Analyses of the Effects of All Ubiquitin Point Mutants on Yeast Growth Rate

Benjamin P. Roscoe¹, Kelly M. Thayer², Konstantin B. Zeldovich², David Fushman³ and Daniel N. A. Bolon¹

 Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA
Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA
Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742, USA

Correspondence to Daniel N. A. Bolon: dan.bolon@umassmed.edu http://dx.doi.org/10.1016/j.jmb.2013.01.032 Edited by S. Koide

Abstract

The amino acid sequence of a protein governs its function. We used bulk competition and focused deep sequencing to investigate the effects of all ubiguitin point mutants on yeast growth rate. Many aspects of ubiguitin function have been carefully studied, which enabled interpretation of our growth analyses in light of a rich structural, biophysical and biochemical knowledge base. In one highly sensitive cluster on the surface of ubiquitin, almost every amino acid substitution caused growth defects. In contrast, the opposite face tolerated virtually all possible substitutions. Surface locations between these two faces exhibited intermediate mutational tolerance. The sensitive face corresponds to the known interface for many binding partners. Across all surface positions, we observe a strong correlation between burial at structurally characterized interfaces and the number of amino acid substitutions compatible with robust growth. This result indicates that binding is a dominant determinant of ubiquitin function. In the solvent-inaccessible core of ubiquitin, all positions tolerated a limited number of substitutions, with hydrophobic amino acids especially interchangeable. Some mutations null for yeast growth were previously shown to populate folded conformations indicating that, for these mutants, subtle changes to conformation caused functional defects. The most sensitive region to mutation within the core was located near the C-terminus that is a focal binding site for many critical binding partners. These results indicate that core mutations may frequently cause functional defects through subtle disturbances to structure or dynamics.

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Analyses of protein sequence–function relationships provide a powerful approach to understand mechanism. Mutational studies provide information on the functional impact of specific chemical changes to the protein. Systematic analyses of point mutations provide a detailed map of chemical space that can be mined to infer mechanism. While it has been possible to generate libraries of point mutants for many years,¹ until recently, it had only been feasible to analyze the function of systematic mutations using amber suppresser strains of *Escherichia coli.*² Functional analyses of mutant libraries in non-suppresser systems can now be performed in high throughput by utilizing deep sequencing to analyze mixtures of multiple mutants simultaneously. In this approach, sequence profiling, originally by microarray³ and now more commonly by deep sequencing, is used to determine the relative abundance of mutants in response to selective pressures.^{4–12}

To measure the fitness effects in cells of all possible point mutants for regions of genes, we developed an approach we refer to as EMPIRIC.^{5,6} This approach utilizes systematic site saturation libraries that incorporate a single degenerate codon (NNN) in an otherwise wild-type (WT) coding sequence. Thus, all possible point mutants are included in the library design and the vast majority can be observed above background in deep sequencing analyses. We analyze libraries of point

mutants in conditional yeast strains that contain a second copy of the gene whose activity can be tightly regulated. This enables the amplification of mutant libraries in yeast under permissive conditions where growth is not dependent on mutant function. Adjusting conditions to turn off the second copy of the gene then initiates growth competition based on mutant fitness. In previous work, we analyzed a 9-amino-acid loop in yeast Hsp90.⁶ Here, we report EMPIRIC fitness analyses for the entire yeast ubiquitin gene.

Ubiquitin is essential in all eukaryotes where it serves multiple functions via its ability to covalently attach to other proteins.¹³ The covalent attachment of the C-terminus of ubiquitin to lysine side chains is mediated by a series of enzymes referred to as E1, E2 and E3.¹⁴ Multiple ubiquitin molecules can be linked through covalent attachment between the Cterminus of one ubiquitin chain and a lysine from another ubiquitin. Lysine 48 in ubiquitin is the only lysine that is essential for yeast growth.¹⁵ K48-linked poly-ubiquitin serves as a degradation signal¹⁶ with four K48-linked ubiquitins sufficient to target substrates for proteasome-mediated degradation.¹⁷

Protein degradation by the ubiquitin-proteasome system is an important regulator of the composition of the proteome.¹⁸ As such, the ubiquitin–proteasome system is required for homeostasis under constant conditions, as well as rapid cellular responses to altered external conditions.¹⁹ Protein degradation is often a critical signal in cells. For example, destruction of cyclins serves as the signal for progression through each step of the cell cycle²⁰ and degradation of IkB serves as a key signal in many immune responses.²¹ Disruptions to protein degradation pathways can lead to a variety of disorders including neurodegeneration²² and cancer.¹⁹ Protein degradation pathways have emerged as promising targets for therapeutic drugs, including proteasome inhibitors that are currently in clinical use as anticancer agents.²³

Because of its central role in mediating eukaryotic physiology, ubiquitin has been carefully analyzed by many approaches providing the opportunity to interpret fitness analyses in regard to a wealth of available structural and biochemical information. In particular, noncovalent binding interactions are critical to ubiquitin function. For example, the covalent attachment of ubiquitin depends on noncovalent interfaces between conjugating enzymes and ubiquitin.²⁴⁻²⁷ After covalent attachment to substrates, most known functions of ubiquitin, including delivery of substrates to the proteasome, are mediated by noncovalent binding to ubiquitin-binding proteins.²⁸⁻³⁰ There are many different ubiquitin-binding proteins in all eukaryotes and binding to ubiquitin is frequently mediated by a set of modular ubiquitin-binding domains (UBDs). The most common UBDs in both yeast and humans²⁹

are the ubiquitin-interacting motif³¹ (UIM) that consists of a single α -helix³² and the ubiquitin-associated³³ domain (UBA) that forms a three-helix bundle.^{34,35} Many UBDs bind to a hydrophobic patch on the surface of ubiquitin that includes residues L8, I44 and V70.^{29,36,37}

Alanine scanning of the surface positions in ubiguitin successfully demarcated hotspots for ubiquitin binding partners by identifying 16 residues where substitutions prevented yeast growth, the majority of these positions located in the proximity of the L8. I44 and V70 hydrophobic patch and the K48 and C-terminal sites of covalent linkage.³⁷ Of note, the alanine scan used a binary scoring of mutants (presence or absence of observed growth) and did not quantify potential intermediate growth defects nor did it sample the full diversity of possible mutations leaving many questions about the sensitivity of ubiquitin to surface mutations unknown. For example, are conservative mutations (e.g., lle to Val) to the L8, I44 and V70 hydrophobic patch functional and at positions where alanine substitutions are functional are more severe mutations also tolerated (e.g., Asp-to-Lys charge reversals)?

In addition to binding, the thermodynamics of ubiquitin folding and unfolding have been subject to careful analysis. Ubiquitin is highly stable to temperature denaturation³⁸ and predominantly populates a folded conformation even when subject to near boiling temperature (90 °C) at pH 4. Though it is a small protein of 76 amino acids, native conformations are sufficiently thermodynamically stabilized relative to unfolded conformations that folding is efficient even for many disruptive mutants in the solvent-inaccessible hydrophobic core including all individual alanine substitutions,³⁹ mutations that increase bulk⁴⁰ and some hydrophobic to polar substitutions.^{41,42} Compared to their influence on protein folding, the impact of core mutations on ubiguitin function has not been thoroughly investigated. Of the few core mutants that have been studied functionally, we have previously analyzed Leu-to-Ser mutations at positions 67 and 69 near the C-terminus of ubiquitin. Both of these substitutions were capable of folding but weakened binding affinity to proteasome receptors, resulted in increased accumulation of high-molecular-weight protein species in cells and failed to support veast growth.⁴¹ These results indicated that small changes to the native structure or dynamics of ubiquitin can impair function.

To comprehensively examine both the sensitivity of the ubiquitin surface to mutation and the impact of core mutations on function, we analyzed the impacts of all ubiquitin point mutants on yeast growth rate. On the surface of ubiquitin, there were 10 ultra-sensitive positions where only the WT amino acid was observed to support robust growth. We also observed a cluster of ultra-tolerant positions on the α - helical face of ubiquitin where virtually all amino acid substitutions were compatible with robust growth. Structural analyses of 44 high-resolution co-crystal structures of ubiquitin bound to different partners indicated that burial at interfaces was a good predictor of sensitivity to mutation at surface positions. In the solvent-inaccessible core of ubiquitin, hydrophobic substitutions were generally tolerated. Comparison of mutant effects on growth with previously determined effects on folding stability indicated that some mutants capable of folding were defective for growth. Functional sensitivity to mutation was asymmetrically distributed in the core. Core positions near the C-terminus where many critical binding interactions occur were the most sensitive to mutation. These findings indicate that binding interactions are a dominant contributor to ubiquitin function, which can be impacted by subtle conformational changes and/or dynamics in the folded state of ubiquitin.

Results and Discussion

Bulk competition of ubiquitin mutants in a shutoff strain

To facilitate the analyses of mutants with varied fitness, we used the Sub328 ubiquitin shutoff strain.¹⁵ In Sub328, the only copy of the ubiquitin gene is expressed from a galactose-regulated promoter that generates sufficient ubiquitin protein to support robust growth in galactose media but that is effectively turned off in dextrose media. These properties enable Sub328 cells to host libraries of ubiquitin mutants in galactose media where growth does not require mutant function and subsequently switch to dextrose media where growth is directly related to mutant function (Fig. 1a). To characterize the timing of the shutoff process, we examined the growth of Sub328 cells harboring either a rescue



Fig. 1. Bulk competition analyses of the effect of ubiquitin mutants on yeast growth. (a) Experimental setup: systematic libraries of ubiquitin point mutants generated using saturation mutagenesis at sequential positions within a 9- to 10-aminoacid window were generated on a plasmid with a constitutive promoter. These libraries were introduced into a ubiquitin strain whose only other source of ubiquitin was regulated by a galactose-dependent promoter. Yeast libraries were amplified in galactose and then competed in dextrose where growth relied on the mutant ubiquitin library. (b) Growth of the ubiquitin shutoff strain is rescued by constitutive expression of WT ubiquitin. (c) Positions 40–48 of ubiquitin point mutants at positions 40–48. (d) Stop codons were rapidly depleted, indicating that they were unable to support growth, while silent substitutions that change the nucleotide sequence without altering the protein sequence persisted in shutoff conditions. (e) Correlation between measured growth effects of mutants (selection coefficient) from full experimental repeats.

plasmid constitutively expressing WT ubiquitin (utilizing a promoter and plasmid system previously developed to analyze ubiquitin mutants in yeast¹⁵) or a control plasmid lacking ubiquitin (Fig. 1b). Cells with the rescue plasmid grow rapidly in dextrose media, but cells with the control plasmid stall in growth after about 10 h in dextrose media. Based on these results, we decided to analyze selection on mutant libraries starting after 12 h in dextrose so that most cells without functional mutants would have stalled in growth.

We examined the robustness of bulk competitions by analyzing a 9-amino-acid region (Fig. 1c) of ubiquitin that included K48, the essential lysine involved in forming poly-ubiquitin chains that target substrates to the proteasome. The size of this region enabled it to be efficiently interrogated by Illumina short-read (36 base) sequencing. We generated site saturation libraries for each position in the region, mixed libraries to create a combined library for the region and used focused deep sequencing⁵ to analyze the relative abundance of each point mutant in the combined library. The library was introduced into yeast, expanded in galactose media for 48 h and then switched to dextrose media for 50 h. Samples from the library of yeast were saved at different time points in the competition and the relative abundance of mutants over time determined by sequencing, providing a direct measure of relative mutant fitness.⁶ The fitness effects of WT synonyms (silent mutations) and stop codons (nonsense mutations) served as important internal positive



Fig. 2. Analyses of the growth effects of mutants across the ubiquitin gene. (a) The gene was subdivided into eight regions of 9-10 amino acids and each region was subject to saturation mutagenesis, bulk competition in yeast and deep sequencing analyses. (b) Heat map representation of the effects of ubiquitin mutants on yeast growth. Mutants that were below a conservative detection limit at the beginning of the competition were omitted from fitness analyses. (c) Bimodal distribution of observed mutant effects on yeast growth indicates that most mutants supported either WT-like or null growth in yeast. (d) Distribution of growth effects for mutations that depleted by more than 2-fold during outgrowth in galactose media but remained sufficiently abundant to quantify fitness. Most depleted mutants had null-like fitness and none were WT-like (s>-0.1). (e) Correlation between the growth rate of a panel of individually analyzed mutants relative to fitness measures from bulk competitions.

and negative controls (Fig. 1d). WT synonyms persisted in the bulk competition consistent with the near-neutral expectation for silent mutations. In contrast, stop codons rapidly decreased in relative abundance consistent with the critical function of the C-terminus of ubiquitin in conjugation to substrates.43 The slope of mutant-to-WT ratio versus time in WT generations was calculated and represents the selection coefficient (s) where s=0 indicates WT growth and s=-1 indicates a null mutant. The rapid drop-off in abundance of strongly deleterious mutants meant that this class of mutant could not be quantified as precisely as mutants that persisted in the culture. We performed a full experimental repeat to judge the reproducibility of our bulk fitness measurements. Excluding strongly deleterious mutants (s < -0.5), we observed a strong correlation between repeat measures of the effects of ubiguitin amino acid substitutions on yeast growth rate (Fig. 1e). Compared to fit mutants, strongly deleterious mutants (s < -0.5) showed larger differences in the experimental repeat (Supplementary Fig. 1a). Excluding strongly deleterious mutants, the correlation between repeat measurements indicates that we can accurately resolve growth differences of about 7%. This level of resolution is valuable for investigating the physical constraints on ubiquitin function. However, it is not sufficient to distinguish the full spectrum of selection that would act on natural populations where fitness effects on the order of the inverse of the effective population size (estimated at 10^{-7} in yeast^{44,45}) are subject to effective selection.

Analyzing mutants across the ubiquitin coding sequence

By investigating multiple different regions in parallel (Fig. 2), we were able to analyze all positions in ubiguitin at the same time. We separated the ubiguitin gene into eight regions each encoding 9-10 amino acids that were amenable to our sequencing-based approach (Fig. 2a and Supplementary Table 1). For each region, we generated site saturation libraries that we introduced into shutoff yeast and analyzed by bulk competition and sequencing. Utilizing this approach, we determined fitness effects across the ubiquitin coding sequence (Fig. 2b). In order to assess reproducibility and selection in each region, we analyzed both WT synonyms and stop codons (Supplementary Fig. 1b). In all regions, WT synonyms were consistently highly fit ($s \approx 0$) with narrow distributions (standard deviations ranging from 0.005 to 0.03). These observations indicate that highly fit mutants are accurately interrogated by our procedure as expected because they persist in the culture and are sampled throughout the competition experiment. The average fitness effects of stop codons is similar in each region (Supplementary Fig. 1b), but with

increased measurement variation. In all regions, the average stop codon is highly deleterious (s < -0.65), indicating that selection is strong across all regions. Because highly deleterious mutants rapidly deplete from the culture, they are not sampled as extensively as other mutants and measurement accuracy has an increased dependence on the synchronization of selection pressure across the yeast culture and the number of sequence reads at the early selection time points. The variation in measurements of stop codons is approximately 10 times greater than WT synonyms (standard deviation ranging from 0.07 to 0.1 for seven of the regions). Because of unintended variations in sequencing depth, one region (positions 49-58) had markedly lower number of reads for the early time points (Supplementary Table 1). In this region, the average stop codon was strongly deleterious (s=-0.75), but as expected, measurement variation was large (standard deviation of 0.37). With the exception of this region, the experimental measure of highly deleterious mutants is reasonably precise across the data set.

For amino acids encoded by multiple codons, we calculated fitness as the average over all synonyms (Supplementary Table 2). Across the entire data set, the fitness effects of all nucleotide changes that encode the same amino acid were similar, indicating that protein sequence had a dominant impact on fitness compared to nucleotide sequence. We were able to directly analyze selection coefficients for the majority of possible amino acid substitutions (85%, colored in Fig. 2b). These quantified mutants exhibited a bimodal distribution of fitness effects (Fig. 2c), which has been commonly observed in many different fitness studies.^{2,46–49}

At the first time point analyzed in shutoff selection, the relative abundance of some mutants was below our threshold for accurate analysis (colored gray in the heat map). We considered two potential explanations for this low mutant abundance: poor representation in the saturation mutagenesis and/or depletion during growth in galactose where WT ubiquitin was co-expressed. We deep sequenced the plasmid pool and found that virtually all point mutants (99%) were represented in the plasmid library at relative abundances above our threshold for analysis (Supplementary Table 3). Sequencing of veast samples obtained immediately following amplification in galactose media revealed that many mutations were depleted, indicating that they had a dominant negative growth defect. We observed greater than 2-fold depletion for 95% of mutants that were below the threshold level for fitness analysis (gray boxes in Fig. 2b). We took advantage of mutants that were highly represented in the plasmid library to provide a data set of mutants that depleted during co-expression with WT ubiquitin but whose relative abundance after outgrowth was sufficient to enable accurate fitness measurements. These depletion-prone mutations are universally unfit (Fig. 2d), indicating that depleted mutations that were not overrepresented in the plasmid library (gray filled boxes in Fig. 2b) are likely unfit. The dominant negative growth effects of ubiquitin mutants are both intriguing and mechanistically unclear as they occur at multiple different structural locations. Experiments focused on these mutants will likely be an exciting area of future research. As indicated in Fig. 2b, a small number of mutations were low in abundance in the plasmid library or introduced an internal restriction site that interfered with sample processing (further described in Materials and Methods).

The bulk fitness measurements are in agreement with the known function of ubiquitin. For example, positions 48 and 76 are sites of critical for covalent ubiquitin–ubiquitin linkages and are known to be sensitive to mutation.¹³ In our EMPIRIC analyses, only the WT amino acids at positions 48 and 76 (outlined in red broken lines in Fig. 2b) are compatible with robust growth. There are three amino acid substitutions between the yeast and human versions of ubiquitin (outlined in maroon in



Fig. 3. Effects of mutants on the solvent-accessible surface of ubiquitin on yeast growth. (a) Distribution of the number of amino acids observed to support growth within 90% or greater of WT ubiquitin. Many positions in ubiquitin are either highly sensitive to mutation (4 or less amino acids support robust growth) or highly tolerant (17 or more amino acids support robust growth). (b) Space-filling representations of ubiquitin structure (based on PDB ID: 1UBQ⁵⁰) with sensitive positions colored blue, tolerant positions colored yellow and intermediate positions colored green. (c) Heat map representations of sensitive, intermediate and tolerant positions on the ubiquitin surface. One-dimensional maps on the bottom compare our analyses with a previous alanine scan.³⁷ At positions where the WT amino acid is alanine, glycine substitutions are shown for both the EMPIRIC and previous alanine scan.

Fig. 2b). Consistent with the strong functional conservation in ubiquitin, all of these substitutions support robust yeast growth. To further probe the accuracy of our bulk competition measurements, we constructed 17 individual point mutations spread across the ubiquitin coding sequence and analyzed their growth rate in monoculture (Supplementary Fig. S1). We find that monoculture growth rate was strongly correlated with EMPIRIC fitness measurements (Fig. 2e).

Sensitivity to mutation on the surface of ubiquitin

We examined the tolerance of each position by quantifying the number of amino acids compatible with robust growth, defined here as having less than a 10% growth defect. This cutoff definition was chosen both because it is larger than our measurement precision and because it encompasses the main peak of near-WT fitness mutants (Fig. 1c). This definition should include essentially all experimentally fit mutants in the robust class and minimize the number of mutants whose classification would switch with a small change in the cutoff. For amino acids located at or near the solvent-accessible surface, the observed tolerance was bimodal (Fig. 3a). The majority of positions permitted either greater than 16 amino acids (and were classified as tolerant) or less than 5 amino acids (and were classified as sensitive). Of note, position 12 was classified as intermediate despite having only four fit amino acids because unusually poor mutant representation in the plasmid library hindered fitness quantification of 10 amino acids. Mapping to the mono-ubiguitin structure revealed that sensitive and tolerant positions clustered completely on opposite faces (Fig. 3b). The fitness-sensitive positions are located on the β -sheet face, which contains the hydrophobic patch³⁶ including L8, I44 and V70. This region is known to bind to many ubiquitin receptors.

We compared the EMPIRIC fitness map for sensitive, intermediate and tolerant positions on the surface of ubiquitin to a previous alanine scan of the ubiquitin surface $^{\rm 37}$ (Fig. 3c). Overall, the EMPIRIC results correspond very well to the previous alanine scan. Of the 16 alanine scan mutants identified as growth defective, all have at least a 20% growth deficiency in our analyses. Of the 41 alanine scan mutants identified as supporting growth, 36 exhibited robust growth in our analyses while 5 exhibited growth defects ranging from 11% to 47%. The small number of discrepancies could be due to differences in experimental detail including the strains analyzed and the growth conditions. The alanine scan effectively identified sensitive sites but did not distinguish tolerant sites from sites with intermediate sensitivity. By quantifying the effects of mutants and examining all possible substitutions, the EMPIRIC analyses presented here define a continuous spectrum of mutational sensitivity from ultrasensitive to ultra-tolerant.

The most tolerant positions in ubiquitin can accommodate almost any amino acid substitution without disturbing observed function (Fig. 2c). These ultra-tolerant positions cluster on one side of ubiquitin. The only mutations other than stop codons observed to reduce function were proline mutations within the α -helix that are known to disrupt helical structure, as well as minor defects for a small fraction of the possible hydrophobic substitutions (S28F, S28I, Q31I, P37W, P37F, P37Y, D39W, T55W) and a small number of charge reversals (D21K, D21R, D39R). Charge reversals are generally well tolerated on this face of ubiquitin (e.g., E16, E18, D39, D52, R54 and K63). The tolerance to charge reversal mutants, which dramatically change the interaction potential of a protein surface, suggests that this face of ubiquitin is not involved in critical binding interactions. The entire amino acid sequence of ubiquitin is highly conserved (there are only three substitutions between yeast and human ubiquitins), indicating that mutations throughout the protein impact fitness on a magnitude that is selectable in natural populations. Of note, the WT residues at the ultra-tolerant positions are never aliphatic or aromatic. Based on these properties (polar and tolerant of individual mutations), we speculate that this region of ubiguitin may perform a solubility promoting role as similar properties have been observed in solubility promoting regions of the Hsp90 chaperone.^{51,4} These studies in Hsp90 support the idea that stringent natural selection for stability can result in highly optimized sequences that are so soluble that they are robust to individual mutations when measured with experimentally limited sensitivity. Ubiquitin is highly soluble (>100 mg/ml) and has been shown to impart solubility on genetically fused partner proteins.⁵³ In principle, ubiquitin solubility could provide a functional benefit by influencing the solubility of covalent complexes with substrates. The ultra-tolerance of this surface of ubiquitin to our experimental analyses is notable and motivated our discussion of solubility, which we acknowledge is highly speculative.

At the most sensitive positions to mutation (G10, R42, G47, K48, H68, R72, L73, R74, G75, G76), only the WT amino acid was observed to support robust growth. All 10 ultra-sensitive positions are located on the surface of the ubiquitin structure. Seven of these positions are either at or adjacent in primary sequence to the sites of critical covalent attachment (G47, K48, R72, L73, R74, G75, G76) consistent with the known role of these positions in conjugation to substrates and promoting recognition by the proteasome.¹³ Of the other three ultrasensitive positions, G10 is located in a β -turn in structural proximity to the C-terminus and has a positive main-chain φ angle that is incompatible with other amino acids. The other two (R42 and H68) are both structurally adjacent to the hydrophobic patch formed by L8, I44 and V70 that is at the interface with many ubiquitin receptors. L8, I44 and V70 can all tolerate conservative substitutions (e.g., Leu to IIe) in our assay. These results are consistent with both the binding of partner proteins to this patch and the relatively low specificity of hydrophobic interactions relative to polar interactions such as hydrogen bonds that are highly directional. Of note, the contributions of H68 to function were unclear from alanine scanning as it has only a partial (20%) growth defect in our analyses (Fig. 3c) and was positive for growth in plasmid swap experiments.³⁷

Mapping fitness sensitivity to interfaces

Analyzing the effects of all ubiquitin mutants combined with the available structural information on interfaces with many different binding partners provides a unique opportunity to investigate how binding interactions influence sensitivity to mutation. Based on chemical intuition, it has long been posited that interaction surfaces will impose evolutionary constraints based on the prediction that many mutations located at interfaces will disrupt binding. Consistent with previous observations, the most sensitive sites to mutation map to known binding interfaces including those with UBA and UIM domains (Fig. 4a and b). To further examine the relationship between binding interfaces and sensitivity to mutation, we analyzed the surface area buried by each position in ubiquitin across 44 highresolution crystal structures (Supplementary Table 4). The average fraction of surface area buried was greater for positions that were sensitive to mutation in our screen compared to positions that were tolerant to mutation (Fig. 4c). Across all surface positions, the fraction of surface area buried at structurally characterized interfaces predicted about 60% of the variance in observed tolerance to mutation (Fig. 4d). Of note, the relationship between surface burial and tolerance to substation appears to have multiple phases. At low fraction surface burial, many mutations are tolerated, and at a threshold around 0.3, amino acid tolerance tends to decrease. Positions with a fraction of surface area buried greater than 0.4 are universally sensitive in our analyses. Consistent with these observations, the data fit well to a transition switching model related to those used in chemical denaturation of proteins (Fig. 4d). These observations demonstrate that binding is a dominant determinant of sensitivity to mutation for ubiguitin and provide guantitative support for a long-standing intuition in molecular evolution.

We further analyzed how sensitive and tolerant positions mapped to the structure of K48-linked tetra-ubiquitin because this is the minimal signal for



Fig. 4. Relating fitness sensitivity on the surface of ubiquitin to binding interfaces. Structural representations of ubiquitin bound to common binding domains: (a) UBA (PDB ID: $200B^{55}$) and (b) UIM domain (PDB ID: $1QOW^{56}$). Top images show binding domains in magenta and ubiquitin as space-filled spheres with fitness-sensitive positions in blue, fitness-tolerant positions in yellow and intermediate positions in green. Bottom images illustrate the underlying ubiquitin secondary structure. (c) Fraction of surface area buried per sensitive or tolerant residue on the surface of ubiquitin in 44 high-resolution co-crystal structures. Error bars represent standard deviations (*N*= 18 and 19, respectively). (d) Correlation between fitness tolerance to amino acid substitution and burial at structurally characterized interfaces.

proteasome targeting.¹⁷ The structure of tetraubiquitin in the proteasome-bound state is currently unavailable, as is the structure of poly-ubiquitin attached to a substrate. However, the structure of unanchored K48-linked tetra-ubiquitin representing the predominant conformation at physiological conditions is available.⁵⁷ Structural analyses of this "closed" conformation of tetra-ubiquitin showed that fitness-sensitive positions all cluster on the interior of tetra-ubiquitin while fitness-tolerant positions all cluster on the exterior (Fig. 5). This result suggests that the structural arrangement mediated by interubiquitin contacts in closed tetra-ubiquitin may be biologically important. For example, by presenting a



Fig. 5. Relating fitness sensitivity to structure of tetra-ubiquitin. (a) Structural image of K48-linked tetraubiquitin (PDB ID: 206V⁵⁷). Top images show space-filling representation with fitness-sensitive positions colored blue, tolerant positions colored yellow and intermediate positions colored gray. Different color shades were used to distinguish subunits. Bottom image illustrates the underlying secondary structure. (b) Fraction surface area buried per sensitive or tolerant residue in tetraubiguitin. Error bars represent standard deviations (N=18 and 19, respectively).

molecular surface that is almost entirely polar, the closed conformation of tetra-ubiquitin may enhance solubility. In addition, the sequestration of binding sites for UBDs in the closed conformation may be important for modulating access to these interfaces by different effectors.⁵⁸ This potential mechanism is consistent both with an observed "open" conformation of K48-linked tetra-ubiquitin⁵⁹ and from NMR studies demonstrating that binding by UBDs can require access to interfaces unavailable in the "closed" conformation.⁶⁰

Effects of mutations in the solvent-inaccessible core

The majority of positions (10 of 16) located in the core of ubiquitin tolerated 3-6 mutations (Fig. 6a and b) with substitutions between hydrophobic amino acids of different geometry frequently resulting in robust growth at all positions. This distribution of tolerated mutations is consistent with solvent-inaccessible residues contributing to a well-packed, largely hydrophobic core to energetically distinguish the native state from unfolded conformations. Aliphatic (Val, Leu, Ile) to aromatic (Phe, Tyr, Trp) substitutions that increase core bulk were generally poorly tolerated. For example, we did not observe any aliphatic to Trp substitutions that were compatible with robust growth. In the core, mutations to Trp were only tolerated at positions where the WT amino acid was aromatic (F45, Y59). These observations indicate that large increases in core over packing, which are likely to alter the native conformation and dynamics, commonly result in functional defects to ubiquitin.

Polar amino acids in the core are generally incompatible with efficient ubiquitin function. Q41 is

the only polar WT amino acid in the core. The side chain of Q41 hydrogen bonds to a solvent-inaccessible and otherwise unsatisfied main-chain carbonyl oxygen. While most polar substitutions at position 41 exhibited a strong growth defect, multiple aliphatic substitutions (Leu, Ile, Met) were tolerated. These findings are consistent with the energetic penalty for burying unsatisfied hydrogen bonding atoms in the core of proteins and the energetic benefit from burial of hydrophobic atoms.^{61,62} The energetic penalty for burial of charged amino acids with unsatisfied hydrogen bonds in the core of proteins is especially severe and we observe growth defects at all core positions for substitutions to R, K, D or E. The general trends that we observe in the core are consistent with a large body of work demonstrating that the interior of proteins is important for governing protein folding and dynamics.63

Within the core of ubiquitin, we observe that the sensitivity to mutation is unevenly distributed with the most tolerant positions (L15, V17, I23, F45) all clustered in one structural region (Fig. 6c). We considered two potential explanations for this observation. First, protein folding may be a dominant determinant of the function of ubiquitin core mutants and the tolerant regions are less important for folding. Second, core mutations may impact ubiquitin function by subtle changes to the folded conformation and/or dynamics that affect critical binding interactions. The contribution of core positions to ubiquitin folding has been determined by φ analysis, which uses alanine mutations to identify residues that contribute to the folding transition state.³⁹ Of note, a broader folding transition state of ubiquitin was reported in studies using engineered double histidine substitutions and divalent metals as probes.⁶⁴ By φ analysis, the transition state for



Fig. 6. Mutant effects in the solvent-inaccessible core of ubiquitin. (a) Heat map indicating the fitness of mutations at core positions indicates that substitutions among aliphatic amino acids are generally well tolerated. (b) Positions in the core exhibit an intermediate tolerance to mutation with most positions having 3–6 different amino acids that support growth rates similar to the WT sequence ($s \ge -0.1$). (c) Structural representation of ubiquitin showing the WT side chains of core positions. Positions that tolerate more than 8 amino acids ($s \ge -0.1$) are colored yellow. (d) Relationship between core mutant impacts on folding stability^{39,42} and yeast growth. Previously measured effects on $\Delta\Delta G$ of folding are plotted such that negative numbers represent destabilization. The amount of destabilization estimated to abolish folding is indicated as a gray broken line on the left. Mutations to Q41, the only WT core polar amino acid, are shown in gray. All mutations in this panel are estimated to populate the unfolded state less than 1% in the absence of elevated temperature or denaturant based on the stability of WT ubiquitin.³⁹ Mutants that are destabilized by more than 2 kcal/mol are shown in orange if they are highly fit ($s \ge -0.12$), blue if they are strongly deleterious ($s \le -0.49$) or purple for those with intermediate fitness. (e) Structure of ubiquitin indicating the location of destabilized and highly fit (yellow) as well as destabilized and deleterious (cyan) mutations.

folding was found to occur in the same core region where we observe relatively high functional tolerance to mutation. In particular, L15, V17 and I23 all have ϕ -values \geq 0.5 (F45 was not analyzed), indicating that they have a large energetic contribution to the folding transition state. In contrast, positions located near the C-terminal tail in the structure of ubiguitin all were observed to have φ-values close to 0, indicating that they provide minimal energetic contributions to the folding transition state. These observations indicate that critical positions for ubiquitin folding can functionally tolerate more mutations than other core positions. Because protein folding to native conformations of ubiquitin should be required for function, this finding indicates either that the transition state for folding does not correlate with

mutant effects on folding efficiency or that ubiquitin core mutants may also have important impacts on protein behavior other than folding (e.g., binding).

We compared growth effects of core mutations to previously reported^{39,42} observations of individual ubiquitin mutant impacts on the thermodynamic stability difference between folded and unfolded conformations (Fig. 6d and Supplementary Table 5). All of the mutants in this panel were able to fold efficiently and populated native conformations under physiologically relevant conditions but exhibited growth rates from WT to null in our assays. This observation indicates that some core ubiquitin mutants able to fold efficiently are functionally defective. Our previous work demonstrated that the L67S and L69S core mutations are capable of folding under physiological conditions but are defective for binding to proteasome receptors and do not support yeast growth.⁴¹ Thus, core ubiquitin mutations capable of folding can be defective for binding to important receptors. In future studies, it will be interesting to determine the specific biochemical defects in core ubiquitin mutants that impair function. Of note, mutations that resulted in less than a 2 kcal/mol destabilization of the folded relative to unfolded states exhibited small to no observable growth defects. Within this stability region, mutations at Q41 had more pronounced functional defects than mutations at positions where the WT amino acid was hydrophobic. This observation is consistent with the observation that polar interactions in protein interiors can have a large influence on protein dynamics.⁶²

Among the analyzed panel of core mutations that destabilized the folded state by 2 kcal/mol or greater, there was a large variance in functional effects spanning from null to WT growth rates (Fig. 6d). Within this stability regime, we mapped mutants with either minimal or severe growth effects onto structure (Fig. 6e). The destabilized mutants that supported the most efficient growth included substitutions at three of the most functionally tolerant positions (L15A, V17N, I23A; mutants at F45 were not in the stability data set). Mutations with a similar range of destabilization that exhibited severe growth defects were clustered near the β-sheet surface and C-terminus in the structure of ubiquitin (Fig. 6e). Thus, core positions that are relatively sensitive to mutation were located adjacent to surface positions that make critical binding interfaces while tolerant core positions were located distant to known binding interfaces. Based on these observations, we speculate that many core ubiquitin mutations may impact function by affecting binding affinities. Interestingly, NMR studies have demonstrated that the C-terminal region of ubiquitin exhibits uncorrelated conformational dynamics^{65,66} suggesting that it is capable of sampling many different conformations that could be important for binding to diverse UBDs. Indeed, recent reports indicate that distinct ubiquitin conformations mediated by core amino acids are important for affinity with different binding partners.⁶⁷ NMR studies on linked ubiquitin also observed conformational dynamics in the C-terminal region, 66 consistent with a potential role in mediating receptor binding to poly-ubiquitin.

Conclusions

The comprehensive analysis of ubiquitin mutants presented here provides a rigorous examination of the physical constraints at each position in the protein. Consistent with previous observations, our results strongly support binding as a dominant functional constraint for ubiquitin both for surface positions and for many core positions. Indeed, location at structurally characterized interfaces alone is a good predictor of the tolerance of surface positions to mutation. One face of ubiquitin that is not commonly at structurally characterized interfaces tolerates almost all substitutions without causing detectable growth defects. The experimental tolerance that we observed for this surface indicates that its contribution to function is relatively insensitive to mutation. This mutational profile together with the polar composition of this region is consistent with a role in promoting solubility. We also find that the functional sensitivity to core mutations is asymmetrically distributed. Sites involved in the folding transition state are the most tolerant to mutation while sites in structural proximity to critical binding sites are the most sensitive. These results are consistent with an important role for ubiquitin conformational dynamics in mediating binding to critical partner molecules. The structural separation of folding centers and regions important for dynamics that influence binding events may be a design principle utilized by other proteins to balance the requirement for both folding and dynamics required for function.

Materials and Methods

High-throughput EMPIRIC fitness measurements

We constructed ubiquitin mutant libraries in a KanMX4marked yeast high-copy shuttle vector (p427) with ubiquitin expression driven by the GPD1 promoter. To aid in cloning, we removed the ampicillin resistance gene from the vector and performed selection during bacterial cloning with kanamycin (resistance provided by the KanMX4 marker). Of note, this high-copy plasmid system is important for expressing ubiquitin at near-physiological levels because the ubiquitin gene is present at multiple chromosomal locations in WT yeast. Libraries of saturated single codon substitutions in yeast ubiquitin were generated using a cassette ligation strategy in p427GPD as previously described.⁶ To facilitate cloning and subsequent sequencing analyses, we subdivided the ubiquitin gene into eight regions of 9-10 amino acids. Pools of saturated point mutants within each region that could be accurately and efficiently interrogated with short-read Illumina sequencing were generated. To distinguish the growth properties of ubiquitin mutants, we utilized the Sub328 yeast strain.¹⁵ The sole ubiquitin in these cells is supplied from a galactose-regulated promoter.

Pooled plasmid libraries of mutants for each region of ubiquitin were transformed separately into the same batch of Sub328 cells as previously described.⁶⁸ In order to minimize doubly transformed cells,⁶⁹ we transformed a total of 1 µg of plasmid DNA per 100 µl of competent yeast cells, and after recovery, the cells were grown for 48 h at 30 °C in 30 ml of liquid SRGal (synthetic 1% raffinose and 1% galactose) media supplemented with G418 (200 µg/ml). Cultures were diluted to maintain log-phase growth. After 48 h of selection for G418 resistance in SRGal media, late-log cells were

collected by centrifugation, then washed and resuspended in SD (synthetic 2% dextrose) media with G418 and ampicillin (50 µg/ml) to hinder bacterial contamination. Cultures of 100 ml were grown in a shaking incubator at 30 °C for 3 days with dilution to maintain the culture in logarithmic growth. Time point samples were collected throughout this period by centrifuging 2×10⁸ cells, resuspending with 1 ml of water, transferring to a microfuge, pelleting, removing the supernatant and freezing the pellet at -80 °C. Fitness analyses were performed on samples isolated after 12, 15, 19, 23, 29 and 45 h in dextrose. These fitness analyses were performed over 11 generations as the doubling time of the yeast harboring the WT rescue plasmid was 3 h under these conditions. A full experimental repeat was performed for the ubiquitin region encompassing amino acids 40-48, including preparation of competent yeast from a separate colony, transformation, growth and time point sampling.

Plasmid DNA was isolated from yeast pellets and processed for deep sequencing. Frozen pellets were thawed, lysed using zymolyase and plasmid DNA isolated and prepared for sequencing as previously described.⁶ An initial PCR reaction was performed to amplify the library version of ubiquitin utilizing primers specific to the p427GPD vector. This PCR product was separated on an agarose gel, excised and purified using a silica column (Zymo Research). A second round of PCR was performed with region-specific primers that added a Mmel cut site immediately upstream of the randomized region and a universal primer binding site 250 bases downstream. The resulting PCR product was purified on a silica column, digested with Mmel and purified again on a silica column. Barcoded adapters were then ligated, and samples were processed and analyzed as previously described.⁶

Adapter cassettes with a sticky-end complementary to the resulting Mmel overhang were attached using T4 DNA ligase (New England Biolabs). Ligation adapters included a binding site for universal Illumina primers and a barcode to distinguish the time point and sample. Ligation reactions were separated on an agarose gel, and the ligation product was excised and purified. A final round of PCR was performed with Illumina primers. Cycles were limited and the high-fidelity Pfusion enzyme (New England Biolabs) was utilized throughout to minimize PCR errors. For each region, a processing control that started with a purified plasmid with WT ubiquitin was included and was processed identical with time point samples (same number of PCR cycles, etc.). To distinguish misreads during sequencing, we also included a sequencing control in all deep sequencing samples. The sequencing control consisted of a region of the Sec61 gene cloned into a plasmid with flanking Illumina primer binding sites. Eight cycles of PCR from this plasmid generated a deep sequencing sample with minimal sequence heterogeneity.

Short-read (36 base) Illumina sequencing was utilized to analyze all time point samples. The resulting FastQ files were analyzed as previously described⁶ in order to determine the fitness effects of mutants. Processing and sequencing errors were directly estimated in each sequencing reaction from the number of apparent mutants observed in the internal processing and sequencing controls. The average per base error rate including PCR processing and sequencing was 0.005 per base. The majority of these errors (~90%) result in apparent double

mutants that were filtered out of the data resulting in accurate abundance measurements of the underlying distribution of mutants in the library. Mutants with a relative abundance (mutant/WT) below 2⁻¹⁰ at the initial time point (12 h in dextrose) under selection were deemed too noisy for accurate fitness measurements based on visual inspection of mutant versus time trajectories and were omitted from further analysis. In addition, mutants that created an internal Mmel site that would complicate processing were omitted from fitness analyses. The first three time points during library selection (corresponding to 12, 15 and 19 h in dextrose) were utilized to analyze the rapid drop-off in stop codons and other null-like mutants. For mutations that persisted in the population, fitness effects were determined using all time points in selection. The residuals for each fit were determined to identify problematic mutants that were subsequently plotted and eliminated from consideration if they showed multi-phasic behavior (<1% of the data were eliminated using this approach). Depletion ratios were calculated from the relative abundance of mutants observed in sequencing of the plasmid library compared to cells collected at the end of growth in galactose. Mutants with a relative abundance (mutant/WT) below 2⁻⁸ in the plasmid library were deemed too noisy for depletion analyses based on visual inspection of trajectories.

Growth rate of individual mutants in monoculture

We measured the growth rate of 19 different point mutants in monoculture after shutoff in dextrose (Supplementary Fig. S1). These mutants were chosen to span a range of EMPIRIC fitness values and were from regions spanning the entire ubiquitin gene. Single mutants were generated in p427GPD and transformed into Sub328 yeast. Growth rates were determined by monitoring OD_{600} during growth under identical conditions to the EMPIRIC bulk competitions (SD media with G418 and ampicillin at 30 °C).

Structural analyses of ubiquitin

Ubiquitin positions were characterized as at or near the solvent-accessible surface or in the solvent-inaccessible core based on the crystal structure of mono-ubiquitin [Protein Data Bank (PDB) ID: 1UBQ⁵⁰] and the classification algorithm of Dahiyat and Mayo.70 We analyzed ubiquitin interfaces by quantifying the surface area buried⁷ by each amino acid in atomic resolution structures in the PDB. We searched the PDB for entries with high sequence identity to ubiquitin, at least two chains in the biological assembly and X-ray resolution of 2.5 Å or better. We identified PDB files that had a BLAST E-value of less than 10⁻²⁰ and length of alignment between 50 and 100 residues as the selection criteria. We manually curated this list to exclude structures of identical protein complexes (keeping only the highest-resolution structure in each case) and structures of mono-ubiguitin, resulting in 117 ubiguitin chains in 44 PDB structures. We used areaimol from the CCP4 v.6.2 package⁷² to calculate ASA (solvent-accessible surface areas) for each residue of every ubiquitin molecule both in the complex and in isolation. The resulting changes in ASA upon complex formation are provided in Supplementary Information (Table S4).

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2013.01.032

Acknowledgements

We are appreciative to C. R. Matthews, F. Massi, O. Rando, R. Gilmore, E. Baehrecke, A. Ballesteros and M. Moore for useful comments. The Sub328 yeast strain was kindly provided by D. Finley. This work was supported in part by grants from the National Institutes of Health (R01-GM083038) and the American Cancer Society (RSG-08-17301-GMC) to D.N.A.B. and by a National Institutes of Health grant (R01-GM065334) to D.F.

> Received 20 December 2012; Received in revised form 22 January 2013; Accepted 27 January 2013 Available online 30 January 2013

Keywords:

fitness; structure; function; thermodynamic stability; sequence

Abbreviations used:

PDB, Protein Data Bank; UBD, ubiquitin-binding domain; UIM, ubiquitin-interacting motif; UBA, ubiquitin-associated domain; WT, wild type.

References

- Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E. & Hutchison, C. A., 3rd. (1989). Complete mutagenesis of the HIV-1 protease. *Nature*, 340, 397–400.
- Rennell, D., Bouvier, S. E., Hardy, L. W. & Poteete, A. R. (1991). Systematic mutation of bacteriophage T4 lysozyme. *J. Mol. Biol.* 222, 67–88.
- Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S. *et al.* (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.
- Fowler, D. M., Araya, C. L., Fleishman, S. J., Kellogg, E. H., Stephany, J. J., Baker, D. & Fields, S. (2010). High-resolution mapping of protein sequence–function relationships. *Nat. Methods*, 7, 741–746.
- Hietpas, R., Roscoe, B., Jiang, L. & Bolon, D. N. (2012). Fitness analyses of all possible point mutations for regions of genes in yeast. *Nat. Protoc.* 7, 1382–1396.
- Hietpas, R. T., Jensen, J. D. & Bolon, D. N. (2011). Experimental illumination of a fitness landscape. *Proc. Natl Acad. Sci. USA*, **108**, 7896–7901.

- DeBartolo, J., Dutta, S., Reich, L. & Keating, A. E. (2012). Predictive Bcl-2 family binding models rooted in experiment or structure. *J. Mol. Biol.* 422, 124–144.
- 8. Pitt, J. N. & Ferre-D'Amare, A. R. (2010). Rapid construction of empirical RNA fitness landscapes. *Science*, **330**, 376–379.
- McLaughlin, R. N., Jr., Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. *Nature*, **491**, 138–142.
- Araya, C. L., Fowler, D. M., Chen, W., Muniez, I., Kelly, J. W. & 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.
- Whitehead, T. A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S. J., De Mattos, C. *et al.* (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **30**, 543–548.
- Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E. M. *et al.* (2012). Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*, 332, 816–821.
- Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* 78, 477–513.
- 14. Hershko, A. & Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479.
- Spence, J., Sadis, S., Haas, A. L. & Finley, D. (1995). A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol. Cell. Biol.* 15, 1265–1273.
- Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K. & Varshavsky, A. (1989).
 A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, 243, 1576–1583.
- Thrower, J. S., Hoffman, L., Rechsteiner, M. & Pickart, C. M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19**, 94–102.
- Rock, K. L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L. *et al.* (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell*, **78**, 761–771.
- 19. Goldberg, A. L. (2007). Functions of the proteasome: from protein degradation and immune surveillance to cancer therapy. *Biochem. Soc. Trans.* **35**, 12–17.
- Goebl, M. G., Yochem, J., Jentsch, S., McGrath, J. P., Varshavsky, A. & Byers, B. (1988). The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme. *Science*, **241**, 1331–1335.
- Alkalay, I., Yaron, A., Hatzubai, A., Orian, A., Ciechanover, A. & Ben-Neriah, Y. (1995). Stimulation-dependent IκBα phosphorylation marks the NF-κB inhibitor for degradation via the ubiquitin– proteasome pathway. *Proc. Natl Acad. Sci. USA*, 92, 10599–10603.
- Dennissen, F. J., Kholod, N. & van Leeuwen, F. W. (2012). The ubiquitin proteasome system in neurodegenerative diseases: culprit, accomplice or victim? *Prog. Neurobiol.* 96, 190–207.

- Orlowski, R. Z. & Kuhn, D. J. (2008). Proteasome inhibitors in cancer therapy: lessons from the first decade. *Clin. Cancer Res.* 14, 1649–1657.
- Whitby, F. G., Xia, G., Pickart, C. M. & Hill, C. P. (1998). Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes. *J. Biol. Chem.* **273**, 34983–34991.
- Lee, I. & Schindelin, H. (2008). Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell*, **134**, 268–278.
- Burch, T. J. & Haas, A. L. (1994). Site-directed mutagenesis of ubiquitin. Differential roles for arginine in the interaction with ubiquitin-activating enzyme. *Biochemistry*, **33**, 7300–7308.
- Miura, T., Klaus, W., Gsell, B., Miyamoto, C. & Senn, H. (1999). Characterization of the binding interface between ubiquitin and class I human ubiquitinconjugating enzyme 2b by multidimensional heteronuclear NMR spectroscopy in solution. *J. Mol. Biol.* 290, 213–228.
- Hurley, J. H., Lee, S. & Prag, G. (2006). Ubiquitinbinding domains. *Biochem. J.* **399**, 361–372.
- Hicke, L., Schubert, H. L. & Hill, C. P. (2005). Ubiquitin-binding domains. *Nat. Rev., Mol. Cell Biol.* 6, 610–621.
- Fushman, D. & Wilkinson, K. D. (2011). Structure and recognition of polyubiquitin chains of different lengths and linkage. *F1000 Biol. Rep.* 3, 26.
- Hofmann, K. & Falquet, L. (2001). A ubiquitininteracting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem. Sci.* 26, 347–350.
- Fisher, R. D., Wang, B., Alam, S. L., Higginson, D. S., Robinson, H., Sundquist, W. I. & Hill, C. P. (2003). Structure and ubiquitin binding of the ubiquitin-interacting motif. *J. Biol. Chem.* 278, 28976–28984.
- Hofmann, K. & Bucher, P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem. Sci.* 21, 172–173.
- Ohno, A., Jee, J., Fujiwara, K., Tenno, T., Goda, N., Tochio, H. *et al.* (2005). Structure of the UBA domain of Dsk2p in complex with ubiquitin molecular determinants for ubiquitin recognition. *Structure*, **13**, 521–532.
- Dieckmann, T., Withers-Ward, E. S., Jarosinski, M. A., Liu, C. F., Chen, I. S. & Feigon, J. (1998). Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr. *Nat. Struct. Biol.* 5, 1042–1047.
- Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M. & Pickart, C. (1996). Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc. Natl Acad. Sci. USA*, 93, 861–866.
- Sloper-Mould, K. E., Jemc, J. C., Pickart, C. M. & Hicke, L. (2001). Distinct functional surface regions on ubiquitin. *J. Biol. Chem.* 276, 30483–30489.
- Wintrode, P. L., Makhatadze, G. I. & Privalov, P. L. (1994). Thermodynamics of ubiquitin unfolding. *Proteins*, **18**, 246–253.
- Went, H. M. & Jackson, S. E. (2005). Ubiquitin folds through a highly polarized transition state. *Protein Eng., Des. Sel.* 18, 229–237.

- Benitez-Cardoza, C. G., Stott, K., Hirshberg, M., Went, H. M., Woolfson, D. N. & Jackson, S. E. (2004). Exploring sequence/folding space: folding studies on multiple hydrophobic core mutants of ubiquitin. *Biochemistry*, 43, 5195–5203.
- Haririnia, A., Verma, R., Purohit, N., Twarog, M. Z., Deshaies, R. J., Bolon, D. & Fushman, D. (2008). Mutations in the hydrophobic core of ubiquitin differentially affect its recognition by receptor proteins. *J. Mol. Biol.* 375, 979–996.
- Loladze, V. V., Ermolenko, D. N. & Makhatadze, G. I. (2002). Thermodynamic consequences of burial of polar and non-polar amino acid residues in the protein interior. *J. Mol. Biol.* **320**, 343–357.
- Pickart, C. M., Kasperek, E. M., Beal, R. & Kim, A. (1994). Substrate properties of site-specific mutant ubiquitin protein (G76A) reveal unexpected mechanistic features of ubiquitin-activating enzyme (E1). *J. Biol. Chem.* 269, 7115–7123.
- Lynch, M. & Conery, J. S. (2003). The origins of genome complexity. *Science*, **302**, 1401–1404.
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A. *et al.* (2009). Population genomics of domestic and wild yeasts. *Nature*, **458**, 337–341.
- Sanjuan, R., Moya, A. & Elena, S. F. (2004). The distribution of fitness effects caused by singlenucleotide substitutions in an RNA virus. *Proc. Natl Acad. Sci. USA*, **101**, 8396–8401.
- Carrasco, P., de la Iglesia, F. & Elena, S. F. (2007). Distribution of fitness and virulence effects caused by single-nucleotide substitutions in tobacco etch virus. *J. Virol.* 81, 12979–12984.
- Domingo-Calap, P., Cuevas, J. M. & Sanjuan, R. (2009). The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet.* 5, e1000742.
- Perisic, O., Xiao, H. & Lis, J. T. (1989). Stable binding of *Drosophila* heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. *Cell*, **59**, 797–806.
- Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531–544.
- Pursell, N. W., Mishra, P. & Bolon, D. N. (2012). Solubility-promoting function of Hsp90 contributes to client maturation and robust cell growth. *Eukaryotic Cell*, **11**, 1033–1041.
- Wayne, N. & Bolon, D. N. (2011). Charge-rich regions modulate the anti-aggregation activity of Hsp90. *J. Mol. Biol.* 401, 931–939.
- Butt, T. R., Jonnalagadda, S., Monia, B. P., Sternberg, E. J., Marsh, J. A., Stadel, J. M. *et al.* (1989). Ubiquitin fusion augments the yield of cloned gene products in *Escherichia coli. Proc. Natl Acad. Sci. USA*, **86**, 2540–2544.
- 54. King, J. L. & Jukes, T. H. (1969). Non-Darwinian evolution. *Science*, **164**, 788–798.
- Peschard, P., Kozlov, G., Lin, T., Mirza, I. A., Berghuis, A. M., Lipkowitz, S. *et al.* (2007). Structural basis for ubiquitin-mediated dimerization and activation of the ubiquitin protein ligase Cbl-b. *Mol. Cell*, **27**, 474–485.
- Swanson, K. A., Kang, R. S., Stamenova, S. D., Hicke, L. & Radhakrishnan, I. (2003). Solution structure of Vps27 UIM–ubiquitin complex important

for endosomal sorting and receptor downregulation. *EMBO J.* **22**, 4597–4606.

- Eddins, M. J., Varadan, R., Fushman, D., Pickart, C. M. & Wolberger, C. (2007). Crystal structure and solution NMR studies of Lys48-linked tetraubiquitin at neutral pH. *J. Mol. Biol.* **367**, 204–211.
- Varadan, R., Walker, O., Pickart, C. & Fushman, D. (2002). Structural properties of polyubiquitin chains in solution. *J. Mol. Biol.* **324**, 637–647.
- Cook, W. J., Jeffrey, L. C., Kasperek, E. & Pickart, C. M. (1994). Structure of tetraubiquitin shows how multiubiquitin chains can be formed. *J. Mol. Biol.* 236, 601–609.
- Varadan, R., Assfalg, M., Raasi, S., Pickart, C. & Fushman, D. (2005). Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Mol. Cell*, **18**, 687–698.
- Bolon, D. N., Marcus, J. S., Ross, S. A. & Mayo, S. L. (2003). Prudent modeling of core polar residues in computational protein design. *J. Mol. Biol.* 329, 611–622.
- Bolon, D. N. & Mayo, S. L. (2001). Polar residues in the protein core of *Escherichia coli* thioredoxin are important for fold specificity. *Biochemistry*, 40, 10047–10053.
- Cordes, M. H., Davidson, A. R. & Sauer, R. T. (1996). Sequence space, folding and protein design. *Curr. Opin. Struct. Biol.* 6, 3–10.
- Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J. Mol. Biol.* 337, 463–475.

- Massi, F., Grey, M. J. & Palmer, A. G., 3rd (2005). Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR R1rho relaxation experiments. *Protein Sci.* 14, 735–742.
- Fushman, D., Varadan, R., Assfalg, M. & Walker, O. (2004). Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements. *Prog. Nucl. Magn. Reson. Spectrosc.* 44, 189–214.
- Zhang, Y., Zhou, L., Rouge, L., Phillips, A. H., Lam, C., Liu, P. *et al.* (2013). Conformational stabilization of ubiquitin yields potent and selective inhibitors of USP7. *Nat. Chem. Biol.* 9, 51–58.
- Gietz, R. D. & Woods, R. A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* 350, 87–96.
- Scanlon, T. C., Gray, E. C. & Griswold, K. E. (2009). Quantifying and resolving multiple vector transformants in *S. cerevisiae* plasmid libraries. *BMC Biotechnol.* 9, 95.
- Dahiyat, B. I. & Mayo, S. L. (1997). De novo protein design: fully automated sequence selection. *Science*, 278, 82–87.
- Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709–713.
- Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr., Sect. D: Biol. Crystallogr. 50, 760–763.