

Reviewed Preprint

v1 • February 24, 2026

Not revised

For correspondence:

jhdavis@mit.edu

keating@mit.edu

Competing interests: No competing interests declared

Funding: See [page 20](#)

Reviewing editor: Noboru Mizushima, The University of Tokyo, Japan

© 2026, Kosmatka et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

A proteome-wide biochemical screen defines binding determinants of the core autophagy protein LC3B

Jennifer Kosmatka¹, Cong Liu², Jackson C Halpin¹, Daniel Lim¹, Joseph H Davis^{1,2,3}✉, Amy E Keating^{1,2,3,4}✉

¹Department of Biology, Massachusetts Institute of Technology, Cambridge, United States • ²Broad Institute of MIT and Harvard, Massachusetts Institute of Technology, Cambridge, United States • ³Computational and Systems Biology Graduate Program, Massachusetts Institute of Technology, Cambridge, United States • ⁴Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, United States

eLife Assessment

This study is a **valuable** contribution that comprehensively identifies and characterizes LC3B-binding peptides through a bacterial cell-surface display screen covering approximately 500,000 human peptides. The data presented are **solid**, although this approach has limitations (e.g., it cannot assess the effects of post-translational modifications, which are often relevant to LIR-mediated interactions). Validation of the newly identified binding peptides by demonstrating their interactions with full-length proteins in cells would further strengthen this manuscript.

<https://doi.org/10.7554/eLife.110318.1.sa3>

Abstract

Human MAP1LC3B (LC3B) binds proteins involved in autophagy and other cellular processes using a degenerate four-residue short linear motif known as the LC3-interacting region (LIR). Biochemical and structural studies have identified LIRs in many LC3B interaction partners, but the sequence features that contribute to binding have not been systematically explored. To discover peptides that interact with LC3B and deeply profile the key binding determinants, we screened a library of ~500,000 36-residue peptides derived from the human proteome using bacterial cell-surface display. Analysis of the screening data, coupled with structural studies and site-directed mutagenesis, revealed exceptions to the reported LIR motif and a strong preference for negatively charged residues adjacent to the LIR, which we visualized in a newly determined structure of LC3B bound to a peptide isolated in our screen. Guided by our screening data, we designed synthetic LIR-containing peptides that bind LC3B with affinities comparable to the tightest measured natural binder. Finally, we determined that mutations in LC3B commonly thought to abrogate binding of LIR-containing peptides instead alter LC3B binding specificity, leading to enhanced binding of some LIR-containing sequences. Taken together, our results refine the LIR motif definition, expand the network of candidate LC3B interaction partners, and highlight how mutations at the LC3B-LIR interface can modulate affinity and specificity.

Introduction

Intrinsically disordered regions (IDRs) in proteins participate in key cellular processes via interactions with globular protein domains (Dyson and Wright, 2005). Short linear motifs (SLiMs) mediate these interactions using contiguous stretches of 3 – 10 amino acids that undergo disorder-to-order transitions upon binding, typically resulting in interactions with equilibrium dissociation constants (K_D) in the micromolar range (Tomba *et al.*, 2014; Wright and Dyson 2015; Diella 2008; Davey *et al.*, 2012; Van Roey *et al.*, 2014). For most SLiMs, the sequence features necessary and sufficient for binding are incompletely understood. Cataloged regular

expressions, such as those in the Eukaryotic Linear Motif (Kumar *et al.*, 2022) and Motif Map of the Proteome (<https://slim.icr.ac.uk/momap/>) databases, capture the features that best typify known binders but also highlight the limitations of existing motif specifications, which are satisfied by many sequences that fail to bind (Bugge *et al.*, 2020). Although the number of experimentally defined SLiMs has expanded in recent years through the targeted assessment of individual binding partners, more comprehensive, high-throughput approaches are required to refine these recognition patterns and to support the development of prediction tools (Davey *et al.*, 2023; Kumar *et al.*, 2024).

SLiMs mediate interactions in the conserved proteostasis pathway macroautophagy (Popelka, 2020), hereafter referred to as autophagy. This pathway involves the *de novo* formation of a double-membrane organelle known as the autophagosome, which transports cellular contents to the lysosome for degradation (Mizushima *et al.*, 2008). Microtubule-associated protein 1A/1B light chain 3B (LC3B) and its homologs are critical at multiple stages of the autophagic process (Rogov *et al.*, 2023) and participate via interaction with many partners that contain a SLiM termed the LC3-interaction region (LIR) (Birgisdottir *et al.*, 2013). During autophagy, hundreds of LC3B proteins are covalently linked to the growing autophagosome through reversible conjugation to phosphatidylethanolamine embedded in the autophagosomal membrane (Kabeya *et al.*, 2004). LC3B•LIR complexes, enhanced by multivalency and avidity at the membrane surface (Sawa-Makarska *et al.*, 2014; Wurzer *et al.*, 2015; Lee and Davis, 2024), tether cellular components to the autophagosome either directly as cargo or indirectly using selective autophagy receptors to ensure efficient degradation (Gubas and Dikic, 2022). LC3B•LIR interactions are also critical for autophagosome biogenesis, trafficking, and lysosomal fusion (Johansen and Lamark, 2020). Mutations that alter LC3B•LIR interactions have been linked to the development of many diseases, including neurodegenerative disorders, aging, and cancer (Ramesh Babu *et al.*, 2008; Park *et al.*, 2020; Fas *et al.*, 2021; Brennan *et al.*, 2022). LC3B also has important roles in LC3-associated forms of phagocytosis (Florey and Overholtzer, 2012), endocytosis (Heckmann *et al.*, 2019), and micropinocytosis (Sønder *et al.*, 2021).

LC3B is one of six hAtg8 proteins that are homologous to the yeast Atg8 protein and that share a common architecture consisting of a ubiquitin-like β -grasp fold with two additional N-terminal α -helices (Sugawara *et al.*, 2004). The 14 kDa LC3B and its paralogs bind to the information-poor consensus LIR [FWY]₀-X₁-X₂-[LVI]₃ (Chatzichristofi *et al.*, 2023), where X can be any amino acid, via two conserved hydrophobic pockets that form the LIR-Docking site (LDS). In its canonical binding mode, the aromatic residue [FWY]₀ engages the first hydrophobic pocket (HP1), and the aliphatic residue [LVI]₃ engages the second hydrophobic pocket (HP2) (Rogov *et al.*, 2023). This degenerate LIR motif occurs ~170,000 times in the human proteome, with ~19,000 occurrences falling in regions of predicted disorder. It is unlikely that all of these sequences represent functional interaction sites; indeed, several proteins containing canonical LIRs fail to bind LC3B with measurable affinity (Chatzichristofi *et al.*, 2023). Moreover, recent discoveries of binders that deviate from the canonical LIR motif have expanded the repertoire of hAtg8 binding modes, and thus increased the number of potential interactors. For example, Li *et al.*, reported that LC3B binds to ankyrin-G using the LIR sequence ¹⁹⁹¹WIEF¹⁹⁹⁴, wherein an aromatic Phe residue engages HP2 (Li *et al.*, 2018), and Knævelsrud *et al.*, found that sorting nexin-18 (SNX18) binds hAtg8 paralogs using an expanded five-residue motif (¹⁵⁴WDDEW¹⁵⁸) (Knævelsrud *et al.*, 2013). The proteome-wide prevalence of binders that lack the canonical [FWY]₀-X₁-X₂-[LVI]₃ motif remains unknown.

Given that not all instances of the LIR motif bind to LC3B, and that not all LC3B interactors contain a LIR motif, we sought to understand determinants beyond the core LIR that contribute to binding. To discover these sequence determinants, we used a bacterial display assay to screen ~500,000 peptides derived from the human proteome for those capable of binding to LC3B. We determined the binding affinity of 51 peptides from the screen and elucidated binding mechanisms that support SLiM-LC3B interactions for known and novel binders through structural and biochemical analysis of identified peptides and mutated variants. Our results expand the number and types of

residues that can engage HP2 on LC3B and strongly support the role of N-terminal acidic residues in enhancing LIR•LC3B binding affinity, allowing us to design a synthetic LIR peptide with affinity comparable to that of the tightest known native LC3B binder.

Results

A high-throughput bacterial display screen identifies thousands of LC3B-binding peptides

To rapidly identify peptides capable of binding LC3B, we used bacterial surface display in combination with fluorescence-activated cell sorting (FACS). Following Hwang *et al.*, we expressed ~500,000 36-mer peptides spanning the human proteome on the cell surface of *Escherichia coli* via fusion to circularly permuted OmpX (Rice *et al.*, 2006; Larman *et al.*, 2011; Hwang *et al.*, 2022).

To compensate for the low micromolar affinity of typical SLiM interactions, and to better mimic the avid interactions enabled by the high local concentration of LC3B conjugated to autophagosomal membranes in cells (Zaffagnini and Martens, 2016), we bound N-terminally biotinylated LC3B to tetravalent streptavidin-phycoerythrin (SA-PE). Using fluorescence-activated cell sorting (FACS), we quantified peptide expression levels using an allophycocyanin (APC)-conjugated antibody against an encoded N-terminal FLAG tag. Cells displaying peptides that bound to LC3B were identified and isolated based on their high SA-PE-to-APC ratio (Figure 1A). We validated our screening approach using well-characterized LC3B-binding peptides of various affinities (Supplementary Figure 1). Following five rounds of positive selection and one round of counterselection against non-specific binding to SA-PE (Figure 1B), cells in the binding gate were enriched ~80-fold, from ~0.7% of the naïve library to ~57% of all peptide-expressing cells (Supplementary Figure 2).

In all, 12,158 peptides derived from 5,578 unique proteins were identified in the final round of library sorting, with individual peptides displaying a wide range of behaviors as assessed by their calculated enrichment ratio (ER) (Rubin *et al.*, 2017) across successive sorts (Figure 1C, see Methods). To define a high-confidence set (HC-set) of LC3B binders for further analysis, we selected peptides with an mean ER at least 1.70 standard deviations (z -score ≥ 1.70 , $n = 427$ peptides) above the mean ER calculated over the final three rounds of sorting (Figure 1D-E). This criterion captured four of the six peptides in our input library that were reported to bind LC3B with dissociation constants below 5 μ M, as compiled in the LIRcentral database (Chatzichristofi *et al.*, 2023) (Supplementary Figure 3A). In contrast, peptides in our input library annotated by LIRcentral as non-binding ($K_D > 50 \mu$ M) were strictly excluded from this group and were progressively depleted during sorting (Supplementary Figure 3B). Beyond these known interactions, this HC-set included peptides from sixteen proteins previously reported to co-immunoprecipitate with LC3B (Stark *et al.*, 2006), most of which lacked mapped interaction sites (Figure 2A). We tested four such peptides (MAP1B⁸²³⁻⁸⁵⁸, ATG4A³⁶³⁻³⁹⁸, SCYL1⁶⁴⁰⁻⁶⁷⁵, and HEAT3³⁷⁷⁻⁴¹²) for binding to monomeric LC3B by bio-layer interferometry (BLI) and observed robust binding with dissociation constants ranging from approximately 0.5 to 10 μ M (Figure 2B, Supplementary Table 1).

Strongly enriched peptides include novel candidate LC3B interactors

To identify peptide hits from our screen most likely to represent biologically relevant LC3B interaction partners, we selected proteins with shared Gene Ontology (GO) annotations (The Gene Ontology Consortium, 2023) with LC3B that contained peptides in our HC-set (Figure 2A, Supplementary Figure 4). From this list, we tested LC3B for binding to two candidate peptides derived from HERC1³⁰⁷⁵⁻³¹¹⁰ and LRRK2⁸⁵⁸⁻⁸⁹³, noting that these proteins are reported to impact autophagy (Montes-Fernández *et al.*, 2020; Pérez-Villegas *et al.*, 2020; Park *et al.*, 2016;

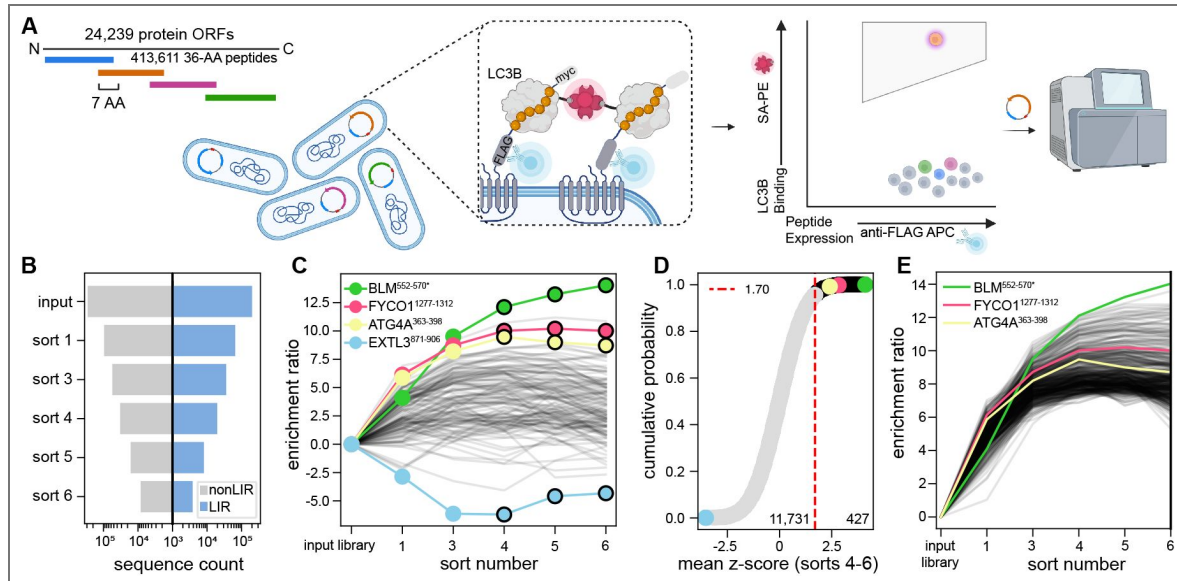


Figure 1. Human peptidome library screening enriches known and novel LC3B-binding peptides.

(A) Schematic of FACS-based bacterial display enrichment screening. Peptides, 36 amino acids in length, that tile the human proteome were expressed as FLAG-tagged fusions to eCPX on the surface of *E. coli*. Expression was detected using the fluorescence signal from an allophycocyanin (APC)-conjugated anti-FLAG antibody. Binding to LC3B (grey), which was tetramerized through binding to streptavidin conjugated to phycoerythrin (SA-PE, pink), was detected using PE fluorescence. Peptide-expressing cells were sorted based on APC and PE fluorescence and sequenced. Panel created using *BioRender.com*. **(B)** Diverging bar chart plotting the number of unique sequences detected in each round of sorting that lacked (grey, left axis) or contained (right, blue) a LIR motif. Sort 2, a negative sort used to eliminate peptides nonspecifically bound to SA-PE, was not sequenced. **(C)** Peptide enrichment profiles across enrichment sorts. Black lines show trajectories for a random subset (0.1%) of all peptides that persisted to sort 6. Overlaid are the enrichment profiles for the best-enriching peptide (BLM^{552-570*}), positive controls (FYCO1¹²⁷⁷⁻¹³¹² and ATG4A³⁶³⁻³⁹⁸), and the worst-enriching peptide that reached sort 6 (EXTL3⁸⁷¹⁻⁹⁰⁶) colored green, pink, yellow, and blue, respectively. Peptides sequences detailed in Supplementary Table 1. **(D)** Cumulative density function plot of the mean z-scores across sorts 4, 5, and 6. Mean z-scores for the 427 peptides of 12,158 that surpassed the threshold of 1.70 that were used to define the HC-set are colored in black, and select peptides are colored as in panel C. **(E)** Peptide enrichment profiles for the 427 peptides in the HC-set (black), with select peptides colored as in panels C-D.

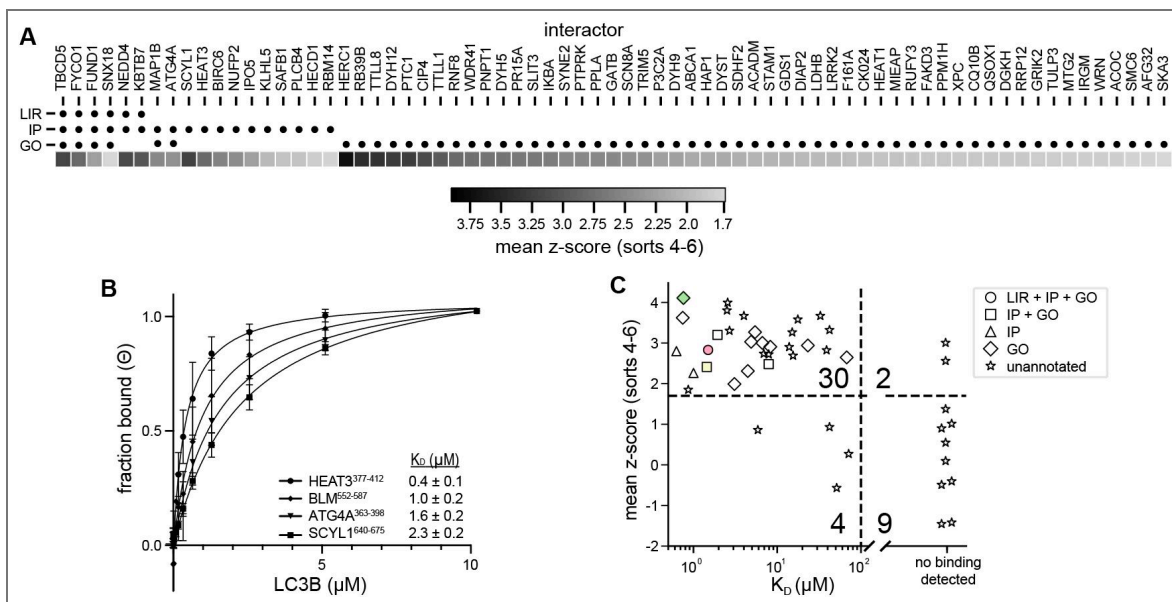


Figure 2. Highly enriched peptides share annotations with LC3B and bind LC3B tightly.

(A) Peptides in the HC-set, annotated with various lines of evidence of binding to LC3B. For each protein, circles indicate: LIR - experimentally validated LIR motifs annotated in LIRcentral (Chatzichristofi *et al.*, 2023); IP - proteins co-purified with LC3B, from BioGrid (Stark *et al.*, 2006); GO - peptides with GO-annotations shared with LC3B, from Uniprot (The UniProt Consortium, 2023), with heatmap of enrichment z-score plotted at bottom. **(B)** Binding to monomeric LC3B assayed by BLI for peptides ATG4A³⁶³⁻³⁹⁸ (VPPAKPEVTTTGAEFIDSTEQLLEFDLEEDFEILSV), SCYL1⁶⁴⁰⁻⁶⁷⁵ (TADRWDDDEDWGSLEQEAESVLAQQDDWSTGGQVSRA), HEAT3³⁷⁷⁻⁴¹² (EDPSDDEWEELSSSDESDAFMENSFSECGGQLFSPL), and BLM⁵⁵²⁻⁵⁸⁷ (DIDNFDIDFDDDDWEDIMHNLAASKSSTAAYQPI). Assay performed in triplicate, with error bars indicating standard deviation. Data fit to a standard binding isotherm: $\theta = \frac{[LC3B]}{[LC3B] + K_d}$. **(C)** Relationship between mean z-score over rounds 4-6 and affinity for monomeric LC3B as determined by BLI. Four peptides are colored as in Figure 1C. Shapes indicate five classes of peptides, as annotated in panel A. Dotted lines mark thresholds delineating high vs. low z-scores and binding vs. non-binding, with the number of peptides in each quadrant listed. Enrichment of binders among high z-score peptides was found to be statistically significant, as assessed using Fisher's exact test ($p < 0.001$).

Roosen and Cookson 2016 [↗](#); Madureira *et al.*, 2020 [↗](#); Boecker and Holzbaur 2021 [↗](#)), but have not been shown to directly interact with LC3B. Each peptide bound with a dissociation constant in the low micromolar range (Supplementary Table 1).

We additionally identified peptide hits without reported links to autophagy or LC3B. Many of these peptides bound LC3B with moderate affinity ($K_D \leq 50 \mu\text{M}$) and thus represent candidates for new LC3B-interacting proteins (Supplementary Table 1). In total, we determined monomeric LC3B binding affinities for 45 peptides spanning varying degrees of enrichment in our sort and found that the z-score cutoff of 1.70 used to define our HC-set effectively distinguished binders ($K_D < 50 \mu\text{M}$) from non-binders ($K_D > 60 \mu\text{M}$) (Figure 2C [↗](#), Supplementary Table 2).

Enriched LIR-containing sequences exhibit a preference for Trp in the X_0 position, preceded by acidic residues

To identify sequence features associated with binding, we generated a sequence logo (Schneider and Stephens, 1990 [↗](#); O'Shea *et al.*, 2013 [↗](#)) for peptides in the HC-set that contained a LIR motif (see Methods). This logo highlighted enrichment of tryptophan in the first position (X_0), over tyrosine and phenylalanine, and a preference for acidic residues in positions X_{-1} , X_{-2} , and X_{-3} , as well as for glutamate in the X_1 position of the core LIR motif and at the C-terminal X_7 position (Figure 3A [↗](#)). Consistent with this logo capturing features related to binding affinity, a synthetic peptide defined by the consensus sequence observed in the logo, pCONS_{LIR}, bound to LC3B with a dissociation constant of $\sim 60 \text{ nM}$ (Figure 3B [↗](#)). This affinity is ~ 25 -fold tighter than a well-studied peptide from FYCO1¹²⁷⁷⁻¹³¹² (Olsvik *et al.*, 2015 [↗](#); Cheng *et al.*, 2016 [↗](#)), ~ 25 -fold tighter than a chimeric peptide designed to target α -synuclein for autophagic degradation (Tong *et al.*, 2023 [↗](#)), and is comparable in affinity to the tightest known LC3B-binding peptide, ANK2¹⁵⁷⁸⁻¹⁶¹³ (Li *et al.*, 2018 [↗](#)).

To test the role of the hydrophobic residue in position X_0 , we partitioned the sequence alignments based on the W/F/Y residue identity and generated three new peptides (pCONS_W, pCONS_F, and pCONS_Y) based on the consensus sequences (Supplementary Figure 5). Because Y-type LIR sequences were frequently located near the C-terminus in our HC-set, we lacked sufficient sequence diversity to generate a logo with C-terminal flanking sequences for this region, and the LIR motif of pCONS_Y was placed at the C-terminus of the peptide to reflect this positional bias. Peptide pCONS_W bound to LC3B with similar affinity ($K_D \sim 90 \text{ nM}$) as peptide pCONS_{LIR} (Supplementary Table 3). The affinities of pCONS_F and pCONS_Y were both weaker ($K_D \sim 400 \text{ nM}$, $K_D \sim 10 \mu\text{M}$, respectively) (Supplementary Table 3), consistent with tryptophan in the first position (X_0) favoring binding.

We next tested the contributions of the highly enriched N- and C-terminal acidic residues in the context of the tightly binding pCONS_W peptide. Truncation of C-terminal residues beyond position X_5 had modest effects on binding, but removing residues N-terminal to position X_2 weakened affinity ~ 7 -fold (Figure 3C [↗](#), Supplementary Table 3). Most of this reduced affinity could be restored by reintroducing either a single N-terminal Asp or a 6-residue EDDDDA sequence that lacked this Asp (Figure 3C [↗](#), Supplementary Table 3). These data support an important role for flanking N-terminal acidic residues in facilitating high-affinity binding to LC3B, without a strict requirement for a specific residue at a specific site. Consistent with the key determinants of binding being proximal to the LIR, removing 7 residues from either the N- or C-terminus of the 36-residue peptide identified in the screen did not significantly impact affinity (Figure 3C [↗](#), Supplementary Table 3).

Given the efficacy of this approach in predicting tightly binding peptides, we next evaluated whether a related method could distinguish LC3B-binding from non-binding peptides that contained LIR motifs. Briefly, we assembled a test set of 79 binders and 56 non-binders as annotated in LIRcentral (Chatzichristofi *et al.*, 2023 [↗](#)) and compared the discriminatory performance of the previously reported iLIR model (Kalvari *et al.*, 2014 [↗](#)) with new PSSM-based models derived from our screening data (see Methods). We also constructed higher-order random forest classifiers using our screening data that aimed to capture potential residue-residue

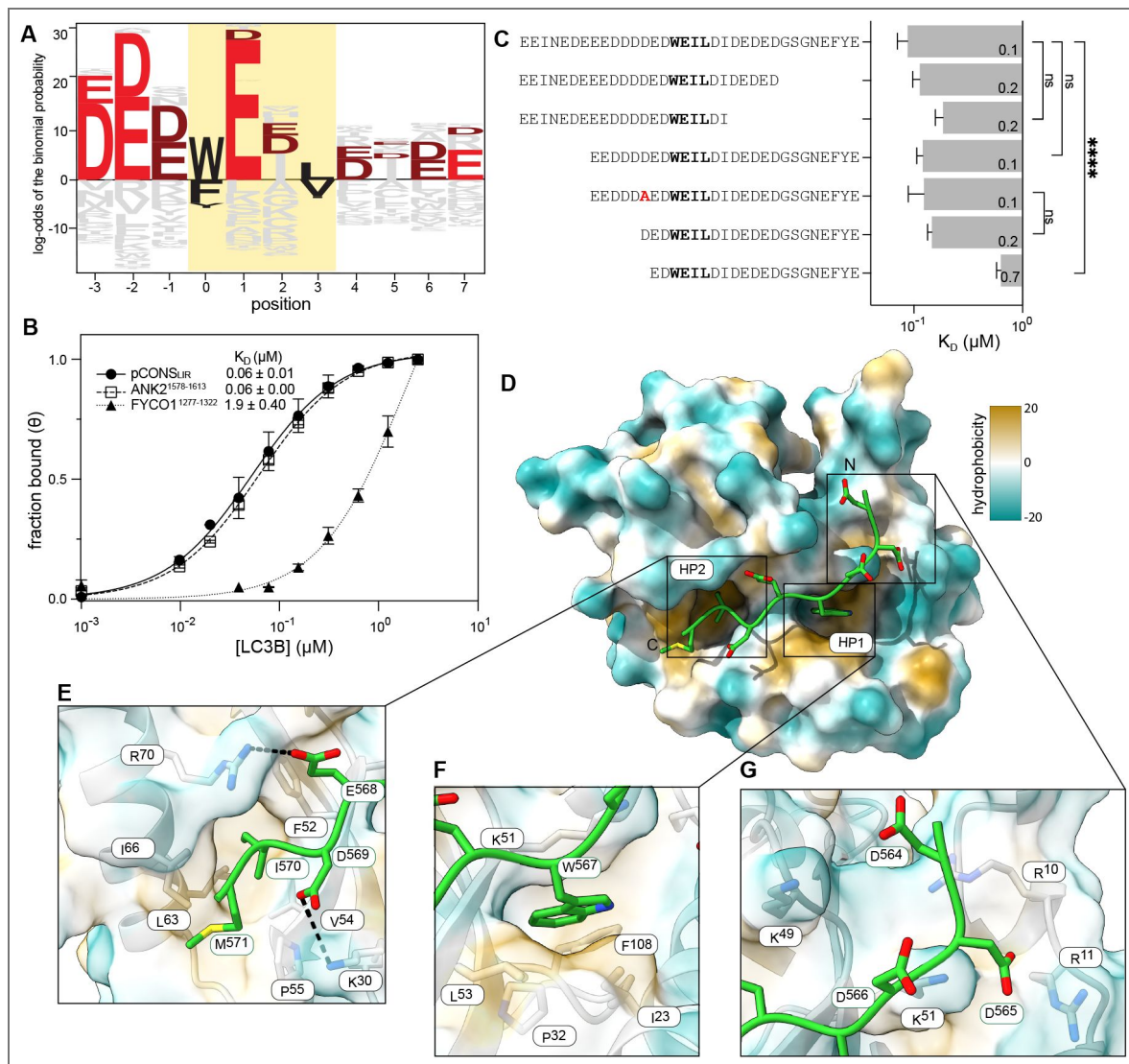


Figure 3. Highly enriched peptides feature W-type core LIR motif flanked by acidic residues.

(A) Sequence logo derived from analysis of the HC-set peptides (see Methods), and plotted using logomaker (Tareen and Kinney, 2019). Acidic residues are colored maroon and those colored red are significantly enriched compared to the input library. Other residues are colored gray with [FWY]₀ and [LVI]₃ in black. The core LIR is highlighted in yellow. (B) BLI measurements of binding affinity between LC3B and peptides pCONSw (EVEVEKEEDDDDEWEILDIEEGSDSEQLISE), ANK2¹⁵⁷⁸⁻¹⁶¹³ (VQSSRSERGLVEEWWIVSDDEIEEARQKAPLEITE), and FYCO1¹²⁷⁷⁻¹³²² (DAVFDIITDEELCQIQESGSLPETPTETDSDLPNA). Error bars report the standard deviation of two or more technical replicates. Data fit to a standard binding isotherm, as in Figure 1. (C) Bar chart plotting binding affinities (K_D) of sequential truncations of pCONSw measured via BLI, with error bars reporting standard error of the mean. Brackets report statistical significance of pairwise t-tests (**** $p \leq 0.0001$; ns not significant). (D) Structure of BLM^{552-571*} bound to LC3B, resolved to 2.2 Å resolution (see Supplementary Table 4). Here, and in panels E-G, the BLM peptide is displayed in stick representation (green backbone) and the LC3B surface is colored by hydrophobicity, as computed by ChimeraX (Meng *et al.*, 2023). Hydrophobic pockets HP1 and HP2, and the N-terminal flanking residues, are indicated. (E) Inset highlighting contact between BLM E568 and LC3B R70, and BLM D569 and LC3B K30, near HP2 (black dashed lines). LC3B residues that form HP2 are annotated and shown in stick representation. (F) Inset highlighting docking of BLM W567 in LC3B HP1, with residues forming that pocket annotated and shown in stick representation. (G) Inset highlighting interactions between BLM N-terminal acidic residues D564 and D566 with R11, K49, and K51 of LC3B. Putative contacts within 4.5 Å marked with dashed black lines.

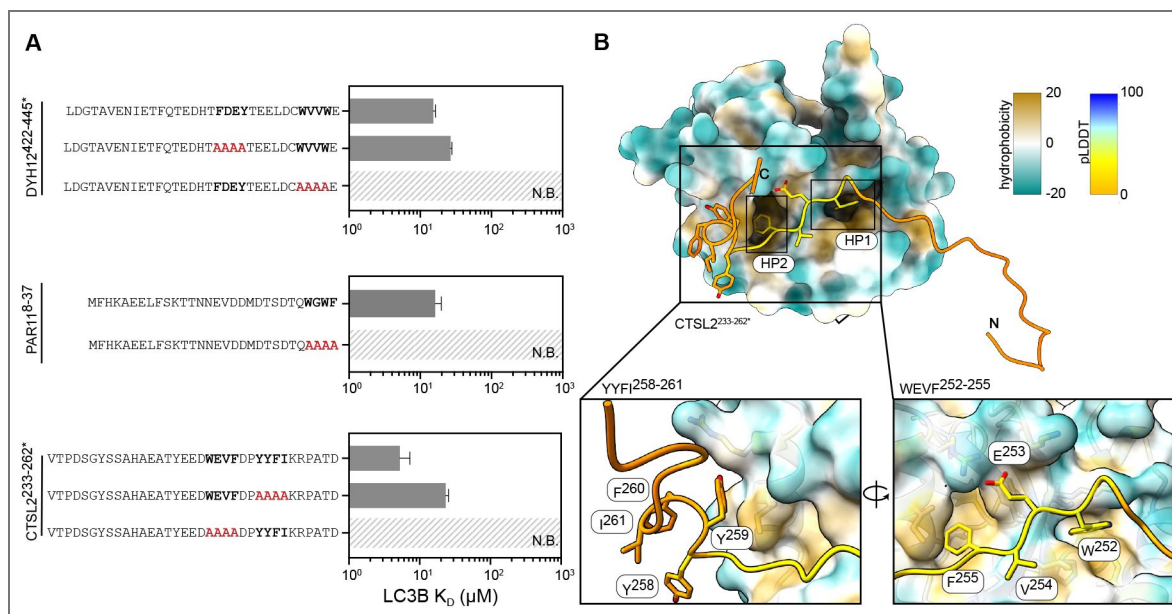


Figure 4. LIR+ motifs support binding to LC3B.

(A) Affinities of LC3B for peptides from PAR11⁸⁻³⁷, CTSL2^{233-262*}, and DYH12^{422-445*}, as measured by BLI, plotted as a bar chart (mean $\pm$ s.e.m, $n \geq 3$). Candidate binding sites that were mutated to Ala are red in the mutated sequence. No detectable binding, up to 40 μM LC3B, is indicated with N.B. **(B)** AlphaFold3 (Abramson *et al.*, 2024) structural prediction of CTSL2^{233-262*} (colored by pLDDT) in complex with LC3B (colored by hydrophobicity). Insets: canonical LIR YYF²⁵⁸⁻²⁶¹ is not predicted to engage the LDS (left), in contrast to WEVF²⁵²⁻²⁵⁵ (right).

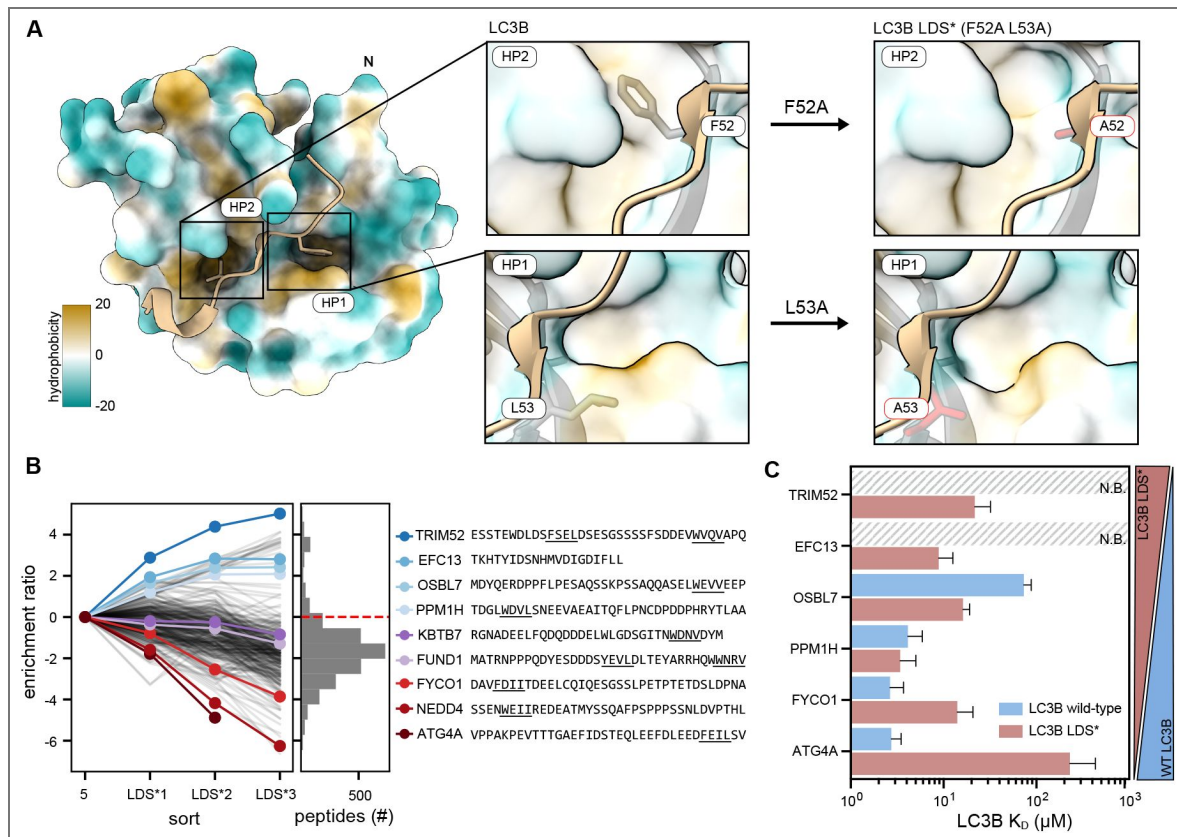


Figure 5. Mutations at the LC3B LIR docking site alter LIR-peptide binding specificity.

(A) Structure of LC3B bound to FYCO1¹²⁷⁶⁻¹²⁸⁸ LIR peptide (PDB 5d94) (Olsvik *et al.* 2015) shown as a surface representation colored by hydrophobicity. Insets depict hydrophobic pockets formed by residues F52, L53 (stick representation) in wild-type LC3B (left), or a predicted surface of the LDS* mutant (right), with each alanine substitution colored red. **(B)** Enrichment profiles across three rounds of the LC3B LDS* enrichment sort, with nine peptides spanning the range of observed profiles colored and sequences provided. Four peptides that were enriched and bound with measurable affinity to LC3B LDS* are in blue. Known, canonical LC3B LDS binders that were depleted in the LDS* sort, are colored purple and red. Black lines indicate a random sample of 10% of peptides that persisted through LC3B LDS* sort 3. The histogram shows the frequencies of peptides with a given enrichment ratio in LDS* sort 3, with the dashed red line marking an enrichment ratio of 0. **(C)** Affinities of LC3B (blue) and LC3B LDS* (light red) for peptides FYCO1¹²⁷⁷⁻¹³¹² and ATG4A³⁶³⁻³⁹⁸ and for four peptides found to enrich in across the LDS* sorts: PPM1H⁴³⁶⁻⁴⁷¹, OSBL7¹⁻³⁶, EFC13^{262-278*}, TRIM52⁸⁸⁻¹²⁰. Error bars indicate standard error of the mean across three or more replicates. Hashed bars mark peptides without detectable binding at the highest measured LC3B concentration (40 μ M).

coupling effects. On this test set (see Methods), no model performed better than the iLIR model (Supplementary Figure 6). Evaluating the models using held-out screening data gave auROC = 0.87 for recognizing LIR peptides with z-score > 2.3, but poor performance for less enriched peptides (Supplementary Figures 7), suggesting that the test set and/or screening data may be too limited, sparse, or noisy, or that key binding determinants depend on higher-order structure that our models fail to capture.

Acidic residues N-terminal of the core LIR directly contribute to LC3B engagement and binding affinity

Many of our top-scoring peptides exhibited features common to both pCONS_{LIR} and pCONS_W, namely the presence of Trp in X₀, Glu in X₁, and multiple N-terminal acidic residues. Indeed, the top-scoring peptide contained these features, and bound tightly to LC3B (K_D ~ 0.7 μM; Supplementary Table 1). Many previously reported high-affinity LC3B binders, including FYCO1¹²⁷⁶⁻¹²⁸⁸, ANK2¹⁵⁸⁸⁻¹⁶¹³, and ANK3¹⁹⁸⁵⁻²⁰¹⁰, use an acidic residue in position X₇ of the C-terminal extension to make an affinity-enhancing contact to LC3B residue R70 (Olsvik *et al.*, 2015 [\[3\]](#); Cheng *et al.*, 2016 [\[4\]](#); Li *et al.*, 2018 [\[5\]](#)) (Supplementary Figure 8), yet our top scoring peptide, derived from the DNA helicase BLM, lacks such a residue. This observation prompted us to investigate how it achieved comparable binding.

We determined the structure of LC3B fused to BLM⁵⁵²⁻⁵⁷¹ at 2.2 Å resolution using X-ray crystallography (Figure 3D-G [\[6\]](#), Supplementary Table 4, PDB 9p3e). BLM⁵⁵²⁻⁵⁷¹* engaged the LDS of LC3B as expected, and the interactions between the peptide and LC3B were consistent with those observed for other canonical LIR motifs: the aromatic W567 (X₀) and hydrophobic residue I570 (X₃) docked in HP1 and HP2, respectively (Figure 3E-F [\[6\]](#)). Additionally, we observed an intermolecular β-sheet between the LC3B residues 51-53 and BLM residues 568-570, which occupy the 1-3 positions of the [FWY]₀-X₁-X₂-[LVI]₃ LIR motif. Notably, the sidechain of BLM E568 (X₁) contacted LC3B R70 (Figure 3E [\[6\]](#)), seemingly substituting for the contacts often found between R70 and glutamates C-terminal to the core LIR (Supplementary Figure 8). Acidic residues N-terminal to the LIR, D565-D567 (X₃ – X₁), were in proximity to positively charged residues R10, R11, K49, and K51 of LC3B (Figure 3G [\[6\]](#)), though they adopted multiple conformations as analyzed by ensemble refinement (Burnley *et al.*, 2012 [\[7\]](#)) (Supplementary Figure 9, see Methods). Together, these observations indicated that acidic residues N-terminal to the core LIR contribute directly to LC3B engagement and binding affinity.

Highly enriched peptides lacking a canonical LIR can bind LC3B

We next asked whether sequences from our HC-set that lacked a canonical LIR could also bind LC3B, focusing on candidates with residues we predicted could engage the HP1 and HP2 pockets. We specifically examined peptides containing a motif we termed LIR+ ([FWY]₀-X₁-X₂-[FWY]₃), as Li *et al.*, had previously reported an X-ray crystal structure showing a peptide bound to LC3B with aromatic residues W and F engaged with the hydrophobic pockets (Li *et al.*, 2018 [\[5\]](#)). To this end, we assessed LC3B binding for three peptides in our HC-set bearing LIR+ sequences: (1) DYH12^{422-445*}, which contains two LIR+ sequences; (2) PAR11⁸⁻³⁷, which contains a single LIR+ sequence, and (3) CTSL2^{233-262*}, which contains both a canonical LIR and a LIR+ motif. We found that each bound LC3B with sub-20 μM affinity as measured by BLI. We then introduced alanine substitutions into each LIR+ sequence and observed reduced affinity for LC3B in each peptide, consistent with a role for the LIR+ motif in supporting LC3B association (Figure 4A [\[8\]](#)). Notably, in CTSL2^{233-262*}, mutation of the LIR+ sequence ²⁴⁶WEVF²⁴⁹ to AAAA strongly reduced binding, whereas analogous alanine substitutes in the canonical ²⁵²YFIF²⁵⁵ LIR motif had a modest effect, consistent with the LIR+ site serving as the dominant binding determinant. Supporting this interpretation, an AlphaFold3 (Abramson *et al.*, 2024 [\[9\]](#)) prediction of a CTSL2^{233-262*}•LC3B complex structure showed the LIR+ motif engaging the LC3B LDS (Figure 4B [\[8\]](#)).

Considering the apparent binding efficacy of such LIR+ sequences, we assessed the relative frequency of canonical LIR, LIR+, and related motifs in our input library and in our HC-set. We found strong evidence for enrichment of LIR sequences, but not for these closely related motifs

(Supplementary Figure 10), indicating that whereas such LIR-related motifs may support binding in some sequence contexts, they are not sufficiently strong binding determinants to be enriched in our screen.

Common mutations to the LIR docking site of LC3B modulate but do not ablate peptide binding

LC3B with mutations F52A and L53A in the LDS, here called LC3B LDS*, is commonly used to assess binding at the LDS, with these substitutions in HP2 and HP1 presumed to disrupt peptide binding at this interface (Figure 5A) (Behrends *et al.*, 2010; Kraft *et al.*, 2014; Qiu *et al.*, 2017; Skytte Rasmussen *et al.*, 2017; Marshall *et al.*, 2019). Using LC3B LDS*, we performed three additional bacterial-display sorts with the round-5-enriched library as input (Supplementary Figure 11, see Methods). The well-studied peptide FYCO1¹²⁷⁷⁻¹³¹² was depleted over the three rounds of LDS* sorting, consistent with its canonical LDS-dependent binding mechanism. Similar behavior was observed for other known LC3B-binding peptides, including KBTB⁶³⁹⁻⁶⁷⁴, FUND1¹⁻³⁶, NEDD4⁶⁸¹⁻⁷¹⁶, and ATG4A³⁶³⁻³⁹⁸ (Figure 5B). Unexpectedly, several other peptides were further enriched over these sorts (Figure 5B), consistent with increased affinity for the altered LC3B LDS* interface.

To investigate this further, we selected six peptides spanning the enrichment patterns observed in the LDS* sorts and quantified their binding to wild-type LC3B or LC3B LDS* proteins using BLI. The results revealed diverse binding behaviors: some peptides bound more tightly to wild-type LC3B (*e.g.*, ATG4A³⁶³⁻³⁹⁸), whereas others exhibited strong selectivity towards the LDS* mutant (*e.g.*, TRIM52⁸⁸⁻¹²⁰) (Figure 5C). Notably, structural modeling indicated that although the LDS* mutations alter the binding interface, the hydrophobic pockets persist (Figure 5A). Therefore, we propose that, rather than abolishing LDS-mediated interactions entirely, the F52A/L53A substitutions subtly remodel the docking surface, creating an alternative hydrophobic interface that modulates peptide binding in a sequence-dependent manner.

Discussion

LC3B and other hAtg8 paralogs often interact with binding partners via the LIR motif [FWY]₀-X₁-X₂-[LVI]₃, an information-poor SLIM that can mediate critical interactions in all major stages of bulk and selective autophagy (Rogov *et al.*, 2023), as well as in non-autophagic processes such as LC3-associated phagocytosis (Florey and Overholtzer 2012) and LC3-associated endocytosis (Heckmann *et al.*, 2019). Here, we used high-throughput bacterial-surface-display screening to broadly assess the LC3B-binding potential of peptides derived from the human proteome, and to define sequence determinants of LC3B recognition. We focused our analysis on a set of strongly enriched peptides (*i.e.*, the HC-set), recognizing that we would miss many authentic but weakly interacting partners. Within the HC-set, we observed roughly threefold enrichment of peptides bearing traditional LIR sequences. Many peptides contained multiple LIR motifs, which may enhance apparent affinity through avid interactions with LC3B oligomerized through the tetravalent streptavidin linkage used in our assay, or through membrane conjugation in cells.

Overall, the vast majority (~90%) of peptides in the HC-set contained a canonical LIR or a related sequence bearing hydrophobic or aromatic residues positioned to plausibly interact with hydrophobic pockets HP1 and HP2 on LC3B. We biochemically characterized one such motif [FWY]₀-X₁-X₂-[WFY]₃, which we termed “LIR+”, finding that it supported robust LC3B binding. Although a small number of LC3B-binding peptides bearing this LIR+ motif have been reported (~1% of cataloged hAtg8-binding peptides) (Li *et al.*, 2018; Chatzichristofi *et al.*, 2023), their prevalence appears underappreciated given their relatively high frequency in the HC-set (Supplementary Figure 10). Further, the binding of LIR+ peptides observed here suggests that other motif variants, such as [ILV]₀-X₁-X₂-[WFY]₃ or [ILV]₀-X₁-X₂-[ILV]₃, may also engage LC3B and, consistent with this notion, such sequences were prevalent in our HC-set (Supplementary Figure 10).

The remaining peptides lacking any LIR-like motifs (~10%) partitioned between those in which the two hydrophobic residues were separated by either 3 or 4 acidic amino acids (3 peptides), similar to the LC3B-binding “WDDEW” motif in SNX18 (Knævelsrud *et al.*, 2013 [↗](#)), those bearing the sequence “HPQ” (3 peptides), which is reported to bind to streptavidin (Weber *et al.*, 1992 [↗](#); Katz 1995 [↗](#)), and a set of peptides lacking linear motifs identifiable through expert-guided inspection or the automated motif discovery tool XSTREME (Grant and Bailey, 2021 [↗](#)).

A sequence logo-based analysis of LIR-bearing peptides in our HC-set revealed enrichment of acidic residues, and biochemical measurements confirmed that acidic residues N-terminal to the core LIR enhance binding affinity without strict positional dependence. These observations complement and expand on previous studies of individual peptides derived from FYCO1 and p62/SQSTM1, for which LC3B binding is supported by N-terminal acidic residues (Pankiv *et al.*, 2007 [↗](#)). They are also aligned with a reported affinity-enhancing role for phosphorylated serine and threonine residues in the N-terminal flanking sequence of the core LIR motif (Richter *et al.*, 2016 [↗](#); Wirth *et al.*, 2021 [↗](#); Kliche *et al.*, 2023 [↗](#)). Indeed, given the impact of N-terminal acidic residues, it is plausible that phosphorylation of target proteins could dynamically modulate LIR•LC3B binding affinity, effectively linking kinase-dependent environmental signaling cascades to the regulation of autophagic cargo selection (Birgisdottir *et al.*, 2013 [↗](#); Kliche and Ivarsson 2022 [↗](#); Rogov *et al.*, 2023 [↗](#)).

LIR databases catalog experimentally validated motifs (Chatzichristofi *et al.*, 2023 [↗](#)) and have been used to define a six-residue PSSM model for predicting binders to hAtg8 proteins and other Atg8 orthologs (Kalvari *et al.*, 2014 [↗](#)). Guided by sequence information from our screening data, we designed a 34-residue synthetic peptide with an affinity comparable to the strongest known interaction partners (Figure 3B [↗](#), Supplementary Table 1). The high-affinity pCONS_{LIR} peptide is characterized by an extended negatively charged N-terminal flanking sequence. Notably, the peptides exhibiting the strongest enrichment in our screen (*i.e.*, *z*-score ≥ 2.3) and those described in the iLIR set of known binders shared this prominent enrichment of acidic N-terminal residues, as evidenced by these sequences co-clustering in UMAP space (Supplementary Figure 6D). However, peptides lacking these specific features can also bind LC3B, and others with similar sequence features do not bind LC3B, underscoring the complex relationship between peptide sequence and LC3B recognition that was not fully captured by the PSSM-based classifiers or by the random forest classifiers we deployed (Supplementary Figures 6-7).

Recently, the Vierstra group reported an alternate binding interface on Atg8 homologs (hAtg8) that they termed the Ubiquitin-interaction-motif Docking Site (UDS). This interface, which occurs on the face opposite that of the LIR docking site, is reported to interact with an alternative SLiM known as the Ubiquitin interacting motif (UIM), with the sequence Ψ-ζ-X-A-Ψ-X-X-S, where Ψ, ζ, and X denote small hydrophobic, hydrophilic, and any amino acids, respectively (Marshall *et al.*, 2019 [↗](#)). Peptides bearing this motif were present in our input library (2,452 peptides) yet only 4 were found in our HC-set, indicating that if an LC3B UDS exists, it was either inaccessible in our assay format or did not support sufficiently high-affinity binding.

Notably, we did not observe enrichment for UIM-bearing peptides even when screening in the context of LC3B constructs bearing mutations at the LIR docking site (LDS*). Instead, screening against LC3B LDS* revealed that peptides bearing canonical LIR sequences exhibited altered affinity for the mutant LC3B domain (Figure 5 [↗](#)), consistent with these often-utilized mutations changing specificity but not ablating binding, contrary to what has been assumed (Behrends *et al.*, 2010 [↗](#); Kraft *et al.*, 2014 [↗](#); Qiu *et al.*, 2017 [↗](#); Skytte Rasmussen *et al.*, 2017 [↗](#); Marshall *et al.*, 2019 [↗](#)).

Together, the LC3B and LC3B LDS* bacterial display datasets presented here constitute a rich resource for the study of SLiMs and autophagy. These datasets reveal preferred residues flanking the canonical LIR motif, expand the repertoire of LC3B-binding sequences, and highlight important caveats in using the LC3B F52A L53A LDS* mutant to disrupt LDS-dependent interactions. Although peptides in our bacterial display system cannot fully capture the complexity of LC3B binding in cells—where SLiM accessibility, subcellular localization, and post-translational modifications may play key roles—our results provide a quantitative and structural framework for

understanding LC3B recognition. Moreover, this work nominates new candidate LC3B partners for validation in cellular and physiological contexts and establishes a generalizable framework for high-throughput analysis of SLiM-peptide interactions with Atg8-family proteins.

Materials and methods

Expression and bacterial display vector construction

LC3B expression vectors

Coding sequences for LC3B and the LDS* mutant (F52A, L53A) were cloned into plasmid pDW363 (Tsao *et al.*, 1996 [\[1\]](#)) downstream of an N-terminal Biotin Acceptor Peptide (BAP) and 6×His tag (GLNDIFEAQKIEWHE-DTGGSS-HHHHHH-GSGSG-[coding sequence]). The resulting plasmids (pL_JD239, pL_JD362) also expressed the BirA biotin ligase for *in vivo* biotinylation of the BAP-fused LC3B. For BLI experiments, the BAP tag was removed to yield 6×His-tagged, non-biotinylated variants pL_JD361 (LC3B wild-type) and pL_JD363 (LC3B LDS*) (M-GSS-HHHHHH-GSGSG-[coding sequence]). For crystallization of LC3B bound to the BLM⁵⁵²⁻⁵⁷¹ peptide, DNA constructs were designed to fuse the sequence DNFIDDDFDDDDDDWEDIM N-terminal to LC3B, separated by a two-residue GS linker, and cloned into a pGEX vector containing an N-terminal GST and 3C cleavage site using Gibson assembly (Gibson *et al.*, 2009 [\[2\]](#)), resulting in plasmid pL_JD371.

Peptide expression vectors

The coding sequence of each peptide was synthesized either as a gene block (IDT) or was encoded on primers (Eton Bio), each used to insert this coding sequence into a modified pDW363 vector using Gibson (Gibson *et al.*, 2009 [\[2\]](#)) or ‘round-the-horn’ insertional cloning (Liu and Naismith, 2008 [\[3\]](#)). The resulting products fused each peptide to the C-terminus of a BAP-6×His-SUMO tag via a short (SG)₂ linker.

Bacterial display vectors

The human peptidome bacterial display library is described by Hwang *et al.*, and the peptide-encoding DNA was originally provided by the Elledge laboratory based on work of Larman *et al.*, Briefly, each peptide is fused to the C-terminus of circularly permuted *E. coli* OmpX (eCPX) (Rice *et al.*, 2006 [\[4\]](#)), with the peptide flanked by an N-terminal FLAG tag and a C-terminal cMyc tag. The resulting amino-acid sequence is:

```

_SRTAS-GGGSGGGG-DYKDDDDK-GGGSGGGSGIPLR-
[ VARIABLE_PEPTIDE ]-RIARGSGS-EQKLISEEDL-*
_eCPX-[linker]-[ FLAG ]-[ linker ]-
[variable peptide]-[linker]-[ cMyc ]-[stop]

```

Individual peptides analyzed as shown in Supplementary Figure 1 were cloned using Gibson assembly (Gibson *et al.*, 2009 [\[2\]](#)) to replace the variable peptide sequence described above.

Protein expression and purification

Expression of LC3B proteins

LC3B proteins were expressed in *E. coli* Rosetta2(DE3) cells (Novagen) grown at 37 °C with aeration via shaking in 2×YT medium supplemented with ampicillin (100 µg/mL) and D-(+)-biotin (0.05 mM). When cultures reached OD₆₀₀ ≈ 0.8, protein expression was induced by addition of IPTG (1 mM), and cells were shifted to 25 °C and incubated overnight with aeration before harvesting by centrifugation (5000×g, 15 minutes, 4 °C). Cell pellets were flash-frozen and stored at -80 °C until purification. The GST-3C-BLM⁵⁵²⁻⁵⁷¹-GS-LC3B fusion protein was expressed as above using *E. coli* BL21(DE3) cells.

Expression of peptides

Peptide constructs were expressed in *E. coli* BL21(DE3) cells (New England Biolabs) cultured with aeration at 37°C in 20 mL Luria Broth (LB) supplemented with ampicillin (100 µg/mL) and D-(+)-biotin (0.05 mM). Expression was induced at an OD₆₀₀ ≈ 0.8 using IPTG (1 mM) and, after 4-5 hours of expression at 37 °C, they were harvested and frozen as above.

Purification of LC3B proteins

Cell pellets were resuspended in buffer NB (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.5 mM TCEP) supplemented with 0.2 mM phenylmethylsulfonyl fluoride (PMSF). Cells were lysed by 20 passes through a Dounce homogenizer followed by sonication on ice (4×2.5min cycles, each alternating 20 sec at 30% amplitude, 10 sec off). The lysate was clarified by centrifugation (15,000×g, 15 min, 4 °C) and ultracentrifugation in a 60Ti rotor (50,000 rpm, 1 hr). The supernatant was loaded onto a 5 mL Bio-Scale Mini Nuvia IMAC Ni-charged cartridge (BioRad 780-0811) equilibrated in buffer NB and eluted with a 0-100% linear gradient of buffer EB (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 300 mM imidazole). Fractions containing LC3B, as identified by SDS-PAGE, were pooled, concentrated (3 kDa cutoff spin concentrator; Millipore), and further purified by size-exclusion chromatography on a Superdex 75 16/600 (Cytiva) column equilibrated with buffer GFB (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 10 % glycerol). Purified protein was concentrated to ~1.3 mM, flash-frozen in liquid nitrogen, and stored at -80 °C.

Cell pellets bearing BLM⁵⁵²⁻⁵⁷¹-GS-LC3B were resuspended in buffer GLB (20 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 140 mM NaCl, 2.7 mM KCl, 0.2 mM PMSF; pH 7.3) and lysed and clarified as above. Supernatant was then loaded onto a GSTPrep FF 16/10 10 mL column (Cytiva) pre-equilibrated with buffer GB lacking PMSF, washed, and eluted with buffer GEB (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 30 mM reduced glutathione). Fractions were analyzed by SDS-PAGE, and those containing the correct molecular-weight species were pooled. Cleavage was performed overnight at 4 °C using a 1:100 (w/w) ratio of 3C protease (gift of Dr. Andrew V. Grassetti) to substrate. The cleavage product was then purified via size-exclusion chromatography as above, concentrated to ~1.4 mM, flash-frozen in liquid nitrogen, and stored at -80 °C in 50 µL aliquots.

Purification of peptides for BLI assays

Cell pellets were thawed on ice and lysed using 4 mL B-PER reagent (ThermoFisher) per gram of cells, supplemented with 0.2 mM PMSF. Lysed cells (~1 mL) were rotated at 25 °C for 15 min before pelleting by centrifugation at 15,000 rpm for 15 min. Supernatant was added to 250 µL Ni-Sepharose High Performance resin (Cytiva) pre-equilibrated with buffer NB in a gravity flow column (Bio-Rad). Resin was washed three times with 1 mL buffer NB before elution with 2 mL of buffer EB. Purified products were flash-frozen and stored at -80 °C.

FACS sample preparation

FACS samples were prepared as previously described (Hwang *et al.*, 2022 [DOI](#)) with slight modifications detailed below.

Bacterial display sample preparation

Briefly, MC1061 *E. coli* cells (Casadaban and Cohen, 1980 [DOI](#)) were transformed either with a control plasmid by chemical transformation or with the human peptidome library by electroporation. Transformed cells were cultured overnight with aeration at 37 °C in LB (5 mL) supplemented with chloramphenicol (25 µg/mL) and glucose (0.2% w/v). The following morning, ~1×10⁷ cells per sample to be analyzed (as estimated using OD₆₀₀, assuming 5×10⁸ cells/mL/OD₆₀₀) were pelleted (3,000×g, 5 min), and resuspended in Terrific Broth supplemented with chloramphenicol (25 µg/mL) to an OD₆₀₀ ≈ 0.1. Cultures were grown with aeration at 37 °C until reaching an OD₆₀₀ ≈ 0.5, at which point peptide expression was induced with 0.04 % (w/v) arabinose for 1.5 hr at 37 °C with aeration. Following induction, cells (~1×10⁷ per sample) were centrifuged (3,000×g, 5 min), washed once with phosphate-buffered saline (PBS; 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 2.7 mM KCl, 138 mM NaCl; pH 7.4) containing 0.1 bovine serum albumin (0.1% w/v; BSA), and resuspended in PBS + BSA (0.1 %) to a final cell density of 4×10⁸ cells/mL.

An anti-FLAG antibody conjugated to allophycocyanin (anti-FLAG-APC; PerkinElmer) was diluted 1:100 in PBS + BSA (0.1%), and 30 µL of solution was added per 1×10⁷ cells (25 µL/sample) for each sample to be analyzed. Sample was rotated at 4 °C for 30 min in foil-wrapped 1.5 mL tubes, pelleted (3,000×g, 5 min) and washed with 500 µL PBS + BSA (0.1 %) to remove unbound antibody, and resuspended in 25 µL PBS per sample to a final cell density of ~4×10⁸ cells/mL.

LC3B sample preparation

Biotinylated BAP-6×His-LC3B proteins were bound to streptavidin-conjugated-phycoerythrin (SA-PE, ThermoFisher Scientific) at a 4.2:1 molar ratio in PBS+BSA (1.0 %) containing 4 mM dithiothreitol (DTT). Complexes were incubated in the dark with rotation for 15 min at 4 °C and prepared at 2× the desired final concentration. For all sorting experiments, LC3B-SA-PE was prepared to yield a final concentration of 1.68 μM monomeric LC3B (0.42 μM tetramerized LC3B•SA-PE).

FACS sample preparation

To prepare final samples, 25 μL of cells (1×10^7) in PBS was mixed 1:1 with 25 μL of pre-tetramerized LC3B (2×stock concentration) for a final volume of 50 μL, and incubated with rotation in the dark at 4 °C for 1 hr. Note that for the initial library, 200 μL of cells ($\sim 1 \times 10^8$) were mixed 1:1 and treated as above. Samples were transferred to a 0.22 μm 96-well Multi-Screen HTS GV sterile filtration plate (Millipore) pre-washed with 500 μL PBS + BSA (0.5%) per well. Cell suspensions were vacuum filtered, washed with 200 μL PBS + BSA (0.1%) and resuspended in 400 μL PBS + BSA (0.1%) for control experiments shown in Supplementary Figure 1 or 4.2 mL PBS + BSA (0.1%) for library samples prior to FACS analysis or sorting.

Bacterial display sorting

Bacterial display libraries were analyzed and sorted on the BD FACSCanto II (analysis) and BD FACSARIA 4 (sorting) instruments. To establish gating parameters for enrichment of LC3B-binding peptides, two positive controls (FYCO1¹²⁶⁴⁻¹²⁹⁹ and ATG4B³⁷²⁻⁴⁰⁷) and a negative control (lacking the variable peptide segment) were analyzed. The selection gate was set to maximize recovery of positive cells while maintaining ≤ 0.4 % background from the negative population. Following fluorescence compensation to correct for spectral overlap between channels, the gate was held constant throughout all positive enrichment sorts.

In the initial sort, $\sim 8.6 \times 10^7$ cells were sorted – sufficient to oversample the input library by ~ 200 -fold. Sorted cells were collected into Super Optimal broth with Catabolite repression (SOC) and incubated for 1 hr at 37 °C with rotation before being transferred to 8 mL LB medium containing chloramphenicol (25 μg/mL) and glucose (0.2 % w/v). Cultures were grown at 37 °C with aeration until OD₆₀₀ $\approx$ 0.4–0.8. Plasmid DNA was purified from these cells using a Monarch Miniprep Kit (New England Biolabs).

For each subsequent round of sorting, *E. coli* MC1061 cells were transformed via electroporation with 50 ng of the purified plasmid library DNA and analyzed as above, with sufficient cells collected to oversample the binding pool by ~ 100 -fold. For the second round (sort 2), which served as a counter-selection against nonspecific peptide binding to the assay components, D-(+)-biotin was incubated with SA-PE at a 4.2:1 molar ratio as described above, and a gate was selected that minimized the capture of positive-control peptides (Supplementary Figure 1).

LC3B LDS* samples were prepared in a similar manner as wild-type LC3B. To enable direct comparison, the binding pool from sort 5 was split and sorted in parallel on the same day against wild-type LC3B (sort 6) and LC3B LDS* (LDS* 1). Both binding and nonbinding populations from each experiment were collected for sequencing. Two additional LC3B LDS* enrichment sorting rounds were then performed to further enrich for LDS*-binding peptides, and both the binding and nonbinding populations were collected and miniprep to isolate the plasmid DNA.

Next-generation sequencing (NGS) sample preparation

Miniprep plasmid libraries from each round of sorting were assigned a unique 6-nucleotide index. The variable region of each library was PCR-amplified using a forward primer (ol_JD806) bearing the 5' Illumina adaptor sequence (AATGATACGGCGACCACCGAGATCTACAC) and a sequence (GTGGCTCGGGAATCCGCTGCGC) designed to anneal immediately 5' of the variable peptide region of our bacterial display vector, and a reverse primer (ol_JD807-813) designed to anneal

immediately 3' of the variable peptide region that also contained one of the fourteen 6-nucleotide index sequences and a 3' Illumina adaptor sequence. The amplicon preparation scheme and all primers are listed in Supplementary File 1.

Eleven cycles of amplification were performed using Phusion polymerase (NEB) with 1.25 μ M primers under the following thermocycling conditions: initial denaturation at 98 °C for 45 s; denaturation at 98 °C for 15 s; annealing at 68 °C for 30 s; and extension at 72 °C for 60 s, followed by a final extension at 72 °C for 5 min. To minimize heteroduplex formation, two additional cycles of reconditioning PCR were performed with 1 μ M primers using the same thermocycling conditions. PCR products were purified using a double-sided AMPure XP (Beckman Coulter) bead size selection (0.85X/0.5X) to exclude large amplicons and small adapter or primer dimers, retaining fragments within the desired library size range for Illumina sequencing. Samples were eluted in 11 μ L of buffer NGS (10 mM Tris-HCl, pH 8.0).

DNA amplicon library purity, size, and quantity were assessed using a Bioanalyzer (Agilent Technologies). Indexed samples were pooled to generate a multiplexed library bearing samples from all wild-type LC3B and LC3B LDS* sorts, which was submitted to the MIT BioMicro Center for sequencing on a NextSeq500 (Illumina) platform using 150 bp paired-end reads.

NGS data processing and analysis

Demultiplexed sequencing data were merged with BBMerge (Bushnell *et al.*, 2017) using an average quality score threshold of 20. Merged reads were used to quantify the number of clonal cells (*i.e.*, cells displaying the same peptide) recovered across consecutive sorts. The frequency of each clone i collected in sort x , denoted as $f_{i,x}$, was calculated as $(c_{i,x}/N_x)$ where $c_{i,x}$ is the number of sequencing reads mapped to sequence i in sort x and N_x is the total number of sequencing reads mapped in sort x , *i.e.*, $N_x = \sum_i c_{i,x}$. The enrichment ratio (ER) for each clone across the sorting trajectory was then calculated as $ER_{i,x} = \log_2(f_{i,x}/f_{i,input})$ where $f_{i,input}$ represents the frequency of the same sequence in the naïve input library (Rubin *et al.*, 2017).

The T7-pep library contains many apparent point mutations and frameshifts. To group related variants, sequences were clustered using ALFAT-Clust (Chiu and Ong, 2022) with default parameters. Within each cluster, the variant that persisted to the latest sort was selected as the representative sequence; when multiple variants persisted, the sequence with the highest read count in that sort was chosen. Clones with fewer than 10 reads in the input library were excluded from all analyses. Sequences were translated using Biopython (Cock *et al.*, 2009) and mapped to the human proteome using NCBI BLASTP (Camacho *et al.*, 2009).

All read count data for each sequence in each sort are available at the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA1276872. Supplementary File 2 contains the pre-collapsed data, and Supplementary File 3 contains the post-collapsed data used in this study, including amino acid sequences, raw counts, ER values, BLAST results, and z-scores. Supplementary File 4 is filtered for peptides that enriched through sort 6 and includes AlphaFold2 (Jumper *et al.*, 2021) pLDDT scores estimating structural disorder within each peptide region.

LC3B and LC3B LDS* scoring metrics

Following NGS data processing described above, 487,021 unique peptide sequences were identified in the input library, with 12,158 remaining in the terminal LC3B sort 6. Among these, clones were placed in the HC-set if the mean z-score across sorts 4-6 was ≥ 1.70 . The z-score for clone i in sort x was defined as: $z_{i,x} = [(ER_{i,x} - \mu_x) / \sigma_x]$, where $ER_{i,x}$ is the enrichment ratio of peptide i in sort x and μ_x and σ_x are the mean and standard deviation ER values for all peptides in that sort. The threshold of 1.70 was chosen based on BLI affinity measurements to minimize the false positives (Figure 2C). Z-scores for all clones that reached sort 6 are provided (Supplemental File 4).

To evaluate peptide behavior in the LC3B LDS* sorts, three complementary metrics were applied. First, enrichment trajectories were examined to identify clones that had enriched through the canonical LC3B sort 5, and that continued to enrich through the three LDS* sorts despite disruption of the canonical LIR-docking site. Second, the enrichment ratio was calculated using

sort 5 as the input library to compare peptide enrichment between the first LDS* sort ($ER_{LDS^*,1} = \text{freq}(LDS^*_{i,1}) / \text{freq}(\text{sort } 5)$) and the canonical sort 6 ($ER_{i,6} = \text{freq}(\text{sort } 6) / \text{freq}(\text{sort } 5)$). Promising clones were those that were enriched in the LDS* condition and depleted in the canonical condition ($ER_{LDS^*,1} / ER_{i,6} > 1$), consistent with binding in the absence of the wild-type LDS pocket. Finally, the ratio of clone frequencies in the binding versus non-binding gates (B/NB) was calculated for each LDS* sort and for the canonical LC3B sort 6. Clones showing progressive increase in the B/NB ratio across the LDS* sorts were considered selectively enriched for binding to LC3B LDS*. Top-scoring clones across these metrics were manually selected and tested for binding to LC3B and LC3B LDS* by BLI, as detailed below. Peptides that bound LC3B LDS* scored highly across all three metrics. All LDS* enrichment data and associated scoring metrics are provided (Supplemental File 5).

Identification of reported interaction partners of LC3B

To assemble a set of high-confidence LC3B-binding peptides and their corresponding human proteins, the LIRcentral database (Chatzichristofi *et al.*, 2023) was used to extract entries ($n=115$) experimentally verified to bind one of the hAtg8 proteins. Of these, 27 were reported to bind LC3B, with 10 such peptides reported to bind with $K_D \leq 50\mu\text{M}$ being present in our input library. The six peptides from this subset that appeared in our HC-set are indicated in the “LIR row” of Figure 2A.

In parallel, human proteins reported to co-immunoprecipitate with LC3B ($n=550$) were retrieved from the BioGrid database (Stark *et al.*, 2006), and peptides related to those proteins that were present in our HC-set are annotated in the “IP row” in Figure 2A. To assess whether candidates in the HC-set may have biologically relevant interactions with LC3B, Gene Ontology (GO) terms for proteins corresponding to each peptide present in sort 6 were extracted from UniProt (The UniProt Consortium, 2025) and cross-referenced with GO terms associated with LC3B. Those proteins sharing a GO term with LC3B are noted in the GO row of Figure 2A, with the GO terms shown in Supplementary Figure 4.

Biolayer interferometry (BLI)

BLI experiments were performed on an Octet Red96 instrument (ForteBio). Purified, biotinylated SUMO-fused peptides were immobilized on Octet SA Biosensors (Sartorius) and loaded until a response level of at least 0.3nm was reached. The ligand-loaded tips were then incubated with increasing concentrations of target protein (*e.g.*, LC3B) in buffer BLI (25 mM Tris, pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 0.05 % Tween-20, 0.5 mg/mL BSA) in a 200 μL reaction conducted in a 96-well flat-bottom polypropylene microplates (Greiner, #655209). Each measurement was repeated 2 – 4 times at 25 °C at an agitation speed of 1000 RPM. Equilibrium dissociation constants (K_D) for the measured interactions were determined by plotting the equilibrated signal, after subtracting the negative control (biotinylated SUMO-only), as a function of protein concentration. The data were fit by nonlinear regression to a one-site binding model, according to the equation:

$$Y = B_{\text{max}}[\text{protein}] / (K_D + [\text{protein}]).$$

Structure determination by X-ray crystallography

Crystals of the purified BLM⁵⁵²⁻⁵⁷¹-GS-LC3B fusion protein were grown in hanging drops containing buffer BXC (0.1 M HEPES, pH 7.3, PEG-3350 (30 % w/v), and 0.32 M MgCl_2 , by mixing 1 μL of protein with an equal volume of reservoir solution; crystals appeared within 7 days. X-ray diffraction data were collected on a Rigaku Micromax-007 rotating anode equipped with Osmic VariMax-HF mirrors and a Rigaku Saturn 944 detector. Diffraction data were processed with the XDS suite (Kabsch, 2010).

The structure was solved by molecular replacement in Phaser (McCoy *et al.*, 2007) using LC3B (pdb 3vtu) (Rogov *et al.*, 2013) as the search model. The molecular replacement solution was refined in PHENIX (Liebschner *et al.*, 2019) with manual fitting in Coot (Emsley *et al.*, 2010), and further geometry optimization in Rosetta (DiMaio *et al.*, 2013). The model was refined to

$R_{\text{work}}/R_{\text{free}}=0.216/0.262$. The X-ray data collection and refinement statistics are summarized in Supplementary Table 4, and the coordinates have been deposited in the Protein Data Bank (pdb 9pe3). Ensemble refinement as implemented in PHENIX (Burnley *et al.*, 2012) was carried out using pTLS values (fraction of atoms included in TLS fitting) ranging from 1.0 to 0.5, with 0.6 yielding the lowest R_{free} value.

Sequence logo-based analysis

We used pLogo (O'Shea *et al.*, 2013) to visualize the enrichment of residues in peptide binders, compared to the input library background, and designed a series of consensus-sequence peptides based on this analysis. pLogo estimates the statistical significance of residue enrichments using foreground and background distributions. To define the background, we identified input-library sequences matching the LIR motif and included 15 residues on either side of the motif. For peptides that lacked 15 flanking residues around the LIR, we appended the N- or C-terminal flanking sequence that was present in the experimental display construct. To define a set of strongly enriching binders for the foreground, peptides that reached sort 6 were clustered ($k=10$), based on their enrichment profiles, using Clust (Abu-Jamous and Kelly, 2018). The sequences in the highest-enriching cluster were used as the foreground, using the LIR plus 15 flanking residues on either side, as above. The foreground and background sequences were aligned on the LIR and used as inputs to pLogo. We repeated this analysis four times, for the canonical $[\text{FWY}]_0\text{-X}_1\text{-X}_2\text{-}[\text{LVI}]_3$ LIR motif and for LIR motifs with position X_0 restricted to F, W, or Y, generating the images in Supplementary Figure 5. For consensus peptide design, the residue with the highest log-odds of the binomial probability at each position of the corresponding 34-residue logo was selected, generating peptides $\text{pCONS}_{\text{LIR}}$, pCONS_{F} , pCONS_{W} , and pCONS_{Y} , which are listed in Supplementary Table 3.

LIR sequence sets for PSSM and random-forest model building

To build and assess models for classifying LC3B-binding and non-binding peptides, five different sequence sets were defined, each matching the 7-residue motif $\text{X}_3\text{-X}_2\text{-X}_1\text{[FWY]}_0\text{X}_1\text{X}_2\text{[FWYILV]}_3$: *screen_binder* and *screen_nonbinder* from our screening data, *test_binder* and *test_nonbinder* sequences from LIRcentral (test sets) (Chatzichristofi *et al.*, 2023), and *iLIR_binder* sequences from iLIR (Kalvari *et al.*, 2014). To define the *iLIR_binder* set, LIR-containing sequences were downloaded from the iLIR webserver. Because the iLIR motifs are only 6 residues long, we extended the 6-mers to 7-mers using full-length sequences from the UniProt reference proteome (The UniProt Consortium, 2025). To define the *test_binder* set, we extracted 76 motif-matching 7-mer sequences annotated to bind LC3B in LIRcentral and that were not present in the *iLIR_binder* set, then added three peptides from our screen that we measured to bind LC3B with $K_D \leq 14 \mu\text{M}$ using BLI (Supplementary Table 1). To define the *test_nonbinder* set, we used 49 motif-matching 7-mers reported by LIRcentral to not bind LC3B (also absent from the *iLIR_binder* set), then added seven peptides from our screen with no detectable binding to LC3B by BLI (Supplementary Table 2). In total, test set contained 79 binder peptides and 56 non-binder peptides.

To define the *screen_binder* and *screen_nonbinder* sets, we selected input-library sequences bearing exactly one LIR motif (as above), appended “PLR” to the N-terminus, which was present in the experimental display construct, extracted 7-mers centered on the LIR motif, and removed any sequences also present in the *iLIR_binder*, *test_binder*, or *test_nonbinder* sets. Remaining 7-mers in the HC-set formed the *screen_binders* set; those that were depleted in sorts 1 and 3 and then dropped out of the screen during any subsequent sort formed the *screen_nonbinders* set. Note that any duplicate 7-mers found within or between the *screen_binder* and *screen_nonbinder* sets were removed from both sets. The final sequence set from the screening data consisted of 151 *screen_binder* sequences and 2,844 *screen_nonbinder* sequences. For modelling, we additionally used ranges of z-scores to define *screen_binder* subsets: *screen_binder_all* (z-score ≥ 1.7 , 151 sequences), *screen_binder_low* ($1.7 \leq \text{z-score} < 2.3$, 101 sequences), and *screen_binder_high* (z-score ≥ 2.3 , 50 sequences). Logos for the screening sets are shown in Supplementary Figure 6C.

For the model building and evaluation, screening data were repeatedly (100×) split into train/test partitions. Each iteration selected 7-mers at random without replacement to form test sets, with 15 from the *screen_binder_high* test subset, and 30 from the *screen_binder_low* test subset; we designated the combined 45 sequences as the “all” *screen_binder* test set. An equal number of nonbinder sequences from the *screen_nonbinder* set was selected to complete the test sets. For models built from the screening data, the sequences that were not selected for the test sets were used to train models or build PSSMs. Thus, for each train-test iteration, three completely non-overlapping test and training sets were defined, which were used to train and evaluate models.

PSSM classifier model construction

PSSM classifier models were generated using a 7-residue spanning the three N-terminal flanking residues and the core LIR motif (*i.e.*, $X_3X_2X_1[FWY]_0X_1X_2[FWYILV]_3$). Each model used one of the binder sets (defined above), as the foreground and the amino-acid frequencies from the *Homo sapiens* reference proteome (UP000005640) as the background distribution. Position-specific scores were calculated as log-odds values:

$$w_{i,c} = \begin{cases} \log_2 \frac{p_{i,c}}{b_c}, & \text{if } p_{i,c} \neq 0 \\ 0, & \text{if } p_{i,c} = 0 \end{cases}$$

where $w_{i,c}$ is the position-specific score for residue c at position i , $p_{i,c}$ is the frequency of residue c at position i in the set of foreground sequences, and b_c is the background frequency of residue c in the proteome. The total PSSM score for a sequence was the sum of position-specific values across the 7-residue window. An analogous model (iLIR₂₇) was generated using the 27 binders that formed the basis of the published iLIR model (Kalvari *et al.*, 2014) as the foreground sequences.

Random-forest classifier model construction

Random-forest and balanced random-forest classifiers were trained using *RandomForestClassifier* from scikit-learn (Pedregosa *et al.*, 2011) and *BalancedRandomForestClassifier* from imbalance-learn (Lemaître *et al.*, 2017), where we noted that unlike the PSSM model from above, these models have the potential to capture higher-order effects, including coupling between sites. The training set consisted of the 2,844 nonbinder and 151 binder sequences defined above. We leveraged the PSSM matrix detailed above to encode each sequence. A PSSM-encoded sequence (X) is generated by replacing the one-letter amino acid code in a sequence with its corresponding PSSM score in the matrix. Mathematically, the i^{th} encoded position of a sequence (X_i) is:

$$X_i = M_{ij}^{\text{PSSM}}$$

where i denotes the i^{th} position in an encoded peptide sequence, j represents the column in the PSSM matrix that corresponds to the amino acid appearing at i^{th} position in a sequence. To handle data imbalances between screening binders and nonbinders, a downsampling technique was used to randomly select nonbinders to match the total number of binders. Cross-validation (5-fold) was applied to select the best hyperparameters from the parameter space for random forest (*n_estimators*: 4-200, *criteria*: [gini, entropy], *ccp_alpha*: 0.004-0.014) and balanced random forest models (*n_estimators*: 4-200, *criteria*: [gini, entropy], *ccp_alpha*: 0.004-0.014, *sampling_strategy*: all, *replacement*: True, *bootstrap*: False). The best model was selected based on the mean ROC-AUC and retrained with all training data.

Acknowledgements

We thank A. Ghanbarpour for help with crystallographic structure refinement. We thank L. Kinman, D. Cui, and B. Powell for helpful discussions. Research reported in this publication was supported by the National Institutes of Health under Award Numbers R35GM149227 (AK), R01GM144542 (JHD), R00AG050749 (JHD) and the Smith Family Odyssey Award (JHD).

Additional information

Data availability

Datasets generated during this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA). Raw reads can be found under BioProject ID PRJNA1276872. The structure of LC3B bound to BLM552-571 is deposited in the Protein Data Bank with accession number 9p3e. Supplementary Files 1-5 are available at Zenodo (10.5281/zenodo.17943925). Python processing scripts are available upon request.

Author contributions

J.E.K., J.H.D., and A.E.K. designed research; J.E.K. performed screening, data analysis, and biochemical experiments, and contributed new reagents and analytic tools; C.L., J.E.K., and J.C.H. developed and tested models; D.L. collected and refined crystallography data; J.E.K., J.H.D., and A.E.K. wrote the paper.

Funding

Funder	Grant reference number	Author
HHS National Institutes of Health (NIH)	R35GM149227	Amy E Keating
HHS National Institutes of Health (NIH)	R01GM144542	Joseph H Davis
HHS National Institutes of Health (NIH)	R00AG050749	Joseph H Davis
Smith Family	Odyssey Award	Joseph H Davis

Author ORCID iDs

Jennifer Kosmatka: <https://orcid.org/0000-0001-5337-3298>

Cong Liu: <https://orcid.org/0000-0002-1687-2220>

Jackson C Halpin: <https://orcid.org/0000-0003-4085-1064>

Daniel Lim: <https://orcid.org/0000-0002-8861-4837>

Joseph H Davis: <https://orcid.org/0000-0002-8858-8907>

Amy E Keating: <https://orcid.org/0000-0003-4074-8980>

Additional files

[Supplementary Information](#) 

References

- Abramson J**, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, *et al.* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**:493-500
- Abu-Jamous B**, Kelly S. (2018) Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome Biol* **19**:172
- Behrends C**, Sowa ME, Gygi SP, Harper JW (2010) Network organization of the human autophagy system. *Nature* **466**:68-76
- Birgisdottir AB**, Lamark T, Johansen T. (2013) The LIR motif - crucial for selective autophagy. *J Cell Sci* **126**:3237-3247
- Boecker CA**, Holzbaur ELF (2021) Hyperactive LRRK2 kinase impairs the trafficking of axonal autophagosomes. *Autophagy* **17**:2043-2045

- Brennan A**, Layfield R, Long J, Williams HEL, Oldham NJ, Scott D, Searle MS (2022) An ALS-associated variant of the autophagy receptor SQSTM1/p62 reprograms binding selectivity toward the autophagy-related hATG8 proteins. *J Biol Chem* **298**:101514
- Bugge K**, Brakti I, Fernandes CB, Dreier JE, Lundsgaard JE, Olsen JG, Skriver K, Kragelund BB (2020) Interactions by Disorder – A Matter of Context. *Front Mol Biosci* **7**:110
- Burnley BT**, Afonine PV, Adams PD, Gros P. (2012) Modelling dynamics in protein crystal structures by ensemble refinement ed. A.T. Brunger. *eLife* **1**:e00311 <https://doi.org/10.7554/eLife.00311>
- Bushnell B**, Rood J, Singer E. (2017) BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* **12**:e0185056
- Camacho C**, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421
- Casadaban MJ**, Cohen SN (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *Journal of Molecular Biology* **138**:179-207
- Chatzichristofi A**, Sagris V, Pallaris A, Eftychiou M, Kalvari I, Price N, Theodosiou T, Iliopoulos I, Nezis IP, Promponas VJ (2023) LIRcentral: a manually curated online database of experimentally validated functional LIR motifs. *Autophagy* **19**:3189-3200
- Cheng X**, Wang Y, Gong Y, Li F, Guo Y, Hu S, Liu J, Pan L. (2016) Structural basis of FYCO1 and MAP1LC3A interaction reveals a novel binding mode for Atg8-family proteins. *Autophagy* **12**:1330-1339
- Chiu JKH**, Ong RT-H. (2022) Clustering biological sequences with dynamic sequence similarity threshold. *BMC Bioinformatics* **23**:108
- Cock PJA**, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**:1422-1423
- Davey NE**, Simonetti L, Ivarsson Y. (2023) The next wave of interactomics: Mapping the SLiM-based interactions of the intrinsically disordered proteome. *Current Opinion in Structural Biology* **80**:102593
- Davey NE**, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ (2012) Attributes of short linear motifs. *Mol Biosyst* **8**:268-281
- Devlin JJ**, Panganiban LC, Devlin PE (1990) Random Peptide Libraries: a Source of Specific Protein Binding Molecules. *Science* **249**:404-406
- Diella F**. (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci* **6580**
- DiMaio F**, Echols N, Headd JJ, Terwilliger TC, Adams PD, Baker D. (2013) Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat Methods* **10**:1102-1104
- Dyson HJ**, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**:197-208
- Emsley P**, Lohkamp B, Scott WG, Cowtan K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**:486-501
- Fas BA**, Maiani E, Sora V, Kumar M, Mashkoo M, Lambrugh M, Tiberti M, Papaleo E. (2021) The conformational and mutational landscape of the ubiquitin-like marker for autophagosome formation in cancer. *Autophagy* **17**:2818-2841
- Florey O**, Overholtzer M. (2012) Autophagy proteins in macroendocytic engulfment. *Trends Cell Biol* **22**:374-380
- Fumagalli F**, Noack J, Bergmann TJ, Cebollero E, Pisoni GB, Fasana E, Fregno I, Galli C, Loi M, Soldà T, et al. (2016) Translocon component Sec62 acts in endoplasmic reticulum turnover during stress recovery. *Nat Cell Biol* **18**:1173-1184

- Genau HM**, Huber J, Baschieri F, Akutsu M, Dötsch V, Farhan H, Rogov V, Behrends C. (2015) CUL3-KBTBD6/KBTBD7 Ubiquitin Ligase Cooperates with GABARAP Proteins to Spatially Restrict TIAM1-RAC1 Signaling. *Molecular Cell* **57**:995-1010
- Gibson DG**, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 2009 6:5 6:343-345
- Grant CE**, Bailey TL (2021) XSTREME: Comprehensive motif analysis of biological sequence datasets. *bioRxiv*
- Gubas A**, Dikic I. (2022) A guide to the regulation of selective autophagy receptors. *Febs J* **289**:75-89
- Heckmann BL**, Teubner BJW, Tummers B, Boada-Romero E, Harris L, Yang M, Guy CS, Zakharenko SS, Green DR (2019) LC3-Associated Endocytosis Facilitates β -Amyloid Clearance and Mitigates Neurodegeneration in Murine Alzheimer's Disease. *Cell* **178**:536-551.e14
- Hwang T**, Parker SS, Hill SM, Grant RA, Ilunga MW, Sivaraman V, Mouneimne G, Keating AE (2022) Native proline-rich motifs exploit sequence context to target actin-remodeling Ena/VASP protein ENAH. *eLife* **11**:e70680 <https://doi.org/10.7554/eLife.70680>
- Johansen T**, Lamark T. (2020) Selective Autophagy: ATG8 Family Proteins, LIR Motifs and Cargo Receptors. *J Mol Biol* **432**:80-103
- Jumper J**, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583-589
- Kabeya Y**, Mizushima N, Yamamoto A, Oshitani-Okamoto S, Ohsumi Y, Yoshimori T. (2004) LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *J Cell Sci* **117**:2805-2812
- Kabsch W**. (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**:125-132
- Kalvari I**, Tsompanis S, Mulakkal NC, Osgood R, Johansen T, Nezis IP, Promponas VJ (2014) iLIR: A web resource for prediction of Atg8-family interacting proteins. *Autophagy* **10**:913-925
- Katz BA** (1995) Binding to protein targets of peptidic leads discovered by phage display: crystal structures of streptavidin-bound linear and cyclic peptide ligands containing the HPQ sequence. *Biochemistry* **34**:15421-15429
- Kliche J**, Garvanska DH, Simonetti L, Badgujar D, Dobritzsch D, Nilsson J, Davey NE, Ivarsson Y. (2023) Large-scale phosphomimetic screening identifies phospho-modulated motif-based protein interactions. *Mol Syst Biol* **19**:e11164
- Kliche J**, Ivarsson Y. (2022) Orchestrating serine/threonine phosphorylation and elucidating downstream effects by short linear motifs. *Biochem J* **479**:1-22
- Knævelsrud H**, Carlsson SR, Simonsen A. (2013) SNX18 tubulates recycling endosomes for autophagosome biogenesis. *Autophagy* **9**:1639-1641
- Kraft LJ**, Nguyen TA, Vogel SS, Kenworthy AK (2014) Size, stoichiometry, and organization of soluble LC3-associated complexes. *Autophagy* **10**:861-877
- Kumar M**, Michael S, Alvarado-Valverde J, Mészáros B, Sámano-Sánchez H, Zeke A, Dobson L, Lazar T, rd M, Nagpal A, *et al.* (2022) The Eukaryotic Linear Motif resource: 2022 release. *Nucleic Acids Research* **50**:D497-D508
- Kumar M**, Michael S, Alvarado-Valverde J, Zeke A, Lazar T, Glavina J, Nagy-Kanta E, Donagh JM, Kalman ZE, Pascarelli S, *et al.* (2024) ELM—the Eukaryotic Linear Motif resource—2024 update. *Nucleic Acids Research* **52**:D442-D455
- Larman HB**, Zhao Z, Laserson U, Li MZ, Ciccio A, Gakidis MAM, Church GM, Kesari S, LeProust EM, Solimini NL, *et al.* (2011) Autoantigen discovery with a synthetic human peptidome. *Nat Biotechnol* **29**:535-541
- Lee A**, Davis JH (2024) NCOA4 initiates ferritinophagy by binding GATE16 using two highly avid short linear interaction motifs. *bioRxiv*

- Lemaître G, Nogueira F, Aridas CK (2017) Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res* **18**:559-563
- Lex A, Gehlenborg N, Strobel H, Vuilleumot R, Pfister H. (2014) UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph* **20**:1983-1992
- Li J, Zhu R, Chen K, Zheng H, Zhao H, Yuan C, Zhang H, Wang C, Zhang M. (2018) Potent and specific Atg8-targeting autophagy inhibitory peptides from giant ankyrins. *Nat Chem Biol* **14**:778-787
- Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung L-W, Jain S, McCoy AJ, et al. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**:861-877
- Liu H, Naismith JH (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology* **8**:91
- Lv M, Wang C, Li F, Peng J, Wen B, Gong Q, Shi Y, Tang Y. (2017) Structural insights into the recognition of phosphorylated FUNDC1 by LC3B in mitophagy. *Protein & Cell* **8**:25-38
- Madureira M, Connor-Robson N, Wade-Martins R. (2020) LRRK2: Autophagy and Lysosomal Activity. *Front Neurosci* **14**
- Marshall RS, Hua Z, Mali S, McLoughlin F, Vierstra RD (2019) ATG8-Binding UIM Proteins Define a New Class of Autophagy Adaptors and Receptors. *Cell* **177**:766-781.e24
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**:658-674
- McInnes L., Healy J., Melville J. (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*
- Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE (2023) UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* **32**:e4792
- Mizushima N, Levine B, Cuervo AM, Klionsky DJ (2008) Autophagy fights disease through cellular self-digestion. *Nature* **451**:1069-1075
- Montes-Fernández MA, Pérez-Villegas EM, Garcia-Gonzalo FR, Pedrazza L, Rosa JL, de Toledo GA, Armengol JA (2020) The HERC1 ubiquitin ligase regulates presynaptic membrane dynamics of central synapses. *Sci Rep* **10**:12057
- Olsvik HL, Lamark T, Takagi K, Larsen KB, Evjen G, vervatn A, Mizushima T, Johansen T. (2015) FYCO1 Contains a C-terminally Extended, LC3A/B-preferring LC3-interacting Region (LIR) Motif Required for Efficient Maturation of Autophagosomes during Basal Autophagy. *J Biol Chem* **290**:29361-29374
- O'Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, Schwartz D. (2013) p ogo: a probabilistic approach to visualizing sequence motifs. *Nat Methods* **10**:1211-1212
- Pankiv S, Clausen TH, Lamark T, Brech A, Bruun J-A, Outzen H, vervatn A, Bjørkøy G, Johansen T. (2007) p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *J Biol Chem* **282**:24131-24145
- Park H, Kang J-H, Lee S. (2020) Autophagy in Neurodegenerative Diseases: A Hunter for Aggregates. *Ijms* **21**:3369
- Park S, Han S, Choi I, Kim B, Park SP, Joe E-H, Suh YH (2016) Interplay between Leucine-Rich Repeat Kinase 2 (LRRK2) and p62/SQSTM-1 in Selective Autophagy. *PLoS One* **11**:e0163029
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**:2825-2830
- Pérez-Villegas EM, Pérez-Rodríguez M, Negrete-Díaz JV, Ruiz R, Rosa JL, de Toledo GA, Rodríguez-Moreno A, Armengol JA (2020) HERC1 Ubiquitin Ligase Is Required for Hippocampal Learning and Memory. *Front Neuroanat* **14**:592797
- Popelka H. (2020) Dancing while self-eating: Protein intrinsic disorder in autophagy. *Prog Mol Biol Transl Sci* **174**:263-305

- Qiu Y, Zheng Y, Wu K-P, Schulman BA (2017) Insights into links between autophagy and the ubiquitin system from the structure of LC3B bound to the LIR motif from the E3 ligase NEDD4. *Protein Sci* **26**:1674-1680
- Ramesh Babu J, Lamar Seibenhener M, Peng J, Strom A, Kemppainen R, Cox N, Zhu H, Wooten MC, Diaz-Meco MT, Moscat J, *et al.* (2008) Genetic inactivation of p62 leads to accumulation of hyperphosphorylated tau and neurodegeneration. *Journal of Neurochemistry* **106**:107-120
- Rice JJ, Schohn A, Bessette PH, Boulware KT, Daugherty PS (2006) Bacterial display using circularly permuted outer membrane protein OmpX yields high affinity peptide ligands. *Protein Sci* **15**:825-836
- Richter B, Sliter DA, Herhaus L, Stolz A, Wang C, Beli P, Zaffagnini G, Wild P, Martens S, Wagner SA, *et al.* (2016) Phosphorylation of OPTN by TBK1 enhances its binding to Ub chains and promotes selective autophagy of damaged mitochondria. *Proc Natl Acad Sci U S A* **113**:4039-4044
- Rogov VV, Nezis IP, Tsapras P, Zhang H, Dagdas Y, Noda NN, Nakatogawa H, Wirth M, Mouilleron S, McEwan DG, *et al.* (2023) Atg8 family proteins, LIR/AIM motifs and other interaction modes. *Autophagy Rep* **2**:2188523
- Rogov VV, Suzuki H, Fiskin E, Wild P, Kniss A, Rozenknop A, Kato R, Kawasaki M, McEwan DG, Löhr F, *et al.* (2013) Structural basis for phosphorylation-triggered autophagic clearance of Salmonella. *Biochemical Journal* **454**:459-466
- Roosen DA, Cookson MR (2016) LRRK2 at the interface of autophagosomes, endosomes and lysosomes. *Mol Neurodegener* **11**:73
- Rozenknop A, Rogov VV, NYu Rogova, Löhr F, Güntert P, Dikic I, Dötsch V. (2011) Characterization of the Interaction of GABARAPL-1 with the LIR Motif of NBR1. *Journal of Molecular Biology* **410**:477-487
- Rubin AF, Gelman H, Lucas N, Bajjalieh SM, Papenfuss AT, Speed TP, Fowler DM (2017) A statistical framework for analyzing deep mutational scanning data. *Genome Biol* **18**:150
- Sawa-Makarska J, Abert C, Romanov J, Zens B, Ibric I, Martens S. (2014) Cargo binding to Atg19 unmasks additional Atg8 binding sites to mediate membrane-cargo apposition during selective autophagy. *Nat Cell Biol* **16**:425-433
- Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**:6097-6100
- Skytte Rasmussen M, Mouilleron S, Kumar Shrestha B, Wirth M, Lee R, Bowitz Larsen K, Abudu Princely Y, O'Reilly N, Sjøttem E, Tooze SA, *et al.* (2017) ATG4B contains a C-terminal LIR motif important for binding and efficient cleavage of mammalian orthologs of yeast Atg8. *Autophagy* **13**:834-853
- Sønder SL, Häger SC, Heitmann ASB, Frankel LB, Dias C, Simonsen AC, Nylandsted J. (2021) Restructuring of the plasma membrane upon damage by LC3-associated macropinocytosis. *Sci Adv* **7**:eabg1969
- Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. (2006) BioGRID: a gen al repository for interaction datasets. *Nucleic Acids Res* **34**:D535-539
- Stadel D, Millarte V, Tillmann KD, Huber J, B-C Tamin-Yecheskel, Akutsu M, Demishtein A, Ben-Zeev B, Anikster Y, Perez F, *et al.* (2015) TECPR2 Cooperates with LC3C to Regulate COPII-Dependent ER Export. *Molecular Cell* **60**:89-104
- Sugawara K, Suzuki NN, Fujioka Y, Mizushima N, Ohsumi Y, Inagaki F. (2004) The crystal structure of microtubule-associated protein light chain 3, a mammalian homologue of *Saccharomyces cerevisiae* Atg8. *Genes to Cells* **9**:611-618
- The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, *et al.* (2023) The Gene Ontology knowledgebase in 2023. *Genetics* **224**:iyad031
- The UniProt Consortium, Bateman A, Martin M-J, Orchard S, Magrane M, Adesina A, Ahmad S, Bowler-Barnett EH, Bye-A-Jee H, Carpentier D, *et al.* (2025) UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Research* **53**:D609-D617

- Tompa P**, Davey NE, Gibson TJ, Babu MM (2014) A Million Peptide Motifs for the Molecular Biologist. *Molecular Cell* **55**:161-169
- Tong Y**, Zhu W, Chen J, Zhang W, Xu F, Pang J. (2023) Targeted Degradation of Alpha-Synuclein by Autophagosome-Anchoring Chimera Peptides. *J Med Chem* **66**:12614-12628
- Tsao KL**, DeBarbieri B, Michel H, Waugh DS (1996) A versatile plasmid expression vector for the production of biotinylated proteins by site-specific, enzymatic modification in *Escherichia coli*. *Gene* **169**:59-64
- Van Roey K**, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, Gibson TJ, Davey NE (2014) Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem Rev* **114**:6733-6778
- Weber PC**, Pantoliano MW, Thompson LD (1992) Crystal structure and ligand-binding studies of a screened peptide complexed with streptavidin. *Biochemistry* **31**:9350-9354
- Wirth M**, Mouilleron S, Zhang W, Sjøttem E, Princely Abudu Y, Jain A, Lauritz Olsvik H, Bruun J-A, Razi M, Jefferies HBJ, *et al.* (2021) Phosphorylation of the LIR Domain of SCOC Modulates ATG8 Binding Affinity and Specificity. *J Mol Biol* **433**:166987
- Wright PE**, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* **16**:18-29
- Wurzer B**, Zaffagnini G, Fracchiolla D, Turco E, Abert C, Romanov J, Martens S. (2015) Oligomerization of p62 allows for selection of ubiquitinated cargo and isolation membrane during selective autophagy. *eLife* **4**:e08941 <https://doi.org/10.7554/eLife.08941>
- Zaffagnini G**, Martens S. (2016) Mechanisms of Selective Autophagy. *J Mol Biol* **428**:1714-1724

Peer reviews

Reviewer #1 (Public review):

Summary:

This study uses high-throughput bacterial cell-surface display to identify LC3B-interacting peptides in the human proteome. The screen is unbiased, and this type of assay has not previously been used for selecting LC3B-interacting peptides. The screen was done with a library of 500,000 peptides, and they ended up with 427 peptides that they scored as high-confidence LC3B binders. The experiments performed are solid, and data are analyzed using well-documented methods and statistics.

The aim of the authors was to isolate LC3B-interacting peptides from the human proteome, and the screen succeeded in doing so. The selected set of peptides included several previously reported LIR motifs, but also many novel LC3B binding peptides that either contained or did not contain the canonical core LIR motif [WFY]xx[LVI].

Another aim was to identify binding determinants important for the LC3B interaction, and they made an interesting sequence logo based on selected LIR-containing peptides. However, this study does not really extend our knowledge related to binding determinants essential for LIR motifs in LC3B binding. They basically verify known characteristics, including the importance of varied types of electrostatic interactions supporting the docking of the core LIR into the LDS of LC3B.

Strengths:

The approach used here (high-throughput bacterial-surface-display) is new. The screen is unbiased, and the fact that peptides are directly tested for LC3B binding may facilitate the discovery of non-canonical LIR motifs. The screen appears to be highly selective and manages to distinguish between peptides that interact with LC3B and peptides that do not interact.

Weaknesses:

It is a limitation that no proteins are analyzed in this study. Further work is therefore needed to verify that identified LIR motifs are functional in full-length proteins and in cells.

<https://doi.org/10.7554/eLife.110318.1.sa2>

Reviewer #2 (Public review):

Summary:

To discover peptides that interact with autophagy-related protein LC3B and profile the key binding determinants, the authors screened a library of ~500,000 36-residue peptides derived from the human proteome using bacterial cell-surface display. Analysis of the screening data revealed exceptions to the reported LIR motif and a strong preference for negatively charged residues adjacent to the LIR.

These results support a refinement of the LIR motif definition and expand the network of candidate LC3B interaction partners.

Strengths:

High-throughput approach.

Weaknesses:

Lack of in vitro data and molecular dynamics simulations.

<https://doi.org/10.7554/eLife.110318.1.sa1>

Reviewer #3 (Public review):

Summary:

The LC3 family of proteins, which includes LC3B, are ubiquitin-like proteins that are covalently linked to phosphatidylethanolamine in the expanding autophagosomal membrane during autophagy. LC3 family members bind to short sequences of amino acids that reside within dynamic regions in a wide variety of proteins. These sequences, termed LC3 Interacting Regions (LIRs), were initially thought to function primarily to link LIR-containing autophagy cargo receptors to LC3 family members to help facilitate their capture during autophagy. However, the functional importance of LIRs in autophagy has broadened to include more general functions in autophagy as well. While a general consensus for LIR sequences has been described as [FWY]0-X1-X2-[LVI]3, recent work has suggested that additional sequences outside of the canonical LIR sequence can bind LC3 family members and play important roles in autophagy. In this manuscript by Kosmatka et al, the authors perform a high-throughput screen using bacterial surface display coupled with fluorescence-associated cell sorting to identify which human sequences can bind to LC3B. They identify a variety of peptides capable of binding LC3B, including peptides from proteins that have not previously been described as LC3B-binding proteins. The results from the bacterial surface display were then used to guide sequence analysis, mutational analysis, and structural studies to further characterize the range of LIR sequences that are capable of binding LC3B. Taken together, this work adds to the growing knowledge of how LIR sequences interact with LC3 family members and demonstrates which amino acids both inside and outside of the LIR sequence aid in binding. This work also identifies new potential LC3 binding proteins, which may play unknown roles in autophagy regulation. Lastly, this work reinforces the importance

of alternative LIR sequences such as the [WFY]0-X1-X2-[WFY]3 sequence, which the authors have dubbed the LIR+ sequence.

Strengths:

The manuscript uses a robust approach to identify and characterize different peptide sequences that can interact with LC3B. They validate a large number of sequences using biolayer interferometry (BLI) and attempt to correlate different amino acids with their binding affinity for LC3B. The large number of LC3B binding sequences and their dissociation constants adds significant new information to the field that will help others understand what sequences can bind to LC3B. The authors are also very careful to accurately report on their data and not overly interpret their findings.

Weaknesses:

After the authors identify proteins from their bacterial display assay, the remainder of the manuscript is focused on characterizing the different types of sequences that are identified in addition to validating the LC3B-LIR interactions using biochemical approaches, including BLI and X-ray crystallography. However, it's not entirely clear if the screen identified novel LC3B binders that interact with LC3B in cells. While I acknowledge that the focus of the manuscript is on the characterization of LIR sequences that can bind LC3B, it seems like a missed opportunity not to validate a few of the novel LC3B binders in vivo. This may result in the demonstration of novel binders of LC3B in cells and may further demonstrate the strength of this approach for identifying LC3 family member binding partners. Therefore, it would be helpful to look at a few proteins identified in the HC set that have not previously been identified as LC3B binders in cells to determine if they CO-IP with LC3B or interact with LC3B using a different approach.

<https://doi.org/10.7554/eLife.110318.1.sa0>