## ML-based prediction of partial atomic charges



## Motivation Example 3: Atom-centered Atom Pairs (AP) fingerprint [3] O(N<sup>3</sup>) computational cost of DFT; Partial charges were derived using TPSSh/def2-Cumulative fraction of samples • ESP-based methods have a relatively high conformational dependence; 1.0 TZVP electron densities and DDEC method; The inability to a priori designate partial charges. AP fingerprints (atom type, number of heavy-atom $_{0.8}$ Example 1: Pfizer Charge Assignment Method (PCAM)<sup>[1]</sup> neighbors, distances, etc.) represent atoms; 9.0 **Fraction** • A **2D topological representation** eliminates HCNOPSCI Br conformational dependency and ensures R<sup>2</sup>=0.98 • An atom's PCAM charge depends only on its transferability; MUE=0.03 neighbors' configuration; 0.5 0.5 0.0 0.0 0.0 0.0 0.5 It is limited to H, C, N, O, F, P, S, Cl, Br, I atoms; 0.2 It is limited to H, C, N, O, F, S, Cl atoms; RF generates accurate predictions of unseen Each element has a charge model that uses 0.0 0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 compounds. radical and angular symmetry functions to represent local surroundings; Reprinted from J. Chem. Inf. Model. 2018, 58, 579 under ACS AuthorChoice License. © 2018 American Chemical Society. • It employs random forest ML-method and the B3LYP/6-31G\* Example 4: Atom-Path-descriptor (APD) [4] -1.5 -1.5 charge is calculated by: $q_j^0 = \frac{\sum_{i=1}^b T_i(A_j)}{R}$ , where B is -1.0 -0.5 0.0 0.5 10 15 RMSE [e\*10-3] ESP-fit charges Element Using APD descriptors (2 and 3D the tree count; information) to predict partial atomic [3] APD-RF APD-XGB • The formula calculates corrected charges: $q_i = q_i^0$ charges; 93 72 63 Н Atoms are represented by multiple layers, Reprinted with permission from J. Comput. Chem. 2013, 34, 1661. © 2013 John Wiley and Sons. with each layer encoding information such Ρ 685 561 372 Example 2: ContraDRG<sup>[2]</sup> as atom type, distance, bond type, and so С 220 142 114 Molecules are represented as cyclic undirected on. graphs with additional information like the number Bioinformatics 2020. 36, 4721. of nested circles, distances between nearby Example 5: Multiple-Choice Knapsack Problem ( $\varepsilon$ -MCKP)<sup>[5]</sup> atoms, etc.; • Atomic positions and adjacency matrix (bonds) are Molecules are considered as graphs (G = (V, E, t)); needed to encode molecules; Create a database with a record for each isomorphism It is limited to H, C, N, O, F, P, S, Cl, Br, I atoms; class in the subgraph list for each atom; Each element has a charge model; Each isomorphism class's partial charges are condensed A detailed benchmark shows KR is the best model. into a histogram. Reprinted from Front. Genet. 2019, 10, 990 under CC-BY OA license. © 2019 Martin and Heider Reprinted from Algorithms Mol. Biol. 2019, 14, 1 under CC-BY 4.0 OA license.