GENE REGULATION AND OVERLOOKED CODING REGIONS: THE CASE FOR TRANSLATION PAST THE 3'-UGA CODON IN CD4

by

BRADY NEAL WARE

(Under the Direction of E. Will Taylor)

ABSTRACT

Purpose. Understanding the mechanism involved in translating the UGA codon as selenocysteine instead of "STOP" is important for understanding the human genome. Selenocysteine provides a biological basis for the growing evidence that selenium is an important micronutrient. Low serum levels of selenium is linked to development and poor prognosis of diseases ranging from AIDS to cancer. A stem loop structure (SECIS element) in the 3' untranslated region of the mRNA is necessary for selenocysteine incorporation. There are variants of the SECIS, and proteins that interact with the SECIS. The process of selenocysteine insertion may be highly regulated and is complex. Proteins may be expressed in multiple forms and some presumed to end at a UGA codon may actually contain selenocysteine. CD4 was chosen as a model protein to study these possibilities and is assumed to end at a single UGA codon. **Methods.** Structural studies on the gene sequence are carried out using programs including BLAST, Scanps, ClustalW, NetPhos, and NetOGlyc. Peptides corresponding to the region past the UGA codon and the cytoplasmic tail of CD4 are prepared. Antibodies used for labeling experiments are prepared against

these peptides. Microscopy studies are done using a laser confocal microscope and a TEM to locate the antibodies within H9 T cells. **Results.** The region past the UGA has plausible structure including a consensus sequence for protein kinase PKA. Antibodies against the cytoplasmic tail of CD4, and the putative region past the UGA codon were prepared. A suitable method for preparing the H9 T cells for antibody labeling was developed. Antibody labeling microscopy studies are underway. **Conclusion**. Structural studies on the sequence of the CD4 gene support the possibility that translation continues beyond the UGA codon. Satisfactory methods for antibody labeling of the CD4 receptor using both laser confocal and immunogold electron microscopy was developed.

INDEX WORDS: UGA, Codon, CD4, Antibody Labeling, Termination, SECIS, Selenocysteine

GENE REGULATION AND OVERLOOKED CODING REGIONS: THE CASE FOR TRANSLATION PAST THE 3'-UGA CODON IN CD4

by

BRADY NEAL WARE

B. S. Biochemistry, Virginia Polytechnic Institute and State University, 1997

B. S. Biology, Virginia Polytechnic Institute and State University, 1997

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2003

GENE REGULATION AND OVERLOOKED CODING REGIONS: THE CASE FOR TRANSLATION PAST THE 3'-UGA CODON IN CD4

by

BRADY NEAL WARE

Major Professor: E. Will Taylor

Committee: Bi-Cheng Wang
Mark Farmer
Diane Hartle
J. Warren Beach

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2003

DEDICATION

I would like to dedicate this dissertation to the two people I most admire, my wife, Emily, and my dad, Connel.

"It is not the critic who counts, nor the man who points out how the strong man stumbled or where the doer of deeds could have done them better. The credit belongs to the man who is actually in the arena; whose face is marred by dust and sweat and blood; who strives valiantly; who errs and comes up short again and again; who knows the great enthusiasms, the devotions, and spends himself in a worthy cause; who at the best knows in the end the triumph of high achievement; and who at the worst, if he fails, at least fails while daring greatly so that his place shall never be with those cold and timid souls who know neither defeat nor victory."

- Theodore Roosevelt

"… I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I-
I took the one less traveled by,
And that has made all the difference."

Robert Frost, "The road not taken"

ACKNOWLEDGEMENTS

I would like to thank those who helped me along the way. Thanks to my family for being supportive. Thanks to my wife Emily for being patient and supportive and tolerating my unconventional style. Thanks to my lab mates for always providing lively conversation. Thanks to God for the wonderful creation we all study. Thanks to the UGA Monoclonal Antibody Facility who supported this research by helping with the preparation of monoclonal antibodies, especially Ruth Davis. In addition, I'm indebted to the six mice, especially mouse #3-025. This project would not have gotten off the ground without their sacrifice. Also, thanks to my committee and especially my major professor, Dr. Taylor, for their support, positive attitude, and willingness to allow me to pursue science without regard to the constraints of conventional thinking. Finally, I would like to thank the American Foundation for Pharmaceutical Education (AFPE) for supporting me with a fellowship for two years while I conducted this research.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Data on the genetic sequences of many organisms are accumulating at a rapid rate. In order to make sense of this data, and use it productively, a thorough understanding of the structure of genes and regulatory functions related to gene expression is necessary. Genes are dispersed within the DNA of prokaryotic and eukaryotic organisms and are delineated by start codons and stop codons. Eukaryotic genes begin with the methionine (met) codon, AUG, and end with one of three stop codons, UGA, UAA, or UAG.[1-3] Of these three, UGA is unique. For in addition to functioning as a stop codon, UGA is also the codon for the twenty-first amino acid selenocysteine.[4] The mechanisms involved in regulating the function of the UGA codon are complex, and only partly understood. In order to make use of the large amounts of genetic data that are now available it is vital to understand the function of the UGA codon.

Eukaryotic genes are broken up into exons separated by introns that do not encode protein. Most of the eukaryotic genome is thought to be composed of non-coding DNA.[5] Only a few percent of the total DNA encodes proteins or is thought to have regulatory functions. Typically, searches for genes look for start codons to delineate the beginning of genes, and stop codons to mark the end of genes. Often very small open reading frames are ignored. In most studies, the assumption is made that a UGA codon always results in translation termination. However, UGA does not always encode termination. The multifunctional nature

of the UGA codon greatly complicates the search for functional genes. First, reading frames that are ignored because they are thought to be small that end in a UGA codon may be functional if the UGA does not result in termination. Second, large pieces of functional genes may be ignored because they are encoded past a read through UGA codon. The result is that some potential genes are ignored and other genes are truncated by the assumption that they end at a UGA codon.

Those studying the protein encoded by the gene amplify this mistake. Recombinant protein is often used to study the function of the native protein in the living cell. Most proteins are simply not abundant enough to be thoroughly studied directly using current methods. In order to prepare the recombinant protein using an organism such as bacteria, the gene must first be copied out of the original cell, placed into a vector, and transfected into the bacteria where it can be transcribed. When the gene is copied, it is usually truncated near the UGA codon. All future studies done on a protein such as this may be tainted by this error. The potential function or functions of the part left out will not contribute to the study. Once the structure of the protein is determined using the recombinant gene, it too will be missing the piece past the UGA codon, and valuable structural and functional data will be missing. Potential drug targets could be missed this way, in addition to the useful knowledge lost. Hundreds of thousands of dollars could be saved by including the missing piece as well because research aimed at discovering how a protein carries out it's known function could come up empty handed if the missing piece is involved, and the

money spent on the research wasted. Then other groups searching for the basis of the same function could fail to find it, and then more money could be wasted. The alternative to this is that if the entire gene is included, the basis of the function of the protein could be found more quickly, and other groups would not have to go looking for it, thus saving money. Even if the entire gene sequence is included in the piece of recombinant DNA expressed in bacteria, not all of the necessary machinery may be present in the bacteria for efficient read through of the UGA stop codon.[6, 7] The mechanisms for read through or recoding of the UGA codon are complex, and differ substantially between eukaryotes and prokaryotes.[8, 9]

Studying the UGA codon is fundamentally important for another reason. The mechanisms controlling the function of UGA are very complex, and the biological functions associated with the differing functions of UGA are likely to be important. Any trouble in this system could potentially lead to or contribute to disease. Many diseases whose exact cause or pathophysiologies are not fully understood may be related to an aberration in the mechanism of UGA translation.

A brief review of the UGA codon provides an appreciation of the complexity involved in understanding the translation of genes ending in UGA. Translation normally proceeds until one of the three stop codons is encountered by the ribosome.[10] The three stop codons are UAA, UAG, and UGA. Once a stop codon is encountered, translation is usually ended by one of two processes. The simplest of these results from the absence of a termination factor binding to

the stop codon, which allows protein synthesis to continue. The ribosome can then dissociate from the mRNA, ending translation. The second way translation is ended involves a specific protein called a release factor that binds to the stop codon and facilitates the release of the growing polypeptide from the ribosome.[11] Release factors are covalently attached to GTP. When they bind to the ribosome and one of the stop codons, the GTP is hydrolyzed to GDP, the carboxyl group of the nascent polypeptide is transferred to water, and the polypeptide dissociates from the ribosome.

Although termination usually occurs when one of the three stop codons is encountered, termination may be suppressed. Mutations in the genes encoding tRNAs sometimes results in the production of functional tRNAs in which the anticodon is complementary to one of the three stop codons. Often, these mutant tRNA's are still charged with the appropriate amino acid by aminoacyl transferase.[12, 13] These are called nonsense suppressor tRNA because they suppress the nonsense codons. The stop codons were originally called nonsense codons because they were the only codons found that did not seem to code for a specific amino acid. When the mutant charged tRNA binds to the stop codon during translation, the aminoacid is transferred to the growing polypeptide chain.[13] Translation thus continues past the stop codon. In E.coli, there are two known nonsense suppressor tRNAs that bind to and facilitate translation past UGA codons.[12, 14] The nonsense suppressor tRNAs in eukaryotes are less well characterized. Both of the nonsense suppressor tRNAs in E.coli. that bind to and suppress the UGA codon are charged with tryptophan. Thus, a tryptophan is

added to the growing polypeptide instead of ending translation. Interestingly, one of the two UGA suppressor tRNAs has a wild type anticodon, and a mutation in the D stem. The suppressor activity is thought to arise from the formation of an A-U pair as opposed to a weaker G-U pair in the wild type D stem. This is thought to induce a conformational change allowing the anticodon CCA to form a pair with UGA as well as its native UGG codon. However, because the mutant tRNAs are present in such low numbers compared to the wild type tRNA, termination suppression by this mechanism is a rare event. The nonsense suppressor tRNAs must compete with release factors for the binding site on the ribosome and the RNA strand. Termination suppression of UGA codons may also result from mutations in the ribosome.[15-17] There is also some evidence for regulatory control of nonsense suppression.[13]

In one example in E.coli, synthesis of protein from RNA mutated to contain a UAG in place of a UGA stop codon that is read through by termination suppression, does not exceed fifty percent of the synthesis rate of the native protein.[18] In contrast, protein synthesis from RNA mutated to contain a UAA stop codon is reduced to 15%. UAA is the most common termination codon, and perhaps release factor has a higher affinity for it than the other two termination codons. Therefore, UGA and UAG are “leaky” stop codons when compared to UAA.

Although these mechanisms may occur without regard to the context of the stop codon, there is a basis for specificity.[19] Recent work investigating the relative efficiency of read through has found a correlation between the context of

the termination codon and the relative read through efficiency[20]. The amino acids encoded by the codons just 5' of the termination codon may also have an influence on read through efficiency [21-23]. In S. cerevisiae, basic amino acids facilitate more read through [24], while in E. coli, acidic and hydrophobic amino acids increase read-through [22]. The secondary structural characteristics of the amino acid can also have an important influence on read-through efficiency. One study reports that the read-through is more efficient if the terminal amino acid is likely to be associated with an alpha helix or beta sheet. Release factors recognize the termination codon as well as the adjacent nucleotides [21, 25]. Some adjacent nucleotides may be bound more favorably by release factors than other nucleotides. This provides a basis for the importance of the immediate context of the termination codon in read through efficiency through the mechanisms involving termination suppression. By weakening the binding of release factor to the termination codon, the equilibrium is shifted toward termination suppression. This hypothesis has been experimentally verified.[26] Harrell et.al. investigated the context of 91 viral sequences that were identified as containing a termination codon that was read through. In this study, 90% were found to have only one of six codons just downstream of the termination codon that was read through. Of the sixty-four possible codons, that one of only six are present in these read through viral sequences is statistically significant. Those six codons are CAA, CGG, GGG, GGA, GUA, and CUA. Although identified and studied as whole codons for convenience, the effects influencing read through are likely exerted on a nucleotide level. In addition, five codons in the 2+ position

past the termination codon were found in 70.3% of the read through sequences. For all but one (CUA) of the six+1 position codons, there was one of the five +2 codons most frequently associated with it. These pairs of codons, along with the percentage association of the +2 codon with the +1 codon are: CAA-UAA (88.0%), CGG-UUU (55.0%), GGG-UGC (52.9%), GGA-GGC (66.3%), and GUA-GAC (80.0%).[26]

In addition to finding codons associated with read through by sequence analysis, Harrell et. al. also studied the relative efficiency of read through of these sequences using a luciferase reporter gene construct. The most efficient read through construct contained the sequence UAG-CAA-UUA. This sequence is reported to facilitate an average read through of 3.2%. The read through of several of the native sequences increased when the termination codon was mutated to be UGA. No consistent folding structures were identified in the sequences that appeared to facilitate read-through.[26]

The aforementioned mechanisms are not unique to UGA. They occur for all three-stop codons, and can be nonspecific. The last mechanism of termination suppression that is mentioned here is unique to UGA and is very specific. As mentioned previously, in addition to its function as a termination codon, UGA is also the codon for selenocysteine, the twenty-first amino acid.

At this point, a brief review of selenobiology will be helpful. Although selenium was recognized many years ago to be an essential nutrient [27], the biological basis for its function remained unknown. When selenocysteine was discovered in glutathione peroxidase, the important role for selenium in biology

began to come into the light.[28] Further investigation on the mechanism by which selenocysteine became incorporated into glutathione peroxidase ultimately revealed that the selenocysteine occurred in the sequence at the place corresponding with a UGA codon in the RNA. This result was striking for a couple of reasons. First, the termination codon UGA appeared to also encode selenocysteine. Second, this observation strongly suggested that selenocysteine was inserted into the growing polypeptide by a co-translational mechanism, and not through a post-translational modification. These assertions were proven correct by further experimentation. Once it became clear that UGA was the codon that encoded selenocysteine, work began to discover the mechanism by which it's function was changed between termination to selenocysteine. Most of the early studies were done on prokaryotic selenoproteins. The context of the UGA codon was proven to be critical for the UGA to encode selenocysteine.[29] The critical area in prokaryotes is found immediately 3' after the UGA codon. Analysis of the sequence in the critical area by mutational analysis, computer folding studies, and nuclease protection assays revealed a folded hairpin structure that is required for selenocysteine insertion. This sequence is called a selenocysteine insertion sequence element (SECIS element). After the identification of a SECIS in bacteria, the eukaryotic and archaic SECIS was found much farther 3' of the UGA codon in the untranslated region.[29-31] However, the SECIS may be within 51-111 bases of the UGA codon.[29] Selenocysteine became recognized as the twenty-first amino acid, but the potential role for UGA codon regulation has not received the attention it

deserves. This is particularly true in the field of bioinformatics and sequence analysis. Until the process of selenocysteine insertion is completely understood and it's principles applied to the study of sequence analysis, genes will continue to be misidentified and overlooked.

Selenium is an essential nutrient whose deficiency can lead to or contribute to many diseases and supplementation of selenium intake may conversely prevent certain diseases.[32-39] Particularly interesting is the link between selenium status and viral disease state.[40-46] Consulting the literature will produce many more references supporting these facts. This topic is reviewed more thoroughly in Appendix 1. In summary, the only logical conclusion from analyzing much of the data seems to be that viruses encode their own selenoproteins. However, the only virus yet proven to encode a selenoprotein is Molluscum Contagiosum although there is evidence that there may be many more.[47] Conventional SECIS elements are missing from the other viruses and no one has demonstrated conclusively that the UGA codons present are transcribed as selenocysteine. However, some unconventional SECIS structure may be present.[48, 49] Just as a lack of appreciation for the UGA codon may cause sections of genes to be overlooked, so too can an incomplete understanding of the mechanisms underlying UGA codon function cause viral selenoproteins to be missed.

To date, there are two basic forms of the eukaryotic SECIS element.[50, 51] The eukaryotic deiodinase SECIS element can function in selenocysteine insertion when it is placed between 111 and 2.7k bases past the UGA codon

though it's natural position is 1.2kb past the UGA codon.[29] Type I deiodinase, GPx, and selenoprotein P loop 2 SECIS elements are typical of Type I. Type 3 deiodinase, phospholipid hydroperoxide, plasma GPx, and selenoprotein P loop 1 SECIS elements are typical of Type II. An unusual four base pair sequence at the base of the stem is critical for SECIS element function.[52-54] The base pairs are A-N, U-A, G-G, and A-U. The non-Watson Crick base pairing may be the most critical feature of the SECIS element, as one study found that they comprise the minimum structure necessary to form a functional SECIS element.[55] However, Kollmus et.al. have found that the exact sequence of the paired nucleotides may be unimportant as long as the combination of length and thermodynamic stability of the loop structure is maintained.[30] There is surprising variation in the structure of functional prokaryotic SECIS elements as well, and a true functional consensus is difficult to identify.[56]

Further complexity emerged when the mechanism by which the SECIS functions was explored. As stated previously, in bacteria, the SECIS is found immediately past the UGA that is recoded as selenocysteine. In eukaryotes, the SECIS is normally around one thousand bases past the UGA that is recoded. Presumably, the eukaryotic SECIS must be brought closer to the UGA codon by some adapter protein in order to carry out it's function. In bacteria, a protein was identified that binds both to the SECIS element and the ribosome.[57] This protein is called selB. The search for the eukaryotic selB homologue has turned up at least two, and perhaps three proteins that may each possess some functional equivalent of the selB protein. There are at least two accessory

protein adapter molecules that are necessary for the SECIS element to function in eukaryotes. One of these binds to the SECIS and the sequence in close proximity to the UGA codon. The other binds to the SECIS and to the ribosome.

The biological bases for the dual function of the UGA codon are not entirely clear. Selenocysteine plays a critical role in physiology, and perturbations of selenium intake and metabolism can have disastrous consequences. Clinical data has accumulated that links selenium intake and physiological concentrations to the severity and prognosis of illnesses that range from viral infections to cancer. A more thorough review of the role of selenium in viral diseases is provided in Appendix I. Given the criticality of proper function, it seems like a significant burden for living organisms to maintain a complex system such as this to regulate the function of a single codon. Of the sixty-four possible codons, there are only twenty-one different amino acids encoded by them, and three stop codons. This analysis is of course, counting UGA twice. On the surface, it seems that there is plenty of potential to use a single codon exclusively for selenocysteine. In fact, it would seem advantageous to simply abandon UGA as a termination codon entirely. After all it is not a very good termination codon. It is very leaky. The biological retention of the complexity bacterial system is difficult to explain. However, that of eukaryotes overshadows the complexity of the bacterial system. The biological purpose of placing the SECIS so far from the UGA codon, and then using at least two adapter molecules to bring it close is not clear.

One consideration is that the genetic code is not as flexible as may be assumed by analyzing the numbers of unique codons versus the number of different amino acids and termination codons encoded by it. Wobble in the third base position of the codons severely limits the number of unique components the three base genetic code can encode. Usually, the first two bases only affect the entity encoded by the codon specifically, and then whether the third base is a pyrimidine or a purine base. In the case of valine, serine, proline, threonine, alanine, and glycine, only the first two bases in the triplet are used to encode the amino acid. There are only two cases in which the function of the codon is changed depending upon which purine base is in the third position. In the first case, AUA encodes isoleucine, whereas AUG encodes methionine. In the second case, UGG encodes tryptophan whereas UGA encodes termination. The redundancy of the genetic code as a result of wobble in the third position may present a significant obstacle to dedicating a specific codon to a unique function.

It may seem logical for the cell to simply dedicate UGA to selenocysteine. However, because UGG (the purine pair to UGA) is the only codon that encodes tryptophan, and most mutant nonsense suppressor tRNAs that suppress termination at UGA codons are tryptophan tRNAs, there is probably significant pressure for the cell to dedicate UGA to encoding tryptophan.

These competing pressures may be enough to account for the current mechanisms involved in determining UGA codon function. However, there is one other important consideration. Is there some outright advantage for the cell in maintaining such a complex system of recoding UGA? By modulating the

function of a termination codon, the cell could make modular proteins (a fundamental idea of protein evolution). The cell could produce proteins that, under certain conditions, were terminated at the UGA codon. When conditions changed, the cell could produce the same protein with an extra domain or tail attached to the C-terminus.

A termination codon is uniquely suited for this sort of functional switching, because by definition, one of the functions should be termination. In addition, UGA is specifically well suited for this application because, as mentioned previously, UGA is a leaky termination codon. By being a leaky stop codon to begin with, the cell does not incur as much risk by modulating its function as would be incurred by modulating the function of less leaky stop codons. Because it is leaky, the cell must already contend with the effects of read through of the UGA codon. This is in contrast to, for example, the UAA codon that is a very effective termination codon. If tampered with, the cell risks losing a very effective stop codon, and may have to contend with unintended read through of other genes. The unintended read through could have disastrous consequences.

In addition, genes that end in a UGA codon are most suitable for adaptation as modular genes. Because of the already leaky termination in these genes, there is always some read through of the UGA and translation of the gene past the functional domains. Obviously, it is very disadvantageous for this additional transcript to be toxic or have some other deleterious effect to the cell. Perhaps equally, it is advantageous for this extra transcript to have some positive

effect for the cell. The extra transcript is simply junk or worse, until it is adapted to have some positive function.

Modular genes could be adapted to have many signal and regulatory functions for the cell. Modular gene structure provides another layer of translation regulation and controlling the proteome makeup of the cell. At least one study of a protein ending in a UGA codon has turned up different functions for the various read through products of the gene.[58] Although these read through products do not appear to be selenoproteins, the various functions associated with the different read through products suggests that the cell has adapted what may have been junk protein to have a positive function for the cell.

Given the possibility for specific regulation of the UGA codon and the production of modular genes outlined above, and the scant attention this possibility has received, there may be genes that have been assumed to end in a UGA codon that are actually translated past this region under certain conditions. Potentially, many modular genes exist that have been overlooked. Many of the genes presently thought to end at a UGA codon may in fact be modular genes. As mentioned previously, because the recognition that UGA does not always function as a termination codon is relatively new, much of the work done using recombinant protein may have missed a potential portion of the gene.

The hypothesis that is tested in this thesis is this: Many proteins thought to end at a UGA codon are actually transcribed past the UGA, and are modular proteins in the sense that the read through of the UGA can be regulated. The ideal protein for studying this hypothesis has several important characteristics. It

should be a protein that was discovered more than twenty years ago and is presumed to end in a UGA codon so that much of what is known about this protein was learned before the multifunctional nature of the UGA codon was discovered. It should be a protein that received significant research attention early on, so that information such as the gene sequence, intron, and exon locations were identified before an appreciation of the complex function of the UGA codon were developed. It should have a relatively short reading frame past the UGA codon, so that molecular weight studies on the native protein may have missed the extra tail. It should also have a plausible amino acid sequence past the UGA codon. Lastly, its function should be complex, so that all of its activity is difficult to account for with the known amino acid sequence.

Cluster of differentiation antigen four (CD4) was chosen as a model protein to study this hypothesis. Studies on the lineage and development of thymocytes identified CD4 as an antigen that defined a certain population. It was originally found as a marker that defines the T-helper subset of lymphocytes by using monoclonal antibodies.[59] This major component of thymocytes is now identified as CD4+ T-helper cells, or just CD4+ T cells. It is a transmembrane protein of approximately 59Kda weight, has four transmembrane domains, and, if translation ends at the first UGA codon, a relatively short 34 aa cytoplasmic tail. The CD4 gene is located on chromosome 12, consists of seven exons, six introns and is presumed to end at a single UGA codon.[60] CD4 is only expressed at low levels (~1000 copies/cell), and this may vary depending upon the state of the cell. It is expressed in such low quantities that it is very difficult to

study in the natural state, and would not have been identified originally without the help of monoclonal antibodies. Because of the difficulties inherent in studying a protein in the natural state that is present in such low quantities, all of the detailed structural studies to date have been done using recombinant CD4 prepared using the assumption that translation ends at the UGA codon. The CD4 gene is the first protein product on chromosome 12.[60] The mRNA is transcribed from bases 1524 to 32821 of the chromosome. The CD4 gene is accepted to comprise nine introns, and encode 458 amino acids. The poly-A signal is located between bases 32790 and 32795 of the chromosome, and the polyA site is at base 32821. It is generally accepted that CD4 ends in a single UGA codon. In fact, the UGA codon presumed to end translation of CD4 is totally conserved in mammals. There are a total of three in frame UGA codons before another stop codon is encountered (TAG). If all three UGA codons were read through, the CD4 gene could extend another 627 bases, corresponding with 209 amino acids.

CD4 is a critical component of the function of CD4+ T-helper cells (Th).[59] CD4 forms part of a receptor complex that is involved in T cell activation, the T cell receptor complex. Once antigen is recognized through the T cell receptor, a series of events occur mediated in part by CD4 that result in the activation of the T cell. T cells recognize processed protein antigen in association with major histocompatibility (MHC) class 2 antigen. CD4 binds to class II MHC receptors, and brings the antigen displayed with the MHC receptor into close proximity to the T cell. The CD4 associated with the MHC class II

receptor diffuses through the T cell membrane until antigen displayed with the MHC class II receptor is recognized by and binds to CD3, another component of the T cell receptor. Once this occurs, a series of intracellular events are set in motion that result in the activation of the T cell. P56lck, that is associated with the cytoplasmic tail of CD4, phosphorylates a tyrosine residue.[61, 62] It then dissociates from CD4, and diffuses through the cell and activates phospholipase C. Phospholipase C degrades phosphoinositol triphosphate into diacylglycerol and inositoltriphosphate. Diacylgylcerol in turn, activates protein kinase C, which activates NFkB. NFkB binds to the promoter region of many genes, and activates gene transcription. The inositol triphosphate released by phospholipase C diffuses through the cell, and causes the release of calcium from the endoplasmic reticulum and the cell membrane. The increased intracellular calcium activates calcium dependent protein kinases, which are able to activate other transcription promoter proteins in the cell.

The large size of CD4 makes the resolution of 5000 Daltons difficult using gel electrophoresis. Normal glycosylation of CD4 has made exact molecular weight determinations of the native protein difficult and has been used to explain higher than expected molecular weight measurements of the native protein. Of course, molecular weight determinations using recombinant protein are useless in providing information about the possible translation of CD4 past the UGA codon as they are typically made using truncated versions of the native CD4 mRNA. Early measurements of the native antigen identified as CD4 reported weights as high as 62Kda. 55Kda to 59Kda has become the accepted molecular

weight range, with 59Kda considered the weight of the glycosylated protein product.

The possibility that CD4 is translated past the UGA has been discussed in the literature previously.[63] Most of the literature focuses on the region past the second 3'-UGA in the +1 frame. Facchiano first reported evidence of the potential read through by showing that there is some similarity between the HIV GP120 gene and the +1 frame of the 3' end of the CD4 gene past the second UGA codon.[64] Facchiano further investigated this similarity, and also suggests that Eukaryotic DNA may contain occasional polycistronic regions, or that they once did.[65] In order to support this possibility, in this region, the reading frame must be shifted from 0 to +1 somewhere between the known CD4 gene, and the beginning of the gene region in the next frame. In fact, a potential frameshift site has been identified that is closely associated with the second UGA codon. The potential frameshift sequence is GGUUGA.[63] A potential SECIS structure was also identified in this reference.

Given the nature of the UGA codon as leaky, the obvious question is what happens to proteins in which translation erroneously proceeds past the UGA codon? Incorrectly translated protein is quickly degraded within the cell, for example, through the ubiquitin pathway.[66, 67] Translation past the UGA codon erroneously could significantly increase the length of the protein, and may alter its folding process. This would be a very large mistake of translation when compared to a single mutation. Incorrectly translated protein should be a very small minority of the total transcripts, and because the degradation process is

efficient, it is highly unlikely that a detectable number of faulty proteins of this type could make it to their final destination within the cell.

The research strategy employed to investigate the possibility of translation past the UGA codon in CD4 is unique. Typically, read through of a UGA codon is proven using engineered plasmid constructs using reporter genes or incorporation of selenium 75. Neither of these strategies is desirable for this research project. In order to prove that a gene such as CD4 is actually translated past the UGA codon, it is necessary to study the protein in a native system with as little external manipulation as possible. If engineered plasmid constructs are used without significant in vivo evidence to support them, then the evidence they could provide will be very weak. One could always argue that any read through observed in such a system was due to the manipulation of the experimental system, and does not occur in nature. Therefore, a unique strategy was developed to study CD4 in a living cell. A living cell was chosen for study in order to investigate the potential read through of the UGA codon in as nearly natural state as possible. For this project, cancerous CD4+ human T cell lines are grown and studied as the source of CD4. Cancerous cells are used because producing large amounts of protein using them is relatively simple as compared to using peripheral blood monocytes isolated from human subjects. This is potentially important because CD4 is a relatively rare protein within the cell. There is no evidence that the cancerous cells would be more or less likely to have non-specifically read through UGA codons compared to normal cells. The genetic structure of the CD4 gene is not manipulated experimentally within these

cell lines. The overall research strategy is as follows. Synthetic peptides are produced which mimic both the translated region before the UGA codon as well as the hypothetical region past the UGA codon. Monoclonal antibodies are then developed against the synthetic peptides. The monoclonal antibodies are used in a variety of labeling experiments comparing the relative binding of the antibodies against the region prior to the UGA codon to the antibodies against the region after the UGA codon. Labeling by the antibodies against the region after the UGA codon is evidence that that the UGA codon in CD4 is read through in a living cell without experimental modification of the gene structure. Engineered plasmid constructs are used in future experiments in order to bolster the evidence obtained using antibody labeling. The plasmid construct experiments are described in the future experiments chapter.

CHAPTER 2

SEQUENCE ANALYSIS

Nucleic Acid Sequence Analysis

Introduction

Additional sequence analysis is needed to further investigate the possibility that CD4 is transcribed past the UGA codon, and to determine the best sequences to use for making synthetic peptides for antibody production. In particular, no functional SECIS has been conclusively identified in the CD4 3' untranslated region (UTR). This sequence analysis study takes a direct approach towards finding potential SECIS elements. A survey of the entire genome for potential SECIS elements is conducted in an effort to identify other proteins as well that are presumed to end in UGA that potentially encode selenocystine. Although a SECIS is not sufficient to cause a UGA to encode selenocystine, the SECIS is likely necessary and the presence of a SECIS is very strong evidence of a selenoprotein.[50] Selenoproteins have been identified using similar techniques in which a search of folding structures similar to SECIS elements was done.[68]

Methods

For this analysis, a search of potential SECIS elements was done online using Blast 2.2.1 available at the NCBI.[69] Default options designed to find short, nearly exact sequence matches were utilized so that the short sequence

similarities would not be ignored. The sequences of several of the known SECIS elements were used as the query sequence with the hope of finding sequences in other genes including CD4 that may be undiscovered SECIS elements. The sequences of the human selenoprotein W, rat deiodinase, S. mansoni glutathione peroxidase, and rat glutathione peroxidase SECIS elements were used as query sequences. Graphical representations of those SECIS elements along with the sequences used are shown in Figure 2.1.[50]

Results and Discussion

Human SelW

Not surprisingly, a search using the human selW SECIS sequence returned many hits corresponding to human selW, and selW analogues from other species including Mus musculus, Ovis aries, Sus scrofa, and Macaca mulatta. Many of the sequence similarities identified with the human SelW SECIS also have critical features of functioning SECIS elements. Several interesting sequence similarities were identified, however, that have equal to or better expectation values and sequence similarities to the human SelW SECIS than does the mus musculus SelW homologue. However, only sequence similarities to human genome sequences will be discussed in this analysis, because inclusion of all the potentially interesting sequence similarities would increase the length of this section greatly. In addition, only sequence similarities containing a significant portion of the SECIS, for example the three to four adenine sequence, will be discussed. One of the most critical features of the SECIS element that appears to be a requirement for its function is the sequence

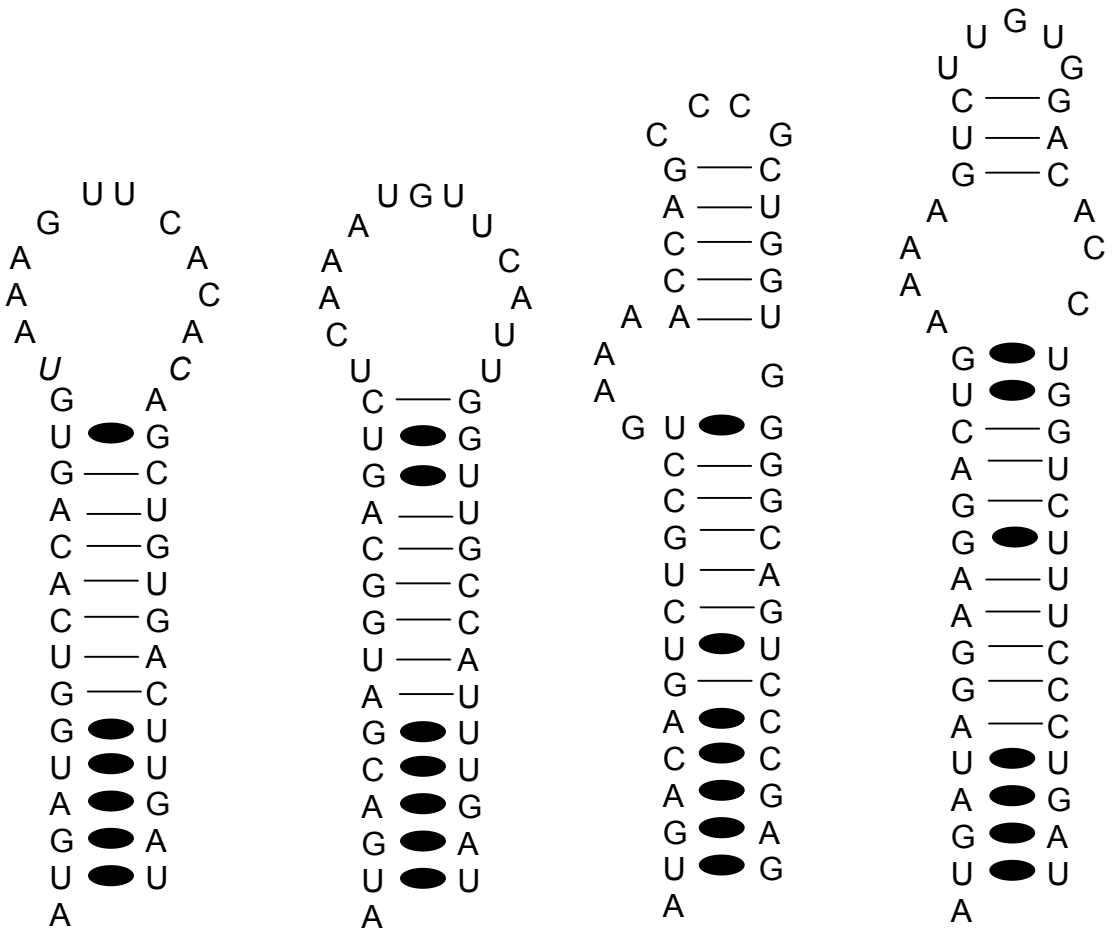


Figure 2.1. Predicted Secondary Structures of Several Known SECIS Elements, (From Low, Berry, 1996)[50]

of adenines present in the loop structure, although as with other parts of the SECIS structure, this too can be variable.[70] Significant sequence similarities were found on chromosomes 4,5,8,10, 12, 19, and 21. Human selenoprotein W itself is located on chromosome 19. The sequence similarities to the human SelW SECIS, which contain important portions of a functional SECIS element, are shown in Figure 2.2. Part of sequence number 4938316 of chromosome 1 shared 95% sequence similarity to the human SelW SECIS sequence, and returned an expectation value of 1.4 whereas that of the Mus musculus selW homologue is 5.7. The sequence similarity was found between base 8 and 29 of the query, and base 100694 and 100715 of sequence number 4938316. Of these 22 overlapping bases, 21 were identical. This region of sequence number 4938316 corresponds to the 3' part of a tyrosine phosphatase. This is interesting because SECIS elements are typically found in the 3' region of known selenoproteins, and this sequence contains the four-adenine region.

There was a region of sequence similarity identified on chromosome 12 as well. This sequence similarity was identified between base 6 and 23 of the SelW sequence and base 4241 and base 4258 of sequence number 15055223 that covers the front half of the SECIS structure including the sequence of adenines. Of the 18 bases in this region, all 18 are identical, resulting in an expectation score of 1.4. This sequence of chromosome 12 has only been analyzed for repeat sequences, and does not correspond to any of the identified repeat regions. However, no sequence similarity could be identified between this region and the sequence of chromosome 12 entry number U47924 (complete

>gi|15055223|gb|AC026364.35| Length = 35916

Score = 36.2 bits (18), Expect = 1.4
Identities = 18/18 (100%) Strand = Plus / Plus
Query: 6 aggaaggactgaaaagtc 23
Sbjct: 4241 aggaaggactgaaaagtc 4258

>gi|14726524|ref|XM_046111.1| Length = 2664
Score = 36.2 bits (18), Expect = 1.4
Identities = 18/18 (100%) Strand = Plus / Plus
Query: 18 aaagtcttgtggacacct 35
Sbjct: 2507 aaagtcttgtggacacct 2524

>gi|14249840|gb|BC008294.1|BC008294 Length = 2686
Score = 36.2 bits (18), Expect = 1.4
Identities = 18/18 (100%) Strand = Plus / Plus
Query: 18 aaagtcttgtggacacct 35
Sbjct: 2508 aaagtcttgtggacacct 2525

>gi|4938316|emb|AL049570.11|HSJ437I16 Length = 120385
Score = 36.2 bits (18), Expect = 1.4
Identities = 21/22 (95%) Strand = Plus / Plus
Query: 8 gaaggactgaaaagtcttgtgg 29
Sbjct: 100694 gaaggactgaat agtcttgtgg 100715

>gi|3136000|emb|AL008637.1|HS833B7 Length = 86574
Score = 34.2 bits (17), Expect = 5.7
Identities = 17/17 (100%) Strand = Plus / Plus
Query: 8 aggaaggactgaaaagt 22
Sbjct: 8964 aggaaggactgaaaagt 8980

Figure 2.2. Sequence similarities Identified with the Human SelW SECIS

sequence), so its potential location on chromosome 12 relative to the CD4 gene could not be determined.

The sequence similarity identified with sequence number14726524 from base 2507 to base 2524 is particularly interesting. The sequence similarity begins with and includes the three-adenine sequence, and continues across the upper portion of the SECIS loop. Of the 18 bases in this region, all 18 are shared between the two sequences resulting in an expectation score of 1.4. Sequence number 14726524 is the sequence of a hypothetical gene supported by mRNA sequencing. However, the chromosome location or other information about this hypothetical protein is not revealed, so a BLASTn search was done using sequence number 14726524 as the query sequence. This search revealed several different sequences of the same or a totally homologous gene with expectation values of 0. Representative of these is sequence number 19911075. This sequence reveals that the gene is TIFA, or is homologous to TIFA, an adapter protein that links TNF receptor-associated factor 6 (TRAF6) to interleukin-1 (IL-1) receptor-associated kinase-1 (IRAK-1) in IL-1 receptor signaling. This gene is presumed to end at a single UGA codon. Further, the sequence similarity identified with the SelW SECIS is found 1686 bases after the UGA codon that is presumed to end the gene sequence. This region in the 3' UTR of the mRNA is within the region where known SECIS elements are known to reside.

Another interesting sequence similarity was identified with sequence number 14249840 between bases 2508 and 2525.[71] Of the 18 bases in this

region, all 18 are shared between the two sequences including the adenine sequence, resulting in an expectation score of 1.4. Sequence number 14249840 is that of another unknown protein identified from mRNA harvested from an endometrial adenocarcinoma. Again, significantly, the gene is presumed to terminate at a single UGA codon at base 524 of the sequence. The sequence similarity identified with the SelW SECIS element, is found within the 3' UTR of the mRNA, 1984 bases past the UGA. This is well within the region that a SECIS element could be found. Because both were unidentified, and the region of sequence similarity was so similar, it seemed that the two genes identified by sequences 14726524 and 14249840 were likely to be homologous. However, the amino acid sequences reported for the two genes are very different. The fact that these two sequences shared the same significant portion of the SelW SECIS was striking. This prompted a further analysis of this region of these two genes to see if there was further sequence similarity between the potential SECIS elements of these two genes. BLAST analysis at the nucleotide level revealed that the sequences were almost totally identical, with 14249840 having an extra adenine at position 498 of the sequence. This made the very different amino acid sequences reported for the gene perplexing. The mutation is within the coding region of the gene. Interestingly, even with the mutation of an extra adenine, the gene still terminates in a single UGA codon. Protein-protein BLAST revealed that the sequences do in fact share a portion of their sequence. The amino acid sequence of the gene reported in 14249840 starts well before the methionine start codon reported in 14726524. The result is that the beginning of

sequence number 14726524 can be found in the middle of number 14249840, and the amino acid sequence of 14249840 terminates in the gene sequence well before the amino acid sequence of 14726524 terminates. The beginning and end of 14249840 is very different than anything found in 14726524, and the end of 14726524 cannot be found in 14249840. Protein-protein BLAST using each of the aforementioned amino acid sequences suggests that they are probably the same gene product, and that the amino acid sequence reported in 14249840 is probably incorrect. The reason for the difference in reported gene start position is unknown. The possibility that these could be two different gene products seems unlikely.

Another sequence similarity found in human DNA to the human SelW sequence having the sequence of adenines is that in sequence number 4938316. Of the 22 bases contained between base 100694 and base 100715, 21 of them are shared between the two sequences resulting in an expectation score of 1.4. The sequence of adenines is in the middle of this region. This region of is also reported to be in the 3' region from a gene for a protein-tyrosine phosphatase. The portion of sequence number 3136000 between base 8964 and 8980 is homologous to a portion of the human SelW sequence between bases 6 and 22. This sequence encodes the NCF4 gene for cytosolic neutrophil factor 4, not a known selenoprotein. This sequence also contains the adenines and covers the first half of the SECIS loop structure. Of the 17 bases in this region, all 17 of them are shared between the two sequences, resulting in an expectation score of 5.7.

Rat DI

A search for potential SECIS elements was also done using the rat deiodinase SECIS element as the query sequence. Sequence similarities were identified on human chromosomes 2, 7, 8, 10, 11, 12, 14, 19, 20, and 21. Fewer significant novel sequence similarities were found using this search than were found using the human selW SECIS sequence. Of the sequence similarities identified between the rat DI SECIS element and human DNA sequences, several have critical features of a functioning SECIS element. The sequence similarities that contain important portions of a functional SECIS element are shown in Figure 2.3. There were two sequence similarities identified on chromosome 7, sequence numbers 5732182 and 4309817.[72] Both sequence similarities cover the sequence of adenines. That with sequence number 5732182 is in a region that is not identified as a coding region, and returned a significant expectation score of 1.1. This sequence similarity was identified between bases 18 and 35 of the query sequence and bases 67409 and 67392 of sequence number 5732182 and begins at the adenine sequence, including two of three of them, and continues across the top of the loop structure and down the backside of the loop. The sequence similarity with chromosome 7, sequence number 4309817 between bases 75800 and 75817 begins immediately after the AUGA sequence at the base of the stem and continues to the top of the SECIS loop structure including all three adenines. This sequence similarity on sequence number 4309817 is mostly contained within a region identified as an AT rich repeat region and returned an expectation score of 1.1. This region is between

>gi|5732182|gb|AC006454.3|AC006454 Length = 153201
Score = 36.2 bits (18), Expect = 1.1
Identities = 18/18 (100%) Strand = Plus / Minus
Query: 18 aagttcacacagctgtga 35
Sbjct: 67409 aagttcacacagctgtga 67392

>gi|4309817|gb|AC006315.2|AC006315 Length = 127526
Score = 34.2 bits (18), Expect = 1.1
Identities = 18/18 (100%) Strand = Plus / Plus
Query: 5 tggtcacagtgtaaagtt 22
Sbjct: 75800 tggtcacagtgtaaagtt 75817

>gi|7408076|gb|AC010967.2|AC010967 Length = 197328
Score = 34.2 bits (17), Expect = 4.4
Identities = 17/17 (100%) Strand = Plus / Plus
Query: 18 aagttcacacagctgtg 34
Sbjct: 191357 aagttcacacagctgtg 191373

>gi|8247025|emb|AL121928.13|HSBA18I14 Length = 167480
Score = 34.2 bits (17), Expect = 4.4
Identities = 20/21 (95%) Strand = Plus / Minus
Query: 14 tgtaaagttcacacagctgtg 34
Sbjct: 73117 tgtaaagatcacacagctgtg 73097

>gi|5805137|emb|AL035078.32|HS17K7 Length = 130336
Score = 34.2 bits (17), Expect = 4.4
Identities = 17/17 (100%) Strand = Plus / Minus
Query: 18 aagttcacacagctgtg 34
Sbjct: 4741 aagttcacacagctgtg 4725

Figure 2.3. Rat DI SECIS Sequence similarities

bases 5 and 22 of the query sequence and bases 75800 and 75817 of sequence number 4309817. This sequence has been identified as part of the glucokinase gene in sequence number 190109 and is identified as a repeat polymorphism associated with early onset type-2 diabetes.

However, the sequence similarity found with sequence 4309817 and shared with 190109 highlights one of the difficulties in finding and studying potential selenoproteins. In this area of their sequences, they share 100% sequence similarity, which returned an expectation score of 1.1. This is a significant score for a sequence this short. This region of sequence number 190109 is identified as a repeat polymorphism within the human glucokinase gene. Therefore, legitimate SECIS element sequences may share significant sequence similarity with legitimate repeat sequences, thus making them difficult to differentiate in all cases. Even if this sequence is correctly identified as a repeat polymorphism, the fact that it shares significant sequence similarity with the rat DI SECIS highlights this potential pitfall. The other possibility that could explain this sequence similarity is even more disturbing. Conventional wisdom in dealing with gene sequence analysis does not allow for the existence of SECIS elements in sequences not yet identified as selenoproteins. The natural tendency of researchers identifying potential gene sequences and ascribing functions to sequences outside coding regions may be to simply assume they are polymorphisims of known repeat sequences rather than investigate the possibility that they possess some other function such as that of a SECIS element.

The sequence similarity with chromosome 2, sequence 7408076, between bases 191357 and 191373 begins at the adenine sequence and includes two of three of them.[72] The sequence similarity continues down the backside of the SECIS loop. This sequence similarity returned an expectation score of 4.4 and this region of sequence 7408076 is identified as belonging to the MIR repeat family. The sequence similarity to chromosome 10, sequence number 8247025, between bases 73117 and 73097 also includes all three adenines. The sequence similarity includes the entire upper portion of the SECIS loop structure. This sequence is identified as a member of the AluSx repeat family. The sequence similarity with chromosome 11, sequence number 5805137, from base 4741 and 4725 begins at the adenine sequence. The sequence similarity includes two of the three adenines, and continues down the backside of the SECIS loop. This region is not identified as anything in particular in this sequence submission. None of the sequence similarities found with the rat DI SECIS included both the critical adenines and the AUGAU sequence at the base of the SECIS stem. It may be worth noting that the sequence similarity with chromosome 20, sequence 9408738 between bases 44159 and 44143 included the UGAU sequence at the base of the backside of the SECIS stem. This sequence is identified as part of the coding region for a novel whey acidic protein type gene.

Rat GPx

Several of the sequence similarities identified between human DNA sequences and the rat GPx SECIS also possess important characteristics of

>gi|12025623|gb|AC011005.7|AC011005 Length = 217615

Score = 36.2 bits (18), Expect = 1.3
Identities = 18/18 (100%) Strand = Plus / Plus

```
Query: 9        ctgcctgaaaaccagccc  26
                 ||||||||||||||||||
Sbjct: 75889    ctgcctgaaaaccagccc  75906
```

>gi|13811889|gb|AC011912.7|AC011912 Length = 175264
Score = 34.2 bits (17), Expect = 5.3
Identities = 17/17 (100%) Strand = Plus / Plus

```
Query: 1        atgacagtctgcctgaa  17
                 |||||||||||||||||
Sbjct: 109131   atgacagtctgcctgaa  109147
```

>gi|13560021|emb|AL391379.12|AL391379 Length = 116800
Score = 34.2 bits (17), Expect = 5.3
Identities = 17/17 (100%) Strand = Plus / Plus

```
Query: 8        tctgcctgaaaccagc  24
                 ||||||||||||||||
Sbjct: 37211    tctgcctgaaaccagc  37227
```

Figure 2.4. Rat GPx SECIS Sequence similarities

functional SECIS elements. The sequence similarities that contain important portions of a functional SECIS element are shown in Figure 2.4. The sequence similarity identified with chromosome 7, sequence number12025623, base 75889 to 75906, includes the sequence of adenines as well as the upper portion of the SECIS loop. The 18 bases in the region are all identical, with an expectation value of 1.3. This sequence of chromosome 7 is not analysed, so it is unknown whether or not this sequence similarity is close to a gene. The sequence similarity found on chromosome 15, sequence number 13811889, base 109131 to 109147, includes the AUGA sequence at the front base of the SECIS stem, and includes up to two of the three critical adenines. Of 17 bases in this region, all 17 are identical, producing an expectation value of 5.3. This sequence of chromosome 15 has not been analyzed. The sequence similarity on chromosome X, sequence number 13560021, base 37211 to 37227 covers the front side of the SECIS loop, including the critical adenine sequence. Of 17 bases in this region, all 17 are identical, resulting in an expectation score of 5.3. No genes have been identified in this sequence of chromosome X, and this region has not been identified as a repeat region.

S. mansoni

The S. mansoni SECIS element was also used as a query sequence. Sequence similarities having an expectation score of less than 4.6 were identified on chromosomes X, 1, 2, 3, 10, 14, and 18. Several of the sequence similarities found between the S.mansoni GPx SECIS and human DNA also have critical features of a functional SECIS. The sequence similarities that contain important

>gi|13273693|emb|AL355606.14|AL355606 Length = 163283
Score = 38.2 bits (19), Expect = 0.29
Identities = 19/19 (100%) Strand = Plus / Minus
Query: 18 aaatgttcattggttgcca 36
Sbjct: 19949 aaatgttcattggttgcca 19931

>gi|11136696|gb|AC026884.5|AC026884 Length = 150583
Score = 34.2 bits (17), Expect = 4.6
Identities = 17/17 (100%) Strand = Plus / Plus
Query: 11 cagtctcaaatgttcat 27
Sbjct: 52049 cagtctcaaatgttcat 52065

>gi|6449519|gb|AC006961.16|AC006961 Length = 171419
Score = 34.2 bits (17), Expect = 4.6
Identities = 20/21 (95%) Strand = Plus / Plus
Query: 17 caaatgttcattggttgccat 37
Sbjct: 68582 caaatgttcattgattgccat 68602

>gi|9211657|emb|AL132716.6|CNS01DTF Length = 155619
Score = 34.2 bits (17), Expect = 4.6
Identities = 17/17 (100%) Strand = Plus / Plus
Query: 5 cgatggcagtctcaaat 22
Sbjct: 8964 cgatggcagtctcaaat 8980

Figure 2.5. S. mansoni GPx SECIS Sequence similarities

portions of a functional SECIS element are shown in Figure 2.5. A very significant sequence similarity was identified between bases 18 and 36 of the query sequence and bases 19949 and 19931 of chromosome X, sequence number 13273693. This sequence similarity returned an expectation score of 0.29, very significant, especially for a query sequence of only 42 letters. This sequence similarity begins at and includes the three adenine sequence and continues across the top of the SECIS loop and down the back side. Of the 19 bases in this region, all 19 are shared between the sequences, resulting in an expectation score of .29, significant. This sequence is on chromosome X in a region identified as an L1P and L1P3 repeat. The sequence similarity found on chromosome 10, sequence number 11136696 bases 52049 to base 52065 also includes the three-adenine sequence. The three-adenine sequence is in the middle of this region, and of the 17 bases, all 17 are identical between the two sequences, resulting in an expectation score of 4.6. Sequence number 11136696 has not been analyzed for gene sequences. The sequence similarity on chromosome 18, sequence 6449519 bases 68582 to 68602, also contains the three adenine sequence. The sequence similarity begins just before the adenine sequence, and continues across the top of the SECIS loop. Of the 21 bases in this region, 20 are shared between the two sequences, resulting in an expectation score of 4.6. This region is identified as an L2 repeat. The sequence similarity on chromosome 14, sequence number 9211657, base 151124 to 151140 begins near the base of the SECIS stem, and includes the

three adenine sequence. Of the 17 bases in this region, all 17 are shared between the two sequences, resulting in an expectation score of 4.6.

Conclusion

Although several interesting sequence similarities were identified throughout the human genome, on many different chromosomes, none were identified in the 3'UTR of the CD4 gene on chromosome 12. Because of the problems associated with sequence similarity searching with short sequences such as that of a SECIS element, and the possibility of undiscovered variants of functional SECIS elements, the inability to identify a potential SECIS in the 3' UTR of the CD4 gene does not necessarily mean that CD4 is not a potential selenoprotein. However, the lack of an identifiable SECIS element supports the assertion that the UGA is read through by some mechanism other than selenocysteine insertion. Add notes about possible similarities to repeat regions.

Protein Sequence Analysis

Introduction

To further investigate the possibility that CD4 is transcribed past the UGA codon, BLAST analyses were done using various parts of the polypeptide that would result from translation of the region past the UGA, in all three reading frames. This study will identify similarities at the amino acid level, which may turn up significant results even if the two proteins are not very similar at the nucleotide level. Therefore, this study is more likely to identify functional as opposed to evolutionary similarities.

Methods

From First UGA to Next Stop Codon

The first BLAST search that was done was using the zero frame, in frame with the CD4 gene, as though the UGA were read through without any change in frame. From the UGA presumed to end translation of the CD4 gene to the next stop codon are 168 nucleotides, enough for 56 codons for 56 amino acids. This sequence, Figure 2.6, was used as a query for a BLASTp search. In addition, settings for the BLAST search were chosen specifically for doing a search using short sequences, and are designed to find short, nearly exact matches. Again, short sequences are challenging for the statistics of the BLAST search design.

BLASTp Analysis From Potential +1 Frameshift Site to Next Stop

A BLASTp search was also done using a portion of the +1 frame of the region past the UGA in CD4. The region chosen for the BLAST analysis begins at the second UGA codon, the first one after the UGA presumed to end CD4, which happens to be another UGA, and continues to the next non-UGA stop codon. This region is illustrated in Figure 2.6. There are 486 bases in this region, potentially encoding 162 amino acids, which were used for this analysis. This BLAST analysis was also done using options designed to find short, nearly exact matches.

BLASTp Analysis of –1 Reading Frame

A BLAST search was also done using the –1 reading frame of the region past the UGA presumed to end translation of the CD4 gene. The area chosen as the query sequence is the region from the presumed end of the CD4 gene to the

3’ End of CD4 Gene, Zero Frame, From First UGA to Second UGA

gtrpgrshlqppqvsaprflpadqmnvadpqplascspplqfaivspglg pgftg

3’ End of CD4 Gene, +1 Frame, From Potential Frameshift Site to Next in Frame, Non-UGA Stop codon

eccslvsrglitpsstpfpfpsslallslflsdplptahldprgvfraspgwhgg gwvsgsmehgtvllqdrtlgpqragtctkshsqasqgwmqiqrflaasts cpmlpashpmwvgpqthiltlhkqplwtqphvhglkgcltssvylrlrkilq gwqqn

3’ End of CD4 Gene, -1 Frame, From First UGA, to Next in Frame Stop Codon

hearqiplaaspgvcpafpacgpdecsrspasgllfasstichcfswvrprl hwlsval

Figure 2.6. Query Sequences

next, in frame stop codon. This region is 177 nucleotides long, potentially encoding 59 amino acids, Figure 2.6. The 59 amino acid long translated sequence was used as the query sequence. This BLAST analysis was also done using options designed to find short, nearly exact matches.

Sequence Similarity Analysis Between The Translated Region Past The 3'-UGA of CD4, and Glutathione Peroxidase

Sequence similarity was assessed using the BlockAlign program (Zhang, Kececioglu, and Taylor, unpublished). The program produces an optimal gapped alignment of the query sequence against an existing grouped sequence alignment. By random shuffling of the query sequence, the average similarity score for optimally aligned random sequences of identical composition can be calculated. Significance is calculated as the distance of the actual score from the average random score in standard deviations. For this analysis, the alignment is shuffled 100 times to calculate the average score expected by random chance, the PAM 250 matrix is used, the gap insertion penalty is 7, and the gap elongation penalty is 1.

Results and Discussion

From First UGA to Next Stop Codon

Despite the challenges of using short sequences, eight sequence similarities were identified with expectation scores of less than 9, and two of these had expectation scores of 0.59. The results are shown in Figure 2.7a and Figure 2.7b. Both of the most significant alignments were with BarH homeobox proteins, or BarH-like homeobox proteins from various species. Sequence

>gi|14149728|ref|NP 064448.1|; gi|29336920|sp|Q9BZE3|BRH1 HUMAN;
gi|12276173|gb|AAG50279.1|AF325688 1; gi|16549407|dbj|BAB70807.1|;
gi|14029398|gb|AAK52674.1|AF321618 1; gi|9506423|ref|NP 062319.1|;
gi|16923956|ref|NP 476450.1|; gi|29336955|sp|Q9JK26|BRH1 MOUSE;
gi|8217344|emb|CAB92529.1|; gi|11463945|dbj|BAB18599.1|;
gi|11463947|dbj|BAB18600.1|; gi|14279212|gb|AAK58534.1|AF264026 1;
Length = 327 Score = 34.1 bits (73), Expect = 0.59
Identities = 20/40 (50%), Positives = 25/40 (62%), Gaps = 10/40 (25%)
Query: 7 SHLQPPQVSAP - - - R - - - - - FLPADQMNVADPQPLASCSP 38
SHLQP Q+SAP R FL D +AD +PLA+ C+P
Sbjct: 76 SHLQPGQLSAPAQSRTVTSSFL I RD - - I LADCKPLAACAP 113

>gi|14165239|gb|AAK55444.1|AF380350 1 Length = 416
Score = 32.9 bits (70), Expect = 1.4
Identities = 14/26 (53%), Positives = 17/26 (65%), Gaps = 3/26 (11%)
Query: 4 PGRSHLQPPQVSAPRFLPADQMNVAD 29
P RSHL P AP FLPA+ +N +D
Sbjct: 182 PTRSH LL P - - -RAPSFLPAEDVNHS D 204

>gi|28544693|ref|XP 287640.1| Length = 100
Score = 32.0 bits (68), Expect = 2.6
Identities = 12/16 (75%), Positives = 12/16 (75%), Gaps = 1/16 (6%)
Query: 3 RPGRSHLQPPQVSAPR 18
RP RS PPQVSAPR
Sbjct: 61 RP- -RSRAPPPQVSAPR 75

>gi|21665898|emb|CAD20302.1| Length = 394
Score = 31.2 bits (66), Expect = 4.6
Identities = 13/18 (72%), Positives = 13/18 (72%), Gaps = 5/18 (27%)
Query: 29 DPQ - - P- -LAS - CSPPLQ 41
DPQ P LAS CSPPLQ
Sbjct: 297 DPQALPASLASVCSPPLQ 314

Figure 2.7a. CD4 Tail, Zero Frame

>gi|27717549|ref|XP 234900.1| Length = 537
Score = 30.8 bits (65), Expect = 6.2
Identities = 8/9 (88%), Positives = 9/9 (100%)
Query: 25 MNVADPQPL 33
M+ VADPQPL
Sbjct: 1 MDVADPQPL 9

>gi|27670188|ref|XP 239515.1| Length = 288
Score = 30.8 bits (65), Expect = 6.2
Identities = 11/19 (57%), Positives = 11/19 (57%), Gaps = 6/19 (31%)
Query: 16 APRFLPADQM - - - - - - NVA 28
APRFLP QM NVA
Sbjct: 183 APRFLPPSQMETVNGWNVA 201

>gi|15806969|ref|NP 295694.1|
gi|7472435|pir|D75331|
gi|6459764|gb|AAF11528.1|AE002035 11 Length = 400
Score = 30.3 bits (64), Expect = 8.3
Identities = 17/38 (44%), Positives = 21/38 (55%), Gaps = 14/38 (36%)
Query: 20 LPADQMNVAD - - PQPLASCSPPL - - - - Q - - - FAIVSPG 48
LP M+VA+ P PL+ SPPL Q FA+V G
Sbjct: 5 LP - - - MSVAEPAPAPLS - - SPPLAPPPQP LTFAVVGAG 37

Figure 2.7b. CD4 Tail, Zero Frame

number 14149728 represents the human BarH-like homeobox protein sequence.[73] A sequence similarity was identified with this sequence between amino acid 76 and 113 corresponding with amino acids 7 to 38 of the query sequence. Including gaps in the alignment inserted into each sequence, this region corresponds to 40 amino acids. Of the 40 amino acids contained in this region, 20 of them are identical between the two sequences, and 25 (62%) are positive matches. Gaps were 10 amino acids total, corresponding to 25% of the sequence region. This resulted in an expectation score of 0.59, significant for a query sequence this length. The alignment is near the middle of the BarH-like homeobox domain. The other alignments with expectation scores of less than 9 were all with proteins from other species and included the following: several BarH domains, expectation scores of 0.59; interferon regulatory factor 10 (Gallus gallus), expectation score 1.4; hypothetical protein XP_287640 (Mus musculus), expectation score 2.6; Alpha 2B adrenergic receptor (Tachyoryctes sp.), expectation score 4.6; serine/threonine kinase II-like protein (Mus musculus and Rattus norvegicus), expectation score 6.2; hypothetical protein XP_239515 (Rattus norvegicus) expectation score 6.2; and hypothetical protein 15806969 (Deinococcus radiodurans), expectation score 8.3.[74] The most significant findings are those with the lowest expectation scores, particularly those less than 1.

BLASTp Analysis From Potential +1 Frameshift Site to Next Stop

This region was chosen because this frame is potentially a viable reading frame in this region because of a putative slippery sequence at the next UGA

after the one presumed to end translation of the CD4 gene. If translation continues into the region through this UGA, the ribosome could shift one base, resulting in a +1 frame shift as translation continues. Therefore, if any region is translated in the +1 frame, this should be the place. The results of the BLAST search do not refute this hypothesis, Figure 2.8. Although no very long sequence similarities were identified with known proteins, there were 6 alignments identified with expectation scores of less than 9, and one of those has an expectation score of 0.27. This most significant alignment, sequence number 29738242, amino acid 104 to 128 corresponding with amino acids 6 to 36 of the query, is to a hypothetical protein identified using an automated “genome scan” of a portion the sequence of chromosome 17, hypothetical protein XP_296973. There are gaps totaling 10 amino acids (30%), causing the resulting region of sequence similarity to be 33 amino acids long. Of the 33 amino acids in this region, 18 (54%) of them are shared between the two sequences. Twenty of them (60%) are positive matches. This results in an expectation score of 0.27, which is significant for a relatively short query sequence as this. No other information is available concerning this hypothetical protein. Other alignments with expectation scores less than 9 included the following: hypothetical protein 23105643 (Azptobacter vinelandii), expectation score 1.2; b114189 (Bradyrhizobium japonicum), expectation score 5.1; CG4593-PA (Drosophila melongaster), expectation score 6.8; hypothetical protein XP_297729 (Homo sapiens), expectation score 6.8; and RE66240p (Drosophila melongaster), expectation score 6.8.

>gi|29738242|ref|XP 296973.1| Length = 132

Score = 37.5 bits (81), Expect = 0.27
Identities = 18/33 (54%), Positives = 20/33 (60%), Gaps = 10/33 (30%)
Query: 6 VSRGLITPSSTPFPFPSSLAL- -LSLFLSDPLP 36
VSR TP+ TPFP +L L LSLF PLP
Sbjct: 104 VSR- - - TPAPTPFP - - -NLGLRVLSLFF- -PLP 128

>gi|23105643|ref|ZF 00092099.1| Length = 419
Score = 35.4 bits (76), Expect = 1.2
Identities = 20/38 (52%), Positives = 20/38 (52%), Gaps = 18/38 (47%)
Query: 24 LA- LLSLFL- - - SDPLPTAHLDPRGV- - - - F- RASPGW 52
LA LLSLFL SDPLP GV F RA GW
Sbjct: 138 LAILLSLFLWRASDPLP - - - - - - - GVLNIRFERA - - GW 166

>gi|27379300|ref|NP 770829.1| b114189
gi|27352451|DBJ|BAC49454.1| b114189 Length = 325
Score = 33.3 bits (71), Expect = 5.1
Identities = 18/41 (43%), Positives = 21/41 (51%), Gaps = 16/41 (39%)
Query: 32 SDPLPTAHLDPRGVFRASPGWHGGGWVSGSMEHGTVLLQDR 72
SDP PT VF +HGGGWV+ G +E T DR
Sbjct: 77 SDPAPTV - - - - - - VF - - - - - FHGGGWVAGDLE - - T - - -HDR 101

>gi|24640291|ref|NP 572377.2| CG4593-PA
gi|7290790|gb|AAF46235.1| CG4593-PA
gi|17946406|gb|AAL49236.1| RE66240p Length = 209
Score = 32.9 bits (70), Expect = 6.8
Identities = 13/21 (61%), Positives = 13/21 (61%), Gaps = 7/21 (33%)
Query: 136 HVHGLKGCLTSS - - VYLRLRK 154
HVH L SS VYLRLRK
Sbjct: 40 HVHDL - - - - - SSAHVYLRLRK 55

>gi|29744718|ref|XP 297729.1| Length = 294
Score = 32.9 bits (70), Expect = 6.8
Identities = 14/25 (56%), Positives = 17/25 (68%), Gaps = 4/25 (16%)
Query: 21 PSSLALLSLFLSDPLP - - - TAHLDP 42
P+ LALL LF+S+P P AHL P
Sbjct: 235 PAGLALLQLFISEP- PVPGAAHLSP 258

Figure 2.8. CD4 Tail, +1 Frame

BLASTp Analysis of –1 Reading Frame

There is no experimental or theoretical evidence that supports translation of this reading frame, so this BLAST analysis is a control. This search should not turn up hits with as significant expectation scores as those using the 0 and +1 frames. The results of this blast analysis compared to those of the 0 and +1 frames supports the hypothesis that the 0 and +1 frames may be translated whereas the –1 frame is not, Figure 2.9. Although there were five alignments found with expectation scores of less than 9, none were found with a statistically significant score of less than 1 as in the analyses of the 0 and +1 frame. The most significant expectation score found in this analysis were four alignments that had a score of 4.0. The sequence similarities with expectation scores less than 9 included the following: Hemagglutinin (Influenza A), expectation score 4.0; arrow (Drosophila melanogaster), expectation score 4.0; LDL-related protein LRP6 (Drosophila melanogaster), expectation score 4.0; serotonin receptor (Rattus norvegicus, Homo sapiens), expectation score 5.3; and CLC-NT2 protein (Nicotiana tabacum), expectation score 5.3.

Sequence Similarity Analysis Between The Translated Region Past The 3'-UGA of CD4, and Glutathione Peroxidase

A detailed analysis of the potential sequence similarity between the 3' end of the CD4 gene and glutathione peroxidase reveals striking similarities. The region containing the first 3'-UGA codon in the CD4 gene, when translated, and if the UGA is presumed to encode selenocysteine, corresponds very well with the

>gi|18496100|emb|CAD20336.1| Length = 566
Score = 31.6 bits (67), Expect = 4.0
Identities = 15/34 (44%), Positives = 16/34 (47%), Gaps = 14/34 (41%)
Query: 22 GPDECSRSPASGLLFASSTICHCFSWVRPRLHWL 55
G D C R PASG F FS RL+WL
Sbjct: 151 GSDACRRGPASG- -F - - - - - - - - FS - - - - RLNWL 170

>gi|15418693|gb|AAF91072.1| ; gi|24653390|ref|NP 524737.2|
gi|21627234|gb|AAF58373.2| CG5912-PA
gi|6760453|gb|AAF28358.1|AF223365 1 Length = 1678
Score = 31.6 bits (67), Expect = 4.0
Identities = 13/21 (61%), Positives = 15/21 (71%), Gaps = 3/21 (14%)
Query: 15 CPAFPACGPDE - - CSRSPASG 33
C AFPACGPD C+ +P SG
Sbjct: 1313 CGAFPACGPDHFTCA - APVSG 1332

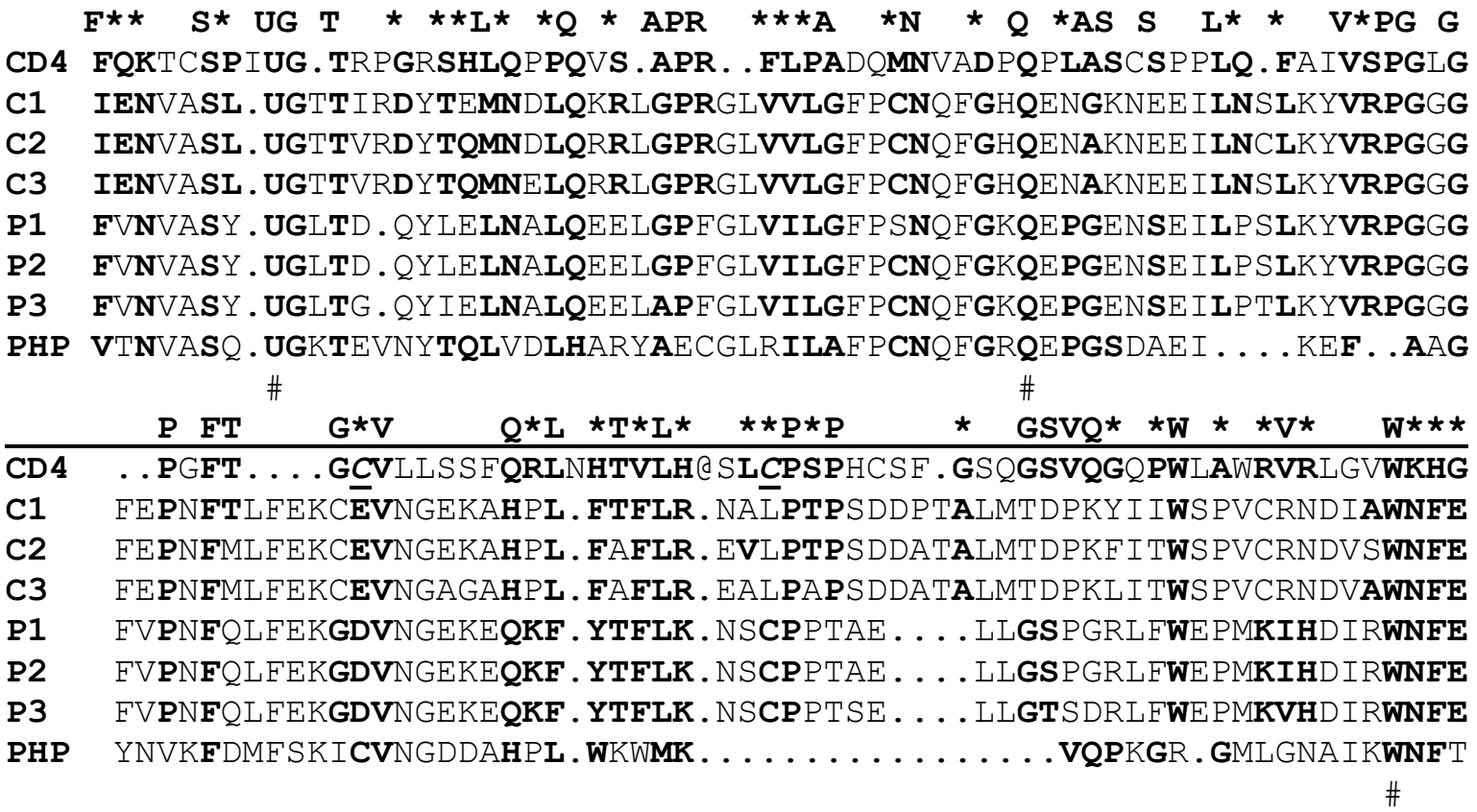
>gi|27676266|ref|XP 213869.1| ; gi|543730|sp|P35365|5H5B RAT
gi|310075|gb|AAA40616.1| Length = 370
Score = 31.2 bits (66), Expect = 5.3
Identities = 14/24 (58%), Positives = 18/24 (75%), Gaps = 3/24 (12%)
Query: 10 ASPGVCPAFPACGPDECSRSPASG 33
A+PG+ AFP GP+ CS SP+SG
Sbjct: 9 ATPGI - - AFPP - GPESCSDSPSSG 29

>gi|543453|pir||S38744 ; gi|435817|gb|AAB28595.1|
gi|737711|prf||1923268A Length = 369
Score = 31.2 bits (66), Expect = 5.3
Identities = 14/24 (58%), Positives = 18/24 (75%), Gaps = 3/24 (12%)
Query: 10 ASPGVCPAFPACGPDECSRSPASG 33
A+PG+ AFP GP+ CS SP+SG
Sbjct: 9 ATPGI - - AFPP - GPESCSDSPSSG 29

>gi|4768916|gb|AAD29679.1|AF133209 1 Length = 786
Score = 31.2 bits (66), Expect = 5.3
Identities = 13/27 (48%), Positives = 14/27 (51%), Gaps = 9/27 (33%)
Query: 11 SPGVCPAFPACGPDECSRSPASGLLFA 37
S GVC AF RSP G+LFA
Sbjct: 254 SSGVCAAF - - - - - - - - - RSPVGGVLFA 271

Figure 2.9. CD4 Tail, -1 Frame

“SLUG” motif commonly found in GPx genes. The sequence here is “SPIUG”. An alignment between the translated 3’-end of the CD4 gene including and past the first UGA codon, and including the +1 frameshift site, versus various glutathione peroxidase genes is shown in Figure 2.10. The sequences share several characteristics that are not expected by chance. First, the sequence containing the first UGA in CD4 is significantly homologous to the SLUG motif as mentioned previously. Also, the sequences share two totally conserved tryptophan residues. Tryptophan is the most unusual amino acid expected by chance, because there is only one codon for it. There are three primary active site regions covered in this alignment. The first is the “SLUG” motif site, which has already been mentioned as a site of potential sequence similarity. The second is partially made up of the sequence “CNQFGHQ”, of which the translated region in the 3’ end of the CD4 gene contains the asparagines and one of the glutamine residues. The third portion of the active site is centered on a conserved tryptophan, which is also found in the CD4 sequence. Especially significant is the fact that the residues found in the 3’ end of CD4 can be aligned to the same location where they are found in the other glutathione peroxidase genes without inserting large gaps or deleting large regions. The overall significance score of the alignment is 4.6 standard deviations above that expected by random chance using the parameters outlined in the methods section.



Types of GPx sequences: C1-C3 = cellular GPx;
P1-P3 = plasma GPx; PHP = phospholipid hydroperoxide GPx
IN CD4, @ = AISFSFKPSPSLII; ***C*** = **U** = potential SeCys
# indicates conserved residues of catalytic triad (U, Q, W)

Figure 2.10. Potential GPx gene encoded as an extension of CD4 protein

Conclusion

That the CD4 gene could be translated past the UGA codon seems plausible. Several significant sequence similarities were identified. Particularly significant are the sequence similarities identified between the translated sequence past the first UGA codon and BarH-like homeobox proteins, and the sequence similarity found between the translated sequence in the +1 reading frame past the second UGA codon and a hypothetical protein identified on chromosome 17. The detailed analysis of the similarities between the translated region past the UGA and several glutathione peroxidase genes uncovered striking similarities. In addition, the negative control in these experiments (the –1 reading frame analysis) failed to identify significant sequence similarities, supporting the positive results obtained in the other sequence sequence similarity studies. These results suggest that translation of the CD4 gene does not end at the UGA codon.

CHAPTER 3

SYNTHETIC PEPTIDE PREPARATION AND PURIFICATION

Introduction

The research plan is to use synthetic peptides as antigens for the production of antibodies against both the accepted cytoplasmic tail of CD4 and the region that is expressed if translation continues past the UGA codon. In order to further analyze the possibility that CD4 is translated past the UGA codon, as well as determine the best possible region of the sequence to mimic with the synthetic peptides, additional sequence analysis studies were undertaken.

At this point, a decision was made about the reading frame and general region that was to be targeted using monoclonal antibodies. Even though there is some evidence of a slippery sequence, and read through of the +1 frame past the second UGA (see sequence analysis section), in order for translation to occur in this region, presumably translation must first proceed through the first UGA and up to the slippery sequence associated with the second UGA. For this reason, further analysis of the peptide that may be produced after read through of the UGA codon will be limited to the zero frame in the region between the first and second UGA codons. The resultant synthetic peptides and monoclonal antibodies will then also be confined to this region.

Structural Analysis and Peptide Production

Introduction

In the interest of producing a good antigen for antibody production, stable structure is desirable in the region that is mimicked with a synthetic peptide structure. Not only will stable structure make the antigen more complex and perhaps more immunogenic, the three-dimensional epitope produced should make the resultant antibody more antigenically specific. Also, when using a small peptide sequence to mimic a larger structure, a structural epitope should be chosen to try and mimic as closely as possible part of the overall structure of the entire region. In addition, other special features of the potential polypeptide sequence such as phosphorylation and glycosylation sites are investigated.

Methods

The analyses conducted to identify special structural features of the polypeptide sequence past the UGA are done using software available online. The analyses are conducted on the translated CD4 0 frame past the UGA codon presumed to end the CD4 gene to the next stop codon, Figure 2.6. The analyses were conducted using the following software available on line: nnpredict software available at www.cmpharm.ucsf.edu/ggi-bin/nnpredict, PredictProtein software suite available at http://cubic.bioc.Columbia.edu/predictprotein, NetOGlyc 2.0 available at www.cbs.dtu.dk/htbin/nph-webface, PhosphoBase available at www.cbs.dtu.dk/htbin/pbase_predict.p1, NetPhos 2.0 available at http://genome.cbs.dtu.dk/htbin/nph-webface, SAPS software suite available at

www.isrec.isb-sib.ch/cgi-bin/SAPS, and GOR4 available at http://phil.ibcp.fr/cgi-bin/secpred_gor4.p1.

After carefully analyzing the results of the sequence and structural analyses, regions of the sequence are chosen to be mimicked by synthetic peptides that are ultimately used to develop monoclonal antibodies. One region is chosen that is before the first UGA codon, one region is chosen past the first UGA codon. Once identified, the peptides are ordered from the Molecular Genetics Instrumentation Facility (MGIF) at the University of Georgia.

Results and Discussion

Secondary Structure Analysis

The nnpredict software analysis predicted a strand associated with the lqfa sequence in the peptide, amino acids 40-43, Figure 3.1. A GOR4 analysis of the same peptide sequence also predicts an extended strand associated with the qfai sequence, amino acids 42-45.[75] In addition, the GOR4 analysis predicts an alpha helix composed of amino acids 22-25, the padq sequence. Although the nnpredict analysis does not show an alpha helix in the vicinity of amino acids 22-25, a careful analysis of this sequence in the nnpredict data reveal that although an alpha helix was not assigned to this region, the amino acids in this region were assigned a higher probability of being an alpha helix than any other structural motif.

Amino Acid Composition Analysis

This sequence of amino acids were also analyzed using two software suites designed to predict the characteristics of the polypeptide and determine if

Sequence *0fme1*:
ugtrpgrshlqppqvsaprflpadqmnvadpqplascspplqfaivspglgpgftg

Secondary structure prediction (H = helix, E = strand, - = no prediction):
----------EEEE----------

Figure 3.1. Results of nnpredict query

the sequence has a typical amino acid composition. The first analysis was done using SAPS. The sequence was too short for most SAPS analyses, but the amino acid composition was determined. The results are shown in Figure 3.2. Because of the short length of this query sequence, the composition as compared to typical human proteins may not be comparable. However, the only striking finding is that this sequence is made up of 20% prolines. In addition, the percentage of glycine in this sequence is 10.9%. The percentages of the other amino acids are less than 10%.

Motif and Secondary Structure Analysis

The sequence was also analyzed using the Predict Protein analysis software suite.[76] This analysis was conducted using a sequence of 85 amino acids corresponding to the region from the presumed end of the CD4 gene, through the next UGA codon assuming it is a frameshift site, and continuing in the +1 frame until the next stop codon is reached. This suite of programs conducted several detailed analyses of the amino acid sequence. A PROSITE motif search identified three potential N-myristoylation sites.[77] They are found at amino acid 2, GTRPGR; 49, GLGPGF; and 100, GVFRAS. (Figure 3.3). A ProDom domain search identified two protein domains that significantly resemble the query sequence.[78] These are sequence number PD138707 and PD037090, Figure 3.3. The probabilities associated with these sequences are 0.00011 and 0.00085 respectively, which is significant. According to a PHD analysis, the overall structure of this sequence can be classified as mixed.[79] Figure 3.3. PHD predicts the following structural characteristics for the protein

Number of Residues: 55 Molecular Weight: 5.7 kDaltons

1 GTRPGRSHLQ PPQVSAPRFL PADQMNVADP QPLASCSPPL QFAIVSPGLG PGFTG

Compositional Analysis (extremes relative to: HUMAN.q)

A : 5 (9.1%) ; C : 1 (1.8%); D : 2 (3.6%); E : 0 (0.0%); F : 3 (5.5%);
G : 6 (10.9%); H : 1 (1.8%); I : 1 (1.8%); K : 0 (0.0%); L : 5 (9.1%);
M : 1 (1.8%); N : 1 (1.8%); P :11(20.0%); Q : 5 (9.1%); R : 3 (5.5%);
S : 5 (9.1%); T : 2 (3.6%); V : 3 (5.5%); W : 0 (0.0%); Y : 0 (0.0%)

KR : 3 (5.5%); ED : 2 (3.6%); AGP : 22 (40.0%);
KRED : 5 (9.1%); KR-ED : 1 (1.8%); FIKMNY : 6 (10.9%);
LVIFM: 13 (23.6%); ST : 7 (12.7%)

Charge Distributional Analysis

1 00+00+0000 0000000+00 00-00000-0 0000000000 0000000000 0000

Figure 3.2. SAPS Amino Acid Analysis

PROSITE motif search

Pattern-ID: **MYRISTYL** PS00008 PDOC00008
Pattern-DE: N-myristoylation site
Pattern: G[^EDRKHPFYW] . {2} [STAGCN] [^P]
2 GTRPGR
49 GLGPGF
100 GVFRAS

ProDom domain search

	prot	(#)	ppOld.	default:	single …	score	P(N)	N	100.00%
1.	PD138707	p2000.1	(2)	O43177 (1)	O4343…	47	0.00011	3	28.4%
2.	PD037090	p2000.1	(2)	Rel (2)	/ / CRE…	52	0.00085	3	29.5%

PHD predictions for predict_h4587

- PHDsec summary overall your protein can be classified as: mixed given the following classes:
  - 'all-alpha': %H>45% AND %E<5%
  - 'all-beta': %H<5% AND %E>45%
  - 'alpha-beta': %H>30% AND %E>20%
  - 'mixed': all others

- Predicted secondary structure composition for your protein:
  %H: 0.0; %E: 11.7; %L: 88.3

```
XGTRPGRSHLQPPQVSAPRFLPADQMNVADPQPLASCSPPLQFA I VSPGLGPGFTGECCSLVSRGLITPSSTP
                                              EEE              EEEEE
```

GLOBE prediction of globularity

nexp = 74 (number of predicted exposed residues)
nfit = 56 (number of expected exposed residues)
diff = 18.00 (difference of nexp-nfit)
→ Your protein may be globular, but it is not as compact as a domain

Figure 3.3. Structure Analysis

sequence: Helix 0%, Extended strand 11.7%, Random Loop 88.3%. PHD predicts extended strands at both the FAI sequence identified previously, as well as the ECCSL sequence at amino acid 57-61. This region is just after the UGA/slippery sequence area. Much of the remainder of the sequence could be assigned neither as an alpha helix or a strand. A GLOBE prediction of globularity predicts that the polypeptide may be globular, but it is not as compact as a domain [80], Figure 3.3. The number of predicted exposed residues was 74 whereas the expected number of exposed residues was 56.

Phosphorylation Analysis

A phosphorylation site prediction was done using PhosphoBase using the same 111 amino acid sequence, Figure 3.4. Several potential phosphorylation sites were identified. The potential sites for protein kinase CKI, along with the consensus sequences after the sites are the following: S-63, SLVSRGL; T-72, TPSSTPFP; S-88, SLFLSDPL. The Potential site for protein kinase SSK3, along with the consensus sequence after the site is S-84, ALLSLFLS. The potential site for protein kinase PKA along with the consensus sequence after the site is S-105, RASPGW. Sites for other known protein kinases were not found. According to the authors of this software, the prediction of phosphorylation site is only a guide. Many false positive and false negative results are possible. A second search for potential phosphorylation sites was conducted using NetPhos 2.0. This analysis identified two potential serine phosphorylation sites, S-8, RPGRSHLQP, and S105, VFRASPGWH. Two potential sites of threonine phosphorylation were also identified. They are T-68,

XGTRPGRSHLQPPQVSAPRFLPADQMNVADPGPLASCSPPLQFAIVSPGLGPG
FTGECCSLVSRGLITPSSTPFPFPSSLALLSLFLSDPLPTAHLDPRGVFRASPGW
HGG

The prediction is based on the consensus sequence motifs. For the list of the used motifs see here.

The consensus sequence for 10 different protein kinases are scanned for potential phosphorylation sites.

No phosphorylation sites for protein kinase CaMII found.

Potential phosphorylation sites for protein kinase CKI (N-terminal S/T must be prephosphorylated):

- S-63 SLVSRGL
- T-72 TPSSTPFP
- S-88 SLFLSDPL

No phosphorylation sites for protein kinase CKII found.

Potential phosphorylation sites for protein kinase GSK3 (C-terminal +4 S must be phosphorylated):

- S-84 ALLSLFLS

No phosphorylation sites for protein kinase MLCK found.

No phosphorylation sites for protein kinase p34cdc2 found.

No phosphorylation sites for protein kinase p70s6k found.

Potential phosphorylation sites for protein kinase PKA:

- S-105 RASPGW

No phosphorylation sites for protein kinase PKC found.

No phosphorylation sites for protein kinase PKG found.

Figure 3.4. Phosphobase phosphorylation site prediction

RGLITPSST, and T-72, TPSSTPFPF. The scores associated with each site are shown in Figure 3.5. The closer the score is to 1, the more likely that the site is a bona fide phosphorylation site. The highest score was associated with S-105, and the score for this residue is 0.995. Only S-8 is found within the region between the first and second UGA codons, the remainder of the potential sites is found after the potential slippery sequence. The most likely site of phosphorylation, S-105, is found near the end of this polypeptide region.

Glycosylation Analysis

A search for potential glycosylation sites was done using NetOGlyc 2.0. Again, a few potential sites were identified. The results are summarized in Figure 3.6. There were two potential sites of threonine glycosylation found, T-68, and T-72. There were three potential sites of serine glycosylation identified, S-36, S-70, and S-71. The most likely site of glycosylation is T-72, associated with a potential of 0.9901. Because this region, if expressed would likely remain within the cell, glycosylation would not be expected.

Identification of Regions Used For Monoclonal Antibody Development

In order to choose the exact region to mimic with a peptide, the above analyses were taken into account as well as several other factors. The peptides will be 20mers, so a region of the polypeptide must be chosen so that there is as much structure as possible in a 20 amino acid region. In addition, solubility of the peptide must be considered because it must be made soluble in a mild buffer such as PBS in order to be injected into a mouse for antibody production. Therefore, a region with some hydrophilic residues must be chosen, as well as a

XGTRPGRSHLQPPQVSAPRFLPADQMNVADPQPLASCSPPLQFAIVSPGLGPG
FTGECCSLVSRGLITPSSTPFPFPSSLALLSLFLSDPLPTAHLDPRGVFRASPGW
HGG
…….S……..T…T……..S
……

Serine predictions

Name	Pos.	Content	Score	Pred.
0fme1fme	8	RPGRSHLQP	0.834	*S*
0fme1fme	16	PPQVSAPRF	0.024	.
0fme1fme	36	QPLASCSPP	0.121	.
0fme1fme	38	LASCSPPLQ	0.030	.
0fme1fme	47	FAIVSPGLG	0.025	.
0fme1fme	60	GECCSLVSR	0.004	.
0fme1fme	63	CSLVSRGLI	0.008	.
0fme1fme	70	LITPSSTPF	0.032	.
0fme1fme	71	ITPSSTPFP	0.097	.
0fme1fme	78	FPFPSSLAL	0.004	.
0fme1fme	79	PFPSSLALL	0.009	.
0fme1fme	84	LALLSLFLS	0.015	.
0fme1fme	88	SLFLSDPLP	0.004	.
0fme1fme	105	VFRASPGWH	0.995	*S*

Threonine predictions

Name	Pos.	Context	Score	Pred.
0fme1fme	3	--XGTRPGR	0.092	.
0fme1fme	55	GPGFTGECC	0.036	.
0fme1fme	68	RGLITPSST	0.808	*T*
0fme1fme	72	TPSSTPFPF	0.799	*T*
0fme1fme	93	DPLPTAHLD	0.110	.

Figure 3.5. NetPhos 2.0 Prediction Results

XGTRPGRSHLQPPQVSAPRFLPADQMNVADPQPLASCSPPLQFAIVSPG
LGPFTGECCSLVSRGLITPSSTPFPFPSSLALLSLFLSDPLPTAHLDPRGV 80
FRASPGWHGG
……..S……T.SST…… 80
……

Name	Residue	No.	Potential	Threshold	Assignment
0fme1fme	Thr	3	0.0102	0.4967	.
0fme1fme	Thr	55	0.0097	0.6311	.
0fme1fme	Thr	68	0.7430	0.6053	T
0fme1fme	Thr	72	0.9901	0.5983	T
0fme1fme	Thr	93	0.3901	0.5775	.

Name	Residue	No.	Potential	Threshold	Assignment
0fme1fme	Ser	8	0.4047	0.5490	.
0fme1fme	Ser	16	0.5524	0.6094	.
0fme1fme	Ser	36	0.6241	0.5942	S
0fme1fme	Ser	38	0.0470	0.6023	.
0fme1fme	Ser	47	0.4301	0.6538	.
0fme1fme	Ser	60	0.0018	0.6709	.
0fme1fme	Ser	63	0.0009	0.6072	.
0fme1fme	Ser	70	0.6652	0.5625	S
0fme1fme	Ser	71	0.7518	0.5683	S
0fme1fme	Ser	78	0.3019	0.5913	.
0fme1fme	Ser	79	0.0431	0.6411	.
0fme1fme	Ser	84	0.0041	0.7059	.
0fme1fme	Ser	88	0.0491	0.6069	.
0fme1fme	Ser	105	0.0005	0.6160	.

Figure 3.6. NetOGlyc 2.0 Prediction Results

region that does not have an extreme PI. In addition, when choosing the region to be used within the known cytoplasmic tail of CD4, the transmembrane segment must be avoided for the region to be accessible to the monoclonal antibodies. The peptides must be large enough to be sufficiently antigenic to cause the mice to produce antibodies against them. The largest size peptides that MGIF will produce without a large increase in price are 20mers. The synthetic procedure and risk of error also grows with the increase in size. Because of this, 20 amino acid long peptides will be made. These were thought to be large enough to be antigenic. With these considerations in mind, the following regions were chosen to mimic with synthetic peptides: amino acids 18-38 past the UGA that may end CD4, and the last 20 amino acids of the accepted CD4 cytoplasmic tail. The peptide chosen for the region past the UGA includes the predicted alpha helix sequence PADQ. The sequences chosen are shown in Figure 3.7. The predicted molecular weight of the peptide for the known region of cytoplasmic tail is 2334.9 Daltons. The predicted molecular weight of the peptide for the region past the UGA codon is 2160.6 Daltons. The peptide for the known cytoplasmic tail of CD4 has several more hydrophilic amino acids than the peptide for the region past the UGA. The peptide for the known cytoplasmic tail has 15 hydrophillic residues whereas the peptide for the region past the UGA only has 10 hydrophillic residues. This is because I was unable to find a sequence of 20 amino acids past the UGA that had as many hydrophilic amino acids as that of the known cytoplasmic tail peptide. The ends of the peptides can be modified so that they better mimic the protein region they imitate. The C-

Peptide “024”, From The Region Just Prior to The First UGA Codon; Includes Part of “SLUG” Motif

LSEKKTCQCPHRFQKTCSPI

Peptide “025”, From The Region Past The First UGA Codon

RFLPADQMNVADPQPLASCS

Figure 3.7. Peptide Sequences

terminal end of the peptide that imitates the known cytoplasmic tail was left as a carboxylate because if translation ends at the UGA, then the end of the peptide will be most similar to this region in the cell. The C-terminal end of the peptide past the UGA codon was modified as an amide, because if translation continues into this region, then translation will continue past this region and there will be no free c-terminal there in the cell.

Conclusion

The peptides were ordered on the 0.02 M scale. When ordered, the peptides were given the designations 024 for the peptide imitating the known cytoplasmic tail, and 025 for the peptide imitating the region past the UGA codon. The first attempt at producing the 025 peptide failed, and MGIF had to repeat its synthesis. MGIF was able to produce 29 mg of 024 (lyophilized weight), and ultimately 17 mg of 025 (lyophilized weight). Both peptides were analyzed by MGIF using reverse phase HPLC and MALDI mass spectrometry. This is to verify their identity and purity. The MGIF HPLC analysis of peptide 024 is shown in Figure 3.8, and the MGIF HPLC analysis of peptide 025 is shown in Figure 3.9. Neither peptide is totally pure. Both the HPLC and MALDI mass spectrometry analysis show that both peptides are contaminated with other substances, probably pieces of the desired peptides as well as peptide pieces with erroneous sequences. This seems to be particularly true for 025. Given the already reduced amount of 025 available, more of this peptide will have to be ordered later.

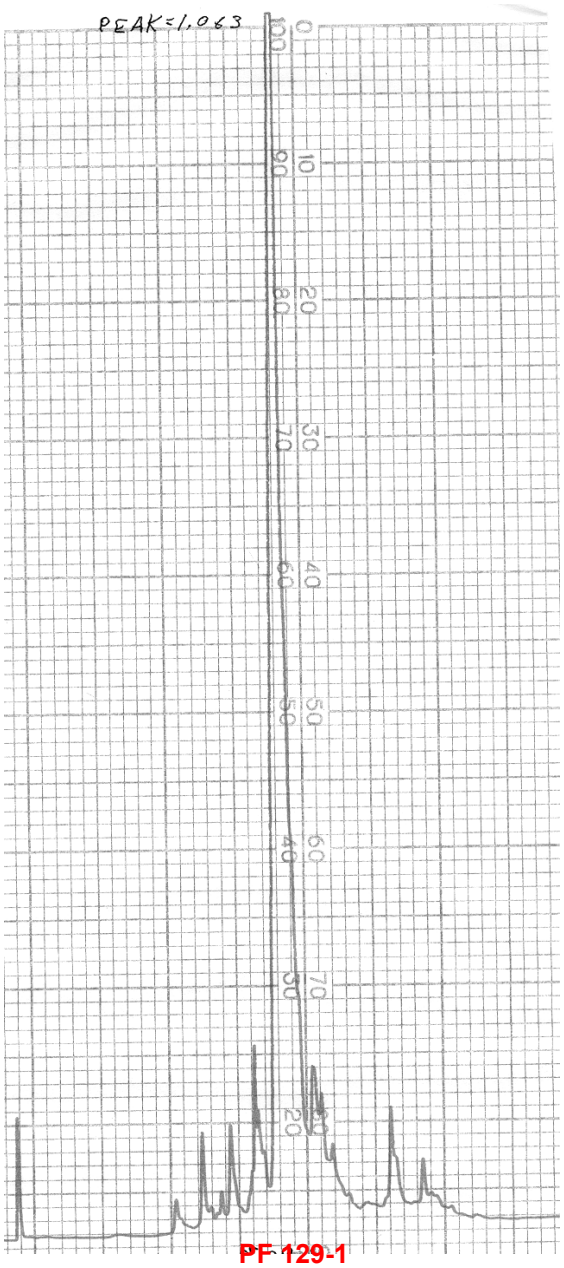


Figure 3.8. 024 Peptide Purity, C-18 Reverse Phase

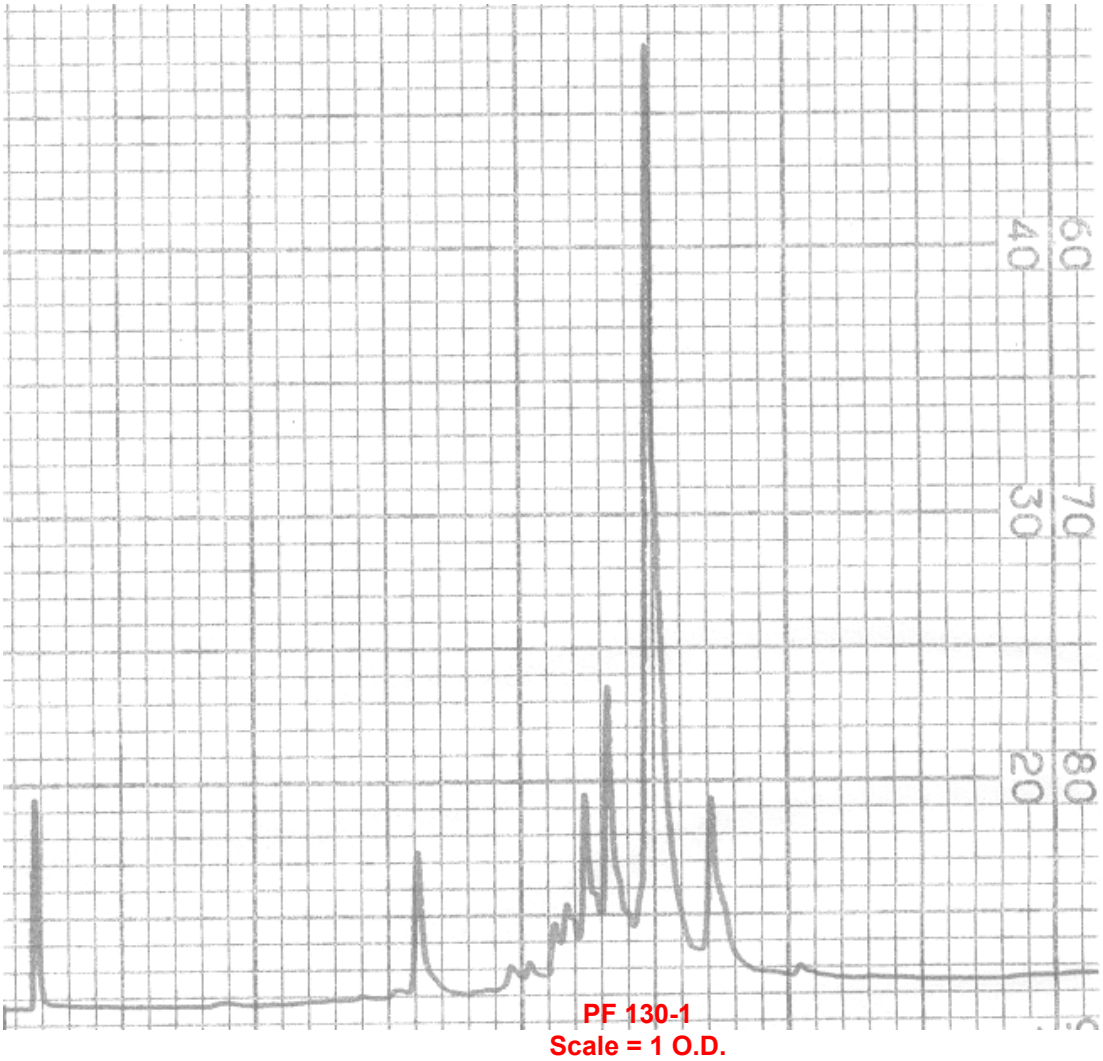


Figure 3.9. 025 Peptide Purity, C-18 Reverse Phase

Peptide Purification

Introduction:

Although absolute purity is not attainable, the highest attainable purity peptide should be used. Because these peptides are used to produce monoclonal antibodies, any impurity has the potential to ruin the resultant antibody. Particularly if the impurity is somehow more immunogenic than the true peptide, then the antibodies isolated have the potential to react to an epitope that is not present in the native protein. There would be no way to verify this problem, except that the antibody would not bind to the native region. Since one of these antibodies will be used to label a region that may or may not be produced in nature, any problem that could cause it to not bind to this native region even if present, must be minimized as much as possible. For these reasons, purifying the peptides further is attempted.

Monoclonal antibody development requires at least 500ul of 2mg/ml peptide in PBS.[81] Because milligram quantities of the peptides are required, purification using analytical C-18 reverse phase HPLC is impractical. Therefore, a different purification strategy was developed. The two most obvious properties of each peptide that can easily be exploited for a purification strategy are molecular weight and charge. 024 is 2335 Daltons, 025 is 2161 Daltons. The calculated pI of 024 is 8.825, with nine ionizable groups. The calculated pI of 025 is 6.64, with six ionizable groups. Pharmacia HiTrap desalting columns filled with Sephadex G25 are used to purify the peptides based on molecular weight. Pharmacia sulpho-propyl (SP) ion exchange columns are used to purify the

peptides based upon molecular charge. In order for the peptides to interact with the negatively charged SP column, the pH of the solution they are in must be at least 1 pH unit below their respective pI.[82] For 024, the pH of the buffer is 7.0, for 025, the pH of the buffer is 5.5. A pH gradient is used to elute the peptides from the column. Detecting the peptides is not straightforward. Because of their very small size, they do not have a large number of each amino acid. In particular, they do not have enough of the aromatic amino acids for detection at 280nm to be an option. Instead, detection of the peptide bond at 220nm must be done, even though it is not very specific for peptides. Concentrating the peptides is also a significant challenge. Convenient centrifuge filters for concentrating molecules or changing buffers are only available in sizes down to a 3000 molecular weight cutoff. This is well above the molecular weight of each peptide, so they would pass through the filter and be lost. In order to concentrate the peptides, the samples will be centrifuged under vacuum in order to evaporate the solvent. This process also concentrates any salt in the samples; therefore minimal salt concentration must be used.

Methods

pH Gradient Ion Exchange: 024 [82]

024 is dissolved in 1ml phosphate pH 7.0 (all 29 mg). The method for the ion exchange of 024 is as follows. A 5ml SP column available from Pharmacia is used. A Hitachi 6200 gradient HPLC pump is used along with an L-4000 UV detector, and a D-2000 integrator. A 2.5ml/min flow rate of 50mM phosphate pH 7.0 is used to rinse the column after sample injection for 10

minutes (a total of 5 column volumes). After 10 minutes, a gradient of 50mM phosphate pH 10.5 is introduced from 0-100% over 15 minutes (7.5 column volumes). This elutes the peptide from the column. The mobile phase is then returned to 50mM phosphate pH 7.0 over five column volumes (10 minutes). The absorbance is monitored at 220nm.

pH Gradient Ion Exchange: 025 [82]

Another Pharmacia SP column is used. 50mM MES pH 5.0 starting buffer pumped through column for 20 minutes. Then a gradient to 100% 50mM TRIS pH 8.5 is introduced over 15 minutes. Flow rate is 2.5ml/min. Absorbance is monitored at 220nm.

Size Exclusion Chromatography [83]

The final product must be about 2mg/ml in concentration, in a biologically compatible buffer such as PBS. However, the purification strategy dilutes the sample significantly. After concentrating the sample peaks many times using the speedvac, the salt concentration of the original mobile phase will be multiplied many times. Therefore, minimal salt is maintained in the mobile phase buffer. PBS has a total salt concentration of 150mM. PBS diluted 15X is used as the mobile phase. The result is a total salt concentration of 10mM, which is the minimum amount that can be used and still avoid ionic interactions with the column. Also, the sample may be concentrated 15X and still be in 1X PBS. The pH of the dilute PBS is also raised to about 8.0, which is near to the pI of 024. This will reduce the number of positive charges on the peptide, which will prevent the ionic interactions with the Sephadex. The eluent is monitored for the

peptides at 220nm. A 5ml Pharmacia Sephadex G25 desalting column is used. PBS diluted 1:15; pH 8.0 is the mobile phase. The flow rate is 2.5ml, and the method is run for 50 minutes in order to insure that everything is off of the column.

Sample Concentration

The collected peak samples are concentrated in a DNA speedvac centrifuge by evaporation under vacuum. The concentrated samples are pooled and concentrated repeatedly until the entire sample is contained in as small a volume as practical.

Peptide Concentration Assay

The amount wasted during the assay must be minimized. The small size of the peptide is problematic. The Lowry method is not used because the Folin reagent contained in the Lowry reacts with tryptophan and tyrosine.[84] There is not a representative amount of these amino acids in the peptide, so it would appear much more dilute than it actually is when compared to some standard protein such as BSA. The Biuret method is not sensitive enough. The Bradford method, which interacts with positive charges on proteins, is used instead. Because each peptide has a number of positively charged residues, there should be enough positive charges on the peptide for the reagent to interact with.

Results and Discussion

Purification of 024, pH Gradient Ion Exchange

Because of the problems that may be encountered because of high salt concentrations, a pH gradient was used to elute the sample from the column. This method utilizes a 50mM phosphate pH 10.5 gradient. A representative chromatogram is shown in Figure 3.10. Several peaks are visible in this chromatogram, although baseline separation was not achieved. Each peak is collected and concentrated using the methods outlined above. The purified peptide should be contained in the largest peak. This is based on the fact that the largest peak in the chromatogram that MGIF produced was tested and found to be the peptide. All peaks will be analyzed.

Sample Concentration

The collected peaks are concentrated using the “DNA speedvac” using the method outlined above. This process takes between 12 and 24hours. The original volume of the major peak in 0244 is 4.5 ml. The final volume is about 115ul, which is nearly a 45-fold concentration. Even though a pH gradient was utilized minimizing the salt concentration, the phosphate concentration could still reach 2.25M, which explains why some salt crystal formation was observed. Instead of trying to dissolve all of this phosphate, the saturated liquid containing the sample is removed from the salt crystals and diluted to 130ul with water. This will prevent the sample from being totally saturated with salt when it is injected onto the G25 column.

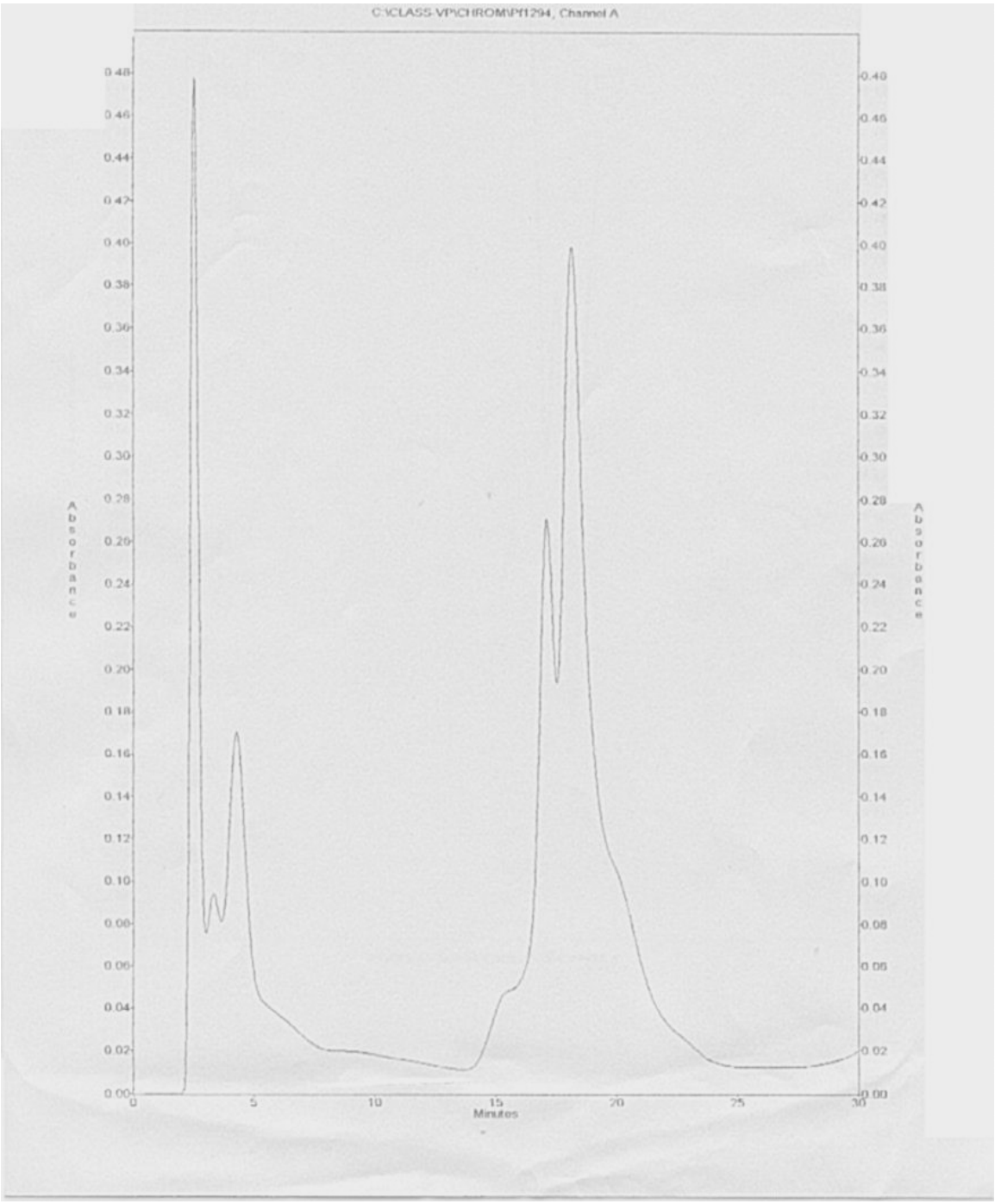


Figure 3.10. pH Gradient Purification of 024

Size Exclusion Chromatography

The concentrated sample corresponding to the largest peak from run 0244 is further purified using the 5ml Pharmacia Sephadex G25 desalting column. The largest peak should be purified 024 as this is the peak identified by MALDI mass spectrometry analysis as 024 in the chromatogram attained by MGIF. The entire sample is injected onto the column, and the results are shown in Figure 3.11. The major peak was collected in a total volume of 7.47ml. It is concentrated down to 500ul using the Speedvac, which is a 15 fold concentration.

Conclusion

The concentrate was assayed for protein concentration, and is only 0.71mg/ml. More of the 024 peptide is concentrated from the major peaks of other ionic exchange separations, separated on the G-25 column, and concentrated again in the same manner as before. The concentrated samples are analyzed using MALDI mass spectrometry analysis. Nothing identifiable as the peptides is in the mass spectra of any of the samples. The salt concentration could still be too high for this method. LC-MS analysis will be done on all samples after the 025 peptide is purified.

Results and Discussion

Purification of 025, pH Gradient Ion Exchange

025 is dissolved into 500ul water plus 250ul of 50mM MES pH 5.0. As before, the first step is to separate the peptide on a SP ion exchange column, and the second step is the buffer change and further purification using a G-25

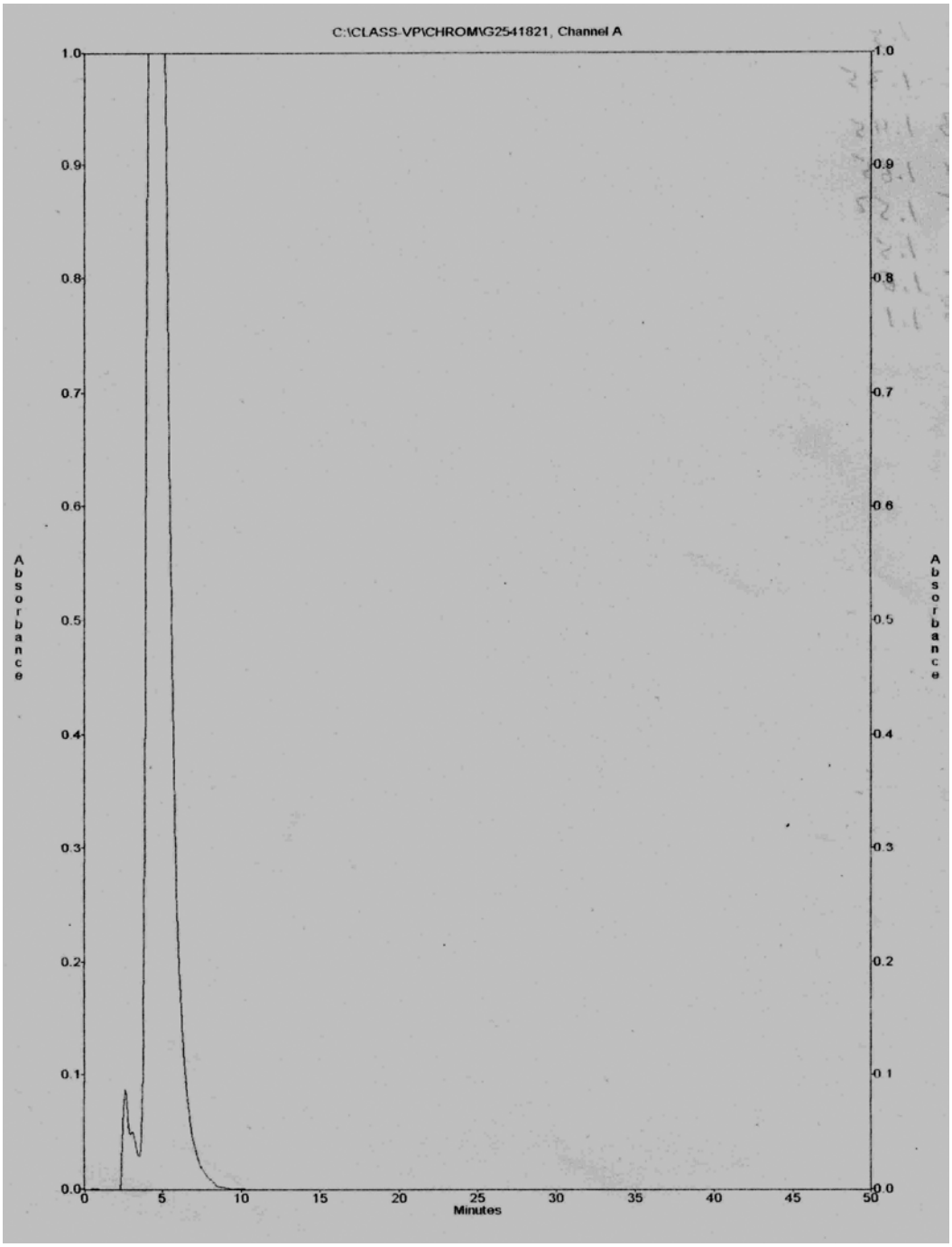


Figure 3.11. Size Exclusion Purification of 024

size exclusion column. The PI of 025 is calculated to be 6.64. The pH of the starting buffer is 5.0. 50mM MES is the elution buffer at pH 8.5. 250ul of the 025 concentrate is separated on the SP column using this method. During the MES gradient, a large peak elutes and is collected in 31 1.5ml tubes (total volume about 46.5 ml). The chromatogram is shown in Figure 3.12. The samples are concentrated in the speedvac for many hours until the total volume is reduced to about 700ul.

Size Exclusion Chromatography

The method is identical to the size exclusion method used for 024 purification with some changes. Two 5ml G-25 columns are used in series with a flow rate of 2.5ml/min. Water is used instead of dilute PBS in order to try and avoid increased salt concentration in the final product. The sample concentrate of the sample collected during the ion exchange chromatography (approximately 700ul) is fractionated using this method, and the peaks are collected. The chromatogram is shown in Figure 3.13.

Another run was conducted using a water to PBS gradient over 19 minutes in order to insure that the peptide did not remain attached to the column in the absence of enough salt. A peak did elute from the column, Figure 3.14, but this also seems too small to account for a meaningful proportion of the substance isolated in the large peak in Figure 3.12. The peaks obtained by cleaning the G-25 column with increased salt concentrations are also collected.

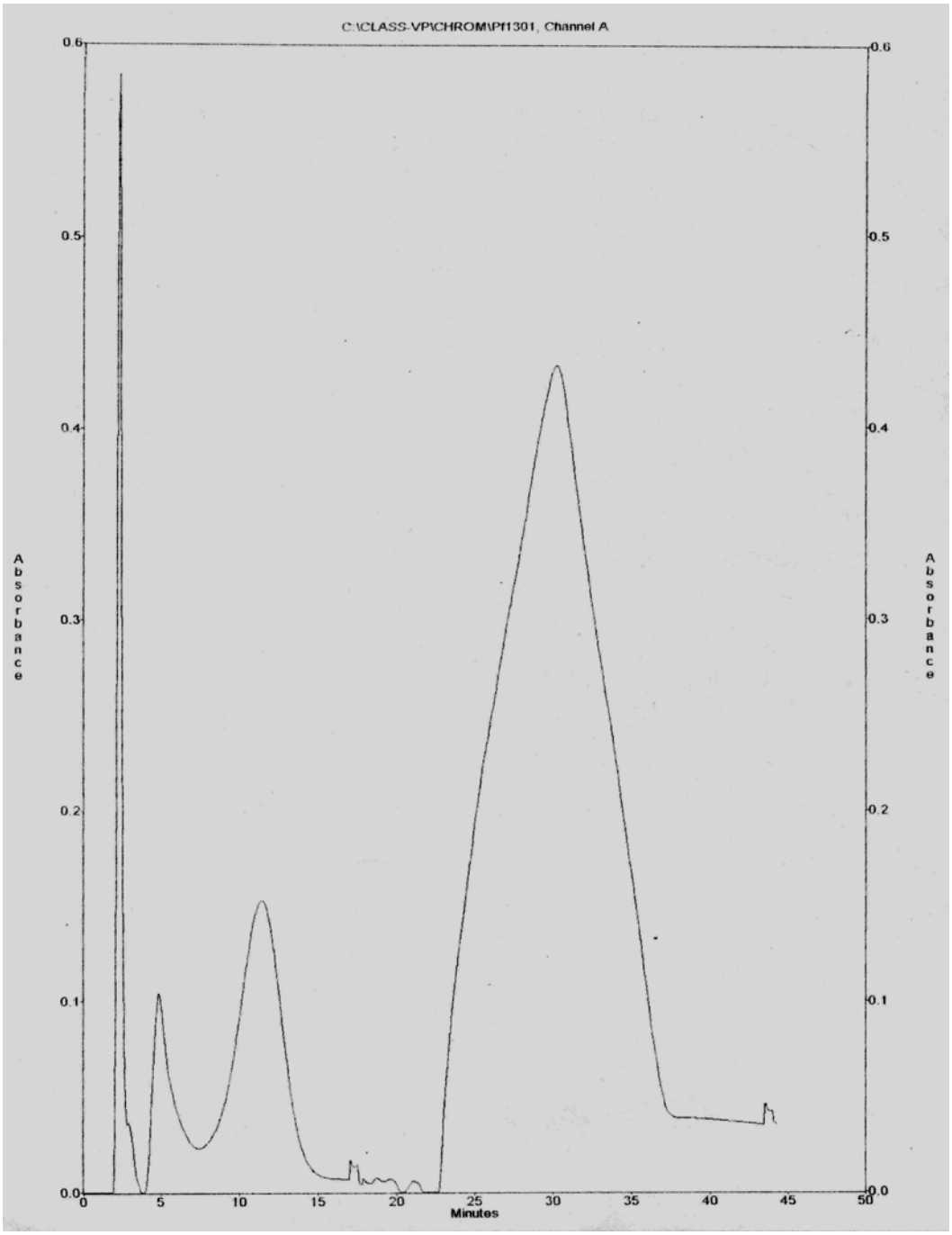


Figure 3.12. pH Gradient Purification of 025

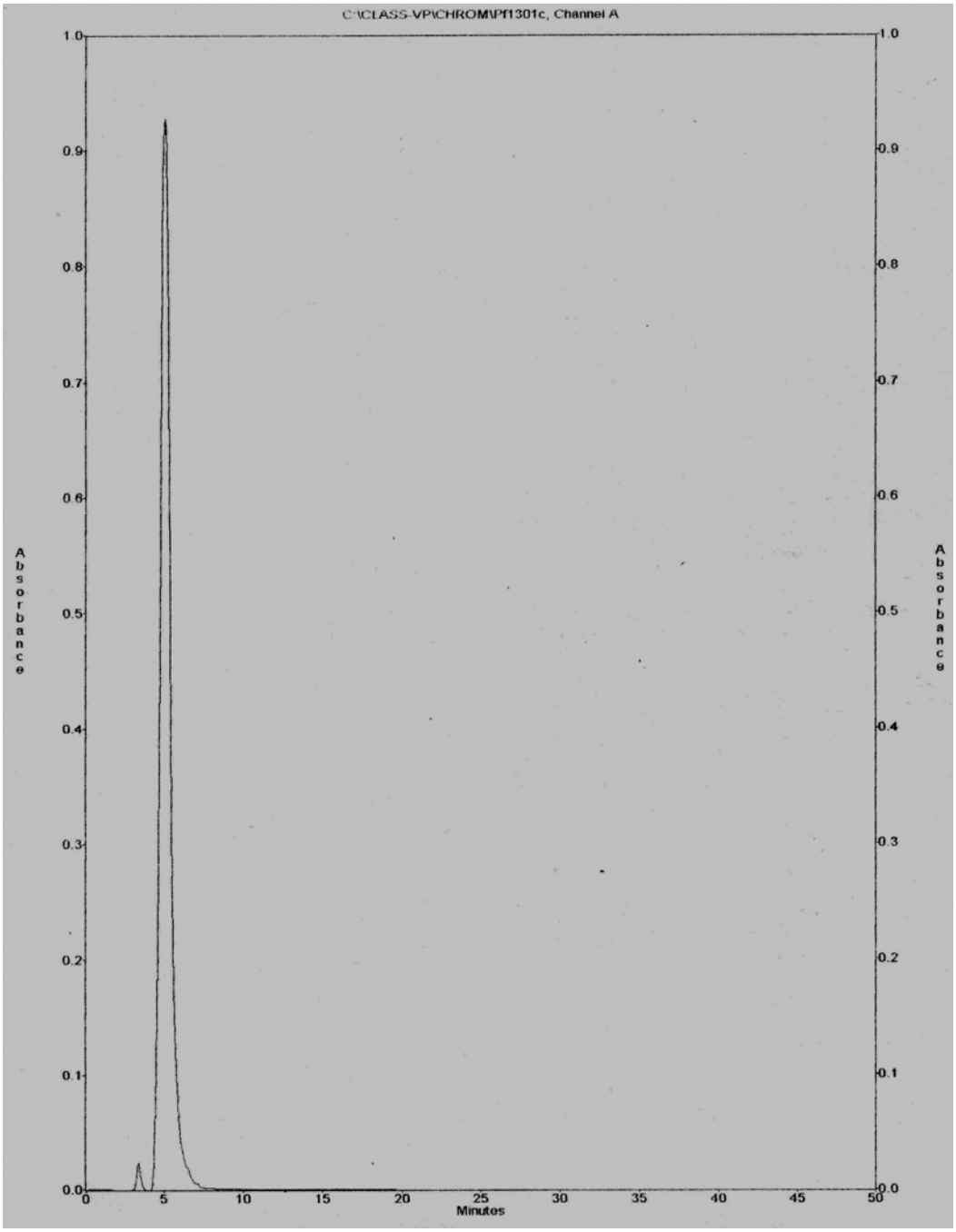


Figure 3.13. Size Exclusion Purification of 025

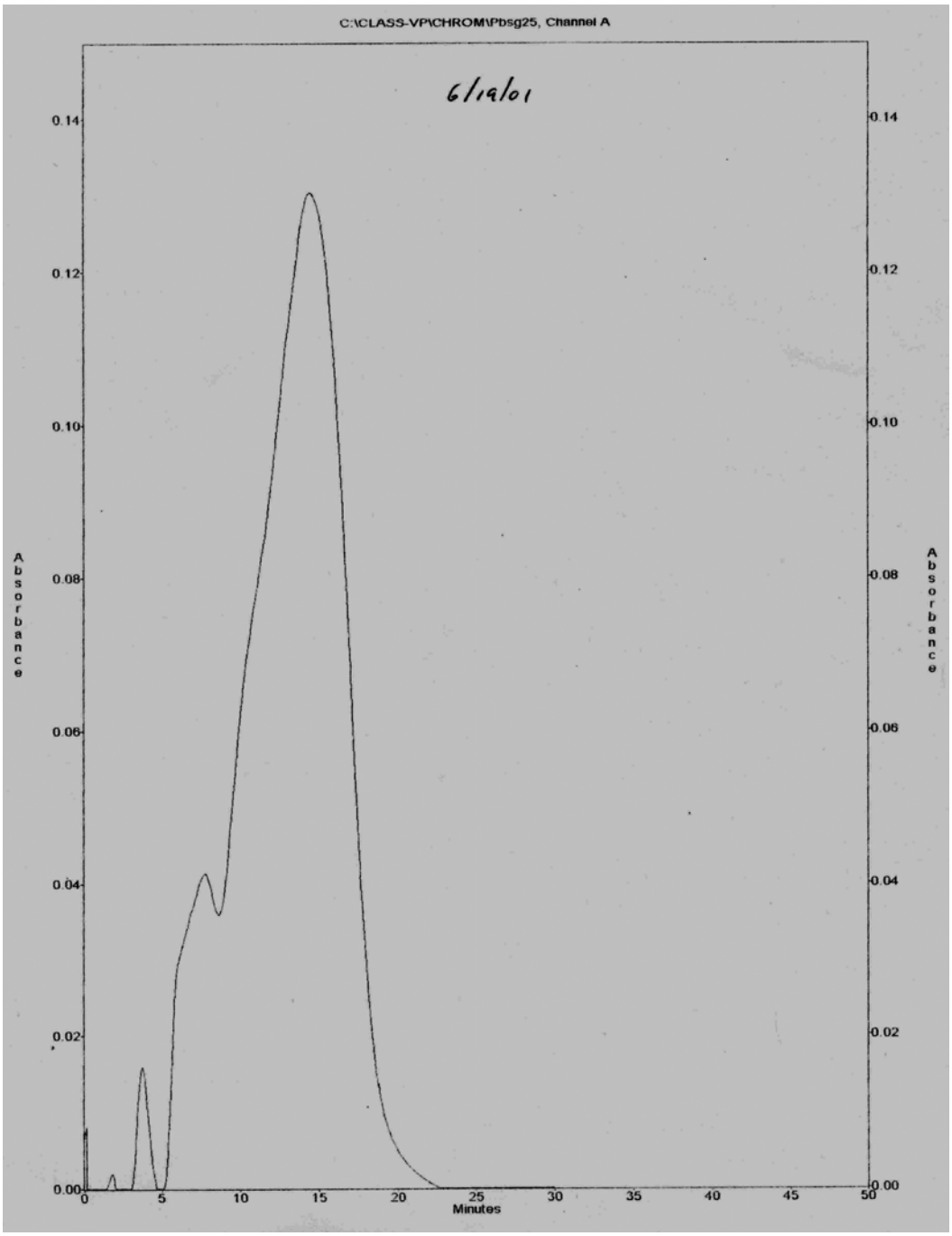


Figure 3.14. PBS Gradient on Size Exclusion Column

Sample Concentration

The peak samples are concentrated as before using a DNA speedvac. After being concentrated using the speedvac, they are set aside for verification of their identity using mass spectrometry.

Sample Analysis

Liquid chromatography-mass spectrometric analysis is carried out on all of the concentrated large peak samples from both the 024 purification and the 025 purification. A HPLC is connected in tandem with the mass spectrometer so that as peaks are separated on the HPLC, they are injected into the mass spectrometer for mass analysis. The samples are separated using a Phenomenex Bondclone 300mmX3.9mm C-18 reverse phase HPLC column at 0.5ml/min. The starting solvent is water. A gradient of 80% acetonitrile is used to elute the samples from the column. Three 10ul injections are made of the purified 024 sample using the autosampler attached to the HPLC. However, the autoinjector failed to pickup sample from the sample tube so a manual injector was installed.

In order to try and avoid further potential problems with the analysis, two standards were analysed in an effort to validate the ability of the instrument to identify peptides. 75ug/ml B endorphin, and 80ug/ml insulin are prepared as standards in water with a small amount of TFA to facilitate dissolution. The samples were manually injected in order to avoid problems with the autosampler. Several injections of the B endorphin were done, and peaks identified on the LC chromatograms. However, B endorphin could not be identified in the mass

spectra. The only thing resolved in the mass spectra was several different molecular weight sizes of polyethylene glycol polymer. The origin of this substance is unknown. Several injections of the standard insulin solution were also done, and peaks were found on the LC chromatogram. Again, though, no insulin was identified in the mass spectra. Another 025 sample is injected, this time manually, and peaks were found on the LC chromatogram. One peak was identified as a polymer of MES. However, no positive identification of the 025 or 024 peptides could be made. However, a very weak 1080 Dalton peak, which corresponds to one of the predicted fragment sizes of the 025 peptide, was found.

Conclusion

The inability to identify the peptides could have been caused by salt in the samples when using the MALDI analysis, although the LC-MS method is not nearly as sensitive to salt contamination. However, adjusting the mass spectrometer to analyze peptide sized masses is much more challenging using LC-MS than with MALDI. This is probably where the problem is, as supported by the inability to identify either B endorphin or insulin.

Several more aliquots of the 024 and 025 peptides were purified using the methods outlined above. However, several things became clear. First, there is no acceptable method to verify what has been purified. Salt in the samples ruins MALDI mass spectrometric analysis, and adjusting the LC-MS to visualize the peptides could be difficult even if time on the instrument were easy to get. Second, the differences between the pure peptide and the contaminants could be

very small. Gel electrophoresis is not an acceptable option because it is unlikely to be able to resolve these small differences, perhaps even single or double amino acid deletions/ additions/ mistakes. Mass spectrometric analysis is also the only method that may be able to accurately analyze the mass quickly without wasting a large amount of the already scarce sample. Third, after conversing with the Monoclonal Antibody Facility in the Vet School, it is also clear that in order to insure that the peptides are sufficiently immunogenic they will likely have to be conjugated to something larger such as BSA.[81] This may also necessitate another round of purification and the problems that go along with it. Even though the peptides are theoretically large enough at about 2000 Daltons each, there are no guarantees because of differences in folding, epitope presentation, complexity, etc. There will be no way to verify the antigenicity of the peptides until months of work are done which would put the project further behind if they are not. These considerations make it unacceptable to proceed with these peptides.

MAP Peptides Are The Solution

A new type of peptide is available from MGIF that will solve some of these problems. It is called a multiple antigenic peptide (MAP), and is four or eight identical peptide chains attached to a central poly-lysine backbone. The peptides are synthesized using the poly-lysine backbone as the base. The peptides are re-synthesized as a four chain MAP peptides. The data from their synthesis is shown in Figure 3.15 and Figure 3.16. The peptide corresponding to

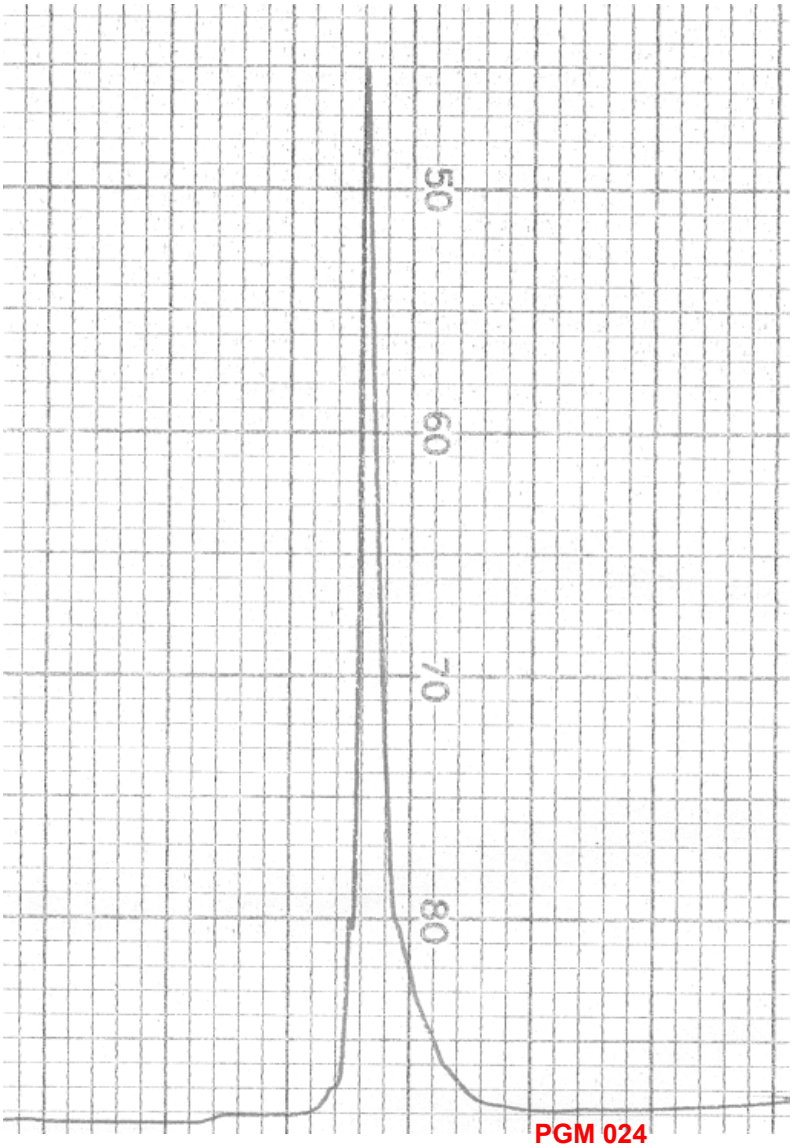


Figure 3.15. 024 MAP Peptide Purity, C-18 Reverse Phase

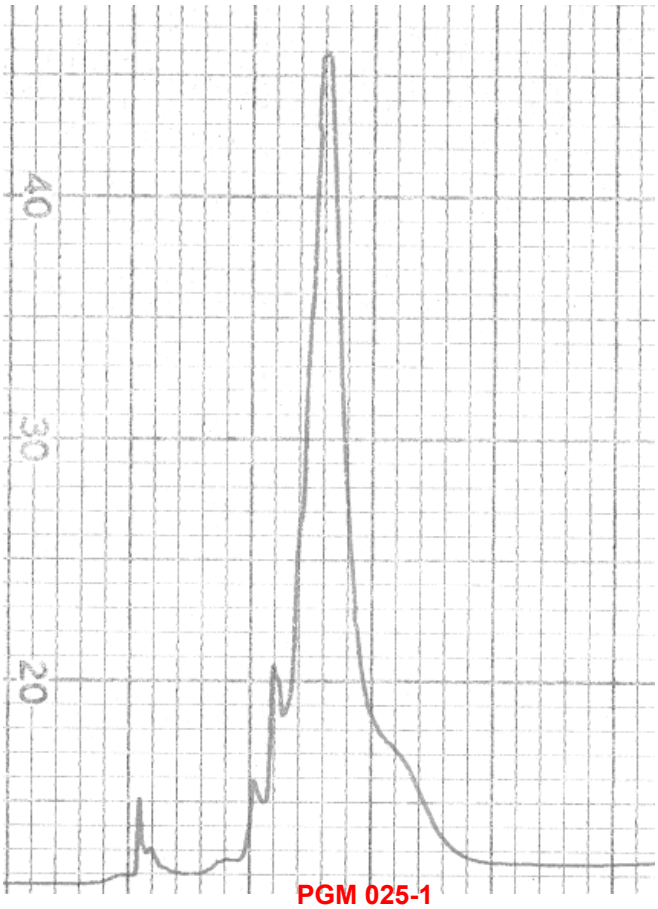


Figure 3.16. 025 MAP Peptide Purity, C-18 Reverse Phase

the peptide for the region before the UGA codon is designated PGM-024 or just 024. The peptide corresponding to the region past the first UGA is PGM-025 or just 025. The synthesis of both peptides is problematic this time. The reported yield is less than 10mg for each, and MGIF repeats their synthesis in order to prepare more. The purity of the peptides appears to be more homogeneous this time than before. In addition, the fact that there are four peptides linked together now

makes it more difficult to isolate only those molecules that are comprised of four identical, perfect peptide chains. For these reasons, as well as the problems associated with purifying the peptides encountered before, additional purification of the MAP peptides is not attempted.

The peptides are weighed this time instead of relying on the weight reported by MGIF. MGIF reported that they were giving me 6mg of 024 and 7mg 025. After carefully weighing the lyophilized peptides, it is clear that the weight reported by MGIF is overstated. The 024-peptide preparation weighs 1.8mg and the 025-peptide preparation weighs 5.2mg. The peptides are dissolved in PBS in order to achieve near a 2mg/ml concentration. 024 is dissolved in 2.6ml, 025 is dissolved in 2.0ml. They are dissolved in PBS because they are used directly to immunize mice, and 2mg/ml is the preferred concentration.

CHAPTER 4

MONOCLONAL ANTIBODY DEVELOPMENT AND PREPARATION

Introduction

This portion of research was done in collaboration with the UGA Monoclonal Antibody Facility in the veterinary school. In particular, Ruth Davis performed the work that was done directly with the mice. An overview of the monoclonal antibody preparation procedure is shown in Figure 4.1. The mice are first immunized against the antigen to which monoclonal antibodies are being prepared against. The antibody response of each mouse is measured using ELISA screening. The absorbance of the solution is a direct result of the amount of alkaline phosphatase conjugated secondary antibody bound to the plate, which in turn is linked to the amount of primary antibody bound to antigen in the well. It is the primary antibody that binds directly to antigen that is of interest, and developing monoclonal primary antibodies that bind to the synthetic peptides is the primary goal of this portion of research. After some time and boosting, B cells are isolated from the spleens of the mice, which are responding to the antigen. The B cells are fused with an immortal hybridoma cell line producing immortalized cells that secrete the antibodies of interest. The hybridoma cells are divided into several cultures and they are each screened for antibody production by testing an aliquot of the media they are grown in for antibody activity using ELISA. If antibodies are detected, the culture is diluted serially and used to seed many new cultures. The new cultures are screened and the dilution

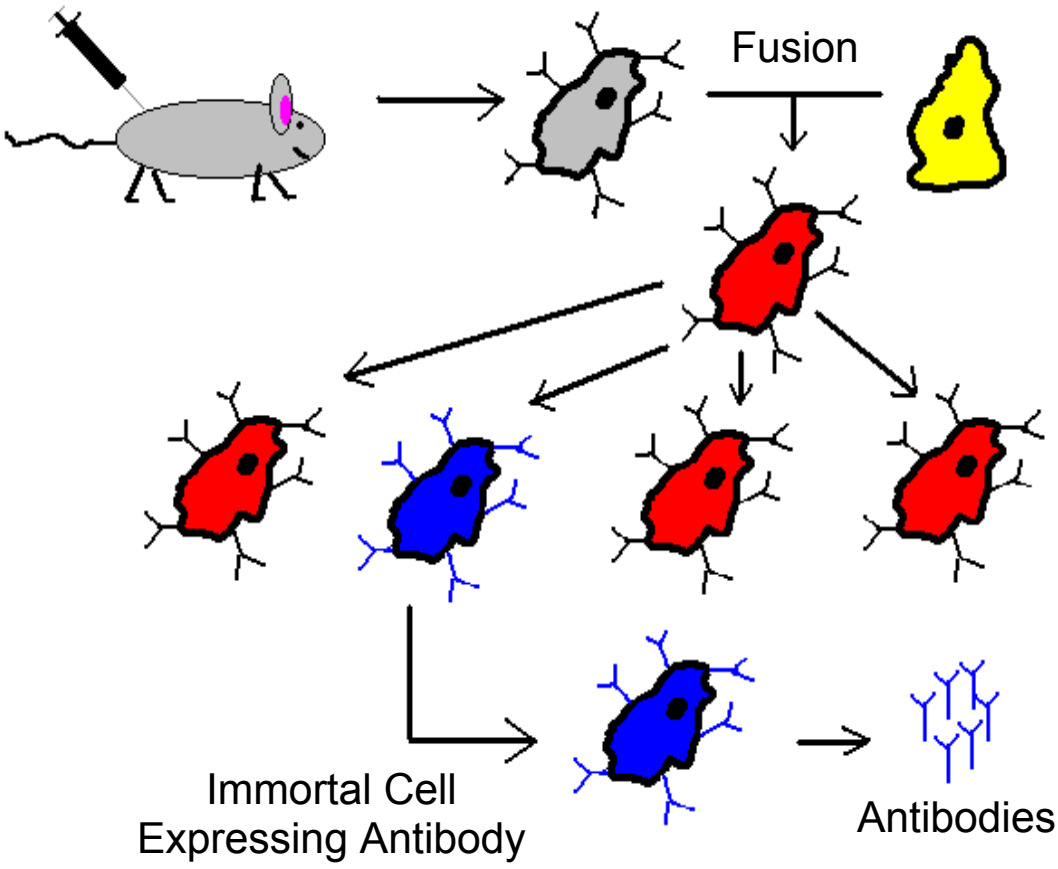


Figure 4.1. Monoclonal Antibody Development

process repeated if antibody activity is detected. Eventually, after repeating the process many times, the dilutions and seedings result in a monoclonal culture producing the antibody of interest. The cells can then be grown when needed and the antibody isolated from the culture media.

Materials and Methods

Most of the methods used in this portion of research that are not described herein are typical for monoclonal antibody preparation. They can be found in the book “Antibodies - a Laboratory Manual" by Ed Harlow and David Lane, published by Cold Spring Harbor Laboratory in 1988.[85]

Screening of Immunization and Boosting

The 96 well plates are coated with antigen by placing 50ul of a solution of 10ug/ml of the appropriate peptide antigen into the wells. When an entire plate is not needed for each antigen, some of the wells are coated with one antigen; others are coated with the other antigen. After adding antigen, the plate is gently agitated to insure that the antigen solution completely covers the bottom of each well. After incubating in the refrigerator overnight, the antigen solution is aspirated from each well, the wells are fully rinsed with PBS, and 200ul 3% BSA 0.2% Tween 20 is added to each well. The plate is then incubated at 37C overnight. After removing the BSA blocking solution and rinsing the wells with PBS, 50ul of the proper sera sample is added to the appropriate well. The plate was incubated for one hour, covered, at 37C. The sera samples were then aspirated from the plate. The wells were rinsed completely with PBS 3 times by adding the PBS, gently agitating, and then removing the PBS. A 50ul aliquot of

1:6000 diluted goat anti-mouse IgG alkaline phosphatase conjugated secondary antibody is added to each well, and the plate is incubated again at 37C for one hour. The secondary antibody is aspirated from the wells, and they are again completely rinsed with PBS 3X. For performing the alkaline phosphatase assay, 200ul of alkaline phosphatase reagent is added to each of the wells, the plate is placed into a reader, and data on the absorbance of the solution in each well at 405nm is collected at 5, 10, 15, 20 (3X), and 25 min.

Screening of Hybridoma Cultures

The protocol developed was as follows. A 50ul aliquot of 10ug/ml of the appropriate antigen (024 or 025 peptide solution) or PBS is added to the wells of the plate. The plate is covered and incubated in the 4C refrigerator for at least 1.5 hours. The antigen solution is removed, and the plate is rinsed with 200ul PBS. 200ul of 1% BSA is added to each well, and the plate is incubated at 4C overnight. The next morning, the BSA is removed; the plate rinsed with 200ul PBS, and 50ul aliquots of the hybridoma culture media is placed into the appropriate wells. After incubating at room temperature for 1 hour, the media is removed, and the wells are rinsed with 200ul PBS. A 50ul aliquot of 1:600 goat anti-mouse IgG alkaline phosphatase conjugated secondary antibody is added to each well. After incubating for 1 more hour, the secondary antibody is removed, and the plate is rinsed 3X with 200ul PBS. A 50ul sample of alkaline phosphatase assay reagent is placed in each well, and the plate is protected from light by covering with aluminum foil. After incubating at room temperature for 20 minutes, the absorbance of each well is recorded at 405nm using an

ELISA plate reader. The absorbencies of the control wells containing PBS are subtracted from that of the other wells.

Results and Discussion

Immunization

The peptides are used as antigens in this portion of research, and will be referred to as antigens in this section. Three mice were injected with a solution of each peptide. Immunization of the mice against the synthetic peptides was begun on 9/17/01. Sera was collected from each of the six mice first on 10/16/01, and secondly on 11/6/01. Approximately six weeks after the mice are first immunized with the peptides, ELISA screening is conducted using the sera collected during each of the first two bleeds to ascertain the relative immune response of each mouse to the peptide. Twelve sera samples were analyzed. Six were from bleed 1 six were from bleed 2. Of the six samples collected during each bleed, three were from mouse #1-3 immunized using 024, and three were from mouse #1-3 immunized using 025. Three dilutions of each of the twelve sera samples was prepared using PBS; 1:300, 1:600, and 1:900. The absorbance of the wells coated with PBS instead of antigen is subtracted from the absorbance of the wells coated with antigen as a control. The data are represented graphically in Figure 4.2-4.7. The data on the left side of each graph represents the first sera sample, collected on 10/16/01. The data on the right side of each graph represents the second sera sample, collected on 11/6/01. In each case, the response was much greater using the sera samples collected during the second bleed. The mice are producing more and more antibody

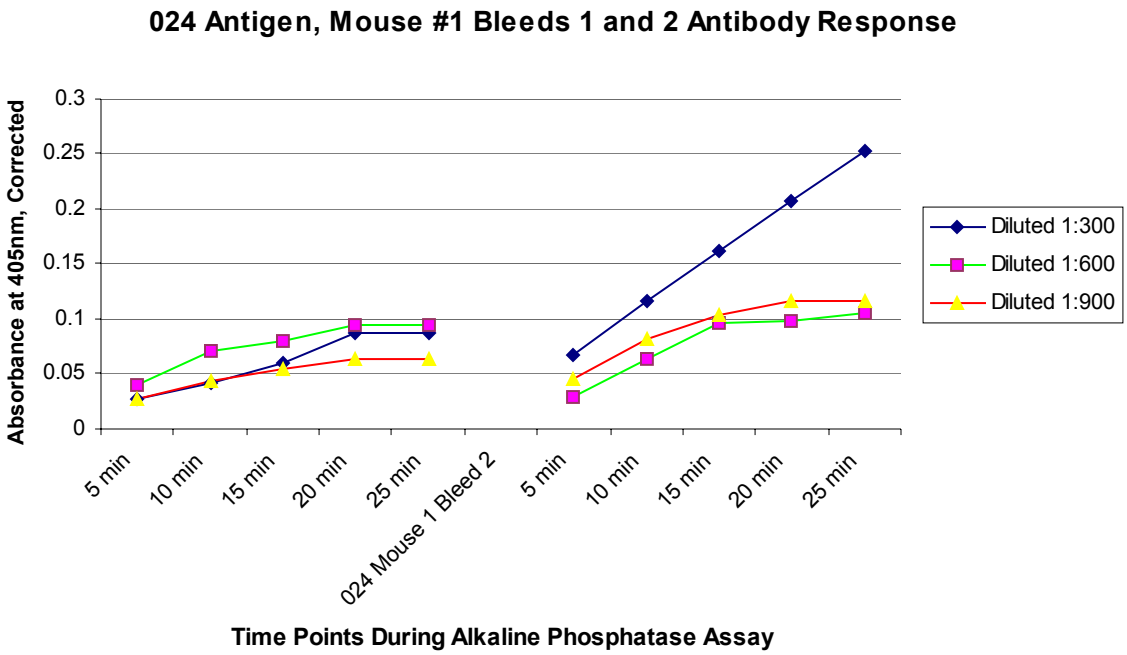


Figure 4.2. Mouse 024 #1, Antigen Response

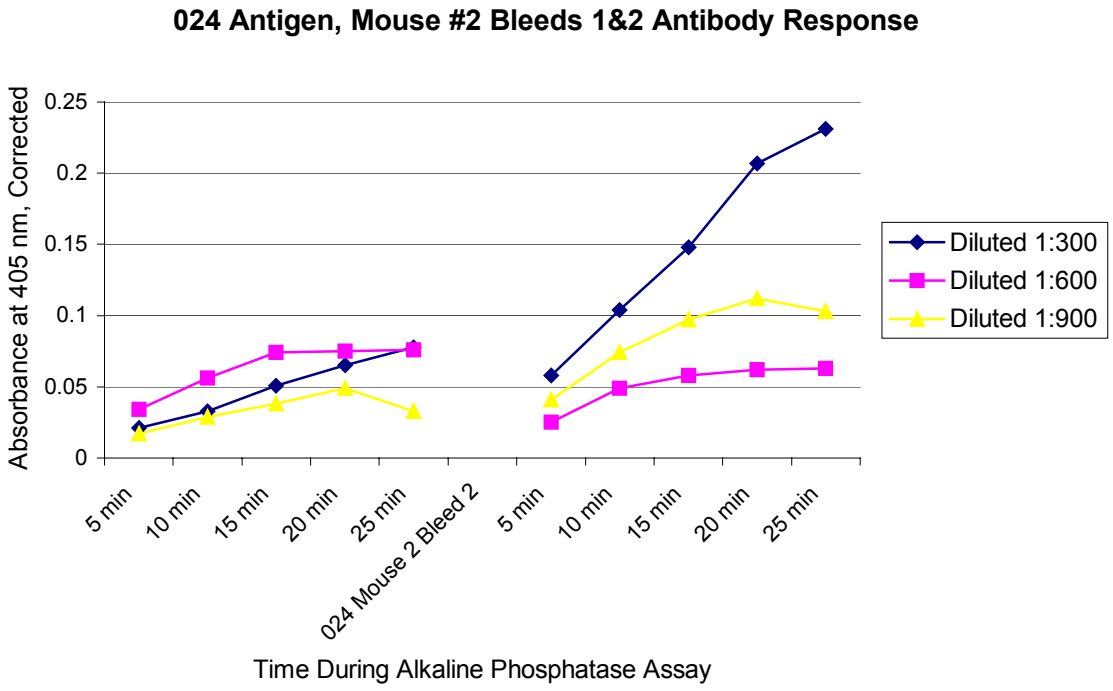


Figure 4.3. Mouse024 #2, Antigen Response

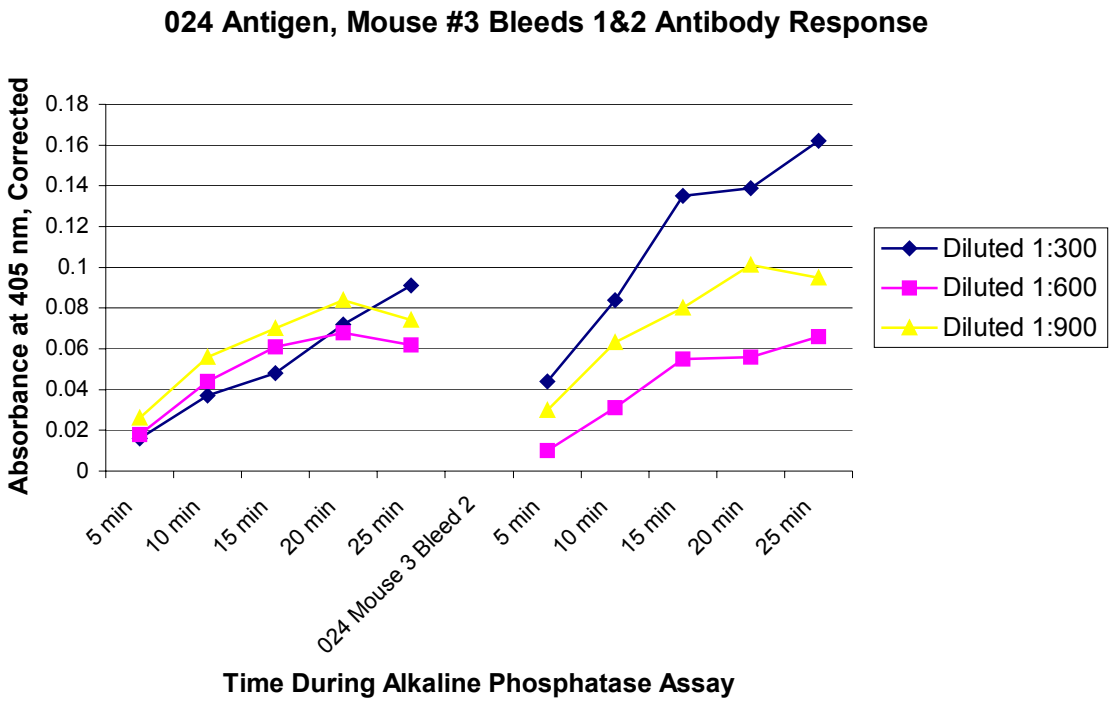


Figure 4.4. Mouse 024 #3, Antigen Response

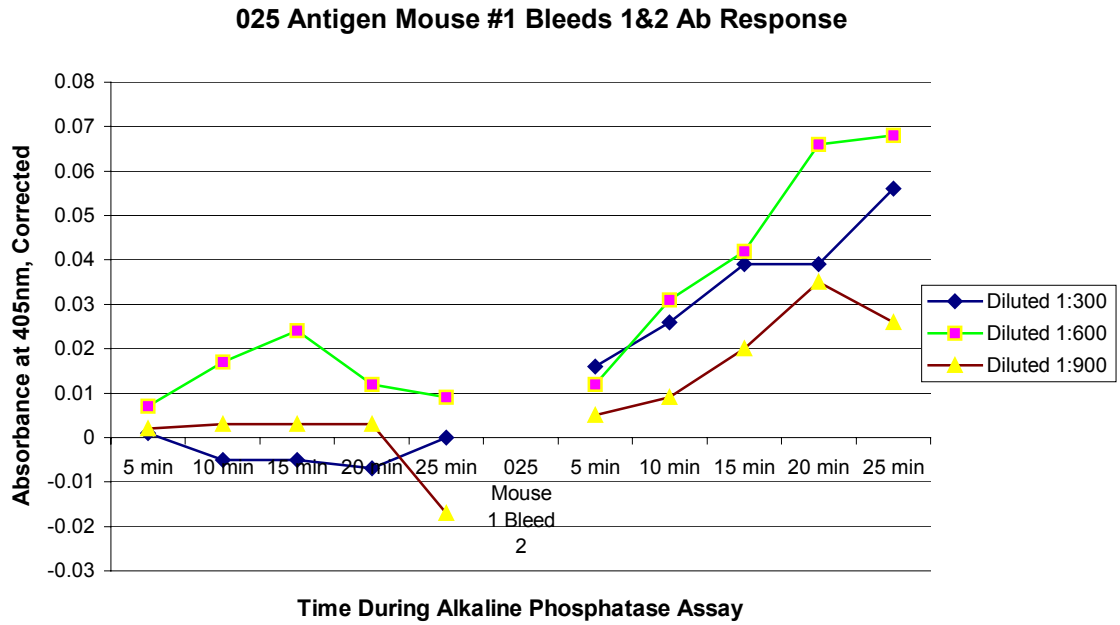


Figure 4.5. Mouse 025 #1, Antigen Response

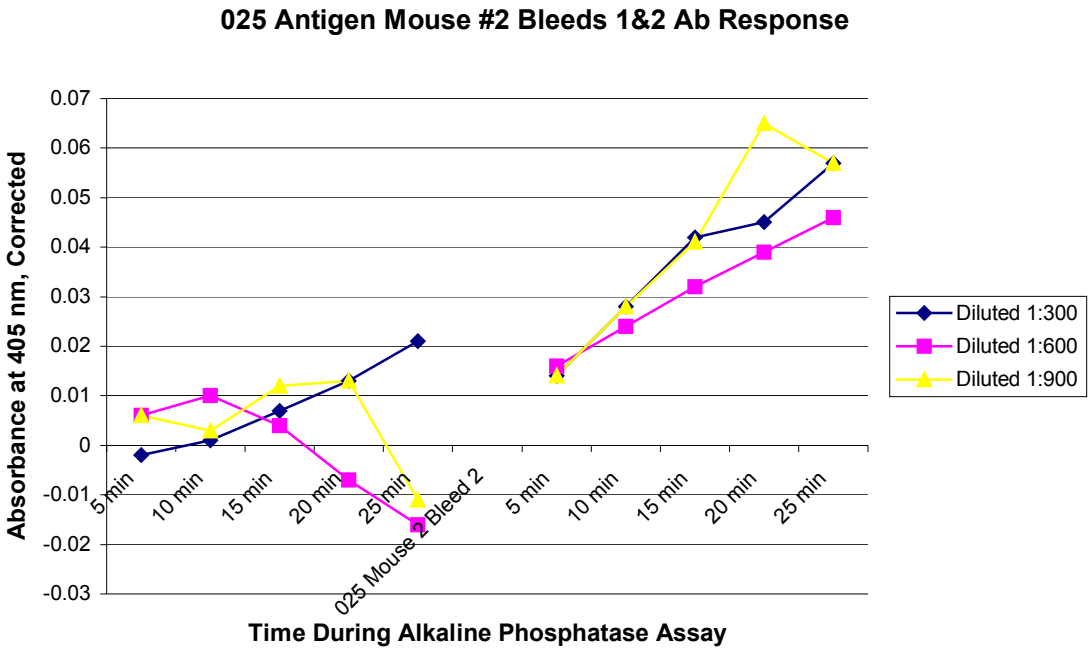


Figure 4.6. Mouse 025 #2, Antigen Response

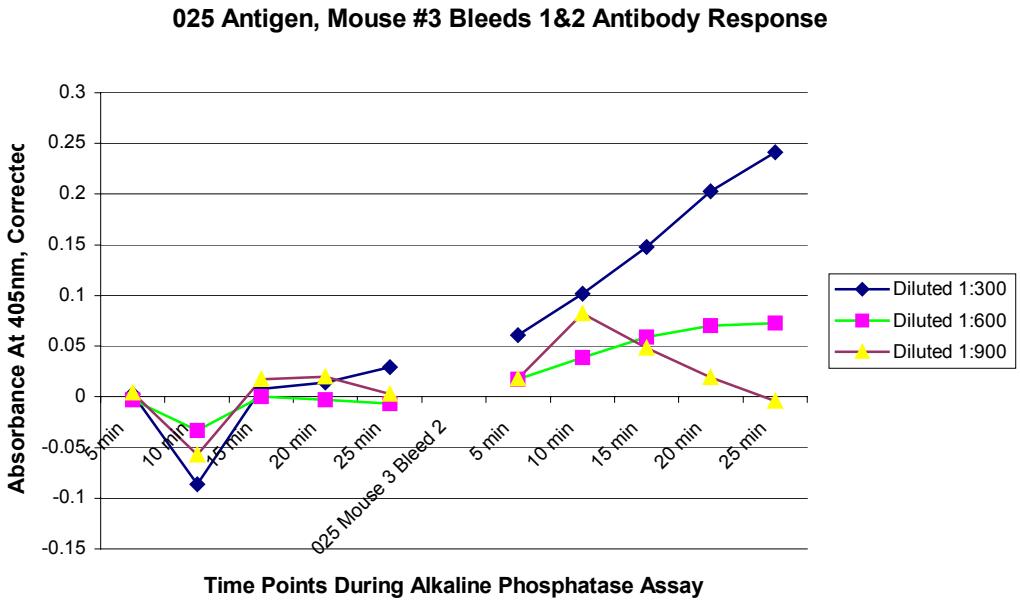


Figure 4.7. Mouse 025 # 3, Antigen Response

directed against the synthetic peptides as time goes by and they continue to be boosted with antigen. The mice responded to the antigens. Had the mice not responded to the antigens, monoclonal antibody production would have been impossible. However, not every mouse is responding in a strongly enough to be useful for producing monoclonal antibodies. All three mice immunized against the 024 peptide responded strongly with antibodies. The order of this response from strongest to weakest is: 1>2>3. Only mouse number 3 responded to the 025 peptide. Again, 024 corresponds to the known CD4 cytoplasmic tail and is being used to prepare a positive control antibody. 025 corresponds to the polypeptide that is expressed if the UGA codon is read through. The anti-025 antibody is absolutely critical for labeling experiments to determine whether or not this UGA is read through. 024 must be more immunogenic than 025, given the weak response from all but one of the mice. The monomer peptides of 025 that were purified previously would have almost certainly not been immunogenic enough to cause the mice to respond with antibodies. The non-curvilinear nature of the data for mouse 1 and 2 of 025 is probably due to the very low absorbencies recorded, and noise in the assay.

Boosting

The mice are boosted with more antigens on 11/20/01, and more sera are collected on 11/27/01. Again, the sera are assayed for antibody activity as before. The results are shown in Table 4.1. Consistent with the last assay, only mouse number 3-025 is responding to the 025 peptide. Boosting this time has

Table 4.1. 11/27/01 Boosted Sera Collection

		Mouse 1	Mouse 2	Mouse 3	Control Sera
024	1:300	0.297	0.134	0.146	0.116
	1:600	0.130	0.133	0.126	0.104
	1:900	0.215	0.145	0.123	0.110
025	1:300	0.110	0.128	0.245	0.116
	1:600	0.107	0.114	0.168	0.119
	1:900	0.107	0.108	0.105	0.106

not increased the response appreciably. Although the response is strong, it is still not as high as desired before fusion is done. The mice are boosted again in December and January, sera are collected, and ELISA analysis is repeated in order to gauge the response of the mice to the antigens. Repeated boosting is producing diminishing returns, as the response of the mice to the antigens appears to be decreasing. Mouse number 3-025 is the only mouse responding to the 025 antigen. Mice numbers 1 and 2 exposed to the 024 antigen are responding with antibodies.

Because of the extra boosting, the supply of 024 and 025 MAP peptides is running low. More of the peptides are ordered from MGIF, and the data relating to their synthesis and purity is similar to that presented previously in Figure 3.15 and Figure 3.16. MGIF reports that there is 23mg of the 025-type peptide and 25mg of the 024-type peptide. Purity for the 024-type peptide is not as good as the last synthesis, but the problems associated with continued purification render it unfeasible. The impurities are not a large proportion of the sample either.

Fusion

Because repeated boosting is failing to increase the response of the mice to the antigens, the fusion between B cells isolated from the spleen and hybridoma cells is scheduled for mice 1-024, 2-024, and 3-025. The fusion for 1-024 and 2-024 occurs on 1/18/02. The fusion for 3-025 occurs on 1/25/02. Soon after fusion, on 1/28/02, intense screening of the immortalized cell lines was begun.

Hybridoma Screening

During the intense screening period, the ELISA protocol was optimized to increase sensitivity and speed in order to deal with the high number of plates that must be screened each day. A high absorbance in the nPP assay is correlated with a high primary antibody activity in the hybridoma culture represented by that particular well in the plate. As outlined previously, the cultures that give a strong response are serially diluted and tested again until a monoclonal culture of the antibody producing cells is attained.

Because of the high volume of plates that had to be prepared and processed each day, there was a strong need for any device or method that could make the process more efficient. Rinsing the plate was the procedure that had to be repeated most frequently, and was accounting for a significant amount of time spent processing the plates. A special device was made to make the rinsing operation more efficient (Figure 4.8). The device was made from 96 1ml syringes, 2 96 well plates, one piece of acrylic sheet, and epoxy. The bottoms of the wells were drilled out of both plates and the holes in the plates were slightly enlarged in the process. The tops of the syringes were removed including the flange. The top of the syringe plungers were trimmed so that they could fit very closely side by side with other syringe plunger tops. The syringes were inserted into the holes into one of the 96 well plate (tops coated with epoxy), the plungers were inserted through the holes in the other plate and into the syringe bodies, and the acrylic sheet was used to cover the syringe plunger tops and the bottom of the second plate. The syringe bodies and acrylic sheet plate were glued into

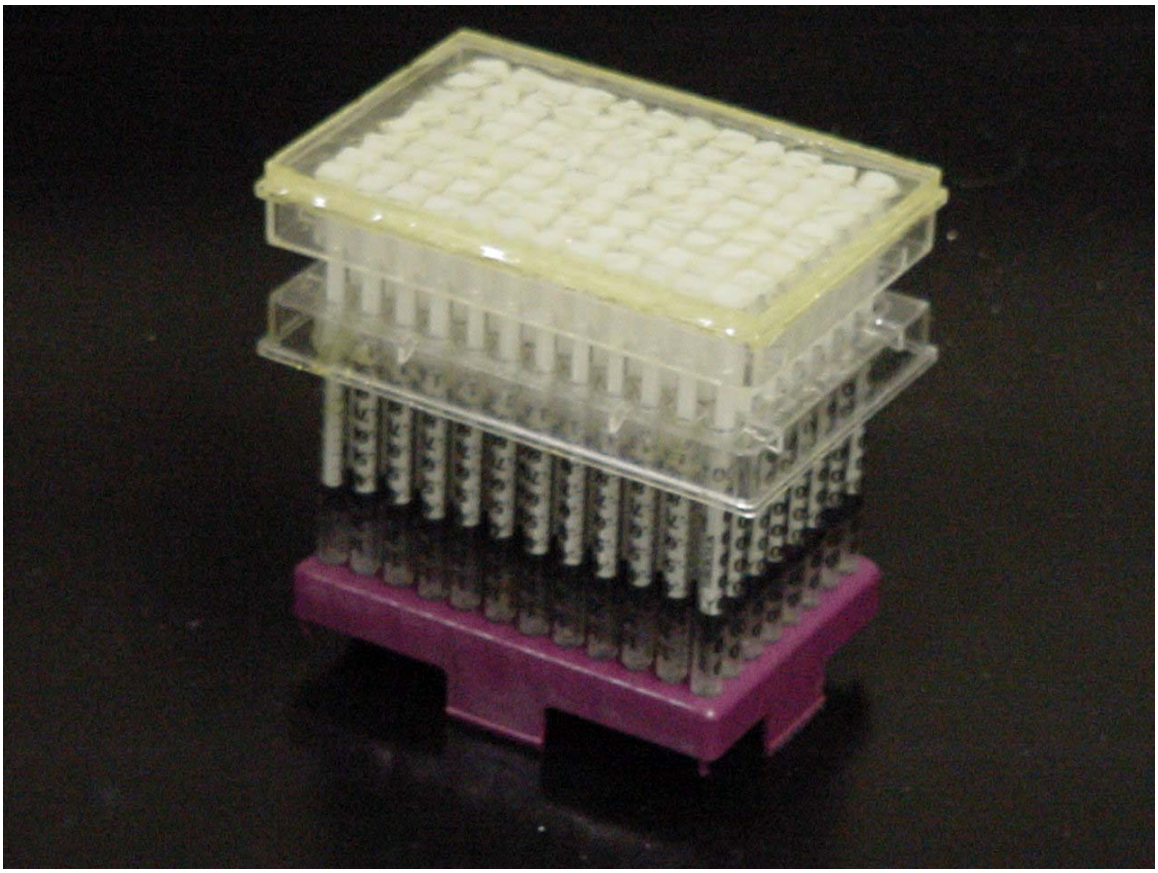


Figure 4.8. Rinsing Apparatus For Rapid ELISA Screening

place using epoxy. After the epoxy cured, the device was loaded with PBS by submerging the tips of the syringes in PBS, and pulling the two plates away from one another, drawing the syringe plungers up the syringe bodies. The syringes were loaded with 1ml of PBS at a time, and it was carefully dispensed 200ul at a time by placing the tips of the syringes into the wells of the plate and pressing the two plates together causing the syringe plungers to move down the syringe bodies. This device rinsed all 96 wells at one time, and greatly reduced the amount of time spent rinsing the plates.

Well over 100 96 well plates were screened for antibody activity using the ELISA procedure outlined above. It is impractical to include the data from every assay. However, the data from the last few screens will be included both as a representative sample and as verification of the pedigree of the cell lines chosen to produce the antibodies for later experiments (Table 4.2- Table 4.4) Table 4.2- Table 4.3 are for anti-024 peptide antibodies, corresponding to the known region of the cytoplasmic tail of CD4. Table 4.4 is for anti-025 peptide antibodies, corresponding to the region past the UGA in CD4. The large tables shown in each figure represent the layout of a 96 well plate. The letters in the boxes represent the location of a well in a cell culture plate from which media was taken and used to coat that well of the assay plate. At this point in the screening, the highly active cell lines have been given designations that relate to the location from which they were isolated through the screening process. These designations are written along the top of the table to show that those wells under the designation are derived from this particular culture. The + and – designations

Table 4.2. Antibody Screening PGM 024

	+ 14H11F11 -	
D7	0.279	0.292
F2	0.462	0.433
F4	0.424	0.224
G9	0.507	0.437

	+ 5C2B9 -	
B7	0.233	0.110
C3	0.177	0.095
D4	0.258	0.098
F12	0.255	0.134

	+	-
Normal Serum	3.452	0.116
10k	1.760	0.090
30k	1.135	0.090
100k	0.588	0.090

Table 4.3. Antibody Screening 024

	+	-		+	-
	10D6B3			**10D6 Clone**	
A4	3.898	0.624	B3	3.916	0.684
D12	3.936	0.238	C1	3.846	0.232
G9	3.847	0.194	C5	3.880	0.212
H9	3.863	0.175	F3	3.786	0.173
Normal Serum	0.242	0.151	A9	**10C12** 2.344	0.186
10k	1.506	0.127	B9 clone	2.482	0.228
30k	0.805	0.126	D3	2.536	0.181
100k	0.480	0.199	D4	2.102	0.216

+	+	+	+	+
		10C12B9		
A3 3.244	**B5** 3.306	**D6** 3.065	**F3** 3.175	**H2** 3.082
A5 2.908	**B6** 3.104	**D10** 2.976	**F4** 3.093	**H3** 2.944
A6 2.617	**C3** 2.637	**D11** 2.752	**F7** 3.050	**H5** 2.985
A11 2.460	**C5** 2.592	**E2** 2.475	**G3** 2.825	**H6** 2.611
A12 2.261	**C9** 2.490	**E4** 2.353	**G5** 2.592	**H8** 2.468
B1 2.392	**D1** 2.349	**E12** 2.258	**G6** 2.520	
B2 2.528	**D3** 2.345	**F12** 2.401	**G11** 2.505	
B10 2.475	**D4** 2.393	**F1** 2.315	**G12** 2.616	

Table 4.4. Antibody Screening 025

	+	-		+	-
	8B8F4F2			**8A5B10**	
C8	1.762	0.873	A5*	3.067	0.111
D8	2.211	0.652	G5	3.003	0.101
E8	1.537	0.491	A52	2.804	0.128
H7	1.504	0.665	G5	2.947	0.196
E3	**3F5E2E3** 1.968	0.926	Normal Serum	0.085	0.076
D10	2.415	0.932	3k	0.917	0.102
D3	1.307	0.618	10k	0.426	0.086
B6.C2	1.579	1.157	30k	0.272	0.080

along the top of the plate layout represent whether the column above was coated with antigen or PBS. Positive columns were coated with antigen as outlined above, and negative columns were coated with PBS. The numbers represent the absorbance at 405nm of each well after incubation with the alkaline phosphatase assay reagent for 20 minutes. The absorbencies of the strongly positive wells are very high compared to the initial screens, demonstrating the success of the fusion and screening processes.

Isotyping

After isolating colonies of hybridoma cells producing antibodies that have high activity against the synthetic peptide antigens, isotyping is performed to determine the class of the antibodies. This information is potentially useful for future experiments where the antibody is labeled with a secondary antibody. The antibody isotyping data is shown in Table 4.5. Most of the antibodies obtained are IGM. None of the anti-025 antibodies are IgG. Only the 10D6B3H9 and 10C12B9A5 anti-024 antibodies are clearly IgG.

Conclusion and Preparation of Concentrated Stocks

The IgM antibodies are cross-reactive as IgG antibodies to the secondary antibody used in these experiments, which is not uncommon. In fact, the entire screening process has been done using anti-mouse IgG secondary antibodies, so if they were not sufficiently cross reactive as IgG, they would not have been found in the ELISA screening process. Although cross reactivity is not uncommon, it is still an issue that has to be monitored. Some secondary antibodies may be cross-reactive whereas others are not. For example, the

Table 4.5. Isotyping

		poly	IgM	IgA	IgG_1	IgG_{2a}	IgG_{2b}	IgG_3	κ	λ
024	5C2	1.188	1.978	0.057	0.199	0.105	0.070	0.060	1.638	0.091
	14H11F11	1.070	1.920	0.081	0.481	0.112	0.162	0.071	1.585	0.079
025	8A5B10A5	1.005	1.949	0.073	0.428	0.146	0.133	0.067	1.720	0.073
	3F5E2	1.381	2.197	0.137	0.905	0.319	0.437	0.082	1.847	0.127
	8B8F2	0.508	1.081	0.115	0.779	0.317	0.198	0.103	0.092	1.209
	10C5F4A7	1.498	2.117	0.059	0.439	0.309	0.092	0.054	1.819	0.077
	10A11F11E7	1.282	2.024	0.311	1.227	0.477	0.482	0.153	1.631	0.116
	19C2C7C5	1.352	2.124	0.246	1.125	0.421	0.429	0.128	1.847	0.109
	18G7G9	1.098	2.138	0.126	0.712	0.343	0.317	0.133	1.627	0.122

secondary antibody used in the ELISA screening process was Sigma product A-1293, a goat anti-mouse IgG alkaline phosphatase conjugate antibody. This antibody is Fab specific. The fact that the IgG and IgM classes of antibodies share the light chain isotypes κ and λ, and that the light chain is contained in the Fab region could explain the cross reactivity of this antibody. In fact, this antibody may be cross-reactive to many different antibody classes because all share either the κ or λ light chain isotypes. In any case, the antibodies isolated during the screening process are labeled by the anti-mouse IgG alkaline phosphatase conjugate antibody (Sigma product #A-1293).

Many of the hybridoma cell lines producing very reactive antibodies are saved in liquid nitrogen by adding 5% glycerol or to their media and then freezing them in a step wise process from –20C to –80C and finally to liquid nitrogen, -195C. Three of the most reactive cell lines were chosen for immediate antibody production for use in labeling experiments. The three cell lines chosen for immediate antibody production are, for the 024 antigen, 10D6B3H9 and 10C12B9A5, and for the 025 antigen, 8A5B10A5. One liter of each of these three hybridoma cell lines were grown and the soluble antibody in their media was isolated using an ammonium sulfate precipitation procedure. The concentrated antibody solutions are stored at 4° C.

CHAPTER 5

LABELING EXPERIMENTS

Introduction

The antibodies isolated as outlined previously were used to label various cells and cell extracts. As mentioned previously, the 024 antigen is a synthetic peptide that mimics the cytoplasmic tail of CD4 just before the UGA codon that is presumed to end translation. There is no disagreement that this region is translated into an amino acid sequence, so this antibody is a positive control. The 025 antigen is a synthetic peptide that mimics the amino acid sequence that would result from translation continuing through the UGA codon that is presumed to end the CD4 gene. There is no direct experimental evidence that this region is translated as part of the CD4 gene although the possibility is supported by theoretical evidence outlined previously. The goal of this portion of research is to show whether or not translation continues through the UGA codon resulting in a CD4 protein that is larger than expected and a cytoplasmic tail that has been studied incompletely.

Several different antibody labeling methodologies were used including dot blots, Western blots, ELISA, and immuno-fluorescence microscopy. Each of these methods has specific advantages, disadvantages, and limitations. An illustration of the overall strategy employed using each of these techniques is shown in Figure 5.1. Western blotting is more specific than dot blotting and has the potential to identify a single protein source that the antibody recognizes. This

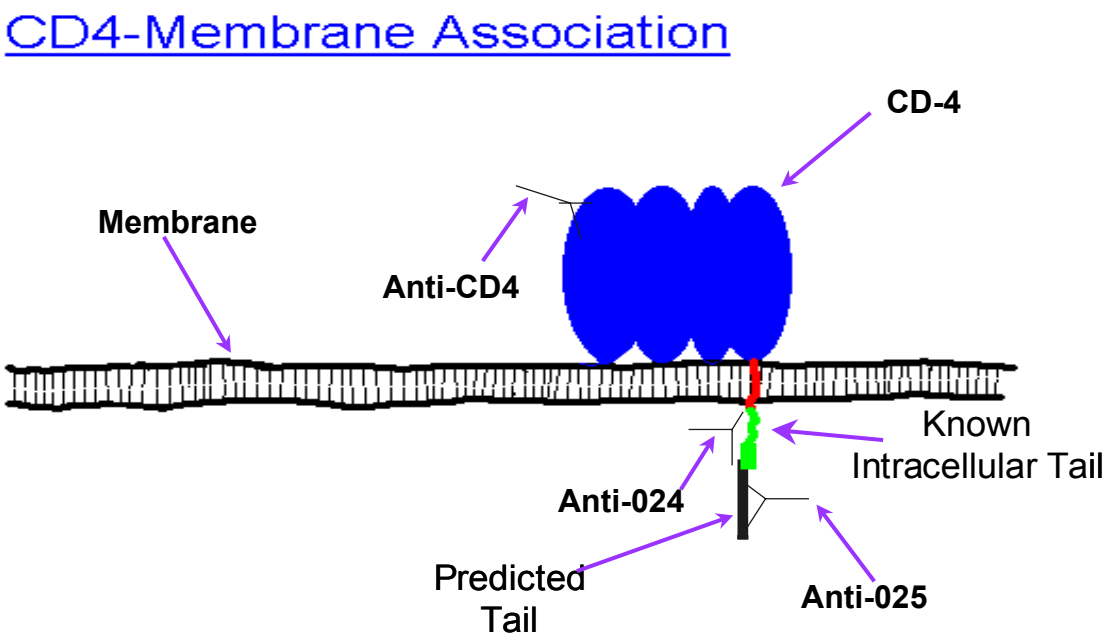


Figure 5.1. Antibody Labeling Strategy

is in contrast to dot blots, which are never able to determine with certainty which protein within a mixture is actually being recognized by the antibody. ELISA techniques are very sensitive, often making it possible to identify the presence of very rare target antigen. However, similar to dot blotting, ELISA techniques can only identify whether or not there is an antigen present that the antibody binds to, but cannot conclusively identify the exact identity of this substance within a mixture.

An overview of Western blotting technique is shown in can be found in the second edition of Basic Methods in Molecular Biology.[86] First, the protein mixture containing the target is separated using gel electrophoresis, sometimes native, sometimes SDS. Electrophoresis separates molecules based upon their molecular weight and charge. SDS gel electrophoresis is usually preferred because this technique normalizes the distance the protein travels through the gel in relation to its molecular weight. SDS is a negatively charged detergent that denatures the protein and binds to it in an approximately per molecular weight basis. This gives the protein a negative charge proportional to its molecular weight, and imparts a constant charge to mass ratio on the proteins in the mixture. When using native gel electrophoresis, the charge to mass ratio of the protein can vary greatly, and neither is normalized in relation to the other as in SDS gels. Further, in native gel electrophoresis, the charge carried by the protein is, in turn, dependent both on the amino acid composition of the protein and the pH of the buffer it is in. The combination of these factors makes it difficult to identify with certainty that a band on a gel is the target of interest and

not something else. In Western blotting, this problem can result in circular reasoning where a band on the gel is identified as the target because the antibody bound to it, and the antibody is thought to be specific to the target because it preferentially binds to a particular band on the gel. The use of buffers containing SDS eliminates much of this uncertainty by allowing the target to be identified based solely upon its molecular weight. However, because SDS denatures the protein and binds to it extensively giving it an overall negative charge, it may damage the epitope recognized by some antibodies, making it difficult to identify the target by preventing antibody binding. The sensitivity of Western blotting technique depends on this and many other factors including the sensitivity of the antibodies and composition of the reagents used during the blotting. Typically, a protein such as BSA and a detergent such as Tween 20 are added to the reagents in order to reduce background binding. The BSA covers places on the blot where protein may nonspecifically bind, and Tween 20 is a detergent that reduces nonspecific binding.

Each of the methodologies employed has several things in common. In the labeling experiments, typically both a primary and a secondary antibody are used. In each experiment, the primary antibodies used are the anti-024 antibody, the anti-025 antibody, and a commercial anti-CD4 antibody available from Sigma (#C-1805). During some of the whole cell labeling experiments, the commercial anti-CD4 antibody conjugated to the fluorescent tag FITC is used in the place of primary and secondary antibody as a positive control. Several secondary antibodies are used depending on the particular experiment.

In general there are two positive controls used in these experiments, the commercial anti-CD4 antibody available from Sigma (product #C-1805) and the antibodies isolated against the 024 antigen. The commercial anti CD4 antibody binds to an external portion of CD4 between loops 1 and 2, blocking the region where HIV binds. The anti-024 antibody binds to an intracellular region of CD4 described previously in the structural analysis section of the peptide purification chapter. Presumably, the extra cellular portion of CD4 is more accessible to antibody labeling than is the intracellular cytoplasmic tail during whole cell labeling experiments. Because the anti-025 antibody also binds to a potential intracellular region, the anti-024 antibody is a more comparable positive control than is the commercial antibody. In addition, since neither of the antibodies isolated using the synthetic peptide antigens have been proven conclusively to bind to their targets in actual cell preparations, the commercial anti-CD4 antibody is used as an overall positive control to show that CD4 is present and the methods used are sensitive enough to label it.

The search for an ideal negative control for use in these experiments is more problematic. However, several negative controls are used. PBS is used as a negative control in the place of cell extract where appropriate. The secondary antibody alone is used as a negative control in each experiment. In addition, CD4 negative cells available from the ATCC and isolates from cultures of these cells is used as a negative control. Although this cell line sounds like an ideal negative control, questions were raised about whether or not this cell line is truly CD4 negative or simply produces a defective CD4 receptor that is not labeled

using certain antibodies. These questions are addressed later, particularly in the ELISA section. Finally, HeLa cells are used as a negative control in the ELISA experiments. Other controls have been incorporated in the experiments when available and appropriate.

Blotting

Materials and Methods

Antibodies

The secondary antibodies used include the anti-mouse IgG (Fab specific) alkaline phosphatase conjugate available from Sigma used previously (product #A-1293). One of the positive control antibodies utilized is the anti-CD4 antibody available from Sigma (product #C-1805). Both the anti-024 and anti-025 antibodies are utilized in these experiments.

Materials

Nitrocellulose membrane was obtained from Micron Separations, Inc. and PVDF membrane was obtained from Amersham Pharmacia. Cells cultured to produce cellular extracts, including H9, Jurkat (+), and Jurkat (-) were obtained from the ATCC. Reagents used for buffer preparation were obtained from Sigma and JT Baker. Blots were developed using Western Blue stabilized alkaline phosphatase substrate available from Promega. 12X30ul lane 4-20% pre-cast gels used for both native and SDS gel electrophoresis were obtained from Gradipore.

Methods

Cell Extract Preparation

Cell extract for blotting was prepared from cell cultures as as follows. The cells were grown to a density of about $1X10^6$ cells/ml, and centrifuged in 50ml aliquots at 3500XG for 7 minutes. The culture supernatant is discarded, and the pellet is rinsed by re-suspending in 500ul PBS and repeating the pelleting procedure after the cell slurry is transferred to a 1.5ml tube. The PBS is then removed, and the pellet is re-suspended in 150ul of re-suspension solution consisting of 20mM sodium phosphate pH 7.0 containing 5mM EDTA, 5mM DTT, and 2mM PMSF. After sonicating on ice for 60 pulses of approximately 0.5 seconds each, the slurry is centrifuged at 20,800XG for 20 minutes at 4C. Supernatant is separated from the pellet, and they are stored at –80C before use.

Protein Assay

The protein concentration of the cell extract preparations is assayed using the Lowry Method.[87] Four solutions are prepared as follows.

1. Solution A, 100ml

   0.5g $CuSO_4$, 5 H_20

   1g $Na_3C_6H_5O_7$-($2H_2O$)

   Add distilled water to 100ml

2. Solution B, 1 liter

   20g Na_2CO_3

   4g NaOH

Add distilled water to 1 liter

3. Solution C, 51ml

1ml solution A

50ml solution B

4. Solution D, 20ml

10ml Folin-Ciocalteu phenol reagent

10ml distilled water

The sample volume is brought to 166ul using distilled water as necessary, and 833ul of solution C is added. The mixture is vortexed, and incubated at room temperature for 5-10 minutes. An 83ul aliquot of solution D is then added, and after incubating at room temperature for 20-30 minutes, the absorbance is recorded at 750nm. Simultaneously, a standard curve is prepared using the following amounts of BSA; 5ug, 10ug, 25ug, 50ug, 75ug, and 100ug.

SDS Gel Electrophoresis

The following stock solutions are prepared. (From Basic methods in Molecular Biology)[88]

1. 10% SDS: 50g SDS dissolved in H_2O to 500ml.

2. 0.5M Tris-Cl: 6g Tris base dissolved into 40ml H_2O plus 40ml 1N HCl. pH adjusted to 6.8.

3. 1% bromophenol blue: 0.1g bromophenol blue dissolved into 10ml H_2O.

The gel electrophoresis unit is assembled and filled with an SDS running buffer consisting of, per liter, 0.025M Tris, 0.19M glycine, 10ml 10% SDS stock. The pre-cast gels used for blotting (12 well, 4-20% acrylamide) had a maximum well

size of 30ul, so only 25ul can be loaded into each well to avoid spilling samples across the tops of the wells. Accounting for the 1:2 dilution of sample in SDS loading buffer, only 12.5ul of sample is loaded into the wells of the gel. The 2X SDS loading buffer consists of, per 10ml, 1M DTT, 2ml 0.5M tris-Cl stock pH 6.8, 4ml 10% SDS stock, 2ml 1% bromophenol blue stock, and 2ml glycerol. The pellet samples are prepared differently. In this case, 75ul of SDS loading buffer is added to each pellet sample, and they are heated to 100C for 10 minutes to denature and extract as much protein as possible from the pellet sample. The samples are centrifuged at 14,000XG for 5 minutes to pellet the insoluble fraction. Then, 25ul of the soluble fraction is loaded into the wells of the gel. After the samples are loaded, gel electrophoresis is carried out using pre-cast gels with 12 lanes of 30ul each. Gels are run at 120V for 1.5 hours, and typically draw 25mA (3W) of current. By this time, the dye in the sample-loading buffer is close to the end of the gel.

Native Gel Electrophoresis [84]

Samples are diluted 1:2 using a 2X-loading buffer consisting of, per 10ml, 5ml glycerol and 100mg bromophenol blue. Only supernatant samples are used. The gel electrophoresis unit is assembled and filled with a Tris-Glycine running buffer prepared as needed from a 10X stock solution consisting of, per liter, 6g tris base and 29g glycine at pH 8.8. As before, 25ul of the diluted sample is loaded into the lanes of the gel. Gels are run at 120V for 1.5 hours, and typically draw 25mA (3W) of current. By this time, the dye in the sample-loading buffer is close to the end of the gel.

Blotting

Transfer from the gel to a membrane is accomplished using an electroblotter apparatus. A blotting sandwich is prepared after soaking membrane and blotting paper in transfer buffer. Either nitrocellulose membrane or PVDF membrane are used. When performing an SDS Western blot, the transfer buffer consists of (per 1000ml) 5.8g tris base, 2.9g glycine, 0.37g SDS, and 200ml MeOH. When performing a native Western blot, and certain dot blots, the SDS is left out of the transfer buffer. Once assembled and placed into the electroblotter, a current of 100mA maximum is maintained through the sandwich for 4 hours. The voltage varies from 12 to about 50. Ice blocks are placed on the blotter over the sandwich in order to prevent overheating the membrane.

Probing Membranes With Antibody

The exact methods used for probing the membranes changed significantly many times, and the details of these changes are best left within the results and discussion section. However, all of the probing experiments have several things in common. In each case, the primary antibodies used are the anti-CD4 commercial antibody, the anti-024 antibody, and the anti-025 antibody. The secondary antibody is the goat anti-mouse IgG alkaline phosphatase conjugate antibody used previously. The anti-CD4 commercial antibody and the anti-024 antibody are used as positive controls, whereas the secondary antibody is often used alone as a negative control. Labeling by the anti-025 experimental antibody indicates the presence of the region produced from translation past the UGA codon. This will show that the UGA codon is read through and translation

continues past the UGA codon. The membranes are developed using the Western Blue stabilized substrate.[89] Once bands are visible or the background is becoming significant, the development is ended by removing the Western Blue reagent, rinsing the membrane strips with water, and allowing them to dry out.

Dot Blotting

The exact methods used for performing dot blots changed significantly several times, and the details of those methods and the changes they underwent are more clearly understood if they are left within the results and discussion section. However, all the dot blots have some things in common. Both the nitrocellulose and PVDF membranes are utilized. Strips of membrane are prepared, and 1-3 ul of sample, both pellet, supernatant, and whole cell extract, are placed onto the membrane strips and are allowed to remain for a while, sometimes drying down, sometimes remaining moist. After this point, the dot blotting protocol closely follows the Western blotting techniques and are probed with the alkaline phosphatase conjugate antibody and are developed using the Western Blue stabilized substrate. Once bands are visible or the background is becoming significant, the development is ended by removing the Western Blue reagent, rinsing the membrane strips with water, and allowing them to dry out.

Dot Blotting

Results and Discussion

In order to fully appreciate these results, they should be considered in conjunction with those of the Western blotting experiments described in the next section. The focus of performing the dot blots was two fold. First, dot blots using

cellular extracts were done to test the antibody reactivity to cellular extracts and to help develop protocols for use in Western blotting. Second, dot blots are conducted to prepare suitable protocols for Western blotting. Accordingly, the methods for performing dot blots are similar to typical Western blotting methods. In doing Western blots with PVDF membrane, the membrane must first be soaked in methanol, and while still wet placed into transfer buffer to soak for at least 30 minutes. If this procedure is not followed, acceptable electrical conduction through the membrane will not be achieved, and the blotting process will fail. Also, when performing Western blots, the membrane must be wet when the blot is performed, and is exposed to a transfer buffer in doing so. These general guidelines are followed except when one of these parameters or reagents is tested experimentally, and it is the goal of the dot blotting procedure to develop successful protocols that incorporate these general conditions.

A: PVDF Membrane, Using SDS

The first method used to perform dot blots is as follows. The transfer buffer used in this experiment contains SDS. The primary buffer in the antibody, blocking, and rinsing solutions is TBS. Five strips of PVDF blotting membrane are prepared by wetting them with methanol for 10 minutes before being moved to transfer buffer solution where they were soaked for an additional hour. The strips were transferred to a clean petri dish and a few drops of transfer buffer were placed along the edge of each so they would not dry out. Cell extracts were prepared as before, and 1ul each of 3% BSA 0.2% Tween, cell pellet, and cell supernatant samples was placed onto each membrane strip. The protein

concentration of the cell extracts for this experiment was 1.7ug/ul. The strips incubate at room temperature for 45 minutes. After rinsing in TBS with gentle agitation for 15 minutes, they are transferred to a 3% BSA, 0.2% Tween TBS blocking solution. The strips are incubated in the blocking solution overnight at 4C. One strip is placed into each of the following solutions of antibody with 0.2% Tween 0.1% BSA and TBS: 1:1000 anti-CD4 commercial antibody; 1:1000 anti-024 antibody and 1:1000 anti-025 antibody. One strip is left in the blocking solution for now and will only be incubated in the secondary antibody, and another strip is left in the blocking solution and will not be exposed to any antibody. The strips incubate in the antibody solutions for one hour at room temperature with gentle agitation. The strips were first rinsed in TBS 0.2% Tween for 10 minutes and then were transferred to fresh TBS 0.2% Tween and left for 1 hour at room temperature with gentle shaking. The strips, including one left in blocking solution up to this point, were transferred to a solution of 1:1000 anti-mouse IgG alkaline phosphatase conjugate antibody in TBS with 0.2% Tween and 0.1% BSA. After incubating at room temperature for 1 hour, the strips of membrane are rinsed as before for 10 minutes and 1 hour. All five strips are included in the rinsing. All are then transferred to a petri dish containing Western Blue reagent and were covered with aluminum foil. The strips are observed at 20 minutes and allowed to continue developing until 50 minutes at which time they are rinsed in water and allowed to air dry. Surprisingly, the membrane soaked in anti-025 antibody developed slightly more color than the others, and the membrane only exposed to secondary antibody developed color

similar in intensity to the membranes exposed to anti-CD4 commercial antibody and anti-024 antibody. The membrane not exposed to any antibody developed only very faint color. The color developed on the dot blots is not due to endogenous alkaline phosphatase. If it were, the strip of membrane not exposed to any antibody would have produced significant color. Both the pellet and supernatant spots developed significant color. These results indicate some non-specific binding of the anti-mouse IgG antibody during these procedures. This is in spite of the fact that Tween and BSA were added to all reagents in an effort to reduce non-specific background binding.

B: PVDF 2^0 Dilutions w/SDS

Another series of dot blots were performed using a series of dilutions of secondary antibody, anti-mouse IgG alkaline phosphatase conjugate, without exposing the blots to any primary antibody. The purpose of this experiment was to determine the dilution of secondary antibody required to reduce the background to near zero. The blots were repeated using the same protocol used above for the membrane strips above, except they are only exposed to secondary antibody. Strips of membrane, after being prepared as before with spots of 3% BSA 0.2% Tween in TBS, cell extract supernatant, and cell extract pellet (1.5ug/ul protein), were exposed to a series of dilutions of the anti-mouse IgG alkaline phosphatase conjugate antibody. The series of dilutions were 1:1000, 1:1000, 1:3000, 1:5000, 1:7000, 1:10000, 1:20000, and 1:30000. The results after 36 minutes of incubation in Western Blue reagent, rinsing in water, and air-drying, are shown in Figures 5.2-5.3. The color intensity seems to drop

Figure 5.2. DotBlots, Secondary Dilutions. S=Supernatant, P=Pellet

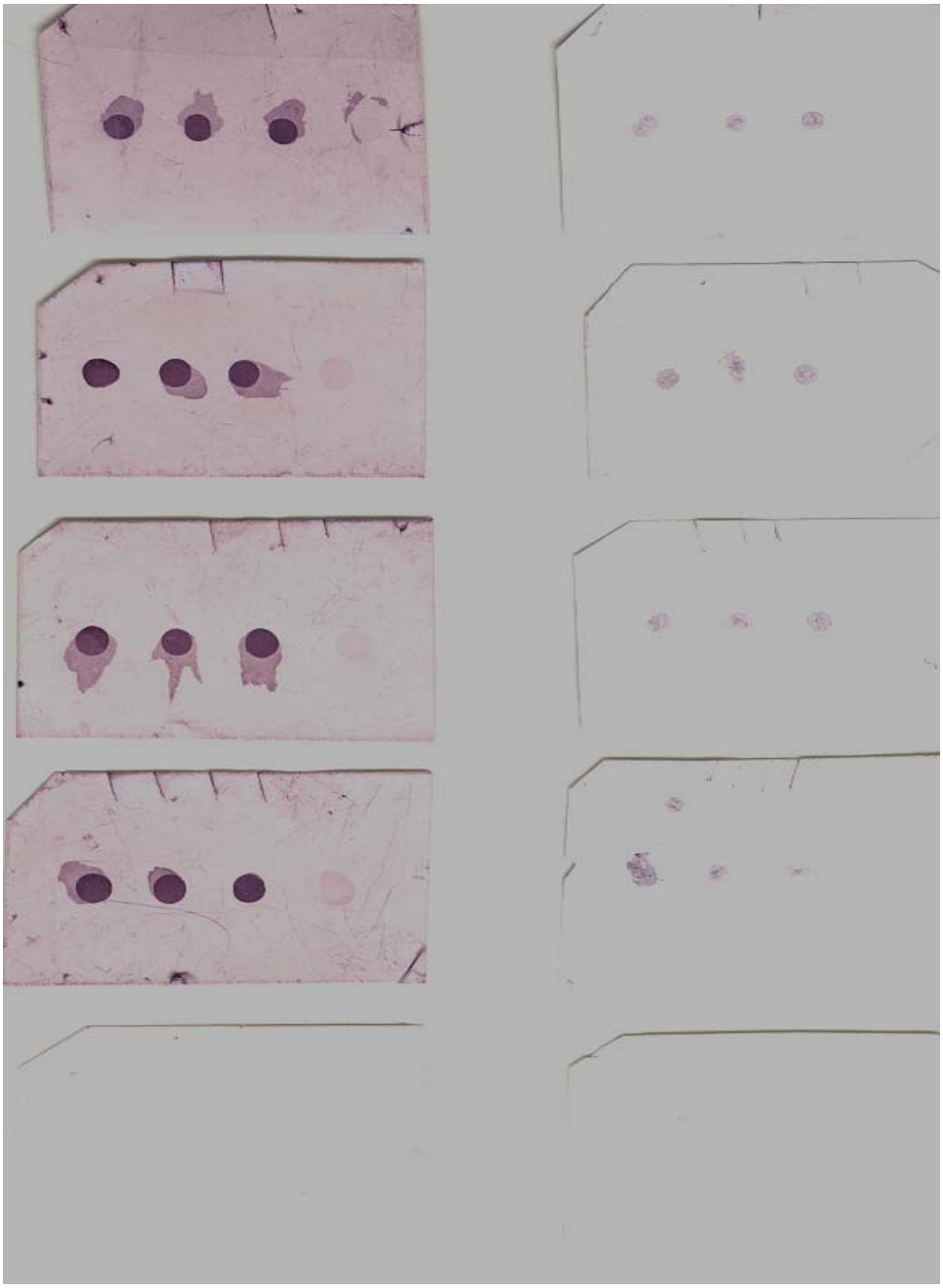
Figure 5.3. DotBlots, Secondary Dilutions. S=Supernatant, P=Pellet

rapidly from the 1:1000 dilution to the 1:5000 dilution, but changes little from the 1:5000 dilution to the 1:30000 dilution. The dark color in the pellet spots of the 1:7000, 1:10000, and 1:30000 blots develops over an extended period of time. The nonspecific labeling in the dot blots is not completely eliminated despite using protocols and reagents designed specifically for this purpose. 3% BSA and 0.2% Tween are considered to be high concentrations of these reagents.

Different antibodies are known to have varying sensitivities to the presence of various reagents. For example, some are known to no longer function properly when the target has been fixed with formalin. Others are sensitive to reagents such as SDS, various salts of certain concentrations, and pH changes. The secondary antibody used may simply be incompatible for use with the PVDF membrane or this set of conditions and reagents in general.

C: PVDF w/o SDS, w/ TBS

The dot blotting experiment is repeated without using SDS in the transfer buffer. This protocol is very similar to that used previously except for this change, and won't be discussed in more detail here. There were five strips used this time, each loaded with 3ul spots, in the usual manner, of the following extracts, H9, Jurkat (+), Jurkat (-), and JM 109 E.coli. The protein concentration of these extracts was about 1ug/ul. These strips were exposed to antibody solutions as follows: 1) anti-024 + 2^0, 2) anti-025 + 2^0, 3) anti-CD4 commercial + 2^0, 4) 2^0 only, and 5) nothing. The blots were developed for 20 minutes in Western Blue reagent. The results are shown on the left side of Figure 5.4.



With Transfer Buffer Without Transfer Buffer (PBS)

Figure 5.4. DotBlots, Comparison of Transfer Buffer and PBS. From the top, the membranes are treated with the following antibodies: anti-024, anti-025, anti-CD4 commercial, secondary antibody only (alkaline phosphatase conjugate), and neither. The spots are from left to right: H9, (+) Jurkat, (-) Jurkat, BSA.

Each blot, except the one not exposed to any antibody, appears to have developed the same amount of color. Apparently, all of the color developed on each blot can be attributed to nonspecific binding by the secondary antibody. Interestingly, however, the JM 109 spots did not develop significant color. The nonspecific interaction by the secondary antibody appears to be limited to the lymphocyte extracts

D: PVDF w/o Transfer buffer, w/ TBS

The experiment is repeated using the following change. This experiment is designed to determine whether or not there is something in the transfer buffer that is interfering with the blotting protocol. This time, the five membrane strips are not exposed to transfer buffer at all. The membrane is loaded with cell extract as in the previous experiment; however, this time the membrane is dry. The remainder of the protocol has not changed. The results after 20 minutes of development are shown in Figure 5.4. Again, each blot except the one not exposed to any antibody developed color. The color intensity of the blot only exposed to secondary still appears to be just as great as those exposed to both primary and secondary antibody. Again, the JM 109 E.coli spot did not develop significant color. Apparently, transfer buffer is not the source of the nonspecific interaction.

E: Nitrocellulose Membrane w/ Transfer Buffer, w/o SDS, w/ TBS

This experiment was conducted to determine if the use of the PVDF membrane was the source of the nonspecific labeling. The protocol is the same as that employed in C, except nitrocellulose membrane is used. This time, they

are also loaded with a spot of re-suspension solution and re-suspension solution with PBS. This will test whether or not there is something in these reagents that is causing the nonspecific labeling. The results are shown in Figure 5.5. The use of nitrocellulose membrane did not make a difference. Again, the blot only exposed to secondary antibody developed color intensity comparable to those exposed to both primary and secondary antibodies. Neither the JM 109 spot, the re-suspension solution spot, or the re-suspension solution plus PBS spot developed color. The nonspecific labeling is not due to the reagents in the re-suspension solution or PBS.

F: Nitrocellulose and PVDF, PBS

This experiment is conducted differently. Two strips each of nitrocellulose and PVDF are prepared and spotted with 1ul aliquots of the cell extracts and re-suspension solution plus PBS. One set of the strips is prepared by soaking them in PBS; the other is prepared by soaking them in the transfer buffer solution used previously without SDS. After the strips are loaded with cell extract, all four are processed the same way. All reagents including blocking, antibody, and rinsing solutions contain PBS instead of TBS as the primary buffer. All four of the membrane strips are only exposed to secondary antibody. The results after 20 of development are shown in Figure 5.6. This time, none of the blots developed significant background. The use of PBS instead of TBS appears to have eliminated the nonspecific labeling. There is no significant difference between the nitrocellulose and PVDF membrane. Exposing one of

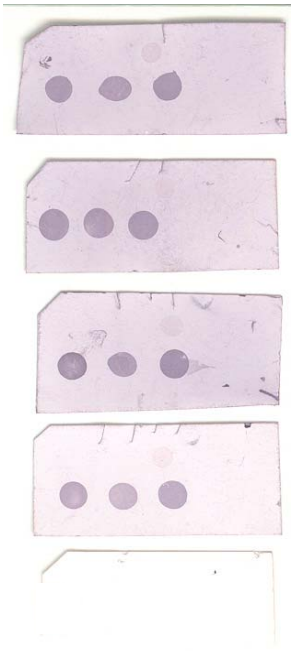


Figure 5.5. Dot Blots, Using Nitrocellulose Membrane From the top, the membranes are treated with the following antibodies: anti-024, anti-025, anti-CD4 commercial, secondary antibody only (alkaline phosphatase conjugate), and neither. The spots are from left to right: H9, (+) Jurkat, (-) Jurkat, BSA.

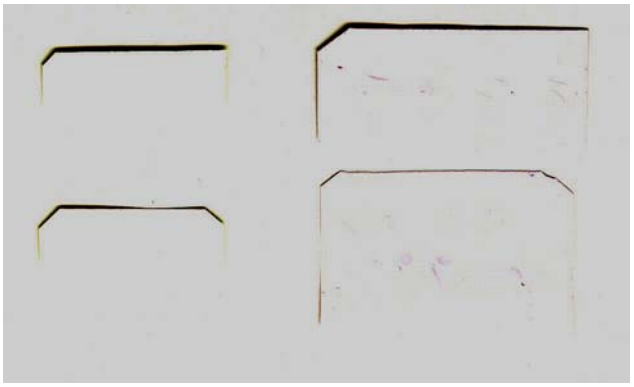


Figure 5.6. Dot Blots, PBS versus Transfer Buffer. Left column= nitrocellulose. Right column= PVDF. Top row =transfer buffer. Bottom row= PBS. The same spot pattern was used as in Figure 5.5.

each pair to transfer buffer containing tris early in the protocol, before they were loaded with sample, does not appear to have made a difference either.

G: PVDF, with PBS

The dot blotting experiment is repeated once more. Again, five strips are prepared with cellular extract as before, and are processed using reagents in which PBS replaces TBS as the primary buffer. However, they are all initially exposed and soaked in transfer buffer up until they are loaded with sample. The results after 12 hours of development are shown in Figure 5.7. The difference between this and earlier dot blots is remarkable. Even after 12 hours of development versus the typical 20 minutes, the amount of nonspecific labeling is not extreme. However, general background color development makes further development of the blots unproductive. However, there is still no appreciable difference between the blots exposed to both primary and secondary antibody and those only exposed to secondary antibody.

Western Blotting

Results and Discussion

Western blotting experiments were carried out in conjunction with the dot blots. As stated previously, the Western blotting technique has the potential to identify an exact antigen within the cellular extract mixture that the antibody labels. The protocol used for probing the membranes with antibody changed significantly as efforts to label CD4 in the blot continued.

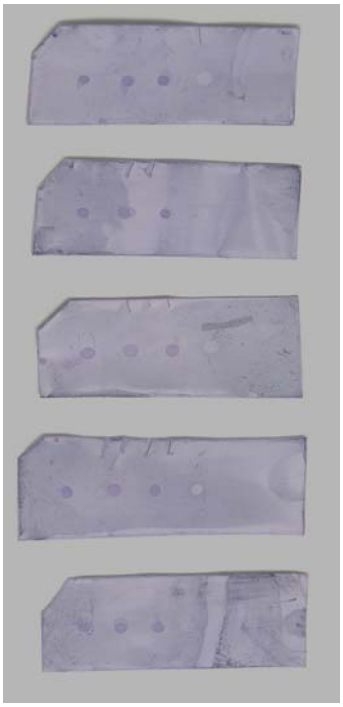


Figure 5.7. Dot Blots Using PBS as Primary Buffer. From the top, the membranes are treated with the following antibodies: anti-024, anti-025, anti-CD4 commercial, secondary antibody only (alkaline phosphatase conjugate), and neither. The spots are from left to right: H9, (+) Jurkat, (-) Jurkat, BSA.

SDS PAGE Western Blotting, TBS

Approximately 210ug of cell extract protein was loaded into each lane of the gel. Initially, TBS is used in all reagents instead of PBS. In order to reduce the non-specific labeling, 0.2% Tween was added to all reagents including the antibody solutions. 3% BSA was also used in the blocking solution, a high percentage. This protocol was repeated with a blot using the anti-024, anti-025, anti-CD4 commercial antibody, and a control blot using only secondary alkaline phosphatase conjugate antibody. Representative blots are shown in Figure 5.8. All four of these blots were developed for 1 hour, 45 minutes. The background color is much improved in this series of blots, and dot blots verified that the difference is mostly due to the absence of SDS in the transfer buffer. The increased rinsing time was apparently not responsible. There is no band in any of the four blots that is conclusively identified as CD4 in any of the four blots. Unlike the previous results, there is also no band labeled in the anti-025 blot that is darker than similar bands in other blots. More disappointing, the labeling in the blot using only secondary anti-mouse alkaline phosphatase conjugate looks similar to the labeling in all three of the other blots. The blots were developed much longer than the suggested 20 minutes, and perhaps they were developed so long that background labeling was inevitable. However, when checked at 10 and 20 minutes, the blots looked similar to the way they do after 1-2 hours of development, just lighter in color. The first bands to appear are the dark ones labeled as A in Figure 5.8. If the blot development were ended at 20 minutes, this band would be one of the only visible, and the darkest. Although this band is

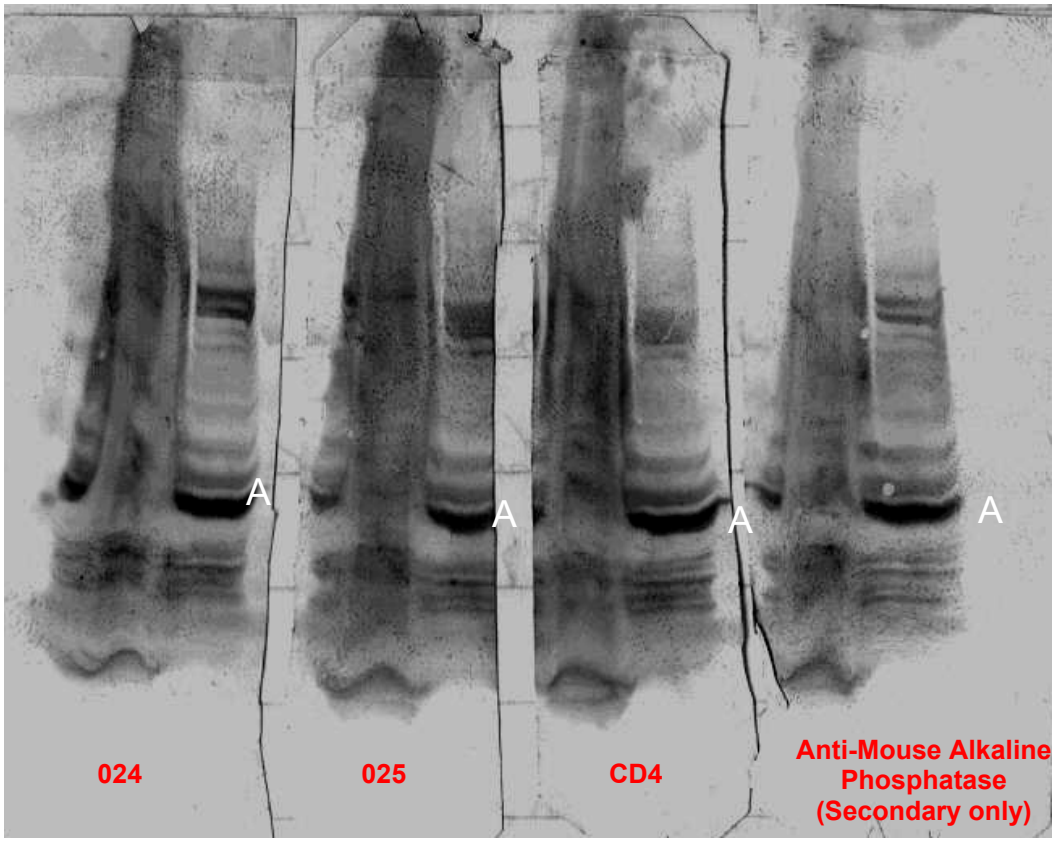


Figure 5.8. SDS PAGE Western Blot, A is approximately 42,000 Daltons

close in molecular weight to CD4 (55KD), the band is also labeled in the blot when only secondary antibody was used. The extended development times of the blots was chosen as apposed to a shorter development time for several reasons. First, as stated previously, if developed for only 20 minutes, the only band clearly visible is A in Figure 5.8, with the rest of the blot looking smeared with very light bands. Second, it was hoped that even with all of the background labeling, the CD4 band would stand out after such a long development, especially when compared to the blot that was only labeled with secondary antibody. Unfortunately, this result was not realized. Instead, there are a couple of possible explanations for the results of the blots so far. First, SDS gel electrophoresis of the extract may ruin the proteins for blotting when using the Sigma # C-1805 anti-mouse IgG alkaline phosphatase conjugate antibody. This may still be a problem, but the dot blotting experiments all but disprove this possibility. The antibody may naturally bind to a negatively charged epitope, and the presence of SDS coating the polypeptide chains of the cell extract may render the labeling of this antibody almost completely nonspecific. SDS coating a peptide chain may reproduce the epitope recognized by this antibody to the extent that the epitope becomes ubiquitous. The second possibility is that CD4 is rare in these cell extracts, and the detection of it using this Western blotting technique is simply beyond the sensitivity limits of this assay. This explanation is consistent with the dot blotting experiments. When developing the blots for long periods of time, the background labeling is just as strong as the specific labeling because the target is rare, and the sensitivity of the assay is too low.

Native Gel Western Blotting, PBS

Following the discoveries from the dot blotting experiments, native gel electrophoresis was also done followed by blotting and antibody labeling. Approximately 200ug of cell extract protein was loaded into each lane of the gel. The protocol closely follows dot blotting protocol G. Again, a 4-12% gel is used, but this time a tris buffer with no SDS is used as the running buffer. In addition, the loading buffer contains no SDS during these experiments. The electrophoresis and blotting proceed as in the SDS gel electrophoresis experiments. In order for the proteins to be charged so that they have some electrophoretic mobility, the pH of the running buffer is 8.7. In order for the protein to migrate towards the positive pole of the gel, it must carry a negative charge. Therefore, a gel of this pH will be useful in separating proteins that have a pI below 8.7, and preferably below 8.0. Several Western blots of native gels of this type were performed using this procedure. In addition to the cell extract supernatant and molecular weight markers, a small amount of 024 and 025 peptide preparations are loaded onto the gel as a control. Western blots performed in this manner are shown in Figure 5.9. The experiment is repeated using only anti-CD4 commercial antibody, and the result is shown in Figure 5.10. None of the blots clearly labeled CD4 above the level of background. However, a smeared band is clearly visible in the 025 lane of the blot labeled using anti-025 antibody. That this band is clearly visible is verification that the overall Western blotting protocol employed is satisfactory, if not sufficiently sensitive to detect the CD4 in the cell extract lane.

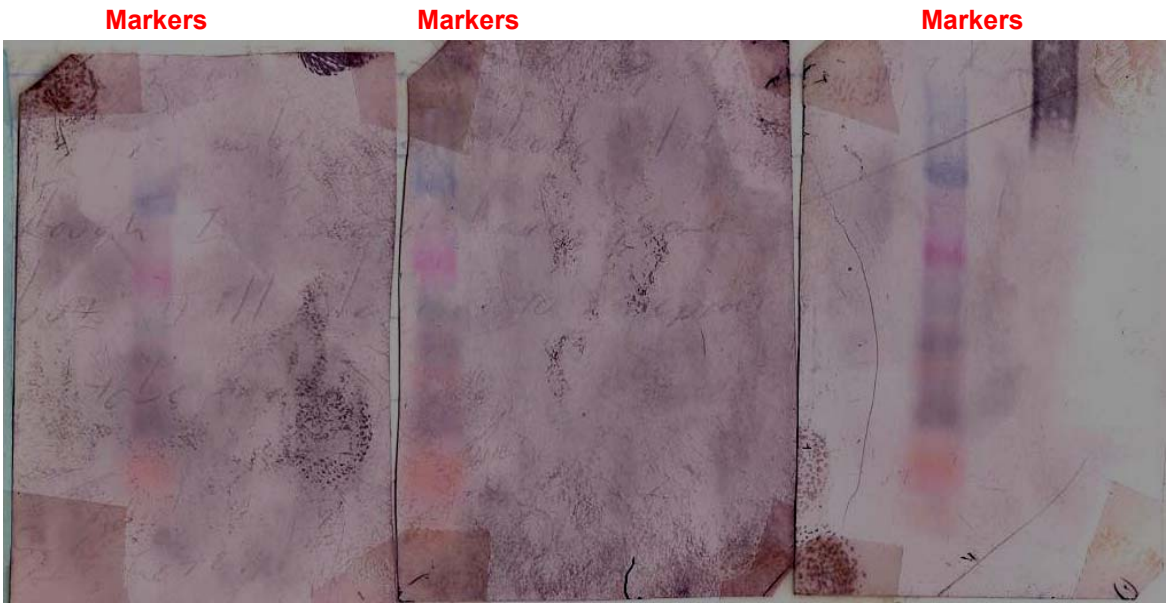


Figure 5.9. Native PAGE Western Blot, The colored bands are molecular weight markers. The markers are, from the top of the blot to the bottom; Blue 204KD, Magenta 124KD, Green 83KD, Violet 42KD, Orange 32KD, Red 18KD.

Figure 5.10. Native PAGE Western Blot. The colored bands are molecular weight markers. The markers are, from the top of the blot to the bottom; Blue 204KD, Magenta 124KD, Green 83KD, Violet 42KD, Orange 32KD, Red 18KD.

The absence of a band in the 024 lane of the blot labeled using anti-024 antibody can be readily explained. The calculated pI of the 025 peptide is 6.6; not counting the contribution of the poly lysine backbone that is used to link four of the monomers together. The pH of the running buffer is sufficiently higher than 6.6 so as to impart a significant negative charge onto the 025 peptide. However, the calculated pI of the 024 peptide is 8.3, just 0.4 pH units below the pH of the running buffer. This charge is probably not sufficient to cause the 024 peptide to move into the running gel, and insufficiencies of the method used to calculate the pI can easily be incorrect by this margin. If the true pI is above 8.7, the peptide would not only fail to move into the gel, it would migrate out of the gel the wrong direction. That nothing is clearly labeled in the cell extract lanes is disappointing. However, there appears to be a very faint band labeled in the cell extract lane of the anti-025 and anti-CD4 commercial blots, Figure 5.9. If this is in fact a labeled band, it is close in molecular weight to CD4, but appears to be lighter. However, background color prevents a positive identification, particular with the anti-CD4 commercial antibody labeled blot, as there is a large smudge of background color in the region where this faint band appears. Developing these blots for a longer time period would not be useful because of background color. The background color of the blot may be increased because PBS was used in the reagents for this experiment, and is known to increase the background of an alkaline phosphatase labeled reaction when the Western Blue reagent is used. Nothing is clearly labeled in the cell extract lanes.

Conclusion

The ability to label CD4 using the Western blotting technique on the cellular extracts is beyond the sensitivity limits of this assay. The nonspecific labeling by the secondary antibody was greatly reduced by using reagents in which PBS was the primary buffer as opposed to TBS. As stated previously, the fact that the 025 antigen sample was successfully labeled is strong evidence that the overall technique is satisfactory. The Tween concentration was greatly reduced in the native Western blotting protocol, so this should not have masked any potential CD4 present in the sample.

Enzyme-Linked Immunosorbant Assay (ELISA)

Introduction

The ELISA assay, similar to Western blotting, takes advantage of the ability to screen for primary antibody labeling using a second antibody that is tagged. The target substance is immobilized onto a solid matrix, usually a polystyrene well, and is probed using an antibody prepared against the desired target. After rinsing to remove unbound antibody, the target is probed using a second antibody. The second (secondary) antibody is prepared against the first antibody so that the primary antibody is the antigenic target of the secondary antibody. In addition, the secondary antibody is labeled with some sort of tag that renders its presence visible to the researcher. The tag is an enzyme that can participate in a colorimetric reaction, typically alkaline phosphatase. After rinsing again to remove unbound antibody, the reagent for the colorimetric reaction is added to the well, and the absorbance of the solution in the well is

monitored and correlated with enzyme activity and therefore the amount of secondary antibody and in turn primary antibody and in turn target substance that is bound to the well. ELISA has some key advantages and some limitations. ELISA can be performed in a very small well in a polystyrene plate making it possible to carry out many ELISA experiments simultaneously on the same small plate. This makes ELISA a very efficient assay in terms of reagent use. In addition, ELISA is also typically a rapid assay, which combined with the ability to perform the assay in a very small well using a minimum amount of reagent makes it very fast and efficient. ELISA is typically performed in a 96 well polystyrene plate, so 96 separate experiments can be carried out simultaneously making it possible to screen a large number of targets and try different assay conditions very quickly and efficiently. The disadvantage of ELISA is that it is unable to identify a specific target within the mixture that is used to coat the well of the plate. ELISA can only identify the presence of a substance within the mixture that the primary antibody recognizes.

To review, the anti-024 antibody was prepared against the cytoplasmic tail of CD4 in the region before the UGA codon. This antibody is a positive control, as it should interact with a region of CD4 that is known to be expressed. The anti-025 antibody was prepared against a region that is expressed only if the UGA presumed to end translation of CD4 is read through in some manner. This antibody should label the cell extracts only if CD4 does not end at the UGA codon, and this region is transcribed into an extended part of the cytoplasmic tail of CD4.

Materials

Immulon 4-HBX 96 well polystyrene plates used for the assays, nitrophenylphosphate tablet sets, anti-mouse IgG alkaline phosphatase conjugate antibody #A-1293, monoclonal anti-human CD4 clone Q4120 #C-1805, and 1L prepared PBS powder sets were obtained from Fisher Scientific (P-3813). Tween 20 was obtained from BioRad. BSA was obtained from Sigma. Para-nitrophenylphosphate assay solution was prepared using tablet sets obtained from Sigma. Anti-024 and anti-025 antibodies were prepared as described in the Monoclonal Antibody Preparation chapter. Cell extracts were prepared using H9, Jurkat E6-1, and Jurkat CD4- cells obtained from the ATCC and JM-109 bacterial cells obtained from Promega. Cell lysis buffer containing PMSF was prepared as described in the Western Blotting chapter. A sonic dismembrator 60 from Fisher Scientific was used to disrupt the cells, and a Multiskan Ascent ELISA plate reader was used to record the absorbance of the assay solutions at 405nm.

Methods

Cell extracts were prepared as outlined previously. Human cell cultures were maintained at 37C in 5% CO2. H9 cells were grown in DMEM media with l-glutamine, sodium bicarbonate and supplemented with 10% NCS. Jurkat cells, both CD4+ and CD4- were grown in RPMI 1640 media containing l-glutamine, sodium bicarbonate, and supplemented with 20% NCS. JM109 E.coli. cells were maintained in LB media at 37C with shaking. Cells were grown in suspension and harvested periodically after reaching a density of about 1X106 cells/ml for

human cell lines, near maximum density for JM109 bacterial cells. Cell cultures were pelleted 50ml at a time by centrifugation at 3500XG for 5 minutes. The supernatant was discarded, and the pellet rinsed using 200ul of ice cold PBS. The cells were pelleted again by centrifugation at 3500XG for 5 minutes and the supernatant was discarded. The pellets were stored at –80C until needed. After thawing the pellets on ice, they were weighed and resuspended in a volume of lysis buffer relative to the weight of the pellet so that the extract is standardized according to the weight of the pellet. 100ul of resuspension solution containing 20mM sodium phosphate, 5mM EDTA, 5mm DTT, and 2mM PMSF pH 7.0 was added for each 46mg of cell pellet weight. The pellet suspended in the lysis buffer was sonicated for 60 pulses of about 0.5 seconds each on ice. Cell disruption was verified visually using the microscope. To prepare cell extract supernatant from the whole cell extract, the whole cell extract was centrifuged for 10 minutes at 14,000XG at 4C. The supernatant was carefully removed and saved.

Much of the ELISA Protocol is adapted from “Antibodies-a Laboratory Manual”.[85] ELISA plates were prepared as follows: 50ul of target antigen solutions including cell extracts were placed into the appropriate wells and the plate was incubated at 4C overnight. The antigen solutions were aspirated out of the well and replaced with 100ul of 1% BSA in PBS. After incubating the plate at 4C for 1 hour, the BSA solution was removed and the wells were rinsed with 200ul of PBS 0.2% Tween. 50ul of the appropriate primary antibody solution was added to each well, and the plate was incubated at room temperature for 1 hour

at which time the primary antibody solutions were removed, and the wells rinsed with 200ul of PBS 0.2% Tween. 50ul of 1:600 diluted anti-mouse Ig alkaline phosphatase conjugate was added to the wells of the plate, and it was incubated at room temperature for 1 hour. After rinsing the plate twice in 200ul PBS 0.2% Tween, 50ul of the alkaline phosphatase reagent containing para-nitrophenylphosphate (prepared using Sigma Fast tablet sets) was added to each well to begin the assay. Alkaline phosphatase reacts with the nitrophenylphosphate yielding a yellow solution. The absorbance of the solution within each well was monitored and measurements were made periodically at 405nm, as soon as 5 minutes, and as late as 24 hours after the start of the reaction.

The Lowry method is used to measure the protein concentrations of extracts and antibody solutions.[87] A standard curve is first constructed using 5ug, 10ug, 25ug, 50ug, 75ug, and 100ug samples of BSA. After 30 minutes, the absorbance of each sample is measured at 750nm. The absorbance at 750nm is graphed versus the micrograms of BSA in the sample, and an equation describing this relationship is derived. The concentrations of the extracts and antibodies are determined by solving the equation for micrograms of protein starting with the absorbance at 750nm of the samples.

Results and Discussion

Anti-024, anti-025 antibody screen

The wells of an ELISA plate were coated as outlined above using either 024 or 025 antigen at a concentration of 10ug/ml in PBS. The wells were then

screened using each of the anti-024 and anti-025 antibodies developed during monoclonal antibody production that showed promise. The results after 10 minutes are shown in Table 5.1. The primary antibody designations are at the top of each column. The primary antibody solutions were used straight, with no further dilutions. However, because of favorable results during monoclonal antibody development, some of the antibodies were grown in larger quantities and were more concentrated than others. The ones undergoing this process were the 10C12, 10D6, 8A5, and 8B8. The first four antibodies, 5C2, 14H11, 10C12, and 10D6 were developed against the 024 antigen corresponding to the known region of the CD4 cytoplasmic tail. The remainder was developed against the 025 antigen corresponding to the potentially translated region past the UGA codon. The 10D6 antibody and 8A5 antibody appear to be the most effective against the 024 and 025 antigens respectively.

Whole Cell Extract, all antibody screen

Cell extracts were prepared using the methods outlined above. The weight of the cell pellets isolated and amount of re-suspension solution used as lysis buffer are shown in Table 5.2. The whole cell extract was used in these experiments without first centrifuging to separate it into pellet and supernatant. An ELISA plate was processed as outlined above using 50ul aliquots of the whole cell extract solutions diluted 1:100 in PBS were used as the target antigen solutions. The whole cell extracts were then screened using 50ul aliquots of each of the monoclonal antibody solutions tested above. The absorbance of the wells at 405nm after 2 hours is displayed in Tables 5.3 and 5.4. Wells screened

Table 5.1. 024-025 Antibody Screen, 10 minutes, Averages

Antibody					
024	5C2 0.089	14H11 0.093	10C12 0.714	10D6 1.635	18G7 0.187
024	3F5 0.081	18C2 0.080	8A5 0.282	10A11 0.112	8B8 0.075
025	5C2 0.061	14H11 0.068	10C12 0.068	10D6 0.060	18G7 0.074
025	3F5 0.069	18C2 0.106	8A5 0.945	10A11 0.109	8B8 0.186

Table 5.2: Whole Cell Extract Preparation

Cell Type	Cell Pellet Weight (g)	Re-suspension Solution (ul)
H9 (+)	0.046	100
CD4+ Jurkat E6-1	0.112	243
CD4- Jurkat D1.1	0.320	695

Table 5.3: Anti-024 Antibody-Whole Cell Extract Screen, Absorbance (405nm), 2 hours, Averages

	14H11	5C2	10C12	10D6	Anti-IgG conjugate
H9 (+)	0.199	0.114	0.190	0.184	0.088
(+) Jurkat	0.211	0.116	0.214	0.176	0.084
(-) Jurkat	0.233	0.136	0.197	0.160	0.084

Table 5.4: Anti-025 Antibody-Whole Cell Extract Screen, Absorbance (405nm), 2 hours, Averages

	10C5	18G7	10A11	18C2	3F5	8B8	8A5	Anti-IgG conjugate
H9 (+)	0.110	0.250	0.140	0.177	0.143	0.143	0.227	0.088
(+) Jurkat	0.114	0.264	0.154	0.208	0.149	0.161	0.211	0.084
(-) Jurkat	0.129	0.259	0.170	0.208	0.156	0.161	0.236	0.084

using only the alkaline phosphatase conjugate secondary antibody are used as negative controls.

When compared to the negative control lane where the extracts were labeled using the anti-IgG conjugate antibody alone, it is clear that all antibodies are labeling something in the cell extract mixtures. Special attention was paid to the results of screening using the 10D6 and 8A5 antibodies because these have a demonstrated ability to bind to their target antigen effectively (Table 5.1). However, the 14H11, 5C2, and 10C5 antibodies may also be promising candidates. In the anti-024 antibody screen, the 10D6 antibody labeled the CD4- Jurkat cell line slightly less than the two CD4+ cell lines. However, the 14H11 and 5C2 antibodies labeled the CD4- Jurkat cell extracts most strongly. The trend of labeling the CD4- Jurkat extracts most strongly is most obvious in the anti-025 antibody screen data (Table 5.4). All seven of the anti-025 antibodies labeled the CD4- Jurkat extracts either equally or more strongly than the CD4+ Jurkat extracts.

This unexpected result prompted an investigation into the pedigree of the CD4- cell line D1.1 used in these experiments. This cell line was obtained from the ATCC, and is advertised as CD4- Jurkat. A thorough literature search revealed no references to this cell line. The only reference listed by the ATCC, and the only one that could be found, is US patent #5,474,771.[90] This patent describes the development of this cell line as spontaneously arising from their cultures of CD4+ Jurkat cells. The inventors performed a FACS analysis of the cell line and found that the cell line appeared to be a CD4+ T helper type cell,

with the exception that the cells were not labeled using a commercial anti-CD4 antibody. The cell line was only screened using one anti-CD4 antibody developed against one of the extra cellular domains of CD4. This is not very convincing proof of the absence of CD4 expression in this cell line. A mutation in the region where the epitope for this antibody is located could cause the antibody to fail to bind although CD4 expression may still be present. In addition, a mutation in one of the many enzymes responsible for glycosylation of extra cellular proteins could cause a change in the epitope recognized by the antibody and result in its failure to label the cell. The presence of this type of mutation may explain why the commercial anti-CD4 antibody used in these experiments labels the CD4- cells less than the CD4+ cells because the epitope recognized by this antibody is also located on the extra cellular domain of CD4. Further, cultures of CD4+ Jurkats are known to spontaneously develop subpopulations of CD4- cells when maintained in culture for long periods, not activated for long periods of time, and over grown. Therefore, this CD4- cell line may be a less than ideal negative control.

Cell Supernatant ELISA

The 10D6 and 8A5 antibodies were chosen for use in continued testing. Both the CD4- Jurkat cells and wells screened using only the anti-mouse alkaline phosphatase conjugate antibody are still used as negative controls. For these experiments, the ELISA plates were coated using 50ul aliquots of 1:100 cell supernatant diluted in PBS as outlined in the methods section. The result of one such experiment after 1 hour 15 minutes is shown in Table 5.5. For this

Table 5.5. Cell Supernatant ELISA, 1 hour 15 minutes

	H9 CD4+ 1:100 Supernatant	Jurkat CD4 (+) 1:100 Supernatant	Jurkat CD4 (-) 1:100 Supernatant	PBS
Anti-024	0.181	0.174	0.196	0.132
Anti-025	0.359	0.363	0.371	0.139
Commercial Anti-CD4	0.219	0.221	0.144	0.138

experiment, the primary antibodies were diluted using PBS as follows: anti-CD4 commercial, 1:100; anti-024 10D6, 1:10; anti-025 8A5 1:10. The secondary antibody was diluted 1:600 in PBS. All antibodies clearly labeled the cells in all wells except that the anti-CD4 commercial antibody did not label the CD4- Jurkat cell extract. The anti-CD4 commercial antibody labeled both the H9 and CD4+ Jurkat cell extracts approximately equally. Both the anti-024 and anti-025 antibodies labeled the CD4- cell line slightly more strongly than either of the CD4+ cell lines, although the difference may not be statistically significant. The fact that the CD4- cell line extract was labeled may not be significant after all (see discussion in the Whole Cell Extract, All Antibody Screen). The anti-025 antibody labeled each of the cell extracts more strongly than either of the other two antibodies. That this occurred may be due to several factors that are unrelated to amount of target antigen present. The anti-025 antibody may be more concentrated than the other antibodies as a result of the hybridoma cell culture secreting more of this antibody than the hybridoma cell line secreting the anti-024 antibody. The anti-025 antibody may naturally bind more strongly to its recognized epitope. In addition, the anti-025 antibody may be less sensitive to the assay conditions than the other antibodies. Perhaps more significantly, the anti-025 antibody is IgM whereas the others are all IgG. IgM is made up of a pentamer of antibodies linked by the heavy chains. The epitope recognized by the secondary antibody is the Fab region of the antibody, part of which is similar between IgG and the IgM monomers. Therefore, each IgM antibody has five times the binding sites for the secondary antibody as each IgG molecule. Each

IgM bound to target antigen may be labeled by as many as five secondary antibodies labeled with alkaline phosphatase. Clearly, on a per labeled antigen basis, the anti-025 antibody will result in a stronger alkaline phosphatase secondary antibody response than will either of the other two antibodies used in these experiments. These results are highly suggestive of transcription of CD4 past the UGA codon. However, the possibility of nonspecific labeling by these antibodies is not yet completely eliminated.

This experiment is repeated a number of times using undiluted anti-024 antibody, undiluted anti-025 antibody, and 1:100 diluted anti-CD4 commercial antibody. Because of expense, it is impractical to use a less dilute stock of the commercial anti-CD4 antibody. In addition, coating wells using only secondary antibody was used as a negative control. This will make the data defensible in light of the failed Western blotting procedures because of the nonspecific binding by the secondary antibody in that assay. The anti-024 antibody used is 10C12 instead of 10D6 because it also appeared to be a promising candidate. The results of these experiments are shown in Table 5.6. As opposed to the Western blotting experiments, clearly the primary antibodies are able to bind to something in the cell extract supernatant. The difference between the binding of the primary antibodies together with the secondary antibodies compared to the binding of the secondary antibodies alone is significant.

Table 5.6: Cell Extract Supernatant ELISA Absorbencies at 405nm After 1 Hour (Averages)

	H9 (+) extract supernatant	Jurkat (+) extract supernatant	Jurkat (-) extract supernatant	PBS only
Anti-024	0.144	0.148	0.157	0.112
Anti-025	0.16	0.178	0.184	0.118
Anti-CD4	0.132	0.105	0.088	0.101
AP-Conjugate 2' Only	0.075	0.067	0.068	0.072

Protein Assay and Reagent Standardization

Although the cell extracts have been standardized based upon cell pellet weight, the protein concentration has not been determined. Even though levels of expression of CD4 may not be constant across different cell lines, running the cell extract ELISA experiments using cell extract standardized for protein concentration may yield new insights. In addition, another negative control was found, human adenocarcinoma HeLa cells, which should not be CD4+. Because they are a different cell type (not T-Lymphocytes), comparing the extract solutions by protein concentration may be more relevant than comparing them by cell pellet weight. Therefore, it is necessary to measure the protein concentrations of each of the extract preparations so that they can be standardized. Also, the concentrations of the anti-024 and anti-025 antibody solutions are measured so that the same concentration of each antibody will be used in the next ELISA.

A standard curve of protein concentration using the Lowry method and BSA standards was constructed as outlined in the methods section. The standard curve, and equation derived from it are shown in Figure 5.11. The 1:10 diluted whole cell extracts and antibody stock solutions were also assayed using the Lowry method, and the concentration of each is calculated using the equation derived from the standard curve as outlined in the methods section. The results of the protein assay are shown in Tables 5.7 and 5.8.

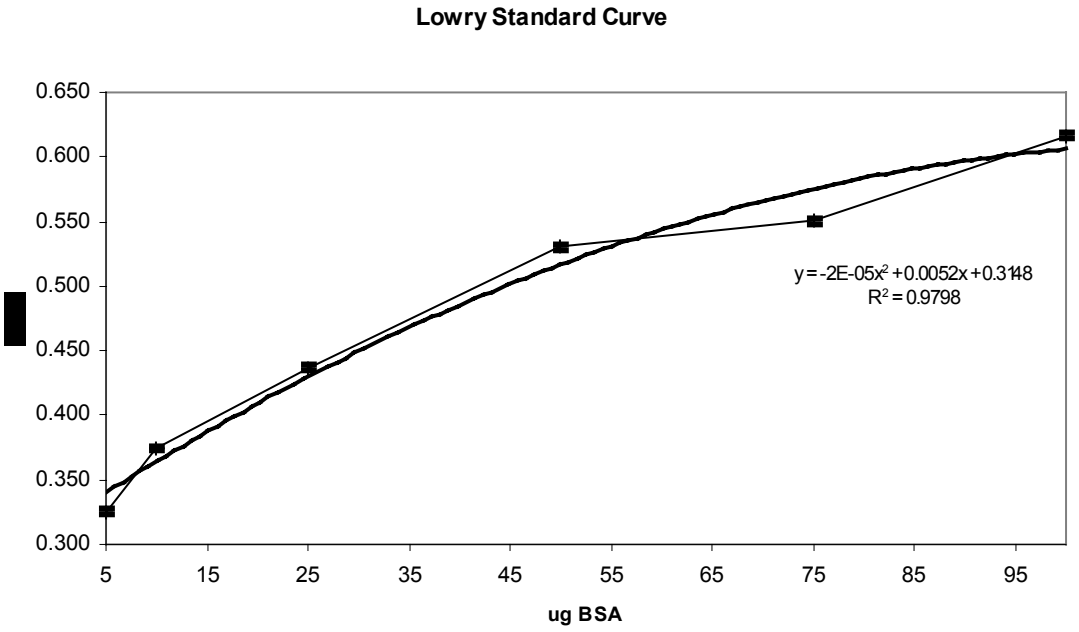


Figure 5.11. Lowry Standard Curve

Table 5.7. Lowry protein assay data. Includes dilutions and amount added to each well of the ELISA plate (50ul). (Averages)

	ug Protein	ug/50ul Stock	Dilution Factor	ug/50ul Standardized Stock
10C12	1374	3435	1.35	2542
10D6	1017	2542	1.00	2542
8A5	1539	3846	1.51	2542

Table 5.8. Lowry protein assay data. Includes dilutions and amount added to each well of the ELISA plate (50ul). (Averages)

	ug Protein	ug/50ul 1:10 Diluted Stock	Dilution Factor	ug/50ul 1:100 Diluted Stock
Jurkat (+) extract	325	325	1.12	29.1
Jurkat (-) extract	606	606	2.08	29.1
H9 extract	707	707	2.43	29.1
HeLa extract	145	291	1	29.1

The antibody stock solutions are very concentrated. They range from approximately 60mg/ml to 80mg/ml. The 8A5 anti-025 antibody solution is the most concentrated of the three antibody stocks prepared against the 024 and 025 peptide antigens. The 10D6 anti-024 antibody solution is the least concentrated. All three are standardized so that the same concentration of antibody is added to each well during the ELISA procedure. In order to maximize labeling, and because there is no shortage of these three antibodies, the only dilution done is that to reduce the concentrations of the 10C12 anti-024 antibody and 8A5 anti-025 antibody solutions to that of the 10D6 anti-024 antibody solution. After preparing standardized samples by dilution, 2.5mg of each antibody will be added to each of the appropriate wells of the ELISA plate. This is a high concentration. The anti-CD4 commercial antibody solution is much more dilute, but because it is very expensive, it is impractical to try and use a more concentrated stock solution. During this experiment, the 1:100 diluted preparation of the antibody solution obtained from Sigma is used as in previous ELISA experiments.

The cell extracts, although standardized for cell pellet weight previously, vary widely in protein concentration. The HeLa cells are adherent cells and were isolated independently from the rest of the cell extracts and were not standardized according to cell pellet weight. The HeLa extract was isolated by Tunde Olubajo and given to me for this experiment. The Hela cell extract is the most dilute followed by the (+) Jurkats, (-) Jurkats and finally the H9 (+) extract. As with the antibody solutions, the extracts are standardized for protein

concentration by diluting the more concentrated samples to match the concentration of the least concentrated sample. After standardization, 29.1ug of cell extract protein is loaded into the wells of the ELISA plate with each 50ul of extract solution.

Standardized ELISA

An ELISA plate is coated using 50ul aliquots of the standardized whole cell extract solutions including the standardized HeLa extract solution and is processed as before except that standardized solutions of the anti-024 10C12, anti-024 10D6, and anti-025 8A5 antibodies are used in place of their un-standardized stocks. Again, wells probed using only secondary AP-antibody conjugate are used as negative controls along with wells coated only with PBS. The results of this experiment are shown in Table 5.9. All four of the antibodies prepared against CD4 labeled the T-lymphocyte extracts significantly more than the HeLa extract or the wells coated only with PBS. In addition, the labeling signal from the antibodies prepared against CD4 on the T-lymphocyte extract was significantly higher than the labeling obtained using only AP-conjugated secondary antibody. Standardizing the protein concentrations smoothed out some of the differences in labeling obtained in earlier experiments, but some slight differences remain. Both the anti-024 10C12 antibody and the anti-025 8A5 antibody labeled the Jurkat (-) extract slightly more strongly than the others. This is suggestive of these two antibodies both labeling an epitope on either the same protein, or separate proteins whose levels of expression follow each other closely. Surprisingly, the anti-CD4 commercial antibody labeled all three cell

Table 5.9. ELISA on whole cell extracts standardized for protein concentration. 1 hour, 15 minutes. Absorbance measured at 405nm. (Averages)

	H9 (+) standardized	Jurkat (+) standardized	Jurkat (-) standardized	HeLa (-) standardized	PBS
Anti-024 10C12	0.159	0.154	0.165	0.130	0.083
Anti-024 10D6	0.130	0.108	0.130	0.098	0.086
Anti-025 8A5	0.151	0.155	0.160	0.112	0.105
Anti-CD4 commercial	0.102	0.096	0.101	0.087	0.090
AP-Conjugate only	0.077	0.079	0.083	0.076	0.069

extracts nearly equally. This is not consistent with many previous experiments, and cannot be accounted for by the dilutions that were done. In fact, the Jurkat (-) extract was diluted more than the rest, as was the H9 (+) extract. The concentration of the Jurkat (+) extract was effectively raised in relation to the other two. Because of the dilutions, this antibody was expected to label the Jurkat (+) extract even more strongly relative to the others than previously, and the Jurkat (-) extract relatively less when compared to previous experiments.

The anti-024 10D6 antibody labeled the Jurkat (-) extract equally well as the H9 (+) extract and more than the Jurkat (+) extract. The relative intensities of labeling of the anti-024 10C12 and anti-025 8A5 antibodies followed each other closely in this experiment. This trend is also found in the unstandardized experiment data in Tables 5.5 and 5.6, but cannot be found in the data from the experiment represented by the data in Tables 5.3 and 5.4. However, Tables 5.5, 5.6, and 5.9 represent three separate experiments, whereas Tables 5.3 and 5.4 collectively represent a single experiment. In addition, the experiment that produced the data in Tables 5.3 and 5.4 was only conducted once, whereas the experiments that produced the data in Tables 5.5, 5.6, and 5.9 were each repeated several times, verifying the trends found in the data.

Statistical Analysis

Replicates of the standardized whole cell extract ELISA experiments were subjected to standard statistical analysis to determine the strength of the confidence interval on the average values obtained. The experimental data included in this analysis were collected during independent experiments, usually

conducted on different days. The results for the standardized whole cell ELISA experiment are shown in Figure 5.12-Figure 5.15. None of the antibodies labeled the HeLa extract significantly more than did the secondary antibody alone, providing evidence of their specificity. The anti-024 and anti-025 antibodies label the T cell extracts more than the secondary antibody only negative control with a 90% confidence interval in all experiments. The only antibody not able to label the cell extract more than this negative control at a 90% confidence interval level was the anti-CD4 commercial antibody. This may be due in part to the very dilute nature of this antibody. However, except for the CD4- Jurkat extract, the anti-CD4 antibody consistently labeled the cell extract more than did the negative control. But the difference is just too small to give statistical confidence to this result without repeating the experiment many more times. Although each of the antibodies labeled the lymphocyte extract more than the HeLa extract, there is slight overlap between the 90% confidence intervals between the T-cell extract and HeLa cell extract samples for the anti-025 antibodies, and extensive overlap using the anti-CD4 commercial antibody. Again, more replicates of the experiment must be conducted in order to improve the statistical confidence.

Conclusion

The results of the ELISA experiments strongly support the hypothesis that CD4 is translated past the UGA codon. In every case with the exception of the anti-CD4 commercial antibody labeling the Jurkat (-) cell extract, the

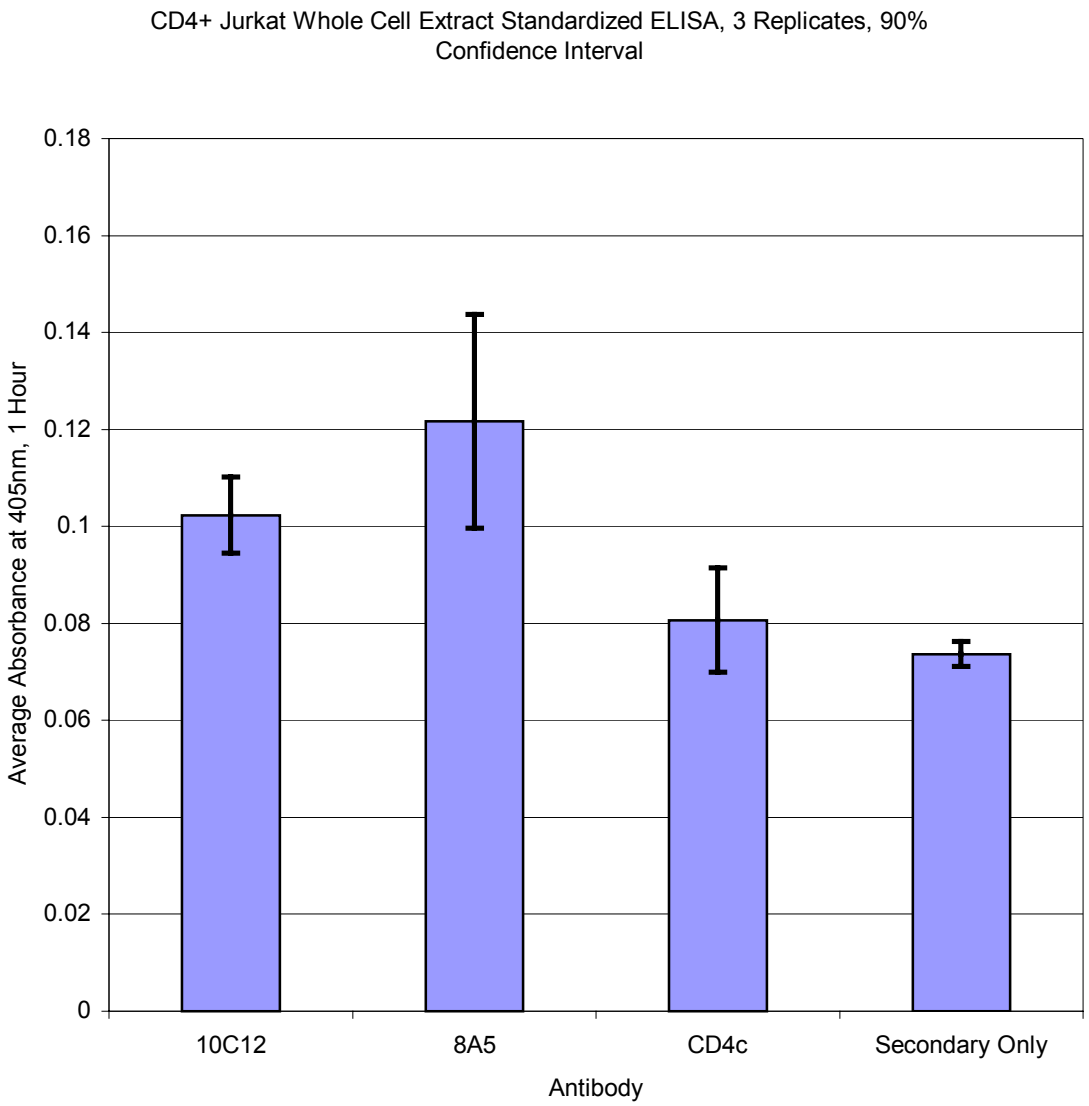


Figure 5.12. CD4+ Jurkat Standardized ELISA

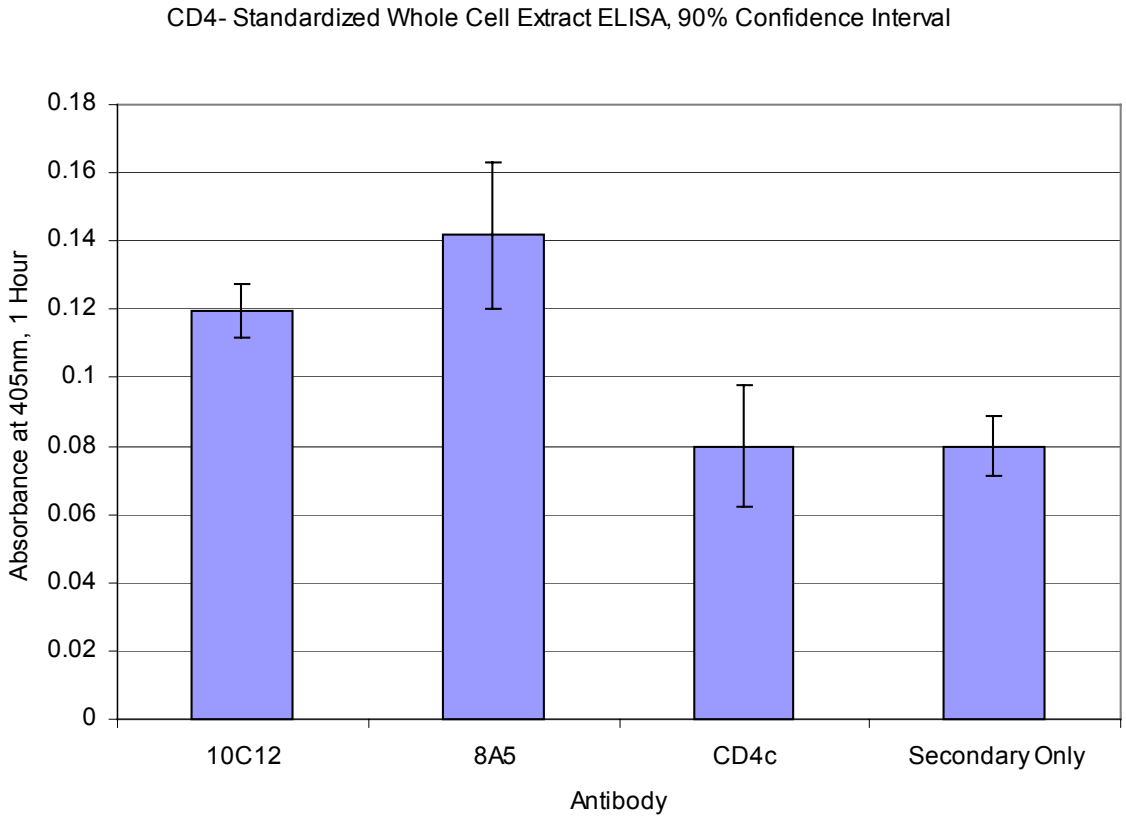


Figure 5.13. CD4- Jurkat Standardized ELISA

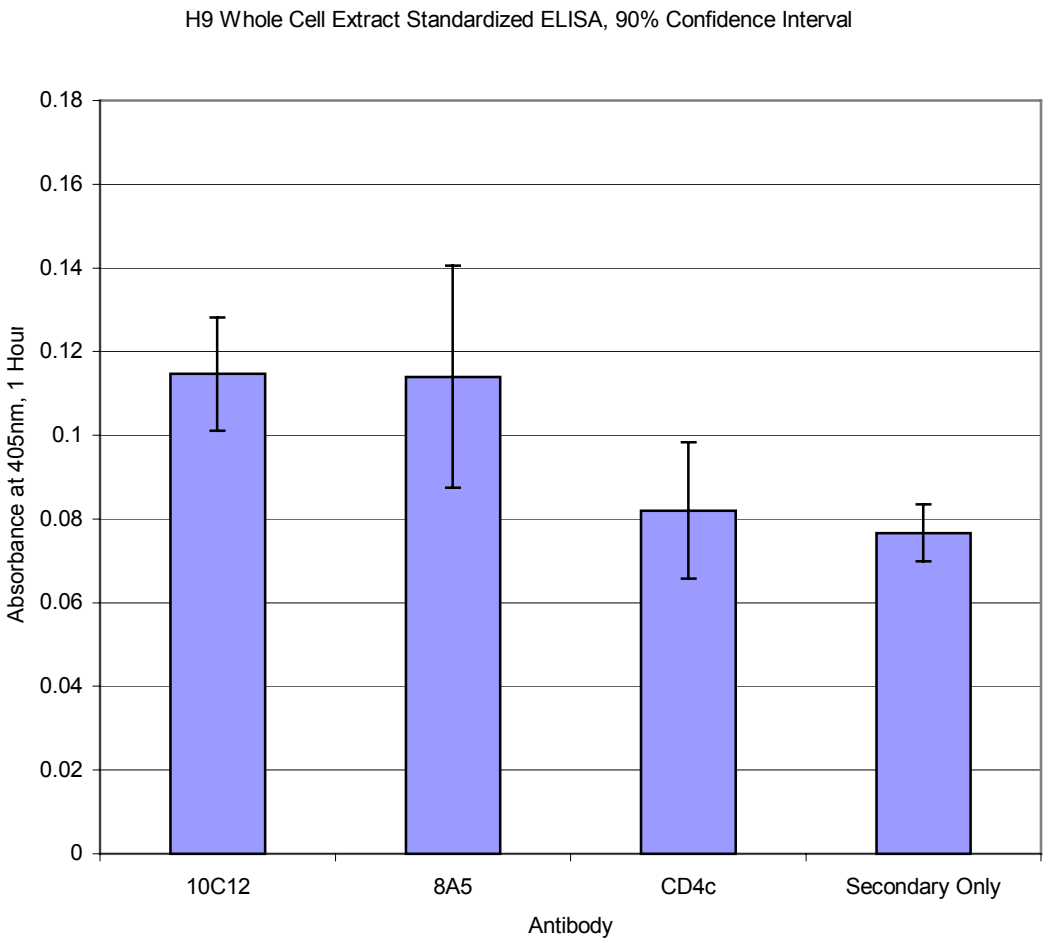


Figure 5.14. H9 Standardized ELISA

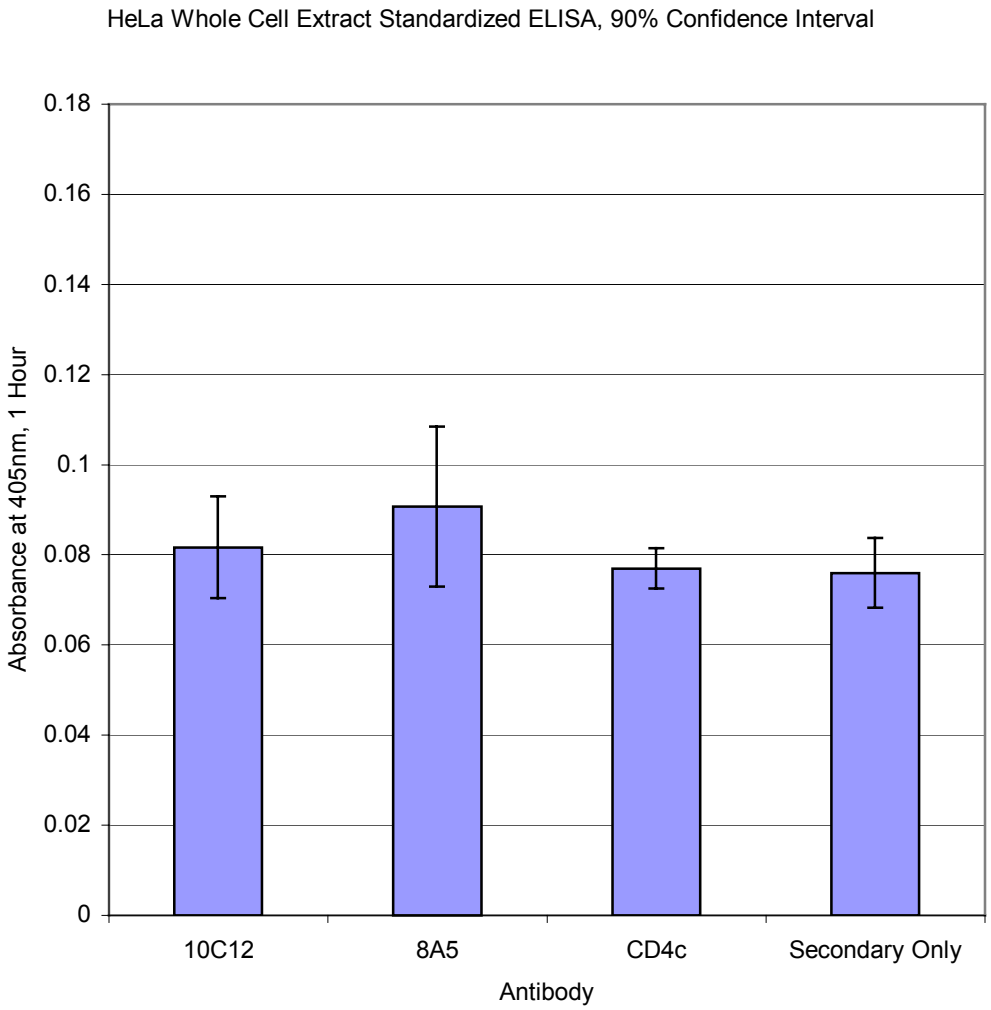


Figure 5.15. HeLa Standardized ELISA

antibodies bound to something in the cell extract producing a signal over and above that observed when AP-conjugate secondary antibody is used alone. Further, the labeling appears to be specific to the T-lymphocyte extract. The antibodies did not label the HeLa cell extract nearly as strongly even though the same protein concentrations of each were used to coat the wells. The statistical analysis provides some support for these observations, except that much of the data is too weak for differences between the 90% confidence intervals because of the low number of experimental replicates used. This is especially true of the data relating to the anti-CD4 commercial antibody. The fact that the difference between labeling the HeLa and T-cell extracts was not resolved at the 90% confidence level does not mean that there is no difference, just that the difference cannot be determined at 90% confidence without repeating the experiment many more times.

The only significant anomaly repeatedly observed, that of labeling the presumed CD4 (-) Jurkat cell extract, can readily be explained by several factors as discussed previously including the presence of a mutant form of the CD4 gene or cellular glycosylation enzyme, and the misidentification of the cell line as CD4(-) instead of a spontaneous subculture of temporarily CD4(-) cells.

As with most experimental methods where very high sensitivity is needed in order to observe a given result, noise and variation in the data can sometimes make trends difficult to find in some experimental data. The inconsistencies, including the unusual labeling pattern of the anti-CD4 commercial antibody in Table 5.9 along with the absence of the connection

between the relative strengths of the anti-024 and anti-025 antibody labeling in the all antibody whole cell extract screen, probably represent noise and variation in the data that can be expected because of the low signal level and sensitivity required to detect CD4 and any potential read through products in this experiment. The trends identified after repeating the experiment several times and several different ways probably represent significant observations.

Coating the wells of the ELISA plate using the whole cell extract as opposed to supernatant did not seem to increase the labeling signal level. The labeling signal might be expected to increase if CD4 molecules were still trapped in the membrane fragments that are contained in the pellet fraction of the sonicated cell extract. However, the membrane fragments may not have attached to the plate efficiently. The polystyrene plates are optimized to bind to protein, not lipid, and the washing method used may have dislodged any membrane fragment that was loosely associated with the plate.

Both the anti-024 and anti-025 antibodies consistently label the Jurkat (-) extract more strongly than the rest of the extracts. This is especially true of the anti-025 antibodies. In fact, the difference between the labeling intensity of the HeLa and CD4- Jurkat cell extract when using the anti-025 and anti-024 antibodies can be resolved even at the 90% confidence interval. The differences in amount of labeling from one type of cell extract to another can be explained by factors other than noise in the data such as differences in the absolute level of CD4 expression in the cell culture. Other factors may also account for this including differences in quality of the cell extract preparations and state of the

cells when harvested. Because the cell normally degrades CD4 by endocytosis, the cytoplasmic tail portion of the receptor would not be exposed to degradation as quickly as the extra cellular domains. This would result in reduced labeling by the anti-CD4 commercial antibody when compared with that of the anti-024 and anti-025 antibodies. In addition, the epitope recognized by the anti-025 antibody may be less fragile, and more stable during degradation of the CD4 receptor.

If all antibodies are labeling CD4, then the labeling signal from each should rise and fall together. In fact, this is observed to a degree. The anti-024 10C12 antibody labels the Jurkat (-) whole cell extract most strongly, and the anti-024 10D6 antibody labels the Jurkat(-) extract more than the Jurkat(+) extract and equal to the H9(+) extract. This trend is still evident in the data from the standardized ELISA experiment in which the amount of protein from each extract that was used to coat the wells of the ELISA plate was held constant. The only antibody that clearly violates this trend is the anti-CD4 commercial antibody. This antibody consistently labels the Jurkat (-) extract significantly less than the CD4+ cell extracts. Although, as discussed previously, there are not enough replicates of the experiments to show a 90% confidence interval difference in the data for this antibody. As also discussed previously, there are serious questions about whether or not this cell line is actually CD4-. However, because this cell line was defined by the fact that it is not labeled by a commercial antibody against a region of the external portion of CD4, it is not

surprising that this commercial antibody that labels a similar region of CD4 also fails to label this cell line.

As a result of economic considerations, the anti-CD4 commercial antibody used was remarkably more dilute than the anti-024 and anti-025 antibodies. The concentration of antibody protein in the anti-CD4 commercial preparation used to probe wells of the ELISA plate was only 0.5ug/ml whereas that of the other antibodies was over 50mg/ml. The dilute nature of the anti-CD4 commercial antibody may account for the decreased signal from labeling using this antibody preparation as opposed to the other antibodies. This may also account for the weak nature of the data relating to this antibody when subjected to statistical analysis. However, the difference in labeling between the different antibody preparations is not proportional to antibody concentration. The anti-024 and anti-025 antibodies although much more concentrated, only produced slightly more labeling. The target antigen is most likely the limiting factor in this analysis, not labeling antibody concentration.

Surprisingly, the labeling intensity of the anti-025 antibody was consistently stronger than that from the other antibodies. The very low concentration of the anti-CD4 commercial antibody stock can explain why the labeling of the anti-025 antibody was stronger than that of the anti-CD4 commercial antibody, but does not account for how it could be stronger than the labeling by the anti-024 antibody. For translation of the region that the anti-025 antibody is prepared against to occur, the UGA codon must be circumvented in some way. Each of the possible mechanisms by which this is known to occur are

relatively inefficient and slow. Only if there were a very active SECIS element present directing the read through could the expression level of this region be expected to approach that of the region prior to the UGA codon targeted by the anti-024 antibody. This is still a possibility, although no definitive SECIS structure could be identified. However, as discussed briefly in the results section, this observation may also be explained by the nature of the anti-025 antibodies themselves. All of the anti-025 antibodies are IgM, not IgG like the other antibodies used in these experiments. The secondary antibody used in these experiments, the alkaline phosphatase conjugated anti-IgG antibody, binds to the Fab fragment of the primary antibodies. Although there are significant differences between IgG and IgM antibodies, the Fab region of both are very similar. Although sold as an anti-IgG antibody, the secondary used in these experiments can also label IgM. However, IgM antibodies are made up of five individual antibody structures linked by a unique structure near the base of the heavy chains. Each IgM antibody molecule has five binding sites for the secondary antibody whereas each IgG antibody molecule only has one. Each IgM molecule bound to antigen in the ELISA could be labeled by up to five of the secondary antibody molecules resulting in a signal that appears to be much larger than the actual amount of labeling relative to the IgG antibodies. If the labeling signal were divided by five, the amount of labeling by the anti-025 antibodies would be very low indeed and more consistent with what was expected. Perhaps the signal from the anti-025 antibodies would be undetectable if they were IgG.

Whole Cell Labeling Immunofluorescence Microscopy

Introduction

Like the other antibody labeling experiments used, this one also takes advantage of the ability of a secondary tagged antibody to bind to and label a primary antibody that is bound to target antigen. The technique applied to labeling whole cells is shown in Figure 5.1. A review of the function of the primary antibodies is useful. The anti-CD4 commercial antibody was developed against the extra cellular portion of the CD4 receptor between loops one and two and is used as a positive control. The anti-024 antibody was developed against the cytoplasmic tail of CD4 just prior to the UGA codon, and is also used as a positive control. The anti-025 antibody was developed against a region of CD4 that is only produced if the UGA is read through. The anti-025 antibody is used to determine whether or not the first UGA codon in the 3' region of CD4 is read through in vivo. For these experiments, fluorescent secondary antibodies are used to label the primary antibodies so that the position of the primary antibodies can be located using a fluorescent microscope.

The advantage of performing whole cell labeling experiments visualized using immunofluorescence microscopy is that the intracellular location of the target antigen can be determined. Whereas the ELISA experiments were only able to determine that the antibodies recognized an antigen present within the T-cell extract, this technique has the potential to provide another level of verification that the antibodies are actually labeling CD4. This is particularly important since the Western blotting technique was unable to localize the antibody labeling to a

particular protein within the cell extract mixture. Although it is impossible to prove that the antibodies do not bind to any antigen other than CD4 and it's potential read through product, any additional evidence that they are binding to CD4 is welcome. Because CD4 is a cell surface receptor, the labeling should be localized on the cell membrane. The exception to this is if the cytoplasmic tail of CD4 is cleaved under certain circumstances allowing the antigen recognized by the anti-024 or anti-025 antibodies to diffuse through the cell cytoplasm. Other antibodies have been developed against the cytoplasmic tail of CD4, but no reference could be found where these were used for cell labeling experiments. Even if this occurs, some other localization may be identified. However, no other experimental or theoretical evidence supports this possibility.

For labeling, the cells are processed in a similar fashion as the whole cell extract is processed during ELISA, with particular changes incorporated for using whole cells, and changes associated with differences in the secondary antibody tag and the way the labeling is visualized. First, the cells are first collected and immobilized onto a glass slide. After fixing the cells with paraformaldehyde, nonspecific binding sites are blocked using BSA. After this, they are exposed to a solution containing the primary antibody. After rinsing, they are exposed to the appropriate secondary antibody. After further rinsing and sealing with coverslips, the cells are viewed using a fluorescence microscope. The position of the secondary antibodies, which should correlate with the position of the secondary antibodies, is visible as a glowing region.

Labeling the extra cellular portion of CD4 with the commercial antibody should be relatively straightforward. However, labeling the intracellular region of CD4 may present a challenge. In order to label the intracellular portion of CD4, the antibodies must efficiently pass through the cells membrane. Triton X-100 is used to facilitate this by creating holes in the cell membrane either during or after fixation using paraformaldehyde.

Other special problems that must be monitored include potential sensitivity of the antibody to paraformaldehyde fixation of the target epitope, and because the anti-025 antibody is IgM, potential inability of an anti-IgG secondary to label the anti-025 antibody. The anti-CD4 antibody is reported to be sensitive to formalin fixation of the target antigen meaning that if the cells are formalin fixed, the antibody is unable to label the cells. Formalin is a 30% solution of paraformaldehyde, which is much higher than the 2-5% that is used in these experiments. Hopefully, the epitope recognized by the anti-CD4 commercial antibody is not destroyed by fixation in such low concentrations of paraformaldehyde. Obviously, no information is available on the potential sensitivity of the anti-024 or anti-025 antibodies.

Materials and Methods

Materials

Human T cell cultures used for labeling experiments were obtained from the ATCC including the D1.1 Jurkat CD4- cell line (please see the Conclusion of the ELISA section for a discussion on the CD4 status of this cell line), the E6-1 Jurkat CD4+ cell line, the H9 CD4+ cell line, and the CEM CD4+ cell line. The

anti-CD4 commercial antibody (Clone 4120), the anti-IgG (Fab) FITC conjugate antibody, the anti-IgG (whole molecule) FITC conjugate antibody, the anti-IgM Cy-5 conjugate antibody, cell media, and the reagents used in buffer preparation were obtained from Sigma. Paraformaldehyde was obtained from JT Baker. Cell culture flasks were obtained from Fisher.

In order to prevent the cells drying out during the labeling process, a humidor is constructed using a styrofoam cooler. Water is kept in the reservoir in the bottom of the humidor, and the cells are placed along the shelf constructed using a microcentrifuge rack while they are being processed.

For viewing the labeled cells, two microscopes were utilized. One is a Nikon Fluorophot using a B2A filter set to view cells labeled with an FITC conjugated secondary antibody. The other microscope is a Biorad 600 confocal laser microscope with a krypton-argon laser. The confocal microscope produces better resolution than the standard fluorescence microscope, and also produces digital images, which are easier to manipulate in a quantitative fashion.

Methods

Cells were grown in media recommended by the ATCC at 37C and 5% CO2 as in previous experiments. Cell cultures were not allowed to exceed about 1X106 cells/ml before adding or changing the media.

Slides were prepared as needed by cleaning with ethanol and placing a 20ul drop of 0.1% poly-L-lysine onto the slide where cells are intended to adhere. The drop of poly-L-lysine remains on the slide without drying out for 1 hour. The slides are then rinsed in de-ionized water and allowed to dry.

Cells were harvested as needed by removing a 20-30ul aliquot of the culture and placing it onto the slide where the poly-L-lysine was placed. The cell culture aliquot remains on the slide for 30-45 minutes to allow the cells to settle onto the poly-L-lysine coated slide. The slides are then gently rinsed with PBS to remove the media either by dipping the slide or removing the media with a pipette and rinsing with a 20-30ul aliquot of PBS. The PBS rinsing solution is removed, and 20-30ul of paraformaldehyde fixative consisting of 3.7% paraformaldehyde, 0.2% triton x-100 in PBS is added. The triton x-100 may also be contained in a separate solution, and the cells treated with it separately after paraformaldehyde fixation. The cells are fixed for 12-15 minutes before the fixative is removed and the cells rinsed by either dipping the slides in PBS or removing the fixative with a pipette and rinsing with a 20-30ul aliquot of PBS. After the PBS rinsing solution is removed, the slides are either placed into a solution of 1% BSA in PBS or a 20-30ul aliquot of the BSA solution is added to the cells on each slide. The cells incubate in the BSA blocking solution for 1 hour before it is removed and they are rinsed as before with PBS. Then, a 20-30ul aliquot of the appropriate primary antibody is added to each slide. The slide is incubated at room temperature for 1 hour in the humidor without drying out. The cells are rinsed again with PBS as before except that the PBS is left on the cells for at least 15 minutes. A 20-30ul aliquot of the secondary antibody is then added to the cells, and they incubate at room temperature for 1 hour in the humidor without drying out. The cells are rinsed twice again with PBS as before

except that the first rinse lasts about 15 minutes, and the second rinse lasts 30-45 minutes.

The cells are then covered with coverslips, sealed with VALAP and viewed using the fluorescence microscopes. Various images of the cells are collected and compared for labeling pattern and intensity.

Results and Discussion

The methods described above were arrived at only after much careful experimentation. A few of those experiments are mentioned here for completeness. First, after several references were consulted, the ideal concentration of the paraformaldehyde fixative was determined to be 3.7% for 12-15 minutes. In arriving at this conclusion, concentrations of paraformaldehyde from 1-5% were tried. The fixed cells were compared to normal cells, and the highest concentration of paraformaldehyde that did not change the cells was used. The amounts of each reagent added to the cells on the slide were determined by trial and error. Using amounts smaller than the 20-30ul aliquots made it difficult to prevent the cells drying out. Using amounts larger than this was difficult because the aliquot would spread across the slide and drip off the edge. Many different dilutions of the antibodies were tried, from 1:1 to 1:10,000. Generally, the 1:10-1:100 dilutions were best.

H9 Cell Labeling

The H9 cells were difficult to work with. Initially, difficulties in immobilizing them to the slides were the problem. Later, they were difficult to identify on the slide because of their irregular shape (Figure 5.16). The

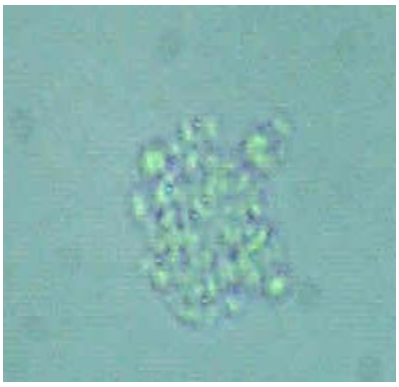


Figure 5.16. 80X Unlabeled, Unfixed H9 Cell; 10um

presence of some dust or dirt on the slide was unavoidable, and because these cells resembled the typical dirt so closely, their use in the labeling studies was ended.

Jurkat Cell Labeling

The Jurkat cells were easier to maintain and visualize than the H9 cells (Figure 5.17). However, the labeling experiments were unable to demonstrate membrane-labeling levels over and above the control labeled only with secondary antibody. As an example, the Jurkat cells were grown and prepared as outlined above using 3.7% paraformaldehyde and 0.2% triton x-100. 1:100 dilutions of antibody were used to label the cells. A comparison of Figure 5.18 with Figure 5.19 reveals that although the labeling appears localized to the membrane when the cells are labeled using the anti-CD4 commercial antibody, the same is true of the cells labeled only with secondary antibody. Labeling using the anti-024 and anti-025 antibodies produced similar results. At the time this experiment was performed, however, I didn't have an appreciation for the slight difference that might be expected between the experimentals and the controls, so the gain and black levels were adjusted to provide the best view of the cells. Therefore, it can only be said that the relative intensity of labeling between the two groups were similar, and no quantitative comparison can be made from their images.

Figure 5.17. Jurkat (-) 80x-2-5-19-03

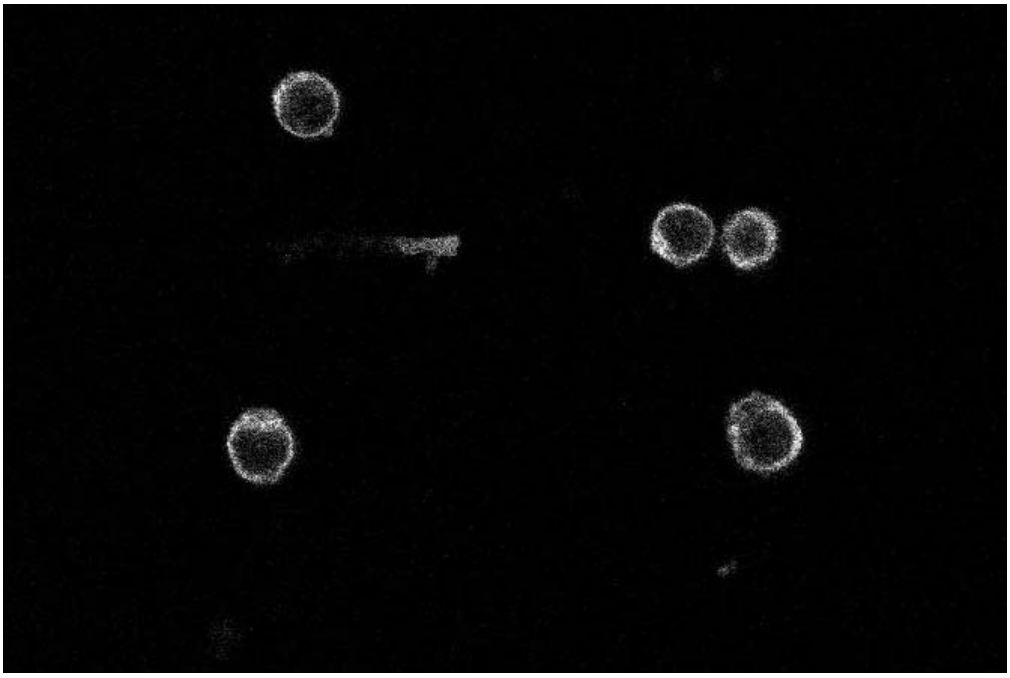


Figure 5.18. Jurkat (+) Labeled Using Anti-CD4 Commercial Antibody

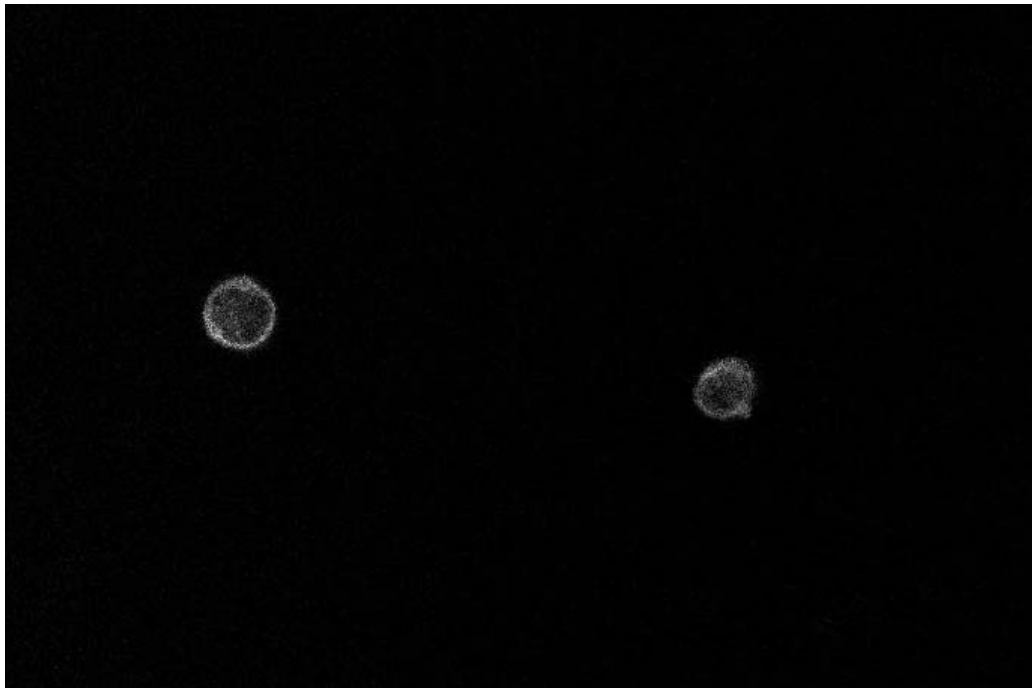


Figure 5.19. Jurkat (+) Labeled Using Only FITC Conjugated Secondary Antibody

CEM Cell Labeling

Because of the inconclusive nature of the results obtained using the Jurkat cells, particularly before the problems with using the CD4- Jurkats as a negative control were identified; CEM cells were used in these experiments.

The CEM cells were prepared as outlined above. 1:10 dilutions of the anti-024 10C12 and anti-025 8A5 antibodies and undiluted anti-CD4 commercial antibody in PBS were used. 1:10 diluted anti-IgG (Fab) FITC conjugate in PBS is used as the secondary antibody. Cells labeled using only secondary antibody are used as the negative control. Cells were imaged using the BioRad 600 confocal laser microscope. The BHS filter set was installed into the microscope, providing a 488nm excitation band compatible with FITC. The CEM cells labeled only with secondary antibody were viewed in the microscope at 60X magnification, and the gain and black level were adjusted so that these cells were barely visible (Figure 5.20). The gain is 9.0, and the black level is 5.0. The image shown in the figure is taken very quickly in order to avoid any effect of bleaching, within a few seconds. This is faster than the images of the cells labeled using experimental antibodies were taken. The other cells were viewed without changing these settings.

A cell labeled with the anti-CD4 commercial antibody is shown in Figure 5.21. These cells are clearly brighter than the negative control. By refocusing on a slice through the middle of these cells, it is evident that the labeling is stronger along the cell membrane, indicating that the antigen labeled by this antibody is associated with the cell membrane. These cells are not perfectly

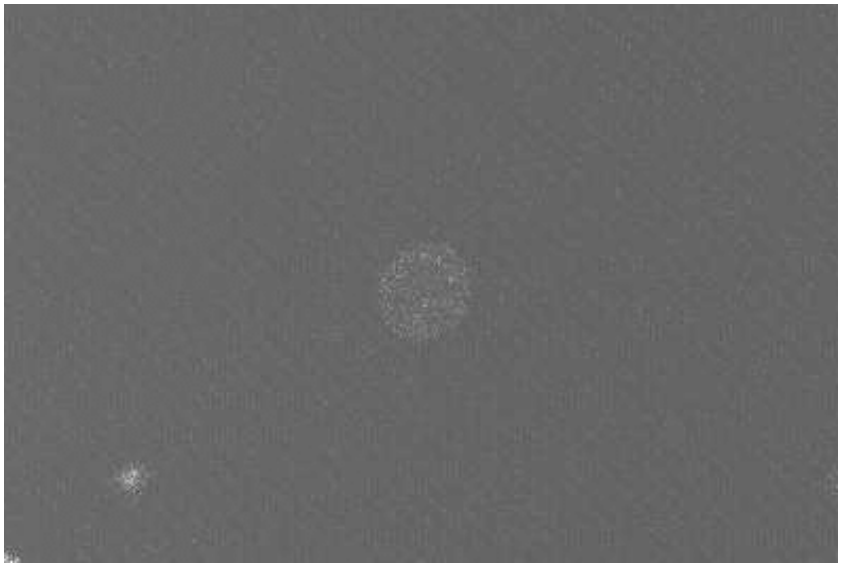


Figure 5.20. CEM Cells Labeled Using Only FITC Conjugated Secondary Antibody

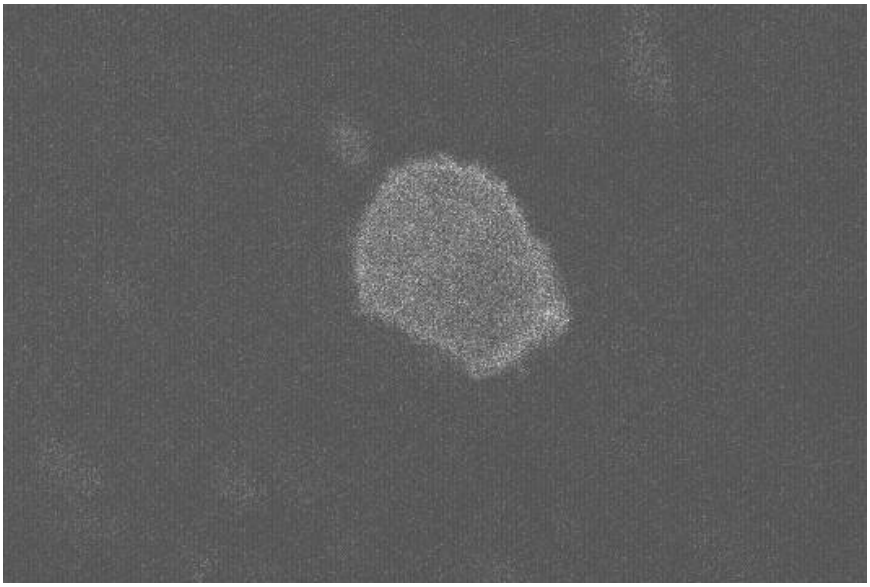


Figure 5.21. CEM Cells Labeled Using Both The Anti-CD4 Commercial Antibody and FITC Secondary Antibody

circular; they often have outgrowths as demonstrated in Figure 5.22. A slice across the base of these outgrowths is shown in Figure 5.23.

A cell labeled with the anti-024 antibody is shown in Figure 5.24. The labeling of this cell is also clearly brighter than the negative control. This view is along the middle plane of the cell, however membrane labeling is not as obvious as with the anti-CD4 commercial antibody. Another view, Figure 5.25, is along the top surface of the cell.

A cell labeled with the anti-025 antibody is shown in Figure 5.26. This view is along the top surface of the cell. Another view, Figure 5.27 is taken closer to the middle plane of the cell. Again, these cells are brighter than the negative control, however membrane labeling is not as obvious as those labeled with the anti-CD4 commercial antibody.

Conclusion

Cell labeling experiments finally yielded results when applied to the CEM cell line using the method outlined above. In fact, the CEM cell line was chosen because of relatively high level of CD4 expression reported in this cell line. The labeling of the CEM cells using each of the three experimental antibodies is clearly more intense than the labeling produced by using the secondary antibody alone. Further, membrane labeling is evident in the cells labeled using the anti-CD4 commercial antibody and the cells labeled using the anti-024 antibody. There is a hint of membrane labeling in the cells labeled using the anti-025 antibody. Cells labeled with each of the experimental antibodies have a fair amount of background labeling of cell cytoplasm. However, membrane labeling

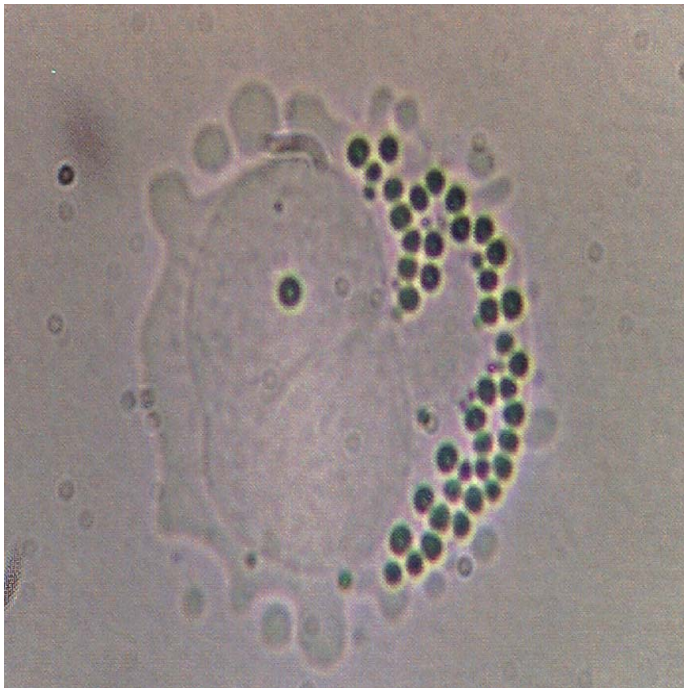


Figure 5.22. Unlabeled, Unfixed CEM Cell. 200x (17.5um)

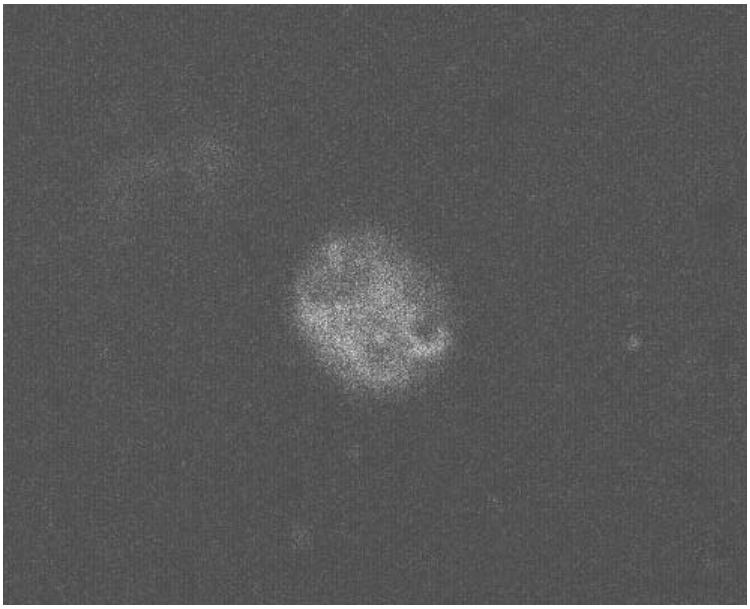


Figure 5.23. CEM Cell Labeled Using Anti-CD4 Commercial Antibody

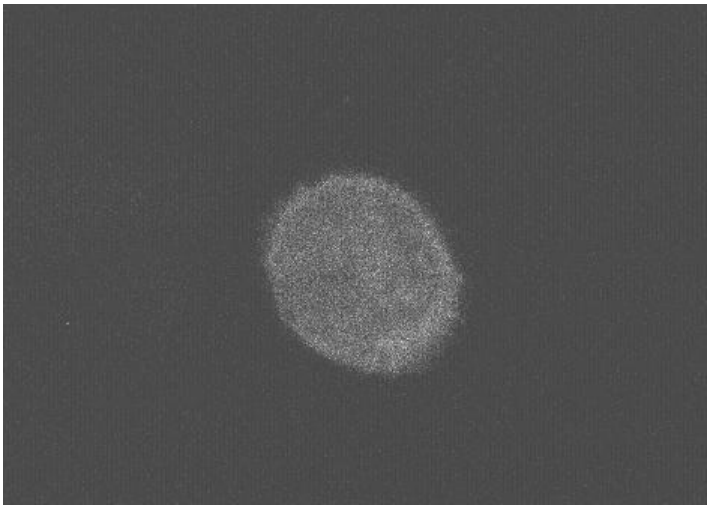


Figure 5.24. CEM Cell Labeled Using The Anti-024 Antibody

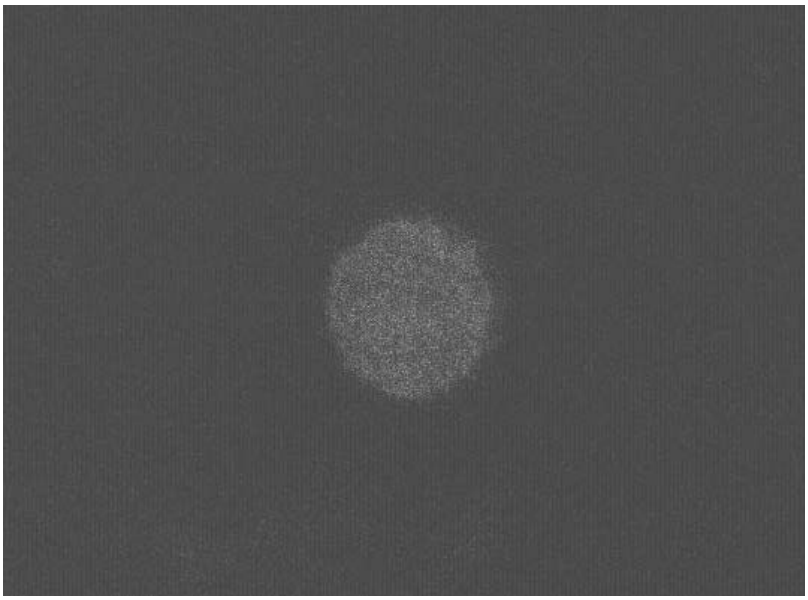


Figure 5.25. CEM Cell Labeled Using The Anti-024 Antibody

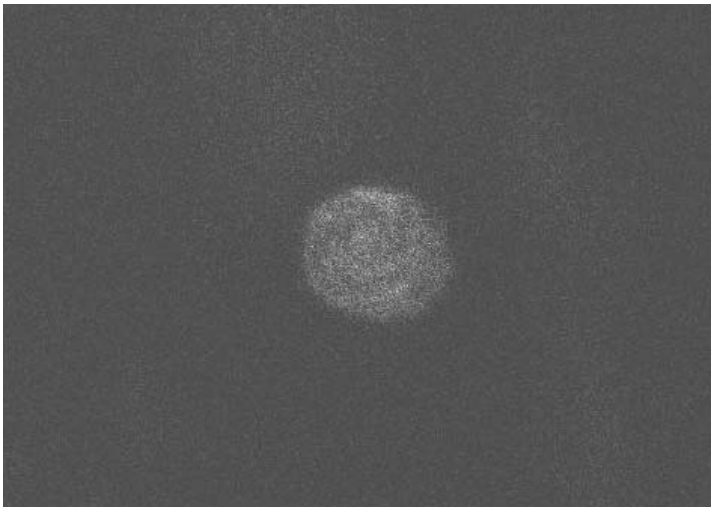


Figure 5.26. CEM Cell Labeled Using the Anti-025 Antibody

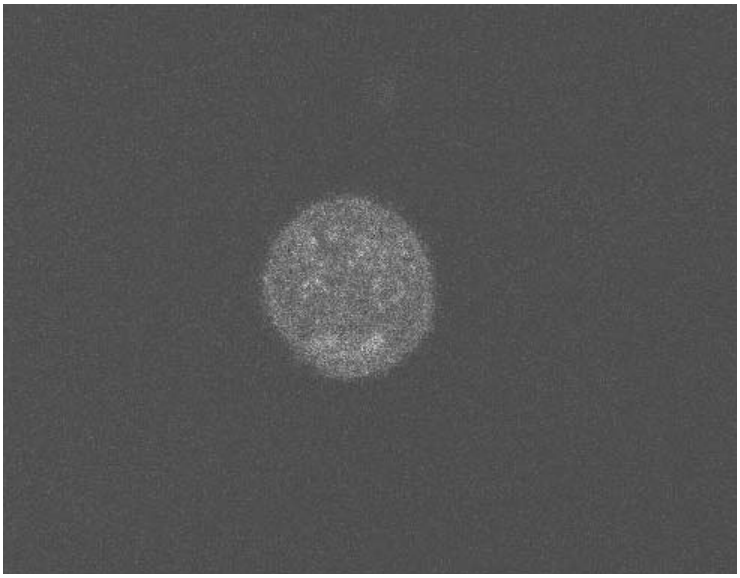


Figure 5.27. CEM Cell Labeled Using the Anti-025 Antibody

is indicated by an increase in intensity of labeling along the membrane. The slight membrane labeling in the cell labeled using the anti-025 antibody is so slight that it may be difficult to visualize in a printed version. This result is not surprising given the expected rare nature of the post-UGA transcript recognized by the anti-025 antibody. However, this result does not explain the ELISA results in which labeling by this antibody was most intense. However, the ELISA experiments were conducted using H9 and Jurkat cell extracts and this cell labeling experiment was conducted using CEM cells. The IgM nature of the anti-025 antibody did not seem to increase the labeling intensity of this antibody as in the ELISA experiments. The results of this experiment can be summed up as follows. Both positive controls, the anti-CD4 commercial antibody which recognizes an extra cellular portion of CD4, and the anti-024 antibody which was developed in Chapter 3 and recognizes the cytoplasmic tail of CD4, clearly labeled the membrane of the CEM cell line over and above the intensity of the CEM cells labeled using only the secondary antibody (the negative control). In addition, the experimental antibody, the anti-025 antibody that recognizes a region produced only if the 3'-UGA in CD4 is read through, also labeled the cells more intensely than the negative control, and the labeling is very slightly more intense along the cell membrane. Taken together, these results also support the hypothesis that translation of CD4 does not end at the first 3'-UGA.

CHAPTER 6

CURRENT AND FUTURE EXPERIMENTS

Introduction

The results of the antibody labeling experiments described in previous chapters are approaching the sensitivity limits of the techniques employed. Further refinement of those methods may produce some insights, particularly repeating the ELISA experiments to improve the statistics of those results, and performing flow cytometry analysis on the cells to verify those results. However, in order to further bolster the evidence obtained using antibody-labeling experiments, an entirely new approach is taken. Experiments have begun at the DNA level to try and show that the 3'-UGA in CD4 is read through in a systematic manner. A cDNA library is first prepared of the T-cells from which the fragment of the CD4 gene can be amplified using PCR.

To approach this problem in this way, a reporter gene is constructed having a multiple cloning site situated between a LacZ and a luciferase gene. A fragment of the CD4 gene containing the 3'-UGA codon is then inserted in between these genes. LacZ is located before the potential stop codon, and luciferase is located after the potential stop codon within the test sequence. An overview of the plasmid constructs are illustrated in Figure 6.1. In order to produce luciferase, any stop codon between the two genes must be circumvented. The LacZ gene is provided as an internal control in order to account for differences in transfection efficiency and copy number of the plasmid

Plasmid Construct

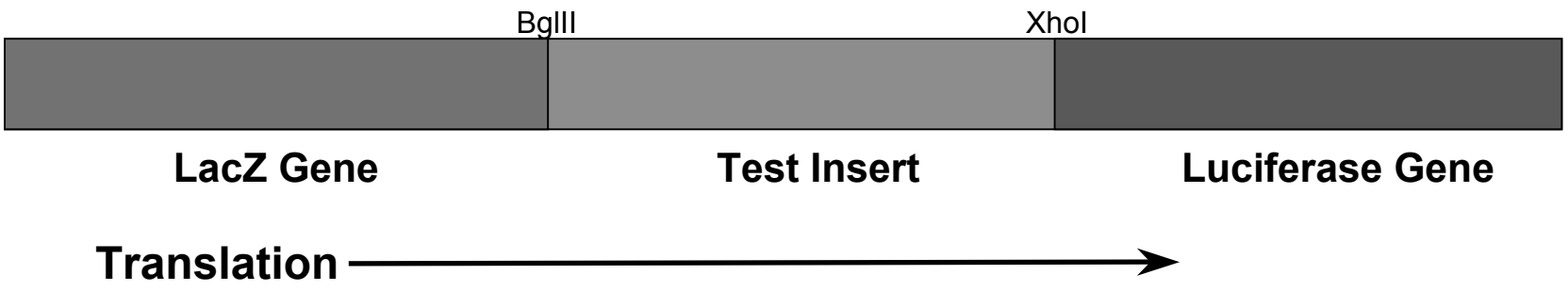


Figure 6.1. Plasmid Construction, Showing Reporter Gene and Test Sequence Insert Region

within the cells. The relative amounts of LacZ and luciferase will indicate read through efficiency.

Several controls are planned, only one of which has been started so far. One negative control fragment is constructed in which the UGA codon is mutated to UAA, which is a more efficient stop codon. Another negative control fragment is constructed in which a random sequence of DNA is constructed in which the only in frame stop codon is UGA. A positive control fragment is also constructed in which the UGA is mutated to UGC, which is cysteine, and will provide a baseline for 100% read through. The relative amounts of LacZ and luciferase are assayed using a fluorometer and the beta-galactosidase assay and compared to one another. The construct containing UAA is taken as 0% read through, the construct containing UGC as 100% read through. The relative amount of luciferase produced using the CD4 gene construct versus the random sequence construct will indicate whether or not the sequence surrounding the UGA in CD4 has an effect on it's read through efficiency. If so, this could provide positive evidence that the 3'-UGA in CD4 is read through by design and not randomly as sometimes occurs. In order to screen transfected cells for the plasmid, a kanamycin resistance gene is also contained in the plasmid.

The plasmid is first transfected into bacterial cells, which are screened for the correct plasmid, then harvested and transfected into Jurkat T-cells for the assay. This way, if there is something endogenous to the human Jurkat T-cells, which would improve the read through efficiency of this UGA codon, it is present in the assay.

Materials and Methods

Materials

Supplies for mRNA isolation and RT-PCR were obtained from Ambion. Materials for isolation of plasmid and PCR products from gels were obtained from Qiagen. The plasmid construct containing the multiple cloning site, the kanamycin resistance gene, the luciferase gene, and the green fluorescent protein gene is obtained from Lijun Zhao. He constructed this plasmid by adding the luciferase gene to the pEGFP-N3 plasmid available from Promega. Enzymes for PCR, restriction, ligation, and transfection are available from Promega and Sigma. Materials for gel electrophoresis were obtained from Sigma. Primers for PCR are obtained from Integrated DNA Technologies, in Iowa. JM-109 cells for transfection are obtained from Promega, and human T-cells including CD4+ Jurkat, CD4- Jurkat and H9, are obtained from the ATCC. Cell culture reagents are available from Sigma.

Methods

Cells were cultured as described in previous chapters. Methods not explicitly described here may be found by consulting the literature available with the kits used for each of the following protocols. mRNA is isolated from a mature culture of the T-cells using a kit obtained from Ambion containing a poly-T resin. 1-1.5ml of the T-cell cultures near 1X106 cells/ml are used for the mRNA isolation. The mRNA is converted to cDNA using a first strand synthesis kit from Ambion using both random decamers and poly-T as primers and reverse transcriptase as the polymerase. The conditions for reverse transcription are as

follows. The template and primers were preheated to 85C for 3 minutes in reaction buffer. RT and dNTP are then added and the reaction is incubated at 43C for 1 hour. To inactivate the RT, the sample is incubated at 92C for 10 minutes. The resulting first strand cDNA is further amplified by PCR using the same primers and Taq polymerase. The desired fragment of the CD4 gene is amplified from this library by PCR using primers developed as described below. The conditions for PCR not explicitly described in the kit are as follows. 1uM of each primer and 2ul of the RT reaction are used. Reagents available form Ambion including Taq and reaction buffer are used. As a positive control, the kit contains provisions for the RT-PCR of a mouse liver gene.

The desired fragment of the CD4 gene containing the UGA codon is chosen by analyzing the sequence and finding regions easily mutatable to contain restriction sites for BglII and XhoI. PCR primers are designed to contain these mutations.

The PCR products and plasmid are analyzed by gel electrophoresis in 1-2% agarose and visualization using ethidium bromide. The plasmid and PCR products are isolated from the gel using kits available from Qiagen.

Methods for restriction, ligation, and transfection were adapted from Promega's Protocols and Applications Guide.[91]

Results and Discussion

Jurkat CD4+, Jurkat CD4-, and H9 cells were grown for mRNA harvesting. The mRNA isolation and RT-PCR proceeded according to the instructions contained in the kits for this purpose from Ambion. The preparation

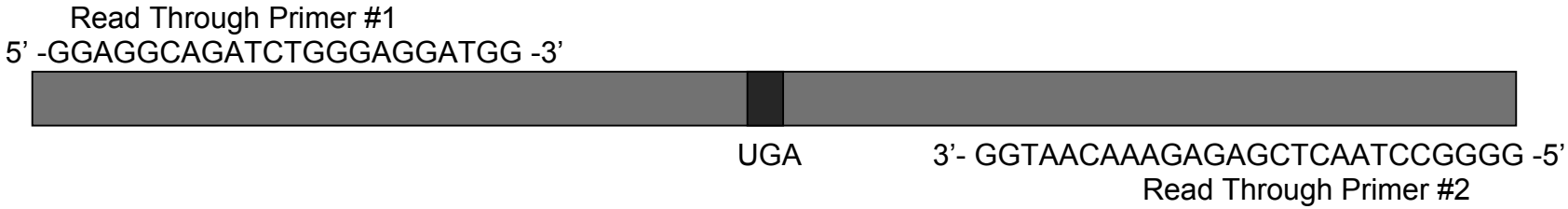
of the cDNA libraries was successful as verified by the ability to PCR this part of the CD4 gene and part of another T-cell gene, Lef-1, from them. The region of the CD4 gene amplified along with the mutation fragments prepared is shown in Figure 6.2. The PCR products using poly-dT in the first strand synthesis are shown in Figures 6.3 2-24-03 and 6.4 2-28-03. The first mutation fragment is visible in the gel in Figure 6.4, but nothing is visible in the second mutation lane. Because production of the second mutation fragment failed using the poly-dT first strand sample, the process is repeated using the random decamers first strand synthesis. The mutation fragments obtained using the random decamers first strand synthesis is shown in Figure 6.5 3-18-03. Because the second mutation fragment has not yet been produced, the mutation process cannot be completed.

The bands containing the fragment of the CD4 gene as well as the first mutation fragment were cut out of the gel, and the DNA isolated using kits from Qiagen made for this purpose. The restriction reaction of the fragment of CD4 containing the UGA and the plasmid is assembled as recommended by Promega, and incubates overnight at 37C. Gel electrophoresis of this reaction is shown in Figure 6.6 4-30-03. The plasmid is visible, but the fragment containing UGA is not. Apparently the purification process before the restriction failed.

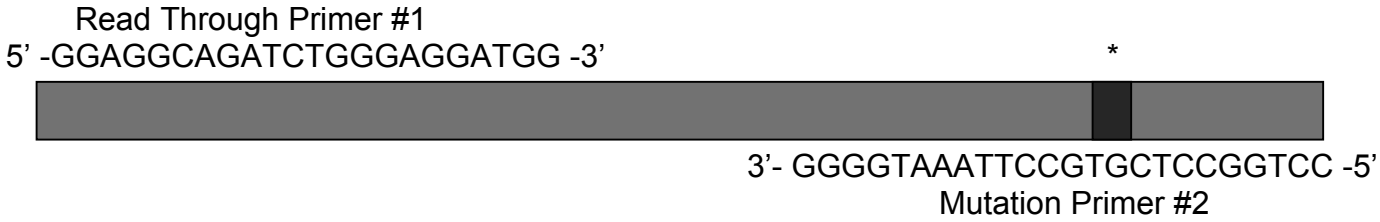
Conclusion

The cDNA library production from the Jurkat CD4+, the Jurkat CD4-, and the H9 cell lines was successful. These libraries are stored at –80C for later use.

PCR of 3' End of CD4 Including UGA, 366bp



PCR of First Mutation Fragment, 209bp



PCR of Second Mutation Fragment, 163bp

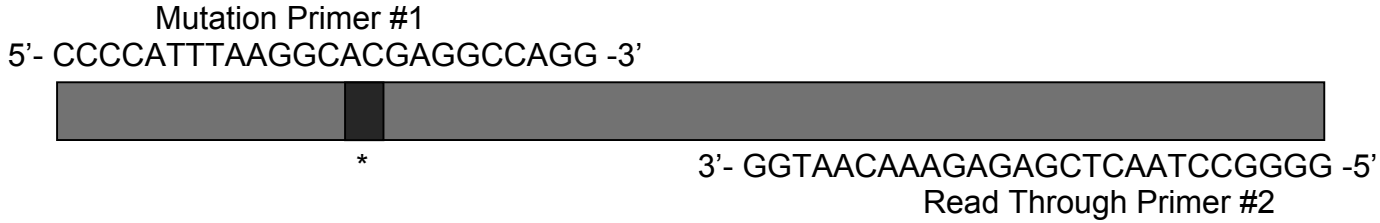


Figure 6.2. PCR Amplifications. The mutation fragments are complementary to one another and a subsequent PCR using these as primers of each other produces a new, mutated 366bp end of the CD4 gene. The asterisk denotes the location of the mutated UGA codon.

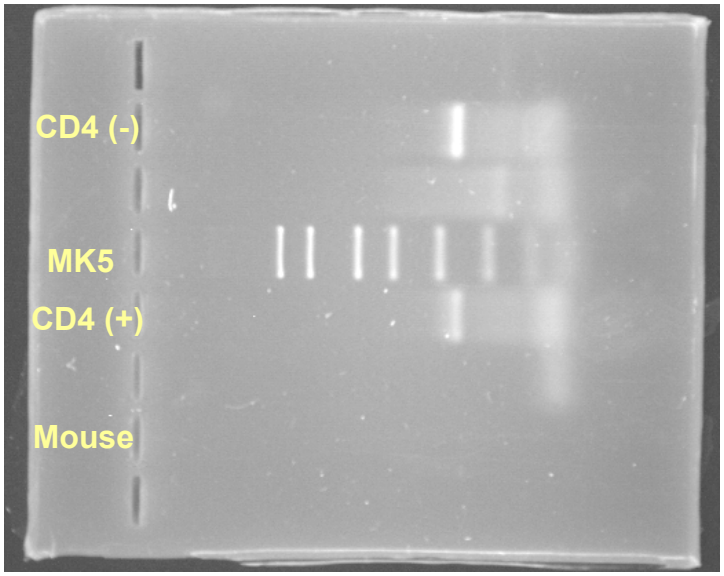


Figure 6.3. PCR products using poly-dT in the first strand synthesis. The visible markers are, from left to right, 2000bp, 1500bp, 1000bp, 750bp, 500bp, 300bp, 150bp, and 50bp. The unlabeled lanes are PCR negative controls.

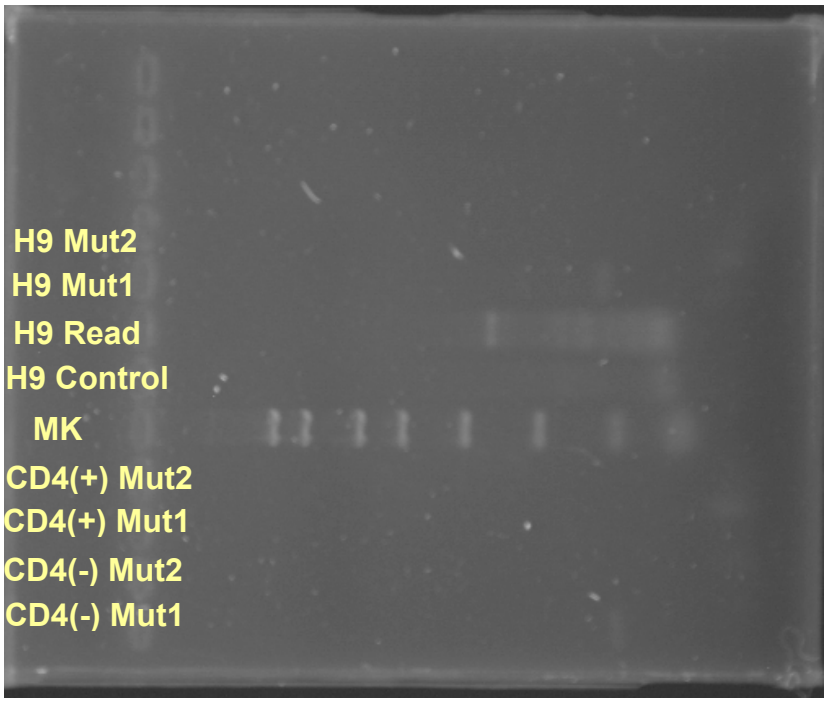


Figure 6.4. PCR products using poly-dT in the first strand synthesis. The same markers were used in this gel as in Figure 6.3.

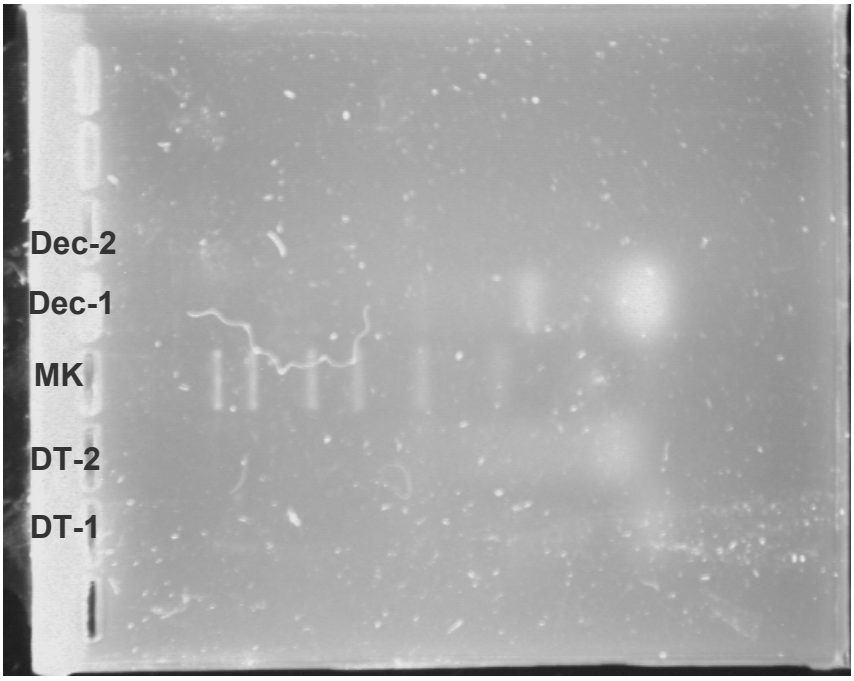


Figure 6.5. Mutation fragments obtained using the random decamers. The markers are the same as in Figure 6.3.

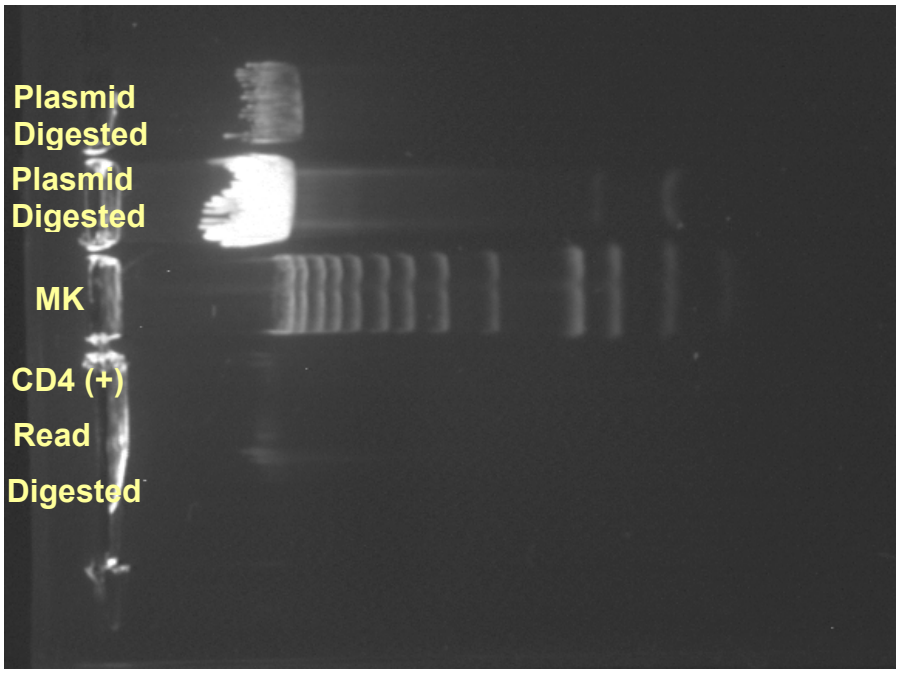


Figure 6.6. Gel electrophoresis of restriction reaction. The markers, from left to right are, in Kb: 10, 8, 6, 5, 4, 3, 2.5, 2, 1.5, 1, 0.8, 0.6, 0.4, and 0.2.

Amplification of the fragment of CD4 containing the UGA codon from the CD4- Jurkat cell line was also successful. This further bolsters the assertion first discussed in the conclusion of the ELISA section that this cell line may not be CD4- after all. These results are not conclusive of this, however. These results only indicate that CD4 mRNA is being produced within these cells. There may still be a mutation within the mRNA sequence that prevents functional CD4 from making it to the surface of the cell. The fact that some antibodies against the outer loops of CD4 may not recognize CD4 on this cell line indicates that there must be a mutation somewhere within the sequence, even if it does not prevent expression of a CD4 receptor. This portion of research is not yet complete. Construction of the control fragments continues as does the process of putting them in plasmids and transfection. The ultimate results of this process will provide useful insights into whether or not the 3'UGA in CD4 is read through. One potential problem is that of the relatively small size of the fragment of CD4 inserted into the reporter gene construct. There may be other structures within the DNA sequence, such as an unidentified SECIS that may be necessary for the read through of this UGA that are not contained within the final construct. If the results of this read through experiment are negative, then the experiment should be repeated using other fragments of CD4 containing the UGA, particularly longer ones containing more of the 3' end of the mRNA.

CHAPTER 7

CONCLUSIONS

Much of the data reported here support the hypothesis tested. Although many of the results are not indisputable, taken together, they present a strong case for continued experimentation on genes such as CD4. In fact, from the outset, it was expected that the read through product would be rare and difficult to find even if expressed. The key results are reviewed here. The sequence analysis produced three different results, all supportive of the hypothesis. First, several other potential selenoproteins were identified throughout the human genome that may have been overlooked in the same way as the CD4 gene. Second, when the region past the UGA codon is translated in the zero and +1 frames (supported by theoretical evidence), the resulting protein sequences are significantly homologous to known proteins. When translated in the -1 frame (not supported by theoretical evidence), equally significant sequence similarities are not identified. Third, a more detailed look at the potential sequence similarity between the zero and +1 frames of the end of the CD4 gene revealed striking similarities. This sequence similarity covers three active site regions including a tryptophan residue, least expected by random chance. The overall sequence similarity score is 4.6 standard deviations above that expected by chance. The antibody labeling studies also produced promising results. Monoclonal antibodies were successfully produced that recognize synthetic peptides that mimic both the recognized cytoplasmic tail of CD4, and the potentially transcribed region past the UGA codon. Unfortunately, the sensitivity of the

Western blotting technique is not able to resolve the low amount of CD4 contained in the T-cell extracts. ELISA experiments using T-cell extracts produced results strongly supporting the possibility that translation continues past the 3'-UGA in CD4. Both the positive control anti-024 antibody which was developed to interact with the recognized cytoplasmic tail of CD4 and the anti-025 antibody which was developed to interact with the potentially translated region past the UGA codon interacted with the T-cell extracts. The whole cell labeling experiments were less conclusive, but membrane labeling is evident when the labeling technique is applied to CEM cells using the anti-CD4 commercial antibody and the anti-024 positive control antibody. In addition, membrane labeling just visible above the background labeling of the cytoplasm may be present in the cells labeled using the anti-024 antibody. Finally, a cDNA library of three of the T-cell lines was successfully developed, and experiments at the DNA level on the 3'-UGA codon in CD4 are ongoing. The ultimate results of these experiments will help resolve any uncertainty about whether or not translation continues past the 3'-UGA in the CD4 gene.

REFERENCES

1. Yanofsky, C., *Gene Structure and Protein Structure.* Sci. Am., 1967. **216**(5): p. 80-94.
2. Davies, M.V. and R.J. Kaufman, *The Sequence Context of the Initiation Codon in the Encephalomyocarditis Virus Leader Modulates Efficiency of Internal Translation Initiation.* Journal of Virology, 1992. **66**(4): p. 1924-1932.
3. Kozak, M., *Initiation of translation in prokaryotes and eukaryotes.* Gene, 1999. **234**(2): p. 187-208.
4. Berry, M.J., et al., *Functional-Characterization of the Eukaryotic Secis Elements Which Direct Selenocysteine Insertion at Uga Codons.* Embo Journal, 1993. **12**(8): p. 3315-3322.
5. Venter, J.C., Adams, M.D., Myers, E.W., et.al., *The Sequence of the Human Genom.* Science, 2001. **291**: p. 1304-1351.
6. Arner, E.S.J., et al., *High-level expression in Escherichia coli of selenocysteine-containing rat thioredoxin reductase utilizing gene fusions with engineered bacterial-type SECIS elements and co-expression with the selA, selB and selC genes.* Journal of Molecular Biology, 1999. **292**(5): p. 1003-1016.
7. Copeland, P.R., et al., *A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs.* Embo Journal, 2000. **19**(2): p. 306-314.

8. Martin, G.W. and M.J. Berry, *Eukaryotic selenocysteine incorporation: Mechanistic insights.* Phosphorus Sulfur and Silicon and the Related Elements, 1998. **136**: p. 309-320.

9. Tujebajeva, R.M., et al., *Decoding apparatus for eukaryotic selenocysteine insertion.* Embo Reports, 2000. **1**(2): p. 158-163.

10. Arkov, A.L. and E.J. Murgola, *Ribosomal RNAs in translation termination: Facts and hypotheses.* Biochemistry-Moscow, 1999. **64**(12): p. 1354-1359.

11. Mottagui-Tabar, S., *Quantitative analysis of in vivo ribosomal events at UGA and UAG stop codons.* Nucleic Acids Research, 1998. **26**(11): p. 2789-2796.

12. Chan, T.S. and A. Garen, *Amino Acid Substitutions Resulting from Suppression of Nonsense Mutations .5. Tryptophan Insertion by Su9+ Gene, a Suppressor of Uga Nonsense Triplet.* Journal of Molecular Biology, 1970. **49**(1): p. 231-&.

13. Kuchino, Y. and T. Muramatsu, *Nonsense suppression in mammalian cells.* Biochimie, 1996. **78**(11-12): p. 1007-1015.

14. Smiley, B.K. and F.C. Minion, *Enhanced Readthrough of Opal (Uga) Stop Codons and Production of Mycoplasma-Pneumoniae P1 Epitopes in Escherichia-Coli.* Gene, 1993. **134**(1): p. 33-40.

15. Goringer, H.U., et al., *Mutations in 16s Ribosomal-Rna That Affect Uga (Stop Codon)-Directed Translation Termination.* Proceedings of the

National Academy of Sciences of the United States of America, 1991. **88**(15): p. 6603-6607.

16. Jemiolo, D.K., F.T. Pagel, and E.J. Murgola, *UGA suppression by a mutant RNA of the large ribosomal subunit.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(26): p. 12309-12313.
17. Prescott, C.D., B. Kleuvers, and H.U. Goringer, *A Ribosomal-Rna-Messenger Rna Base-Pairing Model for Uga-Dependent Termination.* Biochimie, 1991. **73**(7-8): p. 1121-1129.
18. Konan, K.V. and C. Yanofsky, *Role of ribosome release in regulation of tna operon expression in Escherichia coli.* Journal of Bacteriology, 1999. **181**(5): p. 1530-1536.
19. Muramatsu, T. and Y. Kuchino, *Translational control through nonsense suppression.* Seikagaku, 1997. **69**(1): p. 36-39.
20. Poole, E.S., C.M. Brown, and W.P. Tate, *The Identity of the Base Following the Stop Codon Determines the Efficiency of in-Vivo Translational Termination in Escherichia-Coli.* Embo Journal, 1995. **14**(1): p. 151-158.
21. Mottaguitabar, S., A. Bjornsson, and L.A. Isaksson, *The 2nd to Last Amino-Acid in the Nascent Peptide as a Codon Context Determinant.* Embo Journal, 1994. **13**(1): p. 249-257.

22. Zhang, S.P., M. RydenAulin, and L.A. Isaksson, *Functional interaction between release factor one and P-site peptidyl-tRNA on the ribosome.* Journal of Molecular Biology, 1996. **261**(2): p. 98-107.

23. Tate, W.P. and S.A. Mannering, *Three, four or more: The translational stop signal at length.* Molecular Microbiology, 1996. **21**(2): p. 213-219.

24. MottaguiTabar, S. and L.A. Isaksson, *Influence of the last amino acid in the nascent peptide on EF-Tu during decoding.* Biochimie, 1996. **78**(11-12): p. 953-958.

25. Bonetti, B., et al., *The Efficiency of Translation Termination Is Determined by a Synergistic Interplay between Upstream and Downstream Sequences in Saccharomyces-Cerevisiae.* Journal of Molecular Biology, 1995. **251**(3): p. 334-345.

26. Harrell, L., U. Melcher, and J.F. Atkins, *Predominance of six different hexanucleotide recoding signals 3 ' of read-through stop codons.* Nucleic Acids Research, 2002. **30**(9): p. 2011-2017.

27. Levander, O.A., *A Global View of Human Selenium Nutrition.* Annual Review of Nutrition, 1987. **7**: p. 227-250.

28. Fuchs, O., *Selenoproteins.* Chemicke Listy, 1996. **90**(7): p. 444-450.

29. Martin, G.W., J.W. Harney, and M.J. Berry, *Selenocysteine incorporation in eukaryotes: Insights into mechanism and efficiency from sequence, structure, and spacing proximity studies of the type 1 deiodinase SECIS element.* Rna-a Publication of the Rna Society, 1996. **2**(2): p. 171-182.

30. Kollmus, H., L. Flohe, and J.E.G. McCarthy, *Analysis of eukaryotic mRNA structures directing cotranslational incorporation of selenocysteine.* Nucleic Acids Research, 1996. **24**(7): p. 1195-1201.

31. Rother, M., et al., *Selenoprotein synthesis in archaea.* Biofactors, 2001. **14**(1-4): p. 75-83.

32. Arvilommi, H., et al., *Selenium and Immune Functions in Humans.* Infection and Immunity, 1983. **41**(1): p. 185-189.

33. Kiremidjianschumacher, L. and G. Stotzky, *Selenium and Immune-Responses.* Environmental Research, 1987. **42**(2): p. 277-303.

34. Odell, J.R., et al., *Serum Selenium Concentrations in Rheumatoid-Arthritis.* Annals of the Rheumatic Diseases, 1991. **50**(6): p. 376-378.

35. Kavanaughmchugh, A.L., et al., *Selenium Deficiency and Cardiomyopathy in Acquired-Immunodeficiency-Syndrome.* Journal of Parenteral and Enteral Nutrition, 1991. **15**(3): p. 347-349.

36. Delilbasi, E., et al., *Selenium and Behcets-Disease.* Biological Trace Element Research, 1991. **28**(1): p. 21-25.

37. Mei, W.D., et al., *Study of Immune Function of Cancer-Patients Influenced by Supplemental Zinc or Selenium-Zinc Combination.* Biological Trace Element Research, 1991. **28**(1): p. 11-20.

38. Ip, C. and G. White, *Mammary-Cancer Chemoprevention by Inorganic and Organic Selenium - Single Agent Treatment or in Combination with Vitamin-E and Their Effects on Invitro Immune Functions.* Carcinogenesis, 1987. **8**(12): p. 1763-1766.

39. Langman, M. and P. Boyle, *Chemoprevention of colorectal cancer.* Gut, 1998. **43**(4): p. 578-585.

40. Manteroatienza, E., et al., *Selenium Status and Immune Function in Asymptomatic Hiv-1 Seropositive Men.* Nutrition Research, 1991. **11**(11): p. 1237-1250.

41. Dworkin, B.M., *Selenium Deficiency in Hiv-Infection and the Acquired-Immunodeficiency-Syndrome (Aids).* Chemico-Biological Interactions, 1994. **91**(2-3): p. 181-186.

42. Schrauzer, G.N. and J. Sacher, *Selenium in the Maintenance and Therapy of Hiv-Infected Patients.* Chemico-Biological Interactions, 1994. **91**(2-3): p. 199-205.

43. Beck, M.A., *The role of nutrition in viral disease.* Journal of Nutritional Biochemistry, 1996. **7**(12): p. 683-690.

44. Pace, G.W. and C.D. Leaf, *The Role of Oxidative Stress in Hiv Disease.* Free Radical Biology and Medicine, 1995. **19**(4): p. 523-528.

45. Baum, M.K., et al., *High risk of HIV-related mortality is associated with selenium deficiency.* Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 1997. **15**(5): p. 370-374.

46. Look, M.P., et al., *Serum selenium, plasma glutathione (GSH) and erythrocyte glutathione peroxidase (GSH-Px)-levels in asymptomatic versus symptomatic human immunodeficiency virus-1 (HIV-1)-infection.* European Journal of Clinical Nutrition, 1997. **51**(4): p. 266-272.

47. Zhang, W., et al., *Selenium-dependent glutathione peroxidase modules encoded by RNA viruses.* Biological Trace Element Research, 1999. **70**(2): p. 97-116.

48. Grate, L., *Potential SECIS elements in HIV-1 strain HXB2.* Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 1998. **17**(5): p. 398-403.

49. Taylor, E.W., et al., *A Basis for New Approaches to the Chemotherapy of Aids - Novel Genes in Hiv-1 Potentially Encode Selenoproteins Expressed by Ribosomal Frameshifting and Termination Suppression.* Journal of Medicinal Chemistry, 1994. **37**(17): p. 2637-2654.

50. Low, S.C. and M.J. Berry, *Knowing when not to stop: Selenocysteine incorporation in eukaryotes.* Trends in Biochemical Sciences, 1996. **21**(6): p. 203-208.

51. Grundner-Culemann, E., et al., *Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes.* Rna-a Publication of the Rna Society, 1999. **5**(5): p. 625-635.

52. Mizutani, T., K. Tanabe, and K. Yamada, *G-U base pair in the eukaryotic selenocysteine tRNA is important for interaction with SePF, the putative selenocysteine-specific elongation factor.* Febs Letters, 1998. **429**(2): p. 189-193.

53. Fourmy, D., E. Guittet, and S. Yoshizawa, *Structure of prokaryotic SECIS mRNA hairpin and its interaction with elongation factor SelB.* Journal of Molecular Biology, 2002. **324**(1): p. 137-150.

54. Walczak, R., P. Carbon, and A. Krol, *An essential non-Watson-Crick base pair motif in 3 ' UTR to mediate selenoprotein translation.* Rna-a Publication of the Rna Society, 1998. **4**(1): p. 74-84.

55. Martin, G.W., J.W. Harney, and M.J. Berry, *Functionality of mutations at conserved nucleotides in eukaryotic SECIS elements is determined by the identity of a single nonconserved nucleotide.* Rna-a Publication of the Rna Society, 1998. **4**(1): p. 65-73.

56. Sandman, K.E., et al., *Revised Escherichia coli selenocysteine insertion requirements determined by in vivo screening of combinatorial libraries of SECIS variants.* Nucleic Acids Research, 2003. **31**(8): p. 2234-2241.

57. Forchhammer, K., K.P. Rucknagel, and A. Bock, *Purification and Biochemical-Characterization of Selb, a Translation Factor Involved in Selenoprotein Synthesis.* Journal of Biological Chemistry, 1990. **265**(16): p. 9346-9350.

58. Chittum, H.S., et al., *Rabbit beta-globin is extended beyond its UGA stop codon by multiple suppressions and translational reading gaps.* Biochemistry, 1998. **37**(31): p. 10866-10870.

59. Elgert, K.D., *Immunology: Understanding the Immune System*. Prepublication Manuscript. 1994, Blacksburg, VA: Virginia Polytechnic Institute and State University. 622.

60. Ansari-Lari, M.A., Muzny, D.M., Lu, J., Lu, F., et.al., *A gene-rich cluster between the CD4 and triosehosphate isomerase genes at human chromosome 12p13.* Genome Research, 1996. **6**(4): p. 314-326.

61. Holdorf, A.D., et al., *Regulation of Lck activity by CD4 and CD28 in the immunological synapse.* Nature Immunology, 2002. **3**(3): p. 259-264.

62. Foti, M., et al., *p56(Lck) anchors CD4 to distinct microdomains on microvilli.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(4): p. 2008-2013.

63. Taylor, E.W., *Selenium and Cellular-Immunity - Evidence That Selenoproteins May Be Encoded in the +1-Reading-Frame Overlapping the Human Cd4, Cd8, and Hla-Dr Genes.* Biological Trace Element Research, 1995. **49**(2-3): p. 85-95.

64. Facchiano, A., F. Facchiano, and J. Vanrenswoude, *Divergent Evolution May Link Human-Immunodeficiency-Virus Gp41 to Human Cd4.* Journal of Molecular Evolution, 1993. **36**(5): p. 448-457.

65. Facchiano, A., *Investigating Hypothetical Products from Noncoding Frames (Hypnofs).* Journal of Molecular Evolution, 1995. **40**(6): p. 570-577.

66. Hershko, A., Ciechanover, A.,, *The ubiquitin system for protein degradation.* Annu. Rev. Biochem, 1992. **61**: p. 761-807.

67. Voet, D., Voet, Judith, *Biochemistry*. 1995, New York: John Wiley & Sons.

68. Lescure, A., et al., *Novel selenoproteins identified in silico and in vivo by using a conserved RNA structural motif.* Journal of Biological Chemistry, 1999. **274**(53): p. 38147-38154.

69. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., et.al., *Gapped BLAST and PSI-BLAST: an new generation of protein database search programs.* Nucleic Acids Research, 1997. **25**: p. 3389-3402.

70. Fagegaltier, D., et al., *Structural analysis of new local features in SECIS RNA hairpins.* Nucleic Acids Research, 2000. **28**(14): p. 2679-2689.

71. Strausberg, R.L., Feingold, E.A., Grouse, L.H., et.al., *Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.* Proceedings of the National Academy of Sciences USA, 2002. **99**(26): p. 16899-16903.

72. Sulston, J.E., and Waterston, R., *Toward a complete human genome sequence.* Genome Research, 1998. **8**(11): p. 1097-1108.

73. Bulfone, A., Menguzzato, E., Broccoli, V., et.al., *Barhl 1, a gene belonging to a new subfamily of mammalian homeobox genes, is expressed in migrating neurons of the CNS.* Hum. Mol. Genet., 2000. **9**(9): p. 1443-1452.

74. White, O., Eisen, J.A., Heidelberg, J.F., et.al., *Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.* Science, 1999. **286**(5444): p. 1571-1577.

75. Garnier, J., Gibrat, J.F., Robson, B., *GOR secondary structure prediction method version IV.* Methods in Enzymology, 1996. **266**: p. 540-553.

76. Rost, B., *PredictProtein.* Methods in Enzymology, 1996. **266**: p. 525-539.

77. Hofmann, K., Bucher, P., Falquet, L., Bairoch, A., *The PROSITE database, its status in 1999.* Nucleic Acids Research, 1999. **27**: p. 215-219.

78. Corpet, F., Servant, F., Gouzy, J., Kahn, D., *ProDom and ProDom-CG: tools for protein domain analysis and qhole genome comparisons.* Nucleic Acids Research, 2000. **28**: p. 267-269.

79. Rost, B., *PHD: predicting one-dimensional protein structure by profile based neural networks.* Methods in Enzymology, 1996. **266**: p. 525-539.

80. Rost, B., *Short yeast ORFs: Expressed protein or not?* Unpublished, 2000.

81. Davis, R., *Personal Communication.* Unpublished, 2001.

82. Pharmacia, *Hitrap SP Column Literature*. 2000.

83. Pharmacia, *G25 Column Literature*. 2000.

84. Robyt, J.F., White, Bernard J., *Biochemical Techniques Theory and Practice*. 1990, Prospect Heights, Illinois: Waveland Press, Inc. 407.

85. Harlow, E., Lane, D., *Antibodies-a Laboratory Manual*. 1988: Cold Spring Harbor Laboratory.

86. Davis, B., Keuhl, *Basic Methods in Molecular Biology*. 1993.

87. Lowry, O.H., Rosebrough, N.J., Farr, A.L., Randall, R.J., *Protein measurement with the Folin phenol reagent.* J.Biol. Chem., 1951. **193**: p. 265.

88. Davis, B., Keuhl, *Basic Methods in Molecular Biology*. Second ed. 1994, Norwalk, CT: Appelton & Lange.

89. Corporation, P., *ProtoBlot II AP System Technical Manual*. 2001, Promega.

90. Lederman, S., Chess, L., Yellin, M.J., *Murine monoclonal antibody 5c8 recognizes a human glycoprotein on the surface of T-lymphocytes, compositions containing same.*, in *USPTO*. 1991, The Trustees of Columbia University in the City of New York: USA.

91. Doyle, K., ed. *Protocols and Applications Guide*. Third ed. 1996, Promega Corporation: USA. 404.

APPENDIX

SELENIUM AND VIRAL DISEASES: WHICH COMES FIRST?

Introduction and Scope

The purpose of this article is to review the research examining the link between the micronutrient selenium and viral disease etiology. There is much data on the biological roles of selenium in many different forms of life. The challenge now facing scientists interested in the biological effects of selenium lies in integrating experimental data and clinical observations on a cellular and molecular level. A large volume of veterinary data on the effects of selenium on animal health exists, much of it related to immune system function and viral disease etiology, and may provide further insight into the possible mechanisms of how selenium levels affect mammalian health.[92-94] However, because of the large volume of selenium research, the scope of this review is limited to humans and the viruses that infect them, and is largely limited to recent work. A link between selenium nutritional status and many disease states is well established.[95] In addition, clinical and experimental results linking selenium deficiency to poor prognosis are accumulating at a rapid rate. Although there is much data concerning the health benefits of dietary selenium, the biological basis for the diverse effects of selenium have not yet been fully explained. This review begins with a brief overview of the importance of selenium to mammalian physiology. Next, the biological basis for the effects of selenium, the twenty-first amino acid, selenocysteine, is presented and the complex process by which selenocysteine adds to growing polypeptide chains within the cell will be explored. Clinical data highlighting the connection between body levels of

selenium and viral disease progression is presented next. Some experimentally verified effects of selenium deficiency on both the host and the virus are outlined as part of the explanation for the profound link between selenium and disease progression. The possible explanations for the dramatic reduction of serum selenium levels seen during certain viral infections are outlined next. Finally, data concerning the possible existence of presently unknown or unverified selenoproteins in the host and viral genomes is presented. These potential selenoproteins could account for much of the link between selenium and disease progression.

Selenium in Human Physiology

As reviewed by Rayman and Holben, selenium plays a vital role in human physiology.[96, 97] Selenium deficiency in an adult can lead to defective sperm development in the male, cardiac muscle disease, and Kashin-Beck disease, a severe deforming form of arthritis. Selenium deficiency during early development can lead to unique problems such as incomplete development of the immune system, nervous system, and congenital cardiac defects. In addition, selenium deficiency is a cofactor in the development of Keshan disease, a cardiomyopathy.[98-100] Keshan disease and Kashin-Beck disease are endemic to certain regions of China where the soil is deficient in selenium.[101] The diet of the local population is therefore also selenium deficient because crops such as rice grown on the local soil does not contain an adequate amount of selenium.[102] As reviewed by Rayman, less severe selenium deficiency can also have negative physiological effects.[96] Low dietary intake or body level of

selenium is linked to higher incidence or poor outcome of many diseases. These topics will be discussed in more detail later in this review.

Selenium is a unique nutritional requirement. Many micronutrients are required as enzyme cofactors. However, selenium is incorporated into proteins as the twenty-first amino acid, selenocysteine.[103] The incorporation of selenocysteine is a co-translational process, and is thus an integral part of protein synthesis, unlike the binding of cofactors to apoenzymes which occurs after protein synthesis is complete.[104]

The mechanism by which selenocysteine is incorporated into a growing polypeptide chain is thus unique. Despite significant similarities, the mechanism is somewhat different in bacteria and humans, and the focus of this review is limited to mammalian systems and viruses that infect them. The UGA codon, which normally encodes a stop signal, is also used to encode selenocysteine. The process of translating a UGA codon as selenocysteine is not a random process, as simple read through suppression may be (e.g. read through associated with mutant “suppressor” tRNAs of conventional amino acids).[26, 58] Additional specific elements appear to be necessary for translation of UGA as selenocysteine to occur.[30, 54, 104-117] An RNA stem-loop structure in the 3’ untranslated region (3’ UTR) is necessary for selenocysteine insertion to occur. This structure is referred to as a selenocysteine insertion sequence (SECIS) element. In eukaryotic organisms, the SECIS structure is typically found hundreds of bases away from the UGA codon. Some selenoproteins, such as selenoprotein P contain several selenocysteine residues, and experimental

evidence shows that in eukaryotes a SECIS element in the 3'UTR can serve to direct selenocysteine insertion at more than one UGA codon.[4, 118] Much research on the SECIS element structure is ongoing. One of the major goals of this research is to elucidate the minimal structure necessary to direct the incorporation of selenocysteine. The hope is that a consensus sequence can be identified that can be used to find additional SECIS elements in the genome (an approach that has been used with some success by Gladyshev's group.[119]). Although several distinct classes of SECIS element have been identified [4], one recent study reported that the only totally conserved *specific* sequence element of the SECIS is four non-Watson Crick base pairs that occur near the base of the SECIS stem, but this must be in a specific secondary structural context.[119] In addition, recent studies also reveal that there are at least two proteins that interact with the eukaryotic SECIS element and render it functional.[7, 108, 117, 120-122] This fact helps explain why in vitro translation systems utilizing prokaryotes often fail to translate eukaryotic UGA codons as selenocysteine, even when a functional SECIS element is inserted into the 3' UTR.[6, 123] Work on the SECIS element structure and mechanism is ongoing, and there is no reason to expect that all SECIS element structures have been identified. In addition, different proteins may be required to interact with the different classes of SECIS element. Thus, the complexity involved in translation of UGA as selenocysteine is considerable.

In mammals, selenium is a key component of glutathione peroxidases (GPx), thioredoxin reductases (TRx), deiodinases, selenoprotein P, selenoprotein

W, selenophosphate synthase, and other yet to be characterized proteins (section 8). Most appear to play a role in controlling redox tone. Selenium is typically found in the active site in the form of selenocysteine, and is involved in catalyzing the reaction. Substitution of cysteine in the place of selenocysteine in these proteins leads to dramatic reduction of enzymatic activity.[124] In addition, mutation of cysteine to selenocysteine may enhance the enzymatic activity of certain enzymes involved in redox reactions.[125] Selenium is clearly a critical component of these enzymes. The critical role of selenium over sulfur in these enzymes may be based in part on the fact that selenocysteine has a lower pKa than cysteine, and thus is predominantly ionized at physiological pH. In addition, selenium is more nucleophilic than sulfur.

Selenoproteins are therefore essential for controlling oxidative stress within the cell, which is commonly associated with many viral infections.[126] The physiological role of reactive oxygen species has been reviewed recently by Norberg and Arner.[127] Reactive oxygen species (ROS) are produced in vivo during many biochemical processes within the cell. These include mitochondrial electron transport, as second messengers within signaling pathways, apoptosis, and cytotoxic immune system responses. Production of ROS in excess of the ability of antioxidant mechanisms to scavenge them leads to oxidative stress. Oxidative stress has been shown to damage lipids [128], proteins [129], and nucleic acids.[130] The buildup of this damage over time may lead to various problems associated with aging including increased cancer risk and impaired immune function.[131]

Selenium plays many diverse roles in the immune system and is necessary for proper immune system function. Selenium deficiency has been shown to impair both the humoral and cellular immune response in animal studies.[132, 133] (For an excellent review on the physiological effects of selenium, see Combs, Gray, 1998. [134]) Selenium is particularly important for the cellular immune system. Selenium supplementation of cell cultures results in greater expression of the high affinity interleukin-2 receptor.[135, 136] Because IL2 activity leads to a greater proliferation of T cells, both Th and Tc, IL2 is an important activator of the cellular immune response. Other studies have found selenium supplementation results in greater clonal expansion of T cell clones after exposure to antigen.[136, 137] This observation may be linked to the increased expression of the IL2 receptor, but may involve other effects of selenium. In addition, activation of T cells results in an increase in the expression of the enzyme responsible for incorporating selenium into selenocysteine. This observation strongly suggests that selenoproteins are necessary for carrying out an effective immune response. Because most selenoproteins identified to date are scavengers of reactive oxygen species, selenoprotein expression may be particularly important for the cytotoxic responses of macrophages and Tc cells. Selenoproteins such as GPx and TRx are up-regulated in activated macrophages along with genes that are responsible for effecting cytotoxic responses.[138] This may be due to their ability to reduce oxidative stress within the cell that is caused by the cytotoxic response to engulfed particles. The cytotoxic response is affected by imposing severe

oxidative stress on the damaged or intruding cell. Without adequate levels of reactive oxygen species scavengers, the reactive oxygen species produced by this process may damage the cytotoxic effecter cell. Also, selenium deficiency may reduce the cytotoxicity of NK cells, [139, 140], whereas selenium supplementation may enhance the cytotoxicity of NK cells.[141-144]

Clinical Data

3.1 It is with this background that we approach the clinical data that associates low selenium levels with increased incidence and severity of many viral diseases. A brief review of the clinical observations that have accumulated will further illuminate the diverse role of selenium in mammalian systems and provide a basis for further analysis.

As introduced in section 2, Keshan disease is a severe cardiomyopathy that is often fatal. It primarily occurs in certain areas of China where the soil is deficient in selenium which results in a selenium deficiency in the diet of the indigenous people.[101, 102] Because of the link between disease occurrence and a selenium deficient diet, it was hypothesized that selenium deficiency was the cause of the disease. However, because selenium deficient diets are not unique to parts of China, and the occurrence of Keshan disease varied on an annual and seasonal basis, a cofactor mechanism was proposed.[145] Subsequently, coxsackievirus was isolated from the heart tissue of several victims of the disease.[100] Current understanding of the disease is that coxsackievirus can be a cofactor in disease progression. It is possible that other virus could independently act in a similar manner, as various viruses are known

to cause myocarditis (e.g. HCV) or cardiomyopathy.[146, 147] Coxsackievirus appears to be able to take advantage of the reduced selenium status of the host and induce a serious form of coxsackievirus infection. Mice infected with an amyocarditic (benign) strain of coxsackievirus develop cardiomyopathy when fed a diet low in selenium.[99] Analysis of the heart tissue of the mice that were effected by the cardiomyopathy revealed that the coxsackievirus in these mice had mutated into a more virulent (myocarditic) strain.[148] Furthermore, the virus isolated from the hearts of these mice could induce severe cardiomyopathy in mice fed a diet with adequate amounts of selenium. This may be the first example of a virus mutating to a more virulent form in response to a specific dietary factor.[100] Analysis of the mutated coxsackie virus revealed that there were six mutations present that are characteristic of previously known virulent forms of the virus.[149] The mutations were not random, but were very specific.

Much of the data linking selenium and disease progression concerns RNA viruses. Selenium supplementation can prevent mouse mammary tumorogenesis and inhibit the replication of some oncogenic RNA viruses.[150] However, the most striking link between selenium status and poor disease prognosis is found in individuals infected with HIV. As reviewed previously [151], a striking selenium deficiency (as defined by low serum Se levels) is manifest among patients infected with HIV and particularly those suffering from symptoms of the disease.[152, 153] The selenium deficiency is even more severe in patients co-infected with HCV.[46] Several studies demonstrate a correlation between low serum concentration of selenium and a low number of circulating T4

cells.[46, 152, 154] In addition, selenium status is a reliable, independent predictor of disease progression.[155] In a cohort of 125 HIV infected drug users, patients with low serum selenium were 20 times more likely to die from AIDS related causes than patients with adequate levels of serum selenium. In fact, selenium status was a sixteen fold greater indicator of risk of mortality from AIDS related causes than low T-cell count defined as <200 cells/□l serum.

These observations strongly suggest the direct involvement of selenium in the progression of AIDS. Further support for this hypothesis can be found in work done to determine the effects of selenium supplementation on disease progression. Even before the causative agent of AIDS was identified as the HIV virus, it was shown that low molecular weight selenium compounds could inhibit the replication of other retroviruses in vivo and in vitro, as referenced in Schrauzer and Sacher.[156] Apparently, these results were never given the attention they deserved. Many years passed before the potential for selenium supplementation to inhibit replication of the HIV virus was investigated. Several studies suggest that selenium supplementation can slow the progression of HIV infection in vivo, and inhibit replication of the virus in vitro.[42, 156-159] However, a large-scale clinical study on the effects of selenium supplementation on AIDS progression has never been conducted.

Clinical studies have been done using selenium supplementation to treat viral infections. However, given the strong correlation of selenium status and disease progression found in AIDS, it is surprising to note that a large scale clinical trial of selenium supplementation on HIV infected patients has not been

conducted. In China, a four year trial of selenium supplementation on the population in selenium deficient regions where hepatitis B infection is endemic produced striking results.[160] Incidence of both hepatitis B infection and liver cancer was significantly reduced by selenium supplementation. Hepatitis B, a hepadnavirus, is a close relative of the retroviruses and encodes a reverse transcriptase. Also, clinical trials have established the benefit of dietary selenium supplementation on the prevention of hepatocellular carcinoma (HCC) among patients with hepatitis C virus (HCV) infection.

Selenium supplementation is also an effective treatment for Hantaan virus infection. During an outbreak of Hantaan virus infection involving 80 patients in China, selenium supplementation of 2mg/ day for nine days was sufficient to reduce mortality from 38% in the untreated group to 7% in the treated group.[161] Thus, selenium doses near thirteen times the RDA was able to reduce mortality by a dramatic 80%. Hantaan virus, like ebola, is a hemorrhagic fever associated with rupture of capillaries caused by aberrant blood clotting. Although selenium status or supplementation has never been examined in patients infected with the ebola virus, this study strongly suggests that selenium supplementation may be an effective treatment for this disease as well. This is an important observation as there are no treatments for these diseases that are recognized by conventional medical science.

One more example of the effects of selenium supplementation on viral infection is that of the flu virus. At least one study reports that selenium

supplementation can inhibit the replication of the flu virus. In addition, selenium deficiency may enhance the virulence of the influenza virus.[162, 163]

Experimentally Verified Effects of Selenium on the Host That May Directly Account For the Clinical Observations

The weight of clinical data linking selenium status to viral infection progression and outcome compels a careful analysis of the potential explanations for these observations. In this section, host factors that could play a role will be analyzed. Viral factors that could explain the link between selenium levels and disease occurrence and progression will then be discussed in section 5. Some areas of possibility that have not yet been fully examined experimentally will be included in this discussion.

Because of the benefits of proper levels of selenium on immune system function as outlined previously (sec. 2.6), it is believed that much of the clinical data can be explained by the potentiating effect selenium can have on the immune system, and cellular immunity in particular. This general observation is consistent with the fact that the cellular immune system is primarily responsible for eliminating viral infections, and the cellular immune system appears to be particularly sensitive to selenium levels.

In addition, many symptoms associated with the diseases outlined above may be potentiated by effects of selenium on the host. In the case of Keshan disease for example, severe selenium deficiency can cause acute cardiomyopathy apparently in the absence of any viral cofactor.[164-166] In these cases, selenium supplementation can alleviate the disease caused by the

deficiency.[167] However, instead of the virus causing a severe selenium deficiency that results in cardiomyopathy, research by Beck conclusively demonstrates that the virus is able to take advantage of a preexisting selenium deficiency.[148] The preexisting selenium deficiency facilitates mutations within the viral genome that lead to a more virulent form. Once the virus has mutated, it is able to cause acute cardiomyopathy even in the absence of a selenium deficiency. The general mechanism behind this etiology may be applicable to many different viral diseases, and may help explain why selenium deficiency worsens viral disease progression. One possible explanation involves the fact that a decrease in serum selenium is known to lower the effectiveness of the cellular immune system as reviewed in section 2.6. Once the immune system is weakened, viral replication may increase which could lead to an increase in the amount of mutant virus in the body. An increase in the mutant virus pool in turn increases the likely hood of selecting for a more virulent form of the virus. That selenium deficiency may facilitate viral mutation through a general mechanism applicable to many virus is supported by the finding outlined in section 3.7 that selenium deficiency leads to more virulent, mutant forms of influenza virus.[162, 168, 169]

A group of symptoms that manifest themselves as a hypothyroid-like condition is common in patients suffering from AIDS.[170-172] Low levels of serum selenium may lead to or exacerbate this hypothyroid-like condition. Low levels of T3 could be caused by an acute selenium deficiency through attenuation of deiodinase activity.[173-175] Deiodinase, the enzyme responsible

for converting T4 to T3, is a selenoprotein. Reduced levels of selenium may inhibit production of the deiodinase enzyme, and result in reduced levels of T3.

Selenium is involved in the blood clotting cascade.[176-178] Although body levels of selenium has apparently never been examined in patients suffering from viral hemorrhagic fever, the primary cause of death associated with these diseases can be logically linked to a malfunction in a process dependant upon proper levels of selenium. A severe, acute reduction in selenium levels may contribute to problems with the blood-clotting cascade. This hypothesis is strongly supported by, and may help explain the clinical observation outlined above in section 3.6 that selenium supplementation greatly reduces mortality associated with hemorrhagic fever caused by Hantaan virus infection.[151]

The recognized importance of selenium in controlling oxidant tone may play a critical role in controlling HIV disease progression (sec. 3.4). This may account for a portion of the clinical results outlined above. Increased oxidative stress in individuals infected with HIV is well documented.[179] The increase in oxidative stress is consistent with the role of selenium in antioxidant proteins and the demonstrated selenium deficiency in patients infected with HIV. Much of this may be associated with the reduction in glutathione peroxidase activity that occurs in HIV infected individuals.[46] The increase in oxidant tone may make T cells more vulnerable to other stresses or facilitate cell death via apoptosis contributing to T cell losses. Addition of lipid hydroperoxides can induce

apoptosis of T cells that have a deficiency of GPx activity associated with HIV infection.[180]

Experimentally Verified Effects of Selenium on the Virus That May Directly Account For Many Clinical Observations

Now that the direct effects of selenium on the host that relate to disease progression have been reviewed, we turn to the direct effects of selenium on viral systems as it relates to disease progression. Some of these were alluded to in the last section, but will be expanded here. Given the central role selenium plays in controlling oxidative stress as outlined in sections 2.4 and 2.5, it is worth noting that oxidative stress may have a direct effect on the replication of some viruses. HIV replication is increased by oxidative stress. As reviewed in J.AIDS, HIV replication may be linked to redox status through the activity of NFkB, a transcriptional activator that is sensitive to oxidant tone.[181] NFkB is activated upon dissociation of the inhibitor IkB. Upon activation, NFkB dimers migrate to the nucleus where they act as a transcriptional activator of certain genes associated with immune function. The dissociation of IkB is facilitated by UV light, peroxides, and inflammatory cytokines. In each of these, oxidative stress is a common result. Oxidative stress therefore leads to the dissociation of IkB and the activation of NFkB, and a reduction in oxidant tone suppresses the activity of NFkB. NFkB is able to directly activate transcription of HIV related genes through the recognition sequence for NFkB binding present in the promoter region of the HIV genome.[182] Activation of NFkB activity therefore stimulates

the promoter region of HIV and induces HIV replication.[157] This is consistent with the fact that HIV replication is stimulated by oxidative stress.[183-186]

The connection between HIV replication and NFkB activity may be even more complicated than outlined in section 5.1. GPx is an inhibitor of NFkB activation, by reducing reactive oxygen species within the cell.[183, 187] Further, increasing glutathione peroxidase activity decreases HIV replication after first being stimulated by oxidative stress.[183] However, one of the major functions of thioredoxin reductase, through it's substrate thioredoxin, is the activation of NFkB.[188, 189] The thiol group on an active site cysteine must be reduced in order for NFkB to bind DNA.[188] Thioredoxin provides the reducing power to maintain the thiol group in the reduced state. Therefore NFkB activity is regulated by GPx and TRx. The activity of cellular GPx and TRx can be directly altered by selenium status, and therefore effect the activation of NFkB and the replication of HIV in vitro.[157] However, it is important to also consider the possibility that the virus encodes GPx and TRx homologues, which would allow the viral genome to better control its own destiny. The possibility that HIV encodes GPx and TRx homologues is considered in section 9.

As mentioned in section 3.2, selenium deficient conditions promote mutations in the genome of the coxsackie virus.[190] Although the exact mechanism for this effect remains a mystery, the mutations directly facilitate progression of Keshan disease in a mouse model. Without the mutations, the coxsackie B virus is not a virulent strain. However, once mutated, it is able to

cause severe cardiomyopathy whether or not a selenium deficient diet is maintained.

In addition, as a potential explanation to the clinical data outlined in section 3.7, selenium deficiency has a similar effect on the influenza virus. Selenium deficiency facilitates mutations within the influenza genome leading to more virulent strains.[163]

Serum Selenium Levels in Viral Infection

Given the overwhelming evidence for a decrease in serum selenium status in patients suffering from many viral infections including HIV, what could be the cause of the reduction in selenium levels? The strongest evidence for an association between a reduction in serum selenium and viral disease progression is that documented for HIV infection, as discussed previously in section 3.3. The conventional explanation for the apparent selenium deficiency is the wasting syndrome associated with AIDS progression.[191] Poor absorption through the gut is blamed for loss of lean muscle mass and a reduction in the body levels of several nutrients.[192, 193] It is hypothesized that the selenium deficiency is simply a byproduct of disease progression caused by mal-absorption of selenium through the intestinal tract. Under the framework of this hypothesis, the reduction in serum selenium is just a surrogate marker of progression of the disease and is not indicative of any particular role of selenium. However, as briefly reviewed by Taylor, et al., this hypothesis does not completely account for experimental observation.[194] In particular, an acute selenium deficiency is observed in patients infected with HIV early in the disease before the wasting

syndrome takes hold.[195] Also, selenium depletion occurs even in the absence of diarrhea or mal-absorption.[196] In addition, urinary excretion of selenium, which is indicative of selenium intake, is similar in HIV-1 infected individuals and controls.

An alternative explanation for the apparent selenium deficiency is acute phase effects. Best documented in the case of iron, “acute phase effects” refers to the redistribution or sequestration of body minerals during infection.[197-200] The proposed physiological basis for acute phase effects is the body’s need to deprive the pathogen of critical nutrients. Measurements of body levels of nutrients affected by acute phase effects can vary widely depending on the source of the measurement even though total body levels of the nutrient may not change. Although some have interpreted this possibility to be evidence against a specific role for selenium in disease progression, the opposite conclusion is a logical possibility. That selenium may be sequestered or redistributed in response to infection as a defense mechanism implies that selenium may be a critical element for the replication of the disease agent. The only other logical explanation is that selenium is sequestered to parts of the body that have an increased need for this essential nutrient during infection.

Assessment of the Clinical Data and Potential Explanations Outlined in Section 6

There is more than one possible explanation for the unique role of selenium in human viral disease etiology reviewed above. The antioxidant benefits of simple Se compounds and known selenoenzymes may be so critical

to host homeostasis and immune system function that any change in selenium status leads to problems. This suggestion has been used to explain all of the observations outlined above. It is a fact that simple selenium compounds have antioxidant activity that sometimes mimics that of selenoenzymes. However, some of the antioxidant benefits of simple selenium compounds may in part be caused by the subsequent increase of the metabolically available selenium for production of selenoproteins.

The discovery of selenocysteine and selenoproteins are relatively recent. The mechanism of transcribing UGA codons as selenocysteine instead of termination is an even more recent discovery, and details of this mechanism continue to be revealed. These observations lead to another possible explanation for the effects of selenium on disease progression. There may be many more selenoproteins that have yet to be discovered and these could be involved in diverse biological processes related to disease progression. Several facts lead to this conclusion. Mentioned previously, they deserve to be explicitly outlined again. The facts are: 1) Insertion of selenocysteine into a growing polypeptide chain is a complex process. 2) Variations of the SECIS continue to be found. 3) Only fragmentary details about the proteins that interact with the SECIS and the UGA codon facilitating insertion of selenocysteine is known. These potential selenoproteins could be expected to carry out physiologically important functions. Because the production of known selenoproteins is affected by selenium levels, it is logical to conclude that the production of other potential selenoproteins could

also be expected to be effected by selenium levels. Further, new selenoproteins may be expressed by either the host or the virus (sections 8 and 9).

Obviously the known selenoproteins such as GPx, TRx, and deiodinase play a role in disease etiology as outlined above. If there are yet to be discovered selenoproteins that are involved in disease etiology, their effects are in addition to those of the known selenoproteins. The presence of yet to be discovered selenoproteins in both the host and the viral genomes could help explain the complex role of selenium in disease progression outlined above.

Known and Potential Host Selenoproteins

Gel electrophoresis studies conducted using a radioactive isotope of selenium, selenium-75, have been conducted to find selenoproteins in cell extracts. Between 14 and 28 selenoproteins have been identified in cell extracts of various types.[201, 202] Many have not been positively identified.[203] Other studies have identified potential selenoproteins in silico.[68] These results strongly suggest that there are at least some yet to be identified selenoproteins expressed in mammalian systems. The function of these unidentified selenoproteins is unknown, and some of them may play an important role in explaining the complex clinical effects linked to selenium outlined above.

Potential Viral Selenoproteins

The potential existence of selenoproteins encoded within viral genomes has been reported in several studies, and reviewed by others.[151, 204-206] The most widely accepted example of a viral selenoprotein is that found in the Molluscum contagiosum virus genome. Dr. Bernard Moss's group at the NIH

sequenced the M. contagiosum genome, and analyzed its sequence.[207] They identified a gene that shares a 76% homology with glutathione peroxidase at the amino acid level. However, this protein has not yet been confirmed by functional studies.

Given the evidence presented above, it is not surprising that many labs are investigating the possibility that there are undiscovered selenoproteins encoded in the genome of the host and or the virus. As outlined in section 5, viral gene expression is closely tied to host gene expression and cellular oxidant tone. Cellular oxidant tone, in turn, is affected by selenoproteins such as GPx and TRx. Therefore, it could make logical sense for the virus to encode selenoprotein modules capable of affecting oxidant tone and gene expression. In fact, this would not be the first example of a virus encoding proteins capable of altering expression of certain genes. A human pox virus encodes chemokine modules.[208-210]

The HIV genome has been scrutinized for potential selenoproteins since Taylor's theoretical analysis in 1994 [204], and this data has been reviewed previously.[211] Comparative sequence analysis was used to identify potential genes and each reading frame of the viral genome was carefully analyzed. Three potential selenoproteins were found in the HIV-1 genome.[49] One of these potential selenoproteins is in a region overlapping the envelope gene in the –1 reading frame, and has an in frame UGA codon. Recently, this gene was identified as a homologue of GPx containing a GPx active site consensus sequence.[194] The similarity score for the alignment of this gene and known

GPx sequences is five standard deviations above the expected score for a random alignment of sequences of like composition. Thus, the chances of this being a random alignment are less than one in one million. More recently, catalytic GPx activity has been experimentally verified when expressed in vitro.[212] Although the importance of the virally encoded GPx module remains speculative, it is interesting to note that Sandstrom et al. have reported that the addition of a specific inhibitor of GPx actually decreases viral replication.[180] This result is not what may be expected from conventional analysis of the link between viral replication and oxidative tone. GPx is the primary enzyme responsible for reducing oxidative tone, and its inhibition would seem to lead to increased viral replication because of increased oxidative stress. This is not what is observed, and a virally encoded GPx module may help explain this paradoxical result.

Interestingly, the same gene has been identified in the hepatitis C virus (HCV).[206] The GPx gene homologue in HCV is also in the –1 reading frame, overlaps the NS4a gene, and also has an in frame UGA codon. This putative HCV GPx gene is also highly similar to known GPx gene sequences. The similarity score for the alignment of this gene and known GPx sequences is 6.2 standard deviations above the expected score for a random alignment of sequences of like composition. When compared only to mammalian plasma GPx sequences, the alignment is even more similar at 6.7 standard deviations above a random alignment of similar composition.

In the case of both of these putative GPx genes, the similarity alignment spans the active site of known GPx genes.[204, 206] Furthermore, the UGA codons that span the active site are highly conserved in HIV-1 and HCV-1b. Interestingly, HCV strain 1b is most highly associated with progression to cirrhosis, hepatocellular carcinoma, and resistance to interferon treatment. As mentioned in section 3.3, the progressive decline in serum selenium levels is more rapid in patients who are co-infected with HCV.[46] This observation bolsters the hypothesis that there is a link between selenium and HCV infection.

Other potential selenoproteins were identified in the HIV-1 nef coding region and the protease coding region.[204] As in the case of the GPx homologues, these putative genes are found in the –1 reading frame and have highly conserved in frame UGA codons. The putative gene found in the nef region bears significant similarity to thioredoxin reductase, and the putative gene in the protease region bears significant similarity to NFkB.

In addition to potential genes, several pseudoknots and slippery sequences were found in areas where frame shifting would be necessary to express the potential selenoproteins.[204, 213] A novel –1 frameshifting site in the nef coding region associated with the highly conserved UGA codons has been experimentally verified. An RNA pseudoknot that was predicted to overlap the coding region of the HIV-1 reverse transcriptase active site has also been experimentally verified by another lab. Also, potential SECIS elements have been independently identified.[48]

The predictions made in Taylor's analysis in 1994 are supported by both recent clinical data and experimental observations. However, data challenging the validity of the HIV selenoprotein hypothesis has also been presented. Gladyshev et al. reported that cells infected with HIV grown in radioactive selenium-75 enhanced media did not express a selenoprotein that would match the predicted molecular weight of the putative HIV selenoproteins.[214] Careful analysis of the data presented in the paper, however, reveals that there is a band on the gel that is apparently ignored by the authors.

Other viral genomes may also encode selenoproteins. An interesting example is found in the Ebola Zaire virus. A potential gene has been identified in Ebola Zaire that has as many as 16 UGA selenocysteine codons, and potentially the structural features required for translating these as selenocysteine.[215] If this gene is expressed as part of the replication cycle of the Ebola Zaire virus, it could place a severe burden on the selenium stores of the host. Such an acute reduction in the selenium available to the host could contribute to the severe symptoms suffered by patients infected with this virus. As mentioned previously, selenium is necessary for the blood-clotting cascade to function normally (section 4.5). Unfortunately, serum selenium levels have never been tested in persons suffering from Ebola infection.

Potential selenoprotein genes have been identified in the Coxsackie virus B3 genome as well.[213] This is the same strain studied by Beck et al. in their model for Keshan disease mentioned previously. Interestingly, one of these genes may encode a highly truncated GPx module.

Reconciliation of the Observed Selenium Abnormalities With Viral Selenoprotein Theory

As reviewed by Taylor, 1997, although it may seem counter intuitive that a severe selenium deficiency could be linked to increased viral replication, and selenium appears to inhibit viral replication while at the same time the virus requires selenium, there is a possible explanation for these observations. If the viral selenoprotein hypothesis proves to be correct, it would not be the first time that an infection was associated with a reduction in the amount of some nutrient required by the infectious agent. The classic example of this phenomenon in relation to viral replication is the induction of retroviral expression in arginine deficient cell cultures even though arginine is an essential component of most viral proteins. This effect is probably due to some sort of repressor mechanism analogous to the tryptophan repressor in bacteria. The data reviewed here are highly consistent with the possibility that viral infection may respond to selenium deficiency in a similar fashion. Also of note, the potential GPx selenoproteins encoded by viral genomes could also have a repressive effect on viral replication. Because oxidative stress can induce viral replication, additional GPx would be expected to reduce oxidative tone and repress viral replication. This observation could be very important because potential selenium dependent glutathione peroxidase genes have, as reviewed above, been found in several viral genomes. This molecular switch could enable the virus to lie dormant until oxidative stress occurs. The chances of successful replication and survival of the virus may be maximized in this way.

Conclusion

That selenium status is linked to, or is at least a marker of viral disease progression has been proven. This fact may not be entirely accounted for by the known role of selenium as an antioxidant. Although viral infections may result in oxidative stress, and the body may benefit from a boost in its antioxidant defenses, the selenium deficiency that occurs during many infections is not easily explained. One possibility is that viral selenoprotein synthesis sequesters a significant amount of selenium. Given the large amount of clinical data that links a reduction in serum selenium to poor disease progression, it seems logical, even obvious, that a large-scale clinical trial of the benefit of selenium supplementation should be conducted. It seems likely that selenium nutritional therapy can be an inexpensive and effective partner to current antiviral regimens.

Because viral gene transcription can be closely linked to redox status within the host cell, and redox status is partially regulated by the activities of selenoproteins, it is not surprising that the virus may encode selenoprotein modules capable of affecting oxidant tone. As pointed out in section 9, some viruses are known to encode chemokine analogues which have similar purposes and effects. In addition, GPx modules have been identified in several viral genomes by sequence analysis, and one has been shown to be enzymatically active experimentally.

Selenoprotein synthesis is a complex process that has only recently been elucidated. Current sequence analysis methods may easily miss potential SECIS elements, and potential selenoprotein modules. In addition, viral

genomes are very efficient, and may encode truncated versions of the elements necessary to cause the translation of UGA as selenocysteine. This may cause them to be even more difficult to find, particularly when the minimal structure for a functional SECIS is uncertain. The potential role of viral selenoproteins in viral gene regulation makes them important potential targets of antiviral therapy.

In summary, several key facts about selenium and selenoproteins illustrate the importance of continued, careful examination of the potential role of selenoproteins in viral disease progression. 1) The role of selenium in viral disease progression is complex but clear. 2) The cellular machinery required for translating the UGA codon as selenocysteine is considerable. This may make some cellular selenoproteins difficult to identify. Many of the potential viral selenoproteins identified so far are associated with frame shifting mutations in the reading frame of another gene, do not have individual start codons, and may be highly truncated. This may make viral selenoproteins even more difficult to identify. A better understanding of these potential selenoproteins could reveal new targets for drug therapy, as well as illuminate the benefit of nutritional therapy. 3) Selenoproteins are intimately involved in the regulation of viral gene transcription through redox tone. It therefore could make logical sense for the virus to encode selenoprotein modules capable of effecting redox tone. 4) A large clinical trial of the benefits of selenium supplementation has yet to be conducted, yet holds great promise.

REFERENCES

1. Yanofsky, C., *Gene Structure and Protein Structure.* Sci. Am., 1967. **216**(5): p. 80-94.
2. Davies, M.V. and R.J. Kaufman, *The Sequence Context of the Initiation Codon in the Encephalomyocarditis Virus Leader Modulates Efficiency of Internal Translation Initiation.* Journal of Virology, 1992. **66**(4): p. 1924-1932.
3. Kozak, M., *Initiation of translation in prokaryotes and eukaryotes.* Gene, 1999. **234**(2): p. 187-208.
4. Berry, M.J., et al., *Functional-Characterization of the Eukaryotic Secis Elements Which Direct Selenocysteine Insertion at Uga Codons.* Embo Journal, 1993. **12**(8): p. 3315-3322.
5. Venter, J.C., Adams, M.D., Myers, E.W., et.al., *The Sequence of the Human Genom.* Science, 2001. **291**: p. 1304-1351.
6. Arner, E.S.J., et al., *High-level expression in Escherichia coli of selenocysteine-containing rat thioredoxin reductase utilizing gene fusions with engineered bacterial-type SECIS elements and co-expression with the selA, selB and selC genes.* Journal of Molecular Biology, 1999. **292**(5): p. 1003-1016.
7. Copeland, P.R., et al., *A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs.* Embo Journal, 2000. **19**(2): p. 306-314.

8. Martin, G.W. and M.J. Berry, *Eukaryotic selenocysteine incorporation: Mechanistic insights.* Phosphorus Sulfur and Silicon and the Related Elements, 1998. **136**: p. 309-320.

9. Tujebajeva, R.M., et al., *Decoding apparatus for eukaryotic selenocysteine insertion.* Embo Reports, 2000. **1**(2): p. 158-163.

10. Arkov, A.L. and E.J. Murgola, *Ribosomal RNAs in translation termination: Facts and hypotheses.* Biochemistry-Moscow, 1999. **64**(12): p. 1354-1359.

11. Mottagui-Tabar, S., *Quantitative analysis of in vivo ribosomal events at UGA and UAG stop codons.* Nucleic Acids Research, 1998. **26**(11): p. 2789-2796.

12. Chan, T.S. and A. Garen, *Amino Acid Substitutions Resulting from Suppression of Nonsense Mutations .5. Tryptophan Insertion by Su9+ Gene, a Suppressor of Uga Nonsense Triplet.* Journal of Molecular Biology, 1970. **49**(1): p. 231-&.

13. Kuchino, Y. and T. Muramatsu, *Nonsense suppression in mammalian cells.* Biochimie, 1996. **78**(11-12): p. 1007-1015.

14. Smiley, B.K. and F.C. Minion, *Enhanced Readthrough of Opal (Uga) Stop Codons and Production of Mycoplasma-Pneumoniae P1 Epitopes in Escherichia-Coli.* Gene, 1993. **134**(1): p. 33-40.

15. Goringer, H.U., et al., *Mutations in 16s Ribosomal-Rna That Affect Uga (Stop Codon)-Directed Translation Termination.* Proceedings of the

National Academy of Sciences of the United States of America, 1991. **88**(15): p. 6603-6607.

16. Jemiolo, D.K., F.T. Pagel, and E.J. Murgola, *UGA suppression by a mutant RNA of the large ribosomal subunit.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(26): p. 12309-12313.
17. Prescott, C.D., B. Kleuvers, and H.U. Goringer, *A Ribosomal-Rna-Messenger Rna Base-Pairing Model for Uga-Dependent Termination.* Biochimie, 1991. **73**(7-8): p. 1121-1129.
18. Konan, K.V. and C. Yanofsky, *Role of ribosome release in regulation of tna operon expression in Escherichia coli.* Journal of Bacteriology, 1999. **181**(5): p. 1530-1536.
19. Muramatsu, T. and Y. Kuchino, *Translational control through nonsense suppression.* Seikagaku, 1997. **69**(1): p. 36-39.
20. Poole, E.S., C.M. Brown, and W.P. Tate, *The Identity of the Base Following the Stop Codon Determines the Efficiency of in-Vivo Translational Termination in Escherichia-Coli.* Embo Journal, 1995. **14**(1): p. 151-158.
21. Mottaguitabar, S., A. Bjornsson, and L.A. Isaksson, *The 2nd to Last Amino-Acid in the Nascent Peptide as a Codon Context Determinant.* Embo Journal, 1994. **13**(1): p. 249-257.

22. Zhang, S.P., M. RydenAulin, and L.A. Isaksson, *Functional interaction between release factor one and P-site peptidyl-tRNA on the ribosome.* Journal of Molecular Biology, 1996. **261**(2): p. 98-107.

23. Tate, W.P. and S.A. Mannering, *Three, four or more: The translational stop signal at length.* Molecular Microbiology, 1996. **21**(2): p. 213-219.

24. MottaguiTabar, S. and L.A. Isaksson, *Influence of the last amino acid in the nascent peptide on EF-Tu during decoding.* Biochimie, 1996. **78**(11-12): p. 953-958.

25. Bonetti, B., et al., *The Efficiency of Translation Termination Is Determined by a Synergistic Interplay between Upstream and Downstream Sequences in Saccharomyces-Cerevisiae.* Journal of Molecular Biology, 1995. **251**(3): p. 334-345.

26. Harrell, L., U. Melcher, and J.F. Atkins, *Predominance of six different hexanucleotide recoding signals 3 ' of read-through stop codons.* Nucleic Acids Research, 2002. **30**(9): p. 2011-2017.

27. Levander, O.A., *A Global View of Human Selenium Nutrition.* Annual Review of Nutrition, 1987. **7**: p. 227-250.

28. Fuchs, O., *Selenoproteins.* Chemicke Listy, 1996. **90**(7): p. 444-450.

29. Martin, G.W., J.W. Harney, and M.J. Berry, *Selenocysteine incorporation in eukaryotes: Insights into mechanism and efficiency from sequence, structure, and spacing proximity studies of the type 1 deiodinase SECIS element.* Rna-a Publication of the Rna Society, 1996. **2**(2): p. 171-182.

30. Kollmus, H., L. Flohe, and J.E.G. McCarthy, *Analysis of eukaryotic mRNA structures directing cotranslational incorporation of selenocysteine.* Nucleic Acids Research, 1996. **24**(7): p. 1195-1201.

31. Rother, M., et al., *Selenoprotein synthesis in archaea.* Biofactors, 2001. **14**(1-4): p. 75-83.

32. Arvilommi, H., et al., *Selenium and Immune Functions in Humans.* Infection and Immunity, 1983. **41**(1): p. 185-189.

33. Kiremidjianschumacher, L. and G. Stotzky, *Selenium and Immune-Responses.* Environmental Research, 1987. **42**(2): p. 277-303.

34. Odell, J.R., et al., *Serum Selenium Concentrations in Rheumatoid-Arthritis.* Annals of the Rheumatic Diseases, 1991. **50**(6): p. 376-378.

35. Kavanaughmchugh, A.L., et al., *Selenium Deficiency and Cardiomyopathy in Acquired-Immunodeficiency-Syndrome.* Journal of Parenteral and Enteral Nutrition, 1991. **15**(3): p. 347-349.

36. Delilbasi, E., et al., *Selenium and Behcets-Disease.* Biological Trace Element Research, 1991. **28**(1): p. 21-25.

37. Mei, W.D., et al., *Study of Immune Function of Cancer-Patients Influenced by Supplemental Zinc or Selenium-Zinc Combination.* Biological Trace Element Research, 1991. **28**(1): p. 11-20.

38. Ip, C. and G. White, *Mammary-Cancer Chemoprevention by Inorganic and Organic Selenium - Single Agent Treatment or in Combination with Vitamin-E and Their Effects on Invitro Immune Functions.* Carcinogenesis, 1987. **8**(12): p. 1763-1766.

39. Langman, M. and P. Boyle, *Chemoprevention of colorectal cancer.* Gut, 1998. **43**(4): p. 578-585.

40. Manteroatienza, E., et al., *Selenium Status and Immune Function in Asymptomatic Hiv-1 Seropositive Men.* Nutrition Research, 1991. **11**(11): p. 1237-1250.

41. Dworkin, B.M., *Selenium Deficiency in Hiv-Infection and the Acquired-Immunodeficiency-Syndrome (Aids).* Chemico-Biological Interactions, 1994. **91**(2-3): p. 181-186.

42. Schrauzer, G.N. and J. Sacher, *Selenium in the Maintenance and Therapy of Hiv-Infected Patients.* Chemico-Biological Interactions, 1994. **91**(2-3): p. 199-205.

43. Beck, M.A., *The role of nutrition in viral disease.* Journal of Nutritional Biochemistry, 1996. **7**(12): p. 683-690.

44. Pace, G.W. and C.D. Leaf, *The Role of Oxidative Stress in Hiv Disease.* Free Radical Biology and Medicine, 1995. **19**(4): p. 523-528.

45. Baum, M.K., et al., *High risk of HIV-related mortality is associated with selenium deficiency.* Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 1997. **15**(5): p. 370-374.

46. Look, M.P., et al., *Serum selenium, plasma glutathione (GSH) and erythrocyte glutathione peroxidase (GSH-Px)-levels in asymptomatic versus symptomatic human immunodeficiency virus-1 (HIV-1)-infection.* European Journal of Clinical Nutrition, 1997. **51**(4): p. 266-272.

47. Zhang, W., et al., *Selenium-dependent glutathione peroxidase modules encoded by RNA viruses.* Biological Trace Element Research, 1999. **70**(2): p. 97-116.

48. Grate, L., *Potential SECIS elements in HIV-1 strain HXB2.* Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 1998. **17**(5): p. 398-403.

49. Taylor, E.W., et al., *A Basis for New Approaches to the Chemotherapy of Aids - Novel Genes in Hiv-1 Potentially Encode Selenoproteins Expressed by Ribosomal Frameshifting and Termination Suppression.* Journal of Medicinal Chemistry, 1994. **37**(17): p. 2637-2654.

50. Low, S.C. and M.J. Berry, *Knowing when not to stop: Selenocysteine incorporation in eukaryotes.* Trends in Biochemical Sciences, 1996. **21**(6): p. 203-208.

51. Grundner-Culemann, E., et al., *Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes.* Rna-a Publication of the Rna Society, 1999. **5**(5): p. 625-635.

52. Mizutani, T., K. Tanabe, and K. Yamada, *G-U base pair in the eukaryotic selenocysteine tRNA is important for interaction with SePF, the putative selenocysteine-specific elongation factor.* Febs Letters, 1998. **429**(2): p. 189-193.

53. Fourmy, D., E. Guittet, and S. Yoshizawa, *Structure of prokaryotic SECIS mRNA hairpin and its interaction with elongation factor SelB.* Journal of Molecular Biology, 2002. **324**(1): p. 137-150.

54. Walczak, R., P. Carbon, and A. Krol, *An essential non-Watson-Crick base pair motif in 3 ' UTR to mediate selenoprotein translation.* Rna-a Publication of the Rna Society, 1998. **4**(1): p. 74-84.

55. Martin, G.W., J.W. Harney, and M.J. Berry, *Functionality of mutations at conserved nucleotides in eukaryotic SECIS elements is determined by the identity of a single nonconserved nucleotide.* Rna-a Publication of the Rna Society, 1998. **4**(1): p. 65-73.

56. Sandman, K.E., et al., *Revised Escherichia coli selenocysteine insertion requirements determined by in vivo screening of combinatorial libraries of SECIS variants.* Nucleic Acids Research, 2003. **31**(8): p. 2234-2241.

57. Forchhammer, K., K.P. Rucknagel, and A. Bock, *Purification and Biochemical-Characterization of Selb, a Translation Factor Involved in Selenoprotein Synthesis.* Journal of Biological Chemistry, 1990. **265**(16): p. 9346-9350.

58. Chittum, H.S., et al., *Rabbit beta-globin is extended beyond its UGA stop codon by multiple suppressions and translational reading gaps.* Biochemistry, 1998. **37**(31): p. 10866-10870.

59. Elgert, K.D., *Immunology: Understanding the Immune System*. Prepublication Manuscript. 1994, Blacksburg, VA: Virginia Polytechnic Institute and State University. 622.

60. Ansari-Lari, M.A., Muzny, D.M., Lu, J., Lu, F., et.al., *A gene-rich cluster between the CD4 and triosehosphate isomerase genes at human chromosome 12p13.* Genome Research, 1996. **6**(4): p. 314-326.

61. Holdorf, A.D., et al., *Regulation of Lck activity by CD4 and CD28 in the immunological synapse.* Nature Immunology, 2002. **3**(3): p. 259-264.

62. Foti, M., et al., *p56(Lck) anchors CD4 to distinct microdomains on microvilli.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(4): p. 2008-2013.

63. Taylor, E.W., *Selenium and Cellular-Immunity - Evidence That Selenoproteins May Be Encoded in the +1-Reading-Frame Overlapping the Human Cd4, Cd8, and Hla-Dr Genes.* Biological Trace Element Research, 1995. **49**(2-3): p. 85-95.

64. Facchiano, A., F. Facchiano, and J. Vanrenswoude, *Divergent Evolution May Link Human-Immunodeficiency-Virus Gp41 to Human Cd4.* Journal of Molecular Evolution, 1993. **36**(5): p. 448-457.

65. Facchiano, A., *Investigating Hypothetical Products from Noncoding Frames (Hypnofs).* Journal of Molecular Evolution, 1995. **40**(6): p. 570-577.

66. Hershko, A., Ciechanover, A.,, *The ubiquitin system for protein degradation.* Annu. Rev. Biochem, 1992. **61**: p. 761-807.

67. Voet, D., Voet, Judith, *Biochemistry*. 1995, New York: John Wiley & Sons.

68. Lescure, A., et al., *Novel selenoproteins identified in silico and in vivo by using a conserved RNA structural motif.* Journal of Biological Chemistry, 1999. **274**(53): p. 38147-38154.

69. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., et.al., *Gapped BLAST and PSI-BLAST: an new generation of protein database search programs.* Nucleic Acids Research, 1997. **25**: p. 3389-3402.

70. Fagegaltier, D., et al., *Structural analysis of new local features in SECIS RNA hairpins.* Nucleic Acids Research, 2000. **28**(14): p. 2679-2689.

71. Strausberg, R.L., Feingold, E.A., Grouse, L.H., et.al., *Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.* Proceedings of the National Academy of Sciences USA, 2002. **99**(26): p. 16899-16903.

72. Sulston, J.E., and Waterston, R., *Toward a complete human genome sequence.* Genome Research, 1998. **8**(11): p. 1097-1108.

73. Bulfone, A., Menguzzato, E., Broccoli, V., et.al., *Barhl 1, a gene belonging to a new subfamily of mammalian homeobox genes, is expressed in migrating neurons of the CNS.* Hum. Mol. Genet., 2000. **9**(9): p. 1443-1452.

74. White, O., Eisen, J.A., Heidelberg, J.F., et.al., *Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.* Science, 1999. **286**(5444): p. 1571-1577.

75. Garnier, J., Gibrat, J.F., Robson, B., *GOR secondary structure prediction method version IV.* Methods in Enzymology, 1996. **266**: p. 540-553.

76. Rost, B., *PredictProtein.* Methods in Enzymology, 1996. **266**: p. 525-539.

77. Hofmann, K., Bucher, P., Falquet, L., Bairoch, A., *The PROSITE database, its status in 1999.* Nucleic Acids Research, 1999. **27**: p. 215-219.

78. Corpet, F., Servant, F., Gouzy, J., Kahn, D., *ProDom and ProDom-CG: tools for protein domain analysis and qhole genome comparisons.* Nucleic Acids Research, 2000. **28**: p. 267-269.

79. Rost, B., *PHD: predicting one-dimensional protein structure by profile based neural networks.* Methods in Enzymology, 1996. **266**: p. 525-539.

80. Rost, B., *Short yeast ORFs: Expressed protein or not?* Unpublished, 2000.

81. Davis, R., *Personal Communication.* Unpublished, 2001.

82. Pharmacia, *Hitrap SP Column Literature*. 2000.

83. Pharmacia, *G25 Column Literature*. 2000.

84. Robyt, J.F., White, Bernard J., *Biochemical Techniques Theory and Practice*. 1990, Prospect Heights, Illinois: Waveland Press, Inc. 407.

85. Harlow, E., Lane, D., *Antibodies-a Laboratory Manual*. 1988: Cold Spring Harbor Laboratory.

86. Davis, B., Keuhl, *Basic Methods in Molecular Biology*. 1993.

87. Lowry, O.H., Rosebrough, N.J., Farr, A.L., Randall, R.J., *Protein measurement with the Folin phenol reagent.* J.Biol. Chem., 1951. **193**: p. 265.

88. Davis, B., Keuhl, *Basic Methods in Molecular Biology*. Second ed. 1994, Norwalk, CT: Appelton & Lange.

89. Corporation, P., *ProtoBlot II AP System Technical Manual*. 2001, Promega.

90. Lederman, S., Chess, L., Yellin, M.J., *Murine monoclonal antibody 5c8 recognizes a human glycoprotein on the surface of T-lymphocytes, compositions containing same.*, in *USPTO*. 1991, The Trustees of Columbia University in the City of New York: USA.

91. Doyle, K., ed. *Protocols and Applications Guide*. Third ed. 1996, Promega Corporation: USA. 404.

92. Suttle, N.F. and D.G. Jones, *Recent developments in trace element metabolism and function: trace elements, disease resistance and immune responsiveness in ruminants.* J Nutr, 1989. **119**(7): p. 1055-61.

93. Larsen, H.J., K. Moksnes, and G. Overnes, *Influence of selenium on antibody production in sheep.* Res Vet Sci, 1988. **45**(1): p. 4-10.

94. Odiawo, G.O., *The relationship between selenium deficiency and the development of pulmonary and subcutaneous emphysema in bovine ephemeral fever virus- infected cattle.* Onderstepoort J Vet Res, 1989. **56**(2): p. 123-5.

95. Sammalkorpi, K., et al., *Serum selenium in acute infections.* Infection, 1988. **16**(4): p. 222-4.

96. Rayman, M.P., *The importance of selenium to human health.* Lancet, 2000. **356**(9225): p. 233-41.

97. Holben, D.H. and A.M. Smith, *The diverse role of selenium within selenoproteins: a review [see comments].* J Am Diet Assoc, 1999. **99**(7): p. 836-43.

98. Group, K.D.R., *Epidemiologic studies on the etiologic relationship of selenium and Keshan Disease.* Chinese Medical Journal, 1979. **92**: p. 477-482.

99. Beck, M.A., et al., *Benign human enterovirus becomes virulent in selenium-deficient mice.* J Med Virol, 1994. **43**(2): p. 166-70.

100. Levander, O.A. and M.A. Beck, *Interacting nutritional and infectious etiologies of Keshan disease - Insights from Coxsackie virus B-induced myocarditis in mice deficient in selenium or vitamin E.* Biological Trace Element Research, 1997. **56**(1): p. 5-21.

101. Tan, J., *The atlas of endemic diseases and their environments in the People's Republic of China*. 1989, Beijing: Science Press.

102. Fordyce, F.M., *Soil, grain and water chemistry and human selenium imbalances in Enshi District, Hubei Province, China.* British Geological Survey Overseas Geology Series Technical Report, 1998. **WC/96/54**.

103. Berry, M.J., et al., *Selenocysteine Confers the Biochemical-Properties Characteristic of the Type-I Iodothyronine Deiodinase.* Journal of Biological Chemistry, 1991. **266**(22): p. 14155-14158.

104. Berry, M.J., et al., *Recognition of Uga as a Selenocysteine Codon in Type-I Deiodinase Requires Sequences in the 3' Untranslated Region.* Nature, 1991. **353**(6341): p. 273-276.

105. Kurata, H., et al., *The 5' Untranslated Region of the Human Cellular Glutathione-Peroxidase Gene Is Indispensable for Its Expression in Cos-7 Cells.* Febs Letters, 1992. **312**(1): p. 10-14.

106. Shen, Q.C., F.F. Chu, and P.E. Newburger, *Sequences in the 3'-Untranslated Region of the Human Cellular Glutathione-Peroxidase Gene Are Necessary and Sufficient for Selenocysteine Incorporation at the Uga Codon.* Journal of Biological Chemistry, 1993. **268**(15): p. 11463-11469.

107. Berry, M.J., et al., *Selenocysteine Insertion or Termination - Factors Affecting Uga Codon Fate and Complementary Anticodon-Codon Mutations.* Nucleic Acids Research, 1994. **22**(18): p. 3753-3759.

108. Shen, Q.C., P.A. McQuilkin, and P.E. Newburger, *RNA-binding proteins that specifically recognize the selenocysteine insertion sequence of human cellular glutathione peroxidase mRNA.* Journal of Biological Chemistry, 1995. **270**(51): p. 30448-30452.

109. Shen, Q.C., J.L. Leonard, and P.E. Newburger, *Structure and Function of the Selenium Translation Element in the 3'-Untranslated Region of Human Cellular Glutathione-Peroxidase Messenger-Rna.* Rna-a Publication of the Rna Society, 1995. **1**(5): p. 519-525.

110. Hubert, N., et al., *A protein binds the selenocysteine insertion element in the 3'-UTR of mammalian selenoprotein mRNAs.* Nucleic Acids Research, 1996. **24**(3): p. 464-469.

111. Lesoon, A., et al., *An RNA-binding protein recognizes a mammalian selenocysteine insertion sequence element required for cotranslational*

incorporation of selenocysteine. Molecular and Cellular Biology, 1997. **17**(4): p. 1977-1985.

112. Buettner, C., J.W. Harney, and P.R. Larsen, *The 3 '-untranslated region of human type 2 iodothyronine deiodinase mRNA contains a functional selenocysteine insertion sequence element.* Journal of Biological Chemistry, 1998. **273**(50): p. 33374-33378.
113. Weiss, S.L. and R.A. Sunde, *Cis-acting elements are required for selenium regulation of glutathione peroxidase-1 mRNA levels.* Rna-a Publication of the Rna Society, 1998. **4**(7): p. 816-827.
114. Fujiwara, T., et al., *A SECIS binding protein (SBP) is distinct from selenocysteyl-tRNA protecting factor (SePF).* Biochimie, 1999. **81**(3): p. 213-218.
115. Low, S.C., et al., *SECIS-SBP2 interactions dictate selenocysteine incorporation efficiency and selenoprotein hierarchy.* Embo Journal, 2000. **19**(24): p. 6882-6890.
116. Mizutani, T. and T. Fujiwara, *SBP, SECIS binding protein, binds to the RNA fragment upstream of the Sec UGA codon in glutathione peroxidase mRNA.* Molecular Biology Reports, 2000. **27**(2): p. 99-105.
117. Fagegaltier, D., et al., *Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation.* Embo Journal, 2000. **19**(17): p. 4796-4805.

118. Motchnik, P.A. and A.L. Tappel, *Multiple Selenocysteine Content of Selenoprotein-P in Rats.* Journal of Inorganic Biochemistry, 1990. **40**(3): p. 265-269.

119. Korotkov, K.V., et al., *Mammalian selenoprotein in which selenocysteine (Sec) incorporation is supported by a new form of Sec insertion sequence element.* Molecular and Cellular Biology, 2002. **22**(5): p. 1402-1411.

120. Hubert, N., et al., *RNAs mediating cotranslational insertion of selenocysteine in eukaryotic selenoproteins.* Biochimie, 1996. **78**(7): p. 590-596.

121. Fagegaltier, D., et al., *The selenocysteine insertion sequence binding protein SBP is different from the Y-box protein dbpB.* Biochimie, 2000. **82**(2): p. 117-122.

122. Shen, Q.C., et al., *Identification and molecular cloning of a human selenocysteine insertion sequence-binding protein - A bifunctional role for DNA-binding protein B.* Journal of Biological Chemistry, 1998. **273**(10): p. 5443-5446.

123. Wen, W., S.L. Weiss, and R.A. Sunde, *UGA codon position affects the efficiency of selenocysteine incorporation into glutathione peroxidase-1.* Journal of Biological Chemistry, 1998. **273**(43): p. 28533-28541.

124. Berry, M.J., et al., *Substitution of Cysteine for Selenocysteine in Type-I Iodothyronine Deiodinase Reduces the Catalytic Efficiency of the Protein but Enhances Its Translation.* Endocrinology, 1992. **131**(4): p. 1848-1852.

125. Hazebrouck, S., et al., *Substituting selenocysteine for catalytic cysteine 41 enhances enzymatic activity of plant phospholipid hydroperoxide glutathione peroxidase expressed in Escherichia coli.* Journal of Biological Chemistry, 2000. **275**(37): p. 28715-28721.

126. Peterhans, E., *Reactive oxygen species and nitric oxide in viral diseases.* Biol Trace Elem Res, 1997. **56**(1): p. 107-16.

127. Norberg, J.A., E.S.J., *Reactive oxygen species, antioxidants, and the mammalian thioredoxin system.* Free Radical Biology and Medicine, 2001. **31**(11): p. 1287-1312.

128. Yla-Herttuala, S., *Oxidized LDL and Atherogenesis.* Ann. N. Y. Acad. Sci., 1999(874): p. 134-137.

129. Stadtman, E.R., Levine, R. L., *Protein Oxidation.* Ann. N. Y. Acad. Sci., 2000. **899**: p. 191-208.

130. Marnett, L.J., *Oxyradicals and DNA Damage.* Carcinogenesis, 2000. **21**: p. 361-370.

131. Finkel, T., Holbrook, N.J., *Oxidants, oxidative stress and the biology of ageing.* Nature, 2000. **408**: p. 239-247.

132. Mulhern, S.A., Taylor, G.L., Magruder, L.E., Vessey, A.R., *Deficient levels of dietary selenium suppress the antibody response in first and second generation mice.* Nutr. Res., 1985. **5**: p. 201-210.

133. Marsh, J.A., Combs, G.F., Jr., Whitacre, M.E., Dietert, R.R., *Effect of selenium and vitamin E dietary deficiencies on chick lymphoid organ development.* Proc. Soc. Exp. Biol. Med., 1986. **182**: p. 425-436.

134. Combs, G.F., Gray, W.P., *Chemopreventive Agents: Selenium.* Pharmacol. Ther., 1998. **79**(3): p. 179-192.

135. Roy, M., Kiremidjian-Schumacher, L., Wishe, H.I., Cohen, M.W., Stotzky, G., *Effect of selenium on the expression of high-affinity interleukin 2 receptors.* Proc. Soc. Exp. Biol. Med., 1992. **200**: p. 36-43.

136. Roy, M., Kiremidjian-Schumacher, L., Wishe, H.I., Cohen, M.W., Stotzky, G., *Supplementation with selenium and human immune cell functions. I. Effect of lymphocyte proliferation and interleukin 2 receptor expression.* Biol. Trace Elem. Res., 1994. **41**: p. 103-113.

137. Kiremidjianschumacher, L., et al., *Selenium and Immune Cell Functions .1. Effect on Lymphocyte-Proliferation and Production of Interleukin-1 and Interleukin-2.* Proceedings of the Society for Experimental Biology and Medicine, 1990. **193**(2): p. 136-142.

138. Ebert-Dumig, R., et al., *Expression of selenoproteins in monocytcs and macrophages - Implications for the immune system.* Medizinische Klinik, 1999. **94**: p. 29-34.

139. Talcott, P.A., Exon, J.H., Koller, L.D., *Alteration of natural killer cell-mediated cytotoxicity in rats treated with selenium, diethylnitrosamine and ethylnitrosurea.* Cancer Lett., 1984. **23**(313-322).

140. Meeker, H.C., Exkew, M.L., Scheuchenzuber, W., Scholz, R.W., Zarkower, A., *Antioxidant effects on cell mediated immunity.* J. Leukoc. Biol., 1985. **38**: p. 451-458.

141. Koller, L.D., Exon, J.H., Talcott, P.A, Osborne, C.A., Henningsen, G.M., *Immune responses in rats supplemented with selenium.* Clin. Exp. Immunol., 1986. **63**: p. 570-576.

142. Petrie, H.T., Klassen, L.W., Klassen, P.S., O'Dell, J.R., Kay, H.D., *Selenium and the immune response. 2. Enhancement of murine cytotoxic T-lymphocyte and natural killer cell cytotoxicity in vivo.* J. Leukoc. Biol., 1989. **45**: p. 215-220.

143. Kiremidjian-Schumacher, L., Roy, M., Wishe, H.I., Cohen, M.W., Stotzky, G., *Supplementation of selenium and human immune cell fulctions. II. Effect on cytotoxic lymphocytes and natural killer cells.* Biol. Trace Elem. Res., 1994. **46**: p. 1831.

144. Kiremidjian-Schumacher, L., Roy, M., Wishe, H.I., Cohen, M.W., Stotzky, G., *Supplementation with selenium augments the functions of natural killer and lymphokine-activated killer cells.* Biol. Trace Elem. Res., 1996. **52**: p. 227-239.

145. Bai, J., S. Wu, K. Ge, X. Deng, C. Su, Acta Acad. Bed. Sin., 1980. **2**: p. 29-31.

146. Okabe, M., et al., *Chronic variant of myocarditis associated with hepatitis C virus infection.* Circulation, 1997. **96**(1): p. 22-24.

147. Matsumori, A., et al., *Hepatitis C virus from the hearts of patients with myocarditis and cardiomyopathy.* Laboratory Investigation, 2000. **80**(7): p. 1137-1142.

148. Beck, M.A., et al., *Increased virulence of a human enterovirus (coxsackievirus B3) in selenium-deficient mice.* J Infect Dis, 1994. **170**(2): p. 351-7.

149. Beck, M.A., et al., *Rapid genomic evolution of a non-virulent coxsackievirus B3 in selenium- deficient mice results in selection of identical virulent isolates [see comments].* Nat Med, 1995. **1**(5): p. 433-6.

150. Schrauzer, G.N., et al., *Effect of simulated American, Bulgarian, and Japanese human diets and of selenium supplementation on the incidence of virally induced mammary tumors in female mice.* Biol Trace Elem Res, 1989. **20**(1-2): p. 169-78.

151. Ramanathan, C.S. and E.W. Taylor, *Computational genomic analysis of hemorrhagic fever viruses. Viral selenoproteins as a potential factor in pathogenesis.* Biol Trace Elem Res, 1997. **56**(1): p. 93-106.

152. Beck, K.W., et al., *[Trace element concentrations in HIV infected patients].* Onkologie, 1989. **12 Suppl 3**: p. 43-G.

153. Dworkin, B.M., et al., *Abnormalities of blood selenium and glutathione peroxidase activity in patients with acquired immunodeficiency syndrome and aids-related complex.* Biol Trace Elem Res, 1988. **15**: p. 167-77.

154. Beck, K.W., et al., *Serum trace element levels in HIV-infected subjects.* Biol Trace Elem Res, 1990. **25**(2): p. 89-96.

155. Baum, M.K., et al., *High risk of HIV-related mortality is associated with selenium deficiency.* J Acquir Immune Defic Syndr Hum Retrovirol, 1997. **15**(5): p. 370-4.

156. Schrauzer, G.N. and J. Sacher, *Selenium in the maintenance and therapy of HIV-infected patients [published erratum appears in Chem Biol Interact 1995 Feb;94(2):167].* Chem Biol Interact, 1994. **91**(2-3): p. 199-205.

157. Makropoulos, V., T. Bruning, and K. Schulze-Osthoff, *Selenium-mediated inhibition of transcription factor NF-kappa B and HIV- 1 LTR promoter activity.* Arch Toxicol, 1996. **70**(5): p. 277-83.

158. Hori, K., et al., *Selenium supplementation suppresses tumor necrosis factor alpha-induced human immunodeficiency virus type 1 replication in vitro.* Aids Research and Human Retroviruses, 1997. **13**(15): p. 1325-1332.

159. Sprietsma, J.E., *Cysteine, glutathione (GSH) and zinc and copper ions together are effective, natural, intracellular inhibitors of (AIDS) viruses.* Medical Hypotheses, 1999. **52**(6): p. 529-538.

160. Yu, S.Y., Y.J. Zhu, and W.G. Li, *Protective role of selenium against hepatitis B virus and primary liver cancer in Qidong.* Biol Trace Elem Res, 1997. **56**(1): p. 117-24.

161. Hou, J.C., *Inhibitory effect of selenite and other antioxidants on complement-mediated tissue injury in patients with epidemic hemorrhagic fever.* Biol. Trace Elem. Res., 1997. **56**: p. 125-130.

162. Beck, M.A., et al., *Selenium deficiency increases the pathology of an influenza virus infection.* Faseb Journal, 2001. **15**(6): p. U32-U52.

163. Nelson, H.K., et al., *Host nutritional selenium status as a driving force for influenza virus mutations.* Faseb Journal, 2001. **15**(8): p. U488-U499.

164. Vanvleet, J.F., V.J. Ferrans, and G.R. Ruth, *Ultrastructural Alterations in Nutritional Cardiomyopathy of Selenium-Vitamin-E Deficient Swine .1. Fiber Lesions.* Laboratory Investigation, 1977. **37**(2): p. 188-200.

165. Goldman, I.S. and N.E. Kantrowitz, *Cardiomyopathy Associated with Selenium Deficiency.* New England Journal of Medicine, 1981. **305**(12): p. 701-701.

166. Collipp, P.J. and S.Y. Chen, *Cardiomyopathy and Selenium Deficiency in a 2-Year-Old Girl.* New England Journal of Medicine, 1981. **304**(21): p. 1304-1305.

167. Dimand, R.J., et al., *Selenium Deficient Cardiomyopathy Responding to Oral Selenium in a Patient with Wilms Tumor.* Clinical Research, 1984. **32**(1): p. A93-A93.

168. Nelson, H.K., et al., *Selenium deficiency enhances the pathogenicity of influenza virus infection.* Faseb Journal, 2000. **14**(4): p. A536-A536.

169. Nelson, H.K., et al., *Host selenium status as a driving force for influenza virus mutations.* Faseb Journal, 2001. **15**(5): p. A966-A966.

170. Mollison, L.C., et al., *Hypothyroidism Due to Destruction of the Thyroid by Kaposis-Sarcoma.* Reviews of Infectious Diseases, 1991. **13**(5): p. 826-827.

171. Derry, D.M., *Thyroid hormone therapy in patients infected with human immunodeficiency virus: A clinical approach to treatment.* Medical Hypotheses, 1996. **47**(3): p. 227-233.

172. Koutkia, P., E. Mylonakis, and R.M. Levin, *Human immunodeficiency virus infection and the thyroid.* Thyroid, 2002. **12**(7): p. 577-582.

173. Arthur, J.R., F. Nicol, and G.J. Beckett, *The Role of Selenium in Thyroid-Hormone Metabolism and Effects of Selenium Deficiency on Thyroid-Hormone and Iodine-Metabolism.* Biological Trace Element Research, 1992. **34**(3): p. 321-325.

174. Chanoine, J.P., et al., *Effects of Selenium Deficiency on Thyroid-Hormone Economy in Rats.* Endocrinology, 1992. **131**(4): p. 1787-1792.

175. Chanoine, J.P., et al., *Selenium Deficiency and Type-Ii 5'-Deiodinase Regulation in the Euthyroid and Hypothyroid Rat - Evidence of a Direct Effect of Thyroxine.* Endocrinology, 1992. **131**(1): p. 479-484.

176. Meydani, M., S.N. Meydani, and J.B. Blumberg, *Modulation by Dietary Vitamin-E and Selenium of Clotting Whole-Blood Thromboxane-A2 and Aortic Prostacyclin Synthesis in Rats.* Journal of Nutritional Biochemistry, 1993. **4**(6): p. 322-326.

177. Herz, W.C. and G.F. Combs, *Excess Selenium Increases Ca++-Induced Clotting Times in Chicks and Rats.* Faseb Journal, 1991. **5**(4): p. A714-A714.

178. Low, S.C., J.W. Harney, and M.J. Berry, *Clotting and Functional-Characterization of Human Selenophosphate Synthetase, an Essential Component of Selenoprotein Synthesis.* Journal of Biological Chemistry, 1995. **270**(37): p. 21659-21664.

179. Coutellier, A., et al., *[Oxidative stress in 29 HIV seropositive patients. Two-year results of a double-blind trial of diethyldithiocarbamate versus placebo].* Presse Med, 1992. **21**(38): p. 1809-12.

180. Sandstrom, P.A., et al., *Antioxidant defenses influence HIV-1 replication and associated cytopathic effects.* Free Radic Biol Med, 1998. **24**(9): p. 1485-91.

181. Taylor, E.W., et al., *Nutrition, HIV, and drug abuse: The molecular basis of a unique role for selenium.* Journal of Acquired Immune Deficiency Syndromes, 2000. **25**: p. S53-S61.

182. Mondal, D., J. Alam, and O. Prakash, *Nf-Kappa-B Site-Mediated Negative Regulation of the Hiv-1 Promoter by Ccaat Enhancer-Binding Proteins in Brain-Derived Cells.* Journal of Molecular Neuroscience, 1994. **5**(4): p. 241-258.

183. Sappey, C., et al., *Stimulation of Glutathione-Peroxidase Activity Decreases Hiv Type-1 Activation after Oxidative Stress.* Aids Research and Human Retroviruses, 1994. **10**(11): p. 1451-1461.

184. Israel, N., Gougerot-Pocidalo, MA, Aillet, F, et al., *Redox status of cells influences constitutive or induced NF-kB translocation and HIV long terminal repeat activity in human T and monocytic cell lines.* J. Immunol, 1992. **149**: p. 3386-3393.

185. Israel, N., Gougerot-Pocidalo, MA, *Oxidative stress in human immunodeficiency virus infection.* Cell Mol Life Sci, 1997. **53**: p. 864-870.

186. Allard, J., Aghdassi, E, Chau J, et al., *Oxidative stress and plasma antioxidant micronutrients in humans with HIV infection.* Am J Clin Nutr, 1998. **67**: p. 143-147.

187. Brigelius-Flohe R, F.B., Mourer S, et al., *Interleukin-1 induced nuclear factor kappa B activation is inhibited by overexpression of phospholipid hydroperoxide glutathione peroxidase in a human endothelial cell line.* Biochem J, 1997. **328**: p. 199-203.

188. Mitomo K, N.K., Fujimoto K, et al., *Two different cellular redox systems regulate the DNA-binding activity of the p50 subunit of NF-kappa B in vitro.* Gene, 1994. **145**: p. 197-203.

189. Gorlatov, S., Stadtman TC,, *Human thioredoxin reductase from HeLa cells: selective alkylation of selenocysteine in the protein inhibits enzyme activity and reduction with NADPH influences affinity to heparin.* Proc. Natl. Acad. Sci., 1998. **95**: p. 8520-8525.

190. Beck, M.A., *Increased virulence of coxsackievirus B3 in mice due to vitamin E or selenium deficiency.* Journal of Nutrition, 1997. **127**: p. S966-S970.

191. Belec, L., et al., *The Hiv Wasting Syndrome.* Muscle & Nerve, 1992. **15**(7): p. 856-857.

192. Keusch, G.T., et al., *Persistent Diarrhea Associated with Aids.* Acta Paediatrica, 1992. **81**: p. 45-48.

193. Coodley, G.O., et al., *Micronutrient Concentrations in the Hiv Wasting Syndrome.* Aids, 1993. **7**(12): p. 1595-1600.

194. Taylor, E.W., et al., *HIV-1 encodes a sequence overlapping env gp41 with highly significant similarity to selenium-dependent glutathione peroxidases [letter].* J Acquir Immune Defic Syndr Hum Retrovirol, 1997. **15**(5): p. 393-4.

195. Mantero-Atienza, E., et al., *Selenium status of HIV-1 infected individuals [letter; comment].* JPEN J Parenter Enteral Nutr, 1991. **15**(6): p. 693-4.

196. Dworkin, B.M., *Selenium deficiency in HIV infection and the acquired immunodeficiency syndrome (AIDS).* Chem Biol Interact, 1994. **91**(2-3): p. 181-6.

197. Graziadei, I., et al., *Modulation of iron metabolism in monocytic THP-1 cells and cultured human monocytes by the acute-phase protein alpha 1-antitrypsin.* Experimental Hematology, 1998. **26**(11): p. 1053-1060.

198. Maes, M., et al., *Lower serum zinc in major depression in relation to changes in serum acute phase proteins.* Journal of Affective Disorders, 1999. **56**(2-3): p. 189-194.

199. Uzunkoy, A., et al., *Trace element changes: a stress response to laparotomy and laparoscopy.* Trace Elements and Electrolytes, 2000. **17**(4): p. 169-172.

200. Andonova, M., I. Borissov, and L. Sotirov, *Changes in some factors of the innate immunity and serum zinc and iron concentrations in pigs following intravenous administration of Escherichia coli lipopolysaccharide.* Onderstepoort Journal of Veterinary Research, 2001. **68**(2): p. 91-99.

201. Davidson, W.B. and D.G. Kennedy, *Synthesis of [Se-75] Selenoproteins Is Greater in Selenium-Deficient Sheep.* Journal of Nutrition, 1993. **123**(4): p. 689-694.

202. Behne, D., et al., *Newly found selenium-containing proteins in the tissues of the rat.* Biological Trace Element Research, 1996. **55**(1-2): p. 99-110.

203. Behne, D. and A. Kyriakopoulos, *Mammalian selenium-containing proteins.* Annual Review of Nutrition, 2001. **21**: p. 453-473.

204. Taylor, E.W., et al., *A basis for new approaches to the chemotherapy of AIDS: novel genes in HIV-1 potentially encode selenoproteins expressed by ribosomal frameshifting and termination suppression.* J Med Chem, 1994. **37**(17): p. 2637-54.

205. Chen, C.Y., J.Y. Zhou, and H.B. Xu, *Advances in the studies of relationship between selenium and human immunodeficiency virus.* Progress in Biochemistry and Biophysics, 1997. **24**(4): p. 327-330.

206. Zhang, W.Q., A.G. Cox, and E.W. Taylor, *Hepatitis C virus encodes a selenium-dependent glutathione peroxidase gene - Implications for oxidative stress as a risk factor in progression to hepatocellular carcinoma.* Medizinische Klinik, 1999. **94**: p. 2-6.

207. Shisler, J.L., et al., *Ultraviolet-induced cell death blocked by a selenoprotein from a human dermatotropic poxvirus.* Science, 1998. **279**(5347): p. 102-105.

208. Senkevich, T.G., et al., *Genome sequence of a human tumorigenic poxvirus: Prediction of specific host response-evasion genes.* Science, 1996. **273**(5276): p. 813-816.

209. Damon, I., P.M. Murphy, and B. Moss, *Broad spectrum chemokine antagonistic activity of a human poxvirus chemokine homolog.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(11): p. 6403-6407.

210. Murphy, P.M., *Viral exploitation and subversion of the immune system through chemokine mimicry.* Nature Immunology, 2001. **2**(2): p. 116-122.

211. Taylor, E.W., *Selenium and viral diseases: facts and hypotheses.* Journal of Orthomolecular Medicine, 1997(fourth quarter).

212. Zhao, L.J., et al., *Molecular modeling and in vitro activity of an HIV-1-encoded glutathione peroxidase.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(12): p. 6356-6361.

213. Taylor, E.W., R.G. Nadimpalli, and C.S. Ramanathan, *Genomic structures of viral agents in relation to the biosynthesis of selenoproteins.* Biol Trace Elem Res, 1997. **56**(1): p. 63-91.

214. Gladyshev, V.N., et al., *Levels of major selenoproteins in T cells decrease during HIV infection and low molecular mass selenium compounds increase.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(3): p. 835-839.

215. Ramanathan, C.S. and E.W. Taylor, *Computational genomic analysis of hemorrhagic fever viruses - Viral selenoproteins as a potential factor in pathogenesis.* Biological Trace Element Research, 1997. **56**(1): p. 93-106.