Methods for Isolation and Sequencing of Topoisomerase II gene in *Gossypium hirsutum*

By

R. Madison Green

A Thesis Submitted to the

Faculty of the College of Graduate Studies at

Middle Tennessee State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science in Chemistry

Middle Tennessee State University

December 2015

Thesis Committee:

Dr. Donald Burden, Director

Dr. Paul Kline

Dr. B. G. Ooi

ACKNOWLEDGEMENTS

I cannot convey how much appreciate the help and support of Dr. Burden. Research is supposed to be a daunting task, this is understood when you are entering a science graduate program. Without the continuing encouragement of Dr. Burden, I don't think I could have made it. His guidance and support made it possible for me to overcome what most have not been able to do.

I would also like to thank the entire Dr. Altman lab. The capable scientists in this lab gave me insight to how modern cloning and DNA sequencing is understood today. I am also thankful for this lab allowing me to use their reagents and instruments. Through their guidance, I was able to overcome many obstacles set before me.

Finally, I would like to thank my entire committee. Through their thoughts and questions, I was able to see my problems in a different perspective and accomplish them in a manner I did not consider.

ABSTRACT

Topoisomerase II (topo II) is an enzyme that is essential in cell division. Topo II alters DNA supercoiling, thus affecting every aspect of DNA function. Gossypol is a chemical produced by *Gossypium hirsutum* that is known to inhibit topo II. *G. hirsutum* has two complete genomes (an A and a D), making it difficult to sequence any particular gene.

Through data bank queries and BLAST nucleotide searches, probable partial cDNA sequences of topoisomerase II for *G. hirsutum* for both A and D genome were identified. A common start primer was used for both genomes. The reverse primers were specific for the 3' end of the coding sequence for topo II. A middle primer was design based on topoisomerase II amino acid sequences near the middle of the protein from several green plant.

Gossypim hirsutum seeds were germinated and allowed to sprout. Total RNA was taken from the leaves. A cDNA library was made from the total RNA. Polymerase chain reaction (PCR) was used to amplify the topoisomerase II sequence. Bands at sizes of around 4000 (start primer - reverse primer) and 2000 (middle primer – reverse primer) base pairs were expected. The PCR reactions resulted in products of these sizes, and the PCR products were recovered from the gel. Extracted DNA was put into a cloning vector and then the DNA was sequenced. Cloning and sequencing produced no obvious topo II sequences, possibly because the extracted DNA had very low concentrations.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

LIST OF EQUATIONS

EQUATION PAGE

CHAPTER 1

INTRODUCTION

Deoxyribose nucleic acid (DNA) is the blueprint for life. DNA codes for proteins which sustain all forms of life. Erwin Chargaff observed that DNA has an adenine that will base pair with thymine and guanine with cytosine. In 1953, James Watson and Francis Crick discovered the structure of DNA, a double helix. Watson and Crick found that two chains of nucleic acid wound around each other with the strands antiparallel to each other.[1] They showed that it is comprised of deoxyribose sugar/ phosphate backbone with a base attached to the sugar. The phosphate attaches to the deoxyribose at the 5' carbon; the phosphate then attaches to the 3' carbon of the next sugar and this lineage repeats itself. At the 1' carbon of the deoxyribose a base is attached. Figure 1 shows the structure of deoxyribose sugar with the carbons numbered.

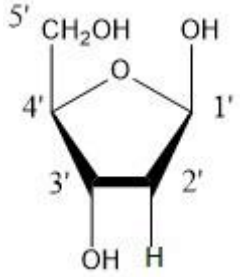


Figure 1. Deoxyribose sugar.

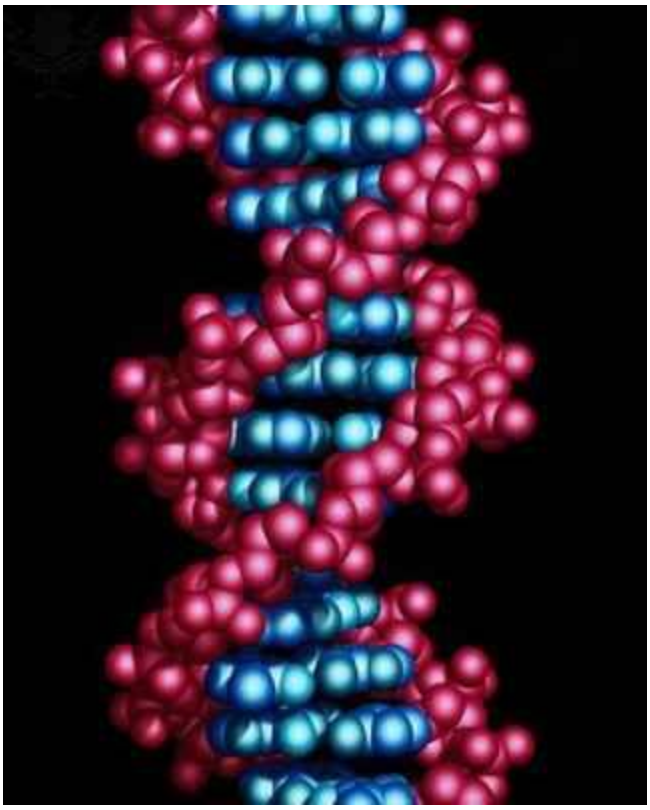
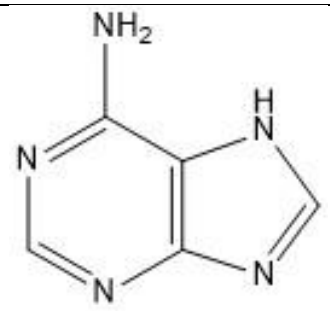
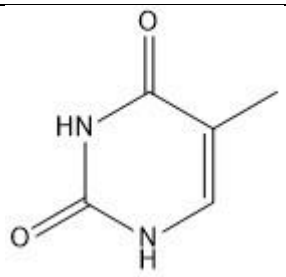
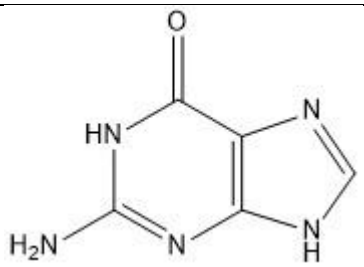
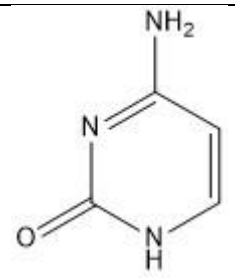


Figure 2. DNA in a double-helix form.

There are four possible bases in DNA which are divided into two categories, purines and pyrimidines. The purines are adenine (A) and guanine (G); the pyrimidines are cytosine (C) and thymine (T). Table 1 shows the structures of the bases. Purines and pyrimidines will base pair through hydrogen bonding. They form a double helix structure with major and minor groves.[1] Purines and pyrimidines will base stack. Base stacking is when cyclic rings, in this case the purines and pyrimidines, stack like one plate on top of another plate.

Table 1. Table of bases.

Purine	Pyrimidine
 **Adenine**	 **Thymine**
 **Guanine**	 **Cytosine**

Adenine will form two hydrogen bonds with thymine and guanine will form three hydrogen bonds with cytosine. Nucleotides are bonded with a deoxyribose sugar at a nitrogen. Nucleotides are bonded at the 5' and 3' carbon of the deoxyribose sugar giving orientation to a sequence of nucleotides. Every three nucleotides in the sequence of a gene, called a codon, represent a specific amino acid. Some amino acids have multiple codons while others have just one.

Amino acids have a core structure consisting of an amine group (Figure 3) and a carboxylate group on the other end (Figure 3).

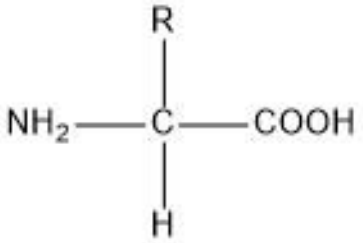


Figure 3. Structure of amino acid. The R group determines which amino acid it is.

As with nucleotides, polymers of amino acids have an orientation in a direction going from N terminus to C terminus. After many amino acids polymerize, a polypeptide is formed. The sequence of amino acids is called the primary structure. Once the primary structure starts to fold on itself through hydrogen and disulfide bonds, ionic and van der waals forces, as well as hydrophobic interactions the protein forms a secondary structure. The complete 3-D shape is the tertiary structure. When two or more polypeptide structures come together they form the quaternary structure. The tertiary and quaternary structures are what allow proteins to have biological activity. Some proteins function as structural components such as proteins found in hair and nails, but most function as enzymes that catalyze biological reactions.[1]

An RNA sequences can be found from amino acid sequences as well. Amino acids can have multiple codons that code for that specific amino acid. By using the National Center for Biotechnology Information Basic Local Alignment Search Tool (BLAST)[2] search engine, a DNA or RNA sequence can possibly be found from the amino acid sequence. By putting the protein sequence of *Arabidopsis thaliana* for topoisomerase II into a BLAST search searching only *Gossypium hirsutum* DNA and RNA sequences, it is possible to find a plausible partial or complete nucleotide sequence for topo II from *G. hirsutum.*

DNA is tightly coiled, supercoiled, in order to pack it into the nucleus of the cell. In order for a cell to replicate itself, it has to make a copy of its DNA to pass on to the daughter cells. There are enzymes that uncoil, relax, DNA to allow for replication. A DNA polymerase goes through the DNA strand and makes copies of it going from 5’ to 3’. There are potential problems when the polymerase makes a copy of the DNA when it is supercoiled. It would be as though two strings were intertwined and then a stick is pushed along between the strings. It necessary to make breaks in the double stranded DNA in order to control supercoiling. There are a family of enzymes that achieve this. Enzymes that cut DNA and alter its supercoiling are called topoisomerases.

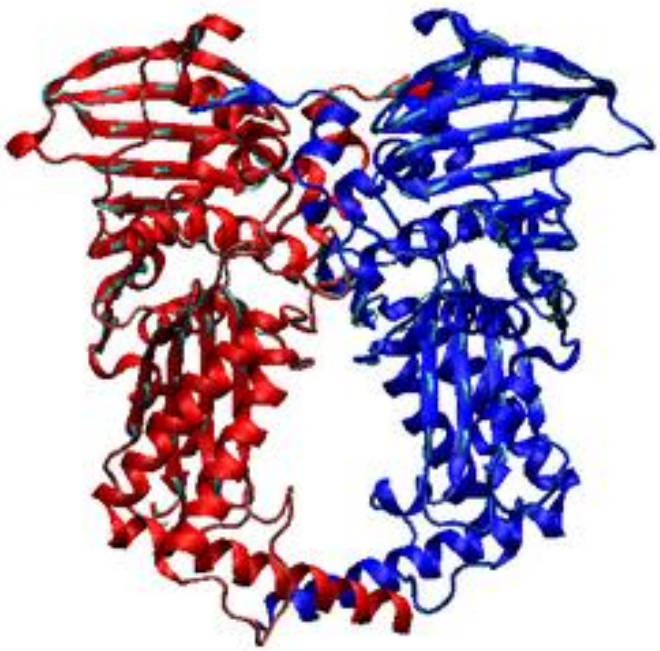


Figure 4. X-ray crystal structure of topoisomerase II.

Topoisomerases are part of the enzyme family that can relax supercoiled DNA. There are two types of topoisomerases. Topoisomerase I (topo I) cuts one strand of the double-

stranded DNA and rotates around the intact strand and then religates (reseals) the broken strand. Topoisomerase II (topo II) makes double-stranded breaks, passes another double-stranded DNA strand through the breaks, and religates the broken DNA. Topo II is a homodimer consisting of about 1500 amino acids for each monomer. This means that the quaternary structure is comprised of two identical tertiary structures. There is a series of lysine amino acids, called the K loop, which allows the topoisomerase II to "sense" binding of DNA. The phosphate backbone of DNA has a negative charge associated with it. The amines at the end of the lysine side-chain have a positive charge, allowing the DNA and enzyme to interact. It is theorized that the enzyme interacts with DNA when two strands cross each other. Once the enzyme has both strands (the G segment and T segment from Figure 5), two adenosine triphosphates (ATP) enter the ATPase site causing a confirmation change in topo II and clamping down onto the two strands. Topo II uses tyrosine residues to cleave the phosphodiester backbone of the G segment. One ATP is hydrolyzed causing a conformational change and forces the T segment through the ligated G segment. The G segment is then religated and another ATP is hydrolyzed, forcing another conformational change and pushing out the G segment and returning topo II to its initial state.[1]

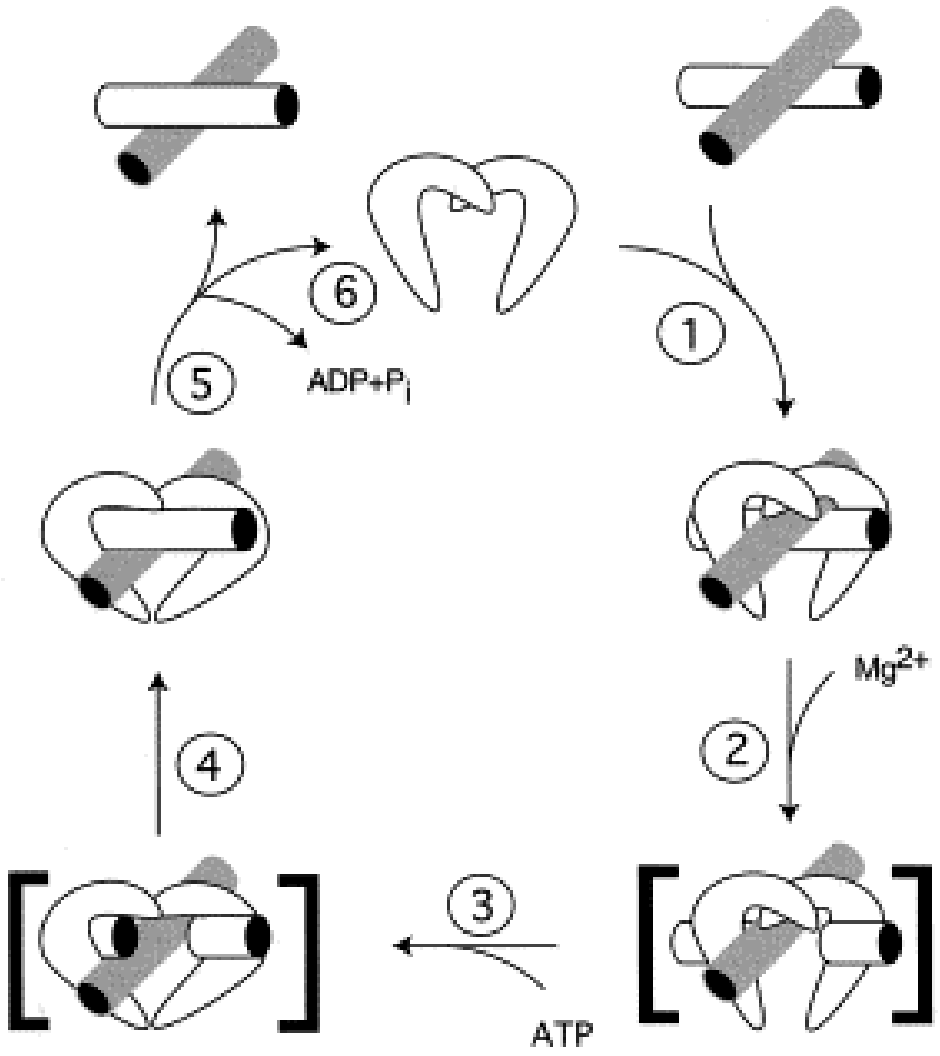


Figure 5. Topoisomerase II mechanism. Mechanism of action of eukaryotic topoisomerase II and drugs targeted to the enzyme. *Biochimica et Biophysica Act.* **1998.** *Vol. 1400,* 16. Figure 2. Reprint with permission from Elsevier.

If the catalytic mechanism of topo II is disrupted, the cell cannot divide properly.[3] This has made topo II an ideal target for anticancer drugs. Cancer cells replicate much faster than normal cells. Since topo II is required for cell division, it tends to be found in higher concentrations in cancer cells. The actions of topo II poisons result in DNA strand breaks. If the cell detects strand breaks, it initiates apoptosis. Topo II poisons will permanently inactivate the

protein. Topo II poisons work by stabilizing the DNA and topo II bond, leaving a permanent double-stranded break. The most common of the topo II poisons is etoposide, which is used as an antitumor drug. Etoposide binds with topo II and the DNA.[4] This prevents the DNA from religating, leaving a double-stranded break. Etoposide stabilizes the step of the mechanism when the G segment is ligated. This prevents a helicase enzyme from separating the strands of DNA. Helicase attempts to push through the enzyme and ultimately breaking the enzyme into its two monomers, leaving DNA with a double stranded break, and causing apoptosis.

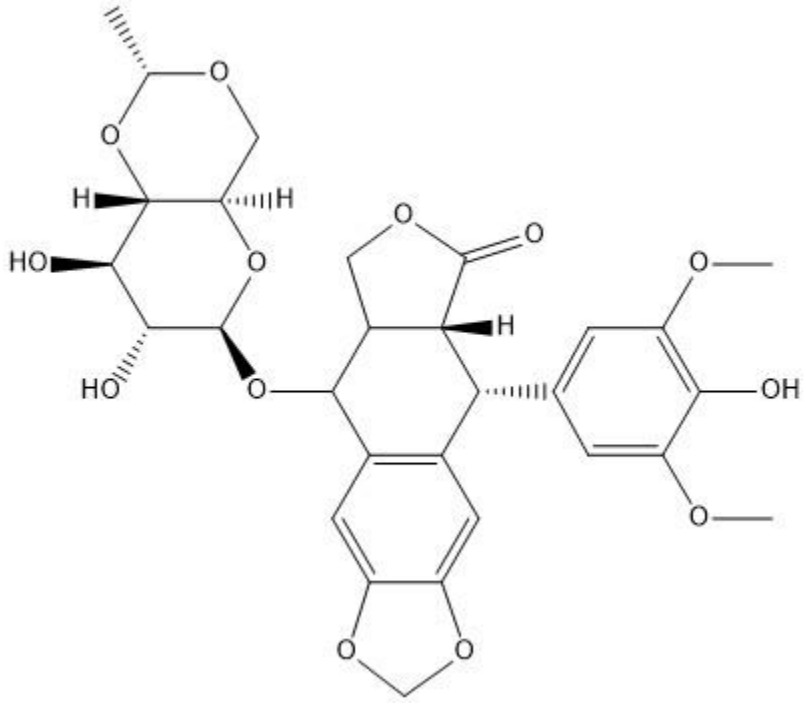


Figure 6. Chemical structure of etoposide.

There are also catalytic inhibitors of topo II. Inhibitors have multiple mechanisms of stopping the catalytic activity. Aclarubicin inhibits topo II by preventing the binding of DNA to topo II.[4] Aclarubicin attacks topo II at the S phase of the cell cycle. The S phase is when a cell relaxes supercoiled DNA when topo II is most abundant. Suramin also inhibits binding of DNA.

The problem with suramin is prolonged exposure can lead to neuropathy and lymphopenia.[4] Novobiocin blocks the ATP binding, stopping topo II from clamping onto DNA.[4] Gossypol is also a known inhibitor of topo II, but little of its mechanism is known.

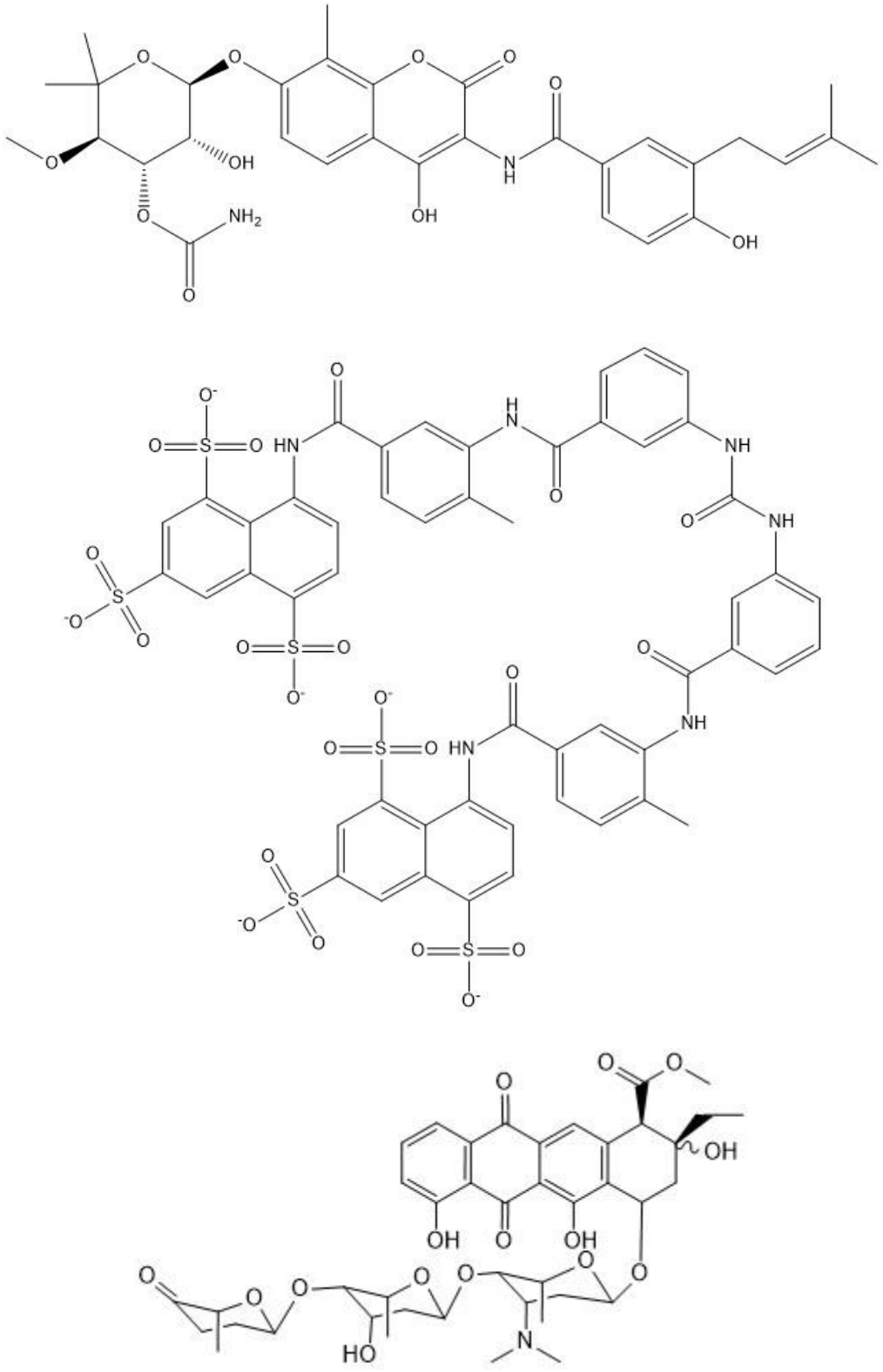


Figure 7. Topoisomerase II inhibitors. From top to bottom: novobiocin, suramin, aclarubicin

Gossypol is a chemical that is produced by cotton, *Gossypium hirsutum.* It is known to inhibit topo II. *Gossypium hirsutum*, upland cotton, has two entire genomes. The paternal genome (D genome) is closely related to *G. raimondii* and the maternal genome (A genome) is closely related to *G. arboreum* and *G. herbaceum*. This means that *G. hirsutum* has two sets of every protein. Many studies have been done into polyploidy genomes to understand how they work and the purpose of having two complete genomes.[5] Recent studies show that in many polyploidy species, as in *G. hirsutum*, one genome accelerates in evolution.[5] This is the D genome for *G. hirsutum*. This can cause one genome to become silenced and ultimately become a psuedogene, a silenced gene.[5]

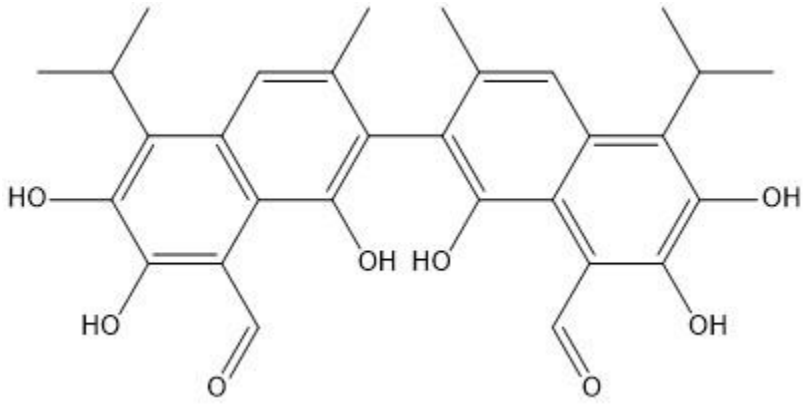


Figure 8. Structure of gossypol.

Gossypol was discovered in the late 19th century. Gossypol is a yellow pigment, so early uses for the compound were as dyes. Early studies quickly found that the compound was unstable in light. In the 1960s, gossypol was found to have antitumor properties.[6] Further studies were done to improve the potency, which led to the discovery that gossypol also caused male infertility. This led to even more studies, but interest soon diminished because gossypol was shown to have irreversible male infertility effects.[6] Gossypol is a polyphenolic compound

with two orientations (+ and -). Its chemical formula is $C_{30}H_{30}O_8$ and it has a molecular weight of 518.5 g/mol. The + enantiomer is thought to be the active form.[6] The two aldehydes appear to be the main contributors to the toxicity of gossypol. Earlier cotton processors would add free lysine amino acids to cotton meal to retain the gossypol in cotton oil. When gossypol reacts with an amine, it creates a Schiff base.[6] The Schiff base mechanism is shown on Figure 9.

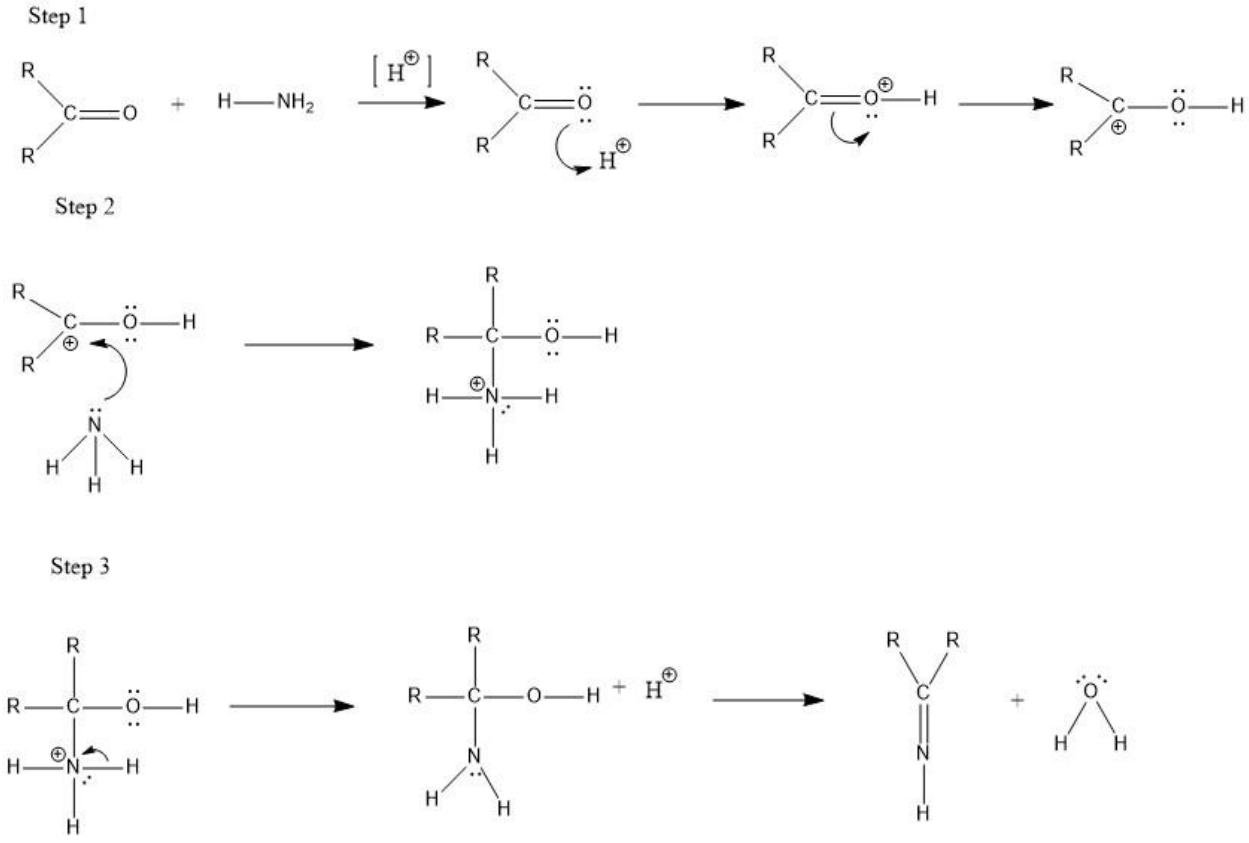


Figure 9. Schiff base mechanism.

A Schiff base is when an amine becomes double bonded to a carbonyl,[7] in this case the aldehyde. As noted previously, topo II binding to DNA involves lysines. Lysines have an amine at the end of its side group as shown in Figure 10.

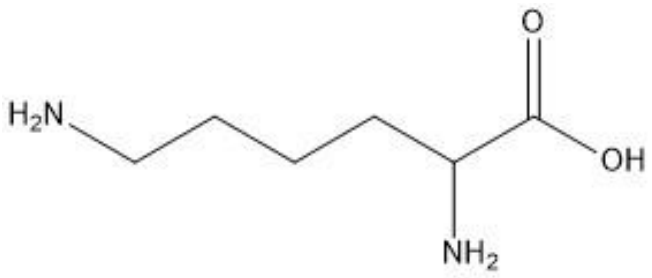


Figure 10. Amino acid lysine.

Binding of gossypol to these lysines may be involved in the mechanism of inhibition.

The topoisomerase II gene from *Gossypium hirsutum* gene has not been sequenced at the start of this project. The purpose of this project is to isolate the topoisomerase II gene and sequence it from *G. hirsutum* species. This information can then be used to compare and contrast cotton topoisomerase II with enzymes from other species. This information will also allow future work on expressing the gene in a recombinant system.

CHAPTER 2

MATERIALS AND METHODS

Instrumentation

Primers for the polymerase chain reaction were developed by using MEGA6, Query Banks, and NCBI BLAST search engines. Polymerase chain reaction (PCR) was performed by using either thermal cycler by Perkin Elmer GeneAmp PCR system 2400 or a Bio-Rad S1000 Thermal Cycler. Agarose gel images were captured using a Gel Doc-It ® 2 310 Imager with an M-26XV manual lens, model P/N 97-0138-04 and Fisher BioTech Electrophoresis Systems, and 312 nm Variable Intensity Transilluminator FBTIV-88. DNA concentration was determined by UV spectrophotometry using a Thermo Scientific Nanodrop Lite spectrophotometer. DNA was sequenced using the Hitachi 3130xl Genetic Analyzer model 628-0040.

Materials and Reagents

The following materials and reagents were used: *Gossypium hirsutum* seeds (generous gift of Dr. Song Cui of the MTSU School of Agribusiness and Agriscience). Table 2 shows the primers used in the PCR reactions.

Table 2. Table of primers used for PCR. All primers were in 1X TE buffer (1 mM EDTA, 10 mM Tris pH 7.9). All primers were stored in -80˚C freezer. Start primer was purchased from Thermo Fisher. All other primers were provided by Eurofins MWG Operon.

Primers		
Primer Name	Primer Sequence (5' → 3')	Concentration (mM)
Start	CCCCTAAGTCACCATGGTGG	1
A_Stop 1	TTAGGTGCCACCGGTCGAGCC	1
A_Stop 2	TTAGGTGCCACCGGTCTTCG	1
D_Stop	TCAATCCTCGTCTTCCTCGAAATCT	1
Bridge +	TCTCAGGATCATGATGGTTCTCAT	1
Bridge -	ATGAGAACCATCATGATCCTGAGA	1

Agarose gel was prepared using 1% (w/v) molecular biology agarose (Bio-Rad, Richmond, CA) in 1X TAE (40 mM Tris at pH 7.9, 1 mM EDTA, 20 mM glacial acetic acid). LB agar plates (10 g trypone, 5 g yeast extract, 5 g NaCl, 1 mL 1 N NaOH, and 15 g agar) with ampicillin (50 µg/mL) was used for selective growth. LB broth (10 g trypone, 5 g yeast extract, 5 g NaCl, and 1 mL 1 N NaOH) with ampicillin (50 µg/mL) were used for transformation stage of cloning. GTE (50 mM glucose, 25 mM Tris-Cl pH 8.0, 10 mM EDTA), potassium acetate pH 4.8, 5 M (29.5 mL glacial acetic acid, KOH pellets to pH 4.8, water to 100 mL) and NaOH/SDS solution (0.2 M NaOH and 1% SDS) were used for the plasmid minipreps. DNA Precipitating Solution A (DNA

Precip A) consisting of 2.1 mL 3 M NaAcetate and 60 mL 95% ethanol was used for DNA precipitation. pMiniT vector was used for cloning purposes (New England Biolabs). Restrictions enzymes AatII and BamHI (New England Biolabs) were used to determine if inserted DNA was present in the vector. Table 3 shows where the restriction enzymes cut DNA sequence.

Table 3. Table of restriction enzymes used.

Restriction Enzymes	
Enzyme	**Sequence and cleavage (')**
AatII	5'…GACGT ' C…3' 3'…C , TGCAG…5'
BamHI	5'…G ' GATCC…3' 3'…CCTAG , G…5'

Germination

Standard protocols were used for germination.[8] *Gossypium hirsutum* seeds were washed with a 1:1 bleach water solution for 15 minutes. The seeds were then rinsed with water for another 15 minutes. Paper towels were wetted and 10 seeds were placed between the paper towels. The seeds were then placed into a plastic container with a lid. The seeds were checked every 3 days to make sure no contamination had occurred. Once the meristem had grown to about 20 cm or leaves sprouted, the meristems and/or leaves were cut to be used for the next procedure.

Total RNA Extraction

Total RNA was extracted from the meristems and/or leaves using Qiagen RNeasy Micro kit for plant tissue with a few minor adjustments to the manufacture's protocol. About 100 mg of tissue was ground up with liquid nitrogen using a mortar and pestle and then transferred to a "Qiashredder" column. According to the protocol, the ground up tissue was to be transferred to a microfuge tube and then lysis buffer added. Instead of transferring from the mortar, the lysis buffer was added directly to the mortar and then transferred to the "Qiashredder" column. RNA samples were quantified using a nanodrop spectrophotometer. The extracted RNA was stored at -80˚C in 10 µL aliquots.

Complete cDNA Synthesis

First strand cDNA was synthesized using New England Biolab's ProtoScript ® II First Strand cDNA Synthesis Kit. The "easy protocol" and "standard protocol" were followed as per the manufacturer's instructions. Template mRNA (1 µg) was used for the cDNA synthesis. Samples were quantified using a nanodrop spectrophotometer. The cDNA was stored at -80˚C in 10 µL aliquots.

PCR

Polymerase chain reaction (PCR) is a process to amplify specific sequences of DNA using a forward and reverse primer.[9] PCR was performed using Thermo Scientific Maxima Hot Start PCR Master Mix (2X). Forward and reverse primers (0.5 µM as well as varying concentrations) were used in the reactions. cDNA template (1 µg) was used for the reactions. All samples were brought to 50µL with nuclease free water. The touchdown PCR method was used to achieve better specificity. Touchdown PCR is when the annealing temperature is raised 5-10 degrees above the melting temperature of the primer and dropped 1 degree until the temperature is 2-5 degrees below the melting temperature.[10] First, the whole sequence was amplified using the start primer and the three stop primers determined by analyzing the A genome of *G. arboreum* and the D genome of *G. raimondii*. Second, the sequence was split into two smaller sequences using a common middle sequence designated the bridge. Bridge (-) is the reverse primer going to the start and bridge (+) is the forward primer going to the end. Tables 4 and 5 show the PCR parameters for the full length sequence and Bridge sequence, respectively.

Table 4. PCR Parameters used with the thermal cycler with full length sequence of topoisomerase II. * indicates for 15 cycles the annealing 1 temperature was dropped by 1 degree. ** indicates for 13 cycles the annealing 1 temperature was dropped by 1 degree. All denaturation 2, annealing 2, and extension 2 were set for 15 cycles.

Parameters for Thermal Cycler for Start and End Primers						
Forward – Reverse Primers	**Start – A1_Stop**		**Start – A2_Stop**		**Start – D_Stop**	
	Temperature (˚C)	**Time**	**Temperature (˚C)**	**Time**	**Temperature (˚C)**	**Time**
Initial denaturation/enzyme activation	**95.0**	**4 min.**	**95.0**	**4 min.**	**95.0**	**4 min.**
Denaturation 1	**95.0***	**30 sec.***	**95.0****	**30 sec.****	**95.0****	**30 sec.****
Annealing 1	**79.5***	**30 sec.***	**72.9****	**30 sec.****	**72.9****	**30 sec.****
Extension 1	**72.0***	**6 min.***	**72.0****	**6 min.****	**72.0****	**6 min.****
Denaturation 2	**95.0**	**30 sec.**	**95.0**	**6 min.**	**95.0**	**6 min.**
Annealing 2	**64.5**	**30 sec.**	**59.5**	**30 sec.**	**59.5**	**30 sec.**
Extension 2	**72.0**	**6 min.**	**72.0**	**6 min.**	**72.0**	**6 min.**
Final Extension	**72.0**	**10 min.**	**72.0**	**10 min.**	**72.0**	**10 min.**
Hold Temps	**4.0**	**∞**	**4.0**	**∞**	**4.0**	**∞**

Table 5. PCR parameters used with bridge primers. All Annealing 1 temperatures were dropped by 1˚C for 12 cycles. Denaturation 2, Annealing 2, and Extension 2 were held for 15 cycles.

Parameters for Thermal Cycler for Start to Bridge to End Primers						
Forward – Reverse Primers	**Start - Bridge (-)**		**Bridge (+) – A2_Stop**		**Bridge (+) – D-Stop**	
	Temperature (˚C)	**Time**	**Temperature (˚C)**	**Time**	**Temperature (˚C)**	**Time**
Initial denaturation/enzyme activation	**95.0**	**4 min.**	**95.0**	**4 min.**	**95.0**	**4 min.**
Denaturation 1	**95.0**	**30 sec.**	**95.0**	**30 sec.**	**95.0**	**30 sec.**
Annealing 1	**71.2**	**30 sec.**	**71.2**	**30 sec.**	**71.2**	**30 sec.**
Extension 1	**72.0**	**6 min.**	**72.0**	**6 min.**	**72.0**	**6 min.**
Denaturation 2	**95.0**	**30 sec.**	**95.0**	**30 sec.**	**95.0**	**30 sec.**
Annealing 2	**59.2**	**30 sec.**	**59.2**	**30 sec.**	**59.2**	**30 sec.**
Extension 2	**72.0**	**6 min.**	**72.0**	**6 min.**	**72.0**	**6 min.**
Final Extension	**72.0**	**10 min.**	**72.0**	**10 min.**	**72.0**	**10 min.**
Hold Temps	**4.0**	**∞**	**4.0**	**∞**	**4.0**	**∞**

DNA Electrophoresis

A 1% (w/v) agarose gel was prepared using BioRad Molecular agarose and a 1X TAE buffer. The gel ran at 70 V for 45 minutes. The gel was stained with ethidium bromide for 20 minutes and destained in 1X TAE buffer for 30 minutes. DNA was visualized using a UV illuminator at a wavelength of 260 nm.[11]

DNA Gel Extraction

The band was cut out and placed into a marked 1.5 mL microfuge tube. The volume of the gel was measured using the markings on the microfuge tube and a 1:1 volume of phenol was added and heated to 70˚C until the agarose gel melted, about 10 minutes. Sample was vortexed and microcentrifuged at max speed for 5 minutes. The top layer was transferred to another 1.5 mL microfuge tube and a 1:1 volume of phenol-chloroform was added. The sample was vortexed 15 seconds and sat for 15 seconds. This was done twice and then the sample was microcentrifuged for 5 minutes. The aqueous layer was transferred to another 1.5 mL microfuge tube and 3 volumes of precipitation solution and placed at -80˚C for 30 minutes. The sample was then thawed on ice and microcentrifuged at 4˚C at max speed for 15 minutes. The supernatant was removed by pipetting and filled with 70% ethanol. Sample was vortexed and microcentrifuged for 5 minutes. All but 50 µL was removed. The rest of the ethanol was evaporated under vacuum at 25˚C. The pellet was resuspended in 10 mM Tris at pH 8.0 and vortexed. The sample was then placed at 4.0 ˚C overnight.[12]

Cloning

Cloning was accomplished using a PCR Cloning Kit (New England Biolabs). For the ligation protocol, a negative control was used. During the transformation protocol, the outgrowth step was at 37˚C for 1.5 hours. All cells were plated on LB agar plates with ampicillin. Cells were incubated at 37.0˚C for 16 hours. The cells were then incubated at 25.0 ˚C for another 8 hours.

Minipreps of Plasmid DNA

One bacteria colony was inoculated in 5 mL of LB broth. The samples were placed on a shaker at 300 rpm for 24 hours at 37.0˚C. Cells (1.5 mL) were microcentrifuged for 20 seconds and the pellet was resuspended in 100 µL of GTE solution. NaOH/SDS (200 µL) solution was added, vortexed, and placed on ice for 5 minutes. Potassium acetate (150 µL) was added, vortexed, and placed on ice for 5 minutes. Solution was microcentrifuged for 3 minutes and 400 µL of supernatant to a new tube. 95-100% ethanol (800 µL) added to solution and incubated at room temperature for 2 minutes. The solution was microcentrifuged for 3 minutes, the pellet was resuspended with 70 % ethanol and microcentrifuged again for 3 minutes. The sample was dried with a speed vacuum at 25˚C. The pellet was resuspended with 1X TE buffer, vortexed and stored at 4˚C overnight.[15] Samples were quantified by nanodrop spectrophotometer.

Restriction Digest

*Bam*HI and *Aat*II restriction enzymes were used to digest the pMiniT plasmid. The manufacture's protocol was followed. The sample was analyzed by electrophoresis on a 1 % (w/V) agarose gel at 70 V for 45 minutes.

DNA Sequencing

DNA sequencing was done using a Hitachi 3130xl Genetic Analyzer model 628-0040. The samples were prepared by adding 1 µL water, 3.5 µL of 2.5x Big Dye Dilution buffer, 1 µL of 4 µM primer (forward and reverse used in separate reactions), 1 µL Big Dye, and 3.5 µL of plasmid template. All samples were put on a thermal cycler with the following parameters: 96˚C for 10 seconds, 50˚C for 5 seconds, and 60˚C for 4 minutes for 39 cycles. The samples were held at 15˚C. The samples were stored at -20˚C overnight and then given to the Biology Department to be analyzed by the DNA sequencer. The files were viewed using the Chromas Lite program. Samples that provided enough sequence were put in a Blast database to see if there were matches for topoisomerase II DNA sequence.

CHAPTER 3

RESULTS AND DISCUSSION

Primer Design

Arabidopsis thaliana topo II was used as the model for the amino acid sequence for *G. hirsutum* topo II. The first 106 amino acid residues were put in a BLAST search using tblastn and using the express sequence tags. Figure 11 show the results.

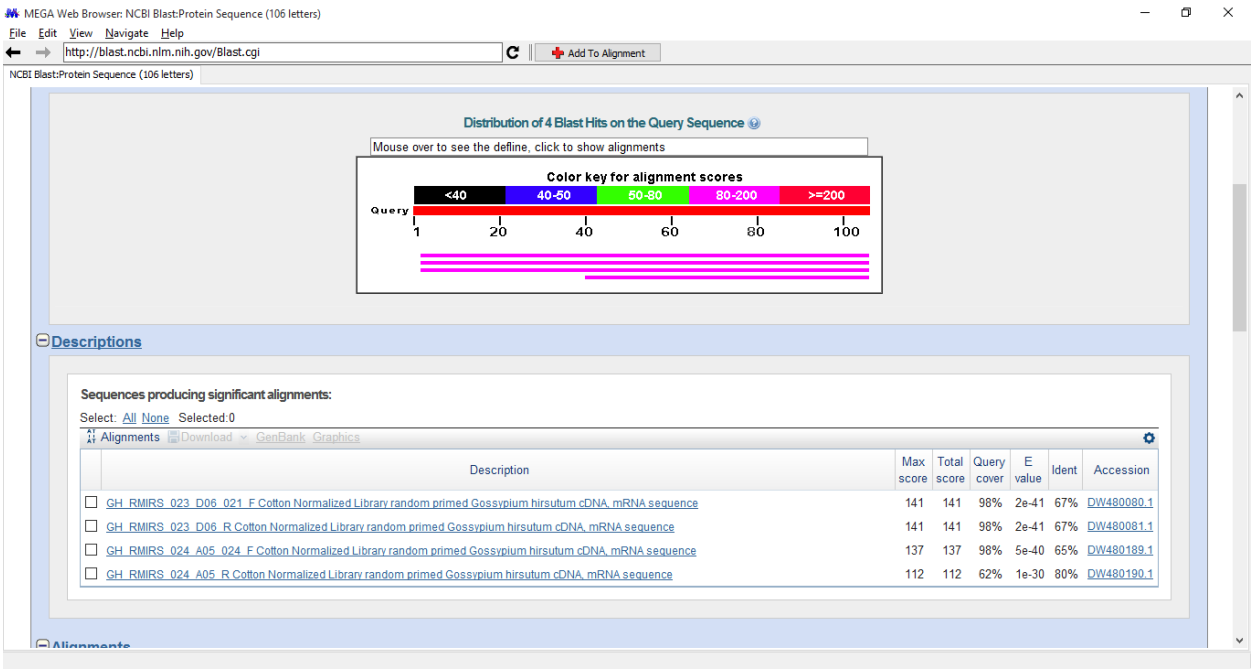


Figure 11. Results from BLAST search with *Arabidopsis thaliana* first 106 amino acid residues from topoisomerase II.

The first hit was used for the D genome and the third hit was used for the A genome. The sequences were entered into MEGA6. The A genome was shifted until both sequences were aligned. Figure 12 shows the two expressed sequence tag (EST) hits aligned.

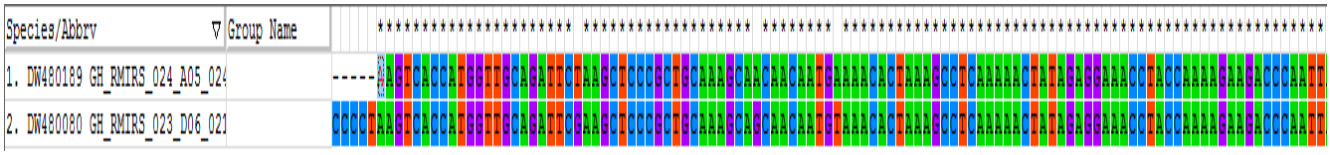


Figure 12. Alignment of the A and D EST hits. Asterisk indicated identical matches.

The first 20 nucleotides of the D EST hit were used as the Start primer. The first nucleotide from each sequence was deleted so the sequence would be an open reading frame. The sequences were then translated into protein sequence. Figure 13 shows the alignment of the two EST hits.

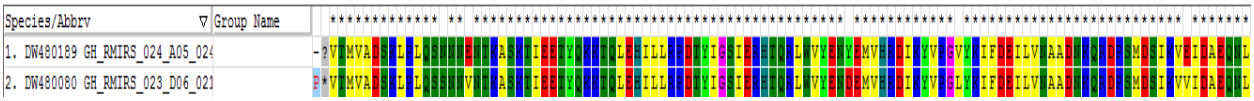


Figure 13. Alignment of the protein sequence of the A and D EST hits. Asterisk indicates identical matches.

The D genome protein sequence was then used to verify that it was in fact a topoisomerase II sequence using the SwissProt database. There were 84 hits and Figure 14 shows the tops hits are topoisomerase II.

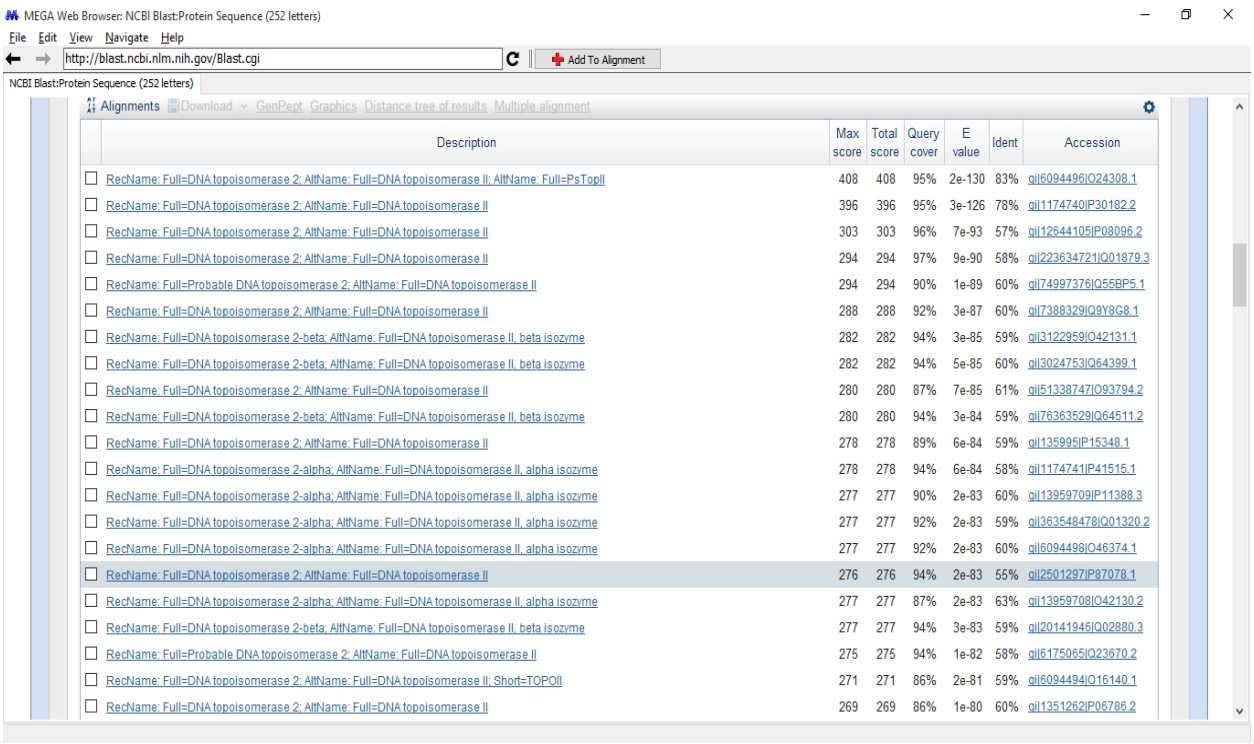


Figure 14. The top hits from a BLAST search using the D EST hits.

For the reverse primer, the last 100 amino acid residues were taken from *Theobroma cacao*. A BLAST search using tblastn using the Transcriptome Shotgun Assembly (TSA) database looking at only *G. hirsutum* was used to find the reverse primer. For the D genome, the first hit's sequence was put into MEGA6 and the sequence was adjusted until an open reading frame was found. Figure 15 shows the amino acid sequence with the D_Stop primer translated.

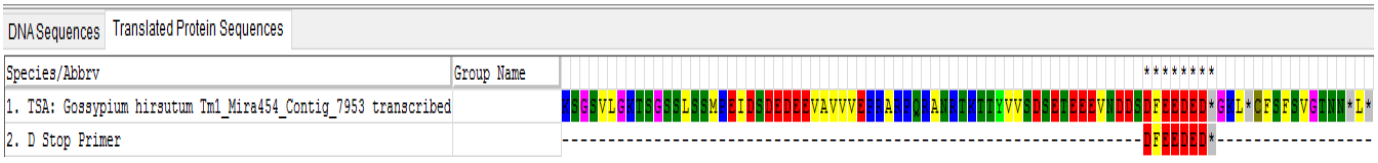


Figure 15. C terminus for D TSA hit with translated D stop primer.

The sequence was put through a BLAST search using the Swiss-Prot database and 42 hits came back, of which were all topoisomerase II proteins. From the stop codon, going 25 bases upstream and taking the reverse compliment is how the D_Stop primer was designed. Figure 16 shows the DNA sequence with the reverse compliment of the D primer.

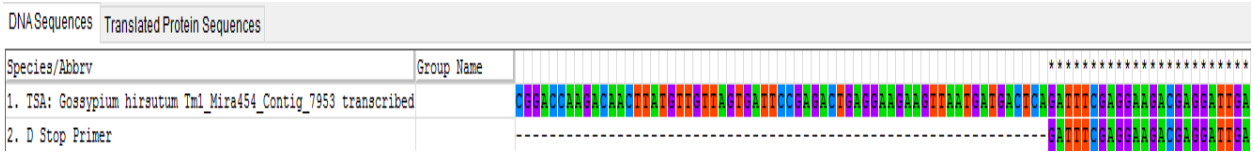


Figure 16. 3' for D TSA hit with reverse compliment D stop primer.

The same procedure was done with the second hit. When there was an open reading frame, the amino acid sequence was put into a BLAST search with the Swiss-Prot database. The search came back with 79 hits, all consisting of either topoisomerase II or a DNA gyrase enzyme. Figure 17 shows the amino acid sequence with the translated A primer. Figure 18 shows the DNA sequence with the reverse compliment of the A stop primer.

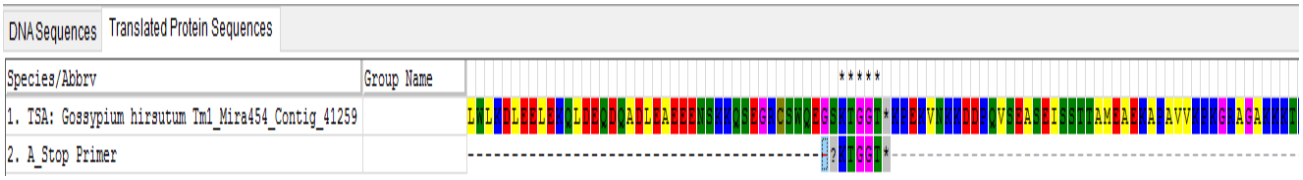


Figure 17. C terminus A TSA hit with translated A stop primer.

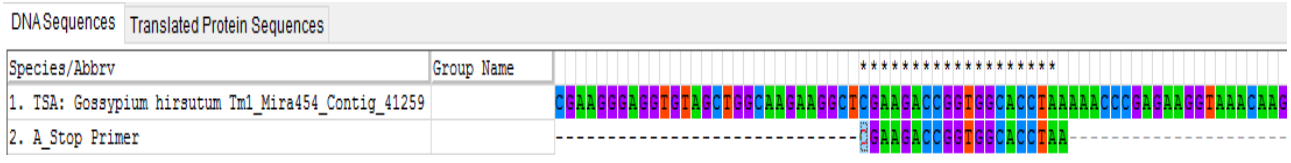


Figure 18. 3' A TSA hit with reverse compliment A stop primer.

For the Bridge primers, several green plant topoisomerase II protein sequences were aligned. Amino acid residues 596 to 602 is a highly conserved sequence among green plants. Figure 19 shows the sequences aligned.

Figure 19. Alignment of green plant topoisomerase II amino acid sequence.

The melting temperatures of the primers provided by the manufacturers were discarded. There was no consistency among the manufacturers. Using Equation 1, the melting temperatures were determined. The manufacturers' were significantly different from Equation 1 results.

$$Tm = 81.5 + [(-28.2) + 0.41(\%GC) - \left(\frac{500}{N}\right)]$$

Equation 1. Melting temperature of PCR primers. Tm is the melting temperature, GC is the number of guanine and cytidine, and N is the total number of nucleotides.

RNA Concentrations from *G. hirsutum* Seeds

A nanodrop spectrophotometer was used to measure the UV absorbance of RNA. Table 6 shows the absorbance of RNA from the total RNA extraction from the *G. hirsutum* seeds. Equation 2 shows the calculation of RNA concentration based on absorbance at 260 nm.

$$Concentration\ of\ RNA\ \left(\frac{\mu g}{mL}\right) = A_{260} \times \frac{dilution\ factor}{0.022}$$

Equation 2. Spectrophotometric determination of amount of RNA.[13]

Each sample is from a different set of germinated seeds. Sample 3 and 4 were used for further experiments. Samples 1 and 2 were stored at -80˚C.

Table 6. Absorbance and concentrations of total RNA extracted from *G. hirsutum*. 1X TE buffer was used as the blank. The absorbance ratio determines how pure the RNA is. An absorbance of approximately 2.00 is considered pure.

UV Absorbance of RNA in Germinated Seeds				
Sample	**Volume (µL)**	**ABS_{260}**	**ABS_{260}/ABS_{280}**	**Concentration (ng/µL)**
1R	2	0.931	1.89	0.03724
2R	2	0.419	1.74	0.01676
3R	2	7.647	2.03	0.30588
4R	2	21.363	2.05	854.5
5R	2	1.077	2.09	43.1

mRNA Extracted from Total RNA Concentrations

After mRNA was extracted from samples 3 and 4, absorbance was measured to determine concentrations of the mRNA. OW2 Buffer supplied by Qiagen RNA kit was used as a blank. Table 7 shows the absorbance and concentrations of the mRNA samples. Concentrations were determined using equation 1. Sample 1m3R was taken from total RNA sample 3R. Samples 2m4R and 3m4R were taken from total RNA sample 4R.

Table 7. Concentrations of mRNA extraction from total RNA.

UV Absorbance and Concentration of mRNA				
Sample	**Volume (µL)**	**ABS_{260}**	**ABS_{260}/ABS_{280}**	**Concentration (ng/µL)**
1m3R	2	0.019	0.52	0.7
2m4R	2	0.007	0.57	0.3
3m4R	2	0.009	1.21	0.4

cDNA Concentrations

Total RNA samples and mRNA samples where then used to make a cDNA library. A nanodrop spectrophotometer was used to quantify cDNA samples. The single stranded setting for DNA was used when measuring the absorbance. Table 8 shows the readings from the spectrophotometer.

Table 8. Absorbance and concentrations of cDNA**.** Sample 1D3m4R came from Total RNA sample 3m4R. Samples 2D2m4R and 3D3m4R came from mRNA samples 2m4R and 3m4R respectively. 3D4R came directly from 4R. A260/A280 ratio of around 1.80 is considered pure.[12]

UV Absorbance and concentrations of cDNA from Total RNA and/or mRNA				
Sample	**Volume (µL)**	**ABS_{260}**	**ABS_{260}/ABS_{280}**	**Concentration (ng/µL)**
1D3m4R	2	33.142	1.71	1093.7
2D2m4R	2	32.657	1.87	1077.7
3D3m4R	2	29.502	1.86	973.6
4D4R	2	27.215	1.86	898.1
5D5R	2	29.008	1.85	957.3

Concentrations of cDNA were determined using Equation 3.

$$Concentration\ of\ DNA\ \left({}^{\mu g}/_{mL}\right) = A_{260} \times \frac{dilution\ factor}{0.02}$$

Equation 3. Spectrophotometric determination of DNA concentration.[12]

PCR Results

PCR results were viewed by running electrophoresis on a 1% agarose gel. Figures 20 through 23 show the results from the reactions. Contamination controls were used with all reactions.

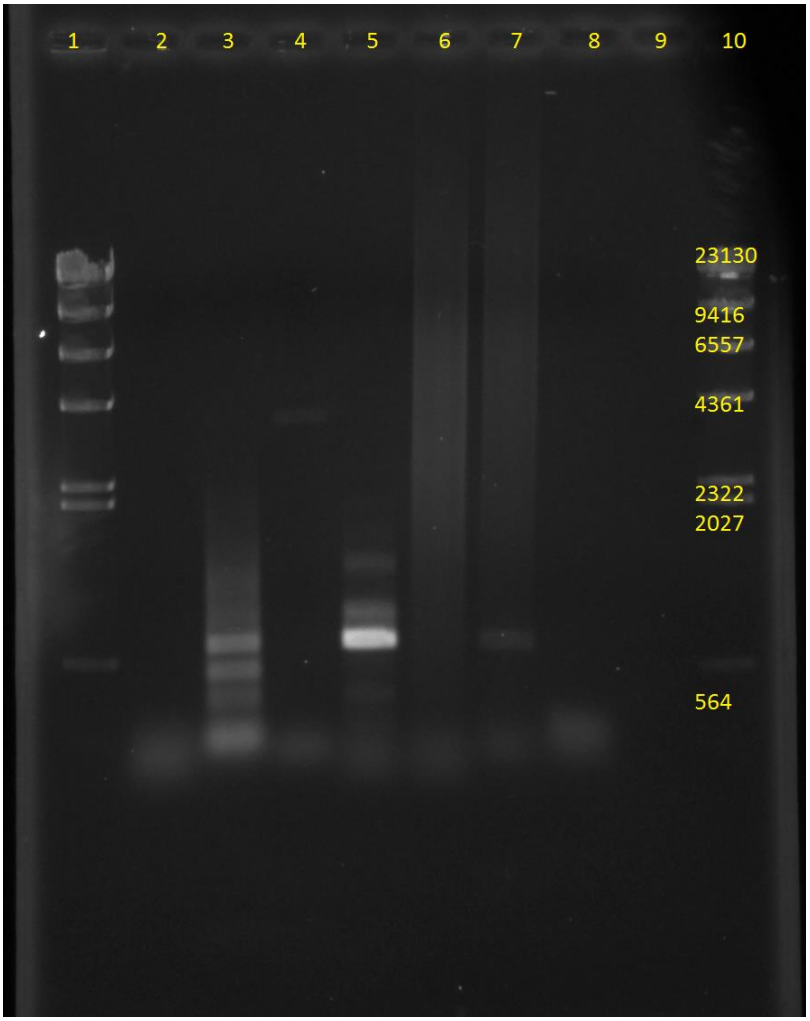


Figure 20. Gel results using all primers. Lanes 1 and 10 contain the ladder for base pair size. The vertical numbers are the base pair size of each band. Lane 2 was the contamination control. Lanes 4, 6, and 8 used sample 1D3m4R. Lanes 3, 5, and 7 used 5D5R sample. Lanes 3 and 4 used Start and A1_Stop primers. Lanes 5 and 6 used start and A2_Stop primers. Lanes 7 and 8 used Start and D_Stop primers. The band in Lane 4 around 4000 base pairs was gel extracted and used for further experiments.

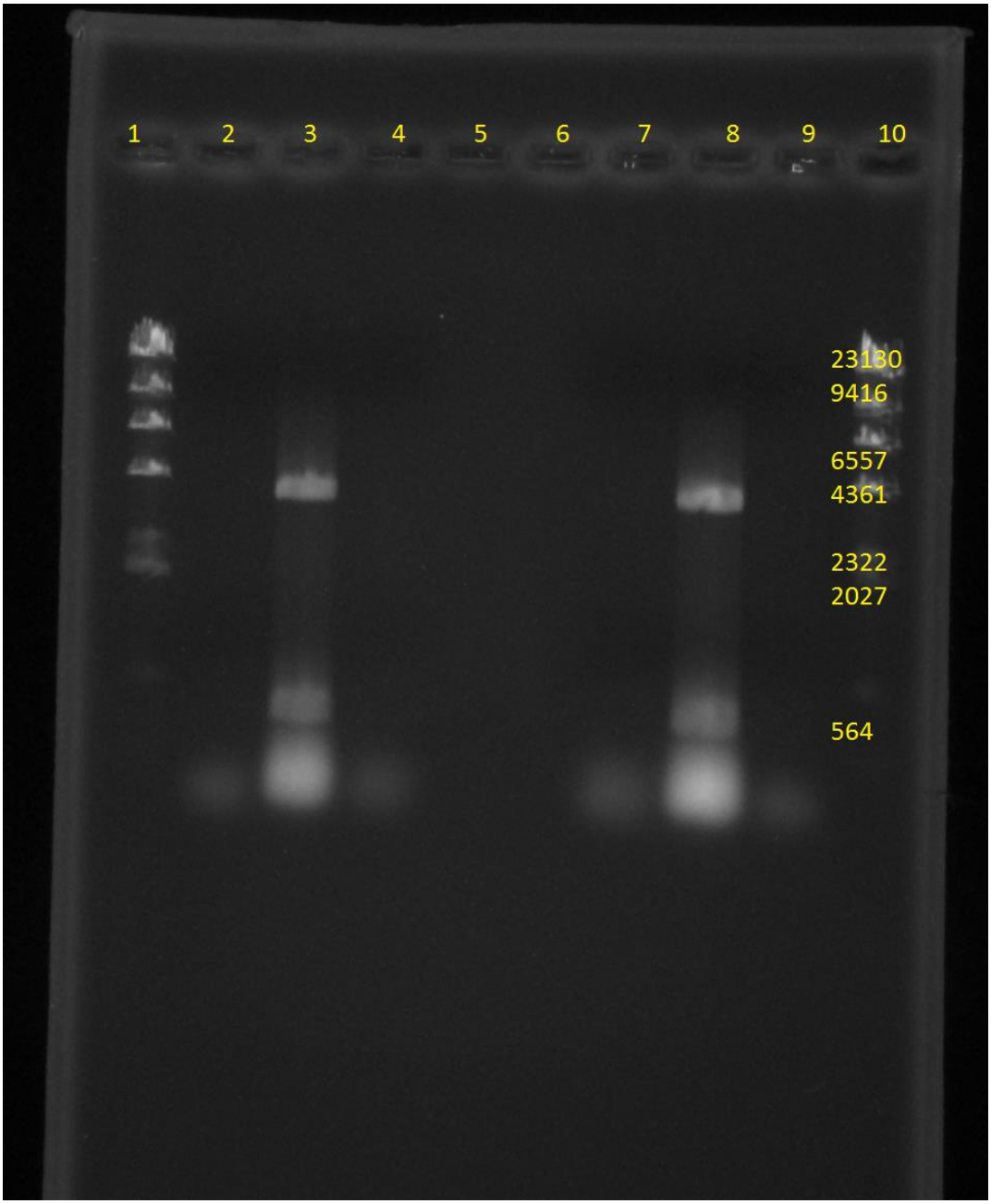


Figure 21. PCR of the gel extracted DNA from Figure 21. Lanes 1 and 10 contain the ladder for base pair size. The vertical numbers indicate the size of each band. Lanes 2 and 9 contain the contamination control. Lanes 3 and 8 contain the DNA extracted from Figure 21 lane 4. Lanes 3 and 8 contain the DNA extracted from Figure 21 lane 4.

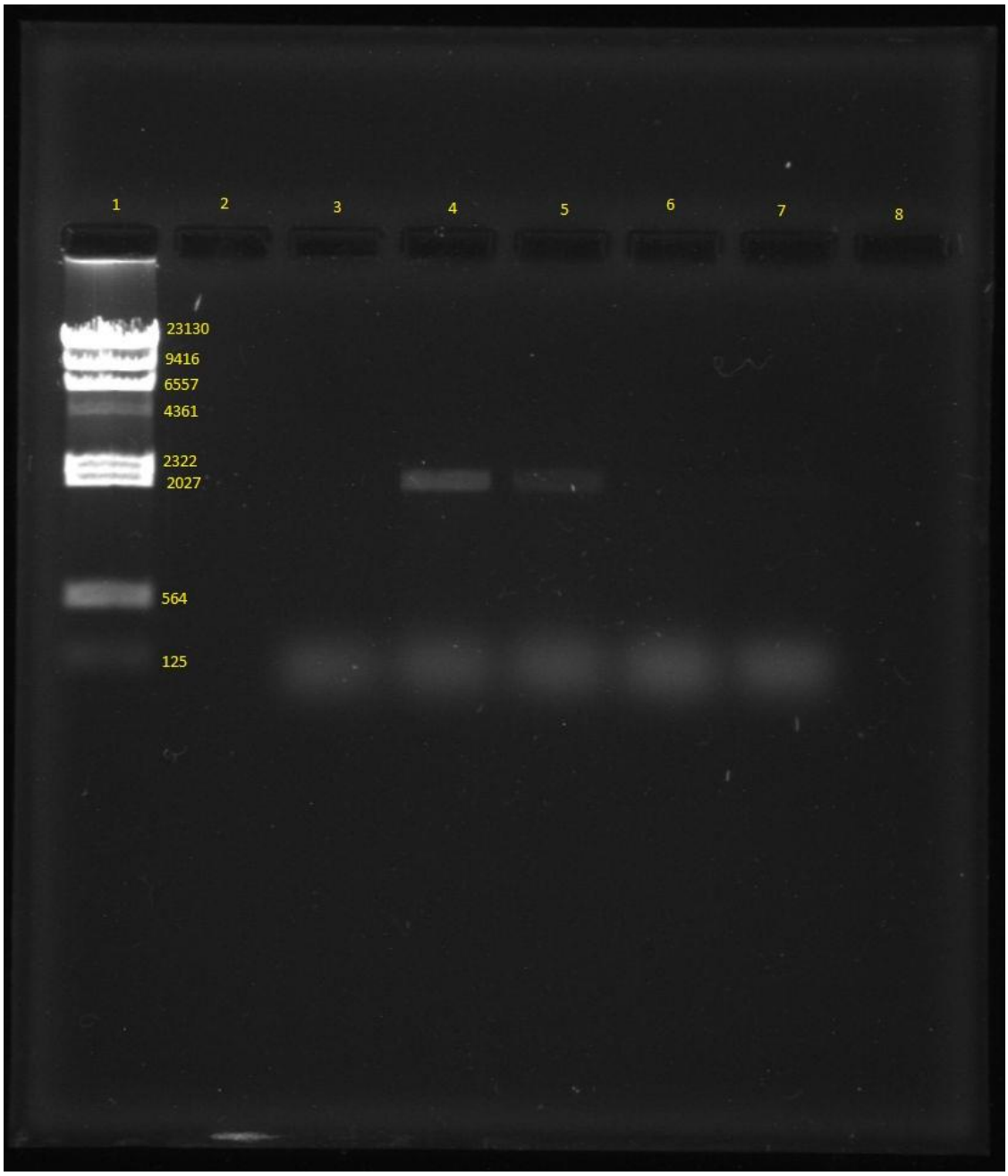


Figure 22. PCR result using bridge and stop primers on calibrated thermal cycler. Lane 1 contained the ladder. Lane 3 contained the contamination control. Lane 4 contained 4D4R. Lane 5 contained 5D5R. Lane 6 contained 3D3m4R. Lane 7 contained 2D3m4R. Bridge + and A_Stop 2 was used as the forward and reverse primers.

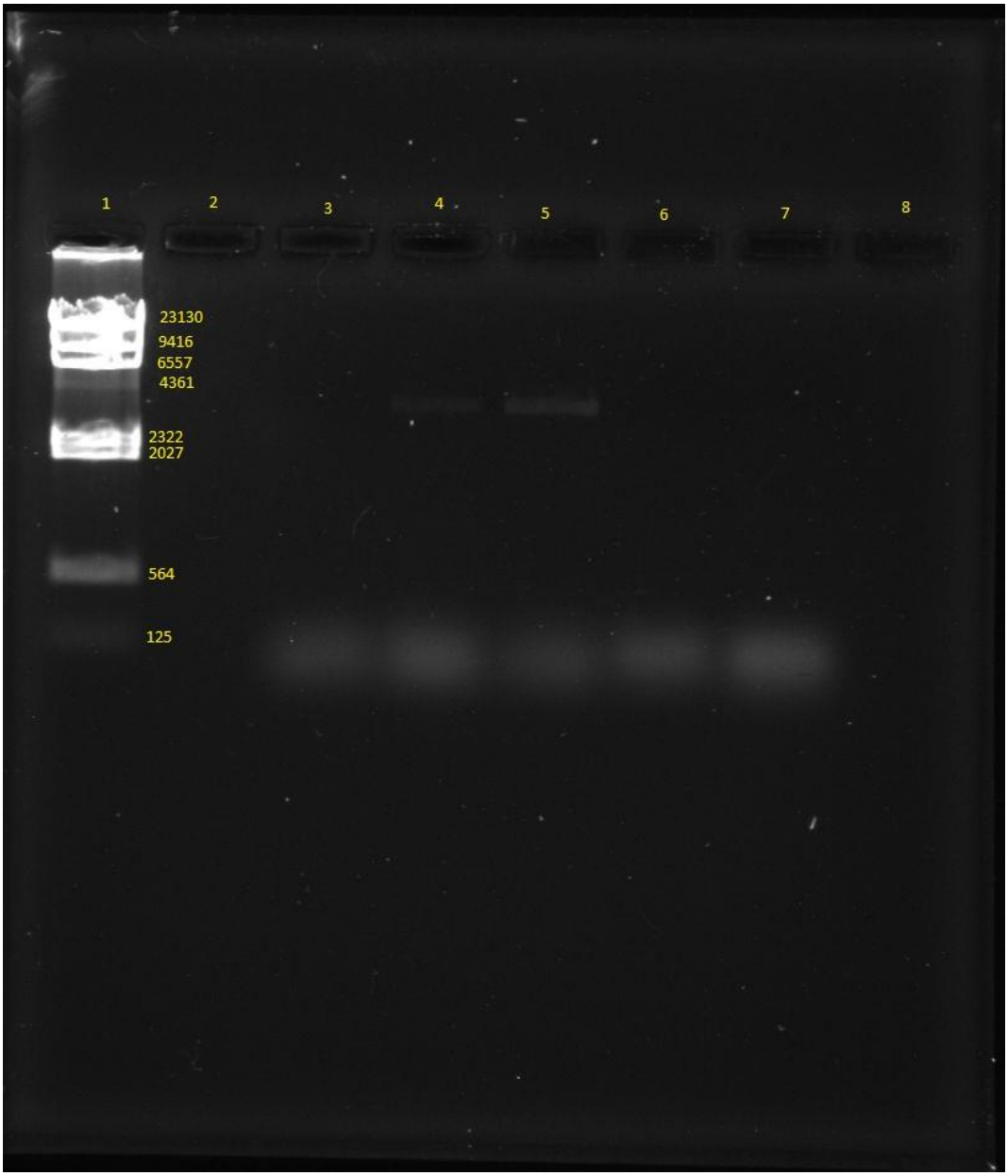


Figure 23. PCR reactions using Start primer and A2_Stop 2 primer. The lanes were the same as Figure 21.

Table 9. Base pair size of extracted agarose gel

Band Size	
Figure and Lane	**Size (bps)**
Figure 21, Lane 4	3944
Figure 23, Lanes 4 & 5	2058
Figure 24, Lane 4 & 5	3908

.

Gel Extraction

DNA was extracted from all lanes that showed bands around 2322 and 4361 base pairs. Table 6 contains the results.

Table 10. Absorbance and concentration of DNA extracted from agarose gel. Sample 1G came from Figure 21 Lane 4. Samples 2G and 3G came from Figure 21 Lanes 3 and 8 respectively. Sample 3G was used for transformation and cloning.

UV Absorbance and Concentration of DNA Extracted from Gel				
Sample	**Volume (µL)**	**ABS_{260}**	**ABS_{260}/ABS_{280}**	**Concentration (ng/µL)**
1G	2	0.00	0	0
2G	2	0.079	2.92	4.0
3G	2	0.110	2.18	5.5

All other attempts at gel extraction produced concentrations below 5.0 ng/µL which cannot be used for cloning. Sample 3G was used cloning.

Sequencing

Figure 25 shows the alignment from 3 clones. The first row is the forward primer from the vector. The second and third rows are the reverse primer. The reverse compliment was used for the alignment of rows two and three. The asterisks above the nucleotide means all three sequences align.

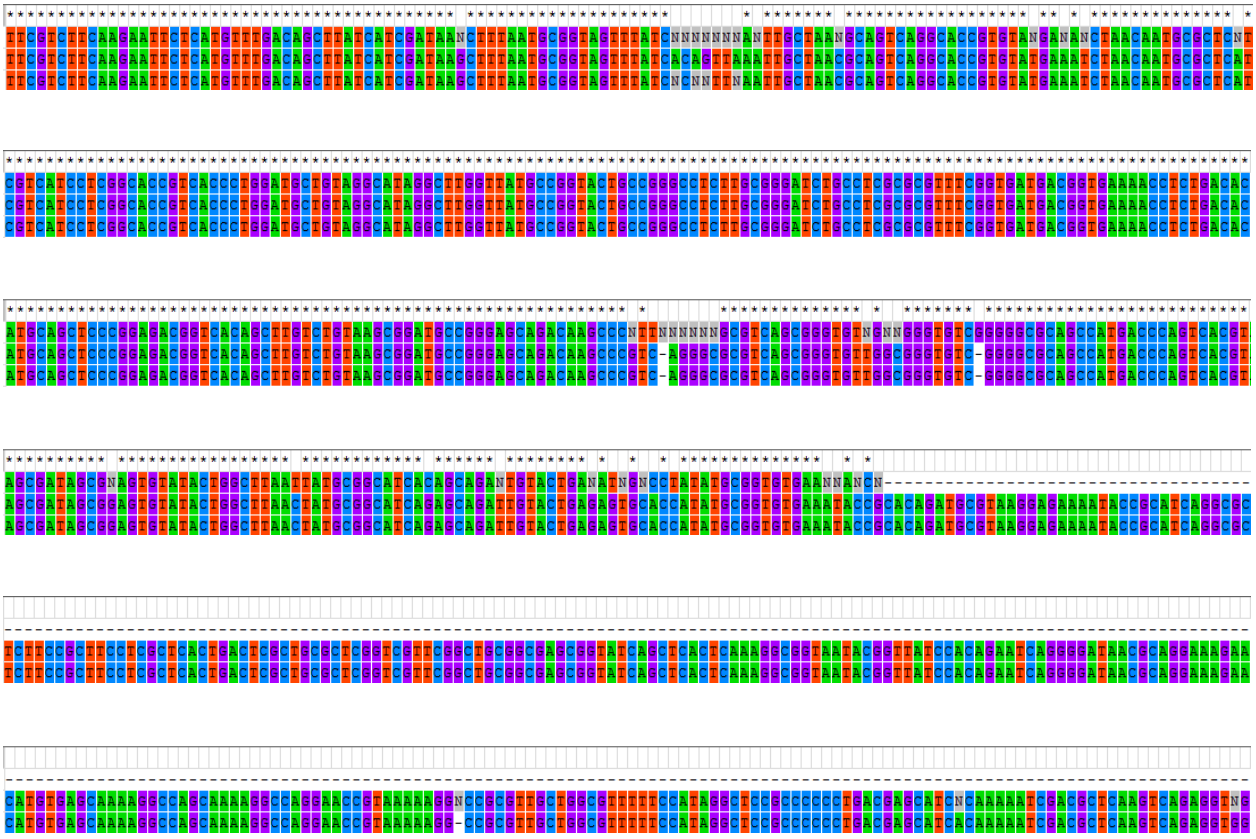


Figure 24. Alignment of three clones.

A BLAST search was done on the second rowing looking at expressed sequence tags.

Figure 26 is the top ten hits from the search.

Description	Max score	Total score	Query cover	E value	Ident	Accession
209p08.p1 AtM1 Arabidopsis thaliana cDNA clone MPMGp2011P08209 5-PRIME, mRNA sequence	963	963	48%	0.0	99%	gi|47828160|CK117844.1
EST 00017 Bark cDNA library of ramie Boehmeria nivea (L.) Gaud Boehmeria nivea cDNA clone SW01A0017 5', mRNA sequence	656	656	32%	0.0	99%	gi|124263280|EH667200.1
207j04.p1 AtM1 Arabidopsis thaliana cDNA clone MPMGp2011J04207 5-PRIME, mRNA sequence	635	635	30%	3e-178	100%	gi|47830676|CK120360.1
212o09.p1 AtM1 Arabidopsis thaliana cDNA clone MPMGp2011O09212 5-PRIME, mRNA sequence	635	635	30%	3e-178	100%	gi|47829713|CK119397.1
AL714571 Danio rerio embryonic inner ear subtracted cDNA Danio rerio cDNA clone BN0AA007ZC02 5', mRNA sequence	592	592	28%	1e-165	100%	gi|20179174|AL714571.1
AL715724 Danio rerio embryonic inner ear subtracted cDNA Danio rerio cDNA clone BN0AA018ZF12 5', mRNA sequence	589	589	28%	2e-164	100%	gi|20180327|AL715724.1
ACL5_09H05 AcL5 cDNA Library Angiostrongylus cantonensis cDNA, mRNA sequence	585	585	28%	3e-163	99%	gi|60292629|DN190868.1
327-384-46_B07_M13-FP Nematostella vectensis normalized cDNA library 327 Nematostella vectensis cDNA clone 327-384-46_B07_M13-FF	581	581	28%	4e-162	100%	gi|82873716|DV092323.1
BES1824105a23 BES1824 Hordeum vulgare subsp. vulgare cDNA clone MPMGp2010A23105 5-PRIME, mRNA sequence	556	556	34%	1e-154	95%	gi|44805683|CK122681.1
Y391B07 Hevea brasiliensis YLX600 library Hevea brasiliensis cDNA clone Y391B07 5', mRNA sequence	554	554	26%	5e-154	100%	gi|358646522|JG012149.1

Figure 25. Top ten hits from clones.

The top hit came back as an *Arabidopsis thaliana* clone for a protein kinase. The hits were so scattered that not one source could be considered. A second cloning attempt gave only the sequence for the cloning vector pMiniT.

CHAPTER 4

CONCLUSIONS

As abundant as *Gossypium hirsutum* is in the United States, there is little genetic information on it. Having two complete genomes makes it a daunting task. With both genomes just recently sequenced (published near the end of the work described in this thesis)[16], the task becomes easier to manage.

RNA was successfully extracted from cotton seeds using the Qiagen total RNA extraction kit. After some troubleshooting, two large batches of total RNA was extracted at concentrations of 43.1 and 854.5 ng/µL. The reason for the concentration of 854.5 ng/µL is due to not diluting the extraction a second time. Extracting mRNA from total RNA was not necessary for further analysis, but attempted with little success. mRNA concentrations did not even reach 1 ng/µL.

cDNA was synthesized from total RNA with little difficulty. Several attempts of quantifying samples were attempted but nothing was reliable. There was not a reliable source to zero out the instrument from background interference.

Optimizing PCR conditions was the most time consuming process. After trying several different methods of PCR, touchdown PCR, the optimal melting temperature was found at 48.6˚C. This temperature worked well with the Start primer as well as the A_Stop2 primer. Using the A_Stop1 primer and the start primer yielded a band around 3944 base pairs. The Start and A_Stop2 primers yielded a band around 3908. Finally, the Bridge + and A_Stop2 primer

yielded a band around 2057 base pairs. The full length sequence is around the size of most topoisomerase II found in similar species.[17] The Bridge + primer being in the middle of the sequence is consistent with it being about half the base pairs of the full length sequence. The biggest problem with PCR was trying to produce enough DNA to be used for sequencing. A solution to this problem could be running many PCR reactions in the thermal cycler and pooling the end products together. Doing this has one issue, more reactions means more gel. If the issue before was there was too little DNA in the gel slice, then trying to separate the gel from DNA using a kit would yield the same problem. This means too much DNA is being left behind on the column filter of the DNA gel extraction kit.

Cloning and DNA sequencing data was not reliable. The sequencing of the inserted DNA had too many mixed results from the BLAST search. Figure 23 shows the top BLAST hits. With so little DNA to work with, it's hard for the bacteria to take up the insert.

Description	Max score	Total score	Query cover	E value	Ident	Accession
209p08.p1 AtM1 Arabidopsis thaliana cDNA clone MPMGp2011P08209 5-PRIME, mRNA sequence	963	963	48%	0.0	99%	gi|47828160|CK117844.1
EST 00017 Bark cDNA library of ramie Boehmeria nivea (L.) Gaud Boehmeria nivea cDNA clone SW01A0017 5', mRNA sequence	656	656	32%	0.0	99%	gi|124263280|EH667200.1
207j04.p1 AtM1 Arabidopsis thaliana cDNA clone MPMGp2011J04207 5-PRIME, mRNA sequence	635	635	30%	3e-178	100%	gi|47830676|CK120360.1
212o09.p1 AtM1 Arabidopsis thaliana cDNA clone MPMGp2011O09212 5-PRIME, mRNA sequence	635	635	30%	3e-178	100%	gi|47829713|CK119397.1
AL714571 Danio rerio embryonic inner ear subtracted cDNA Danio rerio cDNA clone BN0AA007ZC02 5', mRNA sequence	592	592	28%	1e-165	100%	gi|20179174|AL714571.1
AL715724 Danio rerio embryonic inner ear subtracted cDNA Danio rerio cDNA clone BN0AA018ZF12 5', mRNA sequence	589	589	28%	2e-164	100%	gi|20180327|AL715724.1
ACL5_09H05 AcL5 cDNA Library Angiostrongylus cantonensis cDNA, mRNA sequence	585	585	28%	3e-163	99%	gi|60292629|DN190868.1
327-384-46_B07_M13-FP Nematostella vectensis normalized cDNA library 327 Nematostella vectensis cDNA clone 327-384-46_B07_M13-FF	581	581	28%	4e-162	100%	gi|82873716|DV092323.1
BES1824105a23 BES1824 Hordeum vulgare subsp. vulgare cDNA clone MPMGp2010A23105 5-PRIME, mRNA sequence	556	556	34%	1e-154	95%	gi|44805683|CK122681.1
Y391B07 Hevea brasiliensis YLX600 library Hevea brasiliensis cDNA clone Y391B07 5', mRNA sequence	554	554	26%	5e-154	100%	gi|358646522|JG012149.1

Figure 26. Top hits from clone DNA Blast search.

Future work could be in isolation of the topoisomerase II protein from *Gossypium hirsutum* and purifying it. With the purified enzyme, decantation assays could be performed and further look at enzyme kinetics.

REFERENCES

1. Voet, Donald; Voet, Judith; Pratt, Charlotte. Nucleic Acid Structure. *Fundamentals of Biochemistry*. 3rd ed; John Wiley & Sons, Inc.; New Jersey, **2008**; 850-865.
2. Altschul, Stephen; Gish, Warren; Miller, Webb; Myers, Eugene; Lipman, David. Basic local alignment search tool. *Journal of Molecular Biology.* **1990**. *Vol. 215*, 403-410.
3. Gota, T.; Wang, J.C. Yeast DNA Topoisomerase-II – An ATP-dependent type-II topoisomerase that catalyzes the catenation, decatenation, unknotting, and relaxation of double-stranded DNA rings. *Journal of Biological Chemistry.* **1982.** *Vol. 257*, 5866-5872.
4. Larsen, Annette; Escargueil, Alexandre; Skladanowski, Andrzej. Catalytic topoisomerase II inhibitors in cancer therapy. *Pharmacology & Therapeutics.* **2003**, *Vol. 99*, 167-181.
5. Adams, Keith L.; Wendel, Jonathan F. Exploring the genomic mysteries of polyploidy in cotton. *Biological Journal of the Linnean Society*. **2004.** *Vol. 82*, 573-581.
6. Kenar, James. Reaction chemistry of gossypol and its derivatives. *JAOCS.* **2006**, *Vol. 83*, 269-296
7. Wildsmith, K. R., C. J. Albert, et al. Myeloperoxidase-derived 2-chlorohexadecanal forms Schiff bases with primary amines of ethanolamine glycerophospholipids and lysine. *Chemistry and Physics of Lipids*. **2006**. *Vol. 139*, 157-170.
8. Tort, N. Effects of light, different growth media, temperature and salt concentrations on germination of cotton seeds (*Gossypium hirsutum* L cv Nazilli 87). *Journal of Agronomy and Crop Science-Ziet Schrift Fur Acker und Pflanzenbau.* **1996.** *Vol. 176,* 217-221.

9. Mullis, K.; Faloona, F.; Scharf, S.; Saiki, R.; Horn, G.; Erlich, H. Specific enzymatic amplification of DNA *in vitro* – the polymerase chain-reaction. *Cold Spring Harbor Symposia on Quantitative Biology.* **1986.** *Vol. 51,* 263-273.
10. Don, R.; Cox, P.; Wainwright, B.; Baker, K.; Mattick, J. "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **1991.** *Vol. 19,* 4008.
11. Kuroiwa, T.; Suzuki, T.; Kawano, S. Staining of DNA-containing area (nucleus) in phages, bacteria, blue green-algae and organelles by ethidium-bromide. *Cell Structure and Function*. **1979.** *Vol. 4,* 399-399.
12. Finklestein, M.; Rownd, R. Rapid method for extracting DNA from agarose gels. *Plasmid.* **1978.** *Vol. 1,* 557-562.
13. Huss, Volker; Festl, Herbert; Schelifer, Karl. Studies on the spectrophotometric determination of DNA hybridization from renaturation rates. *Systematic and Applied Microbiology.* **1983.** *Vol. 4*, 184-192.
14. Roskams, Jane; Rodgers, Linda. *Lab Ref: A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench.* Cold Spring Harbor: New York, **2002,** 47.
15. Ausubel, F. M. *Current Protocols in Molecular Biology.* John Wiley & Sons: New York, **1987.**
16. Li, Fuguang; Fan, Guangyi; et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature Biotechnology.* **2015.** *Vol. 3,* 524-530.
17. Xie, S. P. and E. Lam . Abundance of Nuclear-DNA Topoisomerase II is correlated with proliferation in *Arabidopsis thaliana*. *Nucleic Acids Research* **1994,** *Vol. 22*, 5729-5736.