Rapid determination of Calcium charge in varying calmodulin environments using machine learning

Pengzhi Zhang, Jaebeom Han, Piotr Cieplak, Margaret. S. Cheung

There is a critical need for developing an accurate computational model for the Calcium ion that advances the understanding of calcium-driven protein-protein interactions and that establishes the principle of target selectivity in a network of signaling proteins. The implementation of polarizable force fields (PFF) in molecular dynamics (MD) simulations appears an excellent choice to address the correct ion-protein interactions in a protein with an affordable computational cost. Among the PFF models, the isotropic polarizable model has limited development to bare ions in aqueous solution and there is no suitable PFF for calcium ions in protein media. Investigation of the mechanism of target binding for calcium sensing protein requires configuration-dependent force fields for divalent-protein pairs, which is usually prohibitively time-consuming using quantum calculations. We developed a method to determine the atomic charge of the calcium ion in CaM in varying solution environments based only on the atomic positions of the calcium ion, nearest water molecule and the corresponding calcium binding loop using random forest regression (RF). The supervised machine-learning model was first trained with about 300 individual loops conformations containing calcium ions that adopt the most populated ternary structure types from MD simulations and then validated using an additional 100 structures. The charges of the training set were generated by using quantum chemistry calculations followed by i-RESP fitting of the charges. The algorithm improves class discrimination and indicates that several variables including relative position of the nearby water molecule, distance and angle between chelating oxygen atoms and the calcium ion influence the calcium atomic charge preference.