

# Optimizing MALDI-TOF MS Sequencing by Changing Reaction Temperatures, Concentrations and Times

Ami Greene  
scyrion@aol.com

Justine Langman  
jasmine9409@yahoo.com

Mark Martinez  
coolmark@yahoo.com

## Abstract

We researched potential optimization of MALDI-TOF by varying the temperature of the sequencing reaction and the concentration of the solutions used. Additionally, we created a computer program to facilitate the sequencing of DNA from the MALDI-TOF data. By lowering the buffer concentration, increasing the dNTP concentration and changing the annealing temperature, we were able to increase the intensity of our mass peaks.

## Introduction

With recent technologies and scientific advances, genomics is a growing and thriving field. Its applications span from the detection of genetic disorders to crime scene analysis. Another important use of genomics can be seen in medicine. For example, analyzing an individual's p53 gene allows scientists to assess his or her risk for tumors. Advancements like this make genomics a field with incredible potential and power. Therefore, there is an increasing need for technologies that sequence DNA efficiently. Current techniques include Sanger Sequencing,

pyrosequencing. A recently developed method utilizes Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) to find the mass of DNA fragments, and use mass differences to calculate the DNA sequence. Whereas Sanger Sequencing and pyrosequencing return unclear data when sequencing hybrid mutations or sections of DNA that are rich in guanine or cytosine nucleotides, MALDI-TOF MS has no such issues. This makes the method quite promising. However, it can also be inconsistent. Obtaining a usable result might require two weeks or more of fine-tuning concentration, temperatures, and other variables. Additionally, lab workers must analyze the data provided by the MALDI-TOF machinery by hand in order to translate it into a DNA sequence. Our project attempts to improve and optimize the MALDI-TOF analysis of gene p53 by changing the concentrations of various reactants, as well as the heat and time settings for the sequencing cycle. We also focused on increasing the efficiency of MALDI-TOF by writing a program to automatically convert the data into the

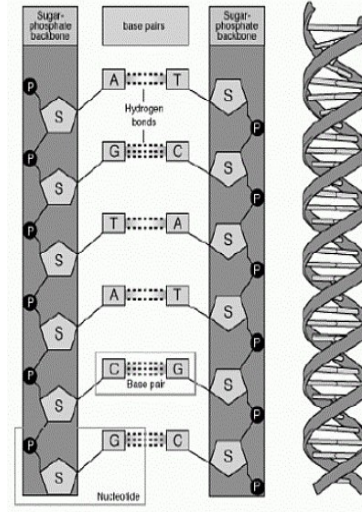
gene's sequence instead of analyzing the data by hand.

### Background

Gene p53 is a tumor-suppressor gene, which is critical to the prevention of cancer. Not only does it play a role in DNA repair, it also inhibits the growth of defective cells and initiates the apoptosis, or programmed cell death, of mutant cells. As such, it actively prevents the growth of tumors. With only one functioning copy of gene p53, an individual will probably develop tumors by early adulthood. More than half of tumors in humans correspond with an error in gene p53. Insuring the effectiveness of sequencing methods that analyze p53 mutations is thus very important.

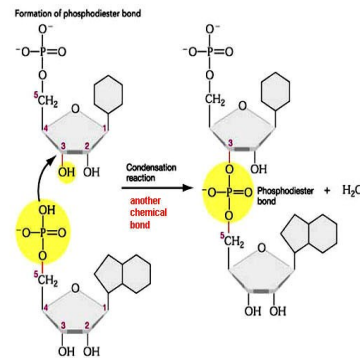
To understand the analysis of p53 via MALDI-TOF, one must first understand certain basic concepts about DNA and its structure. DNA is a molecule formed by nucleotides, or dNTPs. These are smaller molecules that are made of a phosphate group, a deoxyribose sugar, and one of four nitrogenous bases: adenine, thymine, guanine, or cytosine (A, T, G, and C). It is shaped as a twisted helix, or "ladder", where the outside "rails" are formed by the phosphate group and sugar of the joined nucleotides, while the center "rungs" are composed of the four bases. While the order of these bases is very complex, they are always paired such that A complements T, and G complements C, due to the molecular structure of each of the bases. (See Diagram 1). These nitrogenous bases are held together by hydrogen bonds. The sugar-phosphate backbone is held together by phosphodiester bonds between the hydroxyl group on the third

carbon of each sugar and the hydrogen on the end of the phosphate group of the next nucleotide. (See Diagram 2). This becomes important during DNA replication.



**Diagram 1:**

This diagram elucidates the structure of DNA by demonstrating its twisted helix shape and the manner in which nucleotides are linked and arranged and connected.

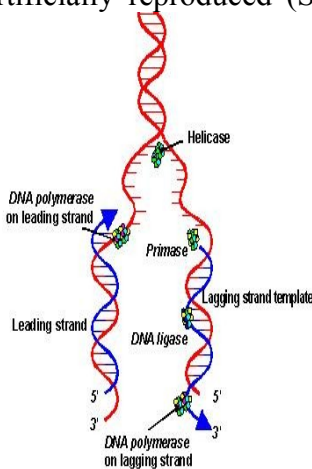


**Diagram 2:**

This image helps to clarify the nature of phosphodiester bonds, which is critical to understanding the function of ddNTPs. The hydroxyl group from the fifth carbon of one nucleotide's sugar bonds with the hydrogen from the hydrogen on the third carbon of the next. They join, forming a molecule of water and leaving the two nucleotides bonded together.

DNA is replicated when a helicase enzyme "unzips" the two

complementary strands of DNA by breaking the bonds between nucleotide pairs. Another enzyme called primase adds an RNA primer to each strand of DNA, which allows another enzyme, DNA polymerase, to begin adding new nucleotides to the primer. This elongation of the new strand of DNA by DNA polymerase is called extension. Other enzymes proofread and refine this process. This replication can be done automatically through the polymerase chain reaction (PCR). A solution containing DNA, primers, dNTPs and the necessary enzymes is cycled through a series of temperatures that separate the strands of DNA, allow the primer to anneal to the DNA strand, and optimize the performance of the enzymes. Respectively, these steps are known as denaturation, annealing and extension. When this cycle is repeated several times by a machine called a thermocycler, massive amounts of DNA can be artificially reproduced (See Diagram 3)

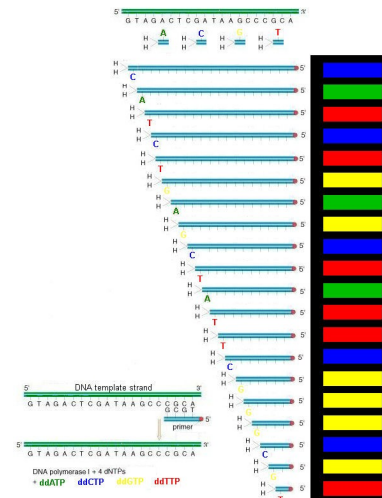


**Diagram 3:**

This diagram helps to clarify DNA replication, and thus the sequencing reaction used for MALDI-TOF MS. It shows how replication is semi-conservative, with one daughter strand being formed from each of the parent strands. Helicase is shown separating the

two parent strands, and DNA polymerase is also depicted as it extends the new strands of DNA.

Many methods of gene sequencing, including MALDI-TOF, utilize this replication process. For MALDI-TOF, nucleotides called ddNTPs are added to the PCR reaction. Whereas dNTPs have a deoxyribose sugar, ddNTPs have a dideoxyribose sugar. Unlike deoxyribose, dideoxyribose doesn't have a hydroxyl group at the third carbon, making it incapable of bonding to the next nucleotide. Thus, when a polymerase adds a ddNTP to a growing strand of DNA, the replication of that strand terminates. Since ddNTPs are incorporated randomly and DNA is replicated numerous times during the PCR protocol, there is a high probability that the reaction shall produce DNA fragments that stopped at every base of the DNA sequence. (See Diagram 4) The labeling of ddNTPs corresponds with the base each nucleotide includes. For example, a ddNTP that had adenine for its nitrogenous base would be called ddATP, while one that contained guanine would be called ddGTP.



**Diagram 4:**

This diagram clarifies the sequencing reaction used in Sanger sequencing and, to a certain extent, in MALDI-TOF sequencing. One can see how the terminating ddNTPs end

replication prematurely, and how these fragments eventually end at every nucleotide in the sequence. If the fragments can be sorted by length, and if the terminating ddNTP of each fragment can be read, then the DNA can be sequenced.

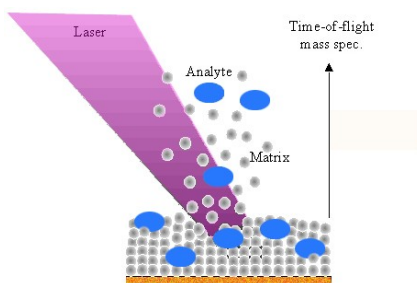
Where Sanger sequencing uses fluorescence to detect these bases after gel electrophoresis sorts them by mass, MALDI-TOF MS determines the mass of each of these strands, and then uses the difference between them to calculate the correct sequence. Therefore, in order to use this method, the DNA fragments must first be separated from enzymes and other molecules that were used in the reaction.

The filtration of DNA fragments is an intricate process. This separation can be achieved since DNA is a polar molecule and carries a small fraction of a negative electrical charge, specifically in the sugar-phosphate backbone. Because of this, it is attracted to molecules with positive charges. Zip-tips, commonly used for filtration, take advantage of this. Zip-tips consist of a pipette tip with a small, slightly positive filter that binds to DNA. First, the filter is rinsed with the enzyme AZT which causes the filter to drop any molecules that it contains.

ionizing the DNA fragments and releasing them from the matrix.

A buffer is then drawn through the filter, which optimizes the DNA's binding. The DNA sample is then drawn through the filter, which is followed by more buffer. At this point, the DNA is caught in the filter of the Zip-tip and separated from the rest of the solution. Water is then drawn through the filter to remove any molecules that didn't bind to the filter. The purified DNA is then released into a separate tube with AZT.

MALDI-TOF analysis is also complex. The purified DNA fragments are combined with a matrix and dried quickly on a plate, forming crystals. When a laser hits these crystals, it frees the particles mixed with the matrix. Furthermore, these particles receive electrons from the matrix and become ionized. The now-freed particles fly up a vacuum tube that has voltage running through it. The speed of their flight is determined by their mass-to-charge ratio: the larger a molecule is, the more slowly it travels up the tube. At the other end of the vacuum tube, the particles hit a detector. A computer then uses the detected times of flight to calculate the mass per charge of the particles, and provides a graph of mass per charge vs. intensity (See Diagram 5).



**Diagram 5:**  
MALDI-TOF MS

This diagram clarifies the process of MALDI-TOF. It shows how the laser functions in

## Methods

Due to the difficulty of obtaining p53, we used the Rhodopsin gene as a model during our experiments. The section of Rhodopsin had a known sequence of GCATGCCCTCAGGGAT GTACCTGGACCAGCCAACGATGG GGGAGCAGCACAGGCCAACGCCA TGATCCAGGTGAAGA and a primer with the sequence TCTTCACCTGGAT

CATGGCGTT. This foreknowledge of the DNA's sequence allowed us to differentiate between accurate, useful results and misleading ones.

To optimize the MALDI-TOF process, we attempted to increase the output of the sequencing reaction. Thus our efforts focused on changing thermocycling times and temperatures, as well as the concentrations of reactants, instead of filtration methods.

Our first set of reactions was a control, containing the standard

concentrations of all reactants. Specifically, we used 1.0  $\mu\text{l}$  of a 10  $\mu\text{M}$  DNA solution, 2.0 $\mu\text{l}$  of a 10  $\mu\text{M}$  primer solution, 1.0  $\mu\text{l}$  of a dNTP solution with 2.5  $\mu\text{M}$  of each dNTP, 5 $\mu\text{l}$  each of a 5.0  $\mu\text{M}$  solution ddATP, ddGTP, ddCTP and ddTTP, 2.0 $\mu\text{l}$  of buffer, 1.0 $\mu\text{l}$  of enzyme, and 12.5 $\mu\text{l}$  of  $\text{H}_2\text{O}$ . The annealing temperature of the solution was 57°C and the time for annealing and extension was 30 seconds.

The various conditions that we tested can be seen in Table 1.

**Table 1: Tested Conditions**

<b>Condition:</b>	<b>Change:</b>
<b>1</b>	Standard Conditions; Control
<b>2</b>	Increase annealing time from 30 seconds to 40 seconds
<b>3</b>	Increase extension time from 30 seconds to 40 seconds
<b>4</b>	Increase annealing and extension time from 30 seconds to 40 seconds
<b>5</b>	Increase annealing temperature from 57 C to 64 C
<b>6</b>	Decrease amount of buffer from 2 $\mu\text{l}$ to 1.6 $\mu\text{l}$
<b>7</b>	Increase amount of dNTP from 1 $\mu\text{l}$ to 2 $\mu\text{l}$
<b>8</b>	Change amount of ddNTPs from .5 $\mu\text{l}$ each to .6 $\mu\text{l}$ ddGTP, .62 $\mu\text{l}$ ddCTP, .225 $\mu\text{l}$ ddATP and .55 $\mu\text{l}$ ddTTP

To combat this problem, the reactant concentrations were increased. This should subsequently increase the probability that the molecules would collide with each other and react. Other trials focused on tailoring the annealing temperature to our primer and giving the annealing and elongation parts of the reaction more time to progress. For the 8<sup>th</sup> reaction, we tailored the concentrations of the ddNTPs to correspond with the ratios of those nucleotides in the DNA sequenced.

After all of the solutions were mixed and the sequencing reaction was run in the thermocycler, the DNA

fragments were filtered using zip-tips. The zip-tips were flushed four times with 10  $\mu\text{l}$  of AZT, four times with the buffer solution, ten times with the DNA sample, twice again with the buffer solution, and lastly six times with water. AZT was then drawn through the zip-tip a final time, and the solution was deposited in a clean test-tube.

These test tubes, which now contained the filtered DNA fragments, were placed in a vacuum centrifuge for 15 minutes. This dried the solutions, leaving the DNA at the bottom of each test tube. The DNA from each tube was re-suspended in 1  $\mu\text{l}$  of water, and mixed with 1  $\mu\text{l}$  of 3-Hydroxypicolinic acid

matrix. Two 0.5  $\mu$ l samples from each sample were then placed on a steel template for the MALDI-TOF machine and analyzed.

The output of the MALDI-TOF machine does not give a direct nucleotide sequence. Because of the shortcomings of the actual results of the MALDI-TOF, we have constructed a program that can directly read the data from the MALDI-TOF into a nucleotide sequence. The numerical data from the MALDI-TOF is directly inputted into the program and the nucleotide sequence is given as an output. Along with the nucleotide sequence, the area where the mutation is located is verified and the type of complications from the mutation is also identified.

## Results

Our experiments met with mixed success. For the control samples in Figure 1, three small peaks could be observed: one corresponding to the primer at 6690 g/z and two peaks that represent the first two nucleotide extensions of the sequence (GG). The second peak is higher than the other peaks, while the others are small and difficult to distinguish from the surrounding noise. Additionally, what appears to be a second peak does not actually correspond to a nucleotide. This lack of height and length were the two factors we attempted to remedy.

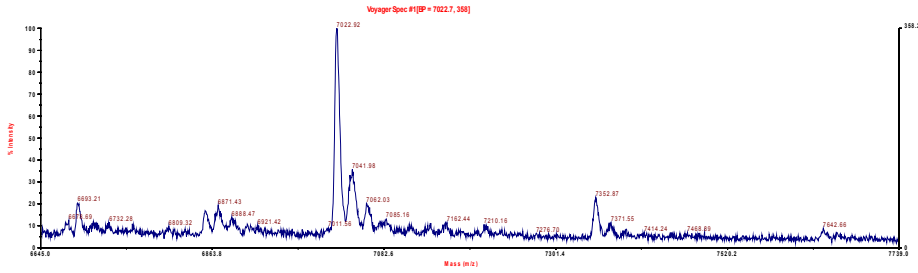
We classified negative results as data that was either tantamount to the control data or less useful than the

control data. Negative results were received in conditions 2, 3, 4 and 8. Not only are less peaks apparent in these conditions, but noise is also increased, making the graphs even less readable. This is a regression from the control data. Figure 2 displays examples of negative results.

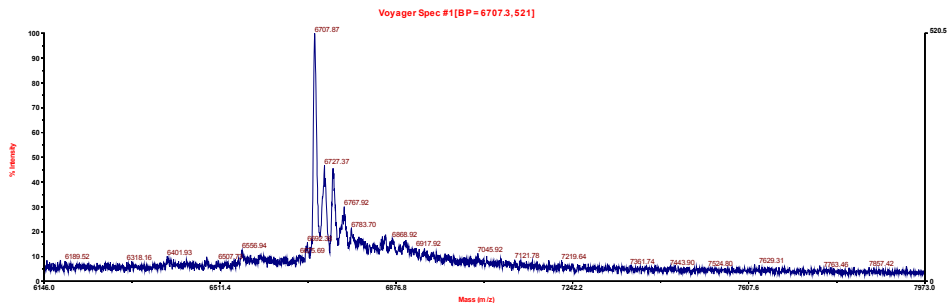
Results in conditions 5 and 6 showed improvement in the intensity of peaks, though this does not imply that the data was easier to analyze. Condition 5, in which the annealing temperature was raised, displayed an increase in peak size, although a simultaneous increase in noise renders the peaks less distinguishable. The peak size increase is still an improvement despite the noise because it signifies that more reaction took place. In condition 6, four peaks of greater intensity than those of the control were visible. Once again, though, the peaks were difficult to distinguish amidst increased noise. In addition, the peaks were wider than in the control sample, making the nucleotide more difficult to identify. Figure 3 shows the data from conditions 5 and 6.

Condition 7 provided the most positive results, meaning that the data was easier to analyze than the control data. The peaks are of higher intensity than those of the control sample. As opposed to conditions 5 and 6, noise is low in condition 7, making peaks more distinct and easily readable. This data can be seen in figure 4.

**Figure 1:** Here in Condition 1, four mass peaks were observed: the primer at 6690 g/z and three more at 6811 g/z, 7022 g/z, and 7352.81 g/z. These correspond to the bases cytosine, cytosine, and guanine. The four peaks are rather wide and relatively small, with the exception of the third.

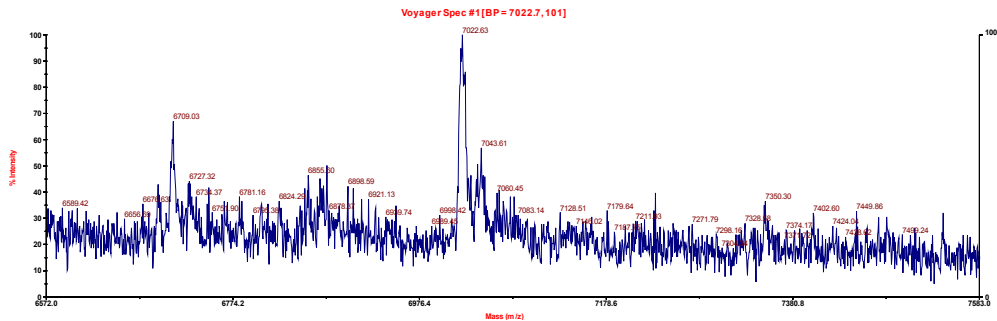


**Figure 2:** This data demonstrates the general results from conditions 2, 3, 4 and 8. Peaks are less distinguishable, and noise is increased.

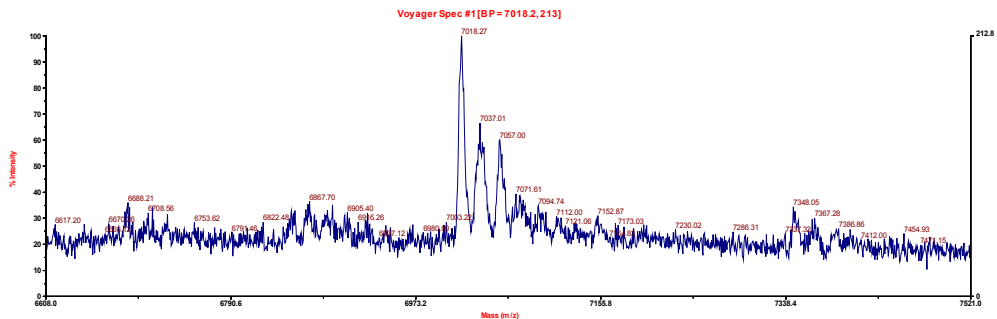


**Figure 3:** a. The results from condition 5 show high intensity peaks but an increased amount of noise. b. Condition 6 also displays high intensity peaks such as in Figure 3a; while noise is reduced compared to Figure 3a, the noise still detracts from the analysis of the data.

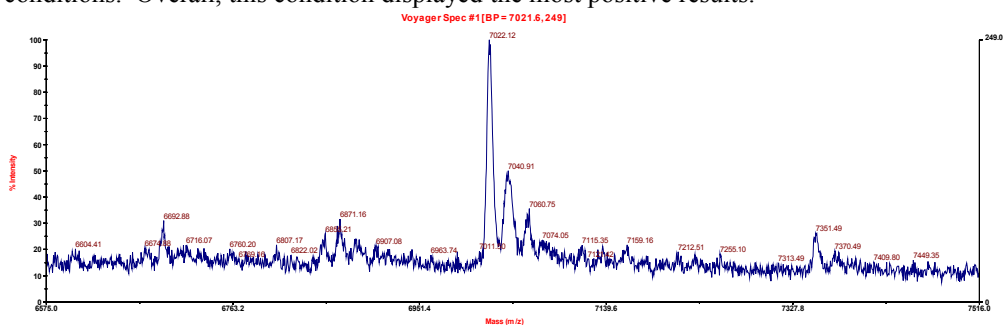
a.



b.



**Figure 4:** Condition 7 shows distinct peaks, more intense peaks, and reduced noise than in the control conditions. Overall, this condition displayed the most positive results.



## Related Works

Our research dealt with the optimizing of sequencing nucleotides and the detection of the sequences using MALDI-TOF by changing various variables of the DNA solutions. However, many other scientists have endeavored to optimize the sequencing process using a variety of methods. Our mentor Professor Kim (2003) increased the clarity of data from the MALDI-TOF by annealing large organic dyes onto each nucleotide increasing the mass difference between nucleotides, therefore allowing a clearer read from the MALDI-TOF. In our work, we did not use biotinylated nucleotides, which have large organic dyes annealed to them as Professor Kim had; however, we attempted to increase the length of the nucleotide sequence which, in conjunction with biotin, will allow greater data to be acquired.

Another researcher that increased the detection of nucleotides using the MALDI-TOF was Professor Hung (1998). Professor Hung did not test changes of the standard constitution of suspended DNA, but instead tested the effects of using different materials as a reading template for the MALDI-TOF. Hung found that using a hydrophobic

material such as paraffin wax instead of a standard steel template the amount of read nucleotides increased.

## Conclusions

In our research we found that decreasing buffer concentration, increasing annealing temperature, and increasing of dNTP concentration allowed us to obtain MALDI-TOF data with greater peak intensity. However, scope of our research was not large enough to provide truly conclusive data. Instead, our data could be used as a base for further research which could range the variables we used to obtain more definitive results. Combining the results of this research with the techniques explored by Professor Kim and Professor Hung would significantly improve the utility of MALDI-TOF MS in gene sequencing.

## Acknowledgements

We would like to acknowledge Rutgers University, Rutgers School of Engineering, the Motorola Foundation, Morgan Stanley, PSEG, Silver Line Building Products, families of 2001-2008 alumni, NJ Governor's School of Engineering and Technology and NJ Governor's School Board of Overseers for funding and supporting our research.

In particular, Professor Sobin Kim, Dan Duffield, Diana Arellano were incredibly generous with their

### References

1. "Entrez Gene: TP53 tumor protein p53 [Homo sapiens]." National Center for Biotechnology Information. 16 July 2009. 19 July 2009 <<http://www.ncbi.nlm.nih.gov/>>.
2. Kim, Sobin, et al. "Digital genotyping using molecular affinity and mass spectrometry." Nature Reviews 4 (Dec. 2003): 1001-1010.
3. Ruparel, Hameer, et al. "Digital Detection of Genetic Mutations Using SPC-Sequencing." Genome Research 14 (2004): 296-300. 18 July 2009 <<http://www.genome.org/cgi/doi/10.1101/gr.1344104>>.
4. The TP53 Web Site. 18 July 2009 <<http://p53.free.fr/index.html>>.
5. Watson, James D., and Andrew Berry. DNA: The Secret of Life. N.p.: Knopf, 2003.
6. Young, Sandra. MALDI-TOF Mass Spectrometry. 18 July 2009 <<http://www.psrc.usm.edu/mauritz/maldi.html>>.