1096-D5-402 **Majid Masso*** (mmasso@gmu.edu), 10900 University Blvd. MS 5B3, Manassas, VA 20109.

Predicting Drug Resistance: Probability and Statistics Meet the Building Blocks of Proteins.

Put away the dice and bag of balls, and consider structural biology examples! This talk begins with the 20-letter amino acid alphabet, which are protein building blocks. A protein has hundreds of consecutively linked amino acids, selected with replacement, forming a unique sequence of letters. The linear chain folds like a balled-up thread into a precise 3D structure, held together by interactions between spatially close amino acids often distant in the sequence. These basic ideas are sufficient to begin work on methods of enumeration, probability distributions, and log-likelihoods.

We developed a technique that generates a distinct feature vector to represent each mutated form of a protein due to amino acid changes. This approach was applied to nearly 500 mutants of the HIV-1 protease enzyme whose altered susceptibilities to the FDA approved drug ritonavir were quantified with a time consuming and expensive assay. By implementing classification and regression statistical learning algorithms, the data set was used to train models for accurately predicting ritonavir resistance in new protease mutants isolated from patient viruses. Since the models are fast and cost-free, they could be used as supplementary diagnostic tools. Our strategy extends to other commercial drugs. (Received September 02, 2013)