New life for "classical" force fields in the age of machine learning

Krzysztof Szalewicz
University of Delaware

Since inception, dynamics of clusters and molecular modelling have been based on potential energy surfaces (PESs), also called force fields (FFs). In particular, in biomolecular modelling, several empirical FFs (empFFs) have been widely used (e.g., AMBER, GROMOS, CHARMM), where "empirical" means that parameters of these FFs were fitted to experimental observables in molecular dynamics (MD) simulations. Such empFFs are relatively simple functions of geometric parameters of a molecular cluster, and are often called "classical" FFs despite the fact that their form is derived from quantum mechanics. For systems with smaller molecules, one can circumvent the need for FFs by calculating potential energies on-the-fly in MD using density-functional theory (DFT), leading to methods known as *ab initio* MD (AIMD). In recent years, empFFs and AIMD are often replaced by machine-learned (ML) PESs fitted to *ab initio* data, using methods such as for example neural networks (NNs). While early ML PESs were suffering of low accuracy and failures at certain regions of configuration space, some of more recent ML PESs are the most accurate ones published for a given system.

Work of our group shows that it is possible to develop FFs with functional form not significantly different from that of empFFs, but fitted exclusively to *ab initio* data, called aiFFs, with the same or better accuracy as the best ML-FFs. For small clusters, such as water or hydrogen fluoride dimers or trimers, our fully automated method for developing aiFFs, called autoPES [1], uses an order of magnitude less grid points than needed by ML-FFs and predicts the most accurate rovibrational spectra of these clusters to date. This progress was achieved by using novel analytic functions with several off-atomic sites and accurate asymptotics based on monomer properties and consistent with symmetry-adapted perturbation theory (SAPT) [2, 3] interaction energies at large separations. The intermolecular two- and three-body components, i.e., inter-aiFFs, depend on both intermonomer and intramonomer degrees of freedom (DoFs), i.e., are fully coupled. The intramonomer FFs, intra-aiFFs, depend on intramonomer DoFs only, in accordance with the principles of many-body expansion.

For much larger systems, with dozens of atoms in each monomer, full coupling in inter-aiFFs is not possible due to huge number of DoFs (dimensionality "curse"). Therefore, inter-aiFFs depend in our approach on monomer geometries by atom-following only, which is possible since the potentials are expressed only via atom-atom distances and fits are performed to training sets that include deformed monomers. For intra-aiFFs, we adopt the functional form of empFFs, but with all parameters fitted to *ab initio* data. While for small clusters the best criterion of quality of an FF are rovibrational spectra, for large monomers the most stringent evaluation comes from crystal structure predictions (CSPs). Empirical FFs fail badly in CSPs. For crystals of rigid monomers, our inter-aiFFs have been shown some time ago to provide very reliable predictions [4, 5]. In our initial extension to flexible monomers, i.e., monomers with soft DoFs such a rotations involving dihedral angles, we first tried to use inter-aiFFs with intra-empFFs, but this approach was not successful [6]. We

found that intra-empFFs predict *ab initio* monomer energies encountered in CSPs so poorly that the results are anticorrelated and the *ab initio* equilibrium monomer has a high energy in empFFs. One may wonder how this failure of empirical FFs impacts all biomolecular simulations. In contrast, our intra-aiFFs agree very well with *ab initio* data. When inter-aiFFs plus intra-aiFFs are used in CSPs, predictions are of good quality and require orders of magnitude less computer resources than CSPs performed using periodic DFT.

---

[1] M. P. Metz and K. Szalewicz. Automatic generation of flexible-monomer intermolecular potential energy surfaces. *J. Chem. Theory Comput.*, 16:2317–2339, 2020. doi:10.1021/acs.jctc.9b01241.

[2] J. Garcia, R. Podeszwa, and K. Szalewicz. SAPT codes for calculations of intermolecular interaction energies. *J. Chem. Phys.*, 152:184109–(1:23), 2020. URL: https://doi.org/10.1063/5.0005093.

[3] K. Szalewicz and B. Jeziorski. Physical mechanisms of intermolecular interactions from symmetry-adapted perturbation theory. *J. Mol. Model.*, 28:273–(1:29), 2022. doi:10.1007/s00894-022-05190-z.

[4] K. Szalewicz. Determination of structure and properties of molecular crystals from first principles. *Acc. Chem. Res.*, 47:3266–3274, 2014. doi:10.1021/ar500275m.

[5] R. Nikhar and K. Szalewicz. Reliable crystal structure predictions from first principles. *Nature Comm.*, 13:3095–(1:9), 2022. doi:10.1038/s41467-022-30692-y.

[6] R. Nikhar, R. Podeszwa, A. U. Rehman, O. Ishaque, A. Jing, J. O. Melkumov, R. Hong, and K. Szalewicz. Crystal structure predictions for molecules with soft degrees of freedom using intermonomer force fields derived from first principles. *Acta Cryst. B*, 80:628–655, 2024. doi:10.1038/s41467-022-30692-y.