**Titles and Abstracts for Poster Session on 10/27/2012**
**NMSU/UTEP Workshop**

1. **Student Presenter: Clemente Aguilar**
   **Faculty Mentors: Drs. Igor Almeida and Ming-Ying Leung**

Title

Computational Tool for Large-Scale GPIomic Analysis

Authors and affiliations

Aguilar-Bonavides, C.[1,2], Leung, M-Y.,[1,2,3,4] Nakayasu, E.S.,[1], Felipe Gazos-Lopes[1] and Almeida, I.C.[1,3]

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

Abstract

Glycosylphosphatidylinositol (GPI) anchored proteins are involved in many biological processes of a large variety of living beings and are of medical importance. The identification and analysis of the entire collection of free and protein linked GPIs within an organism (i.e., GPIomics) requires highly sensitive instruments. At present, liquid chromatography-tandem mass spectrometry (LC-MS/MS or $MS^n$) is the most efficient laboratory technique for these tasks. As a typical LC-MS/MS experiment produces hundreds of thousands of spectra, the data analysis creates a major bottleneck in high-throughput GPIomic projects. Yet, no computational tool for characterizing the chemical structures of GPI is available to date. We propose a library-search algorithm to identify GPIs by matching fragment peaks in the spectra with molecular masses derived from a collection of theoretical GPI structures constructed based on properties of all currently known GPIs. The algorithm involves matching the mass-to-charge (*m/z*) ratio of the parent ions and fragments obtained from the observed spectra to those of the theoretical structures in the library. A theoretically possible GPI structure is assessed by a scoring scheme that incorporates its fitness values for individual observed spectra as well as its frequency of being considered as a good fit. The algorithm has been tested on a set of experimentally confirmed GPIs for the parasite *Trypanosoma cruzi*. The final list of 4686 predicted GPI candidates contains 76 out of the 78 known structures in the test set. Over 70% of the known structures have fitness values among the top 647. This computational tool is expected to quicken the discovery and characterization of GPI molecules, increasing the number of experimentally confirmed GPI structures. This will, in turn, help us further develop the algorithm to reduce the number of false positives.

2. **Student Presenter: Raphael Alarcon**
**Faculty Mentors: Drs. Igor Almeida and Ming-Ying Leung**

Title

Comparison of proteome analysis software

Authors and affiliations

Alarcon R.[1], Pessini-Rodriguez, S.[2], Arigi, E.[2] Aguilar-Bonavides,C.[2,3], Leung M-Y.[1,2,3,4], and Almeida I.[1,2]

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

Abstract

A mass spectrometer is an analytical instrument in which ions produced by a sample are separated by electric or magnetic fields according to their ratios of charge to mass. In a typical proteomics experiment thousands of MS/MS spectra per hour can be generated. Manual interpretation of the spectrum is not a feasible option and a number of computational and software tools have been developed. The identification of peptides from acquired MS/MS is most often performed using the database search approach. We provide a detailed description of the peptide identification process using this approach and review two common database search tools. Specifically, we will show the difference between Sorcerer 2 and Proteome Discoverer. These packages use SEQUEST (Eng et al.) as the core of their algorithm; we will provide a description of SEQUEST and review the differences in the evaluation of peptide assignment for both scoring schemes. Statistical approaches for validating peptide assignments to MS/MS spectra included the concepts of decoy database searching and false discovery rate (FDR). In this context we will explore the problem of false-discovery identifications that can significantly diminish the value of published data. The aims of this study are to gain a better understanding of the reliability of the software and to show the difference in analysis results, thus providing insights to help scientists select appropriate software tools to use for their investigations.

3. **Student Presenter: Priyanka Bodepudi**
   **Faculty Mentor: Dr. Sundararajan**

Title

Study of heavy metal removal through depuration from green mussel (*Perna viridis*)

Authors and affiliations

Priyanka Bodepudi*, Sindhu Adapa, and Dr. Sundararajan

Department of Marine Biotechnology, National Institute of Ocean Technology, India

*Currently with the Bioinformatics Program, The University of Texas at El Paso.

Abstract

Environmental pollution is currently a major concern due to unorganized urban development and increased industrialization. The sea water faces serious pollution problems which affects the marine aquatic animals. Large numbers of mollusks especially Green mussel Perna virids were found to accumulate high concentrations of metals in their soft tissues and can be lethal when consumed by human beings. The present study aimed to test the heavy metal removal technology from Green Mussel, Perna viridis by depuration process. This project work reports the depuration of Cd, Cr, Cu, Zn, Ni and Pb in green mussel. This analysis indicates the level of heavy metals in green mussels and the results are compared with the standard permissible limit values. Depuration involves the transfer of the contaminated green mussel into plastic trays of clean water of different salinity concentrations (0 ppt, 10 ppt, 20 ppt, and 25 ppt) which leads to expulsion of heavy metals by osmosis. More amounts of accumulated heavy metals are removed at 0ppt concentration. It leads to death of the mussel, but it can be preserved in the ice and can be exported.

**4. Student Presenter: Kyle Boone**
**Faculty Mentor: Dr. Jiannong Xu**

Title

Microbial metabolic pathway analysis in the gut of *Anopheles gambiae* during blood digestion

Authors

Kyle Boone[1], Michael Best[2], Phanidhar Kukutla[2], and Jiannong Xu[2]

[1]Bioinformatics Program, The University of Texas at El Paso.
[2]Biology Department, New Mexico State University.

Abstract

Since mosquitoes are vectors for fatal diseases, such as malaria, controlling their life cycle or numbers can save human lives. Blood digestion in mosquitoes is an essential part of their life cycle. A previously published metagenomic study shows that mosquito gut bacteria family composition changes following a blood meal [Wang 2011]. The current study seeks to explain one of the reasons for the composition change: oxidative stress that is associated with blood meals This experiment collects and analyzes the metagenome in the midguts from sugar fed and blood fed mosquitoes. The four data collection points in chronological order are Sugar Meal #1, Blood Meal #1, Sugar Meal #2 and Blood Meal #2. Metagenomic DNA-seq and RNA-seq were done using Illumina sequencing. A metagenomic reference was built based on the de novo assembly of Illumina reads. Metabolic pathways were reconstructed upon the reference. Metagenomic RNA reads from the four datasets were mapped to the pathways. In comparison between sugar-fed and blood-fed gut microbial communities, the microbes in the blood-fed guts had more transcripts added to the pentose phosphate pathway (PPP) than to the glycolysis pathway, suggesting the PPP is more active than glycolysis in the blood-fed microbial community than in the sugar-fed community. PPP is a process that generates NADPH, providing more reducing power to tackle oxidative stress. The metabolic reconfiguration in the microbes upon a bloodmeal benefits the mosquitoes by contributing to the gut oxidative homeostasis.

**5. Student Presentor: Ranjan Das**
**Faculty Mentor: Dr. Ming-Ying Leung**

Title
IEDB: A bioinformatics tool for epitope analysis

Authors and affiliations
Ranjan Das[1], Venkatesh Rajamanikkam[1] and Ming Ying Leung[1,2]

[1]Bioinformatics Program and [2]Department of Mathematical Sciences,
The University of Texas at El Paso.

Abstract
An epitope is a small portion of an antigen protein (a foreign protein molecule in the host system), which usually comprises a 9 – 12 amino acid sequence segment that is recognized by antibodies of the host. Epitope prediction on antigen protein surfaces can be helpful for designing effective vaccines. In the present study we explore the bioinformatics tools in the Immune Epitope Database (IEDB) [www.immuneepitope.org] for *in silico* epitope prediction, using four cattle tick vaccine proteins from different parts of the world. These vaccines protect the cattle from tick bites that cause severe diseases affecting the health of our livestock and threatening our food security. IEDB allows for T-cell and B-cell epitope prediction. T-cell epitope prediction requires defining the peptide binding specificity of class I or class II MHC alleles in the host (the cow in our case), and then applying the "netmhcpan" program. B-cell epitope prediction is made by correlating properties such as hydrophobicity, flexibility, accessibility, turns, exposed surface, polarity with the antigenic propensity of polypeptides chains. We are currently comparing the prediction results from different prediction methods to identify the most likely epitope segments for further experimental analysis. We also plan to study the different computational prediction methods and compare their accuracies.

**6. Student Presenters: Rene Duran and Andres Ortiz**
**Faculty Mentors: Drs. Vladik Kreinovich and Mahesh Narayan**

Title

Developing a platform to enhance the oral bioavailability of novel pharmacophores via a protein drug delivery system for treatment of cancerous and neurodegenerative diseases"

Authors and affiliations

Rene Duran[1], Andres Ortiz[2], Mahesh Narayan[1], Jose E. Nunez[1], and Vladik Kreinovich[3]

[1]Department of Chemistry, [2]Department of Mathematical Science, and [3]Department of Computer Science, The University of Texas at El Paso.

Abstract

Nitrosative stress has recently been demonstrated as a crucial causal in the pathogenesis of Parkinson's and Alzheimer's diseases. Specifically, increased levels of NO disrupt the redox activity of protein-disulfide isomerase which leads to aggregation of misfolded proteins. We have demonstrated *in vitro* that polyphenolic phytochemicals, curcumin and 3,5-Bis(2-fluorobenzylidene)-4-piperidone (EF24), can rescue S-nitroso-PDI formation by scavenging NOx adducts. In our current research, we will explore the ligand-protein binding of 4 novel pharmacophores (pyridineamine platinum (II) complexes) which exhibit similar anti-cancerous effects. Spectrofluorimetry will determine the binding efficacy of these novel pharmacophores to β-lactoglobulin and human serum albumin. Binding of these novel pharmacophores to the milk protein and human transport protein ensures delivery to the microvilli of the intestine where it can be absorbed and transported directly into the systemic blood system. Based on a theoretical model, we propose a method for extracting the binding constants and corresponding binding site concentrations. According to this model, protein-ligand binding curves are solutions to a given linear second order partial differential equation. This method could be an alternative to the standard method of finding dissociation constants and concentrations using least squares methods. The results from this technique will allow us to guide protein-ligand docking experiments.

**7. Student Presenter: Sankara Aditya Garimella**
**Faculty Mentor: Dr. Ming-Ying Leung**

Title

Prediction of protein-protein interaction sites using bioinformatics tools

Authors and affiliations

Sankara Aditya Sarma Garimella[1], Venkatesh Rajamanickam[2], and Ming-Ying Leung[1,2]

[1] Bioinformatics program, [2] Department of Mathematical Sciences,
The University of Texas at El Paso.

Abstract:

The In silico structure prediction of protein-protein complex is also called protein-protein docking. It aims at predicting the orientation of one protein against a second protein when bound to each other to form a complex. Once we know the orientation of the molecules we can predict the protein-protein interaction sites and also the strength of association in order to identify the energetically most favorable orientation using scoring functions. Protein-protein interactions play a central role in biochemistry because the formation of complex is critical to signal transduction (process of eliciting a physiological response). Hence docking helps in the prediction of type, strength of signal produced and also in the structure based drug design which is considered by many to be the most efficient approach for discovering compounds. These days it is mostly used for discovering potential inhibitors against cancer-related proteins that play a critical role in the formation of various tumors. The program we have selected for protein-protein docking is Global Range Molecular Matching (GRAMM). It uses a global search algorithm to perform an exhaustive 6 dimensional search through the relative transitions and rotations of the molecules. To view the molecules from various perspectives after docking by GRAMM it requires molecular graphics visualization tools such as PY-MOL and ICM Browser. In addition, the online server SPPIDER is used to identify the amino-acids at the protein-protein interaction interface. With this collection of tools, we are exploring the possible interactions between a kinase receptor called Janus kinase (JAK-3) and potential inhibitor anaplastic lymphoma kinase (ALK) since the over expression of the JAK-3 was implicated in causing various cancerous disorders like adult T-cell lymphoma/leukemia(ATLL).

**8. Student Presenter: Rishabh Jain**
**Faculty Mentor: Dr. Chuan (River) Xiao**

Title

RIVEM 2: A Comprehensive Software tool for creating 3D to 2D plots of Protein Surface.

Authors and affiliations

Rishabh Jain[1] and Chuan Xiao[2]

[1]Bioinformatics Program and [2]Department of Chemistry,
The University of Texas at El Paso.

Abstract

Although 3Dimensional analysis of electron density map is important, we also need to quantitatively analyze the maps. For this reason 2-dimensional projection of the maps is required. RIVEM2 will be developed to convert 3 dimensional electron density maps obtained from Cryo Electron Microscope (EM) to 2 dimensional plots, which will be presented as a stereographic diagram. This will help us to allow clear visualization and interpretation of high resolution 3D maps obtained from Cryo EM. RIVEM2 will be a newer version of previously developed tool : RIVEM (Radial Interpretation of Viral Electron Density maps). RIVEM was only useful to analyze spherical structures. RIVEM2 would allow scientists to analyze helical and other non spherical structures. RIVEM2 will include a wide range of features to make it much more flexible and usable than RIVEM.

**9. Student Presenter: Harish Kalyani**
**Faculty Mentor: Dr. Surekha Rani**

Title

Immunoinflammatory response in coronary artery disease

Authors and Affiliations

Harish Kalyani[1,2], Rajesh Kumar[1], Surekha Rani[1]

[1] Department of Genetics and Biotechnology, Osmania University, India,

[2] The University of Texas at El Paso, El Paso.

Abstract

Atherosclerosis, formerly considered an insipid lipid storage disease, actually involves an ongoing inflammatory response in Coronary Artery Disease. Immune responses participate in every phase of atherosclerosis. Immune cells dominate early atherosclerotic lesions, their effectors molecules accelerate progression of the lesions, and activation of inflammation can elicit acute coronary syndromes.

Previous clinical studies have shown the role of inflammation in pathogenesis of atherosclerosis which applies directly to human patients. Elevation in markers of inflammation predicts the outcome of patients with acute coronary syndrome, independent of myocardial damage. With improved understanding of the pathogenesis of heart disease, new insights into potential markers of underlying atherosclerosis and cardiovascular risk have been gained. It is studied that both systemic and local inflammation plays a key role in the development of atherosclerosis. There is a strong association between systemic inflammation and coronary artery disease (CAD), and this association is thought to be perfunctory, implying inflammation increases the risk of disease. A G>C single-nucleotide polymorphism at position -174 in the promoter region of *IL-6* gene has been associated with prevalence, incidence and prognosis of cardiovascular disease. Understanding some of the genetic mechanisms controlling *interleukin-6* levels may be of benefit further. The present study was undertaken to understand the *IL-6* polymorphism and to estimate the CRP & ADA levels to evaluate their role as marker in the prediction of CAD and found significant results when compared with controls.

**10. Student Presenter: Joe Knapka**
**Faculty Mentor: Dr. Ming-Ying Leung**

Title
Review Presentation: Probing the deep phylogeny of the eukaryotes

Presenter and affiliation
Joe Knapka, Bioinformatics Program, The University of Texas at El Paso.

Abstract
The deep phylogeny of the eukaryotes is a subject of ongoing investigation. For the past decade, Ivan Rogozin, Miklos Csuros,and Eugene Koonin, along with various co-investigators, have applied innovative statistical analyses of molecular evidence to this question. This poster will summarize findings from two of their recent papers: Rogozin et al 2009, which applies an analysis of rare genetic changes to the initial radiation of the eukaryotes and finds support for multiple endosymbiosis events; and Csuros et al 2011, which applies a large-scale Markov Chain Monte-Carlo simulation to the process of intron gain and loss, and finds support for an intron-rich genome at the root of the eukaryotic tree.

**11. Student Presenter: April Lewis**
**Faculty Mentor: Dr. Ming-Ying Leung**

Title:
Review presentation: "De novo computational prediction of non-coding RNA genes in prokaryotic genomes,"

Presenter and affiliation
April Lewis, Bioinformatics Program, The University of Texas at El Paso.

Abstract
This poster is a review of the paper "De novo Computational prediction of non-coding RNA genes in prokaryotic genomes" by Tran *et al.* (2009) published in the journal *Bioinformatics*. The abstract, as originally published by the authors, is as follows.

"Motivation:
The computational identification of non-coding RNA (ncRNA) genes represents one of the most important and challenging problems in computational biology. Existing methods for ncRNA gene prediction rely mostly on homology information, thus limiting their applications to ncRNA genes with known homologues.

Results:
We present a novel de novo prediction algorithm for ncRNA genes using features derived from the sequences and structures of known ncRNA genes in comparison to decoys. Using these features, we have trained a neural network-based classifier and have applied it to Escherichia coli and Sulfolobus solfataricus for genome-wide prediction of ncRNAs. Our method has an average prediction sensitivity and specificity of 68% and 70%, respectively, for identifying windows with potential for ncRNA genes in E.coli. By combining windows of different sizes and using positional filtering strategies, we predicted 601 candidate ncRNAs and recovered 41% of known ncRNAs in E.coli. We experimentally investigated six novel candidates using Northern blot analysis and found expression of three candidates: one represents a potential new ncRNA, one is associated with stable mRNA decay intermediates and one is a case of either a potential riboswitch or transcription attenuator involved in the regulation of cell division. In general, our approach enables the identification of both cis- and trans-acting ncRNAs in partially or completely sequenced microbial genomes without requiring homology or structural conservation."

**12. Student Presenter: Divya Mandyam**
**Faculty Mentor: Dr. Stephen B. Aley**

Title

Towards a model of foreign gene integration in Mycobacterium phage

Authors and Affiliations

Divya Mandyam[1] and Stephen B. Aley[1,2]

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

Abstract

The acquisition of new genes by lateral gene transfer (LGT) expands genomic diversity within a species. Newly acquired genes may integrate into the recipient genome by recombination or replicate independently as a plasmid. As a consequence of this process, the recipient organism expresses traits not necessarily shared by other members of the same species. Previous studies suggest that the mosaic architecture of Mycobacterium phage genomes' resulted from numerous LGT events between phage and their hosts or different species of phage. However, the mechanism by which these foreign genes integrated into their recipient genomes is unclear. The Boundary Sequence model predicts that integration occurs by homologous recombination where the regions of homology are provided by short sequence stretches conserved in the source and recipient genome. On the other hand, the Illegitimate-recombination model postulates that a foreign gene from a source genome inserts quasi-randomly into any region of a recipient genome. Our investigation aims to elucidate which of the two models, if any, can explain the distribution of LGT genes in Mycobacterium phage. By finding the correct model, given a foreign gene, we can predict if and where it will insert into a Mycobacterium phage Genome. Using Phamerator, a tool designed to perform comparative genome analysis on Mycobacterium phage, we examined putative LGT genes and their flanking genes in cluster A6 phage. This allowed us to assess whether foreign genes integrate randomly or at specific regions. We found that in most cases, the genes flanking an LGT insert are the same across all A6 members containing the insert. Additionally, the direction of an insert and its flanking genes are always the same. Both results suggest that integration may be occurring by the Boundary Sequence model, where the regions of homology required for recombination may come from the flanking genes.

**13. Student Presenters: Abigail Muniz and Jesus Loya**
**Faculty Mentor: Dr. Hamide Dogan**

Title
Matrix Product Application to Fibonacci Sequence

Authors and Affiliations
Abigail Muniz, Jesus Loya, and Hamide Dogan,

Department of Mathematical Sciences, The University of Texas at El Paso.

Abstract
We will discuss a procedure that leads to the computation of the "nth" Fibonacci number applying matrix product. We will provide a specific 2x2 symmetric matrix and demonstrate matrix product connection to Fibonacci sequence through diagonal matrices and the power operation on them. Ultimately, we will reveal how matrix product and power operation on diagonal matrices allow one to calculate any value in the Fibonacci Sequence with ease.

**14. Student Presenter: Alexandria Ogrey**
**Faculty Mentor: Dr. Ming-Ying Leung**

Title
Review Presentation: "Cell++ - simulating biochemical pathways"

Presenter and affiliation
Alexandria Ogrey, Bioinformatics Program, The University of Texas at El Paso.

Abstract
This poster is a review of the paper "Cell++ - simulating biochemical pathways" by Sanford et al. (2006) published in the journal *Bioinformatics.* The abstract, as originally published by the authors is as follows:

"Motivation:
With the generation of a wealth of information, detailing cellular components, their functions and interactions, there is a growing need for the development of new computational tools capable of interpreting these data within spatial and dynamic contexts. Here, we introduce Cell++, a novel stochastic simulation environment with the capacity to study a wide variety of biochemical processes within a spatial context.

Results:
Focusing on three case studies, we highlight the potential impact of spatial organization in the evolution and engineering of signaling and metabolic pathways. In addition to altering signaling and metabolic efficiency, simulations also demonstrated features consistent with the phenomenon of metabolic channeling.

Availability:
Cell++ is licensed under the GNU general public license (GPL) and has been successfully implemented under Linux and IRIX operating systems. Source code together with a simple tutorial is available at http://www.compsysbio.org/CellSim/."

**15. Student Presenter: Venkatesh Rajamanickam**
**Faculty Mentor: Dr. Ming-Ying Leung**

Title:

*Rhipicephalus microplus* Bm86 cattle tick vaccine structure prediction and its GPCR target identification

Authors and affiliations

Venkatesh Rajamanickam[1], Sankara Aditya Sarma Garimella[1], Ming-Ying Leung[1,2], and Felix Guerrero[3]

[1]Bioinformatics program and [2] Department of Mathematical Sciences, The University of Texas at El Paso; [3]USDA-ARS Knipling-Bushland US Livestock Insects Research Laboratory, Kerrville, Texas

Abstract

GAVAC is a strain of the Bm86 tick protein used in vaccine against cattle ticks. It terminates the occurrence of cattle fever in Cuba, but it fails to works well against *Rhipicephalus (Boophilus) microplus* from Texas and Mexico. It is hypothesized that the Texas tick Bm86 protein, called Bm86Deutsch, differs from GAVAC in structure, epitope regions and target proteins. In this study, we apply various bioinformatics tools to explore the similarity and differences in sequence and structure among GAVAC, Bm86Deutsch, and two other Bm86 proteins Bm86TICKGARD (Australian Bm86 vaccine protein) and Bm86CampoGrande (Brazilian Bm86). We have found the hydrophobicity profiles and predicted the protein secondary structure using programs such as PORTER, I-Tasser, and the Prime module from the Schrödinger package. Currently others in our group are attempting to predict the T-cell and B-cell epitope regions using IEDB analysis resource. With the available expressed sequence tags (EST) from the *Boophilus microplus* genome sequence, we are interested in finding the potential G-Protein coupled receptor transcripts from the EST data using the publicly available PRED-GPCR online server at http://athina.biol.uoa.gr/bioinformatics/PRED-GPCR/ to predict the potential GPCR proteins based on a family specific profile hidden Markov model algorithm along with an exhaustive discrimination method to construct a library of highly selective GPCR family signatures. Our goal is to elucidate the molecular mechanisms by which these vaccine proteins interact with the immune system of the cattle.

**16. Student Presenter: Vignesh Ravichandran**
**Faculty Mentor: Professor Karthi S.**

Title

Dissecting the binding affinities of natural anti-obese molecules with potent receptors

Authors and affiliations

Vignesh R.*, Gayathri G., Karthi S.

Department of Bioinformatics, School of Chemical and Biotechnology, SASTRA University, Thanjavur, India.

*Currently with the Bioinformatics Program, The University of Texas at El Paso.

Abstract

Obesity, addressed by the Ayurveda practitioners as 'Medoroga', is associated with many diseases, particularly diabetes, osteoarthritis, hypertension and heart disease. The obesity is increasing at an alarming rate in recent years, becoming a global epidemic. The risk factor involved in the synthetic drugs and the costs involved in their synthesis and treatments have made the drug practitioners to incline towards the natural plant products. The objective of this work is aimed at analyzing the affinity of the four naturally occurring compounds Hydroxycitric acid, Gugglusterone E and Z and Mangiferrin, which are proved anti-obese in crude experimental analysis, towards most of the obesity related biological receptors and computationally measures their potential to bind to them resulting in anti-obesity activity. The scores obtained from the computational docking studies serve as a direct measure of the affinity of the ligands for the receptors. Mangiferin exhibits the highest affinity for PPAR-alpha followed by Insulin Receptor and 11-beta-HSD while Hydroxycitric acid exhibits the highest affinity for Insulin receptor followed by 11-beta-HSD and FXR. Guggulsterone E and Guggulsterone Z have higher affinity for FXR and 11-beta-HSD.Overall binding kinetics is inclined towards 11 beta HSD and FXR. 11 beta HSD is a key player in the anorectic pathway whereas FXR plays an important role in the regulation of lipid metabolism. Hence these natural compounds control obesity either by inhibiting appetite or altering the rate of lipid metabolism.

**17. Student Presenter: Janeth Rodriguez**
**Faculty Mentor: Dr. Jianying Zhang**

Title

Protein-protein interaction of HSP70 and HSP60 with other cancer-related proteins

Authors and affiliations

Janeth Rodriguez[1], Ming-Ying Leung[2,3], Jianying Zhang[1,2]

[1]Department of Biological Sciences, [2]Bioinformatics Program, and [3]Department of Mathematical 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. PPI Networks require the use of in-silico methods to find other cancer-related proteins, called interactors, which interact with the target proteins HSP70 and HSP60. 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 so that we can better understand their roles in cancer. The Needleman and Wunsch dynamic programming algorithm was used to analyze the interacting proteins. Eighty-eight cancer-related proteins that interact with the HSP60 were found in the Human Protein Reference Database (HPRD), of which 34 were expressed in cancer, and 31 interactors were found for the HSP70 in the same database. HPRD database employed the Paralogous Verification Method (PVM), Expression Profile Reliability (EPR) and Domain Pair Verification (DPV) as well as other methods. These interactors need to be analyzed in the future to see which of them are the ones that interact the most with the target proteins. The computational analysis results can be used as a guide to help design wet-lab experiments and is expected to save valuable time and lab resources.

**18. Student Presenter: Eric Vest**
**Faculty Mentors: Drs. Kyle L. Johnson and Ming-Ying Leung**

Title

Understanding the influence of RNA secondary structures within Nodamura Virus RNA 1 on the synthesis of sub-genomic RNA3

Authors and affiliations

[1]Eric Vest, [1]Joshua Frederick, [1]Mireya Ramos, [1]Iliana Juarez, [2,3]Ming-Ying Leung, and [1,2]Kyle L. Johnson

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

Abstract

*Nodamura virus* (NoV) is the type species of the *alphanodavirus* Genus of the virus Family *Nodaviridae*, which also includes *Flock House virus* (FHV). NoV is the first virus described that is lethal for both mammalian and invertebrate organisms. The positive-sense NoV genome is comprised of two RNA molecules, RNA1 and RNA2. RNA1 encodes the viral RNA-dependent RNA polymerase (RdRp), while RNA2 encodes the capsid protein precursors. Subgenomic RNA3, which is identical to the 3'-most 470 nt of RNA1, is transcribed by the RdRp from an RNA1 template during RNA replication. The exact mechanism of RNA3 synthesis remains unclear. For example, it is not known whether positive strand RNA3 is synthesized by internal initiation on a negative strand template or by premature termination of a negative strand template that is copied into the positive strand. Computational analysis using the PseudoknotsRG minimum free energy and Unified Nucleic Acid Folding (UNIFold) algorithms has identified two regions in NoV RNA1 that are predicted to form a long-range interaction (LRI): a proximal sequence located near the RNA3 start site (nt 2516 to 2727) and a distal region (nt 1249 to 1234) approximately 1500 nt upstream. Because a similar structure is involved in RNA3 synthesis for the related FHV, we wondered whether this structure plays a role in subgenomic RNA synthesis for NoV, We used a well-defined reverse genetic system for NoV RNA replication to test the effect of mutating the nt involved in this predicted LRI. We therefore introduced mutations into a full-length RNA1 cDNA clone and tested the ability of the resulting transcripts to replicate in transfected mammalian cells that also express the wildtype viral RdRp. Mutations were designed to disrupt base-pairing interactions within the proximal region, the distal region, or both regions of the LRI in a strand-specific manner, thereby allowing us to determine the polarity of the template used to synthesize positive-strand RNA3. Clarifying the mechanism of RNA3 synthesis will provide key insight into the regulatory roles that RNA structures play during RNA replication.

**19. Student presenter: Lauren Waldrop**
**Faculty mentor: Dr. Ming-Ying Leung**

Title
Review Presentation: “Tumor evolution inferred by single-cell sequencing”

Presenter and affiliation
Lauren Waldrop, Bioinformatics Program, The University of Texas at El Paso.

Abstract
This poster is a review of the paper "Tumor evolution inferred by single-cell sequencing," by Navin *et al.* (2011) in the journal *Nature*. The abstract, as originally published by the authors, is as follows:

“Genomic analysis provides insights into the role of copy number variation in disease, but most methods are not designed to resolve mixed populations of cells. In tumours, where genetic heterogeneity is common1–3, very important information may be lost that would be useful for reconstructing evolutionary history. Here we show that with flow-sorted nuclei, whole genome amplification and next generation sequencing we can accurately quantify genomic copy number within an individual nucleus. We apply single-nucleus sequencing to investigate tumour population structure and evolution in two human breast cancer cases. Analysis of 100 single cells from a polygenomic tumour revealed three distinct clonal subpopulations that probably represent sequential clonal expansions. Additional analysis of 100 single cells from a monogenomic primary tumour and its liver metastasis indicated that a single clonal expansion formed the primary tumour and seeded the metastasis. In both primary tumours, we also identified an unexpectedly abundant subpopulation of genetically diverse ‘pseudodiploid’ cells that do not travel to the metastatic site. In contrast to gradualmodels of tumour progression, our data indicate that tumours grow by punctuated clonal expansions with few persistent intermediates.”

**20. Student presenter: Daniel Yehdego**
**Faculty mentor: Drs. Kyle L. Johnson and Ming-Ying Leung**

Title

Secondary Structure Predictions for Long RNA Sequences Based on Inversion Excursions

Authors and affiliations

Daniel Yehdego[1],Vikram Kumar Reddy Kodimala,[2] Boyu Zhang,[5] Sameera Viswakula,[3*] Rahul Vegesna,[2*] Michela Taufer,[5] Kyle L. Johnson[2,4], and Ming-Ying Leung[2,3,4]

[1]Computational Science Program, [2]Bioinformatics Program, [3]Department of Mathematical Sciences,[4]Border Biomedical Research Center, Department of Biological Sciences, The University of Texas at El Paso; [5]Department of Computer and Information Sciences, University of Delaware, Newark, DE,

*Sameera Viswakula is currently at the Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA., and Rahul Vegesna is currently at The University of Texas M. D. Anderson Cancer Center, Houston, TX,

Abstract

The tremendous demand on computer memory and computing time for prediction of complex secondary structures limits the applicability of most RNA secondary structure prediction programs available to short RNA sequences. We propose to approach this problem by segmenting a long RNA sequence into shorter non-overlapping chunks, predicting the secondary structures of each chunk individually, and then assembling the prediction results to give the structure of the original sequence. The selection of cutting points is a crucial component of the approach. Noting that stem-loops and pseudoknots always contain an inversion, we developed two cutting methods, the centered and optimized methods, for segmenting long RNA sequences based on inversion excursions. For the majority of the sequences in a dataset of 50 RNAs from the RFAM database, the prediction algorithm PKnotsRG used with these cutting methods produces more accurate secondary structures than those predicted for the whole sequence without segmentation. Both the centered and optimized cutting methods outperform the naïve regular segmentation. These results support our claim that cutting is a promising approach for the prediction of long RNA sequences, and choosing the cutting points intelligently by considering sequence features such as inversion excursions can further enhance prediction accuracy.

**21. Student presenter: Meng Zhang**
**Faculty mentor: Dr. Jianying Zhang**

Title

Three-dimensional Reconstruction of Protein IMP2/P62

Authors and affiliations

[1] Meng Zhang, [3]Gary G. Ren, [3]Lei Zhang, and [1,2]Jianying Zhang

[1]Bioinformatics program and [2]Department of Biological Sciences, The University of Texas at El Paso

[3]Molecular Foundry, Lawrence Berkeley National Laboratory

Abstract

The growing matured technique termed single particle reconstruction is used to visualize protein structure at an intermediate resolution (1-2nm). The advantage of this technique is the combined use of the transmission electron microscopy (TEM) and bioinformatics tools in which the 2D structure of the protein is first obtained by TEM and then the 3D structure is subsequently reconstructed by bioinformatics analysis. In this project, we aim to solve the structure of a previously unknown protein called p62/IMP2, which is a tumor-associated autoantigen and is overexpressed in various types of cancer tissues. By using an optimized Uranyle Formate Negative Staining (UF-NS), TEM particle images of p62/IMP2 were obtained and subjected to single particle analysis. With the help of EMAN-1 and SPIDER-20.16 software packages, the images were classified and used for 3D structure reconstitution. The preliminary analysis shows that the folded p62/IMP2 has an average diameter of around 5nm with 6 similar subunits tightly packed together.