Analytical Biochemistry 504 (2016) 30–37

Contents lists available at ScienceDirect

Analytical Biochemistry

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

A dual protease approach for expression and affinity purification of recombinant proteins

Sreejith Raran-Kurussi, David S. Waugh*

Protein Engineering Section, Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA

ARTICLE INFO

Article history:
Received 8 February 2016
Received in revised form
1 April 2016
Accepted 7 April 2016
Available online 19 April 2016

Keywords:
Maltose-binding protein
Tobacco etch virus protease
Rhinovirus 3C protease
His-tag
Affinity chromatography

ABSTRACT

We describe a new method for affinity purification of recombinant proteins using a dual protease protocol. *Escherichia coli* maltose binding protein (MBP) is employed as an N-terminal tag to increase the yield and solubility of its fusion partners. The MBP moiety is then removed by rhinovirus 3C protease, prior to purification, to yield an N-terminally His_6-tagged protein. Proteins that are only temporarily rendered soluble by fusing them to MBP are readily identified at this stage because they will precipitate after the MBP tag is removed by 3C protease. The remaining soluble His_6-tagged protein, if any, is subsequently purified by immobilized metal affinity chromatography (IMAC). Finally, the N-terminal His_6 tag is removed by His_6-tagged tobacco etch virus (TEV) protease to yield the native recombinant protein, and the His_6-tagged contaminants are removed by adsorption during a second round of IMAC, leaving only the untagged recombinant protein in the column effluent. The generic strategy described here saves time and effort by removing insoluble aggregates at an early stage in the process while also reducing the tendency of MBP to "stick" to its fusion partners during affinity purification.

Published by Elsevier Inc.

Rapid advances in genomics and proteomics during the past three decades have revolutionized the fields of biotechnology and human medicine, particularly when recombinant DNA technology joined hands with structural biology. Currently, samples of proteins for structural and functional studies are routinely obtained by bioengineering [1]. Even so, protein purification remains the principal bottleneck. Conventional methods of protein purification have been almost completely supplanted by affinity-based methods that employ protein or peptide affinity tags [2–4]. The popularity of these affinity-based methods can be attributed to their generic nature in comparison with traditional approaches, which are rather protein specific. Other purification platforms such as ion exchange, hydrophobic interaction, and size exclusion chromatography are used as auxiliary steps to further enhance the purity of the sample if necessary.

In our laboratory and many others, a dual His_6–MBP (maltose-binding protein) tag is used in an initial immobilized metal ion affinity chromatography (IMAC) step [5–7]. The reason for employing the dual tag has been discussed in detail elsewhere [8]. Briefly, MBP enhances the solubility and improves the yield of its fusion partners during overproduction but is not a particularly effective affinity tag for protein purification. Hence, the His_6 tag is included to allow affinity chromatography by IMAC. The N-terminal His_6–MBP tags are subsequently removed by tobacco etch virus (TEV) protease to generate a tag-free protein. Although in general this approach has been very successful, it is not without its problems. For instance, a significant fraction of aggregation-prone proteins that are rendered soluble by fusing them to MBP subsequently precipitate when the fusion proteins are cleaved by TEV protease [9,10]. These proteins presumably either are not properly folded or exist as soluble aggregates in partially folded forms. Usually it cannot be ascertained whether or not this will be a problem until after affinity purification. A second potential pitfall is the tendency of some proteins to stick to MBP after TEV digestion, making it difficult to separate them from each other. The interaction between MBP and its fusion partners may be related to the mechanism of solubility enhancement [11]. In these situations, it is usually necessary to employ an MBPTrap (or

Abbreviations used: MBP, maltose-binding protein; IMAC, immobilized metal ion affinity chromatography; TEV, tobacco etch virus; PCR, polymerase chain reaction; ORF, open reading frame; ChikV, Chikungunya virus; DHFR, dihydrofolate reductase; DUSP14, dual-specificity phosphatase 14; MERS-CoV $3CL^{pro}$C148A, Middle East respiratory syndrome coronavirus 3C-like protease; GFP, green fluorescent protein; RBS, ribosome-binding site; IPTG, isopropyl β-D-1-thiogalactopyranoside; SDS–PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TVMV, tobacco vein mottling virus.

* Corresponding author.
E-mail address: waughd@mail.nih.gov (D.S. Waugh).

http://dx.doi.org/10.1016/j.ab.2016.04.006
0003-2697/Published by Elsevier Inc.

amylose) column, or another IMAC step to remove the MBP, thereby making the purification process more labor intensive. In the current study, we attempted to circumvent both of these problems by using a dual protease approach to achieve sequential removal of the MBP and His_6 tags.

Materials and methods

Materials

pBAD24–sfGFPx1 was a gift from Sankar Adhya and Francisco Malagon (Addgene plasmid no. 51558) [12]. The 3C protease expression vector pET/3C was a gift from Ari Geerlof (EMBL, Heidelberg, Germany). The pBLN200–GFPmut2–Car9 plasmid was a gift from François Baneyx [13].

All chemicals of the highest available purity were purchased from Sigma–Aldrich (St. Louis, MO, USA), American Bioanalytical (Natick, MA, USA), Thermo Fisher Scientific (Rockford, IL, USA), Roche Diagnostics (Indianapolis, IN, USA), or EMD Millipore (Billerica, MA, USA) unless otherwise stated. Restriction endonucleases were obtained from New England Biolabs (Ipswich, MA, USA). Fast-Link T4 DNA Ligase was purchased from Epicentre (Madison, WI, USA). All polymerase chain reactions (PCRs) were carried out using either the PfuUltra II Fusion HS DNA Polymerase (Agilent Technologies, Santa Clara, CA, USA) or the Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific).

Construction of pSRK2721

The Gateway destination vector pSRK2721 (see Fig. S1 in online supplementary material) was constructed as follows. First, a *Sna*BI restriction site was inserted at the end of the open reading frame (ORF) encoding MBP in pKM596 [14] using the QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies) with primers PE-2852 and PE-2853 (Table 1) to generate the plasmid pSRK2704. In the next step, the C terminus of MBP in pSRK2704 was further extended to include an in-frame 3C protease recognition site followed by a His_6 tag by PCR restriction cloning using PE-42, PE-2886, and PE-2887 primers (Table 1). The PE-42 primer was designed to anneal just proximal to the *Bgl*II site in the *malE* gene, whereas PE-2886 and PE-2887 were partially overlapping reverse primers that were employed to engineer the desired insertions (3C site and His_6 tag) followed by a *Sna*BI restriction site. The PCR amplicon, consisting of a C-terminal portion of the MBP ORF from pKM596 joined in-frame to a sequence encoding GGGLEVLFQ/GPHHHHHHYV, was digested with *Bgl*II and *Sna*BI and then cloned between the same sites in pSRK2704 to yield the destination vector pSRK2721 (Table 2). The "YV" residues shown above are cloning artifacts derived from the *Sna*BI site. The rhinovirus 3C protease recognition site is underlined, and the cleavage site is marked by a forward slash (see above).

Construction of entry clones

The sfGFP ORF was amplified from pBAD24–sfGFPx1 [12] by a single PCR using primers PE-277 (Forward-1), PE-2829 (Forward-2), and PE-2830 (Reverse) (Table 1). The PCR product contained the N- and C-terminal attB recombination sites and a TEV protease recognition site appended to the N terminus of the sfGFP ORF. The PCR amplicon was used in the Gateway BP reaction with pDONR221 (Thermo Fisher Scientific) to generate the entry clone pSRK2661 (Table 2). Similarly, the Chikungunya virus (ChikV) protease ORF was amplified from a ChikV cDNA clone using the primers PE-277 (Forward-1), PE-2645 (Forward-2) and PE-2646 (Reverse) (Table 1). The PCR amplicon was used in the Gateway BP reaction with pDONR221 to generate the entry clone pKK2483 (Table 2). Additional entry clones encoding *Francisella tularensis* IglC, human dihydrofolate reductase (DHFR), human dual specificity phosphatase 14 (DUSP14), the human SUMO-conjugating enzyme Ubc9, catalytically inactive Middle East respiratory syndrome coronavirus 3C-like protease (MERS-CoV $3CL^{pro}$C148A), and wild-type green fluorescent protein (GFP) were constructed as described previously [8,15–17].

Construction of fusion protein expression vectors

The destination vector (pSRK2721) was recombined in a Gateway LR reaction (Thermo Fisher Scientific) with various entry clones to generate the dual protease expression vectors for further study. All of the expression vectors are listed in Table 2. A schematic representation of the MBP fusion proteins they produce is shown in Fig. 1A. The second protease recognition site corresponds to that of TEV protease, which was incorporated into the entry clones.

Construction of an IPTG-inducible, untagged 3C protease expression vector

Gateway recombinational cloning (Thermo Fisher Scientific) was used to make an untagged rhinovirus 3C protease expression vector for co-lysis experiments. The 3C protease ORF was amplified by PCR from pET/3C using primers PE-2856 and PE-2857 (Table 1). These primers incorporated the attB recombination sites along with an appropriately positioned ribosome-binding site (RBS) in the PCR product. Subsequently, the PCR product was used in a BP reaction with pDONR221 (Thermo Fisher Scientific) to generate the entry clone pSRK2703. The entry clone was recombined in an LR reaction with pDEST42 (Thermo Fisher Scientific) to generate the

Table 1
Primer sequences.

Primer	Sequence (5′–3′)
PE-2852	GAA AGA CGC GCA GAC TAA TTC GTA CGT AAT CAC AAG TTT GTA CAA AAA AGC
PE-2853	GCT TTT TTG TAC AAA CTT GTG ATT ACG TAC GAA TTA GTC TGC GCG TCT TTC
PE-42	GGC ACA CGA CCG CTT TGG TGG CTA C
PE-2886	ATG CAT TAC GTA GTG ATG GTG ATG GTG ATG CGG ACC CTG GAA CAG AAC TTC
PE-2887	ACC CTG GAA CAG AAC TTC CAG ACC ACC ACC CGA ATT AGT CTG CGC GTC TTT C
PE-277	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC GGA GAA CCT GTA CTT CCA G
PE-2829	GAG AAC CTG TAC TTC CAG ATG CGT AAA GGC GAA GAG
PE-2830	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT ATT TGT ACA GTT CAT CCA TAC
PE-2645	GGC TCG GAG AAC CTG TAC TTC CAG AGT AAC GCA TTC CAA AAC AAA GCC AAC GTT TGT
PE-2646	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT ATC CTA CAA AGG CTG CAT TCA GTT GAT
PE-2856	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT TAA GAA GGA GAT ATA CAT ATG GGA CCA AAC ACA GAA TTT G
PE-2857	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT ATT GTT TCT CTA CAA AAT A
pSAMRFFwd	CAA AGG ATC TTC TTG AGA TCC TGG CTT CTG TTT CTA TCA GCT GTC C
pSAMRFRev	CGT GAG CAT CCT CTC TCG TTT CGC GGG GCA TGA CTA ACA TGA GAA