Protein Expression and Purification 121 (2016) 17–21

Contents lists available at ScienceDirect

Protein Expression and Purification

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

Elimination of truncated recombinant protein expressed in *Escherichia coli* by removing cryptic translation initiation site

Matthew J. Jennings [1], Adam F. Barrios [1, 2], Song Tan [*]

Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 23 August 2015
Received in revised form
25 November 2015
Accepted 3 December 2015
Available online 29 December 2015

Keywords:
Recombinant protein expression
Cryptic initiation
Escherichia coli expression

ABSTRACT

Undesirable truncated recombinant protein products pose a special expression and purification challenge because such products often share similar chromatographic properties as the desired full length protein. We describe here our observation of both full length and a truncated form of a yeast protein (Gcn5) expressed in *Escherichia coli*, and the reduction or elimination of the truncated form by mutating a cryptic Shine-Dalgarno or START codon within the Gcn5 coding region. Unsuccessful attempts to engineer in a cryptic translation initiation site into other recombinant proteins suggest that cryptic Shine-Dalgarno or START codon sequences are necessary but not sufficient for cryptic translation in *E. coli*.

1. Introduction

Heterologous overexpression in *Escherichia coli* is a common technique for producing recombinant proteins [1–3]. Although this mature technology has many advantages such as speed, low cost and simplicity, complications can include insoluble or inactive products, low levels of expression or the occurrence of truncated products. Truncated products may pose a particular problem for the subsequent purification of the recombinant protein because the truncated protein is shorter but otherwise identical in amino acid sequence to the desired full length protein. Truncated polypeptides produced in *E. coli* likely result from limited proteolysis of the heterologous protein or from improper initiation of translation.

Our understanding of the mechanism of translational initiation has been significantly enhanced by recent important structural investigations of the ribosome [4–8]. The key steps in prokaryotic translational initiation are the binding of mRNA containing a Shine-Dalgarno or ribosomal binding sequence and a START codon to a complex of the 30S ribosomal subunit, initiation factors and formyl-Met tRNA, adaptation of the mRNA to the ribosome 30S subunit mRNA channel which exposes the START codon to bind to the fMet-tRNA, and subsequent binding of the 50S ribosomal subunit.

The prokaryotic Shine-Dalgarno sequence forms the ribosome binding site on the mRNA through base pairing with the complementary sequence at the 3′ end of the ribosomal 16S rRNA [9]. In *E. coli*, the consensus Shine-Dalgarno sequence is AGGAGGT but a Shine-Dalgarno site does not need to match this consensus to be functional [10,11]. The 3′ edge of this sequence is usually located 3 to 7 nucleotides from the first base of the START codon. Although ATG is the canonical START codon, it accounts for only 83% of the START codons in *E. coli* genes [12]. The GTG codon, which otherwise codes for Val, is found at the start of 14% of *E. coli* genes, while the TTG Leu codon is found at the start of 3% of *E. coli* genes. The ATG, GTG or TTG START codons are all recognized by the formyl-Met tRNA, resulting in a formyl-Met at the N-terminus of the newly synthesized polypeptide.

Given the relatively degenerate sequence requirements for the Shine-Dalgarno site and the START codon and the variable distance possible between these two sequence elements [10,13], one might expect these sequences to be found internal to coding regions in addition to the canonical location at the 5′ end of the coding region. For example, such potential cryptic initiation sites might be found in the coding regions of heterologous genes since there would be no evolutionary pressure to avoid such occurrences. If such cryptic initiation sites occurred in frame with the reading frame of the full length gene, a truncated protein product could result.

We describe here our finding that a cryptic initiation site

* Corresponding author. Center for Eukaryotic Gene Regulation, Department of Biochemistry & Molecular Biology, 108 Althouse Laboratory, The Pennsylvania State University, University Park, PA 16802, USA.
E-mail address: sxt30@psu.edu (S. Tan).
[1] Authors made equal contributions to this work.
[2] Current affiliation: DesigneRx Pharmaceuticals, Vacaville, CA, USA.

http://dx.doi.org/10.1016/j.pep.2015.12.001
1046-5928/

comprised of a near consensus Shine-Dalgarno site coupled with a GTG START codon accounted for a truncated product when a particular yeast gene (Gcn5, a histone acetyltransferase) was expressed in *E. coli*. The truncated product could be nearly eliminated by silent mutations that removed the cryptic initiation sequence. However, installing a consensus Shine-Dalgarno sequence spaced appropriately from a GTG potential START codon in a different context was not sufficient to cause translation of a truncated product.

2. Results

2.1. Truncated yeast Gcn5 polypeptide coexpressed in E. coli

We have previously described methods to express and purify recombinant protein complexes by coexpression from polycistronic vectors in *E. coli* [14–16]. While we were purifying recombinant yeast Ada2/Ada3/Gcn5 SAGA histone acetyltransferase subcomplexes produced by coexpression, we observed a 41 kD polypeptide which copurified with the desired Ada2/Ada3/Gcn5 complexes (hexahistidine tagged on the Ada3 subunit) over multiple chromatography steps including metal affinity, cation-exchange, anion-exchange and size exclusion chromatography (data not shown). For example, the 41 kD polypeptide was present when we coexpressed a particular deletion Ada2/Ada3/Gcn5 variant and partially purified the complex by metal affinity chromatography (Fig. 1). The same truncated Gcn5 product was observed with the BL21(DE3)pLysS or BL21-CodonPlus(DE3)-RIL *E. coli* host strains, but for unknown reasons, significantly better purification over Talon metal affinity resin was obtained using BL21-CodonPlus(DE3)-RIL cells [17].

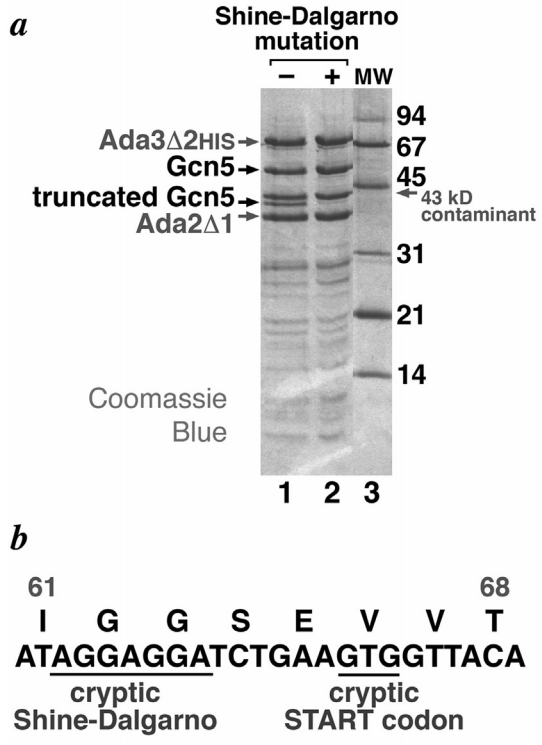

Fig. 1. Truncated Gcn5 product in recombinant yeast Ada2/Ada3/Gcn5 complex expressed in *E. coli*. (a) SDS-PAGE gel of metal-affinity purified yeast Ada2Δ1/Ada3Δ2HIS/Gcn5 complex expressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells of wild-type Gcn5 sequence (lane 1) and Gcn5 sequence mutated to remove cryptic Shine-Dalgarno sequence in codons 62 and 63 (lane 2). The prominent 43 kD contaminant was determined to be *E. coli* translation elongation factor EFTu [17]. Molecular weight markers are show in lane 3 (b) Amino acid and DNA sequence around cryptic Shine Dalgarno and cryptic START codon in native yeast Gcn5.

Since it was unlikely that an *E. coli* contaminant would copurify through all these chromatographic procedures, we suspected that the 41 kD polypeptide was a proteolytic degradation product of Ada3 or Gcn5 (the 41 kD polypeptide was unlikely to be a proteolytic product of Ada2 since it migrated slower on SDS-PAGE than the Ada2Δ1 deletion). To determine the identity of the 41 kD polypeptide, we therefore analyzed this polypeptide by N-terminal protein sequencing. The sequencing results indicated that the 41 kD polypeptide contained an N-terminus starting from Gcn5 position 67, although the limited sample amount and/or potential variability due to partial processing by methionine aminopeptidase precluded unambiguous identification of the first three amino acids (the following 5 residues were identified unambiguously and matched the corresponding residues in the Gcn5 sequence). We concluded that the 41 kD polypeptide corresponds to Gcn5 from Val67 through to its C-terminus, i.e. Gcn5(67-439) because of the good agreement between the expected and observed molecular weights (41.4 kD vs 41 kD). We surmised that truncation of the Gcn5 polypeptide occurred in the cell and not during purification because the 41 kD polypeptide was observed in Western blots of crude cell extracts prepared by boiling the recombinant *E. coli* cells in SDS-PAGE loading buffer. The explanation we favored at this time was that the truncated product resulted from proteolysis *in vivo*.

2.2. Cryptic initiation of translation produces truncated yeast Gcn5 coexpressed in E. coli

An alternate interpretation of the previous observations was prompted by discussions with David Garboczi (NIAID, NIH) who had noted translation from cryptic initiation sites in *E. coli*, including ones from Val GTG codons. These discussions spurred us to examine the coding sequence around the truncated Gcn5 product. We found the sequence AGGAGGA, a near perfect match to the consensus *E. coli* Shine-Dalgarno AGGAGGT, positioned 6 bp upstream of GTG, the codon for Val67. This suggested the possibility that the truncated product, Gcn5(67-439) resulted not from proteolysis but instead from initiation of translation from a cryptic initiation site at Val67.

To test this hypothesis, we engineered translationally silent mutations in yeast Gcn5 that removed the cryptic Shine-Dalgarno site, the cryptic START GTG codon or both. We coexpressed these translationally silent Gcn5 mutants together with Ada2 and hexahistidine tagged Ada3, and prepared crude extracts as well as partially purified the complex by metal affinity chromatography. The crude extract samples showed us truncations that presumably occurred in the cell and not during the purification process, while the metal affinity purified samples allowed us to distinguish between Gcn5 polypeptides in the tagged complex from polypeptides in the crude extract that cross-react with the anti-Gcn5 antibodies used for the Western blot. As Fig. 2 shows, the truncated Gcn5 product is found in both the crude extract and in the metal affinity purified complex (Fig. 2, lanes 1 and 2), consistent with it copurifying with the Ada2/Ada3/Gcn5 complex tagged on the Ada3 subunit. When the cryptic Shine-Dalgarno site was removed by silently mutating the natural Gcn5 AGGAGGA sequence to AGGCGGC, the band corresponding to the truncated Gcn5 polypeptide is almost completely removed and only a faint band at this position on the gel remains in both the crude extract and in the partially purified complex (Fig. 2 lanes 3 and 4). Mutating the cryptic START GTG site to a GTT codon appears to be even more effective at eliminating the truncated product since even the faint band observed with the Shine-Dalgarno mutation is not evident