Identification of endogenous SsrA-tagged proteins reveals tagging at positions corresponding to stop codons

Eric D. Roche and Robert T. Sauer‡

Department of Biology
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

* This work was supported by National Institute of Health grant AI-16892.

‡ Corresponding author: E-mail bobsauer@mit.edu, fax 617-258-0673

running title: Identification of endogenous SsrA-tagged proteins
The SsrA•SmpB quality control system adds a C-terminal degradation peptide (AANDENYALAA) to nascent chains on stalled ribosomes, thereby freeing the ribosome and ensuring proteolysis of the tagged protein. An SsrA mutant with the tag sequence AANDEHHHHHH was used to slow degradation and facilitate Ni\(^{2+}\)-NTA affinity purification. Display of affinity-purified *E. coli* proteins on 2D gels revealed small quantities of a diverse set of SsrA-H\(_6\) tagged proteins, and mass spectroscopy identified LacI repressor, λ cI repressor, YbeL, GalE, RbsK, and a SlyD-kan\(^R\) fusion protein as members of this set. For λ repressor and YbeL, the SsrA-H\(_6\) tag was added after the natural C-terminus of the protein, suggesting that tagging occurred while the ribosome idled at the termination codon of these genes. Potential causes of tagging for the other proteins include interference from translation of downstream reading frames, rare codons, and gene disruption. These and previous results support a broad role for the SsrA•SmpB system in freeing stalled ribosomes and in directing degradation of the products of these frustrated protein synthesis reactions.
INTRODUCTION

Cells have evolved multiple mechanisms to prevent or correct errors in macromolecular synthesis. In bacteria, the SsrA\textbullet SmpB quality control system helps to insure proper protein synthesis by rescuing stalled ribosomes and by marking the incomplete proteins associated with these ribosomes for destruction (1). SsrA is an RNA molecule (363 bp in \textit{Escherichia coli}) with a tRNA-like domain that can be charged with alanine and a short mRNA-like reading frame encoding the peptide sequence ANDENYALAA (2-4; Fig. 1A). Together with SmpB, an associated and required protein factor, SsrA appears to be recruited when ribosomes stall on mRNAs during translation (5,6). A cotranslational tagging process then occurs in which SsrA first acts as tRNA, donating its charged alanine to the stalled peptide chain, then displaces the problematic message, and finally serves as mRNA, directing synthesis of the last ten residues of the AANDENYALAA peptide tag. SsrA recruits the termination factors required to release the tagged protein, and the modified protein is subsequently recognized and degraded by specific proteases (5,7,8).

The SsrA\textbullet SmpB system is universally conserved among bacteria, and a variety of phenotypes are observed in mutant strains (1,9). Deletions of SsrA or SmpB are lethal in some bacteria and cause slow growth under stress conditions in other bacteria (10-12). In \textit{E. coli}, for instance, the growth defects of strains lacking SsrA are magnified at high temperatures or under carbon starvation (2,13), suggesting that quality assurance becomes more important when the incidence of translational errors is elevated. Other phenotypes of SsrA-defective \textit{E. coli} include reduced motility (2), increased protease activity (14), defects in plating of phage \textit{λimmP22} and induction of temperature-sensitive lysogens of phage Mu (6,15), and increased levels of LacI, \textit{λ cI}, and LexA repression (16).

What features of mRNAs lead to ribosome stalling and SsrA-recruitment? SsrA-tagging has been demonstrated for proteins translated from truncated mRNAs that lack in-frame stop codons (5). A message of this type is translated to its 3’ end, where the ribosome stalls because there is no termination codon to recruit release factors. Tagging of \textit{E. coli} LacI repressor appears to occur by this mechanism, with truncation of the mRNA occurring because Lac repressor binds to...
an operator site within the lacI gene and blocks completion of transcription (17). In a SsrA-defective cell, active Lac repressor missing a short, non-essential C-terminal region is synthesized from the truncated transcript. This presumably explains why Lac repressor activity is elevated in SsrA-deletion strains. SsrA-tagging of model substrates has also been shown to occur when their mRNAs contain rare-codon repeats (18). In this case, the principal site of tagging corresponded to the first rare codon of the repeat, and tagging could be suppressed by overproduction of the cognate tRNA, suggesting that ribosome stalling at the rare codon repeat leads to recruitment of SsrA and subsequent tagging.

At present, Lac repressor is the only normal E. coli protein for which SsrA-tagging has been demonstrated (17). Moreover, although it seems likely that prolonged ribosome stalling may cause SsrA-tagging in cases other than truncated mRNAs or runs of rare codons, this has not been demonstrated. SsrA-tagging of proteins in vivo is difficult to probe directly because the tagged proteins are degraded. To obviate this problem, a mutant (SsrA-DD) with base substitutions that change the C-terminus of the peptide tag from AA to DD has been used, with antibodies specific for the mutant tag allowing detection of tagged proteins (17,18). Studies of this type have revealed low-level tagging of at least 20-30 proteins in E. coli, but the actual number and identities of these tagged proteins are unknown. In this paper, we have constructed a novel SsrA mutant that adds a His$_6$ affinity tag to modified proteins. Affinity purification of these tagged proteins followed by separations on two-dimensional gels has allowed us to estimate the extent of tagging and to identify additional proteins that are subject to SsrA-tagging. In two cases, characterization of the tagged product by mass spectroscopy has shown that the tag is added immediately after the normal C-terminal residue of the protein, suggesting that the ribosome stalls and tagging occurs at a position corresponding to the normal termination codon.

**EXPERIMENTAL PROCEDURES**

Microbiology—E. coli strain ER1000A is ssrA::cat slyD::kan ara Δ(lac pro) nalA argEam rif thi-1 / F' lac$^+$ pro$^+$ lacI$^+$ and was constructed by P1 transduction of the ssrA::cat gene
disruption from strain X90 srrA::cat (5) into strain X90 slyD::kan (BB100; Bronwen Brown, unpublished; 19). Strain W3110 srrA::kan has been previously described (2), as have plasmids pKW1 (control), pKW11 (wild-type SsrA), pKW22 (SsrA-DD), and pKW23 (SsrA-DD), each of which contains a p15a origin of replication and a tetracycline resistance gene (18). Plasmid pKW24 (SsrA-H$_{6}$) is identical to pKW11 except for the mutations indicated in Figure 1 and was constructed in collaboration with Kelly Williams. Plasmid pPW500 expresses the λ-cI-N-trpAt message with no in-frame stop codon (5). Plasmid pTMK1 (Theresa Karplus, unpublished) is a pBR322 derivative (amp$^{R}$) that encodes the λ cI repressor gene under lac promoter control from pKB280 (20). Plasmid pTrc-99a was obtained from Amersham Pharmacia Biotech. Phage λimmP22 dis c2-5 (15) was plated on strain W3110 srrA::kan transformed with pKW1, pKW11, pKW22, or pKW24. Growth at 37 °C or 43 °C of strain W3110 srrA::kan containing pKW1, pKW11, or pKW24 was assayed by changes in OD$_{600}$; the values from two independent cultures were averaged and fit to an exponential function for determination of growth rates. All strains were grown in Luria-Bertani (LB) media containing 24 µg/ml tetracycline unless indicated.

SsrA-H$_{6}$ Tagging of λ-cI-N-trpAt—Strain X90 srrA::cat transformed with pPW500 and either pKW1, pKW11, pKW23, or pKW24 was grown in LB broth containing 100 µg/ml ampicillin and 50 µg/ml tetracycline to an OD$_{600}$ of roughly 0.8 and expression of λ-cI-N-trpAt was induced by addition of IPTG to 1 mM. After one hour, samples were removed to ice and the cells were collected and frozen at −80 °C. Thawed cell pellets were lysed with SDS sample buffer, heated to >95°C, and electrophoresed on a 15% polyacrylamide SDS Tris-tricine gel, followed by transfer to a PVDF membrane. The blot was probed with M2 anti-FLAG antibodies (Eastman Kodak) to detect the FLAG epitope present in the λ-cI-N-trpAt protein.

Detection and Purification of Tagged λ Repressor—Strain ER1000A transformed with pKW1 and pTMK1, pKW24 and pTrc-99a, or pKW24 and pTMK1 was grown in LB broth containing 100 µg/ml ampicillin and 20 µg/ml tetracycline to an OD$_{600}$ of roughly 0.5 and 1 mM IPTG was added to induce expression of λ repressor from the lac promoter. After two hours, samples were removed, processed as described above, electrophoresed on a 12% polyacrylamide
SDS Laemmli gel, and transferred to a PVDF membrane. Different portions of the resulting blot were probed with either His probe H-15 HRP antibodies (Santa Cruz Biotechnology) or 51F 7G11 antibodies to λ cI repressor (21). SsrA-H₆ tagged λ repressor was purified by Ni²⁺-NTA chromatography under denaturing conditions according to the Qiagen protocol with minor modifications. All purification steps were performed in 6 M guanidine•Cl (Mallinckrodt). Binding and wash buffers contained 10 mM imidazole, and the elution buffer contained 0.1 M acetic acid. The pH of the eluate was increased by addition of 1/10 volume of 2 M Tris (pH 9.5), and the eluted proteins were concentrated by a second round of Ni²⁺-NTA chromatography. The final eluate was dialyzed against 10 mM Tris•Cl (pH 8) and assayed by SDS-PAGE. The SsrA-H₆ tagged repressor was analyzed by MALDI-TOF mass spectrometry using a PerSeptive Biosystems Voyager instrument with sinapinic acid as the matrix and horse heart myoglobin and chymotrypsinogen A (Sigma) as internal standards. Mass fingerprinting of tryptic peptides from in-gel digests of tagged repressor was performed as described below and without alkylation.

Purification and 2D Analysis of SsrA-tagged E. coli Proteins—For experiments comparing SsrA-H₆ (pKW24) cells to those lacking SsrA (pKW1), multiple 1 L cultures were grown to an OD₆₀₀ of 0.8-0.9 in 6 L flasks and purified through two rounds of Ni²⁺-NTA chromatography. To obtain more concentrated proteins for identification, 10 L fermentor cultures of ER1000A/pKW24 were grown to an OD₆₀₀ of 1.8 and three rounds of Ni²⁺-NTA purification were performed. The pattern of tagged proteins obtained by both methods was similar. Purifications performed either in 6 M guanidine•Cl or in deionized 8 M urea (Mallinckrodt or Amersham Pharmacia Biotech, both >99.5% purity) also gave similar results.

Cells from a typical fermentor culture were suspended in 200 ml lysis buffer (6 M guanidine•Cl, 1% triton, 5 mM β-mercaptoethanol (βME), 100 mM NaH₂PO₄, 10 mM Tris, pH 8) and lysed by stirring for an hour. Cell debris was removed by centrifugation for 30 min at >30,000 g, and the supernatant was mixed with 6 ml of a Ni²⁺-NTA slurry equilibrated in lysis buffer and allowed to bind with rocking for one hour. The resin was collected by centrifugation and washed twice with 40 ml of lysis buffer. Proteins were eluted with 16 ml of elution buffer (6 M
guanidine•Cl, 1% triton, 1 mM βME, 0.1 M acetic acid) and the pH was raised by adding 1/10 volume of 2 M Tris (pH 9.5). This eluate was rebound to 1.5 ml Ni\(^{2+}\)-NTA slurry in wash buffer (lysis buffer containing 1 mM βME) by rocking for an hour. The resin was collected and washed twice with 10 ml of wash buffer, bound protein was eluted with 4 ml elution buffer, and the pH was raised by Tris addition. In the final round of purification, proteins were bound to 1 ml of resin slurry, washed twice with 6 ml of wash buffer, and eluted in small steps resulting in recovery of the majority of protein in a single 0.4 ml fraction.

Two-dimensional electrophoresis was performed according to protocols in the Amersham Pharmacia manuals. Equal volumes of Ni\(^{2+}\)-NTA purified protein and rehydration solution were mixed with 18 cm, pH range 4-7 IPG strips. Following rehydration, isoelectric focusing was performed on a Multiphor II unit. Proteins were alkylated by soaking the IPG strips in SDS equilibration solution containing iodoacetamide as described. Purification in the second dimension was performed by electrophoresis on a 10-18% gradient polyacrylamide gel with SDS Laemmli buffer using the Hoefer Dalt system. Silver staining was performed essentially as described (22). 2D western blots were also performed on the Dalt system, and were probed with His probe H-15 HRP antibodies.

Identification of SsrA-tagged Proteins— Silver-stained gels were dried, and protein spots were excised and processed for peptide mass fingerprinting as described (22) except alkylation was not repeated and digestions of proteins in the gel used 80 ng/µl of modified trypsin (Roche Diagnostics). Tryptic peptides were eluted from the gel by multiple extractions with 5% formic acid and 50% acetonitrile, dried under vacuum, resuspended in 0.1% TFA, and bound to C18 ZipTips (Millipore). Samples were eluted with a 50-75% acetonitrile solution containing α-cyano-4-hydroxy-cinnamic acid and were spotted directly on target plates for MALDI-TOF mass spectrometry using a PerSeptive Biosystems Voyager instrument. Calibration of mass spectra was typically performed using trypsin autodigestion peaks as internal standards. Spectra were compared to controls obtained by processing an empty portion of the gel, and only peaks significantly above background were selected. Peptide fingerprint searches were performed using
MS-Fit (http://prospector.ucsf.edu), searching the SwissProt database of *E. coli* sequences of all molecular weights and pIs. The PAWS program (ProteoMetrics) was used to help identify junction peptides containing the SsrA-H<sub>6</sub> tag.

RESULTS

*SsrA-H<sub>6</sub> is a Functional Mutant*—The reading frame of the SsrA degradation tag was modified by directed mutagenesis to encode HHHHHH as its C-terminal residues (Fig. 1A). This mutant RNA, called SsrA-H<sub>6</sub>, was designed to have the same length of coding sequence as wild-type SsrA and to maintain base-pairing within the stem-loop region that begins with the final codons of the tag sequence (Fig. 1A). We expected that the His<sub>6</sub> epitope on proteins tagged by SsrA-H<sub>6</sub> would slow degradation and also allow affinity purification.

SsrA-H<sub>6</sub> RNA showed biological activity in several different assays. Active SsrA is required for efficient plating of the hybrid phage λimmP22 (15). An SsrA-deletion strain expressing SsrA-H<sub>6</sub> RNA from a plasmid supported λimmP22 growth at about 50% of the efficiency of a comparable strain expressing plasmid-borne wild-type SsrA, whereas the deletion strain showed a 10,000-fold reduction in plaque formation (Fig. 1B). Expression of SsrA-H<sub>6</sub> RNA also largely complemented the slow growth of SsrA-deficient *E. coli* at 43 °C (Fig. 1C) and at 37 °C, where the growth defect is subtle (Fig. 1C, inset).

Co-expression of SsrA-H<sub>6</sub> RNA and λ-ci-N-trpAt mRNA, an artificial substrate expressing the N-terminal domain of λ repressor without in-frame termination codons (5), resulted in tagging of the λ-ci-N-trpAt protein (Fig. 1D, lane a, upper band). This tagged band was not observed in the absence of SsrA-H<sub>6</sub> (Fig. 1D, lane c). As judged by the ratio of tagged to untagged products, SsrA-H<sub>6</sub> tagging of λ-ci-N-trpAt was at least as efficient as tagging by the SsrA-DD mutant (Fig. 1D, lane b). However, tagging mediated by these plasmid-borne SsrA mutants is less efficient than tagging mediated by plasmid-borne or chromosomal wild-type SsrA (18; essentially no untagged protein appears in Fig. 1D, lane d). This experiment also shows that proteins with the AANDEHHHHHH tag are not subject to rapid degradation, in contrast to λ-ci-N-trpAt protein
tagged by wild-type SsrA (5; Fig. 1D, lane d). Hence, although SsrA-H<sub>6</sub> tagged proteins are not degraded rapidly, the SsrA-H<sub>6</sub> mutant is still biologically active. This result is consistent with previous studies that have shown that the ribosome-rescue functions of SsrA are more important for biological function than the subsequent degradation of tagged proteins (23).

*SsrA-H<sub>6</sub> Tags λ Repressor—* As noted above, λ repressor activity is increased in SsrA-defective strains (16). To test if this phenotype might be a consequence of tagging by SsrA, we co-expressed SsrA-H<sub>6</sub> RNA and wild-type λ repressor from an overproducing plasmid. Western blots of cell lysates were then probed with anti-His<sub>6</sub> antibodies, revealing a major tagged band (Fig. 2, lane b). This band was shown to be tagged λ repressor by several criteria. First, it was not observed if either SsrA-H<sub>6</sub> or λ repressor was absent (Fig. 2, lanes a & c). Second, a band with the same electrophoretic mobility also cross reacted with anti-λ repressor antibodies (Fig. 2, lane e) and was missing if either SsrA-H<sub>6</sub> or λ repressor was not expressed (Fig. 2, lanes d & f). The SsrA-H<sub>6</sub> tagged λ repressor was present at roughly 3-5% of the level of untagged λ repressor as judged by the intensities of both bands in western blots. Expression of SsrA-H<sub>6</sub> also resulted in the tagging of several additional proteins (Fig. 2, lanes b & c) that were missing in the absence of SsrA-H<sub>6</sub> (Fig. 2, lane a).

Cell lysates containing the SsrA-H<sub>6</sub> tagged λ repressor were purified by several cycles of chromatography on Ni<sup>2+</sup>-NTA resin under denaturing conditions. SDS gels of the final eluate showed a major band corresponding in size to SsrA-H<sub>6</sub> tagged λ repressor (data not shown). A few proteins of lower molecular were present in the eluate, but these species were also observed in a parallel purification from cells lacking SsrA (data not shown). MALDI-TOF mass spectroscopy of the eluate gave one major peak above 15 kD with a mass of 27437 ± 42 Da, which corresponded within error to that expected for the complete λ repressor sequence plus the 11-residue AANDEHHHHHHH tag (27404 Da; Fig. 2). The tagged repressor was digested in-gel with trypsin and the masses of eluted peptides were determined. A tryptic peptide of mass 2687.7 ± 0.8 Da was observed that did not occur in the normal repressor sequence but matched a junction peptide including the C-terminal tryptic fragment of λ repressor plus the SsrA-H<sub>6</sub> tag (2687.8 Da; Fig. 2).
These observations indicate that tagging occurs after the normal C-terminal residue of the λ repressor protein. Tagging at this position suggests that when ribosomes stop translation at the UGA codon at the 3’ end of the λ cI gene, SsrA-tagging competes with the termination and release mechanism mediated by release factor 2.

**Complexity of SsrA-tagging in E. coli**— How many *E. coli* proteins are tagged by the SsrA•SmpB system? To address this question, Ni$^{2+}$-NTA purifications involving two to three rounds of chromatography under denaturing conditions were performed in parallel, starting with lysates of log-phase cultures of cells containing or lacking SsrA-H$_6$. The resulting eluates were further purified by two-dimensional gel electrophoresis with the focusing dimension ranging from pH 4 to 7, which includes the isoelectric points of most *E. coli* proteins. Silver staining of the 2D gel from the control purification (no SsrA-H$_6$) showed a number of spots (Fig. 3A) even though the gene for SlyD, a His-rich protein that binds to Ni$^{2+}$-NTA, was disrupted in the strain used for these studies. By comparison, however, far more proteins were observed in the 2D gel from the strain containing SsrA-H$_6$ (Fig. 3B). Moreover, most proteins observed in the SsrA-H$_6$ strain but not the control strain cross reacted with anti-His$_6$ antibodies in a Western blot (Fig. 3C). These results indicate that a complex mixture of tagged proteins is generated in *E. coli*, even under non-stress conditions where nutrients are plentiful.

The mass of SsrA-H$_6$ tagged proteins obtained by affinity purification represented roughly 0.01 to 0.05% of the total mass of cellular proteins as judged by the intensities of stained gels (data not shown). However, estimates of the abundance and number of tagged proteins may be somewhat low, because tagged proteins of high molecular weight seemed to be selectively lost during purification (data not shown), and SsrA-H$_6$ is less active in tagging than wild-type SsrA (Fig. 1D). The presence of shadow spots at slightly lower pI’s for many of the proteins in the control and SsrA-H$_6$ samples (Fig. 3) is also a complicating factor. Nevertheless, whether these shadow spots are produced by modifications *in vivo* or during purification, it is clear that a large number of species with different isoelectric points and molecular weights are generated by tagging mediated by the SsrA•SmpB quality control system.
Identification of Tagged Proteins—Roughly 20 of the most prominent SsrA-H$_6$ tagged proteins from the 2D gels were excised, digested with trypsin, and analyzed by MALDI-TOF mass spectrometry. The resulting sets of peptide masses were then used to search a database for potential matches. Several differences from a standard proteomics analysis are noteworthy. First, tagging at internal positions will lead to truncated proteins, and fewer peptides will therefore be available to match the wild-type sequence. Second, although characterization of the junction peptide containing both the protein and the tag sequences could determine the site of tagging, the mass of this peptide cannot be predicted a priori and it is possible that tagging could occur at different sites resulting in a heterogeneous set of junction peptides. In the end, we were able to identify five SsrA-tagged proteins, including YbeL (spot A; Fig 3B), LacI (spot B; Fig. 3B), SlyD (spot C; Fig. 3B), GalE (spot D; Fig. 3B), and RbsK (spot E; Fig. 3B) with a high degree of confidence using the search program MS-Fit (http://prospector.ucsf.edu). Table I lists the peptide matches, fractional coverage, and identified peptide closest to the C-terminus for each of these proteins.

The peptide matches for YbeL (residues 4-155), which is encoded by an open reading frame in the LeuS-GltL region, extended to within five residues of its C-terminus (Table I). Two prominent peaks in the mass spectrum did not match the YbeL primary sequence (Fig. 4A) but had masses expected for junction peptides containing YbeL residues 156-160 or 148-160 plus the AANDEHHHHHHH tag. These results indicate that YbeL tagging, like $\lambda$ repressor tagging, occurs when ribosomes stop at the normal termination codon (Fig. 4A).

Junction peptides were not identified for LacI, GalE, RbsK, or SlyD. For the first three proteins, however, peptide matches extending very close to the C-terminus were identified (Table I). Thus, tagging is also likely to occur near the C-terminal ends of the LacI, GalE, and RbsK proteins. As noted above, tagging of LacI repressor has been demonstrated previously (17). Although some tagging of LacI must occur near the C-terminus because peptides within 9 residues of this site are recovered (Table I), tagged LacI proteins are present in our 2D gels as a ladder of spots of decreasing molecular weight with similar isoelectric points (Fig. 3B; spots B). This observation suggests that tagging at sites more distant from the C-terminus of LacI is also likely to occur. The
most C-terminal peptide matching tagged SlyD (Fig 3B; spots C) occurred roughly two-thirds of the way through the slyD gene (Table I), near the position of a kanamycin resistance insertion present in the strain used for these studies.

Two *E. coli* proteins, NikR and Fur, that were present in both the control and SsrA-H₆ 2D gels were also identified (Table I). In each case, the normal C-terminal peptide of the protein was identified. Histidine-rich sequences are present in both NikR (HHHH) and Fur (HHHDH), which presumably accounts for the binding of these proteins to the Ni²⁺-NTA resin.

**DISCUSSION**

The studies presented here demonstrate SsrA-tagging following the C-terminal residues of two different proteins, λ repressor and YbeL. SsrA-tagging generally occurs at sites of ribosome stalling (1,5,18). Our results indicate that tagging can also ensue when the ribosome stops at the termination codons of the λ cI and *E. coli* ybeL genes. Normally, translational release factors are recruited when a ribosome reaches the termination codon of a gene (24). If this process is inefficient for a particular gene, however, then some fraction of the nascent chains may be tagged and degraded rather than being released. In the case of λ repressor, stop-codon dependent tagging and subsequent degradation of the tagged repressor molecules could explain why decreased λ repressor activity is observed in strains in which the SsrA•SmpB system is active (16). In the absence of SsrA, a higher level of release-factor catalyzed termination (or possibly suppression or frameshifting at the stop codon) would presumably produce additional λ repressor. In the studies reported here, SsrA-H₆ tagging of λ cI occurred at a level of only 3-5% of untagged protein, and it may seem unlikely that this level of modification would be detectable in a functional assay. However, tagging by wild-type SsrA is more efficient than tagging by SsrA-H₆ (the exact difference is uncertain because protein tagged by wild-type SsrA is rapidly degraded), and even small increases in λ repressor concentration could have significant effects on activity because operator binding by λ repressor is a highly cooperative process (25). The biosynthesis of λ repressor is known to be controlled tightly by mechanisms that affect transcriptional and translational initiation.
Regulating translational termination may provide another means of ensuring that intracellular λ repressor levels are maintained within narrow limits.

The efficiency of translation termination is affected by the identity of the stop codon and the following 3’ base and by the final sense codons of the gene and the 5’ bound tRNA (26-28, and references therein). The λ repressor cI gene terminates with a glycine sense codon (GGC) and an opal (UGA) stop codon (29). Opal codons are generally less efficient than either amber (UAG) or ochre (UAA) codons in mediating termination (27). The ybeL gene terminates with a proline sense codon (CCG) and an ochre stop codon (30). Ochre codons are generally thought to be efficient in termination but CCG_UAA is underrepresented in E. coli genes (31). Interestingly, P22 c2 repressor—the phage P22 ortholog of λ repressor—also ends with a proline codon and ochre stop codon (32). We assume that ribosomal stalling at the ybeL or λ cI stop codons is mediated, at least in part, by nearby mRNA sequence context. Another speculative possibility is that termination efficiency might be modulated by the binding of additional protein factors. This strategy might allow cells to regulate the cotranslational tagging and degradation of specific gene products under suitable conditions. It is also possible, of course, that stop-codon dependent tagging simply occurs at a low level on many reading frames in competition with normal termination. In this regard, many studies of translation termination use suppression of the stop codon by tRNAs to assay efficiency. If SsrA-tagging occurs more readily than suppression, however, then some inefficient stop codons might appear to be effective despite a low level of proper termination.

GalE (UDP-galactose 4-epimerase) is encoded by the first of 4 tightly spaced genes in the galactose operon (30), and a site of SsrA-tagging is within 7 residues of the C-terminus of the GalE protein (Table I). The stop codon of the galE gene and the start codon of the next gene, galT, are separated by 9 bases. In fact, the Shine-Dalgarno sequence for galT overlaps the stop codon of galE (Fig. 4B). Because translational initiation is usually slow relative to elongation (33), it is possible that a ribosome initiation complex formed on the ribosome binding site of galT blocks ribosomes translating the 3’ region of galE, resulting in the stalling of these ribosomes and subsequent truncation and SsrA-tagging of some GalE proteins. RbsK (ribokinase), which is
tagged within 2 residues of the C-terminus, is part of the ribose operon. The \textit{rbsK} gene is separated from the downstream gene, \textit{rbsR}, by just 3 bases (30). Hence, interference between ribosomes initiating translation of RbsT and terminating translation of RbsK could also explain tagging in this case. However, the terminal and antepenultimate codons of the \textit{rbsK} gene are AGG (Fig. 4C), which encodes arginine and is the rarest codon in \textit{E. coli}. Thus, ribosome stalling at the terminal AGG might also explain SsrA-tagging of RbsK. The GalE and RbsK proteins could also be tagged at their C-termini because of ribosome stalling at the termination codons of the corresponding mRNAs.

Our studies confirm those of Abo et al. (2000), showing that LacI repressor is subject to SsrA-tagging. One additional bit of information added by our work is that tagging of some Lac repressor molecules occurs within 9 residues of the C-terminus. The \textit{lacI} gene contains an operator site that overlaps its termination codon, and inactivation of Lac repressor with IPTG prevents SsrA-tagging (17). Binding of Lac repressor to this operator site is proposed to lead to transcriptional termination before the normal termination codon, with SsrA-tagging then occurring when ribosomes stall at the 3’ end of the mRNA. The heterogeneity that we observe for SsrA-tagged LacI probably results from heterogeneity in the population of mRNA molecules, caused either by transcription termination at different positions or by exonuclease digestion. Although the underlying causes of SsrA-tagging seem to be different for Lac repressor and \textit{\lambda} repressor, it is interesting that the level of repressor activity, in each case, appears to be reduced by tagging and degradation.

The final protein identified as a substrate for SsrA-tagging was SlyD. We were initially surprised by this result because a \textit{slyD::kan} strain had been used to avoid purification of the histidine-rich, wild-type SlyD protein. The peptide fingerprint of SsrA-H$_6$ tagged SlyD matched most of the N-terminal 140 residues of the protein but none of the final 56 residues (Table I). The \textit{slyD::kan} disruption was constructed by inserting a kanamycin resistance cassette into the gene at a position near codon 140 (Bill Roof and Ry Young, personal communication). It appears, therefore, that tagging occurs near the site of insertion. Although tagging can be viewed as an artifact in this
instance, it also serves as a reminder of a potentially important function for SsrA. In the wild, some non-essential genes in bacterial genomes are presumably interrupted by phage and transposon insertions or by other genetic rearrangements. SsrA-tagging of the partial proteins resulting from these events would lead to their degradation, preventing any deleterious consequences of the accumulation of such products.

The SsrA-H<sub>6</sub> system described here has some advantages over the SsrA-DD assay system. First, proteins containing His<sub>6</sub> tags can be readily detected and purified using commercially available antibodies and affinity resins. Second, the SsrA-H<sub>6</sub> RNA appears to more closely resemble wild type in some activity assays; for example, λimmP22 plates at roughly 5-fold higher efficiency and with plaque morphology closer to normal on a strain expressing SsrA-H<sub>6</sub> than on a strain expressing SsrA-DD (data not shown). It should be noted, however, that C-terminal His<sub>6</sub> tags may not block proteolysis as efficiently as certain other polar sequences (34).

Our studies suggest that SsrA-tagged proteins represent a relatively small fraction of total protein in cells grown under non-stress conditions but encompass a diverse set of protein species. Most of the proteins identified here were also tagged close to their C-terminus. One caveat is that such proteins might be more soluble or protease resistant because they include complete structural domains, and may be easier to identify by peptide masses because more protein sequence is available to match. It may be dangerous, therefore, to draw conclusions based on the relative frequency of recovery of certain types of tagged proteins or the absence of others. Nevertheless, the addition of stop codons to the list of known causes of tagging supports a broad role for the SsrA•SmpB system in protein quality control.

Acknowledgements—We thank Kelly Williams for help in the cloning of the SsrA-H<sub>6</sub> mutant, Brenda Luciano for resources and assistance with proteomics, Ivan Correia for help with mass spectrometry, and Wali Karzai for useful discussions.
REFERENCES

1. Karzai, A. W., Roche, E. D., and Sauer, R. T. (2000) *Nat Struct Biol* **7**(6), 449-55
2. Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K., and Inokuchi, H. (1994) *Proc Natl Acad Sci USA* **91**(20), 9223-7
3. Ushida, C., Himeno, H., Watanabe, T., and Muto, A. (1994) *Nucleic Acids Res* **22**(16), 3392-6
4. Tu, G. F., Reid, G. E., Zhang, J. G., Moritz, R. L., and Simpson, R. J. (1995) *J Biol Chem* **270**(16), 9322-6
5. Keiler, K. C., Waller, P. R., and Sauer, R. T. (1996) *Science* **271**(5251), 990-3
6. Karzai, A. W., Susskind, M. M., and Sauer, R. T. (1999) *EMBO J* **18**(13), 3793-9
7. Gottesman, S., Roche, E., Zhou, Y., and Sauer, R. T. (1998) *Genes Dev* **12**(9), 1338-47
8. Herman, C., Thevenet, D., Bouloc, P., Walker, G. C., and D'Ari, R. (1998) *Genes Dev* **12**(9), 1348-55
9. Keiler, K. C., Shapiro, L., and Williams, K. P. (2000) *Proc Natl Acad Sci USA* **97**(14), 7778-83
10. Hutchison, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O., and Venter, J. C. (1999) *Science* **286**(5447), 2165-9
11. Huang, C., Wolfgang, M. C., Withey, J., Koomey, M., and Friedman, D. I. (2000) *EMBO J* **19**(5), 1098-107
12. Muto, A., Fujihara, A., Ito, K. I., Matsuno, J., Ushida, C., and Himeno, H. (2000) *Genes Cells* **5**(8), 627-35
13. Oh, B. K., and Apirion, D. (1991) *Mol Gen Genet* **229**(1), 52-6
14. Kirby, J. E., Trempy, J. E., and Gottesman, S. (1994) *J Bacteriol* **176**(7), 2068-81
15. Retallack, D. M., Johnson, L. L., and Friedman, D. I. (1994) *J Bacteriol* **176**(7), 2082-9
16. Retallack, D. M., and Friedman, D. I. (1995) *Cell* **83**(2), 227-35
17. Abo, T., Inada, T., Ogawa, K., and Aiba, H. (2000) *EMBO J* **19**(14), 3762-9
18. Roche, E. D., and Sauer, R. T. (1999) *EMBO J* **18**(16), 4579-89
19. Parsell, D. A., Silber, K. R., and Sauer, R. T. (1990) *Genes Dev* **4**(2), 277-86
20. Nelson, H. C., and Sauer, R. T. (1985) *Cell* **42**(2), 549-58
21. Breyer, R. M., and Sauer, R. T. (1989) *J Biol Chem* **264**(22), 13355-60
22. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) *Anal Chem* **68**(5), 850-8
23. Withey, J., and Friedman, D. (1999) *J Bacteriol* **181**(7), 2148-57
24. Nakamura, Y., Ito, K., and Isaksson, L. A. (1996) *Cell* **87**(2), 147-50
25. Ptashne, M. (1986) *A Genetic Switch: Gene Control and Phage λ*, Blackwell Scientific Publications & Cell Press, Cambridge, Ma.
26. Bjornsson, A., Mottagui-Tabar, S., and Isaksson, L. A. (1996) *EMBO J* **15**(7), 1696-704
27. Mottagui-Tabar, S., and Isaksson, L. A. (1998) *Gene* **212**(2), 189-96
28. Zhang, S., Ryden-Aulin, M., and Isaksson, L. A. (1998) *J Mol Biol* **284**(5), 1243-6
29. Sauer, R. T. (1978) *Nature* **276**(5685), 301-2
30. Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science* **277**(5331), 1453-74
31. Arkov, A. L., Korolev, S. V., and Kisselev, L. L. (1993) *Nucleic Acids Res* **21**(12), 2891-7
32. Sauer, R. T., Pan, J., Hopper, P., Hehir, K., Brown, J., and Poteete, A. R. (1981) *Biochemistry* **20**(12), 3591-8
33. de Smit, M. H., and van Duin, J. (1994) *J Mol Biol* **235**(1), 173-84
34. Milla, M. E., Brown, B. M., and Sauer, R. T. (1993) *Protein Sci* **2**(12), 2198-205
35. Williams, K. P., and Bartel, D. P. (1996) *RNA* **2**(12), 1306-10
36. Felden, B., Himeno, H., Muto, A., McCutcheon, J. P., Atkins, J. F., and Gesteland, R. F. (1997) *RNA* **3**(1), 89-103
FIGURE LEGENDS

FIG. 1. **Sequence and activity of SsrA-H<sub>6</sub> RNA.** (A) Base sequence of the SsrA-H<sub>6</sub> RNA in the region encoding the peptide tag. Mutant bases are shown in small letters; wild-type bases are shown in capital letters. The secondary structure of SsrA RNA consists of the alanine-charged tRNA domain, four pseudoknots, and the tag-encoding region which ends in a stem-loop (35,36). (B) Plaque formation by λimmP22 phage (15) on strain W3110 ssrA::kan containing no SsrA (vector plasmid pKW1), wild-type SsrA (pKW11), or SsrA-H<sub>6</sub> (pKW24). (C) Log-phase growth of strain W3110 ssrA::kan containing no SsrA (vector plasmid pKW1), wild-type SsrA (pKW11), or SsrA-H<sub>6</sub> (pKW24) at 43 °C. The inset table gives the relative rates of growth at 43 °C and 37 °C. (D) Western blot probed with antibodies to an epitope in the λ-cI-N-trpAt protein (expressed from plasmid pPW500) of cell lysates of strain X90 ssrA::cat containing SsrA-H<sub>6</sub> (lane a, pKW24), SsrA-DD (lane b, pKW23), no SsrA (lane c, pKW1), or wild-type SsrA (lane d, pKW11).

FIG. 2. **SsrA-H<sub>6</sub> tagging of λ cI repressor.** Western blots of cell lysates expressing SsrA-H<sub>6</sub> (lanes c and f), λ cI repressor (lanes a and d), or SsrA-H<sub>6</sub> and λ cI repressor (lanes b and e) were probed with anti-H<sub>6</sub> antibodies (lanes a-c) or anti-λ repressor antibodies (lanes d-f). The positions of SsrA-H<sub>6</sub> tagged λ repressor and untagged repressor are marked by arrows. The experimental mass of purified, SsrA-H<sub>6</sub> tagged λ repressor and of the C-terminal tryptic peptide of this protein are given below the protein label. The lower portion of the figure shows the C-terminal codons of the repressor message and gives the expected masses of the intact repressor, the repressor plus the SsrA-H<sub>6</sub> tag, and the C-terminal tryptic tag peptide.

FIG. 3. **SsrA-H<sub>6</sub> tagged proteins were purified through two cycles of Ni<sup>2+</sup>-NTA chromatography and analyzed by two-dimensional electrophoresis.** (A) Silver-stained gel of *E. coli* proteins purified from a strain without SsrA-H<sub>6</sub>. (B) Silver-stained gel of proteins
purified from a strain containing SsrA-H₆. (C) Western blot using anti-H₆ antibodies of proteins purified from cells containing SsrA-H₆. Labels A-E and U-Z mark the same positions in all three panels. Identified proteins listed in Table I are (A) YbeL, (B) LacI, (C) SlyD, (D) GalE, (E) RbsK, (X) NikR, and (Y) Fur.

FIG. 4. C-terminal sequences of proteins identified as substrates for SsrA-tagging and the corresponding regions of their genes. (A) A portion of the MALDI-TOF mass spectrum that provided the peptide mass fingerprint for YbeL given in Table I. The spectrum is overlaid with a background spectrum (the lower spectrum), and the two most prominent non-background peaks (1968.99 and 2998.05) are marked by arrows that show matching junction peptides which include the SsrA-H₆ tag. The C-terminal coding region of the ybeL gene is shown for comparison. (B) The 3’ end of the galE gene overlaps the Shine-Dalgarno sequence of the galT gene. (C) Two of the three final sense codons of the rbsK gene are AGG codons, the rarest codon in E. coli. Synthesis of the RbsR protein begins at a codon three bases away from the rbsK stop codon.
TABLE I

Identification of proteins by peptide mass fingerprinting

Individual proteins are marked with the letter of the corresponding spot from the 2D gels in Figure 3. The pI and molecular weight listed are the theoretical values for the normal, untagged proteins. Peptide masses are for peaks selected as significant when the mass spectrum of the peptide mixture was compared to a background spectrum. The number of peptide masses matching the identified protein, the fraction of the protein residues covered by the matched peptides, and the last residue matched in the protein were determined with MS-Fit settings of 2 missed cleavages allowed and 1000 ppm uncertainty for the masses. The identified proteins consistently score higher than all other *E. coli* proteins with a wide variety of search settings, and in most cases a majority of the experimental peptide masses were within 300 ppm of the masses for the matched peptides.

| Identified Proteins | Untagged | Peptide Masses (Da) | Peptides %Coverage | Last Residue Matched |
|---------------------|----------|---------------------|--------------------|----------------------|
| *SsrA*-tagged       |          |                     |                    |                      |
| (A) YbeL 5.11       | 18.8     | 15/20               | 77                 | 155/160              |
| (B) LacI 6.39       | 38.6     | 26/31               | 68                 | 351/360              |
| (C) SlyD 4.86       | 20.9     | 11/16               | 69                 | 140/196              |
| (D) GalE 5.89       | 37.3     | 24/27               | 81                 | 331/338              |
| (E) RbsK 4.99       | 32.3     | 12/12               | 79                 | 307/309              |
| Untagged            |          |                     |                    |                      |
| (X) NikR 5.77       | 15.1     | 9/12                | 90                 | 133/133              |
| (Y) Fur 5.68        | 16.8     | 8/8                 | 72                 | 148/148              |
Figure 1

A

UAGUCGCAAACGACGAAAACUAC
Ala Asn Asp Glu Asn Tyr Ala

UAGUCGCAAACGACGAAAACUAC
Ala Asn Asp Glu His His

UAGUCGCAAACGACGAAAACUAC
Ala Asn Asp Glu His His

B

\[
\begin{array}{c|c|c}
\text{SsrA Variant} & \text{None} & \text{WT} & \text{H}_6 \\
\hline
\text{Phage Plaques} & 10^5 & 10^7 & 10^6 \\
\end{array}
\]

C

| SsrA   | Growth Rate 43°C | Growth Rate 37°C |
|--------|------------------|------------------|
| WT     | 1.00             | 1.00             |
| H\textsubscript{6} | 0.87             | 0.97             |
| None   | 0.65             | 0.85             |

D

\begin{itemize}
  \item \(\lambda\)-cl-N-trpAt: + + + +
  \item SsrA Type: H\textsubscript{6} DD None WT
\end{itemize}

\[\text{OD}_{600}\]

43°C

\[
\begin{array}{c|c|c|c|c}
\text{Time (min)} & 0 & 20 & 40 & 60 & 80 & 100 & 120 \\
\hline
\text{WT} & 0 & 0.5 & 1 & 1.5 & 2 & 2.5 & 3 \\
\text{H}_6 & 0 & 0.4 & 0.8 & 1.2 & 1.6 & 2 & 2.4 \\
\text{None} & 0 & 0.2 & 0.4 & 0.6 & 0.8 & 1 & 1.2 \\
\end{array}
\]
**Figure 2**

- **SsrA-H**
  - + + - + +
- **λ Repressor**
  - + - + + -

- **Tagged λ Repressor**
- Whole protein
  - 27437 ± 42 Da
- Tryptic digest peptide
  - 2687.7 ± 0.8 Da

**λ Repressor**

- **Whole protein**
  - 27437 ± 42 Da
- **Tryptic digest peptide**
  - 2687.8 Da

**SsrA-H**

- 27404 Da
- NH$_2$STKKKP...KVIA$\underline{SQWPEETFGAANDEHHHHH}$-COOH

**λ Repressor**

- 26081 Da
- NH$_2$STKKKP...KVIA$\underline{SQWPEETFG}$-COOH

- Glu Glu Thr Phe Gly Stop
- GAA GAG ACG UUU GGC UGA UCG
Figure 4

A

RPFEPAANDEH₆

CGHDQFRRPFEPAANDEH₆

YbeL

Lys  Cys  Gly  His  Asp  Gln  Phe  Gln  Arg  Arg  Pro  Phe  Glu  Pro  Stop  Stop
AAA  TGT  GGT  CAT  GAC  CAG  TTC  CAG  AGA  CGC  CCG  TTT  GAG  CCG  TAA  TAG  TCT

GalE

Ser  Arg  His  Pro  Gln  Gly  Tyr  Pro  Asp  Stop
TCA  CGC  CAT  CCA  CAG  GGA  TAT  CCC  GAT  TAA  GGA  ACG  ACC  ATG  ACG  CAA
Shine-Dalgarno

GalT

Met  Thr  Thr

B

C

RbsK

Ala  Phe  Leu  Asp  Arg  Gln  Arg  Stop
GCA  TTT  TTA  GAC  AGG  CAG

RbsR

Met  Ala  Thr  Met  Lys  Asp  Val
TGA  CGC  TTG  GCT  ACA  ATG  AAA  GAT  GTT
Identification of endogenous SsrA-tagged proteins reveals tagging at positions corresponding to stop codons

Eric D. Roche and Robert T. Sauer

J. Biol. Chem. published online May 23, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M103864200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts