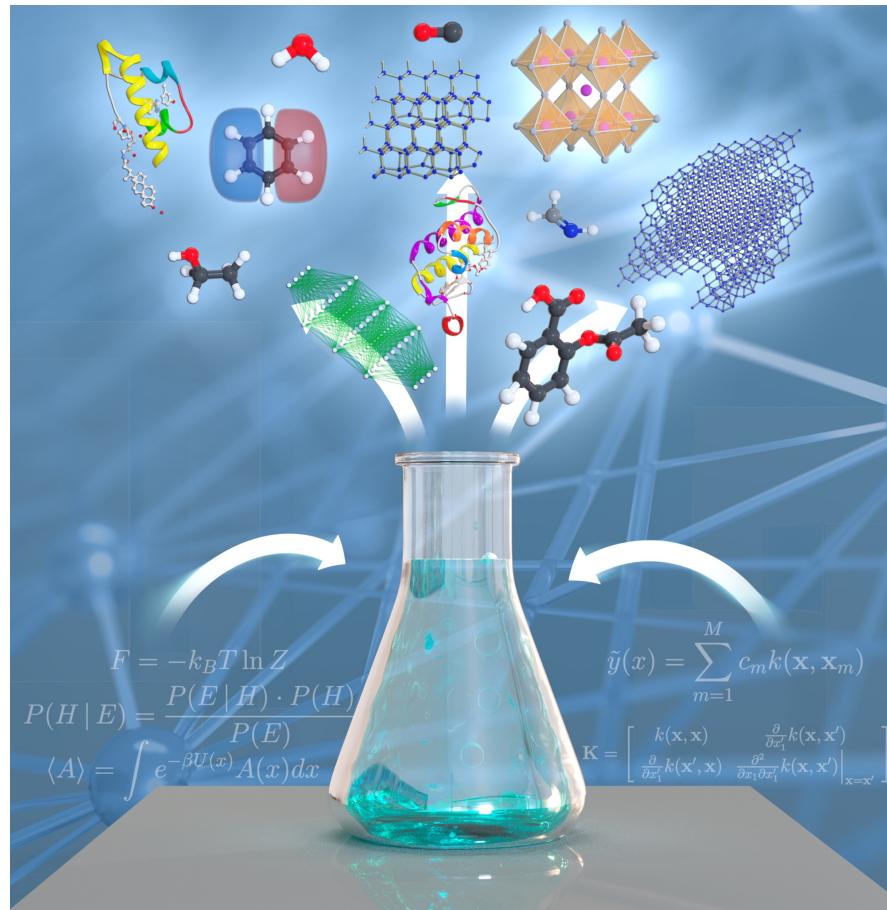


Machine Learning for Molecular Physics

Cecilia Clementi

Freie Universität Berlin
cecilia.clementi@fu-berlin.de

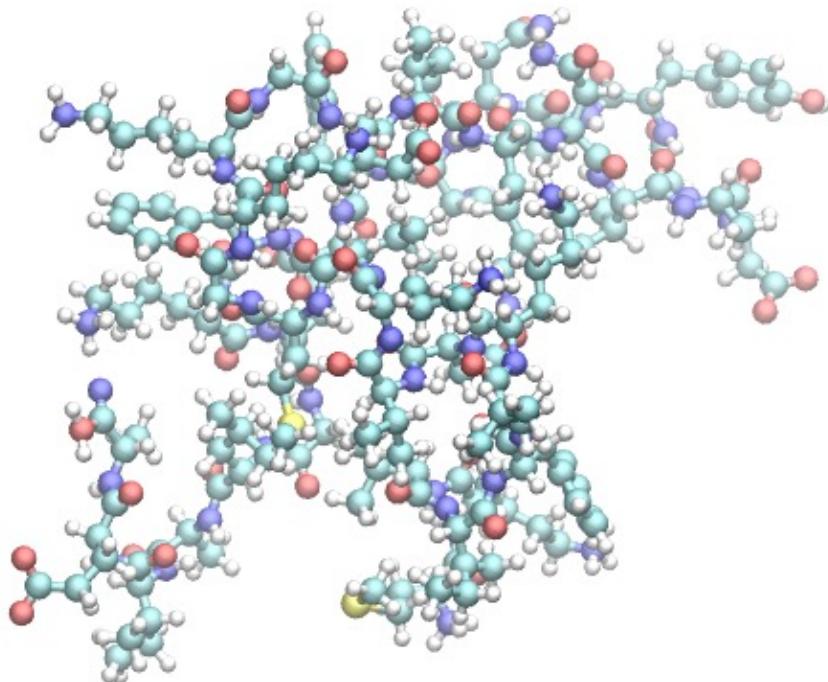


How we have been doing Molecular Simulation before ML

1. Force fields

$$\mathcal{H}\psi_{\mathbf{x}}(\mathbf{r}, \mathbf{s}) = U(\mathbf{x})\psi_{\mathbf{x}}(\mathbf{r}, \mathbf{s})$$

Born-Oppenheimer approximation



Use (approximate)
quantum mechanics
to calculate energy
in specific configurations
and then fit to
classical potentials

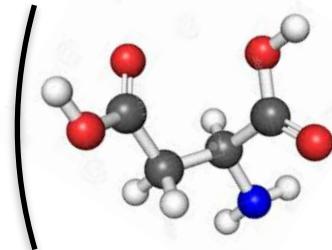
Force fields How can we do it with ML

$$\hat{H}\psi_R(\mathbf{r}_1, \dots, \mathbf{r}_N) = E_R\psi_R(\mathbf{r}_1, \dots, \mathbf{r}_N)$$

1. Highly accurate solution
for small molecules

Quantum Chemistry
(DFT, Coupled-cluster, ...)

Database



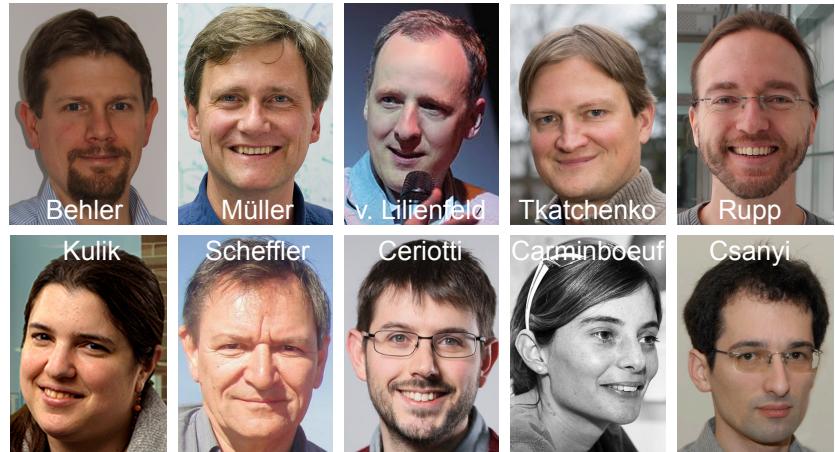
Chemical Elements
Atom coordinates \mathbf{R}
Energy
Forces

$E(\mathbf{R})$
 $-dE(\mathbf{R}) / d\mathbf{R}_i$

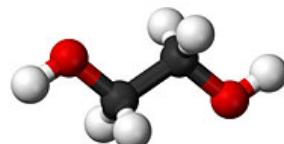
, . . .

2. Predict for new molecule
(Elements and atom coordinates \mathbf{R})

Machine learning (supervised)

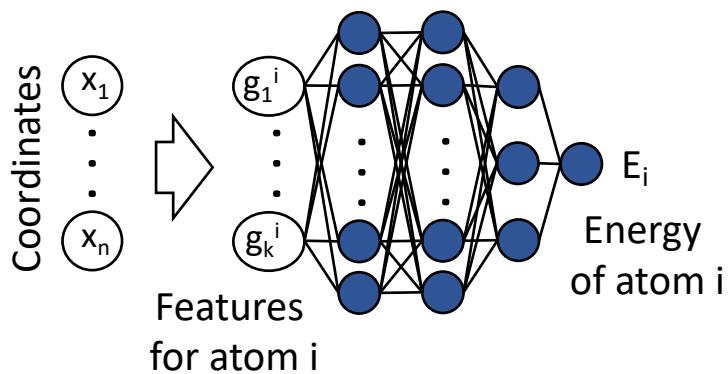


Energy $E(\mathbf{R})$
Forces $-dE(\mathbf{R}) / d\mathbf{R}_i$



Force fields How can we do it with ML

a)

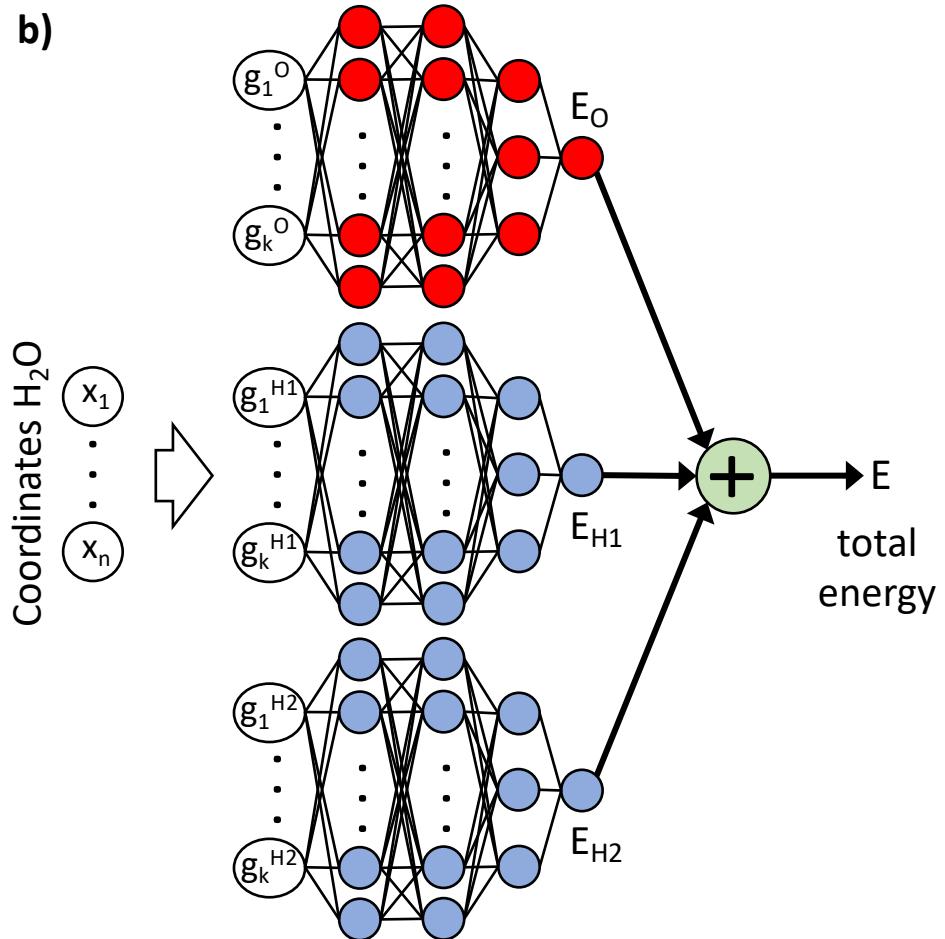


Behler-Parrinello
networks for
quantum mechanical
energies

$$L_{\text{ene}} = \sum (\hat{U}(\mathbf{x}_i, \theta) - U_i)^2$$

$$L_{\text{force}} = \sum_i \left\| \nabla \hat{U}(\mathbf{x}_i, \theta) + \mathbf{f}_i \right\|^2$$

b)



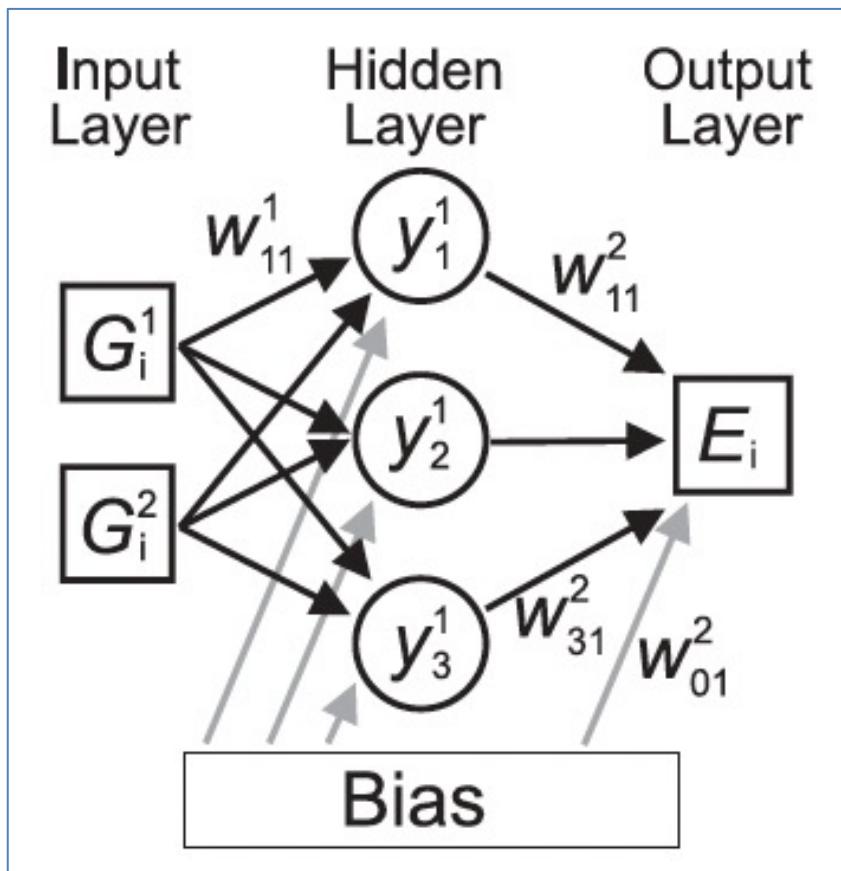
Behler & Parrinello, *Phys. Rev. Lett.* 98:146401 (2007)
Smith, et al. *Chem. Sci.*, 8:3192-3203 (2017)
Smith, et al. *chemrxiv.6744440.v1* (2018)
Han et al. *Phys. Rev. Lett.*, 120:14300 (2018)

Behler-Parrinello network

Original paper: **PRL 98, 146401 (2007)**

[...] Generalized neural-network (NN) method for constructing DFT-based PESs which have ab initio accuracy and are capable of describing all types of bonding.”

“The resulting many-body potentials are a function of all atomic coordinates and can be used in systems of arbitrary size.”



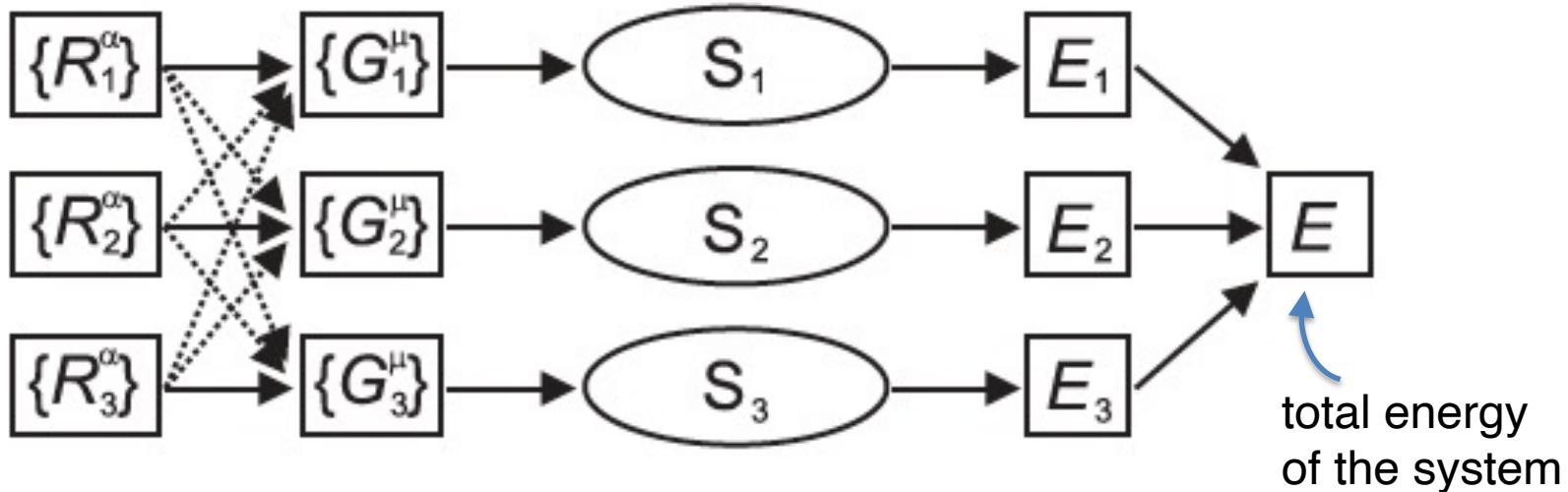
PES for bulk **Silicon**, valid across the phase diagram

$$E_i = f_a^2 \left[w_{01}^2 + \sum_{j=1}^3 w_{j1}^2 f_a^1 \left(w_{0j}^1 + \sum_{\mu=1}^2 w_{\mu j}^1 G_i^\mu \right) \right]$$

activation function is hyperbolic tangent

Behler-Parrinello network

specific network architecture



R_i^α are the Cartesian coordinates of atom i

G_i^μ are a set of μ generalized coordinates, called "**symmetry functions**", describing the local geometric environment of atom i

S_i is the subnetwork of atom i

E_i is the energy of atom i

Behler-Parrinello network

symmetry functions

Two different kinds of symmetry functions:

1. Radial symmetry functions

$$G_i^1 = \sum_{j \neq i}^{\text{all}} e^{-\eta(R_{ij} - R_s)^2} f_c(R_{ij})$$

R_{ij} distance between atoms i and j

η, R_s hyperparameters

$$f_c(R_{ij}) = \begin{cases} 0.5 \times \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1 \right] & \text{for } R_{ij} \leq R_c \\ 0 & \text{for } R_{ij} > R_c \end{cases} \quad \text{cutoff function}$$

Behler-Parrinello network

symmetry functions

Two different kinds of symmetry functions:

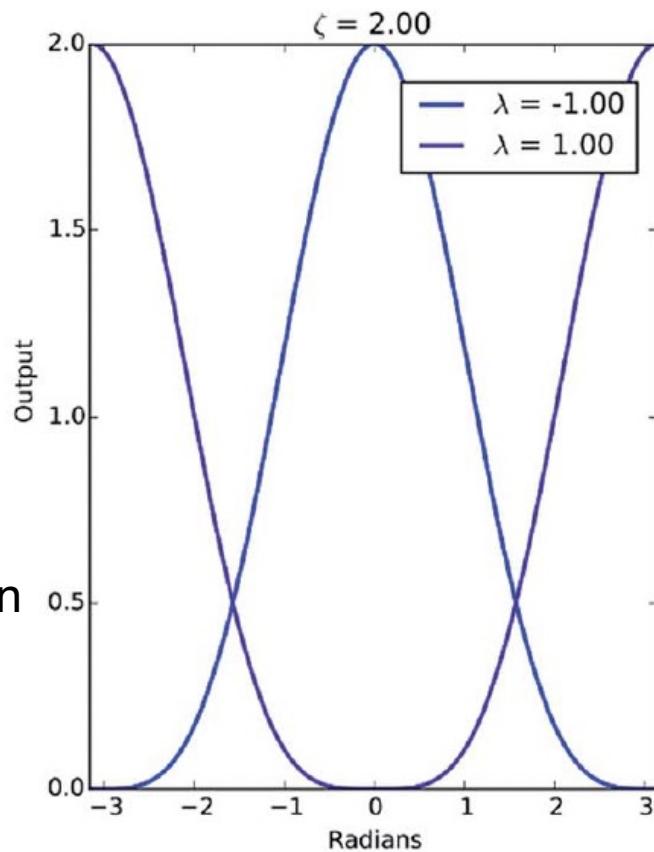
2. Angular symmetry functions

$$G_i^2 = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{all}} (1 + \lambda \cos \theta_{ijk})^\zeta \times e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk})$$

R_{ij} distance between atoms i and j

$$f_c(R_{ij}) = \begin{cases} 0.5 \times \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1 \right] & \text{for } R_{ij} \leq R_c \text{ cutoff} \\ 0 & \text{for } R_{ij} > R_c \end{cases}$$

$\eta, \zeta, \lambda (= \pm 1)$ are hyperparameters



Behler-Parrinello network implementation

- trained on DFT calculations on system of 64 Silicon atoms
- dataset contains 9000 structures in different phases (different P, T)
(8200 used for optimization and 800 for testing)
- subnets with 2 hidden layers, 40 nodes each
- 48 symmetry functions, with different values of the hyperparameters
- few thousands fitting parameters (network weights)

Behler-Parrinello network

results

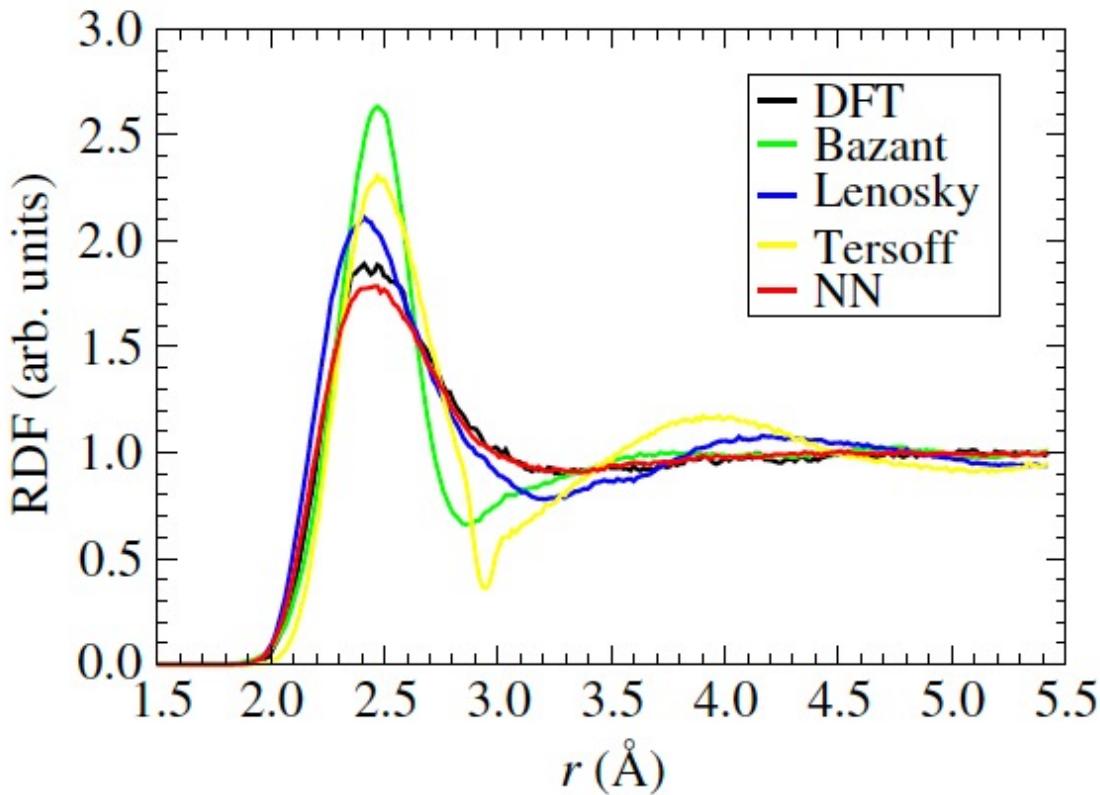


FIG. 3 (color online). Radial distribution function (RDF) of a silicon melt at 3000 K as obtained using a cubic 64 atom cell ($a = 20.526$ bohr). The curves shown were obtained from the Bazant [17,19], the Lenosky [15,19], the Tersoff [16,20], a neural network (NN) potential, and from density-functional theory (DFT) [18].

Behler-Parrinello network results

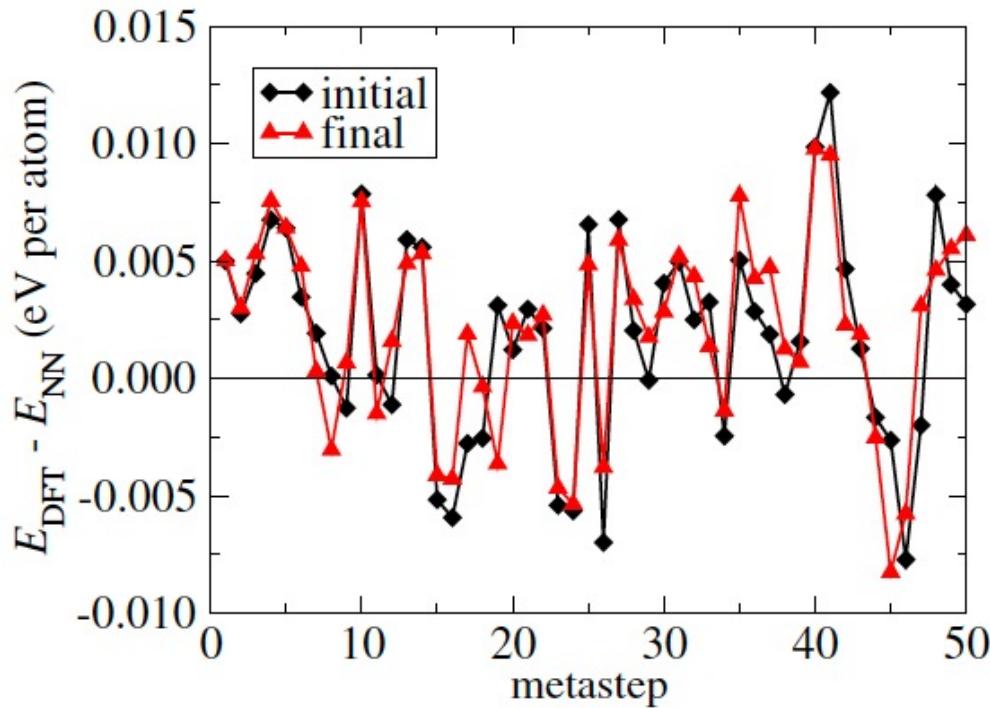


FIG. 4 (color online). Difference between the energies predicted by the neural network (NN) and recalculated energies obtained from density-functional theory (DFT) for the initial and final structures in each step of a metadynamics simulation [14] of bulk silicon. Each metastep involves a molecular dynamics simulation of 2 ps. The metadynamics simulation for the 64 atom cell starts from the β -tin structure with a pressure of 15 GPa at 300 K.

Behler-Parrinello network limitations

“A limitation of the NN is the lack of extrapolation capability to structures very different from the structures included in the training set. We thus do not expect our current parametrization for bulk silicon to yield good results for silicon clusters.”

“Because of the cutoff applied to the symmetry functions, long-range interactions are not included in the present implementation.”

“An extension to multicomponent systems is straightforward, but requires the incorporation of the corresponding cross terms in the symmetry functions and an extended set of DFT energies.”

Coulomb matrix with kernel ridge regression

PRL 108, 058301 (2012)

PHYSICAL REVIEW LETTERS

week ending
3 FEBRUARY 2012

Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning

Matthias Rupp,^{1,2} Alexandre Tkatchenko,^{3,2} Klaus-Robert Müller,^{1,2} and O. Anatole von Lilienfeld^{4,2,*}

¹*Machine Learning Group, Technical University of Berlin, Franklinstr 28/29, 10587 Berlin, Germany*

²*Institute of Pure and Applied Mathematics, University of California Los Angeles, Los Angeles, California 90095, USA*

³*Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany*

⁴*Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439, USA*

(Received 15 June 2011; published 31 January 2012)

We introduce a machine learning model to predict atomization energies of a diverse set of organic molecules, based on nuclear charges and atomic positions only. The problem of solving the molecular Schrödinger equation is mapped onto a nonlinear statistical regression problem of reduced complexity. Regression models are trained on and compared to atomization energies computed with hybrid density-functional theory. Cross validation over more than seven thousand organic molecules yields a mean absolute error of ~ 10 kcal/mol. Applicability is demonstrated for the prediction of molecular atomization potential energy curves.

Coulomb matrix with kernel ridge regression

“[...] A nonlinear regression ML model for computing molecular atomization energies in chemical compound space.”

“Our model is based on a measure of distance in compound space that accounts for both stoichiometry and configurational variation.

“[...] test bed consisting of all 7165 organic molecules [...], with up to seven “heavy” atoms that contain C, N, O, or S, being saturated with hydrogen atoms.”

Coulomb matrix with kernel ridge regression

“Descriptors”: A molecule is represented by a “Coulomb matrix”

$$M_{IJ} = \begin{cases} 0.5Z_I^2 & \text{for } I = J, \\ \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & \text{for } I \neq J. \end{cases}$$

{ \mathbf{R}_I } nuclear coordinates
{ Z_I } nuclear charges

Distance between two molecules is

$$d(\mathbf{M}, \mathbf{M}') = d(\boldsymbol{\epsilon}, \boldsymbol{\epsilon}') = \sqrt{\sum_I |\epsilon_I - \epsilon'_I|^2}$$

$\boldsymbol{\epsilon}$ are the eigenvalues of \mathbf{M} in order of decreasing absolute value

Energy of a molecule is expressed as a sum of weighted Gaussians

$$E^{\text{est}}(\mathbf{M}) = \sum_{i=1}^N \alpha_i \exp\left[-\frac{1}{2\sigma^2} d(\mathbf{M}, \mathbf{M}_i)^2\right],$$

Coulomb matrix with kernel ridge regression

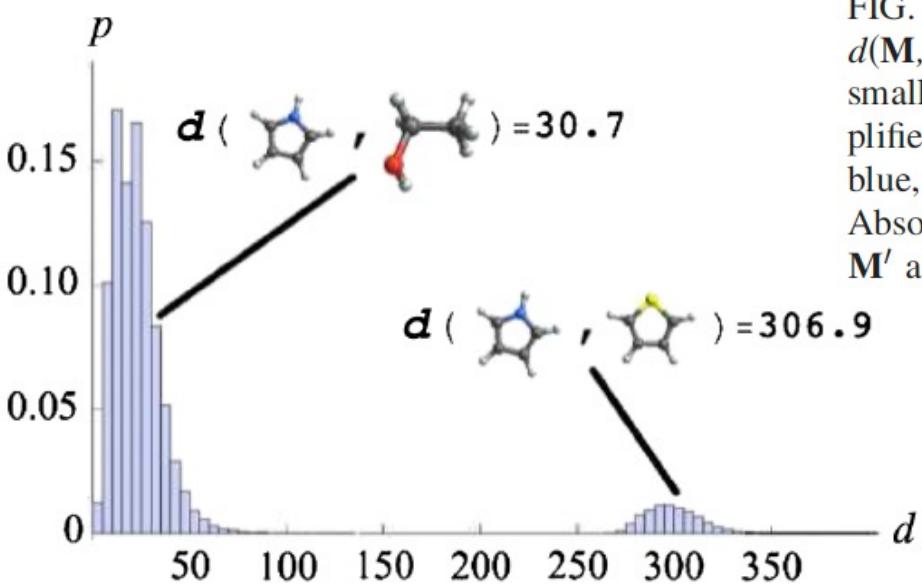
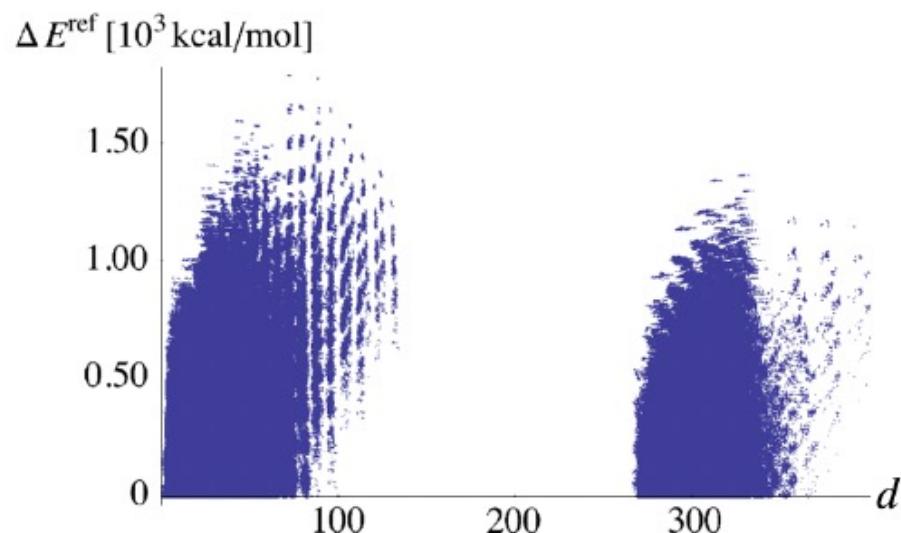


FIG. 1 (color online). Top: Distribution of distances, $d(\mathbf{M}, \mathbf{M}')$, for all molecular pairs occurring in the first 7165 small organic molecules from the GDB [15]. The inset exemplifies two distances, pyrrol/ethanol and pyrrol/thiophene (N: blue, O: red, S: yellow, C: black, and H: white). Bottom: Absolute differences in atomization energies between \mathbf{M} and \mathbf{M}' as a function of $d(\mathbf{M}, \mathbf{M}')$.



Coulomb matrix with kernel ridge regression

Energy of a molecule is expressed as a sum of weighted Gaussians

$$E^{\text{est}}(\mathbf{M}) = \sum_{i=1}^N \alpha_i \exp\left[-\frac{1}{2\sigma^2} d(\mathbf{M}, \mathbf{M}_i)^2\right],$$

Coefficients $\{\alpha_i\}$ are solutions of the minimization problem

$$\min_{\boldsymbol{\alpha}} \sum_i (E^{\text{est}}(\mathbf{M}_i) - E_i^{\text{ref}})^2 + \lambda \sum_i \alpha_i^2$$

DFT energy

Solution is given by kernel ridge regression

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{E}^{\text{ref}}, \quad K_{ij} = \exp[-d(\mathbf{M}_i, \mathbf{M}_j)^2 / (2\sigma^2)]$$

λ, σ are hyperparameters determined by cross-validation

Coulomb matrix with kernel ridge regression

Results

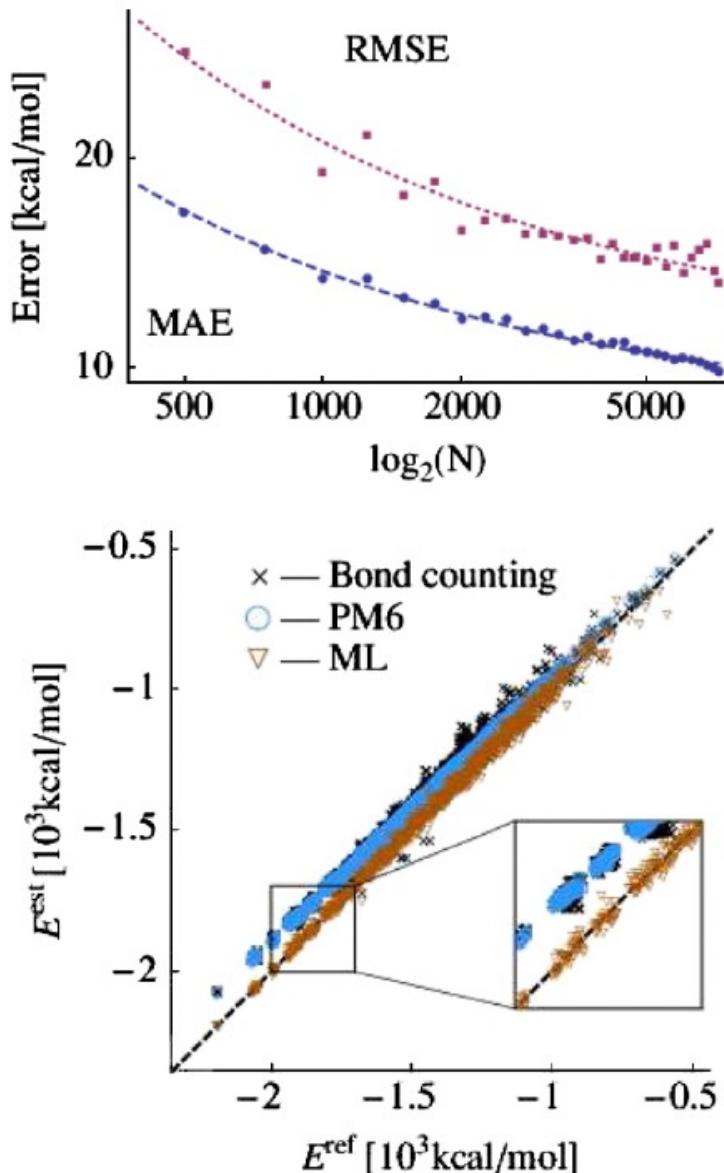


FIG. 2 (color online). Top: Cross-validated ML errors as a function of the number of molecules in the training set, N . Bottom: For $N = 1000$, correlation of DFT-PBE0 [19,20] results (E^{ref}) with ML (cross-validated) based estimates (E^{est}) of atomization energies. Correlations for bond counting [32] and semi-empirical quantum chemistry (PM6 [33]) are also shown. Corresponding root mean square error (RMSE)/mean absolute error (MAE) for bond counting, PM6, and ML are 75.0/71.0, 75.1/73.1, and 30.1/14.9 kcal/mol, respectively.

Coulomb matrix with kernel ridge regression

Results

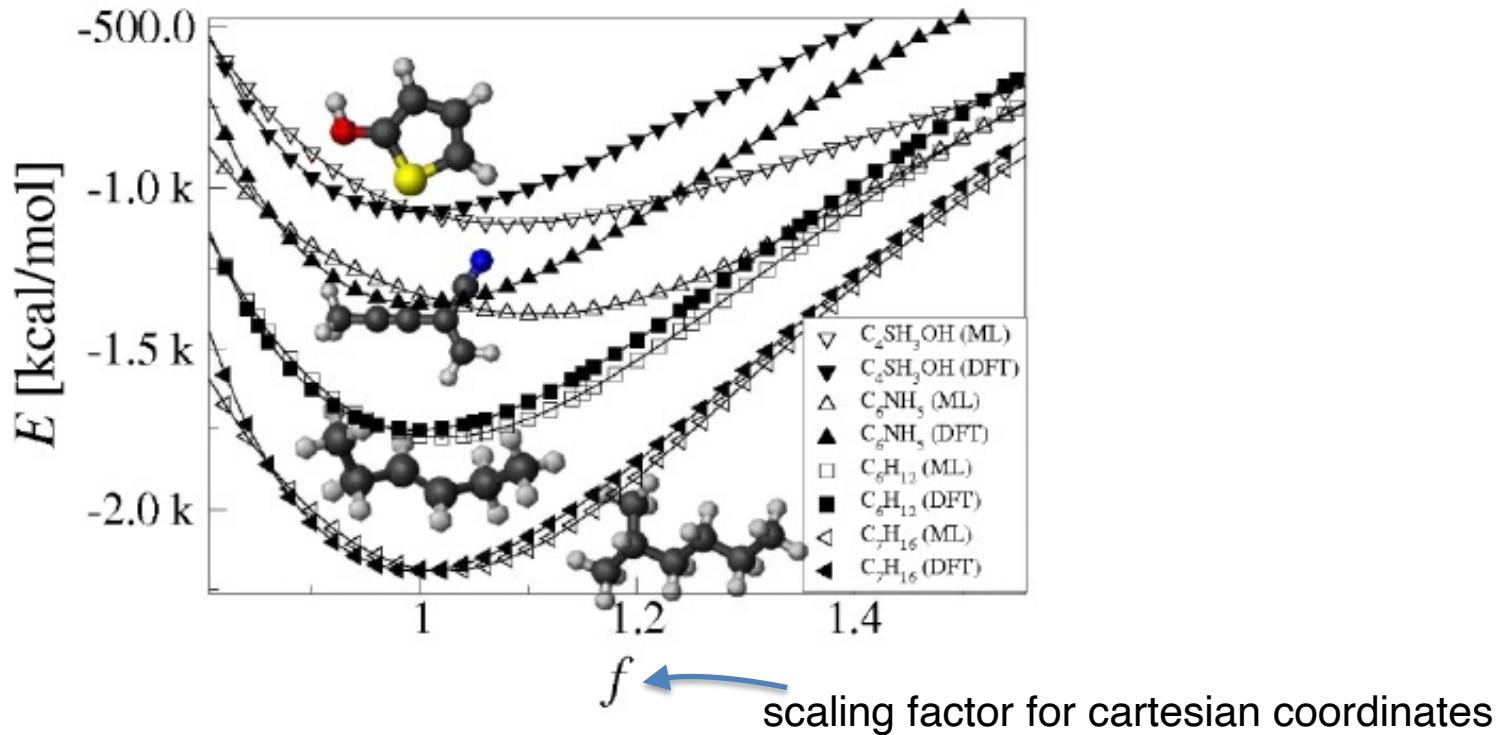


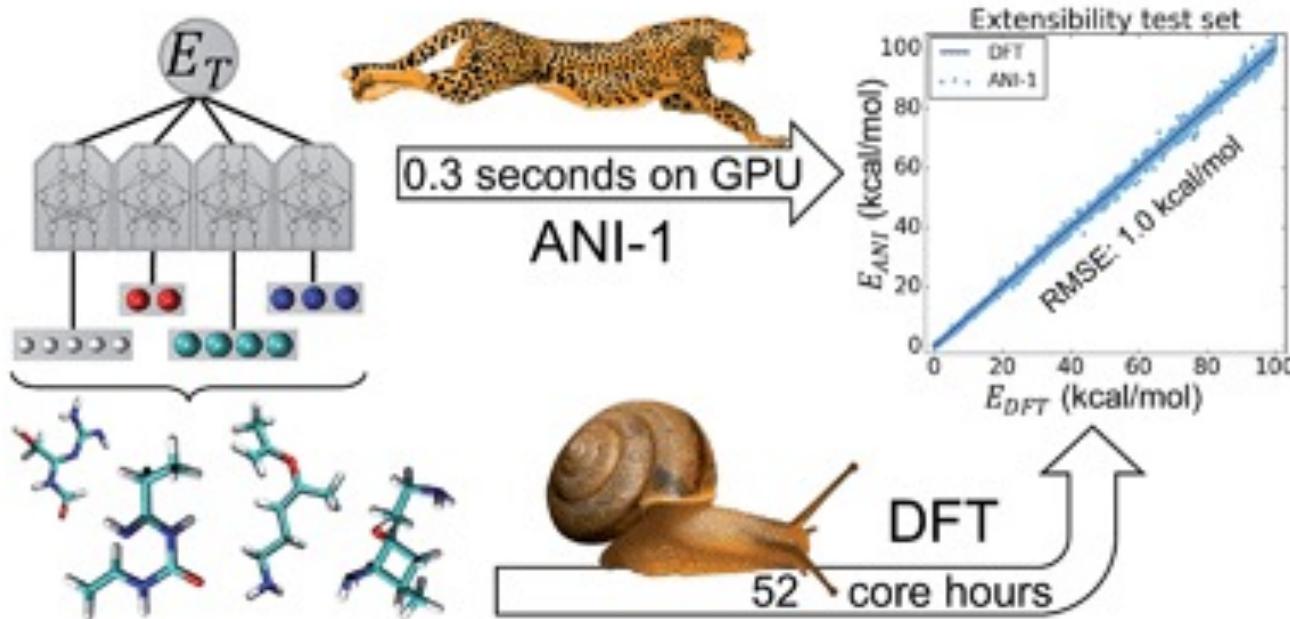
FIG. 3 (color online). Energy of atomization curves of four molecules containing single bonds and branching only (C_7H_{16}), a double bond (C_6H_{12}), triple bonds including nitrogen (C_6NH_5), and a sulfur-containing cycle with a hydroxy group ($\text{C}_2\text{SH}_3\text{OH}$). (From bottom to top in insets: black: Carbon; blue: Nitrogen; yellow: Sulfur; red: Oxygen; white: Hydrogen) (DFT-PBE0 and ML model 1k).

ANI-1

J. S. Smith, O. Isayev, and A. E. Roitberg

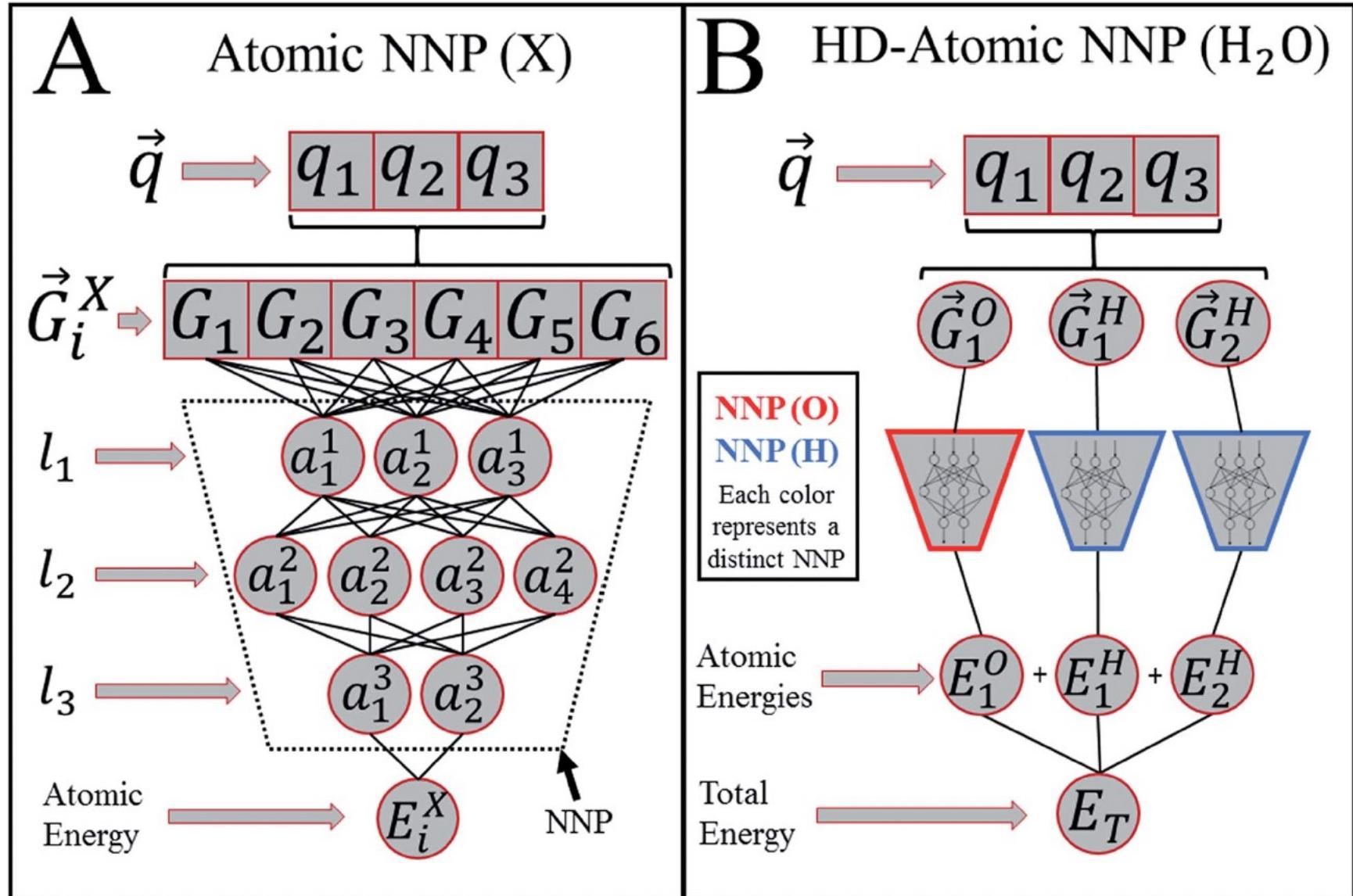
“ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost”, Chem. Sci. 8, 3192 (2017)

“First truly transferable neural network potential (NNP) for organic molecules based on a deep learning architecture”



ANAKIN-ME or ANI for short
(Accurate NeurAI networK englNe for Molecular Energies)

ANI-1



ANI-1

Two different kinds of symmetry functions:

1. Radial symmetry functions

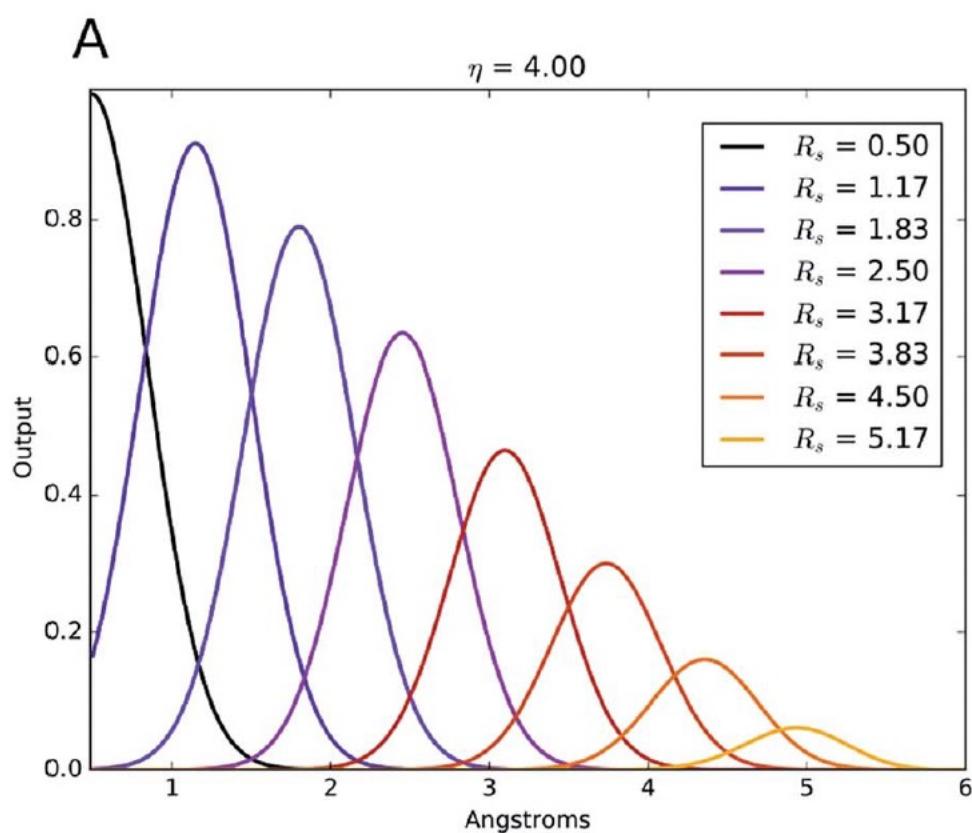
$$G_m^R = \sum_{j \neq i}^{\text{all atoms}} e^{-\eta(R_{ij}-R_s)^2} f_C(R_{ij})$$

$$f_C(R_{ij}) = \begin{cases} 0.5 \times \cos\left(\frac{\pi R_{ij}}{R_C}\right) + 0.5 & \text{for } R_{ij} \leq R_C \\ 0.0 & \text{for } R_{ij} > R_C \end{cases}$$

R_{ij} distance between atoms i and j

η, R_s hyperparameters

In practice, only a single η is used to produce thin Gaussian peaks and multiple R_s are used to probe outward from the atomic center



ANI-1

Two different kinds of symmetry functions:

1. Angular symmetry functions

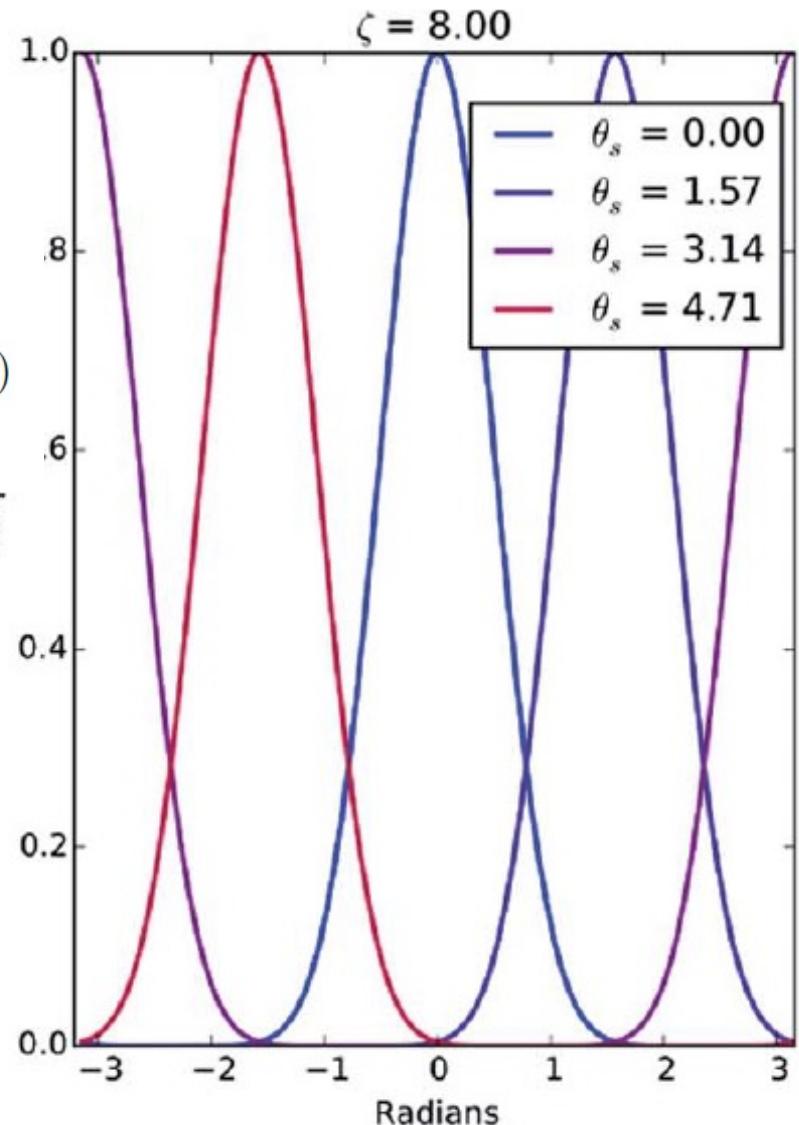
$$G_m^{\text{A}_{\text{mod}}} = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{all atoms}} (1 + \cos(\theta_{ijk} - \theta_s))^{\zeta} \times \exp \left[-\eta \left(\frac{R_{ij} + R_{ik}}{2} - R_s \right)^2 \right] f_C(R_{ij}) f_C(R_{ik})$$

$$f_C(R_{ij}) = \begin{cases} 0.5 \times \cos\left(\frac{\pi R_{ij}}{R_C}\right) + 0.5 & \text{for } R_{ij} \leq R_C \\ 0.0 & \text{for } R_{ij} > R_C \end{cases}$$

R_{ij} distance between atoms i and j

η, ζ, θ_s R_s are hyperparameters

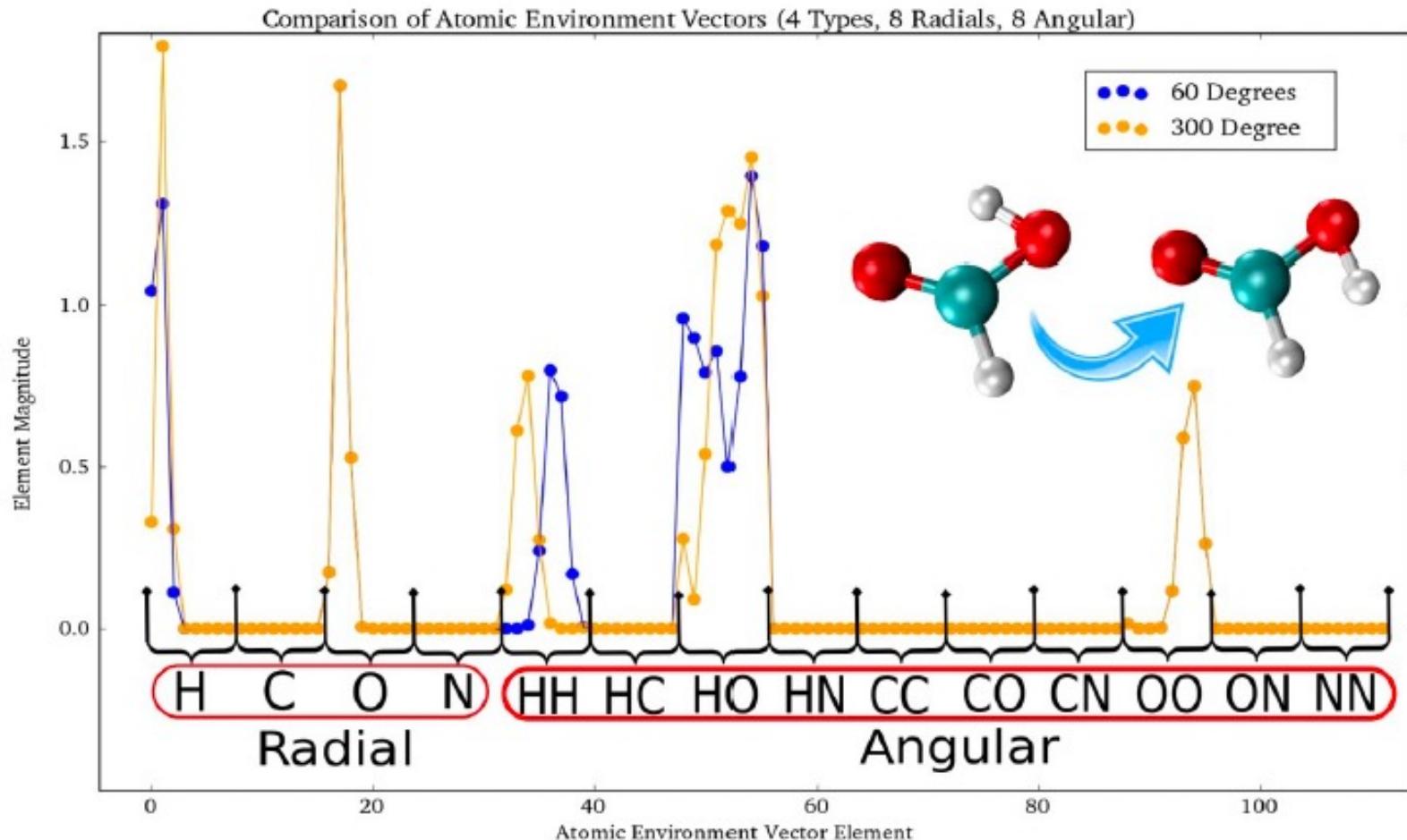
There is a different radial part for each atomic number and a different angular part for each atomic number pair in the local chemical environment.



ANI-1

A visualization of atomic environment vectors for the carbon atom in formic acid

Carbon Atomic Environment Vector



There is a different radial part for each atomic number and a different angular part for each atomic number pair in the local chemical environment.

32 evenly spaced radial shifting parameters are used for the radial part and a total of 8 radial and 8 angular shifting parameters are used for the angular part.

ANI-1

Data

57951 molecules selected as a subset of the GDB-11 database: they are all possible molecules containing up to 8 atoms of the atomic numbers H, C, N, O, filtered by chemical stability, synthetic feasibility considerations, and simple valency rules

Multiple configurations are generated for each molecule by Normal Mode Sampling.

The exact number of structures per molecule is determined using $K = S(3N - 6)$. S is an empirically determined value dependent on the number of heavy atoms in the molecule and N is the total number of atoms in the molecule, including hydrogens.

80% of data is used for training, 10% for validation, and 10% for testing.

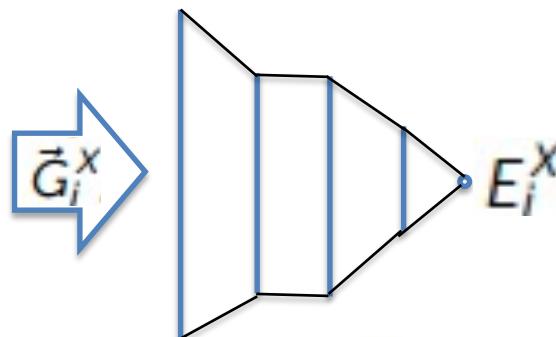
Number of heavy atoms	Total Molecules	Max Temperature	S value	Total data points	ANI-1 test set RMSE per atom (kcal/mol/atom)
1	3	2,000.0	500	8800	7.33×10^{-2}
2	13	1,500.0	450	39370	5.96×10^{-2}
3	20	1,000.0	425	128,880	4.16×10^{-2}
4	63	600.0	400	535,660	3.41×10^{-2}
5	275	600.0	200	1,444,890	3.71×10^{-2}
6	1,408	600.0	30	1,309,620	4.36×10^{-2}
7	7,850	600.0	20	5,276,930	6.65×10^{-2}
8	48,319	450.0	5	8,472,200	7.43×10^{-2}
Total	57,951	-	-	17,216,350	6.66×10^{-2}

Table S1: List of information and parameters used to generate the ANI-1 data set. The first column represents the number of heavy atoms per molecule in the test set. Total represents a combination of all test sets. The molecules are obtained from the GDB-11 database.

ANI-1

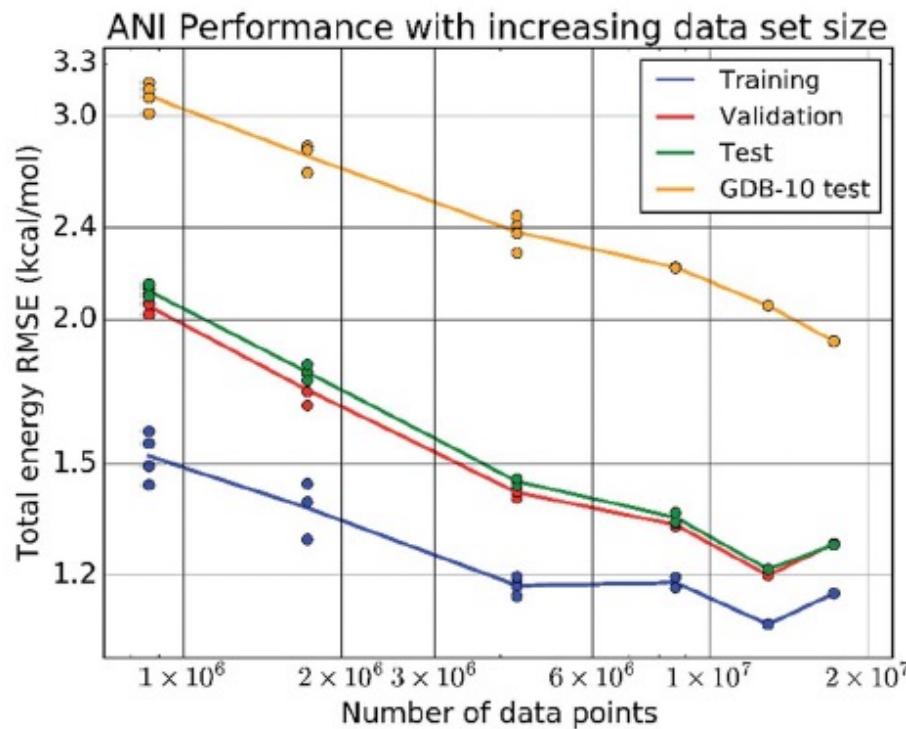
Atomic Network

pyramidal architecture
768 : 128 : 128 : 64 : 1



Gaussian activation function

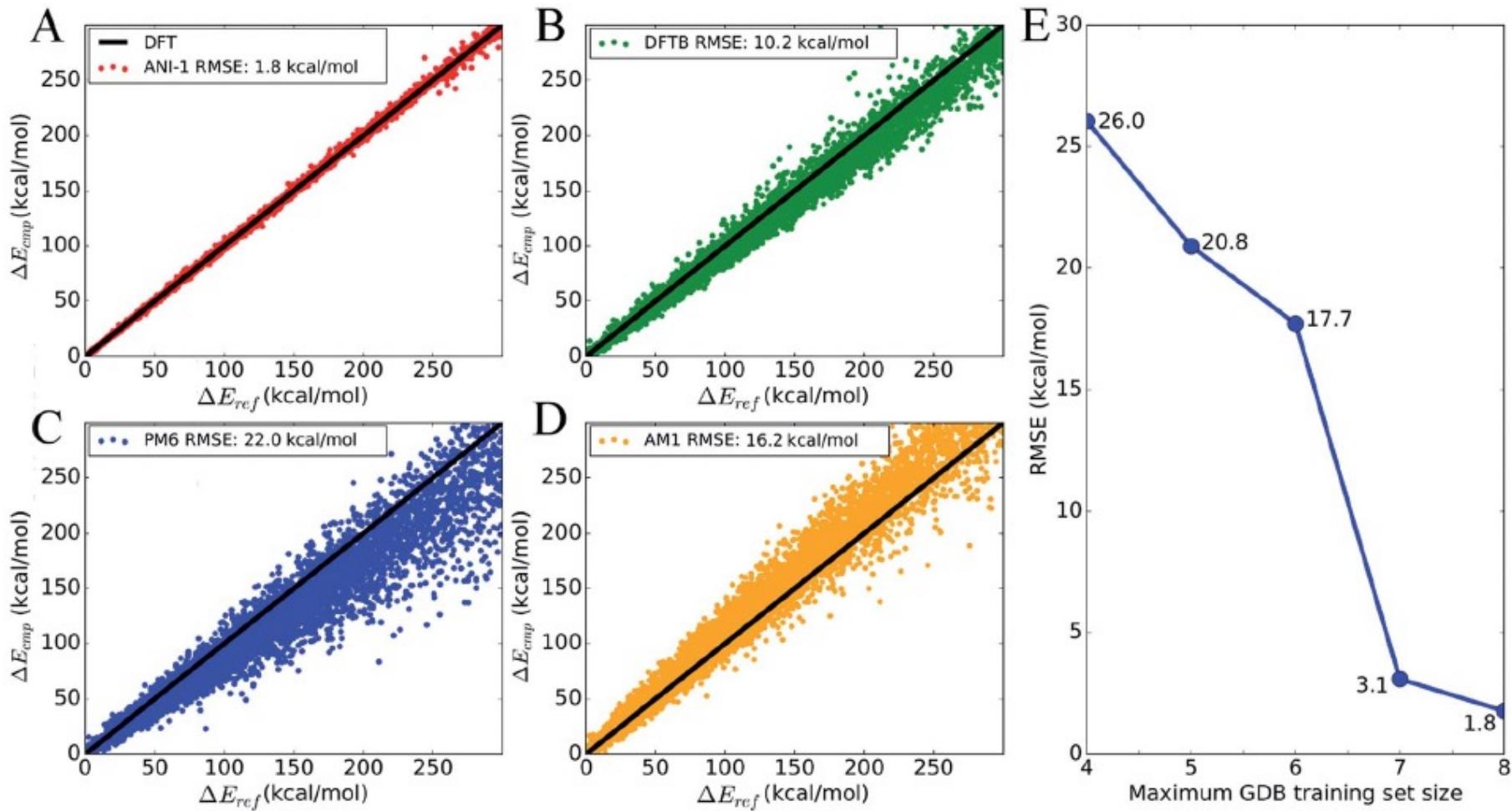
Loss function: $C(\vec{E}^{\text{ANI}}) = \tau \exp\left(\frac{1}{\tau} \sum_j (E_j^{\text{ANI}} - E_j^{\text{DFT}})^2\right)$



ANI-1

Results

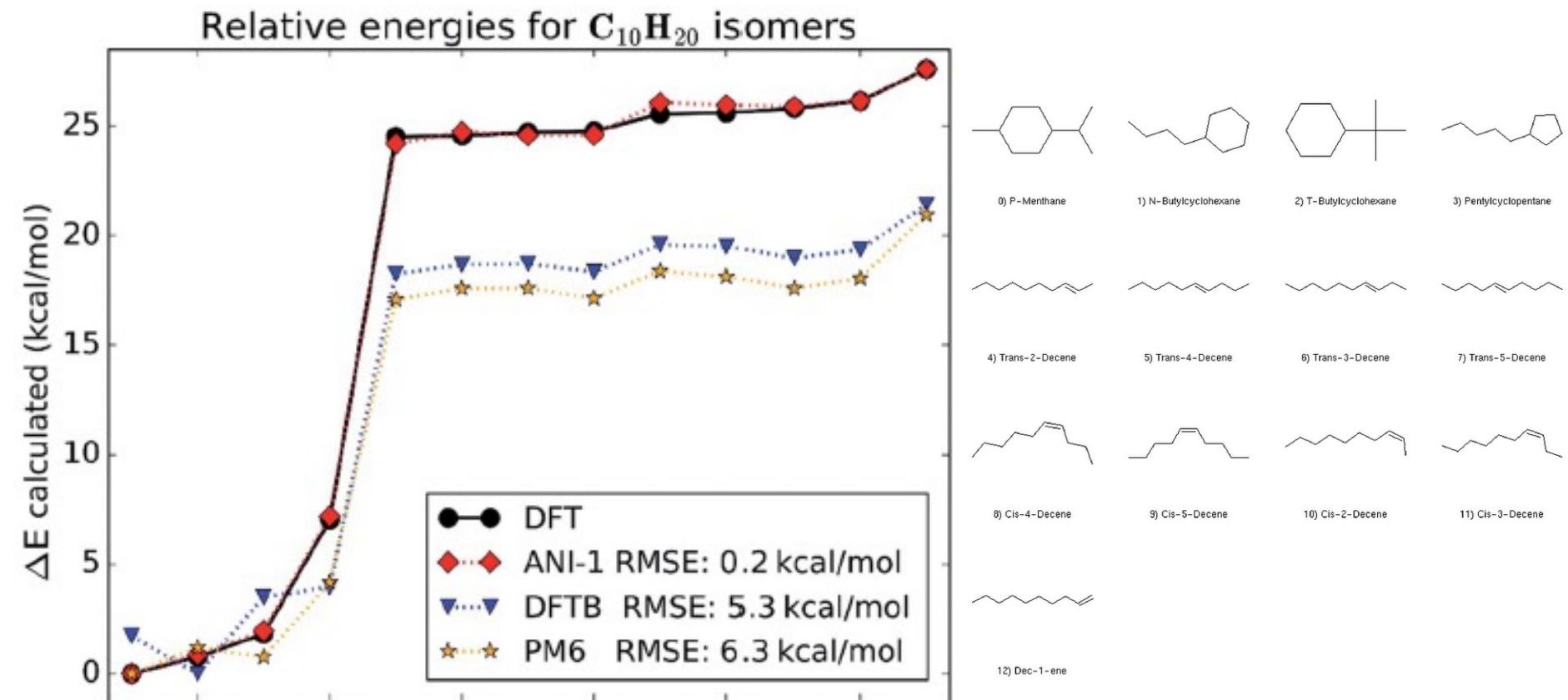
test of performance on 134 molecules from GDB-11 all with 10 heavy atoms, with an average of 62 conformations per molecule.



ANI-1

Results

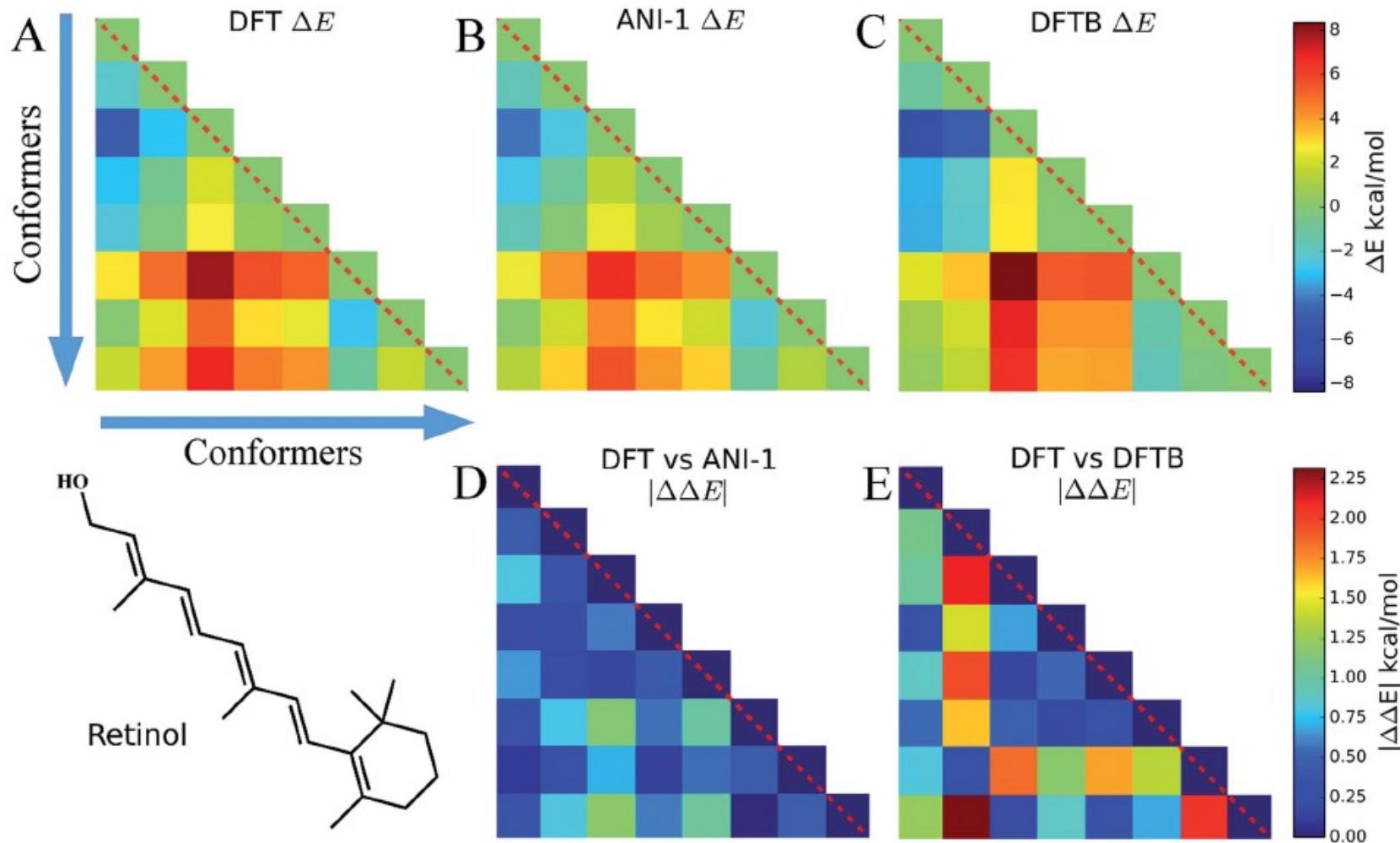
test of performance on relative stabilities of structural and geometric isomers



ANI-1

Results

test of performance on energy difference between pairs of configurations of Retinol



ANI-1

Results potential energy surfaces

