DNA Sequence Analysis of Intrastrain variants of Herpes Simplex Virus-Type 2.

by

Chisaroka W. Onunwor

Submitted in Partial Fulfillment of the Requirements for the Degree of Masters of Science in the Chemistry Program

YOUNGSTOWN STATE UNIVERSITY

May, 2002

DNA Sequence Analysis of Intrastrain variants of Herpes Simplex Virus-Type 2.

Chisaroka W. Onunwor

I hereby release this thesis to the public. I understand this thesis will be housed at the Circulation Desk of the University Library and will be available for the public access. I also authorize the University of other individuals to make copies of this thesis as needed for scholarly research.

Signature: Chisaroka Onunwor 9/30/05

Chisaroka W. Onunwor, Student Date

Approvals:

Thomas D Kim 10/5/05

Dr. Thomas D. Kim, Thesis Advisor Date

Jeffrey A Smiley 10/5/05

Dr. Jeffrey A. Smiley, Committee Member Date

Daryl W Mincey 10/5/05

Dr. Daryl Mincey, Committee Member Date

Peter J Kavinsky 10/24/05

Peter J. Kavinsky, Dean of Graduate Studies Date

Abstract

PCR products from intrastrain variants of HSV-1 vary in size. The length of the gene product determines the length of protein, and hence the number of Proline-Alanine-Threonine (PAT) repeats that are present. Some strains of HSV-2 have a 13-nucleotide (13-mer) repeat, which are unlike the 9-mer repeats found in HSV-1. The repeat sequence of the 13-mer would cause the end of the ICP34.5 gene in HSV-2 to be out of frame such that the encoded protein will be different. One possibility is that the repeat is deleted in the messenger RNA (mRNA). Another possibility is that there is a three-fold repeat in the number of 13-mer copies, which would then result in a gene that continues in frame. In this project, the reasons why the 13-mer in HSV-2 does not affect the ICP34.5 gene and cell infection are investigated. The size and sequence of genetic DNA and the cDNA obtained from PCR will be compared. Amino acid sequences deduced from genes with the 13-mer repeats are also determined for the following HSV-2 strains – 333, 414, 427, 443, 471, 519, 564, HSV-2, 392, 472 and MS.

Acknowledgements

I would like to thank my thesis advisor Dr. Kim for giving me the opportunity to conduct research under his tutelage. I would also like to thank Dr. Ken S. Rosenthal for being a mentor and also for making his resources available to our research. I thank my thesis committee for their time in reviewing and critiquing my work. My appreciation also goes out to all those who have made my time at YSU very enjoyable, especially Dr. Sherri Lovelace-Cameron for her encouragements and friendship. I would like to acknowledge Ray Hoff, John Caplinger, Diana Arnett and Irene Panagapoulos for their help in lab. I would also like to acknowledge Dr. Bruce Levison and the center for Biomedical and Environmental research. I thank Dr. H.E Lorimer for making Molecular Virology interesting, and for helping in the formation of my latest interests.

My friends and family have been the greatest source of inspiration to me. I want to thank Dr. Manuel Pomales for his unwavering belief in me, incessant words of encouragement and for creating ways for me to progress. I thank Dr. and Mrs. Nnewihe for their support and guidance. Also "Uncle Emma" for inspiring me to achieve, and for always believing in me. I appreciate the love from all my brothers and sisters, and the incredible breaks from reality I got from "Ne pas de femme."

I dedicate this to my wonderful parents HRH, Sir and Lady W.W Onunwor for their tremendous support in all my undertakings. This is for encouraging me to always find satisfaction in my pursuit of success and ambition, and for teaching me persistence.

Most importantly, I thank God for all His mercies, and for granting me grace and divine favor in all my deliberations.

Table of Contents **Page**

List of Figures **Page**

List of Symbols and Abbreviations

cDNA	Complementary DNA
CNS	Central Nervous System
COSY	Correlational Sprectroscopy
ddNTP	DideoxyNucleotide Triphosphate
dNTP	Deoxy Nucleotide Triphosphate
DNA	Deoxyribonucleic Acid
DRG	Dorsal Root Ganglia
GC	Guanine Cytosine
HSV	Herpes Simplex Virus
HSV-1	Herpes Simplex Virus-Type 1
HSV-2	Herpes Simplex Virus-Type 2
ICP 34.5	Infected Cell Protein 34.5
LP5	Large Plaque 5
mRNA	Messenger RNA
NMR	Nuclear Magnetic Resonance
PAT	Proline Alanine Threonine
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

SP7	Small Plaque 7
fmol	Femtomole
µL	Microliter
mL	Milliliter
ng	Nanogram
rpm	Revolutions per minute

Chapter 1: INTRODUCTION TO HERPES SIMPLEX VIRUSES, INTRASTRAIN AND INTERTYPIC VARIATIONS OF HSV-2

Section 1.1 Introduction

Herpes Simplex Virus (HSV) is ubiquitous, contagious and causes lifelong infections in humans. HSV is usually transmitted by contact with infected material. HSV, type 1 (HSV-1) is mostly transmitted by oral means (kissing, sharing toothbrushes, sharing drinking glasses, etc), while HSV-2 is generally transmitted by sexual contact. The risks for severe disease are higher with immunosuppressed people such as neonates and geriatrics. Current methods of controlling HSV infection involve inhibition of DNA synthesis (replication). HSV infects by attaching its glycoproteins to the cell surface and eventually emptying its contents (nucleic acids) into the host cell. The infection is lytic in most cells, and latent in neuronal cells. While lytic infections kill the host cell and release the virus into the matrix, some strains of HSV have developed methods of escaping this route. Giant multinucleated cells (synctycia) are formed by some such strains as a method of escaping antibody recognition and immune response. They do this by fusing two or more cell membranes such that virus particles are transferred from cell to cell without being exposed to the blood matrix, which contains antibodies. Latent infections are reactivated by high stress conditions or compromised immunity.

HSV is characterized by double stranded DNA in an electron-opaque core, surrounded by an icosadeltahedral capsid and an envelope with a space between the capsid and outer envelope called the tegument. HSV is easily disinfected by detergents, weak acids and other relatively harsh conditions. The HSV genome is comprised of a long sequence and a short contiguous sequence of DNA. This encodes at least 80 genes, only half of which are needed in viral replication. The remaining genes enhance

interactions between the virus and the host cells. HSV also encodes structural, fusion and immune response proteins (Ward and Roizman, 1994).

The first step of HSV replication occurs when viral glycoproteins recognize and attach to a cell surface receptor. The virus envelope then fuses with the plasma membrane of the host cell, causing release of the nucleocapsid into the cell matrix. The nucleocapsid attaches to the nuclear membrane and the viral genome is transferred into the host's nucleus. Early gene products such as DNA-dependent polymerases are produced and function as catalysts in replication. Circular contiguous DNA is formed, and as replication continues, the circular piece continues to wind over a "reel." The DNA is drawn into a procapsid, resulting in individual genomes. Late genes are transcribed after genome replication, and serve to encode several structural proteins. Capsid proteins are taken into the nucleus, where they form procapsids filled with DNA. In the endoplasmic reticulum, glycoproteins receive a glycan precursor and are released in the nuclear membrane. The capsids (containing DNA) assemble and bud from the nuclear membrane, specifically from areas that are modified by viral glycoproteins. The virus buds and is transferred into the endoplasmic reticulum, processed in the Golgi apparatus, and released either by exocytosis or host cell lysis (lytic infection), or the virus can be transferred through synctytia formation, avoiding antibodies (Murray, et al. 1998).

Section 1.2 Intrastrain variants of HSV-2

HSV has been shown to cause localized infections as well as disseminated disease in infants (Bower, et al. 1999). Neonates generally contract disease from infected mothers during pregnancy. Infected neonates sometimes go blind, but more often, neonate infection results in death. Virulence or disease severity is a function of host

immunity, entry site, and type of viral strain. HSV-1 causes "cold sores" and primary encephalitis at peripheral sites. HSV-2 symptoms are characterized by genital warts and sores, which reflect the viral entry site. The virus has also been shown to cause disease in the central nervous system (neuronal tissue) of a mouse model. This is known as neurovirulence, and is associated with viral genes that encode thymidine kinase, ICP 34.5 and other activities necessary for viral replication. This is yet another mechanism of immune escape. Antibodies are restricted from attacking neuronal tissues, even when these tissues are harboring HSV. The virus finds a safe haven in cells which are generally not susceptible to an attack by the immune system. Neurovirulence allows the virus to initiate tissue damage before immune detection.

HSV characterization comprises a description of plaque morphology, protein structure as well as genetic classification. Two intrastrain variants of HSV-1 have been thoroughly characterized by Bower and coworkers. The variants were obtained post-mortem from an infant who suffered fatal disseminated disease—a small plaque strain and a large plaque strain (Bower et. al, 1999). Plaque assays were used to measure the amount of infectious particles per unit volume, and also show different infection characteristics based on the source of the viral strain. The small plaque strain (SP7) was predominantly found in tissues of the central nervous system (CNS), while the large plaque strain (LP5) was primarily in the lungs and gastrointestinal tract. Plaque phenotype appears to be relevant to virulence—the SP7 harvested from the brain and spinal cord are lethal, while the LP5 strain has lower mortality in mice. The isolates from different tissues have been characterized, and differences in both strains have also been studied.

The virus was assayed on Vero (transformed African Green Monkey Kidney) cells. Cell extracts from infected cells were electrophoresed, and restriction enzyme digests were used to observe genetic relationships between both strains. Recombination analyses were also conducted to study transfer of marker for large plaque characteristics (i.e glycoprotein processing.) Polymerase Chain Reaction (PCR) was used to amplify viral DNA for further sequence analysis.

Mice were inoculated in brain and periphery (foot pad). Three groups of mice inoculated in periphery were sacrificed 0, 1, 3, 5, 7 or 9 days post infection. Their feet, dorsal root ganglia (DRG), spinal cord, brain and sciatic nerve were removed and virus titers were determined from tissues. Plaques from the SP7 strain were approximately 0.5mm in diameter, and showed preferential migration to neuronal tissue. The LP5 strain however was about twice as large as the SP7, and was found primarily in the gastrointestinal tract and lungs.

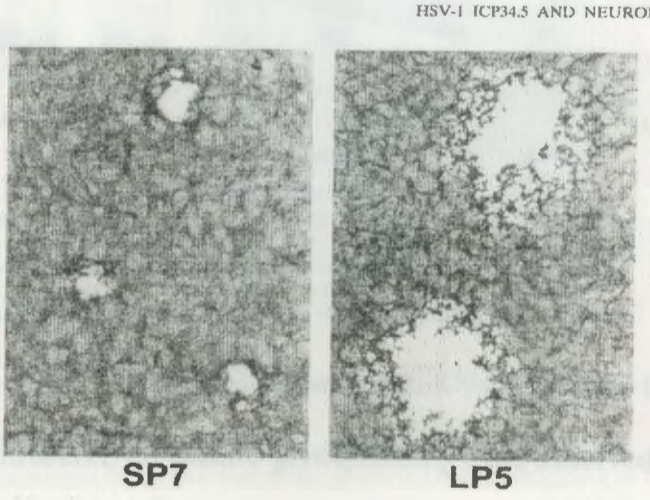


Fig 1.1: Bower, et al. J.Virol 1999:73 (May)

The clear regions shown in the figure above are plaques, and represent areas of viral control and host cell lysis. As shown, the LP5 strain produced 10-fold more viral

particles than the SP7 strain. The ratio of extracellular:cell-associated virus was 0.9 for LP5 and 0.06 for SP7. The small amount of viral release from SP7 shows that the strain is capable of replication, yet highly cell associated.

Glycoprotein maturation was also investigated in SP7 and LP5 using sodium dodecyl sulfate-polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western blot. The small plaque strain showed only precursor forms of both glycoprotein C (gC) and glycoprotein D (gD).

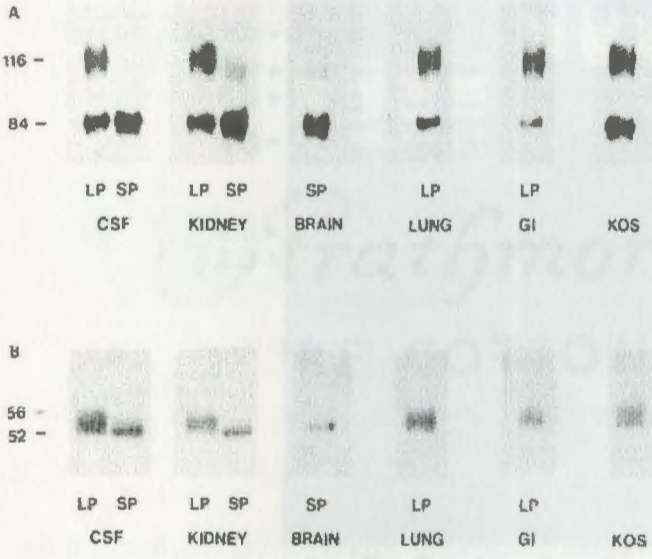


FIG. 2. HSV-1 glycoprotein differences associated with large-plaque and small-plaque variants. Whole-cell detergent extracts from Vero cells infected with large- or small-plaque variants from the CSF, kidney, brain, lung, or GI tract were examined by SDS-PAGE and Western blot analysis using either polyclonal anti-gC (A) or anti-gD (B). The laboratory strain KOS was included as a control. and molecular weights of the major glycoforms (kilodaltons) are indicated.

Fig 1.2: Bower et. al. J.Virol 1999:73 (May)

At 34 hours post infection, no mature glycoproteins were observed on the gel. These results indicate that glycoprotein processing in SP7 is blocked, not slowed at precursor level. Decreased virion release and glycoprotein processing correlates to decreased viral progression through cell.

Restriction fragment length polymorphism was used to determine if SP7 and LP5 are from the same origin. Cell extracts containing viral DNA were digested with BamHI and EcoRI, and electrophoresed in 0.8% agarose gel. The cleavage patterns were similar for both strains, indicating that the small plaque and large plaque isolates are derived from the same parental strain.

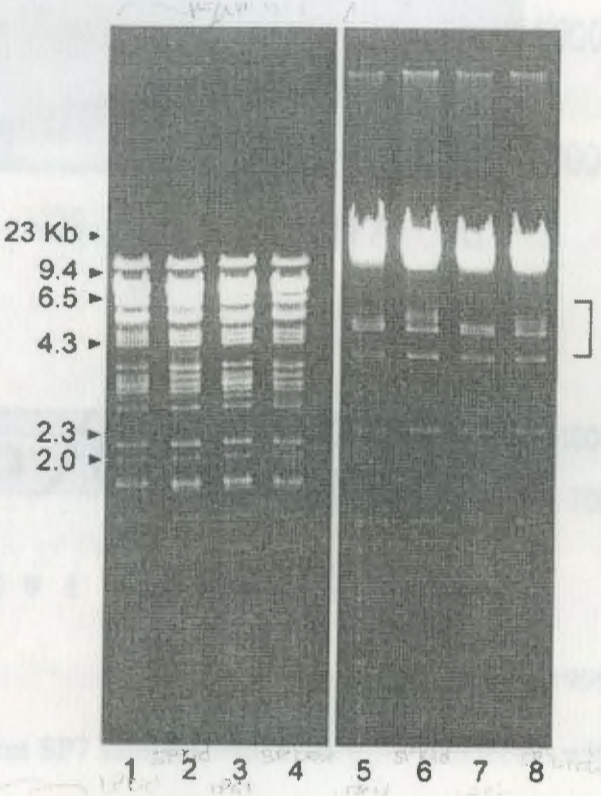


Fig 1.3: Bower et al. J. Virol 1999:73 (May)

Lanes 1, 3, 5 and 7 are large plaque strains, while 2, 4, 6 and 8 and small plaque stains. Both strains appear to be genetically distinct yet related viruses as shown by their similar cleavage patterns.

Marker transfer experiments (recombination) were also used to determine the genes responsible for phenotypic changes observed in small plaque to large plaque transformations. Small plaque virus was co-transfected with EcoRI digests of the KOS

strain, which exhibits large plaque characteristics such as efficient glycoprotein processing, plaque phenotype, etc. Analysis of the sample by PCR illustrated the occurrence of recombination as the only bands present corresponded with that of the KOS strain.

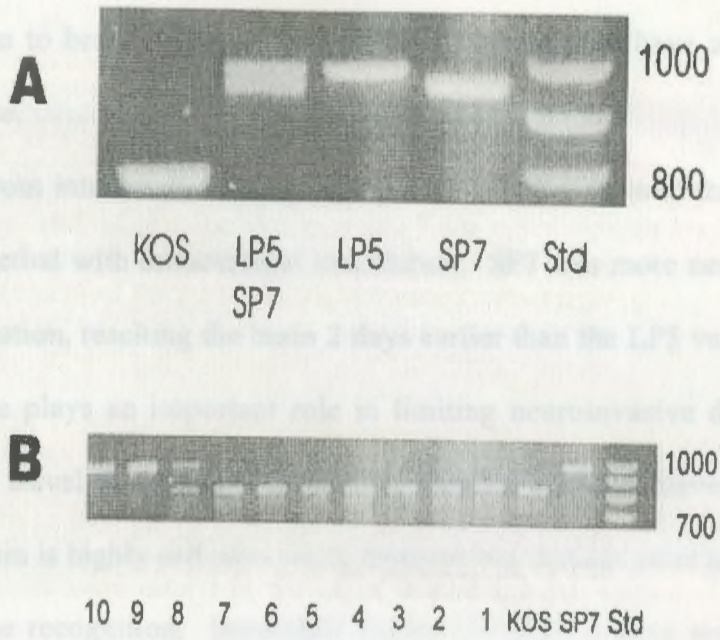


Fig 1.4: Bower et al. J. Virol. 1999:73 (May)

Lanes 1 through 8 represent SP7 samples which were co-transfected with KOS. Lanes 8 and 9 are SP7 bands only. These results show that characteristics of intrastrain variants can be manipulated via recombination.

PCR amplification of the ICP34.5 gene was accomplished using commercially available kits which are able to unzip GC-rich sequences, and minimize formation of secondary loops. PCR products of SP7 and LP5 varied significantly in size. SP7 product contained about 900 base pairs, while the LP5 contained about 950 base pairs. The sequence of the amplified Unique Long 34.5 (UL34.5) gene showed homology within

ICP 34.5 of SP7 and LP5. The major difference between these strains is in the number of CCGGCGACC sequence repeats, which encode Proline-Alanine-Threonine (PAT). The LP5 strain contained 22 PAT repeats while the SP7 strain had only 18. A few deletions were observed as well at different locations for both strains. Results suggest that differences in ICP34.5 of both strains must be in their ability to extend from site of primary infection to brain. The number of PAT repeats may have an impact in the severity of disease.

Results from intracerebral and peripheral inoculation showed that both SP7 and LP5 strains are lethal with intracerebral inoculation. SP7 was more neurovirulent with peripheral inoculation, reaching the brain 2 days earlier than the LP5 variant. Although immune response plays an important role in limiting neuroinvasive disease, the SP7 strain appears to travel to the brain faster than immune response development. Also, since the SP7 strain is highly cell associated, there are not enough virus particles released for quick immune recognition. Intrastrain variants of HSV-2 have not been properly characterized. The project at hand is centered around genetic characterization of several variants of HSV-2 through DNA sequence analysis.

Section 1.3 References

1. Ward, P., Roizman, B. (1994) *Trends Genet.* **10,** 267-274.

2. Bower, J.R., Mao, H., Durishin, C., Rozenbom, E., Detwiler, M., Rempinski, D., Karban, T.L., and Rosenthal, K.S., (1999) *J. Virol.* **73,** 3843-3853.

3. Murray, P., Rosenthal, K., Kobayashi, G., Pfaller, M. (1998) Medical Microbiology, Mosby Inc., St. Louis, Third Edition, p. 405-430.

Chapter 2: STRUCTURAL DETERMINATION OF INTERTYPIC HSV-2 STRAINS USING 2-DIMENSIONAL NMR SPECTROSCOPY AND DNA SEQUENCE ANALYSIS

Section 2.1 Abstract

PCR products from intrastrain variants of HSV-1 vary in size. The length of the gene product determines the length of protein, and hence the number of Proline-Alanine-Threonine (PAT) repeats that are present. Some strains of HSV-2 have a 13-nucleotide (13-mer) repeat, which are unlike the 9-mer repeats found in HSV-1. The repeat sequence of the 13-mer would cause the end of the ICP34.5 gene in HSV-2 to be out of frame such that the encoded protein will be different. One possibility is that the repeat is deleted in the messenger RNA (mRNA). Another possibility is that there is a three-fold repeat in the number of 13-mer copies, which would then result in a gene that continues in frame. In this project, the reasons why the 13-mer in HSV-2 does not affect the ICP34.5 gene and cell infection are investigated. The size and sequence of genetic DNA and the cDNA obtained from PCR will be compared. Amino acid sequences deduced from genes with the 13-mer repeats are also determined for the following HSV-2 strains – 333, 414, 427, 443, 471, 519, 564, HSV-2, 392, 472 and MS.

Section 2.2 Background and Significance

In the past 20 years, nuclear magnetic resonance has made significant contributions in the structural determination of molecules. More recently, the dynamics of protein folding and tertiary structure of biological molecules can be derived from multidimensional NMR spectroscopy. This technique is unparalleled in its ability to characterize the 3-dimensional structures of soluble proteins (Wright, P.E., 1989). Samples are placed in a magnetic field where individual nuclei absorb electromagnetic radiation. Individual molecule assignments are summed to give a series of peaks, which

represent the molecule in one dimension. Cross peaks and couplings are derived from the single dimension structural data to produce a series of dots which can easily be interpreted in 2D. These "dots" give detailed information about the surroundings of individual atoms since the amount of radiation absorbed by each nuclei is a function of its environment. The data provides useful information for determination of thermodynamic and kinetic parameters such as proton exchange rates, pKa values, etc. of proteins.

NMR technology is preferred over the widely used x-ray crystallography method of characterization for multiple reasons: As stated previously, NMR allows for structural determination of proteins in solution, where crystals are needed for x-ray. In some cases, only a fraction of desired material is recovered from crystallization, and the sample treatment could get burdensome. Although advancements in computer technology have greatly improved collection of x-ray data, the production of single crystals remains very tedious. Also unlike x-ray data, NMR produces high resolution peaks representing mobile regions of a given molecule. Proteins generally assume thousands of conformations due to bond rotations, but the most stable conformations can be studied using NMR. Molecular modeling and methods in computational chemistry use this property to calculate stability constants and other order parameters for different (especially the most stable) conformations of a protein.

The significance of multidimensional NMR at an atomic level is manifested in detailed assignments of _H, ^{13}C, ^{15}N, ^{31}P, etc at high resolution. When applied, nuclear Overhauser effect also points out relationships between neighboring protons (coupling). Coupling constants derived from NMR data also give information about the existence of chemical bonds between neighboring atoms (Wright, P.E., 1989). On a molecular level,

structural roles can be assigned to different proteins as a result of their interactions in the organism.

Structural determination of biological molecules is relevant to the understanding of cellular and molecular basis of diseases. Knowledge of the molecular interactions that cause disease serves as an important tool in synthesis of drugs. The efficiency of NMR does not hold up for very large biomolecules. For example, it would be too cumbersome to structurally determine the order of bases in a sample of deoxyribonucleic acid (DNA), containing thousands of bases. To this end, DNA sequencing technology plays an important role. This technique is based on a combination of PCR amplification and the Sanger-dideoxy method.

PCR is a method of amplifying small amounts of DNA. It can be used to increase the number of DNA copies in a mixture of DNA molecules, where the desired material is in low concentration. This method is useful in that large amounts of tissue samples are not needed to extract sufficient DNA for experiments. Small amounts of DNA can be transcribed exponentially using PCR. The technique is not trivial as the sequence of some oligonucelotide fragment of the template DNA is required to construct template primers. Oligonucleotide primers which are complimentary to opposite strands of the DNA are needed to initiate the formation of new strands.

PCR technique is based on an understanding of the biochemistry of DNA replication. DNA is extracted from its source (e.g. a virus-infected cell culture) using commercially available kits. The DNA molecule is buffered, and mixed with excess oligonucleotides, primers, and dideoxynucleotides (ddNTPs). The template DNA is "unzipped" by heating at a temperature above its melting temperature. The temperature

is reduced to a lower annealing temperature, where primers selectively bind to complimentary sites of single strand from the denatured DNA. In the presence of DNA polymerase and excess oligonucleotides, the primers extend to form a new DNA molecule.

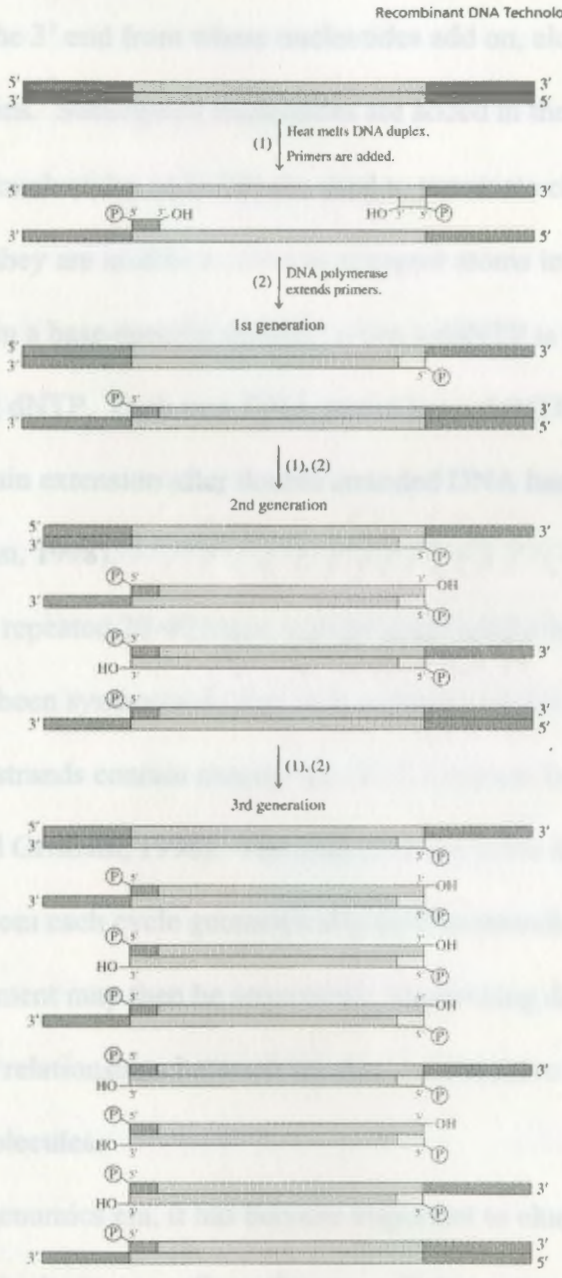


Fig 2.1: Principles of Biochemistry (1996, p. 715)

The primers are simple oligonucleotides that selectively bind and base pair with single stranded DNA, and stretch to form double stranded DNA in the presence of DNA building blocks (dNTPs), and a DNA polymerase (catalyst). The primers have a free hydroxyl group at the 3' end from where nucleotides add on, elongating the chain as polymerization begins. Subsequent nucleotides are added in the 5' → 3' direction.

2,3-dideoxynucleotides (ddNTP) are used to terminate chain elongation. Since they lack a 3'-OH, they are unable to serve as acceptor atoms in 5'-chain elongation. The termination occurs in a base-specific manner, when a ddNTP is incorporated in the place of its corresponding dNTP. Each new DNA strand has a ddNTP at its 3' end. Hence, the ddNTPs prevent chain extension after double stranded DNA has been synthesized (Garrett and Grisham, 1998).

The cycle is repeated 20-40 times in most cases until a large amount of DNA with defined strands has been synthesized. For each molecule of DNA duplicated after the first cycle, the new strands contain exactly the DNA fragment between the 5' ends of the primers (Garrett and Grisham, 1998). The ends of the primers define the length of newly synthesized DNA from each cycle geometrically until thousands of desired fragment are produced. The fragment may then be sequenced. Sequencing data is important in determining genetic relationships between species, or to resolve other structural parameters of biomolecules.

In the post-genomics era, it has become important to elucidate structure-function relationships for multiple reasons. Gene therapy will thrive from such knowledge as the cures to many gene-related illnesses will be much closer. This is clear in the growing

interest in structural and computational biology/bioinformatics. Scientists are developing methods that utilize information from DNA sequence analysis to figure out functions of encoded amino acids and larger proteins. It is also important to know the sequence of bases in viral or bacterial nucleic acids that encode infectious proteins. Methods in recombinant DNA technology or other avenues of studying gene regulation can be applied to control virulence. Also, classification of herpesviridae will be augmented based on the knowledge of these sequences. Gene conservation can be studied as viruses evolve from one species to the other. This would be important in determining the proteins that are important in sustaining infections in the presence of mutations.

Section 2.3 Hypothesis

At the primary level, proteins are simply linear arrangements of amino acids. These primary structures serve as blueprints in protein synthesis, and represent a small aspect of the whole molecule. The sequence and conformation of proteins at this level provide information about their stability and genetic peculiarities. Spatial arrangements of amino acids are a function of their pKa's and pH of their surroundings. At physiological pH levels for example, some amino acids are anionic, others are cationic while others may carry no charge. Under the same circumstances, amino acids with similarly charged side chains will repel each other, causing bond rotations which result in distortion of spatial arrangements. Base configuration in nucleic acids affect the order of proteins, which in turn influences genetic and phenotypic characteristics of organisms. In the same manner, the arrangement of bases in viral DNA will affect the coding sequence of proteins in HSV. A single amino acid residue is encoded by three nucleic acid base pairs. In HSV-2 however, there appears to be a sequence of 13 bases expected to encode PAT. The 13-

mer (clearly not a multiple of 3) does not affect ICP34.5 possibly because of deleted introns. Another possibility is a 3 x 13-mer sequence where the gene becomes in frame, and each protein can be accounted for. This is a possible explanation for intrastrain and intertypic variation in HSV characteristics.

Section 2.4 Rationale

As infectious organisms evolve, they are likely to conserve the genes which aid more efficient interactions with host cells without being destroyed themselves. The genetic composition of different herpes viruses can be analyzed using modern DNA sequencing methodology. Intrastrain and intertypic differences can be observed to study the conservation of genes. Protein conformation can be studied using computational modeling techniques as well as NMR. Most importantly, sequence analysis will be useful in determining the arrangement of bases, and the proteins they encode for intrastrain variants of HSV-2. Results from these will help in analyzing gene interactions, and the functions of different proteins in viral infectivity and host immune detection.

Section 2.5 References

1. Wright, P.E. (1989) *Trends Biochem. Sci.*, **14**, 255-259.
2. Horton, H.R., Moran, L.A., Ochs, R.S., Rawn, J.D., Scrimgeour, K.G. (1996) Principles of Biochemistry, Prentice Hall, Second Edition, p. 714-715.
3. Garrett and Grisham. (1999) Biochemistry, Saunders College Publishing, Second Edition, p. 356-359.
4. Sanders J.K.M., Hunter, B.K. (1990) Modern NMR Spectroscopy – a guide for chemists, Oxford University Press, Fifth Edition, p. 93-160.
5. White, D.O., Fenner, F.J. (1994) *Herpesviridae*, In Medical Virology, Academic Press, 1994; Fourth Edition, p. 317-329.

Chapter 3: SEQUENCE ANALYSIS OF HSV-2 STRAINS: 333, 392, 414, 427, 443, 471, 472, 519, 564, HSV-2 AND MS.

Section 3.1 Specific Aims

We report the determination of DNA sequences of the following HSV-2 strains: 333, 414, 427, 443, 471, 519, 564, HSV2, 392, 472 and MS. The sequence and genetic content of each strain will be analyzed and compared with each other. Genes encoding PAT will be investigated, as well as the number of PAT repeats in each genome. The structure of a PAT trimer will also be analyzed using two dimensional correlational spectroscopy (COSY).

Section 3.2 Materials and Methods

NMR Sample Preparation.

The PAT trimer was synthesized and HPLC purified by Quality Controlled Biochemicals, a division of BioSource International. The solid was reconstituted in 200μL phosphate buffer, pH 5.2, and retrieved from solution by freeze-drying in a lyophalizer. The sample was then dissolved in 1 mL deionized water. Dissolved $(PAT)_3$ is placed in an NMR tube, ready for data collection.

Proton sensitive NMR reagents and 40% dioxane (carbon-sensitive) were used to calibrate the system. The "Glide" command was used for data acquisition, selecting the proper solvent system (D_2O) and running a ^1H and COSY.

DNA purification and Sequencing.

The following viral strains were purified using the QIAGEN kit—333, 392, 414, 427, 443, 471, 472, 519, 564, HSV-2 AND MS.

The QIAquick PCR purification protocol was followed: 5 volumes of PB buffer was added per volume of genomic DNA (PCR sample). The samples were placed in

QIAquick spin columns and centrifuged for 1 minute. The buffer was collected in a tube and discarded. The tube was reused for second wash with PE buffer. 0.75 mL of PE buffer was used to wash the sample, which was then centrifuged for one minute. The QIAquick column was then placed in a 1.5 ml microcentrifuge tube and 50 µl of buffer EB (10 mM Tris.Cl, PH 8.5) was then added to elute the purified DNA.

2.0 µL 10X sequencing buffer, 1.0 µL dNTP mix, and 2.0 µL each of the dye terminators (ddNTP's) were mixed into a 0.2 mL thin-walled tube. DNA polymerase in glycerol was centrifuged to consolidate pellet. 1.0 µL of the polymerase was added along with 0.5µL of the PCR-purified template DNA. The primer was added at an approximate ratio of 4:1 (primer: template). For GC-rich sequences the CEQ-DTCS quick start kit was later used. In a 0.2 mL thin-walled tube, 0.5 µL of template DNA was mixed with 2.0 µL of the appropriate HS2 primer (forward, reverse or middle), 8.0 µL of premix buffer and 9.5 µL of sterilized deionized water The contents of the tube were mixed thoroughly, centrifuged, and placed in thermal cycler. A polymerase chain reaction occured in the cycler, with a denaturing temperature of 96°C. The annealing temperature was set at 55°C, and the extension temperature was set at 60°C. The samples were denatured and annealed for 20 seconds each, and the extension temperature was held for 4 minutes. Each cycle was followed by holding the sample at 4°C for a total of 40 cycles.

A sterile 0.5 mL microfuge was labeled for each sample. For each row (8 samples), 10 µL of glycogen, 20 µL of 3M NaOAc pH 5.2, 16 µL of sterilized deionized water and 4 µL 0.5M EDTA, pH 7.0 were mixed in a centrifuge tube to make a stop solution. The stop solution was always prepared from the stock immediately before use.

For each trial, 5 μL of the stop solution was added to a clean 0.5 mL thick wall tube. Template DNA from produced from PCR was mixed with stop solution to end the polymerization reaction. 60 μL of 95% (v/v) cold ethanol was added to the mixture. The samples were mixed thoroughly using a vortexer, and centrifuged immediately at 13,000 rpm at 4°C for 15 minutes. Supernatant was removed using a micropipette, and the pellet was rinsed twice with 200 μL 70% ethanol (v/v) to remove residual salts. Immediately after each rinse, sample was centrifuged at 13,000 rpm at 4°C for 4 minutes, and supernatant was discarded. Pellets were vacuum-dried for 45 minutes, and resuspended (vortex) in 40 μL sample loading solution. The resuspended samples are transferred into CEQ tubes, and overlaid with a drop each of mineral oil. Samples were then loaded into the CEQ2000 sequencer.

50 fmol of template is needed for polymerase chain reaction (PCR). The templates are doubly purified PCR products and are approximately 66.7 ng/μL. The following was used to determine the volume of template to add for optimum signal strength:

$$V(\mu L) = (649)(10^{-6})(1000)(fmol)/conc.$$

Where 649 is the approximate amount of base pairs in double stranded DNA
1000 is the approximate amount of base pairs in the template DNA
fmol = 50, and
concentration = 66.7 ng/μL

$$\text{So } V(\mu L) = (649)(10^{-6})(1000)(50)/66.7$$

$$= 0.49\ \mu L.$$

Template DNA was heated in water at 96°C for one minute in thermal cycler, and cooled to room temperature. All other components of the sequencing reaction were added and

mixed thoroughly using a pipet. Tubes are sealed and placed in thermal cycler at 60°C for 4 minutes for 40 cycles.

The sample was spun down by centrifuging at 13,000 rpm for 2 minutes. 5 μL of freshly prepared stop solution was added along with 60 μL of 95% ethanol. The tubes were sealed, and inverted several times to mix contents, and then stored at -20°C for 10 minutes. The samples were centrifuged at 4°C and 13,000 rpm for 10 minutes. The supernatant is discarded, and the pellet rinsed twice with 200 _L of cold 70% ethanol. The pellet is vacuum dried for 35 minutes, and resuspended in 40 _L of sample loading solution. A drop of mineral oil was used to overlay samples before samples were loaded into CEQ2000 sequencer.

Sequence Alignment

The HS2F primer reads the sequence in the coding direction, whereas the HS2R and HS2M primers read in the template direction. The sequences obtained from all primers are aligned using the FASTA format from the European Bioinformatics Institute's website. The sequences with the reverse and middle primers were reversed using the appropriate command from the Baylor College of Medicine search launcher. Using the same program, reverse and middle sequences were reverse-complemented in order to get their corresponding 5' → 3' format. These sequences were then aligned with the forward sequences and multiple sequence alignments were observed.

Section 3.3 Results and Interpretation

CEQ data are chromatograms with color codes corresponding to each of the 4 nucleotides – A, C, T and G (See Appendix). The intensity and distinction of each peak illustrates resolution. The following comprise data, which are relatively well resolved and have aligned to some degree with primers from different regions of the template. The following alignment data have shaded regions, which emphasize aligned bases.

Section 3.4 References

1. European Bioinformatics Institute. Retrieved July 9, 2002, from http://www.ebi.ac.uk/clustalw/
2. BCM Search Launcher—Baylor College of Medicine Human Genome Sequencing Center. Retrieved July 9, 2002, from http://www.searchlauncher.bcm.tmc.edu/

Sample chromatograms showing base sequences for (a) 427F, (b) 427R and (c) 427M were obtained using default sequence analysis parameters.

Total length of capillary = 33.0cm

Length of capillary to detector = 30.0cm

Internal diameter of capillary = 75.0mm

Number of runs = 11

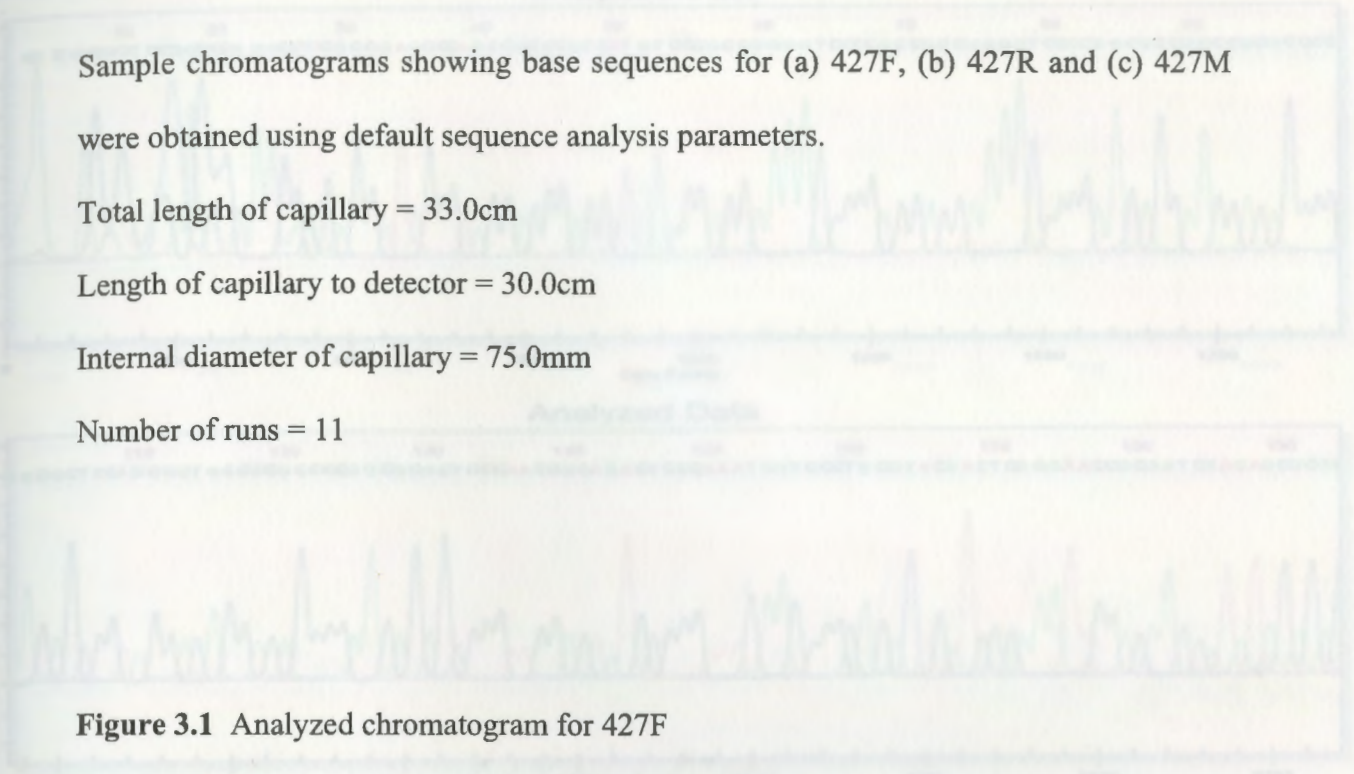
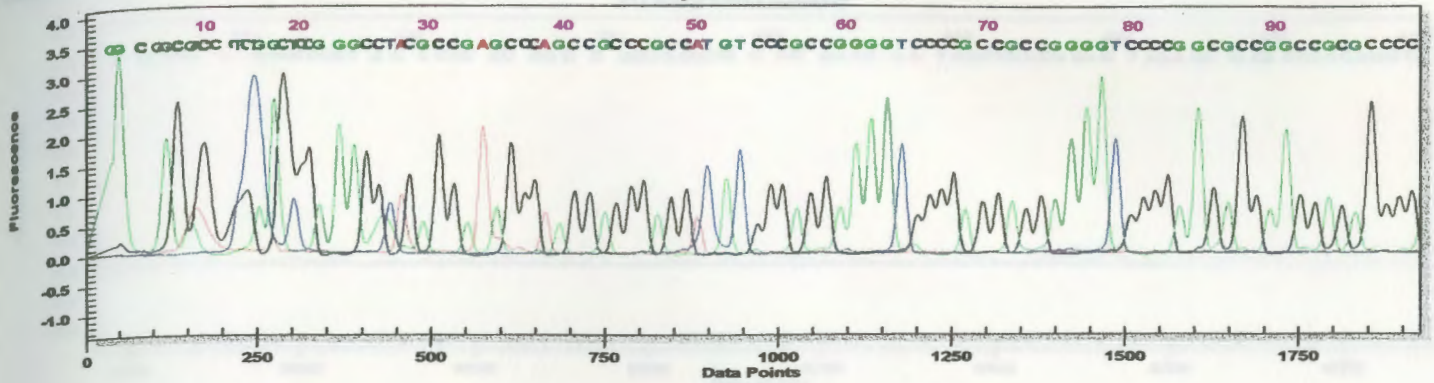
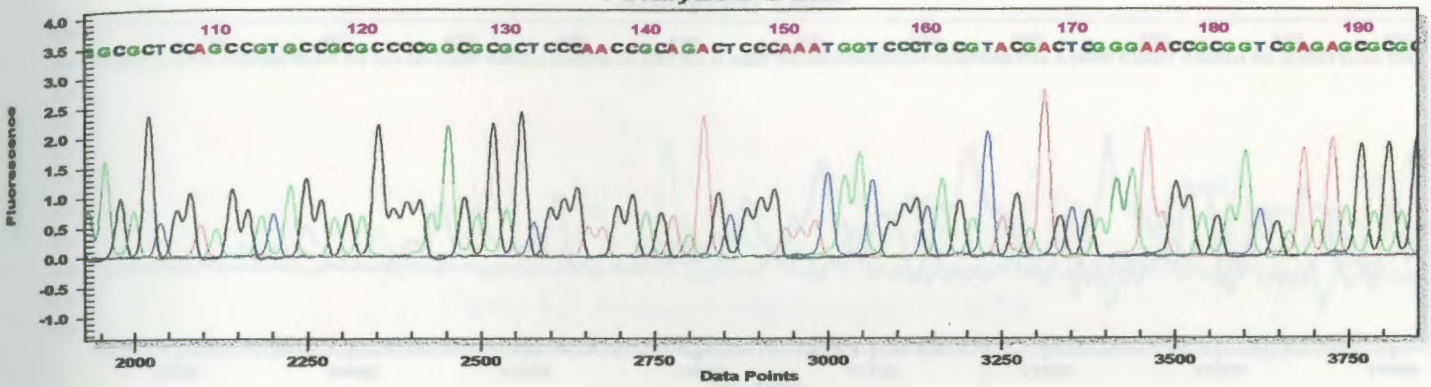


Figure 3.1 Analyzed chromatogram for 427F

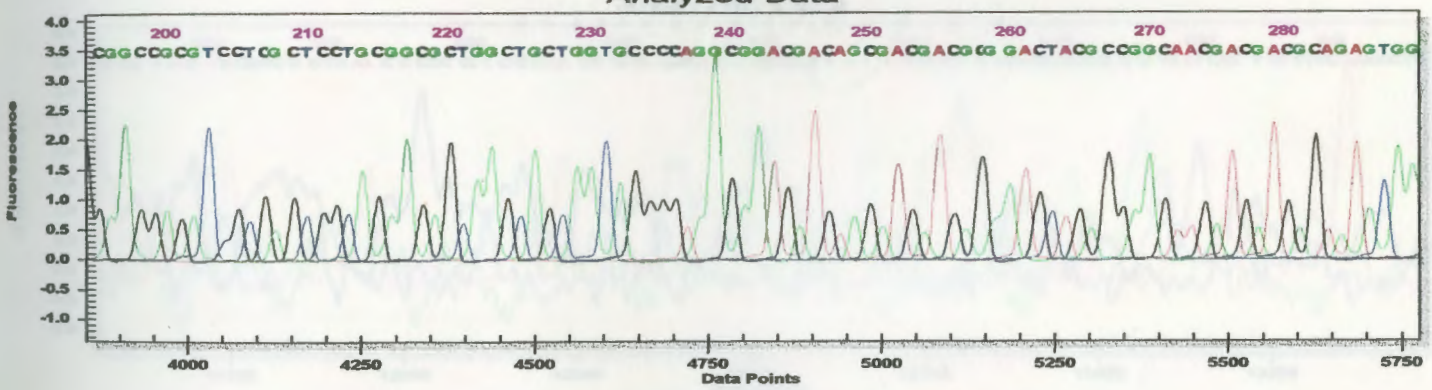
Analyzed Data



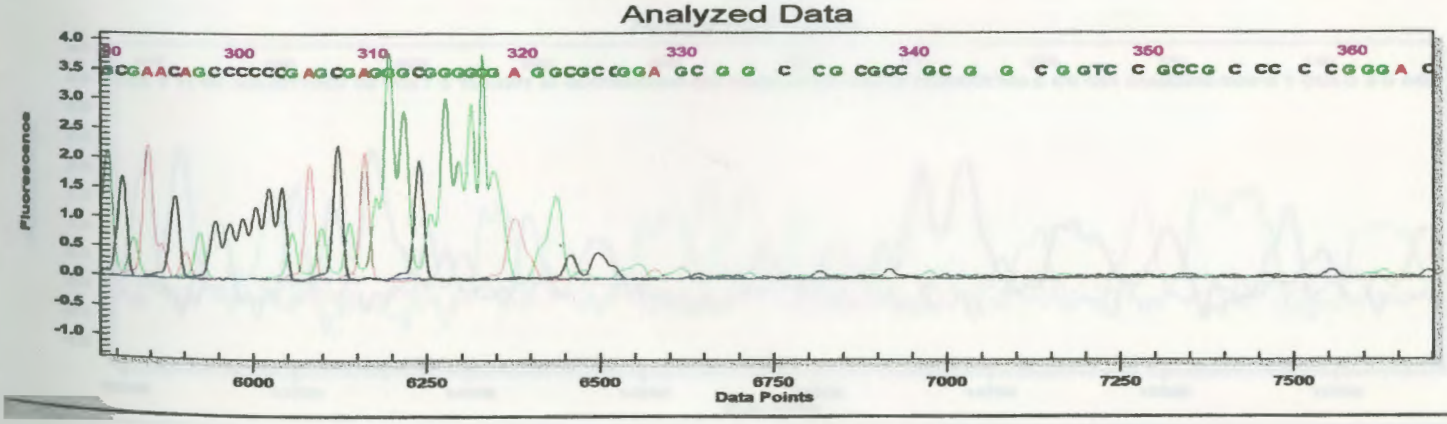
Analyzed Data



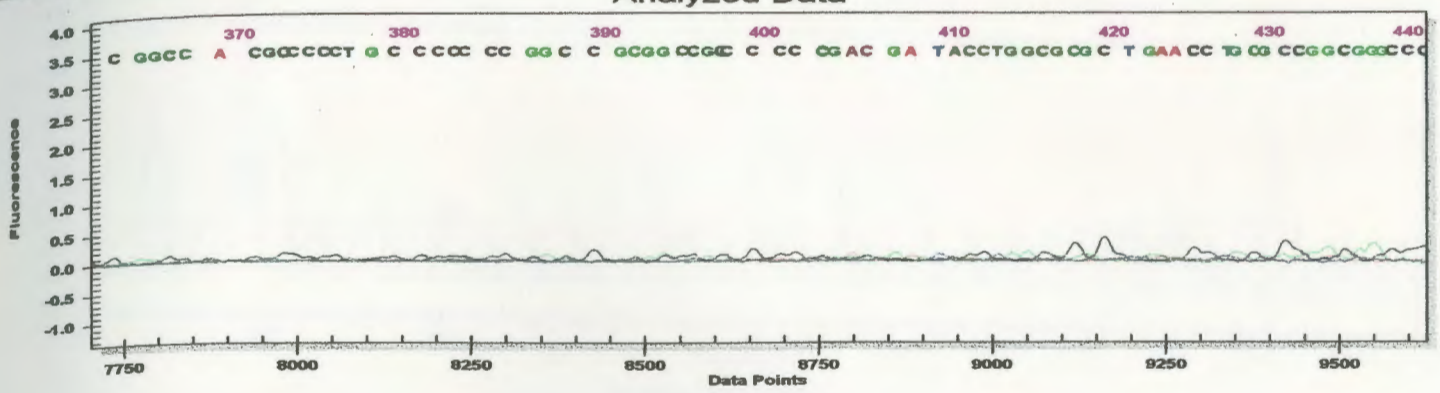
Analyzed Data



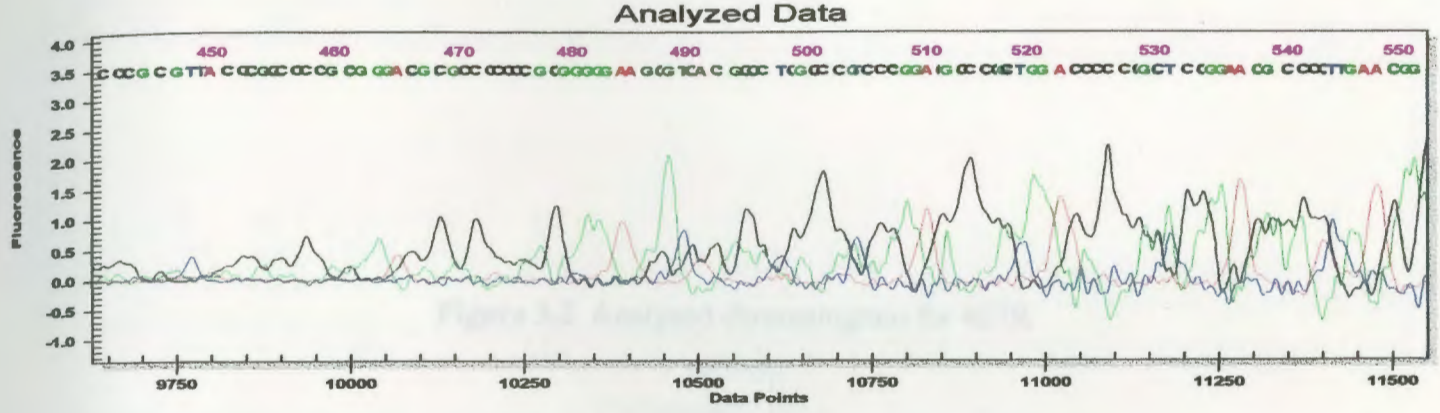
Analyzed Data



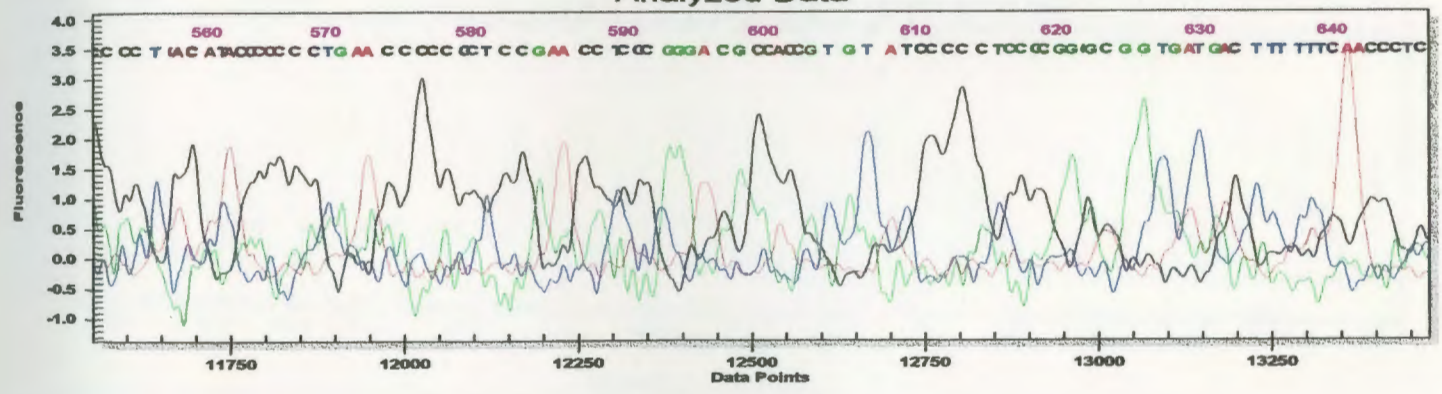
Analyzed Data



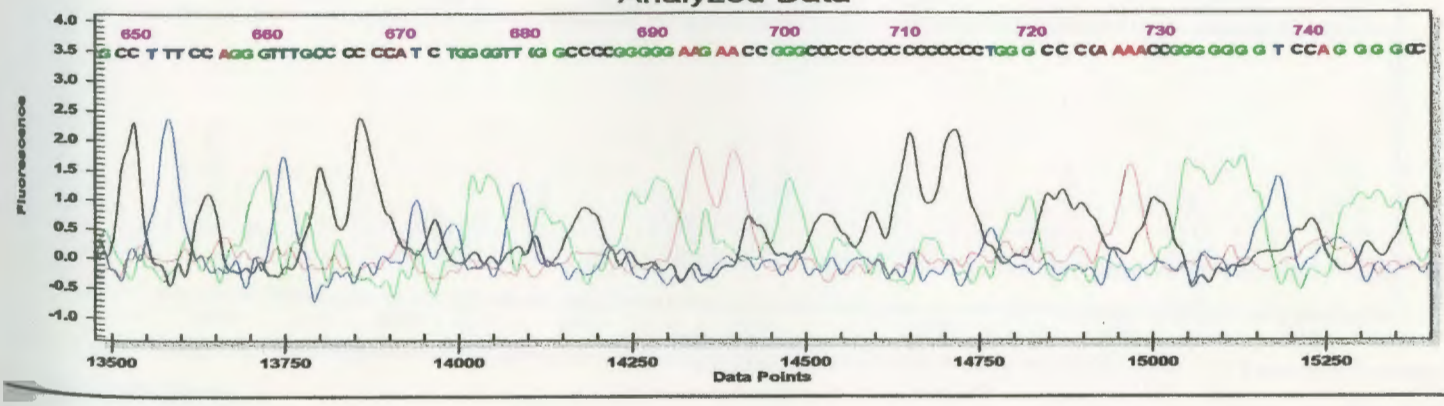
Analyzed Data



Analyzed Data



Analyzed Data



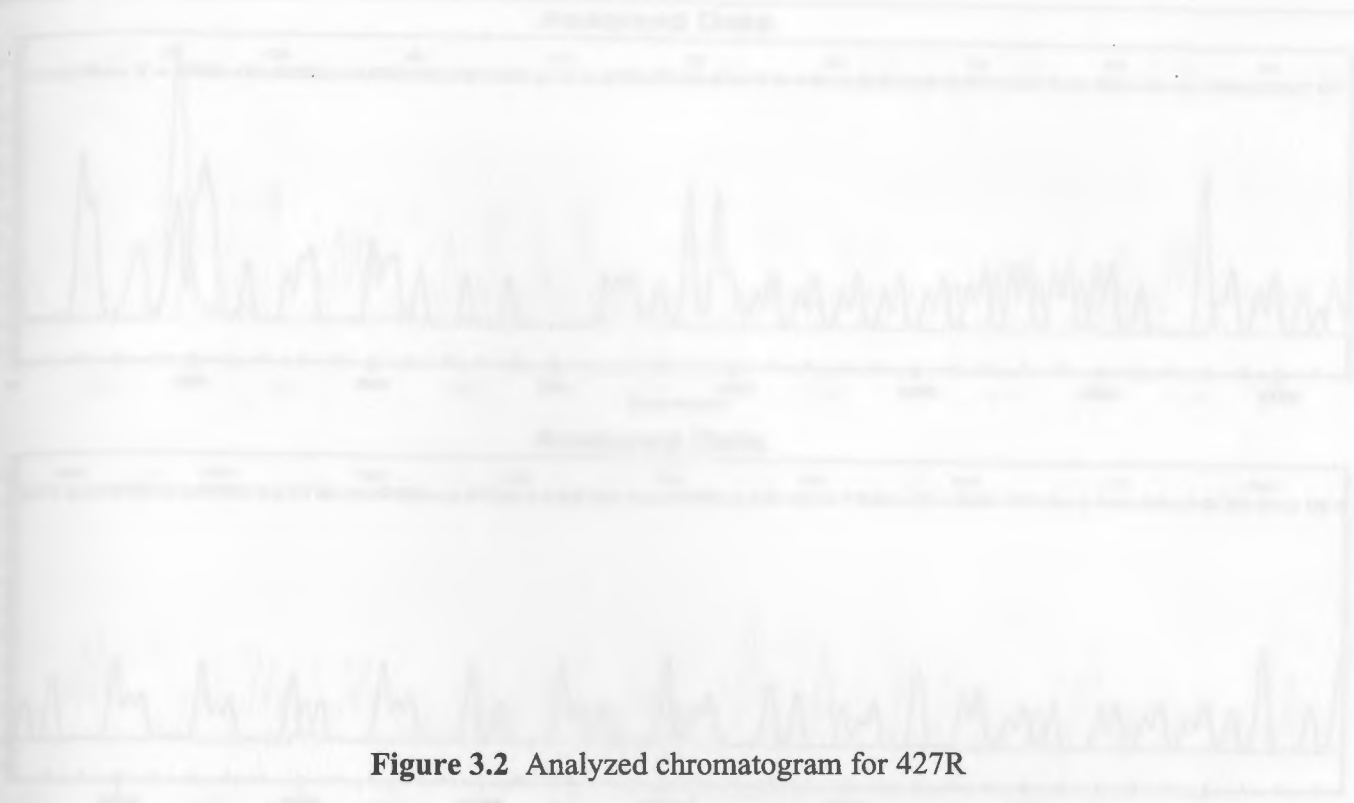
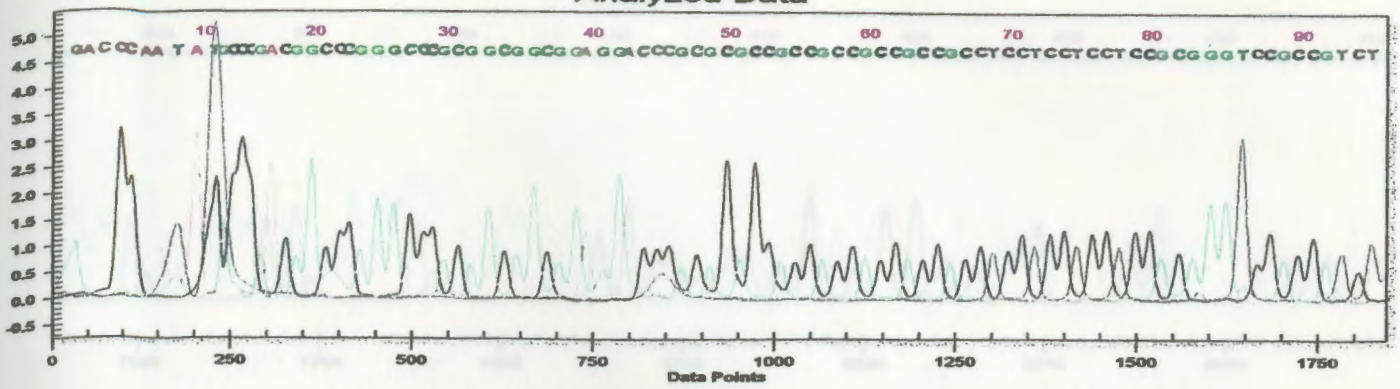
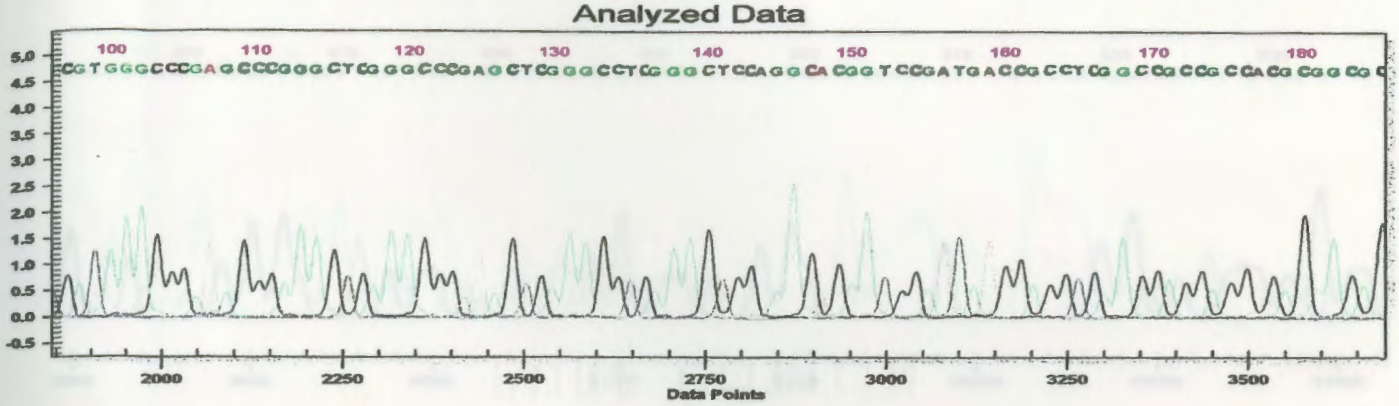


Figure 3.2 Analyzed chromatogram for 427R

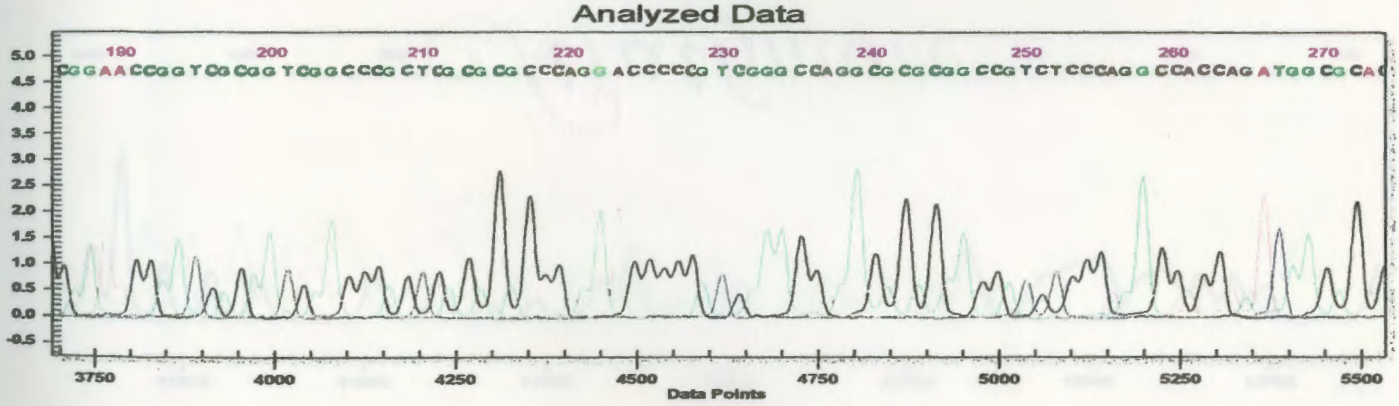
Analyzed Data



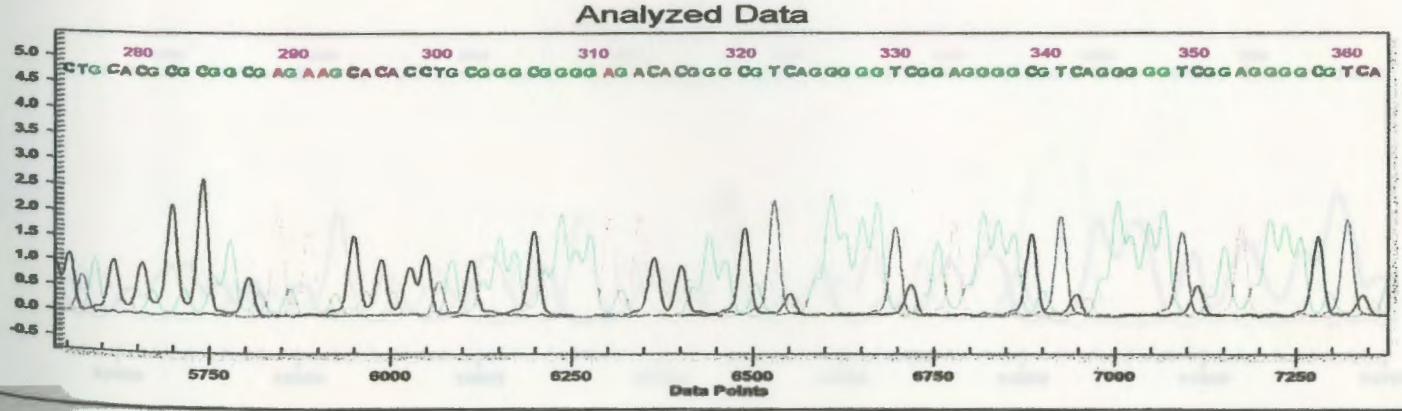
Analyzed Data



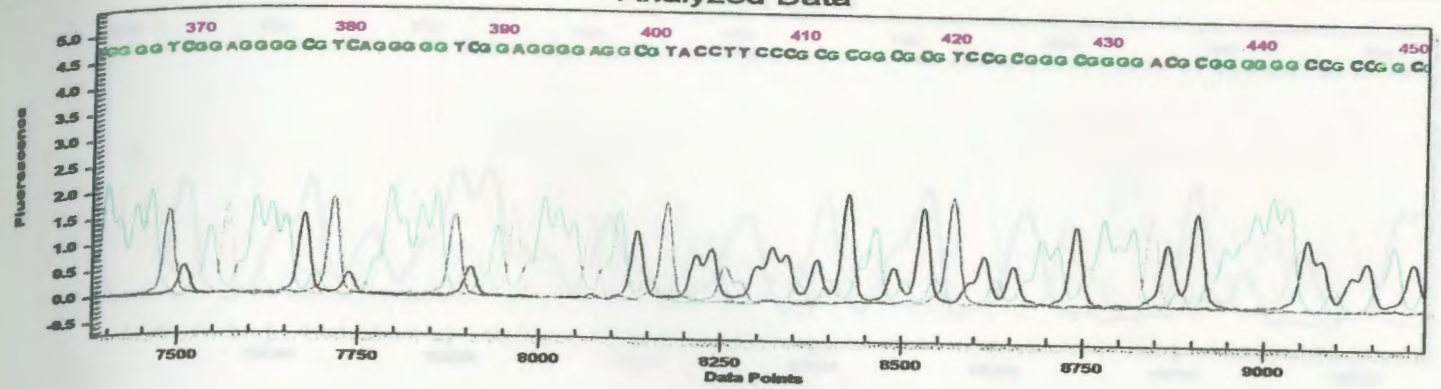
Analyzed Data



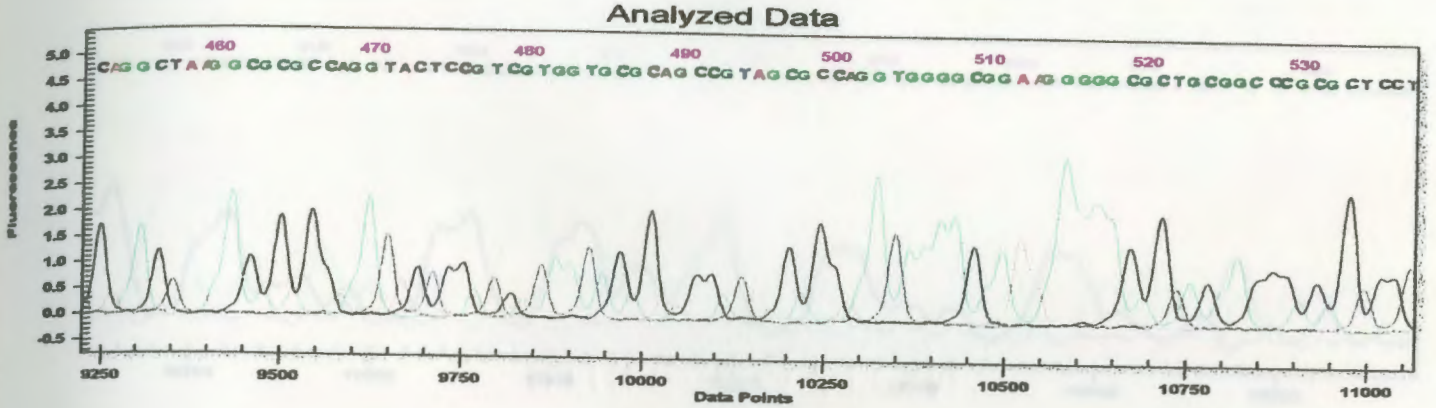
Analyzed Data



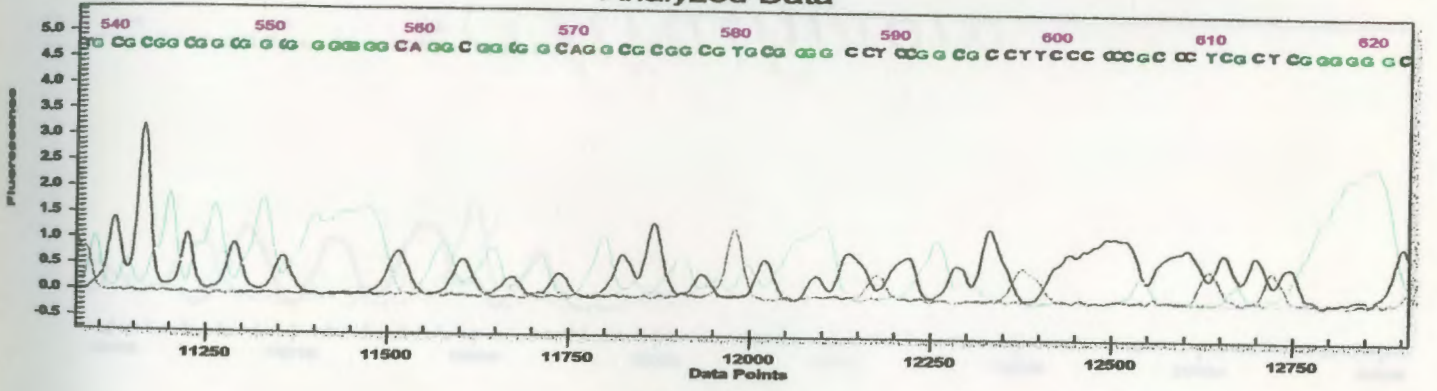
Analyzed Data



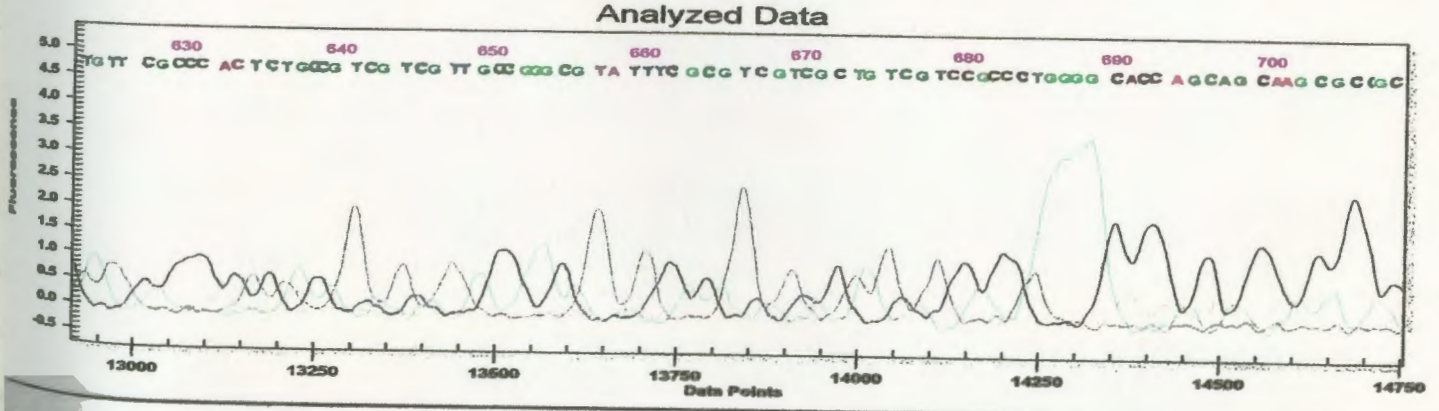
Analyzed Data



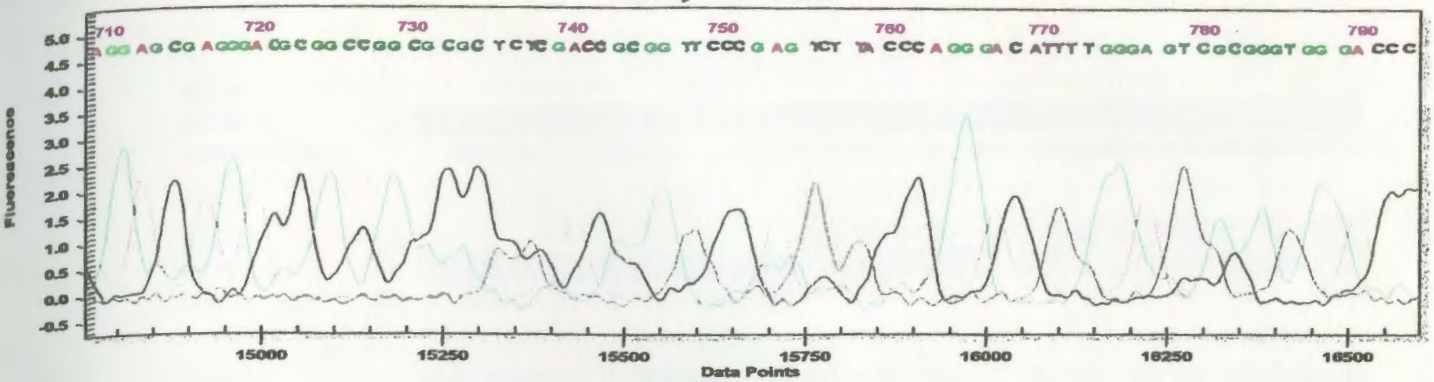
Analyzed Data



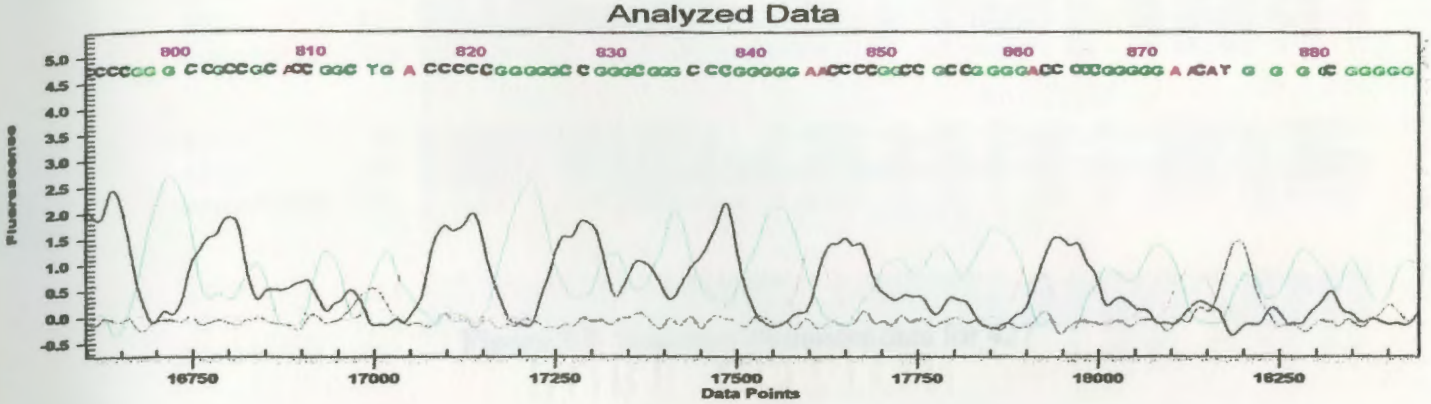
Analyzed Data



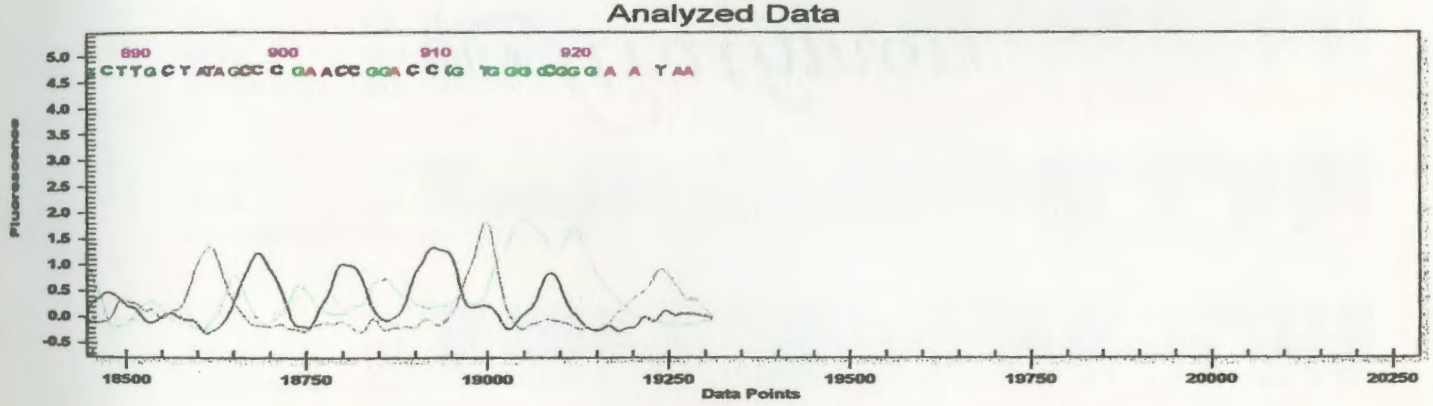
Analyzed Data



Analyzed Data



Analyzed Data



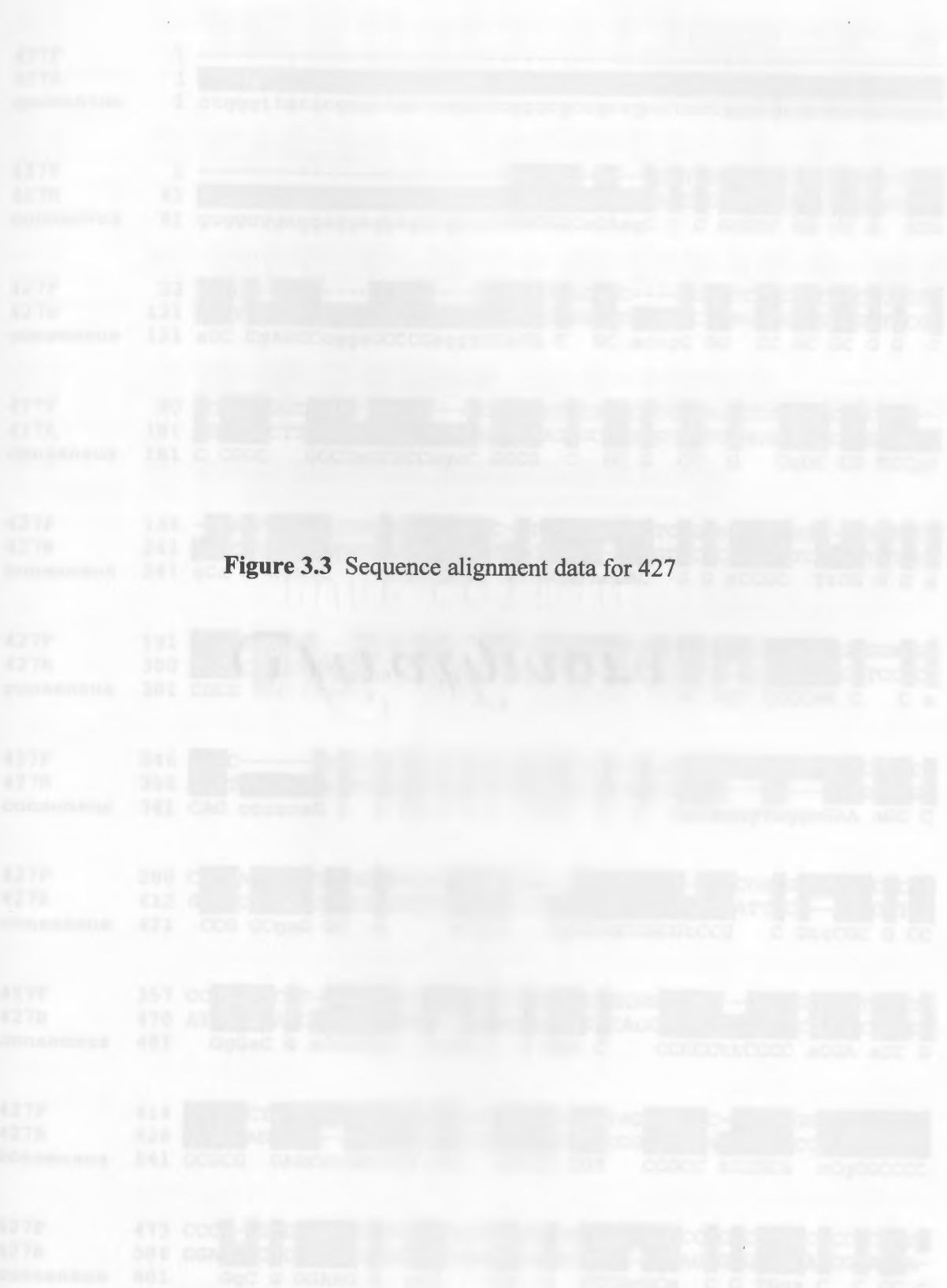
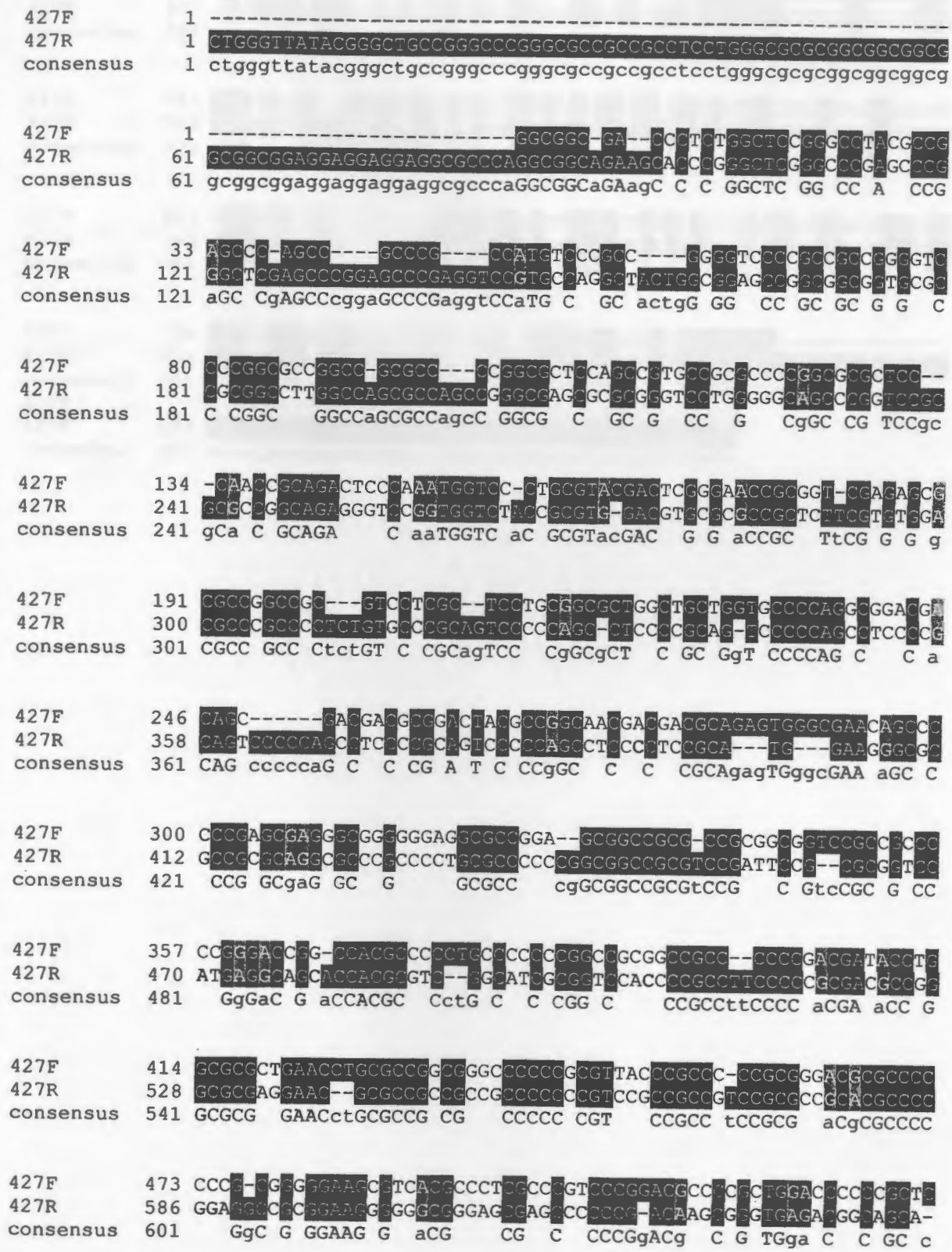


Figure 3.3 Sequence alignment data for 427


427F 1 ----------
427R 1 CTGGGTTATACGGGCTGCCGGGCCCGGGCGCCGCCGCCTCCTGGGCGCGCGGCGGCGGCG
consensus 1 ctgggttatacgggctgccgggcccgggcgccgccgcctcctgggcgcgcggcggcggcg
427F 1 ----------GGCGGC-GA--CCCTCTGGCTCCGGGCCTACGCCG
427R 61 GCGGCGGAGGAGGAGGAGGCGCCCAGGCGGCAGAAGCAGCCGGGGCTGGGCCCGAGCCCG
consensus 61 gcggcggaggaggaggaggcgcccaGGCGGCaGAagC C C GGCTC GG CC A CCG
427F 33 AGCCC-AGCC----GCCCG----CCATGTCCGGC----GGGGTCCCCGCCGCCGGGGGTC
427R 121 GGCTCGAGCCCGGAGCCCGAGGTCCGTGCCAGGGTACTGGCGCAGCCGGGGGCGGTGCGC
consensus 121 aGC CgAGCCcggaGCCCGaggtCCaTG C GC actgG GG CC GC GC G G C
427F 80 CCCGGCGCCGGCC-GCGCC---CCGGCGCTCCAGCCGTGCCGCGCCCCGGGCGCGCTCC--
427R 181 CGCGGCCTTGGCCAGCGCCAGCCGGGGCGAGCGCGCGGGTCCTGGGGGCAGGCCGGTCCGC
consensus 181 C CGGC GGCCaGCGCCagcC GGCG C GC G CC G CgGC CG TCCgc
427F 134 -CAACCGCAGACTCCCAAATGGTCC-CTGCGTACGACTCGGCAACCGCGGGT-CGAGACCG
427R 241 GCGCCGGCAGAGGGTCCGGTGGTCTACCGCGTG-GACGTGCGCCGCCGTCTTCGTGTGGA
consensus 241 gCa C GCAGA C aaTGGTC aC GCGTacGAC G G aCCGC TtCG G G g
427F 191 CGCCGGCCGC---GTCCTCGC--TCCTGCGGCGCTGGCTGCTGGTGCCCCAGGCGGACGG
427R 300 CGCCCGCCCCTCTGTGCCCGCAGTCCCCCAGC-CTCCCCGCAG-TCCCCCAGCCTCCCCG
consensus 301 CGCC GCC CtctGT C CGCagTCC CgGCgCT C GC GgT CCCCAG C C a
427F 246 CAGC------GACGACGCGGACTACGCCGGCAACGACGACGCAGAGTGGGCGAACAGCCG
427R 358 CAGTCCCCCAGCGTCCCCGCAGTCCCCCAGCCTCCCTCCGCA---TG---GAAGGGCGC
consensus 361 CAG cccccaG C C CG A T C CCgGC C C CGCAgagTGggcGAA aGC C
427F 300 CCCGAGCGAGGCGGGGGGAGGCGCCGGA--GCGGCCGCG-CCGCGGCGGTCCGCCGCCC
427R 412 GCCGCGCAGGCGCCCGCCCCTGCGCCCCCGGCGGCCGCGTCCGATTCCG--CGCGGTCC
consensus 421 CCG GCgaG GC G GCGCC cgGCGGCCGCGtCCG C GtcCGC G CC
427F 357 CCGGGACCGG-CCACGCCCCCTCCCCCCGGCCGCGGCCGCC--CCCCGACGATACCTG
427R 470 ATGAGGCAGCACCACGGGTG--GGCATCGCGGTCCACCCGCCTTCCCCCGCGACGCCGG
consensus 481 GgGaC G aCCACGC CctG C C CGG C CCGCCttCCCC aCGA aCC G
427F 414 GCGCGCTGAACCTGCGCCCGGCGGGCCCCCCGCGTTACCCGCCC-CCGCGGGACGCGCCCC
427R 528 GCGCGAGGAAG--GCGCCGCCGCCGCCCCCCCGTCCGCCGCCGTCCGCGCCGCCACGCCCC
consensus 541 GCGCG GAACctGCGCCG CG CCCCC CGT CCGCC tCCGCG aCgCGCCCC
427F 473 CCCC-CGGGGGAAGCGTCACGCCCTCGCCCGTCCCGGACGCCCCGCTGGACCCCCCGCTC
427R 586 GGAAGGCGCGGAAGGGGGCGGGAGCGAGCCCCCG-ACAAGGGGTGAGAGGGAGGA-
consensus 601 GgC G GGAAG G aCG CG C CCCGgACg C G TGga C C GC c

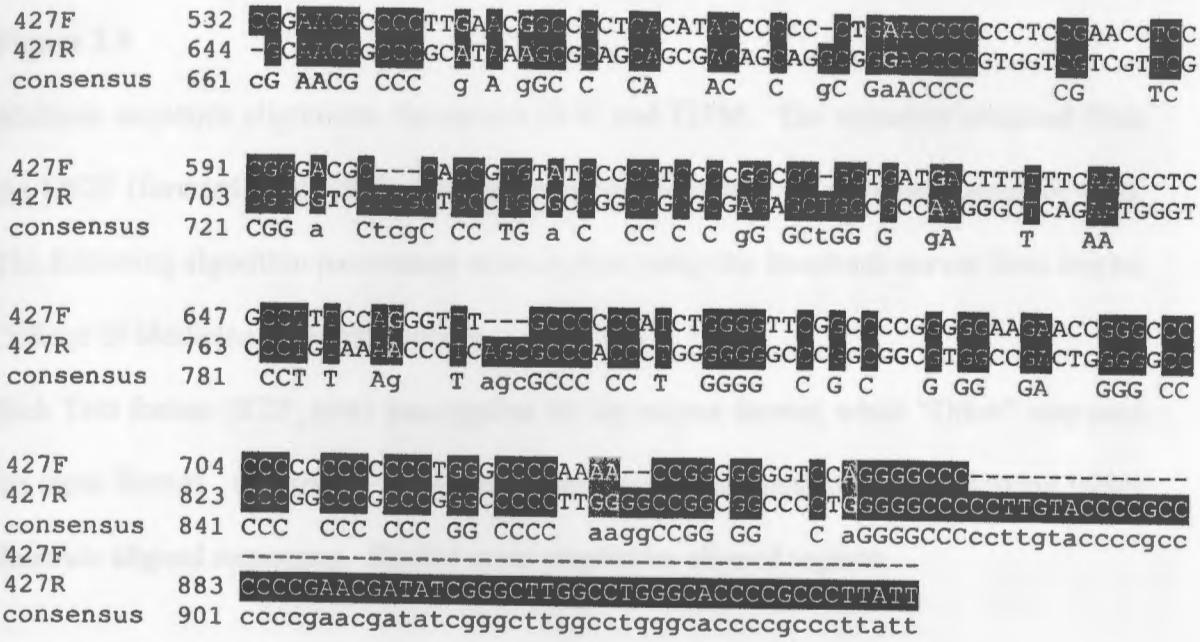
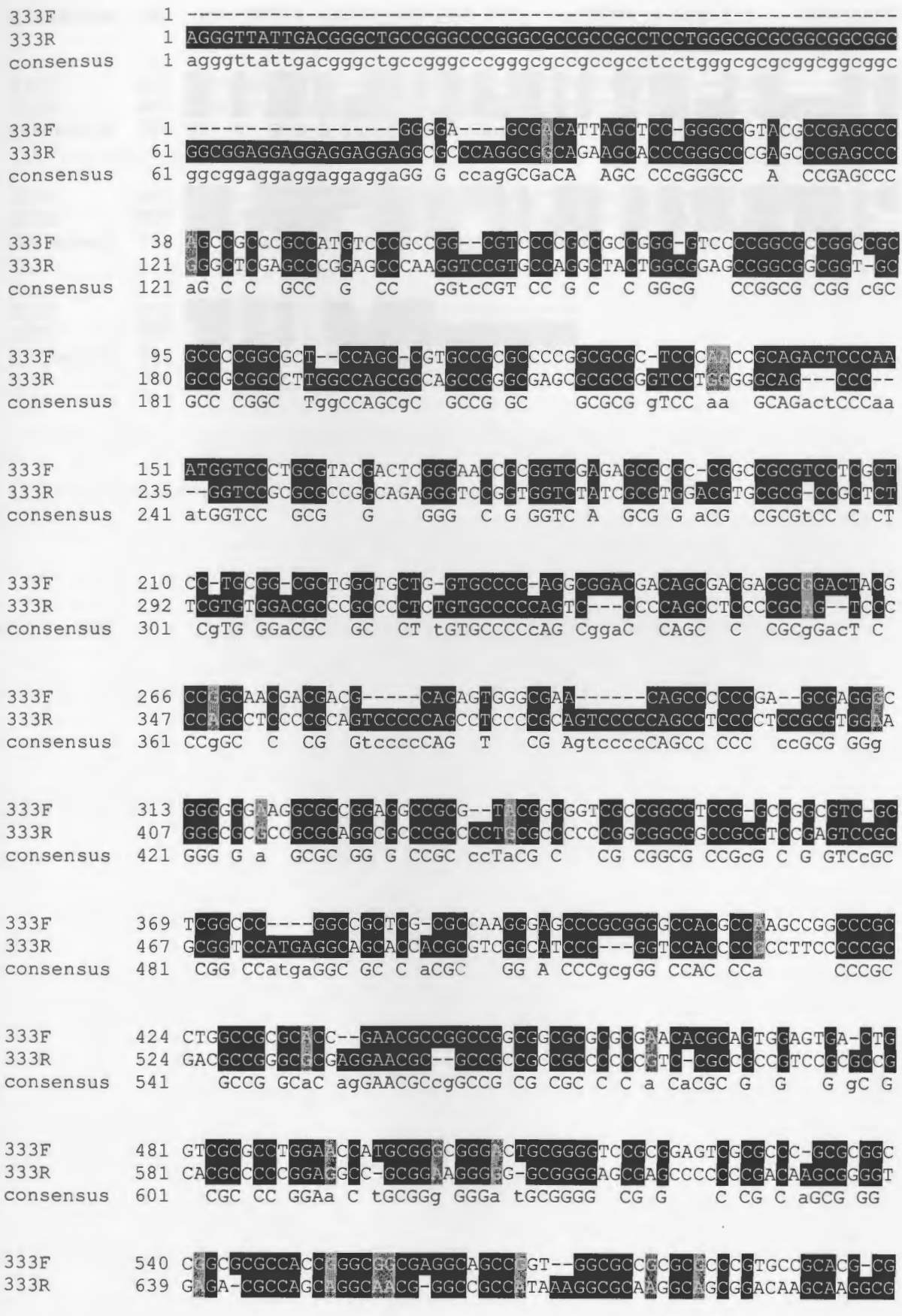

427F 532
427R 644
consensus 661 cG AACG CCC g A gGC C CA AC C gC GaACCCC CG TC
427F 591
427R 703
consensus 721 CGG a CtcgC CC TG a C CC C C gG GCtGG G gA T AA
427F 647
427R 763
consensus 781 CCT T Ag T agcGCCC CC T GGGG C G C G GG GA GGG CC
427F 704
427R 823
consensus 841 CCC CCC CCC GG CCCC aaggCCGG GG C aGGGGCCCccttgtaccccgcc
427F
427R 883 CCCCGAACGATATCGGGCTTGGCCTGGGCACCCCGCCCTTATT
consensus 901 ccccgaacgatatcgggcttggcctgggcaccccgcccttatt

Figure 3.4

Multiple sequence alignments for strains 333F and 333M. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 45%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.

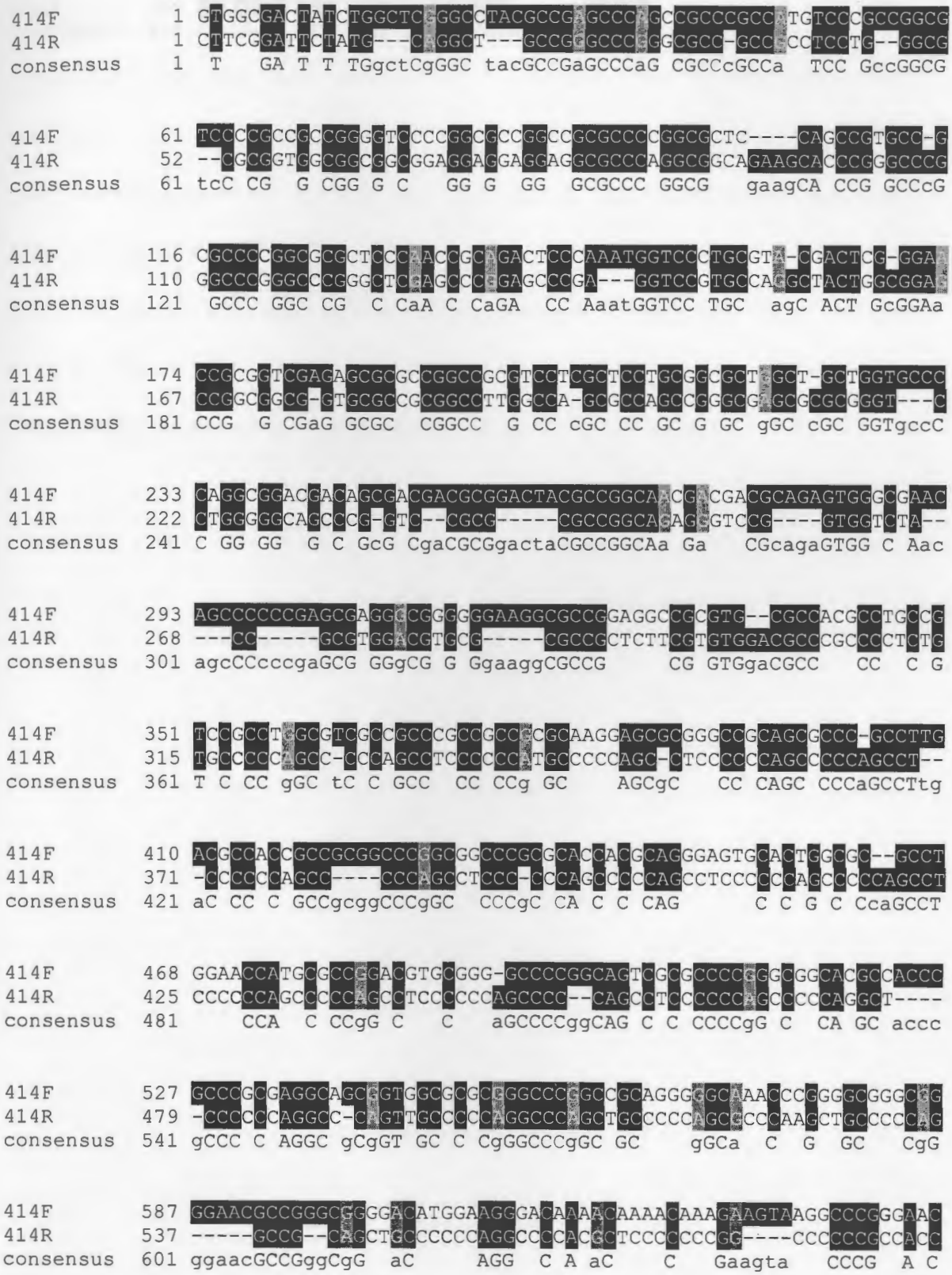

333F 1 ----------
333R 1 AGGGTTATTGACGGGCTGCCGGGCCCGGGCGCCGCCGCCTCCTGGGCGCGCGGCGGCGGC
consensus 1 agggttattgacgggctgccgggcccgggcgccgccgcctcctgggcgcgcggcggcggc
333F 1 ----------GGGGA----GCGGCATTAGCTCC-GGGCCGTACGCCGAGCCC
333R 61 GGCGGAGGAGGAGGAGGAGGCGCCCAGGCGGCAGAAGCACCCGGGCCCGAGCCCGAGCCC
consensus 61 ggcggaggaggaggaggaGG G ccagGCGaCA AGC CCcGGGCC A CCGAGCCC
333F 38 AGCCGCCCGCCATGTCCCGCCGG--CGTCCCCGCCGCGCGGG-GTCCCCGGCGCCGGCCGC
333R 121 CGGCTCGAGCCCGGAGCCCAAGGTCCGTGCCAGGCTACTGGCGGAGCCGGCGGCGGT-GC
consensus 121 aG C C GCC G CC GGtcCGT CC G C C GGcG CCGGCG CGG cGC
333F 95 GCCCCGGCGCT--CCAGC-CGTGCCGCGCCCCGGCGCGC-TCCCAACCGCAGACTCCCAA
333R 180 GCCGCGGCCTTGGCCAGCGCCAGCCGGGCGAGCGCGCGGGTCCTCCGGGCAG---CCC--
consensus 181 GCC CGGC TggCCAGCgC GCCG GC GCGCG gTCC aa GCAGactCCCaa
333F 151 ATGGTCCCTGCGTACGACTCGGGAACCGCGGTCGAGAGCGCGC-CGGCCGCGTCCTCGCT
333R 235 --GGTCCGCGCGCCGGCAGAGGGTCCGGTGGTCTATCGCGTGGACGTGCGCG-CCGCTCT
consensus 241 atGGTCC GCG G GGG C G GGTC A GCG G aCG CGCGtCC C CT
333F 210 CC-TGCGG-CGCTGGCTGCTG-GTGCCCC-AGGCGGACGACAGCGACGACGCCGACTACG
333R 292 TCGTGTGGACGCCCGCCCCTCTGTGCCCCCAGTC---CCCCAGCCTCCCCGCCG--TCCC
consensus 301 CgTG GGaCGC GC CT tGTGCCCCcAG CggaC CAGC C CGCgGacT C
333F 266 CCCGCAACGACGACG-----CAGAGTGGGCGAA-------CAGCCCCCGGA--GCGAGGCC
333R 347 CCAGCCTCCCCGCAGTCCCCCAGCCTCCCCGCAGTCCCCCAGCCTCCCCTCCGCGTGGCA
consensus 361 CCgGC C CG GtccccCAG T CG AgtccccCAGCC CCC ccGCG GGg
333F 313 GGGGGGAAGGCGCCCGGAGGCCGCG--TACGGCGGTCGCCGGCGTCCG-GCCGGCGTC-GC
333R 407 GGGCGCGCCGCGCAGGCGCCCGCCCCTCGCCCCCGGCGGCGGCCGCGTCCGAGTCCGC
consensus 421 GGG G a GCGC GG G CCGC ccTaCG C CG CGGCG CCGcG C G GTCcGC
333F 369 TCGGCCC----GGCCGCTCG-CGCCAAGGGAGCCCGCGGGGCCACGCCCAAGCCGGCCCGC
333R 467 GCGGTCCATGAGGCAGCACCACGCGTCGGCATCCC---GGTCCACCCCCCCTTCCCCCGC
consensus 481 CGG CCatgaGGC GC C aCGC GG A CCCgcgGG CCAC CCa CCCGC
333F 424 CTGGCCGCGCCACC--GAACGCCGGCCGGCGGCGCGCGAACACGCAGTGGAGTGA-CTG
333R 524 GACGCCGGGCGCCGAGGAACGC--GCCGCCGCCGCCCCCGTC-CGCCGCCGTCCGCGCCG
consensus 541 GCCG GCaC agGAACGCcgGCCG CG CGC C C a CaCGC G G G gC G
333F 481 GTCGCGCCTGGAACCATGCGGGCGGGCCTGCGGGGGTCCGCGGAGTCGCGCCC-GCGCGGC
333R 581 CACGCCCCCGGAGGCC-GCGGAAGGGGG-GCGGGGAGCGAGCCCCCCCGACAAGCGGGGT
consensus 601 CGC CC GGAa C tGCGGg GGGa tGCGGGG CG G C CG C aGCG GG
333F 540 CGGCGCGCCACCGGCGCCGAGGCAGCCGGT--GGCGCCGCGCCCCGTGCCGCACG-CG
333R 639 GAGA-CGCCAGCAGGCAACG-GGCCGCCATAAAGGCGCAAGGCAGCGGACAAGCAAGGCG

```
consensus   661  gG gCGCCA CgGGCggCGaGGC GCCg  aaGGCGC g GCg C G    GCA GgCG

333F        597 GGCACAACCGCGC--CGGCCCGGGCA-CCGCCTGGACCGCGAA--CTATG-----GGACC
333R        697 GGACCCCCGTGGTTCGGTCGGGTCGGCCGGCGGTCCTCGCCTCCCTTGCGCCCGGCCC
consensus   721 GG  C  CCG G ttCGG C GG CagCCG C G  CC CGa tcC  TGcgcccGG CC

333F        647 GCCGG-GGAGAGCCCCTGGCCGATCACGCCTGGAGACCCCGGTAATCGCGCTCCCGCGCG
333R        757 GCCGCCGGAGAGC--TTGGGCGC---CCCAAGGGGCTCCAAGCAATGGCG-TCCCCCT--
consensus   781 GCCG cGGAGgGCcc TGG CGatcaC C  GGaG  CC gG AgT GCGcTCCC C cg

333F        706 CGGGGAGTAGTTCC--CAAAA----------
333R        809 -GGGTAAAACCCCCTCCAGAACGCCCCAACCCC
consensus   841 cGGG Ag A   CCtcCAaAAcgccccaacccc
```

Figure 3.5

Multiple sequence alignments for strains 414F and 414R. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 37%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.


414F 1 GTGGCGACTATCTGGCTCGGGCCTACGCCGAGCCCAGCCGCCCGCCATGTCCCGCCGGCG
414R 1 CTTCGGATTCTATG---CAGGCT---GCCGGGCCCCGGCGCC-GCCCCTCCTG--GGCG
consensus 1 T GA T T TGgctCgGGC tacGCCGaGCCCaG CGCCcGCCa TCC GccGGCG
414F 61 TCCCCGCCGCCGGGGTCCCCGGCGCCGGCCGCGCCCCGGCGCTC----CAGCCGTGCC-G
414R 52 --CGCGGTGGCGGCGGCGGAGGAGGAGGAGGCGCCCAGGCGGCAGAAGCACCCGGGCCCG
consensus 61 tcC CG G CGG G C GG G GG GCGCCC GGCG gaagCA CCG GCCcG
414F 116 CGCCCCGGCGCGCTCCCAACCGCAGACTCCCAAATGGTCCCTGCGTA-CGACTCG-GGA
414R 110 GGCCCGGGCCCGGGCTCCAGCCCCGAGCCCGA---GGTCCGTGCCAGCGTACTGGCGGAG
consensus 121 GCCC GGC CG C CaA C CaGA CC AaatGGTCC TGC agC ACT GcGGAa
414F 174 CCGCGGTCGAGAGCGCGCCGGCCGCGTCCTCGCTCCTGCGGCGCTGGCT-GCTGGTGCCC
414R 167 CCGGCGGCG-GTGCGCCGCGGCCTTGGCCA-GCGCCAGCCGGGGGAGCGCGGCGGGT---C
consensus 181 CCG G CGaG GCGC CGGCC G CC cGC CC GC G GC gGC cGC GGTgccC
414F 233 CAGGCGGACGACAGCGACGACGCGGACTACGCCGGCAACGACGACGCAGAGTGGGCGAAC
414R 222 CTGGGGGCAGCCCG-GTC--CGCG-----CGCCGGCAGAGGGTCCG----GTGGTCTA--
consensus 241 C GG GG G C GcG CgaCGCGgactaCGCCGGCAa Ga CGcagaGTGG C Aac
414F 293 AGCCCCCCGAGCGAGGGCGGGGGGAAGGCGCCGGAGGCCGCGTG--CGCCACGCCTGCCG
414R 268 ---CC-----GCGTGGACGTGCG-----CGCCGCTCTTCGTGTGGACGCCCGCCCCTCTG
consensus 301 agcCCcccgaGCG GGgCG G GgaaggCGCCG CG GTGgaCGCC CC C G
414F 351 TCCGCCTGGCGTCGCCGCCCGCCGCCGCCGCAAGGAGCGCGGGCCGCAGCGCCC-GCCTTG
414R 315 TGCCCCCAGCC-CCCAGCCTCCCCCCATGCCCCCAGC-CTCCCCCCAGCCCCCAGCCT--
consensus 361 T C CC gGC tC C GCC CC CCg GC AGCgC CC CAGC CCCaGCCTtg
414F 410 ACGCCACCGCCGCGGCCCGGCGGCCCGCGCACCACGCAGGGAGTGCACTGGCGC--GCCT
414R 371 -CCCCCCAGCC----CCCAGCCTCCC-CCCAGCCCCCAGCCTCCCCCCCAGCCCCCAGCCT
consensus 421 aC CC C GCCgcggCCCgGC CCCgC CA C C CAG C C G C CcaGCCT
414F 468 GGAACCATGCGCCGGACGTGCGGG-GCCCCGGCAGTCGCGCCCCGGGCGGCACGCCACCC
414R 425 CCCCCCAGCCCCCAGCCTCCCCCCAGCCCC--CAGCCTCCCCCCAGCCCCCAGGCT----
consensus 481 CCA C CCgG C C aGCCCCggCAG C C CCCCgG C CA GC accc
414F 527 GCCCGCGAGGCAGCGGTGGCGCGCCGGCCCCGGCCGCAGGGGGCAAACCCGGGGCGGGCGG
414R 479 -CCCCCCAGGCC-CAGTTGCCCCCCAGGCCCCGCTGCCCCCAGCGCCCCAAGCTGCCCCCAG
consensus 541 gCCC C AGGC gCgGT GC C CgGGCCCgGC GC gGCa C G GC CgG
414F 587 GGAACGCCGGGCGGGGACATGGAAGGGACAAAACAAAACAAAGAAGTAAGGCCCGGGAAC
414R 537 -----GCCG--CAGCTGCCCCCCAGGCCCCACGCTCCCCCCCGG----CCCCCCGCCACC
consensus 601 ggaacGCCGggCgG aC AGG C A aC C Gaagta CCCG A C

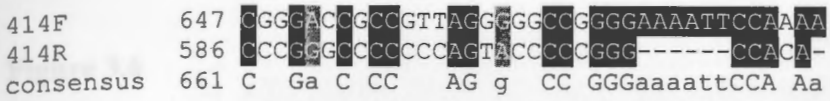
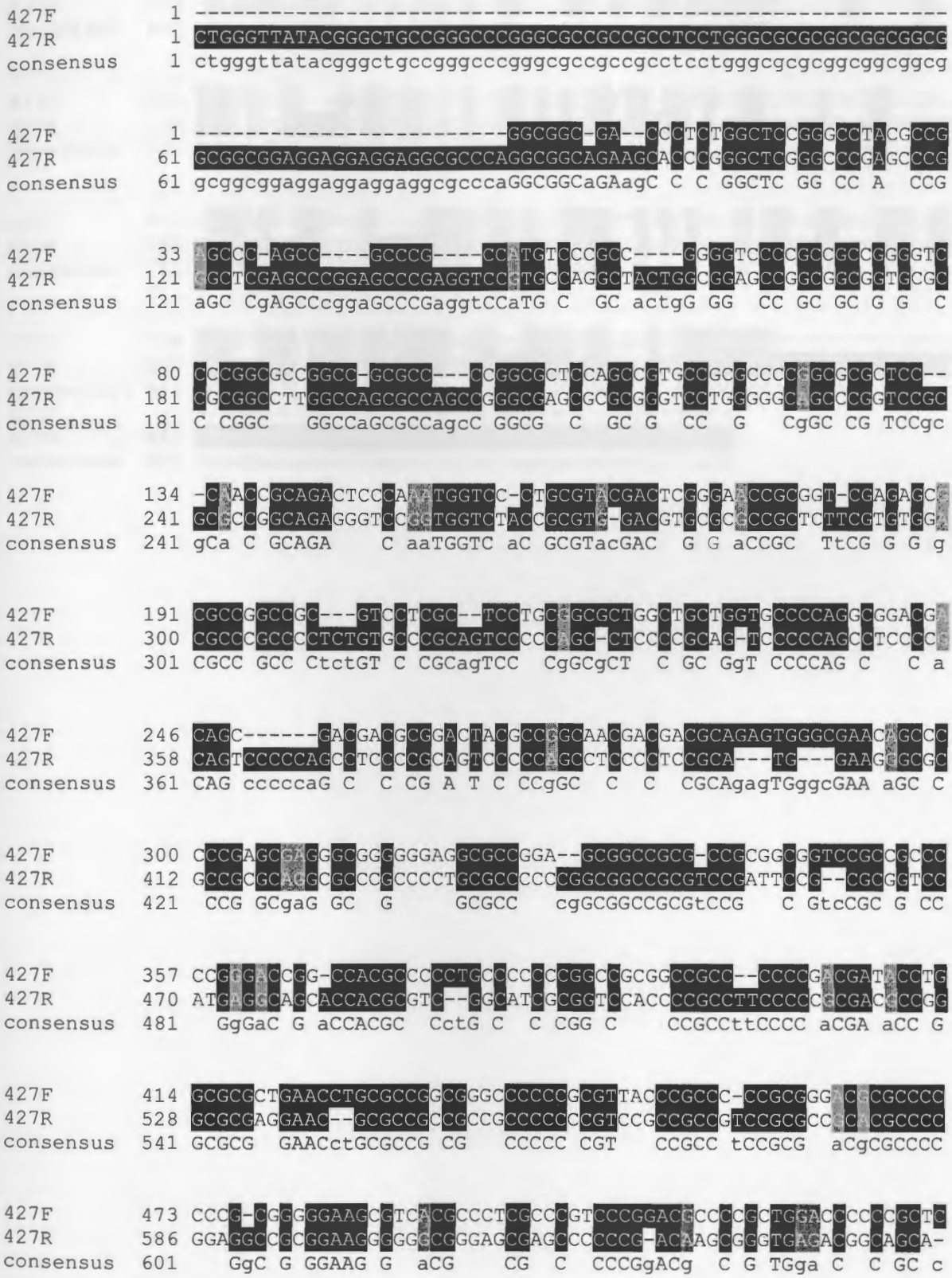

414F 647 CGGGACCGCCGTTAGGGGCCGGGGAAAATTCCAAA
414R 586 CCCGGGCCCCCCAGTACCCCGGG------CCACA-
consensus 661 C Ga C CC AG g CC GGGaaaattCCA Aa

Figure 3.6

Multiple sequence alignments for strains 427F and 427R. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 63%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.


427F 1 ----------
427R 1 CTGGGTTATACGGGCTGCCGGGCCCGGGCGCCGCCGCCTCCTGGGCGCGCGGCGGCGGCG
consensus 1 ctgggttatacgggctgccgggcccgggcgccgccgcctcctgggcgcgcggcggcggcg
427F 1 ----------GGCGGC-GA--CCCTCTGGCTCCGGGCCTACGCCG
427R 61 GCGGCGGAGGAGGAGGAGGCGCCCAGGCGGCAGAAGCACCCGGGCTCGGGCCCGAGCCCG
consensus 61 gcggcggaggaggaggaggcgcccaGGCGGCaGAagC C C GGCTC GG CC A CCG
427F 33 AGCCC-AGCC----GCCCG----CCATGTCCCGGC-----GGGGTCCCCGGCCGCCGGGGTC
427R 121 GGCTCGAGCCCGGAGCCCGAGGTCCGTGCCAGGCTACTGGCGGAGCCGGCGGCGGTGCGC
consensus 121 aGC CgAGCCcggaGCCCGaggtCCaTG C GC actgG GG CC GC GC G G C
427F 80 CCCGGCGCCGGCCC-GCGCC---CCGGCGCTCCAGCCGTGCCGCGCCCCCGGCGCGCTCC--
427R 181 CGCGGCCTTGGCCAGCGCCAGCCGGGCGAGCGCGCGGGTCCTGGGGGCAGCCCGGTCCGC
consensus 181 C CGGC GGCCaGCGCCagcC GGCG C GC G CC G CgGC CG TCCgc
427F 134 -CAACCGCAGACTCCCAAATGGTCC-CTGCGTACGACTCGGGACCGCGGT-CGAGAGCC
427R 241 GCCCCGGCAGAGGGTCCCGTGGTCTACCGCGTG-GACGTGCGCCCCGCTCTTCGTGTGGC
consensus 241 gCa C GCAGA C aaTGGTC aC GCGTacGAC G G aCCGC TtCG G G g
427F 191 CGCCGGCCGC---GTCCTCGC--TCCTGCCGCGCTGGCTGGTGGTGCCCCAGGCGGACGA
427R 300 CGCCCGCCCCTCTGTGCCCGCAGTCCCCCAGC-CTCCCCGCAG-TCCCCCAGCCTCCCCC
consensus 301 CGCC GCC CtctGT C CGCagTCC CgGCgCT C GC GgT CCCCAG C C a
427F 246 CAGC------GACGACGCGGACTACGCCGCAACGACGACGCAGAGTGGGCGAACAGCCC
427R 358 CAGTCCCCCAGCCTCCCCGCAGTCCCCCAGCCTCCCCTCCGCA---TG---GAAGCGCGC
consensus 361 CAG cccccaG C C CG A T C CCgGC C C CGCAgagTGggcGAA aGC C
427F 300 CCCGAGCGAGGGCGGGGGGAGGCGCCGGA--GCGGCCGCG-CCGCGGCGGTCCGCCGCCC
427R 412 GCCGCGCAGGCGCCCGCCCCTGCGCCCCCGGCGGCCGCGTCCGATTCCG--CGCGGTCC
consensus 421 CCG GCgaG GC G GCGCC cgGCGGCCGCGtCCG C GtcCGC G CC
427F 357 CCGTGACCGG-CCACGCCCCCTGCCCCCCGGCCGCGGCCGCC--CCCCGACGATACCTG
427R 470 ATGAGGCAGCACCACGCGTC--GGCATCGCGGTCCACCCCGCCTTCCCCCCGCGACGCCGG
consensus 481 GgGaC G aCCACGC CctG C C CGG C CCGCCttCCCC aCGA aCC G
427F 414 GCGCGCTGAACCTGCGCCGGCGGGCCCCCCGCGTTACCCGCCC-CCGCGGGACGCGCCCCC
427R 528 GCGCGAGGAAC--GCGCCGCCGCCGCCCCCCCGTCCGCCGCCGTCCGCGCCCCCCCCCCC
consensus 541 GCGCG GAACctGCGCCG CG CCCCC CGT CCGCC tCCGCG aCgCGCCCC
427F 473 CCCG-CGGGGGAAGCGTCACGCCCTCGCCCGTCCCGGACGCCCCGCTGGACCCCCCGGCTC
427R 586 GGAGGCCGCGGAAGGGGGGCGGGAGCGAGCCCCCCG-ACAAGCGGGTGACACGGCAGCA-
consensus 601 GgC G GGAAG G aCG CG C CCCGgACg C G TGga C C GC c

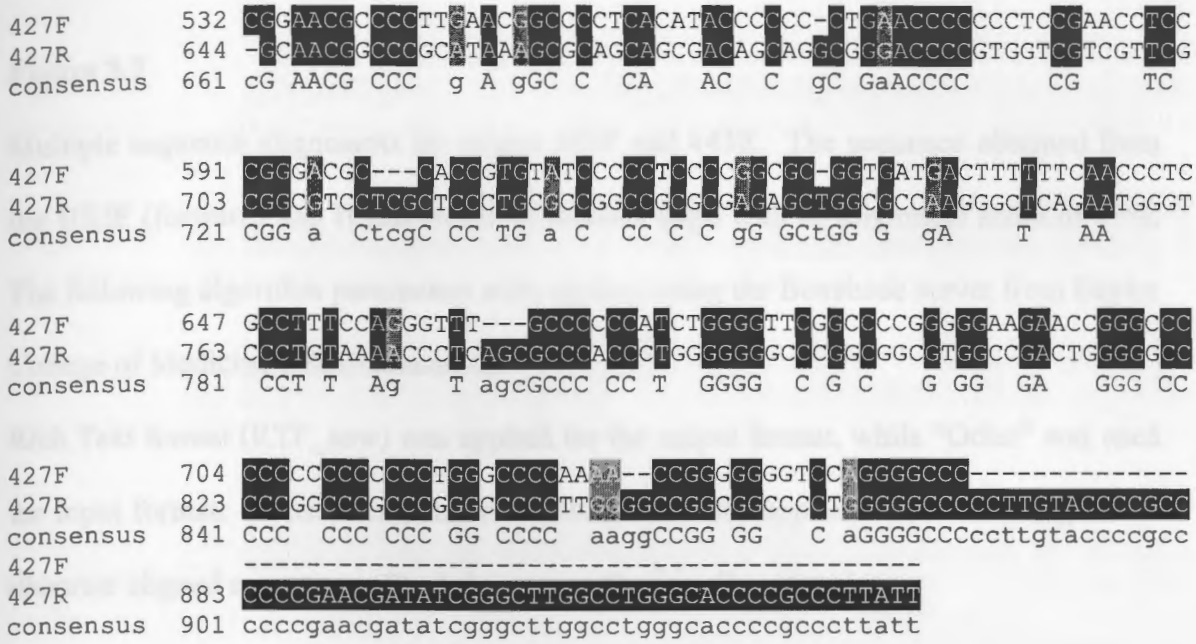
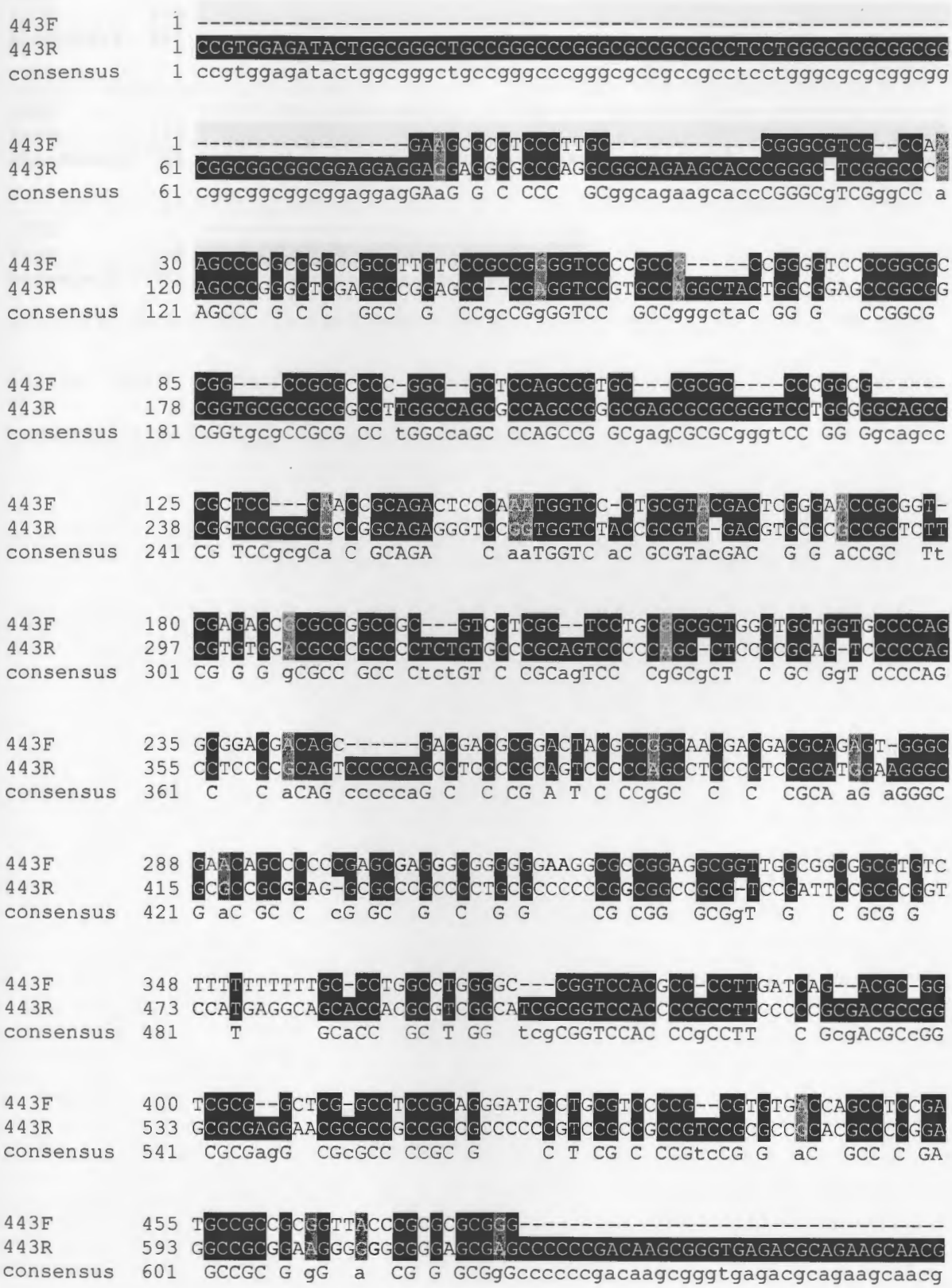

427F 532 CGGAACGCCCCTTGAACGGCCCTCACATACCCCC-CTGAACCCCCCCTCCGAACCTCC
427R 644 -GCAACGGCCCGCATAAAGCGCAGCAGCGACAGCAGGCGGGACCCCGTGGTCGTCGTTCG
consensus 661 cG AACG CCC g A gGC C CA AC C gC GaACCCC CG TC
427F 591 CGGGACGC---CACCGTGTATCCCCCTCCCCGGCGC-GGTGATGACTTTTTTCAACCCTC
427R 703 CGGCGTCCTCGCTCCCTGCGCCGGCCGCGCGAGAGCTGGCGCCAAGGGCTCAGAATGGGT
consensus 721 CGG a CtcgC CC TG a C CC C C gG GCtGG G gA T AA
427F 647 GCCTTTCCAGGGTTT---GCCCCCCATCTGGGGTTCGGCCCCGGGGGAAGAACCGGGCCC
427R 763 CCCTGTAAAACCCTCAGCGCCCACCCTGGGGGGGCCCGGCGGCGTGGCCGACTGGGGGCC
consensus 781 CCT T Ag T agcGCCC CC T GGGG C G C G GG GA GGG CC
427F 704 CCCCCCCCCCCCTGGGCCCCAAAA--CCGGGGGGGTCCAGGGGCCC----------
427R 823 CCCGGCCCGCCCGGGCCCCCTTGGGGCCGGCGGCCCCTGGGGGCCCCCTTGTACCCCGCC
consensus 841 CCC CCC CCC GG CCCC aaggCCGG GG C aGGGGCCCccttgtaccccgcc
427F ----------
427R 883 CCCCGAACGATATCGGGCTTGGCCTGGGCACCCCGCCCTTATT
consensus 901 ccccgaacgatatcgggcttggcctgggcaccccgcccttatt

Figure 3.7

Multiple sequence alignments for strains 443F and 443R. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 45%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.

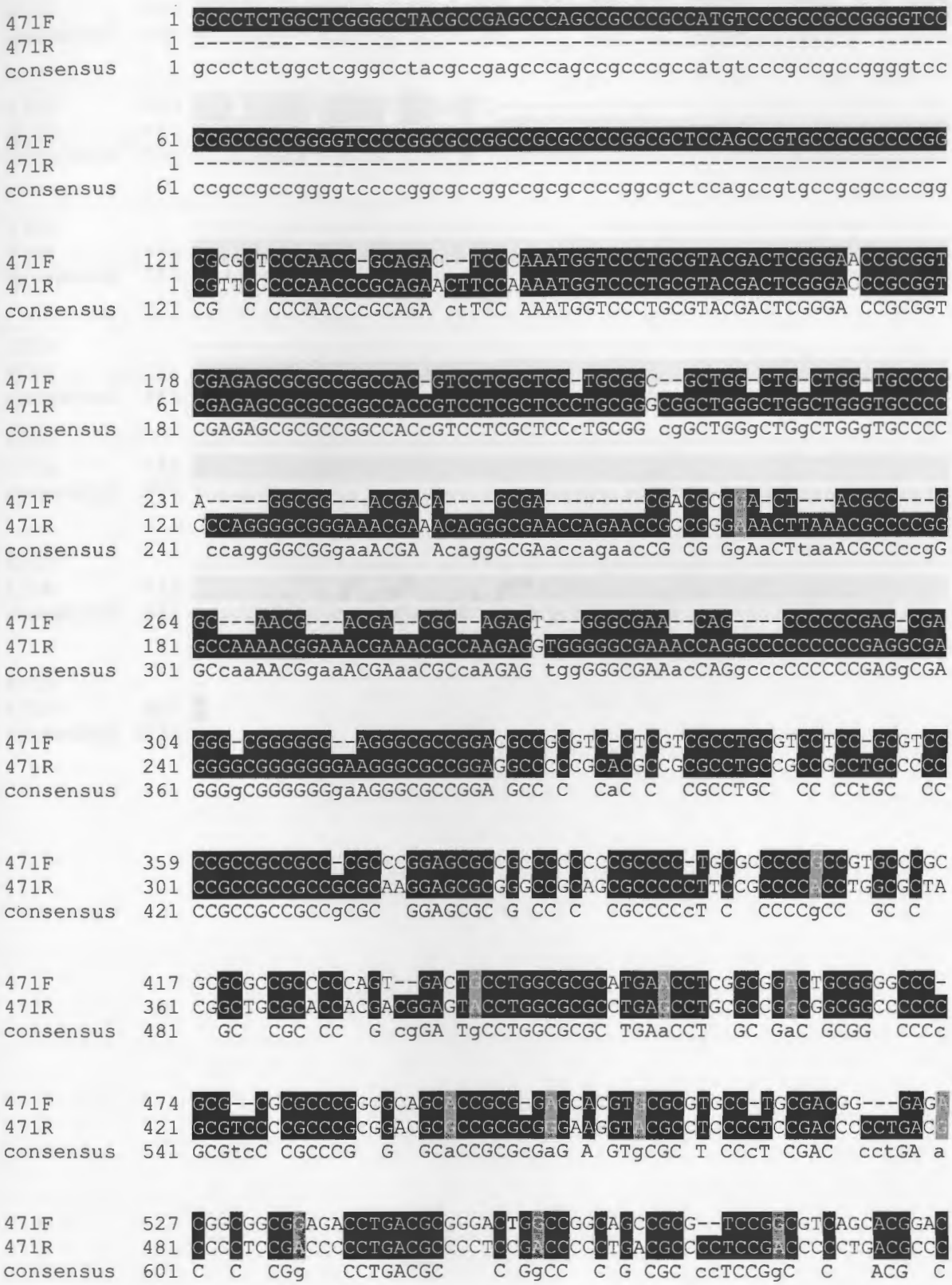

443F 1 ----------
443R 1 CCGTGGAGATACTGGCGGGCTGCCGGGCCCGGGCGCCGCCGCCTCCTGGGCGCGCGGCGG
consensus 1 ccgtggagatactggcgggctgccgggcccgggcgccgccgcctcctgggcgcgcggcgg
443F 1 ----------GAAGCGCCTCCCTTGC----------CGGGCGTCG--CCAA
443R 61 CGGCGGCGGCGGAGGAGGAGGAGGCGCCCAGGCGGCAGAAGCACCCGGGC-TCGGGCCC
consensus 61 cggcggcggcggaggagGAaG G C CCC GCggcagaagcaccCGGGCgTCGggCC a
443F 30 AGCCCCGCCGCCCGCCTTGTCCCGCCGCGGTCCCCGCCC-----CCGGGGTCCCCGGCGC
443R 120 AGCCCGGGCTCGAGCCCGGAGCC--CGAGGTCCGTGCCAGGCTACTGGCGGAGCCGGCGG
consensus 121 AGCCC G C C GCC G CCgcCGgGGTCC GCCgggctaC GG G CCGGCG
443F 85 CGG----CCGCGCCCC-GGC--GCTCCAGCCGTGC---CGCGC----CCCGGCG------
443R 178 CGGTGCGCCGCGGCCTTGGCCAGCGCCAGCCCGGCGAGCGCGCGGGTCCTGGGGGCAGCC
consensus 181 CGGtgcgCCGCG CC tGGCcaGC CCAGCCG GCgagCGCGCgggtCC GG Ggcagcc
443F 125 CGCTCC---CAACCGCAGACTCCCAATGGTCC-CTGCGTACGACTCGGGAACCGCGGT-
443R 238 CGGTCCGCGCGCCGGCAGAGGGTCCGCTGGTCTACCGCGTG-GACGTGCGCCCGCTCTT
consensus 241 CG TCCgcgCa C GCAGA C aaTGGTC aC GCGTacGAC G G aCCGC Tt
443F 180 CGAGAGCGCGCCGGCCGC---GTCCTCGC--TCCTGCGGCGCTGGCTGCTGGTGCCCCAG
443R 297 CGTGTGGACGCCCGCCCCTCTGTGCCCGCAGTCCCCCAGC-CTCCCCGCAG-TCCCCCAG
consensus 301 CG G G gCGCC GCC CtctGT C CGCagTCC CgGCgCT C GC GgT CCCCAG
443F 235 GCGGACGACAGC------GACGACGCGGACTACGCCGGCAACGACGACGCAGAGT-GGGC
443R 355 CCTCCCCGCAGTCCCCCAGCCTCCCCGCAGTCCCCCAGCCTCCCCTCCGCATGGAAGGGC
consensus 361 C C aCAG cccccaG C C CG A T C CCgGC C C CGCA aG aGGGC
443F 288 GAACAGCCCCCGAGCGAGGGCGGGGGGAAGGCGCCGGAGGCGGTTGGCGGCGGCGTGTC
443R 415 GCGCCGCGCAG-GCGCCCGCCCCTGCGCCCCCCGGCGGCCGCG-TCCGATTCCGCGCGGT
consensus 421 G aC GC C cG GC G C G G CG CGG GCGgT G C GCG G
443F 348 TTTTTTTTTTGC-CCTGGCCTGGGGC---CGGTCCACGCC-CCTTGATCAG--ACGC-GG
443R 473 CCATGAGGCAGCACCACGCGTCGGCATCGCGGTCCACCCCGCCTTCCCCCGCGACGCCGG
consensus 481 T GCaCC GC T GG tcgCGGTCCAC CCgCCTT C GcgACGCcGG
443F 400 TCGCG--GCTCG-GCCTCCGCAGGGATGCCTGCGTCCCCG--CGTGTGACCAGCCTCCGA
443R 533 GCGCGAGGAACGCGCCGCCGCCGCCCCCCGTCCGCCGCCGTCCGCGCCCCACGCCCCGGA
consensus 541 CGCGagG CGcGCC CCGC G C T CG C CCGtcCG G aC GCC C GA
443F 455 TGCCGCCGCGGTTACCCGCGCGCGGG----------
443R 593 GGCCGCGGAAGGGGGGCGGGAGCGAGCCCCCCGACAAGCGGGTGAGACGCAGAAGCAACG
consensus 601 GCCGC G gG a CG G GCGgGccccccgacaagcgggtgagacgcagaagcaacg

```
443F                ------------------------------------------------------------
443R           653  GCCGCATAAGGCGCAGCAGGGACAGAAGGCGGACCCCGTGTCGTTCGGTGGCGGAGTCCT
consensus      661  gccgcataaggcgcagcagggacagaaggcggaccccgtgtcgttcggtggcggagtcct

443F                ------------------------------------------------------------
443R           713  TGGTCCTGGGCCGGCCCGCGCGAGAGTTGGCCCCAAGGGTTAAAAAGGGGGCCCCTGGTAA
consensus      721  tggtcctgggccggcccgcgcgagagttggccccaagggttaaaaagggggccccctggtaa

443F                --------------------------------
443R           773  CCCCTAAACCCCCAACCCTGGGGGCCCGGCC
consensus      781  cccctaaacccccaaccctgggggcccggcc
```

Figure 3.8

Multiple sequence alignments for strains 471F and 471R. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 49%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.

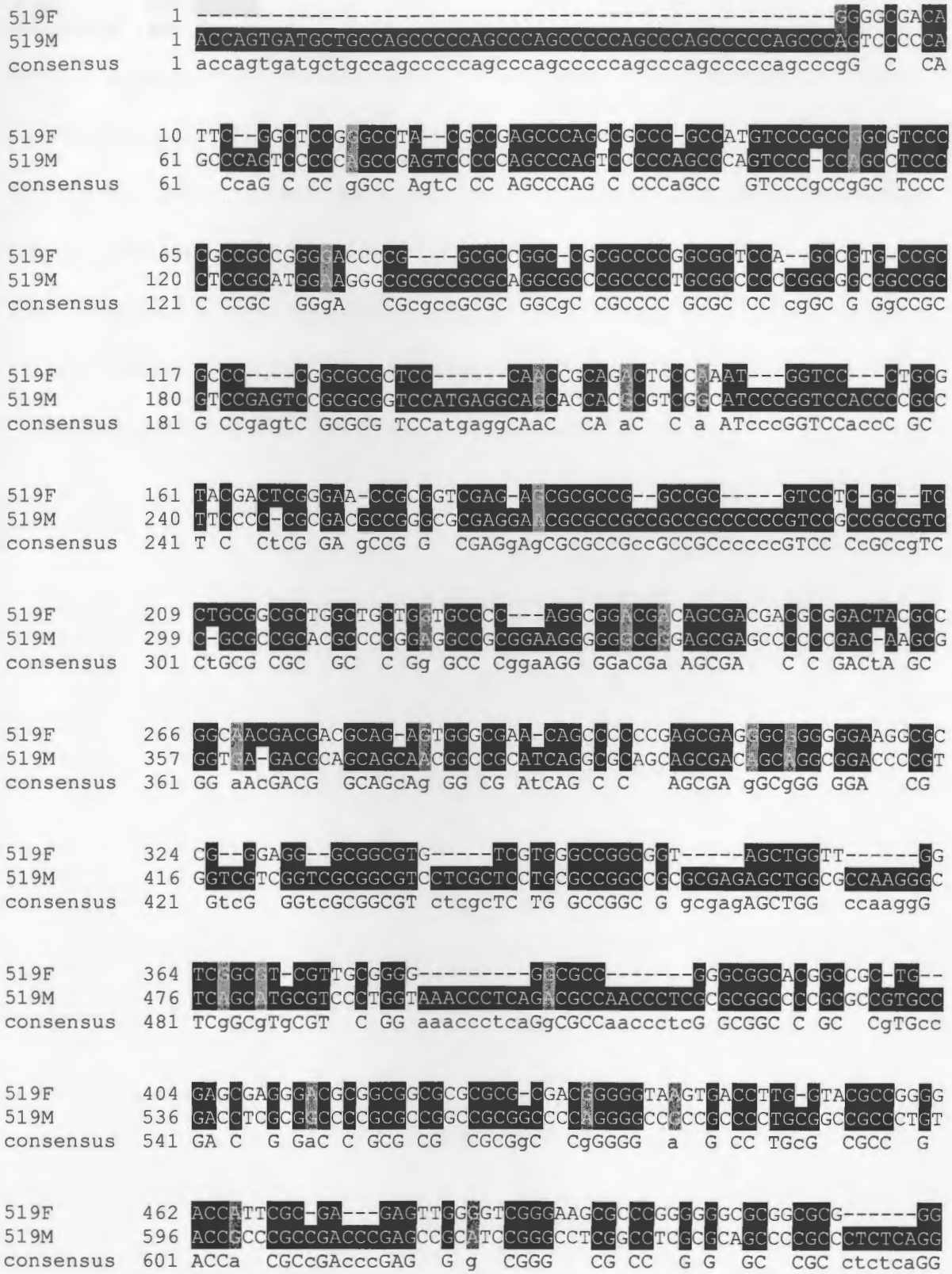

471F 1 GCCCTCTGGCTCGGGCCTACGCCGAGCCCAGCCGCCCGCCATGTCCCGCCGCCGGGGTCC
471R 1 ----------
consensus 1 gccctctggctcgggcctacgccgagcccagccgcccgccatgtcccgccgccggggtcc
471F 61 CCGCCGCCGGGGTCCCCGGCGCCGGCCGCGCCCCGGCGCTCCAGCCGTGCCGCGCCCCGG
471R 1 ----------
consensus 61 ccgccgccggggtccccggcgccggccgcgccccggcgctccagccgtgccgcgccccgg
471F 121 CGCGCTCCCAACC-GCAGAC--TCCCAAATGGTCCCTGCGTACGACTCGGGAACCGCGGT
471R 1 CGTTCCCCCAACCCGCAGAACTTCCAAAATGGTCCCTGCGTACGACTCGGGACCCGCGGT
consensus 121 CG C CCCAACCcGCAGA ctTCC AAATGGTCCCTGCGTACGACTCGGGA CCGCGGT
471F 178 CGAGAGCGCGCCGGCCAC-GTCCTCGCTCC-TGCGGC--GCTGG-CTG-CTGG-TGCCCC
471R 61 CGAGAGCGCGCCGGCCACCGTCCTCGCTCCCTGCGGGCGGCTGGGCTGGCTGGGTGCCCC
consensus 181 CGAGAGCGCGCCGGCCACcGTCCTCGCTCCcTGCGG cgGCTGGgCTGgCTGGgTGCCCC
471F 231 A-----GGCGG---ACGACA----GCGA--------CGACGCGGA-CT---ACGCC---G
471R 121 CCCAGGGGCGGGAAACGAAACAGGGCGAACCAGAACCGCCGGGAAACTTAAACGCCCCGG
consensus 241 ccaggGGCGGgaaACGA AcaggGCGAaccagaacCG CG GgAaCTtaaACGCCcgG
471F 264 GC---AACG---ACGA--CGC--AGAGT---GGGCGAA--CAG----CCCCCCGAG-CGA
471R 181 GCCAAAACGGAAACGAAACGCCAAGAGGTGGGGGCGAAACCAGGCCCCCCCCCGAGGCGA
consensus 301 GCcaaAACGgaaACGAaaCGCcaAGAG tggGGGCGAAacCAGgcccCCCCCCGAGgCGA
471F 304 GGG-CGGGGGG--AGGGCGCCGGACGCCGCGTC-CTCGTCGCCTGCGTCCTCC-GCGTCC
471R 241 GGGGCGGGGGGGAAGGGCGCCGGAGGCCCCCGCACGCCGCGCCTGCCGCCGCCTGCCCCC
consensus 361 GGGgCGGGGGGgaAGGGCGCCGGA GCC C CaC C CGCCTGC CC CCtGC CC
471F 359 CCGCCGCCGCC-CGCCCGGAGCGCCGCCCCCCCGCCCC-TGCGCCCCCGCCGTGCCCGC
471R 301 CCGCCGCCGCCGCGCAAGGAGCGCGGGCCGCAGCGCCCCCTTCCGCCCCCCCTGGCGCTA
consensus 421 CCGCCGCCGCCgCGC GGAGCGC G CC C CGCCCCcT C CCCCgCC GC C
471F 417 GCGCGCCGCCCCCAGT--GACTGCCTGGCGCGCATGAACCTCGGCGGGACTGCGGGGCCC-
471R 361 CGGCTGCGCACCACGACGGAGTACCTGGCGCGCCTGAGCCTGCGCCGGCGGCGGCCCCCC
consensus 481 GC CGC CC G cgGA TgCCTGGCGCGC TGAaCCT GC GaC GCGG CCCc
471F 474 GCG--CGCGCCCGGCGCAGCACCGCG-GAGCACGTACGCGTGCC-TGCGACGG---GAGA
471R 421 GCGTCCCCGCCCGCGGACGCGCCGCGCGGAAGGTACGCCTCCCCTCCGACCCCCTGACG
consensus 541 GCGtcC CGCCCG G GCaCCGCGcGaG A GTgCGC T CCcT CGAC cctGA a
471F 527 CGGCGGCGGAGACCTGACGCGGGACTGGCCGGCAGCCGCG--TCCGCGGTCAGCACGGAC
471R 481 CCCCTCCGACCCCCTGACGCCCCTCCGACCCCCTGACGCCCCTCCGACCCCCTGACGCCC
consensus 601 C C CGg CCTGACGC C GgCC C G CGC ccTCCGgC C ACG C

```
471F        585 CTC-GACCGCCTGAGGGCCGTACAATCCCCAGACCCCTTCCGACCCCCGTTTTTTCCCG
471R        541 CTCCGACCCCCTGACGCCCCTCCGACCCCCTGACGCCCCTCCGACCCCCGTGTCTCCCCG
consensus   661 CTCcGACC CCTGA G CC T CaA CCCC GAC CCC TCCGACCCCCGT T T CCCG

471F        644 CCC-CAGGTTTGCTTTTCAAAATC----------------------------------
471R        601 CCCGCAGGTGTGCTTCTCGCCCCGCGTGCAGGTGCGCCATCTGGTGGCCTGGGAGACGGC
consensus   721 CCCgCAGGT TGCTT TCa  a  cgtgcaggtgcgccatctggtggcctgggagacggc

471F            ------------------------------------------------------------
471R        661 CGCGCGCCTGGCCCGACGGGGGTCCTGGGCGCGCGAGCGGGCCGACCGCGACCGGTTCCG
consensus   781 cgcgcgcctggcccgacgggggtcctgggcgcgcgagcgggccgaccgcgaccggttccg

471F            ------------------------------------------------------------
471R        721 GCGCCGCGTGGCGGCGGCCGAGGCGGTCATCGGACCGTGCCTGGAGCCCGAGGCCCGAGC
consensus   841 gcgccgcgtggcggcggccgaggcggtcatcggaccgtgcctggagcccgaggcccgagc

471F            ------------------------------------------------------------
471R        781 TCGGGCCCGAGCCCGAGCCCGGGCCCACGAAGACGGCGGACCCGCGGAGGAGGAGGAGGC
consensus   901 tcgggcccgagcccgagcccgggcccacgaagacggcggacccgcggaggaggaggaggc

471F            ------------------------------------------------------------
471R        841 GGCGGCGGCGGCGCGCGGGTCCTCCGCCGCCGCGGGCCCGGGCCGTCGGGCGTACTAGGG
consensus   961 ggcggcggcggcgcgcgggtcctccgccgccgcgggcccgggccgtcgggcgtactaggg

471F            -
471R        901 T
consensus  1021 t
```

Figure 3.9

Multiple sequence alignments for strains 519F and 519M. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 50%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.


519F 1 ----------GGGGCGACA
519M 1 ACCAGTGATGCTGCCAGCCCCCAGCCCAGCCCCCAGCCCAGCCCCCAGCCCAGTCCCCA
consensus 1 accagtgatgctgccagcccccagcccagcccccagcccagcccccagcccgG C CA
519F 10 TTC--GGCTCCGGGCCTA--CGCCGAGCCCAGCCGCCC-GCCATGTCCCGCCGGCGTCCC
519M 61 GCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCC-CCAGCCTCCC
consensus 61 CcaG C CC gGCC AgtC CC AGCCCAG C CCCaGCC GTCCCgCCgGC TCCC
519F 65 CGCCGCCGGGGACCCCG----GCGCCGGC-CGCGCCCCGGCGCTCCA--GCCGTG-CCGC
519M 120 CTCCGCATGGAAGGGCGCGCCGCGCAGGCGCCCGCCCCTGCGCCCCCGGCGGCGGCCGC
consensus 121 C CCGC GGgA CGcgccGCGC GGCgC CGCCCC GCGC CC cgGC G GgCCGC
519F 117 GCCC----CGGCGCGCTCC------CAACCGCAGACTCCCCAAT---GGTCC---CTGCG
519M 180 GTCCGAGTCCGCGCGGTCCATGAGGCAGCCACCACCGTCGGCATCCCGGTCCACCCCGCC
consensus 181 G CCgagtC GCGCG TCCatgaggCAaC CA aC C a ATcccGGTCCaccC GC
519F 161 TACGACTCGGGAA-CCGCGGTCGAG-ACCGCGCCG--GCCGC-----GTCCTC-GC--TC
519M 240 TTCCCC-CGCGACGCCGGGCGCGAGGACCGCGCCGCCGCCGCCCCCCGTCCGCCGCCGTC
consensus 241 T C CtCG GA gCCG G CGAGgAgCGCGCCGccGCCGCcccccGTCC CcGCcgTC
519F 209 CTGCGGCGCTGGCTGCTGGTGCCCC---AGGCGGACGACAGCGACGACGCGGACTACGCC
519M 299 C-GCGCCGCACGCCCCGGAGGCCGCGGAAGGGGGGCGGGAGCGAGCCCCCCGAC-AAGCG
consensus 301 CtGCG CGC GC C Gg GCC CggaAGG GGaCGa AGCGA C C GACtA GC
519F 266 GGCAACGACGACGCAG-AGTGGGCGAA-CAGCCCCCCGAGCGAGGGCGGGGGAAGGCGC
519M 357 GGTGA-GACGCAGCAGCAACGGCCGCATCAGGCGCAGCAGCGACGCGGCGGACCCCGT
consensus 361 GG aAcGACG GCAGcAg GG CG AtCAG C C AGCGA gGCgGG GGA CG
519F 324 CG--GGAGG--GCGGCGTG-----TCGTGGCCGGCGGT-----AGCTGGTT------GG
519M 416 GGTCGTCGGTCGCGGCGTCCTCGCTCCTGCGCCGGCCGCGCGAGAGCTGGCGCCAAGGGC
consensus 421 GtcG GGtcGCGGCGT ctcgcTC TG GCCGGC G gcgagAGCTGG ccaaggG
519F 364 TCGGCGT-CGTTGCGGGG---------GGCGCC-------GGGCGGCACGGCCGC-TG--
519M 476 TCAGCATGCGTCCCTGGTAAACCCTCAGACGCCAACCCTCGCGCGGCCCCGCGCCGTGCC
consensus 481 TCgGCgTgCGT C GG aaaccctcaGgCGCCaaccctcG GCGGC C GC CgTGcc
519F 404 GAGCGAGGGTCGCGGCGGCGCGCGCG-CGACGGGGGTAAGTGACCTTG-GTACGCCGGGG
519M 536 GACGTCGCGCCCCGCGCCGGCCGCGGCCCCAGGGGCCGCCGCCCCTGCGGCCGCCCTGT
consensus 541 GA C G GaC C GCG CG CGCGgC CgGGGG a G CC TGcG CGCC G
519F 462 ACCATTCGC-GA---GAGTTGGGGTCGGGAAGCGCCCGGGGGGCGCGGCGCG------GG
519M 596 ACCGCCCGCCGACCCGAGCCGCATCCGGGCCTCGGCCTCGCGCAGCCCCGCCCTCTCAGG
consensus 601 ACCa CGCcGAcccGAG G g CGGG CG CC G G GC CGC ctctcaGG

```
519F        512 G----
519M        656 GTGAT
consensus   661 Gtgat
```

Figure 3.10

Multiple sequence alignments for strains 564F and 564M. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 25%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.

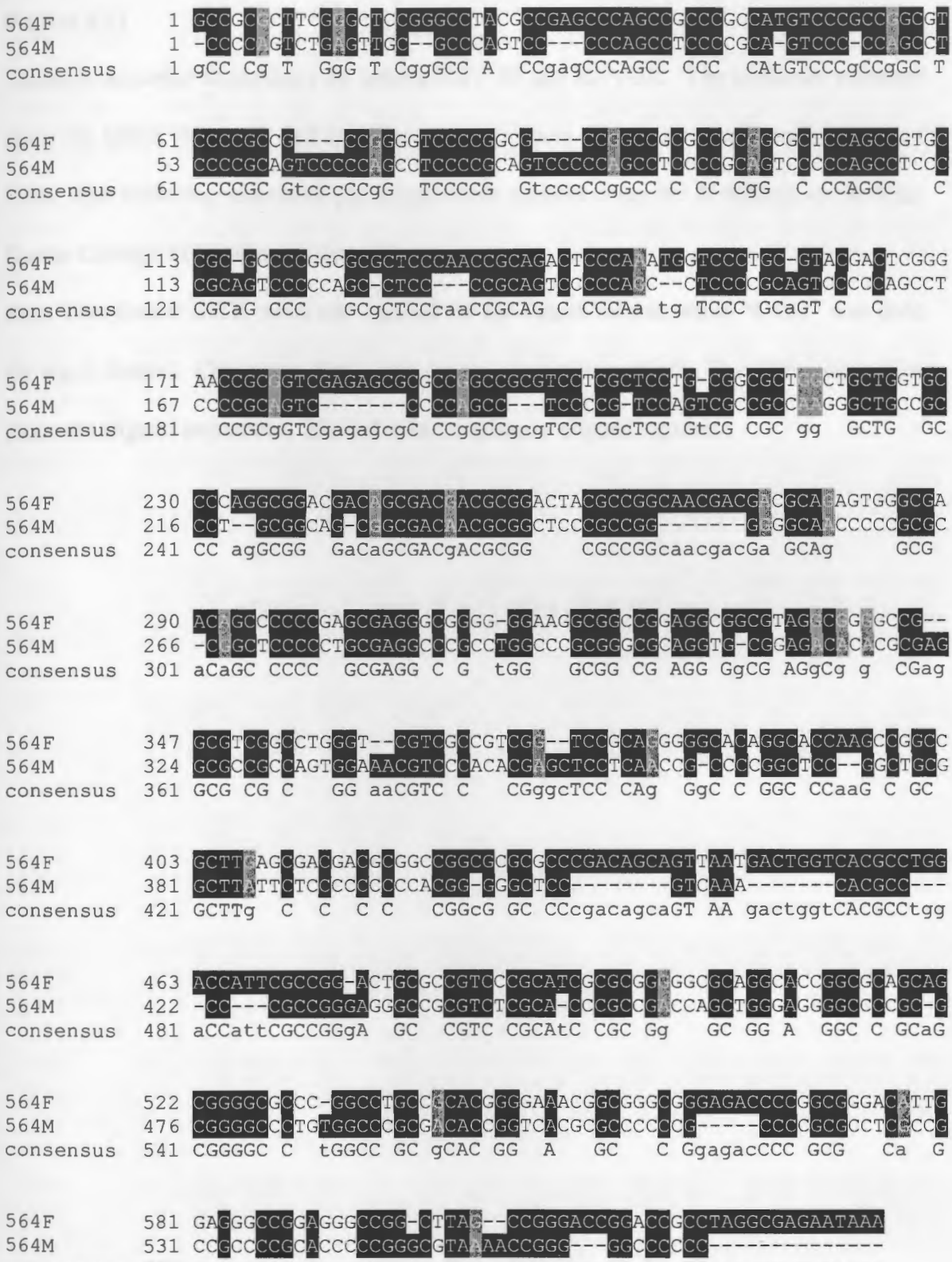
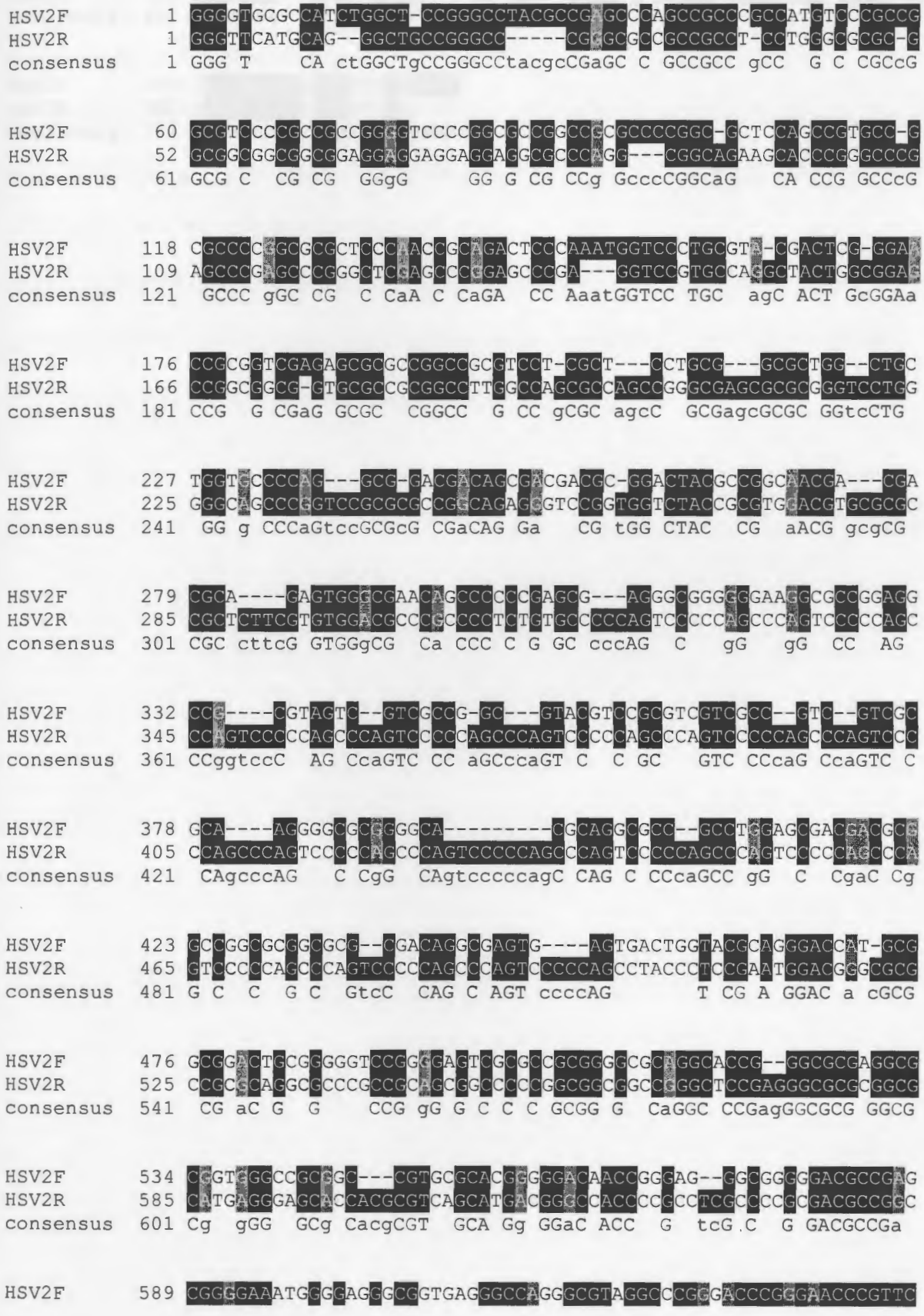

564F 1 GCCGCCCTTCGGGCTCCGGGCCTACGCCGAGCCCAGCCGCCCGCCATGTCCCGCCGGCGT
564M 1 -CCCCAGTCTGAGTTGC--GCCCAGTCC---CCCAGCCTCCCCGCA-GTCCC-CCAGCCT
consensus 1 gCC Cg T GgG T CggGCC A CCgagCCCAGCC CCC CAtGTCCCgCCgGC T
564F 61 CCCCGCCG----CCGGGGTCCCCGGCG----CCGCCGCGCCCCGGCGCTCCAGCCGTGC
564M 53 CCCCGCAGTCCCCCAGCCTCCCCGCAGTCCCCCAGCCTCCCCGCAGTCCCCCAGCCTCCC
consensus 61 CCCCGC GtcccCCgG TCCCCG GtcccCCgGCC C CC CgG C CCAGCC C
564F 113 CGC-GCCCCGGCGCGCTCCCAACCGCAGACTCCCAAATGGTCCCTGC-GTACGACTCGGG
564M 113 CGCAGTCCCCCAGC-CTCC---CCGCAGTCCCCCAGC--CTCCCCGCAGTCCCCCAGCCT
consensus 121 CGCaG CCC GCgCTCCcaaCCGCAG C CCCAa tg TCCC GCaGT C C
564F 171 AACCGCGGTCGAGAGCGCGCCCGGCCGCGTCCTCGCTCCTG-CGGCGCTGGCTGCTGGTGC
564M 167 CCCCGCAGTC-------CCCCAGCC---TCCCCG-TCCAGTCGCCGCCAGGGCTGCCGC
consensus 181 CCGCgGTCgagagcgC CCgGCCgcgTCC CGcTCC GtCG CGC gg GCTG GC
564F 230 CCCAGGCGGACGACAGCGACGACGCGGACTACGCCGGCAACGACGACGCACAGTGGGCGA
564M 216 CCT--GCGGCAG-CGGCGACAACGCGGCTCCCGCCGG-------GGGGCAACCCCCGCGC
consensus 241 CC agGCGG GaCaGCGACgACGCGG CGCCGGcaacgacGa GCAg GCG
564F 290 ACAGCCCCCCGAGCGAGGGCGGGG-GGAAGGCGGCCGGAGGCGGCGTAGGCCGGGCCG--
564M 266 -CGGCTCCCCCTGCGAGGCCCGCCTGGCCCGCGGGCGCAGGTG-CGGAGACACACGCGAG
consensus 301 aCaGC CCCC GCGAGG C G tGG GCGG CG AGG GgCG AGgCg g CGag
564F 347 GCGTCGGCCTGGGT--CGTCGCCGTCGG--TCCGCAGGGGGCACAGGCACCAAGCCGGCC
564M 324 GCGCCGCCAGTGGAAACGTCCCACACGAGCTCCTCAACCG-CCCCGGCTCC--GGCTGCG
consensus 361 GCG CG C GG aaCGTC C CGggcTCC CAg GgC C GGC CCaaG C GC
564F 403 GCTTCAGCGACGACGCGGCCGGCGCGCGCCCGACAGCAGTTAATGACTGGTCACGCCTGG
564M 381 GCTTATTCTCCCCCCCCACGG-GGGCTCC--------GTCAAA-------CACGCC---
consensus 421 GCTTg C C C C CGGcG GC CCcgacagcaGT AA gactggtCACGCCtgg
564F 463 ACCATTCGCCGG-ACTGCGCCGTCCCGCATCGCGCGGGGGCGCAGGCACCGGCGCAGCAG
564M 422 -CC---CGCCGGGAGGGCCGCGTCTCGCA-CCCGCCGCCAGCTGGGAGGGGCCCCGC-G
consensus 481 aCCattCGCCGGgA GC CGTC CGCAtC CGC Gg GC GG A GGC C GCaG
564F 522 CGGGGCGCCC-GGCCTGCCCCACGGGGAAACGGCGGGCGGGAGACCCCGGCGGGACATTG
564M 476 CGGGGCCCTGTGGCCCCGCGACACCGGTCACGCGCCCCCCG-----CCCCGCGCCTCCCCG
consensus 541 CGGGGC C tGGCC GC gCAC GG A GC C GgagacCCC GCG Ca G
564F 581 GAGGGCCGGAGGGCCGG-CTTAC--CCGGGACCGGACCGCCTAGGCGAGAATAAA
564M 531 CCGCCCCGCACCCCCGGGGCGTAAAACCGGG---GGCCCCCC--------------
consensus 601 G CCG A CCGGgC TAgaaCCGGGaccGG CC CCtaggcgagaataaa

Figure 3.11

Multiple sequence alignments for strains HSV-2F and HSV-2R. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 18%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.

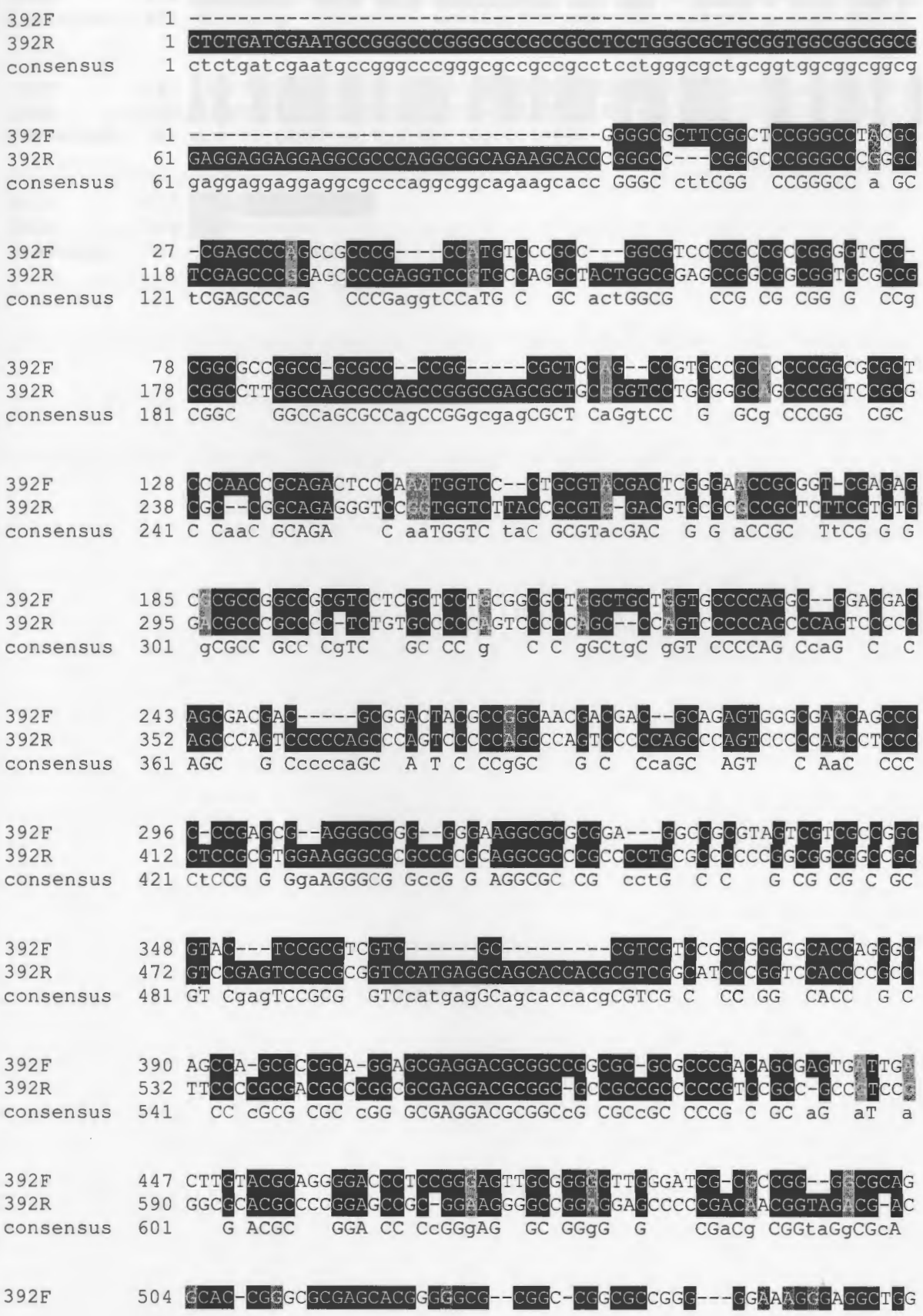

HSV2F 1 GGGGTGCGCCATCTGGCT-CCGGGCCTACGCCGAGCCCAGCCGCCCGCCATGTCCCGCCG
HSV2R 1 GGGTTCATGCAG--GGCTGCCGGGCC-----CGGGCGCCGCCGCCT-CCTGGGCGCGC-G
consensus 1 GGG T CA ctGGCTgCCGGGCCtacgcCGaGC C GCCGCC gCC G C CGCcG
HSV2F 60 GCGTCCCCGCCGCCGGGGGTCCCCGGCGCCGGCCGCGCCCCGGC-GCTCCAGCCGTGCC-G
HSV2R 52 GCGGCGGCGGCGGAGGAGGAGGAGGAGGCGCCCAGG---CGGCAGAAGCACCCGGGCCCG
consensus 61 GCG C CG CG GGgG GG G CG CCg GcccCGGCaG CA CCG GCCcG
HSV2F 118 CGCCCCGGCGCGCTCCCAACCGCAGACTCCCAAATGGTCCCTGCGT-CGACTCG-GGA
HSV2R 109 AGCCCGAGCCCGGGCTCCAGCCCCGAGCCCGA---GGTCCGTGCCAGGCTACTGGCGGAG
consensus 121 GCCC gGC CG C CaA C CaGA CC AaatGGTCC TGC agC ACT GcGGAa
HSV2F 176 CCGCGGTCGAGAGCGCGCCGGCCGCGTCCT-CGCT---CCTGCG---GCGCTGG--CTGC
HSV2R 166 CCGGCGGCG-GTGCGCCGCGGCCTTGGCCAGCGCCAGCCGGGCGAGCGCGCGGGTCCTGG
consensus 181 CCG G CGaG GCGC CGGCC G CC gCGC agcC GCGagcGCGC GGtcCTG
HSV2F 227 TGGTGCCCCAG---GCG-GACGACAGCGACGACGC-GGACTACGCCGGCAACGA---CGA
HSV2R 225 GGGCAGCCCGGTCCGCGCGCCGGCAGAGCGTCCGGTGGTCTACCGCGTGGACGTGCGCGC
consensus 241 GG g CCCaGtccGCGcG CGaCAG Ga CG tGG CTAC CG aACG gcgCG
HSV2F 279 CGCA----GAGTGGGCGAACAGCCCCCCGAGCG---AGGGCGGGGGGAAGGCGCCGGAGG
HSV2R 285 CGCTCTTCGTGTGGACGCCCGCCCCTCTGTGCCCCCAGTCCCCCAGCCCAGTCCCCCAGC
consensus 301 CGC cttcG GTGGgCG Ca CCC C G GC cccAG C gG gG CC AG
HSV2F 332 CCG----CGTAGTC--GTCGCCG-GC---GTACGTCCGCGTCGTCGCC--GTC--GTCGC
HSV2R 345 CCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGGCCAGTCCCCCAGCCCAGTCCC
consensus 361 CCggtccC AG CcaGTC CC aGCccaGT C C GC GTC CCcaG CcaGTC C
HSV2F 378 GCA----AGGGGCGCGGGGGCA---------CGCAGGCGCC--GCCTGGAGCGACGACGCG
HSV2R 405 CCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCA
consensus 421 CAgcccAG C CgG CAgtcccccagC CAG C CCcaGCC gG C CgaC Cg
HSV2F 423 GCCGGCGCGGCGCG--CGACAGGCGAGTG----AGTGACTGGTACGCAGGGACCAT-GCG
HSV2R 465 GTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCTACCCTCCGAATGGACGGGCGCG
consensus 481 G C C G C GtcC CAG C AGT ccccAG T CG A GGAC a cGCG
HSV2F 476 GCGGACTGCGGGGGTCCGGGGAGTCGGCGCCGCGGGGCGCAGGCACCG--GGCGCGAGGCG
HSV2R 525 CCGCGCAGGCGCCCGCCGCAGCGGCCCCCGGCGGCGGCCGGGCTCCGAGGGCGCGCGGCG
consensus 541 CG aC G G CCG gG G C C C GCGG G CaGGC CCGagGGCGCG GGCG
HSV2F 534 CGGTGGGCCGCGGC---CGTGCGCACGGGGGACAACCGGGAG--GGCGGGGGACGCCGAG
HSV2R 585 CATGAGGGAGCACCACGCGTCAGCATGACGGGCCACCCCGCCTCGCCCCGCGACGCCGGC
consensus 601 Cg gGG GCg CacgCGT GCA Gg GGaC ACC G tcG C G GACGCCGa
HSV2F 589 CGGGGAAATGGGGAGGGCGGTGAGGGCCAGGGCGTAGGCCCGGGACCCGGGAACCCGTTC

```
HSV2R      645 CGGAGAACGCGCGCCGCCGCCCCCG-CCGCCGCGTCG-CGCGACGCCCGAGG-CCCG--C
consensus  661 CGGgGAA   G G  G CG     GgCCa  GCGT GgC CGg aCCCGgGaaCCCGttC

HSV2F      649 CGGGGCCGGGAAAATTCAAAAA
HSV2R      700 AGGGGCCGACGAGCCCCC----
consensus  721  GGGGCCGg aAa   C aaaa
```

Figure 3.12

Multiple sequence alignments for strains 392F and 392R. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 27%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.


392F 1 ----------
392R 1 CTCTGATCGAATGCCGGGCCCGGGCGCCGCCGCCTCCTGGGCGCTGCGGTGGCGGCGGCG
consensus 1 ctctgatcgaatgccgggcccgggcgccgccgcctcctgggcgctgcggtggcggcggcg
392F 1 ----------GGGGCGCTTCGGCTCCGGGCCTACGC
392R 61 GAGGAGGAGGAGGCGCCCAGGCGGCAGAAGCACCCGGGCC---CGGGCCCGGGCCCCGGGC
consensus 61 gaggaggaggaggcgcccaggcggcagaagcacc GGGC cttCGG CCGGGCC a GC
392F 27 -CGAGCCCAGCCGCCCG----CCTGTCCCGCC---GGCGTCCCGCCGCCGGGGTCCC-
392R 118 TCGAGCCCGGAGCCCCGAGGTCCTGCCAGGCTACTGGCGGAGCCGGCGCGGTGCGCCG
consensus 121 tCGAGCCCaG CCCGaggtCCaTG C GC actGGCG CCG CG CGG G CCg
392F 78 CGGCGCCGGCC-GCGCC--CCGG-----CGCTCCAG--CCGTGCCGC CCCCGGCGCGCT
392R 178 CGGCCTTGGCCAGCGCCAGCCGGGCGAGCGCTGCTGGTCCTGGGGGCAGCCCGGTCCGCG
consensus 181 CGGC GGCCaGCGCCagCCGGgcgagCGCT CaGgtCC G GCg CCCGG CGC
392F 128 CCCAACCGCAGACTCCCAAATGGTCC--CTGCGTACGACTCGGGAACCGCGGT-CGAGAG
392R 238 CGC--CGGCAGAGGGTCCCGTGGTCTTACCGCGTG-GACGTGCGCCCCGCTCTTCGTGTG
consensus 241 C CaaC GCAGA C aaTGGTC taC GCGTacGAC G G aCCGC TtCG G G
392F 185 CCCGCCGGCCGCGTCCTCGCTCCTCCGGCGCTGGCTGCTGGTGCCCCAGGC--GGACGAC
392R 295 GACGCCCGCCCC-TCTGTGCCCCCAGTCCCCCCAGC--CCAGTCCCCCAGCCCAGTCCCCC
consensus 301 gCGCC GCC CgTC GC CC g C C gGCtgC gGT CCCCAG CcaG C C
392F 243 AGCGACGAC-----GCGGACTACGCCGGCAACGACGAC--GCAGAGTGGGCGAACAGCCC
392R 352 AGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCTCCC
consensus 361 AGC G CccccaGC A T C CCgGC G C CcaGC AGT C AaC CCC
392F 296 C-CCGAGCG--AGGGCGGG--GGGAAGGCGCGCGGA---GGCCGCGTAGTCGTCGCCGGC
392R 412 CTCCGCGTGGAAGGGCGCGCCGCGCAGGCGCCCGCCCCTGCGCCCCCCGGCGGCGGCCGC
consensus 421 CtCCG G GgaAGGGCG GccG G AGGCGC CG cctG C C G CG CG C GC
392F 348 GTAC---TCCGCGTCGTC------GC---------CGTCGTCCGCCGGGGGCACCAGGGC
392R 472 GTCCGAGTCCGCGCGGTCCATGAGGCAGCACCACGCGTCGGCATCCCGGTCCACCCCGCC
consensus 481 GT CgagTCCGCG GTCcatgagGCagcaccacgCGTCG C CC GG CACC G C
392F 390 AGCCA-GCGCCGCA-GGAGCGAGGACGCGGCCGGCGC-GCGCCCGACAGCGAGTGATTGA
392R 532 TTCCCGCGACGCCCGGCGCGAGGACGCGGC-GCCGCCGCCCCCGTCCGCC-GCCCTCCC
consensus 541 CC cGCG CGC cGG GCGAGGACGCGGCcG CGCcGC CCCG C GC aG aT a
392F 447 CTTGTACGCAGGGGACCCTCCGGGAGTTGCGGGCGTTGGGATCG-CGCCGG--GGCGCAG
392R 590 GGCGCACGCCCCGGAGCCGC-GGAAGGGGCCGGAGGAGCCCCCGACAACGGTAGACG-AC
consensus 601 G ACGC GGA CC CcGGgAG GC GGgG G CGaCg CGGtaGgCGcA
392F 504 GCAC-CGGGCGCGAGCACGGGGGCG--CGGC-CGGCGCCGGG---GGAAAGGGAGGCTGG

```
392R       648 ACACGCGATAGCGA-CACGCGAGCGACCGGTGCGTCGCTCTGTCTGCCCGGAGTGGCAGT
consensus  661 gCACgCGg  GCGAgCACG GgGCGacCGG gCG CGC   GtctG a aGgG GGC G

392F       557 CGACGCCGGGCGGGAAATGGGGGGGCGGAGGGCCAGGCGTTGCCGGACCGGAACCCGTCG
392R       707 ACAGTCTGAACTAGCACTGGGCCG-CGCATGGCGCG-CGCTGCCCCTGCCTACCCAGCAG
consensus  721 g A  C GggC gG A TGGG  GgCG A GGC  GgCG TGCC   aC  A CC G  G

392F       617 GGCGGGAAATTAAAA
392R       765 GGCT-----------
consensus  781 GGC ggaaattaaaa
```

Figure 3.13

Multiple sequence alignments for strains 472F and 472R. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 14%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.

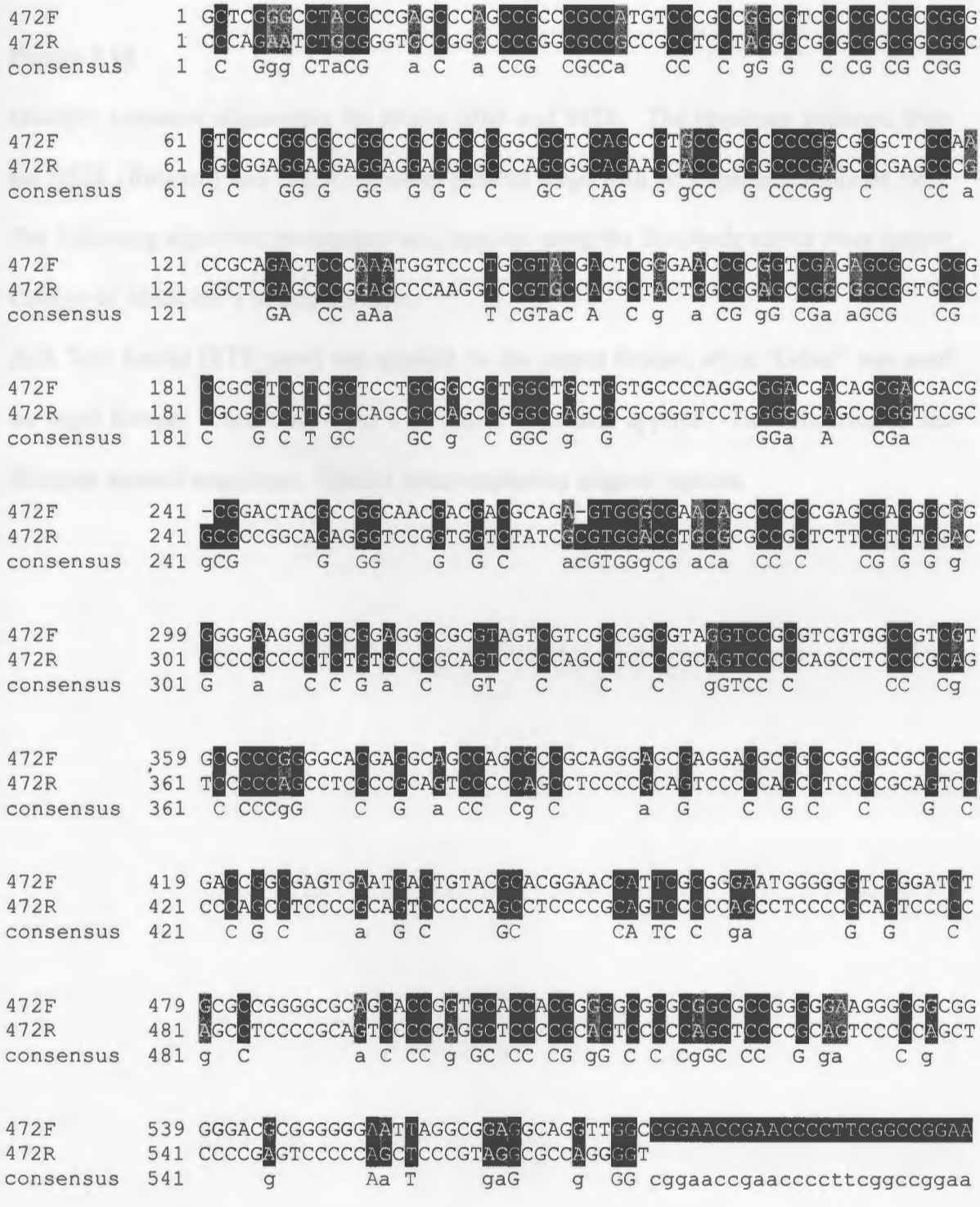
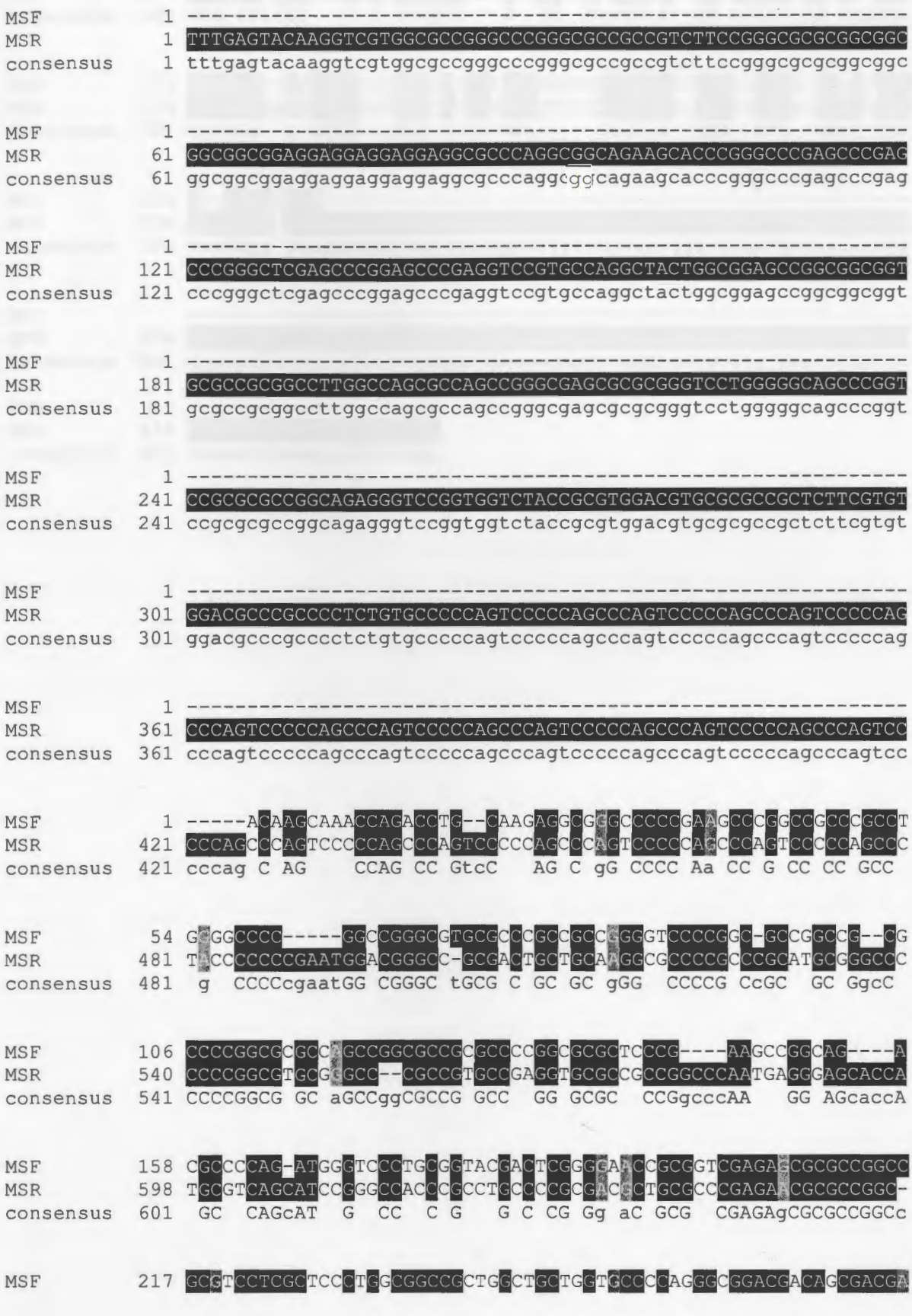

472F 1 GCTCGGGCCTACGCCGAGCCCAGCCGCCCGCCATGTCCCGCCGGCGTCCCCGCCGCCGGG
472R 1 CCCAGAATCTGCGGGTGCCGGGCCCGGGCGCCGCCGCCTCCTAGGGCGCGCGGCGGCGGC
consensus 1 C Ggg CTaCG a C a CCG CGCCa CC C gG G C CG CG CGG
472F 61 GTCCCCGGCGCCGGCCGCGCCCCGGCGCTCCAGCCGTGCCGCGCCCCGGCGCGCTCCCAA
472R 61 GGCGGAGGAGGAGGAGGAGGCGCCCAGCGGCAGAAGCACCCGGGCCCGAGCCCGAGCCCG
consensus 61 G C GG G GG G G C C GC CAG G gCC G CCCGg C CC a
472F 121 CCGCAGACTCCCAAATGGTCCCTGCGTACGACTCGGGAACCGCGGTCGAGAGCCGCCGG
472R 121 GGCTCGAGCCCGGAGCCCAAGGTCCGTGCCAGGCTACTGGCGGAGCCGGCGGCGGTGCGC
consensus 121 GA CC aAa T CGTaC A C g a CG gG CGa aGCG CG
472F 181 CCGCGTCCTCGCTCCTGCGGCGCTGGCTGCTGGTGCCCCAGGCGGACGACAGCGACGACG
472R 181 CGCGGCCTTGGCCAGCGCCAGCCGGGCGAGCGCGCGGGTCCTGGGGGCAGCCCGGTCCGC
consensus 181 C G C T GC GC g C GGC g G GGa A CGa
472F 241 -CGGACTACGCCGGCAACGACGACGCAGA-GTGGGCGAACAGCCCCCCGAGCGAGGGCGG
472R 241 GCGCCGGCAGAGGGTCCGGTGGTCTATCGCGTGGACGTGCGCGCCGCTCTTCGTGTGGAC
consensus 241 gCG G GG G G C acGTGGgCG aCa CC C CG G G g
472F 299 GGGGAAGGCGCCGGAGGCCGCGTAGTCGTCGCCGGCGTAGGTCCGCGTCGTGGCCGTCGT
472R 301 GCCCGCCCCTCTGTGCCCGCAGTCCCCCAGCGTCCCCGCAGTCCCCCAGCCTCCCCGCAG
consensus 301 G a C C G a C GT C C C GgTCC C CC Cg
472F 359 GCGCCCGGGGCACGAGGCAGCCAGCGCCGCAGGGAGCGAGGACGCGGCCGGCGCGCGCGC
472R 361 TCCCCCAGCCTCCCCGCAGTCCCCCAGCCTCCCCGCAGTCCCCCAGCCTCCCCGCAGTCC
consensus 361 C CCCgG C G a CC Cg C a G C G C C G C
472F 419 GACCGGCGAGTGAATGACTGTACGCACGGAACCATTCGCGGGAATGGGGGGTCGGGATCT
472R 421 CCCAGCCTCCCCGCAGTCCCCCAGCCTCCCCGCAGTCCCCCAGCCTCCCCGCAGTCCCCC
consensus 421 C G C a G C GC CA TC C ga G G C
472F 479 GCGCCGGGGCGCAGCACCGGTGCACCACGGGGGCGCGCGGCGCCGGGGGAAGGGCGGCGG
472R 481 AGCCTCCCCGCAGTCCCCCAGGCTCCCCGCAGTCCCCCAGCTCCCCGCAGTCCCCCAGCT
consensus 481 g C a C CC g GC CC CG gG C C CgGC CC G ga C g
472F 539 GGGACGCGGGGGGAATTAGGCGGAGGCAGGTTGGCCGGAACCGAACCCCTTCGGCCGGAA
472R 541 CCCCGAGTCCCCCAGCTCCCGTAGGCGCCAGGGGT
consensus 541 g Aa T gaG g GG cggaaccgaaccccttcggccggaa

Figure 3.14

Multiple sequence alignments for strains MSF and MSR. The sequence obtained from the HS2F (forward) and HS2R (reverse) primers align with an alignment score of 18%. The following algorithm parameters were applied using the Boxshade server from Baylor College of Medicine's Searchlauncher:

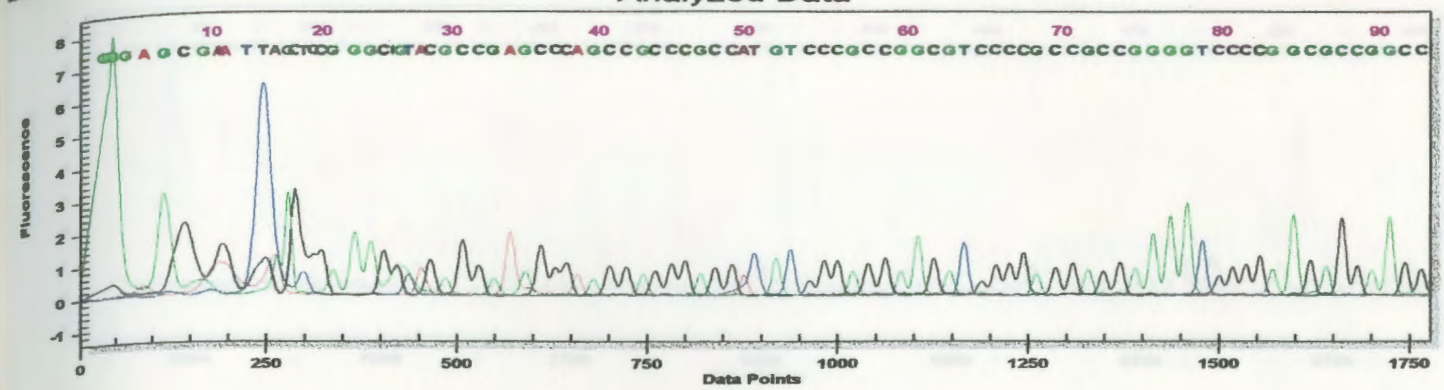
Rich Text format (RTF_new) was applied for the output format, while "Other" was used for input format. Consensus lines with letters were also applied. The following tables illustrate aligned sequences. Shaded areas emphasize aligned regions.


MSF 1 ----------
MSR 1 TTTGAGTACAAGGTCGTGGCGCCGGGCCCGGGCGCCGCCGTCTTCCGGGCGCGCGGCGGC
consensus 1 tttgagtacaaggtcgtggcgccgggcccgggcgccgccgtcttccgggcgcgcggcggc
MSF 1 ----------
MSR 61 GGCGGCGGAGGAGGAGGAGGAGGCGCCCAGGCGGCAGAAGCACCCGGGCCCGAGCCCGAG
consensus 61 ggcggcggaggaggaggaggaggcgcccaggcggcagaagcacccgggcccgagcccgag
MSF 1 ----------
MSR 121 CCCGGGCTCGAGCCCGGAGCCCGAGGTCCGTGCCAGGCTACTGGCGGAGCCGGCGGCGGT
consensus 121 cccgggctcgagcccggagcccgaggtccgtgccaggctactggcggagccggcggcggt
MSF 1 ----------
MSR 181 GCGCCGCGGCCTTGGCCAGCGCCAGCCGGGCGAGCGCGCGGGTCCTGGGGGCAGCCCGGT
consensus 181 gcgccgcggccttggccagcgccagccgggcgagcgcgcgggtcctgggggcagcccggt
MSF 1 ----------
MSR 241 CCGCGCGCCGGCAGAGGGTCCGGTGGTCTACCGCGTGGACGTGCGCGCCGCTCTTCGTGT
consensus 241 ccgcgcgccggcagagggtccggtggtctaccgcgtggacgtgcgcgccgctcttcgtgt
MSF 1 ----------
MSR 301 GGACGCCCGCCCCTCTGTGCCCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAG
consensus 301 ggacgcccgcccctctgtgcccccagtccccccagcccagtccccccagcccagtccccccag
MSF 1 ----------
MSR 361 CCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCC
consensus 361 cccagtccccccagcccagtccccccagcccagtccccccagcccagtccccccagcccagtcc
MSF 1 -----ACAAGCAAACCAGACCTG--CAAGAGGCGGGCCCCCGAAGCCCGGCCGCCCGCCT
MSR 421 CCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCCAGTCCCCCAGCCC
consensus 421 cccag C AG CCAG CC GtcC AG C gG CCCC Aa CC G CC CC GCC
MSF 54 GCGGCCCC-----GGCCGGGCGTGCGCCCGCCGGCCGGGTCCCCGGC-GCCGGCCG--CG
MSR 481 TCCCCCCCCGAATGGACGGGCC-GCGACTGCTGCAAGGCGCCCCGCCCGCATGCGGGCCC
consensus 481 g CCCCcgaatGG CGGGC tGCG C GC GC gGG CCCCG CcGC GC GgcC
MSF 106 CCCCGGCGCGCCGCCGGCGCCGCGCCCCGGCGCGCTCCCG-----AAGCCGGCAG----A
MSR 540 CCCCGGCGTGCGGCCC--CGCCGTGCCGAGGTGCGCCGCCGGCCCAATGAGGGAGCACCA
consensus 541 CCCCGGCG GC aGCCggCGCCG GCC GG GCGC CCGgcccAA GG AGcaccA
MSF 158 CGCCCCAG-ATGGGTCCCTGCGGTACGACTCGGGGCACCGCGGTCGAGACGCGCCGGCC
MSR 598 TGCGTCAGCATCCGGGCCACCCGCCTGCCCCGCGACGCTGCGCCCGAGACGCGCCGGC-
consensus 601 GC CAGcAT G CC C G G C CG Gg aC GCG CGAGAgCGCGCCGGCc
MSF 217 GCGTCCTCGCTCCCTGGCGGCCGCTGGCTGCTGGTGCCCCCAGGGCGGACGACAGCGACGA

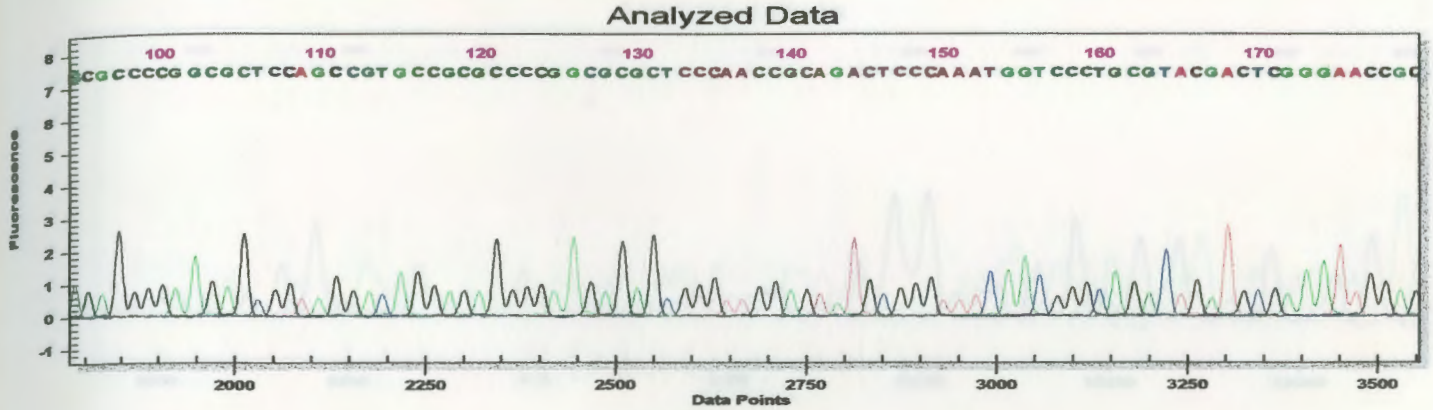
```
MSR          657 GCAGCC-CGCCGTCGGCCGGCCGTCCGTCGCGCGAACCGCCCGGGGGACGCCACG--CGG
consensus    661 GCg CCtCGC   C G CGGCCG   G  GC  G gCC C  GG GGACG CA CgaCGa

MSF          277 CGCCGGACTACGCCGGCCAACGACGACGCAGAGGTGGGCGACCAGGCCCCCCCGA-GCGAG
MSR          714 CACAGGGAGGGCCGCGGAGCCAGGAGGCGCCCCCGGACAAAGAGGGGCCCAGGACGAGAG
consensus    721 CgCgGa  a GCCG   AaC A GA GCa     GGgCgA  AGG  CCC  GAcG GAG

MSF          336 G--GCGGGGGG----------------------------------------------------
MSR          774 GAAGCAACGGGGCAAAATAAAGGGAAGAAATGACAGAAGCGGACACGGTGATGGCCCCGG
consensus    781 GaaGCgg GGGgcaaaataaagggaagaaatgacagaagcggacacggtgatggccccgg

MSF          ------------------------------------------------------------
MSR          834 GGGGCTGCTGTGGCTTGCTGTGACTTGCGCGAGGCTGGGAGAGGCTTGGAAACCCTATGC
consensus    841 ggggctgctgtggcttgctgtgacttgcgcgaggctgggagaggcttggaaaccctatgc

MSF          ----------------------
MSR          894 CACACCTGAAGGGGCCGTGGG
consensus    901 cacacctgaaggggccgtggg
```

APPENDIX: Sample chromatograms

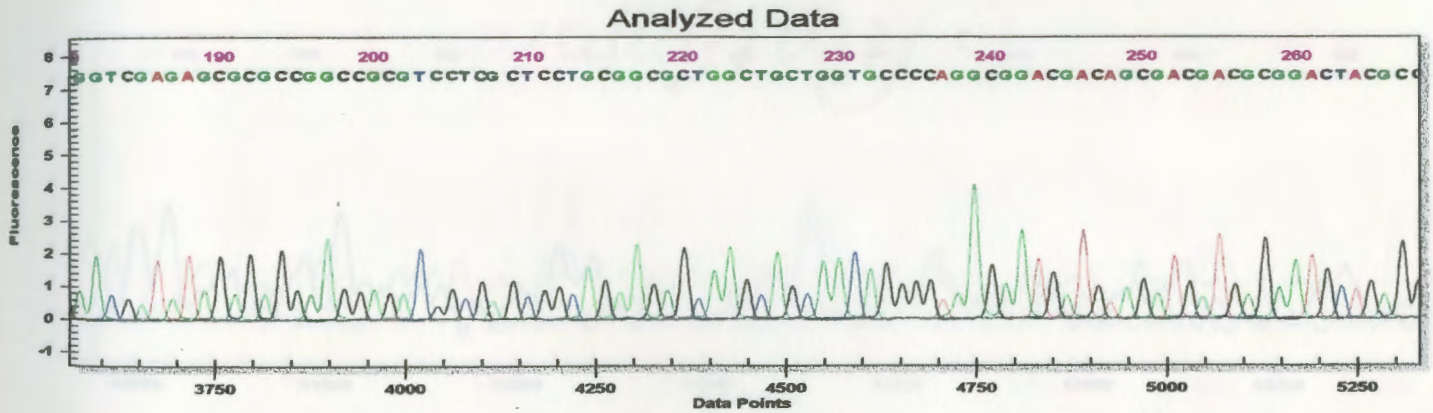
Analyzed Data



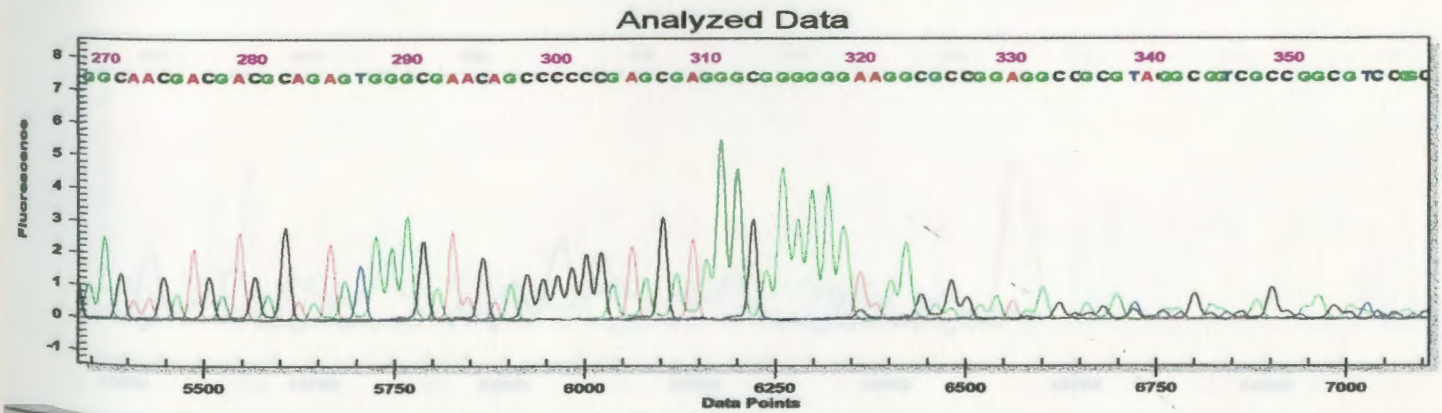
Analyzed Data



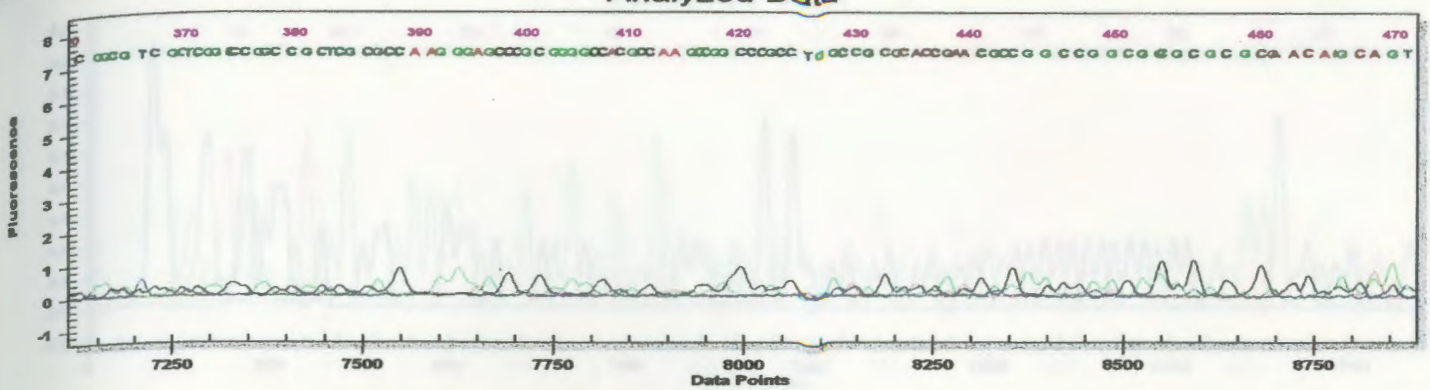
Analyzed Data



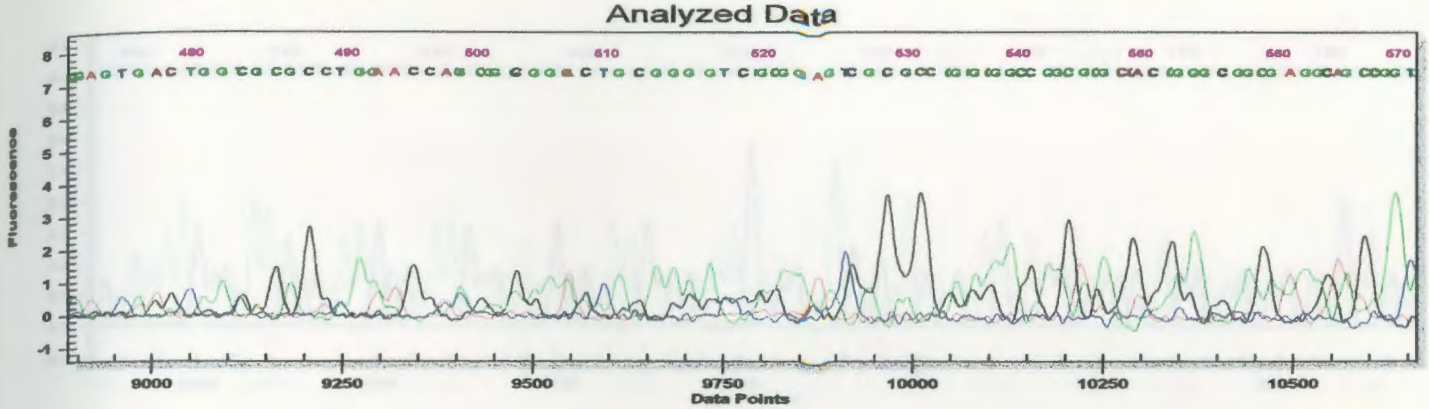
Analyzed Data



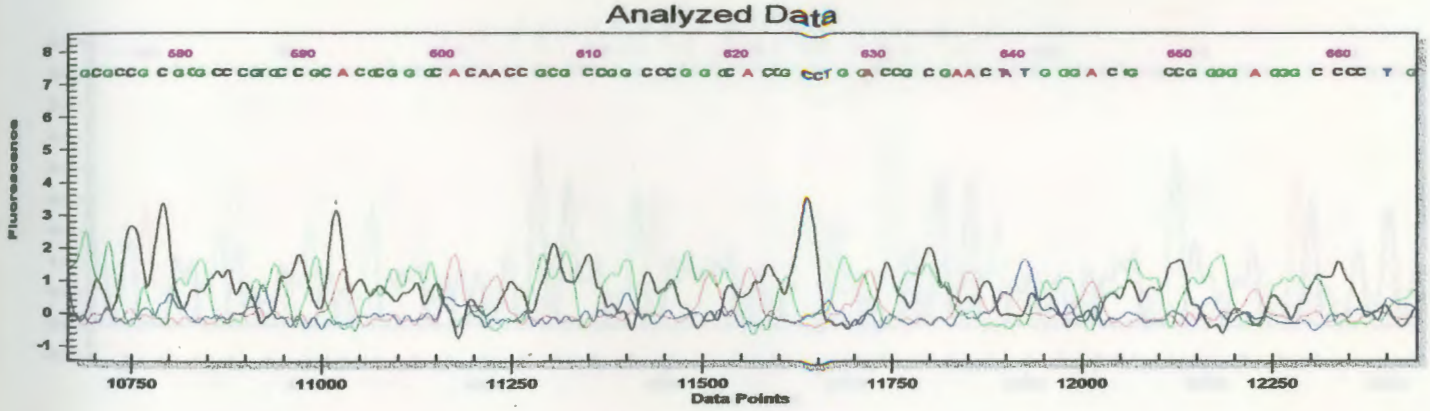
Analyzed Data



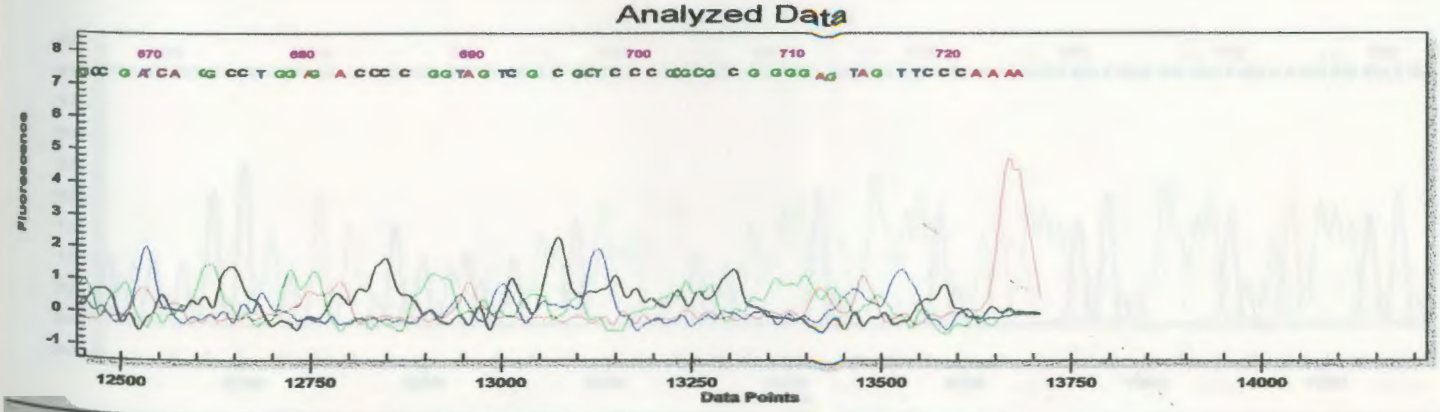
Analyzed Data



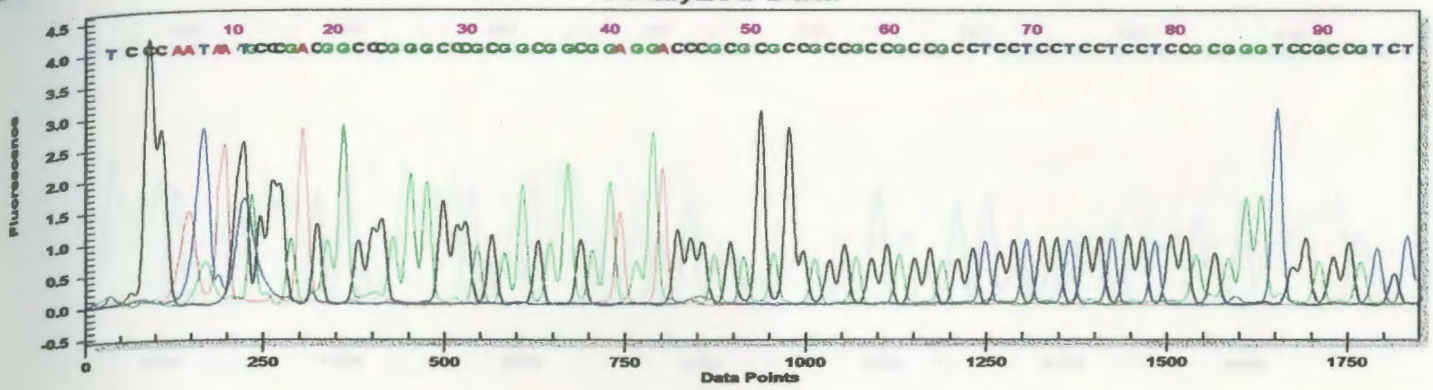
Analyzed Data



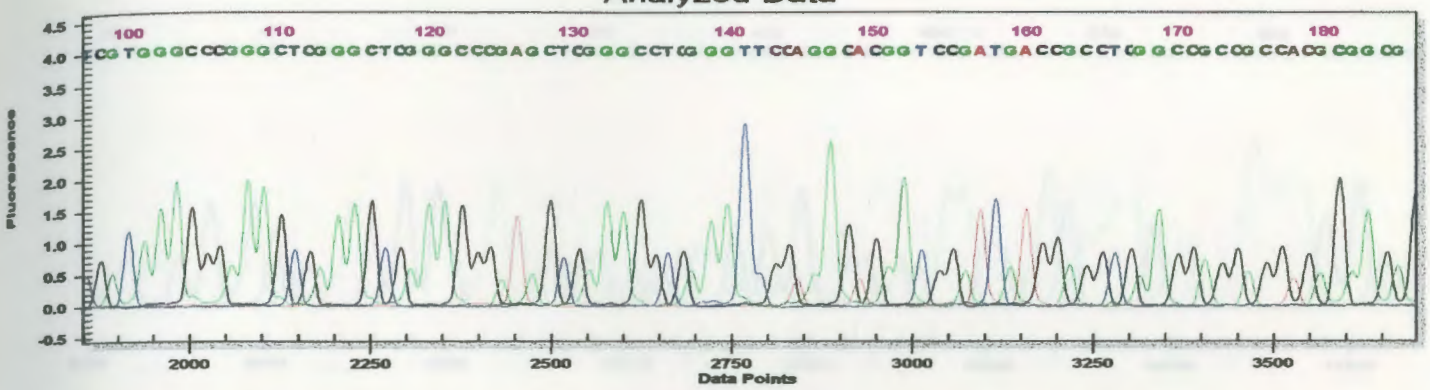
Analyzed Data



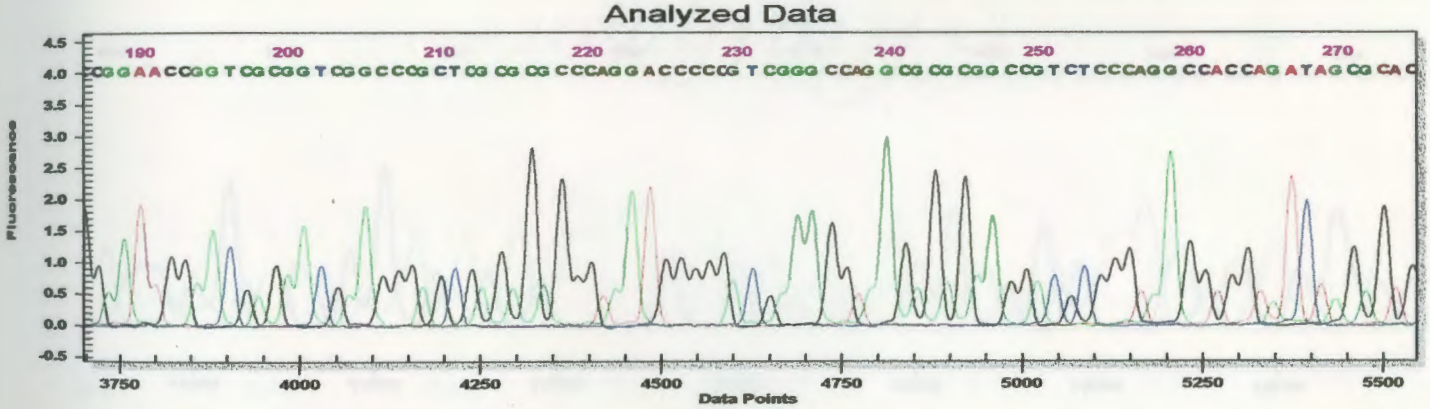
Analyzed Data



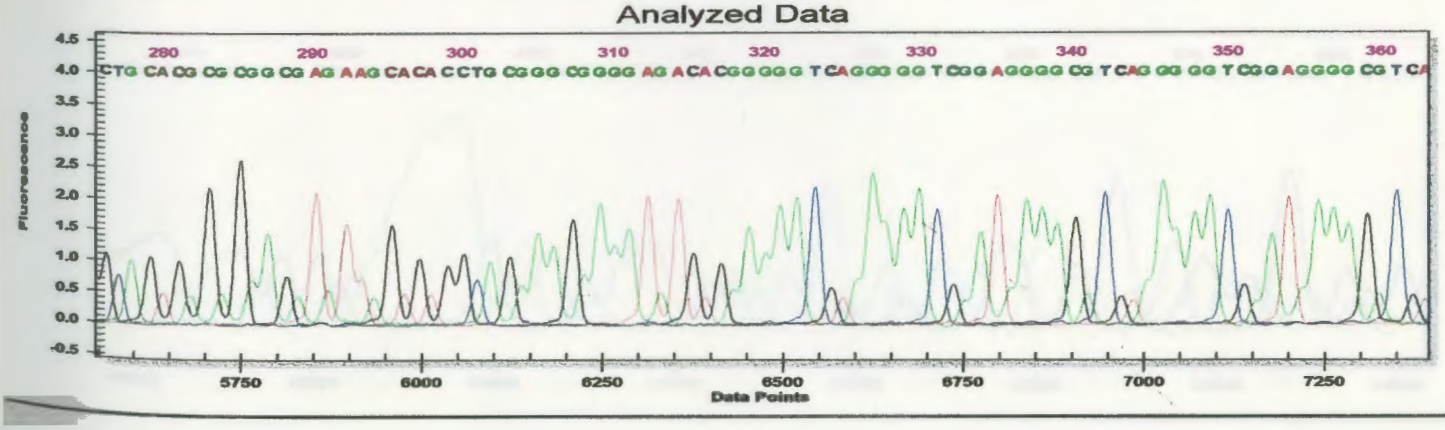
Analyzed Data



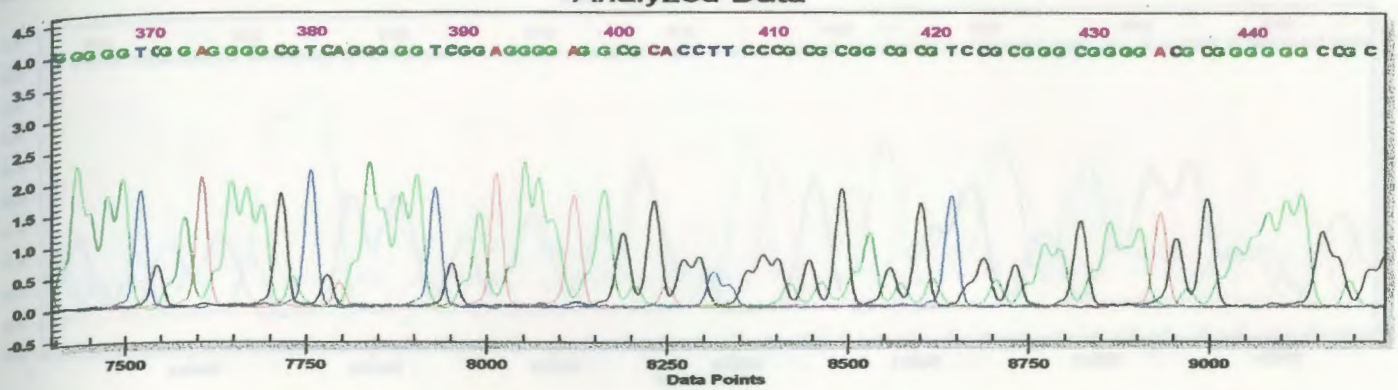
Analyzed Data



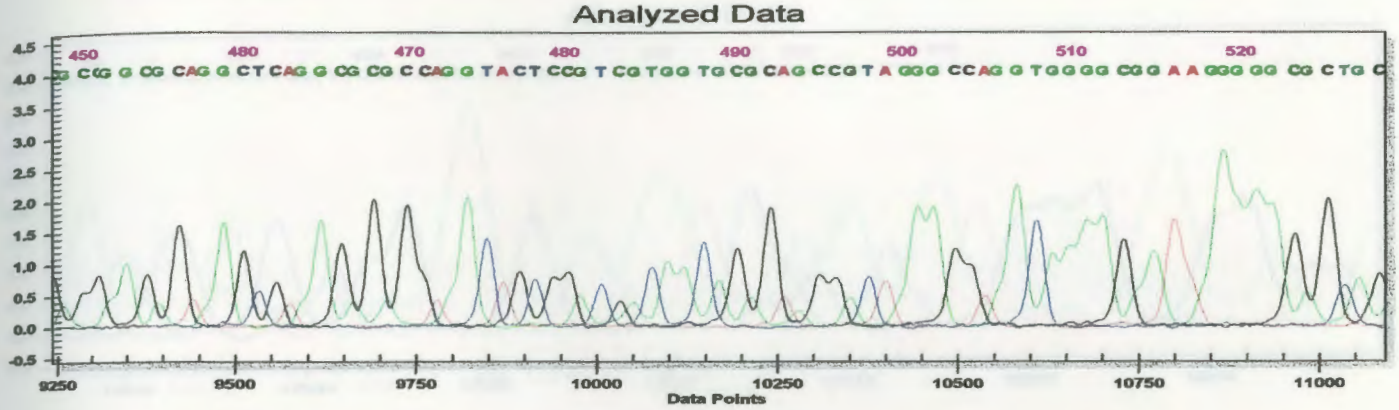
Analyzed Data



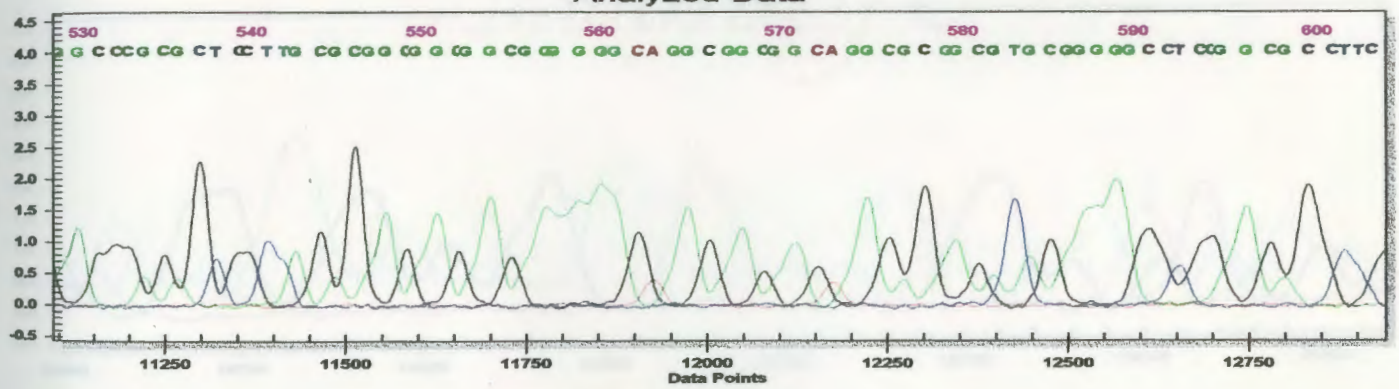
Analyzed Data



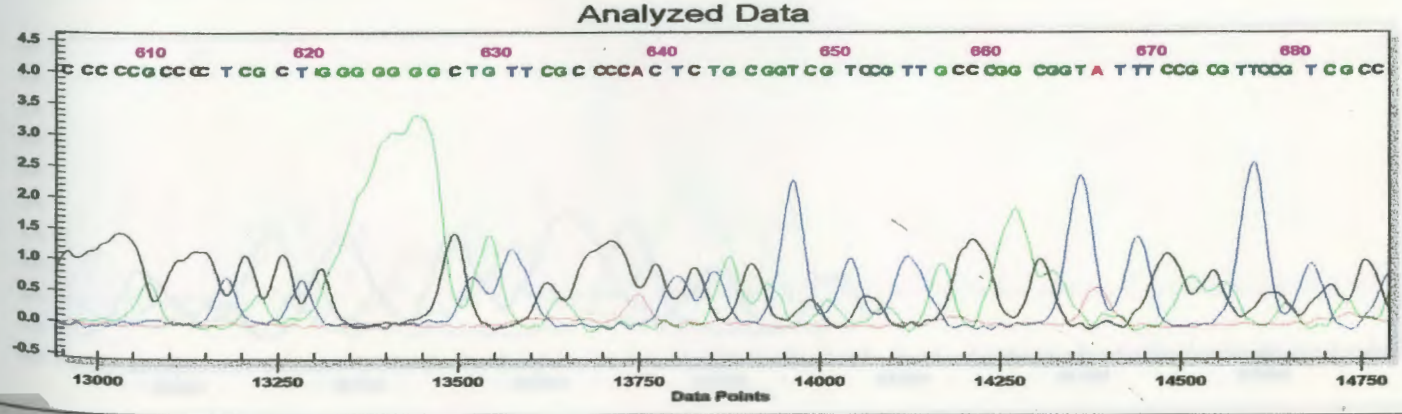
Analyzed Data



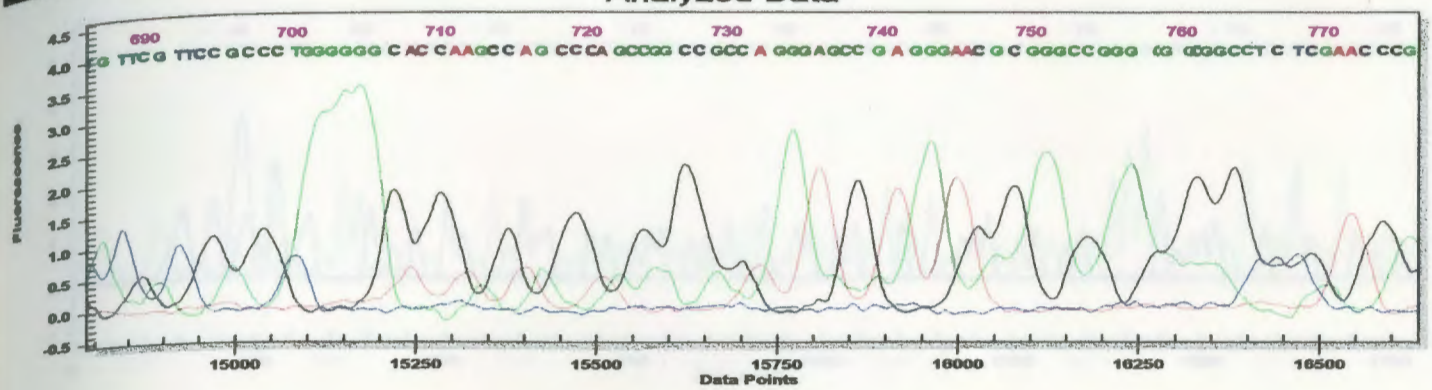
Analyzed Data



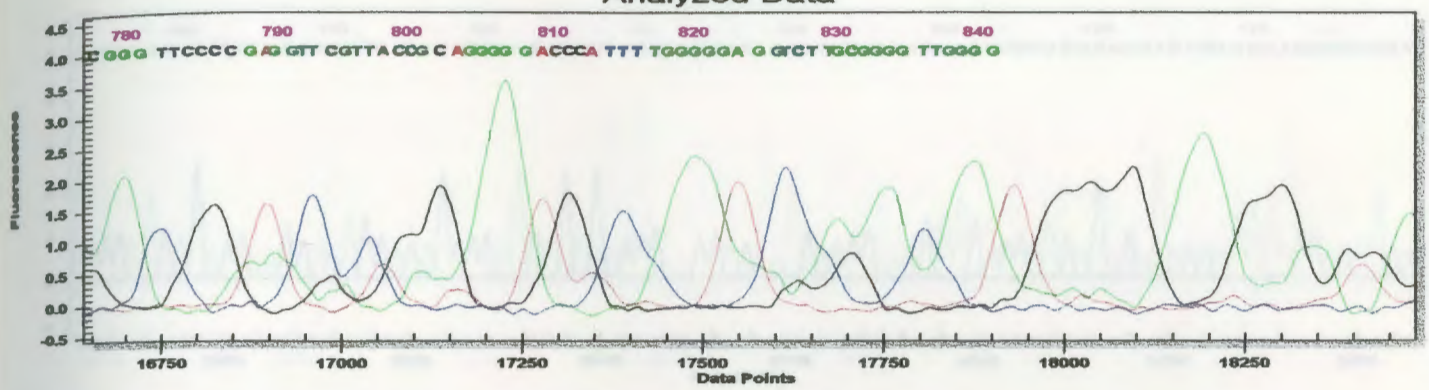
Analyzed Data



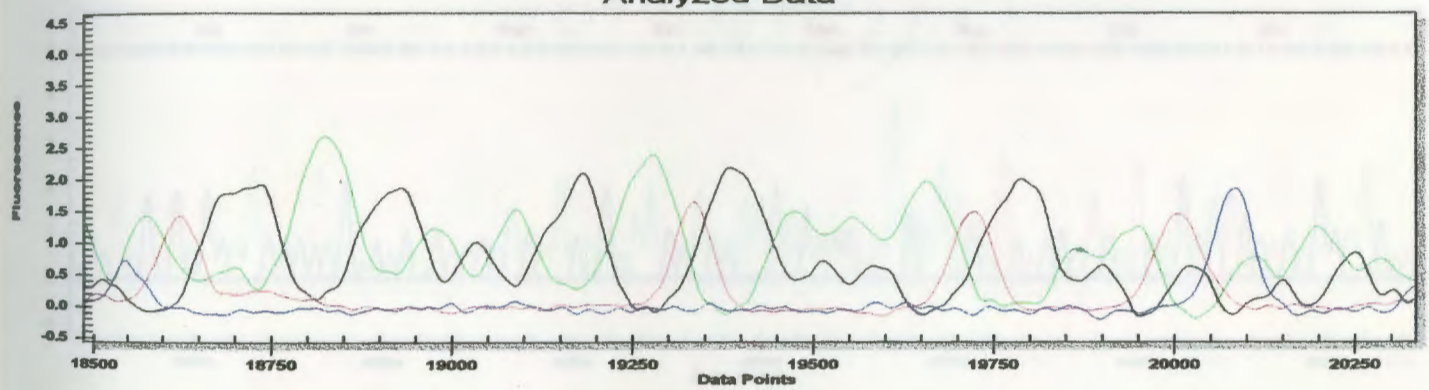
Analyzed Data



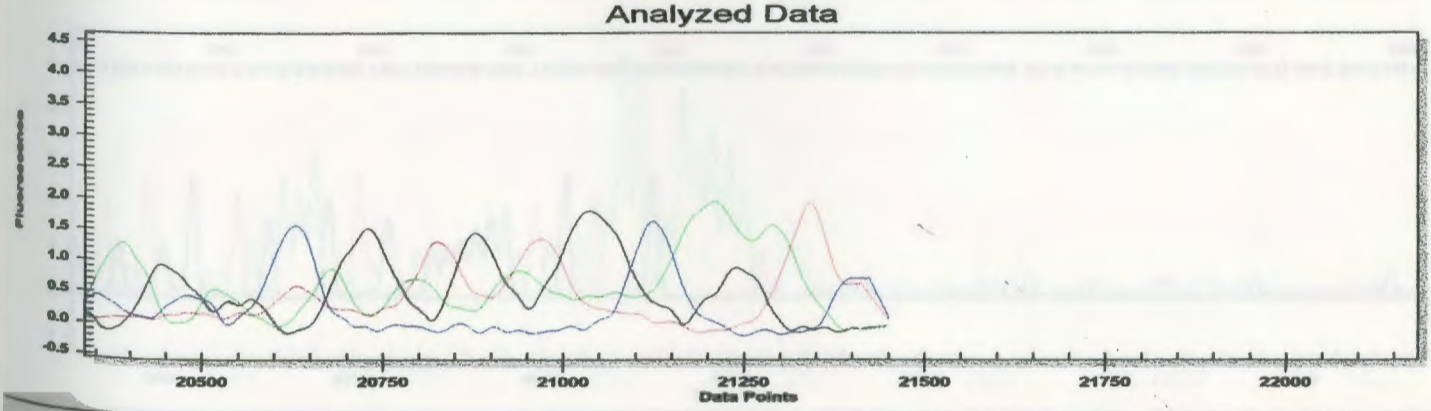
Analyzed Data



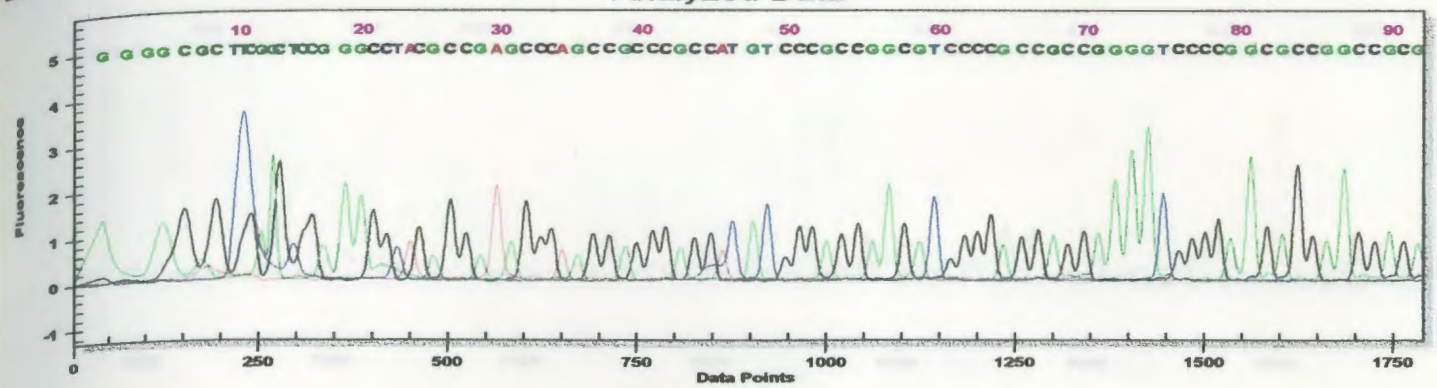
Analyzed Data



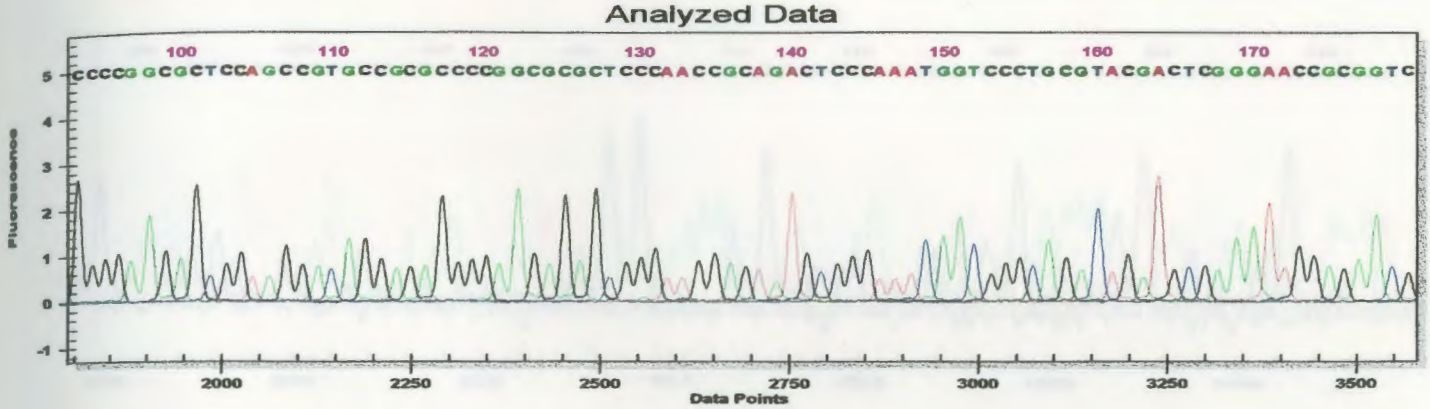
Analyzed Data



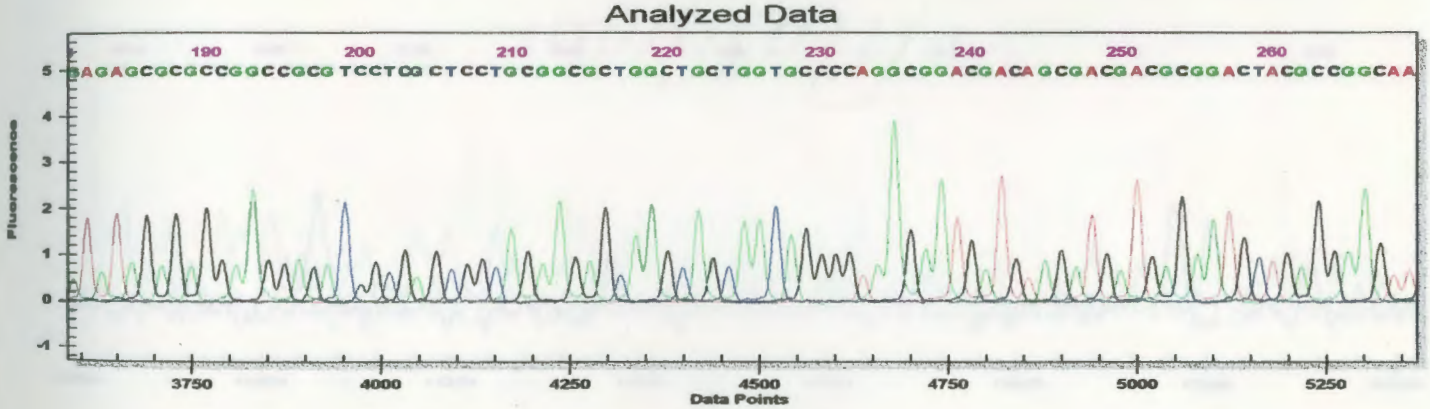
Analyzed Data



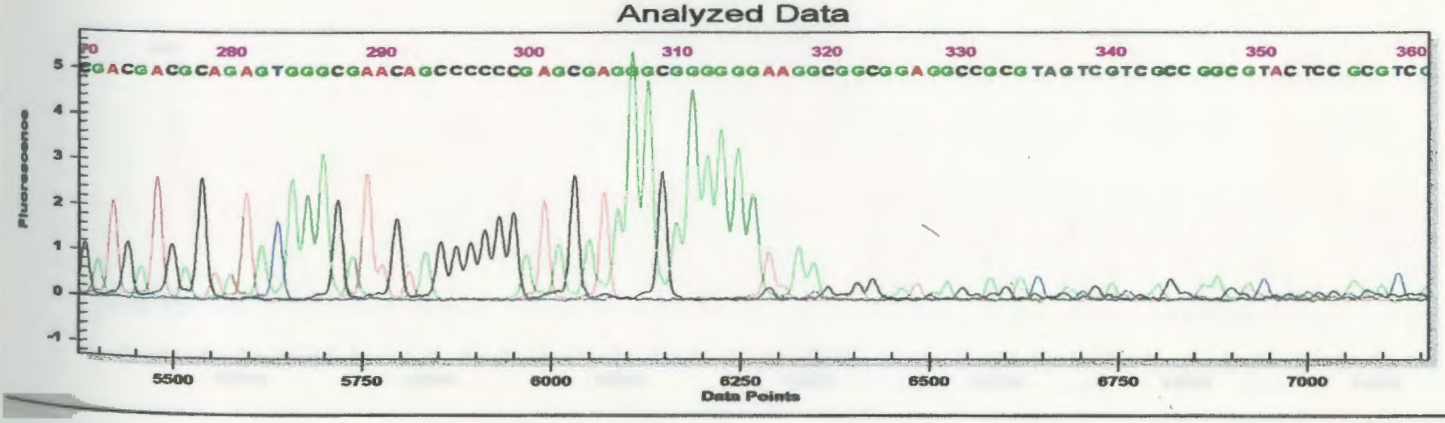
Analyzed Data



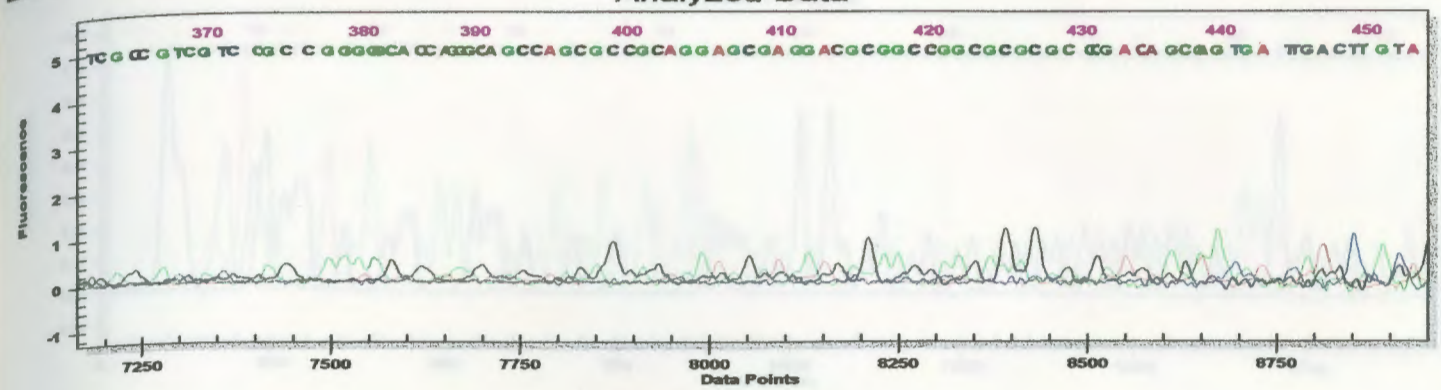
Analyzed Data



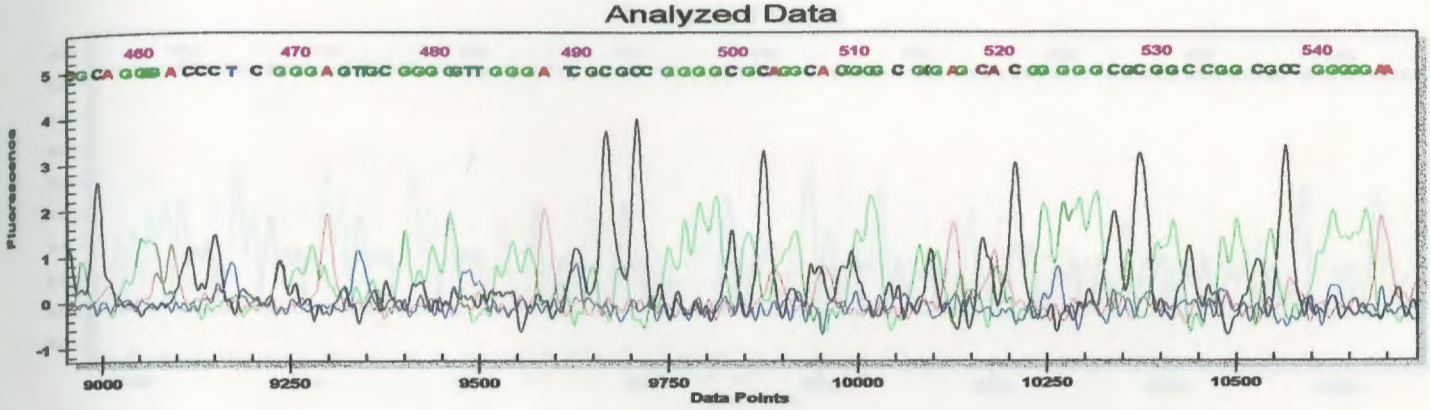
Analyzed Data



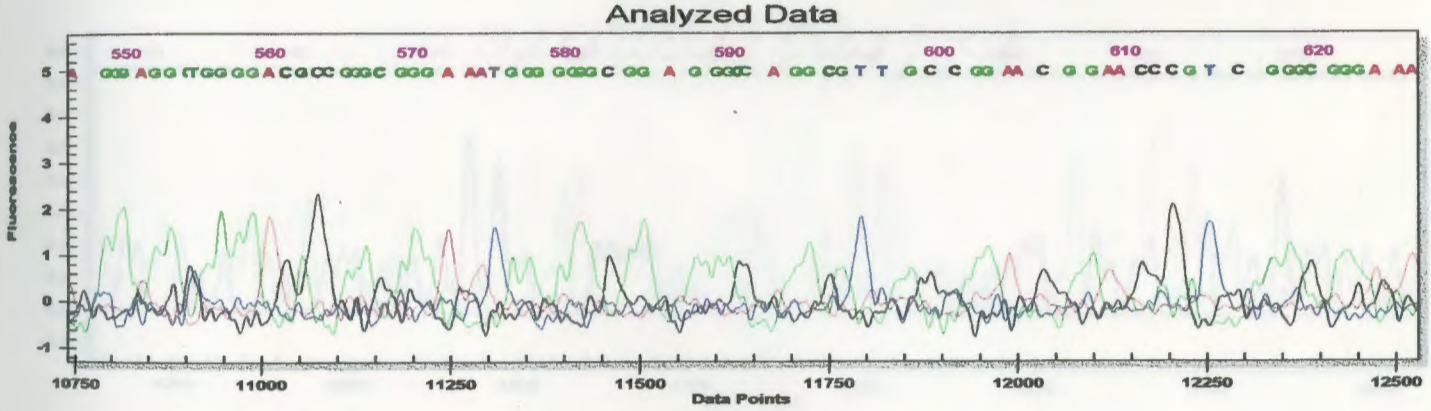
Analyzed Data



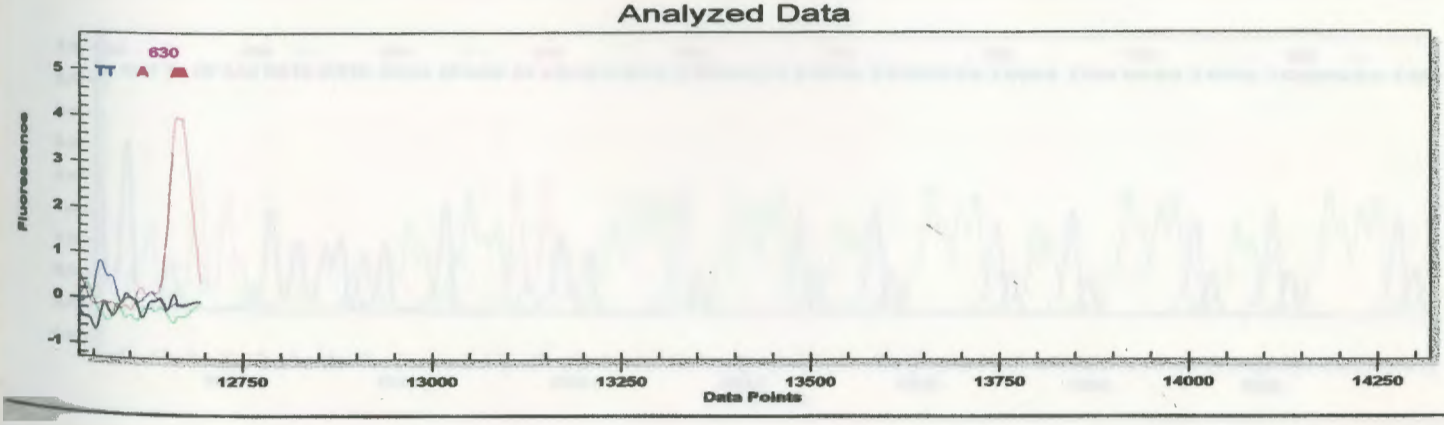
Analyzed Data



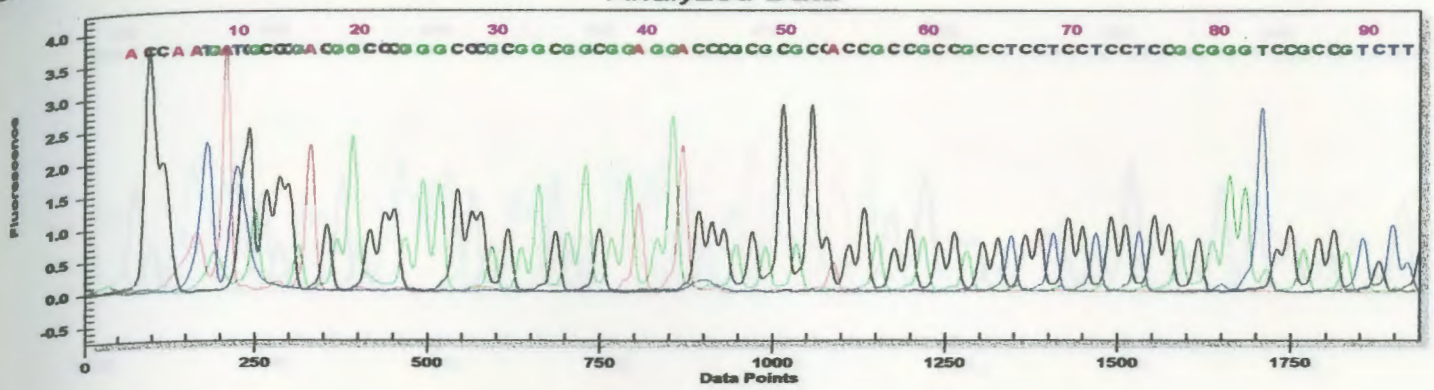
Analyzed Data



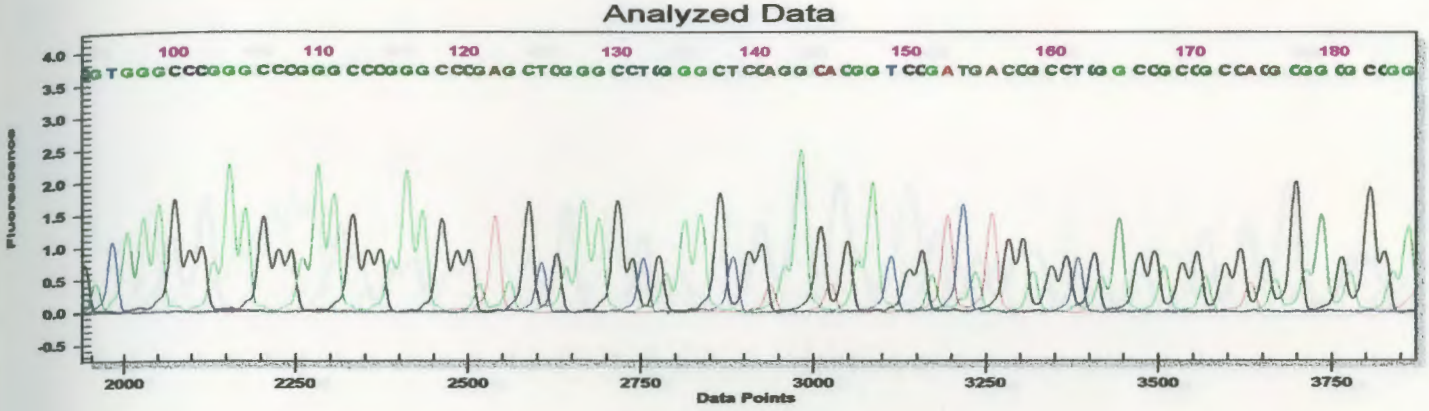
Analyzed Data



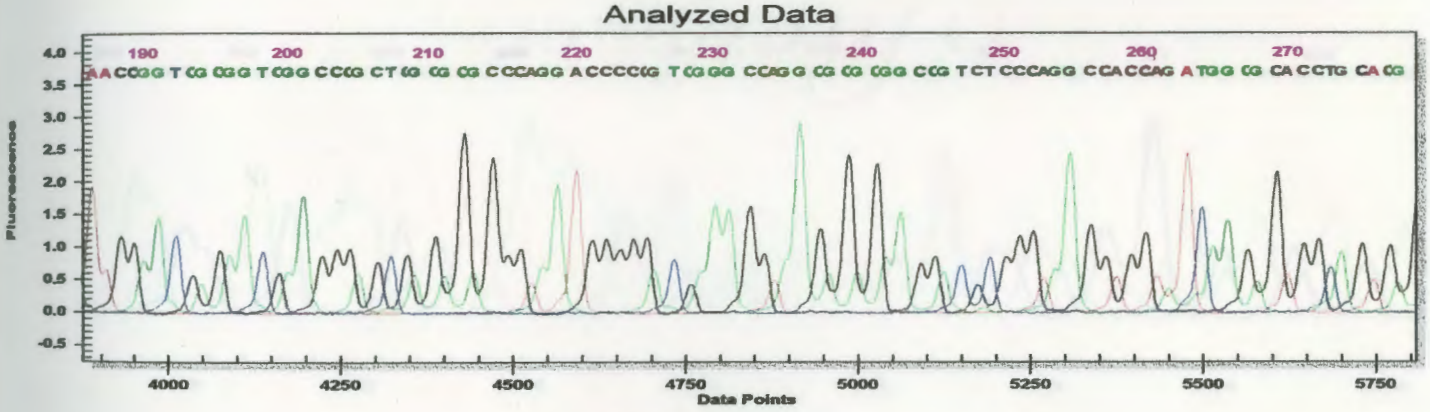
Analyzed Data



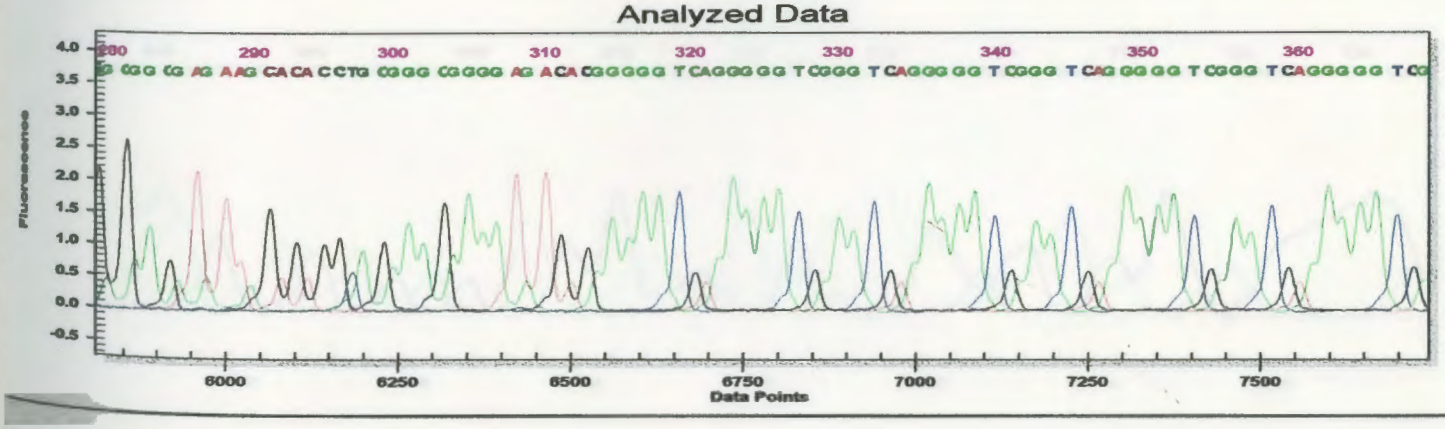
Analyzed Data



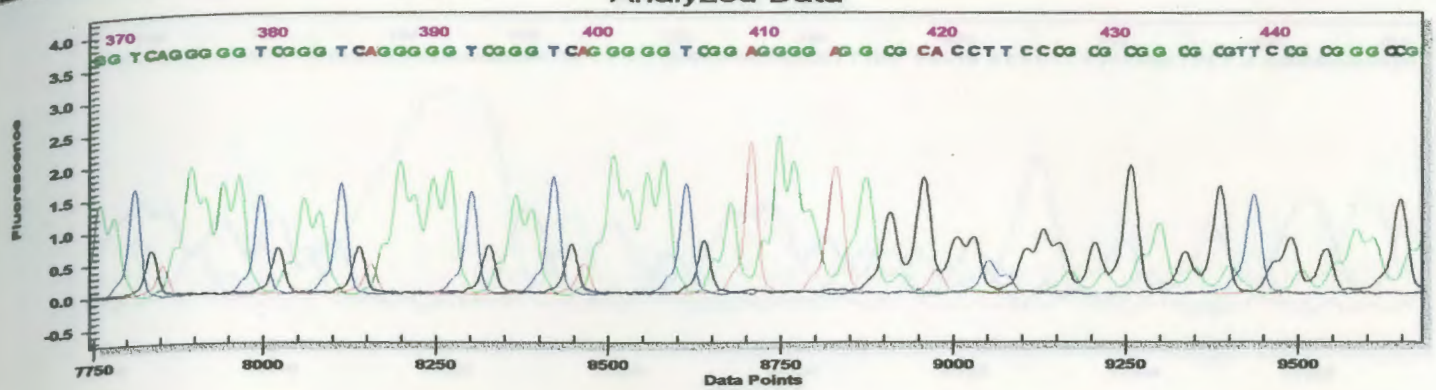
Analyzed Data



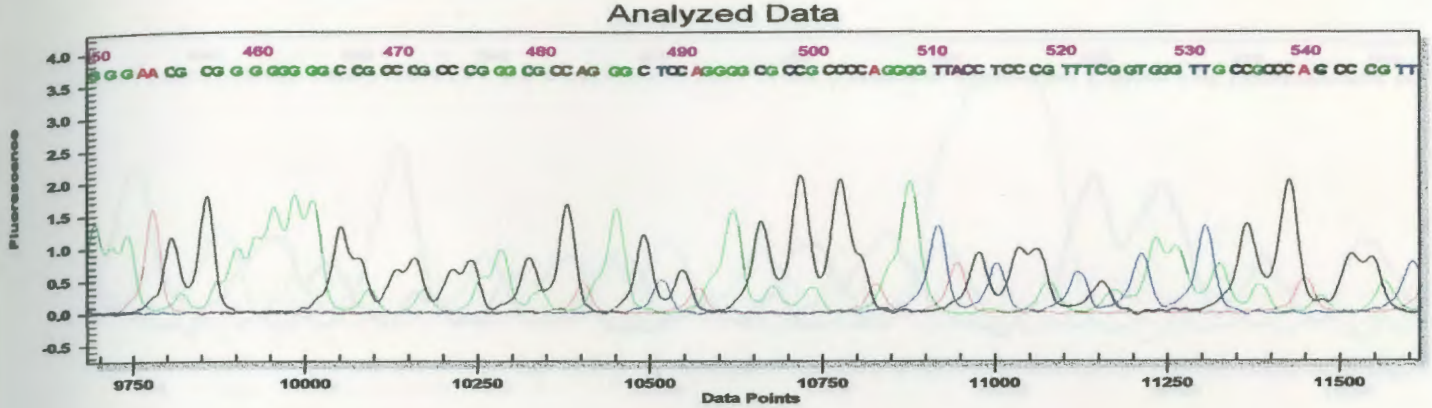
Analyzed Data



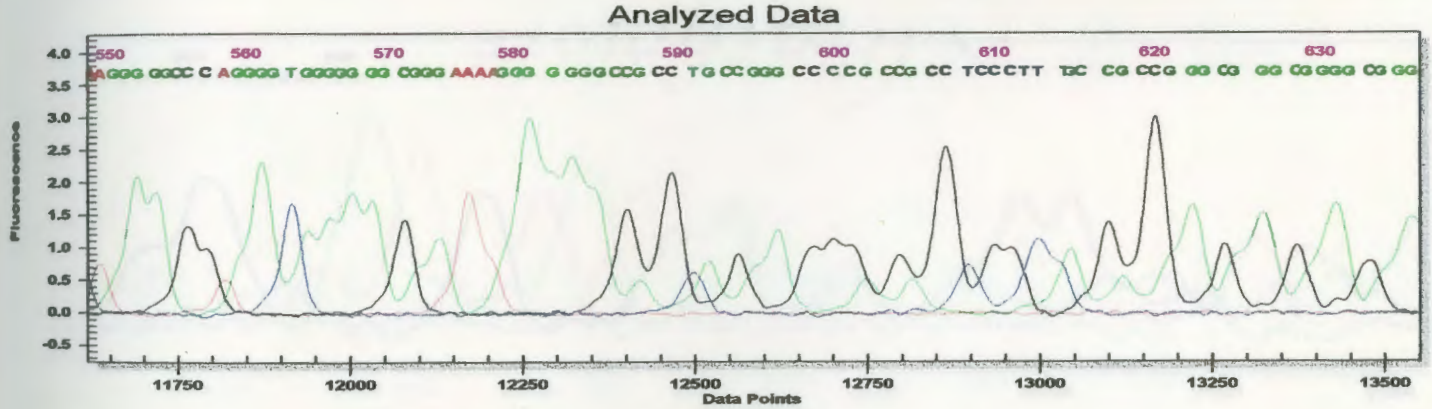
Analyzed Data



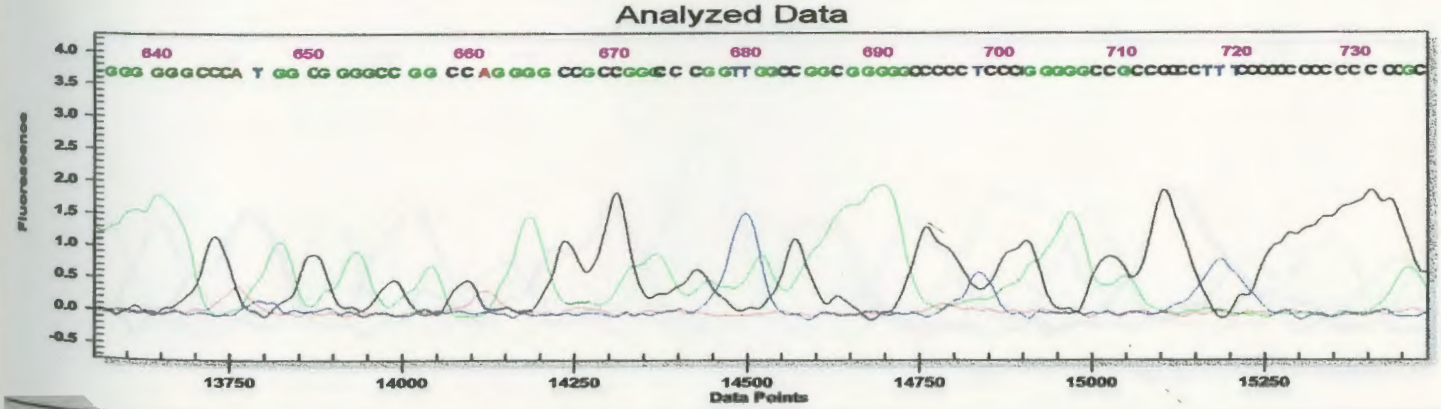
Analyzed Data



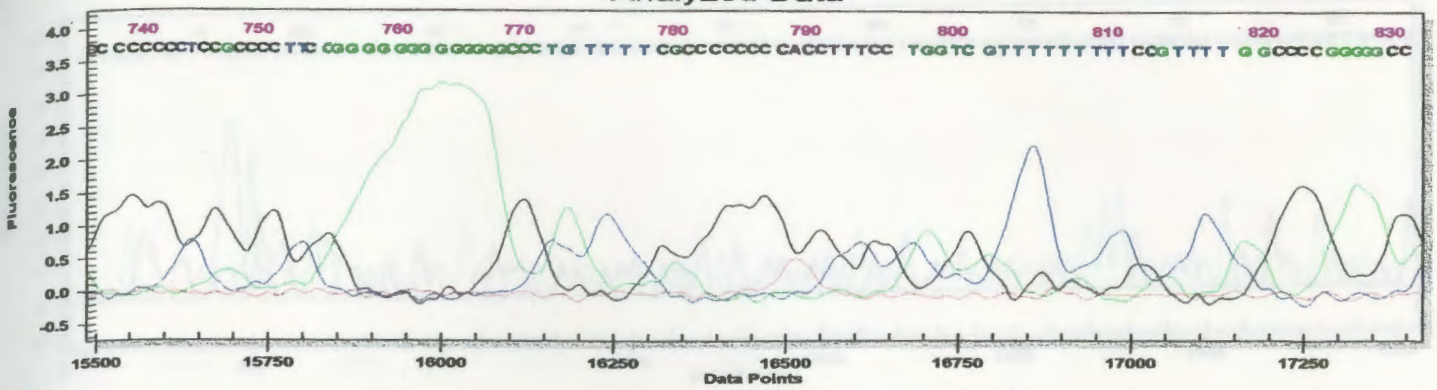
Analyzed Data



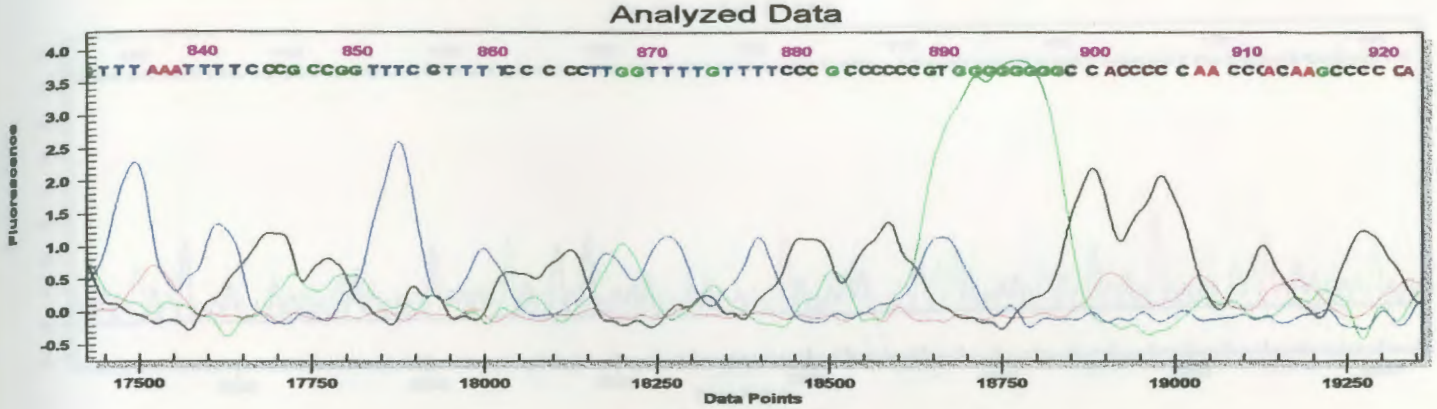
Analyzed Data



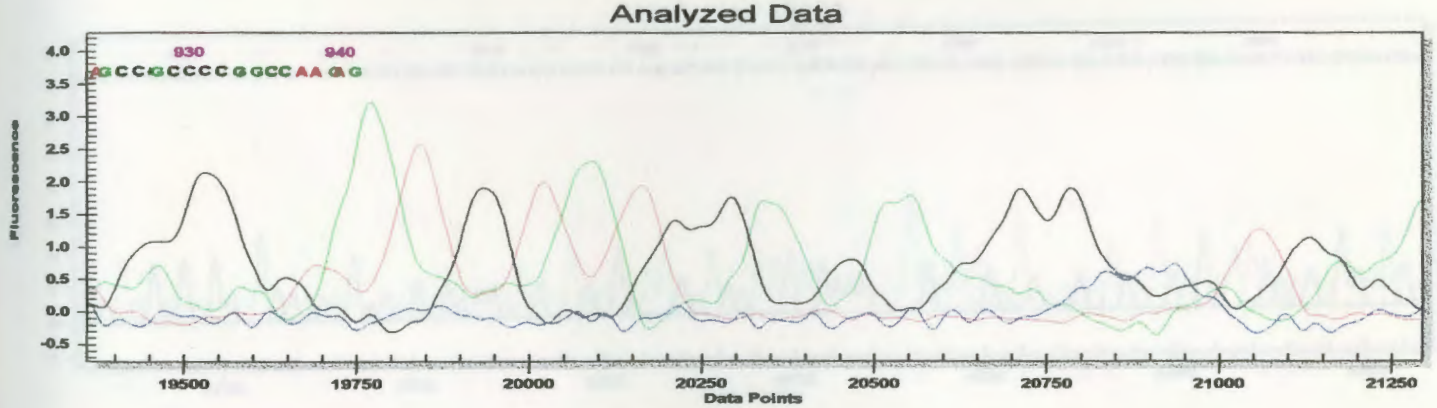
Analyzed Data



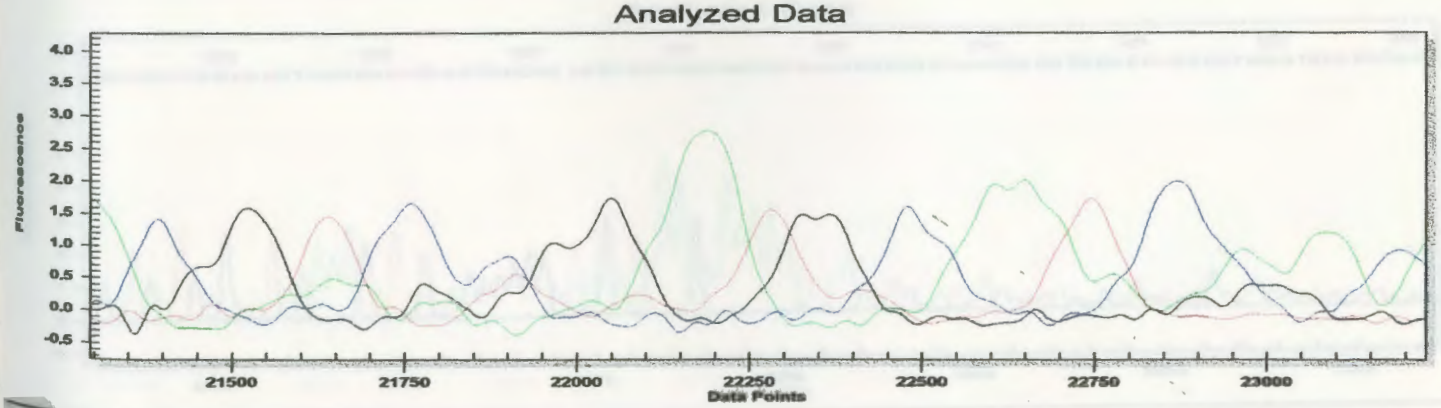
Analyzed Data



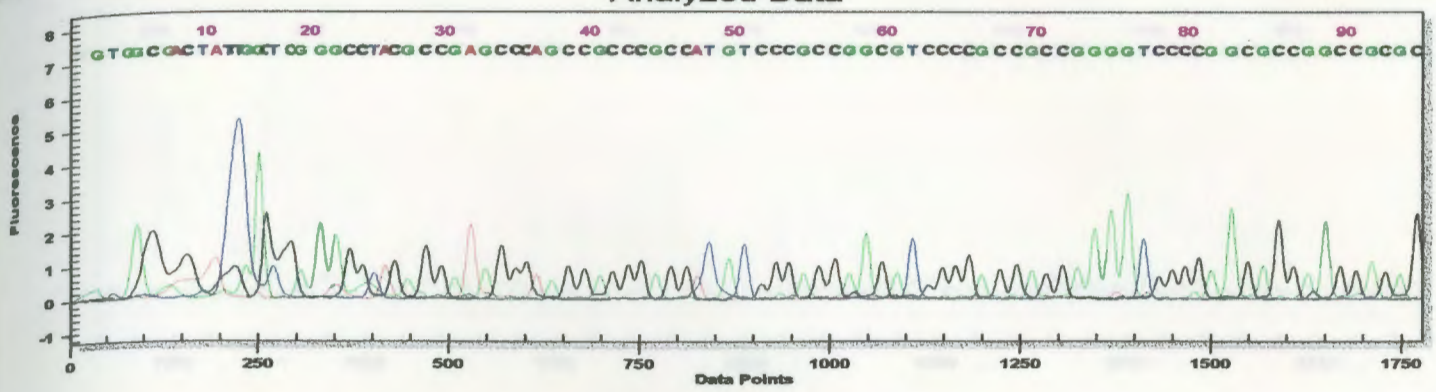
Analyzed Data



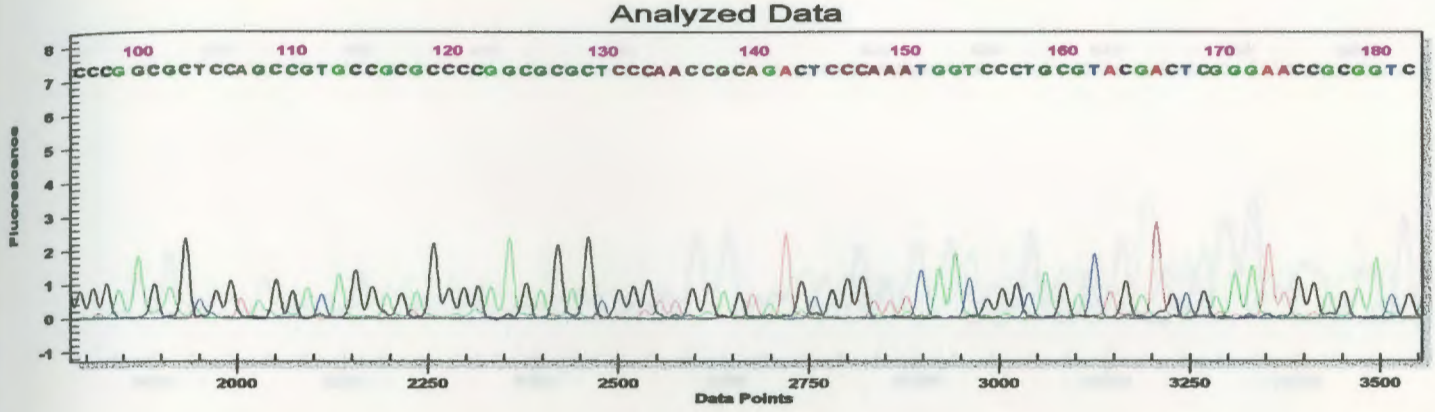
Analyzed Data



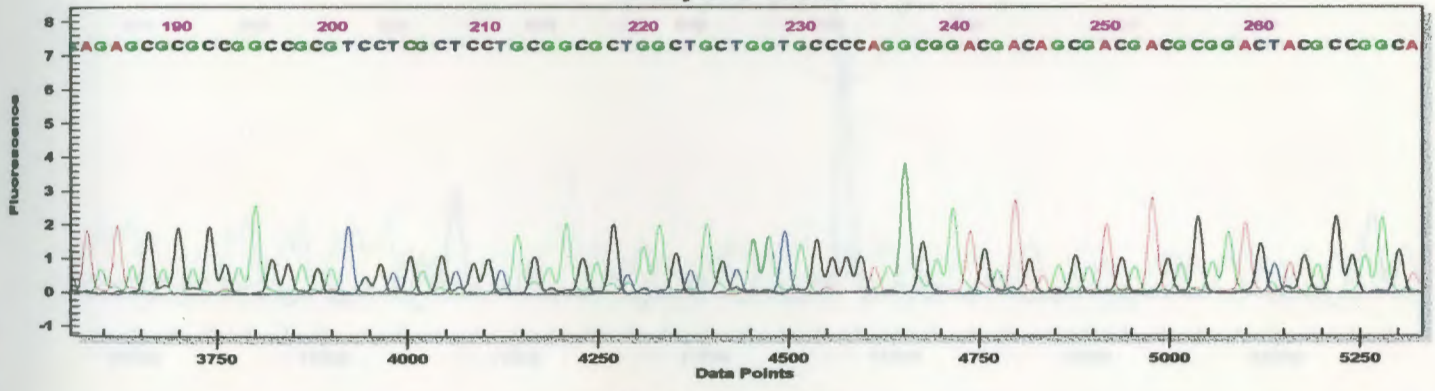
Analyzed Data



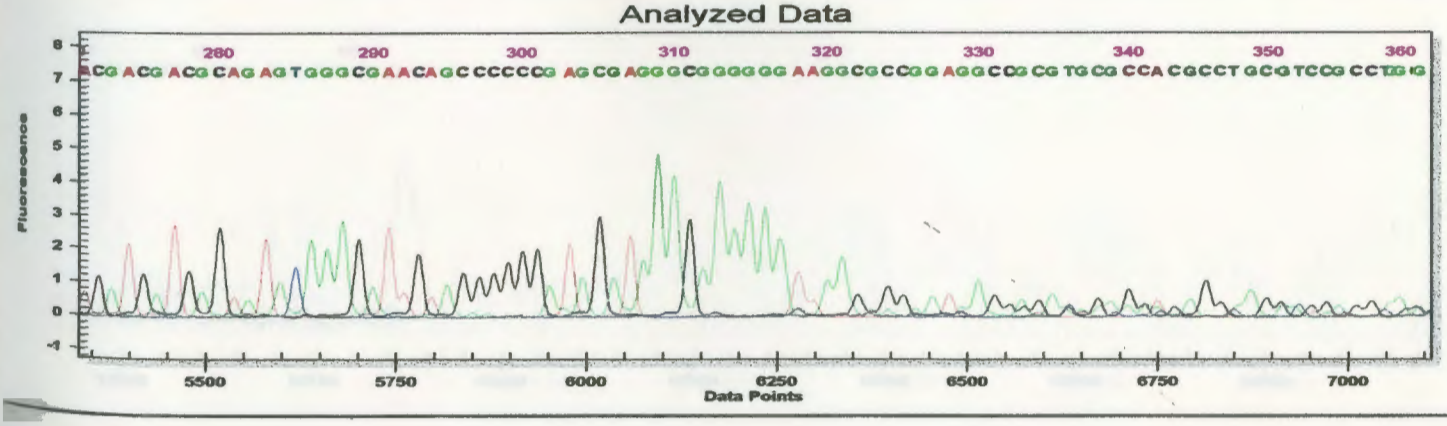
Analyzed Data



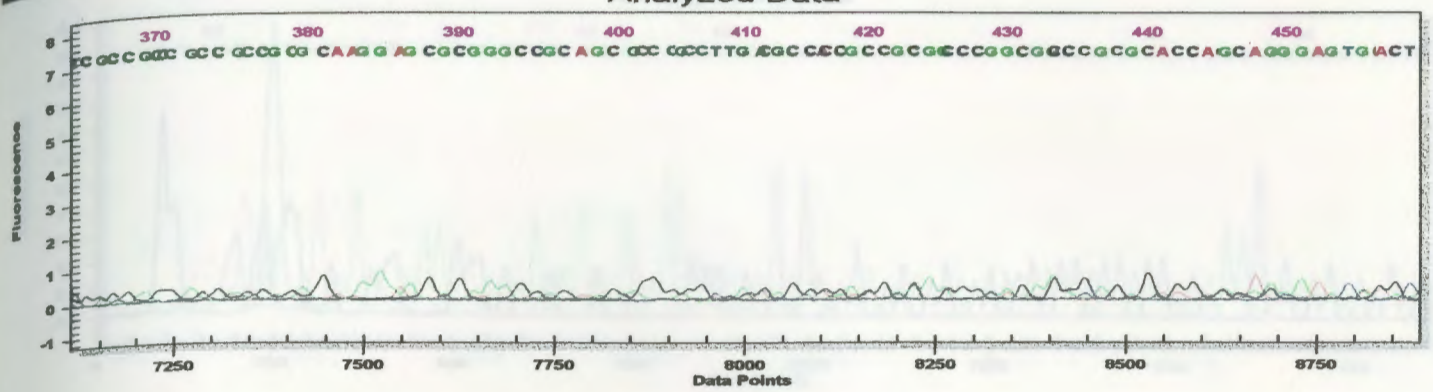
Analyzed Data



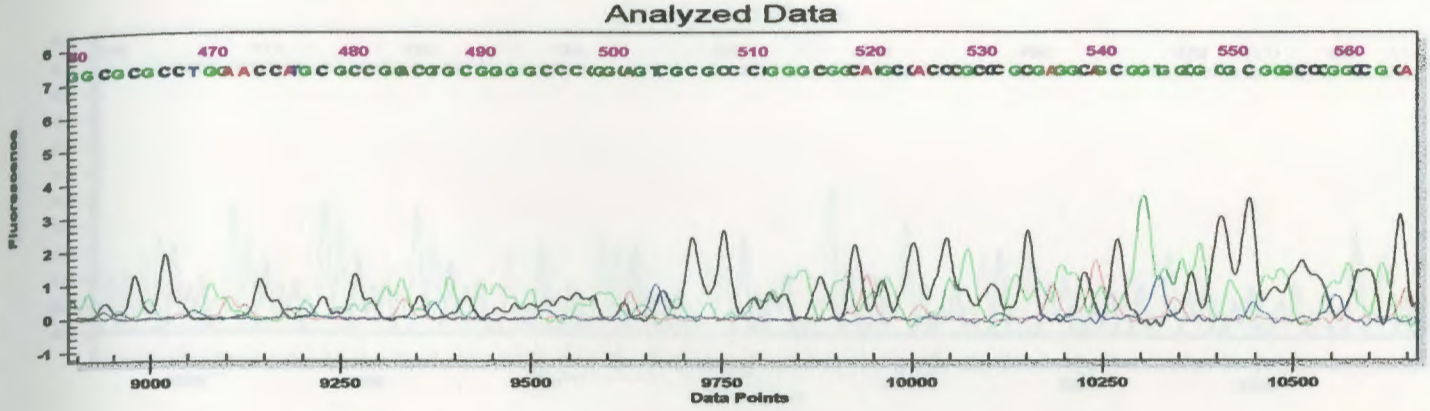
Analyzed Data



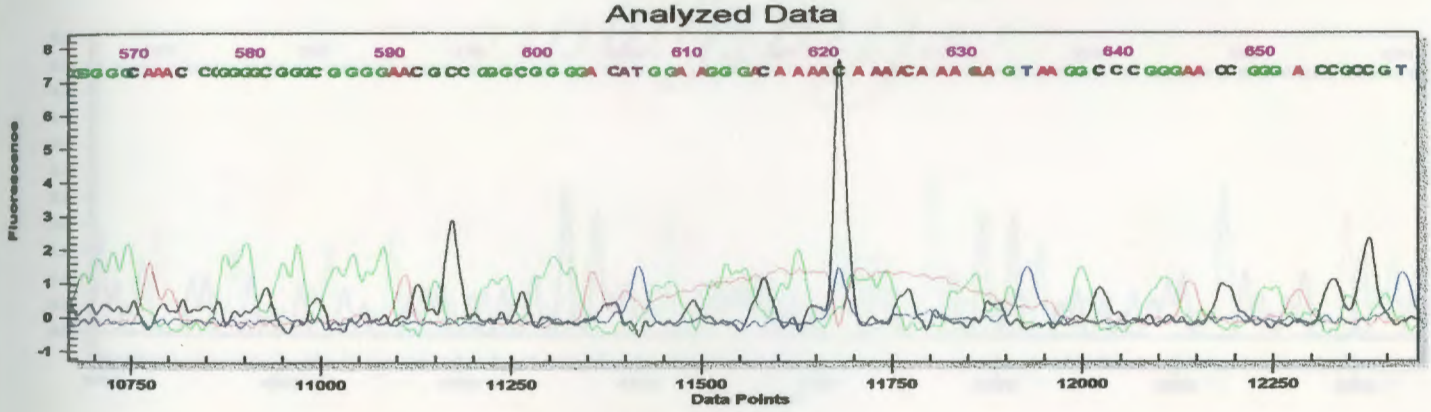
Analyzed Data



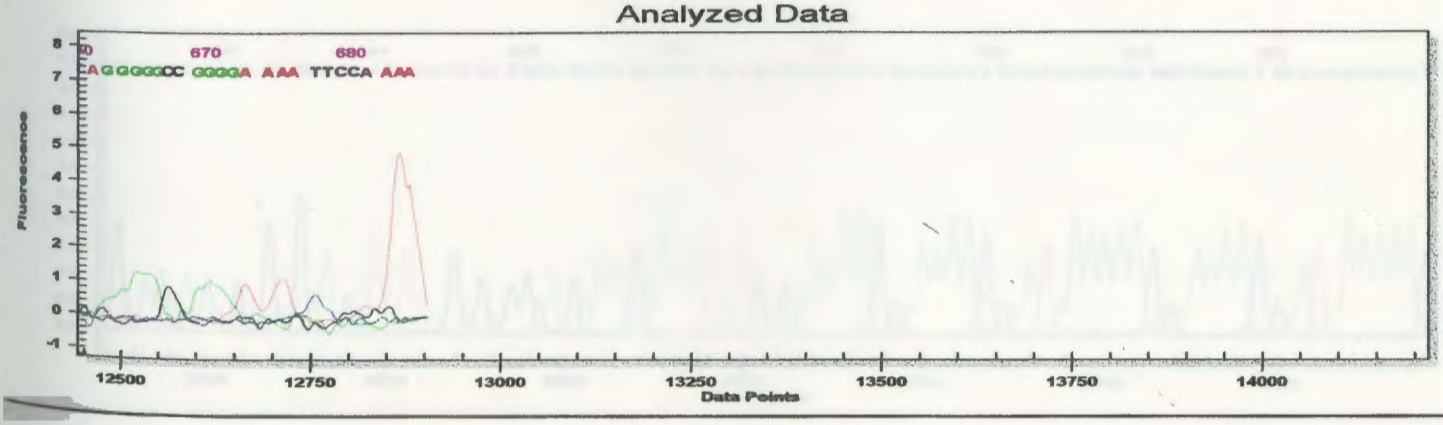
Analyzed Data



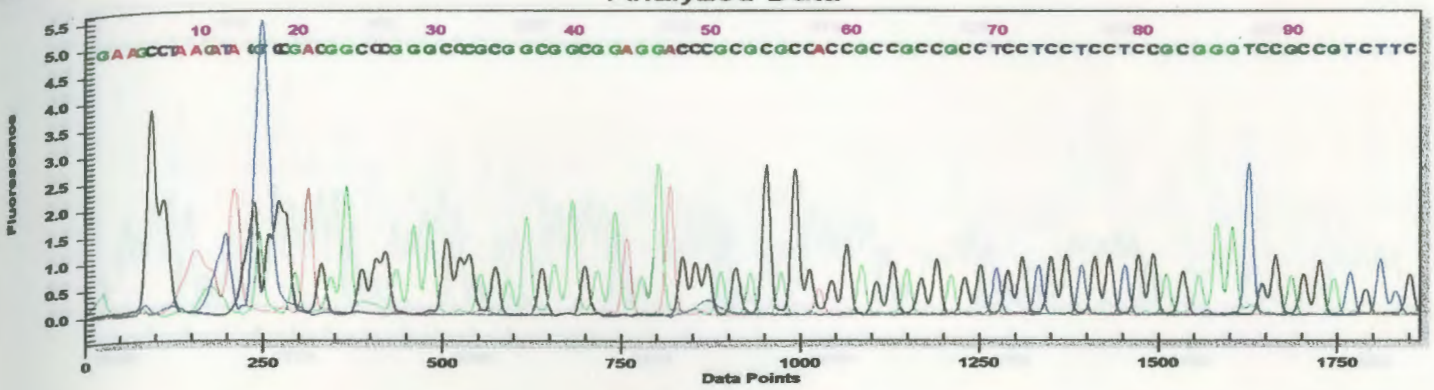
Analyzed Data



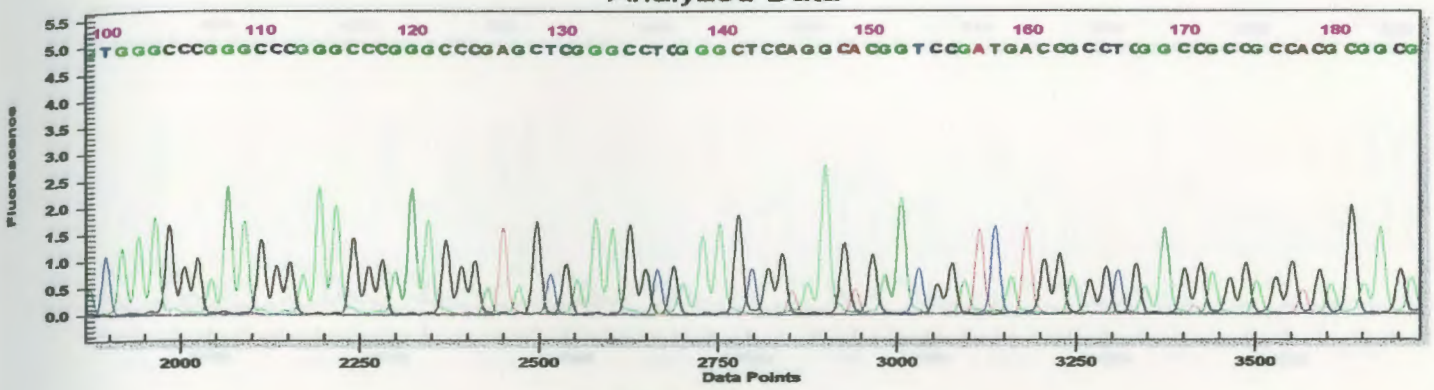
Analyzed Data



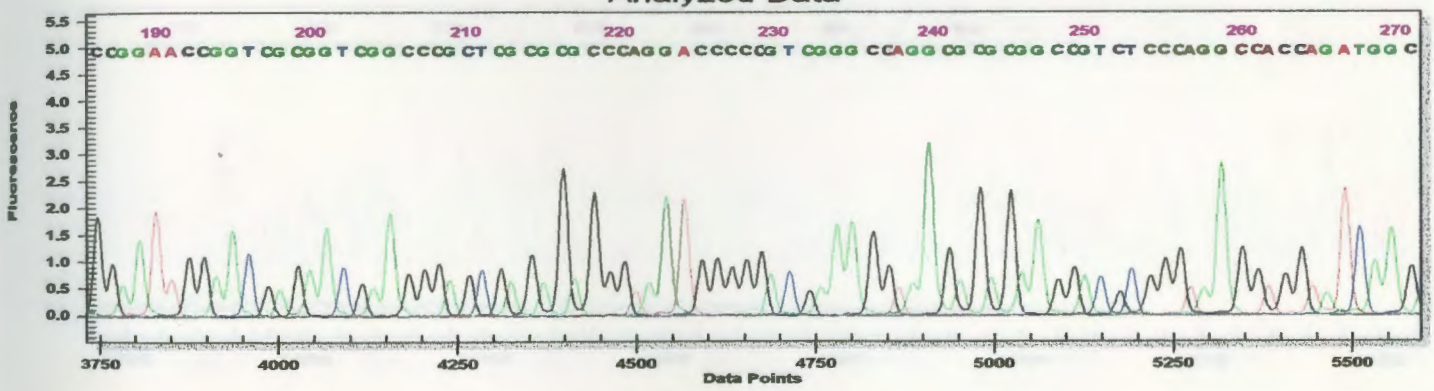
Analyzed Data



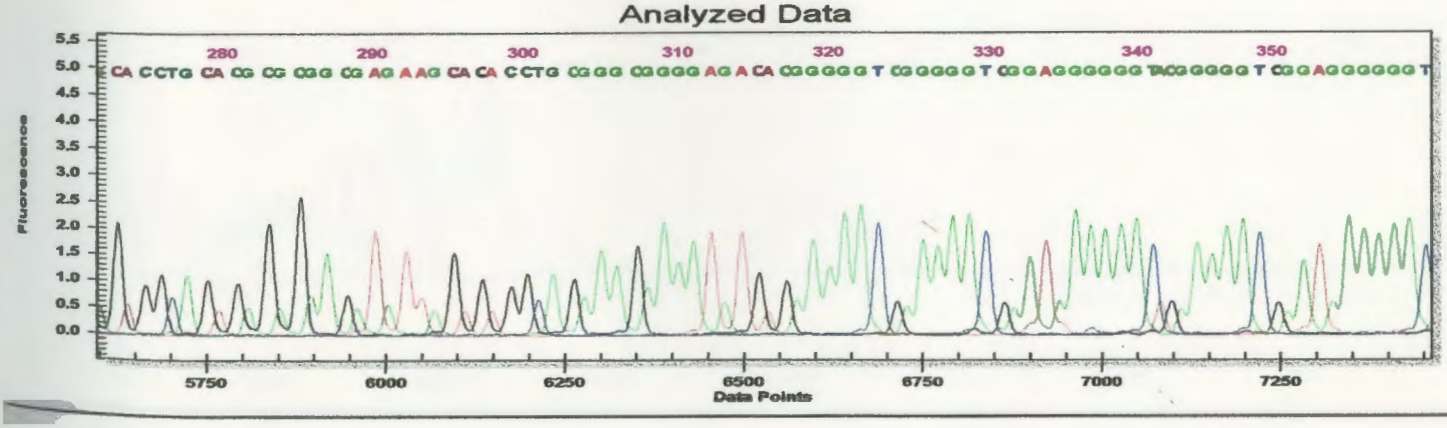
Analyzed Data



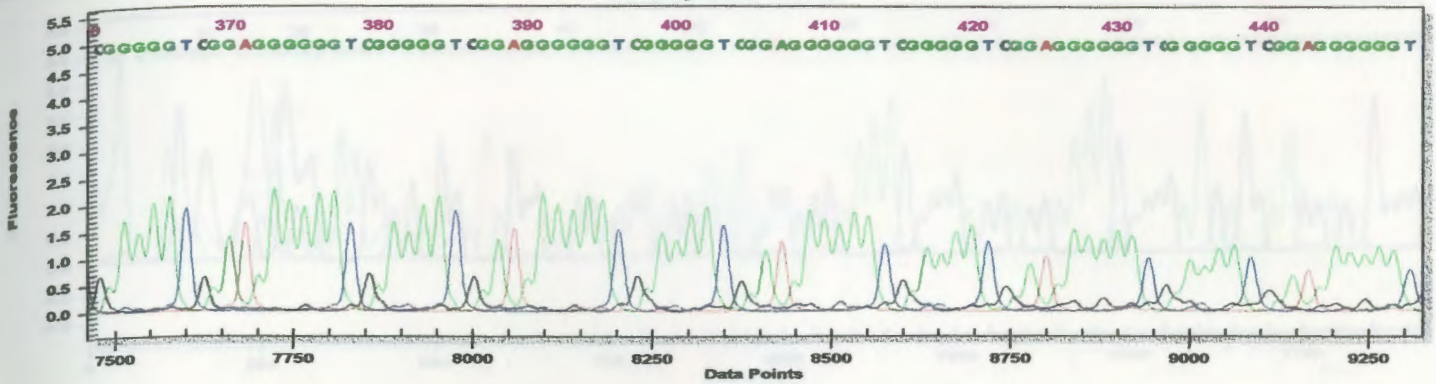
Analyzed Data



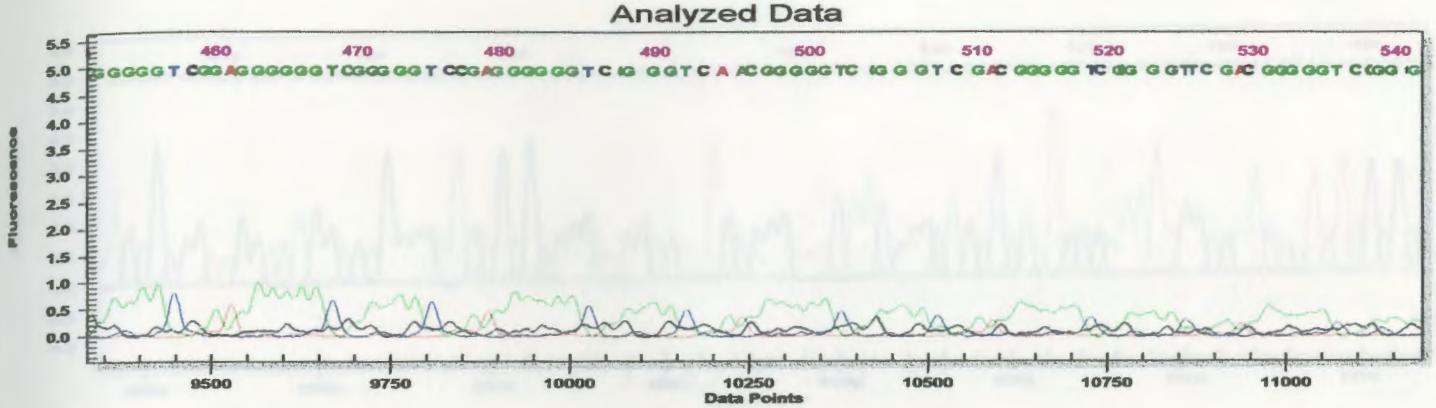
Analyzed Data



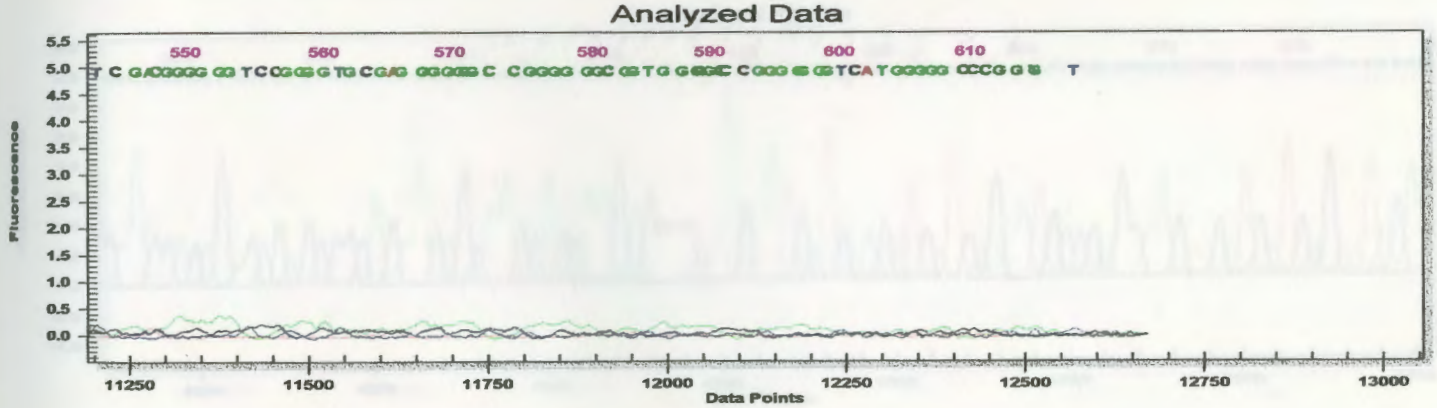
Analyzed Data



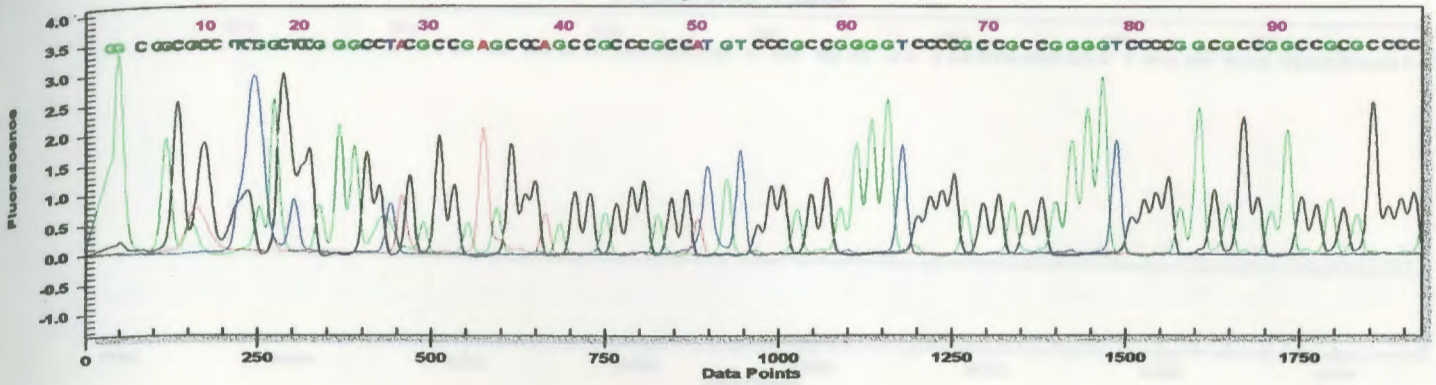
Analyzed Data



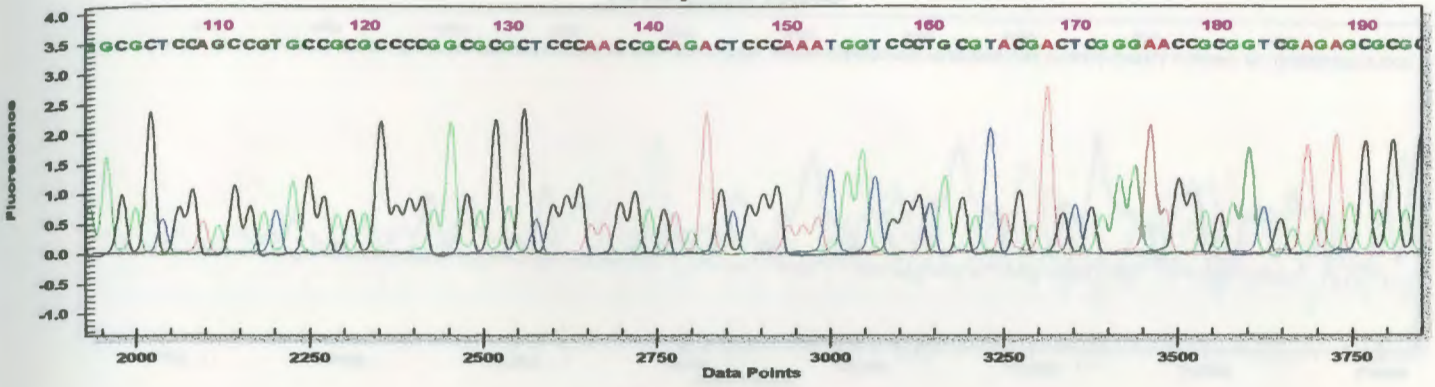
Analyzed Data



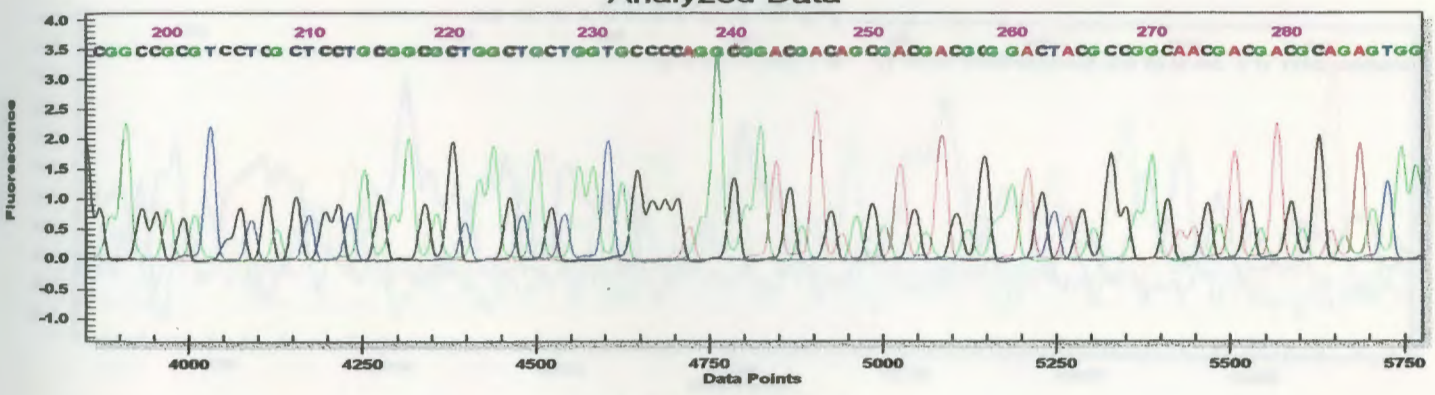
Analyzed Data



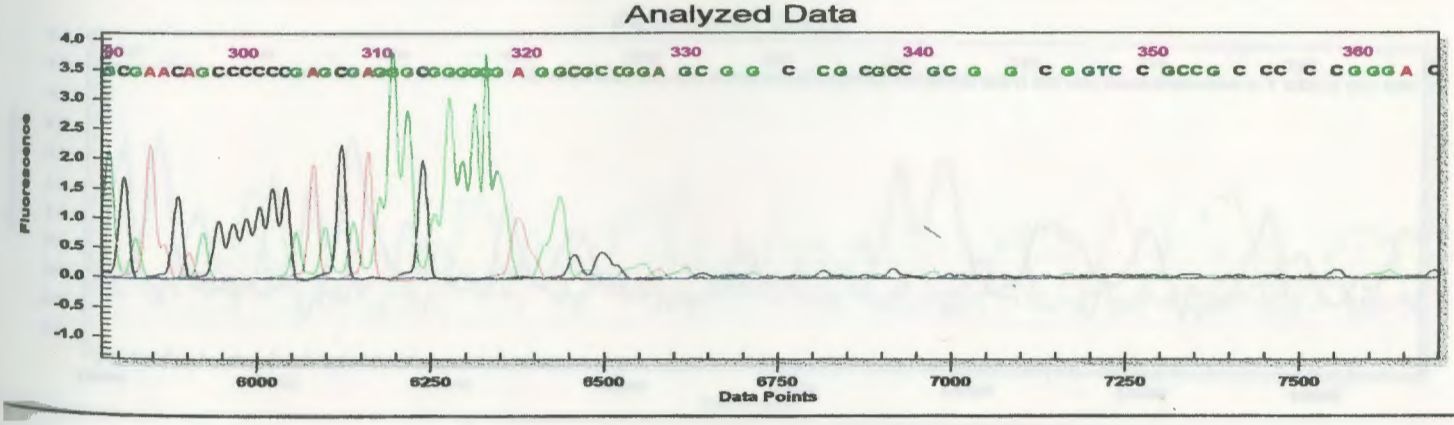
Analyzed Data



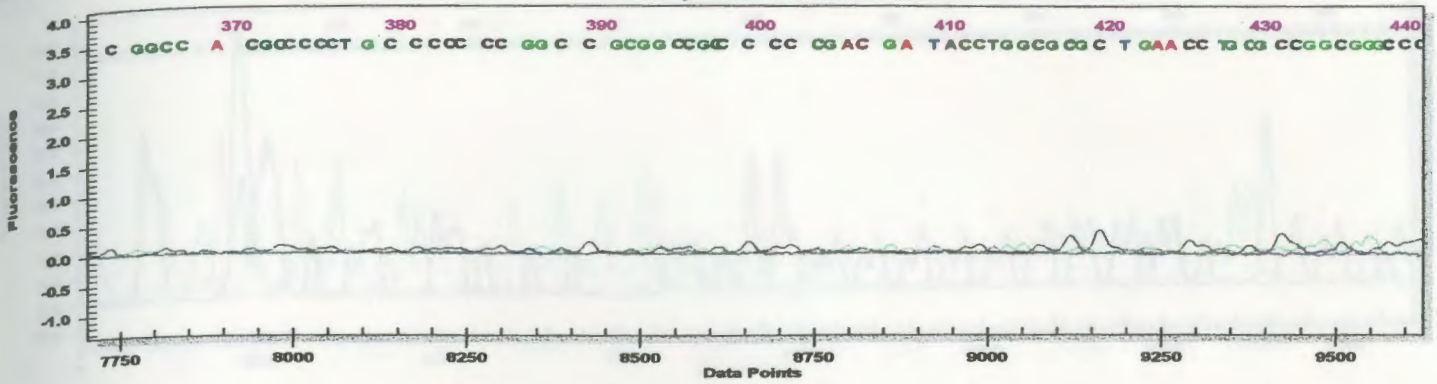
Analyzed Data



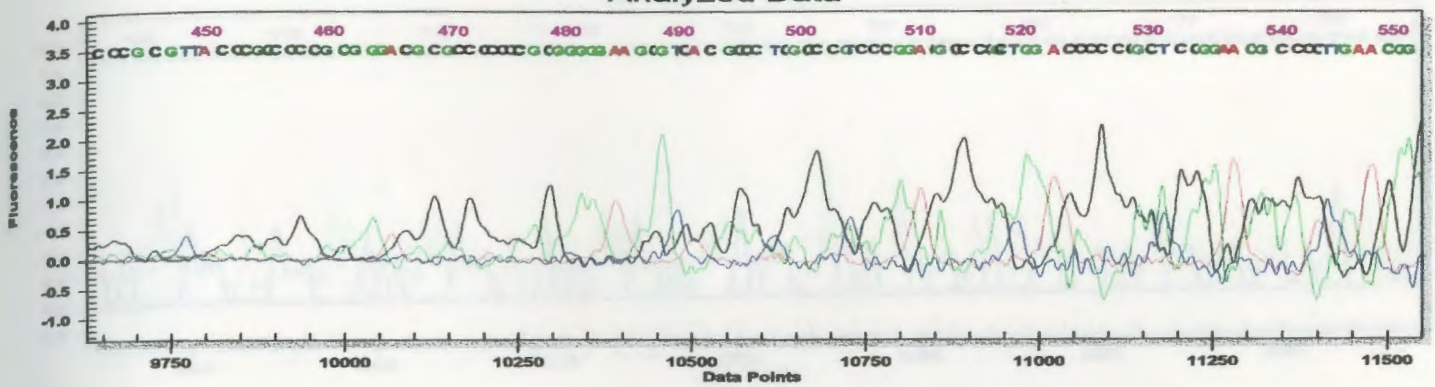
Analyzed Data



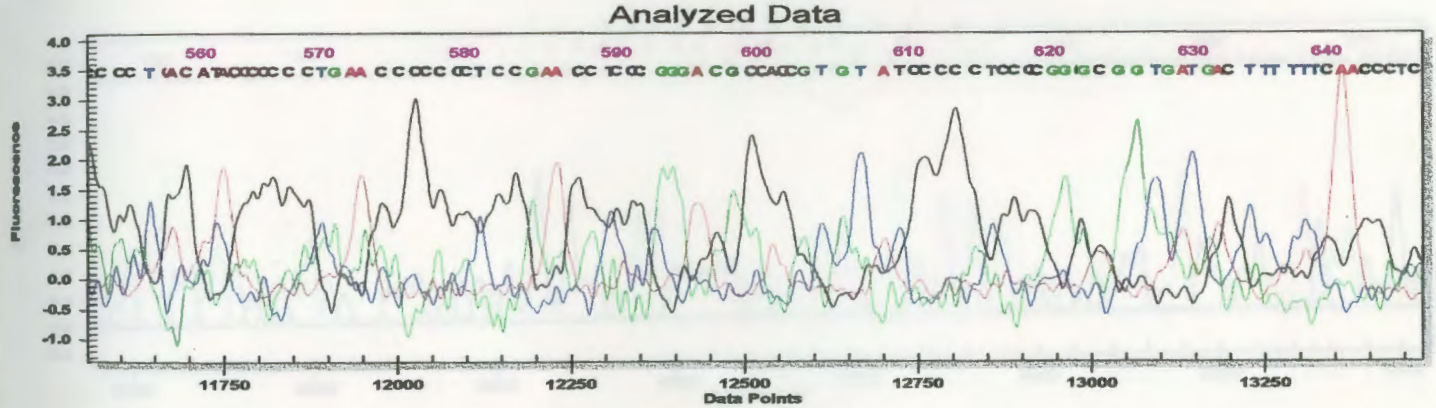
Analyzed Data



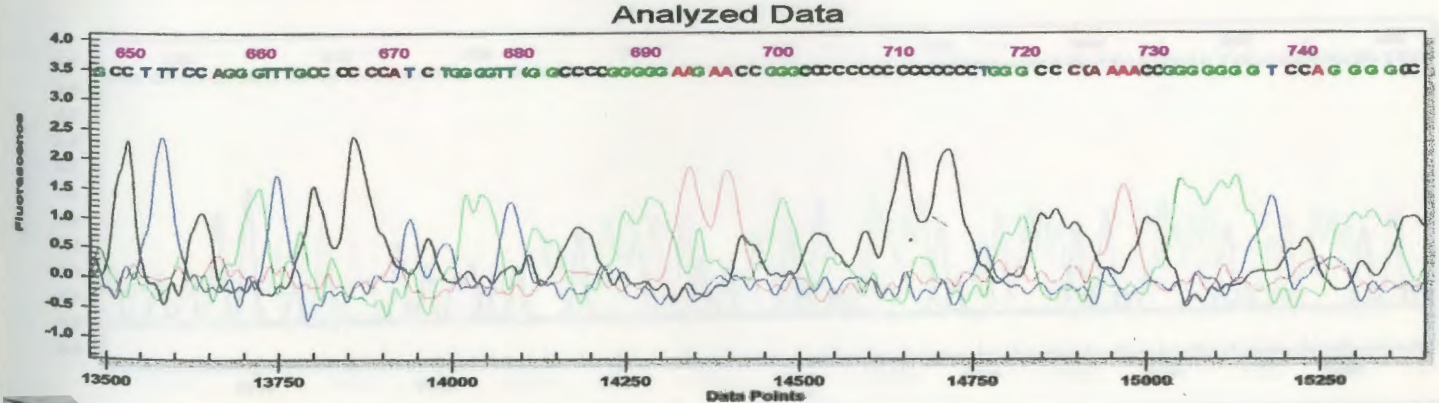
Analyzed Data



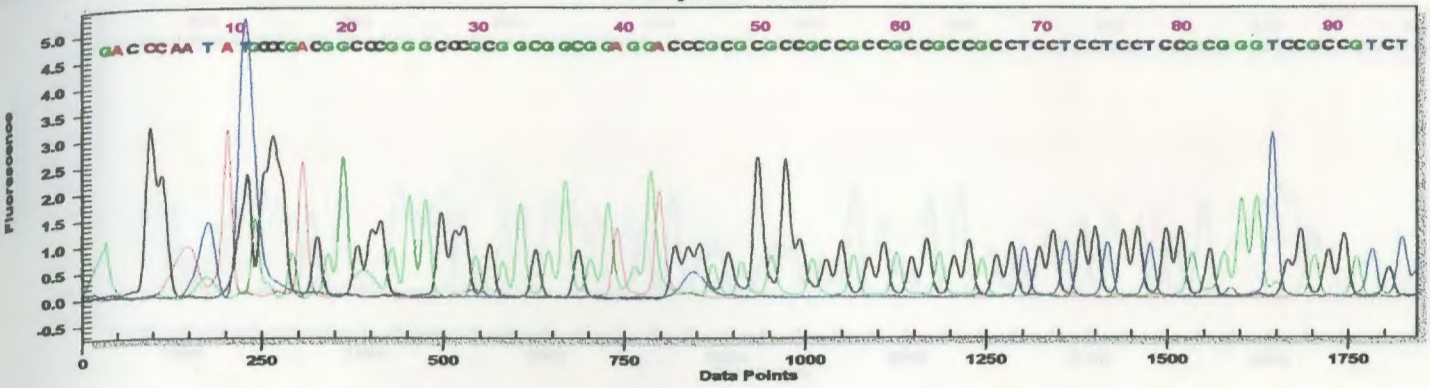
Analyzed Data



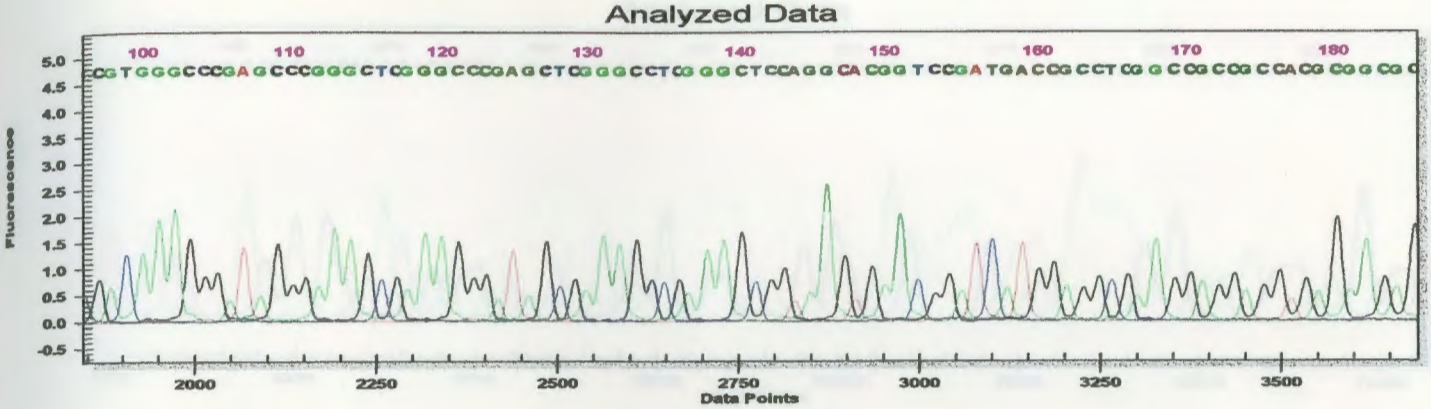
Analyzed Data



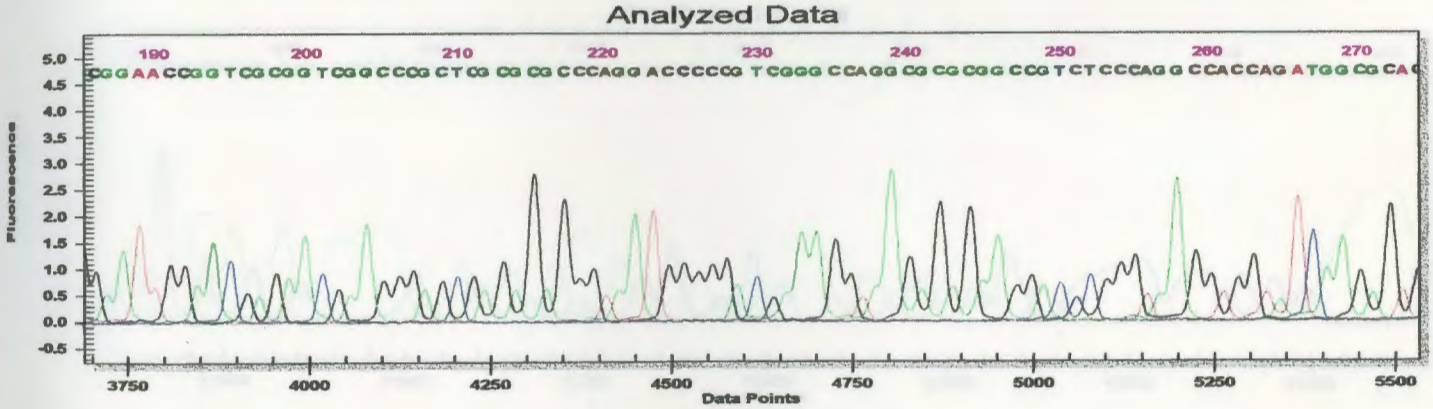
Analyzed Data



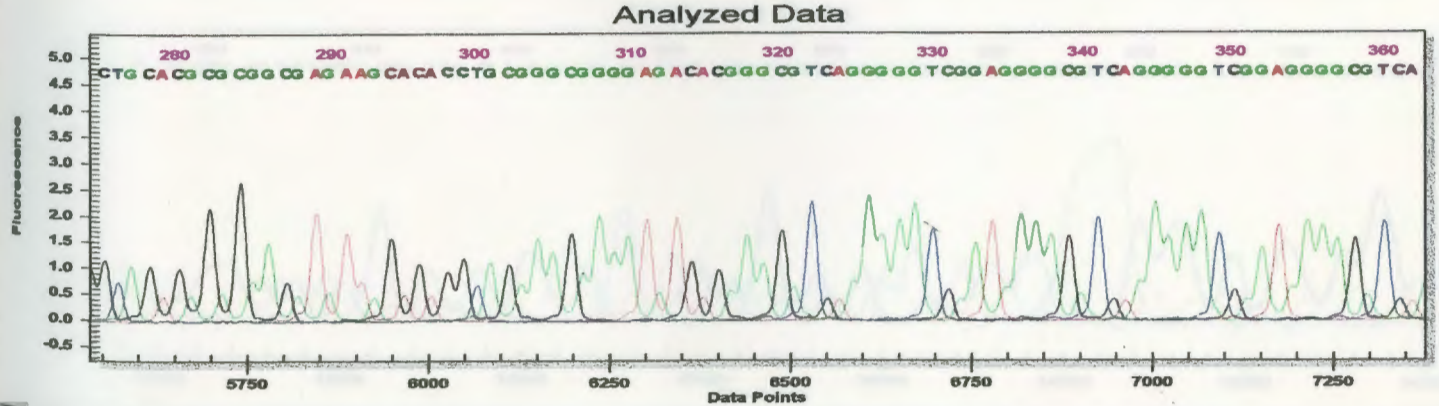
Analyzed Data



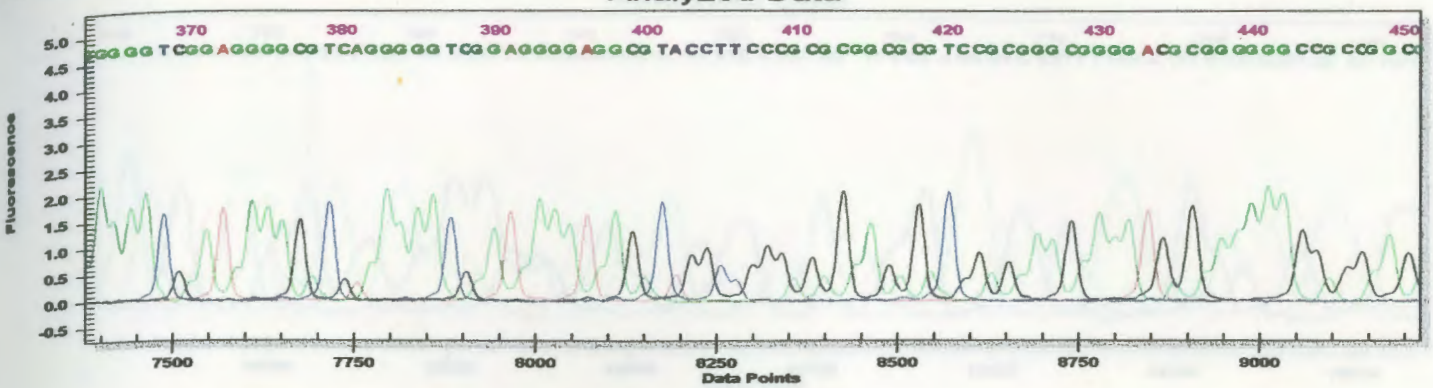
Analyzed Data



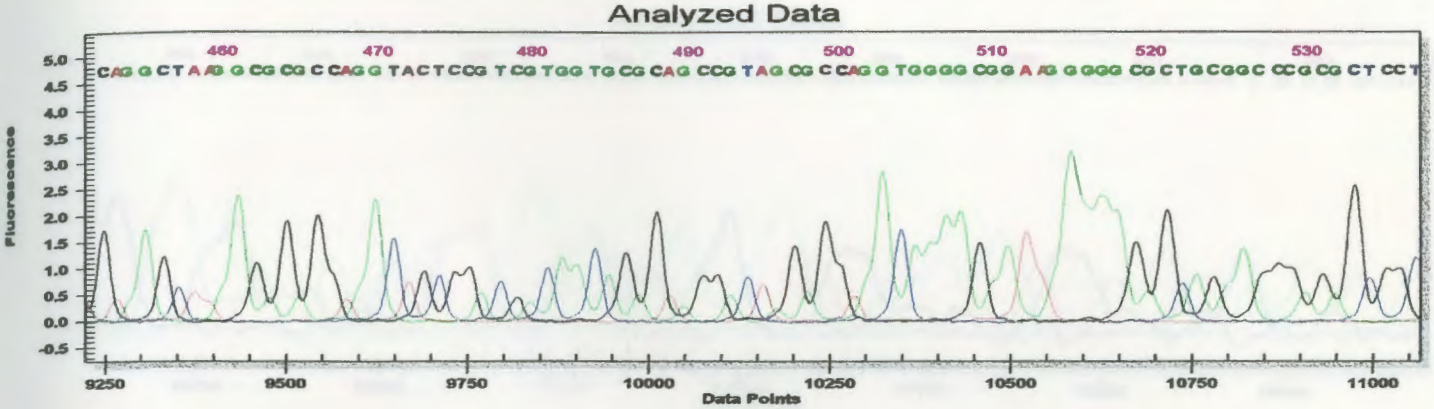
Analyzed Data



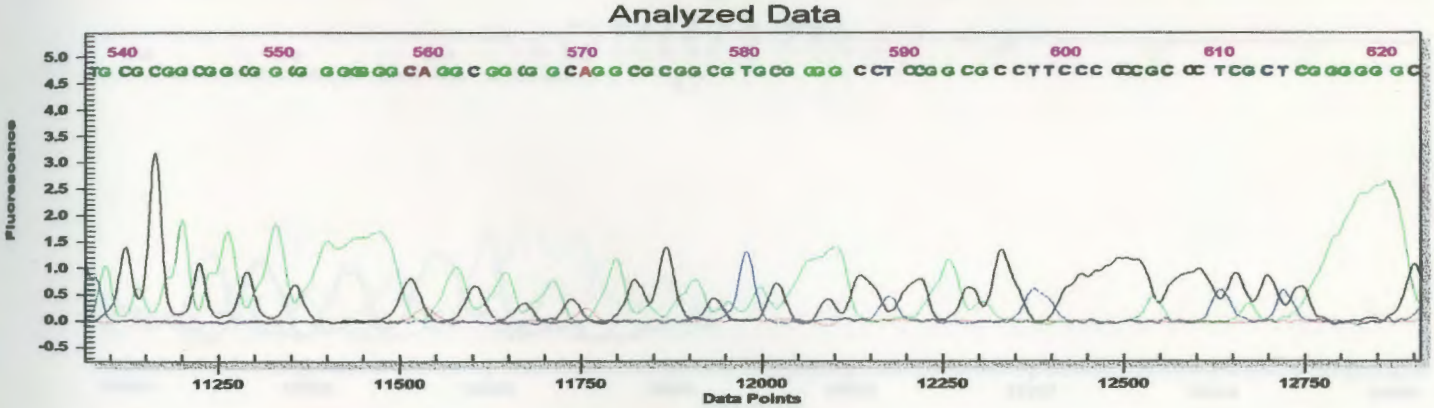
Analyzed Data



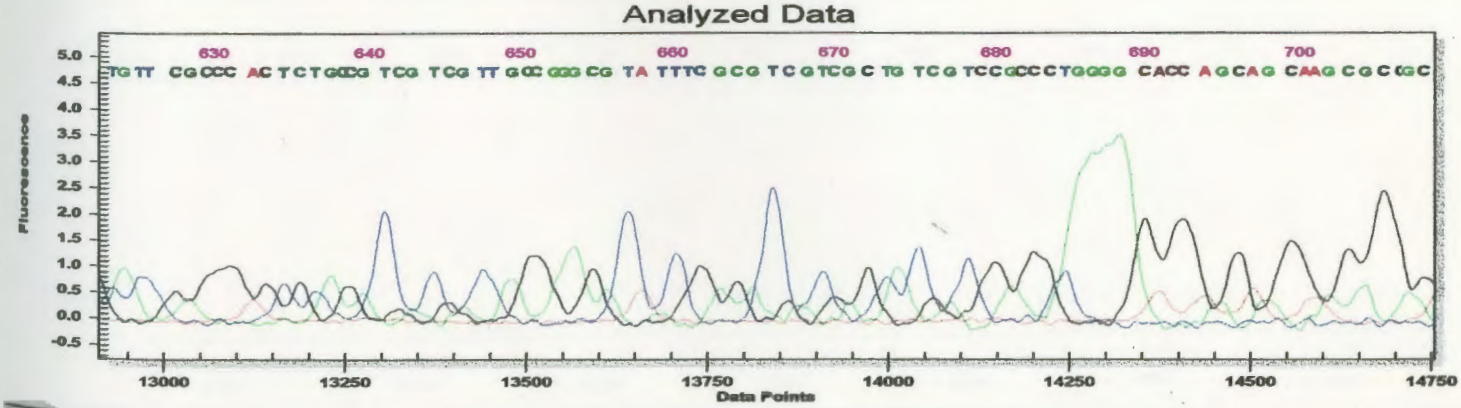
Analyzed Data



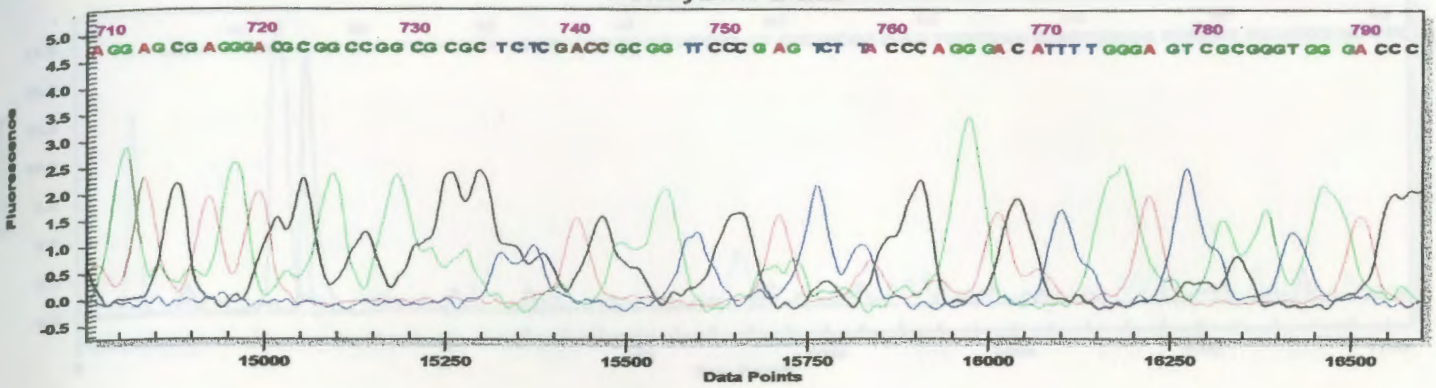
Analyzed Data



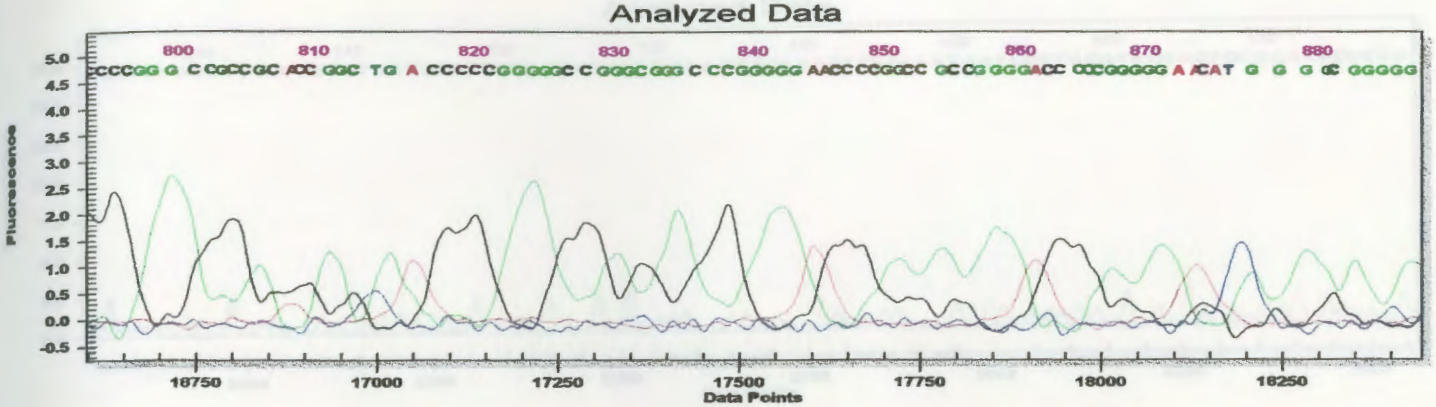
Analyzed Data



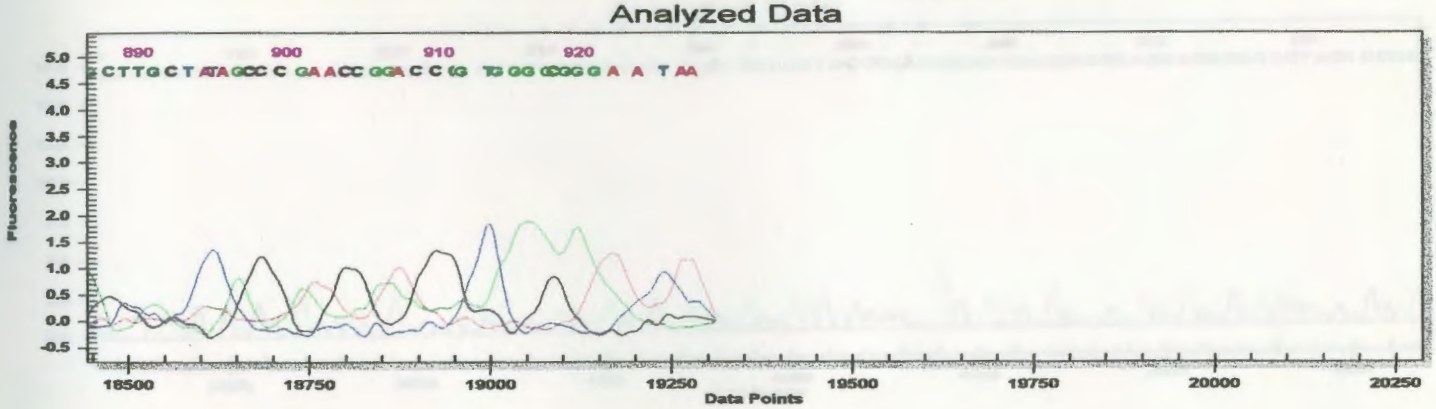
Analyzed Data



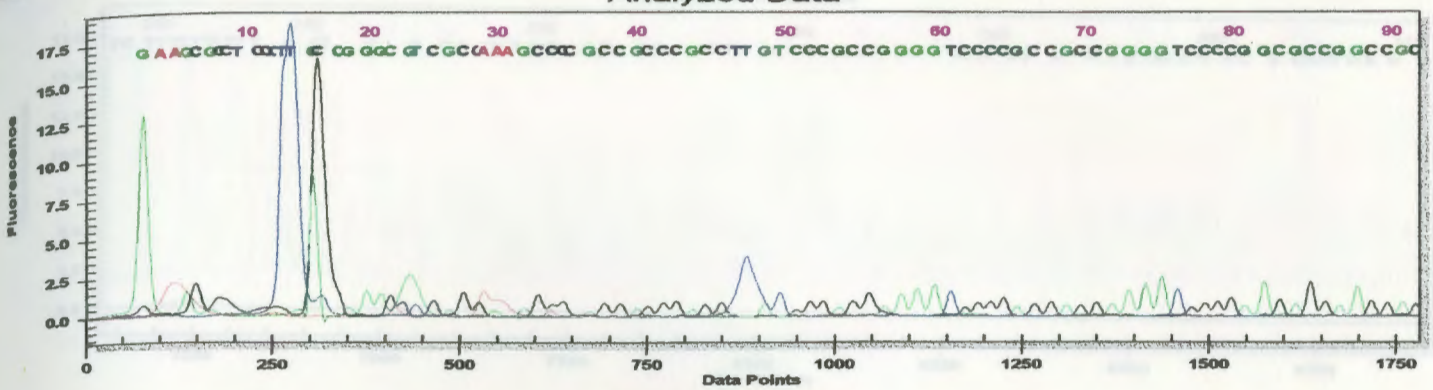
Analyzed Data



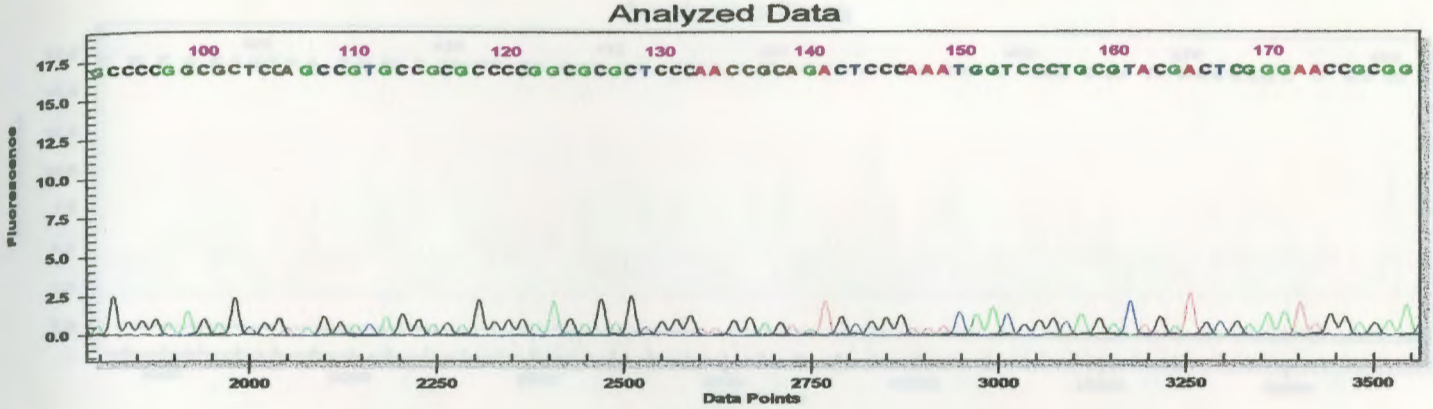
Analyzed Data



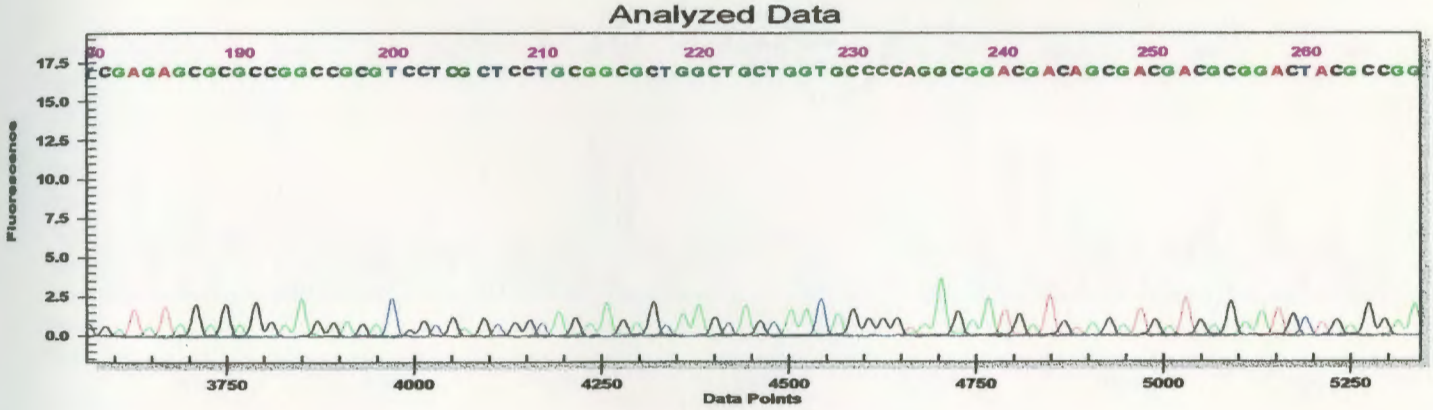
Analyzed Data



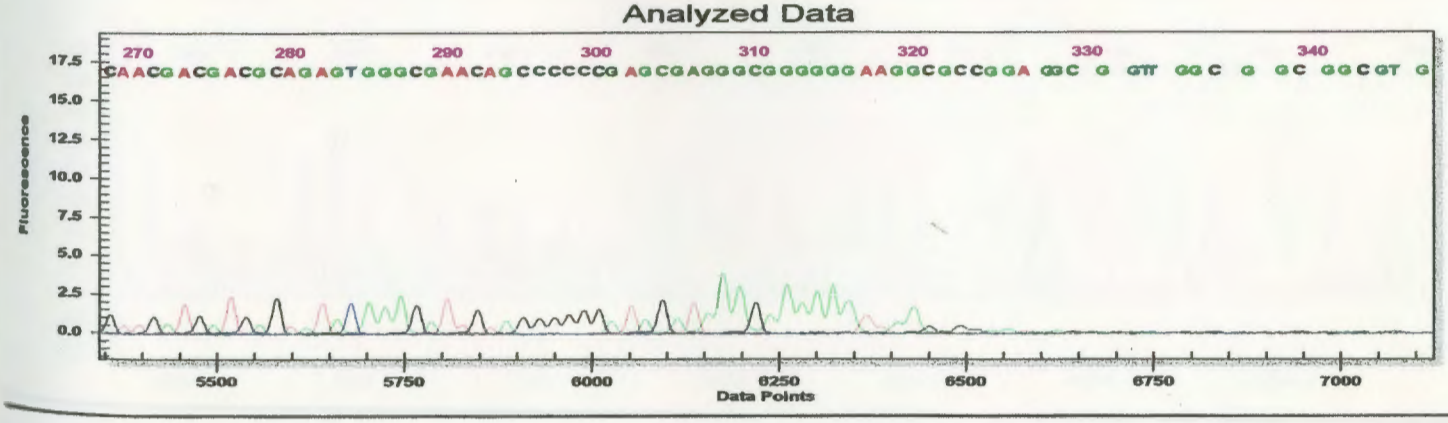
Analyzed Data



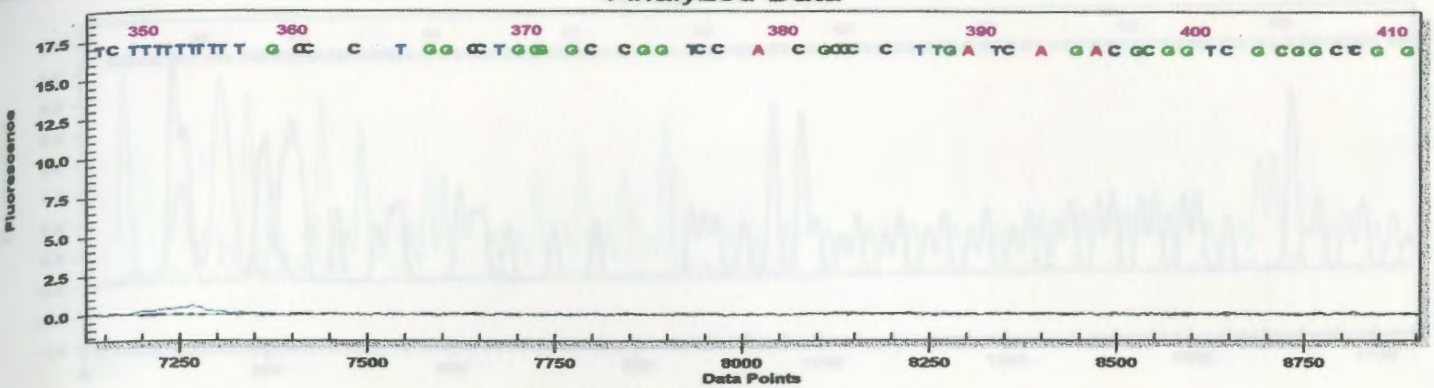
Analyzed Data



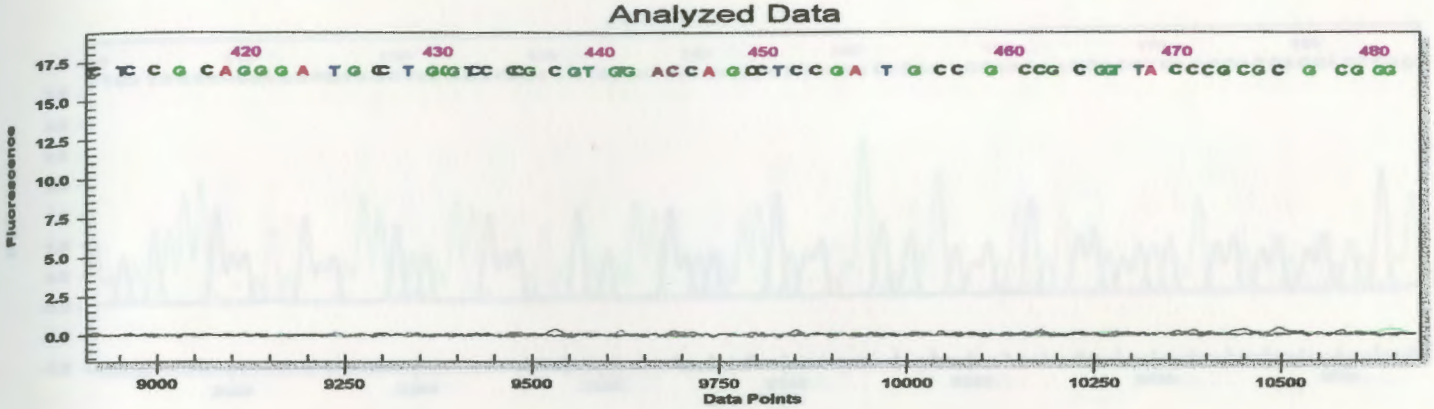
Analyzed Data



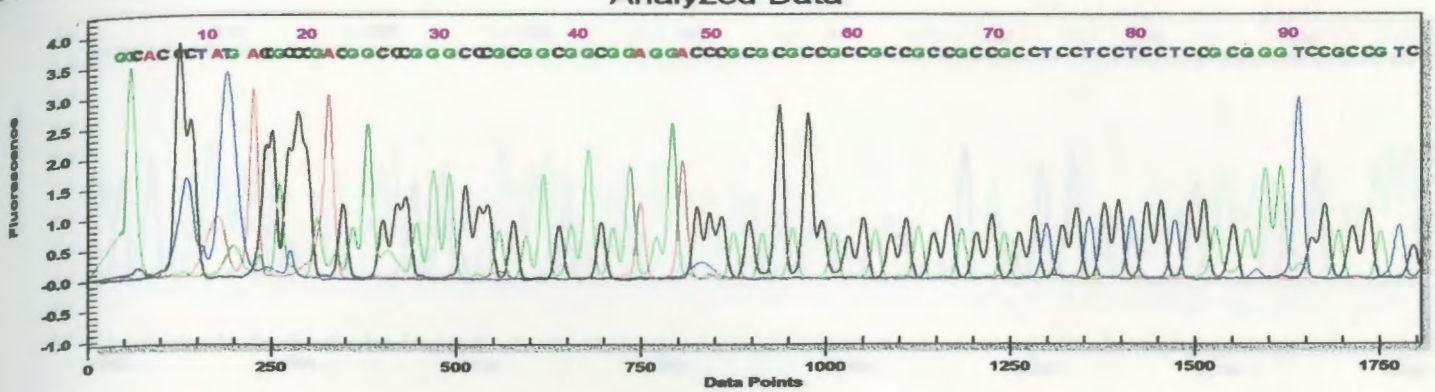
Analyzed Data



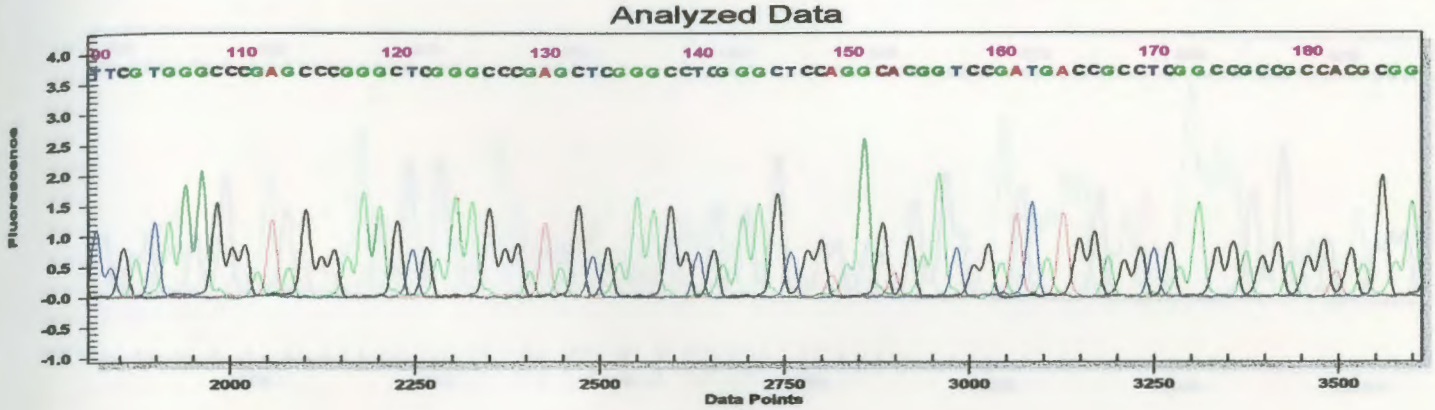
Analyzed Data



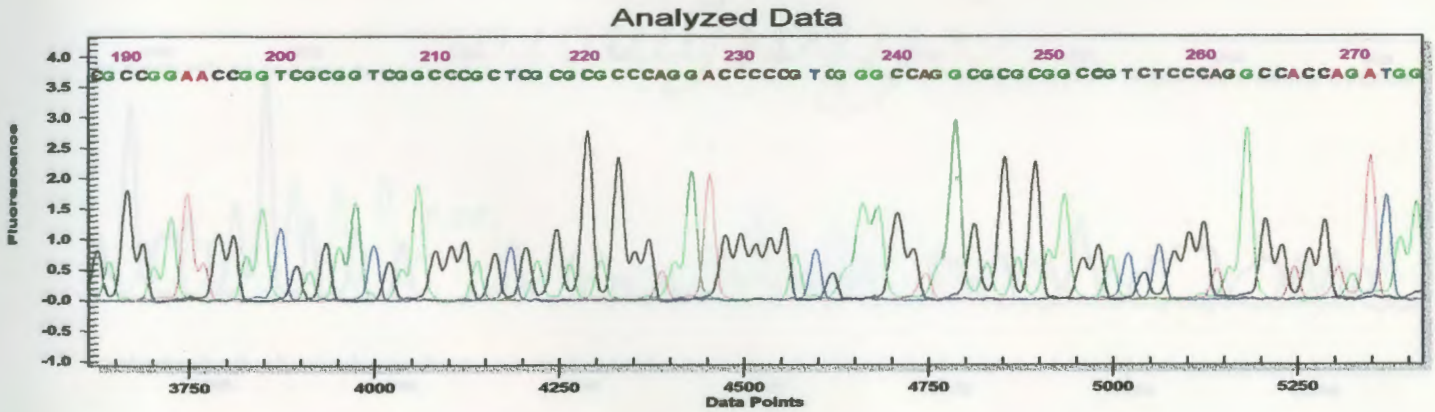
Analyzed Data



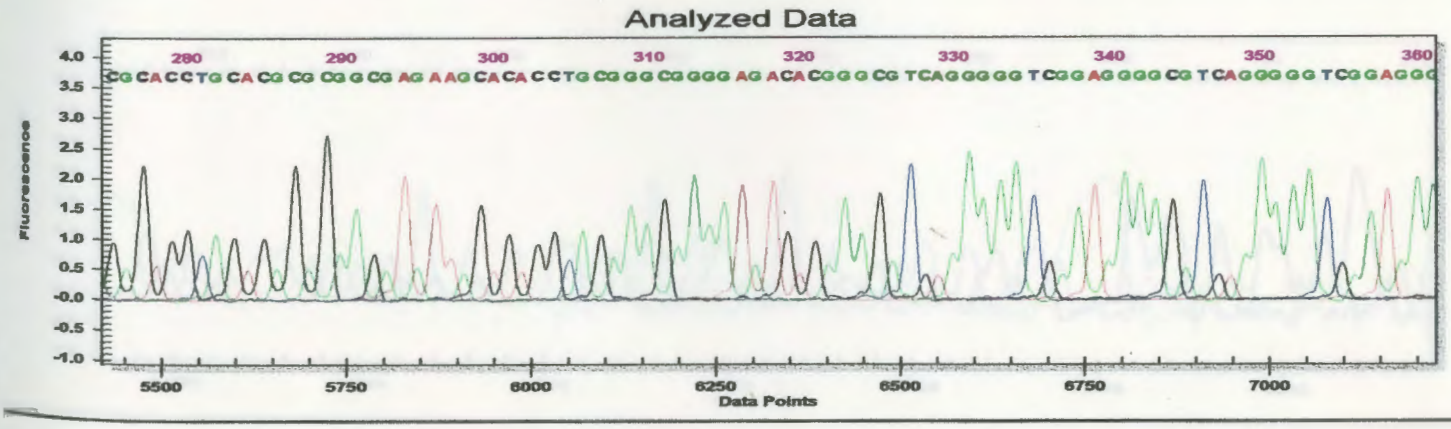
Analyzed Data

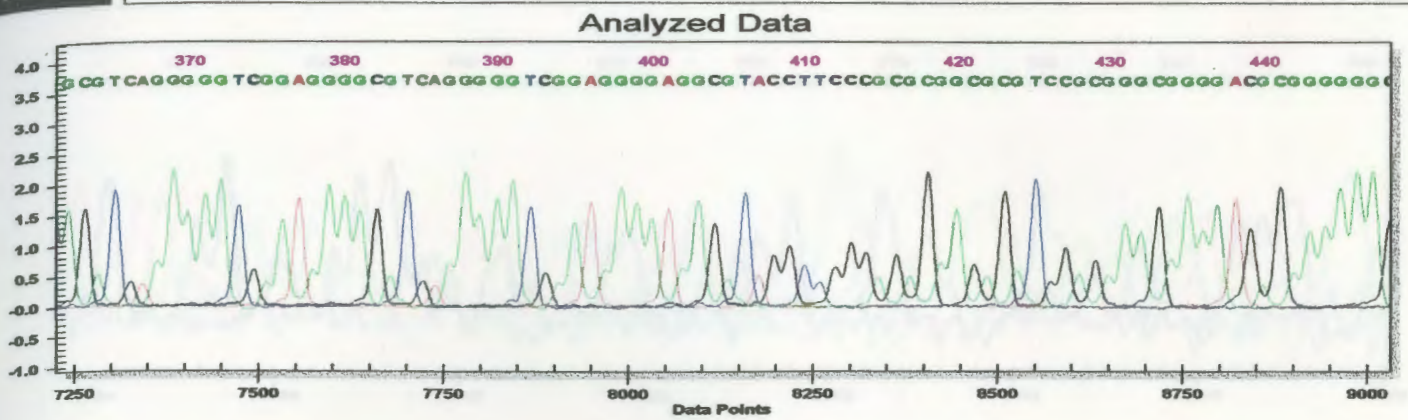


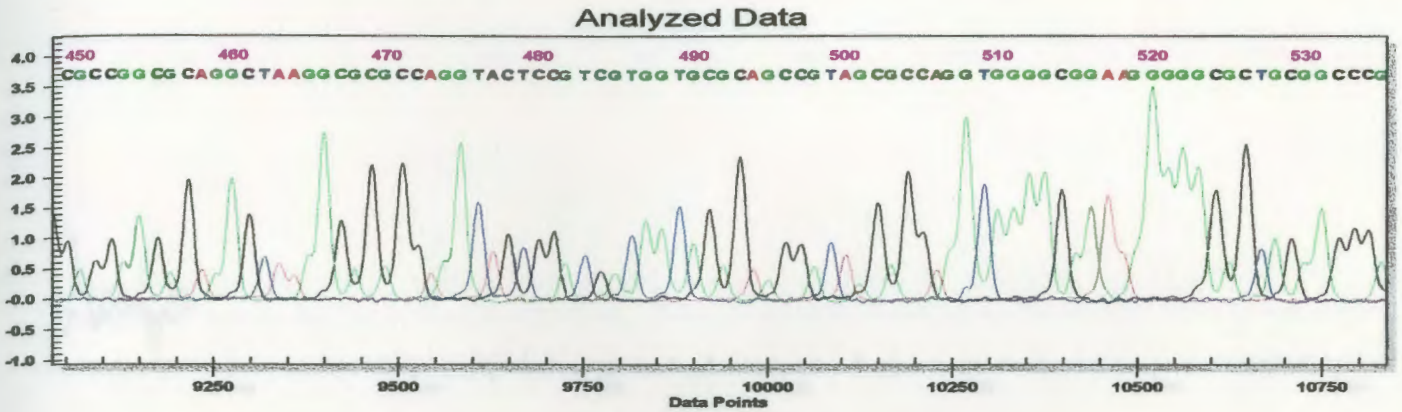
Analyzed Data

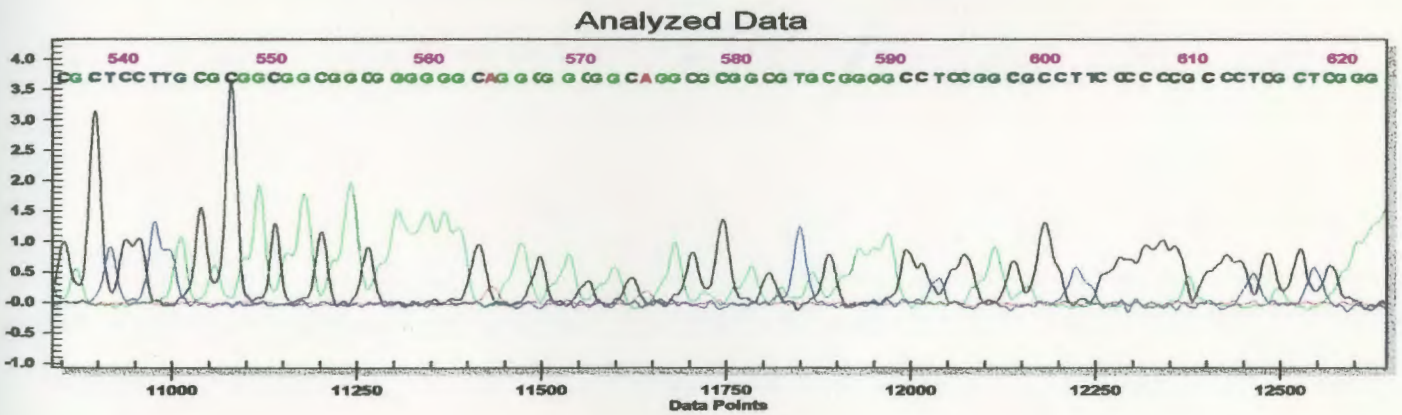


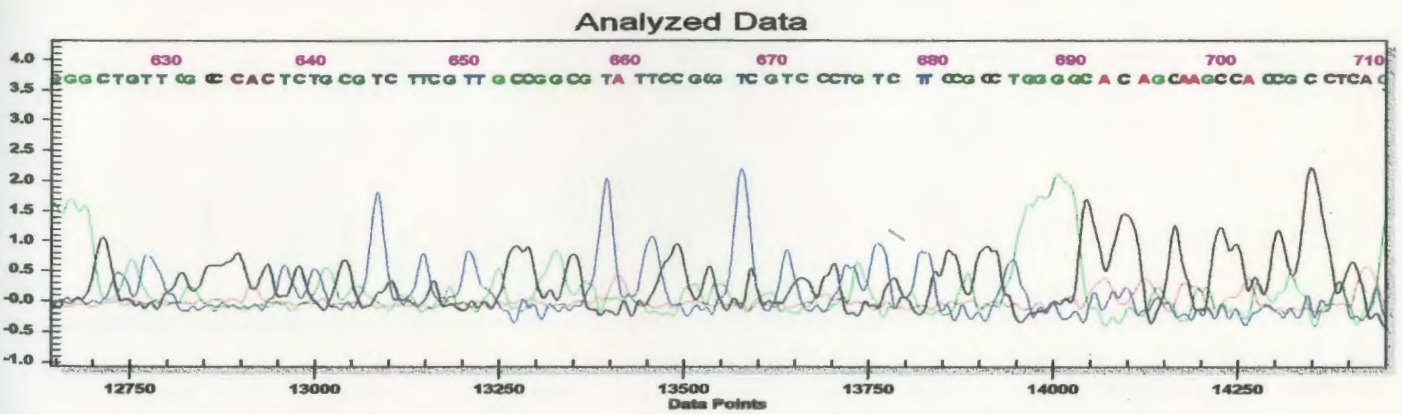
Analyzed Data



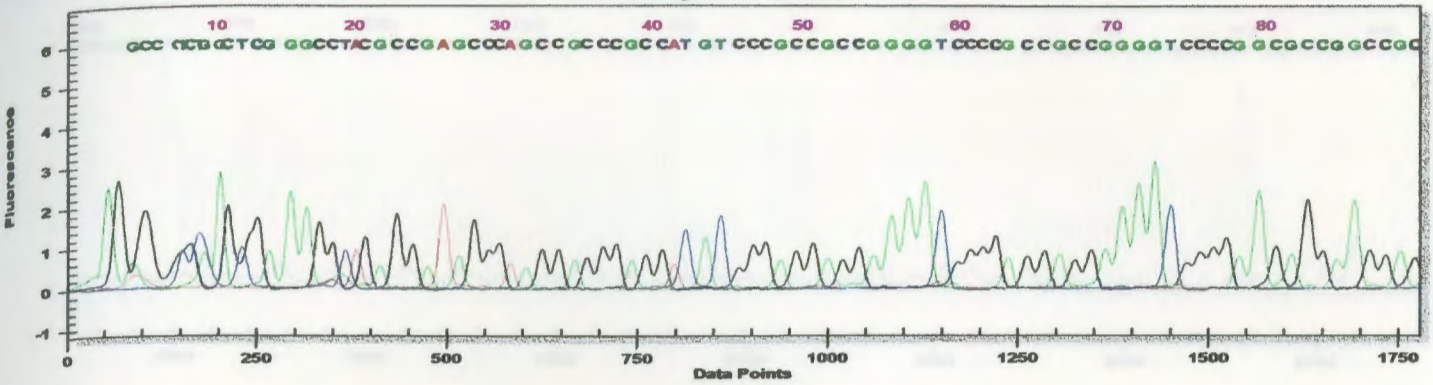

Analyzed Data
Fluorescence
Data Points


Analyzed Data
Fluorescence
Data Points

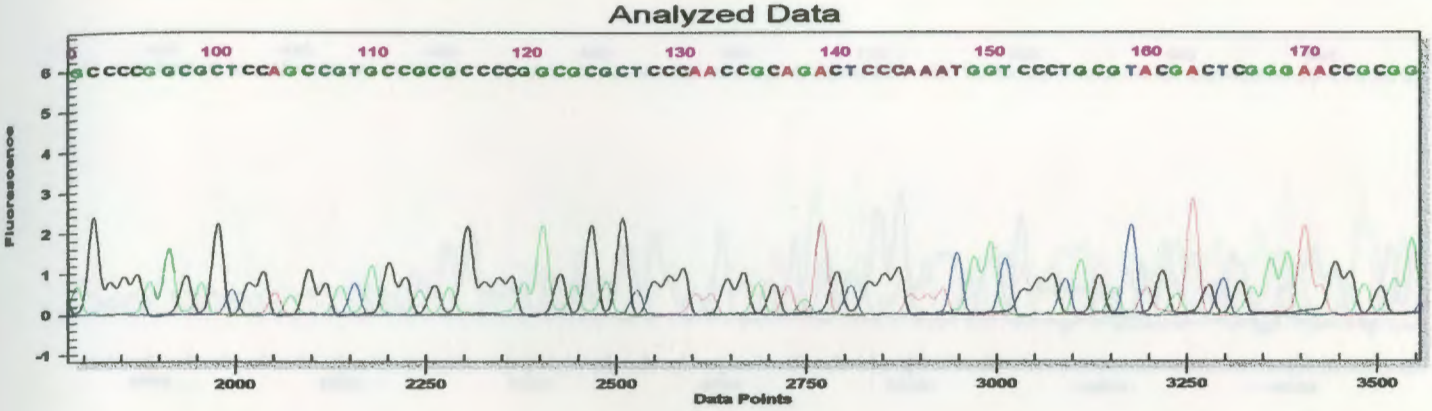

Analyzed Data
Fluorescence
Data Points


Analyzed Data
Fluorescence
Data Points

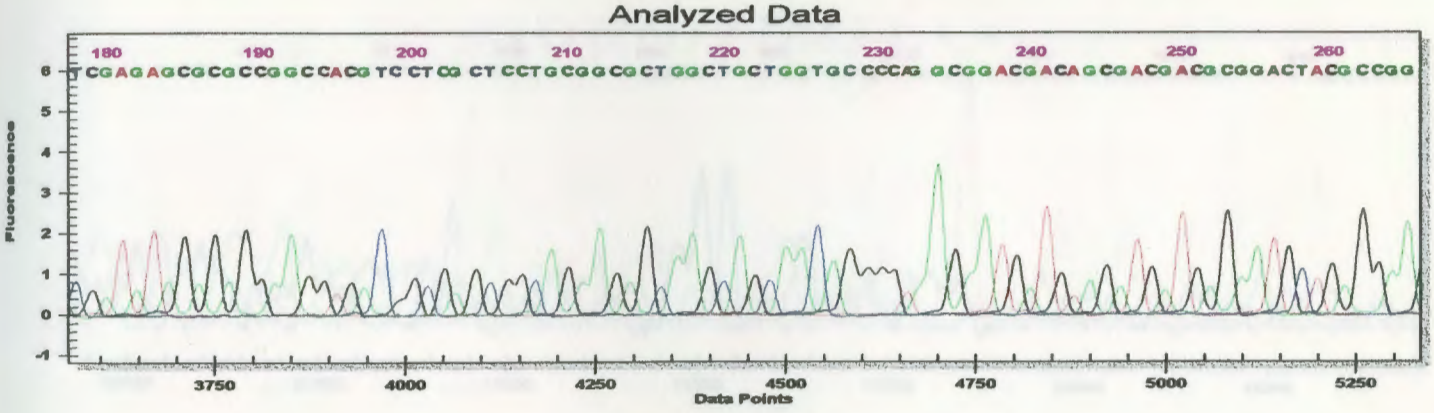
Analyzed Data



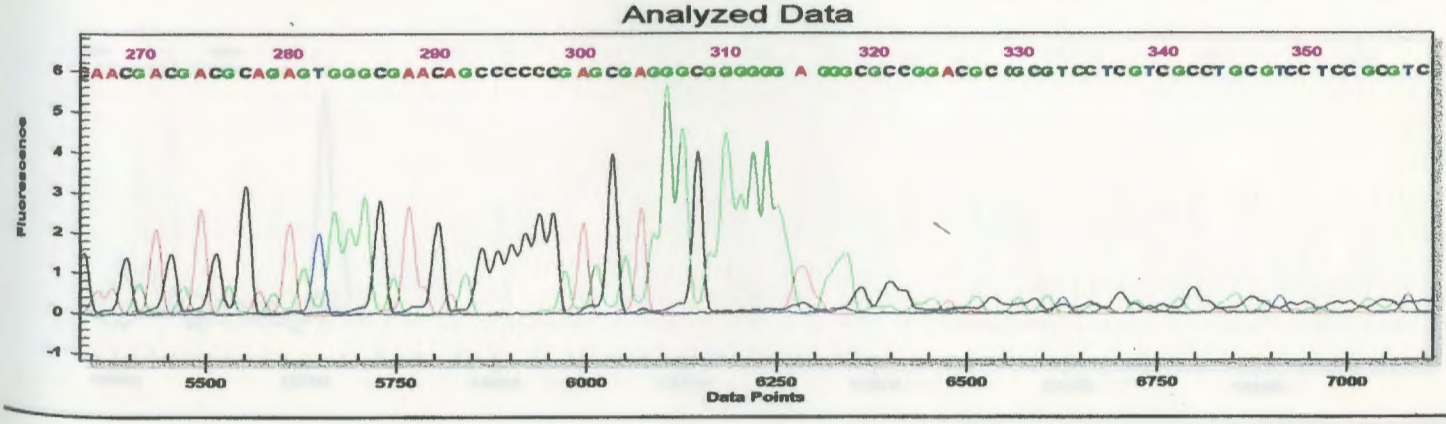
Analyzed Data



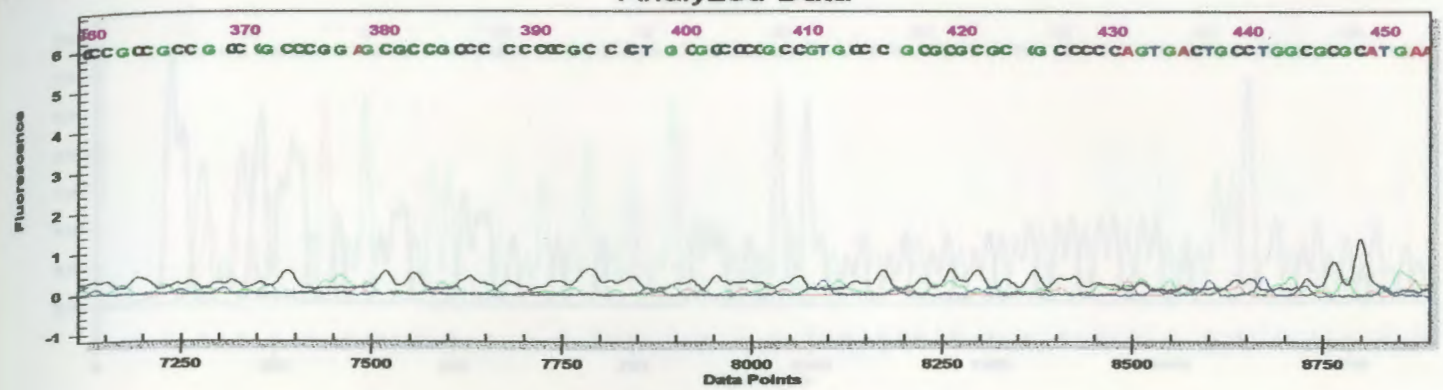
Analyzed Data



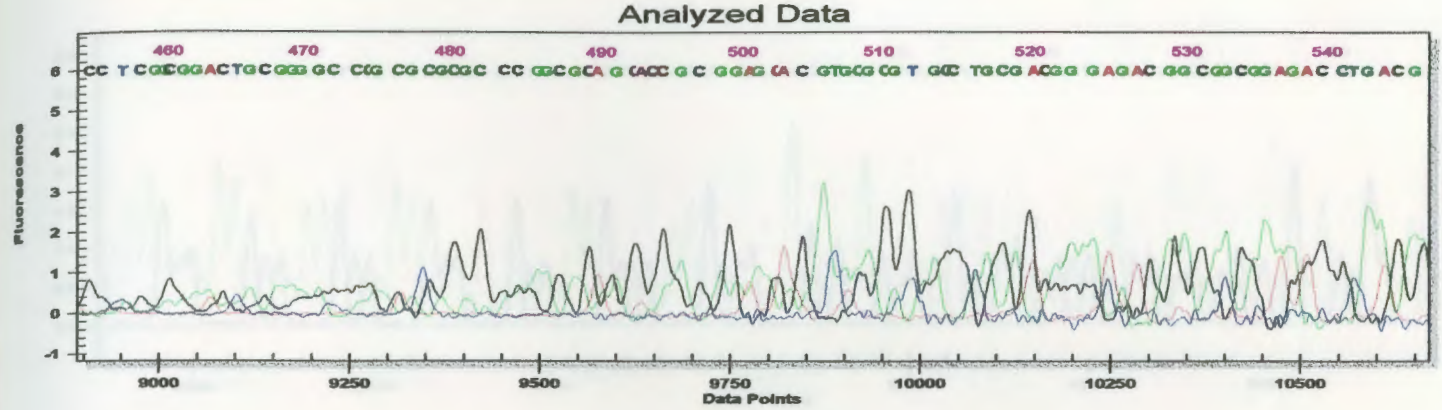
Analyzed Data



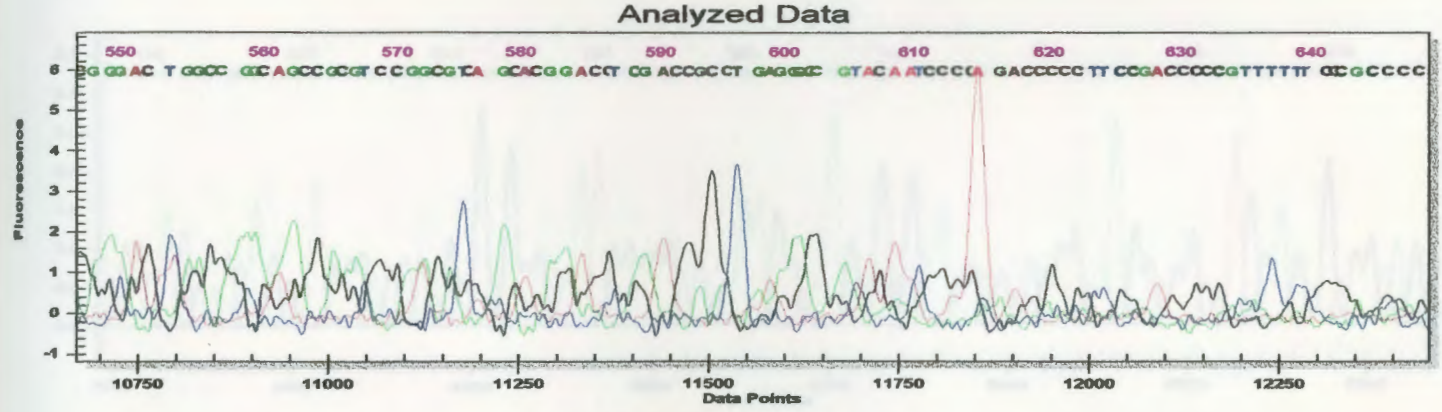
Analyzed Data



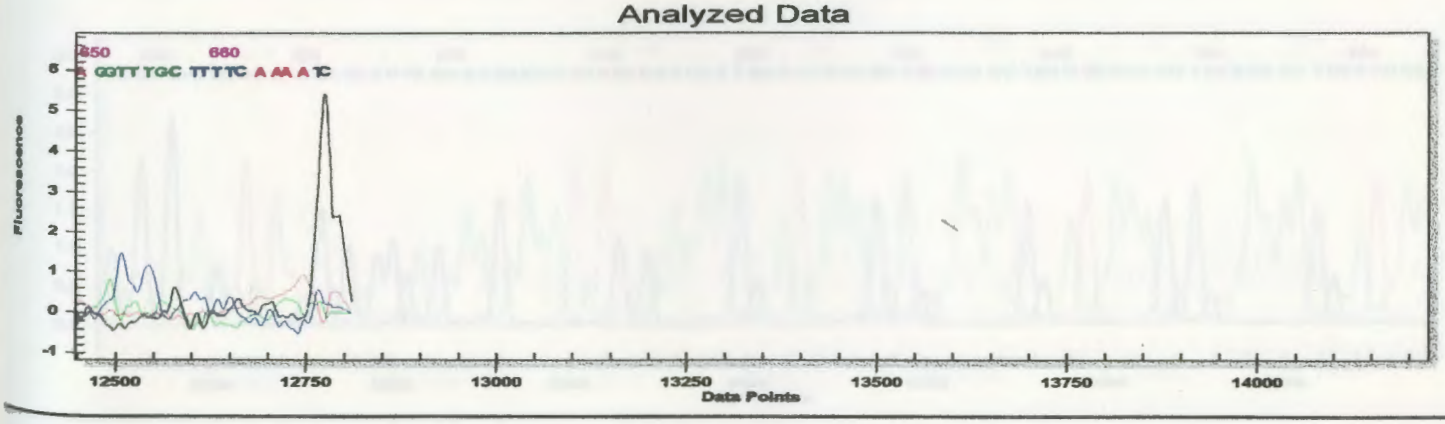
Analyzed Data



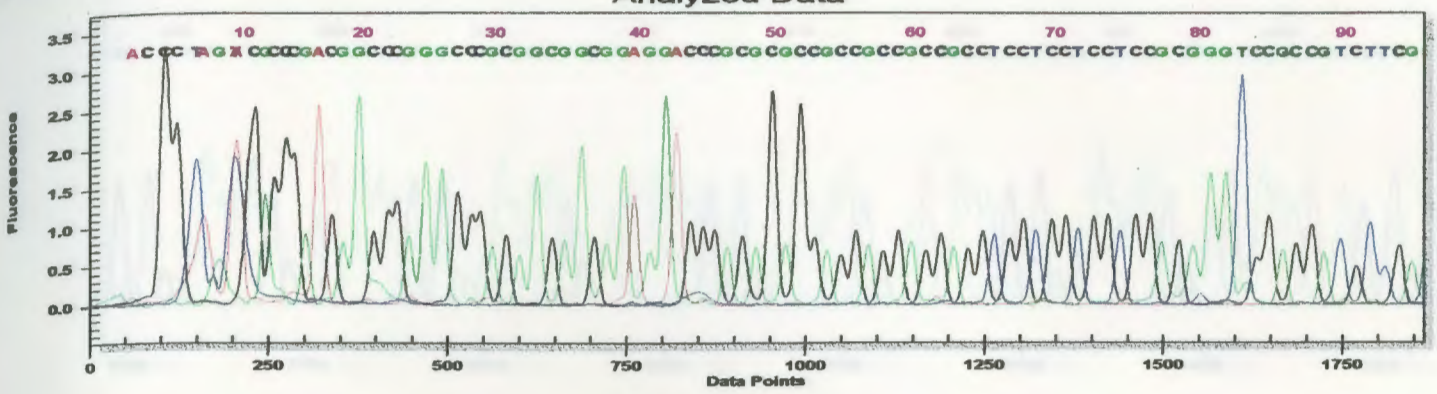
Analyzed Data



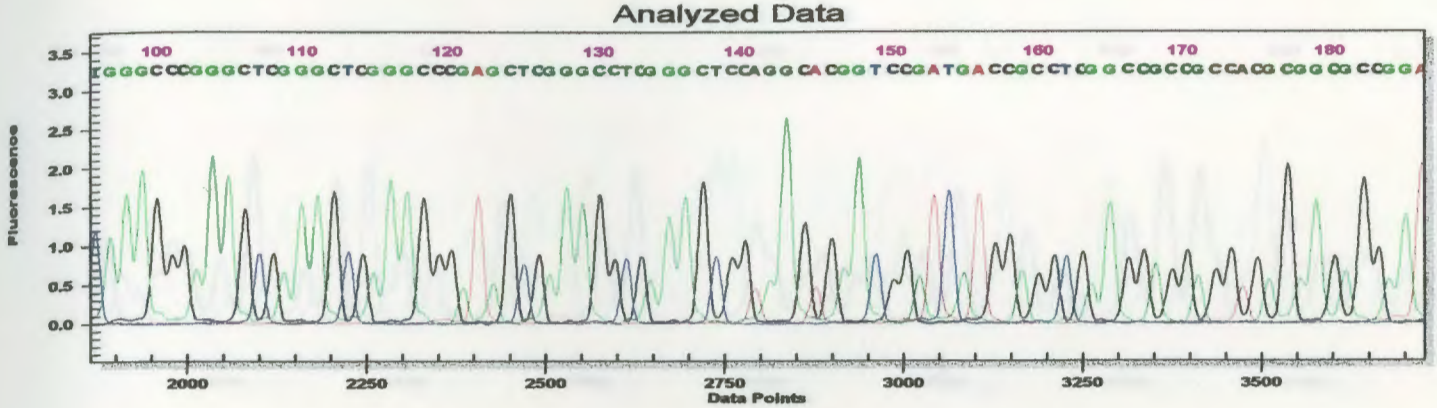
Analyzed Data



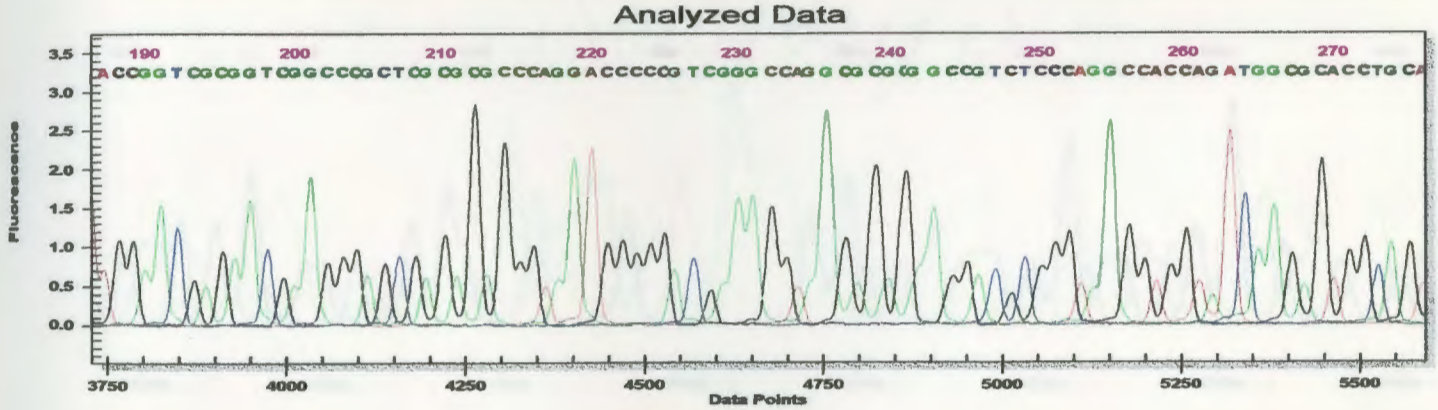
Analyzed Data



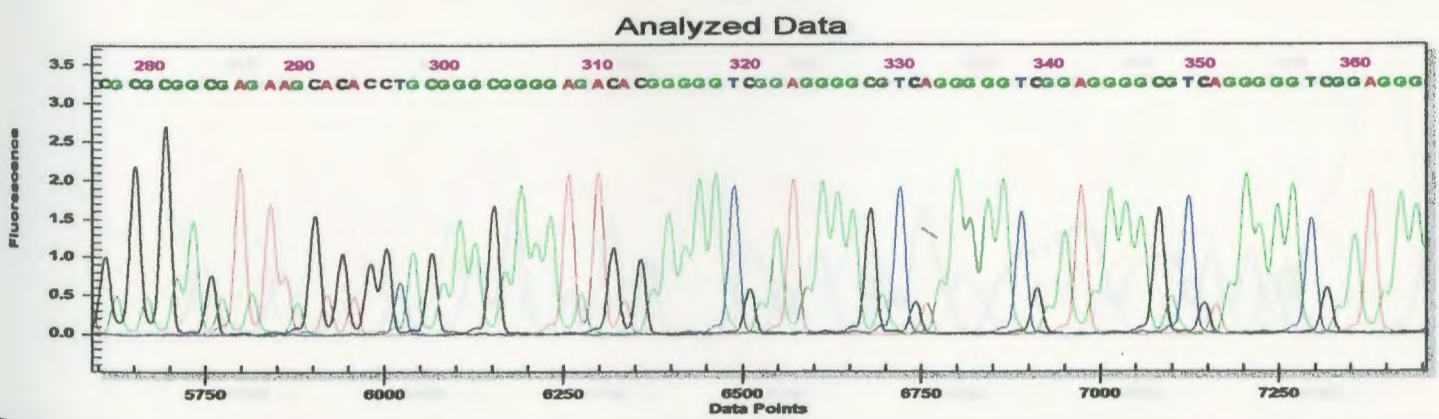
Analyzed Data



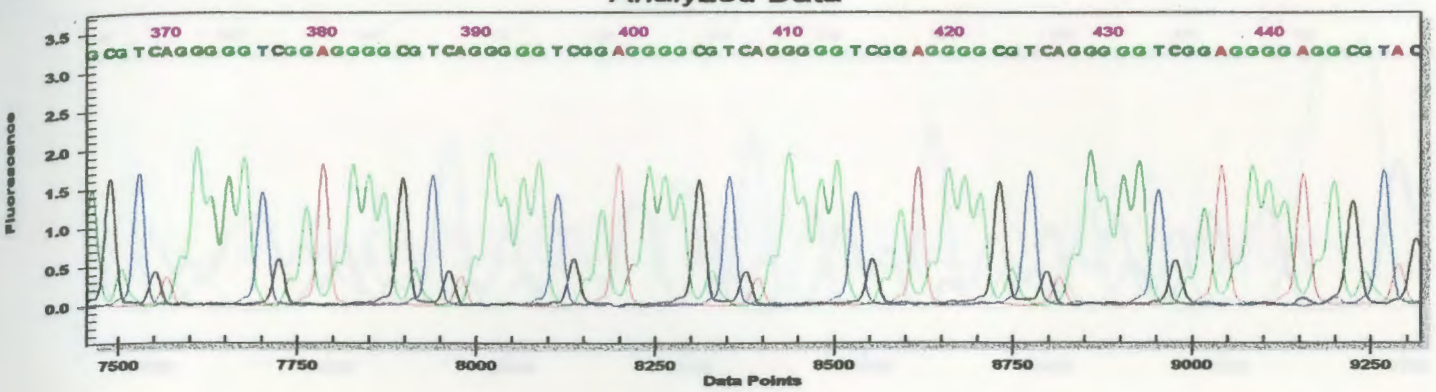
Analyzed Data



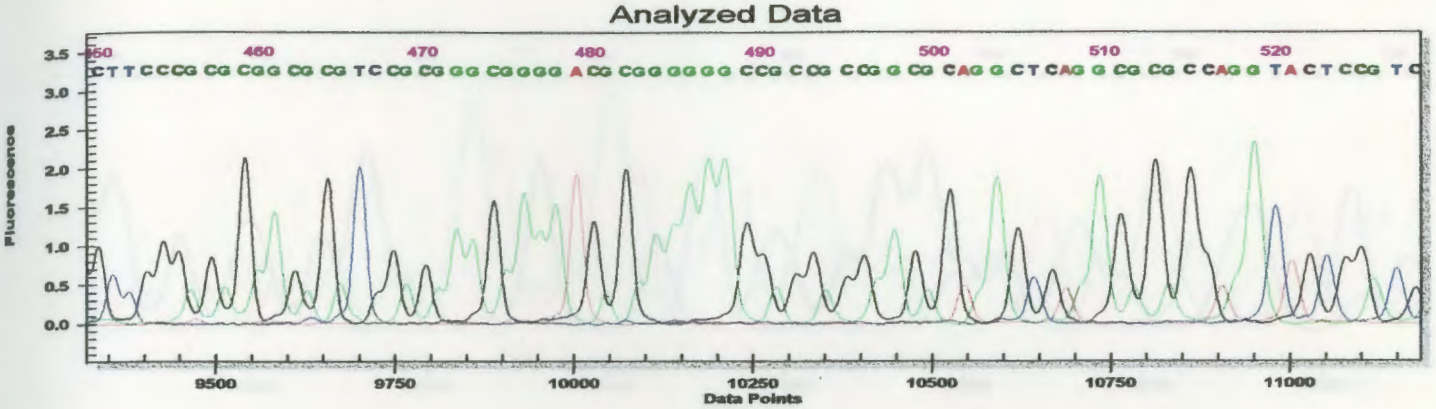
Analyzed Data



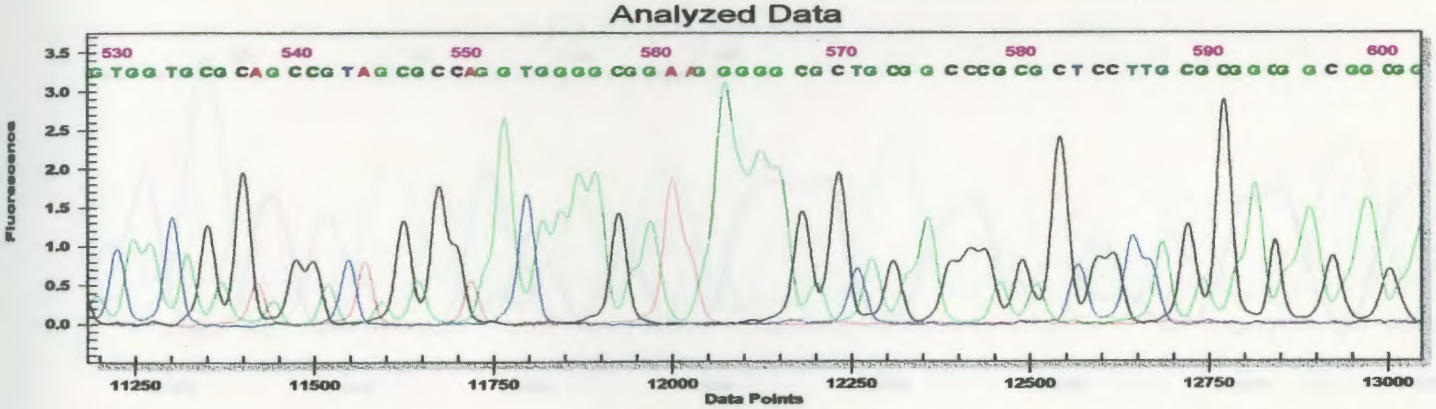
Analyzed Data



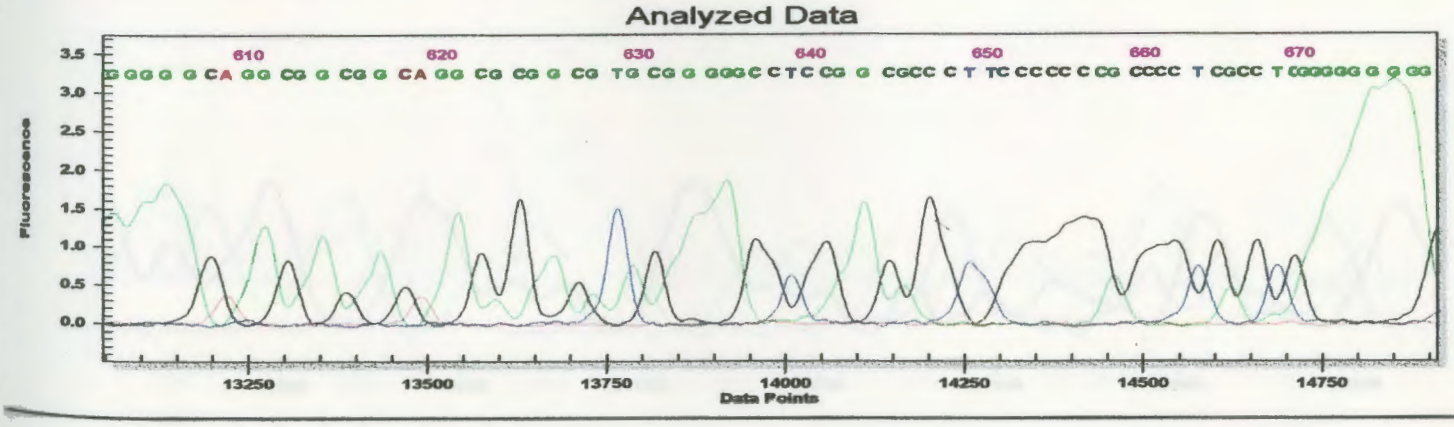
Analyzed Data

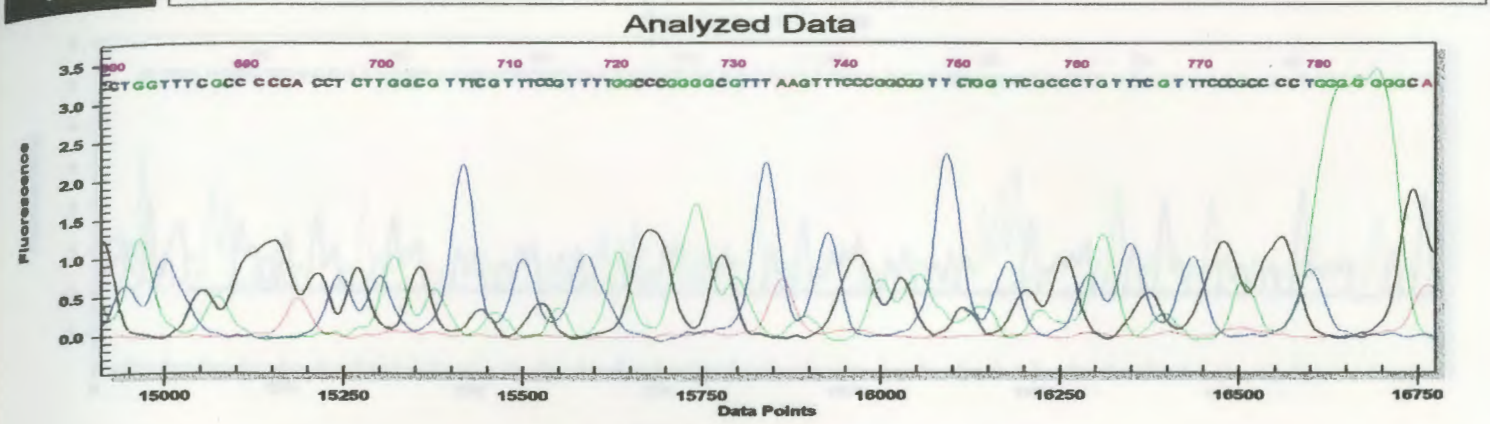


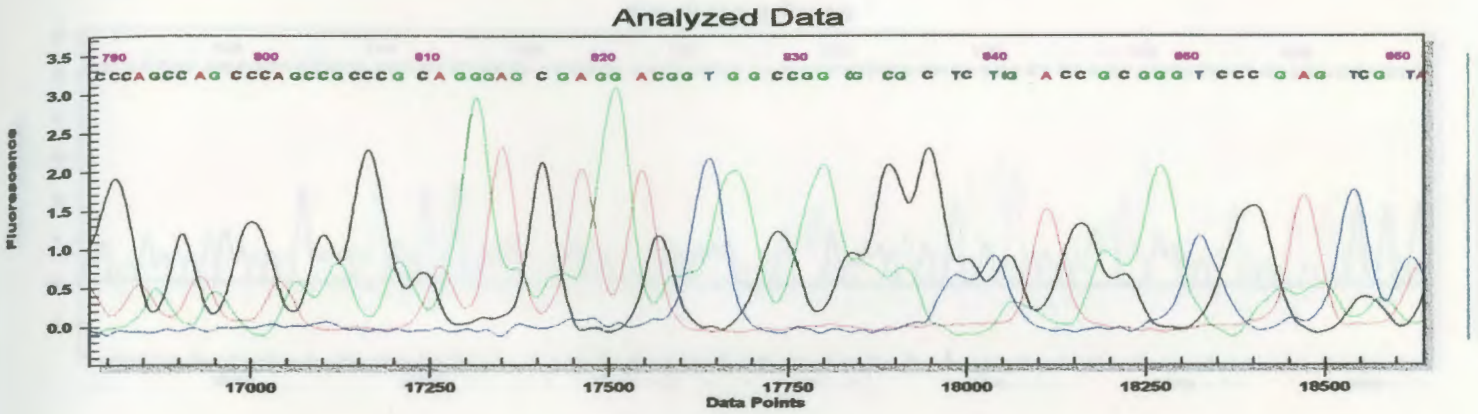
Analyzed Data

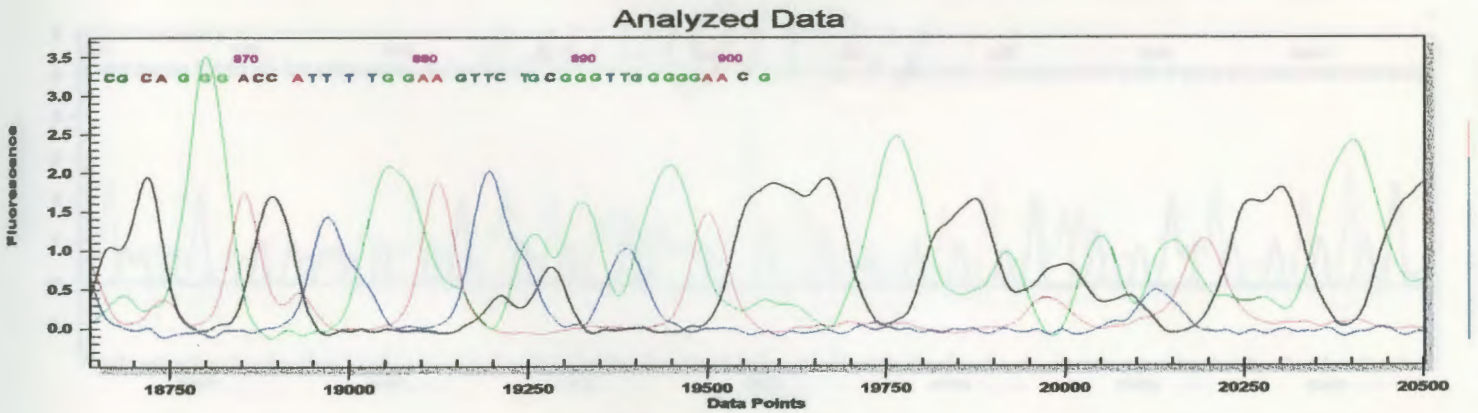


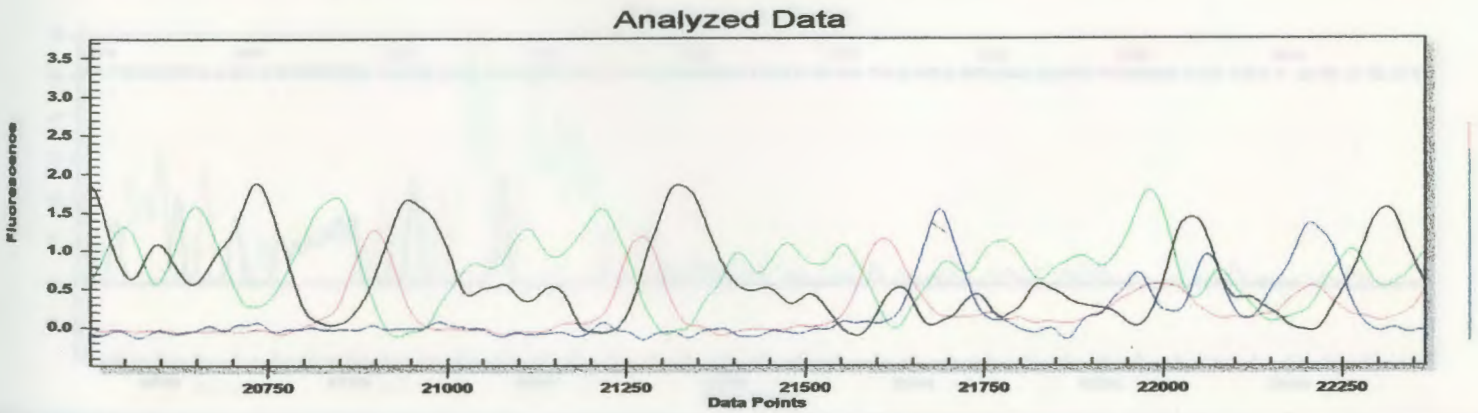
Analyzed Data



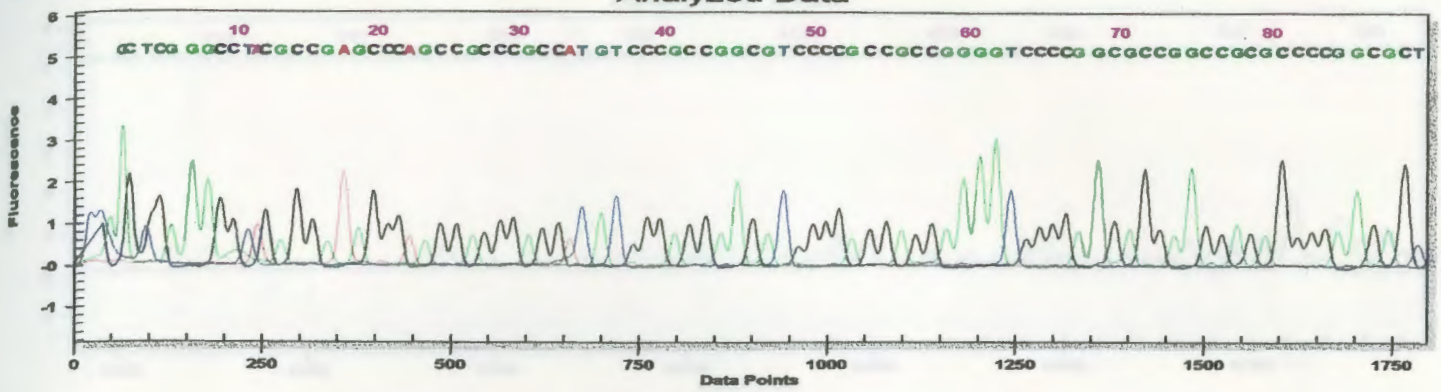

Analyzed Data
Fluorescence
Data Points
15000
15250
15500
15750
16000
16250
16500
16750


Analyzed Data
Fluorescence
Data Points
17000
17250
17500
17750
18000
18250
18500

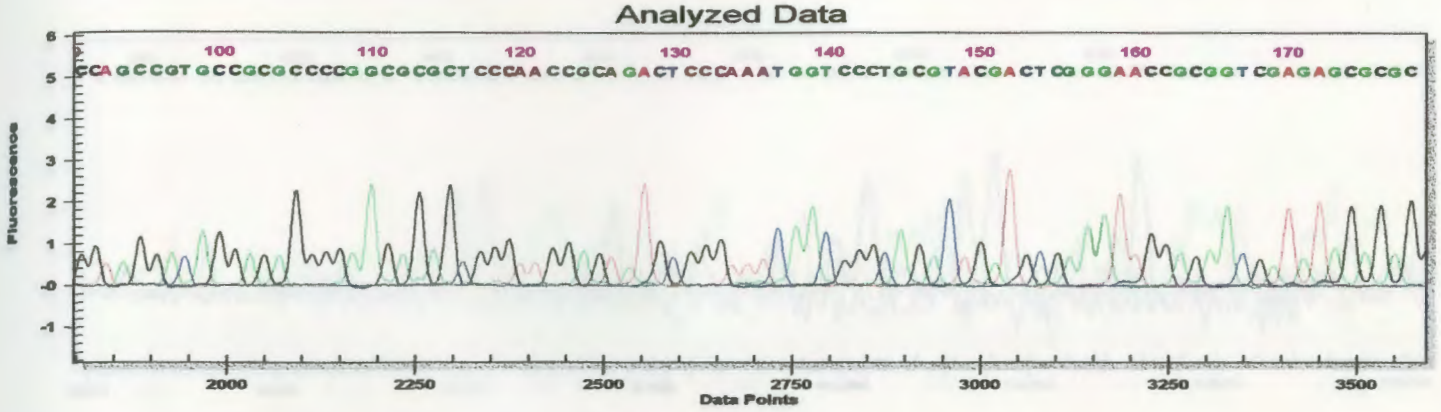

Analyzed Data
Fluorescence
Data Points
18750
19000
19250
19500
19750
20000
20250
20500


Analyzed Data
Fluorescence
Data Points
20750
21000
21250
21500
21750
22000
22250

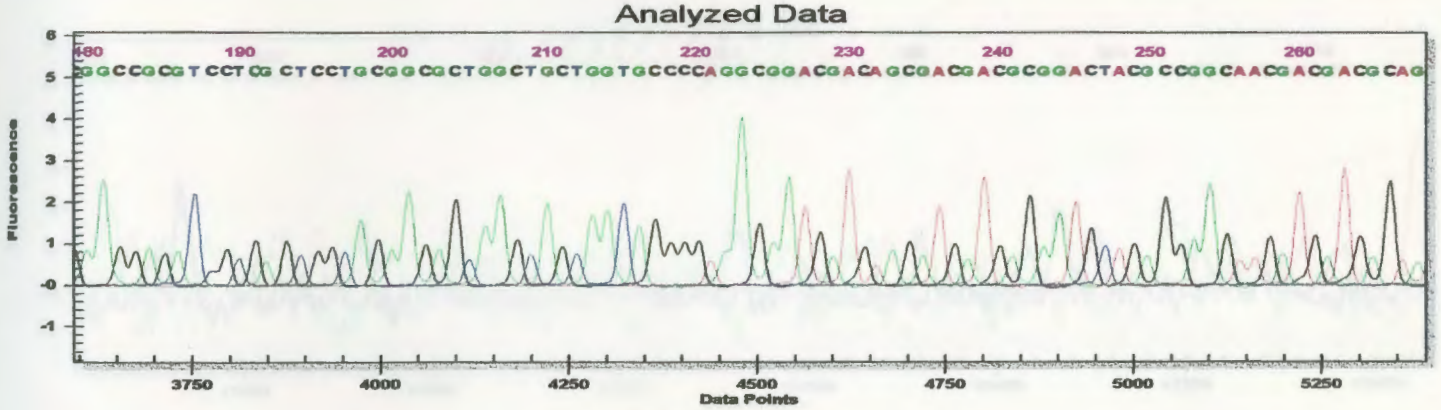
Analyzed Data



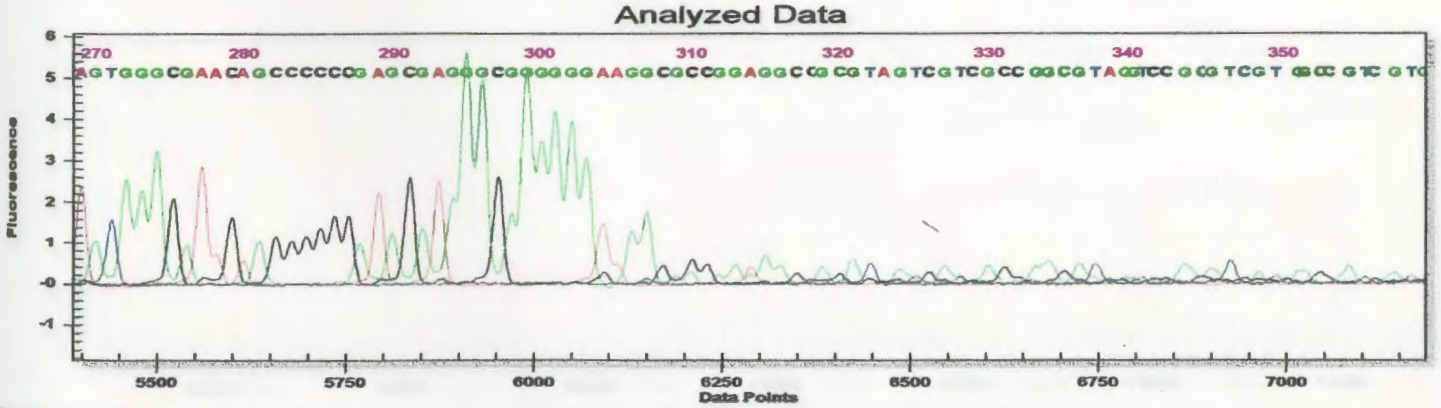
Analyzed Data

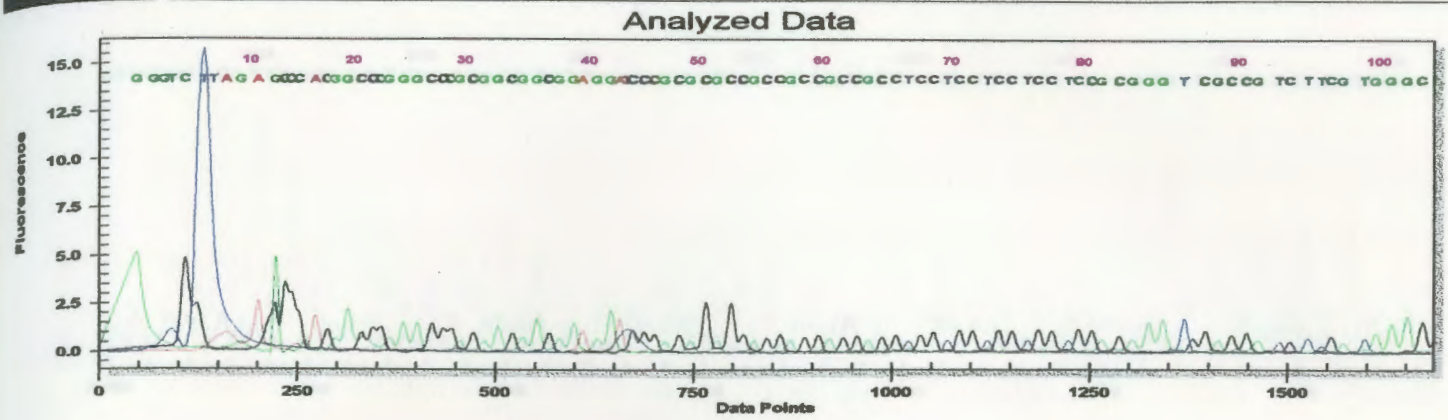


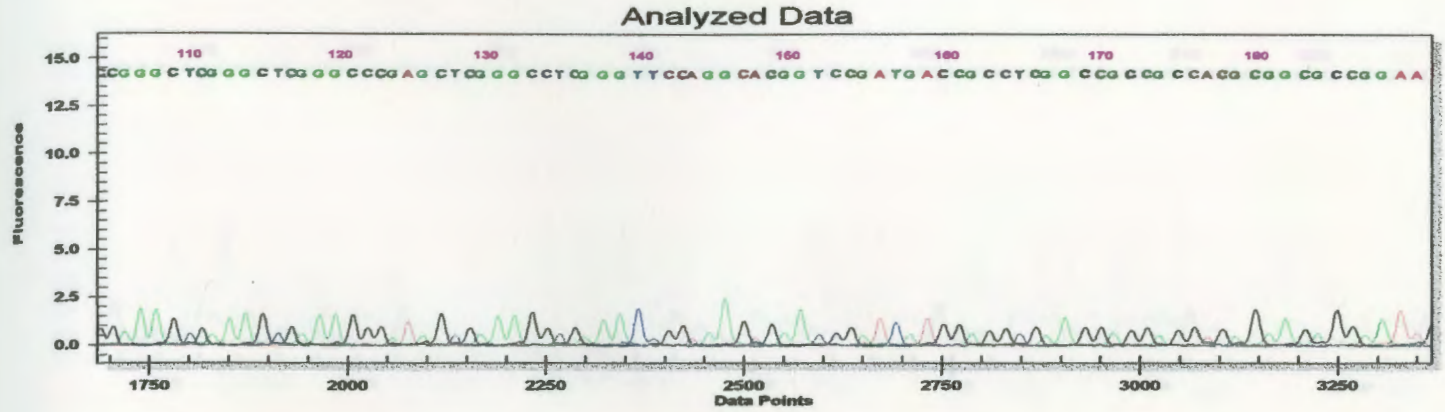
Analyzed Data

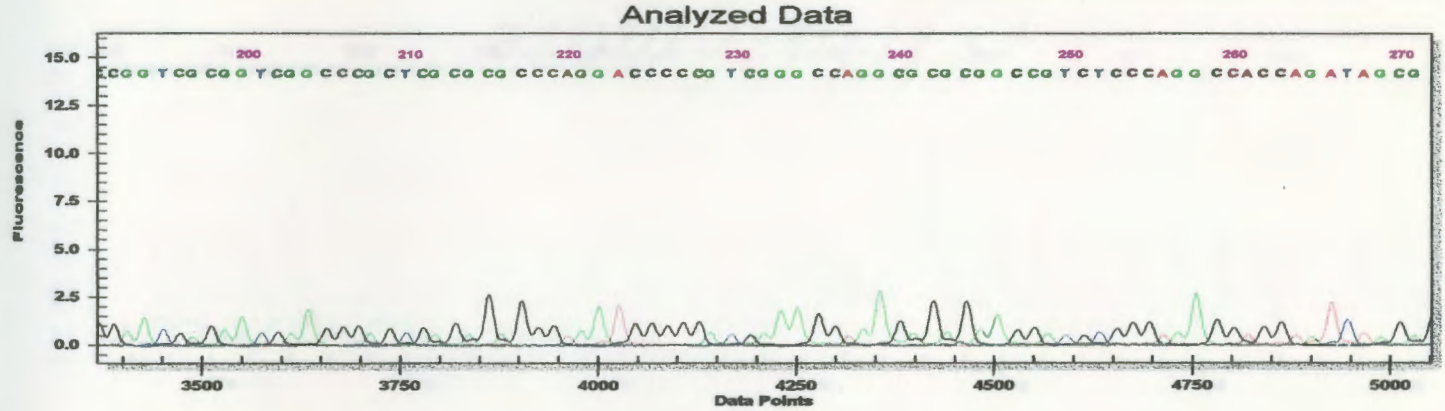


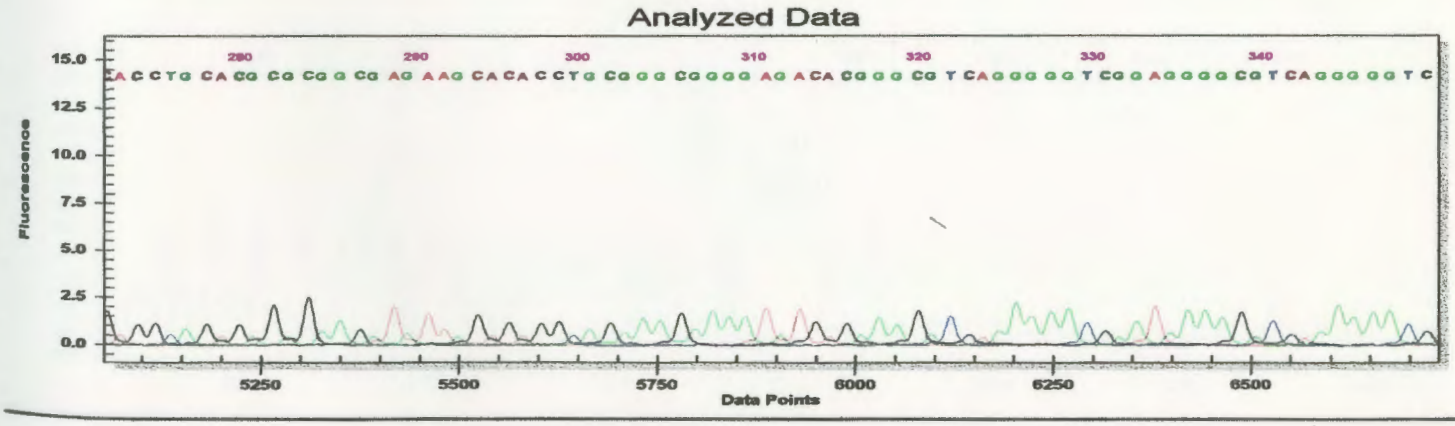
Analyzed Data



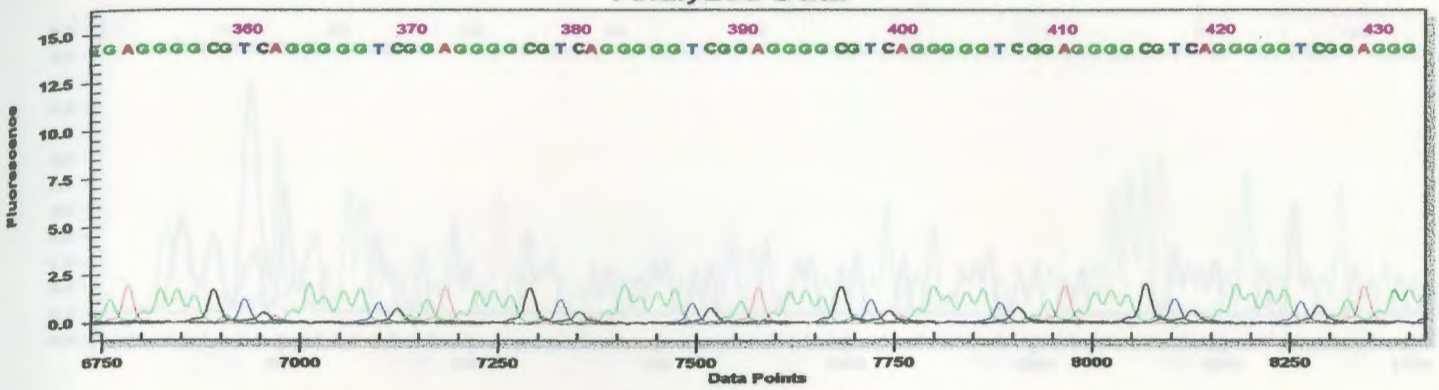

Analyzed Data
Fluorescence
Data Points


Analyzed Data
Fluorescence
Data Points

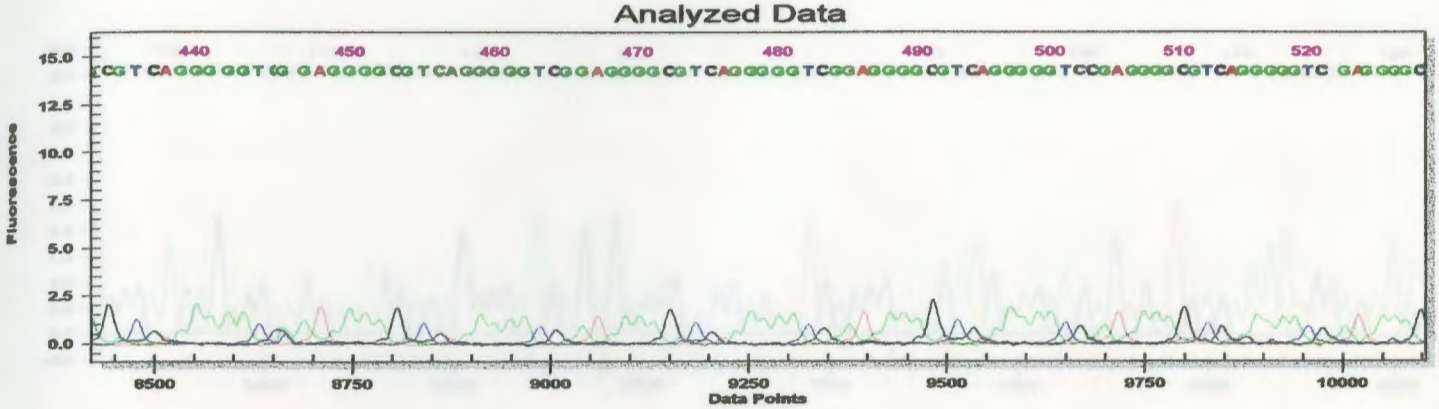

Analyzed Data
Fluorescence
Data Points


Analyzed Data
Fluorescence
Data Points

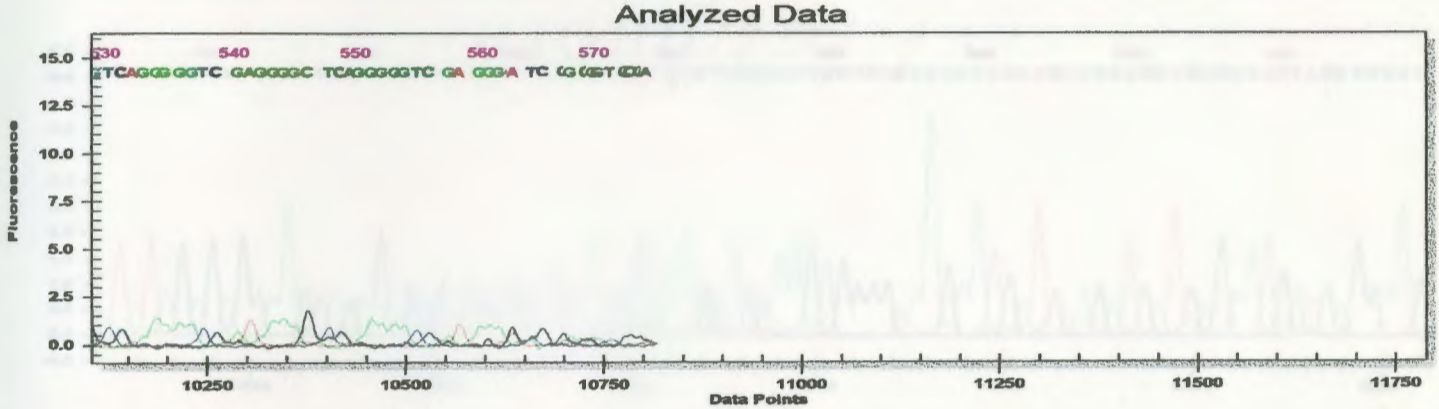
Analyzed Data



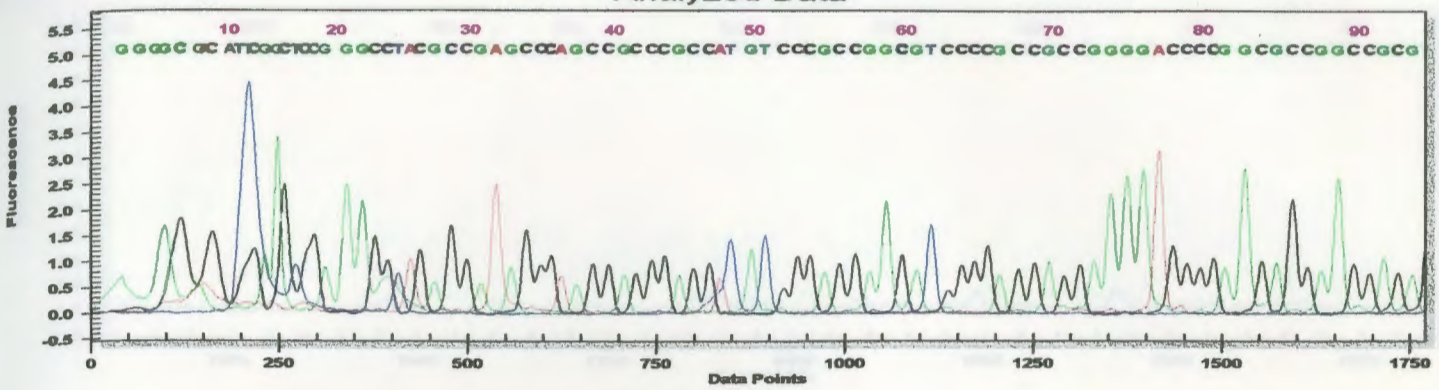
Analyzed Data



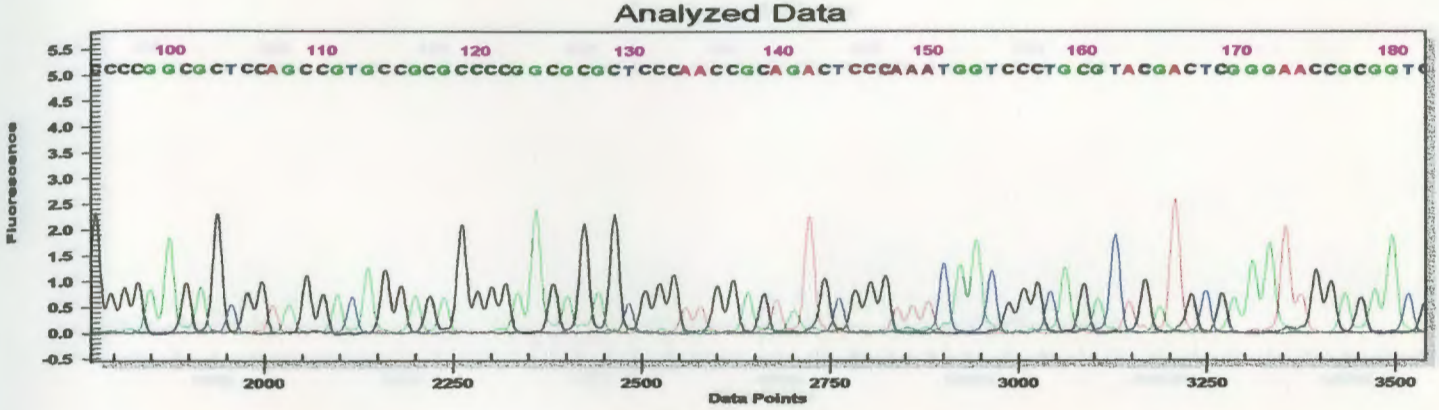
Analyzed Data



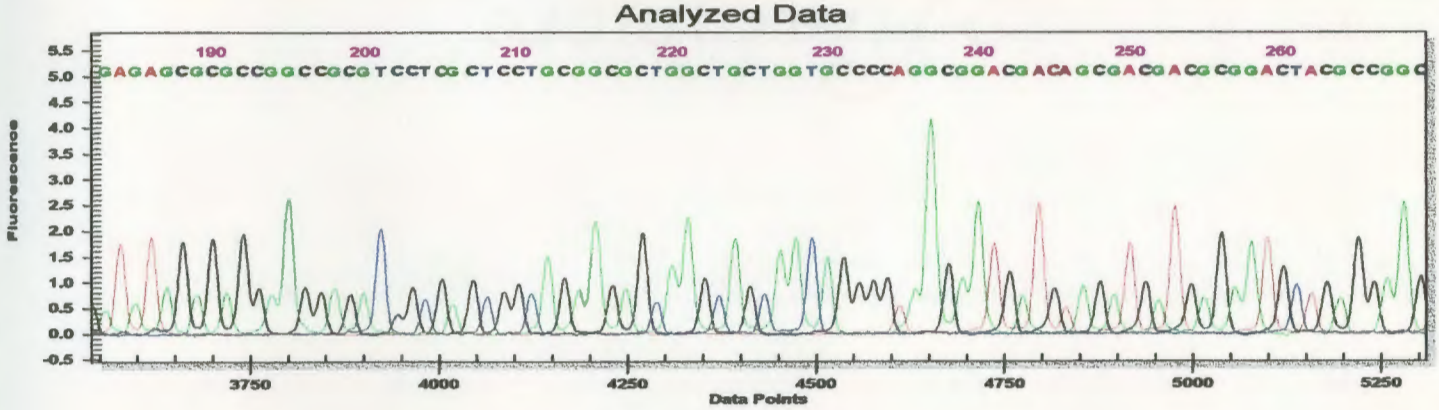
Analyzed Data



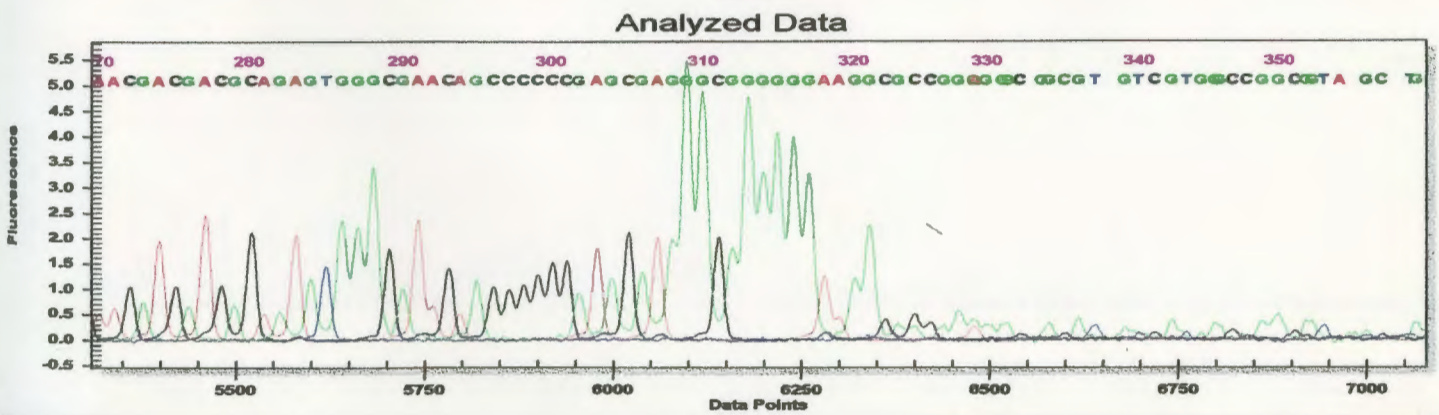
Analyzed Data



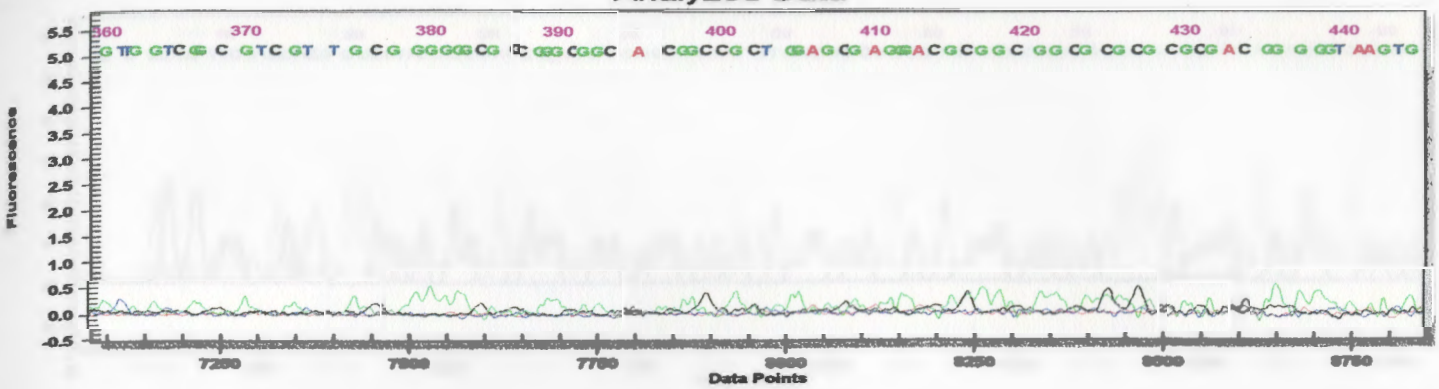
Analyzed Data



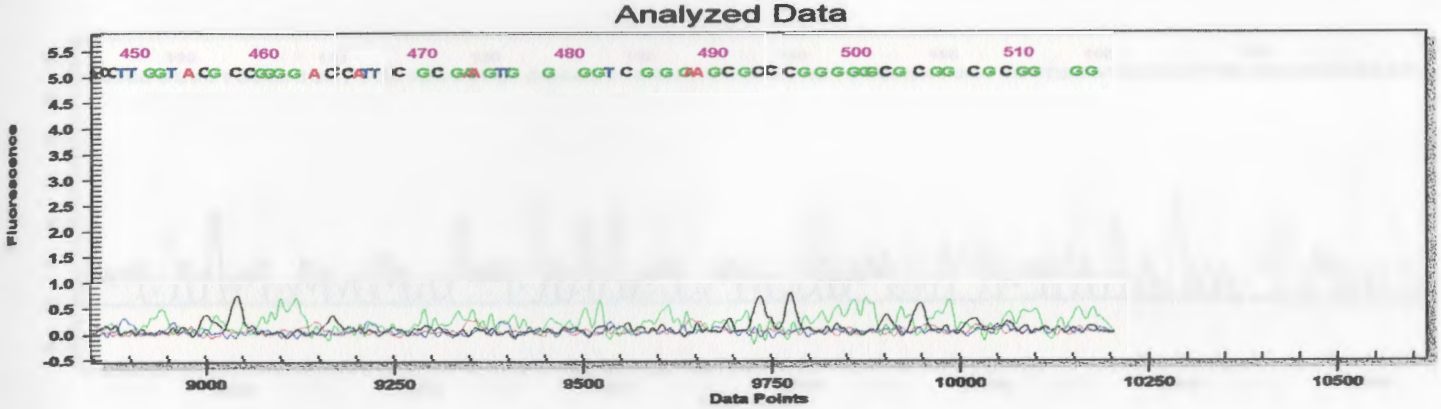
Analyzed Data



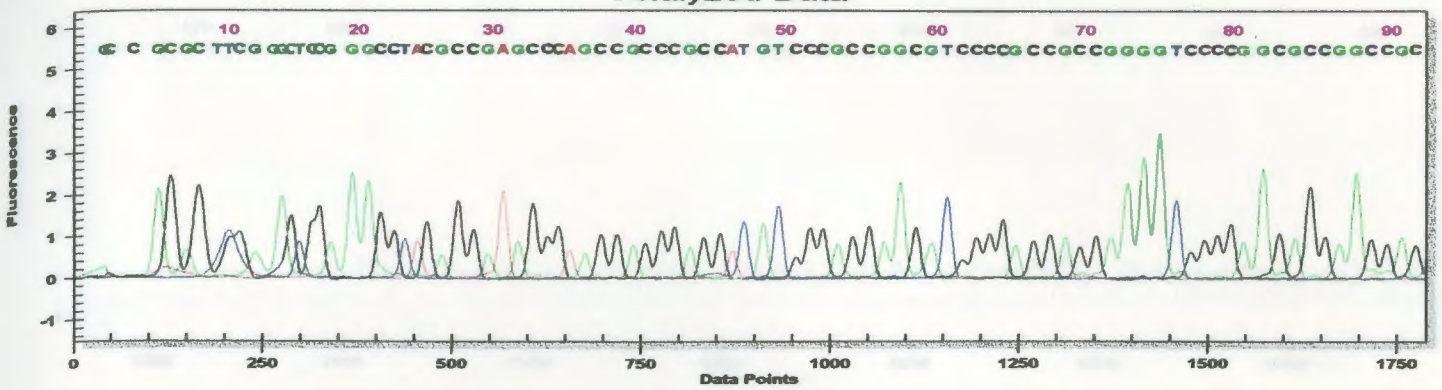
Analyzed Data



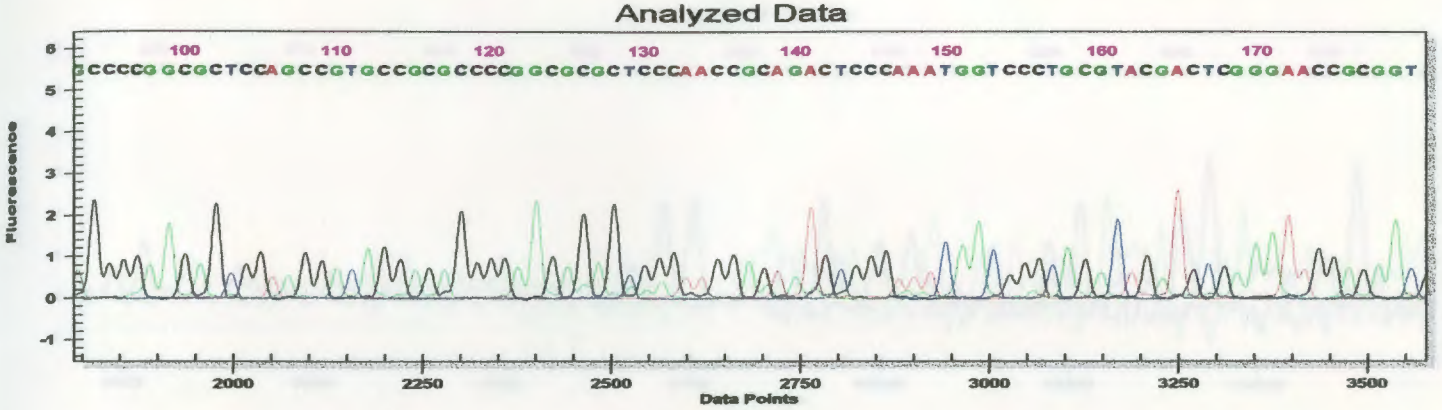
Analyzed Data



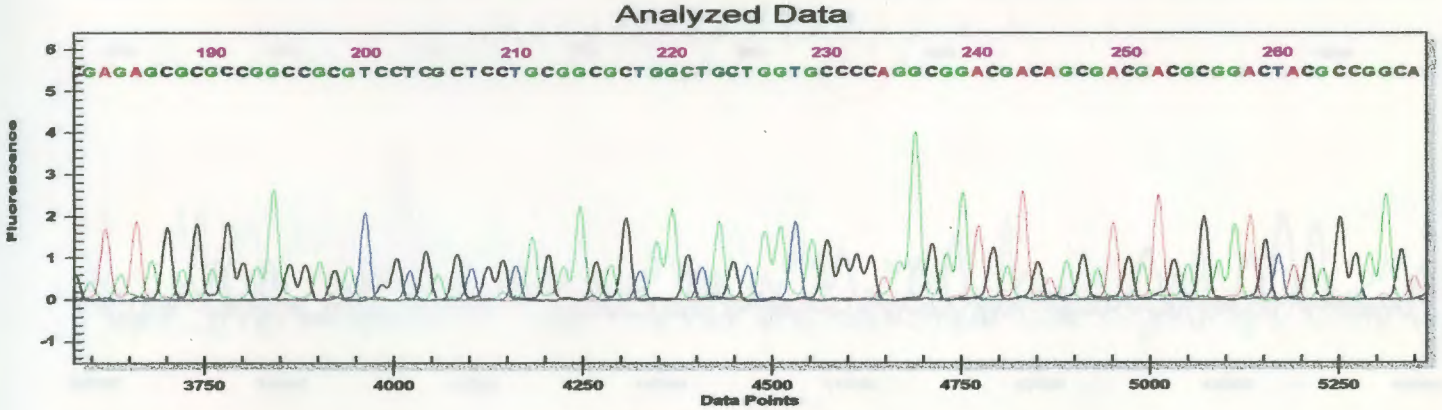
Analyzed Data



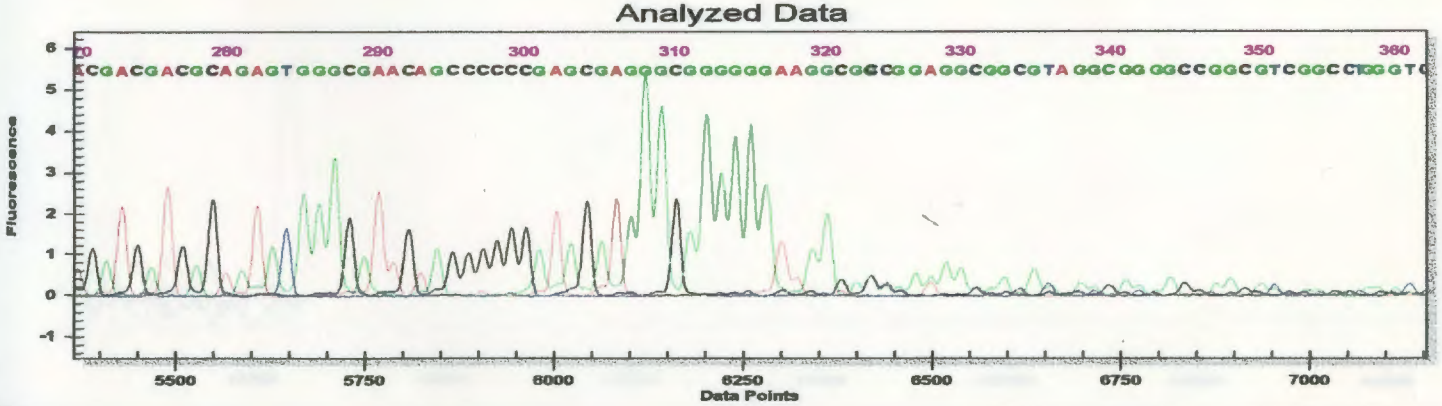
Analyzed Data



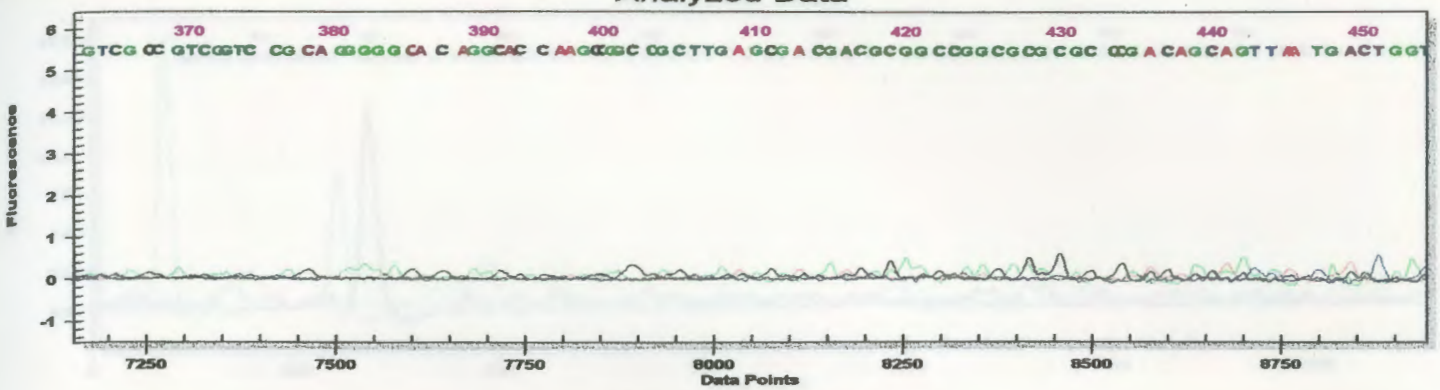
Analyzed Data



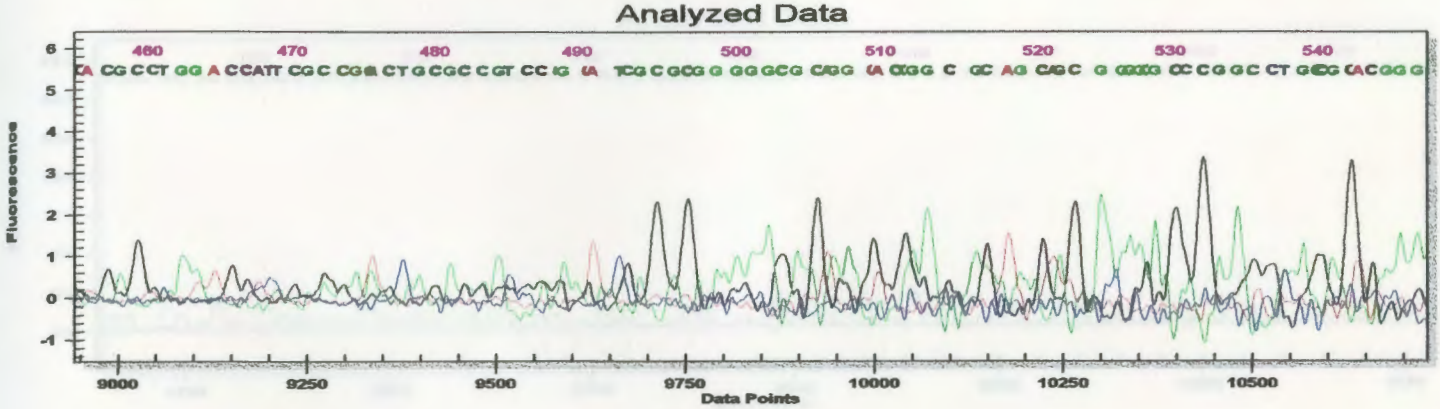
Analyzed Data



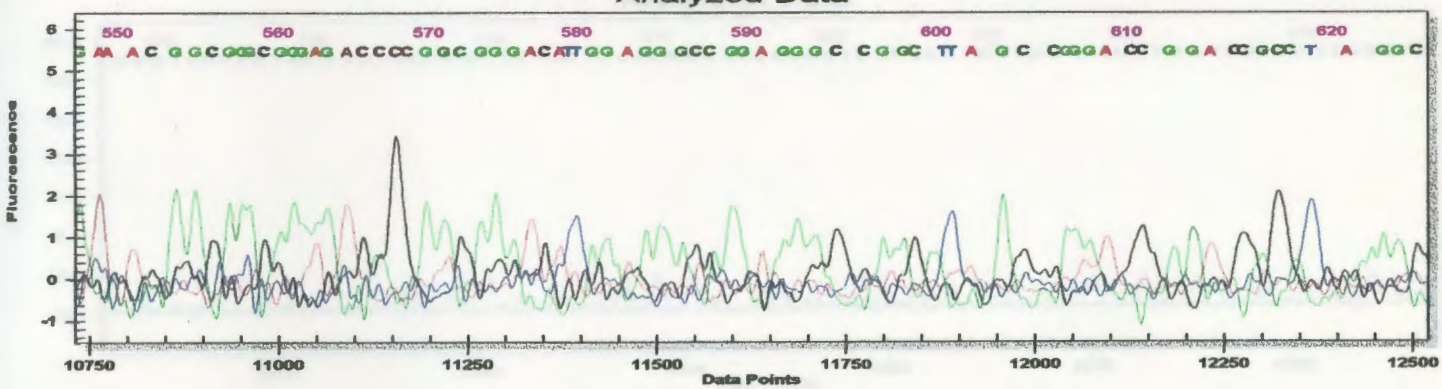
Analyzed Data



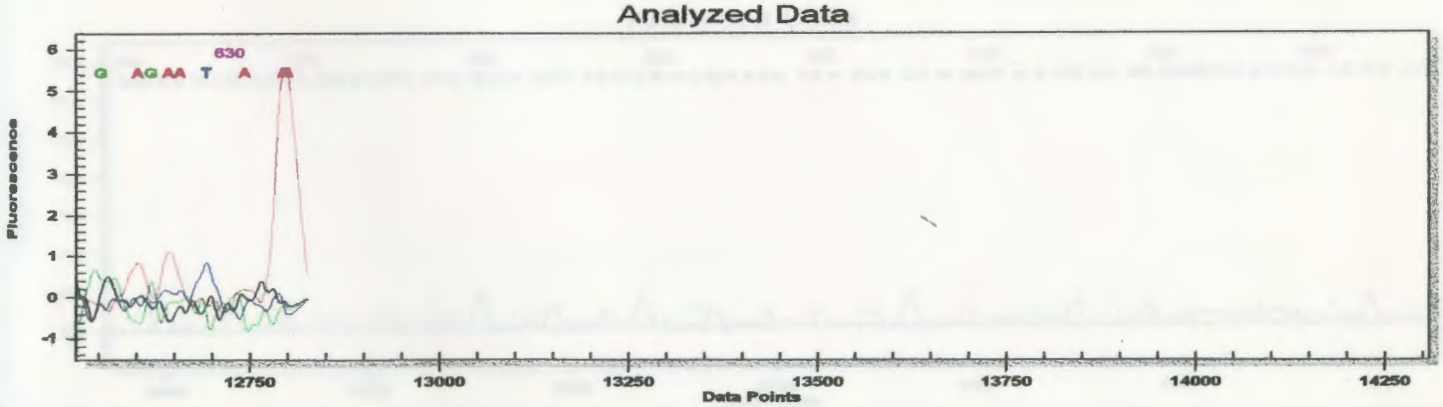
Analyzed Data



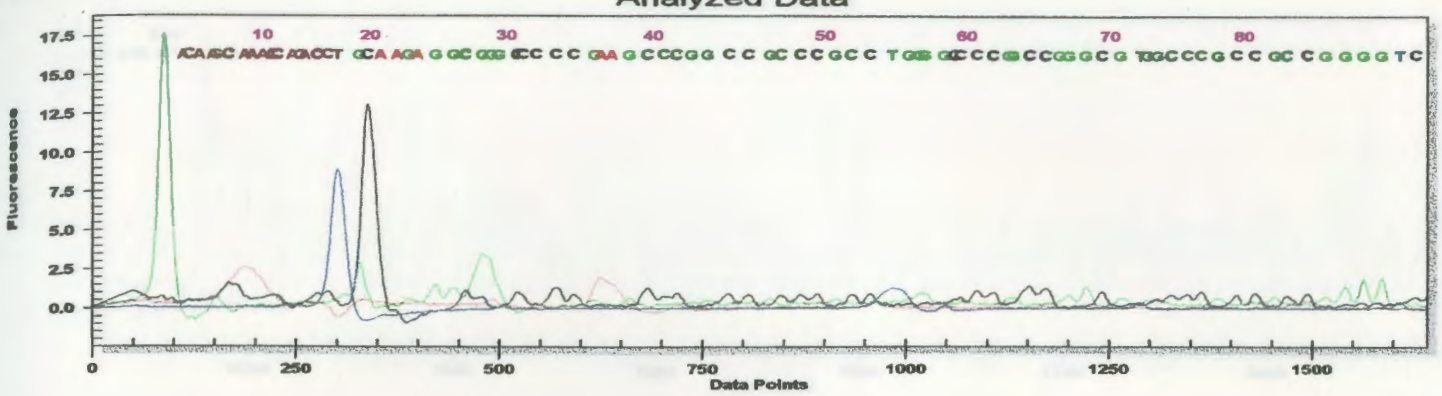
Analyzed Data



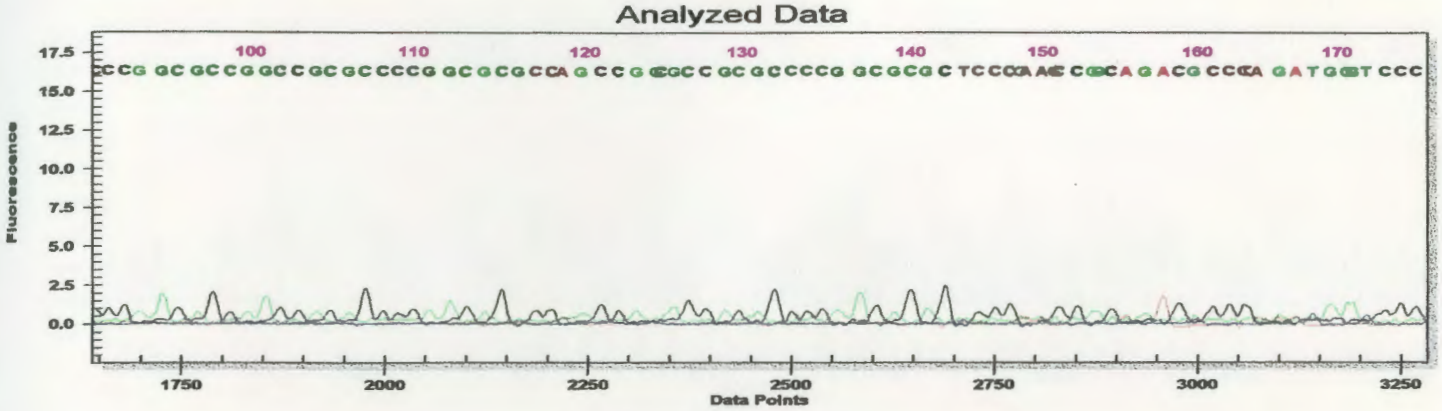
Analyzed Data



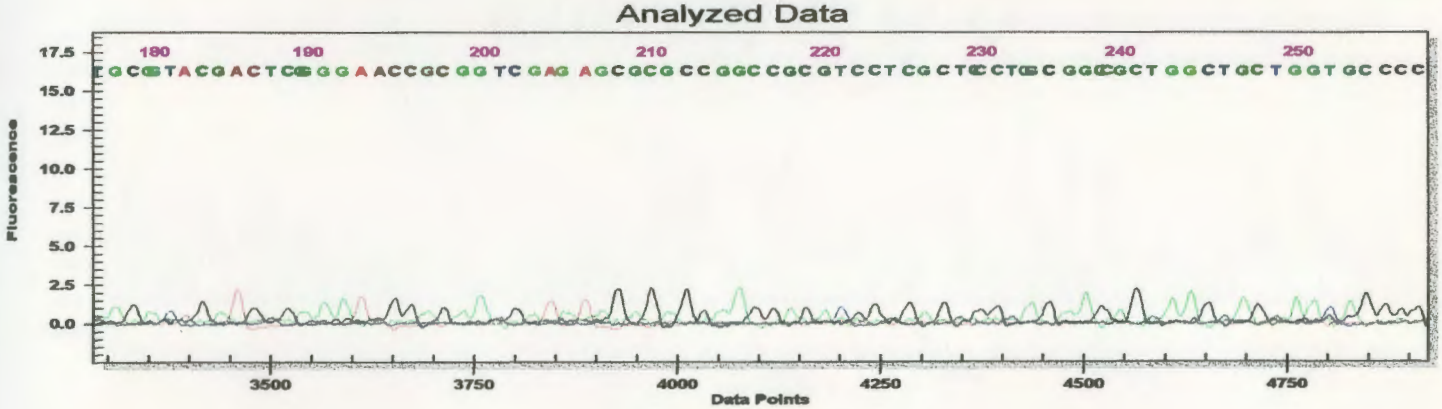
Analyzed Data



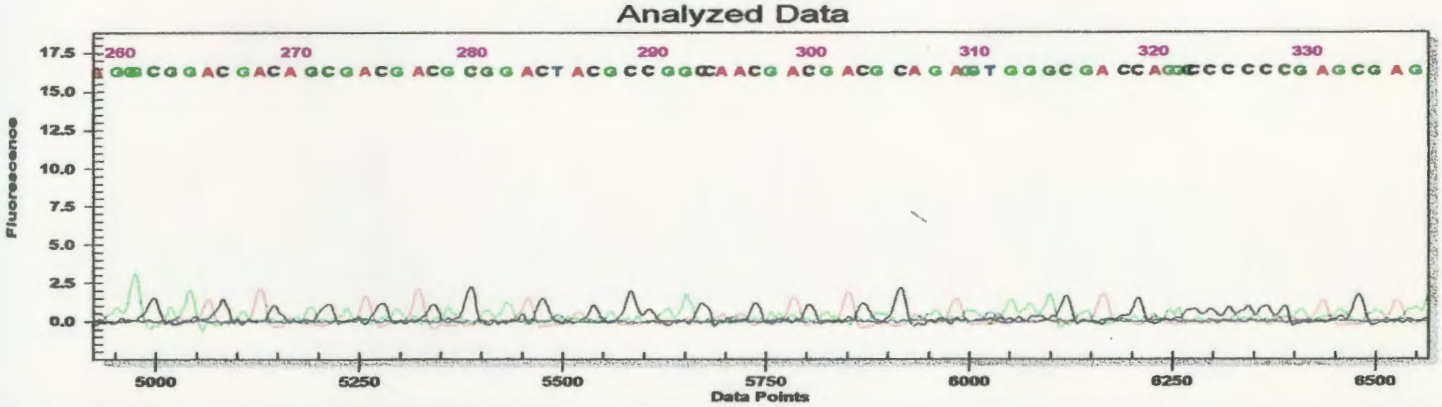
Analyzed Data



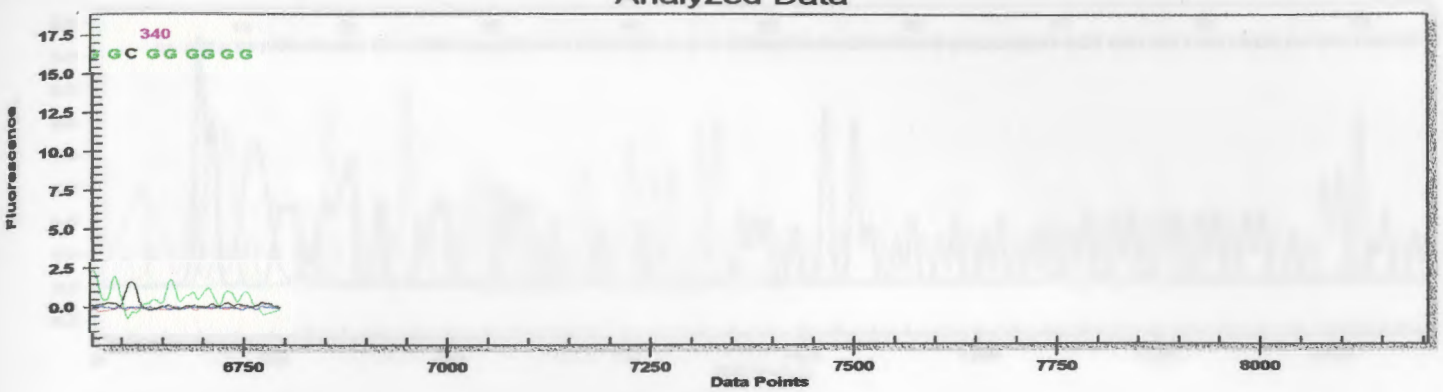
Analyzed Data



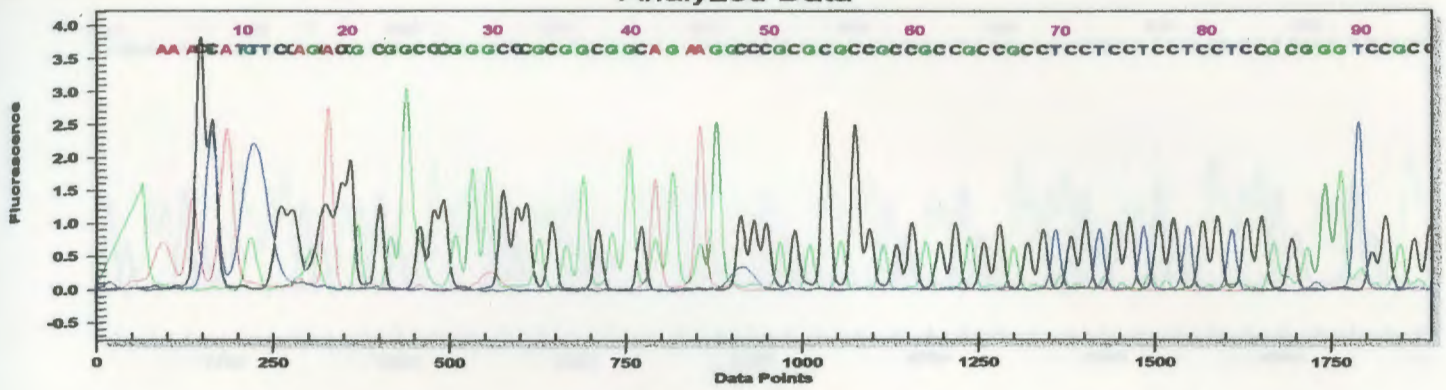
Analyzed Data



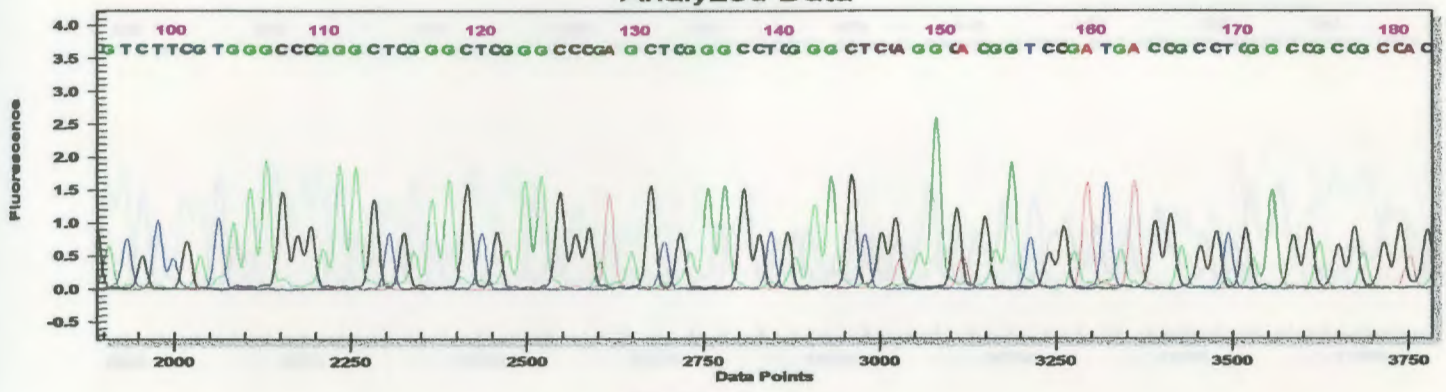
Analyzed Data



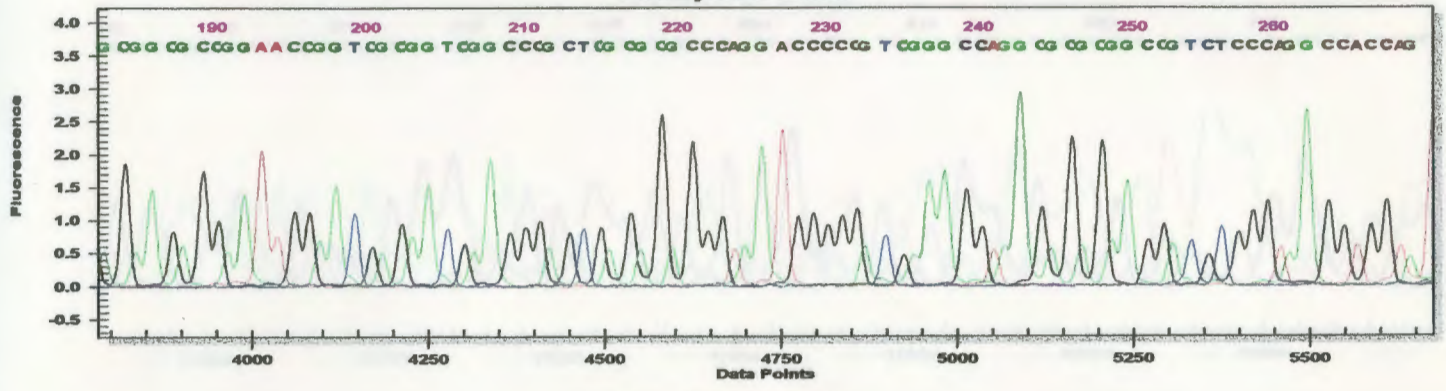
Analyzed Data



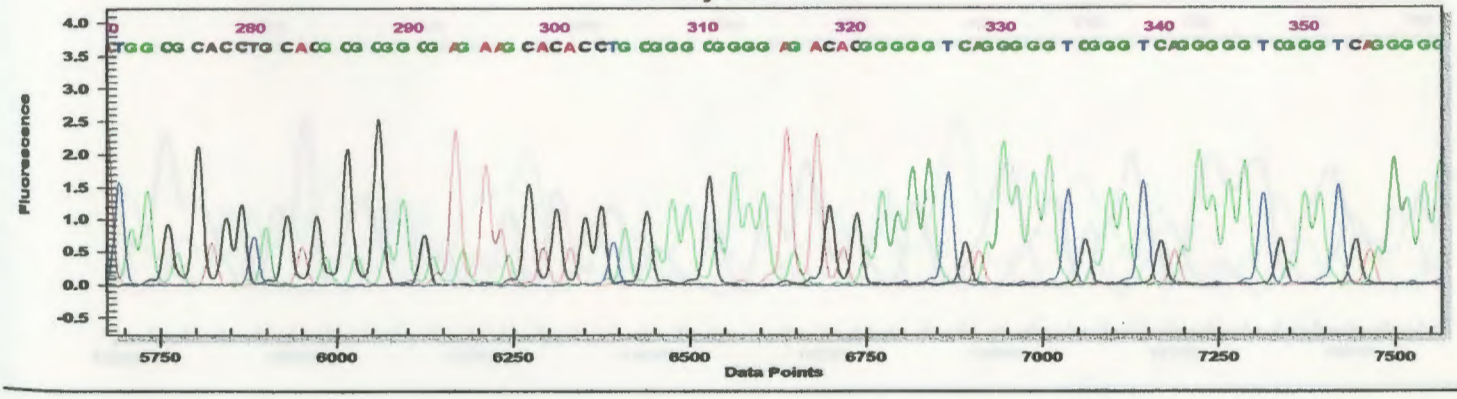
Analyzed Data



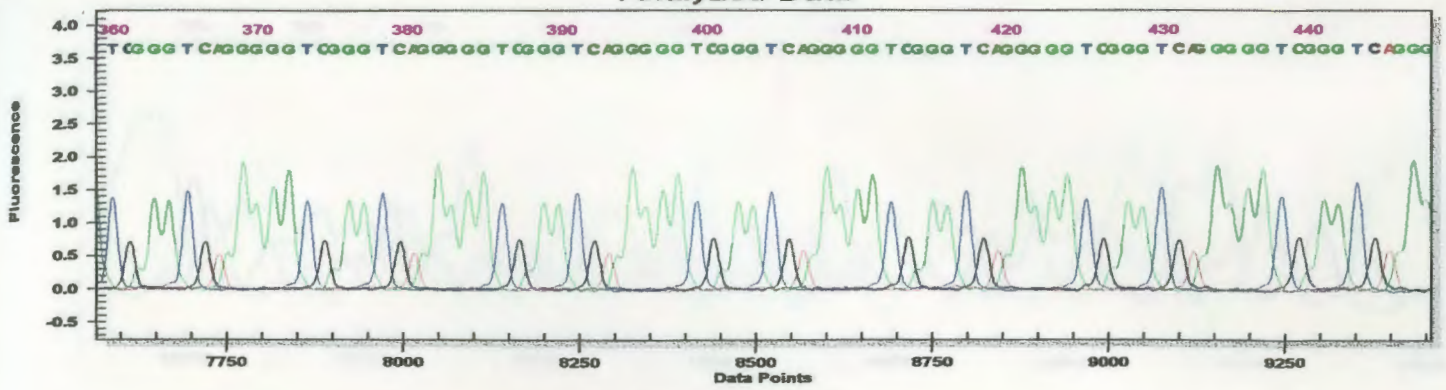
Analyzed Data



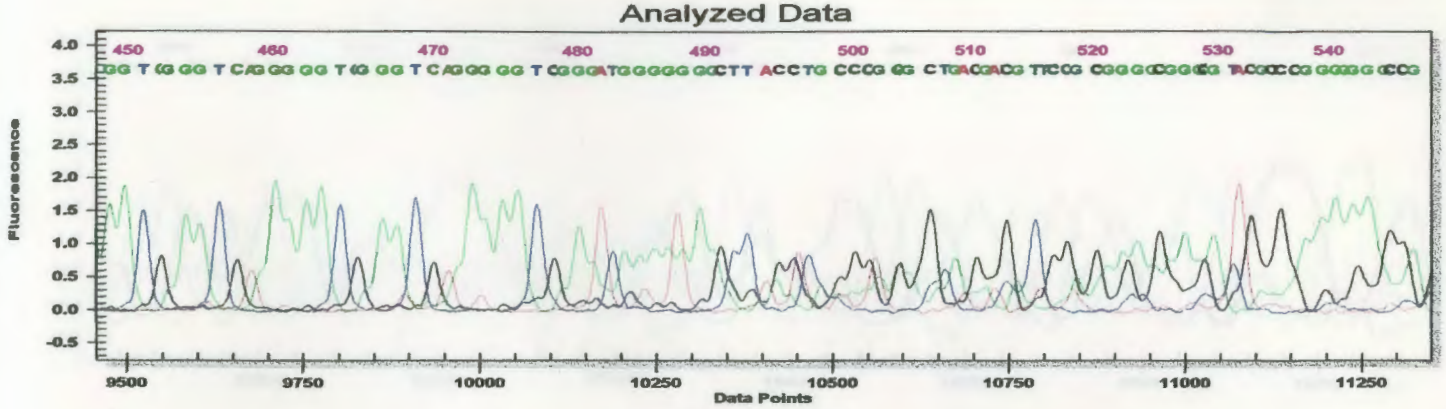
Analyzed Data



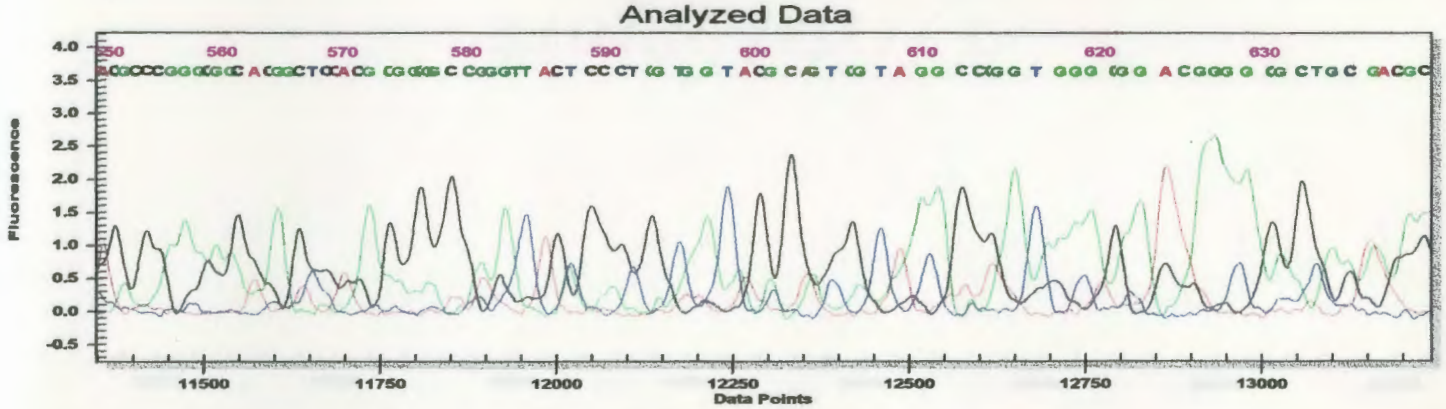
Analyzed Data



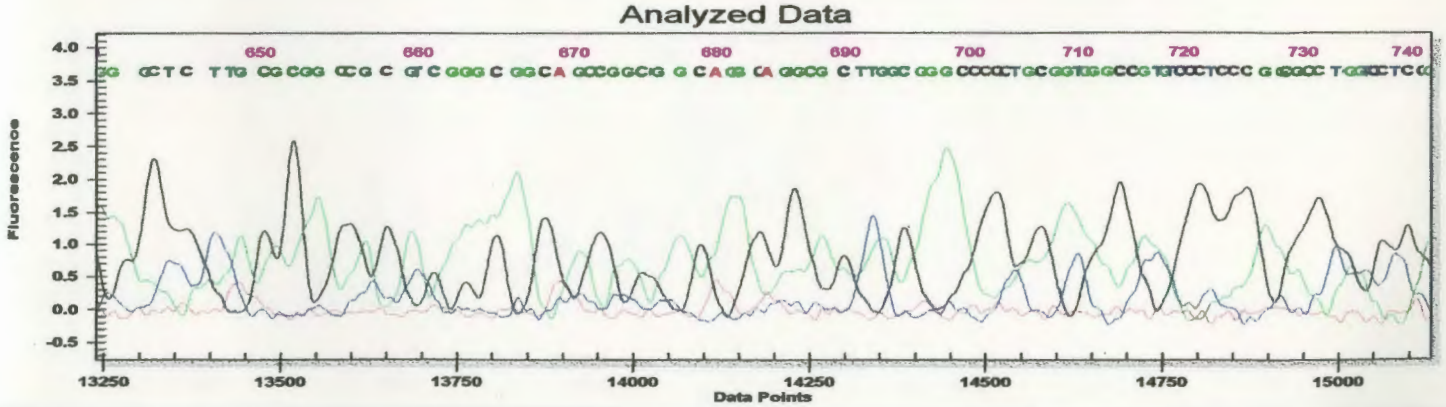
Analyzed Data



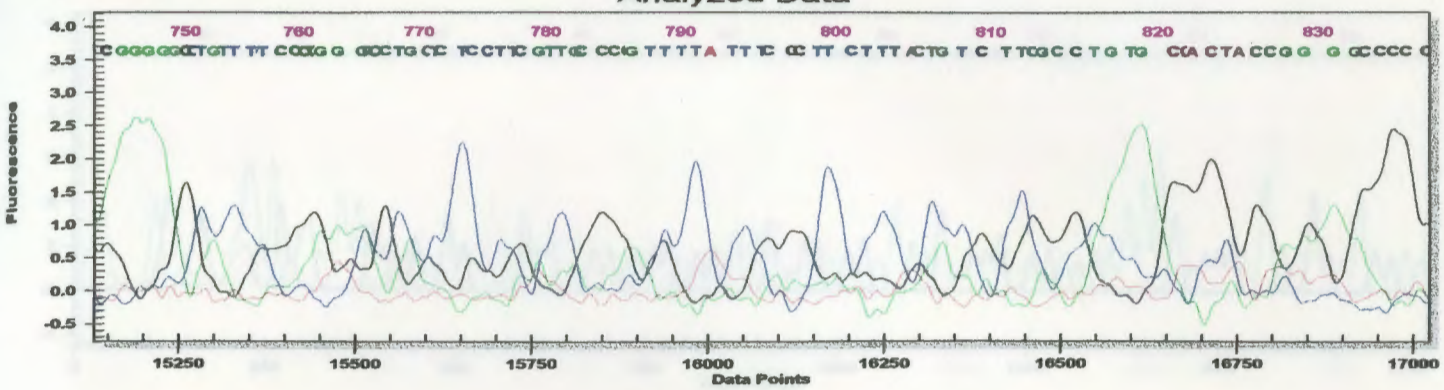
Analyzed Data



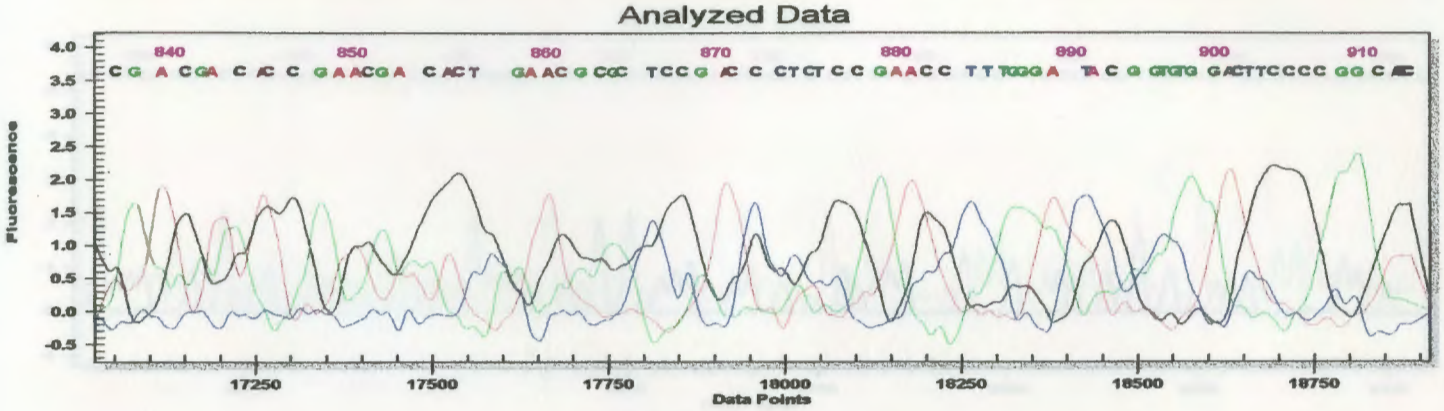
Analyzed Data



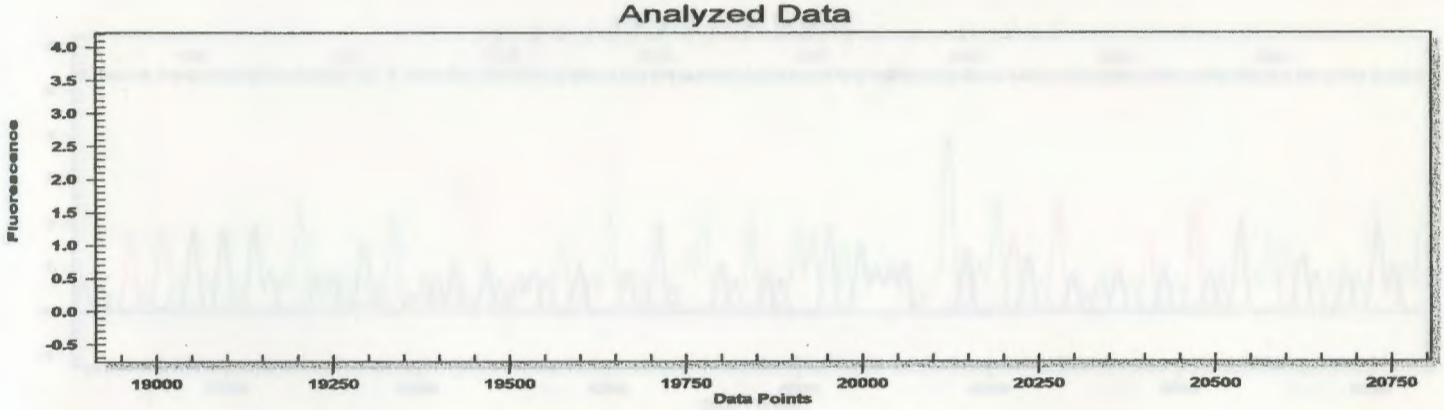
Analyzed Data



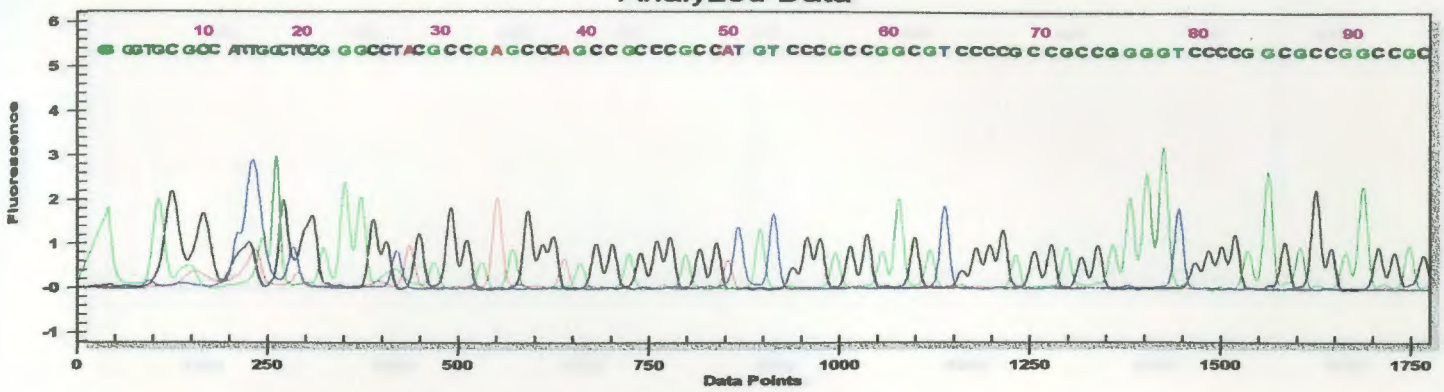
Analyzed Data



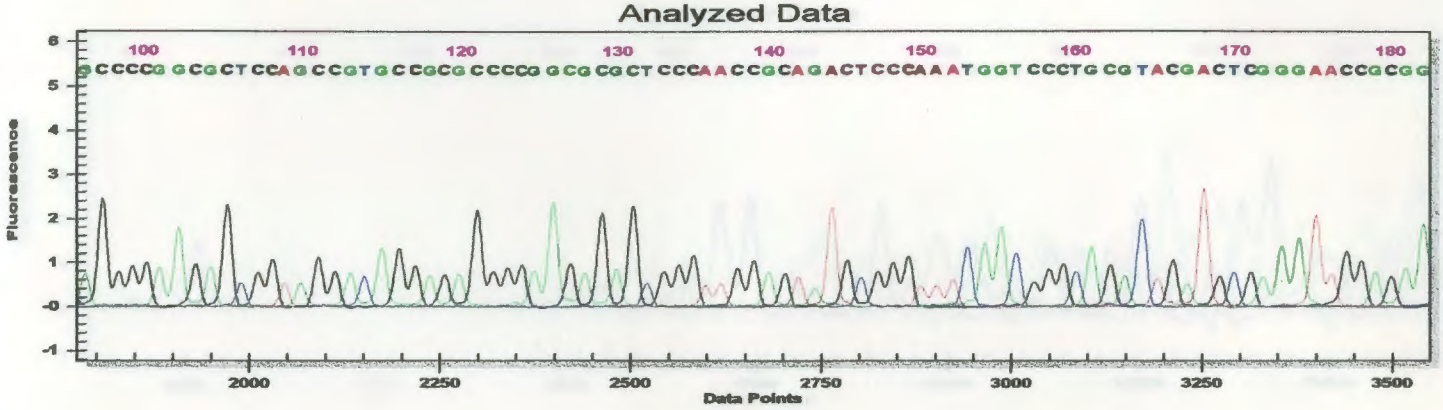
Analyzed Data



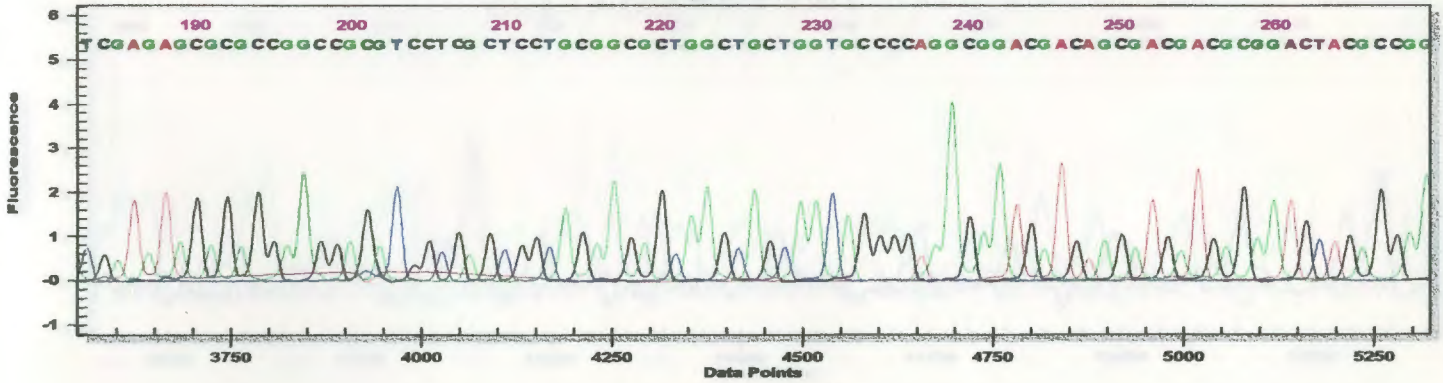
Analyzed Data



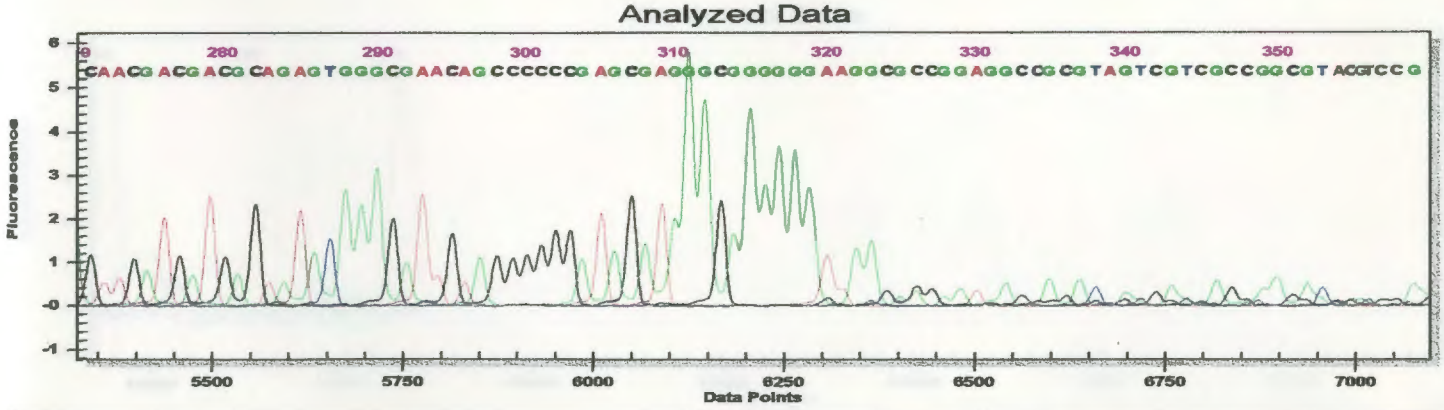
Analyzed Data



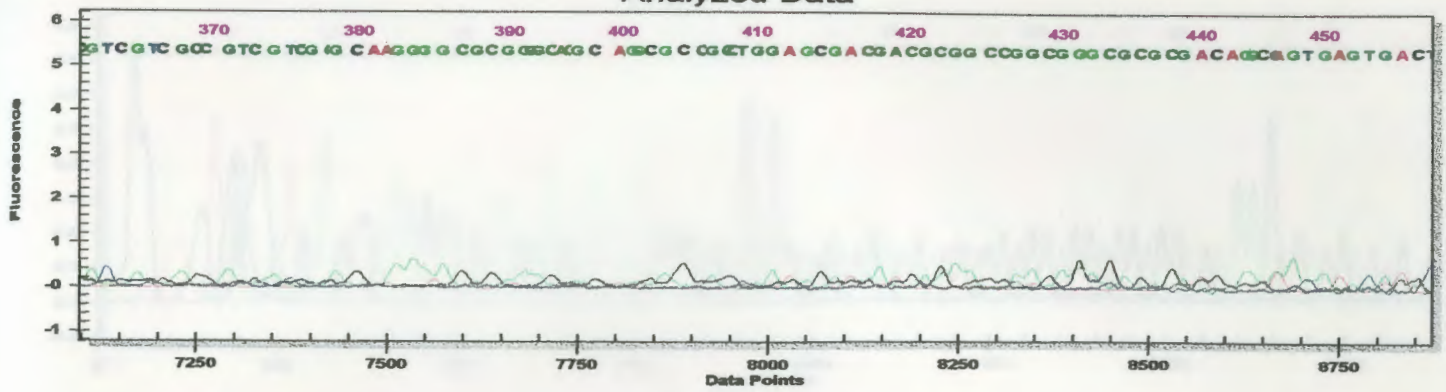
Analyzed Data



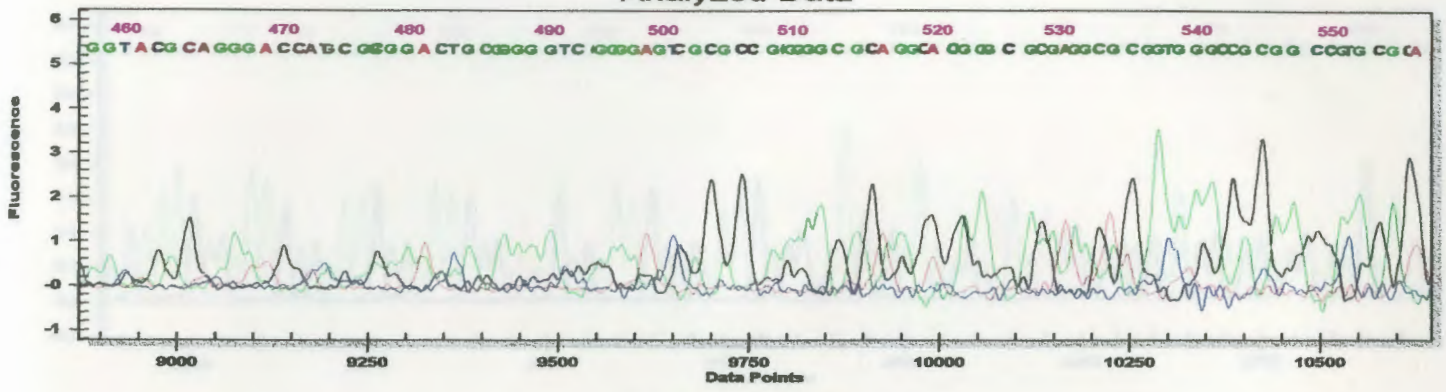
Analyzed Data



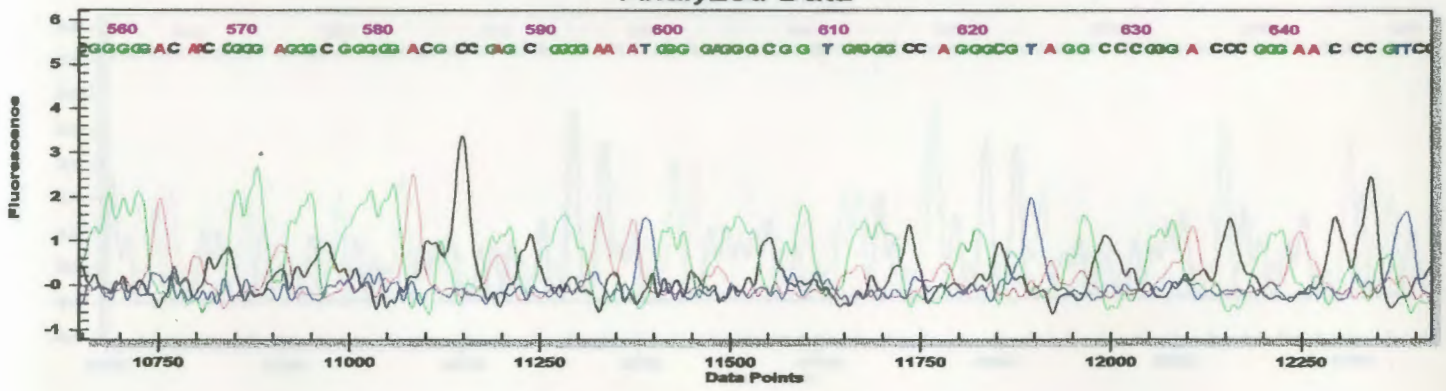
Analyzed Data



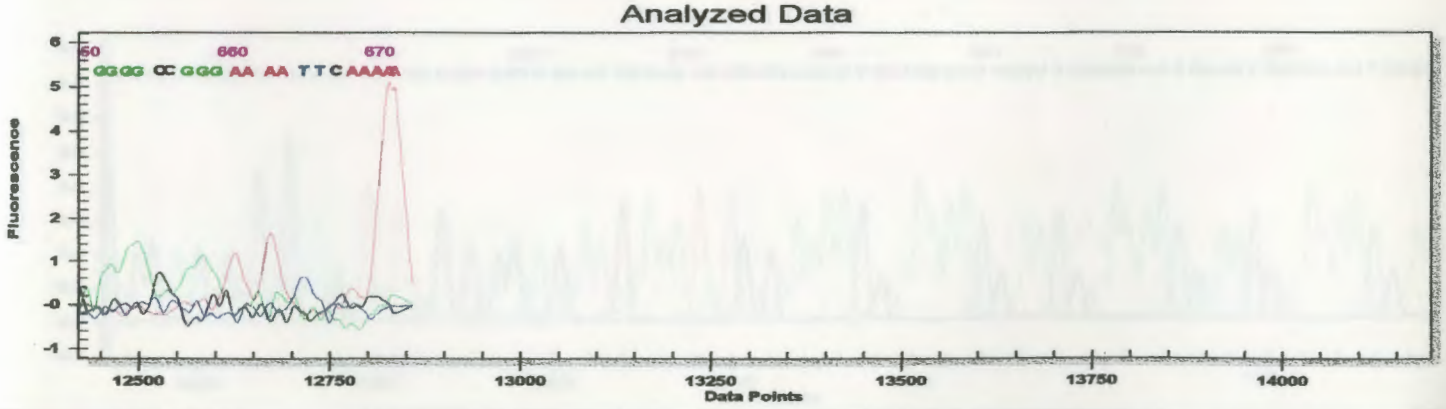
Analyzed Data



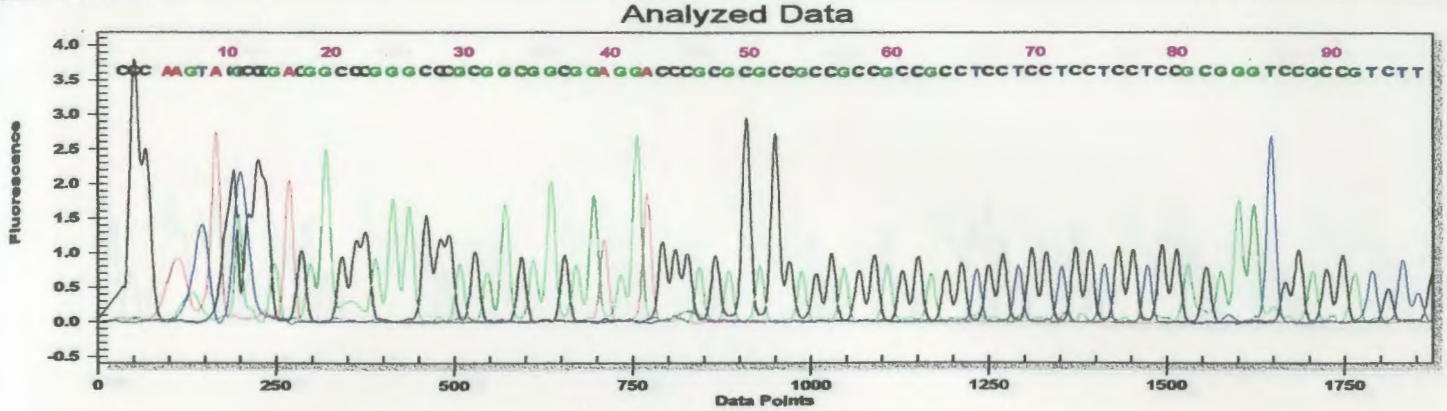
Analyzed Data



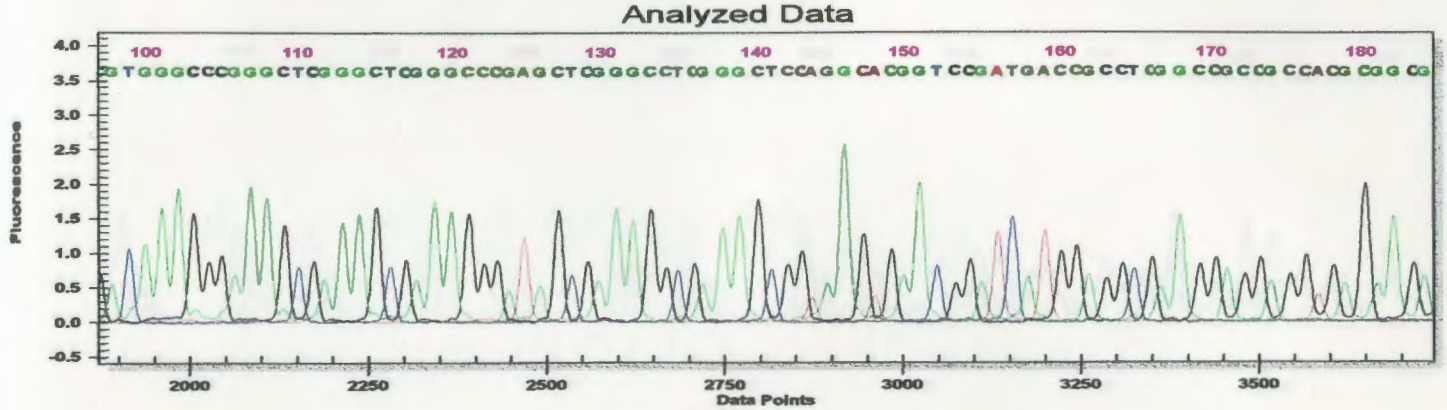
Analyzed Data



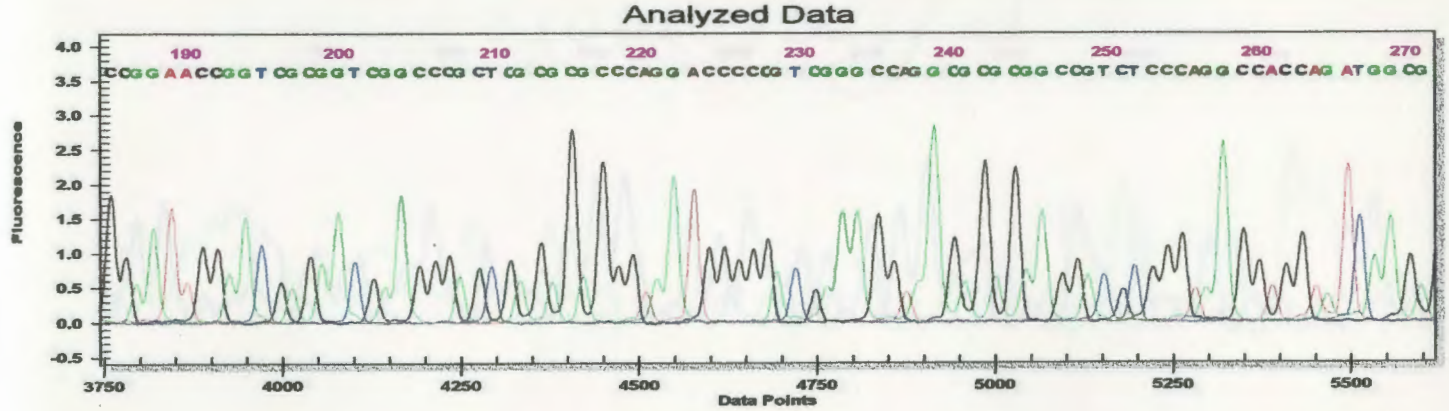
Analyzed Data



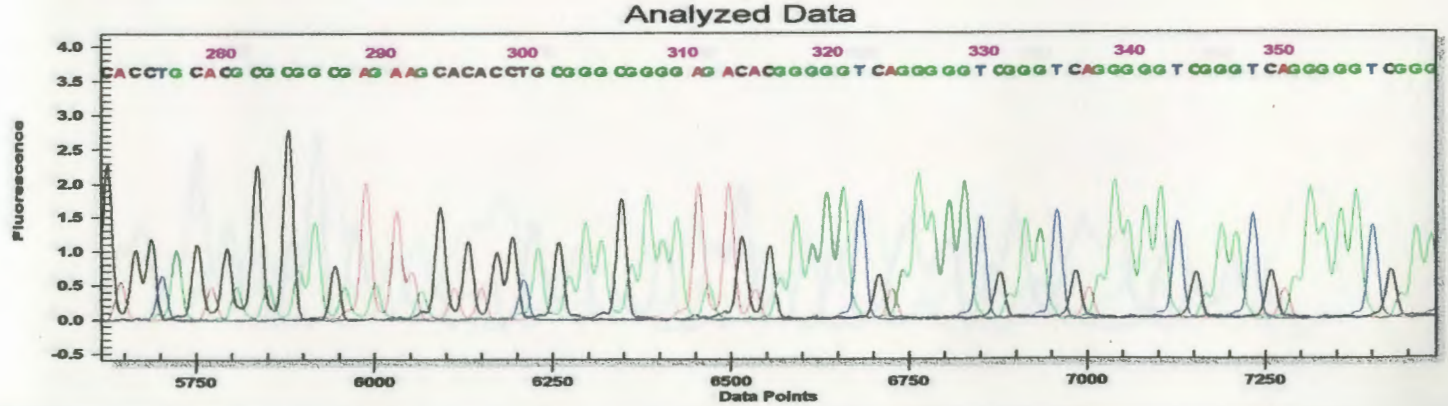
Analyzed Data



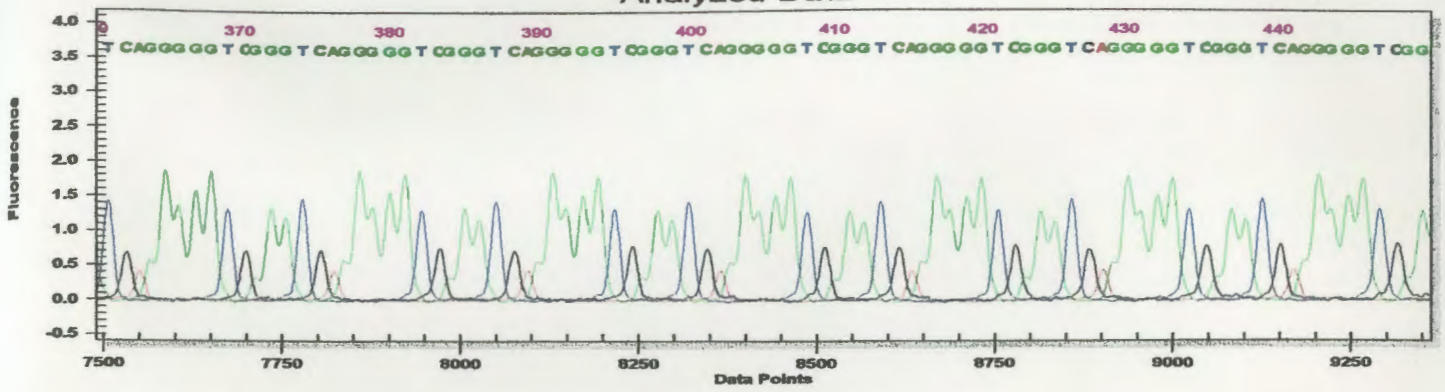
Analyzed Data



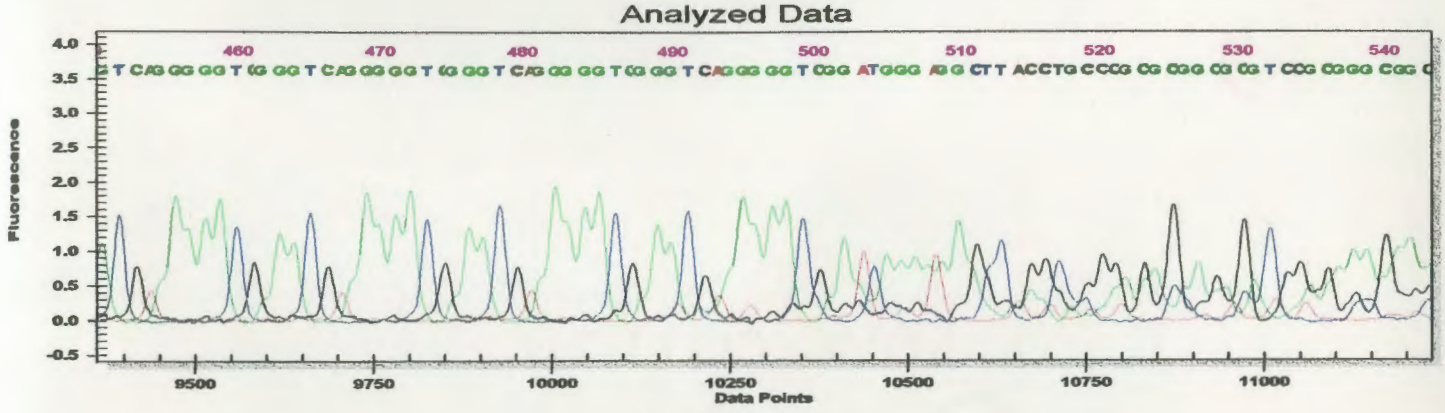
Analyzed Data



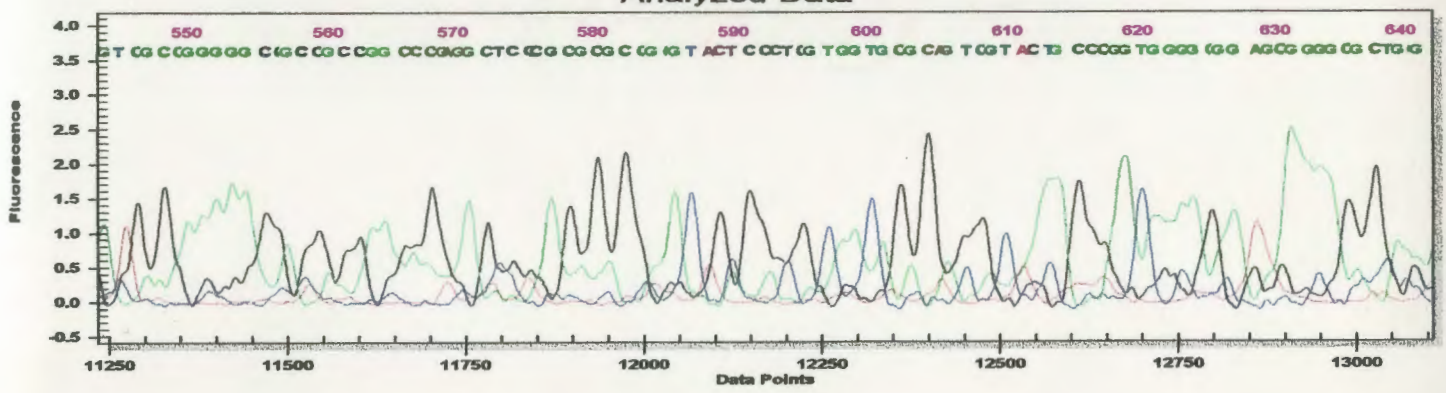
Analyzed Data



Analyzed Data



Analyzed Data



Analyzed Data