A strategy for dissecting the architectures of native macromolecular assemblies

Yi Shi[1], Riccardo Pellarin[2–5], Peter C Fridy[6], Javier Fernandez-Martinez[6], Mary K Thompson[6], Yinyin Li[1], Qing Jun Wang[7], Andrej Sali[2–4], Michael P Rout[6] & Brian T Chait[1]

It remains particularly problematic to define the structures of native macromolecular assemblies, which are often of low abundance. Here we present a strategy for isolating complexes at endogenous levels from GFP-tagged transgenic cell lines. Using cross-linking mass spectrometry, we extracted distance restraints that allowed us to model the complexes' molecular architectures.

Chemical cross-linking with mass spectrometric readout (CX-MS) is emerging as particularly useful for providing distance restraints between amino acid residues in protein assemblies. Such data can be readily combined with other structural information to provide 'integrative' structural models[1,2]. Because the majority of such studies rely on recombinant protein complexes[3–6], the reach and application of CX-MS would be greatly extended if the technique were optimized for the analysis of native protein complexes isolated directly from their endogenous cellular milieus[7,8]. Major challenges to the analysis of such complexes relate to their often low abundance (necessitating ultrahigh-affinity reagents for their efficient and pristine capture[9]), their dynamic nature and their compositional and structural heterogeneity. Here we present a pipeline that incorporates ultrahigh-affinity reagents tuned to perform optimally with improved CX-MS analyses, in order to provide the requisite sensitivity to analyze macromolecular assemblies expressed at native levels.

Our pipeline (**Fig. 1**) was tuned for affinity capture of GFP-tagged protein complexes (with the tagged proteins expressed under control of their endogenous promoters) to take advantage of this most widely used tag for visualizing proteins *in vivo*. The extensive worldwide resource of transgenic organisms and animals that contain GFP-tagged proteins should be of great utility in this application[10,11].

In our method, the cells or tissues of interest are cryogenically milled into micrometer-sized particles to maximize the efficiency of solvent extraction of the protein complexes[12]. The resulting solubilized complexes are affinity captured on magnetic beads conjugated with anti-GFP V_HH nanobodies[13], which we have engineered to preclude reaction with amine-specific chemical cross-linking reagents such as disuccinimidyl suberate (DSS); these nanobodies are based on previously described subnanomolar-affinity LaG-16 (ref. 9) and 3K1K or GBP1 (PDB 3K1K)[14] reagents, but with lysine residues substituted by either arginine or glutamine residues (**Supplementary Figs. 1** and **2**). Because these substitutions are distal from the nanobodies' GFP-binding epitopes, they do not alter their binding affinities (**Supplementary Figs. 1b** and **2b–d**). These nanobodies are also chemically treated to minimize residual reactivity (for example, at their amino termini) toward amine cross-linkers. The resulting high-affinity, 'cross-linking inert' reagents allow rapid, efficient affinity capture of complexes followed by on-bead cross-linking, without cross-linking of the complexes to the bead-immobilized nanobodies. This property enables efficient removal of the cross-linked complex from the nanobody-conjugated beads (**Fig. 1** and **Supplementary Fig. 3**), circumventing the otherwise formidable challenge of efficiently proteolyzing the highly rigid cross-linked complexes while they are still chemically conjugated to the beads. This 'on-bead' cross-linking strategy also simplifies sample handling while providing a means for the removal of complexes that are not cleavable from the resin, including most GFP-transgenic systems.

After on-bead cross-linking by DSS, the complexes are eluted from the nanobody beads by denaturing buffer, resolved on SDS-PAGE and proteolyzed in-gel by trypsin (Online Methods). The proteolyzed products are fractionated, and the cross-linked peptides are identified by MS. Finally, the resulting distance restraints are used to depict architectures or generate structural models of the protein complexes[15].

To assess our overall workflow, we examined the *Saccharomyces cerevisiae* exosome, whose core structure has been elucidated at atomic resolution[16,17]. Exosomes are multiprotein assemblies that process and degrade cellular RNA[18]. An 11-subunit exosome complex (exo11) forms a barrel-like structure composed of two stacked rings: an upper trimeric ring of RNA binding domain–containing proteins (Rrp40, Rrp4 and Csl4), and a lower hexameric ring of PH-domain proteins (Rrp45, Rrp46, Rrp42, Rrp43, Mtr3 and Ski6), as well as two catalytic subunits (Rrp6 and Rrp44).

We affinity-captured the complex via a genomically GFP-tagged component of the core (*RRP46*-GFP). SDS-PAGE and MS analysis identified the 11 components of exo11, as well as two stoichiometric

[1]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York, USA. [2]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, USA. [3]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California, USA. [4]California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, California, USA. [5]Institut Pasteur, Paris, France. [6]Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, New York, USA. [7]Department of Molecular & Cellular Biochemistry, University of Kentucky, Lexington, Kentucky, USA. Correspondence should be addressed to B.T.C. (chait@rockefeller.edu).

RECEIVED 20 MARCH; ACCEPTED 20 AUGUST; PUBLISHED ONLINE 5 OCTOBER 2015; DOI:10.1038/NMETH.3617

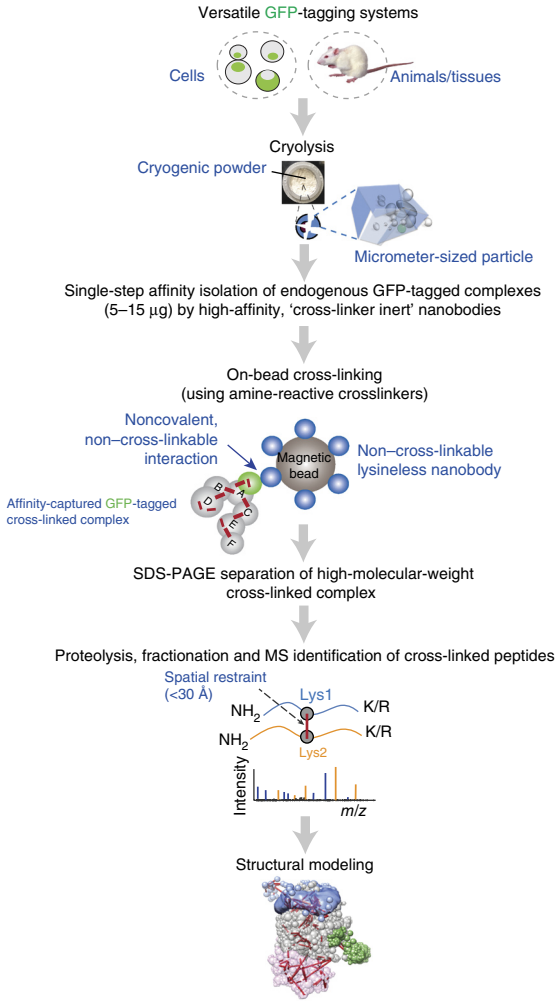


Figure 1 | Workflow for isolation and CX-MS analysis of GFP-tagged protein complexes. Cells or animal tissues are cryogenically milled into micrometer-sized particles (powders), which are dissolved in appropriate buffers to efficiently extract the proteins while preserving the architectures of the native protein assemblies. The GFP-tagged protein complex (5–15 µg) is then affinity-captured by ultrahigh-affinity, non–cross-linkable nanobody-conjugated magnetic beads. The protein complex is on-bead cross-linked by DSS, after which the cross-linked complex is eluted by heated LDS (lithium dodecyl sulfate) denaturing buffer and separated by SDS-PAGE electrophoresis. The efficiently cross-linked complex is in-gel proteolyzed to generate the cross-linked peptides, which are identified by high-resolution MS. The resulting information can be translated into residue-specific spatial restraints for computing integrative structural models.

subunits (Lrp1 and Ski7) and a substoichiometric component (Mpp6) that were not studied by X-ray crystallography[18] (**Fig. 2a** and **Supplementary Table 1**). The resulting connectivity map (**Fig. 2b**) largely recapitulated the previously determined architecture of the complex. For example, the catalytic subunit Rrp6 was located proximal to Rrp43, Csl4 and Rrp4, predominantly via its C-terminal domain (CTD) (residues 421–676), whereas its N-terminal domain (NTD) (residues 45–379) was extensively cross-linked to Lrp1 to potentially form a heterodimer[19] (**Fig. 2b**,**c**). The other catalytic subunit, Rrp44, was cross-linked to the periphery of the hexameric (lower) ring via Rrp45 and Ski6. Interestingly, the fact that the majority of the Rrp6 and Ski7 intersubunit cross-links shared identical lysine residues on Rrp4, Csl4 and Rrp43 (**Fig. 2c**) suggests that Rrp6 and Ski7 are potentially mutually exclusive and exist in two different complexes[16]. Consistent with this notion, GFP localization and affinity capture coupled to MS (**Fig. 2c** and **Supplementary Fig. 4**) showed that Rrp6-Lrp1-Mpp6-exo10 (exo11 minus Rrp6) formed a nucleus-localized complex, whereas the Ski7-exo10 complex was restricted to the cytoplasm (**Fig. 2d**, **Supplementary Fig. 5** and **Supplementary Tables 2** and **3**).

To evaluate the cross-link data, we measured the Euclidean Cα-Cα distances between cross-linked lysine residues in the published X-ray structure of exo10 and the CTD of Rrp6 (PDB 4IFD), as well as in the structures of the heterotrimer Rrp45, Ski6 and Rrp44 (ref. 20) (PDB 2WP8). We found that 34% of the cross-linking data could be mapped to the X-ray structures (**Supplementary Fig. 6** and **Supplementary Table 4**). Of these, >90% spanned ≤30 Å (the expected maximum reach), whereas four cross-links in PDB 4IFD and three in PDB 2WP8 mapped to much longer distances in the X-ray structures (44–65 Å) (**Fig. 2e**,**f** and **Supplementary Fig. 7**). Notably, all of these seemingly 'violated' cross-links (red) were located on Rrp44 (**Fig. 2g**,**h**), the essential catalytic subunit that exhibits both 3′-5′ RNA exoRNase and endoRNase activities, supporting the notion that this enzyme is highly dynamic, forming different conformers in the cell[21].

The relatively rich information about cross-linking of nuclear Rrp6-Lrp1 and cytosolic Ski7 to the exo10 core allowed us to utilize integrative modeling[15] to determine the architectures of these subcellular-specific exosome complexes (Online Methods). As shown in **Figure 2i**, the nuclear Rrp6-Lrp1 heterodimer is localized on top of the trimeric ring, which likely allows recruitment of RNA substrates for Rrp6 exoRNase activity. The position of the corresponding cytosolic component Ski7 with respect to exo10 was resolved with an average residue precision of ~9 Å (root-mean-square fluctuation; **Supplementary Fig. 8** and Online Methods); it docked on the upper ring with its C-terminal GTPase domain above Csl4 and Rrp4, close to the central channel (RNA entrance) (**Fig. 2j**). Intriguingly, its disordered NTD (**Fig. 2j**) (residues 90–264) was curved like a necklace that wrapped around Csl4 and the N terminus of Rrp4 before emerging on top of the upper ring, where it connected to the folded GTPase domain, reminiscent of the architecture of the nuclear Rrp6 CTD, but with the opposite chirality.

To assess the utility of our approach for examining low-abundance native complexes (tens to hundreds of copies per cell)[22,23], we studied the *S. cerevisiae* anaphase-promoting complex (also known as the cyclosome (APC/C)), a ubiquitin E3 ligase essential for cell cycle progression[24]. The architecture of the yeast APC/C (13 unique polypeptides) with its cofactor Cdh1 bound to a D-box peptide has been studied by cryo–electron microscopy and solved using recombinant overexpressed proteins at a resolution of ~10 Å (ref. 22) and recombinant human APC/C at 7.4 Å, resulting in characterization of the packing of secondary-structure segments for all subunits[25].

We affinity-captured the stoichiometric subunits of the endogenous complex together with the substoichiometric component Cdh1 (**Fig. 3a** and **Supplementary Table 5**). The resulting connectivity map (**Fig. 3b** and **Supplementary Table 6**) was consistent with the architecture revealed by electron microscopy[22,25]. For example, the tetratricopeptide repeat (TPR) domain–containing proteins (Cdc27, Cdc16 and Cdc23) were extensively cross-linked, revealing the TPR lobe to which accessory subunits (Apc13, Cdc26 and the yeast-specific subunit Apc9) were bound. Cross-links within the base/platform module (Apc1,

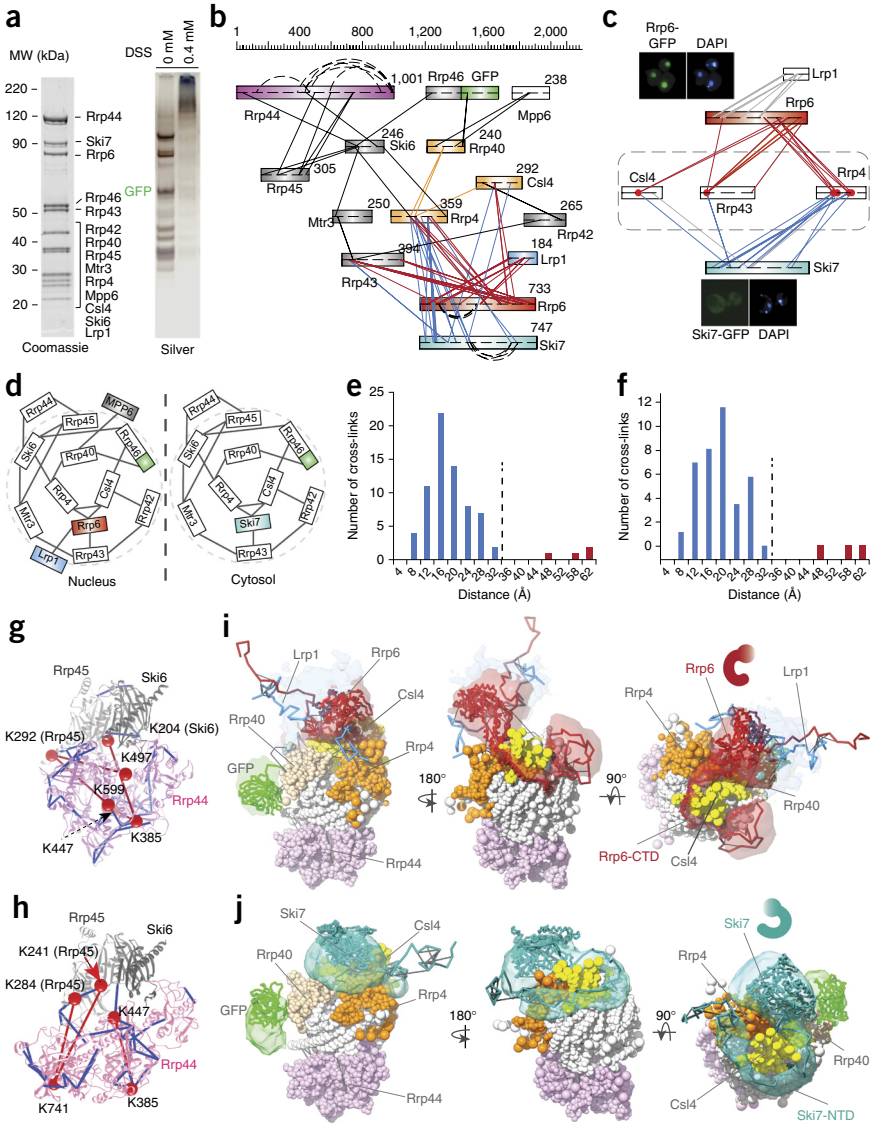


Figure 2 | CX-MS analysis and integrative structural modeling of the eukaryotic exosome complexes. (**a**) SDS-PAGE analyses of affinity-captured and on-bead cross-linked exosome complexes. The protein components were identified by MS. Approximately 30 pmol (~15 µg) of this affinity-captured material was used for each CX-MS analysis. MW, molecular weight. (**b**) Cross-link map for exosome complexes. Straight lines represent intersubunit cross-links, and curved dashed lines represent distal intrasubunit cross-links. (**c**) Heterogeneity of the exosome complexes revealed by CX-MS. Red dots represent lysine residues (on Rrp4, Rrp43 and Csl4) cross-linked to both Rrp6 and Ski7, indicating the presence of two different complexes in the affinity-captured material. Subcellular localizations of Rrp6-GFP (nucleus) and Ski6-GFP (cytosol) were imaged by fluorescence microscopy. (**d**) Models for the exosome complexes based on CX-MS data. (**e**) Euclidean Cα-Cα distance distributions of all measured cross-links on the crystal structure of PDB 4IFD . The *y*-axis provides the number of cross-links that were mapped to the crystal structures. Blue bars represent DSS cross-links shorter than 30 Å, and red bars represent those longer than 44 Å. (**f**) Euclidean Cα-Cα distance distributions of all measured cross-links on the crystal structure PDB 2WP8. (**g**,**h**) Observed cross-links on the crystal structures of exosome complexes (PDB 4IFD (**g**) and PDB 2WP8 (**h**)). Blue and red colors are coded as for **e** and **f**. (**i**,**j**) Structural models of nuclear Rrp6-Lrp1-exo10 and cytosolic Ski7-exo10 exosome complexes. Localization density maps of the 500 best scoring models are shown.

Apc2, Apc4 and Apc5) were prevalent. The TPR lobe and base modules were bridged by the cross-links Cdc23-Apc1, Cdc16-Apc1 and Cdc27-Apc2, defining a structure that seemed to be conserved from yeast to human. Another yeast-specific subunit, Mnd2, was cross-linked to both the TPR lobe (Cdc23 and Cdc16) and the scaffold (Apc4) via its N-terminal region, further connecting the two modules.

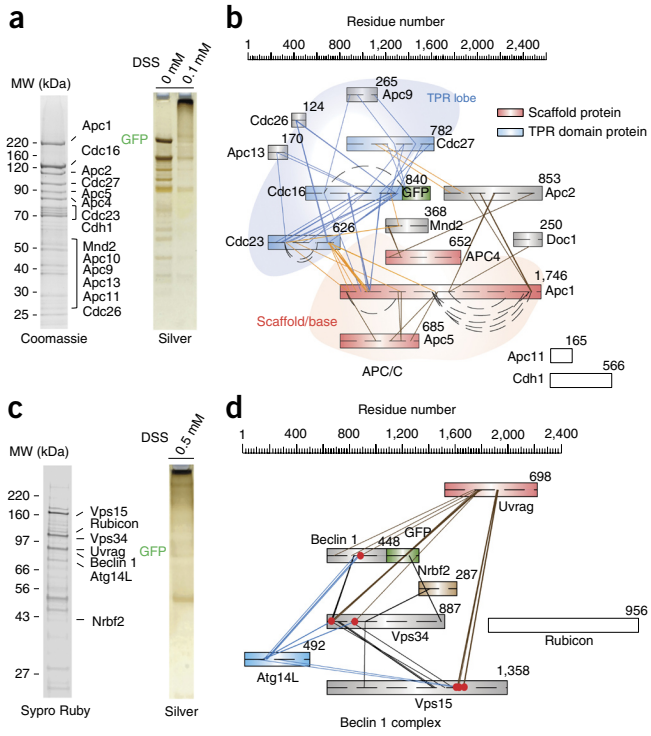
Although intersubunit cross-links provide the most useful distance restraints, certain intrasubunit cross-links may also reveal the architecture of a single subunit; for example, cross-links between the C-terminal and mid-domains of Apc1 constrain these regions so that they are spatially close and proximal to the central region of Apc2 (as in the high-resolution human APC/C structure). We also identified 'same-residue' cross-links on Cdc27 and Cdc16 (**Supplementary Fig. 9** and **Supplementary Table 7**), consistent with homodimer formation[25].

Finally, we explored the possibility of architecturally dissecting native GFP-tagged complexes isolated from a single transgenic mouse (*Mus musculus*) liver, investigating Beclin 1–associated complexes, which are central effectors of autophagy known to regulate aging and neurological disorders[26]. Despite the biological importance of these complexes, their overall architectures remain to be eludicated[27–29]. We applied our pipeline to map the arrangement of the complexes' seven major subunits that we had previously affinity isolated[30,31] (**Fig. 3c**,**d** and **Supplementary Tables 7** and **8**). Multiple cross-links were identified between Beclin 1–EGFP and its binding partner Vps34 through the coiled-coil domain of Beclin 1 (residues 142–267) and the N-terminal C2 domain of Vps34 (residues 1–255). Cross-links also showed that the C2 domain of Vps34 and the Vps15 CTD are in close proximity. Interestingly, Atg14L and Uvrag shared many intersubunit cross-links to the same lysine residues on Beclin 1, Vps34 and Vps15 triad, indicating that Atg14L and Uvrag interact in a mutually exclusive manner with the triad to form two different complexes[32] (**Fig. 3c**,**d**).

In summary, our approach yielded useful distance restraints in three exemplary low- to medium-abundance multisubunit complexes and elucidated complexes displaying heterogeneity, which is common in endogenous assemblies. Together with integrative modeling, this approach provides an effective means to model structures of proteins that contain extensive disordered or flexible regions, which present a problem for most other structural approaches.

Note added in proof: While this manuscript was in production, Makino et al.[33] *reported the X-ray crystal structures of a recombinant yeast nuclear exosome containing Exo9, Rrp44 and the Rrp6-Lrp1 heterodimer in complex with different RNAs. The authors observed*

Figure 3 | Architectural analyses of eukaryotic APC/C and tissue-specific (mouse liver) Beclin 1 complexes. (**a**) SDS-PAGE analyses of affinity-captured and on-bead cross-linked APC/C. Protein subunits were identified by MS. Approximately 15 pmol (15 µg) of the complex was purified from ~6 l of yeast culture for CX-MS analysis. (**b**) Cross-link map for eukaryotic APC/C. APC/C subunits are schematically represented and color-coded as follows: red, scaffold subunits of APC1, APC4 and APC5; orange, APC2; blue, TPR lobe subunits of Cdc27, Cdc16 and Cdc23; and gray, all other subunits. Straight lines represent intersubunit cross-links, and curved dashed lines represent distal intrasubunit cross-links. (**c**) SDS-PAGE analyses of affinity-captured and on-bead cross-linked Beclin 1 complexes. Components of the complexes were identified by MS. Affinity-captured and on-bead cross-linked complexes were stained by Sypro Ruby and silver stain, respectively. (**d**) Cross-link map for tissue-specific Beclin 1 complexes. Protein subunits are schematically represented and color coded as follows: red, Uvrag; blue, Atg14L; brown, Nrbf2; and gray, Beclin 1, Vps34 and Vps15. Red dots represent lysine residues (on Uvrag, Beclin 1 and Atg14L) cross-linked to both Uvrag and Atg14L. No intersubunit cross-links were identified on Rubicon.



substantial structural differences between catalytic subunits of the nuclear exosome (i.e., Rrp44 and the Rrp6-Lrp1 heterodimer) bound to two different RNAs. Our cross-link data from the endogenously isolated complex (for which any co-isolated RNAs were not characterized) are not completely satisfied by either of these structures, in agreement with the notion that these catalytic subunits exist in different conformers (and are RNA specific) and/or are dynamic in vivo. However, our cross-link data can be explained if one simultaneously considers our models and the structures from Makino et al.[33].

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. DNA sequence information has been submitted to GenBank under accession codes KT795402 and KT795403.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful to A.N. Krutchinsky (The Rockefeller University, New York, New York, USA) for providing the yeast GFP-CDC16 strain; S. Obado for technical assistance; E. Jacobs, J. Lacava, S.J. Kim and B. Webb for discussions and assistance; and Z. Yue (Icahn School of Medicine at Mount Sinai, New York, New York, USA) for sharing *Becn1-EGFP*/+ mice. Y.S. acknowledges M. Chen (R.G. Roeder lab, The Rockefeller University, New York, New York, USA) for generating a reagent that was not used for this study. This work was funded by the US National Institutes of Health (grants P41 GM103314 (to B.T.C.), R01 GM083960 (to A.S.) and P41 GM109824 (to A.S., M.P.R. and B.T.C.)) and the Ellison Medical Foundation (Q.J.W.).

AUTHOR CONTRIBUTIONS

Y.S. and B.T.C. conceived the research, with input from M.P.R., J.F.-M., R.P., Q.J.W., A.S. and P.C.F. Y.S., J.F.-M., P.C.F. and M.K.T. cloned and purified the nanobodies. P.C.F. performed the surface plasmon resonance measurements. Y.S. and J.F.-M. carried out the microscopic imaging. Y.S. performed biochemical and mass spectrometric analyses. R.P. performed the modeling analyses. Q.J.W. contributed reagents and discussed the results. P.C.F. and Y.L. discovered the LaG-16 nanobody. Y.S. and B.T.C. wrote the paper with input from M.P.R., J.F.-M., R.P., Q.J.W., A.S. and P.C.F. All authors reviewed the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Ward, A.B., Sali, A. & Wilson, I.A. *Science* **339**, 913–915 (2013).
2. Walzthoeni, T., Leitner, A., Stengel, F. & Aebersold, R. *Curr. Opin. Struct. Biol.* **23**, 252–260 (2013).
3. Erzberger, J.P. *et al. Cell* **158**, 1123–1135 (2014).
4. Knutson, B.A., Luo, J., Ranish, J. & Hahn, S. *Nat. Struct. Mol. Biol.* **21**, 810–816 (2014).
5. Gordiyenko, Y. *et al. Nat. Commun.* **5**, 3902 (2014).
6. Cevher, M.A. *et al. Nat. Struct. Mol. Biol.* **21**, 1028–1034 (2014).
7. Lasker, K. *et al. Proc. Natl. Acad. Sci. USA* **109**, 1380–1387 (2012).
8. Shi, Y. *et al. Mol. Cell. Proteomics* **13**, 2927–2943 (2014).
9. Fridy, P.C. *et al. Nat. Methods* **11**, 1253–1260 (2014).
10. Tsien, R.Y. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
11. Huh, W.K. *et al. Nature* **425**, 686–691 (2003).
12. Oeffinger, M. *et al. Nat. Methods* **4**, 951–956 (2007).
13. Muyldermans, S. *Annu. Rev. Biochem.* **82**, 775–797 (2013).
14. Kirchhofer, A. *et al. Nat. Struct. Mol. Biol.* **17**, 133–138 (2010).
15. Russel, D. *et al. PLoS Biol.* **10**, e1001244 (2012).
16. Makino, D.L., Baumgartner, M. & Conti, E. *Nature* **495**, 70–75 (2013).
17. Wasmuth, E.V., Januszyk, K. & Lima, C.D. *Nature* **511**, 435–439 (2014).
18. Houseley, J., LaCava, J. & Tollervey, D. *Nat. Rev. Mol. Cell Biol.* **7**, 529–539 (2006).
19. Feigenbutz, M., Jones, R., Besong, T.M., Harding, S.E. & Mitchell, P. *J. Biol. Chem.* **288**, 15959–15970 (2013).
20. Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E. & Conti, E. *Cell* **139**, 547–559 (2009).
21. Liu, J.J. *et al. Nat. Struct. Mol. Biol.* **21**, 95–102 (2014).
22. Schreiber, A. *et al. Nature* **470**, 227–232 (2011).
23. Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. *Nat. Methods* **11**, 319–324 (2014).
24. Peters, J.M. *Nat. Rev. Mol. Cell Biol.* **7**, 644–656 (2006).
25. Chang, L., Zhang, Z., Yang, J., McLaughlin, S.H. & Barford, D. *Nature* **513**, 388–393 (2014).
26. Vidal, R.L., Matus, S., Bargsted, L. & Hetz, C. *Trends Pharmacol. Sci.* **35**, 583–591 (2014).
27. Li, X. *et al. Nat. Commun.* **3**, 662 (2012).
28. Miller, S. *et al. Science* **327**, 1638–1642 (2010).
29. Baskaran, S. *et al. eLife* doi:10.7554/eLife.05115 (2014).
30. Zhong, Y. *et al. Nat. Cell Biol.* **11**, 468–476 (2009).
31. Zhong, Y. *et al. J. Biol. Chem.* **289**, 26021–26037 (2014).
32. Itakura, E., Kishi, C., Inoue, K. & Mizushima, N. *Mol. Biol. Cell* **19**, 5360–5372 (2008).
33. Makino, D.L. *et al.* RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature* **524**, 54–58 (2015).

ONLINE METHODS

Cloning, purification and chemical conjugation of anti-GFP V_HH nanobodies. Codon-optimized DNA sequences of lysineless V_HH nanobodies were synthesized (Genscript) and cloned into pET21-pelB-V_HH vector using BamHI and XhoI restriction sites (**Supplementary Note**). A flexible linker sequence of GGGSGGC was inserted at the carboxyl terminus of the proteins to allow chemical conjugation and proper folding of the proteins. The nanobodies were expressed and were purified by nickel–nitrilotriacetic acid beads as previously described[9]. Purified nanobodies were then conjugated to epoxy-activated Dynabeads[34] (Invitrogen). For each 1 mg of epoxyl Dynabeads, ~5 µg of nanobodies was used for the conjugation. The lysineless nanobody beads (LaG-16-2K/R or 3K1K-3K/R) were further chemically blocked by 0.5 mM DSS for 30–45 min at 25 °C with constant agitation on a shaker and were subsequently quenched by a 100 mM final concentration of Tris-HCl (pH 8). The reagents were stored at 4 °C and were used within 3 weeks of conjugation.

Comparison of efficacies of three different anti-GFP affinity reagents for on-bead cross-linking. Equal amounts of llama polyclonal antibody (Poly, generated in-house), wild-type LaG-16 and lysineless LaG-16-2K/R (2K/R) nanobodies were conjugated in parallel to epoxy magnetic beads. The antibody-conjugated magnetic beads were pretreated with DSS cross-linker to block residual reactive amines and were then used for the affinity isolation of the APC/C. The isolated complex was on-bead cross-linked by 0.1 mM DSS for 25 min and eluted by hot LDS buffer. The complex was subjected to SDS-PAGE electrophoresis; gel regions corresponding to the GFP-tagged Cdc16 (~70–160 kDa for non–cross-linked Cdc16 and >350 kDa for cross-linked Cdc16) were excised from the gel and digested by trypsin.

The purified peptides were analyzed by an Orbitrap Fusion mass spectrometer coupled online to the Easy LC system (Thermo). For each analysis, peptides corresponding to ~20–40 ng of the purified complex were loaded onto an Easy-Spray column heated at 35 °C (C18, 3-µm particle size, 200-Å pore size, and 50 µm × 15 cm; Thermo Fisher) and eluted using a 25-min liquid chromatography gradient (2% B–7% B, 0–3 min; 7% B–42% B, 3–21 min; 42% B–100% B, 21–25 min; mobile phase A consisted of 0.1% formic acid (FA), and mobile phase B consisted of 0.1% FA in acetonitrile (ACN)). The flow rate was ~325 nl/min. The instrument was operated in the data-dependent mode, where the top eight most abundant ions were fragmented by higher energy collisional dissociation (HCD) (normalized HCD energy, 30) and analyzed in the Orbitrap mass analyzer. The target resolution for MS1 was 60,000, and for MS2 it was 15,000. Peptides (300–1,500 m/z) with charge state of 2–5 were isolated by the ion trap (isolation window, 1.8 Th) and detected by the Orbitrap analyzer. The experiments were repeated three times, and each repeated sample was measured three times to minimize the LC-MS analytical variations.

For label-free quantification, the raw data were searched against the yeast ORF database (01/05/2010) and quantified by MaxQuant[35] (version 1.5). The search parameters included the following: initial MS accuracy, <20 p.p.m.; main search MS accuracy tolerance, <4.5 p.p.m.; cysteine carbamidomethyl as a fixed modification; and protein N-terminal acetylation and methionine oxidation as variable modifications. A maximum of two trypsin missed-cleavage sites was allowed. More than 30 unique peptides from Cdc16 were identified and used for the quantification. For label-free quantification, ion intensities of Cdc16 (which in essence represents the relative signals of the complex) from the cross-linked APC/C (isolated by three different anti-GFP affinity reagents) were normalized by those of input signals; the resulting ratios were normalized to 2K/R and are plotted in **Supplementary Figure 3**.

K_d measurements by surface plasmon resonance (SPR). SPR measurements were obtained on a Proteon XPR36 Protein Interaction Array System (Bio-Rad). Recombinant GFP was immobilized on a ProteOn GLC sensor chip. The chip surface was first activated with 50 mM sulfo-NHS and 50 mM EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide), run at a flow rate of 30 µl/min for 300 s. The ligand was then diluted to 5 µg/ml in 10 mM sodium acetate, pH 5.0, and injected at 25 µl/min for 180 s. Finally, the surface was deactivated by 1 M ethanolamine-HCl (pH 8.5) run at 30 µl/min for 300 s. This led to immobilization of approximately 200 response units of ligand.

We determined the K_d values of recombinant nanobodies by injecting four concentrations of each protein, in triplicate, with a running buffer of 20 mM Na-HEPES, pH 8.0, 150 mM NaCl, 0.01% Tween-20. Proteins were injected at 100 µl/min for 90 s, followed by a dissociation time of 660 s. Between injections, residual bound protein was eliminated by regeneration with 4.5 M $MgCl_2$ in 10 mM Tris-HCl, pH 7.5, run at 100 µl/min for 90 s. Binding sensorgrams from these injections were processed and analyzed using the ProteOn Manager software. Binding curves were fit to the data with a Langmuir model using grouped k_a, k_d and R_{max} values.

GFP transgenic strains, cell culturing and cryogenic grinding of yeast cells and mouse liver. The GFP-tagged *S. cerevisiae* strains of *RRP6* (*YOR001W*), *SKI7* (*YOR076C*) and *RRP46* (*YGR095C*) were purchased from Invitrogen. The GFP-*CDC16*-3 × Flag-His6 (YKL022C) strain (MATα BY4742) was generated as previously described[36]. All yeast cells were cultured in yeast extract peptone dextrose medium (MP Biomedicals) at 30 °C to a density of 3×10^7 to 4×10^7 cells/ml before harvesting. The *Becn1-EGFP*/+ bacterial artificial chromosome (BAC) transgenic mice express Beclin 1–EGFP under endogenous transcriptional control[30]. One mouse (female, C57BL/6 background, age 5 months) was killed; its liver was collected and immediately frozen in liquid nitrogen for storage. Yeast strains and the mouse liver were cryogenically milled with a planetary ball mill[12,37] (Retsch PM 100), and the cryogenic powders were stored at −80 °C until use. Animal procedures were approved by the Institutional Animal Care and Use Committees at the University of Kentucky, where the mouse was housed.

Fluorescence microscopy. *S. cerevisiae* strains *RRP6*-GFP (YOR001W) and *SKI7*-GFP (YOR076C) were grown to mid-log phase, stained with 2.5 µg/ml DAPI for 20 min, washed with PBS and visualized *in vivo* by a fluorescence microscope (Axioplan 2, Carl Zeiss) equipped with a cooled charge-coupled device (CCD) camera (ORCA-ER, Hamamatsu). The images were analyzed with Openlab software (PerkinElmer) and processed using Adobe Photoshop CS5 (Adobe).

Affinity capture of protein complexes. To identify the compositions of the yeast exosome and APC/C, we dissolved 200 mg and 500 mg of yeast grindate powder in 0.75 ml and 1.5 ml yeast lysis buffer, respectively (100 mM Na-HEPES, pH 7.9, 0.2% Tween-20, 350 mM sodium citrate, protease inhibitor cocktail (Roche)) and briefly sonicated the mixture on ice. The protein extracts were centrifuged at 14,000 r.p.m. for 12 min, and the supernatants were collected. We affinity-isolated the protein complexes by adding ~2 mg of lysineless nanobody-conjugated Dynabeads to the supernatant for 30 min (exosome complex) or 60 min (APC/C) at 4 °C on a rotating wheel. The beads were then washed four times with 0.5–1 ml of yeast pullout buffer and collected using a magnetic stand.

To identify the Beclin 1–EGFP complex, we dissolved ~400 mg of the mouse liver grindate powder in 1 ml of pullout buffer (100 mM Na-HEPES, pH 7.9, 0.2% Tween-20, 150 mM sodium chloride, 350 mM sodium citrate, and protease inhibitor cocktail) and briefly sonicated it on ice. The protein extract was centrifuged at 14,000 r.p.m. for 20 min, and the supernatant was transferred and incubated with ~2 mg of the nanobody-conjugated Dynabeads for ~2 h at 4 °C on a rotating wheel. The beads were washed four times and collected using a magnetic stand before elution.

On-bead crosslinking. We affinity-isolated several hundred nanograms of the complexes and cross-linked them at various concentrations of DSS cross-linker to empirically decide the optimal reaction conditions—that is, the minimal concentration of cross-linker that would enable all component proteins to migrate to the high-molecular-weight regions in the gel (**Supplementary Fig. 10**). We used ~40 mg of the nanobody-Dynabeads (3K1K-3K/R or Lag-16-2K/R) to isolate the protein complexes for CX-MS analyses from 3–4 g yeast grindate powder of Rrp46-GFP (exosome), ~12–15 g grindate powder of GFP-Cdc16 (APC/C), and the liver from a single transgenic GFP-Bcln1 mouse.

For CX-MS analysis, the purified complex (~5–15 µg) was on-bead cross-linked by isotopically labeled DSS (d0:d12 = 1:1, Creative Molecules) in 100 mM K-HEPES, pH 7.9. (We note that the isotopic label was not needed in the present experiments and can be omitted.) The cross-linked complex was quenched and eluted from the beads by LDS buffer (containing 100 mM DTT and 100 mM Tris-HCl, pH 8.5) at 85 °C for 10 min. The sample was cooled at room temperature for cysteine alkylation and separated by electrophoresis in either 4–12% Bis-Tris or 3–8% Tris-acetate SDS-PAGE gel (Invitrogen). SDS-PAGE allows (i) separation and enrichment of highly cross-linked materials (which run as high-molecular-weight products) to reduce the complexity of the digest and (ii) elimination of the detergents needed for solubilization of the complex, as well as chemical 'bleed' products from the magnetic beads, both of which can greatly reduce MS sensitivity. In addition, the presence of SDS helps denature and disperse the rigid, cross-linked complex, greatly facilitating proteolysis. We also tested an alternative possibility for obtaining distance restraints by treating whole-cell lysate with DSS cross-linker before affinity capture of the GFP-tagged protein complexes, but we found that this procedure provided insufficient intersubunit cross-linking yield to be useful while producing a high level of nonspecific protein background (**Supplementary Fig. 11** and **Supplementary Table 9**).

Affinity isolation of the yeast exosome complex from whole-cell lysates cross-linked with different concentrations of DSS cross-linker. One gram of yeast cryomilled powder (Rrp46-GFP) was resuspended in 3.5 ml exosome complex buffer (100 mM Na-HEPES, pH 7.9/8.0, 0.2% Tween-20, 350 mM sodium citrate, protease inhibitor cocktail; Roche). Cell lysate was centrifuged at 13,000–14,000 r.p.m. for 15 min, and the supernatant was collected. Equal amounts of the soluble whole-protein lysate were then cross-linked for 2 h at 4 °C by different concentrations of DSS cross-linker (0, 0.04, 0.2, 1 and 5 mM (because of the solubility of DSS, we did not intend to further increase its concentration in the whole-cell lysate)). After cross-linking of the whole-protein lysates, the reaction was quenched by 50 mM Tris buffer (final concentration). Cross-linked protein lysates were then centrifuged to remove the precipitated proteins, and the supernatants were collected for affinity capture of the exosome complex (using the protocol described above). Approximately 0.5 mg of 2K/R nanobody-conjugated magnetic beads was used for each affinity capture (from ~200 mg of cryogenic yeast powder).

The affinity-isolated proteins were loaded onto a 4–12% Bis-Tris SDS-PAGE gel and digested by trypsin. After proteolysis, the purified peptides were analyzed by LC-MS (two analytical replicates; a 25-min LC gradient was used for each analysis) with a Q Exactive (QE) Plus mass spectrometer (Thermo). The proteins were identified by X! Tandem (http://www.thegpm.org/tandem/), and the relative amounts were quantified by peptide spectra counts using unique peptide identifications (**Supplementary Table 9**).

Proteolysis. The cross-linked complex was digested in-gel with trypsin to generate cross-linked peptides. The gel region above 220 kDa was sliced, crushed into small pieces and incubated overnight with trypsin at a 1:20 (wt/wt) ratio in 50 mM ammonium bicarbonate buffer containing 0.1% Rapigest (Waters) at 37 °C. The cross-linked complex was then re-digested by a second bolus of the same amount of trypsin for 4 h. The digestion mixture was acidified by FA, and peptides were extracted twice by the addition of a stepwise gradient of ACN and vigorous shaking at room temperature. The peptide extracts were pooled and centrifuged at 13,000 r.p.m. for 10 min. The supernatant was collected, flash-frozen and lyophilized to ~30 µl. The peptide mixture was then reconstituted in 5% MeOH, 0.2% FA and desalted by a C18 SPE column (Waters). After lyophilization (to ~5 µl), the resulting peptides were reconstituted in 20 µl of a solution containing 30% ACN and 0.2% FA and fractionated by peptide size-exclusion chromatography[38] (Superdex Peptide PC 3.2/30, GE Healthcare) using off-line HPLC separation and an autosampler (Agilent Technologies). Two or three size-exclusion chromatography fractions covering the molecular mass range of ~2.5–8 kDa were collected and analyzed by LC-MS.

LC-MS analysis and MS data interpretation. To characterize the composition of the affinity-captured material, we loaded a portion of the purified proteolytic peptides (corresponding to 100–200 ng of the purified complex) onto a self-packed PicoFrit column (5 cm × 75 µm inner diameter, 3 µm, New Objective). The column was packed with 5 cm of reversed-phase C18 material (3 µm porous silica, 200-Å pore size, Dr. Maisch). Mobile phase A consisted of 0.5% acetic acid, and mobile phase B consisted of 70%

doi:10.1038/nmeth.3617

ACN with 0.5% acetic acid. The peptides were eluted in a linear LC gradient using an HPLC system (Agilent) and analyzed with an LTQ Velos Orbitrap Pro mass spectrometer (Thermo Fisher). The instrument was operated in the data-dependent mode, where the top 20 most abundant ions were fragmented by collision-induced dissociation (normalized energy: 35) and analyzed in the linear ion trap. The target resolution for MS1 was 60,000. Other instrumental parameters included lock mass at 371.1012 Da and a minimal threshold of 3,000 to trigger an MS/MS event. Ion trap accumulation limits (precursors) were 1×10^5 and 1×10^6, respectively, for the linear ion trap and Orbitrap.

Raw data from the affinity capture–MS experiments were converted to an mzXML file and searched online by X! Tandem[39]. Database search parameters included mass accuracies of <10 p.p.m. for MS1 and <0.4 Da for MS2 (or <20 p.p.m. for MS2 for the sample analyzed by QE MS); cysteine carbamidomethylation as a fixed modification; and methionine oxidation, N-terminal acetylation, and phosphorylation (at serine, threonine and tyrosine) as variable modifications. A maximum of one trypsin missed-cleavage site was allowed. The false positive rate was estimated as about 1%. The search results are provided in **Supplementary Tables 1**–**3**, **5** and **7**.

For cross-link identifications, the purified peptides were dissolved in the sample loading buffer (5% MeOH, 0.2% FA) and analyzed by an LTQ Velos Orbitrap Pro mass spectrometer (for most analyses) or an Orbitrap QE Plus mass spectrometer (Thermo Fisher). For the analysis by the Velos Orbitrap mass spectrometer, the dissolved peptides were pressure-loaded onto a self-packed PicoFrit column with an integrated electrospray ionization emitter tip (360-µm outer diameter, 75-µm inner diameter with a 15-µm tip; New Objective). The column was packed with 8 cm of reversed-phase C18 material (3-µm porous silica, 200-Å pore size, Dr. Maisch). Mobile phase A consisted of 0.5% acetic acid, and mobile phase B consisted of 70% ACN with 0.5% acetic acid. The peptides were eluted in a 120-min LC gradient (8%–50% B, 0–93 min, followed by 50%–100% B, 93–110 min, and equilibrated with 100% A until 120 min) using an HPLC system (Agilent) and analyzed with an LTQ Velos Orbitrap Pro mass spectrometer. The flow rate was ~200 nL/min. The spray voltage was set at 1.9–2.3 kV. The capillary temperature was 275 °C, and ion transmission on Velos S lenses was set at 35%. The instrument was operated in the data-dependent mode, where the top eight most abundant ions were fragmented by HCD[40] (HCD energy: 27–29; 0.1-ms activation time) and analyzed in the Orbitrap mass analyzer. The target resolution was 60,000 for MS1 and 7,500 for MS2. Ions (370–1,700 m/z) with a charge state of >3 were selected for fragmentation. A dynamic exclusion of (15 s/2/55 s) was used. Other instrumental parameters included lock mass at 371.1012 Da, a mass exclusion window of 1.5 Th and a minimal threshold of 5,000 to trigger an MS/MS event. Ion trap accumulation limits (precursors) were 1×10^5 and 1×10^6, respectively, for the linear ion trap and Orbitrap. For MS2, the Orbitrap ion accumulation limit was 5×10^5. The maximum ion injection time for the Orbitrap was 500–700 ms. The QE Plus instrument was directly coupled to an Easy LC system (Thermo Fisher), and the experimental parameters used were similar to those for the Velos Orbitrap. The cross-linked peptides were loaded onto an Easy-Spray column heated at 35 °C (C18, 3-µm particle size, 200-Å pore size, 50 µm × 15 cm; Thermo Fisher) and eluted using a 120-min LC gradient (2%–10% B, 0–6 min, 10%–35% B, 6–102 min, 35%–100% B, 102–113 min, followed by equilibration, where mobile phase A consisted of 0.1% FA and mobile phase B consisted of 0.1% FA in ACN). The flow rate was ~300 nl/min. The spray voltage was 2.0 kV, and the top ten most abundant ions (with charge stages of 3–7) were selected and fragmented by HCD.

The cross-link raw data were transformed to MGF (Mascot generic format) by pXtract 1.0 and searched by pLink[41] using a FASTA database containing protein sequences of the identified subunits of each complex. For the exosome cross-linking database search, we included the budding yeast protein sequences of 14 known exosome subunits of Rrp44 (YOL021C), Rrp40 (YOL142W), Rrp46 (YGR095C), Rrp4 (YHR069C), Mtr3 (YGR158C), Csl4 (YNL232W), Rrp6 (YOR001W), Rrp45 (YDR280W), Ski7 (YOR076C), Rrp43 (YCR035C), Rrp42 (YDL111C), LRP1 (YHR081W), Ski6 (YGR195W), and Mpp6 (YNR024W), together with GFP and two other relatively abundant proteins (RPS3 (YNL178W) and PBP1 (YGR178C), which were identified with more than ten unique peptides by affinity-capture MS (**Supplementary Table 1**)). For the APC/C cross-linking database search, we included the budding yeast protein sequences of 11 known APC/C subunits of Apc11 (YDL008W), Cdc16 (YKL022C), Cdc23 (YHR166C), Cdc27 (YBL084C), Swm1 (YDR260C), Apc1 (YNL172W), Apc2 (YLR127C), Apc4 (YDR118W), Apc5 (YOR249C), Doc1 (YGL240W), Cd26 (YFR036W), Mnd2 (YIR025W), Cdh1 (YGL003C), Apc9 (YLR102C), and GFP. For the mouse (*M. musculus*) Beclin 1 complex cross-linking database search, seven subunits (Beclin 1, Uvrag, Atg14L, Vps34, Vps15, Rubicon and Nrbf2) and the GFP protein were used. An initial MS1 search window of 5 Da was allowed to cover all isotopic peaks of the cross-linked peptides. The data were automatically filtered using a mass accuracy of MS1 ≤10 p.p.m. and MS2 ≤20 p.p.m. of the theoretical monoisotopic mass (A0) and other isotopic masses (A + 1, A + 2, A + 3, and A + 4) as specified in the software. Other search parameters included cysteine carboxymethylation as a fixed modification and methionine oxidation as a variable modification. A maximum of two trypsin missed-cleavage sites was allowed. The initial search results were obtained using the default 5% false discovery rate, estimated using a target-decoy search strategy. In our analysis, we treated the 5% false discovery rate as a rough initial filter of the raw data (albeit quite permissive). Next, we manually applied additional filters to remove potential false positive identifications from our data set[8]. For positive identifications, both peptide chains needed to contain four or more amino acids. In addition, for both peptide chains, the major MS/MS fragmentation peaks had to be assigned and follow a pattern that contained a continuous stretch of fragmentations. The cross-link maps were generated using AUTOCAD (Autodesk, Inc.; educational version). A comparison of cross-link data from the three different complexes is provided in **Supplementary Table 10**.

Annotated high-resolution cross-link MS/MS spectra are provided in the **Supplementary Data**.

Relative stoichiometry quantification of affinity-captured exosome complexes by SYPRO Ruby stain. To quantify the relative stoichiometries of the complexes, we stained SDS-PAGE protein gels with SYPRO Ruby according to the manufacturer's instructions (Life Technologies) and visualized them using an LAS-3000

system (linear detection range; Fujifilm). The intensities of the different protein bands on the gels were measured using ImageJ software (National Institutes of Health) and were normalized according to the predicted molecular weights of the quantifiable subunit proteins. The stoichiometries of the exosome components were calculated relative to that of Rrp6-GFP or Ski7-GFP. Note that Mpp6 has the exact same molecular weight as Lrp1 (21 kDa) but was much less abundant in the affinity-capture complex (the ratio of normalized ion intensities[35,42] of Mpp6:Lrp1 was ~1:10; data not shown). We therefore did not consider Mpp6 in the stoichiometry determination.

Determining the structures of the subcellular specific exosome complexes with Integrative Modeling Platform (IMP). Our integrative approach for determining the exosome complex structure proceeds through four stages[7,43–46]: (1) gathering of data, (2) representation of subunits and translation of the data into spatial restraints, (3) configurational sampling to produce an ensemble of models that optimally satisfies the restraints, and (4) analysis and assessment of the ensemble. The modeling protocol was scripted using the Python Modeling Interface, a library for modeling macromolecular complexes based on the open-source IMP[15] package, release 2.3.0. Briefly, we represented the protein subunits of the exosome by beads of varying sizes, arranged into rigid or flexible strings, on the basis of the available crystallographic structures of the exosome complex (PDB 4IFD and 2HBJ) and GFP (PDB 1GFL), as well as a comparative model for Ski7(259–747) generated by Phyre2 (**Supplementary Fig. 12**). The relative stoichiometries of the key components were determined to be in 1:1 ratios in both the nuclear and cytosolic complexes[16,47] (**Supplementary Fig. 13**). The cross-linking data were then encoded into a Bayesian scoring function that restrained the distances spanned by the cross-linked residues[3,8,48]. Cytosolic Ski7-exo10 and nuclear Rrp6-Lrp1-exo10 complexes were computed separately. The 500 best scoring models for each subcellular complex were clustered to yield the localization density maps (solutions) shown in **Figure 2i,j**. The resulting solutions satisfied 95% of the input cross-link restraints (**Supplementary Fig. 14**). The precisions of the Ski7 and Rrp6 solutions (average r.m.s. deviation with respect to the cluster center) were, respectively, 22.7 Å and 40.0 Å, even though these two proteins contain long disordered regions (**Supplementary Fig. 12**).

PDB 2WP8 and 3Q8T were used in this study to map cross-links onto the atomic structures of the exosome and the Beclin 1 complex, respectively (**Fig. 2** and **Supplementary Figs. 6** and **15**). Cross-links mapping to the crystal structure of the Beclin 1 coil-coil domain were prepared using UCSF Chimera[49].

Representation of subunits. The domains of the two exosome complexes were represented by beads arranged into either a rigid body or a flexible string on the basis of the available crystallographic structures and comparative models. To balance the thoroughness of the configurational sampling and precision of model representation, we represented the structures in a multi-scale fashion. Sequence segments missing in the crystal structures were substituted by one or multiple beads of the corresponding size. For each domain and interface with an atomic model, the beads representing a structured region were kept rigid with respect to one another during configurational sampling (i.e., rigid bodies). The rigid bodies included the exo10 complex; GFP, which is fused to the C terminus of Rrp46; Rrp6 (residues 127–516); and Ski7 (residues 259–747). Segments without a crystallographic structure or comparative model (i.e., with an unknown structure) were represented by a flexible string of beads corresponding to a maximum of five residues each (**Supplementary Fig. 12**).

Bayesian scoring function. The Bayesian approach estimates the probability of a model, given information available about the system, including both prior knowledge and newly acquired experimental data. The model $M \equiv (X, \{\alpha_i\})$ includes the structure coordinates X and additional parameters $\{\alpha_i\}$. Using Bayes' theorem, the posterior probability $p(M|D, I)$, given data D and prior knowledge I, is $p(M|D, I) \propto p(D|M, I)p(M, I)$, where the likelihood function $p(D|M, I)$ is the probability of observing data D, given I and M, and the prior is the probability of model M, given I. To define the likelihood function, one needs a forward model that predicts the data point (i.e., the presence of a cross-link between two given residues) given any model M and a noise model that specifies the distribution of the deviation between the observed and predicted data points. The Bayesian scoring function is the negative logarithm of $p(D|M, I)p(M|I)$, which ranks the models identically to the posterior probability.

Briefly, the forward model f_n is computed as the probability of randomly picking two points $\tilde{r}_i$ and $\tilde{r}_j$ within the spheres centered on the Cα atoms of the cross-linked residues, with coordinates r_i and r_j, with unknown radii σ_i and σ_j, such that the distance between them $\tilde{r}_{ij}$ is less than the maximum cross-linker length l_{XL}; the radii σ_i and σ_j are proxies for the uncertainty of forming a cross-link, given structural model X. To reduce the number of parameters in the model, we used a single uncertainty parameter σ for all residues. We imposed l_{XL} = 21 Å for the DSS cross-linker.

The likelihood function for a cross-link d_n is $p(d_n|X, I) = \psi(1 - f_n(X)) + f_n(X)(1 - \psi)$, where ψ is the uncertainty of observing a cross-link and is approximately equal to the expected fraction of cross-links that are inconsistent with the structure X. We set ψ equal to 5%. The joint likelihood function $p(D|M, I)$ for a data set $D = \{d_n\}$ of N_{XL} independently observed cross-links is the product of the likelihood functions for each data point.

The model prior $p(M|I)$ is defined as a product of the priors $p(X)$ and $p(\sigma)$ on the structural coordinates X and uncertainty σ, respectively. The prior $p(X)$ is composed of the excluded volume restraints, the sequence connectivity restraints, and a weak restraint whose score depends linearly on the distance between cross-linked residues, with a slope of 0.01 Å^{-1}. $p(\sigma)$ is a uniform distribution over the interval [0, 100].

Sampling model configurations. Structural models of the Rrp6-Lrp1-exo10 and Ski7-exo10 complexes were computed by Replica Exchange Gibbs sampling, based on Metropolis Monte Carlo sampling[48]. This sampling was used to generate configurations of the system as well as values for the uncertainty parameters. The Monte Carlo moves included random translation and rotation of rigid bodies (1 Å and 0.025 rad maximum, respectively), random translation of individual beads in the flexible segments (1 Å maximum), and a Gaussian perturbation

doi:10.1038/nmeth.3617

of the uncertainty parameters. The sampling was run on 64 replicas, with temperatures ranging between 1.0 and 2.5. Two independent sampling calculations were run for each complex, each one starting with a random initial configuration, for a total of 100,000 models per complex. We divided this set of models into two ensembles of the same size to assess sampling convergence (data not shown).

Analysis of the model ensemble. For each ensemble, the 500 best scoring models (i.e., the solutions) were grouped by *k*-means clustering[50] on the basis of the r.m.s. deviation between the Rrp6-Lrp1 and Ski7 subunits, after the superposition of the exo10 subunits. Two dominant clusters of similar structure were identified, with one cluster scoring better than the other. The precision of a cluster was calculated as the average r.m.s. deviation with respect to the cluster center (i.e., the solution that had the minimal r.m.s. deviation distance with respect to the others). The per-residue precision of a cluster (the root-mean-square fluctuation; **Supplementary Fig. 9**) was calculated as the average r.m.s. deviation distance of a residue in a solution with respect to the cluster center. The solutions of a cluster, superposed on the exo10 structure, were converted into the probability of any volume element being occupied by Rrp6, Ski7, Lrp1 and the GFP tag (localization density)[43,45].

Code availability. The scripts and models are available at http://salilab.org/exosome.

34. Cristea, I.M. & Chait, B.T. *Cold Spring Harb. Protoc.* **2011**, pdb.prot5611 (2011).
35. Cox, J. & Mann, M. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
36. Deng, C. & Krutchinsky, A.N. *Methods Mol. Biol.* **1163**, 75–95 (2014).
37. Hakhverdyan, Z. *et al. Nat. Methods* **12**, 553–560 (2015).
38. Leitner, A. *et al. Mol. Cell. Proteomics* **11**, M111.014126 (2012).
39. Craig, R. & Beavis, R.C. *Bioinformatics* **20**, 1466–1467 (2004).
40. Olsen, J.V. *et al. Nat. Methods* **4**, 709–712 (2007).
41. Yang, B. *et al. Nat. Methods* **9**, 904–906 (2012).
42. Schwanhäusser, B. *et al. Nature* **473**, 337–342 (2011).
43. Alber, F. *et al. Nature* **450**, 683–694 (2007).
44. Alber, F. *et al. Nature* **450**, 695–701 (2007).
45. Fernandez-Martinez, J. *et al. J. Cell Biol.* **196**, 419–434 (2012).
46. Algret, R. *et al. Mol. Cell. Proteomics* **13**, 2855–2870 (2014).
47. Synowsky, S.A., van den Heuvel, R.H., Mohammed, S., Pijnappel, P.W. & Heck, A.J. *Mol. Cell. Proteomics* **5**, 1581–1592 (2006).
48. Rieping, W., Habeck, M. & Nilges, M. *Science* **309**, 303–306 (2005).
49. Pettersen, E.F. *et al. J. Comput. Chem.* **25**, 1605–1612 (2004).
50. MacQueen, J.B. Some methods for classification and analysis of multivariate observations. in *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1 (Eds. Le Cam, L.M. & Neyman, J.), 281–297 (Univ. of California Press, 1965).