DETECTION OF BAXΔ2 READING FRAME SHIFT USING A DUAL LUCIFERASE REPORTER SYSTEM

BY

EVAN BEATTY

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science in Biology
in the Graduate College of the
Illinois Institute of Technology

Approved Jialijng Xiang
Advisor

Chicago, Illinois
May 2019

ACKNOWLEDGMENT

First and foremost: I would like to thank Dr. Jialing Xiang for her endless support and invaluable insight. As an advisor, professor, supervisor, and friend, she has nurtured my growth as a biologist and pushed for my exploration of fields outside of my comfort zone, some of which I have grown to enjoy immensely. She has proven to be a source of continuous inspiration over these past two years, and I hope to one day consider myself a scientist of equal merit. Finally, I'd like to thank her for her limitless patience, a quality that is too often overlooked in the academia of today.

I would also like to thank: all of the current and former members of the Xiang Lab at IIT, whose hard work and experience have made the research behind this thesis possible; the professors of the Department of Biology at IIT, who have made obtaining my Master's degree an enjoyable experience; and the members of my committee, who took the time out of their busy schedules to read and revise my thesis. Thanks, boss.

Last but not least, I would like to thank the members of my immediate and extended family that have supported me in my pursuit of a higher education. To Tom, Jamie, Colin, and Bear, thank you for everything. I dedicate this thesis to you.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

While initial studies of the pro-apoptotic Bcl-2-associated X protein isoform Δ2 (BAXΔ2) identified the combination of an alternative splicing event and a gene-level mutation as the prerequisites for biosynthesis in microsatellite unstable (MSI+) human colon cancer cells, no similar explanation existed to explain the presence of this protein in normal and normal adjacent tissues. To identify an alternative to the gene-level mutation in the absence of an MSI+ phenotype, we utilized a dual luciferase reporter assay designed to observe epigenetic recoding. Plasmid constructs containing the first two exons encoding BAXΔ2 were either transcribed and translated in vitro or transfected into *BAX*-negative human colon cancer cells. In both cases, assay of the protein products of the reporter genes demonstrate that a low level (2.82% in vitro, 4.43% in vivo) of all translational events which produce the protein product of an upstream reporter gene also produce the protein product of a downstream reporter gene. This occurs despite the two existing in different reading frames as a result of the *BAX* exons cloned between them. These results confirm that an epigenetic recoding event is able to salvage the *BAX* reading frame in cases where exon 2 has been excised, and further narrow down the potential mechanism involved to either transcriptional slippage or programmed ribosomal frameshifting.

CHAPTER 1

INTRODUCTION

1.1 Apoptosis and the Bcl-2 Family of Proteins

1.1.1 Bcl-2 and BAX. Apoptosis, or programmed cell death, is controlled by a tightly regulated set of genes and their corresponding protein products [44]. The Bcl-2-associated X protein (BAX) gene encodes several of these peptides, the most notable of which is isoform α (BAXα) [11, 16, 37, 40]. Alternative splicing of transcripts of the *BAX* gene allow for the production of a number of other proteins, known collectively as members of the BAX subfamily. Most members of the BAX subfamily, including BAXα, are encoded by some combination of six exons conserved within *BAX*. For example, transcripts containing the full set of six exons produce a functional BAXα protein upon translation. BAXα is a pro-apoptotic Class I member of the Bcl-2 family of proteins, as it contains the Bcl-2 homology (BH) domains BH1, BH2, and BH3, encoded in exon 4, exon 5, and exon 3, respectively (see Figure 1.1 B) [22]. The BH3 domain is especially important, as it is required for dimerization with other pro-apoptotic members of the Bcl-2 family of proteins. Homo- and hetero-dimerization of BAXα occurs upon activation by smaller proteins, including BH3 interacting-domain death agonist (BID) and BIM [37, 43]. These small proteins fall into a unique class within the Bcl-2 family, as the only conserved domain that they contain is that of BH3. Dimerization is followed by localization and interaction with the outer mitochondrial membrane, resulting in membrane permeabilization. In the case of BAXα, interaction with the mitochondrial membrane is likely mediated by a mitochondrial localization signal located near the N-terminus of the protein and encoded by exon 2 (see Figure 1.1 B, MTS: mitochondrial targeting sequence) [5, 36]. Membrane permeabilization causes molecules located within the mitochondrial intermembrane space to be released into the cytosol. Of these molecules, cytochrome *c* is of highest

importance; whether it be through the formation of heptameric caspase-activating complexes known as apoptosomes, or through the arrest of energy production caused by the absence of cytochrome *c* in the intermembrane space, release of this protein is followed by eventual cell death (see Figure 1.2) [21, 29].

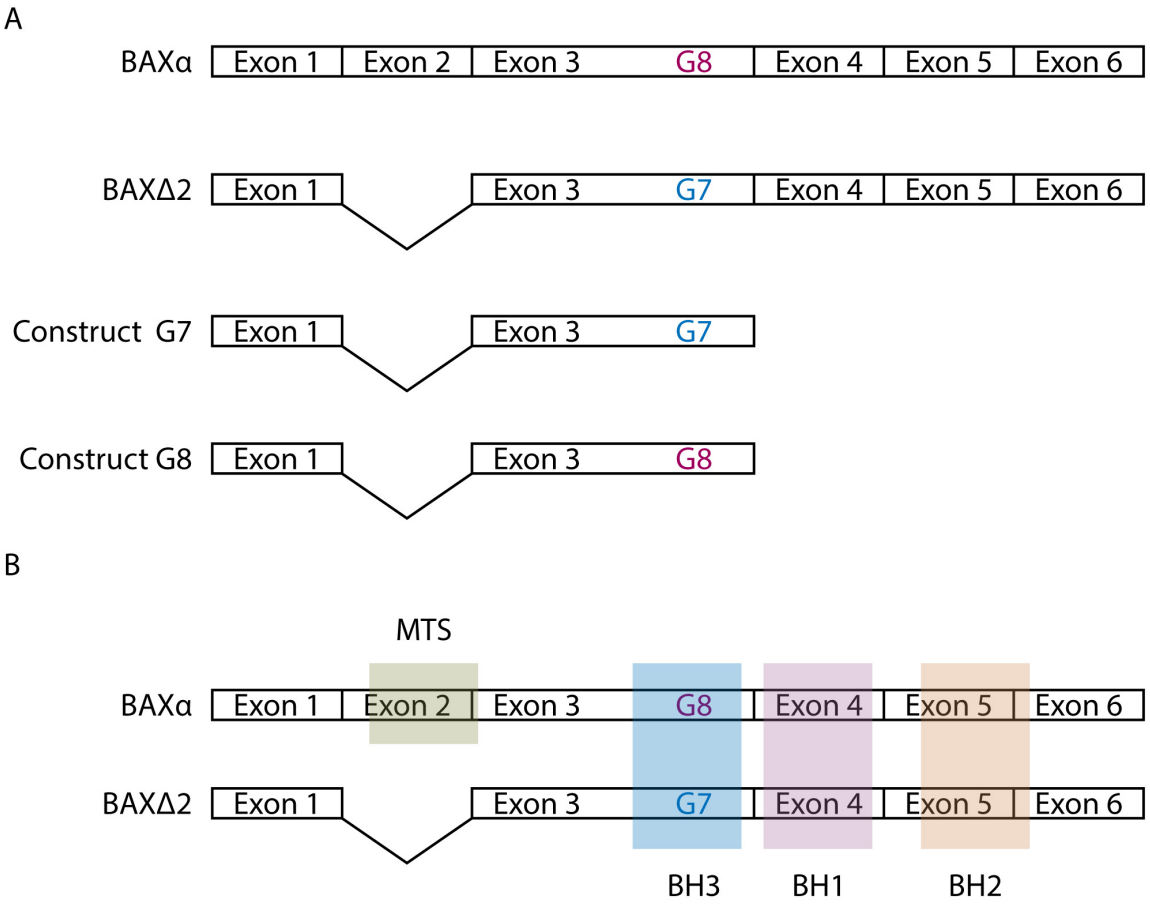

Figure 1.1. Graphical representation of important BAX transcripts and regions encoding functional domains

As mentioned previously, alternative splicing of *BAX* pre-mRNA can yield transcripts that encode proteins that are both functionally and structurally different from BAXα. A number of isoforms have been identified to date, with the majority falling into one of two categories. Isoforms such as BAX-δ and BAXΔ2 result from the translation of transcripts that lack one or more of the six constitutive *BAX* exons

[16, 31]. A second, less common group of isoforms result from the translation of transcripts that contain unexcised introns, or partial regions of them (BAX-ω) [46]. While none of the BAX isoforms have been studied to the same extent as BAXα, an increasing amount of evidence suggests that many retain some role in apoptosis. Not all, however, remain pro-apoptotic; in at least one case, the case of BAX-ω, alternative splicing results in an anti-apoptotic protein similar in function to Bcl-2.

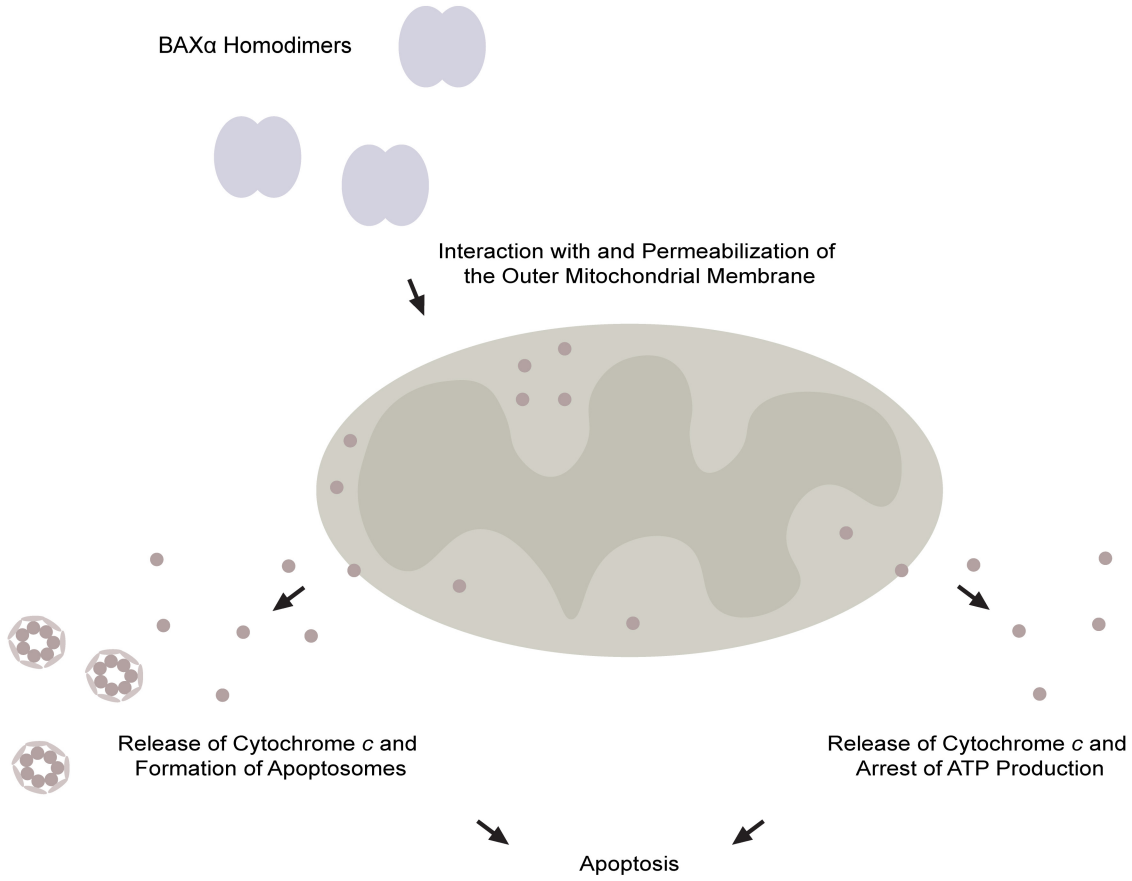

Figure 1.2. Graphical representation of BAXα-mediated cell death

1.1.2 BAXΔ2. While mutations occur naturally during replication, the mismatch repair system exists to ensure that few single nucleotide insertion/deletion mutations make it through each cycle. However, inactivating mutations within the genes of proteins related to the DNA mismatch repair system cause a cell phenotype characterized

by high rates of mutation, especially within repetitive stretches of nucleotides known as microsatellites. This phenotype, referred to as either microsatellite instability high (MSI-H) or microsatellite instability low (MSI-L) depending on the prevalence of mutations, has been used as a marker for the diagnosis and treatment planning of several types of cancer [1, 3, 17]. Among other genes, MSI-H colorectal tumor cells have been shown to contain a high rate of mutation at the point of a tract of repeated nucleotides within coding exon 3 of *BAX* [38]. This mutation typically exists as a single nucleotide insertion or deletion within a stretch of eight guanine residues; throughout the remainder of this paper, this stretch of guanines in exon 3 will be referred to as the *BAX* microsatellite. Translation of either variety of mutant transcript containing the six constitutive *BAX* exons results in a nonfunctional truncated protein, as the shift in reading frame induced by the mutation allows alternative stop codons to be shifted into frame. In the case of a mononucleotide insertion or deletion, translation is halted before the end of exon 3 by stop codons that would have normally been out of frame. MSI-H cells containing this mutation are thus expected to be BAX-negative, in the sense that functional BAXα proteins should not be able to be expressed. The mutation of the *BAX* microsatellite and the proceeding knockout of BAXα may therefore promote tumorigenesis and be advantageous to the survival of MSI+ cells, at least in the absence of BAXα alternatives.

BAXΔ2 is a member of the BAX subfamily of proteins originally identified in colorectal tumors exhibiting the MSI phenotype [16]. The BAXΔ2 transcript results from an alternative splicing event that causes excision of a large portion of exon 2 of *BAX* (see Figure 1.1 A), and was originally hypothesized to produce a functional protein only in the context of the previously described MSI-specific microsatellite mutation [15]. This MSI-specific mutation, a single nucleotide deletion within the *BAX* microsatellite, causes a shift out of the normal reading frame followed by premature termination during translation of the prototypical BAXα transcript. Excision of the

majority of exon 2, however, provides a mechanism by which the zero reading frame is able to be restored. Through the use of an acceptor site not normally utilized during the splicing of BAXα transcripts, all but the final two nucleotides of the canonical *BAX* exon 2 are spliced out [15]. At 50 nucleotides in length, the excised portion of exon 2 does not contain a perfect set of codon triplets; excision thus causes a shift into the -1 reading frame. This shift corrects for the downstream shift into the +1 reading frame, theoretically resulting in a protein with C and N termini identical to those of BAXα (reading frames will be explored in more depth in the following section). Unlike BAXα, however, the corrected protein lacks the amino acids encoded by the first 50 nucleotides of exon 2, and contains a short stretch of ten amino acids that are entirely novel to this isoform [14, 15, 16]. The region encoding the ten new amino acids is hypothesized to occur immediately following the point of the first (-1) shift in reading frame, and end as soon as the reading frame is restored at the mutated exon 3 microsatellite. In this way, alternative splicing acts to rescue the mutated *BAX* gene, allowing for the production of a protein similar, at least structurally, to BAXα in cells that would otherwise be considered BAX-negative.

BAXΔ2 shares similarities with BAXα not only in amino acid sequence, but in functionality as well. While the two proteins utilize different pathways, the end results are the same: activation of caspases and triggering of cell death. Unlike many of the members of the BAX family of proteins, BAXΔ2 has a weak relationship with the mitochondria. It does not act to directly permeabilize the outer mitochondrial membrane, and has been shown to avoid mitochondrial targeting and colocalization despite retention of the BH domains responsible for Bcl-2 family oligomerization. Instead, a BAXΔ2-positive phenotype corresponds with the formation of large protein agglomerates within the cytosol. While it has been determined that caspases (particularly caspases 3 and 8) play a role in BAXΔ2-mediated cell death, the complete pathway has yet to be determined. It has been suggested that these large cytosolic

agglomerates are, through some yet undetermined mechanism, able to interact with and activate the two caspases. The C-terminal α-helix is important for activation of caspase 8, and this activation sets off a cascade of caspase activations, but the specifics of this interaction remain to be elucidated. Further, while not caused by a direct interaction with the mitochondrial membrane, ectopic BAXΔ2 expression does correspond with the presence of cytosolic cytochrome c [16]. Experimental evidence from the use of a selective caspase inhibitors in BAXΔ2-positive cells supports the hypothesis that release from the mitochondrial intermembrane space is mediated by at least caspase 8, likely through the cleavage of BID and the interaction of tBID with the mitochondria [16, 28]. Thus, while straying somewhat from the typical mitochondria-mediated apoptotic pathway, BAXΔ2 allows for salvation of an apoptotic pathway in the absence of BAXα.

Because BAXΔ2 relies upon the presence of a mutation in the *BAX* microsatellite, it wouldn't be expected to be produced outside of MSI+ cells. The alternatively spliced transcript retains a low level of expression in cells lacking the *BAX* microsatellite deletion, but translation following the normal rules is expected to result in a truncated protein that is signaled for degradation [16]. Unexpectedly, recent data has shown that cells containing a protein recognized by a BAXΔ2-specific antibody (which recognizes the ten amino acids unique to BAXΔ2) are present at a low level throughout the body [16, 33, 34, 45]. These positive cells have been observed in otherwise normal and normal adjacent tissue samples, in which a gene-level mutation would not be expected at a statistically significant frequency [32]. Several mechanisms have been proposed to explain this phenomenon, but previous attempts to generate supporting data have not had much success.

1.2 Reading Frames

1.2.1 Introduction. Translation is the complex process by which the information encoded in cytosolic mRNA is converted into polypeptides of predictable sequence. The majority of this process is handled by large protein/RNA complexes called ribosomes, which move stepwise along an mRNA molecule. Assumption of adherence to the strict traditional translational guidelines allows for three distinct reading frames to be defined for any given mRNA. A first reading frame, deemed the 0 (zero) reading frame, consists of a start codon and all consecutive downstream nucleotide triplets; this is the collection of codons that a ribosome would read under normal circumstances. The other two reading frames, deemed the -1 and +1 reading frames, consist of the nucleotide triplets which would have been translated had the ribosome started one nucleotide upstream or downstream of the start codon, respectively. The definition of the three frames is dependent upon the ability of the ribosome to interpret nucleotide triplets (as opposed to duplets, quadruplets, etc.), as well as the ability of the ribosome to move exactly three nucleotides downstream per step. In this way, a single normal translational event can be expected to follow a single reading frame; definition is relative to the point of translation initiation, so standard translational events follow a standard zero reading frame. In practice, however, various cellular events allow for cases of modification and movement between reading frames. Further, departure from the standard translational guidelines is not necessarily a requirement of these events, as they are possible along every step of the central dogma of molecular biology.

1.2.2 Genetic Mutations. As the most basal mechanism of reading frame modification, DNA mutations have the potential to cause the greatest amount of damage. Mutations within the coding region of a gene are able to cause changes to the translational reading frame in cases where nucleotides are either inserted or

deleted. If the number of nucleotides does not correspond to a whole number of codons (cases of n $\equiv$ 1 (mod 3) or n $\equiv$ 2 (mod 3), with n being the number of inserted or deleted nucleotides), the perceived reading frames will experience a change from the point of the mutation onward. Mutations of this sort, if left unaccounted for, have the potential to change the identity of the N-termini of all proteins expressed from an affected gene. Accordingly, the rate of spontaneous mutation is kept negligibly low, especially when compared to that of errors at the level of translation [26]. An upper limit has been determined in the absence of the mechanisms usually in place to ensure replication fidelity, estimated to be between 10^{-7} to 10^{-8} mutations per base pair per replication cycle [25]. Under conditions in which these mechanisms are present, however, a somewhat consistent rate of approximately 0.003 spontaneous mutations per genome per replication cycle has been observed across both prokaryotic and eukaryotic organisms [8]. This number has been determined to be, somewhat surprisingly, independent of genome size. As such, spontaneous mutations do not contribute significantly to reading frame modifications in healthy cells, even after many rounds of replication.

The mismatch repair (MMR) system is responsible for regulation and repair of spontaneous mutations that occur during replication. However, MMR genes themselves are not immune to damage; unrepaired mutations that act to knockout these genes result in a phenotype marked by a heightened rate of spontaneous mutations per replication cycle. A hypothesized two mutations across the MMR locus are accompanied by a 10- to 10000-fold increase in the mutation rate [41]. Mutations in these genes have correspondingly been attributed to tumorigenesis, as the heightened rate increases the chances of further mutation to tumor suppressor genes. Thus, unlike cases involving healthy cells, spontaneous mutations may contribute substantially to reading frame modifications in tumor cells.

1.2.3 Transcriptional Errors and Slippage. Similar to the way that the insertion or deletion of nucleotides during replication has the potential to modify reading frames, the insertion or deletion of nucleotides during transcription does as well. The RNA polymerase class of proteins is almost entirely responsible for errors of this sort, with each of the individual members (RNA polymerase I, RNA polymerase II, etc.) having different characteristics and rates of error. When dealing with the production of proteins by eukaryotic organisms, RNA polymerase II is of highest importance. As RNA polymerase II is responsible for the production of the bulk of pre-mRNA (yielding true mRNA after processing), errors caused by it directly affect proteins at an early stage of production. Further, individual errors are amplified at the level of translation, as a single transcript may be translated a number of times before being degraded. RNA polymerase II has been estimated to have a total error rate of approximately 3.9×10^{-6} errors per base pair, calculated from studies of the *S. cerevisiae* transcriptome [12]. Single nucleotide substitution mutations comprise the bulk of these errors, while, together, insertion and deletion mutations are responsible for approximately one quarter ($\sim 9.5 \times 10^{-7}$ indel mutations per base pair) of the total error rate [12]. The transcription-level indel mutation rate, while still quite low, resembles that of the estimated gene-level mutation upper limit. Taking into consideration the observed genetic mutation rate per replication cycle, it can be suggested that the effect of transcriptional errors on translational reading frames is much greater than that of genetic mutations.

The rate of error increases and decreases dependent upon the context in which a specific RNA polymerase exists. Nucleotide sequence plays a role in this rate, with certain mutations favoring certain sequences and individual mutations each having their own rates of occurrence. For example, among substitution mutations, the replacement of guanine with adenine is (by an order of magnitude) the most common form [12]. Similarily, in the case of insertion and deletion mutations in nascent tran-

scripts, favor is given to regions of mono- and di-nucleotide repeats [12]. The term "transcriptional slippage" has been used to describe the mechanism responsible for indel mutations in these regions, as it has been suggested to be caused by the abnormal movement of RNA polymerase during transcriptional elongation. Early evidence pointed towards denaturation at the point of the RNA/DNA interface, followed quickly by rehybridization in either the upstream (causing nucleotide insertion) or downstream (causing nucleotide deletion) direction [42]. While the RNA polymerase would have physically moved along the template strand, base pair interactions would have the potential to remain unchanged as a result of the repetitive nature of regions containing these mono- and di-nucleotide repeats. This occurs most commonly along repeats of the nucleotide adenine or thymine (or a combination of the two), as the weaker A/U and T/A interactions allow for the least amount of resistance to denaturation [42].

Due to the destructive nature of frameshift mutations, there are mechanisms in place to ensure that many frameshifted transcripts never complete translation. The eukaryotic nonsense-mediated decay (NMD) pathway acts to identify and dispose of transcripts containing premature stop codons [18]. These premature stop codons normally exist outside of the zero-reading frame (the frequency of stop codons in non-coding regions is quite high), but become shifted in due to errors that result in frameshifting. The origin of a particular mutation is irrelevant, allowing the NMD pathway to correct for errors at the level of both replication and transcription. Different organisms follow different guidelines for premature stop codon identification, but all share some key similarities and are dependent upon the process of translational termination. Premature stop codons are identified by their proximity to exon-exon junctions; in mammals, an NMD complex composed of a number of proteins is able to form around a junction only when it is further than 50 nucleotides upstream of the interpreted stop codon [18, 27]. Distance from this stop codon corresponds to distance

from the ribosome, suggesting that complex formation may be sterically hindered when the junction is near. The complex acts to dissociate any proteins bound to the transcript (including the ribosome), and follows with complete mRNA degradation [18].

1.2.4 Translational Errors and Programmed Frameshifting. The ribosome exists solely to mediate translation of mRNA into peptide molecules. This large, multisubunit complex dates back to the time of the last common universal ancestor, and is one of the few constructs conserved in both structure and function across all branches of life [10]. While minor divergences have occurred, the ribosome remains largely unchanged; a large subunit composed of both proteins and ribosomal RNA contains a peptidyltransferase center, while a smaller subunit contains the decoding center responsible for mRNA/tRNA hybridization. Aminoacyl-tRNA molecules from the surrounding environment are able to freely enter the first of three sites (referred to as the A site) at the peptidyltransferase center, in which interaction with a bound mRNA molecule is possible. When the mRNA exposed at this site comes into contact with a complementary anti-codon tRNA loop, hybridization is able to occur. Movement of the ribosome along an mRNA molecule occurs in steps of three nucleotides (the length of a single codon), and is accompanied by the movement of any hybridized tRNA molecules. In the case of an A site hybridized aminoacyl tRNA, movement of the ribosome causes movement into a second site (referred to as the P site) within the peptidyltransferase center. In this second site, the aminoacyl portion of the aminoacyl-tRNA molecule is able to form a peptide bond with the C-terminal amino acid of a nascent peptide chain. Further translocation of the ribosome shifts the tRNA and new C-terminal amino acid, now separated, to a third site (referred to as the E site) when present. Not all organisms have a well-defined E site, and its primary role is to simply house the C-terminal amino acid of the nascent peptide chain before further movement shifts it into the exit channel of the ribosome. Ter-

mination of translation is triggered by interactions between release factors and stop codons in the A site. Like aminoacyl tRNA molecules, the release factors contain regions that are able to specifically recognize the nucleotides which encode the three stop codons. Termination completes with dissociation of the nascent peptide, mRNA template, and now-free ribosome.

Unlike mutations at the genetic and transcriptional levels, translational errors are confined to individual nascent peptides. The ribosome has no way of introducing new mutations to existing transcripts, and thus errors during one translational event are not conveyed to later translation events. Despite this, the total translational error rate has been determined to be far higher than that of either replication or transcription. Missense errors, those resulting from the incorporation of incorrect aminoacyl-tRNA molecules, have been hypothesized to occur at a rate of approximately 5×10^{-4} errors per codon in bacteria [6, 26]. While several orders of magnitude higher than pre-translational error rates, the effect of such errors to the function of a protein tends to be quite limited, as reading frames remain conserved [26]. A second category of mechanisms which cause lapses in translational fidelity and translation of portions of non-zero reading frames, however, have the potential to cause a substantial amount more change to a nascent peptide. Mechanisms of this sort are known generally as translational recoding events, and occur to some extent in every known organism [7, 23]. Early estimates place the total rate of these errors between approximately 5×10^{-5} and 5×10^{-4} errors per codon in both bacteria and yeast, but the individual rates of each of these mechanisms varies drastically [4, 24, 26]. The most well-studied of these events are programmed ribosomal frameshifting and translational initiation at alternative start codons, both of which have the potential to cause the ribosome to translate codons in non-zero reading frames. While unregulated movement between reading frames has the potential to cause massive changes to the proteome of an organism, the regulation of these mechanisms has been shown to be beneficial under

certain circumstances. Several examples of this can be seen upon analysis of certain viral genomes, in which the minimization of genetic material through the incorporation of dual encoding regions and the regulation of protein expression have been suggested to be accomplished to some extent through these means [7, 20].

Programmed ribosomal frameshifting (PRF) is a collection of conceptually related events which result in the abnormal movement of the ribosome along a template mRNA strand. During these events, it has been hypothesized that the bonds between the tRNA and mRNA molecules in the A and P sites break transiently, allowing for reformation of the bonds at some different point along the mRNA molecule. The most common of these events are the -1 programmed ribosomal frameshifts, which cause a shift of the ribosome by a single nucleotide in the upstream (5') direction. Functionally, nascent peptides that result from these events are identical to those that might have been formed by a single nucleotide insertion during either replication or transcription. Translation continues normally following the event, but the ribosome reads the -1 frame from the point of the shift onward.

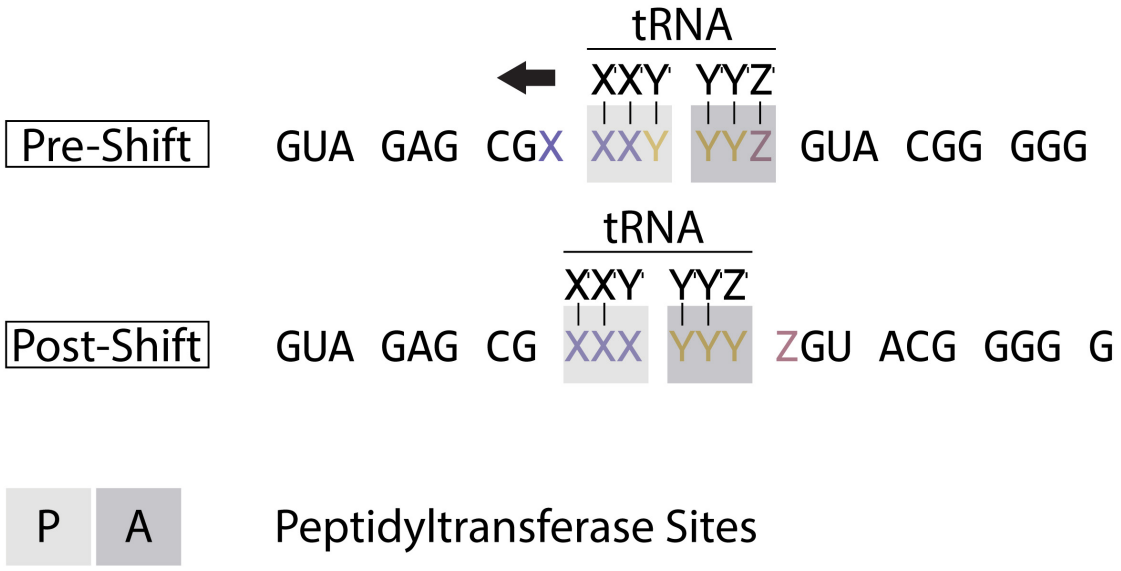

Figure 1.3. Graphical representation of -1 PRF during translation of a slippery sequence

As the name suggests, these events do not occur entirely randomly during translation, and instead have a set context in which the chance of occurrence is heightened. For -1 PRF, this context is dependent upon the primary and secondary structures of the template mRNA molecule, and most commonly relies upon the presence of two well-defined motifs. The first, a sequence of seven nucleotides, is the region of the mRNA molecule that is exposed in the decoding center of the small ribosomal subunit during the actual frameshifting event. This motif is known as a slippery sequence, and matches the pattern X XXY YYZ. The first triplet (XXY) is exposed to the ribosomal P site and the second triplet (YYZ) is exposed to the ribosomal A site [4, 7, 9]. While X can be any nucleotide, Y and Z have been shown to favor adenine and either uracil or cytosine, respectively [7]. This motif allows for stable rehybridization of tRNA and mRNA at the decoding center, as two out of the three nucleotide-nucleotide interactions at each site remain unchanged. Upon shifting into the -1 reading frame, the triplet XXX is exposed to the ribosomal P site and the triplet YYY is exposed to the ribosomal A site. As can be seen upon comparison to the original triplets exposed at these sites, only the 3'-most nucleotide has the potential for a change in identity (see Figure 1.3 for a graphical representation of a -1 PRF event). While changes of this sort to the first two nucleotides might cause the inhibition of rehybridization, the interaction of the third nucleotide is somewhat more flexible. "Wobble" base pairing at this point will still, in many cases, allow for hybridization to occur. The second motif is an mRNA secondary structure that occurs a short distance downstream of the slippery sequence. While a defined sequence for this motif is not conserved across sites of -1 PRF, these downstream regions most commonly form either a pseudoknot or a stem loop. These large secondary structures have been hypothesized to sterically hinder the ribosome, resulting in temporary translational stalling while the structure is cleared [7, 35]. Translational arrest of this sort is followed by the deformation of the peptidyltransferase center, which in

turn has the potential to cause temporary RNA de-hybridization [35]. When the two motifs are combined, frameshifted proteins are able to be produced at a consistent (but typically low) rate.

Less common than -1 programmed ribosomal frameshifting events, and far less well studied, +1 programmed ribosomal frameshifting events have also been observed. As opposed to shifting the ribosome by a single nucleotide in the upstream direction, +1 PRF events result in a shift of the ribosome by a single nucleotide in the downstream direction. This results in a single nucleotide being functionally skipped by the ribosome, and produces a nascent peptide similar to that which might have been produced by a single nucleotide deletion. Events of this sort do not rely on the slippery sequence and secondary structure dependent mechanism commonly observed in cases of -1 PRF, and instead vary quite drastically between cases. One of the most well studied of these cases, that of mammalian ornithine decarboxylase (ODC) and ODC antizyme, has been shown to be reliant upon the presence of the small cytosolic polyamines. Spermidine, one of these polyamines, has been shown to interact with the ribosome in a way which induces a +1 PRF event during translation of the antizyme gene [39]. Frameshifting occurs at a low rate ($\sim$3%) in vitro in the absence (or near absence, as some is likely present in the reticulocyte lysate) of spermidine, but this rate can be increased tenfold upon artificial supplementation [13]. As well as by regulation by these small molecules, a role in this +1 PRF event is played by the antizyme transcript. Similar to the typical -1 PRF events, the antizyme transcript has been shown to contain a short nucleotide sequence crucial to frameshifting. This sequence has been posited to be similar in function to a slippery sequence, despite an identity not matching the X XXY YYZ pattern. Containing the first stop codon of the antizyme gene, as well as the two codons which directly precede it, deletion of this region results in the complete destruction of PRF efficiency. As interpretation of stop codons has the tendency to stall the ribosome, the zero-frame antizyme stop

codon might then play a role similar to that of the large downstream mRNA secondary structures present in many cases of -1 PRF. Stalling of any sort, whether stop codon or mRNA secondary structure dependent, creates a window in which RNA de-hybridization has the ability to occur.

1.3 Dual Reporter System for Frameshift Analysis

Studies of translational recoding events have utilized reporter assays from a very early date. However, until the 1990's, a standardized system that could be assayed in vitro simply did not exist. In 1998, a team of scientists at the University of Utah published a paper describing a plasmid which aimed to solve this problem. Called p2luc, this plasmid contained two genes encoding luciferase proteins which were luminescent in the presence of substrate [13]. *Renilla* luciferase, encoded by a gene located just upstream of a cloning site, catalyzes the oxidation of luciferin. Similarly, firefly luciferase is encoded by a gene located just downstream of the same cloning site, and acts to catalyze the oxidation of coelenterazine. Genes encoding proteins that are either radioactive or quantifiable through some enzymatic assay had historically been chosen for these assays, so incorporation of luciferase genes allowed for a level of convenience. Measurement of the light released by the reactions between enzyme and substrate allowed for an accurate measurement of the relative amounts of each protein present. Both oxidation reactions produce light across a broad spectrum, so measurements needed to be separated temporally. By cloning a portion of the coding region of a gene of interest between the two luciferase genes, an interpretation of the rate of translation of the coding region could be made. Somewhat more interestingly, by cloning a coding region into p2luc in such a way that caused the movement of the downstream firefly luciferase gene into a non-zero reading frame, measurements of frameshifting could be inferred. While this system was originally designed to measure levels of programmed ribosomal frameshifting, the assay does

not by itself rule out the possibility of transcriptional errors.

The p2luc-based dual luciferase reporter assay was not without fault, as protein products were never produced as separate entities. Complete translation starting at the start codon of the *Renilla* luciferase gene and ending at the termination codon of the firefly luciferase gene resulted in production of a large conjoined protein. When present, the protein product of the cloned region between the luciferases was also part of this single polypeptide. The potential effect of the cloned region on the activity of the luciferases, or of one luciferase on the activity of the other, is something that remains hard to measure but likely contributed to error in early measurements of translational recoding. In 2017, Dr. Gary Loughran and Dr. John Atkins published a paper which described a derivative plasmid which solved several problems encountered by the original, including the problem of protein separation [30]. While the overall structure has been conserved, 2A sequences have been added directly downstream and upstream of the *Renilla* luciferase and firefly luciferase genes, respectively [30]. These sequences encode a short stretch of amino acids which are able to self-cleave as soon as they've been translated, separating both luciferase proteins from each other, as well as from the product of the cloned gene [2]. This, in theory, should result in a more accurate measurement of the levels of each luciferase protein, as they are now produced in a form closer to that of their natural states. Likewise, a change of this sort likely increases the level of consistency possible over a wide range of genes.

1.4 Research Overview

The aim of the experiment described in this thesis was to determine whether or not a reading frameshift can be introduced during transcription or translation of the first ∼190 nucleotides of the hypothesized BAXΔ2 coding region. The observation of (or of the absence of) an event of this sort could lead to a better understanding of the prerequisites required for BAXΔ2 biosynthesis in the cells of normal and normal

adjacent tissue. Prior studies by Mañas et al. have led to the suggestion that these events could provide a mechanism by which an earlier frameshift introduced by an alternative splicing event could be corrected [32, 36]. Here, we employ a dual luciferase reporter assay designed to observe events of this sort.

CHAPTER 2

MATERIAL AND METHODS

2.1 Material

Single stranded DNA oligomers for use as primers in PCR were provided by Integrated DNA Technologies. The DNA purification kit used to purify PCR products was purchased from Feldan Therapeutics, and a magnetic bead-based gel extraction kit used to purify DNA from polyacrylamide gels was purchased from AvanBio. The TnT Quick Coupled Transcription/Translation System and accompanying GloMax 20/20 Luminometer were purchased from Promega. New England Biolabs provided all restriction enzymes, as well as the buffers used in the restriction reactions. All buffers used during DNA miniprep and DNA large prep, as well as the QIAGEN-tip 100 plasmid purification kit, were purchased from Qiagen. Lipofectamine 3000, P3000 Enhancer Reagent, and Opti-MEM Reduced Serum Media were purchased from Thermo Fisher Scientific. Sanger sequencing of DNA samples was done by the University of Chicago Comprehensive Cancer Center DNA Sequencing Facility. Previously generated constructs were developed by members of the Xiang Lab, in the Department of Biology at IIT. The human colon cancer cell line HCT116 was originally obtained from the American Type Culture Collection (ATCC), and the BAX-negative subline #28 used in this study was generated previously by the Xiang lab. pSGDluc, the plasmid used in the dual luciferase reporter assay, was designed by G. Loughran et al., and provided by the laboratory of Dr. John Atkins at the University College Cork, Ireland [30].

2.2 Cloning

To clone BAXΔ2 constructs into the dual luciferase reporter vector pSGDluc, we designed a set of two restriction site-containing primers (labeled 888U and 888R,

see Appendix A). These two primers were used to amplify a region of two BAXΔ2 cDNA constructs that were chosen for use in the assay as a result of the identity of their exon 3 microsatellite. Construct-containing plasmids were diluted in clean TE Buffer to a final concentration of 10 ng/μL. For each construct, 25 μL of PCR master mix, 2 μL of 10 mM forward primer, 2 μL of 10 mM reverse primer, 2 μL of 10 ng/μL plasmid DNA, and 19 μL of ddH_2O were combined in a PCR tube. PCR reactions were carried out in a thermocycler using a standard 30 cycle PCR protocol with the following modifications: 65 °C annealing temperature, 30 second extension time. PCR products were purified using a PCR Purification Kit, as instructed by the manual. 27 μL of each purified PCR product was combined with 3.2 μL of 10X NEBuffer 2, 1 μL of 5 unit/μL PspXI, and 1 μL of 5 unit/μL HindIII. At the same time, 2 μL of the 1000 ng/μL pSGDluc was combined with 3.2 μL of NEBuffer 2, 1 μL of 5 unit/μL PspXI, 1 μL of 5 unit/μL HindIII, and 26 μL of dH_2O. These three samples were heated at 37 °C for one hour to allow for digestion. The entirety of each digested sample was run on a 1% TAE gel, and digested samples were extracted and purified using the AvanBio gel extraction kit, as instructed by the manual. Ligation of digested constructs into the plasmid pSGDluc was accomplished through the use of T4 Ligase. Two μL of digested construct, 1.2 μL of 10X Ligation Buffer, 1 μL of T4 Ligase, 1 μL of digested vector, and dH_2O to a total reaction volume of 12 μL were combined in a microcentrifuge tube and incubated at 14 °C overnight to allow for ligation.

2.3 DNA Large-Scale Preparation

Two pSGDluc constructs were prepared in Section 2.2. These two constructs were transformed into supercompetent *E. coli* and grown on LB agar plates containing ampicillin for selection of transformants. Transformation followed a standard heat-shock protocol, starting with 50 μL of frozen DH5α *E. coli*. Cells were thawed on ice,

then transferred to a sterile test tube along with 3 μL of the appropriate plasmid construct. Cells were incubated on ice for 30 minutes, then placed in a beaker containing approximately 500 mL of water at 42 °C for exactly 42 seconds. Following heat shock, cells were allowed to recover on ice for five minutes before addition of 1 mL of Super Optimal Broth with Catabolite repressor (SOC) medium. Transformed cells were incubated at 37 °C while shaking at approximately 250 RPM for one hour. Following this incubation period, the transformed cells were pelleted to allow for removal of 900 μL of supernatant (leaving approximately 150 μL of supernatant behind). Pellets were resuspended in the remaining supernatant, plated on LB agar plates containing ampicillin, and allowed to grow at 37 °C overnight.

In each case, four colonies were translocated into 2 mL of fresh LB media containing ampicillin, and grown overnight in a 37 °C shaking incubator. A portion of each was spun down, and plasmid DNA was extracted from the pellets using a protocol designed by Qiagen and described in the miniprep section of their Plasmid Purification Handbook. The protocol uses a three-buffer system: Buffer P1 (50 mM Tris-Cl, pH 8.0, 10 mM EDTA, 100 μg/ml RNase A) to resuspend the pelleted cells, Buffer P2 (200 mM NaOH, 1% SDS (w/v)) to lyse the cells, and Buffer P3 (3.0 M potassium acetate, pH 5.5) to stop the lysis reaction and precipitate genomic DNA and cellular debris. Separation of supernatant from precipitate was possible after centrifugation, and further cleaning and purification of plasmid DNA was accomplished through repeated isopropanol washes. After a final wash, the reaction tubes were placed on a 37 °C heating block to allow all remaining isopropanol to evaporate. DNA was dissolved in 10 μL of clean TE, and a small portion of each sample was digested in PspXI and HindIII for verification of transformation. Success was further confirmed via Sanger sequencing.

One confirmed transformant from each of the two groups was chosen for large-

scale DNA preparation. Five μL of each transformant, stored previously in a 4 °C refrigerator, was plated on a fresh LB agar plate containing ampicillin. Plates were grown overnight in a 37 °C shaking incubator, so that colonies could be picked up the following day. Individual colonies were transferred to sterile test tubes containing 2 mL of fresh liquid LB media and 2 μL of 100 mg/mL ampicillin. Samples were allowed to grow in a 37 °C shaking incubator (shaking at roughly 250 RPM) for 6 hours. After incubation, the entirety of each sample was transferred to an Erlenmeyer flask containing approximately 150 mL of sterile LB media and 150 μL of 100 mg/mL ampicillin. Samples were allowed to grow in a 37 °C shaking incubator (shaking at roughly 250 RPM) overnight. After incubation, the entirety of each sample was centrifuged in a JA-14 rotor for 10 minutes at 4 °C and 4000 RPM to pellet cells. The same three-buffer system designed by Qiagen was again used for large preparation of plasmid constructs. Pelleted cells were resuspended in 4 mL of buffer P1 (containing 100 μg/mL RNase) and lysed with 4 mL of buffer P2. Incubation in lysis buffer P2 was done at room temperature for five minutes. Following this incubation, 4 mL of buffer P3 was added to precipitate genomic DNA and cellular debris. Incubation in precipitation buffer P3 was done on ice for approximately fifteen minutes. Following this second incubation, the liquid portion of each sample was centrifuge at 12000 RPM using the JA-20 rotor for 30 minutes at 4 °C. Plasmid DNA was purified from the supernatant through the use of the QIAGEN-tip 100 column purification kit. Being careful to transfer only the liquid portion of the supernatant to the column, samples were washed twice with 10 mL of QC buffer and plasmid DNA eluted with 5 mL of QF buffer, as instructed by the Qiagen Plasmid Purification Handbook. Plasmid DNA was further purified through the addition of 3.5 mL of 99% isopropanol, centrifugation at 12000 RPM in the JA-20 rotor for 30 minutes at 4 °C, and repeated washing of the pellet with 75% ethanol. All remaining ethanol was evaporated through incubation on a 37 °C heating block, and pellets containing purified plasmid DNA were dissolved

in 50 μL of clean TE.

2.4 Cell Culture and Transfection

Live HCT116 cells were transfected through the use of Lipofectamine 3000 and a standard lipofection protocol. Cells were grown in a 6-well plate until approximately 80% confluent before transfection. For each well, two tubes were prepared: one contained 125 μL of Opti-MEM Reduced Serum Media and 3.75 μL of Lipofectamine 3000, while the other contained 125 μL of Opti-MEM Reduced Serum Media, 5 μL of P3000 Enhancer Reagent, and approximately 2500 ng of plasmid DNA. Each tube was mixed separately, then combined with the other and mixed again. After the second mixing step, the combined solutions were allowed to incubate at room temperature for ten minutes. While incubating, the cells were prepared for transfection. Old media was removed, and the cells were washed once with phosphate buffer solution (PBS). Upon removal of PBS, 250 μL of the previously prepared Lipofectamine solution was added drop-wise to each individual well. Fresh Opti-MEM Reduced Serum Medium was added to a final volume of 1.5 mL in each well, and the cells were allowed to incubate at 37 °C for 24 hours. Following incubation, lysis was accomplished through the use of a passive lysis buffer (PLB) provided with the dual luciferase reporter assay kit. Old media was removed from each well of the 6-well plate, and cells were washed once with PBS. Following the wash, 500 μL of PLB was added to each well, and the cells were allowed to incubate for 15 minutes. Cells were then scraped using a plastic cell lifter and transferred to microcentrifuge tubes so that they could easily be subjected to two freeze-thaw cycles. Each cycle consisted of a period of freezing at -80 °C followed by a period of thawing on ice. The use of freeze-thaw cycles here was done to ensure complete lysis of the cells in our sample. Following the second freeze-thaw cycle, the samples were considered ready for assay without any further cleaning or manipulation.

2.5 In Vitro Transcription and Translation

In Vitro Transcription and Translation (IVTT) was accomplished through the use of a TnT Quick Coupled Transcription/Translation System. This system was dependent on a proprietary TnT T7 Quick Master Mix containing, at the very least, reticulocyte lysate, T7 RNA Polymerase, dNTP's, and each of the essential amino acids with the exception of methionine. A modified version of the official protocol was used: 20 μL of TnT T7 Quick Master Mix, 0.5 μL of 1 mM Methionine, 3.5 μL of nuclease-free water, and 1 μL of plasmid DNA were combined in a microcentrifuge tube. This solution was then heated to 30 °C for 90 minutes on an AccuBlock Digital Dry Bath to allow for the transcription and translation reactions to be carried out. These samples, after the completion of the heating phase, were considered ready for assay without any further cleaning or manipulation.

2.6 Dual Luciferase Reporter Assay

Substrates were provided as part of a Dual Luciferase Reporter Assay System. The luciferase assay was performed on a GloMax 20/20 Luminometer, using the pre-programmed DUALGLO protocol. On days in which an assay was performed, enough 50X Stop & Glo Substrate was combined with Stop & Glo Buffer in a 1:50 (v/v) ratio to run all planned assays. Luciferase Assay Reagent II was prepared once per kit by combining the entirety of the provided sample of lyophilized Luciferase Assay Substrate with 10 mL of Luciferase Assay Buffer II. The prepared reagent was then aliquoted and frozen at -80 °C, and enough to complete the planned assays was thawed each day. For the assay, 100 μL of Luciferase Assay Reagent II was added to a microcentrifuge tube containing 1 μL of translation product (described in Section 2.4 and Section 2.5). The solution was pipetted up and down to mix, and immediately assayed for firefly luciferase activity. The microcentrifuge tube was then removed from the luminometer so that 100 μL of 1X Stop & Glo Reagent could be added. The

solution was vortexed for three seconds to mix, and immediately assayed for *Renilla* luciferase activity. Translation products from each of the two preparation methods were run through the assay in triplicate on three separate occasions. Three values were calculated by the luminometer: firefly luciferase luminescence, *Renilla* luciferase luminescence, and the ratio of *Renilla* luciferase luminescence to firefly luciferase luminescence. The values were entered into a Microsoft Excel spreadsheet so that they could be analyzed. Entered values were verified by calculation of the *Renilla* luciferase to firefly luciferase luminescence ratio and comparison to the ratio given by the luminometer.

CHAPTER 3

RESULTS

3.1 Design and Construction of Luciferase Reporter Vectors

Two variants of the BAXΔ2 protein coding region were chosen to be cloned into the vector pSGDluc. In the first construct (referred to as both G8/pSGDluc and the Construct G8 throughout the remainder of this paper), the first 191 nucleotides of the BAXΔ2 coding region were cloned into pSGDluc. A mutated variant of exon 3 contained a silent T to C point mutation (N21N) to prevent ribosome assembly and translation initiation at an out of frame start codon without affecting the zero-frame amino acid sequence. In the second construct (referred to as both G7/pSGDluc and the Construct G7 throughout the remainder of this paper), the first 190 nucleotides of the BAXΔ2 coding region were cloned into pSGDluc. A mutated variant of exon 3 contained a single-nucleotide deletion at the point of the *BAX* microsatellite, decreasing its length from eight nucleotides to seven. In both constructs, the start codon of exon 1 was converted to a phenylalanine codon through two individual point mutations, changing ATG to TTC (M1P), to prevent ribosome assembly and translation initiation at a point other than the start codon of the pSGDluc *Renilla* luciferase gene. The first construct, G8/pSGDluc, was designed in a way that placed a second (downstream) luciferase gene out of frame with a first (upstream). This allowed for observation of the presence or absence of events which caused a change in the reading frame resulting in translation of the second luciferase gene. The second construct, G7/pSGDluc, was designed in a way that placed both luciferase genes in frame with each other, and acted as a positive control for the experiment.

The polymerase chain reaction (PCR) was used to amplify each of the two insert regions from previously prepared constructs (Construct G8 from construct 897 and Construct G7 from construct 896) using the primers 888U and 888R (see Table

Table 3.1. Primers

Primer Name	Sequence
888U	CCCTACTCGAGCTTCGACGGGTCCGGGGAGCAGC
888R	GGGAAGCTTCCTCTGCAGCTCCATGTTACTGTCCAG
F2AU	GTGAAACAGACTTTGAATTTTGACCTTCTTA

3.1). More information on these primers can be found in Appendix A. Amplification allowed for the incorporation of the restriction sites PspXI (on the 5'-end of the insert) and HindIII (on the 3'-end of the insert) for use in ligation. Following aplification, the expression vector pSGDluc and each of the two PCR products were digested by the restriction endonucleases PspXI and HindIII. Digested DNA was run on a 1% agarose TAE gel against a 1 KB Plus DNA Ladder (Figure 3.1) for validation of size and purity of the digested DNA. From this gel, an approximation was made of the relative concentrations of each sample, which was used to determine the amounts to use in the following ligation reaction. White arrows indicate the location of digested PCR fragments, and the size of the ladder fragments are indicated on the left of the image (in units of base pairs). Each of the two digested insert samples (columns 2 and 3, from the left) contain a band near the 200 base pair ladder band, as would be expected after digestion with PspXI and HindIII. The two digested inserts were expected to be 190 and 191 base pairs in length. The digested expression vector sample (column 5, from the left) contains a single band above the 5000 base pair ladder band but below the 10000 base pair ladder band, as would be expected from digestion of the vector with PspXI and HindIII. The digested vector was expected to be 5844 base pairs in length.

Following ligation, DNA miniprep was done on four transformed colonies per construct (designated $G7_{1-4}$ and $G8_{1-4}$, in reference to the plasmid constructs with which they were transformed). Approximately 15% of the purified plasmid DNA

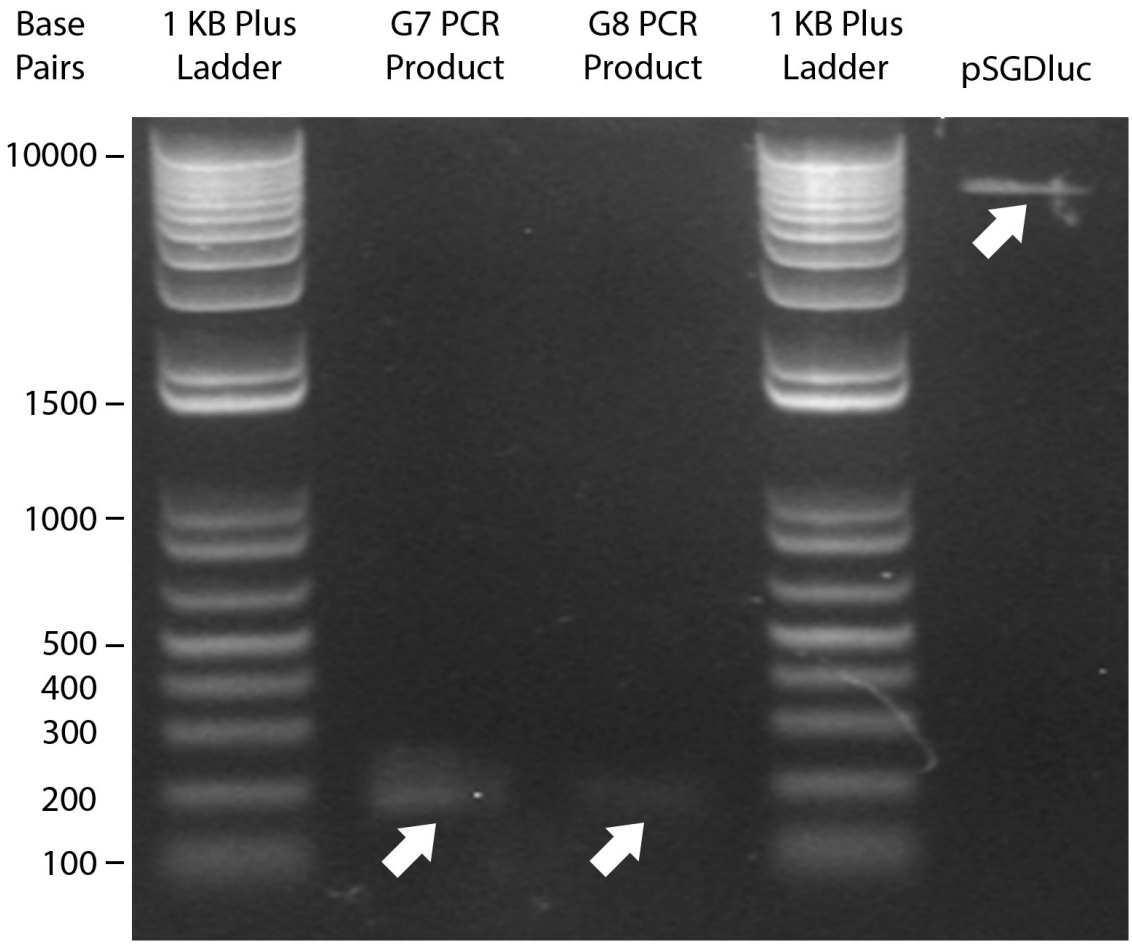

Figure 3.1. Validation of digested vector and BAXΔ2 PCR products for ligation by 1% agarose TAE gel electrophoresis

from each colony was digested by the restriction endonucleases PspXI and HindIII, and digested samples were run in their entirety on a 1% TAE gel against a 1 KB Plus Ladder (Figure 3.2). Arrows indicate the location of digested fragments, and the size of the ladder fragments are indicated on the left of the image (in units of base pairs). The digested plasmid constructs were expected to appear as bands approximately 200 and 5000 base pairs in length, corresponding to the lengths of the ligated inserts and empty vector. In all eight samples, ligation and DNA miniprep appeared to be successful as fragments appeared at these expected points relative to the ladder DNA. Two samples from each set, $G7_{1,3}$ and $G8_{2,3}$, were sent to the University of Chicago

for sequencing along with a primer (F2AU) designed specifically for this purpose. Sequencing results confirmed transformation success in all four colonies (Figure 3.3), and sample $G7_3$ and $G8_3$ were chosen to be used for DNA large prep. The arrows in Figure 3.3 indicate positions along the *BAX* gene relative to the location of the start codon (which has, for this assay, been mutated to prevent ribosome assembly). Position 182 corresponds to the end of exon 3, and therefore the end of the region cloned into the vector pSGDluc.

The vector pSGDluc is used in the dual luciferase reporter assay, as it hosts the *Renilla* and firefly luciferase genes. Production of a functional firefly luciferase protein is dependent upon the conservation of the *Renilla* luciferase reading frame through the multiple cloning site. In the case of the empty vector, the firefly luciferase gene is intentionally positioned one nucleotide out of frame relative to the *Renilla* luciferase reading frame, and thus expression of the firefly luciferase gene is not expected. In the case of Construct G7, however, firefly luciferase is in the same reading frame as *Renilla* luciferase, and expression of both genes is expected as a result. In the case of Construct G8, the firefly luciferase gene is again out of frame relative to the *Renilla* luciferase gene, and thus expression of the firefly luciferase gene is not expected. However, in the case of either a single nucleotide deletion during transcription or a shift of the ribosome by a single nucleotide in the downstream direction during translation, the reading frame could theoretically be restored, allowing for expression of the firefly luciferase gene (see Figure 3.4 for a graphical representation of these constructs). The purpose of the assay, and of the constructs that we have created, is to determine whether or not this theorized recoding event is occurring during translation of BAXΔ2, and, if it is, to obtain a measurement of the rate at which it occurs.

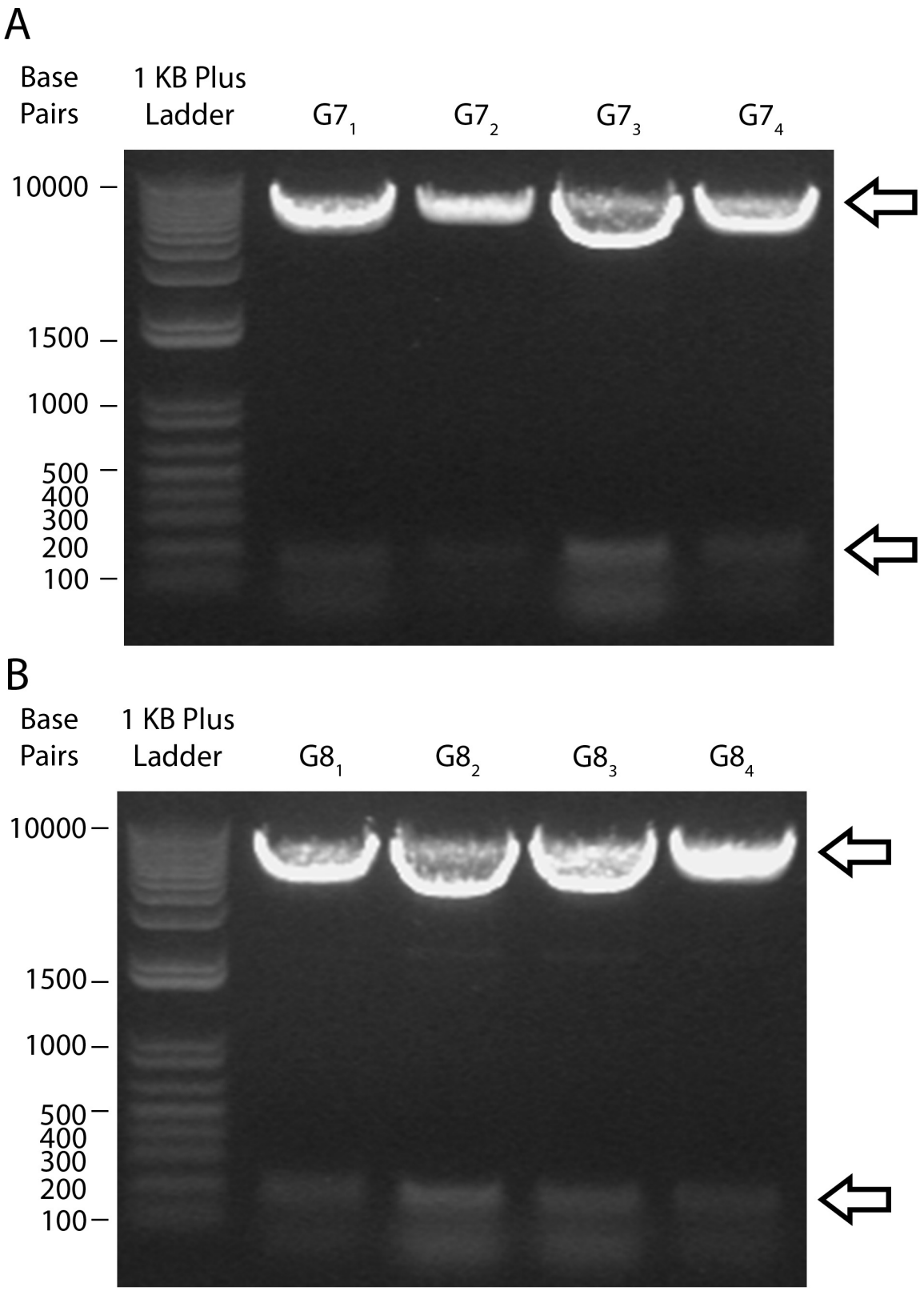

Figure 3.2. Validation of ligation and DNA miniprep via restriction digestion of plasmid DNA and 1% agarose TAE gel electrophoresis

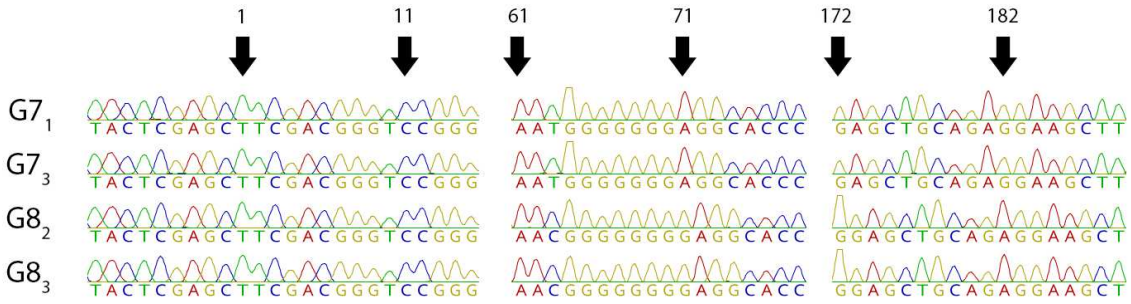

Figure 3.3. Sequencing results of selected DNA miniprep samples

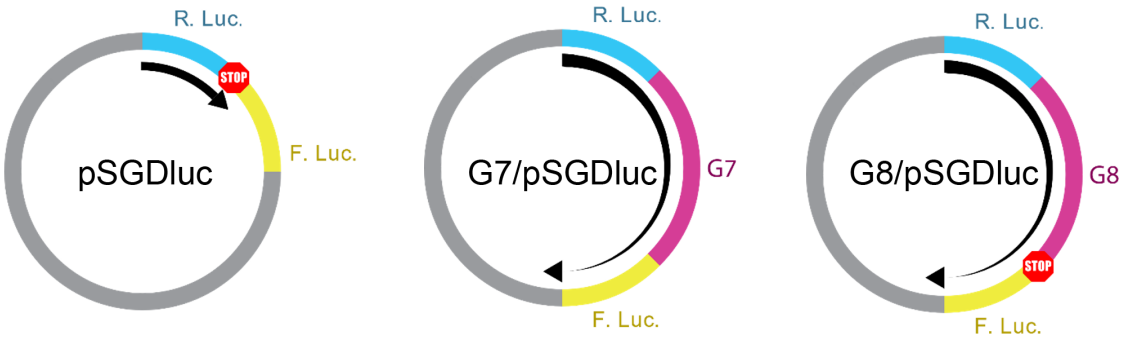

Figure 3.4. Graphical representation of relevant constructs

3.2 Arrangement of a Dual Luciferase Reporter Assay System

Luminescence was measured on a GloMax 20/20 Luminometer, produced by Promega in the late 2000's. The machine is a stand-alone unit, able to function without any additional hardware or software; however, several optional hardware add-ons were purchasable from Promega, including an internal pump (up to two per luminometer) used to transfer reagent from a storage tube to the sample tube. One was included with the unit we had access to, and an attempt was made to work it into the assay protocol to improve efficiency and reproducibility. This was not successful, at least for the dual luciferase assay, as the dual luciferase assay required two reagents to be added to each sample (and a single pump could only handle one). A second reason the pump was excluded from the final protocol is that it required a relatively large amount of stock reagent for it to work properly; this would have been wasteful for the Stop & Glo reagent, as it needed to be made fresh every day, and was only ever prepared in small aliquots (usually less than 2 mL).

The luminometer contained a number of pre-programed protocols for use with certain Promega kits. The Promega protocol DUALGLO was chosen for use with the dual luciferase assay, as it allowed for the measurement of two values of luminescence (the first after the addition of a first luciferase reagent, and the second after the addition of a second luciferase reagent), it calculated the ratio of the two measurements, and it didn't utilize pumps. Relative luminescence was measured for one second after the addition of each reagent. Protocols with longer measurement times were available, but luminescence was observed to drop off quickly (within 5-10 seconds) after addition of the Stop & Glo reagent. Using this protocol, it was determined that 1 μL of translation product (from either in vitro translation or transfection) was enough to produce measurement numbers within the range that the machine was able to measure. For the initial tests, luciferase control DNA provided as a part of an in vitro

translation kit was used. This DNA was transcribed and translated using the in vitro kit (as per the kit instructions), and 5 μL of the product was combined with 100 μL of LARII either with or without mixing via pipetting. Luminescence was measured on the luminometer both before and after combination of the reagent and the translation product, allowing for the determination of the background relative luminescence levels of both. Luminescence was also measured when no sample was present in the machine, as well as when a microcentrifuge tube containing dH_2O was present in the sample compartment. Background luminescence levels in all cases were observed to be lower than 100 relative luminescence units (RLU's). Luminescence of the control sample, after combination with LARII, was observed to be approximately 70,000,000 RLU's without mixing and approximately 240,000,000 RLU's after 3 seconds of gentle pipetting. As these numbers were quite high relative to the background, it was decided that 1 μL of translation product would be used for all future assays. This allowed for the consumption of less translation product during the assay so that the remaining product could be frozen and stored indefinitely at -80 °C.

The first round of the actual assay used 1 μL of G7, G8, or pSGDluc in vitro translation product, and was done in triplicate. Consistency between measurements was not initially at a level deemed ideal, so the protocol was modified slightly for all further assays. These modifications included the determination of a set vortex time (3 seconds) for use after the addition of Stop & Glo reagent, the preparation of a logical working environment that allowed for movement from left to right (ultimately reaching the luminometer) as the assay progressed, and the setting of more points at which reagents and translation products were to be mixed. All further assays produced more consistent measurements than the first, as can be seen upon analysis of standard deviations.

3.3 In Vitro Transcription and Translation of Luciferase Reporter Vectors and Assay of Products for Luciferase Activity

A dual luciferase reporter assay kit was used to detect the presence of "read-through" activity from both firefly luciferase and *Renilla* luciferase in samples containing the products of construct translation. The translational products of the plasmid pSGDluc were used as a negative control for the assay, as without an insert in the multiple cloning site no expression of the downstream luciferase gene was expected (Figure 3.4 A). The translational products of Construct G7 (Figure 3.4 B) were used as a positive control for the assay, as the insert was cloned into the vector pSGDluc in such a way as to allow for the expression of both luciferase genes in the absence of frameshifting. In other words, the first and second luciferase genes were presented in-frame with each other, allowing for translation of both during a single translation event under normal circumstances. The translational products of Construct G8 were used to test for frameshifting, as the insert was cloned into the vector pSGDluc in such a way as to allow for the expression of both luciferase genes only in cases in which a +1 frameshifting event had occurred (see Figure 3.5 B for a graphical representation of IVTT of this construct). In the absence of a +1 frameshifting event, expression of only the upstream luciferase gene was expected as the two genes were presented out of frame with each other. Luciferase expression levels were measured as a function of relative luminescence (measured in RLU's) in the presence of the appropriate substrate. Relative expression levels were calculated as the ratio of luminescence in the presence of the firefly luciferase substrate to luminescence in the presence of the *Renilla* luciferase substrate to determine the frequency of completed translational events. The relative expression ratio of the positive control was then normalized to a value of 100%, with the other relative expression ratios adjusted accordingly.

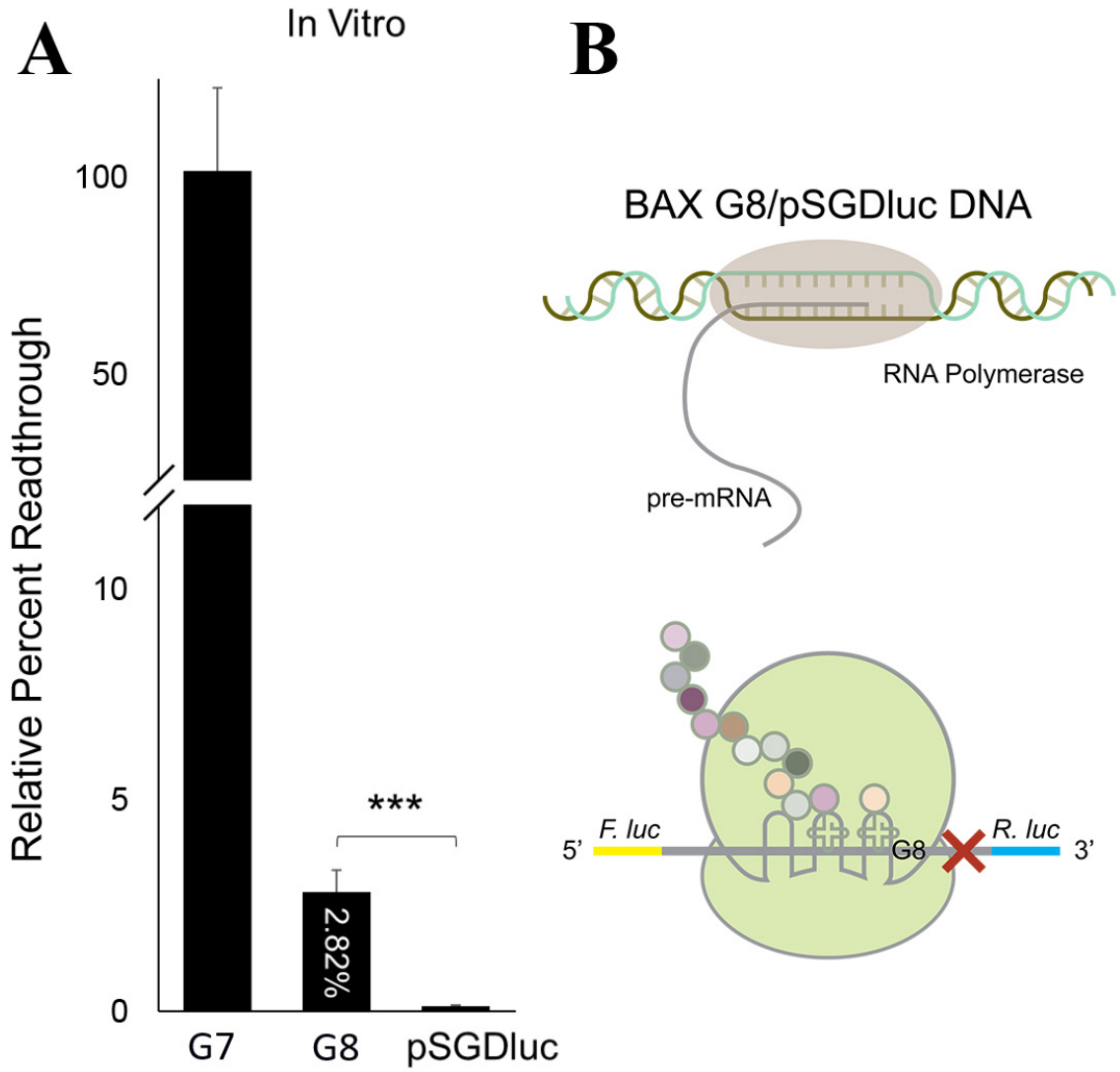

Figure 3.5. Luciferase assay of in vitro transcribed and translated Construct G8 results in a low level of expression of the out of frame firefly luciferase gene

In vitro translation allows for the rapid preparation of translation products immediately ready for assay (without the need for any further purification or manipulation). Approximately 1000 ng of DNA of each of the following was subjected to in vitro transcription and translation in triplicate: pSGDluc, Construct G7 and Construct G8. Of the in vitro translation products, 1 μL of each was subjected to the dual luciferase assay, again in triplicate. As can be seen in Figure 3.5 A, a statistically significant level of expression of the downstream luciferase gene was observed when assaying Construct G8. In the case of Construct G8, the downstream firefly luciferase gene was translated at a rate of 2.82% that of Construct G7. This suggests that epigenetic frameshifting events of the +1 variety are possible during either the transcription or translation of BAXΔ2, allowing for a low level of production of the functional firefly luciferase protein.

3.4 Transfection of Luciferase Reporter Vectors and Assay of Cell Lysates for Luciferase Activity

Transfection of plasmids into living cells through the use of a lipofectamine protocol allows for translation of our constructs in an environment more natural than that of in vitro translation. Approximately 2000 ng of DNA of each of the following was subjected to lipofection into HCT116 cells in triplicate: pSGDluc, Construct G7 and Construct G8. Following growth overnight, the cells were lysed and subjected to two freeze-thaw cycles to ensure all membranes had been disrupted. One μL of each lysate was subjected to the dual luciferase assay, again in triplicate. As can be seen in Figure 3.6 A, a statistically significant level of expression of the downstream luciferase gene was observed when assaying Construct G8 (see Figure 3.6 B for a graphical representation of transfection of this construct). In the case of Construct G8, the downstream firefly luciferase gene was translated at a rate of 4.43% that of Construct G7. This suggests that epigenetic recoding events of the +1 variety are possible during either the transcription or translation of BAXΔ2 in a cellular

environment, allowing for a low level of production of the functional firefly luciferase protein. These results mimic the results observed previously in the case of assay of in vitro translation products. A larger level of background luminescence was observed, however, likely as a result of unspecified auto-luminescent cellular debris.

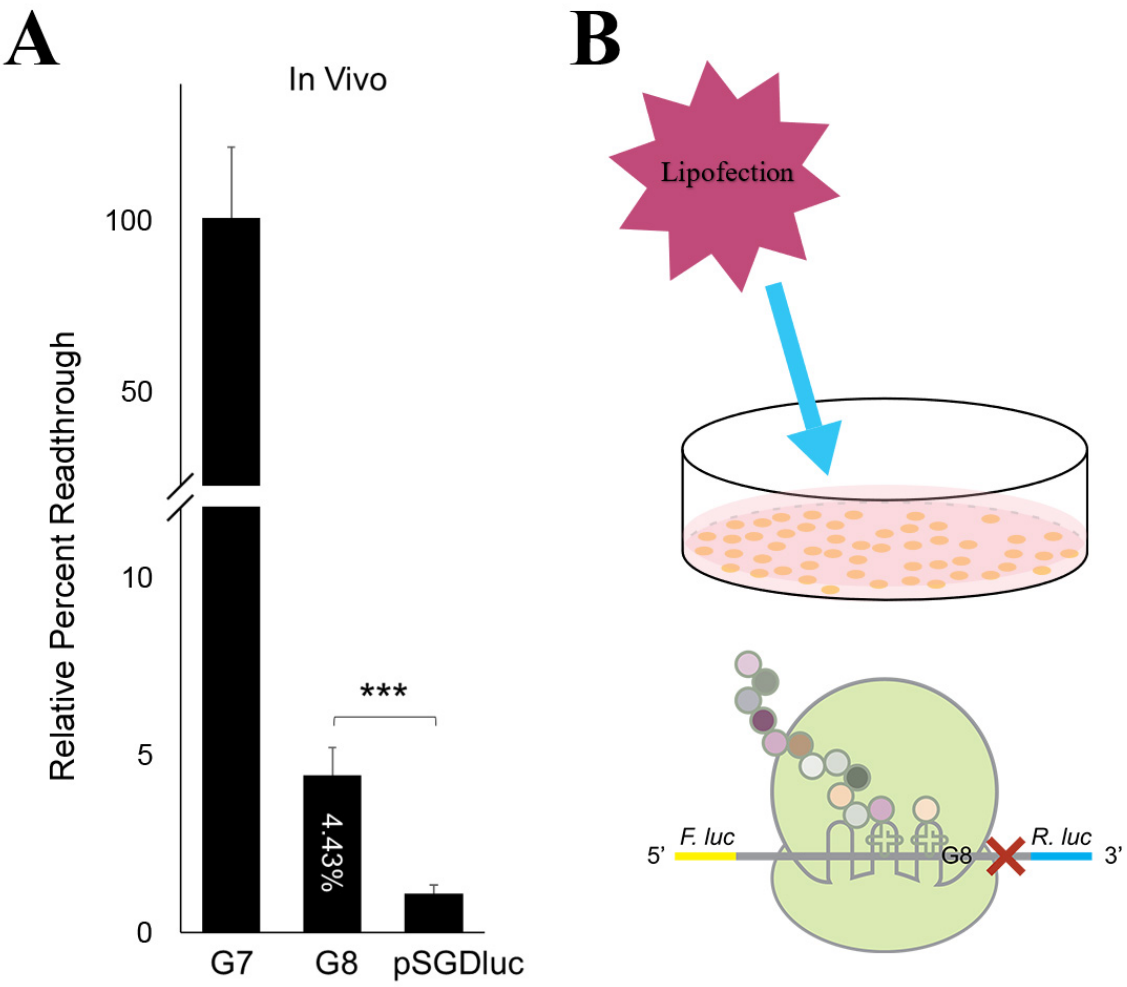

Figure 3.6. Transfection of Construct G8 into HCT-116 cells results in a low level of expression of the out of frame firefly luciferase gene

CHAPTER 4

DISCUSSION

While initial studies of the pro-apoptotic protein BAXΔ2 identified the combination of an alternative splicing event and a gene-level mutation as the prerequisites for biosynthesis in MSI+ human colon cancer cells, no similar explanation existed to explain the presence of this protein in normal and normal adjacent tissues. To identify a conceptual alternative to the gene-level mutation in the absence of an MSI+ phenotype, we utilized a dual luciferase reporter assay designed to observe epigenetic recoding. Plasmid constructs containing the first and third genes of *BAX* were either transcribed and translated in vitro or transfected into *BAX*-negative human colon cancer cells. In both cases, assay of the protein products of the reporter genes demonstrate that a low level (2.82% in vitro, 4.43% in vivo) of all translational events which produce the protein product of an upstream reporter gene also produce the protein product of a downstream reporter gene, despite the two existing in different reading frames as a result of the *BAX* exons cloned between them. These results confirm that an epigenetic recoding event is able to salvage the *BAX* reading frame in cases where exon 2 has been excised, and further narrow down the potential mechanism involved to either transcriptional slippage or programmed ribosomal frameshifting of the +1 variety.

The particular dual luciferase reporter assay that was chosen for use does not distinguish between frameshifting as a result of transcriptional slippage and frameshifting as a result of +1 programmed ribosomal frameshifting events. To determine which of the two potential recoding events are occuring, further experimentation is required. It should be noted here that the two recoding events are not mutually exclusive; +1 programmed ribosomal frameshifting and transcriptional slippage could potentially work in tandem to produce salvaged BAXΔ2 proteins at an appropriate

level in the absence of a gene-level mutation. Further, the two could be induced under differing circumstances, allowing for selective expression of the protein BAXΔ2 in a context-dependent manner.

In the case of transcriptional slippage, the most concrete way to prove its occurrence would be through a study of the transcriptome of cells known to express BAXΔ2 in the absence of a *BAX* microsatellite mutation. Should transcripts be identified that exhibit an exon 3 microsatellite composed of any number of guanines other than the canonical eight, it can be suggested that an epigenetic recoding event is able to occur at the level of transcription. If this were to occur, the next logical step would be to identify the point along the *BAX* gene that RNA polymerase is able to most efficiently undergo slippage. From what is already known about the process of transcriptional slippage and its dependence upon regions of mono- and di-nucleotide repeats, it can already be hypothesized that the process would occur during transcription of the exon 3 microsatellite. Despite the strong bonds expected to form between the DNA and nascent RNA transcript, slippage in this region would allow for re-dimerization to occur in a way that conserves the identity of some (or all) of the DNA/RNA bonds. However, should no transcripts be identified which contain any number of guanine residues other than the canonical eight in the region of the *BAX* microsatellite, and should it be shown that the cells do in fact produce the protein BAXΔ2, then it can be suggested that any transcriptional slippage that does occur does so at a rate below that which would be required for BAXΔ2 production at the levels observed. Further, this would identify programmed ribosomal frameshifting as the likely alternative prerequisite to the *BAX* microsatellite mutation observed in human colon cancer cells.

In the case of programmed ribosomal frameshifting, it's somewhat more difficult to directly prove or disprove its occurrence. Indirect evidence supporting its

occurrence could be generated by the previously described studies of the transcriptomes of BAXΔ2-positive cells, in the case where no abnormal transcripts have been identified. However, even if that were to occur, we would still be left with a number of options as to the exact type of ribosomal movement taking place. Due to the nature of ribosomal frameshifting events, it is difficult to distinguish between the different forms without the ability to quickly and efficiently sequence the protein products. As an example, both +1 and -2 programmed ribosomal frameshifting events should theoretically appear identical when observed using a dual luciferase reporter assay like the one employed here. Both events act to shift the ribosome back into the prototypical *BAX* reading frame, and would accordingly result in the production of the protein product of the downstream luciferase reporter gene. Identification of the type would require that the protein products of the translated region of *BAX* be sequenced, as a -2 ribosomal frameshifting event would result in the production of a protein containing one amino acid more than the protein produced by a +1 ribosomal frameshifting event. When we consider the fact that, in vitro, less than 3% of all *BAX* translational events undergo a recoding event, production of enough frameshifted protein for sequencing becomes the initial concern. While this is a question that does need to be answered, it will likely require some alternative approach.

If it is determined that programmed ribosomal frameshifting is responsible for production of BAXΔ2 in cells lacking a *BAX* microsatellite mutation, consideration must also be given to the motifs responsible for its regulation. The gene encoding the ornithine decarboxylase antizyme could likely be used as a starting point for these studies, as it is one of the only documented mammalian genes known to undergo a +1 ribosomal frameshifting event. It has been determined that a short sequence of seven nucleotides (similar in a sense to the slippery sequences found in cases of -1 programmed ribosomal frameshifting) must be conserved for the antizyme frameshift to occur. This sequence, UCC-UGA-U, doesn't share the predictable pattern of a

typical slippery sequence. As such, it leads to the possibility that the region of *BAX* that acts as the site of the actual frameshifting event might not be the exon 3 microsatellite. However, based on the results of the dual luciferase reporter assay, we can safely conclude that the site (should one exist) is located in either exon 1 or exon 3 of *BAX*. It's also likely that a second motif would exist to regulate the rate of frameshifting. In various documented cases of programmed ribosomal framshifting, large mRNA secondary structures have been shown to be critical for efficient frameshifting. Preliminary data suggesting the presence of a large stem loop upstream of the second half of exon 3 has already been generated through the use of the RNA folding program Knotty [19]. This program is designed to identify all RNA secondary structures, including the highly complex pseudoknots common among cases of -1 programmed ribosomal frameshifting. Further, this stem loop is only hypothesized to exist in transcripts that lack exon 2. Predictions of the secondary structure of the prototypical BAXα transcript show a complete lack of any stable secondary structures, at least in the regions prior to exon 4. Taking these predictions into consideration, it might be worth narrowing the initial search for a site of frameshifting to the second half of exon 3 (corresponding to the region not hypothesized to be part of the stem loop).

Taken together, the data generated by the dual luciferase reporter assay lays the foundation for the future of our studies into BAXΔ2. Whether it's determined that the recoding event observed is occuring at the level of transcription or translation, there is still a huge amount of work to be done before a comprehensive description of the prerequisites for BAXΔ2 biosynthesis in normal cells can be made. The combination of an alternative splicing event with a transcriptional or translational recoding event has the potential to produce, in an entirely novel way, a functional isoform of BAXα. It's yet to be determined if this is the only pathway through which BAXΔ2 can be produced in these cells, but the unique combination alone justifies further

explorations into the processes involved.

APPENDIX A

SEQUENCES

Table A.1. Primers

Primer Name	Sequence
888U	CCCTACTCGAGCTTCGACGGGTCCGGGGAGCAGC
888R	GGGAAGCTTCCTCTGCAGCTCCATGTTACTGTCCAG
F2AU	GTGAAACAGACTTTGAATTTTGACCTTCTTA

All sequences are given 5' to 3'. Underlined sequences indicate sites recognized by restriction endonucleases. The primer 888U contains a restriction site recognized by the enzyme PspXI, while the primer 888R contains a restriction site recognized by the enzyme HindIII. The primers 888U and 888R were used as upper and lower primers for PCR amplification, respectively, and correspond to regions of *BAX*. The primer F2AU was used for sequencing of our plasmid constructs, and corresponds to a region just upstream of the multiple cloning site of pSGDluc.

Table A.2. Plasmids

Construct Name	Sequence
pSGDluc	...TACTCGAGCAGAAGCTTAGATCTGAGGCACGGC...

All sequences are given 5' to 3'. Only the multiple cloning site (MCS) of the plasmid is shown. Underlined sequences indicate sites recognized by restriction endonucleases. The MCS of the plasmid pSGDluc contains restriction sites recognized by both PspXI and HindIII.

BIBLIOGRAPHY

[1] Christian N. Arnold, Ajay Goel, Hubert E. Blum, and C. Richard Boland. Molecular pathogenesis of colorectal cancer: implications for molecular diagnosis. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 104(10):2035–2047, 2005.

[2] John F. Atkins, Norma M. Wills, Gary Loughran, Chih-Yu Wu, Krishna Parsawar, Martin D. Ryan, Chung-Hsiung Wang, and Chad C. Nelson. A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA*, 13(6):803–810, 2007.

[3] Piero Benatti, Roberta Gafà, Daniela Barana, Massimiliano Marino, Alessandra Scarselli, Monica Pedroni, Iva Maestri, Laura Guerzoni, Luca Roncucci, Mirco Menigatti, et al. Microsatellite instability and colorectal cancer prognosis. *Clinical Cancer Research*, 11(23):8332–8340, 2005.

[4] Ian Brierley. Ribosomal frameshifting on viral RNAs. *Journal of General Virology*, 76(8):1885–1892, 1995.

[5] Pierre-François Cartron, Muriel Priault, Lisa Oliver, Khaled Meflah, Stephen Manon, and François M. Vallette. The N-terminal end of Bax contains a mitochondrial-targeting signal. *Journal of Biological Chemistry*, 278(13):11633–11641, 2003.

[6] Julio E. Celis and John D. Smith, editors. *Nonsense Mutations and Transfer Ribonucleic Acid Suppressors*. Academic Press Inc, 1979.

[7] Jonathan D. Dinman. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdisciplinary Reviews: RNA*, 3(5):661–673, 2012.

[8] John W. Drake. Spontaneous mutation. *Annual review of genetics*, 25(1):125–146, 1991.

[9] Philip J. Farabaugh. Programmed translational frameshifting. *Annual review of genetics*, 30(1):507–528, 1996.

[10] George E. Fox. Origin and evolution of the ribosome. *Cold Spring Harbor perspectives in biology*, 2(9):1–18, 2010.

[11] Nai Yang Fu, Sunil K. Sukumaran, Sze Yen Kerk, and C. Yu Victor. Baxβ: a constitutively active human Bax isoform that is under tight regulatory control by the proteasomal degradation mechanism. *Molecular cell*, 33(1):15–29, 2009.

[12] Jean-Francois Gout, Weiyi Li, Clark Fritsch, Annie Li, Suraiya Haroon, Larry Singh, Ding Hua, Hossein Fazelinia, Zach Smith, Steven Seeholzer, et al. The landscape of transcription errors in eukaryotic cells. *Science advances*, 3(10):1–11, 2017.

[13] Guido Grentzmann, Jennifer A. Ingram, Paul J. Kelly, Raymond F. Gestaland, and John F. Atkins. A dual-luciferase reporter system for studying recoding signals. *RNA*, 4(4):479–486, 1998.

[14] Bonnie Haferkamp. *Anti-tumor Splicing: Restoring the Tumor Suppressor Bax in Microsatellite Unstable Tumors*. PhD thesis, Illinois Institute of Technology, 2011.

[15] Bonnie Haferkamp, Honghong Zhang, Samuel Kissinger, Xin Wang, Yuting Lin, Megan Schultz, and Jialing Xiang. BaxΔ2 family alternative splicing salvages Bax microsatellite-frameshift mutations. *Genes & cancer*, 4(11-12):501–512, 2013.

[16] Bonnie Haferkamp, Honghong Zhang, Yuting Lin, Xinyi Yeap, Alex Bunce, Juanita Sharpe, and Jialing Xiang. BaxΔ2 is a novel Bax isoform unique to microsatellite unstable tumors. *Journal of Biological Chemistry*, 287(41):34722–34729, 2012.

[17] Heather Hampel, Wendy L. Frankel, Edward Martin, Mark Arnold, Karamjit Khanduja, Philip Kuebler, Mark Clendenning, Kaisa Sotamaa, Thomas Prior, Judith A. Westman, et al. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. *Journal of Clinical Oncology*, 26(35):5783–5788, 2008.

[18] Nele Hug, Dasa Longman, and Javier F. Cáceres. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research*, 44(4):1483–1495, 2016.

[19] Hosna Jabbari, Ian Wark, Carlo Montemagno, and Sebastian Will. Knotty: efficient and accurate prediction of complex RNA pseudoknot structures. *Bioinformatics*, 34(22):3849–3856, 2018.

[20] T. Jacks and H. Varmus. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science*, 230(4731):1237–1242, 1985.

[21] J. M. Jurgensmeier, Z. Xie, Q. Deveraux, L. Ellerby, D. Bredesen, and J. C. Reed. Bax directly induces release of cytochrome c from isolated mitochondria. *Proceedings of the National Academy of Sciences*, 95(9):4997–5002, 1998.

[22] Ameeta Kelekar and Craig B. Thompson. Bcl-2-family proteins: the role of the BH3 domain in apoptosis. *Trends in Cell Biology*, 8(8):324–330, 1998.

[23] Robin Ketteler. On programmed ribosomal frameshifting: the alternative proteomes. *Frontiers in Genetics*, 3, 2012.

[24] Emily B. Kramer, Haritha Vallabhaneni, Lauren M. Mayer, and Philip J. Farabaugh. A comprehensive analysis of translational missense errors in the yeast Saccharomyces cerevisiae. *RNA*, 16(9):1797–1808, 2010.

[25] Thomas A. Kunkel. DNA replication fidelity. *Journal of Biological Chemistry*, 279(17):16895–16898, 2004.

[26] C. G. Kurland. Translational accuracy and the fitness of bacteria. *Annual Review of Genetics*, 26(1):29–50, 1992.

[27] Tatsuaki Kurosaki and Lynne E. Maquat. Nonsense-mediated mRNA decay in humans at a glance. *Journal of Cell Science*, 129(3):461–467, 2016.

[28] Honglin Li, Hong Zhu, Chi jie Xu, and Junying Yuan. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell*, 94(4):491–501, 1998.

[29] Peng Li, Deepak Nijhawan, Imawati Budihardjo, Srinivasa M. Srinivasula, Manzoor Ahmad, Emad S. Alnemri, and Xiaodong Wang. Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell*, 91(4):479–489, 1997.

[30] Gary Loughran, Michael T. Howard, Andrew E. Firth, and John F. Atkins. Avoidance of reporter assay distortions from fused dual reporters. *RNA*, 23(8):1285–1289, 2017.

[31] Sara Maia, W. Nicholas Haining, Sascha Ansén, Zhinan Xia, Scott A. Armstrong, Nilufer P. Seth, Paolo Ghia, Monique L. den Boer, Rob Pieters, Stephen E. Sallan, Lee M. Nadler, and Angelo A. Cardoso. Gene expression profiling identifies BAX-δ as a novel tumor antigen in acute lymphoblastic leukemia. *Cancer Research*, 65(21):10050–10058, 2005.

[32] Adriana Mañas, Aislinn Davis, Evan Beatty, Honghong Zhang, Qi Yao, Jiajun Li, Adam Nelson, Sana Basheer, Huaiyuan Zhang, and Jialing Xiang. Unique distribution of BaxΔ2 in human tissues: from normal to highly-malignant independent of prerequisite microsatellite mutation. Manuscript submitted for publication.

[33] Adriana Mañas, Aislinn Davis, Sydney Lamerand, and Jialing Xiang. Detection of pro-apoptotic BaxΔ2 proteins in the human cerebellum. *Histochemistry and Cell Biology*, 150(1):77–82, 2018.

[34] Adriana Mañas, Sheng Wang, Adam Nelson, Jiajun Li, Yu Zhao, Huaiyuan Zhang, Aislinn Davis, Bingqing Xie, Natalia Maltsev, and Jialing Xiang. The functional domains for BaxΔ2 aggregate-mediated caspase 8-dependent cell death. *Experimental Cell Research*, 359(2):342–355, 2017.

[35] Olivier Namy, Stephen J. Moran, David I. Stuart, Robert J. C. Gilbert, and Ian Brierley. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, 441(7090):244–247, 2006.

[36] Adriana Mañas Núñez. *BAXΔ2: From Functional Domains To Cancer Correlation*. PhD thesis, Illinois Institute of Technology, 2018.

[37] Zoltán N. Oltval, Curt L. Milliman, and Stanley J. Korsmeyer. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programed cell death. *Cell*, 74(4):609–619, 1993.

[38] Nicholas Rampino, Hiroyuki Yamamoto, Yurij Ionov, Yan Li, Hisako Sawai, John C. Reed, and Manuel Perucho. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science*, 275(5302):967–969, 1997.

[39] E. Rom and C. Kahana. Polyamines regulate the expression of ornithine decarboxylase antizyme in vitro by inducing ribosomal frame-shifting. *Proceedings of the National Academy of Sciences*, 91(9):3959–3963, 1994.

[40] Estelle Schmitt, Claudie Paquet, Myriam Beauchemin, Jessica Dever-Bertrand, and Richard Bertrand. Characterization of Bax-σ, a cell death-inducing isoform of Bax. *Biochemical and Biophysical Research Communications*, 270(3):868–879, 2000.

[41] I. P. M. Tomlinson, M. R. Novelli, and W. F. Bodmer. The mutation rate and cancer. *Proceedings of the National Academy of Sciences*, 93(25):14800–14803, 1996.

[42] Lori A. Wagner, Robert B. Weiss, Robert Driscoll, Diane S. Dunn, and Ray F. Gesteland. Transcriptional slippage occurs during elongation at runs of adenine or thymine in Escherichia coli. *Nucleic Acids Research*, 18(12):3529–3535, 1990.

[43] D. Westphal, R. M. Kluck, and G. Dewson. Building blocks of the apoptotic pore: how Bax and Bak are activated and oligomerize during apoptosis. *Cell Death & Differentiation*, 21(2):196–205, 2013.

[44] Gwyn T. Williams and Christopher A. Smith. Molecular regulation of apoptosis: Genetic controls on cell death. *Cell*, 74(5):777–779, 1993.

[45] H. Zhang, Y. Lin, A. Manas, Y. Zhao, M. F. Denning, L. Ma, and J. Xiang. BaxΔ2 promotes apoptosis through caspase-8 activation in microsatellite-unstable colon cancer. *Molecular Cancer Research*, 12(9):1225–1232, 2014.

[46] Mei Zhou, Susan D. Demo, Thida N. McClure, Roberto Crea, and Catherine M. Bitler. A novel splice variant of the cell death-promoting protein BAX. *Journal of Biological Chemistry*, 273(19):11930–11936, 1998.