UNC
ESHELMAN
SCHOOL OF PHARMACY

Artificial Intelligence (AI) Solutions for Computational Chemistry & Organic Chemistry

@olexandr

Olexandr Isayev
University of North Carolina at Chapel Hill
olexandr@olexandrisayev.com
http://olexandrisayev.com

Artificial Intelligence (AI) Solutions for Computational Chemistry & Organic Chemistry

@olexandr

Olexandr Isayev
Department of Chemistry, Carnegie Mellon University
olexandr@olexandrisayev.com
http://olexandrisayev.com

January 2020 Stay tuned for even more exciting science!

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Mariya Popova
Roman Zubatyuk
Daniel Korn
Kyle Bowler
Hatice Gockan

Adrian Roitberg
Justin S. Smith
Christian Devereux
Kavindri Ranasinghe

UF UNIVERSITY of FLORIDA

Funding:

DEPARTMENT OF THE NAVY ONR Science & Technology

NSF

CHE-1802789

UNC ESHELMAN INSTITUTE FOR INNOVATION

Collaborators:

Nicholas Lubbers
Ben Nebgen
Andrew Sifain
Kipton Barros
Sergei Tretiak

Center for Integrated Nanotechnologies CINT

Center for Nonlinear Studies

HPC Computing:

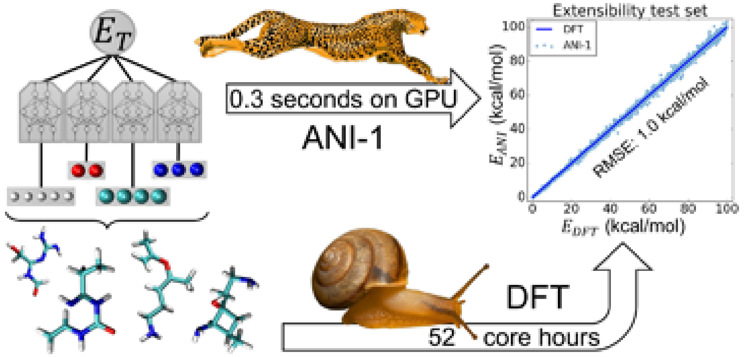
XSEDE Extreme Science and Engineering Discovery Environment

Los Alamos NATIONAL LABORATORY EST. 1943

Open Science Grid

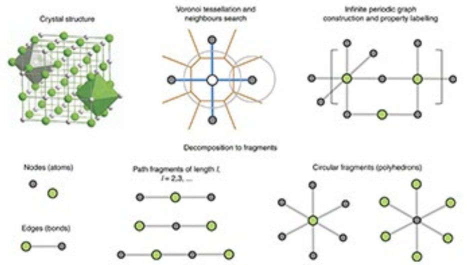
NVIDIA

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY



J. Chem. Phys. **2018**, 148, 241733

Chem. Sci., **2017**, 8, 3192-3203



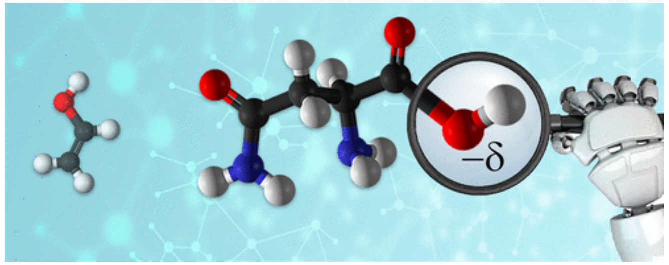
Crystal structure
Voronoi tessellation and neighbours search
Infinite periodic graph construction and property labelling
Decomposition to fragments
Nodes (atoms)
Edges (bonds)
Path fragments of length *l*, *l* = 2,3, ...
Circular fragments (polyhedrons)

Nature Commun. **2017**, 8, 15679

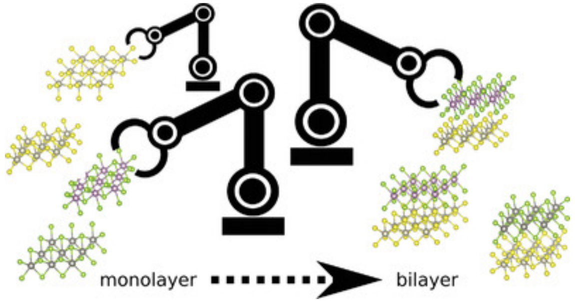
Comp. Mater. Sci., 152, **2018**, 134-145

AFLOW

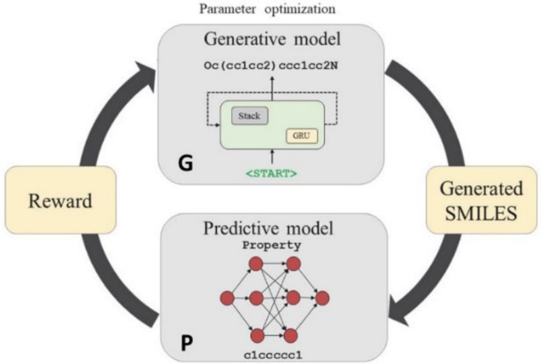
Automatic - FLOW for Materials Discovery



J. Phys. Chem. Lett., **2018**, 9 (16), pp 4495–4501

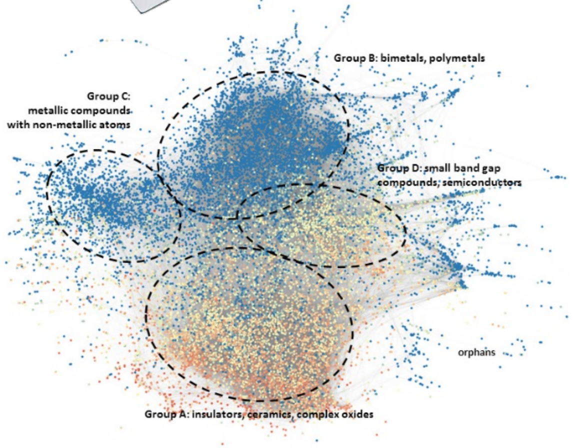


Adv. Theory Simul., **2019**, 2: 1800128

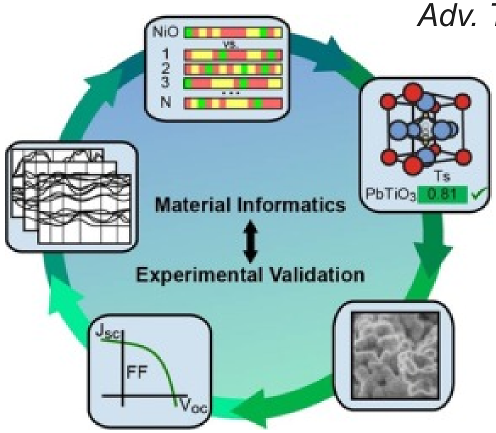


ACS Med. Chem. Lett. **2018**. 9, 1065–1069

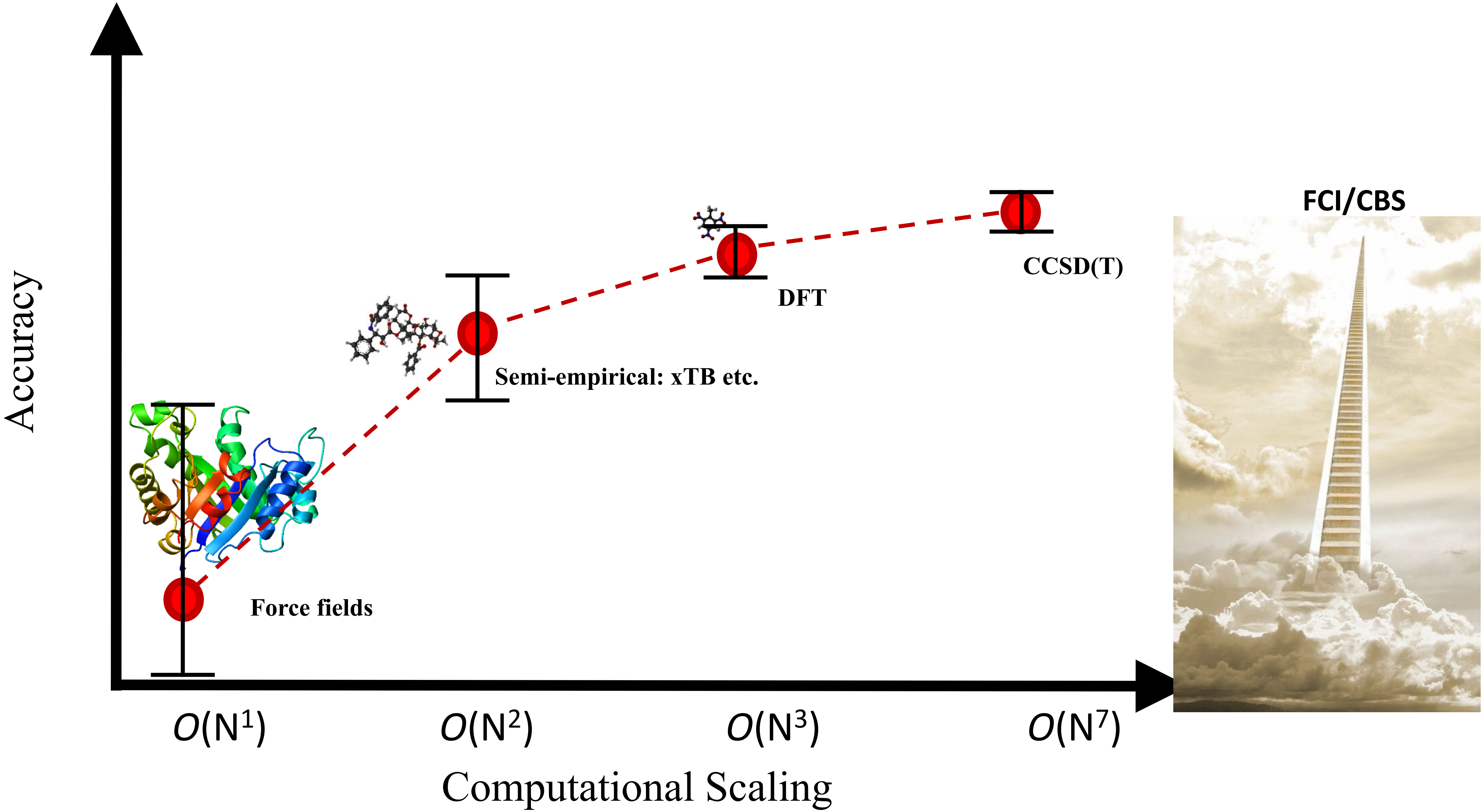
Science Advances, **2018**, 4 (7) ,eaap7885



Chem. Mater., **2015**, 27, 735-742.



Materials Discovery. **2017**, 6*, 9-16*

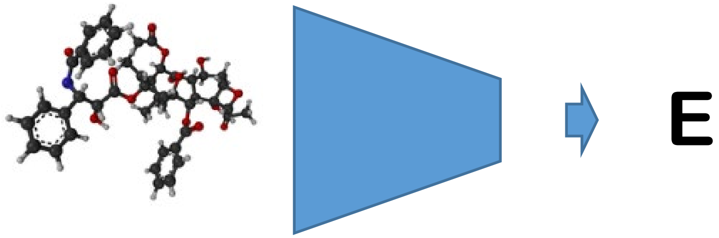

Accuracy
FCI/CBS
CCSD(T)
DFT
Semi-empirical: xTB etc.
Force fields
$O(\mathrm{N}^1)$
$O(\mathrm{N}^2)$
$O(\mathrm{N}^3)$
$O(\mathrm{N}^7)$
Computational Scaling

Quantum Mechanics 101

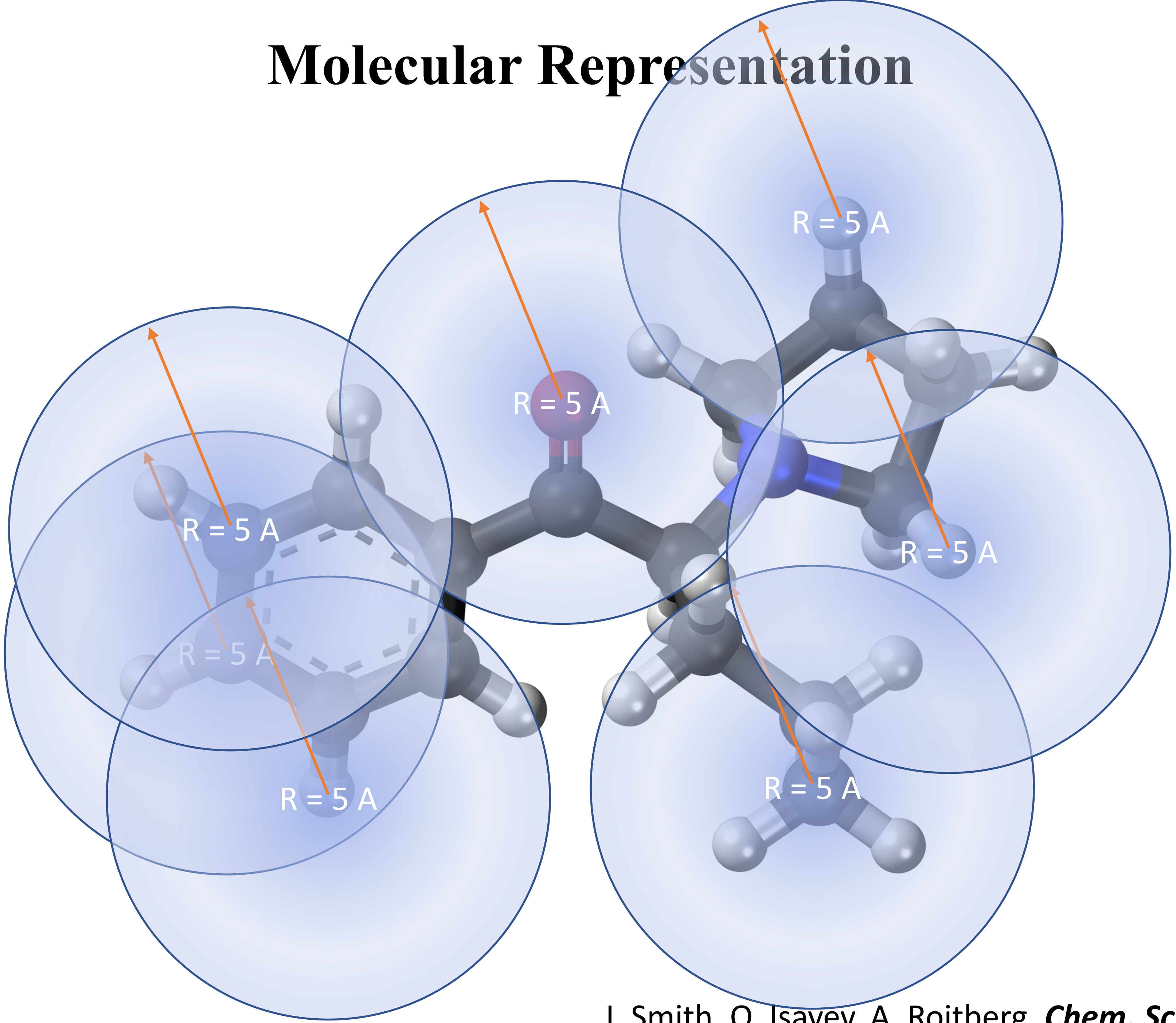
$$\hat{H}\psi = E\psi$$

The Schrodinger equation was discovered in 1926 by Erwin Schrodinger, an Austrian theoretical physicist.
It is an important equation that is fundamental to quantum mechanics.

$$E = f(R_{vector})$$



Molecular Representation



J. Smith, O. Isayev, A. Roitberg. ***Chem. Sci.***, 2017, **8**, 3192-3203

Emergence of ‘hybrid’ ML/NN force field

We use mostly DFT as a reference QM!

ANI-1: E = E(NN) + E(vDW), vDW = D2, D3, D3(BJ)

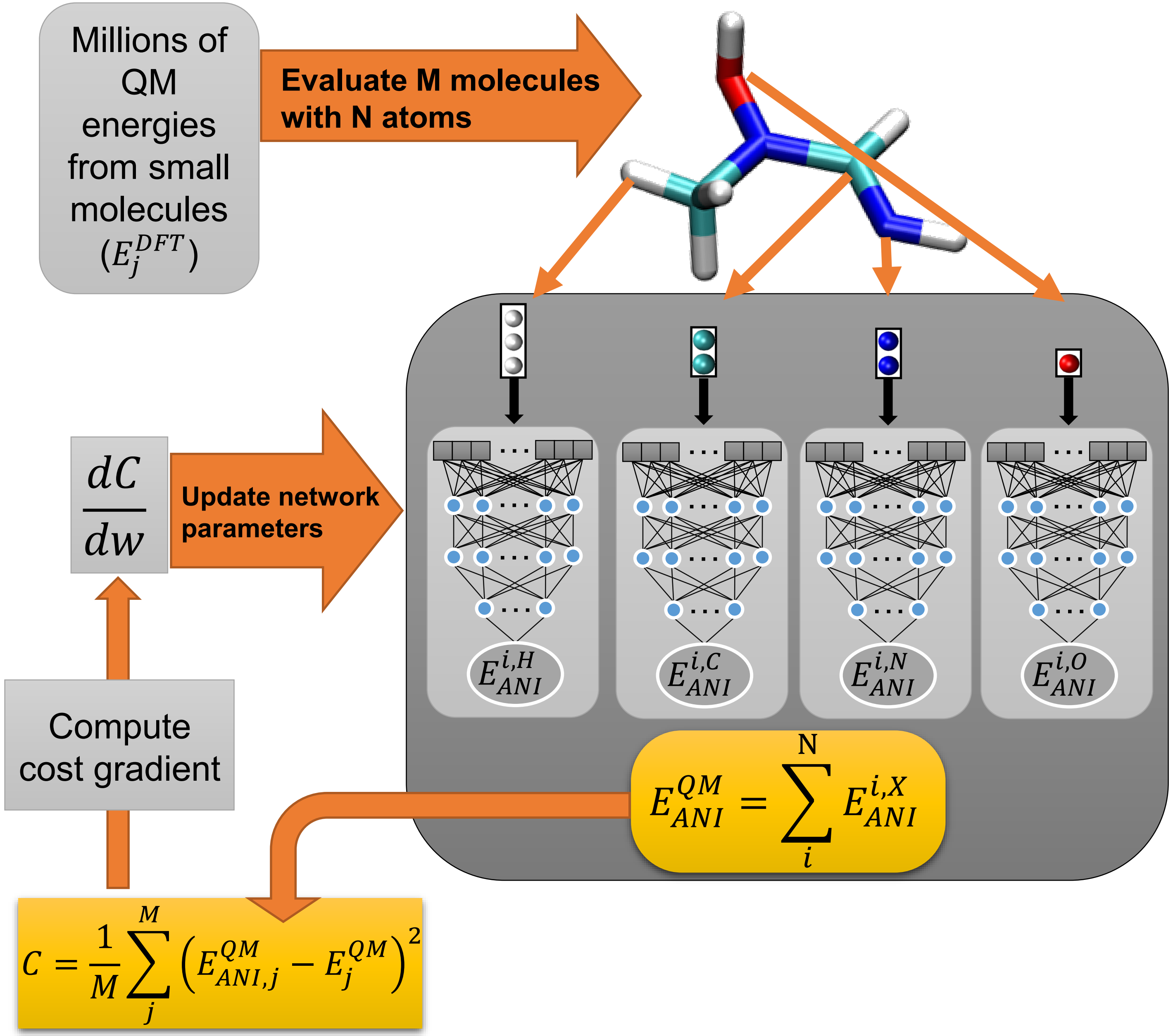
Now we could predict dynamic charges, volumes, C6 coefficients, etc.

ANI-2: E = E(NN) + E(vDW) + E(LR)

vDW = D3, D4, TS, MBD
LR = electrostatics, …

AIMNet: E = E(NN) Dispersion & LR are implicit

Neural Network molecular potential - training



Millions of QM energies from small molecules (E_j^{DFT})

Evaluate M molecules with N atoms

$\frac{dC}{dw}$

Update network parameters

Compute cost gradient

$$E_{ANI}^{QM} = \sum_{i}^{N} E_{ANI}^{i,X}$$

$$C = \frac{1}{M}\sum_{j}^{M}\left(E_{ANI,j}^{QM} - E_j^{QM}\right)^2$$

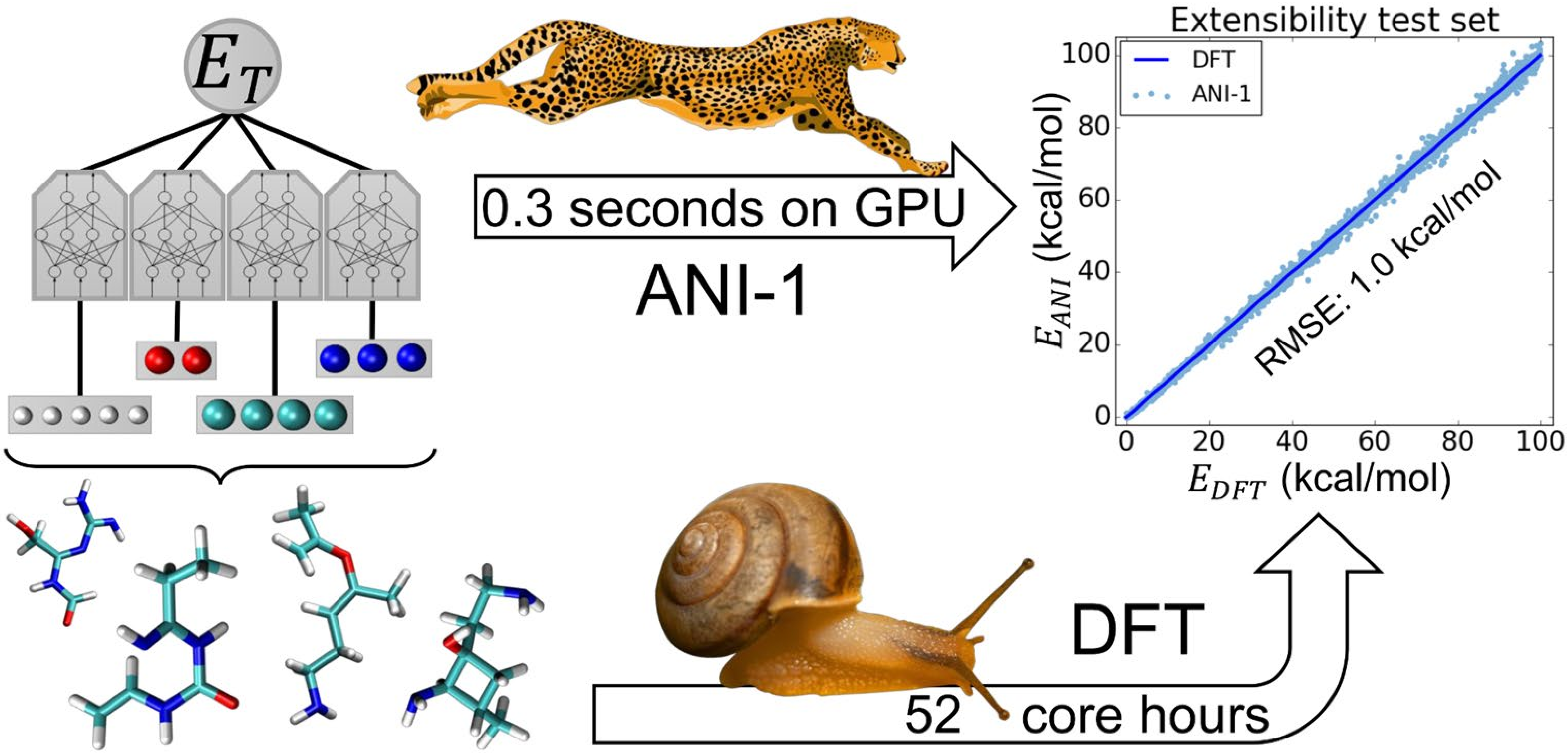
Currently available:
CHNOSFCl

P, Si, Br, I, Se, B …
in progress

2018:
ωB97x/DZ -> ωB97x/TZVPP

2019:
ωB97M/Def2-TZVPP and CCSD(T)*/CBS

ANI Deep Neural Network



0.3 seconds on GPU

ANI-1

DFT

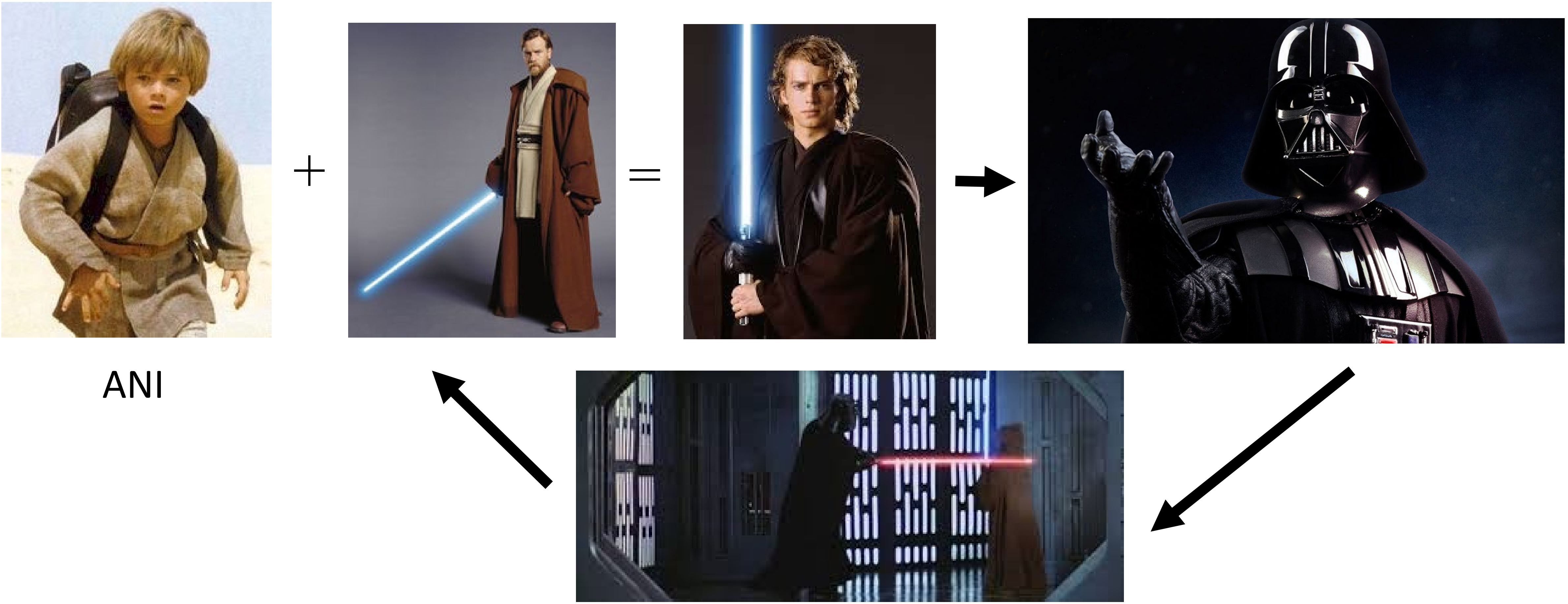
52 core hours

Chem. Sci., 2017, **8**, 3192-3203

ANAKIN-ME

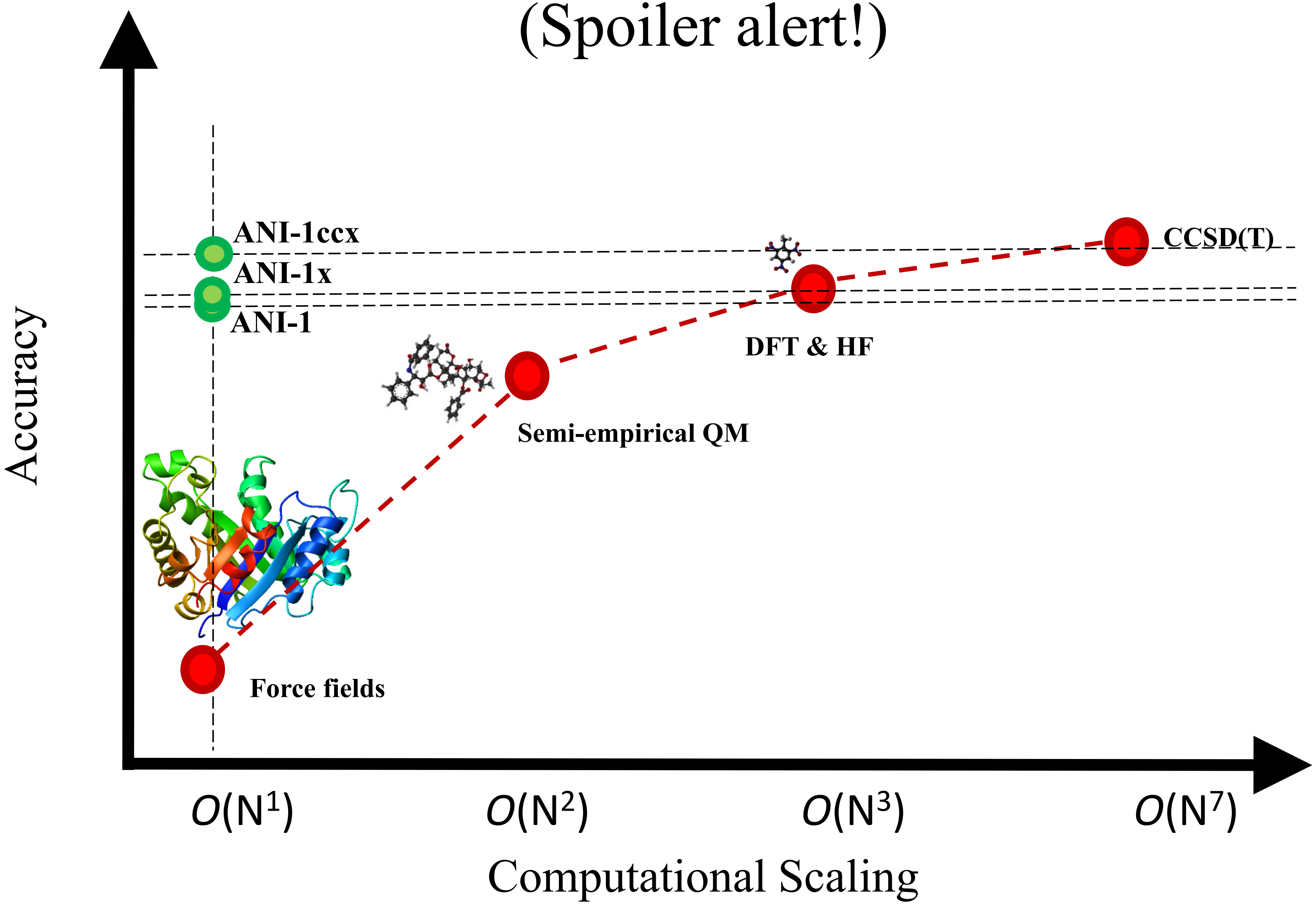
Accurate NeurAl networK engINe for Molecular Energies

We want to train a padawan network to become a DFT jedi master



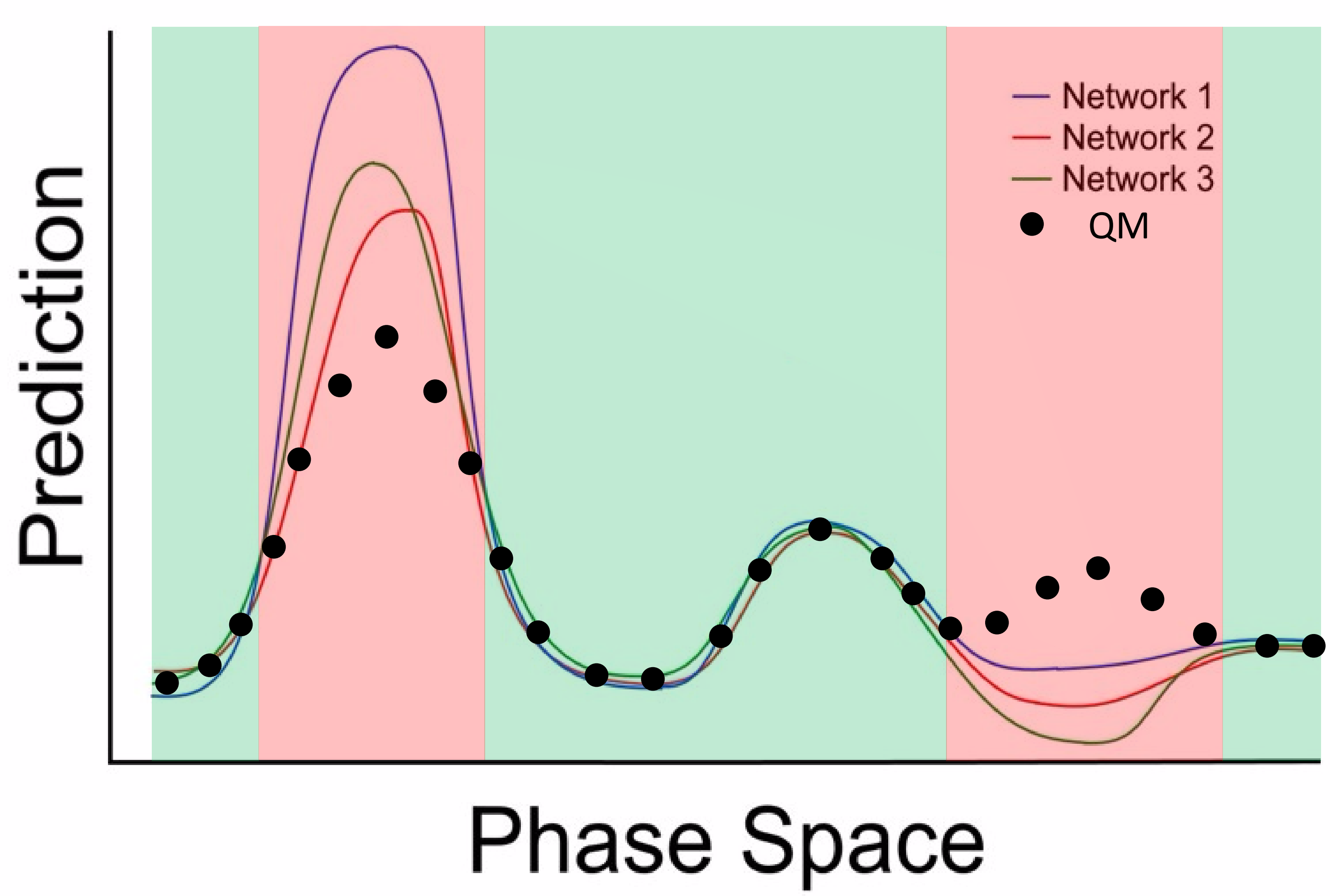
Where do we fit?

(Spoiler alert!)



Can we predict when the model is wrong?

Ensemble disagreement can drive data generation



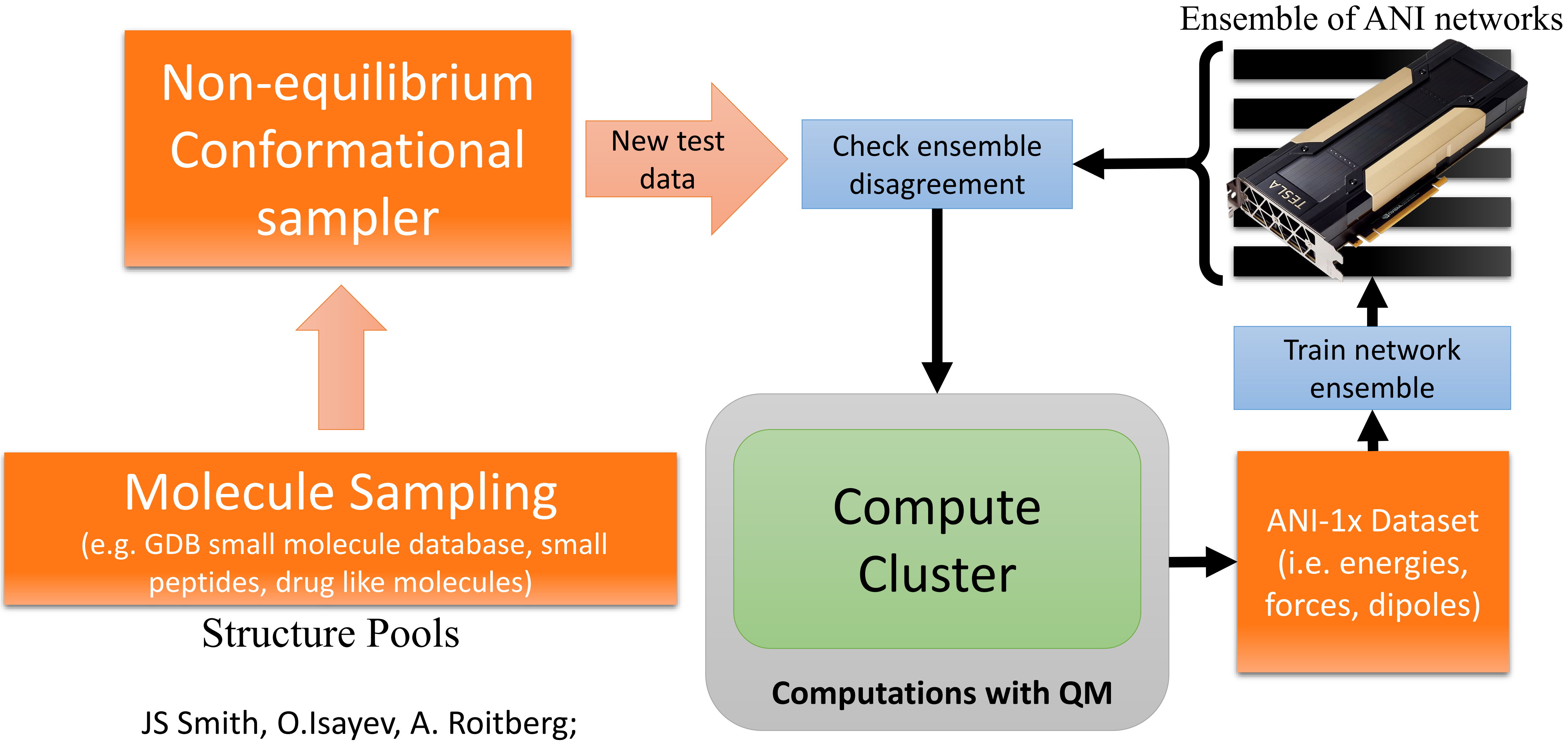
Good data coverage

Bad data coverage

JS Smith, O.Isayev, A. Roitberg; *Journal of Chemical Physics*, (**2018**), 148 (24), 241733

Active Learning - The Big Picture

An automated and self-consistent data generation framework



Ensemble of ANI networks

Non-equilibrium Conformational sampler

New test data

Check ensemble disagreement

Train network ensemble

Molecule Sampling

(e.g. GDB small molecule database, small peptides, drug like molecules)

Structure Pools

Compute Cluster

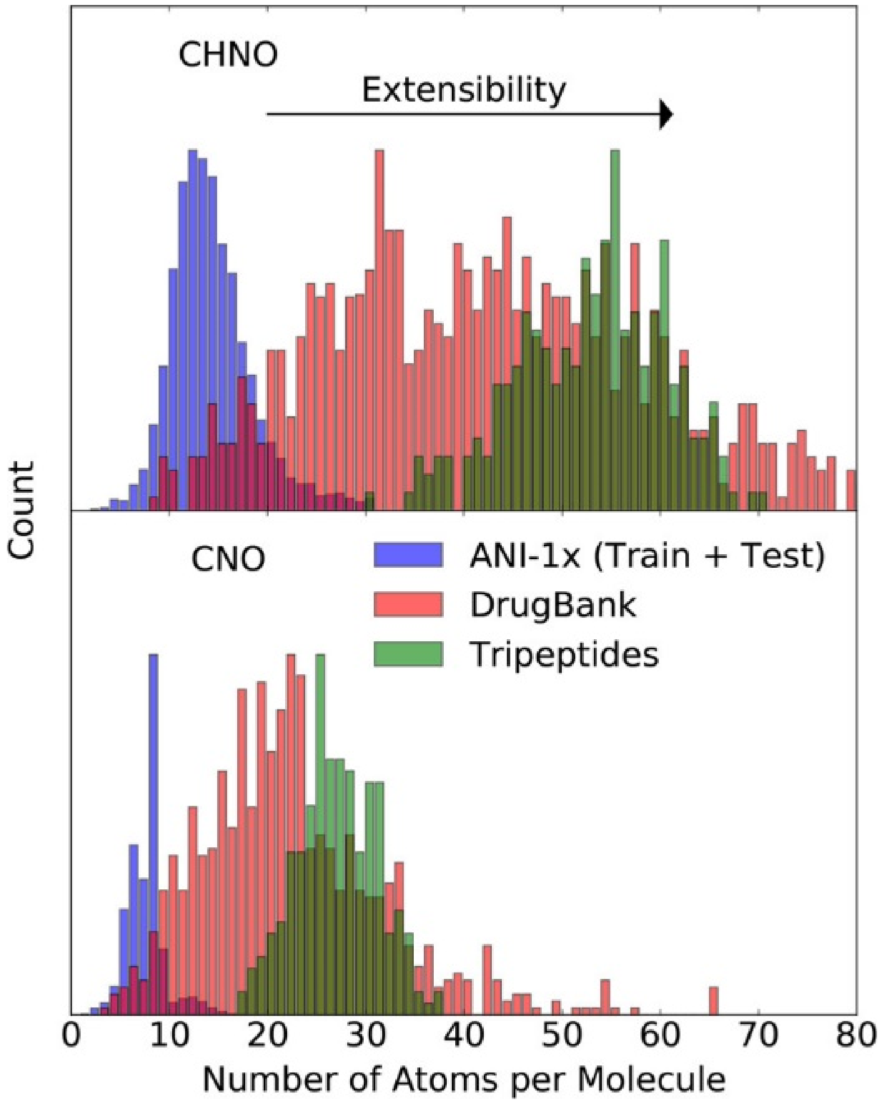
Computations with QM

ANI-1x Dataset (i.e. energies, forces, dipoles)

JS Smith, O.Isayev, A. Roitberg;
Journal of Chemical Physics, (**2018**), 148 (24), 241733

What do you need?

- ANI requires TONS of data
  - For ANI-1 we run ~20M DFT data points @ wB97x/DZ.
  - Available to anyone!
  - Molecules with 1 to 15 heavy atoms from various databases
  - Out-of-equilibrium geometry sampling with NMS, MD
- Train network on a fraction of available data, validate on independent data
- Test on 'known sizes' (Molecules with <= # max heavy atoms per molecule in training set)
  - Interpolation
- Test on 'unknown sizes' (Molecules larger than any in the training set)
  - Extrapolation



Datasets

- Original ANI-1 dataset (Soon to be Deprecated!!!)
  - Random sampling
  - 60K organic molecules, ~25M DFT datapoints

- ANI-1x (CHNO)
  - AL sampling
  - 5M DFT datapoints
  - 0.5M CCSD(T)/CBS

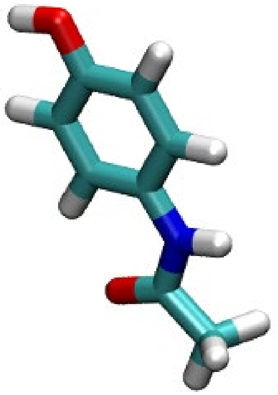
- ANI-1x (+SFCl)
  - AL sampling
  - 4M DFT datapoins
  - CCSD(T)/CBS is being computed now

ANI-1: *Sci. Data*, 2017, **4**, 170193 DOI: 10.1038/sdata.2017.193

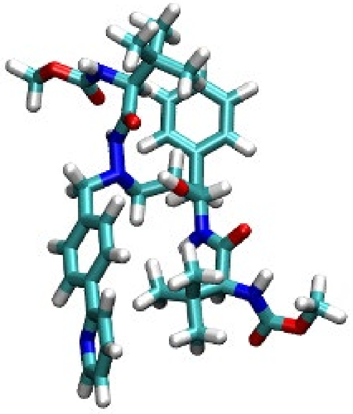
ANI Data set Python library
Available at: https://github.com/isayev/ANI1_dataset

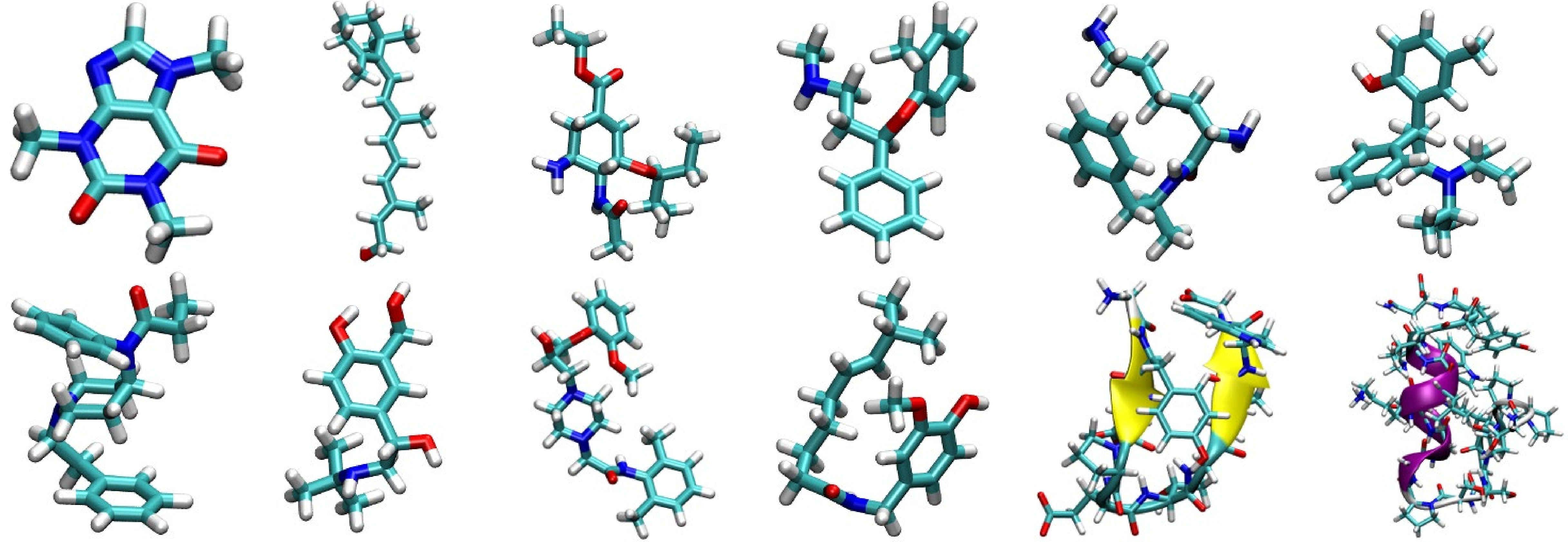
ANI-1x: To be released soon.

ANI-MD Benchmark // COMP6

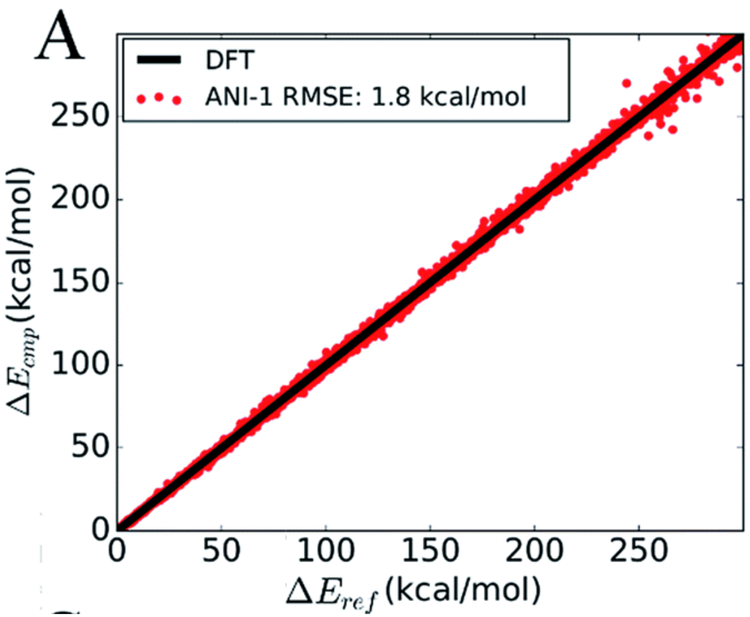


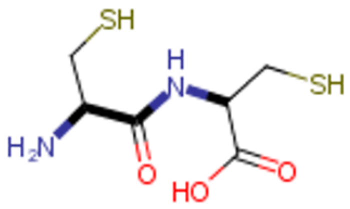
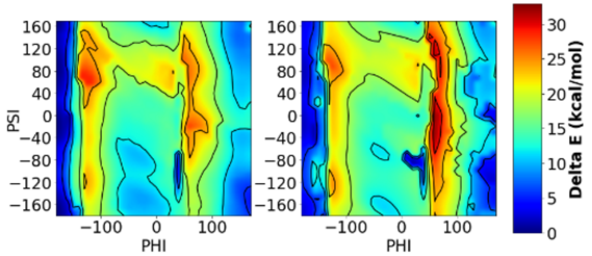
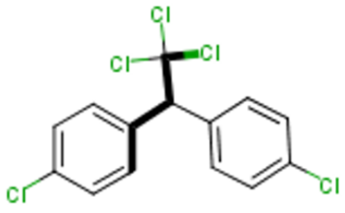
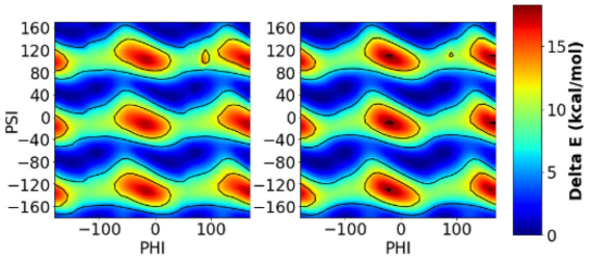
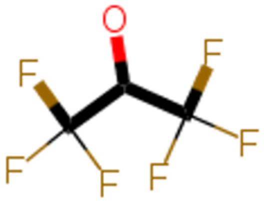
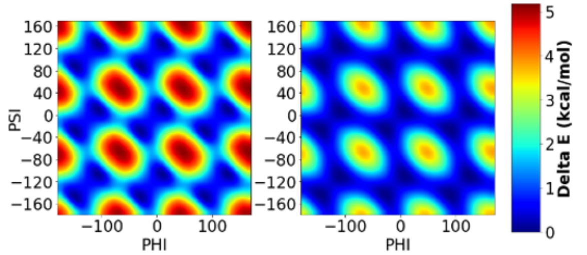
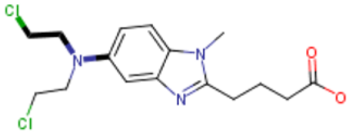
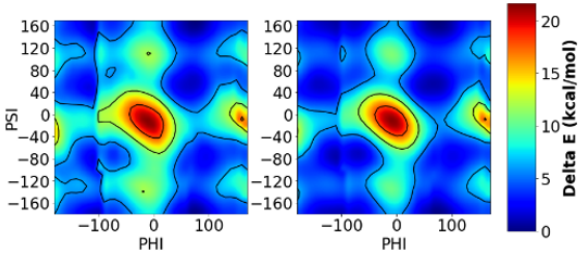
- 12 drug molecules and 2 proteins
- Mean size 75 atoms (max 312 atoms)
- 1ns of molecular dynamics (MD)
- Dynamics at 300K
- MD ran on ANI-1x potential
- 128 randomly sampled frames



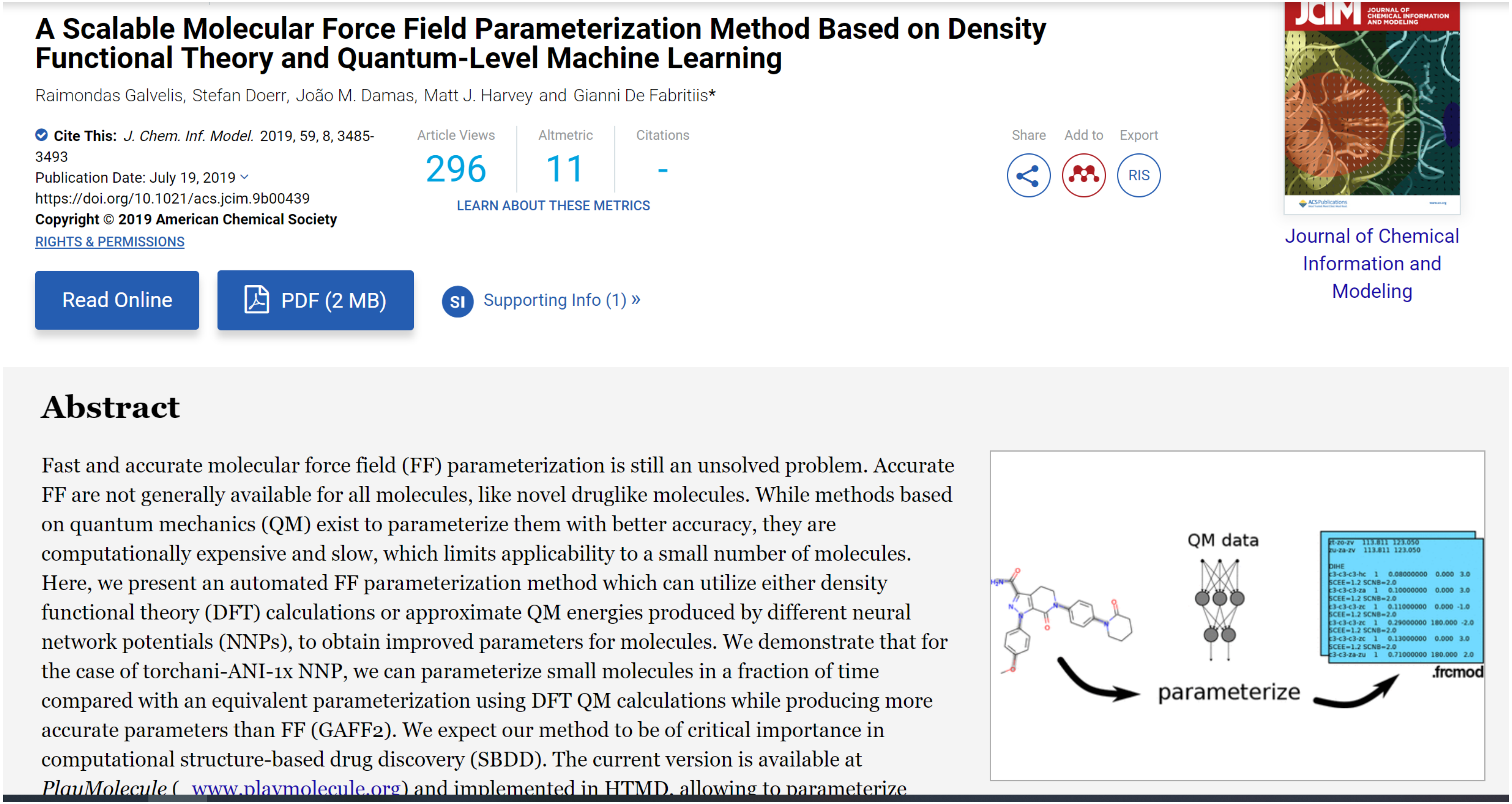


Accuracy of Energy & PES Prediction



Name	Molecule	MAE	RMSE	Scan (Left:ANI Right:DFT)
Cysteine-Dipeptide		2.18	2.96	
DDT		0.58	0.71	
Hexafluoroacetone		0.92	1.05	
Bendamustine		1.16	1.38	

Relaxed 2D torsion scans for ANI-2x (left) and DFT (right).



A Scalable Molecular Force Field Parameterization Method Based on Density Functional Theory and Quantum-Level Machine Learning

Raimondas Galvelis, Stefan Doerr, João M. Damas, Matt J. Harvey and Gianni De Fabritiis*

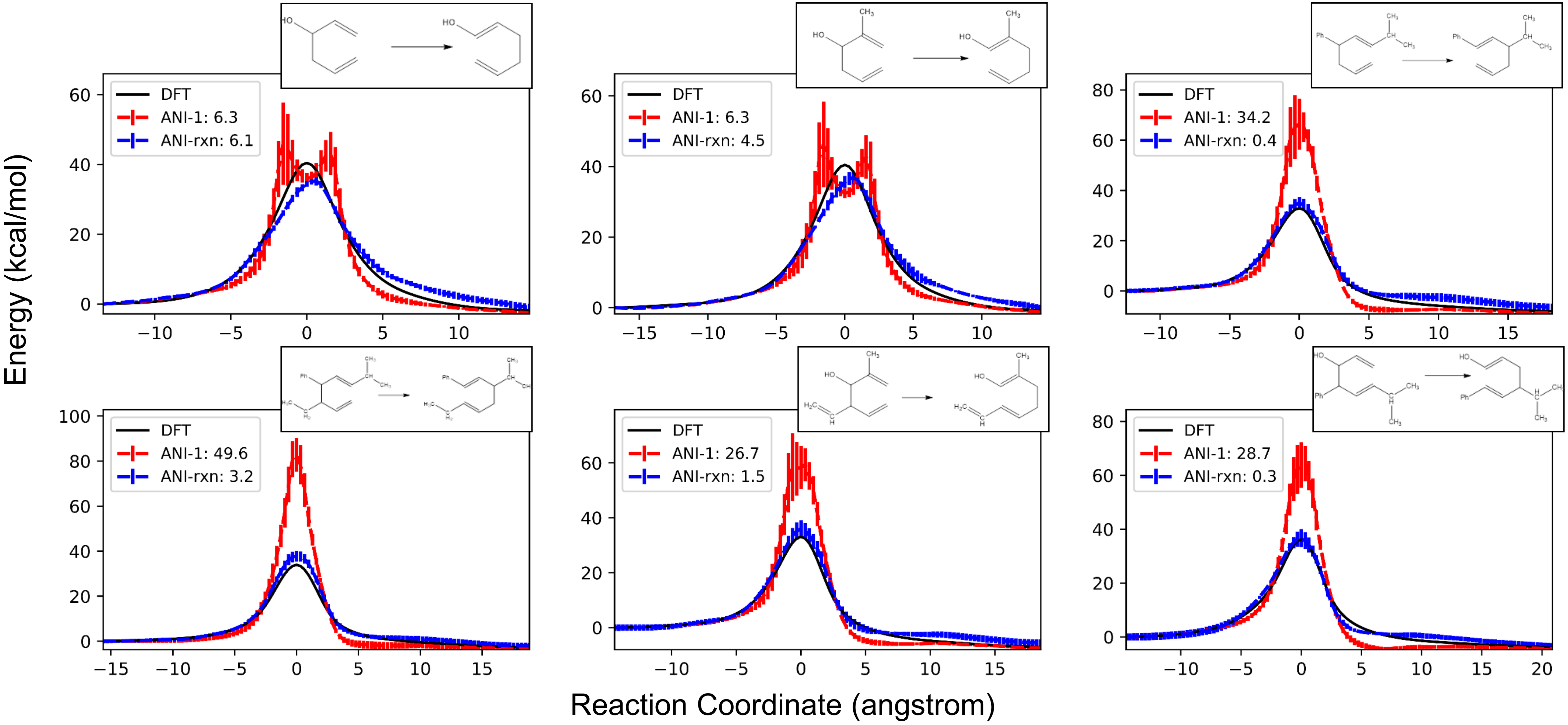
Cite This: *J. Chem. Inf. Model.* 2019, 59, 8, 3485-3493
Publication Date: July 19, 2019
https://doi.org/10.1021/acs.jcim.9b00439

Abstract

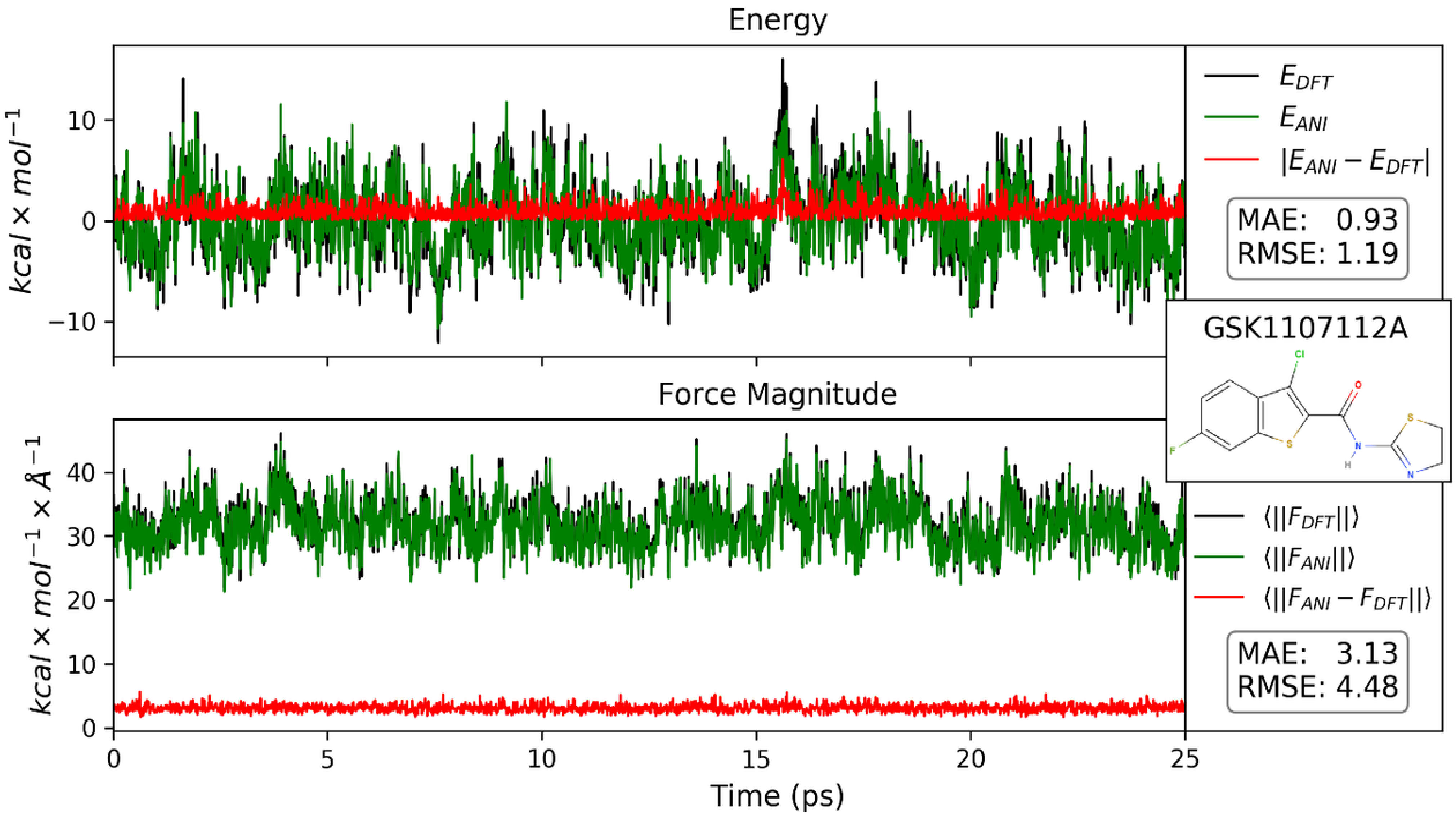
Fast and accurate molecular force field (FF) parameterization is still an unsolved problem. Accurate FF are not generally available for all molecules, like novel druglike molecules. While methods based on quantum mechanics (QM) exist to parameterize them with better accuracy, they are computationally expensive and slow, which limits applicability to a small number of molecules. Here, we present an automated FF parameterization method which can utilize either density functional theory (DFT) calculations or approximate QM energies produced by different neural network potentials (NNPs), to obtain improved parameters for molecules. We demonstrate that for the case of torchani-ANI-1x NNP, we can parameterize small molecules in a fraction of time compared with an equivalent parameterization using DFT QM calculations while producing more accurate parameters than FF (GAFF2). We expect our method to be of critical importance in computational structure-based drug discovery (SBDD). The current version is available at *PlayMolecule* (www.playmolecule.org) and implemented in HTMD, allowing to parameterize

```
zt-zo-zv   113.811  123.050
zu-za-zv   113.811  123.050

DIHE
c3-c3-c3-hc   1   0.08000000   0.000  3.0
SCEE=1.2 SCNB=2.0
c3-c3-c3-za   1   0.10000000   0.000  3.0
SCEE=1.2 SCNB=2.0
c3-c3-c3-zc   1   0.11000000   0.000 -1.0
SCEE=1.2 SCNB=2.0
c3-c3-c3-zc   1   0.29000000 180.000 -2.0
SCEE=1.2 SCNB=2.0
c3-c3-c3-zc   1   0.13000000   0.000  3.0
SCEE=1.2 SCNB=2.0
c3-c3-za-zu   1   0.71000000 180.000  2.0
```

Active-learning reactions : Cope rearrangement



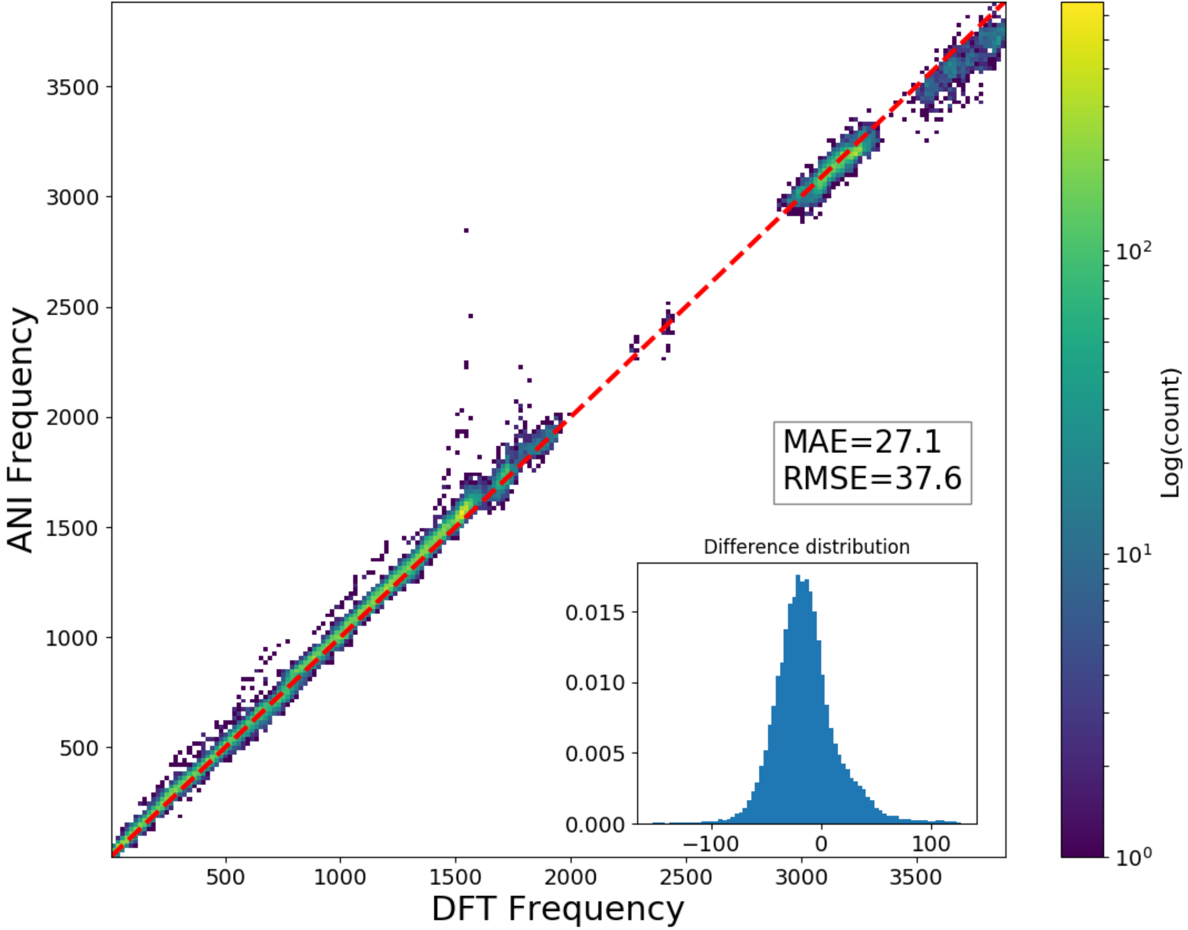
Accuracy of Molecular Dynamics



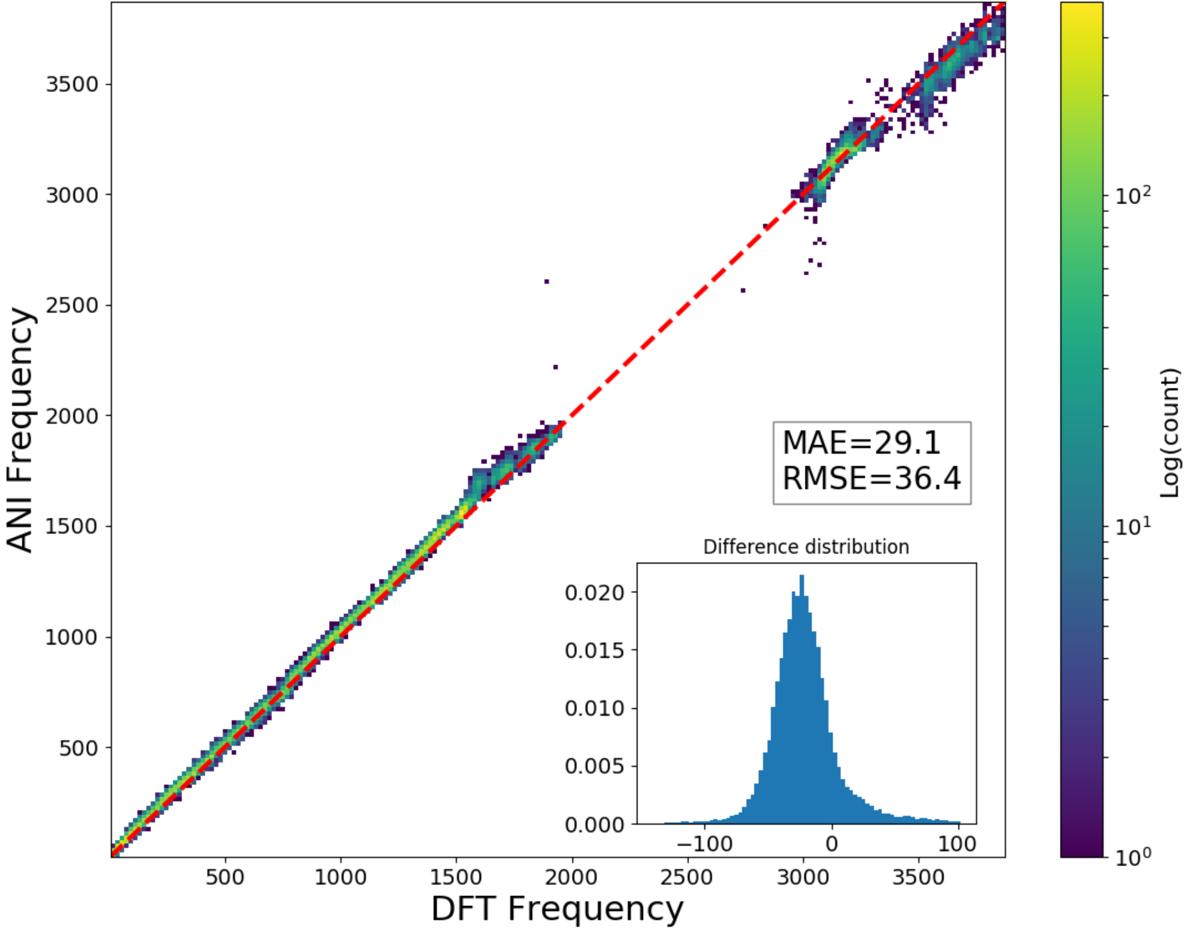
ANI-1x predicted harmonic frequencies

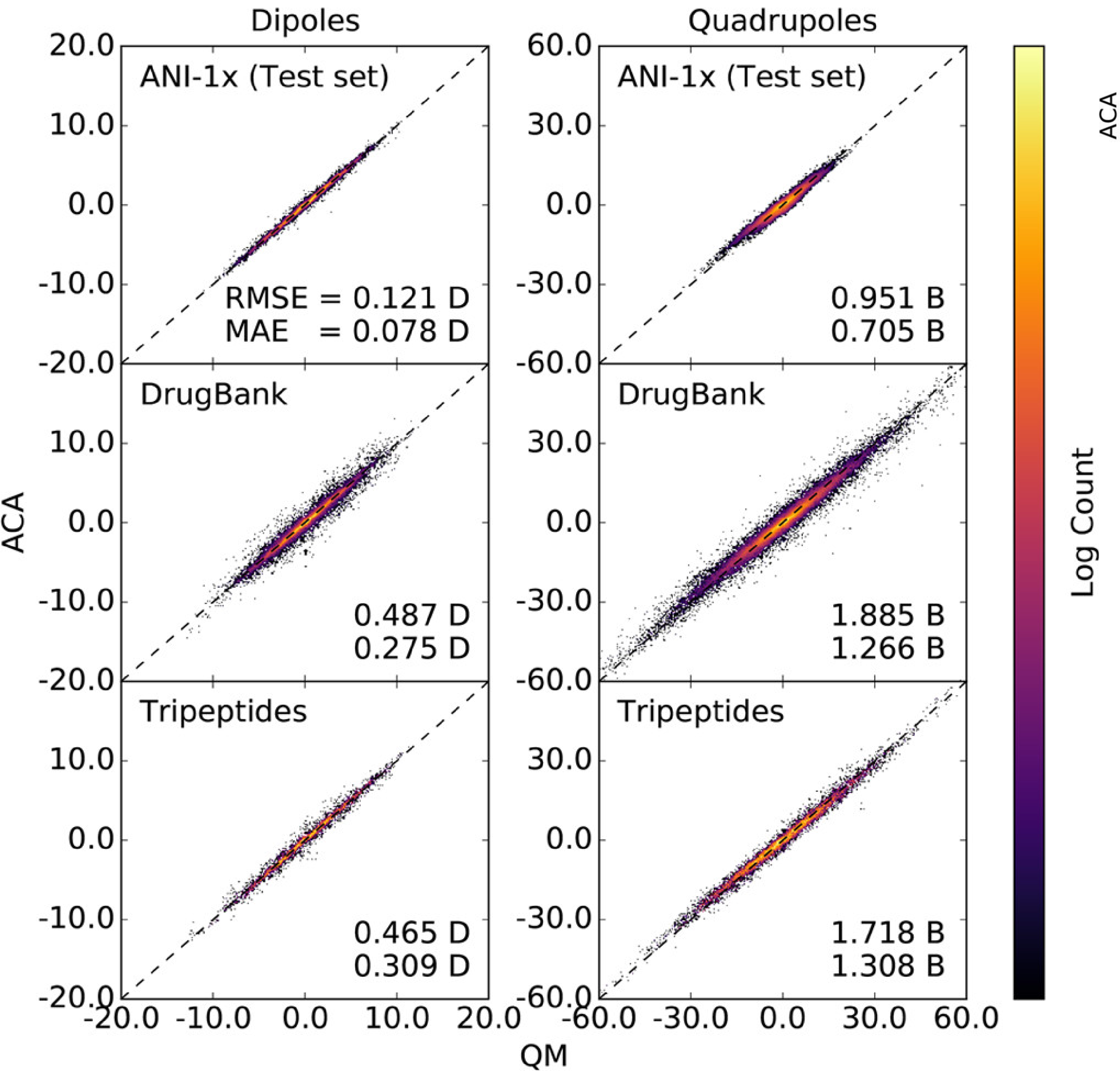
Work in progress with Christian Devereux @ UF

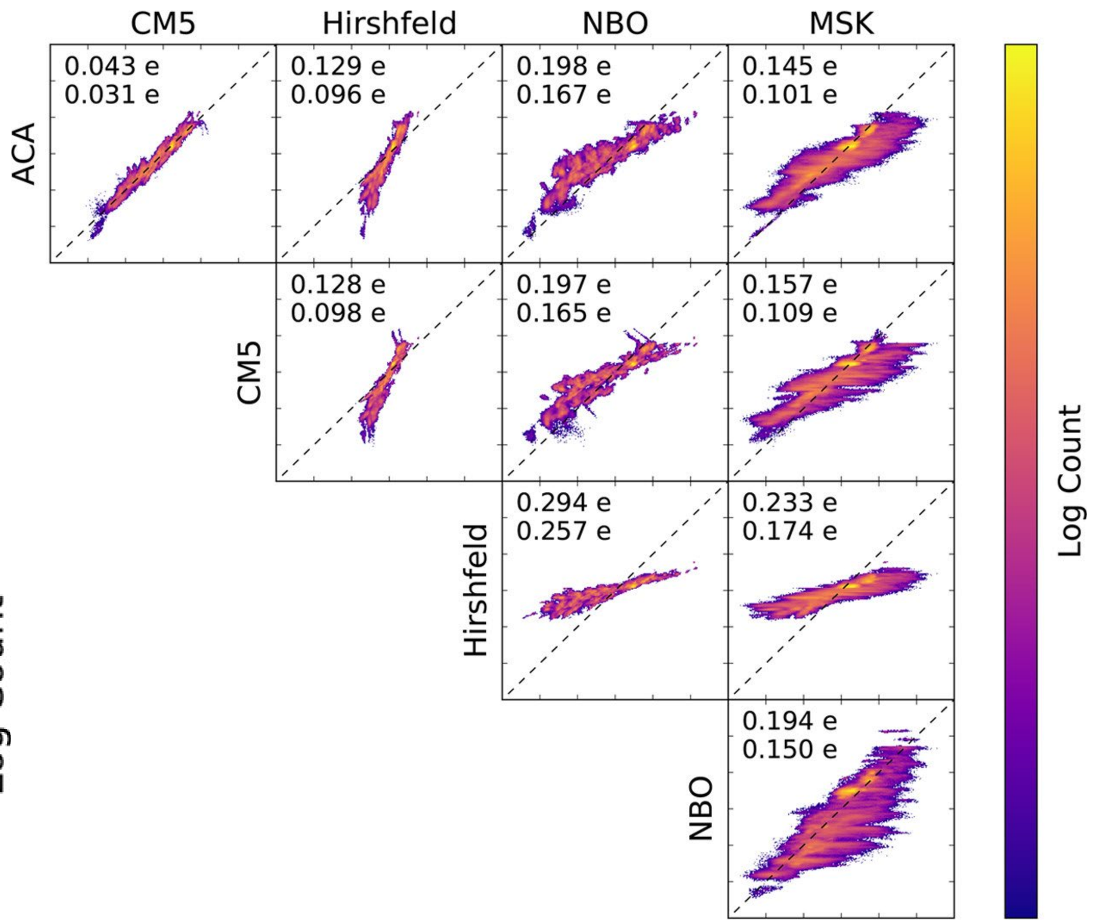
DrugBank Benchmark (cm^{-1})



Tripeptide Benchmark (cm^{-1})





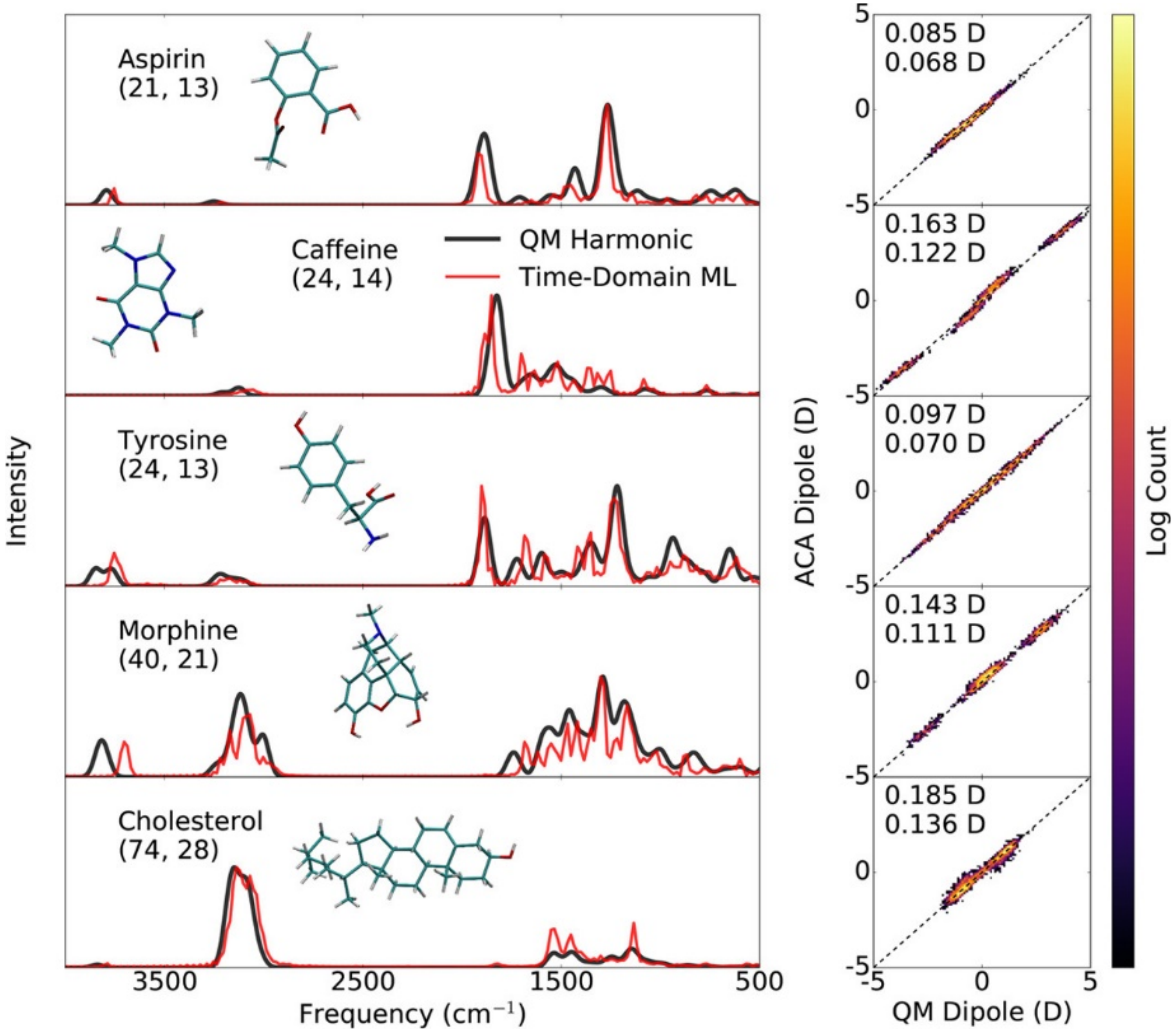


Discovering a Transferable Charge Assignment Model Using Machine Learning

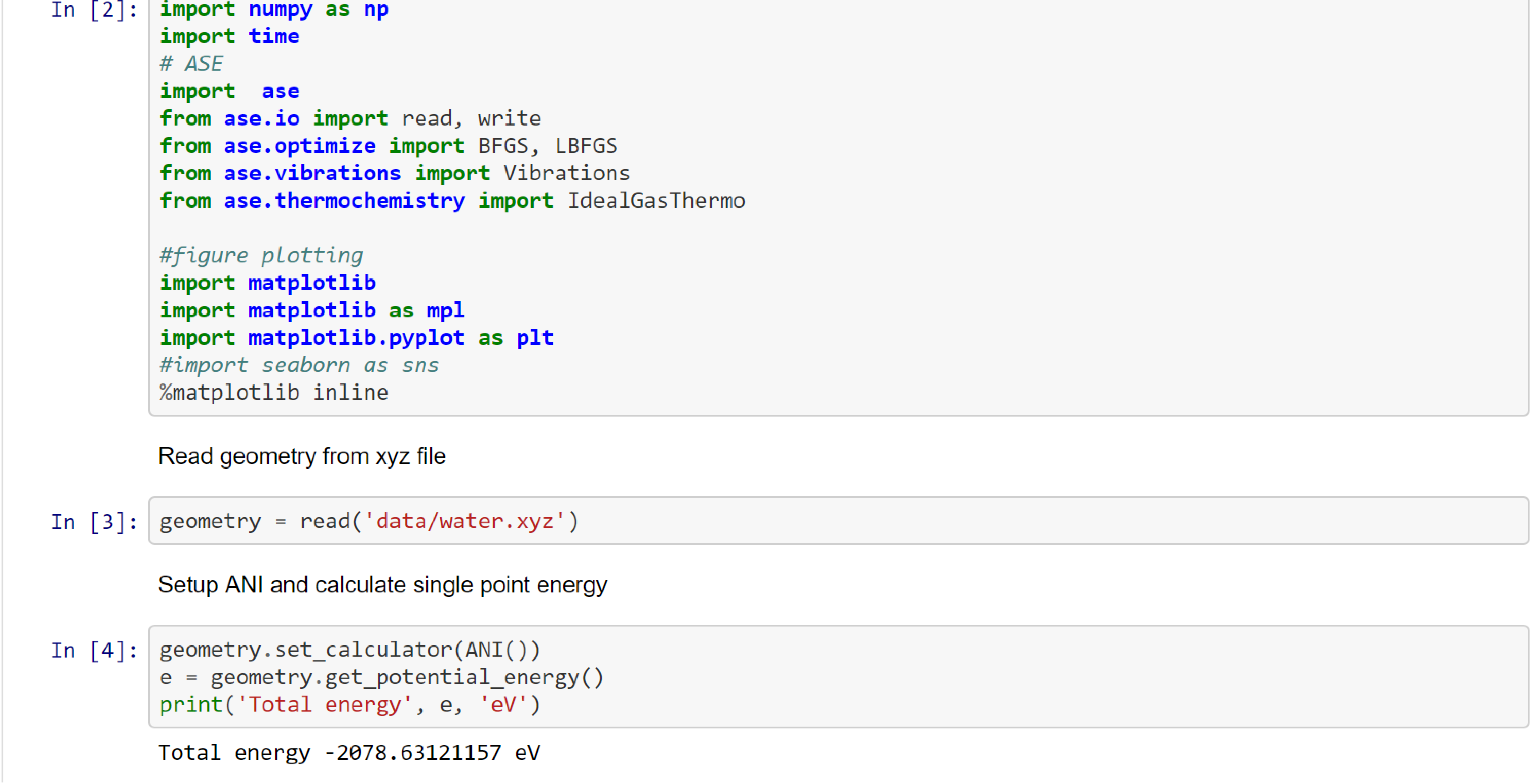
A.E. Sifain, N. Lubbers, B.T. Nebgen, J.S. Smith, A.Y. Lokhov, O. Isayev, A. E. Roitberg, K. Barros, S. Tretiak.

J. Phys. Chem. Lett. 9, **2018**, 4495-4501

Accurate IR spectra simulation with time-domain ML



A.E. Sifain, N. Lubbers, B.T. Nebgen, J.S. Smith, A.Y. Lokhov, O. Isayev, A. E. Roitberg, K. Barros, S. Tretiak.
J. Phys. Chem. Lett. **2018**. DOI: 10.1021/acs.jpclett.8b01939



In [2]:
```
import numpy as np
import time
# ASE
import  ase
from ase.io import read, write
from ase.optimize import BFGS, LBFGS
from ase.vibrations import Vibrations
from ase.thermochemistry import IdealGasThermo

#figure plotting
import matplotlib
import matplotlib as mpl
import matplotlib.pyplot as plt
#import seaborn as sns
%matplotlib inline
```

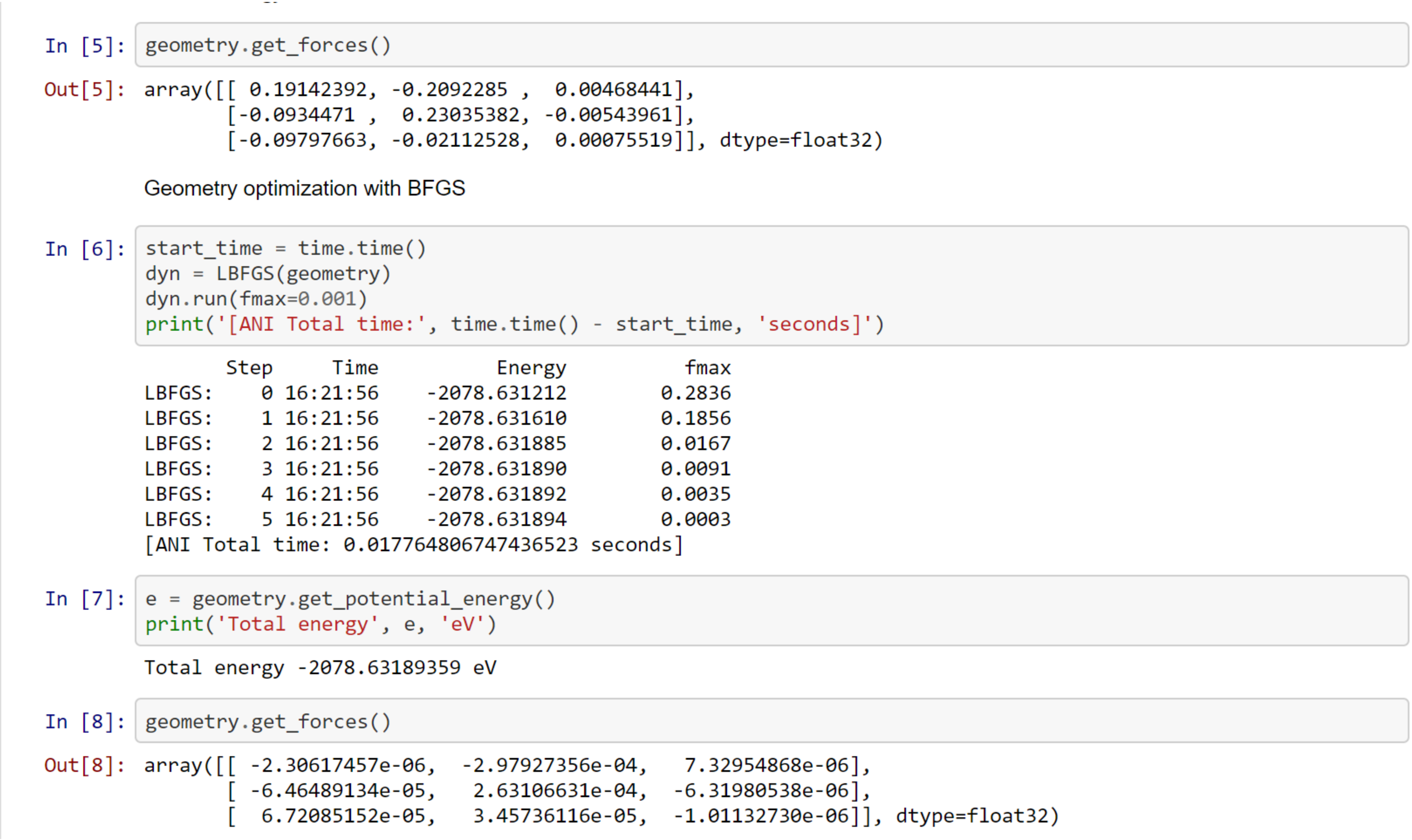
Read geometry from xyz file

In [3]:
```
geometry = read('data/water.xyz')
```

Setup ANI and calculate single point energy

In [4]:
```
geometry.set_calculator(ANI())
e = geometry.get_potential_energy()
print('Total energy', e, 'eV')
```

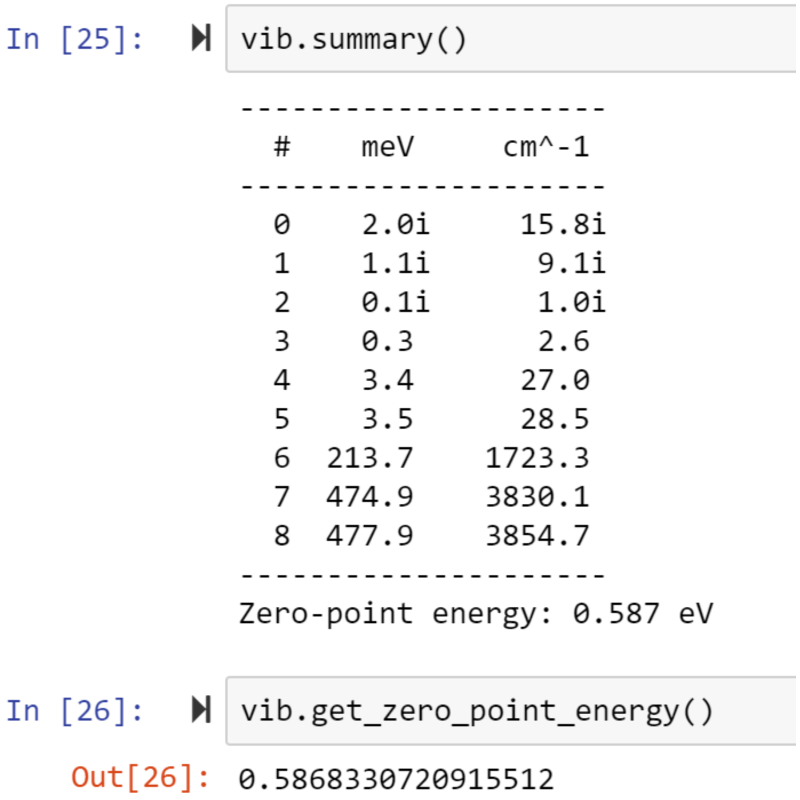
```
Total energy -2078.63121157 eV
```



```
In [5]: geometry.get_forces()

Out[5]: array([[ 0.19142392, -0.2092285 ,  0.00468441],
               [-0.0934471 ,  0.23035382, -0.00543961],
               [-0.09797663, -0.02112528,  0.00075519]], dtype=float32)
```

Geometry optimization with BFGS

```
In [6]: start_time = time.time()
        dyn = LBFGS(geometry)
        dyn.run(fmax=0.001)
        print('[ANI Total time:', time.time() - start_time, 'seconds]')

               Step     Time          Energy         fmax
        LBFGS:    0 16:21:56    -2078.631212        0.2836
        LBFGS:    1 16:21:56    -2078.631610        0.1856
        LBFGS:    2 16:21:56    -2078.631885        0.0167
        LBFGS:    3 16:21:56    -2078.631890        0.0091
        LBFGS:    4 16:21:56    -2078.631892        0.0035
        LBFGS:    5 16:21:56    -2078.631894        0.0003
        [ANI Total time: 0.017764806747436523 seconds]

In [7]: e = geometry.get_potential_energy()
        print('Total energy', e, 'eV')

        Total energy -2078.63189359 eV

In [8]: geometry.get_forces()

Out[8]: array([[ -2.30617457e-06,  -2.97927356e-04,   7.32954868e-06],
               [ -6.46489134e-05,   2.63106631e-04,  -6.31980538e-06],
               [  6.72085152e-05,   3.45736116e-05,  -1.01132730e-06]], dtype=float32)
```



In [25]:

```
vib.summary()
```

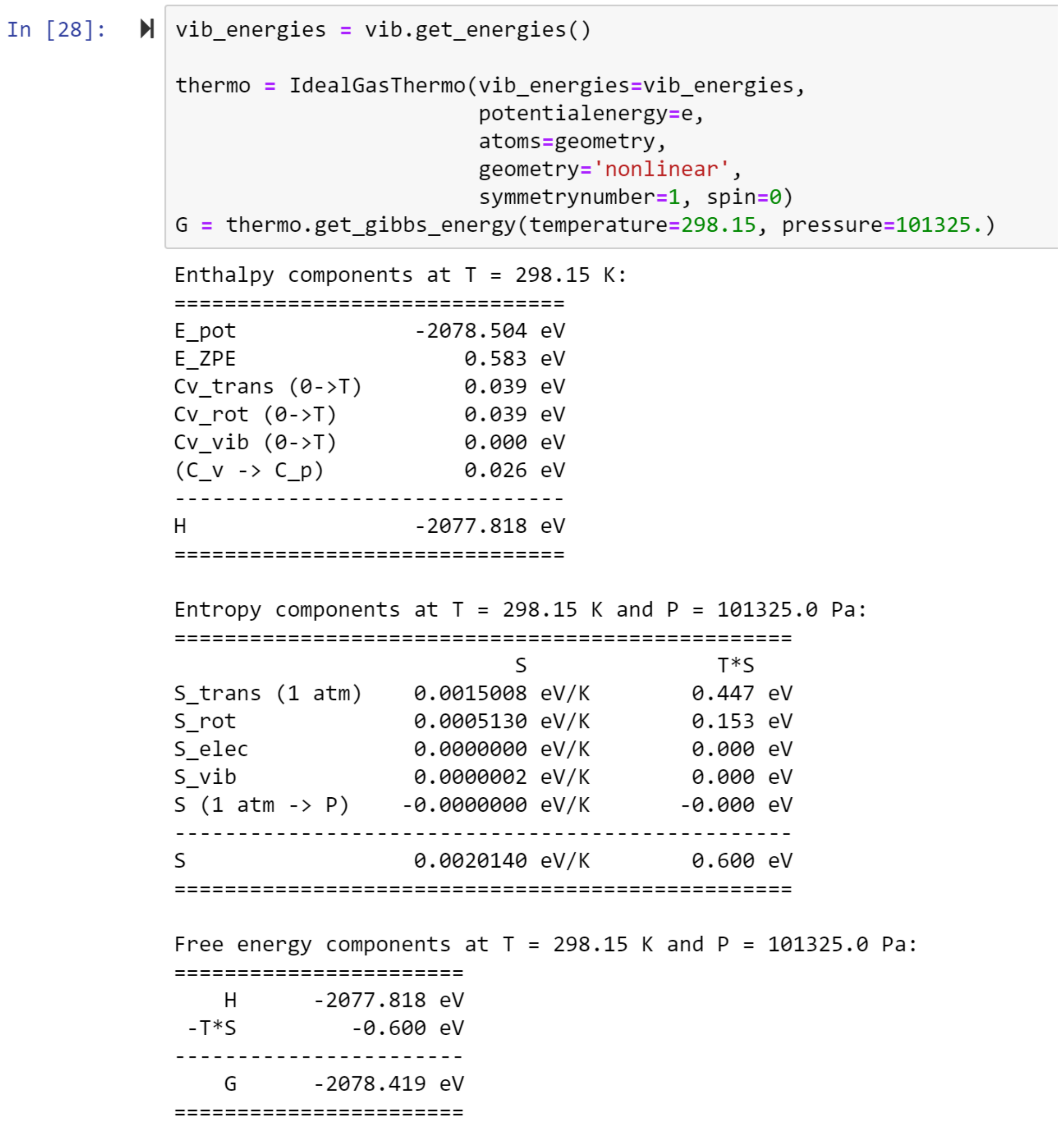
```
---------------------
  #    meV     cm^-1
---------------------
  0    2.0i     15.8i
  1    1.1i      9.1i
  2    0.1i      1.0i
  3    0.3       2.6
  4    3.4      27.0
  5    3.5      28.5
  6  213.7    1723.3
  7  474.9    3830.1
  8  477.9    3854.7
---------------------
Zero-point energy: 0.587 eV
```

In [26]:

```
vib.get_zero_point_energy()
```

Out[26]:

```
0.5868330720915512
```



In [28]:

```
vib_energies = vib.get_energies()

thermo = IdealGasThermo(vib_energies=vib_energies,
                        potentialenergy=e,
                        atoms=geometry,
                        geometry='nonlinear',
                        symmetrynumber=1, spin=0)
G = thermo.get_gibbs_energy(temperature=298.15, pressure=101325.)
```

```
Enthalpy components at T = 298.15 K:
===============================
E_pot              -2078.504 eV
E_ZPE                  0.583 eV
Cv_trans (0->T)        0.039 eV
Cv_rot (0->T)          0.039 eV
Cv_vib (0->T)          0.000 eV
(C_v -> C_p)           0.026 eV
-------------------------------
H                  -2077.818 eV
===============================

Entropy components at T = 298.15 K and P = 101325.0 Pa:
=================================================
                           S               T*S
S_trans (1 atm)    0.0015008 eV/K        0.447 eV
S_rot              0.0005130 eV/K        0.153 eV
S_elec             0.0000000 eV/K        0.000 eV
S_vib              0.0000002 eV/K        0.000 eV
S (1 atm -> P)    -0.0000000 eV/K       -0.000 eV
-------------------------------------------------
S                  0.0020140 eV/K        0.600 eV
=================================================

Free energy components at T = 298.15 K and P = 101325.0 Pa:
=======================
    H      -2077.818 eV
 -T*S         -0.600 eV
-----------------------
    G      -2078.419 eV
=======================
```

Can we go beyond DFT?

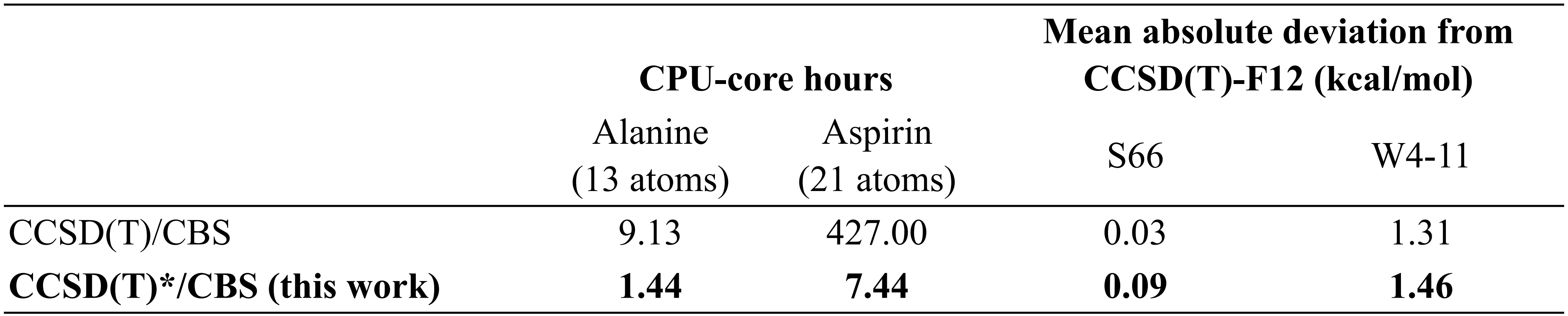
High Throughput CCSDT(T)/CBS

$$E_{total}^{CBS} \approx E_{HF}^{CBS} + E_{MP2}^{CBS} + \left(E_{CCSD(T)}^{cc-pVTZ} - E_{MP2}^{cc-pVTZ}\right)$$

$$E_{CCSD(T)}^{cc-pVTZ} \approx E_{Normal-DPLNO-CCSD(T)}^{cc-pVTZ} + (E_{Tight-DPLNO-CCSD(T)}^{cc-pVDZ} - E_{Normal-DPLNO-CCSD(T)}^{cc-pVDZ}\Big)$$

J.S. Smith et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature Comm.* ***2019**,* 10*,* 2903.

Accuracy Benchmark

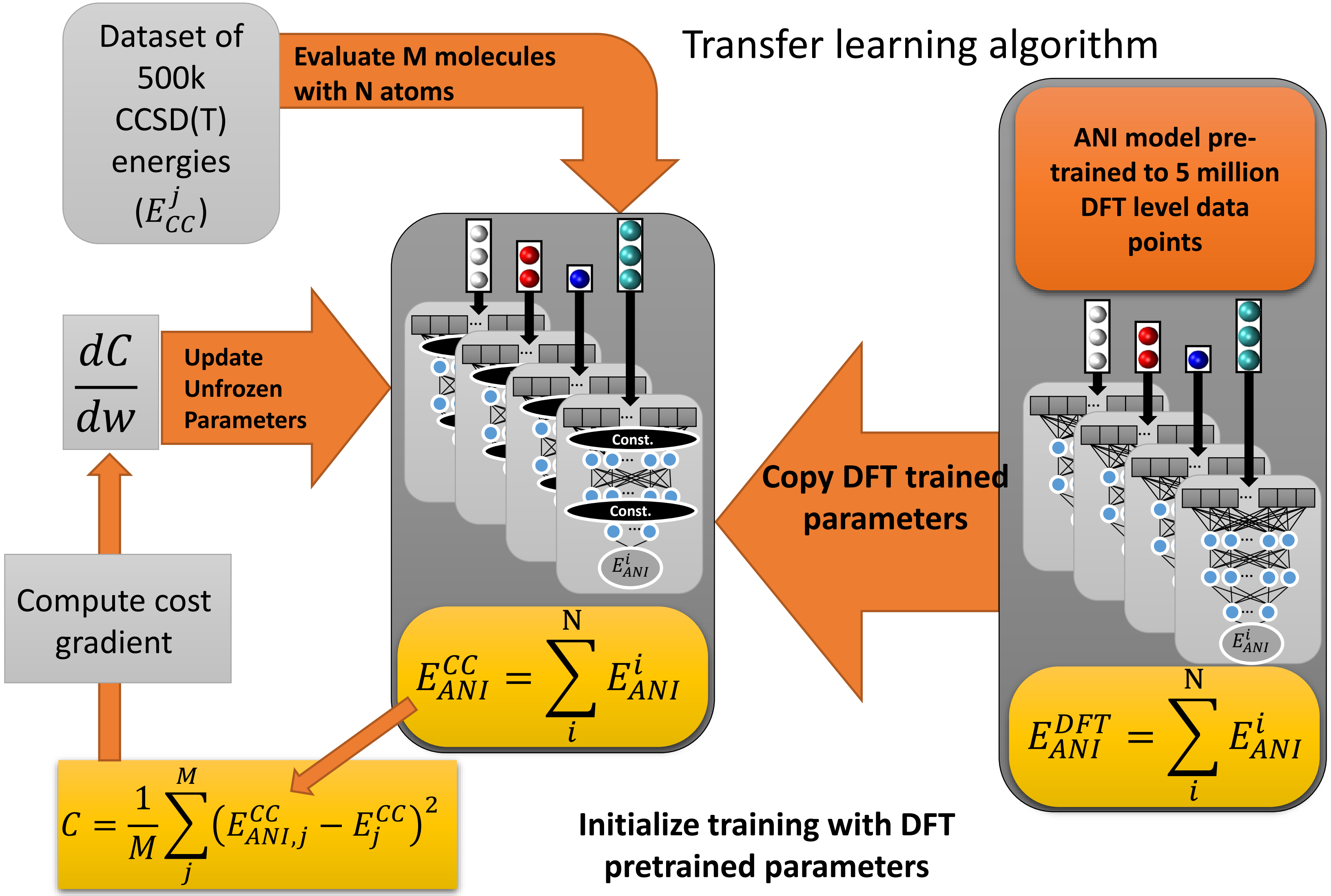


	CPU-core hours		**Mean absolute deviation from CCSD(T)-F12 (kcal/mol)**	
	Alanine (13 atoms)	Aspirin (21 atoms)	S66	W4-11
CCSD(T)/CBS	9.13	427.00	0.03	1.31
CCSD(T)*/CBS (this work)	**1.44**	**7.44**	**0.09**	**1.46**

J.S. Smith et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature Comm.* ***2019****,* 10*,* 2903.

Transferring knowledge of CCSD(T)/CBS

- Regenerate 10% of ANI-1x training data (0.5M of 5M)
- For high-level reference we use CCSD(T)/CBS accurate QM model
- We only retrain 60k of 400k neural network parameters
- Results show clear improvement over DFT trained model
- New models are exceeding the DFT in accuracy



Transfer learning algorithm

Dataset of 500k CCSD(T) energies (E^{j}_{CC})

Evaluate M molecules with N atoms

ANI model pre-trained to 5 million DFT level data points

Copy DFT trained parameters

$$E^{DFT}_{ANI} = \sum_{i}^{N} E^{i}_{ANI}$$

$$E^{CC}_{ANI} = \sum_{i}^{N} E^{i}_{ANI}$$

$$C = \frac{1}{M}\sum_{j}^{M}\left(E^{CC}_{ANI,j} - E^{CC}_{j}\right)^2$$

Compute cost gradient

$$\frac{dC}{dw}$$

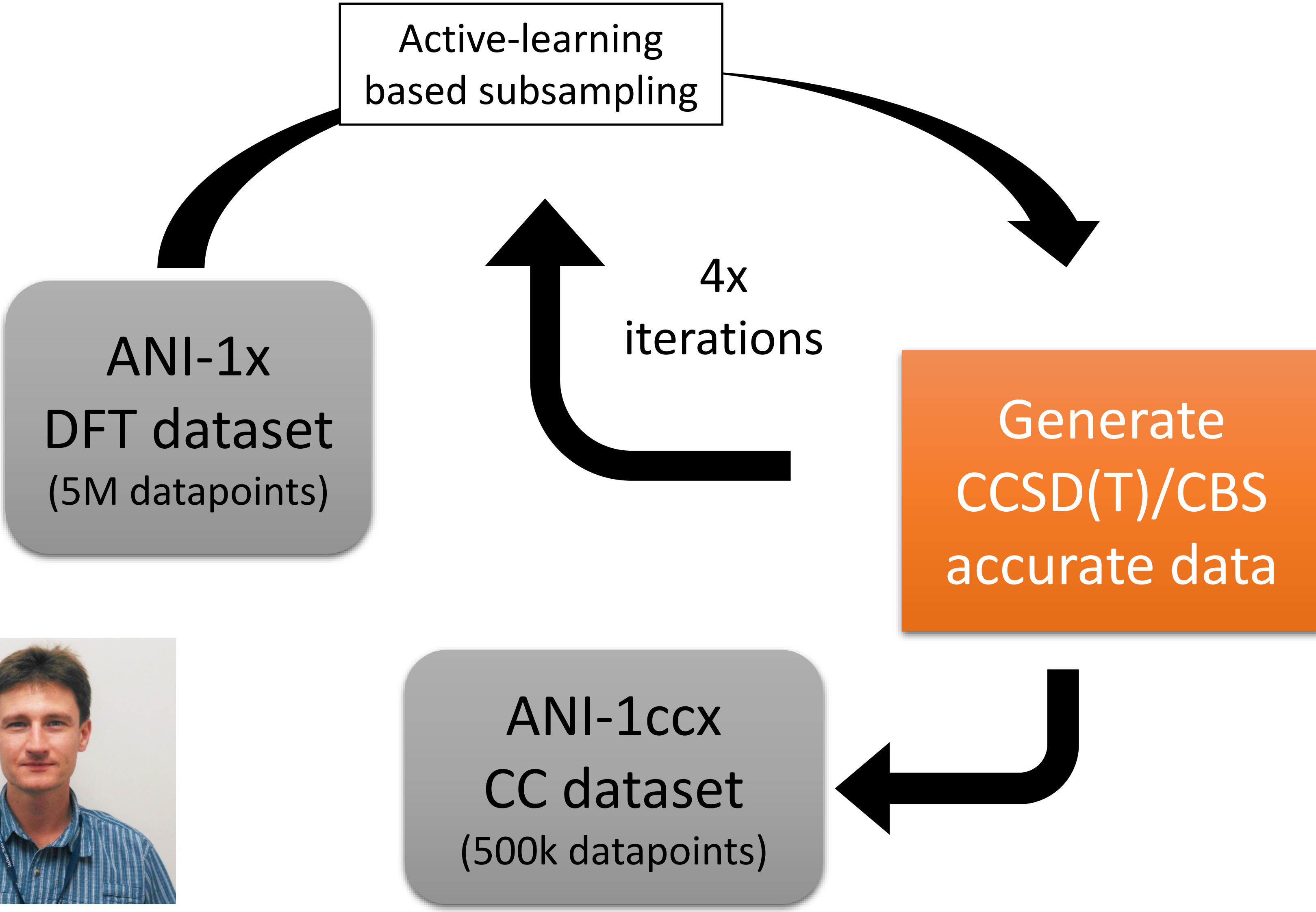
Update Unfrozen Parameters

Initialize training with DFT pretrained parameters

J.S. Smith et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature Comm.* ***2019***, 10, 2903.

Transferring knowledge of CCSD(T)/CBS

Method	Avg. Time/data point
CCSD(T)	24h
DFT	6m
ANI-1ccx	2µs



Active-learning based subsampling

ANI-1x
DFT dataset
(5M datapoints)

4x
iterations

Generate
CCSD(T)/CBS
accurate data

ANI-1ccx
CC dataset
(500k datapoints)

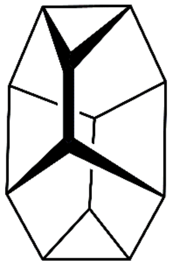
LANL
Ben Nebgen

UF -> LANL
Justin S. Smith

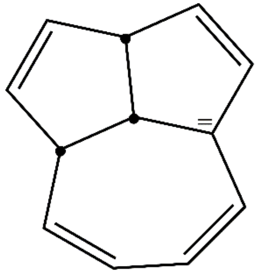
UNC
Roman Zubatyuk

15M of HPC computer hours at LANL. To be released soon

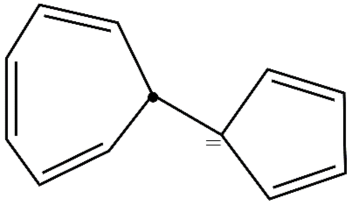
Hydrocarbon reaction energy benchmark, DFT vs CCSD(T)



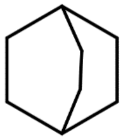
E1 (1)



E2 (22)



E3 (31)

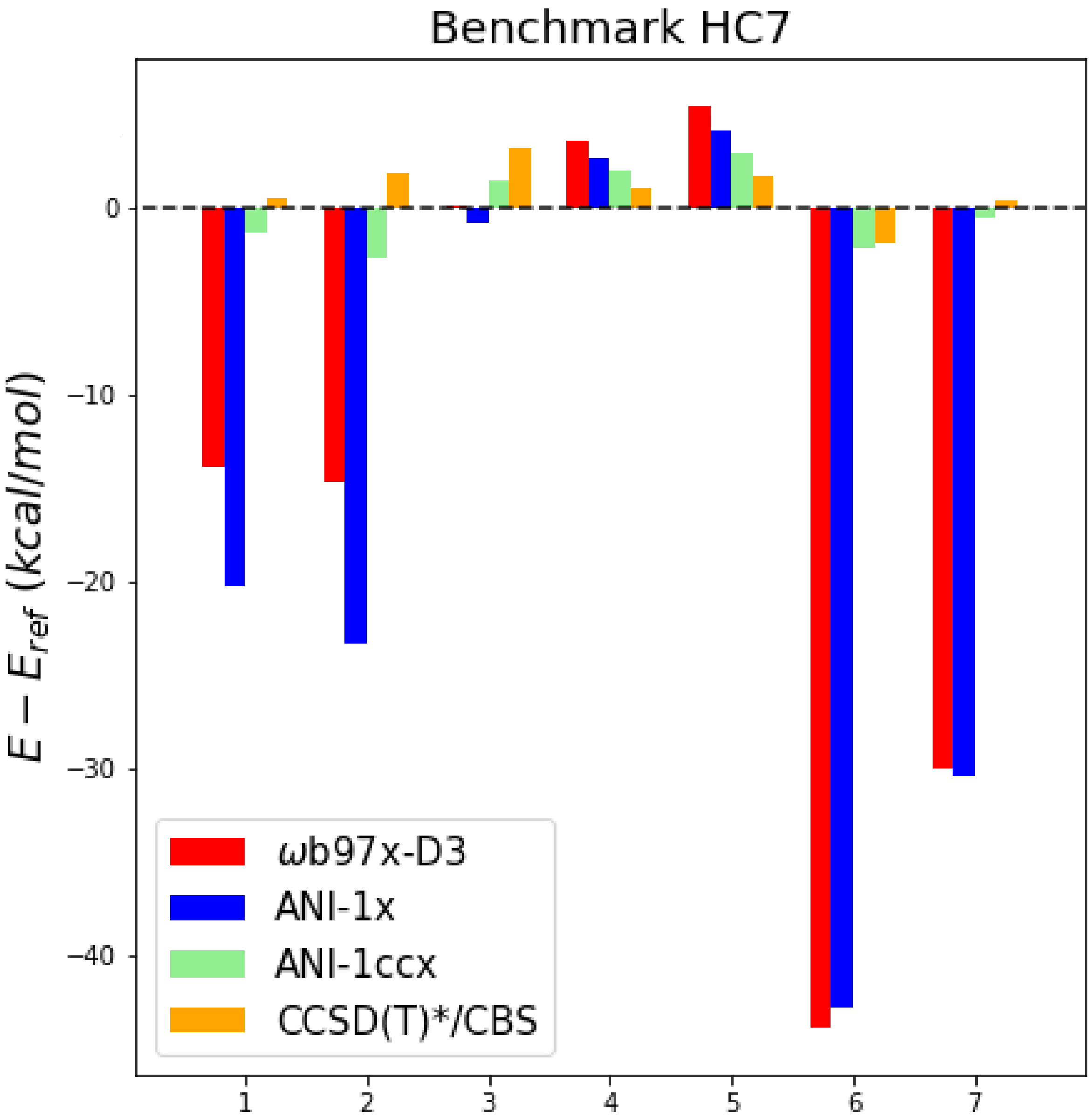


E4
(Bicyclo[2.2.2]octane)

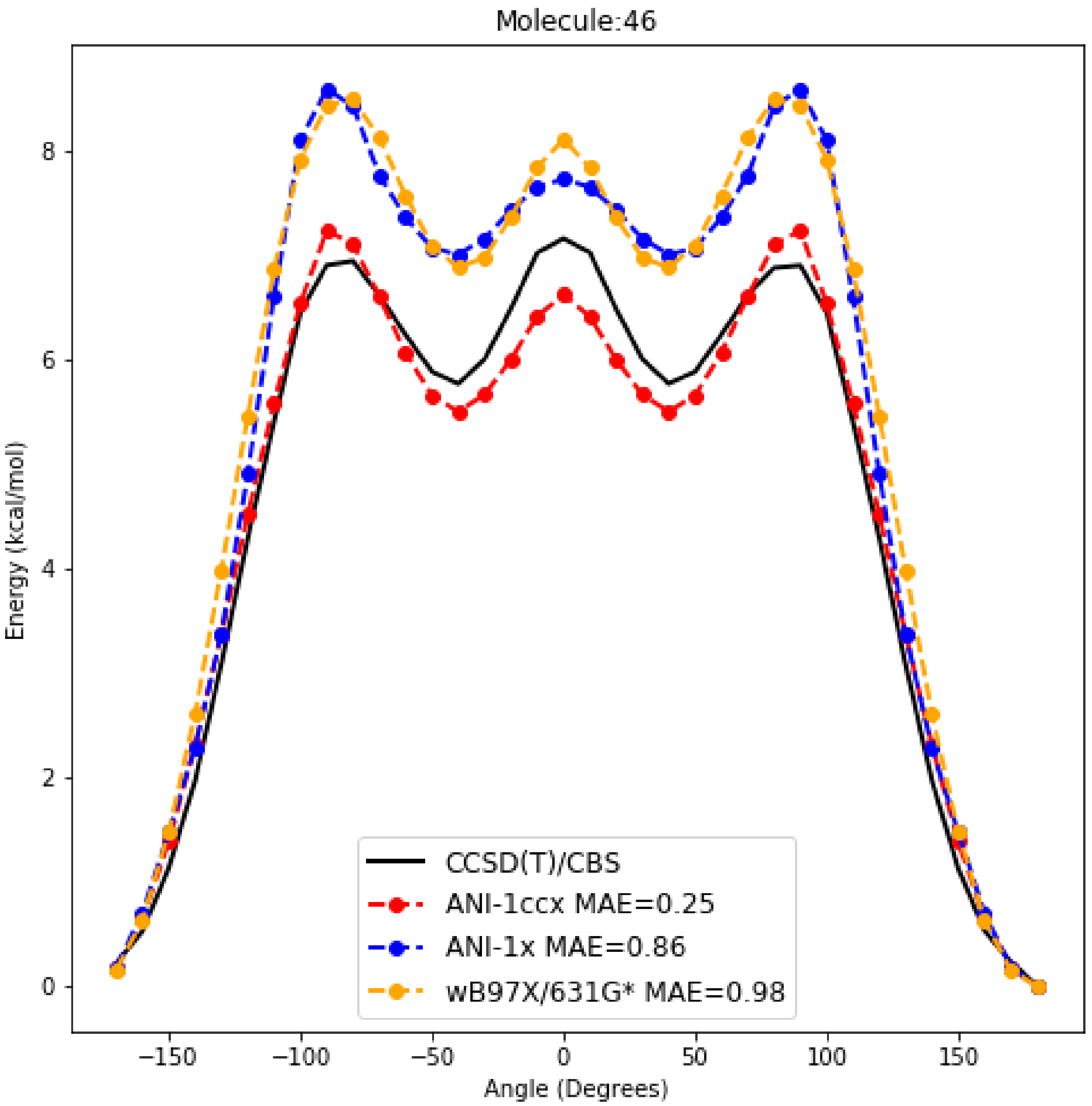
Units: kcal/mol

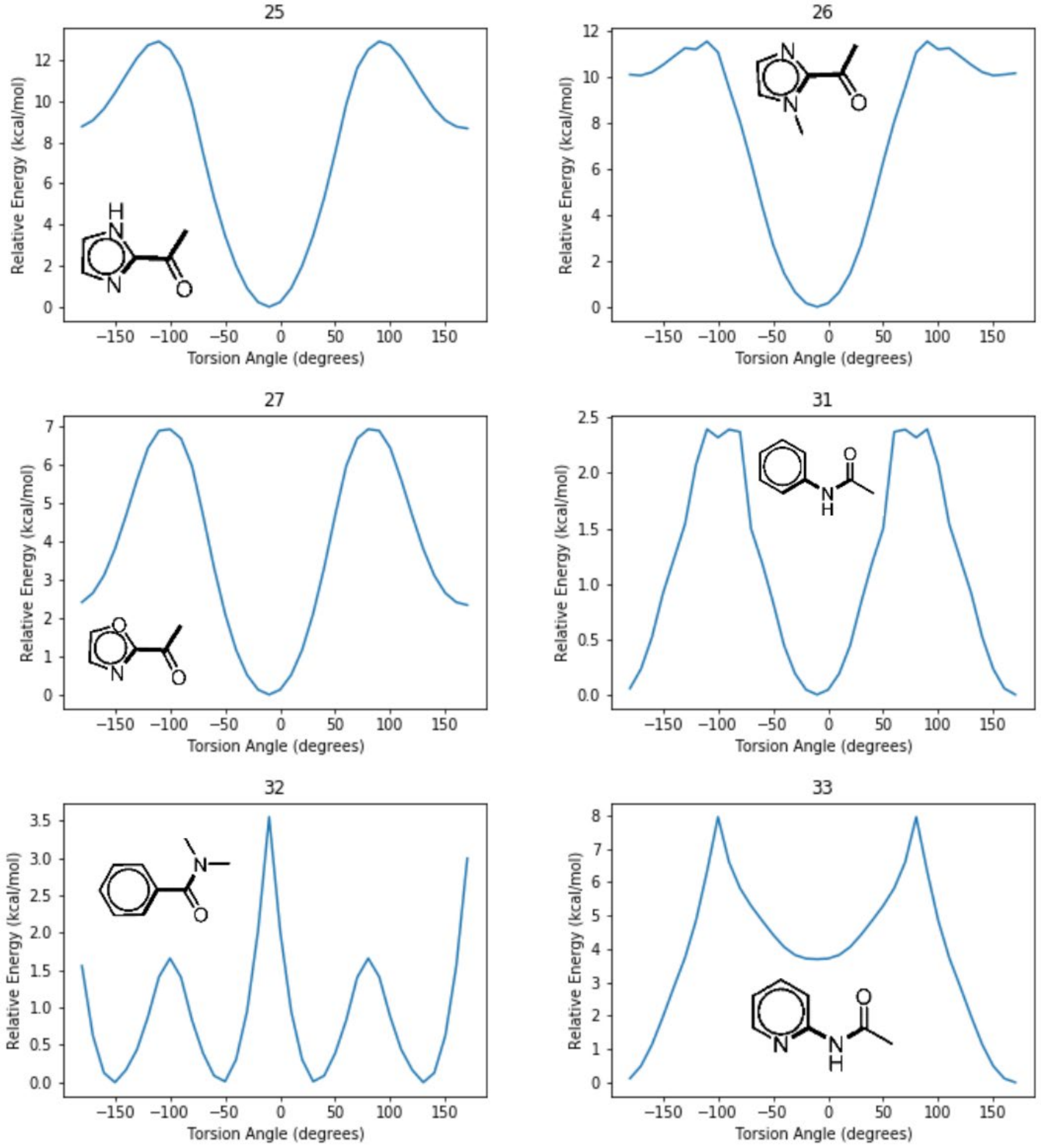
Reaction	Ref.	ANI-1ccx	CCSD(T)*/CBS	ANI-1x	ωb97x
1) E**1** → E**2**	14.3	15.6	13.8	34.6	28.2
2) E**1** → E**3**	25.0	27.7	23.1	48.3	39.7
3) Octane-a → Octane-b	1.9	0.4	-1.3	2.7	1.7
4) $4CH_4 + C_6H_{14} \rightarrow 5C_2H_6$	9.8	7.9	8.7	7.2	6.2
5) $6CH_4 + C_8H_{18} \rightarrow 7C_2H_6$	14.8	11.9	13.1	10.8	9.3
6) Adamantane → $3CH_4 + 2C_2H_2$	194.0	196.2	195.9	236.8	238.0
7) E4 → $3CH_4 + 2C_2H_2$	127.2	127.8	126.9	157.7	158.0

Reference data: Peverati, R.; Zhao, Y.; Truhlar, D. G., *J. Phys. Chem. Lett.* **2011**, *2* (16), 1991–1997.



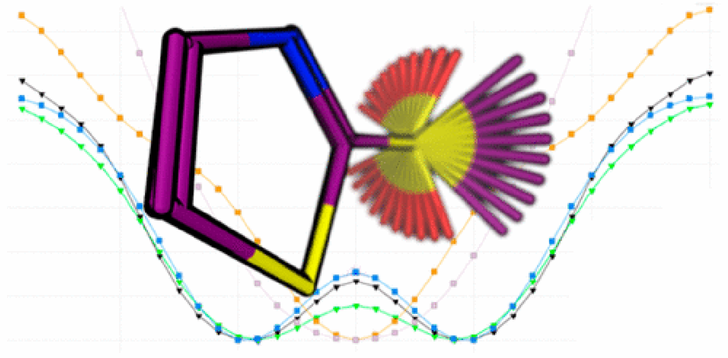
J.S. Smith et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature Comm.* ***2019***, 10, 2903.

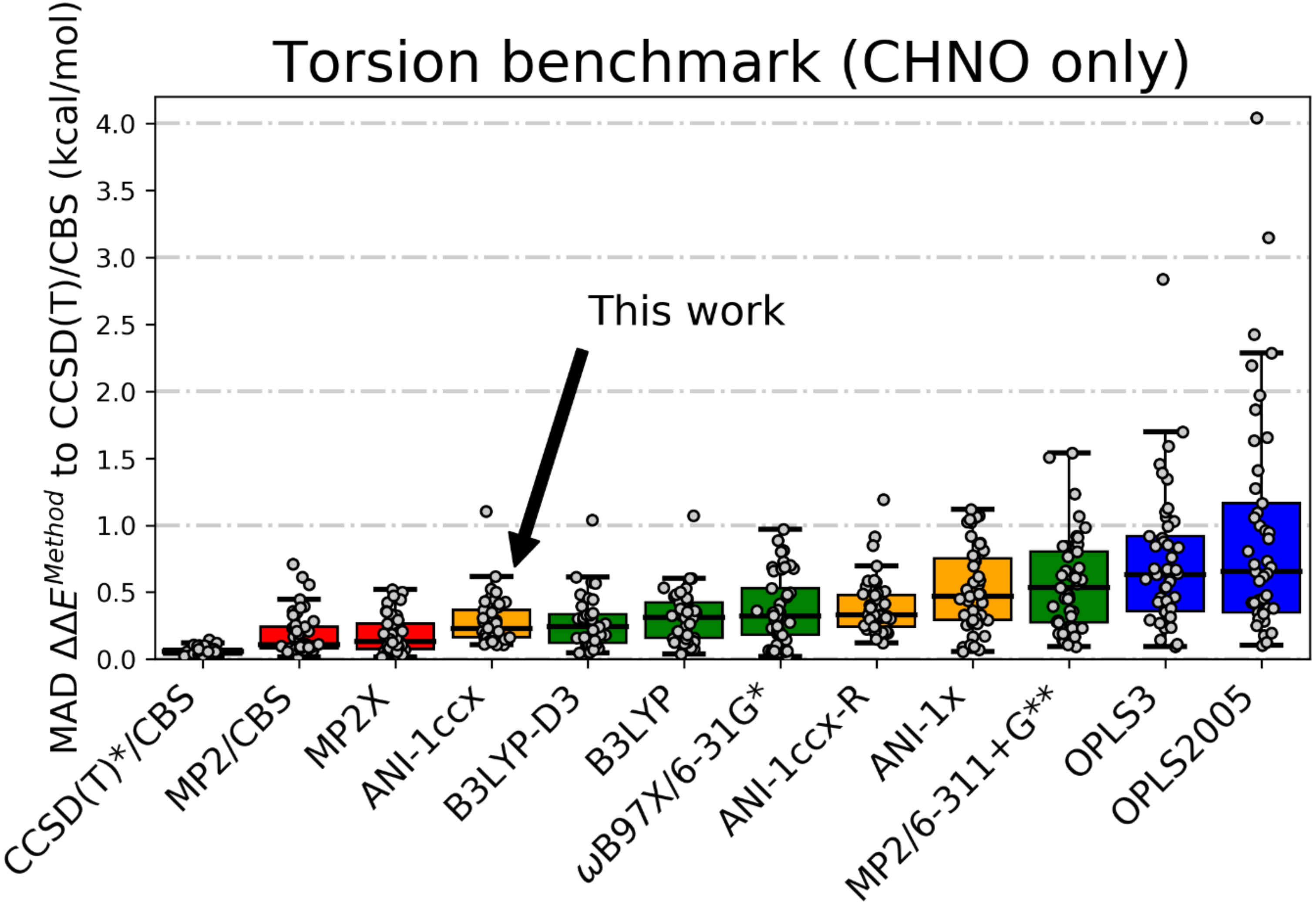




Sellers, B. D.; James, N. C.; Gobbi, A. A Comparison of Quant
Estimate Strain Energy in Druglike Fragments. *J. Chem. Inf. M*

Accurate Dihedral Profiles for Drug-like Molecules (Genentech Benchmark)

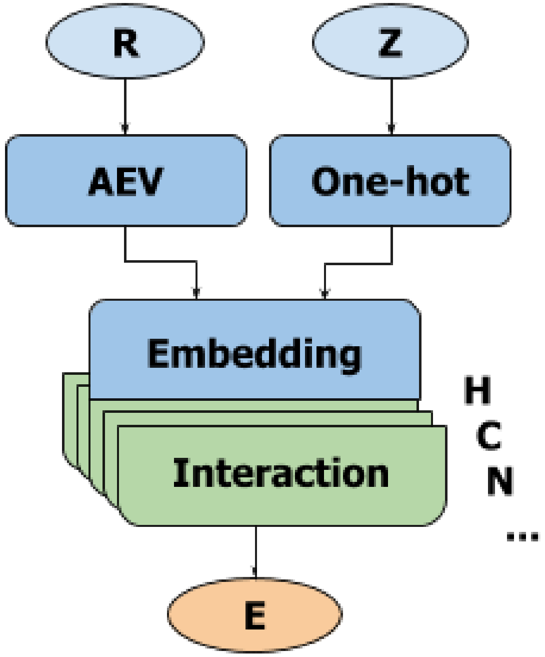




Sellers, B. D.; James, N. C.; Gobbi, A. A Comparison of Quantum and Molecular Mechanical Methods to Estimate Strain Energy in Druglike Fragments. *J. Chem. Inf. Model.* **2017**, *57* (6), 1265–1275.

Can we go beyond simple energies?

Bird's Eye View on Architecture

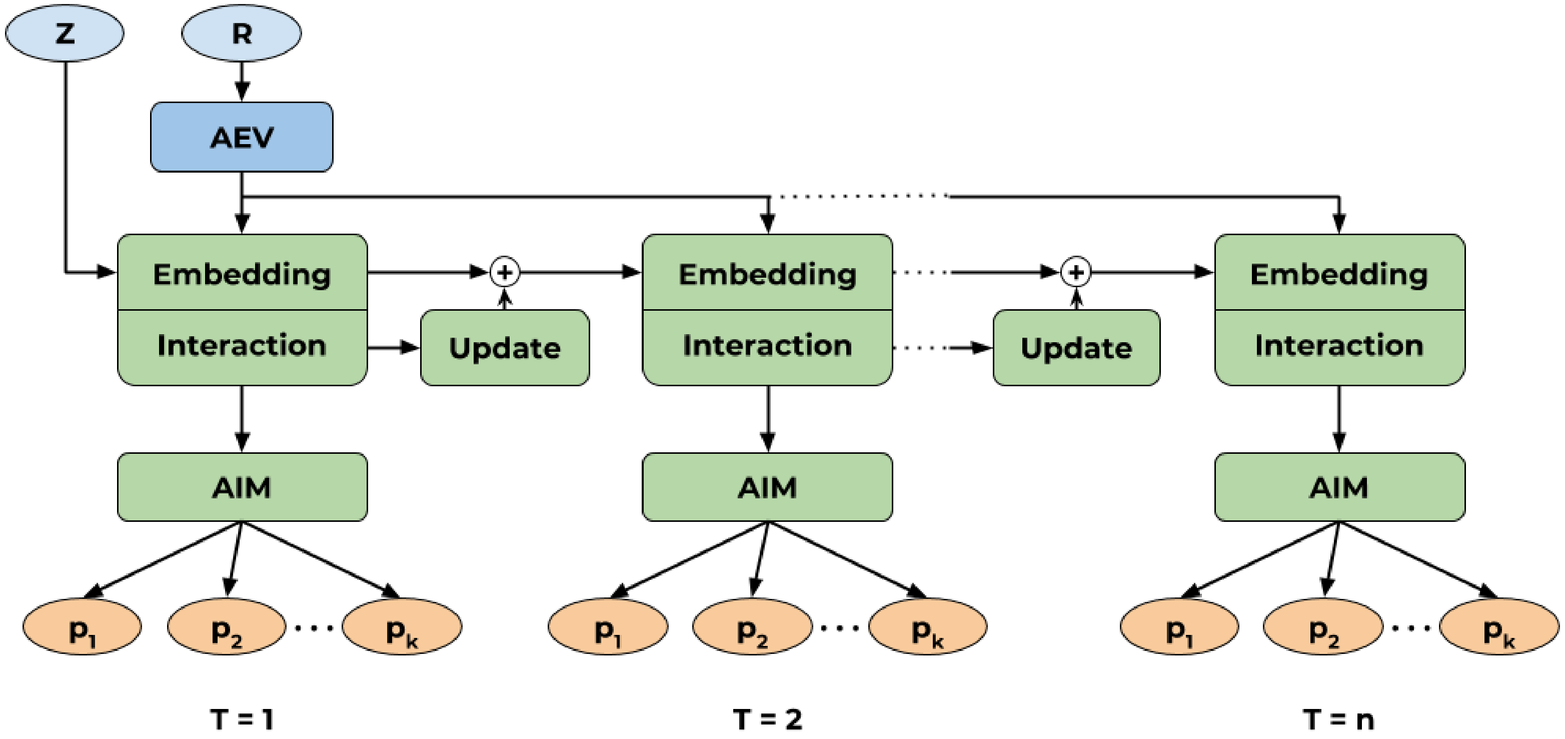


Rethinking Network Architecture: AIMNet

Atoms-in-molecules neural net

Iterative “SCF-like” update for better accuracy and
Long range interactions

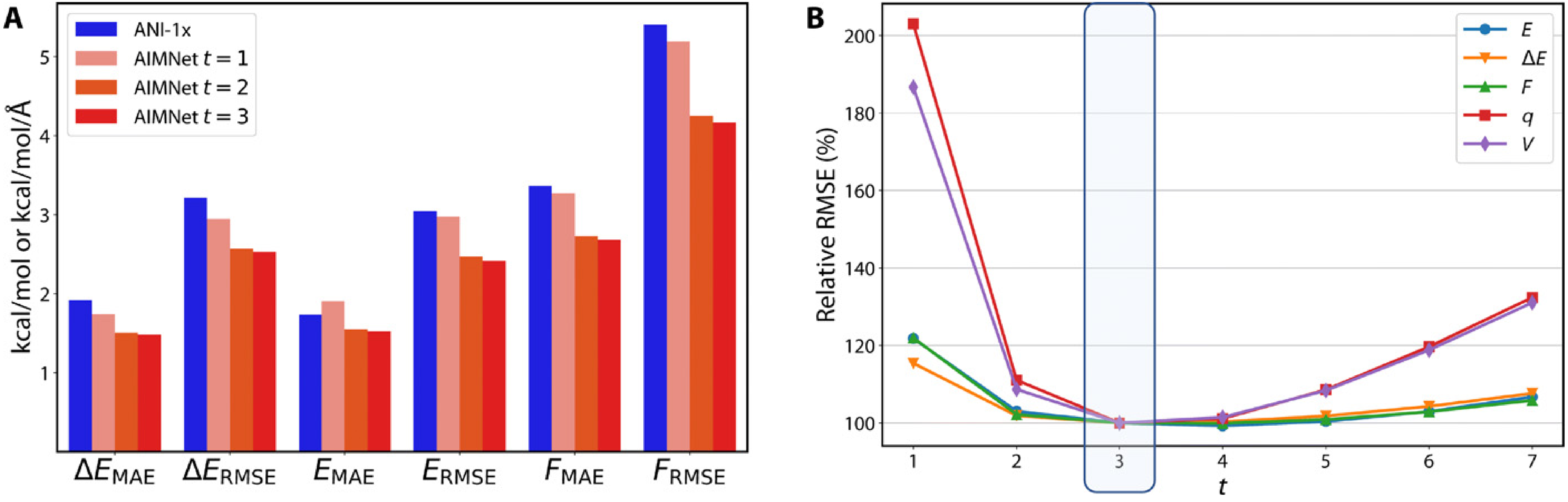
Multimodal and
multi-task learning: gas phase energy, charges, atomic volumes, continuum solvent (SMD) Correction



Deep NN network, AIMNet with T=3:
33 hidden layers, ~1M parameters

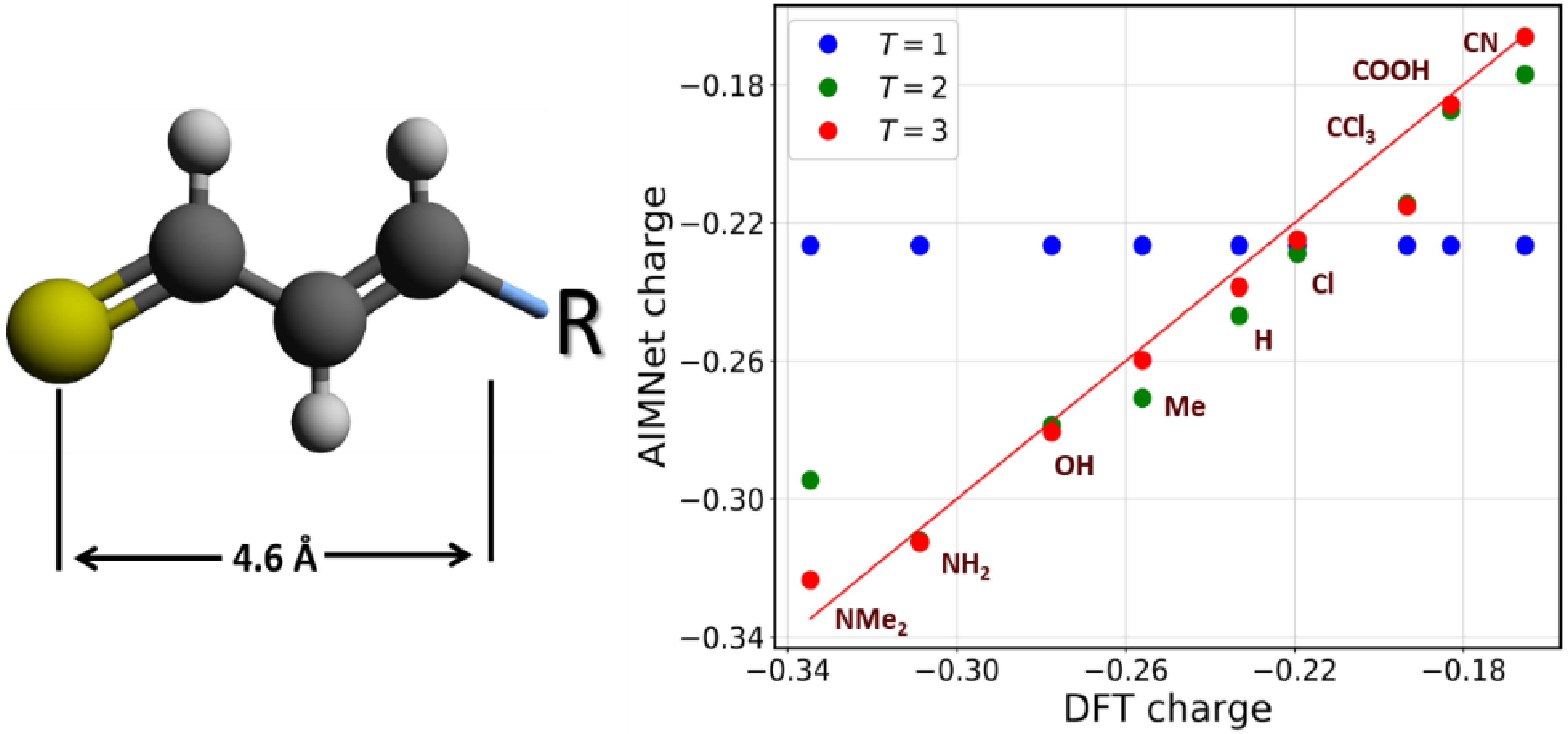
R. Zubatyuk, J.S. Smith, J. Leszczynski, O. Isayev, Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecule Neural Network. *Science Advanced*, **2019**, 5, eaav6490.

Accuracy vs NNet Iterations



R. Zubatyuk, J.S. Smith, J. Leszczynski, O. Isayev, Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecule Neural Network. *Science Advanced*, **2019**, 5, eaav6490.

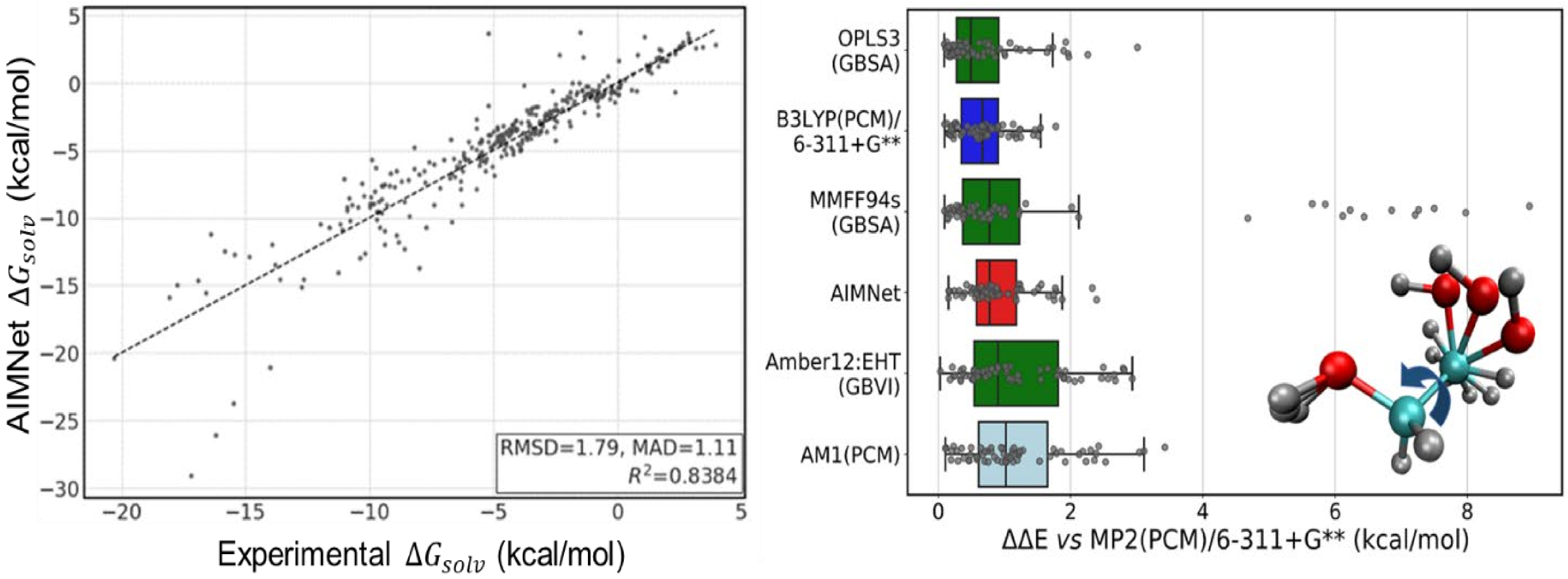
Importance of LR descriptor for atomic charges



DFT ωB97x/def2-TZVPP atomic charges on the sulfur atom of substituted thioaldehyde and AIMNet prediction with a different number of iterative passes T.

R. Zubatyuk, J.S. Smith, J. Leszczynski, O. Isayev, Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecule Neural Network. *Science Advanced*, **2019**, 5, eaav6490.

Fast & Accurate Solvation Free Energies with AIMNet

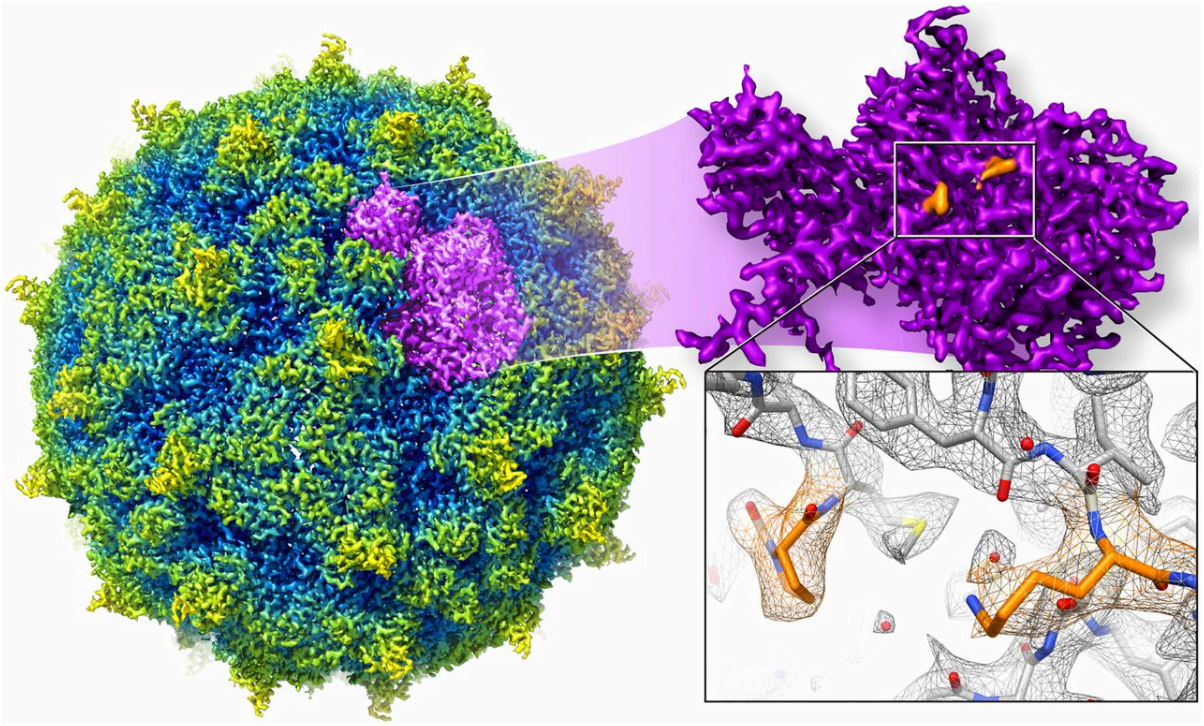


a) Experimental versus predicted with AIMNet solvation free energies (kcal/mol) for 414 neutral molecules from MNSol database. b) performance of AIMNet and other solvation models on torsion benchmark of Sellers et al.

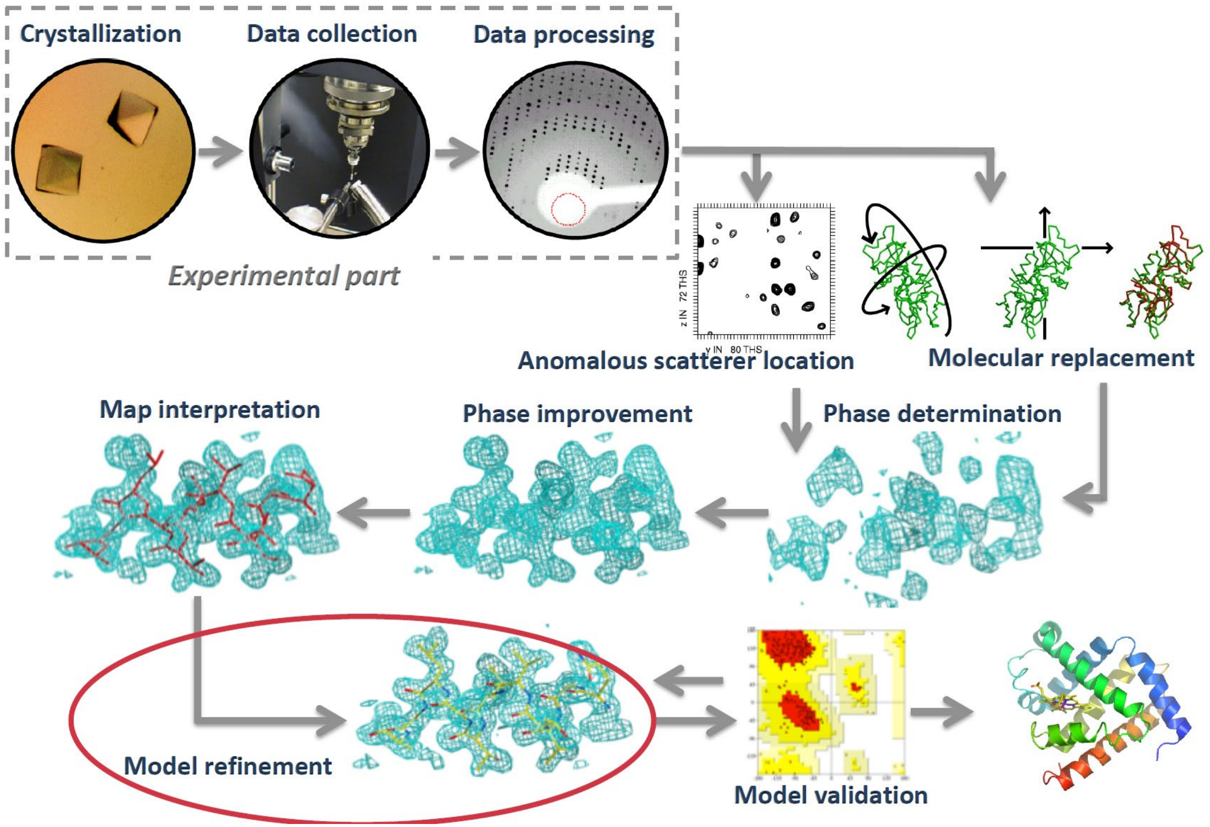
R. Zubatyuk, J.S. Smith, J. Leszczynski, O. Isayev, Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecule Neural Network. *Science Advanced*, **2019**, 5, eaav6490.

Major future developments

Quantum Refinement: next generation method for bio-crystallography and Cryo-EM



Structure determination workflow: crystallography

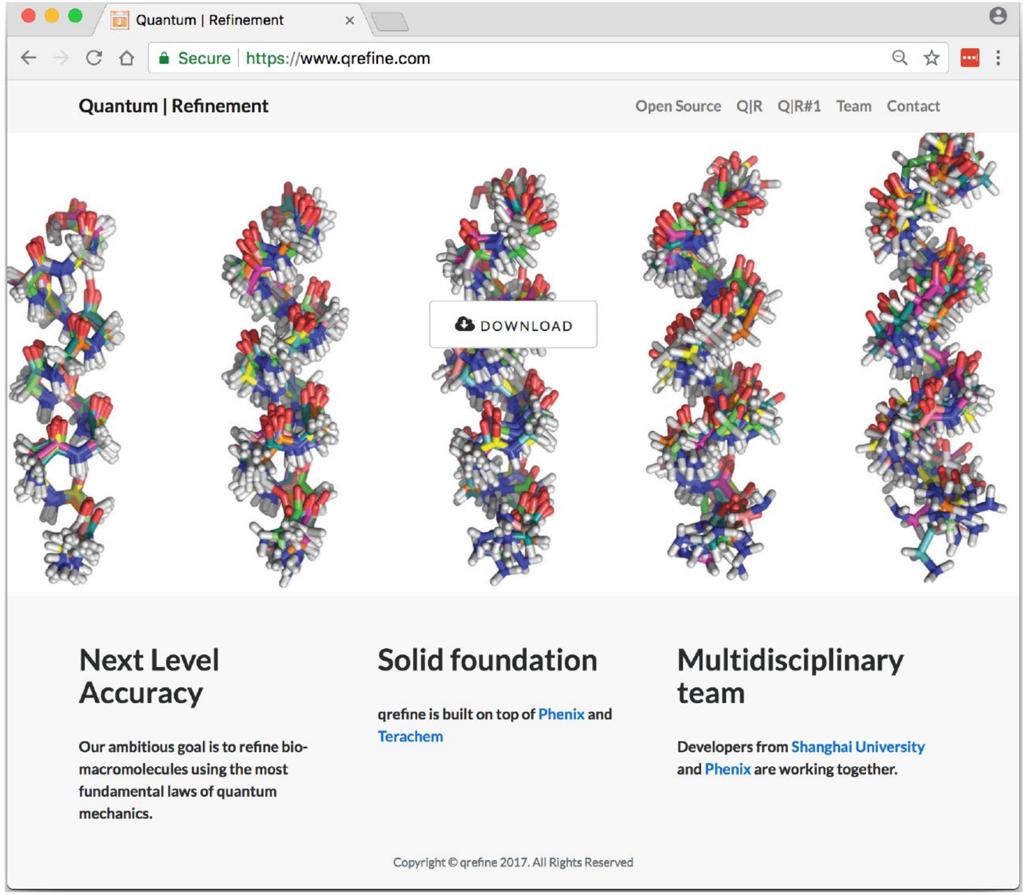


Slide courtesy of Pavel Afonine & PHENIX Q|R team

Phenix

Quantum Refinement & ANI

Q|R – Home



Next Level Accuracy

Our ambitious goal is to refine bio-macromolecules using the most fundamental laws of quantum mechanics.

Solid foundation

qrefine is built on top of Phenix and Terachem

Multidisciplinary team

Developers from Shanghai University and Phenix are working together.

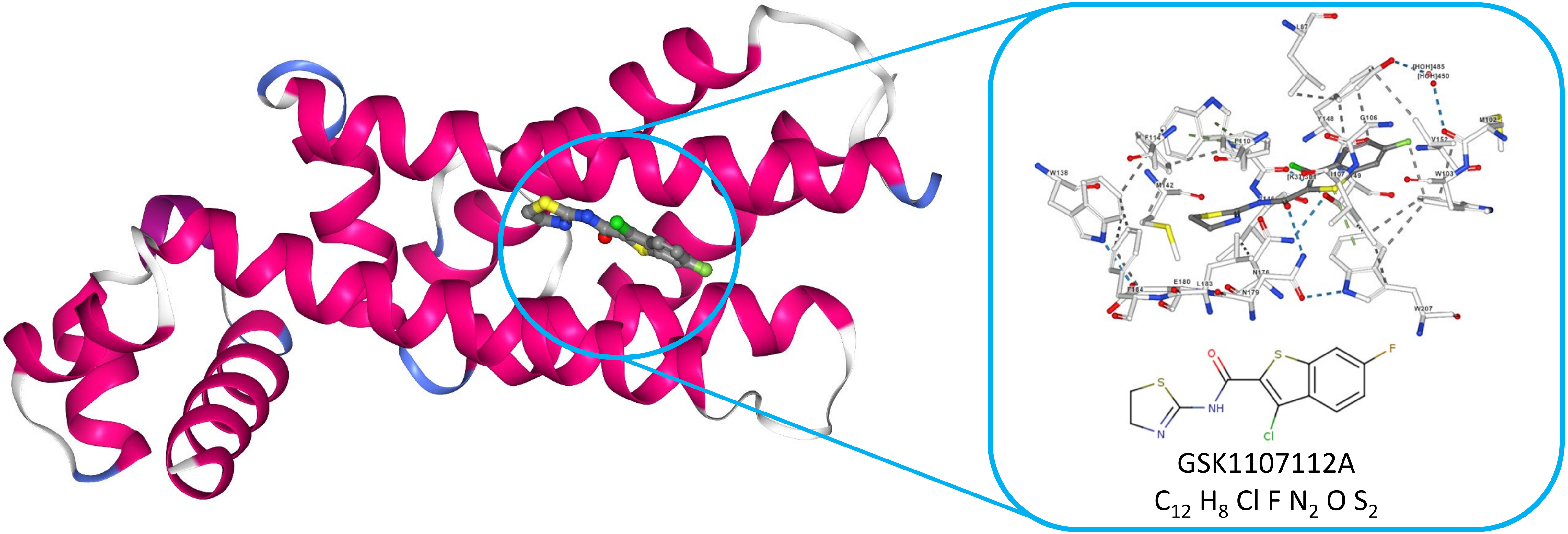
Copyright © qrefine 2017. All Rights Reserved

- TeraChem is very expensive!
- Need for special hardware (GPU)
- Takes day to a week on HPC

- Free for academia!
- Optional special hardware (GPU)
- Seconds to minutes on laptop

Slide courtesy of Pavel Afonine & PHENIX Q|R team

Toward Realistic Macromolecular Simulations

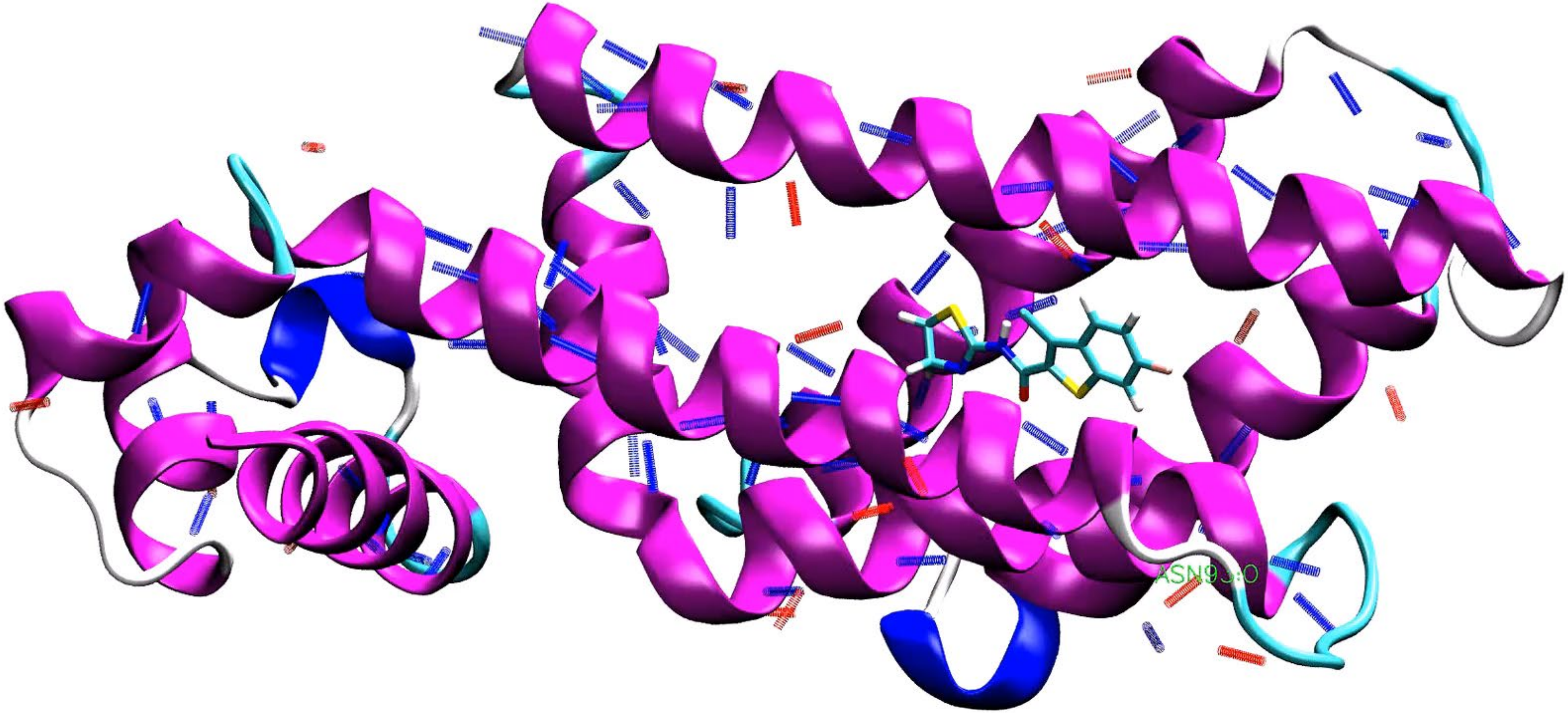


GSK1107112A

$C_{12}\,H_8\,Cl\,F\,N_2\,O\,S_2$

- ~35K atoms
- Explicit water
- No ions
- S, F and Cl in ligand

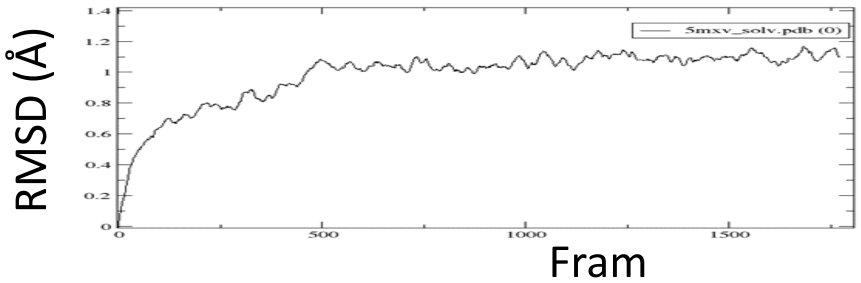
Mycobacterium tuberculosis (5MXV) in explicit water
Simulated with ANI-2 (CHNOSFCl)

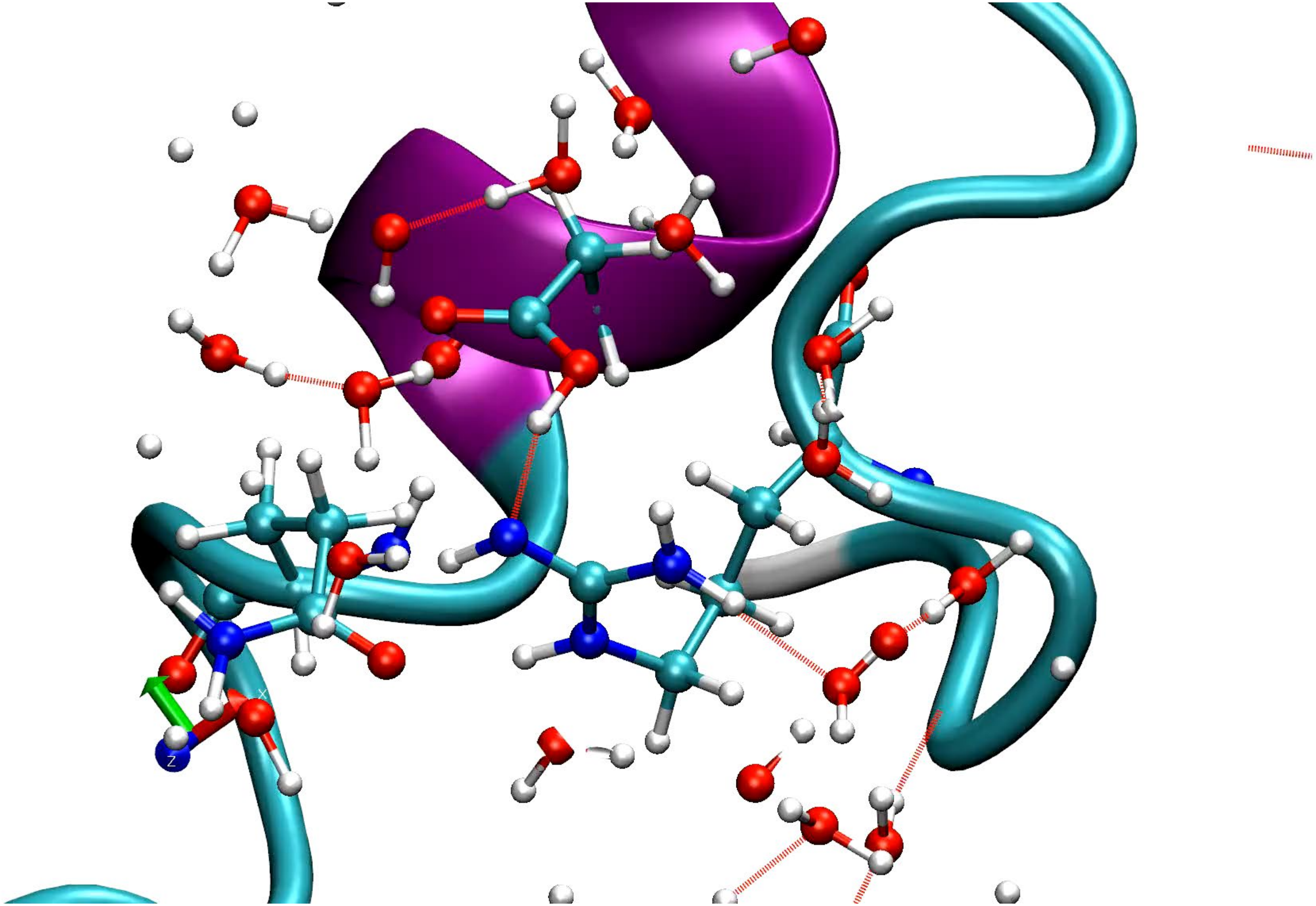


Timings for a 5x ensemble prediction for ANI-2x

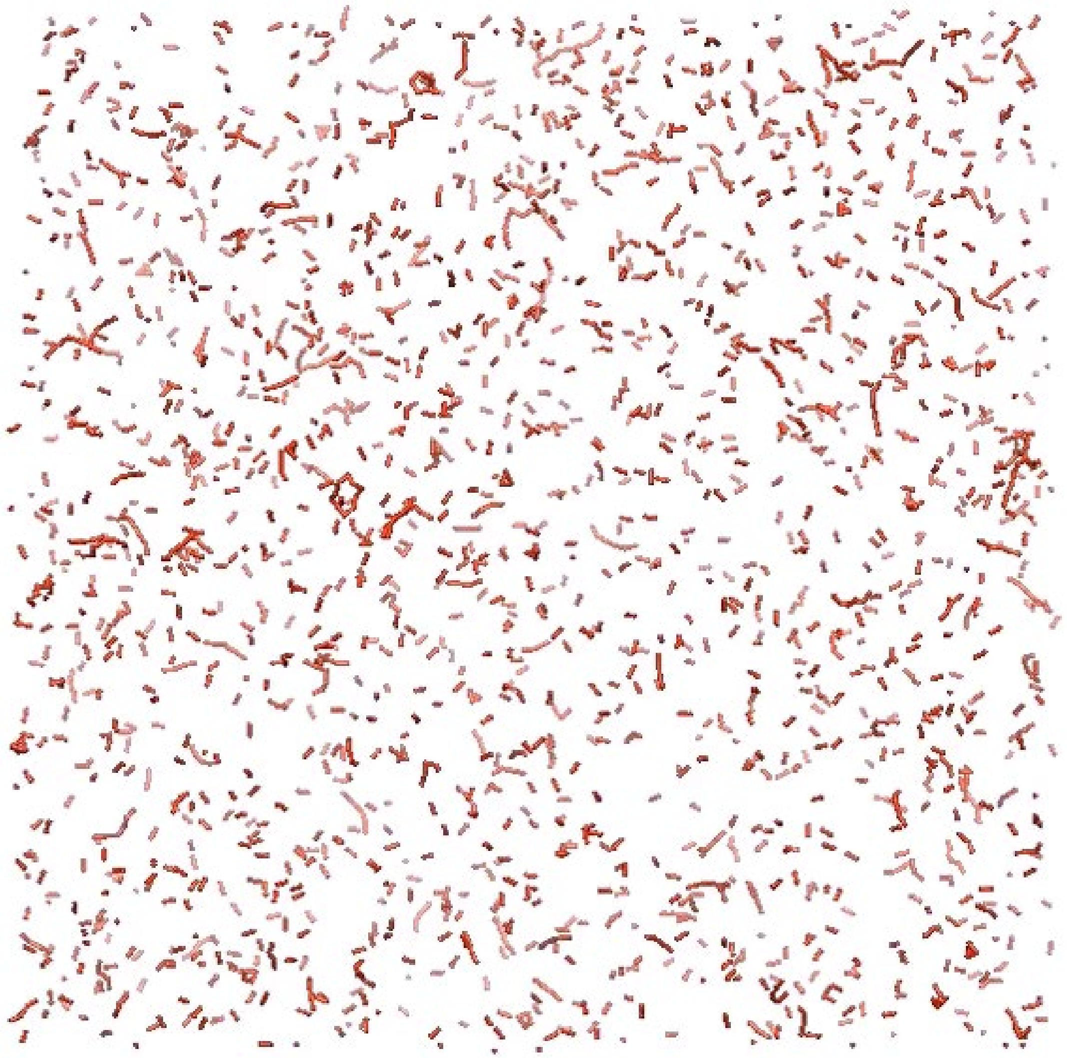
GPU	ANI-2x time per step	Total time per step	Steps per day
Tesla V100	297ms	317ms	**272k**

5ns simulation time




X
Z

Simulation of Complex Chemical Reactions



https://youtu.be/DRVMH5u8EA0

Carbon nanoparticles/sheets nucleation [4000 atoms in 60A box at 2500K, 5ns MD simulation]

Use the ANI-1x potential:

ANI-1x interfaced to ASE Python library
Available at: https://github.com/isayev/ASE_ANI

ANI-1x implementation in PyTorch
Available at: https://github.com/aiqm/torchani

Coming soon to AMBER, OpenMM & LAMMPS

Use the AIMNet:

Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecule Neural Network. **ChemRxiv**, 2018.

AIMNet implementation in Pytorch & ASE calculator:
Available at: https://github.com/aiqm/aimnet

Use the ANI-1 dataset:

ANI-1: A data set of 20M off-equilibrium DFT calculations for organic molecules
Sci. Data, 2017, **4**, 170193 DOI: 10.1038/sdata.2017.193

ANI Data set Python library
Available at: https://github.com/isayev/ANI1_dataset

Users:

academic labs:

- Stanford
- U Pitt
- CMU
- USF
- NCSU
- Barcelona
- Helsinki
- Tel Aviv

Government labs, companies etc.

Los Alamos
NATIONAL LABORATORY
EST. 1943

BERKELEY LAB
Lawrence Berkeley National Laboratory

NIH
National Institutes of Health

BASF
We create chemistry

SCHRÖDINGER.

Pfizer

Genentech

OpenEye
Scientific Software