Supplementary Information:

The TensorMol0.1 Model Chemistry: a Neural Network Augmented with Long-Range Physics.

Kun Yao,[1] John E. Herr,[1] David W. Toth,[1] Ryker Mckintyre,[1] and John Parkhill[1, a)]

Dept. of Chemistry and Biochemistry, The University of Notre Dame du Lac

(Dated: 31 December 2017)

[a)]Electronic mail: jparkhil@nd.edu

I. Baseline

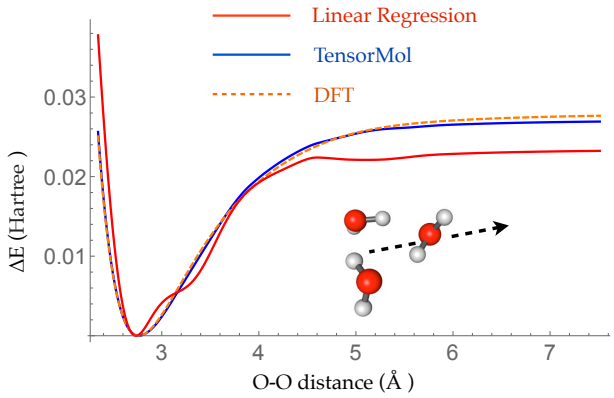

FIG. S1. Harmonic IR spectrum of morphine simulated by ωB97X-D/6-311G**(dashed orange line) and linear regression over symmetry functions (solid yellow line).

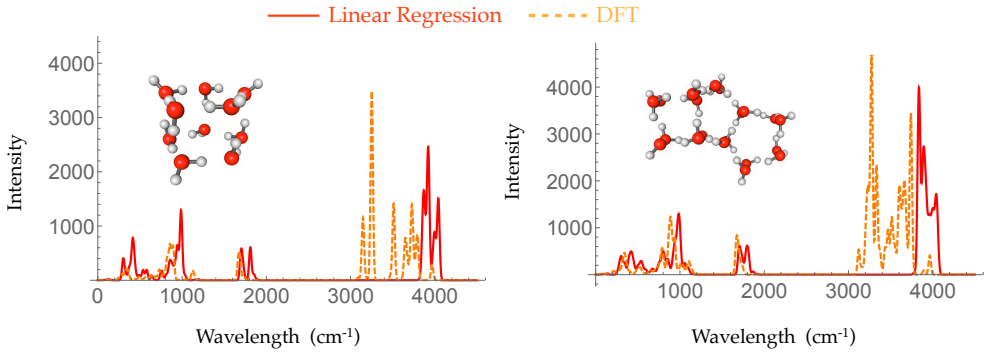

FIG. S2. Harmonic IR spectrum 10 water cluster (left panel) and 20 water cluster (right panel) simulated by ωB97X-D/6-311G**(dashed orange line) and linear regression over symmetry functions (solid red line) . The MAE of the linear regression frequencies is 132.5 cm^{-1} for 10 water cluster and 133.0 cm^{-1} for 20 water cluster

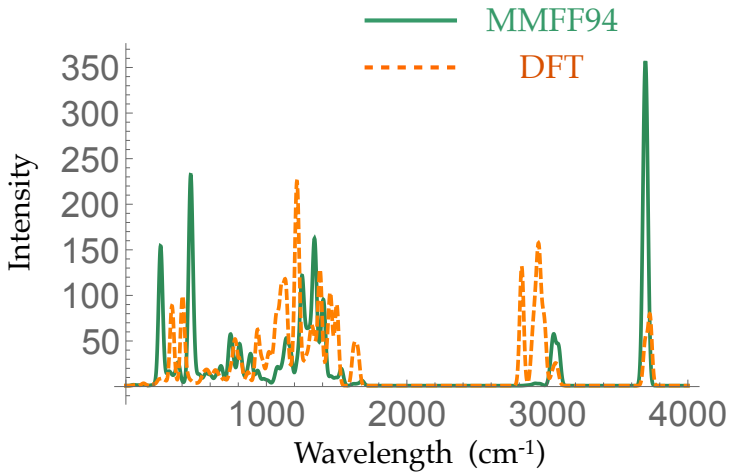

FIG. S3. Harmonic IR spectrum of morphine simulated by ωB97X-D/6-311G**(dashed orange line) and MMFF94 functions (solid green line). MMFF94 frequencies is calculated using RDKit[1].The DFT frequencies are scaled by a scaler factor of 0.957[2]. The MAE of MMFF94 frequencies is 28.4 cm^{-1}

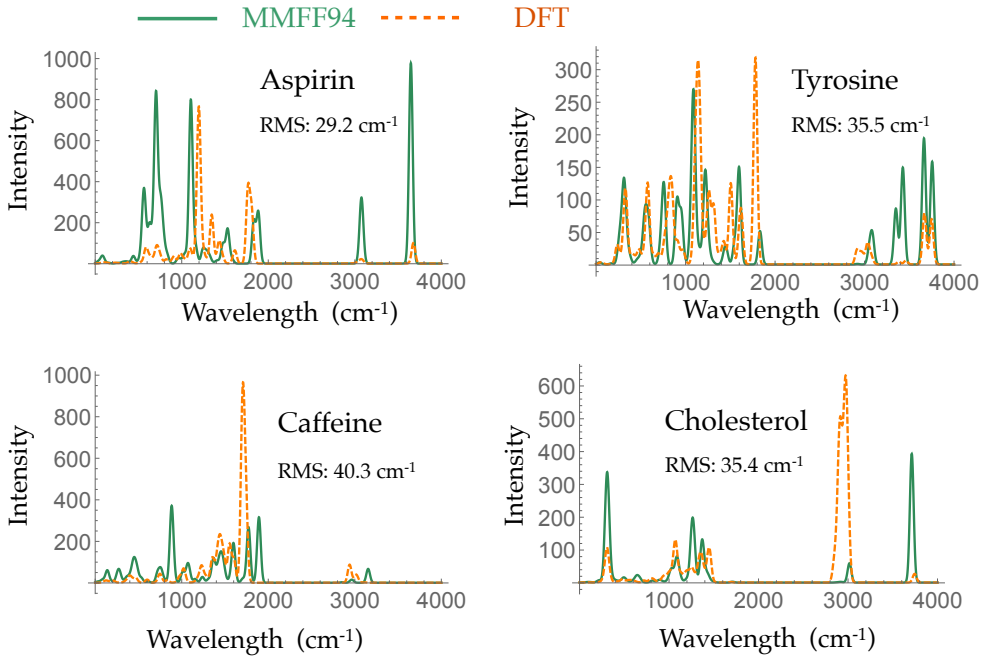

FIG. S4. Harmonic IR spectrum of four different molecules simulated by ωB97X-D/6-311G**(dashed orange line) and MMFF94 (solid green line). MMFF94 frequencies is calculated using RDKit[1]. The DFT frequencies are scaled by a scaler factor of 0.957[2].

II. Water hexamer cluster and geometry optimization with/without electrostatic energy

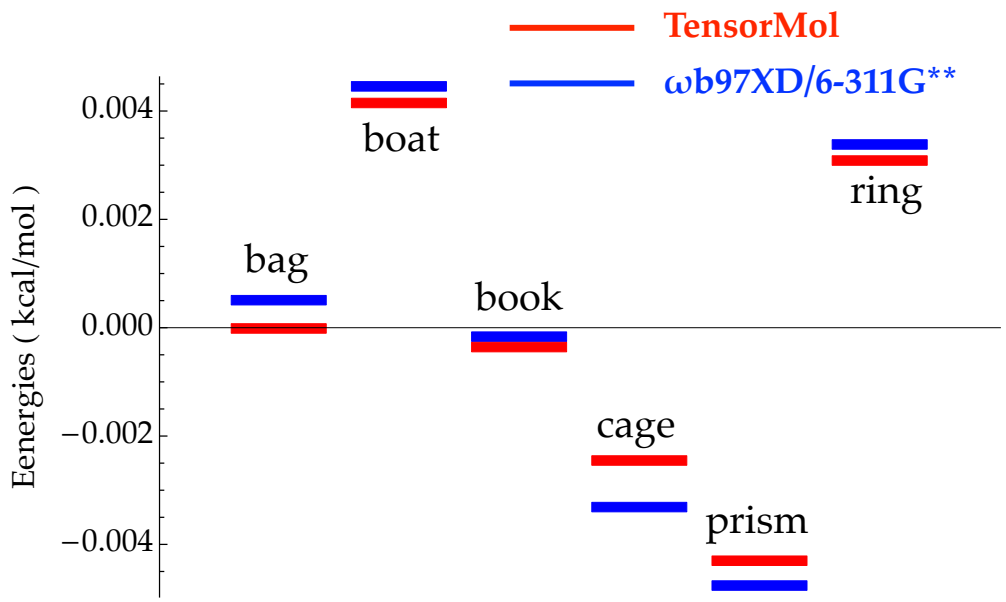

FIG. S5. Relative energies of different conformers of water hexamer cluster[3]. TensorMol predicts the same order of energies as the target method, ωB97X-D/6-311G** and the MAE of the relative energies predicted by TensorMol is 0.44 kcal/mol.

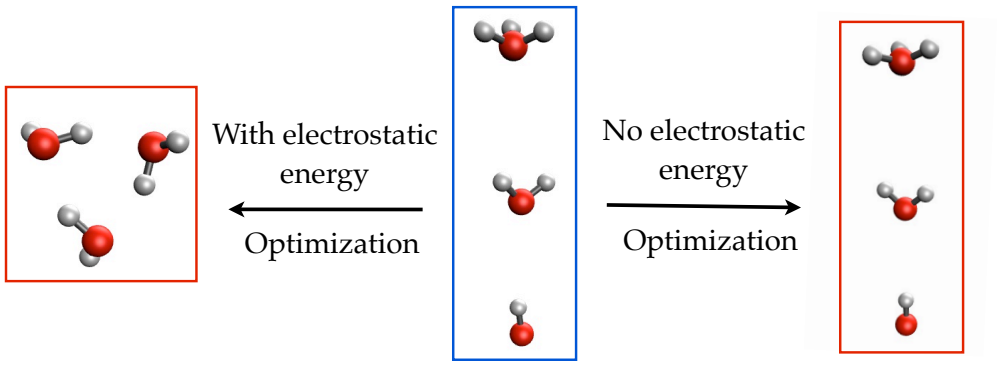

FIG. S6. Geometry optimization of a water trimmer that contains OH^- and H_3O^+. The neural network that includes electrostatic interaction successfully bring the cluster to the correct global minimal while the one that does not consider electrostatic interaction is stuck at local minimal because of the lack of long range interaction.

III. Hyperparamters Search

TABLE S1. Test RMSE of each learning target for water networks trained with different hyperparameters.The unit of energy RMSE, gradient RMSE and dipole RMSE is kcal/mol per atom,kcal/mol/Å per atom and Debye per atom, respectively

Hidden layers	Neurons per layer	Act. function	Energy	Gradient	Dipole
3	500	Softplus ($\alpha = 100$)	0.054	0.49	0.0082
3	100	Softplus ($\alpha = 100$)	0.058	0.56	0.0090
3	200	Softplus ($\alpha = 100$)	0.059	0.52	0.0086
3	1000	Softplus ($\alpha = 100$)	0.066	0.48	0.0082
1	500	Softplus ($\alpha = 100$)	0.065	0.69	0.0086
2	500	Softplus ($\alpha = 100$)	0.093	0.54	0.0085
4	500	Softplus ($\alpha = 100$)	0.054	0.50	0.0083
3	500	Softplus ($\alpha = 10$)	0.089	0.80	0.0090
3	500	Softplus ($\alpha = 1$)	0.61	3.4	0.011
3	500	Tanh	0.24	1.1	0.010
3	500	Sigmoid	0.38	2.6	0.011
3	500	Guassian	0.075	0.66	0.0098

IV. Timing

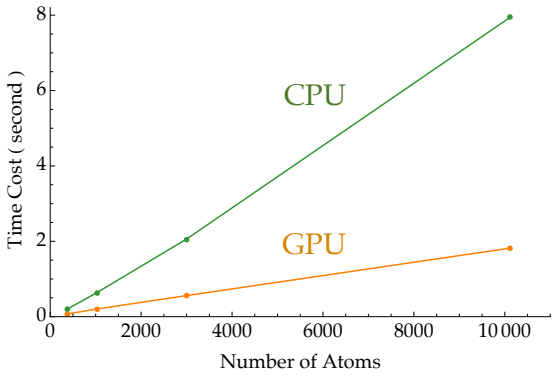

FIG. S7. Aperiodic timings of an energy, charge, force call for cubic water clusters at a density of 1 gm/cm without considering neighbor list building which is not implemented in GPU yet. The GPU timing is measured on single Nvidia K40 CPU and CPU timing is measure on a 8 thread Intel Xeon CPU E5-1620 v2. The GPU is 3 times to 4 times faster than CPU per call.

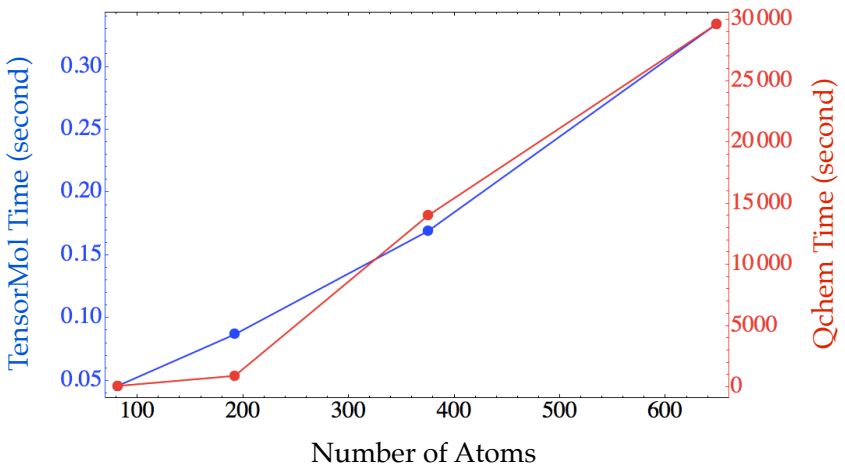

FIG. S8. Timing of TensorMol force field and Qchem for different size of water clusters. The GPU timing is measured on single Nvidia K40 CPU and CPU timing is measure on two 16 thread Intel Xeon CPU E5-2667 v4.

V. Datasets

Both water datasets and chemspider datasets are avaiable at https://drive.google.com/drive/folders/1IfWPs7i5kfmErIRyuhGv95dSVtNFo0e_?usp=sharing. To load the datasets, copy the *.pdb file to the "./datasets" folder, using following script to view the datasets.

```
a = MSet("chemspider12_clean_maxatom35") # for chemspider network
#a = MSet("H2O_wb97xd_1to21_with_prontonated") # for water network
a.Load()
for mol in a.mols:
    print (mol.coords, mol.atoms, mol.properties)
```

VI. Procedures of using TensorMol0.1

- Install Tensorflow, Python.
- Download TensorMol from https://github.com/jparkhill/TensorMol. Checkout master branch. Download the trained water network and chemspider network from https://drive.google.com/drive/folders/1IfWPs7i5kfmErIRyuhGv95dSVtNFo0e_?usp=sharing. Copy the trained networks (network.tar.bz2) into TensorMol folder. Unzip trained networks.

- Copy the test script test_tensormol01.py in folder "./samples" to into the Tensormol folder. Run the script test the geometry optimization, molecular dynamic, harmonic IR spectrum and realtime IR spectrum.

- Demo of training a neural network force field using TensorMol: Copy the training script training_sample.pyinto the tensormol folder. Run the script. This will train a network force field for water.

VII. XYZ of Morphine and 10 water cluster

```
40
Comment: morphine
C   -0.26128864264  -2.75777726029  0.853585170695
C   -1.5731577245  -2.9331214008  0.438597292241
C   -2.28783536382  -1.90359422499  -0.174056162063
C   -1.60675107221  -0.720684675887  -0.397707544461
C   -1.05885310175  1.45422839253  -0.87996473547
C   -1.37934163468  2.51495028937  0.18555012868
C   -0.994593859686  2.0096024107  1.56217625666
C   0.207845084401  1.4778667849  1.75826057187
C   2.23312771495  0.21623088604  0.843421285889
C   1.75697954809  -1.2106173412  1.19755022153
C   0.405861195872  -1.55167473654  0.628798710766
C   -0.264287220653  -0.639282633463  -0.153523501842
C   0.260061289681  0.681799321679  -0.587581030167
C   1.09631637014  1.22119060851  0.574973727893
C   1.14465054951  0.602799716836  -1.82847073625
C   2.42196487767  -0.1950960018  -1.55504427211
C   4.41665573803  -0.301410817383  -0.163494885385
O   -3.57784584995  -2.08142645743  -0.509105514634
O   -2.7261830319  2.88856809595  0.1105939134
O   -2.13654226164  0.458725543352  -0.8204936497
N   3.08636691844  0.250896158649  -0.337245552508
H   0.230321332519  -3.57422440962  1.36953282915
```

```
H   -2.07863974118  -3.87663230172  0.610582560046
H   -1.0856978507  1.91077278592  -1.86398278382
H   -0.779006714497  3.40909210108  -0.0304774509977
H   -1.69917069508  2.17224020205  2.36455233088
H   0.552920535637  1.23754632997  2.75341805043
H   2.82189961453  0.598923188177  1.67726907999
H   2.48045519306  -1.93243664483  0.813520005582
H   1.74042618376  -1.35446085791  2.28267687069
H   1.57670373336  2.15015129495  0.247450380732
H   0.591190495438  0.157195222901  -2.65425989337
H   1.4109229804  1.62795751438  -2.10674556104
H   2.18868725402  -1.27195036279  -1.50715561626
H   3.1147653044  -0.0621217571847  -2.38767702983
H   5.04635367629  0.012651048409  -1.00011486156
H   4.43727603061  -1.40116877603  -0.128010221059
H   4.86256552386  0.0833435016803  0.75675528593
H   -3.9029405557  -1.26425773557  -0.895167792724
H   -3.22326677646  2.07816903532  -0.0396507665797


30
Comment: water_10
O   -0.803145962132  -2.32393801483  1.16375849822
H   -1.18462906878  -3.05755112744  1.64596595709
H   -1.41094425413  -1.55728862437  1.26122589388
O   1.76316446851  -1.35483999607  1.35848784018
H   0.883056024321  -1.76283112263  1.42152122382
H   1.98700707921  -1.40838429497  0.413981932003
O   -2.31576361507  -0.152161413299  1.25703925349
H   -3.14787224055  0.0115259283642  1.70329279532
H   -1.76855977549  0.678653068469  1.36990820607
O   1.66415663864  1.21932807468  1.67708251707
```

```
H   1.69501152689   0.22471780342   1.62689913853
H   2.29885848428   1.480527248   2.34465546867
O   2.05553942977   -1.01547237817   -1.36304051433
H   1.97721657597   -0.0232968652168   -1.32734021811
H   2.81972137334   -1.21024778461   -1.9059363719
O   1.86755697058   1.56082730517   -1.07012502147
H   0.966853805132   1.86072013507   -1.28970628174
H   1.91324535184   1.62958645215   -0.102525288322
O   -2.07800279215   -0.0191239847135   -1.56421376894
H   -2.36940060924   -0.160982444156   -0.65135902534
H   -1.49411323095   -0.787603919374   -1.74207555598
O   -0.765179898359   2.19883402485   -1.34369372442
H   -1.14310672039   2.85562540923   -1.92879723686
H   -1.26380858081   1.3492439109   -1.50371897263
O   -0.923280792683   2.0100619437   1.40499510393
H   -0.000391985469245   1.799684217   1.63705815974
H   -0.88025306128   2.28941902312   0.47480960141
O   -0.431194827639   -2.15105598676   -1.62596938578
H   0.474155852145   -1.79700142427   -1.66132649704
H   -0.544798481692   -2.41713915834   -0.701790997279
```

References

[1]P. Tosco, N. Stiefl, and G. Landrum, J. Cheminform. **6**, 37 (2014).

[2]"Nist precomputed vibrational scaling factors," http://cccbdb.nist.gov/vibscale.asp, accessed: 2017-12-23.

[3]E. E. Dahlke, R. M. Olson, H. R. Leverentz, and D. G. Truhlar, The Journal of Physical Chemistry A **112**, 3976 (2008).