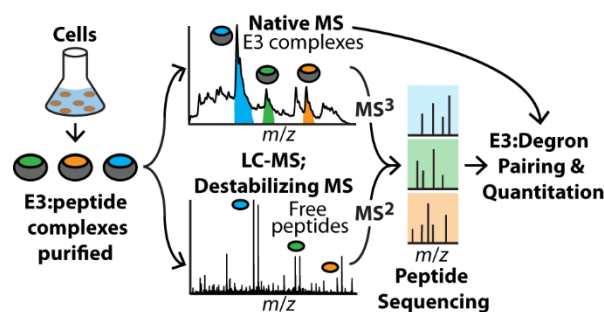


Degronomics: Mapping the Interacting Peptidome of a Ubiquitin Ligase Using an Integrative Mass Spectrometry Strategy



Original manuscript submitted to *Anal. Chem.* on 5/19/19 as ac-2019-02331b

Revised manuscript submitted on 9/15/19

Daniele Canzani,^a Domnița-Valeria Rusnac,^b Ning Zheng,^b Matthew F. Bush^{a*}

^a Department of Chemistry, University of Washington, Seattle, WA 98195, USA

^b Howard Hughes Medical Institute, Department of Pharmacology, University of Washington, Seattle, WA 98195, USA.

* Address correspondence to: mattbush@uw.edu

Abstract

Human cells make use of hundreds of unique ubiquitin E3 ligases to ensure proteome fidelity and control cellular functions by promoting protein degradation. These processes require exquisite selectivity, but the individual roles of most E3s remain poorly characterized in part due to the challenges associated with identifying, quantifying, and validating substrates for each E3. We report an integrative mass spectrometry (MS) strategy for characterizing protein fragments that interact with KLHDC2, a human E3 that recognizes the extreme C-terminus of substrates. Using a combination of native MS, native top-down MS, MS of destabilized samples, and liquid chromatography MS, we identified and quantified a near complete fraction of the KLHDC2-binding peptidome in *E. coli* cells. This degronome includes peptides that originate from a variety of proteins. Although all identified protein fragments are terminated by diglycine, the preceding amino acids are diverse. These results significantly expand our understanding the sequences that can be recognized by KLHDC2, which provides insights into the potential substrates of this E3 in humans. We anticipate that this integrative MS strategy could be leveraged more broadly to characterize the degronomes of other E3 ligase substrate receptors, including those that adhere to the more common N-end rule for substrate recognition. Therefore, this work advances “degronomics,” *i.e.*, identifying, quantifying, and validating functional E3:peptide interactions in order to determine the individual roles of each E3.

Introduction

The ubiquitin-proteasome system (UPS) regulates intracellular protein degradation.¹ E3 ubiquitin ligases are essential UPS enzymes that transfer ubiquitin from an E2 conjugating enzyme to a protein substrate² and ensure proteome fidelity by selectively eliminating aberrant proteins that may have translation, folding, or other errors.³ There are hundreds of E3s in humans, but the individual roles of most remain poorly characterized.⁴ One tremendous hurdle for determining the roles of individual E3s has been identifying their substrates. Several techniques have been adapted for characterizing E3-substrate pairing, including affinity-based capturing approaches,⁵ yeast two-hybrid methods,⁶ high-throughput microscopy screens,⁷ and phage display assays.⁸ Bottom-up, mass spectrometry (MS) based proteomics,⁹ in which proteins are enzymatically digested into peptides, peptides are separated using liquid chromatography (LC), and peptides are sequenced using tandem MS, has been used to investigate several aspects of the UPS. For example, targets of the entire UPS have been determined by identifying ubiquitination sites across the proteome.^{10,11} Quantitative proteomics has been used to characterize protein turnover in mammalian cells in response to expression of a selected E3 ligase, which identified several substrates for those E3s.^{12,13}

One emerging technique for characterizing the UPS is global protein stability profiling (GPS), which is a high throughput approach for characterizing the stability of individual proteins.¹⁴ Recently, GPS was used to identify hundreds of potential substrates of the Skp1–CUL1–F-box (SCF) family of ubiquitin ligases. Specific E3-substrate interactions are often dictated by a short linear sequence, known as a degron, in the substrate that is specifically recognized by the E3.¹⁵ Identification of degrons can help predict and validate substrate targeting by an E3. For instance, GPS revealed that one family of E3 ligases, the CUL2-RING ligases

(CRL2), recognizes degrons at the extreme C-terminus of substrates.^{16,17} GPS analysis of KLHDC2, a member of the CRL2 family, revealed binding to selenoprotein fragments that contain diglycine at the C terminus.¹⁸ Subsequent characterization using *in vitro* binding assays and X-ray crystallography showed that KLHDC2 can recognize degron peptides with nanomolar affinities via the top surface pocket of its β -propeller domain.¹⁹

Despite substantial and sustained efforts to identify E3 ligase substrates, substrates have been validated for relatively few E3s. One challenge is that established methods, including conventional MS-based proteomics and GPS, do not directly probe E3:substrate binding. Native MS, in which samples are prepared in aqueous solutions with physiological pH and ionic strength, can preserve noncovalent interactions during transfer into the gas phase.²⁰ Due to this advantage, native MS has been used to identify protein cofactors^{21–23} and to characterize the stoichiometry and composition of protein complexes.^{24–26} More recently, native top-down MS, in which ions of intact proteins and protein complexes are subjected to fragmentation to yield sequence information, has been used to identify proteoforms²⁷ and link that information to higher-order protein structure.^{28,29} Native MS has been used to construct structural models of the CRL5^{SOCS2} ligase³⁰ and to characterize cofactor binding to the SCF^{FBXL3} and SCF^{TIR1} E3s.^{31,32} We recently used native top-down MS to corroborate interactions between KLHDC2 and a selenoprotein fragment identified through GPS analysis;¹⁹ otherwise native MS and native top-down MS have not been used previously to characterize E3:substrate interactions.

In this study, we report an integrative MS strategy that combines results from native and non-native experiments to characterize KLHDC2:peptide complexes that were copurified from cells. We describe the first application of native top-down MS to identify unknown copurified peptides. Our results provide direct evidence that KLHDC2 recognizes cellular peptides

containing C-terminal glycylglycine or glycylalanine, and demonstrates high-affinity binding for a diverse profile of peptides that contain those features. These results significantly expand our understanding of degrons that can be recognized by KLHDC2, which aids in identifying potential substrates of this E3 in humans. More generally, this integrative MS strategy advances “degronomics,” *i.e.*, identifying, quantifying, and validating functional E3:peptide interactions in order to determine the individual roles of each E3.

Methods

Molecular Biology and Protein Purification for Mass Spectrometry Experiments.

The kelch repeat domain of human KLHDC2 (UniProt: Q9Y2U9, amino acids 22–362) was subcloned into the pET vector with an N-terminally fused His-elongation factor Ts (TSF) and a TEV-cleavage site. The His-TSF-KLHDC2 protein was overexpressed and purified from BL21 (DE3) *E. coli* cells. Bacterial cells transformed with the pET-based expression plasmid were grown in LB broth to an OD₆₀₀ of 0.8-1 and induced with 0.5 mM IPTG. Cells were harvested, re-suspended and lysed in lysis buffer (20 mM Tris, pH 8.0, 200 mM NaCl, 20 mM imidazole) in the presence of protease inhibitors (1 µg mL⁻¹ leupeptin, 1 µg mL⁻¹ pepstatin and 100 µM phenylmethylsulfonyl fluoride) using a microfluidizer. The His-TSF-KLHDC2 protein was isolated from the soluble cell lysate by HisPurTM Ni-NTA Superflow Agarose (Thermo Fisher Scientific, Waltham, Massachusetts). After TEV cleavage of the His-TSF, KLHDC2 was further purified by Q Sepharose High Performance resin (GE Healthcare, Chicago, Illinois). The NaCl eluates were subjected to Superdex-200 size-exclusion chromatography (GE Healthcare). All samples were flash frozen in liquid nitrogen for storage prior to use.

Native MS. KLHDC2 purified from *E. coli* cells was buffer exchanged into aqueous ammonium acetate (200 mM, pH 7) using four cycles of dilution and re-concentration with a 10K MWCO Spin-X UF centrifugal concentrator (Corning Inc., Corning, New York) at 14,000 g and 4 °C. The final concentration of KLHDC2 used for experiments was approximately 20 μ M. 2-3 μ L of the prepared solution was added into the tip of a pulled borosilicate glass capillary. A platinum wire electrode was inserted into the capillary and placed in direct contact with the solution. A potential between 0.5 and 1.0 kV was applied to the electrode to achieve ionization.³³ Those ions were then analyzed using a hybrid electrospray/quadrupole/ion-mobility/time-of-flight mass spectrometer (Waters Synapt G2 HDMS, Milford, Massachusetts). Native mass spectra were acquired using a 45 V bias between the sampling and extraction cones in the atmospheric-pressure interface that was operated at room temperature, and using a 3 V bias between the quadrupole mass filter and the trap collision cell that contained ~20 mTorr of argon gas. Activation in the trap collision cell was performed using those conditions, except that the bias between the quadrupole and trap collision cell was increased to 45 V. Activation in the atmospheric-pressure interface was achieved by increasing the bias between the sampling and extraction cones to 120 V. Tandem MS of peptide ions released during activation in the atmospheric-pressure interface was performed by increasing the bias between the quadrupole mass filter (used to isolate a released peptide ion) and the trap collision cell to 30-45 V. Mass spectra were calibrated externally using spectra obtained from electrospray ionization of 30 mg mL⁻¹ aqueous CsI.

MS from Destabilizing Conditions. Thermal destabilization was achieved by transferring approximately 5 μ L of the sample prepared for native MS into a 1.5 mL snap-cap vial, which was then placed in 55 °C heat block for 5 minutes. Chemical destabilization was

achieved by adding 15% acetonitrile and 1% formic acid by volume into the sample prepared for native MS. The thermally and chemically destabilized samples were then each analyzed using the parameters described for native MS. Tandem MS was performed by increasing the bias between the quadrupole mass filter (used to isolate the selected peptide ion) and the trap collision cell to 25-45 V.

LC-MS². The sample prepared for native MS was adjusted to pH 4 with the addition of trifluoroacetic acid, loaded onto a Pierce C₁₈ tip with a 10 µL bed (Thermo Fisher Scientific), washed, and eluted using an aqueous solution with 70% acetonitrile and 0.1% formic acid by volume. Peptides were separated using a nanoAcquity UPLC (Waters) using a 150 mm fused-silica emitter that was packed with reversed-phase ProntoSIL AQ 120-C18 5 µm resin (Bischoff Chromatography, Germany). A 60-minute linear gradient from 10-30% organic phase (acetonitrile with 0.1% formic acid), followed by a 1-minute gradient to 80% organic phase, was used with a flow rate of 0.3 µL/min. The eluent was analyzed using a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) and a data-dependent acquisition method. Full-scan MS spectra were acquired in positive ion mode with the resolution set to 70k and with a scan range of 400 to 2,500 *m/z*. MS² spectra were acquired on the top 20 precursors present in the MS¹ scan, and MS² spectra were acquired at 17.5k resolution with a 1.5 *m/z* isolation window.

Peptide Identification. Peptide sequences were identified using unrestricted database searches and downstream validation. A protein database was constructed that contained the *E. coli* BL21 proteome (Proteome ID UP000002032),³⁴ the His-TSF-KLHDC2 fusion protein that introduced with molecular biology (Figure S3), and reverse-sequence protein decoys. The experimental data was searched against this database using both Comet³⁵ and MSFragger.³⁶

Searches were performed using standard settings for accurate mass data and allowing for a wide range of isotope errors (-1, 0, +1, +2, +3 isotope offsets). The results of both searches were each analyzed with PeptideProphet³⁷ using only the expect score as the discriminate and using the decoys to establish the negative distribution. Those analyses were then combined using iProphet.³⁸ This workflow was performed within the Trans-Proteomic Pipeline,³⁹ except the MSFragger search and the subsequent PeptideProphet analysis were performed using FragPipe.³⁶ Quantification based on the LC-MS data was performed using XPRESS⁴⁰ in label-free mode within the Trans-Proteomic Pipeline. PEAKS Studio 8.5 (Bioinformatics Solutions, Waterloo, Ontario) was used to sequence spectra obtained from native top-down and destabilizing MS experiments. Spectra from those experiments were signal averaged using MassLynx v4.1, converted into a single mzML file, then searched using a 0.1 Da precursor mass tolerance and a 0.1 Da fragment ion tolerance with no enzyme selected for *de novo* sequencing. For PEAKS DB analysis, the UniProt *E.coli* BL21 and K12 proteomes were combined with the FASTA entries for human KLHDC2 and the His-TSF tag.

Determining Relative Abundances of Bound Peptides. The native mass spectra were resampled with 4 equally spaced points between each m/z value. The signal profile for apo KLHDC2 was used to represent the contributions from each KLHDC2:peptide complex. A template was created by first isolating the feature for apo KLHDC2 in the resampled spectrum. Since the tailing edge of this feature was obscured by the peptide-bound portion of the spectrum, a tail was added by reflecting the leading edge of the profile and then reducing the intensity of this tail to reach the baseline of the spectrum. The relative intensity (I_n) for each complex was optimized by minimizing the total residual (Equation 1) using Powell's conjugate gradient method.⁴¹ A notebook illustrating this method is included with the *Supporting Information*.

Results and Discussion

KLHDC2 is a human E3 ligase that is of considerable interest because it recognizes the C-terminus of substrates,^{16–19} rather than the more common N-end rule⁴² for substrate recognition. As part of our effort to characterize the molecular biophysics of KLHDC2, we used native MS to analyze this protein purified from *E. coli* cells. Several high-intensity peaks were observed at m/z values expected for the 11+ to 13+ charge states of apo KLHDC2. To our surprise, many additional features were also observed (Figure 1). The unexpected features correspond to ions that are approximately 400 to 1500 Da greater in mass than apo KLHDC2 and suggest that many different peptides had complexed with KLHDC2. Since exogenous peptides were not added at any point, these peptides must have bound to KLHDC2 during their expression or purification from *E. coli* cells. Therefore, the bound peptides represent bacteria-produced protein fragments that copurified with KLHDC2. To the best of our knowledge, these spectra show the first evidence of copurified protein fragments bound directly to an E3 using MS. Here we report the integrative mass spectrometry strategy that we developed to identify and quantify the protein fragments that comprise the interacting peptidome of this E3 ligase, *i.e.*, the degrome of KLHDC2 in these cells. We then discuss the properties of this degrome and the potential for using this strategy to determine the roles of individual E3 ligases.

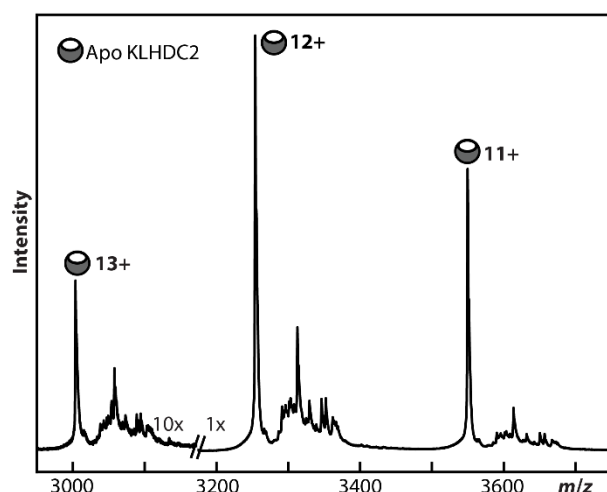


Figure 1. Native mass spectrum of KLHDC2 expressed in *E. coli* cells. In addition to features expected for apo KLHDC2, additional features corresponding to ions that are approximately 400-1500 Da larger in mass relative to the apo protein are also observed. These features are attributed to KLHDC2:peptide complexes. Intensities below 3200 m/z are increased tenfold to aid in visualization.

Native Top-Down MS. To determine the identity of the copurified peptides, we first used native top-down MS, *i.e.*, subjecting intact KLHDC2:peptide complex ions to multiple stages of activation in order to sequence the bound peptides. Peptides were released using collision-induced dissociation (CID), which was achieved at the atmospheric-pressure interface or in the trap collision cell of the mass spectrometer. When CID was performed in the trap collision cell, the transmission profile of the quadrupole (positioned immediately prior to the trap cell and operated as an RF-only guide) was set to favor higher- m/z ions corresponding to apo or peptide-bound KLHDC2. This filtered away lower- m/z contaminants, including any peptides that were unbound in solution, and provides high confidence that the peptide ions observed after CID were released from complexes with KLHDC2. When CID was performed at the atmospheric-pressure interface, it is possible that unbound peptides from solution may also be transmitted. CID performed in the trap collision cell (Figure S1) or at the atmospheric-pressure interface (Figure 2A) both yielded identical peptide ions, which corroborates that the peptide ions in both experiments originated from complexes with KLHDC2. Peptide ions released at the atmospheric-pressure interface were used for acquiring peptide fragmentation spectra because those ions could be quadrupole selected prior to subsequent fragmentation.

Seven interacting peptides were identified using native top-down MS; representative fragmentation spectra are shown in Figure 2B and S2. In order to minimize bias, the fragmentation spectra were first analyzed using unconstrained *de novo* sequencing and the candidate sequences were compared with a protein library, as described in the *Methods*. Five of the peptides are assigned to a series of N-terminal truncations of ASDEGEVIVFGG (Figure 2B), which is a fragment of KLHDC2. This assignment was validated by subjecting synthetic ASDEGEVIVFGG to tandem MS on the same instrument (Figure 2C). The origin of the N-terminal truncations are not yet understood, but are presumably the result of digestion or degradation that occurred *in vivo*. The other two peptides were assigned to EKGGGSGGG and GGGSGGG, which are fragments of the His-elongation factor Ts (TSF) tag that was fused to the N-terminus of KLHDC2 for expression and purification purposes (Figure S3). These peptides all feature C-terminal diglycine, which is present in the majority of the reported degrons of KLHDC2.¹⁷ These seven degron-like peptides were all identified through the direct analysis of KLHDC2:peptide complexes that were formed in cells or during cell lysis and survived extensive purification.

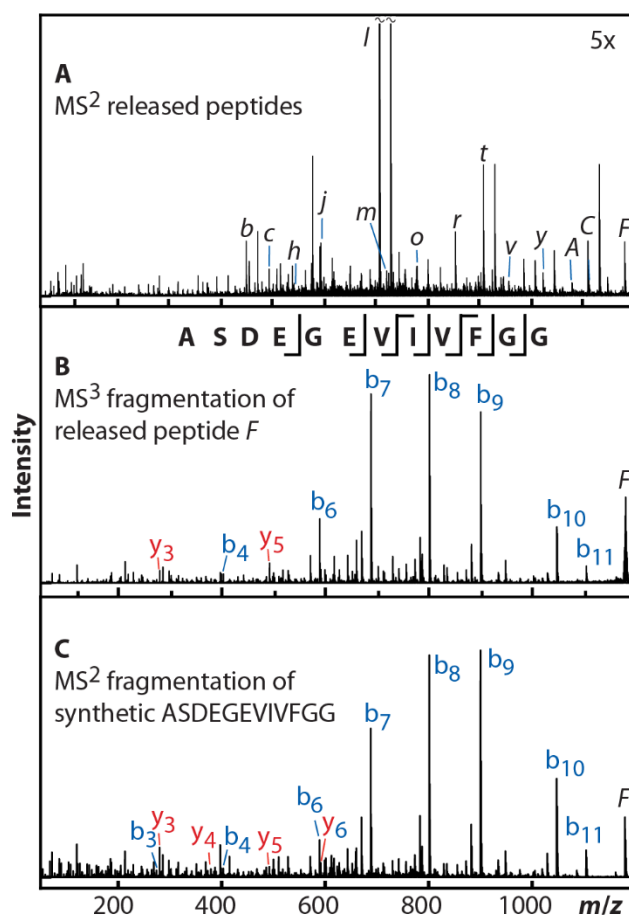


Figure 2. Native top-down MS of KLHDC2:peptide complexes. (A) Complexes were ionized and subjected to collision-induced dissociation (CID) at the atmospheric-pressure interface of the mass spectrometer, resulting in the release of peptide ions that are labeled using the IDs in Table 1. Intensities are increased fivefold to aid in visualization. Additional ions are annotated in Figure S2. (B) Ion *F* was isolated and subjected to CID in the trap collision cell of the instrument. CID fragments (labeled using the scheme of Roepstorff and Fohlman)⁴³ were used to assign ion *F* to ASDEGEVIVFGG, which is a fragment of KLHDC2. (C) Synthesized ASDEGEVIVFGG was resuspended in aqueous 200 mM ammonium acetate, ionized using electrospray, and subjected to CID in the trap collision cell of the same instrument. The fragmentation spectra in B and C are nearly identical, which corroborates the assignment of peptide *F* to ASDEGEVIVFGG.

Table 1. The degronome of KLHDC2 in *E. Coli* Cells based on observations (○) and sequencing (●) in Native, Destabilizing, and Liquid Chromatography MS experiments.

Protein of Origin	Interacting Peptide	ID	RA ^a	N	D	LC
KLHDC2	VSDGRHMFVWGG	<i>J</i>	1.7(0.4)	—	○	●
	TVGNRGFVFGG	<i>A</i>	0.3(0.2)	○	○	●
	VGNRGFVFGG	<i>h</i>	1.4(0.4)	—	—	●
	NRGFVFGG	<i>D</i>	0.4(0.4)	○	○	●
	FVFGG	<i>x</i>	0.5(0.1)	○	○	●
	SSDHLFLFGG	<i>r</i>	1.8(0.5)	○	—	●
	FLFGG	<i>f</i>	0.6(0.4)	○	—	●
	HTACASDEGEVIVFGG	<i>U</i>	0.2(0.1)	—	—	●
	ACASDEGEVIVFGG	<i>K</i>	0.3(0.4)	—	○	●
	CASDEGEVIVFGG	<i>H</i>	2.0(0.5)	○	—	●
	ASDEGEVIVFGG	<i>F</i>	5.7(1.1)	●	●	●
	SDEGEVIVFGG	<i>C</i>	5.5(0.8)	●	●	●
	DEGEVIVFGG	<i>y</i>	1.7(0.2)	●	●	●
	EGEVIVFGG	<i>t</i>	4.3(0.5)	●	●	●
	GEVIVFGG	<i>o</i>	3.1(0.5)	○	○	●
	EVIVFGG	<i>m</i>	1.5(0.8)	—	○	●
	VIVFGG	<i>j</i>	5.4(1.7)	—	—	●
	IVFGG	<i>c</i>	1.9(1.1)	●	○	●
His-TSF solubility tag	AWSHQPFEKGGGSGGGSGG	<i>X</i>	0.8(0.3)	—	—	●
	WSHPQFEKGGGSGGGSGG	<i>W</i>	0.9(0.7)	—	—	●
	SHQPFEKGGGSGGGSGG	<i>R</i>	0.3(0.4)	—	—	●
	SSAWSHQPFEKGGGSGGG	<i>Y</i>	0.4(0.2)	—	—	●
	AWSHQPFEKGGGSGGG	<i>T</i>	0.5(0.1)	—	—	●
	WSHPQFEKGGGSGGG	<i>Q</i>	0.3(0.1)	—	—	●
	SHQPFEKGGGSGGG	<i>I</i>	3.7(0.8)	—	—	●
	HPQFEKGG	<i>s</i>	1.5(0.5)	○	○	●
	EKGGGSGGG	<i>l</i>	19.5(2.6)	●	●	—
	GGGSGGG	<i>b</i>	4.5(2.0)	●	○	—
UPF0441 protein YgiB	QRSATGTSSRSMGG	<i>N</i>	1.7(0.3)	—	—	●
Cell division protein FtsZ	TNDAAIKVIGVGGGG	<i>L</i>	0.6(0.3)	—	—	●
	VIKVIKVGGGG	<i>v</i>	0.8(0.1)	○	—	●
Glutamate dehydrogenase	SSAIGPYKGG	<i>u</i>	0.8(0.3)	○	—	●
	SAIGPYKGG	<i>q</i>	0.2(0.2)	○	—	●
Outer membrane protein F	SDDFFVGRVGG	<i>E</i>	0.8(0.1)	—	—	●
	DDFFVGRVGG	<i>z</i>	0.8(0.5)	—	—	●
Malate dehydrogenase	MKVAVLGAAGG	<i>w</i>	0.9(0.2)	—	—	●
	KVALGAAGG	<i>p</i>	1.7(0.2)	—	—	●
Superoxide dismutase	TAFEGKLEEIRSEGG	<i>Z</i>	1.0(0.5)	—	—	●
30S ribosomal protein S11	TDRQGNALGWATAGG	<i>P</i>	0.2(0.1)	—	—	●
	LGWATAGG	<i>n</i>	1.1(0.9)	○	—	●
30S ribosomal protein S12	IGGEGHNLQEHSVILIRGG	<i>α</i>	0.5(0.2)	—	—	●
30S ribosomal protein S5	VFMQPASEGTGHIAGGA	<i>V</i>	0.5(0.2)	—	—	●
	MQPASEGTGHIAGGA	<i>M</i>	0.6(0.3)	—	—	●
Phage Shock operon rhodanese PspE	AEHWIDVRVPEQYQQEHVQGA	<i>δ</i>	1.6(1.8)	—	—	●
	IDVRVPEQYQQEHVQGA	<i>β</i>	0.8(0.6)	—	—	●
17 kDa surface antigen	IQQGDDSNVIGAIGGA	<i>O</i>	0.9(0.2)	—	—	●
Phenylalanine-tRNA ligase α subunit	LRELPPERPAAGA	<i>S</i>	0.5(0.2)	—	—	●
DNA- binding protein HU-β	KSQIDKIAAGA	<i>G</i>	0.5(0.3)	—	—	●
	SQLIDKIAAGA	<i>B</i>	0.1(0.1)	—	—	●
Ambiguous origin (See SI)	^b LPPGG	<i>e</i>	1.2(0.6)	○	—	●
	LVIYGG	<i>d</i>	2.3(0.9)	—	○	●
	LPEFGG	<i>k</i>	4.9(0.7)	—	—	●
	PEFGG	<i>g</i>	0.0(0.0)	○	○	●
	LIIGGG	<i>i</i>	3.0(0.5)	—	○	●
	^b LYFGG	<i>a</i>	1.2(0.9)	○	—	●

^aRelative Abundances are reported at the 95% confidence level.

^bAdditional sequences with leucine/isoleucine substitutions are also possible.

Integrating Complementary MS-Based Measurements.

Although roughly half of the KLHDC2 containing ions observed using native MS were complexed with peptides (Figure 1), the abundance of each individual peptide was inherently lower than that of KLHDC2. This ultimately limited the number of peptides that could be observed in native MS² experiments (Figure 2A) and sequenced in native MS³ experiments (Table 1). Since it was apparent that many other peptides were bound to KLHDC2 than were identified using native top-down MS alone, we developed an integrative MS-based strategy to identify additional bound peptides (Figure 3).

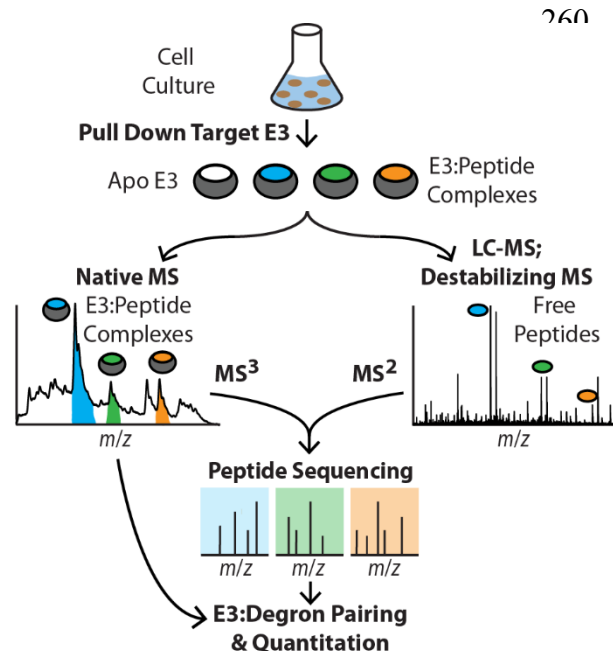


Figure 3. Integrative MS for degromics. Native MS enabled the direct observation and mass determination of E3:peptide complexes that were purified from cells. Native top-down MS provided primary amino acid sequences for a subset of the bound peptides. Complementary MS-based approaches, including destabilized-sample MS² and LC-MS² provided additional peptide identifications. Together, these methods provide an integrative strategy to identify and quantify protein fragments interacting with an E3 in cells that could be used to study E3:degrom interactions more broadly.

Destabilizing samples, by the addition of acetonitrile and formic acid (15% and 0.1% by volume, respectively) or heating the sample to 55 °C, caused peptides to be released from KLHDC2 in solution. MS¹ of these destabilized samples and MS² of the native-like complexes yielded similar spectra, but destabilized-sample MS yielded more intense signals for peptide ions than native MS². The highest-intensity ions from the destabilized samples were quadrupole selected for MS² (Figure S4), which confirmed the peptide sequences identified using native top-

down MS (Table 1). However, analysis of the destabilized samples was limited by congested MS¹ spectra; some peptide ions could not be isolated completely using the quadrupole mass filter and others exhibited inadequate intensity for MS². Ultimately, no additional peptides were sequenced from the destabilized samples, but analysis of those samples corroborated many of the results from native top-down MS. LC-MS², without enzymatic digestion, was used to determine the sequences for the greatest number of interacting peptides (Table 1, probability scores from peptide-spectrum matches in Table S1).

Although the destabilized-sample MS and LC-MS² methods offer certain advantages, *i.e.*, greater precursor ion intensities and greater number of identified peptides, respectively, those methods both destroy the KLHDC2:peptide interactions prior to analysis. The native top-down and destabilized sample MS benefited from the use of direct infusion of the sample into the ion source. Relative to the LC-MS² experiments, adjusting the extent of activation for individual peptide ions contributed to higher quality fragmentation spectra and acquiring fragmentation spectra for more time enabled higher signal-to-noise spectra after averaging. Consequently, the average of the significance scores from PEAKS DB analysis ($-10\log_{10}P$) for the peptides identified using all three approaches are 51, 44, and 30 for native MS³, destabilized-sample MS², and LC-MS², respectively. Interesting, two of the most abundant interacting peptides (*b* and *l*) were not identified in the LC-MS² experiments. The absence of *b* and *l* in the LC-MS² data may be attributable to the hydrophobicity of each peptide, which is significantly lower than the peptides that were identified using LC-MS² (Table S2), and poor retention during solid-phase extraction and LC. Therefore, the data generated using all three approaches are complementary and enable the most comprehensive overview of the interacting peptidome of KLHDC2 in *E. coli* cells.

Relative Abundances. The peptides identified through the integrative MS strategy were used to assign features in the native mass spectrum (Figure 4A) that had been tentatively assigned to KLHDC2:peptide complexes (Figure 1). Fundamental studies indicate that native-like ions are formed through the charged-residue model for electrospray ionization, in which the final ions are the result of desolvation of charged droplets containing the biomolecule.^{33,44,45} Since the ionization efficiency of KLHDC2 and its complexes should be largely independent of peptide binding,⁴⁶ we hypothesize that the relative abundances of the KLHDC2:peptide complex ions accurately reflect the relative abundance of those complexes in solution prior to ionization. In contrast, the ionization efficiency of peptides from electrospray can depend strongly on peptide length, sequence, hydrophobicity, solubility, and other factors.⁴⁷ For example, hydrophobic peptides have greater affinity for surface of the electrospray droplet,⁴⁸ which can result in their preferential ionization⁴⁹ and suppress the ionization of less hydrophobic peptides.⁵⁰ Therefore, it is normally extremely difficult to quantify peptides without using internal standards or peptide labelling, which is challenging to implement in discovery experiments. Label-free quantitation of the peptides discovered in these experiments was attempted based on their LC-MS elution profiles. The relative abundances of the peptides from that analysis are reported in Table S1, but due to the absence of highly abundant peptides such as *b* and *l* (Table 1) in the LC-MS experiments and the predicted differences in the ionization of peptides as described above, those values are not correlated with the relative abundances determined from the native MS experiments.

327
328

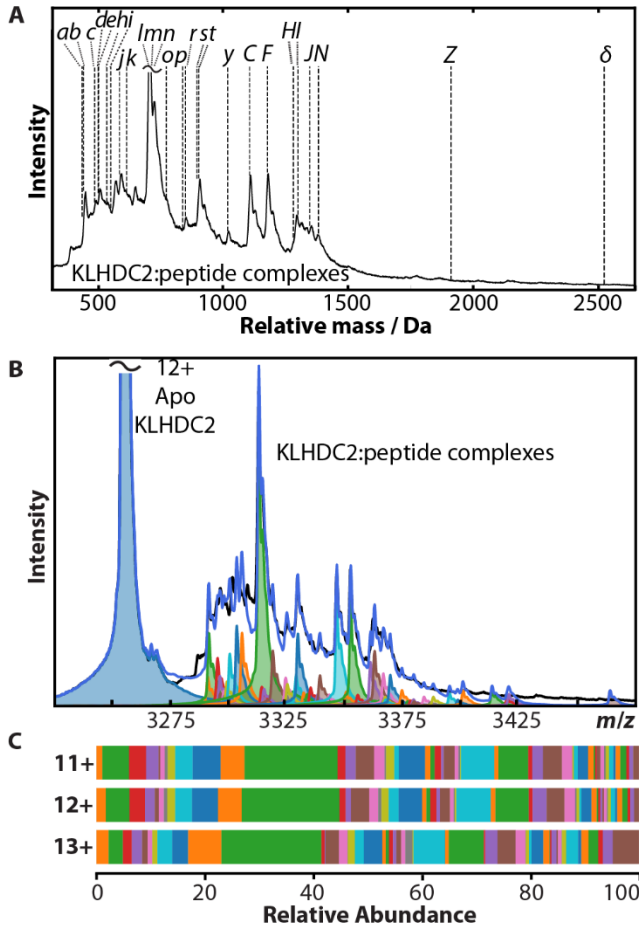


Figure 4. Identification and quantification of interacting peptides. (A) Native mass spectrum of the 12+ KLHDC2:peptide complex ions, plotted using a false mass axis that is relative to apo KLHDC2. Vertical bars are plotted using the masses of the interacting peptides that had a 1% or greater relative abundance and are labeled using the IDs in Table 1. For comparison, this region is plotted with all peptide IDs as a function of m/z in Figure S5. (B) The total residual (Equation 1) of the experiment (*black trace*) and the sum of modelled components (*blue trace*) was minimized by optimizing the relative intensity of each component. The contribution from each KLHDC2:peptide complex is represented using a different color. (C) The relative abundances of the interacting peptides determined using this approach for the 11+, 12+, and 13+ ions.

Congestion in these native mass spectra

imposes some additional challenges in

determining the relative abundance of the bound

peptides. For example, the signals for apo KLHDC2 (Figure 1) are broad and appear to depend on the presence of non-specific adducts that persisted through electrospray ionization⁵¹ as well as scattering with background gas in the time-of-flight mass analyzer. However, factors that broaden the MS signals for apo KLHDC2 are expected to be common to all KLHDC2:peptide complexes. From this, we hypothesize that the profile for apo KLHDC2 can be used as template to simulate the signals originating from each of the KLHDC2:peptide complexes. To test this hypothesis, the feature for apo KLHDC2 was extracted from the experimental spectrum as described in the *Methods*. For each of the n peptides in Table 1, this template was reproduced and shifted by τ , which is the neutral mass of the interacting peptide divided by the charge state

of the complex. Next, the relative intensity (I_n) for each simulated feature was optimized to minimize the total residual:

$$\text{Total Residual} = \sqrt{\sum_{m/z} (\text{Experiment}(\frac{m}{z}) - \sum_0^n I_n \times \text{Template}(\frac{m}{z} + \tau_n))^2} \quad (1)$$

Note that I_n is a scalar that is relative to that for the original template (I_0). Additional details of this process are described in the *Methods* and *Supporting Information*.

Figure 4B shows results from this analysis of the 12+ ions. The sum of the components is strongly correlated with the experimental spectrum. The derivatives of these traces are shown in Figure S6, which highlight the fidelity and the limitations of this approach. The differences between the model and experiment are most likely due to limitations in the template and unidentified peptides; for instance, the experiment appears to include contributions from KLHDC2 bound to peptides with masses of approximately 398, 565, and 648 Da that were not identified and are therefore not included in the model. However, the model reproduces nearly 98% of signal observed for KLHDC2:peptide ions, suggesting that the integrative MS strategy identified the vast majority of the interacting peptides. The relative abundances of the interacting peptides determined from the analysis of the 11+, 12+, and 13+ KLHDC2:peptide ions are shown in Figure 4C. Overall, the relative abundances determined for different charge states and for technical replicates (Figure S6) are similar to each other, demonstrating the robustness of this approach. The relative abundances, averaged over all charge states and replicates, are reported in Table 1. Note that the accuracy of this approach depends on the extent to which the original native mass spectrum (Figure 1) reports the distribution of interacting peptides in solution; different extents of peptide dissociation during protein purification or MS may bias these results.

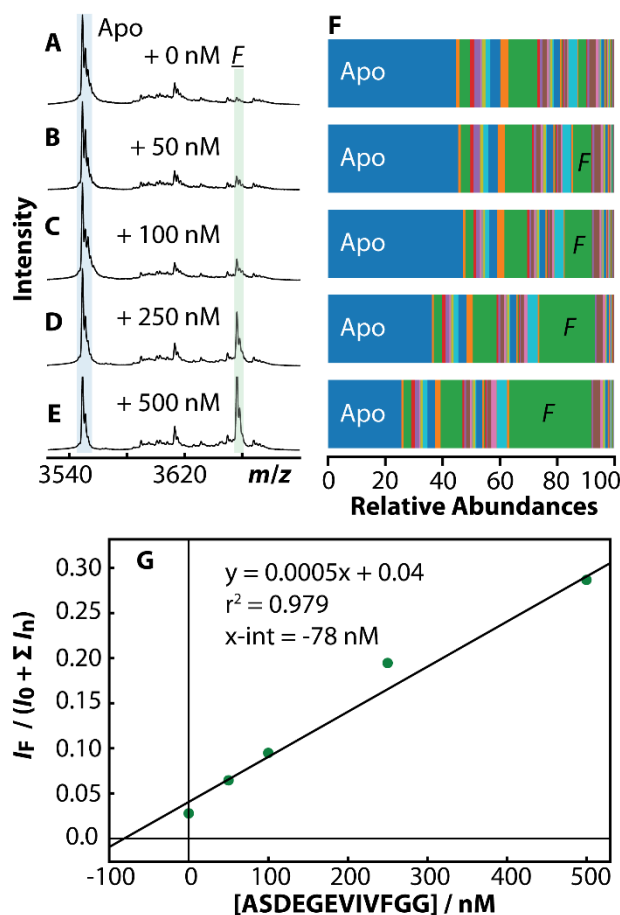


Figure 5. The peptide *F* (ASDEGEVIVFGG) was synthesized and added at various concentrations to the sample of KLHDC2 expressed in *E. coli* cells. (A-E) The intensity of the feature for the KLHDC2:*F* complex increases with peptide concentration, which is consistent with the intensities in the native mass spectra depending on the abundances of the corresponding complexes in solution. (F) The relative abundances of apo KLHDC2 and the KLHDC2:peptide complexes were determined for the spectra shown in A to E using the same method used for Figure 4. (G) The abundance of the KLHDC2:*F* complex relative to apo KLHDC2 and all KLHDC2:peptide complexes as function of the concentration of additional ASDEGEVIVFGG after it was spiked into the sample.

404

405 In order to characterize the relationship between the relative abundances determined
 406 using this quantitation approach and the abundances of the complexes in solution, we performed
 407 a titration experiment by adding the synthesized peptide *F* (ASDEGEVIVFGG). We added that
 408 peptide to the original sample of KLHDC2 from *E. Coli* cells and monitored the change in the
 409 resulting native mass spectra, which is shown in Figure 5A-E. The original sample was diluted
 410 twofold with the addition of the standard; the concentration of the additional standard is reported
 411 after this dilution step. The intensity of the feature assigned to the KLHDC2:*F* complex
 412 increased with the concentration of the spiked peptide. Figure 5F shows that the increase in the
 413 relative abundance of that complex is concomitant with a decrease in the relative abundance of
 414 the other components of the sample, especially apo KLHDC2. For comparison, Figure S7 shows

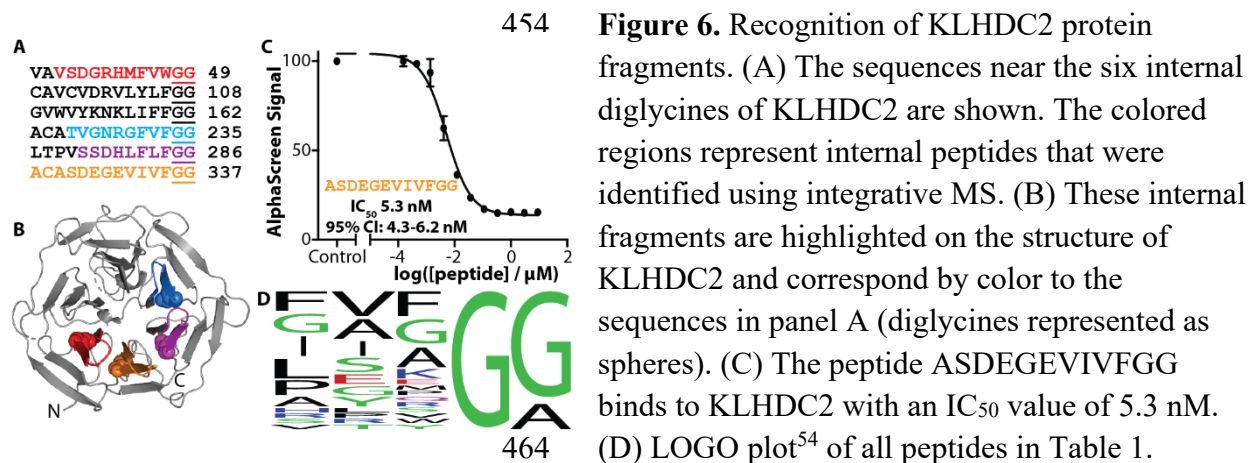
the relative abundance of the other components, excluding apo KLHDC2 and KLHDC2:*F*. This analysis suggests that the concentration of the KLHDC2:*F* complex in sample was 164 ± 44 nM, prior to the addition of the standard (Figure 5G). These results also illustrate the advantage interpreting native mass spectra using an experimentally derived template. Even though the different experiments yield slightly different extents of nonspecific adducts following electrospray ionization (Figures 5A to 5E), those differences are inherently accounted for using this approach for quantitation.

We anticipate that this approach for quantitation could be used to quantify the relative abundance of other molecules that interact with proteins, including those that bind endogenously and/or lack standards. Additionally, this templating approach could be used to facilitate measurements of K_d values directly from native MS experiments. Although simulating native mass spectra in order to quantify the relative abundance of noncovalent complexes has been reported previously,^{25,52} to the best of our knowledge this is the first report the use of an integrative mass spectrometry strategy to inform the identities and masses of the contributing complexes as well the use of an experimentally derived template to represent peak shapes.

The Degronome of KLHDC2 in *E. coli* Cells. This degronome includes peptides that originate from a variety of proteins, including those from *E. coli* and those introduced using molecular biology (Table 1). *E. coli* doesn't have a UPS or encode KLHDC2, so the peptides that originate from *E. coli* proteins may not directly inform the role of KLHDC2 in humans. However, these *E. coli* derived peptides may provide insights into KLHDC2 function in humans. For instance, peptide *N* is the extreme C-terminal fragment of UPF0441 protein YgiB, whereas all other peptides originate from the interior or close to the N-termini of the associated proteins

(Table S3). This suggests that KLHDC2 may recognize proteins that natively feature diglycine at their extreme C-terminus, in addition to protein fragments.

Interestingly, many of the peptides originated from KLHDC2. There are six interspersed diglycines within KLHDC2 (Figure 6A, one for each of the Kelch repeats of the protein) and interacting peptides terminating in four of the six diglycines were identified (Figure 6B). To determine the binding affinity of ASDEGEVIVFGG, which is the most-abundant peptide originating from KLHDC2, we used a luminescence-based competition assay (AlphaScreen, see *Supporting Information* and Figure S8). In this experiment, free ASDEGEVIVFGG peptide competed for binding to KLHDC2 against immobilized HLRGSPPPMAGG, which enables a direct comparison with the other degron peptides that have been characterized.¹⁹ Because the concentration of immobilized components is extremely low relative to that of the free peptide, the IC₅₀ values determined here are approximately equal to the corresponding K_d values.⁵³ Strikingly, this internal KLHDC2 peptide binds almost as tightly (IC₅₀ = 5.3 nM) as the recently characterized selenoprotein fragment (HLRGSPPPMAGG, IC₅₀ = 3.4 nM).¹⁹ An intriguing, albeit speculative, possibility is that the high affinity between KLHDC2 and its internal fragments may enable KLHDC2 to clear its own proteolytic products through the UPS.



Although all identified peptides are terminated by diglycine or glycyllalanine (Table 1), the preceding amino acids in those peptides are diverse (Figure 6D). This finding is consistent with the extremely tight binding of both ASDEGEVIVFGG and HLRGSPPPMAGG, as well as the previous report that KLHDC2 binds degron peptides through non-covalent interactions between the peptide backbone carbonyls and the binding pocket of KLHDC2, rather than specific interactions with amino acid side chains of the peptide.¹⁹ However, if only the presence of diglycine or glycyllalanine were required, the *E. coli* peptides in this degronome are much less diverse than might have been considered possible. The *E. coli* proteome contains ~4300 proteins, ~2700 of which contain a total of nearly 7000 diglycines and ~3000 of which contain ~8000 glycyllalanines. The peptides identified here originate from only 12 of those proteins, which based on comparisons with a meta-analysis of results for *E. coli* K-12,⁵⁵ have cellular concentrations that likely span several orders of magnitude (Figure S9). This finding indicates that KLHDC2 has particularly high affinity for these degrons or that only a small number of the candidate proteins undergo proteolytic digestion that exposes these C-terminal motifs. Therefore, this degronome may be useful for extending the profile of sequences that can be recognized by KLHDC2, *i.e.*, treat the *E. coli* proteome as a peptide library that can be recognized by this E3. For instance, if the last 5 amino acids from each peptide originating from an *E. coli* protein represents a functional degron, KLHDC may have hundreds of potential protein substrates across the human proteome (Table S4). In the future, applying the integrative MS strategy reported here to study the degronomes of E3s in human cells may enable even more direct insights into the roles of those E3s in human biology.

Conclusions

We demonstrated an integrative MS strategy for identifying and quantifying cellular peptides that interacted with an E3 ubiquitin ligase (Figure 3). Unlike other methods for identifying E3 substrates, native MS provides direct evidence of E3:peptide binding. Combining results from native MS (Figure 1), native top-down MS (Figure 2), MS² of destabilized samples, and liquid chromatography MS² revealed a near complete fraction of the KLHDC2-binding peptidome from *E. coli* (Figure 4 and Table 1). These results demonstrate that KLHDC2 binds to peptides containing C-terminal diglycine or glycylalanine, and that a wide profile of sequences preceding the C-terminus are recognized (Figure 6D). Using the native mass spectra and the comprehensive list of identified peptides, we demonstrated a novel and robust method for determining the relative abundance of interacting molecules (Figure 4B). The direct and quantitative data enabled by this integrative MS strategy offers many advantages relative to existing tools for characterizing E3-substrate pairing. Given the persistent challenges associated with identifying and validating substrates for the hundreds of E3 ligases in humans, we anticipate that MS-based degromics will be leveraged more broadly to determine the individual roles of other E3s, including those that adhere to the more common N-end rule⁴² for substrate recognition.

Supporting Information

Determining the binding affinity of ASDEGEVIVFGG, Discussion of ambiguous peptide assignments, Figures S1 to S9, Tables S1 to S4, and Appendix: Determining Relative Abundances.

511 **Conflicts of Interest**

512 The authors declare no competing financial interest.

513

514 **Acknowledgements**

515 Research reported in this publication was supported by the National Science Foundation
516 under award number CHE-1807382 (M. F. B.), the National Institutes of Health under award
517 number T32GM008268 (training grant to D. C.), and the Howard Hughes Medical Institute (N.
518 Z.). The authors thank David Shteynberg (Institute for Systems Biology) for useful discussions.
519 and Priska Von Haller and the University of Washington Proteomics Resource for assistance
520 with LC-MS experiments.

References

- (1) Goldberg, A. L. Protein Degradation and Protection against Misfolded or Damaged Proteins. *Nature* **2003**, 426 (6968), 895–899. <https://doi.org/10.1038/nature02263>.
- (2) Deshaies, R. J.; Joazeiro, C. A. P. RING Domain E3 Ubiquitin Ligases. *Annu. Rev. Biochem.* **2009**, 78 (1), 399–434. <https://doi.org/10.1146/annurev.biochem.78.101807.093809>.
- (3) Glickman, M. H.; Ciechanover, A. The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiol. Rev.* **2002**, 82 (2), 373–428. <https://doi.org/10.1152/physrev.00027.2001>.
- (4) Iconomou, M.; Saunders, D. N. Systematic Approaches to Identify E3 Ligase Substrates. *Biochem. J.* **2016**, 473 (22), 4083–4101. <https://doi.org/10.1042/BCJ20160719>.
- (5) Mark, K. G.; Loveless, T. B.; Toczyski, D. P. Isolation of Ubiquitinated Substrates by Tandem Affinity Purification of E3 Ligase–Polyubiquitin-Binding Domain Fusions (Ligase Traps). *Nat. Protoc.* **2016**, 11, 291. <https://doi.org/10.1038/nprot.2016.008>.
- (6) Guo, Z.; Song, E.; Ma, S.; Wang, X.; Gao, S.; Shao, C.; Hu, S.; Jia, L.; Tian, R.; Xu, T.; et al. Proteomics Strategy to Identify Substrates of LNX, a PDZ Domain-Containing E3 Ubiquitin Ligase. *J. Proteome Res.* **2012**, 11 (10), 4847–4862. <https://doi.org/10.1021/pr300674c>.
- (7) Benanti, J. A.; Cheung, S. K.; Brady, M. C.; Toczyski, D. P. A Proteomic Screen Reveals SCFGrr1 Targets That Regulate the Glycolytic–Gluconeogenic Switch. *Nat. Cell Biol.* **2007**, 9, 1184. <https://doi.org/10.1038/ncb1639>.

- 542 (8) Guo, Z.; Wang, X.; Li, H.; Gao, Y. Screening E3 Substrates Using a Live Phage Display
543 Library. *PLoS One* **2013**, *8* (10), e76622–e76622.
544 <https://doi.org/10.1371/journal.pone.0076622>.
- 545 (9) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., 3rd. Protein Analysis by
546 Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394.
547 <https://doi.org/10.1021/cr3003533>.
- 548 (10) Kim, W.; Bennett, E. J.; Huttlin, E. L.; Guo, A.; Li, J.; Possemato, A.; Sowa, M. E.; Rad,
549 R.; Rush, J.; Comb, M. J.; et al. Systematic and Quantitative Assessment of the Ubiquitin-
550 Modified Proteome. *Mol. Cell* **2011**, *44* (2), 325–340.
551 <https://doi.org/10.1016/j.molcel.2011.08.025>.
- 552 (11) Wagner, S. A.; Beli, P.; Weinert, B. T.; Nielsen, M. L.; Cox, J.; Mann, M.; Choudhary, C.
553 A Proteome-Wide, Quantitative Survey of in Vivo Ubiquitylation Sites Reveals
554 Widespread Regulatory Roles. *Mol. Cell. Proteomics* **2011**, *10* (10), M111.013284-
555 M111.013284. <https://doi.org/10.1074/mcp.M111.013284>.
- 556 (12) Burande, C. F.; Heuzé, M. L.; Lamsoul, I.; Monsarrat, B.; Uttenweiler-Joseph, S.; Lutz, P.
557 G. A Label-Free Quantitative Proteomics Strategy to Identify E3 Ubiquitin Ligase
558 Substrates Targeted to Proteasome Degradation. *Mol. Cell. Proteomics* **2009**, *8* (7), 1719–
559 1727. <https://doi.org/10.1074/mcp.M800410-MCP200>.
- 560 (13) Hör, S.; Ziv, T.; Admon, A.; Lehner, P. J. Stable Isotope Labeling by Amino Acids in Cell
561 Culture and Differential Plasma Membrane Proteome Quantitation Identify New
562 Substrates for the MARCH9 Transmembrane E3 Ligase. *Mol. Cell. Proteomics* **2009**, *8*
563 (8), 1959–1971. <https://doi.org/10.1074/mcp.M900174-MCP200>.

- 564 (14) Yen, H.-C. S.; Xu, Q.; Chou, D. M.; Zhao, Z.; Elledge, S. J. Global Protein Stability
565 Profiling in Mammalian Cells. *Science* **2008**, 322 (5903), 918.
566 <https://doi.org/10.1126/science.1160489>.
- 567 (15) Yen, H.-C. S.; Elledge, S. J. Identification of SCF Ubiquitin Ligase Substrates by Global
568 Protein Stability Profiling. *Science* **2008**, 322 (5903), 923.
569 <https://doi.org/10.1126/science.1160462>.
- 570 (16) Koren, I.; Timms, R. T.; Kula, T.; Xu, Q.; Li, M. Z.; Elledge, S. J. The Eukaryotic
571 Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell* **2018**,
572 173 (7), 1622-1635.e14. <https://doi.org/10.1016/j.cell.2018.04.028>.
- 573 (17) Lin, H.-C.; Yeh, C.-W.; Chen, Y.-F.; Lee, T.-T.; Hsieh, P.-Y.; Rusnac, D. V.; Lin, S.-Y.;
574 Elledge, S. J.; Zheng, N.; Yen, H.-C. S. C-Terminal End-Directed Protein Elimination by
575 CRL2 Ubiquitin Ligases. *Mol. Cell* **2018**, 70 (4), 602-613.e3.
576 <https://doi.org/10.1016/j.molcel.2018.04.006>.
- 577 (18) Lin, H.-C.; Ho, S.-C.; Chen, Y.-Y.; Khoo, K.-H.; Hsu, P.-H.; Yen, H.-C. S. CRL2 Aids
578 Elimination of Truncated Selenoproteins Produced by Failed UGA/Sec Decoding. *Science*
579 **2015**, 349 (6243), 91–95. <https://doi.org/10.1126/science.aab0515>.
- 580 (19) Rusnac, D.-V.; Lin, H.-C.; Canzani, D.; Tien, K. X.; Hinds, T. R.; Tsue, A. F.; Bush, M.
581 F.; Yen, H.-C. S.; Zheng, N. Recognition of the Diglycine C-End Degron by
582 CRL2KLHDC2 Ubiquitin Ligase. *Mol. Cell* **2018**, 72 (5), 813-822.e4.
583 <https://doi.org/10.1016/j.molcel.2018.10.021>.
- 584 (20) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J. Am. Soc.*
585 *Mass Spectrom.* **2017**, 28 (1), 5–13. <https://doi.org/10.1007/s13361-016-1545-3>.

- 586 (21) Marcoux, J.; Wang, S. C.; Politis, A.; Reading, E.; Ma, J.; Biggin, P. C.; Zhou, M.; Tao,
587 H.; Zhang, Q.; Chang, G.; et al. Mass Spectrometry Reveals Synergistic Effects of
588 Nucleotides, Lipids, and Drugs Binding to a Multidrug Resistance Efflux Pump. *Proc.*
589 *Natl. Acad. Sci. USA* **2013**, *110* (24), 9704. <https://doi.org/10.1073/pnas.1303888110>.
- 590 (22) Dyachenko, A.; Gruber, R.; Shimon, L.; Horovitz, A.; Sharon, M. Allosteric Mechanisms
591 Can Be Distinguished Using Structural Mass Spectrometry. *Proc. Natl. Acad. Sci. USA*
592 **2013**, *110* (18), 7235. <https://doi.org/10.1073/pnas.1302395110>.
- 593 (23) Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler,
594 H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L.; et al. Top-down
595 Characterization of Endogenous Protein Complexes with Native Proteomics. *Nat. Chem.*
596 *Biol.* **2017**, *14*, 36. <https://doi.org/10.1038/nchembio.2515>.
- 597 (24) Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.;
598 Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W.; et al. Mass Spectrometry
599 Reveals Modularity and a Complete Subunit Interaction Map of the Eukaryotic
600 Translation Factor EIF3. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (47), 18139.
601 <https://doi.org/10.1073/pnas.0801313105>.
- 602 (25) Stengel, F.; Baldwin, A. J.; Bush, M. F.; Hilton, G. R.; Lioe, H.; Basha, E.; Jaya, N.;
603 Vierling, E.; Benesch, J. L. P. Dissecting Heterogeneous Molecular Chaperone Complexes
604 Using a Mass Spectrum Deconvolution Approach. *Chem. Biol.* **2012**, *19* (5), 599–607.
605 <https://doi.org/10.1016/j.chembiol.2012.04.007>.
- 606 (26) Snijder, J.; van de Waterbeemd, M.; Damoc, E.; Denisov, E.; Grinfeld, D.; Bennett, A.;
607 Agbandje-McKenna, M.; Makarov, A.; Heck, A. J. R. Defining the Stoichiometry and

608 Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. *J. Am.*
609 *Chem. Soc.* **2014**, *136* (20), 7295–7299. <https://doi.org/10.1021/ja502616y>.

610 (27) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.;
611 Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; et al. How Many Human
612 Proteoforms Are There? *Nat. Chem. Biol.* **2018**, *14* (3), 206–214.
613 <https://doi.org/10.1038/nchembio.2576>.

614 (28) Skinner, O. S.; Havugimana, P. C.; Haverland, N. A.; Fornelli, L.; Early, B. P.; Greer, J.
615 B.; Fellers, R. T.; Durbin, K. R.; Do Vale, L. H. F.; Melani, R. D.; et al. An Informatic
616 Framework for Decoding Protein Complexes by Top-down Mass Spectrometry. *Nat.*
617 *Methods* **2016**, *13*, 237. <https://doi.org/10.1038/nmeth.3731>.

618 (29) Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A. An
619 Integrated Native Mass Spectrometry and Top-down Proteomics Method That Connects
620 Sequence to Structure and Function of Macromolecular Complexes. *Nat. Chem.* **2018**, *10*,
621 139. <https://doi.org/10.1038/nchem.2908>.

622 (30) Bulatov, E.; Martin, E. M.; Chatterjee, S.; Knebel, A.; Shimamura, S.; Konijnenberg, A.;
623 Johnson, C.; Zinn, N.; Grandi, P.; Sobott, F.; et al. Biophysical Studies on Interactions and
624 Assembly of Full-Size E3 Ubiquitin Ligase: SUPPRESSOR OF CYTOKINE
625 SIGNALING 2 (SOCS2)-ELONGIN BC-CULLIN 5-RING BOX PROTEIN 2 (RBX2). *J.*
626 *Biol. Chem.* **2015**, *290* (7), 4178–4191. <https://doi.org/10.1074/jbc.M114.616664>.

627 (31) Tan, X.; Calderon-Villalobos, L. I. A.; Sharon, M.; Zheng, C.; Robinson, C. V.; Estelle,
628 M.; Zheng, N. Mechanism of Auxin Perception by the TIR1 Ubiquitin Ligase. *Nature*
629 **2007**, *446*, 640. <https://doi.org/10.1038/nature05731>.

- (32) Xing, W.; Busino, L.; Hinds, T. R.; Marionni, S. T.; Saiffee, N. H.; Bush, M. F.; Pagano, M.; Zheng, N. SCFFBXL3 Ubiquitin Ligase Targets Cryptochromes at Their Cofactor Pocket. *Nature* **2013**, *496*, 64. <https://doi.org/10.1038/nature11964>.
- (33) Davidson, K. L.; Oberreit, D. R.; Hogan, C. J.; Bush, M. F. Nonspecific Aggregation in Native Electrokinetic Nanoelectrospray Ionization. *Int. J. Mass Spectrom.* **2017**, *420*, 35–42. <https://doi.org/10.1016/j.ijms.2016.09.013>.
- (34) Lucas, S.; Copeland, A.; Lapidus, A.; Glavina Del Rio, T.; Dalin, E.; Tice, H.; Bruce, D.; Goodwin, L.; Pitluck, S.; LaButti, K. M.; et al. Complete Sequence of Escherichia Coli BL21(DE3). EMBL/GenBank/DBJ databases July 2009.
- (35) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: An Open-Source MS/MS Sequence Database Search Tool. *PROTEOMICS* **2013**, *13* (1), 22–24. <https://doi.org/10.1002/pmic.201200439>.
- (36) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry–Based Proteomics. *Nat. Methods* **2017**, *14*, 513.
- (37) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.* **2002**, *74* (20), 5383–5392. <https://doi.org/10.1021/ac025747h>.
- (38) Shteynberg, D.; Deutsch, E. W.; Lam, H.; Eng, J. K.; Sun, Z.; Tasman, N.; Mendoza, L.; Moritz, R. L.; Aebersold, R.; Nesvizhskii, A. I. IProphet: Multi-Level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. *Mol. Amp Cell. Proteomics* **2011**, *10* (12), M111.007690. <https://doi.org/10.1074/mcp.M111.007690>.

- 653 (39) Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Slagel, J.; Sun, Z.; Moritz, R. L. Trans-
654 Proteomic Pipeline, a Standardized Data Processing Pipeline for Large-Scale
655 Reproducible Proteomics Informatics. *PROTEOMICS – Clin. Appl.* **2015**, *9* (7–8), 745–
656 754. <https://doi.org/10.1002/prca.201400164>.
- 657 (40) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. Quantitative Profiling of Differentiation-
658 Induced Microsomal Proteins Using Isotope-Coded Affinity Tags and Mass Spectrometry.
659 *Nat. Biotechnol.* **2001**, *19* (10), 946–951. <https://doi.org/10.1038/nbt1001-946>.
- 660 (41) Powell, M. J. D. An Efficient Method for Finding the Minimum of a Function of Several
661 Variables without Calculating Derivatives. *Comput. J.* **1964**, *7* (2), 155–162.
662 <https://doi.org/10.1093/comjnl/7.2.155>.
- 663 (42) Tasaki, T.; Sriram, S. M.; Park, K. S.; Kwon, Y. T. The N-End Rule Pathway. *Annu. Rev.*
664 *Biochem.* **2012**, *81* (1), 261–289. [https://doi.org/10.1146/annurev-biochem-051710-](https://doi.org/10.1146/annurev-biochem-051710-093308)
665 [093308](https://doi.org/10.1146/annurev-biochem-051710-093308).
- 666 (43) Roepstorff, P.; Fohlman, J. Letter to the Editors. *Biomed. Mass Spectrom.* **1984**, *11* (11),
667 601–601. <https://doi.org/10.1002/bms.1200111109>.
- 668 (44) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of
669 Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9.
670 <https://doi.org/10.1021/ac302789c>.
- 671 (45) Allen, S. J.; Schwartz, A. M.; Bush, M. F. Effects of Polarity on the Structures and Charge
672 States of Native-Like Proteins and Protein Complexes in the Gas Phase. *Anal. Chem.*
673 **2013**, *85* (24), 12055–12061. <https://doi.org/10.1021/ac403139d>.

- 674 (46) Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. Reliable Determinations of
675 Protein–Ligand Interactions by Direct ESI-MS Measurements. Are We There Yet? *J. Am.*
676 *Soc. Mass Spectrom.* **2012**, *23* (3), 431–441. <https://doi.org/10.1007/s13361-011-0311-9>.
- 677 (47) Mirzaei, H.; Regnier, F. Enhancing Electrospray Ionization Efficiency of Peptides by
678 Derivatization. *Anal. Chem.* **2006**, *78* (12), 4175–4183.
679 <https://doi.org/10.1021/ac0602266>.
- 680 (48) Cech, N. B.; Enke, C. G. Practical Implications of Some Recent Studies in Electrospray
681 Ionization Fundamentals. *Mass Spectrom. Rev.* **2002**, *20* (6), 362–387.
682 <https://doi.org/10.1002/mas.10008>.
- 683 (49) Iribarne, J. V.; Dziedzic, P. J.; Thomson, B. A. Atmospheric Pressure Ion Evaporation-
684 Mass Spectrometry. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *50* (3), 331–347.
685 [https://doi.org/10.1016/0020-7381\(83\)87009-0](https://doi.org/10.1016/0020-7381(83)87009-0).
- 686 (50) Hirabayashi, A.; Ishimaru, M.; Manri, N.; Yokosuka, T.; Hanzawa, H. Detection of
687 Potential Ion Suppression for Peptide Analysis in Nanoflow Liquid Chromatography/Mass
688 Spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21* (17), 2860–2866.
689 <https://doi.org/10.1002/rcm.3157>.
- 690 (51) Felitsyn, N.; Peschke, M.; Kebarle, P. Origin and Number of Charges Observed on
691 Multiply-Protonated Native Proteins Produced by ESI. *Int. J. Mass Spectrom.* **2002**, *219*
692 (1), 39–63. [https://doi.org/10.1016/S1387-3806\(02\)00588-2](https://doi.org/10.1016/S1387-3806(02)00588-2).
- 693 (52) Morgner, N.; Robinson, C. V. Massign: An Assignment Strategy for Maximizing
694 Information from the Mass Spectra of Heterogeneous Protein Assemblies. *Anal. Chem.*
695 **2012**, *84* (6), 2939–2948. <https://doi.org/10.1021/ac300056a>.

- 696 (53) Yung-Chi, C.; Prusoff, W. H. Relationship between the Inhibition Constant (KI) and the
697 Concentration of Inhibitor Which Causes 50 per Cent Inhibition (I50) of an Enzymatic
698 Reaction. *Biochem. Pharmacol.* **1973**, 22 (23), 3099–3108. [https://doi.org/10.1016/0006-](https://doi.org/10.1016/0006-2952(73)90196-2)
699 2952(73)90196-2.
- 700 (54) Crooks, G. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, 14 (6), 1188–
701 1190. <https://doi.org/10.1101/gr.849004>.
- 702 (55) Wang, M.; Herrmann, C. J.; Simonovic, M.; Szklarczyk, D.; von Mering, C. Version 4.0
703 of PaxDb: Protein Abundance Data, Integrated across Model Organisms, Tissues, and
704 Cell-Lines. *PROTEOMICS* **2015**, 15 (18), 3163–3168.
705 <https://doi.org/10.1002/pmic.201400441>.
706
707