

Title

Review Presentation: “Effect of temperature and sorbitol in improving the solubility of carboxylesterases protein CpCE-1 from *Cydia pomonella* and biochemical characterization”

Presenter

Faiyez Ahmed, Bioinformatics Program, The University of Texas at El Paso.

Abstract:

This poster is a review of the paper "Effect of temperature and sorbitol in improving the solubility of carboxylesterases protein CpCE-1 from *Cydia pomonella* and biochemical characterization" by Yang and Zhang (2013) published in the journal *Appl. Microbiol. Biotechnol.* The abstract, as originally published by the authors, is as follows:

“Carboxylesterases (CEs) are enzymes responsible for the detoxification of insecticides in insects. In the *Cydia pomonella*, CE s are involved in synthetic pyrethroid, neonicotinoid, carbamate, and organophosphate detoxification. However, functional over expression of CE s proteins in *Escherichia coli* systems often results in insoluble proteins. In this study, we expressed the fusion protein CpCE-1 in *E. coli* BL21 (DE3). This recombinant protein was over expressed as inclusion bodies at 37 °C whereas it produced a higher percentage of soluble protein at lower growth temperatures. Production of soluble proteins and enzyme activity increased in the presence of sorbitol in the growth medium. Production of soluble proteins and enzyme activity increased in the presence of sorbitol in the growth medium. The fusion protein was purified from the lysate supernatant using a Ni²⁺-NTA agarose gel column. The enzyme exhibited a higher affinity and substrate specificity for α -naphthyl acetate (α -NA), with k_{cat}/K_m of 100 s⁻¹ μ M⁻¹ for α -NA, and the value is 29.78 s⁻¹ μ M⁻¹ for β -naphthyl acetate. The V_{max} and K_m were also determined to be 12.9 μ mol/min/mg protein and 13.4 μ M using substrate α -NA. The optimum pH was 7.0 and temperature was 25 °C. An enzyme inhibition assay shows that PMSF and DEPC strongly inhibit the enzyme activity, while the metal ions Cu²⁺ and Mg²⁺ significantly activated the activity. More importantly, cypermethrin, methomyl, and acephate were found to suppress enzyme activity. The data demonstrated here provide information for heterologous expression of soluble protein and further study on insecticide metabolism in *C. pomonella* in vitro. This is the first report of the characterization of CE s protein from *C. pomonella*.”

Title

Dissecting the gene expression pattern in gastric cancer

Authors

Priyanka Bodepudi, Bioinformatics Program, The University of Texas at El Paso.

Abstract

Gastric cancer, as the second leading cause of cancer-related death worldwide, remains a serious health problem even though its overall incidence is decreasing in many countries. Gastric carcinogenesis is a multi-step process in which environmental and genetic factors interact. Among the genetic changes observed in cancerous cells, genomic DNA amplification is a well-known alteration that is involved in gastric cancer. Amplification is often associated with increased expression levels of the oncogenes contained in the amplified loci. It is possible that there exist genes whose amplification in gastric cancer has not been revealed to date. This study is a brief overview of the analysis of gene expression patterns to diagnose gastric cancer. Micro arrays have been used to examine the expression patterns of some 25,000 genes in tissues from gastric cancer patients. Computer cluster analysis of the patterns led to the identification of marker genes that can correctly identify the cancer which would eventually develop metastasis. The specific objective of this research is to find if there are any marker genes that help in identifying the cancer cells in the patients who are more likely to experience aggressive progression of cancer to test the hypothesis that there are marker genes that identify the cancer cells which would eventually develop metastases.

Title

Potential enhancer elements for the mushroom body neurons critical for learning and memory in *Drosophila*.

Authors

Carolina Guerra,¹ Ming Ying Leung,^{1,2} and Kyung-An Han³

¹Bioinformatics Program, ²Department of Mathematical Sciences, ³Department of Biological Sciences, The University of Texas at El Paso.

Abstract

In this study, we are interested to use bioinformatics-based tools and algorithms to identify enhancer elements that allow selective expression of genes in a brain structure important for learning and memory. This information can be used to identify additional genes that are expressed in the neurons under study facilitating the study of the mechanism underlying learning and memory. For the study, we used *Drosophila* as a model system. *Drosophila* is a powerful model organism due to its well characterized genetics, fully sequenced genome, and sophisticated nervous system comparable to that of humans. We are interested in aligning the sequence of the genomic region that drives expression of the reporter Green Fluorescent Protein (GFP) in the mushroom body gamma lobe neurons, a principal brain structure for learning and memory. The general idea behind this study is to find sequences that are common in all DNA segments that drive GFP expressions in γ neurons. We are more interested in γ neurons because they mediate the dopamine and octopamine signals for learning and memory. Abnormal dopamine functions are responsible for various diseases like ADHD, autism, schizophrenia, Parkinson's disease, and drug abuse/addiction. We are developing a bioinformatics tool to facilitate the study of the pathogenesis mechanisms and intervention strategy.

Title

Isolation, amplification, incorporation of ASPM (Abnormal Spindle like Microcephaly Associated - Exon 18) into suitable vector pUC 18.

Authors

Akshita Gurram^{1,2}, Sachin Anumula², and Nitin Kumar Reddy²

¹Bioinformatics Program, The University of Texas at El Paso.

²Department of Biotechnology, Jawaharlal Nehru Technological University, India.

Abstract

The role of intelligence in bringing about a huge change in the development of human race over other animal races on this planet is remarkably great. A vast morphological change in the increase of human cerebral hemisphere volume from the ancestral races to the present one is quite noticeable. The credit for this is given to a single gene called ASPM (Abnormal Spindle like Microcephaly associated). The present work is mainly concentrated on the amplification of exon 18 of the ASPM gene followed by the incorporation of it into a suitable vector for further studies. Out of the 28 exons present in ASPM gene, one exon differing from other animals and supposed to contribute towards variation in human development is thought to be exon 18. Student blood DNA samples were used for the amplification of the ASPM Exon 18. Thus obtained PCR products are subjected to purification for further modification with suitable nitrogen for cloning into pUC 18 vector. The multiple cloning sites present in pUC 18 facilitate the incorporation of exon 18 into this vector. The selection of the incorporation site was done depending upon the future availability of that gene for further research. The Sma I restriction cleaving site was effectively selected for this purpose. The alteration in the gene sequence in the pUC 18 vector is to be analyzed after the isolation of transformed bacterial *E. coli* colonies.

Title

Review Presentation: “West Nile alternative open reading frame (N-NS4B/WARF4) is produced in infected West Nile Virus (WNV) cells and induces humoral response in WNV infected individual.”

Presenter

Amal Hassan, Bioinformatics Program, The University of Texas at El Paso.

Abstract

This poster is a review of the paper "West Nile alternative open reading frame (N-NS4B/WARF4) is produced in infected West Nile Virus (WNV) cells and induces humoral response in WNV infected individual." by Faggioni *et al.* (2012) published in the *Virology Journal*. The abstract, as originally published by the authors, is as follows:

“Background:

West Nile Virus (WNV) is a flavivirus that requires an efficient humoral and cellular host response for the control of neuroinvasive infection. We previously reported the existence of six alternative open reading frame proteins in WNV genome, one of which entitled WARF4 is exclusively restricted to the lineage I of the virus. WARF4 is able to elicit antibodies in WNV infected horses; however, there was no direct experimental proof of the existence of this novel protein. The purpose of this study was to demonstrate the *in vitro* production of WARF4 protein following WNV infection of cultured VERO cells and its immunity in WNV infected individuals.

Results:

We produced a monoclonal antibody against WARF4 protein (MAb 3A12) which detected the novel protein in WNV lineage I-infected, cultured VERO cells while it did not react with WNV lineage II infected cells. MAb 3A12 specificity to WARF4 protein was confirmed by its reactivity to only one peptide among four analyzed that cover the full WARF4 amino acids sequence. In addition, WARF4 protein was expressed in the late phase of WNV lineage I infection. Western blotting and bioinformatics analyses strongly suggest that the protein could be translated by programmed –1 ribosomal frameshifting process. Since WARF4 is embedded in the NS4B gene, we rename this novel protein N-NS4B/WARF4. Furthermore, serological analysis shows that N-NS4B/WARF4 is able to elicit antibodies in WNV infected individuals.

Conclusions:

N-NS4B/WARF4 is the second Alternative Reading Frame (ARF) protein that has been demonstrated to be produced following WNV infection and might represent a novel tool for a better characterization of immune response in WNV infected individuals. Further serological as well as functional studies are required to characterize the function of the N-NS4B/WARF4 protein. Since the virus might actually make an extensive use of ARFs, it appears important to investigate the novel six ARF putative proteins of WNV. ”

Title

A genomics approach to arsenic detoxification pathways in rotifera – a model aquatic organism

Authors

Kayla I. Hinson, Edgar Jauregui, and Elizabeth J. Walsh

Bioinformatics Program and Department of Biological Sciences, The University of Texas at El Paso.

Abstract

Arsenic, though found naturally, is prevalent in many water systems due to various anthropogenic activities. At high concentrations it has been shown to not only be toxic to aquatic organisms residing in these waters, but also to humans through contaminated drinking water. In response to this toxicity, organisms have developed detoxification mechanisms and similarities among the pathways are expected. Here we specifically investigate pathways involved in arsenic resistance. We used bioinformatics approaches to discover and identify genes involved in arsenic metabolism in the euryhaline model rotifer, *Brachionus plicatilis*. Using known pathways identified in other species (e.g., *Loa loa*, *Ascaris suum*, *Danio rerio*) we verified these detoxification enzymes in the transcriptome of the bdelliod rotifer *Adenita ricciae*, using Trinotate: Trinity Transcriptome Assembly software. This analysis revealed 15 potential arsenic-associated enzymes including ATPase ANA1, an *arsA* homolog in bacteria. Further investigation of ATPase ANA1 through multiple sequence alignment (MAFFT), identification of potential motifs (MEME), and phylogenetic reconstruction (Neighbor-Joining), showed strong conservation among these taxa, including rotifers. Future work will involve referencing confirmed transcripts back to *B. plicatilis* by performing BLAST searches against the EST libraries for this species. These results will enhance our understanding of the pathways lower trophic level organisms use to metabolically remove arsenic found in waterways.

Title

Isolation, amplification, purification, and cloning of hepatitis B-virus surface gene

Authors

Harish Kalyani,^{1,2} Naresh,¹ and Madhavi Chandra¹

¹Centre for Liver Research and Diagnostics, Hyderabad, India,

²Bioinformatics Program, The University of Texas at El Paso.

Abstract

“Hepatitis” implies inflammation of the liver characterized by the presence of inflammatory cells in the tissues of organ. It may occur with limited or no symptoms (sub clinically), but often leads to jaundice. It is acute when it lasts less than six months and chronic when it persists longer. The hepatitis B virus (HBV) is a small DNA virus with unusual features similar to retroviruses. (Ganem *et al.* 2001). The infectious HBV virion (Dane particle) has a spherical, double-shelled structure 42 nm in diameter, consisting of a lipid envelope containing HBsAg that surrounds an inner nucleocapsid composed of hepatitis B core antigen (HBcAg) complexed with virally encoded polymerase and the viral DNA genome (Liang 2009). The HBV nucleocapsid or surface antigen (HBsAg) is extremely immunogenic during infection and after immunization. HBsAg has a unique three dimensional folding that interacts with immunoglobulin’s outside the classical antibody binding site, has the CD4 T cell epitopes and the encapsulated nucleic acid that can efficiently interact and activate presenting cells, especially naive B cells. HBsAg also perform pivotal function of mediating release of sub viral particles, regulation of super coiled DNA amplification and transcriptional transactivation. The hepatitis B surface gene was cloned by using top 10 culture cells. The cloned gene can be used to study the mutational analysis of the disease, diagnostic purposes and to know the expression of the gene.

Title:

Review Presentation: “The histone deacetylase SIRT6 is a tumor suppressor that controls cancer metabolism.”

Presenter

Sivakumar Kareti, Bioinformatics Program, The University of Texas at El Paso.

Abstract

This poster is a review of the paper "The histone deacetylase SIRT6 is a tumor suppressor that controls cancer metabolism" by Sebastian *et al.* (2012) published in the journal *Cell*. The abstract, as originally published by the authors, is as follows:

“Reprogramming of cellular metabolism is a key event during tumorigenesis. Despite being known for decades (Warburg effect), the molecular mechanisms regulating this switch remained unexplored. Here, we identify SIRT6 as a tumor suppressor that regulates aerobic glycolysis in cancer cells. Importantly, loss of SIRT6 leads to tumor formation without activation of known oncogenes, whereas transformed SIRT6-deficient cells display increased glycolysis and tumor growth, suggesting that SIRT6 plays a role in both establishment and maintenance of cancer. By using a conditional SIRT6 allele, we show that SIRT6 deletion *in vivo* increases the number, size, and aggressiveness of tumors. SIRT6 also functions as a regulator of ribosome metabolism by corepressing MYC transcriptional activity. Lastly, Sirt6 is selectively downregulated in several human cancers, and expression levels of SIRT6 predict prognosis and tumor-free survival rates, highlighting SIRT6 as a critical modulator of cancer metabolism. Our studies reveal SIRT6 to be a potent tumor suppressor acting to suppress cancer metabolism.”

Title

An investigation of GPCRs in the brain of *Rhipicephalus microplus*

Authors

April Lewis^{1*}, Alexandria Ogrey^{1*}, Ming-Ying Leung^{1,2}, and Felix Guerrero³

¹Bioinformatics program, ²Department of Mathematical Sciences, University of Texas at El Paso,

³USDA-ARS Knippling-Bushland US Livestock Insects Research Laboratory, Kerrville, Texas.

*These authors contributed equally to this study.

Abstract

Rhipicephalus microplus, commonly known as the cattle tick, is an economically important agricultural pest that parasitizes a variety of livestock species. Heavy tick burdens on animals can decrease agricultural production for both crops and animal products. This research focuses on identifying G-Protein coupled receptors (GPCRs) that could potentially be vaccine targets in the development of new measures to protect against *R. microplus*. The main function of GPCRs is to sense molecules outside the cell, activate signal transduction pathways inside the cell, and ultimately, to create cellular responses. Vaccines targeting GPCRs can be developed to stimulate the cattle's immune system to recognize the parasite as a foreign agent, kill the foreign agent, and respond to it quickly in future encounters. Utilizing DNA sequence data from the tick synganglia, a series of Python scripts were written to help identify potential GPCRs. As GPCR proteins are known to contain seven transmembrane helices, putative coding regions were run through the Transmembrane Hidden Markov Model (TMHMM), an online server that predicts the likelihood of transmembrane helices in the protein. With the information obtained from TMHMM, a script was written to predict whether or not there were GPCRs based on length and number of predicted helices. Since the assembled DNA was not continuous, the possibility of partial length GPCRs were considered. Preliminary results show approximately 118 potential GPCRs were collected. Further analysis is required to determine if these proteins are true GPCRs.

Title

Review Presentation: “The utility and limitations of current web-available algorithms to predict peptides recognized by CD4 T Cells in response to pathogen infection.”

Presenter

Sergio L. Muñoz, Bioinformatics Program, The University of Texas at El Paso.

Abstract

This poster is a review of the paper "The utility and limitations of current web-available algorithms to predict peptides recognized by CD4 T Cells in response to pathogen infection" by Chaves *et al.* (2012) published in the *Journal of Immunology*. The abstract, as originally published by the authors, is as follows:

“The ability to track CD4 T cells elicited in response to pathogen infection or vaccination is critical because of the role these cells play in protective immunity. Coupled with advances in genome sequencing of pathogenic organisms, there is considerable appeal for implementation of computer-based algorithms to predict peptides that bind to the class II molecules, forming the complex recognized by CD4 T cells. Despite recent progress in this area, there is a paucity of data regarding the success of these algorithms in identifying actual pathogen-derived epitopes. In this study, we sought to rigorously evaluate the performance of multiple Web- available algorithms by comparing their predictions with our results—obtained by purely empirical methods for epitope discovery in influenza that used overlapping peptides and cytokine ELISPOTs—for three independent class II molecules. We analyzed the data in different ways, trying to anticipate how an investigator might use these computational tools for epitope discovery. We come to the conclusion that currently available algorithms can indeed facilitate epitope discovery, but all shared a high degree of false- positive and false-negative predictions. Therefore, efficiencies were low. We also found dramatic disparities among algorithms and between predicted IC50 values and true dissociation rates of peptide–MHC class II complexes. We suggest that improved success of predictive algorithms will depend less on changes in computational methods or increased data sets and more on changes in parameters used to “train” the algorithms that factor in elements of T cell repertoire and peptide acquisition by class II molecules.”

Title

Waste heat recovery with low temperature combustion

Author

Dylan Ott, Department of Mechanical Engineering, The University of Texas at El Paso.

Abstract

Low temperature combustion (LTC) engines are investigated for both their lower exhaust emissions of nitrogen oxides and fixed carbon soot and the potential of being more efficient than conventional combustion engines. Waste heat recovery (WHR) systems are also investigated for their potential to turn presently-wasted thermal energy into useful energy (or work energy). A heat transfer model of an LTC engine exhaust was used to estimate exhaust pipe surface temperatures, which are used to assess the potential of two different WHR systems to recover thermal energy: an Organic Rankine Cycle concept and a Thermo Electric Generator concept. It is determined that both systems could work with an LTC engine and convert about 15% of the presently wasted thermal energy into useful work.

Title

Investigating micro RNA expression profiles and corresponding pathways in breast cancer cells

Author

Vignesh Ravichandran, Bioinformatics Program, The University of Texas at El Paso.

Abstract

Breast cancer is one of the leading causes of death in women in North America. The etiology of breast cancer involves a combination of genetic and environmental factors. In previous studies marker genes for various types of cancers by analyzing micro array data have been successfully identified. Micro RNA gene expression profiles have been particularly instrumental in diagnosing uncontrolled cell proliferation. Lu *et al.* (2005) found that total miRNA levels are significantly lower in cancer cells relative to normal cells. Vrba *et al.* (2013) identified a subgroup of miRNAs whose promoters are hypermethylated in breast cancer cells. These findings motivated us to critically investigate if there is a relation between promoter hypermethylation in miRNA genes and miRNA expression in breast cancer cells. This study will provide an insight into the reduced miRNA expression observed in breast cancer cell lines based on the promoter methylation. This investigation is based on the hypothesis that promoter hypermethylation in miRNA genes are the key epigenetic factor dictating the reduced miRNA expression levels in cancer cell lines and this investigation is about to be carried out on the existing data. We predict that a negative correlation exists between the extent of promoter methylation and the amount of miRNA expressed whereby high levels of the former are associated with the low levels of the later. The project aims at identifying differentially expressed miRNAs in breast cancer cells and identifying the cellular pathway they are involved to understand the regulation of these miRNAs in breast cancer.

Title

Protein-Protein Interaction of HSP70 and HSP60 with Other Cancer-related Proteins

Authors

Janeth Rodriguez¹, Ming-Ying Leung^{2,3}, Jianying Zhang^{1,3}

¹Bioinformatics Program, ²Department of Mathematical Sciences, ³Department of Biological Sciences, The University of Texas at El Paso.

Abstract

The purpose of this project is to use Protein-Protein Interaction (PPI) Networks to help elucidate the roles of HSP70 and HSP60 in cancer. Since different proteins often interact with one another to perform a biological function, it is important to find other cancer-related proteins that interact with HSP70 and HSP60. PPI Networks require the use of in-silico methods to find other cancer-related proteins, called interactors, which interact with the target proteins. Examples include the interaction retrieval, protein-protein docking, and visualization methods used in the software packages like HPRD, DECOMP, GRAMM, Chimera, PISA, and Python Molecule Viewer. Eighty-eight cancer-related proteins that interact with the HSP60 were found in the Human Protein Reference Database, of which 28 were expressed in cancer. Thirty-one interactors were found for HSP70 in the same database, of which 9 were expressed in cancer. These interactors need to be further analyzed in terms of their 3-dimensional structures to see exactly how they interact with the target proteins and whether they can be used as biomarkers for cancer detection. The computational analysis results can be used as a guide to help design wet-lab experiments to study HSP70 and HSP60 and is expected to save valuable time and lab resources. By using the same methods a new target cancer-related protein named p62 is going to be elucidated to find its role in cancer.

Title

Elucidating multidrug resistance in *Mycobacterium tuberculosis* using molecular modeling

Authors

Shivangi Sharma, Bioinformatics Program, The University of Texas at El Paso,
Rituraj Purohit, Vellore Institute of Technology, India.

Abstract

The problem of tuberculosis (TB) drug resistance and the continuing rise in the disease incidence has prompted the research on new drug development as well as on increasing the understanding of the mechanisms of drug resistance. Molecular docking and molecular dynamics (MD) simulations were performed to study the binding activity of isoniazid (INH; a first line anti-TB drug) onto the active site of enoyl-acyl carrier protein reductase (InhA) in an effort to increase the understanding of the action and resistance of INH in this bacterium. The results support the theory that the activation of INH to isonicotinic acyl-NADH (INADH) by KatG enzyme is required for its ultimate activity. It is shown that INADH has tremendously high binding affinity towards InhA by forming more hydrogen bonds, hydrophobic, van der Waals, electrostatic and π - π interactions, compared to the parent drug (INH). These studies show that molecular modeling is a reliable method to investigate the binding properties of an acceptor and a receptor at the molecular level as it can reliably reproduce experimental data. The molecular modeling approach is also useful for exploring novel compound classes directed against a given target, especially when the experimental information about the binding mode is not available. This study is to utilize molecular docking and molecular dynamics (MD) simulations to investigate the binding of INH onto the active site of *Mycobacterium tuberculosis* enoyl-acyl carrier protein (ACP) reductase (InhA) mutants and fluoroquinolones onto active sites of gyrA mutants in an attempt to address the mycobacterial resistance against INH and fluoroquinolones.

Title

Review Presentation: "Multi-target QSAR modeling in the analysis and design of HIV-HCV co-inhibitors: an in-silico study."

Presenter

Juan Soto, Bioinformatics Program, The University of Texas at El Paso.

Abstract

This poster is a review of the paper "Multi-target QSAR modelling in the analysis and design of HIV-HCV co-inhibitors: an in-silico study" by Liu et al. (2011) published in the journal *BMC Bioinformatics*. The abstract, as originally published by the authors, is as follows:

“Background

HIV and HCV infections have become the leading global public-health threats. Even more remarkable, HIV-HCV co-infection is rapidly emerging as a major cause of morbidity and mortality throughout the world, due to the common rapid mutation characteristics of the two viruses as well as their similar complex influence to immunology system. Although considerable progresses have been made on the study of the infection of HIV and HCV respectively, few researches have been conducted on the investigation of the molecular mechanism of their co-infection and designing of the multi-target co-inhibitors for the two viruses simultaneously.

Results

In our study, a multi-target Quantitative Structure-Activity Relationship (QSAR) study of the inhibitors for HIV-HCV co-infection were addressed with an in-silico machine learning technique, i.e. multi-task learning, to help to guide the co-inhibitor design. Firstly, an integrated dataset with 3 HIV inhibitor subsets targeted on protease, integrase and reverse transcriptase respectively, together with another 6 subsets of 2 HCV inhibitors targeted on NS3 serine protease and NS5B polymerase respectively were compiled. Secondly, an efficient multi-target QSAR modelling of HIV-HCV co-inhibitors was performed by applying an accelerated gradient method based multi-task learning on the whole 9 datasets. Furthermore, by solving the L_1 -infinity regularized optimization, the Drug-like index features for compound description were ranked according to their joint importance in multi-target QSAR modelling of HIV and HCV. Finally, a drug structure-activity simulation for investigating the relationships between compound structures and binding affinities was presented based on our multiple target analysis, which is then providing several novel clues for the design of multi-target HIV-HCV co-inhibitors with increasing likelihood of successful therapies on HIV, HCV and HIV-HCV co-infection.

Conclusions

The framework presented in our study provided an efficient way to identify and design inhibitors that simultaneously and selectively bind to multiple targets from multiple viruses with high affinity, and will definitely shed new lights on the future work of inhibitor synthesis for multi-target HIV, HCV, and HIV-HCV co-infection treatments.”

Title

Using bioinformatics to determine which proteolytic enzymes cut GPI-anchored proteins into peptides of ideal length for mass spectrometry analysis

Authors

Andrea Wurm,^{1,2*} Sara Yazdi,^{2,3*} Clemente Aguilar,^{1,4,5} Gloria Polanco¹, Ming-Ying Leung,^{1,2,3,4} and Igor Almeida^{1,2}

¹Border Biomedical Research Center, Dept. of Biological Sciences, ²Bioinformatics Program, ³Dept. of Mathematical Sciences, ⁴Computational Science Program, The University of Texas at El Paso.

⁵Currently at the National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville.

*These authors equally contributed to this study.

Abstract

The focus of this project is to determine which enzymes cut proteins into peptide sequences of ideal lengths for mass spectrometry (MS) analysis in order to better identify biomarkers for cancer detection. Glycosylphosphatidylinositol (GPI) is a universal posttranslational modification of proteins. GPI-anchored proteins (GPI-APs) can possibly be used to identify biomarkers for prostate cancer, a leading cause of cancer deaths in men. Studies have shown that GPI-transamidase (TA), the enzyme that attaches the GPI anchor to the protein, is overexpressed in prostate cancer tissue. To prepare GPI-APs for MS, three enzymes –Asp-N endopeptidase, Lys-C lysyl endopeptidase, and trypsin– were chosen to virtually cleave the GPI-APs. Each enzyme follows a specific set of rules on where to cut a protein. We have written a series of Perl scripts that use the rules of each enzyme to virtually cut protein sequences into theoretical peptides to identify which enzyme, or combination thereof, will provide the greatest number of ideal length peptides. The ideal peptide length for detection by MS falls between six and twenty amino acids. We are further exploring the effect that the small probability to miss a cut at each possible site has on the choice of enzymes.