Stiv Llenga | 20.07.2023

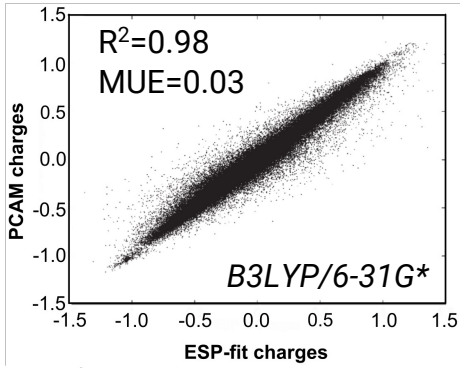
ML-based prediction of partial atomic charges

Motivation

- *O(N³) computational cost of DFT;*
- ESP-based methods have a relatively high conformational dependence;
- The inability to a priori designate partial charges.

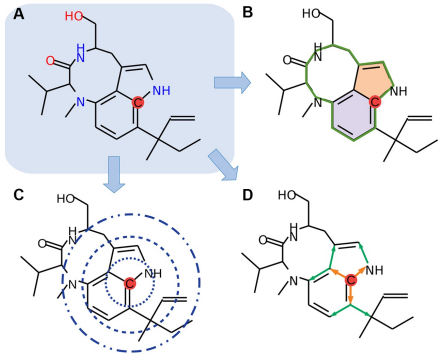
Example 1: Pfizer Charge Assignment Method (**PCAM**) [1]

- An atom's PCAM charge depends only on its neighbors' configuration;
- It is limited to H, C, N, O, F, S, Cl atoms;
- Each element has a charge model that uses **radical and angular symmetry functions** to represent local surroundings;
- It employs random forest ML-method and the charge is calculated by: $q_j^0 = \frac{\sum_{i=1}^{b} T_i(A_j)}{B}$, where B is the tree count;
- The formula calculates corrected charges: $q_j = q_j^0 - \frac{q_j^0 * \sigma_j * \Delta Q}{Q_{abs}}$.



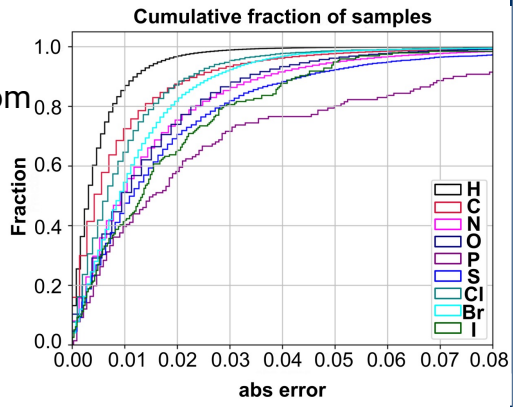
Example 2: **ContraDRG** [2]

- Molecules are represented as **cyclic undirected graphs** with additional information like the number of nested circles, distances between nearby atoms, etc.;
- Atomic positions and adjacency matrix (bonds) are needed to encode molecules;
- It is limited to H, C, N, O, F, P, S, Cl, Br, I atoms;
- Each element has a charge model;
- A detailed benchmark shows KR is the best model.



Example 3: Atom-centered Atom Pairs (**AP**) fingerprint [3]

- Partial charges were derived using TPSSh/def2-TZVP electron densities and DDEC method;
- AP fingerprints (atom type, number of heavy-atom neighbors, distances, etc.) represent atoms;
- A **2D topological representation** eliminates conformational dependency and ensures transferability;
- It is limited to H, C, N, O, F, P, S, Cl, Br, I atoms;
- RF generates accurate predictions of unseen compounds.



Example 4: Atom-Path-descriptor (**APD**) [4]

- Using APD descriptors (**2 and 3D information**) to predict partial atomic charges;
- Atoms are represented by multiple layers, with each layer encoding information such as atom type, distance, bond type, and so on.

Element	RMSE [e*10⁻³]		
	[3]	APD-RF	APD-XGB
H	93	72	63
P	685	561	372
C	220	142	114

Bioinformatics **2020**, *36*, 4721.

Example 5: Multiple-Choice Knapsack Problem (**ε-MCKP**) [5]

- Molecules are considered as **graphs** (G = (V , E, t));
- Create a database with a record for each isomorphism class in the subgraph list for each atom;
- Each isomorphism class's partial charges are condensed into a histogram.