Coarse graining autoencoders and evolutionary learning of atomistic potentials

IPAM, Los Angeles,
September 24 2019

Rafael Gómez-Bombarelli
rafagb@mit.edu
http://gomezbombarelli.mit.edu/

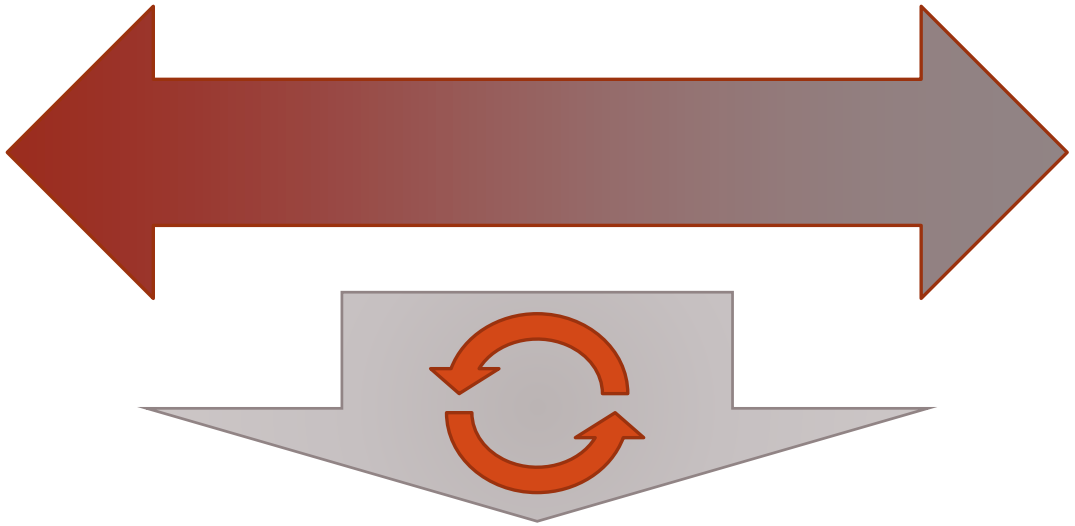
Computational spectrum - virtuous cycle

There is essentially a continuum of higher parametrization and statistical learning connecting first principles (theory-based simulations) to black-box statistical learning over experiments.

First Principles

Out-of-the-box
No/Little fitting
Extrapolates
High-throughput
Sometimes cheap

- but -
As good as model
Sometimes costly



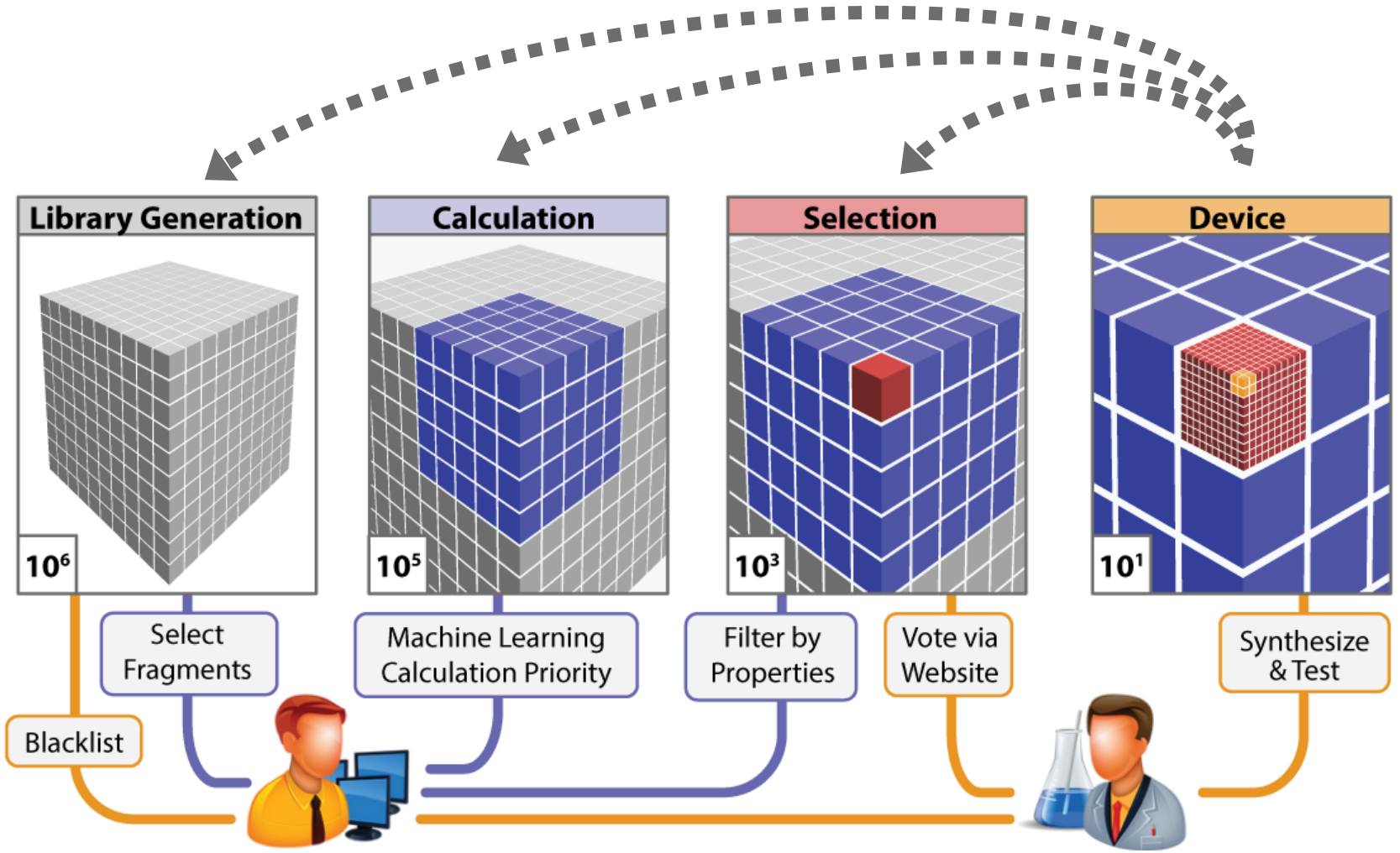
Discovery

Machine learning

Fast
Uncanny performance
Leverage large data

- but -
As good as training data
Expensive / proprietary data

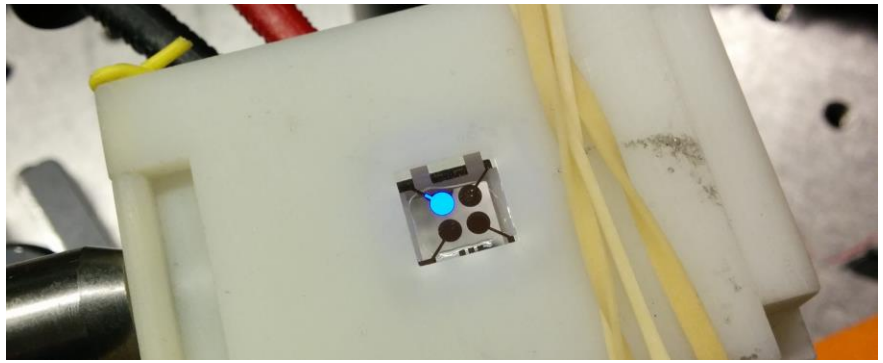
High-throughput virtual screening

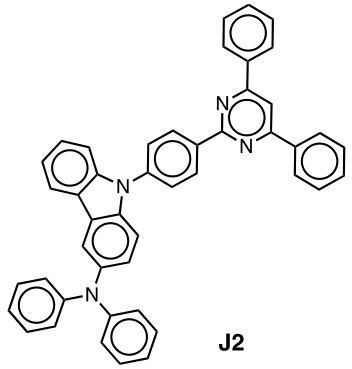


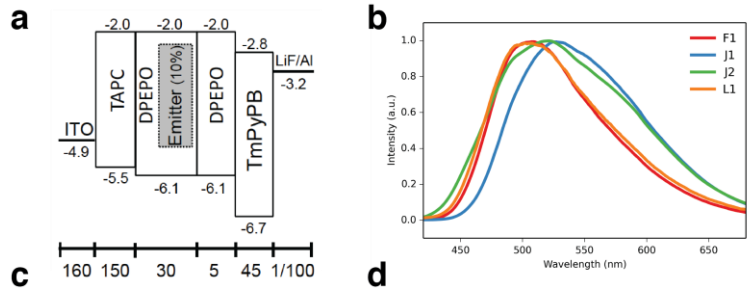
Successful applications

Organic Light Emitting Diodes

- High end displays, potentially lighting.
- Lightweight, flexible, transparent, high contrast, low power

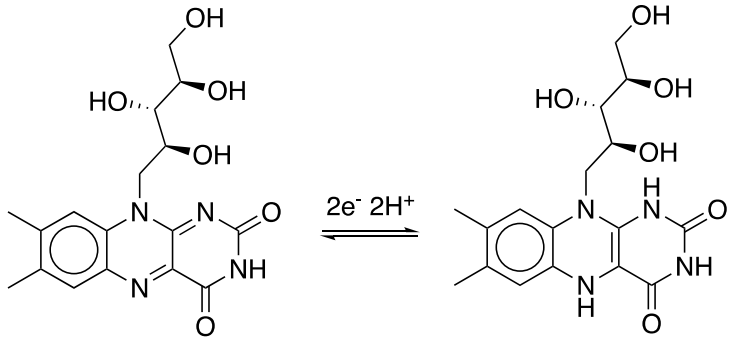






Organic Flow battery electrolytes

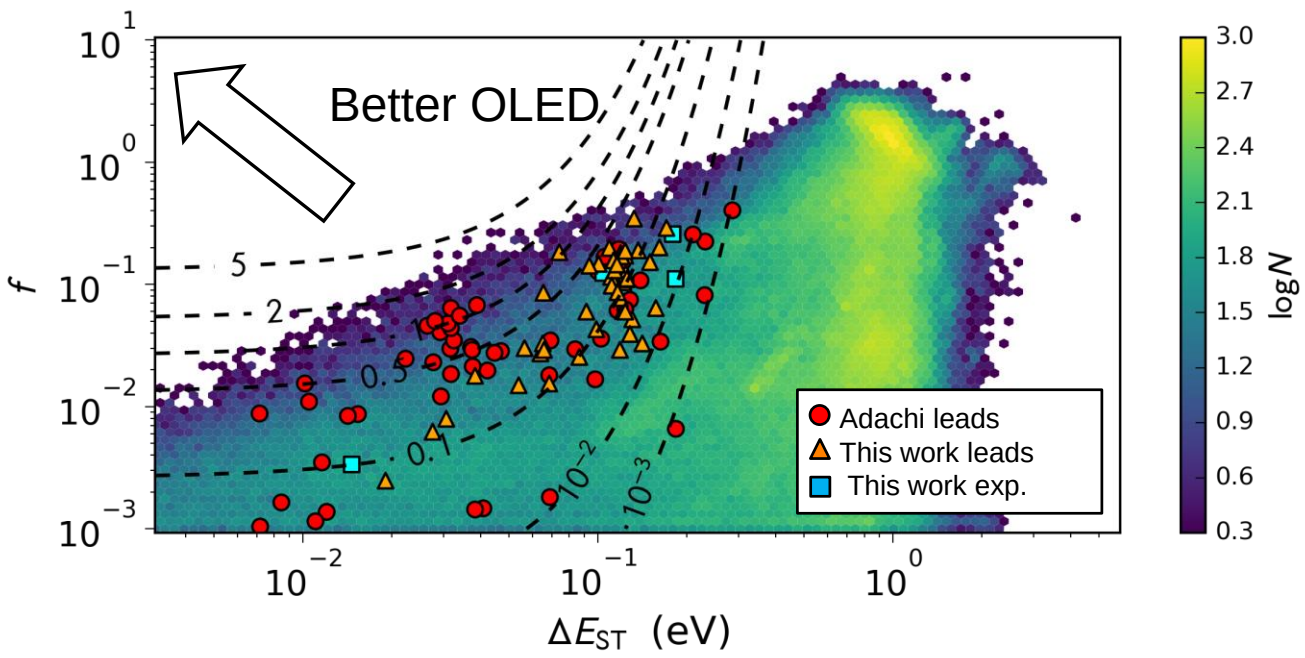
- High-scale energy storage
- Emerging technology, promising low-cost

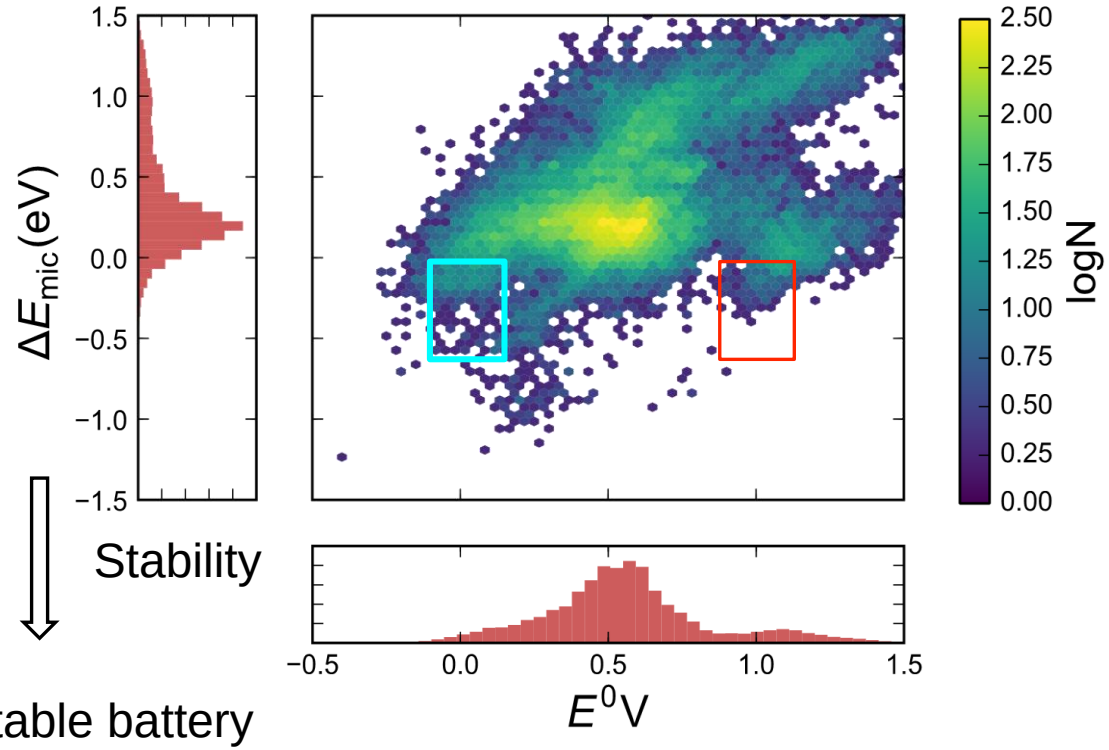


RGB et al. *Nature materials* **2016** *15,* 1120

Lin, K. *et al. Nature Energy* **2016** *1,* 16102

Charting chemical space





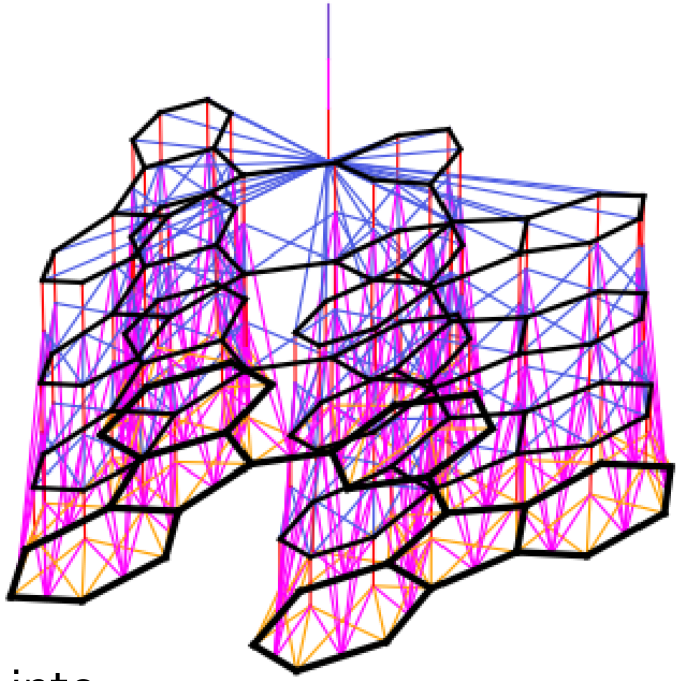
GCNN for interpolating

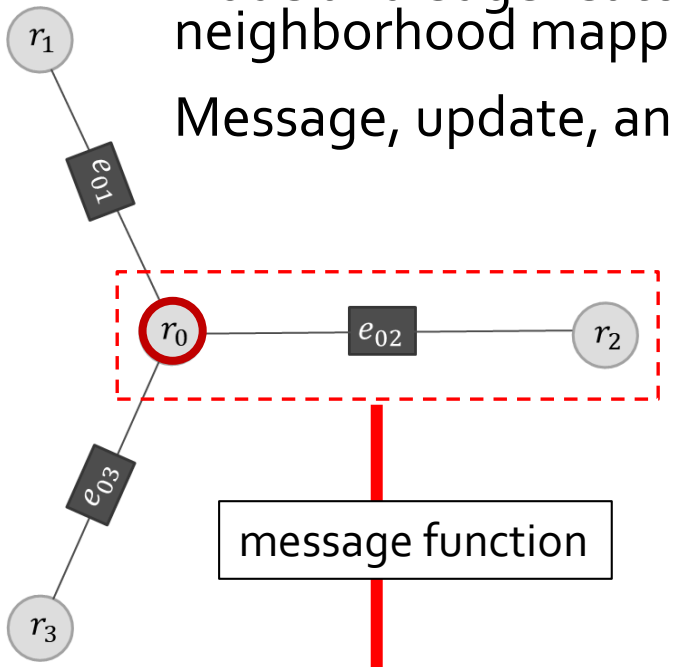
Cheap surrogate function to save useless computations

Atoms represented as nodes and bonds as distance-labeled edges

Node and edge features updated iteratively based on learned neighborhood mappings

Message, update, and embedding functions are neural networks





message function

m_{02}

Combine messages from all neighbors

$$m_0 = m_{01} + m_{02} + m_{03}$$

update function

r_0'

Pool atomic into molecular feature

$$\{r_0', r_1', r_2', r_3'\}$$

embedding function

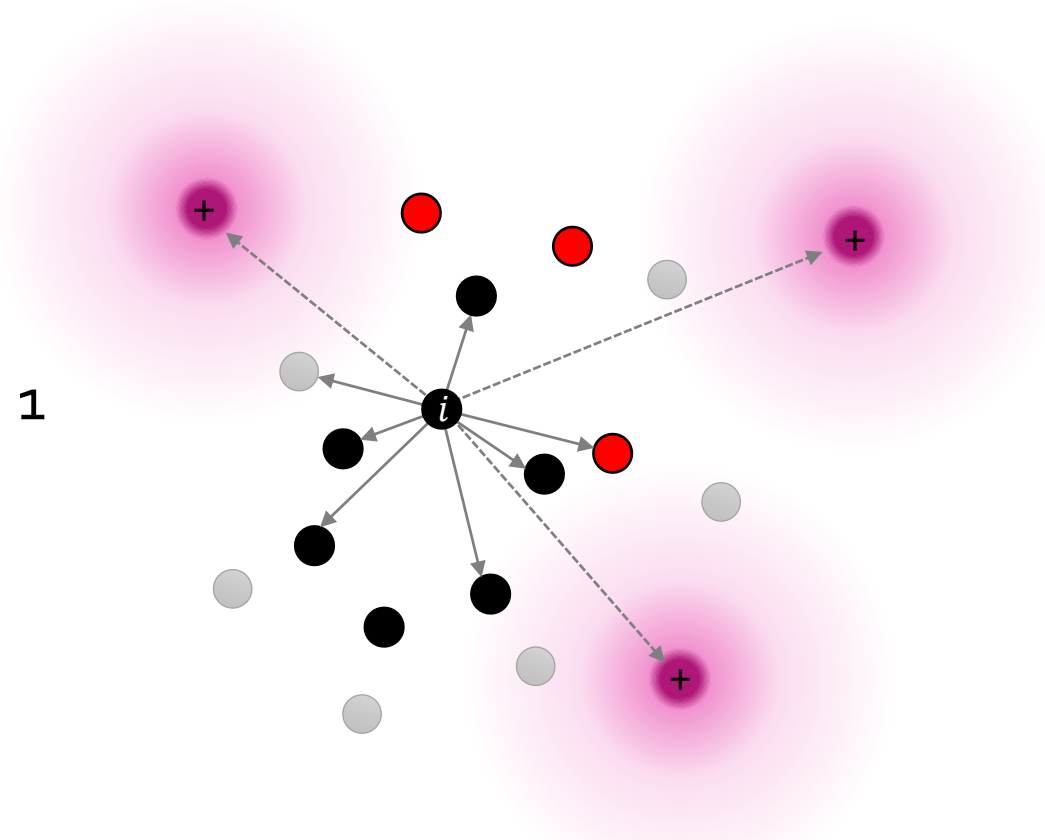
f'

3D vs 2.5D ?

Duvenaud *et al.* arXiv:1509.09292, J Gilmer et al arXiv:1704.01212v2

GCNN and Neural Potentials

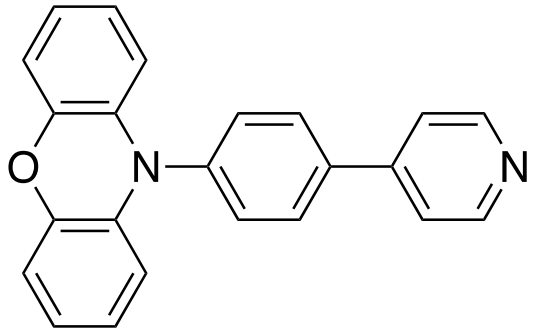
- Typically achieve state of the art performance over topological QSPR regression problems in chemistry
- Require 10^4 or more to be truly effective
- As interatomic potentials < 1 kcal/mol energies and 1 kcal/mol/A forces (as low as 0.1)
- Can overfit in chemical and configurational space



Smith, J. S. et. al.. *Sci. Data* **4**, 170193 (2017); Smith, J. S *Chem. Sci.* **8**, 3192–3203 (2017). Smith, J. S. et. al. *J. Chem. Phys.* **148**, 241733 (2018); Schüt et al. *J. Chem. Phys*. 148, 241722 (2018); Hansen, K. et al. *J. Phys. Chem. Lett.* **6**, 2326–2331 (2015); Chmiela, S. et al. *Sci. Adv.* 3, e1603015 (2017); Schütt, K. T. *Nat. Commun.* **8**, 13890 (2017); Chmiela, S. *Nat. Commun*. **9**, 3887 (2018).

Inverse design

Progress in predicting performance given candidate

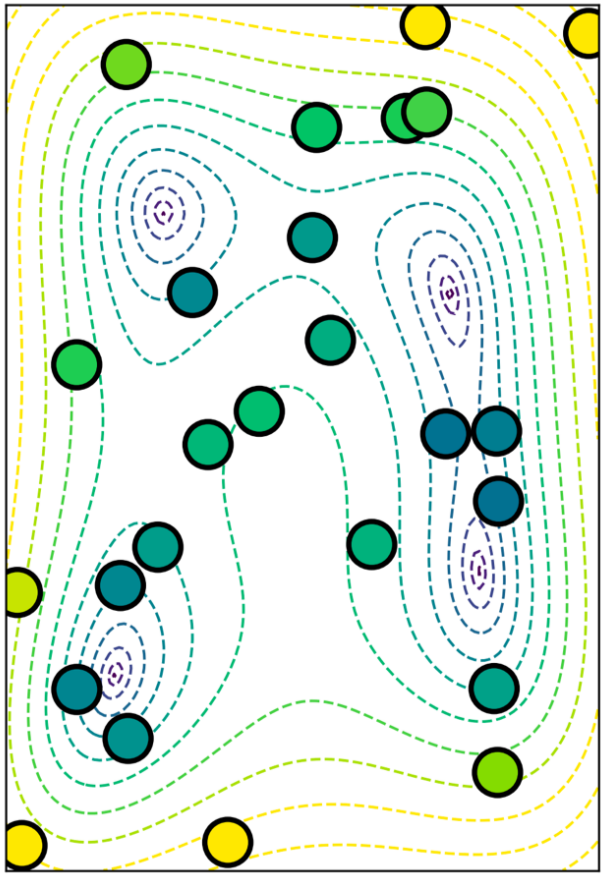


Can we generate candidate based on design targets?

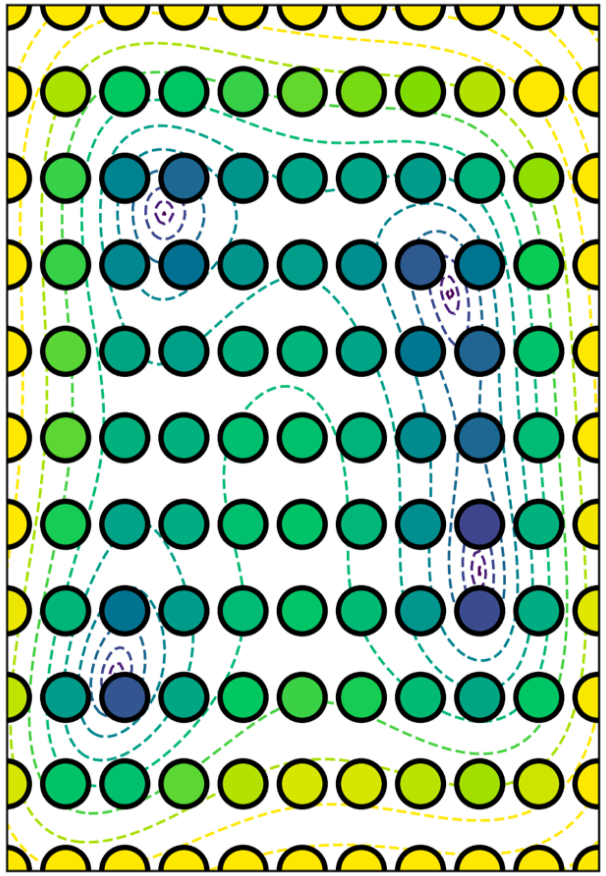
?

Inverse design

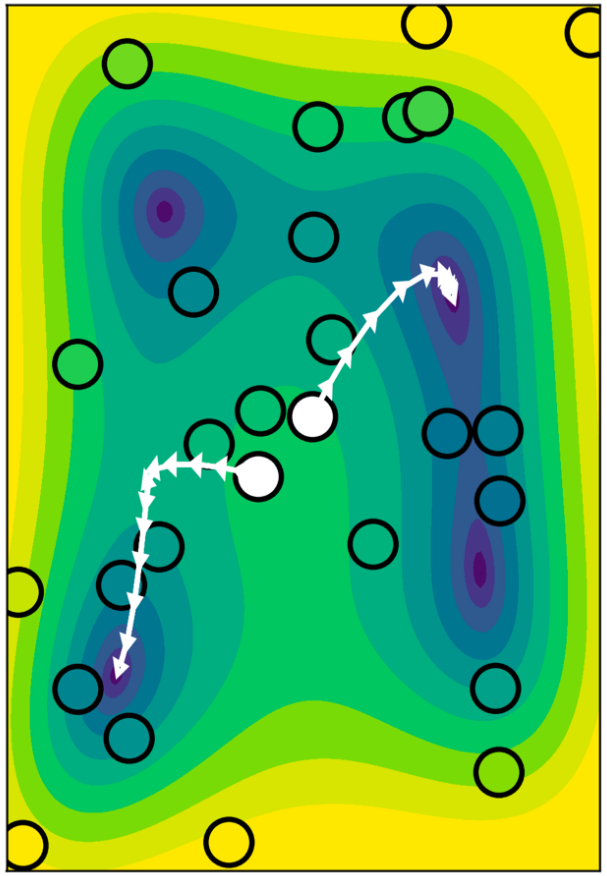
Trial and error



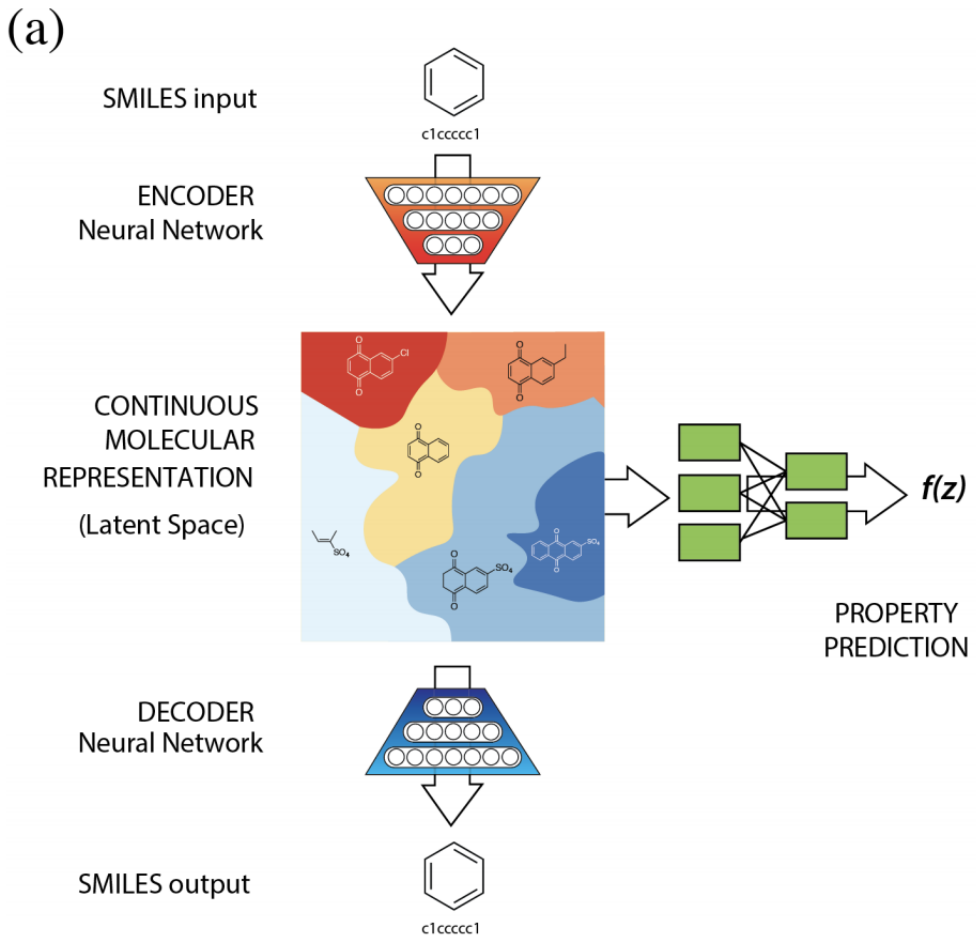
High-throughput



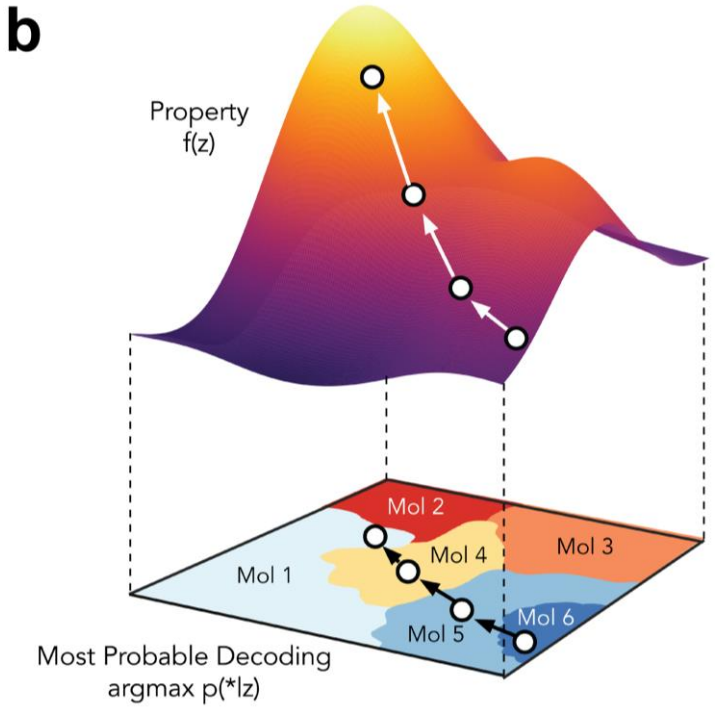
Inverse design



Semisupervised Molecular VAE



The latent representation now encodes mapping to one or more properties.



Gomez-Bombarelli et al. ACS Central Science **2018**, *4*(2), 268–276

Chemical space is different

IN THE PIPELINE

Derek Lowe's commentary on drug discovery and the pharma industry. An editorially independent blog from the publishers of *Science Translational Medicine*. All content is Derek's own, and he does not in any way speak for his employer.

By Derek Lowe

CHEMICAL NEWS

Calculating A Few Too Many New Compounds

By Derek Lowe | 8 November, 2016

In chemistry, one's ideas, however beautiful, logical, elegant, imaginative they may be in their own right, are simply without value unless they are actually applicable to the one physical environment. (Woodward)

Is there a distribution we want to draw from?

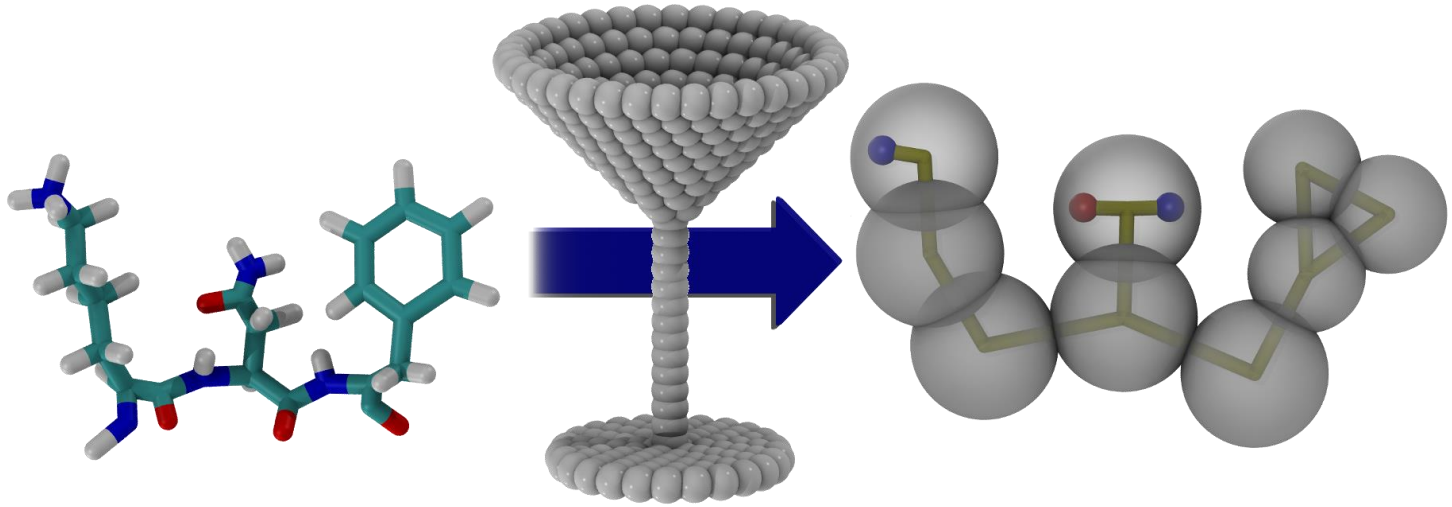
Coarse-graining MD

TOWARDS INVERSE DESIGN IN 3D

Coarse Grained Methods

Coarse Graining MD simulates coarse grained variables that represents **slow collective atomistic motions** derived from full atomistic simulations

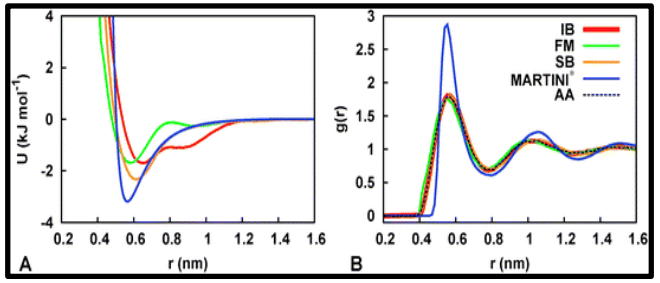
Coarse Graining methods find the "effective" coarse grained potential **given** a predetermined coarse graining mapping



S. J. Marrink et al. J. Phys. Chem. B **111,** 7812 (2007).

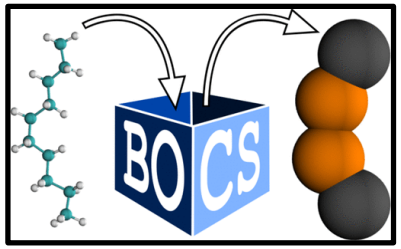
Coarse Grained Methods

Extensively studied how to find the coarse graining potentials that reproduces ***equilibrium*** structural correlation function from atomistic simulations given a pre-determined CG mapping



[1]

Methods to approximate Coarse Grained Force Fields: Relative Entropy, Force Matching, g-YBG (implemented in VOTCA, BOCS, etc.)



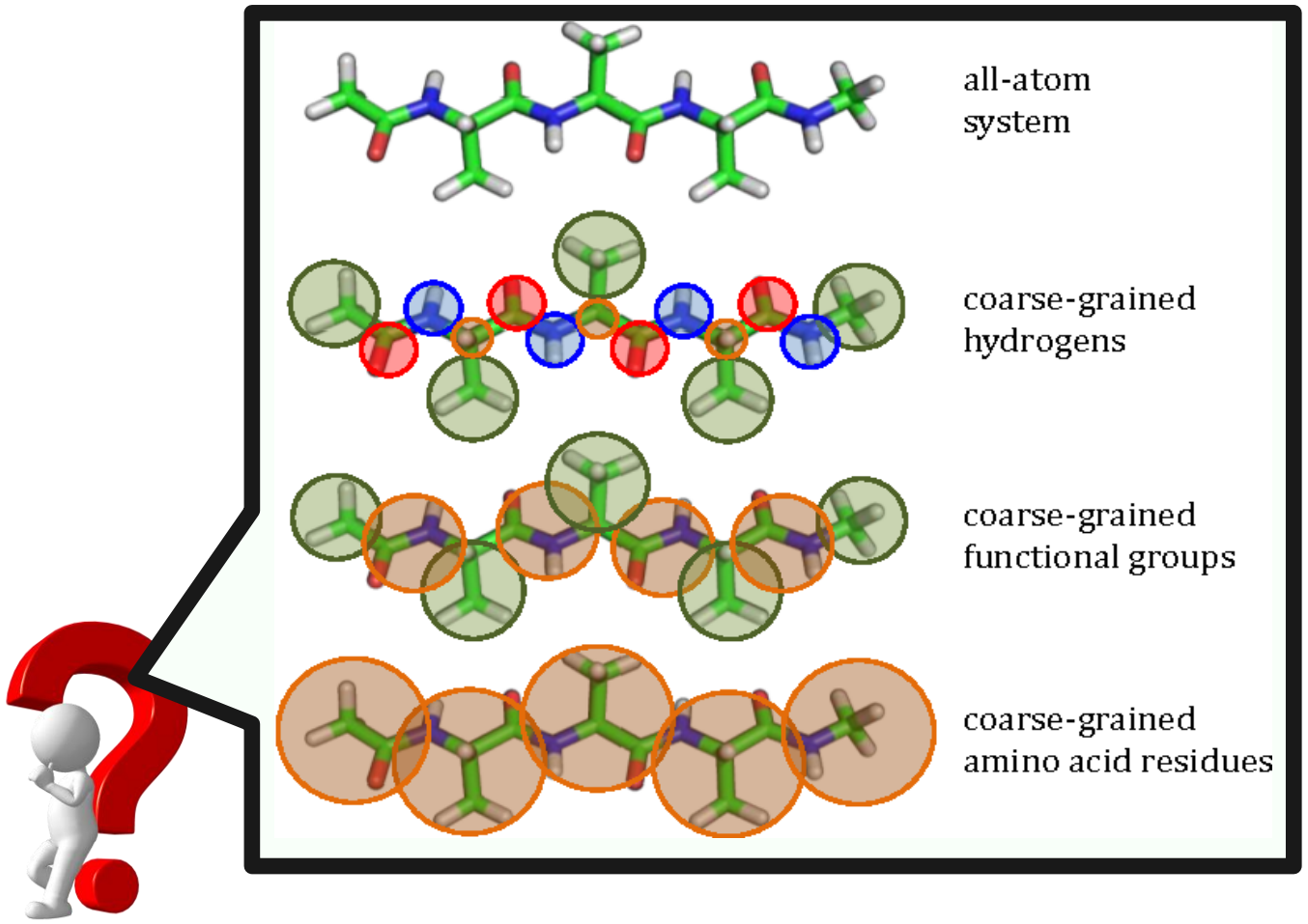
[2]

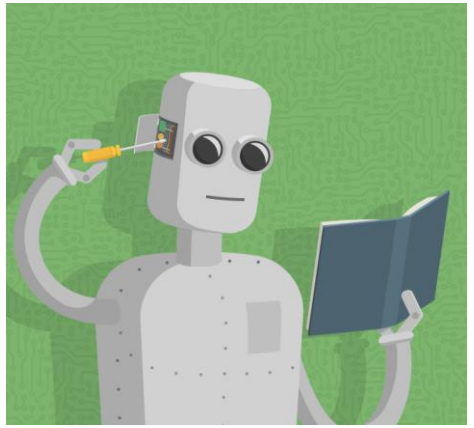
Systematic Coarse Grained force fields for Biomolecules: MARTINI

[3]

[1] A. J. Rzepiela et al. Phys. Chem. Chem. Phys. **13**, 10437 (2011). [2] N. J. H. Dunn et al. J. Phys. Chem. B **122**, 3363 (2018). [3] S. J. Marrink et al. J. Phys. Chem. B **111**, 7812 (2007).

Learning to Coarse-Grain

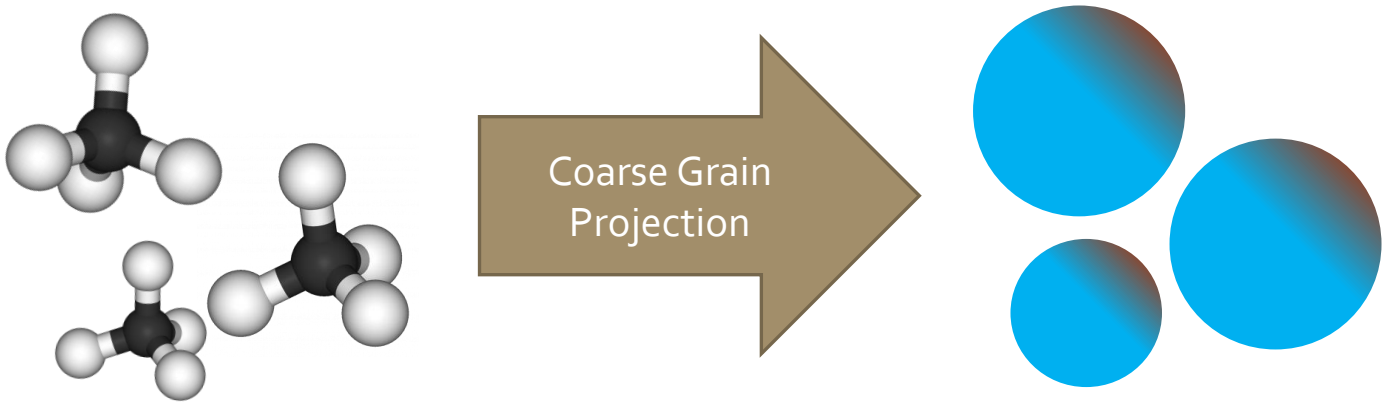




A learning problem

M. S. Shell, in Adv. Chem. Phys. (Wiley-Blackwell, 2016), pp. 395–441.

Coarse-graining framework



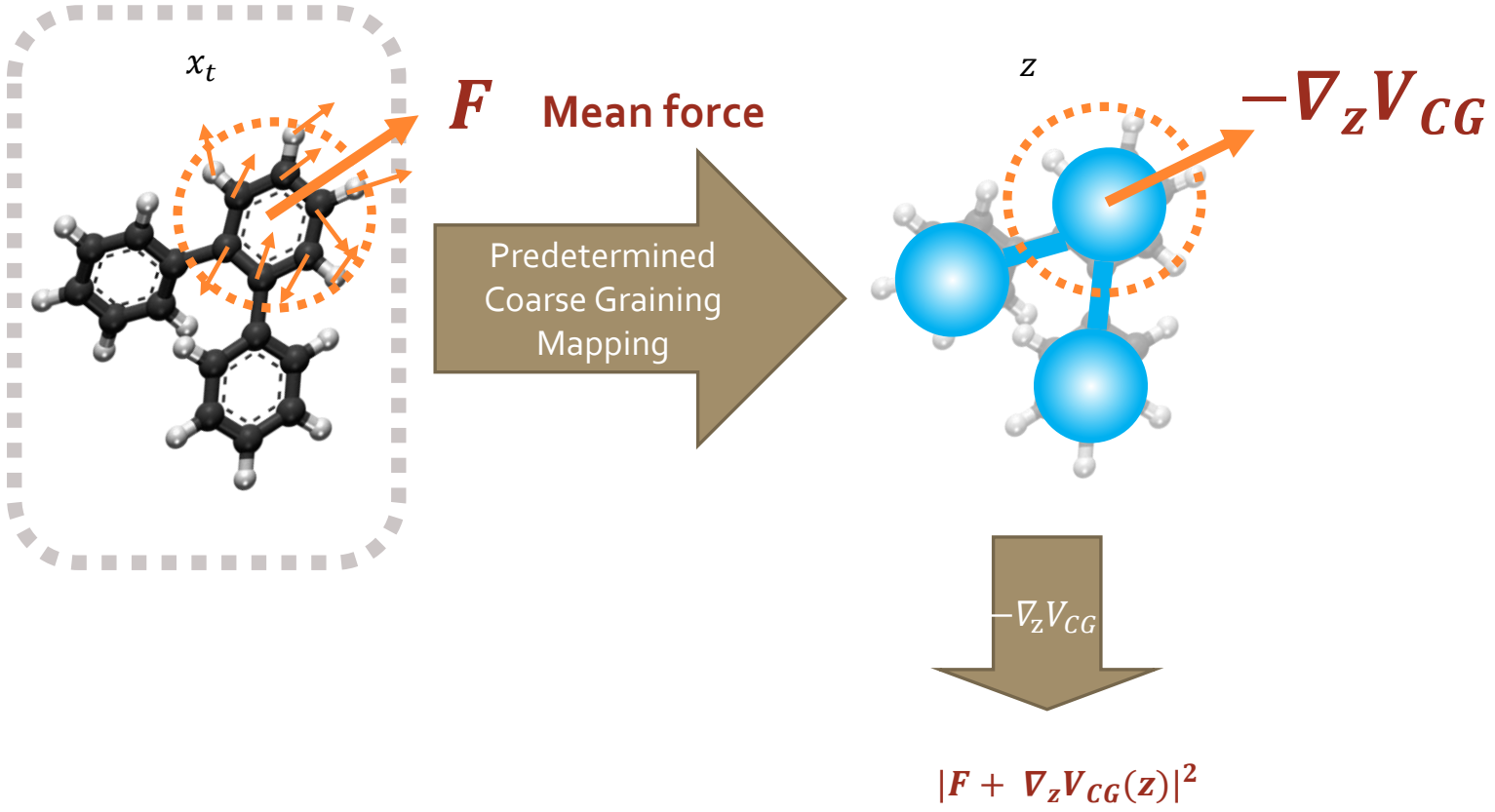
$\{x, V(x)\}$ $\quad z = E(x)$ $\quad \{z, V_{CG}(z)\}$

$\mathbf{x}$: atomistic coordinates
$\mathbf{V(x)}$: All-Atom Potential
$\boldsymbol{z}$: coarse grained coordinates
$\boldsymbol{V_{CG}(z)}$: coarse grained Potential

Instead of
given $\boldsymbol{x}, \boldsymbol{V(x)}, \boldsymbol{E(x)} \rightarrow$ find $\boldsymbol{V_{CG}(z)}$
We propose
given $\boldsymbol{x}, \boldsymbol{V(x)}$ $\rightarrow$ find $\boldsymbol{E(x)}$ and $\boldsymbol{V_{CG}(z = E(x))}$

$V_{CG}(z)$ can have an arbitrary functional form
- Classical
  $\boldsymbol{V_{CG}(z) = V_{bonded}(z) + V_{non-bonded}(z)}$
- Neural (using MPNN)

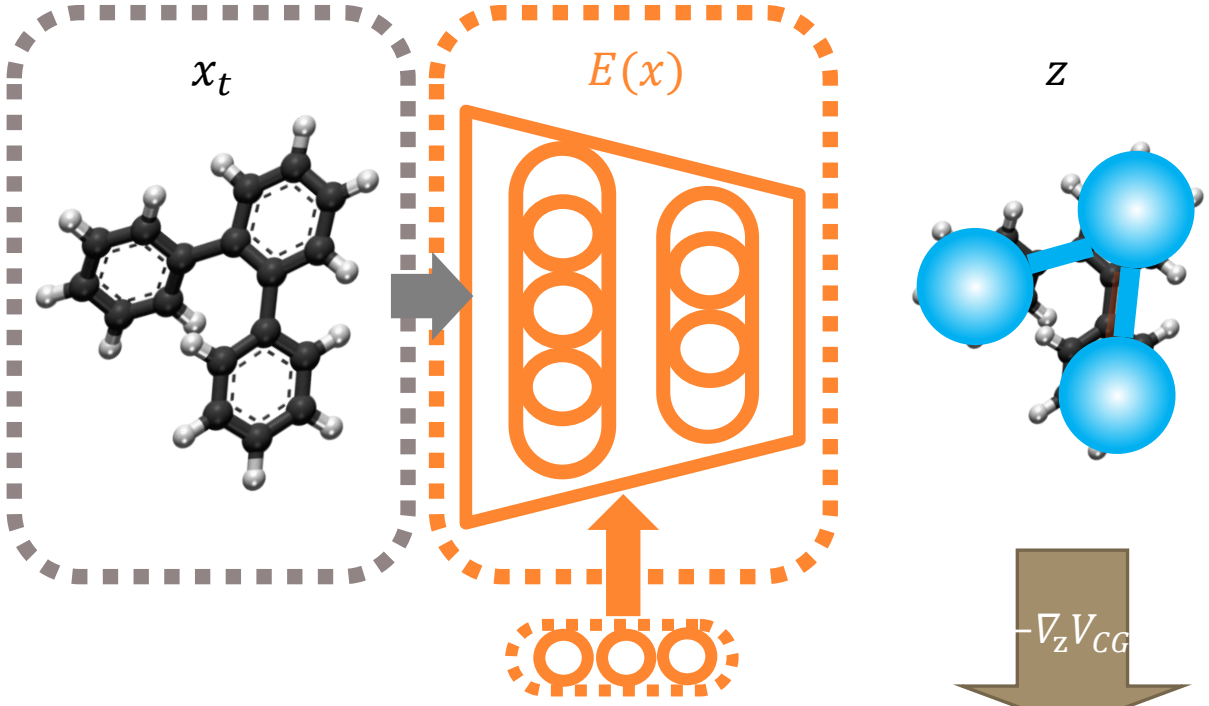
Force matching



- Force matching finds $\nabla V_{CG}(z)$ that best approximates the mean force F.
- We want to optimize the optimization target L :

$$L = |F + \nabla_z V_{CG}(E(x))|^2$$

The encoding function



[1] I: 1 … N coarse grained atom index
j: 1 … n atom index

We propose a neural network like encoding with the following constraints

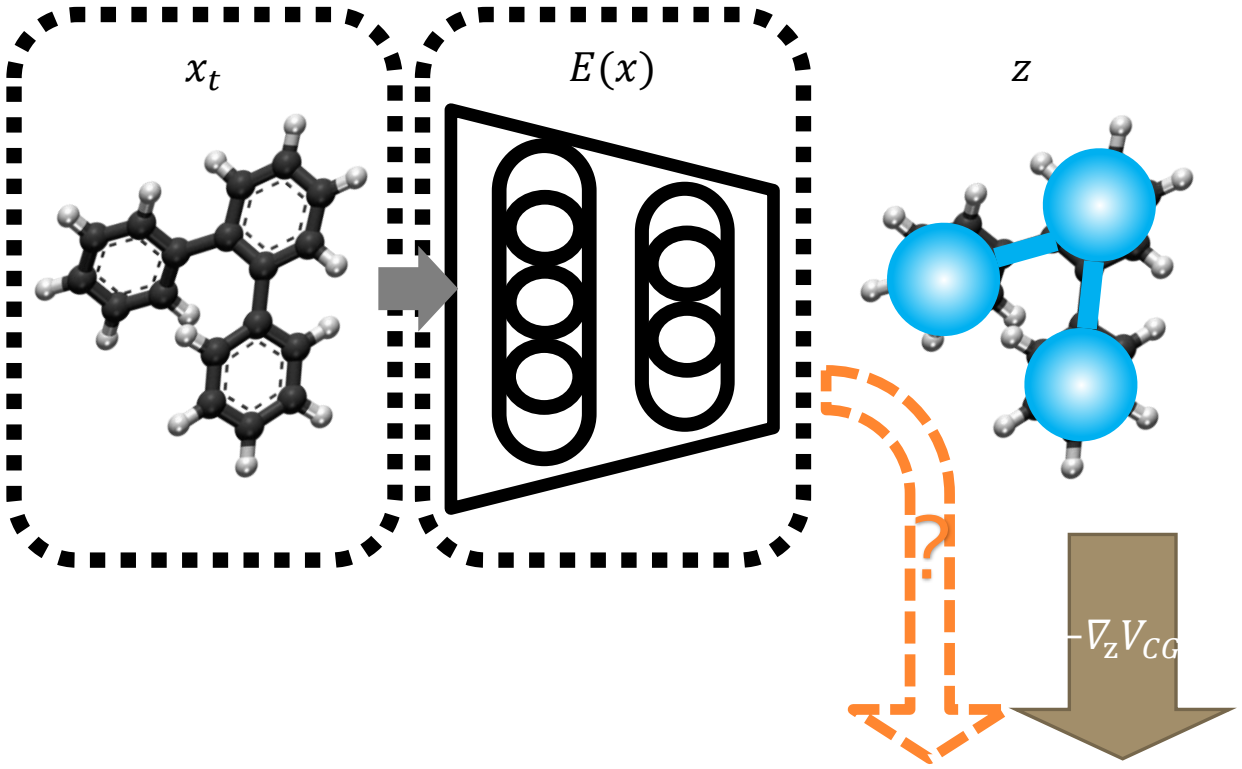
1) $z_I = E_I(x) = \sum_j E_{Ij}\, x_j$ [1]

2) $\sum_j E_{Ij} = 1$ and $E_{Ij} \geq 0$

3) $M_I = (\nabla E^{-1})^T M_j \nabla E^{-1}$

4) We use Gumbel-softmax during training to enforce the learning of **discrete** coarse graining variables to ensure that each atom only contributes to one CG atom

“Coarse Graining” Forces

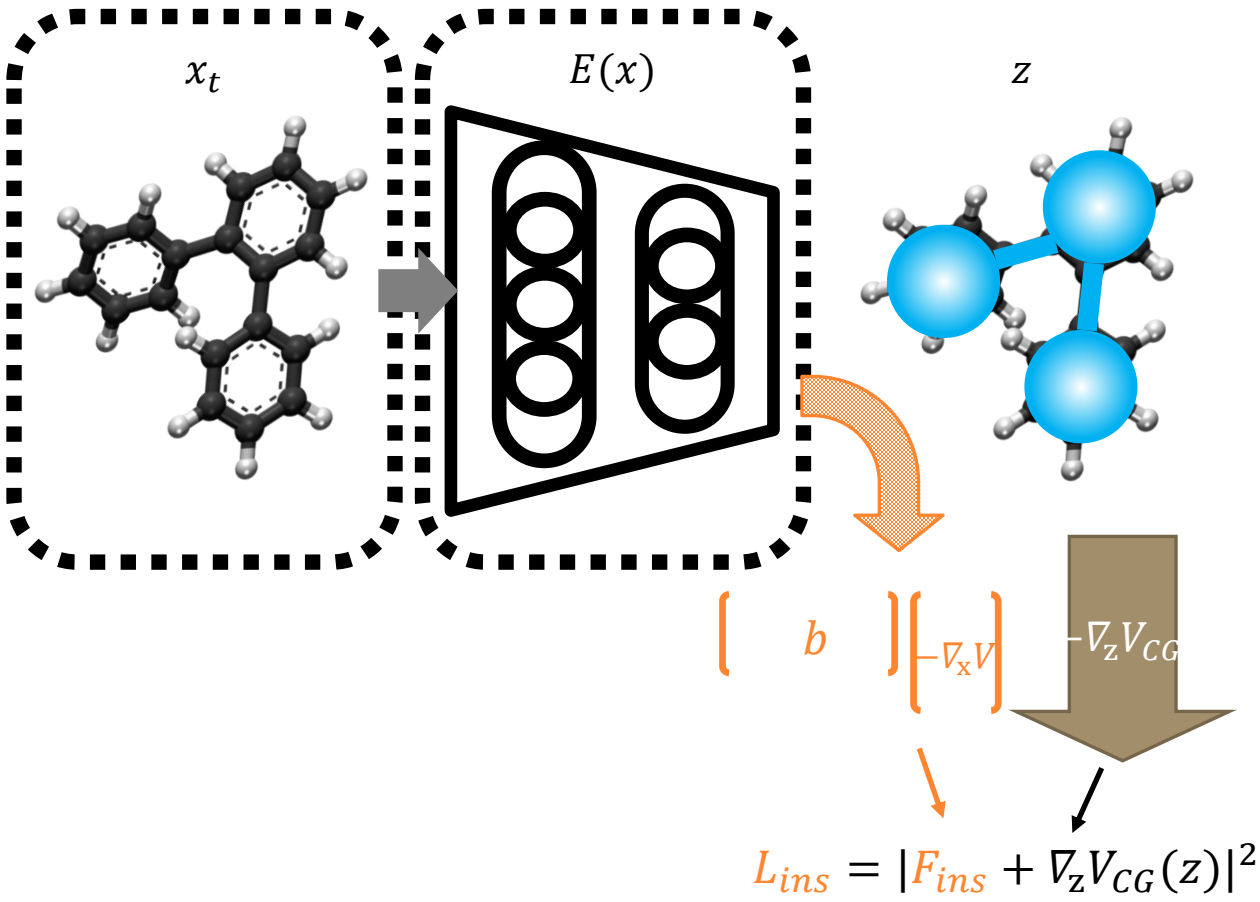


$$|\langle -\mathrm{b}\cdot\nabla V\rangle + \nabla_{\mathrm{z}}\,\mathrm{V_{CG}}(\mathrm{E(x)})|^2$$

- We also need a function that variationally determines F
- $F = \langle -\mathrm{b}\cdot\ \nabla V(x)\rangle_{E(x)=z}$ where b is the **force coarse graining function**
- A consistent choice for b from statistical mechanics: $\mathrm{b} = \frac{\nabla \mathrm{E(x)}^{\mathrm{T}}}{\nabla \mathrm{E(x)}^{\mathrm{T}}\cdot\nabla \mathrm{E(x)}}$
- Computing the mean force F requires constrained dynamics.
- However, we want a one-shot optimization stack without running extra MD simulations.

[1] E. Darve, *Numerical Methods for Calculating the Potential of Mean Force*.

Stochastic Force Matching



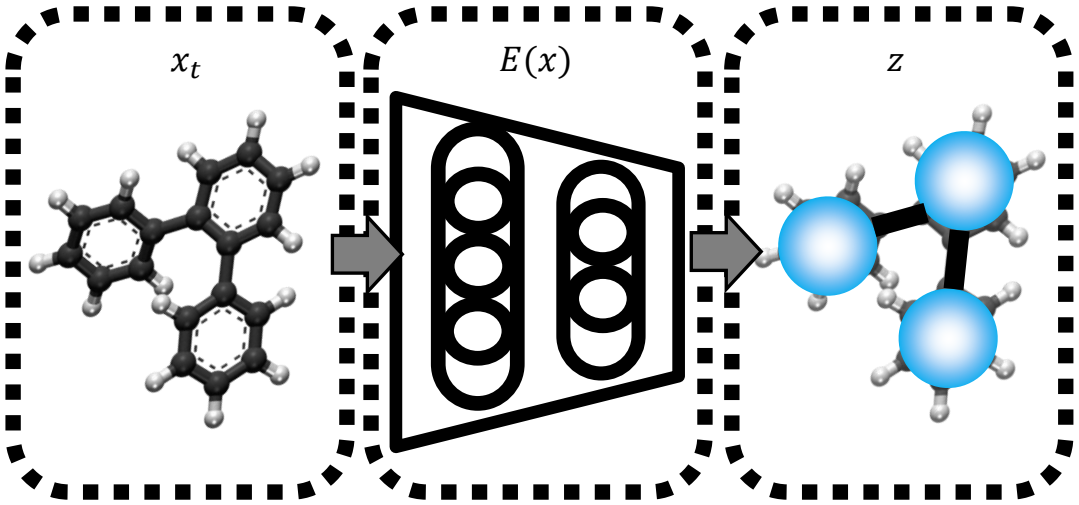
- We compute the instantaneous stochastic "coarse grained" force F_{ins} [1]
- $F_{ins} = -\mathrm{b} \cdot \nabla V(x)$ (instantaneous force)
- $F = \langle F_{ins} \rangle_{E(x)=z}$ (mean force)

$$L = |F + \nabla_z V_{CG}(E(x))|^2$$

$$L_{ins} = |F_{ins} + \nabla_z V_{CG}(E(x))|^2$$

[1] L. Zhang et al. J. Chem. Phys. **149,** 034101 (2018).

Coarse Graining Auto-Encoding Framework

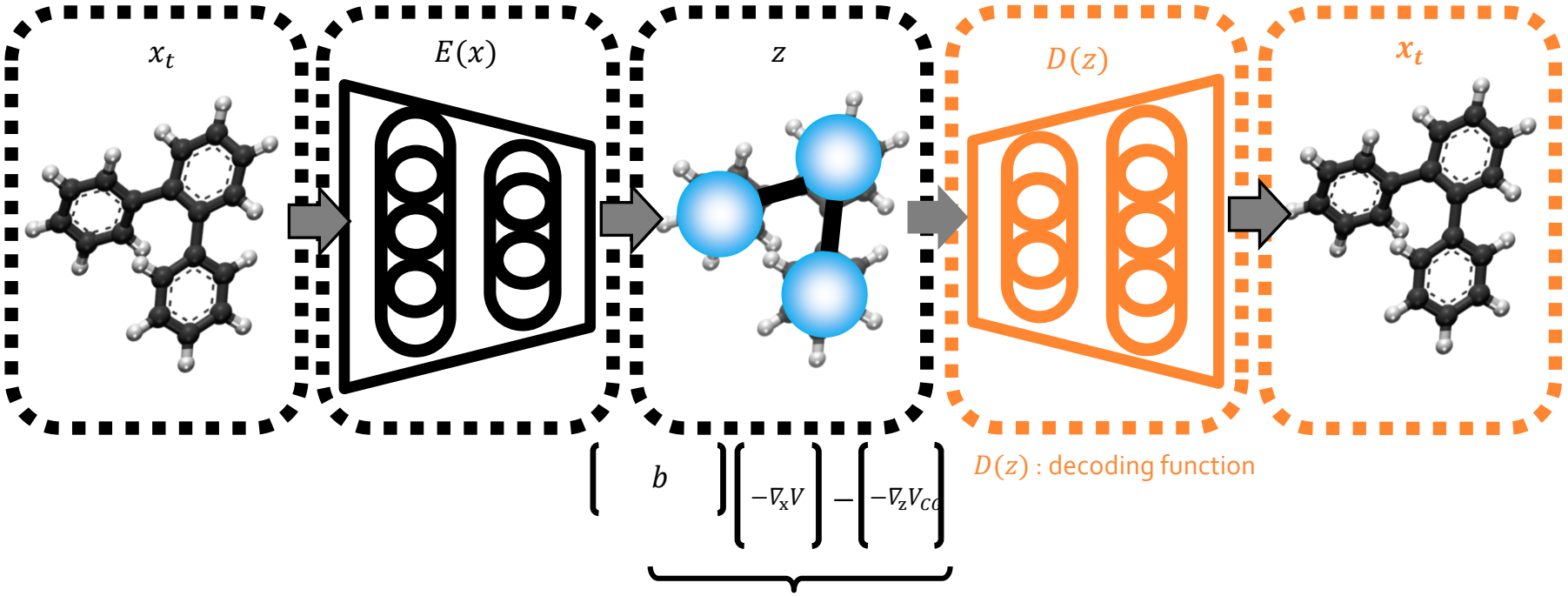


$$\begin{bmatrix} b \end{bmatrix} \begin{bmatrix} -\nabla_x V \end{bmatrix} - \begin{bmatrix} -\nabla_z V_{CG} \end{bmatrix}$$

$$\boldsymbol{L_{VCGE} = |F_{ins} + \nabla_z V_{CG}(E(x))|^2}$$

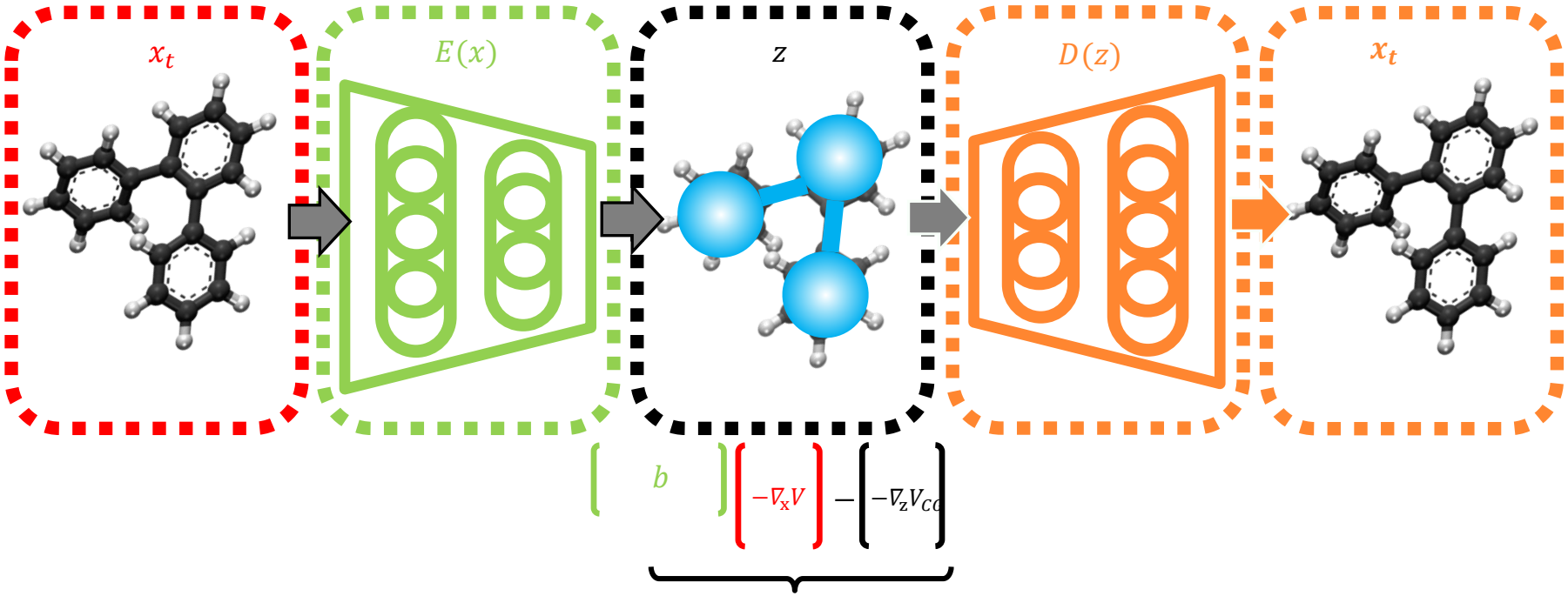
- **Supervised** force matching conditioned on E(x) learns coarse graining mapping field and the potential in CG.

Coarse Graining Auto-Encoding Framework



- **AutoEncoder** automatically coarse-grains atomistic coordinates to CG coordinates in a data-driven way
- **Force matching** also helps to shape the learning of CG and obtain $V_{CG}(z = E(x))$ for CG simulations

Coarse Graining Auto-Encoding Framework



$$L_{VCGE} = | - \nabla_x V \cdot b + \nabla_z V_{CG}(E(x))|^2 + |D\,E\,(x_t) - x_t|^2$$

- **AutoEncoder** automatically coarse-grains atomistic coordinates to CG coordinates in a data-driven way
- **Force matching** also helps to shape the learning of CG and obtain $V_{CG}(z = E(x))$ for CG simulations

Pipeline

Run short all-atom MD trajectory on small system.

Shuffle frames

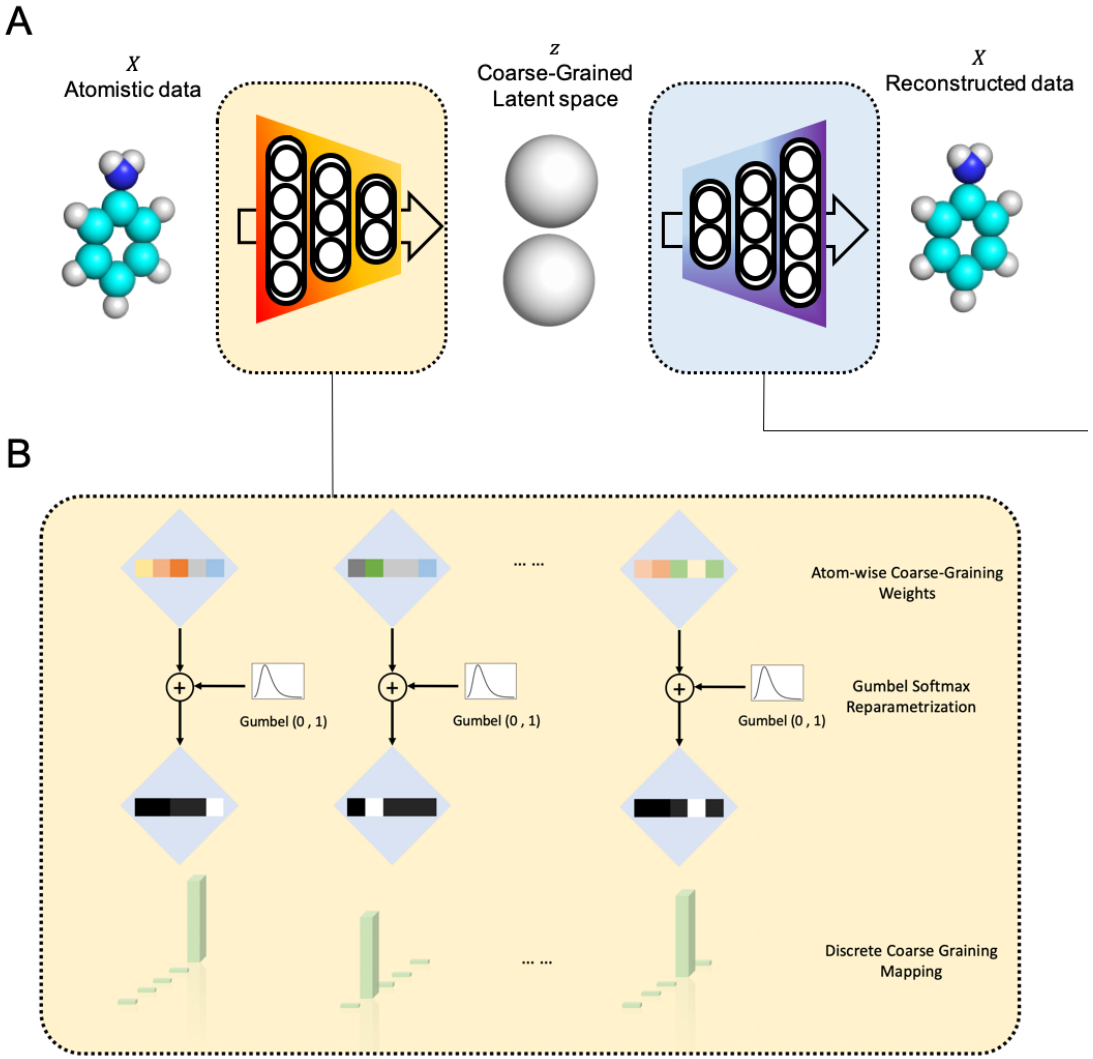
Decide degree of compression, functional form of V_{CG}

Train E and V_{CG} simultaneously

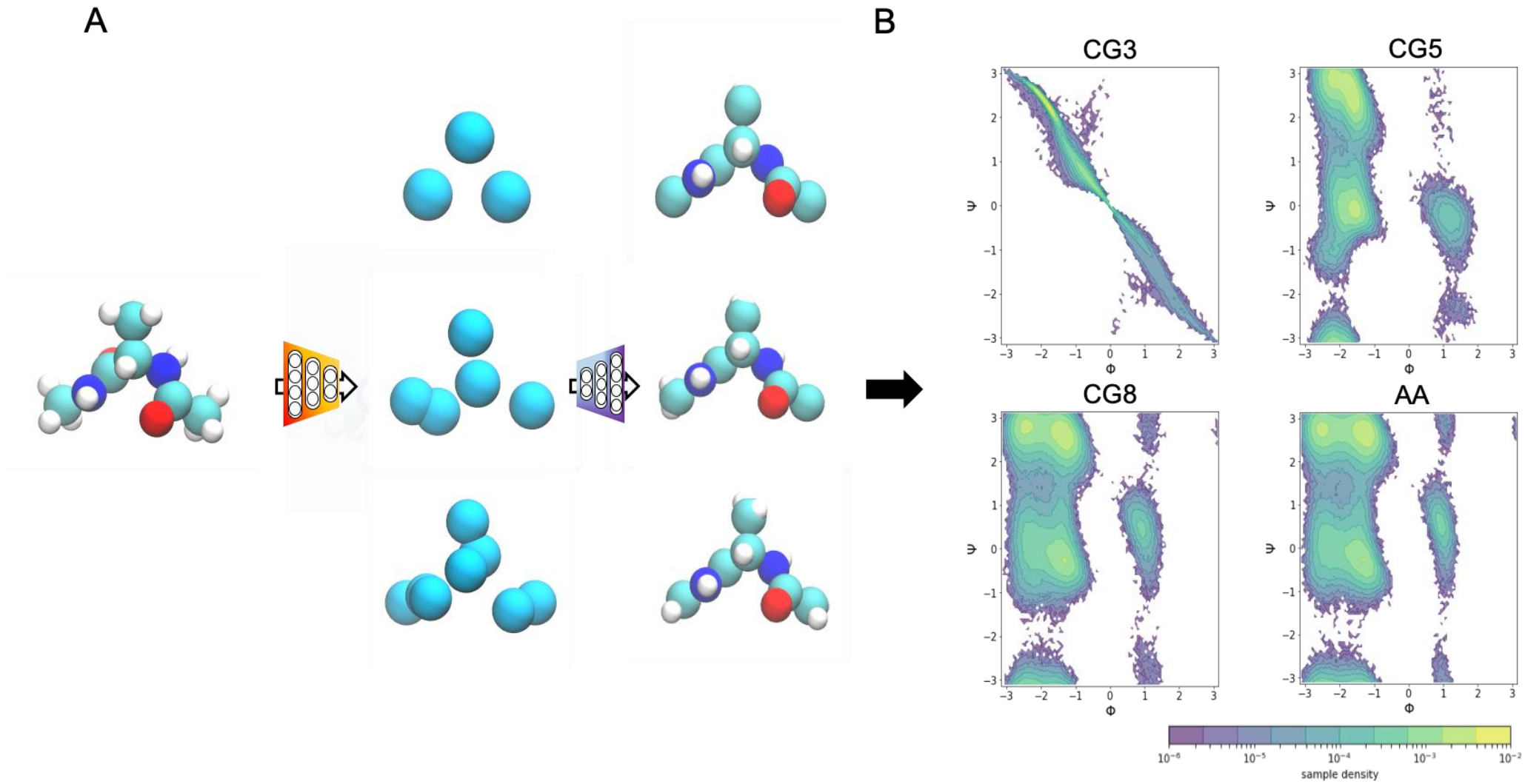
Deploy CG trajectory on larger system

Compare statistics

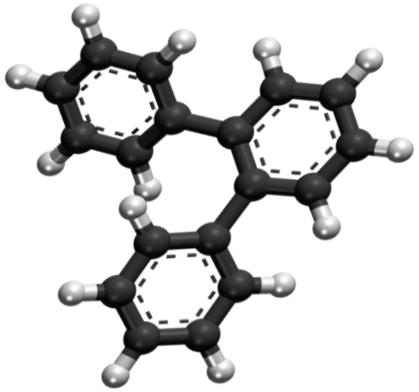
- Encoded AA vs. CG MD sampling

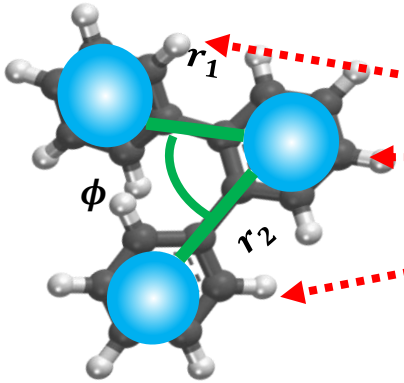


Alanine Dipeptide fully unsupervised



Automatic CG for small molecules

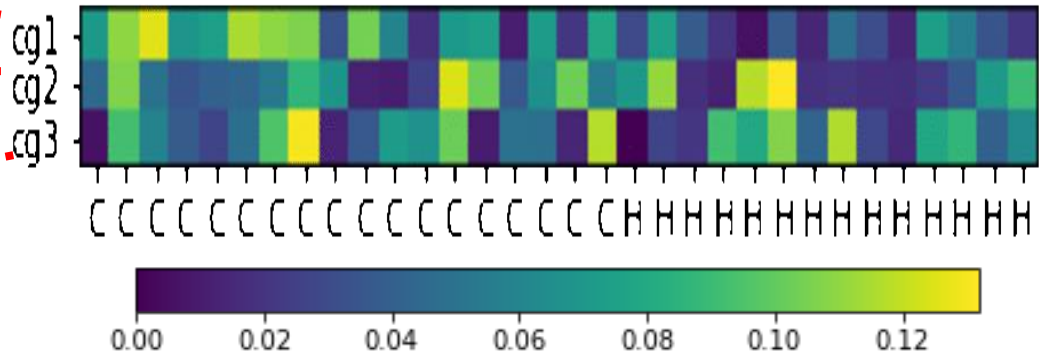


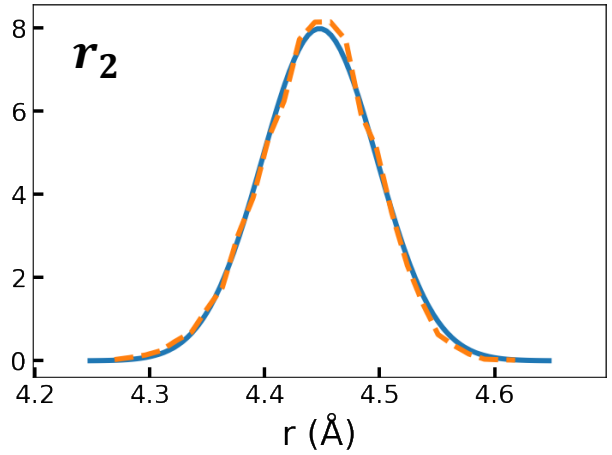


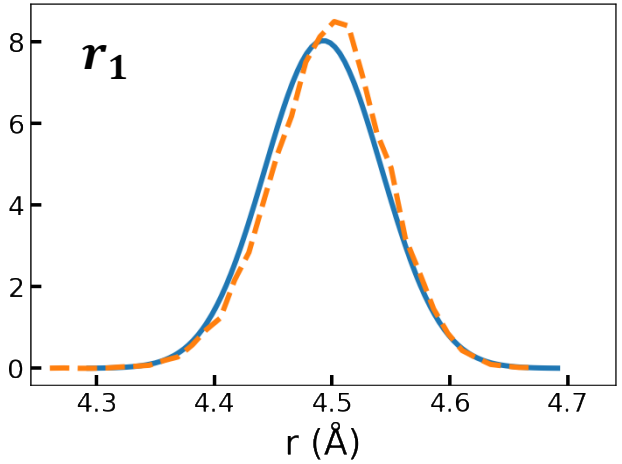
$V(x) = V_{bond}(\mathrm{x}) + \mathrm{V_{angle}}(\mathrm{x}) + \mathrm{V_{dihe}}(\mathrm{x}) + \mathrm{V_{pair}}(\mathrm{x}) + \mathrm{V_{coulomb}}(\mathrm{x})$

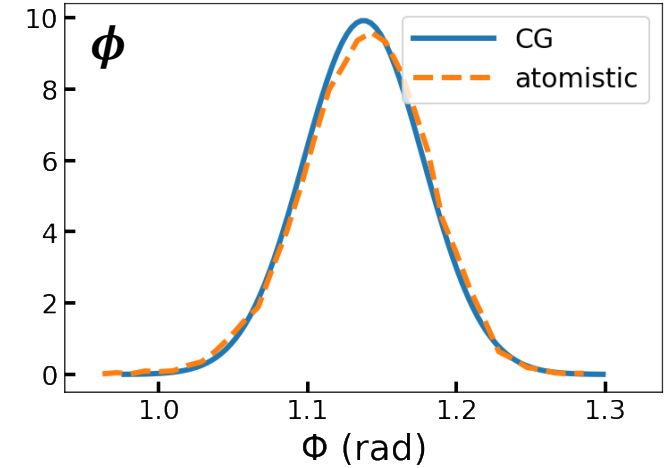
$V_{CG}(z) = V_{bond}(\mathrm{z}) + \mathrm{V_{angle}}(\mathrm{z})$

CG Encoder In Training

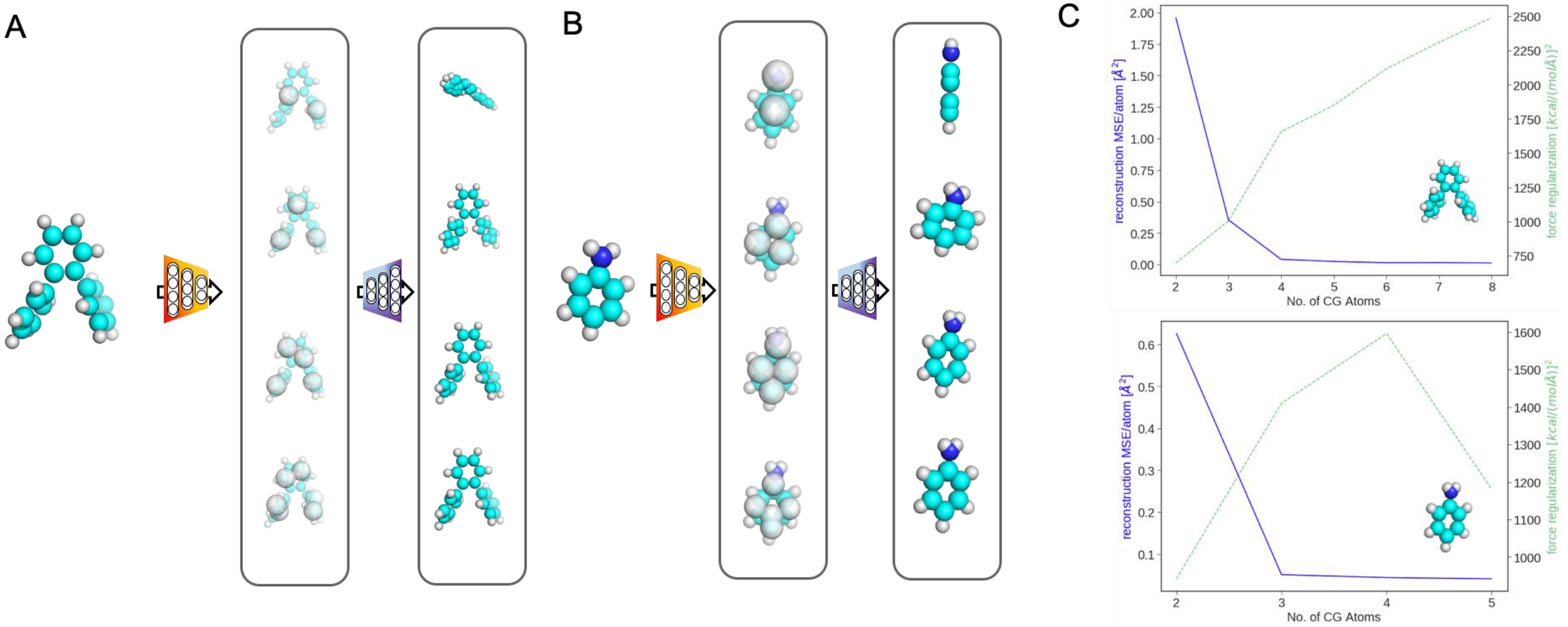




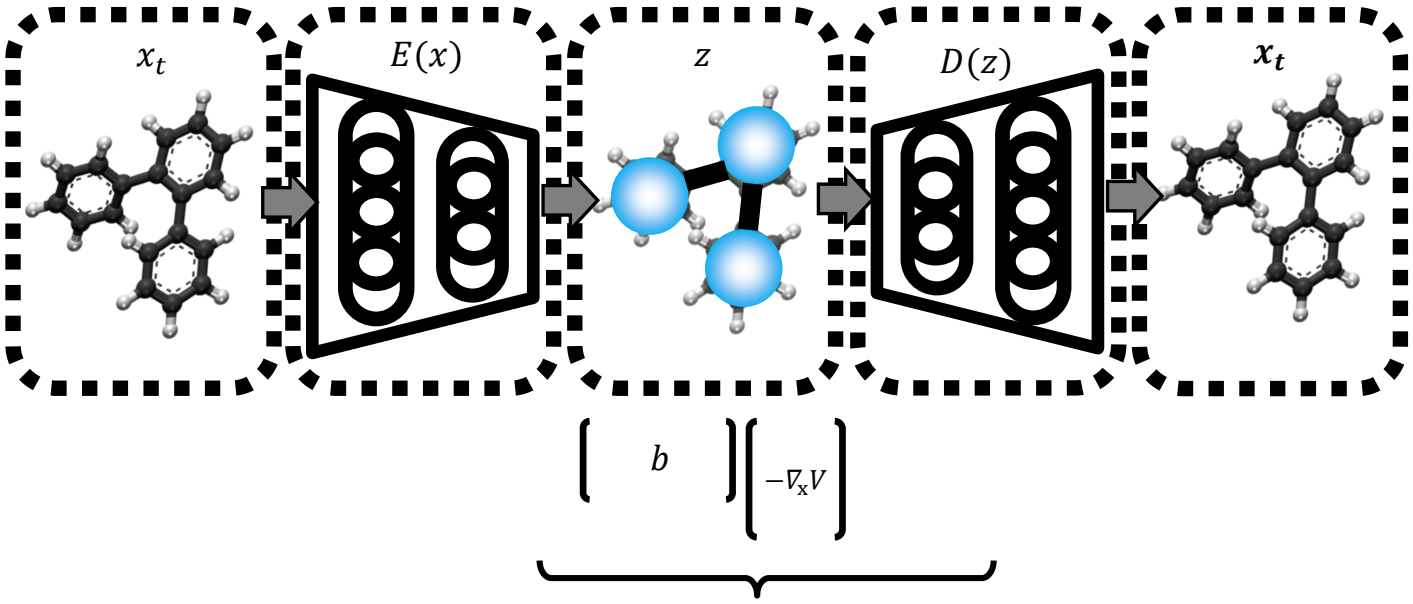




Compression vs. fluctuations



Coarse Graining Auto-Encoding Framework

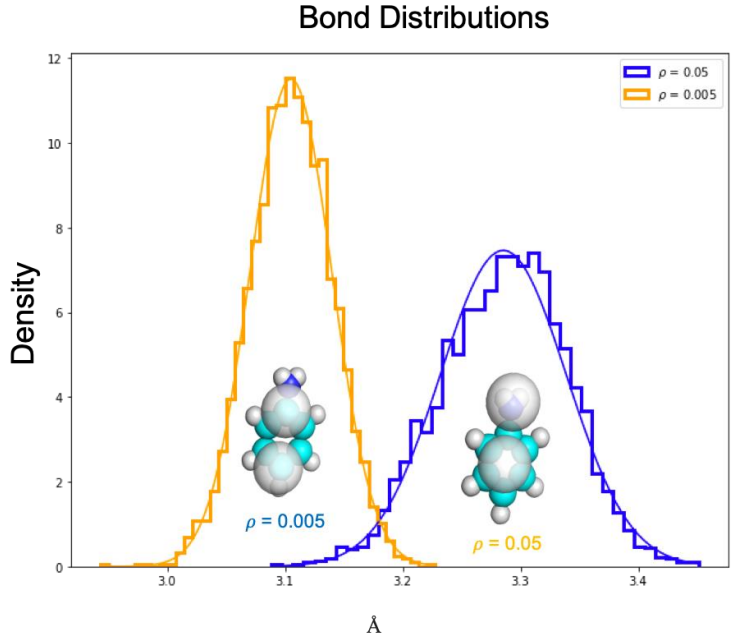


$$L_{VCGE} = |D\,E\,(x_t) - x_t|^2 + \rho F_{ins}\left(E(x)\right)^2$$

A pre-training step that is unsupervised (i.e. no V_{CG} is learnt) followed by adding supervision

The difference between F_{ins} and F is related to the friction, memory and fluctuations in the GLE.

We choose to add additional regularization to minimize $(F_{ins})^2$

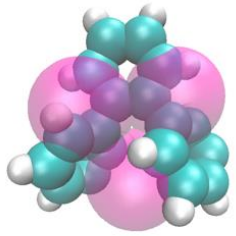


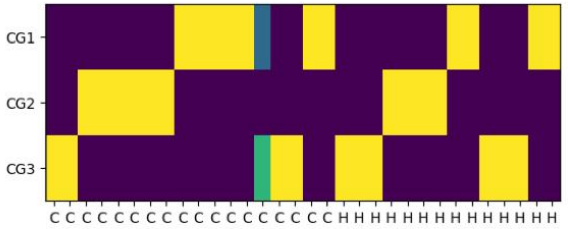
Pipeline

Split training

- into pre-train with force regularization and reconstruction
- fully supervised training of E and V_{CG}

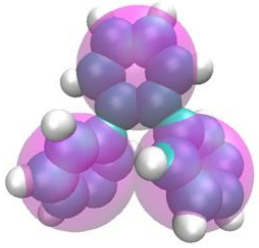
A

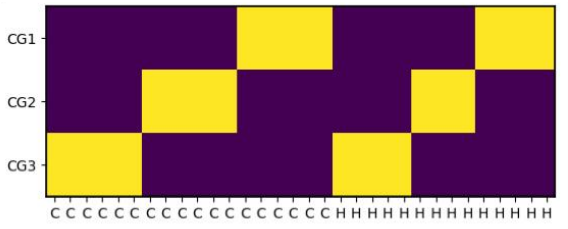




ρ	0.0
Force regularization	4498.0
Reconstruction	0.36

B

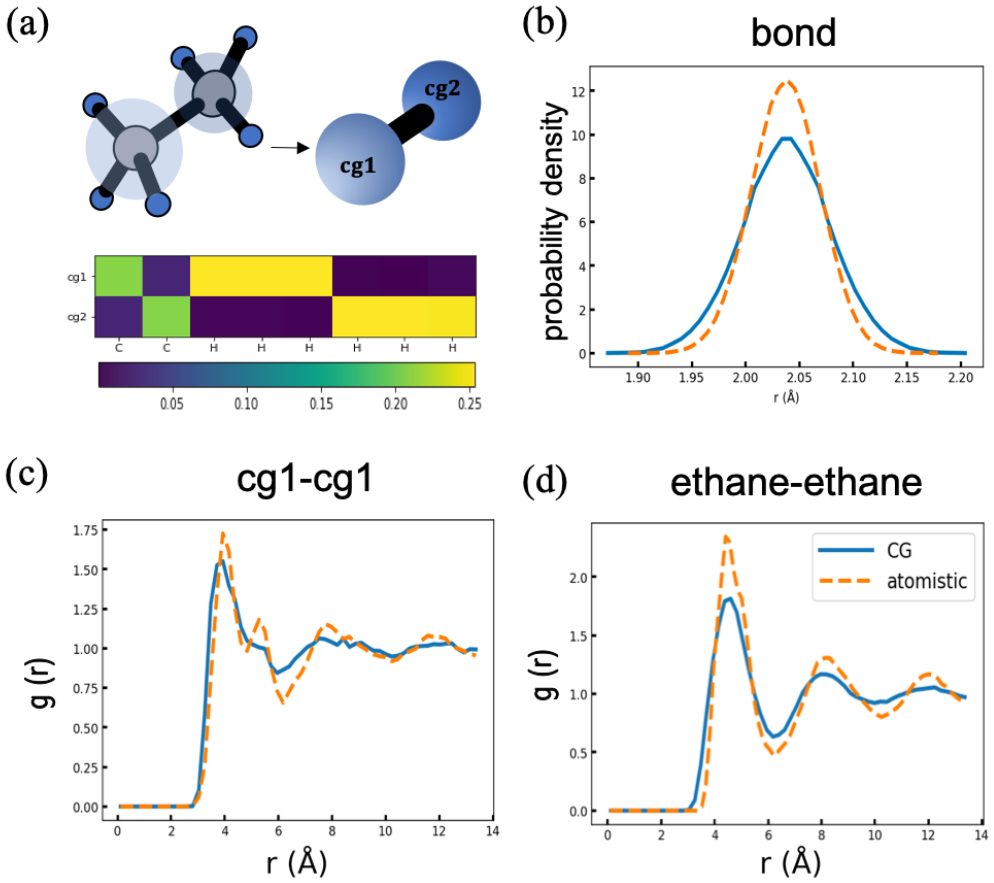




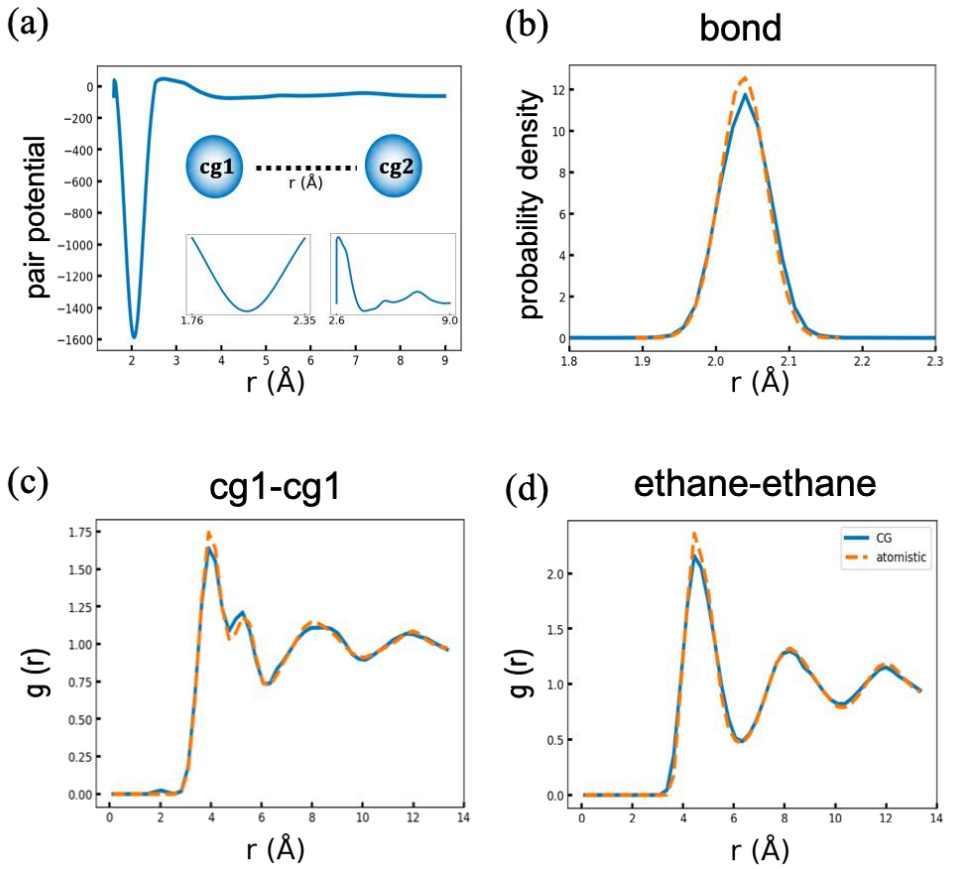
ρ	0.01
Force regularization	1008.0
Reconstruction	0.36

CG of liquid ethane

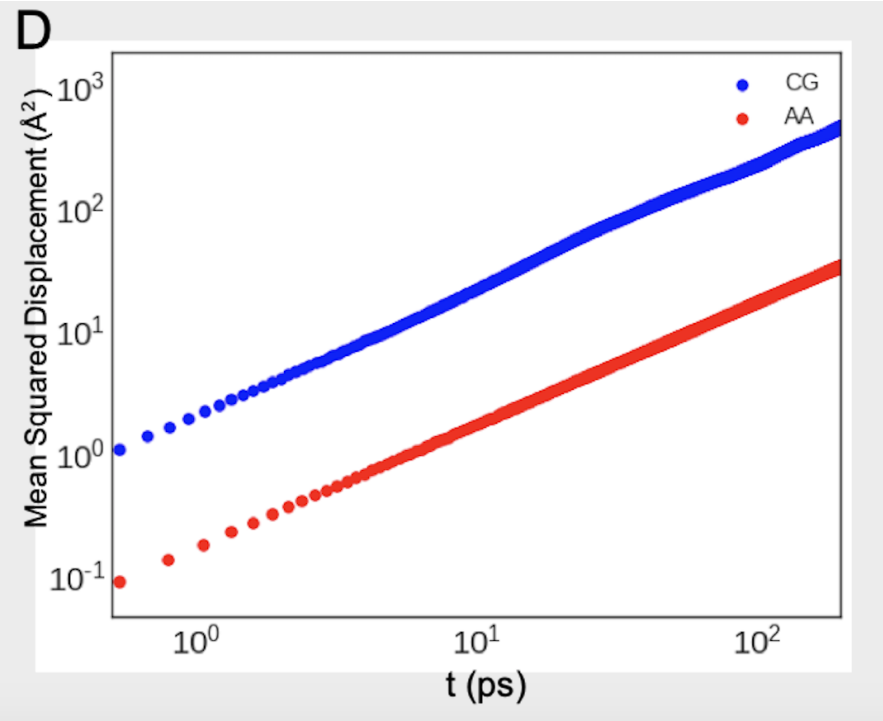
Classical potential



MPNN neural potential

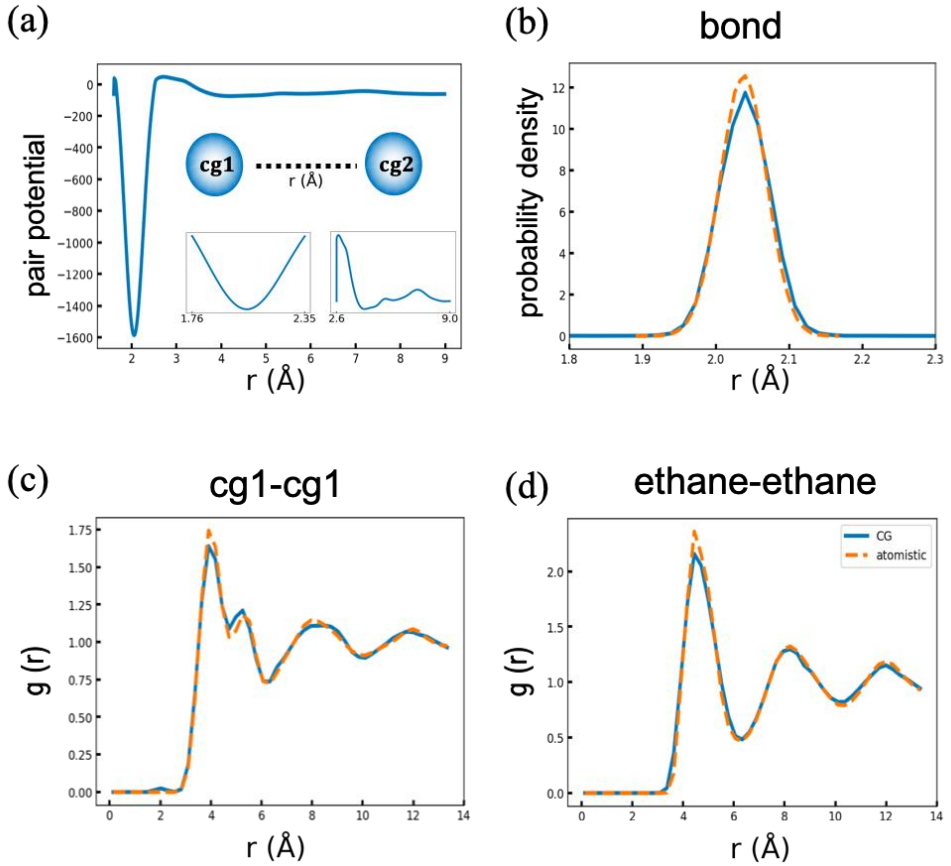


CG of liquid ethane



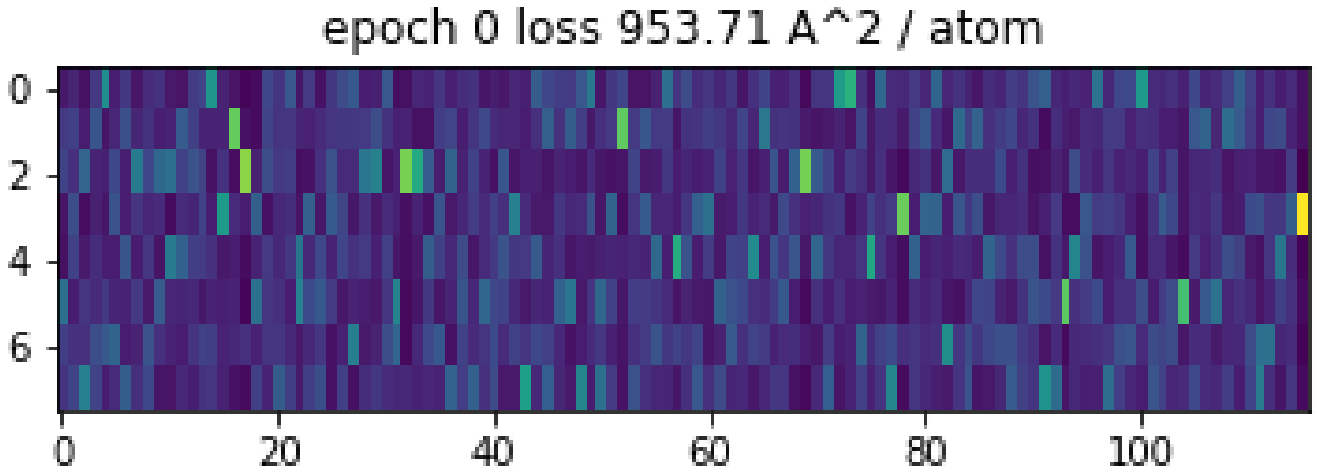
Accelerated diffusion

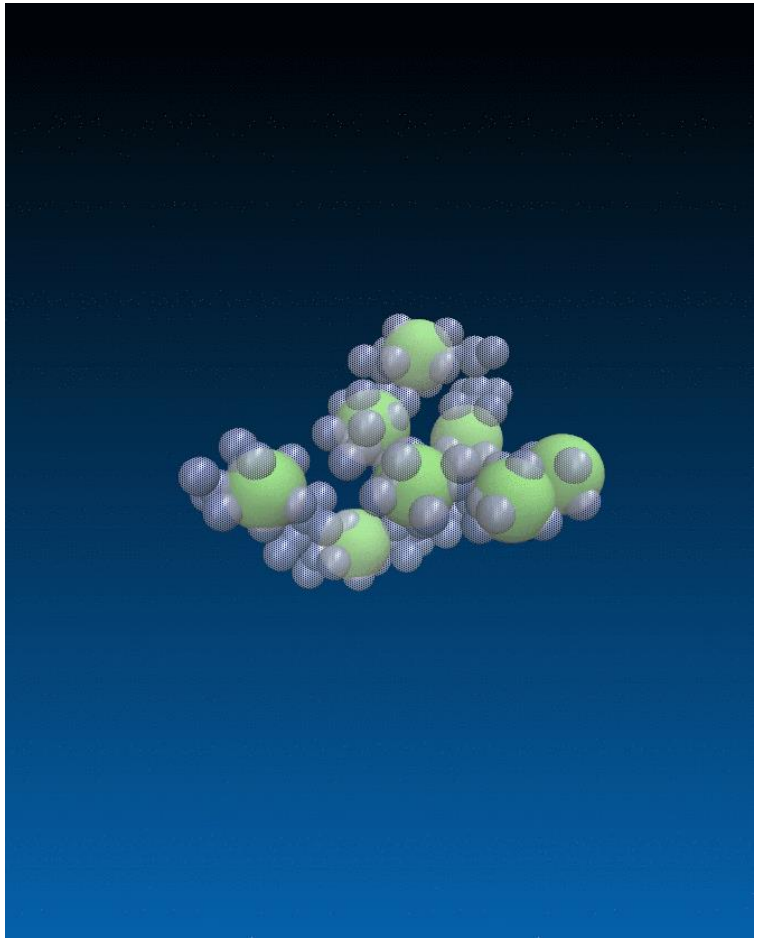
MPNN neural potential



CG of polyethylene

120 atoms of polyethylene



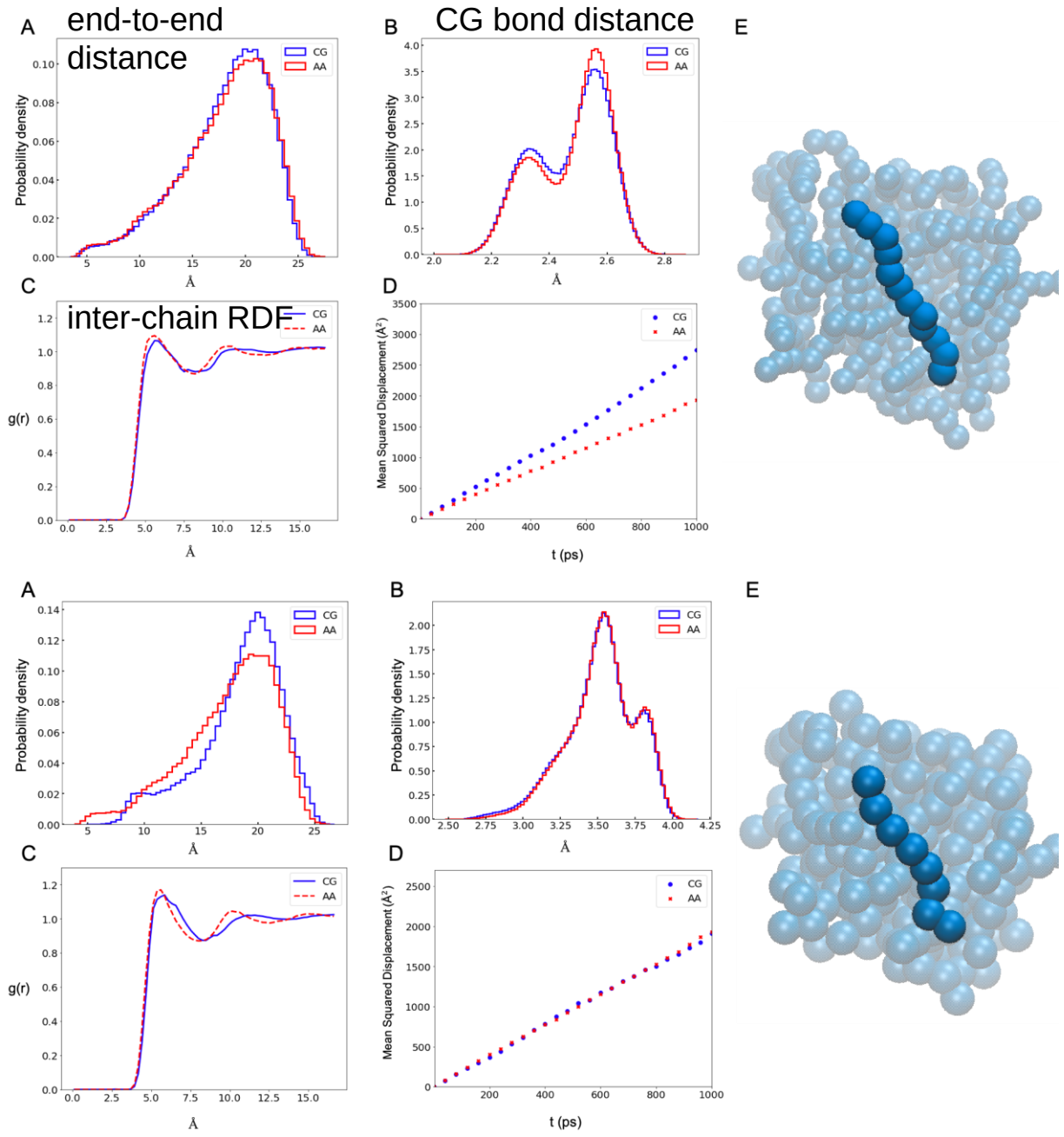


C24 in bulk

1:2 and 1:3 compression

Structure correlations are recovered correctly.

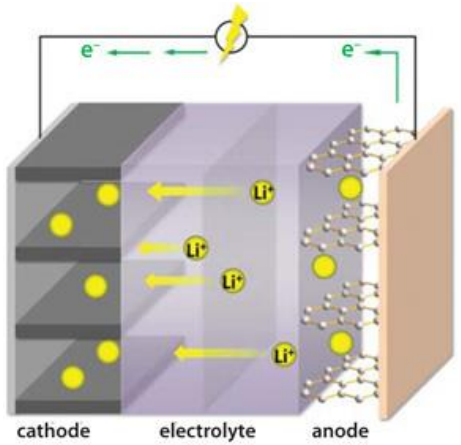
Diffusion kinetics are modified (fortuitous agreements)



Evolutionary Neural Potentials

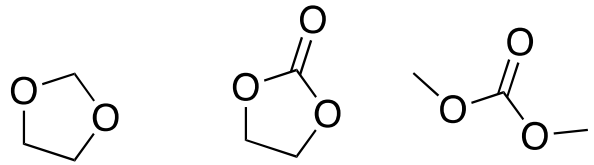
Lithium chelation

Lithium ion batteries require electrolytes to shuffle Li cations between electrodes.

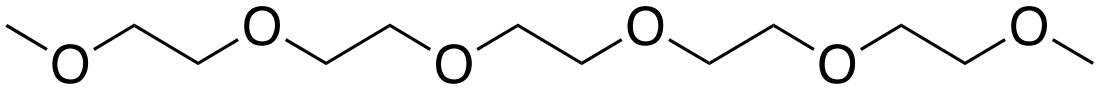


Organic electrolytes

- have a great design space
- Liquids have high conductivity can catch fire.



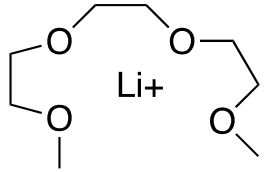
- Polymer solids are safer but lower conductivity. Polyethylene oxides



Explore the design space of lithium-binding moieties

Chemistries and configurations

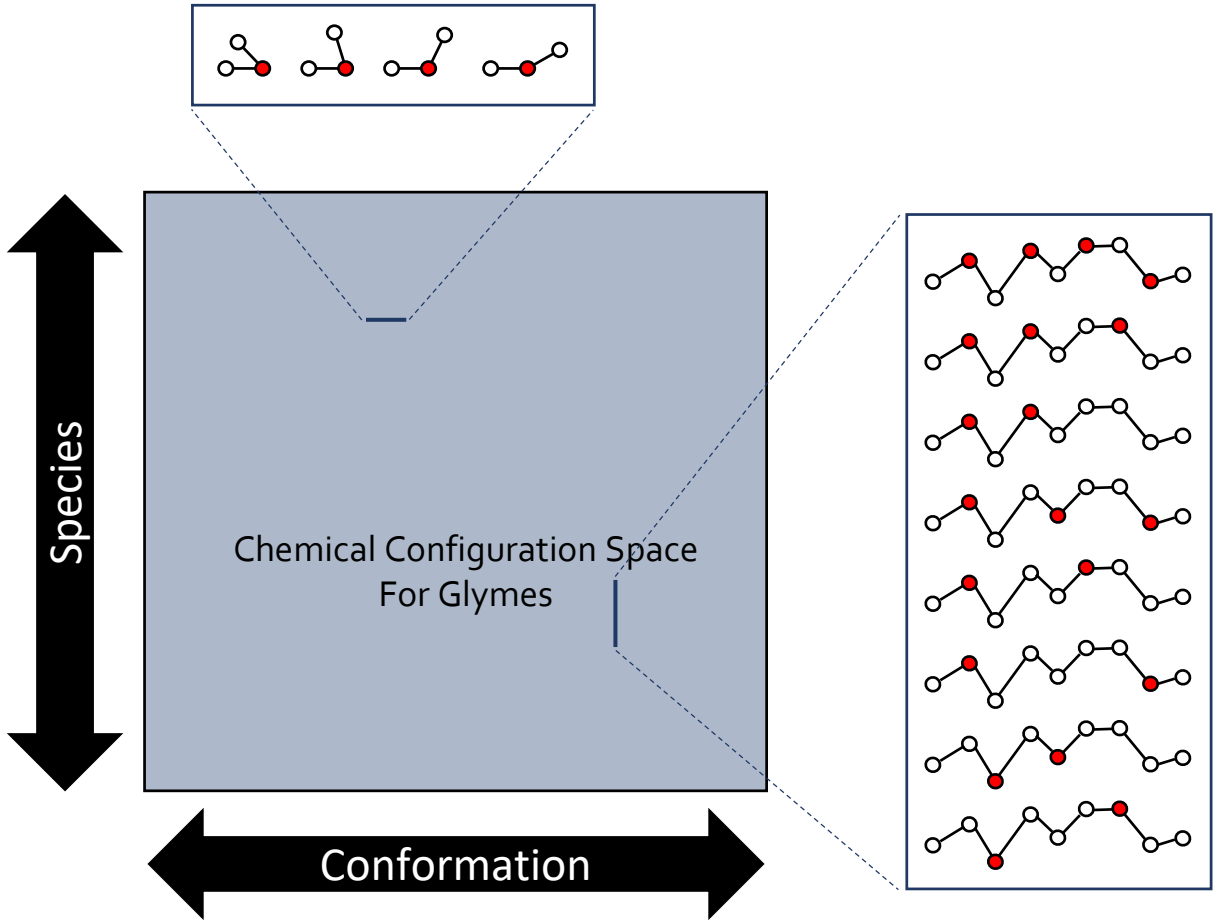
We are interested in Li-organic binding enthalpies (and free energies too, …)



Global energy minimization requires good sampling of supramolecular clusters.

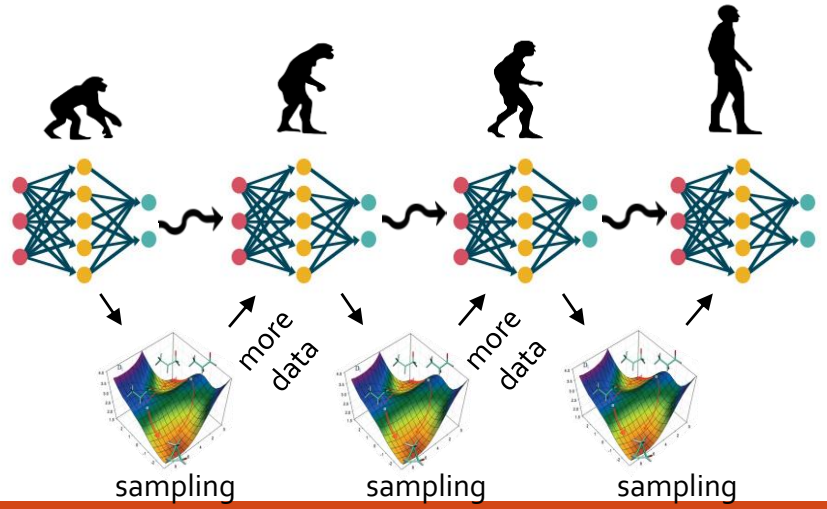
- A sampling problem!
- Can access with MD with some degree.

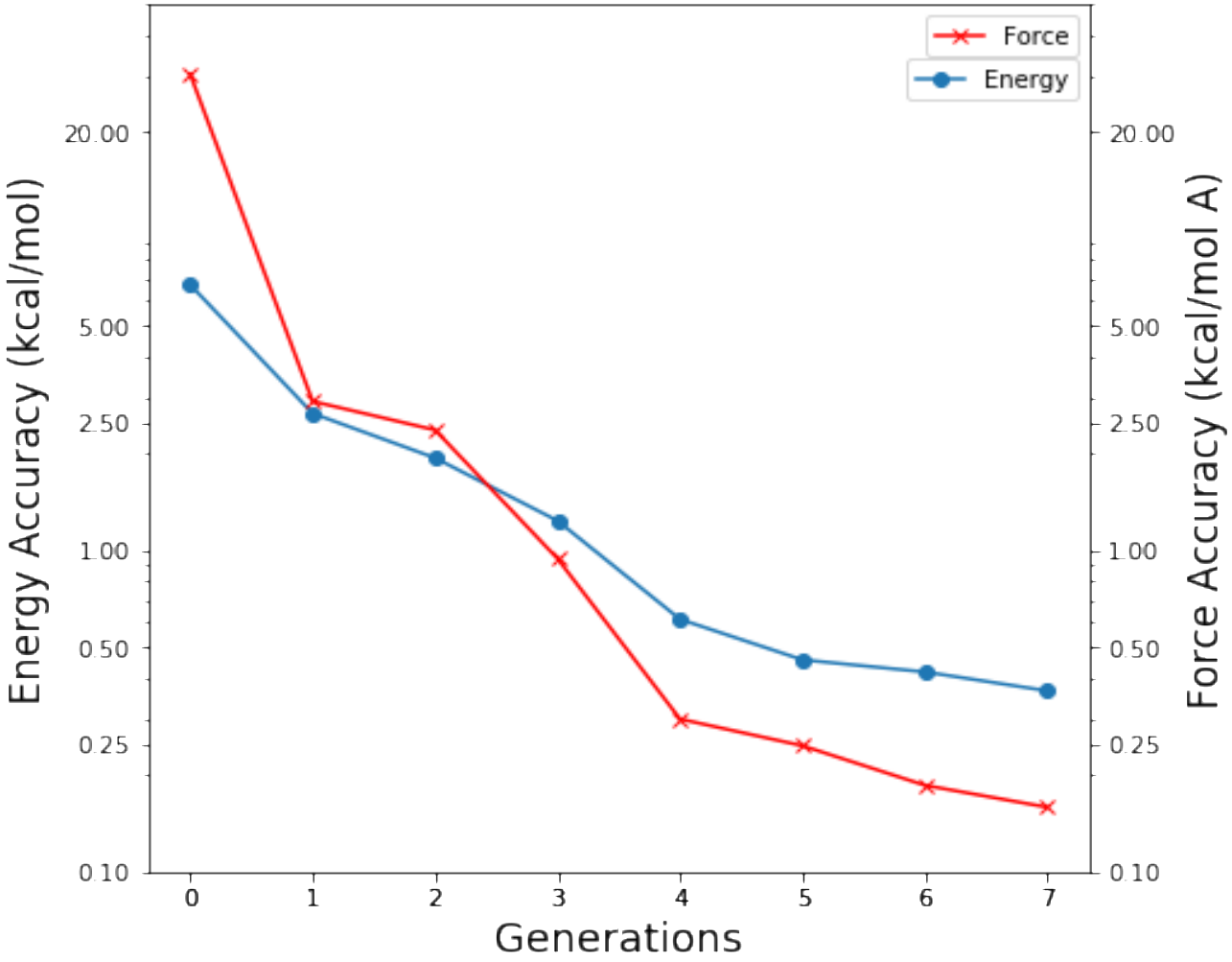
First-principles is too costly, and the chemistry is fairly constrained: create a potential.



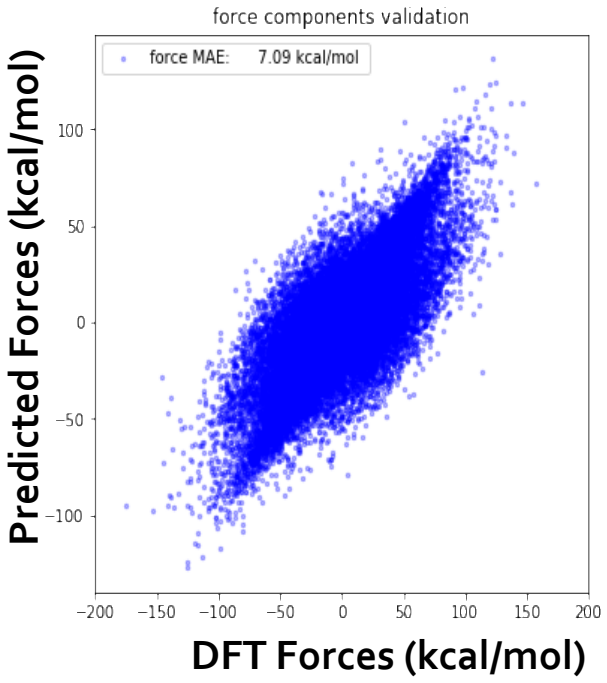
Evolutionary NNs

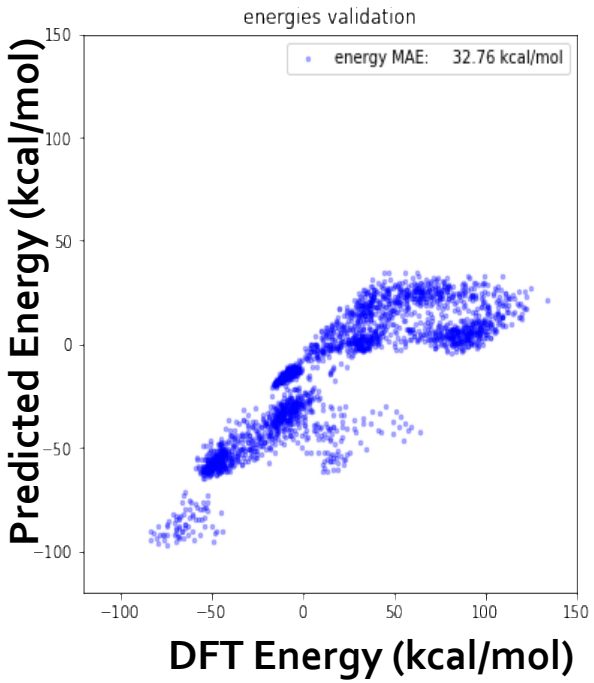
- Initialize with normal—mode data only and using neural network to sample new configurations via MD.
- Obtain DFT energies on neural network sampled data and validate on accuracy
- Train on <= 30000 points BP86/SVP-D3. (E and forces) across 14 chemistries with <= 20 heavy atoms, validated on 40 larger ones
- Decouple sampling for data acquisition.
- **Evolve** by incorporating neural network sampled configurations into the next generation of training data

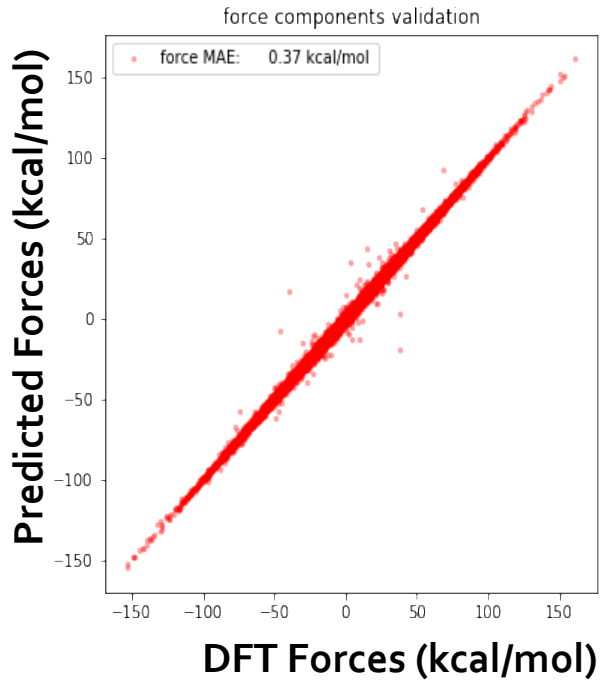


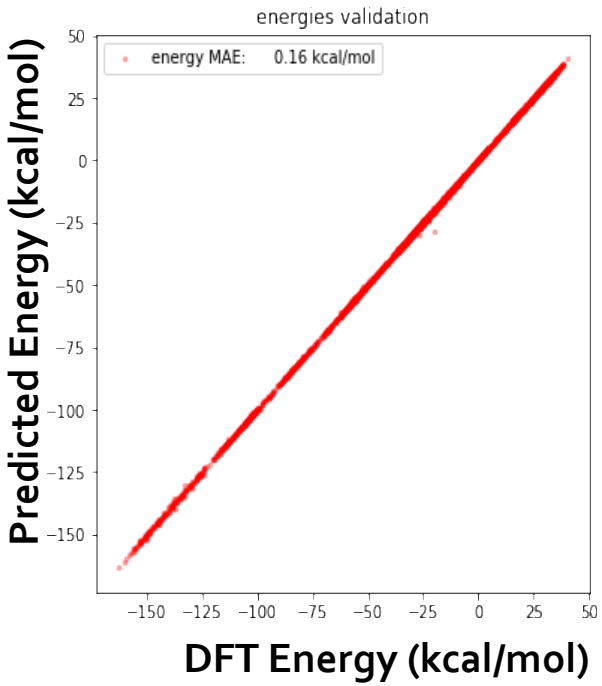


Evolutionary NNs



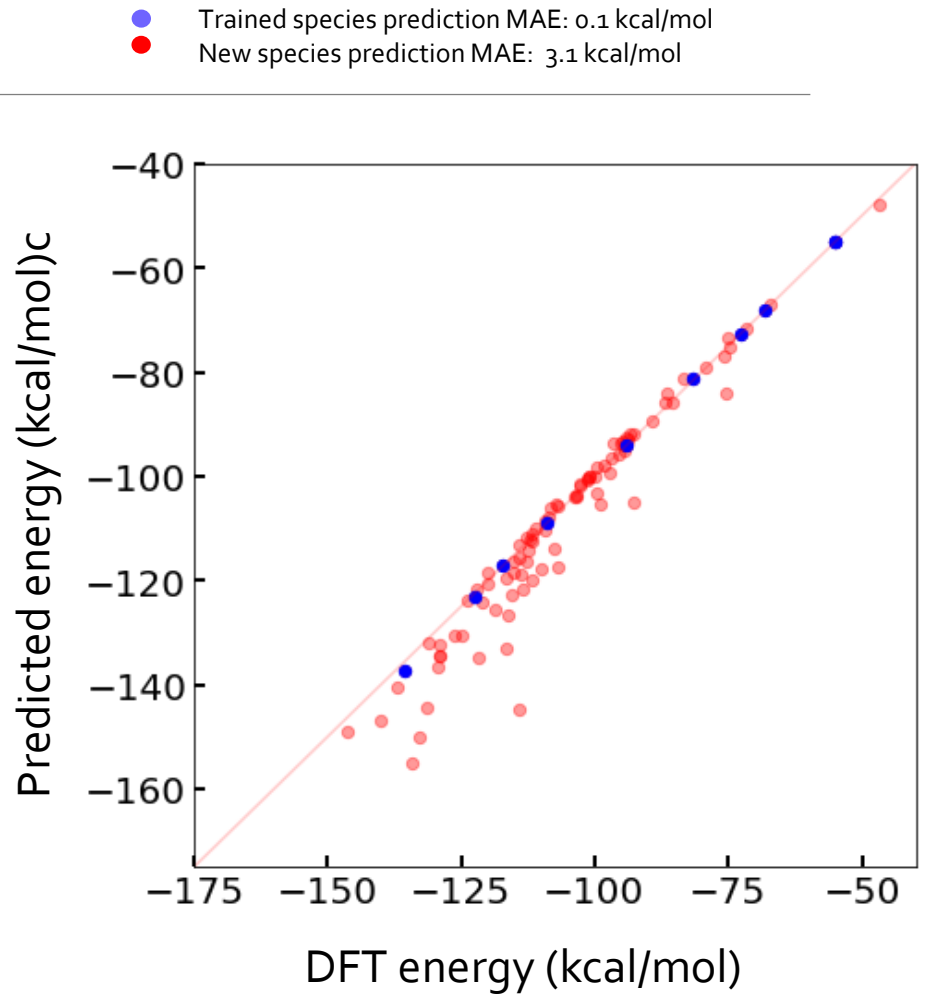






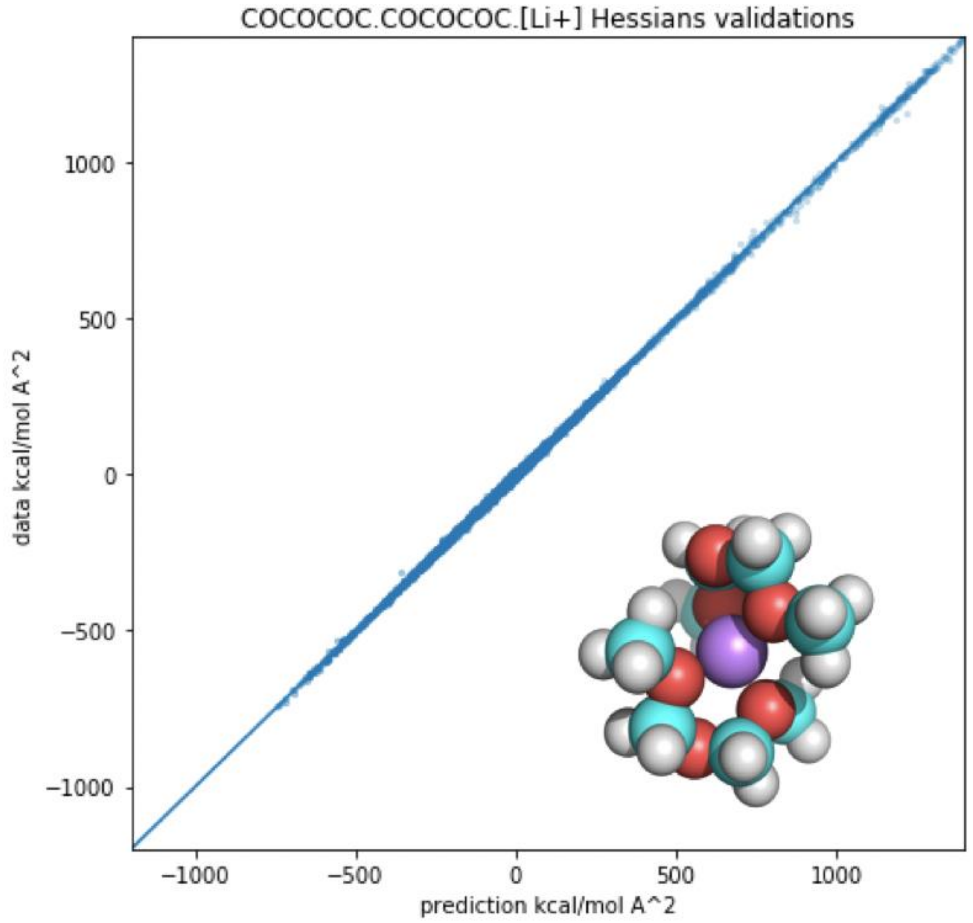
Validation

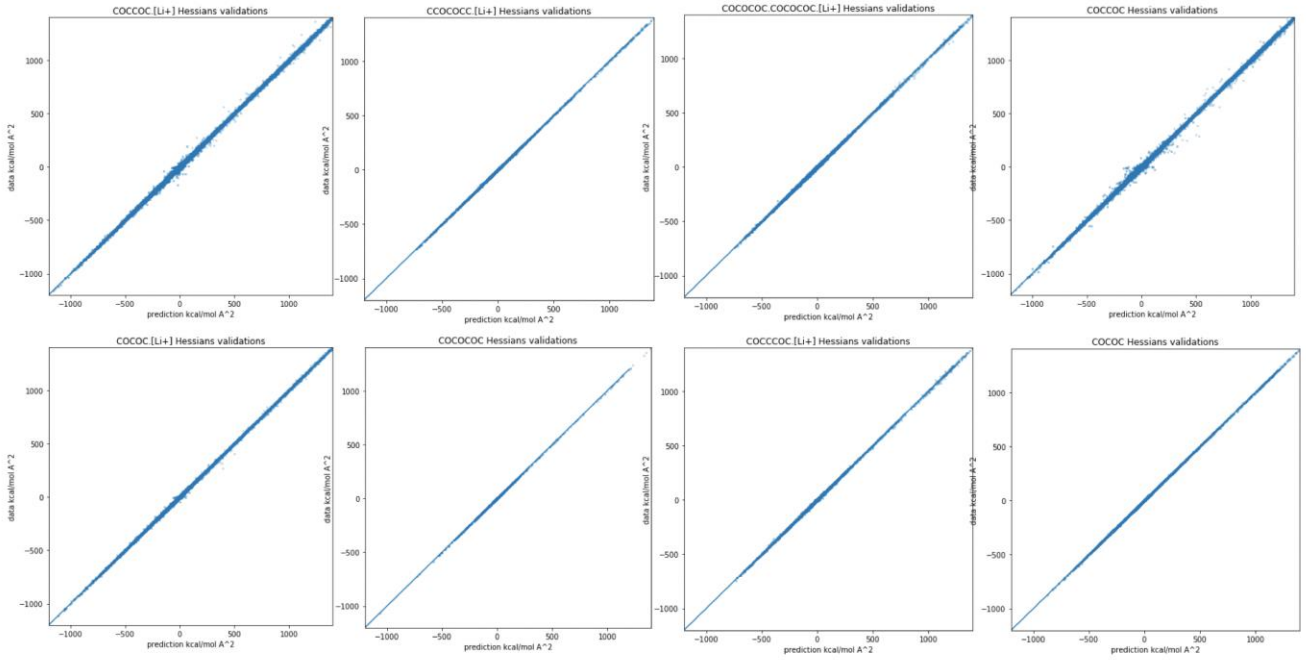
- Validated on seen and unseen chemicals, of up to 2x the size.
- It achieves good accuracy when validated on converged geometries on species that are not trained by the model -> **good transferability over ether chemical space.**



Error: 3.0 kcal/mol

Validating Hessians

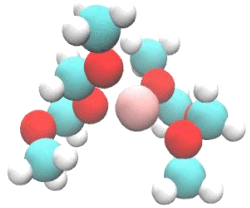




Exploring ether space

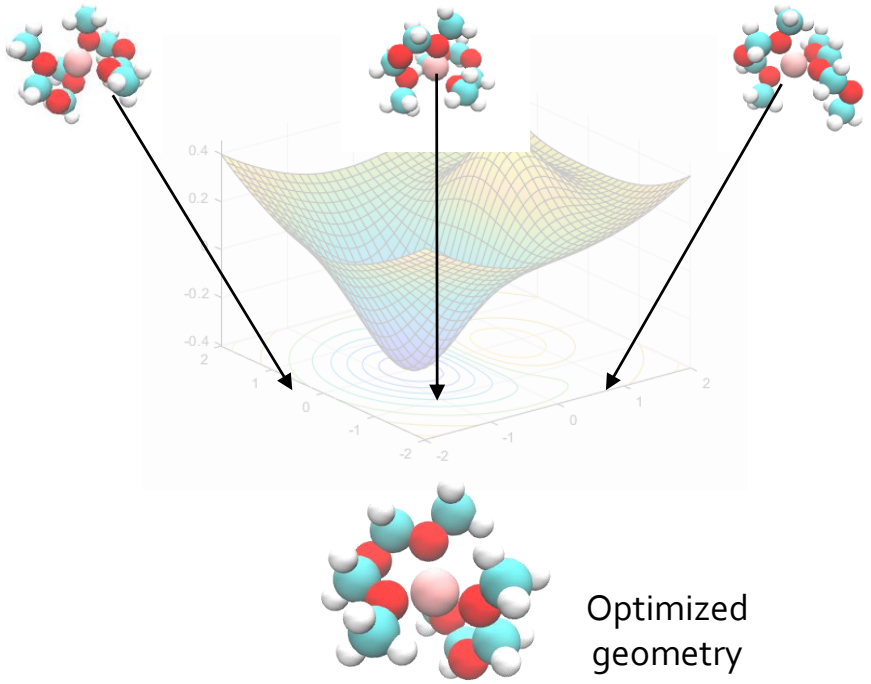
Predicting binding affinity:

- Run neural NVE MD
- Gradient descent at some frames

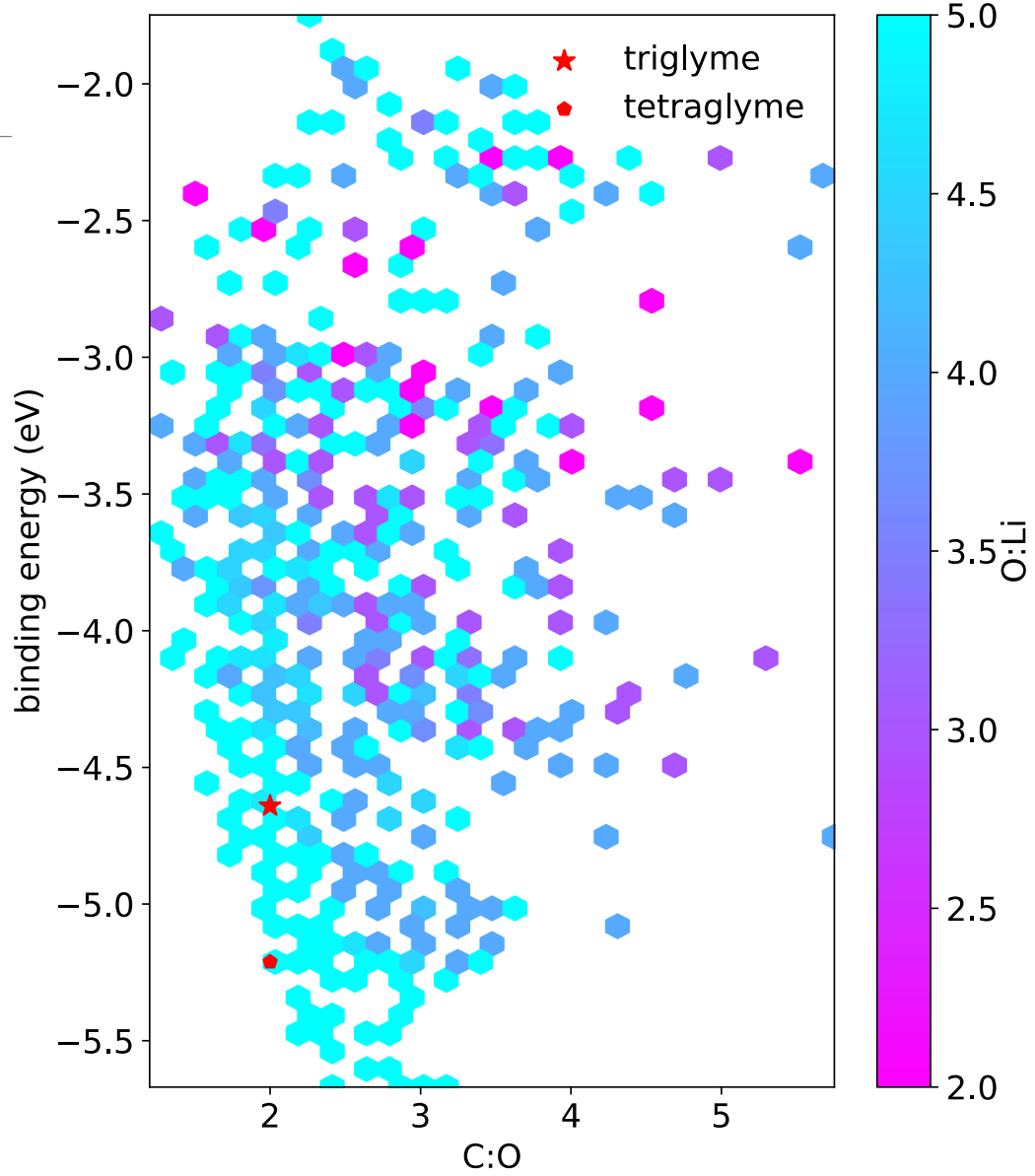


<2ms second per energy/force calculations

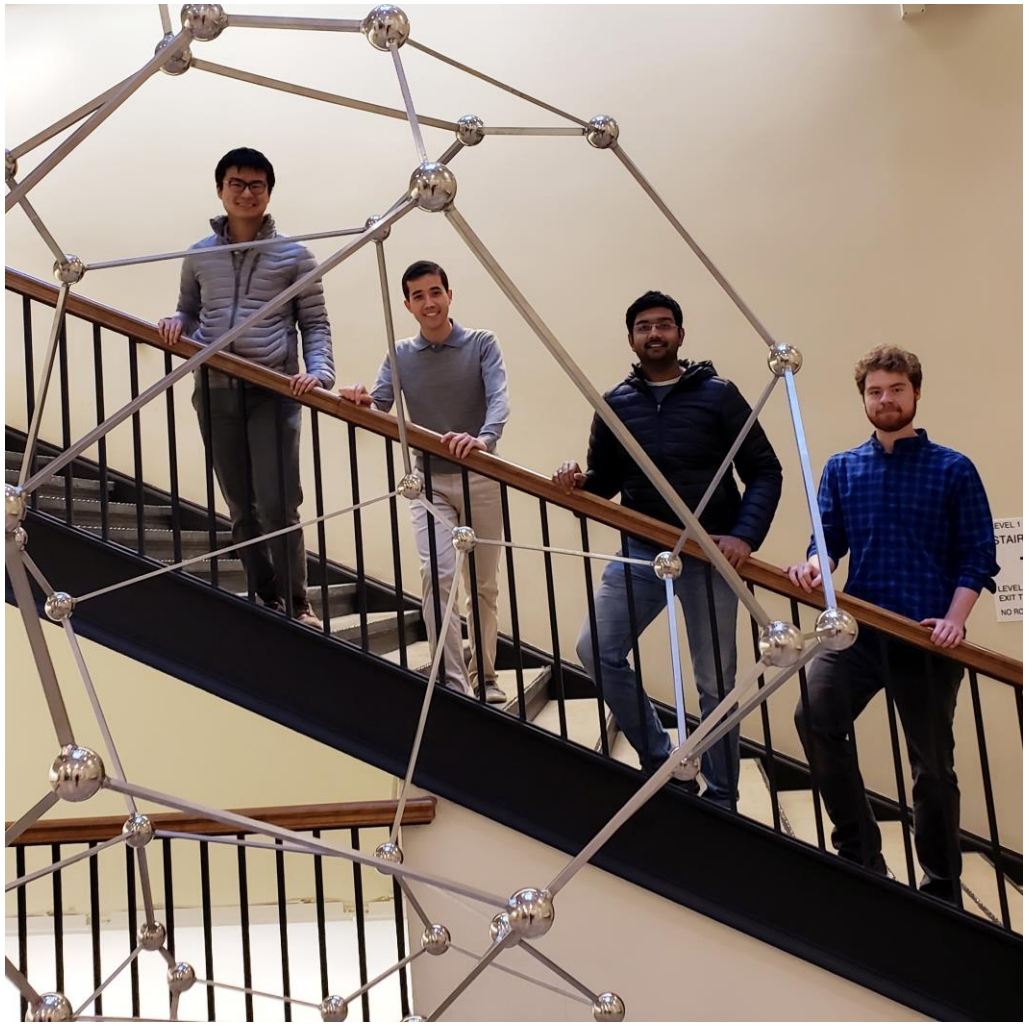
O(N^2) and embarrassingly parallel



Optimized geometry



Thanks!



Wujie Wang (CG, Evolving NN)
Wil Harris (Graph NN)
Daniel Schwalbe Koda
Somesh Mohapatra
James Damewood
Shi Jun Ang