are sensitized, and become less fit than in DMSO. Generally this increased sensitivity is localized





199  
200  
201  
202

**Figure 4 - Figure Supplement 1) Growth curves.** We determined the concentration to inhibit SUB328 growth by 25% by monitoring optical density. Error bars represent standard deviation of multiple measurements.



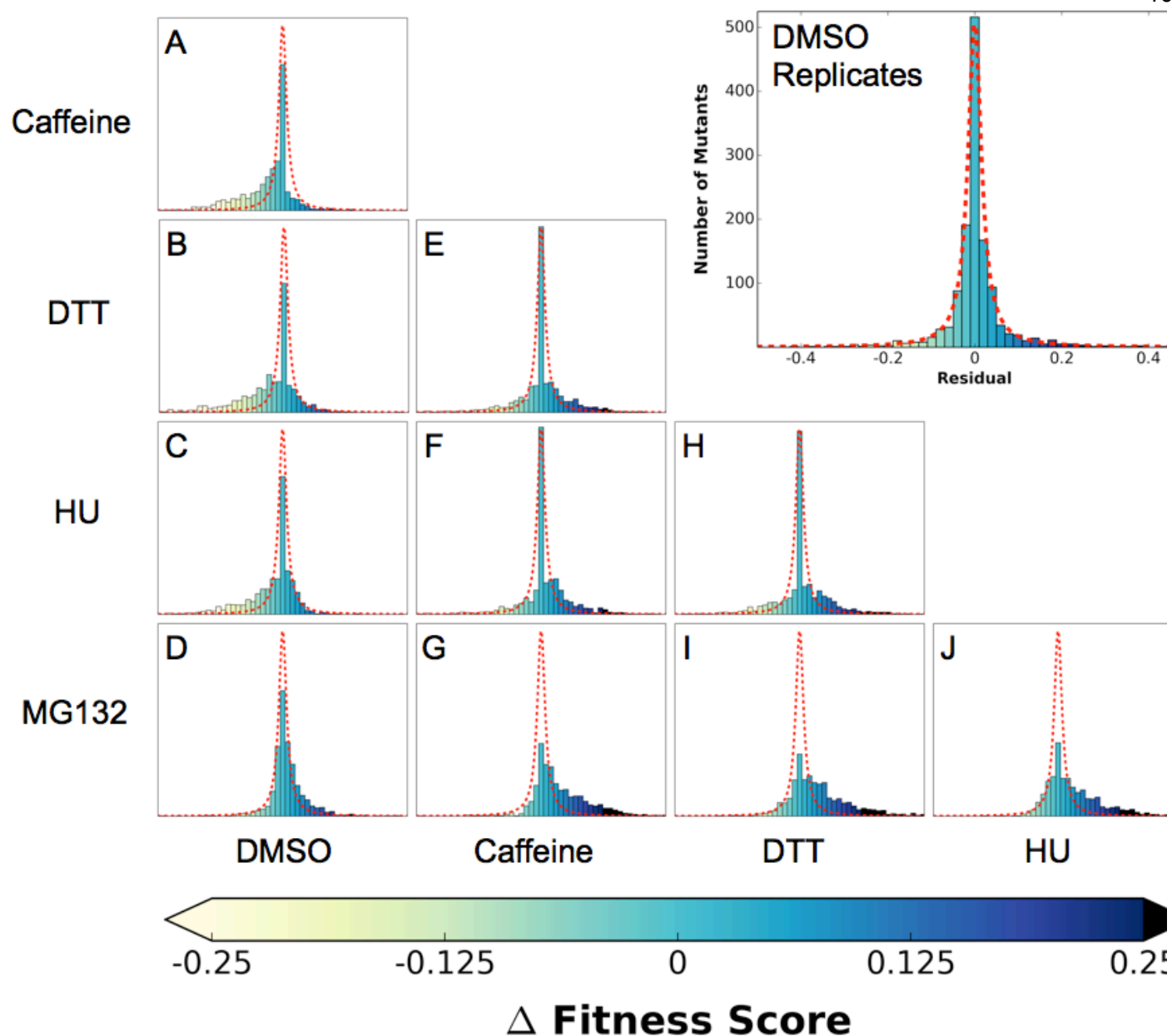
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223

**Figure4 - Figure Supplement 2) Difference maps relative to fitness measured after DMSO treatment.**

Perturbation fitness of each Ub allele under : **(A)** Caffeine, **(B)** DTT **(C)** Hydroxyurea **(D)** MG132. Wild type amino acids are shown in red and mutations without fitness values (due to lack of barcode or competition sequencing reads) are shown in grey. Interactive versions of these figures will appear with the final article.

To compare the responses to each perturbation, for each pairwise comparison we plotted the fitness scores for each mutant as a scatter plot and calculated the residual to the identity line. We compared the distribution of these residuals to the distribution of residuals calculated by the DMSO self comparison (**Figure 5**). Caffeine, DTT and HU generally sensitize the protein to mutation, which is evident in the enrichment of mutations with reduced fitness compared to the DMSO self distribution. These newly sensitized mutations are largely shared between these different chemical perturbations.

In contrast to the sensitizing effects of DTT, Caffeine, and HU, the proteasome inhibitor MG132 increases mutational robustness throughout the protein. This effect can be seen in the slight shift of the residuals distribution to the right when compared to the DMSO self distribution (**Figure 5 D**). The effect is small at the MG132 concentration we assayed, which is likely due to the poor penetrance of MG132 in yeast cells containing a wild type allele of *ERG6* (Lee and Goldberg, 1996). This alleviating interaction is likely because MG132 directly perturbs proteasome, reducing the impact of defects related to Lys48 linked poly-Ub chains and leaving functions related to other, non-degradative poly-Ub topologies unperturbed.



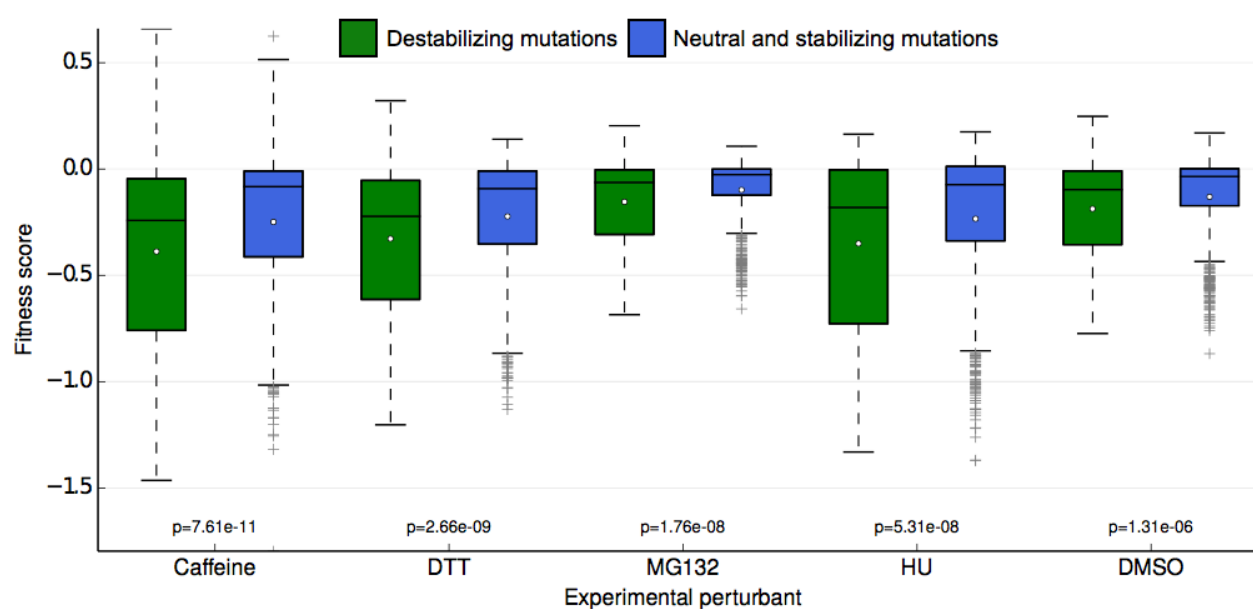
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240

**Figure 5) Residual distributions highlight a shared mutational response between Caffeine, DTT and HU.** The residuals between datasets shows are shown with the Lorentzian representing the biological replicates of DMSO in red. when compared to DMSO, three perturbations (Caffeine, DTT and HU) shift the distributions to the left, which highlights the increased sensitivity to mutation. In contrast, MG132 slightly shifts the distribution to the right, which highlights the alleviating interaction between MG132 and deleterious ubiquitin alleles. Comparisons between Caffeine, Hydroxyurea and DTT are symmetric but with longer tails than the control experiments. This result suggests a shared response comprised of many sensitized residues and a smaller number of perturbation-specific signals.

### **Rosetta $\Delta\Delta$ G Modeling Indicate that Sensitive Mutants Mildly Perturb Stability**

One potential explanation of the buffer unmasked by the chemical perturbations is the stability of the Ub protein itself. Although Ub is highly stable (Ibarra-Molero et al., 1999; Wintrode et al., 1994), mutations that destabilize it may lead to misfolding or perturb Ub/protein interactions important for UPS function. To assess the degree to which mutational destabilization of ubiquitin itself is predictive of a decrease in mean fitness for each perturbation, we used the macromolecular modeling software Rosetta to estimate changes in protein stability (Kellogg et al.,

241 2011; Kortemme and Baker, 2002) for every mutation in our library. With the resulting  
242 predictions, we classified each ubiquitin mutation as either destabilizing (change in Rosetta Energy  
243 Units (R.E.U.)  $\geq 1.0$ ) or neutral/stabilizing (change in R.E.U.  $< 1.0$ ). We observed a significant  
244 difference in experimental fitness between the two predicted classes for all conditions (**Figure 6**).  
245 This result holds independently of the absolute mean experimental fitness score of each  
246 perturbation, meaning that the difference in mean experimental fitness between predicted  
247 destabilizing and neutral mutations is not simply the result of lower mean destabilizing fitness  
248 scores. These results suggest that ubiquitin stability is more important for fitness in each of the  
249 perturbed conditions than in unperturbed yeast. Under stress, subtle changes in Ub stability could  
250 induce fitness defects that are otherwise buffered under control (DMSO) conditions. Furthermore,  
251 even small changes to ubiquitin stability could induce considerable changes to the Ub  
252 conformational ensemble that could destabilize Ub/protein complexes (Lange et al., 2008; Phillips  
253 et al., 2013). Adaptability within the UPS could buffer these defects in DMSO, but they can be  
254 revealed upon chemical stress.  
255



256  
257 **Figure 6) Fitness score data binned by Rosetta stability predictions.** Fitness scores for each of the 5 sets of  
258 experimental conditions are shown along the y-axis as boxplots. Scores are grouped first by their respective  
259 experimental condition, and then by the change in stability of the ubiquitin monomer of the mutation estimated by  
260 Rosetta. Mutations that Rosetta predicts to be neutral or stabilizing (R.E.U. (Rosetta Energy Units)  $< 1.0$ ) are shown in  
261 blue boxes; mutations predicted to be destabilizing (R.E.U.  $\geq 1.0$ ) are shown in green boxes. The mean of each  
262 fitness score distribution is shown as a white dot. The p-value of the two-sided T test between the fitness mean of  
263 mutations predicted to be stabilizing and those predicted to be neutral/stabilizing is shown at the bottom of the plot.  
264 Experimental conditions are arranged from left to right along the x-axis in order of decreasing p-value.  
265

### 266 **Mutational Sensitivity is Primarily Localized to Three Regions of Ub**

267 To assess the role of specific positions in Ub we averaged the fitness score of each amino  
268 acid mutation at a given position. We then binned each position into sensitive ( $\leq -0.35$ ),  
269 intermediate ( $-0.35$  to  $-0.075$ ) and tolerant ( $\geq -0.075$ ) and examined the distribution of average  
270 fitness in each condition (**Figure 7A**). These distributions again show that most positions in Ub are  
271 robust to mutation in DMSO, but many positions are sensitized upon chemical perturbation.

272 In DMSO only residues with well-established roles are sensitive: Arg42 (E1 activation),  
273 Ile44 (hydrophobic patch hotspot), Lys48 (essential Lys48 linked poly-Ub) and Gly75-Gly76 of the  
274 C-terminus (E1 activation). The face opposite the hydrophobic patch is mostly tolerant and the  
275 protein core and residues adjacent to the sensitive residues are mostly intermediate (**Figure 7B -**  
276 **i**). When treated with Caffeine, DTT or HU, a shared set of residues become sensitive (**Figure 7B**  
277 **ii- iv, Figure 7C**). These residues are either: located adjacent to DMSO sensitive residues (e.g.  
278 Leu73, which is in the C-terminal tail); residues with important biological functions that of  
279 intermediate sensitivity in DMSO (e.g. Leu8, Val70, which are important hydrophobic patch  
280 residues); or core residues (e.g. Ile36, Leu71). These positions tolerated a small set of  
281 substitutions in DMSO but upon perturbation became only tolerant of mutations that share physical  
282 chemistry with the wild type residue.

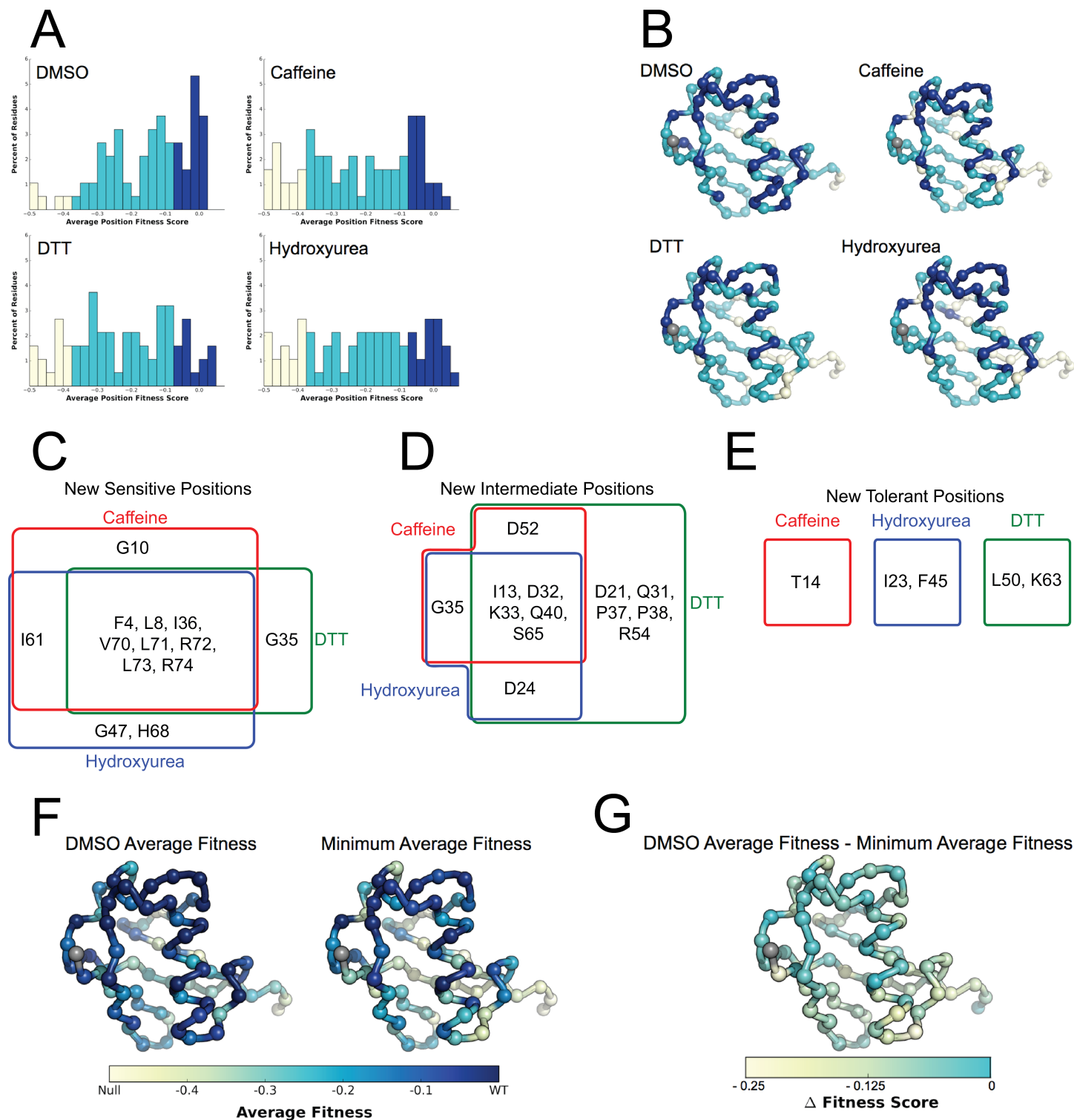
283 Examining the positions made intermediate by the perturbations highlights the similarities  
284 and differences between the DTT and Caffeine/HU datasets (**Figure 7D**). All three perturbations  
285 shift a shared set of residues to the intermediate bin. These residues are mostly surface residues  
286 on the tolerant face of Ub. In DMSO they tolerate a wide range of physiochemistries. Upon  
287 perturbation the mutational tolerance is reduced to physiochemistries generally compatible with  
288 surface residues. For example in DMSO, Asp32 is tolerant to any substitution except Proline. Upon  
289 perturbation, this position is restricted to polar and negative substitutions.

290 Additionally, DTT uniquely shifts five positions into the intermediate bin. This is due to subtle  
291 changes in the tolerance of positions that are otherwise highly tolerant. For example, mutations at  
292 Arg54 are well tolerated in all other conditions. However, in DTT mutations to negative residues  
293 become deleterious while all other substitution remain tolerated. This suggests that Arg54 may  
294 participate in a salt bridge during a protein-protein interaction that is involved in mediating the  
295 cellular response to DTT.

296 We also uncovered newly tolerant positions, which are uniquely tolerant to each of the  
297 perturbations (**Figure 7E**). These positions tend to be mildly sensitive to most mutations in DMSO,  
298 suggesting that these residues are involved in biological pathways that are important for cellular  
299 function, but not essential. When perturbed, these positions are mildly desensitized to mutation,  
300 with little regard for mutant physiochemistry. The most striking example is at Lys63 in DTT. In all  
301 other conditions any mutation of this residue is mildly deleterious. Because Lys63 linked poly-Ub  
302 chains are important for efficient cargo sorting in the endosome, this sensitivity is likely due to an  
303 endocytic defect. DTT treatment causes position 63 to become robust to mutations suggesting that  
304 an endocytic defect is protective against DTT treatment. Average difference maps showing the  
305 (DMSO - perturbation) fitness score highlight the features that underlie the sensitized and  
306 desensitized positions (**Figure 4 – Figure Supplement 2**).

307 In a final effort to resolve the dichotomy between the Ub fitness landscape and the  
308 evolutionary record, we visualized the average fitness of each position in DMSO and compared it  
309 to the minimum of the average fitness of each position for all perturbations (**Figure 7F-G**). The  
310 data in DMSO again shows that biologically relevant positions are sensitive, the face opposite the  
311 hydrophobic patch is extremely tolerant to mutation, and that core residues are intermediately  
312 tolerant. Perturbations dramatically increase mutational sensitivity at the C-terminus, around the  
313 hydrophobic patch and at some core positions. However, much of the tolerant face of the protein  
314 remains robust to mutation in all of the perturbations. By exploring a wider array of perturbations

315 we should be able determine the environmental pressures that constrain these tolerant positions  
 316 and explain the extreme conservation of Ub.



317  
 318

319 **Figure 7) Average fitness values show sensitization by the perturbations at each position in ubiquitin. (A)**  
 320 Based on the average fitness score, positions were binned into tolerant ( $\geq -0.075$  - Dark Blue), intermediate ( $< -0.075$   
 321 to  $> -0.35$  - Light Blue) and sensitive ( $\leq -0.35$  - Bone). (i) DMSO (ii) Caffeine (iii) DTT (iv) Hydroxyurea show a shift  
 322 from tolerant to intermediate and sensitive positions. (B) Positions binned by average fitness score mapped onto the  
 323 ubiquitin structure. C-alpha atoms are shown in spheres and the residues are colored as in A. Met1 is colored grey. (C)  
 324 New sensitive positions induced by the perturbation describe a shared response to perturbation with 8 of 13 positions  
 325 shared between Caffeine, DTT and HU. (D) New intermediate positions highlight the similarity between HU and

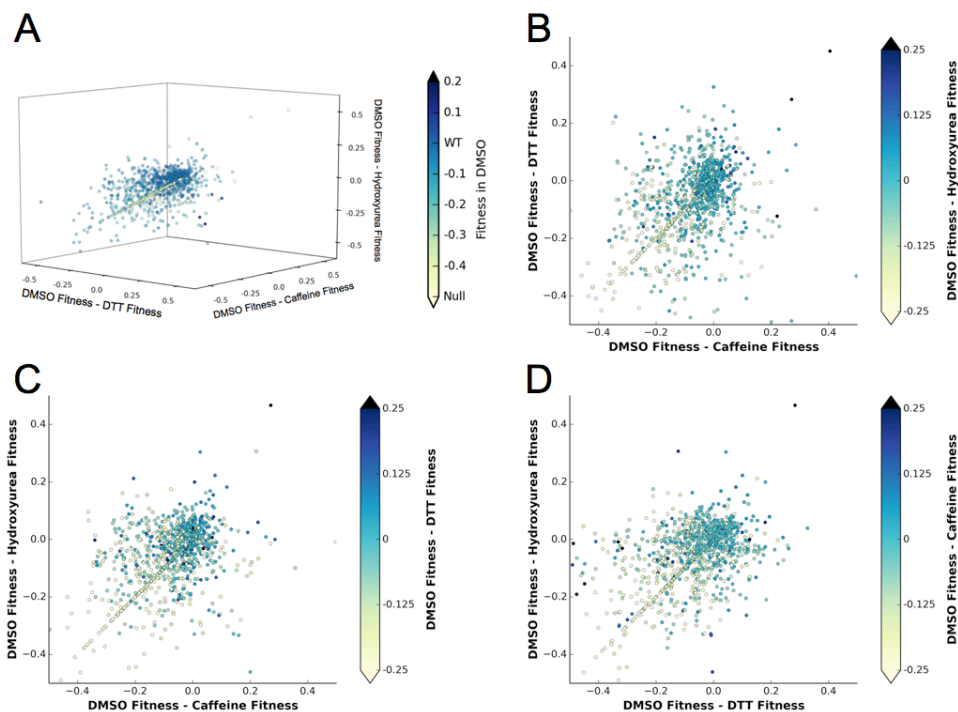
326 Caffeine, with DTT sensitizing a unique set of residues. **(E)** New tolerant positions are unique to each perturbation.  
327 **(G)** Average position fitness scores mapped onto ubiquitin. **(i)** DMSO **(ii)** Minimum average fitness score in all  
328 perturbations. C-alpha atoms are shown in spheres and the residues are colored, as in A. Met1 is colored grey. **(F)**  
329 DMSO average fitness scores – minimum average fitness scores mapped onto ubiquitin. C-alpha atoms are shown in  
330 spheres and the residues are colored as in A. Met1 is colored grey. With this small set of perturbations most positions  
331 are sensitized.  
332

332

### 333 **Specific Elements of the Shared Response to Perturbations**

334 To determine the elements of the shared response to HU, Caffeine and DTT, we defined  
335 “shared sensitizing mutations” as those that were both sensitizing ( $\Delta$  fitness  $\leq -0.2$  for all  
336 perturbations) and consistent between perturbations (within 0.1 of the regression line) (**Figure 8A**  
337 and **Figure 8 – Table Supplement 1**). Most of these mutations change from being mildly  
338 deleterious to being nearly null upon chemical stress. For example, in DMSO Ub tolerates mutation  
339 to small hydrophobics and other polar residues at Thr7. However, chemical stresses causes  
340 mutations of small hydrophobics or charged residues at this position to be deleterious. As Thr7 is  
341 adjacent to the hydrophobic patch residue Leu8, this sensitization is likely due to non-polar  
342 substitutions disrupting Ub adaptor protein binding and poly-Ub packing (Komander and Rape,  
343 2012). Additionally, typically destabilizing substitutions such as Proline or Tryptophan generally  
344 become more deleterious under perturbation.  
345

345



346

347 **Figure 8) A shared response to different chemical perturbations. (A)** Comparing the difference between fitness in  
348 either Caffeine, DTT or HU and DMSO shows both the shared response and mutations that are specifically affected by  
349 the perturbations. Points are colored based on the mutant fitness in DMSO. **(B - D)** Projections of the 3D scatter plot  
350 shown in A. **(B)** DMSO fitness - Caffeine fitness vs. DMSO fitness - DTT fitness. The markers are colored based on  
351 DMSO fitness - Hydroxyurea fitness. **(C)** DMSO fitness - Caffeine fitness vs. DMSO fitness - Hydroxyurea fitness. The  
352 markers are colored based on DMSO fitness - DTT fitness. **(D)** DMSO fitness - DTT fitness vs. DMSO fitness -  
353 Hydroxyurea fitness. The markers are colored based on DMSO fitness - Caffeine fitness. Interactive versions of these  
354 figures will appear with the final article.  
355

355

356  
357  
358  
359  
360  
361

**Figure 8 – Table Supplement 1) Shared response mutants representing mutations that are equally perturbed by all three sensitizing perturbations.** Mutants in the shared response were determined by fitting a line to the fitness scores. The distance from each point to that line was calculated. If the distance was less than 0.1 and the average  $\Delta$  (DMSO - Perturbation) fitness was less than -0.2 the mutant was considered part of the shared response. E1 activity relative to WT Ub (Roscoe and Bolon, 2014) is listed and may explain the sensitization of some of the shared response mutants.

Wild Type	Mutant	Type	Average $\Delta$ (DMSO - Perturbation) fitness	Relative E1 reactivity
Gln2	Aspartate	Polar to negative	-0.46	1.06
Val5	Aspartate	Hydrophobic to negative	-0.20	0.00
Lys6	Proline	Positive to proline	-0.35	0.08
Thr7	Methionine	Polar to hydrophobic	-0.41	1.01
Thr7	Glutamine	Polar to negative	-0.25	1.02
Leu8	Tyrosine	Hydrophobic to aromatic	-0.25	0.84
Leu8	Histidine	Hydrophobic to positive	-0.28	0.66
Leu8	Aspartate	Hydrophobic to negative	-0.72	0.21
Thr12	Valine	Polar to hydrophobic	-0.39	0.95
Ile13	Tyrosine	Hydrophobic to aromatic	-0.31	0.95
Val26	Arginine	Hydrophobic to positive	-0.31	0.09
Lys27	Serine	Positive to polar	-0.30	0.55
Ile30	Glycine	Hydrophobic to glycine	-0.28	0.46
Asp32	Phenylalanine	Negative to aromatic	-0.26	1.00
Asp32	Isoleucine	Negative to hydrophobic	-0.29	0.99
Glu34	Leucine	Glycine to hydrophobic	-0.32	1.03
Pro37	Tyrosine	Proline to aromatic	-0.31	0.96
Gln41	Proline	Polar to proline	-0.24	0.42
Arg42	Cystine	Positive to cystine	-0.27	-0.13
Arg42	Proline	Positive to proline	-0.23	0.31
Leu43	Tyrosine	Hydrophobic to aromatic	-0.32	0.87
Gly47	Phenylalanine	Glycine to aromatic	-0.38	0.14
Gly47	Threonine	Glycine to polar	-0.21	0.47
Gln49	Tyrosine	Polar to aromatic	-0.28	-0.61
Leu50	Glycine	Hydrophobic to glycine	-0.40	0.15
Asp58	Tyrosine	Negative to aromatic	-0.26	0.94

Asp58	Proline	Negative to proline	-0.30	0.52
Ile61	Tyrosine	Hydrophobic to aromatic	-0.36	0.20
Ile61	Glycine	Hydrophobic to glycine	-0.26	0.06
Ile61	Lysine	Hydrophobic to positive	-0.28	-0.02
Thr66	Tyrosine	Polar to aromatic	-0.37	0.93
Thr66	Isoleucine	Polar to hydrophobic	-0.24	0.94
Thr66	Arginine	Polar to positive	-0.24	0.98
Leu67	Glycine	Hydrophobic to glycine	-0.21	0.90
Leu69	Arginine	Hydrophobic to positive	-0.27	0.94
Val70	Tyrosine	Hydrophobic to aromatic	-0.25	-0.07
Leu71	Serine	Hydrophobic to polar	-0.22	1.02
Arg74	Isoleucine	Positive to hydrophobic	-0.37	1.00
Gly75	Phenylalanine	Glycine to aromatic	-0.41	0.11
Gly75	Valine	Glycine to hydrophobic	-0.26	0.11
Gly75	Asparagine	Glycine to polar	-0.24	0.17

362  
363  
364  
365  
366

**Figure 8 – Table Supplement 2) Outlier mutations represent alleles that are differentially affected by Caffeine, DTT and Hydroxyurea.** Outlier points were determined by fitting a line to the delta (DMSO - perturbation) fitness scores. The distance from each point to that line was calculated. If the distance was greater than 0.35 the point was called an outlier.

Wild Type	Mutant	Type	Notes	$\Delta$ fitness score in Caffeine	$\Delta$ fitness score in DTT	$\Delta$ fitness score in Hydroxyurea
Lys11	Asparagine	Positive to polar	Lys11 linked poly-Ub	0.55	-0.32	-0.03
Thr14	Valine	Polar to hydrophobic	Surface exposed beta strand, adjacent to Phe4	0.16	-0.01	-0.33
Val17	Tyrosine	Hydrophobic to aromatic	Core residue	0.27	-0.49	-0.01
Ser20	Tyrosine	Polar to aromatic	Surface exposed loop	0.5	-0.33	-0.01
Asp21	Phenylalanine	Negative to aromatic	Surface exposed loop	0.11	-0.41	-0.01
Thr22	Histidine	Polar to positive	Putative helix cap	0.51	-0.45	-0.15
Asp39	Isoleucine	Negative to hydrophobic	Surface exposed loop	0.20	0.00	-0.46
Gln40	Asparagine	Shortened by one carbon	Surface exposed loop	0.20	-0.49	-0.09
Leu50	Threonine	Hydrophobic to polar	Core residue	-0.34	0.22	0.00
Leu56	Methionine	Extension of	Core residue	0.57	-0.48	-0.19

		hydrophobic group				
Glu64	Tyrosine	Negative to aromatic	Surface exposed beta strand	0.12	-0.33	-0.30
His68	Tyrosine	Positive to aromatic	Surface exposed beta strand	0.17	-0.30	-0.28

367

368

**Figure 8 – Table Supplement 3: Specific information regarding highlighted mutants**

Mutant	Perturbation	Average of the barcode fitness scores	Standard deviation of barcode scores	Number of barcodes	Initial number of observations of each barcode
Lys11Asn	Caffeine	0.60	0.038	4	140, 24, 93, 105
	DTT	-0.27	0.033	4	1120, 188, 647, 698
	HU	0.02	0.034	5	485, 1878, 69, 299, 287
Glu64Arg	Caffeine	< -0.50	0.105	25	115, 3, 32, 91, 12, 174, 37, 27, 7, 4, 101, 15, 18, 66, 8, 52, 21, 34, 36, 4, 40, 14, 20, 36, 24
	DTT	-0.24	0.160	28	10, 102, 28, 38, 29, 101, 7, 172, 18, 39, 113, 21, 18, 65, 15, 9, 3, 68, 30, 44, 41, 8, 34, 37, 11, 27, 18, 24
	HU	-0.48	0.155	29	6, 96, 24, 28, 15, 83, 8, 153, 46, 3, 20, 7, 3, 76, 16, 12, 42, 5, 13, 51, 30, 31, 15, 15, 25, 9, 17, 11, 16
His68Tyr	Caffeine	-0.03	0.100	3	16, 24, 18
	DTT	< -0.50	0.188	3	23, 26, 26
	HU	-0.48	0.111	3	15, 12, 20

369

370

### Specific Residues Connect Different Stresses to Ub Protein-Protein Interactions

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

We also investigated specific signals outside of the shared sensitizing response (**Figure 8B**). We identified outlier mutations by comparing the change in fitness scores of each of the sensitizing perturbations (**Figure 8 – Table Supplement 2, 3**). Because these mutants are not sensitized by all of the perturbations they likely alter binding to specific adaptors, conjugation machinery, or substrates. Most of these outliers represent mutants that are tolerated in Caffeine and HU, but sensitized by DTT treatment. However, the H68Y mutation differs as DTT and HU treatments sensitize this mutation whereas Caffeine treatment does not. His68 is an important position found at the interface between Ub and adaptor domains such as UIM and UBA domains. These domains are important for the trafficking of ubiquitinated proteins. His68 lies adjacent to the hydrophobic patch and binding to UIM domains is reduced when it is protonated (Fujiwara et al., 2004). In contrast, when His68 is mutated to Val in Ub, the binding to UIM domains is increased, likely mimicking the deprotonated state that forms a hydrophobic surface (Fujiwara et al., 2004).

Lys11 is similarly important for Ub biology and shows a specific sensitization to DTT. Lys11 linked poly-Ub chains are the second most abundant linkage in yeast. These chains likely signal for degradation at the proteasome, like Lys48 linked chains, and have been implicated in the response to ER stress (Xu et al., 2009). In DMSO all substitutions, except to negative and aromatic residues, are tolerated. However, substitutions to Leu, Ile, His and Asn are sensitized uniquely in DTT. These data suggest that Lys11 is mediating an interaction to DTT induced stress. Although previous studies have indicated a synthetic lethal interaction between Lys11Arg and DTT (Xu et

390 al., 2009), in our experiments, at lower DTT concentrations, the relatively high fitness of  
391 Lys11Arg suggests that the structural role of the positively charged residues and not poly-Lys11  
392 Ub linkages may dominate the physiological response.

393 In addition to fitness defects that are likely due to perturbing Ub/protein interfaces, we also  
394 observed defects due to perturbing poly-Ub chain structure and dynamics. Lys63 linked poly-Ub  
395 chains exist in three distinct conformations in solution (Liu et al., 2015). The populations of these  
396 conformational states help determine binding partner selection between Lys63 linked chains and  
397 adaptor proteins. Mutating Glu64 to Arg biases the chains towards the open conformation (Liu et  
398 al., 2015). In DMSO, the mutation of Glu64 to a positive residue caused a fitness defect. In  
399 Caffeine and HU these mutants are sensitized and the fitness defect is further increased. However,  
400 DTT treatment increased the tolerance to positive mutations at this position, again suggesting an  
401 interaction between Lys63 linked poly-Ub and DTT treatment.

402  
403

## 404 **DISCUSSION**

405

406 We have determined the fitness landscape of Ub in yeast grown in the presence of five  
407 chemical perturbations. We identified newly sensitized positions in the protein, which supports the  
408 hypothesis that the Ub sequence is highly constrained by its role in a wide array of environmental  
409 stress responses. Although each perturbation had some unique features, we observed a general  
410 buffering effect that may have obscured mutational sensitivity in the previously determined Ub  
411 fitness landscape.

412 Perhaps the most surprising result in our study was the failure to recapitulate the synthetic  
413 lethal interaction between Lys11Arg and DTT (Xu et al., 2009). This interaction was observed  
414 using the same strain (SUB328), however fitness was determined through a dilution spot assay on  
415 an agar plate containing 30mM DTT. Our experiments were conducted in liquid culture with 1mM  
416 DTT refreshed every sampling period. It is likely that we did not achieve a stress regime where  
417 Lys11 poly-Ub is essential for DTT tolerance. The Lys11Arg mutation induces the upregulation of  
418 proteins involved in ERAD including Ubc6, the ERAD E2. Also, the turnover of known ERAD  
419 substrates is unaffected by the Lys11Arg mutation, suggesting that Lys48 linked chains can be  
420 substituted for Lys11 linked chains (Xu et al., 2009). These adaptations could be sufficient to  
421 counteract the loss of Lys11 poly-Ub in our experiments, but are insufficient at higher  
422 concentrations of DTT. It would be interesting to explore these two regimes and determine the  
423 concentration of DTT that induces the lethality of the Lys11Arg mutant.

424 Taken together, these data represent a step towards understanding the apparent dichotomy  
425 between the Ub conservation and the previously determined Ub fitness landscape. While much of  
426 the protein is tolerant to mutation when cells are grown with traditional laboratory conditions, stress  
427 conditions reveal hidden mutational sensitivity. We show that thirteen new positions are extremely  
428 sensitized in at least one stress condition with an additional thirteen new positions intermediately  
429 sensitized. While the incorporation of these new stresses provides a rationale for an additional 1/3  
430 of the protein, we cannot currently explain the conservation of some positions in the “tolerant” face  
431 of the protein. Expanding the set of chemical perturbations assayed may begin to address this  
432 dichotomy further. It is also possible that mutations at tolerant positions create fitness defects that  
433 are too subtle to be determined by our methods. These subtle defects can lead to the sequence

434 conservation observed in Ub when a large population undergoes selection over a longer  
435 evolutionary time (Boucher et al., 2014).

436 These experiments also demonstrate the success of graduate-level project based courses  
437 (Vale et al., 2012) as key components of a first-year curriculum. Our students were able to  
438 generate high quality data and useable computational pipelines during the 8 weeks of class time.  
439 These successes are notable because few students began the class with a background in both  
440 areas. By creating a project lab environment that encouraged team based learning and teaching,  
441 we enabled students to quickly acquire relevant skills within the context of an active research  
442 project. The wide variety of stress responses that Ub mediates and the vast chemical space that  
443 can be safely and economically addressed in a classroom make yeast and Ub ideal systems for  
444 continuing these studies. It is our hope that other graduate programs can similarly offer project  
445 based classes in their curriculums and we will make our reagents and pipeline available for use to  
446 further that goal.

447

## 448 **ACKNOWLEDGEMENTS**

449

450 We acknowledge: administrative support from Rebecca Brown, Julia Molla, and Nicole Flowers;  
451 technical support from Jennifer Mann and Manny De Vera; gifts from David Botstein, and Illumina;  
452 and helpful discussions with Hana El-Samad, Nevan Krogan, Danielle Swaney, and Ron Vale. The  
453 Project Lab component of this work is specifically supported by an NIBIB T32 Training Grant,  
454 “Integrative Program in Complex Biological Systems” (T32-EB009383). UCSF iPQB and CCB  
455 Graduate programs are supported by US National Institutes of Health (NIH) grants EB009383,  
456 GM067547, GM064337, and GM008284, HHMI/NIBIB (56005676), UCSF School of Medicine,  
457 UCSF School of Pharmacy, UCSF Graduate Division, UCSF Chancellors Office, and Discovery  
458 Funds. S.T., W.C., and L.S.M. are supported by NSF Graduate Research Fellowships. E.M.G. is  
459 supported by a Kellogg Chancellor Fellowship. D.N.B. is supported by NIH GM112844. J.S.F. is a  
460 Searle Scholar, Pew Scholar, and Packard Fellow, and is supported by NIH OD009180.

461

462 The authors declare that no competing interests exist.

463

## 464 **METHODS**

465

466 Additional material is available on our website ([www.fraserlab.com/pubs\\_2014](http://www.fraserlab.com/pubs_2014)) and GitHub  
467 (<https://github.com/fraser-lab/PUBS2014>).

468

### 469 **Yeast Library:**

470 Yeast strain SUB328 (MATa *lys2-801 leu2-3,2-112 ura3-52 his3-Δ200 trp1-1 ubi1-Δ1::TRP1 ubi2-*  
471 *Δ2::ura3 ubi3-Δub-2 ubi4-Δ2::LEU2* (pUB146) (pUB100)) was used, which expresses ubiquitin  
472 from a galactose-inducible promoter in pUB146. pUB100 expresses the Ub1 tail. A library of  
473 ubiquitin genes was saturated with point mutations (Roscoe et al., 2013). Barcodes were added by  
474 ligating N18 oligos flanked by EagI and AscI sites into each of the eight previously create Ub  
475 libraries. These libraries were bottlenecked by transformation into *E. coli* and then pooled to create  
476 the single N18BC-UbLib. This pooled library was transformed into *E. coli* to create the final N18BC-  
477 UbLib.

478

479 **Barcode Association PCR/Library/Sequencing:**

480 To associate the N18BCs to a given Ub allele, we performed a paired end read on the Illumina  
481 MiSeq. Because Ub is a small gene, we were able to read the entire ORF with a 260 bp read and  
482 the associated N18BC with a 30 bp read. To prepare the library for sequencing, plasmid DNA was  
483 extracted from e. coli using the Omega Bio-Tek mini-prep kit. A ~700 bp product was amplified with  
484 primers containing the Illumina PE1 and PE2 primer sequences for 9 cycles to minimize PCR  
485 recombination. These products were separated on agarose gel, and excised products were  
486 purified by silica column. This library was prepared for sequencing on the Illumina MiSeq.

487

488 **Drug Concentration:**

489 The concentration to reduce the growth rate of SUB328 (WT Ub) by 25% was determined by  
490 monitoring the growth of cells by optical density measurements at 600nm over 8 hours. MG132  
491 and DMSO did not affect SUB328 (WT Ub) growth rate at any tested concentration. Hydroxyurea  
492 treatment induces a lag-phase followed by WT like growth.

493

494 **EMPIRIC-BC**

495 **Transformation:**

496 SUB328 strain was independently transformed three times with the barcoded Ub library. Two of  
497 these transformations (LibA, LibB) were transformed with the LiAc method described previously  
498 (Gietz and Woods, 2002). The third library (LibC) was transformed using the hybrid  
499 LiAc/electroporation protocol described previously (Benatuil et al., 2010). Libraries were grown in  
500 log phase for 48h @ 30°C in SRGal (synthetic, 1% raffinose, 1% galactose) + G418 and ampicillin  
501 and then flash frozen in LN2 at late log phase and stored at -80°C as 1 mL aliquots.

502

503 **Library Growth and Sample Collection:**

504 Frozen aliquots were thawed and grown in 50 mL SRGal +G418 in log-phase for 48h. The library  
505 was transferred into SD (synthetic, 2% glucose) + G418 as described (Roscoe et al., 2013). The  
506 library was maintained in log-phase for 12 hours in SD + G418, at which time an initial sample was  
507 collected as described (Roscoe et al., 2013). The libraries were then maintained in log-phase  
508 growth by diluting cells into fresh SD +G418 every 12 hours, in the presence of the perturbation.  
509 The perturbation was refreshed with each dilution. Samples were taken every 2-3 SUB328 (WT  
510 Ub) generations.

511

512 **PCR and Miniprep:**

513 Plasmid DNA was extracted from yeast and prepared for deep sequencing. Yeast pellets were  
514 thawed and lysed and plasmid DNA recovered as previously described (Roscoe et al., 2013). A  
515 268 bp product was amplified from the plasmids by PCR, using only 9 cycles of amplification. This  
516 product contained the N18BC. PCR products were separated on agarose gel, and excised  
517 products were purified by silica column. A second round of PCR was performed to add unique  
518 indices (Illumina TruSeq) to barcode each sample.

519

520 **Sequencing and Data analysis**

521 Each PCR product was quantified by qBit and diluted to 4nM. The samples were then pooled and  
522 the pooled libraries prepared for sequencing on the Illumina HiSeq. The N18BCs were sequenced

523 with a single end HiSeq run with a custom primer (TGCAGCGGCCCTGAGTCCTGCC) that read  
524 directly into the N18BC. Samples were indexed using the HiSeq indexing read and the Illumina  
525 TruSeq indices.

526

527 **Pipeline:**

528 Module 0: Sub assembly

529 Script1:

530 01\_sele\_BC.py paired\_end\_read1.fasq > good\_BC\_reads.fastq

531

532 This script takes a raw fastq file (Illumina output) and checks each sequence to see if it matches  
533 the expected Ub-Library vector sequence after the N18 bar-code Input file should be the Read1  
534 output file of a paired end Illumina read. The output is matched sequences in fastq format printed  
535 to the terminal. The script will also write a log file named "Script01\_logfile.txt" by default

536

537 02\_pair\_reads.py good\_BC\_reads.fastq paired\_end\_read2.fastq -o pair\_dict.pkl

538

539 This script takes the output fastq from 01\_sele\_BC.py and creates a dictionary keyed on the  
540 sequence sample ID. It then takes the full raw read2 fastq and associates the sample N18 bar-  
541 code with the Ub sequence from read2. The output is a dictionary keyed on the sample ID with  
542 values as a 2 item list. the first entry is the N18 bar-code (pair\_dict[identity\_key][0]). The second is  
543 a list of the Ub sequence from read2 (pair\_dict[identity\_key][1]).

544

545 pair\_dict.pkl

546 {'SampleID':['Barcode', 'Ub\_sequence'], ...}

547

548 03\_sequences\_assigned\_to\_barcode.py pair\_dict.pkl -o barcode\_to\_Ub.pkl

549

550 This script takes the output from 02\_pair\_reads.py and associates a given N18 barcode with all the  
551 related ubiquitin sequences. It then returns a dictionary that is keyed on the barcode with values of  
552 a list of all associated ubiquitin reads.

553

554 barcode\_to\_Ub.pkl

555 {'Barcode': ['Ub\_sequence1', 'Ub\_sequence2', ...] ...}

556

557 04\_generate\_consensus.y barcode\_to\_Ub.pkl -o Allele\_Dictionary.pkl

558

559 This script takes the output from 03\_sequences\_assigned\_to\_barcode.py and generates a  
560 consensus sequence from the list of Ub sequences associated with a given barcode. The mutant  
561 in the consensus sequence is identified and associated with the barcode. A barcode must be  
562 observed at least 3 times and the consensus sequence must contain only one mutation to be  
563 included. The output is a dictionary keyed on the barcode with a tuple value of  
564 (int(amino\_acid\_position), str(mutant\_codon))

565

566 Allele\_Dictionary.pkl

567 {"barcode" : (aa\_position\_in\_Ub, Mutant\_Codon)}

568 “AGCTCTA” : (74, AUU)

569 “AGCCCTA” : (5, GCU)}

570

571

572 Module 1: Extract BC counts from fastq with Hamming error correction:

573 Requires seqmatch.py to be present in the working directory. The below scripts use function

574 imported from this file

575

576 Script1:

577 fastq\_index\_parser\_v4.py data.fastq --indices barcodes.txt -o indexed\_data.pkl --index\_cutoff 2 --

578 const\_cutoff 2

579

580 This will take the fastq files from a sequence run as input and create dictionaries that are keyed on  
581 the sample index and have values of barcode:quality score. The index and const cutoffs are  
582 Hamming distance cutoffs for the sample index (2 is acceptable because all TS BCs used are  
583 greater than 2 Hamming distance apart) and constant region of the vector (again 2 is acceptable  
584 because we are matching to a known constant region of length 8)

585

586 data.fastq - fastq formatted file directly from the sequencer

587

588 barcodes.txt - a table delimited file with 2 columns, the first being the name of the index and the  
589 second being the DNA sequence of the index

590 TS1 CGTGAT

591 TS2 ACATCG

592 TS3 GCCTAA

593

594 indexed\_data.pkl

595 {TS1:{barcode:quality-score, ...}, TS2:{barcode:quality-score, ...}, ..., }

596 {TS1: {'AGCTCTA': '\*55CCF>', ...} ...}

597

598 Script2:

599 pkl\_barcode\_parser.py indexed\_data.pkl --out\_pickle indexed\_data\_counts.pkl --allele\_pickle

600 Allele\_Dictionary.pkl --fuzzy\_cutoff 2

601

602 This script takes in the pkl file produced by the previous script for each sample and checks the  
603 fastq quality scores and matches the sequenced barcodes to those identified by the assembly of  
604 the library and counts the number of times a barcode is observed. This script also uses the  
605 Hamming distance between an observed barcode and members of the Allele\_Dictionary to assign  
606 counts to previously observed barcodes even if a sequencing error occurred in a given read. The  
607 “fuzzy\_cutoff” parameter sets the max Hamming distance considered.

608

609 indexed\_data\_counts.pkl

610 {TS1:{barcode:number-of-reads, ...}, TS2:{barcode:number-of-reads, ...}, ...}

611 {TS1: {'AGCTCTA': 147, ...} ...}

612

613 Script 3:

614 pickleread.py indexed\_data\_counts\_1.pkl ... indexed\_data\_counts\_N.pkl --out\_dir output\_files/ --  
615 pkl\_basename TS\_ --allele\_pickle Allele\_Dictionary.pkl

616

617 This script takes the output from multiple runs of "pkl\_barcode\_parser.py" and combines the  
618 counts. This will result in one dictionary for each sample index with barcodes sequence as key and  
619 the number of reads as values.

620

621 The output files (pkl) will be as follows (30 pkl files):

622 TS1:{barcode:number-of-reads, ...}

623 TS2:{barcode:number-of-reads, ...}

624 ...

625 TSN:{barcode:number-of-reads, ...}

626

627 Module 2: Initial scoring - Barcodes, initial counts cutoff = 3

628 Script1:

629 pickle\_condensing.py TS\_1.pkl TS\_2.pkl TS\_3.pkl perturbation replicate -o Barcode\_Counts.pkl

630

631 This script simply takes the counts from the dictionaries created by Module 0 Script 3 and  
632 combines them into a single dictionary that contains the counts for a given barcode for all 3  
633 samples that describe an experiment. The perturbation and replicate inputs are used in naming the  
634 output dictionary.

635

636 Script2:

637 Score\_BCs.py Barcode\_Counts.pkl Allele\_Dictionary.pkl time1 time2 -o Barcode\_scores.pkl

638

639 Barcode\_Counts.pkl

640 { "barcode" : [count\_sample1, count\_sample2...]}

641 "AGCTCTA" : [15, 3, 1]

642 "AGCCCTA" : [222, 23, 21]}

643

644

645 This script takes a .pkl of a dictionary (Barcode\_Counts.pkl) keyed on the sample barcodes with  
646 values of a list of counts at each time point. The scoring function uses these counts and scores  
647 them based on the time values (in WT generations) - the relative fitness is compared to wild type  
648 barcodes, which are distinguished in Allele\_Dictionary.pkl. Output is a dictionary keyed on sample  
649 barcode with values of the fitness score. Fitness scores are determined by calculating the slope of  
650 the regression line of the three counts for each barcode. The score reported is  $\log_2(\text{Mutant Slope}/$   
651  $\text{WT Slope})$ . Any barcode that is observed three or less times in the initial sample is not used in the  
652 fitness score calculation

653

654 Barcode\_scores.pkl

655 { "barcode" : float(Fitness\_score)}

656 "AGCTCTA" : -0.56

657 "AGCCCTA" : -0.1}

658

659 Module 3: Outlier detection and removal

660 `toss_outliers.py Barcode_scores.pkl codon 4 -o clean_BCs.pkl`

661

662 `clean_BCs.pkl`

663 `{“dirty_barcodes”: [(aa_position_in_Ub, Mutant_Codon, barcode) ...]}`

664 `“clean_barcodes”: [(aa_position_in_Ub, Mutant_Codon, barcode) ...]}`

665 `“dirty_barcodes”: [(74, AUU, “AGCTCTA”), (21, CCC, “ACTTCTA”) ...]`

666 `“clean_barcodes”: [(5, GCU, “AGCCCTA”), (21, UUU, “GCATTTTC”) ...]}`

667

668 This script compares the scores of barcodes mapping to the same allele. The median absolute  
669 deviation (MAD) is calculated for the barcodes that map to the same allele. Outlier barcodes are  
670 determined as scores that are greater than or equal to 1.5 times the interquartile range of the  
671 distribution and removed. The codon flag tells the script to compare all scores mapping to the  
672 same codon. The 4 flag sets the minimum number of barcodes before the MAD will be performed.  
673 The output is a pkl of barcodes to be kept in the dataset.

674

675 `remove_bad_BCs.py Barcode_scores.pkl clean_BCs.pkl Allele_Dictionary.pkl -o`

676 `Barcode_scores_outliers_removed.pkl`

677 This script checks the Barcode\_scores dictionary against the clean\_BCs returned by  
678 `toss_outliers.py` and removes those BCs that are not in the clean\_BCs.pkl. The script returns a  
679 dictionary keyed on sample barcode with fitness scores as values but with outlier BCs removed.

680

681 `Barcode_scores_outliers_removed.pkl`

682 `{ “barcode” : float(Fitness_score)}`

683 `“AGCCCTA” : -0.1`

684 `“GCATTTTC” : -0.67}`

685

686 Module 4: Create matrix

687 `heatmap_BCs.py Barcode_scores_outliers_removed.pkl Allele_Dictionary.pkl -o`

688 `Barcode_scores_outliers_removed_matrix.pkl`

689 This script takes the Barcode scores and averages them to the amino acid level. It then outputs  
690 these scores as a heatmap and as a numpy matrix pkl.

691

692 `Barcode_scores_outliers_removed_matrix.pkl`

693 `masked_array(data = [21X76 matrix containing fitness scores for each aa substitution])`

694

695 `[[- -0.631791406846642 -0.5397724613430753 ..., -0.3530569856099873`

696 `-0.4209070611721436 --]`

697 `[-- -0.04155544432657808 -- ..., -- -0.4672829444990562`

698 `-0.015863306980341812]`

699 `[-- -0.06881685222913404 -0.3000996826283508 ..., -0.09038257104760622`

700 `-0.5060198247122988 --]`

701

702 `[-- -0.2136845391374962 -0.5846954623554699 ..., -0.5201027046981986 -- --]`

703 [-- -0.037103840372513595 -0.6445621511224743 ..., -0.6398418859301449  
704 -0.5009308293341608 --]  
705 [-- -0.03813077561871194 -0.5946959237696324 ..., -0.5684471534238328  
706 -0.3959407495759722 --]]

707

### 708 **Rosetta predictions of ubiquitin stability changes upon point mutations:**

709 We used Rosetta version number 55534 for all simulations. The Rosetta software can be  
710 downloaded at [www.rosettacommons.org](http://www.rosettacommons.org).

711 Using the crystal structure of human ubiquitin (1UBQ) as input, we first introduced three mutations  
712 to match the *S. cerevisiae* sequence using Rosetta fixed backbone design:

713

#### 714 **Command line:**

715 `fixbb.linuxgccrelease -s 1UBQ.pdb -resfile UBQ_to_yeast.res -ex1 -ex2 -extrachi_cutoff 0 -`  
716 `nstruct 1 -overwrite -linmem_ig 10 -minimize_sidechains`

717

#### 718 **UBQ\_to\_yeast.res file contents:**

719 NATRO

720 start

721 19 A PIKAA S

722 24 A PIKAA D

723 28 A PIKAA S

724

725 We then followed a protocol described by Kellogg & coworkers (Kellogg et al., 2011) for estimating  
726 stability changes in monomeric proteins in response to point mutations. For documentation of the  
727 protocol and file formats (`mut_file`, `cst_file`), see

728 [https://www.rosettacommons.org/docs/latest/application\\_documentation/analysis/ddg-monomer](https://www.rosettacommons.org/docs/latest/application_documentation/analysis/ddg-monomer)

729

730 The first step minimizes the input structure (the model of yeast ubiquitin generated above,  
731 1UBQ\_0001.pdb):

732

#### 733 **Command line** (weights file `sp2_paper_talaris2013_scaled.wts` in supplement):

734 `minimize_with_cst.static.linuxgccrelease -s 1UBQ_0001.pdb -in:file:fullatom -`  
735 `ignore_unrecognized_res -fa_max_dis 9.0 -ddg::harmonic_ca_tether 0.5 -`  
736 `ddg::constraint_weight 1.0 -ddg::out_pdb_prefix min_cst_0.5 -ddg::sc_min_only false -`  
737 `score::bonded_params 300 150 40 40 40 -scale_d 1 -scale_theta 1 -scale_rb 1 -`  
738 `score:weights sp2_paper_talaris2013_scaled.wts`

739

740 The second step performs the stability calculations:

741

#### 742 **Command line:**

743 `ddg_monomer.static.linuxgccrelease -in:file:s 1UBQ_minimized.pdb -ddg::mut_file (mutfile) -`  
744 `constraints::cst_file (cst_file) -ignore_unrecognized_res -in:file:fullatom -fa_max_dis 9.0 -`  
745 `ddg::dump_pdbs true -ddg::suppress_checkpointing true -ddg:weight_file soft_rep_design -`  
746 `ddg::iterations 50 -ddg::local_opt_only false -ddg::min_cst true -ddg::mean false -ddg::min`  
747 `true -ddg::sc_min_only false -ddg::ramp_repulsive true -score::bonded_params 300 150 40`

748 40 40 -scale\_d 1 -scale\_theta 1 -scale\_rb 1 -ddg:minimization\_scorefunction  
749 sp2\_paper\_talaris2013\_scaled.wts

750

## 751 REFERENCES

- 752 Balch, W.E., Morimoto, R.I., Dillin, A., and Kelly, J.W. (2008). Adapting proteostasis for disease  
753 intervention. *Science* 319, 916-919.
- 754 Benatuil, L., Perez, J.M., Belk, J., and Hsieh, C.M. (2010). An improved yeast transformation  
755 method for the generation of very large human antibody libraries. *Protein engineering, design &  
756 selection : PEDS* 23, 155-159.
- 757 Boucher, J.I., Cote, P., Flynn, J., Jiang, L., Laban, A., Mishra, P., Roscoe, B.P., and Bolon, D.N.  
758 (2014). Viewing protein fitness landscapes through a next-gen lens. *Genetics* 198, 461-471.
- 759 Erpapazoglou, Z., Walker, O., and Haguenuer-Tsapis, R. (2014). Versatile roles of k63-linked  
760 ubiquitin chains in trafficking. *Cells* 3, 1027-1088.
- 761 Finley, D., Ulrich, H.D., Sommer, T., and Kaiser, P. (2012). The ubiquitin-proteasome system of  
762 *Saccharomyces cerevisiae*. *Genetics* 192, 319-360.
- 763 Fowler, D.M., Stephany, J.J., and Fields, S. (2014). Measuring the activity of protein variants on a  
764 large scale using deep mutational scanning. *Nature protocols* 9, 2267-2284.
- 765 Frand, A.R., and Kaiser, C.A. (1998). The ERO1 gene of yeast is required for oxidation of protein  
766 dithiols in the endoplasmic reticulum. *Molecular cell* 1, 161-170.
- 767 Friedlander, R., Jarosch, E., Urban, J., Volkwein, C., and Sommer, T. (2000). A regulatory link  
768 between ER-associated protein degradation and the unfolded-protein response. *Nature cell biology*  
769 2, 379-384.
- 770 Fujiwara, K., Tenno, T., Sugasawa, K., Jee, J.G., Ohki, I., Kojima, C., Tochio, H., Hiroaki, H.,  
771 Hanaoka, F., and Shirakawa, M. (2004). Structure of the ubiquitin-interacting motif of S5a bound to  
772 the ubiquitin-like domain of HR23B. *The Journal of biological chemistry* 279, 4760-4767.
- 773 Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and  
774 Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental  
775 changes. *Molecular biology of the cell* 11, 4241-4257.
- 776 Gietz, R.D., and Woods, R.A. (2002). Transformation of yeast by lithium acetate/single-stranded  
777 carrier DNA/polyethylene glycol method. *Methods in enzymology* 350, 87-96.
- 778 Hietpas, R., Roscoe, B., Jiang, L., and Bolon, D.N. (2012). Fitness analyses of all possible point  
779 mutations for regions of genes in yeast. *Nature protocols* 7, 1382-1396.
- 780 Ibarra-Molero, B., Loladze, V.V., Makhatadze, G.I., and Sanchez-Ruiz, J.M. (1999). Thermal  
781 versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from  
782 charge-charge interactions to protein stability. *Biochemistry* 38, 8138-8149.
- 783 Jensen, T.J., Loo, M.A., Pind, S., Williams, D.B., Goldberg, A.L., and Riordan, J.R. (1995). Multiple  
784 proteolytic systems, including the proteasome, contribute to CFTR processing. *Cell* 83, 129-135.

- 785 Jiang, L., Mishra, P., Hietpas, R.T., Zeldovich, K.B., and Bolon, D.N. (2013). Latent effects of  
786 Hsp90 mutants revealed at reduced expression levels. *PLoS genetics* 9, e1003600.
- 787 Kellogg, E.H., Leaver-Fay, A., and Baker, D. (2011). Role of conformational sampling in computing  
788 mutation-induced changes in protein structure and stability. *Proteins* 79, 830-838.
- 789 Koc, A., Wheeler, L.J., Mathews, C.K., and Merrill, G.F. (2004). Hydroxyurea arrests DNA  
790 replication by a mechanism that preserves basal dNTP pools. *The Journal of biological chemistry*  
791 279, 223-230.
- 792 Komander, D., and Rape, M. (2012). The ubiquitin code. *Annual review of biochemistry* 81, 203-  
793 229.
- 794 Kortemme, T., and Baker, D. (2002). A simple physical model for binding energy hot spots in  
795 protein-protein complexes. *Proceedings of the National Academy of Sciences of the United States*  
796 *of America* 99, 14116-14121.
- 797 Lange, O.F., Lakomek, N.A., Fares, C., Schroder, G.F., Walter, K.F., Becker, S., Meiler, J.,  
798 Grubmuller, H., Griesinger, C., and de Groot, B.L. (2008). Recognition dynamics up to  
799 microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320, 1471-  
800 1475.
- 801 Lee, D.H., and Goldberg, A.L. (1996). Selective inhibitors of the proteasome-dependent and  
802 vacuolar pathways of protein degradation in *Saccharomyces cerevisiae*. *The Journal of biological*  
803 *chemistry* 271, 27280-27284.
- 804 Lindquist, S.L., and Kelly, J.W. (2011). Chemical and biological approaches for adapting  
805 proteostasis to ameliorate protein misfolding and aggregation diseases: progress and prognosis.  
806 *Cold Spring Harbor perspectives in biology* 3.
- 807 Liu, Z., Gong, Z., Jiang, W.X., Yang, J., Zhu, W.K., Guo, D.C., Zhang, W.P., Liu, M.L., and Tang,  
808 C. (2015). Lys63-linked ubiquitin chain adopts multiple conformational states for specific target  
809 recognition. *eLife* 4.
- 810 Peng, J., Schwartz, D., Elias, J.E., Thoreen, C.C., Cheng, D., Marsischky, G., Roelofs, J., Finley,  
811 D., and Gygi, S.P. (2003). A proteomics approach to understanding protein ubiquitination. *Nature*  
812 *biotechnology* 21, 921-926.
- 813 Petermann, E., Orta, M.L., Issaeva, N., Schultz, N., and Helleday, T. (2010). Hydroxyurea-stalled  
814 replication forks become progressively inactivated and require two different RAD51-mediated  
815 pathways for restart and repair. *Molecular cell* 37, 492-502.
- 816 Phillips, A.H., Zhang, Y., Cunningham, C.N., Zhou, L., Forrest, W.F., Liu, P.S., Steffek, M., Lee, J.,  
817 Tam, C., Helgason, E., *et al.* (2013). Conformational dynamics control ubiquitin-deubiquitinase  
818 interactions and influence in vivo signaling. *Proceedings of the National Academy of Sciences of*  
819 *the United States of America* 110, 11379-11384.
- 820 Powers, E.T., Morimoto, R.I., Dillin, A., Kelly, J.W., and Balch, W.E. (2009). Biological and  
821 chemical approaches to diseases of proteostasis deficiency. *Annual review of biochemistry* 78,  
822 959-991.

- 823 Reinke, A., Chen, J.C., Aronova, S., and Powers, T. (2006). Caffeine targets TOR complex I and  
824 provides evidence for a regulatory link between the FRB and kinase domains of Tor1p. The  
825 *Journal of biological chemistry* *281*, 31616-31626.
- 826 Rock, K.L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D., and Goldberg, A.L.  
827 (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation  
828 of peptides presented on MHC class I molecules. *Cell* *78*, 761-771.
- 829 Roscoe, B.P., and Bolon, D.N. (2014). Systematic exploration of ubiquitin sequence, E1 activation  
830 efficiency, and experimental fitness in yeast. *Journal of molecular biology* *426*, 2854-2870.
- 831 Roscoe, B.P., Thayer, K.M., Zeldovich, K.B., Fushman, D., and Bolon, D.N. (2013). Analyses of  
832 the effects of all ubiquitin point mutants on yeast growth rate. *Journal of molecular biology* *425*,  
833 1363-1377.
- 834 Sharp, P.M., and Li, W.H. (1987). Ubiquitin genes as a paradigm of concerted evolution of tandem  
835 repeats. *Journal of molecular evolution* *25*, 58-64.
- 836 Sloper-Mould, K.E., Jemc, J.C., Pickart, C.M., and Hicke, L. (2001). Distinct functional surface  
837 regions on ubiquitin. *The Journal of biological chemistry* *276*, 30483-30489.
- 838 Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the  
839 polyubiquitin proteolytic signal. *The EMBO journal* *19*, 94-102.
- 840 Vale, R.D., DeRisi, J., Phillips, R., Mullins, R.D., Waterman, C., and Mitchison, T.J. (2012).  
841 Graduate education. Interdisciplinary graduate training in teaching labs. *Science* *338*, 1542-1543.
- 842 Wanke, V., Cameroni, E., Uotila, A., Piccolis, M., Urban, J., Loewith, R., and De Virgilio, C. (2008).  
843 Caffeine extends yeast lifespan by targeting TORC1. *Molecular microbiology* *69*, 277-285.
- 844 Wintrode, P.L., Makhatadze, G.I., and Privalov, P.L. (1994). Thermodynamics of ubiquitin  
845 unfolding. *Proteins* *18*, 246-253.
- 846 Xu, P., Duong, D.M., Seyfried, N.T., Cheng, D., Xie, Y., Robert, J., Rush, J., Hochstrasser, M.,  
847 Finley, D., and Peng, J. (2009). Quantitative proteomics reveals the function of unconventional  
848 ubiquitin chains in proteasomal degradation. *Cell* *137*, 133-145.
- 849 Zhang, W., Qin, Z., Zhang, X., and Xiao, W. (2011). Roles of sequential ubiquitination of PCNA in  
850 DNA-damage tolerance. *FEBS letters* *585*, 2786-2794.
- 851 Zolk, O., Schenke, C., and Sarikas, A. (2006). The ubiquitin-proteasome system: focus on the  
852 heart. *Cardiovascular research* *70*, 410-421.
- 853 Zuin, A., Isasa, M., and Crosas, B. (2014). Ubiquitin signaling: extreme conservation as a source of  
854 diversity. *Cells* *3*, 690-701.
- 855  
856