

# Purification of Recombinant Proteins by Chemical Removal of the Affinity Tag

CAROLE RAIS-BEGHDADI,<sup>1</sup> MARIO A. ROGGERO,<sup>2</sup>  
NICOLAS FASEL,<sup>2,\*</sup> AND CHRISTOPHE D. REYMOND<sup>2</sup>

<sup>1</sup>Institut de Biochimie, Université de Lausanne, ch. des Boveresses 155, 1066 Epalinges, Switzerland; and <sup>2</sup>Institut de Biologie cellulaire et de Morphologie, Université de Lausanne, Rue du Bugnon 9, 1005 Lausanne, Switzerland  
E-mail: Nicolas.Fasel@ib.unil.ch

Received October 2, 1997; Revised February 11, 1998; Accepted February 23, 1998

## ABSTRACT

The efficient removal of a N- or C-terminal purification tag from a fusion protein is necessary to obtain a protein in a pure and active form, ready for use in human or animal medicine. Current techniques based on enzymatic cleavage are expensive and result in the presence of additional amino acids at either end of the proteins, as well as contaminating proteases in the preparation. Here we evaluate an alternative method to the one-step affinity/protease purification process for large-scale purification. It is based upon the cyanogen bromide (CNBr) cleavage at a single methionine placed in between a histidine tag and a *Plasmodium falciparum* antigen. The C-terminal segment of the circumsporozoite polypeptide was expressed as a fusion protein with a histidine tag in *Escherichia coli* purified by Ni-NAT agarose column chromatography and subsequently cleaved by CNBr to obtain a polypeptide without any extraneous amino acids derived from the cleavage site or from the affinity purification tag. Thus, a recombinant protein is produced without the need for further purification, demonstrating that CNBr cleavage is a precise, efficient, and low-cost alternative to enzymatic digestion, and can be applied to large-scale preparations of recombinant proteins.

**Index Entries:** Recombinant polypeptide; polypeptide purification; affinity chromatography; CNBr cleavage.

\*Author to whom all correspondence and reprint requests should be addressed.

## INTRODUCTION

Recombinant proteins have been produced in various organisms. Bacteria, yeast, insect and mammalian cells, and viral systems are among the most commonly used systems today. Optimal expression is often obtained by compromising the choice of host cells with the expression system, while evaluating advantages and drawbacks. However, in all cases, the production at low cost of relevant proteins depends not only on the expression system, but also on the required purification step(s). Indeed, reducing the number of downstream processing steps is essential to render the commercial use of recombinant proteins more attractive. Recombinant proteins have been purified based on either their chemical properties or by immunoaffinity using specific antibodies. New purification procedures have been based on the addition of a peptidic tag to these proteins allowing affinity or immunoaffinity purification. For example, glutathion-S-transferase- (1) or histidine tags (2) have allowed successful one-step purifications of recombinant proteins. In these cases, the purified products have been used directly, or the tags have been removed prior to use. In the latter case, cleavage sites for thrombin (1), enterokinase (3), or Factor X (4) have been inserted in between the tag and the desired polypeptide, which then enables proteolytic cleavage and purification on an affinity column to remove the tag. Although recombinant proteins are purified rapidly using this strategy, the end product usually contains several extra amino acids and is contaminated by the proteolytic enzyme. Additional purification steps could be necessary. Furthermore, these extra amino acids on the polypeptide of interest as well as the presence of the proteolytic enzyme could influence the activity and/or possibly trigger unwanted immune responses in humans or animals, rendering the recombinant protein inadequate for in vivo studies. Furthermore, these purification steps involving enzymatic cleavage steps are expensive and, thus, inappropriate for large-scale production of proteins.

All these considerations taken together prompted us to investigate a simple, rapid, and efficient procedure for the large-scale purification of recombinant polypeptides. Cyanogen bromide (CNBr), under precise conditions, has been known for its ability to cleave polypeptides specifically at methionine and has been used specifically for the cleavage of two proinsulin molecules expressed in *Escherichia coli* (5). We have first performed purification by immobilized metal ion-affinity chromatography (IMAC) (6), which is based on the presence of a histidine tag. It was followed by CNBr cleavage of the tag at a methionine placed between the tag and the polypeptide to be purified. In this article, we present the purification of a C-terminal segment of the circumsporozoite protein (CSP) of the malaria agent, *Plasmodium falciparum* produced in *E. coli*.

## MATERIALS AND METHODS

### Reagents

All restriction and modification enzymes used were purchased from Boehringer Mannheim (Mannheim, Germany) or Gibco BRL (Gaithersburg, MD). Oligonucleotides were obtained from MWG-Biotech (GmbH, Ebersberg, Germany). Bacterial media and culture conditions have been previously described (7).

### DNA Construct

To obtain expression of the C-terminal segment (103 amino acids) of *P. falciparum* CSP, we used the vector pET9a (Promega, Madison, WI). This vector contains a T7 RNA polymerase gene, which is under the control of the LacUv5 promoter. Such a promoter can be induced by isopropyl-thio- $\beta$ -D-galactoside (IPTG) and, thus, controls the expression of the T7 RNA polymerase. To obtain T7-regulated expression of the region coding for the C-terminal segment of CSP (*P. falciparum* strain NF-54) (8), we amplified the segment of the CSP gene corresponding to amino acid 282–383 by PCR using specific oligonucleotides. The 5'-amplimer hybridized to 38 nucleotides of the CSP gene (8). Upstream of the CSP sequence, the 5'-amplimer contained the *NheI* restriction site and six nucleotides (ATG-GAA), coding for methionine and aspartic acid. The 3'-amplimer hybridized to 23 nucleotides of the CSP gene which enabled the insertion of an in-frame UAA and *SacI* restriction site into the gene. This 3'-amplimer also allowed us to change methionine 379 (8). After amplification and digestion with *NheI* and *SacI*, the DNA fragment was inserted into the pET9a vector. This region was completely sequenced in order to confirm the presence of the correct open reading frame. As expected, the open reading frame contained the initiation codon, the six histidine residues, the methionine residue at position 15, two valine residues at position 28 and 114, and a stop codon. Thus, this open reading frame could code for a polypeptide of 118 amino acids in its complete form and of 103 amino acids polypeptide after cleavage of the affinity tag (Fig. 1A).

### Cell Lysis

A single colony containing the expression plasmid was used to inoculate 2.5 L of LB broth medium containing 50  $\mu$ g/mL of kanamycin. The culture was grown at 37°C with vigorous shaking until the  $A_{600}$  reached 0.8. IPTG was added to a final concentration of 1 mM, and the culture was incubated for an additional 3 h at 37°C. Cells were harvested by centrifugation at 4000 rpm for 15 min, resuspended in 10 mL 0.1 M  $\text{Na}_2\text{HPO}_4$ , 0.01

**A**

1 | 40

MRGSHHHHHH **GMASMEKNNQ** GNGQGHN**VPN** DPNNRV**DENA**

NANSAVKNNN NEEPSDKHIK EYLNKI**QNSL** STEWSPCS**VT**

CGNGIQ**VR**IK PGSANKPKDE LDYANDIE**KK** ICK**VE**KCS

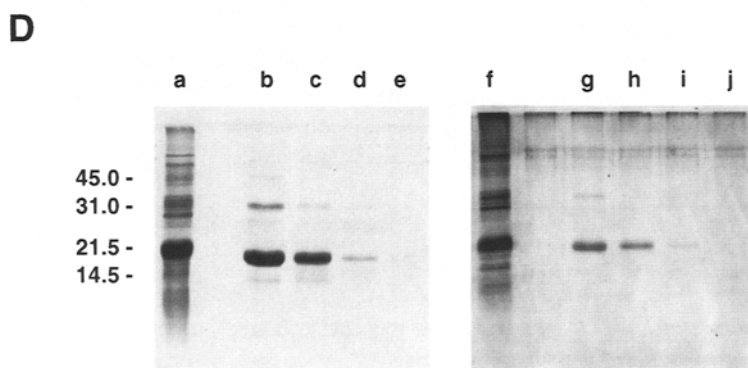
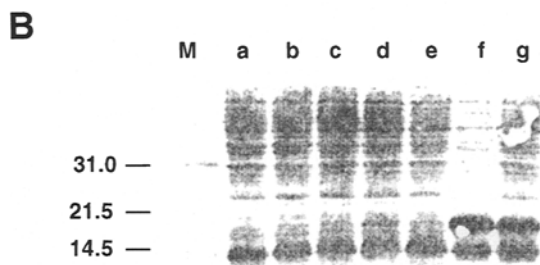


Fig 1. (A) Amino acid sequence of the *P. falciparum* C-terminal segment of CSP (PfCS C-ter) fused to a purification tag for expression in *E. coli*. The methionine residues (positions 12 and 15) present in the purification tag sequence and the two valine residues (positions 28 and 105) replacing the methionines of the CSP are shown in bold. (B) Analysis of different clones expressing PfCS C-ter. Protein lysates were separated on SDS PAGE 15%, transferred onto nitro cellulose, and stained with Ponceau red. Total lysate of an *E. coli* colony transformed with the control vector (pET9a) is shown in lane a. Total lysates of seven independent bacterial colonies transformed with PfCS C-ter construct are shown in lanes b–g. The molecular-mass markers 14.4, 21.5 and 31.0 kDa (Bio-Rad, Hercules, CA, low range) are indicated. (C) Immunodetection of the PfCS C-ter in the *E. coli* clones shown in (B). Immunodetection was performed using the S5III serum (lanes a–g) at a dilution of 1:2000. The molecular-mass markers 14.4, 21.5, and 31.0 kDa (Bio-Rad, low range) are indicated. (D) Purification of the PfCS C-ter on the Ni-NTA agarose resin. Total lysate of an *E. coli* colony transformed with the PfCS C-ter vector (lane a) and fractions eluted at 10 mM (lane b), 30 mM (lane c), and 50 mM (lane d) were analyzed by SDS-PAGE. Coomassie blue (lanes a–d) and silver nitrate (lane e–h) stainings are shown. The molecular-mass markers 14.4, 21.5, 31.0, and 45.0 kDa (Bio-Rad, low range) are indicated.

M Tris, pH adjusted to 8.0 with  $\text{H}_3\text{PO}_4$ , and lysed in a French press. Urea was added to 8 M (final concentration) and the obtained solution centrifuged at 9780 g for 15 min. The supernatant was collected, and  $\beta$ -mercaptoethanol added to a final concentration of 10 mM.

## IMAC and CNBr Cleavage

The cell lysate was incubated in batch for 1 h at room temperature with 25 mL Ni-NTA agarose resin (Qiagen Inc., Chatsworth, CA, USA). The resin had been previously equilibrated with buffer A (8 M urea, 0.1 M  $\text{Na}_2\text{HPO}_4$ , 0.01 M Tris, pH adjusted to 8.0 with  $\text{H}_3\text{PO}_4$ ). The column was prepared as follows: lysate-coupled resin was poured at a rate of 60 mL/h flow rate, washed with 120 mL of buffer A, and then the material was eluted (flow rate 60 mL/h) with buffer B (8 M urea, 0.1 M  $\text{Na}_2\text{HPO}_4$ , 0.01 M Tris, pH adjusted to 6.3 with  $\text{H}_3\text{PO}_4$ ) containing increasing concentrations of imidazole (10, 30, 50, 250 mM).

The eluted material was desalted by a Sephadex G25 column (50  $\times$  2.5 cm using 50% acetic acid/ $\text{H}_2\text{O}$  as mobile phase), lyophilized, and treated for 8 h at room temperature at a concentration of 20 mg/mL in 70% TFA using a 100-fold molar excess of CNBr. The digested material was lyophilized, solubilized in buffer A, and loaded again on the Ni-NTA agarose column. The flow through of the column was desalted by a Sephadex G25 column (50  $\times$  2.5 cm using 50% acetic acid/ $\text{H}_2\text{O}$  as mobile phase) and lyophilized. Approximately 200 mg of over 95% pure CSP (as estimated by SDS-PAGE) was recovered from the 2.5-L culture.

## Mass Spectrometry

MALDI-TOF mass analysis was performed using a time-of-flight mass spectrometer Voyager-DE (PerSeptive Biosystem, Framingham, MA) equipped with a nitrogen laser ( $\lambda = 337$  nm) to desorb and ionize the sample. The accelerating voltage used was 30 kV. The sample (1  $\mu\text{L}$ ; 1 mg/mL) was mixed with 24  $\mu\text{L}$  of a saturated solution of *trans*-3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) (Sigma, St. Louis, MI) in 1:2 acetonitrile/1.0% TFA, and 1.0  $\mu\text{L}$  of this solution was placed into a well of the Voyager sample plate and air-dried at room temperature. Ion masses were assigned based on an external mass calibration using two points that bracketed the mass range of interest. Relative signal abundances in the MALDI spectra are not directly comparable to the relative abundances of the species owing to potential differences in sensitivities of different substances.

## RESULTS AND DISCUSSION

### Expression of the C-terminal Segment of CSP in *E. coli*

Using a commercially available prokaryotic expression system, we expressed the C-terminal segment of the CSP of *P. falciparum* (PfCS C-ter) in the *E. coli* DE3 bacterial strain using the vector pET9a. The open reading frame codes for a polypeptide of 118 amino acids in its complete form, and 103 amino acids after cleavage of the affinity tag by CNBr (Fig. 1A). A polypeptide of approx 19 kDa is detectable in different bacterial colonies after SDS-PAGE separation, blotting onto nitrocellulose, and Ponceau red staining of whole cell lysates (Fig. 1B). As exemplified, the 19-kDa polypeptide is present at very various levels in the different clones. In certain clones, the presence of the 19-kDa polypeptide is barely detectable (lanes b and c), whereas in others, it is one of the predominant polypeptides (lanes f and g). It was also confirmed as a CSP-containing polypeptide by immunodetection using an antiserum directed specifically against the C-terminal segment of the CSP protein (Fig. 1C). No signal was observed in *E. coli* DE3 transformed with the plasmid alone (Fig. 1C, lane a).

### Purification by IMAC of PfCS C-ter Polypeptide

The presence of initiation of translation sequences and a histidine tag allows a one-step purification of the fusion protein by IMAC. Bacterial cell lysates were loaded onto a Ni-NTA agarose resin. After washing, the fusion protein was eluted in the presence of urea and increasing imidazole concentrations. The material was analyzed by SDS-PAGE Coomassie blue and silver nitrate stainings. The starting material (lanes a and f) and the material purified (lanes b–e and g–j) at 10, 30, 50, and 250 mM imidazole have been separated on SDS gels, and their pattern is compared (Fig. 1D). As evidenced in the figure, (Fig. 1D), elution was most efficient at 10 and 30 mM imidazole, unlike conventional elution profiles, which use up to 250 mM imidazole. Such a discrepancy in the elution profiles is probably the result of the influence of the amino acid sequence of the polypeptide of interest on the affinity of the histidine tail for the Ni-NTA agarose column. The identity of the material was verified by immunoblotting using an anti-CSP antibody (data not shown). Additional bands correspond to degraded products.

### CNBr Cleavage and Analytical Characterization of Purified PfCS C-ter

In order to cleave only the protein of interest at a specific site, we inserted a single methionine residue next to the PfCS C-ter coding region (Fig. 1A). This insertion provides a CNBr cleavage site between the tag and

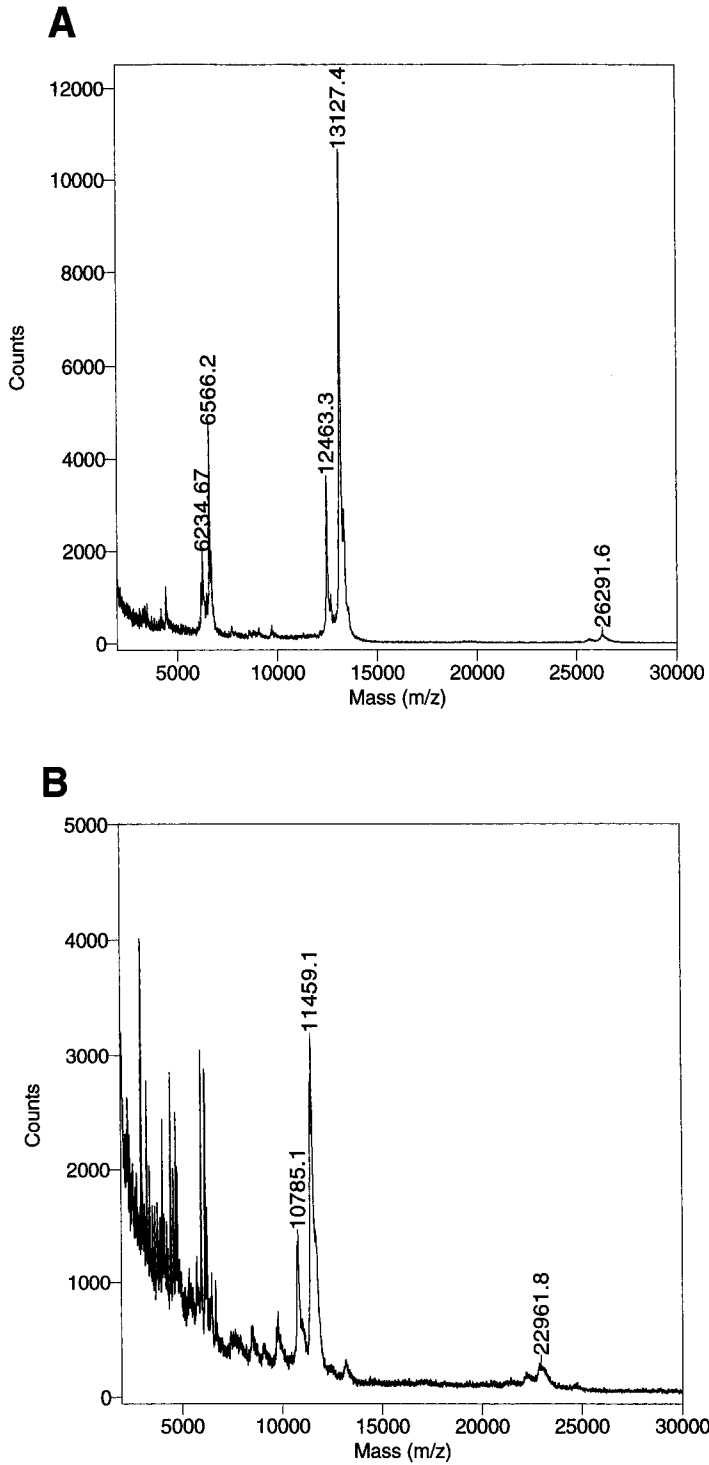


Fig. 2. Mass spectra of PfCS C-ter polypeptide before CNBr treatment and after CNBr treatment.

the CSP. Although other chemical cleavage methods can be used, a high yield of protein is obtained when 20- to 100-fold molar excess of CNBr is used. After IMAC purification, CNBr cleavage of CSP yielded approx 95% pure CSP C-ter based on the integration of the peak areas at 214 nm.

The histidine tag was removed by a second IMAC purification, and the product analyzed by chromatography and mass spectrography (Fig. 2). A mol/wt of 13'127.1 was determined by mass spectrography analysis before CNBr and of 11'459.4 after CNBr treatment. The difference in molecular weight between these two proteins corresponds exactly to the molecular weight of the amino acids composing the cleaved sequence, demonstrating the specificity and high yield of cleavage at the methionine residue interposed between the affinity tag and the desired polypeptide.

Although methionine is among the less-frequent residues and is often removed from the N-terminus of proteins *in vivo*, its presence within the amino acid sequence of the polypeptide will result in cleavage at this site and could limit the use of this method. This limitation can be circumvented by replacing internal methionines by alanines, valines, or isoleucines, which are known, in most instances, to be conservative changes and not to alter drastically the properties of the recombinant polypeptide. We replaced two such methionines from the C-terminal CSP sequence by valines, using site-directed mutagenesis. It should be noticed that internal cleavage sites also represent a major problem in enzyme cleavage procedures, like thrombin, sometimes precluding the isolation of the protein. Furthermore, it should be pointed out that cystein may be slowly oxidized and aspartic acid-proline bonds cleaved after prolonged CNBr exposure. Finally, excess reagent can cause degradation of Trp and Tyr side chains. However, using the conditions we describe, these modifications should be minor.

The chemical removal of the purification tag, as exemplified by the mass spectrum profiles, could be applied to other amino acids. As an alternative to Met/X cleavage, it could be mentioned that Trp/X cleavage is also possible in glacial acetic acid and 9 M HCl in the presence of DMSO followed by a CNBr treatment (9). In this case, methionine is not cleaved, but is oxidized to methionine sulfoxide. These residues can be then reduced by overnight incubation at 37°C with 10%  $\beta$ -mercaptoethanol to obtain original methionine (9).

The combination of methionine insertion and CNBr cleavage was recently described for long synthetic peptide (10). It is potentially applicable to any recombinant protein production system ranging from viruses to eukaryotic expression systems. It is also applicable to tags that can alter the chemical properties of the protein, such as hydrophobic tags, polyethylene tags, and so forth, circumventing the limitation of using defined



peptidic sequence. Considering the high level of expression that is now obtained in various systems and the simple affinity purification protocols available, one key step in the production of a recombinant protein is removal of tag elements to obtain the desired polypeptide precisely. We have shown here such specific and high yield expression and recovery of a recombinant polypeptide using a usual affinity chromatography and a simple chemical cleavage reaction. Over 95% of the recombinant polypeptide was devoid of the purification tag sequence. This low-cost and straightforward approach may easily be used for a variety of recombinant proteins.

## ACKNOWLEDGMENTS

We acknowledge C. Desponds for technical expertise, G. Corradin for helpful discussion, and S. Belli for careful reading of the manuscript. This work was supported by Grant No. 3218.1 of the Commission for Technology and Innovation (CTI) of the Swiss government to N. F. and C. R. and by RMF Dictagene.

## REFERENCES

1. Guan, K. and Dixon, J. E. (1991), *Anal. Biochem.* **192**, 262–267.
2. Lindeberg, G., Bennich, H. and Engstrom, A. (1991), *Int. J. Peptide Prot.* **38**, 253–259.
3. Collins-Racie, L. A., McColgan, J. M., Grant, K. L., DiBlasio-Smith, E. A., McCoy, J. M., and LaVallie, E. R. (1995), *Bio/technology* **13**, 982–987.
4. Nagai, K. and Thøgersen, H. C. (1984), *Nature* **309**, 810–811.
5. Shen, S. H. (1984), *Proc. Natl. Acad. Sci., USA.* **81**, 4627–4632.
6. Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975), *Nature* **258**, 598–599.
7. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989), *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
8. Caspers, P., Gentz, R., Matile, H., Pink, J. R., and Siniglia, F. (1989), *Mol. Biochem. Parasitol.* **35**, 185–190.
9. Huang, H. V., Bond M. W., Hunkapillar, M. W., and Hood, L. E. (1983), *Methyl Enzymol.* **91**, 324–332.
10. Roggero, M. A., Servis, C., and Corradin, G. (1997), *FEBS Lett.* **408**, 285–288.