Genetic Sequencing Lab Final Report: SNP Detection

Jonathan Ward

Introduction and Background

The purpose of this experiment was to investigate various techniques to detect single nucleotide polymorphisms (SNPs). For this, we worked on detecting SNPs in the SOR1 A gene, which are associated with Alzheimer's disease.

One way to detect a SNP is to design a primer that anneals to a section of single stranded DNA right next to the site of a possible SNP. After annealing, a polymerase chain reaction is performed to incorporate a single base onto the primer, hence the name single base extension (SBE). After denaturing, this primer with the extra nucleotide can be analyzed using various techniques to determine the identity of the unknown base extension.

One can increase the efficiency of SNP detection by using a multiplex single base extension (MSBE) process. This involves annealing multiple primers onto the target DNA and performing single base extension on each of them at all at once. This allows for the detection of multiple SNPs, but complicates the experimental procedure.

One technique for identifying the unknown base is MALDI-TOF mass spectrometry. MALDI-TOF stands for matrix assisted laser desorption/ionization time of flight mass spectrometry. In operation, a laser is used to ionize the sample and a high intensity electric field gradient propels the ions to an electrode sensor, which can precisely measure the ion current with respect to time. It is called time of flight analysis because having an accurate measure of time is critical to sort the sample's components by mass. According to Newton's law, the acceleration that a particle undergoes when subjected to an outside force is proportional to its mass. Assuming the charge on all the ions is the same, they will each experience the same force in an electric field. More massive ions will accelerate more slowly towards the sensor electrode, causing them to take longer to reach it. The ionic current peak associated with their motion will occur later than for

lighter ions. If the force on the ions and their time of arrival are precisely known, their mass can be accurately determined.

To detect SNPs, MALDI-TOF mass spectrometry is used to measure the weight of SBE DNA strands. By comparing the SBE DNA molecule's mass to that of the primer, one can figure out which base was incorporated.

In order to achieve this accuracy, the testing samples need to be ultra-pure. One technique known as zip tip purification works by essentially filtering out unwanted molecules. Another technique called biotinylated solid phase capture purifies the sample by extracting only the wanted molecules, which have been tagged with a magnetic marker.

Materials and Methods

The SBE primers were synthesized using solid-phase synthesis with an Expedite 8909 Nucleic Acid Synthesizer. The Expedite 8909 cycled through all the synthesis reactions (coupling, detritylation, capping, oxidation) automatically while cleavage and deprotection was performed manually.

OPC purification was then completed to remove chemical impurities and uncompleted chains. Following purification, the concentration of the primer solution must be determined to be able to use it successfully in future reactions. To accomplish this, a spectrophotometer was used to measure the solution's absorbance of 260 nm light. A solution of DNA will have a unique absorption based on its concentration and composition. If the composition is known, then the concentration can be calculated using an equation (figure 1 in the appendix).

Touchdown PCR was used with the SOR1 A gene to reduce amplification of undesired regions of DNA. The process starts off with a higher than normal annealing temperatures and is gradually reduced for subsequent cycles until it reaches the desired ending annealing temperature. The higher temperature makes the amplification less efficient, but it is more selective. As the

annealing temperature is lowered for each cycle, the amplification efficiency increases, but selectivity decreases. Because the previous amplification cycles were more selective, the desired products are much more abundant than the undesired.

Following Touchdown PCR, the products were purified and checked with gel electrophoresis (figure 2). Because a band appears in the negative control, this PCR is probably contaminated. Four different SBE and MSBE procedures were then performed. The first three procedures used exon 8 as a template and the last one used exon 11, 25, and 34 from the SOR1 A gene. The sequences for these exons can be in figure 3 in the appendix. The primers used for SNP detection can be seen in figure 4. For the first procedure, an SBE reaction was performed with primer B and the products were precipitated and purified using zip tip purification. Using a zip tip pipette tip concentrates and purifies the sample, improving the mass spectrometer results. Any salt present in the sample will greatly alter the mass spectrum. Using a zip tip removes this salt, which yields more accurate results.

Like the first procedure, the second procedure used precipitation with zip tip purification, but the second procedure used MSBE with both primers A and B. The third reaction was also an MSBE reaction with primers A and B, but used biotinylated ddNTPs followed by solid-phase capture to extract only the SBE molecules. The final reaction, with the Sor1 A gene used MSBE with three primers followed by solid-phase capture of cleavable biotinylated ddNTPs.

After the various purification methods used (zip tipping, solid phase capture), the products were spotted onto a plate and analyzed using a MALDI-TOF mass spectrometer.

Results and Discussion

The 260 nm absorption of our primer solution was measured to be .549. Because the solution was diluted 20 to 1 we need to multiply by 20 to get the actual absorption. Using the formula in the appendix, the concentration was calculated to be 66.6 pmol/uL.

A MALDI-TOF mass spectroscopy measurement (figure 5) was taken of the primer sample and the mass was measured to be 4610 Da. This is within the normal error range of the calculated mass (4608 Da) of the primer.

For the SBE experiment, the mass spectrum (seen in figure 6) shows that the highest peak is the primer, even after all of the reaction steps. The next highest peak is the single base extension molecule. After subtracting a constant offset of 3 Da and comparing the mass to Table 1, it appears that the single base extension molecule was extended with ddATP. This is a little ambiguous, because the mass is close to other nucleotides as well. Better accuracy was achieved with biotinylated nucleotides.

The mass spectrum for the MSBE experiment performed with solid phase capture can be seen in figure 7. This is a very clean spectrum. No traces of of unextended primer can be detected, because of the solid phase capture. When comparing the mass peaks in this plot with the expected masses tabulated in Table 3, it appears that primer A was extended with Biotin-ddUTP while primer B was extended with Biotin-ddATP. The measured masses fit very closely to the calculated masses.

The mass spectrum for MSBE with biotinylated solid phase capture with a cleavable linker can be seen in figure 8. This plot is a little more baffling than the previous ones. There are only two peaks when there should be three and the masses of the peaks do not correspond to the calculated masses of extended primers in Table 4. There could be a large mass offset from salt contamination that is skewing the results. Also, one of the smaller peaks could be the SBE from the third primer, but it is hard to tell with the mass offset and the existence of many sub peaks of equal intensity.

Conclusion

Single Base Extension reactions with solid phase capture work very well to identify SNPs. The solid phase capture is a key step to improving the accuracy of this technique. Using just zip tip

purification of the reaction products does not produce very accurate results. This is because the possible mass error sources from technique, contamination, etc. can add up to be greater than the mass difference between ddNTPs. Solid phase capture also works really well to separate out the reaction products, selecting only the fully extended primers and leaving unwanted products behind. The extra mass of the biotinylated bases also helps improve their discernibility. Using a cleavable linker for the biotinylated bases streamlines the post-processing required after the SBE reactions

The disadvantage of all forms of SBE is that it is a screening procedure only. To use it to detect SNPs, we must know exactly where one is likely to occur. Because SBE essentially sequences only one unknown base pair, we must know a lot about the surrounding sequence before use.

Other ways to detect SNPs, such as direct sequencing, do not have this limitation. With direct sequencing, we can scan for SNPs that don't know about yet and also detect many SNPs at once. With SBE, we can only detect as many SNPs as we design primers for. The downside to sequencing is that it is very expensive.

A Sanger sequencing plot can be seen in the appendix (figure 9). By examining the sequences, we can see that there are SNPs at location 108 and 112 in the middle sequence. By looking at the upper and lower sequence, we can see that the SNP at location 108 is an A-G polymorphism and the SNP at location 110 is a G-C polymorphism.

In addition to direct sequencing and SBE, allele specific oligonucleotide hybridization (ASO) can also be used to detect SNPs. In ASO, a set of labeled primers are designed that surround the SNP site. By carefully controlling the annealing temperature, the labeled primers will bind only to their specific alleles. SNPs can then be detected using any of the various primer labeling techniques.

Appendix

Figure 1: Equation for determining primer concentration

[DNA] =A260x100/[1.54(nA) + 0.75(nC) +1.17(nG) + 0.92(nT)]

where

[DNA] = DNA concentration in pmol/µL;

A260 = absorbance at 260 nm in UV/vis spectrum;

nA =number of nucleotide A in DNA sequence;

nC =number of nucleotide C in DNA sequence;

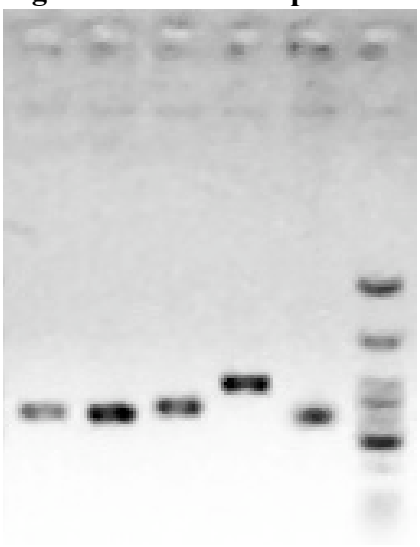
nG =number of nucleotide G in DNA sequence;

nT =number of nucleotide T in DNA sequence;

Sample used for concentration analysis:

5'TCTAGCAGTGCTGGA3' (MW: 4608)

Figure 2: Gel Electrophoresis of Touchdown PCR products



From left to right: water w/ exon 11 primer (negative control), genomic DNA with exon 11 primer (positive control), amplified exon 34, amplified exon 25, amplified exon 11, marker

Figure 3: Exons for SNP Analysis

Exon 8

ACGGAGAACGAAGAGAAAAGGATAGGACTCATCACCATTAGATGACCCTG

CCTTGTCGAACTCCACGCACAAACACGGACAGGACCCTCTCTGGCCGCGT

GTCTCCTTC

Exon 11

gaaagcattcttagttgctagatactacgcagaatttctggctgggcaag
gtgattatctgacatattcttgaaattaaaaataattatttctcttgcat
tttagGCTCAGTGGGAAAGAACTTGGCTAGCAAGACAAACGTGTACATCT
CTAGCAGTGCTGGAGCCAGGTGGCGAGAGgtcagccccctcccccaatcc
cgtcccctccaccctcattcccatttgtgtgagaatgtagactgtgcctt
ggcatttcagggacacaaatggtccatggcagagaccagtaatgccagat
atgactataaagtaaagaaaggtacctgaagccactgatttatatccaca

Exon 25

tctctgtggatcattggtgatcacgggtccatctccatcctttatgagag
ctttgacaccagagacaaaattctgaacaagcttttgtcctcacctctct
gtttatggtctcacctgcagAAATGCACCAGTGCCGGAGTGACGAGTACA
ACTGCAGTTCCGGCATGTGCATCCGCTCCTCCTGGGTATGTGACGGGGAC
AACGACTGCAGGGACTGGTCTGATGAAGCCAACTGTACCGgtcagtactt
cctggactcagttgacagcactcatccgttcatgcagtggttaacattat
agctttaaacgatcggaaaatctaggctctgagaattagcttggacctta

Exon 34

tttgcccgcctcggcctcccaaagtgctaggattacaggcgtgagccacc
acgcccagccaagatgtaactttgtttgtctttttctcctttatctctct
agATGAGTTGACTGTGTACAAAGTACAGAATCTTCAGTGGACAGCTGACT
`TCTCTGGGGATGTGACTTTGACCTGGATGAGGCCCAAAAAAATGCCCTCT
GCTTCTTGTGTATATAATGTCTACTACAGgtaggtcccatctttgtcatg
ggagatgaaaatgatgtcttcattgccccaggatgttctgtaaactctct
agtggcatgcaccatgctattaacggagtcagacatctcagccccccatg

Figure 4: Primers for SNP Detection

Primers for SBE and MSBE of exon 8.

Primer A: 5’-GATAGGACTCATCACCA-3’ (MW: 5163)

Primer B: 5’-TAGATGACCCTGCCTTGTCG-3’ (MW: 6084)

Primers for Cleavable MSBE with PCR products from SorlA gene exon 11, 25 and 34.

SorlA SBE primer 1: 5’TCTAGCAGTGCTGGA3’ (MW: 4608)

SorlA SBE primer 2: 5’TACAGTTGGCTTCATC_3’ (MW: 4847)

SorlA SBE primer 3: 5’ AAAAAAATGCCCTCTGC-3’ (MW: 5147)

Table 1: SBE

		Mass of Single Bass Extension Products (Da)			
Primer	Mass (Da)	ddATP	ddGTP	ddTTP	ddCTP
B	6084	6381	6397	6372	6457

Table 2: MSBE

		Mass of Single Bass Extension Products (Da)			
Primer	Mass (Da)	ddATP	ddGTP	ddTTP	ddCTP
A	5163	5460	5476	5451	5436
B	6084	6381	6397	6372	6457

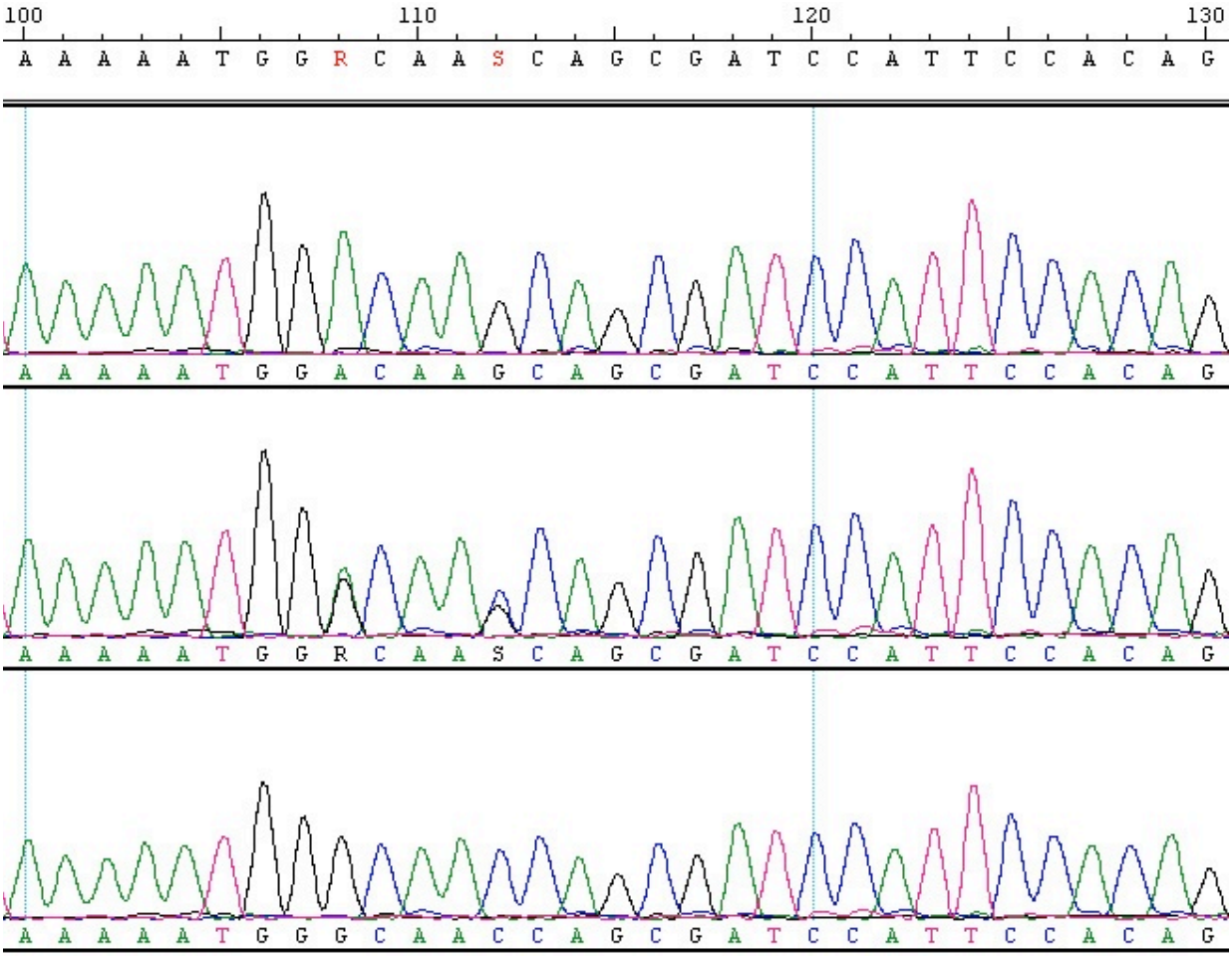
Table 3: MSBE with Biotin Solid-Phase Capture

		Mass of Single Bass Extension Products (Da)			
Primer	Mass (Da)	Biotin-ddATP	Biotin-ddGTP	Biotin-ddUTP	Biotin-ddCTP
A	5163	5851	5867	5917	5828
B	6084	6772	6788	6838	6749

Table 4: MSBE with Solid-Phase Capture with Cleavable Biotin Linker

		Mass of Single Bass Extension Products (Da)			
Primer	Mass (Da)	Biotin-N3-ddATP	Biotin-N3-ddGTP	Biotin-N3-ddUTP	Biotin-N3-ddCTP
1	4608	5233	5362	5324	5210
2	4847	5472	5601	5563	5449
3	5147	5772	5901	5863	5749

Figure 9: Sanger Sequencing Plot of SNPs



References

Facts about oligonucleotides. http://www.generi-biotech.com/oligo/pages/udaje_o_oligonukleotidech_e.htm. Accessed May 10, 2010.

Kim, S., Edward, J.R., Deng, L,. Chung, W., Ju, J. (2002) Solid phase capturable dideooxynucleotides for multiplex genotyping using mass spectometry. *Nucleic Acid Res.* **30.**

Kim, S., Hameer D., Ruparel, T., Gilliam, C., Ju, J., (2003) Digital genotyping using molecular affinity and mass spectrometry. *Nature.* **4.** 1001-1008

Acknowledgements

I would like to thank Prof. Jingyue Ju, Shundi Shi, Jim Russo, and everyone else who made CHEN E4760 possible.