The molecular genetics of colon cancer: hereditary non-polyposis colorectal cancer and the mutations concerning the MLH1 gene in relation to PMS2 binding as studied in 11 patients under laboratory conditions

Timothy J-P Ducey
BIOS 201 General Biology Lab-Section 6B
University of Notre Dame
26 November, 2001

Abstract

Analyzing the two major types of hereditary colon cancer, familial adenamotous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC), it has been shown that HNPCC, a syndrome characterized by autosomal dominance, inheritance, and the predisposition of early onset colorectal cancer, accounts for between 5% and 10% of colon cancer cases among Western populations. The mutation of the MLH1 gene often triggers the onset of HNPCC. During this study, patients were screened for HNPCC and MLH1 mutation using established criteria, the locations of the mutations were identified, and the mutations were analyzed to determine their effect on the function of the MLH1 protein. In this analysis, patients were screened for MLH1 mutations using PCR, cloning, and sequencing, and the proteins were examined to determine proper binding activity with the PMS2 protein. Of the 11 patients screened for an MLH1 mutation, 9 showed a mutation in the MLH1 gene. All 4 regions of the gene were affected over the course of the 11 patients, and the mutations were numerous, with deletion framshift, missense substitution, nonsense substitution, insertion frameshift, and silent substitutions present. Subjected to binding assay, six of the patients showed abnormal binding with the PMS2 binding region as a result of the MLH1 gene mutation, while 3 patients showed normal binding in spite of the MLH1 mutation. These data confirm studies that show a link between the MLH1 and PMS2 genes in relation to HNPCC, and further the scientific community to undertake research that will determine the exact link between MLH1 and PMS2 in DNA mismatch repair and help forge an eventual cure for HNPCC.

Introduction

Cancer is a group of diseases whose primary characteristics include the proliferation and growth of cells which lack normalcy (American Cancer Society 2001). Such abnormalities are most often the result of genetic mutations, and these mutations many times are the end product of environmental factors, namely chemical carcinogens, physical mutagens like X-rays, or specific viruses (Campbell et al. 1999). In general, as Campbell et al. (1999) discuss, the genetic mutations that cause proto-oncogenes, or normal cellular genes, to become oncogenes, or cancer-causing genes, can be divided into three categories: movement of DNA within the genome,

amplification of a proto-oncogene, and point mutation in a proto-oncogene. The need to increase the knowledge concerning cancer and its causes stems from the estimate that approximately 8.7 million people in the United States have a history of cancer, and the prediction that around 1.3 million new cancer cases would arise in 2001 in the U.S. alone (American Cancer Society 2001).

The mutations that cause changes in proto-oncogenes also stem from hereditary predispositions. While any body system can be the target of hereditary cancer-causing agents, the colon is very often affected. Estimates show that nearly one-half of all Americans will develop some form of a tumor in their colon during the course of their lifetime, and that one in ten persons will have colon tumors that develop a malignancy (American Cancer Society 2001). Of the various types of colon cancer, there are two major types of hereditary colon cancer: familial adenamotous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC) (Komaromy 2001). Accounting for five to ten percent of colon cancer cases among Western populations, HNPCC is autosomal dominant, inherited, and characterized by the predisposition of early onset colorectal cancer (Wang et al. 1997). In families in which colon cancer spans several generations, indicating a potential hereditary link, HNPCC is diagnosed as the cause if a certain set of criteria, called *Amsterdam criteria II*, is met (Wang et al. 1997). These criteria are the following: 1) three or more relatives with verified colorectal cancer (CRC), 2) one or more of these is a first degree relative, 3) one or more generations is affected with CRC, and 4) one or more of the affected relatives is less than 50 years old (Bocker et al. 1999). One must also note that HNPCC is not a cancer, rather it is a syndrome that increases a person's risk of acquiring colon cancer, and mutations in the genes that cause HNPCC will further increase a person's chances of contracting colon cancer (Komaromy 2001).

Of the HNPCC genes, one that frequently undergoes a mutation leading to the development of colon cancer is the mutL homolog 1 gene, also known as MLH1 (Li and Modrich 1995). A DNA mismatch repair gene, MLH1 is comprised of 19 exons, and its cDNA is 2.7 kb is length (Bocker et al. 1999). While it has been detected that four human genes, MSH2, MLH1, PMS1, and PMS2, have been known to undergo germline mutations correlated with HNPCC, the majority of the cases center around mutations to the MSH2 and MLH1 genes (Wang et al. 1997). Research has shown that HNPCC-afflicted individuals are heterozygous for a repair defect gene such as MLH1, and the inactivation of this gene, effectively a loss of heterozygosity, triggers the onset of colon cancer (Li and Modrich 1994).

During the course of my study, the following goals were carried out. First, using *Amsterdam criteria II*, suspected HNPCC patients were screened for MLH1 mutation. Second, the locations of mutations on the MLH1 gene were identified. Finally, it was determined whether the mutations were likely to have a significant effect on MLH1 protein function, and therefore potentially trigger the onset of colon cancer. To carry out these goals, patients were screened for MLH1 mutations using PCR, cloning, and sequencing, and the patients' MLH1 proteins were examined to determine proper binding activity.

Materials and Methods

Patient Selection-

To begin the study on the molecular genetics of colon cancer, selected patients were screened and identified as probable candidates for inclusion in the study. After viewing the case histories of 20 selected patients, the 11 most likely to meet the *Amsterdam criteria II*, or the clinical classification of HNPCC, were selected for further analysis. Patient 1, #7006, was 47 years of age, had a single malignant tumor of the colon, had replication error phenotype (REP)

present, and had a grandparent, parent, and sibling diagnosed with colon cancer. Patient 2, #6439, was aged 60, had a single malignant tumor of the colon, had REP present, and had a parent and sibling in which colon cancer was present, the sibling being diagnosed at age 47. Patient 3, #2164, was aged 40, had a single malignant tumor of the colon, had REP present, and had a grandparent and younger sibling with colon cancer and a parent with uterine cancer. Patient 4, #5878, was aged 55, had a single malignant tumor of the colon, had REP present, and had a parent, uncle, and sibling with colon cancer, the sibling having been diagnosed at age 45. Patient 5, #1869, was aged 42, had a single malignant tumor of the colon, had REP present, and had a parent, uncle, and sibling with colon cancer. Patient 6, #7035, was aged 32, had a single malignant tumor of the colon, had REP present, and had a grandfather and aunt with colon cancer. Patient 7, #8551, was 37 years of age, had a single malignant tumor of the colon, had REP present, and had a parent with ovarian cancer. Patient 8, #8184, was aged 47, had a single malignant tumor of the colon, had REP present, and had a parent and sibling with colon cancer. Patient 9, #2600, was aged 41, had a single malignant tumor of the colon, had REP present, and had a sibling with uterine cancer and an unknown parental history. Patient 10, #1396, was aged 52, had a single malignant tumor of the colon, had REP present, and had a grandfather and parent with colon cancer. Patient 11, #2386, was aged 27, had a single malignant tumor of the colon, had REP present, and had a grandparent with an unknown type of cancer and a sibling diagnosed with stomach cancer at age 35.

Polymerase Chain Reaction-

After patient selection, amplification of the MLH1 gene was performed using the polymerase chain reaction, or PCR. The MLH1 gene was broken into four separate regions, or primers, for amplification, and each region had a plus primer and a minus primer. For region 1,

the plus primer consisted of 21 base pairs, and its sequence 5' to 3' was ATCTAGACGTTTCCTTGGCTC. Region 1's minus primer consisted of 20 base pairs, and its sequence was TGAGGCATTGGGTAGTGTCC. The expected size of this region was 667 base pairs, from –34 on the plus primer to 633 on the minus primer. For region 2, the plus primer was 21 base pairs and its sequence was GCTGATGTTAGGACACTACCC. The minus primer was 18 base pairs in length and its sequence was TCCCCTTTGTTGTATCCC. The expected size of this region was 760 base pairs, from 604 on the plus primer to 1363 on the minus primer. For region three, the plus primer consisted of 24 base pairs, and its sequence was CAAAGGGGACTTCAGAAATGTCAG. The minus primer consisted of 23 base pairs, and its sequence was AAGGGGTAATCCAATTCAGGTTCC. The expected size of this region was 569 base pairs, from 1355 on the plus primer to 1923 on the minus primer. For region 4, the plus primer consisted of 18 base pairs, and its sequence was TTGATGAGGAAGGGAACC. The minus primer was 21 base pairs in length, and its sequence was AGAAGGAACACACATCCCACAGTG. The expected size of the region was 419 base pairs, from 1889 on the plus primer to 37 on the minus primer. Thirty cycles of PCR amplification were run in the thermocycler for the four regions of the MLH1 gene. The four regions were denatured at 94^{o}C for 1.5 minutes. Regions 1, 2, and 4 were annealed at 48 ^{o}C for 1.5 minutes, and region 3 was annealed at 57 ^{o}C for 1.5 minutes. Finally, all regions of the MLH1 gene were extended at 72 ^{o}C for 1.5 minutes.

Cloning of PCR products-

After amplification, the PCR products were cloned. 2 ul of PCR product was utilized in the ligation, and PCR 2.1 TOPO vector was the plasmid used in this process. In the course of the ligation, the TA-TOPO cloning kit (Invitrogen Corporation) was used. After cloning, 6 ul of

ligated plasmid were added to *E. coli* competent cells ensuring proper bacterial transformation. Following transformation, 100 ul of cells were plated onto an LB-ampicillin medium containing an additional layer of X-gal solution, and these plates were incubated.

Plasmid Isolation-

Upon completion of the transformation and incubation, the plasmid was isolated. 3 cultures of white colonies containing the recombinant plasmid and 1 culture of blue colony containing a recircularized plasmid without the PCR insert were grown up from each transformation using LB/Ampicillin broth. For the plasmid isolation, 5 ul of cells were used. The plasmid isolation was performed using the Wizard Miniprep Kit (Promega Corporation).

DNA Sequencing and Analysis-

Following the miniprep isolation of plasmid DNA from bacteria, a complete DNA sequence for each individual in the study was produced, and the sequences were analyzed to determine where the mutations occurred in the gene. In order to sequence the DNA, electropherogram results were produced by an ABI automated DNA sequencer. This method is a modification of the PCR reaction in that only one primer is used, and a portion of the nucleotides has been modified so that they are missing the 3'-OH group. After performing an initial sequence analysis of the first 90 nucleotides by hand, a computer-based nucleotide analysis was performed. Using the National Center for Biotechnology Information web site (www.ncbi.nlm.nih.gov), the patients' DNA sequences were checked against the normal MLH1 sequence. Following the nucleotide analysis, a computer-based protein analysis was performed using the Protein Translation Tool (http://tw.expasy.org/tools/dna.html). This tool allowed for the automatic display of amino acid sequences, making it very easy to see the possible effects of the MLH1 mutations.

MLH1 Protein Binding Assay-

After the DNA sequencing and analysis, a protein binding assay was performed to determine whether the protein produced by the MLH1 allele functioned normally. To carry out this operation, gel filtration chromatography was utilized. This technique could be utilized to detect binding because the MLH1 protein forms a dimer, or a structure in which two molecules are bound, in order to function properly. In the chromatography, molecules are run through a gel column, and if the molecules are larger than the exclusion limit of the column, then they will pass quickly through the column, while those molecules smaller than the exclusion limit pass slowly through the column. In the case of the study, if an MLH1 dimer was formed, the protein molecule would run quickly, an indication of normal binding. For the chromatography, 150 ul of protein sample was utilized, and the exclusion size of the resin was 120 kda, or 120,000 daltons. Following the chromatography, the protein assay was analyzed using a plate reader to form an elution profile, which indicated normal versus abnormal binding.

Results

The sizes of the DNA fragments amplified by PCR were determined using gel electrophoresis. As indicated in figure 1, the sizes of the fragments varied by region. For region 1, the size was approximately 700 base pairs. For region 2, the size was approximately 800 base pairs. For region 3, the size was approximately 500 base pairs, and for region 4, the size was approximately 400 base pairs. These approximations indicate that the correct DNA regions were amplified, as the exact size of region 1 was 667 base pairs, the exact size of region 2 was 760 base pairs, the exact size of region 3 was 569 base pairs, and the exact size of region 4 was 419 base pairs.

Among the 11 patients screened for MLH1 mutations, 9 showed an actual mutation of the MLH1 gene. Table 1 shows a complete analysis of the data for each patient. In every case, the patients in whom a mutation occurred were heterozygous for the mutation, and between the nine patients affected, all regions of the DNA segment showed a mutation. In three of the cases, region 1 was affected. Two cases showed a mutation in region 2. Two patients showed a mutation in region 3, and two patients showed a mutation in region 4. Analyzing the nine mutations, all forms of possible mutations were seen. Two patients had a deletion frameshift mutation, a mutation in which the number of nucleotides deleted is not a multiple of 3, causing an improper grouping of codons. Three patients had a missense substitution, a mutation in which a single amino acid is changed. Two patients had a nonsense substitution, resulting in a premature stop signal. One patient showed a frameshift insertion, a mutation in which the number of nucleotides inserted is not a multiple of 3, causing, like the frameshift deletion, an improper grouping of codons. Finally, one patient showed a silent substitution, a mutation in which there is no change in the amino acid sequence.

In carrying out a binding assay of MLH1 protein function with the PMS2 DNA repair protein, as the sixth column of table 1 shows, 6 of the patients showing a MLH1 mutation showed an abnormality in the formation the MLH1/PMS2 dimer, while the remaining three patients showing a MLH1 mutation had normal MLH1/PMS2 binding. After the filtration gel chromatography was performed for each patient, a binding assay elution profile was compiled to indicate normal or abnormal MLH1/PMS2 binding. As figure 2 shows, normal binding presents an elution curve closely resembling the control curve, very steep in its curvature leading to and from the apex of the curve. Abnormal binding, however, produces a curve in which the

curvature to and from the apex is very gradual, and the absorbance of light at 595 nm is much less than compared to the normal binding curve.

Discussion

In an effort to determine whether mutations on the MLH1 gene could potentially have an effect on MLH1 protein function and thereby possibly trigger the onset of colorectal cancer, 11 patients inflicted with colon cancer were tested to examine such a hypothesis. In theory, a mutation at the DNA level, on the gene, could have a significant effect on the reading and expression of the gene and subsequent protein. In the case of cancer, such a mutation could trigger the onset of a malignancy, which in turn could be fatal. In the case of my study, 9 of the 11 patients showed a mutation in the MLH1 gene. According to Guerrett et al. (1999), it has been shown that germline mutations in two human mismatch repair (MMR) genes, one of which is MLH1, are present in nearly 70% of HNPCC cases. Furthermore, it has been shown that there are significant protein interactions between the MLH1 and PMS2 human MutL homologs (Matton et al. 2000). Considering the results of the protein assay, it becomes very apparent that a relationship exists between MLH1 and PMS2. In 6 of the 9 cases of MLH1 mutation, the amino acid change as a result of the MLH1 mutation made it impossible for the MLH1 protein to form a dimer with the DNA mismatch repair protein PMS2. Therefore, it can be inferred that the absence of the formation of this dimer made DNA repair on the MLH1 gene impossible, and therefore, a mutation was caused which in turn triggered HNPCC colorectal cancer in these patients.

However, in 3 of the cases, PMS2 binding was not affected by the MLH1 mutation. Therefore, in these cases, conclusive evidence cannot be offered that the cancer was caused by a mutation in the MLH1 gene. In these cases, the onset of colorectal cancer could have been the

result of other factors such as the person's environment or the contraction of a virus. Furthermore, 2 of the patients in which there is normal MLH1/PMS2 binding show incomplete *Amsterdam II* criteria for HNPCC. Here, it is seen that even though a mutation in the MLH1 gene exists, a definitive clinical diagnosis of the colon cancer is nearly impossible as there is proper binding and no complete criteria indicating HNPCC. In this instance, because of the ambiguities surrounding the cancer, potential treatment plans may be more difficult for a doctor to formulate.

The relevance of this study becomes very apparent as it first of all solidifies the cause of colon cancer as HNPCC in six of the patients. In further confirming an already reported link between the MLH1 and PMS2 mismatch repair genes, it solidifies the findings of others. Yet, it also points out the strict clinical classification of HNPCC in the *Amsterdam II criteria*, and indicates that further research is needed to clarify and potentially amend these HNPCC criteria. In addition to forging a more in depth analysis of HNPCC criteria, the study also magnifies the relationship between MLH1 and PMS2 and the lack of knowledge surrounding this interaction. As Guerrette et al.(1999) indicate, the function of MLH1 and PMS2 as they relate to human mismatch repair is yet unknown. By forging ongoing research in this area, a greater knowledge of HNPCC will be acquired, and a cure for this cancer could, and hopefully will, be found.

Literature Cited

American Cancer Society. 2001. Cancer Facts and Figures 2001. American Cancer Society. Available: http://www.cancer.org/downloads/STT/F&F2001.pdf. [2001, November 5]

Bocker, T., Ruschoff, J., and Fishel, R. 1999. Molecular diagnostics of cancer predisposition: hereditary non-polyposis colorectal carcinoma and

mismatch repair defects. Biochimica et Biophysica Acta **1423**: 01-010.

Campbell, N.A., Reece, J.B., and Mitchell, L.G. 1999. Biology. 5th ed. Menlo Park: Addison Wesley Longman, Inc.

Guerrette, S., Acharya, S., and Fishel, R. 1999. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. Journal of Biological Chemistry **274**: 6336-6341.

Komaromy, M. 2000. What is HNPCC? Genetic Health. Available: http://www.genetichealth.com/CRC_HNPCC_A_Hereditary_Syndrome.shtml. [2001, November 5]

Li, G.M., and Modrich, P. 1995. Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. Proc. Natl. Acad. Sci. USA **92**: 1950-1954.

Matton, N., Simonetti, J., Williams, K. 2000. Identification of mismatch repair protein complexes in HeLa nuclear extracts and their interaction with heteroduplex DNA. Journal of Biological Chemistry **275**: 17808-17813.

Wang, Q., Desseigne, F., Lasset, C., Saurin, J.C., Navarro, C., Yagci, T., Keser, I., Bagci, H., Luleci, G., Chayvialle, J.A., Puisieux, A., and Ozturk, M. 1997. Germline hMSH2 and hMLH1 gene mutations in incomplete HNPCC families. International Journal of Cancer **73**: 831-836.

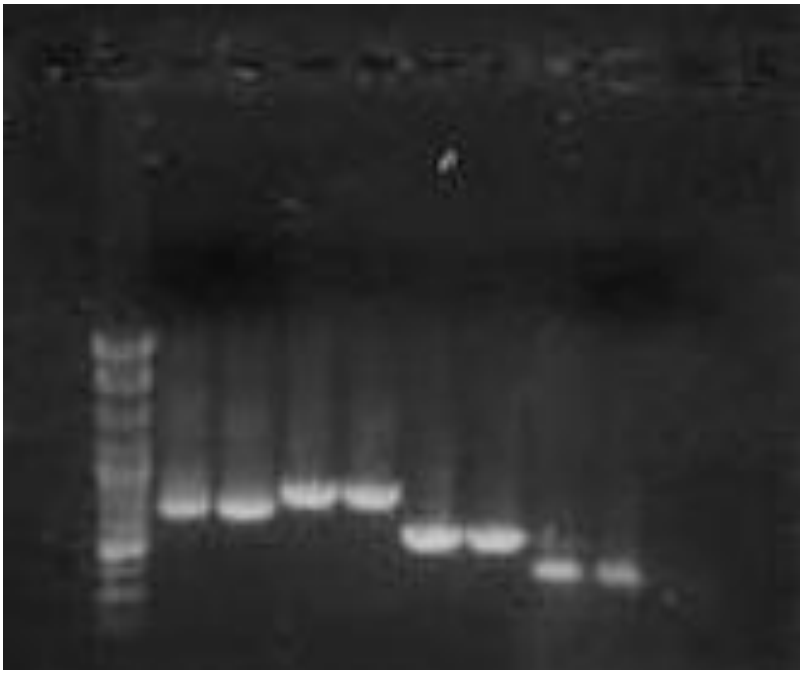

Figure 1. Gel electrophoresis showing the size of DNA fragments amplified by PCR. The sizes of DNA standard bands are in the leftmost lane, with the top band indicating the largest and the bottom band indicating the smallest. The topmost band is at 2680 base pairs, and the lowest band is at 100 base pairs. The bright band is at 500 base pairs. The amplified regions are loaded in order from left to right. The 2 leftmost bands show region 1, the third and fourth bands show region 2, the fifth and sixth bands show region 3, and the seventh and eighth bands show region 4. The approximate sizes of the amplified DNA regions are as follows: Region 1- 700 bp, Region 2- 800 bp, Region 3- 500 bp, Region 4- 400bp. These gel approximations confirm the exact sizes of the regions which are as follows: Region 1- 667 bp, Region 2- 760 bp, Region 3- 569 bp, Region 4- 419 bp.

Table 1. Summary of sequencing results for nine patients in which mutations were detected. Upon completion of mimiprep plasmid isolation, DNA for the eleven test cases was sequenced using an ABI automated sequencer. Computer analysis was used to analyze the full region, and mutations were detected in nine of the eleven patients. Using this analysis, the region of the mutation, the codon of the mutation, the nucleotide change involved, and the consequence of the mutation were determined. A represents the nitrogenous base adenine. C represents the nitrogenous base cytosine. G represents the nitrogenous base guanine. T represents the nitrogenous base thymine. The codon on which the mutation occurred was determined by dividing the nucleotide number by three. The column labeled "binding activity" indicates the result of the gel filtration chromatography which was used to assay protein function. If the MLH1 protein formed a dimer with PMS2, creating a dimer of 180.4 kda, then binding was normal, and the protein moved through the gel column, which had an exclusion limit of 120 kda, quickly. If a dimer was not formed, binding was abnormal, and the proteins would move through the gel column slowly as their volume was below the exclusion limit.

Patient ID	Region	Codon	Nucleotide Change	Consequence	Binding Activity
7006	1	25	Deletion of A at 73	Frameshift	Abnormal
6439	3	582	C to G at 1744	Leu to Val	Normal
2164	1	128	G to C at 382	Ala to Pro	Abnormal
5878	4	659	C to T at 1975	Arg to Stop	Abnormal
7035	2	342	Insert of G at 1026	Frameshift	Abnormal
8184	3	616	Deletion of AAG in 1846 to 1848	Deletion of Lys	Abnormal
2600	2	269	C to G at 806	Ser to Stop	Abnormal
1396	1	67	G to A at 199	Gly to Arg	Normal
2386	4	653	G to T at 1959	Leu to Lwu	Normal

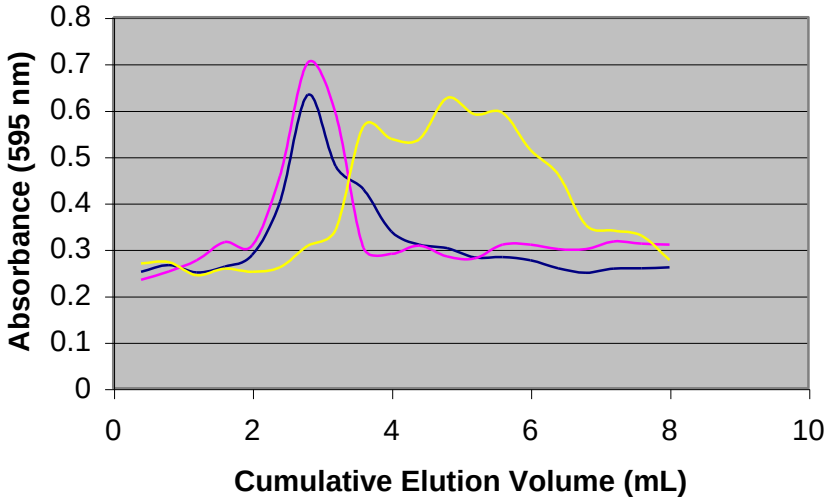

Figure 2. Absorbance (595nm) versus cumulative elution volume. Measurements of elution volume for the 9 patients showing an MLH1 mutation were taken using a plate reader to determine normal versus abnormal MLH1/PMS2 binding activity. This graph shows the normal control, indicated in blue, and the graphs for 2 example patients. The pink curve, which closely resembles the control curve, shows a patient with normal binding. Contrasting the normal binding, the yellow curve indicates abnormal MLH1/PMS2 binding. The curve for the abnormal binding has a greater cumulative elution volume because of the relative slowness of elution for the abnormal binding compared to the normal binding. Because more buffer was added with time as the two unbound proteins, MLH1 and PMS2, made their way through the gel column, and the proteins did not proceed through the gel quickly due to their volume being below the exclusion limit of 120 kda, a greater elution volume was produced compared to the proteins which demonstrated normal binding and moved quickly.