Available online at www.sciencedirect.com
SCIENCE DIRECT
Protein Expression and Purification 48 (2006) 1–13
Protein Expression & Purification
www.elsevier.com/locate/yprep

Review

Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins

José Arnau *, Conni Lauritzen, Gitte E. Petersen, John Pedersen

Unizyme Laboratories A/S, Dr. Neergaards vej 17, DK-2970 Hørsholm, Denmark

Available online 28 December 2005

Abstract

Affinity tags are highly efficient tools for protein purification. They allow the purification of virtually any protein without any prior knowledge of its biochemical properties. The use of affinity tags has therefore become widespread in several areas of research e.g., high throughput expression studies aimed at finding a biological function to large numbers of yet uncharacterized proteins. In some cases, the presence of the affinity tag in the recombinant protein is unwanted or may represent a disadvantage for the projected application of the protein, like for clinical use. Therefore, an increasing number of approaches are available at present that are designed for the removal of the affinity tag from the recombinant protein. Most of these methods employ recombinant endoproteases that recognize a specific sequence. These process enzymes can subsequently be removed from the process by affinity purification, since they also include a tag. Here, a survey of the most common affinity tags and the current methods for tag removal is presented, with special emphasis on the removal of N-terminal histidine tags using TAGZyme, a system based on exopeptidase cleavage. In the quest to reduce the significant costs associated with protein purification at large scale, relevant aspects involved in the development of downstream processes for pharmaceutical protein production that incorporate a tag removal step are also discussed. A comparison of the yield of standard vs. affinity purification together with an example of tag removal using TAGZyme is also included.

Keywords: Affinity purification; Affinity tag; His-tag; IMAC; Recombinant protein; Protein purification; Tag removal; Endoproteases; Exopeptidases; Aminopeptidases; Therapeutic protein purification

With the modern advances in genomics, proteomics and bioinformatics, the number of proteins being produced using recombinant techniques is exponentially increasing. High throughput screening approaches are being performed to rapidly identify proteins with a potential application as therapeutic, diagnostic or industrial enzymes [1]. For this purpose, different expression hosts (e.g., *Escherichia coli*, *Saccharomyces cerevisiae*, *Pichia pastoris*, insect and mammalian cell lines) have been developed to express heterologous proteins [2–7]. Additionally, genomic approaches are being pursued to solve the structure of numerous proteins [8,9].

The above-mentioned screening approaches would not be feasible if specific purification procedures were to be developed for each individual protein. Here, the use of affinity tags enables different proteins to be purified using a common method as opposed to highly customized procedures used in conventional chromatographic purification.

When designing a downstream processing strategy for a protein, the inclusion of an affinity tag might be attractive for a number of additional reasons. In many cases, the protein candidate may exist as a version that includes an affinity tag from its early research stages where no biochemical characterization or functional assay is yet available. For structural studies, more than 60% of the proteins produced include a polyhistidine tag (his-tag, [10]). Additionally, the fact that affinity purification normally results in high yields—often over 90%—makes this

* Corresponding author.
E-mail address: ja@unizyme.dk (J. Arnau).

1046-5928/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2005.12.002

Table 1
Affinity and solubility tags for recombinant proteins

Tag	Size (aa)	Comments	References[a]
His-tag	5–15	Purification under native or denaturing conditions	[10,38,85,109]
FLAG	8	Calcium-dependent, mAb-based purification	[42,43,45]
Streptag II	8	Modified streptavidin, elution with biotin analog	[49,51,52,110]
HA-tag	9	Influenza virus hemagglutinin tag, Ab-based purification	[36]
Softag1, Softag 3	13, 8	Recognized by polyol-responsive mAb	[46–48]
c-myc	10	mAb-based purification	[31]
T7-tag	11–16	mAb-based purification	[56]
S-tag	15	S-protein resin affinity purification	[45]
Elastin-like peptides	18–320	Protein aggregation by temperature shift, intein used to remove tag	[64–66]
Chitin-binding domain	52	Binds only insoluble chitin (see intein, Table 3)	[68,111]
Thioredoxin	109	Affinity purification with modified resin	[17,26,89,112]
Xylanase 10A	163	Cellulose based capture, elution with glucose	[113,114]
Glutathione *S*-transferase	201	Glutathione or GST-Ab affinity	[9,26,87]
Maltose binding protein	396	Amylose affinity purification	[17,26,53,54,57,60]
NusA	495	Increased solubility in *E. coli*. Affinity tag needed for purification	[19]

[a] Only a few relevant references are included.

alternative economically favorable. Other economical and practical issues like the number of unit operations needed and the timesaving resulting from a reduction in chromatographic steps are also relevant aspects for consideration.

Introducing an affinity tag may have a positive effect in the biochemical properties of the target protein. A literature survey reveals that affinity tags have been observed to: (i) improve protein yield [11,12], (ii) prevent proteolysis [13], (iii) facilitate protein refolding [14], (iv) protect the antigenicity of the fusion protein [15], and (v) increase solubility [16–19]. Affinity tags have also been used to increase the sensitivity of binding assays for tagged ScFv [20].

On the other hand, adding a tag has also been reported to negatively affect the target protein resulting in e.g., (i) a change in protein conformation [21], (ii) lower protein yields [22], (iii) inhibition of enzyme activity [23,24], (iv) alteration in biological activity [25], (v) undesired flexibility in structural studies [26] and (vi) toxicity [27].

Due to the somehow unpredictable changes that adding a tag may introduce in a protein and its behavior, it is usually desirable to remove the tag. This reflects on the design of the protein fusion. Importantly, removal of the tag needs to be considered when designing a process for the production of a recombinant protein that is intended for human use to enable production of a 'native' (i.e., tagless) protein. And consequently, both the enzyme(s) used to cleave the tag and the cleaved fusion partner need to be removed from the purified protein.

Here, a review of affinity tags commonly used for recombinant protein production and the methods available for tag removal are discussed. A comparison of purification processes for a recombinant enzyme with and without affinity tag is also presented together with an overview of a downstream process that incorporates affinity purification and tag removal. Finally, an example of process for tag removal is presented for a his-tag thioredoxin (Trx).[1]

An overview of affinity tags and the design of the protein fusion

Recent reports have included several overviews of the currently available affinity tags for protein production and purification [28–32]. Nevertheless, since the choice of affinity tag and the method for tag removal are mutually dependent, an introduction to affinity tags is given herein for clarity.

Affinity tags can be defined as exogenous amino acid (aa) sequences with a high affinity for a specific biological or chemical ligand. A major group of affinity tags consists of a peptide or protein that binds a small ligand linked on a solid support (e.g., his-tags bind to immobilized metals, discussed below). Another group includes tags that bind to an immobilized protein partner such as an antibody or antibody purification using protein A affinity chromatography ([29], Table 1). The protein A-based methodology, used for e.g., purification of monoclonal antibodies (mAb), has been extensively reviewed elsewhere [33,34] and will not be discussed here.

His-tags are the most widely used affinity tags. Purification of his-tagged proteins is based on the use of chelated

[1] *Abbreviations used:* Trx, thioredoxin; 6×his, a stretch containing six consecutive histidine residues; aa, amino acid(s); AAP, *Aeromonas proteolytica* aminopeptidase; APM, aminopeptidase M; *B.*, *Bacillus*; CPA, carboxypeptidase A; CPB, carboxypeptidase B; DAPase, recombinant dipeptidyl peptidase, part of TAGZyme; *E.*, *Escherichia*; ELP, elastin-like polypeptides; FMN, flavin mononucleotide; FP, green fluorescent protein; GST, glutathione *S*-transferase; His-tag, a polyhistidine tag; IMAC, immobilized metal–ion affinity chromatography; mAb, monoclonal antibodies; MBP, maltose-binding protein; NMR, nuclear magnetic resonance; pGAP, recombinant pyroglutamyl aminopeptidase; pGAPase, an engineered version of recombinant pGAP used in TAGZyme; PHB, polyhydroxybutyrate; Qcyclase, recombinant glutamine cyclotransferase, part of TAGZyme; ScFv, single chain antibodies.