Accelerating electrostatics-driven pK_a predictions with fast and interpretable deep learning models

Pedro B. P. S. Reis, Marco Bertolini, Floriane Montanari, Walter Rocchia, Diogo Vila-Viçosa, Miguel Machuqueiro, and Djork-Arne Clevert

Ionizable residues greatly influence the physicochemical properties of proteins, such as lipophilicity or polarity, and their function since many biological processes are triggered by changes in the ionization state of protein residues side-chains [1,2]. Due to their central role in protein folding, stability, and function, many computational methods have been developed to understand their pH dependency. Arguably the most popular classes of pK_a predictors are methods based on continuum electrostatics (CE) and empirical ones. In Poisson-Boltzmann-based methods, proteins are represented by point charges in a low dielectric medium while the solvent is implicitly described. Empirical methods rely on statistically fitting parameters over large datasets of experimental pKa values. These are much faster than the physics-based methods, although at the cost of less microscopic insights and unknown predictive power on mutations and proteins dissimilar to those in the training set.

In this work, we have developed deep learning pK_a predictors that combine the best features of CE models – accuracy and interpretability – with the speed of classical empirical methods. These models have been trained on a database of 3 million theoretical p*K*a values estimated from 50 thousand structures using a CE method. With this approach, we can retrieve the physics-based predictions with an average error below 0.4 p*K* units while being up to 1000x faster. Despite not being explicitly trained on experimental pKa values, by adding a regularization term to the cost function that penalizes overestimation, the deep learning method outperforms the original model. Furthermore, we are able to show that these models implicitly learn most of the required energetic contributions such as Coulomb interactions, desolvation, and Hydrogen-bonding.

1. Warshel A, Aqvist J. Electrostatic energy and macromolecular function. Annu Rev Biophys Biophys Chem. 1991;20: 267–298.

2. Kim J, Mao J, Gunner MR. Are acidic and basic groups in buried proteins predicted to be ionized? J Mol Biol. 2005;348: 1283–1298.