THE ROLE OF EXON B OF THE 5’ UNTRANSLATED REGION OF
SURFACTANT PROTEIN A IN THE REGULATION OF EXPRESSION

A Thesis in
Genetics
by
Brett Patrick Simmons

© 2008 Brett Patrick Simmons

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Master of Science

August 2008
The thesis of Brett P. Simmons was reviewed and approved* by the following:

**Joanna Floros**  
Evan Pugh Professor of Pediatrics and Obstetrics and Gynecology  
Thesis Adviser

**Ralph L. Keil**  
Associate Professor of Biochemistry and Molecular Biology

**Kristin A. Eckert**  
Professor of Pathology

**Sarah K. Bronson**  
Associate Professor of Cellular and Molecular Physiology  
Co-Chair of the Intercollege Graduate Degree Program in Genetics

*Signatures are on file in the Graduate School.
Abstract

Gene regulation is a complex and multifaceted process, and both coding and non-coding sequences are capable of regulatory function. The 5’ untranslated region (UTR) of a gene can have a significant effect on regulation at the levels of transcription and translation. Post-transcription, the 5’ UTR is capable of conferring mRNA stability, providing binding sites for translation machinery, containing degradation and/or localization signals, and other functions dependent on tertiary RNA structure and RNA-protein structure. One gene in which the 5’ UTR is essential for proper control of protein expression is Surfactant Protein A (SP-A).

SP-A is one of four pulmonary surfactant-associated proteins, all of which are encoded by separate genes. A thin liquid layer covers the distal airspaces or alveolar spaces. Pulmonary surfactant is found between the air-liquid interface of the distal airspaces and its function is to reduce surface tension at low lung volumes and prevent alveolar collapse. Pulmonary surfactant is comprised of proteins and lipids that together play a vital role in lung function and host defense. Furthermore, aberrant expression of SP-A is a characteristic of many pulmonary diseases and disorders, including asthma, respiratory distress syndrome, and cancer, among others.

Previous studies have indicated that exon B of the SP-A 5’ UTR may be a potent cis-acting element of post-transcriptional and translational regulation. Using the luciferase assay and quantitative real-time PCR, the translational enhancer capacity of this 5’ UTR exon was assessed to investigate the hypothesis that the 30-basepair exon B sequence is a cis-acting translational enhancer. Transcription was normalized by using an SV40 promoter and through the calculation of a translational index (activity divided by relative mRNA content). To investigate the function of exon B in the context of the whole SP-A 5’ UTR, exon B was deleted from the 5’ UTR. In another construct, the exon B sequence was replaced with a 30-basepair random sequence. Compared with the whole SP-A 5’ UTR construct, these two constructs, which lack the exon B sequence, showed reduced luciferase activity. This indicated that exon B is capable of increasing translation activity and could have resulted from an increase in stability, altered localization patterns, or an increase in translation efficiency.

To determine whether exon B acts as a translational enhancer in the context of other 5’ UTR sequences, constructs that each contained one of two independent guest 5’ UTRs cloned downstream of the B exon were used in a guest 5’ UTR study. The SP-B and SP-D 5’ UTRs
functioned as guest 5’ UTRs. While the size of the leader sequence appeared to have little effect on translation, constructs containing the complete exon B sequence were found to have activities twice as high as the activities of 5’ UTR constructs containing only the guest 5’ UTRs. Constructs containing exon B also exhibited twice as much activity as those containing random 30-nucleotide sequences in place of exon B. This indicated that exon B is a translational enhancer and functions in a sequence-specific manner. Furthermore, the combination of the expression and real time PCR data suggest that SP-A 5’ UTR exon B enhances expression by improving translational efficiency. A finding unrelated to SP-A 5’ UTR exon B was that the SP-D 5’ UTR is stronger than the SP-B 5’ UTR, meaning that it enhanced expression by a larger amount when the luciferase assay was used to measure activity.

Using a series of deletion mutations, as well as other site-directed mutagenesis changes, the region of this noncoding exon that is responsible for the expression enhancement was investigated. For this, 5 basepairs were deleted at a time within the B exon of a complete SP-A 5’ UTR-luciferase construct, and these deletion mutation constructs were transfected into NCI-H441 cells. Luciferase assay results indicated that a 15-nucleotide region appeared to be important for exon B’s translational enhancement.

Translation is a complex and dynamic biological process and is regulated at several levels and by various means. While elucidating the precise mechanisms of SP-A translational control is beyond the scope of the present study, the resources and technology required to advance our understanding of these processes are available. Future studies may reveal the trans-acting factors necessary for SP-A translation and how the process of regulating SP-A expression works from nuclear mRNA export to translation termination.
## Table of Contents

List of Tables ......................................................... Page vii
List of Figures ......................................................... Page viii
Acknowledgements ..................................................... Page ix

I. Introduction and Background ...................................... Page 1
   1. Lungs, Pulmonary Surfactant .................................. Page 1
   2. Surfactant-Associated Genes .................................. Page 1
   3. The 5’ UTR and Translation .................................. Page 5
   4. Summary ....................................................... Page 8

II. Objectives .......................................................... Page 9

III. Methods ............................................................ Page 11
   1. Preparation of Plasmid Constructs ......................... Page 11
   2. Site-Directed Mutagenesis .................................. Page 16
   3. Isolation of DNA Plasmids from Bacteria Cultures ...... Page 18
   4. Large Scale DNA Preparations ......................... Page 18
   5. Transformations ............................................. Page 19
   6. Quantification of Nucleic Acid Yield and Purity ...... Page 19
   7. Cell Culture with NCI-H441 Cells ..................... Page 20
   8. Preparation of Selective Antibiotic Plates .......... Page 21
   9. Preparation of LB Media .................................. Page 21
  10. Transient NCI-H441 Cell Transfections ............... Page 21
  11. Luciferase Assay ............................................ Page 22
  12. Quantification of mRNA by Real Time RT-PCR .... Page 23
  13. Agarose Gels ................................................. Page 24
  14. PAGE Gels .................................................... Page 25
  15. Statistics ..................................................... Page 26
IV. Results

1. Whole SP-A 5’ UTR Study

2. Guest 5’ UTR Studies

3. Deletion Mutation Study

V. Discussion

VI. Conclusions and Future Directions

Appendix: RNA Structures

References
# List of Tables

1. Table of Oligonucleotides Used to Produce Constructs  
   Page 13

2. Table of Primers Used in Site-Directed Mutagenesis Reactions  
   Page 16

3. Table of Primers Used in Real-Time RT-PCR Reactions  
   Page 24
### List of Figures

1. Schematic of SP-A Genomic, Genetic, and 5’ UTR Structure  
   Page 3
2. Original and Modified Plasmid Structures  
   Page 11
3. Constructs Used in Whole 5’ UTR Experiment  
   Page 29
4. Luciferase Assay Results (Whole 5’ UTR)  
   Page 30
5. Constructs Used in Enhancer Experiment (SP-B 5’ UTR)  
   Page 32
6. Luciferase Assay Results (SP-B 5’ UTR)  
   Page 33
7. Constructs Used in Enhancer Experiment (SP-D 5’ UTR)  
   Page 35
8. Luciferase Assay Results (SP-D 5’ UTR)  
   Page 36
9. Combined Enhancer Experiment Luciferase Assay Results (SP-B and SP-D)  
   Page 39
10. Real-Time PCR Results  
    Page 41
11. Translational Index  
    Page 43
12. Deletion Mutation Constructs  
    Page 46
13. Deletion Mutation Luciferase Assay Results  
    Page 47
14. Relative Activity vs. Stability Graph 1  
    Page 53
15. Relative Activity vs. Stability Graph 2  
    Page 54
16. Relative Activity vs. 5’ UTR Length Graph  
    Page 56
Acknowledgements

I express my sincere gratitude to Dr. Joanna Floros, Ph.D. and Dr. Guirong Wang, Ph.D. as well as the personnel of the laboratory, including Susan DiAngelo, Faizah Bhatti, M.D., Anatoly Mikorev, Ph.D., Mudit Vaid, Ph.D., Mehboob Ali, Ph.D., Rizwanul Haque, Ph.D., and Anamika Singh, Ph.D., for their insightful and helpful advice and support. I also thank my doctoral graduate committee, comprised of Dr. Gavin Robertson, Ph.D., Dr. Kristin Eckert, Ph.D., Dr. Victor Canfield, Ph.D., and Dr. Christopher Herzog, Ph.D. The guidance and critique provided by my master’s committee, which included Dr. Joanna Floros, Ph.D., Dr. Ralph Keil, Ph.D., and Dr. Kristin Eckert, Ph.D., were crucial to the completion of this thesis and I appreciate the time and effort spent on guiding my project to completion. I thank Dr. Anita Hopper, Ph.D. for her administrative support as co-chair of the Genetics program and Robert Brucklacher for assistance with the RT-PCR work. Most importantly, I wish to thank my parents for their frequent assistance and counsel during my tenure as a graduate student at the Milton S. Hershey School of Medicine.
I. Introduction and Background

1. Lungs, Pulmonary Surfactant

The lungs are the largest and most studied organs of the respiratory system. In mammals and other higher organisms, the lungs serve three essential functions: to transfer oxygen from the external environment to the blood, to remove carbon dioxide and other gaseous wastes, and to defend the organism from airborne pathogens. The anatomical structure and biochemical metabolism and composition of the lungs have been optimized for these purposes throughout evolution.

Air enters the respiratory tract through the nasal or oral passages and proceeds down the trachea. From the trachea, oxygen-rich air enters the right or left lung by way of the primary bronchi, from which the secondary and tertiary bronchi branch in an inverted tree-like structure. Bronchioles stem from tertiary bronchi and lead to tiny sac-like structures known as alveoli. The alveoli are the sites of gas exchange, and contain a substance called pulmonary surfactant.

While surfactants are a part of everyday life for the developed world as a component of soaps, detergents, and other household chemicals, pulmonary surfactant is no less important to the biology of higher organisms. This substance reduces surface tension in the alveoli and prevents the lungs from collapsing. Through its ability to drastically reduce the surface tension at the liquid-gas interface, surfactant allows the lungs to inflate more easily and greatly reduces the work associated with breathing. Surfactant also keeps fluid accumulation in the lungs at a minimum.

Pulmonary surfactant is composed of approximately 90% lipid and 10% protein (reviewed in Hawgood, 2001). While a diverse range of lipids are present in surfactant, 60% of the lipid fraction is comprised of 2 lipids known as dipalmitoylphosphatidylcholine (DPPC) and phosphatidylglycerol (PG). Approximately half of the protein is comprised of plasma proteins, while the other half is the 4 surfactant-associated proteins SP-A, SP-B, SP-C, and SP-D (reviewed in Hawgood, 2001).

2. Surfactant-Associated Genes

While the 4 surfactant-associated proteins have unique structures, SP-D and SP-A are involved in innate immunity, while SP-B and SP-C function in enhancing surfactant properties (Clark, 1995; Clark, 1997; and Glasser, 2001). SP-A and SP-D genes are located on the q arm of
chromosome 10 and the SP-C gene is found on chromosome 8 and SP-B on chromosome 2 (Hoover, 1998; Bruns, 1987; Fisher, 1988 and Pilot-Matias, 1989). The 5’ UTRs of SP-A (137 nucleotides in the longest transcript) and SP-C (158 nucleotides) are longer than those of the other surfactant-associated genes. The SP-B 5’ UTR is only 14 bases in length, while the 5’ UTR of SP-D is 43 bases long (Pilot-Matias, 1989 and Crouch, 1993).

Surfactant protein A is the most abundant of the surfactant-associated proteins and functions in host defense, modulating inflammatory responses, and regulating surfactant metabolic processes (reviewed in Crouch, 2001; McCormack, 2002; reviewed in Floros, 2002 and reviewed in Phelps, 2001). SP-A is expressed in alveolar epithelial type II cells, tracheal and bronchial submucosal gland cells, Clara cells, and other tissues outside of the lungs (Madson, 2003; Floros, 1988 and reviewed in Phelps, 2001).

At the protein level, the structure of SP-A consists of 4 domains, including an N-terminal piece, a collagen-like domain, a neck domain that is relatively hydrophobic, and a carbohydrate recognition domain (CRD). SP-A is considered to be part of the collectin family of proteins, and many proteins in this class enhance phagocytosis processes in an antibody-independent manner (Hoppe, 1994). While an SP-A monomer consists of 248 amino acids (including the signal peptide), human SP-A undergoes post-translational modifications and exists as an approximately 30-35 kDa sialoglycoprotein in its modified form. Mature SP-A resembles a bouquet of flowers, as 3 SP-A molecules form a trimer and 6 trimers form an octadecameric flower-like structure (Voss, 1991). Oligomerization takes place at the amino termini and collagen-like domains (Voss, 1991 and reviewed in McCormack, 1998). It was thought that only SP-A1 and SP-A2 heterodimers existed naturally, but recent in vitro studies have suggested that homodimers exist and that these are functional (Mikerov, 2007; Wang, 2000, 2004, 2008 and Oberley, 2003). The carbohydrate-binding domain of SP-A interacts with lung pathogens by binding to carbohydrates on the cell membrane.

The human SP-A locus is found on the long arm of chromosome 10 and consists of 2 functional genes, SP-A1 and SP-A2, and a nonfunctional pseudogene (Bruns, 1987 and Hoover, 1998). The functional genes, while only 40 kilobases apart, are transcribed in opposite orientation, toward the chromosome’s telomeres (Hoover, 1998). The genomic sequence of the SP-A1 and SP-A2 genes are available, along with the cDNA sequence for each gene (White, 1985; Floros, 1986 and Katyal, 1992). While SP-A1 and SP-A2 are structured similarly, there are several single-nucleotide polymorphisms (SNPs) that occur in these 2 genes, including some that
result in amino acid changes. The core amino acid changes that distinguish SP-A1 from SP-A2 are located within the collagen-like domain of SP-A (Karinch, 1995). As a result of several SNP combinations in SP-A1 and SP-A2, there are many commonly found alleles of these 2 genes (DiAngelo, 1999). The result of the differences among the variants on immunity and surfactant structure and metabolism is not well understood. The genomic, genetic, and 5’ UTR structure of an SP-A gene (representative of SP-A1 or SP-A2) is shown in Figure 1. In Figure 1, “C” represents the centromeric side, while “T” represents the telomeric side. The 5’ UTR exons are labeled alphabetically, while the coding exons are labeled with Roman numerals.

**Figure 1**

**SP-A Genomic, Genetic, and 5’ UTR Structure**

Figure 1 shows the genomic (top), genetic (middle), and 5’ UTR (bottom) structure of an SP-A gene. Exon B, the focus of the present study, is represented as the red 5’ UTR exon. Other 5’ UTR exons are represented as light-red blocks and are labeled alphabetically. Known 5’ UTR exons and their sizes are shown below the 5’ UTR structure. Coding exons are colored blue in the genetic and 5’ UTR structures and are labeled with Roman numerals. The figure is not drawn to scale.

The 5’ untranslated region (5’ UTR) of the SP-A genes consists of 4 exons, or regions (A, B, C, and D). The fourth 5’ UTR exon is continuous with the first coding exon. The 3’ untranslated region (3’ UTR) consists of only 1 exon, which is continuous with the fourth coding
exon. The 5’ UTR of these genes is complex, as the exons in this region undergo alternative splicing to form a large number of 5’ UTR configurations (Karinch, 1995). While variations of all 5’ UTR exons are found in SP-A1 transcripts, only the A, B, and D exons are found in SP-A2 transcripts (Karinch, 1995).

The lengths of the 5’ UTR exons also differ among the splice-variants. There are 3 SP-A1 transcription start sites, resulting in 3 different possible sizes for the A 5’ UTR exon: 44, 39, and 34 nucleotides. These exons are labeled A, A’, and A”, respectively. Among SP-A1 and SP-A2 transcripts, exon B is 30 nucleotides long while exon B’ is 70 nucleotides long. Unlike exon A, exon variants B and B’ differ in length at the 3’ end. Although exon C has only been found in SP-A1 transcripts, its isoforms differ at the 5’ end of the exon by 3 nucleotides. Exon C can be either 60 (C’) or 63 (C) nucleotides long. Like exon C, exon D varies in length at the 5’ end by 3 nucleotides. It is found in transcripts as 26 (D) and 23 (D’) nucleotide segments (Karinch, 1995).

The formation of SP-A 5’ UTR splice-variants is not a random process, as there are major, minor, and rare splice-variants for SP-A1 and SP-A2 transcripts. A study of SP-A1 and SP-A2 transcripts found that the predominant splice-variant configurations and their frequencies differ between the 2 genes, and they likely also differ among individuals. The major splice-variant for SP-A1 is AD’ with a frequency of 81%. ACD’ and AB’D’ splicing patterns were also observed, and each has a relative frequency of approximately 7%. Other less commonly found splice-variants comprised the remaining 5%. For the SP-A2 gene, the major observed splice-variants are ABD (44%) and ABD’(49%). Others made up the remaining 7% (Karinch, 1995).

SP-A expression is controlled at the levels of transcription and translation. SP-A is regulated by tissue specificity and is controlled developmentally and by hormones (Boggaram, 1988 and reviewed in Mendelson, 2000). Transcription is turned on during the second trimester and increases steadily until birth (Stray-Pedersen, 2007). Several transcription factors are key determinants of surfactant protein expression. Expression of SP-A1 and SP-A2 is differentially regulated by glucocorticoids and other modulators, but the physiological roles of these factors is unclear (Boggaram, 1991, Iannuzzi, 1993, Hoover, 1999, Karinch, 1998, Kumar, 1998 and Wang, 2003).

Little is known about the functional role of the 5’ UTR splice-variants. The 5’ UTR splice-variants of SP-A have been shown to be translated \textit{in vitro} and \textit{in vivo} (Karinch, 1995 and
Karinch, 1997). Studies on the impact of the 5’ UTR variants on translation efficiency and RNA content and stability indicate that the SP-A2 5’ UTR and transcripts may be more active and abundant than those from the SP-A1 gene (Wang, 2005).

In humans, several diseases and complications are correlated with altered SP-A levels and functionality. Patients with cystic fibrosis and pneumonia commonly have reduced levels of SP-A (McCormack, 1995 and Noah, 2003). Most smokers have reduced SP-A concentrations, which could at least partially explain their elevated risk for respiratory infections (Kida, 1997). Functionally impaired SP-A protein or reduced SP-A expression is a leading cause of asthma (Cheng, 2000). Respiratory distress syndrome (RDS) was at one time the leading cause of death for premature infants, and exhibits a very strong correlation with low SP-A levels (deMello, 1989; Pryhuber, 1991 and Ramet, 2000). RDS in infants is still a major cause of death, especially in developing regions of the world. More recently, specific variants of SP-A have been associated with lung cancer (Seifart, 2005). While a reduction in function is associated with many diseases, high surfactant levels can also lead to complications. For example, pulmonary alveolar proteinosis (PAP) is characterized by large accumulations of surfactant in the alveoli (Brasch, 2004).

3. The 5’ UTR and Translation

The 5’ UTR, or leader sequence, of a gene is the DNA or RNA sequence upstream of the coding sequence that is not translated but is present in the mRNA. The 5’ UTR and 3’ UTR are important determinants of RNA trafficking, RNA stability, and translation initiation efficiency (Pesole, 1997). Databases of 5’ UTR and 3’ UTR sequences have been available for some time and offer insights into their functional role in modulating expression patterns in various species (Pesole, 2000). Untranslated regions, both 5’ and 3’, can vary greatly in length between genes. In general, 5’ UTRs are much shorter than 3’ UTRs. In humans, 5’ UTR lengths range from just a few nucleotides to a maximum known length of 2803 nucleotides (Pesole, 2000). The average human 5’ UTR consists of approximately 210 nucleotides. Other vertebrate species, by comparison, typically have shorter 5’ UTRs, with an average of 100 to 200 nucleotides (Pesole, 2000).

While the vast majority of 5’ UTRs contain only one exon, the genomic sequence from which 5’ UTRs are derived can include one or more introns. It is relatively uncommon for genes to contain larger numbers of exons in their UTRs (5’ and 3’ UTRs) (Pesole, 2000). Like other
noncoding sequences, 5’ and 3’ UTRs can contain several types of repeats in both intronic and exonic sequences. Simple repeats, LINEs, SINEs, and long terminal repeats have been found in 5’ UTR sequences. In some cases, these repeats may have a role in regulating expression through RNA stability or translational mechanisms (reviewed in Smit, 1999). 5’ UTRs generally consist of highly conserved sequences, which are often comparable in conservation parameters to coding sequences (Makalowski, 1998). This fact not only underscores the importance of gene regulation, but indicates that 5’ UTR sequences may be nearly as important as protein coding sequences in terms of gene function. While some sequence- and structure-based motifs are overrepresented in 5’ and 3’ UTR sequences, more advanced motif algorithms are needed to discover and classify functional 5’ UTR sequences.

The process of converting an mRNA sequence to an amino acid sequence is known as translation (Matthaei, 1961). Translation is a highly regulated, dynamic process that takes place in 3 stages: initiation, elongation, and termination. The predominant point of regulation occurs at initiation. While the process of transcription initiation is largely dependent on the primary sequence of the DNA, the step of translation initiation relies on both the primary and secondary structures of the RNA (Pesole, 2001). Essentially, the ribosome and a number of associated proteins bind at the 5’ UTR of an mRNA. In most cases, circularization occurs through interactions with the 5’ UTR, translation initiation factors, and the polyadenine tail added to the 3’ UTR. Under the currently accepted scanning model of translation initiation, the 40S ribosomal subunit and various initiation factors bind to the 5’ UTR and this complex then scans in a 5’ → 3’ direction until the first AUG codon (Kozak, 1978). With the help of other initiation factors, the 60S ribosomal subunit attaches in a GTP-dependent manner and the newly formed 80S ribosome along with charged tRNA residues catalyze the formation of peptide linkages between amino acids. Translation initiation requires dozens of general (and in some cases specific) proteins for translation to initiate and proceed in an efficient manner (reviewed in Dever, 2002).

While regulation of cellular metabolic processes can take place at the levels of transcription, post-transcription, translation, and post-translation, translational regulation is advantageous for several reasons. First, regulation at the point of translation is more energy efficient than post-translational regulation. Polypeptide synthesis is one of the top 3 energy consuming processes in cells (along with DNA/RNA synthesis and Na+/Ca2+ cycling). Second, translational regulation is direct and rapid, as control at this point is independent of upstream processes and of the time involved in transcription and nuclear export. Third, the majority of
translational controls are accomplished through reversible mechanisms, which makes this a highly reversible process. For example, phosphorylation or dephosphorylation of translation factors is usually very rapid and reversible. Fourth, translational regulation offers finer control than upstream regulatory mechanisms. While most genes are regulated transcriptionally and translationally, changes in transcription are typically greater in magnitude. Fifth, translational regulation offers spatial control of gene expression. Translation factors or mRNA can be sequestered and translated where it is needed within the cell. Finally, regulation of translation allows for both global and specific protein regulation. By activating different modes of control, the cell can regulate all protein synthesis, translation of a subset of mRNAs, or the production of individual proteins.

Within the cytosol, messenger RNA (mRNA) is not observed as a simple linear strand of ribonucleotides. The primary structure of mRNAs determines the unique secondary structures that form through sequence-dependent means. Computer-based algorithms for predicting the secondary structure of mRNAs are available. One program, widely used by biologists to predict the most likely secondary structures given a specific RNA sequence, is mFold (Zuker, 2003). The secondary RNA structures formed in and near the 5’ UTRs of genes are important for translation initiation and may have enhancing or inhibitory effects on translation rates (Kozak, 1989).

A translational enhancer is defined as an mRNA sequence capable of increasing the rate of protein synthesis. These elements can act as binding sites for trans-acting factors that directly or indirectly interact with the translational machinery. They can also affect localization of mRNAs by tagging them for transport to cytosolic regions of high or low translational activity (Kwon, 1999). Translational enhancers are commonly found in the 5’ or 3’ UTRs of mRNAs (Komarova, 2005; Chizhikov, 2000; Takahashi, 2007 and Turner, 1999). Few translational enhancers have been described in mammalian systems (Kwon, 1999; Stein, 1998 and Rubtsova, 2003). A 21-nucleotide RNA trafficking signal in the 5’ UTR of the myelin gene has been identified as a translational enhancer and functions by facilitating RNA transport along microtubules (Kwon, 1999). Internal ribosome entry sequences (IRES) are present in the 5’ UTRs of the human Hsp70 and VEGF genes (Stein, 1998). The Hsp70 IRES sequence enhanced translation in a cap-independent manner more than 100-fold over an empty vector control (Rubtsova, 2003). In theory, mammals can use translation enhancement mechanisms similar to those employed by bacteria and viruses. Translational enhancers have been more
thoroughly studied in these lower organisms. In *E. coli* and other bacteria, A-U rich regions in 5’ UTRs enhance translation and stabilize mRNAs (Komarova, 2005). Viruses have evolved novel mechanisms to achieve a maximum rate of protein synthesis and multiplication inside their hosts. Viral translation enhancers are more often found in the 3’ UTR, but 5’ UTR enhancers are also common (Chizhikov, 2000). Translational enhancers in viruses have been reported to enhance translation efficiency by up to 30-fold (Turner, 1999).

4. Summary

Pulmonary surfactant, and especially a protein component of surfactant known as SP-A, are critical for proper respiratory function in humans and other higher organisms. The 5’ UTRs of the SP-A1 and SP-A2 genes are complex because of they each consist of 4 exons with unique patterns of splice variation. Splice variation within the 5’ UTR likely affects expression of transcripts because the 5’ and 3’ UTRs play a central role in regulating translational processes. Translational regulation is advantageous for numerous reasons. Sequences that increase the translational activity of a transcript are known as translational enhancers. Translational enhancers function through a wide range of mechanisms and the available literature on translational enhancers suggests that much about these fascinating sequences remains to be discovered.
II. Objectives

Several common 5’ UTR splice-variants of the SP-A gene have been studied with regards to the translatability and mRNA stability of downstream sequences (Wang, 2005). Various SP-A 5’ UTR splice-variants (ABD, A’D’, AB’D’ and A’CD’) were cloned upstream of a firefly luciferase gene and downstream of an SV40 promoter. Using the luciferase assay, it was determined that the ABD splice-variant, which is found only among SP-A2 transcripts, is approximately 4 times more active than the SP-A1 splice-variants A’D’, AB’D’, or A’CD’. The A’D’ and A’CD’ splice-variant constructs had half of the activity of the AB’D’ construct. When the relative mRNA content of the luciferase transcripts was measured with real-time RT-PCR, the SP-A2 transcript (ABD) was found to be approximately twice as abundant as the 3 SP-A1 transcripts. No significant differences in the mRNA content were found among the A’D’, AB’D’, and A’CD’ variants (Wang, 2005).

Based on the luciferase assay and real-time PCR data, a translational index was calculated by dividing the relative luciferase assay activity by the relative mRNA content (Wang, 2005). The translational index is thus an indirect measure of translation efficiency. Translational efficiency was found to be increased at least 2-fold by the two 5’ UTRs containing some form of exon B (the ABD and AB’D’ variants). Translation efficiency of the ABD splice-variant was improved by approximately 50% over the AB’D’ splice-variant. The A’D’ and A’CD’ splice-variant constructs, which did not contain any B exon sequence, exhibited translational efficiencies below the luciferase control vector, which did not contain a 5’ UTR sequence. Based on the results of these experiments, it was hypothesized that exon B sequence, through some unknown mechanism, confers increased translational capabilities to a transcript.

The objectives of the present study are:

1) **Determine whether SP-A 5’ UTR exon B is a cis-acting translational enhancer.** Due to splice-variant and 5’ UTR isoform differences in the previous study, the direct effects of exon B on translation and mRNA stability could not be determined and its translational enhancement capacity could not be quantified. Exon B function can be studied in the context of the whole SP-A 5’ UTR by deleting this sequence or by replacing this sequence with a 30-nucleotide random sequence. By cloning the 30-nucleotide exon B sequence into vectors containing 2 different guest 5’ UTRs, then comparing the activity and mRNA levels of these vectors with guest 5’ UTR-containing constructs that lacked the exon B sequence, the function of SP-A 5’ UTR exon B can be analyzed in a more direct way.
2) If SP-A 5’ UTR exon B is a cis-acting translational enhancer, identify segments of this untranslated exon essential for enhancement. By developing a series of constructs with 5-nucleotide deletions in the B exon, and measuring the activity of these constructs with the luciferase assay, it may be possible to determine the critical region of exon B for translational efficiency. Moreover, although this is beyond the current scope, this approach may identify regions where trans-acting factors bind to enhance translation.
III. Methods

1. Preparation of Plasmid Constructs

Experimental plasmids used in the experiments were derived from a single modified pcDNA3 vector (Invitrogen, Figure 2 Panel A), termed rpcDNA3/5’-UTR/LUC (Wang, 2005). To obtain the rpcDNA3/5’-UTR/LUC plasmid, pcDNA3 was first digested with BamHI and EcoRI restriction enzymes, which removed a 33-bp fragment containing a putative transcription start site. The large fragment was purified and used as the template in a polymerase chain reaction (PCR) containing primers 1046 and 1047. Primer 1046 contains HindIII and BamHI restriction enzyme recognition sites at its 5’ end. In contrast, primer 1047 contains XhoI and BamHI restriction enzyme recognition sites at its 5’ end. The PCR product was purified and digested with BamHI. Finally, the vector was ligated to remove a 133-bp fragment, and the plasmid was purified and sequenced.

Figure 2

Original and Modified Plasmid Structures

A. Original pcDNA3 Plasmid

A) The original pcDNA3 vector is shown in Figure 2 panel A. The multisite cloning region consists of the restrictions sites between the T7 and Sp6 promoters. “P CMV” is a cytomegalovirus promoter, which was replaced with an SV40 promoter. “BGH pA” represents the bovine growth hormone polyadenylation signal and “f1 ori” is the origin of replication. Ampicillin and neomycin selection genes are shown. The Colicin E1 resistance gene is
represented by “ColE1”. Boxes with “SV40” represent SV40 promoter sequences. B) The modified plasmid contained the selection genes and other elements of the pcDNA3 vector, but contained an SV40 promoter (SV40 P), 5’ UTR sequence, a firefly luciferase gene, and a polyadenylation signal. Relevant restriction sites are shown.

A firefly luciferase gene was inserted into the modified pcDNA3 plasmid. Restriction enzymes XhoI and BamHI were used to cut the 1.7 kb firefly luciferase gene from a pGEM-luc DNA plasmid. The same enzymes were used to digest the modified rpcDNA3 vector, and the luciferase gene was ligated into this plasmid. Sequencing confirmed the presence of the firefly luciferase insert.

An ABD SP-A2 5’ UTR splice-variant was inserted into the modified rpcDNA3 plasmid. The ABD insert was 100 nucleotides in length plus flanking restriction sites. The ABD insert was generated by PCR of an SP-A cDNA clone. HindIII and BamHI enzymes were used to clone the 5’ UTR sequence into the plasmid.

Finally, the CMV promoter of the rpcDNA3 construct was replaced with the weaker SV40 promoter. Although both promoters are constitutively active, SV40 induces a lower level of transcription than CMV. The pSVLSV40CAT vector was used to generate the SV40 promoter sequence by PCR. The SV40 promoter PCR product was digested with restriction enzymes NruI and HindIII prior to cloning into the modified rpcDNA3 plasmid. Sequencing confirmed the orientation of the SV40 promoter and downstream ABD SP-A 5’ UTR variant. No further modifications were necessary for the ABD construct used in subsequent experiments.

The modified construct is shown in Figure 2 Panel B. The luciferase (LUC) construct was identical to the ABD construct but did not have a cloned 5’ UTR upstream of the firefly luciferase gene. To generate the other plasmids, the ABD construct was digested with HindIII and BamHI restriction enzymes in reactions containing the following: 6 μL ABD plasmid DNA (373 ng/μL), 6 μL buffer 2 (New England Biolabs), 0.6 μL BSA (10 mg/mL), 1 μL (10 units) BamHI enzyme (Promega), 1 μL (20 units) HindIII (New England Biolabs), and 35 μL deionized water. This reaction mix was incubated at 37°C overnight.

The digestion products were extracted from a 0.8% agarose gel. A Qiagen gel extraction kit was used for this purpose. Briefly, the band containing the linearized plasmid was excised from the gel. Buffer QG (Qiagen) (3 volumes) was added to each gel slice. Tubes containing the buffer and agarose were incubated for 10 minutes at 50°C with occasional vortexing.
Isopropanol (1 volume) was added to each tube and the solution was added to Qiaquick spin columns (Qiagen). The spin columns were centrifuged for 1 minute at 15700 x g and the flow-through was discarded. To each column, 750 μL buffer PE (Qiagen) was added and the columns were centrifuged again. The flow-through was discarded and another centrifugation was done to remove residual alcohol. The columns were placed in clean tubes, and 50 μL of deionized water were added to each column prior to a final centrifugation. The purified DNA was then analyzed with a 0.8% agarose gel.

Small (<70 bp) insert DNA fragments were generated by annealing 2 single-stranded oligonucleotides. The 2 oligos were allowed to anneal (as described below) and the resulting double stranded inserts were ligated into the digested rpcDNA3 vector. Inserts for Constructs 1, 2, 3, and 5 (Table 1) were generated this way. Larger (>70 bp) insert fragments were made using overlapping PCR. Larger overlapping oligonucleotides were annealed at their overlapping regions and served as a template for the extension step. Overlap primers were used in 33-cycle PCR reactions to amplify the double-stranded insert DNA. Inserts for Constructs 4, 6, 9, and 10 (Table 1) were generated this way. Oligonucleotides were produced by the core facilities at Penn State University College of Medicine.

Table 1

<table>
<thead>
<tr>
<th>Construct</th>
<th>Oligos Used (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. SP-A 5’ UTR Exon B with SP-B 5’ UTR</strong></td>
<td>GW 1615 &lt;br&gt;GGGAAGCTTGTCGATTTTCTTGGAGCCTGA &lt;br&gt;AAAGAAGGCTGAGGATGCCGGATCCAGG &lt;br&gt;GW 1616 &lt;br&gt;CCTGGATCCGGCACCTCTGCAGCAAGCTTCCC</td>
</tr>
<tr>
<td>5. SP-D 5’ UTR Only</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td><strong>OVERLAP PCR PRIMERS:</strong></td>
<td></td>
</tr>
<tr>
<td>BS 1620</td>
<td></td>
</tr>
<tr>
<td>CCTGGATCCGGCAGGTTTTC</td>
<td></td>
</tr>
<tr>
<td>BS 1621</td>
<td></td>
</tr>
<tr>
<td>GGGAAAGCTTGTGCTGCTGAT</td>
<td></td>
</tr>
<tr>
<td>BS 1622</td>
<td></td>
</tr>
<tr>
<td>CCTGGATCCGGCAGGTTTTC</td>
<td></td>
</tr>
<tr>
<td>BS 1623</td>
<td></td>
</tr>
<tr>
<td>GGGAAAGCTTGTGCTGCTGAT</td>
<td></td>
</tr>
<tr>
<td>BS 1624</td>
<td></td>
</tr>
<tr>
<td>CCTGGATCCGGCAGGTTTTC</td>
<td></td>
</tr>
<tr>
<td>BS 1625</td>
<td></td>
</tr>
<tr>
<td>GGGAAAGCTTGTGCTGCTGAT</td>
<td></td>
</tr>
<tr>
<td>BS 1626</td>
<td></td>
</tr>
<tr>
<td>CCTGGATCCGGCAGGTTTTC</td>
<td></td>
</tr>
<tr>
<td>BS 1627</td>
<td></td>
</tr>
<tr>
<td>GGGAAAGCTTGTGCTGCTGAT</td>
<td></td>
</tr>
<tr>
<td>BS 1628</td>
<td></td>
</tr>
<tr>
<td>CCTGGATCCGGCAGGTTTTC</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6. Random Sequence with SP-D 5’ UTR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OVERLAP PCR PRIMERS:</strong></td>
</tr>
<tr>
<td>BS 1629</td>
</tr>
<tr>
<td>GGGAAAGCTTGTGCTGCTGAT</td>
</tr>
<tr>
<td>BS 1630</td>
</tr>
<tr>
<td>CCTGGATCCGGCAGGTTTTC</td>
</tr>
<tr>
<td>BS 1631</td>
</tr>
<tr>
<td>GGGAAAGCTTGTGCTGCTGAT</td>
</tr>
<tr>
<td>BS 1632</td>
</tr>
<tr>
<td>CCTGGATCCGGCAGGTTTTC</td>
</tr>
<tr>
<td>BS 1633</td>
</tr>
<tr>
<td>GGGAAAGCTTGTGCTGCTGAT</td>
</tr>
<tr>
<td>BS 1634</td>
</tr>
<tr>
<td>CCTGGATCCGGCAGGTTTTC</td>
</tr>
<tr>
<td>BS 1635</td>
</tr>
<tr>
<td>GGGAAAGCTTGTGCTGCTGAT</td>
</tr>
<tr>
<td>BS 1636</td>
</tr>
<tr>
<td>CCTGGATCCGGCAGGTTTTC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. SP-A 5’ UTR AD Variant (Exon B Knockout)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OVERLAP PCR PRIMERS:</strong></td>
</tr>
<tr>
<td>BS 1637</td>
</tr>
<tr>
<td>GGGAAAGCTTAAACTTGGAGGAGCAGAGACC</td>
</tr>
<tr>
<td>BS 1638</td>
</tr>
<tr>
<td>CCTGGATCCGGCTCTGGAGGAGCAGAGACC</td>
</tr>
<tr>
<td>BS 1639</td>
</tr>
<tr>
<td>GGGAAAGCTTAAACTTGGAGGAGCAGAGACC</td>
</tr>
<tr>
<td>BS 1640</td>
</tr>
<tr>
<td>CCTGGATCCGGCTCTGGAGGAGCAGAGACC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. SP-A 5’ UTR A- Random Sequence-D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OVERLAP PCR PRIMERS:</strong></td>
</tr>
<tr>
<td>BS 1641</td>
</tr>
<tr>
<td>GGGAAAGCTTAACTTGGAGGAGCAGAGACC</td>
</tr>
<tr>
<td>BS 1642</td>
</tr>
<tr>
<td>CCTGGATCCGGCTCTGGAGGAGCAGAGACC</td>
</tr>
</tbody>
</table>
Underlined sequences identify the primers used for the overlapping PCR reactions.

For inserts made by simply annealing 2 DNA strands, 10x annealing buffer was made by adding 2 mL of 0.5 M Tris HCl (0.5 M), 2 mL of 5 M NaCl, and 200 μL of 0.5 M EDTA to 6 mL of deionized water. Each oligonucleotide (1 μL at 1 μg/μL), 5 μL of 10x annealing buffer, and 43 μL deionized water were then added to a 1.5 mL tube and incubated at 95° C for 5 minutes. The tube was allowed to cool to room temperature for 1 hour and then placed on ice.

For inserts made by overlap PCR, initial PCR reactions were set up containing: 5 μL of each oligonucleotide (1 μg/μL), 2 μL dNTPs (10 mM of each nucleotide) (Roche), 0.5 μL (2.5 units) high fidelity Taq polymerase (Roche), 2.5 μL buffer 1 (Roche), 2.5 μL buffer 2 (Roche), and 32.5 μL deionized water. A GeneAmp PCR System 9700 PCR machine was used to extend the sequence using the following program: 2 minutes at 94° C, 30 seconds at 94° C*, 30 seconds at 60° C*, 1 minute at 72° C*, end at 4° C. Steps with an asterisk were repeated for 2 cycles. The double-stranded overlap PCR sequences were then amplified in a PCR reaction. Reactions contained the following: 5 μL overlap PCR product, 2 μL of each overlap PCR primer (1 μg/μL), 5 μL dNTPs (10 mM of each nucleotide) (Roche), 2.5 μL buffer 1 (Roche), 2.5 μL buffer 2 (Roche), 1 μL (2.5 units) high fidelity Taq polymerase (Roche), and 30 μL deionized water. The following PCR program was used to amplify the insert sequences: 2 minutes at 94° C, 30 seconds at 94° C*, 30 seconds at 58° C*, 1 minute at 72° C*, 7 minutes at 72° C, end at 4° C. Steps with an asterisk were repeated for 33 cycles.

Following the generation of insert sequences, the double-stranded insert DNA samples were digested with HindIII and BamHI enzymes. This provided the staggered ends for more efficient annealing. With the following reagents, 50 μL reactions were set up: 6 μL double-stranded insert DNA, 6 μL buffer 2 (New England Biolabs), 1 μL (10 units) BamHI enzyme (Promega), one μL (20 units) HindIII (New England Biolabs), and 35 μL deionized water. This reaction mix was incubated at 37° C overnight.

The digested insert DNA was then column purified. Larger insert sequences (>60 bp) were purified with G-50 STE Midi Select-D columns (Shelton Scientific). Briefly, the columns were prepared by resuspending the gel and removing the top and bottom closures. Columns were placed in collection tubes and centrifuged at 15700 x g for 90 seconds. The columns were placed in new collection tubes and 50 μL of insert DNA were added. After 2-3 minutes, the tube and
column were centrifuged at 15700 x g for 90 seconds. The insert DNA elution was then analyzed with a 2% agarose gel.

Shorter insert sequences (<59 bp) were purified with G-25 STE Select-D columns (Shelton Scientific). The columns were prepared by removing the closures and draining the buffer from the columns. Columns were placed in 15 mL conical tubes and centrifuged at 1100 x g for 1 minute. A Beckman GS-6R centrifuge was used. Buffer was removed and the centrifugation was repeated. The column was then placed in a collection tube and 50 μL of insert DNA were added to the column gel bed. The columns were spun down for 4 minutes at 1100 x g. The eluted DNA in the collection tubes was analyzed with a 2% agarose gel.

The insert DNA fragments were ligated into the HindIII/BamHI digested vector. The vector was the same for all constructs. Ligation reactions were arranged with the following reagents: 2 μL vector DNA, 4 μL insert DNA, 1 μL ligase buffer (Roche), 1 μL (1 unit) T4 ligase (Roche), and 2 μL deionized water. Control ligation reactions containing no insert DNA were set up in parallel to all ligations. Ligation reactions were incubated at 14° C overnight.

Ligated DNA samples were transformed with XL-10 Gold Ultracompetent cells (Stratagene) to amplify the newly formed constructs. Individual colonies were picked from transformation plates and used to inoculate minicultures of LB-Ampicillin. Plasmid DNA was isolated from minicultures with a miniprep kit (Qiagen). Nucleic acid content was quantified and plasmid DNA was sent to the Molecular Biology Core Facility at Penn State College of Medicine for sequencing analysis. Large scale preparations were completed for samples with perfect sequences.

2. Site-Directed Mutagenesis

Site-directed mutagenesis was carried out using the Quickchange II XL Site-Directed Mutagenesis kit (Stratagene). Appropriate sense and antisense primers were designed to contain the 5-nucleotide deletions of interest (both primers contained the same mutations and annealed to the same sequence on opposite strands of the plasmid DNA). The mutagenesis primers were optimized with melting temperatures less than 78° C according to the following equation:

\[ Tm = 81.5 + 0.41(GC\%) - 675/N \]

where N is the total number of nucleotides.
and ranged in length from 41 to 46 nucleotides. Attempts were made to center the deleted sequences in the primers and to maintain at least 15 bases of correct sequence on both sides of the deletions. The primers and their sequences are listed in Table 2.

The primers were produced by the Core Facility at the Penn State College of Medicine. Upon receipt of the oligonucleotides, they were diluted to a concentration of 1 μg/μL. To make the ABD splice-variant deletion mutations, the following primers were used in the site-directed mutagenesis PCR reactions (described below).

**Table 2**

**Primers Used in Site-Directed Mutagenesis Reactions**

<table>
<thead>
<tr>
<th>Deletion</th>
<th>Sense Primer (5’ → 3’)</th>
<th>Antisense Primer (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GW 1525</td>
<td>GW 1526</td>
</tr>
<tr>
<td></td>
<td>CAAGCAGCTGGAGCTCTGTG</td>
<td>CAGGCTCCAAGAAATCAGCGA</td>
</tr>
<tr>
<td></td>
<td>GTCGCTGATTTCCTGGAGGCTG</td>
<td>CCACAGAGGCTCCAGCTGCTTG</td>
</tr>
<tr>
<td>2</td>
<td>GW 1527</td>
<td>GW 1528</td>
</tr>
<tr>
<td></td>
<td>GCTGAGGGCTCTGTGTTGTTG</td>
<td>CTTTCAGGCTCCAAGAAATCA</td>
</tr>
<tr>
<td></td>
<td>GATTTCCTGGAGGCTGAAAG</td>
<td>CCACACACACACACACACAC</td>
</tr>
<tr>
<td>3</td>
<td>GW 1529</td>
<td>GW 1530</td>
</tr>
<tr>
<td></td>
<td>GAGGCTCTGTGTTGTTGTCGTC</td>
<td>CTGCTGTTTTTCAGGCTCCAAGA</td>
</tr>
<tr>
<td></td>
<td>GCTGAGCCGGGAAAAGAAGCAG</td>
<td>GCGACACACACACACACAC</td>
</tr>
<tr>
<td>4</td>
<td>GW 1531</td>
<td>GW 1532</td>
</tr>
<tr>
<td></td>
<td>CTGTGTGTGGTGGTGCGTGATTTG</td>
<td>CTCTCGTTCTCTCTCAAGCTCA</td>
</tr>
<tr>
<td></td>
<td>AGCCTGAAAAGAAGCAGGAG</td>
<td>ATCAGGCAACCCACACAC</td>
</tr>
<tr>
<td>5</td>
<td>GW 1533</td>
<td>GW 1534</td>
</tr>
<tr>
<td></td>
<td>GTGTTGGCTCCTGATTTTTGTTG</td>
<td>CGCTGTCCTCGTCTTTTTCAC</td>
</tr>
<tr>
<td></td>
<td>GAAAGAAGACAGAGGACAGG</td>
<td>AAGAAATCGACGAGACACAC</td>
</tr>
<tr>
<td>6</td>
<td>GW 1535</td>
<td>GW 1536</td>
</tr>
<tr>
<td></td>
<td>GGTGCCTGATTTCCTGGAGCCCA</td>
<td>CCAGTGCCCTGCTCGTCTTCTTG</td>
</tr>
<tr>
<td></td>
<td>GAAGCAGAGCAGGACTGG</td>
<td>GCTCCAAGAAATCAGGACC</td>
</tr>
<tr>
<td>7</td>
<td>GW 1537</td>
<td>GW 1538</td>
</tr>
<tr>
<td></td>
<td>GTGATTTCCTGGAGCCTGAAAAC</td>
<td>GGTCCAGCTCAGTGCTCTGTT</td>
</tr>
<tr>
<td></td>
<td>AGGAGCAGCCAGCTGGACC</td>
<td>TCCAGGCTCCAAGAAATCAGC</td>
</tr>
<tr>
<td>8</td>
<td>GW 1539</td>
<td>GW 1540</td>
</tr>
<tr>
<td></td>
<td>GATTTCCTGGAGCGCTGAAAAGA</td>
<td>GCTCTGGCTCCAAGTGCTGCTC</td>
</tr>
<tr>
<td></td>
<td>GGCAGCGACTGGAGCACCAGAGC</td>
<td>TCTTTCAGGCTCCAAGAAATC</td>
</tr>
</tbody>
</table>

The site-directed mutagenesis primers from Table 2 were used in mutagenesis PCR reactions. The following reagents were combined in PCR tubes: 5 μL 10x reaction buffer (Stratagene), 1 μL (10 ng) DNA template (ABD splice-variant luciferase construct), 1 μL (1 μg)
of each site-directed mutagenesis primer (see above), 1 μL dNTP mix (10 mM of each nucleotide) (Stratagene), 3 μL QuikSolution (Stratagene), 37 μL dH2O, and 1 μL (5 units) PfuUltra high fidelity DNA polymerase (Stratagene). Total reaction volume for each reaction was 50 μL.

A GeneAmp 9700 thermocycler (Applied Biosystems) was used to amplify the plasmid sequence and introduce the desired deletion mutations. Cycling parameters used for the reactions were: 95° C for 1 minute, 95° C for 50 seconds*, 60° C for 50 seconds*, 68° C for 8 minutes*, and 68° C for 7 minutes. Steps with an asterisk were repeated for 18 cycles.

To remove the DNA with the original sequence, the PCR products were digested with DpnI enzyme, which cuts only methylated and hemi-methylated DNA. The new PCR products were not digested because they were not methylated. DpnI enzyme (1 μL or 20 units) was added to each reaction and briefly mixed with a pipet. The reactions were centrifuged for 1 minute and then incubated at 37° C for 1 hour to digest the non-mutated DNA. Following the restriction digestion reaction, the digested PCR product plasmid DNA was used to transform XL-10 Gold Ultracompetent cells. Following a plasmid purification, sequences for the region of interest were obtained to confirm the efficacy of the mutagenesis reaction.

3. Isolation of DNA Plasmids from Bacteria Cultures

When isolating plasmid DNA from bacterial cultures, miniprep kits (Qiagen) were employed. Each miniculture was grown in 3 mL of LB-Ampicillin media, and 0.5-1 mL of this miniculture medium was transferred to a microcentrifuge tube. These tubes were spun down for 2 minutes at 15700 x g. The supernatant was removed and the pellet was resuspended in 250 μL of chilled resuspension buffer (Buffer P1) containing RNase. To each tube, 250 μL of lysis buffer (Buffer P2) were added and the tubes were inverted 5-10 times. Approximately 350 μL of neutralization buffer (Buffer N3) were then added to each tube and the tubes were again inverted several times. This was followed by a 10-minute centrifugation at 15700 x g. The supernatant from this spin was transferred to a QIAprep spin column within a microcentrifuge tube by pipetting, and the columns were centrifuged for 1 minute at 15700 x g. To the column, 500 μL of binding buffer (Buffer PB) were added before another 1-minute spin at 15700 x g. The flow-through was discarded. Prior to a 1-minute spin, 750 μL of wash buffer (Buffer PE) were added. The flow-through was decanted off and another spin was executed in order to remove any residual wash buffer in the columns. The columns were transferred to new 1.5 mL carrier tubes.
and 50 μL of deionized water were applied to the center of the columns. The tubes with columns were centrifuged for 1 minute to elute the DNA. This final centrifugation was repeated following the addition of another 30 μL of deionized water.

4. Large Scale DNA Preparations

Large scale DNA preparations were completed with a Qiagen Maxiprep kit (Qiagen). The contents of a 16-20 hour miniculture were decanted into 200 to 500 mL LB-Ampicillin media. This large scale culture was then incubated for 12-16 hours at 37° C with shaking (approximately 250 RPM). The cultures were then split into 2 centrifugation tubes and spun down at 6000 x g for 15 minutes at 4° C. A Beckman J2-M1 centrifuge was used for all centrifugation steps. The supernatant was decanted and 1 bacterial pellet was saved for the maxiprep while the other tube was stored in a -20° C freezer. The pellet for maxiprep was resuspended in 10 mL P1 buffer (Qiagen). P2 buffer (Qiagen) in the amount of 10 mL was then added and the tube was inverted several times to mix the contents. After 5 minutes, 10 mL of 4° C buffer P3 (Qiagen) were added. The tube was then incubated on ice for 20 minutes. The contents of the tube were then centrifuged at 20000 x g for 30 minutes at 4° C. The supernatant was centrifuged again for 15 minutes at the same speed and temperature. A Qiagen-tip 500 (Qiagen) was equilibrated and the contents of the second centrifugation were added to the column. The column was then washed twice with 30 mL wash buffer (Qiagen). The DNA was then eluted with elution buffer (Qiagen) into a polypropylene centrifuge tube.

DNA was precipitated with 10.5 mL 2-propanol and centrifuged for 30 minutes at 15000 x g and 4° C. The supernatant was removed and the pellet was washed with 70% ethanol solution. The tube was then centrifuged at 15000 x g for 10 minutes and the ethanol was poured off. The DNA pellet was briefly air-dried and redissolved in 500 μL 0.1x TE buffer. The DNA that remained in the tube was obtained by washing again with 250 μL 0.1x TE buffer. Both fractions of DNA were then quantified with an agarose gel and nanodrop.

5. Transformations

Transformations were accomplished using either competent E. coli (HL2) stock cultures (prepared by G. Wang) or XL-10 Gold Ultracompetent cells (Stratagene). The procedures differed slightly at a few steps, but were otherwise similar. When transforming XL10 Gold Ultracompetent cells, 45 μL of cells per transformation were gently thawed on ice for several minutes, and aliquoted to pre-chilled microcentrifuge tubes. β-mercaptoethanol mix (1-2 μL) was
added to the cells, and the competent bacteria were incubated on ice for 10 minutes. Ligation or PCR product DNA (10-100 ng) was then added to the cells, followed by a 30-minute incubation on ice. Next, the cells were heat-pulsed for 50 seconds in a 42°C waterbath and cooled on ice for 2 minutes. To the cells, 800 μL of S.O.C. media (Invitrogen) were added. The transformation tubes were incubated at 37°C with shaking at 200-250 RPM for 1 hour. Finally, the cells were concentrated by centrifuging at 15700 x g for 2 minutes, pipeting off 300-500 μL of supernatant, and redissolving the pelleted cells. A fraction of these cells, usually 200 μL, was then spread on LB-Ampicillin plates. The plates were inverted after 10 minutes and incubated at 37°C overnight. A negative control transformation, which contained no DNA, was done in parallel to all experimental transformations.

6. Quantification of Nucleic Acid Yield and Purity

Single-stranded DNA oligonucleotides, plasmid DNA, and RNA samples were analyzed using a Nanodrop ND-1000 spectrophotometer. This full-spectrum spectrophotometer is capable of measuring 1 μL samples with high accuracy and reproducibility. Before each use, the pedestal was cleaned twice with 1.5 μL dH2O and initialized with dH2O. The instrument was blanked with the solvent that each nucleic acid was dissolved in. To the receiving fiber optic cable on the pedestal, 1 μL of each sample was applied, and the readings were taken and stored on a Dell Latitude D520 laptop. Readings were taken with the cursor position at 230 nm. After use, the pedestal was cleaned twice with dH2O.

For real-time RT-PCR work, a Beckman Coulter DU640B spectrophotometer was used to measure RNA, primer, and probe concentration and purity. A cuvette with a pathlength of 1.0 cm was used. The instrument was blanked with the appropriate solvent, either dH2O or 0.1x TE buffer, and 100 μL of a 1:100 or 1:50 dilution of sample were used in each measurement. The optical density at 260 nm was recorded and used to determine the oligo or RNA concentration. RNA concentration was calculated using the formula: concentration = O.D.260 x dilution factor x 40. Primer and probe concentrations were calculated according the formula: concentration (M) = (O.D.260 x dilution factor) / (sum of extinction coefficients x cuvette pathlength in centimeters).

7. Cell Culture with NCI-H441 Cells

NCI-H441 cells were originally obtained from the American Type Culture Collection (Manassas, Virginia). These cells were cultured in Invitrogen RPMI-1640 medium containing 1% antibiotic/antimycotic (penicillin, streptomycin, and amphotericin B) solution (Sigma-
Aldrich), 1% L-glutamine solution (Sigma-Aldrich), and 10% heat-inactivated fetal bovine serum (FBS) (Invitrogen). The cells were maintained in a 37° C, 5% CO₂ atmosphere (Nuaire water-jacketed incubator).

The cells were subcultured weekly or twice per week, when confluency of the cell culture plates reached 90%. Subculturing was accomplished by the following procedure. The excess medium was vacuumed from the plates, and the plates were briefly washed with 3-5 mL of phosphate-buffered saline (PBS) (Invitrogen). The PBS was removed and 1 mL of 0.25% Trypsin-EDTA solution (Sigma) was added to each plate. After 5 minutes of occasional tilting of the plates, 6 mL of RPMI-1640 medium were added to each plate to dilute the trypsin. Any loose cells were removed with a pipet and the cells were transferred to 15 mL conical tubes. The cells were then centrifuged at 1500 x g for 5 minutes with a Sorvall Legend RT centrifuge. The medium was vacuumed off and 5-12 mL of new RPMI medium were added to resuspend the cells. To each new cell culture plate (Falcon) or 6-well plate (Becton-Dickinson) well, 1 mL of cells was added. RPMI medium (approximately 15 mL) was added to each standard plate, and 2 mL of medium were added to each well of any 6-well plates.

8. Preparation of Selective Antibiotic Plates

Ampicillin was used as a selective antibiotic during the cloning process. LB-Ampicillin plates were prepared by dissolving 10 grams of sodium chloride (Fisher Scientific), 5 grams of yeast extract (Fisher Scientific), and 10 grams of tryptone (Fisher Scientific) in 600 mL deionized water. Agar (BD Biosciences) (15 g) was added and the mixture was brought to 1 L with deionized water. This mixture was then autoclaved and allowed to cool to 55° C before adding ampicillin stock solution to achieve an ampicillin concentration of 100 mg/liter. The liquid agar was distributed to approximately 35 plates (Falcon) by pipetting (27.5 mL per plate). The agar plates were allowed to solidify at room temperature. Selective plates were then packaged and stored at 4° C.

9. Preparation of LB Medium

LB medium was prepared by dissolving 10 grams of sodium chloride (Fisher Scientific), 5 grams of yeast extract (Fisher Scientific), and 10 grams of tryptone (Fisher Scientific) in 600 mL deionized water in a large flask. The flask was swirled and the solution was brought to 1 L with deionized water. The mixture was autoclaved. After cooling, the LB medium was used immediately or aliquoted to sterile storage containers.
10. Transient NCI-H441 Cell Transfections

All transfection work was done in a sterile cell culture hood. When the cells were approximately 90% confluent, the cells were trypsinized and plated to 6-well culture plates. All cells from 1 standard cell culture plate (100 mm diameter) were split to the 6 wells of a 6-well culture dish 24 hours before transfection. RPMI 1640 medium with antibiotics/antimycotics and serum was used.

The medium was vacuumed off and replaced with Dulbecco’s Modified Eagle’s Medium (DMEM) without serum or antibiotics (Invitrogen) 4 hours before transfection.

Transfections were completed in duplicate. DNA from large-scale plasmid preps was used after spectrophotometric analysis. For each transfection, 1 μg of experimental plasmid was added to 100 μL of DMEM in a 5 mL polystyrene tube. *Renilla* luciferase control plasmid (pRL-SV40) (0.05 μg) was added to the same solution. To this DNA solution, 6 μL of Plus reagent (Invitrogen) were added. The tube was lightly swirled and incubated at room temperature for 15 minutes. In another polystyrene tube, 6 μL of Lipofectamine reagent (Invitrogen) were mixed with 100 μL of DMEM. This tube was then combined with the tube containing the DNA and Plus reagent and incubated at room temperature for 15 minutes. Following the incubation, 2 mL of DMEM were added to the complexed DNA solution, and the resulting volume was added to each cell-containing well of the 6-well plates. The transfection plates were cultured in a 5% CO₂ environment at 37°C.

The transfection medium was vacuumed from the wells and new DMEM with 10% FBS and antibiotics/antimycotics was added 4 hours after transfection. The cells were placed back into the 37°C incubator until used for an RNA extraction or luciferase assay.

11. Luciferase Assay

Luciferase assays were carried out by first performing transient transfections of NCI-H441 cells in duplicate. After 36 hours, the medium was vacuumed off of the cells and the cells were washed with phosphate-buffered saline (PBS). Passive lysis buffer (1x) (Promega) was prepared by dilution and 500 μL were added to each well of each 6-well plate. The plates were then gently shaken for 15 minutes at room temperature. The 500 μL lysates were then transferred to new 1.5 mL tubes and immediately placed on ice. The lysates were then centrifuged at 9300 x g for 1 minute.
Luciferase assay reagent II (Promega) and “Stop and Glo” reagent (Promega) were prepared prior to each experiment. Luciferase assay reagent II was distributed to 1 luminometer cuvette (BD Biosciences) per transfection. Background readings were taken prior to each set of readings. A volume of 20 μL of cell lysate was added to the LARII and activity (in relative light units per second) was measured by the luminometer (Zylux Corporation). The delay time was set at 3 seconds and read time was set at 10 seconds. Following the firefly luciferase reading, the Renilla luciferase reading was obtained by adding 100 μL of Stop and Glo reagent, briefly vortexing, and measuring activity with the luminometer. The delay and read times were the same as for the firefly luciferase readings. This protocol was followed for each lysate sample, which were kept on ice for the duration of the experiment. Following the luciferase assay readings of each set of lysates, 2 additional sets of readings were taken, so that the activity of each lysate sample was measured in triplicate.

12. Quantification of mRNA by Real Time RT-PCR

Prior to quantitative Real-Time PCR analysis, RNA samples were obtained by harvesting NCI-H441 cells 30 hours after transfection. Cells were briefly washed with 1 mL PBS. RNA-Bee solution (Tel-Test) (1 mL) was then added to each transfection well. Homogenization was accomplished by pipetting, and the cell lysates were transferred to chilled 1.5 mL tubes. To each tube, 0.2 mL chloroform was added. Tubes were centrifuged at 12000 x g for 15 minutes at 4° C. The aqueous phases were then transferred to new tubes and 0.5 mL 2-propanol was added to each tube. Following a 10-minute room temperature incubation, the tubes were centrifuged at 12000 x g for 5 minutes at 4° C. The supernatants were decanted off and 1 mL of 75% ethanol was added to each tube. The tubes were centrifuged at 7500 x g at 4° C for 5 minutes. The ethanol was removed and the pellets were air-dried for 10-15 minutes. Pellets were redissolved in 50 μL of nuclease-free water (Ambion). A Nanodrop 1000 spectrophotometer and Beckman Coulter 640B spectrophotometer were used to measure the concentrations of the RNA samples. Prior to use in real-time PCR reactions, the RNA samples were diluted to 100 ng/μL. RNA samples were stored at -80° C until used.

The purity of the RNA samples was first checked with a 2100 bioanalyzer spectrophotometer (Agilent Technologies) by the Functional Genomics Core Facility at the Penn State College of Medicine. Samples that passed this first quality control checkpoint were checked with PCR to determine if any contaminating DNA was present. The following reagents were added to PCR tubes: 5 μL RNA sample, 2 μL primer 1152 (sense primer for luciferase gene
– GCCCGCGAACGACATTTA, 2 μL primer 1153 (antisense primer for luciferase gene- 
TTTGCAACCCCTTTTTGGAA), 5 μL dNTPs (10 mM of each nucleotide) (Roche), 2.5 μL 10x 
buffer 1 (Roche), 2.5 μL 10x buffer 2 (Roche), 30 μL deionized water, and 0.5 μL (2.5 units) high 
fidelity Taq DNA polymerase (Roche). A GeneAmp 9700 thermocycler (Applied Biosystems ) 
was then used to amplify any contaminating DNA: 95° C for 2 minutes, 95° C for 30 seconds*, 
60° C for 30 seconds*, 72° C for 1 minute*, and 72° C for 7 minutes. Steps with an asterisk were 
repeated for 33 cycles. PCR products were then run on a 2% agarose gel.

A Taqman One-Step RT-PCR Master Mix Reagent kit (Applied Biosystems) was used to 
perform real-time PCR reactions. All probes and primers used with RT-PCR reactions targeted 
the luciferase gene, since this would eliminate any influences from the cells’ endogenous SP-A 
gene. Renilla luciferase (pRL-SV40) was used as the endogenous standard. A 1 μL volume (100 
ng mass) of RNA was combined with 19 μL of real-time PCR mix in each well. The mix 
included either primers 1152 and 1153 or 1156 and 1157 in 50 nM concentrations and a dual 
labeled probe at a concentration of 300 nM. Sequences of these primers are shown in Table 3. 
The probes for firefly (6-fam-CATTTCGAGCCTACCGTGGTTC-tamra) and 
Renilla (hex-TATCATGGCCTCGTGAAATCCCGTAGTAA-tamra) luciferase were obtained from Sigma- 
Aldrich. The fluorescent labels 6-fam and hex produced the signal in the PCR reaction when the 
quincher (tamra) was not present due to hybridization of probe and PCR product. An ABI 
PRISM 7900 sequence detection system was used to thermocycle the RT-PCR reactions on 384- 
well plates and detect the fluorescence during cycling. All of the 48 RNA samples were included 
on the same 384-well plate, in triplicate. The reverse transcription was carried out first and 
included 2 steps: 48° C for 30 minutes and 95° C for 10 minutes. Amplification of the cDNA 
then consisted of 43 cycles, each at 95° C for 15 seconds and 60° C for 1 minute. The cycle time 
at which fluorescence became statistically different than the baseline fluorescence was termed the 
threshold cycle number. The control firefly luciferase DNA construct was used to generate 2 
duplicate standard curves of copy number (10⁶ to 1 ug/μL) vs. threshold cycle number (1-43). 
DNA concentrations of 10⁶ through 1 μg/μL were included in the standard curves. The PRISM 
7900 software used these standard curves to determine copy number (of reverse-transcribed 
cDNA) for the experimental RNA samples. The mRNA level of each experimental sample was 
then normalized to the LUC control to obtain a relative mRNA content value for each sample.
### Table 3

**Primers Used in Real-Time RT-PCR Reactions**

<table>
<thead>
<tr>
<th>Primer Number and Name</th>
<th>Sequence (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1152 – Forward Primer for Firefly Luciferase</td>
<td>GCCCGCGAACGACATTTA</td>
</tr>
<tr>
<td>1153 – Reverse Primer for Firefly Luciferase</td>
<td>TTTGCAACCCCTTTTTGGAA</td>
</tr>
<tr>
<td>1156 – Forward Primer for Renilla Luciferase</td>
<td>GCAGCATATCTTGAACCATTCAAA</td>
</tr>
<tr>
<td>1157 – Reverse Primer for Renilla Luciferase</td>
<td>CATCACTTGCACGATAAAGCATTATA</td>
</tr>
</tbody>
</table>

### 13. Agarose Gels

Agarose gels were used for quality control purposes during the preparation of plasmid constructs and to check the quality of DNA prior to transfections. Agarose gels were prepared with 1x TAE buffer and run in 1x TAE buffer. Gels were made by adding an appropriate mass of agarose (Sigma) to 100 or 150 mL TAE (percentage of agarose was calculated on a weight/volume basis). This mixture was microwaved for 2.5 minutes on high setting, and 5-7 μL of ethidium bromide (EtBr) were added prior to pouring each gel. Following the solidification of the agarose gel, it was transferred to a gel platform and loaded. A voltage source (Whatman Biometra Model 250EX) was connected to the gel apparatus. Following the migration of DNA, gels were imaged with a UV Transilluminator (Mitsubishi). Images were saved as digital images or printed on photo paper.

### 14. PAGE Gels

Polyacrylamide gel electrophoresis (PAGE) (8%) gels were used to confirm the presence of insert sequences following ligations. PAGE gels were prepared with the following reagents: 1 mL 10x TBE buffer, 2 mL 40% bis-acrylamide, 7 mL deionized water, 40 μL 10% ammonium persulfate, and 10 mL tetramethylethylenediamine (TEMED). Reagents were mixed in a conical tube and added to the assembled gel apparatus. PAGE gels were run in 1% TBE buffer at 110 volts. A loading volume of 1 μL was used. The gels were imaged by removing the apparatus and soaking the gels in a dilute ethidium bromide solution. Gels were imaged with a UV Transilluminator (Mitsubishi), and images were either saved as digital files or printed directly onto photo paper.
15. Statistics

Data were sorted with a Microsoft Excel spreadsheet. For the luciferase assay and real-time PCR data, readings from each transfection were averaged so the sample size (n) is equal to the number of transfections of each construct. The error bars on all graphs represent the standard error (standard deviation divided by the square root of the sample size). The Student’s t-test was used to determine p-values for all comparisons. Comparisons with a p-value greater than 0.05 are considered significant. Lower p-values are indicated as p<0.01 and p<0.001.
IV. Results

To investigate the role of exon B in the SP-A2 5’ UTR, 3 studies were performed. First, exon B was examined in the context of the complete SP-A 5’ UTR. In the second study, 2 guest 5’ UTRs were cloned downstream of exon B sequence to investigate the translational enhancement capacity of exon B. A guest 5’ UTR is simply the 5’ UTR sequence of another human gene. Guest 5’ UTRs from the SP-B and SP-D genes were used because these 2 genes are expressed in NCI-H441 cells and their 5’ UTRs are relatively short. In the third study, a series of deletion mutations were made in the exon B region in order to identify a functional region of exon B sequence.

1. Whole SP-A 5’ UTR Study

Figure 6 shows the 4 constructs that were designed for the whole 5’ UTR study. Construct 9 (AD; Table 3) was identical to Construct 8 (containing the ABD SP-A 5’ UTR splice-variant) but did not contain the 30-nucleotide exon B. This construct was referred to as the AD splice-variant, although AD splice-variants are rarely found in vivo (and have not been detected among SP-A1 variants). It should be noted that the AD’ splice-variant, which differs from the AD sequence by only 3 nucleotides at the 5’ end of exon D, is the most commonly observed SP-A1 splice-variant (Karinch, 1995). Construct 10 was also similar to Construct 8 but contained the random 30-basepair sequence in place of the 5’ UTR exon B. The random sequence was the same as the one used in Constructs 3 and 6, and was simply a 30-nucleotide sequence from within the human SP-B gene’s coding region. Constructs 7 through 10 were each co-transfected into H441 cells with the Renilla luciferase control plasmid, which was used to standardize transfection efficiency in this study. The construct designs are shown in Figure 3, and the results of the luciferase assay are displayed in Figure 4. The luciferase assay results were normalized to the ABD splice-variant construct (Construct 8).

The results of the transfection experiment with the whole SP-A 5’ UTR constructs confirm that SP-A 5’ UTR exon B plays a vital role in modulating the activity of the 5’ UTR. Construct 8, the complete SP-A 5’ UTR construct, had an activity significantly higher than the construct that lacked the SP-A 5’ UTR exon B sequence (Construct 9). The construct containing the 30-basepair random sequence from the SP-B coding region (Construct 10) displayed a reduced level of activity, approximately 4 times lower than the wild type construct (Construct 8) (n=6). Comparison of the data from Constructs 8, 9, and 10 indicates that the size alone of the
ABD variant does not account for its increased activity (at least not at its entirety) since Construct 10 exhibited lower activity than the smaller Construct 9. All constructs had higher activity than the LUC control construct (Construct 7). The specific sequence of exon B appears to be essential for increasing activity, as the construct of identical size to Construct 8, but which contains a random sequence instead of exon B, exhibits activity of only 24% that of Construct 8. The AD construct, Construct 9, displayed activity of approximately 67% of Construct 8 activity. These data together indicate that replacing exon B with the random sequence disrupts the translatability of the transcript much more than simply deleting exon B. Thus, the random sequence actually plays a negative role when used in place of exon B.
Several constructs (7, 8, 9, and 10) were used to test the effects of SP-A 5’ UTR exon B in the context of the whole SP-A 5’ UTR (ABD SP-A2 splice-variant). The SV40 promoter is indicated by red circles, and poly A signals are illustrated with yellow circles. The 30-basepair exon B sequence is shown as a gray box between the A and D exons in Construct 8. The 5’ UTR in Construct 9 consists of 70 nucleotides and contains only the A and D exons. This plasmid was termed the AD splice-variant. Construct 10 is identical to Construct 8 except that Construct 10 contains a sequence from the SP-B gene’s coding region in place of exon B, and this is referred to as the “random” sequence. This sequence is identical to the 30-nucleotide random sequence used in Constructs 3 and 6. The Figure is not drawn to scale.
The 4 constructs for studying SP-A 5’ UTR exon B in the context of the whole SP-A 5’ UTR were co-transfected into NCI-H441 cells with the pRL-SV40 plasmid. Firefly and Renilla luciferase activities were measured 36 hours after transfection. Each experiment included 2 transfections of each construct, and 3 independent experiments were performed. The firefly and Renilla luciferase activities were measured in triplicate. The ratio firefly/Renilla luciferase was then calculated. The relative activity values were normalized to Construct 8 (SP-A 5’ UTR), which was set to a value of 1. Error bars indicate the standard errors of the ratios. *Construct 8 displayed significantly higher activity than Constructs 9 and 10 (p<0.01). #Construct 9 activity was significantly lower than the activities of Constructs 8 and 10 (p<0.01). ^The activity seen with Construct 10 was significantly lower than the activity of Construct 8 (p<0.01). LUC (Construct 7) activity was significantly lower than all other constructs (p<0.001).
2. Guest 5’ UTR Studies

a) With SP-B 5’ UTR Guest Sequence

In order to determine the translational enhancement capacity of SP-A 5’ UTR exon B, the 30-basepair exon B sequence was cloned immediately upstream of the guest 14-basepair SP-B 5’ UTR (noted as B in Figure 5), which was inserted in front of the luciferase gene in a recombinant pcDNA3 plasmid construct (Figure 5, Construct 1). Additional constructs, one containing a random coding sequence from the SP-B gene in place of exon B (Construct 3), and one containing only the SP-B 5’ UTR with no additional upstream sequence (Construct 2), were produced. These 3 plasmids, along with control plasmids, a firefly luciferase (LUC) plasmid that did not contain any 5’ UTR sequence (Construct 7) and a construct containing the ABD SP-A2 5’ UTR splice-variant upstream of firefly luciferase (Construct 8), were transiently transfected into H441 cells. Luciferase activity was measured 36 hours after transfection. The 5 plasmids are diagrammatically shown in Figure 5 and the luciferase assay results are displayed in Figure 6.

The activity values shown in Figure 6 are statistically significant compared to one another with p<0.05, except for the comparison of Constructs 2 and 3, which were not significantly different. Construct 1, containing exon B, displayed an approximately 2-fold increase in activity over Constructs 2 and 3, which did not contain the untranslated exon B. Activity seen with Construct 7, which did not have any 5’ UTR sequence, was significantly lower than that of each of the 4 other constructs. No construct reached a level of activity as high as the wild type, or ABD SP-A 5’ UTR splice-variant construct (n=6).

The mean activity ratio for Construct 1 for all 3 experiments was 0.46 ± 0.07. This activity was 175% higher than the mean activity of Construct 2, which was 0.17 ± 0.02. Activity of Construct 1 was 120% greater than the mean activity of Construct 3, which was 0.21 ± 0.02. The average activity ratio of the LUC construct, Construct 7, was 0.04, while activity of the complete 5’ UTR construct, Construct 8, was 1.52 ± 0.19.
The 5 constructs that were used to test the translation enhancing ability of SP-A 5’ UTR exon B are shown. In this series the 14-nucleotide SP-B 5’ UTR sequence served as a guest 5’ UTR and is depicted by blue “B” boxes in Constructs 1, 2, and 3. The SV40 promoter is indicated by red closed circles, and the poly A signals are illustrated with yellow open circles. The 30-basepair SP-A 5’ UTR exon B is shown as a white box in Construct 1, while the random sequence (from the SP-B coding region) is indicated by the light green box in Construct 3. The luciferase gene is represented as “LUC” (green boxes) in all constructs. The complete SP-A 5’ UTR (ABD splice-variant of SP-A2) is shown as a purple box. Constructs 7 and 8 are the same LUC (Construct 7) and whole 5’ UTR (Construct 8) plasmids that were used in the whole 5’ UTR study (Section 1). Each construct was assigned a number (1, 2, 3, 7, and 8) and a title (ie. Exon B + SP-B 5’ UTR, SP-B 5’ UTR, LUC, etc.). The Figure is not drawn to scale.
The luciferase assay results for the constructs in Figure 5 are shown. The 5 constructs were co-transfected into NCI-H441 cells with the pRL-SV40 plasmid. Firefly and Renilla luciferase activities were measured 36 hours later. Each experiment included 2 transfections of each construct, and 3 independent experiments were performed. The firefly and Renilla luciferase activities were measured in triplicate. The ratio firefly/Renilla luciferase was then calculated. Error bars indicate the standard errors of the ratios. *Significance bars indicate the significant comparisons (p<0.05) of Construct 1 with each of the other constructs under study. The activities of Constructs 2 and 3 were not significantly different from each other.
b) With SP-D 5’ UTR Guest Sequence

In order to confirm the enhancement ability of SP-A exon B, the experiments with the SP-B 5’ UTR were repeated using the SP-D 5’ UTR in place of the SP-B 5’ UTR. New constructs were made containing the SP-D 5’ UTR in the place of the SP-B 5’ UTR. Constructs 1, 2, and 3 were replaced with the new plasmid Constructs 4, 5, and 6. The structures of the constructs are shown in Figure 7, and the luciferase assay results are displayed in Figure 8.

The luciferase assay results seen in Figure 8 with the SP-D 5’ UTR as the guest 5’ UTR mirror the results in which the SP-B 5’ UTR was used. Again, the construct containing exon B displayed double the activity of Constructs 5 and 6, which did not contain SP-A 5’ UTR exon B sequence. All comparisons were statistically significant with p<0.05, except for the comparison between Constructs 5 and 6, which failed to reach statistical significance with the t-test (n=6). None of the constructs containing the guest 5’ UTR (SP-D) reached a level of activity as great as that seen with the ABD splice-variant construct (Construct 8).

The activity of Construct 4 for all 3 experiments was 0.73 ± 0.09. This activity was 77% higher than the activity of Construct 5, which was 0.41 ± 0.05. Activity of Construct 4 was 117% greater than the activity of Construct 6, which was 0.34 ± 0.05. The average activity of the LUC construct, Construct 7, was 0.03, while activity of the complete 5’ UTR construct, Construct 8, was 1.29 ± 0.10.
The 5 constructs that were used to test the translation enhancing ability of SP-A 5’ UTR exon B are shown. In this series the 43-nucleotide SP-D 5’ UTR sequence served as a guest 5’ UTR and is depicted by orange “D” boxes in Constructs 4, 5, and 6. The SV40 promoter is indicated by red closed circles, and the poly A signals are illustrated with yellow open circles. The 30-basepair SP-A 5’ UTR exon B is shown as a white box in Construct 4, while the random sequence (from the SP-B coding region) is indicated by the light green box in Construct 6. The luciferase gene is represented as “LUC” (green boxes) in all constructs. The complete SP-A 5’ UTR (ABD splice-variant of SP-A2) is shown as a purple box. Each construct was assigned a number (4, 5, 6, 7, and 8) and a title (ie. Exon B + SP-D 5’ UTR, SP-D 5’ UTR, etc.). The Figure is not drawn to scale.
Figure 8

Exon B is an Enhancer of Translation
When in the SP-D Guest 5’ UTR Model

The luciferase assay results for the constructs in Figure 7 are shown. The 5 constructs were co-transfected into NCI-H441 cells with the pRL-SV40 plasmid. Firefly and Renilla luciferase activities were measured 36 hours after transfection. Each experiment included 2 transfections of each construct, and 3 independent experiments were performed. The firefly and Renilla luciferase activities were measured in triplicate. The ratio firefly/Renilla luciferase was then calculated. Error bars indicate the standard errors of the ratios. *Significance bars indicate the significant comparisons (p<0.05) of Construct 4 with each of the other constructs under study. The difference in activities between Constructs 5 and 6 was not statistically significant.
c) Comparison of SP-B and SP-D 5' UTR Activity

While the luciferase assay results support the hypothesis that SP-A 5’ UTR exon B is a translational enhancer, they also reveal that the activity of the SP-D 5’ UTR is higher than that of the SP-B 5’ UTR in this in vitro system. Figure 9 displays the luciferase assay results from both guest 5’ UTR experiments in a single graph. For Constructs 1 through 6, the level of activity seen with the constructs containing the SP-D 5’ UTR was nearly twice that of the respective constructs containing the SP-B 5’ UTR. Statistical comparisons of constructs containing the SP-B and SP-D 5’ UTRs were significant with p<0.05 by t-test. It was observed that constructs with the SP-D 5’ UTR had higher activity than those with the SP-B 5’ UTR. This is consistent with previous studies where longer 5’ UTR sequences in the range of the SP-D 5’ UTR size lead to more stable mRNAs and higher translatability (Hirata 2004). In the 2 guest 5’ UTR experiments, the complete SP-A 5’ UTR construct (Construct 8) and the LUC construct (Construct 7) displayed consistent levels of activity (n=6).

Comparing Constructs 1 and 4 between the 2 experiments reveals that activity of Construct 4 is 61% higher than that of Construct 1. Activity of the construct containing only the SP-D 5’ UTR (Construct 5) was 150% higher than the activity of Construct 2, which contained only the SP-B 5’ UTR. Construct 6 displayed activity approximately 60% higher than Construct 3. It seems that the SP-D 5’ UTR’s enhancement over the SP-B 5’ UTR is dissipated when upstream 5’ UTR sequences are present. Alternatively, the lower activity of the shorter size SP-B 5’ UTR appears to be enhanced more when translation enhancer sequences are placed upstream, compared to the activity of the longer SP-D 5’ UTR.

The ratio of Construct 1 activity to Construct 2 activity is 2.8, while the ratio of Construct 4 activity to Construct 5 activity is 1.8. Thus, exon B enhancement over the guest 5’ UTR only construct appears to be greater when the SP-B 5’ UTR is used. The ratio of Construct 1 activity to Construct 3 activity is 2.2, while the ratio of Construct 4 activity to Construct 6 activity is also 2.2. Thus, exon B enhancement over the random sequence construct seems to be the same no matter which guest 5’ UTR is used. Together, these comparisons indicate that the SP-B and SP-D 5’ UTRs have an effect on the enhancement capacity of exon B. Because the SP-B 5’ UTR in Construct 2 is very small (14 nucleotides), the enhancement seen over the guest SP-B 5’ UTR construct could be due to a stabilizing effect from the extra sequence. Thus, 5’ UTR size may be an important factor in the difference in strength between the 2 guest 5’ UTRs. Furthermore, because the exon B enhancement over the random sequence constructs is identical no matter
which guest 5’ UTR is used, these constructs, which are identical in size (number of basepairs) to the exon B constructs, may be the better comparison tools for measuring exon B’s enhancement potential.
Figure 9

The SP-D 5’ UTR is Stronger than the SP-B 5’ UTR

This Figure depicts the combined data of Figures 6 and 8 to visually assess the activity differences between the SP-B and SP-D 5’ UTRs. The black bars indicate the luciferase activity ratios from the guest SP-B 5’ UTR experiment (Figure 6). The white bars represent the luciferase activity ratios from the guest SP-D 5’ UTR experiment (Figure 8). For each bar, n=6. Error bars indicate the standard errors of the ratios. Asterisks denote the significant differences (p<0.05) between SP-D and SP-B 5’ UTR constructs.
In order to determine whether increased RNA stability or an increase in translation efficiency is responsible for the different expression levels seen with Constructs 1 through 8, real-time RT-PCR was employed to determine the relative luciferase mRNA content in the transfected H441 cells 30 hours after transfection. The real-time RT-PCR results are displayed in Figure 10. Constructs containing SP-B and SP-D 5’ UTRs, as well as the LUC and complete SP-A 5’ UTR constructs, were included in the same experiment. RNA levels from 6 transfections were measured for each construct (n=6). For each sample, 3 real-time RT-PCR reactions were completed and averaged.

The relative mRNA content between the control constructs, the LUC (Construct 7) and the ABD (Construct 8), as determined by real-time RT-PCR in this study, was not consistent with published data from a previous study (Wang, 2005). Among the constructs containing the guest SP-B 5’ UTR, the mRNA level of the construct containing SP-A 5’ UTR exon B was much lower than that of the construct containing only the SP-B 5’ UTR. Levels of the transcript containing the random sequence (Construct 3) were not statistically different than levels of transcripts from Construct 1. No significant differences in RNA levels were seen among the constructs containing the guest SP-D 5’ UTR (Constructs 4, 5, and 6).

The fact that the relative mRNA levels of the control constructs did not match previously reported data is problematic. In the published experiment, the mRNA levels seen with the ABD construct (Construct 8) were approximately 8 times higher than the mRNA levels seen with the LUC construct (Construct 7). According to the present results, Construct 8 mRNA levels were only 50% higher than Construct 7 levels. The experimental setups were nearly identical, although different preparations of DNA were used, different reagent kits were used, and a different real-time detection system was used. Another difference between the present experiment and the published work is that 50 μL reactions were used in the original experiment, while 20 μL reactions were used in the present study. This total reaction volume difference should not affect the measured relative mRNA content. The real-time PCR experiment should be repeated because of this discrepancy.
Figure 10

Real Time PCR Reveals Variable mRNA Levels for SP-B 5’ UTR Constructs

One-step Quantitative real-time RT-PCR was performed with total RNA from transient transfections of NCI-H441 cells. The RNA was extracted 30 hours after transfection. A probe targeting part of the firefly luciferase gene was used to detect the experimental transcripts, and the transfected Renilla luciferase transcript level was used as an endogenous standard. A standard curve with known concentrations of DNA was generated to determine the mRNA levels after adjusting for endogenous control (Renilla luciferase) RNA content. All samples and standards were cycled and read on the same 384-well plate. RNA samples from 6 transfections were cycled in triplicate. The RNA levels were then normalized with Construct 7 (LUC) equal to one. Plus and minus signs under the graph show which sequences were included in the constructs.
e) Translational Index

The translational index, a measure of translation efficiency, was calculated for each construct by dividing the relative expression (determined by the luciferase assay and normalized to LUC) by the relative mRNA content (determined by real-time RT-PCR normalized to LUC). Translational indexes are shown in Figure 11. The translation efficiency of the complete SP-A 5’ UTR construct (Construct 8) was 27 times higher than the LUC plasmid, highlighting the importance of the 5’ UTR in translation initiation. In fact, translational indexes of all plasmids were higher than the translational index of the LUC plasmid, which was expected because the LUC plasmid does not have any 5’ UTR sequence. Among the constructs containing the guest SP-B 5’ UTR, the construct containing SP-A 5’ UTR exon B (Construct 1) had a translation index 10 times higher than the construct lacking this 30-nucleotide sequence (Construct 2), and a 4-fold higher efficiency than the random sequence control construct (Construct 3). Among the constructs containing the 43-nucleotide SP-D 5’ UTR, the one containing SP-A 5’ UTR exon B (Construct 4) again had the highest translation efficiency. However, the difference was not as large as that seen in the SP-B 5’ UTR constructs.

The calculation of a translational index is dependent on both the luciferase assay data and the real-time PCR data. An inconsistent result was seen with the controls in the real-time PCR experiment, and any problems in this dataset would also be observed in the translational index values. For this reason, emphasis will not be placed on the real-time PCR data or the translational index values. The real-time PCR experiment should be repeated and a new set of translational indexes should be calculated and analyzed.
Figure 11

Constructs Containing Exon B Exhibit Higher Translation Efficiency

Translation index is an indirect measure of translation efficiency. Translation indexes were calculated by first normalizing all luciferase assay activity results to Construct 7 (LUC). These normalized luciferase activities were then divided by the mRNA levels from Figure 10, which were also normalized to Construct 7. The result can be expressed as translation activity per mRNA transcript.
3. Deletion Mutation Study

According to the luciferase assay and real-time PCR experiments, SP-A 5’ UTR exon B acts as a translational enhancer. In order to determine which sequence segment of this exon is responsible for the increase in activity, a series of deletion mutations was made from the ABD splice-variant template (Construct 8) by site-directed mutagenesis. In each mutation construct, 5 basepairs were deleted. Deletion Mutation Construct 1 contained a deletion of 5 nucleotides immediately upstream of exon B in the wild type cDNA sequence (the deletion occurred at the 3’ end within exon A). Deletion Mutation Construct 8 contained a deletion of the 5 nucleotides immediately downstream of exon B, within exon D. The 6 remaining deletion mutations occurred within the B exon and were non-overlapping. The nucleotide positions of all deletion mutations are shown in Figure 12. In this Figure, exon B sequence is represented by an underline, and the deleted sequence is shown in lowercase letters. The deletion mutation constructs were co-transfected with the pRL-SV40 (Renilla luciferase) vector into NCI-H441 cells, and the luciferase activity was measured after 36 hours. Constructs 7 and 8 (LUC and positive controls) were included. The results of the luciferase assay are shown in Figure 13. The luciferase activity results were normalized to the ABD splice-variant construct (Construct 8).

The luciferase assay results with the deletion mutations reveal that no 5-nucleotide deletion mutation construct achieves a level of activity as high as the complete 5’ UTR construct (Construct 8). Consistent with the results seen in the guest 5’ UTR and whole 5’ UTR studies, the LUC control plasmid displayed activity of approximately 3.6% ± 0.5% of Construct 8 activity. While no single deletion showed a greatly reduced activity, Deletion Mutation Construct 1 displayed the lowest activity with a mean activity of approximately 34% of Construct 8 activity. Although the deletion in this construct occurred within exon A, it is plausible that this sequence at the A-B splice junction is somehow involved in proper secondary RNA structure and/or translation initiation functions. Deletion Mutation Constructs 3 (0.45 ± 0.03), 4 (0.42 ± 0.05), and 5 (0.43 ± 0.03) also displayed reduced activity. Because all 3 of these constructs contain adjacent deletions within exon B, it is likely that nucleotides 5 to 20 are most responsible for SP-A 5’ UTR exon B’s translational enhancement capacity. Contrariwise, Deletion Mutants 2 (0.75 ± 0.04), 6 (0.73 ± 0.09), 7 (0.82 ± 0.10), and 8 (0.79 ± 0.08) showed the smallest decrease in activity compared to Construct 8, which contains the entire ABD variant 5’ UTR of SP-A.

The results of the deletion mutation study indicate that the splice junction at the 3’ boundary within exon A may play a key role in translation initiation or RNA stability. There are
effectively 2 classes of deletion constructs: high activity constructs with mean activities between 0.73 and 0.82 (Deletion Mutation Constructs 2, 6, 7, and 8), and low activity constructs with mean activities between 0.34 and 0.45 (1, 3, 4, and 5). The observation that 4 deletion mutation constructs display activities of approximately 40% of the wild type 5’ UTR construct is particularly striking because, even though these deletion constructs had the lowest activities, deleting the entire exon B sequence yielded an activity of 67% in the whole 5’ UTR study. Furthermore, the finding that no deletion mutation construct reached a level of activity as high as Construct 8 activity may indicate one of two things: a) that the 5’ UTR size is an important determinant of translatability or stability, as a loss of 5 nucleotides causes a non-proportional decrease in activity, or 2) that the ABD 5’ UTR secondary structure has been optimized through selective evolution and disrupting the refined secondary structure with any of the 5-basepair deletions causes a reduction in activity. Together, these data may provide support for a secondary structure-based mechanism of translation initiation at the exon B region.
The 8 deletion mutation constructs were generated using site-directed mutagenesis of the ABD splice-variant construct (Construct 8). Deleted sequences are shown in lowercase letters and the exon B region is underlined. The first deletion mutation is located at the 3’ end within exon A, while the eighth deletion lacks the first 5 nucleotides of exon D. Deletions are non-overlapping. Deletion mutations were labeled in sequential 5’ to 3’ order.
Figure 13

Exon A Flanking Sequence and Specific Exon B Sequence May be Essential to Exon B Function

*See Figure legend on next page.*
The 8 deletion mutants, along with the LUC (negative control) and the complete SP-A 5’ UTR (positive control) constructs were each transfected into NCI-H441 cells with the pRL-SV40 plasmid. Firefly and Renilla luciferase activities were measured 36 hours later. Each experiment included 2 transfections of each construct, and 3 independent experiments were performed. The firefly and Renilla luciferase activities were measured in triplicate. The ratio firefly/Renilla luciferase was then calculated. The activity of the positive control (SP-A 5’ UTR) was set to 1 and the activity of each construct was normalized to the positive control (deletion mutant or LUC activity divided by positive control activity). Error bars indicate the standard errors of the ratios. The Table under the graph indicates significant differences (p<0.05) in activity between constructs (asterisks indicate significant comparisons and shaded boxes represent non-significant comparisons). For example, comparison of LUC and Deletion Construct 2 data indicates a significant (asterisk) difference, whereas comparison of Deletion Construct 8 and 2 data indicates that no significant difference was observed between them (shaded box). Deletion Construct 1 activity is significantly different (p<0.05) than all other constructs except Deletion Constructs 3, 4, and 5. Deletion Mutation Constructs 3, 4, and 5 were not significantly different when compared to each other, but the comparison of these constructs with all others was statistically significant (p<0.05). Deletion Mutation Constructs 2, 6, 7, and 8 were not significantly different when compared to each other, but were significantly different than all other deletion constructs (p<0.05). All constructs except Deletion Mutation Constructs 7 and 8 were significantly different (p<0.05) than the positive control (SP-A 5’ UTR). LUC control construct activity was significantly lower (p<0.001) than all other constructs, as expected. The negative (LUC) and positive (SP-A 5’ UTR) controls have been described previously in Figures 6, 8, and 10 as Constructs 7 and 8, respectively.
V. Discussion

A reporter gene study of several common SP-A 5’ UTR splice-variants revealed differences in activity and translation efficiency among the variants, and indicated that exon B may be important for increasing activity and/or translation efficiency. To investigate this hypothesis, exon B function was tested in the context of the whole ABD 5’ UTR splice-variant by knocking out the exon B sequence and replacing it with a random sequence. Exon B was then cloned upstream of 2 independent guest 5’ UTRs and compared with a 5’ UTR that lacks exon B and a 5’ UTR containing a random sequence in place of exon B. A series of 5-nucleotide deletion mutations were made within exon B to determine which sequence region along its length is more important for its function. The 5’ UTR sequences were tested by completing transient transfections of NCI-H441 cells and measuring activity with the luciferase assay. In general, constructs containing exon B in the 5’ UTR exhibited higher luciferase activity than constructs lacking this sequence or containing a random sequence of the size of exon B. A 15-nucleotide region of exon B was found to reduce activity when deletions occurred there. Together, the findings indicate that exon B is a translational enhancer and about half of this exon sequence appears to be most important for its enhancer function. Based on the findings, SP-A 5’ UTR exon B represents one of the first translational enhancers discovered in a mammalian system.

Gene regulation is an essential, dynamic, and often complex process in virtually all life forms. Accurate regulation of gene function is achieved through a multitude of conserved and unique mechanisms, many of which likely remain undiscovered. In particular, the regulation of human SP-A is a complex biological process (Karinch, 1995 and Wang, 2003, 2005). SP-A is expressed at different levels in various tissues, as well as at different times during development (McCormick, 1994; Lin, 2001; Ballard, 2003 and MacNeill, 2004). The 2 human SP-A genes, SP-A1 and SP-A2, are quite similar at the sequence level, but both functional and structural differences between the 2 gene products have been described (Karinch, 1995; Garcia-Verdugo, 2002; Mikerov, 2005, 2007, 2007 and Wang, 2000, 2004, 2007). In addition, the variants of both genes may account for differences in SP-A expression among individuals (Floros, 1996).

The focus of the present study was on the 5’ untranslated region of the SP-A gene, and within the context of cap-dependent translation. Although evidence supporting a model of cap-independent translation has recently been found, it is thought that the cap-dependent mechanism predominates in the case of SP-A (unpublished data). SP-A1 and SP-A2 5’ UTRs differ in their mRNA stability and translation efficiency potentials (Wang, 2005), with the SP-A2 5’ UTR
exhibiting higher translation efficiency than the SP-A1 5’ UTR, and therefore the focus of the present study was on the SP-A2 sequence or the ABD splice-variant. Between the SP-A1 and SP-A2 genes, no sequence differences are observed within the 30-bases of the B exon, except for a single nucleotide polymorphism at nucleotide 30 (the last nucleotide), which is G (SP-A2) or A (SP-A1) (White, 1985 and Karinch, 1995). This single basepair difference affects splicing. In the present study only a G was used in this position, which is present in the ABD splice-variant of SP-A2.

The data signify that exon B functions in a sequence-dependent manner to enhance translation. For example, in both the whole 5’ UTR experiment and the guest 5’ UTR experiment the presence of exon B enhanced activity. In the whole 5’ UTR study, the construct containing the whole ABD 5’ UTR had an activity higher than the construct with exon B deleted as well as the construct with the random sequence in place of exon B. In the guest 5’ UTR studies, the constructs containing the B exon sequence had activities that were significantly higher than the constructs containing only the guest 5’ UTRs as well as the constructs containing the random sequence with the guest 5’ UTRs. Based on the guest 5’ UTR data, exon B sequence appears to enhance translation by a factor of 2. The results of the deletion mutation study revealed that deleting parts of exon B lead to reduced activity, providing further proof of exon B’s function as a translational enhancer. The results of all 3 studies consistently indicate that exon B is an enhancer of translation. Although exon B’s enhancement capacity does not appear to reach the level of enhancement seen with the previously reported translational enhancers in viruses, bacteria, and mammals, exon B is relatively small in comparison to the other 5’ UTR sequences (Stein, 1998; Turner, 1999; Chizhikov, 2000; Rubtsova, 2003 and Komarova, 2005).

How this exon performs its role as a translational enhancer is not entirely clear, but several features and characteristics of 5’ UTRs are known to impact post-transcriptional modifications, RNA decay rates, and translation efficiency. The 5’ UTR has important roles in posttranscriptional RNA processing, localization, and stability, as well as translation initiation and ribosomal recycling (Pesole, 2001). Several factors associated with the 5’ UTR are known to affect both RNA decay rates and translation efficiency, and these include the 5’ cap, the primary sequence of the 5’ UTR, trans-acting translation factors, secondary mRNA structure, the presence of alternative translation start sites (AUG codons), and length of the 5’ UTR. Some of these factors, and how they relate to exon B, are discussed in greater detail below. Specifically, four
factors are relevant to the present study: secondary mRNA structure, the primary sequence, 5’ UTR length, and trans-acting translation factors.

**Secondary mRNA Structure**

The secondary structure of the 5’ region of the RNA is known to be important for translation. Iron response elements, commonly occurring RNA stem-loop structures within untranslated regions, are known to regulate translation of genes involved in iron metabolism in conjunction with several iron response proteins (Coulson, 1993). IRES elements are also known to form unique secondary structures that are recognized by the cap-independent initiation machinery. It is estimated that more than 6% of human 5’ UTRs contain an IRES element (Pesole, 2001). Secondary RNA structures in cap-dependent translation initiation (the type of translation initiation that was examined in the present study) may function in altering translation efficiency by forming structures that are too stable to be unwound by initiation factors eIF4A/eIF4B (Kozak, 1989). In addition to providing binding domains for translation factors, stem-loop structures may slow the rate of ribosome scanning (Kozak, 1989). In general, a more stable mRNA secondary structure leads to reduced translation efficiency (Kozak, 1986).

Formation of secondary RNA structure is not a random process. The secondary structure is dependent on the primary sequence and can be predicted using computational tools (Zuker, 2000). One commonly used online program for visualizing likely RNA secondary structure is mFold (Zuker, 2003). This program was used to predict secondary structures of all of the 5’ UTRs that were used in this study. The predicted secondary structures of all 5’ UTRs except the LUC control construct, which did not have a 5’ UTR, are shown in the Appendix, along with their relative activities (specific construct luciferase assay activity divided by the activity of the LUC negative control construct). When using mFold to predict secondary structure, 5’ UTR size is a major determinant of the output RNA structure. In the present analysis, only the 5’ UTR sequence of each construct was used to predict RNA structure despite the fact that 5’ UTRs used in these experiments differed substantially in size (14 to 100 nucleotides). For example, the 5’ UTR of Construct 1 is 44 nucleotides long while the 5’ UTR of Construct 2 consists of only 14 nucleotides. For this reason, the structures shown in the Appendix may not accurately reflect structure in the context of the whole mRNA sequence (with the firefly luciferase gene). Complicating matters more is the fact that RNA structure is not static, but may change between several conformations. While the most energetically stable structures are shown in the Appendix, other structures are possible for each 5’ UTR. Furthermore, while many computational tools exist...
for predicting RNA secondary structure, only one algorithm was used to predict the secondary structures shown in the Appendix.

An analysis of the predicted mRNA structures reveals that the predicted structures of all of the deletion mutation 5’ UTRs and the complete SP-A 5’ UTR contain a terminal loop (marked with black arrows in Section B of the Appendix) in the exon B region of the structure. The results of the deletion mutation study revealed 2 classes of activities: 4 constructs that displayed high activity (Deletion Constructs 2, 6, 7, and 8 had activities near 80% of wild type activity), and 4 constructs that displayed low activity (Deletion Constructs 1, 3, 4, and 5 had activities near 40% of wild type activity). The terminal loops (marked with black arrows) within the structures of Deletion Constructs 3, 4, and 5, which are part of the low activity group, contain 10 nucleotides (see secondary structures of deletion mutants in Appendix Section B). However, the terminal loops within the structures of Deletion Constructs 6, 7, and 8, which make up most of the high activity group, contain 12 nucleotides. Deletion Constructs 1 and 2 contain terminal double loop structures, similar to the whole 5’ UTR construct. Deletion Construct 2 differs from Deletion Construct 1 in that two contains a third loop structure (marked with a red arrow) 6 nucleotides from the terminal loops, and this could explain the observed increase in activity seen with Deletion Construct 2. While it is not clear whether the correlation between terminal loop size and activity is meaningful, these findings indicate that activity may be somehow enhanced by larger terminal loop structures. It is more likely that the correlation between terminal loop size and activity is just a coincidence, as Deletion Mutation Constructs 1 and 2 would be expected to have activities near the activity level of the whole SP-A 5’ UTR construct if the terminal loop structure was important. Also, there is no previously reported evidence for 10 and 12-ribonucleotide terminal loops having different functional roles in translation initiation. Although the formation of a 5’ stem-loop structure is known to stabilize prokaryotic transcripts in a structure-specific manner, no such observation has been found in mammalian systems (Emory, 1992).

In summary, among the experimental constructs, no common structural features that correlate with activity were observed with a visual inspection of the mFold output. While this may indicate that secondary structure is not important for exon B’s enhancement capabilities, it is not possible to draw any conclusions from this visual analysis because of the caveats and limitations of the mFold predictions.

The mFold program provides the user with both a schematic of the predicted mRNA structure and its stability (Zuker, 2003). The mFold program calculates stability as a free energy
value (-dG) from a set of thermodynamic parameters. Computed structures with the lowest free energy values are most likely to reflect the actual structure under physiological conditions. It should be noted that the mFold free energy value is a measure of thermostability but is not a melting temperature ($T_m$). To determine whether a correlation between 5’ UTR structure stability and activity, as measured by the luciferase assay, existed with the constructs used in the present studies, 2 stability vs. activity graphs were made (see Figures 14 and 15). Figure 14 consisted of data points for the constructs used in the whole 5’ UTR and guest 5’ UTR studies, except for the LUC control construct, which did not contain a 5’ UTR. A weak positive correlation ($R^2 = 0.2872$) was observed. Figure 15 consisted of data points for the 8 deletion mutation constructs. A very weak negative correlation ($R^2 = 0.0429$) was observed. Because the correlation between stability and activity is extremely weak, it is unlikely that RNA stability of the 5’ UTR impacts SP-A2 translation efficiency or the translation efficiency of the synthetic 5’ UTRs used in this study. Together, these findings indicate that RNA secondary structure stability may not be a primary determinant of 5’ UTR-driven activity. Alternatively, more sophisticated analyses need to be performed to fully assess the role of SP-A 5’ UTR secondary structure on translation.

**Figure 14**

**A Weak Positive Correlation Between RNA Stability and Activity is Seen with the Experimental Constructs**

![Graph showing correlation between RNA stability and activity](image)

Relative activities of Constructs 1 through 10 (excluding Construct 7) from the present experiments (y-axis) and the minimum mFold-computed stability value (-dG) (x-axis) were plotted. A weak positive correlation between these parameters was found ($R^2=0.2872$).
The Primary Sequence

Functionality of mRNAs is defined by both the primary sequence and secondary structure. The lack of correlation with activity that was seen in the examined secondary structures suggests a larger functional role for the primary sequence of exon B. The sequence within the 5’ UTR is important not only for determining the secondary structure of the 5’ UTR, but could determine the efficiency with which the pre-translation complex scans the leader sequence before translation begins at the AUG codon. According to the scanning model of translation initiation, the 40S ribosomal subunit and several translation initiation factors bind to the mRNA near the cap and scan in a 5’ to 3’ orientation for an accessible start codon. Primary sequences that do not allow for efficient and stable scanning may reduce the translation efficiency of downstream coding sequence (Kozak, 1978).

The sequence within the 5’ UTR may also determine the binding efficiencies of trans-acting elements, such as general and specific translation initiation factors. At the level of
transcription, most modulators of transcription bind to enhancer and promoter sequences based on the primary structure, or sequence, of the DNA. While some translation factors selectively bind to mRNA molecules based on unique RNA secondary structures, many factors bind in a purely sequence-specific context (Pesole, 2001).

In the guest 5’ UTR studies, it was found that exon B enhances translation in a sequence-specific manner. The constructs containing the random sequence instead of exon B failed to significantly boost activity to the level of the constructs containing the specific exon B sequence even though the sizes of their 5’ UTRs were similar. In the whole 5’ UTR study, it was again observed that the B exon enhances translation in a sequence-specific manner. The activity of the construct with the complete SP-A 5’ UTR was approximately 4 times higher than activity of the construct with the random sequence in place of exon B and moderately (47%) higher than the activity of the construct with exon B deleted. Thus, the shorter AD 5’ UTR construct had an intermediate level of activity. Although this finding indicates that 5’ UTR length is likely not an important factor in SP-A 5’ UTR activity (or at least not a major factor), it does not differentiate between a primary sequence and secondary structure-based mechanism of enhancement.

5’ UTR Length

The length of the 5’ UTR may affect translatability or mRNA decay. It is thought that longer 5’ UTR sequences typically cause increased translation efficiency (Senanayake, 1999 and Wang, 2005). This may not be surprising, as longer 5’ UTR sequences allow for more sequence for translation factor and ribosome binding. Longer 5’ UTR sequences are also more likely to contain IRES elements and other structural features that could influence translation in a positive manner. To determine whether a correlation between 5’ UTR length and activity exists for the constructs under study, the 5’ UTR length and activity were plotted as shown in Figure 16. Only 1 deletion mutation construct point was included (all deletion mutation constructs have UTRs of the same length). The negative control LUC construct, which did not contain a 5’ UTR, was excluded. A weak positive correlation between 5’ UTR length and activity is observed ($R^2 = 0.2953$). It is important to note that this analysis may be influenced more by the differences in length and activity between the guest SP-B and SP-D 5’ UTRs. With this in mind, it seems that there is no correlation between 5’ UTR length and activity in the present study.
Relative activities of Constructs 1 through 10 (excluding Construct 7) and the average activity of all deletion mutation constructs from the present experiments (y-axis) are plotted with the 5’ UTR length of each construct (x-axis). A weak positive correlation between these parameters was found ($R^2=0.2953$).

One can speculate on how the SP-A 5’ UTR exon B carries out its enhancer function based on the results of the 3 studies. Because little correlation is seen between mRNA stability and activity, and between 5’ UTR length and activity, it is likely that the primary sequence within exon B plays a key role in enhancing posttranscriptional expression. The exon B sequence’s ability to enhance translation appears to be independent of the downstream 5’ UTR sequence, whether this be the SP-B 5’ UTR, the SP-D 5’ UTR, or the D exon of the SP-A 5’ UTR. However, the degree of the enhancement may depend on downstream or surrounding sequence. In the guest 5’ UTR studies, the degree of enhancement was much higher when the SP-B 5’ UTR was used. This may or may not be due to the difference in length between the 2 guest 5’ UTRs (the SP-B 5’ UTR is 14 nucleotides long while the SP-D 5’ UTR is 43 nucleotides in length). Adding sequence to the short SP-B 5’ UTR may stabilize the transcript or allow for more efficient ribosomal loading during translation initiation. Support is provided for this concept by the fact that exon B’s enhancement over the random sequence constructs (which had longer 5’ UTRs) was not dependent on which guest 5’ UTR was used. Thus, 5’ UTR length may be an important
determinant of mRNA decay or translation efficiency up to a certain point, as indicated by the differences in the enhancement between the 14 and the 44 basepair guest 5’ UTRs.

Trans-acting Translation Factors

The availability and correct localization of translation initiation factors is essential for efficient initiation. Dozens of translation factors are known to function in translation initiation by interacting with the 5’ UTR, other initiation factors, and ribosomal subunits. Some eukaryotic initiation factors, eIF4A and eIF4B, are thought to bind the leader sequence on the mRNA strand and unwind any RNA secondary structures (Naranda, 1994). Without these specific and general translation factors, global and/or specific translation may not be possible (Neff, 1999). The cellular dynamics of these trans-acting factors may determine the cellular rate of translation. As an example, factor eIF4GI can be localized in the nucleus or cytosol depending on various cell states (Coldwell, 2004).

It is likely that trans-acting factors play a central role in mediating translation efficiency of the constructs used in the present studies. RNA-binding proteins may bind to 5’ UTRs or other RNA sequences to promote or inhibit translation (Kozak, 1978, 1986, and 1989). RNA-protein interactions may be due to primary RNA sequence or secondary structure (Pesole, 2001). Thus, it is possible that one or more proteins bind to the exon B RNA sequence and either lead to a more stable transcript or upregulate the rate or fidelity of translation. A trans-acting factor model may be validated by the results of the deletion mutation study. In this study, it was observed that the presence of an intact 15-nucleotide region of exon B sequence leads to higher luciferase assay activity. This region represents a hypothetical RNA target for protein binding. Because the NCI-H441 cells used in the experiments were treated the same, subjected to the same stresses and environments, and harvested at identical time points, it is unlikely that availability or localization of translation factors influenced the results.

eIF4A and eIF4B are 2 initiation factor enzymes that are thought to melt the secondary structure of the 5’ UTR during translation initiation (Naranda, 1994), and many transcription and translation factors are known to bind to DNA and RNA based on the primary sequence. It seems likely that exon B contains a recognition sequence, possibly between nucleotides 5 to 20, for a trans-acting translation factor that somehow improves the rate or fidelity of translation initiation. While this mechanism is speculative, future experiments could be devised to test this hypothesis.
One observation that bridges all 3 studies (guest 5’ UTR study, whole 5’ UTR study, and the deletion mutation analysis) is the finding that the activity of the complete SP-A 5’ UTR was higher than every synthetic 5’ UTR used in the experiments. This fact may indicate that, from an evolutionary standpoint, the SP-A2 5’ UTR has undergone natural selection and has been optimized for expression at the level of translation. This finding also highlights the importance of the SP-A genes for survival.

In summary, while understanding the structure and function of the SP-A 5’ untranslated region is not a simple undertaking, the role of this sequence at the DNA and RNA level is becoming clearer. A series of experiments in the present study aimed at elucidating the function of the B exon within this 5’ UTR showed that this exon likely acts as a translational enhancer. When placed upstream of 2 relatively short, independent 5’ UTR sequences, and in the context of the entire SP-A 5’ UTR, this 30-nucleotide exon seemed to enhance the translatability of a downstream coding region (a firefly luciferase gene in the present study). Based on a deletion mutation analysis, it appears that one part of this exon, at nucleotides 5 through 20, may be more important for exon B function in translation than the surrounding sequence. In addition, the 3’ terminus of the A exon may contribute to SP-A translatability. Within exon B, the primary sequence, secondary structure, or a combination of these 2 factors could be important for interactions with translation initiation factor(s), but the basic sequence is likely most important.
VI. Conclusions and Future Directions

Through a series of transient transfection and luciferase assay experiments aimed at studying the posttranscriptional effects of transcripts containing and lacking SP-A 5’ UTR exon B sequence, it was found that exon B functions as a translational enhancer. Exon B was tested in the context of 2 independent guest 5’ UTRs, the SP-B 5’ UTR and the SP-D 5’ UTR. In both instances, the construct containing the exon B sequence enhanced activity by approximately 2-fold. Exon B was found to enhance activity of transcripts in the context of the whole SP-A 5’ UTR.

In the whole 5’ UTR experiment, the 70-nucleotide AD 5’ UTR construct displayed activity 3 times greater than that of the random sequence construct. This, along with other results, indicates that exon B functions in a sequence-specific manner. The sequence specificity with which exon B functions was confirmed by the deletion mutation study, in which it was found that a deletion of any 5 nucleotides in this exon resulted in reduced activity. The results indicate that the 3’ end of exon A may be important. One region within exon B, consisting of nucleotides 5 through 20, may hold the key to exon B’s capacity to enhance translation, and/or provide binding sites for proteins involved in the regulation of translation. Alternatively, this region may determine a vital secondary structure elsewhere in the 5’ UTR.

Translation initiation is the rate-limiting step in translation and is mediated by several proteins known as translation initiation factors. These factors bind to the mRNA and may interact with other proteins to enhance the rate or fidelity of initiation. Some of these general and specific translation factors may bind to the 5’ UTR, while others, such as poly A binding protein (PABP), may bind to the 3’ UTR. Other translation factors interact exclusively with other protein factors. To determine which translation factors bind to the 5’ UTR of SP-A, an electrophoretic mobility shift assay (EMSA) could be done. This experiment would involve generating several short RNA sequences, such as various SP-A 5’ UTR splice-variants, exon B sequence, and control RNA strands, and incubating them in NCI-H441 cytosolic lysate. After running the RNA on a gel, the bound protein(s) could be washed, purified, and identified.

Although previous studies had indicated a potential enhancer role for SP-A 5’ UTR exon B (Wang, 2005), the results of the present study not only provided evidence for this postulate from different sets of experiments but also indicated that exon A sequence may also function to increase translatability. Because no construct displayed activity at a level as high as the ABD
complete 5’ UTR construct, it appears possible that exons A and/or D function in a synergistic way with exon B to increase translatability. It is not clear whether exon A and/or D can act as translational enhancers on their own, but this could be experimentally tested by creating a few new constructs and testing their activity with the luciferase assay. For instance, a construct resembling Constructs 1 and 4, but with exon A sequence in place of exon B, could be easily made. Whole 5’ UTR constructs with exon A and exon D deleted could also be produced and tested.

The deletion mutation analysis identified a 15-nucleotide region that likely accounts for exon B’s enhancing ability. While it is not entirely clear whether the mRNA’s primary sequence or secondary structure in this region accounts for the deletion mutation results, it may be possible to take this analysis one step further by creating smaller deletions, overlapping deletions, or single basepair changes in this region of interest. Basic PCR-based site-directed mutagenesis reactions could be set up to make the constructs, and the new plasmids could be tested with the luciferase assay to determine the effects of the new mutations.

The region of interest within exon B could be confirmed by placing this 15-basepair sequence upstream of the SP-B and/or SP-D 5’ UTRs, in an experiment parallel to the guest 5’ UTR study. By using the luciferase assay with these new conceptual constructs along with the original Constructs 1 through 8, the enhancement capacity of this region of interest could be determined. The results would also reveal how important the remaining exon B sequence is and whether the secondary structure is essential, as exon B’s secondary structure in such a construct would be significantly altered.

To better standardize the methods for experimentally testing translation efficiency, it may be possible to develop a cell-free translation system or utilize a commercially available one. This approach would eliminate any confounding factors associated with the cell culture and transient transfections, and could normalize the amount of RNA in each reaction. This approach may also allow for addition of a known amount of RNA, instead of using a DNA template, into the translation system. A stipulation of this approach is that any specific translation factor that interacts with the SP-A 5’ UTR may not be present in a cell-free system.

Now that the SP-A 5’ UTR and its role in posttranscriptional regulation is beginning to become clearer, the focus of investigation should shift to identifying any trans-acting factors and studying the mechanisms of expression and regulation. The present study represents only a small
fraction of the work that must be done to obtain a detailed picture of SP-A expression in the lungs. From another perspective, the human genome consists of about 3 billion nucleotides, and the focus of the present study was on only 30. SP-A is one of a handful of proteins found in pulmonary surfactant, and one of thousands of genes expressed in lung tissue. Since altered SP-A levels and function are correlated to so many human conditions, however, this gene must be examined in great depth. This work may have identified a “hotspot” for SP-A regulation. As with any scientific finding, the results of the present study have only lead to more questions and hypotheses.
Appendix: RNA Structures

A) Secondary Structures of Experimental Constructs (Whole 5' UTR and Guest 5' UTR Studies)

Construct 1 (Relative Activity = 12.3)

Construct 2 (Relative Activity = 5.5)

Construct 3 (Relative Activity = 6.2)

Construct 4 (Relative Activity = 25.1)
Construct 5 (Relative Activity = 14.1)

Construct 6 (Relative Activity = 11.6)

Construct 8 (Relative Activity = 41.5)

Construct 9 (Relative Activity = 20.9)
Construct 10 (Relative Activity = 7.4)
B) Secondary Structures of SP-A2 Deletion Mutation Constructs

Deletion Mutation 1 (Low Activity)  Deletion Mutation 2 (High Activity)

Deletion Mutation 3 (Low Activity)  Deletion Mutation 4 (Low Activity)
Deletion Mutation 5 (Low Activity)

Deletion Mutation 6 (High Activity)

Deletion Mutation 7 (High Activity)

Deletion Mutation 8 (High Activity)
References


Senanayake, S., Brian, D. Translation from the 5’ Untranslated Region (UTR) of mRNA 1 Is Repressed, but That From the 5’ UTR of mRNA 7 Is Stimulated in Coronavirus-Infected Cells. *J Virol*: 73. October 1999.


