Journal of Structural Biology 192 (2015) 151–158

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Site-specific labeling of proteins for electron microscopy

Corey M. Dambacher, Gabriel C. Lander *

Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA

ARTICLE INFO

Article history:
Received 18 August 2015
Received in revised form 22 September 2015
Accepted 23 September 2015
Available online 25 September 2015

Keywords:
Electron microscopy
Unnatural amino acids
Molecular labeling

ABSTRACT

Electron microscopy is commonly employed to determine the subunit organization of large macromolecular assemblies. However, the field lacks a robust molecular labeling methodology for unambiguous identification of constituent subunits. We present a strategy that exploits the unique properties of an unnatural amino acid in order to enable site-specific attachment of a single, readily identifiable protein label at any solvent-exposed position on the macromolecular surface. Using this method, we show clear labeling of a subunit within the 26S proteasome lid subcomplex that has not been amenable to labeling by traditional approaches.

1. Introduction

Determining the structural architecture of a macromolecular complex is a critical step in understanding its molecular function. While recent technological advances have enabled atomic-resolution visualization of macromolecules by single particle electron microscopy (EM) (Bai et al., 2015), protein complexes that exhibit high degrees of structural or compositional heterogeneity are typically not amenable to high resolution studies. Single particle EM techniques can nonetheless provide important biological information at intermediate to low resolution, although assignment of protein subunit locations, or localization of flexible domains within a macromolecule can be ambiguous in this resolution range (Chowdhury et al., 2015; Jiang et al., 2013; Lander et al., 2012; Tsai et al., 2014). To overcome the issues associated with subunit identification in EM maps, a variety of molecular labeling strategies have been developed to locate regions of interest within complexes, although all have significant weaknesses (reviewed in Table 1 of reference (Oda and Kikkawa, 2013)).

Translational fusion of an identifiable protein label (such as maltose binding protein (MBP), or green fluorescent protein (GFP)) to the N- or C-terminus of a protein subunit is a common labeling strategy (Ciferri et al., 2012; Lander et al., 2013, 2012; Tsai et al., 2014; Wang et al., 2007), although this approach is best suited for identifying single-domain protein subunits whose termini do not extend far from the domain. Additionally, the labeled subunit must tolerate the genetic fusion of a large globular domain without disrupting normal folding, and without hindering incorporation of the subunit into the macromolecular complex. Internal insertion of GFP labels within target proteins has also been performed (Ciferri et al., 2012), but this requires insertion of a peptide linker, significantly altering the target's native sequence, leading to potential folding defects. Posttranslational labeling of natively assembled complexes is possible by attaching antibodies or Fabs (Samso and Koonce, 2004; Tsai et al., 2014), but versatility in epitope mapping by this method is limited to the number of available monoclonal antibodies for a given subunit, and is further complicated by the fact that antibodies vary significantly in binding affinity. Antibody labeling can also be prohibitively expensive due to the high cost of many antibodies. Tagging of specific biotinylated positions with streptavidin also offers a method for internal labeling of subunits, although this technique involves the insertion of a 15 amino-acid Avi tag into the polypeptide backbone at flexible solvent-exposed loops, requiring prior knowledge of target structure, and limiting the number of potential sites for localization (Lau et al., 2012). Furthermore, addition of this lengthy tag to an already flexible loop confounds precise subunit localization, due to a high degree of freedom of the streptavidin label. Labeling by conjugation of gold clusters to –SH, $-NH_2$, or His6 tags can increase labeling precision (Ackerson et al., 2010), although these methodologies suffer from low occupancy of gold labels. Visualization of gold labels using negative stain can also be challenging due to the comparatively strong scattering of the heavy metal ions used for staining. Identification of gold labels by negative stain often requires the use of large gold clusters and very thin stain (Buchel et al., 2001), which can introduce structural artifacts that may negatively impact image analysis.

* Corresponding author.
E-mail address: glander@scripps.edu (G.C. Lander).

http://dx.doi.org/10.1016/j.jsb.2015.09.010
1047-8477/

The field of EM is in desperate need of a site-specific, biocompatible strategy for robust, high occupancy labeling of proteins for identification of subunits within complexes. Here we disclose a technique that can be universally exploited to label proteins at any solvent-exposed, single amino acid location using a globular protein that is readily identifiable by simple negative stain EM imaging. Our strategy utilizes the specificity of a commercially available unnatural amino acid (UAA) for mutagenesis to target single residue positions in proteins for orthogonal bioconjugation to a chemically modified MBP. The technique is performed without the introduction of non-native peptide sequences or labeling "tags" that are required for any internal labeling technology developed to date. The technique involves a very biocompatible, 2-step conjugation reaction that is followed by simple purification steps, resulting in the enrichment of >90% labeled target protein while preserving the native structure.

2. Materials and methods

2.1. Generation of MBP_{Cys}, $Rpn5^{Y13\rightarrow TAG}$ and $Rpn5^{S26\rightarrow TAG}$ by site-directed mutagenesis

To make MBP_{Cys}, the MBP gene was amplified using the pYT7 vector as template DNA in a standard, 50 μL Q5 PCR (NEB), undergoing 35 cycles using the following primers (Integrated DNA Technologies (IDT)): Fwd: 5′-TATTATACTCGAGATGCATCATCATCATCATCATGG GGAAAA CCTGTACTTCCAGTCAAAATCGAAGAAGGTAAACTGGTA ATCTGG-3′ and Rev 5′-ATATATAACTAGTTTACTTGGTGATAC GAGTCTGCGCGTC-3′. During amplification, the Fwd primer was used for appending a 5′ XhoI site, followed by the ATG start codon and bases encoding a 6× His-tag and TEV cleavage site (immediately upstream of the encoded cysteine residue) to the N-terminus of MBP. The Rev primer imparts a 3′ SpeI site downstream of the TAA stop codon. The PCR was purified using a QIAquick PCR Purification Kit (Qiagen) and eluted in 40 μL H_2O. Following a 50 μL digestion with XhoI and SpeI (NEB) restriction enzymes (RE), the PCR product was gel purified using a PureLink Quick Gel Extraction Kit (Invitrogen). The pCDSStrc vector (Shoji et al., 2011) was also digested with XhoI and SpeI, and gel-purified in parallel with the PCR product. Ligation of RE-digested 6× His/TEV/MBP_{Cys} insert and pCDSStrc vector was performed using T4 DNA ligase (NEB) in a 1 h ligation reaction at room temperature. 2 μL of the ligation reaction was used in a 50 μL transformation into electrocompetent Top 10 *Escherichia coli* cells (Invitrogen). Following a 1 h recovery at 37 °C in 2× YT media (Amresco), shaking at 220 rpm, cells were plated on LB agar plates containing 50 μg/mL spectinomycin (G-Biosciences) for overnight selection. Surviving colonies were singly picked, and grown to saturation in 5 mL 2× YT media supplemented with 50 μg/mL spectinomycin, shaking at 220 rpm. Plasmid DNA was purified using a QIAprep Spin Miniprep Kit (Qiagen), and the 6× His-tagged MBP_{Cys} construct containing an N-terminal TEV cleavage site upstream of the introduced cysteine residue (herein referred to as pCDSStrc/MBP_{Cys}) was confirmed by sequencing (GeneWiz).

To remove natural amber codons from Rpn6 and Rpn9 genes, sub-clones were generated for site-directed mutagenesis. In brief, the ~12 kb pETDuet-1 plasmid (Lander et al., 2012) harboring Rpn5, Rpn6, Rpn8, Rpn9 and Rpn11 (referred to herein as lid vector 1 (LV1)) was cleaved with NotI and XhoI restriction enzymes (NEB), and the resulting fragment containing the Rpn5 and Rpn6 genes was cloned into a clean pETDuet-1 vector via NotI and XhoI restriction sites. The LV1 plasmid was also used for sub-cloning of the Rpn9-containing fragment into a clean pUC19 vector via BamHI (NEB). Both sub-clones were then subjected to site-directed mutagenesis in a standard, 50 μL Q5-based PCR as above, but undergoing 25 cycles, and using the following primers (IDT): Rpn6 Fwd 5′-GTGTCTTGTATTAAGGCCGGCCTAATACGACTCACTA TAGGG-3′ and Rpn6 Rev 5′-TATTAGGCCGGCCTTAATACAAGACAC TTGCCTTTTCAAATAG-3′ and Rpn9 Fwd 5′-CCATCTGGGTTTAA GAATTCTAATACGACTCACTATAGGGG-3′ and Rpn9 Rev 5′-CGTATTA GAATTCTTAAACCCAGATGGATTGGCCACGAGCTTC-3′ to generate $Rpn6^{TAG\rightarrow TAA}$ and $Rpn9^{TAG\rightarrow TAA}$, respectively. To make $Rpn5^{Y13\rightarrow TAG}$, the pETDuet/Rpn5-$Rpn6^{TAG\rightarrow TAA}$ sub-clone (sequence-verified; GeneWiz) was then subjected to site-directed mutagenesis in a standard, 50 μL Q5-based PCR as above, undergoing 25 cycles, and using the following primers (IDT): Fwd 5′-GGC TGACAAGGATTAGAGCCAAATTTTGAAGGAAGAGTTTCC-3′ and Rev 5′-CCTTCAAAATTTCCGTCTAATCCTTGTCAGCCTTAATTGGTGC-3′. To make $Rpn5^{S26\rightarrow TAG}$, the same sub-clone was used as template with following primers (IDT): Fwd 5′-TCCTAAGATCGATTAGCTCGCT CAAAATGATTGTAACTCTGC-3′ and Rev 5′-CATTTTGAGCGAGC TAATCGATCTTAGGAAACTCTTCCTTC-3′. Following all site-directed mutagenesis experiments, PCR products were purified using a QIAquick PCR Purification Kit (Qiagen), and template DNA was removed by incubation with Dpn1 (NEB) for 2 h at 37 °C. The Dpn1 digestion was purified using a QIAquick PCR Purification Kit, and eluted in 20 μL H_2O. 2 μL of the freshly purified DNA was used in 50 μL transformations into electrocompetent Top 10 *E. coli* cells as above. Following a recovery at 37 °C as described above, cells were plated on LB agar plates containing 100 μg/mL ampicillin (G-Biosciences) for overnight selection. Surviving colonies were singly picked, grown to saturation in 5 mL 2× YT media supplemented with 100 μg/mL ampicillin, and plasmid DNA was purified as described above for verification by sequencing. Sequencing reactions confirmed the presence of the amber (TAG) codon, replacing the natural tyrosine (TAT) codon at amino acid position 13 in $Rpn5^{Y13\rightarrow TAG}$ clones, and the natural serine codon (TCG) at position 26 in $Rpn5^{S26\rightarrow TAG}$ clones. The $Rpn9^{TAG\rightarrow TAA}$ sub-clone was also verified by sequencing, and was cloned back into the original LV1 parent vector via BamHI; the pETDuet/Rpn5-$Rpn6^{TAG\rightarrow TAA}$ plasmid was cut with NotI and XhoI, and this fragment was cloned back into the original LV1 parent vector (resulting in the generation of LV1 without amber codons in Rpn6 or Rpn9 (LV1-A)). LV1-A was used for all wild-type lid expression and purification. Sequence-verified plasmids were then used for cloning of the $Rpn5^{Y13\rightarrow TAG}$- or $Rpn5^{S26\rightarrow TAG}$-containing fragments (these fragments also contain $Rpn6^{TAG\rightarrow TAA}$) back into the pETDuet/$Rpn9^{TAG\rightarrow TAA}$ plasmid via NotI-XhoI sites (to generate LV1-B and LV1-C, respectively). LV1-B and LV1-C were verified by sequencing and were found to contain the TAA stop codon in both the Rpn6 and Rpn9 genes, as well as the previously verified amber (TAG) codons at Rpn5 Y13 and S26, respectively.

2.2. Protein expression and purification

Wild-type recombinant yeast proteasome lid complex was expressed and affinity purified from *E. coli* lysate as described previously[5], using anti-FLAG M2 resin (Sigma). Prior to expression of UAA-containing lid complex, a fourth compatible vector (pUltra) containing the unnatural aaRS and tRNA pair[14] for incorporation of the pAzF UAA was co-transformed with the three plasmids encoding all 9 proteins of the proteasome lid complex (pCOLADuet/Rpn3, Rpn7, Rpn12 and pACYCDuet/Hsp90, SEM1, ytRNAs, and the engineered LV1-B or LV1-C) into electrocompetent BL21(DE3) cells (Invitrogen). This *Methanococcus jannaschii*-derived tyrosyl aaRS/tRNA pair was originally evolved for orthogonal, site-specific encoding of *p*-cyanophenylalanine (pCNF) in *E. coli*, but was also found to be capable of efficiently incorporating a variety of *para* substituted tyrosine analogs when present in the media. In the absence of UAA in rich media, this pair will incorporate phenylalanine (although at much lower levels) in response to the amber codon

(Young et al., 2011). For expression of UAA-containing lid complex, cells were grown at 30 °C, shaking at 200 rpm in 2× YT media supplemented with 100 μg/mL ampicillin, 50 μg/mL kanamycin, 25 μg/mL chloramphenicol, 50 μg/mL spectinomycin, and 1 mM pAzF (Bachem) or *p*-acetylphenylalanine (pAcF) (SynChem), to test the incorporation of more than one tyrosyl analog. Upon induction with 1 mM IPTG at OD_{600} = 0.6, the temperature was dropped to 16 °C for 18 h. Cells expressing the UAA-containing lid complex were harvested and protein was purified in parallel with cells expressing wild-type lid complex via M2 anti-FLAG affinity resin as described above. Incorporation of both pAcF as well as pAzF, replacing the natural tyrosine at position 13 in Rpn5, were confirmed by nano-LC/MS/MS analysis of excised bands (from SDS gels corresponding to the Rpn5 protein from purified UAA-containing lid samples) at the Mass Spectrometry Proteomics Core Facility at TSRI, La Jolla (data not shown). Compared to yields obtained when expressing/purifying wild-type lid complex, yields of UAA-containing lid complex ranged from 70% to 90% (~2 mg/L for wild-type lid and ~1.4 mg/L for pAcF-containing lid and ~1.8 mg/L for pAzF-containing lid complex).

Top 10 *E. coli* cells harboring pCDSStrc/MBP_{Cys} were grown at 37 °C in 2× YT media supplemented with 50 μg/mL spectinomycin to OD_{600} = 0.4, and induced with 1 mM IPTG. Following induction, cells were grown at 37 °C, shaking at 220 rpm for 16 h prior to harvest. Cells were harvested by centrifugation for 10 min at 5000*g* using a Beckman Coulter Allegra X-12R centrifuge equipped with a SX4750 swinging bucket rotor, and re-suspended in lysis buffer (100 mM $NaPO_4$ pH = 7.0, 150 mM NaCl, 10 mM imidazole) supplemented with 1× Protease Arrest (G-Biosciences), and 2 mg/mL T4 lysozyme (Sigma). Sonication was performed using a Q125 sonicator (Qsonica) for 2 min (20 s pulses) at 50% amplitude. Lysate was cleared by spinning at 18,400 rpm for 30 min in a Beckman 70Ti rotor using a Beckman Optima LE-80 K ultracentrifuge. Cleared lysate was incubated with Ni-NTA agarose resin (Thermo Scientific) that was equilibrated in WashI buffer (100 mM $NaPO_4$ pH = 7.0, 150 mM NaCl, 20 mM imidazole) for 1 h at 4 °C. Incubated resin was collected in a 5 mL polypropylene column (Qiagen), and washed with one column volume of WashI buffer (100 mM $NaPO_4$ pH = 7.0, 150 mM NaCl, 20 mM imidazole), two column volumes of WashII buffer (WashI with 50 mM imidazole), and two column volumes of WashIII buffer (WashI with 75 mM imidazole). MBP_{Cys} protein was then eluted from Ni-NTA resin using Elution buffer (WashI with 300 mM imidazole), diluted to 1–2 mg/mL in WashI buffer with 1 mM DTT, but without imidazole, and immediately transferred to 3 K MWCO dialysis tubing equilibrated in TEV cleavage buffer (100 mM $NaPO_4$ pH = 7.0, 150 mM NaCl, 1 mM DTT). Eluted protein was analyzed by SDS gel (Supplementary Fig. 2, lane 1).

BL21(DE3) *E. coli* cells were co-transformed with pUltra/pCNF and pET101/$GFP^{Y151\rightarrow TAG}$ for expression and purification of the $GFP^{Y151\rightarrow pAzF}$ reporter protein. Cells were grown at 37 °C in 2× YT media supplemented with 100 μg/mL ampicillin, 50 μg/mL spectinomycin, and 1 mM pAzF, shaking at 220 rpm to OD_{600} = 0.4, and were induced with 1 mM IPTG. Cells were harvested and the 6× His-tagged $GFP^{Y151\rightarrow pAzF}$ reporter protein was purified in the same manner as MBP_{Cys} protein described above. For its use in reactions performed on resin, following subjection to WashIII buffer, 6× His-tagged $GFP^{Y151\rightarrow pAzF}$ reporter-bound Ni-NTA resin was washed two additional times in WashI buffer without imidazole prior to undergoing the conjugation reaction.

2.3. Generation of MBP_{DBCO} labeling reagent

An appropriate amount (~1 unit per 20 μg target protein) of acTEV protease (Invitrogen) was added directly to dialysis tubing containing the Ni-NTA-purified MBP_{Cys} protein described above. The TEV cleavage reaction was performed during removal of Imidazole from the sample by dialysis into 4 L of TEV cleavage buffer, while 1 mM DTT was introduced concurrently; dialysis was performed for 2 h at RT, then buffer was replaced with 4 L of fresh, TEV cleavage buffer at 4 °C, and the dialysis was allowed to continue at 4 °C overnight (gel analysis of the TEV cleavage reaction shown in Supplementary Fig. 2, lane 2). This procedure resulted in significantly more observable cleavage by TEV protease than using standard conditions recommended by the manufacturer. The sample was then removed from dialysis tubing, and TEV protease, along with the cleaved 6× His tag (and other contaminating proteins co-purified with MBP_{Cys}) were re-captured on Ni-NTA agarose equilibrated in TEV cleavage buffer without DTT, by incubating at RT for 30 min while turning. Resin was collected in a 2.5 mL PTFE 0.2 μm filter (Millipore) by spinning at 3000 rpm for 5 min at 4 °C. Proteins retained on Ni-NTA resin following re-capture were analyzed by gel (Supplementary Fig. 2, lane 3). Flow-through from Ni-NTA recapture contained only the cleaved MBP_{Cys} protein (Supplementary Fig. 2, lane 4).

To reduce di-sulfides, MBP_{Cys} protein was concentrated to 100 μM using a 10 K MWCO Amicon Ultracel filter (Millipore), and incubated with 5 mM TCEP for 30 min at RT. TCEP was then removed by buffer exchange (desalting) in a 5 mL, 7 K MWCO Zeba column. MBP_{Cys} was then concentrated to 200 μM, and mixed 1:1 with 10 mM DBCO-maleimide or DBCO-PEG_4-maleimide (Click Chem Tools) in 10% DMSO/H_2O. Maleimide-sulfhydryl coupling (final reaction concentrations: 100 μdfM MBP_{Cys}, and 5 mM DBCO-maleimide in 5% DMSO, 75 mM NaCl, 50 mM $NaPO_4$ pH = 7.0) was performed at 4 °C, for 2–4 h. The reaction appears opaque due to the low solubility of the maleimide-DBCO heterobifunctional cross-linker, however the opacity of the solution clears slightly as the reaction proceeds. Unreacted maleimide-DBCO reagent was removed by desalting as described above, while the sample was buffer exchanged into 100 mM $NaPO_4$ pH = 7.0, 150 mM NaCl. The extent of DBCO modification was determined by division of spectroscopic absorbance at 310 nm by absorbance at 280 nm of the conjugated sample (using a coefficient of 12,000 for DBCO, as per the manufacturer's instructions). Modification of MBP_{Cys} to MBP_{DBCO} was ~60% complete after 2 h at 4 °C, and ~70% complete in 2 h at RT. Reactions ran for 4 h at 4 °C, resulting in ~80% conversion of MBP_{Cys} to MBP_{DBCO}. Aliquots of MBP_{DBCO} reagent were then prepared for long-term storage by addition of 5% glycerol then flash-freezing in LN_2.

2.4. Copper-free click reactions

Conjugation of UAA-containing lid complex ($Rpn5^{Y13\rightarrow pAzF}$) to MBP_{DBCO} was allowed to proceed for 12 h at 4 °C, using 50 μM lid ($Rpn5^{Y13\rightarrow pAzF}$) and 250 μM MBP_{DBCO} in 150 mM NaCl, and 100 mM $NaPO_4$ pH = 7.0. Unreacted MBP_{DBCO} was removed by size exclusion chromatography (SEC) using a superose 6 (S6) column. Alternatively, unreacted MBP_{DBCO} could be removed from the sample by desalting using a 40 K MWCO column (Pierce), or by re-capture of lid complex following in-solution reactions on anti-FLAG M2 affinity resin. Following MBP_{DBCO} removal, the sample contained both labeled and unlabeled lid complex, as determined in 2D class averages obtained by negative stain EM (Fig. 2B).

Conjugation of the MBP_{DBCO} label to the $GFP^{Y151\rightarrow pAzF}$ reporter protein was performed under the same conditions as described above, but with the $GFP^{Y151\rightarrow pAzF}$ reporter protein still bound to its purification resin. This reaction (250 μM MBP_{DBCO} incubated with resin-bound $GFP^{Y151\rightarrow pAzF}$ reporter protein for 4 h while turning at 4 °C) was designed to under-label the GFP reporter in order to show the efficacy of the downstream amylose enrichment step.

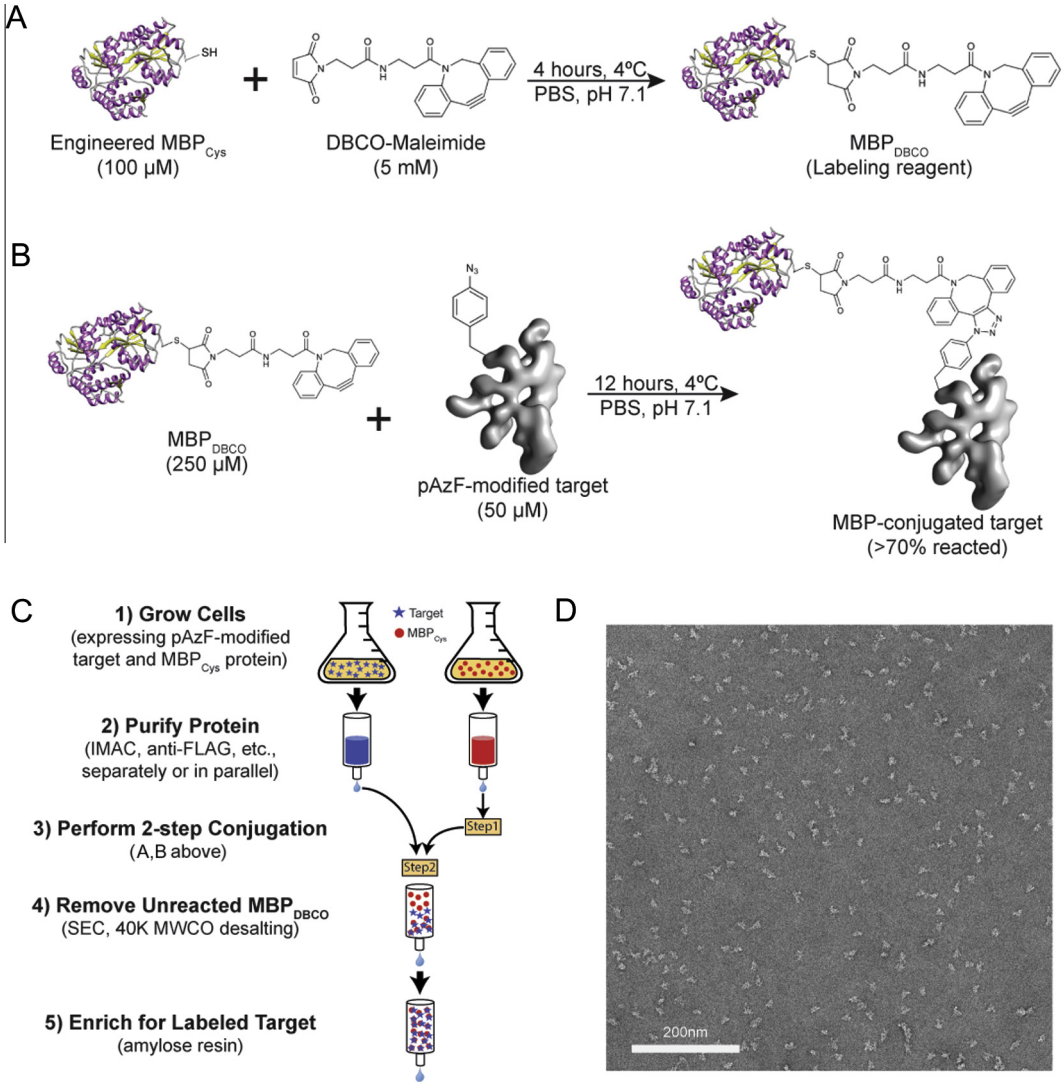

Fig. 1. 2-step reaction for site-specific conjugation of an MBP label. MBP was modified to contain an N-terminal 6X His-tag upstream of a TEV protease site (GNLYFQ/C), followed immediately downstream by an introduced Cys residue (MBP$_{Cys}$). (A) The cysteine sulfhydryl reacts with the maleimide moiety of a zero-length heterobifunctional crosslinker reagent containing a dibenzocyclooctyne at the opposing end. (B) Functionalized MBP (MBP$_{DBCO}$) is used in molar excess in a Cu^{2+} free 3 + 2 cycloaddition ("click") reaction with the UAA-modified target protein. Reaction conditions are included. In (C) we present a general workflow for the labeling of UAA-modified target proteins using commercially available reagents. (1) Cells are grown for expression and (2) purification of the UAA-modified target protein. Modification of MBP$_{Cys}$ to MBP$_{DBCO}$ can be performed off-line or in parallel with purification of the target protein, and in large quantities for long-term storage at −80 °C. The two-step labeling procedure in (3) can therefore be performed directly on resin or in solution, using a wide range of MBP$_{DBCO}$ concentrations. (4) Removal of unreacted MBP$_{DBCO}$ labeling reagent from the conjugation reaction can be accomplished by size exclusion chromatography (SEC), by using a 40,000 MWCO desalting column, or simply by washing resin following reactions performed directly on resin. (5) MBP-conjugated target can then be significantly enriched over unconjugated protein following incubation with amylose resin. The eluted conjugate is now ready for visualization by negative stain or cryo electron microscopy. (D) A representative micrograph containing negatively stained, MBP-conjugated lid complex, labeled at position Y13 in Rpn5.

MBP$_{DBCO}$ was removed from the reaction by washing of the GFP reporter on Ni-NTA resin following the reaction, and captured GFP was eluted using Elution buffer. Eluted GFP reporter protein from this "on-resin" reaction was analyzed by gel (Fig. 2, lane 3), along with the resulting conjugate species, which was enriched over unconjugated GFP$^{Y151\rightarrow pAzF}$ following re-capture on amylose resin (Fig. 2, lane 5).

2.5. Enrichment of labeled target using amylose resin

Regardless of the labeling efficiency observed using a variety of reaction conditions, labeled target protein can be enriched by incubation with amylose resin. This enrichment was exploited following removal of MBP$_{DBCO}$ reagent from the reactions described above. In brief, samples containing labeled and unlabeled target protein were diluted to between 0.1 and 0.2 mg/mL, and incubated with amylose resin (equilibrated in 150 mM NaCl, 100 mM $NaPO_4$) for 1 h at 4 °C. Unlabeled target protein was then removed in washes with the same buffer prior to elution of labeled target protein using 10 mM maltose.

2.6. Mass spectrometry

To confirm site-specific incorporation of UAAs at position 13 in the Rpn5 subunit, bands corresponding to this subunit were excised from coomassie-stained SDS gels that had been loaded with the purified lid complex expressed in the presence of pAcF, and submitted to the TSRI Center for Mass Spectrometry in La Jolla, CA. The gel band was destained, reduced (10 mM DTT), alkylated (55 mM iodoacetamide) and digested with trypsin overnight

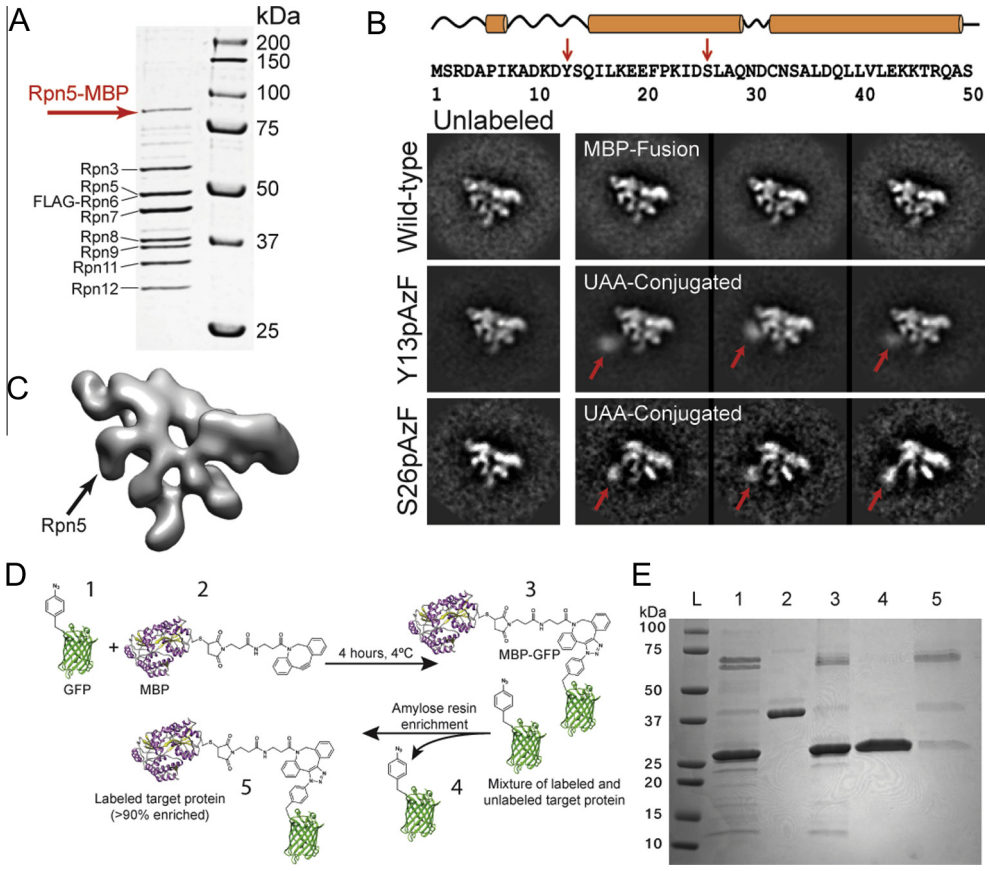

Fig. 2. Visualization of labeled target complex and enrichment of occupancy. The Rpn5 subunit of the wild-type yeast proteasome lid complex was modified by site-directed mutagenesis to contain the amber codon (TAG) at position Y13 ($Rpn5^{Y13\rightarrow TAG}$) and S26 ($Rpn5^{S26\rightarrow TAG}$). UAA-containing lid complex ($Rpn5^{Y13\rightarrow pAzF}$ or $Rpn5^{S26\rightarrow pAzF}$) was then generated by *in vivo* incorporation of the UAA via amber suppression. (A) SDS–PAGE of conjugated lid complex showing a ~90 kDa band corresponding to the Rpn5-MBP conjugate (indicated with a red arrow). This band was excised from the gel and its composition was analyzed by nano-LC/MS/MS (Supplementary Fig. 3). (B) Rpn5 secondary structure prediction (ssPRO4.0) places Y13 within a flexible loop, and S26 within an N-terminal alpha-helix. 2D class averages: the left images show unlabeled wild-type and unlabeled $Rpn5^{Y13\rightarrow pAzF}$ or $Rpn5^{S26\rightarrow pAzF}$ lid particles obtained by negative stain EM. Representative 2D classes from previous[5] attempts to label Rpn5 via N-terminal MBP fusion are shown to the right of the unlabeled wild-type 2D class average. Isolated in the same SEC fractions as the unlabeled lid complexes shown below to the left are: middle; reference-free 2D class averages of MBP-conjugated lid complex, labeled at amino acid position Y13, and bottom; position S26 in Rpn5 using the maleimide-DBCO reagent shown in Supplementary Fig. 1B. Red arrows indicate electron density corresponding to the MBP label. (C) 3D negative stain reconstruction (~20 Å resolution) of wild-type lid complex for visual orientation of Rpn5 subunit location within the lid complex. (D) Cartoon schematic of a conjugation reaction performed using the $GFP^{Y151\rightarrow pAzF}$ target reporter protein while bound to its affinity purification resin. Numbers in this panel correspond to the numbered lanes in (E) SDS–PAGE analysis of the on-resin conjugation of MBP_{DBCO} (42.5 kDa) to $GFP^{Y151\rightarrow pAzF}$ (26.9 kDa) with amylose enrichment. A conjugation reaction was performed immediately following capture of 6× His-tagged $GFP^{Y151\rightarrow pAzF}$ reporter protein on Ni-NTA agarose resin. Following standard washing (see Supplementary Methods), 5 μL of resin was used for elution and visualization by SDS gel (lane 1). For the "on-resin" reaction, 50 μL of 250 μM MBP_{DBCO} was added directly to 20 μL of saturated resin, and the reaction was allowed to proceed for 4 h while turning at 4 °C. Ni-NTA resin was then collected, allowing for separation of unreacted MBP_{DBCO} in the flow-through (lane 2) from a resin-bound mixture of conjugated and unconjugated $GFP^{Y151\rightarrow pAzF}$ reporter protein (eluted and visualized in lane 3). This conjugation reaction was <20% efficient, leaving >80% of the reporter unreacted. Enrichment for MBP-conjugated $GFP^{Y151\rightarrow pAzF}$ reporter was performed using amylose resin, added directly to diluted Ni-NTA eluate from the reaction. Amylose resin was collected, and flow-through contained unconjugated reporter (lane 4), enabling significant (>90%) enrichment of MBP-GFP conjugate (visualized in lane 5 following elution from amylose resin in 10 mM maltose).

before analysis by nano-LC/MS/MS. The raw data were compared against a custom sequence database containing the Rpn5 sequence, and Rpn5 was identified with 21 unique peptides and 45% sequence coverage. The MS/MS data were compared against the Rpn5 sequence for the possible incorporation of pAcF at tyrosine 13. Two peptides were seen and through the presence of b and y fragment ions in the MS/MS spectrum: 9–18 [ADKDYSQILK] and 12–18 [DYSQILK]; both confirmed 1 incorporation site of pAcF at Y13.

To confirm the presence of the Rpn5-MBP conjugate, the ~90 kDa band shown in Fig. 2A was excised and subjected to the same overnight trypsin digestion procedure as described above. Results from nano-LC/MS/MS analysis confirmed the presence of MBP and Rpn5 proteins in this band. MBP and Rpn5 peptides identified by this procedure are summarized in Supplementary Fig. 3.

2.7. Electron microscopy

For negative stain imaging, wild-type and MBP-conjugated lid complex were diluted to 50 nM in EM buffer (50 mM HEPES pH 7.6, 100 mM NaCl, 100 mM KCl, and 1 mM TCEP (Sigma)). A thin layer of carbon was applied to 400-mesh Cu–Rh maxtaform grids (Electron Microscopy Sciences) by chemical vapor deposition, and grids were subsequently exposed to a 95% Ar/5% O_2 plasma for 20 s to charge/activate the carbon surface. 4 μL of sample was applied, and wicked away prior to addition of 4 μL 2% uranyl formate. Images were acquired on a Tecnai Spirit LaB_6 electron microscope operating at 120 keV, with a random defocus range of −0.5 to −1.5 μm and an electron dose of 20 e-/$Å^2$ using the Leginon automated image acquisition software (Suloway et al., 2005). 331, 214 and 491 images of negatively stained wild-type lid, MBP-labeled lid at Rpn5 Y13 and Rpn5 S26, respectively, were

collected at a nominal magnification of 52,000× on an F416 CMOS 4K × 4K camera (TVIPS) with a pixel size of 2.05 Å/pixel at the sample level.

2.8. Image processing

All image preprocessing was performed using the Appion image-processing pipeline (Lander et al., 2009). The contrast transfer function (CTF) was estimated using CTFFIND3 (Mindell and Grigorieff, 2003) and only micrographs having a CTF confidence greater than 80% were used for processing. Particle picking was performed using the template-based FindEM software (Roseman, 2004). Particles were extracted with a box size of 160 pixels and pixel values that were 4.5 sigma above or below the mean were replaced with the mean intensity of the extracted particles. Multiple rounds of iterative particle alignment by iterative stable alignment and clustering (ISAC) (Yang et al., 2012) was used for 2D classification, and class averages containing wild-type, or MBP-labeled particles were selected for display in Fig. 2b . Each of the displayed images represents an average of anywhere between 16 and 34 individual negatively stained particles from the data collection.

2.9. 3D reconstruction of wild-type lid complex

Multiple rounds of reference-free alignment were used for 2D classification and alignment of particles, whereby class averages containing damaged, aggregated, or false particles, were removed. This resulted in a dataset containing 17,680 wild-type lid particles for 3D classification and 3D refinement, which was performed using RELION v1.31 (Scheres, 2012). The 3D reconstruction was resolved to 19.6 Å, according to a gold standard Fourier Shell Correlation at 0.143. Low resolution intensities were dampened using a SPIDER script in order to more clearly visualize domain features. UCSF Chimera (Goddard et al., 2007) was used for visualization of the 3D model, and for generation of Fig. 2c.

3. Results

We employ an *in vivo* method (Chatterjee et al., 2013) that makes use of an evolved, promiscuous aminoacyl tRNA synthetase (Young et al., 2011) to efficiently incorporate the UAA *p*-azidophenylalanine (pAzF) (Chin et al., 2002) (shown in Supplementary Fig. 1A) into a specific site within the peptide backbone of a target complex for downstream conjugation to the label. Fig. 1A depicts the first step of our two-step labeling strategy. The first step affords the functionalization of an engineered MBP (MBP_{Cys}, preparation outlined in Supplementary Fig. 2) with a dibenzocyclooctyne (DBCO) group. At this step, the introduced sulfhydryl in MBP_{Cys} is targeted for covalent attachment of the maleimide-containing heterobifunctional crosslinker reagent shown in Supplementary Fig. 1B. This reaction is well-characterized and can be performed with high efficacy at 4 °C in phosphate or HEPES buffer at physiological pH, allowing for structural preservation of the MBP_{Cys} label during incubation. Following a rapid desalting step to remove unreacted linker reagent, the functionalized MBP can be used in downstream labeling reactions with UAA-modified target proteins. The labeling reagent (MBP_{DBCO}) can be made in large quantities due to the commercial availability of the crosslinker reagent and high expression levels of MBP_{Cys} (>20 mg/L), and is stable for at least a week at 4 °C. Addition of 5% glycerol to the MBP_{DBCO} and flash-freezing in LN_2 allows for long term storage at −80 °C, enabling multiple labeling experiments on many different constructs using a single preparation of the label.

The second step of the labeling strategy is shown in Fig. 1B. In this step, the MBP_{DBCO} reagent is incubated in molar excess with the UAA-containing target protein. Expression and purification of the target proteins in this study are described in Section 2. The DBCO moiety provides a ring-strained alkyne that is suitable for orthogonal, Cu^{2+}-free click reactions via the azide moiety of the incorporated pAzF UAA. The 3 + 2 cycloaddition reaction results in formation of the triazole species shown in Fig. 1B, with moderate to high yield, depending upon the concentration of the reactants, as well as the temperature and duration of incubation. In this study, we used a concentration of 250 μM MBP_{DBCO} incubated with 50 μM target protein (recombinant yeast proteasome lid complex ($Rpn5^{Y13\rightarrow pAzF}$ or $Rpn5^{S26\rightarrow pAzF}$)) in a standard reaction time of 12 h at 4 °C, and observed >70% labeling in both cases (Cheng et al., 2005). Importantly, the target is purified as an intact protein complex prior to labeling, and therefore its assembly is not affected by incubation with the label. Additionally, the absence of Cu^{2+} in the click reaction increases the structural preservation of a broad range of potential target proteins when compared to reactions carried out in the presence of Cu^{2+}, or other established techniques for UAA-mediated bioconjugation using commercially available UAAs (Kim et al., 2012; Lang and Chin, 2014; Lukinavicius et al., 2013). Following the labeling reactions, unreacted MBP_{DBCO} reagent can be removed by size exclusion chromatography (SEC).

Here, we applied this methodology to the recombinant yeast 26S proteasome "lid" complex (Lander et al., 2012) to show its utility for molecular labeling and subunit identification. The proteasome lid, a 370 kDa complex composed of 9 subunits, was chosen for these labeling experiments due to the recalcitrance of one of its subunits (Rpn5) to identification using traditional MBP fusions (Lander et al., 2012) (Fig. 2B). Secondary structure prediction with ssPro4 (Cheng et al., 2005) was used to identify sites of Rpn5 that were contained within differing structural elements. Y13 and S26, predicted to reside within a flexible loop and an alpha-helix, respectively, were chosen for labeling. A 3D negative stain reconstruction of wild-type lid complex is shown in Fig. 2C, with a black arrow indicating the position of the Rpn5 subunit that was targeted for labeling. Using the described reactions (Fig. 1A and B), MBP was attached to these internal positions of Rpn5 (loop residue Y13, and helix residue S26) and imaged by negative stain EM. Single particle 2-dimensional image analysis readily shows the conjugated MBP label, as evidenced by the appearance of a globular density attached to Rpn5 (indicated by red arrows in 2D classes shown in Fig. 2B). Specific conjugation of the MBP_{DBCO} label to the Rpn5 subunit was confirmed by nano-LC/MS/MS of the excised band, indicated with a red arrow in Fig. 2A (peptides identified by nano-LC/MS/MS are shown in Supplementary Figs. 3a and b). These data clearly show the advantage of UAA labeling at specific positions within the polypeptide backbone, as the traditional labeling of the Rpn5 with an N-terminally fused MBP did not enable subunit identification due to the extreme flexibility of the resulting construct. Furthermore, MBP fusions to the C-terminus of Rpn5 were shown to disrupt lid assembly (Lander et al., 2012), an issue that is circumvented using the UAA labeling strategy, which allows for native assembly of the lid prior to labeling. Interestingly, despite using the same labeling strategy for both sites, the MBP label appears to be closer to the Rpn5 subunit in the $Rpn5^{S26\rightarrow pAzF}$ labeling experiment than in $Rpn5^{Y13\rightarrow pAzF}$ experiment (Fig. 2B). This difference may result from an increased flexibility of the MBP label when attached to a loop (Y13) as opposed to a rigid secondary structural element, such as an alpha-helix (S26).

SEC fractions corresponding to intact lid complex contained labeled and unlabeled species, as observed in 2D class averages (Fig. 2B). To remove unlabeled lid complex from the SEC fractions,

the covalently attached MBP was used as a "handle" to capture and enrich for the labeled species on amylose resin. To show the versatility of the labeling strategy and the efficacy of the amylose resin enrichment step, we used $GFP^{Y151\rightarrow TAG}$ as a model reporter protein (Young et al., 2010) in an "on-resin" conjugation reaction (Fig. 2D). This GFP reporter contains an amber stop codon at amino acid position 151, enabling pAzF incorporation by amber suppression (see Section 2). Here, the conjugation reaction was performed following standard washes associated with IMAC affinity purification of the GFP reporter, whereby 250 μM MBP_{DBCO} was added directly to the target while bound to its affinity resin. Regardless of the observed efficiency of the click reaction (the on-resin reaction was <20% complete after 4 h at 4 °C), we are able to enrich for labeled protein following incubation with amylose resin. This novel tandem affinity purification results in isolation of protein containing >90% occupancy of the label at the specified site, as determined by comparison of band intensities corresponding to labeled and unlabeled species following amylose resin enrichment (to remove unlabeled protein), shown in lane 5 of Fig. 2D. A diagram summarizing the entire workflow, including expression and purification of the UAA-containing target protein, preparation of the MBP_{DBCO} labeling reagent for on-resin or in-solution reactions, and enrichment of labeled product suitable for analysis by negative stain or cryoEM is shown in Fig. 1C.

4. Conclusion

Initial structural characterization of large macromolecular assemblies is a challenging task, and investigators are in need of a clear, straightforward strategy to label any specific site within the subunits that comprise the complex in order to accurately outline its architecture. Here, we present a site-specific labeling strategy that will allow unambiguous assignment of protein location and orientation within complexes by EM. The strategy presented here is most useful for 2D analysis of protein complexes, and this report establishes a foundation for further development of UAA-mediated labeling techniques for high-resolution applications. For example, the described reaction chemistry could be used to incorporate smaller and more precisely positioned labels that would be amenable for 3D analyses. Our technique is designed for its broad applicability (*in vivo* incorporation of UAAs has been performed in many organisms to date, reviewed in (Dumas et al., 2015)), and its gentle implementation, allowing for structural preservation of labeled targets. This labeling technique is extremely versatile and precise, requiring only a single, solvent exposed position within the target that is not limited to a flexible loops or to particular regions of the polypeptide. This methodology utilizes commercially available reagents, and allows for expansion to targets expressed by other organisms, as well as to a variety of chemically-modified labels. The disclosed strategy will be extended to include other UAAs exhibiting enhanced reaction rates (Lukinavicius et al., 2013) as they become widely available, and we expect for the tandem affinity purification described here to become useful for a variety of applications in the fields of molecular biology and bioengineering.

Author contributions

C.M.D. and G.C.L. conceptualized the study and designed experiments. C.M.D. carried out the experiments. C.M.D. and G.C.L. wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgments

We thank P. Schultz (The Scripps Research Institute) for providing vectors encoding the unnatural aaRS/tRNA pairs used in this study, and A. Martin (UC Berkeley) for providing expression vectors encoding the wild-type recombinant yeast proteasome lid complex. We thank F. Peters, D. Cayer, and J. Dambacher for discussion, S. Chowdhury, J. Sears, and M. Herzik for help with figure preparation, and L. Nosaka for technical assistance. This work was supported by the Damon Runyon Cancer Research Foundation (DFS-#07-13), the Pew Scholars program, the Searle Scholars program, and NIH Grant DP2 EB020402-01 to G.C.L.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2015.09.010.

References

Ackerson, C.J., Powell, R.D., Hainfeld, J.F., 2010. Site-specific biomolecule labeling with gold clusters. Methods Enzymol. 481, 195–230.
Bai, X.C., McMullan, G., Scheres, S.H., 2015. How cryo-EM is revolutionizing structural biology. Trends Biochem. Sci. 40, 49–57.
Buchel, C., Morris, E., Orlova, E., Barber, J., 2001. Localisation of the PsbH subunit in photosystem II: a new approach using labelling of His-tags with a Ni(2+)-NTA gold cluster and single particle analysis. J. Mol. Biol. 312, 371–379.
Chatterjee, A., Sun, S.B., Furman, J.L., Xiao, H., Schultz, P.G., 2013. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. Biochemistry 52, 1828–1837.
Cheng, J., Randall, A.Z., Sweredoski, M.J., Baldi, P., 2005. Scratch: a protein structure and structural feature prediction server. Nucleic Acids Res. 33, W72–W76.
Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., Schultz, P.G., 2002. Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. J. Am. Chem. Soc. 124, 9026–9027.
Chowdhury, S., Ketcham, S.A., Schroer, T.A., Lander, G.C., 2015. Structural organization of the dynein–dynactin complex bound to microtubules. Nat. Struct. Mol. Biol. 22, 345–347.
Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., Nogales, E., 2012. Molecular architecture of human polycomb repressive complex 2. eLife 1, e00005.
Dumas, A., Lercher, L., Spicer, C.D., Davis, B.G., 2015. Designing logical codon reassignment – expanding the chemistry in biology. Chem. Sci. 6, 50–69.
Goddard, T.D., Huang, C.C., Ferrin, T.E., 2007. Visualizing density maps with UCSF Chimera. J. Struct. Biol. 157, 281–287.
Jiang, J., Miracco, E.J., Hong, K., Eckert, B., Chan, H., Cash, D.D., Min, B., Zhou, Z.H., Collins, K., Feigon, J., 2013. The architecture of Tetrahymena telomerase holoenzyme. Nature 496, 187–192.
Kim, C.H., Axup, J.Y., Dubrovska, A., Kazane, S.A., Hutchins, B.A., Wold, E.D., Smider, V.V., Schultz, P.G., 2012. Synthesis of bispecific antibodies using genetically encoded unnatural amino acids. J. Am. Chem. Soc. 134, 9918–9921.
Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.W., Lyumkis, D., Potter, C.S., Carragher, B., 2009. Appion: an integrated, database-driven pipeline to facilitate EM image processing. J. Struct. Biol. 166, 95–102.
Lander, G.C., Estrin, E., Matyskiela, M.E., Bashore, C., Nogales, E., Martin, A., 2012. Complete subunit architecture of the proteasome regulatory particle. Nature 482, 186–191.
Lander, G.C., Martin, A., Nogales, E., 2013. The proteasome under the microscope: the regulatory particle in focus. Curr. Opin. Struct. Biol. 23, 243–251.
Lang, K., Chin, J.W., 2014. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. Chem. Rev. 114, 4764–4806.
Lau, P.W., Potter, C.S., Carragher, B., MacRae, I.J., 2012. DOLORS: versatile strategy for internal labeling and domain localization in electron microscopy. Structure 20, 1995–2002.
Lukinavicius, G., Umezawa, K., Olivier, N., Honigmann, A., Yang, G., Plass, T., Mueller, V., Reymond, L., Correa Jr., I.R., Luo, Z.G., Schultz, C., Lemke, E.A., Heppenstall, P., Eggeling, C., Manley, S., Johnsson, K., 2013. A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. Nat. Chem. 5, 132–139.
Mindell, J.A., Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347.
Oda, T., Kikkawa, M., 2013. Novel structural labeling method using cryo-electron tomography and biotin–streptavidin system. J. Struct. Biol. 183, 305–311.
Roseman, A.M., 2004. FindEM – a fast, efficient program for automatic selection of particles from electron micrographs. J. Struct. Biol. 145, 91–99.
Samso, M., Koonce, M.P., 2004. 25 Angstrom resolution structure of a cytoplasmic dynein motor reveals a seven-member planar ring. J. Mol. Biol. 340, 1059–1072.
Scheres, S.H., 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530.

Shoji, S., Dambacher, C.M., Shajani, Z., Williamson, J.R., Schultz, P.G., 2011. Systematic chromosomal deletion of bacterial ribosomal protein genes. J. Mol. Biol. 413, 751–761.

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., Carragher, B., 2005. Automated molecular microscopy: the new Leginon system. J. Struct. Biol. 151, 41–60.

Tsai, K.L., Tomomori-Sato, C., Sato, S., Conaway, R.C., Conaway, J.W., Asturias, F.J., 2014. Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. Cell 157, 1430–1444.

Wang, R., Chen, W., Cai, S., Zhang, J., Bolstad, J., Wagenknecht, T., Liu, Z., Chen, S.R., 2007. Localization of an NH(2)-terminal disease-causing mutation hot spot to the "clamp" region in the three-dimensional structure of the cardiac ryanodine receptor. J. Biol. Chem. 282, 17785–17793.

Yang, Z., Fang, J., Chittuluru, J., Asturias, F.J., Penczek, P.A., 2012. Iterative stable alignment and clustering of 2D transmission electron microscope images. Structure 20, 237–247.

Young, T.S., Ahmad, I., Yin, J.A., Schultz, P.G., 2010. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. J. Mol. Biol. 395, 361–374.

Young, D.D., Young, T.S., Jahnz, M., Ahmad, I., Spraggon, G., Schultz, P.G., 2011. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. Biochemistry 50, 1894–1900.