Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

🔓 OPEN ACCESS   ✓ Check for updates

Structure of the human autophagy factor EPG5 and the molecular basis of its conserved mode of interaction with Atg8-family proteins

Yiu Wing Sunny Cheung[a], Sung-Eun Nam[a], Gage M. J. Fairlie[a], Karlton Scheu[a], Jennifer M. Bui[b], Hannah R. Shariati[a], Jörg Gsponer[b], and Calvin K. Yip [iD][a]

[a]Life Sciences Institute, Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, BC, Canada; [b]Michael Smith Laboratories, Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, BC, Canada

ABSTRACT

The multi-step macroautophagy/autophagy process ends with the cargo-laden autophagosome fusing with the lysosome to deliver the materials to be degraded. The metazoan-specific autophagy factor EPG5 plays a crucial role in this step by enforcing fusion specificity and preventing mistargeting. How EPG5 exerts its critical function and how its deficiency leads to diverse phenotypes of the rare multi-system disorder Vici syndrome are not fully understood. Here, we report the first structure of human EPG5 (HsEPG5) determined by cryo-EM and AlphaFold2 modeling. Our structure revealed that HsEPG5 is constructed from helical bundles analogous to tethering factors in membrane trafficking pathways but contains a unique protruding thumb domain positioned adjacent to the atypical tandem LIR motifs involved in interaction with the GABARAP subfamily of Atg8-family proteins. Our NMR spectroscopic, molecular dynamics simulations and AlphaFold modeling studies showed that the HsEPG5 tandem LIR motifs only bind the canonical LIR docking site (LDS) on GABARAP without engaging in multivalent interaction. Our co-immunoprecipitation analysis further indicated that full-length HsEPG5-GABARAP interaction is mediated primarily by LIR1. Finally, our biochemical affinity isolation, X-ray crystallographic analysis, affinity measurement, and AlphaFold modeling demonstrated that this mode of binding is observed between *Caenorhabditis elegans* EPG-5 and its Atg8-family proteins LGG-1 and LGG-2. Collectively our work generated novel insights into the structural properties of EPG5 and how it potentially engages with the autophagosome to confer fusion specificity.

ABBREVIATIONS: ATG: autophagy related; CSP: chemical shift perturbation; eGFP: enhanced green fluoresent protein; EM: electron microscopy; EPG5: ectopic P-granules 5 autophagy tethering factor; GST: glutathione S-transferase; HP: hydrophobic pocket; HSQC: heteronuclear single-quantum correlation; ITC: isothermal titration calorimetry; LDS: LC3 docking site; LIR: LC3-interacting region; MD: molecular dynamics; NMR: nuclear magnetic resonance; TEV: tobacco etch virus

Introduction

Autophagy is an evolutionarily conserved degradation pathway crucial to maintaining cellular homeostasis. Under normal physiological conditions, autophagy operates at low basal levels to selectively remove and degrade misfolded or aggregated proteins, damaged organelles, and intracellular pathogens [1–3]. When cells face stress such as nutrient starvation, autophagy is upregulated to generate basic building blocks for downstream cellular processes through nonselective bulk degradation of cytoplasmic materials. Dysregulation of autophagy has been linked to aging and numerous human diseases ranging from different forms of cancers, neurodegenerative and infectious diseases [4,5].

The multi-step autophagy degradation process begins with the formation of a cup-shaped membrane precursor known as the phagophore, which elongates while sequestering cytoplasmic materials and ultimately self-fuses into a double-membraned transport vesicle termed the autophagosome. The cargo-laden autophagosome is then transported to and fuses with the late endosome/lysosome to form the autolysosome, where the hydrolytic lysosomal enzymes degrade the content. Finally, building blocks such as amino acids are returned to the cytoplasm through permeases for anabolic processes [6–8]. Genetic screens performed in yeast and the nematode *Caenorhabditis elegans* led to the discovery of a group of *Atg* (autophagy related) and *epg* (ectopic P granules) genes and their corresponding proteins that regulate different steps of autophagy [9,10]. Subsequent biochemical and cell biology studies on the core Atg proteins have generated mechanistic insights into the early steps of autophagy including autophagosome biogenesis. Notably, the Atg1/ULK1 kinase complex and the Vps34-Atg14 class III phosphatidylinositol 3-kinase complex coordinate the induction of autophagy and the initiation of phagophore formation. The expansion of the phagophore involves coordinated action between the Atg2 tether and intermembrane lipid transporter and the Atg9 transmembrane protein with a scramblase activity. Last, the Atg12 and Atg8 ubiquitin-like conjugation systems generate Atg8 covalently

---

linked to the lipid phosphatidylethanolamine (PE), which localizes to the growing phagophore membrane and mediates cargo loading through binding selective autophagy receptors [7,11,12]. Lipidated Atg8 is also assumed to mediate autophagosome enclosure through membrane tethering [13]. In contrast, much less is known about the later stages of autophagy and in particular how fusion specificity between the cargo-laden autophagosome and late endosome/lysosome is enforced.

Recent studies in *C. elegans* and mammalian cells have identified several proteins and protein complexes involved in autophagosome-lysosome fusion. Notably, two sets of soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor/SNARE proteins were shown to mediate fusion between autophagosome and late endosome/lysosome: (1) autophagosomal STX17 (syntaxin 17), SNAP29, and late endosomal/lysosomal VAMP7 and VAMP8; (2) the autophagosomal YKT6, SNAP29, and late endosomal/lysosomal STX7 [14,15]. This was followed by the discovery of a diverse group of proteins and protein complexes that serve as tethers to link autophagosomes to lysosomes, including the HOPS complex, ATG14, PLEKHM1, EPG5 (ectopic P-granules 5 autophagy tethering factor; ~290 kDa), TECPR1, and BIRC6/BRUCE [16–21]. Among these autophagic tethers, the large EPG5 functions in conferring fusion specificity. EPG5 deficiency causes autophagosomes to mistarget and incorrectly fuse with other endocytic vesicles, leading to the accumulation of enlarged vesicles that contain both autophagic and endocytic markers [10,19]. Biochemical and colocalization studies suggested that EPG5 is recruited to the lysosomal membrane by the small GTPase RAB7 and engages the autophagosome by interacting with Atg8-family proteins [19]. EPG5 also modestly enhances the formation of the STX17-SNAP29-VAMP8 complex in an *in vitro* fusion assay. These findings led to the proposal that EPG5 functions as a tethering factor in the last step of autophagy. In agreement with this, our previous single-particle negative stain electron microscopy (EM) analysis revealed that human EPG5 (HsEPG5) adopts an extended, "shepherd-staff"-shaped overall architecture [22]. However, the limited resolution of this analysis has precluded further comparison of the structural properties of HsEPG5 with tethering factors in other intracellular transport pathways.

Human cells express six isoforms of Atg8-family proteins that are grouped into the LC3 (MAP1LC3A, MAP1LC3B, MAP1LC3C) and GABARAP (GABARAP, GABARAPL1, GABARAPL2) subfamilies, respectively [23,24]. EPG5 interacts with Atg8 proteins and our recent work showed that HsEPG5 preferentially binds members of the GABARAP subfamily previously demonstrated to be critical to autophagosome-lysosome fusion [22]. Atg8-interacting proteins characterized thus far bind their cognate Atg8 proteins via their LC3-interacting region (LIR) motif which has a consensus core sequence $[\Theta_0-X_1-X_2-\Gamma_3]$ where Θ in position 0 represents an aromatic residue (W/F/Y) and Γ in position 3 represents a hydrophobic residue (L/I/V) [25]. The canonical mode of interaction involves docking the aromatic and hydrophobic residues of the LIR into two hydrophobic pockets (HP1 and HP2) at the LC3 docking site (LDS) on the surface of the Atg8 protein [25]. However, it is thought that other non-canonical modes of interaction remain to be discovered as only a limited number of Atg8 interactors have been characterized. The majority of characterized human LC3/GABARAP interactors contain one main LIR motif. EPG5, however, has two LIR motifs (LIR1 and LIR2) that are arranged in tandem and both of these motifs are required for full function [19]. Exactly why EPG5 requires two LIR motifs to interact with GABARAP proteins and the precise mode of interaction remain unclear. Our previous biochemical analysis revealed that while LIR1 plays a more dominant role in mediating HsEPG5-GABARAP interaction in the context of the full-length protein, a HsEPG5 LIR2 peptide binds at higher affinity to GABARAP than the LIR1 peptide does [22]. Although our crystallographic studies showed that HsEPG5 LIR2 docks to the canonical LDS of GABARAPL1, how LIR1 interacts with GABARAP proteins remains unknown.

Here, we report the first structural model of HsEPG5, determined by cryo-electron microscopy (cryo-EM) and AlphaFold2-based modeling. Our model revealed that HsEPG5 is constructed from helical bundles like members of the complexes associated with tethering containing helical rods (CATCHR) family of tethers [26]. Our model also showed that the tandem LIR motifs involved in autophagosome engagement are located in a more central location of HsEPG5. To gain further insights into the mode of interaction between HsEPG5 and GABARAP, we carried out AlphaFold2-based structural modeling, nuclear magnetic resonance (NMR) spectroscopy, and molecular dynamics (MD) simulations. Our data showed that a HsEPG5 LIR1 peptide alone, like the HsEPG5 LIR2 peptide, binds to the canonical LDS of GABARAPL1. Interestingly, only one of the two LIRs engages in canonical interaction with GABARAPL1 for a peptide that contains the two tandemly arranged LIR motifs. This suggests that the tandem LIR does not engage in multivalent interaction with GABARAPL1. Further AlphaFold3-based structural modeling of the full-length HsEPG5 in complex with GABARAPL1 and co-immunoprecipitation with GABARAP indicated that only LIR1 of the tandem is involved in GABARAP interactions. We also conducted biochemical, biophysical, and structural studies on *C. elegans* EPG-5 and found that although this nematode ortholog binds two *C. elegans* Atg8 proteins LGG-1 and LGG-2 equally well via its tandem LIR, one of the two LIR motifs (LIR1) binds more tightly to LGG-1 and docks into the LDS. These results suggest the mode of interaction between EPG5 and Atg8 is likely conserved across metazoans.

Results

HsEPG5 is constructed from helical bundles and contains a protruding thumb region

Our previous negative stain EM analysis revealed that HsEPG5 adopts an extended architecture. To gain further insights into its structural properties, we attempted to visualize HsEPG5 at higher resolution by cryo-EM. The fact that HsEPG5 is prone to aggregation at high concentrations prevented us from obtaining vitrified specimens free of aggregated and/or stacked particles. By subjecting HsEPG5 to PEGylation and then adsorbing the modified protein to

grids overlaid with a graphene oxide support layer, we were able to minimize sample aggregation and obtain a suitable vitrified specimen for cryo-EM data collection. Multiple rounds of 2D classification of 875,590 particles selected from this dataset and followed by *ab initio* reconstruction and heterogeneous refinement resulted in a density map with an average resolution of 7.85 Å (Figures 1A and S1A). In addition to the curved "hook" domain and an extended "shaft" domain observed in our negative stain EM analysis, our cryo-EM map shows that HsEPG5 has a prominent centrally-located protruding domain we referred to as the "thumb". Despite the modest resolution of this density map, secondary structures are discernable in the thumb and adjacent regions. By contrast, the bulk of the hook and the terminal region of the shaft were less well resolved likely due to their inherent flexibility (Figure S1B).

To construct a structural model of HsEPG5, we first examined the predicted model of this protein in the AlphaFold Protein Structure Database (AlphaFoldDB: Q9HCE0) (Figure 1B) [27,28]. Except for the N-terminal disordered region, this model contains the same three domains (hook, thumb, shaft) as observed in our cryo-EM map (Figure 1A,B). However, we were unable to fit this model into the cryo-EM

density map as the hook and shaft domains of this model curled toward the tip of the thumb, resulting in an overall "ω"-shaped architecture. We thus segmented HsEPG5 into three parts that correspond to each of the three domains (hook: 534–1519, thumb: 288–533,1524–1709, shaft: 1710–2579) and then carried out independent AlphaFold2 predictions of these HsEPG5 fragments (Figure S1C). We could dock the three predicted models into the EM density, with the most optimal fit for the thumb and neighboring regions. By reconnecting the chains with refinement, we were able to generate a composite structural model of HsEPG5 covering the residues 270 to the C-terminal residue 2579, excluding the flexible loop region between residues 1391 to 1523 (Figures 1C and S1C).

Our structural model reveals that the N-terminal region of HsEPG5 forms the coiled-coil and beta sheets of one-half of the thumb domain, followed by the construction of the full hook domain composed of a parallel series of alpha helices. The protein then loops back, through an unstructured region, to form the other half of the thumb before extending laterally toward the C-terminus to form the shaft domain that is composed of alpha-helical bundles (Figure 1C). Although the use of helical bundles to generate extended structures is
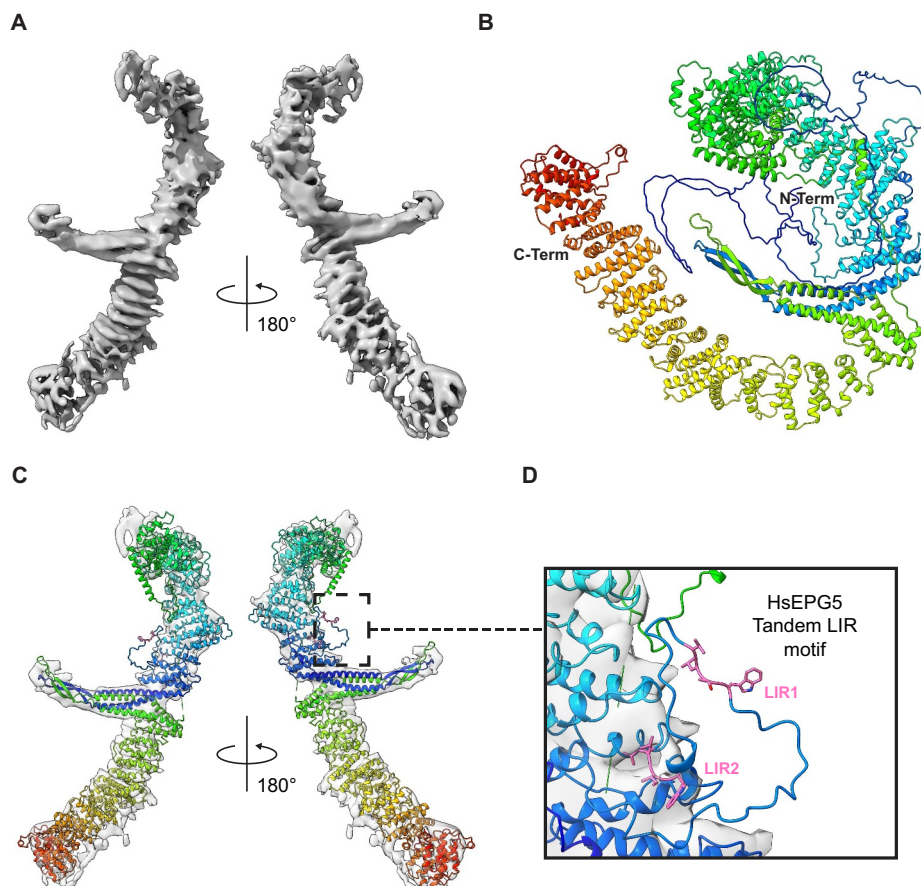

Figure 1. Composite structural model of HsEPG5. (A) Refined cryo-EM density map of HsEPG5. (B) AlphaFold3 predicted model of HsEPG5. Cartoon model is colored in a rainbow color scheme (blue to red) from the N terminus to the C terminus. (C) Fitting of the composite structural model of HsEPG5 into cryo-EM density map. Cartoon representation of HsEPG5 is colored in a rainbow color scheme (blue to red) from the N terminus to the C terminus. The tandem LIR motifs (LIR1 and LIR2) are represented in pink sticks, located on the flexible loop that has missing EM density. (D) Close-up view of the tandem LIR motifs on the flexible loop indicated by the dotted lined box in (C).

also observed for tethering factors such as members of the CATCHR family [26], the intertwined domain formation and the protruding thumb are features unique to HsEPG5.

Moreover, our structural model also shows that the loop region containing the tandem LIR motifs that mediate interaction with Atg8 proteins (residue 521 to residue 581) is located in the hook domain near the center of the protein and directly across the thumb (Figure 1D). Although this loop is not visualized in our cryo-EM map, AlphaFold modeling suggested that LIR1 is present in the unstructured middle region of this loop while part of the LIR2 folds into a short helix that forms coil-coil interactions with the "spine" of the hook.

HsEPG5 LIR1 binds the LDS of GABARAPL1

We next focused on clarifying how the tandem LIR mediates HsEPG5 interaction with GABARAP proteins. Our previous mutagenesis and biochemical affinity isolations showed that although LIR1 in the tandem motifs has a more dominant role in mediating HsEPG5 binding to GABARAP, a peptide corresponding to LIR2 ($_{560}$DEDPETSWILLN$_{571}$) exhibits higher affinity interaction with the GABARAP proteins than a peptide corresponding to LIR1 ($_{546}$GSGTWTLVDEG$_{556}$). Because our previous crystal structure of HsEPG5 LIR2 in complex with GABARAPL1 shows that LIR2 docks to the canonical LDS of this Atg8 protein, we hypothesized that LIR1 would bind to a non-canonical site on GABARAPL1 [22]. However, we were unable to crystallize the HsEPG5 LIR1-GABARAP complex. Recent studies have demonstrated the power of the AlphaFold-Multimer algorithm in predicting, at high accuracy, potential protein-protein binding interfaces between LIR and Atg8 proteins [29]. We conducted a proof-of-concept ColabFold-based AlphaFold-Multimer run of HsEPG5 LIR2 peptide in complex with GABARAPL1 [30]. The predicted structural model resembles our experimentally determined crystal structure [22], with a root-mean-square deviation (r.m.s.d.) of 0.9 Å (Figure S2A,B). We next carried out AlphaFold-Multimer prediction of the structure of HsEPG5 LIR1 peptide in complex with GABARAPL1. Contrary to our hypothesis, the generated structural model shows LIR1 alone similarly binds the LDS of GABARAPL1 as LIR2 (Figure S2C).

To experimentally validate the predicted LIR1-GABARAPL1 structural model, we used NMR spectroscopy to map the HsEPG5 LIR1 binding site on GABARAPL1. More specifically, we titrated unlabeled HsEPG5 LIR1 or LIR2 peptide, into ^{15}N-labeled GABARAPL1 and then determined regions of perturbation by examining the acquired [^{15}N,^{1}H]-heteronuclear single quantum correlation (^{15}N-HSQC) spectra. To facilitate chemical shift assignments, we repeated the NMR spectroscopy experiment, as previously reported by Rozenknop et al., by titrating unlabeled NBR1-LIR into ^{15}N-GABARAPL1 [31]. ^{15}N-HSQC spectra were recorded for each titration point, from a protein:peptide molar ratio of 1:0 to 1:1.5 (Fig. S2D). The chemical shifts we obtained aligned with those published, which enabled pattern-based assignment of the resonances in the

apo-^{15}N-GABARAPL1 HSQC spectrum and subsequently the spectra of the complexes involving LIR1 or LIR2.

As a control, amides showing amide chemical shift perturbations (CSPs) upon titration with a peptide model of LIR2 indeed mapped to the LDS site in GABARAPL1 identified previously through X-ray crystallographic studies (Figure 2A-C). For experiments involving the peptide model of LIR1, GABARAPL1 residues that showed the largest amide CSPs included K24, R28, V33, K46, Y49, L50, S53, L63, R65, K66, Y103, F104, and V107 [32] (Figure 2E,F). Through mapping of these perturbed residues on the surface of GABARAPL1 we found, in agreement with the AlphaFold-Multimer prediction, HsEPG5 LIR1 peptide docks into the canonical LDS site in GABARAPL1 in a similar fashion as HsEPG5 LIR2 (Figure 2C,G). Comparing the CSP values induced by the binding of the LIR1 and LIR2 peptides revealed some differences in the hydrophobic pocket where the C-terminal end of the LIR motifs dock (Fig. S2E and F). In the case of the LIR1 peptide, binding occurred in the fast to intermediate exchange regime (Figure 2E). Therefore, we were able to fit the CSP values of the ten most perturbed protein amides to the equation for a simple 1:1 isotherm and obtain a K_D value of 14 μM (Fig. S2G). This is consistent with the value we previously measured through isothermal titration calorimetry (ITC) (8.5 μM) [22]. In contrast, binding of the LIR2 peptide occurred in the intermediate to slow exchange regime (Figure 2A). Although precluding the determination of a corresponding K_D value, this confirmed that the LIR2 peptide binds GABARAPL1 with a higher affinity or slower exchange rate, than the LIR1 peptide.

HsEPG5 tandem LIR does not engage in multivalent interaction with GABARAPL1

Because HsEPG5 LIR1 and LIR2 are both capable of binding the LDS of GABARAPL1 in isolation, we next asked if a non-canonical binding site is involved when LIR1 and LIR2 are present in a tandem arrangement. To test this hypothesis, we first conducted an AlphaFold-Multimer run to predict how HsEPG5 tandem LIR (HsEPG5[541–581]) interacts with GABARAPL1. This peptide encompasses both LIR1 and LIR2 as well as a putative helical region extending from the C terminus, based on the previous proposal that a helical element C-terminal to the LIR is important for GABARAP interaction [33–38]. Contrary to our hypothesis, the five structural models generated from our run showed that only LIR1 in the tandem binds the LDS of GABARAPL1 while LIR2 is not directly involved in protein-protein interaction (Fig. S2H). We next conducted NMR spectroscopy analysis by titrating purified recombinant tandem LIR peptide (HsEPG5[541–581]) into ^{15}N-labeled GABARAPL1 and acquired the corresponding HSQC spectrum. In agreement with the AlphaFold2 prediction, the residues that experienced the largest CSPs (E17, R28, I32, K46, Y49, L50, L55, T56, L63, I68, Y103, and F104) mapped exclusively to the LDS of GABARAPL1, indicating the tandem LIR only contacts GABARAPL1 at the canonical LDS (Figure 2I-K). In other words, HsEPG5 tandem LIR, at least in the context of this peptide and within the conditions of
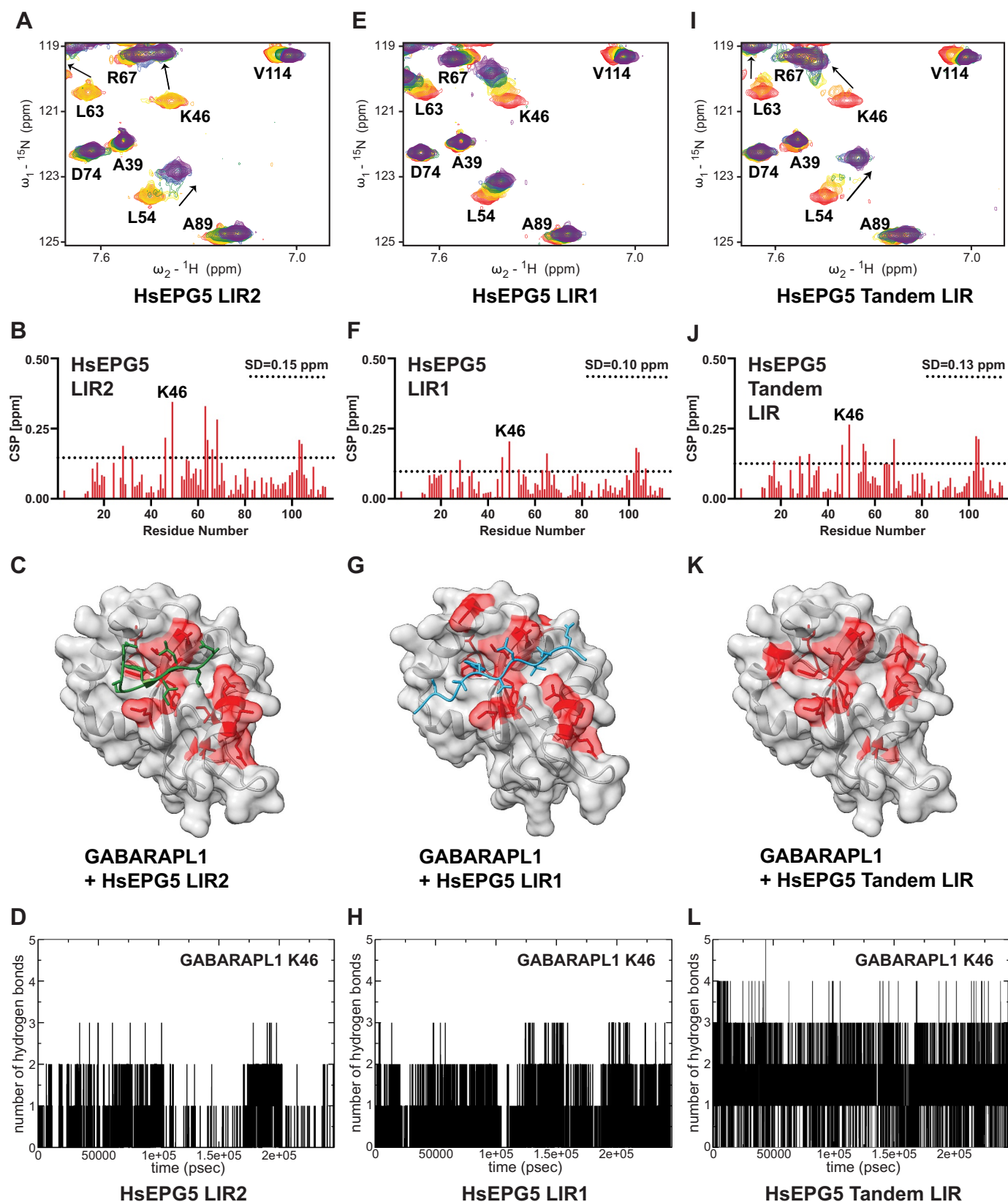
Figure 2. NMR-monitored titrations and MD simulations of GABARAPL1 in complex with HsEPG5 LIR peptides. (A, E, I) Overlaid regions of the ^{15}N-HSQC spectra of ^{15}N-GABARAPL1 titrated with unlabeled peptides corresponding to (A) HsEPG5 LIR2, (E) HsEPG5 LIR1 and (I) HsEPG5 tandem LIR. The rainbow color scheme represents the titration of peptides from a protein:peptide molar ratio of 1:0 (red) to 1:1.5 (purple) for HsEPG5 LIR1 and tandem LIR peptide, whereas the titration of HsEPG5 LIR2 is represented from 1:0 molar ratio (red) to 1:1 molar ratio (purple). (B, F, J) Combined ^{1}H-^{15}N CSPs of GABARAPL1 with and without saturating amounts of the (B) HsEPG5 LIR2 peptide, (F) HsEPG5 LIR1 peptide or (J) HsEPG5 tandem LIR peptide. Calculated CSPs for each residue are represented in bar graphs and the largest CSP residue, LYS46, are labeled. The dotted lines represent the standard deviation (S.D.) for each graph. (C) Crystal structure of GABARAPL1 (ribbon and transparent surface representation in gray) in complex with HsEPG5 LIR2 peptide (ribbon and sticks representation in green) (PDB:7JHX) [25]. Residues that experienced CSPs greater than 1 S.D. upon the titration of HsEPG5 LIR2 peptide are labeled in red. (G) AlphaFold2 predicted model of GABARAPL1 (ribbon and

the NMR experiment, does not engage in multivalent interactions with GABARAPL1. Although we cannot determine which LIR segment within the tandem binds GABARAPL1 from our NMR data, we detected CSPs in the fast to intermediate exchange regime (Figure 2I), similar to what was observed when HsEPG5 LIR1 peptide was titrated into GABARAPL1.

To further examine the nature of the HsEPG5-GABARAP interactions, we next carried out a series of MD simulations using the AlphaFold-Multimer-derived structural models of GABARAPL1 in complex with LIR1 (Fig. S2C), LIR2 (Fig. S2B), and the tandem LIR (Fig. S2I) as starting models. Our data revealed that LIR2 remained docked to the LDS throughout the 2.5 microsecond (μs) MD trajectory although it is highly mobile (Fig. S2J). LIR1 also remained docked to the LDS throughout the simulation but is less mobile than LIR2 (Fig. S2J). For the tandem LIR peptide (HsEPG5[546–571]) that encompasses both LIR1 and LIR2 and the linker region in between, the AlphaFold-predicted model depicted that LIR1 was docked to the LDS. Notably, our data demonstrated that during the MD simulation, only LIR1 remained bound to the LDS, and the higher binding affinity LIR2 did not displace LIR1 at the LDS (Fig. S2I and K).

Our NMR titration experiments showed that K46 of GABARAPL1 exhibited the most significant CSP (Figure 2B, F,J). To further dissect the differences in how the three peptides interact with GABARAPL1, we next analyzed our three sets of simulation data to assess the hydrogen bonding potential with K46 of GABARAPL1. Throughout the simulation, LIR2 displayed the lowest frequency of number of hydrogen bonds with K46 among the three peptides (Figure 2D,H,L). Conversely, LIR1 and tandem LIR show similar hydrogen bond frequencies with GABARAPL1 K46, consistent with their similar CSP patterns observed in our NMR titration experiments (Figure 2E,I). Interestingly, for the tandem LIR, our simulation data showed that the side chain of GLU559 and GLU561 located at the linker region can potentially interact with K46 of GABARAPL1 (Fig. S2L), suggesting that the linker region may play a role in facilitating or enhancing HsEPG5-GABARAPL1 interaction.

Full-length HsEPG5 binds GABARAP primarily through LIR1

The peptides we analyzed in our NMR titration and MD simulation experiments represent only a segment of the HsEPG5 protein. To determine if local structure may affect the accessibility and behavior of the tandem LIR, we used the recently released AlphaFold3 algorithm to predict the structural model of the HsEPG5-GABARAPL1 complex [39]. The five models, which resemble one another, show that HsEPG5 engages GABARAPL1 exclusively via LIR1, while LIR2 was observed to adopt a helical structure that docks the backbone

of the hook and does not directly contact with GABARAPL1 (Figures 3A,B, and S3A,B). We note that while the overall ipTM and pTM scores of these models (0.7 and 0.48) were considered only of medium quality, these values are influenced by the inherent flexibility between the three EPG5 domains and the intrinsically disordered nature of the N-terminal and middle regions. In comparison, the structural model of apo-HsEPG5 obtained from AlphaFold3 only has a pTM score of 0.53 (Figure 1B). Furthermore, when we performed AlphaFold3 modeling using only the sequence corresponding to the hook domain (HsEPG5[425–1519]) where the tandem LIR is located and is devoid of the unstructured regions, we obtained structural models of HsEPG5 hook-GABARAPL1 complex with significantly improved ipTM and pTM scores of 0.77 and 0.73, respectively (Fig. S3C). Consistent with the full-length models, LIR1 was observed to be the only HsEPG5 element involved in binding GABARAPL1 (Fig. S3D and E). Overall, our *in silico* modeling suggests that LIR1 likely serves as the primary GABARAP-interaction motif, with LIR2 playing a more structural role in this region of HsEPG5.

To validate this hypothesis, we generated four HEK293 cell lines that stably express wild-type (WT) HsEPG5, or one of the three HsEPG5 loss-of-function LIR mutants (W550A for LIR1, W567A for LIR2, and W550A W567A double mutations for tandem LIR). We next transiently expressed enhanced green fluorescent protein (eGFP)-tagged GABARAP in each of these four cell lines and performed co-immunoprecipitation to assess the interaction between the different HsEPG5 LIR mutants and GABARAP. Our data showed that GABARAP binds WT HsEPG5 and HsEPG5[W567A] but not with the HsEPG5[W550A] and HsEPG5[W550A,W567A] mutants (Figure 3C). In other words, HsEPG5-GABARAP interaction is severely affected by radical mutation to LIR1 but not LIR2, confirming the importance of LIR1 in mediating this interaction. We did, however, observe that HsEPG5[W567A] shows slightly weaker binding to GABARAP compared to WT HsEPG5 (Figure 3C), suggesting that the tryptophan (W567) mutation to alanine may impact the local structure.

CeEPG5 LIR1 plays a more dominant role in mediating interaction with LGG-1 and LGG-2

The tandem LIR is found in all EPG5 orthologs, including the originally characterized *C. elegans* EPG-5 (CeEPG5). Previous studies showed that CeEPG5 binds the *C. elegans* Atg8 protein LGG-1 via its tandem LIR [19]. However, *C. elegans* expresses two Atg8 isoforms: LGG-1 and LGG-2 that resemble human GABARAP and LC3, respectively [40]. To determine if CeEPG5 can interact with LGG-2, we first produced recombinant CeEPG5 using the baculovirus-insect cell expression system. We also purified recombinant full-length LGG-1 and an N-terminally truncated LGG-2[16–130] which is less prone

---

transparent surface representation in gray) in complex with HsEPG5 LIR1 peptide (ribbon and sticks representation in cyan). Residues that experienced CSPs greater than 1 S.D. upon the titration of HsEPG5 LIR1 peptide are labeled in red. (K) Crystal structure of GABARAPL1 (ribbon and transparent surface representation in gray) (PDB:2R2Q). Residues that experienced CSPs greater than 1 S.D. upon the titration of HsEPG5 tandem LIR peptide are labeled in red. (D, H, L) Plots of number of hydrogen bonds as a function of simulated time formed between LYS46 on GABARAPL1 and the (D) HsEPG5 LIR2 peptide, (H) HsEPG5 LIR1 peptide or (L) HsEPG5 tandem LIR peptide.

A

B

C

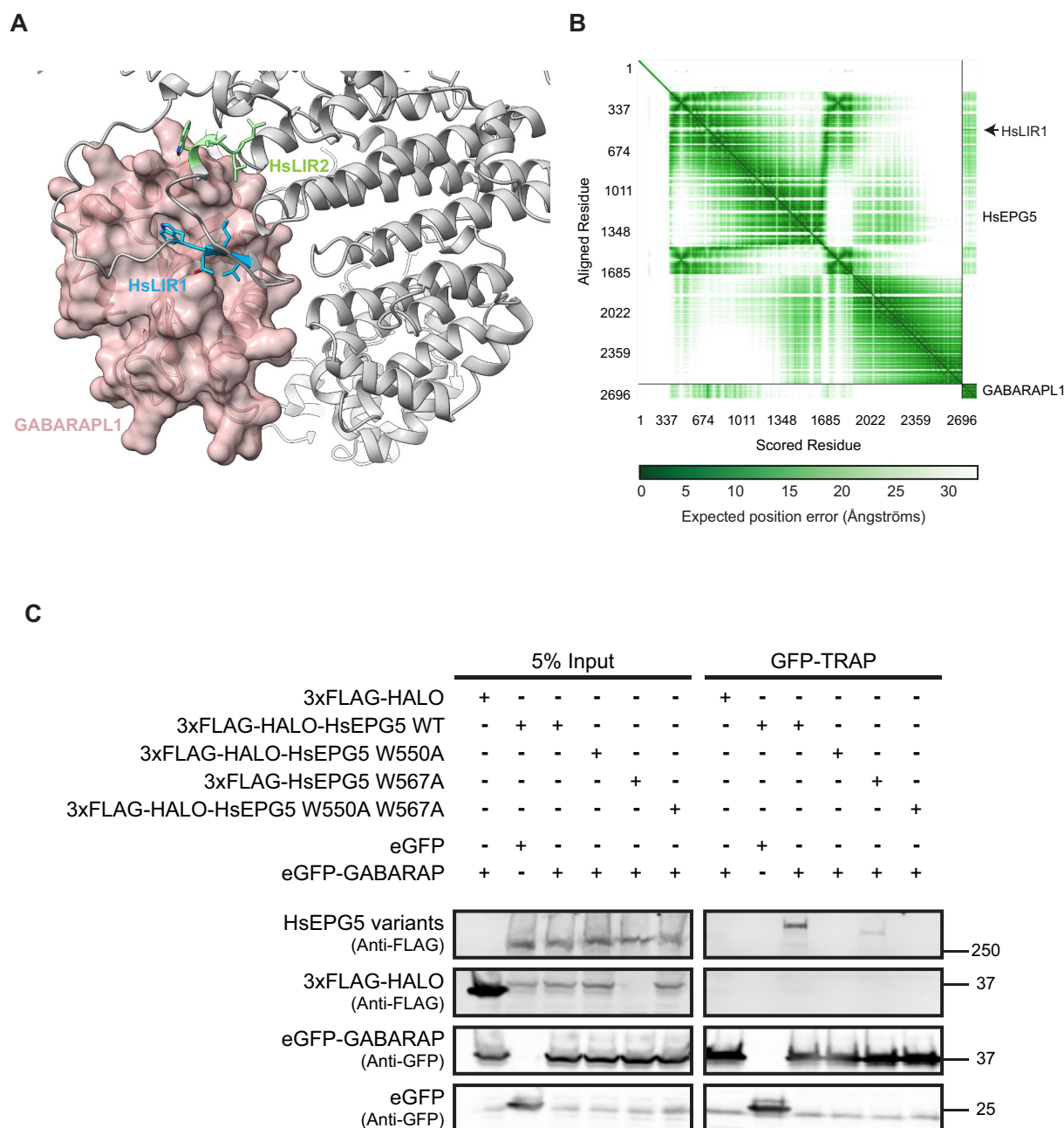|	5% Input						GFP-TRAP					
3xFLAG-HALO	+	-	-	-	-	-	+	-	-	-	-	-
3xFLAG-HALO-HsEPG5 WT	-	+	+	-	-	-	-	+	+	-	-	-
3xFLAG-HALO-HsEPG5 W550A	-	-	-	+	-	-	-	-	-	+	-	-
3xFLAG-HsEPG5 W567A	-	-	-	-	+	-	-	-	-	-	+	-
3xFLAG-HALO-HsEPG5 W550A W567A	-	-	-	-	-	+	-	-	-	-	-	+
eGFP	-	+	-	-	-	-	-	+	-	-	-	-
eGFP-GABARAP	+	-	+	+	+	+	+	-	+	+	+	+

Figure 3. AlphaFold3 prediction and co-immunoprecipitation of full-length HsEPG5 binding to GABARAPL1 and GABARAP. (A) Close-up view of AlphaFold3 predicted structure of GABARAPL1 (ribbon and transparent surface representation in salmon) binding to HsEPG5 LIR1 (ribbon and sticks representation in cyan) of the tandem LIR motifs at the hook region of full-length HsEPG5 (ribbon representation in gray). HsEPG5 LIR2 (ribbon and sticks representation in light green) forms a partial helix and is not involved in GABARAPL1 interaction. (B) Predicted align error (PAE) plot of GABARAPL1 in complex with full-length HsEPG5. (C) GFP-TRAP co-immunoprecipitation of FLAG-tagged wild-type and mutant HsEPG5 using eGFP-tagged GABARAP as bait. HEK293 cells stably expressing 3×FLAG-HALO (control) or 3×FLAG-HALO-HsEPG5, or 3×FLAG-HALO-HsEPG5[W550A], 3×FLAG-HsEPG5[W567A], 3×FLAG-HALO-HsEPG5[W550A,W567A] mutants are co-transfected with eGFP (control) or eGFP-GABARAP for co-immunoprecipitation experiments. Representative western blot was probed by anti-FLAG antibodies and anti-GFP antibodies. Experiments were performed in triplicates.

to precipitation but retained its physiological function using a bacterial expression system [40]. Using these proteins, we conducted glutathione S-transferase (GST) affinity isolation assays. Our data showed that CeEPG5 binds equally well to LGG-2[16–130] compared to LGG-1 (Figure 4A,B). Thus, unlike HsEPG5, CeEPG5 does not show preferential binding to one specific type of Atg8.

To systematically analyze the contribution of the corresponding LIR1 and LIR2 in mediating interaction with LGG-1 and LGG-

2, we first designed CeEPG5 mutants containing alanine replacement mutation to the first aromatic residue of LIR1 ($_{507}$WEIL$_{510}$) or LIR2 ($_{523}$FVTI$_{526}$) or both LIR motifs in the tandem. We next produced recombinant mutant CeEPG5 proteins and subjected them to GST-affinity isolation assays. We found that mutation to the first LIR (W507A) severely diminished CeEPG5 interaction with LGG-1, while the effects of mutation to the second LIR (F523A) were quite mild (Figure 4C,D). A similar overall trend was observed in the affinity isolation experiments involving LGG-
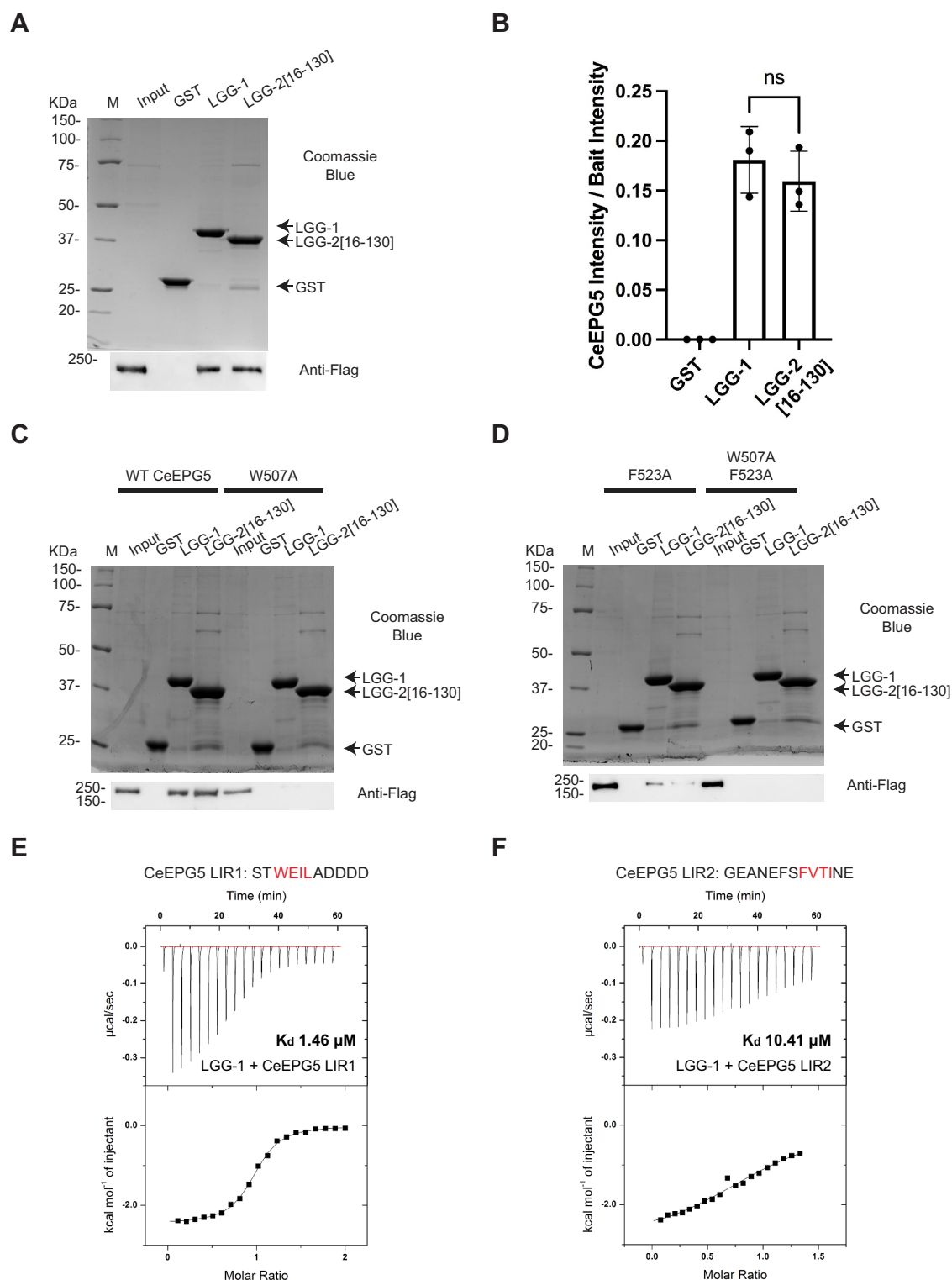
Figure 4. CeEPG5 interacts with LGG-1 and LGG-2[16–130] proteins via its tandem LIR motifs. (A-D) *In vitro* affinity isolation assays of CeEPG5 with LGG-1 and LGG-2 [16–130]. Precipitation of FLAG-tagged wild-type and mutant CeEPG5 using GST (control) and N-terminally GST-tagged LGG-1 and LGG-2[16–130] proteins as baits. Representative SDS – PAGE stained with Coomassie Blue (top panel) shows the input. Representative western blot (bottom panel) was probed by anti-FLAG antibodies. M represents the protein marker. Experiments were performed in triplicates. (A) *In vitro* affinity isolation assays of CeEPG5 with GST (control) and N-terminally GST-tagged LGG-1 and LGG-2(16–130) proteins used as baits. (B) Quantification of CeEPG5 binding to GST (control) and GST-LGG-1 and GST-LGG-2 [16–130]. Data are shown as mean ± SEM of three individual experiments. (C and D) Wild-type and CeEPG5 mutant (CeEPG5[W507A], CeEPG5[F523A], and CeEPG5[W507A, F523A]) affinity isolation analysis using GST-LGG-1 and GST-LGG-2[16–130] as bait. (E-F) Isothermal titration calorimetry (ITC) analysis of CeEPG5 LIR motifs with LGG-1. ITC results of CeEPG5 LIR1 peptide ($_{505}$STWEILADDDD$_{515}$) (600 μM) (E), and LIR2 peptide ($_{516}$GEANEFSFVTINE$_{528}$) (400 μM) (F) titrating into LGG-1 (60 μM). The top diagram in each ITC plot shows the raw data and the bottom diagram shows the integrated data.

2[16–130], but the effect was more pronounced as the CeEPG5^{W507A} mutant nearly abolished interaction while the CeEPG5^{F523A} mutant severely diminished the LGG-2[16–130] interaction (Figure 4C,D). Last, the W507A F523A double mutation completely abolished CeEPG5 interaction with both LGG-1 and LGG-2[16–130] (Figure 4D).

Our affinity isolation results revealed that much like what we have observed for HsEPG5, LIR1 in the CeEPG5 tandem motifs plays a more important role in mediating interaction with LGG-1 and LGG-2. To more quantitatively assess binding affinity, we synthesized two peptides corresponding to CeEPG5 LIR1 ($_{505}$STWEILADDDD$_{515}$) and LIR2 ($_{516}$GEANEFSFVTINE$_{528}$) and used ITC to measure their binding constants with the *C. elegans* LGG-1. Our data showed that the LIR1 peptide has an affinity ten times stronger toward LGG-1 (K_D of 1.46 μM)

than the LIR2 peptide (K_D of 10.41 μM) (Figure 4E,F, and Table S1). This observation could be explained by the fact that the aromatic position "0" of the CeEPG5 LIR2 contains a phenylalanine, which has been previously shown to confer lower binding affinity toward the Atg8 proteins compared to LIR motifs with tryptophan in this position [31]. Thus, the more dominant LIR in the CeEPG5 tandem motifs is the one that represents the high-affinity binding site.

CeEPG5 LIR1 binds to the LDS of LGG-1

Based on the high affinity observed between CeEPG5 LIR1 and LGG-1, we postulated that this LIR docks to the LDS of this Atg8 isoform. To verify this hypothesis, we determined the crystal structure of LGG-1 in complex with CeEPG5 LIR1 at 1.6 Å
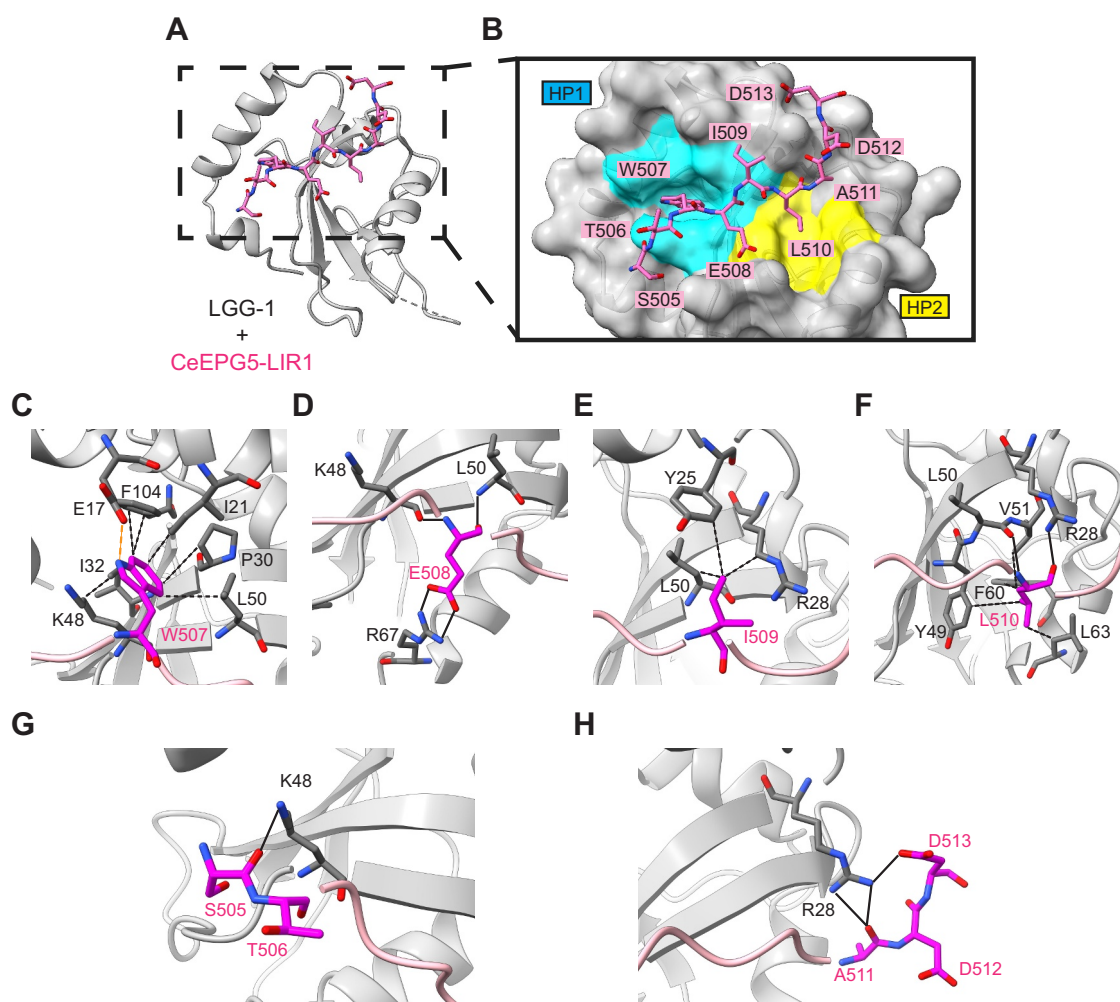

Figure 5. Crystal structure of the CeEPG5 LIR1-LGG-1 complex. (A) Structure of LGG-1 (ribbon representation in gray) in complex with CeEPG5 LIR1 (sticks representation in pink) at 1.6 Å. Dotted line represents the missing residues LEU70, ARG71, and PRO72 at the loop region. (B) Close-up view of CeEPG5 LIR1 peptide (sticks representation in pink) binding to the LDS on LGG-1 (ribbon and transparent surface representation in gray). The aromatic residue TRP507 and the hydrophobic residue LEU510 of CeEPG5 LIR1 are inserted into hydrophobic pocket 1 (HP1 in cyan) and hydrophobic pocket 2 (HP2 in yellow) through hydrophobic interactions, respectively. (C–F) Close-up view of each CeEPG5 LIR1 motif residue (in magenta) interacting with the residues on LGG-1 (in dark gray). Black dotted lines and solid lines represent hydrophobic interaction and hydrogen bonding, respectively; orange dotted lines represent electrostatic interaction. (C) Side chain of TRP507 interacts with side chains of ILE21, PRO30, ILE32, LYS48, LEU50, and PHE104 through hydrophobic interactions, while interacting with the carboxyl group on GLU17 side chain through electrostatic interaction. (D) Carboxyl group on GLU508 side chain forms hydrogen bonds with the guanidinium group on ARG67 side chain, while the main chain of GLU508 forms hydrogen bonds with the main chain of LYS48 and LEU50. (E) Side chain of ILE509 interacts with the side chain of TYR25, ARG28 and LEU50 through hydrophobic interactions. (F) Side chain of LEU510 forms hydrophobic interactions with the side chain of TYR49, VAL51, PHE60 and LEU63, while the main chain of LEU510 forms hydrogen bonds with the guanidinium group on ARG28 side chain and main chain of LEU50. (G-H) Close-up view of each CeEPG5 LIR1 motif flanking residues (in magenta) interacting with the residues on LGG-1 (in dark gray). Black solid lines represent hydrogen bonds. (G) CeEPG5 LIR1 N-terminal residue SER505 main chain forms hydrogen bond with LYS48 side chain. (H) CeEPG5 LIR1 C-terminal residues ALA511 main chain and ASP513 side chain form hydrogen bond with guanidinium group on ARG28 side chain.

resolution (Table S2, Figure 5A,B, and Fig. S4A and B). Our structure confirms that CeEPG5 LIR1 engages LGG-1 at the canonical LDS. Notably, CeEPG5 LIR1 adopts an extended beta-strand conformation with the aromatic residue CeEPG5 W507 inserting into the hydrophobic pocket (HP1) and forming hydrophobic and electrostatic interactions with side chains E17, I21, P30, I32, K48, L50, and F104 of LGG-1 (Figure 5C). Conversely, the hydrophobic residue CeEPG5 L510 docks into the second hydrophobic pocket (HP2), where it contacts side chains Y49, V51, F60, and L63 of LGG-1 through hydrophobic interactions (Figure 5F). The backbone of CeEPG5 L510 also forms hydrogen bonds with the backbone carbonyl group of LGG-1 L50 and the side chain LGG-1 R28. The X_1 and X_2 residues within LIR1 also engage in interaction with LGG-1. The carboxyl group of CeEPG5 E508 hydrogen bonds with the guanidinium group of LGG-1 R67, while the backbone amide group and carbonyl group of CeEPG5 E508 form hydrogen bonds with carbonyl oxygen of LGG-1 K48 and the amide group of L50, respectively (Figure 5D). In contrast, the side chain of CeEPG5 I509 binds the side chain of LGG-1 Y25, R28, and L50 through hydrophobic interactions (Figure 5E). Interactions were also observed in the regions flanking the LIR core. Notably, the backbone of CeEPG5 S505 N-terminal to the core forms a hydrogen bond with the ε-amino group of LGG-1 K48 (Figure 5G) while the carbonyl groups of the C-terminal CeEPG5 A511 and the side chain of CeEPG5 D513 contact the guanidinium group of LGG-1 R28 (Figure 5H).

Next, we aligned our CeEPG5 LIR1-LGG-1 structure to the crystal structure of apo-LGG-1 (Table S2 and Fig. S4C). We found that the side chains of several LGG-1 residues undergo conformational rearrangement to accommodate CeEPG5 LIR1 peptide binding (Fig. S4D). The movement of E17, K48, and F104 increases the solvent-accessible surface area of HP1 from 5.5 Å^2 to 39.1 Å^2 to accommodate CeEPG5 W507 binding (Fig. S4E and F). Similarly, the movement of L63 increases the solvent-accessible surface area of HP2 from 0.2 Å^2 to 3.1 Å^2 for CeEPG5 L510 binding (Fig. S4E and I). A highly conserved residue amongst the Atg8 family proteins (K46 for Atg8, K49 for LC3A/B, K55 for LC3C, K46 for GABARAP/L1/L2, K46 for LGG-1) has been previously shown to mediate interaction with the central X_1 residue of the LIR motif [22,41,42]. Indeed, LGG-1 K46 shifts 7.3 Å outwards to allow CeEPG5 E508 to dock and interact with the sidechain of R67 through hydrogen bonding (Fig. S4D and G). Conversely, LGG-1 Y25 shifts toward the CeEPG5-LIR peptide and coordinates with R28 and L50 to form hydrophobic interactions with CeEPG5 I509 (Fig. S4H). Similar conformation changes were also observed in previously reported crystal structures LGG-1 in complex with the WEEL peptide and the UNC-51 LIR peptide (Fig. S4D) [40].

We then investigated how the CeEPG5 tandem motifs interact with LGG-1 in the full-length context. CeEPG5 is substantially shorter and the structural model available on the AlphaFold structural database shows that it only encompasses the analogous "hook" region observed in HsEPG5 [28] (Figure 6A). This structural model showed that LIR1 is fully accessible at the beginning of the loop region while LIR2 is located at the end of the loop region, adjacent to one of the alpha-helix that constitute the hook region may pose steric hindrance for LGG-1/2 binding (Figure 6B). When we carried out AlphaFold3 run on CeEPG5 together with LGG-1, the generated structural model, with the high pLDDT and PAE scores and an overall ipTM and pTM scores of 0.75 and 0.66, showed the tandem LIR motifs did not engage in multivalent interaction with LGG-1. Rather, LIR1 was selected over LIR2 in binding the LDS of its cognate Atg8 protein, where LIR1 exhibited an extended beta-strand conformation with CeEPG5 W507 and CeEPG5 L510 inserted into the HP1 and 2 of LGG-1, respectively (Figure 6C,D, and Fig. S5A and B). The similar mode of interaction between EPG5 and Atg8 in both humans and *C. elegans* suggests that the accessibility of the LIR motif is important for Atg8 binding.

Discussion

Molecular structure of HsEPG5 and implications on its tethering mechanism

While we were able to visualize HsEPG5 in its solution state, the modest resolution of our cryo-EM map precluded *de novo* model building. AlphaFold2 has facilitated the interpretation of our medium-resolution cryo-EM map and enabled the construction of the first structural model of HsEPG5. Interestingly, although both our cryo-EM map and the available model of full-length HsEPG5 in the AlphaFold Protein Structure Database feature three domains (hook, thumb, shaft), the orientations of these domains are quite different. In particular, the hook and shaft domains are positioned much closer to one another in the predicted model compared to the experimentally derived cryo-EM map. We speculate that an emphasis on searching for potential intramolecular interactions from co-evolution data in the AlphaFold2 algorithm contributed to the generation of a more "bent" artificial overall architecture. This issue was resolved by conducting focused predictions on individual HsEPG5 domains.

Our structural model revealed that HsEPG5, similar to other structurally characterized tethering factors in other intracellular transport pathways, uses alpha-helical bundles to construct extended structural features, most notably the shaft domain [26,43]. However, instead of having three domains arranged tandemly from the N-terminus to the C-terminus, the hook domain of HsEPG5 is actually "sandwiched" between the two halves of the thumb domain based on the primary sequence. Intriguingly, our structural model showed that the loop region containing the autophagosome-binding tandem LIR motifs is located near the center of HsEPG5. In contrary to other tethering factors, HsEPG5 does not appear to fully harness its extended length to link the autophagosome to the lysosome. Instead, HsEPG5 might either exert its tethering function at a shorter range or it might collaborate with other autophagy factors to fill the distance gap in bridging the autophagosome to the lysosome.

Mode of interaction between EPG5 tandem LIR motifs and Atg8-family proteins

The majority of LIR-containing proteins identified to date possess one main LIR that binds to the LDS of their cognate Atg8 proteins. Even for the few others that were found to
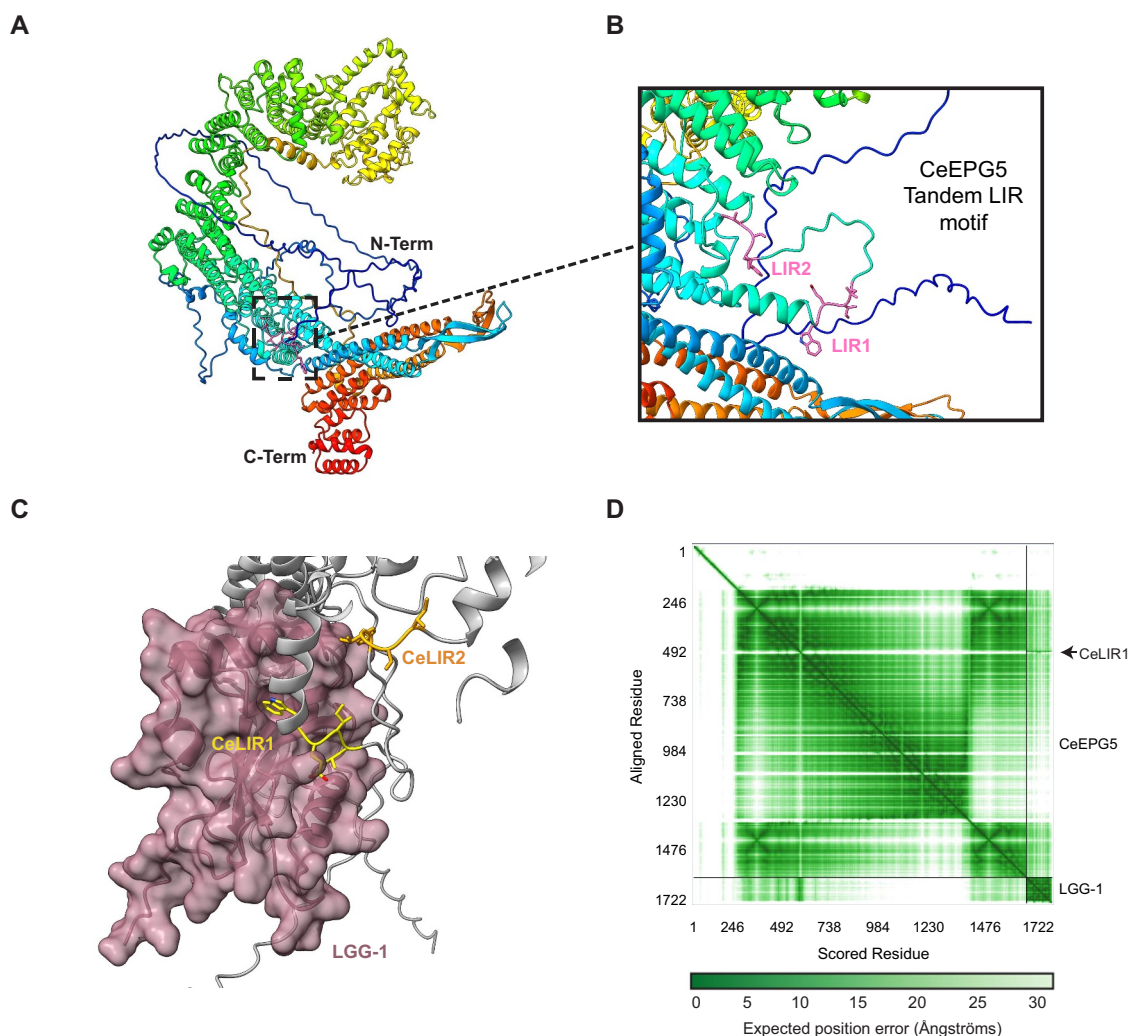
Figure 6. AlphaFold3 prediction model of CeEPG5 tandem LIR motifs binding to LGG-1. (A) AlphaFold3 predicted model of CeEPG5. Cartoon model is colored in a rainbow color scheme (blue to red) from the N terminus to the C terminus. The tandem LIR motifs (LIR1 and LIR2) are represented in pink sticks, located on the flexible loop. (B) Close-up view of the tandem LIR motifs on the flexible loop indicated by the dotted lined box in (A). (C) Close-up view of AlphaFold3 predicted structure of LGG-1 (ribbon and transparent surface representation in mauve) binding to CeEPG5 LIR1 (ribbon and sticks representation in yellow) of the tandem LIR motifs at the hook region of full-length CeEPG5 (ribbon representation in gray). CeEPG5 LIR2 (ribbon and sticks representation in orange) is not involved in LGG-1 interaction. (D) Predicted align error (PAE) plot of LGG-1 in complex with full-length CeEPG5.

contain multiple LIRs, these motifs are positioned distal to one another in sequence and most likely spatially in the context of the folded proteins [25,44]. The autophagy tethering factor EPG5 represents an atypical LIR-containing protein as it has two LIR motifs that are arranged directly adjacent to one another in tandem fashion. The only other protein known to contain a tandem LIR is RavZ, the *Legionella pneumophila* secreted effector protein that binds and cleaves LC3 to inhibit xenophagy [45]. Based on our previous finding that LIR2 of the HsEPG5 tandem motifs docks into the LDS of GABARAPL1, we initially postulated that the neighboring LIR1 would engage GABARAP in a non-canonical manner to enable multivalent interaction with this Atg8 isoform [22]. However, *in silico* AlphaFold-Multimer structural prediction, NMR spectroscopic analyses, and MD simulations showed that LIR1, when isolated, binds the LDS of GABARAPL1 in a canonical fashion and a peptide containing the tandem motifs (LIR1 + LIR2) does not appear to engage in multivalent

interaction with GABARAPL1. Further investigations using AlphaFold3 modeling in conjunction with co-immunoprecipitation in cells suggest that LIR1 represents the primary GABARAP-interacting motif of HsEPG5 while LIR2, which is predicted to be located in an inaccessible region, likely plays a structural role.

The tandem LIR motifs are conserved in all annotated EPG5 protein homologues from *C. elegans* to human but whether its biochemical properties and function are also preserved is not known. Previous studies showed that CeEPG5 binds LGG-1, one of the two Atg8 isoforms found in *C. elegans*, through its tandem LIR [19]. Our study demonstrated that CeEPG5 can interact with both LGG-1 and LGG-2, the GABARAP and LC3 orthologs, respectively. In human cells, the LC3 subfamily of Atg8 proteins plays a central role in phagophore expansion while GABARAP subfamily of Atg8 proteins is important for autophagosome formation and autophagosome-

lysosome fusion [46,47]. In contrast, the "GABARAP-like" LGG-1 acts in the earlier stage of aggrephagy in *C. elegans* by interacting with the UNC-51/EPG-1 complex to mediate autophagosome formation [40] whereas the "LC3-like" LGG-2 is essential for lysosomal recruitment and autophagosome-lysosome fusion [48,49]. Surprisingly, we did not observe preferential binding of CeEPG5 to LGG-2. However, the CeEPG5-LGG-2 interaction is more sensitive to mutations to the tandem LIR motifs than the cEPG-LGG-1 interaction.

Our Alphafold predictions, biochemical affinity isolation and site-specific mutagenesis studies have revealed key similarities between the CeEPG5 tandem LIR and its HsEPG5 counterpart: First, we found that CeEPG5 LIR1 is situated in a flexible loop region, making it potentially available for protein-protein interactions, while CeEPG5 LIR2 likely forms part of the hook backbone structure. Secondly, tryptophan is also present at position 0 of the CeEPG5 LIR1 motif, but it is phenylalanine in the same position as the CeEPG5 LIR2 motif. Thirdly, our findings indicate that CeEPG5 LIR1 is the sole motif responsible for binding LGG-1 and LGG-2. Data from previous functional studies on CeEPG5 also support our observations. In particular, the CeEPG5 LIR1 mutant was unable to revert the LGG-1 puncta and SQST-1 accumulation in *epg-5* mutant embryos, but the CeEPG5 LIR2 mutant can partially rescue CeEPG5's activity, confirming that LIR1 plays a more critical role in CeEPG5 function [20]. These findings collectively confirm that LIR1 serves as the primary Atg8-family protein binding motif and the conservation of tryptophan at position 0 in the interaction between *C. elegans* and humans.

In spite of these new findings, the role of LIR2, and in particular the LIR2 motif in HsEPG5, remains ambiguous. HsEPG5 LIR2 contains a tryptophan at position 0 instead of the phenylalanine observed in CeEPG5 LIR2. Studies on Atg8-binding proteins have demonstrated that LIR motifs with tryptophan at position 0 show higher binding affinity with Atg8 proteins compared to those with phenylalanine or tyrosine [31]. Furthermore, previous cell-based localization studies show that similar to wild-type HsEPG5, HsEPG5 LIR1 mutant could form small puncta that colocalize with lysosomal surface-residing Rab7 in HeLa cells. However, the HsEPG5 LIR2 and tandem LIR mutants could not form these punctae and they were found mainly diffused in the cytoplasm [19]. Although we have shown that HsEPG5 localization to the peri-nuclear region is independent of LC3/GABARAP proteins, these results imply that LIR2 is crucial to HsEPG5 function in cells, albeit in a GABARAP-independent fashion [22]. Future studies should focus on delineating the distinct roles of the two LIRs in mediating GABARAP-dependent and GABARAP-independent functions of HsEPG5.

Importance in characterizing the LIR motif in the context of the full-length protein

Over the past two decades, different biochemical, biophysical, and structural studies involving isolated peptides containing LIR sequences ranging from 10 to 30 residues have significantly advanced our understanding of the molecular bases of how LIR motif-bearing interactors engage their cognate Atg8 proteins [25,33–38]. However, there are potential caveats associated with the approach of examining the "minimal binding motif" of Atg8-interactors. First of all, the 4-residue LIR motif only requires specific residues at positions 0 and 3 and similar motifs are also found in proteins that do not interact with Atg8 proteins. Second, even for confirmed Atg8-interacting partners, the presence of an LIR motif in the sequence does not necessarily mean that the putative LIR mediates Atg8 binding. In particular, the putative LIR may not be located in an accessible region of the full-length protein. Our previous work showed that as an isolated peptide, HsEPG5 LIR2 binds GABARAP with higher affinity compared to HsEPG5 LIR1 [22]. This finding led us to determine the crystal structure of GABARAPL1 in complex with HsEPG5 LIR2. However, the work we presented here, and in particular our AlphaFold3 and co-immunoprecipitation experiments, demonstrated that HsEPG5 LIR2 is not directly involved in binding GABARAP but instead serves a structural role. Our present work underscores the importance of analyzing LIR-Atg8 interaction in the context of the full-length LIR-bearing proteins. Importantly, we should leverage structural modeling tools such as the machine learning-based AlphaFold2/3 algorithms to perform initial "screening" analysis to filter out inaccessible LIR motifs.

Materials and methods

Plasmid generation

HsEPG5 and CeEPG5 cDNA were cloned into a modified pFastBacHTB encoding a His-FLAG tag (gift from Dr. Ji-Joon Song, Korea Advanced Institute of Science and Technology, Korea), using SalI and NotI restriction enzyme sites. Site-directed mutagenesis of CeEPG5 were carried out using plasmid pFastBacHTB-FLAG-CeEPG5 with the following oligos: CeEPG5(W507A), forward 5'- aaagaggagtcaactgcggagattcttgcaga −3' and reverse 5'- tctgcaagaatctccgcagttgactcctctt −3'; HsEPG5(F523A), forward 5'- gaagcaaatgagttttcagctgtcacaatcaacgaatc −3' and reverse 5'- gattcgttgattgtgacagctgaaaactcatttgcttc −3'.

3×FLAG-HsEPG5, 3×FLAG-HsEPG5 W550A, 3×FLAG-HsEPG5 W567A, or 3×FLAG-HsEPG5 W550A W567A were cloned from the pcDNA3.1-3×FLAG-HsEPG5 wild-type or mutant constructs into PB-EF1a-IRES-Neo (gift from Dr. Sheila Teves, The University of British Columbia, Canada), using XhoI and NotI restriction enzyme sites. HALO tag constructs were generated using the plasmid HALO-SOX2 (gift from Dr. Sheila Teves) and the following oligos: forward 5'-gatgacgatgacaagccaagcttggtaccgatggcagaaatcggtactggctttccattc-3' and reverse 5'-tggctgtcgacgctcgagatttccggcagctcggatccgaattcgagctccgtcg-3'.

cDNA of human *GABARAPL1* was cloned into pGEX6P–1 (Cytiva) using BamHI and NotI restriction enzyme sites. HsEPG5[541–581] LIR fragment and NBR1 LIR fragment (GAMG-NBR1[726–738]-S) were cloned into a modified pGEX6P–1 encoding a tobacco etch virus (TEV)-cleavable

linker. The pGEX6P–1-TEV construct was generated using plasmid pGEX6P–1 with the following oligos: forward 5'-ggggagaatctttattttcagggatccccggaattcccggg –3' and reverse 5'-ctgaaaataaagattctcccccaggggcccctggaacagaa –3'. HsEPG5 LIR fragment was cloned from pFastBacHTB-FLAG-HsEPG5 into pGEX6P–1-TEV using BamHI and NotI restriction enzyme site with the following oligos: forward 5'- tataggatcctcctcctcagggcctgggtctgg –3' and reverse 5'- tatagcggccgcc tactgtgctaaaatggtaaccaaatcatcttcattaaggagaatcc –3'. The NBR1 LIR fragment was cloned into pGEX6P–1-TEV in a two-step procedure, first using pGEX6P–1-TEV and the following oligos: forward 5'- ggcgccatgggctctgcttcctcagaggatggatccccg-gaattcccggg –3' and reverse 5'- atcctctgaggaagcagagcccatgg cgccctgaaaataaagattctcccccaggggg –3', followed by the second set of oligos: forward 5'- tacatcatcatcctgcctgagtcatagggatccccg-gaattcccggg –3' and reverse 5'- ctatgactcaggcaggatgatgatg-taatcctctgaggaagcagagcccat –3'.

cDNA of *C. elegans* LGG-1 and LGG-2 proteins were cloned into pQLinkG2, pQLinkH, and pGEX6P–1-TEV, using BamHI and NotI restriction enzyme sites. An LGG-2 N-terminally truncated mutant was generated using plasmid pQLinkG2-LGG-2 with the following oligos: LGG-2[16–130], forward 5'- cgagaatctttattttcagggatccagctttaaagaacgtcgtccgttc –3' and reverse 5'- gaaacggacgacgttctttaaagctggatccctgaaaa-taaagattctcg –3'. pQLinkG2, pQLinkH, and pGEX6P–1-TEV constructs were expressed for GST affinity isolation analysis, ITC analysis, and X-ray crystallography, respectively.

Cell culture

The human embryonic kidney 293 (HEK293; gift from Dr. Sharon Gorski, BC Cancer Research Centre, Canada) cell line was cultured in Dulbecco's Modified Eagle Medium (DMEM) high glucose with L-glu Na-pyruvate (Sigma-Aldrich, D6429) supplemented with 8% (v:v) fetal bovine serum (FBS; Sigma-Aldrich, F1051).

For cell lysis, cells were resuspended and incubated with 500 μL of ice-cold buffer A (50 mm Tris-HCl, pH 7.4, 100 mm NaCl, 1 mm EDTA, 5% glycerol, 0.01% Triton X-100 (Sigma-Aldrich, T8787), 1 mm phenylmethylsulfonyl fluoride [PMSF; BioShop Canada, PMS123], and cOmplete ethylenediaminete-traacetic acid [EDTA]-free protease inhibitor (Roche 5056489001) on ice for 30 min. The cells were lysed by passing cell lysate through a 26-gauge needle attached to a 1 mL syringe 30 times, followed by centrifugation at $12,000 \times g$ at 4°C for 30 min. The total protein concentration of the cell lysate was determined by BCA assay (Thermo Fisher Scientific 23227).

Stable cell generation

HEK293 cells were seeded at 0.25×10^6 cells/well in 6-well plates 24 h prior to co-transfection with 1 μg of construct (PB-EF1a-3×FLAG-HALO, PB-EF1a-3×FLAG-HALO-HsEPG5, PB-EF1a-3×FLAG-HALO-HsEPG5 W550A, PB-EF1a-3×FLAG-HsEPG5 W567A, or PB-EF1a-3×FLAG-HALO-HsEPG5 W550A/W567A) and 1 μg of the Super Piggybac transposase plasmid (gift from Dr. Sheila Teves) using Lipofectamine 3000 (Invitrogen, L3000008) according to the

manufacturer's protocol. After 24 h, the transfection reagents were removed and media containing 500 μg/mL of G418 (Gibco 10131035) was added for selection. Antibiotic media was refreshed every 48 h until all negative control cells were dead.

Transfection

For transient transfection, HEK293 cells stably expressing 3×FLAG-HALO, 3×FLAG-HALO-HsEPG5, 3×FLAG-HALO-HsEPG5 W550A, 3×FLAG-HsEPG5 W567A, or 3×FLAG-HALO-HsEPG5 W550A/W567A were seeded at 0.5×10^6 cells per well in 6-well plates 24 h before transfection. Transfection was performed using PEI Max (Transfection grade linear polyethylenimine hydrochloride MW 40,000; Polyscience 24765–100). 2.5 μg of pcDNA3.1-eGFP DNA per well, or 1.5 μg of pcDNA3.1-eGFP-GABARAP DNA per well, were diluted with 250 μL of DMEM without FBS. PEI was then added to the diluted DNA at a PEI:DNA (w:w) ratio of 3:1, followed by a 15-minute incubation at room temperature before the addition to the cell. After 24 h of transfection, cells were harvested, frozen, and stored at – 70°C.

Protein expression and purification

Wild-type His-FLAG-HsEPG5, wild-type His-FLAG-CeEPG5, and His-FLAG-CeEPG5 mutant constructs were used for generating baculovirus in Sf9 insect cells using the Baculovirus Expression Vector System. The generated baculovirus was then used for large-scale protein expression by infecting a 200 mL culture of Sf9 cells at a density of $1.4–4 \times 10^6$ cells/mL. Cells were harvested after 72 h of expression for purification or stored at – 70°C until use.

Purification of HsEPG5

For HsEPG5 purification, Sf9 cell pellets expressing wild-type His-FLAG-HsEPG5 were resuspended in buffer B (50 mm NaPhosphate, pH 7.4, 150 mm NaCl, 0.1% CHAPS [BioShop Canada, CHA003], 1 mm PMSF, cOmplete EDTA-free protease inhibitor) prior to lysis. The cells were lysed by sonication using Branson Sonicator 450 for five cycles consisting of 30 s sonication followed by 90 s cooling on ice, with duty cycle set to 30% and output control at 3. The cell lysate was centrifuged at $110,200 \times g$ at 4°C for 40 mins. The supernatant was then incubated with an Anti-FLAG M2 affinity gel (Sigma-Aldrich, A2220) pre-equilibrated with buffer B without PMSF and protease inhibitor at 4°C for 1 h, while gently inverting. The resin was washed with buffer C (50 mm NaPhosphate, pH 7.4, 150 mm NaCl) with a decreasing amount of CHAPS, from 0.1% to 0.05% to 0.01%, to remove non-bound species. Proteins were eluted with 3× FLAG peptide (GenScript) in buffer D (50 mm NaPhosphate, pH 7.4, 150 mm NaCl, 0.01% CHAPS), followed by PEGylation using methyl-PEG4-N-hydroxysuccinimide (NHS) ester (TCI America, M2186), concentration and size-exclusion chroma-tography purification on a SEC650 column (Bio-Rad 7801650), equilibrated and run in buffer E (50 mm Tris, pH 7.4, 150 mm NaCl, 0.01% CHAPS, 0.5 mm tris

(2-carboxyethyl)phosphine [TCEP; BioShop Canada, TCE101]). Proteins were subsequently stored at − 70°C.

Purification of CeEPG5

For CeEPG5 purification, Sf9 cell pellets expressing wild-type or mutant His-FLAG-CeEPG5 were resuspended in buffer F (50 mm Tris, pH 8.0, 150 mm NaCl, 0.1% Tween 20 [Fisher Scientific, BP337100], 1 mm PMSF, and cOmplete EDTA-free protease inhibitor) prior to lysis. The cells were lysed by sonication using a Branson Sonicator 450 for four cycles consisting of 30 s sonication followed by 90 s cooling on ice, with duty cycle set to 30% and output control at 3. The cell lysate was centrifuged at $110,200 \times g$ at 4°C for 40 min. The supernatant was then incubated with an Anti-FLAG M2 affinity gel pre-equilibrated with buffer F at 4°C for 1 h, while gently inverting. The resin was washed with buffer G (50 mm Tris, pH 8.0, 150 mm NaCl) with decreasing amount of Tween 20, from 0.1% to 0.05% to 0.01%, and a final wash with buffer G without Tween 20 to remove non-bound species. Proteins were eluted with 3× FLAG peptide in buffer G, subsequently aliquoted and stored at − 70°C. Protein concentration was measured by spectrophotometry.

Purification of GABARAPL1 for NMR spectroscopic analysis

^{15}N-GABARAPL1 was expressed as an N-terminally GST-tagged construct in *Escherichia coli* T7 Express cells (New England Biolabs, C3010I). Cells were grown in minimal M9 media supplemented with 1 g/L ^{15}NH$_4$Cl (Cambridge Isotope Laboratories, Inc., NLM-467), and induced with 1 mm isopropyl β-d-1-thiogalacpyranoside (IPTG; Gold Bio, I2481C50) at 30°C for 4 h. Prior to lysis, cell pellets were resuspended in buffer H (50 mm Tris, pH 8.0, 150 mm NaCl, 2 mm PMSF, 0.1% Triton X-100). Cells were lysed by sonication and the lysate was centrifuged at $20,950 \times g$ at 4°C for 40 min. The supernatant was then incubated with glutathione resin (GenScript, L00206) pre-equilibrated with buffer H at 4°C for 1 h, followed by six washes with buffer G to remove non-bound species. GST-tagged ^{15}N-GABARAPL1 was then cleaved with PreScission protease in buffer G supplemented with 1 mm dithiothreitol [DTT; BioShop Canada, DTT001] and 1 mm EDTA pH 8.0 for 2 h at room temperature. Subsequently, flow through and three washes were collected for purification using reverse glutathione resin chromatography. Target protein was then concentrated and further purified through size-exclusion chromatography using a Sephacryl S-200 column (GE Healthcare 17116601) equilibrated and run in buffer I (50 mm Tris, pH 7.0, 150 mm NaCl). The purified target protein was dialyzed overnight at 4°C and stored in buffer J (50 mm NaPhosphate, pH 6.5, 100 mm NaCl). Protein concentration was measured by spectrophotometry.

Both the HsEPG5 tandem and NBR1 LIR fragments (HsEPG5 541–581 [SSSGPGSGTWTLVDEGGEEDEDPET SWILLNEDDLVTILAQ] and GAMG-NBR1[726–738]-S [GAMG-SASSEDYIIILPE-S], respectively) were expressed as N-terminally GST-tagged constructs containing a TEV-cleavable linker. HsEPG5 tandem LIR and NBR1 LIR fragments were expressed in *E. coli* T7 Express grown in 2×YT media with 1 mm IPTG at 37°C and 30°C for 4 h, respectively. Cells were resuspended, sonicated, and centrifuged, as described

above, using buffer K (50 mm Tris, pH 8.0, 300 mm NaCl, 2 mm PMSF, 0.1% Triton X-100). The supernatant was then incubated with glutathione resin pre-equilibrated with buffer K at 4°C for 1 h. The resin was washed six times with buffer L (50 mm Tris, pH 8.0, 300 mm NaCl) and subsequently eluted with 30 mm reduced glutathione (BioShop Canada, GTH001) in buffer L. The eluted GST-TEV-target proteins were individually concentrated and cleaved with TEV protease in buffer M (50 mm Tris, pH 8.0, 500 mm NaCl, 1 mm DTT, 1 mm EDTA pH 8.0, 0.01% Triton X-100) and dialyzed in buffer N (50 mm Tris, pH 8.0, 300 mm NaCl, 0.5 mm DTT, 0.01% Triton X-100) at 4°C overnight. As a final clean-up step. The targets were collected and further purified by reverse glutathione resin chromatography, followed by concentration and size-exclusion chromatography purification on a Sephacryl S-200 column (GE Healthcare 17116601), equilibrated and run in buffer I. Purified target proteins were dialyzed at 4°C overnight and stored in buffer J. Protein concentration was measured by spectrophotometry.

Purification of LGG-1 and LGG-2

For GST affinity isolation analysis, LGG-1 and LGG-2[16–130] were expressed as N-terminally GST-tagged constructs in *E. coli* (T7 Express) cells. Cells expressing GST-LGG-1 were grown in 2×YT media and induced with 1 mm IPTG (Gold Bio, I2481C50) at 25°C overnight, while cells expressing GST-LGG-2[16–130] were induced with 1 mm IPTG at 37°C for 4 h. Before lysis, cell pellets were resuspended in buffer O (50 mm Tris, pH 8.0, 150 mm NaCl, 2 mm PMSF). Cells were lysed by sonication for four cycles consisting of 1 min sonication followed by 2 min cooling on ice, with duty cycle set to 50% and output control at 5. The cell lysate was centrifuged at $20,950 \times g$ at 4°C for 40 mins. The supernatant was then incubated with glutathione resin pre-equilibrated with buffer G at 4°C for 1 h. The resin was washed with buffer G and subsequently eluted with 30 mm reduced glutathione in buffer G. Free glutathione was removed by dialysis in buffer G and samples were aliquoted and stored at − 70°C. Protein concentration was measured by spectrophotometry.

For ITC studies, LGG-1 was expressed as N-terminally His-tagged constructs in *E. coli* T7 Express cells. Cells expressing His-LGG-1 were grown in 2×YT media and induced with 1 mm IPTG at 25°C for 4 hr. Prior to lysis, cell pellets were resuspended, sonicated, and centrifuged as described above with buffer O. The supernatant was then incubated with HisPur^{TM}Ni-NTA resin (Thermo Scientific 88222) pre-equilibrated with buffer O at 4°C for 1 h. The resin was washed with buffer G five times and subsequently with buffer G containing 30 mm imidazole ten times. The bound proteins were eluted with buffer G containing 100 mm imidazole once, then 300 mm imidazole four times, and 500 mm imidazole once. Eluted proteins were loaded onto a HiPrepQ FF 16/10 column (GE Healthcare 28936543) in buffer P (50 mm Tris, pH 8.0) and 1% of buffer Q (50 mm Tris, pH 8.0, 2 M NaCl). Target proteins in flow-through were collected, concentrated, and further purified by size exclusion chromatography using Sephacryl S-200 column (GE Healthcare 17116601) with buffer G. Purified proteins were aliquoted and stored at −

70°C. Protein concentration was measured by spectrophotometry.

For crystallography, N-terminally GST-tagged LGG-1 containing a TEV-cleavable linker was expressed in *E. coli* T7 Express cells with 1 mm IPTG induction at 25°C overnight. Cells were resuspended, sonicated, and centrifuged, as described above, using buffer K. The supernatant was then incubated with glutathione resin pre-equilibrated with buffer K at 4°C for 1 h. The resin was washed six times with buffer L and subsequently eluted with 30 mm reduced glutathione in buffer L. The eluted GST-TEV-LGG-1 was concentrated and cleaved with TEV protease in buffer M and dialyzed in buffer N at 4°C overnight. As a final clean-up step, the target protein was collected and further purified by reverse glutathione resin chromatography, followed by concentration and size-exclusion chromatography purification on a Sephacryl S-200 column (GE Healthcare 17116601), equilibrated and run in buffer G. Purified proteins were aliquoted and stored at − 70°C. Protein concentration was measured by spectrophotometry.

AlphaFold structural modeling

AlphaFold2 or AlphaFold-Multimer using MMseqs2 of ColabFold notebook, or AlphaFold3 using the AlphaFold server, were utilized for making a series of structural predictions: Full-length HsEPG5 and CeEPG5 from AlphaFold Protein Structure Database; truncated HsEPG5[534–1519], HsEPG5[288–533,1524–1709], and HsEPG5[1710–2579] from AlphaFold2 predictions; GABARAPL1 in complex with HsEPG5[546–556], HsEPG5[560–571], HsEPG5[546–571] and HsEPG5[541–581] from AlphaFold-Multimer predictions; GABARAPL1 in complex with full-length HsEPG5 and truncated HsEPG5[425–1519], and LGG-1 in complex with full-length CeEPG5 from AlphaFold3 predictions. The predicted local distance difference test (pLDDT) confidence values varied across all models, with the pLDDT and predicted aligned error (PAE) scores shown in Fig. S3, S5 and S6.

Cryo-EM sample preparation and data collection

Lacey Carbon grids (Ted Pella 01895-F) were overlaid by a thin graphene oxide (GO) (Sigma-Aldrich 763705) layer as described previously [50]. Prior to vitrification, the GO-grids were glow discharged at 15 mA for 5 s using a Pelco easiGlow glow-discharger (Ted Pella). The GO-grids were then applied with aliquots of 3 uL of purified PEGylated HsEPG5, followed by 1.5 s of blotting with a blot force of −5 and plunge freezing into liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific), operated at a temperature of 4°C with 100% humidity. Automated data collection was performed at the UBC High Resolution Macromolecular Cryo-Electron Microscopy (HRMEM) facility, using a 300 kV FEI Titan Krios transmission electron microscope (Thermo Fisher Scientific) equipped with a Falcon IV direct electron detector. A total of 3042 movies were collected at a pixel size of 1.20 Å and magnification of 105,000 × with a dose rate of 50 electrons/Å^2 over 294 frames.

Cryo-EM image processing

The cryo-EM data was processed using cryoSPARC v.3.3.2 [51]. The movies initially underwent patch motion correction with default settings and Fourier cropping by a factor of 2 for generating the micrographs, followed by patch CTF estimation with default settings. A subset of micrographs was selected for particle picking to generate 2D class averages, which were used as templates for autopicking 875,590 particles from the entire dataset. The particles were then extracted with a box size of 400 pixels and Fourier cropped to a box size of 200 pixels. Multiple rounds of 2D classification were performed to remove poorly defined particles, selecting 180,311 higher-quality particles for a four-class *ab initio* reconstruction and four-class heterogenous refinement. Finally, a homogenous refinement generated a reconstruction with an overall resolution of 7.85 Å, based on the Fourier shell correlation 0.143 criterion. Local resolutions were estimated using the local resolution estimation tool in cryoSPARC. An atomic model of the truncated HsEPG5 was generated using ColaFold notebook to generate the hook (HsEPG5[534–1519]), thumb (HsEPG5[288–533,1524–1709]), and shaft (HsEPG5[1710–2579]) regions, where each segment was fitted into the map using ChimeraX [52]. The structure was built with iterative rounds of refinement using Phenix real space refine [53].

NMR-monitored titrations

All NMR experiments were performed on a cryoprobe-equipped Bruker Avance III 600 MHz NMR spectrometer at 25°C. The HsEPG5 LIR1 peptide ($_{546}$GSGTWTLVDEG$_{556}$) was synthesized by LifeTein whereas the HsEPG5 LIR2 peptide ($_{560}$DEDPETSWILLN$_{571}$) was synthesized by GenScript. Synthesized peptides were dissolved in buffer J. All unlabeled peptides (HsEPG5 LIR1, HsEPG5 LIR2, HsEPG5[541–581] tandem LIR, and NBR1 LIR) were used at a concentration ranging from ~ 9 to 26 mm, determined via UV spectroscopy and manufacturer supplied extinction coefficients. Peptides were individually titrated into ^{15}N-GABARAPL1, initially at ~0.18 mm, in buffer J supplemented with 5% lock D$_2$O. ^{15}N-GABARAPL1 was titrated with (1) HsEPG5 LIR1 and (2) NBR1-LIR separately at protein:peptide ratios of 1:0, 1:0.25, 1:0.5, 1:0.75, 1:1, 1:1.5 and 1:2.5; (3) HsEPG5 LIR2 at ratios of 1:0, 1:0.1, 1:0.25, 1:0.5, 1:0.75, 1:1, and 1:1.5; and (4) HsEPG5 tandem LIR at 1:0, 1:0.25, 1:0.5, 1:0.75, 1:1, 1:1.5 and 1:2.2. ^{15}N-HSQC spectra were recorded at each titration point and processed using NMRPipe [54] and NMRFAM-Sparky [55]. Amide ^{1}H-^{15}N chemical shift assignments were obtained by matching the observed peaks to the previous assigned spectra of GABARAPL1-NBR1-LIR [31]. The combined amide ^{1}H-^{15}N chemical shift perturbations ($\Delta\delta_i$) for each titration point were calculated according to Equation (1) [56].

$$\Delta\delta_i = \sqrt{(\Delta\delta_H)^2 + (0.14 \cdot \Delta\delta_N)^2} \qquad (1)$$

Here $\Delta\delta_N$ and $\Delta\delta_H$ denote the changes in ^{15}N and ^1H chemical shifts in parts per million (ppm), respectively, relative to the initial peptide-free protein sample.

Equilibrium dissociation (K_D) values were determined by fitting the titration data to a two-state binding model (2) [56] using MATLAB.

$$\Delta\delta_i = \Delta\delta_{sat} \frac{\left(2[P]_{T,i} + [L]_{T,i} + K_D\right) - \sqrt{\left(2[P]_{T,i} + [L]_{T,i} + K_D\right)^2 - 4[P]_{T,I}[L]_{T,i}}}{2[P]_{T,i}}$$

(2)

Here the change in chemical shift of any residue $\Delta\delta_i$ is dependent on $[P]_{T,i}$ and $[L]_{T,I}$, the total concentrations of labeled protein and peptide adjusted for dilution effects, respectively, at each titration point i. $\Delta\delta_{sat}$ corresponds to the extrapolated CSP at saturation.

Molecular dynamics simulation

Molecular Dynamics (MD) simulation of the HsEPG5 LIR1 (HsEPG5[546–556]), LIR2 (HsEPG5[560–571]), and tandem LIR (HsEPG5[546–571]) peptide bound to GABARAPL1 were performed using the AlphaFold-Multimer predicted structures. A standard MD simulation protocol was followed [57]. Briefly, protonation states and the effective pKa for all titratable residues were computed and determined using the Web application ProteinPrepare [58]. MD simulations were conducted using the modified AMBER ff19SB force field [59]. The simulations started with 5000 steps of steepest descent energy minimization, followed by the reassignments of velocities from the Maxwell distribution at 300 K every 1 ps for 5000 ps, and a final equilibration of the system for 20.0 ns. Subsequently, the MD trajectories were integrated using the GPU-accelerated pmemd engine in Amber 22 [60]. The simulations were carried out using the isobaric-isothermal ensemble [61] at 300 K and 1 atmosphere and using long-range non-bonded interactions with a 12 Å residue-based cutoff. Long-range electrostatic forces were calculated using the particle-mesh Ewald sum [62]. Additionally, bonds to hydrogen atoms were maintained with the SHAKE algorithm [63], and an integration step size of two femtoseconds was used. Overall, a total of 7.5 microseconds of MD simulations for the HsEPG5 LIR1, LIR2, and tandem LIR peptide bound to GABARAPL1 were obtained.

Isothermal titration calorimetry (ITC)

Purified His-LGG-1 was dialyzed against buffer R (50 mm Tris, pH 8.0, 25 mm NaCl) using 6–8 kDa molecular mass cutoff membrane (Spectrum, 08-670-3). CeEPG5 LIR1 peptide ($_{505}$STWEILADDDD$_{515}$) and LIR2 peptide ($_{516}$GEANEFSFVTINE$_{528}$) were synthesized by GenScript. Synthetic peptides were dissolved in buffer R. His-LGG-1 and CeEPG5 LIR1 peptide concentration was measured by spectrophotometry (NanoDrop, Thermo Scientific). CeEPG5 LIR2 peptide does not contain any tryptophan and tyrosine and hence the concentration was measured by BCA using a standard curve with L-PHE dissolved in buffer R. All ITC experiments were performed using ITC200 (GE Healthcare now Malvern Panalytical). For the ITC experiments, 0.6 mm of CeEPG5 LIR1 or LIR2 peptide was titrated into 0.06 mm of His-LGG-1 at 25°C in 20 steps. The ITC data were analyzed with a one-site binding model using Origin 7.0 (OriginLab) after a point-to-point subtraction of the corresponding reference titration of 0.6 mm of peptides into buffer R. Of note, the measured CeEPG5 LIR2 concentration was lower than the expected concentration, hence the data analysis was adjusted to 0.4 mm of CeEPG5 LIR2 with 0.06 mm of His-LGG-1 for data fitting. Experiments were performed in triplicates. Refer to Table S1 for thermodynamic parameters of the ITC experiments of CeEPG5 LIR1 and LIR2 peptide with His-LGG-1, respectively.

Crystallization and data processing

CeEPG5 LIR1 and LIR2 peptides were dissolved in buffer G. LGG-1 (8.9 mg/mL) was incubated with dissolved CeEPG5 LIR1 (4.0 mg/mL) or LIR2 peptide (6.7 mg/mL) at 4°C for 1 h prior to all crystallization trials. LGG-1/CeEPG5 LIR1 co-crystals grew in a condition containing 0.1 M sodium acetate, pH 4.6, 21% (w:v) PEG6000 (BioShop Canada, PEG600) in a 1:3 protein-to-liquid ratio. LGG-1 CeEPG5 LIR2 co-crystals grew in a condition containing 0.02 M zinc chloride, 0.1 M MES, pH 6.0, and 14% (w:v) PEG6000 in a 1:3 protein-to-liquid ratio. Crystals were harvested, cryo-protected by the crystal growing condition supplemented with up to 30% PEG6000, and frozen in liquid nitrogen prior to data collection.

X-ray diffraction data for both crystals were collected on Beamline 5.0.2 at the Advanced Light Source/ALS and processed using DIALS [64] and Aimless [65]. Molecular replacement was performed using PHASER [66] with the search model 5AZF (chain A only) [40] for solving the crystal structures. A combination of PHENIX [67], CCP4 [65] and Coot [68] was used for performing model building and refinement. It is noteworthy that the electron density for the CeEPG5 LIR2 peptide in the LGG-1 CeEPG5 LIR2 co-crystals was absent; therefore, the crystal structure of apo-LGG-1 was solved instead. CASTp [69] was used for hydrophobic pocket 1 and 2 pocket surface area and volume measurement with a radius probe of 1.4 Å (default setting). Refer to Table S2 for data collection and refinement statistics, Fig. S4A and B for the structural figure of CeEPG5 LIR1 in complex with LGG-1, including the probability ellipsoids.

Affinity isolation assay

Affinity isolations were performed with a 50 μL slurry of glutathione resin (50% slurry; GenScript, L00206) pre-equilibrated with buffer S (50 mm Tris, pH 8.0, 150 mm NaCl, 5% glycerol, 0.01% Tween 20, 0.5 mm DTT). The resins were incubated with equal amounts (240 μg) of baits (purified GST, GST-LGG-1, and GST-LGG-2[16–130]) in 1 mL of buffer S at 4°C for 15 min with gentle inversion. The tubes were centrifuged at 500 × g for 2 min and the supernatant containing excess bait protein were removed. The resins were then incubated with equal amounts (30 μg) of prey (purified His-FLAG-CeEPG5 or its corresponding mutants: W507A, F523A, or W507A F523A) in 50 μL of buffer S at 4°C for 1 h. The tubes were once again centrifuged at 500 × g for 2 min

and the supernatant were removed. The resins were washed with 1 mL buffer S, followed by centrifugation at $500 \times g$ for 2 min and removal of the excess buffer for five times. 25 µL of 2× SDS loading dye was added to each tube after the five times washes and the tubes were heated at 65°C for 10 min. The SDS samples were then loaded onto two 6–15% gradient SDS – PAGE gels and stained with PageBlue Protein Staining Solution (Thermo Scientific, 24620). For western blots, samples were transferred from a 6–15% gradient SDS – PAGE gel to a nitrocellulose membrane, followed by incubation with blocking buffer containing 5% skim milk in PBS-T at room temperature for 1 h. Mouse anti-FLAG was used as the primary antibody (1:2000; Sigma-Aldrich, F1804) and donkey anti-mouse IRDye 680LT as the secondary antibody (1:10000; LI-COR, 926–68022) for visualization on a ChemiDoc Imager (Bio-Rad). Experiments were performed in triplicates.

Co-immunoprecipitation

Co-immunoprecipitation was performed with 20 µL of GFP-TRAP agarose beads (ChromoTek, gta-20), each pre-equilibrated with buffer A. The resins were incubated with an equal amount of total protein in cell lysate in 400 µL of buffer A at 4°C for 2 h with gentle inversion. The tubes were centrifuged at $1000 \times g$ for 1 min, and the supernatant containing excess protein were removed. The resins were then washed with 500 µL of buffer T (50 mm Tris-HCl, pH 7.4, 100 mm NaCl, 1 mm EDTA, 5% glycerol, 0.01% Triton X-100), followed by centrifugation at $1000 \times g$ for 1 min and removal of the excess buffer for three times. SDS loading dye (400 µL of 2×) was added to each tube after the three washes and the tubes were heated at 95°C for 10 min. The SDS samples, along with the input controls, were subjected to western blot analysis. Samples were loaded onto a 6–15% gradient SDS – PAGE gels and transferred to a nitrocellulose membrane, followed by incubation with blocking buffer containing 6% skim milk in PBS at room temperature for 1 h. Mouse anti-FLAG (1:500; Sigma-Aldrich, F1804) and Rabbit anti-GFP (1:2000; Proteintech 50430–2-AP) were used as the primary antibodies, and donkey anti-mouse IRDye 680LT (1:10000; LI-COR, 926–68022) and donkey anti-rabbit IRDye 800CW (1:10000; LI-COR, 926–32213) as the secondary antibodies, respectively, for visualization on a ChemiDoc Imager (Bio-Rad). Experiments were performed in triplicates.

Quantification and statistical analysis

Statistics and graphs for NMR titration experiments were performed and generated using Prism 9 (GraphPad). NMR derived-affinity and graphs were calculated and generated using MATLAB R2020a (MathWorks). Co-immunoprecipitation assays, and CeEPG5-related affinity isolation assays and ITC experiments were performed in triplicates. ITC data represents means or means ± SEM. Intensity of wild type and mutant His-FLAG-CeEPG5, GST (control), and GST-LGG-1 -LGG-2[16–130] in affinity isolation assays was quantified using Bio-Rad Image Lab Software v6.0, with Prism 9 (GraphPad) used for statistics.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The 3D reconstruction of HsEPG5 is available at the Electron Microscopy Data Bank under accession code EMD-44835. The atomic coordinates for the LGG-1-CeEPG5-LIR1 complex and Apo-LGG-1 are available at the Protein Data Bank under accession codes 8TGF and 8TGX, respectively. All other data are available from the corresponding author upon reasonable request.

ORCID

Calvin K. Yip http://orcid.org/0000-0003-1779-9501

References

[1] Galluzzi L, Pietrocola F, Levine B, et al. Metabolic control of autophagy. Cell. 2014;159(6):1263–1276. doi: 10.1016/j.cell.2014.11.006
[2] Mancias JD, Kimmelman AC. Mechanisms of selective autophagy in normal physiology and cancer. J Mol Biol. 2016;428(9 Pt A):1659–1680. doi: 10.1016/j.jmb.2016.02.027
[3] Sánchez-Martín P, Komatsu M. Physiological stress response by selective autophagy. J Mol Biol. 2020;432(1):53–62. doi: 10.1016/j.jmb.2019.06.013
[4] Dikic I, Elazar Z. Mechanism and medical implications of mammalian autophagy. Nat Rev Mol Cell Biol. 2018;19(6):349–364. doi: 10.1038/s41580-018-0003-4

[5] Levine B, Kroemer G. Biological functions of autophagy genes: a disease perspective. Cell. 2019;176(1–2):11–42. doi: 10.1016/j.cell.2018.09.048

[6] Reggiori F, Ungermann C. Autophagosome maturation and fusion. J Mol Biol. 2017;429(4):486–496. doi: 10.1016/j.jmb.2017.01.002

[7] Nakatogawa H. Mechanisms governing autophagosome biogenesis. Nat Rev Mol Cell Biol. 2020;21(8):439–458. doi: 10.1038/s41580-020-0241-0

[8] Zhao YG, Codogno P, Zhang H. Machinery, regulation and pathophysiological implications of autophagosome maturation. Nat Rev Mol Cell Biol. 2021;22(11):733–750. doi: 10.1038/s41580-021-00392-4

[9] Klionsky DJ, Abdel-Aziz AK, Abdelfatah S, et al. Guidelines for the use and interpretation of assays for monitoring autophagy (4th edition)1. Autophagy. 2021;17(1):1–382. doi: 10.1080/15548627.2020.1797280

[10] Tian Y, Li Z, Hu W, et al. C. elegans screen identifies autophagy genes specific to multicellular organisms. Cell. 2010;141 (6):1042–1055. doi: 10.1016/j.cell.2010.04.034

[11] Feng Y, He D, Yao Z, et al. The machinery of macroautophagy. Cell Res. 2014;24(1):24–41. doi: 10.1038/cr.2013.168

[12] Suzuki H, Osawa T, Fujioka Y, et al. Structural biology of the core autophagy machinery. Curr Opin Struct Biol. 2017;43:10–17. doi: 10.1016/j.sbi.2016.09.010

[13] Weidberg H, Shpilka T, Shvets E, et al. LC3 and GATE-16 N termini mediate membrane fusion processes required for autophagosome biogenesis. Dev Cell. 2011;20(4):444–454. doi: 10.1016/j.devcel.2011.02.006

[14] Itakura E, Kishi-Itakura C, Mizushima N. The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. Cell. 2012;151(6):1256–1269. doi: 10.1016/j.cell.2012.11.001

[15] Matsui T, Jiang P, Nakano S, et al. Autophagosomal YKT6 is required for fusion with lysosomes independently of syntaxin 17. J Cell Biol. 2018;217(8):2633–2645. doi: 10.1083/jcb.201712058

[16] Jiang P, Nishimura T, Sakamaki Y, et al. The HOPS complex mediates autophagosome–lysosome fusion through interaction with syntaxin 17. Mol Biol Cell. 2014;25(8):1327–1337. doi: 10.1091/mbc.e13-08-0447

[17] Diao J, Liu R, Rong Y, et al. ATG14 promotes membrane tethering and fusion of autophagosomes to endolysosomes. Nature. 2015;520(7548):563–566. doi: 10.1038/nature14147

[18] McEwan DG, Popovic D, Gubas A, et al. PLEKHM1 regulates autophagosome-lysosome fusion through HOPS complex and LC3/GABARAP proteins. Mol Cell. 2015;57(1):39–54. doi: 10.1016/j.molcel.2014.11.006

[19] Wang Z, Miao G, Xue X, et al. The vici syndrome protein EPG5 is a Rab7 effector that determines the fusion specificity of autophagosomes with late Endosomes/Lysosomes. Mol Cell. 2016;63 (5):781–795. doi: 10.1016/j.molcel.2016.08.021

[20] Chen D, Fan W, Lu Y, et al. A mammalian autophagosome maturation mechanism mediated by TECPR1 and the Atg12-Atg5 conjugate. Mol Cell. 2012;45(5):629–641. doi: 10.1016/j.molcel.2011.12.036

[21] Ebner P, Poetsch I, Deszcz L, et al. The IAP family member BRUCE regulates autophagosome–lysosome fusion. Nat Commun. 2018;9(1):599. doi: 10.1038/s41467-018-02823-x

[22] Nam S-E, Cheung YWS, Nguyen TN, et al. Insights on autophagosome–lysosome tethering from structural and biochemical characterization of human autophagy factor EPG5. Commun Biol. 2021;4(1):291. doi: 10.1038/s42003-021-01830-x

[23] Shpilka T, Weidberg H, Pietrokovski S, et al. Atg8: an autophagy-related ubiquitin-like protein family. Genome Biol. 2011;12(7):226. doi: 10.1186/gb-2011-12-7-226

[24] Mizushima N. The ATG conjugation systems in autophagy. Curr Opin Cell Biol. 2020;63:1–10. doi: 10.1016/j.ceb.2019.12.001

[25] Rogov VV, Nezis IP, Tsapras P, et al. Atg8 family proteins, LIR/AIM motifs and other interaction modes. Autophagy Rep. 2023;2 (1):27694127. doi: 10.1080/27694127.2023.2188523

[26] Ungermann C, Kümmel D. Structure of membrane tethers and their role in fusion. Traffic. 2019;20(7):479–490. doi: 10.1111/tra.12655

[27] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596 (7873):583–589. doi: 10.1038/s41586-021-03819-2

[28] Varadi M, Anyango S, Deshpande M, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50(D1):D439–D444. doi: 10.1093/nar/gkab1061

[29] Ibrahim T, Khandare V, Mirkin FG, et al. AlphaFold2-multimer guided high-accuracy prediction of typical and atypical ATG8-binding motifs. PLOS Biol. 2023;21(2):e3001962. doi: 10.1371/journal.pbio.3001962

[30] Mirdita M, Schütze K, Moriwaki Y, et al. ColabFold: making protein folding accessible to all. Nat Methods. 2022;19 (6):679–682. doi: 10.1038/s41592-022-01488-1

[31] Rozenknop A, Rogov VV, Rogova NY, et al. Characterization of the interaction of GABARAPL-1 with the LIR motif of NBR1. J Mol Biol. 2011;410(3):477–487. doi: 10.1016/j.jmb.2011.05.003

[32] Kleckner IR, Foster MP. An introduction to nmr-based approaches for measuring protein dynamics. Biochim Biophys Acta. 2011;1814(8):942–968. doi: 10.1016/j.bbapap.2010.10.012

[33] Khaminets A, Heinrich T, Mari M, et al. Regulation of endoplasmic reticulum turnover by selective autophagy. Nature. 2015;522 (7556):354–358. doi: 10.1038/nature14498

[34] Olsvik HL, Lamark T, Takagi K, et al. 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. 2015;290(49):29361–29374. doi: 10.1074/jbc.M115.686915

[35] Cheng X, Wang Y, Gong Y, et al. Structural basis of FYCO1 and MAP1LC3A interaction reveals a novel binding mode for Atg8-family proteins. Autophagy. 2016;12(8):1330–1339. doi: 10.1080/15548627.2016.1185590

[36] Sakurai S, Tomita T, Shimizu T, et al. The crystal structure of mouse LC3B in complex with the FYCO1 LIR reveals the importance of the flanking region of the LIR motif. Acta Crystallogr F Struct Biol Commun. 2017;73(3):130–137. doi: 10.1107/S2053230X17001911

[37] Li J, Zhu R, Chen K, et al. Potent and specific Atg8-targeting autophagy inhibitory peptides from giant ankyrins. Nat Chem Biol. 2018;14(8):778–787. doi: 10.1038/s41589-018-0082-8

[38] Wirth M, Zhang W, Razi M, et al. Molecular determinants regulating selective binding of autophagy adapters and receptors to ATG8 proteins. Nat Commun. 2019;10(1):2055. doi: 10.1038/s41467-019-10059-6

[39] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630 (8016):493–500. doi: 10.1038/s41586-024-07487-w

[40] Wu F, Watanabe Y, Guo X-Y, et al. Structural basis of the differential function of the two C. elegans Atg8 homologs, LGG-1 and LGG-2, in autophagy. Mol Cell. 2015;60(6):914–929. doi: 10.1016/j.molcel.2015.11.019

[41] Suzuki H, Tabata K, Morita E, et al. Structural basis of the autophagy-related LC3/Atg13 LIR complex: recognition and interaction mechanism. Structure. 2014;22(1):47–58. doi: 10.1016/j.str.2013.09.023

[42] Lv M, Wang C, Li F, et al. Structural insights into the recognition of phosphorylated FUNDC1 by LC3B in mitophagy. Protein Cell. 2017;8(1):25–38. doi: 10.1007/s13238-016-0328-8

[43] Stanton AE, Hughson FM. The machinery of vesicle fusion. Curr Opin Cell Biol. 2023;83:102191. doi: 10.1016/j.ceb.2023.102191

[44] Sora V, Kumar M, Maiani E, et al. Structure and dynamics in the ATG8 Family from experimental to computational techniques. Front Cell Dev Biol. 2020;8:420. doi: 10.3389/fcell.2020.00420

[45] Kwon DH, Kim L, Kim B-W, et al. A novel conformation of the LC3-interacting region motif revealed by the structure of a complex between LC3B and RavZ. Biochem Biophys Res Commun. 2017;490(3):1093–1099. doi: 10.1016/j.bbrc.2017.06.173

[46] Nguyen TN, Padman BS, Usher J, et al. Atg8 family LC3/GABARAP proteins are crucial for autophagosome–lysosome fusion but not autophagosome formation during PINK1/Parkin mitophagy and starvation. J Cell Biol. 2016;215(6):857–874. doi: 10.1083/jcb.201607039

[47] Weidberg H, Shvets E, Shpilka T, et al. LC3 and GATE-16/GABARAP subfamilies are both essential yet act differently in autophagosome biogenesis. EMBO J. 2010;29(11):1792–1802. doi: 10.1038/emboj.2010.74

[48] Manil-Ségalen M, Lefebvre C, Jenzer C, et al. The C. elegans LC3 acts downstream of GABARAP to degrade autophagosomes by interacting with the HOPS subunit VPS39. Dev Cell. 2014;28(1):43–55. doi: 10.1016/j.devcel.2013.11.022

[49] Djeddi A, Al Rawi S, Deuve JL, et al. Sperm-inherited organelle clearance in C. elegans relies on LC3-dependent autophagosome targeting to the pericentrosomal area. Development. 2015;142(9):1705–1716. doi: 10.1242/dev.117879

[50] Palovcak E, Wang F, Zheng SQ, et al. A simple and robust procedure for preparing graphene-oxide cryo-em grids. J Struct Biol. 2018;204(1):80–84. doi: 10.1016/j.jsb.2018.07.007

[51] Punjani A, Rubinstein JL, Fleet DJ, et al. cryoSPARC: algorithms for rapid unsupervised cryo-em structure determination. Nat Methods. 2017;14(3):290–296. doi: 10.1038/nmeth.4169

[52] Meng EC, Goddard TD, Pettersen EF, et al. UCSF ChimeraX: tools for structure building and analysis. Protein Sci. 2023;32(11):e4792. doi: 10.1002/pro.4792

[53] Afonine PV, Poon BK, Read RJ, et al. Real-space refinement in PHENIX for cryo-em and crystallography. Acta Crystallogr D Struct Biol. 2018;74(6):531–544. doi: 10.1107/S2059798318006551

[54] Delaglio F, Grzesiek S, Vuister G, et al. Nmrpipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6(3). doi: 10.1007/BF00197809

[55] Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics. 2015;31(8):1325–1327. doi: 10.1093/bioinformatics/btu830

[56] Williamson MP. Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc. 2013;73:1–16. doi: 10.1016/j.pnmrs.2013.02.001

[57] Bui JM, Gsponer J. Phosphorylation of an intrinsically disordered segment in Ets1 shifts conformational sampling toward binding-competent substates. Structure. 2014;22(8):1196–1203. doi: 10.1016/j.str.2014.06.002

[58] Martínez-Rosell G, Giorgino T, De Fabritiis G. PlayMolecule ProteinPrepare: a web application for protein preparation for molecular dynamics simulations. J Chem Inf Model. 2017;57(7):1511–1516. doi: 10.1021/acs.jcim.7b00190

[59] Tian C, Kasavajhala K, Belfon KAA, et al. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. J Chem Theory Comput. 2020;16(1):528–552. doi: 10.1021/acs.jctc.9b00591

[60] Case DA, Aktulga HM, Belfon K, et al. AmberTools. J Chem Inf Model. 2023;63(20):6183–6191. doi: 10.1021/acs.jcim.3c01153

[61] Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys. 1980;72(4):2384–2393. doi: 10.1063/1.439486

[62] Darden T, York D, Pedersen L. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. J Chem Phys. 1993;98(12):10089–10092. doi: 10.1063/1.464397

[63] Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23(3):327–341. doi: 10.1016/0021-9991(77)90098-5

[64] Winter G, Waterman DG, Parkhurst JM, et al. DIALS: implementation and evaluation of a new integration package. Acta Crystallogr D Struct Biol. 2018;74(Pt 2):85–97. doi: 10.1107/S2059798317017235

[65] Agirre J, Atanasova M, Bagdonas H, et al. The CCP4 suite: integrative software for macromolecular crystallography. Acta Crystallogr D Struct Biol. 2023;79(6):449–461. doi: 10.1107/S2059798323003595

[66] McCoy AJ, Grosse-Kunstleve RW, Adams PD, et al. Phaser crystallographic software. J Appl Crystallogr. 2007;40(4):658–674. doi: 10.1107/S0021889807021206

[67] Afonine PV, Grosse-Kunstleve RW, Echols N, et al. Towards automated crystallographic structure refinement with phenix. Refine. Acta Crystallogr D Biol Crystallogr. 2012;68(4):352–367. doi: 10.1107/S0907444912001308

[68] Emsley P, Lohkamp B, Scott WG, et al. Features and development of coot. Acta Crystallogr D Biol Crystallogr. 2010;66(4):486–501. doi: 10.1107/S0907444910007493

[69] Tian W, Chen C, Lei X, et al. Castp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res. 2018;46(W1):W363–W367. doi: 10.1093/nar/gky473