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