

Abstract

Post-translational modifications (PTMs) refer to the attachment of molecules (e.g., phosphates, sulfates, glycans, ubiquitins, etc.) to amino acids in proteins. These modifications change the properties of the protein and play a large role in protein function, regulation, and cell signaling. PTMs also play a role in disease (e.g., Alzheimer's disease via Tau hyperphosphorylation, Huntington's disease via hypopalmitoylation of huntingtin protein, and type 2 diabetes mellitus via malonylation of enzymes involved in glucose and lipid metabolism); therefore, understanding PTMs and locating these modifications can provide valuable targets for therapeutics. However, discovery of PTMs and their locations on proteins is very time consuming, laborious, and expensive. Computational methods, such as machine learning, are required to alleviate these challenges. Logistic regression, support vector machines, random forests, and k-nearest neighbors are some traditional machine learning methods that were used with good success in predicting PTMs. As machine learning and AI have advanced tremendously in the last few years, advanced machine learning techniques such as Convolutional Neural networks(CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), Transformer Models, and Graph Neural Networks (GNNs) have been used and outperformed traditional machine learning methods. This work will explore the use of a CNN to predict phosphorylation sites using the Pytorch package in Python and data from UniProt. Structural data, attention heads, PTM crosstalk, and bootstrapping will be considered in future work to improve the model.

Introduction

Figure 1: Ω is the fluid domain with boundary Γ . A no-slip boundary condition is imposed on each body γ_ℓ . On the outer geometry Γ , a Hagen-Poiseuille flow is imposed. The bodies are constrained to the middle third of the channel that is located between the dashed lines. The bodies erode at a rate that is proportional to the shear stress.

Figure 2: The erosion of 50 bodies. The individual snapshots are evenly spaced in time. The color is the vorticity of the fluid whose magnitude is proportional to the rate of erosion.

Numerical Challenge

(a) (b)
(c) (d)

Figure 3: The difference of the velocities using the two quadrature methods. Since the flow vanishes on the bodies, the flow on the red dashed line should be close to zero. (a) Difference of the velocities in the x-direction; (b) Difference of the velocities in the y-direction; (c) The velocities in the x-direction on the red circle; (d) The velocities in the y-direction on the red circle.

Numerical Examples

Figure 4: Tracers move in the fluid from left to right. Since we use highly accurate numerical methods, the tracers can be very close to the bodies.

(a) (b) (c)

Figure 5: (a): The tortuosity (mean distance traveled) of an eroding geometry initialized with 100 bodies. (b): The dispersion (spreading) of the same eroding geometry. (c): The Delaunay triangulation of the porous media.

(a) 100 bodies, 258 gaps (b) 100 gaps, 258 gaps (c) 97 bodies, 249 gaps
(d) 92 bodies, 234 gaps (e) 75 bodies, 189 gaps (f) 34 bodies, 74 gaps

Figure 6: The distribution of the gaps of 100 eroding bodies at six porosities. The black curves are Weibull distributions whose first and second moments agree with the data.

References

- [1] Bryan D. Quaife, and M. Nicholas J. Moore. A boundary-integral framework to simulate viscous erosion of a porous medium *Journal of Computational Physics*, 375:1-21
- [2] Alex Barnett, Bowei Wu, and Shravan Veerapaneni. Spectrally-accurate quadratures for evaluation of layer potentials close to the boundary for the 2D Stokes and Laplace equations. *SIAM Journal on Scientific Computing*, 37(4):B519-B542