Journal of Structural Biology 192 (2015) 159–162

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Protein domain mapping by internal labeling and single particle electron microscopy

Claudio Ciferri [a,*], Gabriel C. Lander [b,1], Eva Nogales [a,b,c]

[a] Department of Molecular and Cell Biology, University of California, Berkeley, United States
[b] Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, United States
[c] Howard Hughes Medical Institute, UC Berkeley, Berkeley, United States

ARTICLE INFO

Article history:
Received 1 August 2015
Received in revised form 26 September 2015
Accepted 28 September 2015
Available online 30 September 2015

Keywords:
Protein labeling
GFP
Electron microscopy
Structure
Domain mapping

ABSTRACT

In recent years, electron microscopy (EM) and single particle analysis have emerged as essential tools for investigating the architecture of large biological complexes. When high resolution is achievable, crystal structure docking and *de-novo* modeling allows for precise assignment of individual protein domain sequences. However, the achievable resolution may limit the ability to do so, especially when small or flexible complexes are under study. In such cases, protein labeling has emerged as an important complementary tool to characterize domain architecture and elucidate functional mechanistic details. All labeling strategies proposed to date are either focused on the identification of the position of protein termini or require multi-step labeling strategies, potentially interfering with the final labeling efficiency. Here we describe a strategy for determining the position of internal protein domains within EM maps using a recombinant one-step labeling approach named Efficient Mapping by Internal Labeling (EMIL). EMIL takes advantage of the close spatial proximity of the GFP's N- and C-termini to generate protein chimeras containing an internal GFP at desired locations along the main protein chain. We apply this method to characterize the subunit domain localization of the human Polycomb Repressive Complex 2.

During the past several years, electron microscopy (EM) and single particle analysis have described the architecture and function of several macromolecular machineries (Nogales and Scheres, 2015). When high resolution is achievable, docking of available atomic coordinates or *de-novo* modeling of protein structures allow for precise assignment of individual components and localization of protein domains (Wiedenheft et al., 2011; He et al., 2013; Chang et al., 2015; Baskaran et al., 2014). When high resolution is not achievable or atomic coordinates are unavailable, additional structural information is required to describe both architectural and mechanistic details. This is particularly true for small or flexible macromolecular complexes. Different approaches have been developed to reconstitute and visualize protein complexes containing individual subunits labeled at specific sites. The majority of labeling studies utilize fusion protein tags, expressed in-frame at either the N- or the C-terminal region of the protein of interest. Successful applications of this technology include Maltose Binding Protein (MBP) (Lander et al., 2012; Ciferri et al., 2012; Baskaran et al., 2014), Green Fluorescence Protein (GFP) (Choy et al., 2009; Ciferri et al., 2012), Dynein Interacting Domain (DID) (Flemming et al., 2010) and actin polymer (Stroupe et al., 2009). While these approaches offer the advantage of reconstituting homogeneously labeled protein complexes, they are also best suited for the labeling of small subunits, where the localization of the N or C-termini matches reasonably well the position of the entire protein (Lander et al., 2012). In contrast, this subunit localization can be ambiguous if the N- and C-termini of the labeled protein are several nanometer away from each other or distant from important functional domains of interest. To overcome this limitation, other additional strategies have been adopted thus far. The first one makes use of monoclonal antibodies raised against specific protein domains (Hutchins et al., 2010; Chittuluru et al., 2011). While this technology has the potential of being very efficient, generating a complete set of monoclonal antibodies for each individual domain is often difficult and, when possible, low labeling efficiency or high-flexibility of the bound antibody could make the detection of the labeling challenging. A second strategy, termed DOLORS, utilizes monovalent streptavidin added post-translationally to an avi-tag sequence positioned within the main chain of the protein of interest (Lau et al., 2012). This method has the great advantage of specifically labeling any desired domain within a protein complex, without using costly

* Corresponding author at: Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, United States.
E-mail address: ciferric@gene.com (C. Ciferri).
[1] Present address: Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd, La Jolla, CA 92037, United States.

http://dx.doi.org/10.1016/j.jsb.2015.09.016
1047-8477/

and labor-intensive antibody production. However, a potential pitfall is represented by the multi-step process utilized for the labeling, which could diminish the overall labeling efficiency and tag-occupancy of the EM images.

In this manuscript, we present a strategy named Efficient Mapping by Internal Labeling (EMIL) to identify and localize internal domains within a multi subunit complex by electron microscopy. This method takes advantage of the close spatial proximity of the N- and C-termini of GFP (Supplementary Fig. 1A) and combines the advantages of fusion protein-based tags with the spatial resolution of the internal labeling. Previous work, utilizing similar concepts, has been used to characterize functional fusion proteins (Kratz et al., 1999; Roberts et al., 2009; Cockrell et al., 2011; Sun et al., 2014).

We designed vectors for *Escherichia coli*, insect cell and mammalian cell expression systems for the production of protein chimeras containing an internal GFP, connected through a short loop, to desired locations along the main protein chain (Supplementary Fig. 1B and C). GFP is a compact 27 kDa protein that can be easily visualized by electron microscopy when attached to the surface of a larger protein complex at defined location (Choy et al., 2009). For this reason, GFP can be inserted inside a polypeptide and serve as a marker for the identification of a specific domain within a protein complex. We use this method to characterize the domain organization of the Polycomb Repressive Complex 2 (PRC2) bound to AEBP2 (Ciferri et al., 2012). The results are presented here with a particular focus on the vector design and cloning strategy used to reconstitute different complexes carrying

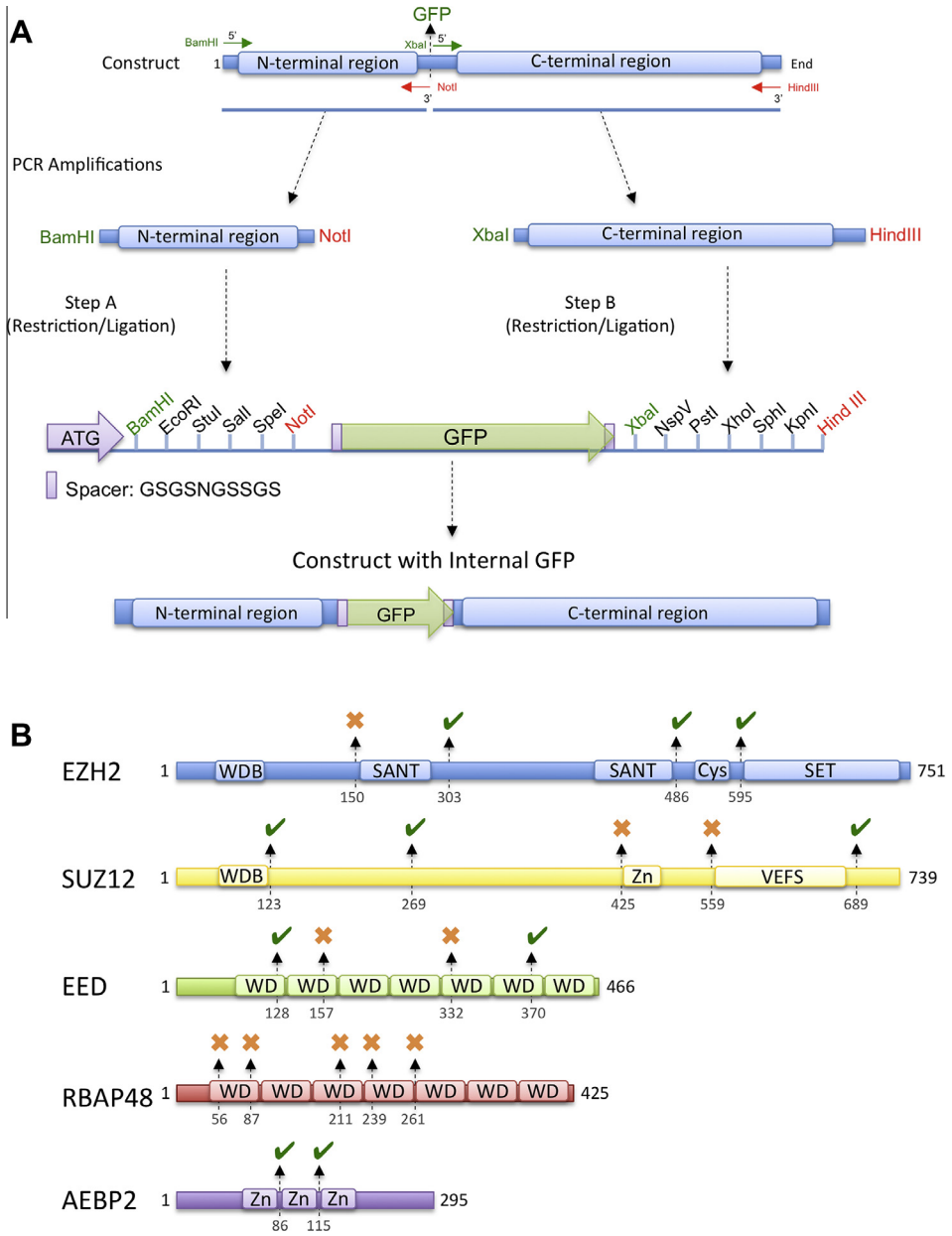

Fig. 1. EMIL tagging strategy and its utilization in the characterization of the PRC2 complex. (A) EMIL tagging cloning strategy. Protein chimeras, carrying GFP at desired locations, can be generated with a two-step cloning procedure. In the first step, the DNA preceding the desired point of GFP insertion is cloned into the first MCS. Successively, the remaining DNA sequence, following the site of GFP insertion, is cloned into the second MCS of a vector already containing the first insert. This procedure generates a chimera composed of the N-terminal portion of the protein of interest fused to GFP through two ten-residue antiparallel spacers, followed by its C-terminal region. (B) EMIL tagging applied to the domain characterization of the PRC2–AEBP2 complex. Black arrows and numbers indicate the position of the GFP insertions into the main chains. A green mark indicates insertions with successful expression used for domain localization. Red marks indicate chimeras that proved to be not amenable for structural studies.