Pattern Recognition of Proteases Using Multiplex Peptoid Arrays

Authors: Avanti Panajkar*[1], Abram Huang[2], Veronica Mendez-Gonzalez[3], Hattie Schunk[4], Haoqi Wang[4], Laura Suggs[4], Mia Markey[4]

[1]North Carolina State University, [2]University of Maryland-College Park, [3]University of Puerto Rico-Mayaguez, [4]The University of Texas at Austin

Introduction: Detection of activity levels of proteases is a promising approach to develop a "liquid biopsy" for cancer screening. However, developing sensing methods that are sufficiently accurate, noninvasive, and inexpensive is challenging. Furthermore, many traditional sensors tend to have high false positive rates and can only detect protease classes, but cannot distinguish between proteases within a class.[1] In this work, we aim to use array-based multiplex sensing to obtain more comprehensive information on biological samples. Our sensor arrays are made of sequence-defined peptoids, which allow for enhanced biostability, reduced cross-reactivity, targeted susceptibility, and improved specificity over traditional sensors.[2] Our objective is to develop a pattern recognition algorithm that can classify proteases from fluorescence readouts of peptoid array sensors.

Materials and Methods: In the experiment, we examined the activity of seven proteases by collecting fluorescent responses for eight peptoid-based sensors for a period of 180 minutes over three trials. We visualized the fluorescent data and performed principal component analysis (PCA), an unsupervised pattern recognition technique, as a means of data exploration and to evaluate the dataset's variability. We then used leave-one-out cross-validation with the supervised technique of k-Nearest-Neighbors (k-NN) to predict labels of test datasets.

Results and Discussion: The fluorescent data showed clear differences among the different types of proteases present in the sample (Figure 1A). Furthermore, our PCA score plot illustrates that the proteases can be distinguished using only the first two principal components (PCs) (Figure 1B). A scree plot was used to determine the number of PCs to use for subsequent analysis, and indicated that 95% of the data variability was explained using the first five PCs. With k-NN, we obtained an accuracy of 100%, which indicates that our array-based peptoid sensing system can successfully classify the seven proteases tested. However, this data collection was performed in a controlled environment using commercially purchased enzymes, and has not yet been translated in vitro. Our next steps are to use k-NN to train the model to predict proteases of future, more biologically relevant datasets, and then ultimately test on liquid biopsies of cancer patients.

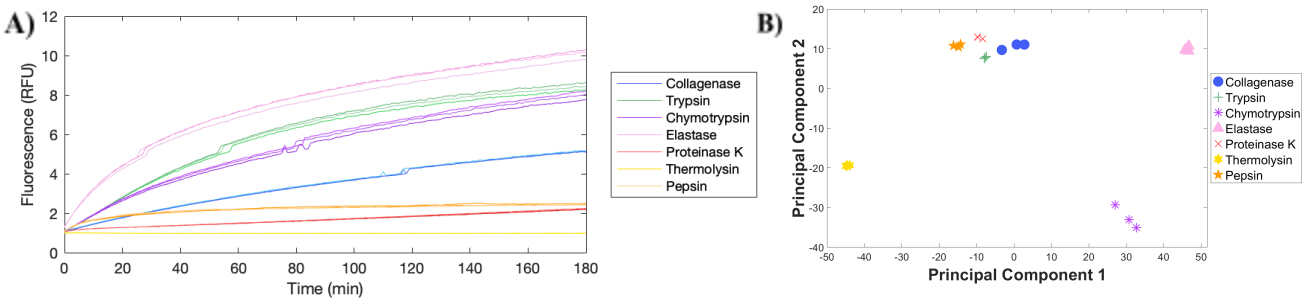

Fig 1. Sensor Data and PCA. A) Fluorescence over time presented for one sensor of the array when screened against seven proteases. **B)** PCA score plot of first two principal components, grouped by protease class.

Conclusion: Accurate detection of proteases present in a liquid biopsy is important for screening and early diagnosis of cancer in primary care. Earlier diagnosis can lead to better prognosis for patients and higher quality of life. Additionally, our model can be applied to different diseases where multiple proteases are active, and aid in identification of biomarkers used for further development of targeted therapies.

References: [1] Rotello, V. et al. ACS Sensors. 2016. 1(11), 1282 - 1285. [2] Knight, A. et al. Adv. Mater. 2015. 27: 5665 - 5691