Pattern Recognition of Proteases Using Multiplex Peptoid Arrays

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.
- Traditional sensors tend to have high false positive rates and cannot distinguish between proteases within a class.[1]
- We aim to use array-based multiplex sensing to obtain more comprehensive information on biological samples.
- Our sensor arrays consist of sequence-defined peptoids, which allow enhanced biostability and improved specificity.[2]
- **Objective:** to develop a pattern recognition algorithm that can classify proteases from fluorescence readouts of peptoid array sensors.

Methods

- Collected and visualized fluorescent responses for eight peptoid-based sensors for 180 minutes over three trials
- Performed principal component analysis (PCA), an unsupervised pattern-recognition technique, as a means of data exploration and to evaluate the dataset's variability
- Used leave-one-out cross-validation with the supervised technique of k-Nearest-Neighbors (k-NN) to predict labels of test datasets.

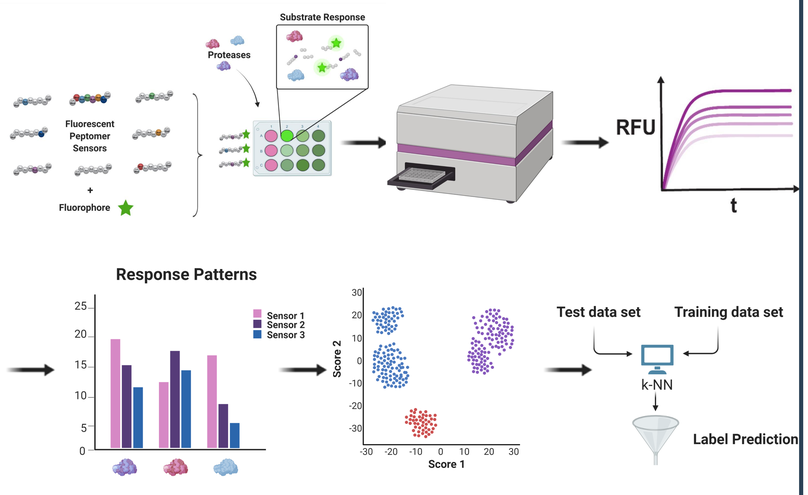

Fig. 1: Experimental Workflow

Figures and Results

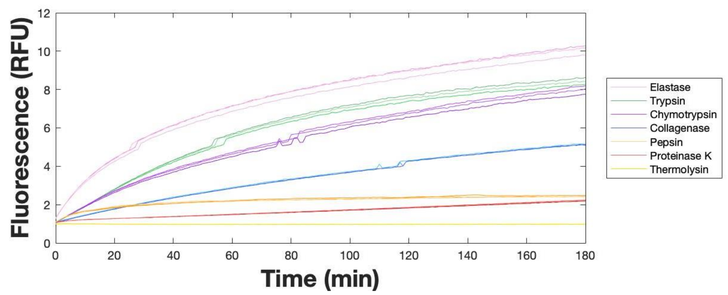

Fig. 2 Fluorescent data, presented for a sensor here, show clear differences in response to all seven types of proteases tested

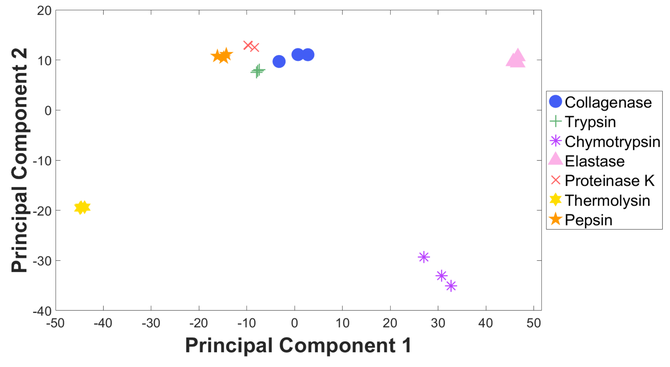

Fig. 3 PCA score plot illustrates that the proteases can be distinguished using only the first two principal components (PCs).

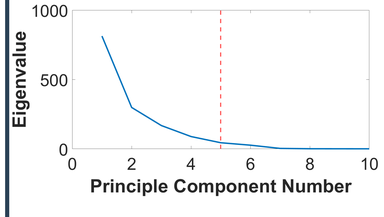

Fig. 4 Scree plot indicates that 95% of data variability can be explained using first five PCs

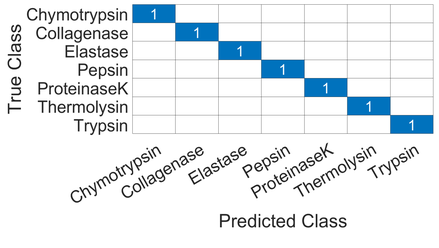

Fig. 5 Confusion matrix for k-NN showing accuracy of 100%

Discussion

- **Outcome:** our classification accuracy indicates that our array-based peptoid sensing system can successfully classify the seven proteases tested.
- **Limitations:** data collection was performed in a controlled environment using commercially purchased enzymes and has not yet been translated *in vitro.*
- **Next steps:** 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.

Conclusions

The 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.

Acknowledgments

Special thanks to NSF Research Experience for Undergraduates Award (1461192) and the Department of Biomedical Engineering at the University of Texas at Austin for supporting our research.

References

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