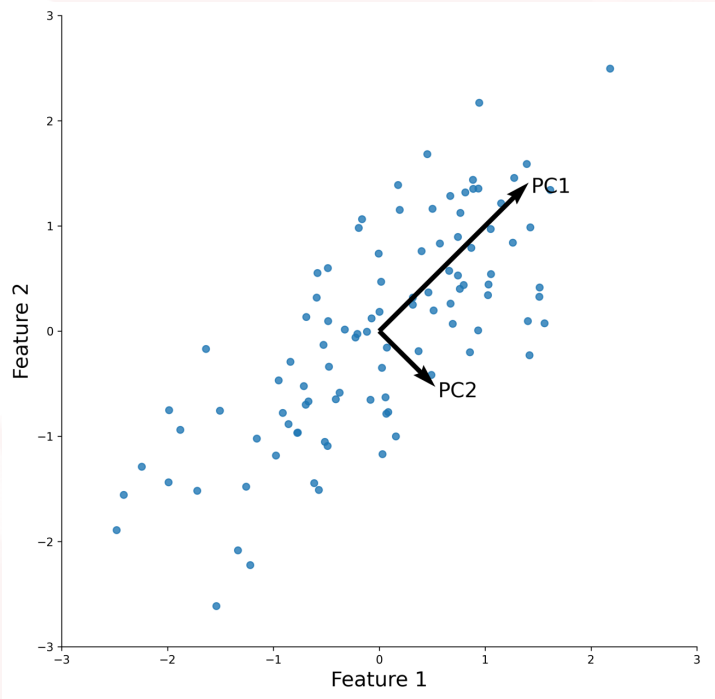
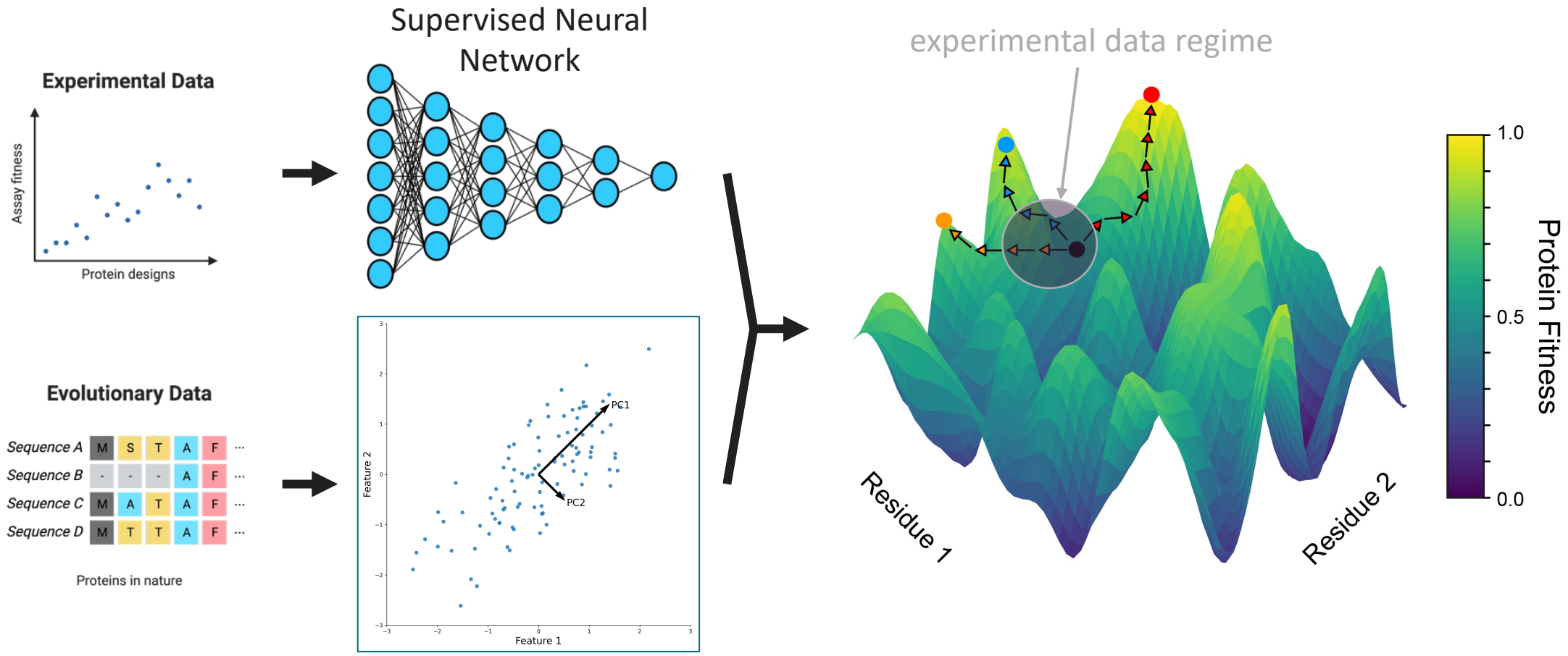
Brief Assessment of Principal Component Analysis for Protein Design



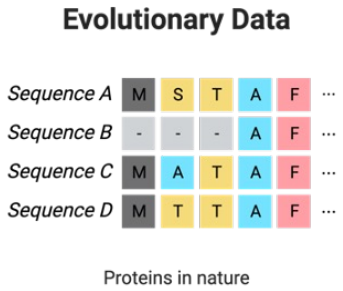
Nathaniel Blalock

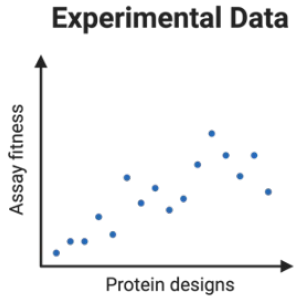
Unsupervised deep learning models can leverage evolutionary data for data-driven protein design



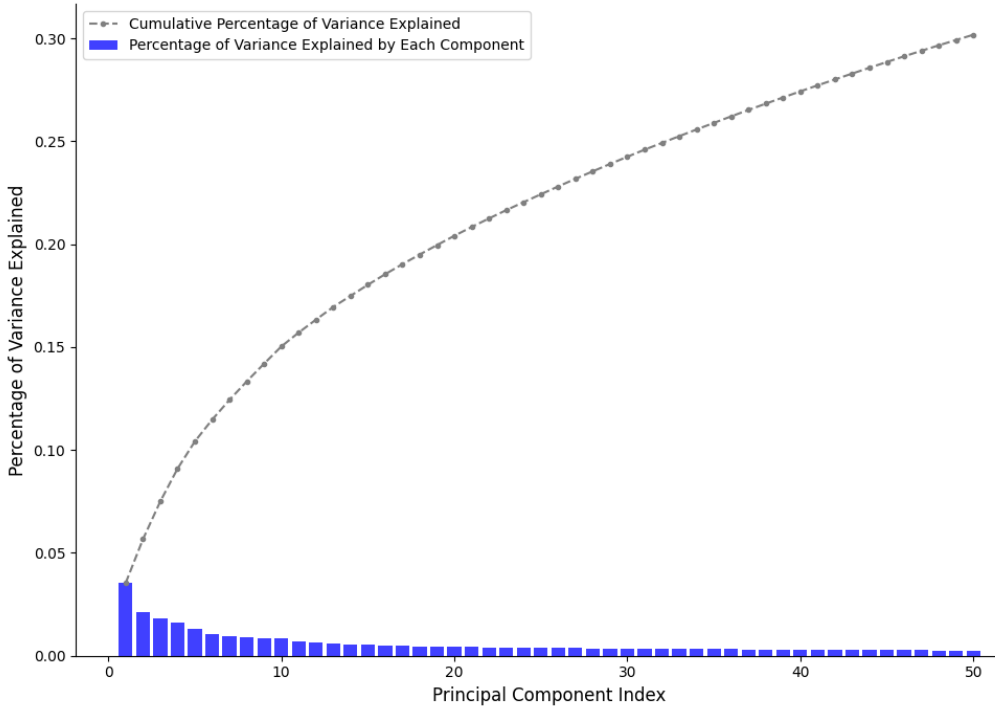
[1] Freschlin, C.R., Fahlberg, S.A. and Romero, P.A., 2022. *Current Opinion in Biotechnology*. [2] Wu et al., 2021. *Current Opinion in Chemical Biology*.

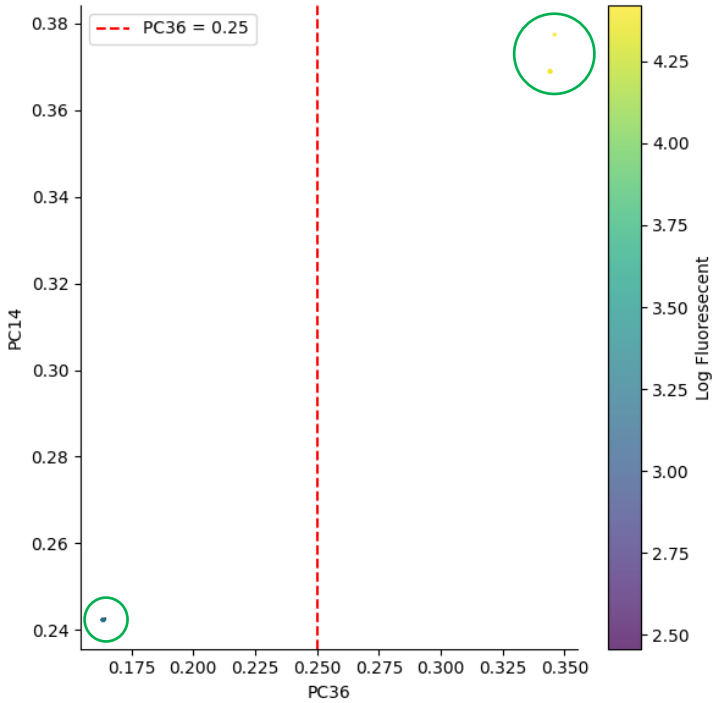
Principal Components from PCA of Multiple Sequence Alignment Discretely Separate Most Fluorescent Proteins

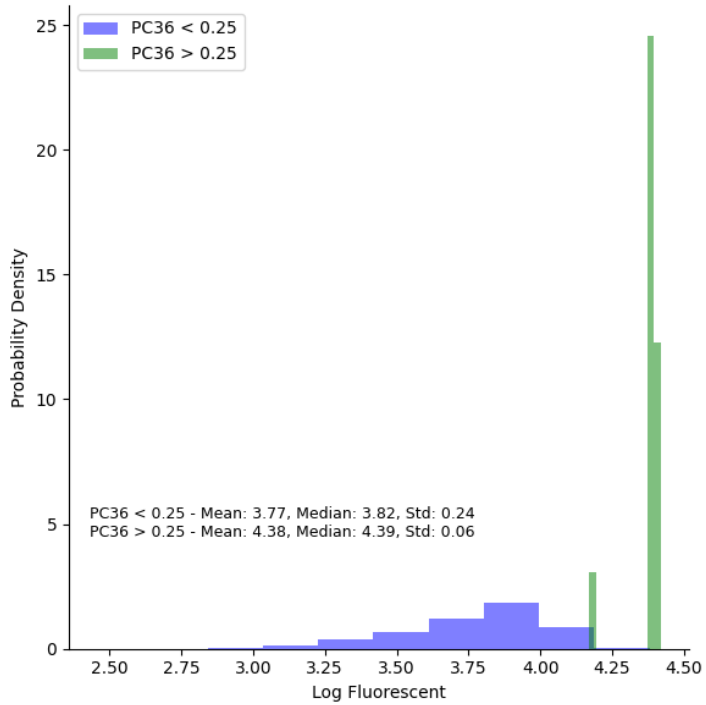




PC36 and PC14 had the greatest spearman correlation with fluorescent measurements



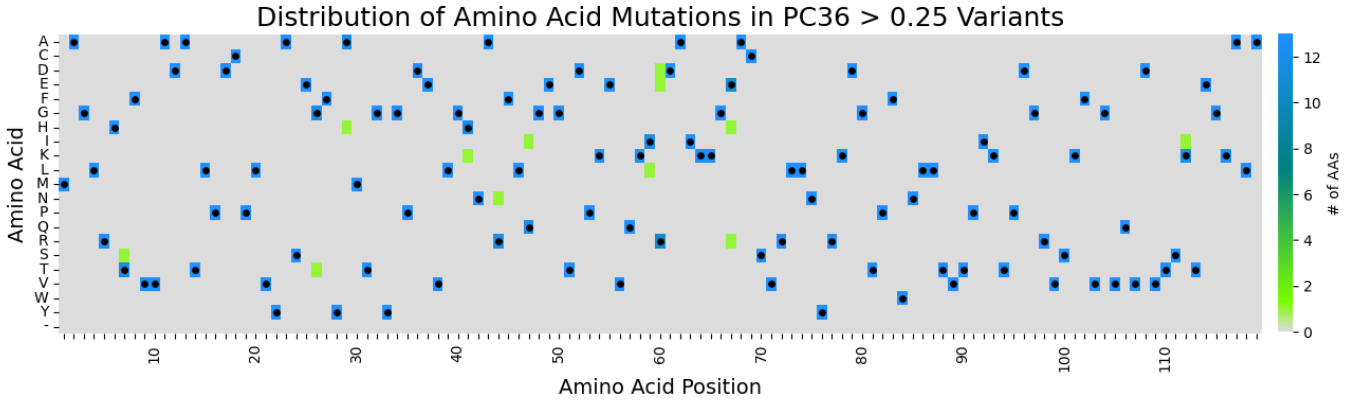




Interpreting the contribution of amino acid positions to principal components may prove meaningful for protein design

Amino Acid Positions Contributing the Most to Correlated Principal Components

Feature Rank	PC36	PC14
1	94	92
2	92	74
3	50	36
4	36	29
5	34	25
6	26	26
7	94	28
8	92	35
9	111	22
10	49	25




WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON