Supporting Information

Improving the Silicon Interactions of GFN-xTB

Leonid Komissarov and Toon Verstraelen*

Center for Molecular Modeling (CMM), Ghent University, Technologiepark-Zwijnaarde 46, B-9052, Ghent, Belgium

E-mail: toon.verstraelen@ugent.be

S1. Extra display items

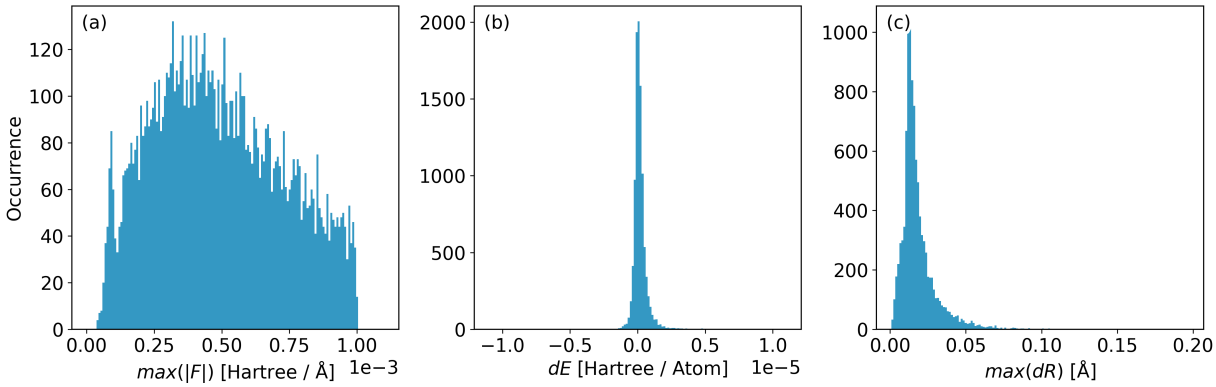

Figure S1: Distribution of convergence criteria at the last optimization step for all calculated systems in the reference data set. Showing (a) the highest absolute component of all nuclear gradients, (b) change in system energy and (c) highest relative atomic displacement.

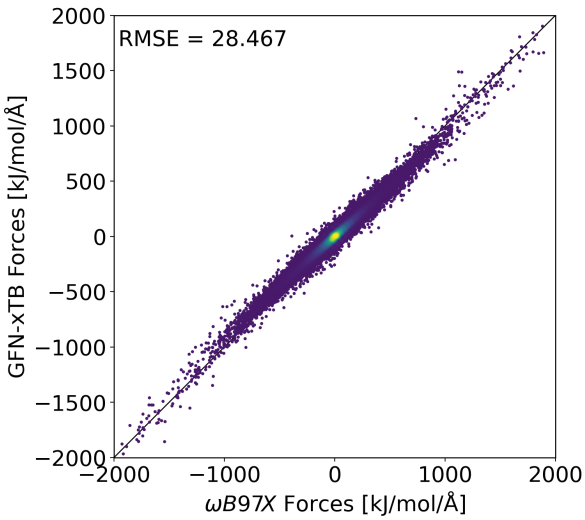

Figure S2: Correlation plot of atomic forces, as calculated with the test set generated from ANI-1x[1] data ($\omega WB97X$[2] reference, x-axis). As no silicon is present in this set, both, the GFN1-xTB and GFN1-xTB-Si parametrizations predict the same forces (y-axis).

Table S1: Format specification for the reference data repository. Each individual geometry optimization trajectory is stored in a NumPy .npz file with available keys listed below. Variables N and R denote the number of geometry optimization steps and the system size respectively.

Data	**Unit**	**Key**	**Array Shape**
Atomic Numbers	-	`numbers`	$(R,)$
Atomic Coordinates	Å	`xyz`	$(N, R, 3)$
Energy	hartree	`energy`	$(N,)$
Nuclear Gradients	hartree/bohr	`gradients`	$(N, R, , 3)$

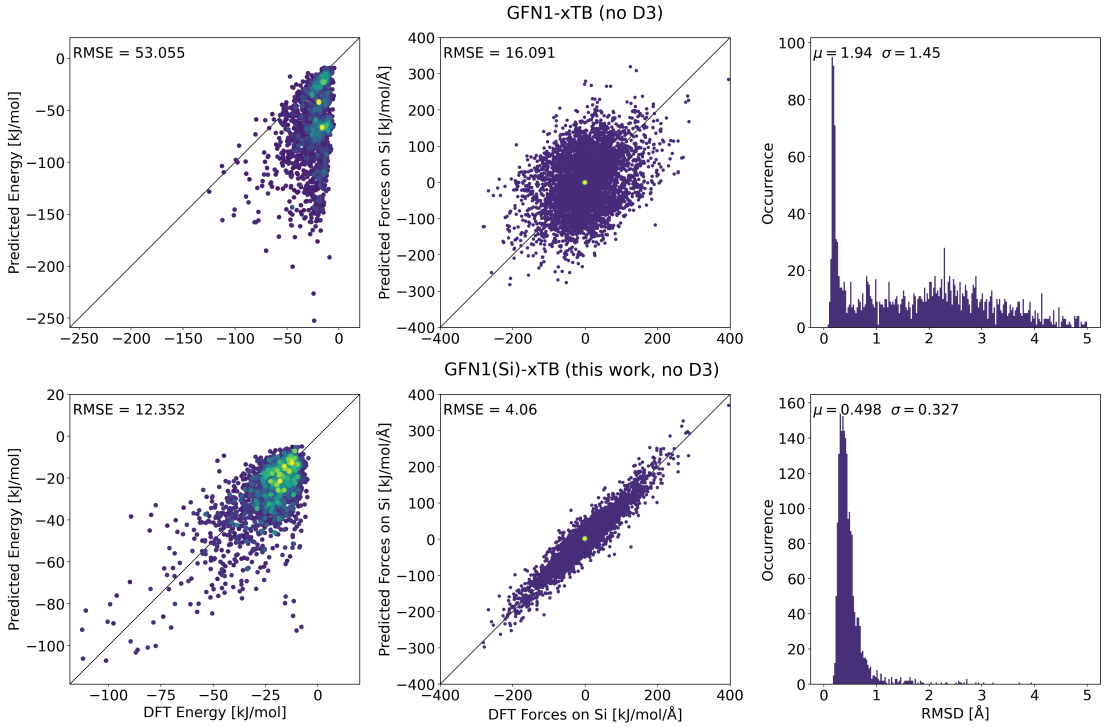

Figure S3: Validation set performance of GFN1-xTB[3] (top) and this GFN1(Si)-xTB (this work) (bottom) with D3-corrections disabled, showing comparable results to Fig. 2 in the main manuscript. Columns, from left to right depict energy differences, force components on the Si atoms, and RMSD of atomic positions (as described in the Methods Section). X and Y values in the first two columns are reference properties and their DFTB predictions respectively. Areas of lower point densities are depicted in dark blue; higher densities in bright green. Root-mean-square error (RMSE) printed in the same units as the axes. Histograms in the right column show the RMSD between geometry-optimized reference and DFTB structures. Mean μ and standard deviation σ printed in ångström.

References

(1) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x Data Sets, Coupled-Cluster and Density Functional Theory Properties for Molecules. *Sci. Data* **2020**, *7*, 1.

(2) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(3) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.