Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 45 (2006) 374–380
Protein Expression & Purification
www.elsevier.com/locate/yprep

Rapid and easy thermodynamic optimization of the 5′-end of mRNA dramatically increases the level of wild type protein expression in *Escherichia coli*

Régis Cèbe, Martin Geiser *

Novartis Institutes for BioMedical Research Basel, Discovery Technologies/Protein Structure Unit, WSJ 88.707, CH-4002 Basel, Switzerland

Received 18 May 2005, and in revised form 28 June 2005
Available online 8 August 2005

Abstract

Low levels of expression in *Escherichia coli* are often observed when using wild type proteins. The addition of an N-terminal His-tag to these same proteins dramatically improves the level of expression. We therefore concluded that post-transcriptional regulation and in particular translational regulation are probably influenced by the presence of the tag. The RNAfold program was used to analyze the 5′-end of the encoding mRNA, and more precisely the area encompassing the Shine–Dalgarno region and the initiation codon ATG. We observed that hairpin loops can be formed and that the stability of these loops correlates with the level of protein expression in *E. coli*. Our recently developed cloning technology by PCR fragment integration allows us to easily and rapidly introduce mutations anywhere within a gene. In our studies, we used this technology to destabilize the predicted hairpin by introducing silent mutations within the first 72 nucleotides of the coding sequence. As a result of the decreased stability of the RNA hairpins, we could significantly increase the level of expression of wild type proteins and without having to rely on the use of tags in *E. coli*. In addition, our studies allow us to predict whether or not a protein will be expressed without additional engineering of its encoding gene.

Keywords: Initiation of translation; mRNA secondary structure; Fusion proteins

The level of protein expression in *Escherichia coli* has a profound influence on the complete "gene to structure" process. Low levels of expression make purification attempts more difficult and usually require either the utilization of tags for more efficient purification or larger fermentation volumes thus making the whole process tedious and less cost-effective. Alternatively, the protein can be expressed in a different host which in almost all the cases tends to be baculovirus-infected insect cells. Such alternative means might facilitate the expression of the recombinant protein but always at the cost of time and would be less amenable to labeling.

The influence of mRNA secondary structures on the efficiency of translational initiation in prokaryotes has been reported with individual proteins [1–7]. We thus decided to analyze the effect the secondary mRNA structures might have on the expression of a selection of proteins which we originally were unable to express well in *E. coli*. If the nucleotide sequence around the ribosome-binding site on the mRNA and the region encoding the N-terminal portion of the protein to be expressed is capable of forming stable hairpins and loops, then the ribosomes will very likely stall and the translation level might be dramatically affected. In the studies presented here, we destabilized these RNA hairpins by introducing silent mutations in the 5′ terminal part of the encoding gene using a simple and very fast PCR method [8] and

* Corresponding author. Fax: +41 61 3242686.
E-mail address: martin.geiser@novartis.com (M. Geiser).

1046-5928/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2005.07.007

Table 1
List of the oligonucleotides utilized in the studies presented here

Oligonucleotides	Sequence
MG379	GAGCGTGACGGCTGAAGGTGCCCAC
MG380	GTGGGCACCTTCAGCCGTCACGCTC
MG466	CACCGCCGCCGCAAGGAATGG
MG1114	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGCAGGGGTGGCAGGCGTTCAAG
MG1115	CGTCGACGTAGGCCTTTGAATTCATTAGTAGGCGTTCTCCAGCTCGG
MG1142	CGTCGACGTAGGCCTTTGAATTCACACAGAGATCCACGGGGGCGCC
MG1143	CTGGAAGTTCTGTTCCAGGGGCCACTCATCGGGCATGGTACTAAGGTC
MG1149	GGCCGCCCGCTGGATCCATGGTATATCTCCTTCTTAAAGTTAAACAAAA
MG1175	GGCCCCAGGGGTGGCAGGCGTTCAAG
MG1176	CATTAGTAGGCGTTCTCCAGCTCGG
MG1177	CCAGGGGTGGCAGGCGTTCAAG
MG1178	AATTCATTAGTAGGCGTTCTCCAGCTCGG
MG1179	GGATCCAGCGGGTGGCCCACGGGGCGTCCTCCCTCGCCCATGCAGAGTGCTGGTGCTGC
MG1192	GGTTCAGCAGCACCAGAACTCTGCAAGGTCTTGGTAGGACTCCTCTTGGGCCTCCTGCTGGATCCATGG
MG1198	CCCCTTGAGCTCCAGAAGCAGTGCCTTTTGAGAACTAGGTACAAGTGGTCTTGCCATGGTATATCTCC
MG1242	CATGGGCCAAGGATGGCAAGCTTTTAAGAATG
MG1243	GGCCAAGGATGGCAAGCTTTTAAGAATG
RS68	GCTAGTTATTGCTCAGCGG

showed that these mutations significantly influenced the level of protein expression in *E. coli.* These results can be achieved without having to rely on the use of N-terminal tags. In this report, we will present three examples.

Materials and methods

Cloning

All the cloning steps were performed using the PCR fragment integration method following the published protocol [8]. In short, the method allows one to insert PCR fragments between any two nucleotides within a target plasmid. The only requirement is that the amplified fragments are embedded between DNA sequences homologous to the site in which the integration is planned. This method is an adaptation of the QuikChange Site-Directed Mutagenesis (Stratagene) protocol. All the DNA constructs were tested for integrity by DNA sequence analysis.

CDC34(1–180)

The clone encoding the full-length CDC34[1] (L22005) was received from F.Hamy (NIBR Basel). The gene was cloned into pET28a between the restriction sites *Nco*I and *Xho*I resulting in plasmid pXI321. A stop codon was introduced in the gene sequence to obtain a new plasmid (pXI329) which is now encoding CDC34(1–180). The insertion of the stop codon was made with the classic procedure of the QuikChange site-directed mutagenesis kit (Stratagene) with the oligonucleotides MG379/MG380. The list of all the oligonucleotides used in these studies is given in Table 1. A PCR fragment encoding for Met-Gly-Ser-Ser-$(His)_6$-PreScission protease cleavage site was integrated at the N-terminus of CDC34(1–180) as described [8]. The resulting plasmid was called pXI330.

Silent mutations were introduced in the sequence encoding the first [13] residues of the gene by integration of a PCR fragment containing all the mutations into pXI329. The mutations were introduced into this PCR fragment with the oligonucleotides MG1198 used together with MG466 with pXI329 as template DNA. The resulting plasmid was called pXI329b.

Sphingosine kinase 1

The gene encoding the full-length Sphingosine kinase 1 (BC030553) was encoded on a plasmid which we received from A. Billich (NIBR Vienna). The gene was recloned into a pET28b vector with Met-Gly-Ser-Ser-$(His)_6$-Stag-PreScission protease cleavage site sequence as an N-terminal fusion. The resulting plasmid was called pXI466a. A PCR fragment derived from pXI466a with the oligonucleotides MG1149/MG466 was integrated again into pXI466a resulting in plasmid pXI466c. In this plasmid, all the different tags were removed.

The modification of the sequence within the 5′-end of the gene was done in two steps. First, a PCR fragment amplified from pXI466c with MG1179/RS68 was integrated back into pXI466c. The resulting plasmid was used as template for a second round of mutations. The PCR fragment amplified from this last plasmid with MG1192/MG466 was back integrated into the same plasmid. The plasmid obtained after the two rounds of mutation was called pXI466b.

[1] *Abbreviations used:* wt, wild type; PreSc, cleavage site for the protease PreScission; CDC34, ubiquitin-conjugating enzyme E2-32 kDa.