Increased peptide deformylase activity for $N$-formylmethionine processing of proteins overexpressed in *Escherichia coli*: application to homogeneous rubredoxin production

Jianzhong Tang, Griselda Hernández, and David M. LeMaster*

Wadsworth Center, New York State Department of Health and Department of Biomedical Sciences, University at Albany—SUNY, Empire State Plaza, P.O. Box 509, Albany, NY 12201-0509, USA

Received 27 January 2004, and in revised form 8 March 2004

Abstract

Deformylation of the initiator $N$-formylmethionine does not always proceed to completion for proteins overexpressed in *Escherichia coli*. To overcome this limitation, the *def* gene encoding the *Escherichia coli* peptide deformylase was cloned into the plysS plasmid under the tetracycline (Tc) promoter control. The efficiency of this constitutive level of peptide deformylase expression was demonstrated for the case of the rubredoxins from both mesophilic and hyperthermophilic organisms which normally retain a majority of their $N$-formyl terminal form. Indicating the potential structural/functional significance of residual formylation, the presence of a highly solvent exposed $N$-formyl group in rubredoxin is discernable in the amide NMR chemical shifts for the active site metal-coordinating cysteines more than 21 Å away.

© 2004 Elsevier Inc. All rights reserved.

Keywords: $N$-formylmethionine; Peptide deformylase; *N*-terminal processing; Rubredoxin

$N$-formylmethionine is used to initiate protein synthesis in eubacteria, mitochondria, and chloroplasts. With rare exceptions such as the Tsr chemotaxis receptor protein [1], the $N$-formyl group is removed by the peptide deformylase encoded by the *def* gene in *Escherichia coli* [2,3]. Depending primarily on the identity of the second residue, the $N$-terminal methionine may also be removed. In such cases, it has been shown that the deformylation step must precede methionine processing [4]. Although deformylation is generally highly efficient for the normal expression levels of the endogenous proteins of *E. coli*, high levels of induction can result in partial retention of the formyl group. Examples include both the α- and β-subunits of *E. coli* tryptophan synthase [5,6], bovine somatotropin [7], human [8], and eel [9] growth hormone, *E. coli* 1-acyl-sn-glycerol-3-phosphate acyltransferase [10], human granulocyte colony-stimulating factor [11], bovine fatty acid-binding protein [12], bovine cytochrome P450 [13], and human S100β [14].

Incubation in 0.6 M HCl at 25 °C provides a useful means of protein deformylation preceding sequence analysis by Edman degradation [15]. However, this approach is of limited generality when biological activity is to be retained. In vivo suppression of the peptide deformylase activity can be achieved by growth of *E. coli* in the presence of high levels of trimethoprim in rich media which serve to provide the other required metabolites arising from $N^{10}$-formyl-tetrahydrofolate [3]. This approach has been combined with overexpression of the methionine aminopeptidase in *E. coli* to generate an archaear histone with over 85% fully processed protein for which normal heterologous expression yielded 50% retention of the $N$-formyl group [16].

Bovine somatotropin possessing a phenylalanine residue at the second position retains the $N$-terminal methionine as well as 5% of the initial $N$-formylmethionine form when expressed in *E. coli* [17]. To achieve a high degree of sample homogeneity suitable for pharmacological use, and perhaps to circumvent the limitations of the trimethoprim growth conditions as well, these Monsanto researchers turned to overexpression of
the peptide deformylase. The def gene was subcloned into their expression plasmid carrying the bovine somatotropin gene with both genes under an inducible trp promoter. The coexpression of these two genes resulted in synthesis of bovine somatotropin containing less than 0.5% of residual N-formylation. As a general approach to N-terminal processing, several aspects of this def coexpression system are potentially amenable to enhanced efficiency. By placing the def gene under the same promoter control as the target gene, the peptide deformylase is likely to be produced in substantial excess of that needed for adequate N-terminal processing of the target protein. On the other hand, at the start of induction the peptide deformylase will be present in only the basal level produced from the chromosomal copy of the gene, so that the initially produced target protein may be inadequately processed before it assumes its native conformation and ceases to be an efficient substrate for deformylation [18]. Finally, combining def and the target gene into the same operon reduces the flexibility of the expression system. Each of these concerns can be readily addressed by development of a constitutive production of the peptide deformylase using a two plasmid expression system.

Rubredoxin from the hyperthermophilic archaeon *Pyrococcus furiosus* (*Pf*) is believed to be the most thermostable protein characterized to date [19]. As such it has proven to have been a valuable model system for probing the structural basis of protein thermal stability. Many of the biophysical studies of *Pf* rubredoxin have been carried out on protein produced from its natural host, demanding specialized fermentation facilities [20]. The gene for *Pf* rubredoxin has been expressed in *E. coli* at a high expression level using a codon optimized sequence [21]. Unfortunately, when expressed in *E. coli* three N-terminal forms are generated in significant amounts. Only ~15% of the product corresponds to the native N-Ala terminal form, while another ~15% retain the initial N-terminal methionine and the remaining 70% retain the formylmethionine terminus. The three *Pf* rubredoxin forms exhibit modest differences in thermal stability [22]. Preparative chromatographic separation of the N-terminal forms of *Pf* rubredoxin is challenging [20,21]. Rubredoxin has provided the model system for development of a two plasmid peptide deformylase expression system.

Materials and methods

Materials

The pLysS plasmid and the BL21(DE3) *E. coli* expression strain were from Novagen. Restriction enzymes were obtained from New England Biolabs. Site-directed mutagenesis reagents were obtained from Stratagene. All other chemicals were of analytical grade.

Cloning the *E. coli* chromosomal def gene into the pLysS expression plasmid

Designing of both primers flanking the *E. coli* def gene was based on the published sequence [2,23]. The forward primer (5'-GGGAATTGTTAACAAGAAGG AGATATACATATGGCTTTTCAAGGTGTTAC) includes the def gene sequence upstream of the initiator codon. In addition, a *HpaI* restriction site is appended for the subsequent cloning step. The reverse primer (5'-TGGGAATTCGTTAACTTAAGCCGGGCTTTCA GACGATCCAG) includes a stop codon for the def gene followed by a second *HpaI* restriction site linker. The def coding sequence was PCR amplified from BL21(DE3) genomic DNA. Cells boiled in water for 5 min were used directly as template. The PCR profile was: denature at 94 °C for 4 min, followed by 30 cycles of 94 °C (30 s), 55 °C (1 min), and 72 °C (30 s). The products were then extended at 72 °C for 10 min. The 500 basepair PCR product was gel purified, digested with *HpaI*, and blunt end ligated to the pLysS plasmid at the *EcoRV* site. The plasmid carrying the def gene orientation under the control of the Tc promoter was designated as pSdef1.

Construction of the A2K variant of Pf rubredoxin

Codon-optimized *Pf* rubredoxin gene in the plasmid pT7-7 [21] was changed from alanine to lysine in the second residue by site-directed mutagenesis (Stratagene). The forward primer (5'-GGAGATATACATATGAAA AAATGGGTGTGCAAATCTCGGG) and reverse primer (5'-CCGCGATTTTGGCAAACTTCTCGG) are highlighted at the modified codon. The mutagenesis reaction was performed for 18 cycles in Perkin-Elmer Gene Amp PCR System (model 9600) according to the manufacturer’s recommended procedure.

N-terminal analysis

The iron form of *Pf* A2K rubredoxin was expressed and purified as previously described [24] with the substitution of 10 mg FeSO₄/L for ZnCl₂ in the growth medium. Following dialysis against 25 mM NH₄HCO₃ and lyophilization, the protein was redissolved in distilled water, loaded onto a Vydac C₄ column (4.6 x 250 mm), and eluted with a 0–100% gradient of acetonitrile with 0.1% trifluoroacetic acid. Peak fractions were dried, redissolved in water, and applied to a Finnigan LCQ Deca MS equipped with an ion trap for mass determination.
NMR analysis

The Zn$^{2+}$ form of rubredoxin from *Clostridium pasteurianum* was expressed and purified as previously described [24] with the variation of 2 g/L of U–2H glycerol instead of unlabeled glucose as carbon source. The NMR samples were equilibrated by dialysis against 100 mM NaCl, 20 mM NaH$_2$PO$_4$, pH 6.00, with 7% D$_2$O. Spectra were collected on a Bruker DRX 500 at 25 °C. TROSY-based experiments [25,26] were used for the 2D correlation spectra. The assignments were confirmed by a 2D version of an $^{15}$N–$^{15}$N–$^1$H HMQC–NOESY–TROSY experiment [27].

Results and discussion

**In vivo deformylation of the A2K variant of Pf rubredoxin**

Much of the biophysical research on *Pf* rubredoxin is focused on contrasting its properties to those of homologous mesophile rubredoxins, most notably that from *Clostridium pasteurianum* (*Cp*). The crystal structures of the isolated N-Ala, N-Met, and N-fMet forms of *Pf* rubredoxin have been determined to 0.95, 1.1, and 1.2 Å [22], respectively, while that of the *Cp* rubredoxin is known to be 1.1 Å [28,29]. As with most known rubredoxins from mesophilic organisms, *Cp* rubredoxin is expressed with the N-terminal methionine retained as expected from the lysine residue in the second position. Taking advantage of the fact that the methyl sidechain of Ala 2 in *Pf* rubredoxin is highly exposed to solvent, this position was converted to lysine for the combined benefit of increased homology to the *Cp* protein and suppression of processing by the endogenous methionine aminopeptidase.

The A2K variant of *Pf* rubredoxin was expressed in *E. coli* strain BL21(DE3)/pLysS and purified via gel filtration and anion exchange chromatography. As illustrated in Fig. 1, analytical reverse phase HPLC on C$_4$ derivatized silica yields two species. Mass spectrometry yielded masses of 6083 and 6112, consistent with the predicted molecular weights of the apo-forms of the N-Met and N-fMet *Pf* A2K rubredoxin. It should be noted that *Pf* rubredoxin did not elute from a C18 column using up to 100% acetonitrile.

To convert the N-fMet form of *Pf* A2K rubredoxin to the Met terminal form, a constitutive expression system for the peptide deformylase was constructed. The *def* gene was PCR amplified from BL21(DE3) using the 5′ and 3′ flanking sequences and introducing HpaI linker sequences at each end. The pLysS plasmid was cut at the unique *Eco*RV site approximately midway between the clockwise directed tetracycline (Tc) promoter and the 3′ end of the counterclockwise oriented T7 lysozyme gene. The PCR product of the *def* gene amplification was digested with HpaI and blunt end ligated into the linearized pLysS plasmid. The resultant plasmid containing the *def* gene oriented under the Tc promoter control was isolated and designated pSdef1. No interference with the T7 lysozyme expression was apparent.

Both the pT7-7-derived *Pf* A2K rubredoxin expression plasmid and pSdef1 were transformed into BL21(DE3), and the protein was purified as before [24]. This rubredoxin sample now gave a single peak on reverse phase HPLC with a mass of 6084, consistent with the apo-form of the N-Met terminal species (Fig. 2). Any residual N-fMet form was reduced to <0.5%. The level of peptide deformylase expression in pSdef1 thus provides at least a ~100-fold reduction in residual N-formylation as compared to that provided by the chromosomal *def* gene. On the other hand, 1D SDS-
PAGE analysis of BL21(DE3)/pSdef1 vs. BL21(DE3)/plysS does not indicate obvious induction of a band at the molecular weight of the peptide deformylase. The still modest level of expression of the peptide deformylase would not be expected to interfere with the maximal expression of the target protein which in the present case is produced at a level of approximately 10 μmol/L of growth culture.

NMR chemical shift effects induced by the N-formyl group of Cp rubredoxin

The benefits of avoiding the growth media constraints of the trimethoprim-based system as well as minimizing excess expression of the peptide deformylase are readily apparent in the application of isotopic labeling of the target protein. Given that the peptide deformylase has a threefold greater mass than rubredoxin, expression at a comparable molar level could seriously increase the material costs. The complications arising from the N-terminal heterogeneity introduced by incomplete processing can also become apparent during NMR analysis as reported for the doubling of numerous resonances in the spectra of the calcium-binding protein S100β [14]. For the case of Cp rubredoxin, previous NMR studies [30] have detected heterogeneity in 2D $^1$H-$^{15}$N correlation spectra which report the chemical shifts of each amide proton and its directly bonded nitrogen. However, only quite recently has this heterogeneity been recognized to arise from a mixture of the N-fMet and N-Met forms of the protein [31].

In Fig. 3 is presented the central portion of the 2D $^1$H-$^{15}$N TROSY correlation spectra [25,26] for U-2H,$^{15}$N Cp rubredoxin expressed in BL21(DE3)/plysS (upper panel) and BL21(DE3)/pSdef1 (lower panel) in the presence of ZnCl$_2$ so as to generate the Zn$^{2+}$-coordinated protein. The narrow resonance linewidths provided by a combination of perdeuteration and the relaxation optimization of the TROSY experiment allow for detection of a large number of peak doublings in the upper panel arising from the presence of both the N-Met and N-fMet forms of the protein. Peak doublings at residues 1–5, 28–32, 50–52, and 54 have been previously identified by NMR analysis of $^{13}$C,$^{15}$N-labeled samples [31]. In addition, well-resolved doublings of residues 13, 14, and 53 as well as partially overlapped doublings of residues 6, 9, 16, 42, and 49 are observed in the present study. In all, the chemical shifts over 40% of the backbone amides are observably perturbed by the substitution of a formyl group on the highly solvent exposed N-terminus. As illustrated in Fig. 4, this includes the amides of Cys 9 and Cys 42 which are hydrogen bonded to the sulfurs coordinating the active site metal 21 Å away from the N-terminal nitrogen.
Production of rubredoxin for NMR studies poses a peculiar potential challenge for the use of peptide deformylase overexpression. Rubredoxin is naturally produced with a single iron atom coordinated to the four cysteine residues. The one-electron transfer between the Fe$^{2+}$ and Fe$^{3+}$ states forms the basis of this protein’s still poorly characterized biological electron transfer function. The paramagnetism of the unpaired electrons of the iron substantially shifts and broadens the $^1$H resonances of the protein [32]. To obtain the high-resolution NMR spectra reported herein, the iron is substituted with zinc, most readily accomplished by expression in an iron-deficient growth medium supplemented with Zn$^{2+}$. However, the peptide deformylase is also an iron–cysteine sulfur cluster protein for which substitution of zinc is inhibitory [33]. Nevertheless, it was found that growth on defined medium with added Zn$^{2+}$ provides a great majority of the rubredoxin in the zinc form with high efficiency processing of the N-formyl group still preserved. The residual iron form of rubredoxin is removed during the ion exchange chromatography step.

Conclusions

Incomplete deformylation occurs moderately frequently for proteins overexpressed in E. coli. Given the difficulties that will generally be involved in separating the $N$-fMet species from the $N$-Met form of the target protein, demonstration that the formyl group does not modify the structural or functional properties of interest is often not straightforward. The constitutive expression system for peptide deformylase described herein offers an easily implementable means for efficient in vivo deformylation in E. coli even for cases in which the formyl derivative constitutes a substantial fraction of the product when only the chromosomal def gene is utilized. This def expression plasmid is compatible with the widely used BL21(DE3)/pET-based expression system and can presumably function with any system compatible with the pLysS vector.

Acknowledgments

The authors thank Marly Eidsness for providing the parental Pyrococcus furiosus and Clostridium pasteurianum rubredoxin expression plasmids used in this work. We acknowledge the use of the Wadsworth Center NMR facility, Mass Spectroscopy facility, and Biochemistry facility. We thank Thomas Ryan and Leslie Eisele for technical assistance. This work was supported in part by NIH Grant GM 64736 (G.H.).

References


Fig. 4. $^{15}$N slice of 2D $^1$H–$^{15}$N TROSY correlation spectra through the amide resonance of Cys 9 in U–$^2$H,$^{15}$N Cp rubredoxin. The second smaller upfield resonance in the left-hand panel arises from the N-formyl species present when protein expression utilized the pLysS plasmid which is eliminated by use of the pSdef1 plasmid (right-hand panel).


