Chapter I

Introduction & Background: Translational Bypassing in Bacteriophage T4 gene *60*

Motivation

In the central dogma of molecular biology, mRNA is depicted as a long, linear molecule whose primary function is to serve as an intermediate in the transfer of information from DNA to protein[1]. Since this model was proposed, however, mRNA has been found to play many other interesting roles in gene expression and regulation. For example, the primary sequence of mRNA can target transcripts to specific regions of a cell for spatial localization of translation[2]. Temporal regulation of translation is also encoded in the primary sequence of mRNA in the form of rare codons. Slower elongation rates or transient stalling at rare codons provides time for cotranslational folding of nascent peptide domains[3,4]. mRNA structures in the 5' untranslated region (UTR), called riboswitches, maintain metabolite homeostasis in the cell by binding to specific metabolites when they are present above threshold concentrations. Metabolite binding induces structural changes that either inhibit or enhance the transcription or translation of genes related to synthesis or uptake of the bound metabolite[5-8]. Occlusion of the Shine-Dalgarno (SD) region within an mRNA structure can block small ribosomal subunit binding (S30 in *E. coli*), inhibiting translation initiation of chaperones and other proteins involved in temperature stress response until a temperature change melts the mRNA structure[8-14]. Finally, pseudoknots or stable hairpins within an open reading frame (ORF) can induce shifts in the reading frame. These shifts can range from small -1[15-21] or +1[17,19] nucleotides, to larger jumps that skip over or re-frame stop codons[19]. In each case, multiple protein products are encoded in overlapping reading frames and the efficiency of the frameshifting event determines the relative ratio at which each protein is expressed[15,17]. In these ways, mRNA provides additional layers of regulation for gene expression[19,20] beyond those at the DNA and protein levels.

A significantly more complicated mechanism for gene regulation is attributed to the mRNA structure of gene *60* in Bacteriophage T4[19,22-28]. In the middle of the gene *60* mRNA ORF there is an in-frame stop codon which may terminate translation. Alternatively, the ribosome may also skip over this stop codon and the subsequent 47 nucleotides and continue decoding the message (Figure I.1 A). This event, termed *translational bypassing*, is unique among traditional frame-shifts in that it is extremely efficient (~50%) and skips over a very large number of nucleotides (+50). The structure of the nucleotides in the untranslated region is thought to play a crucial role in bypassing[19,22-30]. Several structural models of this coding gap have been proposed (Figure I.5), all of which include a hairpin at the 5' edge of the region[19,22-30] (Figure I.1 B). Computational[26] and mutational analyses strongly[23,24,27-29] support the presence of this hairpin, but could not clearly confirm or refute hypotheses about the structure in the rest of the coding gap.

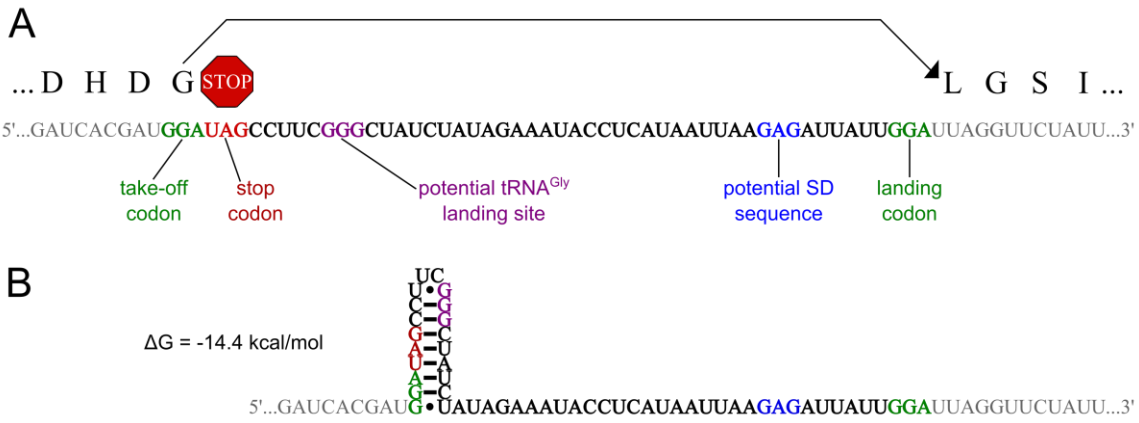


Figure I.1 The coding gap in gene *60* mRNA. (**A**) Nucleotides in gray are translated to the amino acids listed above codons in large black letters. Coding gap nucleotides are black or colored. The coding gap is flanked by take-off and landing sites composed of identical Gly codons (green). The in-frame stop codon is in red. Scanning of the coding gap is disfavored because a closer Gly codon (purple) is never used as a landing site[27]. A short SD-like element (blue) lies just upstream of the landing site Gly. (**B**) A stable, tight-turning UUCG loop hairpin is predicted to exist at the 5' end of the coding gap[19,22-30]. An MFold calculation of the stability of the hairpin is shown[31].

This thesis investigates proposed hypotheses of the mechanism of translational bypassing with a structure-function study of gene *60* mRNA. Probing with structure-sensitive chemicals provides information about the environment of each nucleotide in gene *60* mRNA, enabling us to support or refute various proposed structures for the

coding gap. Translational bypassing competency of wildtype and mutant versions of gene *60* mRNA in a cell-free translation assay allow correlations to be made between mRNA structure and bypassing efficiency. Finally, single-molecule fluorescence experiments are proposed to elucidate aspects of translational bypassing unobservable in the ensemble level assays. These studies provide important new insights about the mechanism of translational bypassing.

The Origin of Gene *60*

A family of bacteriophages with high homology to bacteriophage T4 (phage T4 or T4) are described as "T4-like" or "T-even". The most closely related T-even members infect *E. coli*, but more distantly related T-even phages (< 25) have been isolated from *Acinetobacter, Aeromonas, Klebsiella, Pseudomonas, Shigella, Vibrio* and some marine cyanobacteria hosts. All T-even phages contain a core genome consisting of the genes essential for phage replication under all known conditions. These 31-33 core genes, while having greater than 90% nucleotide sequence identity among T-even phages, only

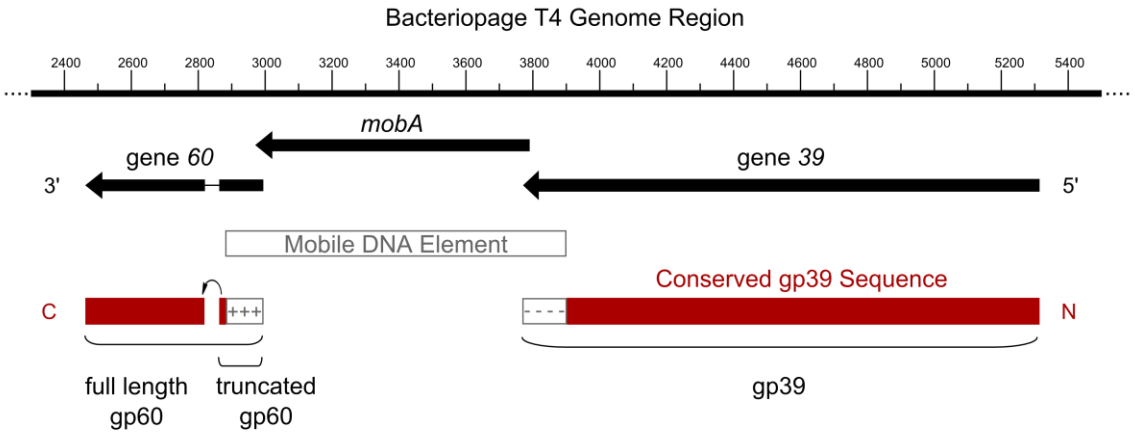


Figure I.2. Generation of gene *60* in Bacteriophage T4. Highly conserved among T-even phages, gene *39* of Bacteriophage T4 was split into two open reading frames (red bars) by the insertion of a mobile DNA element (gray bar), resulting in two separate genes, *39* and *60* (black arrows). A homing endonuclease gene, *mobA*, is encoded on the mobile DNA element. The insertion also added additional amino acids to the N-terminus of the protein product of gene *60* (gp60) and the C-terminus of gp39 (gray boxes). These regions are oppositely charged and likely bring together the separately expressed gp39 and gp60 through electrostatic interactions. An additional insertion of 50 nucleotides is also present in gene *60* (small black line disrupting gene) that is not translated (black curved arrow)[33]. Figure adapted from the Genome Browser view of the Bacteriophage T4 genome.

make up only 12-15% of the phages' proteome. In addition to the core genome, T-even phages infecting the same host or sharing similar genetic backgrounds also contain conserved quasicore genomes. These genes may or may not be essential for phage reproduction and are likely maintained to help a phage survive during times of environmental stress[32].

Bacteriophage T4 contains a very interesting functional genetic variation from other T-even phages: generally T-even phage genomes contain two highly conserved quasicore genes, *52* and *39*, which code for protein subunits of the phage's Type II DNA topoisomerase. These genes are essential below 25°C[35]. Phage T4, however, contains a ~1 kb insertion (gray bar, Figure I.2) that splits the N and C-terminus of gene *39* protein (gp39) into two distinct genes, *39* and *60*, generating a third subunit of the topoisomerase[25,32,35]. Intriguingly, there is also a second insertion of 50 nucleotides within the newly formed gene *60*, which make up the coding gap that is bypassed during translation[25] (Figure I.2 thin black line, curved black arrow).

Collaborative Homing between MobA and the Coding Gap of Gene *60*

The large insertion that splits gp39 into two ORFs contains a homing endonuclease gene, *mobA*[33]. Homing endonucleases are mobile DNA elements that recognize and cleave a specific sequence in a homologous DNA genome lacking the endonuclease gene. Cleavage of genomic DNA triggers DNA repair by recombination and the genome containing the endonuclease gene is used as a template to repair the cleaved one. The resulting transfer of the endonuclease gene to a new host genome is referred to as "homing"[33,36]. Homing endonucleases can insert themselves into introns or inteins, but most found in T-even phages, like *mobA*, are free standing[33,36]. These types of elements are commonly found in or near highly conserved regions of proteins involved in DNA synthesis and repair, probably to increase their likelihood of being preserved in the genome[36]. Homing endonuclease genes are also often near their cleavage site to ensure horizontal transfer of the endonuclease gene during DNA repair[33,36].

Homing endonucleases also require negative regulation so they do not cleave and damage the genome in which they reside[33,36]. This can be accomplished by blocking the endonuclease recognition site in one of several ways. Free standing homing

endonucleases can be regulated by simple point mutations in the cleavage recognition site. Because these mutations are near the point of insertion of the endonuclease, they will be transferred to the cleaved genome along with the endonuclease gene (Figure I.3 A). For endonuclease-containing introns, the intron itself can disrupt the endonuclease recognition site. Thus only genomes lacking the intron (and endonuclease gene) will be

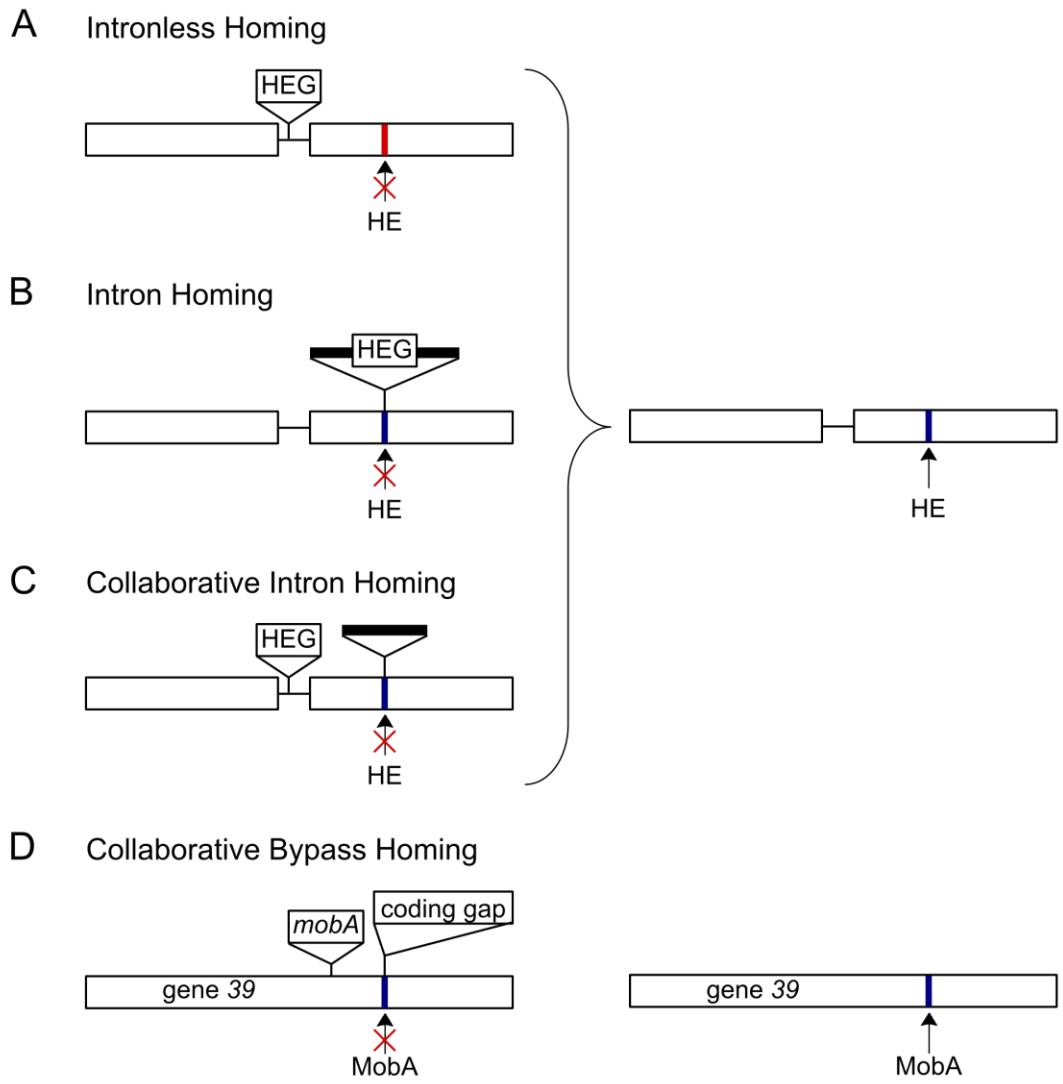


Figure I.3 Mechanisms by which homing endonucleases prevent cleavage of their genome. Homing endonucleases (HE) cleave specific sequences (blue band) in genomes lacking the homing endonuclease gene (HEG) (genome middle, right). (**A**) For intronless homing, point mutations at the cleavage recognition site (red band) prevent cleavage by HE. (**B**) Introns containing a HEG can split the cleavage recognition sequence of the HE. (**C**) Similarly, in collaborative homing a free-standing HEG can use an intron to disrupt the HE cleavage site. (**D**) The coding gap in T4 gene *60* prevents cleavage by the MobA HE in a collaborative manner (left), ensuring only genomes without *mobA* are cleaved (right). Adapted from Bonocora et al.[33].

cleaved (Figure I.3 B). Finally, a free standing homing endonuclease can "collaborate" with an autonomous intron disrupting its recognition site (Figure I.3 C). All of these simple blocking mechanisms prevent homing endonucleases from cleaving their own genomes[33].

Although it is not spliced before tranlsation[25] and does not share any homology to self-splicing introns[25,33], Bonocora et al. hypothesized that the coding gap of gene *60* could be involved in collaborative homing with the inserted endonuclease, MobA. They confirmed *in vitro* that MobA cleaves gene *39* of a related phage, T2, nineteen nucleotides 5' of the site where the coding gap is inserted in gene *60* of T4 (Figure I.4). Additionally, MobA was able to cleave a mutant of T4 gene *60* that completely lacked the coding gap sequence. With the coding gap present in either T2 gene *39* or T4 gene *60*, however, MobA activity was completely inhibited. Thus the coding gap appears to collaborate with MobA by disrupting its cleavage recognition site. *In vivo* experiments also confirmed this hypothesis. *E. coli* cells expressing MobA and containing a plasmid with T4 gene *60* were infected with phage T2. Almost all of the resulting T2 phage progeny acquired the coding gap region, confirming transfer of the coding gap sequence in a MobA dependent fashion[33]. These collaborative elements are also highly mobile in nature. An insertion creating distinct genes *39* and *60* as well as the exact 50 nucleotide insertion in gene *60* has also been found in phages M1, SKX, FS10, RB55, RB59 and Pol[37].

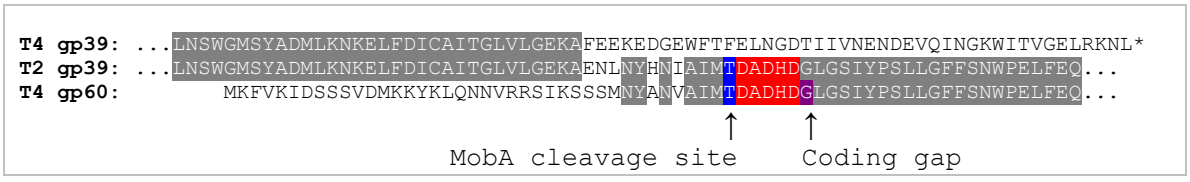


Figure I.4 Sequences of T4 gp39 and T4 gp60 are compared to T2 gp39. The N-terminal region of T4 gp39 is nearly identical to the N-terminus of T2 gp39 (gray shading). Similarly, T4 gp60 C-terminal end is nearly identical to the C-terminal domain of gp39. Ellipses indicate the sequence continues and the stop of T4 gp39 is marked with an asterisk. The purple Gly marks the insertion site of the 50 nucleotide coding gap into T4 gp60. The MobA cleavage site lies within the blue Thr codon[33], just upstream of the conserved Toprim domain of gp39 highlighted in red[33,38,39].

The Two ORFs Composing gp39 in Bacteriophage T4

The *mobA* insertion splits gene *39* into two ORFs very close to a highly conserved Toprim motif required for topoisomerase activity (Figure I.4, highlighted in red)[33,38,39]. The Toprim motif contains four acidic residues: a glutamate and three aspartic acids arranged in a DxDxD configuration that coordinate a catalytic Mg^{2+} ion[38,39]. Insertion of *mobA* into this conserved DNA processing gene likely ensures its retention in the genome[36]. However, it directly disrupts the ORF of gene *39*, which could hinder phage vitality. To circumvent this potential problem, there are additional features in the mobile DNA element to ensure both halves of the subunit are expressed. First, the insertion adds an additional 43 amino acids to the C-terminus of the new gp39 as well as a termination codon. Similarly, translation initiation elements (a SD element and a start codon) and 31 amino acids are added to the N-terminus of gp60 (Figure I.4, compare T4 gp39 and T4 gp60 to the continuous T2 gp39). These modifications coordinate translation of the two protein fragments[33]. Although it would be convenient if these added protein sequences could perform protein ligation to reconnect the fragments of gp39, they do not contain any conserved intein motifs[40,41]. The added amino acid sequences also do not have significant homology to other proteins. Interestingly, the amino acid extensions are oppositely charged at neutral pH: the C-terminal extension of gp39 has a net charge of -8 while the N-terminal extension of gp60 has a net charge of +6 (Figure I.1). The function of these highly charged peptides added at the site of breakage could be to bring the separately expressed polypeptides together through electrostatic interactions.

It is also important to briefly address truncated gp60 in the context of the *mobA* mobile element. Truncated gp60 is produced in about 50% of translation attempts of gene *60* mRNA[29] (Figure I.4, truncated gp60 ends at the purple Gly), yet its role (if any) during phage infection remains unknown. Except for the conserved Toprim domain, truncated gp60 is primarily composed of amino acids inserted by the *mobA* element (Figures I.2 and I.4). Thus truncated gp60 could have a role in regulation of MobA activity. Additional implications of this protein are discussed below in "Role of gp60 Nascent Peptide Chain in Translational Bypassing".

Translational Bypassing of T4 Gene *60* mRNA

The phenomenon of translational bypassing was first discovered by Huang et al. when they attempted to clone and purify gp60. They discovered a 50 nucleotide insertion in the ORF of gene *60* mRNA, beginning with an in-frame stop codon, that was not translated[25] (Figure I.1 A). As this sequence shares no homology with any introns[25,33] and the authors were unable to detect any products of its removal by splicing, they concluded bypassing must occur during translation[25]. The simplest mechanism for bypassing at the translational level is an extension of the basic principles of frameshifting[42]. When a translating ribosome encounters a downstream structure in the mRNA, the peptidyl-tRNA becomes strained and dissociates from the mRNA or "takes off". If there is an overlapping or nearby codon that matches the anti-codon of the peptidyl-tRNA, it can reattach or "land" before the ribosome resumes translation[19]. Similarly, peptidyl-tRNAs can dissociate from the mRNA in the absence of any downstream structure if the amino acid for the subsequent codon is lacking. The ribosome will hold onto the peptidyl-tRNA while scanning the mRNA in the 3' direction until it finds the next codon complementary to the peptidyl-tRNA anti-codon. Once the peptidyl-tRNA is bound to the landing codon, translation resumes[43].

By applying these principles to translational bypassing, the simplest mechanism for this phenomenon would be that either the stop codon or some sort of structure in the coding gap of gene *60* mRNA causes the peptidyl-$tRNA^{Gly}$ to dissociate from the mRNA, and the ribosome scans in the 3' direction over 50 nucleotides before finding the next Gly codon for peptidyl-$tRNA^{Gly}$ reattachment. The extremely high frequency of bypassing (~50%) disfavors this simple mechanism. The efficiency at which the ribosome resumes translation under starvation conditions decreases with distance between the take-off and landing codons. Hopping over up to 15 nucleotides can be as efficient as 20%; hopping over 40 nucleotides is only successful in 1% of attempts[43]. Contrary to this trend, deletions of the coding gap that bring the landing codon closer to the take-off site decrease bypassing efficiency[19,24,28]. Most importantly, however, there is a potential landing site (GGG) for $tRNA^{Gly}$ just 6 nucleotides downstream of the stop codon, while the wildtype landing codon is 47 nucleotides away. If the ribosome was simply scanning the coding gap it would statistically almost always land at the first Gly codon, leading to

extremely low levels of full-length gp60. Mass spectrometry experiments confirm that under wildtype conditions this alternate landing codon is never used[27]. As this simple mechanism is strongly disfavored, it was proposed that the coding gap folds into a structure that induces the ribosome to skip over this entire region at high efficiencies[19,23,25,27,28].

Structural Models of the Coding Gap of Gene *60* mRNA

Several structural models have been proposed for the coding gap of gene *60* mRNA[22,25,26] (Figure I.5). In the initial secondary structure model proposed by Huang et al., the coding gap was folded in such a way as to position the codons before (GGA) and after (UUA) the gap adjacent to one another. By folding into this structure, they proposed that the ribosome could directly translate over the discontinuity in the mRNA at high efficiency[25] (Figure I.5 B). As an extension of this model, Burke-Aguero and Hearst[22] hypothesized that the bonding patterns in the UGGAU pentanucleotide repeat on either side of the coding gap could migrate as in a Holliday junction. This would move the location of the discontinuity in the mRNA around the peptidyl transferase center such that the ribosome would never have to translate over a discontinuity in the mRNA[22] (Figure I.5 C). Computational folding of gene *60* mRNA also supports the presence of a highly stable structure in the coding gap region and several models were proposed[26] (Figure I.5 D). While the 5' edge hairpin is common to all models[22,25,26] (Figure I.5 A), structures in the 3' end vary significantly. Gaining strong support for one model over another through covariation analysis[44-47] is not possible because the coding gap sequence is not homologous to any other known sequence.

To test structural models, several mutants of gene *60* were constructed[23,24,27-29,42,48]. Early tests of bypassing efficiency were performed by adding a *lacZ* gene in-frame and just downstream of the coding gap (Figure I.6 A). β-galactosidase activity of wildtype and mutant versions were compared as a measure of bypassing efficiency[25,28,42]. Efficiency was also monitored by ^{35}S-Met labeling of GST-tagged *in vivo* translation products[23,24,29,49] as well as purification and analysis of protein products by mass spectrometry[27]. With all methods, the mutations that most significantly impacted bypassing were those made to the predicted hairpin at the 5' end of the coding gap region.

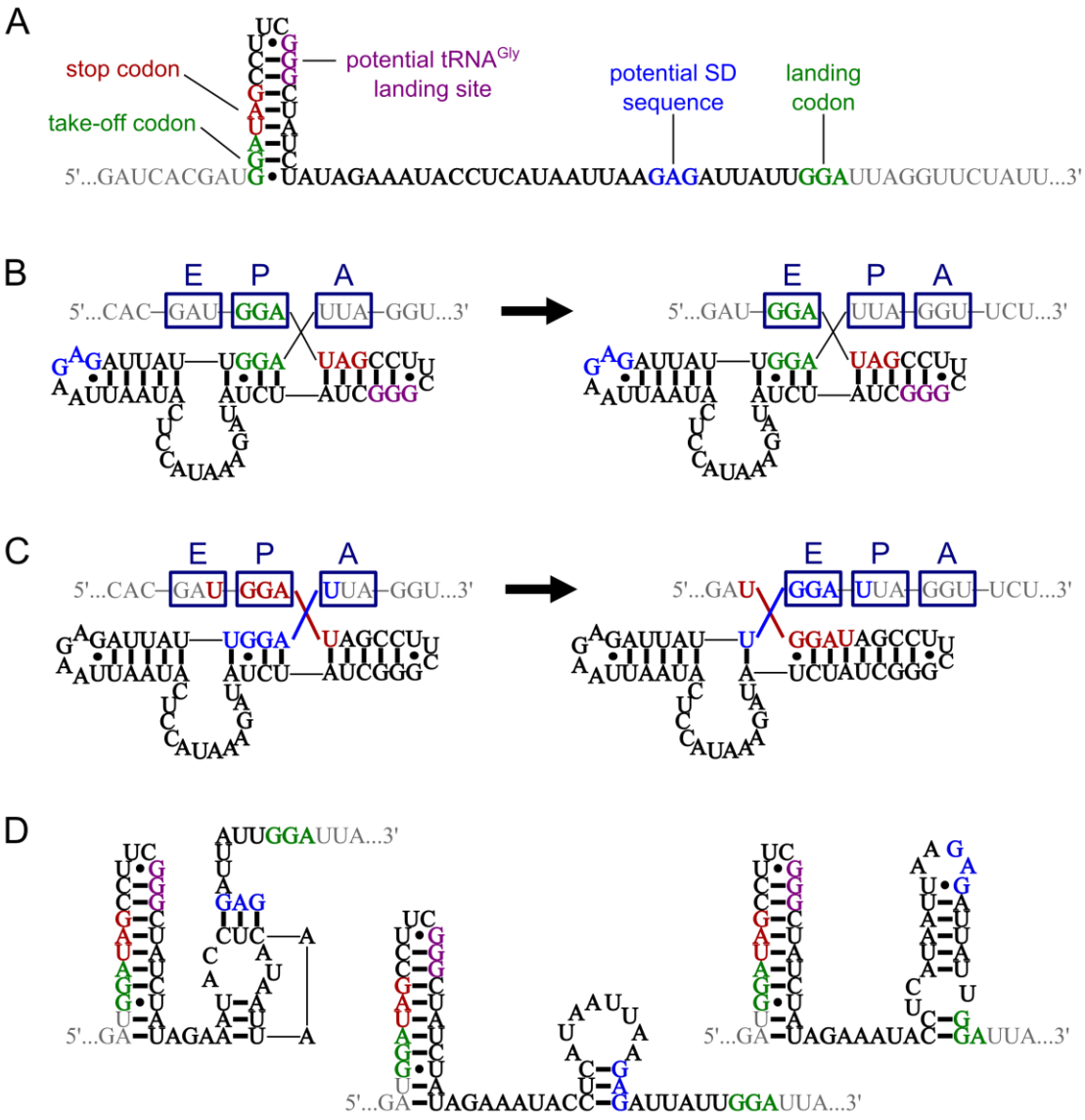


Figure I.5 Proposed models of the gene *60* mRNA coding gap. (**A**) A hairpin in the 5' edge of the coding gap region is present in all proposed models[22,25,26]. Labels and colors as in Figure I.1. (**B**) Original proposed fold in the coding gap region in which the ribosome translates over a discontinuity in the mRNA[25]. E, P and A sites of the ribosome before and after elongation (black arrow) over the mRNA discontinuity are labeled in dark blue. (**C**) The UGGAU pentanucleotide repeat at the 5' edge (red) and the 3' edge (light blue) of the coding gap. Migration of the branch point could circumvent elongation over a discontinuous mRNA template[22]. (**D**) Three additional proposed structures of the coding gap region[26] with colors as in (**A**).

Destabilizing or extending the stem and changing the sequence of or elongating the loop significantly disrupted bypassing[19,23-25,27,28]. Mass spectrometry confirmed that one source contributing to this decrease in efficiency was landing at the closer Gly codon. Thus stability of the 5' hairpin modulates bypassing and scanning behavior of the

ribosome[27]. Portions of structural features proposed in the 3' end of the original model were deleted systematically, generating significant drops in bypassing efficiency. The largest deletion, however, still exhibited ~3% bypassing. Similarly, insertions or substitutions in the 3' half of the coding gap also reduced bypassing efficiency considerably, but could not completely abolish it. This indicates some compositional variation is tolerated in the 3' part of the coding gap[28].

There are several other minor features known to be important for bypassing. The identity of the take-off and landing codons must be identical[28,42]. Additionally, all possible combinations of identical take-off and landing codons were tested for bypassing activity. (Note: in all constructs tested, the stem of the 5' hairpin was maintained by making complementary mutations to match the take-off codon.) The wildtype GGA codon was the most efficient for bypassing. Codons with a C or G in the first or second position favored higher rates of efficiency[42]. Finally, a short, SD-like element (GAG) just upstream of the landing site aided in faithful recognition of the landing codon[27,50].

Role of gp60 Nascent Peptide Chain in Translational Bypassing

In addition to extensive mutational analysis of the coding gap itself, the surrounding sequence in gene *60* was also probed for elements involved in translational bypassing[28]. The earliest tests were performed with an in-frame *lacZ* gene inserted five nucleotides downstream of the untranslated region of gene *60*. Constructs with and without the coding gap exhibited identical β-galactosidase activity, indicating the sequence of gene *60* mRNA 3' of the coding gap is not involved in bypassing (Figure I.6 A). Truncations were also made to the 5' end of the coding sequence of gene *60* mRNA. Results show a steady decrease in bypassing efficiency as larger portions of the 5' end are deleted. Removal of 42 or more nucleotides (14 amino acids) completely abolishes bypassing (Figure I.6 B). Internal insertions 5' of the conserved Toprim domain disrupt bypassing more effectively as insertions increase in length. Similarly, large portions in this region of gene *60* cannot be deleted without severely hampering bypassing. To determine if the region around the Toprim domain is important at the level of mRNA or protein, the reading frame of various parts of this section were shifted by single point insertions and deletions or by randomly varying the protein sequence. Alteration of the

protein sequence from amino acids 16 to 32 reduced bypassing activity 100-fold (Figure I.C), suggesting the nascent peptide sequence plays an important role in translational bypassing. Mutation of this region to synonymous codons had wildtype levels of bypassing, confirming that the crucial aspect of this part of gene *60* is the nascent peptide sequence, not the nucleotide sequence or structure[28].

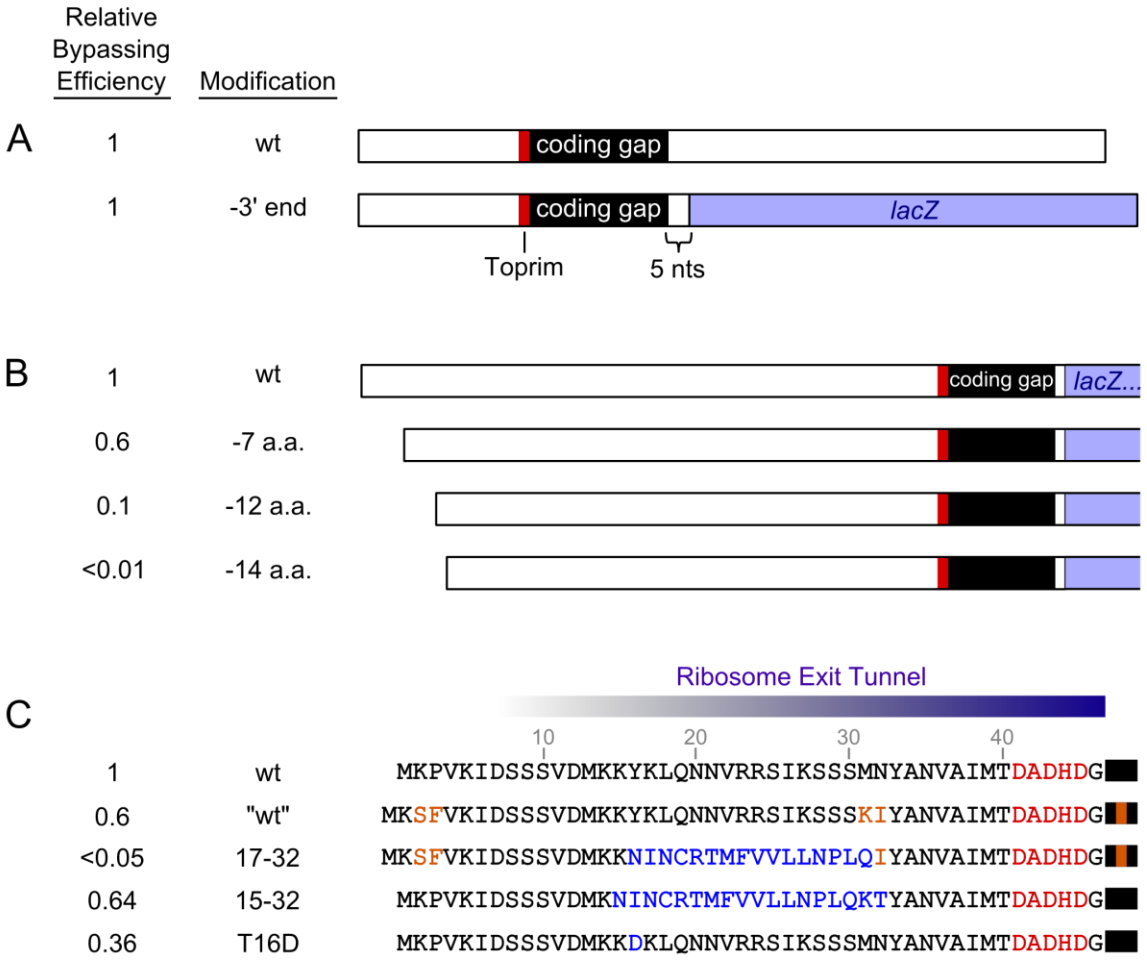


Figure I.6 Regions outside of the coding gap that play a role in bypassing. (**A**) Schematic of wildtype gene *60* mRNA (open bar) and the insertion of the *lacZ* reporter (blue bar) 5 nucleotides (nts) after the coding gap (black bar). The Toprim motif is marked by a red bar. Levels of bypassing are identical for the two constructs indicating the 3' end of gene *60* is not involved in bypassing[28]. (**B**) Indicated truncations from the 5' end of the gene *60* ORF and their bypassing efficiency relative to wildtype[28]. (**C**) Indicated changes were made to gp60 and relative bypassing efficiencies are listed[23,24,28,51]. Amino acids (a.a.) in the Toprim domain are red. Orange a.a. and orange bar in the coding gap highlight mutations made during the original determination of bypassing efficiency[28]. Blue a.a. are indicated mutations to gp60. Region of nascent peptide that lies in the ribosomal exit tunnel is depicted by the blue fading bar[51-54].

Due to the limitations of cloning at the time, the original efficiencies from mutating the nascent peptide sequence were obtained with substrates containing additional mutations to the nascent peptide sequence and also a mutation in the hairpin at the 5' edge of the coding gap[28] (Figure I.6, orange amino acids and orange band). A decade later, these measurements were repeated and the mutated nascent peptide chain alone generated a more modest 36% decrease in bypassing efficiency (Figure I.6 C). Although not as dramatic as originally reported, the nascent peptide sequence still plays a significant role in translational bypassing[23,24]. Another study individually mutated each of the amino acids in the critical peptide sequence to Asp and did not observe any major affects on bypassing efficiency except in the Tyr17Asp mutation. When the Tyr17 residue was changed to an Asp, Gly or Ala, β-galactosidase activity fell to 30% that of wildtype (Figure I.6 C). A Tyr17Phe mutation, however, had no impact on bypassing efficiency, indicating a hydrophobic residue at this location is critical[51]. In sum, these studies strongly imply the wildtype nascent peptide chain plays an important role in translational bypassing.

There are several examples of nascent peptide sequences that can regulate translation. Based on protease protection assays, the exit tunnel of the ribosome can hold 30-40 amino acids of the nascent peptide[51-54] (Figure I.6 C, fading blue bar). The exit tunnel is primarily composed of rRNA, but also has a constricted region generated by L2 and L22[55]. Several peptide sequences are known to regulate gene expression by interacting with these elements. An inducible class of translation regulating peptides responds to metabolites or antibiotics[53]. One example is the 24 amino acid peptide TnaC of the tryptophanase (*tna*) operon of *E. coli*. At high cellular levels of tryptophan, the TnaC peptide interacts with the ribosomal exit tunnel and tryptophan binds in the A-site preventing termination by RF3. The prolonged ribosomal stall induced by tryptophan binding blocks downstream Rho-dependent transcription termination sites, upregulating tryptophan catabolizing genes[54,56,57]. A second class of peptides has intrinsic stalling properties[53]. The SecM (secretion monitor) peptide interacts with rRNA, distorting the peptidyl transferase site and stalling the ribosome. Stalled 70S complexes expose a SD element for the downstream *secA* gene[53], causing upregulation of this protein exporter[54,55]. Finally, the short 2A peptide of the foot-and-mouth disease virus forms an amphipathic

helix and a tight turn containing a self-cleavage site. 2A cleaves its own C-terminus in the ribosomal peptidyl transferase center due to strain generated by the structure of the nascent peptide. After cleavage, the ribosome can resume translation of the ORF with no apparent need for re-initiation[54,58].

Given the interesting regulatory processes of nascent peptides, it was hypothesized that the nascent peptide of gene *60* could play an important functional role in translational bypassing. The critical region of nascent gp60 would reside in the ribosome exit tunnel when the ribosome reaches the bypass site (Figure I.6 C). If this peptide were interacting with the ribosome exit tunnel, addition or deletion of sequence on either side of the crucial residues would alter their position relative to important binding partners in the tunnel. Similarly, mutation of important residues, such as Try17[51], could disrupt interactions between the peptide and the tunnel walls that are necessary for productive bypassing. A crosslinking study exposed interactions between the N-terminus of the gp60 nascent peptide and the mouth of the exit tunnel. Surprisingly, however, crosslinks were also found between the N-terminus of gp60 and the peptidyl transferase center. These data suggest this peptide might fold back up into the exit tunnel[59] or perhaps never use the exit tunnel at all[54]. In the vicinity of the peptidyl transferase center, the nascent peptide could affect structure and/or chemistry of peptide bond formation or termination[59]. Additional crosslinks were found between the N-terminus of the gp60 and small ribosomal subunit proteins S1, S2, S3 and S4[52]. Proteins S3 and S4 are thought to be involved in the RNA helicase activity of the ribosome[60]. These data suggest that the nascent peptide could play a role in inactivating helicase elements, allowing the ribosome to skip over structured elements in the coding gap of gene *60* mRNA.

It is interesting to note that the important amino acids of the gp60 nascent peptide are not part of the conserved topoisomerase. Instead, these residues are part of the N-terminal portion of gp60 that was added by *mobA* insertion (Figure I.2 and I.4). Thus the sequence of the nascent peptide and the coding gap could have evolved in a cooperative manner, ensuring efficient bypassing levels and survival of their host phage.

Additional Mechanistic Insights into Translational Bypassing

The coding gap begins with an in-frame stop codon, UAG, which ordinarily signals termination via release factor 1 (RF1). Herr et al. hypothesized that after decoding the preceding Gly codon, bypassing could compete with normal termination at this stop codon[23]. To test this hypothesis, RF1 was overexpressed *in vivo* and bypassing efficiency decreased, suggesting direct competition between these two processes. Depletion of RF1 with a temperature sensitive mutant, however, did not lead to an increase in bypassing efficiency. Instead, no change was observed in the level of bypassing[23]. This result indicated there is no direct competition between termination and take-off under normal cellular levels of RF1. Furthermore, take-off does not appear to be the limiting factor in determining bypassing efficiency. Rather landing of the peptidyl-tRNA at the correct codon is likely limiting[42].

Once take-off has occurred, how does the ribosome bypass the coding gap? A clever experiment was designed by Wills et al. to probe this question[27]. They replaced the take-off codon with UCC (and made complementary mutations to maintain base-pairing in the stem of the 5' hairpin) and looked at the protein products of bypassing by mass spectrometry. The mass of two of the proteins detected corresponded to forward bypassing by landing at synonymous codons within the coding gap. The third protein detected, however, clearly demonstrated backtracking of the ribosome after take-off by 7 nucleotides to land at a UCC codon in the -1 frame. To explain this phenomenon, the authors constructed the following hypothesis. During decoding of the take-off codon, the ribosome must disrupt the hairpin at the 5' edge of the coding gap. Once the peptidyl-tRNA dissociates from the mRNA, the hairpin and the rest of the coding gap "snap" into a folded structure (perhaps mimicking the structure of a tRNA) and occupy the A-site of the ribosome. By pulling the mRNA into the A-site from both the 5' and 3' directions, the peptidyl-tRNA becomes aligned over a matching codon upstream of the take-off site, resulting in backwards bypassing. Applying these results to wildtype forward bypassing, after the coding gap folds, the peptidyl-tRNA must move past this structure. Then, with the help of the SD-like element, the ribosome aligns the peptidyl-tRNA over the landing codon and continues translating[27].

Although little direct evidence was provided to support the above hypothesis, the coding gap region has been shown to fold in the presence of the 30S ribosomal subunit[30]. 30S subunits were bound to an mRNA containing the coding gap sequence 8 nucleotides downstream of a SD element and the complex was stabilized with $tRNA^{Gly}$. Assuming the SD:aSD interaction was maintained, toeprints of these complexes exhibited a much smaller band than would be expected by a linear mRNA template. In fact, the toeprint was of the correct size for the entire coding gap to be folded between the SD and the $tRNA^{Gly}$ in the P site. Shorter than expected toeprints were also found for other mRNA templates with potential to form structural elements between the SD and P-site or after the P-site[30]. The SD of ribosomal protein S1 provides additional evidence that mRNA can be structured. The SD of S1 mRNA is composed of nucleotides from the loops of two separate hairpins in its 5' UTR. Furthermore, its start codon lies in the loop of a third hairpin[61]. While these examples demonstrate mRNA folds are tolerated in the presence of the S30 subunit, it is unknown whether these structures also fold (have room to fold) in the presence of both ribosomal subunits or in the context of translation.

Thesis Objectives

Based on the literature described above, the structure of gene *60* mRNA appears to play a significant, but poorly understood role in translational bypassing. While there is substantial evidence for the hairpin at the 5' edge of the coding gap, data for the 3' half do not conclusively support or refute any of the proposed structures for this region. Despite its potential importance, to date no direct structural studies have been performed on gene *60* mRNA. The overarching goal of this thesis is to generate structural models of gene *60* mRNA using chemical probing techniques and to correlate these models with translational bypassing activity. With this biochemical and structural foundation, single-molecule fluorescence techniques will be devised to monitor structural changes in the coding gap mRNA during translational bypassing.

Chapter II

Development of an *In Vitro* Translation Assay

Introduction

A future goal of this project is to study translational bypassing of gene *60* mRNA via single-molecule level fluorescence microscopy. In order to observe translation on a slide, we must accurately replicate the environment of the cell and provide all components necessary for translation in the appropriate quantities. One way to accomplish this is to individually purify all proteins required for translation, which is the most optimal scenario as it ensures each component is RNase free and can be added or subtracted separately to study specific steps of translation. However, more than 30 proteins (three initiation factors, three elongation factors, three release factors and a recycling factor, and twenty aminoacyl-tRNA synthases), purified ribosomes and tRNAs, an energy generating system, amino acids, folinic acid, NTPs, T7 RNA polymerase (for coupled transcription/translation), buffers and salts are needed for optimal translation[62-64]. Given this extensive list, purification and development of assays to ensure high activity of each component would be very time consuming (Appendix C). Alternatively, a kit containing all purified components was developed by Dr. Takuya Ueda at the University of Tokyo and aptly named protein synthesis using recombinant elements or PURE system[62,63]. This kit can be purchased from New England Biolabs (E6800S), but is rather expensive (~$10 per 10 μL assay, channel volume of slide requires ~ 100 μL). Also, initial tests with this kit did not produce sufficient protein to be observed by Coomassie staining (data not shown), thus this kit was not used further.

Cellular extracts have been used for decades to probe the mechanism of various steps of translation[65-73]. They are inexpensive, comparatively easy to make and can be readily optimized for efficient translation[65-67,70,73-82]. Furthermore, cell extracts eliminate

the pressure to maintain cell viability enabling overexpression of proteins that are cytotoxic or have low solubility in cytosol, such as membrane proteins[66,70,74,76-79]. Extract composition can be adjusted by preparing extracts from strains where specific components are either deleted, depleted (if an essential component) or upregulated [83]. Additionally, individual protein components can be removed after the extract is prepared by immunoprecipitation[84-86]. Independently purified and fluorescently labeled proteins can be added back to extracts depleted of a particular component to observe their roles during translation at the single molecule level. Thus, although extracts are not precisely defined, they provide sufficient flexibility to study labeled translational components.

A bulk *in vitro* translation reaction in extract is a straightforward method by which to optimize translation conditions and ensure functionality of translational components that have been fluorescently-labeled for single molecule detection. Radiolabeled methionine (^{35}S-Met) or cysteine (^{35}S-Cys) can be added to the extract along with a template of interest, a buffering system, an energy regenerating system and amino acids; products of translation can be resolved by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE)[70]. After drying, bands in the gel can be quantified to compare efficiency of translation under various conditions. The modest changes generated by radiolabeling the products of translation provides a simple, non-intrusive way to follow translation and eliminates the need for large protein purification or signal tags that could alter the stability of the protein products or mRNA transcripts[29].

There are many published methods by which translational extracts can be made and many different variables can be varied to optimize translation of a specific product[65-67,70,73-82]. As such, several different methods and multiple bacterial strains were used to prepare cell extracts (Appendix B). Additionally, components of the translation reaction were varied to determine the best translation and translational bypassing conditions for gene *60*. These factors include pre-mix buffer composition, gel buffer system, acrylamide:bis-acrylamide ratio, gel percentage, gel drying techniques, mRNA concentration, mRNA preparation/purification method, amino acid concentration, concentration of extract, etc. This chapter outlines the steps taken to optimize each component of the translation reaction and initial tests of translational competency of fluorescently-labeled ribosomes for use in single-molecule translation assays.

General Conditions for Bulk *In Vitro* Translation Reactions

General reaction conditions are based on the methods of Dr. Rachel Green's lab at The Johns Hopkins University (Appendix C). All components are stored at -20°C except for the extract and ribosomes, which are kept at -80°C. Extract, ribosomes, template, S30 pre-mix buffer and radioactive amino acids are thawed at room temperature and then kept on ice. Amino acid mixes are thawed at 90°C and kept at room temperature. A master mix of S30 pre-mix, H_2O (as necessary), extract, ribosome (for S100 extracts only) and amino acids was prepared on ice. The master mix was incubated at 37°C for two minutes to free any ribosomes bound to cellular mRNAs. A radiolabeled amino acid (volume depended on age of stock, 0.1 – 1 μL, ~1 μCi per reaction) was then added to the master mix on ice and appropriate amounts of master mix were added to templates of interest. Final reaction volume was 10 μL. Reactions were incubated at 37°C for 30 min and quenched by mixing with 1 μL of 1 M KOH, which cleaved any peptides still bound to tRNAs, and placed on ice. Translation products were precipitated with 50 μL of cold acetone to remove PEG. Proteins were pelleted by centrifugation at 20,000 x g for 10 min at room temperature, the supernatant was removed by pipetting and pellets were air-dried. Samples were resuspended in 1X SDS loading buffer, heated to 90°C, then loaded onto an SDS protein gel. Gels were resolved with 100-200 V for 2-5 hrs, soaked in acid to fix the proteins and soaked in methanol to displace water before being dried onto Whatmann paper at 80°C for 1-2 hrs. Dried gels were exposed to a phosphor storage screen overnight and quantified by ImageJ software[87].

Optimization was carried out with gene *60* mRNA and various mutants thereof as model templates. As previously published, ribosomes pass over the coding gap region in about 50% of translation events[19,23,24,29,49,87]. Bands in our gels corresponded to the expected sizes of the non-bypassed (truncated, 7.8 kDa) and bypassed (full-length, 18.4 kDa) protein (gp60). Efficiency of bypassing was calculated with Equation II.1 and was used as a metric, along with overall translation competency (band intensity), to determine the optimal conditions in which to study this system.

Equation II.1

$$\%\ \text{bypass} = \frac{\text{full length gp60}}{\text{full length gp60} + \text{truncated gp60}}$$

Source of S30 Pre-mix Solution

Translational bypassing experiments carried out during my visit to Rachel Green's lab (Appendix C) were accomplished by adding cellular extracts to a pre-made solution from Promega (part of a translation kit, #L1110) containing buffer, salts, amino acids, NTPs, tRNAs, an enzymatic ATP regenerating system and polyethylene glycol (PEG). However, due to the high cost of this product (~$4 per 10 μL bulk reaction), we wished to prepare our own "pre-mix" solution from individual components. Additionally, use of Promega's pre-mix is not advantageous because its exact composition is unknown and we are unable to adjust the concentration of various components that might play important roles in translational bypassing. Upon recommendation of Hani Zaher (postdoctoral fellow for Rachel Green), a pre-mix solution was prepared as described by Kigawa et al.[75], the composition of which is outlined in Table II.1.

Component	2.5X S30 Pre-mix	Final Concentration
HEPES-KOH pH 7.5	137.5 mM	55 mM
potassium glutamate	525 mM	210 mM
ammonium acetate	68.8 mM	27.5 mM
magnesium acetate	26.8 mM	10.7 mM
dithiothreitol (DTT)	4.3 mM	1.7 mM
ATP pH 7.0	3 mM	1.2 mM
CTP pH 7.0	2 mM	0.8 mM
GTP pH 7.0	2 mM	0.8 mM
UTP pH 7.0	2 mM	0.8 mM
cAMP	1.6 mM	0.64 mM
folinic acid	0.17 mM	0.068 mM
E. coli total tRNA	0.438 mg/mL	0.175 mg/mL
creatine phosphate	200 mM	80 mM
creatine phosphokinase	188 units	75 Units
PEG 8000	10 %	4 %

Table II.1 Composition of 2.5X S30 pre-mix solution[75]. Pre-mix solution contains all buffer, salts, rNTPs for coupled transcription/translation, tRNA and an energy generating system. PEG is added to simulate molecular crowding. Buffer and salt components in the top section of the table denote the composition of "S30 buffer", in which structure probing experiments were carried out (see Chapter II "Degradation of mRNA", and Chapter III "Optimizing Structure Probing Conditions")

To compare our home-made pre-mix solution to that from Promega, translation of gene *60* mRNA was performed in four different extract preparations (Appendix B) with both pre-mix solutions (Figure II.1). Extract M2_S30 does not appear to be capable of bypassing. By contrast, the rest of the extracts produced two protein products corresponding to truncated gp60 and full-length gp60. Bypassing efficiencies between the two buffering systems were very similar (Figure II.1 B), indicating that we could use the recipe from Kigawa et al in further experiments. Additionally, S30 extracts appeared to be more efficient for translation and bypassing than S100 extracts (further comparisons between extracts are made below).

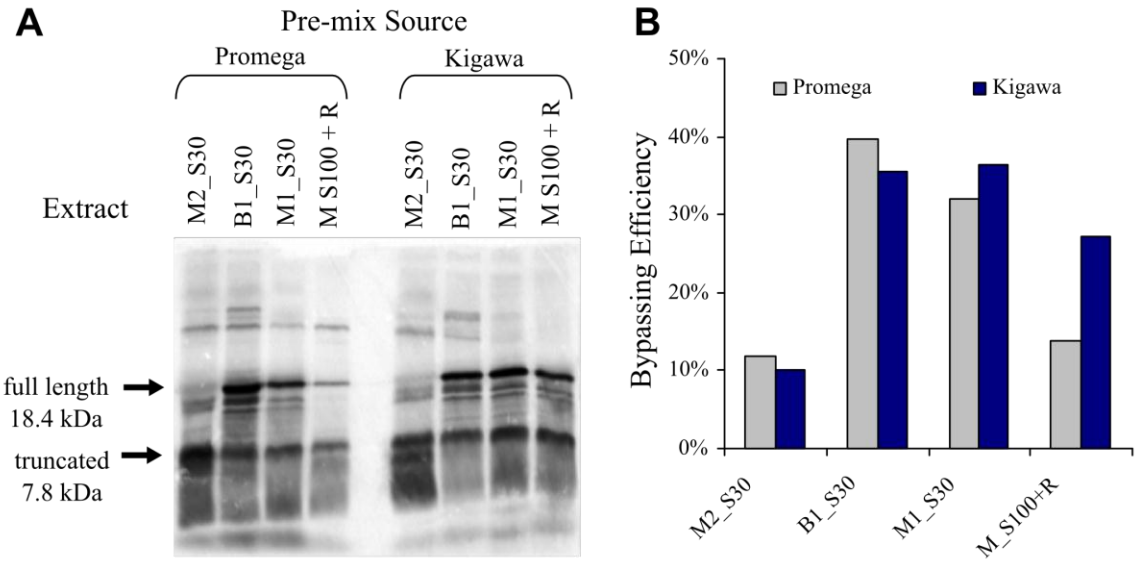


Figure II.1 Comparison of S30 pre-mix solutions. Translation was carried out in 10 µl reactions containing 4 µl of the indicated pre-mix solution, 4 µl of the indicated S30 extract or 3 µl S100 + 1 µl ribosomes, 400 µM amino acids (– Met), 25 pmol gene *60* mRNA and 0.5 µL ^{35}S-Met at 37°C for 30 min. Protein products were resolved on a 16% Tris-glycine gel (**A**). Calculation of bypassing efficiency using Equation II.1 for each extract in Promega (grey) or homemade (blue) S30 pre-mix solutions (**B**).

Protein Gel Composition

Due to the small size of both full-length (18.4 kDa) and truncated gp60 (7.8 kDa), we sought to find a gel condition that would give the best resolution in this range. The gel type recommended by Hani Zaher (Appendix C) for resolving our small proteins was 16% Tris-tricine gels prepared with acrylamide:bis-acrylamide ratio of 29:1, as described by Schagger & von Jagow[88]. To compare this condition to our standard SDS protein gel

protocol, aliquots from large-scale gene *60* translation reactions carried out in different extracts were resolved on 16% Tris-tricine or Tris-glycine gels prepared with acrylamide:bis-acrylamide ratios of 19:1 or 29:1 (Figure II.2). As expected, Tris-tricine gels gave better resolution for lower molecular weight proteins than Tris-glycine gels, a characteristic that is especially important for accurate quantification of the truncated product band. Furthermore, using an acrylamide:bis-acrylamide ratio of 19:1 enhanced resolution of shorter peptide fragments due to its smaller pore size. Thus, all further translation reactions were optimally resolved on Tris-tricine gels prepared from 19:1 acrylamide:bis-acrylamide.

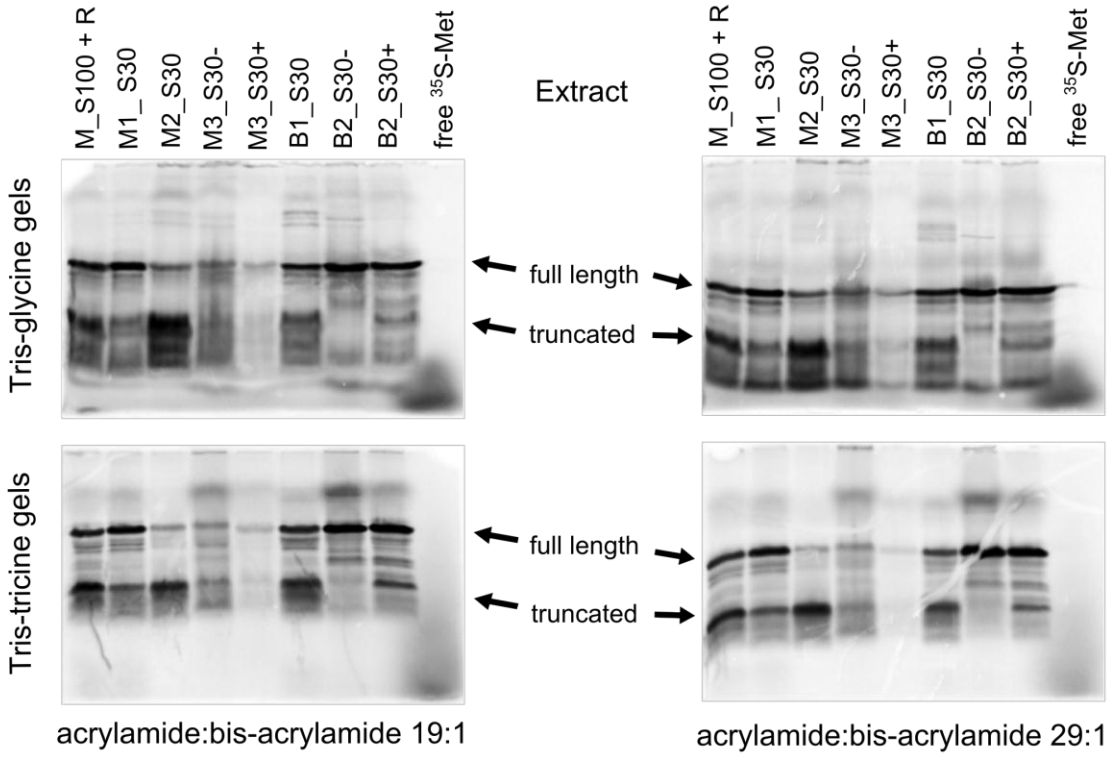


Figure II.2 Buffer composition and acrylamide:bis-acrylamide ratio of gels. Translation was carried out in 44 µl reactions containing 17.6 µl pre-mix solution, 13.2 µl of the indicated S30 extract or 9.9 µl S100 + 3.3 µl ribosomes, 800 µM amino acids (– Met), 100 pmol gene *60* mRNA and 2.2 µl ^{35}S-Met at 37°C for 30 min. 10 µl of each reaction was loaded onto each of the 16% gels with the indicated buffer and ratio of acrylamide:bis-acrylamide.

We note that extracts M2_S30, M3_S30- and M3_S30+ do not undergo significant bypassing, therefore these extract preparations were not used further for study

of translational bypassing. Tests were performed below to assess the benefit of translation carried out in the remaining extracts (Figure II.4). Free ^{35}S-Met runs as a diffuse band below all resolvable bands, indicating resolvable bands are peptide fragments not amino acids (Figure II.2). Additionally, due to its diffuse gel electrophoretic mobility, free ^{35}S-Met was not quantifiable. For further discussion of free ^{35}S-Met see section on "Quantification of gp60 Production and Calculation of Bypassing Efficiency".

Protein Gel Dehydration Conditions

^{35}S is a weak Beta emitter and cannot readily penetrate through a thick aqueous gel[89,90]. Therefore, our protein gels must be dried before exposure to a phosphor storage screen to obtain adequate signal. Gels were dried onto 1 mm Whatman paper using a Biorad Model 583 Gel Dryer attached to a Savant GP110 vacuum pump. Application of heat and a strong vacuum caused the gel to shrink in the z-dimension resulting in a thin film embedded on the filter paper.

After running, the gels were fixed in an acid solution to prevent diffusion of bands during the drying process. As described in Appendix C, gels were initially soaked in 10% trichloroacetic acid solution (TCA) for 10-15 min per wash with two washes. After fixation, gels were immediately dried at 80°C for 1 hr. Although this method worked well in the Green Lab, in our hands these conditions frequently resulted in cracking and shrinking of the polyacrylamide gel to a point where bands were no longer quantifiable.

Several variables were tested to determine the best conditions for reduction of gel cracking during drying. First, the default setting on the gel dryer rapidly heats the gel to the desired temperature and maintains that temperature until the end of the drying cycle. According to the Biorad Model 583 Gel Dryer manual, a slow ramp up to the desired temperature reduces the likelihood of gel cracking. Furthermore, the length of the drying cycle was increased from one hour to one and a half hours, and the maximum temperature reached during the drying cycle was reduced to 75°C. Finally, 3mm Whatmann paper adheres better to gels than 1mm Whatmann paper. Each of these adjustments resulted in mild improvement of gel drying.

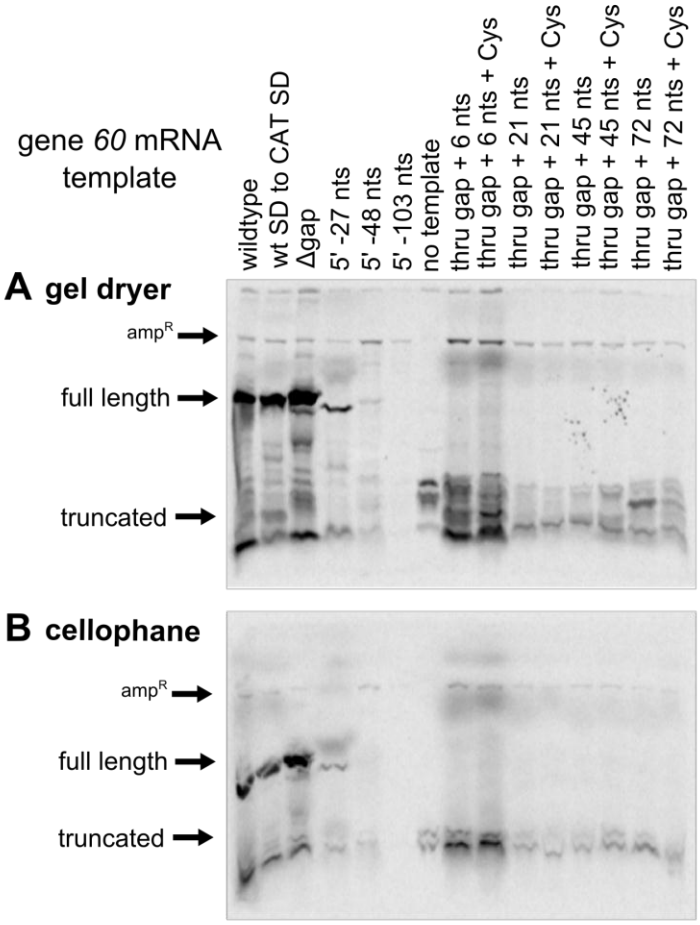


Figure II.3 Comparison of gel drying methods. Translation was carried out in 20 µl reactions containing 8 µl pre-mix solution, 4 µl B2_S30+ extract, 1 µL homemade T7 RNA polymerase, 0.4 µl ^{35}S-Met and 100 ng indicated gene *60* mutant plasmid DNA templates at 37°C for 30 min. 10 µL of each reaction was resolved on each 20% Tris-tricine gel and dried with a gel dryer (**A**) or with porous cellophane (**B**). The laminating layer of the cellophane prevents β particles from interacting with the phosphor screen, significantly reducing the intensity of the bands. Band above full-length corresponds to ampicillin resistance gene expressed from pUC19 backbone.

To further enhance our gel drying procedure, we compared two alterations to soaking procedure for fixing protein samples before drying. First, after soaking gels twice in 10% TCA, a third 15 minute soaking step in 50% (v/v) methanol and 10% (w/v) PEG-8000 was added[90]. The second method combined acid fixation with the methanol exchange into a single soaking step in 5% (v/v) glycerol, 40% (v/v) methanol and 10% (v/v) acetic acid for 1 hr (Biorad Model 583 Gel Dryer manual). The purpose of these additional steps is to replace the water in the gels with methanol, which has a lower boiling point and will therefore evaporate more quickly during the drying process. The glycerol or PEG in these solutions also diffuses into the gels enhancing their drying

properties. Although there was no noticeable difference in resistance to cracking between the two drying methods, using the single soaking solution saved time and reagents and was thus used for drying all further gels.

An alternate method for drying gels was also tried. After soaking, gels were sandwiched between two sheets of a porous cellophane membrane (The Gel Company, #EJA345-050) and clamped around the edges to a glass plate. These gels were left in a well ventilated area overnight (such as a fume hood) and all moisture evaporated through the membrane leaving behind a laminated gel. To test this method for gel drying, a large scale translation reaction was split between two gels. The first gel was dried using the gel dryer as described above and the second gel was dried in the cellophane membrane. Lamination of the gel significantly reduces the ability of ^{35}S beta particles to reach the phosphor screen and low intensity bands are not visible[90] (Figure II.3). Thus, although a simple and convenient way of drying gels, this method was not useable for our optimization experiments.

Finally, according to the Biorad Model 583 Gel Dryer manual, the vacuum pressure required for drying gels should be ~25 torr. Actual vacuum pressure being pulled by our pump, however, was only ~250 torr, or 10 times weaker than required. Our vacuum pump was serviced and the diaphragms replaced. While greatly reducing the frequency and extent of gel cracking, this repair did not completely eliminate cracking. It is likely that a stronger or newer pump is needed to draw adequate vacuum for reproducible gel drying.

Optimization of mRNA Concentration

As the concentration of gene *60* mRNA expressed during Bacteriophage T4 infection is not known, a titration was performed in multiple extracts to determine the optimal concentration of mRNA needed to observe bypassing signal (Figure II.4). Interestingly, each extract showed different patterns of gp60 production as well as background bands (see 0 nM template lanes). M1_S30, B1_S30 and B2_S30- all contained significant background bands that overlapped the full-length gp60 band. Although both B1_S30 and B2_S30+ had background bands that overlapped the truncated gp60 band, B2_S30+ had significantly lower background overall and translated

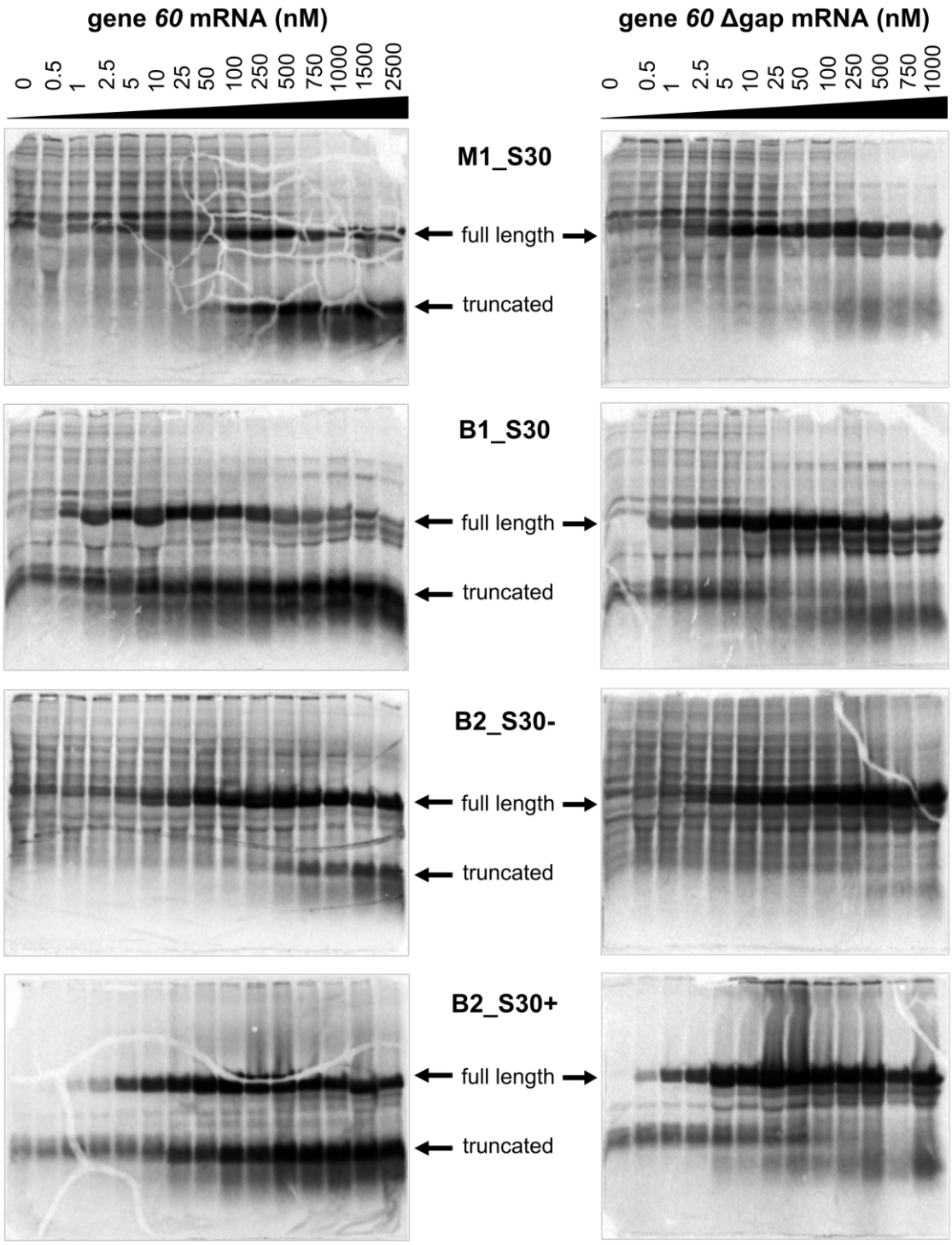


Figure II.4 Titration of mRNA in various extracts. Translation was carried out in 10 µl reactions containing 4 µl pre-mix solution, 3 µl of the indicated S30 extract (Appendix B), 800 µM amino acids (–Met), 0.1 µl ^{35}S-Met and the indicated final concentration of gene *60* mRNA or gene *60* Δgap mRNA (nM) at 37°C for 30 min. Protein products were resolved on a 16% Tris-tricine gel. Extract B2_S30+ (bottom row) had the lowest overall background and was used for all future experiments. (Note: concentrations are estimations as these mRNAs were generated as in Appendix C and still contain linearized DNA template from their transcriptions (see "Purification Method for mRNA Templates"). All extracts generated from

BL21 Star (DE3) cells contain T7 RNA polymerase (Appendix B), which could further increase the concentration of mRNA by producing fresh transcripts.)

efficiently (Figure II.4, gel on the bottom left). Therefore we chose to move forward using B2_S30+ extract for all future experiments (Appendix B).

To accurately estimate the bypassing efficiency of gene *60* in B2_S30+, we needed a way to subtract out the background overlapping the truncated gp60 band. A titration of gene *60* delta gap mRNA (gene *60* mRNA that does not contain the internal stop codon and coding gap region and therefore should only produce full-length gp60, Appendix A) in B2_S30+ (Figure II.4, gel on bottom right) revealed that addition of more than 100 nM of mRNA effectively eliminated the background band intensity, presumably because gene *60* delta gap mRNA was out-competing the contaminating mRNA for translation on the ribosomes. The background band intensity in the gene *60* delta gap gel was subtracted from the sum of the truncated gp60 and the overlapping contaminant in the wildtype gene *60* gel (Figure II.5). These corrected protein levels revealed that truncated gp60 increased continually with increasing mRNA concentration, while the amount of full-length gp60 leveled off above 100 nM mRNA. Correspondingly, the efficiency of bypassing decreased drastically when more than 100 nM of mRNA was present. Based on this trend, 50 nM of mRNA template was used in all future reactions.

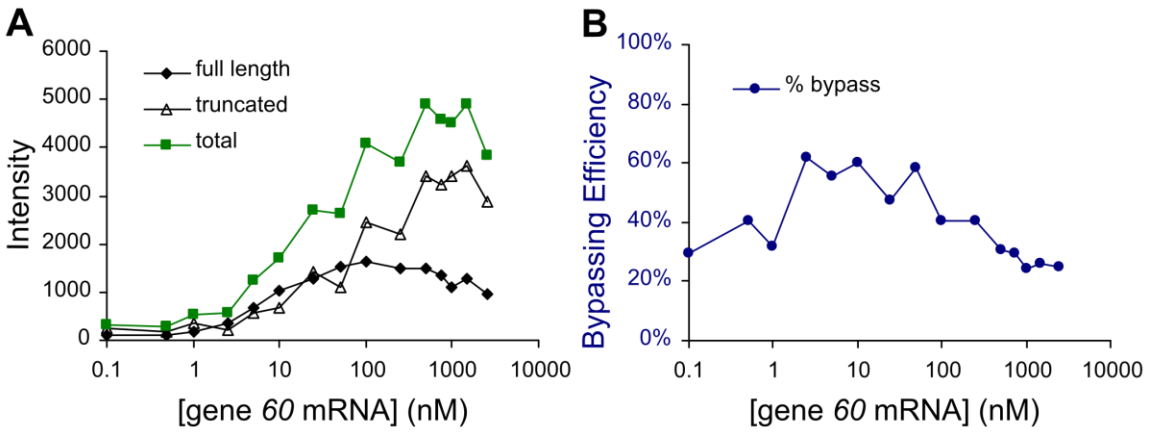


Figure II.5 Quantification of full-length and truncated gp60 as a function of mRNA concentration for translation reactions in B2_S30+ extract. The background band from gene *60* Δgap mRNA titration was subtracted from the sum of truncated gp60 and the background band from the gene *60* mRNA titration (bottom gels, Figure II.4) to estimate intensity of just the truncated gp60 (**A**). Bypassing efficiency was calculated with these corrected values for truncated gp60 (**B**). Note: data from the 0 nM mRNA lane (Figure II.4)was plotted as 0.1 nM due to the prohibitive nature of 0 on a logarithmic x-axis.

The observed decrease in bypassing efficiency could be explained if high concentrations of mRNA are specifically or non-specifically titrating important cellular factor(s) required for the bypassing mechanism. Alternatively, the phage could tune the amount of full-length gp60 produced by inhibiting bypassing at high gene *60* mRNA concentrations, possibly through a multimerization of the mRNA. This hypothesis, however, is ruled out in Chapter IV. Finally, degradation products (below the truncated gp60 band) increase in intensity with increasing mRNA concentration with a wildtype gene *60* template but not as extensively with gene *60* delta gap template. Thus a good portion of the degraded product appears to be due to instability of the truncated gp60, which could lead to an *over-estimation* of bypassing efficiency at higher mRNA concentrations and a corresponding *under-estimation* of the decrease in bypassing efficiency. Further testing is required to determine the cause of this phenomenon.

Purification Method for mRNA Templates

We next wanted to determine if the method by which mRNA is generated affects translation or bypassing efficiency. All mRNAs were transcribed *in vitro* from a linearized plasmid template using homemade T7 RNA polymerase (Appendix A), but potential subsequent purification steps were varied (Figure II.6). In Dr. Green's lab, transcription reactions were simply desalted and precipitated (lanes 11 and 12). These samples still contain linearized DNA template from the transcription reaction, hindering accurate calculations of mRNA concentration. Additionally, because extracts from *E. coli* BL21 strains contain their own T7 RNA polymerase, fresh transcription from the DNA template could further adjust the concentration of mRNA. Although linearized template alone does not seem to generate significant amounts of gp60 (lane 2), we attempted to remove the DNA template with Promega's RQ1 RNase-free DNase (#M610A). All DNase treated samples showed a reduction in translation and bypassing efficiency (lanes 4, 5 and 6), which was due to mRNA degradation (data not shown). On the other hand, mRNA templates purified by denaturing PAGE showed high levels of both translation and bypassing, therefore all further experiments were performed with gel purified mRNA templates (unless otherwise specified).

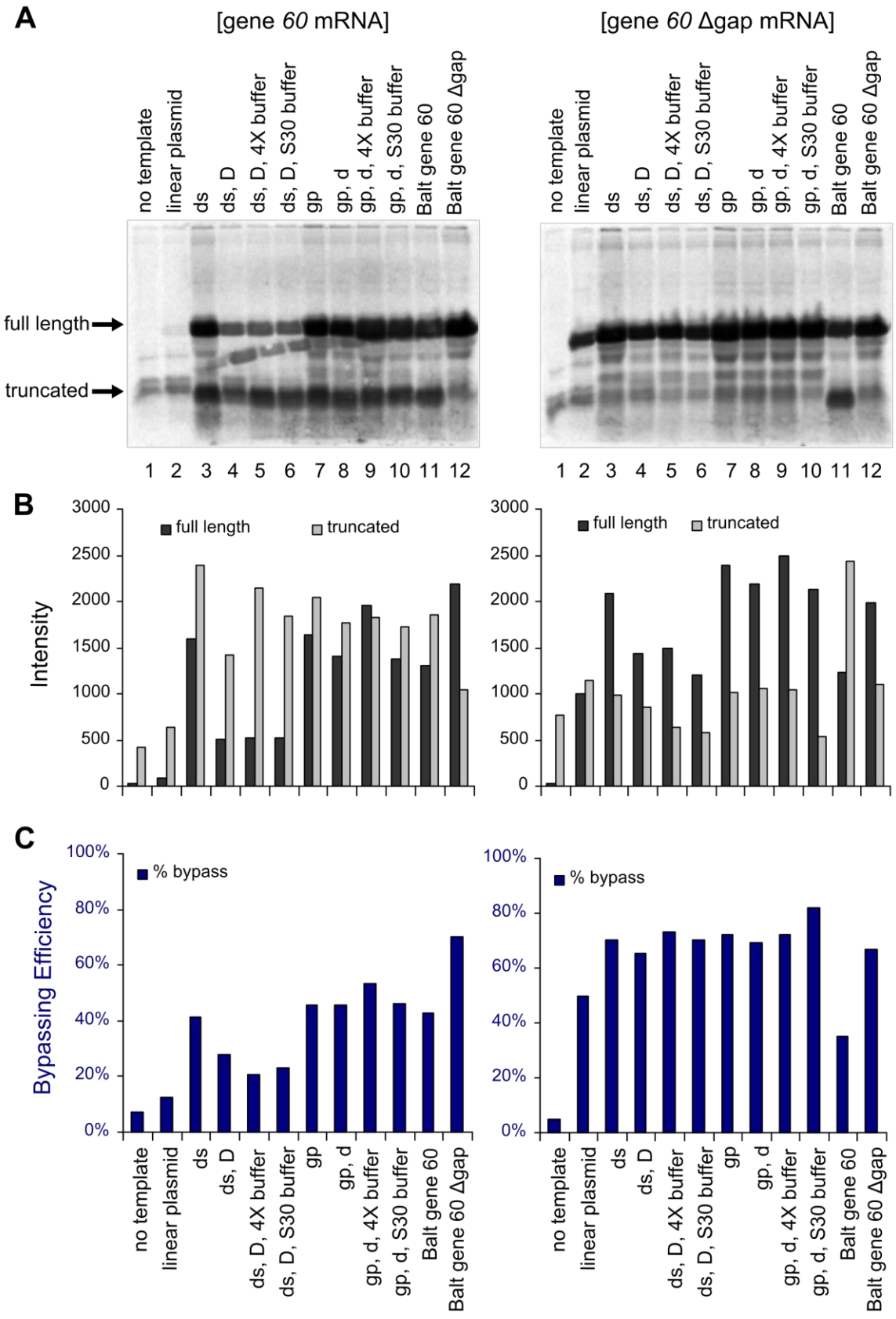

A
[gene 60 mRNA]
[gene 60 Δgap mRNA]
no template
linear plasmid
ds
ds, D
ds, D, 4X buffer
ds, D, S30 buffer
gp
gp, d
gp, d, 4X buffer
gp, d, S30 buffer
Balt gene 60
Balt gene 60 Δgap
full length
truncated
1 2 3 4 5 6 7 8 9 10 11 12
B
Intensity
3000
2500
2000
1500
1000
500
0
full length
truncated
C
Bypassing Efficiency
100%
80%
60%
40%
20%
0%
% bypass

Figure II.6 Effect of purification method and refolding conditions of mRNA on translation. Translation was carried out in 10 µl reactions containing 4 µl pre-mix solution, 3 µl of the indicated 30 extract, 0.3 µl ^{35}S-Met and 0.7 µL template at 37°C for 30 min. Linear plasmid – pUC19 gene *60* digested with BsrBI; ds – desalted; D – DNase treated; 4X buffer – renatured in 100 mM Tris-HCl pH 7.0, 100 mM NaCl, 10 mM $MgCl_2$; S30 buffer – renatured in buffer outlined in top of Table II.1; gp – gel purified; d – denatured; templates for lanes 11 and 12 were prepared in the Green lab (as described in the text). Protein products were resolved on a 16% Tris-tricine gel (**A**). Raw intensities of full-length and truncated bands (**B**) and calculation of bypassing efficiency using Equation II.1 (**C**).

Additionally, we assessed the impact of denaturing and refolding gel purified mRNA to determine if structures in gene *60* mRNA played an important role in either translation initiation or bypassing. No effect was seen from either denaturing or refolding mRNA prior to its addition to the translation reaction (lanes 8, 9 and 10). It is likely that the mRNA structure refolds optimally upon addition of cell extracts regardless of pre-folding conditions. Because no difference was observed by either denaturing or refolding mRNA prior to translation reactions, all further experiments were carried out by simply adding gel purified mRNA directly to the reaction.

Degradation of mRNA

Structure probing experiments on gene *60* mRNA (as described in Chapter III) were performed in the S30 pre-mix solution (Table II.1) used for translation reactions to mimic the buffer conditions gene *60* mRNA experiences during translation. Incubation in this solution, however, resulted in severe degradation of mRNA even in the presence of RNase inhibitor (Applied Biosystems, #N8080119) (Figure II.7, lanes 16 and 17, respectively). mRNA was incubated with each component of the S30 pre-mix solution separately and creatine kinase was found to be the major factor causing degradation in the S30 pre-mix solution (Figure II.7, lane 14).

Creatine kinase is an essential component of the pre-mix because it generates ATP during the translation reaction. To determine if there was another ATP generating system without significant RNase activity, we compared our batch of creatine kinase from Sigma (#C3755, not guaranteed to be RNase-free, email correspondence with Sigma technician) to creatine kinase from Roche (#10127566001) and pyruvate kinase from MP Biomedicals (#151999). mRNA was incubated with each kinase in the presence and absence of an RNase inhibitor (Applied Biosystems, #N8080119).

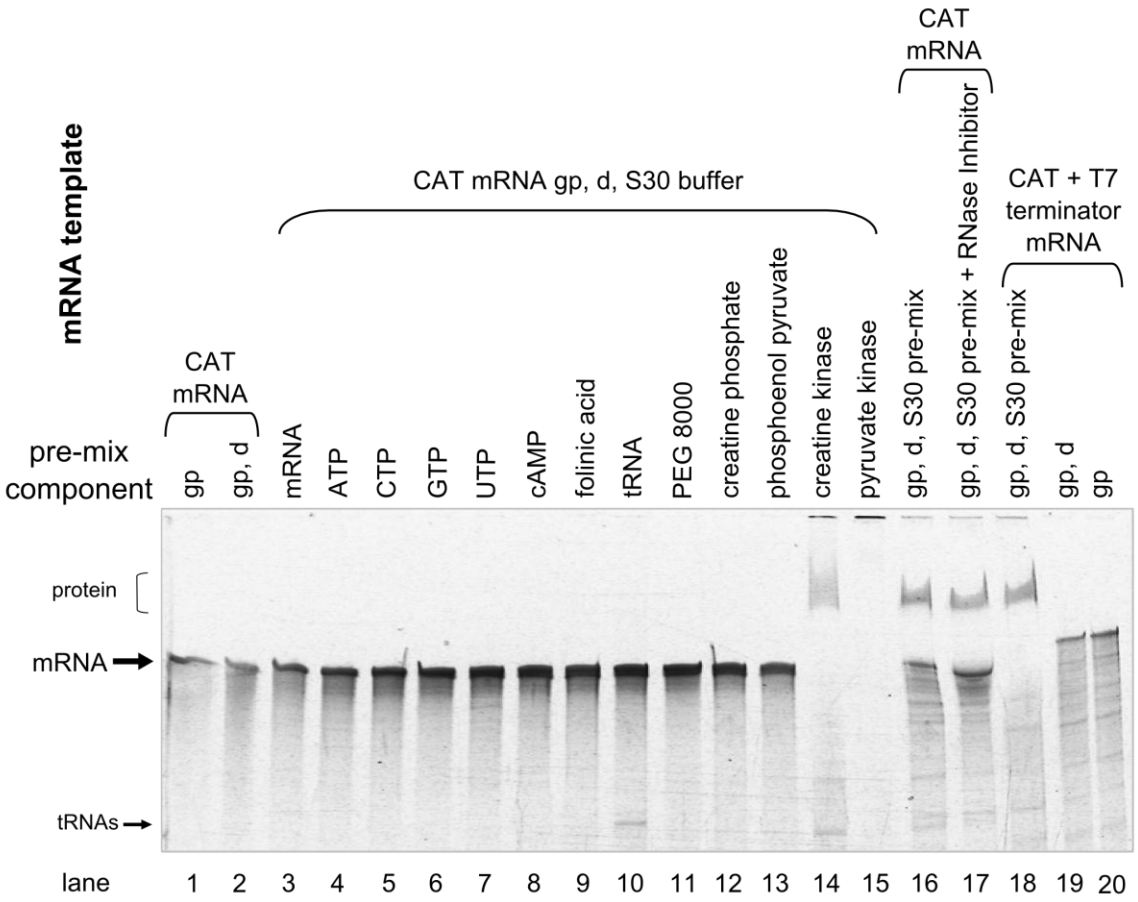


Figure II.7 Determining the component of S30 pre-mix that degrades mRNA. 300 ng of CAT mRNA gel purified (gp), denatured (d), refolded in S30 buffer (S30 buffer) was incubated with the indicated individual component of the S30 pre-mix at 37°C for 45 min (lanes 1-15). 300 ng of CAT mRNA was refolded in S30 pre-mix, with and without RNase inhibitor, at 37°C for 45 min (lanes 16 and 17). 300 ng of CAT mRNA with a 3' T7 terminator hairpin gel purified (gp), denatured (d), refolded in S30 pre-mix (S30 pre-mix) was incubated at 37°C for 45 min. Integrity of mRNA was observed by SYBR Gold staining of a 5% denaturing PAGE run at 25 W for 2 hrs and 45 min.

Compared to gel purified, denatured and refolded mRNA in S30 buffer (Figure II.8, lanes 1-3), incubation with S30 pre-mix containing creatine kinase from Sigma generated extensive degradation of the mRNA that was significantly reduced upon addition of RNase inhibitor (lanes 10 and 11, respectively). mRNA incubated with each kinase generated varying levels of degradation (lanes 4, 6 and 8) that were somewhat abrogated by an RNase inhibitor (lanes 5, 7 and 9). Creatine kinase from Roche exhibited the least RNase activity of the three and was therefore used in all further S30 pre-mix preparations.

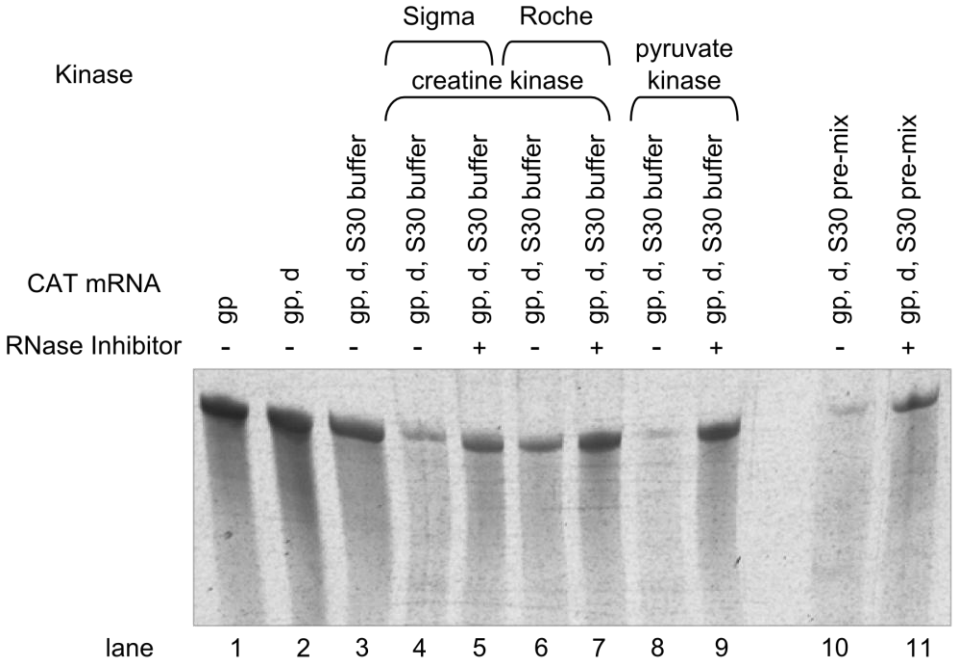


Figure II.8 Testing kinases for RNase activity. 230 ng of CAT mRNA gel purified (gp), denatured (d), refolded in S30 buffer (S30 buffer) or refolded in S30 pre-mix (S30 pre-mix) were incubated with 75 U/mL of the indicated kinase, with and without RNase inhibitor (+/-), at 37°C for 45 min. Integrity of mRNA was observed by SYBR Gold staining of a 5% denaturing PAGE run at 25 W for 4 hrs and 25 min. Note: this batch of mRNA is old and already somewhat degraded. However, further degradation from the RNase contaminants is still readily detectable.

Given that all kinases contained some RNase contaminants, we wanted to determine whether or not addition of an RNase inhibitor to the translation reactions would improve yield (Figure II.9). Surprisingly, however, translation of in the presence of RNase inhibitor resulted in generally lower levels of translation for both full-length and truncated gp60 (Figure II.9, grey and light green bars). Therefore RNase inhibitor was not included in further experiments.

A major source of mRNA degradation comes from 3'-5' exonucleases[91-94]. The action of these nucleases, however, can be mitigated by the presence of a stable hairpin at the 3' end of a transcript. Specifically, the T7 terminator hairpin (Figure II.10 A) seems to provide the best protection; hairpins of non-specific sequence or modified T7 hairpins with different stem lengths (longer or shorter) provide protection to a lesser extent[74]. Although we did not observe enhanced stabilization of CAT mRNA with the T7 terminator hairpin during incubations with S30 pre-mix alone (Figure II.7, lane 18), this modification enhances translational efficiency of transcripts in extracts (Figure II.10 B

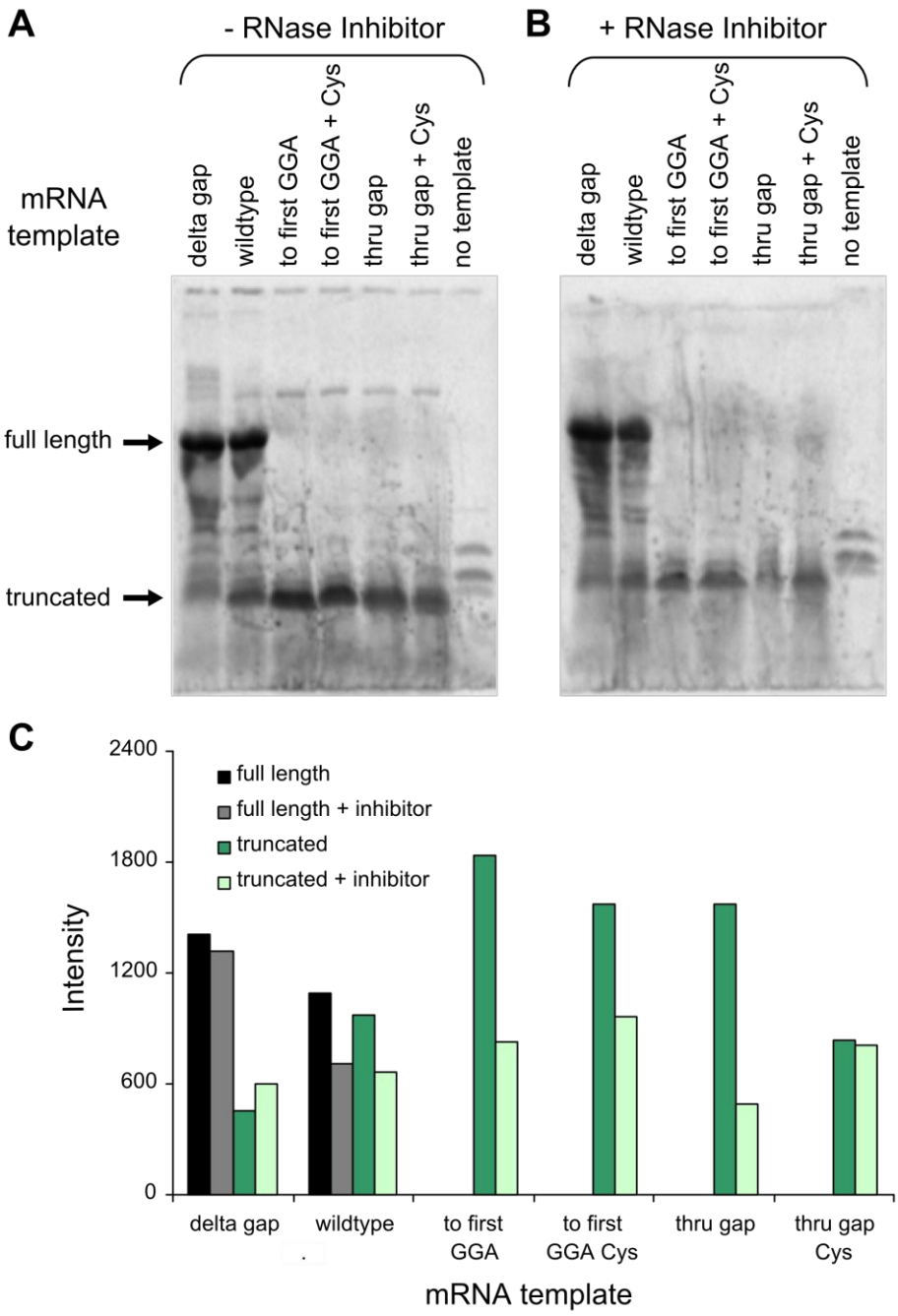


Figure II.9 Translation levels are altered by the addition of an RNase inhibitor. Translation was carried out in 10 µl reactions containing 4 µl pre-mix solution, 2 µl of B2_S30+ extract, 0.5 µl ^{35}S-Met, 50 nM mRNA and either with (**B**) or without (**A**) 0.03 U of Promega RNasin (#N211A) at 37°C for 30 min. Protein products were resolved on a 20% Tris-tricine gel and quantified with ImageJ (**C**).

and C). A 3-fold increase was observed in GFP production and an 8-fold increase was observed for CAT production. The disparity between our pre-mix and translation data for these transcripts could be because the RNase contaminants in the S30 pre-mix are

predominantly endonucleases or 5'-3' exonucleases. Cellular components of the extract may protect the mRNA from these types of nucleases with RNA binding proteins or translating ribosomes.

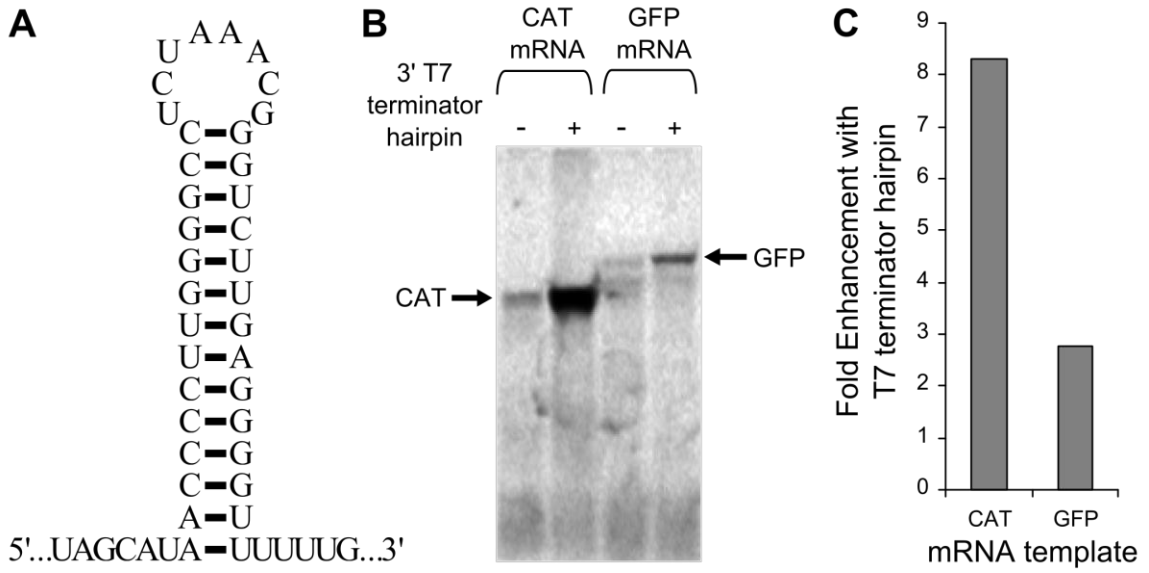


Figure II.10 Addition of a T7 terminator hairpin enhances translation. (**A**) T7 terminator hairpin sequence. (**B**) Translation was carried out in 10 µl reactions containing 4 µl pre-mix solution, 3 µl of B2_S30+ extract, 800 µM amino acids (-Met), 1 µl ^{35}S-Met and 4.6 nM CAT mRNA or 10 nM GFP mRNA at 37°C for 30 min. Protein products were resolved on a 20% Tris-tricine gel. (**C**) Fold enhancement in protein yield with a 3' T7 terminator hairpin on the transcript.

Optimization of Amino Acid Concentration

The amount of ^{35}S-Met added to each reaction varied from 85 nM to 850 nM due to its decaying specific activity over time ($t_{1/2}$ = 87.4 days[95]). The concentration of all other amino acids in the reaction ranged from 400 to 800 µM. As there was ~1,000-fold less Met added to each reaction than the other amino acids, we wanted to determine whether the concentration of ^{35}S-Met was limiting translation. A titration with increasing concentrations of ^{35}S-Met (of the same specific activity) was performed with gene *60* mRNA in B2_S30+ (Figure II.11). Although gp60 bands get slightly more intense with increasing amounts of ^{35}S-Met, they do not increase 5-fold as would be expected if the concentration of Met was limiting translation. This lack of correlation is likely because the cell extract contains an excess stock of charged Met-tRNA^{Met} that were generated during the aminoacylation step of the B2_S30+ preparation (Appendix B). Furthermore,

the bypassing efficiency (or relative amounts of each gp60 band) is also independent of the concentration of ^{35}S-Met. Thus we can safely increase the amount of ^{35}S-Met in the reaction as it decays without affecting translation or bypassing efficiency.

A titration of cold amino acids (-Met) was performed to determine if the levels of charged tRNA generated during the pre-aminoacylation step during preparation of B2_S30+ (Appendix B) were sufficient to fuel our translation reactions (Figure II.12). The extent of protein production varied less than two-fold at each concentration of amino acids (-Met) and bypassing efficiencies were nearly identical. Interestingly, when no additional amino acids (-Met) were added to the translation reaction, no significant decrease in gp60 production or bypassing was visible. Thus additional amino acids are not required for our translation reactions when tRNAs aminoacylation has been performed during preparation of the extract (Appendix B).

In Chapter III the secondary structure of gene *60* mRNA was determined in S30 buffer; however this structure in buffer may differ significantly from the structure that dominates during translational bypassing. We therefore looked for a way to isolate gene *60* mRNA poised or undergoing bypassing as a substrate for structure probing experiments. We hypothesized that withholding individual amino acids (specifically Gly, as the take-off Gly is the first instance of this amino acid in gp60) would be able to stall translation just before bypassing. However, in B2_S30+ translation is unaffected

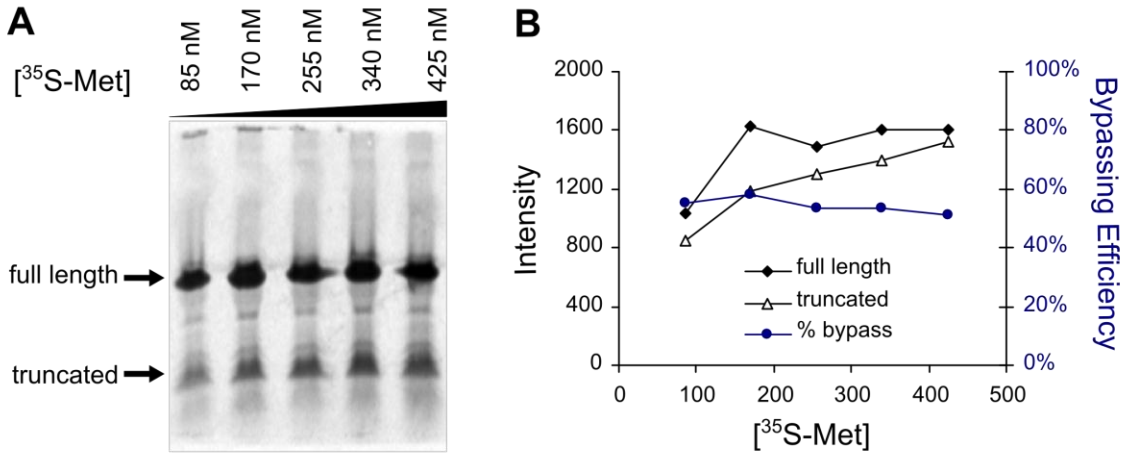


Figure II.11 ^{35}S-Met added to translation reactions is not limiting translation. Translation was carried out in 10 µl reactions containing 4 µl pre-mix solution, 3 µl of B2_S30+ extract, 800 µM amino acids (-Met), 50 nM mRNA and 0.1 to 0.5 µl ^{35}S-Met at 37°C for 30 min. Protein products were resolved on a 16% Tris-tricine gel (**A**). Quantification of bands and calculation of bypassing efficiency with ImageJ (**B**).

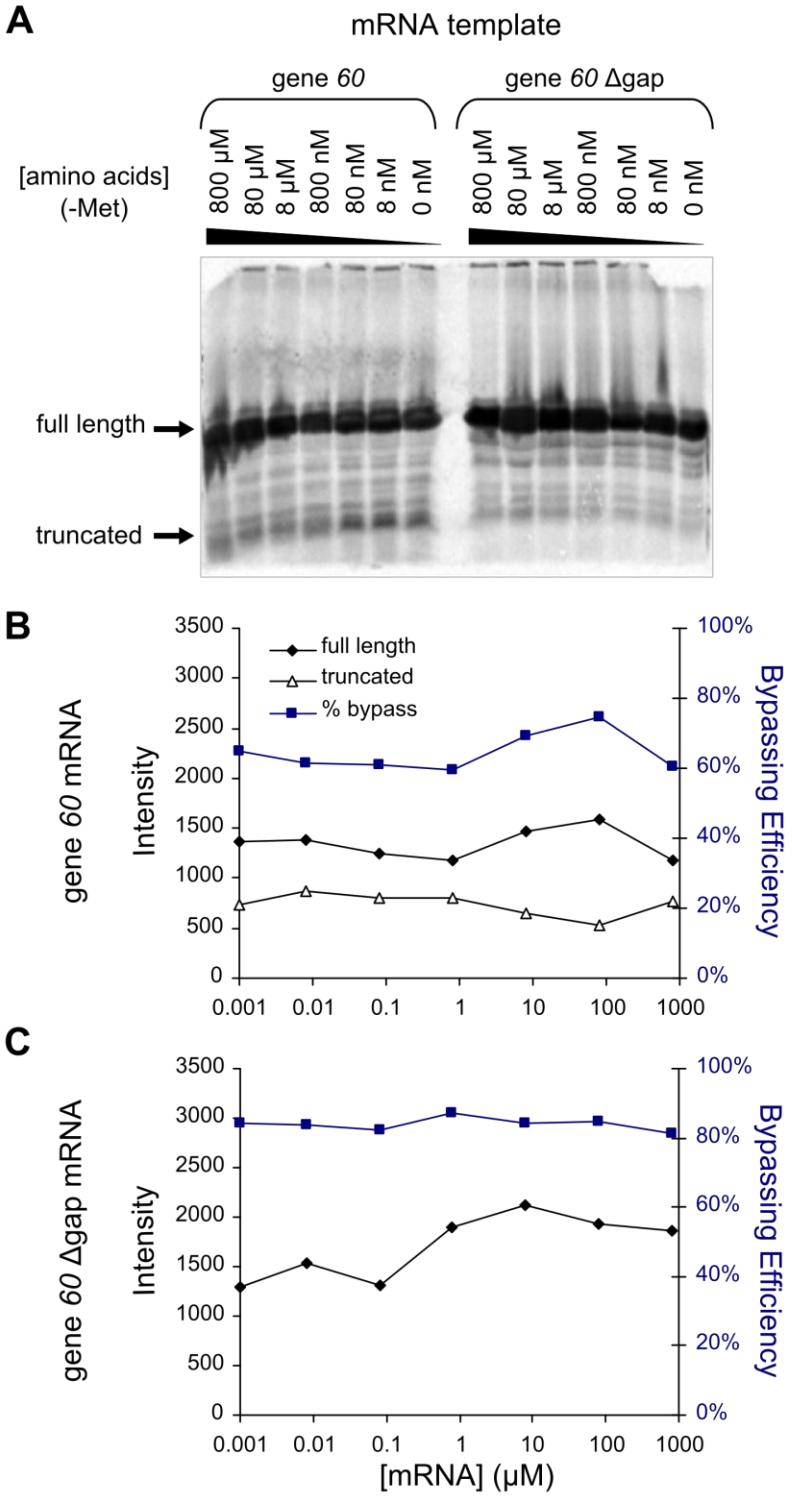


Figure II.12 Additional amino acids are not needed for translation in B2_S30+ extract. Translation was carried out in 10 µl reactions containing 4 µl pre-mix solution, 3 µl of B2_S30+ extract, 50 nM mRNA of the indicated template and 0.1 µl ^{35}S-Met and the indicated amount of amino acids (-Met) at 37°C for 30 min. Protein products were resolved on a 16% Tris-tricine gel (**A**). Quantification of translation and

bypassing efficiency for gene *60* mRNA (**B**) and gene *60* delta gap mRNA (**C**). Note: data from the 0 nM amino acid lanes is plotted as 0.001 nM due to the prohibitive nature of 0 on a logarithmic x-axis.

by the addition of free amino acids to the reaction (Figure II.12) due to the high concentration of charged tRNAs generated during preparation of the extract. For this reason we tested the amino acid dependence of translation in B2_S30- extract, which did not undergo the aminoacylation step (Appendix B) and therefore might allow us to control translation by the addition of amino acids.

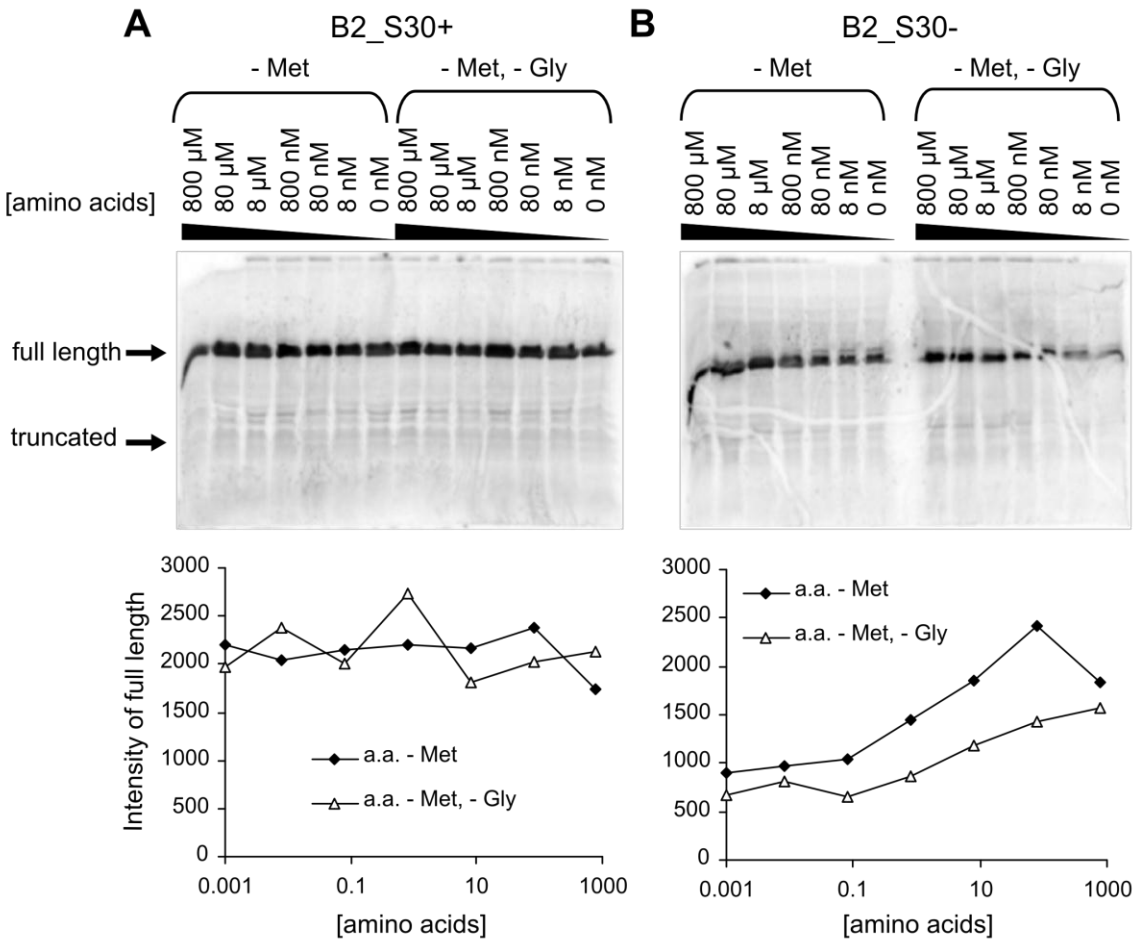


Figure II.13 Translation can be limited by amino acid addition in extract B2_S30-, but not in extract B2_S30+. Translation was carried out in 10 µl reactions containing 4 µl pre-mix solution, 3 µl of B2_S30+ (**A**) or B2_S30- (**B**) extract, 50 nM gene *60* mRNA and 0.1 µl ^{35}S-Met and the indicated amount of amino acids (-Met) or amino acids (-Met and -Gly) at 37°C for 30 min. Protein products were resolved on a 20% Tris-tricine gel and quantified with ImageJ.

A titration of amino acids without Met or without both Met and Gly was performed in B2_S30+ and B2_S30- extracts using ^{35}S-Met labeling (Figure II.13).

Depletion of amino acids without Met or without both Met and Gly did not affect translation in B2_S30+ extract. Thus, this extract already possesses sufficient amounts of charged Met and Gly tRNAs such that amino acids are not limiting during translation. In B2_S30- extract, however, depletion of amino acids without Met or without both Met and Gly reduced overall protein production significantly. There was still some residual translation when no additional amino acids are added, which can be attributed to metabolite synthesizing enzymes that are present in the extract. Based on these results, a simple amino acid starvation is not sufficient to stall translation at crucial points during bypassing. All further experiments were carried out in B2_S30+ extract without adding additional amino acids.

Observing gp60 Production as a Function of Time

Reprogramming of ribosomes for frameshifting can generate a measurable pause during translation[17,20,21,96,97]. Given the apparent complexity of the translational bypassing event, we would also expect a significant delay in translation for the bypassing event. We anticipated this could be observed as a temporary build up of truncated gp60

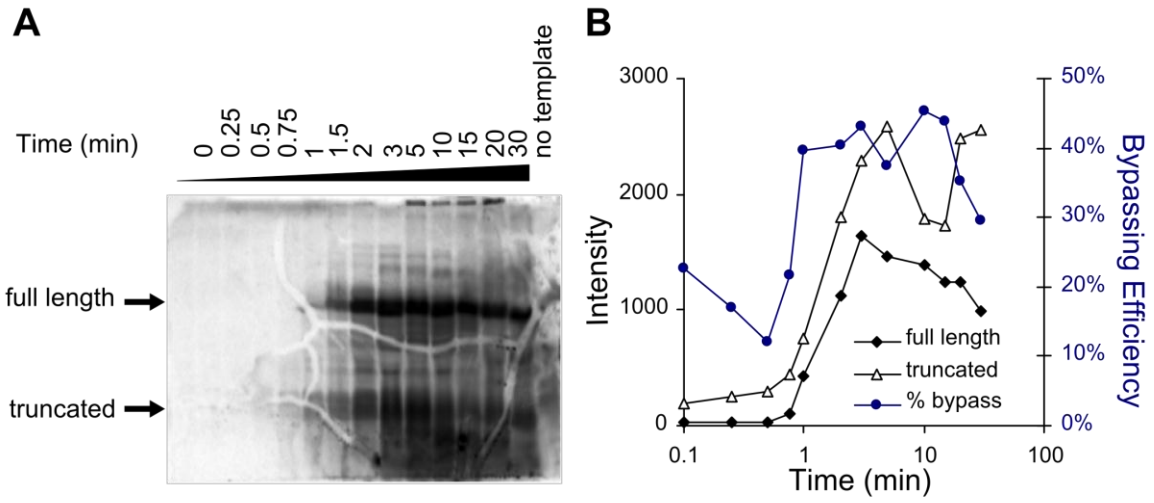


Figure II.14 Production of gp60 as a function of time. (**A**) Translation was carried out in a 160 µl reaction containing 64 µl pre-mix solution, 32 µl of B2_S30+ extract, 50 nM gene *60* mRNA and 4.8 µl ^{35}S-Met at 37°C for 30 min. 10 µL aliquots of the reaction were removed at the indicated time points and quenched with 1 µL of 1 M NaOH. Protein products were resolved on a 20% Tris-tricine gel. (**B**) Quantification of full-length and truncated gp60 and calculation of bypassing efficiency as a function of time.

while the ribosome is undergoing bypassing, followed by a decrease in truncated gp60 as elongation resumes after bypassing.

To estimate the amount of time required for bypassing, aliquots from a large-scale translation reaction were removed and quenched over time (Figure II.14). Quantification of protein bands revealed that the rate of formation of full-length gp60 is only slightly slower than that of truncated gp60, which could be simply due to the different lengths of the two proteins. Thus there is not an exorbitant delay in translation for the bypassing event. Maximum translation of both species occurs after just 5 minutes. Furthermore, significant amounts of degradation (smear below truncated product) begin to accumulate after 3 min. The majority of degradation appears to originate from the truncated gp60 band, indicating that long incubation times could lead to an *over-estimation* of bypassing efficiencies. Thus all further translation reactions could and should be carried out for < 5 min at 37°C.

Although we did not see a delay for bypassing in this assay, it is likely that our time resolution is simply too low. Estimations in the literature suggest that translation occurs at a rate of 10-20 amino acids per second[98]. For the lengths of gene *60* this would mean ~3-5 s for truncated gp60 and 8-16 s for full-length gp60. Further time courses (presented in Chapter IV) were performed on shorter time scales to address this issue.

Another important feature of this time course in Figure II.14 is the ~45 s lag between the start of the reaction and the first detectable protein production. The reaction was assembled on ice and placed at 37°C at time 0. Thus the delay in translation is likely because the reaction took several seconds to warm to 37°C. For further time courses (see Chapter IV) separate tubes containing mRNA/^{35}S-Met and extract/pre-mix were pre-incubated at 37°C for 1 min before combing to start the translation reaction.

Finally, there is no way of distinguishing if the kinetics of this time course were collected under single- or multiple-turnover conditions. The reaction time was long enough and translation fast enough that ribosomes could make multiple passes along the same mRNA. To eliminate this ambiguity, further time courses were performed with an initial pulse of ^{35}S-Met followed by a large excess of cold Met. This allowed us to

observe only the first round of translation of an mRNA under single-turnover conditions. The outcome of these experiments will be discussed further in Chapter IV.

Optimizing the Concentrations of S30 Extract and S30 Pre-mix

To determine the best concentration of B2_S30+ to add to each 10 µl reaction the amount of extract added was varied from 0-60% (v/v) (Figure II.15). Upon quantification of the full-length band, we find that 20% (v/v) extract resulted in the most optimum conditions for translation and was used in all subsequent reactions. In addition, a titration of the 2.5X pre-mix solution was also performed and resulted in an optimum amount of 30% (v/v) instead of the 40% (v/v) expected. The difference in yield is small and could be a result of pipetting error. Because the exact source of this difference is unknown, the S30 pre-mix buffer was used at the published concentration of 40% (v/v) for all future experiments.

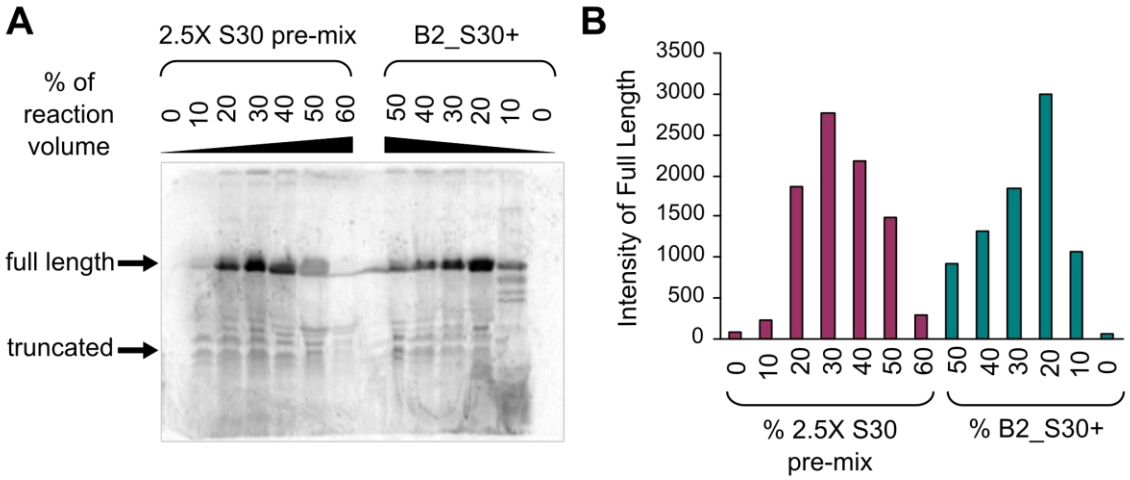


Figure II.15 Optimization of S30 extract and S30 pre-mix concentrations. (**A**) Translation was carried out in 10 µl reactions containing either 4 µl pre-mix solution or the indicated amount, 3 µl of B2_S30+ extract or the indicated amount, 50 nM gene *60* mRNA and 0.1 µl ^{35}S-Met at 37°C for 30 min. Protein products were resolved on a 20% Tris-tricine gel. (**B**) Quantification with ImageJ reveals optimal concentrations for B2_S30+ extract and S30 pre-mix solution.

Quantification of gp60 Production and Calculation of Bypassing Efficiency

Proper quantification of truncated and full-length gp60 bands is required to obtain accurate bypassing efficiencies. Intensity of bands was determined using the software ImageJ[87]. To use these intensity values, we must assume that ^{35}S-Met gets incorporated

into proteins at the same rate as unlabeled Met. And, although not all Met residues in our protein may be radioactive, statistically each site should be occupied to the same extent with ^{35}S-Met. Full-length gp60 contains eight Met residues whereas truncated gp60 has only four so the intensity of full-length gp60 bands was halved for molar comparison with truncated gp60 and for calculation of bypassing efficiency using Equation II.1. Finally, the concentration of tRNAs charged with ^{35}S-Met will increase initially after the ^{35}S-Met is added to the extract at the rate at which tRNAs charged with cold Met are recycled. Thus initial estimates of ^{35}S-Met incorporation may slightly *underestimate* the rate of translation.

Ideally, accurate comparison of bypassing efficiency between different gene *60* mutants can only be made if we are able to account for loading errors between different lanes. One method by which this correction can be done is to normalize band intensities to the total amount of radioactivity added to each lane. However, the total radioactivity is significantly greater in the presence of an added template compared to the background bands when no template is added. Therefore the total radioactivity added to each reaction is not necessarily incorporated into protein products that we can resolve by SDS PAGE.

To account for these differences in protein band intensity, unincorporated ^{35}S-Met must be included in the sum of total radioactivity per lane. Under standard conditions, however, the free ^{35}S-Met is not retained as a band on the gel (Figure II.16, lanes 5-8, Figure II.2). We hypothesized that free ^{35}S-Met could be lost during the acetone precipitation step required to remove PEG from the translation reaction. PEG is used to simulate molecular crowding during translation, but causes irregular migration of protein through the gel (Figure II.16, compare lanes 1-4 with lanes 5-8; Figure C-2 A, left three lanes). To circumvent the acetone precipitation step, but to maintain viscosity in the translation reaction, a pre-mix solution was prepared containing 10% glycerol instead of 10% PEG. Glycerol is a normal component of gel loading buffers (~25% final) and does not affect migration of proteins in an SDS PAGE. Despite elimination of the precipitation step, free ^{35}S-Met was not visible on the gels as a quantifiable band (Figure II.16 lanes 1-4, 9 and 11; Figure II.2) because it migrates quickly and diffuses rapidly.

Thus, we are unable to quantify protein bands a fraction of the total radioactivity in a lane. All quantification of band intensities in this chapter represents raw intensity

measurements and contains error due to loading variability. Calculations of bypassing efficiency, however, represent the fraction of gp60 in the full-length band relative to all gp60 in the lane (full-length + truncated) therefore these values can be compared among different lanes of a gel.

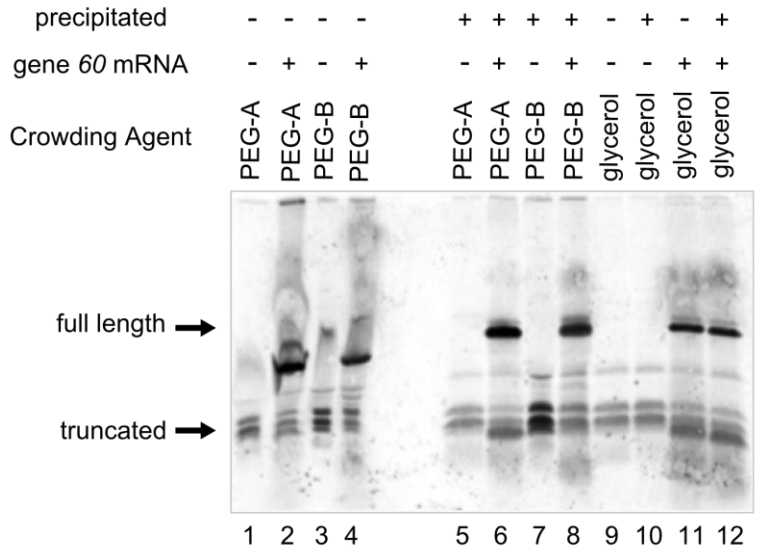


Figure II.16 Comparison of molecular crowding agents with and without precipitation of protein products. Translation was carried out in 40 µl reactions containing 16 µl of S30 pre-mix solution with the indicated crowding reagent, 12 µl of B2_S30+ extract, with or without 50 nM or gene *60* mRNA and 0.8 µl ^{35}S-Met at 37°C for 30 min. 10 µL of each reaction, with or without prior acetone precipitation (+ and -, respectively), were resolved on a 16% Tris-tricine gel. Two preparations of S30 pre-mix with PEG (PEG-A or B) and one preparation with glycerol were tested.

As a substitute for total lane intensity, an internal translation product can also be used as loading control. All cloned gene *60* mutants share a pUC19 backbone, and if these plasmids are added to extracts for coupled transcription/translation the ampicillin resistance gene on the plasmid is expressed and labeled with ^{35}S-Met along with gp60. The ampicillin resistance protein (Amp^R) is 27 kDa[99], which easily separates from gp60 bands. Furthermore, the transcription/translation of Amp^R should be constant regardless of the gene *60* mutant because it has its own promoter sequences. Thus all gp60 bands can be normalized to the intensity of the ampicillin resistance protein band to control for loading differences in each lane (see Figure II.3). Alternatively, a small amount of a second mRNA, whose protein product significantly differs in size from gp60, can be added to the extract master mix before it is split among various templates. We chose to add CAT mRNA with a T7 terminator hairpin at its 3' end, which greatly enhances its

stability in extracts (Figure II.10). Intensities of gp60 bands can be normalized to that of the product for the CAT mRNA (26.5 kDa) allowing for direct comparison between lanes. (Note: these two internal control methods are best used for single time-point translation

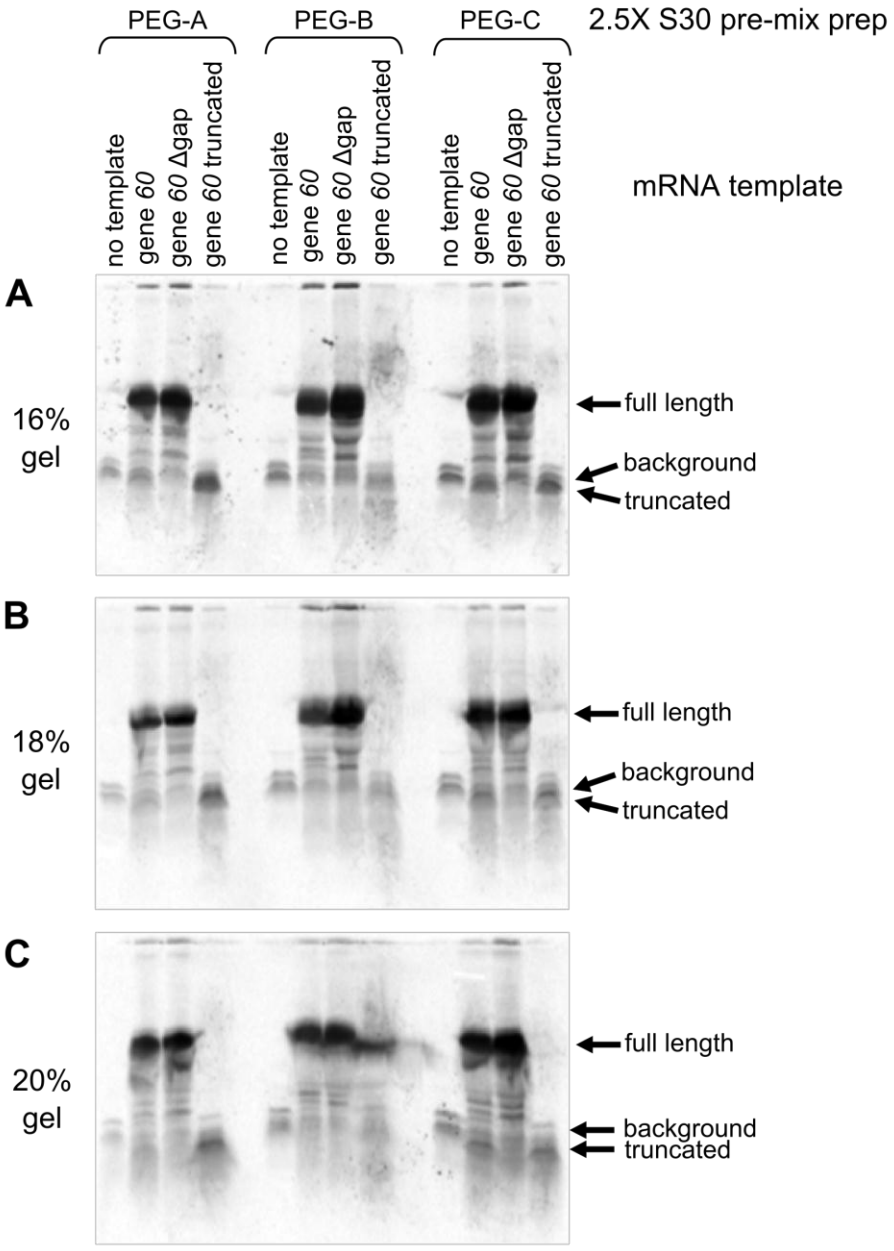


Figure II.17 20% gels are best able to separate truncated gp60 from the background band. Translation was carried out in 35 µl reactions containing 14 µl of the indicated S30 pre-mix solution (three different preparations with PEG – A, B and C), 10.5 µl of B2_S30+ extract, 50 nM of the indicated gene *60* mRNA mutant and 1.1 µl ^{35}S-Met at 37°C for 30 min. 10 µL from each reaction was resolved on gels with 16% (**A**), 18% (**B**) or 20% (**C**) acrylamide.

experiments. The intensity of control bands would change during a time course and therefore are not useful in these types of experiments.)

Finally, difficulty with quantification in B2_S30+ extracts comes from a contaminating mRNA that produces a background band that overlaps with the top of our truncated gp60 band. In an effort to separate these bands from one another (and to avoid the estimations made in Figure II.5), the percentage of the gel was increased from 16 to 20%. At 20% acrylamide we can just separate the background band from that of the truncated product (Figure II.17). Thus 20% acrylamide gels were used to resolve products of all future translation reactions.

Confirmation of gp60 Product Identity

There are two main protein products produced from translation of gene *60* mRNA. Although their molecular weights match those expected for full-length and truncated gp60 relative to other translation products, we wanted to confirm their identity through mutational analysis. Serendipitously, gp60 contains no natural Cys residues. The second to last residue of gp60, Ser 159, was mutated to a Cys to mark translation of the entire ORF of gene *60*. Ser 8 was also mutated to a Cys as a second way to observe both products of translational bypassing. These mutations were introduced into both wildtype gene *60* as well as gene *60* delta gap mRNA (Figure II.18 A). Transcripts were translated in the presence of either ^{35}S-Met, to compare to normal conditions and confirm the templates are translationally active, or ^{35}S-Cys, to determine the identity of protein products. Gene *60* S8C mRNA produced two protein products identical to those of wildtype in both the ^{35}S-Met and ^{35}S-Cys conditions, but only produced full-length gp60 for gene *60* delta gap S8C mRNA in the presence of ^{35}S-Cys as expected (Figure II.18 B and C). For S159C mutants, only a band corresponding to full-length gp60 bands was observed in all conditions. Thus the band we call full-length gp60 is, indeed, a product of translation of the entire gene *60* ORF.

To confirm the identity of the truncated gp60 band, gene *60* mRNA was truncated after the in-frame stop codon at the beginning of the coding gap. A Cys residue was added just after the take-off Gly but before the stop codon for this construct. The resulting protein product was identical in size to that of truncate gp60, confirming the

identity of this band (Figure II.18 B and C). Finally, according to the literature[28], the shortest gene *60* mRNA that can undergo bypassing has been truncated 5 nucleotides after the coding gap and replaced with a *lacZ* gene. To test the bypassing ability of this truncated mRNA, we generated a mutant of gene *60* that stops 6 nucleotides after the coding gap. The product of bypassing in this construct and truncated gp60 cannot be separated due to the very small difference between their molecular weight (2 amino

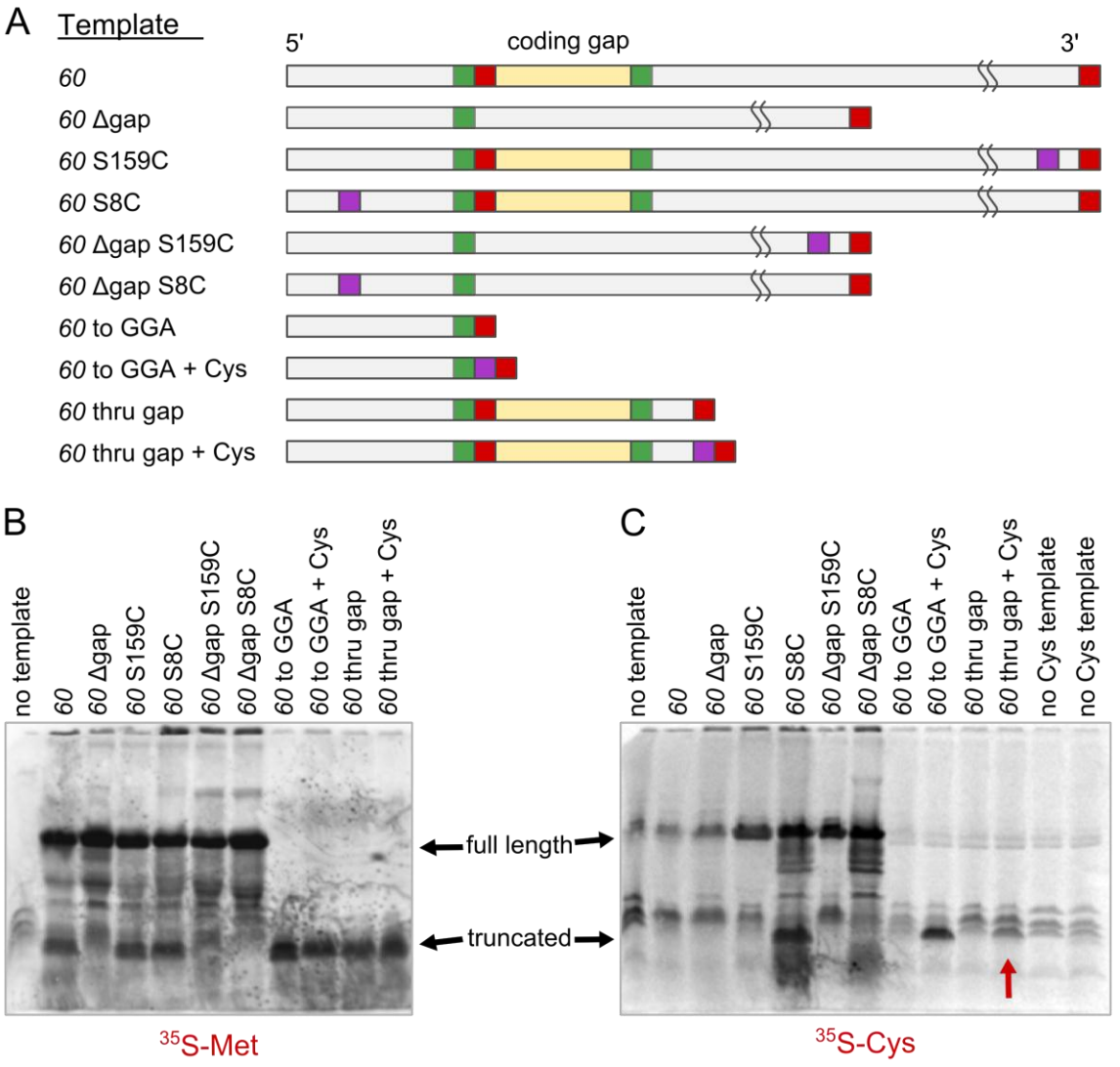


Figure II.18 Cys mutations to confirm the identity of protein products from gene *60* mRNA. (**A**) S159C and S8C mutations were introduced into gene *60* and gene *60* delta gap mRNAs. Gene *60* mRNA was truncated at the internal stop codon and 6 nucleotides after the stop codon. A Cys residue was added to each of these truncations just before the stop codon. Translation products from indicated mRNA templates in the presence of ^{35}S-Met (**B**) and ^{35}S-Cys (**C**). Full-length and truncated gp60 are indicated. The red arrow indicates ambiguity in the ability of gene *60* mRNA truncated 6 nucleotides after the coding gap to undergo translational bypassing.

acids). Therefore, we added a Cys residue after the 6 nucleotides following the coding gap. Translation of this mRNA in the presence of ^{35}S-Cys should only give a product the size of truncated gp60 if the translational bypassing has occurred. However, as indicated by the red arrow in Figure II.18 C, a product of the appropriate size cannot be detected over background. Three bands of similar intensity and location are also apparent in two templates lacking any Cys residues (rightmost lanes, Figure II.18 C). It could be that the efficiency of bypassing in this construct is too low to observe with this assay or because truncated gp60 tends to degrade rapidly once formed (Figure II.14), the bypassed product could have degraded during the 30 min incubation. The original, functional truncations of gene *60* to this point also added a significant amount of in-frame codons of the *lacZ* gene such that the ribosome could continue translating well after the bypassing event. Thus the ribosome may require some additional sequence after the coding gap as a landing platform. Extensions of varying lengths were constructed after the coding gap with wildtype gene *60* sequence and a C-terminal Cys marker. However, due to overlap of background bands with the ^{35}S-Cys-labeled bypassed products, we were unable to determine which mutants were competent in translational bypassing (data not shown).

Translational Competency of Fluorescently-Labeled Ribosomes

The initial plan for single-molecule translation assays involved the use of a mutant ribosome designed in Dr. Phil Cunningham's lab at Wayne State University. The Cunningham lab replaced the natural anti-Shine Dalgarno (aSD) sequence at the 3' end of the small subunit rRNA with its complementary SD sequence. Similarly, they generated mRNAs whose SD was changed to an aSD sequence such that it would only be translated by mutant ribosomes. This swap of SD/aSD sequences has several advantages. *E. coli* has several copies of genes coding for ribosomal RNA, which makes it difficult to generate a homogenous population of ribosomes with a desired mutation. Furthermore, if mutations made to a ribosome reduce translation efficiency they could result in poor growth phenotypes or cell death. Expressing ribosomes with a mutant aSD (mut aSD) from a plasmid, however, allows mutant ribosomes to be produced without affecting translation of normal cellular transcripts. Additional alterations can be made to ribosomes with mut aSD and expression of a mutant SD (mSD) GFP gene can provide a

relative measure of translational efficiency of different ribosomal mutants. Finally, the population of mut aSD ribosomes can be isolated from wildtype ribosomes by passing cell lysate over a column to which short oligonucleotides containing a complementary mSD are bound[100].

The mut aSD ribosomes also would provide many advantages for studying translation at the single-molecule level. First, we were able to insert additional sequence into the non-conserved helix 33a[101] of the mut aSD ribosome plasmid. These ribosomes were expressed in *E. coli* and purified by the method described above. These cells could also be used to generate cellular extracts for translation. The extension of helix 33a was site-specifically labeled by hybridization with a fluorophore-labeled complementary

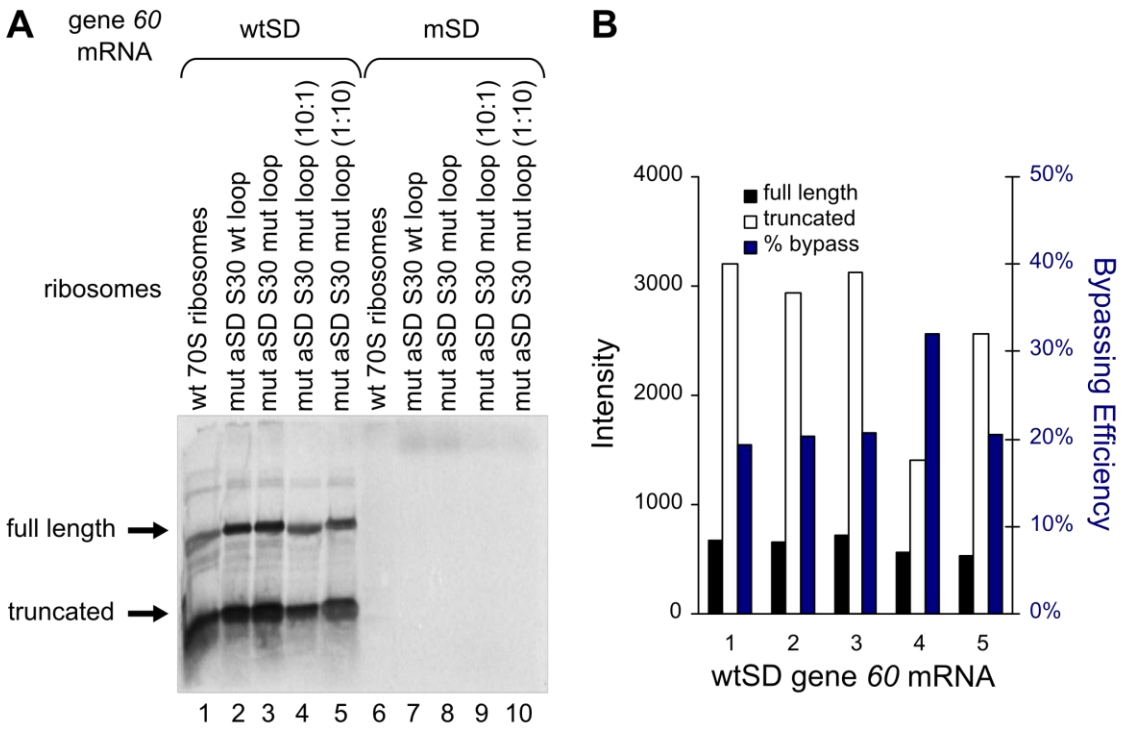


Figure II.19 Testing ribosomes with a mutated anti-Shine-Dalgarno sequence. (**A**)Translation was carried out in 10 µl reactions containing 4 µl of S30 pre-mix solution, 2.25 µl of S100 extract, 10 nM 70S ribosomes or 10 nM 50S subunits with 10 nM of S30 subunits with the indicated mutations, 800 µM amino acids (-Met), 50 nM gene *60* mRNA with wildtype or mutant Shine-Dalgarno sequences and 1.1 µl ^{35}S-Met at 37°C for 30 min. Protein products were resolved on a 16% Tris-tricine gel. wtSD – wildtype Shine-Dalgarno, mSD – mutant Shine-Dalgarno; mut aSD – mutant anti-Shine-Dalgarno; wt loop – wildtype sequence of helix 33a; mut loop – hairpin mutant 5 added to helix 33a[100,101]; 10:1 and 1:10 indicate the ratio of 30S subunit to fluorophore-labeled DNA oligo complementary to the extended sequence added to helix 33a. (**B**) Quantification of translation and bypassing using ImageJ.

DNA oligonucleotide. By using mSD mRNA templates, only mut aSD ribosomes containing the labeled extension of helix 33a would be visible in single molecule experiments.

We obtained purified ribosomal subunits from the Cunningham lab that contained mut aSD elements as well as extended loops for DNA oligo binding. Initial labeling reactions for the ribosome were carried out at ratios of 10:1 and 1:10 of fluorescently labeled DNA probe to ribosomes (work of Matt Marek, Walter Lab). These ribosomes were tested for translation in the presence of gene *60* mRNA with both a wtSD and mSD (prepared by Matt Marek, Walter Lab) (Figure II.19). Analysis of translation products revealed no translation of the mSD template, but translation of the wtSD template was visible for all combinations of ribosomes. This indicates that the mut aSD ribosomes sent to us by the Cunningham lab are not able to actively transcribe mSD mRNAs. Additionally, translation and bypassing of wtSD gene *60* mRNA templates were similar for all combinations of mut aSD ribosomes, indicating the purified ribosomes obtained from the Cunningham lab contained significant contamination of wildtype ribosomes.

As these ribosomal elements were not behaving as expected, we chose to follow an alternative strategy that would allow us to label ribosomes with wildtype aSD elements. This change is also important for enabling study of translational bypassing as a short GAG element just upstream of the landing site in the coding gap of gene *60* was implicated as a potential SD-like element that could help the ribosome align the landing codon in the P-site of the ribosome[27]. Thus the presence of wildtype aSD elements in the ribosome may be crucial for obtaining efficient bypassing.

Conclusions

Based on the results of the experiments presented above, an optimized translation reaction should be carried out as follows. In a 10 μL reaction combine 4 μL of 2.5X S30 pre-mix (composition see Table II.1), 2 μL B2_ S30+ extract, 50 nM template mRNA, 5 nM control mRNA, 0.1 – 1 μL of 10 μCi/μL ^{35}S-Met or ^{35}S-Cys (~1 μCi per reaction) and water (if necessary). Assemble reactions on ice and incubate at 37°C for no longer than 5 min. Quench reactions with 1 μL of 1 M KOH and precipitate with 50 μL cold acetone.

Pellet the protein by centrifugation at room temperature, remove supernatant and allow pellet to air dry. Resuspend pellet in 15 μL of 1X loading buffer (100 mM Tris-HCl pH 6.8, 24% glycerol, 2% (w/v) SDS, 200 mM DTT, 0.02% Coomassie Blue) and resolve on a 20% Tris-tricine gel at 150 V for ~4 hours or until dye runs completely off of the gel. Soak gel for 1 hr in 5% (v/v) glycerol, 40% (v/v) methanol and 10% (v/v) acetic acid and dry onto 3 mm Whatman paper for 1.5 hrs at 75°C with a slow ramping program. Dried gels should be exposed to a phosphor storage screen overnight, or to obtain good background to noise in the linear range of the screen.

For time courses, mRNA and radioactive amino acids in a separate tube from extract and pre-mix should be pre-incubated at 37°C for 1 min before combining to initiate the reaction. Also, the loading control template can be omitted because its intensity will change over time.

Chapter III

Establishment of a High-throughput Chemical Probing Assay for the Determination of RNA Secondary Structure

Introduction

Non-coding RNAs (ncRNAs) participate in a wide range of cellular activities including regulation of gene expression[6,45,102], catalysis[103-106] and guiding large protein complexes to specific RNA targets[107,108]. The function of most annotated ncRNAs, however, is currently unknown. A major hurdle to understanding the roles of ncRNAs is the lack of fast, accurate methods for secondary and tertiary structure determination. Structural information is crucial for understanding how RNA molecules are able to carry out their complex cellular functions[45,46].

Traditionally, RNA secondary structures are determined by probing with small molecule modifiers and site specific cleavage by nucleases. The most common small molecule probes used are dimethyl sulfate (DMS), kethoxal and 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate (CMCT)[109,110]. DMS will react with the N1 of unstructured adenosines and the N3 of unstructured cytidines[109-111]. Kethoxal modifies the N1 or 2-NH_2 of unstructured guanosines and CMCT modifies N3 of unstructured uridines[110]. The most recent suite of small molecule modifiers, developed by Kevin Weeks, are used for what is called selective 2'-hydroxyl acylation analyzed by primer extension SHAPE[46,112-116]. SHAPE reagents are useful as they will acylate the 2' hydroxyl of any unstructured nucleotide, regardless of the identity of its base[46,112,113,117]. Hydroxyl radicals[110] or small metal ions such as lead (II) (Pb^{2+})[110] or terbium (III) (Tb^{3+})[118,119] also react independently of base identity by cleaving the phosphodiester backbone 3' of flexible nucleotides. Finally, there are several sequence and structurally specific nucleases that will cleave either ssRNA or dsRNA at particular nucleotides[110].

Detection of these structurally sensitive modifications can be accomplished via

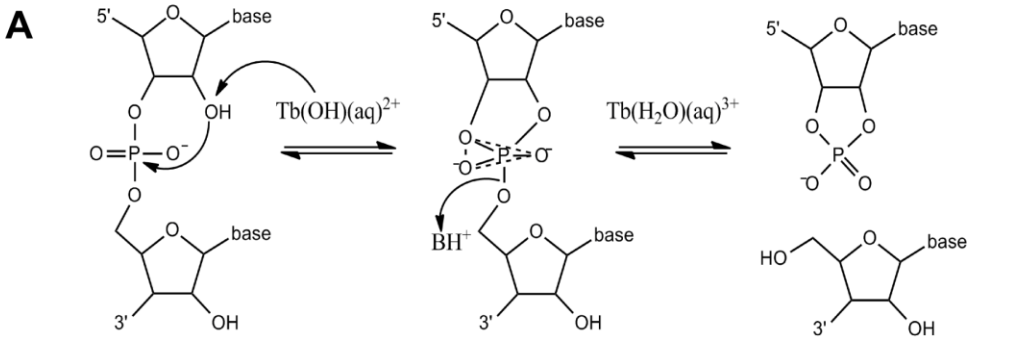
two techniques. For probes that cleave the phosphodiester backbone, fragments of an end-labeled RNA of interest can be resolved by denaturing polyacrylamide gel electrophoresis (PAGE). Intensity of bands correlates to the frequency at which each site was cleaved, giving a quantitative measure of the environment of each nucleotide[110,118,119]. Alternatively, the processivity of a reverse transcriptase will be inhibited by certain chemical modifications to an RNA[46,110,112,113,117]. Stops during primer extension with labeled primers generate cDNA fragments that can be resolved by denaturing PAGE[46,110,112,120] or by capillary electrophoresis (CE)[113-117,121-125]. Again the intensity of bands in a gel or the size of peaks in an electropherogram relate to the frequency of modification at a particular site and provide quantitative information about the structural environment of each nucleotide. Generally, a combination of several probes is used to generate overlapping maps and the sum of these experimental data can be used to generate secondary structure models of an RNA[110].

Probing the Structure of Gene *60* mRNA

The structure of gene *60* mRNA, especially in the coding gap region, is thought to be crucial for translational bypassing (see Chapter I)[19,22-27,29,30]. Mutational studies suggest that the hairpin in the 5' end of the coding gap is especially important for efficient and accurate bypassing[19,23,27,28]. Further hypotheses propose structural features within the rest of the coding gap[22,25-27]. Yet despite the apparent functional importance of the non-coding region of gene *60* mRNA, to date no direct structural work has been conducted. Thus we set out to probe the structure of gene *60* mRNA in an effort to directly confirm or refute proposed structural models.

In order to explore the structural characteristics of gene *60* mRNA, we needed to establish a protocol in the lab for structure probing. We chose to use two complementary structure-sensitive chemical probing techniques. First, hydrated terbium(III) ions were used to deprotonate the 2' hydroxyl of the ribose sugar. Nucleophilic attack of the 3' phosphate by the deprotonated oxygen generates a trigonal bipyramidal transition state that is resolved to form a 2',3'-cyclic phosphate and a 5' hydroxyl (Figure III.1 A). Nucleotides that are flexible, most likely single-stranded and solvent exposed, are more often in a favorable conformation for deprotonation by hydrated terbium ions. Thus, the

frequency of cleavage at a particular location gives a measure of the flexibility of the base 5' of the cleavage site[118,119]. SHAPE reagents were used as a second small molecule probe. These isatoic anhydride derivatives acylate the 2' hydroxyl of flexible nucleotides[46,112,113,117] (Figure III.1 B). The reactivity of both of these reagents is independent of base identity and both types of modifications are detectable by primer extension. SHAPE reagents have the added benefit that they hydrolyze at defined rates in aqueous solutions so no quench is needed[114,117] (Figure III.1 B and C). Tb^{3+} probing reactions must be quenched with EDTA[118,119].



C

SHAPE Reagent	Hydrolysis $t_{1/2}$ (s)	R_1	R_2
1M7	14	CH_3	NO_2
4NIA	101	H	NO_2
NMIA	260	CH_3	H
IA	430	H	H

Figure III.1 Mechanism of structure probing reagents. (**A**) Tb^{3+} cleavage, adapted from Harris et al. 2003[118]. Aqueous terbium ions deprotonate the 2' hydroxyl of the ribose of flexible nucleotides allowing the 2' oxygen to make a nucleophilic attack on the 3' phosphorous. The trigonal bypyrimidal intermediate is resolved by cleaving the phosphodiester backbone leaving a 5' hydroxyl and a 3',2'-cyclic phosphate[118]. (**B**) SHAPE reagent acylation, adapted from Gherghe et al. 2008[117]. SHAPE reagents acylate the 2' hydroxyl of flexible nucleotides, releasing CO_2. (**C**) This table, adapted from Gherghe et al.[117], lists modifications made to an isatoic anhydride scaffold to generate a family of SHAPE reagents with different time scales of self-quenching hydrolysis in aqueous environments.

Gene *60* mRNA is several hundred nucleotides long (~700) and therefore it is difficult to resolve probing data with single nucleotide resolution. Because of its size, terbium cleavage fragments of an end-labeled gene *60* mRNA were not able to be resolved by denaturing PAGE (data not shown). Primer extension reactions could be used to detect probing modifications, but ~7 labeled primers would be required to get accurate data for the entire length of the RNA if analyzed by denaturing PAGE[110]. Thus we chose to pursue a high-throughput capillary electrophoresis (CE) method for analysis of primer extension reactions. CE is fast, requires fluorophore-labeled DNA primers instead of radiolabeled ones and provides high quality data for 250-400 nucleotides at a time[125]. This reduces the number of primer binding sites required to analyze the full-length gene *60* mRNA from 7 to 2, which greatly simplifies data analysis. Furthermore, the DNA Sequencing Core at the University of Michigan (Sequencing Core) is able to perform the capillary electrophoresis in a 96-well plate format for a small fee so an in-house DNA sequencer is not required. Data analysis of electropherograms with ShapeFinder software[126] allows for quantitative analysis of reactivity at each nucleotide position, which can be incorporated into RNA folding algorithms to improve secondary structure models[127,128].

In order to establish a reliable structure probing protocol for our lab, each step was carefully optimized. This chapter describes the optimization process as well as the general method for data processing. Secondary structure models of gene *60* mRNA and their structural implications for translational bypassing will be discussed in Chapter IV.

General Method of Structure Probing and Resolving Primer Extension Reactions by Capillary Electrophoresis (CE)

RNA of interest was prepared by *in vitro* transcription with a T7 RNA polymerase and purified by denaturing PAGE (Appendix A). Structure probing[118,119,124] and reverse transcription were carried out on refolded RNA essentially as described[124] (Figure III.2). Briefly, each structure probing reaction required four pmols of RNA: one for probing with reagent, one to measure background (negative control) and one for each of two dideoxy sequencing reactions. A large pool of RNA was prepared containing all pmols of RNA required for the number of probing reactions to be performed, Z. There were 4

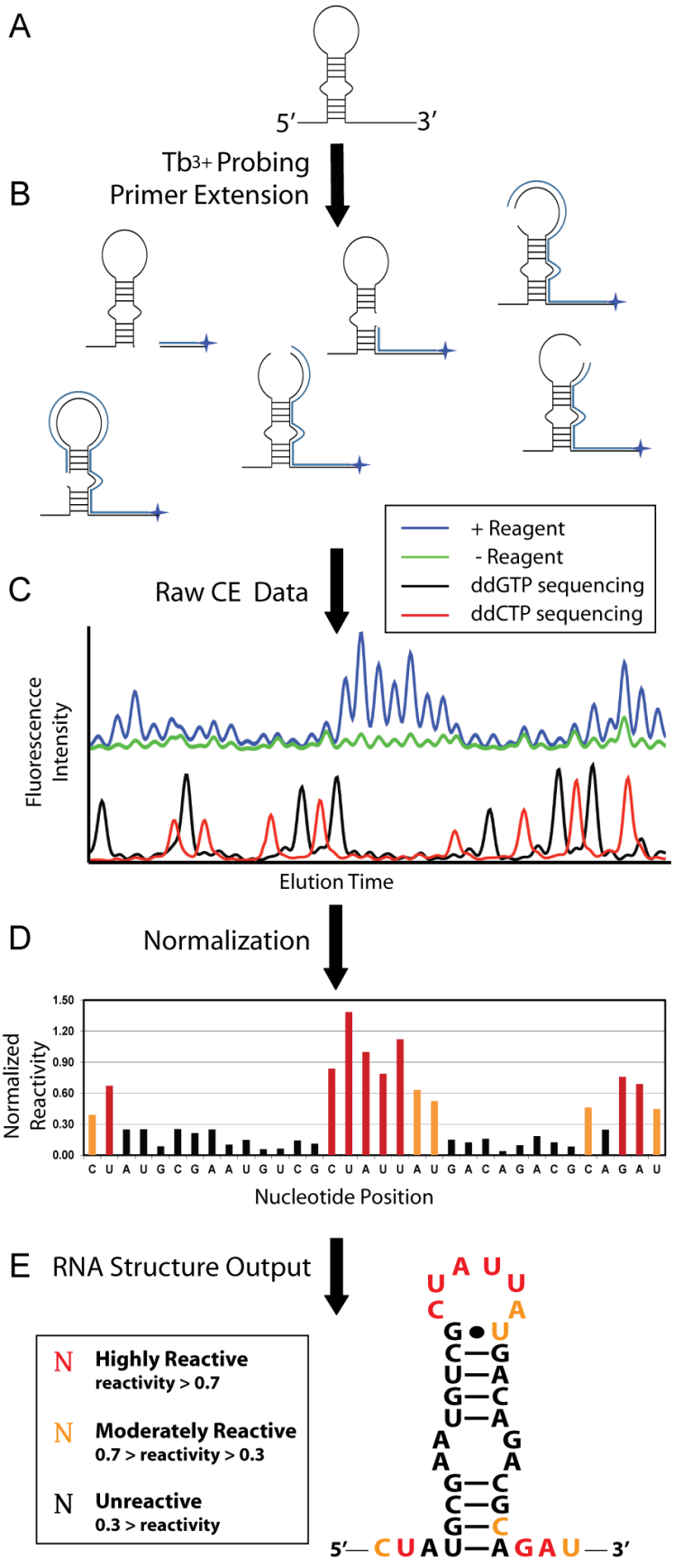

A
5′
3′
B
Tb3+ Probing
Primer Extension
C
Raw CE Data
+ Reagent
- Reagent
ddGTP sequencing
ddCTP sequencing
Fluorescencce Intensity
Elution Time
D
Normalization
Normalized Reactivity
1.50
1.20
0.90
0.60
0.30
0.00
C U A U G C G A A U G U C G C U A U U A U G A C A G A C G C A G A U
Nucleotide Position
E
RNA Structure Output
N Highly Reactive
reactivity > 0.7
N Moderately Reactive
0.7 > reactivity > 0.3
N Unreactive
0.3 > reactivity
5′— C U A U — A G A U — 3′

Figure III.2. Methodology for RNA Secondary Structure Determination. (**A**) RNA of interest was folded in the desired buffer. (**B**) Folded RNA was incubated with modifying reagents that are sensitive to the environment of each nucleotide (see Figure III.1). Modifications to or cleavage of the RNA generated stops during reverse transcription with fluorophore-labeled nucleotides (blue lines). (**C**) cDNA fragments were resolved by capillary electrophoresis (CE). In ShapeFinder software[126], the identity of the nucleotide corresponding to each peak was determined by aligning dideoxy sequencing reactions (red and black) to data traces. The difference in integrated intensity of peaks for RNA incubated with probing reagent (blue) and unprobed RNA (green) was normalized in Excel to obtain a "quantitative" relative measure of reactivity at each nucleotide position (**D**). Normalized reactivity was then used as a pseudo free energy constraint in the RNA folding software RNAstructure[128], generating structures that correlate well with experimental data (**E**).

pmols of RNA per 20 μL of TE (10 mM Tris-HCl pH 7.0, 1 mM EDTA) in the pool and Z * 4 pmols of RNA total. (Note: due to pipetting error, pools contained a few pmols more RNA than needed for the total number of reactions.) From the RNA pool, Z * 5 μL was twice removed and transferred to a separate tube. This RNA was set aside for the two dideoxy sequencing reactions. The remaining pool of RNA was denatured at 90°C for 2 min and quickly cooled on ice. Z * 2 pmols * 4 μL of a 2.5X buffer containing Mg^{2+} were added to the RNA pool and the pool was subsequently incubated at 37°C to allow the RNA to refold (Figure III.2 A). Z * 9 μL of the refolded pool was transferred to a separate tube for the negative control reaction. The rest of the pool was divided into Z separate tubes, 9 μL each, for probing reactions. 100 mM $TbCl_3$ stocks in 5 mM sodium cacodylate pH 5.5 (to prevent precipitation of $Tb(OH)_3$[119]) were diluted in water to 10 times the desired final concentration; similarly, 10X stocks of SHAPE reagents were prepared in anhydrous DMSO. 1 μL of a 10X stock of probing reagent was added to each probing reaction while Z * 1 μL of water or neat DMSO were added to the negative control reaction for Tb^{3+} or SHAPE reagent probing, respectively. Structure probing reactions and negative controls were incubated at 37°C for 30 min for Tb^{3+} probing or for four half-lives of the SHAPE reagent (Figure III.2 B). Reactions were quenched with 1 μL of 50 mM EDTA for each probing reaction or Z * 1 μL of 50 mM EDTA for negative control pools. All RNA pools were precipitated by adding 1 μL or Z * 1 μL 20 mg/mL glycogen (co-precipitant), water to bring the volume up to 100 μL, 10 μL of 3 M NH_4OAc pH 5.2 and 300 μL ethanol. Tubes were incubated at -20°C overnight or at -80°C for ~20 min. RNA was pelleted by centrifugation at 20,000 relative centrifugal force (rcf), 4°C for 30 min, washed with 200 μL of cold 70% ethanol and centrifuged again at 20,000 rcf, 4°C for 30 min. The majority of the supernatant was

removed by decanting and blotting on a paper towel; the rest was removed by vacuum for ~15 min. (Note: it was important to briefly centrifuge samples at ~2,200 rcf before drying by vacuum to ensure the pellet rested at the bottom of the tube.)

Pellets reserved for dideoxy sequencing reactions were resuspended in Z * 9 μL of 0.5X TE pH 8.0, while probing reactions and negative controls were resuspened in 10 μL or Z * 10 μL of 0.5X TE pH 8.0, respectively. 3 μL or Z * 3 μL of 400 nM fluorescently labeled DNA primer was annealed to the RNA by incubation at 65°C for 5 min, 45°C for 2 min and placing the tubes on ice. At this point, Z * 1 μL of the desired dideoxy nucleotide triphosphate (ddNTP) was added to each dideoxy sequencing reaction. A reverse transcription master mix containing Z * 4 μL of 5X SuperScript III First Strand Buffer (Invitrogen 18080-085), Z * 1 μL of 100 mM DTT, Z * 1 μL of 10 mM each dNTPs and Z * 1 μL of SuperScript III Reverse Transcriptase (Invitrogen 18080-085) was prepared on ice for each set of probed reactions or for each pool (negative control or sequencing reactions). The master mix and pre-annealed primer/RNA were incubated at 45°C for 1 minute before the addition of 7 μL or Z * 7 μL of master mix to each reaction. Tubes were quickly transferred to 52°C for 5 min of reverse transcription, 65°C for 5 min to inactivate the enzyme and then placed on ice (Figure III.2 B). 20 μL each of probing reaction, negative control and the two dideoxy sequencing reactions are combined with 40 μL water, 10 μL of 3 M NH_4OAc pH 5.2 and 300 μL ethanol. Reactions were precipitated as described above and resuspended in 10 μL HiDi formamide (Applied Biosystems, 4311320). They were then loaded into 96-well optical plates (Applied Biosystems, 4306737) and submitted to the University of Michigan DNA Sequencing Core for analysis by CE with an Applied Biosystems 3730XL DNA Analyzer. Raw electropherograms were visualized and analyzed with the ShapeFinder software[126] (Figure III.2 C). Integrated peak intensities obtained from the ShapeFinder software were normalized in Excel (Figure III.2 D) and these values were used as pseudo free-energy constraints during folding in RNAstructure[128] (Figure III.2 E). The process of data analysis is described in further detail in section "Generating Models from CE Data with ShapeFinder and RNAstructure Software" below.

Primer Design for Analysis of Reverse Transcription by CE

Structure probing reactions must be carried out under single-hit conditions such that each RNA molecule is modified, on average, far less than once[110,112,119]. This restriction is necessary because the first alteration of an RNA molecule could lead to local and/or global structural changes and further modifications to the same RNA might reflect this altered structure instead of the native state. With a majority of RNAs unmodified after the probing reaction, the most abundant product of reverse transcription is full-length product (Figure III.5, rightmost peak). Because the full-length product is so intense, this peak overlaps with several nucleotides near the end of the trace obscuring information on the environment at the very 5' end of the RNA. Under our assay conditions there is also a considerable amount of unextended primer present, which generates an intense peak that overlaps with data for nucleotides at the 3' end of the RNA[112] (Figure III.5, leftmost peak). In order to obtain data for every nucleotide in gene *60* mRNA, extensions were added to its 5' and 3' ends. The 5' extension consisted of 61 nucleotides of gene *60*'s 5' UTR. The 3' extension included 41 nucleotides of gene *60*'s 3' UTR and another 51 nucleotides of the vector backbone of pUC19 (generated by cleavage with BsrBI, Appendix A). A 28-nucleotide DNA primer complementary to the very 3' end of the transcript was ordered with flourophore labels attached. As CE data only provides quality data for roughly 250-400 nucleotides [125], an additional primer was designed to bind 88 nucleotides 3' of the coding gap of gene *60*.

Although 96-capillaries can be resolved simultaneously, each has a slightly different migration time for identical cDNAs[126,129]. Thus traces between capillaries are not directly comparable. To circumvent this issue, the Weeks' lab developed a system in which the same DNA primer is 5'-labeled with one of four different fluorophores. One color is used for primer extension of probing reactions; a second is used for negative control reactions and the third and fourth are used for two different dideoxy sequencing reactions. All four reactions are combined and run in the same capillary, therefore identical cDNA fragments elute at the same time and traces from each "color" reaction can be compared to one another directly[126]. DNA primers for the reverse transcription of gene *60* mRNA were ordered with 5' 6FAM, HEX, TAMRA and ROX dyes (Figure III.3, Table III.1).

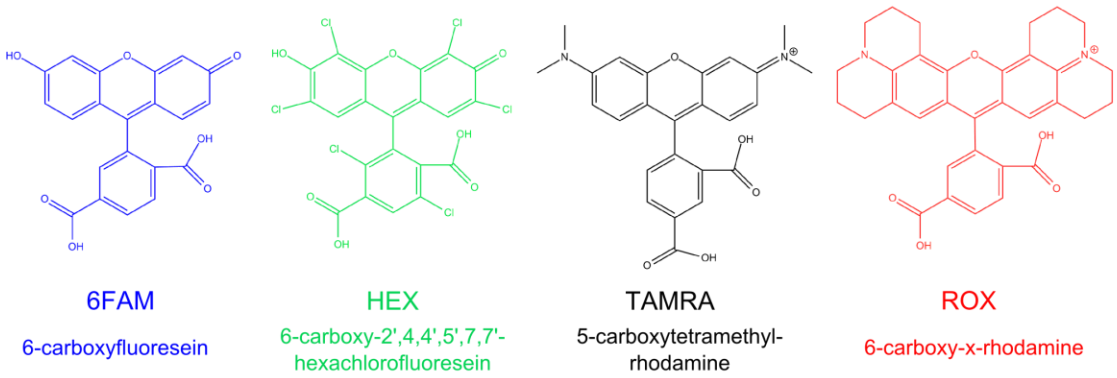


Figure III.3 Structure of fluorescent dyes used to label DNA primers for reverse transcription.

Fluorophore	Absorption λ_{max} (nm)	Emission λ_{max} (nm)	Vendor
6FAM	494	520	Invitrogen[130]
HEX	535	553	Invitrogen[130]
TAMRA	565	580	Invitrogen[130]
NED*	553	575	Applied Biosystems[131]
ROX	576	601	Invitrogen[130]

Table III.1 Absorbance and emission maxima of fluorescent dyes in Figure III.3. These dyes were used to label DNA primers for reverse transcription. *The structure of NED is proprietary information of Applied Biosystems and is therefore not present in Figure III.3.

The emission spectra of these four dyes are overlapping and therefore a matrixing algorithm must be applied to separate intensity contributions from each fluorophore[126]. The University of Michigan Sequencing Core automatically performs a matrixing step before returning the raw data. To check that their matrixing algorithm was functioning properly, dideoxy sequencing reactions with each dye were resolved in separate capillaries. If matrixing has been done correctly, each capillary should only have peaks corresponding to a single dye[126]. This was not the case, however, for our set of dyes (6FAM, HEX, TAMRA and ROX). In Figure III.4 A, TAMRA peaks (black) generated positive peaks in the 6FAM (blue) and ROX (red) channels and negative peaks in the HEX (green) channel. Although the ShapeFinder software contains a manual matrixing tool[126], it was unable to repair the matrixing applied previously by the Sequencing Core. The matrixing algorithm used by the Sequencing Core was generated with Applied

Biosystems Matrix Standard Set DS-30, which contains 6FAM, HEX, *NED* and ROX. NED and TAMRA have slightly different emission spectra and therefore the algorithm generated with NED could not properly distinguish contributions from TAMRA. NED-labeled DNA primers were ordered, and its signal was deconvoluted properly by the Sequencing Core matrixing algorithm for this set of dyes (Figure III.4 B).

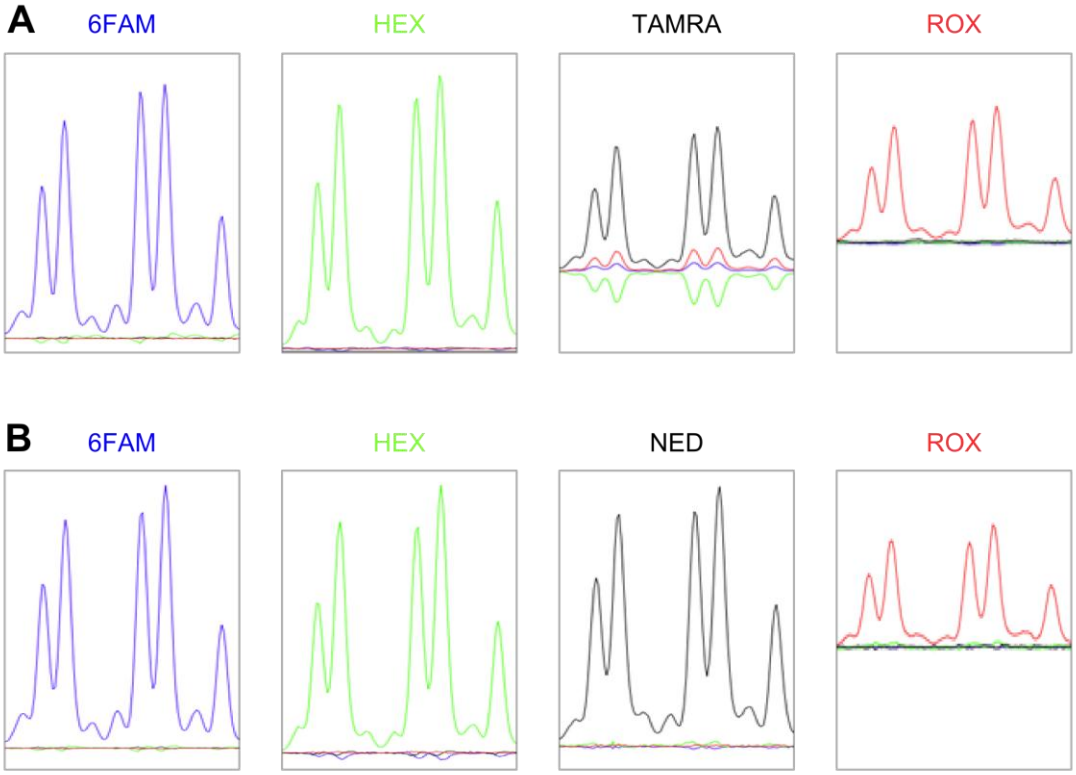


Figure III.4 Obtaining properly matrixed electropherograms. A dideoxy sequencing reaction was extended with DNA primers attached to each color fluorophore and reactions were resolved in separate capillaries. The dye set used in (**A**) were improperly matrixed relative to one another, as can be seen by positive and negative peaks in other channels for the TAMRA dye. For the dye set in (**B**), each capillary only exhibited signal from a single fluorophore, indicating correct matrixing.

Optimization of Primer, Reverse Transcriptase and Dideoxynucleotide Triphosphate Concentrations

To further improve the quality of our CE data, primers, enzyme and dideoxynucleotide triphosphates (ddNTPs) were titrated to find optimal concentrations. In the original protocol used for reverse transcription reactions, a 20% excess of primer was used to ensure that all RNA templates could be extended[124]. However, based on the

high intensity of our primer peak, a significant portion of primers are not being extended (Figure III.5, D leftmost peak). Furthermore, small fluorescently labeled oligonucleotides are easily able to contaminate the capillaries of the DNA sequencer (Bob Lyons, director of the University of Michigan DNA Sequencing Core, personal communication). We hypothesized that a reduction in primer concentration might reduce the intensity of the primer peak, decreasing the likelihood of contaminating the sequencer and allowing us to get more structural information at the 3' end of gene *60* mRNA. Lower ratios of primer to RNA did indeed reduce the primer peak, but also led to significant reduction in our data peaks (Figure III.5, A – C). Thus all further reactions were carried out with a 20% excess of primer (Figure III.5 D) and the Sequencing Core added an extra wash step after running our plates on their sequencer to ensure all small oligonucleotides were removed.

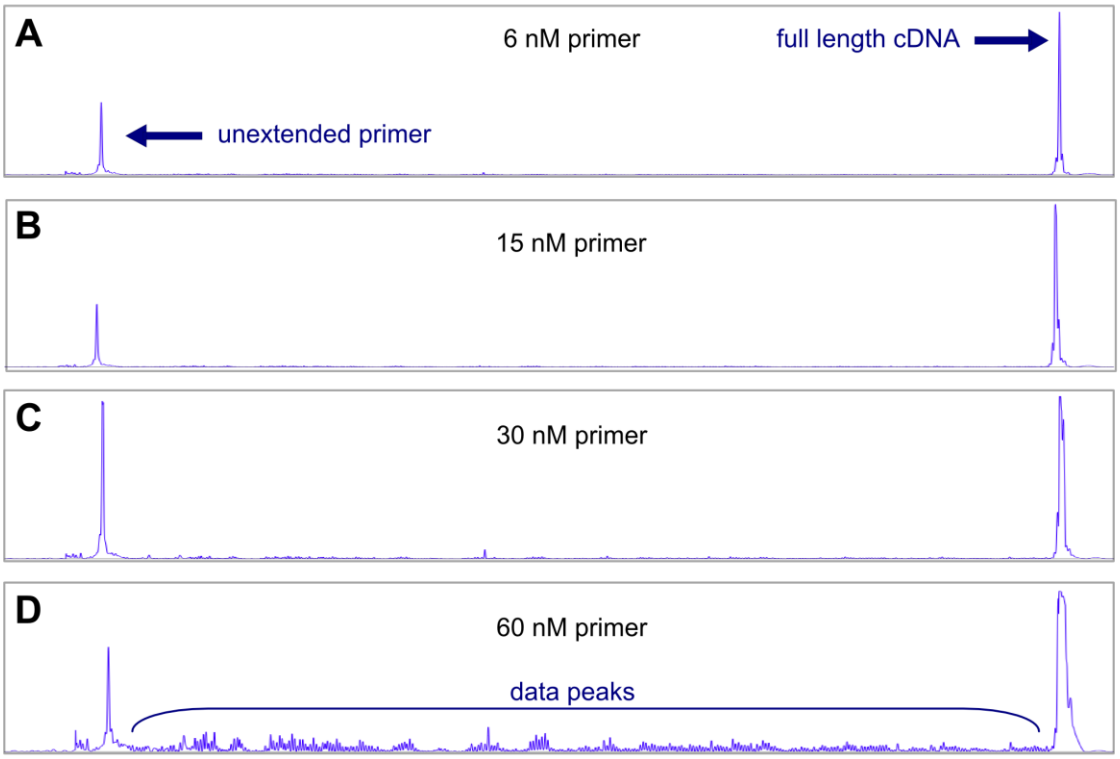


Figure III.5: Optimizing the concentration of primers for reverse transcription reactions. Raw traces of primer extension reactions with unextended primer, full-length cDNA and data peaks indicated. The concentration of fluorophore-labled DNA primers was varied from 6 – 60 nM (**A – D**), with template RNA held constant at 50 nM per reaction. Although lower primer concentrations significantly reduced the intensity of the unextended primer peak, the signal for data peaks was also significantly reduced. Thus 60 nM primer was used for all further reactions.

Another factor that we wanted to vary was the concentration of reverse transcriptase required for each extension reaction. Under the published protocol, 200 U of SuperScript III reverse transcriptase (SSIII RT) was used for each extension reaction[124]. At $5 per 200 U of SSIII RT and with four reactions per capillary, a single experiment consumes $20 worth of enzyme. We hypothesized that perhaps lower concentrations of enzyme would produce similar results, thus reducing the cost of each experiment. Alternatively, adding more enzyme to each reaction might result in more cDNA copies of probed RNA, which would increase the signal to noise ratio for data peaks. To test these hypotheses, enzyme was varied from 50 – 400 U per reaction (Figure III.6). Contrary to what would be expected by the addition of twice as much enzyme (Figure III.6 D), signal for data peaks actually decreased in intensity suggesting inhibition in the presence of too much SSIII RT. Reduction in the amount of enzyme did not produce a consistent trend for peak intensities (Figure III.6 A – C). For very intense peaks, reduction in enzyme

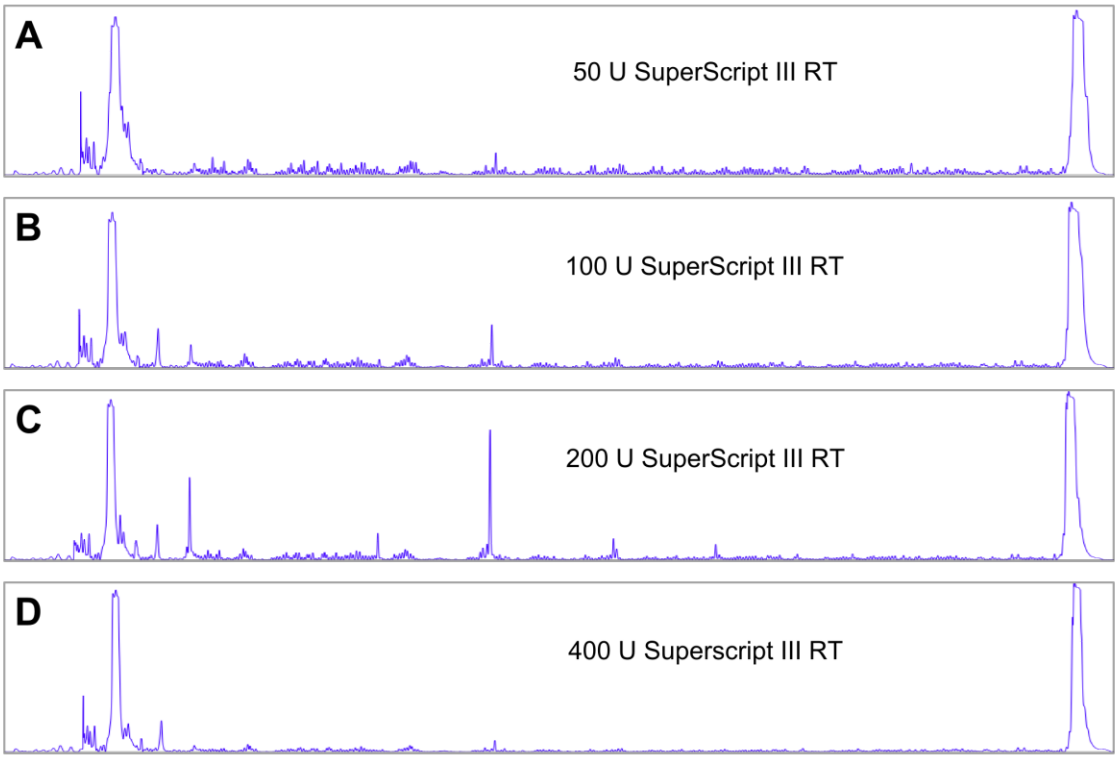


Figure III.6 Optimization of SuperScript III Reverse Transcriptase (SSIII RT). Lowering enzyme concentration (**C** to **A**) did not affect peak intensities proportionally. (**D**) High amounts of SSIII RT decreased signal to noise ratios of data peaks. Therefore 200 U of SSIII RT were used consistently for all further reactions.

reduced peak height. The inverse effect, however, was observed for low intensity peaks. As this trend was inconsistent, we could not perform reactions at lower enzyme concentrations. All further experiments were carried out with 200 U SSIII RT.

Finally, the concentration of ddNTPs required to produce a strong sequencing ladder over the trace range was optimized. If too many ddNTPs are added, the polymerase will not be able to sequence the whole length of the RNA. Inversely, if too few ddNTPs are added then the intensity of sequencing peaks will be too low to accurately assign data peaks. ddNTPs were varied from 250 – 1000 µM (final concentration). Figure III.7 shows that 500 µM of ddNTPs provided the most intense sequencing ladder for our ~700 nucleotide RNA. Thus ddNTPs were used at a concentration of 500 µM for all further experiments.

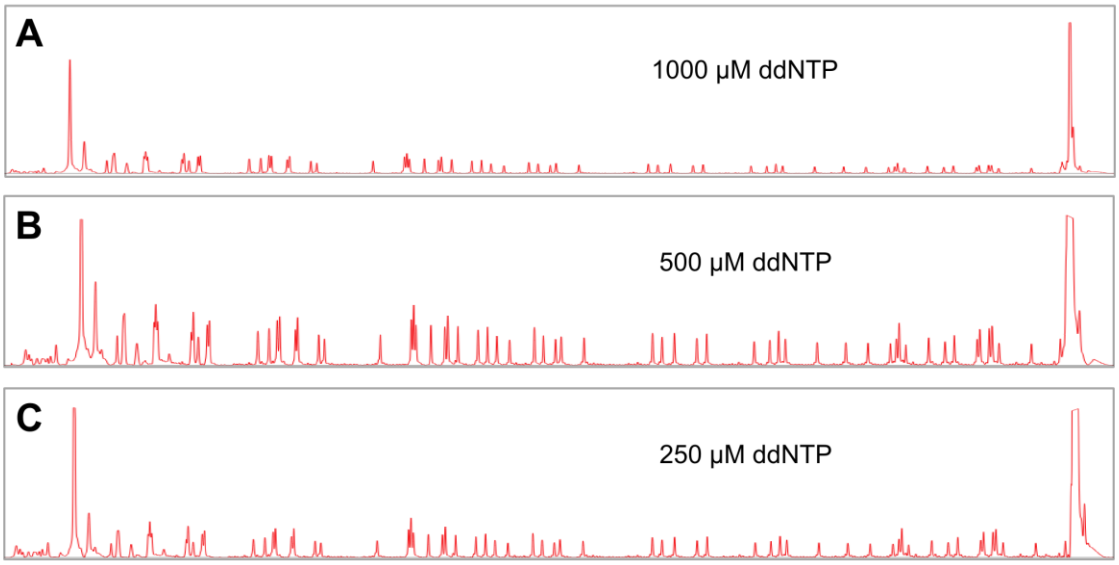


Figure III.7 Optimization of ddNTP concentration. To obtain a sequencing ladder of gene *60* mRNA with good signal to noise, the concentration of ddNTPs was varied from 250 – 1000 µM (final concentration) in the reverse transcription reaction (**C** – **A**). 500 µM ddNTPs (**B**) produced the highest signal to noise peaks and ddNTPs were used at this concentration for all further ladders.

Persistent Band Compression in the Coding Gap Region

CE data for all gene *60* mutants contained significant band compression in the region overlapping the proposed hairpin at the 5' edge of the coding gap (Figure III.9, blue stars). The compressed region consisted of a broad peak with a shoulder where there should have been three distinct peaks corresponding to 5'-GCC-3'. Band compression is

a common problem in PCR and sequencing reactions and is thought to be due to incomplete denaturation of stable structures in the cDNA[132,133] (Figure III.8). When passing through the capillary, these structured cDNAs do not migrate linearly with their proper nucleotide length and therefore cDNA fragments ending in each of the compressed nucleotides elute at overlapping times. To determine whether or not there was a mistake in gene *60* mRNA sequence, the DNA template was resequenced and found to be unambiguously correct (data not shown). Additionally, gene *60* mRNA was sequenced with all four ddNTPs simultaneously to confirm the mRNA sequence was correct and the compressed region was not simply a mistake of transcription (data not shown). Sequencing of the mRNA also confirmed the location and identity of the compressed nucleotides. All possible pairs of ddNTPs were used for dideoxy sequencing of gene *60* mRNA (Figure III.9 B (G & T) and C (A & C), Figure III.10 F (G & C), Figure III.11 B (A & T)) as well as a hydroxide ladder (Figure III.9 D), however individual peak maxima in the compressed region were not better defined with any combination of ddNTPs. As the compression occurs in a critical region of gene *60*, in the coding gap and overlapping the predicted site for the 5' hairpin, we sought ways to improve data resolution in this region.

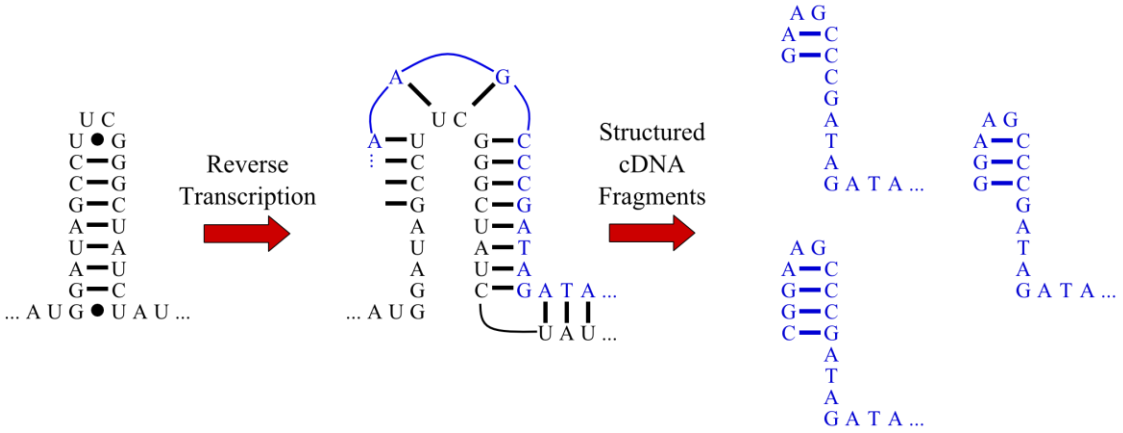


Figure III.8 Stable structures in cDNA induce peak compression. cDNA copies of tight-turning RNA hairpins (see Figure III.11 **C**) are also capable of forming structure. If stable enough, these structures may not get fully denatured during CE and peaks corresponding to the structured nucleotides migrate too closely to one another for clear resolution.

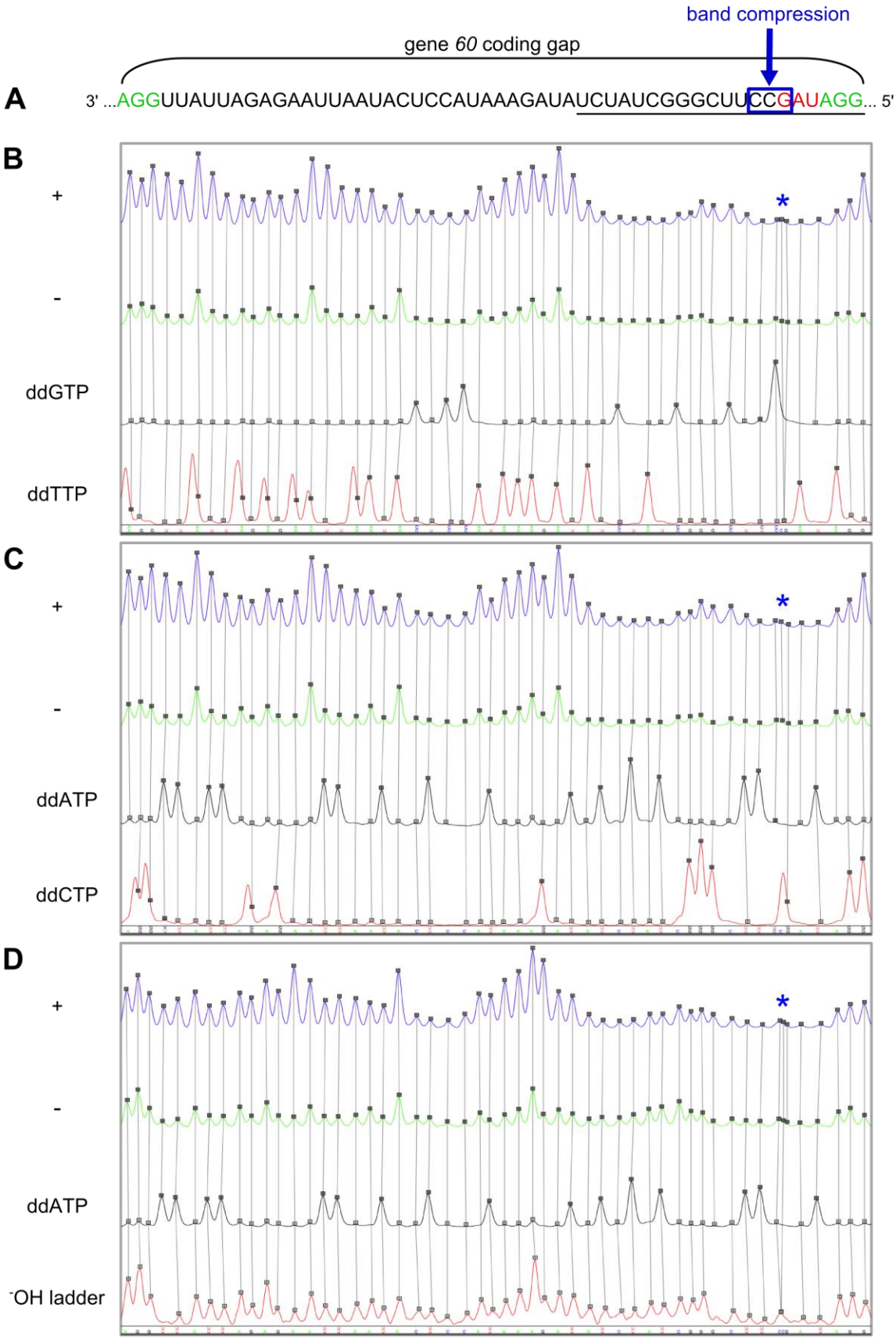

band compression
gene 60 coding gap
A
3' ...AGGUUAUUAGAGAAUUAAUACUCCAUAAAGAUAUCUAUCGGGCUUCCGAUAGG... 5'
B
+
-
ddGTP
ddTTP
C
+
-
ddATP
ddCTP
D
+
-
ddATP
-OH ladder

Figure III.9 Persistent peak compression in the coding gap. (**A**) Sequence of gene *60* mRNA coding gap corresponding to peaks in traces below. Note: trace data runs 3' to 5' from left to right. Take-off and landing codons are in green and the stop codon is in red. Underlined nucleotides are predicted to form a hairpin at the 5' edge of the coding gap. Compressed nucleotides are boxed in blue. Compressed peaks (blue star) are unable to be resolved by sequencing with any of the four ddNTPs (**B** and **C**) or with a hydroxide ladder (**D**). For each dataset, blue traces have been probed with Tb^{3+} (+), green traces are negative controls (-) and black and red traces are the indicated sequencing reaction (or hydroxide ladder). Gray boxes mark the maxima of each peak and gray lines between boxes align peaks from different traces that correspond to the same nucleotide. Figures **B** – **D** as data is presented in ShapeFinder software[126].

2'-deoxyinosine 5'-triphosphate (dITP, Fermentas R1191) is frequently substituted for dGTP to alleviate band compression[46,132]. dITP lacks the amine group of dGTP at the C2 position, thus reducing its base-pairing ability with cytosine by one hydrogen bond. (Figure III.10 A). In our reverse transcription experiments, replacement of dGTP with dITP resolved the compressed region into individual peaks (Figure III.10 E), but generated very intense stops at three consecutive guanosines (red bar) just downstream of the initial compressed site (Figure III.10 D and E). Strong stops in G-rich regions are a known disadvantage of dITP[132]. Because the GGG peaks are 35-fold more intense than peaks in the rest of the trace, accurate comparisons of reactivity of each nucleotide could not be made, and dITP was not used further.

Similarly, dGTP can be substituted with 7-deaza-2'-deoxy-guanosine-5'-triphosphate (7-deaza-dGTP or c7dGTP, Roche 10 988 537 001) to reduce band compression[132,133]. In c7dGTP, the N7 of guanosine has been replaced with a carbon (Figure III.10 B). Substitution of c7dGTP for dGTP in reverse transcription reactions did not significantly improve resolution of the compressed area (blue star, Figure III.10 F) and actually generated a second area of compression (purple star). Thus c7dGTP was also not a good alternative to dGTP.

Based on these results, the band compression appeared to be a permanent feature of cDNA with this particular primary sequence. To work around this issue, we noted that the compressed nucleotides are flanked by a 5' A and a 3' U. Dideoxy sequencing with ddATP and ddTTP allowed us to accurately define peaks on either side of the broad compressed band. Between these boundaries, three "peak maxima" were selected on the compressed peak. These "peaks" represent signal intensity for the nucleotides 5'-GCC-3'. We consider this assumption acceptable for several reasons. First, the intensity of the compressed band in the probed and unprobed traces (Figure III.11 B, blue and green

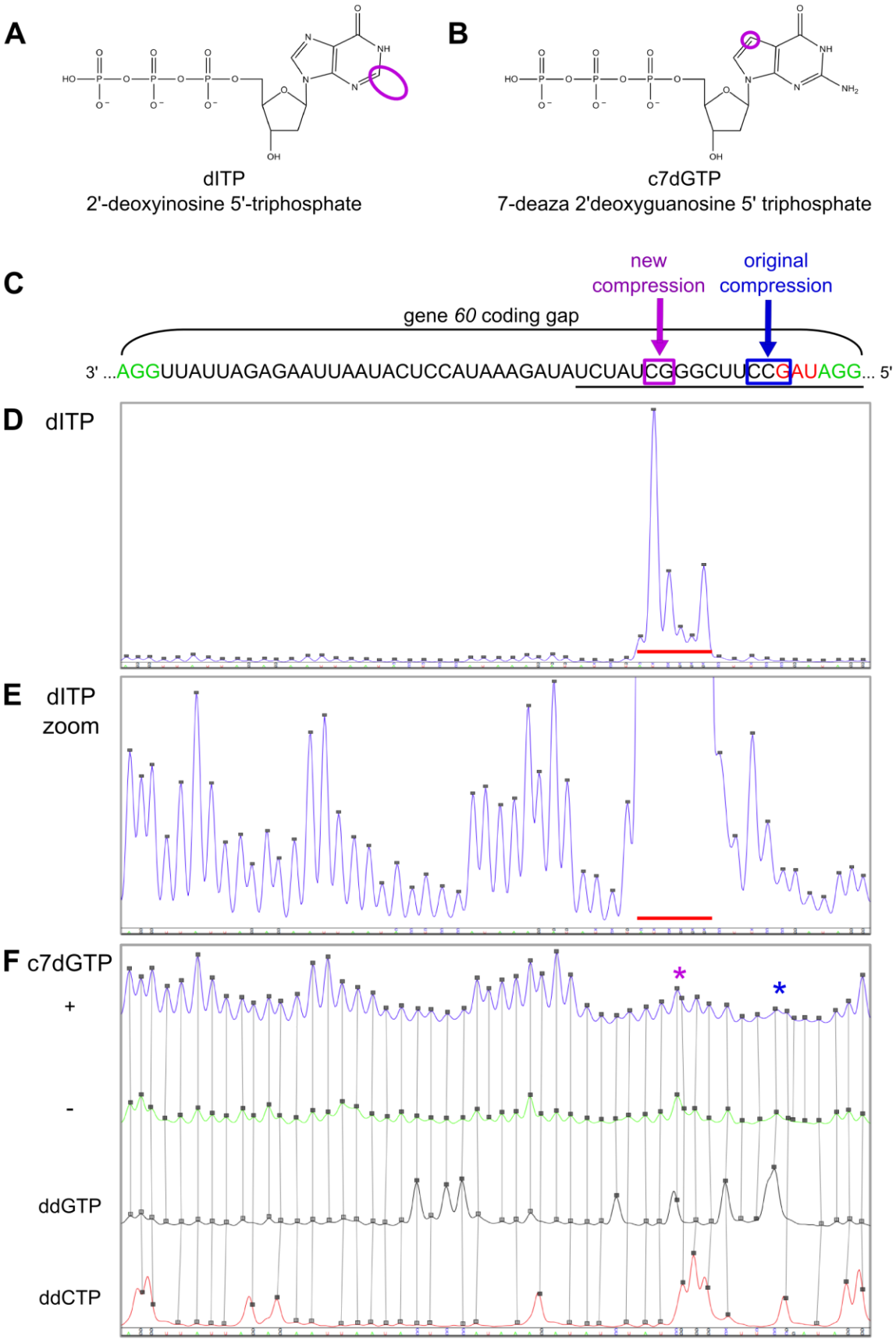

A
dITP
2'-deoxyinosine 5'-triphosphate
B
c7dGTP
7-deaza 2'deoxyguanosine 5' triphosphate
C
new compression
original compression
gene 60 coding gap
3' ...AGGUUAUUAGAGAAUUAAUACUCCAUAAAGAUAUCUAUCGGGCUUCCGAUAGG... 5'
D dITP
E dITP zoom
F c7dGTP
+
-
ddGTP
ddCTP

Figure III.10 dGTP alternatives that reduce band compression. Structure of dITP (**A**) and c7dGTP (**B**); chemical differences between these molecules and dGTP are circled in purple. (**C**) As described in Figure III.9 with newly compressed bands boxed in purple. Subsitution of dGTP with dITP relieved band compression at original location (**E**) but generated extremely large stops just 3' of the original compression (red band, **D**). Substitution of dGTP with c7dGTP (**F**) did not resolve the original region of band compression (blue star) and added another (purple star).

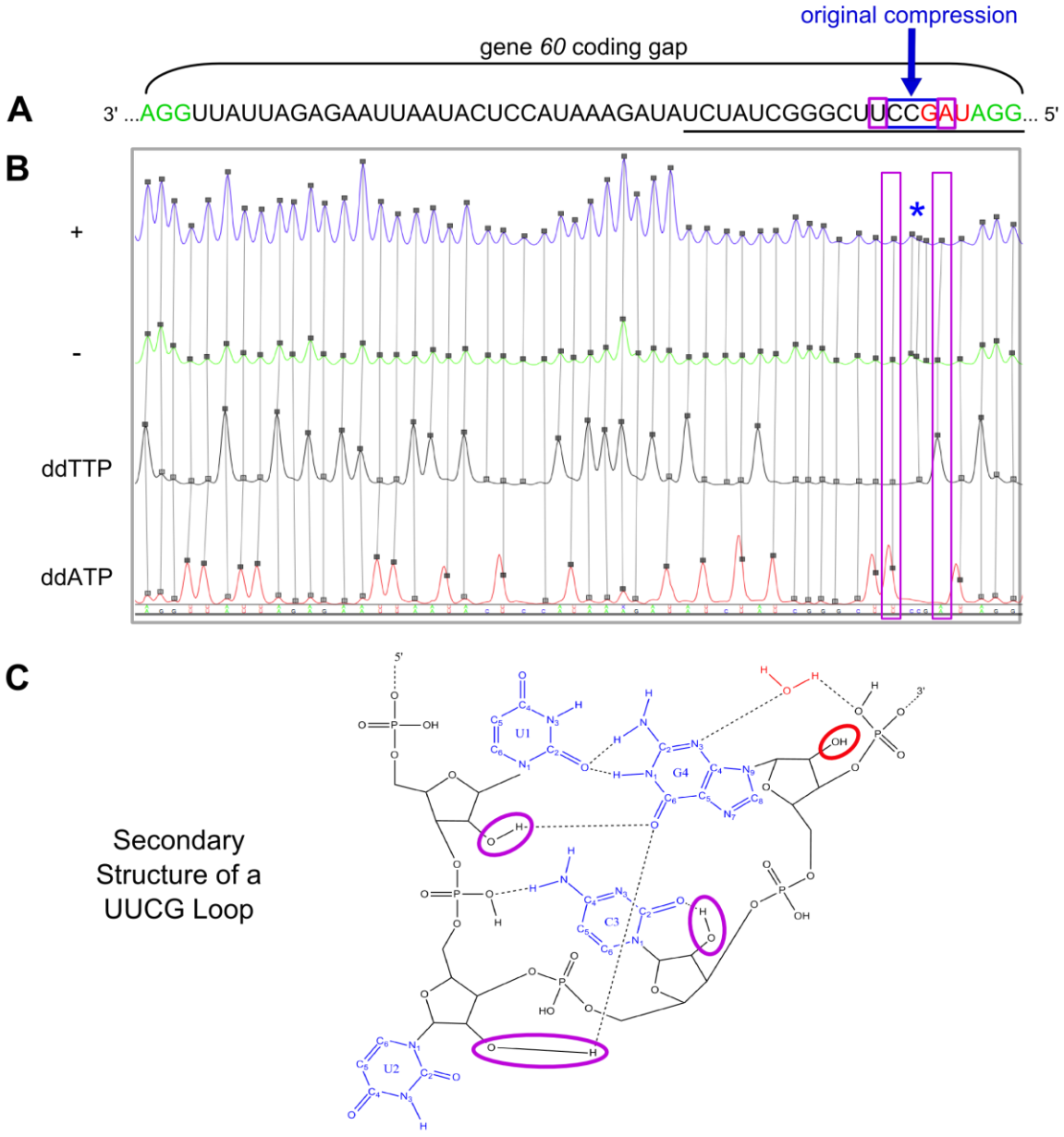


Figure III.11 Defining the region of band compression with ddATP and ddTTP sequencing reactions. (**A**) As described in Figure III.9. Nucleotides boxed in purple flank the original region of band compression and can be accurately assigned with ddATP and ddTTP sequencing ladders (**B**). Three peaks were assigned between the flanking A and T peaks. (**C**) Secondary structure of a UUCG loop[134]. 2' hydroxyls of a UUCG loop (circled in purple) are involved in hydrogen bonding interactions, except for G4 (circled in red), which prevent them from interacting strongly with structure probing reagents.

traces, respectively) is almost identical. This means the compressed nucleotides are highly protected from probing reagents and are very likely structured. Therefore, all "peaks" assigned in this region will have low reactivity regardless of the exact location of their peak maxima. Secondly, the entire sequence of the proposed hairpin at the 5' edge of the coding gap (which includes the compressed nucleotides) is highly protected; blue and green traces for underlined nucleotides in Figure III.11 B have very similar intensities. Strong protection in this region confirms the presence of this hairpin in gene *60* mRNA. Additional support for this hairpin can be drawn from the presence of the compression itself. cDNA copies of hairpins with G-C rich stems and 3-4 nucleotide long loops commonly form short, but strong complementary structures that are not fully denatured during CE[133] (Figure III.8). Finally, there is weak probing of the nucleotides in the UUCG tetraloop of this proposed hairpin. Loop regions are generally more reactive to probing reagents than stem regions, but the UUCG tetraloop is highly structured. A secondary structure representation of the hydrogen bonding interactions from the crystal structure of a UUCG loop reveals that the 2' hydroxyls of three of the four nucleotides are involved in intramolecular hydrogen bonding interactions[134] (Figure III.11 C, purple circles) and we would therefore not expect them to be strongly hit by probing reagents. Based on the sum of this experimental evidence, we can confidently conclude the proposed hairpin in the 5' end of the coding gap does, indeed, exist.

Optimizing Structure Probing Conditions

The structure of RNA can vary greatly depending on the concentration of surrounding solutes and the pH[109,110,114,118,119]. Thus, we sought to perform our structure probing of gene *60* mRNA in conditions similar to that which it experiences during translation. Initially, refolding was performed in S30 pre-mix solution[75] (Table II.1), but this resulted in severe degradation of the RNA due to contaminating RNases in the creatine kinase (Figure II.7 and II.8). While tests were being done to determine the source of this degradation, mRNA was refolded in a simple buffer 100 mM Tris-HCl pH 7.0, 100 mM NaCl, 10 mM $MgCl_2$. All of the initial optimization of CE data (above) was performed on RNA folded in this simple buffer. For optimization of structure probing

conditions and data used to build models, however, RNA was refolded in S30 buffer (top section of Table II.1) to simulate buffer conditions experienced during translation.

A titration of Tb^{3+} was performed on gene *60* mRNA (Figure III.12) to determine the optimal concentration for structure probing. Tb^{3+} is known to cleave the phosphodiester backbone of RNA at μM to mM concentrations[118,119]. To confirm these results, gene *60* mRNA was incubated with 0 – 5 mM Tb^{3+} and primer extension reactions were analyzed by CE. Comparison of each Tb^{3+} concentration to unprobed RNA (blue) mRNA reveals that 1 mM of Tb^{3+} provides the best signal without excessive cleavage, and was therefore used to generate all subsequent models.

Similarly, we varied the concentration of the SHAPE reagents 1M7 and IA between 0 and 10 mM (Figure III.13 and Figure III.14, respectively). Comparison to unprobed RNA (blue) revealed that 2 mM of 1M7 provides the best signal. This concentration was used for further experiments. 10 mM of IA, however, was not sufficient to generate enough signal over background and therefore 25 mM IA was used for probing reactions. Benzoyl cyanide (BzCN), another SHAPE reagent tested, has a

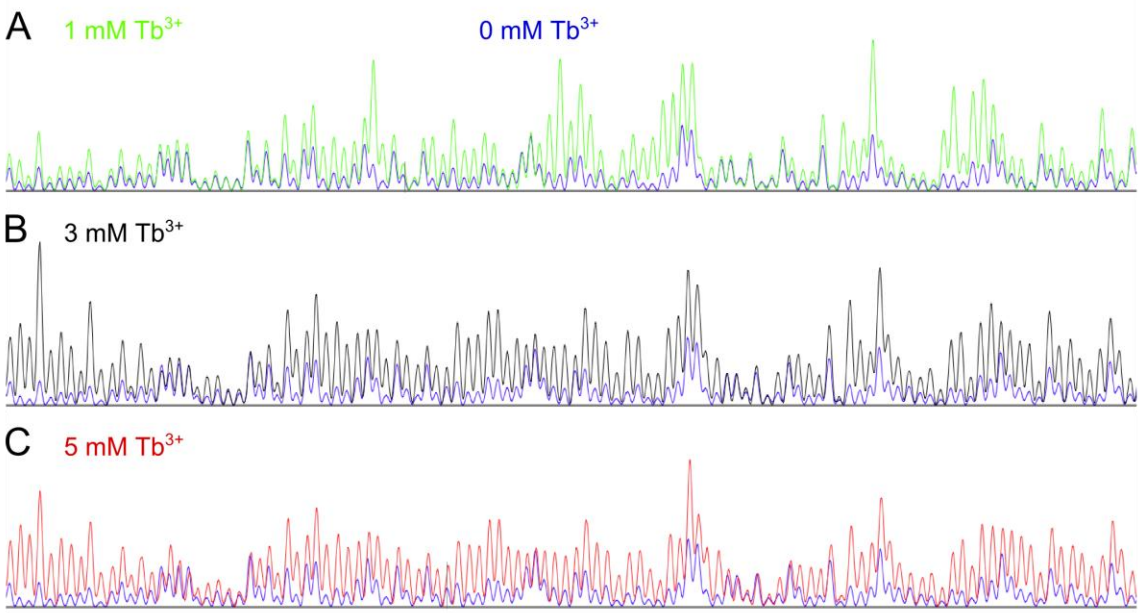


Figure III.12 Optimization of Tb^{3+} concentration for structure probing. Gene *60* mRNA was incubated with 0 (blue), 1 (green), 3 (black) or 5 (red) mM Tb^{3+} for 30 min at 37°C. Comparing each trace to that for the unprobed mRNA (blue) revealed that > 1 mM Tb^{3+} caused excessive cleavage. 1 mM Tb^{3+} was used for all further probing reactions. Note: lower concentrations of Tb^{3+} did not significantly modify the RNA (data not shown).

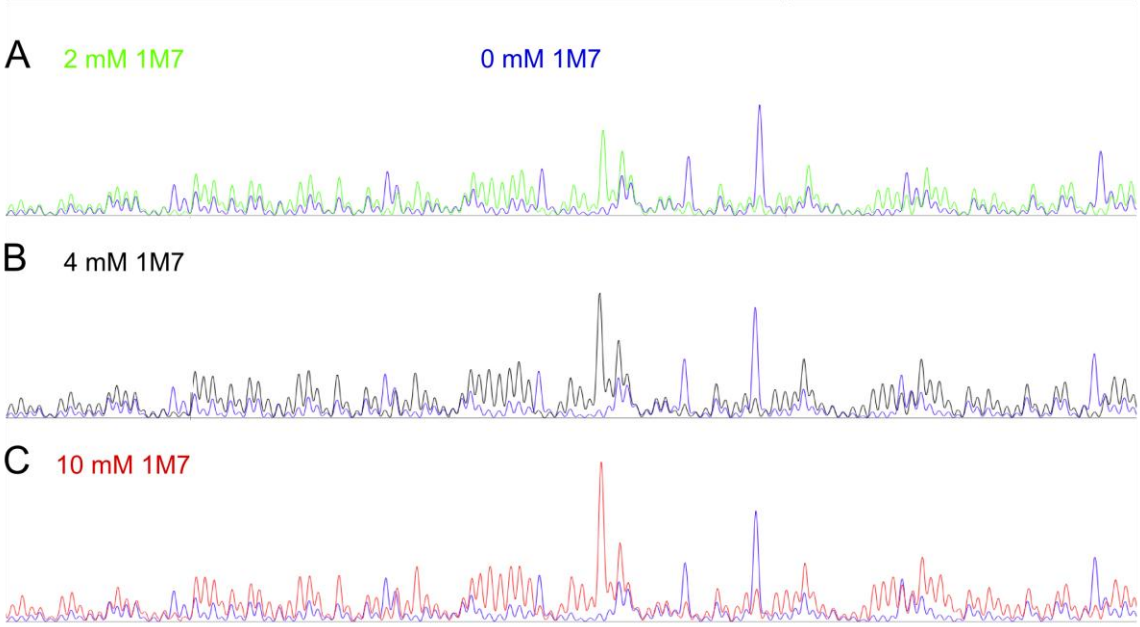


Figure III.13 Optimization of 1M7 concentration for structure probing. Gene *60* mRNA was incubated with 0 (blue), 2 (green), 4 (black) or 10 (red) mM 1M7 for 4 min at 37°C. Comparing each trace to that for the unprobed mRNA (blue) revealed that > 2 mM 1M7 exhibited excessive modification of the RNA. 2 mM 1M7 was used for all further probing reactions. Note: lower concentrations of 1M7 did not significantly modify the RNA (data not shown).

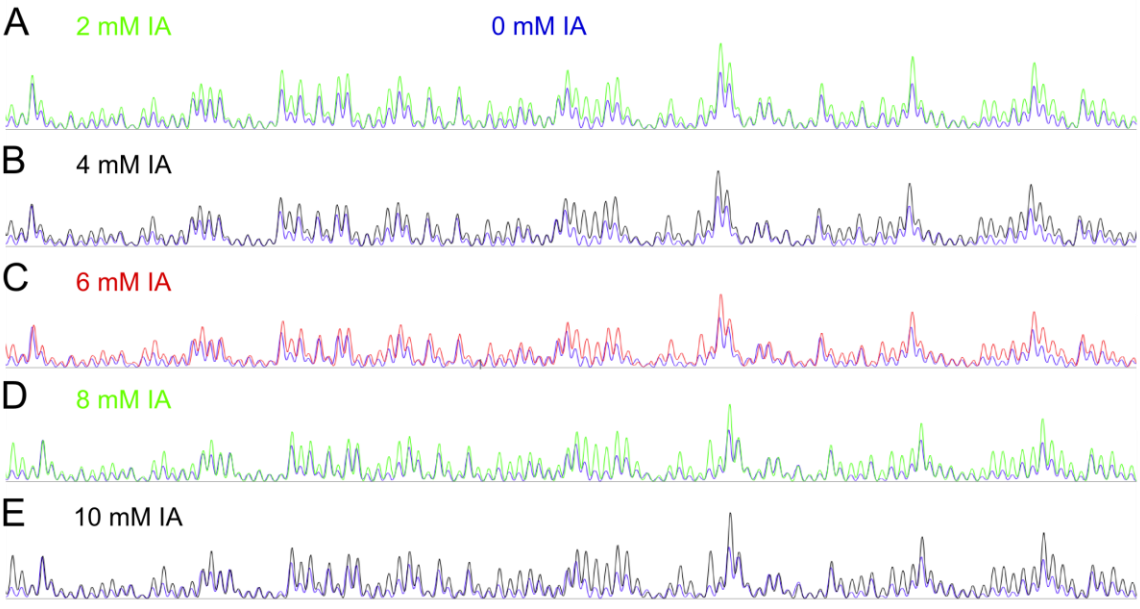


Figure III.14 Optimization of IA concentration for structure probing. Gene *60* mRNA was incubated with 0 (blue), 2 (green), 4 (black), 6 (red), 8 (green) or 10 (black) mM IA for 40 min at 37°C. Comparing each trace to that for the unprobed mRNA (blue) reveals that 10 mM IA is not enough to sufficiently modify the RNA and produce adequate signal above background. 25 mM IA was used in further reactions.

half life of 0.25 s in water. This fast SHAPE reagent was used at a concentration of 40 mM[114].

Generating Models from CE Data with ShapeFinder and RNAstructure Softwares

Electropherograms were analyzed with ShapeFinder software developed by Kevin Weeks and Morgan Giddings as described[126] (Figure III.2 C). Briefly, traces were fitted with a baseline and smoothed. Traces belonging to different fluorophores were shifted horizontally with the mobility shift function to ensure coincidence of peaks for each nucleotide. Traces were corrected for intensity decay over the length of the trace. The vertical scaling of the 6FAM and HEX channels (probing reaction and negative control, respectively) was adjusted such that low intensity 6FAM peaks overlapped peaks in the HEX channel (~2:1 6FAM:HEX). Finally, the difference in integrated intensity between probed and unprobed peaks was calculated providing a quantity proportional to the frequency of modification at a particular nucleotide.

Normalization of data was performed in Excel (Figure III.2 D). The normalization factor was determined by excluding the highest top 2% of hits and averaging the next highest 8%[114]. Data normalized by this factor generally range from roughly 0 to 1.5, where a value of 1 indicates strong reactivity. Probing was performed at least three times with each reagent and averaged, with the standard error (n = 3) in reactivity at each nucleotide being generally < 0.1 normalized units. When two primers were used to analyze one mRNA template, the datasets were joined by averaging values for overlapping nucleotides. Nucleotides near the ends of either primer dataset that had standard error > 0.1 normalized units were excluded as the trace data becomes unreliable in these regions.

The reactivity of gene *60* mRNA with Tb^{3+} was consistently lower than that from 1M7 datasets. Average reactivity values for the length of wildtype gene *60* mRNA were 0.35 and 0.43, respectively. Because the normalization factors differ between the two reagents (~65 for Tb^{3+} and ~47 for 1M7), these datasets were re-normalized such that both had an average reactivity of 0.4 for the entire RNA. This adjustment allowed for direct comparison of reactivity between datasets.

Secondary structure models of gene *60* mRNA were generated with RNAstructure software[128] (Figure III.2 E). Normalized reactivity data collected above were used as pseudo-free energy constraints, with a slope of 2.6 kcal/mol and an intercept of -0.8 kcal/mol, to calculate a partition function for the mRNA. The probability for each base pair as determined by the partition function was used as input for the MaxExpect algorithm to generate the minimum free energy structure compatible with the given constraints. The ProbKnot tool[103] was also tested; however, no pseudoknots were predicted to form in gene *60* mRNA. Without further experimental evidence, the MaxExpect structure with the minimum free energy became our best model for gene *60* mRNA structure (Figures IV.1 and IV.2).

Finally, to evaluate models of gene *60* mRNA, the normalized reactivity of nucleotides was compared to their environment in each model. To do this we assumed that protected nucleotides with reactivity < 0.3 should be structured and nucleotides with reactivity > 0.6 should be single stranded. Nucleotides with reactivity > 0.3 but < 0.6 could be assumed correctly modeled regardless of whether they are single-stranded or base-paired. Alternatively, we could assume that medium reactivity is not sufficient to determine the structure of these nucleotides and thus these nucleotides are excluded from the calculation. Both calculations of the "fit" of experimental data to each model are discussed in Chapter IV.

Conclusions

Analysis of structure probing data by CE has greatly improved the accuracy and speed of RNA structure generation. It is a versatile tool that can rapidly resolve primer extension products generated from any chemical or modifying protein that generates a stop during reverse transcription[46]. To establish this cutting-edge, rapid structure determination capability in our lab, several steps of this procedure were optimized using gene *60* mRNA as a model template. Importantly, the combination of fluorophores used to analyze data must be compatible with the matrixing algorithm applied. The concentration of primers should be kept in 20% excess of template RNA to ensure good signal-to-noise of data peaks and SSIII RT should be maintained at 200 U per reaction to obtain consistent and comparable results among experiments[124]. For templates varying in

length, the concentration of ddNTPs required to produce a strong sequencing ladder throughout the entire trace must be optimized independently. Additionally, the amount of each probing reagent required to generate good signal under less than single-hit conditions must be titrated. The optimal concentration of reagent differs for substrates of varying lengths, with shorter transcripts requiring higher concentrations of reagents[115]. Processing of raw electropherograms into "quantitative" measurements of the environment at each nucleotide position can be easily accomplished using ShapeFinder software[126]. These values can then be incorporated into various RNA folding programs to enhance the accuracy of structural models[109,127,135]. With these tools in hand, our lab is now poised to rapidly determine the structure of any RNA of intriguing function.

Chapter IV

The Role of mRNA Secondary Structure in Translational Bypassing

Introduction

In the middle of the open reading frame (ORF) of Bacteriophage T4 gene *60* mRNA lies an in-frame stop codon that leads to production of two protein products: one corresponds to termination at the stop codon and the other corresponds to translation of the entire ORF except for 50 nucleotides including the in-frame stop codon. The latter product is a single polypeptide chain generated from two distinct regions of the mRNA by a rare phenomenon known as translational bypassing[19,23-25,28,29] (Figure I.1). The untranslated region, or coding gap, of gene *60* mRNA neither shares any homology to self-splicing introns[25,33] nor gives rise to any detectable levels of spliced RNA [25]. The next simplest explanation for bypassing is an extension of traditional frameshifting or starvation hopping: the stop codon or a downstream structure could stimulate peptidyl-tRNA:mRNA dissociation (take-off), allowing the ribosome to scan the mRNA until the peptidyl-tRNA can reattach (or land) at a complementary codon and resume translation[15-17,19,23,24,27,43,97]. This simple scanning mechanism is disfavored because bypassing is extremely efficient (50%) and skips over a very large number of nucleotides (50)[17,29,43]. Also, deletions in the coding gap that bring the take-off and landing sites closer to one another should statistically enhance bypassing if the ribosome is simply scanning the mRNA[43]. By contrast, deletions in this region decrease bypassing efficiency[28]. Most importantly, however, a potential closer landing codon for the peptidyl-$tRNA^{Gly}$ is never used[27]. These data suggest the ribosome is not scanning, but rather completely skipping over the coding gap in gene *60* mRNA.

Several groups have hypothesized that the coding gap RNA plays a functional role in translational bypassing by adopting a compact structure[19,22,23,25-28,30] (Figure I.5).

In the first proposed structural model, codons before and after the coding gap were placed in close proximity to one another by the mRNA structure, such that the ribosome could simply continue translating without ever unfolding and scanning the coding gap region[25]. As an extension of this model, another group hypothesized that the bonding patterns in the UGGAU pentanucleotide repeat on either side of the coding gap could migrate as in a Holliday junction, so the ribosome would never have to translate over a discontinuity in the mRNA[22]. A third model proposed that the peptidyl-tRNA dissociates from the mRNA and bypasses a tRNA-like structure in the coding gap before landing at the wildtype site[27]. Finally, several other structural models for the coding gap were proposed from computational analysis, but have not been tested experimentally[26]. Common to all of these models is a tight-turning UUCG hairpin at the 5' edge of the coding gap[19,22,23,25-28,30] (Figure I.5 A).

Extensive mutational analysis has been performed to test various hypotheses about the structural requirements of the coding gap[23,27,28,42]. Most importantly, the presence of a stable hairpin at the 5' edge of the coding gap is crucial for bypassing. Destabilization of this hairpin or extensions to the loop region greatly decreases bypassing efficiency[23,27,28], and identification of protein products from these mutants reveals that part of the decrease in efficiency arises from landing at the closer Gly codon[27]. Thus the stability of the 5' hairpin modulates bypassing versus scanning of the coding gap region. Other features known to be important for bypassing are matching take-off and landing codons[28,42], an in-frame stop codon[23,28] and a short Shine-Dalgarno (SD) element just upstream of the landing codon[27]. Additionally, the nascent peptide sequence of gene *60* mRNA that lies in the ribosomal exit tunnel as the ribosome approaches the coding gap is thought to be important for stimulating take-off of the peptidyl-tRNA[19,27,28,51,52,59,136].

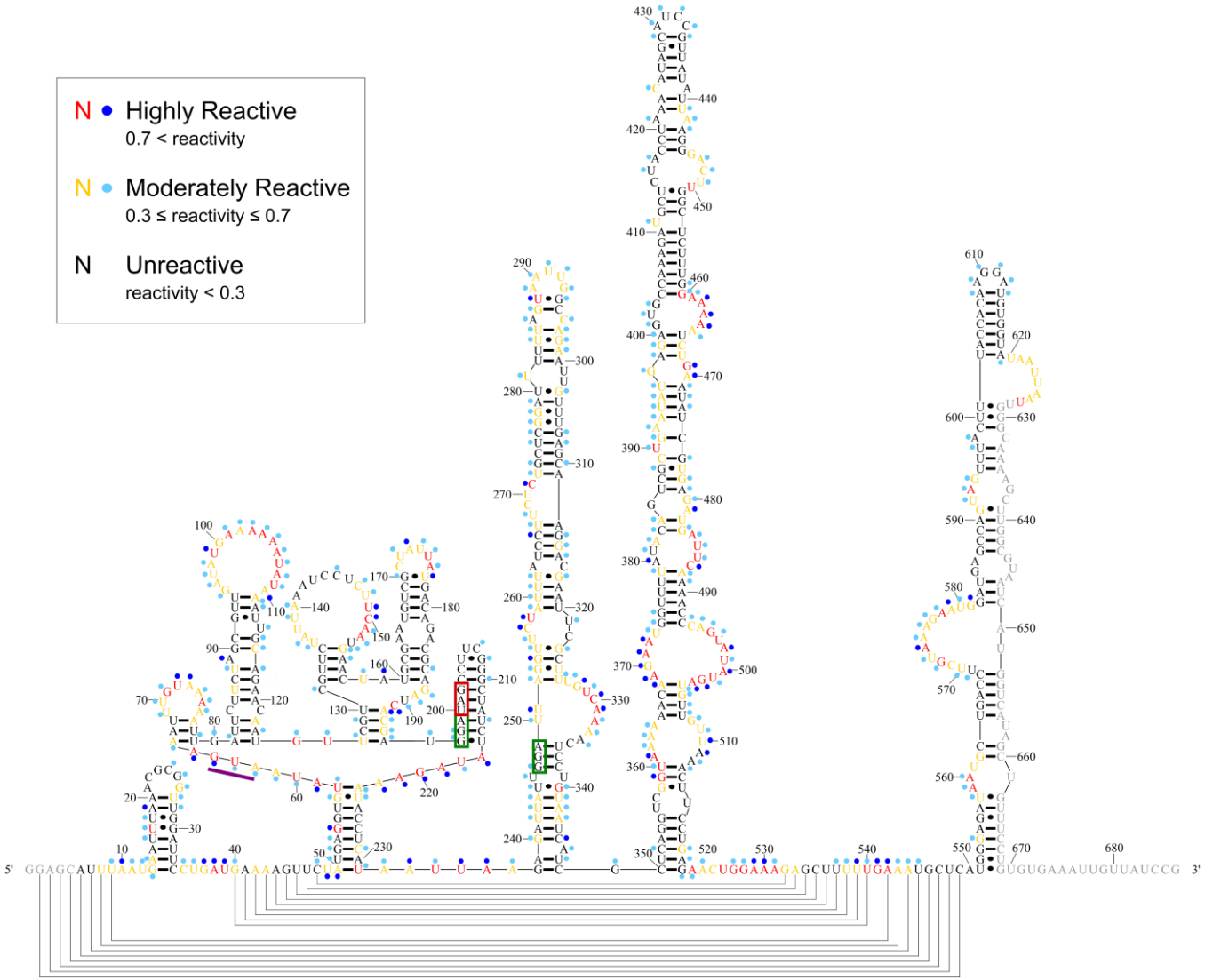
Despite decades of mutational and computational studies suggesting various structural models of the coding gap, to date the structure of gene *60* mRNA has not been directly probed. In an effort to confirm or refute various structural hypotheses, we probed the structure of gene *60* mRNA using the small molecule RNA modifiers Tb^{3+} ions and SHAPE reagents as described in Chapter III. Our results confirm the presence of the tight-turning hairpin at the 5' edge of the coding gap, and reveal that the rest of the

coding gap is mostly unstructured. The *in vitro* translation assay developed in Chapter II was used to confirm translational bypassing activity of gene *60* mRNA and various mutants thereof, and to characterize the pause in translation generated by translational bypassing.

Evaluation of Gene *60* mRNA Structural Models

Gene *60* mRNA was probed with either Tb^{3+} ions or the SHAPE reagent 1M7 as described in Chapter III. Electropherograms were analyzed using ShapeFinder software[126] and reactivity data were normalized in Excel. Overall, 1M7 reacted more strongly than Tb^{3+}, giving average normalized reactivity along the mRNA of ~0.43 compared to ~0.35 for Tb^{3+}. To correct for this difference, the reactivity at each nucleotide position was normalized a second time such that the average reactivity throughout gene *60* mRNA was 0.40 for both reagents. This allowed us to directly compare reactivity at each nucleotide between the two reagents. Minimum free energy structures were generated with the Max Expect function in RNAstructure[128] using normalized reactivity at each nucleotide position as pseudo-free-energy constraints. The RNAstructure folding algorithm is only able to consider one dataset at a time, so gene *60* mRNA was folded with data from each reagent individually. Two separate models were obtained, corresponding to probing with Tb^{3+} and 1M7. These are presented in Figures IV.1 and IV.2, respectively. Nucleotides are color-coded based on reactivity of the reagent used to generate the model and blue dots denote reactivity of each nucleotide with the second reagent for comparison.

We next wanted to evaluate these two structural models to determine how well each matched the probing data. Nucleotides with reactivity greater than 0.7 (red or dark blue circles) were considered single stranded and nucleotides with reactivity less than 0.3 (black or no circle) were considered structured. Two separate calculations were performed, one assuming nucleotides with reactivity between 0.3 and 0.7 (orange or light blue circles) were correct regardless of their environment (single- or double-stranded). In the second calculation, these mid-range nucleotides were deemed unassignable. These results are summarized in Table IV.1. Regardless of the mid-range reactive residues,


N Highly Reactive
0.7 < reactivity
N Moderately Reactive
0.3 ≤ reactivity ≤ 0.7
N Unreactive
reactivity < 0.3
5' GGAGCAUUUAAUCCCUGAUGAAAGUUCUUAAUAGCGAACUGGAAAGAGCUUUUGAAAUGCUCAU GUGUGAAAUUGUUAUCCG 3'

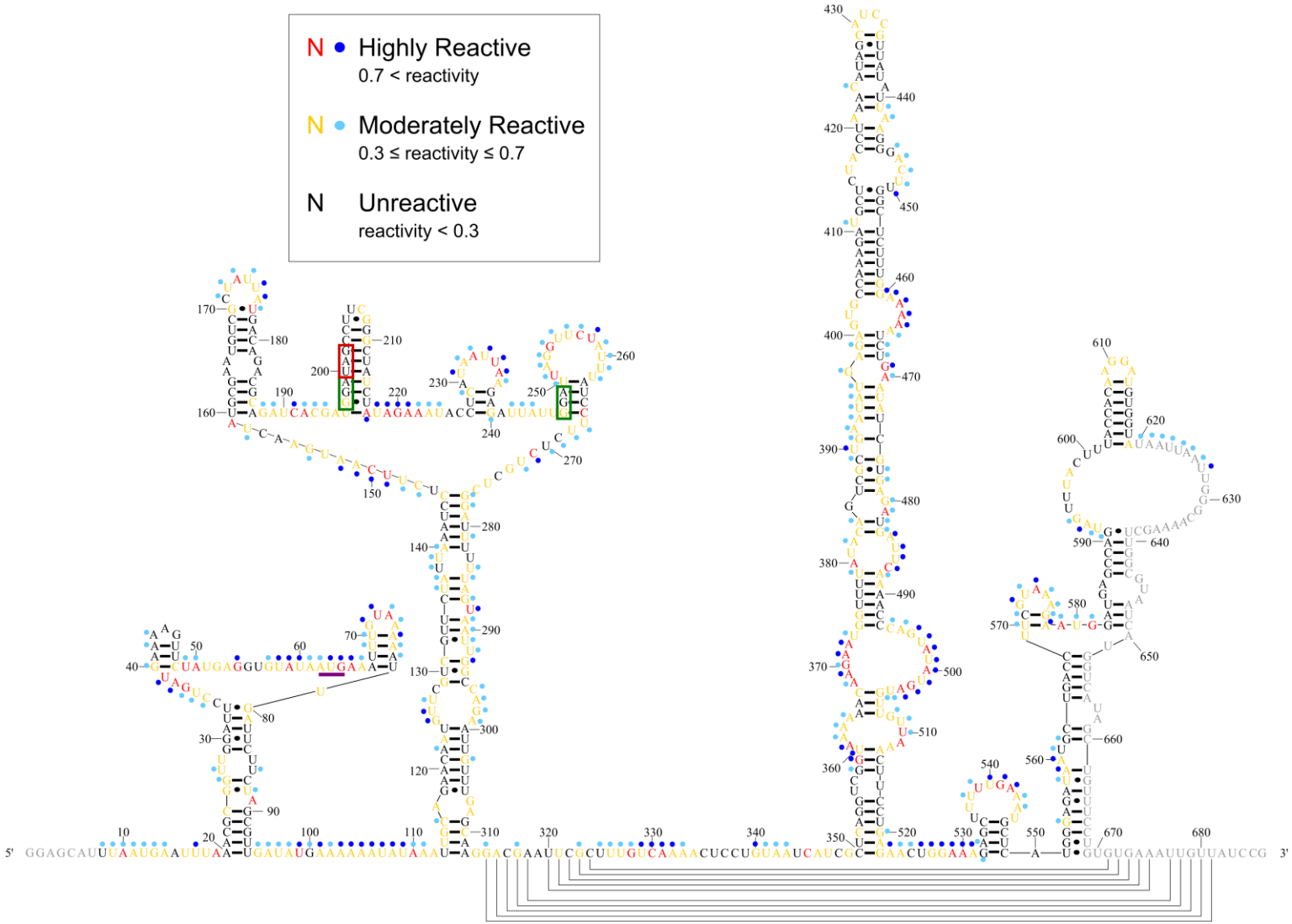

N Highly Reactive
0.7 < reactivity
N Moderately Reactive
0.3 ≤ reactivity ≤ 0.7
N Unreactive
reactivity < 0.3
5' GGAGCAUUUAAUGAAUUUAAUGAUAUGAAAAAAUAUAAAUAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUCGCGAACUGGAAAGCAUGUGUGAAAUUGUUAUCCG 3'

Figure IV.1 Gene *60* mRNA secondary structure model from Tb^{3+} data. (Page 77). Normalized reactivity with Tb^{3+} is annotated by nucleotide color as described in the legend. 1M7 data is overlaid on the Tb^{3+} structure with dark and light blue dots to denote highly and mildly reactive nucleotides, respectively. The coding gap is flanked by the take-off and landing codons boxed in green. The in-frame stop codon is boxed in red and the gene *60* start codon is underlined in purple. No reactivity data was available for nucleotides in gray.

Figure IV.2 Gene *60* mRNA secondary structure model from 1M7 data. (Page 78). Normalized reactivity with 1M7 is annotated by nucleotide color as described in the legend. Tb^{3+} data is overlaid on the 1M7 structure with dark and light blue dots to denote highly and mildly reactive nucleotides, respectively. The coding gap is flanked by the take-off and landing codons boxed in green. The in-frame stop codon is boxed in red and the gene *60* start codon is underlined in purple. No reactivity data was available for nucleotides in gray.

both datasets fit the Tb^{3+} model better than the 1M7 model. Additionally, the free energy of folding of the Tb^{3+} model is significantly lower than that for the 1M7 model, leading to several orders of magnitude preference for the Tb^{3+} model over the 1M7 model in solution. Finally, a minimal free energy structure generated with the average of Tb^{3+} and 1M7 normalized reactivity was not significantly different from that generated by Tb^{3+} data alone (data not shown). Based on these criteria, we chose the Tb^{3+} model as our best estimate of gene *60* mRNA structure.

	Tb^{3+} Model		1M7 Model	
Correctly Modeled Nucleotides	Tb^{3+} Data	1M7 Data	Tb^{3+} Data	1M7 Data
Mid-range Correct	86%	91%	81%	89%
Mid-range Unassignable	57%	46%	52%	44%
Free Energy of Folded RNA	-142.6 kcal/mol		-108.6 kcal/mol	

Assumptions:

reactivity < 0.3	structured
0.3 ≤ reactivity ≤ 0.7	mid-range
reactivity > 0.7	unstructured

Table IV.1 Fit of experimental data to structural models. Nucleotides were characterized as structured, mid-range or unstructured based on the assumptions listed above. Residues with mid-range normalized reactivity (orange or light blue circles) were either considered always correct or completely unassignable. Each experimental dataset was compared to both the Tb^{3+} model and the 1M7 model. The free energy of each structure is listed.

In accordance with the fact that both datasets fit the Tb^{3+} model best, the pattern of nucleotide reactivity along the majority of the transcript is similar for both reagents (Figure IV.3 A). There are a few regions, however, where the patterns are significantly different between reagents. These differences tend to occur at highly reactive, single-stranded nucleotides. Generally, the range of normalized reactivity for each nucleotide from three separate experiments is less than 0.1 normalized units. For highly reactive nucleotides, however, the extent of modification can vary significantly, leading to higher uncertainty in actual values for reactivity. Defined values, however, may not be required as long as these nucleotides are restrained as single-stranded during folding.

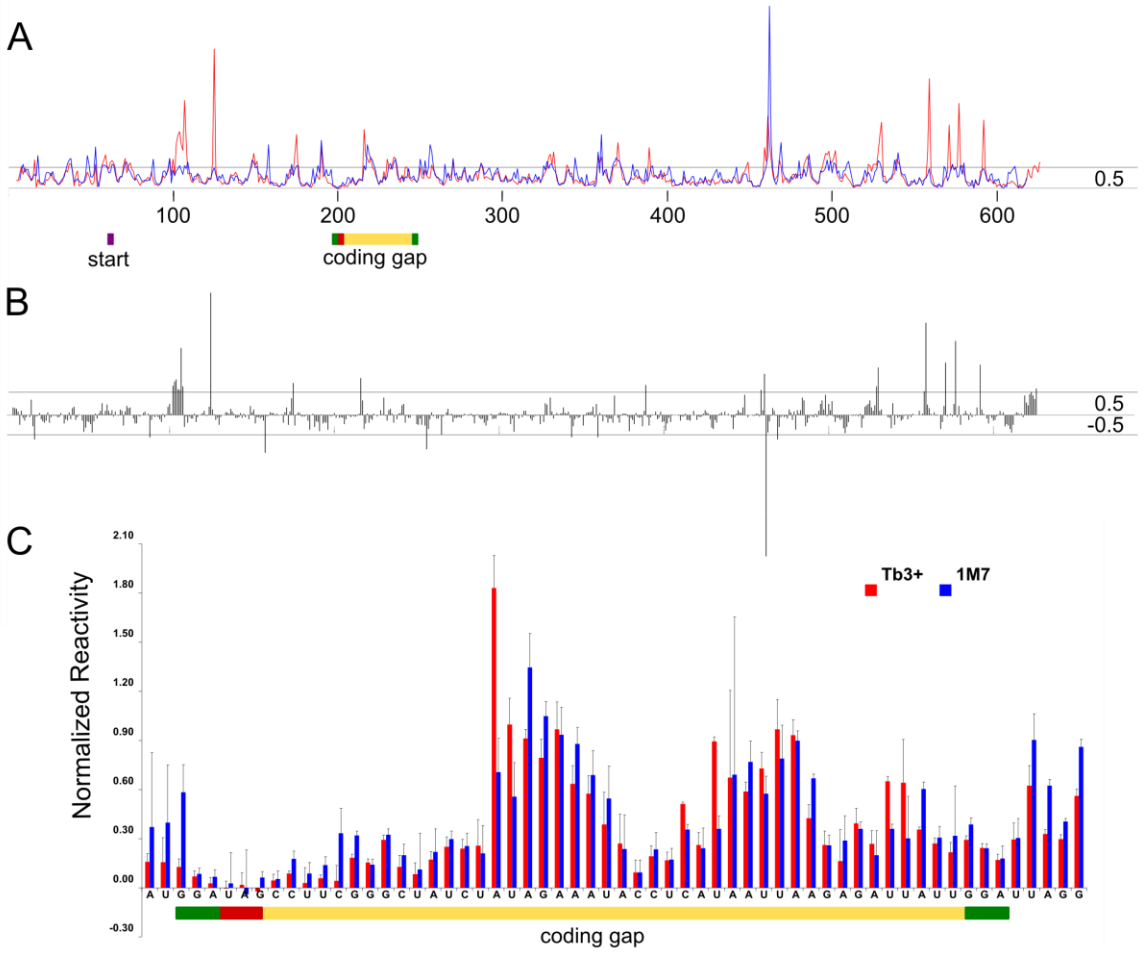


Figure IV.3 Tb^{3+} and 1M7 normalized reactivity for gene *60* mRNA. (**A**) Red and blue lines correspond to Tb^{3+} and 1M7 reactivity, respectively. Nucleotide number from the 5' end is plotted on the x-axis; normalized reactivity is on the y-axis. The position of the gene *60* mRNA start codon and the coding gap are indicated. The gray horizontal line marks a normalized reactivity of 0.5. (**B**) Difference between Tb^{3+} and 1M7 normalized reactivity (red minus blue). Gray horizontal lines indicate differences in normalized reactivity of 0.5 to -0.5. (**C**) Close up of normalized reactivity for the coding gap of gene *60* mRNA. Red – Tb^{3+} data; blue – 1M7 data. Error bars indicate standard error (n = 3).

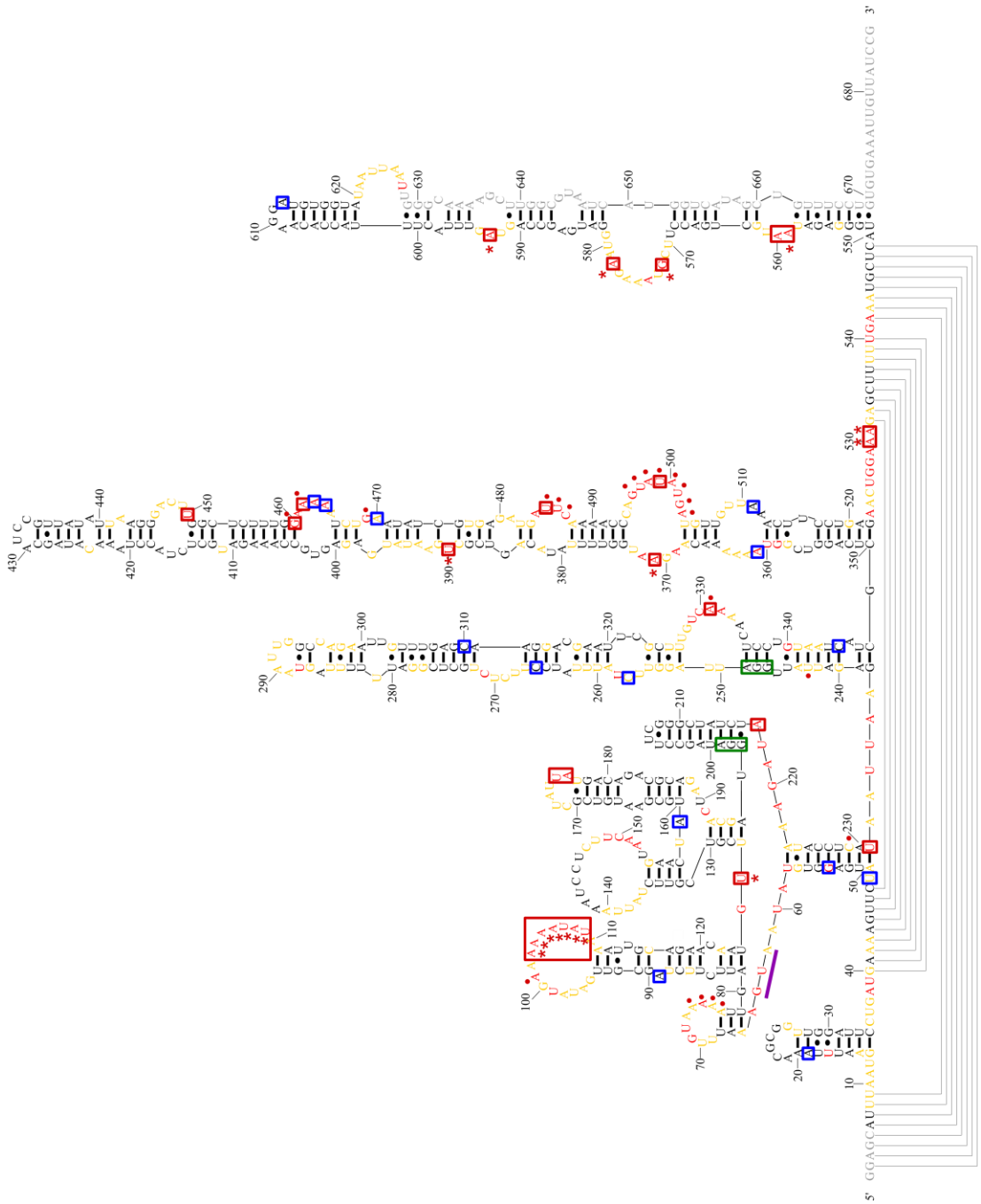


Figure IV.4 Tb^{3+} model with differences between Tb^{3+} and 1M7 data. Tb^{3+} model as described in Figure IV.1. For clarity, the stop codon is not marked and 1M7 reactivity is not shown for each nucleotide. Nucleotides more than 0.5 normalized reactivity units greater for Tb^{3+} than 1M7 are boxed in red. Nucleotides more than 0.5 normalize reactivity units greater for 1M7 than Tb^{3+} are boxed in blue. Red asterisks and solid circles indicate potential strong and weak metal binding sites, respectively.

Variation between the two reagents can be highlighted by plotting the difference between Tb^{3+} and 1M7 reactivity at each nucleotide position (Figure IV.3 B). Residues with 0.5 or more normalized units between the two reagents are boxed in Figure IV.4 and IV.5. Interestingly, all highly hit Tb^{3+} sites are in regions of the 1M7 structure predicted to be single-stranded or at the ends of stems (Figure IV.5, red boxes). A few examples of strong 1M7 hits, on the other hand, occur in the middle of stable stems in the Tb^{3+} structure (Figure IV.4, blue boxes). These few strongly hit nucleotides inconsistent with the Tb^{3+} model could disrupt stems during folding, leading to the significantly lower free energy of the 1M7 model. Given the similarity in mechanisms of SHAPE and Tb^{3+} probing, it was unexpected that specific nucleotides react more strongly with one reagent over the other. This could reflect slight preferences of the two chemicals for specific bases or different conformations of the RNA.

An additional benefit to probing with Tb^{3+} at low concentrations is that it binds primarily to metal binding sites and induces significant cleavage[118,119]. These nucleotides appear to have higher reactivity at low as opposed to high Tb^{3+} concentrations due to the normalization procedure (Figure IV.6). At higher Tb^{3+} concentrations, the signal of all other nucleotides increases due to probing of the RNA secondary structure, which in turn increases the normalization constant for the dataset. In Figure IV.4 and IV.5, potential metal binding sites are labeled with an asterisk if they are strong and a closed circle if they are weak. These locations correspond with many of the nucleotides that are hit more extensively with Tb^{3+} than 1M7, providing an explanation for these differences. For example, the loop between nucleotides 100 and 110 is strongly hit with Tb^{3+} ions (Figure IV.4), but only mildly with 1M7 (Figure IV.5). This likely indicates a metal binding site in this loop. Similarly, A371 and the opposite side of this bulge 496-503 (Figure IV.4 and IV.5) could be involved in metal ion coordination across the stem. Additional examples of individual nucleotides with the characteristics of a metal binding site (indicated in Figure IV.4 and IV.5), which could indicate tertiary folds stabilized by

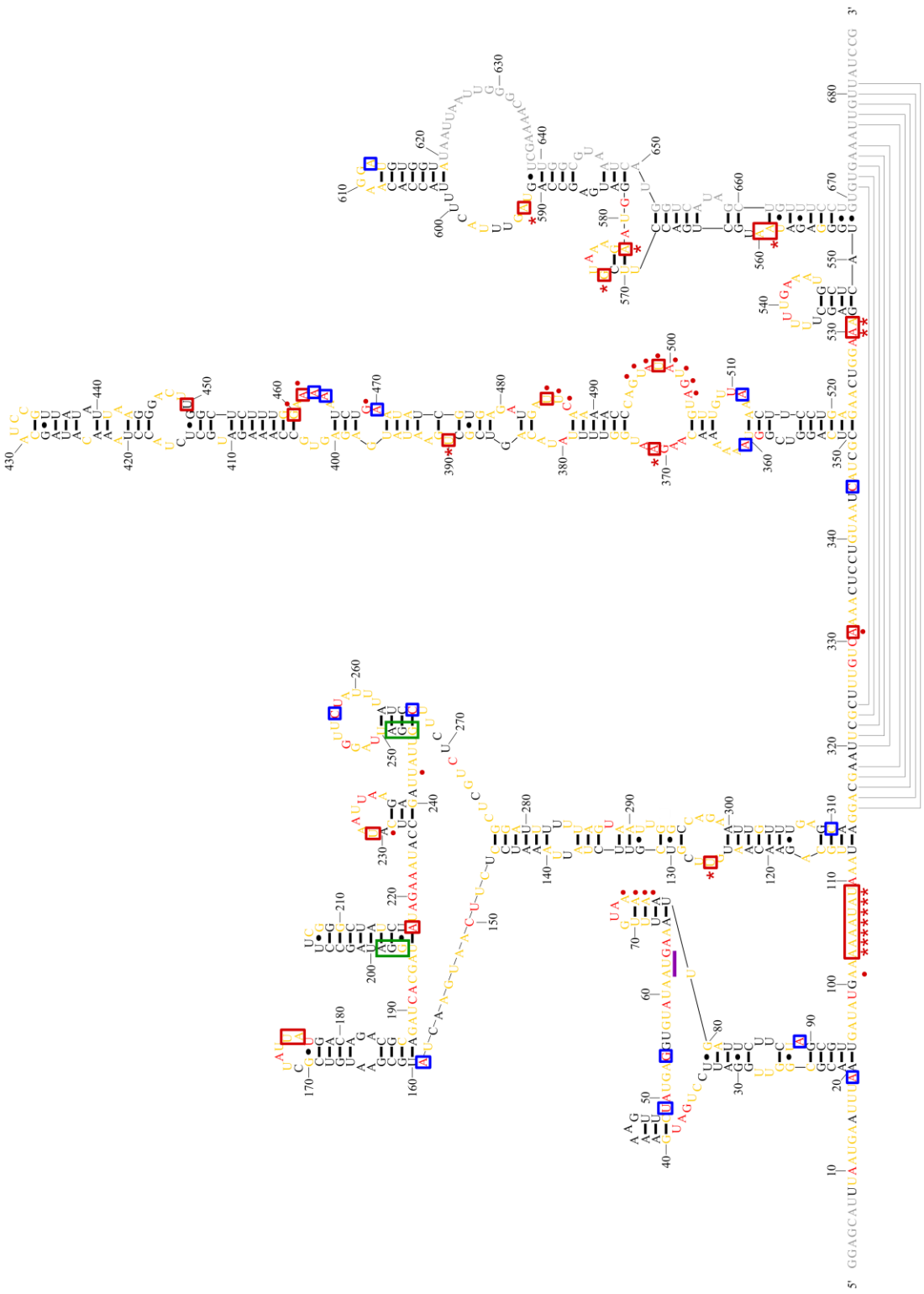


Figure IV.5 1M7 model with differences between 1M7 and Tb^{3+}data. 1M7 model as described in Figure IV.2. For clarity, the stop codon is not marked and Tb^{3+} reactivity is not shown for each nucleotide. Nucleotides more than 0.5 normalized reactivity units greater for Tb^{3+} than 1M7 are boxed in red. Nucleotides more than 0.5 normalize reactivity units greater for 1M7 than Tb^{3+} are boxed in blue. Red asterisks and solid circles indicate potential strong and weak metal binding sites, respectively.

metal ions. It is likely that Tb^{3+} ions stabilize structural features in gene *60* mRNA by shielding its negatively charged backbone. This feature of charged metals could account for the lower free-energy model of gene *60* mRNA and consequently lower average overall reactivity to Tb^{3+} ions[137,138].

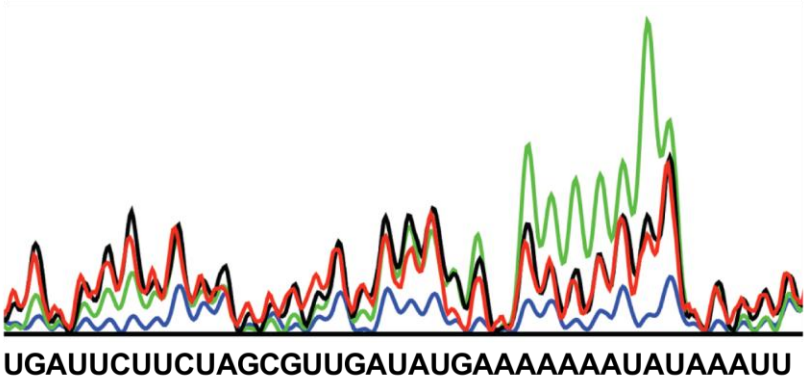


Figure IV.6 Probing with low concentrations of Tb^{3+} exposes metal binding sites. Data for a subset of gene *60* mRNA nucleotides probed with blue – 0 mM, green – 1 mM, black – 3 mM and red – 5 mM Tb^{3+}. Traces represent raw intensity from electropherograms. Nucleotides that are more reactive at low Tb^{3+} concentrations (green peaks are highest) represent metal binding sites (Figure IV.4 and IV.5, nucleotides 101, 103-109).

Coding Regions of mRNA have Less Structure than Regulatory Elements

To prevent significant delays during translation, coding regions of mRNA are relatively unstructured when compared to regulatory elements in the mRNA[124,125]. A prime example of this observation was found in the first ~1,000 nucleotides of the HIV-1 genome with SHAPE reagents. The 5' UTR composes the first 335 nts of the HIV-1 genome and contains several well-known, highly structured regulatory elements for viral replication and packaging. The rest of the probed region codes for viral proteins. In accord with the function of these two distinct regions, the median reactivity in the 5' UTR was 0.13 whereas coding regions have a median reactivity of 0.40[125]. Thus, probing data reflect the fact that structured regions of mRNA play regulatory roles during viral replication and coding regions tend to lack considerable amounts of structure. Another interesting feature of these data is the presence of significant structure between ORFs.

RNA structure in these regions, such as the frameshift signal between *gag* and *pol*, are known to directly impact protein expression[124]. As such, when probing a long mRNA, highly structured regions in an ORF or between ORFs likely indicate critical points for regulation during translation.

These principles were applied to our probing data for gene *60* mRNA to search for important points of regulation. As described above, 1M7 reacted more strongly than Tb^{3+} all along the mRNA giving median normalized reactivity of 0.34 and 0.27, respectively (after normalization such that the average reactivity of both for the entire mRNA was 0.4, as described above). These values are very close to the global median, 0.34, for the entire length of the HIV-1 genome, most of which codes for protein[124]. Thus, as expected for an ORF, gene *60* mRNA is relatively unstructured. Regions with structure are discussed in the next section.

The Coding Gap of Gene *60* mRNA is Only Partially Structured

The predicted hairpin at the 5' edge of the coding gap that is expected to regulate translational bypassing[19,23,27,28] is present in both models (Figure IV.1 and IV.2). Nucleotides 197 to 216 that compose this stem loop are all highly protected from both reagents (Figure IV.3 A and C): the average normalized reactivity of the stem nucleotides is 0.11 and 0.18 for Tb^{3+} and 1M7, respectively. Low reactivity includes the loop region of this hairpin as would be expected by the tight fold of a UUCG loop[134] (Figure III.11 C). Based on reactivity data, it appears the stem only extends through the take-off Gly codon, despite the potential for two additional Watson-Crick base pairs to form. In sum, our structure probing data confirms the presence of the regulatory 5' hairpin in the coding gap region.

The rest of the coding gap reacted strongly with both reagents (Figure IV.3 A and C, average reactivity is 0.54 for Tb^{3+} and 0.50 for 1M7) and, accordingly, has little predicted structure in our models (Figure IV.1 and IV.2). Reactivity in this region is also well above the average reactivity for the whole length of gene *60* mRNA, directly contradicting previous models postulating that the coding gap is highly structured. An overlay of Tb^{3+} reactivity on each of these proposed models reveals that, aside from the 5' hairpin, the only structure compatible with our experimental data is Figure IV.7 D,

which has minimal structure in the 3' end. From our probing data we can conclude the coding gap is mostly unstructured in solution. It is known, however, that the entire coding gap plays a role in translational bypassing because deletions from its 3' end decrease bypassing efficiency[28]. Furthermore, when the bypassing sequence is transferred among genomes by homing, the entire and exact sequence of the coding gap is always replicated[37]. Therefore the 3' end of the coding gap could function as a single-stranded spacer, fold cotranslationally or interact with the ribosome in some way to promote bypassing.

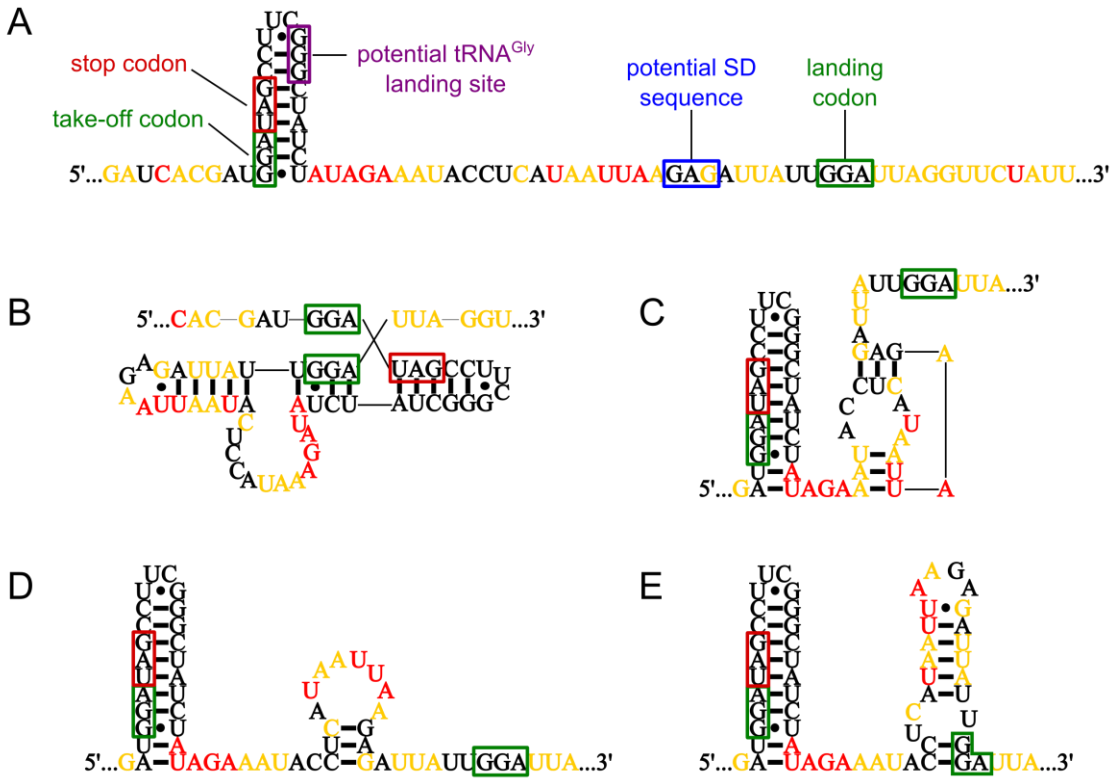


Figure IV.7 Overlay of Tb^{3+} normalized reactivity on proposed structural models. Nucleotides are color coded as in Figure IV.1. (**A**) Important regions of the coding gap are labeled as in Figure I.1. (**B**) Original model of gene *60* mRNA coding gap[22,25]. (**C**-**E**) Additional proposed models[26].

Despite its high reactivity in our Tb^{3+} model, part of the coding gap region pairs extensively with the SD of gene *60* mRNA (Figure IV.1). The potential for this interaction was noted in the original discovery of bypassing[25]. The authors stated that mutation of the SD to reduce its potential to base pair with the coding gap did not affect

bypassing levels, and they concluded this interaction was not relevant for bypassing[25]. We have independently confirmed this result (data not shown). This structure is also highly unlikely to exist during translation as the SD element needs to be single-stranded to initiate translation, and multiple ribosomes translating on the same mRNA would prevent the SD from repairing with the coding gap.

While the reactivity of the coding gap was similar for both reagents (Figure IV.3 C), the SD of gene *60* mRNA was modified more extensively with 1M7 than with Tb^{3+} with the exception of the very reactive G54 in both (Figure IV.4 and IV.5, just before start codon). Because the Max Expect folding algorithm finds the lowest free energy structure, the relatively unreactive SD was paired with the coding gap in the Tb^{3+} model. This interaction, although not required for bypassing *in vivo*, remains a major feature of our Tb^{3+} model. This could be simply an artifact of the folding algorithm, given how unlikely it would be to get a very strong hit in the middle of such a long stem (G54). By forbidding this interaction during folding in RNAstructure, the coding gap forms only minor structures within itself, as in the 1M7 structure, while the SD forms alternative base pairs with nucleotides 322-328 in the 3' end of gene *60* mRNA. Experiments to alleviate this ambiguity are outlined in "Future Directions" in Chapter V.

Gene *60* mRNA is an Ensemble of Structures

Another explanation for why we are not observing any substantial structure in the coding gap could be ensemble averaging. Probing was carried out at the ensemble level and our data is therefore an average of all structures that gene *60* mRNA samples. A native gel was used to estimate the number of major conformational populations for gene *60* mRNA (Figure IV.8). mRNA refolded under the conditions used for structure probing (red asterisk) forms three basic populations: one slower migrating band (C) and two faster migrating, but smeary bands (A and B). This result confirms that our probing data for wildtype gene *60* mRNA is an average of multiple conformational populations. Attempts to probe specific bands or regions of the smear for wildtype gene *60* mRNA were unsuccessful due to low RNA recovery from the gel.

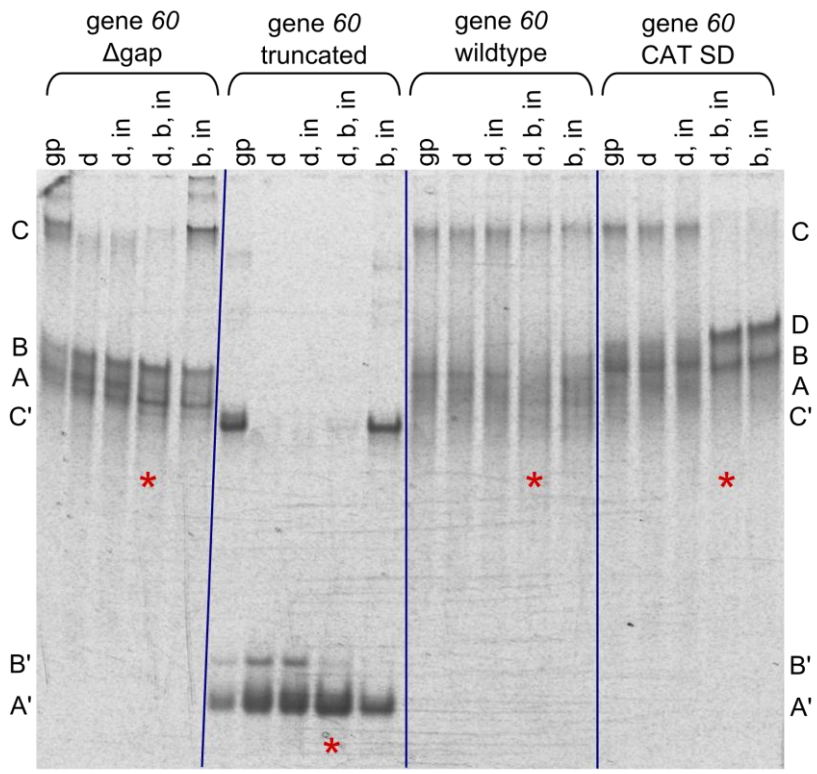


Figure IV.8 Native gel of gene *60* mRNA variants. 1 µg of the indicated RNA was loaded into each lane and resolved at 4°C in a 6% native THEMA gel for 7 hrs at 15 W. Bands were detected with UV light after staining with SYBR Gold for 25 min. Abbreviations: gp – gel purified; d – denatured; b – S30 buffer added; in – incubated at 37°C for 15 min. Brackets denote refolding process for the indicated gene *60* mRNA mutant. Blue lines separate data between each mutant. Red asterisks indicate refolded condition used for structure probing. Letters to the left and right of the gel mark specific bands; an apostrophe after the letter indicates a band for the truncated gene *60* mRNA.

A striking feature in both the Tb^{3+} and the 1M7 models (Figure IV.1 and IV.2, respectively) is apparent independent folding of the 3' and 5' domains of gene *60* mRNA. Aside from a few long-range interactions, the 3' end interacts entirely with itself forming long stem loops with several bulges each. It is known that the entire 3' end of gene *60* mRNA can be replaced by a *lacZ* gene without disrupting translational bypassing[28]. Thus if gene *60* mRNA structure plays a role in translational bypassing, critical structural features or interactions are unlikely to involve the 3' end.

By probing full-length gene *60* mRNA in solution, however, the potential for non-essential, long-range intramolecular interactions could bias our structural data in functional regions. As translational bypassing does not require the 3' end of the mRNA[28], gene *60* mRNA was truncated 6 nucleotides after the coding gap. Refolded truncated

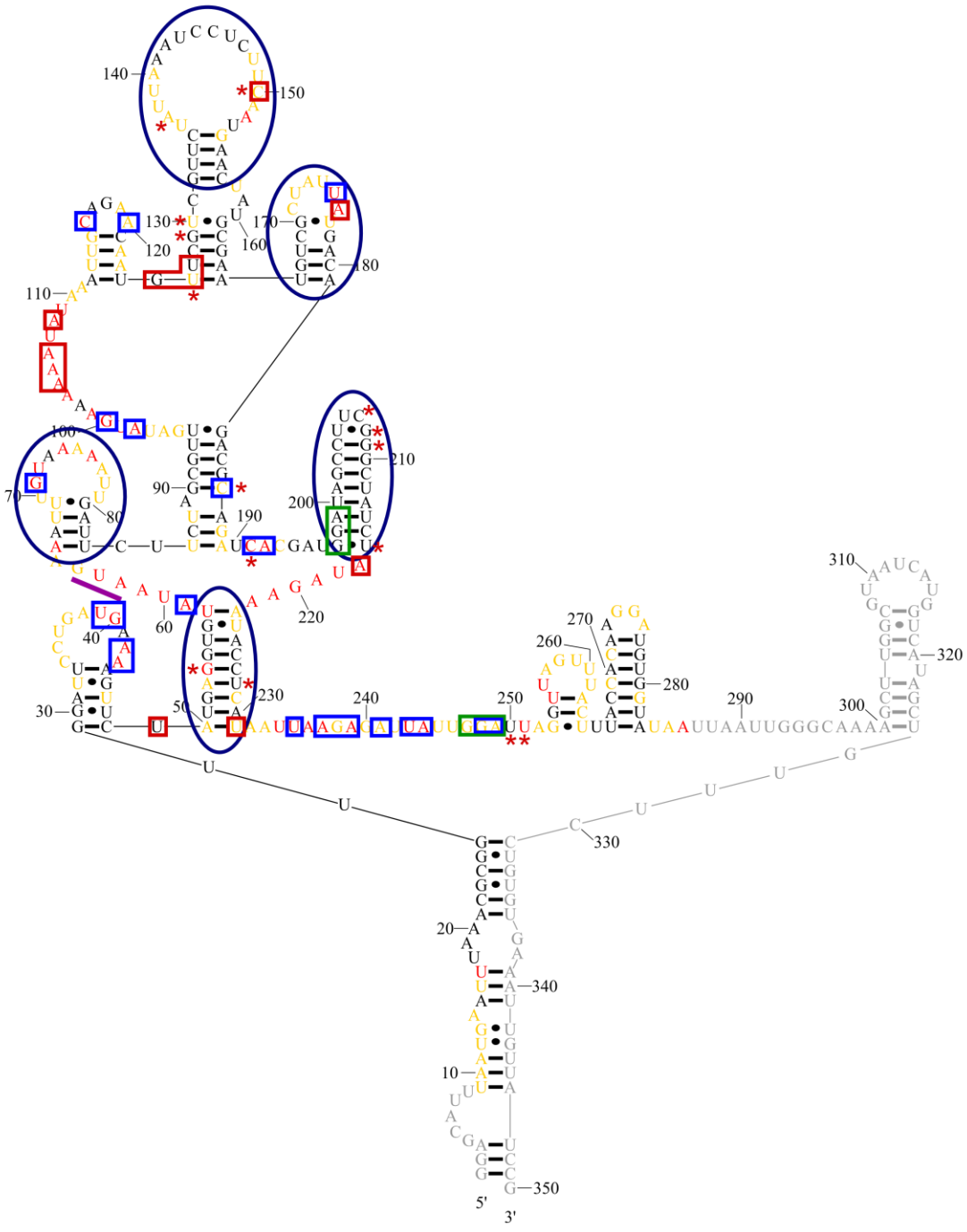


Figure IV.9 Tb^{3+} structural model of truncated gene *60* mRNA. The gene *60* ORF ends 6 nucleotides after the landing site. Nucleotides are colored as in Figure IV.1. Take-off and landing codons are boxed in green and flank the coding gap region. The gene *60* start codon is underlined in purple. Features in common between the truncated and full-length (Figure IV.1) models are circled in dark blue. Nucleotides more than 0.5 normalize reactivity units greater for full-length than truncated mRNA are boxed in red. Nucleotides more than 0.5 normalized reactivity units greater for truncated than full-length mRNA are boxed in light blue. Red asterisks indicate likely metal binding sites in the truncated gene *60* mRNA.

mRNA migrated as a single population through a native gel (Figure IV.8 A') making it a good candidate for structure probing the functional portion of gene *60* mRNA. Truncated gene *60* mRNA was probed with Tb^{3+} and 1M7 and the minimal free energy structures generated for each reagent were extremely similar to the 5' end of the full-length models for their respective reagents (Figure IV.9 and IV.10, 1M7 data not shown). Structural features in common between full-length and truncated gene *60* mRNA probed with Tb^{3+} are circled in dark blue. As in the full-length mRNA, the coding gap of the truncated mRNA is unstructured except for the 5' hairpin and a potential, but non-functional interaction with the SD region. The extensive similarity between the two models validates the 5' end of our model of full-length gene *60* mRNA and suggests the majority of the populational heterogeneity in full-length mRNA arises from dimerization or multiple conformations in the 3' end that do not greatly impact the structure of the functional 5' portion.

A difference plot between reactivity of the truncated and full-length mRNAs emphasizes the similarity between the two structures (Figure IV.10). Although the magnitude of reactivity for individual residues changes slightly for a few nucleotides, the overall pattern is quite similar. Nucleotides that differ by more than 0.5 normalized units are boxed in Figure IV.9. U126 shows the most extreme difference between the two datasets, becoming significantly protected in the truncated mRNA. This nucleotide is predicted to interact with metal ions in both the full-length and truncated gene *60* mRNA (Figure IV.4), yet is in a very different environment in the two structures. Interestingly, the potential metal binding site in the loop nucleotides 103-109 disappears in the truncated gene *60* mRNA probing, and while several new metal binding sites emerge. This indicates the truncated and full-length gene *60* mRNAs may have slightly different tertiary structures, potentially coordinated by metal ions. The 3' end of the coding gap shows some minor differences between the two mRNA templates, but the disparity in reactivity is likely because the sequence immediately downstream of the coding gap has been removed in the truncated mRNA, changing its local environment.

In Figure IV.8, truncated gene *60* mRNA had three populations in the native gel before denaturation (A', B' and C'). After denaturation the slower migrating band

disappeared and redistributed into the two lower bands. Addition of buffer shifted RNA out of band B' and fully into band A'. Addition of buffer without prior denaturation

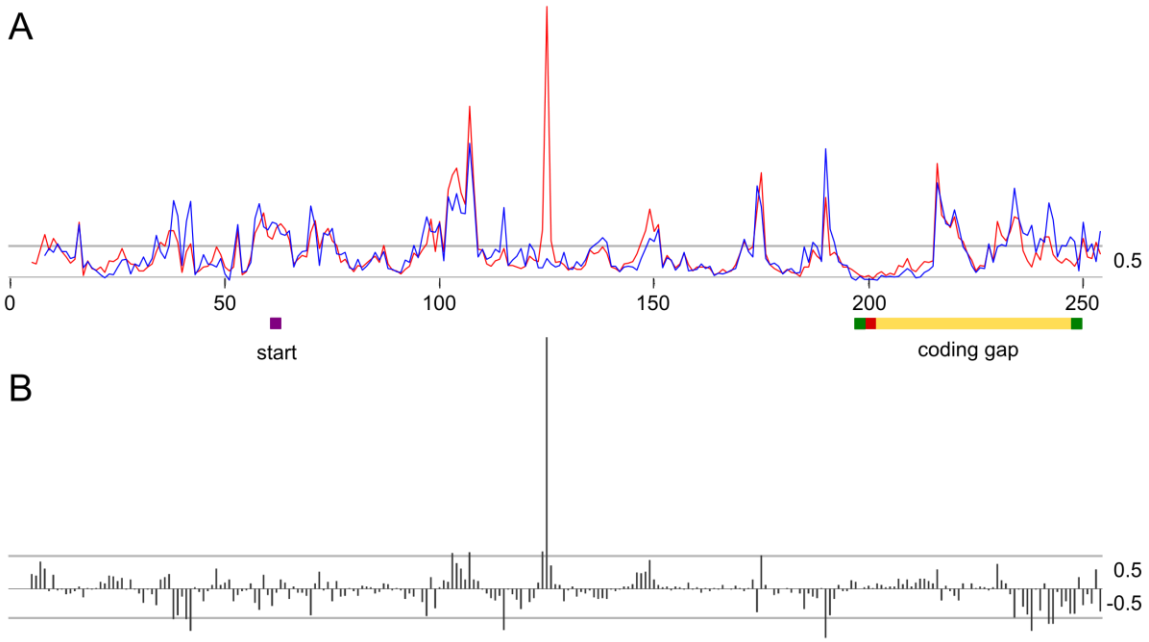


Figure IV.10 Normalized reactivity for full-length and truncated gene *60* mRNA. (**A**) Red and blue lines correspond to normalized reactivity for full-length and truncated gene *60* mRNA, respectively. Nucleotide number from the 5' end is plotted on the x-axis; normalized reactivity is on the y-axis. The position of the gene *60* start codon and the coding gap are indicated. The gray horizontal line marks a normalized reactivity of 0.5. (**B**) Difference between full-length and truncated normalized reactivity (red minus blue). Gray horizontal lines indicate differences in normalized reactivity of 0.5 to -0.5.

retained the slower migrating band C', but removed the population in band B'. These data signify that there is both a temperature dependent (C' to A') and buffer and/or Mg^{2+} dependent (B' to A') step to homogeneous folding of gene *60* mRNA. The full-length gene *60* mRNA did not make these same changes upon refolding. It maintained the same three populations (A, B and C) throughout.

These differences in folding behavior between full-length and truncated gene *60* mRNA were compared to those of gene *60* delta gap and a gene *60* mRNA whose SD had been mutated to that of the chloramphenicol acetyltransferase (CAT) gene (this SD lacks the ability to base pair with the coding gap). The refolding of gene *60* delta gap mRNA initially contained the same three bands as wildtype gene *60* mRNA. Upon denaturation the topmost band redistributed into the two faster migrating bands (C to A and B). Addition of buffer exhibited no further changes to the population. Thus, in the absence of the coding gap region there is a temperature dependent conformational change.

The CAT SD gene *60* also contained the same initial three bands as wildtype gene *60* mRNA, but exhibited completely different patterns upon refolding. Heat denaturation of the mRNA had no effect on the distribution of populations, but addition of buffer with Mg^{2+} caused bands A and C to redistribute into B and a new conformation, D. This indicates there is a buffer and/or Mg^{2+} dependent change into a new conformation involving the new SD.

Altogether, these data suggest that mutants with the wildtype SD have a heat-dependent refolding step while constructs containing the coding gap undergo a buffer and/or Mg^{2+} dependent rearrangement. Wildtype gene *60* mRNA, however, appears trapped in multiple conformations that are not easily convertible by heat, salts or metal ions, which could be due to an extensive interaction between the coding gap and the SD. Although our model of the truncated mRNA also predicts an interaction between these two elements, its conformation is sensitive to heat and buffer and/or Mg^{2+}. This suggests there may be some contacts between the 3' end of gene *60* mRNA and the base-paired SD-coding gap that stabilizes the slower migrating population. Even with the presence of the 3' end, both the coding gap and SD need to be present to fully trap the mRNA in the topmost species. However because the 3' end of gene *60* mRNA is not required for translational bypassing, it is unlikely that this tertiary interaction plays a role in bypassing. Rather it could be a feature of probing the naked RNA in S30 buffer.

Gene *60* mRNA Does Not Dimerize

When varying the mRNA concentration during optimization of our *in vitro* translation assay, we noted that the bypassing efficiency decreased as the mRNA concentration increased above 100 nM (Figure II.4 and II.5). We hypothesized that high concentrations of gene *60* mRNA may dimerize to inhibit bypassing and prevent overproduction of gp60. A dimer of gene *60* mRNA could explain our slow migrating band in the native gel (Figure IV.8, band C). As structure probing was carried out at 100 nM mRNA, it is possible that our data does not accurately model a monomer of gene *60* mRNA, but rather a mixture of monomers and dimers. To test this hypothesis, gene *60* mRNA was radiolabeled at its 5' end with a ^{32}P phosphate. After gel purification, radiolabeled mRNA was folded with varying concentrations of cold gene *60* mRNA.

Complexes were resolved on a 6% native THEMA gel at 4°C for 7.5 hrs. The radiolabeled complexes were very similar to those observed by SYBR Gold staining (Figure IV.8): one slowly migrating band and two faster migrating, but smeary bands. Addition of cold mRNA up to 500 μM did not lead to a significant change in any population of gene *60* mRNA, indicating it is unlikely that gene *60* mRNA dimerizes (data not shown). The decrease in bypassing efficiency at higher concentrations of mRNA could instead indicate that we are titrating out a cellular factor required for bypassing or that we are simply saturating the system with non-physiological concentrations of RNA.

A Protected Loop is Complementary to the Take-off and Landing Sites

As there did not appear to be a straightforward structural mechanism for translational bypassing based on our models, we searched for other areas of the mRNA that might be important. Nucleotides 140-146 reside in a large loop in the Tb^{3+} model, yet are highly protected from both reagents (Figure IV.11 A, circled in red). Nucleotides in large loops are generally highly reactive, so we looked for potential Watson-Crick binding partners of the protected nucleotides. Regions in gene *60* mRNA complementary to the protected nucleotides are boxed in green and its interaction with nucleotides 277-282 in the 1M7 model (Figure IV.2) is boxed in blue. Surprisingly, the protected nucleotides are complementary to both the take-off and landing codons (Figure IV.11 A and B). We hypothesized these nucleotides could interact in some way during translational bypassing.

To test this hypothesis, we mutated the protected loop nucleotides. They reside in the coding region for the critical nascent peptide. Previous mutations of this region to synonymous codon demonstrated that the protein sequence rather than the nucleotide sequence was critical for bypassing (Figure IV.11 B and C)[28]. The synonymous codons chosen completely disrupt the ability of nucleotides in this region to base pair with the take-off or landing codons, strongly suggesting that these sequences are serendipitously complementary and this potential structure does not play a role in bypassing. To confirm this result, we mutated the loop nucleotides to different synonymous codons (Figure

IV.11 C) and compared the bypassing efficiency of the mutant to wildtype gene *60* mRNA. We found that the mutant underwent translational bypassing to the same extent

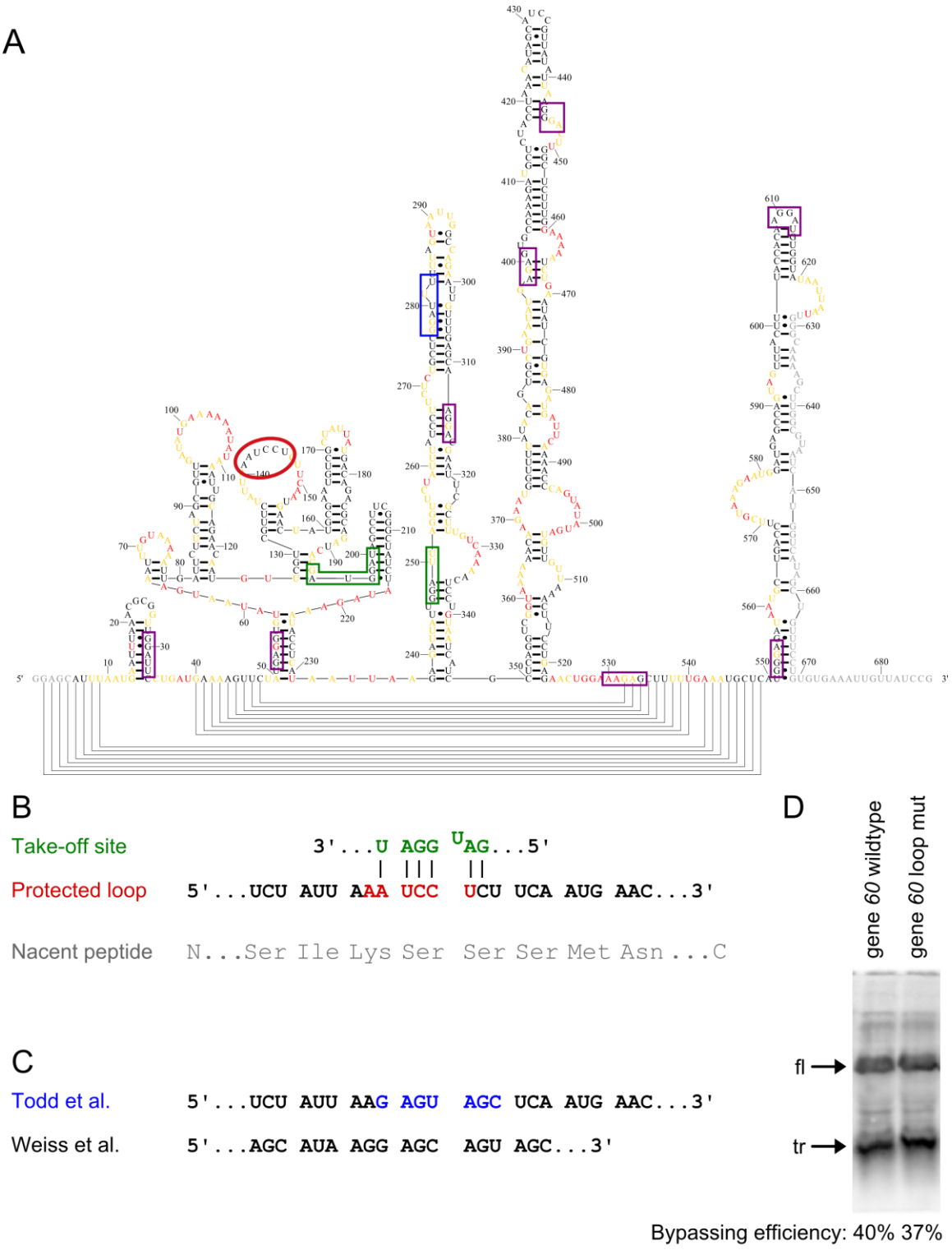


Figure IV.11 The potential interaction between protected loop nucleotides and the take-off and landing sites is not required for translational bypassing. (**A**) Tb^{3+} secondary structure model of gene *60*

mRNA. Nucleotides circled in red are consistently protected from both probing reagents. Boxes mark regions of complementarity to protected nucleotides. The blue box (277-282) marks the pairing partners of the protected nucleotides (in red circle) that are present in the 1M7 model (Figure IV.2). There are green boxes around the complementary take-off and landing regions. All other regions of complementarity to the protected nucleotides (in red circle) are boxed in purple. (**B**) Schematic of possible interaction between the take-off site (green) and the protected loop nucleotides (red). In gray below is the encoded peptide sequence. (**C**) Mutations to synonymous codons in the protected loop that disrupt its potential to base pair with the take-off and landing sites. (**D**) Translational bypassing efficiency of the mutant by Todd et al. and wildtype gene *60* mRNA.

as wildtype (Figure IV.11 D), confirming previous results that nucleotide sequence and structure in this region does not play a role in bypassing.

There are other potential binding partners for loop residues 141-147 (Figure IV.11, boxed in purple). The most highly protected nucleotides complementary to these loop residues are nucleotides 29-33 within the 5' UTR. The current binding partners of this region in the Tb^{3+} model contain the very strongly reactive U17, which disfavors the modeled interaction. Alternatively, instead of residing in Watson-Crick base pairs, the unreactive loop nucleotides could be single-stranded, but locally structured. Similar C and U-rich loop regions in other RNAs have been found to have low reactivity towards SHAPE reagents[139]. These have been explained by the slightly lower reactivity of C and U nucleotides when compared to A and G[140]. However, Tb^{3+} data exhibits the same trend for these nucleotides as 1M7. Single-stranded runs of C can form helices that are significantly more stable than those of other homopolymers. Thus it is possible that there is some local stabilizing structure in this region[141] protecting nucleotides from structure probing reagents.

Kinetics of Translational Bypassing

Hairpins and pseudoknots are known to stall the ribosome, often at points of translation regulation[15-18,20,97]. Because there is a hairpin at the edge of the coding gap in gene *60* mRNA, and this structural element plays a critical role in regulating translational bypassing, we hypothesized a significant pause at this point in translation might be observable. If so, structure probing of gene *60* mRNA during the pause may provide a better understanding of the mechanism of translational bypassing.

Products of translational bypassing were monitored over time with a ^{35}S-Met pulse-chase assay to obtain single turnover conditions. The results are presented in

Figure IV.12. Production rates of truncated and full-length gp60 from wildtype gene *60* mRNA were compared to the rate of production of these proteins independently from

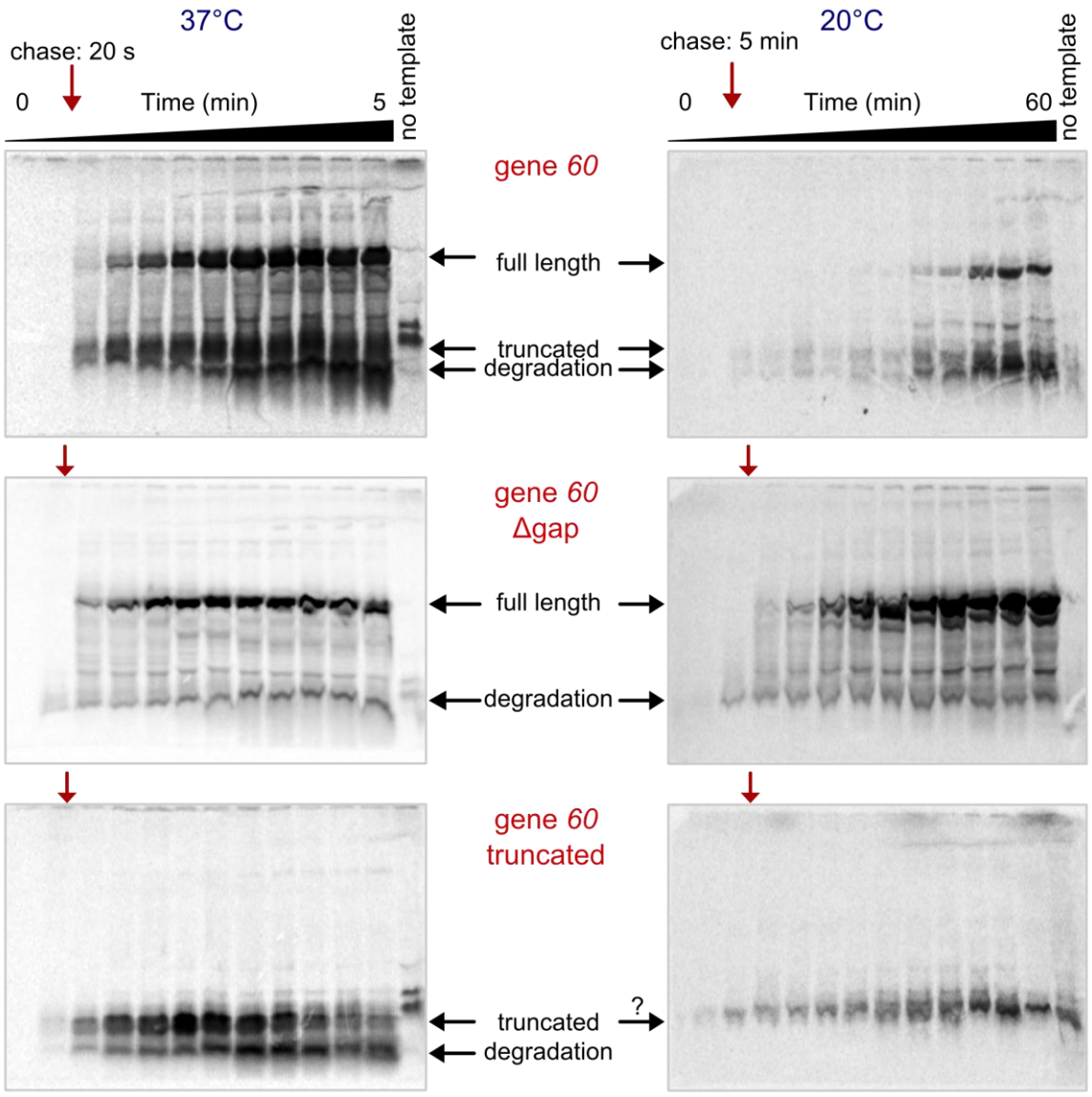


Figure IV.12 Time course of translational bypassing. Products of translation were monitored from gene *60*, gene *60* Δgap and truncated gene *60* mRNA templates as indicated. Aliquots were removed at 0, 10, 30, 50, 70, 90, 110, 140, 170, 200, 240 and 300 s for reactions at 37°C. Aliquots were removed at 0.2, 1, 5, 7.5, 12.5, 15, 20, 25, 30, 40, 50 and 60 min for reactions at 20°C. Full-length and truncated gp60 bands are labeled. Degraded protein is also indicated. Red arrows indicate 20 s or 5 min after the start of the reactions (at 37°C and 20°C, respectively) when the 35S-Met was chased with excess cold Met.

truncated and gene *60* Δgap mRNA, respectively. These mutants provide a measure of the time it takes short and long proteins to form in the absence of translational bypassing. At 37°C there is ~20 s delay in formation of full-length gp60 from gene *60* delta gap

mRNA relative to truncated gp60 from truncated gene *60* mRNA, which is reasonable if we assume that the rate of translation in our *in vitro* assay is slightly lower than that *in vivo* (10-20 amino acids per second) and note their difference in size is 114 amino acids. Interestingly, formation of truncated gp60 from wildtype gene *60* mRNA exhibits a ~10 s delay relative to that from the truncated gene *60* mRNA. This delay could be because the coding gap of wildtype gene *60* mRNA is base-pairing with its SD and this interaction must be disrupted before translation can be initiated. Formation of full-length gp60 from wildtype gene *60* mRNA is delayed ~30 s relative to that produced by gene *60* delta gap mRNA. If we assume 10s of that delay is required for disruption of the SD-coding gap interaction then the remaining 20 s is attributable to translational bypassing. Thus, there is a measurable pause in translation associated with bypassing and conditions could be optimized to probe gene *60* mRNA structure in this window of time during translation.

Temperature Dependence of Translational Bypassing

The genes *52*, *39* and *60* from which the DNA topoisomerase in Bacteriophage T4 is composed are only essential to the phage at temperatures below 25°C[35]. One possible mechanism by which the phage could upregulate production of the topoisomerase during times of temperature stress would be to enhance bypassing efficiency at lower temperatures. This could be accomplished by forming more stable structures in the coding gap region to enhance the rate and perhaps efficiency of bypassing, leading to production of higher quantities of full-length gp60 for incorporation into the topoisomerase. To test this hypothesis, a time course was performed at 20°C. Translation of gene *60* delta gap and truncated gene *60* mRNA proceeded slowly, but produced protein within 10-15 min. Wildtype gene *60*, however, did not generate any full-length gp60 until 25-30 min. It was difficult to assess whether truncated gp60 was forming before 25 min or if these smaller products were part of the background. These results suggest that translational bypassing may become extremely slow at low temperatures, counter intuitive to the requirement for topoisomerases in these stress conditions. Another interpretation could be that translation initiation is inhibited by the extensive base-pairing interactions between the coding gap and the SD of gene *60* mRNA. We know from our structural data that this interaction may form in buffer and at 20°C the

mRNA may not have enough kinetic energy to disrupt this interaction rapidly. Instead, the ribosome might have to wait until the SD becomes free on its own or with the help of slow cellular proteins.

Our hypothesis about higher bypassing efficiencies at lower temperatures could still be possible if gene *60* mRNA is translated cotranscriptionally, preventing any base-pairing between the SD and gene *60* mRNA. This hypothesis could be tested by measuring the rate of translation at 20°C of gene *60* mRNA with coupled transcription translation. Alternatively, we could use CAT SD gene *60* mRNA, a mutant which lacks the ability to interact with the coding gap.

Conclusions

We have probed the structure of gene *60* mRNA with Tb^{3+} and 1M7 SHAPE reagents and obtained models with each dataset. Overall, probing between the two reagents is very similar, as would be expected by their similar mechanism of action[117,118]. Major differences between the two reagents arise mainly at potential metal binding sites as determined by probing at lower concentrations of Tb^{3+}[118,119]. Gene *60* mRNA migrates as multiple populations through a native gel, but the minimal functional portion of gene *60* mRNA[28] migrates as a single band. The structure of this portion alone is remarkably similar to the 5' portion of the full-length gene *60* mRNA model (compare Figure IV.1 and IV.9), further supporting our predicted structures.

Both datasets match the Tb^{3+} model better than the 1M7 model, so we conclude that our Tb^{3+} model is the best estimate of the secondary structure of gene *60* mRNA in solution. In the Tb^{3+} model the coding gap is base-paired with the SD region. This interaction is not essential for translational bypassing[25] (data not shown). Furthermore, nucleotide G54 in the center of this potential 10 base-pair long stem is very strongly hit by both reagents which disfavors the presence of this structure. By forbidding this interaction while folding in RNAstructure, the SD pairs in the 3' end and the coding gap is left relatively unstructured (as proposed in our 1M7 model).

Lack of structure in the coding gap except for the 5' hairpin is consistent with high reactivity in this region from both reagents (Figure IV.3 C) Furthermore, plotting

reactivity data onto previously proposed structures for the coding gap[22,25,26] reveals that none are consistent with our experimental data (Figure IV.7), except for the relatively unstructured model proposed by Le et al.[26] (Figure IV.7 D). Based on our data, the only portion of the coding gap with significant structure is the hairpin at the 5' end. However, we cannot rule out the possibility that the interaction between the SD and the coding gap is present in our probing conditions and occludes other potential structures from forming in the coding gap. Alternatively, the 3' end of the coding gap could fold cotranslationally or interact with the ribosome itself, possibly during the pause generated by translational bypassing (Figure IV.12). Further experimentation is required to test these hypotheses.

Chapter V

Conclusions and Future Directions

Conclusions

Translational bypassing is a unique phenomenon in which the ribosome translates a single protein from a discontinuous ORF. Although first discovered a quarter of a century ago[25], the mechanism by which the ribosome accomplishes this feat is still a mystery. Several mutational and computational studies have been performed to suggest the untranslated region in the ORF folds into a compact structure that is simply passed over by the ribosome[19,22,25,27]. Many structural models have been proposed[22,25,26], but none have been tested experimentally by traditional RNA structure probing techniques.

In this thesis, the contributions of mRNA structure to the mechanism of translational bypassing were examined. Gene *60* mRNA was probed with two chemical modifiers of RNA, Tb^{3+} and the SHAPE reagent 1M7. Based on the chemical environment of each nucleotide as revealed by probing, a structural model of gene *60* mRNA was proposed (Figures IV.1 and IV.2). Overall, gene *60* mRNA has a normalized reactivity of 0.40 indicating that it is only moderately structured. This property was found in other mRNA molecules probed via SHAPE[125] and may be an important feature of mRNA to prevent significant delays during translation.

By contrast, mRNA often contains significant structure at translation regulation points[124]. As translational bypassing represents a regulatory site during translation of gene *60* mRNA[19,25,27,28], several RNA structural motifs have been proposed in the coding gap of gene *60* mRNA. Common to all hypothetical structures is a stable, tight-turning hairpin with a UUCG loop at the 5' end of the coding gap[22,25,26] whose stability and integrity is required for efficient bypassing[23,24,27,28]. The presence of this hairpin was confirmed by strong protection of these nucleotides from both probing reagents (average

reactivity < 0.20) as well as persistent peak compression due to incomplete denaturation of cDNA copies of this hairpin.

Several structures have been proposed for the 3' end of the coding gap, however, our data reveal this region is highly reactive to both probing reagents, with an average reactivity of ~0.50. This value is higher than typical of mRNA coding regions indicating there is little, if any, RNA structure in this region. The only previously proposed model consistent with our data contains a single, short (3 base-pairs) hairpin in the 3' end of the coding gap[26] (Figure IV.7 D).

Under the assumption that only structured mRNAs play a role in translation regulation and given the results of our structure probing experiments, it appears that the 5' hairpin in the coding gap is the only regulatory element required for translational bypassing. Yet the sequence in the 3' end of the coding gap is important for efficient translational bypassing[28] and the entire, exact coding gap sequence is always transmitted by homing[37]. These data suggest the 3' end of the coding gap may play a non-structural role in translational bypassing. Alternatively, this region could fold co-translationally, interact with the ribosome in some other manner to promote bypassing and/or recruit (a) cellular cofactor(s).

Potential longer-range interactions between the coding gap and the rest of the gene *60* mRNA were predicted in the experimentally generated structural models. Strikingly, the coding gap forms extensive base-pairing interactions with the SD of gene *60* mRNA in the Tb^{3+} model (Figure IV.1). The SD is hit more strongly by 1M7 than Tb^{3+}, therefore this interaction is not present in the 1M7 model, and the coding gap of gene *60* mRNA makes only minor interactions within itself (Figure IV.2). Disruption of the SD-coding gap interaction does not impact bypassing efficiency[25] (and data not shown) indicating that, while it is possible it could be present during our *in vitro* probing conditions, this interaction is not relevant for bypassing. Secondly, the highly protected loop nucleotides 141-146 are complementary to the take-off and landing codons (Figure IV.11). These loop nucleotides were mutated to synonymous codons that are not able to base-pair with the take-off and landing sites. This mutation also had no impact on translational bypassing[28] (Figure IV.11 D). Finally, the 3' end of gene *60* mRNA is not essential for translational bypassing[28]. The structure of a truncated version of gene *60*

mRNA was probed (Figure IV.9) and has a very similar structure to the 5' end of the model for full-length gene *60* mRNA, providing additional support for these structural features in the full-length models.

Taken together, our results imply that the coding gap is an independent, minimally structured element that does not obviously interact with any other portion of gene *60* mRNA to orchestrate translational bypassing. The autonomy of this region could be due to the fact that the coding gap is a foreign genetic element[33] and needs to carry out translational bypassing in whatever genetic context into which it is inserted. It is interesting to note that in addition to this RNA element, translational bypassing requires a specific nascent peptide sequence. The nascent peptide sequence is encoded by the additional amino acids added to the N-terminus of gp60 by the MobA insertion (Figures I.2 and I.4). Furthermore, the SD sequence of gene *60* mRNA and the region of protected loop nucleotides are also part of the inserted genetic element[33]. It is unlikely that these regions of complementarity to the coding gap are completely serendipitous, suggesting they may play alternative roles in homing rather than translational bypassing.

A second major accomplishment of this thesis was the establishment of an optimized *in vitro* translation assay. This flexible assay was used to confirm the identity of our two gp60 bands and determine the relative bypassing efficiency of several gene *60* mutants. Translation rates of truncated and full-length gp60 from non-bypassing templates were compared to those produced from wildtype gene *60* mRNA. These results reveal that bypassing causes a ~20 s delay in translation at 37°C.

Future Directions

The biochemical and structural characterization of translational bypassing in gene *60* mRNA described in this thesis provides a strong foundation for further elucidation of its mechanism. Several relatively straight-forward experiments should be performed to probe questions that arose during or could not be answered by this study.

Firstly, while our best estimate of the structure of gene *60* mRNA is the model generated by Tb^{3+} probing, it contains an extensive base-pairing interaction between the coding gap and the SD of gene *60* mRNA. This interaction is not necessary for bypassing[25] (data not shown) and must minimally be disrupted to initiate translation. It is

possible that these base pairs are present during our probing of naked gene *60* mRNA in buffer or are simply predicted because the folding algorithm searches for the lowest free energy structure. To determine whether base-pairing to the SD in buffer is perhaps occluding alternative structures in the coding gap region, the structure of CAT SD gene *60* mRNA should be probed. If the reactivity of nucleotides in the coding gap changes, it may support different structures in the 3' end of the coding gap. If reactivity remains the same, we can conclude the coding gap – SD interaction was an artifact of the folding algorithm. The latter case would support the conclusion that the coding gap is a foreign element that neither interacts with the other parts of gene *60* mRNA nor possesses much inherent structure.

The structure of gene *60* mRNA should also be probed under translation conditions to obtain a better understanding of its structure during bypassing. This could be accomplished by probing *in vivo* during expression of gp60. Cyclohexamide could be added at various times after induction of gp60 expression to stall the ribosome. Subsequently, DMS (and possibly Tb^{3+} or SHAPE reagents) could be added to modify gene *60* mRNA as it is bound to the ribosome during translation. Another method by which to study translational bypassing would be to stall ribosomes with cyclohexamide, purify ribosomal complexes and analyze where they are located along the mRNA by toeprinting. This experiment could help to differentiate between scanning and bypassing mechanisms. Positions on gene *60* mRNA that generate regulatory pauses during translation (and we suspect that a ~20 s pause is induced by translational bypassing at the coding gap at 37°C, Figure IV.12), would have a statistically amplified toeprint indicating the location on the mRNA at which the ribosome pauses. Assuming the ribosome is paused by the hairpin at the 5' edge of the coding gap, if the coding gap is folded up within the ribosome[27,30] we would expect to see a toeprint just 3' of the coding gap. If the ribosome is scanning the coding gap we would expect to obtain a toeprint within the coding gap itself. Alternative toeprints would also provide clues about the mechanism of translational bypassing.

Identical experiments could be carried out during translation of gene *60* mRNA in the *in vitro* translation assay developed in Chapter II. There are several advantages to studying bypassing in extracts as opposed to live cells. Translation of gene *60* mRNA

could be initiated in a "synchronized pulse" by adding pre-formed initiation complexes (Appendix C) directly to extracts. In these experiments, the structural and toeprinting information collected as a function of time should represent a more homogeneous population of gene *60* mRNA at different stages of translation and bypassing instead of averaging over all states the mRNA experiences during multiple rounds of translation. Additionally, if translation is too fast to obtain structural snapshots, reactions could be carried out at lower temperatures[96] without the need to consider cell viability. Translation could be paused at multiple time points during translation by cyclohexamide to isolate complexes for analysis. Alternatively, aliquots of a translation reaction could be probed directly with the SHAPE reagent benzoyl cyanide (BzCN). The half-life of this reagent is 0.25 s in aqueous solutions and thus generates ~1 s snapshots of RNA structure[114,115]. Finally, if degradation of gene *60* mRNA continues to hinder data acquisition during translation, these experiments could be carried out in the presence of an RNase inhibitor or optimized using the completely recombinant PURE translation kit.

Another approach to directly monitor the structure of gene *60* mRNA during bypassing would be to find a mechanism to stall the ribosome in the act of bypassing. Initial attempts to accomplish this by depleting amino acids from translation reactions in extracts (Figure II.13) were unsuccessful due to the presence of amino acid metabolizing enzymes in the extracts. This approach might be successful during translation with the PURE kit, which should not contain significant amounts of any proteins other than those required for translation. As the take-off Gly codon is the first use of this amino acid in gp60 (Figure I.4), translation reactions in the absence of Gly may stall the ribosome just as it is approaching the coding gap. Structure probing of this complex before and immediately after addition of Gly could reveal important structural changes in the coding gap required for translational bypassing. If such stalled complexes can be stably purified, cryo-EM studies may provide an additional dimension of structural information about this mysterious process.

The coding gap of gene *60* mRNA can also be studied by single-molecule fluorescence resonance energy transfer (FRET) experiments (Figure V.1). As the coding gap of gene *60* mRNA is a foreign element[33] and does not appear to have functional interactions with any other regions of the mRNA, it would be useful to study the structure

of this segment in isolation. With FRET pairs on either side of the coding gap (as depicted in Figure V.1), the structure and dynamics of this RNA could be measured under various conditions (+/- salts, Mg^{2+}, etc.). Our ensemble structural information suggests that the coding gap region is highly unstructured, thus we would expect low FRET values in buffer alone. However, single-molecule experiments would reveal any transient excursions into (a) folded state(s). Additionally, Ringquist et al. have previously shown by toeprinting that the entire coding gap can fold into a structure that is protected by the presence of the 30S ribosomal subunit[30]. Based on these results, addition of the 30S subunit to our fluorophore-labeled coding gap construct should generate a high FRET state. Furthermore, it is unknown whether or not mRNA can fold in the presence of both ribosomal subunits and/or during translation. FRET values of our coding gap mRNA in the presence of both 30S and 50S subunits should reveal whether or not there is room for extra mRNA in a 70S complex. These results could be confirmed biochemically by toeprinting. This experiment could be elaborated to observe the structure of the coding gap in actively translating ribosomes.

Finally, the role of truncated gp60 could be probed. The relative ratio of truncated to full-length gp60 is regulated by the efficiency of translational bypassing. Such types of protein ratio regulation by mRNA elements are a common mechanism employed by many viruses to increase the functional diversity of their proteomes[15,17]. This suggests there could be a function for truncated gp60 in endonuclease homing, translational bypassing or topoisomerase activity. To test whether or not it plays a role in translational bypassing, truncated gp60 could be purified and added back to *in vitro* translation assays at various concentrations. Changes in the efficiency of translational bypassing would reveal whether or not this protein plays a role in bypassing. Truncated gp60 is also primarily composed of amino acids added to the N-terminus of gp60 by the MobA insertion (Figure I.2 and I.4). Truncated gp60 contains many positively charged residues suggesting that it may be able to bind to RNA. Gel shift assays with radiolabeled gene *60* mRNA would reveal whether or not this protein binds RNA. If so, truncations of gene *60* mRNA should be made to isolate the minimal binding site. The impact of truncated gp60 on the structure of the coding gap could also be studied in the single-molecule assay described in Figure V.1.

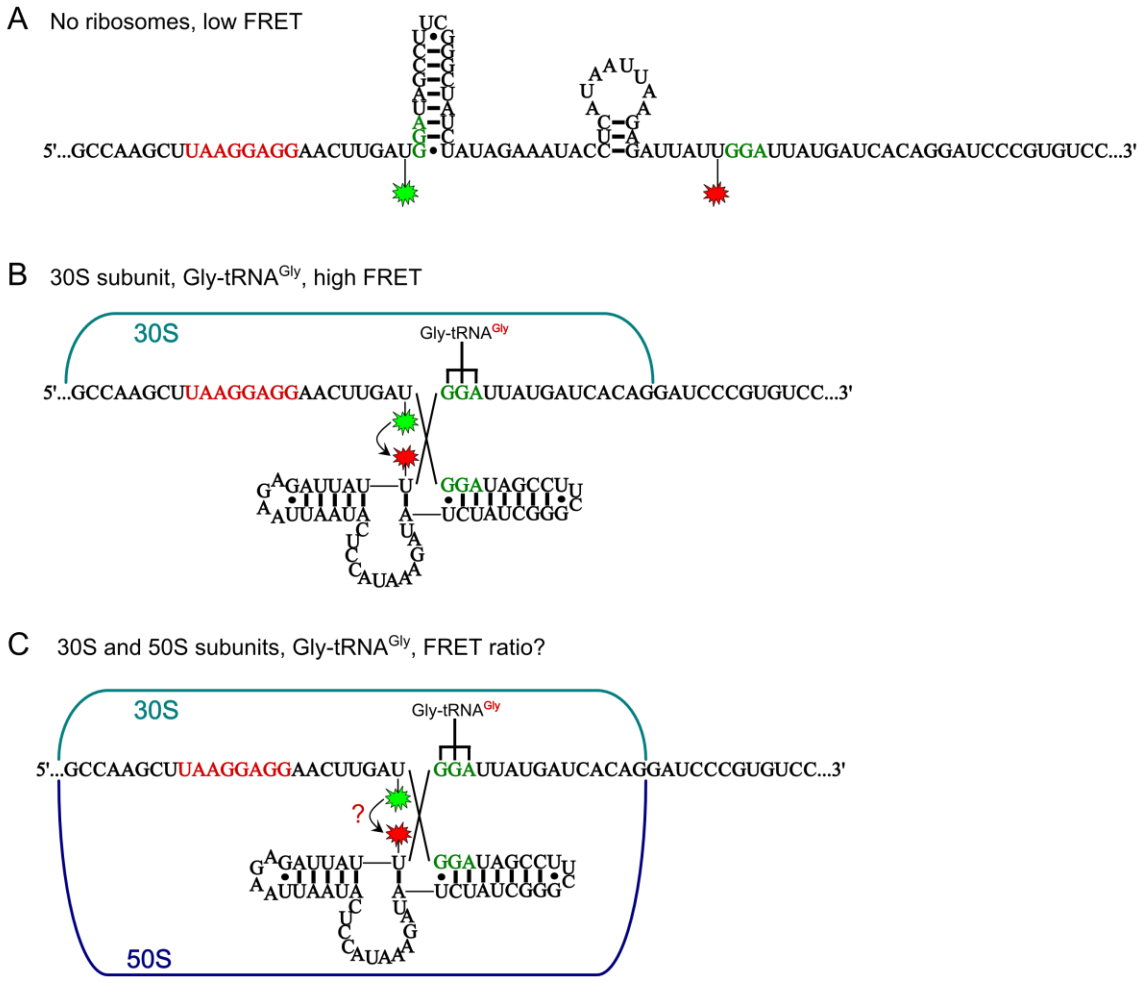


Figure V.1 Single-molecule analysis of coding gap structure. The coding gap, flanked by the take-off and landing Gly codons (green), is inserted downstream of a SD element (red). Fluorophores are added on either end of the coding gap region (red and green stars). (**A**) Based on ensemble structural data, the coding gap should be unstructured in buffer and the fluorophores will be out of FRET range. (**B**) The coding gap folds in the presence of $tRNA^{Gly}$ and the 30S ribosomal subunit[30], which brings the fluorophores close enough for FRET. (**C**) Addition of the 50S subunit will reveal whether or not there is room for structured mRNA in a 70S complex.

Much more work is required to fully elucidate the role played by RNA structure in translational bypassing. The biochemical and structural assays developed in this thesis provide a good foundation for further studies. Simple extensions of these assays (as described above) will yield answers to critical questions remaining about the mechanism of translational bypassing.

Appendix A: Cloning, Stocks and Transcriptions

Cloning Mutants of Gene *60*

A plasmid containing mutant Shine Dalgarno (mSD) gene *60* was obtained from John Hoerter (former graduate student, Walter lab). Flanking the wildtype gene *60* sequence are 61 nucleotides of its natural 5' UTR (aside from the mSD) and 41 nucleotides of its 3' UTR. This sequence was inserted downstream of a T7 RNA polymerase promoter in a pUC19 vector using XbaI and HindIII. Site directed mutagenesis was used to convert the mSD to a wildtype SD, generating a wildtype gene *60* plasmid, pR13. All further mutations of gene *60* were made by site directed mutagenesis from the plasmid pR13 (unless indicated).

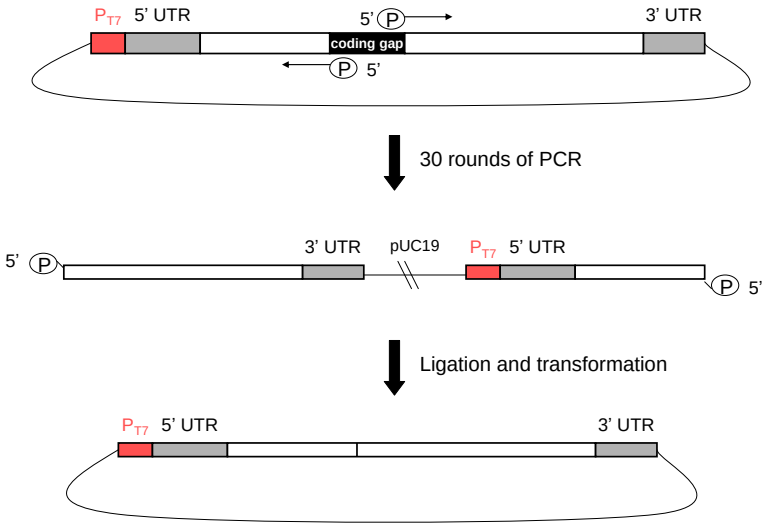


Figure A.1 Method for obtaining gene *60* delta gap (pR14) and other truncations of gene *60*. Two phosphorylated primers were used to amplify the portion of gene *60* from either side of the region to be deleted. PCR products of the correct molecular weight were gel purified, ligated, DpnI digested and transformed. Deletions were confirmed by sequencing.

Truncations were made with two phosphorylated primers as depicted in Figure A-1. A single primer was used to make point mutations. Both of these were performed following the site directed mutagenesis protocol obtained from Professor Roger Sunahara at the University of Michigan in Figure A-2. T4 DNA ligase (NEB #M0202) could be substituted for Taq DNA ligase (NEB #M0208) and Pfu Ultra (Stratagene #600380-52) was substituted for Pfu Turbo. Transformations were performed in JM109 cells (Promega #L1001).

- **Step 1: Design a primer(s) containing the desired mutation.**
  Ideally, primers should have at least 10 bases on either side of the mutation, and primers should terminate on one or more G/C bases
- **Step 2: Phosphorylate the primer(s)**
  Use the following reaction conditions (for 50µl):
  • 300pmol primer (6µM final)
  • 5µl 10mM ATP (1mM final)
  • 5µl 10X NEB Buffer 2 (1X final)
  • 1µl T4 Polynucleotide Kinase
  • add (39 – primer vol.) µl H2O to make 50µl total
  Incubate 30min at 37C
  Heat-inactivate 20min at 65C
- **Step 3: PCR target template with phosphorylated primers**
  Use the following reaction conditions (for 50µl):
  • 1nM (final) plasmid template
  • 2.1µl 6µM phosphorylated primer (250nM final)
  • 5µl 10mM dNTP mix (1mM final)
  • 2.5µl 10X pfuTurbo buffer (0.5X final)
  • 2.5µl 10X Taq DNA ligase buffer (0.5X final)
  • 1µl pfuTurbo (2.5U)
  • 0.5µl Taq DNA ligase (20U)
  • add (36.4 – plasmid vol.) µl H2O to make 50µl
  Use the following cycle parameters:
  • 95C 1min
  • 95C 1min |
  • 55C 1min | 30X
  • 65C 2min/kb plasmid length |
  • 4C hold
- **Step 4: DpnI-treat the PCR**
  Use the following reaction conditions:
  • 5µl PCR reaction
  • 2µl 10X NEB Buffer 4
  • 17µl H2O
  • 1µl DpnI
  Incubate 1hr at 37C
- **Step 5: Transform mutant plasmids**
  Transform 100µl competent E. coli (DH5α) cells with 5µl of the DpnI digest, as per normal protocol
- **Step 6: Pick colonies, miniprep, and sequence**

Figure A.2 Sunahara Lab Protocol for Site-Directed Mutagenesis.

Tables of Stocks

Table A.1 Ribosome Project Plasmids

Stock #	Parent Plasmid	Description of Modifications	Source
pR1	pGEM-4z		Promega
pR2	pAMB-CAT		Ambion Active Pro translation kit
pR3	pUC19		NEB
pR4	pUC19	mSD gene 32	John Hoerter
pR5	pUC19	mSD gene 60	John Hoerter
pR6	pUC19	mSD gene 60 point mut (in coding gap)	John Hoerter
pR7	pRNA		Phil Cunningham, Wayne State
pR8	pRSET		Phil Cunningham, Wayne State
pR9	pRSET	13b (FspI)	primer 13
pR10	pRSET	14b (EcoRV)	primer 14
pR11	pDHFR		NEB PURExpress Protein Synthesis Kit
pR12	pUC19	gene 32	primer 17, generates wt SD
pR13	pUC19	gene 60	primer 16, generates wtSD
pR14	pUC19	gene 60 Δgap	primers 22 and 23
pR15	pAMB-CAT	SDM 13	primer 13, adds site after T7 terminator hairpin
pR16	pUC19	mSD gene 60Δgap	primers 22 and 23
pR17	pIF2	Maxi-prep purified plasmid from bacteria stock	Green Lab's Bacteria Stock - Purified by JRA
pR18	pUC19	gene 60 Ser159Cys	primers 40 and 41, for in vitro tln with 35S-Cys
pR19	pUC19	gene 60 Ser8Cys	primers 42 and 43, for in vitro tln with 35S-Cys
pR20	pUC19	gene 60 Δ gap Ser159Cys	primers 40 and 41, for in vitro tln with 35S-Cys
pR21	pUC19	gene 60 Δ gap Ser8Cys	primers 42 and 43, for in vitro tln with 35S-Cys
pR22	pUC19	gene 60 to first GGA	primers 26 and 28, for in vitro tln with 35S-Cys
pR23	pUC19	gene 60 to first GGA - Cys	primers 26 and 29, for in vitro tln with 35S-Cys
pR24	pUC19	gene 60 thru gap + 6nts	primers 27 and 28, for in vitro tln with 35S-Cys
pR25	pUC19	gene 60 thru gap + 6 nts + Cys	primers 27 and 29, for in vitro tln with 35S-Cys
pR26	pUC19	gene 60 to first GGA - HindIII	primers 26 and 30, for in vitro tln with 35S-Cys
pR27	pUC19	gene 60 thru gap - HindIII	primers 27 and 30, for in vitro tln with 35S-Cys
pR28	pUC19	gene 60 thru gap + 21 nts	primers 28 and 49
pR29	pUC19	gene 60 thru gap + 21 nts-Cys	primers 29 and 49
pR30	pUC19	gene 60 thru gap + 45 nts	primers 28 and 50
pR31	pUC19	gene 60 thru gap + 45 nts-Cys	primers 29 and 50
pR32	pUC19	gene 60 thru gap + 72 nts	primers 28 and 51
pR33	pUC19	gene 60 thru gap + 72 nts-Cys	primers 29 and 51
pR34	pUC19	gene 60_5' - 27 nts	primers 52 and 53
pR35	pUC19	gene 60_5' - 42 nts	primers 52 and 54
pR36	pUC19	gene 60_5' - 48 nts	primers 52 and 55
pR37	pUC19	gene 60, wt SD -> CAT SD	primers 70 and 71, pR13 template
pR38	pUC19	gene 60_5' - 103 nts	random mutation found during sequencing
pR39	pUC19	gene 60 Δgap, wt SD -> CAT SD	primers 70 and 71, pR14 template
pR40	pUC19	gene 60 to first GGA, wt SD -> CAT SD	primers 70 and 71, pR22 template
pR41	pUC19	gene 60 Cys in gap, Ser159Cys	primer 73, pR18 template
pR42	pUC19	gene 60 shift frame BglII	primers 75 and 76, template pR13
pR43	pUC19	gene 60 shift frame	primers 75 and 77, template pR13
pR44	pUC19	gene 60 anti CAT SD in coding gap	primer 72, template pR13
pR45	pUC19	gene 60, wt SD->CAT SD, CAT aSD in coding gap	primer 72, template pR37
pR46	pUC19	gene 60 loop mut	primer 85, template pR13
pR47	pUC19	gene 60 delta gap loop mut	primer 85, template pR14
pR48	pUC19	gene 60 to first GGA loop mut	primer 85, template pR22
pR49	pUC19	gene 60, wt SD -> CAT SD, loop mut	primer 85, template pR37
pR50	pUC19	gene 60 thru gap + 6 nts, wt SD -> CAT SD	primers 70 and 71, pR24 template

Table A.2 Ribosome Project Primers

Primer #	Primer Name	Primer Sequence (5' - 3')	Purpose
1	EmGFP rev KpnI	GTCTGGTACCTTCTGTTACTTGTACAGC	Moving GFP from pRNA to pGEM-4z
2	EmGFP for mSD XbaI	GTAGTCTAGATATCCCTCCGCAAATGG	Moving GFP from pRNA to pGEM-4z
3	EmGFP for wtSD XbaI	GTAGTCTAGATAGGAGGTCCGCAAATGGAGAA	Moving GFP from pRNA to pGEM-4z
4	EmGFP rev KpnI	GTCTGGTACCTTCTGTTACTTGTACAGCTCGTCC	Moving GFP from pRNA to pGEM-4z
5	EmGFP for mSD XbaI	GTAGTCTAGATATCCCTCCGCAAATGGAGAAAAAAATCACTG	Moving GFP from pRNA to pGEM-4z
6	EmGFP for wtSD XbaI	GTAGTCTAGATAGGAGGTCCGCAAATGGAGAAAAAAATCACTGTG	Moving GFP from pRNA to pGEM-4z
7	F_emGFP	TAA TAC GAC TCA CTA TAG GCT CAG CAT CCC TCC GCA AAT GGA GAA	Cheng-Yen Ordered, EmGFP PCR template
8	R_emGFP	TTA CTT GTA CAG CTC GTC CAT GC	Cheng-Yen Ordered, EmGFP PCR template
9	EmGFPT7 w/SphI to pGEM fwd	GCA TGC ATG GTG AGC AAG GGC GA	Matt, for moving EmGFP from pRSET to pGEM-4z
10	EmGFPT7 w/XbaI to pGEM rev	TCT AGA CGC TAT TAC GCC AGA TCC	Matt, for moving EmGFP from pRSET to pGEM-4z
11	EmGFP_T7_pro_rev	GAA CTT CAG GGT CAG CTT GCC	For sequencing T7 promoter of pRSET-EmGFP
12	EmGFP_T7_term_rev	GGA AGG GAA GAA AGC GAA AGG AGC	For sequencing T7 terminator of pRSET-EmGFP
13	T7_term_FspI	GAG GGG TTT TTT GCG CAA AGG AGG AAC TAT ATC C	add an FspI site after T7 terminator (pRSET)
14	T7_termlong_EcoRV	GGA ACT ATA TCC GGA TAT CGC GTA ATA GCG AAG	add an EcoRV site after longer T7 terminator (pRSET)
15	T7_term_rev_seq	CCT CAA GAC CCG TTT AGA GGC C	Sequencing in the reverse direction from a T7 terminator
16	60_mSD_to_wtSD	CCT GAT GAA AAG TTC TAT GAG GTG TAT AAT GAA ATT TG	Convert John's pUC19 mSD gene 60 to wt SD
17	32_mSD_to_wtSD	CCT GAT GAA AAG TTC TAT GAG GTG TAT AAT GTT TAA ACG	Convert John's pUC19 mSD gene 32 to wt SD
18	pUC19_BamHI_to_EcoRV	GCT CGG TAC CCG GAT ATC CTC TAG ATA ATA CG	Converts BamHI in John's pUC19 gene 60/32 plasmids to an EcoRV
19	5_Gly_to_BamHI	CAC GAT GGA TCC CCT TCG GGC TAT C	Converts Gly codon at 5' end of coding gap to a BamHI site
20	3_Gly_to_BamHI	GAG ATT ATT GGA TCC GGT TCT ATT TAT CCT TCT C	Converts Gly codon at 3' end of coding gap to a BamHI site
21	BamHI_to_glyleu	CGC AGA TCA CGA TGG ATT AGG TTC TAT TTA TCC TTC TC	Converts BamHI site to coding sequence without a gap
22	gene60_postgap	GAT TAG GTT CTA TTT ATC CTT CTC TGC TCG	PCR pUC19 gene 60 without gap
23	gene60_pregap	CAT CGT GAT CTG CGT CTG TC	PCR pUC19 gene 60 without gap
24	CAT_mid_for	CGC AAG ATG TGG CGT GTT ACG G	To sequence 3' end of pAMB-CAT
25	CAT_mid_rev	CCG TAA CAC GCC ACA TCT TGC G	To sequence 5' end of pAMB-CAT
26	60_int_stop	TCC ATC GTG ATC TGC GTC TGT C	rev primer for gene 60 up to internal stop codon
27	60_thru_gap	ACC TAA TCC AAT AAT CTC TTA ATT ATG AGG	rev primer for gene 60 including coding gap and 6 nt 3' of landing site
28	60_3'UTR	TAG TTT ACT TTA CCA CAA GGA TGT GG	for primer stop codon of full-length gene 60 and gene 60 3' UTR
29	60_Cys_3'UTR	TGT TAG TTT ACT TTA CCA CAA GGA TGT GG	for primer Cys codon then stop of full-length gene 60 and 3' UTR
30	HindIII_pUC19	AAG CTT GGC GTA ATC ATG GTC ATA GC	for primer with HindIII site into pUC19 (removes gene 60 3' UTR)
31	HindIII_5-FAM_RT	FAM-GCC CAA TTA ATT ATA CCA CAT CCT TGT GG	5' FAM-labeled for RT of gene 60 mRNA HindIII
32	BsrBI_5-FAM_RT	FAM-CGG ATA ACA ATT TCA CAC AGG AAA CAG C	5' FAM-labeled for RT of gene 60 mRNA BsrBI
33	BsrBI_5-TAMRA_RT	TAMRA-CGG ATA ACA ATT TCA CAC AGG AAA CAG C	5' TAMRA-labeled for RT of gene 60 mRNA BsrBI
34	BsrBI_5-ROX_RT	ROX-CGG ATA ACA ATT TCA CAC AGG AAA CAG C	5' ROX-labeled for RT of gene 60 mRNA BsrBI
35	BsrBI_5-HEX_RT	HEX-CGG ATA ACA ATT TCA CAC AGG AAA CAG C	5' HEX-labeled for RT of gene 60 mRNA BsrBI
36	Gene 60 3UTR Label 5p 3	NH2-CTTTACCACAAGGATGTG-NH2	5' and 3' free amine for fluorophore labeling, binds gene 60 3' UTR
37	Gene 60 3UTR Label 5p A	NH2-CTTTACCACAAGGATGTG	5' free amine for fluorophore labeling, binds gene 60 3' UTR
38	3-biotin to 60 5-UTR	CCGCGTTTAAATTCATTAAATGCTCC-biotin	Complementary to gene 60 5' UTR,attach mRNA to a slide
39	BsrBI_5-NED_RT	NED-CGG ATA ACA ATT TCA CAC AGG AAA CAG C	5' NED-labeled for RT of gene 60 mRNA BsrBI
40	Ser159Cys_for	CGTAAAGAATGGATGTGCCAGTAGTTTACTTTACC	SDM - Converts S159 to Cys
41	Ser159Cys_rev	GGTAAAGTAAACTACTGGCACATCCATTCTTTACG	SDM - Converts S159 to Cys
42	Ser8Cys_for	GTAAAAATTGATTGTTCTAGCGTTGATATG	SDM - Converts S8 to Cys

43	Ser8Cys_rev	CATATCAACGCTAGAACAATCAATTTTTAC	SDM - Converts S8 to Cys
44	44_610-632 5,3 amine	GCC CAA TTA ATT ATA CCA CAT CC	binds in gene 60 3' UTR, 5' and 3' amines for labeling with Cy5
45	mid-60_5-FAM_RT	FAM-CCG ACC TGA GCG ATG ATT ACA G	for structure probing of 5' end of gene 60
46	mid-60_5-HEX_RT	HEX-CCG ACC TGA GCG ATG ATT ACA G	for structure probing of 5' end of gene 60
47	mid-60_5-NED_RT	NED-CCG ACC TGA GCG ATG ATT ACA G	for structure probing of 5' end of gene 60
48	mid-60_5-ROX_RT	ROX-CCG ACC TGA GCG ATG ATT ACA G	for structure probing of 5' end of gene 60
49	60 thru gap + 21 nts	AGA AGG ATA AAT AGA ACC TAA TCC AAT AAT CTC	with primers 28 or 29, extend gene 60 21 nts past coding gap
50	60 thru gap + 45 nts	CCA ATT ACT AAA AAA TCC GAG CAG	with primers 28 or 29, extend gene 60 45 nts past coding gap
51	60 thru gap + 72 nts	AAT TCG TCC TTG CTC AAA CAA TTC	with primers 28 or 29, extend gene 60 72 nts past coding gap
52	60 start rev	CAT TAT ACA CCT CAT AGA ACT TTT CAT CAG	with primers 53-55, make truncations in 5' region gene 60
53	60_5' – 27 nts	GTT GAT ATG AAA AAA TAT AAA TTG CAG AAC	with primer 52, remove first 27 nts gene 60 after start
54	60_5' – 42 nts	TAT AAA TTG CAG AAC AAT GTT CGT CG	with primer 52, remove first 42 nts gene 60 after start
55	60_5' – 48 nts	TTG CAG AAC AAT GTT CGT CGT TC	with primer 52, remove first 48 nts gene 60 after start
56	60_SD –> CAT_SD	GATGAAAAGTTC TAA GAA GGA GAT ATACATATGAAA TT TG	SDM to replace gene 60's SD with that of CAT
57	cx6x_1_HEX	HEX - GTT GTG GTT TGT CCA AAC TCA TC	structure probing of cx6c mRNA
58	cx6x_2_HEX	HEX - CAT CAC TTG CAC GTA GAT AAG C	structure probing of cx6c mRNA
59	cx6x_3_HEX	HEX - CAG CGT GAA CTA TTG CTT TGA TC	structure probing of cx6c mRNA
60	HMGA2_1_HEX	HEX - GCC ACT GTG CTG GAT ATC TGC	structure probing of HMGA2 mRNA
61	HMGA2_2_HEX	HEX - GTA GAG AAG TTA GAT GTC CAG TC	structure probing of HMGA2 mRNA
62	HMGA2_3_HEX	HEX - GTA AGG CCC ATC CGA AGA CCC G	structure probing of HMGA2 mRNA
63	HMGA2_4_HEX	HEX - CCT TTT TCA TTG CAT TAG GC	structure probing of HMGA2 mRNA
64	HMGA2_5_HEX	HEX - CTA GCA GGA AAG CTG CCA CAA GCA T	structure probing of HMGA2 mRNA
65	HMGA2_6_HEX	HEX - GGT ACA CAA TGA AGA ACA GGG A	structure probing of HMGA2 mRNA
66	HMGA2_7_HEX	HEX - GGC TAA GTG GCT TTT GGC TGT G	structure probing of HMGA2 mRNA
67	HMGA2_8_HEX	HEX - GGT TTT CTT CAA CCA CAC ACC CT	structure probing of HMGA2 mRNA
68	HMGA2_9_HEX	HEX - GCG GAC TCT TGC GAG GAT GTC	structure probing of HMGA2 mRNA
69	HMGA2_10_HEX	HEX - GAA TCC GCC CAG CAC CTT TC	structure probing of HMGA2 mRNA
70	60 SD – CAT SD for	GAG ATA TAC ATA TGA AAT TTG TAA AAA TTG ATT CTT C	SDM to convert gene 60 SD to CAT SD
71	60 SD – CAT SD rev	CTT CTT AGA ACT TTT CAT CAG GAA TCC	SDM to convert gene 60 SD to CAT SD
72	aSD_CAT_cg	GGC TAT CTA TAG AAA TTT CCT TCT ATT AAG AGA TTA TTG	gene 60 coding gap complementary to the CAT SD
73	C_scan_cg	CCT TCG GGC TAT GTA TAG AAA TAC C	Cys residue in-frame after potential GGG landing site in gap
74	60 + BglII	TCC TCT TCA AAG ATC TAT GCG AAT GTC	BglII site into the 5' region of gene 60
75	nasc_pep_-1	CTA GCG TTG ATA TGA AAA ATA TAA ATT GCA G	Change the reading frame of the nascent peptide
76	nasc_pep_+1_BglII	TCC TCT TCA AAA GAT CTA TGC GAA TGT C	Change the reading frame of the nascent peptide
77	nasc_pep_+1_wt	TCC TCT TCA AAT GAA CTA TGC GAA TGT C	Change the reading frame of the nascent peptide
78	NcoI_6H_N_gp60	CCG ACC ATG GAT GAA ATT TGT AAA AAT TGA TTC	N-terminal 6His-tag protein fragments
79	6H_N_60fl_HindIII	GCA TAA GCT TCT ACT GGC TCA TCC	N-terminal 6His-tag protein fragments
80	6H_N_60tr_HindIII	CAT AAG CTT CTA TCC ATC GTG ATC TG	N-terminal 6His-tag protein fragments
81	6H_N_60nc_HindIII	CAT AAG CTT CTA CAT TGA AGA GGA TTT AAT AG	N-terminal 6His-tag protein fragments
82	NdeI_6H_C_60fl	CCG ACA TAT GAA ATT TGT AAA AAT TGA TTC	C-terminal 6His-tag protein fragments
83	NdeI_6H_C_60cnsv	CCG ACA TAT GAA CTA TGC GAA TGT C	C-terminal 6His-tag protein fragments
84	6H_C_60gp_XhoI	CTA ACT CGA GCT GGC TCA TCC ATT C	C-terminal 6His-tag protein fragments
85	60 loop mut	CGT CGT TCT ATT AAG AGT AGC TCA ATG AAC TAT G	mutate protected loop nt 91-97 in gene 60 mRNA

Table A.3 Ribosome Project Bacterial Stocks

Stock #	Parent Strain	Parent Plasmid	Description of Modifications
B0	MRE600		none
B1	JM109	pGEM-4z	
B2	JM109	pAMB-CAT	
B3	JM109	pUC19	
B4	JM109	pUC19	mSD gene 32
B5	JM109	pUC19	mSD gene 60
B6	JM109	pUC19	mSD gene 60 pt mut
B7	JM109	pRNA	
B8	JM109	pRSET	
B9	JM109	pRSET	13b (FspI)
B10	JM109	pRSET	14b (EcoRV)
B11	JM109	pUC19	wtSD gene 32
B12	JM109	pUC19	wtSD gene 60
B13	JM109	pUC19	wtSD gene 60 pt mut
B14	JM109	pUC19	wtSD gene 60Δgap
B15	JM109	pAMB-CAT	13 (FspI)
B16	JM109	pUC19	mSD gene 60Δgap
B17	?		alaS (Ala synthetase)
B18	?		argS (Arg synthetase)
B19	?		asnS (Asn synthetase)
B20	?		aspS (Asp synthetase)
B21	?		cysS (Cys synthetase)
B22	?		glnS (Gln synthetase)
B23	?		gltX (Glu synthetase)
B24	?		glyS (Gly synthetase)
B25	?		hisS (His synthetase)
B26	?		ileS (Ile synthetase)
B27	?		leuS (Leu synthetase)
B28	?		lysS (Lys synthetase)
B29	?		metG (Met synthetase)
B30	?		pheS (Phe synthetase)
B31	?		proS (Pro synthetase)
B32	?		serS (Ser synthetase)
B33	?		thrS (Thr synthetase)
B34	?		trpS (Trp synthetase)
B35	?		tyrS (Tyr synthetase)
B36	?		valS (Val synthetase)
B37	?		IF1
B38	?		IF2
B39	?		IF3
B40	?		EF-Tu
B41	?		EF-G
B42	?		RF1
B43	?		RF2
B44	?		RF3
B45	BL21 pLys?		*EF-Ts
B46	HB101	pUC19	pR18 gene 60 Ser159Cys
B47	HB101	pUC19	pR19 gene 60 Ser8Cys
B48	HB101	pUC19	pR20 gene 60 Δ gap Ser159Cys
B49	HB101	pUC19	pR21 gene 60 Δ gap Ser8Cys
B50	HB101	pUC19	pR22 gene 60 to first GGA - Stop- 3' UTR - HindIII
B51	HB101	pUC19	pR23 gene 60 to first GGA - Cys - Stop - 3' UTR - HindIII
B52	HB101	pUC19	pR24 gene 60 thru gap - Stop - 3' UTR - HindIII
B53	HB101	pUC19	pR25 gene 60 thru gap - Cys - Stop - 3' UTR - HindIII
B54	HB101	pUC19	pR26 gene 60 to first GGA - HindIII
B55	JM109	pUC19	pR27 gene 60 thru gap - HindIII
B56	JM109	pEF6	pEF6-HMGA2-FL
B57	JM109	pRL	pRL-cx6x
B58	JM109	pUC19	pR28 gene 60 thru gap + 21 nts
B59	JM109	pUC19	pR29 gene 60 thru gap + 21 nts-Cys
B60	JM109	pUC19	pR30 gene 60 thru gap + 45 nts
B61	JM109	pUC19	pR31 gene 60 thru gap + 45 nts-Cys
B62	JM109	pUC19	pR32 gene 60 thru gap + 72 nts
B63	JM109	pUC19	pR33 gene 60 thru gap + 72 nts-Cys
B64	JM109	pUC19	pR34 gene 60_5' - 27 nts
B65	JM109	pUC19	pR35 gene 60_5' - 42 nts
B66	JM109	pUC19	pR36 gene 60_5' - 48 nts

B67	JM109	pUC19	pR37 gene 60 wt SD -> CAT SD
B79	JM109	pUC19	pR39 gene 60 delta gap, wt SD - CAT SD
B80	JM109	pUC19	pR40 gene 60 to first GGA, wt SD - CAT SD
B81	JM109	pUC19	pR50 gene 60 thru gap + 6 nts, wt SD - CAT SD

Transcription Reactions

Plasmid of the desired mutant was digested with BsrBI, which cleaves the pUC19 backbone 51 nucleotides 3' of the HindIII cloning site, which is the end of wildtype gene *60*'s natural 3' UTR. Digested plasmids could be used either with or without phenol/chloroform extraction and ethanol precipitation for transcription reactions. Transcription reactions consisted of 50 nM digested template, 40 mM Tris-HCl pH 8.0, 25 mM $MgCl_2$, 100 mM NaCl, 0.01% Triton-X, 5 mM DTT, 1 mM spermidine, 12 mM ATP, 12 mM CTP, 12 mM GTP, 12 mM UTP, 0.01 U/μL pyrophosphatase and 0.07 mg/mL T7 RNA polymerase. Transcription was carried out at 37°C for 1 hr before an equal volume of formamide loading buffer (90% formamide in TBE) was added to quench the reaction. Transcripts were resolved by denaturing PAGE and correctly sized product bands were detected by UV shadowing. RNA bands were cut from the gel using sterile razor blades, crushed through a syringe and incubated in crush 'n soak buffer (500 mM NH_4OAc, 0.1% SDS, 100 μM EDTA) overnight at 4°C. Eluate was extracted with phenol and chloroform and ethanol precipitated. Purified transcripts were resuspended in TE pH 7.0 and stored at -20°C.

All mutants of gene *60* mRNA contained 5' and 3' extensions to allow for proper translation and as spacers to obtain structure probing information for the entire open reading frame. The 5' extension consisted of 61 nucleotides of gene *60*'s 5' UTR. The 3' extension included 41 nucleotides of gene *60*'s 3' UTR and another 51 nucleotides of the vector backbone of pUC19 due to cleavage with BsrBI. The 28-nucleotide DNA primer used for RT was complementary to the very 3' end of each transcript.

Table A.4 Sequence of mRNAs

pR13 – gene 60 BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA

CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAAUUGGCCAGAA
UUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUCGCUCAGGUCGGU
AAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCCAAAGAUGCUCUACCU
AAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAAUCUGAAUAUCGUGAG
AUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAACUGGAAAGAGCUUUUU
GAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUGAGCCAGUAGUUUACU
UUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCAUGGUCAUAGC
UGUUUCCUGUGUGAAAUUGUUAUCCG

pR14 – gene 60 Δgap BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAA
UUGGCCAGAAUUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUCGC
UCAGGUCGGUAAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCCAAAGA
UGCUCUACCUAAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAAUCUGA
AUAUCGUGAGAUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAACUGGAA
AGAGCUUUUUGAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUGAGCCA
GUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCA
UGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR18 – gene 60 S159C BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAAUUGGCCAGAA
UUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUCGCUCAGGUCGGU
AAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCCAAAGAUGCUCUACCU
AAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAAUCUGAAUAUCGUGAG
AUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAACUGGAAAGAGCUUUUU
GAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUGUGCCAGUAGUUUACU
UUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCAUGGUCAUAGC
UGUUUCCUGUGUGAAAUUGUUAUCCG

pR19 – gene 60 S8C BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU

AAUGAAAUUUGUAAAAAUUGAUUGUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAAUUGGCCAGAA
UUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUCGCUCAGGUCGGU
AAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCCAAAGAUGCUCUACCU
AAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAAUCUGAAUAUCGUGAG
AUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAACUGGAAAGAGCUUUUU
GAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUGAGCCAGUAGUUUACU
UUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCAUGGUCAUAGC
UGUUUCCUGUGUGAAAUUGUUAUCCG

pR20 – gene 60 Δgap S159C BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAA
UUGGCCAGAAUUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUCGC
UCAGGUCGGUAAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCCAAAGA
UGCUCUACCUAAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAAUCUGA
AUAUCGUGAGAUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAACUGGAA
AGAGCUUUUUGAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUGUGCCA
GUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCA
UGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR21 – gene 60 Δgap S8C BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUGUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAA
UUGGCCAGAAUUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUCGC
UCAGGUCGGUAAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCCAAAGA
UGCUCUACCUAAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAAUCUGA
AUAUCGUGAGAUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAACUGGAA
AGAGCUUUUUGAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUGAGCCA
GUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCA
UGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR22 – gene 60 to first GGA BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU

AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGC
AAAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR23 – gene 60 to first GGA + Cys BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUGUUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUG
GGCAAAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR24 – gene 60 thru gap BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAA
GCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR25 – gene 60 thru gap + Cys BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUGUUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCA
AAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR26 – gene 60 to GGA, no 3'UTR BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAA
AUUGUUAUCCG

pR27 – gene 60 thru gap, no 3'UTR BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUG

UUAUCCG

pR28 – gene 60 thru gap + 21 nts BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUUAGUUUACUUUACCACAAGGAUGUGGUAUA
AUUAAUUGGGCAAAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUU
AUCCG

pR29 – gene 60 thru gap + 21 nts + Cys BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUUGUUAGUUUACUUUACCACAAGGAUGUGGU
AUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUU
GUUAUCCG

pR30 – gene 60 thru gap + 45 nts BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAAUUGGUAGUUU
ACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCAUGGUCAU
AGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR31 – gene 60 thru gap + 45 nts + Cys BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAAUUGGUGUUAG
UUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCAUGGU
CAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR32 – gene 60 thru gap + 72 nts BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAAUUGGCCAGAA
UUGUUUGAGCAAGGACGAAUUUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGG
GCAAAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR33 – gene 60 thru gap + 72 nts + Cys BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAAUUGGCCAGAA
UUGUUUGAGCAAGGACGAAUUUGUUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAU
UGGGCAAAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR34 – gene 60 5’ – 27 nts BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGGUUGAUAUGAAAAAAUAUAAAUUGCAGAACAAUGUUCGUCGUUCUAUUAAAUCCUC
UUCAAUGAACUAUGCGAAUGUCGCUAUUAUGACAGACGCAGAUCACGAUGGAUAGCCUUC
GGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAGAUUAUUGGAUUAGGUUCUAUUUAUCCU
UCUCUGCUCGGAUUUUUUAGUAAUUGGCCAGAAUUGUUUGAGCAAGGACGAAUUCGCUUU
GUCAAAACUCCUGUAAUCAUCGCUCAGGUCGGUAAAAAACAAGAAUGGUUUUAUACAGUC
GCUGAAUAUGAGAGUGCCAAAGAUGCUCUACCUAAACAUAGCAUCCGUUAUAUUAAGGGA
CUUGGCUCUUUGGAAAAAUCUGAAUAUCGUGAGAUGAUUCAAAACCCAGUAUAUGAUGUU
GUUAAACUUCCUGAGAACUGGAAAGAGCUUUUUGAAAUGCUCAUGGGAGAUAAUGCUGAC
CUUCGUAAAGAAUGGAUGAGCCAGUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAU
UGGGCAAAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR35 – gene 60 5’ – 42 nts BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGUAUAAAUUGCAGAACAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGC
GAAUGUCGCUAUUAUGACAGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAA
AUACCUCAUAAUUAAGAGAUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUU
UUUAGUAAUUGGCCAGAAUUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUA
AUCAUCGCUCAGGUCGGUAAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGU
GCCAAAGAUGCUCUACCUAAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAA
AAAUCUGAAUAUCGUGAGAUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAG
AACUGGAAAGAGCUUUUUGAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGG

AUGAGCCAGUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGG
CGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR36 – gene 60 5' – 48 nts BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGUUGCAGAACAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGU
CGCUAUUAUGACAGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCU
CAUAAUUAAGAGAUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGU
AAUUGGCCAGAAUUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUC
GCUCAGGUCGGUAAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCCAAA
GAUGCUCUACCUAAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAAUCU
GAAUAUCGUGAGAUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAACUGG
AAAGAGCUUUUUGAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUGAGC
CAGUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAU
CAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR37 – gene 60 SD to CAT SD BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAAGAAGGAGAU
AUACAUAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUG
CAGAACAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUU
AUGACAGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUU
AAGAGAUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAAUUGGC
CAGAAUUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUCGCUCAGG
UCGGUAAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCCAAAGAUGCUC
UACCUAAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAAUCUGAAUAUC
GUGAGAUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAACUGGAAAGAGC
UUUUUGAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUGAGCCAGUAGU
UUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAaagcuuggcguaaucauggucauagcuguuuc
cugugugaaauuguuauccg

pR38 – gene 60 5' – 103 nts BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGUCGCUAUUAUGACAGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAA
UACCUCAUAAUUAAGAGAUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUU
UUAGUAAUUGGCCAGAAUUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAA
UCAUCGCUCAGGUCGGUAAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUG
CCAAAGAUGCUCUACCUAAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAA
AAUCUGAAUAUCGUGAGAUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGA
ACUGGAAAGAGCUUUUUGAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGA

UGAGCCAGUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGC
GUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR39 – gene 60 Δgap, wt SD to CAT SD BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAAGAAGGAGAU
AUACAUAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUG
CAGAACAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUU
AUGACAGACGCAGAUCACGAUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUU
AGUAAUUGGCCAGAAUUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUC
AUCGCUCAGGUCGGUAAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCC
AAAGAUGCUCUACCUAAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAA
UCUGAAUAUCGUGAGAUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAAC
UGGAAAGAGCUUUUUGAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUG
AGCCAGUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGU
AAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR40 – gene 60 to first GGA, wt SD to CAT SD BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAAGAAGGAGAU
AUACAUAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUG
CAGAACAAUGUUCGUCGUUCUAUUAAAUCCUCUUCAAUGAACUAUGCGAAUGUCGCUAUU
AUGACAGACGCAGAUCACGAUGGAUAGUUUACUUUACCACAAGGAUGUGGUAUAAUUAAU
UGGGCAAAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCG

pR46 – gene 60 loop mut BsrBI mRNA

GGAGCAUUUAAUGAAUUUAAACGCGGUUGGAUUCCUGAUGAAAAGUUCUAUGAGGUGUAU
AAUGAAAUUUGUAAAAAUUGAUUCUUCUAGCGUUGAUAUGAAAAAAUAUAAAUUGCAGAA
CAAUGUUCGUCGUUCUAUUAAGAGUAGCUCAAUGAACUAUGCGAAUGUCGCUAUUAUGAC
AGACGCAGAUCACGAUGGAUAGCCUUCGGGCUAUCUAUAGAAAUACCUCAUAAUUAAGAG
AUUAUUGGAUUAGGUUCUAUUUAUCCUUCUCUGCUCGGAUUUUUUAGUAAUUGGCCAGAA
UUGUUUGAGCAAGGACGAAUUCGCUUUGUCAAAACUCCUGUAAUCAUCGCUCAGGUCGGU
AAAAAACAAGAAUGGUUUUAUACAGUCGCUGAAUAUGAGAGUGCCAAAGAUGCUCUACCU
AAACAUAGCAUCCGUUAUAUUAAGGGACUUGGCUCUUUGGAAAAAUCUGAAUAUCGUGAG
AUGAUUCAAAACCCAGUAUAUGAUGUUGUUAAACUUCCUGAGAACUGGAAAGAGCUUUUU
GAAAUGCUCAUGGGAGAUAAUGCUGACCUUCGUAAAGAAUGGAUGAGCCAGUAGUUUACU
UUACCACAAGGAUGUGGUAUAAUUAAUUGGGCAAAAGCUUGGCGUAAUCAUGGUCAUAGC
UGUUUCCUGUGUGAAAUUGUUAUCCG

Appendix B

Preparation of Cell Extracts for Translation

Several different translation extracts were prepared by the methods described below. Two different *E. coli* strains were used to prepare extracts. MRE600 (ATCC 29417) is a common strain used to make extracts because of its low ribonuclease activity[142]. The second strain we used was BL21 Star™ (DE3) cells (Invitrogen C6010-03). This strain contains a mutation in RNase E, a major component of the RNA degradosome, resulting in longer half-lives of mRNA and enhanced protein production.

To different types of extracts were made: S30 and S100. S30 extracts contain all components required for translation including the ribosomes, whereas S100 extracts lack ribosomes. Additionally, the ribosomes pelleted while preparing the S100 extract were purified over a glycerol gradient and isolated as 70S tight-couples.

Variations in extract prep were made in an attempt to optimize protein production. As can be seen in Figure II.2, the behavior of each extract varies significantly depending on the strain, extract type and preparation conditions. Detailed preparation methods for each extract can be founding Table B-1 below.

Table B.1 (first half) Preparation Methods for Extracts

	7/7/2008	8/7/2008	8/11/2008	1/22/2009	1/22/2009
Extract	**M1_S30**	**M_S100**	**Ribosomes**	**M2_S30**	**B1_S30**
Strain	MRE600	MRE600	MRE600	MRE600	BL21 Star DE3
Growth Medium	207 mM potassium phosphate pH 7.3 10 g/L yeast extract 1% glucose	207 mM potassium phosphate pH 7.3 10 g/L yeast extract 1% glucose		2XYT 16.5 g/L bacto-tryptone 10 g/L bacto yeast extract 5 g/L NaCl	2XYT 16.5 g/L bacto-tryptone 10 g/L bacto yeast extract 5 g/L NaCl
Harvest Cells	4 hrs, 37C	4 hrs, 37C		4 hrs, 37C	4 hrs, 37C add IPTG to 1 mM 2 hrs, 37C
Wash cells	~24 g cells 3X 10 mM Tris-acetate, pH 8.0 14 mM Mg(Oac)2 60 mM KCl 6 mM β-ME 1 mM PMSF	~24 g cells 3X 10 mM Tris-acetate, pH 8.0 14 mM Mg(Oac)2 60 mM KCl 6 mM β-ME 1 mM PMSF		3.1 g cells 3X 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 7 mM β-ME 1 mM DTT	2.5 g cells 3X 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 7 mM β-ME 1 mM DTT
Lyse cells	Fierke microfluidizer	Fierke microfluidizer		Fierke microfluidizer	Fierke microfluidizer
Centrifugation	2X Ti45 rotor 30 min 4C 19,500 rpm decant super	Ti45 rotor 20 min 4C 19,000 rpm, 30 krcf decant super Ti45 rotor 60 min 4C 19,000 rpm, 30 krcf decant super Ti45 rotor		3X Ti45 rotor 30 min 4C 19,500 rpm, 30 krcf pipette off super	3X Ti45 rotor 30 min 4C 19,500 rpm, 30 krcf pipette off super

		120 min 4C 34,000 rpm, 30 krcf			
		pipette off super	rinse pellets 3X with lysis buffer resuspend in 30 mL lysis buffer with stir bar overnight Ti45 rotor 120 min 4C 34,000 rpm, 30 krcf rinse pellets 3X with lysis buffer resuspend in 2 mL lysis buffer with stir bar overnight run over gradient collected 70S tight couples		
Aminoacylation	15 mL aminacylation mix 750 mL Tris-acetate pH 8.0 7.5 mM DTT 21.3 mM Mg(OAc)2 75 uM amicase (a.a. source) 6 mM ATP 20 mg/mL phosphoenol pyruvate 50U pyruvate kinase 37C, 2 hrs shaking in the dark			37C, 90 min shaking 100 rpm	37C, 90 min shaking 100 rpm
Dialyze	MWCO 500 Da 10 mM Tris-acetate pH8.0 14 mM Mg(Oac)2 60 mM KOAc 1 mM DTT 1 mM PMSF 2 L, 4C, 3X	MWCO 500 Da 10 mM Tris-acetate pH8.0 14 mM Mg(Oac)2 60 mM KOAc 1 mM DTT 1 mM PMSF 2 L, 4C, 3X		MWCO 10,000 Da 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 7 mM β-ME 1 mM DTT 4X	MWCO 10,000 Da 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 7 mM β-ME 1 mM DTT 4X
Centrifugation				6700 rcf, 10 min, 4C	6700 rcf, 10 min, 4C
Concentration				3-fold with MWCO 10,000	3-fold with MWCO 10,000

Storage	1 mL aliquots snap freeze store -80C	1 mL aliquots snap freeze store -80C	100 uL aliquots snap freeze store -80C	200 uL aliquots snap freeze store -80C	200 uL aliquots snap freeze store -80C

Table B.1 (second half) Preparation Methods for Extracts

Extract	1/29/2010 **M3_S30-**	1/29/2010 **M3_S20+**	1/29/2010 **B2_S30-**	1/29/2010 **B2_S30+**
Strain	MRE600	MRE600	BL21 Star DE3	BL21 Star DE3
Growth Medium	2XYT 16.5 g/L bacto-tryptone 10 g/L bacto yeast extract 5 g/L NaCl		2XYT 16.5 g/L bacto-tryptone 10 g/L bacto yeast extract 5 g/L NaCl	
Harvest Cells	3.5 hrs, 37C		3.5 hrs, 37C add IPTG to 1 mM 1.5 hrs, 37C	
Wash cells	8.7 g cells 3X 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 7 mM β-ME 1 mM DTT		10 g cells 3X 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 7 mM β-ME 1 mM DTT	
Lyse cells	Fierke microfluidizer		Fierke microfluidizer	
Centrifugation	2X Ti45 rotor 30 min 4C 19,500 rpm, 30 krcf pipette off super		2X Ti45 rotor 30 min 4C 19,500 rpm, 30 krcf pipette off super	
Aminoacylation		1.5 U/mL pyruvate kinase 68 mM Tris-acetate pH 8.2 2.1 mM Mg(OAc)2 3 mM ATP, pH 7.0 20 mM phosphoenolpyruvate, pH 7.0 1 mM DTT 9.2 uM each amino acid		1.5 U/mL pyruvate kinase 68 mM Tris-acetate pH 8.2 2.1 mM Mg(OAc)2 3 mM ATP, pH 7.0 20 mM phosphoenolpyruvate, pH 7.0 1 mM DTT 9.2 uM each amino acid

	37C, 80 min, 125 rpm	37C, 80 min, 125 rpm	37C, 80 min, 125 rpm	37C, 80 min, 125 rpm
Dialyze	MWCO 10,000 Da 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 1 mM DTT 4X	MWCO 10,000 Da 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 1 mM DTT 4X	MWCO 10,000 Da 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 1 mM DTT 4X	MWCO 10,000 Da 10 mM Tris-acetate, pH 8.2 14 mM Mg(Oac)2 60 mM KCl 1 mM DTT 4X
Centrifugation	10,000 rcf, 10 min, 4C	10,000 rcf, 10 min, 4C	10,000 rcf, 10 min, 4C	10,000 rcf, 10 min, 4C
Concentration	MWCO 10,000 to 2 mL/g cells	MWCO 10,000 to 2 mL/g cells	MWCO 10,000 to 2 mL/g cells	MWCO 10,000 to 2 mL/g cells
Storage	100 uL aliquots snap freeze store -80C	100 uL aliquots snap freeze store -80C	100 uL aliquots snap freeze store -80C	100 uL aliquots snap freeze store -80C

Appendix C

Internship with Professor Rachel Green at The Johns Hopkins University

Introduction

Funded by a Rackham Research Grant, I spent one month in the lab of Professor Rachel Green at Johns Hopkins University. Under the mentorship of Dr. Hani Zaher, I learned several techniques pertaining to *in vitro* translation

Translational Bypassing of Gene *60* mRNA in S30 Extract

The first experiment performed in the Green lab was a simple test of my gene *60* mRNA templates in S30 extract to see if translational bypassing was possible. RNA was added to extract and translation products were labeled with ^{35}S-Met. The results of the translation reaction are shown in Figure C-1. mRNA for the three size standard lanes on the right was obtained from Dr. Zaher. Gp60 products are shown in the two middle lanes. For wildtype gene *60* mRNA, three bands were present in the gel, whereas for gene *60* Δ gap only two bands were present. The topmost band in each lane runs just about 20 kDa and is present for both templates, as expected for full-length gp60. The middle band runs just below 10 kDa and is not present when gene *60* Δ gap mRNA is used as a template. This band corresponds to truncated gp60. (Note: the theoretical pIs for full-length and truncated gp60 are 8.65 and 9.52, respectively, so they could migrate more slowly than expected for their molecular weights.) The bottommost band in each lane could be degraded or to another short termination product of gp60. The three lanes on the left are translation products from CAT, EmGFP and gene *32* mRNA. Translation of CAT and gene *32* mRNA did not occur. Translation of EmGFP produced a significant amount of a truncated product for unknown reasons. These templates were not studied further.

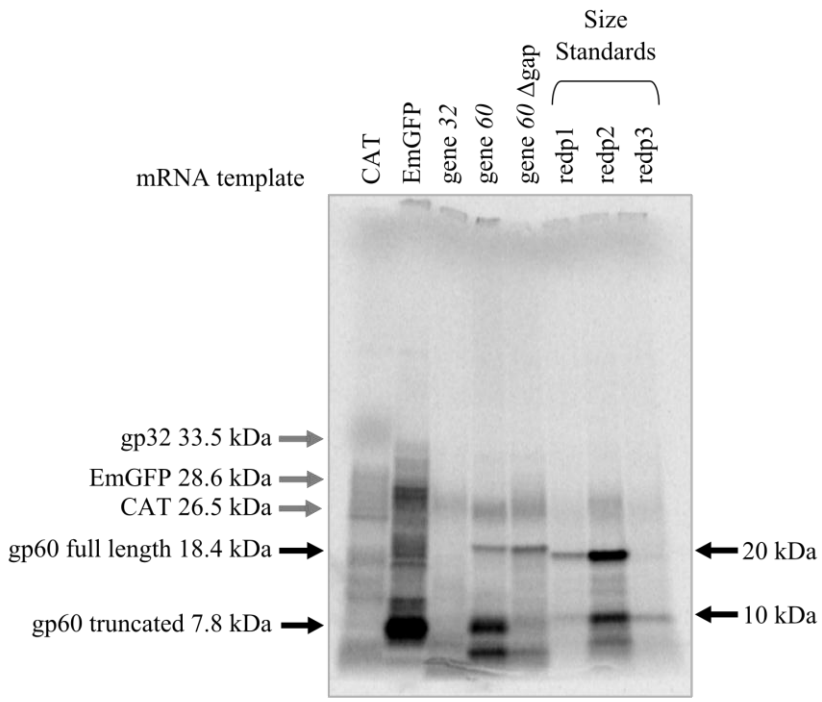


Figure C.1 Testing gene *60* for translational bypassing. 1 µg of each mRNA was translated in a 50 µl reaction. A master mix containing 20 µl of Promega's S30 pre-mix, 15 µl of S30 extract from *E. coli* MRE600 and 5 µl of Promega's amino acid mix minus methionine (1 mM each) per reaction was incubated at 37°C to initiate translation of contaminating mRNAs. 2.5 µl of ^{35}S-Met (~430 nM final) was added to the master mix, which was subsequently divided to bring each reaction up to 50 µl. Translation reactions were incubated at 37°C for 30 min and precipitated in 5 volumes of cold acetone. After centrifugation at room temperature for 5 min, the supernatant was removed and the dried pellet was resuspended in 50 µl of SDS loading buffer. Samples were boiled for 5 min and 10 µl of each reaction were resolved on a pre-cast Criterion XT 4-12% Bis-Tris gel (Biorad).

Preparation of Initiation Complexes

We next wanted to determine if we could observe translation and translational bypassing of gene *60* mRNA with completely recombinant components. To do this we first tested the steps of translation initiation. Charged and formylated $tRNA^{fMet}$ was prepared by incubating $tRNA^{fMet}$ with a formyl donor, ^{35}S-Met and S100 extract, which contains the formylase enzyme. (Alternatively, this step could be performed with a recombinantly purified formylase, but the Green Lab does not have this enzyme.) After incubation for 30 min at 37°C, the reaction was extracted with phenol and chloroform to remove proteins and the RNA was precipitated with ethanol.

Competency of initiation was tested by incubating mRNA with 70S ribosomes, recombinant initiation factors (IFs) 1, 2 and 3 and formyl-Met-tRNA^{fMet}. Functional IFs will assemble the mRNA into the ribosome with the start codon aligned in the P-site and stabilized by fMet-tRNA^{fMet}. This initiation complex (IC) can be purified through a sucrose cushion. The efficiency of initiation can be estimated by calculating the fraction of ^{35}S-Met that pellets with the ribosome because charged tRNAs will only remain bound to the ribosome if they are assembled in an IC.

ICs were formed generally as described by Hani Zaher[143] but using polymix buffer[144] for all steps. Final concentrations in the initiation complex formation assay are as follows: 1X polymix (95 mM KCl, 5 mM NH_4Cl, 5 mM $Mg(OAc)_2$, 0.5 mM $CaCl_2$, 8 mM putrescine, 1 mM spermidine, 5 mM KPO_4, pH 7.5, 1 mM DTT), H_2O as necessary, 10 mM K_2HPO_4, 2 mM GTP, 3 μM IF1, 3 μM IF2, 3 μM IF3, 2.4 μM ^{35}S-fMet-tRNA^{fMet}, 2 μM 70S ribosomes, 6 μM mRNA. Reactions were incubated for 45 min at 37°C. 2 μl of 1 M $MgCl_2$ were added to the reaction to maintain 70S initiation complexes during centrifugation. Initiation complexes were purified through a sucrose cushion (1.1 M sucrose, 20 mM Tris-HCl, pH 7.5, 500 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA) by centrifugation at 258,000 x g in a TLA100.3 rotor for 2 hrs. The resulting pellet was resuspended in 95 mM KCl, 5 mM NH_4Cl, 5 mM magnesium acetate, 0.5 mM $CaCl_2$, 8 mM putrescine, 1 mM spermidine, 15 mM potassium phosphate, pH 7.5, 1 mM DTT and stored in aliquots at -80°C.

The efficiency of IC formation for gene *60* mRNA was ~25%, which is fairly low compared to other templates tested by the Green lab. IC formation for gene *60* Δ gap mRNA was very poor; only 2% of the ^{35}S-Met was incorporated. Therefore gene *60* Δ gap IC was not studied further.

Elongation Capacity of ICs

To determine whether or not our gene *60* ICs were translationally competent, we incubated our ICs with various extracts (Figure C-2). The S30 extracts used previously (Figure C-1) contain ribosomes, we cannot be sure if translation products from these extracts came from our pre-initiated complexes or if the extract dissociates our IC and reuses the ^{35}S-fMet-tRNA^{fMet} for new translation reactions. Thus ICs were also added to

S100 extracts, which lack ribosomes. Translation in S100 extracts ensures products are originating from our ICs. Finally, our ICs were tested for translational competency with completely recombinant translation proteins. The result of these experiments can be seen in Figure C-2.

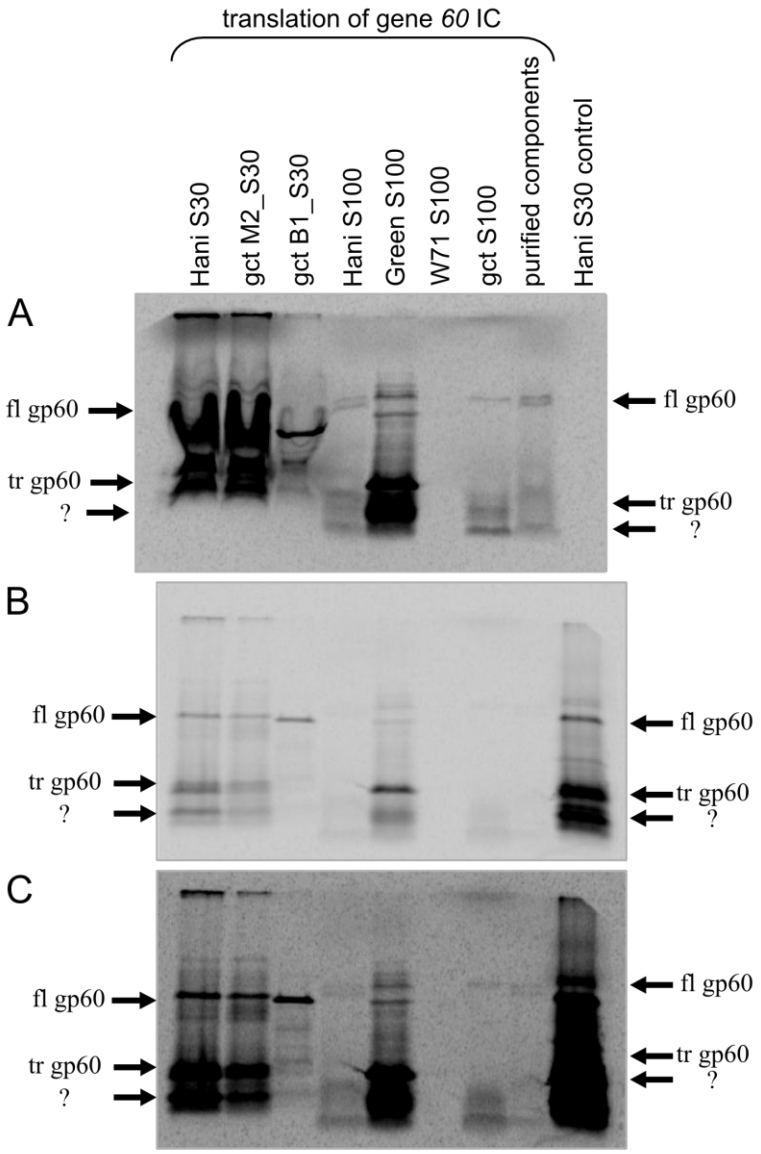


Figure C.2 Testing gene *60* ICs for translational bypassing. Gene *60* ICs were tested for translation and translational bypassing capabilities in the indicated extracts. All gels contain the same translation reactions, except **B** and **C** have an added S30 control of gene *60* wildtype from figure C-1 above. **B** and **C** are the same gel, just presented at different contrasts so the bands from the completely purified system can be observed.

Translation of ICs in S30 extract was carried out in 10 µl reactions: 4 µl of Promega's S30 pre-mix, 1 µl of Promega's amino acid mix (1 mM each) and 3 µl of S30 extract were added to 2 µl of IC. Reactions were incubated at 37°C for 15 min. 0.5 µl of 1 M KOH was added to deacylate the peptidyl tRNA and reactions were precipitated with 50 µL of cold acetone. Following centrifugation at room temperature, dried pellets were resuspended in 20 µl of SDS loading dye. 10 µl of boiled sample were resolved on a 16% Tris-tricine gel.

Translation in S100 extracts or with recombinant components required addition of pre-charged tRNAs. Charging reactions were performed in 95 mM KCl, 5 mM NH_4Cl, 5 mM magnesium acetate, 0.5 mM $CaCl_2$, 8 mM putrescine, 1 mM spermidine, 10 mM potassium phosphate, pH 7.5, 1 mM DTT, 2.1 mM ATP, 200 µM of each amino acid, 150 µM of total *E. coli* MRE600 tRNA, 3.2 mM phosphoenolpyruvate, 40 µg/ml pyruvate kinase, 1 µM each recombinantly purified tRNA synthetase. Charging reactions were incubated at 37°C for 15 min.

5 µl of each tRNA charging reaction and 3 µl of S100 extract were added to 2 µl of IC. Reactions were incubated at 37°C for 15 min and stopped by the addition of .5 µl of 1 M KOH (to deacylate peptidyl tRNAs) and 10 µl of Tris-tricine loading dye. 10 µl of boiled sample were resolved on a 16% Tris-tricine gel.

Alternatively, 5 µl of each tRNA charging reaction were added to recombinantly purified translation factors at the final concentration of 87.2 µM EF-Tu-GTP, 6 µM EF-G, 6 µM RF1, 6 µM RF2, 6 µM RF3. 2 µl of IC was added and the reaction was incubated at 37°C for 15 min. Reactions were stopped by the addition of .5 µl of 1 M KOH (to deacylate peptidyl tRNAs) and 10 µl of Tris-tricine loading dye. 10 µl of boiled sample were resolved on a 16% Tris-tricine gel.

The products of translational bypassing and their relative intensities are very different when different translation conditions are used. Three different S30 extracts were used, two generated from the *E. coli* MRE600 strain (one I prepared and one made by Dr. Zaher) and the third came from BL21 D3 Star cells. The two extracts from MRE600 appear to have similar translation and translational bypassing efficiencies. BL21 DE3 Star extract, however does not generate any truncated gp60 indicating it is extremely efficient at bypassing. Four different S100 extract preparations were tested,

but only one appears translationally competent. In this extract, the amount of truncated gp60 greatly exceeds that of full-length, indicating this extract is not very efficient in bypassing. This experiment does, however, indicate our initiation complexes are indeed competent for elongation.

Finally, a small amount of translation was observed in the fully recombinant translation reaction. There is a second band above the full-length gene *60*, and there is not a defined band for the truncated product. Overall, however, this translation condition was very poor. One possible cause could be inefficient charging of the tRNAs by the recombinantly purified tRNA synthetases (RS). This hypothesis is addressed below. A second potential problem with this assay is the lack of EF-Ts, which is used to recycle EF-Tu during translation elongation. Addition of this protein may enhance efficiency with completely recombinant translation.

Charging Efficiency of Purified tRNA Synthetases

To address the low translational yield in the recombinant translation condition, the efficiency of tRNA aminoacylation by each synthetase was measured with ^{14}C or ^{3}H labeled amino acids. Charging efficiency was generally < 50% (Table C-1). Most of these protein stocks were not purified by Dr. Zaher and were of unknown age. I prepared fresh batches of each tRNA synthetase (except for Asn RS and Phe RS, which were recently purified and had high charging efficiencies) (data not shown). All RS proteins were His-tagged and purified over nickel columns either by gravity or by FPLC.

The freshly prepared synthetases are generally more efficient than the older stocks (highlighting indicates most efficient method for charging each amino acid). Additionally, the charging efficiency of all of the amino acids in S100 extract was tested to see if this could be used for charging instead of the purified synthetases. In many cases the S100 extract worked better than the purified synthetases. Highlighted below in Table C-1 is the condition that gives the best overall charging efficiency for each amino acid. Charging efficiency was generally measured with ^{14}C-labeled or ^{3}H-labeled amino acids as indicated, except for His, which was estimated from a ^{14}C-labeled amino acid mix.

	Amino Acid	Test with Green Lab Synthetases	Test with S100 extract gct	Test of newly purified Synthetases	
	Ala	11%	66%	13%	
	Arg	46%	35%	33%	
3H	Asn	240%	132%	**115%**	Green lab's
	Asp	16%	77%	88%	
	Cys	13%	52%	66%	
	Gln	71%	139%	83%	
	Glu	11%	46%	27%	
	Gly		37%	37%	
a.a. mix	His	150%	1053%	52%	
	Ile	29%	27%	24%	
3H	Leu	6%	18%	39%	
	Lys	98%	86%	108%	
	Met	313%	83%	203%	
	Phe	92%	99%	**73%**	Green lab's
	Pro	24%	62%	55%	
	Ser	51%	96%	71%	
	Thr	99%	93%	114%	
	Trp	56%	41%	116%	
	Tyr	29%	31%	46%	
	Val	41%	33%	53%	

Table C.1 Aminoacylation efficiency of each recombinantly purified tRNA synthetase. Charging efficiency was determined using total *E. coli* tRNAs and the indicated labeled amino acid. Absolute quantities of each tRNA were estimated based on their natural abundance as reported by [144].

Activity of Purified Ribosomes

To test whether ribosomes I prepared in the Walter lab (Appendix B) are active, gene *60* ICs were prepared using both my and Hani Zaher's ribosomes. ICs were incubated with Lys-tRNA^{Lys} and Phe-tRNA^{Phe} (the second and third amino acids of gp60, respectively) and with and without EF-G. In all cases, the dipeptide ^{35}S-Met-Lys should form. Only in the presence of EF-G, however, will the peptidyl tRNA be translocated into the P-site, allowing Phe-tRNA^{Phe} to enter the A-site and add to the peptide chain. Lanes 1 and 2 show the elongation products of Hani Zaher's ICs in the presence and

absence of EF-G. Tripeptide was only formed upon addition of EF-G, but conversion was almost complete. Significant elongation occured with my ICs in the absence of added EF-G indicating they are not completely pure. Upon addition of EF-G to my ICs there was additional conversion to tripeptide over background, but the extent of elongation was not as significant as that catalyzed with Hani Zaher's ICs, indicating further impurities. Future preparations of ribosomes should be carried out as described in Hani Zaher's protocol.

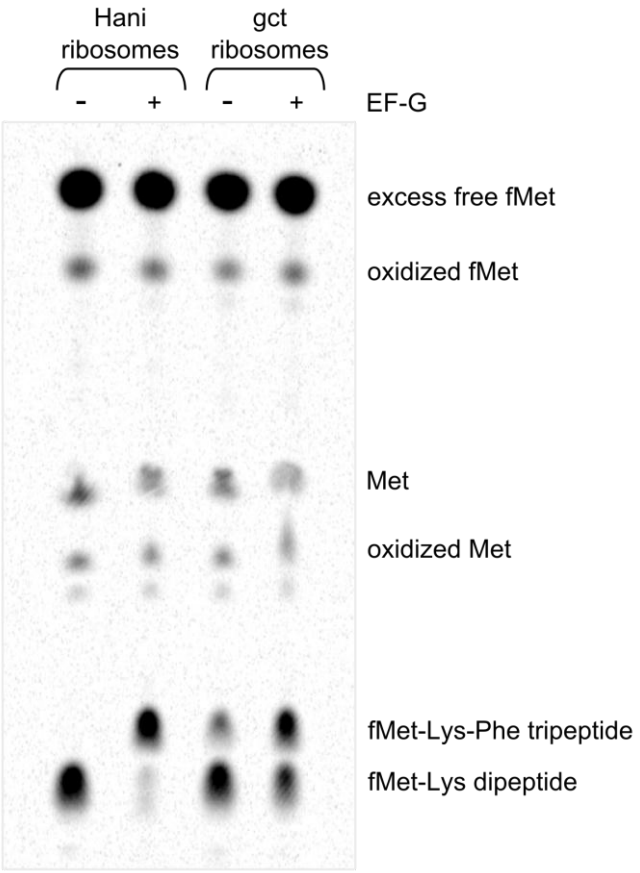


Figure C.3 Testing ribosome activity. Small scale gene *60* ICs were made in 1X polymix (95 mM KCl, 5 mM NH_4Cl, 5 mM $Mg(OAc)_2$, 0.5 mM $CaCl_2$, 8 mM putrescine, 1 mM spermidine, 5 mM KPO_4, pH 7.5, 1 mM DTT), H_2O as necessary, 10 mM K_2HPO_4, 2 mM GTP, 3 μM IF1, 3 μM IF2, 3 μM IF3, 2.4 μM ^{35}S-fMet-tRNA^{fMet}, 2 μM 70S ribosomes, 6 μM mRNA. Reactions were incubated for 45 min at 37°C. GDP bound EF-Tu was exchanged with GTP by incubation in 1X polymix, H_2O as necessary, 10 mM KH_2PO_4, 2 mM GTP and 50 μM EF-Tu by incubation at 37°C for 20 min. Charge Phe-tRNA^{Phe} and Lys-tRNA^{Lys} were added to a final concentration of 6 μM and incubated at 37°C for 25 min to generate ternary complexes between EF-Tu and the charged tRNAs. IC and ternary complexes were incubated with and without EF-G at room temperature for less than 1 min to prevent miscoding. 0.5 μL of 1 M KOH was added to deacylate the tRNA. 0.5 μL of each reaction was spotted onto the center of cellulose TLC paper and resolved by electrophoretic TLC at 1200 V for 35 min. After drying the TLC plate, it was exposed to a storage phosphor screen overnight.

Charging Total tRNAs

Total tRNA charging was performed as described above with gct S100, Green S100, gct purified RS mix, Green purified RS mix and no RS. Results are presented in Table C-2. Translation with completely recombinantely purified proteins was repeated as described above using each of the charge tRNA preparations listed below. Translation was not apparent in any condition and there was not time left in my internship to trouble-shoot this reaction.

RS source	Average fraction charged	STD DEV
gct S100	11.15%	1.26%
Green S100	19.44%	1.75%
gct RS	15.57%	0.48%
Green RS	11.93%	0.36%
no RS	0.02%	0.01%

Table C.2 Charging efficiency of total tRNAs with the indicated extract or purified RS mix. The Green lab's S100 extract provides the best charging of the extracts whereas my freshly prepared RS mix was more efficient than the Green lab's older RS mix.

References

1 Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).

2 Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S. & Amster-Choder, O. Translation-independent localization of mRNA in E. coli. *Science* **331**, 1081-1084, doi:10.1126/science.1195691 (2011).

3 Komar, A. A. A pause for thought along the co-translational folding pathway. *Trends in biochemical sciences* **34**, 16-24, doi:10.1016/j.tibs.2008.10.002 (2009).

4 Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature structural & molecular biology* **16**, 274-280, doi:10.1038/nsmb.1554 (2009).

5 Breaker, R. R. Complex riboswitches. *Science* **319**, 1795-1797, doi:10.1126/science.1152621 (2008).

6 Breaker, R. R. Prospects for riboswitch discovery and analysis. *Molecular cell* **43**, 867-879, doi:10.1016/j.molcel.2011.08.024 (2011).

7 Stoddard, C. D. & Batey, R. T. Mix-and-match riboswitches. *ACS chemical biology* **1**, 751-754, doi:10.1021/cb600458w (2006).

8 Winkler, W. C. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Current opinion in chemical biology* **9**, 594-602, doi:10.1016/j.cbpa.2005.09.016 (2005).

9 Johansson, J. *et al.* An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. *Cell* **110**, 551-561 (2002).

10 Morita, M. T. *et al.* Translational induction of heat shock transcription factor sigma32: evidence for a built-in RNA thermosensor. *Genes & development* **13**, 655-665 (1999).

11 Neupert, J. & Bock, R. Designing and using synthetic RNA thermometers for temperature-controlled gene expression in bacteria. *Nature protocols* **4**, 1262-1273, doi:10.1038/nprot.2009.112 (2009).

12 Neupert, J., Karcher, D. & Bock, R. Design of simple synthetic RNA thermometers for temperature-controlled gene expression in Escherichia coli. *Nucleic acids research* **36**, e124, doi:10.1093/nar/gkn545 (2008).

13 Waldminghaus, T., Heidrich, N., Brantl, S. & Narberhaus, F. FourU: a novel type of RNA thermometer in Salmonella. *Molecular microbiology* **65**, 413-424, doi:10.1111/j.1365-2958.2007.05794.x (2007).

14 Waldminghaus, T., Kortmann, J., Gesing, S. & Narberhaus, F. Generation of synthetic RNA-based thermosensors. *Biological chemistry* **389**, 1319-1326, doi:10.1515/BC.2008.150 (2008).

15 Chou, M. Y. & Chang, K. Y. An intermolecular RNA triplex provides insight into structural determinants for the pseudoknot stimulator of -1 ribosomal frameshifting. *Nucleic acids research* **38**, 1676-1685, doi:10.1093/nar/gkp1107 (2010).

16 Chung, B. Y., Firth, A. E. & Atkins, J. F. Frameshifting in alphaviruses: a diversity of 3' stimulatory structures. *Journal of molecular biology* **397**, 448-456, doi:10.1016/j.jmb.2010.01.044 (2010).
17 Farabaugh, P. J. Programmed translational frameshifting. *Microbiological reviews* **60**, 103-134 (1996).
18 Giedroc, D. P. & Cornish, P. V. Frameshifting RNA pseudoknots: structure and mechanism. *Virus research* **139**, 193-208, doi:10.1016/j.virusres.2008.06.008 (2009).
19 Herr, A. J., Atkins, J. F. & Gesteland, R. F. Coupling of open reading frames by translational bypassing. *Annual review of biochemistry* **69**, 343-372, doi:10.1146/annurev.biochem.69.1.343 (2000).
20 Kontos, H., Napthine, S. & Brierley, I. Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Molecular and cellular biology* **21**, 8657-8670, doi:10.1128/MCB.21.24.8657-8670.2001 (2001).
21 Mazauric, M. H., Seol, Y., Yoshizawa, S., Visscher, K. & Fourmy, D. Interaction of the HIV-1 frameshift signal with the ribosome. *Nucleic acids research* **37**, 7654-7664, doi:10.1093/nar/gkp779 (2009).
22 Burke-Aguero, D. H. & Hearst, J. E. An RNA Holliday junction? Structural and dynamic considerations of the bacteriophage T4 gene 60 interruption. *Journal of molecular biology* **213**, 199-201, doi:10.1016/S0022-2836(05)80177-9 (1990).
23 Herr, A. J., Gesteland, R. F. & Atkins, J. F. One protein from two open reading frames: mechanism of a 50 nt translational bypass. *The EMBO journal* **19**, 2671-2680, doi:10.1093/emboj/19.11.2671 (2000).
24 Herr, A. J., Wills, N. M., Nelson, C. C., Gesteland, R. F. & Atkins, J. F. Drop-off during ribosome hopping. *Journal of molecular biology* **311**, 445-452, doi:10.1006/jmbi.2001.4899 (2001).
25 Huang, W. M. *et al.* A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science* **239**, 1005-1012 (1988).
26 Le, S. Y., Chen, J. H. & Maizel, J. V., Jr. Identification of unusual RNA folding patterns encoded by bacteriophage T4 gene 60. *Gene* **124**, 21-28 (1993).
27 Wills, N. M. *et al.* Translational bypassing without peptidyl-tRNA anticodon scanning of coding gap mRNA. *The EMBO journal* **27**, 2533-2544, doi:10.1038/emboj.2008.170 (2008).
28 Weiss, R. B., Huang, W. M. & Dunn, D. M. A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell* **62**, 117-126 (1990).
29 Maldonado, R. & Herr, A. J. Efficiency of T4 gene 60 translational bypassing. *Journal of bacteriology* **180**, 1822-1830 (1998).
30 Ringquist, S., MacDonald, M., Gibson, T. & Gold, L. Nature of the ribosomal mRNA track: analysis of ribosome-binding sites containing different sequences and secondary structures. *Biochemistry* **32**, 10254-10262 (1993).
31 Zuker, M. On finding all suboptimal foldings of an RNA molecule. *Science* **244**, 48-52 (1989).

32 Petrov, V. M., Ratnayaka, S., Nolan, J. M., Miller, E. S. & Karam, J. D. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Virology journal* **7**, 292, doi:10.1186/1743-422X-7-292 (2010).
33 Bonocora, R. P., Zeng, Q., Abel, E. V. & Shub, D. A. A homing endonuclease and the 50-nt ribosomal bypass sequence of phage T4 constitute a mobile DNA cassette. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 16351-16356, doi:10.1073/pnas.1107633108 (2011).
34 Genomes of the T4-like Phages.
35 Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiology and molecular biology reviews : MMBR* **67**, 86-156, table of contents (2003).
36 Edgell, D. R., Gibb, E. A. & Belfort, M. Mobile DNA elements in T4 and related phages. *Virology journal* **7**, 290, doi:10.1186/1743-422X-7-290 (2010).
37 Repoila, F., Tetart, F., Bouet, J. Y. & Krisch, H. M. Genomic polymorphism in the T-even bacteriophages. *The EMBO journal* **13**, 4181-4192 (1994).
38 Laponogov, I. *et al.* Structural basis of gate-DNA breakage and resealing by type II topoisomerases. *PloS one* **5**, e11338, doi:10.1371/journal.pone.0011338 (2010).
39 Sissi, C. & Palumbo, M. Effects of magnesium and related divalent metal ions in topoisomerase structure and function. *Nucleic acids research* **37**, 702-711, doi:10.1093/nar/gkp024 (2009).
40 Dassa, B., London, N., Stoddard, B. L., Schueler-Furman, O. & Pietrokovski, S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic acids research* **37**, 2560-2573, doi:10.1093/nar/gkp095 (2009).
41 Mootz, H. D. Split inteins as versatile tools for protein semisynthesis. *Chembiochem : a European journal of chemical biology* **10**, 2579-2589, doi:10.1002/cbic.200900370 (2009).
42 Bucklin, D. J., Wills, N. M., Gesteland, R. F. & Atkins, J. F. P-site pairing subtleties revealed by the effects of different tRNAs on programmed translational bypassing where anticodon re-pairing to mRNA is separated from dissociation. *Journal of molecular biology* **345**, 39-49, doi:10.1016/j.jmb.2004.10.037 (2005).
43 Gallant, J. A. & Lindsley, D. Ribosomes can slide over and beyond "hungry" codons, resuming protein chain elongation many nucleotides downstream. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13771-13776 (1998).
44 Noller, H. F. *et al.* Secondary structure model for 23S ribosomal RNA. *Nucleic acids research* **9**, 6167-6189 (1981).
45 Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Y. Understanding the transcriptome through RNA structure. *Nature reviews. Genetics* **12**, 641-655, doi:10.1038/nrg3049 (2011).
46 Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature protocols* **1**, 1610-1616, doi:10.1038/nprot.2006.249 (2006).
47 Woese, C. R. *et al.* Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic acids research* **8**, 2275-2293 (1980).

48 Adamski, F. M., Atkins, J. F. & Gesteland, R. F. Ribosomal protein L9 interactions with 23 S rRNA: the use of a translational bypass assay to study the effect of amino acid substitutions. *Journal of molecular biology* **261**, 357-371, doi:10.1006/jmbi.1996.0469 (1996).
49 Herr, A. J., Atkins, J. F. & Gesteland, R. F. Mutations which alter the elbow region of tRNA2Gly reduce T4 gene 60 translational bypassing efficiency. *The EMBO journal* **18**, 2886-2896, doi:10.1093/emboj/18.10.2886 (1999).
50 Herr, A. J., Wills, N. M., Nelson, C. C., Gesteland, R. F. & Atkins, J. F. Factors that influence selection of coding resumption sites in translational bypassing: minimal conventional peptidyl-tRNA:mRNA pairing can suffice. *The Journal of biological chemistry* **279**, 11081-11087, doi:10.1074/jbc.M311491200 (2004).
51 Larsen, B. *et al.* Upstream stimulators for recoding. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **73**, 1123-1129 (1995).
52 Choi, K. M., Atkins, J. F., Gesteland, R. F. & Brimacombe, R. Flexibility of the nascent polypeptide chain within the ribosome--contacts from the peptide N-terminus to a specific region of the 30S subunit. *European journal of biochemistry / FEBS* **255**, 409-413 (1998).
53 Ito, K., Chiba, S. & Pogliano, K. Divergent stalling sequences sense and control cellular physiology. *Biochemical and biophysical research communications* **393**, 1-5, doi:10.1016/j.bbrc.2010.01.073 (2010).
54 Jha, S. & Komar, A. A. Birth, life and death of nascent polypeptide chains. *Biotechnology journal* **6**, 623-640, doi:10.1002/biot.201000327 (2011).
55 Bhushan, S. *et al.* SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS biology* **9**, e1000581, doi:10.1371/journal.pbio.1000581 (2011).
56 Gong, M., Cruz-Vera, L. R. & Yanofsky, C. Ribosome recycling factor and release factor 3 action promotes TnaC-peptidyl-tRNA Dropoff and relieves ribosome stalling during tryptophan induction of tna operon expression in Escherichia coli. *Journal of bacteriology* **189**, 3147-3155, doi:10.1128/JB.01868-06 (2007).
57 Seidelt, B. *et al.* Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* **326**, 1412-1415, doi:10.1126/science.1177662 (2009).
58 de Felipe, P., Hughes, L. E., Ryan, M. D. & Brown, J. D. Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide. *The Journal of biological chemistry* **278**, 11441-11448, doi:10.1074/jbc.M211644200 (2003).
59 Choi, K. M. & Brimacombe, R. The path of the growing peptide chain through the 23S rRNA in the 50S ribosomal subunit; a comparative cross-linking study with three different peptide families. *Nucleic acids research* **26**, 887-895 (1998).
60 Takyar, S., Hickerson, R. P. & Noller, H. F. mRNA helicase activity of the ribosome. *Cell* **120**, 49-58, doi:10.1016/j.cell.2004.11.042 (2005).
61 Boni, I. V., Artamonova, V. S., Tzareva, N. V. & Dreyfus, M. Non-canonical mechanism for translational control in bacteria: synthesis of ribosomal protein S1. *The EMBO journal* **20**, 4222-4232, doi:10.1093/emboj/20.15.4222 (2001).

62 Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nature biotechnology* **19**, 751-755, doi:10.1038/90802 (2001).
63 Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* **36**, 299-304, doi:10.1016/j.ymeth.2005.04.006 (2005).
64 Swartz, J. A PURE approach to constructive biology. *Nature biotechnology* **19**, 732-733, doi:10.1038/90773 (2001).
65 Chambliss, G. H., Henkin, T. M. & Leventhal, J. M. Bacterial in vitro protein-synthesizing systems. *Methods in enzymology* **101**, 598-605 (1983).
66 Katzen, F., Chang, G. & Kudlicki, W. The past, present and future of cell-free protein synthesis. *Trends in biotechnology* **23**, 150-156, doi:10.1016/j.tibtech.2005.01.003 (2005).
67 Lesley, S. A., Brow, M. A. & Burgess, R. R. Use of in vitro protein synthesis from polymerase chain reaction-generated templates to study interaction of Escherichia coli transcription factors with core RNA polymerase and for epitope mapping of monoclonal antibodies. *The Journal of biological chemistry* **266**, 2632-2638 (1991).
68 Mackow, E. R., Yamanaka, M. Y., Dang, M. N. & Greenberg, H. B. DNA amplification-restricted transcription-translation: rapid analysis of rhesus rotavirus neutralization sites. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 518-522 (1990).
69 Matthaei, J. H. & Nirenberg, M. W. Characteristics and stabilization of DNAase-sensitive protein synthesis in E. coli extracts. *Proceedings of the National Academy of Sciences of the United States of America* **47**, 1580-1588 (1961).
70 Pratt, J. M. in *Transcription and Translation: A Practical Approach* (ed B. D. and Higgins Hames, S. J.) 179-209 (IRL Press, New York, 1984).
71 Rheinberger, H. J., Geigenmuller, U., Wedde, M. & Nierhaus, K. H. Parameters for the preparation of Escherichia coli ribosomes and ribosomal subunits active in tRNA binding. *Methods in enzymology* **164**, 658-670 (1988).
72 Singer, M. F., Jones, O. W. & Nirenberg, M. W. The effect of secondary structure of the template activity of polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **49**, 392-399 (1963).
73 Zubay, G. In vitro synthesis of protein in microbial systems. *Annual review of genetics* **7**, 267-287, doi:10.1146/annurev.ge.07.120173.001411 (1973).
74 Ahn, J. H. *et al.* Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions. *Biochemical and biophysical research communications* **338**, 1346-1352, doi:10.1016/j.bbrc.2005.10.094 (2005).
75 Kigawa, T. *et al.* Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *Journal of structural and functional genomics* **5**, 63-68, doi:10.1023/B:JSFG.0000029204.57846.7d (2004).
76 Kim, D. M. & Choi, C. Y. A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane. *Biotechnology progress* **12**, 645-649, doi:10.1021/bp960052l (1996).
77 Kim, D. M., Kigawa, T., Choi, C. Y. & Yokoyama, S. A highly efficient cell-free protein synthesis system from Escherichia coli. *European journal of biochemistry / FEBS* **239**, 881-886 (1996).

78 Kim, R. G. & Choi, C. Y. A linear function for the approximation of accessible surface area of proteins. *Protein and peptide letters* **13**, 549-553 (2006).
79 Kim, T. W. *et al.* Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *Journal of biotechnology* **126**, 554-561, doi:10.1016/j.jbiotec.2006.05.014 (2006).
80 Kim, T. W., Kim, D. M. & Choi, C. Y. Rapid production of milligram quantities of proteins in a batch cell-free protein synthesis system. *Journal of biotechnology* **124**, 373-380, doi:10.1016/j.jbiotec.2005.12.030 (2006).
81 Kim, T. W. *et al.* Prolonged cell-free protein synthesis using dual energy sources: Combined use of creatine phosphate and glucose for the efficient supply of ATP and retarded accumulation of phosphate. *Biotechnology and bioengineering* **97**, 1510-1515, doi:10.1002/bit.21337 (2007).
82 Savage, D. F., Anderson, C. L., Robles-Colmenares, Y., Newby, Z. E. & Stroud, R. M. Cell-free complements in vivo expression of the E. coli membrane proteome. *Protein science : a publication of the Protein Society* **16**, 966-976, doi:10.1110/ps.062696307 (2007).
83 Roybal, G. A. & Jurica, M. S. Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. *Nucleic acids research* **38**, 6664-6672, doi:10.1093/nar/gkq494 (2010).
84 Kim, S. H. & Lin, R. J. Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Molecular and cellular biology* **16**, 6810-6819 (1996).
85 Warkocki, Z. *et al.* Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. *Nature structural & molecular biology* **16**, 1237-1243, doi:10.1038/nsmb.1729 (2009).
86 Yeh, T. C. *et al.* Splicing factor Cwc22 is required for the function of Prp2 and for the spliceosome to escape from a futile pathway. *Molecular and cellular biology* **31**, 43-53, doi:10.1128/MCB.00801-10 (2011).
87 NIH. *ImageJ: Image Processing and Analysis in Java*, <http://rsb.info.nih.gov/ij/> (
88 Schagger, H. & von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical biochemistry* **166**, 368-379 (1987).
89 Cooper, P. C. & Burgess, A. W. Simultaneous detection of 35S- and 32P-labeled proteins on electrophoretic gels. *Analytical biochemistry* **126**, 301-305 (1982).
90 Moghaddam, A. a. R., Nils *Use of polyethylene glycol for drying polyacrylamide gels to avoid cracking* <http://sciphu.com/2008/03/use-of-polyethylene-glycol-for-drying.html> (2008).
91 Hirao, I., Yoshizawa, S. & Miura, K. Stabilization of mRNA in an Escherichia coli cell-free translation system. *FEBS letters* **321**, 169-172 (1993).
92 Kaberdin, V. R. & Blasi, U. Translation initiation and the fate of bacterial mRNAs. *FEMS microbiology reviews* **30**, 967-979, doi:10.1111/j.1574-6976.2006.00043.x (2006).
93 Rauhut, R. & Klug, G. mRNA degradation in bacteria. *FEMS microbiology reviews* **23**, 353-370 (1999).

94 Uhler, S. A., Cai, D., Man, Y., Figge, C. & Walter, N. G. RNA degradation in cell extracts: real-time monitoring by fluorescence resonance energy transfer. *Journal of the American Chemical Society* **125**, 14230-14231, doi:10.1021/ja036854b (2003).
95 PerkinElmer. *Sulfur-35*, <http://www.perkinelmer.com/Catalog/Category/ID/sulfur%2035> (
96 Somogyi, P., Jenner, A. J., Brierley, I. & Inglis, S. C. Ribosomal pausing during translation of an RNA pseudoknot. *Molecular and cellular biology* **13**, 6931-6940 (1993).
97 Tu, C., Tzeng, T. H. & Bruenn, J. A. Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 8636-8640 (1992).
98 Vanzi, F., Vladimirov, S., Knudsen, C. R., Goldman, Y. E. & Cooperman, B. S. Protein synthesis by single ribosomes. *RNA* **9**, 1174-1179 (2003).
99 Sutcliffe, J. G. Nucleotide sequence of the ampicillin resistance gene of Escherichia coli plasmid pBR322. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 3737-3741 (1978).
100 Saraiya, A. A., Lamichhane, T. N., Chow, C. S., SantaLucia, J., Jr. & Cunningham, P. R. Identification and role of functionally important motifs in the 970 loop of Escherichia coli 16S ribosomal RNA. *Journal of molecular biology* **376**, 645-657, doi:10.1016/j.jmb.2007.11.102 (2008).
101 Dorywalska, M. *et al.* Site-specific labeling of the ribosome for single-molecule spectroscopy. *Nucleic acids research* **33**, 182-189, doi:10.1093/nar/gki151 (2005).
102 Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature reviews. Genetics* **2**, 919-929, doi:10.1038/35103511 (2001).
103 Bellaousov, S. & Mathews, D. H. ProbKnot: fast prediction of RNA secondary structure including pseudoknots. *RNA* **16**, 1870-1880, doi:10.1261/rna.2125310 (2010).
104 Leung, E. K., Suslov, N., Tuttle, N., Sengupta, R. & Piccirilli, J. A. The mechanism of peptidyl transfer catalysis by the ribosome. *Annual review of biochemistry* **80**, 527-555, doi:10.1146/annurev-biochem-082108-165150 (2011).
105 Lilley, D. M. Mechanisms of RNA catalysis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **366**, 2910-2917, doi:10.1098/rstb.2011.0132 (2011).
106 Cochrane, J. C. & Strobel, S. A. Catalytic strategies of self-cleaving ribozymes. *Accounts of chemical research* **41**, 1027-1035, doi:10.1021/ar800050c (2008).
107 Liang, B. & Li, H. Structures of ribonucleoprotein particle modification enzymes. *Quarterly reviews of biophysics* **44**, 95-122, doi:10.1017/S0033583510000235 (2011).
108 Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nature reviews. Genetics* **11**, 597-610, doi:10.1038/nrg2843 (2010).
109 Tijerina, P., Mohr, S. & Russell, R. DMS footprinting of structured RNAs and RNA-protein complexes. *Nature protocols* **2**, 2608-2623, doi:10.1038/nprot.2007.380 (2007).

110 Ziehler, W. A. a. E., D. R. in *Current Protocols in Nucleic Acid Chemistry* (John Wiley & Sons, Inc., 2000).
111 Woodson, S. A. in *Current Protocols in Nucleic Acid Chemistry* (John Wiley & Sons, Inc., 2000).
112 Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *Journal of the American Chemical Society* **127**, 4223-4231, doi:10.1021/ja043822v (2005).
113 Mortimer, S. A. & Weeks, K. M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *Journal of the American Chemical Society* **129**, 4144-4145, doi:10.1021/ja0704028 (2007).
114 Mortimer, S. A. & Weeks, K. M. Time-resolved RNA SHAPE chemistry. *Journal of the American Chemical Society* **130**, 16178-16180, doi:10.1021/ja8061216 (2008).
115 Mortimer, S. A. & Weeks, K. M. Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nature protocols* **4**, 1413-1421, doi:10.1038/nprot.2009.126 (2009).
116 Mortimer, S. A. & Weeks, K. M. C2'-endo nucleotides as molecular timers suggested by the folding of an RNA domain. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15622-15627, doi:10.1073/pnas.0901319106 (2009).
117 Gherghe, C. M., Mortimer, S. A., Krahn, J. M., Thompson, N. L. & Weeks, K. M. Slow conformational dynamics at C2'-endo nucleotides in RNA. *Journal of the American Chemical Society* **130**, 8884-8885, doi:10.1021/ja802691e (2008).
118 Harris, D. A. a. W., N. G. in *Current Protocols in Nucleic Acid Chemistry* (John Wiley & Sons, Inc., 2003).
119 Harris, D. A. a. W. N. G. in *Handbook of RNA Biochemistry* (ed A. Bindereif R. K. Hartmann, A. Schon, E. Westhof) 205-213 (Wiley-VCH, Weinheim, 2005).
120 Badorrek, C. S. & Weeks, K. M. Architecture of a gamma retroviral genomic RNA dimer. *Biochemistry* **45**, 12664-12672, doi:10.1021/bi060521k (2006).
121 Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 97-102, doi:10.1073/pnas.0806929106 (2009).
122 Gherghe, C. *et al.* Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19248-19253, doi:10.1073/pnas.1006897107 (2010).
123 Low, J. T. & Weeks, K. M. SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150-158, doi:10.1016/j.ymeth.2010.06.007 (2010).
124 Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716, doi:10.1038/nature08237 (2009).
125 Wilkinson, K. A. *et al.* High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS biology* **6**, e96, doi:10.1371/journal.pbio.0060096 (2008).

126 Vasa, S. M., Guex, N., Wilkinson, K. A., Weeks, K. M. & Giddings, M. C. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14**, 1979-1990, doi:10.1261/rna.1166808 (2008).
127 Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51-55, doi:10.1038/nature06684 (2008).
128 Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC bioinformatics* **11**, 129, doi:10.1186/1471-2105-11-129 (2010).
129 Mitra, S., Shcherbakova, I. V., Altman, R. B., Brenowitz, M. & Laederach, A. High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic acids research* **36**, e63, doi:10.1093/nar/gkn267 (2008).
130 Invitrogen. *Oligo Modification Options*, <http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Nucleic-Acid-Amplification-and-Expression-Profiling/Oligonucleotide-Design/Oligo-Ordering-Details/Oligo-Modification-Options.html?> (2011).
131 Biosystems, A. *GeneScan Reference Guide*, <http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_040961.pdf> (2011).
132 Li, S., Haces, A., Stupar, L., Gebeyehu, G. & Pless, R. C. Elimination of band compression in sequencing gels by the use of N4-methyl-2'-deoxycytidine 5'-triphosphate. *Nucleic acids research* **21**, 2709-2714 (1993).
133 Yamakawa, H., Nakajima, D. and Ohara, O. Identification of Sequence Motifs Causing Band Compressions on Human cDNA Sequencing. *DNA Research* **3**, 81-86 (1996).
134 Ennifar, E. *et al.* The crystal structure of UUCG tetraloop. *Journal of molecular biology* **304**, 35-42, doi:10.1006/jmbi.2000.4204 (2000).
135 Schroeder, S. J., Stone, J. W., Bleckley, S., Gibbons, T. & Mathews, D. M. Ensemble of secondary structures for encapsidated satellite tobacco mosaic virus RNA consistent with chemical probing and crystallography constraints. *Biophysical journal* **101**, 167-175, doi:10.1016/j.bpj.2011.05.053 (2011).
136 Lovett, P. S. Nascent peptide regulation of translation. *Journal of bacteriology* **176**, 6415-6417 (1994).
137 Walter, N. G., Yang, N. & Burke, J. M. Probing non-selective cation binding in the hairpin ribozyme with Tb(III). *Journal of molecular biology* **298**, 539-555, doi:10.1006/jmbi.2000.3691 (2000).
138 Newby Lambert, M. *et al.* Mg2+-induced compaction of single RNA molecules monitored by tethered particle microscopy. *Biophysical journal* **90**, 3672-3685, doi:10.1529/biophysj.105.067793 (2006).
139 Legiewicz, M. *et al.* The RNA transport element of the murine musD retrotransposon requires long-range intramolecular interactions for function. *The Journal of biological chemistry* **285**, 42097-42104, doi:10.1074/jbc.M110.182840 (2010).

140 Wilkinson, K. A. *et al.* Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA* **15**, 1314-1321, doi:10.1261/rna.1536209 (2009).
141 Seol, Y., Skinner, G. M., Visscher, K., Buhot, A. & Halperin, A. Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Physical review letters* **98**, 158103 (2007).
142 Wade, H. E. & Robinson, H. K. Magnesium ion-independent ribonucleic acid depolymerases in bacteria. *The Biochemical journal* **101**, 467-479 (1966).
143 Zaher, H. S. & Green, R. Quality control by the ribosome following peptide bond formation. *Nature* **457**, 161-166, doi:10.1038/nature07582 (2009).
144 Jelenc, P. C. & Kurland, C. G. Nucleoside triphosphate regeneration decreases the frequency of translation errors. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 3174-3178 (1979).