Carnegie Mellon University

Empowering ML accelerated molecular modeling workflows with OSG Connect

Roman Zubatiuk

Carnegie Mellon University

Carnegie Mellon University

Hatice Gokcan
Phil Gusev
Mariya Popova
Roman Zubatyuk

Collaborators:

Kate Davis
Christian Devereux
Kavindri Ranasinghe
Adrian Roitberg

UF UNIVERSITY of FLORIDA

Funding:

DEPARTMENT OF THE NAVY ONR Science & Technology

NSF

BASF We create chemistry

Kipton Barros
Justin S. Smith
Nicholas Lubbers
Ben Nebgen
Sergei Tretiak

Center for Integrated Nanotechnologies CINT

CHE-1802789

Center for Nonlinear Studies

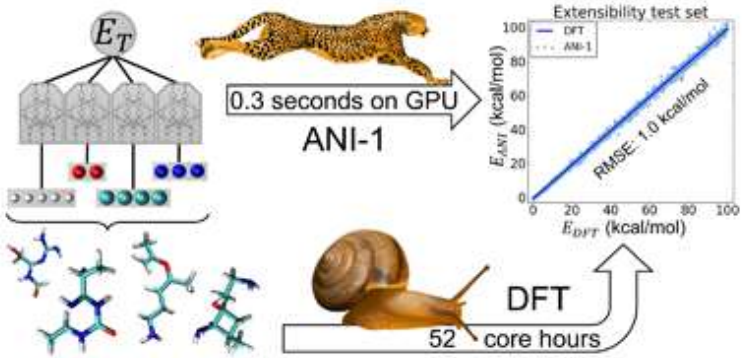
HPC Computing:

XSEDE Extreme Science and Engineering Discovery Environment

Los Alamos NATIONAL LABORATORY EST. 1943

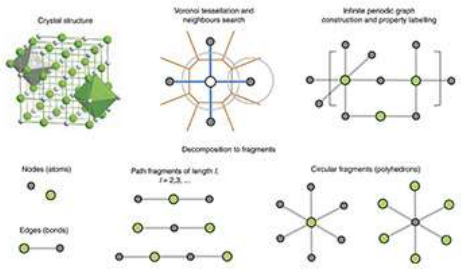
Open Science Grid

NVIDIA



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

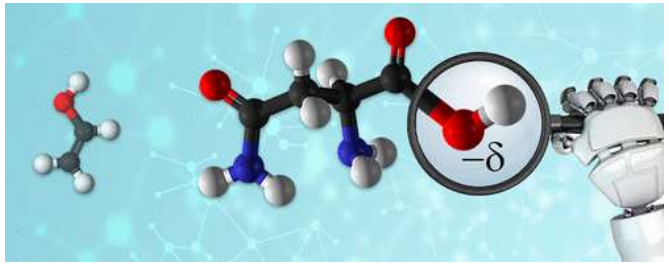
Chem. Sci., **2017**, 8, 3192-3203



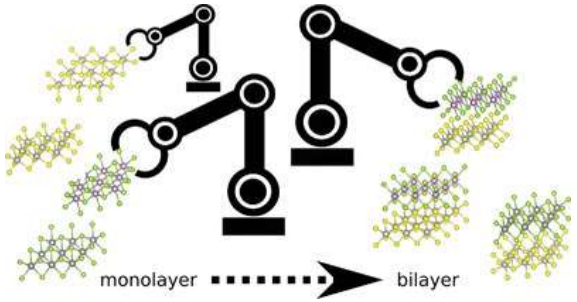
Nature Commun. **2017**, 8, 15679

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

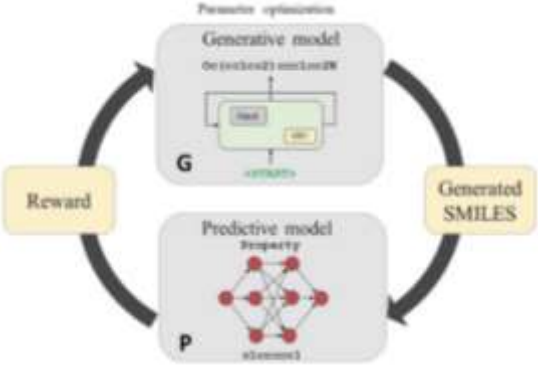
AFLOW
Automatic - FLOW for Materials Discovery



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

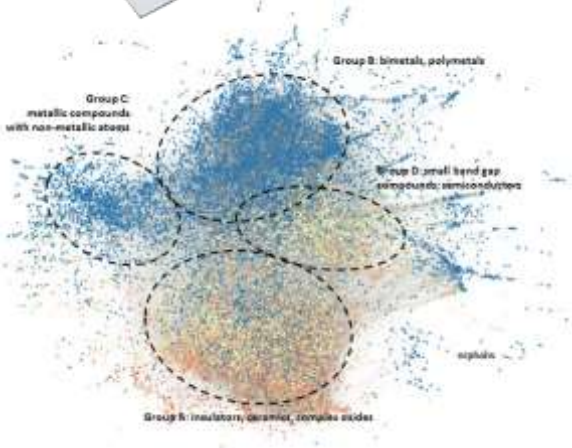


Adv. Theory Simul., **2019**, 2: 1800128

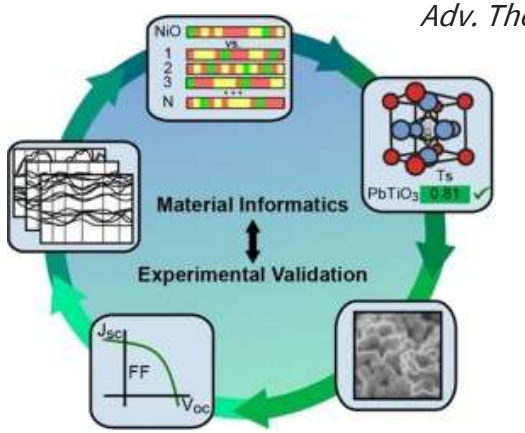


ACS Med. Chem. Lett. **2018**. 9, 1065–1069

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

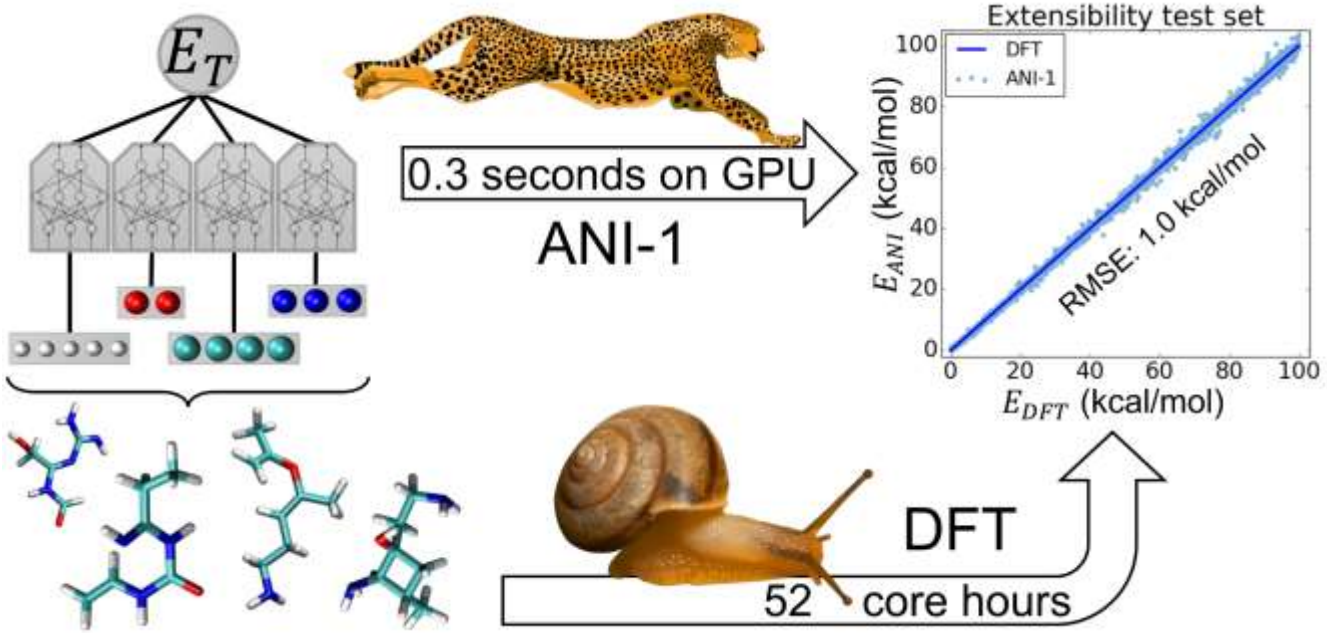


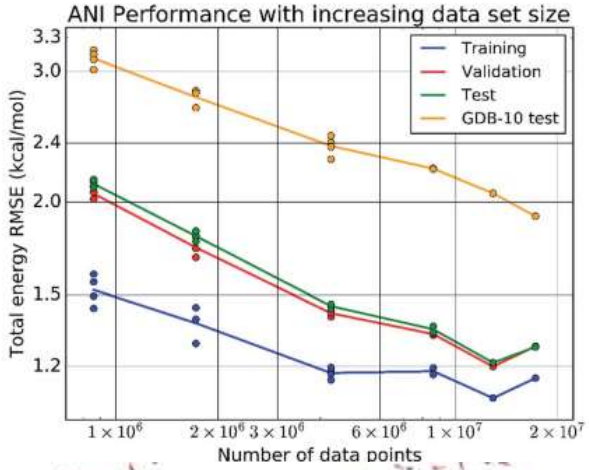
Chem. Mater., **2015**, 27, 735-742.

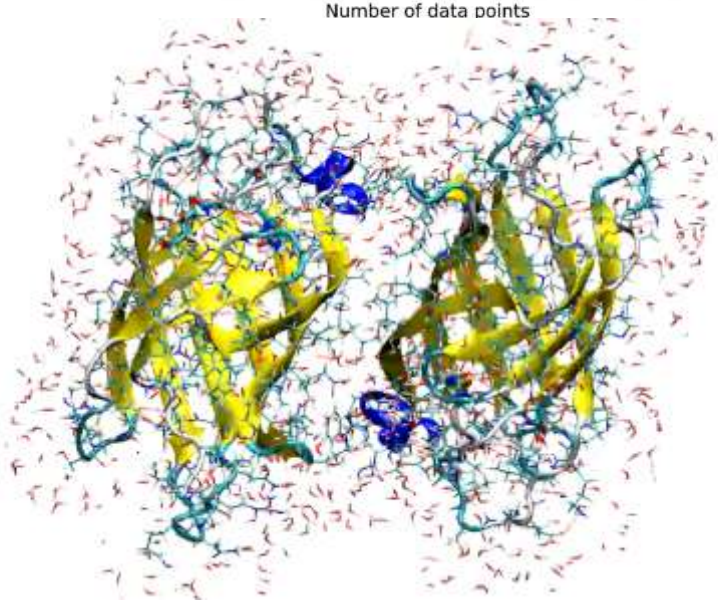


Materials Discovery. **2017**, 6, *9-16*

Fast, accurate, transferable and extensible neural network potentials





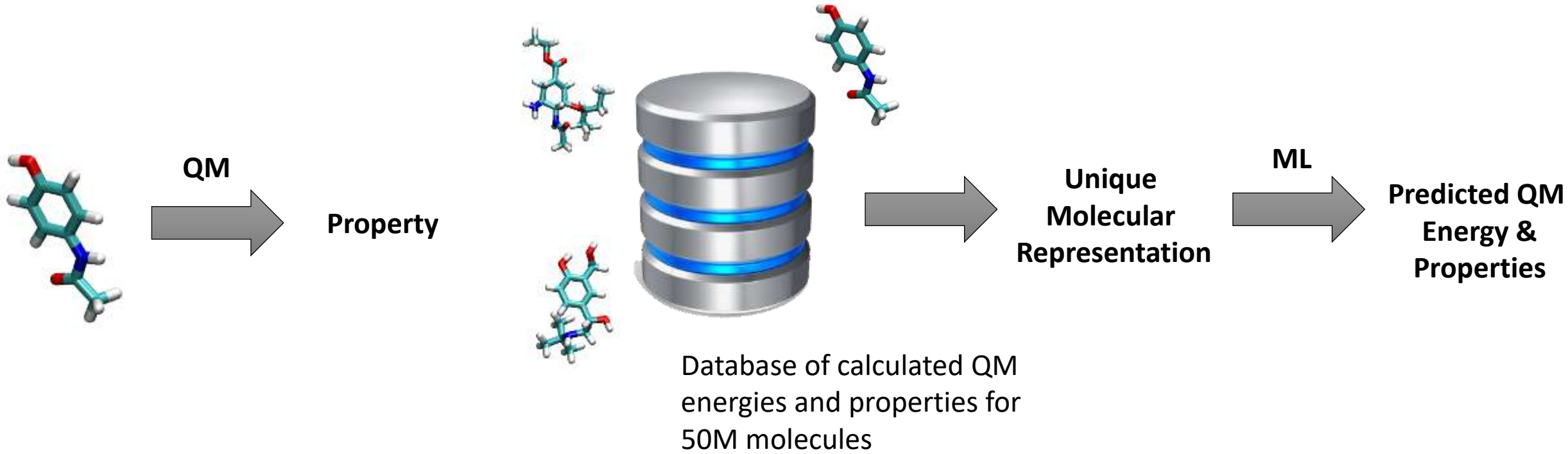


ANI1: 20M DFT calculations (CHNO)
ANI-1x: 5M DFT calculations (CHNO)
ANI-1ccx: 0.5M CCSD(T)/CBS (CHNO)
ANI-2x, AIMNet: 9M DFT calculations (CHNOSFCl)

Current development: more chemical elements, charged molecules.

Smith, Justin S., Olexandr Isayev, and Adrian E. Roitberg. "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost." *Chemical science* 8.4 (2017): 3192-3203.
Smith, Justin S., et al. "Less is more: Sampling chemical space with active learning." *The Journal of chemical physics* 148.24 (2018): 241733.
Smith, Justin S., et al. "Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning." *Nature communications* 10.1 (2019): 2903.
Zubatyuk, Roman et al., “Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network.” Sci. Adv. 2019, 5 (8), eaav6490.

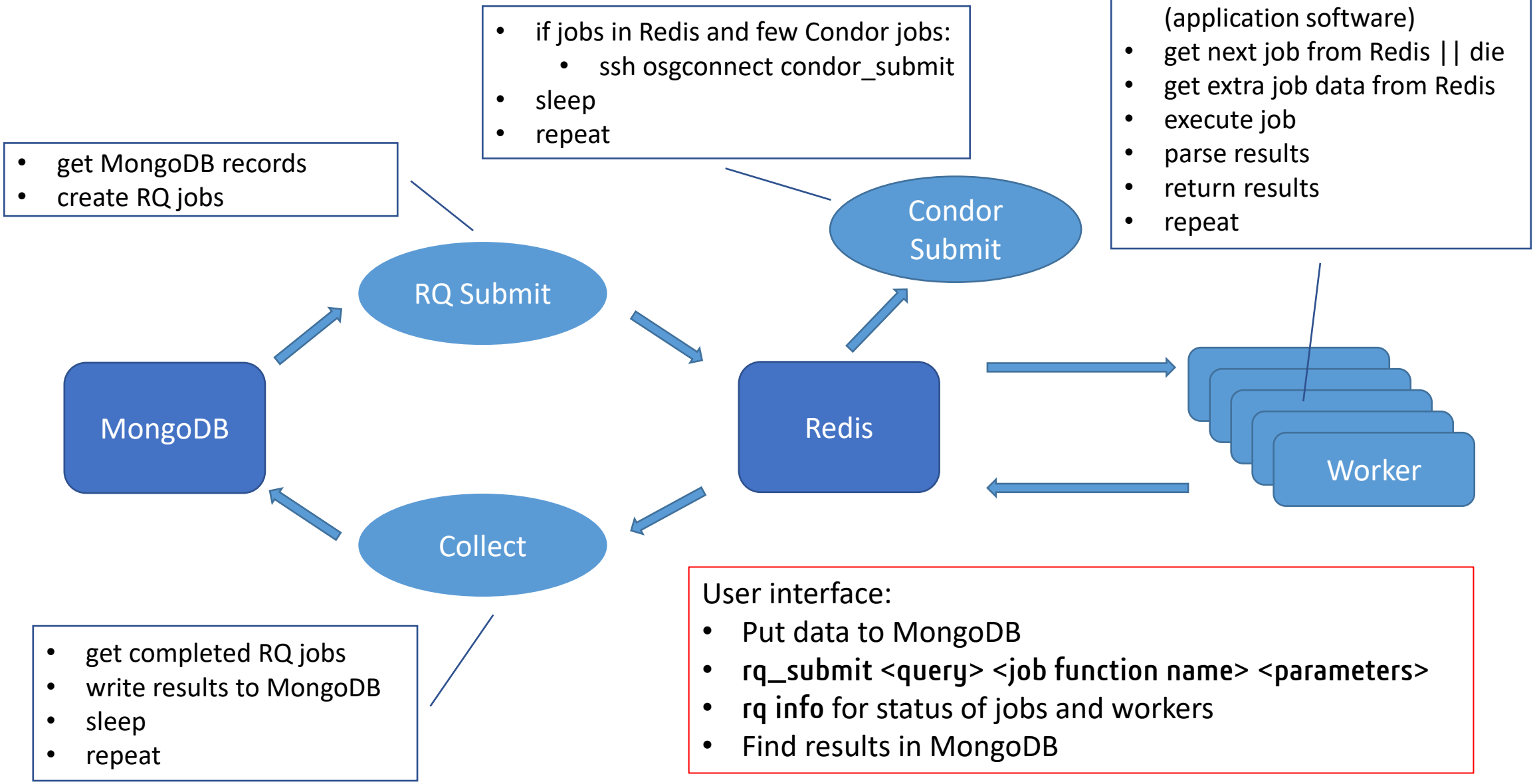
Supervised machine learning on quantum-chemical data



- Traditional quantum mechanics: Slow calculations, one molecule at a time
- QSAR: Statistical modeling of historical experimental data
- Now we could use accumulated historical QM data to train a statistical models that can accurately predict results of quantum mechanics

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

High-throughput calculations on OSG



Current Features of ANI software

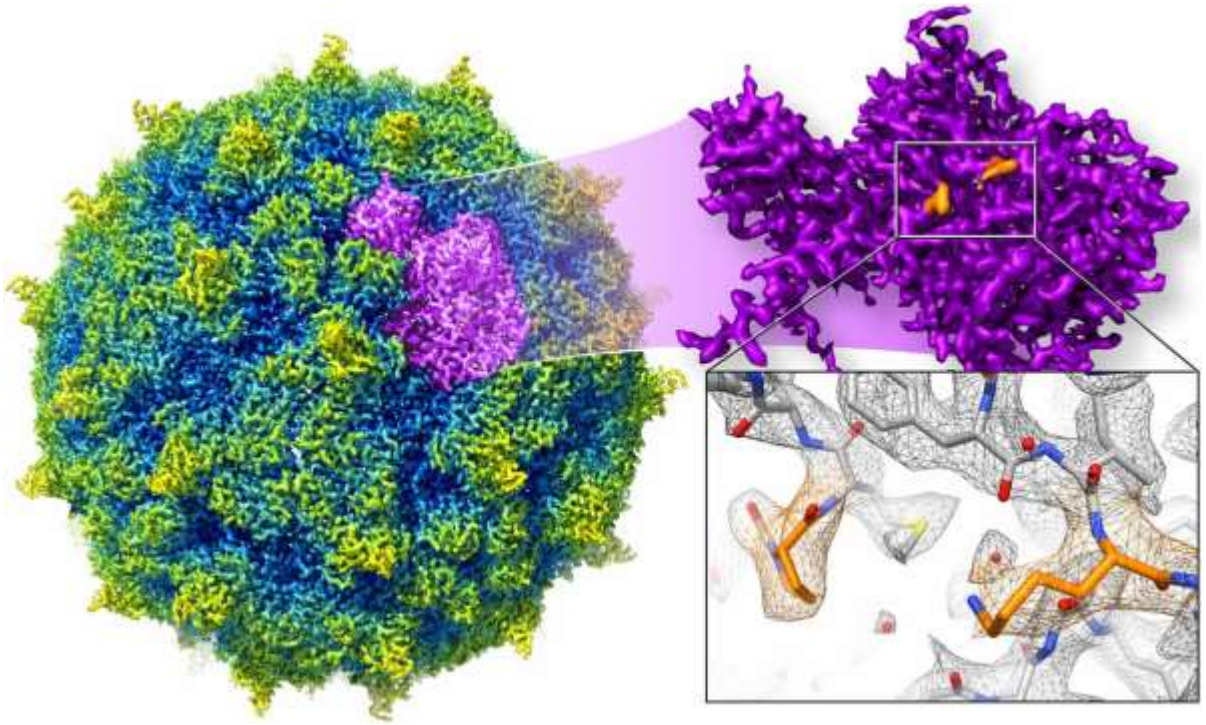
Methods

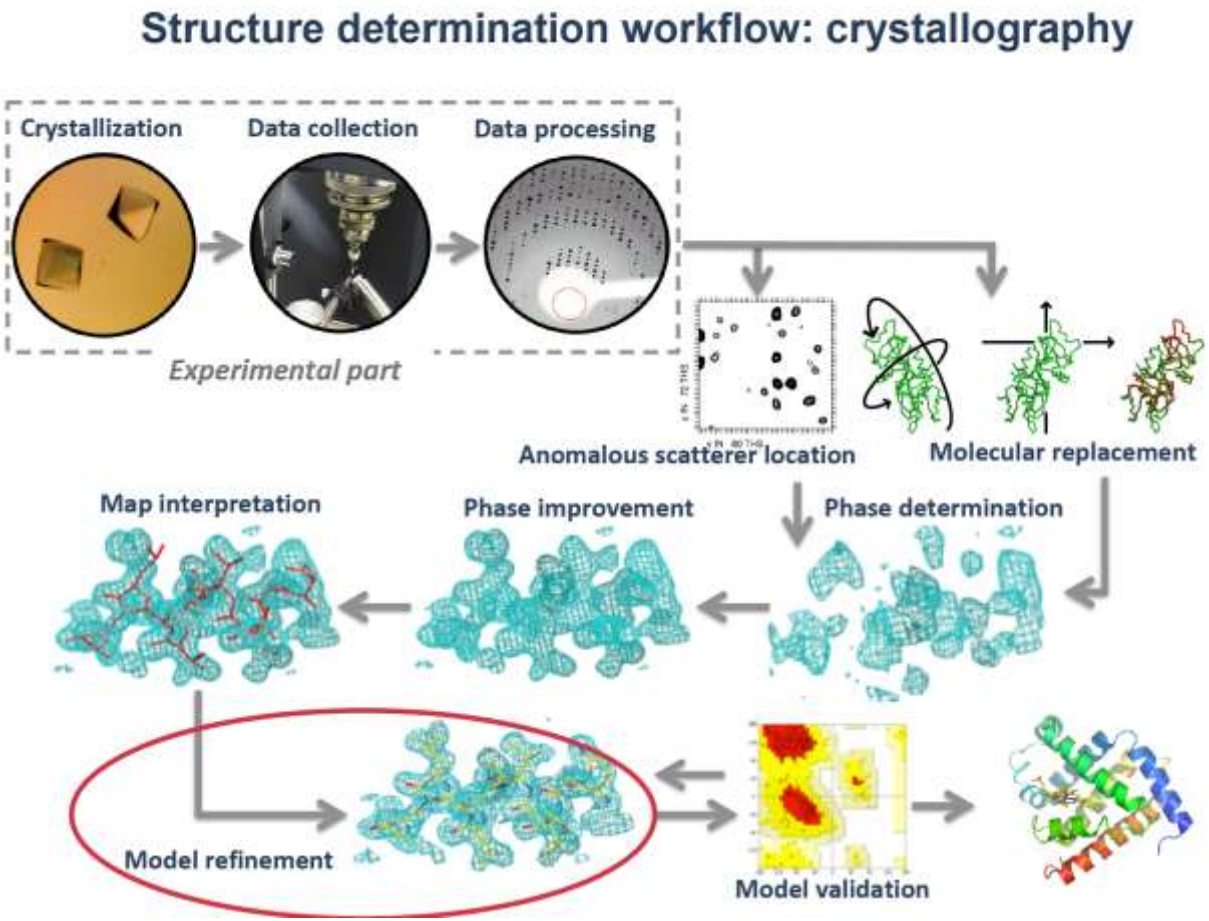
- Currently available elements: CHNOSFCl
  - In progress: P, Br, I, B, Si, Se, Na…
  - Internal prototype with 14 elements
- Two levels of theory:
  - ωB97M/Def2-TZVPP and
  - CCSD(T)*/CBS

Applications

- Geo opt, analytic hessian, MD: NVE, NVT, NPT etc
- Thermochemistry in harmonic and quasi-harmonic approximations
- Accurate conformational search in gas and continuum dielectric
- Full implementation of PBC, domain decomposition on multiple GPUs
- Stress tensor and cell optimization

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





Slide courtesy of Pavel Afonine & PHENIX Q|R team

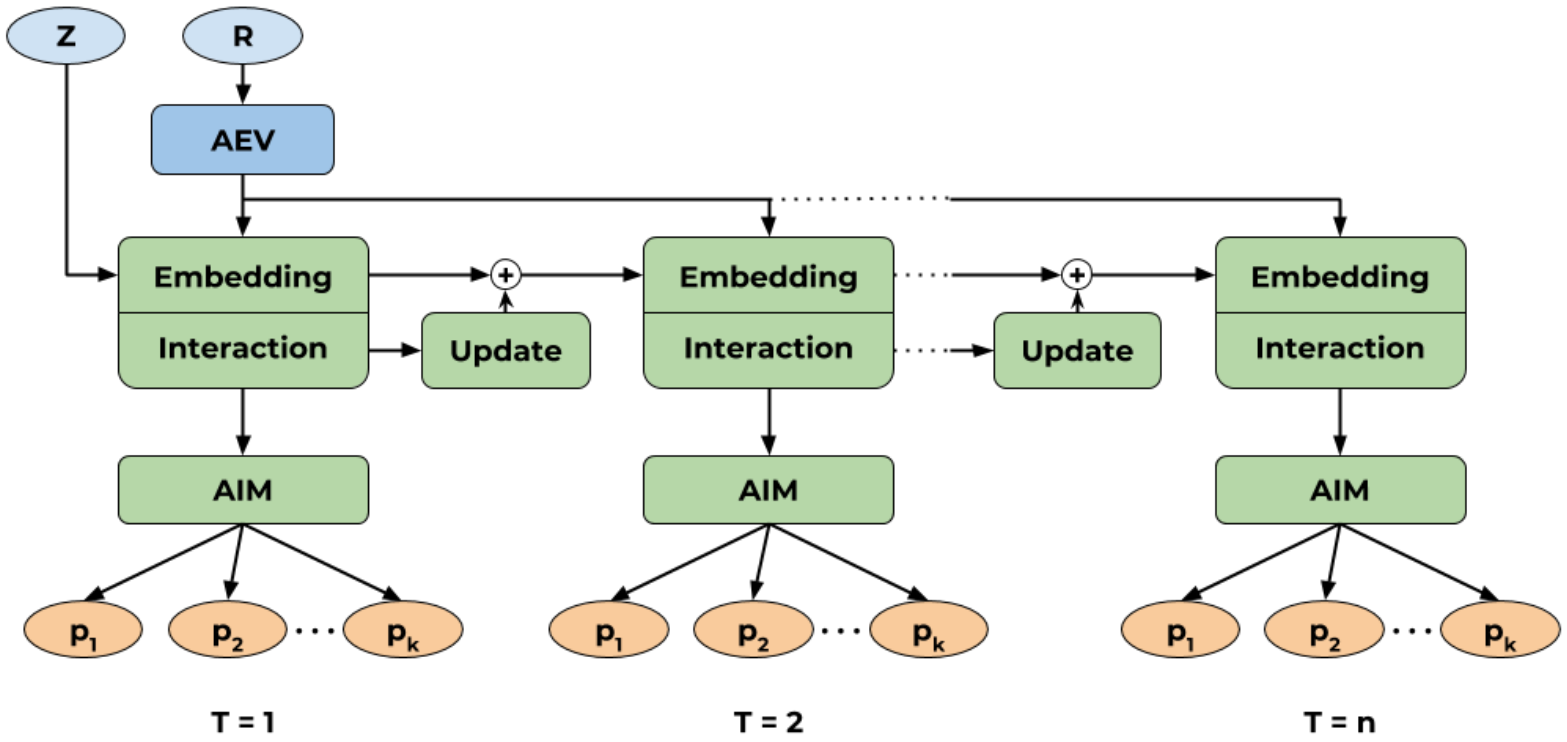
Phenix

Rethinking Network Architecture: AIMNet

Atoms-in-molecules neural net

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

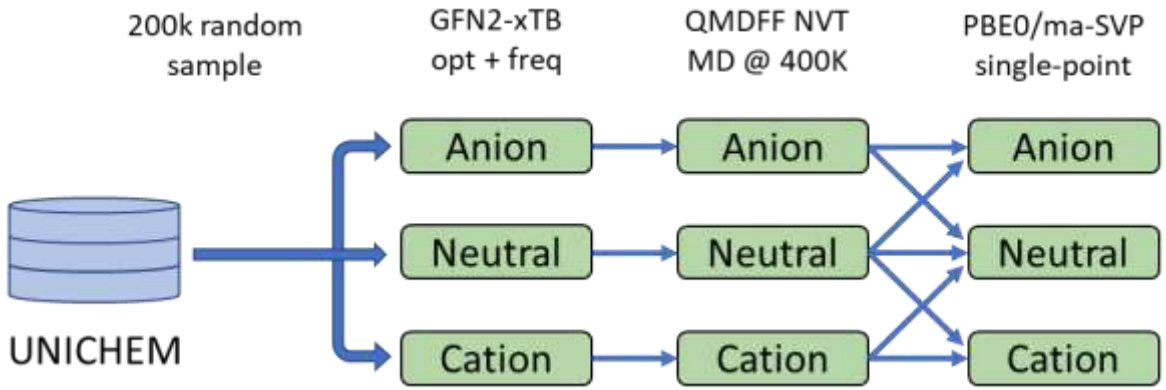
Multimodal and
multi-task learning: gas phase energy, charges, atomic volumes, continuum solvent (SMD) Correction

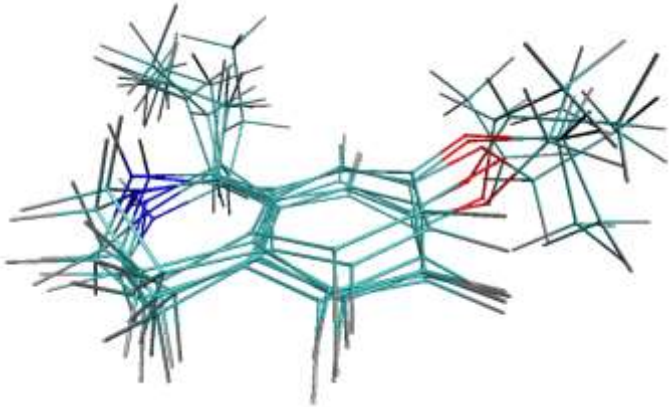


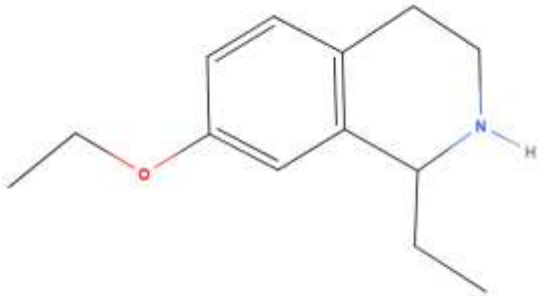
Deep NN network, AIMNet with T=3:
33 hidden layers, ~1M parameters

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 Advances*, **2019**, 5, eaav6490 ChemRxiv:10.26434/chemrxiv.7151435

Ions dataset generation

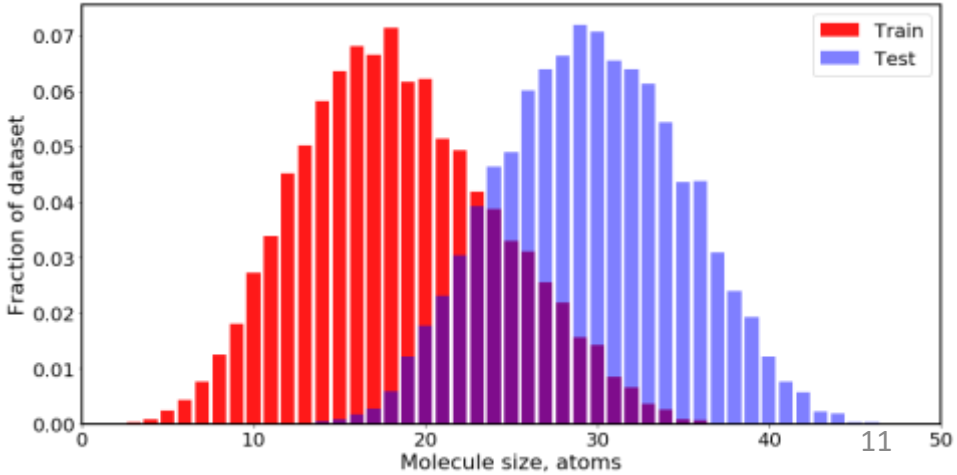






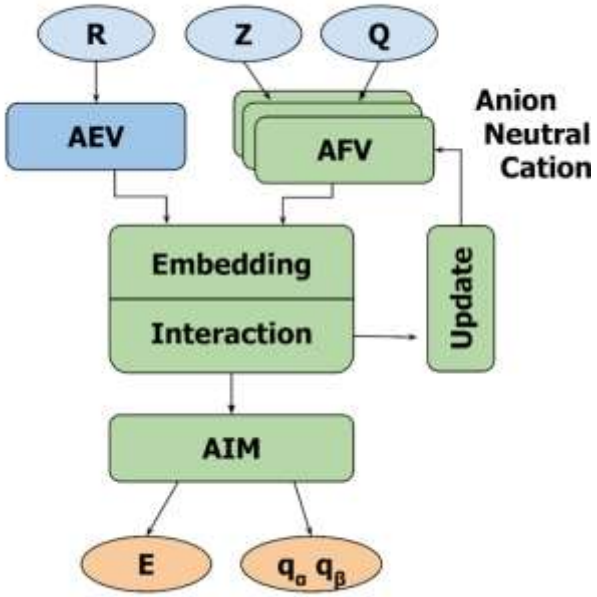
Train and test dataset sizes

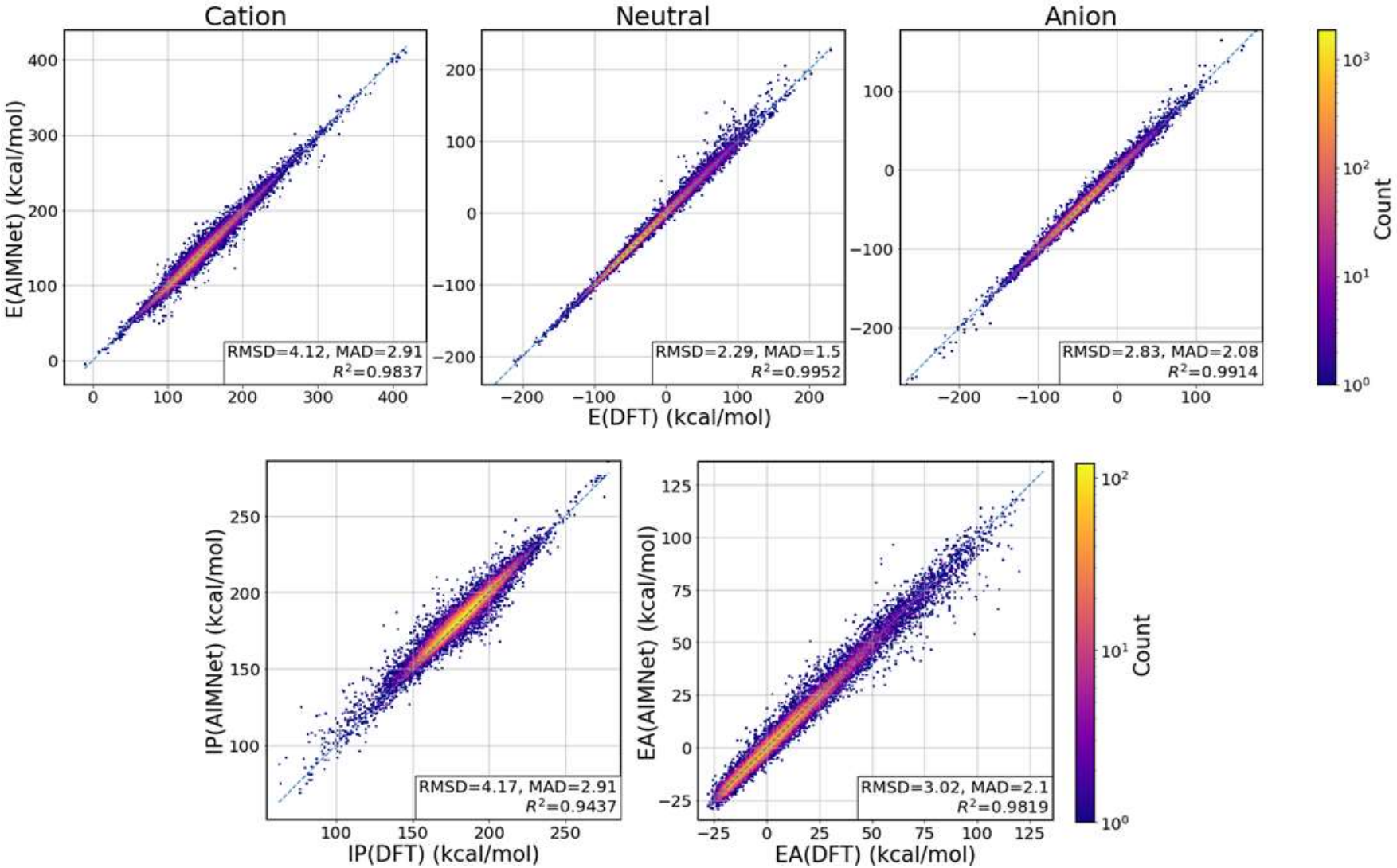
	Anions	Neutrals	Anions
Molecules			
Ions-12	154,381	167,103	155,981
Ions-16	13,607	16,138	13,860
Conformers			
Ions-12	1,896,354	3,273,755	2,150,019
Ions-16	75,654	148,543	85,741



Teaching neural network to attach and detach electrons from molecules

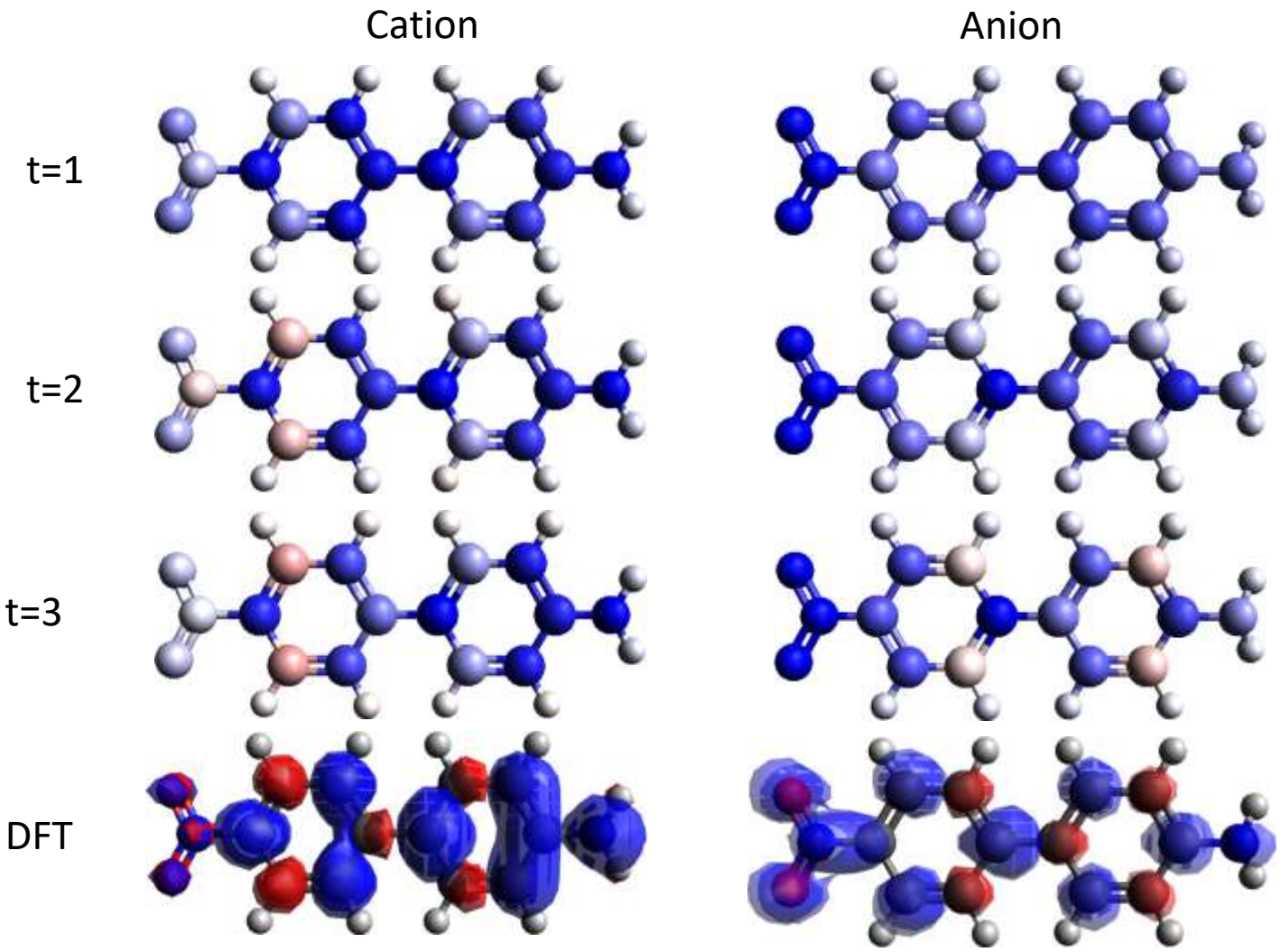
AIMNet-ME

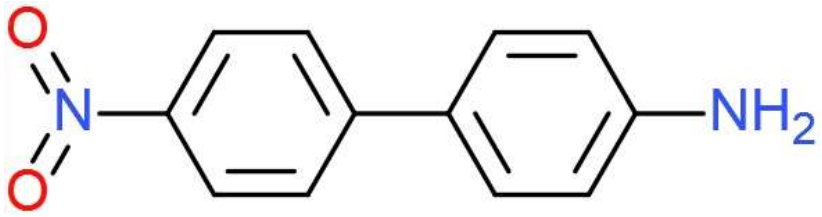




Zubatyuk, Roman; Smith, Justin; Nebgen, Benjamin T.; Tretiak, Sergei; Isayev, Olexandr (2020): Teaching a Neural Network to Attach and Detach Electrons from Molecules. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.12725276.v1

Implicit Spin-Charge Equilibration





Zubatyuk, Roman; Smith, Justin; Nebgen, Benjamin T.; Tretiak, Sergei; Isayev, Olexandr (2020): Teaching a Neural Network to Attach and Detach Electrons from Molecules. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.12725276.v1

Conceptual DFT: reactivity indices

Chemical potential $\mu = \left(\frac{\partial E}{\partial N}\right)_V \approx -\frac{1}{2}(IP + EA)$

Molecular hardness $\eta = \frac{1}{2}\left(\frac{\partial^2 E}{\partial N^2}\right)_V \approx \frac{1}{2}(IP - EA)$

Electrophilicity index $\omega = {}^{\mu^2}/_{2\eta}$

Fukui index $f(r) = \left(\frac{\partial \rho(r)}{\partial N}\right)_V$

$f_a^- = q_C - q_N$; $f_a^+ = q_N - q_A$; $f_a^0 = \frac{1}{2}(q_C + q_A)$

Atomic philicity indexes $\omega_a^k = \omega f_a^k$

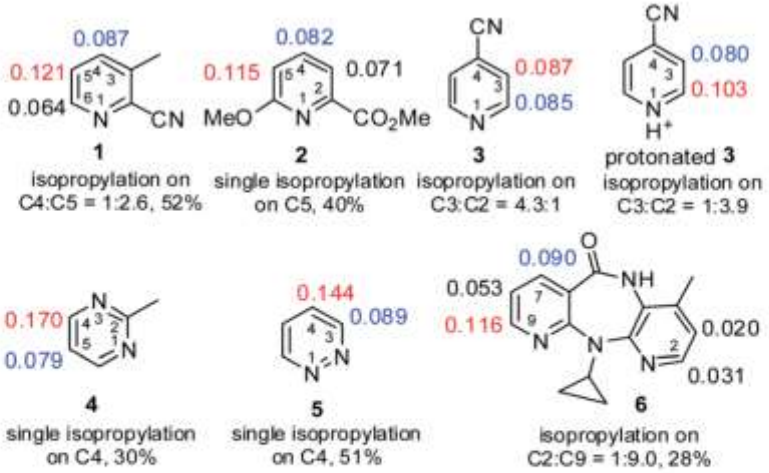
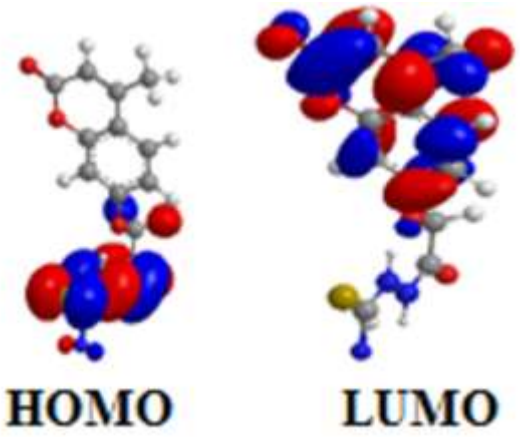
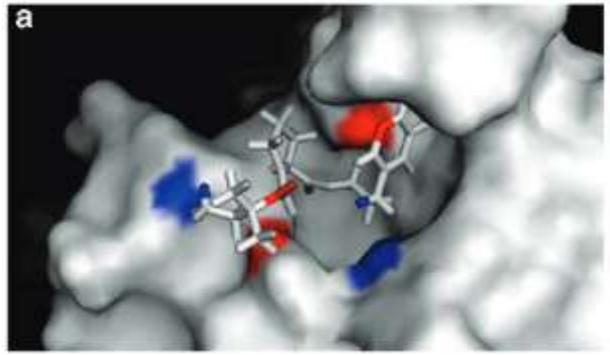


Fig. 1 Calculated f_A^+ of carbon atoms on representative nitrogen containing arenes (highest value in red, second highest in blue) and the reported radical alkylation sites with ratio or yield.

Ma, Y.; Liang, J.; Zhao, D.; Chen, Y.-L.; Shen, J.; Xiong, B. Condensed Fukui Function Predicts Innate C–H Radical Functionalization Sites on Multi-Nitrogen Containing Fused Arenes. *RSC Adv.* **2014**, *4* (33), 17262–17264. https://doi.org/10.1039/C4RA01853B.





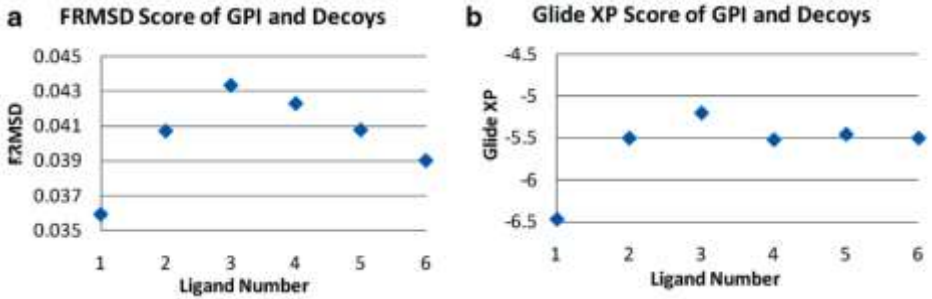
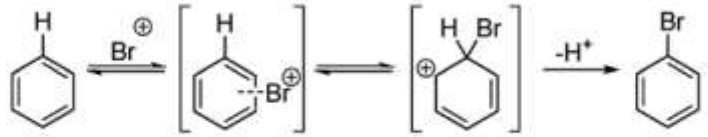


Figure 4. (a) FRMSD scores of the best scoring poses of the GPI molecule and decoy ligands to FKBP12. (b) Glide XP scores. The ligands are numbered as (1) GPI-1046, (2) alpinetin, (3) pinocembrin, (4) pinocembrine chalcone, (5) pinostrobin, and (6) pinostrobin chalcone. Both scores correctly discriminate the known binder from the decoy ligands.

Faver, J.; Merz, K. M. Utility of the Hard/Soft Acid−Base Principle via the Fukui Function in Biological Systems. *J. Chem. Theory Comput.* **2010**, *6* (2), 548–559. https://doi.org/10.1021/ct9005085.

Fast and Accurate prediction of regioselectivity for electrophilic aromatic substitution using reactivity indexes



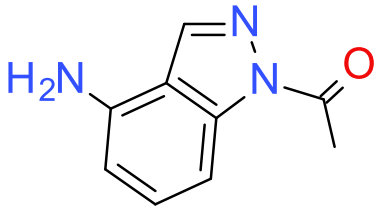
Scheme 1 General EAS bromination mechanism.

Accuracy: 0.900

ω, ω_a^e

AIM

AIMNet



TPSSh/def2-SVP

Opt

Charge

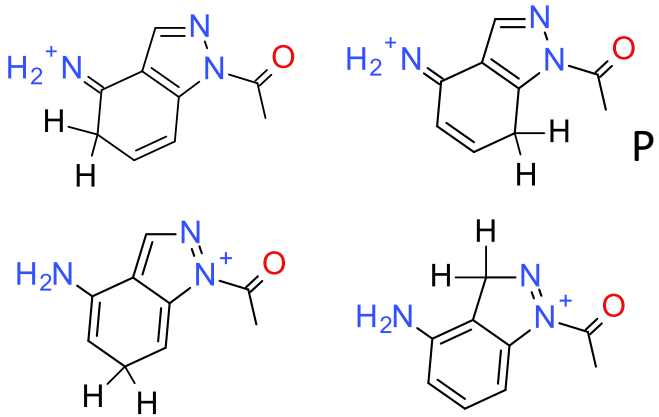
Sum of bond orders

SAS

Fukui function f^-

Accuracy: 0.899

No reaction intermediate complex calculations required



PM3/COSMO

RegioSQM

Accuracy: 0.931

Kromann, Jimmy C., Jan H. Jensen, Monika Kruszyk, Mikkel Jessing, and Morten Jørgensen. "Fast and accurate prediction of the regioselectivity of electrophilic aromatic substitution reactions." *Chemical science* 9, no. 3 (2018): 660-665.

https://github.com/jensengroup/RegioSQM

Use the ANI-1x/2x potential:

ASE Python & CUDA library
Available at: https://github.com/isayev/ASE_ANI

PyTorch
Available at: https://github.com/aiqm/torchani

Plugins: OpenMM, AMBER (dev), NAMD
Coming soon: Tinker-HP, LAMMPS, SCM-ADF

Use the AIMNet:

Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Science Advanced*, **2019**, 5, eaav6490

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

ANI-1x & 1ccx datasets:

The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. Sci Data 7, 134 (2020). https://doi.org/10.1038/s41597-020-0473-z
Available at: https://github.com/aiqm/ANI1x_datasets

Users:

academic labs:

- Stanford
- UNC
- Pitt
- USF
- NCSU
- Barcelona
- Helsinki
- Paris
- Tel Aviv

Government labs, companies etc.