Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

New trends and affinity tag designs for recombinant protein purification

David W Wood

Engineered purification tags can facilitate very efficient purification of recombinant proteins, resulting in high yields and purities in a few standard steps. Over the years, many different purification tags have been developed, including short peptides, epitopes, folded protein domains, non-chromatographic tags and more recently, compound multifunctional tags with optimized capabilities. Although classic proteases are still primarily used to remove the tags from target proteins, new self-cleaving methods are gaining traction as a highly convenient alternative. In this review, we discuss some of these emerging trends, and examine their potential impacts and remaining challenges in recombinant protein research.

Addresses
Department of Chemical & Biomolecular Engineering, 435 Koffolt Laboratories, 140 W Nineteenth Avenue, Columbus, OH 43210, United States

Corresponding author: Wood, David W (wood.750@osu.edu)

Current Opinion in Plant Biology 2014, **26**:54–61

This review comes from a themed issue on **New constructs and expressions of proteins**

Edited by **Junichi Takagi** and **Christopher G Tate**

For a complete overview see the Issue and the Editorial

Available online 12th June 2014

http://dx.doi.org/10.1016/j.sbi.2014.04.006

Purification tag methods

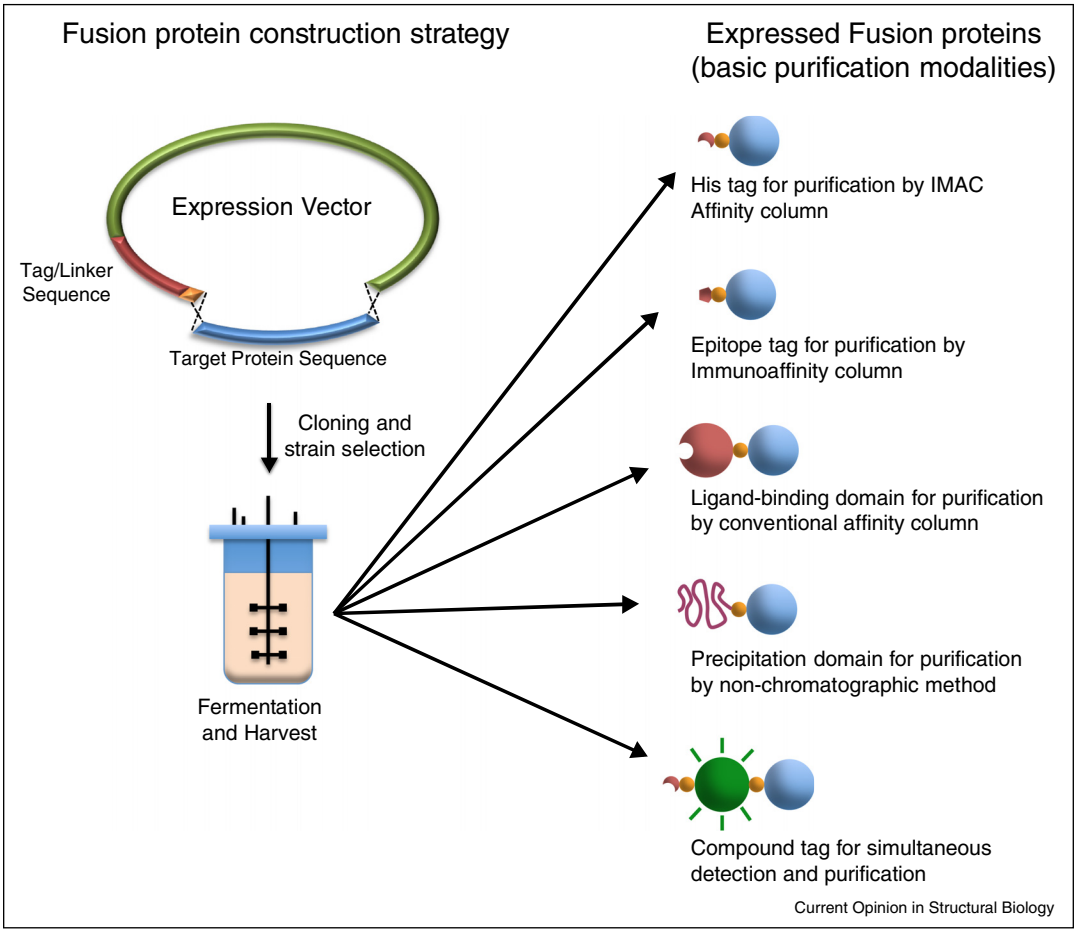
Although there are many applications related to gene fusions, perhaps the most common involves the addition of a purification 'tag', which provides a standardized method to purify the fused target protein [1••]. These tags are usually fused to the N-terminus or C-terminus of a target protein, and commonly allow their partner proteins to be selectively captured and purified through association with a tag-specific affinity resin, or by highly selective tag-dependent precipitation or aggregation (Figure 1). Many tags can also provide additional functions unrelated to purification, such as facilitating detection of the target protein or improving its solubility [2]. In cases where a native untagged protein is required, the tag can usually be removed by a variety of methods once the target is purified.

Despite the fact that tag-based methods have been used in laboratories throughout the world for decades, no universal tag has been identified that can be applied to any protein expressed in any host. In most cases, the identification of an optimal tag for a given host–target combination requires significant trial and error, and for some target proteins there are no effective tags. These limitations continue to drive the development of new options, and this research has led to ingenious and effective innovations over the past several years. Although a truly universal purification strategy is likely unrealistic, short-term objectives include increasing the variety of proteins that can be effectively purified with tag methods, while simultaneously facilitating the rapid identification of a suitable method for each new target. To accomplish these goals, researchers are increasingly using compound tags made up of multiple small domains, while continuing to develop single-domain tags with increased capabilities. Many new tags bind to highly inexpensive substrates, and are made even more economical through a variety of new approaches for tag removal. Taken together, these strategies can provide greater expression, increased purity and simplified detection for a wide range of fused products, along with inexpensive tag removal via simple self-cleaving or conventional proteolytic methods. A comprehensive discussion of even those tags developed in the last three years is far beyond the scope of this review, so I will highlight some interesting examples of each of these trends while providing a more comprehensive list of recently reported tags along with their basic advantages and recent references in Table 1. Notably, Table 1 lists many well-known conventional tags as well, which are included here due to their continued wide-spread use. Indeed, it is these tags that set the bar for newly developing methods in this field.

Popular tag methods

Undoubtedly the most commonly used purification tag in laboratories worldwide is the Polyhistidine tag, or His-tag [3]. The primary characteristics of this tag are that it is small, inexpensive to use, and it typically has minimal or no effect on the target protein structure or function [4]. The small size of the tag allows it to be trivially added to either terminus of a given target protein, and it does not require a specific fold in order to function, making it highly reliable in all major expression systems. Further, several expression vectors are commercially available for fusion to included His-tags, and anti-His antibodies are available for immunoassay detection. Importantly, the His-tag can function under native or denaturing conditions, allowing it to be

Figure 1



Schematic representation of basic cloning strategies and commonly used purification modalities. Although many novel tags have been developed, the majority of them fit one of the basic modalities shown at right. An important advance has been the construction of complex multidomain compound tags, with a simple example shown at lower right. These tags exhibit optimized combinations of behavior, and can provide highly effective methods for large groups of protein targets.

used in protein refolding protocols [5], and it has also become central to recent efforts on the purification of soluble membrane proteins that are stabilized by lipids or detergents [6–8]. Indeed, this unique combination of strength and reliability has made the His tag ubiquitous in research, where many laboratories simply clone new proteins into His tag vectors before any attempt is made to express or purify the native untagged product.

Despite these strengths, the His tag does suffer from a number of limitations. Perhaps the most significant is a tendency for contaminant proteins with external His residues to co-purify with His-tagged targets. The removal of these contaminants can require significant optimization [9], and indeed the *Escherichia coli* strain, LOBSTR (Low Background Strain), has been recently engineered to eliminate the most abundant contaminating host proteins [10•]. Further, in some instances the His tag can interfere with proper folding and activity of the target protein [11–13], and it has recently been reported to be incompatible with secretion in *Streptomyces* expression hosts [14]. In addition, the His-tag is generally ineffective in promoting proper folding of proteins with solubility issues, although it is routinely appended to a variety of solubility-enhancing domains.

Some of the problems associated with His tags have been solved through the use of small epitope tags, with the most widely recognized being arguably the FLAG, c-Myc and HA tags [15,16]. These tags are typically eight to twelve amino acids in length and are very strongly and specifically bind to their corresponding immunoaffinity resins [17]. The small size of these tags allows them to retain many of the advantages of the His tag, while providing superior purity and recovery of the fused target. These tags are also compatible with virtually any expression host and facilitate