Machine Learning for Efficient Prediction of Protein Redox Potential: The Flavoproteins Case

B.G. Galuzzi.[1,2], A. Mirarchi.[1], E. Viganò.[3], L. De Gioia. [1], C. Damiani.[1,2], F. Arrigoni. [1]
E-mail: bruno.galuzzi@unimib.it
[1] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Italy
[2] SYSBIO Centre of Systems Biology/ISBE.IT, 20126 Milan, Italy
[3] Istituto di Ricerche Farmacologiche Mario Negri, 20156 Milan, Italy
Keywords: Flavoproteins - Redox potential – Supervised Learning – Extreme Gradient Boosting – Feature importance

Abstract:

Determining the redox potentials of protein cofactors and how they are influenced by their molecular neighbourhoods is essential for basic research and many biotechnological applications, from biosensors and bio-catalysis to bioremediation and bioelectronics. The laborious determination of redox potential with current experimental technologies pushes forward the need for computational approaches that can reliably predict it. Although current computational approaches based on quantum and molecular mechanics are accurate, their large computational costs hinder their usage. In this work, we explored the possibility of using more efficient models based on machine learning (ML) for the prediction of protein redox potential, as an alternative to classical approaches. As a proof of concept, we focused on flavoproteins, one of the most important families of enzymes directly involved in redox processes. To train and test different ML models, we retrieved a dataset of flavoproteins with a known midpoint redox potential (Em) and 3D structure. The features of interest, accounting for both short- and long-range effects of the protein matrix on the flavin cofactor, have been automatically extracted from each protein PDB file. Our best ML model (XGB) has a performance error below 1 kcal/mol (~36 mV), comparing favourably to more sophisticated computational approaches. We also provided indications on the features that mostly affect the Em value, and when possible, we rationalized them based on previous studies.

Galuzzi, B. G., Mirarchi, A., Viganò, E. L., De Gioia, L., Damiani, C., & Arrigoni, F. (2022). Machine Learning for Efficient Prediction of Protein Redox Potential: The Flavoproteins Case. *Journal of chemical information and modeling*, *62*(19), 4748-4759.