PSC

Carnegie Mellon University

University of Pittsburgh

Bridges-2 Webinar

AIMNet2: Foundation Neural Network Potential for Molecules and Reactions

Olexandr Isayev
Carnegie Mellon University

September 16, 2024

Pittsburgh Supercomputing Center

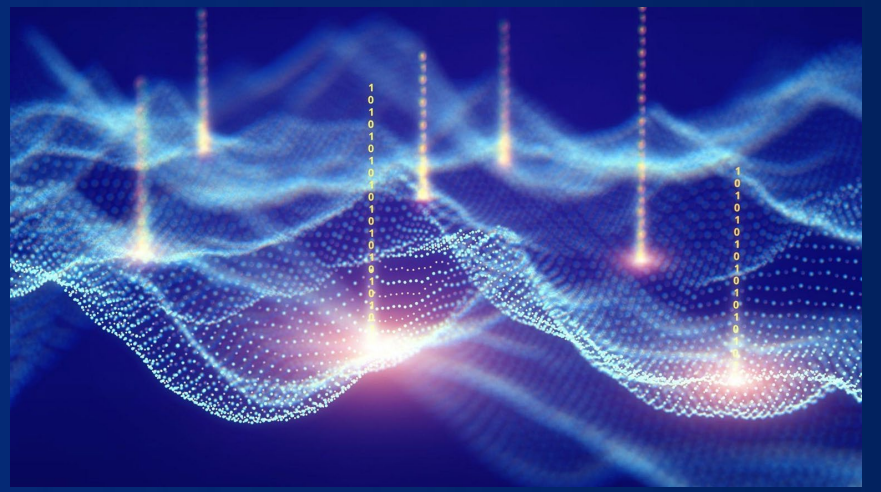
enabling discovery since 1986

The **Pittsburgh Supercomputing Center (PSC)** provides advanced research computing capability, education, and expertise to the national research community.

Since 1986, PSC has provided university, government, and industry researchers with access to some of the most powerful systems available for high-performance computing, enabling discovery across all fields of science.

OUR AREAS OF EXPERTISE

- high-performance and data-intensive computing
- data management technologies
- software architecture, implementation, and optimization
- enabling ground-breaking science, computer science, and engineering
- user support for all phases of research and education
- STEM outreach in data science, bioinformatics, and coding



Welcome!

Hewlett Packard Enterprise is delivering *Bridges-2*

Bridges-2 Leadership Team

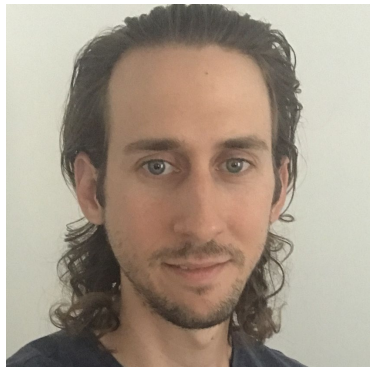
Sergiu Sanielevici
PI & Dir. Support
for Sci. Apps.

Robin Scibek
Dir. Comms.
co-PI

Paola Buitrago
Dir. AI & Big Data
co-PI

Edward Hanna
Dir. Systems & Ops.
co-PI

Tom Maiden
User Services Mgr.
co-PI



Stephen Deems
Project Manager

Andrew Adams
Information Security
Officer

Bridges-2 Webinars

- A forum for the Bridges-2 community to learn and share ideas and achievements: Bridges-2 Webinar series | PSC
- Topics and speakers of interest to work that is being done, or that may be done in future.
- Please suggest future speakers (including from your own team) and/or topics (including your own)!

Just email: sergiu@psc.edu

Introducing today's presenter: Olexandr Isayev

Olexandr is a full-time professor in the Department of Chemistry at Carnegie Mellon University. In 2008, Olexandr received his Ph.D. in computational chemistry. He was a Postdoctoral Research Fellow at Case Western Reserve University and a scientist at the government research lab. Before CMU, he was a faculty member at UNC Eshelman School of Pharmacy, the University of North Carolina at Chapel Hill. Olexandr is a 2023 Scialog Fellow and Associate Editor for the ACS Journal of Chemical Information and Modeling. The research in his lab focuses on connecting artificial intelligence (AI) with chemical sciences.

Q&A Logistics

- **We abide by https://support.access-ci.org/code-of-conduct**
- All of us except Olexandr will be muted during his presentation.
- Please type your questions into the Zoom chat.
- We may be able to address some questions in the chat while Olexandr is presenting.
- When Olexandr finishes his presentation, he will answer questions live during the final ~10 minutes of this webinar.

Carnegie Mellon University

2024 PSC Bridges-2 webinar

AIMNet2: Foundation neural network potentials for molecules and reactions

@olexandr

Olexandr Isayev

Department of Chemistry, Carnegie Mellon University

olexandr@olexandrisayev.com

http://olexandrisayev.com

Carnegie Mellon University

Dylan Anstine
Polina Avdiunina
Hatice Gokcan
Phil Gusev
Zhen (Jack) Liu
Kamal Singh Nayal
Shuhao Zhang
Tanya Zubatiuk
Roman Zubatyuk

Isayev Lab circa 2022

University of Notre Dame

Brett Savoie

Julia Laskin
Jianguo Mae

Purdue University

Funding:

BASF We create chemistry

NSF

CHE-2154447
DMR-2323749

NSF CCI C-CAS

C-CAS Center for Computer Assisted Synthesis

ONR MURI
N00014-21-1-2476

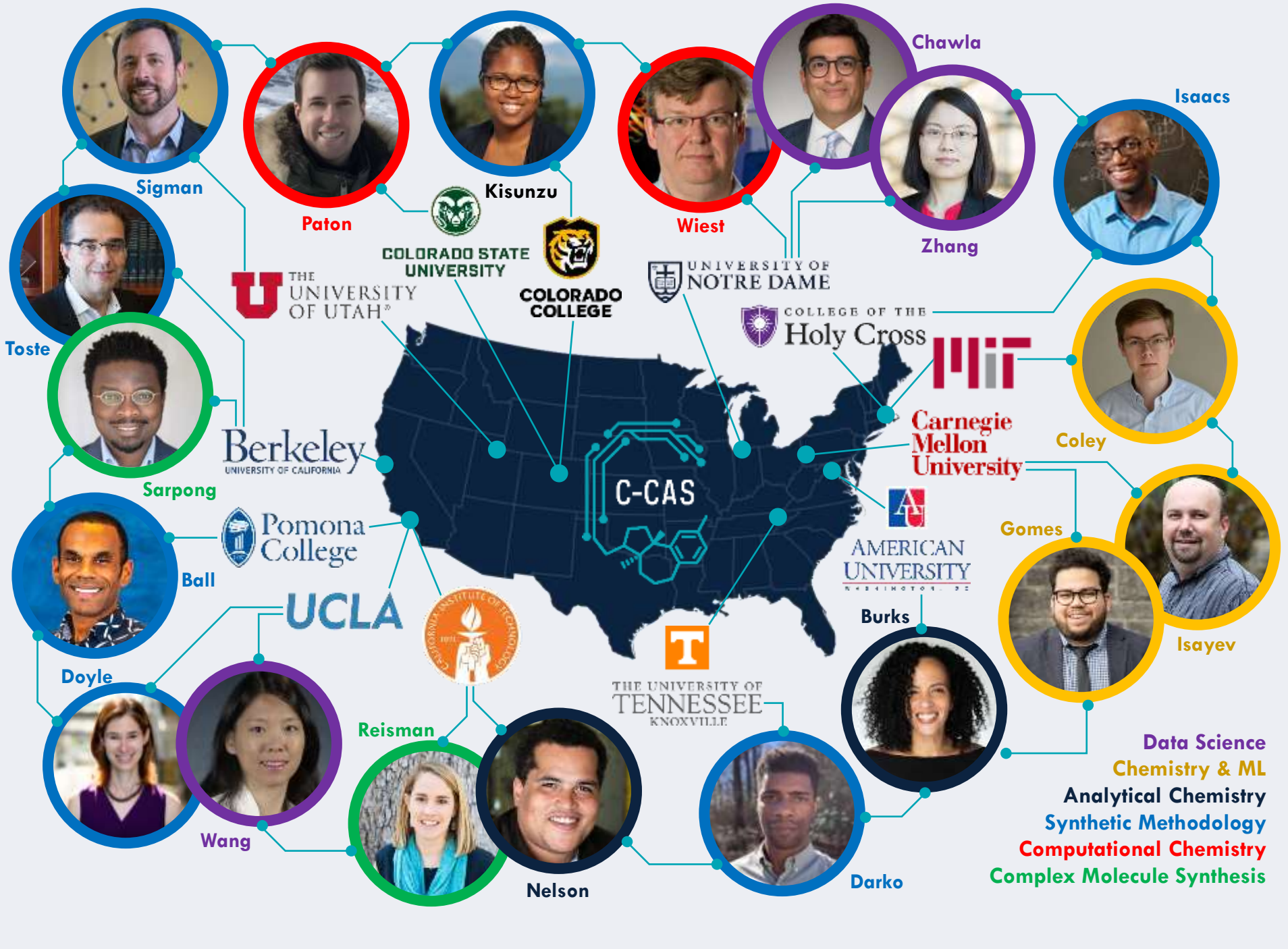
Department of the Navy ONR Science & Technology

NSF | ACCESS | PSC

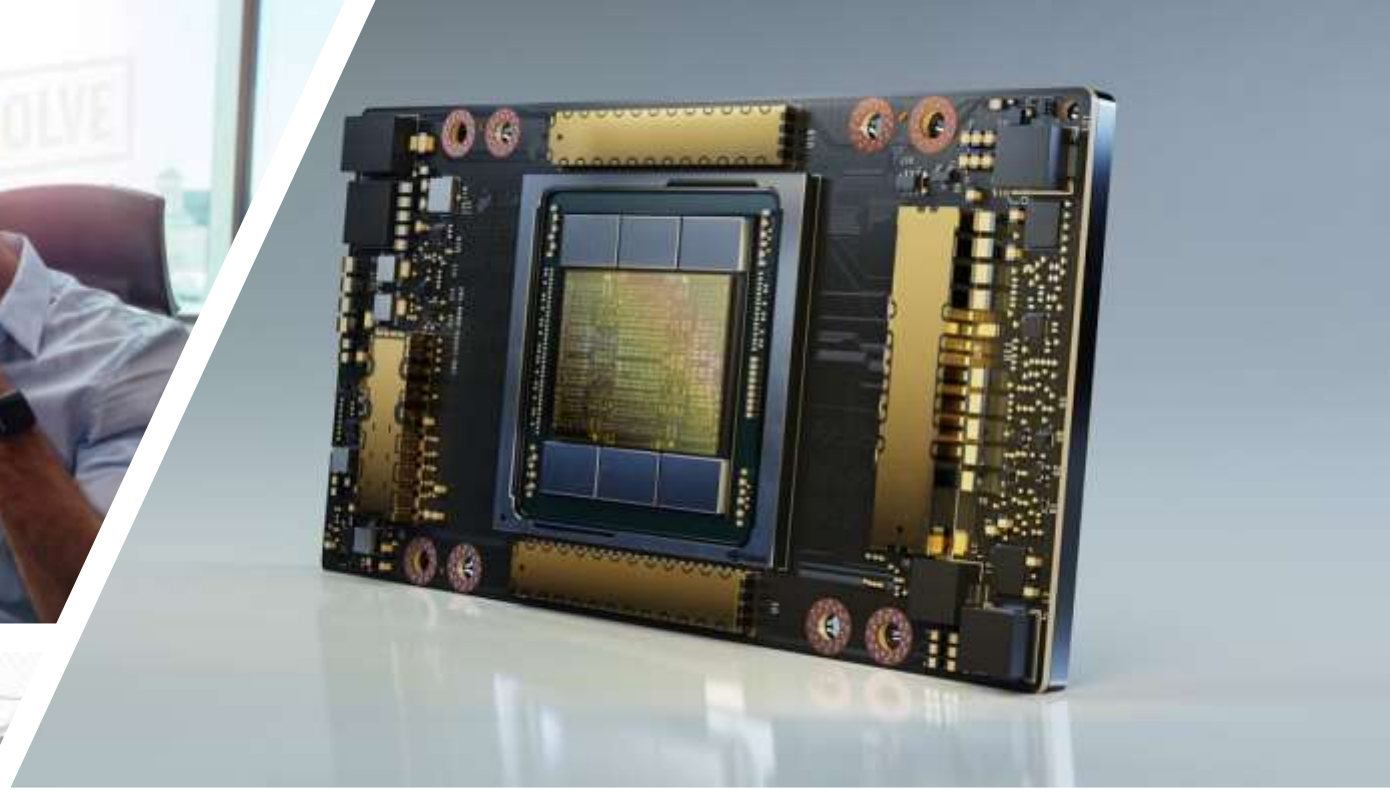
Open Science Grid

NVIDIA

Los Alamos National Laboratory EST. 1943


Sigman
Paton
Kisunzu
Wiest
Chawla
Zhang
Isaacs
Toste
Sarpong
Berkeley
UNIVERSITY OF CALIFORNIA
THE UNIVERSITY OF UTAH
COLORADO STATE UNIVERSITY
COLORADO COLLEGE
UNIVERSITY OF NOTRE DAME
COLLEGE OF THE Holy Cross
MIT
Coley
C-CAS
Carnegie Mellon University
Gomes
Isayev
Ball
Pomona College
AMERICAN UNIVERSITY
Burks
UCLA
Doyle
Wang
Reisman
Nelson
THE UNIVERSITY OF TENNESSEE KNOXVILLE
Darko
Data Science
Chemistry & ML
Analytical Chemistry
Synthetic Methodology
Computational Chemistry
Complex Molecule Synthesis

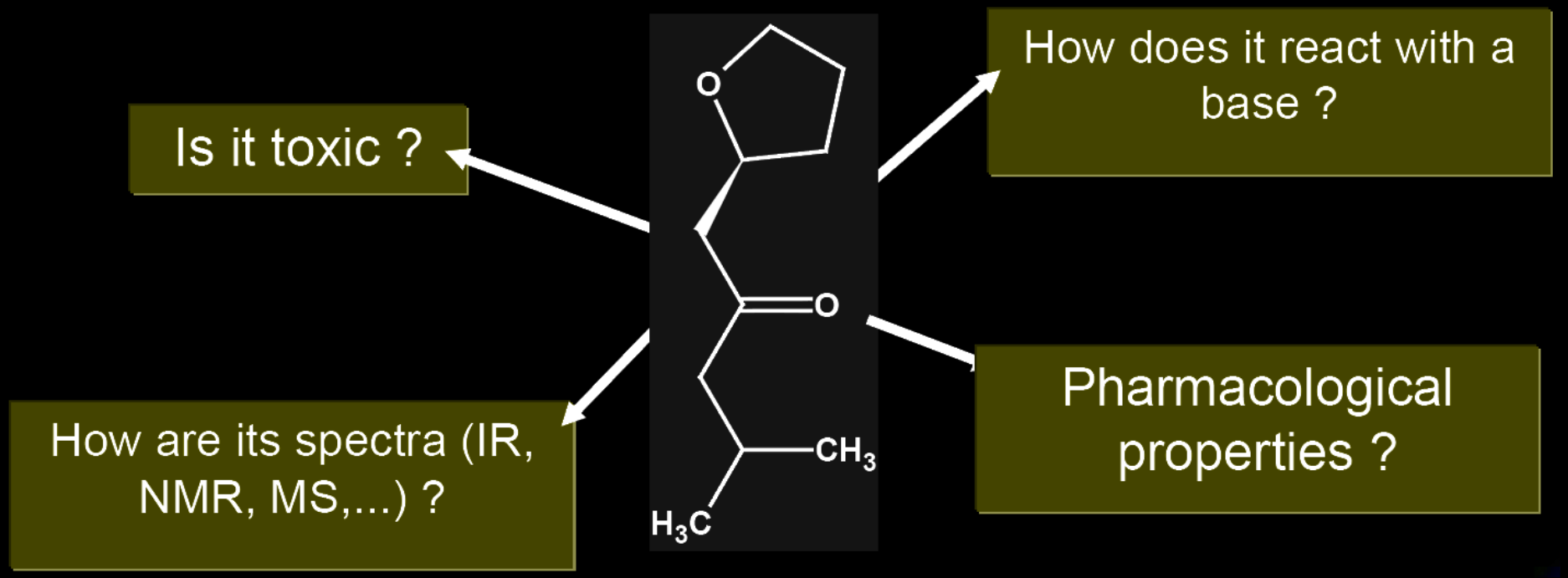
Can you imagine a Chemist?




Emerald Cloud Lab


BRIDGES
NEXT-GENERATION SUPERCOMPUTING
FOR NEW COMMUNITIES AND BIG DATA

Can a computer learn chemistry ?



What's it Good For?

Computational chemistry is a rapidly growing field in chemistry.

- Computers are getting faster.
- Algorithms and programs are maturing.

Some of the almost limitless properties that can be calculated with computational chemistry are:

- Design of novel molecules and materials
- Medicinal chemistry, drug discovery
- Vibrational frequencies, IR and Raman Spectra
- NMR spectra
- Electronic excitations and UV spectra
- Reaction rates and catalyst design
- Thermochemical and high accuracy reference data

Time Independent Schrödinger Equation

We'll be solving the Time-Independent Schrödinger Equation

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

$$\hat{H} = \hat{T} + \hat{V}$$

Your book writes this as: $\hat{H} = \hat{E}_{kinetic} + \hat{E}_{potential}$

For Many electron atoms/molecules:

$$\hat{H} = -\frac{\hbar^2}{2}\sum_{\alpha}\frac{1}{m_\alpha}\nabla_\alpha^2 - \frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 + \sum_\alpha\sum_{\alpha>\beta}\frac{Z_\alpha Z_\beta e'^2}{r_{\alpha\beta}} - \sum_\alpha\sum_i\frac{Z_\alpha e'^2}{r_{i\alpha}} + \sum_j\sum_{i>j}\frac{e'^2}{r_{ij}}$$

Nuclei kinetic energy	Electron kinetic energy	Nuclear-Nuclear repulsion	Nuclear-electron attraction	Electron-electron repulsion
$\hat{T}$		$\hat{V}$		

$$\left(\frac{-\hbar^2}{2m}\frac{d^2}{dx^2}\right)$$

where: $\nabla_i^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$

Motivation

Schrödinger Equation can only be solved exactly for simple systems.

- Rigid Rotor, Harmonic Oscillator, Particle in a Box, Hydrogen Atom

For more complex systems (i.e. many electron atoms/molecules) we need to make some simplifying assumptions/approximations and solve it numerically.

However, it is still possible to get very accurate results (and also get very crummy results).

- In general, the “cost” of the calculation increases with the accuracy of the calculation and the size of the system.

The Born-Oppenheimer Approximation

- The wave-function of the many-electron molecule is a function of electron and nuclear coordinates: $\psi(R,r)$ (R=nuclear coords, r=electron coords).
- The motions of the electrons and nuclei are coupled.
- However, the nuclei are much heavier than the electrons
  - $m_p \approx 2000\, m_e$
- And consequently, nuclei move ***much*** more slowly than do the electrons ($E=1/2mv^2$). To the electrons the nuclei appear fixed.
- Born-Oppenheimer Approximation: to a high degree of accuracy we can separate electron and nuclear motion:

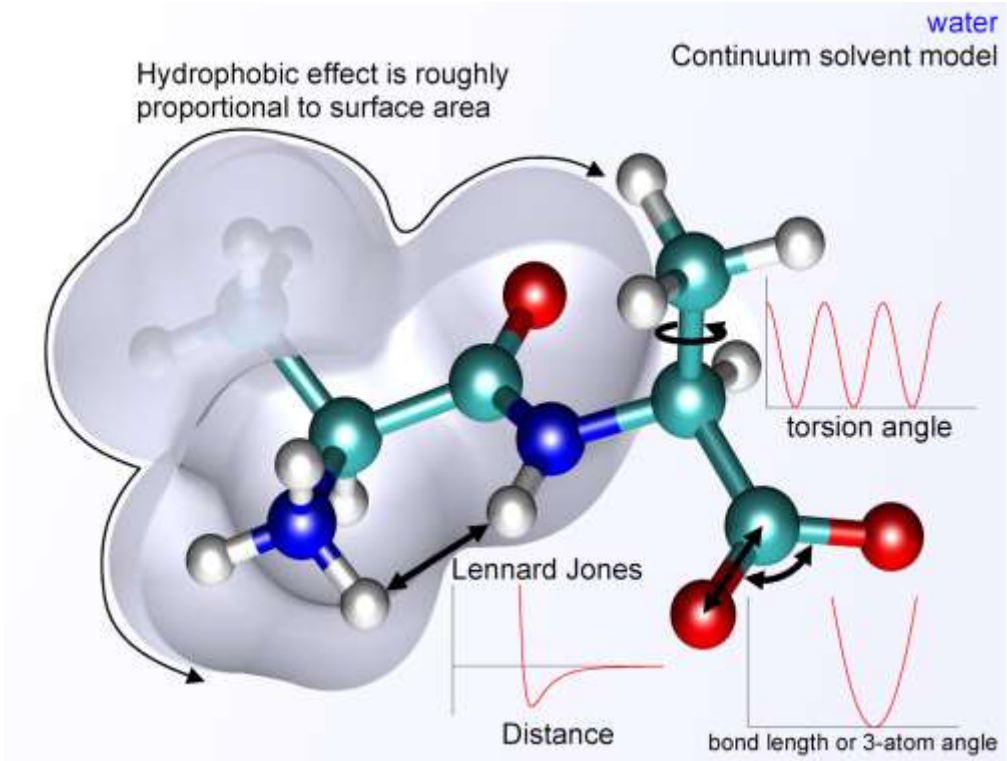
  $\psi(R,r) = \psi_{el}(r;R)\ \psi_N(R)$

Molecular mechanics/Classical force field

$$u(r^N) = \sum_{\text{bonds}} k_i (l_i - l_{i,0})^2 + \sum_{\text{angles}} k_i (\theta_i - \theta_{i,0})^2$$

$$+ \sum_{\text{torsions}} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$

$$+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right)$$



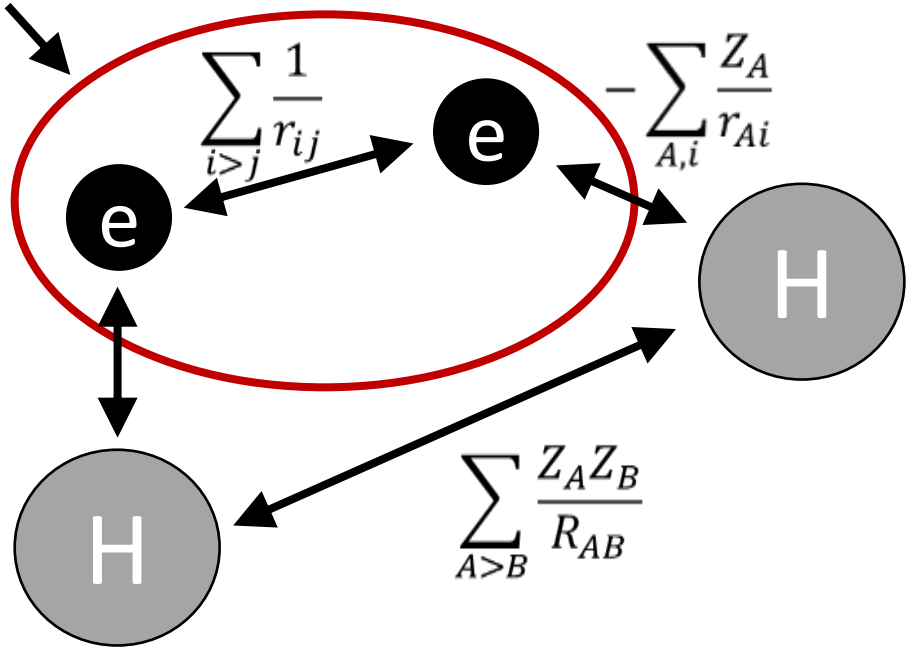
The Electronic Schrödinger Equation (QM)

$$\hat{H}\Psi(r;R) = E_{el}\Psi(r;R)$$

Electronic wave function

Electronic potential energy

$$-\frac{1}{2}\sum_i \nabla_i^2 \qquad \sum_{i>j} \frac{1}{r_{ij}} \qquad -\sum_{A,i} \frac{Z_A}{r_{Ai}} \qquad \sum_{A>B} \frac{Z_A Z_B}{R_{AB}}$$



Supervised Machine Learning Framework

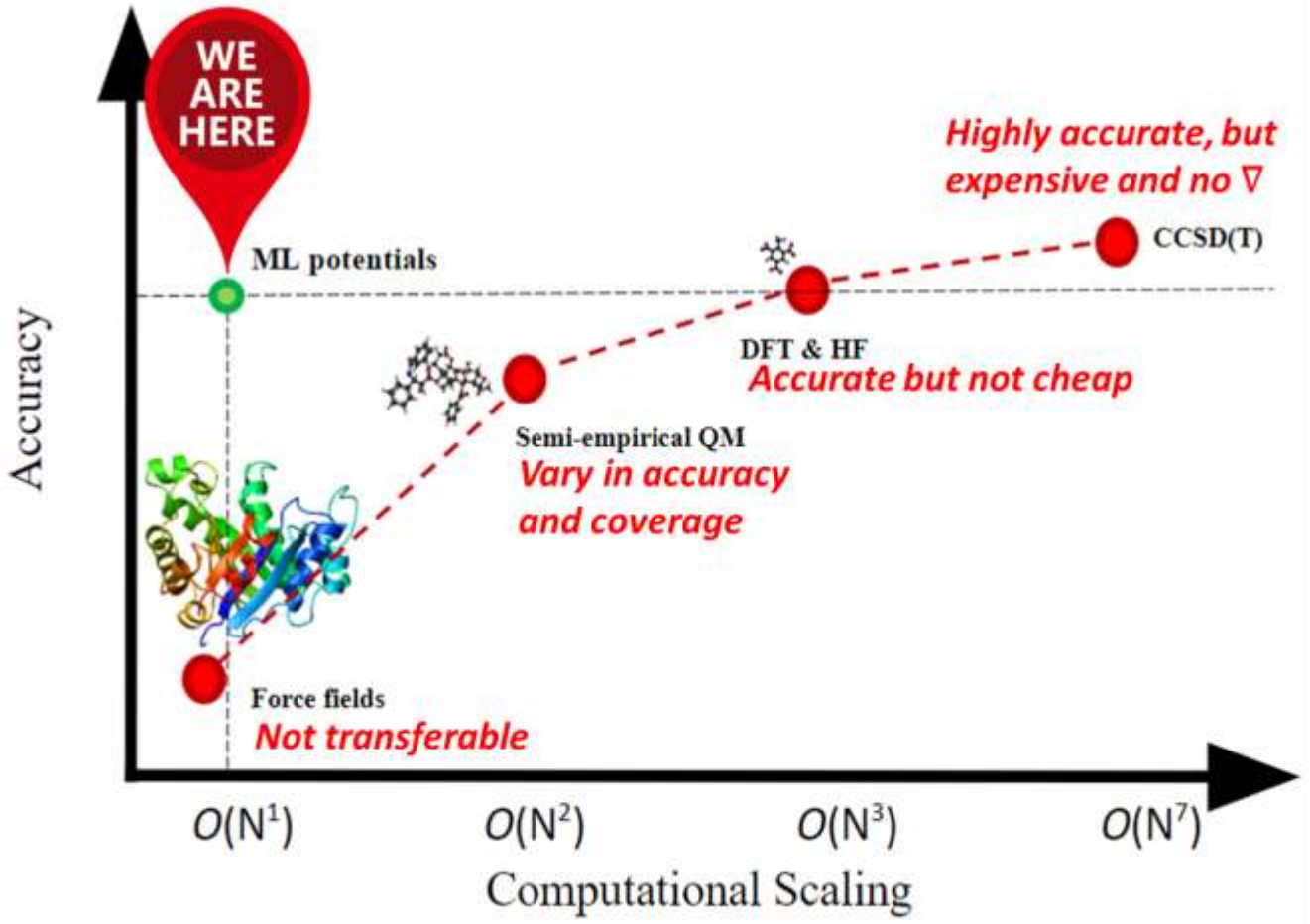
$$y = f(\mathbf{x})$$

output — prediction function — Molecular features

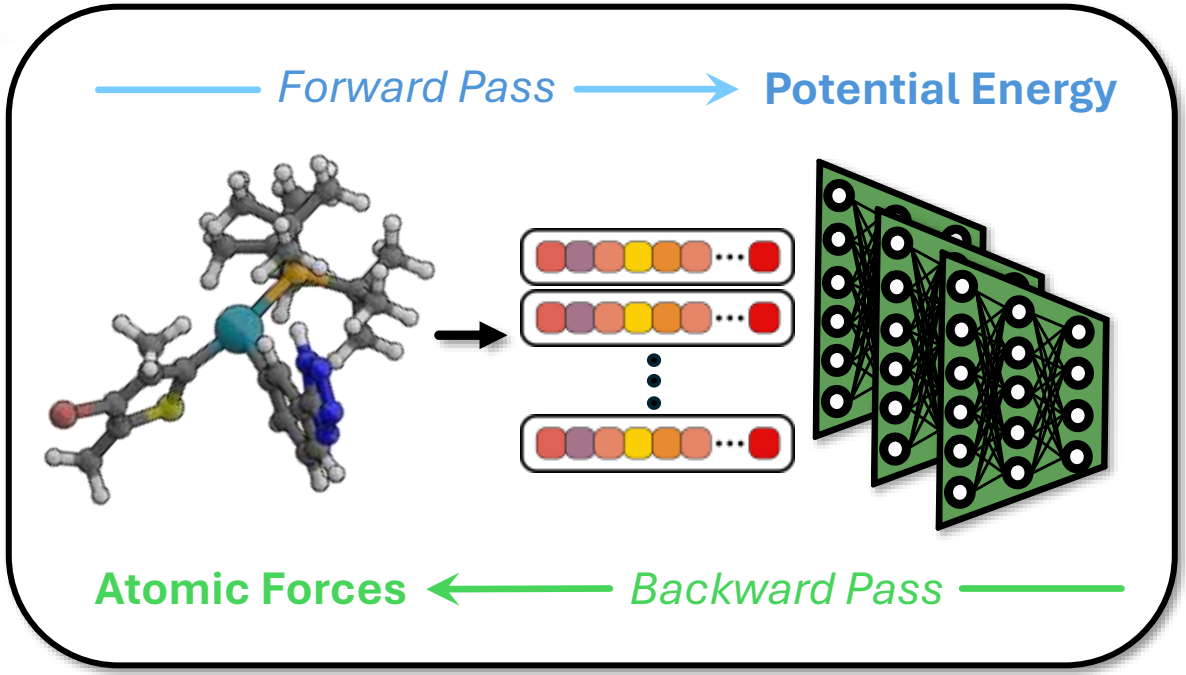
- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set
- **Testing:** apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Machine Learned Interatomic Potential (MLIP) Motivation

(Spoiler alert!)

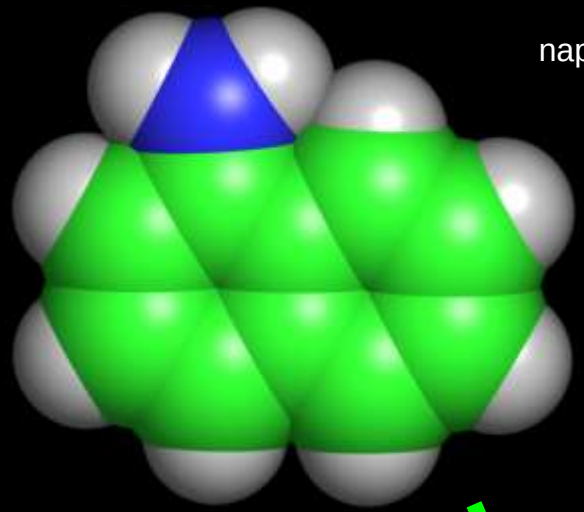


$$E = f(R_{vector})$$

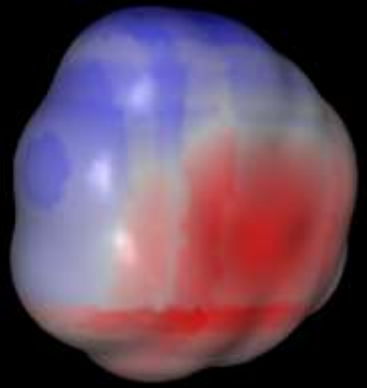


Neural Network Interatomic Potential

Structure representation

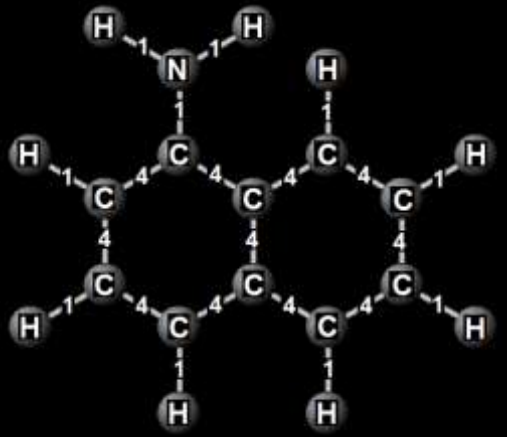


naphtalen-1-amine

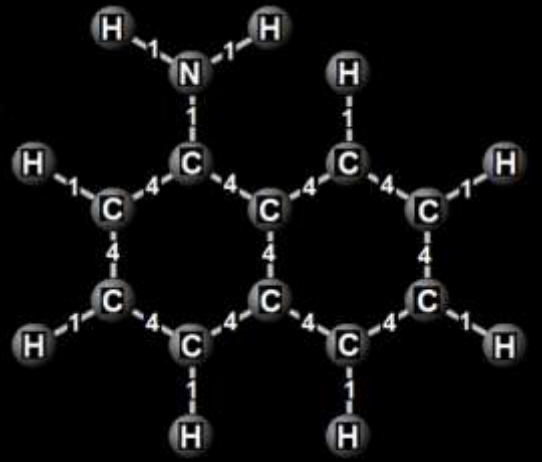


Viewed by another molecule

Viewed by chemists

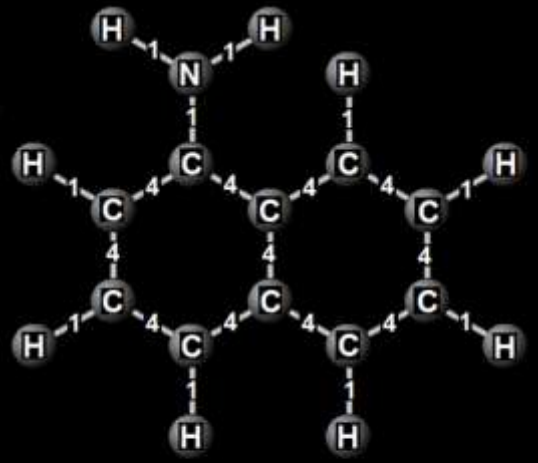


Viewed by computers



Viewed by computers

Structure representation



Viewed by computers

Graphs are widely used to represent and differentiate chemical structures, where atoms are vertices and bonds are expressed as edges connecting these vertices.

MOL File

```
  -ISIS-  11240515102D

 16 15  0  0  0  0  0  0  0  0999 V2000
   -0.1958   -2.9667    0.0000 C   0  0  0
    0.5167   -2.5500    0.0000 C   0  0  0
    0.5125   -1.7250    0.0000 O   0  0  0
    1.2292   -2.9625    0.0000 N   0  0  3
    1.9417   -2.5458    0.0000 C   0  0  0
    2.6542   -2.9583    0.0000 C   0  0  0
    3.3667   -2.5417    0.0000 C   0  0  0
    4.0792   -2.9542    0.0000 C   0  0  0
    4.7917   -2.5375    0.0000 C   0  0  0
    5.5042   -2.9500    0.0000 C   0  0  0
    1.2250   -3.7875    0.0000 C   0  0  0
    0.8083   -4.5000    0.0000 C   0  0  0
    1.3917   -5.0833    0.0000 C   0  0  0
    0.9750   -5.7958    0.0000 C   0  0  0
    1.5583   -6.3792    0.0000 C   0  0  0
    0.9708   -6.9625    0.0000 C   0  0  0
  8  9  1  0  0  0  0
  4  5  1  0  0  0  0
  9 10  1  0  0  0  0
  2  3  2  0  0  0  0
  4 11  1  0  0  0  0
  5  6  1  0  0  0  0
 11 12  1  0  0  0  0
  1  2  1  0  0  0  0
 12 13  1  0  0  0  0
  6  7  1  0  0  0  0
 13 14  1  0  0  0  0
  2  4  1  0  0  0  0
 14 15  1  0  0  0  0
  7  8  1  0  0  0  0
 15 16  1  0  0  0  0
M  END
```

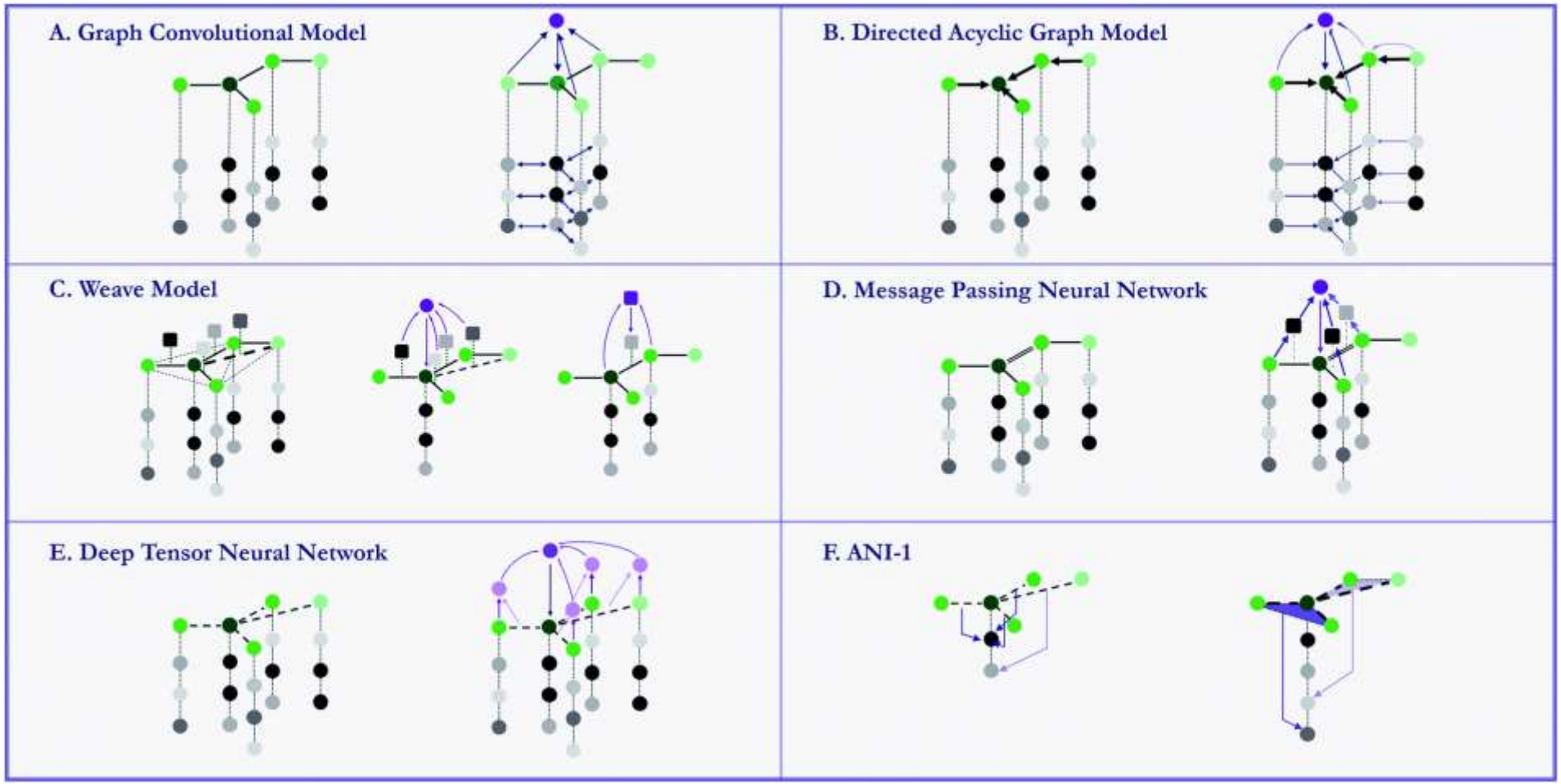
Vertices (atomic type, coordinates etc.)

Edges (connectivity table, label-types of bonds)

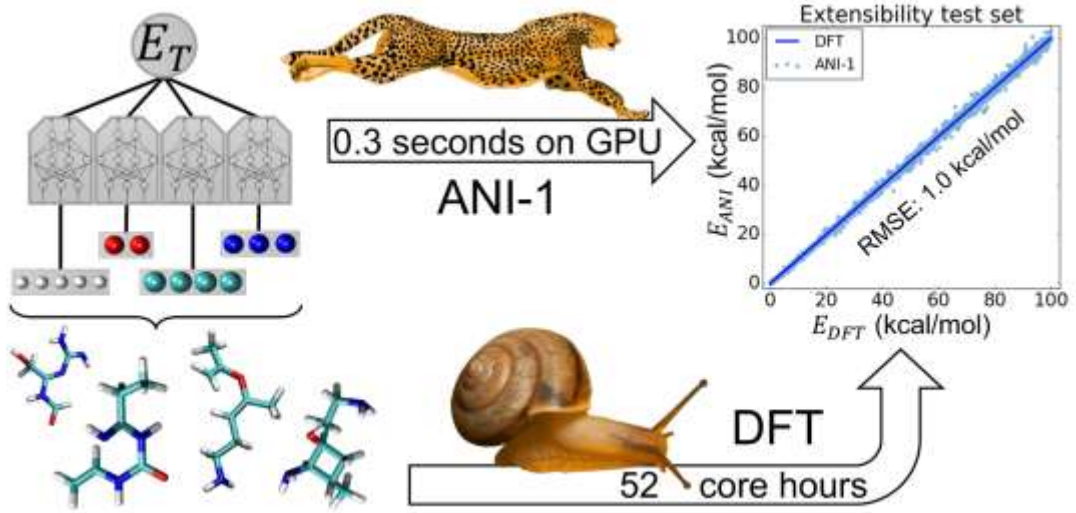
Molecular graphs allow the computation of numerous features to compare them quantitatively.

Downstream ML tasks

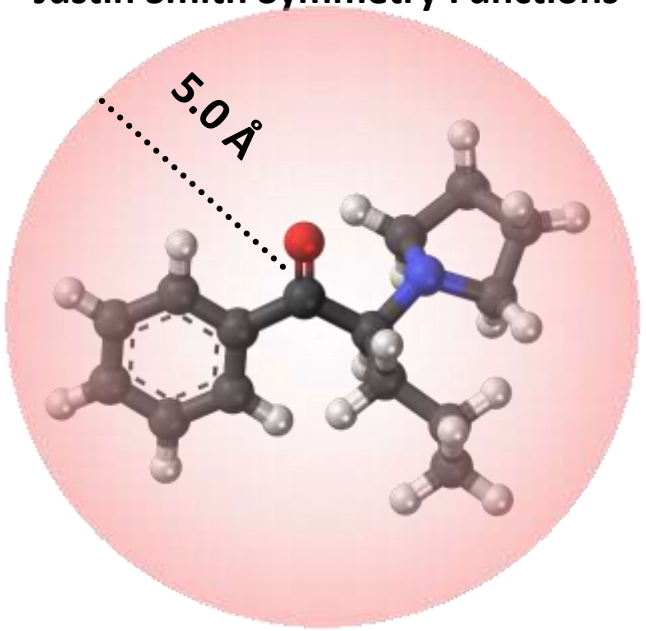
Graph/point cloud-based models for molecules



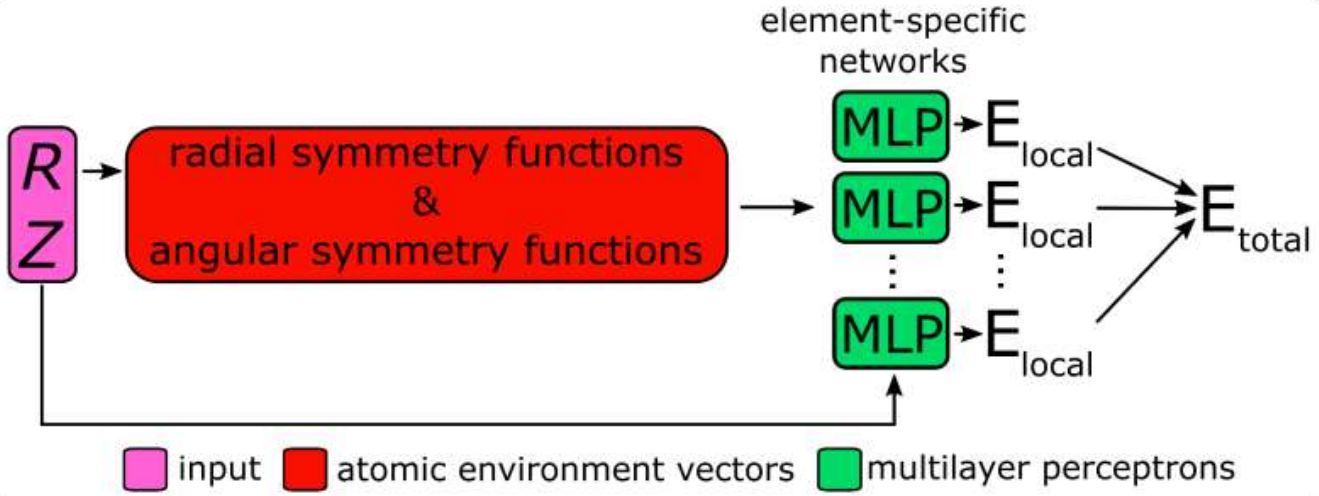
ANI Architecture



Local Atomic Environment:
Justin Smith Symmetry Functions



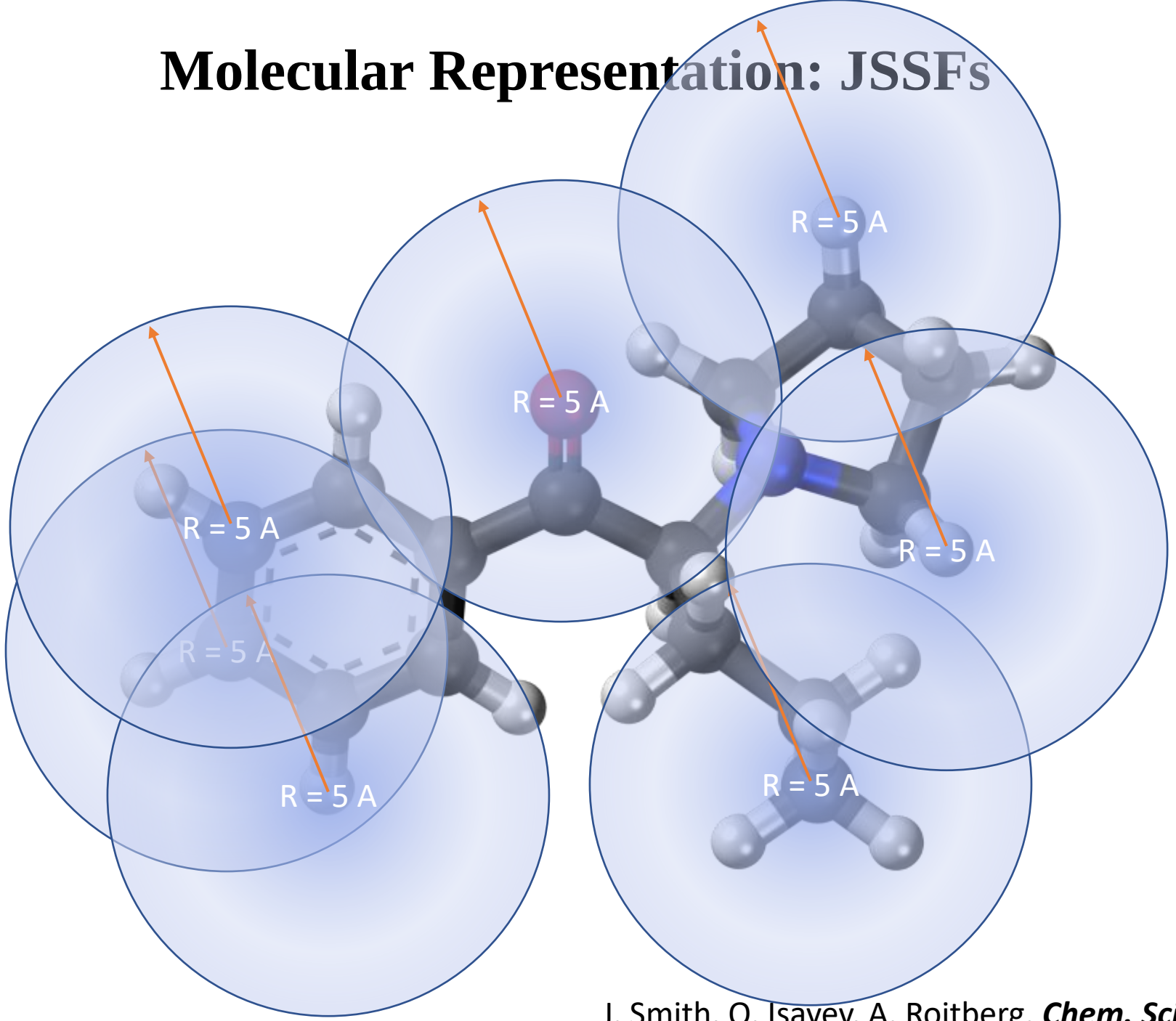
e.g., use element-dependent invariant symmetry functions



The Force is Strong with This One!

Justin Smith, O.I, and Adrian Roitberg. *Chem. Sci.* 8(4), 3192-3203, 2017.

Molecular Representation: JSSFs



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

Current Features of ANI software

https://github.com/aiqm/torchani

Methods

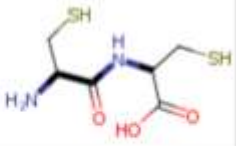
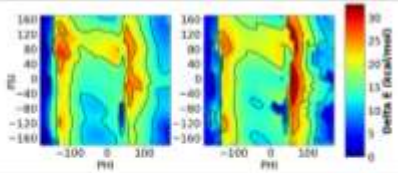
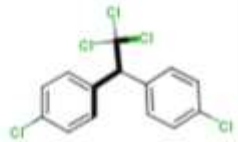
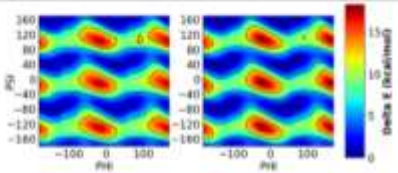
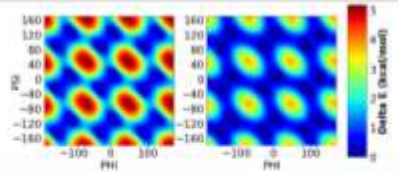
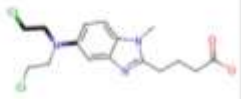
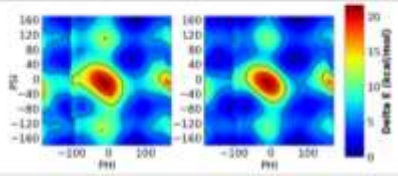
- Currently available elements: CHNOSFCl
  - More advanced model architectures with 14 elements
- Two levels of theory:
  - ωB97M/Def2-TZVPP and
  - CCSD(T)*/CBS

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

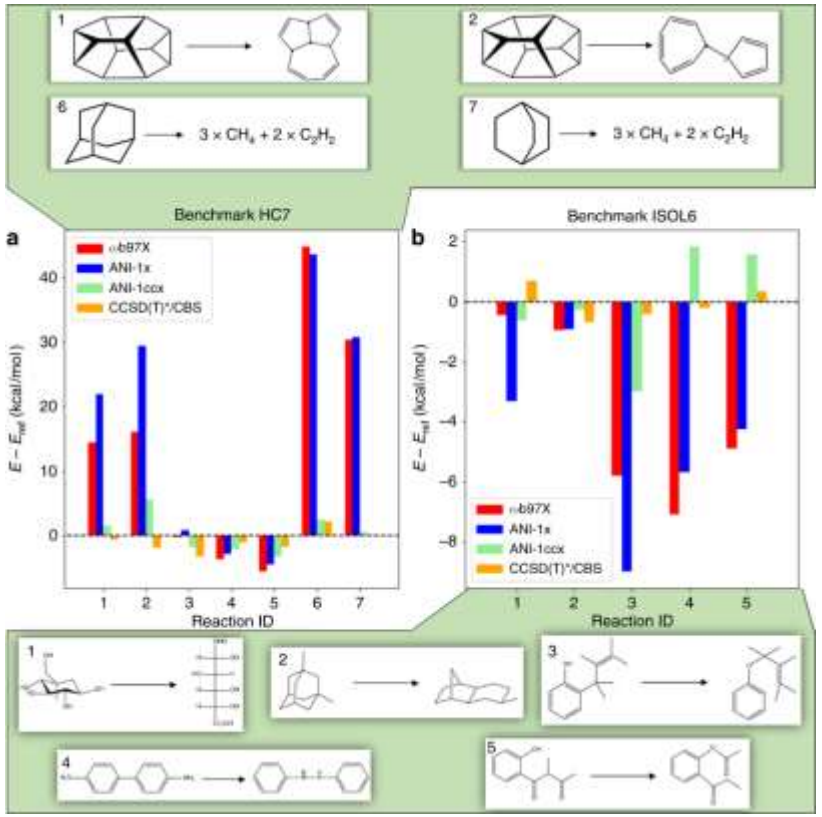
Fast but Accurate Property Predictions with ML

Geometry & Potential Energy Surface

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	

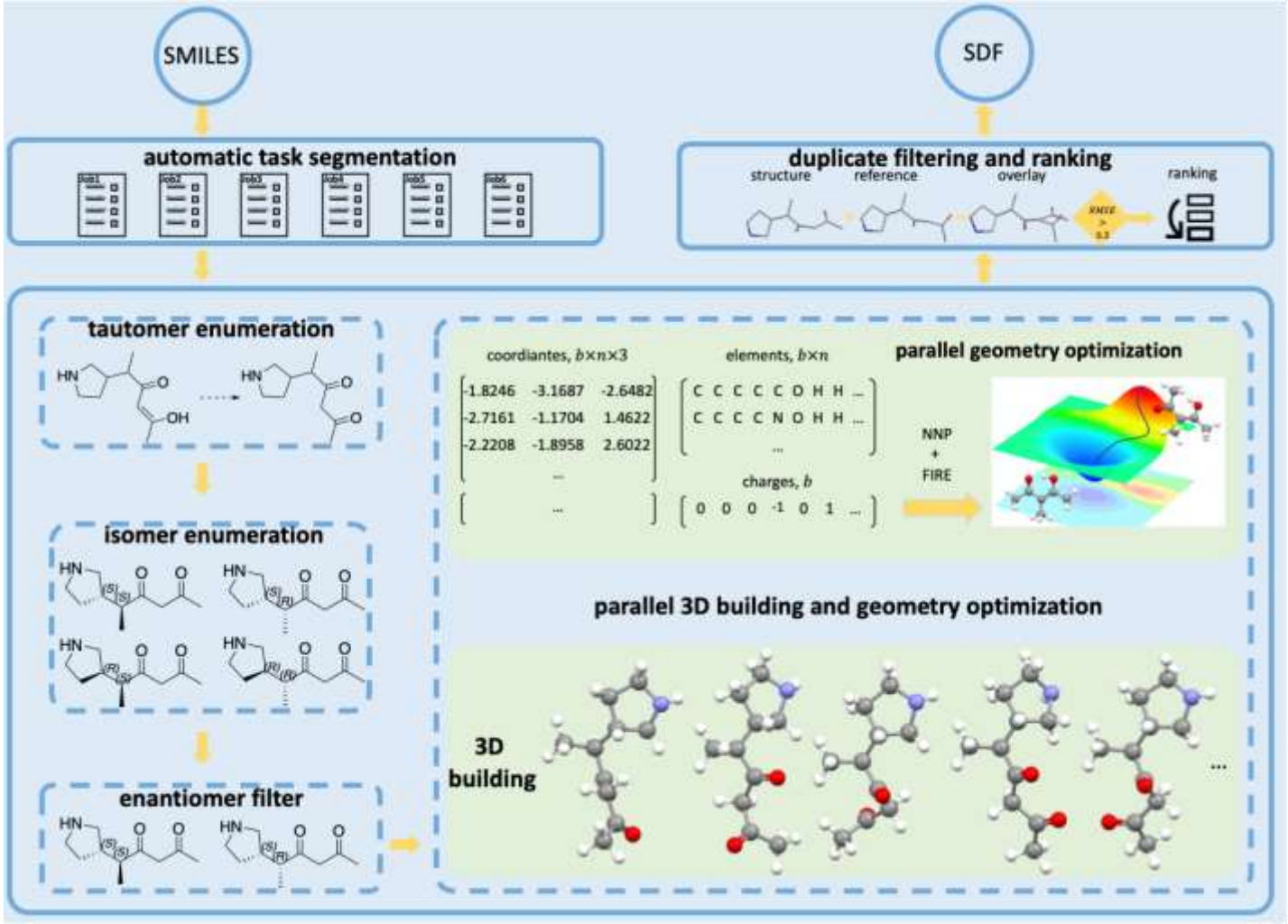
Christian Devereux, Justin S. Smith, Kate K. Davis, Kipton Barros, Roman Zubatyuk, Olexandr Isayev, and Adrian E. Roitberg. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. Journal of Chemical Theory and Computation 2020 16 (7), 4192-4202 DOI: 10.1021/acs.jctc.0c00121

Reaction Thermochemistry



Smith, J.S., Nebgen, B.T., Zubatyuk, R. et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. Nat Commun 10, 2903 (2019). https://doi.org/10.1038/s41467-019-10827-4

Auto3D Package



Input: SMILES

Output: Optimized conformers and energies

Isomer engine: RDKit/OMEGA

Optimizing engine: ANI-2x/AIMNET/ANI-2xt

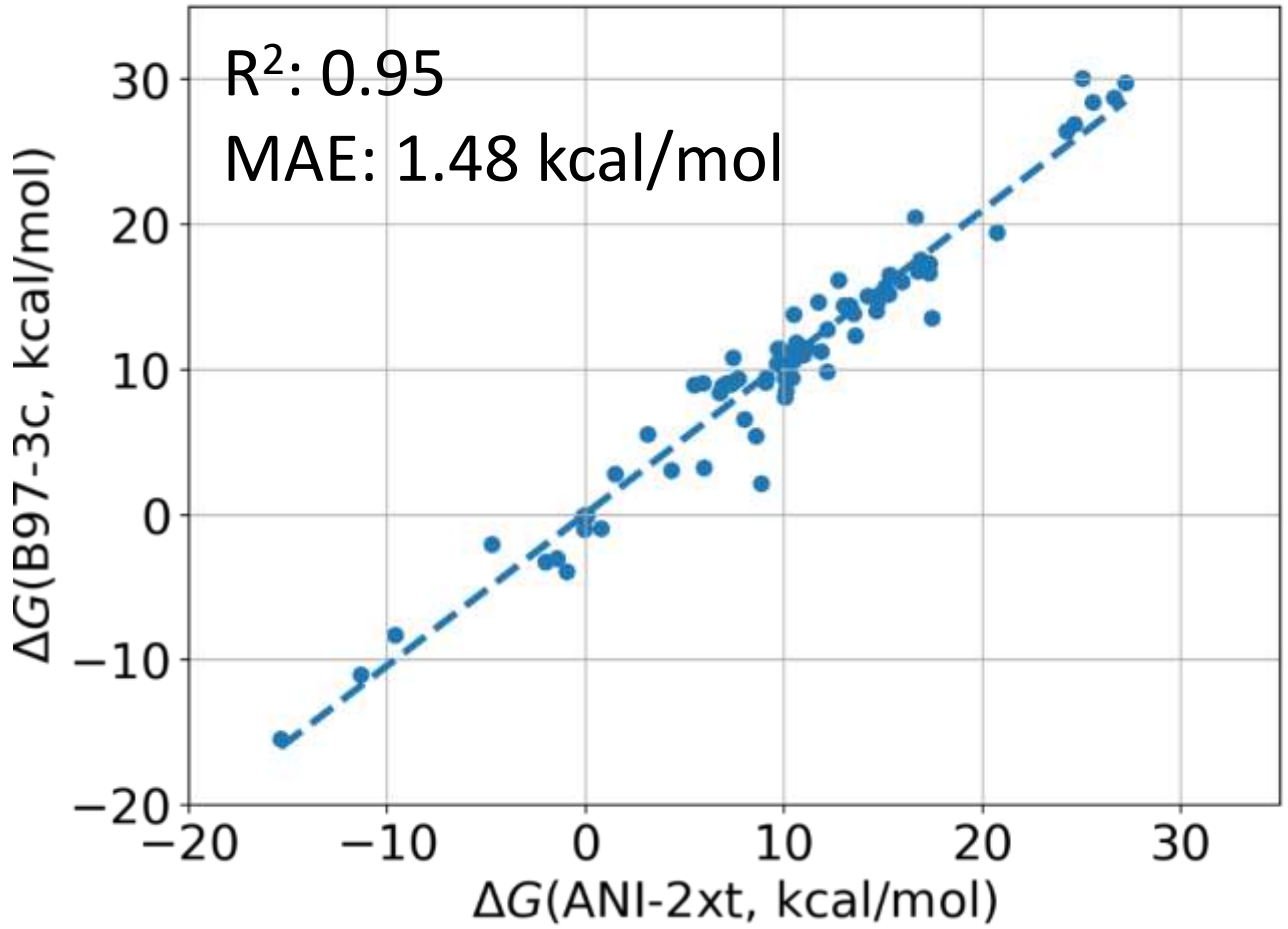
Key features:

- Accurate NNPs
- Highly parallel on GPUs
- Run as a Python library, CLI, yaml

Carnegie Mellon University

Liu, Z.; Zubatiuk, T.; Roitberg, A.; Isayev, O. *J. Chem. Info. Model.* **2022**, *62(22)*, 5373

An example from scratch: Gibbs free energy of tautomerization reactions



- 81 SMILES pairs
- Obtain 3D structure with Auto3D
- Calculate Gibbs free energies with ANI-2xt and B973c, respectively.

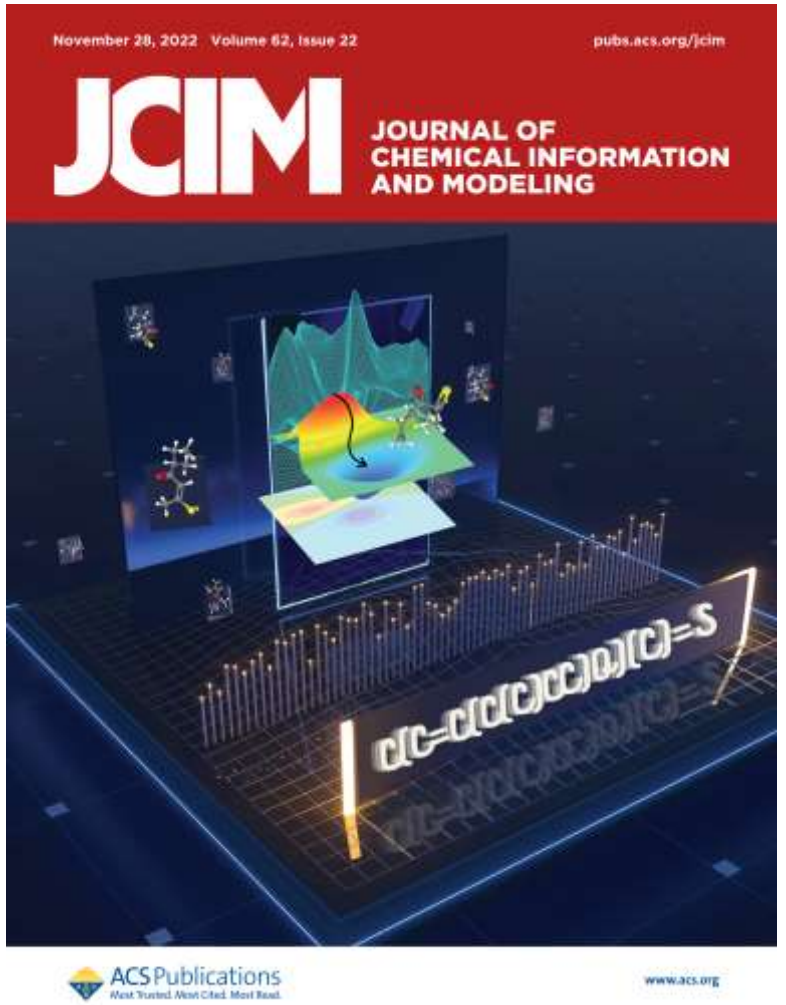
$$\Delta G_{taut}^{gas} = \Delta E_{taut}^{el} + \Delta E(ZPE) + \Delta E^{t} - T\Delta S$$

Carnegie Mellon University

Liu, Z.; Zubatiuk, T.; Roitberg, A.; Isayev, O. *J. Chem. Info. Model.* **2022**, *62(22)*, 5373

Current user base

- Academic labs
- Big pharmaceutical companies
- Chemical Industry
- Drug discovery startups

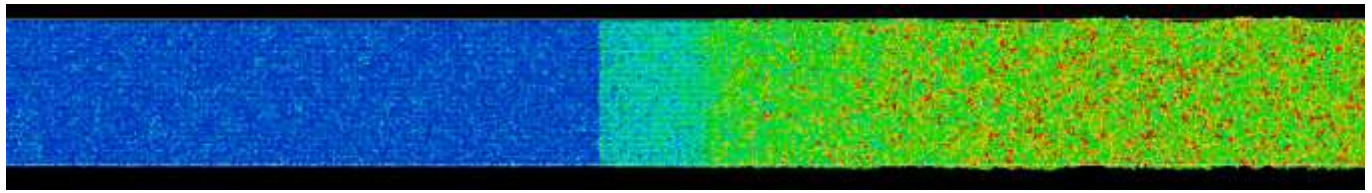


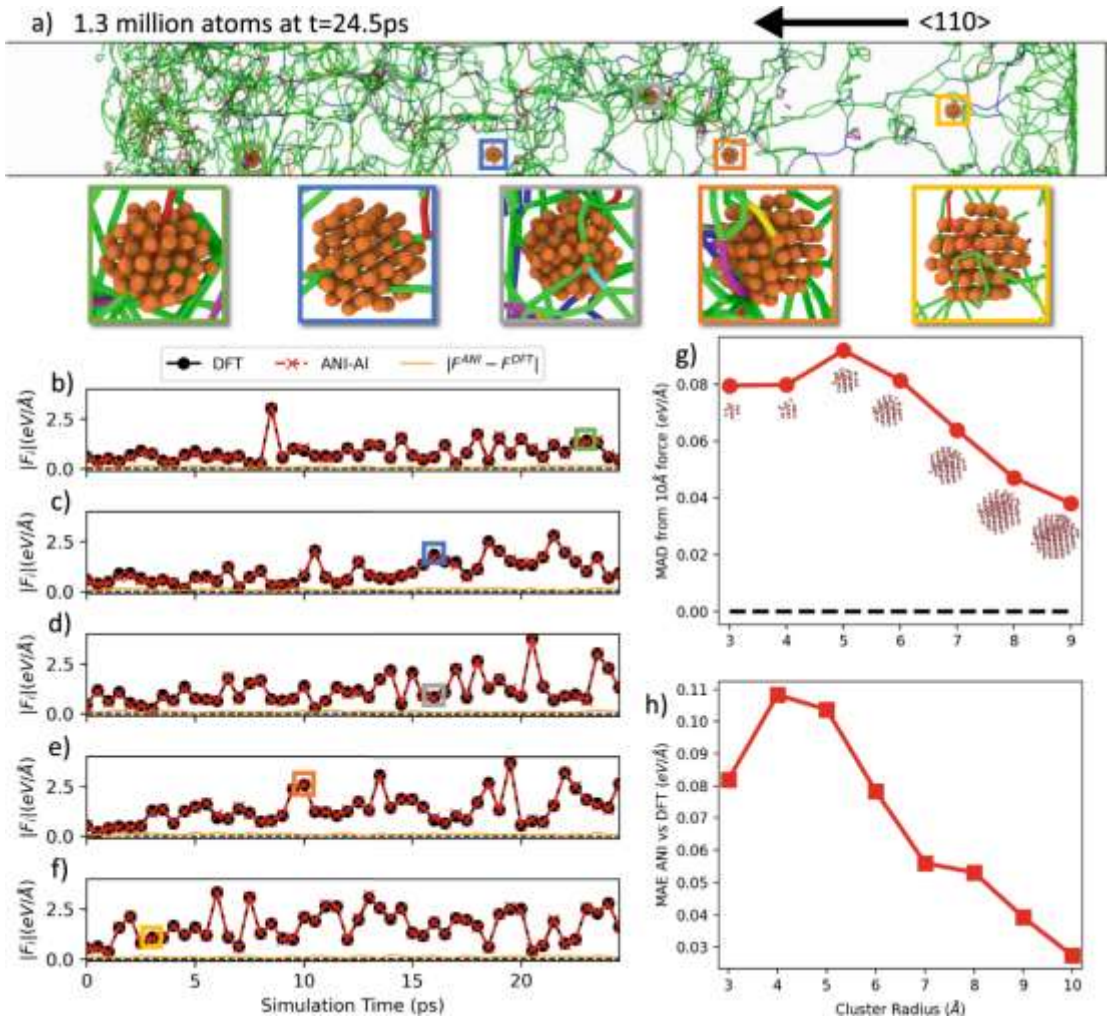
- Installation: pip install Auto3D
- https://github.com/isayevlab/Auto3D_pkg

Carnegie Mellon University

Los Alamos national lab: materials simulations at scale with ANI

- Constructed on the DOE LLNL Sierra supercomputer using V100 GPUs during its open science period
- Active learning automatically samples all phases of interest
- Human knowledge is not required, and may even be inferior to fully automated data selection
- Enabled large scale shock simulations of 1.3 million atoms on 80 NVIDIA GPUs
- Shock simulation can be compared back to EXFAS experimental data for validation





JS Smith, B Nebgen, N Mathew, J Chen, N Lubbers, L Burakovsky, S Tretiak, HA Nam, T Germann, S Fensin, K Barros; *Nature Communications* (12) 1257 (**2021**)

Slide courtesy of Justin Smith

ANI-1xnr: Reactive ML Potential for CHNO Chemistry

nature chemistry

Article

https://doi.org/10.1038/s41557-023-01427-3

Exploring the frontiers of condensed-phase chemistry with a general reactive machine learning potential

Received: 13 April 2023

Accepted: 12 December 2023

Published online: 07 March 2024

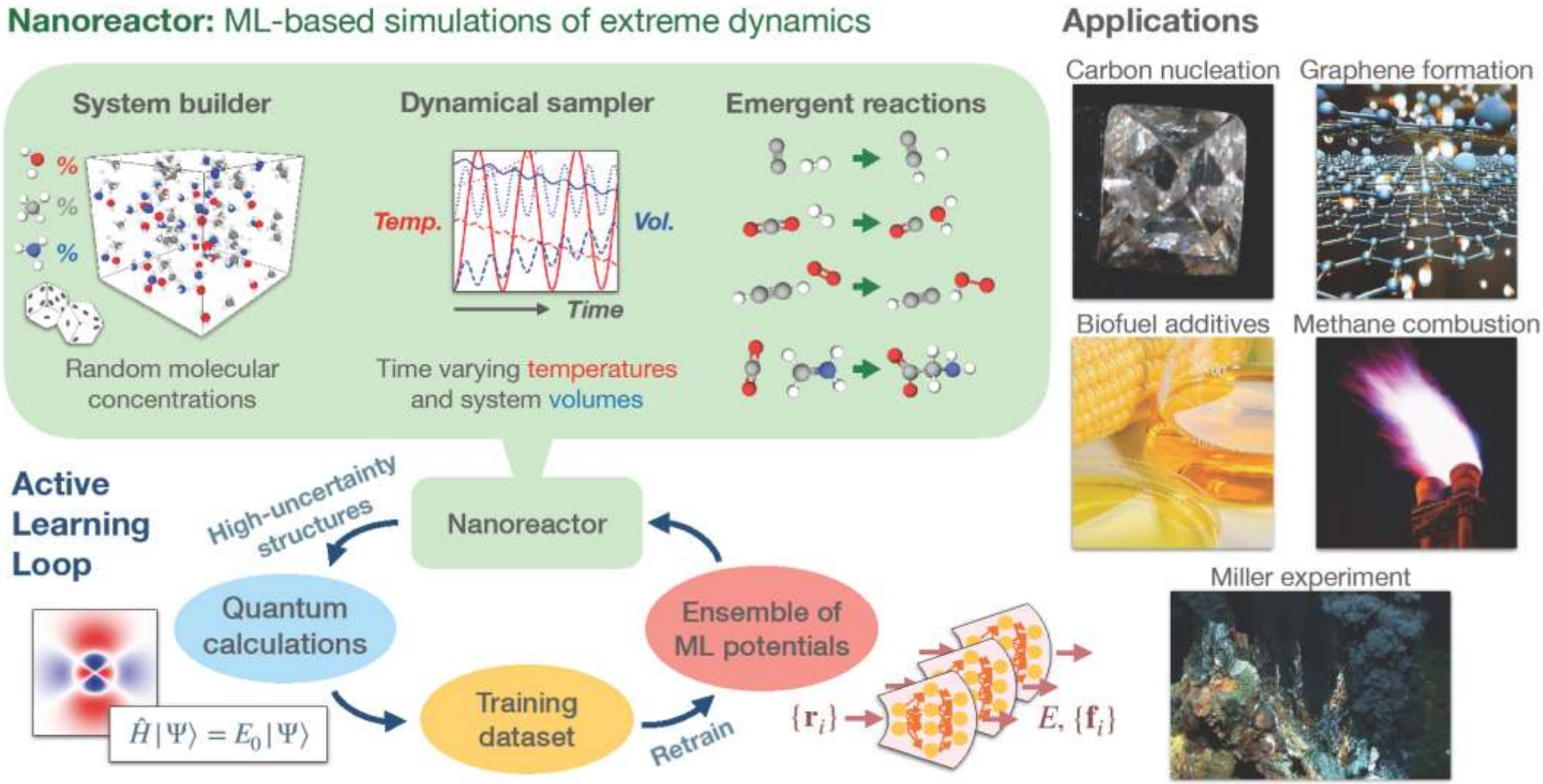
Check for updates

Shuhao Zhang[1,2], Małgorzata Z. Makoś[3,4], Ryan B. Jadrich[2,5], Elfi Kraka[3], Kipton Barros[2,5], Benjamin T. Nebgen[2], Sergei Tretiak[2,5], Olexandr Isayev[1], Nicholas Lubbers[4] ✉, Richard A. Messerly[2] ✉ & Justin S. Smith[2,6] ✉

Atomistic simulation has a broad range of applications from drug design to materials discovery. Machine learning interatomic potentials (MLIPs) have become an efficient alternative to computationally expensive ab initio simulations. For this reason, chemistry and materials science would greatly benefit from a general reactive MLIP, that is, an MLIP that is applicable to a broad range of reactive chemistry without the need for refitting. Here we develop a general reactive MLIP (ANI-1xnr) through automated sampling of condensed-phase reactions. ANI-1xnr is then applied to study five distinct systems: carbon solid-phase nucleation, graphene ring formation from acetylene, biofuel additives, combustion of methane and the spontaneous formation of glycine from early earth

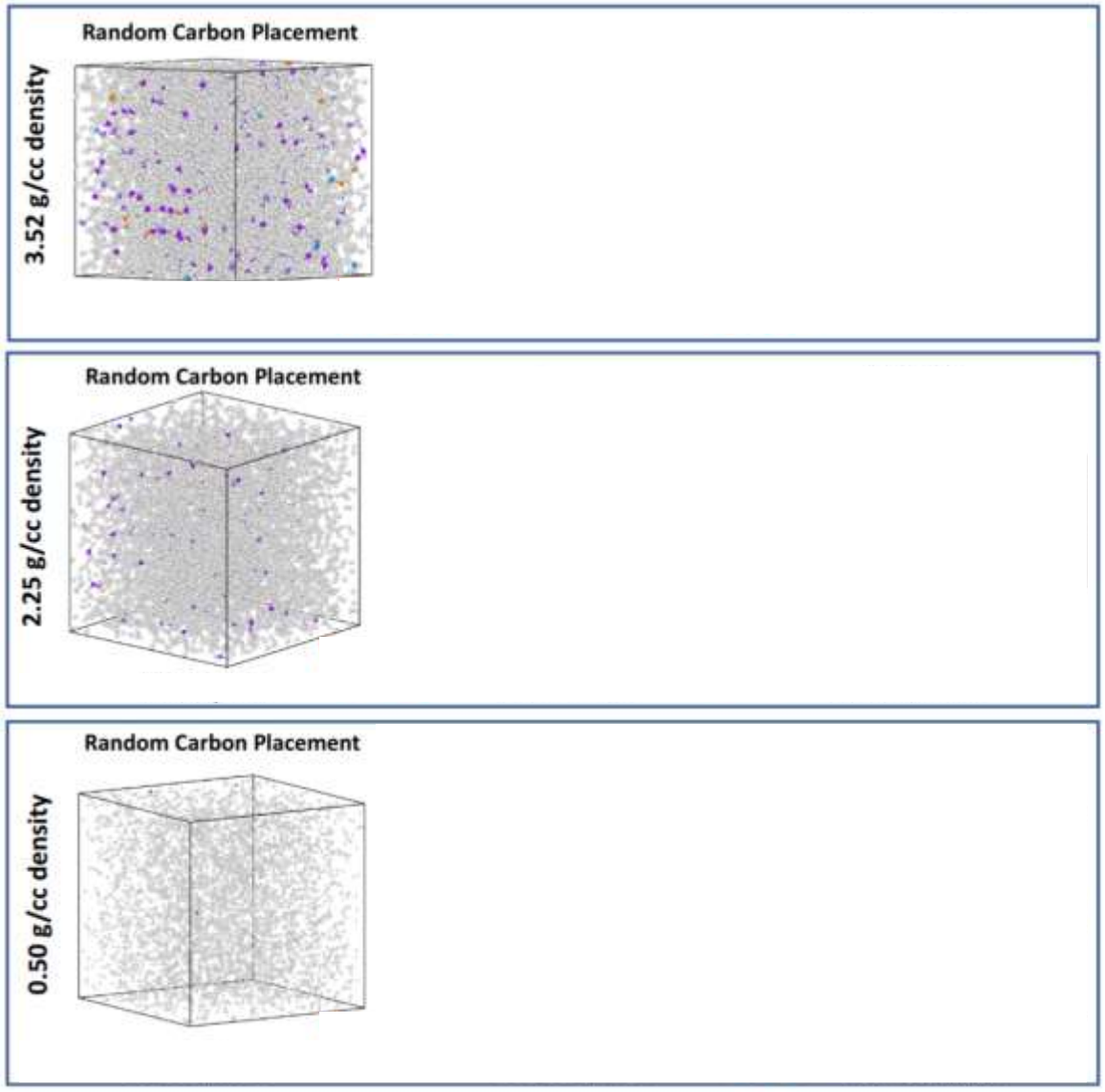
The ANI-1xnr model can be found at:
https://github.com/atomistic-ml/ani-1xnr/

A most recent progress of ANI potential (ANI-1xNR) built by active learning workflow & nanoreactor MD



Zhang, Shuhao, et al. "Exploring the frontiers of chemistry with a general reactive machine learning potential." (2024).

Case Study 1: Carbon solid-phase nucleation



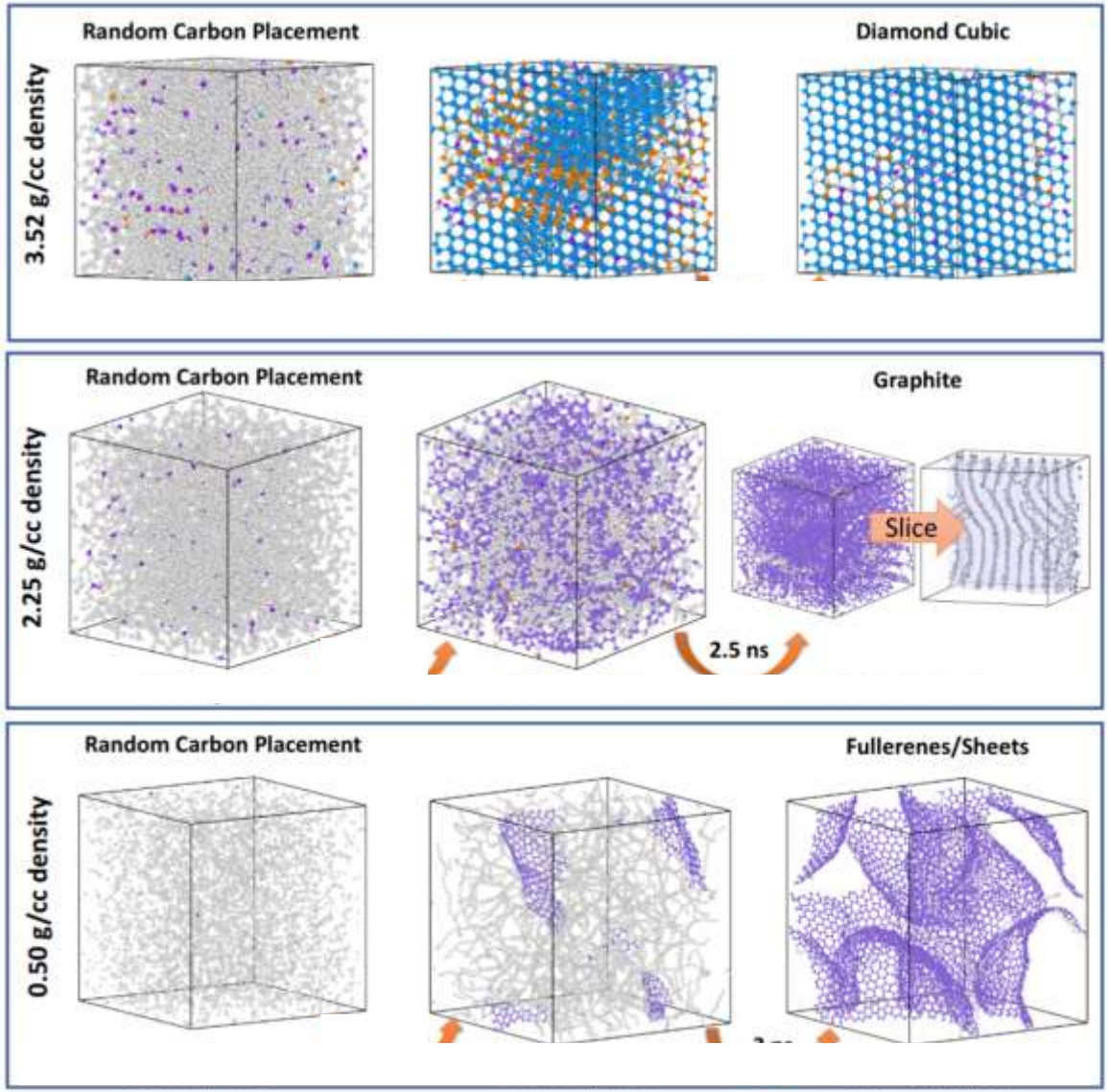
- **Initial simulation set-up:**
  5000 carbon atoms
  0.5-3.5 gm/cm^3
  2500 K

Carbon will form different solid-phase products under different conditions

Test the model's performance on carbon solid-phase nucleation process to see if the physics the model learned can also work on physical process not shown in the training set

Zhang, Shuhao, et al. "Exploring the frontiers of chemistry with a general reactive machine learning potential." (2024).

Case Study 1: Carbon solid-phase nucleation

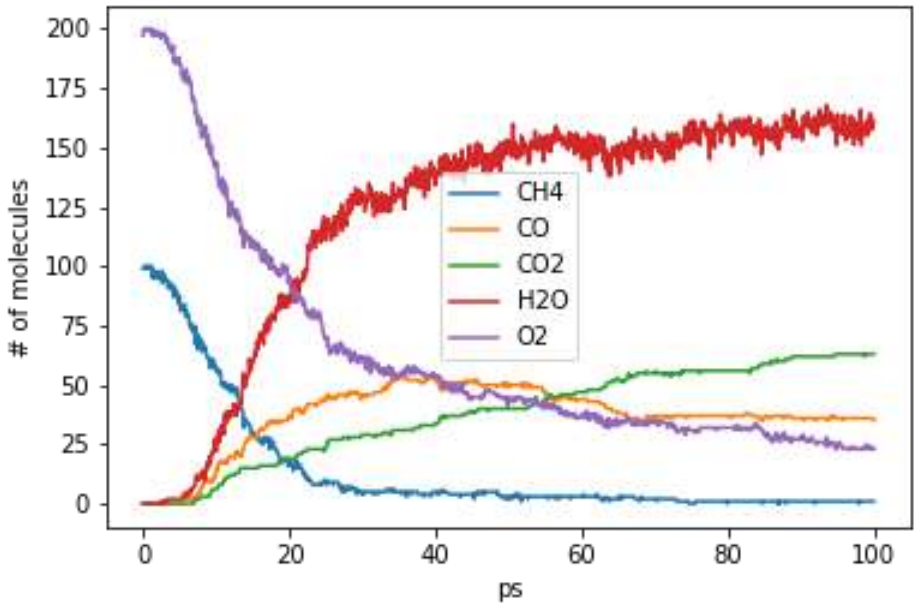


- **ANI-1nr produces the expected bulk structures at each density**
- **ANI-1nr predicts reliable lattice constants for diamond and graphite**

Crystal	Model	*a* (Å)	*b* (Å)	*c* (Å)
Diamond	ANI-1xnr	3.58	3.58	3.58
	ANI-2x	3.75	3.75	3.64
	Exp.	3.57	3.57	3.57
Graphite	ANI-1xnr	2.47	2.47	6.24
	ANI-2x	2.44	2.44	10.0
	Exp.	2.46	2.46	6.71

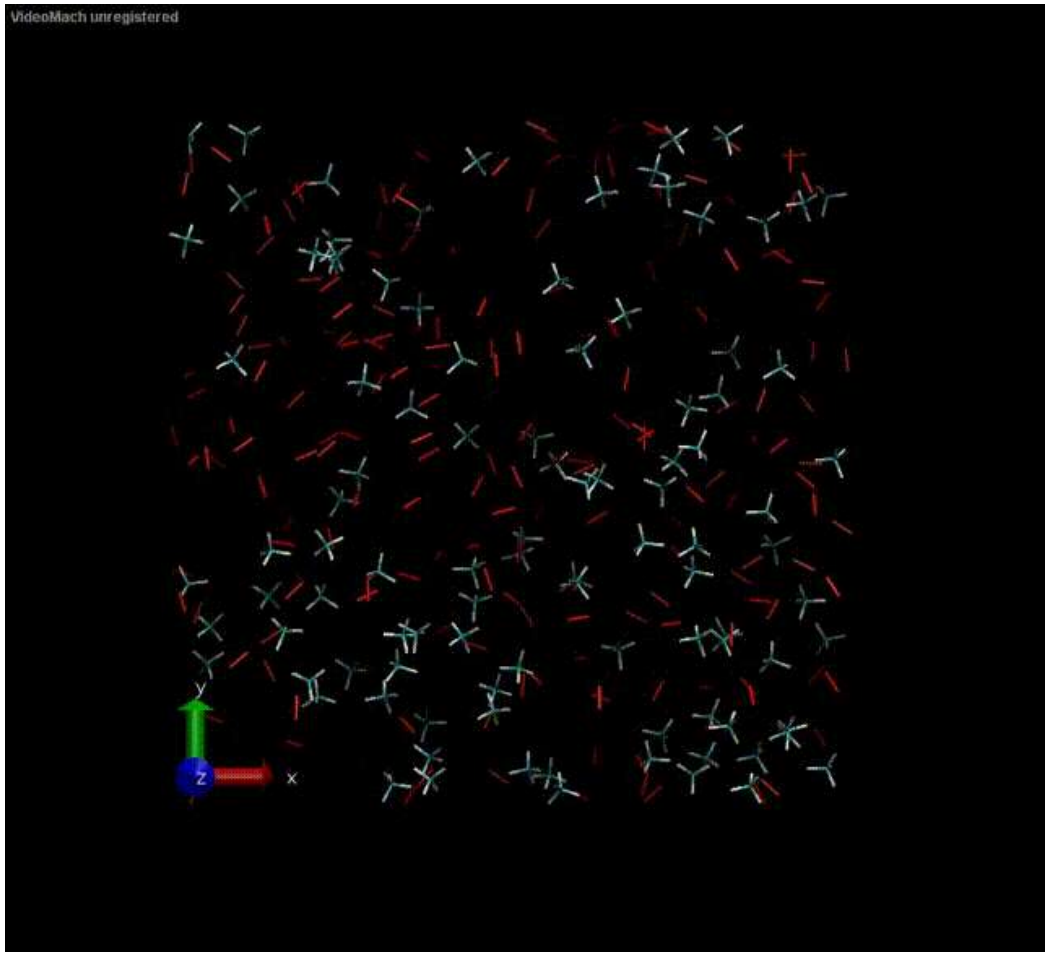
Zhang, Shuhao, et al. "Exploring the frontiers of chemistry with a general reactive machine learning potential." (2024).

Case Study 3: Methane Combustion



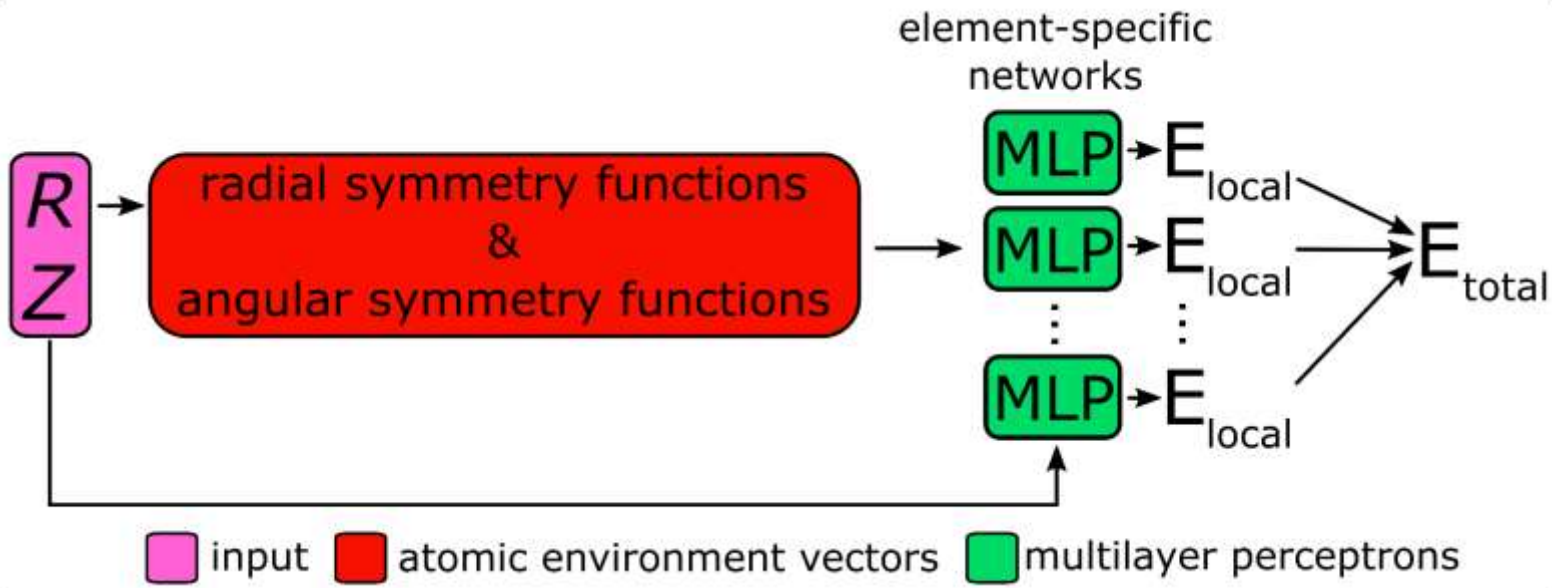
Fragments track during the simulation

Species profiles consistent with literature bespoke ML potential

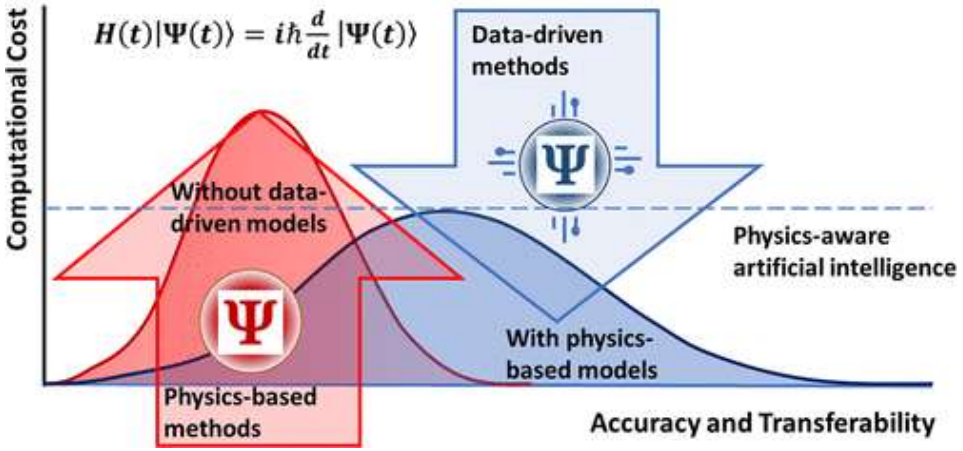


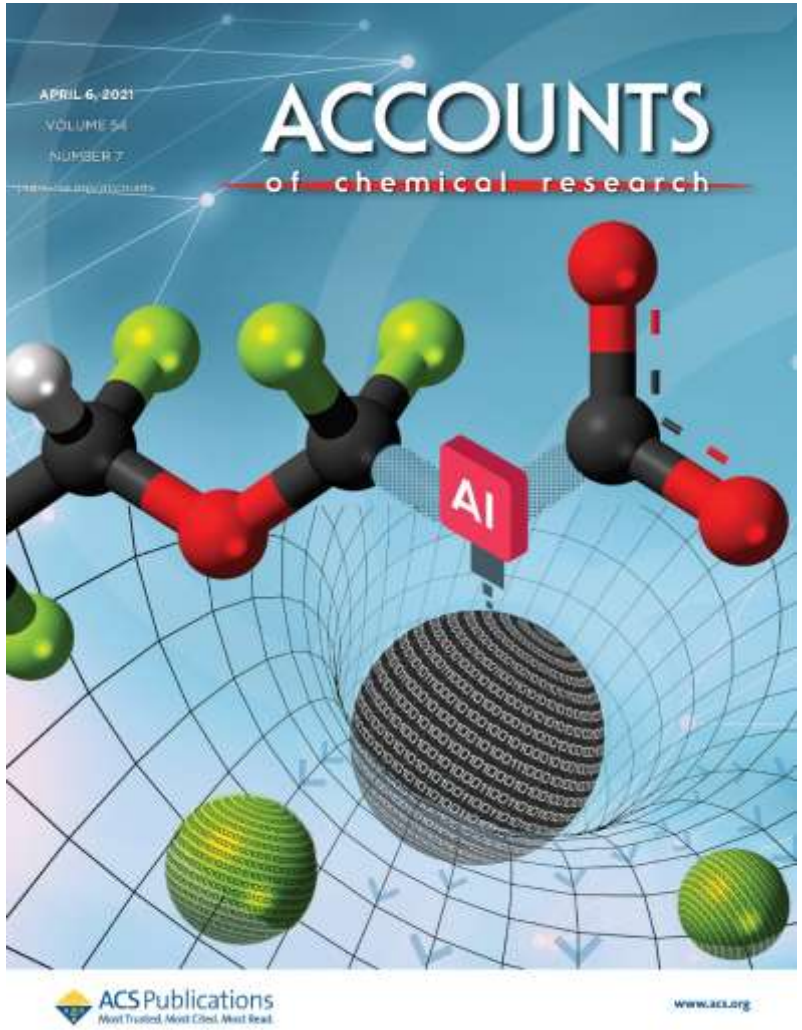
Methane combustion (100 methane + 200 O2) with ANI-RXN reactive potential

2017: ANI


element-specific
networks
R
Z
radial symmetry functions
&
angular symmetry functions
MLP
MLP
MLP
E local
E local
E local
E total
input
atomic environment vectors
multilayer perceptrons

Different NN Architectures





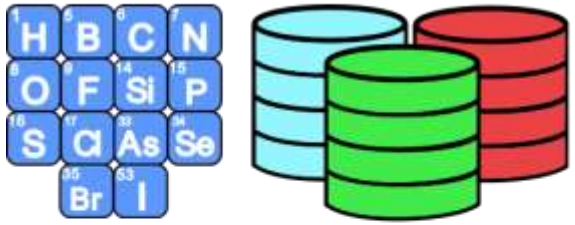
Tetiana Zubatiuk and Olexandr Isayev. Accounts of Chemical Research **2021** 54 (7), 1575-1585 DOI: 10.1021/acs.accounts.0c00868

The 2nd Generation Atoms-in-Molecules ML Potential

AIMNet2 is generalized potential with chemical space coverage of

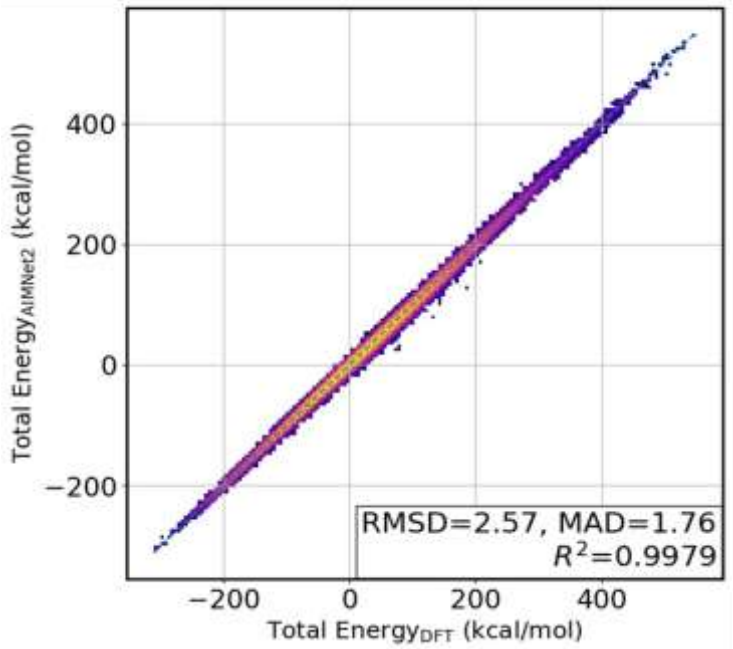
<u>common non-metal and halogen elements with ~hybrid DFT accuracy</u>

20M ωB97m/def2-TZVPP Calculations

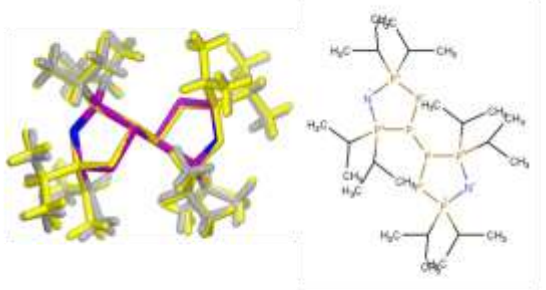


- Existing datasets
- Extraction from CSD
- Constrained Molecular Dyn.
- Normal Mode Sampling
- Active Learning
- Data Distillation
- ...

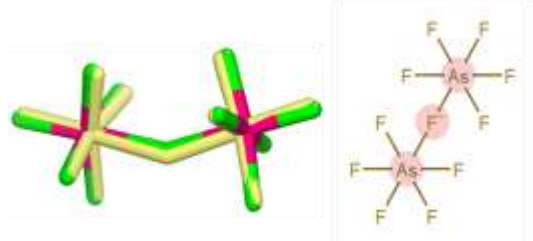
Neutral and Charged Species



zwitterionic bridged 5-membered phosphorous-nitrogen rings



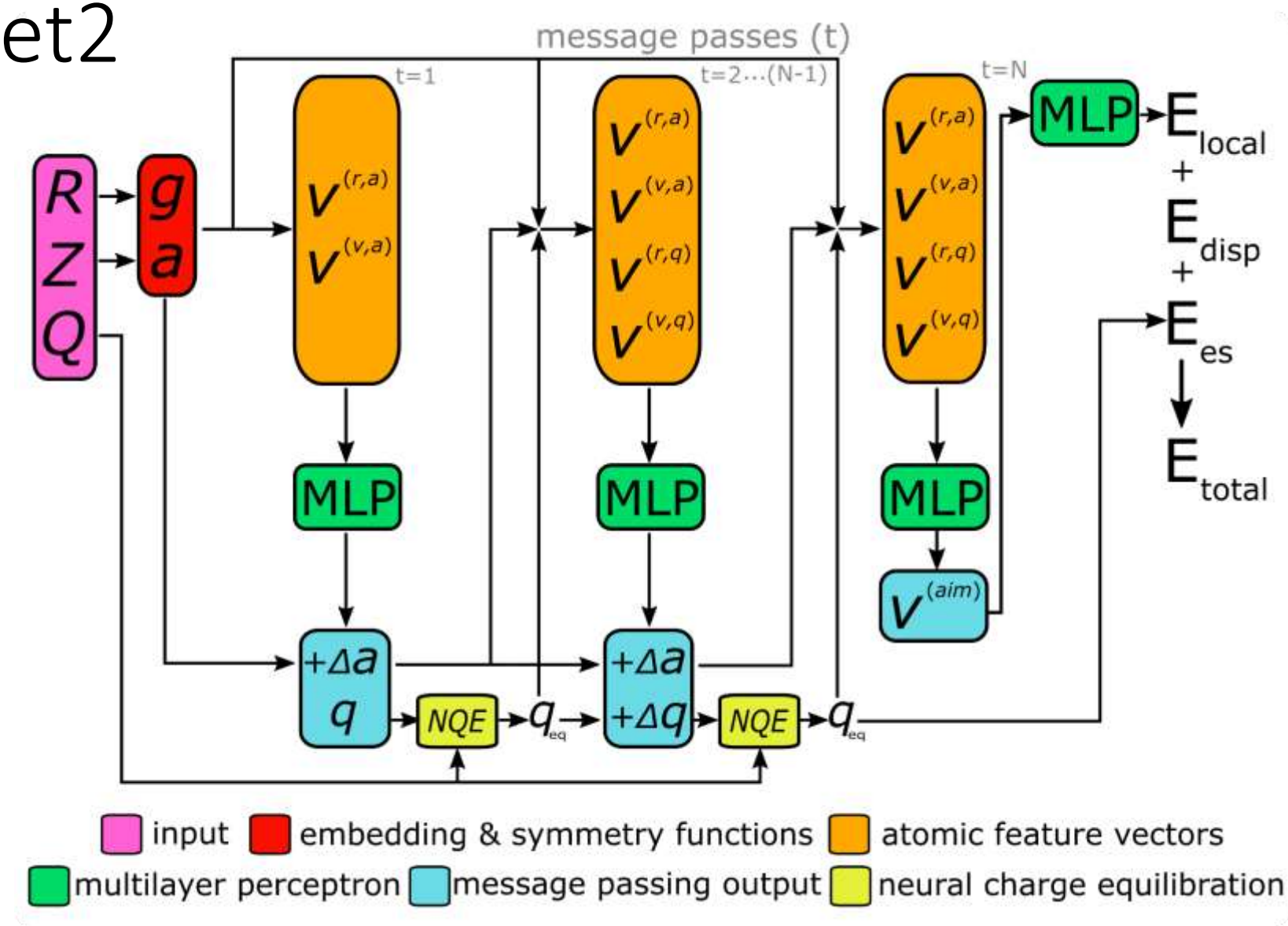
Arsenic pentafluorides bridged by a negative fluorine



<u>Pre-trained Models and Calculators Available</u>: https://github.com/isayevlab/aimnet2

Anstine D, Zubatyuk R, Isayev O. AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs. *ChemRxiv*.; **2023**; DOI: 10.26434/chemrxiv-2023-296ch

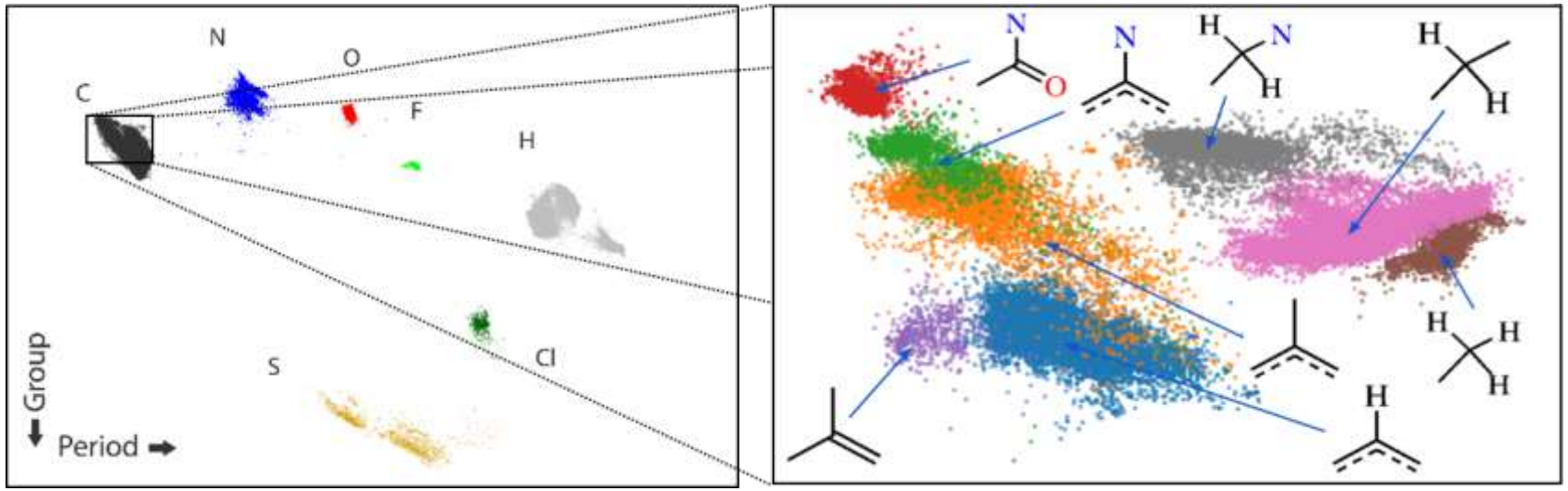
2024: AIMNet2



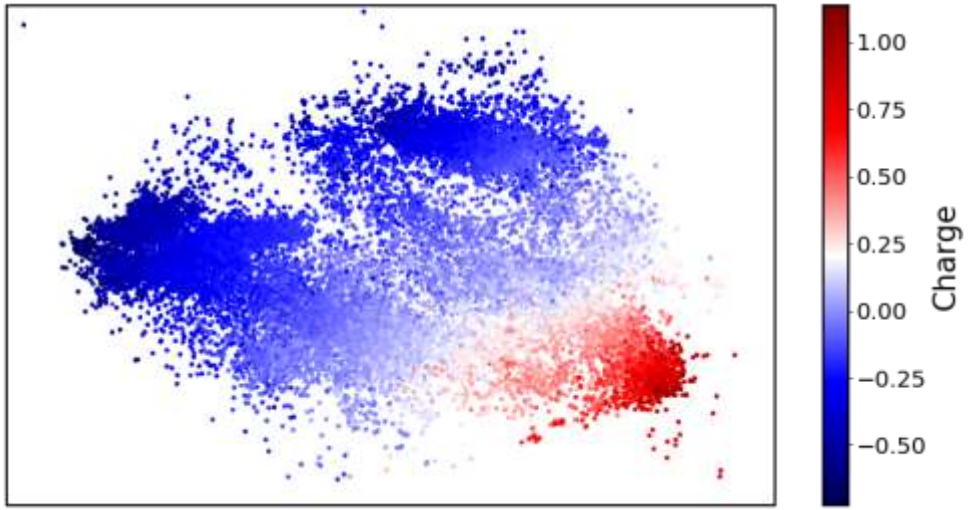
Anstine D, Zubatyuk R, Isayev O. AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs. *ChemRxiv*.; **2023**; DOI: 10.26434/chemrxiv-2023-296ch

Parametric t-SNE projection of Atomic Feature Vectors

Differentiation by chemical environment



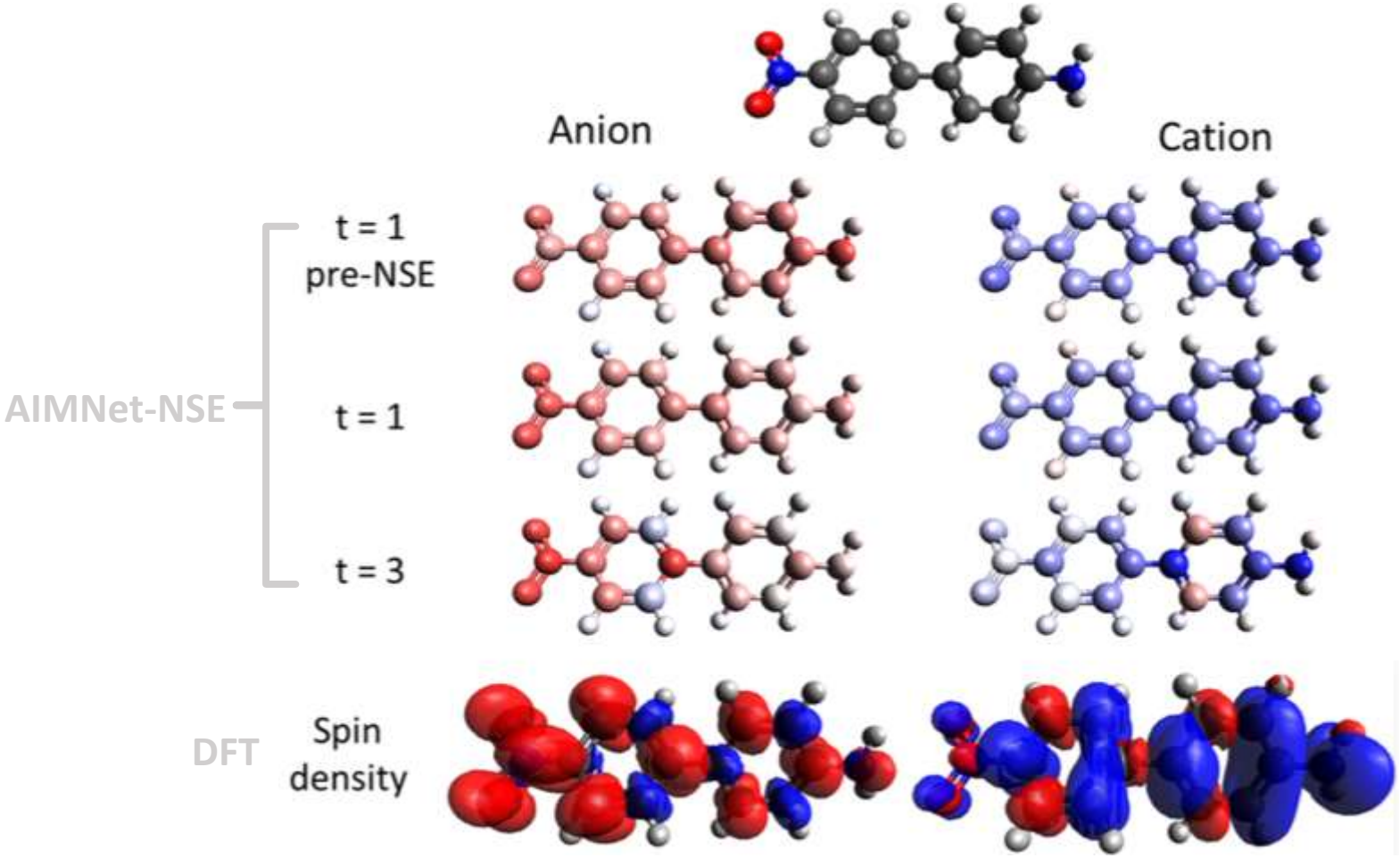
Differentiation by atomic charge



NSE: Neutral Spin and Charge Equilibration

The development of **neutral spin equilibration** improves AIMNet to **cover neutral and charged molecules**.

Zubatyuk et al. Nat. Comm. 12, 4870, 2021.



During message passing, charge normalization factors are predicted to redistribute charge across the system.

$$q_i^s = \tilde{q}_i^s + \frac{f_i^s}{\sum_{j=1}^{N} f_j^s}\left(Q^s - \sum_{j=1}^{N} \tilde{q}_j^s\right)$$

AIMNet-NSE achieves energy accuracy of ~2 kcal mol^{-1} for neutral, anionic, and cationic molecules.

Conceptual DFT: reactivity indices

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

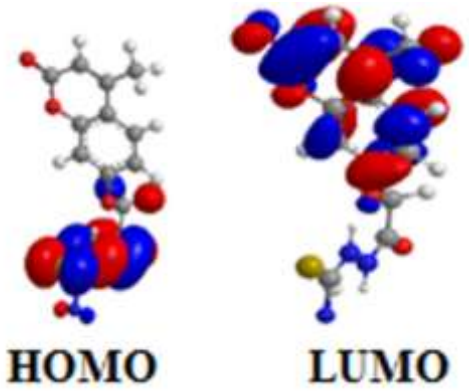
Molecular hardness $\eta = \frac{1}{2}\left(\frac{\partial^2 E}{\partial N^2}\right)_V \approx \frac{1}{2}(IP - EA)$

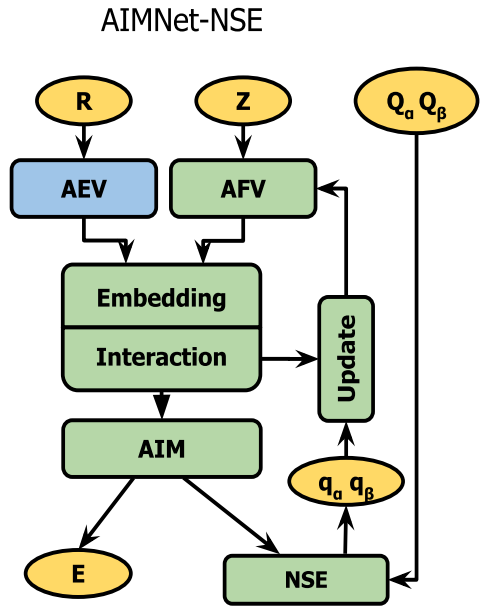
Electrophilicity index $\omega = {}^{\mu^2}/_{2\eta}$

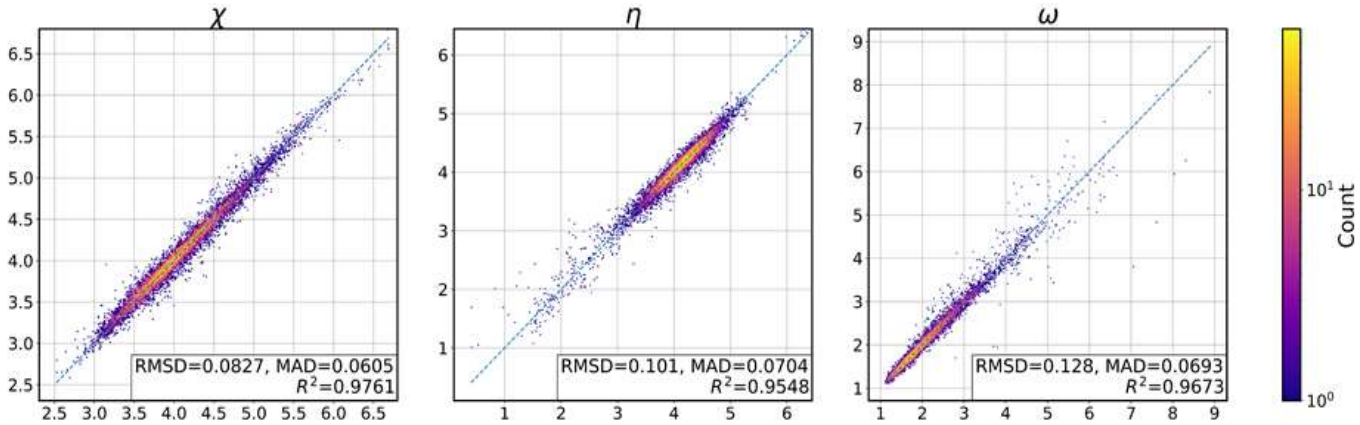
Fukui index $f(r) = \left(\frac{\partial \rho(r)}{\partial N}\right)_V$

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

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





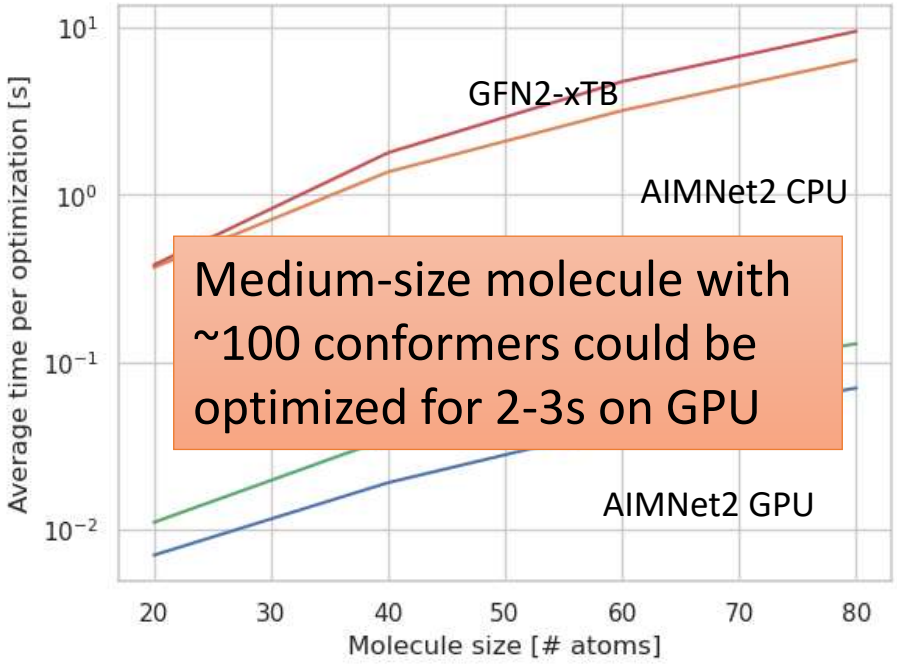


DFT and AIMNet-NSE predictions for electronegativity (χ), chemical hardness (η) and electrophilicity index (ω)

Zubatyuk, Roman; Smith, Justin; Nebgen, Benjamin T.; Tretiak, Sergei; Isayev, Olexandr. Teaching a Neural Network to Attach and Detach Electrons from Molecules. *Nature Communications*, **2021**, 12, 4870. Preprint: https://doi.org/10.26434/chemrxiv.12725276.v2

AIMNet2 computational performance

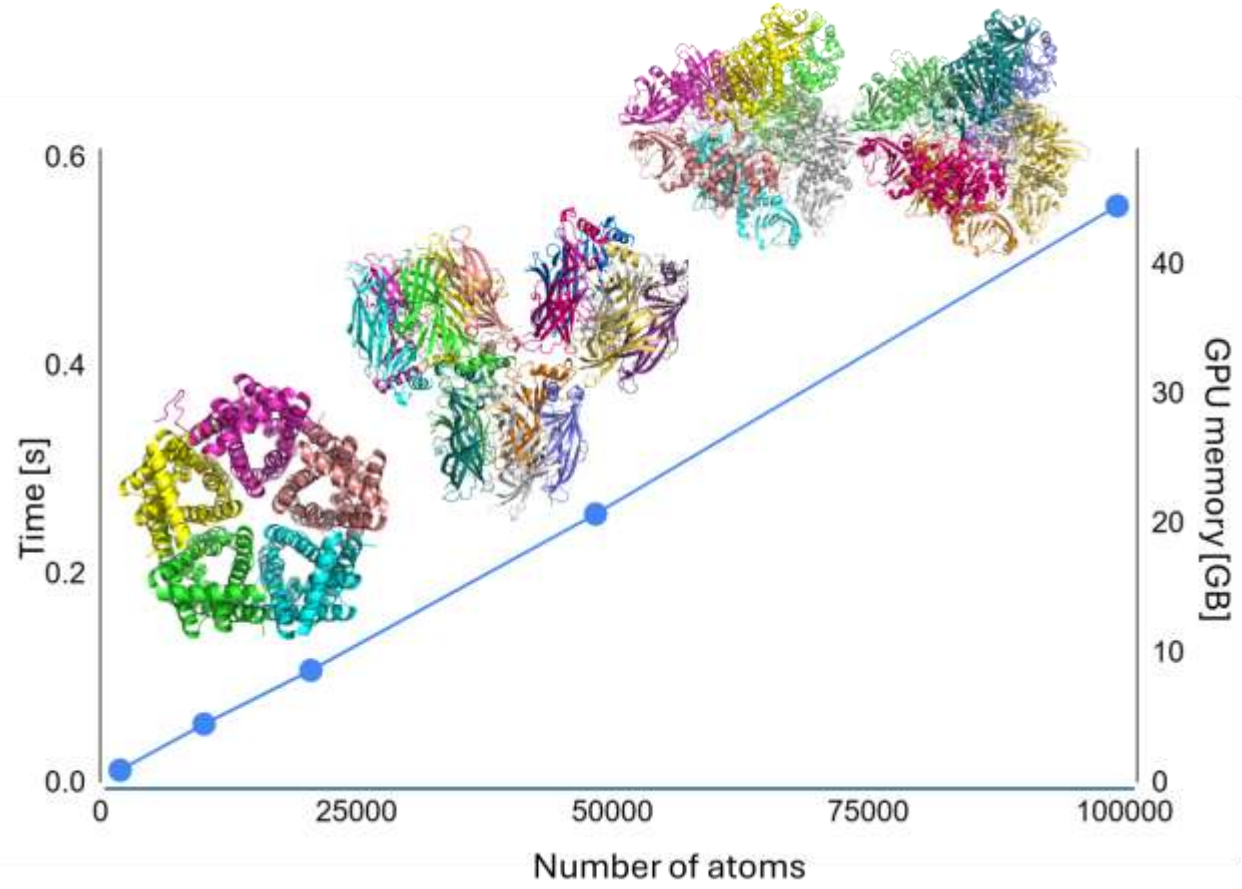
Batch optimization of conformer ensembles



Benchmarks performed on a single core of i7-9700K CPU and Nvidia Titan V GPU.
Reported **total** optimization time per molecule.

AIMNet2 model was used with PyTorch based implementation of FIRE optimizer, which takes 1.5-2x more steps to converge compared to ANC optimizer in XTB.

Protein energy & gradient evaluation

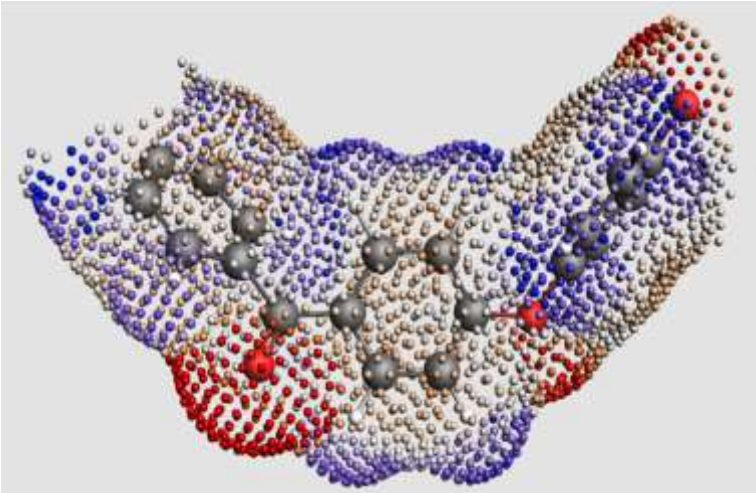


Benchmarks performed on a Nvidia H100 (80GB VRAM)
LBFGS optimizer in ASE has overhead of 4-5% beyond forces evaluation with AIMNet2.

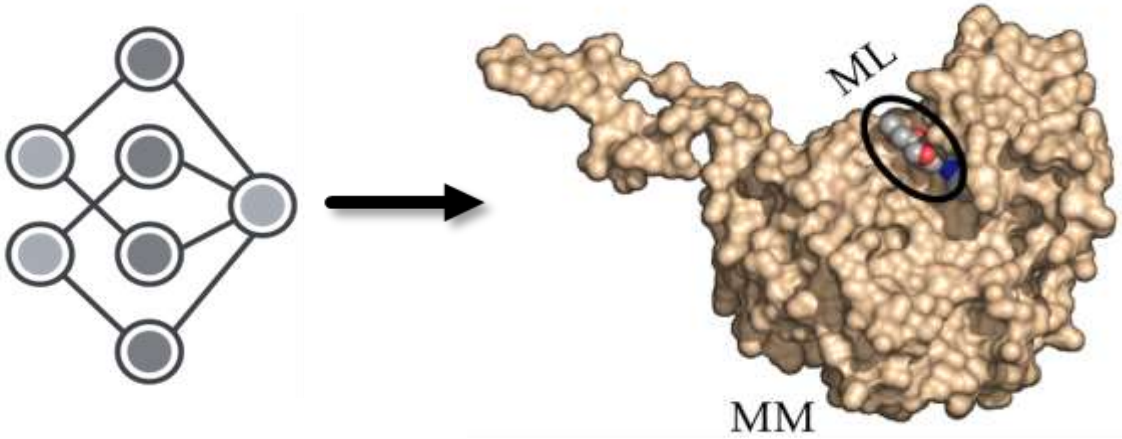
Anstine D, Zubatyuk R, Isayev O. AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs. *ChemRxiv*.; **2023**; DOI: 10.26434/chemrxiv-2023-296ch

AIMNet2 Foundation Model: Enabling Diverse Application

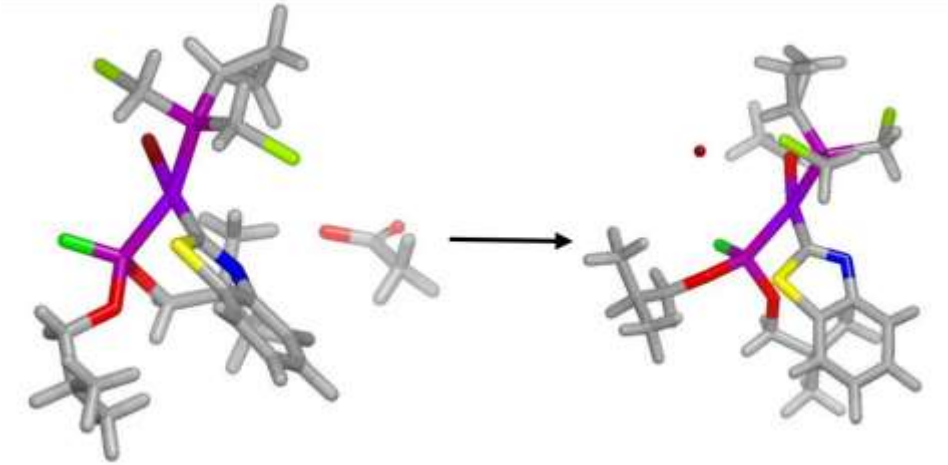
AIMNet2-COSMO-SAC/ RS



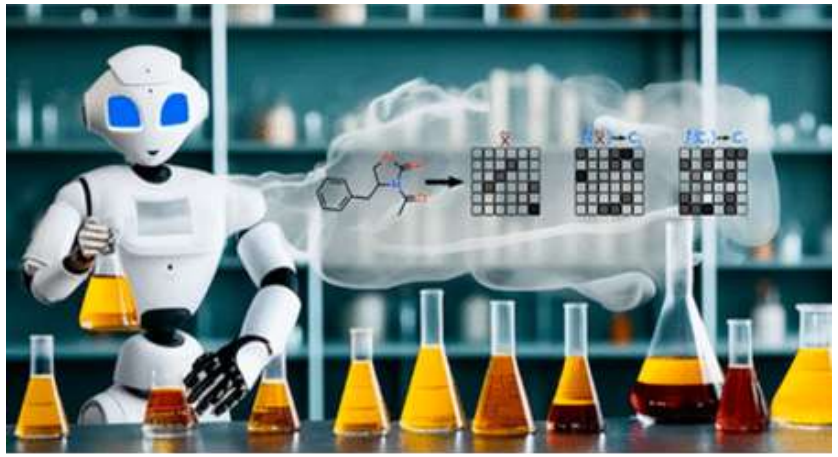
Machine Learning / Molecular Mechanics Simulations



Pd Catalyzed Reactions



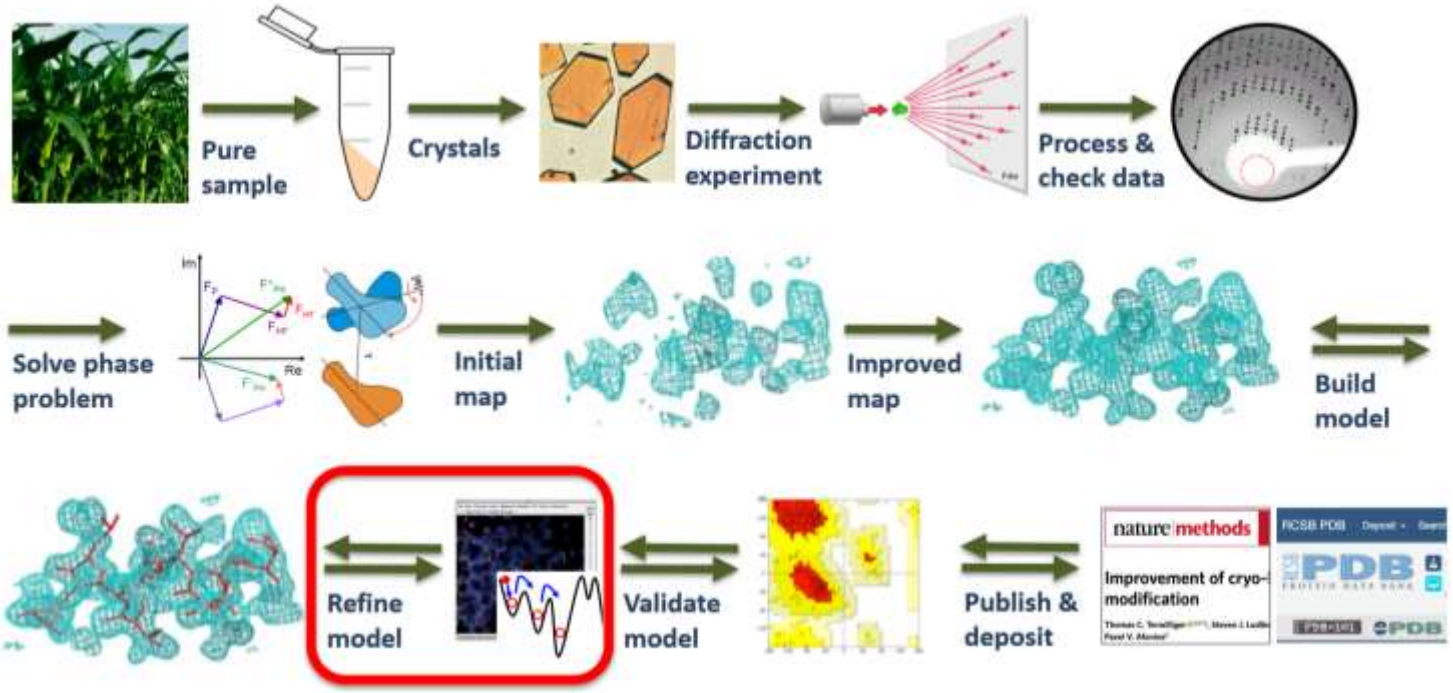
Generative Modeling



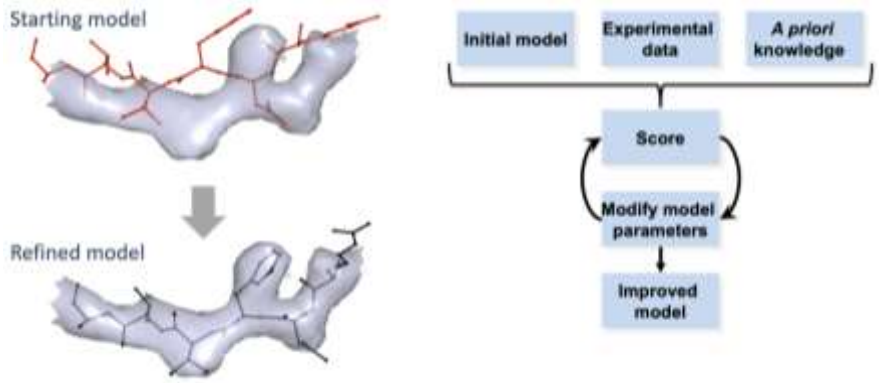
Anstine & Isayev, JACS, 2023

Refinement of protein crystal structure with ML

Optimization process of fitting structural parameters to experimental data



***Priori* knowledge**



Fit atomic model to experimental data as good as possible while making sure the model makes physical and chemical sense

QM computations

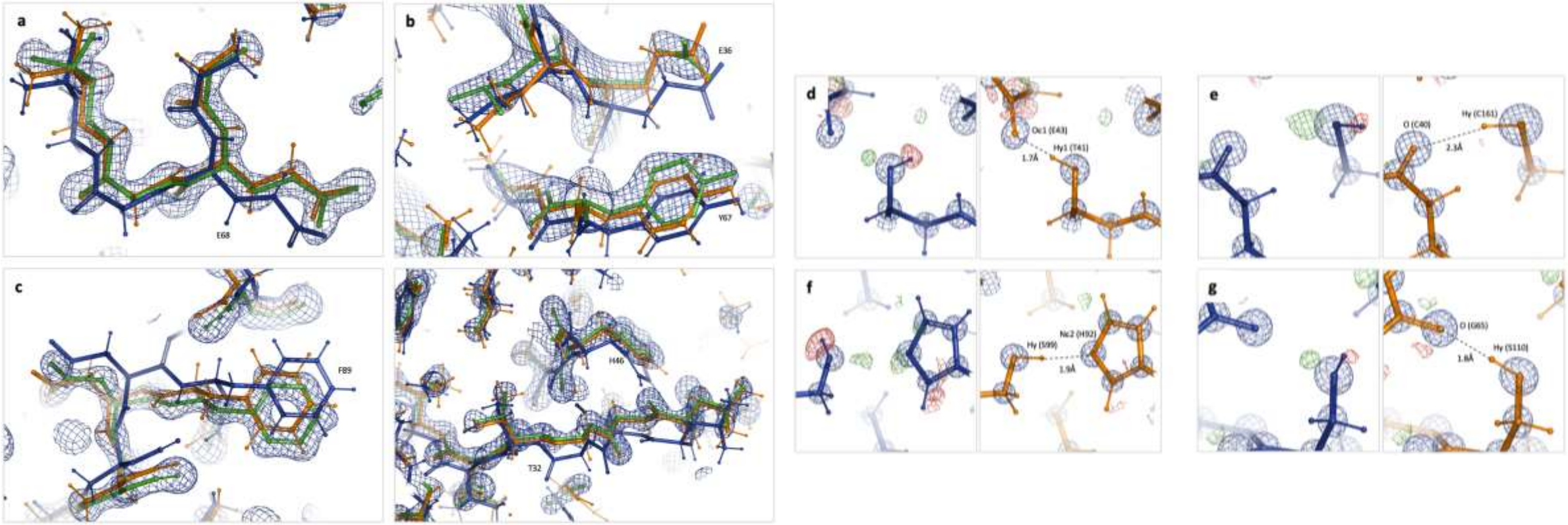
https://github.com/qrefine/

Q|R

Phenix

R. Zubatyuk, et al. AQuaRef: Machine learning accelerated quantum refinement of protein structures. BioRxiv 2024.07.21.604493; DOI: https://doi.org/10.1101/2024.07.21.604493

Default Phenix Refinement vs AQuaRef



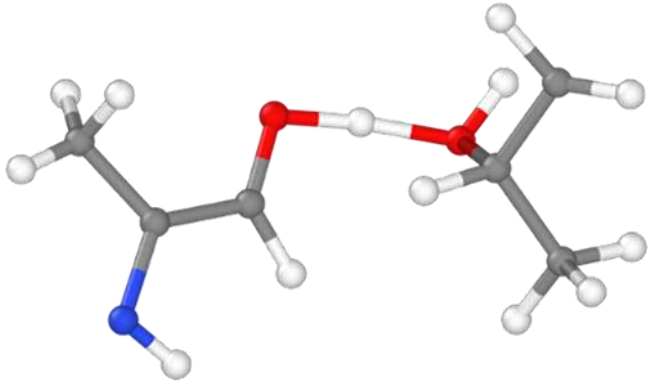
Close-up showing models refined with standard restraints (blue) and AQuaRef restraints (orange) superposed onto their higher-resolution homologous models (green) with their corresponding 2mFo-DFc Fourier maps contoured at 2σ.

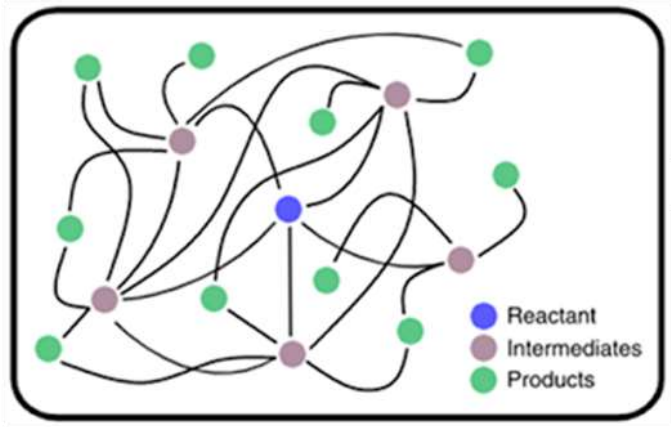
R. Zubatyuk, et al. AQuaRef: Machine learning accelerated quantum refinement of protein structures. BioRxiv 2024.07.21.604493; DOI: https://doi.org/10.1101/2024.07.21.604493

Phenix

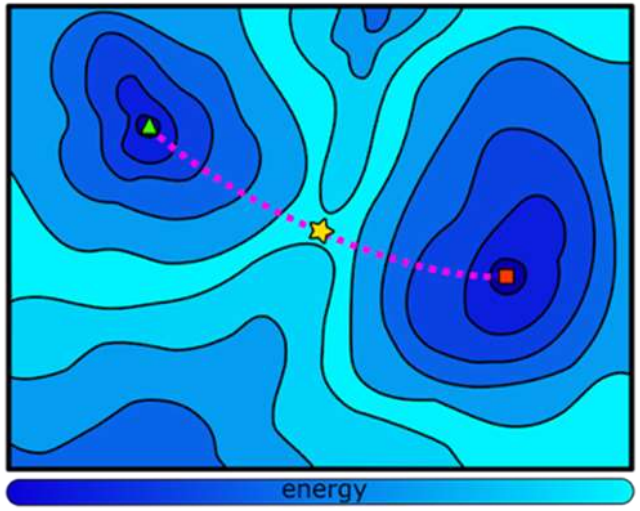
What is Next? AIMNet2 for Transition State Chemistry!

Research Overview: develop Neural Network Potentials that accelerate reaction pathway searches and support robust transition state conformational sampling



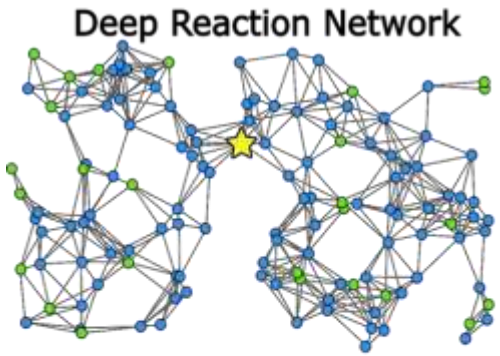


Zhao and Savoie, 2021, Nat. Comp. Sci.

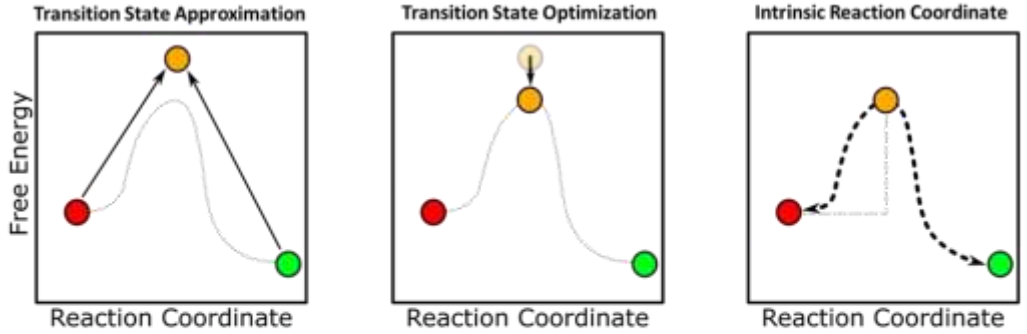


minimum energy pathway (MEP)
reactant/product
transition state

High-Throughput Reaction Characterization with AIMNet2-RXN



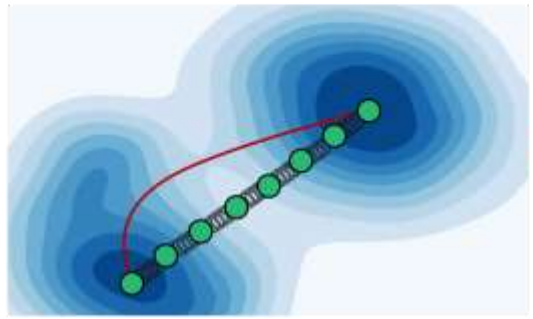
Synthesis Planning
Stressor-Specific Deconstruction
Engineered Degradation
Biochemical Pathways
...

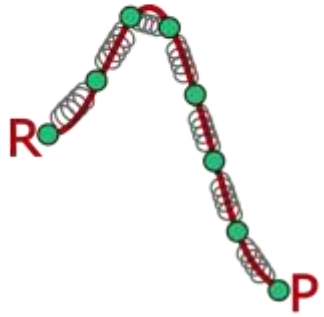


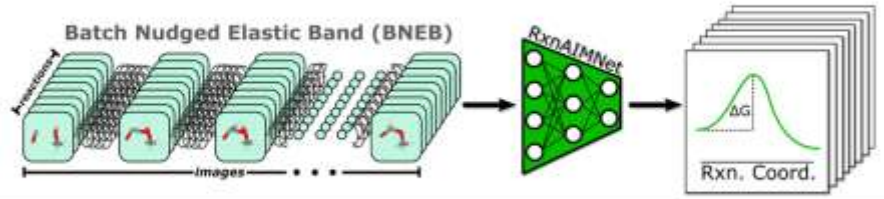
approximate time required per reaction with DFT: hour(s) – several days

Nudged Elastic Band Method

(i.e., minimizing atomic and spring forces)

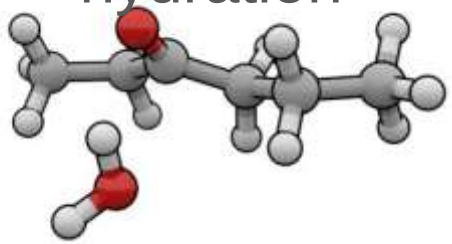




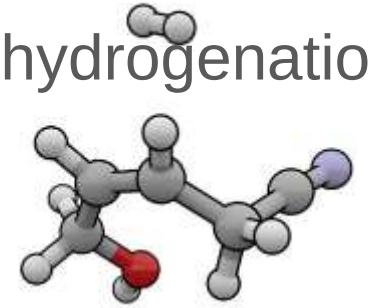


~350,000 Reactions per day w/ 1 RTX3090 GPU

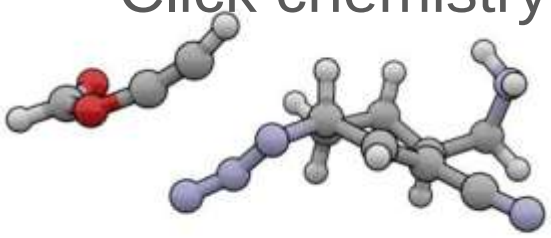
hydration



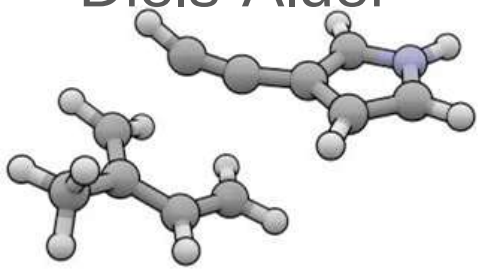
hydrogenation



Click chemistry



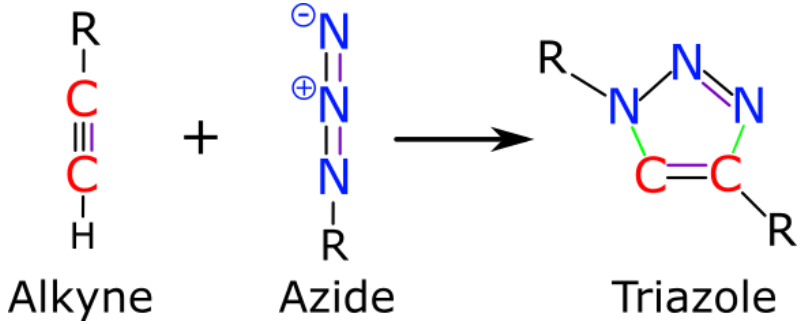
Diels-Alder

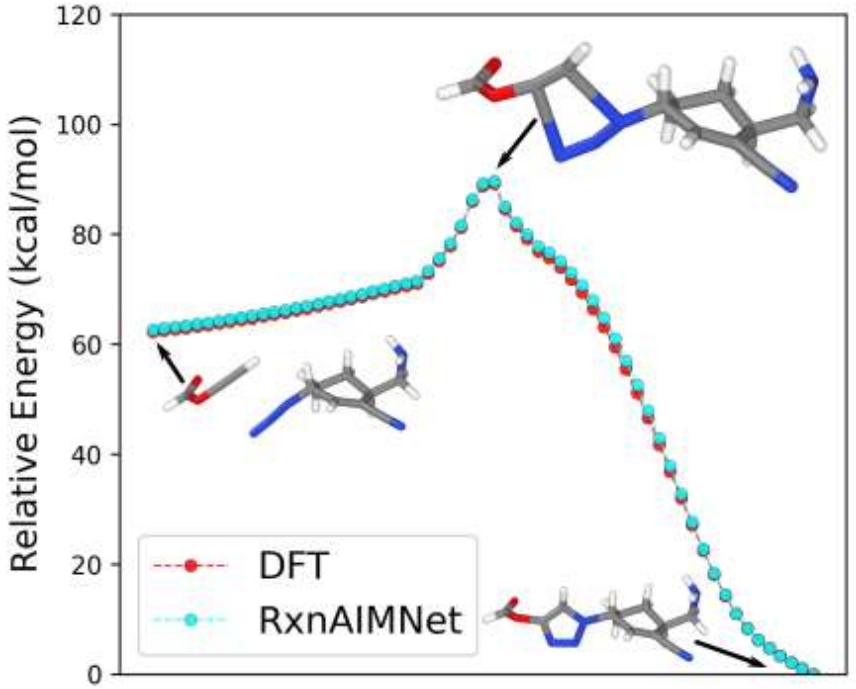


Coming Soon!

AIMNet2-RXN Covers Diverse Mechanisms

Triazole synthesis via click chemistry

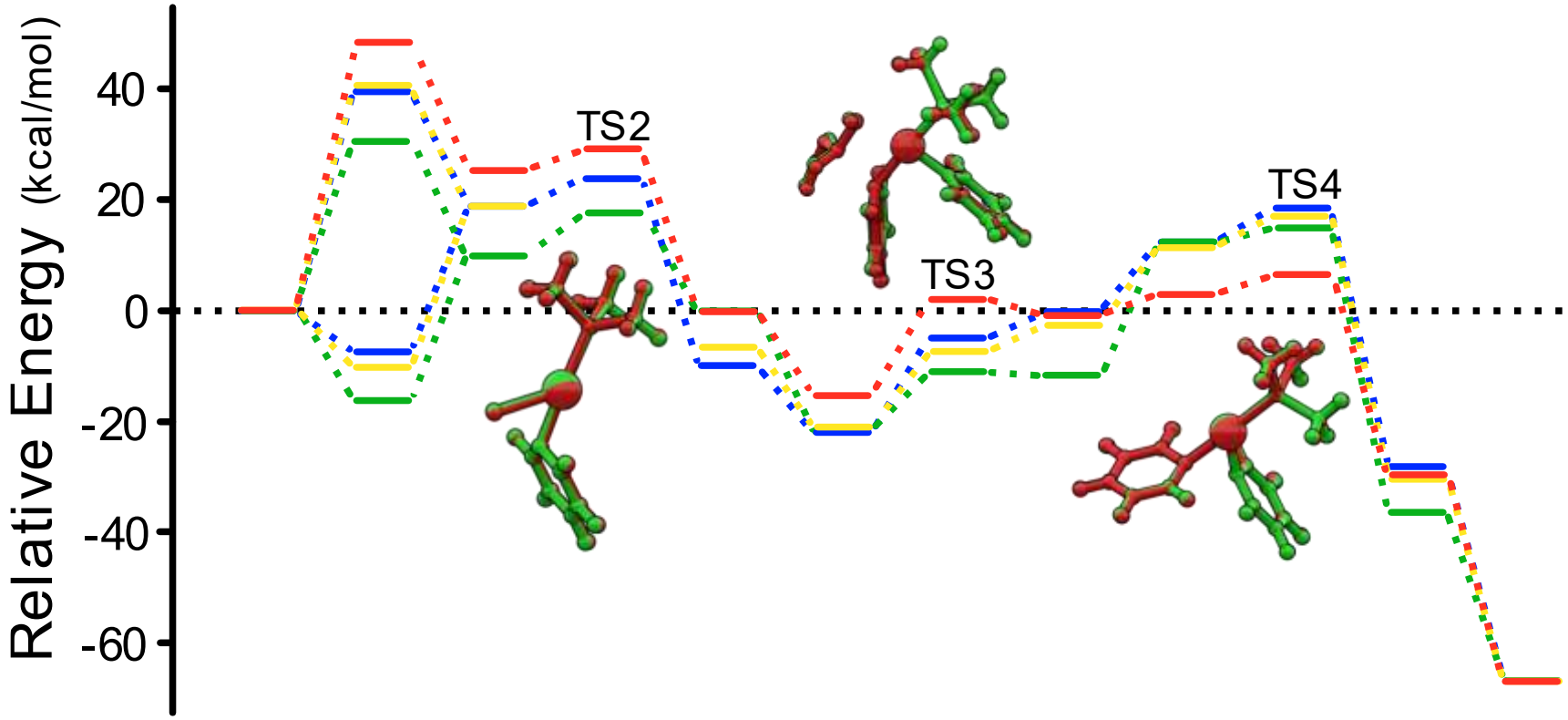




Reaction Mechanisms Explored (thus far)

- Diels-Alder
- Triazole formation (click chemistry)
- Combustion
- Tautomerization
- Hydrogen transfer
- Esterification
- Aldehyde and Ketone formation
- Metathesis
- Ring-closing and ring-opening
- Amine protection-deprotection
- …

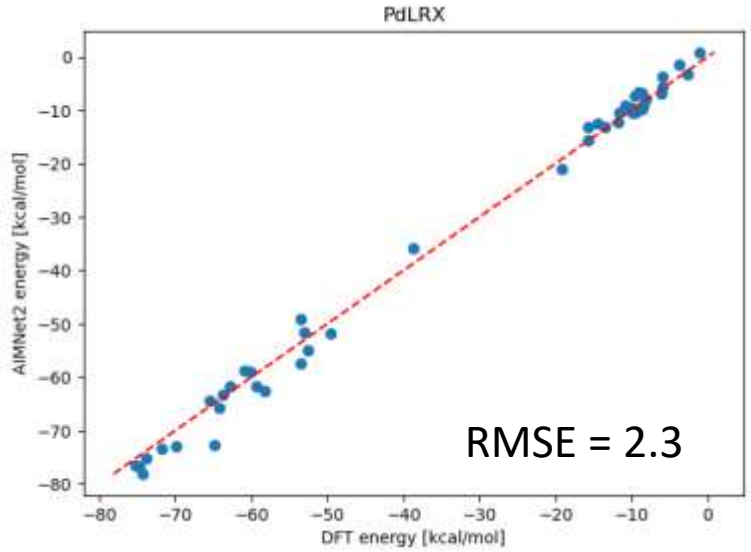
AIMNet2-Pd: Reactive Model for Pd-catalyzed C-C cross-coupling

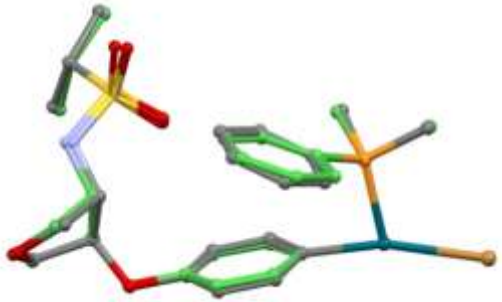


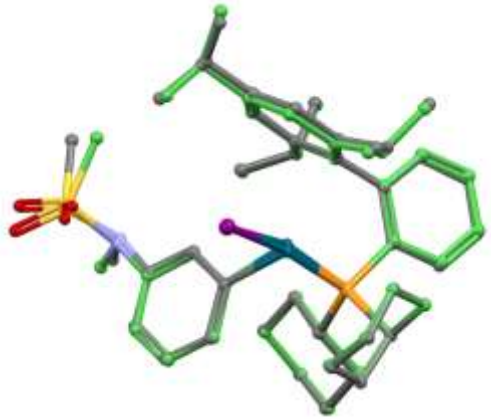
AIMNet2-Pd is a tailored model for **rapidly profiling multi-step mechanisms that define Pd-based Suzuki cross-coupling reactions.**

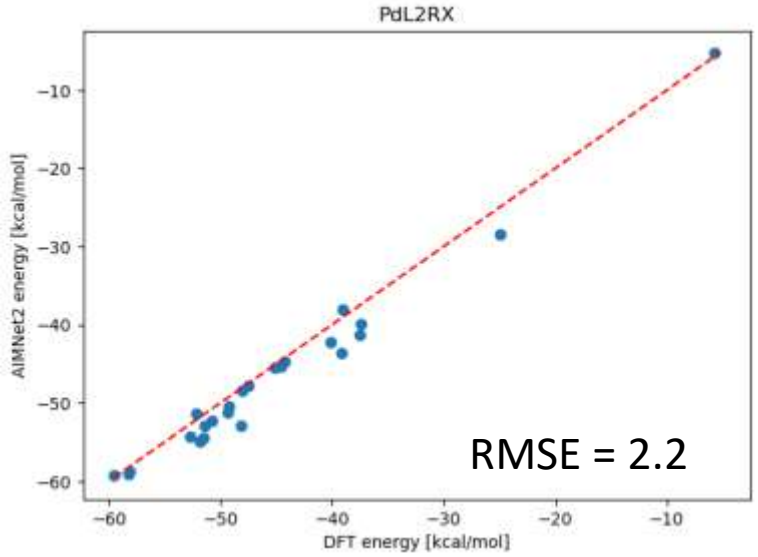
C-CAS Center for Computer Assisted Synthesis

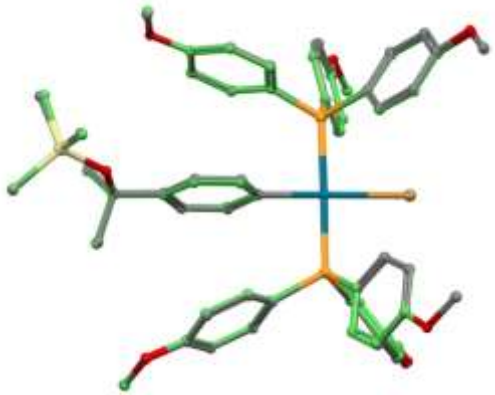
Pd-catalyzed C-C cross-coupling reaction: benchmarking energies of oxidative addition









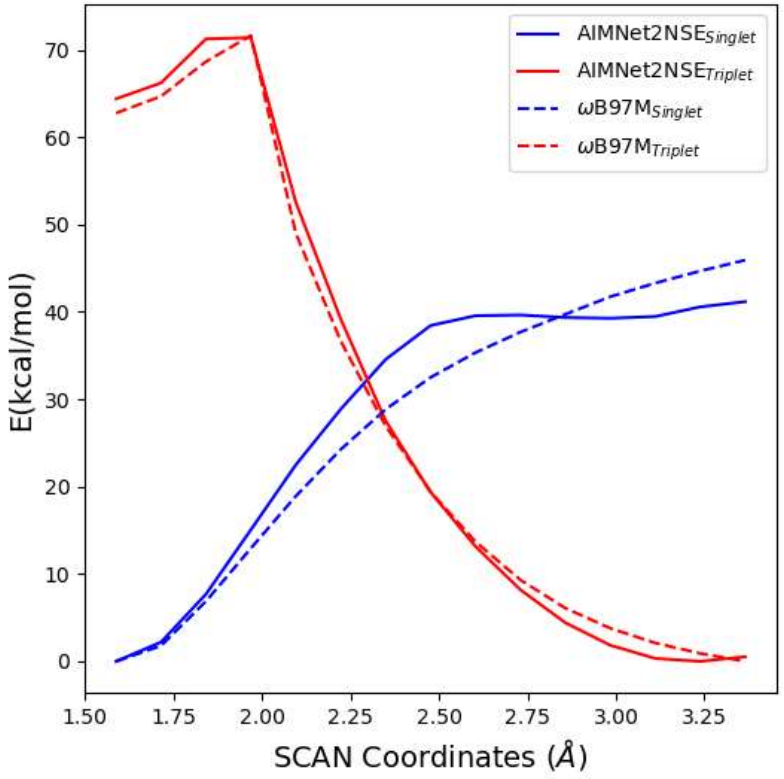


Reference systems:

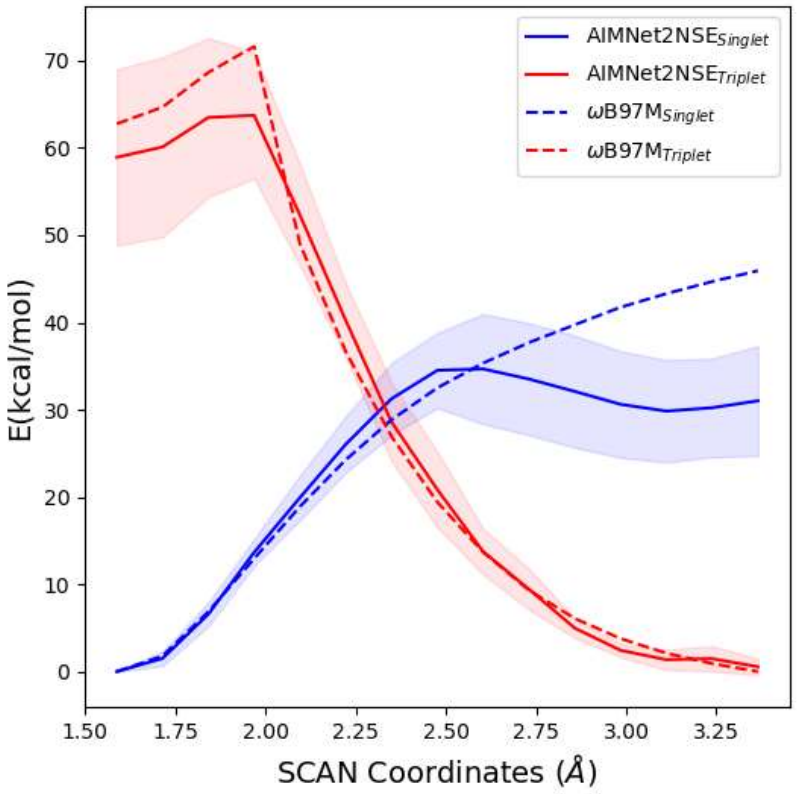
- PdL_2 for mon-P ligands
- PdL for bi-P ligands and P-π ligands
- R_1X

C-CAS Center for Computer Assisted Synthesis

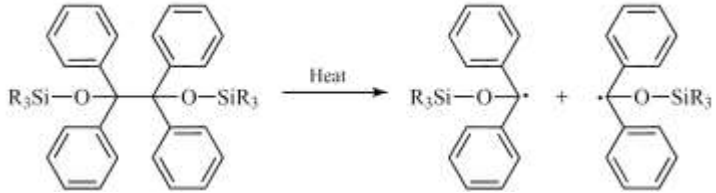
Radical Decomposition of Silylated Benzopinacol



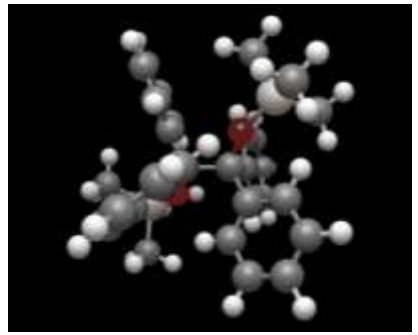
AIMNet2-NSE Single-Point Energy on DFT geometry evaluated with an ensemble average of 4 models



The average energy of geometries calculated with 4 AIMNet2-NSE models with standard deviation and comparison to DFT reference.



R = CH_3



Use the ANI-1x/2x potential:

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

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

Use the AIMNet2:

Anstine D, Zubatyuk R, Isayev O. ChemRxiv Preprint. 2023; https://doi.org/10.26434/chemrxiv-2023-296ch

AIMNet implementation in Pytorch
Available at: https://github.com/isayevlab/aimnet2

Plugins: ASE, LAMMPS, OpenMM

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

ANI-2x dataset & COMP6 benchmarks:
https://zenodo.org/communities/aiqm

Used in Government labs, companies etc.