Supporting Information

Starita et al. 10.1073/pnas.1303309110

SI Materials and Methods

Plasmids and Oligonucleotides. Oligonucleotide sequences are provided in Table S3. Ubiquitination factor E4B (Ube4b) U-box (E4BU, 1,092–1,173) and extended U-box domain (longE4BU, 1,072–1,173) in the pBG102 bacterial expression vector were gifts from W. Chazin (Vanderbilt University, Nashville, TN). The pBG102 construct adds an amino terminal His6-small ubiquitin-related modifier (SUMO)-H3C protease cleavage site to aid in protein solubility. Mutations were created using the Quickchange II mutagenesis kit (Agilent) and primers LS571–574, LS757–758, and LS793–806. The Flag-GFP expression vector was a gift from R. James (Seattle Children's Hospital, Seattle, WA). pcDNA3_Flag-Ube4b and pcDNA3_Flag-Ube4bΔU-box were gifts from R. Leng (University of Alberta, Edmonton, AB, Canada). A cDNA clone harboring human UBE4B variant 2 (NM_006048.4) was purchased from Open Biosystems, PCR-amplified, and cloned into the pcDNA3 vector with an amino-terminal Flag tag. Mutants were created using primers LS757–758, LS768–769, and LS837–844. Mutation P1140T (dsSNP identifier rs15191) was identified in the genome of J. Craig Venter. The full-length ORFs were verified by sequencing. Phosphorylated CMV (pCMV)-neo-p53 and pCMV-Myc3-human double minute 2 (Hdm2) were purchased from Addgene.

T7 Bacteriophage Constructs. The E4BU (1,092–1,173) and longE4BU (1,072–1,173) were PCR-amplified with primers LS569–570 and LS788 and cloned into the genome of T7-Select 10-3b bacteriophage (EMD Millipore), packaged into phage particles, amplified in *Escherichia coli* strain BLT5403, and titered according to the T7-Select Cloning Kit instructions (EMD Millipore). The T7_H3C_HA_longE4BU and E4BU phages were constructed similarly using primer LS789–792.

T7 LongE4BU Variant Library Preparation. Oligonucleotides for cloning the E4BU library were ordered from TriLink Biotechnologies, and all other oligonucleotides were ordered from IDT. The 306 nucleotides that encoded the variable region were synthesized with a 2% error rate, corresponding to 99.325% of the correct base and 0.225% of each incorrect base. Stop codons were inserted in all three frames 3′ of the coding region, followed by the degenerate 18-base barcode. These DNA features were spread over two oligonucleotides, LS714 and LS715. The 3′ oligonucleotide was phosphorylated. One hundred picomoles of each oligonucleotide was annealed and ligated together in the presence of 1 nmol of a guide oligonucleotide (LS699) that was terminated with a dideoxynucleoside to prevent its extension. The oligonucleotides in T4 DNA ligase buffer were heated to 90 °C in a thermocycler set to reduce the temperature 0.1 °C/min until it reached room temperature. T4 DNA ligase (New England Biolabs) and 1 mM DTT were added to the annealed oligonucleotides and incubated overnight at 16 °C. The ligation reaction was purified using the DNA Clean and Concentrator Kit (Zymo Research). To make the full-length construct double-stranded, 200 pmol of short oligonucleotide (LS713) was annealed to a common sequence 3′ of the barcode as above but in phi29 polymerase buffer (New England Biolabs). After annealing and addition of dNTPs, the short oligonucleotide was extended with phi29 polymerase (New England Biolabs) for 3 h at 30 °C to make a dsDNA construct. The double-stranded product was purified, cut with EcoRI and HindIII (New England Biolabs), and ligated into the genome of T7-Select 10-3b bacteriophage. Phage genomic DNA was packaged into phage particles, the number of ligation/packaging events was estimated by titer, and the phages were then amplified in *E. coli* strain BLT5403 according to the T7-Select Cloning Kit instructions.

According to plaque titering assays, 5 million phages were packaged following the ligation and packaging reactions. We PCR-amplified and Sanger-sequenced a small number of individual plaques; of those, 10% were empty phages and nearly 50% had deletions stemming from the use of long oligonucleotides to make the construct. We estimate that there were 2 million ligation events that encoded a uniquely barcoded, full-length longE4BU protein.

Plaque PCR and Sequencing. For bulk population analysis of mixtures of two phage populations, as in Fig. S1*C*, 1 μL of phage lysate was used in a PCR assay to amplify the longE4BU inserts from the phage genome with primers LS826–827. The PCR product was sequenced, and the ratio of the mixed bases coding for the mutant position was quantified using the PeakPicker algorithm (1). To assess the diversity of the population for the deep mutational scanning, the longE4BU insert was PCR-amplified from individual plaques. Agarose containing the plaque was picked with a pipette tip and mixed with 25 μL of LB. One microliter of the mixture was used to amplify the longE4BU insert with Taq polymerase. The amplicons were purified and Sanger-sequenced with primer LS826.

Library Preparation for Tag-Directed Subassembly and Barcode Counting. DNA was purified from 200 μL of amplified phage lysates before and after selection by phenol chloroform extraction and ethanol precipitation. To create the overlapping Illumina libraries for tag-directed subassembly as described in Fig. S2*B*, 800 ng of total DNA from the input sample was used as a template for PCR. Two hundred-microliter PCR reactions were performed over 12 cycles using Phusion polymerase (New England Biolabs) and oligonucleotides LS725, LS727–729, and LS749. The amplicons were purified with Qiaquick PCR purification columns (Qiagen) and again with the DNA Clean and Concentrate Kit (Zymo Research) and quantified using Qbit (Invitrogen). Each PCR-generated Illumina library was subjected to paired-end 101-base reads in separate lanes on a HiSeq2000 (Illumina) using primers LS730–733 for read 1 and primer LS724 for read 3. To create Illumina libraries for sequencing of the 18-base barcodes, 800 ng of total phage DNA was used as a template in 200 μL of Phusion polymerase (New England Biolabs) PCR reactions over 12 cycles using oligonucleotides LS749 and LS737–741. The amplicons were purified and quantified as above. Samples were indexed by addition of an 8-base sequence on the reverse primer LS737–741 and subject to single-end 36-base reads with an 8-base index read using primers LS724 and LS748 on an Illumina Genome AnalyzerIIx.

Tag-Directed Subassembly. The reads and barcodes from each nested library were filtered for quality. Reads were discarded if they contained random "N" bases, more than 9 Q2 bases, or homopolymeric runs of more than 7 bases in the read. Barcodes (the first 18 bases of the reverse read) were discarded if they contained Ns, more than 4 Q2 bases, or homopolymeric runs of more than 7 bases in the read (Table S1). Filtered fastq files generated by sequencing from the ends of the overlapping amplicons described above were combined. Forward reads were grouped by the sequence of the 18-base barcode (the first 18 bases of the reverse read). A Smith–Waterman algorithm was used to align the grouped reads to the WT longE4BU sequence, and a consensus sequence was determined for each barcode group as in the studies by Hiatt et al. (2) and Patwardhan et al. (3). A minimum quality score

of 20 was required for a base to contribute to an assembly. Full-length subassembled longE4BU sequences were filtered for quality by requiring that a given base in the assembly was observed at least twice and was present at an intratag group allele frequency of 1 for positions covered by two or three reads or at a frequency of at least 0.75 for positions covered by four or more reads. If these conditions were not met, the assembly was discarded. A database of barcodes and their associated full-length longE4BU assembly was created to facilitate linking barcodes sequenced from the experimental samples to the full-length subassemblies.

Barcode Sequencing and Conversion to LongE4BU DNA Sequences. Barcodes from the input variant library and after round 3 of selection were PCR-amplified and sequenced by 36-base single-end reads. The reads were trimmed to 18 bases and reverse-complemented to match the polarity of the barcode reads from the subassembly. Barcode reads were discarded if they contained Ns, more than 4 Q2 bases, or homopolymeric runs of more than 7 bases. For every barcode sequence from each sample, we attempted to match it to a barcode and subassembly from the database created above. The barcode matching attempt led to three possible outcomes: (*i*) the barcode did not match any of the barcodes associated with a subassembly and was discarded, (*ii*) the barcode matched an incomplete subassembly and was discarded, or (*iii*) the barcode matched a full-length subassembly and was converted to the longE4BU sequence (Table S2).

Expanded Protein Purification. E4BU or longE4BU in pBG102 expression plasmids was transformed into BL21 (DE3) chemically competent cells (Agilent). pBG102 expression plasmids create a translational fusion of 6× His-SUMO-E4BU. For longE4BU purifications, 500 mL of midlog LB + kanamycin cultures were induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside and grown overnight at 18 °C. Cultures were cooled on ice and spun at 4,000 × *g* for 20 min. Pellets were resuspended in 20 mL of 25 mM Tris, 200 mM NaCl, and 5 mM imidazole (pH 7.6), and were then lysed with 1.2 mL of 20 mg/mL lysozyme and small scoops of powdered RNaseA and DNaseI followed by incubation on ice for 30 min. Lysates were sonicated and cleared by spinning at 12,000 × *g* for 20 min. The E4BU (1,092–1,173) mutants used for CD, NMR, and lysine reactivity assays were grown in 1-L cultures induced as above; however, they were lysed using a French press. His-tag purifications were carried out by incubating cell lysate with 1 mL Talon Agarose Beads (Clontech) for 1 h at 4 °C. Beads were washed two times with 20 mL of binding buffer [25 mM Tris, 200 mM NaCl, 5 mM imidazole (pH 7.6)] and eluted with 25 mM Tris, 200 mM NaCl, and 500 mM imidazole (pH 7.6). Talon column elutions were cleaved overnight with GST-tagged H3C protease while being dialyzed into 25 mM Tris and 200 mM NaCl (pH 7.6). The protein mix was subjected to a GST capture using 1 mL of GE Glutathione Sepharose 4 Fast Flow beads at 4 °C for 45 min to collect the GST-tagged H3C protease. The flow-through was subjected to a His capture with 1 mL of Talon Agarose Beads at 4 °C for 1 h to bind 6× His-SUMO, and the flow-through containing purified E4BU was collected. Samples were concentrated using Amicon Ultra-15 3,000 molecular weight cutoff spin tubes (Millipore). E4BU proteins underwent a final size exclusion chromatography step over an SDX75 column into 25 mM NaPi and 150 mM NaCl at pH 7.0 (for NMR and lysine reactivity assays) or 10 mM NaPi and 30 mM NaCl at pH 7.0 (for CD). Protein purity was assessed as >95% by SDS/PAGE followed by Coomassie staining. Quantification of protein concentration was determined using the extinction coefficient for both E4BU constructs of 6.99 $mM^{-1} \cdot cm^{-1}$.

Lysine Reactivity Assay. UbcH5c∼Ub(K0) conjugate was preformed by incubation of E1, E2, Ub, and ATP/Mg for 30 min. The "K0" Ub construct was HA-tagged and contained arginine or methionine mutations at all lysine positions to prevent Ub chain formation. The conjugate was combined with E4BU (WT or mutant) and free lysine. Reaction samples were quenched in nonreducing SDS sample buffer at the specified time points. Ub transfer to lysine was monitored by disappearance of the UbcH5c∼Ub species and appearance of lysine-linked Ub, as observed in Western blots probed for the HA epitope on Ub.

CD. Far-UV CD spectra were recorded on an Aviv model 420 spectrometer at 25 °C in a 1-mm path-length quartz cell. For each E4BU variant, 50 μM samples were prepared in 10 mM sodium phosphate (pH 7.0) and 30 mM NaCl. Signals were recorded with 5-s averaging times, and thermal melts were performed in increments of 2°.

1. Ge B, et al. (2005) Survey of allelic expression using EST mining. *Genome Res* 15(11): 1584–1591.
2. Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J (2010) Parallel, tag-directed assembly of locally derived short sequence reads. *Nat Methods* 7(2):119–122.
3. Patwardhan RP, et al. (2012) Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol* 30(3):265–270.

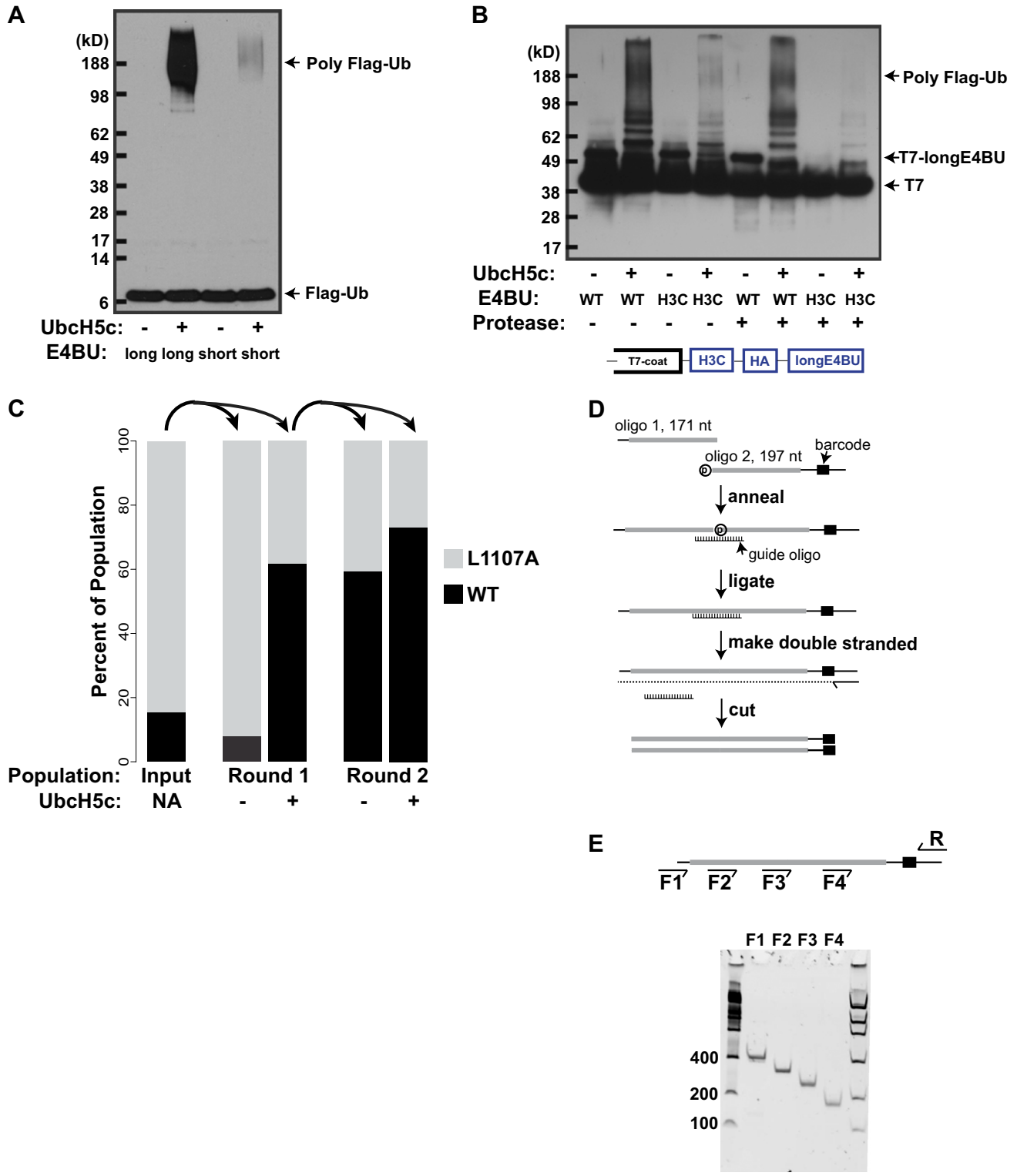
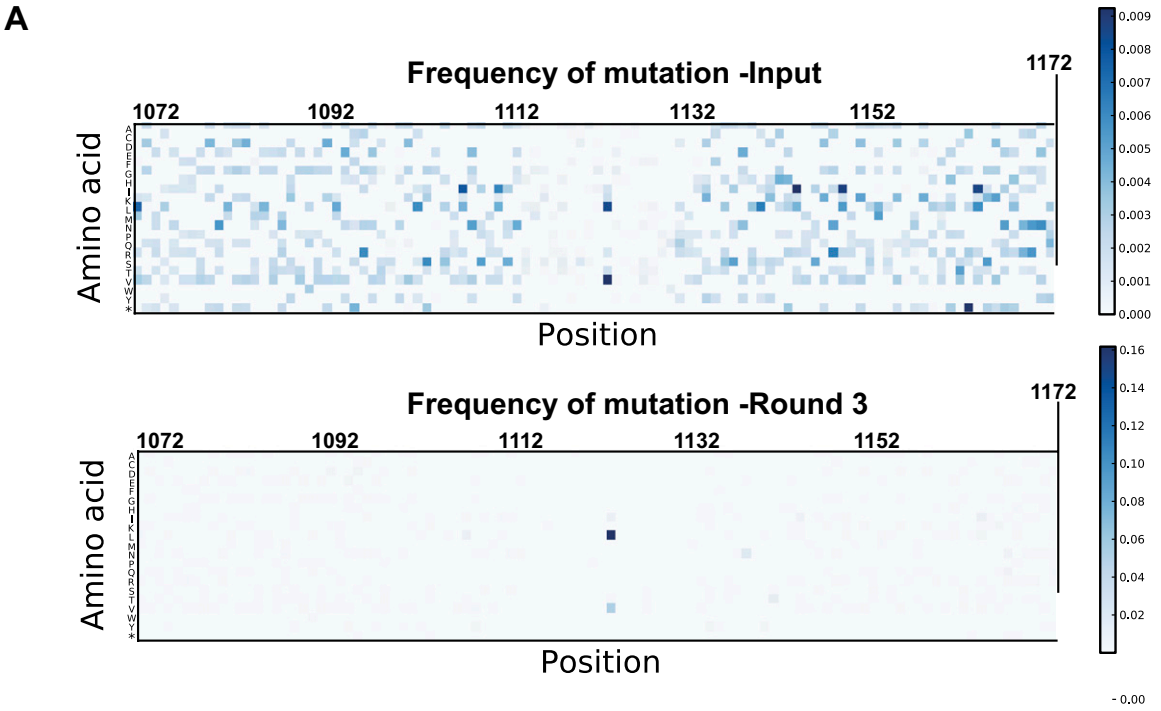


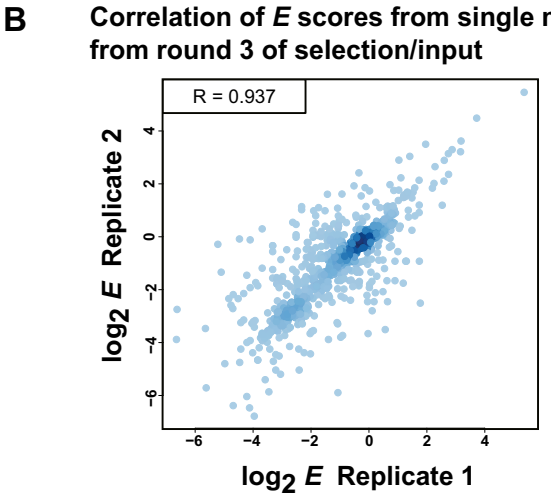
Fig. S1. (*A* and *B*) Evidence for auto-ubiquitination by T7-E4BU. (*A*) Minimal U-box of E4BU does not support robust Ub polymerization. Either longE4BU (1,072–1,173) or minimal E4BU (1,092–1,173) was fused to the coat protein of T7 bacteriophage. The minimal E4BU is lacking four lysine residues that are the presumed to be auto-ubiquitination targets (1). Amplified phage lysate was incubated with purified E1, UbcH5c, ATP, and Flag-Ub. The reaction was resolved by SDS/PAGE, transferred to a PVDF membrane, and probed with anti-Flag antibodies. (*B*) Protease cleavage between the U-box and T7 coat protein reduces high-molecular-weight species. The longE4BU with an HA tag and an H3C protease cleavage site was fused to T7 coat protein as diagrammed. Amplified WT or H3C-HA-E4BU phage lysates were incubated with purified E1, UbcH5c, ATP, and Flag-Ub. Following the ubiquitination reaction, the mixture was incubated with GST-H3C protease. The reaction was resolved by SDS/PAGE, transferred to a PVDF membrane, and probed with monoclonal anti-T7 antibodies. E4BU and H3C cleavage is as indicated. (*C*) WT T7-E4BU is enriched after selection for ubiquitinated T7 bacteriophage. For round 1 of selection, amplified phage lysates from WT and the L1107A variant of T7-longE4BU were mixed at a 1:5 ratio (Input) and incubated with purified E1, ATP, and Flag-Ub with or without UbcH5c as indicated. Flag-ubiquitinated phages were purified with anti-Flag agarose and eluted with Flag peptide. Eluates were reamplified in *E. coli*. The longE4BU

Legend continued on following page

inserts were amplified by PCR from the phage lysate and Sanger-sequenced. The ratio of the mixed bases coding for the mutant position was quantified using the PeakPicker algorithm and is represented in the bar graphs (2). The eluate from the ubiquitination reaction with UbcH5c from round 1 of selection was used as the input for round 2 of selection, as indicated by the arrows. NA, not added. (*D*) Diagram of T7-E4BU variant cloning. Because the longest oligonucleotide we were able to purchase was 200 bases, we spread features of the library across two oligonucleotides whose lengths are indicated. The longE4BU is represented as gray boxes and was synthesized with two mutations per 306 nucleotides encoding the 102 amino acids of longE4BU. Oligo 2 was synthesized with a 5′ phosphate. Oligo 2 also codes for a stop codon in all three frames 3′ of the E4BU ORF, an 18-base degenerate barcode (black box), and a common sequence. The oligos were annealed to a guide oligo synthesized with a 3′ dideoxynucleotide to prevent polymerase extension and ligated together with T4 DNA ligase. This construct was made double-stranded by annealing a short oligo to a common sequence 3′ of the barcode and extending that oligo with the strand displacing polymerase phi29. The double-stranded construct was cut with EcoRI and HindIII and ligated into T7-Select 10-3b bacteriophage arms in the purchased T7-Select Cloning Kit. (*E*) Nested amplicons for tag-directed subassembly. Oligonucleotides for amplification of nested amplicons are represented by arrows. The 101-base sequences of the forward read of each amplicon were designed to overlap by 10 nucleotides. The first 18 bases of the reverse (R)-sequencing read comprise the barcode. Each amplicon was separated on a 10% Tris/borate/EDTA gel and stained with SYBR gold. Sequencing and subassembly statistics are provided in Table S1.

1. Nordquist KA, et al. (2010) Structural and functional characterization of the monomeric U-box domain from E4B. *Biochemistry* 49(2):347–355.
2. Ge B, et al. (2005) Survey of allelic expression using EST mining. *Genome Res* 15(11):1584–1591.





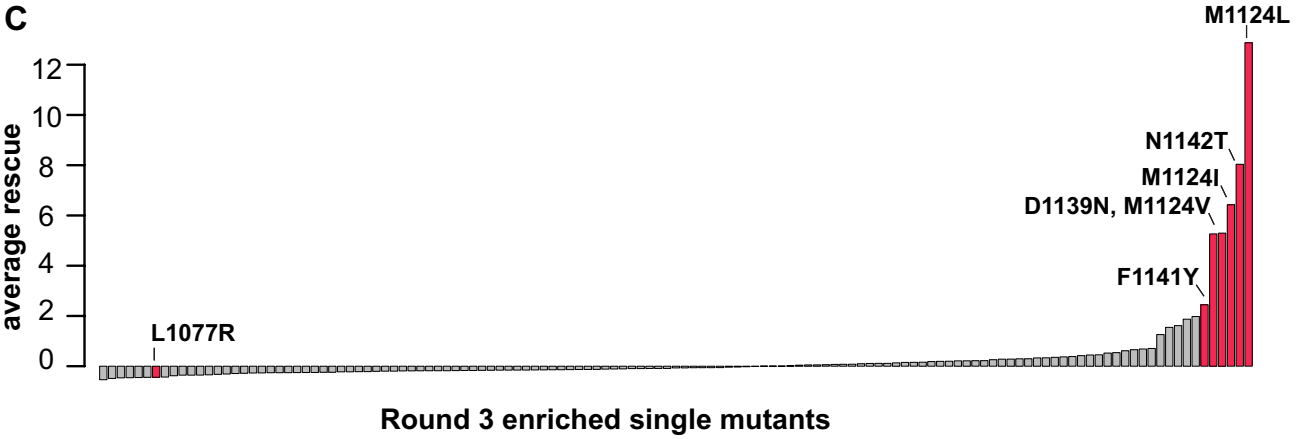


Fig. S2. (*A*) Mutations within longE4BU are distributed throughout the protein. The frequency of occurrence of each amino acid substitution in the input (*Upper*) and round 3 (*Lower*) populations is plotted as a heat map with the position of the longE4BU across the horizontal axis and each possible amino acid change on the vertical axis. Scales are shown to the right. Note that the mutations in the input library are not completely random. (*i*) Mutations are depleted

Legend continued on following page

between amino acid positions 42 and 60. We believe that this occurred because the nucleotides that code for these positions must anneal to the guide oligonucleotide used in the ligation step (as shown in *A*). Some mutations in this region may have caused inefficient annealing and ligation and would therefore be depleted in the library. (*ii*) As expected from random mutagenesis, changes to amino acids that required more than one nucleotide change are less common. For example, an amino acid change to W, which is encoded by only a single codon, occurred only once throughout the whole domain. (*B*) *E* scores for technical replicates of the longE4BU deep mutational scan are highly concordant. Correlation of *E* scores of variants containing a single mutation from two experimental replicates of selection are shown. A Pearson correlation coefficient of 0.93 was calculated for the two datasets. (*C*) Mutations found in highly enriched variants can rescue most other mutations. To assess the functional importance of enriched mutations, we evaluated their ability to rescue detrimental mutations when the two are paired. To this end, we calculated an average rescue score ($\overline{R}$) for variants with $E > 1$ by the following steps. To calculate $\overline{R}$ for a mutation X (where $E_X > 1$), we identified all double mutants that contain X and another mutation Y. We excluded single mutations found in fewer than 50 doubly mutated variants. Next, we subtracted E_Y from $E_{X,Y}$ and averaged all these remainders to determine $\overline{R}$. For 120 of the 132 mutations for which we could calculate an average rescue score, $\overline{R}$ was near zero, indicating that they did not rescue second site mutations because, on average, $E_{X,Y}$ and E_Y were similar. However, when, on average, $E_{X,Y} >> E_Y$, we suggest that E_X was able to rescue E_Y when paired. The average $\overline{R}$s for variants with an *E* score >1 that were observed in combination with ≥50 secondary mutations are shown. Select mutations are indicated.

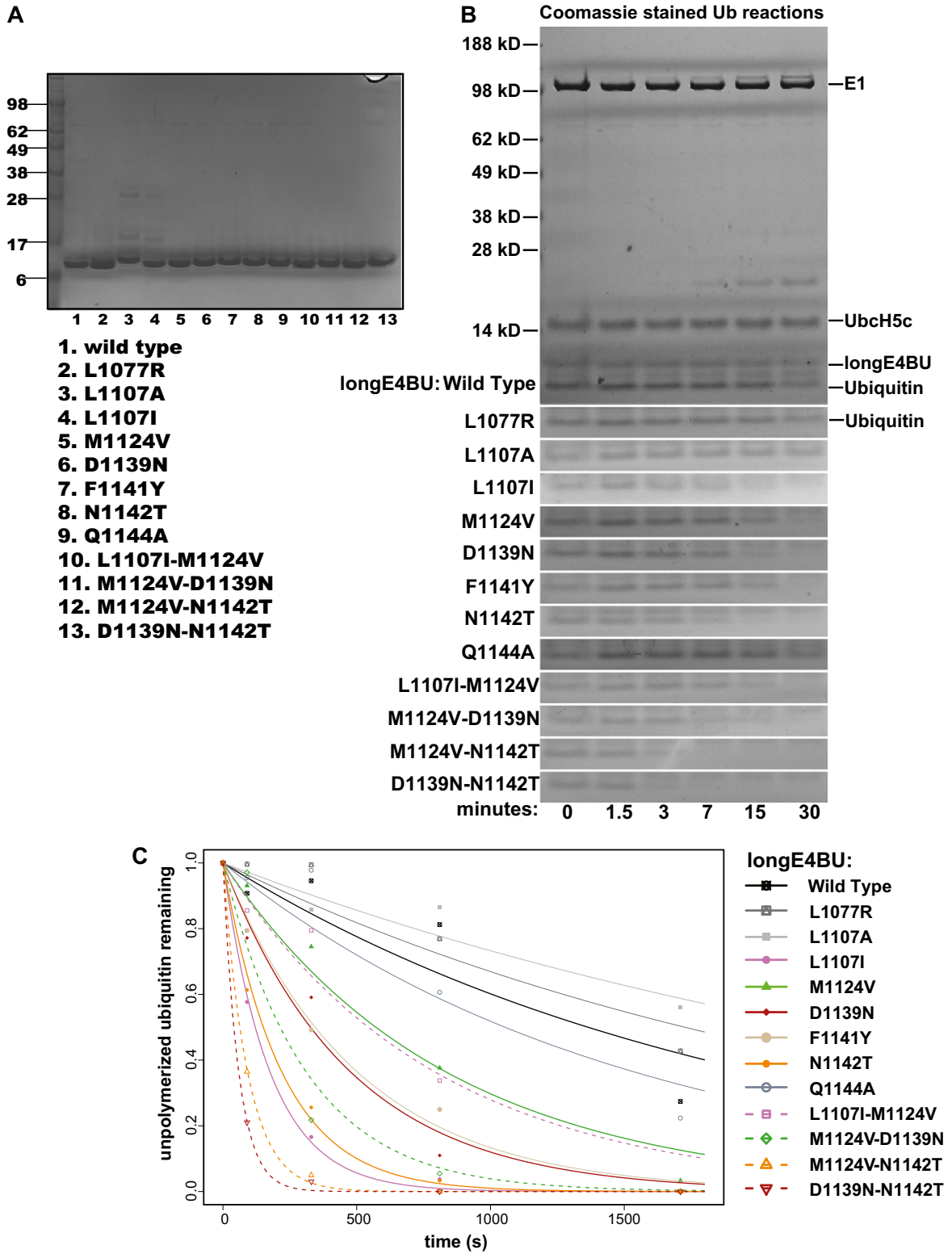


Fig. S3. (*A*) Equal concentrations of longE4BU variants were used in in vitro ubiquitination assays. Two hundred picomoles of each longE4BU variant was electrophoresed on a 4–12% gel and visualized with Coomassie stain. Variants are as indicated. (*B* and *C*) Ub bands from Coomassie-stained SDS/PAGE gels were used to quantify the rate of Ub disappearance. (*B*) Ubiquitination assays. The purified variants of longE4BU were incubated with recombinant E1, UbcH5c, ATP/Mg, and Flag-Ub at 37 °C for the indicated time. Ubiquitination products were separated by SDS/PAGE and visualized by Coomassie staining. The full-size gel for the WT longE4BU reaction is shown, along with only the Ub band for all the indicated variants. The intensities of these bands were quantified by ImageJ (National Institutes of Health) to estimate the rate of disappearance of unmodified Ub. (*C*) Decay curves for the loss of the unmodified Ub band. Intensities from the unmodified Ub bands were normalized to the 1.5-min time point due to the fuzzy nature of the T = 0 Ub bands. The normalized band intensities were used to estimate the rate (k) at which the Ub was lost by fitting the data to a first-order rate law equation $[A] = [A_o]e^{kt}$.

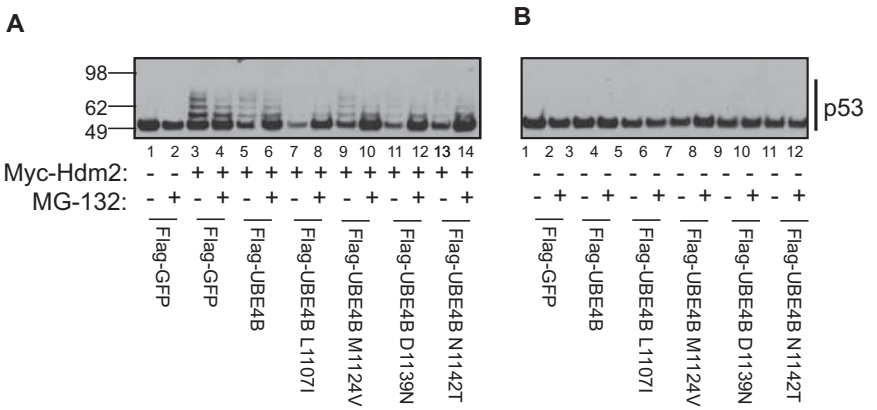


Fig. S4. UBE4B-induced degradation of p53 is proteasome- and Hdm2-dependent. (*A*) H1299 cells were transfected with phosphorylated CMV (pCMV)-neo-p53, and the cells in lanes 3–14 were also transfected with pCMV-Myc3-Hdm2. Flag-tagged, full-length human UBE4B constructs were transfected as indicated. Twenty-five micromolar MG132 or DMSO was added 5 h before cell lysis. Protein concentration was quantified by Bradford assay, and 10 μg of protein was separated by SDS/PAGE and transferred to a PVDF membrane. Blots were probed for p53. (*B*) H1299 cells were transfected with pCMV-neo-p53 and UBE4B in the absence of Hdm2. Twenty-five micromolar MG132 or DMSO was added 5 h before cell lysis. Protein concentration was quantified by Bradford assay, and 10 μg of protein was separated by SDS/PAGE and transferred to a PVDF membrane. Blots were probed for p53.

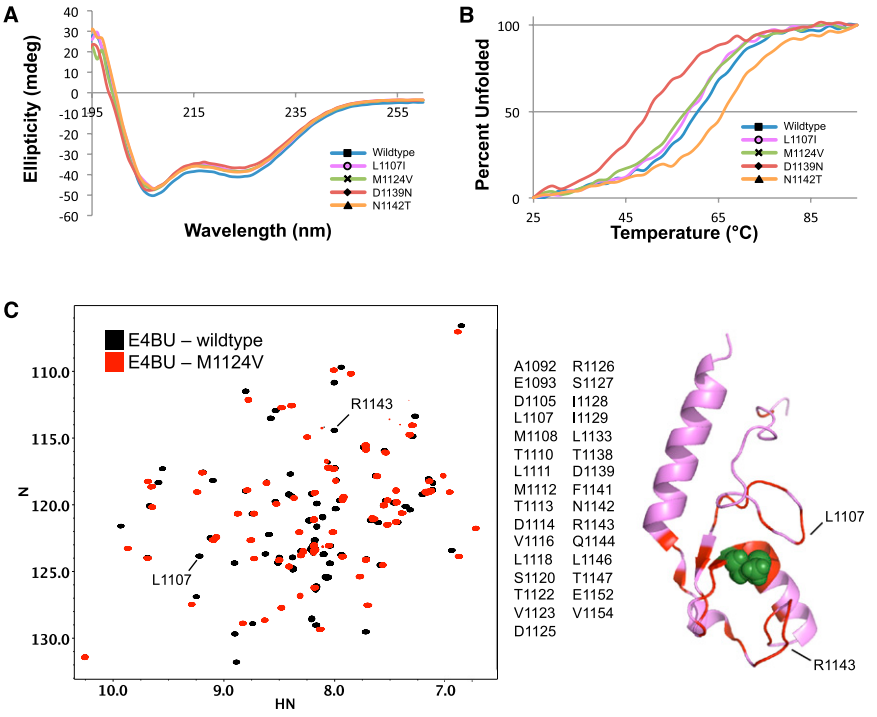


Fig. S5. Variants with activity-enhancing mutations do not have changes in their secondary structure nor are they significantly more thermodynamically stable. (*A*) CD wavelength scans from 195 to 260 nm are shown for 50 μM samples of E4BU WT (blue), L1107I (magenta), M1124V (green), D1139N (red), and N1142T (orange). mdeg, millidegrees. (*B*) Thermal melts following CD signal at 220 nm were used to determine melting temperature (Tm) values for E4BU WT (61 °C), L1107I (59 °C), M1124V (57 °C), D1139N (50 °C), and N1142T (66 °C). The color scheme is the same as in *A*. All ellipticity data were normalized to 0% and 100% unfolded using the first and last data points, respectively. (*C*) ^{1}H, ^{15}N heteronuclear single quantum coherence (HSQC) transverse relaxation-optimized spectroscopy (TROSY) overlay of WT (black) and M1124V E4BU (red). E4BU M1124V resonances that have shifted off the peak in the WT spectrum are listed and mapped onto the E4BU structure in red. The M1124 position is shown in green spheres. Assignments of select resonances that show large perturbations are labeled, and their positions in the structure are marked.

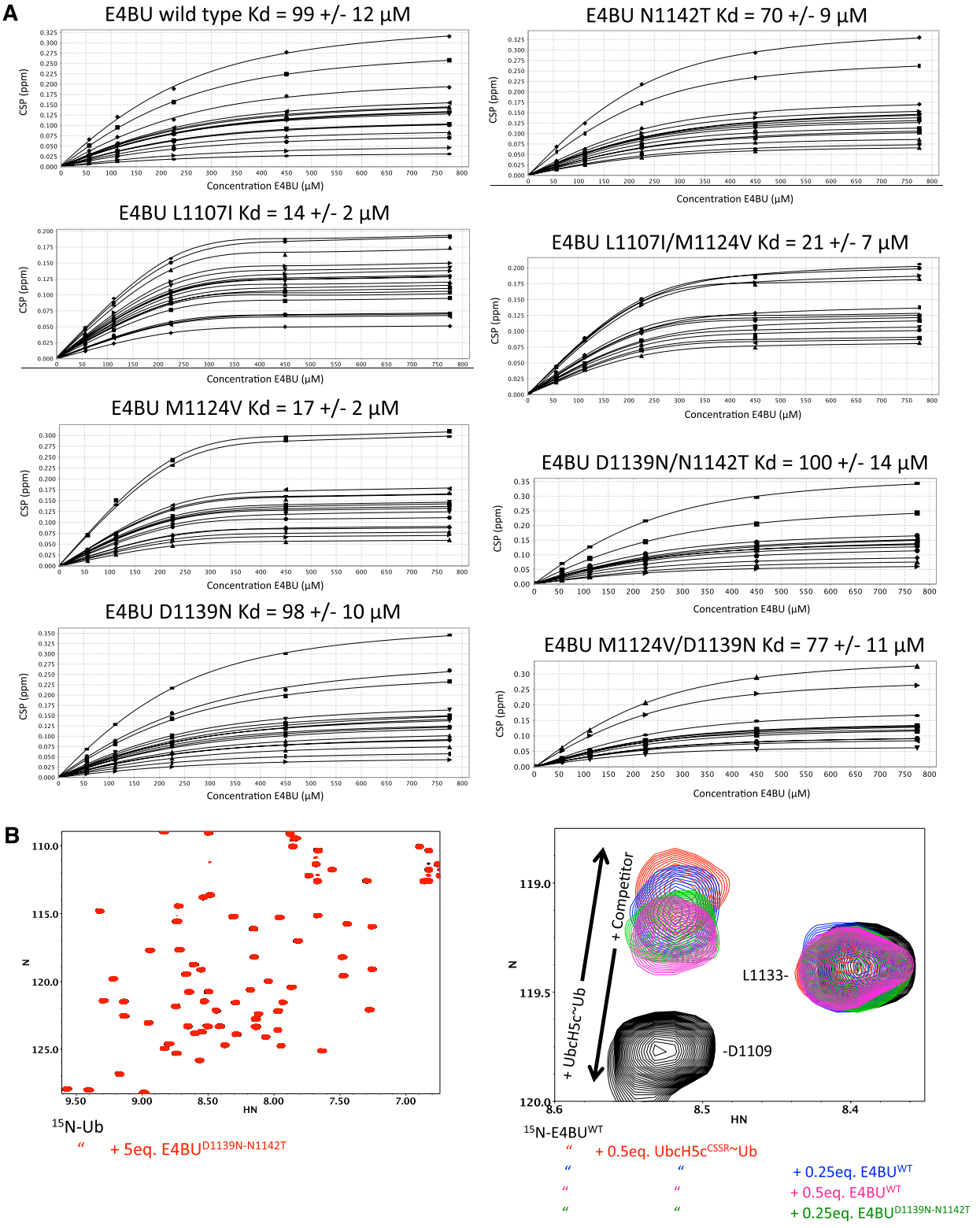


Fig. S6. Individual binding curves for titrations of E4BU variants into 225 µM ^{15}N-labeled UbcH5c. (*A*) NMR View was used to fit the E4BU:UbcH5c-binding data with the quadratic equation f = A + (C – A) * (((p + x + 10^B) – ((p + x + 10^B)^2–4 * p * x)$^{0.5}$))/(2 * p), where p is the concentration of UbcH5c. A minimum of 10 representative titrating peaks were used to determine a K_d, and the error is reported as the average SD from the quadratic fits. Variants as indicated. (*B*) (*Left*) Comparison of HSQC spectra collected on ^{15}N-Ub in isolation (black) and in the presence of 5 mol eq of E4BU$^{D1139N/N1142T}$ (red) showed no observable interaction. (*Right*) Because the UbcH5c~Ub conjugate undergoes appreciable hydrolysis in the presence of E4BU, competition experiments were used in place of traditional NMR titrations to compare the binding of WT and D1139N/N1142T E4BU. Select resonances of ^{15}N-E4BU^{WT} (black) undergo large chemical shift perturbations (CSPs) on addition of 0.5 mol eq of UbcH5c^{CSSR}~Ub (red). Further addition of 0.25 (blue) or 0.5 (magenta) mol eq of unlabeled E4BU^{WT} results in a backward titration because the ^{15}N-labeled and natural abundance E4BU competes for UbcH5c~Ub binding. Addition of 0.25 mol eq of unlabeled E4BU$^{D1139N/N1142T}$ resulted in levels of competition similar to WT, indicating comparable binding affinities. Fresh samples were prepared for each of the HSQC spectra.

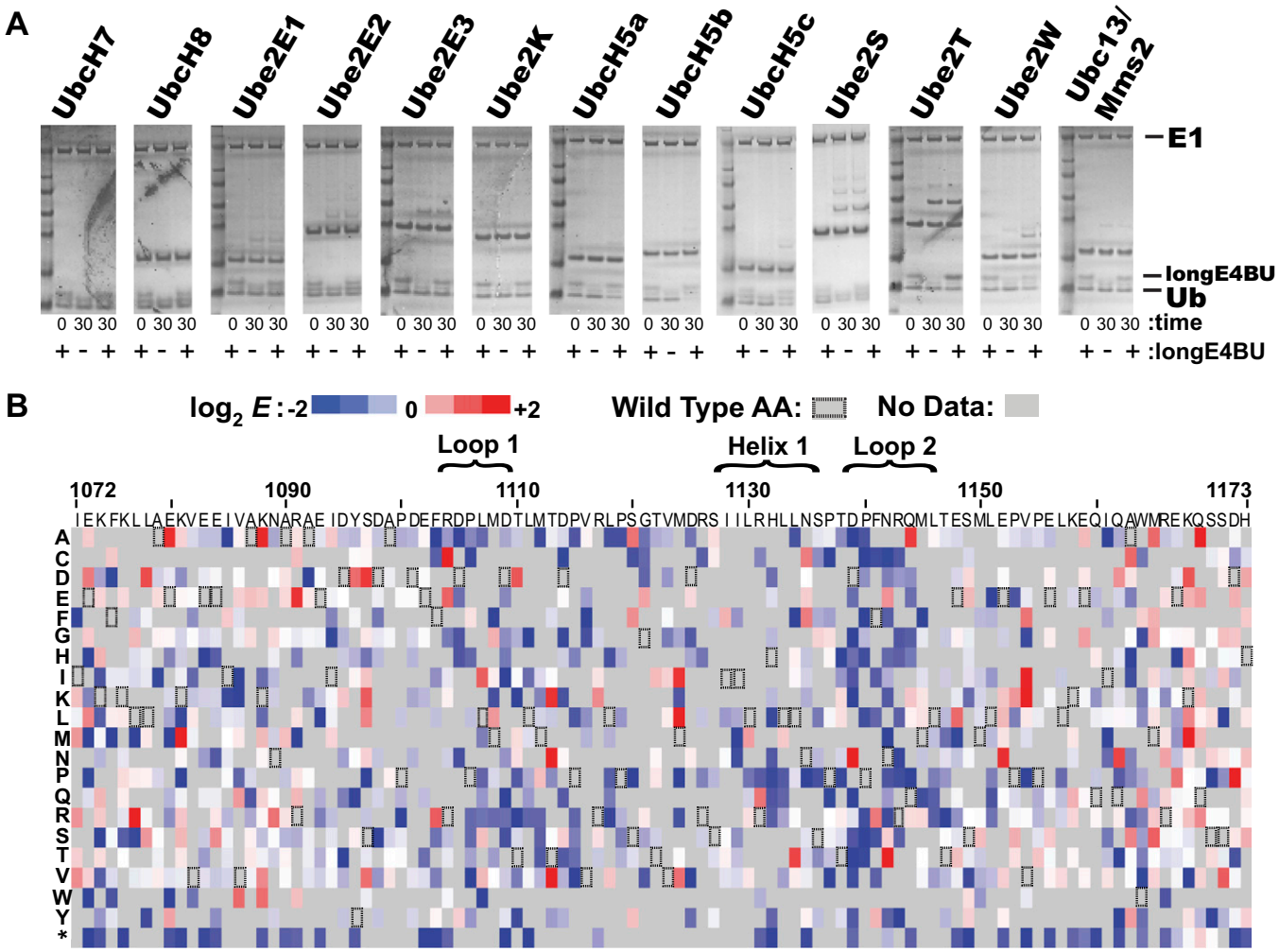


Fig. S7. (*A*) LongE4BU has auto-ubiquitination activity when paired with the UbcH5 family and Ube2w. We tested a panel of E2s to determine if they supported auto-ubiquitination by longE4BU. The longE4BU was incubated with ATP, purified E1, Ub, and the indicated E2 for 0 or 30 min at 37 °C. Reactions were stopped by addition of sample buffer containing β-mercaptoethanol, separated by SDS/PAGE, and stained with Coomassie blue. (*B*) Effects of mutations on longE4BU function when paired with Ube2w are uncovered by deep mutational scanning of longE4BU. A sequence–function map of $\log_2 E$ scores for T7-longE4BU variants with a single amino acid change is shown. Blue, white, and red boxes represent T7-longE4BU variants that were depleted, neutral, or enriched, respectively, during the selection process. Residues that make up loop 1, loop 2, and helix 1 are indicated. The longE4BU sequence is represented on the *x* axis, and the possible amino acid substitutions are represented on the *y* axis.

Table S1. Summary of Illumina sequencing reads for the subassembly of longE4BU DNA sequences

Illumina library	High-quality reads and barcodes	No. of barcode groups	No. of full-length subassemblies	No. of subassemblies passing quality filter
F1	20,719,043			
F2	14,141,828			
F3	17,463,371			
F4	23,150,873			
Total	75,475,115	2,281,364	811,480	573,797

Table S2. Summary of barcode sequencing and conversion to subassembled and quality filtered longE4BU sequences

Sample	Input	Replicate 1	Replicate 2
Round of selection	NA	3	3
Total barcodes*	24.44	19.051	17.804
Total no assembly*	15.271	7.411	6.956
Unique no assembly*	2.327	1.524	1.32
Total Incomplete assembly*	0.167	0.046	0.038
Unique incomplete assembly*	0.01	0.007	0.005
Total assembly match*	9.001	11.594	10.811
Unique assembly match*	0.563	0.49	0.448
Assembly match, %	36.8	60.9	60.7

NA, not applicable.
*Millions.

Table S3. Oligonucleotide sequences

Primer name	Description	Sequence 5′→3′
LS574	mmE4B(1,072–1,173) L1107I C	gtcatcagggtgtccatGaTagggtctctgaactcg
LS575	mmE4B(1,072–1,173) R	aagcttTCAGTGGTCACTGCTCTGTTTCT
LS576	mmE4B(1,072–1,173) L1107A W	gacgagttcagagaccctgcGatggacaccctgatgac
LS577	mmE4B(1,072–1,173) L1107A C	gtcatcagggtgtccatCgcagggtctctgaactcgtc
LS578	mmE4B(1,072–1,173) L1107I W	cgagttcagagaccctAtCatggacaccctgatgac
LS579	mmE4B(1,072–1,173) L1107I C	gtcatcagggtgtccatGaTagggtctctgaactcg
LS714	Variant library oligo 1	CCGGCCGAATTCTatagagaagtttaaacttcttgcagagaaagtggaggaaatcgtggcaaagaatgcgc gggcagaaatagactacagcgatgccccggacgagttcagagaccctctgatggacaccctgatgaccgatccc gtgagactgccctctggcaccgtcat
LS715	Variant library oligo 2	ggaccgttctatcatcctgcggcatctgctcaactccccaccgacccctt caaccgccagatgctgactgagagcatgctggagccagtgccagagc taaaggagcagattcaggcctggatgagagagaaacagagcagtgac cacTGAATAGAATGAnnnnnnnnnnnnnnnnnnAAGCT TACGCCGCACCGG
LS737	E4B-index01_CG-R	CAAGCAGAAGACGGCATACGAGATAAAACCCCCCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS738	E4B-index02_CG-R	CAAGCAGAAGACGGCATACGAGATTTTTCCCCCCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS739	E4B-index03_CG-R	CAAGCAGAAGACGGCATACGAGATCCCCAAAACCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS740	E4B-index04_CG-R	CAAGCAGAAGACGGCATACGAGATGGGGAAAACCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS741	E4B-index05_CG-R	CAAGCAGAAGACGGCATACGAGATCACCACCACCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS742	E4B-index06_CG-R	CAAGCAGAAGACGGCATACGAGATGTGGTGGTCCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS743	E4B-index07_CG-R	CAAGCAGAAGACGGCATACGAGATTCATCATCCCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS744	E4B-index08_CG-R	CAAGCAGAAGACGGCATACGAGATACTGACTGCCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS745	E4B-barcode-seq-F	CTG GAT GAG AGA GAA ACA GAG CAG TGA CCA CTG AAT AGA ATG A
LS757	Ube4b-M1124V-W	CCTCTGGCACCGTCGTGGACCGTTCTATC
LS758	Ube4b-M1124V-C	GATAGAACGGTCCACGACGGTGCCAGAGG
LS768	HS-UBE4B-M1124V-W	CTGCCCTCTGGCACCATCGTGGACCGCTCCATCATCC
LS769	HS-UBE4B-M1124V-C	GGATGATGGAGCGGTCCACGATGGTGCCAGAGGGCAG
LS787	E4BUF-BamHI	ggcccgGGATCCGCAGAAATCGACTACAGCGATG
LS788	T7-E4BU_EcoRI_F	cggcccggaatcctGCAGAAATAGACTACAGCGATG
LS789	T7_H3C_HA_E4B_W	ccggccgaattctCTGGAAGTTCTGTTCCAGGGGCCCTACCCATACGATGTTCC
LS790	T7_H3C_HA_E4B_C	CTTTCTCTGCAAGAAGTTTAAACTTCTCTATAGCGTAATCTGGAACATCGTATGGGTA
LS791	T7_H3C_HA_E4B_F	TACCCATACGATGTTCCAGATTACGCTATAGAGAAGTTTAAACTTCTTGCAGAGAAAG
LS792	T7_H3C_HA_E4B_amplify_F	ccggccgaattctCTGGAAGTTC
LS793	E4B_F1141Y_W	CCCCACCGACCCCTACAACCGCCAGATG
LS794	E4B_F1141_C	CATCTGGCGGTTGTAGGGGTCGGTGGGG
LS795	E4B_Q1144A_W	CCCCTTCAACCGCGCGATGCTGACTGAGAG
LS796	E4B_Q1144A_C	CTCTCAGTCAGCATCGCGCGGTTGAAGGGG
LS799	E4B_D1139N_W	GCTCAACTCCCCCACCAACCCCTTCAACCG

Table S3. Cont.

Primer name	Description	Sequence 5′→3′
LS800	E4B_D1139N_C	CGGTTGAAGGGGTTGGTGGGGGAGTTGAGC
LS801	E4B_N1142T_W	CACCGACCCCTTCACCCGCCAGATGC
LS802	E4B_N1142T_C	GCATCTGGCGGGTGAAGGGGTCGGTG
LS805	E4B_L1078R_W	ATAGAGAAGTTTAAACTTCGTGCAGAGAAAGTGGAGGAAATC
LS806	E4B_L1078R_C	GATTTCCTCCACTTTCTCTGCACGAAGTTTAAACTTCTCTAT
LS837	hsUBE4B_L1107I_W	GATGAGTTCAGAGACCCTATCATGGACACCCTCATGAC
LS838	hsUBE4B_L1107I_C	GTCATGAGGGTGTCCATGATAGGGTCTCTGAACTCATC
LS839	hsUBE4B_D1139N_W	CAACTCCCCCACGAACCCCTTCAACCG
LS840	hsUBE4B_D1139N_C	CGGTTGAAGGGGTTCGTGGGGGAGTTG
LS841	hsUBE4B_N1142T_W	CACGGACCCCTTCACCCGGCAGACGC
LS842	hsUBE4B_D1142N_C	GCGTCTGCCGGGTGAAGGGGTCCGTG
LS843	hsUBE4B_P1140T_W	CTCCCCCACGGACACCTTCAACCGGC
LS844	hsUBE4B_P1140T_C	GCCGGTTGAAGGTGTCCGTGGGGGAG
LS845	hsUBE4B_BsrGI_F	ATTTTTTATTGTACAATACTCTCCCCAGGCGCTTTATG
LS826	T7_UP_v2	CTAACTTCCAAGCGGACCAG
LS827	T7_DOWN_v2	TGAGCGCATATAGTTCCTCCT
LS725	E4B-R	CAAGCAGAAGACGGCATACGAGATTAACTAGTTACTCGAGTGCGGCCG
LS749	New E4B-CG-F	AATGATACGGCGACCACCGAGATCTACACgccGTGATGCTCGGGGATCCGAATTCT
LS727	E4B-90-F	AATGATACGGCGACCACCGAGATCTACACgcAGCGATGCCCCGGACG
LS728	E4B-180-F	AATGATACGGCGACCACCGAGATCTACACgcgcGGACCGTTCTATCATCCTGCG
LS729	E4B-267-F	AATGATACGGCGACCACCGAGATCTACACcccgAGTGCCAGAGCTAAAGGAGCAG
LS730	E4B-FL-seq-F	cgcgATGCTCGGGGATCCGAATTCT
LS731	E4B-90-seq-F	gcAGCGATGCCCCGGACG
LS732	E4B-180-seq-F	gcgcGGACCGTTCTATCATCCTGCG
LS733	E4B-267-seq-F	ccccgAGTGCCAGAGCTAAAGGAGCAG
LS724	E4b-first-read	TAACTAGTTACTCGAGTGCGGCCGCAAGCTT
LS737	E4B-index01_CG-R	CAAGCAGAAGACGGCATACGAGATAAAACCCCCCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS739	E4B-index03_CG-R	CAAGCAGAAGACGGCATACGAGATCCCCAAAACCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS741	E4B-index05_CG-R	CAAGCAGAAGACGGCATACGAGATCACCACCACCC AAG GGG TTA ACT AGT TAC TCG AGT GCG G
LS748	E4B-barcodes_index-seq	AAG CTT GCG GCC GCA CTC GAG TAA CTA GTT AAC CCC TTG GG

Other Supporting Information Files

Dataset S1 (XLSX)