Making life easier for theoretical chemists? Selected applications of machine learning to challenging problems

Alberto Vela, David Ramírez-Palma, Oscar X. Guerrero-Gutiérrez, and Lucio Peña-Zarate

Department of Chemistry, Cinvestav-Zacatenco, Av. Politécnico Nacional 2508, Colonia San Pedro Zacatenco, 07360, CDMX, México

Deciding between improving precision or lowering computational cost is one of the dilemmas when studying complex problems of interest in Chemistry and Physics. To the surprise and skepticism of many, including some of the authors of this work, the emergence of methods based on machine learning points towards the end of this crossroad. In this work, we focus on some applications of the ANI and DM21 neural networks (NN). ANI is an Accurate NeurAl networK engINe for Molecular Energies (ANAKIN-ME), which in its version ANI-1ccx, is trained to reproduce calculations with gold standard quality. On the other hand, DeepMind 2021 (DM21) is a NN representing the energy functional and trained to capture the fractional behavior of electrons aiming to capture aspects of the exact functional. This work uses these NNs in three problems: 1) conformational analysis, 2) chemical space exploration, and 3) supramolecular chemistry.

In the first case, we address the issue of gas phase conformational ordering of amino acids, the fundamental building blocks of proteins. Although they have been extensively studied, theoretically and experimentally, the theoretical prediction of the most stable conformer varies significantly depending on the method used. In many cases, it is mandatory to use very demanding methods. In the second problem, we present an exploration of chemical space. It has been estimated that the total number of possible small organic molecules populating 'chemical space' exceeds 10^{60}. Starting from a minimal formula, the total number of possible conformers grows at an incredible speed; for example, for the minimal formula C_6H_6, the number of potential isomers is 1378, but for C_8H_8, this number increases to 167,379! Thus, we show the prediction of minima using ANI-1ccx, comparing the ordering with CCSD(T) and DFT all-electron calculations. Finally, we present a challenging supramolecular problem: the assembly and movement of a macrocycle inserted in an axis. This study demonstrates that implementing methods trained by machine learning is a viable option for analyzing organic systems, highlighting two aspects: computation time and quality of the results. The neural networks used here allow us a reliable prediction of the conformational ordering of amino acids, a swift exploration of the chemical space of organic molecules, and the possibility of describing the energy profiles and energy barriers involved in the assembly of supramolecular structures.