Protein Expression and Purification **24,** 61–70 (2002)
doi:10.1006/prep.2001.1545, available online at http://www.idealibrary.com on IDEAL®

Processive Degradation of Nascent Polypeptides, Triggered by Tandem AGA Codons, Limits the Accumulation of Recombinant Tobacco Etch Virus Protease in *Escherichia coli* BL21(DE3)

Rachel B. Kapust, Karen M. Routzahn, and David S. Waugh[1]

Protein Engineering Section, Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, P.O. Box B, Frederick, Maryland 21702-1201

Received June 5, 2001, and in revised form August 16, 2001; published online December 10, 2001

Due to its high degree of sequence specificity, the catalytic domain of the nuclear inclusion protease from tobacco etch virus (TEV protease) is a useful reagent for cleaving genetically engineered fusion proteins. However, the overproduction of TEV protease in *Escherichia coli* has been hampered in the past by low yield and poor solubility. Here we demonstrate that the low yield can be attributed to the presence of arginine codons in the TEV protease coding sequence that are rarely used in *E. coli* and specifically to a tandem pair of AGA codons. The yield of protease can be improved by replacing these rare arginine codons with synonymous ones or by increasing the supply of cognate tRNA that is available to the cell. Furthermore, we show that when ribosomes become stalled at rare arginine codons in the TEV protease mRNA, the nascent polypeptides are targeted for proteolytic degradation in BL21(DE3) cells by a mechanism that does not involve tmRNA-mediated peptide tagging.

Affinity tags can facilitate the purification, increase the yield, and even improve the solubility of recombinant proteins (1–5). In the end, however, it is usually desirable to remove the tag from the target protein, and it is this step in the process that has proven to be the Achilles heel of the fusion approach. The main problem is specificity. Affinity tags are usually removed by site-specific proteolysis at a designed site, using factor Xa, enteropeptidase (enterokinase), or thrombin. However, all of these proteases have been observed to cleave fusion proteins at locations other than the intended site (e.g., 6–11). An alternative reagent for removing affinity tags that shows considerable promise is TEV[2] protease, the catalytic domain of the nuclear inclusion protease from tobacco etch virus (12). In contrast to factor Xa, thrombin, and enteropeptidase, no instances of cleavage by TEV protease at noncanonical sites in fusion proteins have been reported.

Although two procedures for overproducing the TEV protease catalytic domain in *Escherichia coli* have been described, the yield of active enzyme was relatively low (1–10 mg per liter of cells) in both cases (13, 14). In the present study, we examined the influence of codon bias on the production of TEV protease in *E. coli* BL21(DE3) cells. A pair of tandem AGA codons in the TEV protease mRNA was shown to be responsible for the low yield of protease. Altering these codons to CGC or increasing the supply of cognate tRNA (*argU*) in the cells caused the protease to accumulate to a very high level. Additional experiments were performed to investigate what happens to nascent polypeptides when ribosomes become stalled at rare arginine codons within the TEV

[1] To whom correspondence should be addressed. Fax: (301) 846-7148. E-mail: waughd@ncifcrf.gov.

[2] Abbreviations used: CAT, chloramphenicol acetyltransferase; MBP, *E. coli* maltose-binding protein; PCR, polymerase chain reaction; TEV, tobacco etch virus; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; *ssrA*, gene encoding tmRNA in *E. coli*; LB, Luria–Bertani; DIG, digoxigenin; EOP, equivalent to a plating efficiency; IPTG, isopropyl β-D-thiogalactopyranoside.

1046-5928/02 $35.00

protease mRNA. The results suggest that pausing at rare arginine codons triggers the processive degradation of partially synthesized polypeptides by a mechanism that does not depend on tmRNA.

MATERIALS AND METHODS

Bacterial Strains

E. coli BL21(DE3) cells (15) were obtained from Novagen (Madison, WI). X90(DE3) cells were described previously (10). X90*ssrA*::CAT cells were obtained from Dr. R. Sauer (MIT). BL21(DE3)*ssrA*::CAT cells were constructed by P1 transduction as described (16). *E. coli* SG1146a, a *clpP*$^-$ derivative of BL21(DE3), was obtained from Dr. S. Gottesman (NCI).

TEV Protease Expression Vectors

pRK508 and pDW484 were described previously (17). pDW533, which produces wild-type *E. coli* maltose-binding protein without its N-terminal signal peptide, has also been described (18). pDC952 (19) was obtained from Dr. J. Walker (University of Texas). pACYC184 was obtained from New England Biolabs, Inc. (Beverly, MA). The following oligodeoxyribonucleotide primers were used to construct the remainder of the TEV protease expression vectors used in this study: PE-29, 5′-GAT GAA GCC CTG AAA GAC GCG CAG-3′; PE-30, 5′-GCA AGG CGA TTA AGT TGG GTA ACG C-3′; PE-398, 5′-GTC CTG GAT CCT CAT TAA AAC AAG TGC TTG TTT GTA ATG ATG-3′; PE-404, 5′-CCT CCC ATC GAT GAG GTG TTG TTG-3′; PE-408, 5′-CCC TCG AAG ACG GAG CTG GAT CCA TTA TTG CGA GTA GAC TAA TTC-3′; PE-409, 5′-CAA CAA CAC CTC ATC GAT GGG AGG GAC-3′; PE-410, 5′-AAG CAC TTG TTT CGC CGC AAT AAT GGA ACA CTG-3′; PE-411, 5′-TGT TCC ATT ATT GCG GCG AAA CAA GTG CTT GTT TG-3′; PE-412, 5′-CTC ATC GAT GGG CGC GAC ATG ATA ATT-3′; PE-413, 5′-AAT TAT CAT GTC GCG CCC ATC GAT GAG-3′; PE-414, 5′-CTG AAA TTT CGC GAG CCA CAA CGC GAA GAG CGC-3′; PE-415, 5′-GCG CTC TTC GCG TTG TGG CTC GCG AAA TTT CAG-3′; PE-416, 5′-TTA GTA TCA ACT CGC GAT GGG TTC ATT-3′; PE-417, 5′-AAT GAA CCC ATC GCG AGT TGA TAC TAA-3′; PE-770, 5′-CCT CAT CGA TGG GAG GAA CAT GAT AAT TAT TCG-3′; and PE-771, 5′-CGA ATA ATT ATC ATG TTC CTC CCA TC-3′. The expression vector encoding the wild-type MBP–TEV protease fusion protein (pRK688) and its derivatives (pRK689-693) were constructed by overlap extension PCR (20). To construct pRK688, first pDW484 was used as the template for two separate PCR reactions, one with primers PE-29 and PE-404 and the other with primers PE-408 and PE-409. These two PCR amplicons then were combined and used as the template for another PCR reaction, this time using primers PE-29 and PE-408. The PCR amplicon was digested with *Sac*I and *Bbs*I and then ligated with the *Sac*I/*Hin*dIII vector backbone of pMal-C2 (New England Biolabs, Inc.) to create pRK688. This expression vector contains unique *Sac*I, *Hin*dIII, *Cla*I, *Sph*I, *Spe*I, *Xmn*I, and *Bam*HI restriction sites within and adjacent to the TEV protease coding sequence (Fig. 1B). To construct pRK689, pRK688 was used as the template for two separate PCR reactions, one with primers PE-29 and PE-411 and the other with PE-410 and PE-30. The two PCR amplicons were combined and used as the template for a third PCR reaction, this time with primers PE-29 and PE-30. The resulting PCR amplicon was digested with *Sac*I and *Cla*I and then ligated with the *Sac*I/*Cla*I vector backbone fragment of pRK688 to create pRK689. A similar strategy was used to construct pRK690. Two PCR reactions were performed with pRK688 as the template, one with primers PE-29 and PE-413 and the other with primers PE-412 and PE-30. The two amplicons were combined and used as the template for a third PCR reaction with primers PE-29 and PE-30. The amplicon was digested with *Hin*dIII and *Spe*I and then ligated with the *Hin*dIII/*Spe*I vector backbone fragment of pRK688 to create pRK690. The same strategy was used to construct pRK691, except that primers PE-415 and PE-414 were used instead of primers PE-413 and PE-412, respectively. To construct pRK692, two PCR reactions were performed with pRK688 as the template, one with primers PE-29 and PE-417 and the other with primers PE-416 and PE-30. The two amplicons were combined and used as the template for another PCR reaction with primers PE-29 and PE-30. The amplicon was digested with *Spe*I and *Bam*HI and then ligated with the *Spe*I/*Bam*HI vector backbone fragment of pRK688 to create pRK692. To construct pRK693, two separate PCR reactions were performed using different templates. In one reaction, pRK690 was the template and the primers were PE-29 and PE-415. In the other reaction, pRK692 was the template and PE-414 and PE-30 were the primers. The two amplicons from these PCR reactions were combined and used as the template for a third PCR reaction, this time using PE-29 and PE-30 as primers. This PCR product was digested with *Cla*I and *Xmn*I and then ligated with the *Cla*I/*Xmn*I vector backbone fragment of pRK689 to create pRK693. pRK685 was constructed by amplifying the TEV protease open reading frame from pDW484 with primers PE-29 and PE-398, cleaving the PCR amplicon with *Sac*I and *Bam*HI, and then ligating it with the *Sac*I/*Bam*HI vector backbone fragment of pMal-C2. To construct pRK1033, first two PCR reactions were performed using pRK508 as the template. In one reaction the primers were PE-29 and PE-771, while in the other reaction they were PE-770 and PE-30. These PCR amplicons were combined and used as the template for a third

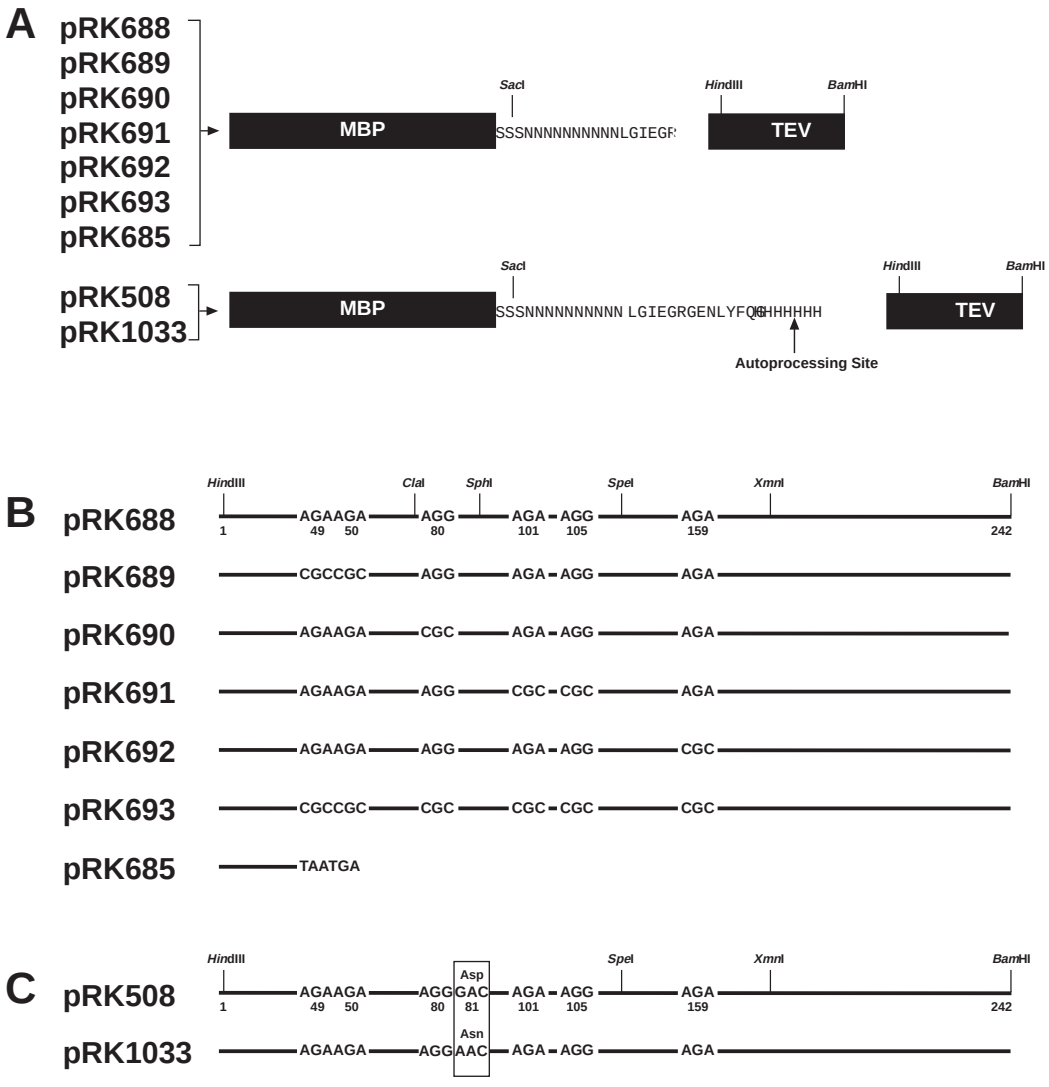

FIG. 1. Schematic representations of the MBP–TEV protease expression vectors used in this study. (A) Amino acid sequences (in single letter code) of the interdomain linkers encoded by various MBP–TEV protease expression vectors. (B) MBP–TEV protease expression vectors with modified arginine codons. pRK688 encodes the wild-type MBP–TEV protease fusion protein. The specific codon replacements in each (otherwise identical) vector are indicated. Unique restriction sites in the TEV protease coding sequence, some of which were engineered by site-directed mutagenesis, are also indicated. (C) Autoprocessing and catalytically inactive MBP–TEV protease fusion proteins. pRK508 produces wild-type TEV protease. pRK1033 produces a protease with a single amino acid substitution at position 81 (Asp to Asn) that renders it catalytically inactive, but is otherwise identical to pRK508.

PCR reaction, this time using primers PE-29 and PE-30. This PCR amplicon was digested with *Sac*I and *Bam*HI and then ligated with the *Sac*I/*Bam*HI vector backbone fragment of pRK508 to create pRK1033. The nucleotide sequences of all the MBP-TEV protease expression vectors (Fig. 1) were confirmed experimentally.

tmRNA Plasmid

The *ssrA* gene and flanking regulatory sequences were amplified from X90 genomic DNA by PCR using primers PE-206 (5′-TAT TAT GAC GTC AGG CTA CAT GGG TGC TAA ATC-3′) and PE-207 (5′-TAT TAT ATC GAT CTT CGC GGG ACA AAT TGA GGG CAC-3′). The amplicon was digested with *Aat*II and *Cla*I and then ligated with the *Aat*II/*Cla*I vector backbone of the low copy number plasmid pZS*24-MCS1 (21) to create pRK593.

Protein Expression and SDS–PAGE

Cells from single colonies were grown to saturation in LB broth (22) supplemented with the appropriate antibiotics (100 μg/ml ampicillin and/or 30 μg/ml chloramphenicol) at 37°C. These cultures were diluted 100-fold in the same medium and grown to early log phase

(A_{600} = 0.3–0.5) at 37°C, at which time IPTG was added to a final concentration of 1 mM. After 3 h of shaking at 37°C, the cells from 10 ml of each culture were recovered by centrifugation and resuspended in 1 ml of 20 mM Tris–HCl (pH 7.6), 1 mM EDTA. The cell suspensions were sonicated to induce lysis. Aliquots of the sonicated cell suspensions were mixed with an equal volume of 2× SDS sample buffer (23) to generate samples of the total intracellular protein for SDS–PAGE. All solutions of protein in sample buffer were heated at 90°C for 4 min and then centrifuged at 14,000*g* for 15 min prior to SDS–PAGE. Precast SDS–polyacrylamide gels (10–20%, Tris–glycine) were purchased from Novex, a subsidiary of Invitrogen (San Diego, CA).

Southern Blot

Samples of genomic DNA from X90, X90*ssrA*::CAT, BL21(DE3), and a putative *ssrA*$^-$ transductant of BL21(DE3) were prepared using Qiagen (Valencia, CA) Genomic-tips in accordance with the manufacturer's instructions. Genomic DNA (3 μg) was digested overnight with *Cla*I or *Pst*I. The reaction products were separated on a 1% agarose gel in TBE buffer (23). The *ssrA*-specific gene probe was generated by PCR amplification of X90 DNA with primers PE-206 and PE-207. The PCR amplicon (25 ng) was labeled by random priming in a reaction containing 2.5 units of Klenow fragment and 1× DIG labeling mix (Roche Molecular Biochemicals, Indianapolis, IN). The reaction mixture was incubated for 1 h at 37°C. Southern transfer was performed in 20× SSC overnight (23). After transfer, the DNA was UV-crosslinked onto the nylon membrane (Tropix, Bedford, MA). The membrane was incubated in Ultrahyb solution (Ambion, Austin, TX) at 42°C for 1 h, before the DIG-labeled *ssrA* gene-specific probe was added to 0.1 ng/ml and incubation was continued overnight. After hybridization, the membrane was washed twice for 5 min with 2× SSC, 0.1% SDS and then twice for 15 min with 0.1× SSC, 0.1% SDS at 42°C. DNA hybrids were detected with anti-DIG–alkaline phosphatase (1:20,000) and the chemiluminescence substrate CDP-Star (Roche Molecular Biochemicals).

Phage Plating Assay

To compare the plating efficiency of phage λimmP22 *dis c*2-5 on BL21(DE3), BL21(DE3)*ssrA*::CAT, and BL21(DE3)*ssrA*::CAT cells containing the tmRNA plasmid pRK593, the cells were grown in LB broth (22) supplemented with 0.2% maltose to an OD_{600} of 0.25. Serial dilutions (between 10^{-2} and 10^{-8}) of the high titer phage stock were made in LB + maltose. One-hundred microliters of each phage dilution was combined with 1 ml of cells and incubated for 20 min at 37°C. A total of 2.5 ml of molten (45°C) top agar (22) was then added to each tube containing cells and phage, and the mixture was spread on an LB plate and incubated overnight at 37°C. Plaques were counted by visual inspection. The phage titer on BL21(DE3) cells (ca. 9×10^9 pfu/ml) was equivalent to a plating efficiency (EOP) of 1.

RESULTS

Although most MBP fusion proteins can be expressed at a high level in *E. coli* (24, 25), this was not the case when the catalytic domain of TEV protease was fused to the C-terminus of MBP (Fig. 2, lane 2). Transcription and translation initiation should be very efficient in the context of an MBP fusion protein vector, and so we wondered if the problem might be at the level of translation elongation, due to differing codon biases in *E. coli* and eukaryotic cells. The TEV protease coding sequence contains two AGG codons (R80 and R105) and four AGA codons (R49, R50, R101, and R159), which are rarely found in highly expressed genes in *E. coli* (Fig. 1). It is well documented that the presence of these

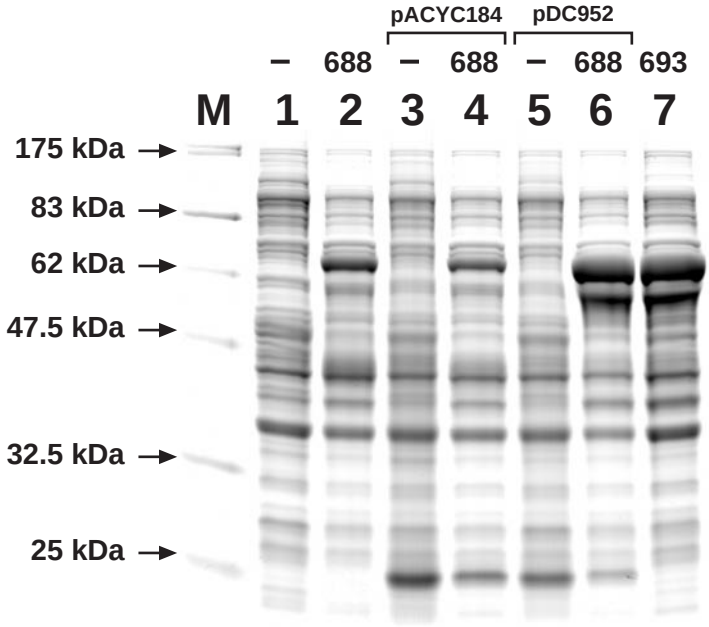

FIG. 2. Rare arginine codons limit the accumulation of TEV protease in *E. coli* BL21(DE3). Cells were grown to early log phase at 37°C and induced with IPTG for 3 h. Samples of the total intracellular protein from an equivalent number of cells were resolved by SDS–PAGE (10–20% Tris–glycine gel, Novex) and stained with GelCode Blue (Pierce). Lane M, molecular weight standards; lane 1, BL21(DE3); lane 2, BL21(DE3) + pRK688; lane 3, BL21(DE3) + pACYC184; lane 4, BL21(DE3) + pACYC184 + pRK688; lane 5, BL21(DE3) + pDC952; lane 6, BL21(DE3) + pDC952 + pRK688; lane 7, BL21(DE3) + pRK693.

codons in an mRNA can have a deleterious effect on the yield of a recombinant protein in *E. coli* (26). In several instances, much higher yields could be obtained when the cells were supplemented with multiple copies of the cognate tRNA gene (*argU*) on a plasmid, suggesting that in these cases the accumulation of recombinant protein was limited by depletion of the cognate aminoacyl tRNA (e.g., 27–32).

To ascertain whether rare arginine codons in the TEV protease mRNA affect the yield of MBP-TEV protease fusion protein, we examined the amount of MBP-TEV protease produced by pRK688 (Fig. 1) in BL21(DE3) cells that also contained pDC952, a derivative of pACYC184 that carries the *E. coli argU* gene (19). Consistent with a translational effect, the yield was far greater when cells contained pDC952 than when they contained pACYC184 (Fig. 2, lanes 6 and 4, respectively). A high yield of the 69-kDa fusion protein, equivalent to that produced when cells contained pDC952, could also be obtained in the absence of this plasmid if all of the AGG and AGA codons were replaced by CGC (Fig. 2, lane 7). There are no rare arginine codons in MBP, so these results demonstrate that AGA and/or AGG codons within the TEV protease coding sequence limit the accumulation of MBP–TEV protease fusion protein in *E. coli*.

Does each AGG or AGA codon contribute equally to the low yield of TEV protease in *E. coli*? This question was addressed by constructing a series of MBP–TEV protease expression vectors (pRK689–pRK692) in which only one or two AGG or AGA codons at a time were changed to CGC (Fig. 1B). As expected, all of these vectors produced a large amount of MBP–TEV protease in the presence of pDC952 (Fig. 3, even lanes), but only the vector in which the tandem AGA codons had been altered to CGC produced a similar amount of fusion protein in the absence of pDC952 (Fig. 3, lane 5). Because pRK691 also contains two codon replacements (R101 and R105), this result implies that it is not simply the number of rare arginine codons that makes the difference; the context is also important. Even so, the simultaneous replacement of codons 101 and 105 appeared to be more beneficial than the single codon replacements. Replacing only the tandem AGA codons restored the yield of MBP–TEV protease to the level seen in the presence of pDC952. Similar results were obtained when the effect of tandem rare arginine codons in other proteins was investigated (33–35).

If the depletion of aminoacyl-tRNA$^{AGA/AGG}$ arrests translation of the fusion protein, then this would explain the low yield of full-length MBP–TEV protease, but what happens to the partially translated fusion proteins? Little or no truncated fusion protein accumulates in the absence of pDC952 (compare Fig. 2, lanes 4 and 6, or Fig. 3, lanes 3 and 4). An interesting result is obtained when an autoprocessing form of the MBP–

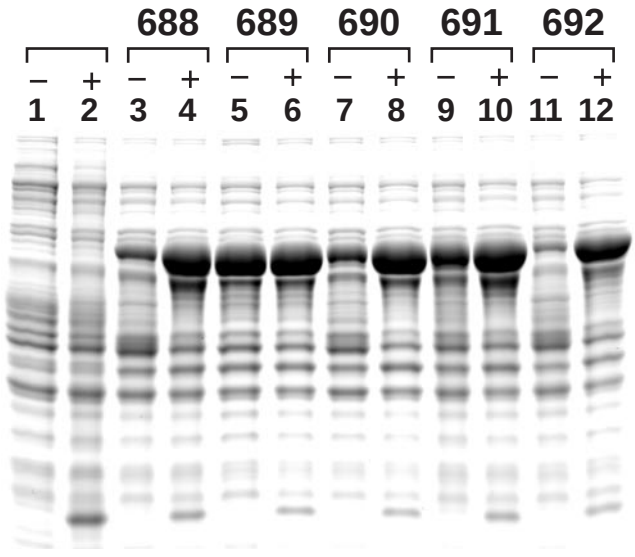

FIG. 3. Tandem AGA codons have the greatest impact on the yield of MBP–TEV protease fusion protein. Cells were grown to early log phase at 37°C and induced with IPTG for 3 h. Samples of the total intracellular protein from an equivalent number of cells were resolved by SDS–PAGE (10–20% Tris–glycine gel, Novex) and stained with GelCode Blue (Pierce). (+) and (−), presence or absence, respectively, of the tRNA accessory plasmid pDC952. Lanes 1 and 2, no MBP–TEV protease expression vector; lanes 3 and 4, pRK688; lanes 5 and 6, pRK689; lanes 7 and 8, pRK690; lanes 9 and 10, pRK691; lanes 11 and 12, pRK692.

TEV protease fusion protein with a canonical TEV protease recognition site in the linker between the two domains (Fig. 1A) is expressed in *E. coli*. Both in the presence and in the absence of pDC952, the fusion protein is processed efficiently in cells to yield separate MBP and TEV protease domains (Fig. 4, lanes 6 and 5, respectively). In the presence of pDC952, stoichiometrically equivalent amounts of both polypeptides accumulate to a high level. However, a different result is obtained when the self-processing fusion protein is expressed in the absence of pDC952. Although the yield of MBP is about the same, much less TEV protease is produced. Again, no truncated translation products are visible on the gel (compare Fig. 4, lanes 5 and 6). Most remarkably, when pDC952 is not present, the yield of the nonprocessing MBP–TEV protease fusion protein is far less than the yield of MBP generated by the autoprocessing fusion protein under identical conditions. This observation strongly suggests that the partially synthesized fusion proteins are unstable and subject to proteolytic degradation *in vivo*. To rule out the possibility that differences between the mRNAs produced by the two vectors (i.e., the presence or absence of a TEV protease recognition site in the linker between the MBP and TEV protease domains) are responsible for the observed effect, the yield of a catalytically inactive form of the MBP–TEV protease fusion protein, produced by

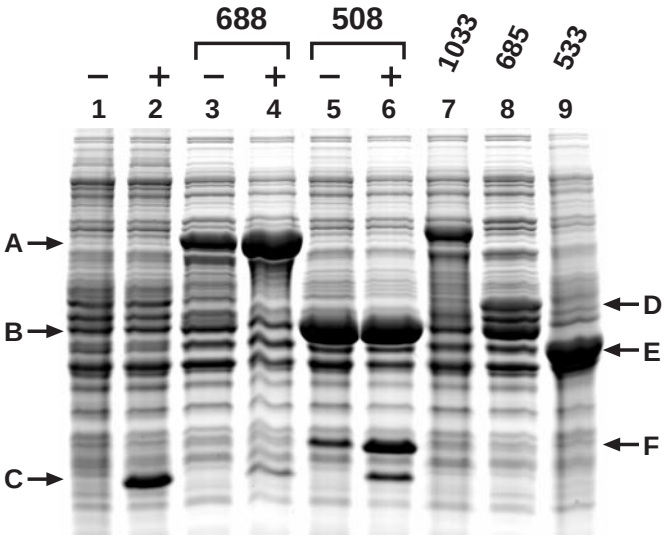

FIG. 4. Processive degradation of nascent MBP–TEV protease polypeptides. Cells were grown to early log phase at 37°C and induced with IPTG for 3 h. Samples of the total intracellular protein from an equivalent number of cells were resolved by SDS–PAGE (10–20% Tris–glycine gel, Novex) and stained with GelCode Blue (Pierce). Lanes 1, BL21(DE3); lane 2, BL21(DE3) + pDC952; lane 3, BL21(DE3) + pRK688; lane 4, BL21(DE3) + pDC952 + pRK688; lane 5, BL21(DE3) + pRK508; lane 6, BL21(DE3) + pDC952 + pRK508; lane 7, BL21(DE3) + pRK1033; lane 8, BL21(DE3) + pRK685; lane 9, (BL21(DE3) + pDW533. Recombinant proteins: A, full-length MBP–TEV fusion protein (lanes 3, 4, and 7); B, MBP domain produced by autoprocessing of the MBP–TEV fusion protein *in vivo* (lanes 5 and 6); C, chloramphenicol acetyltransferase encoded by pDC952 (lanes 2, 4, and 6); D, truncated MBP–TEV fusion protein produced by pRK685 (lane 8); E, MBP with no C-terminal linker peptide (lane 9); F, TEV protease catalytic domain produced by autoprocessing of the MBP–TEV fusion protein *in vivo* (lanes 5 and 6).

pRK1033 (Fig. 1A), was also examined. This fusion protein, which contains a single amino acid substitution (D81N) in the catalytic triad of TEV protease, also accumulated at a low level in the absence of pDC952 (Fig. 4, lane 7). Because the mRNA encoding the catalytically inactive fusion protein is otherwise identical to that encoding the autoprocessing fusion protein, the dramatic difference between the yield of processing and nonprocessing forms of the MBP–TEV protease fusion protein cannot be attributed to the differences in their mRNAs (or protein) that result from the presence or absence of a TEV protease site and hexahistidine tag in the interdomain linker.

We next sought to determine whether the susceptibility of these incompletely synthesized fusion proteins to proteolysis is triggered directly or indirectly by the presence of rare arginine codons in the mRNA. After all, it seemed reasonable to imagine that the partially unfolded fusion proteins might be attractive targets for endogenous proteases after they had been released from ribosomes. There is evidence for premature release when ribosomes encounter rare arginine codons (36). To investigate this possibility, we constructed an MBP–TEV protease expression vector (pRK685) in which the tandem pair of AGA codons was replaced by consecutive termination codons (TAA,TGA) (Fig. 1B). To our surprise, the truncated MBP–TEV protease fusion protein produced by this construct proved to be relatively stable in BL21(DE3) cells (Fig. 4, lane 8). In this experiment, a few extra bands with more rapid mobility than the full-length translation product were observed. All of these were greater or equal in size to native MBP (Fig. 4, lane 9), suggesting that they correspond to fusion proteins from which varying lengths of the TEV protease domain had been removed by intracellular proteolysis. Even so, the combined yield of the truncated MBP–TEV protease fusion proteins produced by pRK685 was comparable to that of the full-length fusion protein produced by the vector with no rare arginine codons and significantly greater than the yield of MBP–TEV protease fusion protein produced from pRK688 in the absence of pDC952. Thus, the degradation of partially synthesized MBP–TEV protease fusion proteins evidently is mechanistically coupled to the pausing of ribosomes at rare arginine codons.

The degradation of partially synthesized MBP–TEV protease fusion proteins is processive, because it results in the disappearance of the MBP domain of the fusion protein as well as the incompletely synthesized TEV protease domain. The cytosolic ATP-dependent proteases in *E. coli*, including ClpAP, ClpXP, and FtsH, are highly processive enzymes (37), and so we suspected that they might be involved in the degradation of nascent MBP–TEV protease polypeptides. Indeed, an important function of these proteases is to degrade incompletely synthesized proteins produced from damaged mRNAs (38). When ribosomes reach the end of an mRNA that does not have a stop codon, such as a truncated mRNA generated by the action of endogenous ribonucleases, the blockade can be alleviated by a trans-translation mechanism involving tmRNA (also called 10Sa RNA in *E. coli*) (39). This curious RNA has properties of both a tRNA and an mRNA. Aminoacylated tmRNA binds to the vacant A site in a stalled ribosome, whereupon it elicits a template switch that enables the ribosome to resume translation by using the tmRNA as a surrogate mRNA. A short open reading frame encoded by tmRNA is followed by a termination codon, which allows the ribosome to dissociate normally from the message. As a result of this process, 11 nonnative amino acids are added to the C-terminus of the partially synthesized polypeptide. The tmRNA-encoded peptide tag is rich in hydrophobic residues, and this has the effect of recruiting ATP-dependent proteases, which selectively attack proteins with hydrophobic residues at their extreme C-termini (40, 41). We reasoned that our results could be explained if tmRNA-mediated peptide tagging occurs when ribosomes become stalled in the

middle of an otherwise normal mRNA, due to depletion of the cognate tRNA. Indeed, tmRNA-mediated tagging at runs of consecutive AGA codons has been demonstrated experimentally (42).

To test this hypothesis, we expressed both the autoprocessing and the nonprocessing forms of the MBP–TEV protease fusion protein in an *ssrA*$^-$ host (*ssrA* is the gene that encodes tmRNA in *E. coli*) and compared the results with those obtained in an otherwise isogenic *ssrA*$^+$ strain. Initially, we chose the *E. coli* K12 strain X90 (43) for these experiments because we already had the *ssrA*::CAT lesion in this genetic background. Unexpectedly, we found that the presence of pDC952 did not improve the yield of MBP–TEV protease fusion protein in X90 cells (Fig. 5, lane 6). A low yield of fusion protein, similar to the amount produced in BL21(DE3) cells that do not contain pDC952, was obtained in X90 irrespective of the presence or absence of the accessory plasmid. The same result was obtained when the experiment was performed in X90(DE3) cells, demonstrating that the lambda prophage does not contribute to the effect (data not shown). This observation underscores the fact that there are significant differences between *E. coli* B and K12 strains.

No effort was made to determine whether or not the *argU* gene product was overproduced and functional in X90 cells. Instead, we decided to perform the experiment *E. coli* in BL21(DE3). Accordingly, the *ssrA*::CAT lesion in X90*ssrA*::CAT cells was introduced into BL21(DE3) by bacteriophage P1-mediated transduction. Genomic DNA from a putative transductant was analyzed by Southern blotting to verify that the *ssrA*::CAT lesion was present (Fig. 6A). The probe for this experiment was prepared by PCR amplification of the complete *ssrA* gene from X90 DNA. When genomic DNA from the putative transductant was digested with *Cla*I, only one fragment hybridized with the probe (Fig. 6A, lane 2). Similarly, the probe hybridized with only one *Cla*I fragment when genomic DNA from X90, X90*ssrA*::CAT, and BL21(DE3) cells was analyzed (Fig. 6A, lanes 3, 4, and 1, respectively). The hybridizing fragment from the putative transductant was exactly the same size (ca. 4.2 kb) as the hybridizing fragment from X90*ssrA*::CAT cells and clearly distinct from the hybridizing fragments produced by X90 and BL21(DE3) (ca. 9.5 and 6.5 kb, respectively). A similar result was obtained when the four genomic DNAs were digested with *Pst*I. However, in this case two fragments of both BL21(DE3)*ssrA*::CAT and X90*ssrA*::CAT DNA (ca. 2.5 and 1.5 kb) hybridized with the probe (Fig. 6A, lanes 6 and 8, respectively). This was expected, because *Pst*I sites bracket the chloramphenicol acetyltransferase (CAT) gene cassette that was used to create the insertional inactivation of the *ssrA* gene (44). Thus, our results clearly demonstrate that there is only one copy of the *ssrA* gene in BL21(DE3) cells and that the *ssrA*::CAT lesion is present in BL21(DE3)*ssrA*::CAT cells.

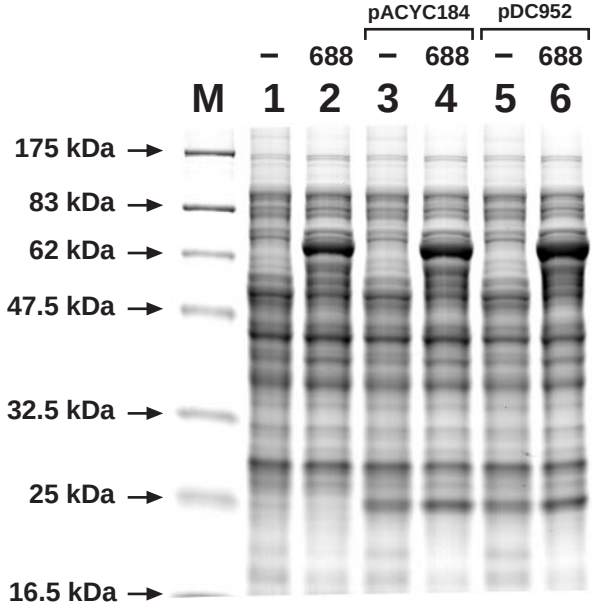

FIG. 5. pDC952 does not improve the yield of MBP–TEV protease fusion protein in *E. coli* X90 cells. Cells were grown to early log phase at 37°C and induced with IPTG for 3 h. Samples of the total intracellular protein from an equivalent number of cells were resolved by SDS–PAGE (10–20% Tris–glycine gel, Novex) and stained with GelCode Blue (Pierce). Lane 1, X90; lane 2, X90+ pRK688; lane 3, X90+ pACYC184; lane 4, X90+ pACYC184 + pRK688; lane 5, X90+ pDC952; lane 6, X90+ pDC952 + pRK688.

To further corroborate this conclusion, we also sought to confirm that BL21(DE3)*ssrA*::CAT cells are phenotypically tmRNA-deficient. It has been observed that certain *λimm*P22 hybrid phage fail to plate efficiently on *E. coli* strains that lack a functional *ssrA* gene (45, 46). Accordingly, we determined the plating efficiency of *λimm*P22 *dis c*2-5 on BL21(DE3) and BL21(DE3)*ssrA*::CAT cells. The results are summarized in Fig. 6B. Compared with its congenic *ssrA*$^+$ counterpart, the plating efficiency was reduced by almost five orders of magnitude in BL21(DE3)*ssrA*::CAT cells, which is consistent with the absence of tmRNA in this strain. This dramatic defect in *λimm*P22 *dis c*2-5 plating on BL21(DE3)*ssrA*::CAT cells could be complemented by a low copy number plasmid encoding a functional tmRNA gene (pRK593).

Curiously, we found that inactivation of the *ssrA* gene did not improve the yield of MBP-TEV protease fusion protein in BL21(DE3) cells (Fig. 7). On the contrary, even less of the fusion protein accumulated under these circumstances (Fig. 7, compare lanes 4 and 5). Furthermore, no truncated fusion proteins were observed to accumulate in BL21(DE3)*ssrA*::CAT cells. These results indicate that tmRNA is not involved in the processive

degradation of nascent MBP–TEV protease polypeptides. The Clp proteases (ClpXP and/or ClpAP) are responsible for most of the degradation of ssrA-tagged polypeptides in the cytosol of *E. coli* (40). To investigate whether the Clp proteases play a role in the degradation of nascent MBP–TEV protease fusion proteins, we also performed the experiment in SG1146a cells, a derivative of BL21(DE3) with an inactive *clpP* gene (*clpP*::CAT). Inactivation of ClpXP and ClpAP also reduced the level of MBP–TEV protease fusion protein (Fig. 7, lane 6) relative to what was obtained in BL21(DE3) cells (Fig. 7, lane 4), and no truncated fusion proteins were observed to accumulate in this genetic background either. It is not clear why the *ssrA* and *clpP* lesions reduce the yield of MBP–TEV protease fusion protein in BL21(DE3) cells. Nevertheless, this effect is exactly the opposite of what one would expect if either tmRNA or the Clp proteases are involved in the processive degradation of nascent MBP–TEV protease fusion proteins.

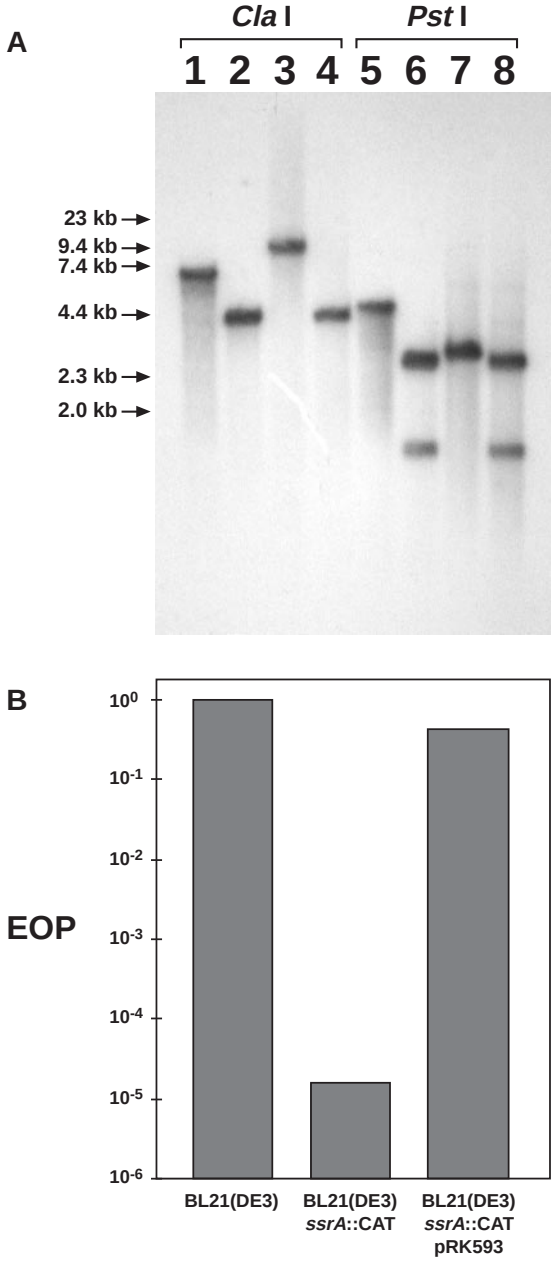

FIG. 6. Inactivation of the *ssrA* gene in BL21(DE3). (A) Genomic DNA from *E. coli* BL21(DE3) (lanes 1 and 5), BL21(DE3)*ssrA*::CAT (lanes 2 and 6), X90 (lanes 3 and 7), and X90*ssrA*::CAT (lanes 4 and 8) was digested with either *Cla*I or *Pst*I (as indicated), resolved by agarose gel electrophoresis, transferred to a nylon membrane, and probed with a labeled PCR product generated by amplification of the *ssrA* gene from X90 DNA. The positions of molecular weight standards are indicated. (B) Plating efficiency (EOP) of *λimm*P22 *dis c*2-5 on BL21(DE3), BL21(DE3)*ssrA*::CAT, and BL21(DE3)*ssrA*::CAT cells harboring a plasmid that produces tmRNA (pRK593).

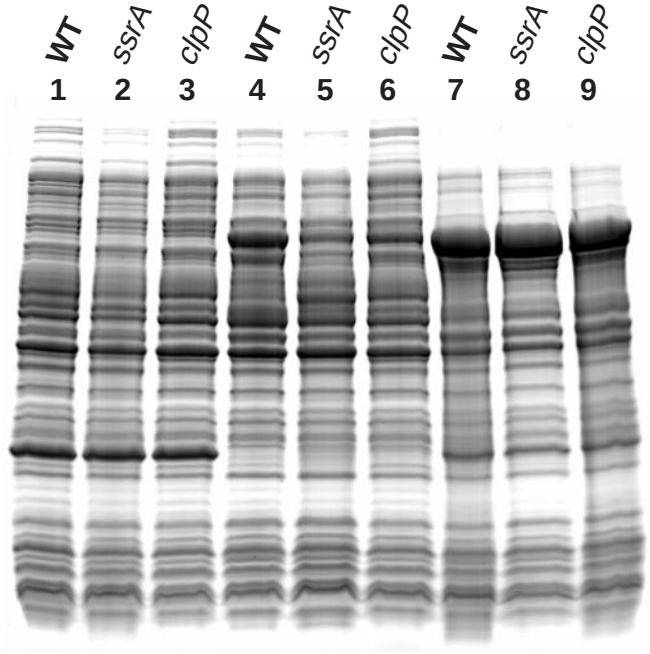

FIG. 7. Inactivation of *ssrA* or *clpP* does not improve the yield of MBP–TEV protease fusion protein in BL21(DE3). Cells were grown to early log phase at 37°C and induced with IPTG for 3 h. Samples of the total intracellular protein from an equivalent number of cells were resolved by SDS–PAGE (10–20% Tris–glycine gel, Novex) and stained with GelCode Blue (Pierce). WT, ssrA, and clpP refer to BL21(DE3), BL21(DE3)*ssrA*::CAT, and SG1146a cells, respectively. Cells contained either no expression vector (lanes 1–3), pRK688 only (lanes 4–6), or both pRK688 and pDC952 (lanes 7–9).

DISCUSSION

We have presented convincing evidence that the yield of TEV protease is limited by the presence in the mRNA of arginine codons that are rarely used in *E. coli*. An expression vector in which all of the AGG and AGA codons were changed to CGC produced a large amount

of TEV protease. The same result was obtained when *E. coli* cells contained, in addition to the TEV protease expression vector, an accessory plasmid carrying a copy of the cognate tRNA gene (*argU*). Together these results demonstrate that the yield of protease is affected by these codons at the translational level. By constructing a series of vectors in which only one or two AGA or AGG codons at a time were altered to CGC, we were able to show that the yield of TEV protease could be improved dramatically if only the tandem AGA codons were altered. Thus, we conclude that the low yield of protease can be attributed mainly to the contextual effect of tandem AGA codons rather than to the density of AGA and AGG codons in the mRNA. Our results lend additional support to the notion that clusters of rare codons are particularly deleterious to protein expression in *E. coli* (26).

We have also uncovered evidence that stalling of ribosomes at the tandem AGA codons in the TEV protease mRNA, which occurs when the supply of cognate aminoacyl tRNA is exhausted, triggers the processive proteolytic degradation of the nascent polypeptides. This effect is dramatic in BL21(DE3) cells but not in the *E. coli* K12 strain X90. Because runs of consecutive rare arginine codons in mRNA have been shown to stimulate tmRNA-mediated peptide tagging and subsequent degradation by ATP-dependent cytosolic proteases (42), we suspected that this mechanism was responsible for the low yield of MBP–TEV protease in BL21(DE3) cells. However, we unexpectedly found that the yield of MBP–TEV protease fusion protein was not affected by inactivation of either the *ssrA* gene or the *clpP* gene in BL21(DE3). A fusion protein produced from a gene in which the tandem AGA codons were replaced by tandem termination codons (pRK685) accumulated to a high level in BL21(DE3) cells, indicating that the truncated fusion protein is not intrinsically prone to processive proteolytic degradation *in vivo*. This result argues that proteolysis of MBP–TEV protease somehow is mechanistically linked to the presence of tandem AGA codons in the mRNA. Because proteolysis occurs in the absence of tmRNA, we believe that another mechanism must exist for coupling translational pausing and proteolytic degradation in BL21(DE3) cells. Further research will be required to illuminate the nature of this mechanism.

ACKNOWLEDGMENTS

We thank Stephanie Banach and Kelly Martin for technical assistance and Dr. Robert Sauer, Dr. Susan Gottesman, and Dr. David Friedman for providing bacterial strains X90*ssrA*::CAT, SG1146a, and phage *λimm*P22 *dis c*2-5, respectively.

REFERENCES

1. Uhlen, M., Forsberg, G., Moks, T., Hartmanis, M., and Nilsson, B. (1992) Fusion proteins in biotechnology. *Curr. Opin. Biotechnol.* **3,** 363–369.
2. Nygren, P.-A., Stahl, S., and Uhlen, M. (1994) Engineering proteins to facilitate bioprocessing. *Trends Biotechnol.* **12,** 184–188.
3. LaVallie, E. R., and McCoy, J. M. (1995) Gene fusion expression systems in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6,** 501–506.
4. Nilsson, J., Stahl, S., Lundeberg, J., Uhlen, M., and Nygren, P.-A. (1997) Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Express. Purif.* **11,** 1–16.
5. Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10,** 411–421.
6. Forsberg, G., Brobjer, M., Holmgren, E., Bergdahl, K., Persson, P., Gautvik, K. M., and Hartmanis, M. (1991) Thrombin and H64A subtilisin cleavage of fusion proteins for preparation of human recombinant parathyroid hormone. *J. Protein Chem.* **10,** 517–526.
7. Forsberg, G., Baastrup, B., Rondahl, H., Holmgren, E., Pohl, G., Hartmanis, M., and Lake, M. (1992) An evaluation of different enzymatic cleavage methods for recombinant fusion proteins, applied on des(1–3) insulin-like growth factor I. *J. Protein Chem.* **11,** 201–211.
8. He, M., Jin, L., and Austen, B. (1993) Specificity of factor Xa in the cleavage of fusion proteins. *J. Protein Chem.* **12,** 1–5.
9. Wagner, C. R., Bergstrom, C. P., Koning, K. R., and Hanna, P. E. (1996) Arylamine N-acetyltransferases: Expression in *Escherichia coli*, purification, and substrate specificities of recombinant hamster monomorphic and polymorphic isozymes. *Drug Metab. Dispos.* **24,** 245–253.
10. Tsao, K.-L., and Waugh, D. S. (1997) Balancing the production of two recombinant proteins in *Escherichia coli* by manipulating plasmid copy number: high-level expression of heterodimeric Ras farnesyltransferase. *Protein Express. Purif.* **11,** 233–240.
11. Stevens, R. C. (2000) Design of high-throughput methods of protein production for structural biology. *Structure Fold Des.* **8,** R177–R185.
12. Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A., and Dougherty, W. G. (1994) Release of proteins and peptides from fusion proteins using a recombinant plant virus protease. *Anal. Biochem.* **216,** 413–417.
13. Parks, T. D., Howard, E. D., Wolpert, T. J., Arp, D. J., and Dougherty, W. G. (1995) Expression and purification of a recombinant tobacco etch virus NIa proteinase: Biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology* **210,** 194–201.
14. Lucast, L. J., Batey, R. T., and Doudna, J. A. (2001) Large-scale purification of a stable form of recombinant tobacco etch virus protease. *Biotechniques* **30,** 544–554.
15. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1991) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185,** 60–89.
16. Waugh, D. S. (1996) Genetic tools for selective labeling of proteins with α-^{15}N-amino acids. *J. Biomol. NMR* **8,** 184–192.
17. Kapust, R. B., and Waugh, D. S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8,** 1668–1674.
18. Fox, J. D., Kapust, R. B., and Waugh, D. S. (2001) Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. *Protein Sci.* **10,** 622–630.

19. Garcia, G. M., Mar, P. K., Mullin, D. A., Walker, J. R., and Prather, N. E. (1986) The *E. coli* dnaY gene encodes an arginine transfer RNA. *Cell* **45,** 453–459.
20. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77,** 51–59.
21. Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Res.* **25,** 1203–1210.
22. Miller, J. H. (1972) "Experiments in Molecular Genetics," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
23. Sambrook, J., and Russell, D. W. (2001) "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Press, Cold Spring Harbor, NY.
24. Pryor, K. D., and Leiting, B. (1997) High-level expression of soluble protein in *Escherichia coli* using a His_6-tag and maltose-binding-protein double-affinity system. *Protein Express. Purif.* **10,** 309–319.
25. Riggs, P. (2000) Expression and purification of recombinant proteins by fusion to maltose-binding protein. *Mol. Biotechnol.* **15,** 51–63.
26. Kane, J. F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli. Curr. Opin. Biotechnol.* **6,** 494–500.
27. Chen, G.-F., and Inouye, M. (1994) Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli. Genes Dev.* **8,** 2641–2652.
28. Imamura, H., Jeon, B.-S., Wakagi, T., and Matsuzawa, H. (1999) High level expression of *Thermococcus litoralis* 4-α-glucanotransferase in a soluble form in *Escherichia coli* with a novel expression system involving minor arginine tRNAs and GroELS. *FEBS Lett.* **457,** 393–396.
29. Hua, X., Want, H., Chen, D., and Zhu, D. (1994) Enhancement of expression of human granulocyte-macrophage colony stimulating factor by *argU* gene product in *Escherichia coli. Biochem. Mol. Biol. Int.* **32,** 537–543.
30. Brinkmann, U., Mattes, R. E., and Bucket, P. (1989) High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product. *Gene* **85,** 109–114.
31. Spanjaard, R. A., Chen, K., Walker, J. R., and van Duin, J. (1990) Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: Assigning a codon to *argU* tRNA and T4 $tRNA^{Arg}$. *Nucleic Acids Res.* **18,** 5031–5035.
32. Zahn, K., and Landy, A. (1995) Modulation of lambda integrase synthesis by rare arginine tRNA. *Mol. Microbiol.* **21,** 69–76.
33. Komar, A. A., Fuillemet, E., Reiss, C., and Cullin, C. (1998) Enhanced expression of the yeast Ure2 protein in *Escherichia coli*: The effect of synonymous codon substitutions at a selected place in the gene. *Biol. Chem.* **379,** 1295–1300.
34. Truscott, K. N., and Scopes, R. K. (1998) Sequence analysis and heterologous expression of the groE genes from *Thermoanaerobacter* sp. Rt8.G4. *Gene* **217,** 15–23.
35. Hu, X., Shi, Q., Yang, T., and Jackowski, G. (1996) Specific replacement of consecutive AGG codons results in high-level expression of human cardiac troponin T in *Escherichia coli. Protein Express. Purif.* **7,** 289–293.
36. Gao, W., Tyagi, S., Kramer, F. R., and Goldman, E. (1997) Messenger RNA release from ribosomes during 5′-translational blockage by consecutive low-usage arginine but not leucine codons in *Escherichia coli. Mol. Microbiol.* **25,** 707–716. [Erratum (1998) **27,** 669]
37. Wickner, S., Maurizi, M. R., and Gottesman, S. (1999) Posttranslational quality control: Folding, refolding, and degrading proteins. *Science* **286,** 1888–1893.
38. Karzai, A. W., Roche, E. D., and Sauer, R. T. (2000) The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nature Struct. Biol.* **7,** 449–455.
39. Keiler, K. C., Waller, P. R., and Sauer. R. T. (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271,** 990–993.
40. Gottesman, S., Roche, E., Zhou, Y., and Sauer. R. T. (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12,** 1338–1347.
41. Herman, C., Thevenet, D., Bouloc, P., Walker, G. C., and D'Ari, R. (1998) Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). *Genes Dev.* **12,** 1348–1355.
42. Roche, E. D., and Sauer, R. T. (1999) SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. *EMBO J.* **18,** 4579–4589.
43. Amann, E., Brosius, J., and Ptashne, M. (1983) Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in *E. coli. Gene* **25,** 167–178.
44. Kirby, J. E., Trempy, J. E., and Gottesman, S. (1994) Excision of a P4-like cryptic prophage leads to alp protease expression in *Escherichia coli. J. Bacteriol.* **176,** 2068–2081.
45. Retallack, D. M., Johnson, L. L., and Friedman, D. I. (1994) Role for 10Sa RNA in the growth of lambda-P22 hybrid phage. *J. Bacteriol.* **176,** 2082–2089.
46. Karzai, A. W., Susskind, M. M., and Sauer, R. T. (1999) SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). *EMBO J.* **18,** 3793–3799.