2014-03-25

Analyses of All Possible Point Mutations within a Protein Reveals Relationships between Function and Experimental Fitness: A Dissertation

Benjamin P. Roscoe
University of Massachusetts Medical School

ANALYSES OF ALL POSSIBLE POINT MUTATIONS WITHIN A PROTEIN REVEALS RELATIONSHIPS BETWEEN FUNCTION AND EXPERIMENTAL FITNESS

A Dissertation Presented

By

BENJAMIN PETER ROSCOE

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

(March 25, 2014)

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

ANALYSES OF ALL POSSIBLE POINT MUTATIONS WITHIN A PROTEIN REVEALS RELATIONSHIPS BETWEEN FUNCTION AND EXPERIMENTAL FITNESS

A Dissertation Presented

By

BENJAMIN PETER ROSCOE

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

Daniel N. Bolon, Ph.D., Thesis Advisor

C. Robert Matthews, Ph.D., Member of Committee

William E. Royer, Ph.D., Member of Committee

Peter M. Pryciak, Ph.D, Member of Committee

Douglas M. Fowler, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

Reid Gilmore, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school

Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Dedication

This work is dedicated to my mother and father, Peter Jack Roscoe and Erika Maria Roscoe.

Acknowledgements

Before matriculating in the UMass Medical School Graduate School of Biological Sciences, I believed a doctorate program was another hoop to jump through for a distinguished career in the biomedical sciences. I have found this outlook to be extremely shortsighted and cannot begin to quantitate my personal growth as a scientist and human. Most importantly, I now realize this growth should always be encouraged and pursued by all people. I could not have started this lifelong journey without the individuals acknowledged below:

I would like to thank my advisor, Daniel N. Bolon, Ph.D., for continually pushing me to always make concrete advances in any tasks I pursue, and for being a constant source of projects and ideas, as well as accepting and encouraging my ideas. Next, I would like to recognize all of my current and former Bolon lab members for constant personal and scientific interactions, Ryan Hietpas, Ph.D., Natalie Wayne Pursell Ph.D., Parul Mishra, Ph.D., Li Jiang, Pamela Cote, Aneth Laban, Lester Pullen, and Dan Virgil.

I would also like to thank my GSBS classmates Paul Nobrega, Kenneth Lloyd, and Sarah Grace Swygert for scientific discussions, support, and editing duties.

Additionally, I must recognize the incredible assistance of the UMass Medical school flow cytometry core staff that helped integrate flow cytometry and cell sorting into the middle of my thesis research almost seamlessly. Most of the work in chapter III would not have been possible without their expertise.

My thesis advisory committee members have given indespensible advice throughout my research, thank you to Reid Gilmore, William Royer, C. Robert Matthews, Francesca Massi and Peter Pryciak. I also want to give special thanks to Ken Knight, GSBS assistant Dean for lots of personal support, and Dean Anthony Carruthers for personal communications regarding my projects.

Abstract

The primary amino acid sequence of a protein governs its specific cellular functions. Since the cracking of the genetic code in the late 1950's, it has been possible to predict the amino acid sequence of a given protein from the DNA sequence of a gene. Nevertheless, the ability to predict a protein's function from its primary sequence remains a great challenge in biology. In order to address this problem, we combined recent advances in next generation sequencing technologies with systematic mutagenesis strategies to assess the function of thousands of protein variants in a single experiment. Using this strategy, my dissertation describes the effects of most possible single point mutants in the multifunctional Ubiquitin protein in yeast. The effects of these mutants on the essential activation of ubiquitin by the ubiquitin activating protein (E1, Uba1p) as well as their effects on overall yeast growth were measured. Ubiquitin mutants defective for E1 activation were found to correlate with growth defects, although in a non-linear fashion. Further examination of select point mutants indicated that E1 activation deficiencies predict downstream defects in Ubiquitin function, resulting in the observed growth phenotypes. These results indicate that there may be selective pressure for the activity of the E1enzyme to selectively activate ubiquitin protein variants that do not result in functional downstream defects. Additionally, I will describe the use of similar techniques to discover drug resistant mutants of the oncogenic protein BRAFV600E in human melanoma cell lines as an example of the widespread applicability of our strategy for addressing the relationship between protein function and biological fitness.

Table of Contents

List of Figures

List of Electronic Suppplementary Table Files

Tables that are too large to be conveniently incorporated into the text of this dissertation are listed below and supplied as Microsoft excel format xlsx files. Tables are named by the chapter in which they are referred, preceding the table number within the chapter.

TableS2_1.xlsx – Number of illumina sequence reads for each mutant during bulk Fitness competition

TableS2_2.xlsx - Relative effects of amino acid substitutions on yeast growth rate

Table S2_3.xlsx – Sequencing counts of mutant reads in the plasmid library and after outgrowth in galactose media with WT ubiquitin co-expression

TableS2_4.xlsx – Change in solvent accessible surface area (d-ASA) of 123 ubiquitin chains in 44 heterocomplexes and the 2O6V.PDB homotetramer

TableS2_5.xlsx – Fitness measurements and changes in free energy of folded states Relative to unfolded states compiled in the literature

TableS3_1.xlsx – Sequencing counts

TableS3_2.xlsx – E1 reactivity estimates from bulk competitions

TableS3_3.xlsx – sequencing counts of E+H region mutants reacted with excess E1

TableS3_4.xlsx – E1 reactivity estimates from bulk competitions with excess E1

Table S5_1.xlsx – Number of illumina sequence reads for each ubiquitin point mutants during competitive growth at 36º C.

Table S5_2.xlsx – Relative growth rate of ubiquitin point mutants at 36º C.

Preface

Chapter II is published as:

Roscoe, B. P., Thayer, K. M., Zeldovich, K. B., Fushman, D. & Bolon, D. N. A. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J. Mol. Biol.* **425,** 1363–77 (2013).

Chapter III was accepted for publication by the Journal of Molecular biology on 05/17/2014, still pending online publication and DOI.

Benjamin P. Roscoe and Daniel N. Bolon. "Systematic exploration of ubiquitin sequence, E1 activation efficiency, and experimental fitness in yeast"

Chapter IV is published as:

Wagenaar, T. R. *et al.* Resistance to vemurafenib resulting from a novel mutation in the BRAFV600E kinase domain. *Pigment Cell Melanoma Res.* **27,** 124–33 (2014).

Chapter V material is unpublished and requires further experiments before consideration for submission

Benjamin P. Roscoe and Daniel N.A. Bolon

Chapter I - Introduction

Cellular function is mediated by biochemical interactions of all the molecules in a cell. The primary amino acid sequence of a given protein determines its cellular function via its structure[1] and the biochemical interactions mediated by its amino acid side chains.. Since the cracking of the genetic code in the late 1950's, it has been possible to predict the primary amino acid sequence of a given protein from the coding DNA sequence.[2,3,4] Although it remains difficult to predict the 3-dimensional structure of a folded protein from a DNA coding sequence, great advances in structural determination of proteins by x-ray crystallography, NMR, and electron microscopy have resulted in over 90,000 solved protein structures deposited in the RCSB Protein Data Bank.[5] Predicting protein function from primary amino acid sequence is an even more daunting task, with the primary methods being phylogenetic comparisons of DNA coding sequences and homology modeling of similar protein domains[6]. Therefore, experimental techniques to systematically determine the role of each amino acid position of a protein on biochemical and cellular function are proving to be very valuable to link protein function with sequence[7]. The following work describes insights gained from studying the consequences on biochemical function and cellular growth by high-throughput measurements of all possible point mutants of a protein.

Analysis of protein function using systematic mutagenesis

Systematic mutagenesis (introducing amino acid substitutions at all positions of an entire protein or region of interest) of proteins has been achieved for many years through error prone PCR or replacing portions protein-coding genes with cassette-based oligonucleotides containing synthetically introduced randomized codons[8,9]. High throughput methods to screen mutants generated by these methods were previously limited by the ability to analyze the function of many variants at the same time. Previous methods included Sanger (inhibition of polymerization reactions with randomly incorporated dideoxy nucleotides) sequencing, amber-codon suppression in *E. coli* to screen for functional variants,[10] and proteomic screens to detect mutant protein sequences after selection.[11] The fairly recent advent of next generation DNA sequencing techniques now allows for the analysis of millions of DNA sequence reads from a single pool of DNA. Within the last 5 years, multiple investigators have reported high throughput analysis of large pools of protein variants using various selection techniques (e.g. growth competition, phage display) analyzed by deep sequencing.[12–17]

Recently, multiple strategies combining high-throughput mutagenesis coupled with deep-sequencing readouts have been developed and utilized by various researchers. Each strategy has inherent advantages and disadvantages that should be carefully considered for each experimental application. Some investigators, including Stan Fields *et al*,[18] and Doug Fowler *et al*[12] utilize error prone PCR and/or error prone oligonucleotide synthesis to generate tens to hundreds of thousands of variants within a

single library. This is a very powerful approach because of the massive number of mutants that can be screened simultaneously. Another advantage of this strategy is that both single and multiple mutations are generated in the library, so the effects of multiple substitutions can be determined. The study of epistatic mutations within a single protein is a current area of immense interest in the field of protein sequence evolution.[19] The disadvantage to random mutagenesis techniques is that it is difficult to capture every single amino acid substitution within a library, as there is no way to prevent multiple substitutions when increasing the PCR or oligonucleotide synthesis error rate. Therefore, the library size would have to be untenably large to ensure all single mutants are covered. For applications such as screening for drug resistant mutants within a target protein, it is advantageous to screen for only single point mutations, as these are probabilistically the most likely mutations to occur in a clinical setting.[20]

Work in our lab and presented throughout this thesis systematically builds a mutant library containing every codon within a protein sequence with all 64 possible codons.[13,14] This is achieved by generating a self-removable fragment from the region of the gene to be mutagenized by the addition of inverted type-II endonuclease (specifically BsaI) sites within the gene (Figure 1.1a,b). Type-II endonucleases cut outside of their DNA recognition sequence and allow for excision of the region to be mutagenized leaving 4 base pair overlapping ends that can then have cassette libraries containing randomized codons ligated within the gene (Figure 1.1c). These mutant libraries are then able to be tested in multiple types of high-throughput experiments, including the ability to support growth as a single copy in yeast (figure 1.2)[13,14], or for specific biochemical

functions by harnessing display technologies such as yeast surface and phage display to display library variants on the surface of a yeast cell or phage particle, and screening for specific interactions. Deep sequencing can then be used to tabulate the overall abundance of point mutants in the initial library compared to specified experimental conditions, whether it is abundance in a cell population over time, or enrichment/depletion after a biochemical screen.

This methodology provides multiple advantages, although less total mutants (and no multiple substitutions within the same protein) are screened than in the approaches described above: One advantage is that by substituting every codon, all WT synonyms and stop codons (missense mutations) are sampled, resulting in a large number of internal controls for WT synonymous and truncated protein sequences that are useful for normalizing enrichment and depletion of point mutants. Another advantage of this strategy is that the effect of protein point mutations can be averaged across synonymous codons, giving a better handle on variation within a single experiment, and greater confidence in measurements. An additional benefit is that this strategy has the potential to differentiate between activities between synonymous codons, although we have not observed this phenomenon within our experiments to date. Perhaps the most important advantage to this method is that by systematically substituting with all possible mutations at each position in a protein, we can obtain detailed information regarding the chemical and physical requirements of each amino acid within a protein. Chapters 2-4 will report on positions of the proteins ubiquitin and BRAF V600E where very modest substitutions (e.g. substitutions with residues of similar physical and chemical properties) at specific

amino acid positions are not tolerated, indicating specific biochemical interactions are perturbed.

Figure 1.1. Steps to generate plasmid libraries of point mutants. (a) Whole-plasmid PCR to generate inverted BsaI vector that can be fully excised with the BsaI endonuclease. (b) Digestion of this vector to generate directional sticky-ends. (c) Cassette ligation to introduce point mutants in frame with the gene.

Figure 1.1. Steps to generate plasmid libraries of point mutants.

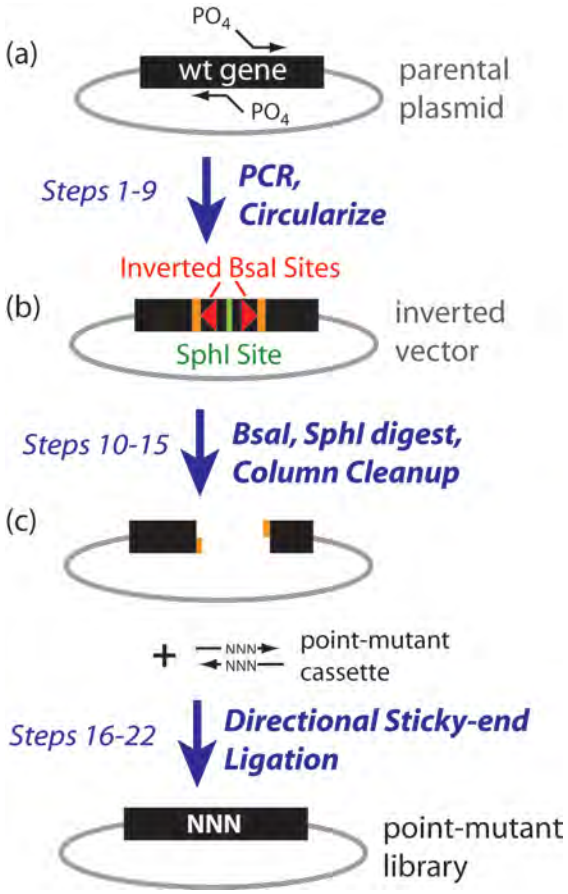
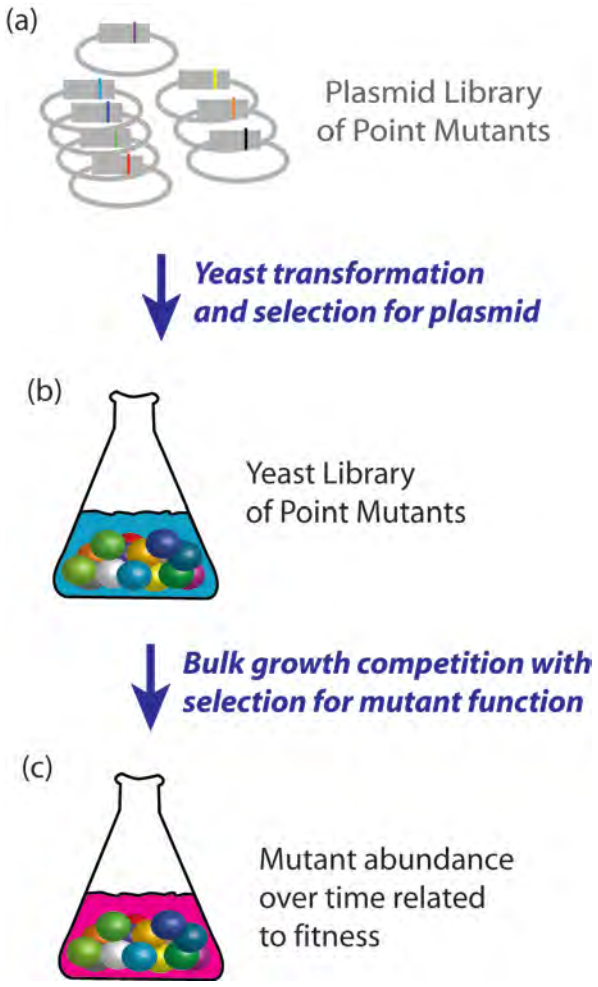


Figure 1.2. Bulk competition of libraries of point mutants in yeast. (a) Plasmid libraries are transformed into yeast. (b) Yeast that have taken up a plasmid are selected for and amplified. (c) Selection pressure is applied to the library copy of the mutated gene and samples are collected over time in bulk-competition.

Figure 1.2. Bulk competition of libraries of point mutants in yeast.



Fitness measurements and application to experimental evolution

Mutations allow adaptation of organisms under selective pressures through natural selection.[21] While this theory has been extensively studied in population genetics studies of natural populations, until recently, there has been little experimental methodology available to scan the mechanisms of mutant accumulation in the laboratory setting. Recent work on hsp90 by Ryan Hietpas, *et al* from our group provided convincing experimental support for the nearly neutral theory of evolution posited by Dr. Mootoo Kimura based on predictions of population genetics models in 1983.[22] This work examined the tolerance of the highly conserved multifunctional hsp90 protein to mutation, which resulted in a biphasic distribution of growth effects for point mutations depending on the position in the protein.[13] Most mutations in this study were either well tolerated (nearly neutral), or highly deleterious. The work presented in chapter 2 of this dissertation demonstrates a similar result for all point mutations of residues 2-76 of ubiquitin in *Saccharomyces cerevisae*.

More insights from the research presented in chapter 2 of this thesis lend support to the theory proposed by Dr. Emile Zuckerlandl in 1976 that the functional density of proteins (i.e. the magnitude of molecular interactions of a given residue) correlates with mutational sensitivity of a given residue.[23] There are over 40 distinct co-crystal structures of ubiquitin and binding partners in the Protein Data Bank, and we were able to strongly correlate the interaction density of Ubiquitin among all of these co-crystals with the mutational sensitivity of mutants at each residue in our screen. Interestingly, the most

sensitive mutants (and those with the most buried surface area in over 40 co-crystal structures) clustered on one side of the 3-dimensional structure of ubiquitin, indicating a common binding motif for many clients.

High throughput analysis of specific function

Proteins contribute to cellular function through biochemical interactions. The specific amino acid composition of a protein dictates its structure and interaction with partner molecules by interaction surfaces created by the three dimensional location and physical properties of its amino acid side chains.[1,24] The complexity of interactions that can be mediated by different combinations of the 20 common amino acids of proteins is staggering, especially with large proteins composed of hundreds or thousands of residues. This makes predicting the structure and function of the majority of proteins from amino acid sequences extremely difficult,[25] and highlights the value of high-throughput approaches to determine the specific effects of amino acid substitutions within a given protein.

Coupling high-throughput mutagenesis techniques with protein display on the surface of phage[26,27] (and later yeast surface display)[28] has been used successfully since the early 1990's to screen for protein variants with increased binding affinities to a particular target. Because of limitations in sequencing technology, these types of experiments were usually focused on mutants that were enriched after selection by repeated cycles of binding to a desired target, washing, and eluting. Alternatively,

proteins displayed on a cell surface could be separated by FACS and then sequenced in low/medium throughput (tens to hundreds of variants). Variants with the desired binding properties could then be identified by sequencing DNA from the enriched phage or yeast populations. The limitation of these strategies is that quantitative analysis in differences of interaction properties is difficult, making it difficult to detect variants with reduced affinity.

The advent of next generation nucleic acid sequencing technologies has made it possible to identify millions of DNA sequences from a single pool of DNA. This capability has made it possible for researchers to scan hundreds of thousands of protein variants in single display experiments and make quantitative estimates about relative binding properties. For example, Doug Fowler and Carlos Araya *et al* while in Stan Fields' laboratory utilized phage display coupled with deep sequencing analysis in a technique they refer to as "deep mutational scanning," to analyze the binding of >600,000 amino acid variants of a human WW domain to its peptide substrate, and were able to report features important to binding that would not be evident from analysis of fewer mutations at each position (e.g. alanine scanning)[12]. These findings were then extended to determine the thermodynamic properties of the WW domain amino acid primary amino acid sequence that affected binding.[16] This group was also able to screen thousands of mutants of the RRM domain of a poly-A binding protein and determine the epistatic effects of substitutions at multiple sites within the same protein. Further evidence of the synergy between these techniques is evidenced by the discovery of activity-enhancing mutations in an E3 ubiquitin ligase by Lea M. Starita *et al*.[18]

It is a significant challenge to determine how a specific biochemical interaction of a given protein affects the overall function of a cell. Some proteins are not essential or only affect the phenotype of a cell under specific conditions (e.g. temperature stress, starvation, or DNA damage), while other proteins perform many functions, all of which can contribute to growth defects. The relationship between a specific biochemical activity and experimental fitness of a protein that performs a single function can be predicted from biochemical flux models.[29,30] In these cases, there is a threshold of specific activity that can be reduced (elasticity function) before observing a decrease in growth rate, and then with further reduction of activity, a decrease in growth rate that can be fit to an exponential function and correlated to growth rate. Nevertheless, many proteins are involved in multiple cellular interactions, and the fitness effect of any given mutation can be explained by either the effect on one rate limiting step or a combination of multiple processes, making predictions of mutant effects from a single function extremely difficult. Therefore, high throughput measurements of mutant libraries for specific biochemical functions can be compared to experimental fitness data to determine the effect of a given function on experimental fitness. Chapters II+III of this dissertation report high-throughput measurements of the relative E1 reactivity of all ubiquitin point mutants, and insights into how this one essential activity correlates with the growth rate in yeast.

Ubiquitin is a 76 amino acid protein that predominately acts as an intracellular signaling molecule by covalently attaching to substrates via its c-terminus or lysine (K)

sidechains.[31,32] Ubiquitin can also form chains of varying length and structure by covalently attaching to other ubiquitin molecules through any of its seven lysines, as well as the amino and carboxyl termini. Many different structural orientations can be formed through the various polyubiquitin linkages mediated by E1, E2, and E3 ubiquitin ligases, and these structures can be specifically recognized by different binding partners and receptors (as reviewed by Daniel Finley in the 2009 Annual Review of Biochemistry).[32] For example, K48-linked tetrameric ubiquitin is required for efficient targeting of substrates to the 26S proteasome for degradation.[33] Attachment of single molecules of ubiquitin to substrates via K63 has been shown to be involved in DNA repair, and multiple other linkages (including K63-linked polyubiquitin) are involved in everything from signaling proteins for protein trafficking to chromatin remodeling in the nucleus to control gene expression.[34,35]

Ubiquitin interacts directly with hundreds of binding partners in the cell, even excluding substrates that are targeted for proteasomal degradation. Most of these interactions are mediated by ubiquitin binding domains (UBD) and ubiquitin interacting motifs (UIM) homology regions found in most binding partners.[36,37] Activation of ubiquitin by the E1 ubiquitin activating enzyme (Uba1p is the sole E1 in yeast) is essential as the first step for the formation of polyubiquitin chains. This activation consists of a two-step process where ubiquitin is first adenylated by E1, and then the catalytic cysteine of E1 forms a thioester bond with the C-terminus of ubiquitin. The attached ubiquitin can then be transferred to an E2 enzyme that associates with the E1 (dozens in eukaoryotes). The different E2 molecules can then transfer ubiquitin to other

molecules with different linkages according to the specific E2 activity (although there is overlap between linkages formed within even the same E2 enzymes, an area of active research)[38] Ubiquitin can then be transferred to specific substrates by multiple mechanisms. Ubiquitin can either be transferred to a thioster-linkage with an E3 ubiquitin ligase (hundreds in eukaryotes) which directly interacts with substrates and transfers the ubiquitin to the substrate, or alternatively, some E3's interact with RING domains that mediate transfer of ubiquitin from E3 to the substrate.[39] Some substrates also react directly with E2's and ubiquitin transfer is mediated by transfer of the ubiquitin to a HECT-domain containing protein followed by transfer to the substrate.[40]

Ubiquitin is an essential protein in all eukaryotes and defects with the ubiquitination and proteasome pathway are associated in multiple human diseases, including cancer and neurodegenerative diseases.[41] Because ubiquitin is involved in so many interactions, it is difficult to determine precisely which functions are perturbed by any defect in ubiquitin or its conjugation system. Any ubiquitin mutants that effect cellular function may be defective at any one or any combination of processes. To begin to determine how ubiquitin point mutants effect cellular function, we made a yeast library containing all ubiquitin point mutants, and determined the competitive growth rate of each single ubiquitin mutant (growth rate compared to wild type in the same culture) of each point mutant. This results from this study is reported in chapter II of this dissertation, with analysis of the mutational sensitivity of each residue of ubiquitin that correlates defects in growth rates with known interactions of many binding partners. To extend this study to determine how one essential interaction, activation of ubiquitin by

the sole E1 in yeast, we expressed the same ubiquitin mutant libraries on the surface of yeast and tested the ability of the yeast libraries to react with E1. By sorting reactive ubiquitin variants, we were able to quantitatively assess the reactivity of most point mutants (described in chapter III of this dissertation). Because E1 activation of ubiquitin is essential for yeast growth, we expected that variants that were not able to be activated would also be growth defective. Surprisingly, we found that the ability of ubiquitin variants to be activated could be reduced up to 50-fold compared to wild-type ubiquitin and still support yeast growth (figure 3.6). Another unexpected finding was that an overwhelming majority of variants that reacted efficiently with E1 did not exhibit growth defects, even with the potential for perturbation of hundreds of other interactions downstream and independent of E1 activation. The combination of these two studies provides a unique examination into the effects of saturation mutagenesis of a whole essential protein that is involved in many cellular processes.

Structural Insights

The primary amino acid sequence of a protein determines its final folded structure within a cell.[1] Ubiquitin is an extremely stably folded small protein that can withstand temperatures of up to 90º C at pH 4.0 without melting.[42] By analyzing the sensitivity of mutations to residues to the core of ubiquitin, we were able to determine which positions were especially sensitive to mutation, even with substitutions of other branched aliphatic residues (isoleucine, leucine, and valine). These substitutions were not predicted to cause unfolding (Figure 2.7d), and tended to be located at regions in the core that were

proximal to surface residues involved in many interactions (Figure 2.7). These findings demonstrate that high-resolution information about the sensitivity of core residues to mutation can give strong insights into protein function, and highlight regions of the protein that may be involved in local structural changes or protein dynamics. Another potential strength of this technique is to determine the requirements of protein folding in cell membranes, which are difficult to purify and crystallize.

Determining drug resistance by systematic mutagenesis of a protein kinase

Single point mutations that interfere with drug binding are a common mechanism of the development of drug resistance.[43] In chapter IV, Using the overall systematic point mutant screening strategy described above, we were able to determine second site resistance mutants to the oncogenic protein BRAFV600E which is present in many human cancers and 80% of human melanomas. The valine to aspartic acid mutation causes BRAF to constitutively phosphorylate downstream targets in the MEK and ERK pathways. We generated mutant libraries in the kinase domains of BRAFV600E, and were able to identify over a dozen resistance mutants that were resistant to a research-tool inhibitor (PLX4700) analogous to the FDA approved inhibitor Vemurafineb. One of these mutants, L505H, is attainable by a single base substitution, which is the most likely mutation in a single protein drug target[43] and was independently discovered by Tian Xu and colleagues from a vemurafenib-resistant tumor cell line derived from a human melanoma patient (unpublished communications with Michael R. Green).

Summary

My main goal in this dissertation is to clearly communicate the value of direct measurements of the effects of systematic point mutations for the functional analysis of proteins. Especially in light of rich structural and biochemical data available for many protein systems, I believe that high-throughput functional characterization can be used to help solve many complex problems in biology that are difficult without applying new technology. Because it is difficult to predict function from amino acid sequence alone, I hope that systematically characterizing specific biochemical interactions and fitness effects of entire proteins will lead to better understanding of how protein sequences evolve to carry out specific functions. This quantitative information, when gathered across many different protein systems, will help synergistically in the efforts of phylogenetic reconstructions in evolution, as well as predictions of protein structure and function.

These approaches are also very powerful for studying mechanisms of drug resistance in human cancer as well as diseases caused by bacteria and viruses, as mutations at the DNA level that cause single amino acid protein variants are the most likely to cause resistance within a drug target. While mutations outside of protein drug targets can cause resistance as well, these methods are not limited to mutagenizing any one protein, and the fast increase in high-throughput sequencing technology may soon lead to being able to study mutations of many proteins within a system in parallel.

Chapter II – Analyses of the effects of all ubiquitin point mutants on yeast growth rate

This work has been previously published as ***Roscoe, B. P.**, Thayer, K. M., Zeldovich, K. B., Fushman, D. & Bolon, D. N. A. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J. Mol. Biol. **425,** 1363–77 (2013).*

The work presented in this chapter was a collaborative effort of all listed authors. Dr. Daniel N. Bolon designed the initial Ubiquitin mutant libraries. I performed all yeast selection, sequencing, and single mutant generation experiments. Dr. Kelly M. Thayer and Dr. Konstantin M. Zeldovich calculated average surface area of ubiquitin contacts from solved structures in the protein data bank. Dr. Daniel N. Bolon and I both analyzed all data and prepared the figures and initial manuscript. Dr. Daniel N. Bolon prepared the final manuscript.

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 ubiquitin point mutants on yeast growth rate. Many aspects of ubiquitin 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.

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[9], until recently it had only been feasible to analyze the function of systematic mutations using amber suppresser strains of E. coli.[10] 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 microarray[44] and now more commonly by deep sequencing is used to determine the relative abundance of mutants in response to selective pressures.[12–17,45–47]

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.[13,14] 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 nine amino acid loop in yeast Hsp90.[13] 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 substrate proteins.[32] 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.[31] Multiple ubiquitin molecules can be linked through covalent attachment between the C-terminus of one ubiquitin chain and a lysine from another ubiquitin. Lysine 48 in ubiquitin is the only lysine that is essential for yeast growth.[34] K48-linked poly-ubiquitin serves as a degradation signal[48] with four K48-linked ubiquitins sufficient to target substrates for proteasome-mediated degradation.[33] Protein degradation by the ubiquitin-proteasome system is an important regulator of the composition of the proteome.[49] As such, the ubiquitin-proteasome system is required for homeostasis under constant conditions as well as rapid cellular responses to altered external conditions.[50] 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 cycle[51]; and degradation of IκB serves as a key signal in many immune responses.[52] Disruptions to protein degradation pathways can lead to a variety of disorders including neurodegeneration[53] and cancer.[50] Protein degradation pathways have emerged as promising targets for therapeutic drugs, including proteasome inhibitors that are currently in clinical use as anti-cancer agents.[54]

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 regards to a wealth of available structural and biochemical information. In particular, non-covalent binding interactions are critical to ubiquitin function. For example, the covalent attachment of ubiquitin depends on non-covalent interfaces between conjugating enzymes and ubiquitin.[55–58] After covalent attachment to substrates most known functions of ubiquitin, including delivery of substrates to the proteasome, are mediated by non-covalent binding to ubiquitin-binding proteins.[36,59,60] 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 humans[36] are the ubiquitin-interacting motif[61] (UIM) that consists of a single α-helix[62], and the ubiquitin-associated[63] (UBA) domain that forms a three-helix bundle.[64,65] Many UBDs bind to a hydrophobic patch on the surface of ubiquitin that includes residues L8, I44 and V70.[36,66,67]

Alanine scanning of the surface positions in ubiquitin 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, V70 hydrophobic patch as well as the K48 and C-terminal sites of covalent linkage.[67] 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. Ile to Val) to the L8, I44, 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 denaturation[42] 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[68], mutations that increase bulk[69] and some hydrophobic to polar substitutions.[70,71] Compared to their influence on protein folding, the impact of core mutations on ubiquitin 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 yeast growth.[70] 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 ten ultra-sensitive positions where only the wild type 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.

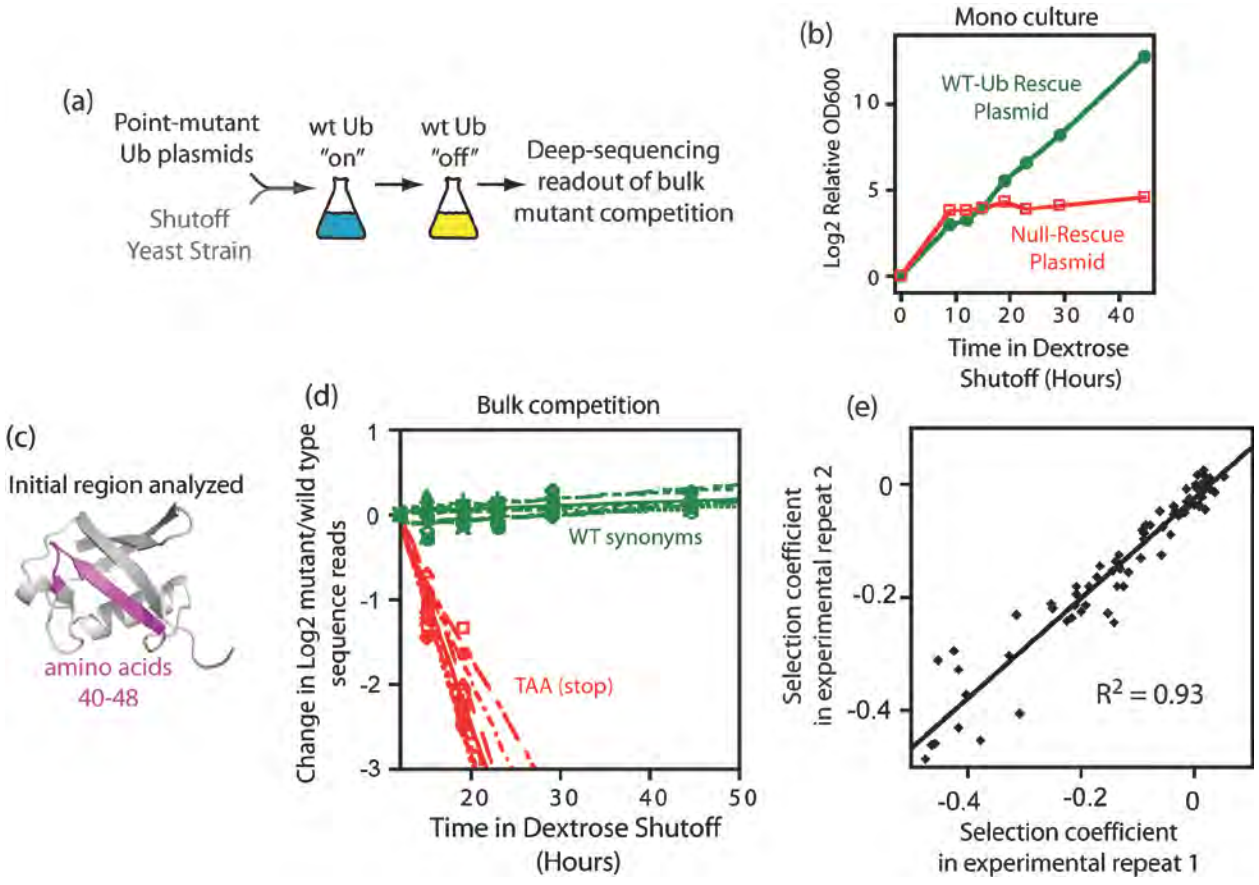
Results and Discussion

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.[34] 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 (Figure 2.1a). To characterize the timing of the shutoff process, we examined the growth of Sub328 cells harboring either a rescue plasmid constitutively expressing WT ubiquitin (utilizing a promoter and plasmid system previously developed to analyze ubiquitin mutants in yeast[34]), or a control plasmid lacking ubiquitin (Figure 2.1b). Cells with the rescue plasmid grow rapidly in dextrose media, but cells with the control plasmid stall in growth after about 10 hours in dextrose media. Based on these results, we decided to analyze selection on mutant libraries starting after 12 hours in dextrose so that most cells without functional mutants would have stalled in growth.

Figure 2.1. 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 9-10 amino acid 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 were selected for initial method development. (d-e) Sequence based analyses of bulk competition of libraries 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.

Figure 2.1. Bulk competition analyses of the effect of ubiquitin mutants on yeast growth.



We examined the robustness of bulk competitions by analyzing a nine amino acid region (Figure 2.1c) 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 sequencing[14] 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 hours, and then switched to dextrose media for 50 hours. 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[13]. The fitness effects of WT synonyms (silent mutations) and stop codons (nonsense mutations) served as important internal positive and negative controls (Figure 2.1d). 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.[72] The slope of mutant to WT ratio versus time in WT generations was calculated and represents the selection coefficient (s) where s=0 indicates wild type 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 ubiquitin amino acid substitutions on yeast growth rate (Figure 2.1e). Compared to fit mutants, strongly deleterious mutants (s<-0.5) showed larger differences in the experimental repeat (Figure 2.2A). 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)[73,74] are subject to effective selection.

Analyzing mutants across the ubiquitin coding sequence

By investigating multiple different regions in parallel (Figure 2.3), we were able to analyze all positions in ubiquitin at the same time. We separated the ubiquitin gene into eight regions each encoding 9-10 amino acids that were amenable to our sequencing-based approach (Figure 2.3a and Supplementary Table 2.1). 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 (Figure 2.3b). In order to assess reproducibility and selection in each region we analyzed both WT synonyms and stop codons (Figure 2.2B). In all regions, WT synonyms were consistently highly

Figure 2.2. Growth properties of select mutants. (A) Correlation between full experimental repeat for amino acids 40-48 across all measurements. (B) EMPIRIC measurements of stop codons and WT synonyms in each region. Error bars indicate standard deviation (N=22-29 for stop codons; N=16-34 for WT synonyms). (C) Growth rate in monoculture measured by changes in optical density over time. SUB328 cells harboring ubiquitin variants on 417GPD plasmids were pre-grown in shutoff conditions for 16 hours, and then monitored for the following 12 hours in shutoff conditions. A total of 17 point mutants were analyzed as well as a positive control (wt) and a negative control (G75D/G76D).

Figure 2.2. Growth properties of select mutants.

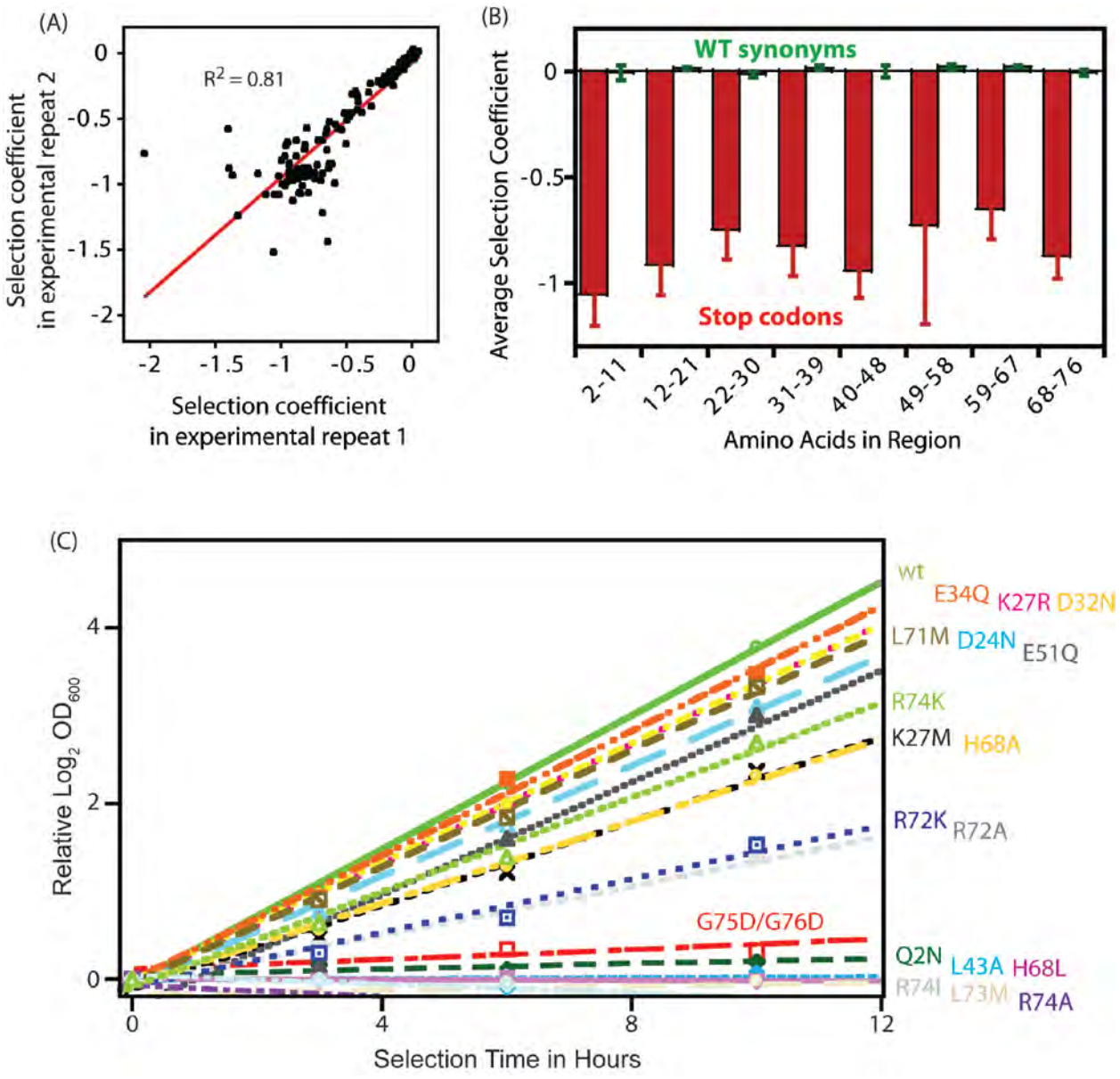
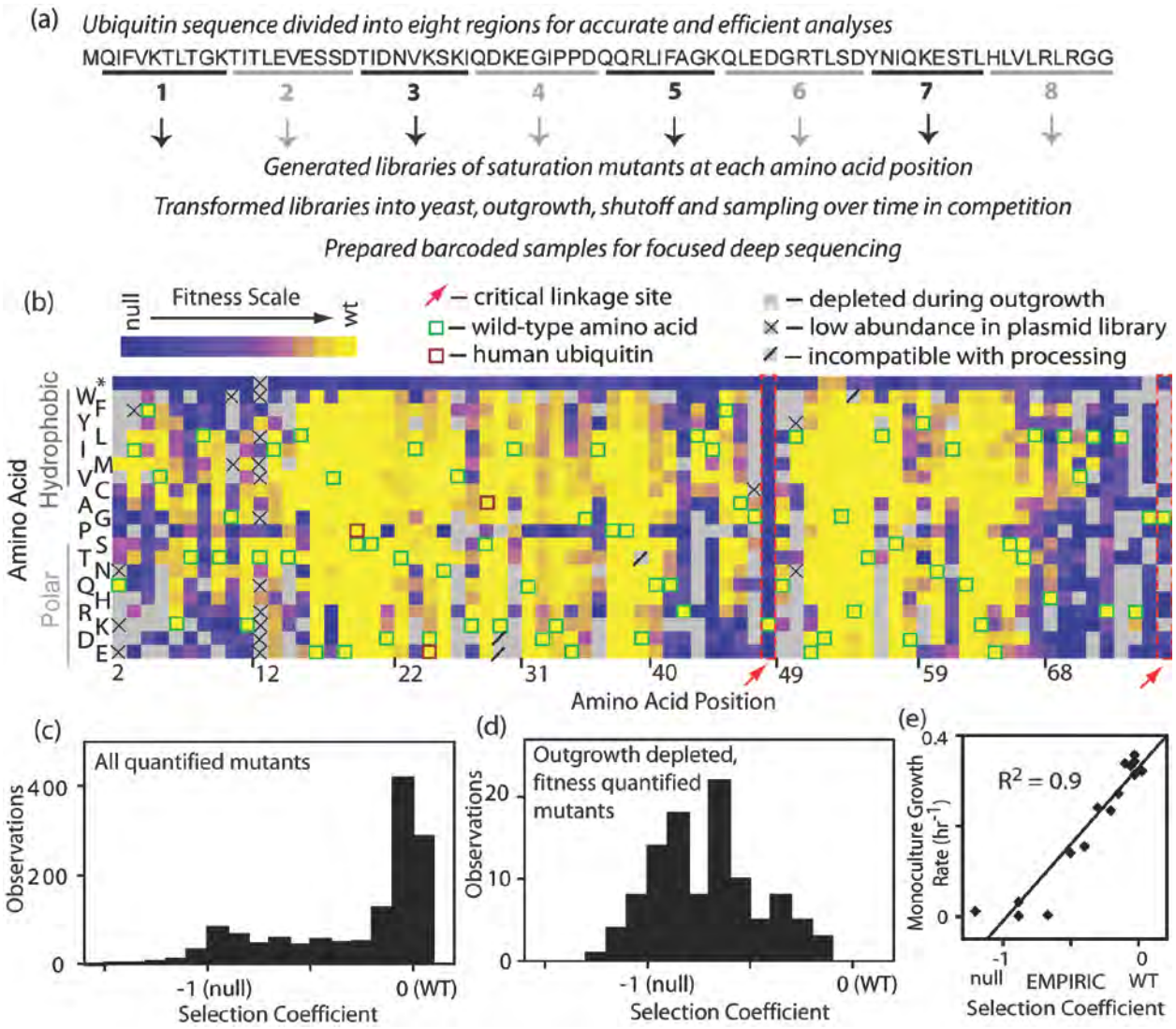


Figure 2.3. Analyses of the growth effects of mutants across the ubiquitin gene. (a) The gene was subdivided into eight regions of 9-10 amino acids and each region was subject to saturation mutagenesis, bulk competition in yeast, and deep sequencing analyses. (b) Heat map representation of the effects 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) Bi-modal 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.

Figure 2.3. Analyses of the growth effects of mutants across the ubiquitin gene.



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 Figure S2.1B), 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 ten 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 timepoints (Supplementary Table 2.1). 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 dataset.

For amino acids encoded by multiple codons, we calculated fitness as the average over all synonyms (Supplementary Table 2.2). 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 2.3b). These quantified mutants exhibited a bi-modal distribution of fitness effects (Figure 2.3c), which has been commonly observed in many different fitness studies.[10,75–78]

At the first time point analyzed in shutoff selection, the relative abundance of some mutants was below our threshold for accurate analysis (colored grey 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 2.3). Sequencing of yeast 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 (grey boxes in Figure 2.3b). We took advantage of mutants that were highly represented in the plasmid library to provide a dataset 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 (Figure 2.3D), indicating that depleted mutations that were not over-represented in the plasmid library (solid grey boxes in Figure 2.3b) 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 Figure 2.3b, 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 the Methods section).

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.[32] In our EMPIRIC analyses, only the WT amino acids at positions 48 and 76 (outlined in dashed red lines in Figure 2.3b) are compatible with robust growth. There are three amino acid substitutions between the yeast and human versions of ubiquitin (outlined in maroon in Fig 2.3b). 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 (Figure 2.2). We find that monoculture growth rate was strongly correlated with EMPIRIC fitness measurements (Figure 2.3e).

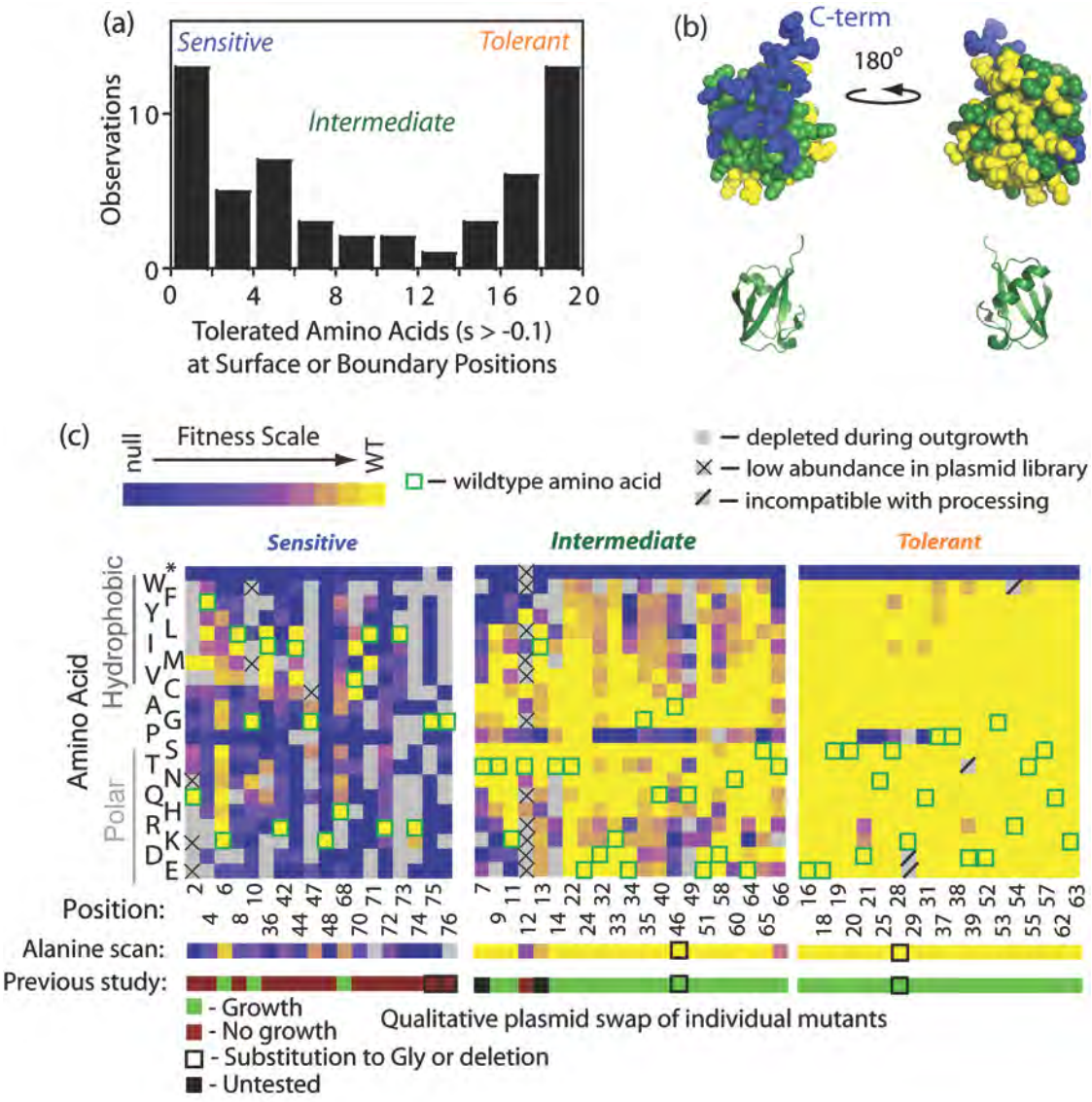
Sensitivity to mutation on the surface of ubiquitin

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 (Figure

2.1c). 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 (Figure 2.4a). 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-ubiquitin structure revealed that sensitive and tolerant positions clustered completely on opposite faces (Figure 2.4b). The fitness-sensitive positions are

Figure 2.4. 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 wild type 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 1UBQ.PDB67) with sensitive positions colored blue, tolerant yellow, and intermediate 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 scan37. At positions where the WT amino acid is alanine, glycine substitutions are shown for both the EMPIRIC and previous alanine scan.

Figure 2.4. Effects of mutants on the solvent-accessible surface of ubiquitin on yeast growth.



located on the β-sheet face, which contains the hydrophobic patch[66] 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[67] (Figure 2.4c). 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 five 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 defines a continuous spectrum of mutational sensitivity from ultra-sensitive to ultra-tolerant.

The most tolerant positions in ubiquitin can accommodate almost any amino acid substitution without disturbing the observed function (Figure 2.4c). 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 ubiquitin), 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 ubiquitin may perform a solubility promoting role as similar properties have been observed in solubility promoting regions of the Hsp90 chaperone[79,80]. 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.[81] 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 ten 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.[32] Of the other three ultra-sensitive 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 Ile) in our assay. These results are consistent with both the binding of partner proteins to this patch as well as 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 was unclear from alanine scanning as it has only a partial (20%) growth defect in our analyses (Figure 2.4c) and was positive for growth in plasmid swap experiments.[67]

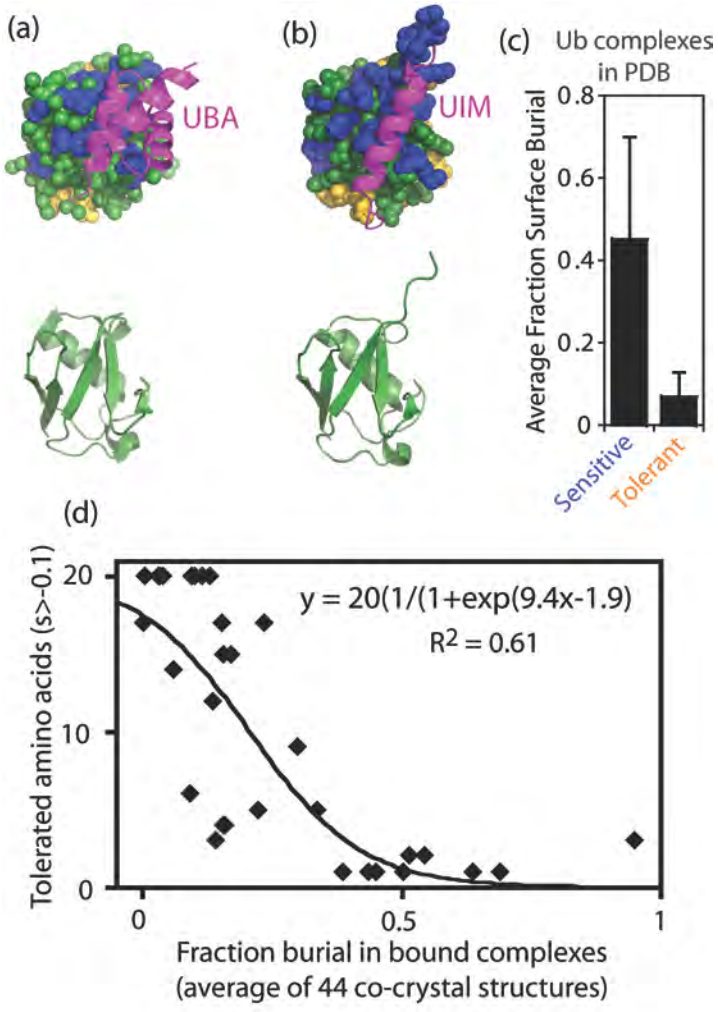
Mapping fitness sensitivity to interfaces

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.[82] Consistent with previous observations, the most sensitive sites to mutation map to known binding interfaces including those with UBA and UIM domains (Figure 2.5a,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 high resolution crystal structures (Supplementary Table S2.4). 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 (Figure 2.5c). 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 (Figure 2.5d). Of note, the relationship between surface burial and tolerance to substitution appears to have multiple phases. At low fraction surface burial many mutations are tolerated and at a threshold around 0.3, amino

Figure 2.5. Relating fitness sensitivity on the surface of ubiquitin to binding interfaces. Structural representations of ubiquitin bound to common binding domains: (a) UBA domain (2OOB.PDB71), and (b) UIM domain (1QOW.PDB72). 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.

Figure 2.5. Relating fitness sensitivity on the surface of ubiquitin to binding interfaces.



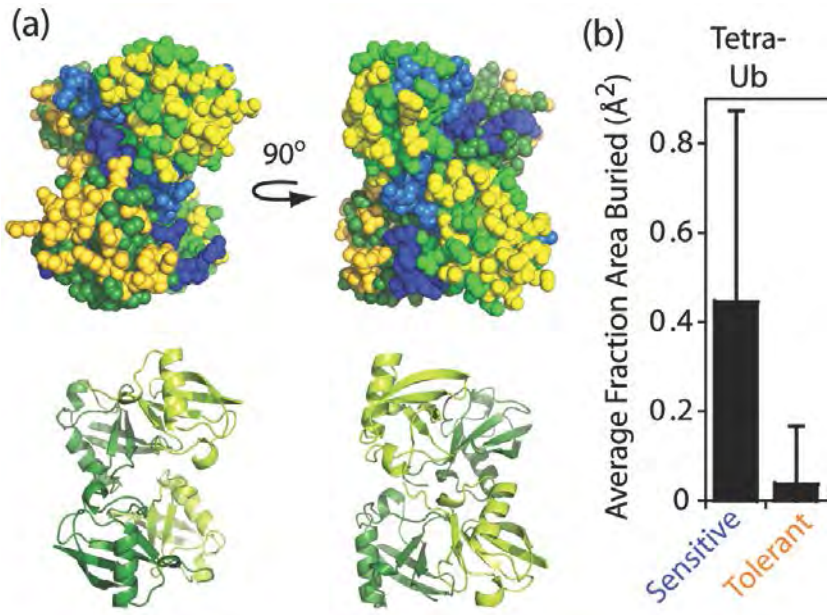
acid tolerance tends to decrease. Positions with a fraction of surface area buried greater than 0.4 are universally sensitive in our analyses. Consistent with these observations, the data fit well to a transition switching model related to those used in chemical denaturation of proteins (Figure 2.5d). These observations demonstrate that binding is a dominant determinant of sensitivity to mutation for ubiquitin and provide quantitative 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 proteasome-targeting.[33] The structure of tetra-ubiquitin 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.[83] 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 (Figure 2.6). This result suggests that the structural arrangement mediated by inter-ubiquitin contacts in closed tetra-ubiquitin may be biologically important. For example by presenting a 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.[84] This potential mechanism is consistent with both an observed "open" conformation of K48-linked tetra-ubiquitin56 and from NMR studies

demonstrating that binding by UBDs can require access to interfaces unavailable in the "closed" conformation. [84]

Figure 2.6. Relating fitness sensitivity to structure of tetra-ubiquitin (a) Structural image of K48-linked tetra-ubiquitin (2O6V.PDB54). Top images show space-filling representation with fitness sensitive positions colored blue, tolerant positions yellow, and intermediate positions grey. Different color shades were used to distinguish subunits. Bottom image illustrates the underlying secondary structure. (b) Fractoin surface area buried per sensitive or tolerant residue in tetra-ubiquitin. Error bars represent standard deviations (N=18 and 19 respectively)

Figure 2.6. Relating fitness sensitivity to structure of tetra-ubiquitin



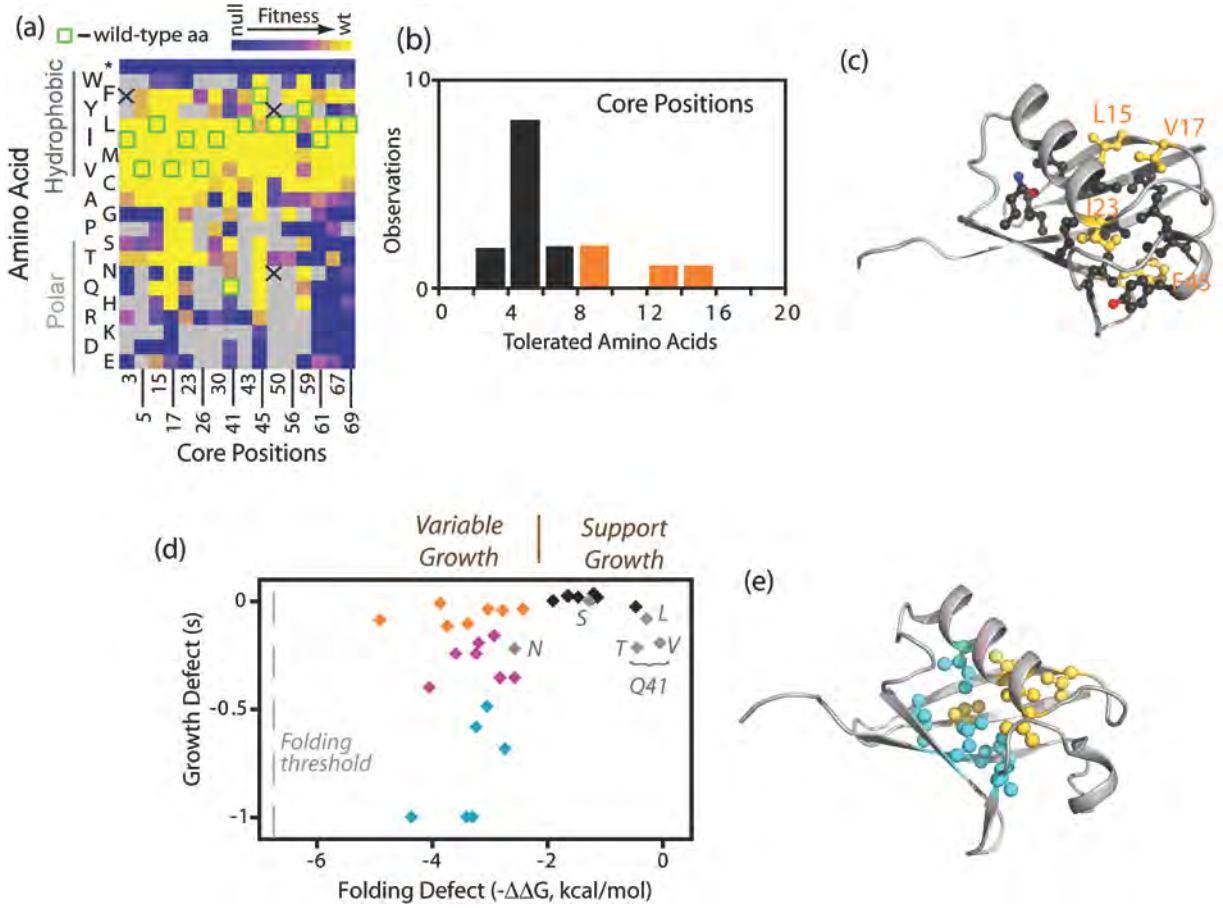
Effects of Mutations in the Solvent-inaccessible Core.

The majority of positions (10 of 16) located in the core of ubiquitin tolerated 3-6 mutations (Figure 2.7a,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 as well as the energetic benefit from burial of hydrophobic atoms.[85,86]

Figure 2.7. 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 wild type sequence (s>-0.1). (c) Structural representation of ubiquitin showing the wild type side chains of core positions. Positions that tolerate more than eight amino acids (s>-0.1) are colored in yellow. (d) Relationship between core mutant impacts on folding stability39; 42 and yeast growth. Previously measured effects on ΔΔG of folding are plotted such that negative numbers represent destabilization. The amount of destabilization estimated to abolish folding is indicated as a dashed grey line on the left. Mutations to Q41, the only WT core polar amino acid, are shown in grey. 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 wild type ubiquitin39. Mutants that are destabilized by more than 2 kcal/mol are shown in orange if they are highly fit (s>-0.12), blue if they are strongly deleterious (s<-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.

Figure 2.7. Mutant effects in the solvent-inaccessible core of ubiquitin.



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.[87]

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 (Figure 2.7c). 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.[68] Of note, a broader folding transition state of ubiquitin was reported in studies using engineered double histidine substitutions and divalent metals as probes.[88] By ϕ analysis, the transition state for 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 ≥ 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 ubiquitin all were observed to have ϕ-values close to zero 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[68,71] observations of individual ubiquitin mutant impacts on the thermodynamic stability difference between folded and unfolded conformations (Figure 2.7d and Supplementary Table S2.5). 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.[89] 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.[86]

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 (Figure 2.7d). Within this stability regime, we mapped mutants with either minimal or severe growth effects onto structure (Figure 2.7e). 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 dataset). 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 (Figure 2.7e). 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[90,91] 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.[92] NMR studies on linked ubiquitin also observed conformational dynamics in the C-terminal region[91], 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 as well as 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 KanMX4-marked yeast high copy shuttle vector (p427) with ubiquitin expression driven by the GPD1 promoter. To aid in cloning, the ampicillin-resistance gene was removed from the vector, and selection during bacterial cloning was performed 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 wild type yeast. Libraries of saturated single codon substitutions in yeast ubiquitin were generated using a cassette ligation strategy in p427GPD as previously described.[14] To facilitate cloning and subsequent sequencing analyses, the ubiquitin gene was subdivided into eight regions of 9-10 amino acids. Pools of saturated point mutants within each region were generated that could be accurately and efficiently interrogated with short-read Illumina sequencing. To distinguish the growth properties of ubiquitin mutants, we utilized the Sub328 yeast strain.[34] 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 described.[93] In order to minimize doubly transformed cells[94], a total of 1 μg of plasmid DNA was transformed per 100 μL competent yeast cells, and after recovery the cells were grown for 48 hours at 30° C in 30 mL of liquid SRGal (synthetic 1% raffinose, 1% galactose) media supplemented with

G418 (200 μg/mL). Cultures were diluted to maintain log-phase growth. After 48 hours 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 $2*10^8$ 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 hours in dextrose. These fitness analyses were performed over 11 generations as the doubling time of the yeast harboring the WT rescue plasmid was 3 hours 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 MmeI 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 MmeI and purified again on a silica column. Barcoded adapters were then ligated and samples processed and analyzed as previously described.[14]

Adapter cassettes with a sticky-end complementary to the resulting MmeI 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 excised and purified. A final round of PCR was performed with Illumina primers. To minimize PCR errors, cycles were limited and the high-fidelity Pfusion enzyme (New England Biolabs) was utilized throughout. For each region, a processing control was included that started with a purified plasmid with wild type ubiquitin and was processed identical to time point samples (same number of PCR cycles, etc.). To distinguish mis-reads during sequencing, a sequencing control was also included 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 described[14] 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 hours 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 MmeI site that would complicate processing were omitted from fitness analyses. The first three time points during library selection (corresponding to 12, 15, and 19 hours 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 was 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 (Figure 2.2). 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 the OD600 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 (1UBQ.PDB67) and the classification algorithm of Mayo and colleagues.[95] We analyzed ubiquitin interfaces by quantifying the surface area buried[96] by each amino acid in atomic resolution structures in the protein data bank. We searched the Protein Data Bank (PDB) for entries with high sequence identity to ubiquitin, at least 2 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-ubiquitin, resulting in 117 ubiquitin chains in 44 PDB structures. We used areaimol from the CCP4 v.6.2 package[97] to calculate ASA for each residue of every ubiquitin molecule both in the complex and in isolation. The resulting changes in ASA

upon complex formation are provided as a supplementary table to this dissertation (Supplementary Table S2.4).

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 an NIH grant (R01-GM065334) to D.F.

Chapter III - Systematic exploration of ubiquitin sequence, E1 activation efficiency, and experimental fitness in yeast

Benjamin P. Roscoe and Daniel N.A. Bolon

Department of Biochemistry and Molecular Pharmacology

University of Massachusetts Medical School

364 Plantation St., Worcester, MA 01605 USA

The manuscript presented in this chapter is currently in submission for publication. I performed all of the experiments described within. Daniel N.A. Bolon and I designed the experiments and prepared the initial manuscript. Daniel N.A. Bolon prepared the final submitted manuscript.

ABSTRACT

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

INTRODUCTION

Determining how genes function together as biological systems is a defining challenge of the genomic era. While genome sequences reveal the DNA blueprint of organisms, deciphering how this blueprint leads to biological function is challenging due in large part to genetic interaction complexity.[98,99] For example, many phenotypes are mediated by multiple genes[100], and numerous genes exhibit pleiotropy.[101] Tremendous progress has been made in mapping the connections (aka edges) between genes and gene products by both genetic[102–104] and biochemical approaches.[105,106] Epistatic analyses of gene knockout combinations have provided a broad understanding of the impacts of node deletions on network function.[102] In addition, approaches have been developed to analyze the effects of disrupting individual network edges by identifying mutations that disrupt a specific interaction.[107–109] However, for most complex biological networks, the elasticity function[110,111] relating network edge strength (e.g. the affinity of a specific protein-protein interaction) to overall network function is poorly understood. To address this challenge, we developed a high throughput strategy to analyze how all point mutations in a central gene impact both an edge to a directly connected node in its network as well as overall network function. Here we report experiments with ubiquitin and the E1 enzyme that provide fundamental insights into regulated protein degradation in eukaryotes.

Systematic investigations of the relationships between gene or protein sequence, function and fitness provide new opportunities to bridge molecular, systems, and evolutionary biology.[7,19,112,113] While a wealth of studies demonstrate that the fitness

effects of mutations are mediated by biochemical changes,[10,29,114–118] most systematic studies of mutants have focused predominantly on either growth effects (e.g. Chapter II of this dissertation)[13,111]or biochemical effects.[12,15,16,108] The relationships between mutant effects on biochemical properties and experimental fitness under defined conditions have been studied using traditional approaches for a handful of genes, almost all of which encode enzymes that catalyze a single critical chemical transformation. In many of these systems,[117,119–121] the experimental fitness effects of a set of mutants can be accurately predicted based on both the proficiency of the mutant enzyme and physiological models of biochemical fluxes.[29] However, for the majority of genes (particularly those that perform multiple functions, or whose functions are not fully appreciated) the relationships between a mutation's impact on biochemical properties and fitness remain unclear. In theory, each activity of a multi-functional protein may contribute independently to fitness and be predicted based on flux models, or the contributions of each activity to function may be interdependent, likely depending on the molecular and evolutionary context of each particular central network node. Distinguishing these possibilities provides insights into network function and can be accomplished by systematically investigating mutant effects on both biochemical function and experimental fitness.

We determined the effects of all possible point mutants in ubiquitin on activation by the E1 enzyme and compared this functional map to a corresponding map of experimental fitness effects in yeast (Figure 2.2). Through its ability to covalently link to other proteins, ubiquitin contributes to multiple important cellular processes including

regulated protein degradation.[31] The covalent attachment of ubiquitin is mediated by a series of enzymes, with E1 activation serving as the first step in this process. E1 activates ubiquitin by first adenylating the C-terminus of ubiquitin and subsequent covalent attachment via a catalytic cysteine in E1.[56,122] We find that most ubiquitin variants that were deficient for E1 activation failed to support robust yeast growth, consistent with the essential role of this reaction.[123] However, our results also demonstrate that activation of wild-type ubiquitin is far more efficient than required to support robust growth and that the relationship between the E1 reactivity of a ubiquitin mutant and yeast growth rate is non-linear. Despite this non-linear elasticity function,[110,111] the effects of most ubiquitin mutants on E1 activation were similar to their effects on yeast growth rate. These observations suggest that most ubiquitin mutations that lead to defects in E1 activation also lead to defects in other ubiquitin-network edges (e.g. binding to the proteasome), and that the combined biochemical defects of these ubiquitin mutants are responsible for the observed fitness defect.

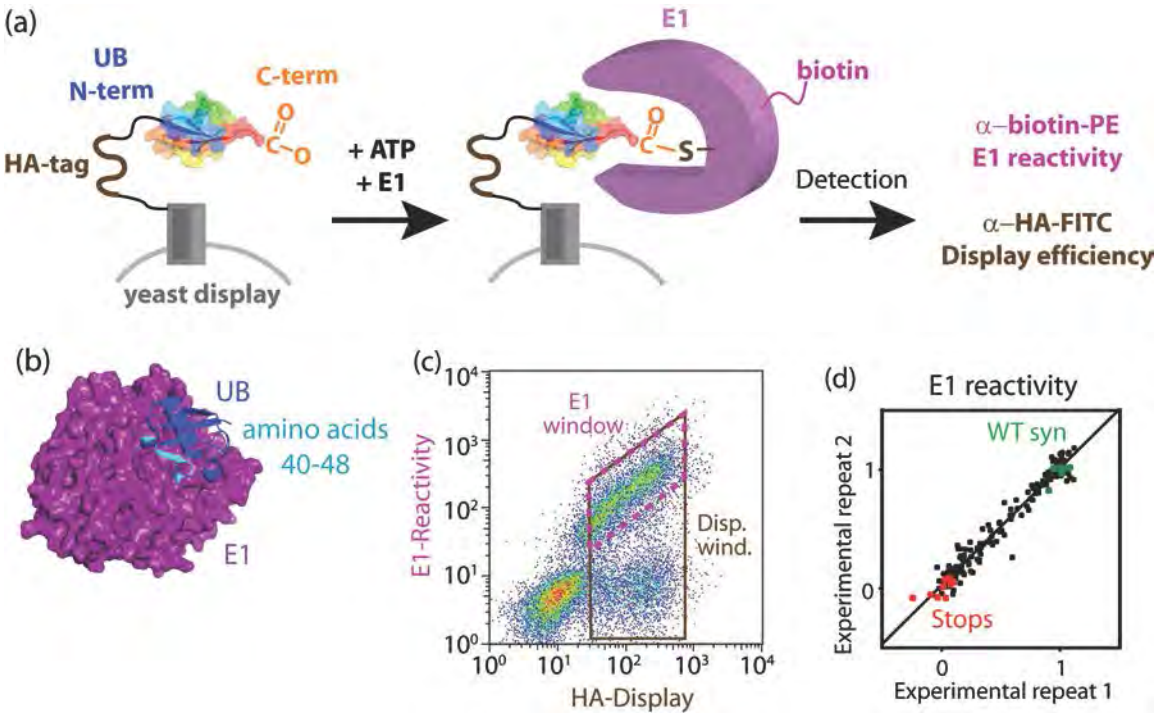
RESULTS AND DISCUSSION

Investigating E1 reactivity

We developed a bulk competition approach to interrogate the E1 reactivity of all possible point mutants in ubiquitin (Fig. 3.1). Comprehensive site saturation libraries of ubiquitin point mutants in eight pools of 9-10 consecutive amino acid positions (Figure 2.2)[14] were transferred to a yeast display system. Importantly, this yeast display setup (Fig. 3.1a) presents ubiquitin molecules with a free C-terminus, which is required for activation by E1. For initial method development, we focused on a region of ubiquitin encoding amino acids 40-48 that form a close contact with E1 in the co-crystal structure[56] (Fig. 3.1b). Display cells were reacted with a limiting concentration of yeast E1 (Uba1) (Fig. 3.2), labeled with fluorescent antibodies targeted to E1 as well as an HA epitope used to control for display level, and separated by flow cytometry into pools of E1 reactive cells and HA displaying cells (Fig. 3.1c). Cells from each pool were subjected to focused deep sequencing[38] and the difference in observed mutant frequency used to assess the E1 reactivity of each ubiquitin point mutant.

Figure 3.1. E1 reactivity of ubiquitin mutants assessed using yeast display and FACS. (a) Experimental setup: the C-termini of displayed ubiquitin variants are free to react with E1 and the upstream HA-tag enables normalization for display level. (b) in amino acids 40-48 that were chosen as an initial test. (c) FACS analysis of pools of ubiquitin point mutants in the test region. Cells in the sort windows were independently collected and analyzed by focused deep sequencing in order to estimate the E1 reactivity of each mutant. (d) E1 reactivity estimates are reproducible in a full experimental repeat.

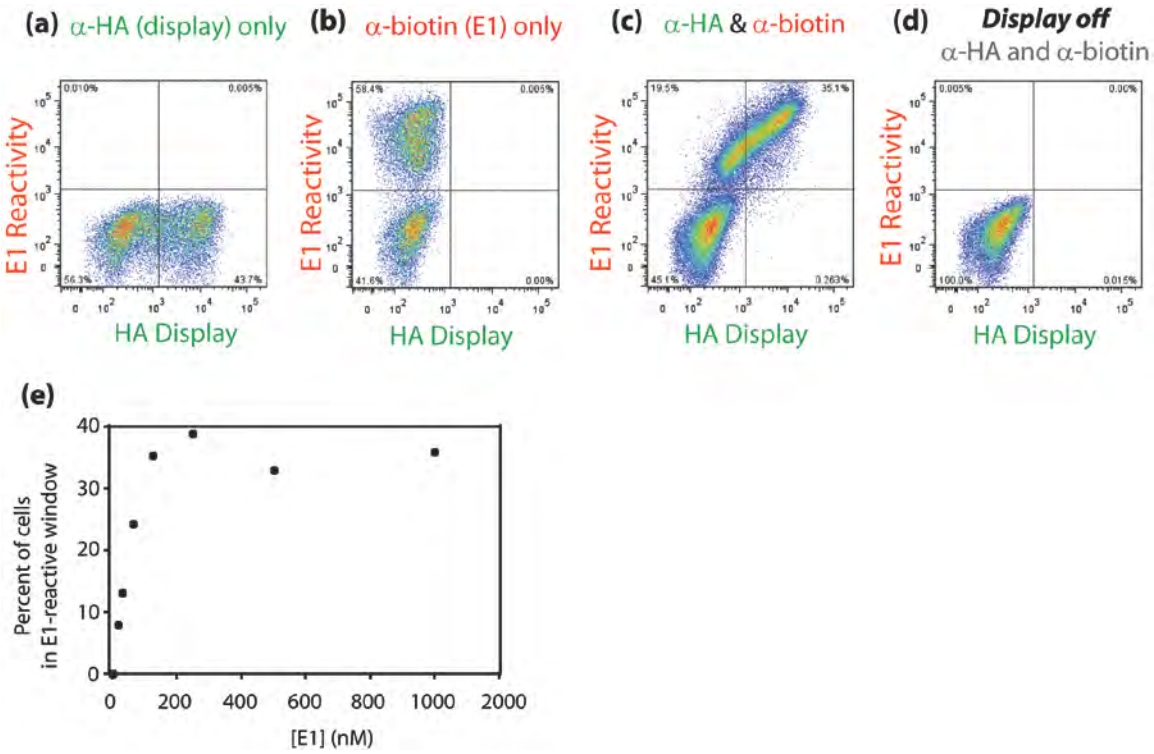
Figure 3.1. E1 reactivity of ubiquitin mutants assessed using yeast display and FACS.



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

Figure 3.2. Flow cytometry controls. Yeast displaying a wild type ubiquitin construct were reacted with a saturating concentration of purified biotin-E1 and analyzed by flow cytometry following antibody labeling directed towards the HA epitope only (a), E1 only (b), or both HA and E1 (c). Control yeast that do not display ubiquitin on the cell surface show negligible antibody labeling (d). As is commonly observed with yeast display, there is a population of cells that do not effectively express this ubiquitin construct on the cell surface. (e) Fraction of yeast displaying wild type ubiquitin in the double positive window as a function of E1 concentration. All further experiments used E1 at sub-saturating concentrations (100 nM).

Figure 3.2. Flow cytometry controls.



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

Given the caveats of our experimental approach, we performed a number of control analyses to assess the quality of our data. The site saturation ubiquitin libraries include all 64 codons at each position and thus encode many wild type synonyms and stop codons. Across all measurements, wild type synonyms exhibit similarly robust E1 reactivity and stop codons exhibit baseline levels of E1 reactivity (Fig. 3.1d), consistent with selection on the ubiquitin amino acid sequence in our screen and the known requirement of the C-terminus of ubiquitin for E1 reactivity. This approach resulted in highly reproducible (R^2=0.96) measurements in a full experimental repeat (Fig. 3.1d),

and was used to interrogate the E1 reactivity of mutants across all of ubiquitin (Fig. 3.3a, Supplementary Tables S3.1 and S3.2). We also developed an independent assay using purified proteins to measure the E1 reactivity of individual mutants relative to wild type (Fig. 3.4). Multiple factors may contribute to distinctions in E1 activation observed in the yeast display and purified protein analyses including: noise in each experiment, variations in selection strength for different regions of ubiquitin in the yeast display experiments, and biochemical distinctions due to yeast surface versus solution reaction conditions. Analyses of a panel of mutants indicated that yeast display E1 reactivity measurements of ubiquitin mutants in different regions correlate (R^2=0.6) with measurements made with purified proteins. Based on these analyses, we conclude that the bulk yeast display studies distinguish highly active from weakly active variants, but that smaller distinctions in relative activation efficiency may not be determined with confidence from the bulk competitions (Fig. 3.4).

Figure 3.3. Mapping the effects of ubiquitin mutants on E1 reactivity to structure. Heat map representation of E1 reactivity for ubiquitin mutants normalized to wild type synonyms and stop codons. (a) Map of entire ubiquitin sequence except for the initiating methionine. Of note, we did not observe any ubiquitin mutants present in our plasmid libraries that failed to efficiently display the HA epitope. (b) Sensitivity of ubiquitin surface positions for E1 reactivity assessed by measuring the number of amino acids compatible with proficient E1 activation within 20% of wild-type ubiquitin. (c) Mapping sensitive (purple) and tolerant (blue) positions on the ubiquitin surface onto the structurally [36] characterized complex with E1 (shown in transparent grey). (d) Correlation between the average impact of substitutions at each ubiquitin position on E1 activation and the fraction of wild-type side-chain surface area buried at the E1 interface. Positions 11, 27, and 35 (colored green in panel d) stand out as sensitive for E1 activation despite not burying side chain surface at the binding interface. (e) The wild-type amino acids at these positions (K11, K27, and G35) all form intra-molecular interactions that likely contribute to the ground state structure and/or dynamics of ubiquitin.

Figure 3.3. Mapping the effects of ubiquitin mutants on E1 reactivity to structure.

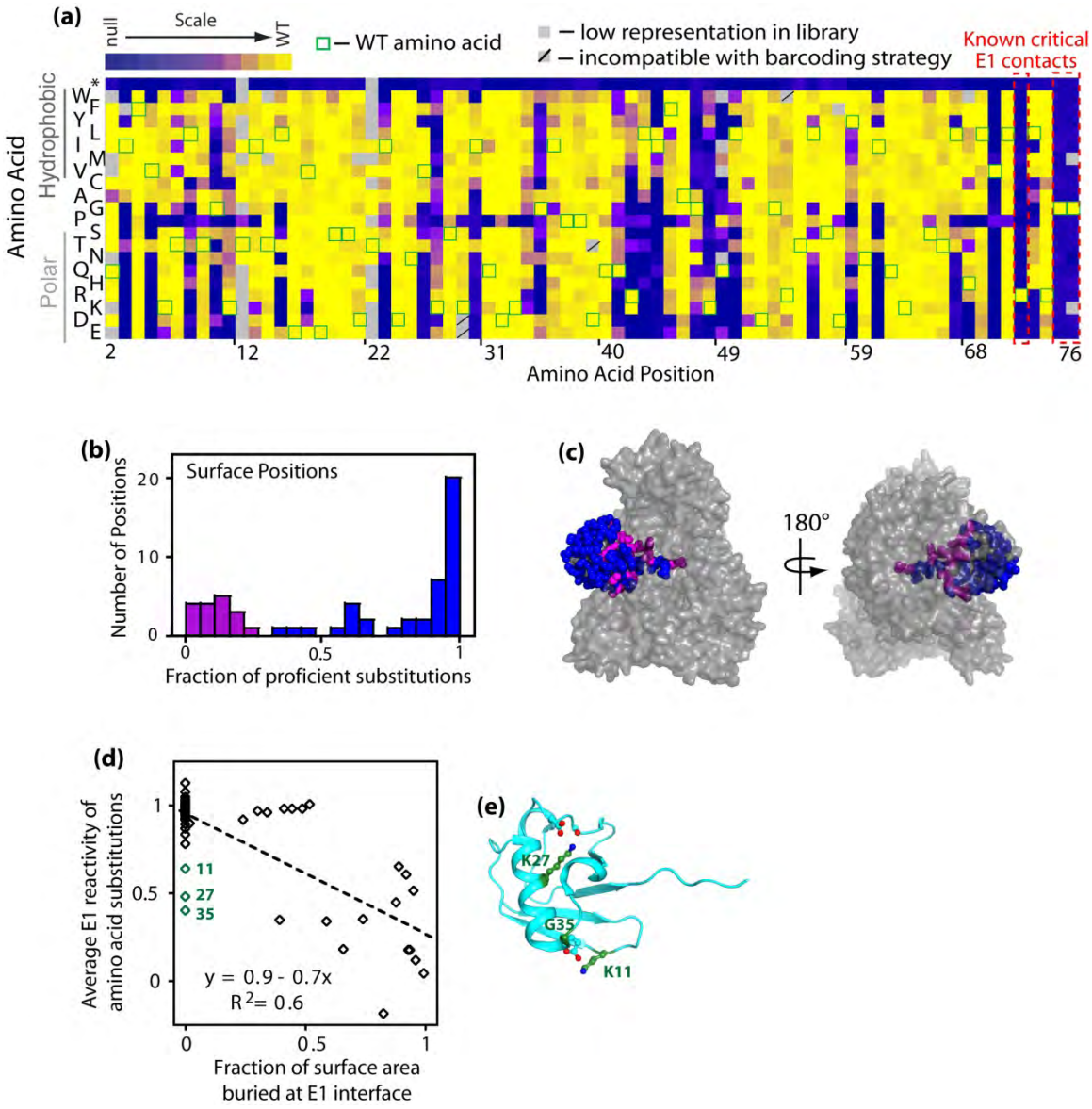
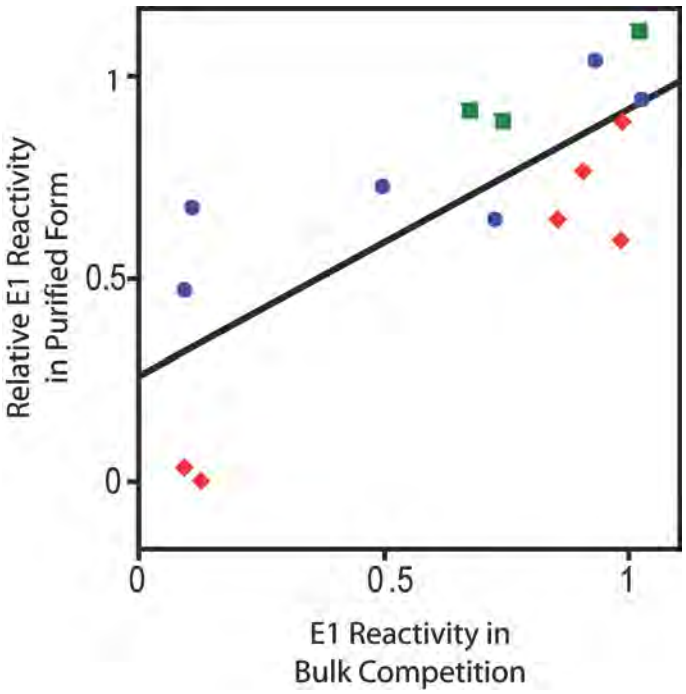


Figure 3.4. Comparison of E1 reactivity estimates from bulk competitions with independent measurements made using purified proteins. Individual mutants from three regions analyzed in separate bulk competitions are distinguished by colors: green squares (K33A, E34G, G35N), blue circles (Q40A, I44M, I44V, K48R), and red diamonds (H68N, H68Q, H68S, L69S, R72S, G75D). Overall estimates from bulk competitions positively correlate (R^2=0.6) with those using purified proteins. Of note, E1 contains two ubiquitin binding sites that would be constrained to interact with identical ubiquitin variants in the bulk competitions, but that would be able to independently access wild type ubiquitin and a mutant in the experiments with purified proteins. Thus, distinctions between surface display and freely diffusible conditions may explain some of the observed discrepancy between E1 reactivity estimates by these two approaches.

Figure 3.4. Comparison of E1 reactivity estimates from bulk competitions with independent measurements made using purified proteins.



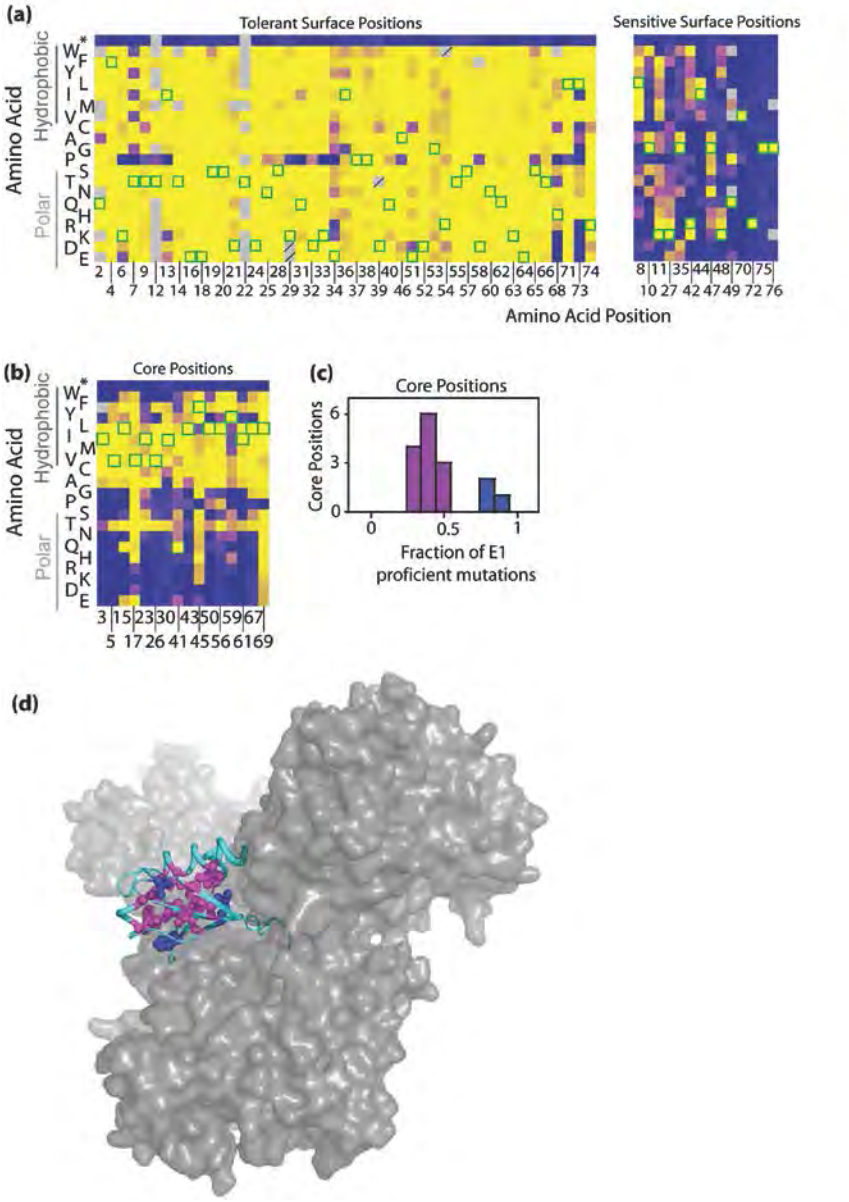
Mapping mutant effects on E1 activation to structure

Structural mapping indicates a general correspondence between E1 reactivity and contact surfaces observed[56] between E1 and ubiquitin. To estimate the sensitivity of each ubiquitin position, we calculated the fraction of mutations at each position in ubiquitin that were proficient for E1 activation (within 20% of the average wild type synonym). Of note, 90% of wild type synonyms, but no stop codons classify as E1 proficient under this definition. Most positions on the surface of ubiquitin either tolerated almost every amino acid substitution or were highly sensitive to mutation (Fig. 3.3b, Fig. 3.5a). Mapping the tolerant surface ubiquitin positions to the structure indicates that sensitive positions were located almost exclusively at the interface with E1 and tolerant positions remained predominantly solvent accessible (Fig. 3.3c). The fraction of ubiquitin side chain surface area buried at the interface with E1 correlates with the observed variation in the average effect of ubiquitin mutations on E1 reactivity (Fig. 3.3d), indicating that surface area burial is a major determinant of mutational sensitivity. All ubiquitin positions that bury greater than 60% of their side chain surface area at the E1 interface are strongly sensitive to mutation. Conversely, the majority of ubiquitin positions that do not bury any side chain surface area at the E1 interface are almost completely tolerant to mutation. Three ubiquitin positions that do not bury side chain surface area at the E1 interface exhibit mutational sensitivity that stands out (positions 11, 27, and 35 shown in Fig. 3.3d&e). The side chains at these three positions all make intramolecular contacts that may impact ubiquitin structure and dynamics: K11 and K27 both form salt bridges between different secondary structure elements, and G35 is part of a turn structure and has a main chain

conformation (positive phi angle) energetically disfavored for non-glycine amino acids (Fig. 3.3e). These structural analyses are consistent with the chemical intuition that the functional sensitivity of a position to mutation is primarily determined by direct binding interfaces[82] as well as structural integrity[124] and dynamics.[70]

Figure 3.5. Representations of E1 reactivity for ubiquitin mutants normalized to wild type synonyms and stop codons. (a) Heatmap representation of positions on the solvent accessible surface of ubiquitin separated into tolerant positions (left panel) and sensitive positions (right panel). (b) Heatmap representation of solvent inaccessible core positions. (c) Distribution of sensitivity to mutation for core positions. Positions 17, 45, and 69 stand out as the most tolerant core positions (colored blue). (d) Structural representation of ubiquitin and E1 illustrating the location of core ubiquitin amino acids (colored purple or blue as in panel c).

Figure 3.5. Representations of E1 reactivity for ubiquitin mutants normalized to wild type synonyms and stop codons.



In the solvent inaccessible core of ubiquitin, most positions exhibited a similar pattern of mutational tolerance for E1 reactivity (Fig. 3.5b-d). 13 of the 16 core positions have aliphatic side chains in wild-type ubiquitin (six Leu, four Ile, and three Val) that form a hydrophobic cluster known to be a driving force for stabilizing native protein structures.[125,126] Consistent with observations that the protein folding stability of wild-type ubiquitin is far greater than required for yeast growth (Figure 2.6)[127], we observe that modest substitutions to other aliphatic side chains are generally well tolerated for activation by E1 (Fig. 3.5b). In contrast, substitutions to polar amino acids are poorly tolerated, suggesting that these substitutions likely disrupt the ground state structure or the dynamics of ubiquitin. Three positions that are exceptions to this rule are all located at the edge of the solvent inaccessible core (Fig.3.5c&d) where long polar amino acid side chains may be able to access solvent without disrupting the structure.

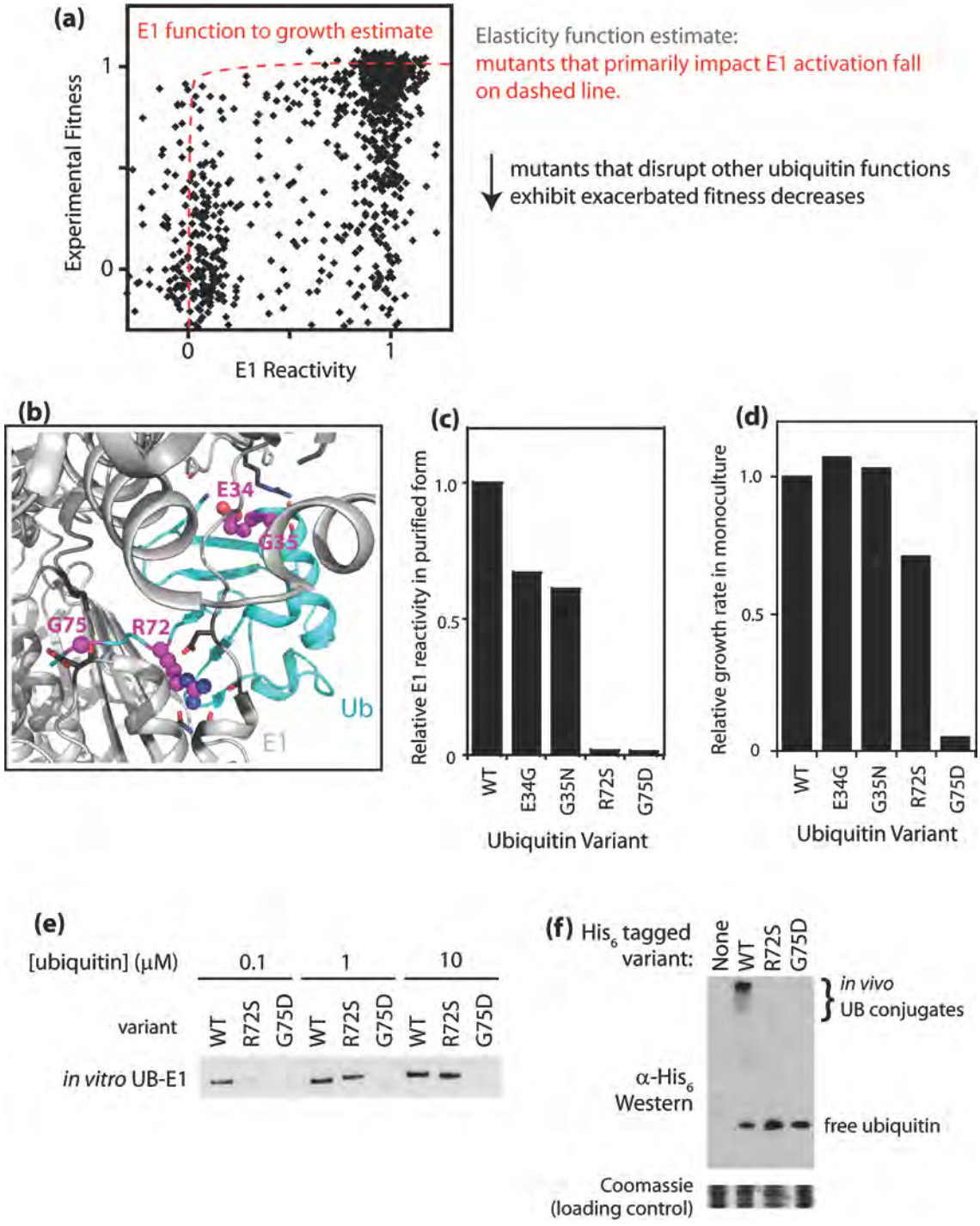
Relationship between ubiquitin mutant effects on E1 activation and experimental fitness.

The fitness effects of ubiquitin mutants integrate over impacts on the entire ubiquitin interaction network. For this reason, ubiquitin mutants with identical impacts on E1 activation can have different fitness effects (e.g., due to different effects of each mutant on binding to and recycling by the proteasome).[70,125,128,129] Comparing the upper-bound of fitness effects to E1 reactivity (see dashed red line in Fig. 3.6a) provides an estimate of the underlying elasticity relationship demarcated by ubiquitin mutations that primarily impact E1 activation. Our systematic scan of ubiquitin mutants indicates that

this elasticity relationship is non-linear and that E1 reactivity can be reduced to baseline levels under the conditions of our screen with minimal impacts on fitness.

Figure 3.6 Relating the effects of ubiquitin mutations on E1 reactivity to experimental fitness. (a) Comparison of ubiquitin mutant effects on E1 reactivity and yeast experimental fitness for all positions in ubiquitin. (b-f) Ubiquitin mutations at four positions located at the E1 interface were analyzed further using purified components. (b) Illustration indicating the location of ubiquitin amino acids E34, G35, R72 and G75. E34 and G35 are located at the periphery of interface between ubiquitin and E1. R72 is located in a deep cavity on the surface of E1 and forms multiple hydrogen bonds across the interface. G75 is located in a narrow cleft adjacent to the active site. (c) E1 reactivity for a panel of mutants at these positions was determined using purified proteins and binary competitions with wild-type. (d) The experimental fitness of this panel of mutants was analyzed by monitoring the growth rate of each mutant in isolation. (e) The E1 activation potential of purified wild-type (WT), R72S and G75D ubiquitin variants analyzed without competition by Western blotting for high molecular weight ubiquitin. (f) The accumulation pattern of epitope tagged ubiquitin variants expressed in yeast co-expressing endogenous untagged ubiquitin.

Figure 3.6 Relating the effects of ubiquitin mutations on E1 reactivity to experimental fitness.



To further assess the minimum level of E1 reactivity for a ubiquitin mutant required to support yeast growth, we investigated a panel of individual ubiquitin mutants (Fig. 3.6b-f). We chose non-conservative mutations located at the structurally determined interface with E1 (Fig. 3.6b) that exhibited activation defects in display competitions. We independently determined the E1 reactivity of each mutant in our panel using purified proteins (Fig. 3.7). Consistent with our bulk experiments, each mutant exhibited reduced E1 reactivity compared to wild type (Fig. 3.6c). The E34G and G35N ubiquitin mutants both reduce E1 activation by roughly 40% but both support yeast growth rates as the sole ubiquitin in cells that are indistinguishable from wild-type (Fig. 3.6c&d). The R72S and G75D mutants were both severely defective (~50-fold) for E1 reactivity relative to wild type. Of these two severely E1 deficient mutants, R72S supported yeast growth albeit at a rate 30% slower than wild type and G75D exhibited null-like growth based on monoculture experiments (Fig. 3.6d, Fig. 3.8).

Figure 3.7. E1 reactivity of ubiquitin variants estimated with purified proteins. (a) Experimental setup: under conditions of limiting E1, fluorescently labeled ubiquitin and unlabeled ubiquitin compete for E1 activation. Titrations of unlabeled competitor ubiquitin variants, separation of reaction products by SDS-PAGE and fluorescent imaging were used to quantify the amount of fluorescent label in E1 conjugates. The resulting plots of fluorescently conjugated E1 as a function of unlabeled competitor were fit to a simple kinetic model in order to estimate reactivity relative to wild type ubiquitin. (b) Fluorescent scan of SDS-PAGE gel with competition between fluorescently labeled and unlabeled wild type ubiquitin. The star denotes a minor contaminant in the FL-UB preparation. (c) Plots of binary competitions for a panel of unlabeled competitor ubiquitin variants. The impacts of these mutants on experimental fitness in yeast are noted on the right side of the panel. (d) Coomassie stained gel of purified unlabeled ubiquitin variants with intervening lanes removed for clarity (indicated with white space). (e) Reactivity measures with purified proteins are highly reproducible for replicate experiments using unlabeled wild type ubiquitin (mean = 0.95, standard deviation = 0.07, N=3).

Figure 3.7. E1 reactivity of ubiquitin variants estimated with purified proteins.

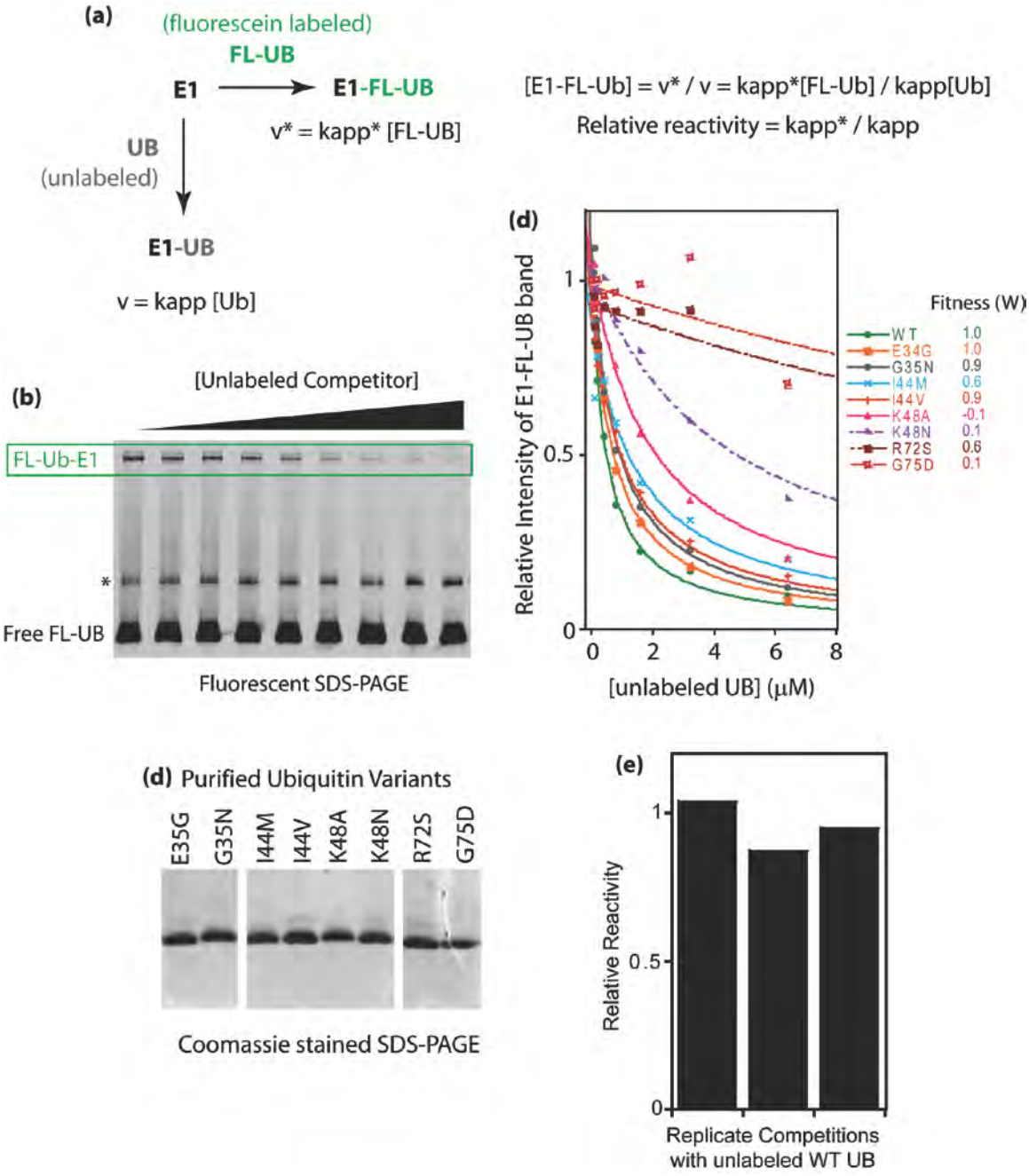
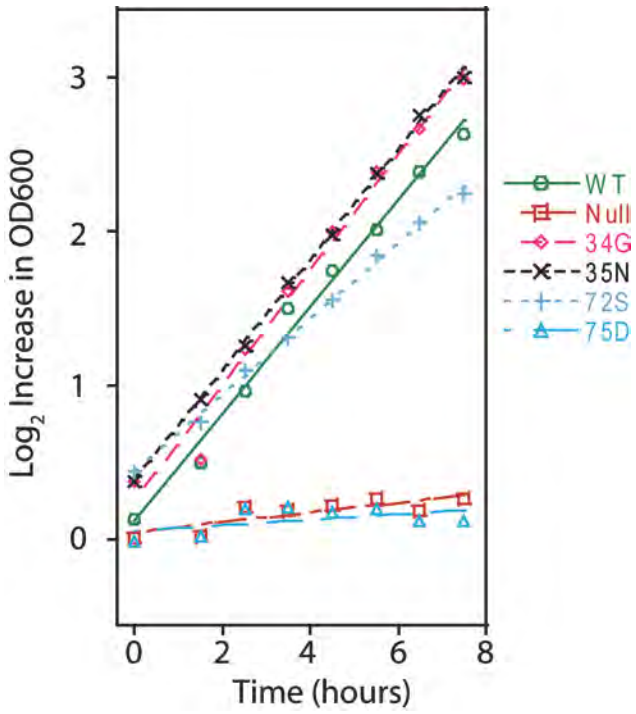


Figure 3.8. Monoculture yeast growth rate supported by a panel of ubiquitin point mutants as the sole expressed ubiquitin.

Figure 3.8. Monoculture yeast growth rate supported by a panel of ubiquitin point mutants as the sole expressed ubiquitin.



To further investigate the E1 activation potential of R72S and G75D, we tested them individually with varying concentrations of E1 (Fig. 3.6e). At low E1 concentrations, both R72S and G75D reacted poorly with E1 compared to wild-type ubiquitin providing an additional confirmation of the activation defects of these mutants. At higher concentrations, R72S was capable of reacting with E1. We did not observe reaction of G75D with E1 even at concentrations 100-fold greater than those where we observed reaction with wild type, or 10-fold greater than for R72S. G75D was recently recovered in a phage display selection for E1 reactivity,[130] which may be due to the use of non-covalent and unstable[131] Fos-Jun mediated association between ubiquitin and phage particles, or other distinctions between the experimental setups. Our observations with purified proteins show that G75D ubiquitin is severely defective for E1 reactivity.

The ability of R72S ubiquitin to support yeast growth, albeit with an approximately 30% defect relative to wild type, was unexpected as position 72 is the main determinant of activation specificity for ubiquitin-like proteins.[55,57,132] Consistent with the importance of R72 in E1 activation, our binary competitions with purified proteins (Fig. 3.6c) indicate that R72S ubiquitin is activated by E1 approximately 2% as efficiently as wild type. This represents an upper estimate on the E1 reactivity of a ubiquitin mutant required to support yeast viability, as we cannot rule out the possibility that the R72S mutation impacts other ubiquitin functions.

To examine how the E1 reactivity that we observed *in vitro* extends to *in vivo* utilization of ubiquitin, we measured the accumulation pattern of the R72S and G75D

ubiquitin variants in yeast cells. In cells, ubiquitin exists primarily in two pools: free ubiquitin monomers of low molecular weight, or covalent conjugates of far greater molecular weight (depending on the mass of the targeted protein and the number of ubiquitin molecules attached). To examine how ubiquitin variants accumulate in these two pools, we co-expressed untagged wild-type ubiquitin with mutant versions tagged with an epitope tag that is compatible with *in vivo* function.[34] The separation of denatured cell lysates by gel electrophoresis followed by Western blotting for the epitope tag enabled estimation of the fraction of the tagged ubiquitin variant incorporated into conjugates while in competition with wild type ubiquitin in cells. These experiments provide a valuable examination of ubiquitin and E1 in the complex cellular environment, but they do not distinguish E1 activation from contributions of other enzymes (e.g. E2's and E3's) in the conjugation process. While epitope tagged wild type ubiquitin readily accumulated as conjugated species, we did not observe appreciable accumulation of conjugates of either R72S or G75D (Fig. 3.6e), consistent with our observations that both of these mutants are at severe competitive disadvantage for E1 activation relative to wild-type ubiquitin.

Investigating activation potential of ubiquitin mutants with excess E1.

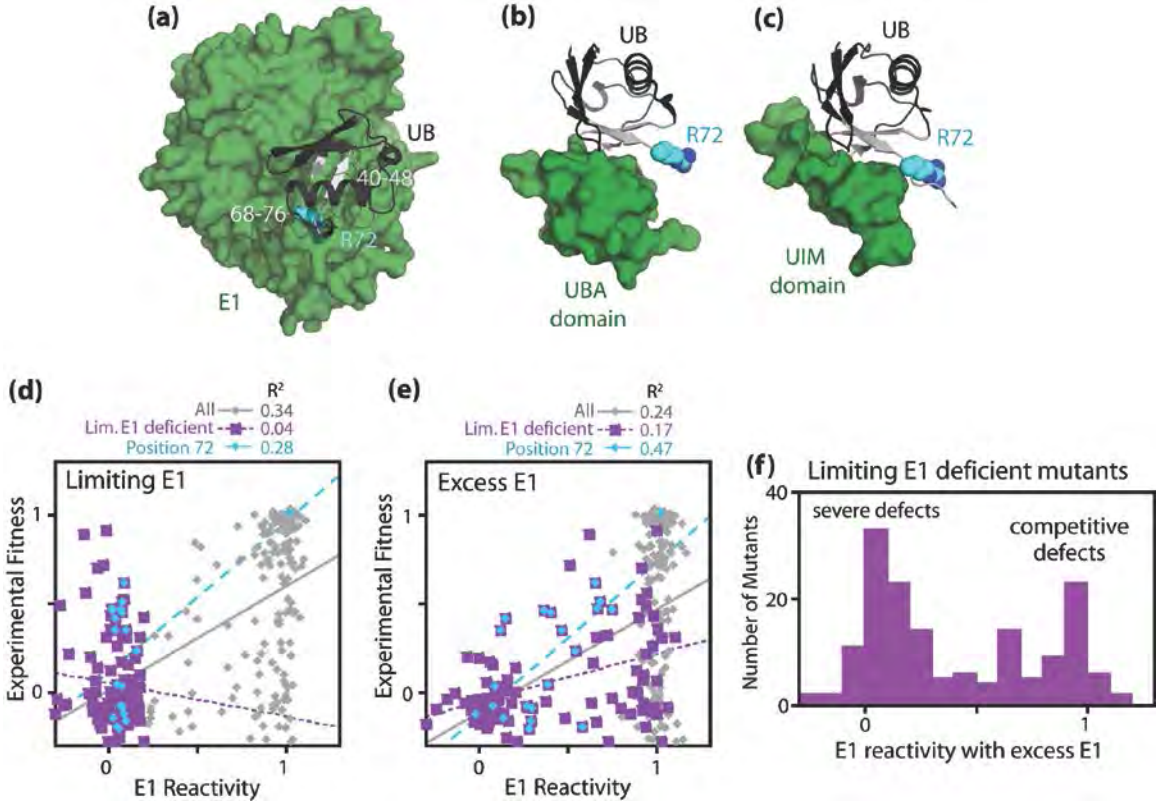
To delineate ubiquitin effects on E1 activity near the threshold required to support robust yeast growth rates, we performed display experiments under conditions of excess E1 for two ubiquitin regions encompassing amino acids 40-48 and 68-76 located at the E1 interface (Fig. 3.9a). Excess E1 in these experiments provides the opportunity for each

displayed mutant to react with minimal competition from other variants and distinguishes ubiquitin mutants with severe E1 activation defects from those with competitive activation defects that may not compromise fitness on their own. The regions we chose to study in these experiments are located at structurally characterized interfaces with other ubiquitin binding partners[36] including ubiquitin-associated domains (UBA) and ubiquitin-interacting motifs (UIM) as illustrated in Fig. 3.9b&c.

The relationship between fitness and activation efficiency with limiting E1 in these two regions (Fig. 3.9d) is similar to the pattern observed across all positions in ubiquitin (Fig. 3.6a). In particular, these regions contain many mutations that cause deficient E1 activation with limiting E1, including some that exhibit growth rates approaching wild-type (Fig. 3.9d). Of note, R72 forms extensive contacts with E1, but is largely exposed in complexes with UBA or UIM proteins such that mutations at this position may primarily impact E1 activation. The contact between arginine 72 and E1 has previously been demonstrated to be important for efficient ubiquitin activation.[55,56,132] As expected based on these previous observations, only the wild-type amino acid at position 72 was compatible with proficient E1 activation under limiting conditions (Fig. 3.9d). In terms of fitness effects, all point mutations at position 72 were deleterious, though they ranged from roughly 40% growth defects to null in estimates from bulk competitions (Fig. 3.9d).

Figure 3.9. Distinguishing the E1 reactivity of ubiquitin mutants near the threshold required to support yeast growth. Two regions of ubiquitin encompassing amino acids 40-48 and 68-76 were analyzed using bulk competitions performed with excess E1. (a) Molecular illustration highlighting the contacts between these two ubiquitin regions (shown in light grey) and E1 (shown in dark green) based on 3CMM.PDB.[56] (b&c) Illustrations of contacts between ubiquitin and two common ubiquitin binding domains (UBA and UIM) based on 1QOW.PDB[133] and 2OOB.PDB[134] respectively. (d&e) The impacts of ubiquitin mutants in these two regions on experimental fitness compared with effects on reactivity with either limiting E1 (d), or excess E1 (e). Ubiquitin mutants deficient for activation with limiting E1 are shown as purple squares and mutations at position 72 are highlighted with a light blue diamond in panels d&e. (f) For ubiquitin mutants deficient for reactivity with limiting E1, the distribution of E1 reactivity observed in experiments with excess E1.

Figure 3.9. Distinguishing the E1 reactivity of ubiquitin mutants near the threshold required to support yeast growth.



With excess E1, the relationship between ubiquitin mutant effects on activation and fitness shifted distinctly (Fig. 3.9e, Supplementary Tables 3.3&3.4). Excess E1 caused an increase in the E1 reactivity observed for many ubiquitin mutants. All ubiquitin mutants with severe activation defects with excess E1 also exhibited deficient growth in yeast, suggesting that E1 activation in this set of ubiquitin mutants is below the level required for growth. While this class of ubiquitin mutants likely has fitness limiting E1 activation defects, they may also have defects in other critical ubiquitin functions due to the structural location of many of these residues at contact sites with other ubiquitin binding domains (Fig 3.9b&c).

We observed many ubiquitin mutants that were activation deficient at limiting E1, but were capable of activating with excess E1 (Fig. 3.9e&f). This class of ubiquitin mutant exhibits competitive activation defects that are not severe enough on their own to prohibit yeast growth, but that would competitively hinder activation in the presence of ubiquitin molecules that are more E1 reactive. Most ubiquitin mutants with competitive activation defects exhibited severely impaired fitness, suggesting that these mutants also caused biochemical defects in other critical ubiquitin properties.

Amino acid substitutions at position 72 resulted in activation efficiencies with excess E1 that correlated positively with fitness effects (Fig. 3.9e). The location of R72 in a deeply buried cleft in the structure with E1 (Fig. 3.9a) and at the periphery of structures of ubiquitin with other binding domains (Fig. 3.9b&c) suggests that mutations at position 72 may primarily impact E1 activation within the ubiquitin interaction network.

Consistent with this structural inference, the activation observed with excess E1 of ubiquitin mutants at position 72 disproportionately correlate with fitness effects compared to mutants at other positions.

Correspondence between ubiquitin mutant effects on E1 activation and fitness.

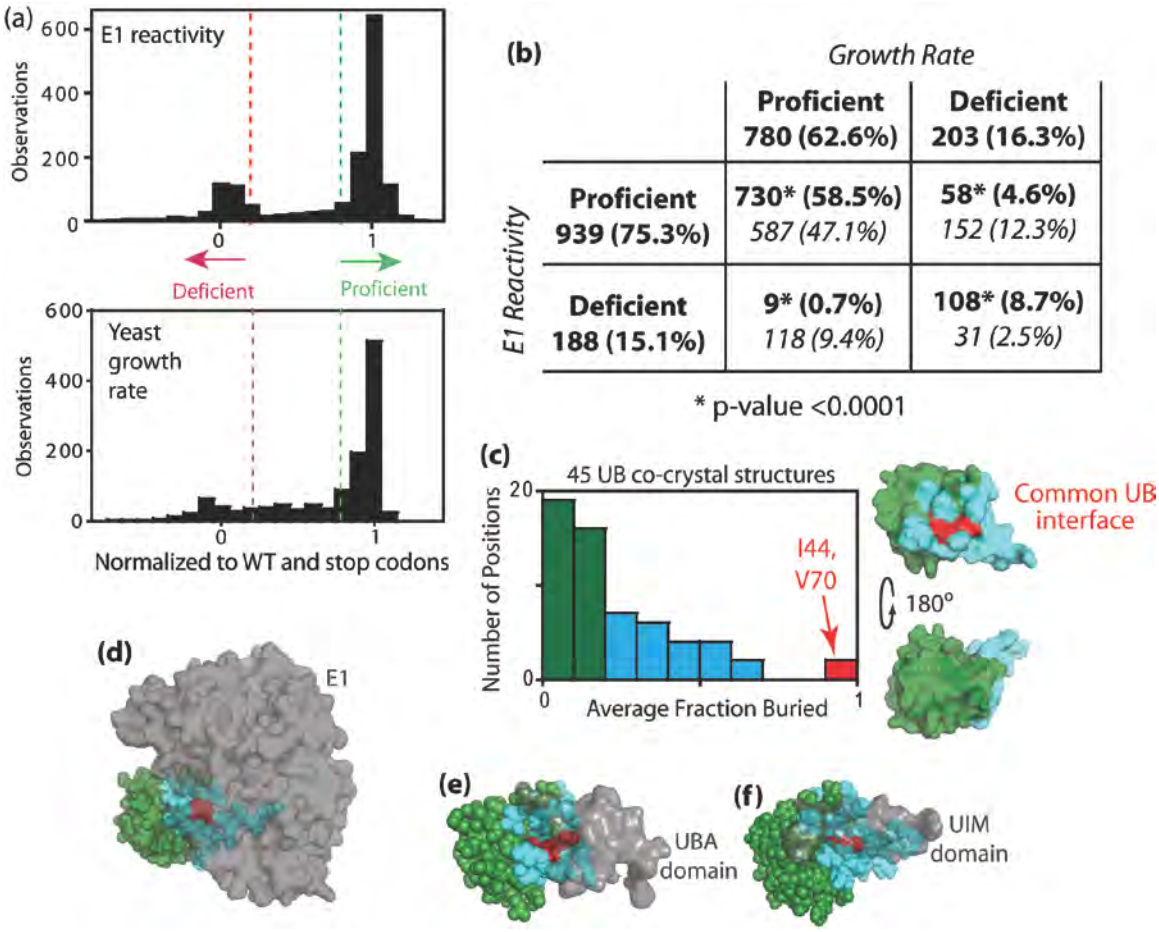
The overall distribution of ubiquitin mutant effects on E1 reactivity under conditions of limiting E1 is similar to the distribution of experimental fitness effects (Fig. 3.10a). Both distributions are bi-modal with a main peak near wild type (defined as 1) and null (defined as 0). These profiles indicate the magnitudes of mutant effects on fitness and this E1 function are similar. In principle, this could be due to a strong or linear relationship where E1 activation is rate-limiting for yeast growth, but this idea is incompatible with our observation that R72S ubiquitin reduces E1 activation 50-fold while impairing growth rate less than two fold. As an alternative hypothesis, we propose that the similar observed distributions of ubiquitin mutant effects may be due to parallel biochemical impacts of many ubiquitin mutants on E1 activation and other ubiquitin functions.

For both experimental fitness and E1 reactivity, we classified each mutant as proficient (within 20% of the average wild type synonym), deficient (within 20% of the average stop codon), or intermediate (Fig. 3.10a&b). To assess the reasonableness of these cutoffs we examined wild type synonyms and stop codons as positive and negative controls. We observe that greater than 90% of wild type synonyms classify as proficient and greater than 90% of stop codons classify as deficient. In addition, these control sets

exhibited no full misclassifications (e.g., proficient misclassified as deficient). The majority of E1 deficient ubiquitin mutations were deficient for yeast growth; and the preponderance of E1 proficient mutations supported robust yeast growth (Fig. 3.10b). The large number of mutants that exhibit robust yeast growth and E1 reactivity indicates that most ubiquitin variants that react efficiently with E1 tend to also function well in all other essential ubiquitin activities. The strong correspondence between E1 reactivity and fitness is consistent with the structurally characterized interfaces between ubiquitin and many binding partners that often center on the isoleucine 44 of ubiquitin.[36]

Figure 3.10. Similar impacts of ubiquitin mutations on E1 reactivity and experimental fitness. (a) Distribution of ubiquitin mutant effects on E1-reactivity with limiting E1 (top) and previously determined[27] effects on yeast growth rate (bottom). Stop codons were not included in the E1 or fitness panels, and four severely depleted reactivity measurements were excluded in order to focus on the main features of the distribution. (b) Contingency table describing the observed overlap of ubiquitin mutants effects on E1-reactivity and yeast growth. All observed frequencies were statistically skewed compared to expectations from independent binomial distributions, which are shown in italics. (c) Analyses of the side-chain surface area buried between ubiquitin and many different binding partners in 45 high resolution co-crystal structures indicate that I44 and V70 are almost always fully buried at the binding interface. Positions surrounding I44 and V70 are buried in a large fraction of interfaces depending on the orientation of the binding partner relative to ubiquitin. (d) The adenylation domain of E1 (shown in gray) almost completely encompasses ubiquitin surfaces that were structurally determined to contribute to binding interfaces with other proteins. (e&f) Domains that commonly mediate binding to ubiquitin bind to smaller surface regions of ubiquitin than E1.

Figure 3.10. Similar impacts of ubiquitin mutations on E1 reactivity and experimental fitness.



Common ubiquitin binding interface

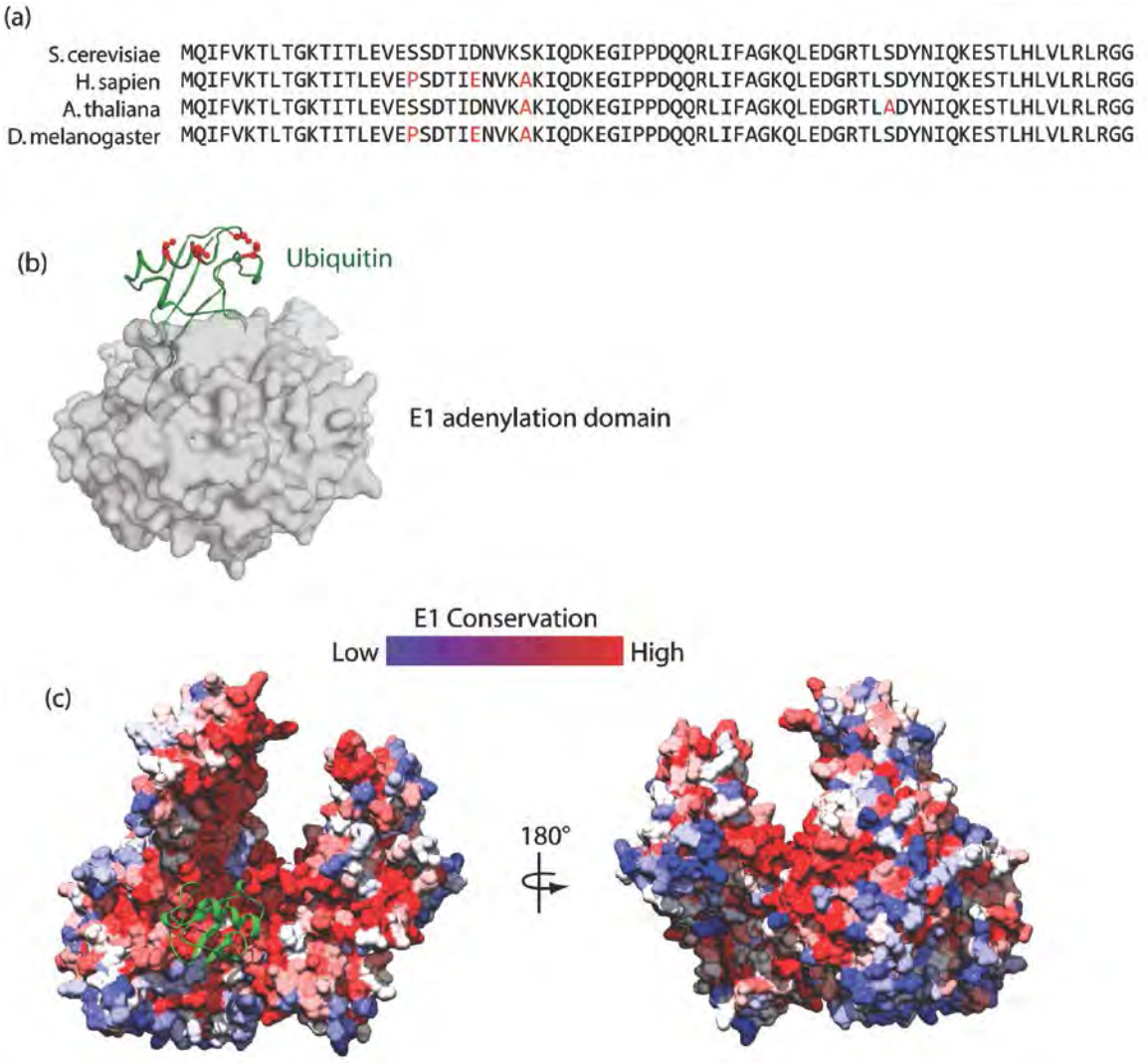
Structural analyses indicate that ubiquitin almost universally contacts partner proteins via a common binding surface (Fig. 3.10c-f). Our own analyses of 45 high resolution co-crystal structures of ubiquitin with a variety of binding partners indicate that the side-chains of I44 and V70 in ubiquitin are almost always fully buried at the interface (Fig. 3.10c) and surrounding positions are buried in a fraction of these structures depending on the shape and orientation of the binding partner (Fig. 3.10d-f, Supplementary Table S3.5). The adenylation domain of E1 forms a very large contact surface with ubiquitin that encompasses nearly all of the structurally characterized contacts between ubiquitin and other common binding domains including ubiquitin-associated domains (UBA) and ubiquitin-interacting motifs (UIM) as illustrated in Fig. 3.10d-f.

One third (3300 $Å^2$) of the total surface area of ubiquitin is buried by contact with E1,[56] including the hydrophobic patch formed by L8, I44, and V70 that is required for binding to proteasomal[66] and many other ubiquitin receptors.[36] This large interface is not a chemical prerequisite for activation, as the chemistry of this reaction is localized to the C-terminal carboxyl group of ubiquitin. The same chemical mechanism is utilized to activate SUMO (a ubiquitin like protein) despite a far smaller interface (1600 $Å^2$)[55]. The strong evolutionary conservation of E1, whose protein sequence is 50% identical between human and yeast[36], suggests that the large ubiquitin-E1 interface has been subject to stringent purifying selection in nature. The structural interface between ubiquitin and E1

is among the surfaces that exhibit the strongest evolutionary conservation in these proteins (Fig. 3.11). Our results indicate that the large E1-ubiquitin interface enables E1 to preferentially activate ubiquitin variants that are functional across the majority of the ubiquitin interaction network.

Figure 3.11. Sequence conservation across species of ubiquitin and E1. (a) Sequences of ubiquitin from diverse eukaryotes. The protein sequence of ubiquitin is ultra-conserved with amino acid substitutions at four positions that are colored red. (b) Illustration of the contacts between ubiquitin and the adenylation domain of E1. Ubiquitin side chains at positions that vary in nature are colored red and cluster away from the interface with E1. (c) Illustration of the structure of E1 and ubiquitin. Ubiquitin is colored green and the E1 is colored based on amino acid conservation observed among diverse fungi [alignment from Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi, Nature 449:54-61]

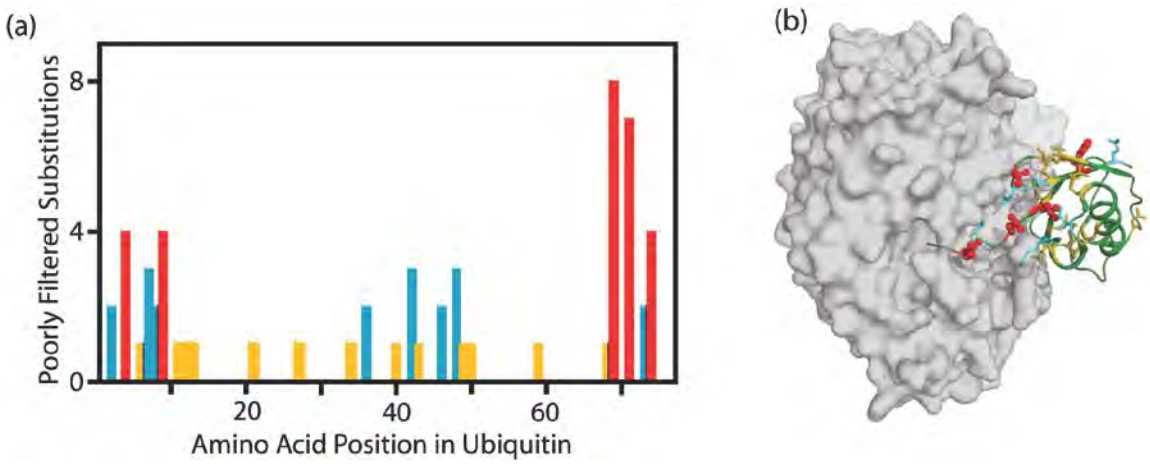
Figure 3.11. Sequence conservation across species of ubiquitin and E1. (a) Sequences of ubiquitin from diverse eukaryotes.



While E1 poorly activates most ubiquitin mutants that are growth deficient, we observed a small fraction of ubiquitin mutants that were activated efficiently by E1, but that were incompatible with robust growth. Of the 939 ubiquitin point mutants that were E1 proficient (Fig. 3.10b), only 59 (6%) were growth deficient (Fig. 3.12). These poorly filtered ubiquitin point mutants were frequently located at the periphery of the interface between ubiquitin and the adenylation domain of E1 (Fig. 3.12), consistent with the chemical intuition that peripheral contacts have smaller contributions to binding and reactivity than central contact points.[135]

Figure 3.12. Ubiquitin point mutants with robust E1 activation, but strong fitness defects. (a) Histogram illustrating the number of point mutants at each position in ubiquitin that were proficient for activation with limiting E1, but that were deficient for yeast growth. Positions are color coded based on the number of poorly filtered substitutions (>3 red; 2-3 blue; 1 yellow). (b) Location of these ubiquitin mutants relative to the interface with the adenylation domain of E1 (shown in grey). Positions are color coded as in panel a.

Figure 3.12. Ubiquitin point mutants with robust E1 activation, but strong fitness defects.

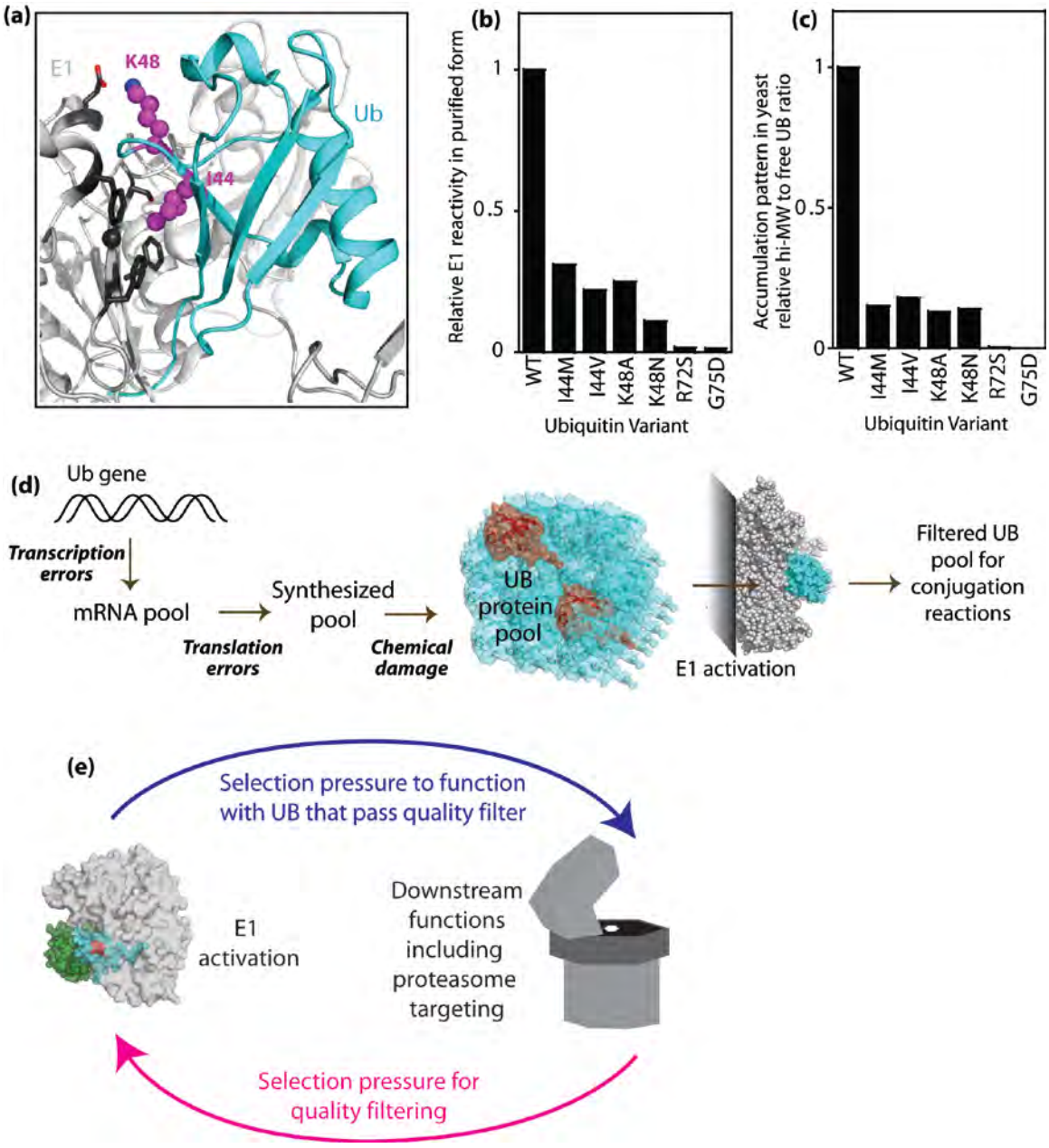


Discriminating Activation by E1

We investigated the effects of ubiquitin mutations at two positions (I44 and K48) that are both critical for many downstream functions, but whose mutational sensitivity for E1 reactivity was not predicted nor discussed in the description of the ubiquitin-E1 crystal structure.[56] I44 is at the center of a hydrophobic patch on the surface of ubiquitin that forms central contacts with most structurally characterized ubiquitin binding domains,[36,66] and K48 is the site of covalent linkage to form ubiquitin polymers that target substrates for proteasome mediated degradation.[33] In the structurally characterized complex with E1, I44 of ubiquitin forms hydrophobic contacts with multiple side chains of E1, while K48 forms a partially solvent accessible salt bridge with E892 from E1 (Fig. 3.13a). While hydrophobic contacts stabilize interfaces, they tend to tolerate slight changes to geometry and often permit conservative substitutions (see Chapter II, Figure 2.6).[12,13,15] Using binary competitions with purified proteins, we observed that E1 reactivity was sensitive to even the most conservative substitutions of Ile to Val at position 44 (Fig. 3.13b and Fig. 3.7). Similarly, partially solvent exposed salt bridges such as the one formed by K48 of ubiquitin with E1 often fail to stabilize protein structures and complexes due to the cost of displacing water molecules from unbound states.[136] However, the ubiquitin K48 mediated salt bridge to E1 is critical for efficient activation as mutations that remove the positive charge at position 48 reduce competitive reactivity with E1 (Fig. 3.13b and Fig. 3.7).

Figure 3.13. E1 inefficiently activates ubiquitin variants with known biochemical defects in downstream pathways. (a) Molecular illustration of contacts between ubiquitin amino acids I44 and K48 with the adenylation domain of E1. Covalent linkage by the ubiquitin K48 side chain is a critical signal for proteasome mediated degradation, and I44 forms direct binding contacts with almost all structurally characterized ubiquitin binding partners. Conservative substitutions to I44 as well as substitutions that removed the positive charge at K48 exhibited decreased E1 reactivity in purified form (b) and a reduced ability to conjugate to other proteins in cells co-expressing wild-type ubiquitin (panel c and Supplementary Fig. S8). (d) The capability of E1 to preferentially activate ubiquitin protein molecules functional in downstream pathways provides the potential for post translational quality control over the pool of ubiquitin protein in wild-type cells that will include ubiquitin protein molecules with errors from synthesis and/or chemical damage. (e) Selection model indicating that pressure for E1 to quality filter ubiquitin protein molecules could lead to parallel selection for downstream processes to function with the pool of ubiquitin variants efficiently activated by E1. This hypothesis is consistent with the correspondence we observe between ubiquitin mutant effects on competitive fitness and E1 activation as well as the extensive binding interface between E1 and ubiquitin that encompasses the interfaces of ubiquitin with other binding partners.

Figure 3.13. E1 inefficiently activates ubiquitin variants with known biochemical defects in downstream pathways.

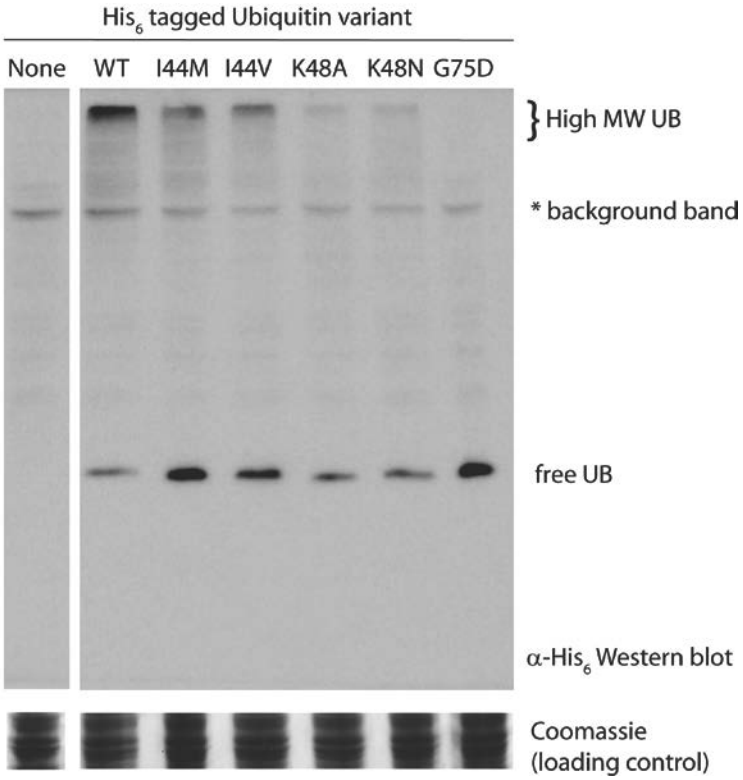


While ubiquitin mutations at position 44 and 48 almost universally cause strong activation defects with limiting E1, many are capable of activation with excess E1. With excess E1, eight substitutions at isoleucine 44 (Q, H, T, A, V, L, M, W) and all substitutions except D and E at lysine 48 could be activated in experiments with excess E1 (Supplementary Table S3.4). These observations indicate that the fitness defects of most mutations at positions 44 and 48 are caused by biochemical defects other than E1 activation, consistent with the known biochemical function of K48 in forming critical polymers and I44 in binding to essential receptors.[36,66]

We investigated how mutants at positions 44 and 48 accumulated *in vivo*. We observed that ubiquitin mutants at I44 or K48 exhibited decreased accumulation as conjugated species in cells co-expressing wild type ubiquitin (Fig. 3.13c, Fig. 3.14). Together with our observation that these mutants exhibit a competitive E1 defect *in vitro*, we infer that E1 contributes to the limited conjugation of these ubiquitin variants *in vivo*.

Figure 3.14. Accumulation pattern of epitope tagged ubiquitin variants expressed in yeast co-expressing endogenous untagged ubiquitin. Yeast lysates were analyzed by Western blotting with anti-His antibody. The star denotes a minor background band present in yeast lysates that cross reacted with the anti-His antibody.

Figure 3.14. Accumulation pattern of epitope tagged ubiquitin variants expressed in yeast co-expressing endogenous untagged ubiquitin.



Post translational quality filtering model

Based on our observations, we propose that E1 can discriminately activate ubiquitin protein molecules that are capable of binding to other partners from the pool of ubiquitin protein in cells that will include molecules with synthetic errors[137] and/or chemical damage (e.g. deamination of glutamine to glutamate)[138] (Fig. 3.13d). Of note, the E1 quality filtering that we propose occurs on the protein pool of ubiquitin generated from the wild-type ubiquitin gene. Estimates of the rate of transcription and translational errors[139] suggest that ~0.1% of ubiquitin protein molecules generated from the wild-type gene will contain an amino acid substitution error. The average observed fitness effect of an amino acid substitution in ubiquitin (s_{avg} = -0.25) leads to a potential fitness benefit from E1 quality filtering of 0.025% (s=0.00025) under the simplifying assumption that all amino acid substitutions are equally probable. In natural populations, selection coefficients above ≈0.0001% ($s \approx 10^{-6}$) would be subject to natural selection based on estimates of the effective population size of yeast[140] and the nearly neutral model.[141] Of note, it is possible that some ubiquitin variants that were efficiently activated by E1, but that were unfit (Fig. 3.10B) may not have been subject to strong selection for filtering due to the amino acid substitution in these variant having a low probability to occur from synthesis errors. These rough approximations of protein synthesis errors and fitness costs indicate that quality filtering by E1 could impact fitness by a magnitude sufficient for selection in natural populations of yeast.

The tendency for partner proteins to bind to a similar ubiquitin surface is consistent with our quality filtering hypothesis, but does not rule out alternative explanations. For example, the ubiquitin surface that binds to E1 has biophysical properties (exposed hydrophobic side chains including I44) that favor macromolecular interactions and likely contribute to this surface serving as a common target of other binding partners. However, the observed biophysical diversity at protein-protein interfaces[24] makes it unlikely that biophysical preferences alone would lead to a near-universal ubiquitin binding surface that is encompassed by the E1-ubiquitin interface, while the quality filtering model would.

In addition, the quality filtering model provides a rationale for the impacts of mutations to K48 on E1 activation efficiency. Our display studies as well as follow up studies with individual ubiquitin mutants demonstrate that E1 selectively filters ubiquitin protein molecules with substitutions at position 48 that cause known biochemical defects when activated and attached to substrates.[33] In future studies, it will be important to further test the quality filtering model. In particular, this model makes the potentially testable prediction that loss of quality filtering by E1 would lead to fitness defects.

The quality filtering model provides a possible rationale for the large contact area observed between E1 and ubiquitin: to enable E1 to extensively interrogate the properties of ubiquitin molecules and discriminate functional ubiquitin variants. All positions at an interface can contribute to relative affinity (e.g. $\Delta\Delta G$ of binding compared to wild-type) and this in turn contributes directly to competitive or relative reactivity for systems under

equilibrium control.[23,82] Thus, all positions at interfaces have the potential to contribute strongly to competitive affinity and reactivity. Consistent with this idea, the interface between ubiquitin and E1 is conserved relative to other surfaces on these proteins (Fig. 3.11). This type of interface-mediated quality filtering may reduce potential toxic consequences from flawed macromolecules in other systems with large contact surfaces (e.g. ribosome assembly). Of note, the expression of flawed proteins can impose a fitness cost even in the absence of aberrant function[142] that is of sufficient magnitude to be under selection in natural populations.[141] Quality filtering of ubiquitin pools by E1 may be particularly important because proteasome-mediated surveillance, an important quality control component for the majority of the proteome,[143] may be unavailable for ubiquitin due to the inherent ubiquitin recycling function of the proteasome.[144,145]

The selection we propose for quality filtering of ubiquitin protein pools by E1 predicts feedback selection such that ubiquitin mutants will often have similar biochemical effects across many ubiquitin functions (Fig. 3.13e). In this model, downstream ubiquitin functions impose selection pressure for quality filtering by E1, and quality filtering imposes feedback selection on downstream ubiquitin functions to be efficient with the set of ubiquitin variants that pass E1 quality filtering. Quality filtering and feedback selection provide a plausible evolutionary rationale for the structurally observed large ubiquitin E1 interface that encompasses interfaces of ubiquitin with other binding partners.

Conclusions

Understanding the connections between function and fitness is a primary goal of many biological disciplines including systems biology and molecular evolution. While sound approaches have been developed to understand the connections between function and fitness for proteins that perform a single function,[29,117,119–121] investigating potential interdependencies in multi-functional proteins had posed daunting technical challenges. Systematic investigations of mutant effects provide a powerful approach to delineate mutations that primarily impact a single network edge from mutations with broad network impacts. Using this approach we find that the activation of ubiquitin by E1 is far greater than required for robust yeast growth, but that most ubiquitin mutations that disrupt E1 activation also disturb other ubiquitin functions. This study demonstrates that systematic mutant analyses provide a powerful approach to investigate how edge-rich protein interaction networks contribute to overall biological function.

MATERIALS AND METHODS

Libraries of ubiquitin point mutants were displayed on the surface of yeast as C-terminal fusions with Aga2-HA similar to previous descriptions.[28,146] Pools of yeast displayed mutants were reacted with E1, labeled with fluorescent antibodies directed to either E1 or the HA tag. FACS was used to isolate E1-reactive cells (E1 and HA positive) and/or HA-displaying (HA positive) cells. Deep sequencing[14] was used to determine the enrichment or depletion of each mutant in E1 reactive cells compared to HA displaying cells. The relative E1 reactivity of a panel of individual ubiquitin variants was

independently determined relative to wild type ubiquitin using purified proteins. The accumulation pattern of His_6-ubiquitin variants in yeast harboring untagged wild type ubiquitin was monitored by inducing expression of the epitope tagged variant followed by Western blotting.

Expression and purification of E1 (Uba1)

The yeast E1 (Uba1) open reading frame was cloned with a biotin ligase acceptor peptide[147,148] encoded at the far C-terminus into a pAC-T7, an expression vector with a T7 promoter and a chloramphenicol resistance marker. This expression plasmid was co-transformed into BLR(DE3) *E. coli* together with pET24-birA to co-express biotin ligase. Cells were grown at 37 °C in 2x YT media to an OD600 of 0.8. Cells were then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 25° C for 6 hours, harvested by centrifugation, and resuspended in IMAC binding buffer (20 mM potassium phosphate, pH 7.2, 300 mM sodium chloride, 10 mM imidazole). Bacterial pellets were lysed with a combination of lysozyme, DNAse I, and sonication in the presence of 1 mM phenylmethanesulfonyl fluoride (PMSF) to inhibit proteolysis. Biotinylated Uba1 was then purified by Cobalt immobilized metal affinity chromatography followed by anion exchange chromatography. Active E1 concentration was estimated by titration with purified wild type ubiquitin, and was routinely 20-40% of the E1 concentration estimated by absorbance at 280 nm. E1 concentrations based on absorbance at 280 nm were more precise and are used throughout the text and figures.

Yeast surface display of ubiquitin point mutants

Systematic libraries of ubiquitin point mutants were generated in the pCTCON2 yeast display plasmid[28] with a galactose dependent promoter driving a fusion of Aga2 with HA followed by a glycine rich linker and ubiquitin with its native C-terminus. Libraries of ubiquitin point mutants were generated in eight pools. Each pool contained mutants in 9-10 consecutive amino acids as described in chapter II (Fig. 2.3). Ubiquitin mutant libraries were transferred into pCTCON2 using SLIC.[149] To facilitate transfer a modified pCTCON2 destination plasmid was constructed with the sequence GCTAGCGATTCTAGAACTAGTAATATGCATGCTCGAGTCATGTAATTAGTTAG GATCC immediately following the HA tag and glycine rich sequence in pCTCON2. This vector was prepared for SLIC by digestion with SphI and treatment with T4 DNA polymerase as described.[149] SLIC inserts were prepared by 8 cycles of PCR with ubiquitin libraries in p427GPD (described in chapter II) as template and forward (GATTCTAGAACTAGTAATATG) and reverse primers (TAACTAATTACATGACTCGAG) that bind immediately upstream and downstream of the ubiquitin open reading frame in this template, followed by treatment with T4 DNA polymerase as described.[149] After annealing of prepared vector and inserts, samples were transformed into competent bacteria and plasmid libraries prepared in bulk as previously described.[14] The library generation procedures were developed to maximize the fraction of the library with relevant point mutations and minimize chances for secondary mutations, especially those outside of the regions directly sequenced and hence undetectable to our enrichment analyses. The starting ubiquitin libraries were generated

using a cassette ligation strategy (described in chapter II) such that all regions outside of those directly sequenced were copied entirely in bacteria where fidelity should virtually eliminate the probability of secondary mutations. In transferring the libraries to the display plasmid, we performed 8 cycles of PCR using Pfusion DNA polymerase (New England Biolabs). According to the manufacturer's estimated error rate for this polymerase (4.4×10^{-7}) and the amplification details, we estimate that less than 1 in 10,000 molecules would have a secondary mutation outside the region that we sequence.

Pooled mutant libraries of each region were transformed separately into the EBY100 yeast-display strain as described in chapter II. Following plasmid transformation, yeast cells were pelleted and washed three times in 1X Tris buffered saline (TBS) to remove extra-cellular plasmid. Each pellet was then resuspended in 50 mL of synthetic dextrose (SD) media lacking tryptophan and uracil to select for transformed cells. Cells were grown for 48 hours (to an OD600 of about 1) at 30 °C in a shaking incubator. Aliquots of approximately 10^8 cells were collected for each library and stored in 20% glycerol at -80°C. Aliquots for each library region were thawed and used to inoculate 50 mL of casamino acid and dextrose media (CAAD). These cultures were grown at 30 °C to near-saturation for 24 hours, and then diluted 50-fold into 50 mL of fresh CAAD. Yeast proliferation was then monitored by OD600 reading and kept in mid-log growth by dilution with fresh CAAD for 16 hours. Cells in log phase were collected by centrifugation and washed 3 times with CAA-RG media (casamino acids media with 1% raffinose and 1% galactose), resuspended in CAA-RG media to an OD600 of 0.5, and grown at 30°C for a further 16 hours. As a control for non-displaying

cells, cultures were also grown in CAAD to repress expression from the gal-inducible promoter.

Labeling and sorting of yeast display cells

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

Labeled samples were sorted for display efficiency and E1 reactivity on a BD FACSVantage DV-1 cell sorter by collecting all FITC-positive cells as one population (HA-display positive), followed by double-positive FITC+PE cells (HA-display positive and E1 reactive). To ensure adequate library coverage, at least 150,000 cells of each population were sorted and collected into sterile SD media. Sorted yeast were amplified

in 50 mL of SD-U-W media (display off) for 24 hours at 30 °C to an OD600 of approximately 1. These yeast samples were collected by centrifugation, washed with TBS, and cell pellets stored at -80 °C.

Quantifying mutant responses to selection by sequencing

Plasmid DNA from yeast pellets was prepared for deep-sequencing as previously described in chapter II. Briefly, plasmid DNA was isolated from yeast and the display ubiquitin open reading frame amplified with primers specific to the pCTCON2 promoter and terminator regions. A second PCR step was used to focus on the randomized region of each library, including adding an MmeI site adjacent to the mutated region. Three-base barcodes each differing by at least two bases were ligated to MmeI-digested samples to differentiate between unsorted cells, HA-positive displaying cells, and double positive E1 and HA-displaying cells. FastQ files from deep sequencing were analyzed as previously described.[14] Raw counts of each mutant were normalized to the wild type ubiquitin sequence count. The relative enrichment or depletion of each mutant in E1 reactive cells to HA displaying cells was calculated in log scale. Because the last amino acid of ubiquitin is strictly required for E1 activation, stop codons at each position should be biochemically null. To normalize for small differences in observed raw enrichment and depletion values for different regions, the apparent E1 reactivity's of mutants in each region were linearly scaled such that the average stop codon was 0 and the average wild type synonym was 1. Of note, analyses of unsorted and HA-displaying cells indicated that all mutations (including stop codons and cysteine substitutions throughout ubiquitin) were displayed with similar efficiency. As described in chapter II, mutations that were

low in abundance in our libraries (mutant:WT ratio less than 2^{-8}) or that introduced an internal MmeI site were omitted from analysis. For wild type amino acids where synonyms were not available or analyzed, E1 reactivity was set to the average of all wild type synonyms in the region (1 by definition).

Monoculture growth rate of yeast with individual ubiquitin mutations

Growth of yeast supported by ubiquitin variants was determined as previously described in chapter II.[70] Briefly, plasmid (p427GPD) encoded ubiquitin variants driven by a constitutive promoter were transformed into a ubiquitin shutoff strain (Sub328[52]). Growth rates at 30 °C were determined in synthetic dextrose media by following the change in OD600 after 12 hours of pre-equilibration in shutoff conditions.

Quantification of ubiquitin activation by E1 using purified proteins

We developed a binary competition assay to determine the E1 reactivity of ubiquitin mutants relative to wild type using purified proteins. We generated and purified wild type ubiquitin with a His_6 tag and a unique cysteine at the N-terminus (MGHHHHHHCGG). Purified protein was reacted with fluorescein iodoacetamide, and fluorescently labeled ubiquitin (FL-UB) further purified by size exclusion chromatography using a superdex-200 column. Competition experiments between fluorescently labeled wild type ubiquitin and unlabeled competitors were setup with 100 nM total E1, 500 nM FL-UB, and a range of competitor concentrations. Reactions were performed at room temperature in E1 reaction buffer (25 mM Tris, pH 7.2, 50 mM sodium chloride, 5 mM magnesium chloride, 5 mM ATP) supplemented with 0.1 mg/mL

bovine serum albumin. After 1 minute, reactions were halted by the addition of sodium dodecyl sulfate to 2%. Reaction products were separated on a non-reducing SDS-PAGE and imaged on a fluorescent imager. The intensity of the FL-UB-E1 band was quantified using the program Multigauge (Fuji) and plotted as a function of the concentration of unlabeled competitor. These plots were fit to a simple kinetic model (Fig. 3.7) to estimate relative E1 reactivity. Relative E1 reactivity was log transformed and normalized (wild type set to 1 and the null mutant G75D set to 0) in order to facilitate comparison to reactivity estimates from display experiments.

Analyzing ubiquitin accumulation in yeast

To examine the accumulation profile of ubiquitin mutants in yeast, we generated inducible epitope tagged ubiquitin constructs. Selected ubiquitin mutants with an N-terminal His_6-Myc epitope were cloned with a galactose inducible promoter[150] into p427 plasmids. These constructs were transformed into W303 yeast cells that express wild type ubiquitin from endogenous loci. Following transformation, single colonies were grown to saturation at 30 °C in synthetic dextrose media, and then grown at 30 °C for 16 hours in synthetic media with 2% raffinose to an OD600 of 1. At this point, a sample of control (uninduced) cells were collected and frozen at -80 °C. The remaining culture was grown in synthetic media with 1% galactose and 1% raffinose for 2 hours at 30 °C. Samples were collected by centrifugation, washed once with TBS and stored at -80 °C. Frozen samples were lysed by vortexing with glass beads and treatment with 2% SDS buffer with 1 mM PMSF at 95 °C for 5 minutes. After removing cell debris by centrifugation, the protein concentration in each sample was determined using a BCA

assay (Pierce). Samples (20 μg of total protein from each lysate) were analyzed by Western blotting with anti-HisG antibody (Invitrogen 46-1008). Multigauge (Fuji) densitometry software was used to quantify both free and conjugated ubiquitin species.

Structural analyses.

Structural analyses were performed with Pymol (Schrödinger) or Chimera (UCSF), and these programs were also used to generate all molecular images. The average surface area buried at structurally characterized ubiquitin interfaces was calculated from surface area measurements described in chapter II (table 2.4).

ACKNOWLEDGEMENTS.

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

Chapter IV - Resistance to vemurafenib resulting from a novel mutation in the BRAFV600E kinase domain

This chapter has been published by Timothy R. Wagenaar*, Leyuan Ma*, Benjamin Roscoe, Sung Mi Park, Daniel N. Bolon and Michael R. Green
Pigment Cell Melanoma Res. 27; 124–133 (2013)

This work was a collaboration between Daniel N Bolon, Ph.D., and Michael R. Green, Ph.D.'s laboratories. It is a novel example of using systematic mutagenesis and deep sequencing to generate inhibitor-resistant mutants in a human oncogene. My personal contributions included the processing of library DNA for deep sequencing and the analysis of deep sequencing data to determine resistant mutants in high-throughput.

Timothy R. Wagenaar, Leyuan Ma, Benjamin Roscoe, Daniel Bolon and Michael R. Green designed the experiments. Daniel N. Bolon constructed the bRAF saturating point mutant libraries. Timothy R. Wagenaar, Leyuan Ma and Sung Mi Park performed all of the *in vivo* mouse and *in vitro* cellular assays. Timothy Waagenar extracted and amplified bRAF DNA from cells after drug selection, and Benjamin Roscoe processed amplified bRAF library DNA for directed deep sequencing. Benjamin Roscoe and Daniel Bolon analyzed the deep sequencing data and identified enriched mutants. Timothy R. Waagenar, Leyuan Ma, Daniel N. Bolon and Michael R. Green prepared the manuscript.

Summary

Resistance to the BRAF inhibitor vemurafenib poses a significant problem for the treatment of BRAFV600E-positive melanomas. It is therefore critical to prospectively identify all vemurafenib resistance mechanisms prior to their emergence in the clinic. The vemurafenib resistance mechanisms described to date do not result from secondary mutations within BRAFV600E. To search for possible mutations within BRAFV600E that can confer drug resistance, we developed a systematic experimental approach involving targeted saturation mutagenesis, selection of drug-resistant variants, and deep sequencing. We identified a single nucleotide substitution (T1514A, encoding L505H) that greatly increased drug resistance in cultured cells and mouse xenografts. The kinase activity of BRAFV600E/L505H was higher than that of BRAFV600E, resulting in cross-resistance to a MEK inhibitor. However, BRAFV600E/L505H was less resistant to several other BRAF inhibitors whose binding sites were further from L505 than that of PLX4720. Our results identify a novel vemurafenib-resistant mutant and provide insights into the treatment for melanomas bearing this mutation.

Significance

The oncogenic BRAF kinase mutation BRAFV600E is found at high frequency in melanomas. Vemurafenib, a BRAF kinase inhibitor, has shown remarkable clinical efficacy in the treatment for BRAFV600E-positive melanoma. Inevitably, however, resistance emerges. Therefore, prospectively identifying possible vemurafenib resistance mechanisms is critical for developing more effective therapeutic approaches. Toward this end, we developed a systematic saturation mutagenesis approach to search for second-site

mutations within BRAFV600E that could confer drug resistance. Using this method, we identify and characterize a novel vemurafenib-resistant BRAFV600E mutant, which arises by a single nucleotide substitution, and provide insights into the potential treatment of melanomas bearing this mutation.

Introduction

BRAF is a serine–threonine kinase that functions as an immediate downstream effector of RAS (reviewed in Dhomen and Marais, 2007).[151] BRAF activates the MAP kinase extracellular signal-regulated kinase (MEK), which in turn phosphorylates and activates extracellular signal-regulated kinases 1 and 2 (ERK1 and ERK2). Oncogenic BRAF mutations are found in a significant number of human cancers, with a particularly high frequency (50–70%) occurring in melanomas.[152] The most frequent oncogenic mutation occurs within the BRAF kinase domain and is the substitution of a valine for glutamic acid at amino acid 600 (V600E). The mutation leads to unchecked kinase activity and constitutive activation of the downstream MEK and ERK kinases.

Vemurafenib (also called PLX4032) is a selective inhibitor of BRAFV600E that can elicit marked melanoma tumor regression, resulting in improved progression-free and overall survival in patients with metastatic disease.[153,154] However, the durability of the vemurafenib response is limited by acquired drug resistance.[155] Thus, elucidating the basis of resistance to vemurafenib, and other BRAF inhibitors, is essential to developing more effective therapies for the treatment of melanoma.

A common mechanism of resistance to small molecule protein kinase inhibitors is the acquisition of a second-site mutation that interferes with drug binding.[43] Such drug-

resistant variants, isolated from patients or in cell- or animal-based experiments, typically arise from a single, non-synonymous nucleotide mutation within the protein kinase domain. Several vemurafenib resistance mechanisms have been described and, in most cases, are due to alternative activation of MEK-ERK signaling .[155] Other vemurafenib resistance mechanisms include amplification of the BRAFT1799A allele (encoding BRAFV600E)[156] and generation of aberrantly spliced BRAFV600E variants.[157] Surprisingly, however, drug-resistant amino acid substitution mutants within the BRAFV600E protein-coding region have not been isolated from vemurafenib-resistant melanomas or melanoma cell lines, suggesting that they are either impossible or improbable relative to other resistant pathways. Here, we describe a systematic, structure-based saturation mutagenesis approach to identify single, second-site nucleotide substitutions within BRAFV600E that can confer vemurafenib resistance.

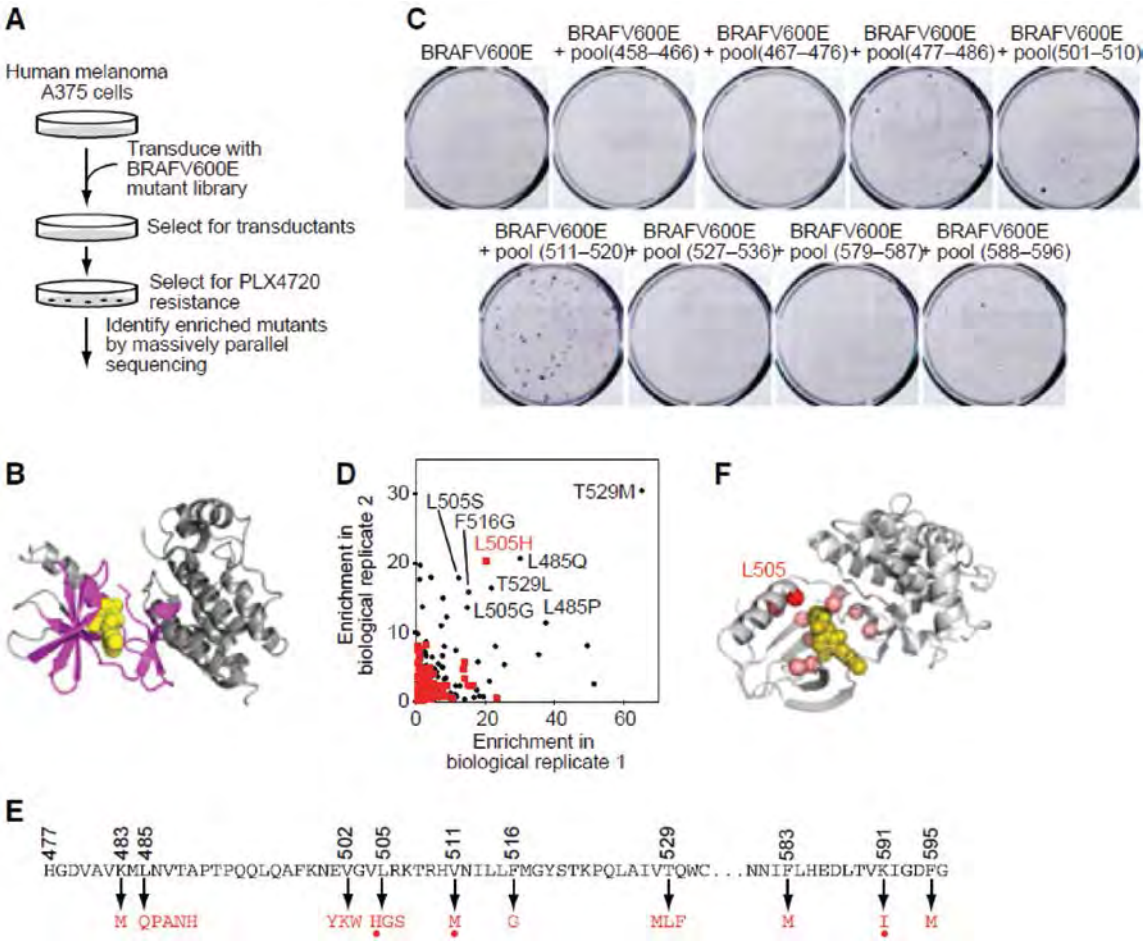
Results

A structure-based, targeted, saturation mutagenesis screen identifies PLX4720-resistant BRAFV600E mutants

Our experimental strategy is summarized in Figure 4.1(A) and discussed below. Guided by the structure of BRAFV600E bound to PLX4720 (PDB: 3C4C and Figure 4.1B), a tool compound for vemurafenib that elicits comparable actions to its clinical-grade counterpart[158], we performed targeted saturation mutagenesis of 77 amino acids surrounding the PLX4720-binding site. Eight mutant pools were generated—corresponding to amino acids 458–466, 467–476, 477–486, 501–510, 511–520, 527–536, 579–587, and 588–596—in which each amino acid was mutated to all possible 64 codons. The mutant pools were transferred into a retroviral vector containing BRAFV600E and then stably transduced into the human melanoma cell line A375, which is homozygous for BRAFV600E and highly sensitive to PLX4720.[159] Cells were cultured for 3 weeks in the presence of 10 μM PLX4720, at which point resistant clones emerged for six of the eight pools (Figure 4.1C). Pools encoding mutants corresponding to amino acids 458–466 and 467–476 did not result in any detectable PLX4720-resistant colonies above background.

Figure 4.1. A structure-based, targeted, saturation mutagenesis screen identifies PLX4720-resistant BRAFV600E mutants. (A) Schematic summary of the PLX4720 resistance screen. (B) Structure of BRAF complexed with PLX4720 (yellow); amino acids surrounding the drug-binding site are colored magenta. (C) Images of A375 cells transduced with retroviruses expressing BRAFV600E or pools of mutagenized BRAFV600E and grown in media containing PLX4720. Cells were fixed and stained with crystal violet. (D) Scatter plot showing enriched BRAFV600E mutants after PLX4720 selection from two independent biological replicates. Mutants that occur due to single nucleotide substitution are indicated by a red dot. (E) Amino acid sequence of the mutagenized region of human BRAFV600E showing the PLX4720-resistant mutants isolated from the screen showing a >fivefold enrichment. Mutants that occur by single nucleotide substitution are indicated by a red dot. (F) Location of the amino acids whose mutants are enriched >fivefold (shown in pink, except for L505 in red) in the BRAF–PLX4720 structure.

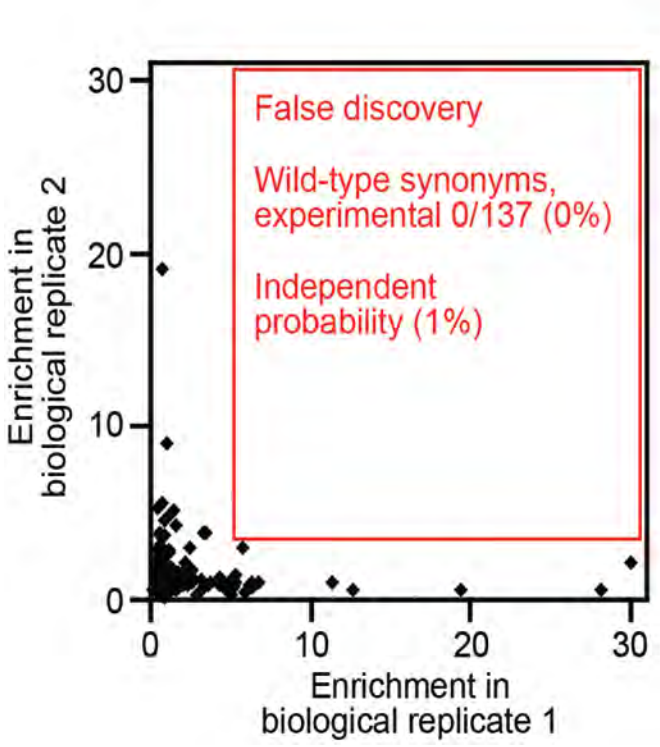
Figure 4.1. A structure-based, targeted, saturation mutagenesis screen identifies PLX4720-resistant BRAFV600E mutants.



To identify the mutated amino acid(s) that conferred PLX4720 resistance, cells from the resistant cell populations were pooled, genomic DNA was isolated, and BRAFV600E variants were identified by massively parallel sequencing. We anticipated that, following drug selection, the sequences of PLX4720-resistant BRAFV600E mutants would be enriched relative to PLX4720-sensitive BRAFV600E variants. To control for possible unequal representation of the variants in the starting population, deep sequencing was also performed on genomic DNA isolated from transduced cells prior to drug selection. Figure 4.1(D) displays the relative enrichment of BRAFV600E variants following PLX4720 selection from two independent biological replicates. The use of replicate experiments enabled us to distinguish between reproducible enrichment, resulting from a drug-resistant mutation, and non-reproducible enrichment, resulting from a random variation in the sequence pool. For example, silent substitutions that did not alter the BRAFV600E protein sequence were occasionally enriched in one replicate, but rarely in both (Figure 4.2). Mutations that were enriched in both replicates across all measured synonymous codons with a net false discovery rate of < 1% were considered statistically significant. The results, summarized in Table S1, identified 25 different amino acids encompassing 55 variants, 18 of which could arise by a single-base substitution; mutants that were enriched >fivefold are shown in Figure 4.1(E). Figure 4.1F displays the positions of the 10 amino acids, whose substitutions were enriched >fivefold, on the BRAF–PLX4720 structure.

Figure 4.2. Drug enrichment of silent mutations that did not change the parental BRAFV600E protein sequence.

Figure 4.2. Drug enrichment of silent mutations that did not change the parental BRAFV600E protein sequence.

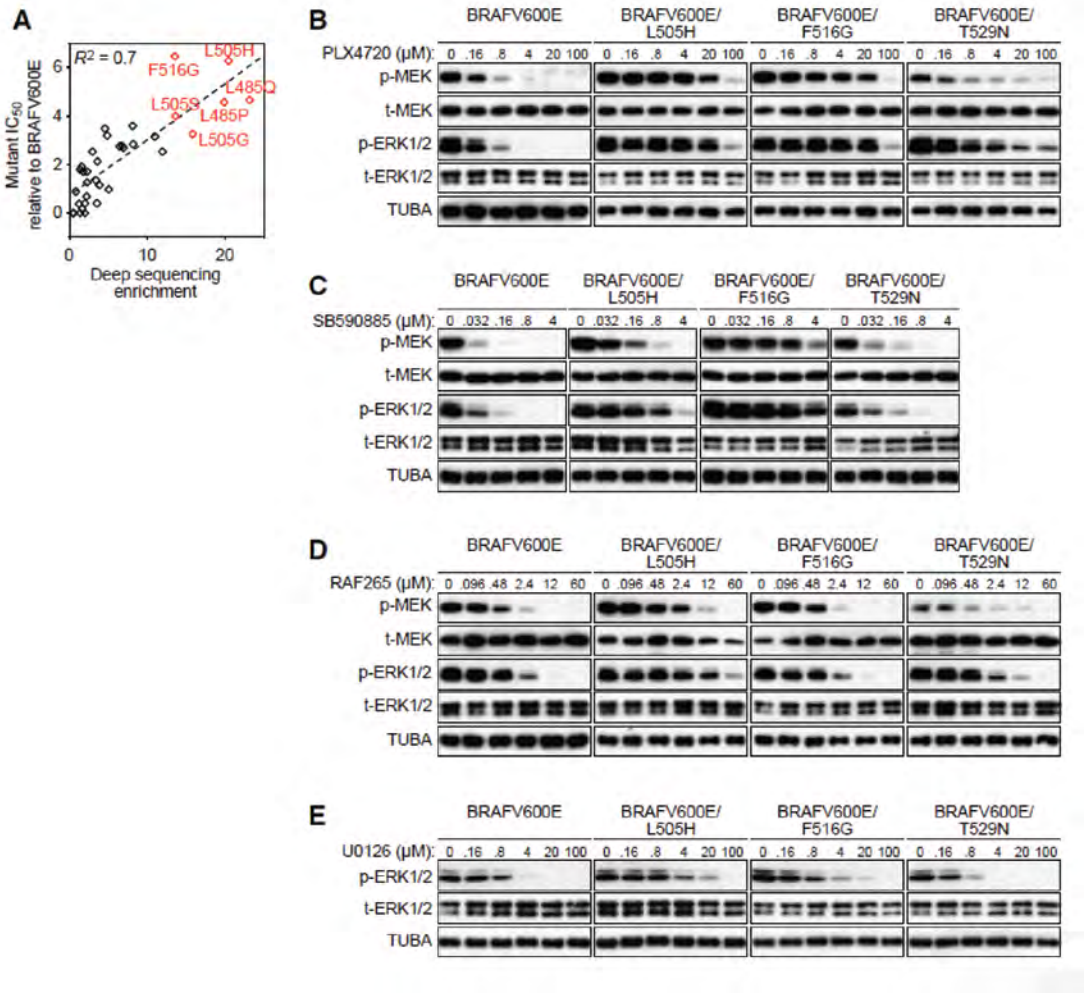


Sensitivity of the BRAFV600E/L505H mutant to BRAF and MEK inhibitors

We analyzed the 12 most enriched mutants and a subset of less enriched mutants, by stable expression in A375 cells followed by determination of the cellular PLX4720 median inhibitory concentration (IC50). Figure 4.3(A) shows that the strongest PLX4720 resistance was associated with substitutions at amino acids L485, L505, and F516. Notably, there was an excellent correlation between the normalized enrichment obtained from deep sequencing and the relative IC50.

Figure 4.3. Sensitivity of the BRAFV600E/L505H mutant to BRAF and MEK inhibitors. (A) Relative cellular IC50 of A375 cells transduced with a subset of candidate PLX4720-resistant mutants plotted against their relative deep sequencing enrichment. (B–E) Immunoblots showing levels of phospho-MEK (p-MEK), total MEK (t-MEK), phospho-ERK1/2 (p-ERK1/2), and total ERK1/2 (t-ERK1/2) in A375 cells transduced with a retrovirus expressing BRAFV600E, BRAFV600E/L505H, BRAFV600E/F516G, or BRAFV600E/T529N and treated with increasing doses of PLX4720 (B), SB590885 (C), RAF265 (D), or U0126 (E). α-Tubulin (TUBA) was monitored as a loading control.

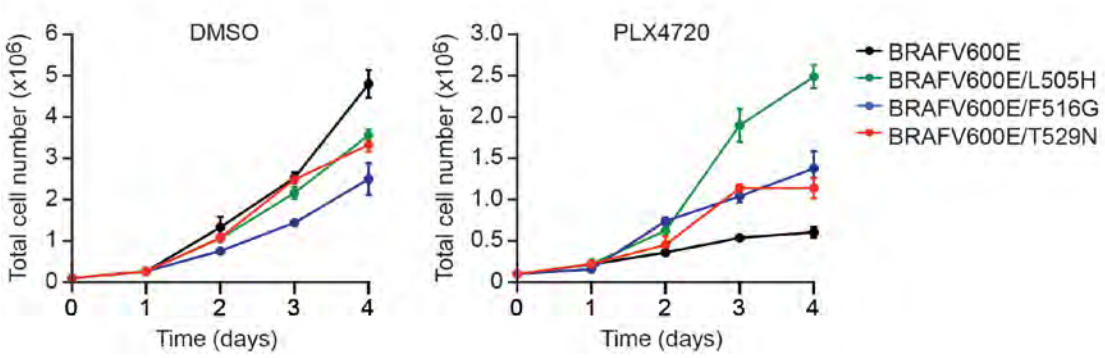
Figure 4.3. Sensitivity of the BRAFV600E/L505H mutant to BRAF and MEK inhibitors.



We elected to further characterize BRAFV600E/L505H, which was the single nucleotide substitution mutant most resistant to PLX4720 (see Supplementary Table S4.1). For comparison, we also analyzed the most PLX4720-resistant mutant identified in our screen, BRAFV600E/F516G (Figure 4.3A), which arises from multiple nucleotide substitutions, and BRAFV600E/T529N, a previously characterized mutant derived by directed mutagenesis of the gatekeeper residue.[160] As expected, the cellular proliferation assay of Figure 4.4 shows that A375 cells expressing BRAFV600E/L505H, BRAFV600E/F516G, or BRAFV600E/T529N were all more resistant to PLX4720 compared with A375 cells expressing BRAFV600E. As expected, in the absence of PLX4720, proliferation of A375 cells expressing BRAFV600E was comparable to that of A375 cells expressing BRAFV600E/L505H, BRAFV600E/F516G, or BRAFV600E/T529N (Figure 4.4).

Figure 4.4. PLX4720 resistance of A375 cell lines expressing BRAFV600E, BRAFV600E/L505H, BRAFV600E/F516G or BRAFV600E/T529N.

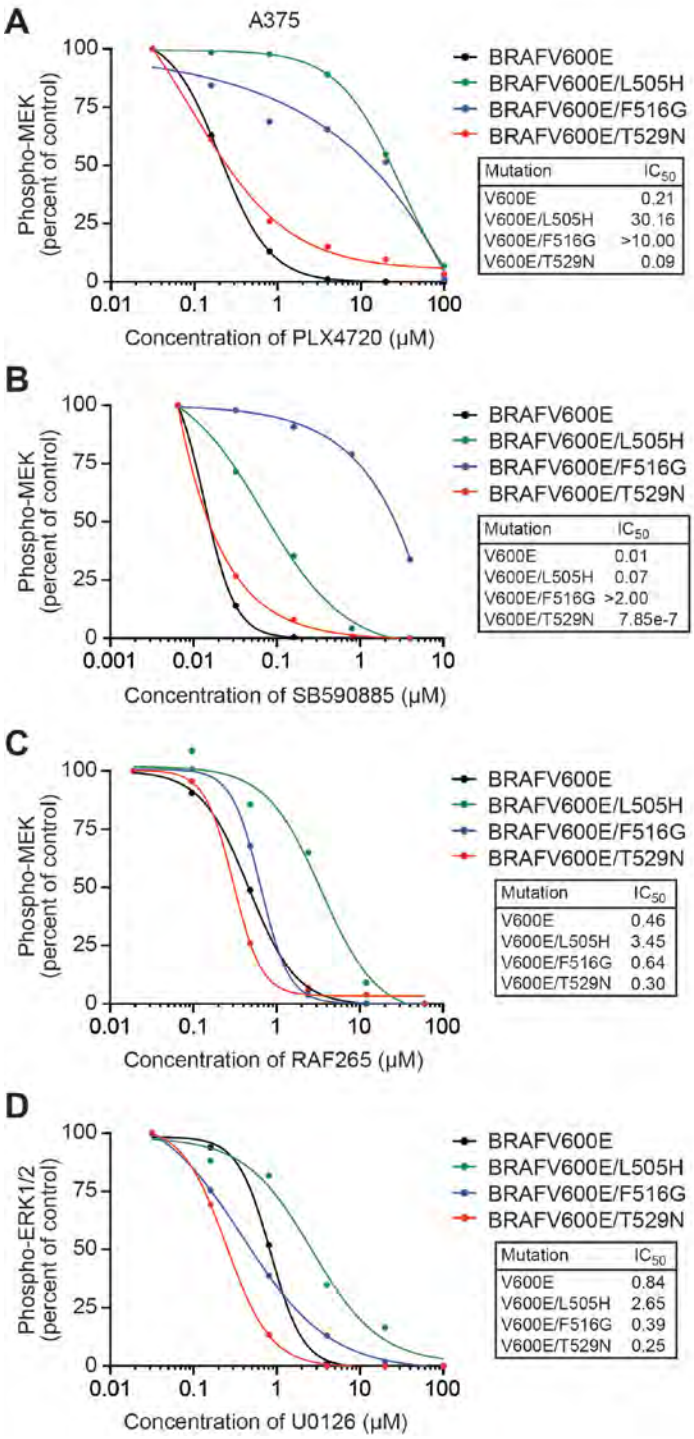
Figure 4.4. PLX4720 resistance of A375 cell lines expressing BRAFV600E, BRAFV600E/L505H, BRAFV600E/F516G or BRAFV600E/T529N.



We next confirmed the relative sensitivity of the various BRAFV600E mutants to PLX4720 by monitoring phosphorylation of the downstream signaling components, MEK and ERK1/2. Figure 4.3(B) shows, as expected, that in A375 cells, PLX4720 potently inhibited BRAFV600E with marked reduction in phosphorylated MEK and ERK1/2 (phospho-MEK and phospho-ERK1/2, respectively) by 0.8 μM PLX4720. By contrast, BRAFV600E/L505H and BRAFV600E/F516G required an approximately 100–150-fold higher concentration of PLX4720 to obtain similar reduction in phospho-MEK and phospho-ERK1/2 (Figure 4.3B and Figure 4.5A). Notably, the PLX4720 sensitivity of BRAFV600E/T529N was comparable to that of BRAFV600E, which likely explains why this mutant was not isolated in our screen. By comparison with the results with PLX4720, the BRAFV600E/L505H mutant had only a modest (5–7-fold) effect on resistance to two other BRAF inhibitors: SB590885 (Takle et al., 2006) and RAF265 (Amiri et al., 2006) (Figure 4.3C, D and Figure 4.5B, C).

Figure 4.5. Phospho-MEK or phospho-ERK1/2 IC50 curves for the immunoblots shown in Figure **2**B–E.

Figure 4.5. Phospho-MEK or phospho-ERK1/2 IC50 curves for the immunoblots shown in Figure **2**B–E.



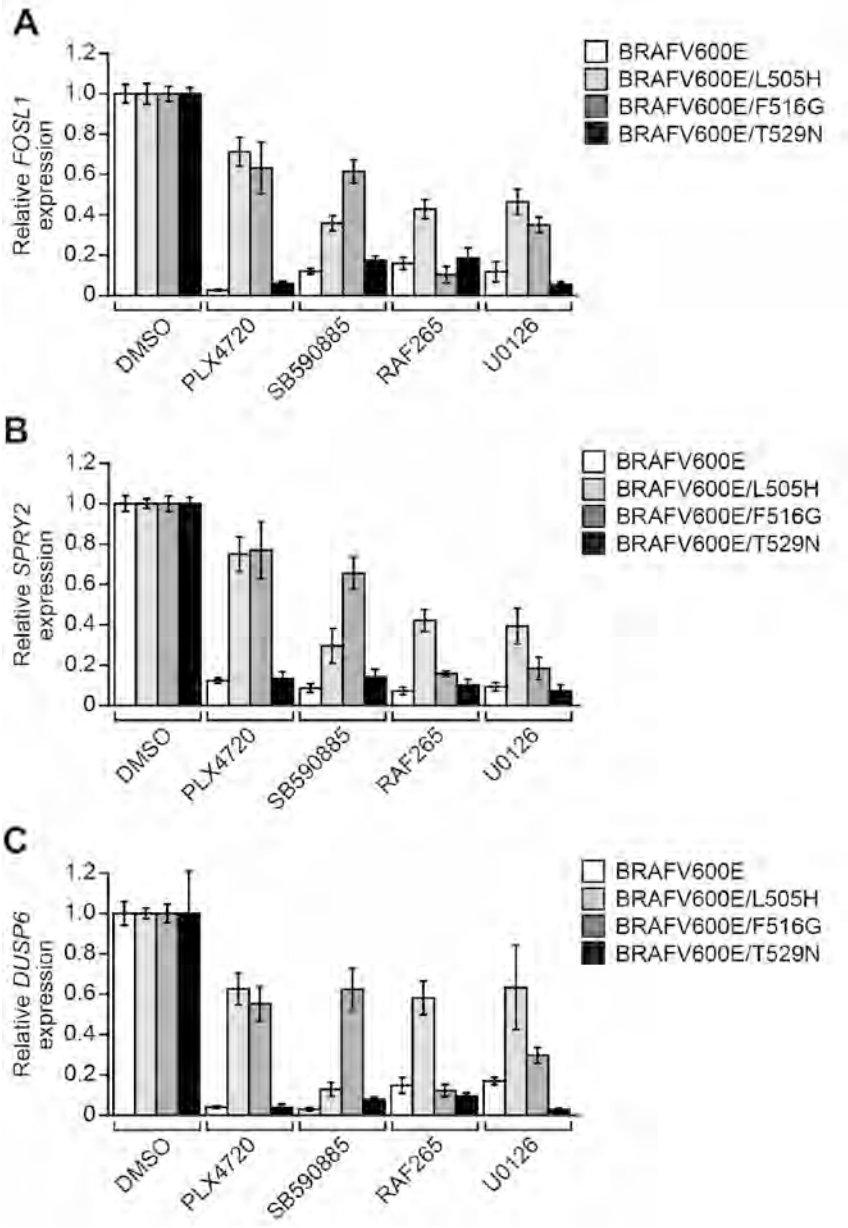
A possible approach for treatment of vemurafenib-resistant melanomas is the use of a MEK inhibitor.[155,161] We therefore analyzed the sensitivity of BRAFV600E/L505H to the MEK inhibitor U0126.[162] Notably, BRAFV600E/L505H exhibited increased resistance to U0126 compared with BRAFV600E, as evidenced by elevated phospho-ERK levels (Figure 4.3E and Figure S4.5D). By contrast, relative to BRAFV600E, the BRAFV600E/F516G mutant was comparably sensitive and the BRAFV600E/T529N mutant was actually more sensitive to U0126 (Figure 4.3E and Figure 4.5D).

To determine whether the differences in MEK/ERK signaling were biologically relevant, we performed two additional experiments. First, we monitored expression of three representative ERK target genes (FOSL1, SPRY2, and DUSP6), a biologically relevant output of the BRAF-MEK-ERK signaling pathway.[163] Gene expression was measured by quantitative RT-PCR (qRT-PCR) following treatment of A375 cells with a drug concentration that differentially affected the various BRAFV600E mutants. Consistent with the results of Figure 4.3(B), at a PLX4720 concentration of 20 μM, expression of FOSL1, SPRY2, and DUSP6 was greatly reduced in cells expressing BRAFV600E or BRAFV600E/T529N compared with cells expressing BRAFV600E/L505H or BRAFV600E/F516G (Figure 4.6). Similarly, treatment with SB590885 (0.8 μM) or U0126 (4 μM) reduced FOSL1, SPRY2, and DUSP6 expression to a greater extent in cells expressing BRAFV600E or BRAFV600E/T529N relative to cells expressing BRAFV600E/L505H or BRAFV600E/F516G. Finally, treatment with RAF265 (2.4 μM) reduced ERK target gene expression to a greater extent in cells

expressing BRAFV600E, BRAFV600E/F516G, or BRAFV600E/T529N relative to cells expressing BRAFV600E/L505H.

Figure 4.6. Analysis of ERK target gene expression following drug treatment in A375 cell lines expressing BRAFV600E, BRAFV600E/L505H, BRAFV600E/F516G or BRAFV600E/T529N.

Figure 4.6. Analysis of ERK target gene expression following drug treatment in A375 cell lines expressing BRAFV600E, BRAFV600E/L505H, BRAFV600E/F516G or BRAFV600E/T529N.

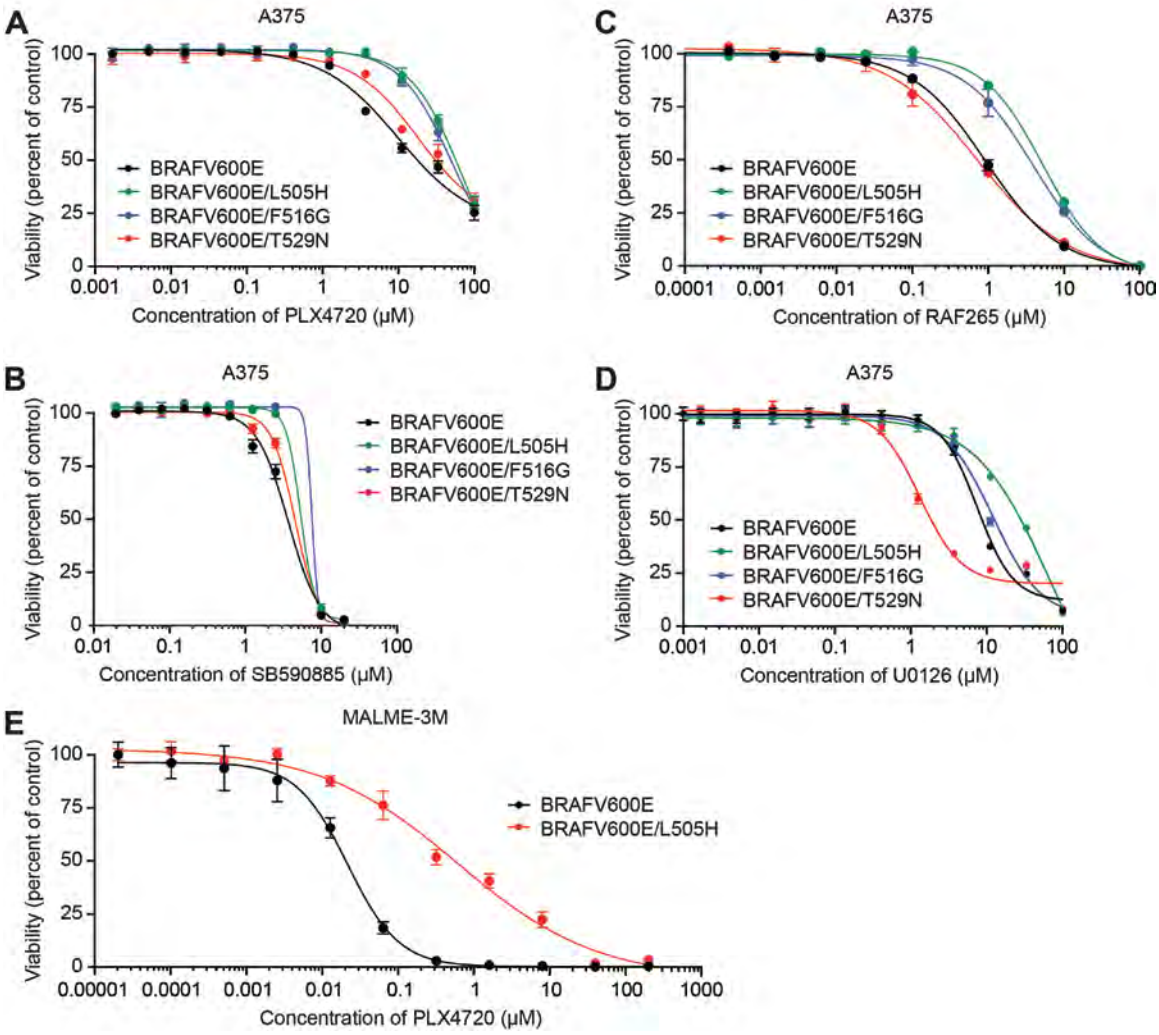


In a second set of experiments, we measured the relative drug resistance of A375 cells expressing the various BRAFV600E mutants. Figure 4.7A–D shows that cells expressing either BRAFV600E/L505H or BRAFV600E/F516G were relatively more resistant to PLX4720, SB590885, RAF265, and U0126 than cells expressing BRAFV600E or BRAFV600E/T529N. Collectively, these results indicate that the differences in MEK-ERK signaling (Figure 4.3B) correlated well with both ERK target gene expression (Figure S4) and relative drug resistance (Figure 4.7A–D) of cells expressing the mutants.

Finally, we also confirmed the PLX4720 resistance of the BRAFV600E/L505H mutant in an additional BRAFV600E-positive human melanoma cell line, MALME-3M. The results show that MALME-3M cells expressing BRAFV600E/L505H were substantially more resistant to PLX4720 than cells expressing BRAFV600E (Figure 4.7E).

Figure 4.7. Relative drug resistance of BRAFV600E mutants in A375 cells, and confirmation of PLX4720 resistance of the BRAFV600E/L505H mutant in an additional BRAFV600E-positive human melanoma cell line.

Figure 4.7. Relative drug resistance of BRAFV600E mutants in A375 cells, and confirmation of PLX4720 resistance of the BRAFV600E/L505H mutant in an additional BRAFV600E-positive human melanoma cell line.



Characterization of the BRAFV600E/L505H mutant in 293T cells and Ba/F3 cells

As described above, the initial characterization of BRAFV600E/L505H was performed in the A375 cell line. However, we found that A375 cells transduced with BRAFV600E were approximately six-fold more resistant to PLX4720 compared with parental A375 cells (Figure 4.8). Consistent with our results, previous reports have shown that BRAFV600E amplification leads to vemurafenib resistance.[156] We therefore considered that the elevated levels of endogenous BRAFV600E in A375 cells might confound an accurate determination of the resistance conferred by PLX4720-resistant alleles, and elected to analyze the BRAFV600E/L505H mutant in two other cell lines that lacked BRAFV600E.

First, we transiently expressed BRAFV600E/L505H in human embryonic kidney 293T cells, which contain wild-type BRAF and have relatively low levels of phospho-MEK and phospho-ERK1/2. As expected, the expression of BRAFV600E resulted in the activation of MEK-ERK signaling, as evidenced by increased levels of phospho-MEK and phospho-ERK1/2 (Figure 4.9A). Interestingly, 293T cells expressing BRAFV600E/L505H had substantially higher levels of phospho-MEK and phospho-ERK1/2 compared with 293T cells expressing BRAFV600E, despite similar levels of BRAF protein, indicating that the L505H substitution increases BRAFV600E kinase activity. Even in the absence of the V600E mutation, the L505H substitution (BRAFL505H) led to elevated levels of phospho-MEK (Figure 4.9A).

Figure 4.8 Increased PLX4720 resistance of A375 cells expressing BRAFV600E.

Figure 4.8. Increased PLX4720 resistance of A375 cells expressing BRAFV600E.

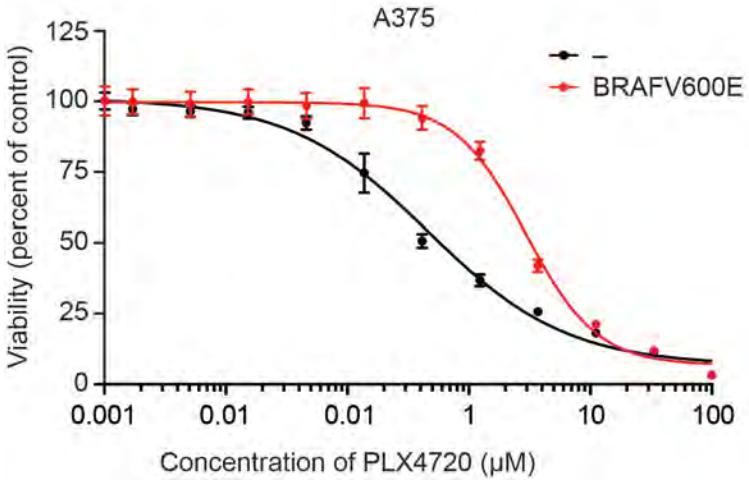
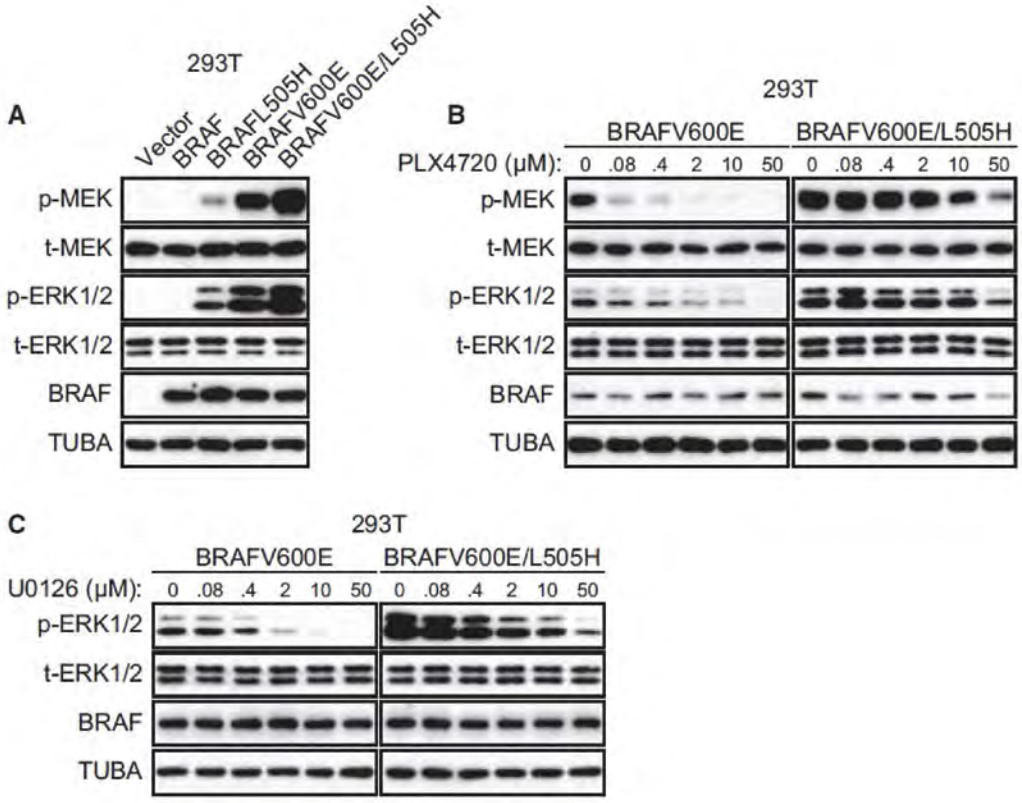


Figure 4.9. Characterization of the BRAFV600E/L505H mutant in 293T cells. (A) Immunoblots showing levels of p- and t-MEK, p- and t-ERK1/2, and myc-tagged BRAF in 293T cells transfected with empty vector, wild-type BRAF, BRAFL505H, BRAFV600E, or BRAFV600E/L505H. (B, C) Immunoblots showing levels of p- and t-MEK, p- and t-ERK1/2, and myc-tagged BRAF in 293T cells transiently transfected with BRAFV600E or BRAFV600E/L505H and treated with PLX4720 (B) or U0126 (C).

Figure 4.9. Characterization of the BRAFV600E/L505H mutant in 293T cells.



Treatment of 293T cells expressing BRAFV600E with PLX4720 resulted in dose-dependent inhibition of MEK phosphorylation, with phospho-MEK levels being nearly undetectable by 2 μM PLX4720 (Figure 4.9B). By comparison, 293T cells expressing BRAFV600E/L505H displayed persistent phospho-MEK levels even at a PLX4720 concentration of 50 μM (see also Figure S4.10A). Finally, consistent with the results in A375 cells, BRAFV600E/L505H was substantially more resistant to U0126 compared with BRAFV600E (Figure 4.9C and Figure 4.10B).

Previous studies have shown that stable expression of BRAFV600E renders Ba/F3 cells, a BRAF wild-type, interleukin-3 (IL-3)-dependent pro-B cell line, dependent on BRAF-MEK-ERK signaling following IL-3 deprivation.[160] To examine the effects of PLX4720 on cellular proliferation, we stably expressed BRAFV600E and BRAFV600E/L505H in Ba/F3 cells, to generate Ba/F3-BRAFV600E and Ba/F3-BRAFV600E/L505H cells, respectively. As expected, expression of either BRAFV600E or BRAFV600E/L505H led to robust activation of MEK-ERK signaling (Figure 4.11A). Interestingly, total levels of BRAF were reduced in Ba/F3-BRAFV600E/L505H cells compared with Ba/F3-BRAFV600E cells, perhaps due to cytotoxicity of the BRAFV600E/L505H mutant. Nonetheless, MEK-ERK signaling was comparable or higher in Ba/F3-BRAFV600E/L505H cells compared with Ba/F3-BRAFV600E cells, again indicating that the BRAFV600E/L505H mutant has elevated kinase activity. As in 293T cells, even in the absence of the V600E mutation, the L505H substitution (BRAFL505H) led to elevated levels of phospho-MEK (Figure 4.11B). Treatment of Ba/F3-BRAFV600E cells with PLX4720 resulted in a dose-dependent reduction in

phospho-MEK and phospho-ERK1/2 levels (Figure 4.11C). Notably, Ba/F3-BRAFV600E/L505H cells required a ~20- to 25-fold higher concentration of PLX4720 to achieve similar reduction in phospho-MEK and phospho-ERK1/2 compared with Ba/F3-BRAFV600E cells (Figure 4.11C and Figure 4.12).

Figure 4.10. Phospho-MEK or -ERK1/2 IC50 curves for the immunoblots shown in Figure 4.9B, C.

Figure 4.10. Phospho-MEK or -ERK1/2 IC50 curves for the immunoblots shown in Figure 4.9B, C.

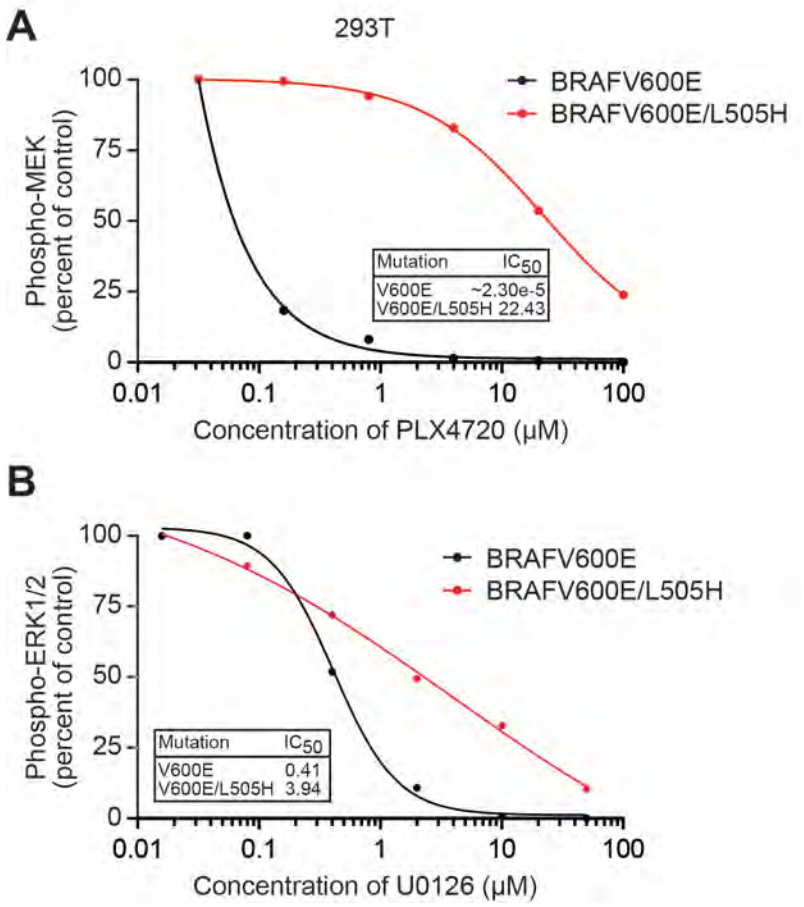


Figure 4.11. Characterization of the BRAFV600E/L505H mutant in Ba/F3 cells. (A) Immunoblots showing levels of p- and t-MEK, p- and t-ERK1/2, and myc-tagged BRAF in parental Ba/F3 cells [(−); cultured with IL-3] or Ba/F3 cells stably expressing BRAFV600E or BRAFV600E/L505H (cultured without IL-3). (B) Immunoblots showing levels of p- and t-MEK, p- and t-ERK1/2, and myc-tagged BRAF in Ba/F3 stably expressing empty vector, wild-type BRAF or BRAFL505H (cultured with IL-3). (C) Immunoblots showing levels of p- and t-MEK, p- and t-ERK1/2, and myc-tagged BRAF in Ba/F3 cells cultured without IL-3, stably expressing BRAFV600E or BRAFV600E/L505H and treated with PLX4720. (D) Growth of parental Ba/F3 cells (−) or Ba/F3 cells stably expressing BRAFV600E or BRAFL505H/V600E and cultured with (+) or without (−) IL-3. (E, F) Cellular IC50 of parental Ba/F3 cells [(−); cultured with IL-3] and Ba/F3 cells stably transduced with BRAFV600E or BRAFV600E/L505H (cultured without IL-3) and treated with PLX4720 (E) or vemurafenib (F). Error bars indicate SEM.

Figure 4.11. Characterization of the BRAFV600E/L505H mutant in Ba/F3 cells.

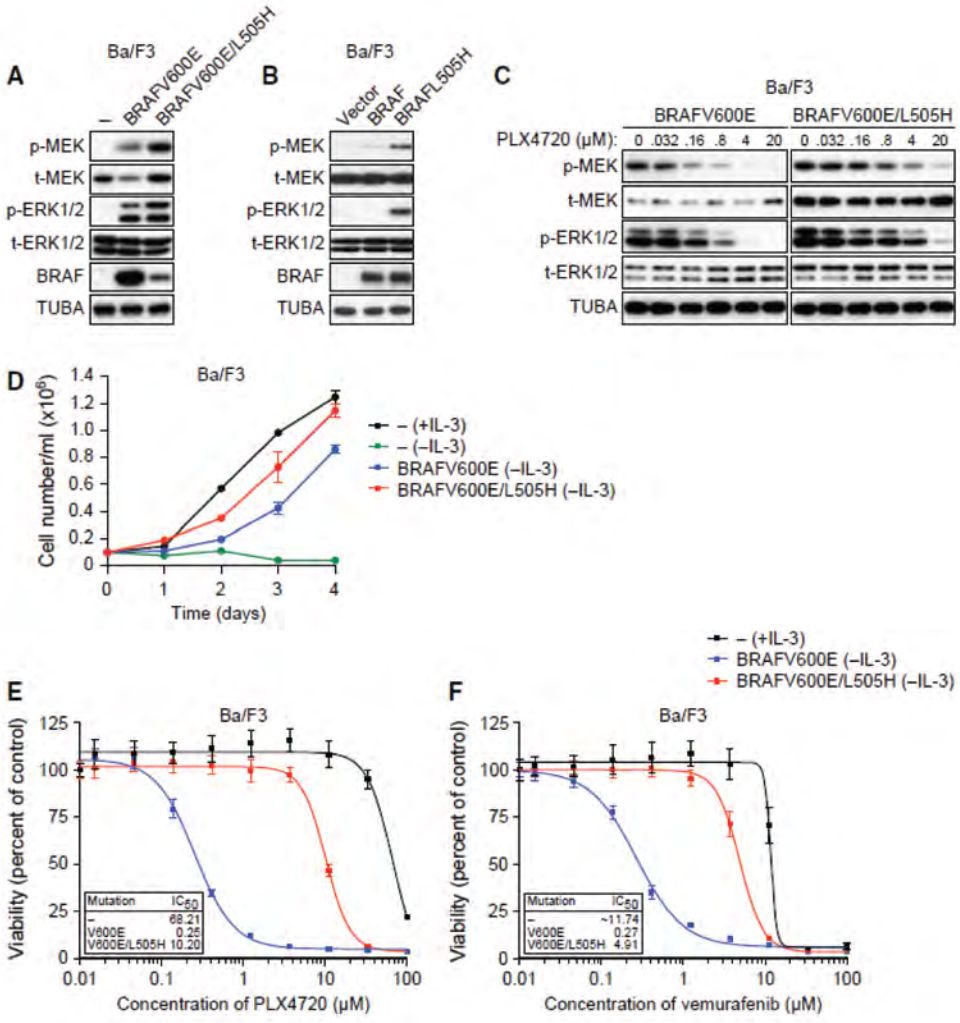
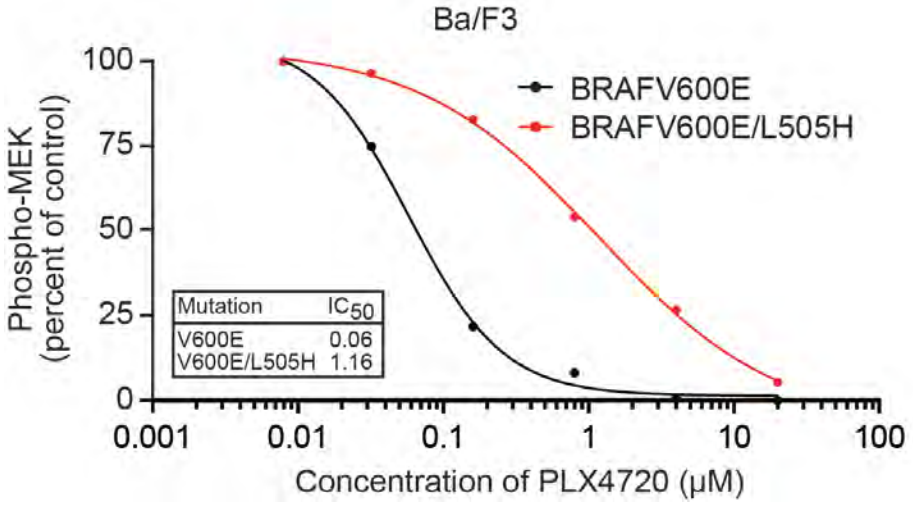


Figure 4.12. Phospho-MEK IC50 curve for the immunoblots shown in Figure 4.11(C).

Figure 4.12. Phospho-MEK IC50 curve for the immunoblots shown in Figure 4.4(C).



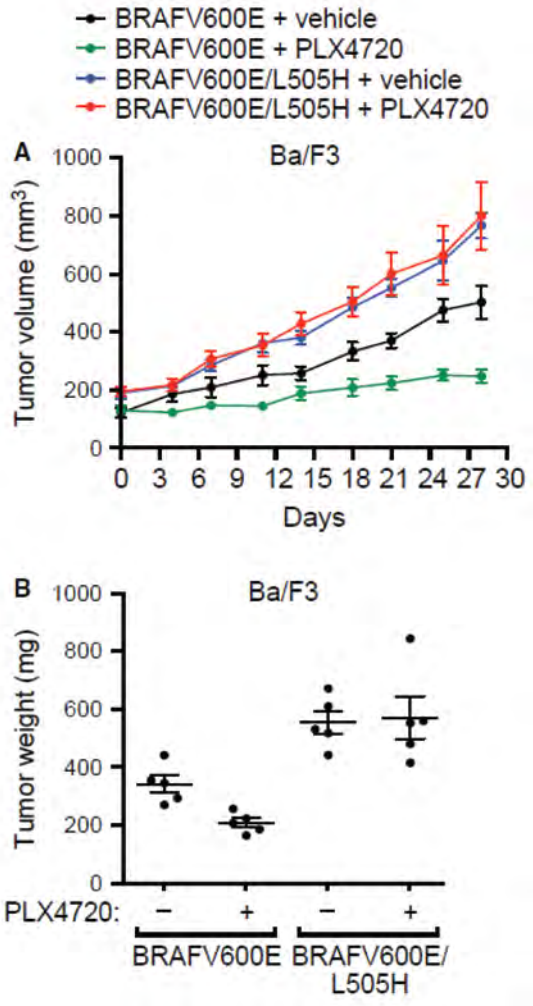
Consistent with previous studies[160], stable expression of BRAFV600E or BRAFV600E/L505H in Ba/F3 cells conferred IL-3-independent proliferation (Figure 4.11D), enabling determination of the effect of PLX4720 on cellular proliferation. Notably, the cellular PLX4720 IC50 was 40-fold higher for Ba/F3-BRAFV600E/L505H cells compared with Ba/F3-BRAFV600E cells (Figure 4.11E). As expected, similar resistance was observed when vemurafenib was used instead of PLX4720 (Figure 4.11F).

Sensitivity of the BRAFV600E/L505H mutant to PLX4720 in mouse xenografts and to other BRAF inhibitors in cell culture

To determine whether the BRAFV600E/L505H mutant also conferred resistance to PLX4720 in vivo, we performed mouse xenograft experiments. Ba/F3-BRAFV600E/L505H cells or Ba/F3-BRAFV600E cells were injected into the flanks of immunocompromised mice, and PLX4720 or control vehicle was administered intraperitoneally. PLX4720 markedly reduced growth of Ba/F3-BRAFV600E tumors, as expected, whereas Ba/F3-BRAFV600E/L505H tumors were resistant to PLX4720 (Figure 4.13A, B). Notably, even in the absence of PLX4720, the growth of Ba/F3-BRAFV600E/L505H tumors was faster than that of Ba/F3-BRAFV600E tumors (Figure 4.13), which is likely due to the elevated kinase activity of BRAFV600E/L505H.

Figure 4.13. Sensitivity of the BRAFV600E/L505H mutant to PLX4720 in mouse xenografts and to other BRAF inhibitors in cell culture. (A) Tumor growth in mice subcutaneously injected with Ba/F3-BRAFV600E or Ba/F3-BRAFV600E/L505H cells and then intraperitoneally injected daily with either vehicle or PLX4720. For BRAFV600E ± PLX4720, $P < 0.05$; for BRAFV600E/L505H ± PLX4720, $P > 0.05$; for BRAFV600E versus BRAFV600E/L505H (vehicle), $P < 0.05$ at day 27. (B) Tumor weight after the final injection of either vehicle or PLX4720. Results are plotted with group mean, and SEM indicated by the horizontal and vertical lines, respectively. For BRAFV600E ± PLX4720, $P < 0.05$; for BRAFV600E/L505H ± PLX4720, $P > 0.05$; for BRAFV600E versus BRAFV600E/L505H (vehicle), $P < 0.05$.

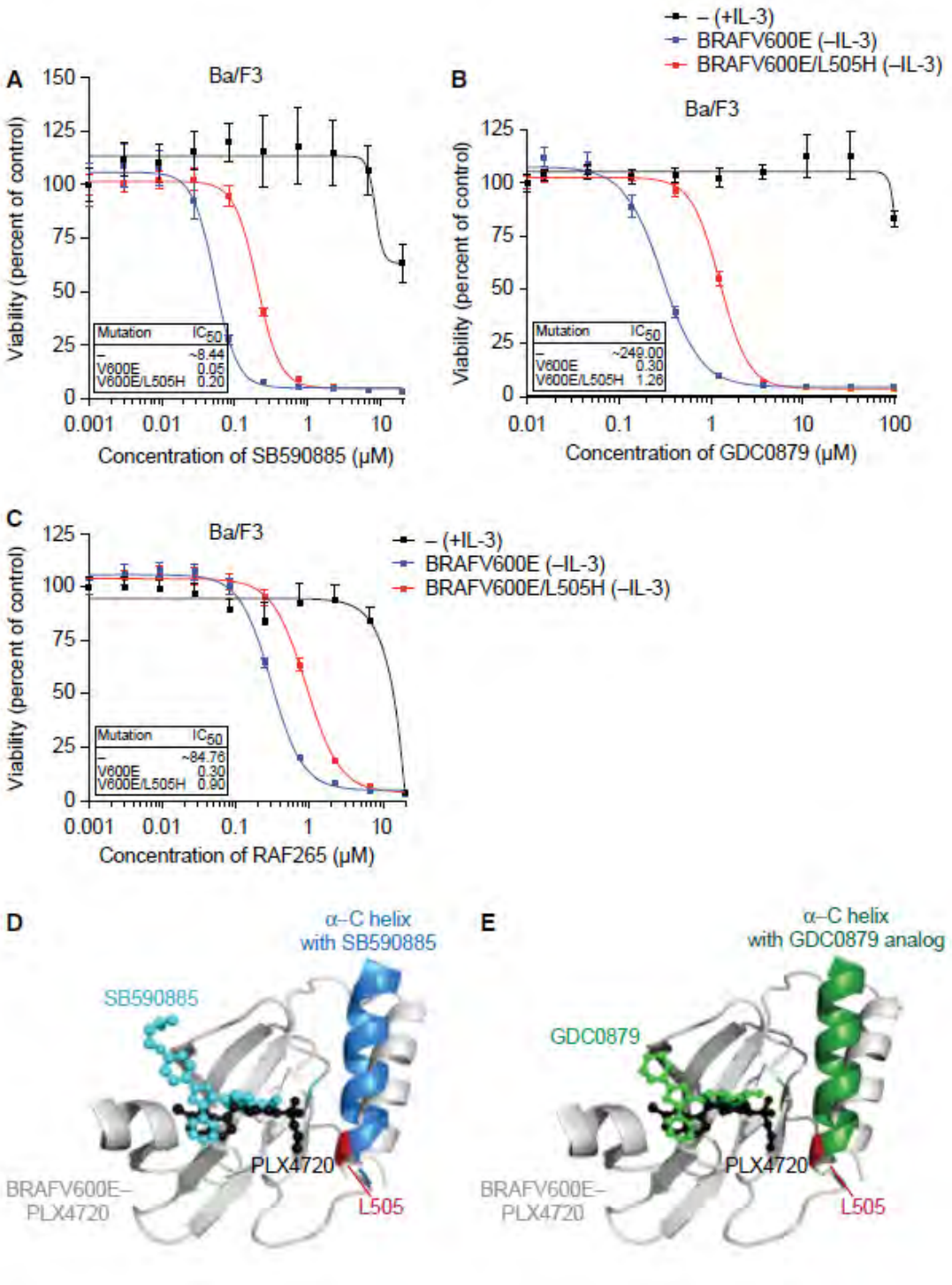
Figure 4.13. Sensitivity of the BRAFV600E/L505H mutant to PLX4720 in mouse xenografts and to other BRAF inhibitors in cell culture.



Finally, the BRAFV600E/L505H mutant conferred a much lower level of resistance to several other BRAF inhibitors in Ba/F3 cells. For example, the cellular IC50s for SB590885 and GDC0879 were only fourfold higher (Figure 4.14A, B) and the cellular IC50 for RAF265 was only threefold higher (Figure 4.14C) in Ba/F3-BRAFV600E/L505H cells compared with Ba/F3-BRAFV600E cells.

Figure 4.14. Sensitivity of the BRAFV600E/L505H mutant to other BRAF inhibitors can be explained by differences in steric clash imposed by the L505H substitution. (A–C) Cellular IC50 of parental Ba/F3 cells [(−); cultured with IL-3] and Ba/F3 cells stably transduced with BRAFV600E or BRAFV600E/L505H (cultured without IL-3) and treated with SB590885 (A), GDC0879 (B), or RAF265 (C). Error bars indicate SEM. (D, E) Position of the α-C helix of BRAFV600E bound to SB590885 (blue; D) or the GDC0879 analog (green; E) superimposed on the structure of the BRAFV600E kinase domain (gray) bound to PLX4720 (black). The position of the L505 residue is shown in red.

Figure 4.14. Sensitivity of the BRAFV600E/L505H mutant to other BRAF inhibitors can be explained by differences in steric clash imposed by the L505H substitution.



Discussion

The comprehensive identification of drug-resistant mutations in cancers is an important step toward the development of improved therapeutic strategies. Here, we describe a systematic, structure-based experimental strategy to identify drug-resistant mutants within the oncogenic protein kinase BRAFV600E to its inhibitor vemurafenib. This method can be broadly applied for evaluating resistance to inhibitors of other proteins and has several important advantages over previous experimental approaches. Traditional drug resistance studies generate mutants by either (i) random methods (e.g., error prone PCR) that produce undesired mutants with multiple amino acid substitutions and do not necessarily generate all relevant single mutants (because of sampling limitations and biases in random mutagenesis), or (ii) directed mutagenesis based on results from related targets (e.g., gatekeeper mutations in protein kinases) that do not broadly sample sequence space[164]. Using systematic, structure-based saturation mutagenesis, we experimentally generate all relevant mutants without complications from undesired multiple amino acid substitutions. Furthermore, coupling the mutagenesis with a comprehensive deep sequencing readout directly quantifies the enrichment of each mutant in response to drug treatment. Thus, our systematic approach not only broadly samples sequence space but also quantifies the contribution of each mutant to drug resistance.

Using this method, we have shown that vemurafenib resistance can occur by a novel second-site mutation, T1514A (encoding L505H), within the BRAFV600E kinase domain. A series of cell proliferation and biochemical assays were used to demonstrate

the PLX4720 resistance of the BRAFV600E/L505H mutant in three human melanoma cell lines, human 293T cells, and mouse Ba/F3 cells. The similarity of the results obtained in diverse cell lines indicates that the PLX4720 resistance of the BRAFV600E/L505H mutant, which is the major focus of our study, is independent of cell type. The L505H mutant increased the kinase activity of both wild-type BRAF and BRAFV600E. Interestingly, a search of the Catalogue of Somatic Mutations in Cancer (COSMIC) database[165] revealed that the T1514A mutation was recently found in an individual with prostate cancer,[166] suggesting that the elevated kinase activity of BRAFL505H may contribute to cancer development.

We found that BRAFV600E/L505H was relatively resistant to MEK inhibition, which is likely due its increased kinase activity. By contrast, the BRAFV600E/L505H mutant had a lesser effect on sensitivity to several other BRAF inhibitors, such as SB590885, RAF265, and GDC0879. Inspection of the crystal structure of the BRAFV600E kinase domain bound to PLX4720 reveals that the propyl group of the sulfonamide moiety of the drug extends toward L505 in the BRAF α-C helix (Figure 4.1E and Figure 4.14D, E). By contrast, the crystal structure of BRAFV600E bound to SB590885 (PDB: 2FB8)[167] (Figure 4.14D) or an analog of GDC0879 (PDB: 3D4Q)[168] (Figure 4.14E) reveals that neither drug is proximal to L505. Thus, resistance of the BRAFV600E/L505H mutant to PLX4720, but not SB590885 or GDC0879, can be explained by differences in steric clash imposed by the L505H substitution. Collectively, these results indicate that such BRAFV600E/L505H-containing melanomas will be more responsive to other BRAF inhibitors than to a MEK inhibitor.

It is important to prospectively identify the various mechanisms that can result in vemurafenib resistance prior to their emergence in the clinic. This information is essential both to accurately assess the growing population of treated patients that will become drug-resistant and to develop better drugs or drug combinations to overcome diverse resistance mechanisms. Based upon our results, we predict that the L505H mutant will be found in some vemurafenib-resistant melanomas particularly if the BRAFV600E-independent resistance pathways can be inhibited. Consistent with this prediction, Tian Xu and colleagues have recently described the isolation of the BRAFV600E/L505H mutant in a melanoma cell line derived from an individual with vemurafenib-resistant melanoma (Choi J., Landrette S., Wang T., Evans P., Bacchiocchi A., Bjornson R., Cheng E., Stiegler A.L., Gathiaka S., Acevedo O., Boggon T.J., Krauthammer M., Halaban R., Xu T., unpublished data). These findings clearly demonstrate the power of our experimental approach for the prospective identification of drug-resistant mutants.

Methods

Cell culture

A375 and Ba/F3 cells were cultured in RPMI-1640 (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% fetal bovine serum (FBS) (Atlanta Biologicals, Flowery Branch, GA, USA), 1 mM sodium pyruvate, and Pen/Strep (Sigma). Media for Ba/F3 cells were additionally supplemented with 10 ng/ml murine IL-3 (Peprotech, Rocky Hill, NJ, USA). The identity of the Ba/F3 cell line was confirmed by IL-3 withdrawal experiments and fluorescence-activated cell sorting (FACS) analysis (Figure S9). 293T cells were cultured in DMEM (Sigma) with 10% FBS, 1 mM sodium pyruvate and Pen/Strep.

Transient and stable transfections

Retroviruses expressing BRAFV600E (Addgene, Cambridge, MA, USA) and all variants were generated by transient transfection of 293T cells using Effectene (Qiagen, Valencia, CA, USA). Retroviral supernatants were collected 48 h post-transfection, filtered, and supplemented with 5 or 10 μg/ml polybrene for infection of Ba/F3 or A375 cells, respectively. For infection of Ba/F3 cells, cells were centrifuged for 90 min at 1000 × g. At 24 h post-infection, cells were selected with puromycin at 2 μg/ml. Transient transfections of 293T cells were performed using Effectene.

Mutagenesis screen

Saturation mutagenesis was used to generate point mutant libraries for individual amino acids. To facilitate mutagenesis, a portion of BRAFV600E (nucleotides 1270–1900) was cloned into pRNDM[14], a minimal cloning plasmid, bracketed by SacI and SphI

restriction sites. Cassette mutagenesis was performed on this plasmid as previously described.[13,14] The mutant pools were then transferred into the retroviral vector pBABE-Puro-BRAFV600E (Addgene #15269) by sequence and ligation-independent cloning (SLIC).[149] Briefly, the mutagenized cassette was excised using SacI and SphI, treated with T4 DNA polymerase to create 5′ single-stranded overhangs (approximately 30 nt) and then purified on a silica column. A pBABE-Puro-BRAFV600E SLIC vector was created by replacing BRAF nucleotides 1300–1870 with an ApaI restriction site. The construct was linearized and treated with T4 DNA polymerase as described above. The treated mutagenized cassette library and pBABE SLIC vector were annealed, and the entire reaction transformed into bacteria. Based on plating of a small fraction of the bacteria, this procedure routinely resulted in > 30 000 transformants, which was sufficient to transfer our libraries without compromising diversity. Focused deep sequencing was used to determine the representation of each mutation in the pBABE plasmid libraries and indicated that all mutations were present well above noise level. Individual mutations were generated by the same method and confirmed by Sanger sequencing.

Retroviral particles were generated by transient transfection of 293T cells using Effectene as described above. A375 cells (3 × 106) were transduced at a multiplicity of infection of three with the retroviral BRAFV600E mutant pools in 100-mm plates, selected for stable integration with puromycin, and cultured in the presence of 10 μM PLX4720 for 3 weeks. Drug-resistant colonies were isolated and pooled, and genomic DNA was extracted.

The library version of BRAF was PCR-amplified from genomic DNA and prepared for Illumina sequencing as described.[14] Briefly, PCR was used to introduce an MmeI site immediately 5′ to randomized regions and a 3′ Illumina primer binding site 250 bases downstream. This product was digested with MmeI and ligated to adapters that included a barcode and a 5′ Illumina primer binding site. This product was PCR-amplified with Illumina Sequencing Primers, gel-purified, and submitted for short single-read (36 bases) analysis using a Solexa-Illumina GA Massively Parallel Deep Sequencer. Raw sequencing files in fastq format were processed essentially as described.[14] All sequences were subjected to quality filtering requiring a PHRED score of ≥20 at all read positions. In addition, sequence reads where more than one codon differed from the parental construct were eliminated from analysis. The remaining sequences were analyzed to determine the number of reads for each point mutant in the plasmid library, as well as in transduced cells prior to and following treatment with drug. Mutations that created internal MmeI sites or that were severely under-represented in cells (100-fold relative to median wild-type synonym) without drug were omitted from analysis. Of the 1160 possible protein point mutations (58*20), all but five (478A, 486D, 486E, 511W, and 580W) were successfully analyzed. The ratio of each codon substitution from cells after and before treatment with drug was used as a metric of drug resistance. The median enrichment of silent substitutions was used to normalize the data (resulting in the median synonymous substitution having an enrichment score of 1). Based on the distribution of enrichments for synonymous substitutions, we set a false discovery threshold for each codon observation. Amino acid substitutions were considered enriched in drug if all

observed synonymous substitutions were above a statistical cutoff such that the net P-value was < 0.01.

Structural images

Images of BRAF bound to PLX4720 were generated from Protein Data Bank (PDB) entry 3C4C[158] using the Pymol software package (Delano Inc., San Carlos, CA, USA)

Drug treatment

PLX4720 (Selleckchem, Houston, TX, USA), SB590885 (Tocris, Minneapolis, MN, USA), RAF265 (Selleckchem), U0126 (Cell Signaling, Beverly, MA, USA), GDC0879 (Selleckchem), and vemurafenib/PLX4032 (Selleckchem) were prepared in DMSO at 20 mM. 1 × 103 A375 or 1.5 × 104 Ba/F3 cells were plated per well of a 96-well plate in 100 μl volume, and 24 h later, 50 μl of drug was added to the cells.

Cellular and pMEK IC50 determination

Cells expressing BRAFV600E or BRAFV600E mutants were cultured in the presence of serially diluted drug for 72 h, and viability was measured by Alamar Blue assay (Invitrogen, Grand Island, NY, USA). In Figure 2A, Trp and Met mutants were excluded from consideration. For Figures 2F, 4E, F and 6A–C, data were plotted in GraphPad Prism and a dose–response curve was fit with nonlinear regression. To

determine p-MEK IC50, the phospho-MEK immunoblots were quantified in ImageJ (NIH) and the densitometry used to plot the data.

Immunoblotting

Cells were lysed with cell lysis buffer [1%NP-40, 10% glycerol, 150 mM NaCl, 50 mM Tris–HCl (pH 7.4), 1 mM EDTA] supplemented with complete protease inhibitor tablet (Roche, Indianapolis, IN, USA) and phophatase inhibitor cocktails 2 and 3 (Sigma). Cell extract was quantified with bicinchoninic acid (Pierce, Rockford, IL, USA), and equal amounts of protein were loaded in each lane. Blots were probed with the following primary antibodies: phosphorylated MEK (S217/221), total MEK, phosphorylated ERK1/2 (T202/Y204), total ERK1/2, myc (all from Cell Signaling Technology), and tubulin.

Cell growth assays

Parental Ba/F3 cells or Ba/F3 cells stably transduced with BRAFV600E or BRAFV600E/L505H and cultured in the presence or absence of 10 ng/ml IL-3 were assessed for growth by counting.

qRT-PCR analysis

Total RNA was isolated from A375 cells treated in the presence or absence of drug using TriPure Isolation Reagent (Roche). Reverse transcription was performed using SuperScript II Reverse Transcriptase (Invitrogen) followed by qPCR using Fast SYBR

Green Master Mix (Applied Biosystems, Foster City, CA, USA) using the following primers: FOS1L forward (5′-CACTCCAAGCGGAGACAGAC-3′) and reverse (5′-AGGTCATCAGGGATCTTGCAG-3′); SPRY2 forward (5′-ATGGCATAATCCGGGTGCAA-3′) and reverse (5′-TGTCGCAGATCCAGTCTGATG-3′); and DUSP6 forward (5′-AGCTCAATCTGTCGATGAACG-3′) and reverse (5′-GCGTCCTCTCGAAGTCCAG-3′). The expression level of each gene was normalized to that of three internal control genes, B2M forward (5′-CGCTCCGTGGCCTTAGC-3′) and reverse (5′-AATCTTTGGAGTACGCTGGATAGC-3′); TBP forward (5′-CACAG GAGCCAAGAGTGAAG-3′) and reverse (5′-CAAGGCCTTCTAACCTTATAGG-3′); and GUSB forward (5′-TTGAGCAAGACTGATACCACCTG-3′) and reverse (5′-TCTGGTCTGCCGTGAACAGT-3′). The qRT-PCR data were normalized to the three internal control genes individually and then combined.

Tumor formation assays

Balb/c nu/nu mice (NCI; n = 5 per group) were injected subcutaneously with 6 × 106 Ba/F3 cells transduced with BRAFV600E or BRAFV600E/L505H. Tumors were allowed to form for 14 days at which point animals received daily intraperitoneal injections of vehicle (water + 5% DMSO) or PLX4720 (50 mg/kg). Tumor dimensions were measured every 3–4 days, and tumor volume was calculated using the formula $\pi/6 \times (\text{length}) \times (\text{width})2$. After the final dose, animals were sacrificed and tumors were excised and weighed. For statistical analysis, Wilcoxon rank sum test was performed in R

(www.r-project.org). Animal experiments were performed in accordance with the Institutional Animal Care and Use Committee (IACUC) guidelines.

Probability determinations

The probabilities of mutation types for idealized libraries were calculated based on independent probabilities (Bayesian theory). For random mutagenesis, these probability estimates assume ideal conditions (a mutation rate of one base per gene). For a protein of N amino acids (3N nucleotides), the probability of a mutation at any base (p) is 1/3N, the probability of any position being wild-type (q) is 1-p, and the probability of all positions being wild-type is q3N. The dependence of this equation on protein length is negligible for proteins > 100 amino acids. The probability of having only one base mutated is p*q(3N−1). The probability of having two or more mutations is 1 – wild-type probability – single-mutant probability.

Acknowledgements

We thank the UMMS Deep Sequencing Core; M. Perkins and S. Bhatnagar for assistance with animal experiments; L. J. Zhu for assistance with statistical analysis; and S. Deibler for editorial assistance. This work was supported in part by a grant from the National Institutes of Health (R01GM083038) to D.N.B. M.R.G. is an investigator of the Howard Hughes Medical Institute.

Chapter V – Systematic analysis of ubiquitin point mutants under temperature stress

Benjamin P. Roscoe and Daniel N. Bolon

The experimental results presented in this chapter are currently unpublished. I performed all of the experiments herein, and Dan Bolon and I analyzed the data and generated figures. Further experiments, including biochemical and biophysical analysis of single substitution ubiquitin variants as well as a repeat of the bulk competition experiments is necessary before preparation of a final manuscript.

Rationale:

Chapter II of this dissertation describes the effect all ubiquitin point mutants on yeast growth rate under permissive conditions (30º C) in synthetic dextrose media.[169] A major conclusion from this study is that many single amino acid substitutions were tolerated throughout the protein sequence, negatively correlated with client interaction density that we defined by analysis of the average surface burial of ubiquitin in 44 co-crystal structures.[169] That finding could not be predicted from phylogenetic comparisons, as ubiquitin is ultra-highly conserved with respect to protein sequence in eukaryotes (Figure 3.5a). The discrepancies between tolerated amino acid substitutions to a protein sequence in laboratory conditions versus that sampled over evolutionary time can have multiple explanations. First, measurement errors in our experimental bulk competition fitness measurements (chapter III,) are higher than the minimum fitness defects that will cause purifying selection in a natural population ($s \approx 10^{-6}$).[74,141] A second explanation is that under different growth conditions there will be less tolerance for mutational change. In previous work from our lab, Ryan Hietpas, Ph.D. *et al* found that a 9 amino acid stretch of hsp90 was less tolerant to amino acid substitution under conditions of high salinity or temperature stress.[170]

In order to determine the effect of temperature stress on the tolerance of ubiquitin to amino acid substitutions, I performed a bulk competition experiment using the same randomized libraries and strategy described in chapter II[169] at 36º C, a temperature that induces stress in yeast.[171]

Results and Discussion

Comparison of ubiquitin amino acid substitutions at 30º C vs. 36º C

The tolerance of yeast to ubiquitin amino acid substitutions was markedly reduced in bulk competition experiments at 36º C compared to growth competitions at 30º C, with 141 tolerant substitutions (fitness > 0.9) at 30º C becoming deleterious (fitness < 0.5) at elevated temperature (table 5.1, figure 5.1&5.2a). Wild type synonyms all remained universally fit and stop codons were all deleterious. The overall pattern of amino acid position tolerance to mutation remained similar, with notable exceptions at positions K6 and K63, with neither tolerating any substitutions at elevated temperature. Lysine 63-linked ubiquitin is involved in DNA damage repair, and the K63R mutant as well as deletions of the Rsp5 HECT domain responsible for K63 linkage specificity in DNA repair have both previously been demonstrated to cause a temperature sensitive (ts) phenotype at 36º C in *S. cerevisae* (figure 5.1b, dashed red line).[172]

Temperature sensitive mutations were distributed throughout the protein, with similar fractions of temperature sensitive mutants in both the solvent-accessible surface/boundary residues and the solvent-inaccessible core residues (Figure 5.2b&c). This suggests that intermolecular interactions with binding partners as well as intramolecular structural perturbations are both likely mechanisms of temperature sensitive phenotypes.

Figure 5.1. Heatmaps of ubiquitin point mutant effects on yeast growth rates at 30º and 36º C. a) The ubiquitin gene was divided into eight separate libraries for bulk growth competition and deep sequencing analysis (described in chapter II). b) Heatmaps representing the amino acid substitution average relative growth rate scaled from null-like (blue) to WT-like yellow at each temperature. Lysine linkage sites (K6 and K63) that were found to be critical at elevated temperatures are highlighted by red dashed rectangles.

Figure 5.1. Comparison of ubiquitin mutant fitness effects at 30º C vs. 36º C.

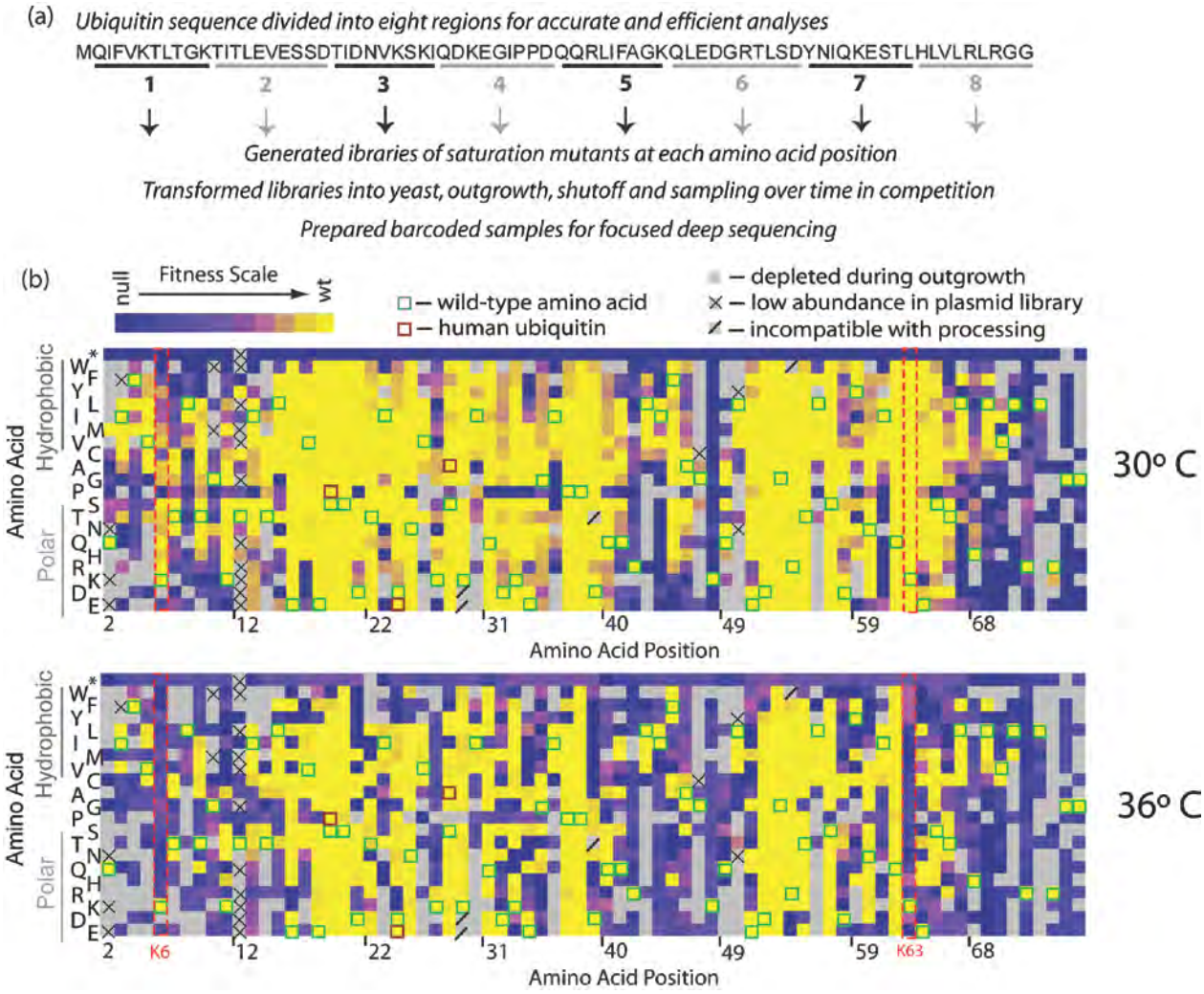
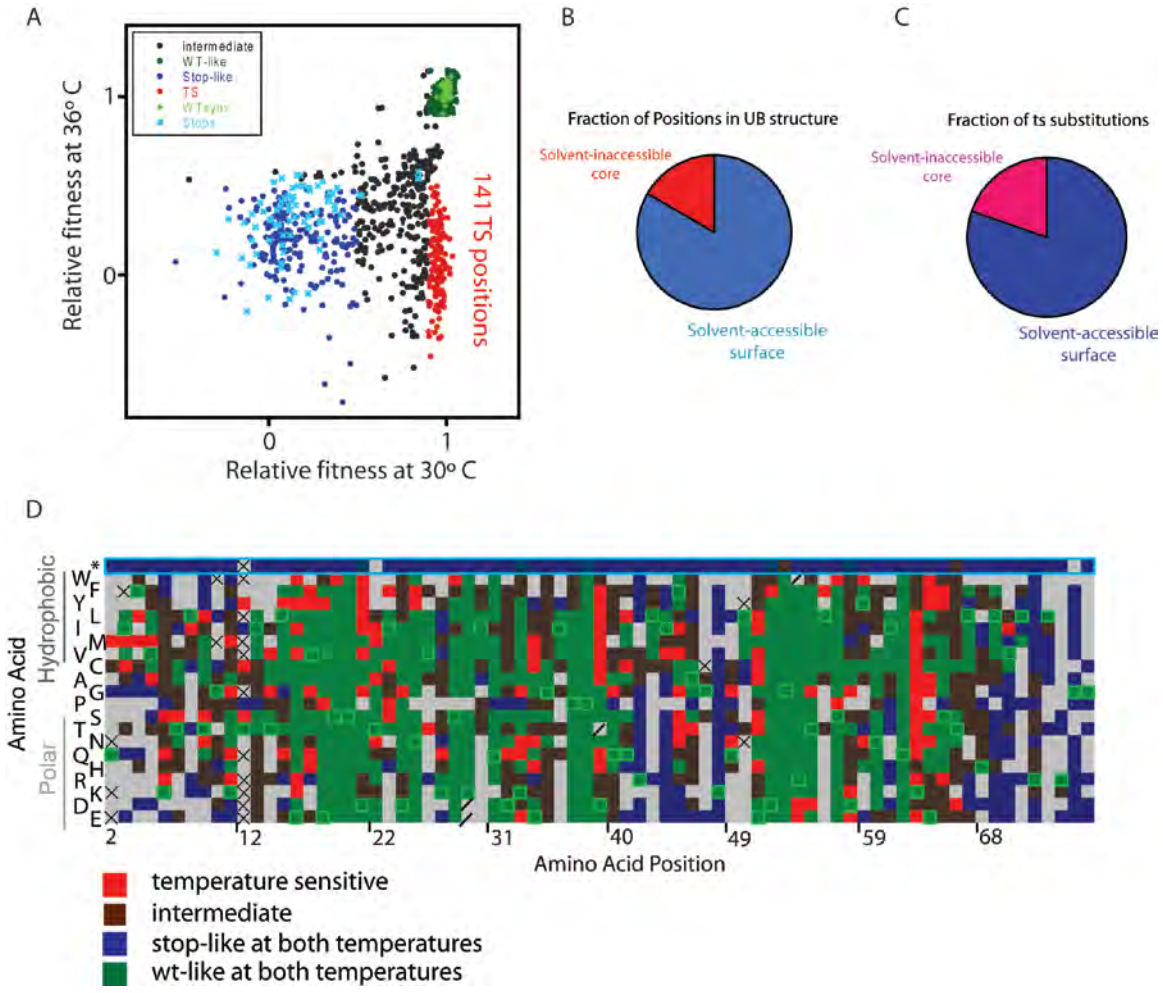


Figure 5.2. Comparison of ubiquitin point mutant effects on yeast growth at 30º vs. 36º C. A. Correlation plot of ubiquitin point mutant relative growth rate at 30º vs. 36º C. B. Pie chart representing fraction of temperature sensitive point mutants on the solvent-exposed surface and boundary residues of ubiquitin. C. Pie chart representing fraction of temperature sensitive point mutants in the solvent-inaccessible core of ubiquitin. D. Difference map of ubiquitin point mutant fitness classifications at 30º vs. 36º C.

Figure 5.2. Comparison of ubiquitin point mutant effects on yeast growth at 30º vs. 36º C.

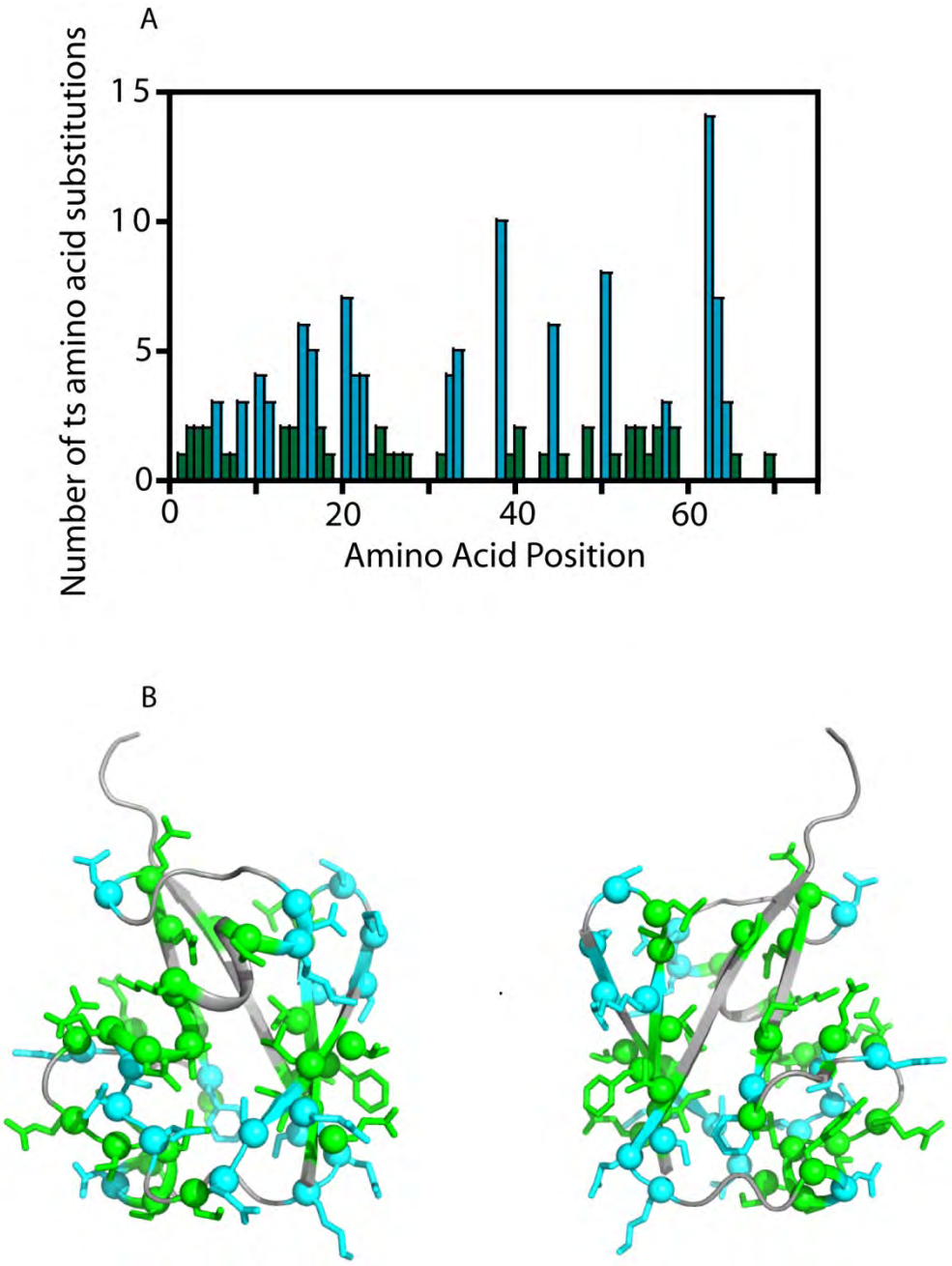


Mapping temperature sensitive mutants to structure

We tabulated the number of temperature sensitive substitutions at each amino acid position of ubiquitin and mapped them to the crystal structure of monoubiquitin (figure 5.3 a&b). Temperature sensitive substitutions appeared to be fairly evenly distributed across the structure (figure 5.3b). This result was unexpected given the sensitivity of a common ubiquitin binding interface described in chapters II+III. Biochemical analysis of individual temperature sensitive mutants for folding defects, binding of known partner proteins (e.g. proteasome receptors) and *in vivo* polyubiqutin formation in yeast at both temperatures will be necessary to further investigate this finding.

Figure 5.3. Mapping temperature sensitive ubiquitin mutants to structure. A. Number of temperature sensitive mutants at each position of ubiquitin. B) Location of temperature sensitive mutants on the ubiquitin crystal structure 1UBQ.pdb. Positions with 1-4 temperature sensitive mutants are colored in green, and positions with >4 temperature sensitive mutants are colored in blue.

Figure 5.3. Mapping temperature sensitive ubiquitin mutants to structure



Conclusions and future directions

This preliminary investigation clearly demonstrates the power of utilizing saturation mutagenesis and deep sequencing to discover temperature sensitive mutations in a single protein in yeast. We identified 141 temperature sensitive amino acid substitutions distributed in both the solvent-accessible surface/boundary residues and solvent-inaccessible core. This suggests that both inter- and intramolecular interactions likely contribute to observed temperature sensitive phenotypes. Future experiments to delineate these effects will be necessary to specify the specific causes of the temperature sensitive phenotype of individual mutants. For example, panels of both core and surface recombinant temperature sensitive mutants should be purified and analyzed by circular dichroism under chemical denaturation for changes in folding properties. Analysis of previously reported thermodynamic folding defects and computational predictions of changes in free energy of ubiquitin point mutants did not correlate strongly with temperature sensitive mutants. Immunoblotting and/or immunoprecipitation of specific complexes for polyubiquitin accumulation of specific mutants at both temperatures may also help pinpoint specific defects.

Chapter VI – Discussion

The research presented throughout this dissertation is unified by a single underlying experimental strategy, systematic mutagenesis of proteins focused on obtaining rich information on biochemical and cellular function through a deep-sequencing readout. The two different biological systems studied in this work (yeast ubiquitin and BRAF V600E from human melanoma) are both supported by a rich body of mechanistic and structural studies, allowing us to comprehensively apply our data to give new insights into function in light of known biochemical mechanisms. This supports the value of systematic mutagenesis and focused deep sequencing as a way to explore the physical and chemical space requirements of the amino acid sequence of a protein. The strong correlation of residue tolerance to mutation in these systems in light of previously described functions in the literature suggests that extending this methodology to other lesser understood systems can result in rich interpretable data. Ongoing work in our group focused on the evolution of drug resistance in viral targets strongly supports this prediction.

The original research presented in chapter II reports the fitness effects of all possible point mutations to ubiquitin in yeast. Perhaps the most intriguing finding from this study is that ubiquitin tolerates many amino acid substitutions under the reported experimental conditions (30º C, synthetic dextrose media). These results demonstrate that it is possible for ubiquitin to acquire many neutral or near neutral amino acid substitutions in permissive conditions, adding to the body of literature supporting the

theory of near-neutral mutations providing a mechanism of adaptation under environmental selective pressures.[13,21,22,76,173]

A second finding from the systematic analysis of ubiquitin point mutants is that specific residue tolerance to mutation correlates strongly with the interaction surfaces of ubiquitin with binding partners (from ASA burial of 44 co-crystal structures in the PDB). This lends strong support to the theory that the mutational tolerance of a a specific residue in a protein is restrained by its molecular contacts, as proposed by E. Zuckerlandl in 1976.[23] While this finding may seem intuitive, because of the previous technical challenges in generating and testing thousands of amino acid substitutions, our studies in ubiquitin are some of the most comprehensive analyses of the fitness effects of amino acid substitutions at each position in a gene. For other proteins with less characterized cellular interaction, these strategies will likely be very useful for discovering regions of proteins involved in previously unidentified processes.

The work in chapter III presents findings from examining the effect of all possible point mutants in ubiquitin residues 2-76 on the essential activation by the yeast E1 ubiquitin activating protein (Uba1p). Because ubiquitin is a multifunctional protein with many steps to activation even for the canonical function of targeting client proteins to the 26S proteasome for degradation,[32] it is near impossible to predict which defects in function explain fitness defects. It can be inferred that because activation of ubiquitin by E1 is essential, that defects that prevent this activation will prevent cellular growth. However, the relationship between partial defects and the rate of growth defects was found to be complex. Mutants that had no discernible E1 activation activity were indeed

very deleterious for growth, however the yeast tolerated a reduction in ubiquitin activation by 98% in at least one mutant (R72S). Many other mutants had smaller but measurable defects in E1 activation, yet most were unable to grow. These results strongly suggests that while E1 activation is not rate limiting for yeast fitness, competitive defects in E1 activation predict downstream functional defects. Analysis of the available yeast E1 : ubiquitin co-crystal structure[56] indicated that E1 has a very large interaction surface compared to many other ubiquitin binding partners downstream. This led us to posit that E1 is sensitive to small competitive defects in ubiquitin activation, consistent with a model that suggests E1 molecules without this filtering capacity may be selected against in a natural yeast population.

Chapter IV of this dissertation describes a collaborative effort of the Daniel N. Bolon and Michael R. Green research groups to extend high throughput mutational screening to predict inhibitor resistant mutants in the oncogenic bRAF V600E kinase from human melanoma. Single point mutations are a common method of drug resistance in drug-binding targets. Therefore, we used the structure of the inhibitor bound to bRAF V600E to target the drug binding portions of bRAF V600E mutagenesis. By selecting mutant library cell lines that were able to grow in the presence of the inhibitor, we were able to discover over a dozen novel mutants that conferred drug resistance. The findings from the high-throughput screen were validated by cell-based screens, probing of the kinase phosphorylation pathway, and tumor resistance in mice. One particularly resistant mutant, V600E L505H is acquirable through a single base substitution, and was recently discovered in a human melanoma cell line (Choi J.,

Landrette S., Wang T., Evans P., Bacchiocchi A., Bjornson R., Cheng E., Stiegler A.L., Gathiaka S., Acevedo O., Boggon T.J., Krauthammer M., Halaban R., Xu T., unpublished data). These findings clearly demonstrate the power of our experimental approach for the prospective identification of drug-resistant mutants.

Selection pressure over evolutionary time encompasses all environmental and competitive pressure to the ancestors of an organism over billions of years. I believe it is presumptuous to believe that we can ever perfectly reconstruct the changing environment and timescales of evolutionary time, but I do believe that directed evolution experiments within the laboratory are able to provide strong evidence into how proteins can evolve under changing environments. In chapter V, the tolerance of ubiquitin in yeast to single point mutations is greatly reduced by a simple change in experimental conditions from permissive to heat stress temperatures. This suggests that the evolutionary conservation of ubiquitin may be able to be explained partially by changing environmental conditions, a hypothesis that is testable using these methods and a current direction of our research. Integrating the fitness effects of many point mutants across many conditions will likely help support population genetics models, especially as the mutational tolerance of many proteins is reported. Ryan Hietpas' work in Hsp90 demonstrated a similar result, including some mutations that were more fit than wild type under conditions of high salinity.[170] It should also be noted that many other factors lead to organismal fitness, including the genetic background of the organism that leads to its entire network of function, including factors such as relative expression level.[111]

While the experiments described in the preceding chapters focus on amino acid substitutions at the protein level, many other applications can be imagined. The increasing throughput we demonstrated by introducing all point mutants to the ubiquitin gene in yeast can be applied to studies of codon preference in organisms, all well as studies of RNA machinery including ribosomal subunits and enzymes composed of RNA. It is also possible to apply these mutagenesis and sequencing strategies to nucleic acid regions of unknown or poorly understood function and test the effects on cells or organisms under many different conditions or genetic backgrounds, potentially discovering epigenetic effects of changes to chromosomal DNA.

In summary, I believe that our work, as well as the discoveries of other researchers investigating the role of amino acid sequence in function, will help bridge the divides between theoretical models of protein evolution and biochemistry. This dissertation provides evidence that this methodology has the potential to be applicable to human medicine in the context of drug resistance in cancer, a mechanism driven by mutation,[20] often leading to amino acid substitutions in proteins involved in cellular signaling. Technology is paramount to furthering discovery in all of the scientific disciplines, but the most important part is applying the proper technology and methodology to the most important scientific problems.

"Be technique agnostic." -Dean Tony Carruthers to the incoming GSBS class during orientation 2008.

Bibliography

1. Anfinsen, C.B. (1972). The formation and stabilization of protein structure. *Biochem. J.* **128**, 737–49

2. Khorana, H.G. (1959). Synthesis and structural analysis of polynucleotides. *J. Cell. Comp. Physiol.* **54**, 5–15

3. Holley, R.W., Apgar, J. & Doctor, B.P. (1960). Separation of amino acid-specific "soluble"-fraction ribonucleic acids. *Ann. N. Y. Acad. Sci.* **88**, 745–51

4. Leder, P., Clark, B.F., Sly, W.S., Pestka, S. & Nirenberg, M.W. (1963). Cell-free peptide synthesis dependent upon synthetic oligodeoxynucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **50**, 1135–43

5. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–42

6. Radivojac, P., Clark, W.T., Oron, T.R., Schnoes, A.M., Wittkop, T., Sokolov, A., Graim, K., Funk, C., Verspoor, K., Ben-Hur, A., Pandey, G., Yunes, J.M., Talwalkar, A.S., Repo, S., Souza, M.L., Piovesan, D., Casadio, R., Wang, Z., Cheng, J., Fang, H., Gough, J., Koskinen, P., Törönen, P., Nokso-Koivisto, J., Holm, L., Cozzetto, D., Buchan, D.W.A., Bryson, K., Jones, D.T., Limaye, B., Inamdar, H., Datta, A., Manjari, S.K., Joshi, R., Chitale, M., Kihara, D., Lisewski, A.M., Erdin, S., Venner, E., Lichtarge, O., Rentzsch, R., Yang, H., Romero, A.E., Bhat, P., Paccanaro, A., Hamp, T., Kaßner, R., Seemayer, S., Vicedo, E., Schaefer, C., Achten, D., Auer, F., Boehm, A., Braun, T., Hecht, M., Heron, M., Hönigschmid, P., Hopf, T.A., Kaufmann, S., Kiening, M., Krompass, D., Landerer, C., Mahlich, Y., Roos, M., Björne, J., Salakoski, T., Wong, A., Shatkay, H., Gatzmann, F., Sommer, I., Wass, M.N., Sternberg, M.J.E., Škunca, N., Supek, F., Bošnjak, M., Panov, P., Džeroski, S., Šmuc, T., Kourmpetis, Y.A.I., van Dijk, A.D.J., ter Braak, C.J.F., Zhou, Y., Gong, Q., Dong, X., Tian, W., Falda, M., Fontana, P., Lavezzo, E., Di Camillo, B., Toppo, S., Lan, L., Djuric, N., Guo, Y., Vucetic, S., Bairoch, A., Linial, M., Babbitt, P.C., Brenner, S.E., Orengo, C., Rost, B., Mooney, S.D. & Friedberg, I. (2013). A large-scale evaluation of computational protein function prediction. *Nat. Methods* **10**, 221–7

7. Harms, M.J. & Thornton, J.W. (2013). Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14**, 559–71

8. Pannekoek, H., van Meijer, M., Schleef, R.R., Loskutoff, D.J. & Barbas, C.F. (1993). Functional display of human plasminogen-activator inhibitor 1 (PAI-1) on

phages: novel perspectives for structure-function analysis by error-prone DNA synthesis. *Gene* **128**, 135–40

9. Loeb, D.D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S.E. & Hutchison, C.A. (1989). Complete mutagenesis of the HIV-1 protease. *Nature* **340**, 397–400

10. Rennell, D., Bouvier, S.E., Hardy, L.W. & Poteete, A.R. (1991). Systematic mutation of bacteriophage T4 lysozyme. *J. Mol. Biol.* **222**, 67–88

11. Fishman, A., Tao, Y., Bentley, W.E. & Wood, T.K. (2004). Protein engineering of toluene 4-monooxygenase of Pseudomonas mendocina KR1 for synthesizing 4-nitrocatechol from nitrobenzene. *Biotechnol. Bioeng.* **87**, 779–90

12. 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**, 741–6

13. Hietpas, R.T., Jensen, J.D. & Bolon, D.N.A. (2011). Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7896–901

14. Hietpas, R., Roscoe, B., Jiang, L. & Bolon, D.N.A. (2012). Fitness analyses of all possible point mutations for regions of genes in yeast. *Nat. Protoc.* **7**, 1382–96

15. McLaughlin, R.N., Poelwijk, F.J., Raman, A., Gosal, W.S. & Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. *Nature* **491**, 138–42

16. 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. U. S. A.* **109**, 16858–63

17. Whitehead, T.A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S.J., De Mattos, C., Myers, C.A., Kamisetty, H., Blair, P., Wilson, I.A. & Baker, D. (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **30**, 543–8

18. Starita, L.M., Pruneda, J.N., Lo, R.S., Fowler, D.M., Kim, H.J., Hiatt, J.B., Shendure, J., Brzovic, P.S., Fields, S. & Klevit, R.E. (2013). Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E1263–72

19. Dean, A.M. & Thornton, J.W. (2007). Mechanistic approaches to the study of evolution: the functional synthesis. *Nat. Rev. Genet.* **8**, 675–88

20. Hochhaus, A., Schenk, T., Erben, P., Ernst, T., La Rosée, P. & Müller, M.C. (2009). Cause and management of therapy resistance. *Best Pract. Res. Clin. Haematol.* **22**, 367–79

21. Nei, M. (2007). The new mutation theory of phenotypic evolution. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12235–42

22. Kimura, M. (1983). Rare variant alleles in the light of the neutral theory. *Mol. Biol. Evol.* **1**, 84–93

23. Zuckerkandl, E. (1976). Evolutionary processes and evolutionary noise at the molecular level. I. Functional density in proteins. *J. Mol. Evol.* **7**, 167–83

24. Jones, S. & Thornton, J.M. (1997). Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.* **272**, 121–32

25. Clark, W.T. & Radivojac, P. (2011). Analysis of protein function and its prediction from amino acid sequence. *Proteins* **79**, 2086–96

26. Clackson, T., Hoogenboom, H.R., Griffiths, A.D. & Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature* **352**, 624–8

27. Lowman, H.B., Bass, S.H., Simpson, N. & Wells, J.A. (1991). Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* **30**, 10832–8

28. Boder, E.T. & Wittrup, K.D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–7

29. Kacser, H. & Burns, J.A. The molecular basis of dominance. *Genetics* **97**, 639–66

30. Grossniklaus, U., Madhusudhan, M.S. & Nanjundiah, V. (1996). Nonlinear enzyme kinetics can lead to high metabolic flux control coefficients: implications for the evolution of dominance. *J. Theor. Biol.* **182**, 299–302

31. Hershko, A. & Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–79

32. Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513

33. Thrower, J.S., Hoffman, L., Rechsteiner, M. & Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19**, 94–102

34. 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**, 1265–73

35. Messick, T.E. & Greenberg, R.A. (2009). The ubiquitin landscape at DNA double-strand breaks. *J. Cell Biol.* **187**, 319–26

36. Hicke, L., Schubert, H.L. & Hill, C.P. (2005). Ubiquitin-binding domains. *Nat. Rev. Mol. Cell Biol.* **6**, 610–21

37. Dikic, I., Wakatsuki, S. & Walters, K.J. (2009). Ubiquitin-binding domains - from structures to functions. *Nat. Rev. Mol. Cell Biol.* **10**, 659–71

38. Ye, Y. & Rape, M. (2009). Building ubiquitin chains: E2 enzymes at work. *Nat. Rev. Mol. Cell Biol.* **10**, 755–64

39. Hatakeyama, S., Yada, M., Matsumoto, M., Ishida, N. & Nakayama, K.I. (2001). U box proteins as a new family of ubiquitin-protein ligases. *J. Biol. Chem.* **276**, 33111–20

40. Kim, H.C. & Huibregtse, J.M. (2009). Polyubiquitination by HECT E3s and the determinants of chain type specificity. *Mol. Cell. Biol.* **29**, 3307–18

41. Petroski, M.D. (2008). The ubiquitin system, disease, and drug discovery. *BMC Biochem.* **9 Suppl 1**, S7

42. Wintrode, P.L., Makhatadze, G.I. & Privalov, P.L. (1994). Thermodynamics of ubiquitin unfolding. *Proteins* **18**, 246–53

43. Sierra, J.R., Cepero, V. & Giordano, S. (2010). Molecular mechanisms of acquired resistance to tyrosine kinase targeted therapy. *Mol. Cancer* **9**, 75

44. Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A.P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.-D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Güldener, U., Hegemann, J.H., Hempel, S., Herman, Z., Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kötter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K.,

Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W. & Johnston, M. (2002). Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387–91

45. 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**, 124–44

46. Pitt, J.N. & Ferré-D'Amaré, A.R. (2010). Rapid construction of empirical RNA fitness landscapes. *Science* **330**, 376–9

47. Fleishman, S.J., Whitehead, T.A., Ekiert, D.C., Dreyfus, C., Corn, J.E., Strauch, E.-M., Wilson, I.A. & Baker, D. (2011). Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Science (80-.).* **332**, 816–821

48. 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**, 1576–83

49. Rock, K.L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D. & Goldberg, A.L. (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**, 761–71

50. Goldberg, A.L. (2007). Functions of the proteasome: from protein degradation and immune surveillance to cancer therapy. *Biochem. Soc. Trans.* **35**, 12–7

51. 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**, 1331–5

52. Alkalay, I., Yaron, A., Hatzubai, A., Orian, A., Ciechanover, A. & Ben-Neriah, Y. (1995). Stimulation-dependent I kappa B alpha phosphorylation marks the NF-kappa B inhibitor for degradation via the ubiquitin-proteasome pathway. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 10599–603

53. Dennissen, F.J.A., Kholod, N. & van Leeuwen, F.W. (2012). The ubiquitin proteasome system in neurodegenerative diseases: culprit, accomplice or victim? *Prog. Neurobiol.* **96**, 190–207

54. Orlowski, R.Z. & Kuhn, D.J. (2008). Proteasome inhibitors in cancer therapy: lessons from the first decade. *Clin. Cancer Res.* **14**, 1649–57

55. 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**, 34983–91

56. Lee, I. & Schindelin, H. (2008). Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* **134**, 268–78

57. 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**, 7300–8

58. Miura, T., Klaus, W., Gsell, B., Miyamoto, C. & Senn, H. (1999). Characterization of the binding interface between ubiquitin and class I human ubiquitin-conjugating enzyme 2b by multidimensional heteronuclear NMR spectroscopy in solution. *J. Mol. Biol.* **290**, 213–28

59. Hurley, J.H., Lee, S. & Prag, G. (2006). Ubiquitin-binding domains. *Biochem. J.* **399**, 361–72

60. Fushman, D. & Wilkinson, K.D. (2011). Structure and recognition of polyubiquitin chains of different lengths and linkage. *F1000 Biol. Rep.* **3**, 26

61. Hofmann, K. & Falquet, L. (2001). A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem. Sci.* **26**, 347–50

62. 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**, 28976–84

63. Hofmann, K. & Bucher, P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem. Sci.* **21**, 172–3

64. Ohno, A., Jee, J., Fujiwara, K., Tenno, T., Goda, N., Tochio, H., Kobayashi, H., Hiroaki, H. & Shirakawa, M. (2005). Structure of the UBA domain of Dsk2p in complex with ubiquitin molecular determinants for ubiquitin recognition. *Structure* **13**, 521–32

65. 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**, 1042–7

66. 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. U. S. A.* **93**, 861–6

67. Sloper-Mould, K.E., Jemc, J.C., Pickart, C.M. & Hicke, L. (2001). Distinct functional surface regions on ubiquitin. *J. Biol. Chem.* **276**, 30483–9

68. Went, H.M. & Jackson, S.E. (2005). Ubiquitin folds through a highly polarized transition state. *Protein Eng. Des. Sel.* **18**, 229–37

69. Benítez-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**, 5195–203

70. 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**, 979–96

71. 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**, 343–57

72. 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**, 7115–23

73. Lynch, M. & Conery, J.S. (2003). The origins of genome complexity. *Science* **302**, 1401–4

74. Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., Tsai, I.J., Bergman, C.M., Bensasson, D., O'Kelly, M.J.T., van Oudenaarden, A., Barton, D.B.H., Bailes, E., Nguyen, A.N., Jones, M., Quail, M.A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R. & Louis, E.J. (2009). Population genomics of domestic and wild yeasts. *Nature* **458**, 337–41

75. Sanjuán, R., Moya, A. & Elena, S.F. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8396–401

76. 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**, 12979–84

77. Domingo-Calap, P., Cuevas, J.M. & Sanjuán, R. (2009). The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet.* **5**, e1000742

78. 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

79. Pursell, N.W., Mishra, P. & Bolon, D.N.A. (2012). Solubility-promoting function of Hsp90 contributes to client maturation and robust cell growth. *Eukaryot. Cell* **11**, 1033–41

80. Wayne, N. & Bolon, D.N. (2010). Charge-rich regions modulate the anti-aggregation activity of Hsp90. *J. Mol. Biol.* **401**, 931–9

81. Butt, T.R., Jonnalagadda, S., Monia, B.P., Sternberg, E.J., Marsh, J.A., Stadel, J.M., Ecker, D.J. & Crooke, S.T. (1989). Ubiquitin fusion augments the yield of cloned gene products in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2540–4

82. King, J.L. & Jukes, T.H. (1969). Non-Darwinian evolution. *Science* **164**, 788–98

83. 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**, 204–11

84. 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**, 687–98

85. 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**, 611–22

86. Bolon, D.N. & Mayo, S.L. (2001). Polar residues in the protein core of Escherichia coli thioredoxin are important for fold specificity. *Biochemistry* **40**, 10047–53

87. Cordes, M.H., Davidson, A.R. & Sauer, R.T. (1996). Sequence space, folding and protein design. *Curr. Opin. Struct. Biol.* **6**, 3–10

88. 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**, 463–75

89. Sims, J.J., Haririnia, A., Dickinson, B.C., Fushman, D. & Cohen, R.E. (2009). Avid interactions underlie the Lys63-linked polyubiquitin binding specificities observed for UBA domains. *Nat. Struct. Mol. Biol.* **16**, 883–9

90. Massi, F., Grey, M.J. & Palmer, A.G. (2005). Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR R1rho relaxation experiments. *Protein Sci.* **14**, 735–42

91. 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

92. Zhang, Y., Zhou, L., Rouge, L., Phillips, A.H., Lam, C., Liu, P., Sandoval, W., Helgason, E., Murray, J.M., Wertz, I.E. & Corn, J.E. (2013). Conformational stabilization of ubiquitin yields potent and selective inhibitors of USP7. *Nat. Chem. Biol.* **9**, 51–8

93. 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

94. 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

95. Dahiyat, B.I. & Mayo, S.L. (1997). De novo protein design: fully automated sequence selection. *Science* **278**, 82–7

96. Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709–13

97. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 760–3

98. Sahni, N., Yi, S., Zhong, Q., Jailkhani, N., Charloteaux, B., Cusick, M.E. & Vidal, M. (2013). Edgotype: a fundamental link between genotype and phenotype. *Curr. Opin. Genet. Dev.* **23**, 649–57

99. Vidal, M., Cusick, M.E. & Barabási, A.-L. (2011). Interactome networks and human disease. *Cell* **144**, 986–98

100. Ideker, T. & Sharan, R. (2008). Protein networks in disease. *Genome Res.* **18**, 644–52

101. Stearns, F.W. (2010). One hundred years of pleiotropy: a retrospective. *Genetics* **186**, 767–73

102. Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., Weissman, J.S. & Krogan, N.J. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–19

103. Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., Andrews, B., Tyers, M. & Boone, C. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–8

104. Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. & Rothberg, J.M. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623–7

105. Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.-A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. & Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–7

106. Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sørensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W. V, Figeys, D. & Tyers, M. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**, 180–3

107. Vidal, M., Brachmann, R.K., Fattaey, A., Harlow, E. & Boeke, J.D. (1996). Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10315–20

108. Kato, S., Han, S.-Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R. & Ishioka, C. (2003). Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8424–9

109. Dreze, M., Charloteaux, B., Milstein, S., Vidalain, P.-O., Yildirim, M.A., Zhong, Q., Svrzikapa, N., Romero, V., Laloux, G., Brasseur, R., Vandenhaute, J., Boxem, M., Cusick, M.E., Hill, D.E. & Vidal, M. (2009). "Edgetic" perturbation of a C. elegans BCL2 ortholog. *Nat. Methods* **6**, 843–9

110. Kacser, H. & Burns, J.A. (1995). The control of flux. *Biochem. Soc. Trans.* **23**, 341–66

111. Jiang, L., Mishra, P., Hietpas, R.T., Zeldovich, K.B. & Bolon, D.N. a (2013). Latent effects of Hsp90 mutants revealed at reduced expression levels. *PLoS Genet.* **9**, e1003600

112. Liberles, D.A., Teichmann, S.A., Bahar, I., Bastolla, U., Bloom, J., Bornberg-Bauer, E., Colwell, L.J., de Koning, A.P.J., Dokholyan, N. V, Echave, J., Elofsson, A., Gerloff, D.L., Goldstein, R.A., Grahnen, J.A., Holder, M.T., Lakner, C., Lartillot, N., Lovell, S.C., Naylor, G., Perica, T., Pollock, D.D., Pupko, T., Regan, L., Roger, A., Rubinstein, N., Shakhnovich, E., Sjölander, K., Sunyaev, S., Teufel, A.I., Thorne, J.L., Thornton, J.W., Weinreich, D.M. & Whelan, S. (2012). The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci.* **21**, 769–85

113. Fraser, J.S., Gross, J.D. & Krogan, N.J. (2013). From systems to structure: bridging networks and mechanism. *Mol. Cell* **49**, 222–31

114. Zhu, G., Golding, G.B. & Dean, A.M. (2005). The selective cause of an ancient adaptation. *Science* **307**, 1279–82

115. Bridgham, J.T., Carroll, S.M. & Thornton, J.W. (2006). Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101

116. Linnen, C.R., Poh, Y.-P., Peterson, B.K., Barrett, R.D.H., Larson, J.G., Jensen, J.D. & Hoekstra, H.E. (2013). Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* **339**, 1312–6

117. Dekel, E. & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–92

118. Eames, M. & Kortemme, T. (2012). Cost-benefit tradeoffs in engineered lac operons. *Science* **336**, 911–5

119. Weinreich, D.M., Delaney, N.F., Depristo, M.A. & Hartl, D.L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–4

120. Lunzer, M., Miller, S.P., Felsheim, R. & Dean, A.M. (2005). The biochemical architecture of an ancient adaptive landscape. *Science* **310**, 499–501

121. Dean, A.M., Dykhuizen, D.E. & Hartl, D.L. (1986). Fitness as a function of beta-galactosidase activity in Escherichia coli. *Genet. Res.* **48**, 1–8

122. Haas, A.L., Warms, J. V, Hershko, A. & Rose, I.A. (1982). Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. *J. Biol. Chem.* **257**, 2543–8

123. McGrath, J.P., Jentsch, S. & Varshavsky, A. (1991). UBA 1: an essential yeast gene encoding ubiquitin-activating enzyme. *EMBO J.* **10**, 227–36

124. Tokuriki, N. & Tawfik, D.S. (2009). Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604

125. Gangadhara, B.N., Laine, J.M., Kathuria, S. V, Massi, F. & Matthews, C.R. (2013). Clusters of branched aliphatic side chains serve as cores of stability in the native state of the HisF TIM barrel protein. *J. Mol. Biol.* **425**, 1065–81

126. Dill, K.A. (1990). Dominant forces in protein folding. *Biochemistry* **29**, 7133–55

127. Lee, S.Y., Pullen, L., Virgil, D.J., Castañeda, C.A., Abeykoon, D., Bolon, D.N.A. & Fushman, D. (2013). Alanine Scan of Core Positions in Ubiquitin Reveals Links between Dynamics, Stability, and Function. *J. Mol. Biol.* **426**, 1377–1389

128. Matyskiela, M.E., Lander, G.C. & Martin, A. (2013). Conformational switching of the 26S proteasome enables substrate degradation. *Nat. Struct. Mol. Biol.* **20**, 781–8

129. Phillips, A.H., Zhang, Y., Cunningham, C.N., Zhou, L., Forrest, W.F., Liu, P.S., Steffek, M., Lee, J., Tam, C., Helgason, E., Murray, J.M., Kirkpatrick, D.S., Fairbrother, W.J. & Corn, J.E. (2013). Conformational dynamics control ubiquitin-deubiquitinase interactions and influence in vivo signaling. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11379–84

130. Zhao, B., Bhuripanyo, K., Schneider, J., Zhang, K., Schindelin, H., Boone, D. & Yin, J. (2012). Specificity of the E1-E2-E3 enzymatic cascade for ubiquitin C-terminal sequences identified by phage display. *ACS Chem. Biol.* **7**, 2027–35

131. Patel, L.R., Curran, T. & Kerppola, T.K. (1994). Energy transfer analysis of Fos-Jun dimerization and DNA binding. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7360–4

132. Walden, H., Podgorski, M.S. & Schulman, B.A. (2003). Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. *Nature* **422**, 330–4

133. 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**, 4597–606

134. Peschard, P., Kozlov, G., Lin, T., Mirza, I.A., Berghuis, A.M., Lipkowitz, S., Park, M. & Gehring, K. (2007). Structural basis for ubiquitin-mediated dimerization and activation of the ubiquitin protein ligase Cbl-b. *Mol. Cell* **27**, 474–85

135. Clackson, T. & Wells, J.A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383–6

136. Hendsch, Z.S. & Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–26

137. Drummond, D.A. & Wilke, C.O. (2009). The evolutionary consequences of erroneous protein synthesis. *Nat. Rev. Genet.* **10**, 715–24

138. Khoury, G.A., Baliban, R.C. & Floudas, C.A. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **1**,

139. Kramer, E.B., Vallabhaneni, H., Mayer, L.M. & Farabaugh, P.J. (2010). A comprehensive analysis of translational missense errors in the yeast Saccharomyces cerevisiae. *RNA* **16**, 1797–808

140. Tsai, I.J., Bensasson, D., Burt, A. & Koufopanou, V. (2008). Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4957–62

141. Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–8

142. Geiler-samerotte, K.A., Dion, M.F., Budnik, B.A., Wang, S.M. & Hartl, D.L. (2010). Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. doi:10.1073/pnas.1017570108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1017570108

143. Joshi, K.K., Chen, L., Torres, N., Tournier, V. & Madura, K. (2011). A proteasome assembly defect in rpn3 mutants is associated with Rpn11 instability and increased sensitivity to stress. *J. Mol. Biol.* **410**, 383–99

144. Yao, T. & Cohen, R.E. (2002). A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419**, 403–7

145. Verma, R., Aravind, L., Oania, R., McDonald, W.H., Yates, J.R., Koonin, E. V & Deshaies, R.J. (2002). Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**, 611–5

146. Shusta, E. V, Holler, P.D., Kieke, M.C., Kranz, D.M. & Wittrup, K.D. (2000). Directed evolution of a stable scaffold for T-cell receptor engineering. *Nat. Biotechnol.* **18**, 754–9

147. Duffy, S., Tsao, K.L. & Waugh, D.S. (1998). Site-specific, enzymatic biotinylation of recombinant proteins in Spodoptera frugiperda cells using biotin acceptor peptides. *Anal. Biochem.* **262**, 122–8

148. Chen, I., Howarth, M., Lin, W. & Ting, A.Y. (2005). Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99–104

149. Li, M.Z. & Elledge, S.J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–6

150. Mumberg, D., Müller, R. & Funk, M. (1994). Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* **22**, 5767–8

151. Dhomen, N. & Marais, R. (2007). New insight into BRAF mutations in cancer. *Curr. Opin. Genet. Dev.* **17**, 31–9

152. Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., Menzies, A., Mould, C., Parker, A., Stevens, C., Watt, S., Hooper, S., Wilson, R., Jayatilake, H., Gusterson, B.A., Cooper, C., Shipley, J., Hargrave, D., Pritchard-Jones, K., Maitland, N., Chenevix-Trench, G., Riggins, G.J., Bigner, D.D., Palmieri, G., Cossu, A., Flanagan, A., Nicholson, A., Ho, J.W.C., Leung, S.Y., Yuen, S.T., Weber, B.L., Seigler, H.F., Darrow, T.L., Paterson, H., Marais, R., Marshall, C.J., Wooster, R., Stratton, M.R. & Futreal, P.A. (2002). Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–54

153. Bollag, G., Tsai, J., Zhang, J., Zhang, C., Ibrahim, P., Nolop, K. & Hirth, P. (2012). Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat. Rev. Drug Discov.* **11**, 873–86

154. Bollag, G., Hirth, P., Tsai, J., Zhang, J., Ibrahim, P.N., Cho, H., Spevak, W., Zhang, C., Zhang, Y., Habets, G., Burton, E.A., Wong, B., Tsang, G., West, B.L., Powell, B., Shellooe, R., Marimuthu, A., Nguyen, H., Zhang, K.Y.J., Artis, D.R., Schlessinger, J., Su, F., Higgins, B., Iyer, R., D'Andrea, K., Koehler, A., Stumm, M., Lin, P.S., Lee, R.J., Grippo, J., Puzanov, I., Kim, K.B., Ribas, A., McArthur, G.A., Sosman, J.A., Chapman, P.B., Flaherty, K.T., Xu, X., Nathanson, K.L. & Nolop, K. (2010). Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467**, 596–9

155. Dummer, R. & Flaherty, K.T. (2012). Resistance patterns with tyrosine kinase inhibitors in melanoma: new insights. *Curr. Opin. Oncol.* **24**, 150–4

156. Shi, H., Moriceau, G., Kong, X., Lee, M.-K., Lee, H., Koya, R.C., Ng, C., Chodon, T., Scolyer, R.A., Dahlman, K.B., Sosman, J.A., Kefford, R.F., Long, G. V, Nelson, S.F., Ribas, A. & Lo, R.S. (2012). Melanoma whole-exome sequencing identifies (V600E)B-RAF amplification-mediated acquired B-RAF inhibitor resistance. *Nat. Commun.* **3**, 724

157. Poulikakos, P.I., Persaud, Y., Janakiraman, M., Kong, X., Ng, C., Moriceau, G., Shi, H., Atefi, M., Titz, B., Gabay, M.T., Salton, M., Dahlman, K.B., Tadi, M., Wargo, J.A., Flaherty, K.T., Kelley, M.C., Misteli, T., Chapman, P.B., Sosman, J.A., Graeber, T.G., Ribas, A., Lo, R.S., Rosen, N. & Solit, D.B. (2011). RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). *Nature* **480**, 387–90

158. Tsai, J., Lee, J.T., Wang, W., Zhang, J., Cho, H., Mamo, S., Bremer, R., Gillette, S., Kong, J., Haass, N.K., Sproesser, K., Li, L., Smalley, K.S.M., Fong, D., Zhu, Y.-L., Marimuthu, A., Nguyen, H., Lam, B., Liu, J., Cheung, I., Rice, J., Suzuki, Y., Luu, C., Settachatgul, C., Shellooe, R., Cantwell, J., Kim, S.-H., Schlessinger, J., Zhang, K.Y.J., West, B.L., Powell, B., Habets, G., Zhang, C., Ibrahim, P.N., Hirth, P., Artis, D.R., Herlyn, M. & Bollag, G. (2008). Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3041–6

159. Mitsiades, N., Chew, S.A., He, B., Riechardt, A.I., Karadedou, T., Kotoula, V. & Poulaki, V. (2011). Genotype-dependent sensitivity of uveal melanoma cell lines to inhibition of B-Raf, MEK, and Akt kinases: rationale for personalized therapy. *Invest. Ophthalmol. Vis. Sci.* **52**, 7248–55

160. Whittaker, S., Kirk, R., Hayward, R., Zambon, A., Viros, A., Cantarino, N., Affolter, A., Nourry, A., Niculescu-Duvaz, D., Springer, C. & Marais, R. (2010). Gatekeeper mutations mediate resistance to BRAF-targeted therapies. *Sci. Transl. Med.* **2**, 35ra41

161. Smalley, K.S.M., Haass, N.K., Brafford, P.A., Lioni, M., Flaherty, K.T. & Herlyn, M. (2006). Multiple signaling pathways must be targeted to overcome drug resistance in cell lines derived from melanoma metastases. *Mol. Cancer Ther.* **5**, 1136–44

162. Favata, M.F., Horiuchi, K.Y., Manos, E.J., Daulerio, A.J., Stradley, D.A., Feeser, W.S., Van Dyk, D.E., Pitts, W.J., Earl, R.A., Hobbs, F., Copeland, R.A., Magolda, R.L., Scherle, P.A. & Trzaskos, J.M. (1998). Identification of a novel inhibitor of mitogen-activated protein kinase kinase. *J. Biol. Chem.* **273**, 18623–32

163. Pratilas, C.A., Taylor, B.S., Ye, Q., Viale, A., Sander, C., Solit, D.B. & Rosen, N. (2009). (V600E)BRAF is associated with disabled feedback inhibition of RAF-MEK signaling and elevated transcriptional output of the pathway. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4519–24

164. Warmuth, M., Kim, S., Gu, X., Xia, G. & Adrián, F. (2007). Ba/F3 cells and their use in kinase drug discovery. *Curr. Opin. Oncol.* **19**, 55–60

165. Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., Teague, J.W., Campbell, P.J., Stratton, M.R. & Futreal, P.A. (2011). COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–50

166. Barbieri, C.E., Baca, S.C., Lawrence, M.S., Demichelis, F., Blattner, M., Theurillat, J.-P., White, T.A., Stojanov, P., Van Allen, E., Stransky, N., Nickerson, E., Chae, S.-S., Boysen, G., Auclair, D., Onofrio, R.C., Park, K., Kitabayashi, N., MacDonald, T.Y., Sheikh, K., Vuong, T., Guiducci, C., Cibulskis, K., Sivachenko, A., Carter, S.L., Saksena, G., Voet, D., Hussain, W.M., Ramos, A.H., Winckler, W., Redman, M.C., Ardlie, K., Tewari, A.K., Mosquera, J.M., Rupp, N., Wild, P.J., Moch, H., Morrissey, C., Nelson, P.S., Kantoff, P.W., Gabriel, S.B., Golub, T.R., Meyerson, M., Lander, E.S., Getz, G., Rubin, M.A. & Garraway, L.A. (2012). Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–9

167. King, A.J., Patrick, D.R., Batorsky, R.S., Ho, M.L., Do, H.T., Zhang, S.Y., Kumar, R., Rusnak, D.W., Takle, A.K., Wilson, D.M., Hugger, E., Wang, L., Karreth, F., Lougheed, J.C., Lee, J., Chau, D., Stout, T.J., May, E.W., Rominger, C.M., Schaber, M.D., Luo, L., Lakdawala, A.S., Adams, J.L., Contractor, R.G., Smalley, K.S.M., Herlyn, M., Morrissey, M.M., Tuveson, D.A. & Huang, P.S. (2006).

Demonstration of a genetic therapeutic index for tumors expressing oncogenic BRAF by the kinase inhibitor SB-590885. *Cancer Res.* **66**, 11100–5

168. Hansen, J.D., Grina, J., Newhouse, B., Welch, M., Topalov, G., Littman, N., Callejo, M., Gloor, S., Martinson, M., Laird, E., Brandhuber, B.J., Vigers, G., Morales, T., Woessner, R., Randolph, N., Lyssikatos, J. & Olivero, A. (2008). Potent and selective pyrazole-based inhibitors of B-Raf kinase. *Bioorg. Med. Chem. Lett.* **18**, 4692–5

169. Roscoe, B.P., Thayer, K.M., Zeldovich, K.B., Fushman, D. & Bolon, D.N.A. (2013). Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J. Mol. Biol.* **425**, 1363–77

170. Hietpas, R.T., Bank, C., Jensen, J.D. & Bolon, D.N.A. (2013). Shifting fitness landscapes in response to altered environments. *Evolution* **67**, 3512–22

171. Finley, D., Ozkaynak, E. & Varshavsky, A. (1987). The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses. *Cell* **48**, 1035–46

172. Kee, Y., Muñoz, W., Lyon, N. & Huibregtse, J.M. (2006). The deubiquitinating enzyme Ubp2 modulates Rsp5-dependent Lys63-linked polyubiquitin conjugates in Saccharomyces cerevisiae. *J. Biol. Chem.* **281**, 36724–31

173. DePristo, M.A., Weinreich, D.M. & Hartl, D.L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **6**, 678–87