CECAM
Centre Européen de Calcul Atomique et Moléculaire

ML molecular dynamics: ground state (II)

Pavlo O. Dral

dral@xmu.edu.cn; dr-dral.com

Xiamen University, China

MLatom.com

XACS

A Package for
Atomistic Simulations
with **M**achine **L**earning

MLQCDyn
reinventing dynamics

7 September 2022

ML-Infrared spectra

To summarize, the steps *Gaussian* uses to perform vibrational analysis are

1. Mass weight the Hessian

$$f_{\mathrm{MWC}ij} = \frac{f_{\mathrm{CART}ij}}{\sqrt{m_i m_j}} \qquad f_{\mathrm{CART}ij} = \left(\frac{\partial^2 V}{\partial \xi_i \partial \xi_j}\right)_0$$

2. Determine the principal axes of inertia

$$\mathbf{I}' = \mathbf{X}^{\dagger}\mathbf{I}\mathbf{X}$$

3. Generate coordinates in the rotating and translating frame

$$\mathbf{S} = \mathbf{D}\mathbf{q}$$

4. Transform the Hessian to internal coordinates and diagonalize

$$\mathbf{f}_{\mathrm{INT}} = \mathbf{D}^{\dagger}\mathbf{f}_{\mathrm{MWC}}\mathbf{D}$$

$$\mathbf{L}^{\dagger}\mathbf{f}_{\mathrm{INT}}\mathbf{L} = \mathbf{\Lambda}$$

5. Calculate the frequencies

$$\tilde{\nu}_i = \sqrt{\frac{\lambda_i}{4\pi^2 c^2}}$$

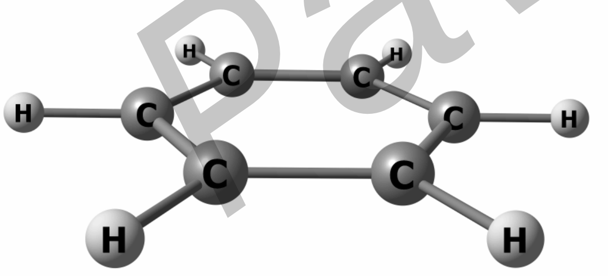
6. Calculate reduced mass, force constants and cartesian displacements

$$\mu_i = \left(\sum_k^{3N} l_{\mathrm{CART}\,k,i}^{\;2}\right)^{-1}$$

$$k_i = 4\pi^2 \tilde{\nu}_i^2 c^2 \mu_i$$

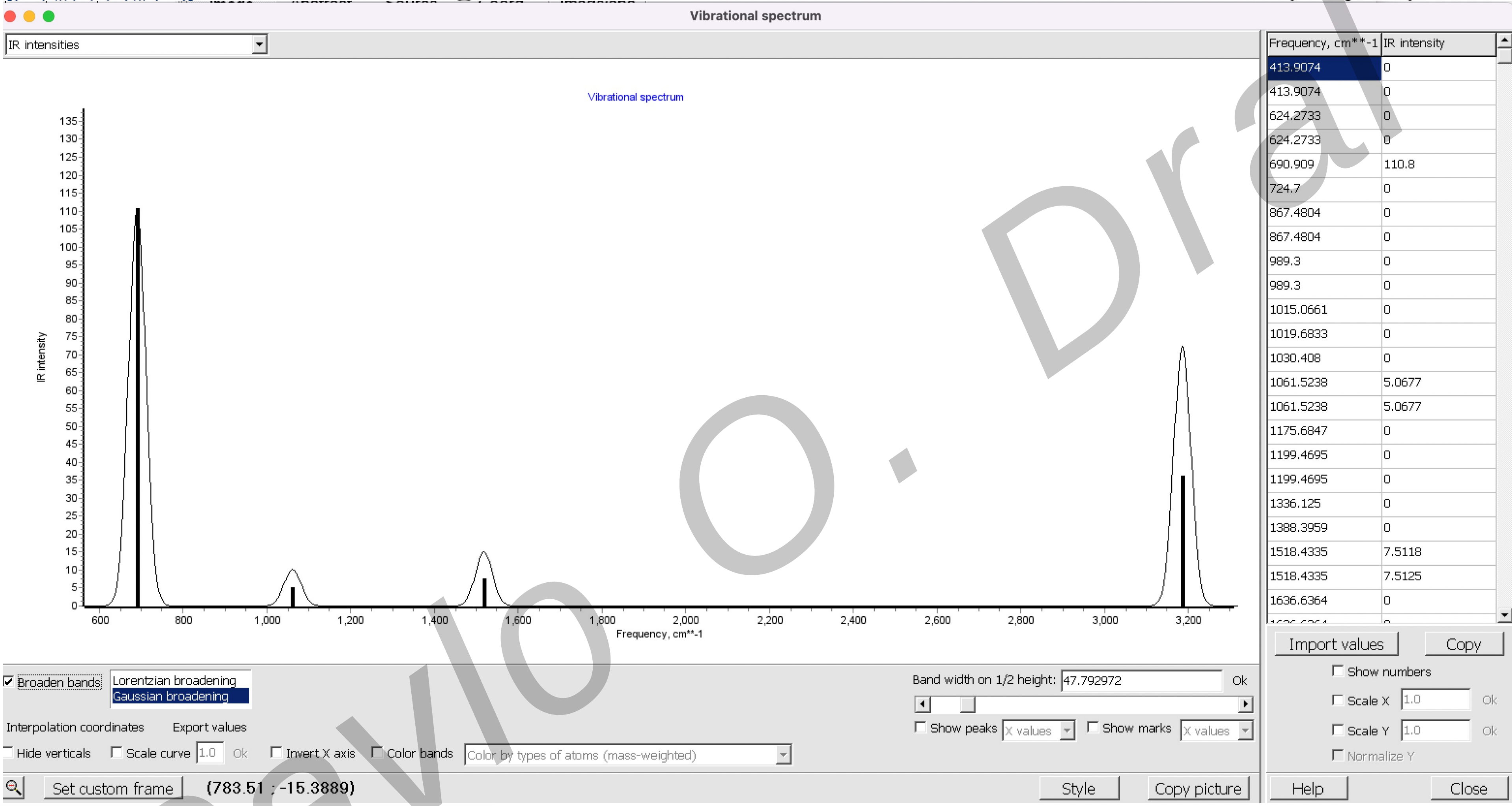
$$\mathbf{l}_{\mathrm{CART}} = \mathbf{MDL}$$

Source: https://gaussian.com/vib/



To summarize, the steps *Gaussian* uses to perform vibrational analysis are

1. Mass weight the Hessian



Frequency, cm**-1	IR intensity
413.9074	0
413.9074	0
624.2733	0
624.2733	0
690.909	110.8
724.7	0
867.4804	0
867.4804	0
989.3	0
989.3	0
1015.0661	0
1019.6833	0
1030.408	0
1061.5238	5.0677
1061.5238	5.0677
1175.6847	0
1199.4695	0
1199.4695	0
1336.125	0
1388.3959	0
1518.4335	7.5118
1518.4335	7.5125
1636.6364	0

☑ Broaden bands

Lorentzian broadening

Gaussian broadening

Band width on 1/2 height: 47.792972

☐ Show peaks X values ☐ Show marks X values

☐ Hide verticals ☐ Scale curve 1.0 ☐ Invert X axis ☐ Color bands Color by types of atoms (mass-weighted)

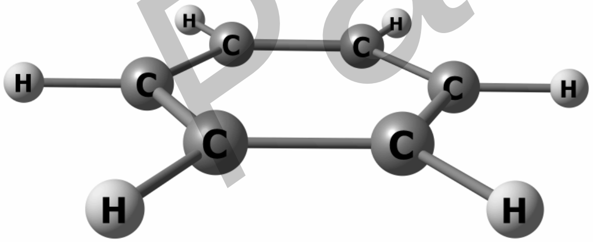
☐ Show numbers

☐ Scale X 1.0

☐ Scale Y 1.0

☐ Normalize Y

(783.51 ; -15.3889)



$$\mu_i = \left(\sum_k^{3N} l_{\mathrm{CART}}{}_{k,i}^2 \right)^{-1}$$

$$k_i = 4\pi^2 \tilde{\nu}_i^2 c^2 \mu_i$$

$$\mathbf{l}_{\mathrm{CART}} = \mathbf{MDL}$$

Variants of vibrational simulations

Vibrational Schrödinger equation:

$$\hat{H}\Psi_\mathrm{n}(\boldsymbol{q}) = \hat{T}\Psi_\mathrm{n}(\boldsymbol{q}) + V(\boldsymbol{q})\Psi_\mathrm{n}(\boldsymbol{q}) = E_n\Psi_n(\boldsymbol{q})$$

$$V(\boldsymbol{Q}) = \frac{1}{2}\sum_{i=1}^{D}\omega_i Q_i^2$$

Easy to solve within harmonic approximation

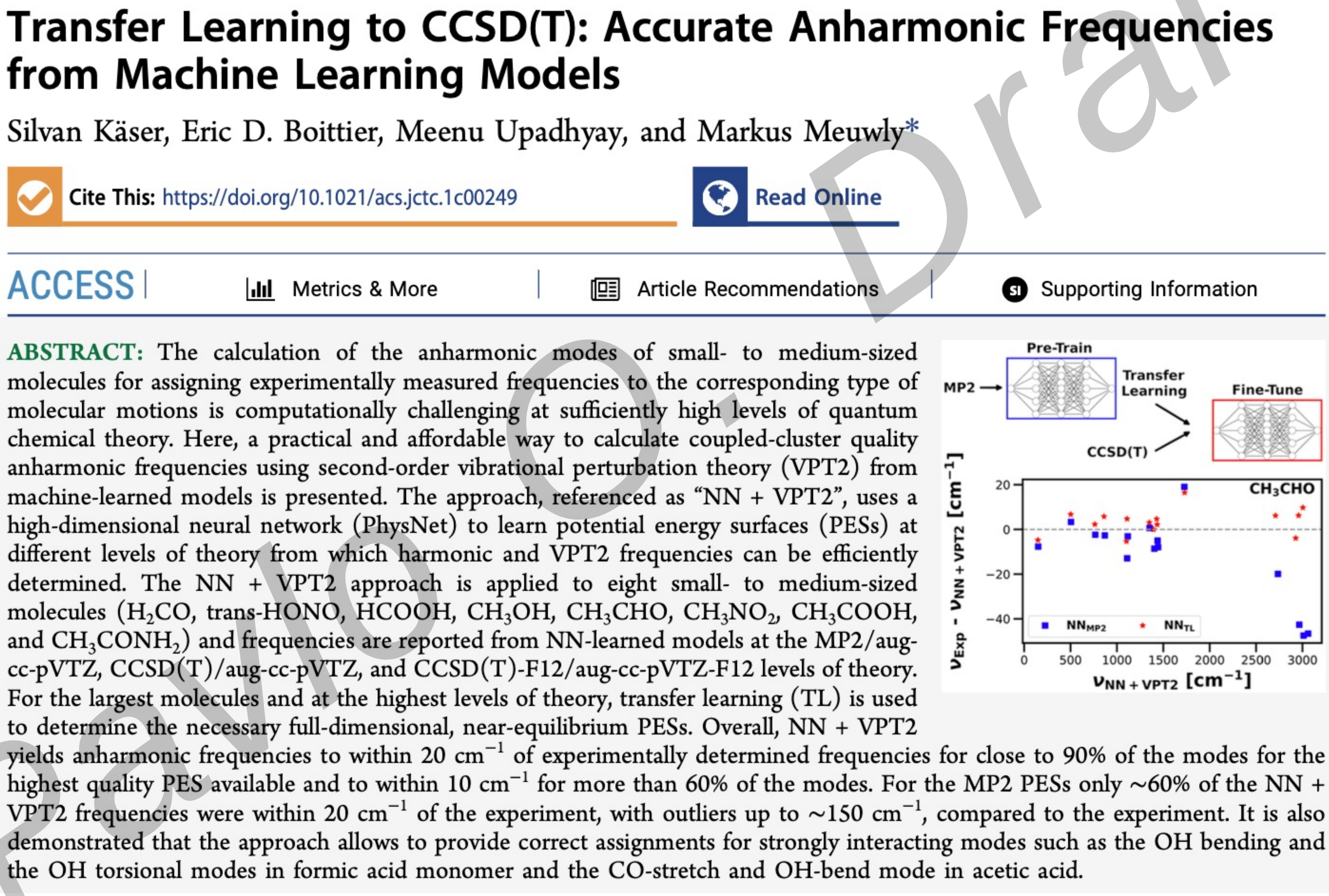
Real potentials are anharmonic:

$$V(\boldsymbol{Q}) = \frac{1}{2}\sum_{i=1}^{D}\omega_i\ \ Q_i^2 + \left(\frac{1}{6}\sum_{i,j,k=1}^{D}\eta_{ijk}Q_iQ_jQ_k + \frac{1}{24}\sum_{i,j,k=1}^{D}\eta_{ijkl}Q_iQ_jQ_kQ_l\right) = V^\mathrm{harm} + V'$$

Methods like VPT2, VCI, VSCF…

S. Manzhos, M. Ihara, T. Carrington, Machine learning for vibrational spectroscopy. In *Quantum Chemistry in the Age of Machine Learning*, P. O. Dral, Ed. Elsevier: 2022, in press. Paperback ISBN: 9780323900492

ML-VPT2



Transfer Learning to CCSD(T): Accurate Anharmonic Frequencies from Machine Learning Models

Silvan Käser, Eric D. Boittier, Meenu Upadhyay, and Markus Meuwly*

Cite This: https://doi.org/10.1021/acs.jctc.1c00249

ABSTRACT: The calculation of the anharmonic modes of small- to medium-sized molecules for assigning experimentally measured frequencies to the corresponding type of molecular motions is computationally challenging at sufficiently high levels of quantum chemical theory. Here, a practical and affordable way to calculate coupled-cluster quality anharmonic frequencies using second-order vibrational perturbation theory (VPT2) from machine-learned models is presented. The approach, referenced as "NN + VPT2", uses a high-dimensional neural network (PhysNet) to learn potential energy surfaces (PESs) at different levels of theory from which harmonic and VPT2 frequencies can be efficiently determined. The NN + VPT2 approach is applied to eight small- to medium-sized molecules (H_2CO, trans-HONO, HCOOH, CH_3OH, CH_3CHO, CH_3NO_2, CH_3COOH, and CH_3CONH_2) and frequencies are reported from NN-learned models at the MP2/aug-cc-pVTZ, CCSD(T)/aug-cc-pVTZ, and CCSD(T)-F12/aug-cc-pVTZ-F12 levels of theory. For the largest molecules and at the highest levels of theory, transfer learning (TL) is used to determine the necessary full-dimensional, near-equilibrium PESs. Overall, NN + VPT2 yields anharmonic frequencies to within 20 cm^{-1} of experimentally determined frequencies for close to 90% of the modes for the highest quality PES available and to within 10 cm^{-1} for more than 60% of the modes. For the MP2 PESs only ~60% of the NN + VPT2 frequencies were within 20 cm^{-1} of the experiment, with outliers up to ~150 cm^{-1}, compared to the experiment. It is also demonstrated that the approach allows to provide correct assignments for strongly interacting modes such as the OH bending and the OH torsional modes in formic acid monomer and the CO-stretch and OH-bend mode in acetic acid.

Variants of vibrational simulations

Also possible from molecular dynamics:

$$I(\omega) \propto \int \langle \boldsymbol{M}(\tau)\boldsymbol{M}(t+\tau)\rangle_\tau e^{-i\omega t} dt$$

where $\langle \boldsymbol{M}(\tau)\boldsymbol{M}(t+\tau)\rangle_\tau$ is the autocorrelation function of the dipole moment $\boldsymbol{M}(t) = \sum_{j=1}^{Np} e_j \boldsymbol{r}_j(t)$, where $\boldsymbol{r}_j$ and e_j are charges of each of N_p particles.

S. Manzhos, M. Ihara, T. Carrington, Machine learning for vibrational spectroscopy. In *Quantum Chemistry in the Age of Machine Learning*, P. O. Dral, Ed. Elsevier: 2022, in press. Paperback ISBN: 9780323900492

IR from MD: Learn PES and dipole moments

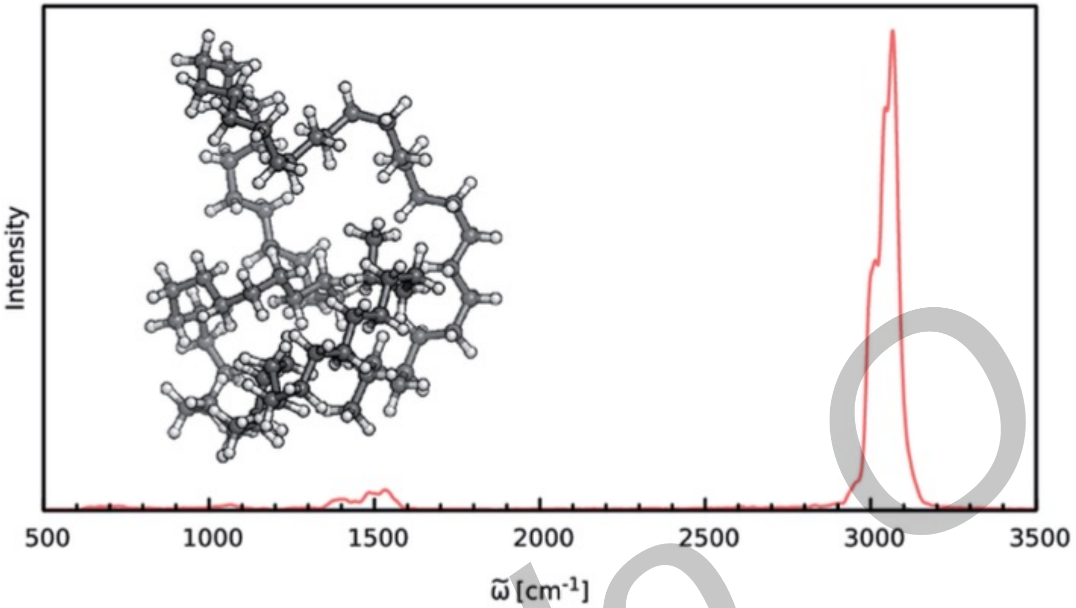
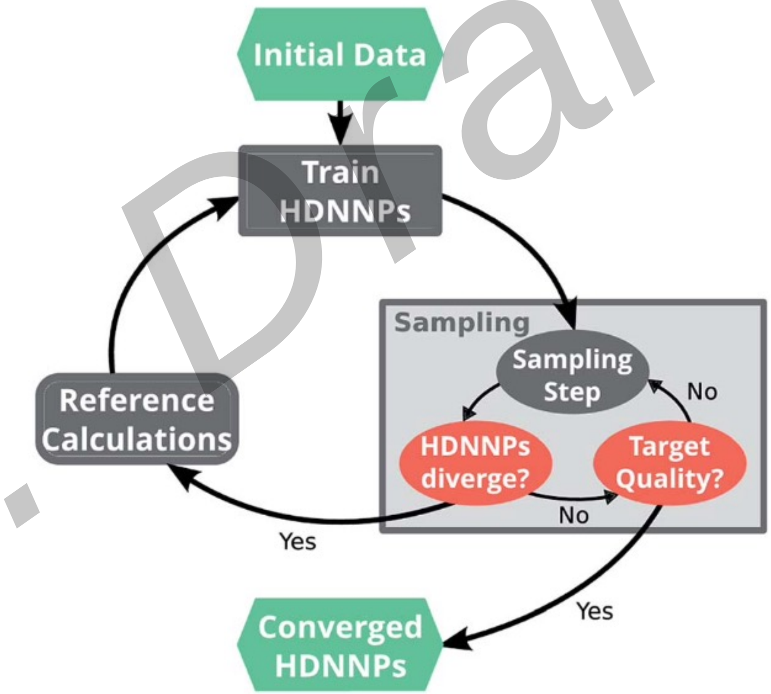
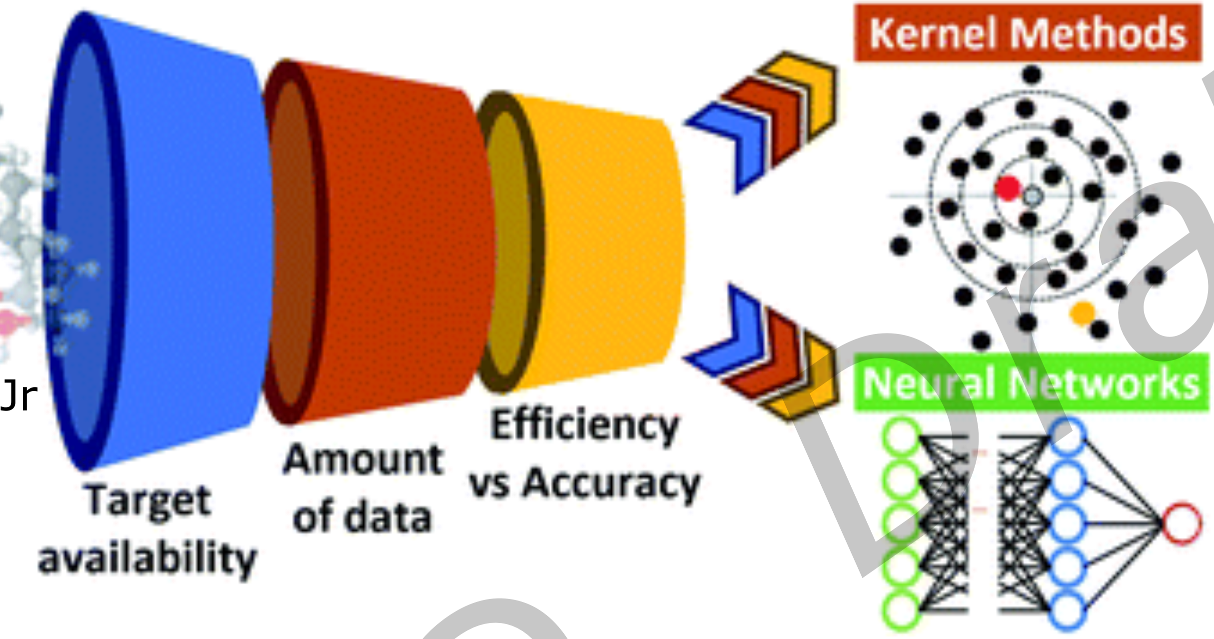


Fig. 6 IR spectrum of the $C_{69}H_{140}$ alkane as predicted by the ML model based on the B2PLYP method.



M. Gastegger, J. Behler, P. Marquetand, *Chem. Sci.* **2017**, *8*, 6924



Mario
Barbatti

Max
Pinheiro Jr

What machine learning potential to choose?

Fuchun Ge

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413
P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro Jr, J. Huang, M. Barbatti. *Top. Curr. Chem.* **2021,** *379*, 27

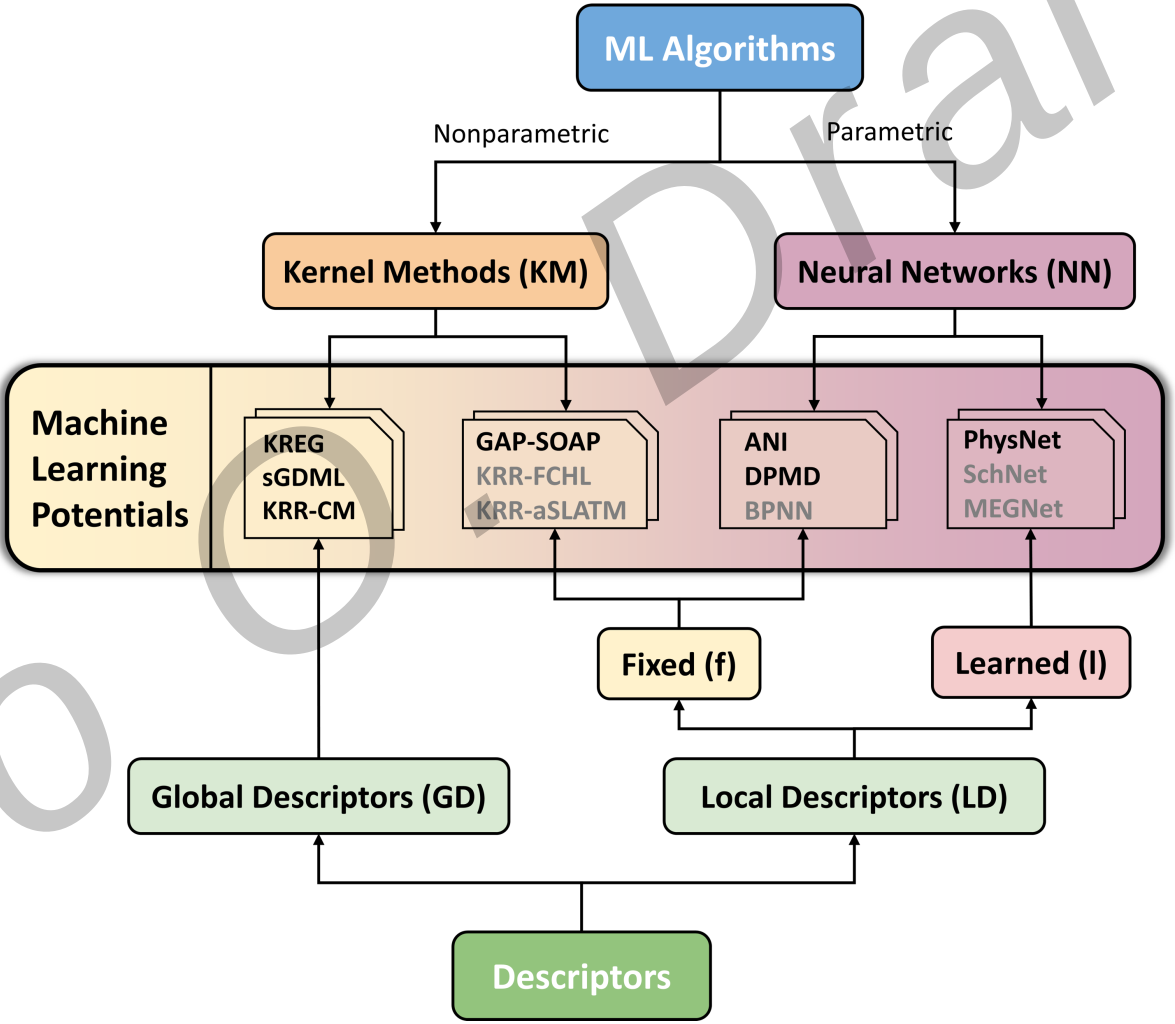
Zoo of machine learning potentials

Normal font – implemented models in MLatom

Greyed – not implemented in MLatom

MLatom.com

XACS



ML Algorithms

Nonparametric

Parametric

Kernel Methods (KM)

Neural Networks (NN)

Machine Learning Potentials

KREG
sGDML
KRR-CM

GAP-SOAP
KRR-FCHL
KRR-aSLATM

ANI
DPMD
BPNN

PhysNet
SchNet
MEGNet

Fixed (f)

Learned (l)

Global Descriptors (GD)

Local Descriptors (LD)

Descriptors

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413

MLatom's 3rd-party interfaces

DeePMD-kit	deepmodeling.com
GAP	libatoms.github.io/GAP
sGDML	sgdml.org
TorchANI	aiqm.github.io/torchani
PhysNet	github.com/MMunibas/PhysNet
ASE	wiki.fysik.dtu.dk/ase
Hyperopt	hyperopt.github.io/hyperopt
Sparrow	scine.ethz.ch/download/sparrow
NX	newtonx.org
Gaussian	gaussian.com
MNDO	mndo.kofo.mpg.de

ML algorithms

Linear regression

$$f(\mathbf{x}_i;\boldsymbol{\beta}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots$$

Kernel ridge regression (KRR)

Number of parameters depends on the training set

$$f(\mathbf{x}_i;\boldsymbol{\alpha}) = \sum_{j=1}^{N_{\mathrm{tr}}} \alpha_j k(\mathbf{x}_i, \mathbf{x}_j)$$

‘nonparametric model’

Examples: KREG, KRR-CM

Neural networks (NN)

Number of parameters is fixed

$$f(\mathbf{x};\boldsymbol{\beta}) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \cdots$$

$$Z_m = \sigma(\alpha_{m0} + \alpha_{m1} x_1 + \alpha_{m2} x_2 + \cdots)$$

‘parametric model’

Examples: ANI-1ccx, AIQM1

The KREG Model

$$f(\mathbf{x}_i) = \sum_{j=1}^{N_{\text{tr}}} \alpha_j k(\mathbf{x}_i, \mathbf{x}_j)$$

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{2\sigma^2}\sum_{s}^{N_x}\left(x_{i,s} - x_{j,s}\right)^2\right)$$

the Gaussian kernel function
(σ is the kernel width)

$$x = \left(\dots \quad \frac{R^{eq}}{R} \quad \dots\right)^T$$

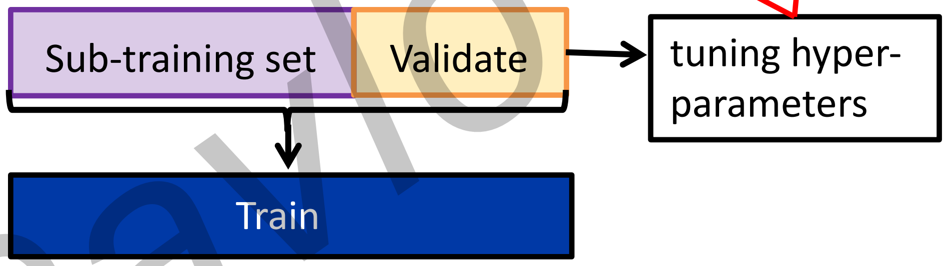
RE descriptor

Analytical solution for the regression coefficients α given N_{tr} training points

$$\begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) + \lambda & \cdots & k\left(\mathbf{x}_1, \mathbf{x}_{N_{\text{tr}}}\right) \\ \vdots & \ddots & \vdots \\ k\left(\mathbf{x}_{N_{\text{tr}}}, \mathbf{x}_1\right) & \cdots & k\left(\mathbf{x}_{N_{\text{tr}}}, \mathbf{x}_{N_{\text{tr}}}\right) + \lambda \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{N_{\text{tr}}} \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_{N_{\text{tr}}} \end{pmatrix}$$

$$(\mathbf{K} + \lambda\mathbf{I})\boldsymbol{\alpha} = \mathbf{y}$$

λ is the regularization parameter ensuring transferability



The **KREG** model (**K**ernel ridge regression [KRR] with **RE** descriptor and the **G**aussian kernel function; RE descriptor stands for Internuclear distances **R**elative to **E**quilibrium) to complete all the ML tasks.

P. O. Dral, A. Owens, S. Yurchenko, W. Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

sGDML

sGDML – a symmetrized variant of the gradient-domain machine learning (GDML). It uses unnormalized permutationally invariant, Matérn kernel.

$$\mathbf{x}^{\mathrm{T}} = \begin{bmatrix} \cdots & \frac{1}{r_{a,b}} & \cdots \end{bmatrix}$$

$$\frac{\partial f(\mathbf{x})}{\partial M_{at}} = \sum_{j=1}^{N_{\mathrm{tr}}} \sum_{b=1}^{N_{\mathrm{at}}} \sum_{u=1}^{3} \alpha_{j,bu} \frac{\partial^2 k(\mathbf{x}, \mathbf{x}_j)}{\partial M_{at} \partial M_{j,bu}}$$

$$\begin{pmatrix} \frac{\partial^2 k(\mathbf{x}_1, \mathbf{x}_1)}{\partial M_{1,11} \partial M_{1,11}} + \lambda & \cdots & \frac{\partial^2 k\left(\mathbf{x}_1, \mathbf{x}_{N_{\mathrm{tr}}}\right)}{\partial M_{1,11} \partial M_{N_{\mathrm{tr}},N_{\mathrm{at}}3}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 k\left(\mathbf{x}_{N_{\mathrm{tr}}}, \mathbf{x}_1\right)}{\partial M_{N_{\mathrm{tr}},N_{\mathrm{at}}3} \partial M_{1,11}} & \cdots & \frac{\partial^2 k\left(\mathbf{x}_{N_{\mathrm{tr}}}, \mathbf{x}_{N_{\mathrm{tr}}}\right)}{\partial M_{N_{\mathrm{tr}},N_{\mathrm{at}}3} \partial M_{N_{\mathrm{tr}},N_{\mathrm{at}}3}} + \lambda \end{pmatrix} \begin{pmatrix} \alpha_{1,11} \\ \vdots \\ \alpha_{N_{\mathrm{tr}},N_{\mathrm{at}}3} \end{pmatrix} = \begin{pmatrix} \frac{\partial y_1}{\partial M_{1,11}} \\ \vdots \\ \frac{\partial y_{N_{\mathrm{tr}}}}{\partial M_{N_{\mathrm{tr}},N_{\mathrm{at}}3}} \end{pmatrix}$$

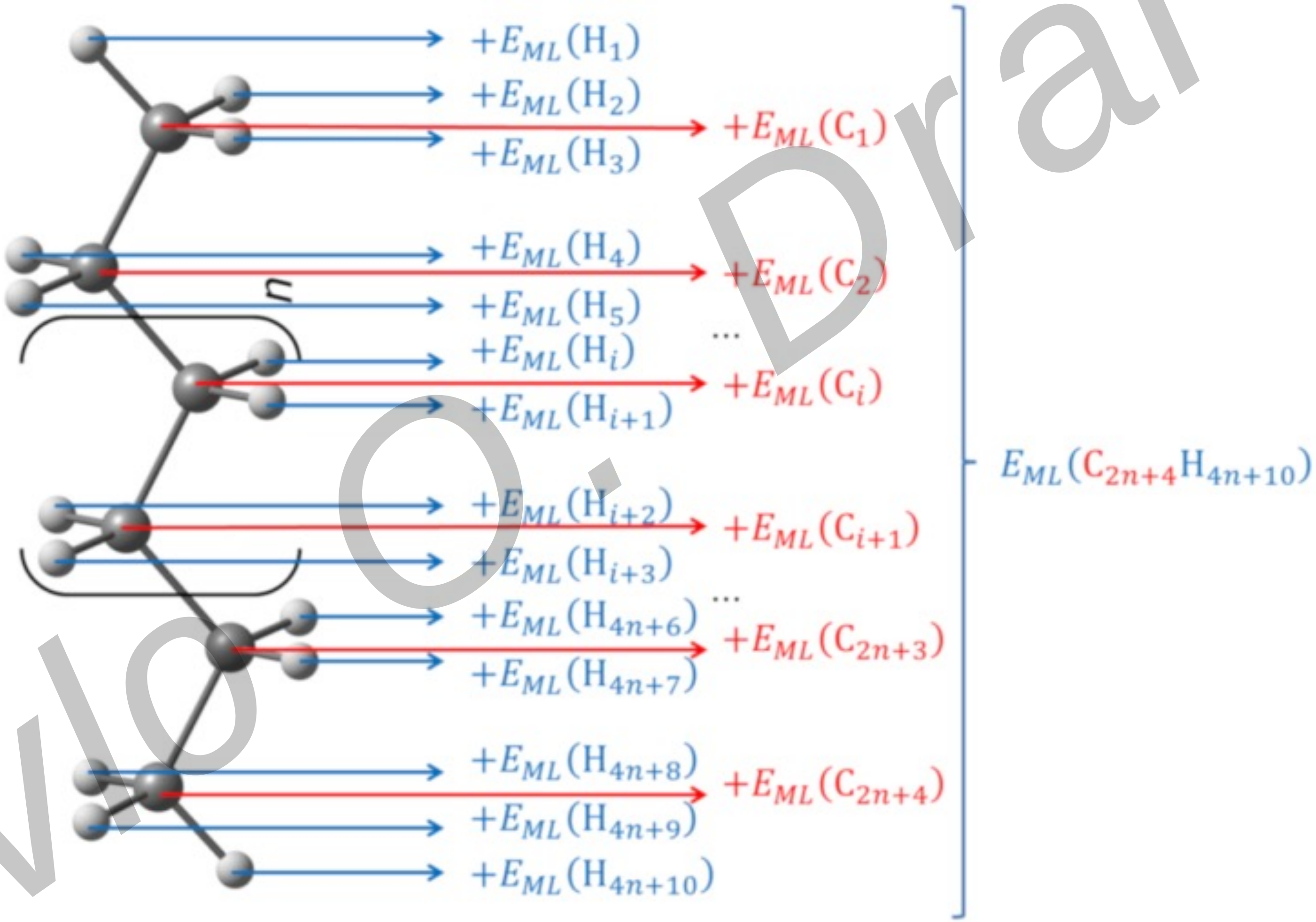
$$f(\mathbf{x}) = const + \sum_{j=1}^{N_{\mathrm{tr}}} \sum_{b=1}^{N_{\mathrm{at}}} \sum_{u=1}^{3} \alpha_{j,bu} \frac{\partial k(\mathbf{x}, \mathbf{x}_j)}{\partial M_{j,bu}}$$

S. Chmiela, H. E. Sauceda, K.-R. Müller, A. Tkatchenko. *Nat. Commun.* **2018,** *9*, 3887

Locality in Quantum Chemistry

Behler–Parrinello
ANI
PhysNet
SchNet
GAP-SOAP
FCHL
aSLATM
…

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413



Approach: Behler, Parrinello, *Phys. Rev. Lett.* **2007**, *98*, 146401

Figure: P. O. Dral, *J. Phys. Chem. Lett.* **2020**, *11*, 2336

GAP-SOAP

Gaussian approximation potential (GAP)[1] with Smooth Overlap of Atomic Positions (SOAP)[2] descriptor

the atomic neighborhood density

$$\rho_i(\mathbf{r}) = \sum_j \exp\left(-\frac{\left|\mathbf{r}-\mathbf{r}_{ij}\right|^2}{2\sigma_{atom}^2}\right) f_{\mathrm{cut}}\left(\left|\mathbf{r}_{ij}\right|\right),$$

the cutoff function

$$f_{\mathrm{cut}}(r) = \begin{cases} 1, & r \le r_{\mathrm{cut}} - r_\Delta, \\ \frac{1}{2}\left(\cos\left(\pi\frac{r - r_{\mathrm{cut}} + r_\Delta}{r_\Delta}\right) + 1\right), & r_{\mathrm{cut}} - r_\Delta < r \le r_{\mathrm{cut}}, \\ 0, & r > r_{\mathrm{cut}}, \end{cases}$$

[1] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, Phys. Rev. Lett. **2010**, *104*, 136403

[2] A. P. Bartók, R. Kondor, G. Csányi, Phys. Rev. B **2013**, *87*, 187115

DPMD and DeepPot-SE

DeePMD-kit[1] implements Deep Potential Molecular Dynamics (DPMD)[2] and Deep Potential - Smooth Edition (DeepPot-SE)[3]

the generalized local environment matrix $\tilde{\mathbf{R}}^{\mathbf{i}}$

$$\left\{ s(r_{ij}), \frac{x_{ij}}{r_{ij}} s(r_{ij}), \frac{y_{ij}}{r_{ij}} s(r_{ij}), \frac{z_{ij}}{r_{ij}} s(r_{ij}) \right\},$$

$$s(r_{ij}) = \begin{cases} \dfrac{1}{r_{ij}}, & r_{ij} < r_{cs} \\ \dfrac{1}{r_{ij}} \left(\dfrac{1}{2} \cos\left(\pi \dfrac{r_{ij} - r_{cs}}{r_c - r_{cs}} \right) + \dfrac{1}{2} \right), & r_{cs} < r_{ij} < r_c \\ 0, & r_{ij} > r_c \end{cases}$$

[1] H. Wang, L. Zhang, J. Han, W. E. Comput. Phys. Commun. 2018, 228, 178
[2] L. Zhang, J. Han, H. Wang, R. Car, W. E. *Phys. Rev. Lett.* **2018,** *120*, 143001
[3] L. F. Zhang, J. Q. Han, H. Wang, W. A. Saidi, R. Car, W. N. E. *Adv. Neural. Inf. Process. Syst.* **2018,** *31*, 4436

PhysNet

PhysNet[1] is using message-passing NN and so called 'learned' local descriptors

the embedding vector

$$\mathbf{x}_i^0 = \mathbf{e}_{z_i}$$

the coordinates are transformed to

$$g_k(r_{ij}) = f_c(r_{ij}) \cdot e^{-\beta_k \left(e^{-r_{ij}} - \mu_k\right)^2}$$

[1] O. T. Unke, M. Meuwly. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019,** *15*, 3678

ANI family of potentials

ANI environment vectors (AEVs) consist of radial and angular atomic terms. Each element has its own subAEV:

$$G_k^R = \sum_{j\neq i} e^{-\eta\left(R_{ij}-R_s^{(k)}\right)^2} f_c\left(R_{ij}\right)$$

$$G_{p,q}^A = 2^{1-\zeta} \sum_{j,k\neq i} \left(1+\cos\left(\theta_{ijk}-\theta_s^{(q)}\right)\right)^\zeta e^{-\eta\left(\frac{R_{ij}+R_{ik}}{2}-R_s^{(p)}\right)^2} f_C\left(R_{ij}\right) f_C\left(R_{ik}\right).$$

the cutoff function

$$f_{\text{cut}}(r) = \begin{cases} 1, & r \leq r_{\text{cut}} - r_\Delta, \\ \frac{1}{2}\left(\cos\left(\pi\frac{r - r_{\text{cut}} + r_\Delta}{r_\Delta}\right) + 1\right), & r_{\text{cut}} - r_\Delta < r \leq r_{\text{cut}}, \\ 0, & r > r_{\text{cut}}, \end{cases}$$

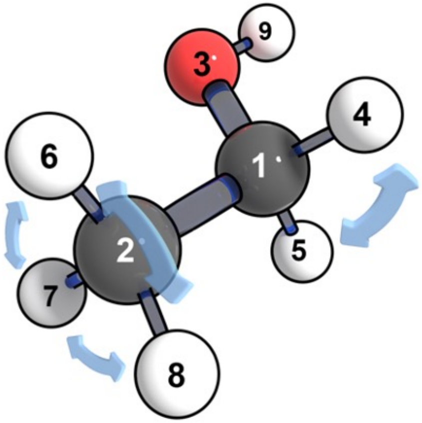
X. Gao, F. Ramezanghorbani, O. Isayev, J. S. Smith, A. E. Roitberg. *J. Chem. Inf. Model.* **2020**, *60*, 3408

KRR-CM vs KREG

KRR-CM – kernel ridge regression with Gaussian kernel and Coulomb matrix descriptor

Test set RMSEs in kcal/mol for models trained on 100 points of MD17 ethanol

KRR-CM	KREG
3.90±0.41	4.45±0.36



M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**. DOI: 10.1039/D1SC03564A

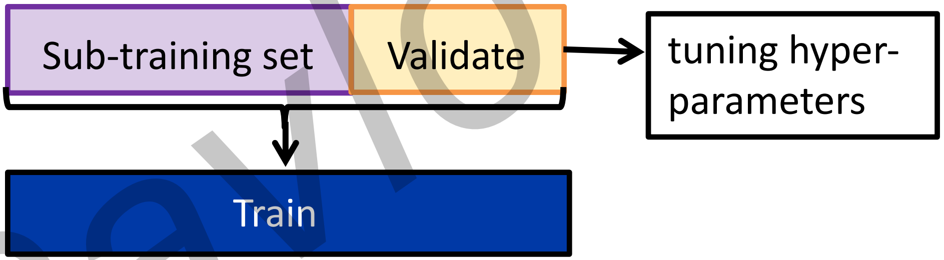
KRR-CM vs KREG

KRR-CM – kernel ridge regression with Gaussian kernel and Coulomb matrix descriptor

Test set RMSEs in kcal/mol for models trained on 100 points of MD17 ethanol

KRR-CM	KREG
3.90±0.41	4.45±0.36

What do we mean by '100 training points?' Is validation set included?
Let's use the term **'sub-training set'**!

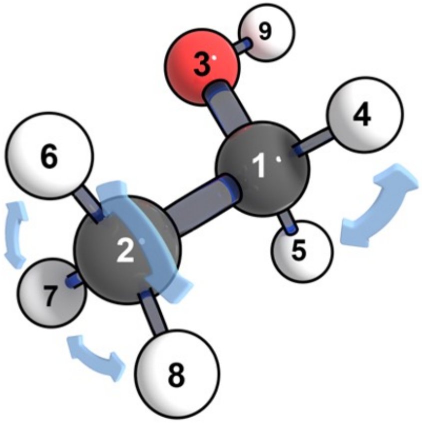


Sub-training set

Validate

tuning hyper-parameters

Train



M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**. DOI: 10.1039/D1SC03564A

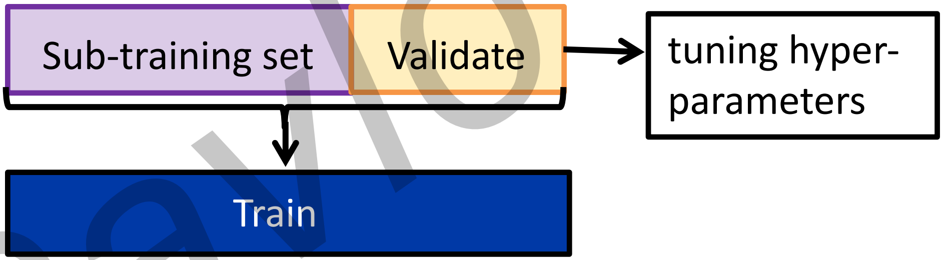
KRR-CM vs KREG

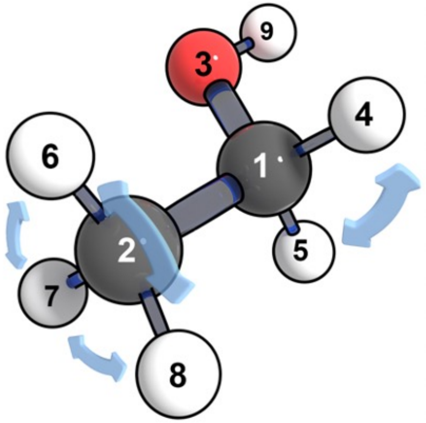
KRR-CM – kernel ridge regression with Gaussian kernel and Coulomb matrix descriptor

Test set RMSEs in kcal/mol

Number of training points	KRR-CM	KREG
100	**3.90±0.41**	**4.45±0.36**
2500	**0.70±0.02**	**0.52±0.01**

What do we mean by '100 training points?' Is validation set included? Let's use the term **'sub-training set'**!

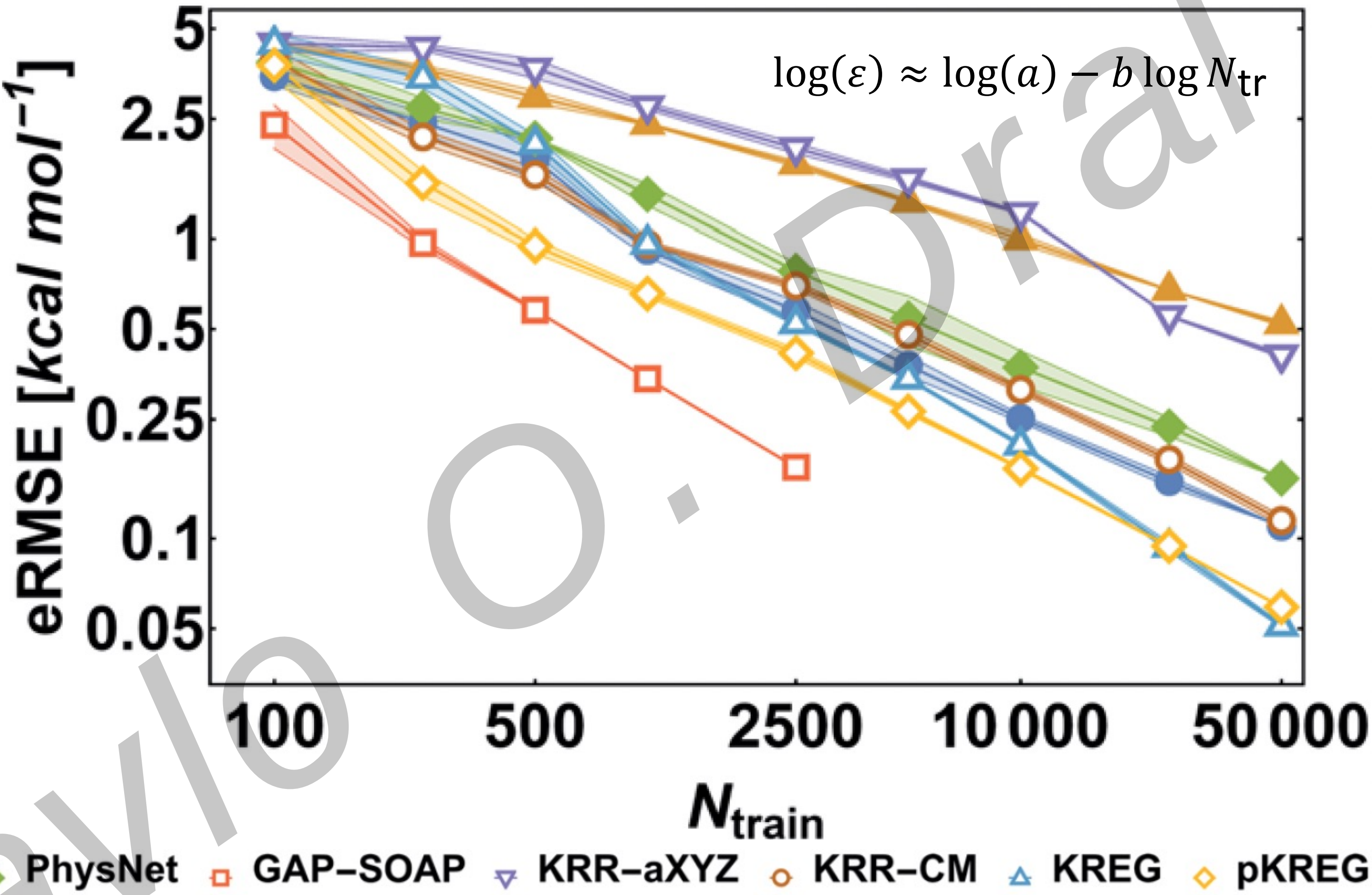

Sub-training set
Validate
tuning hyper-parameters
Train



M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413

Learning curves (energies-only)

$$\varepsilon = \varepsilon_{\mathrm{a}} + \frac{a}{N_{\mathrm{tr}}^{b}}$$



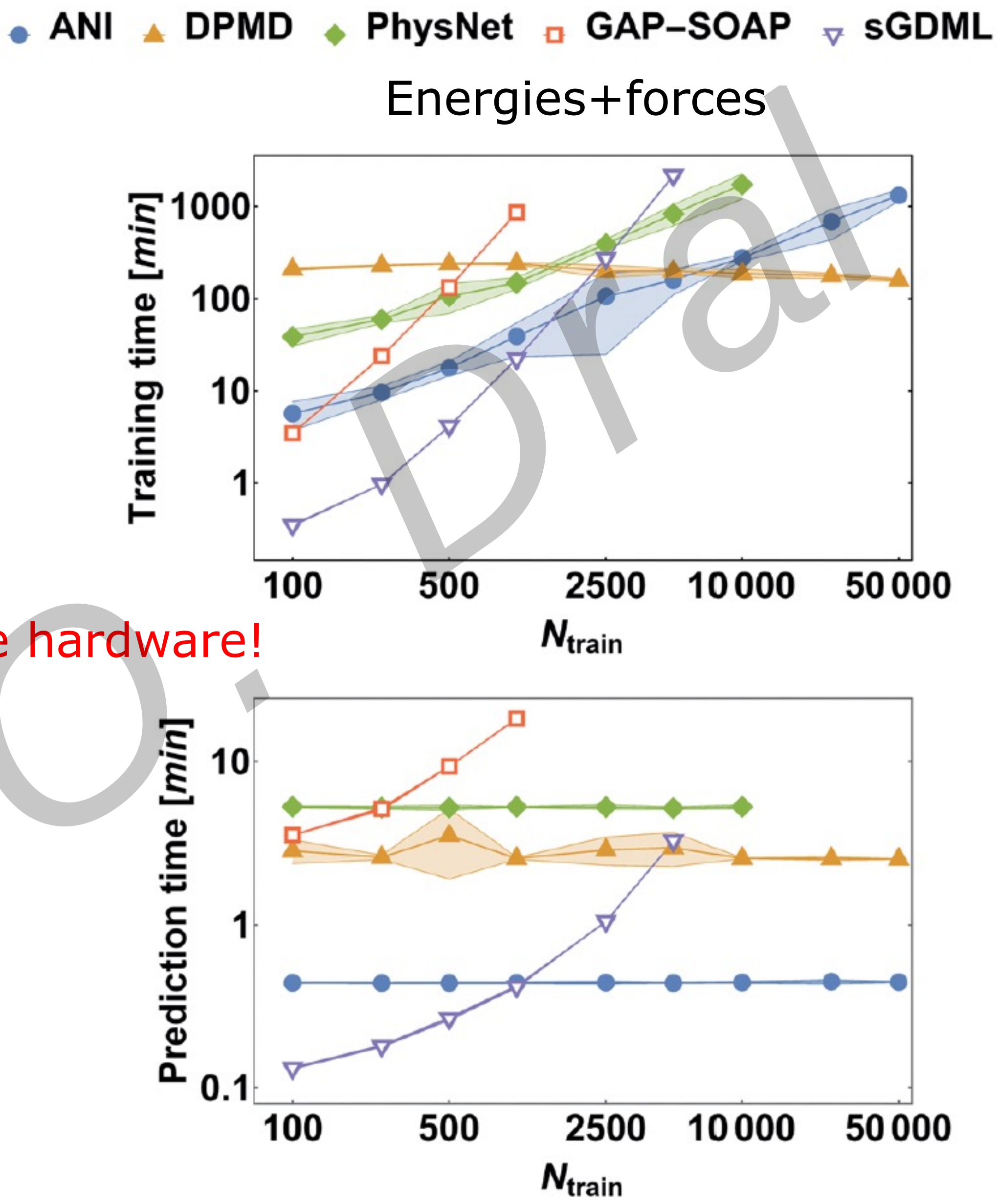
$$\log(\varepsilon) \approx \log(a) - b \log N_{\mathrm{tr}}$$

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413
P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro Jr, J. Huang, M. Barbatti. *Top. Curr. Chem.* **2021,** *379*, 27

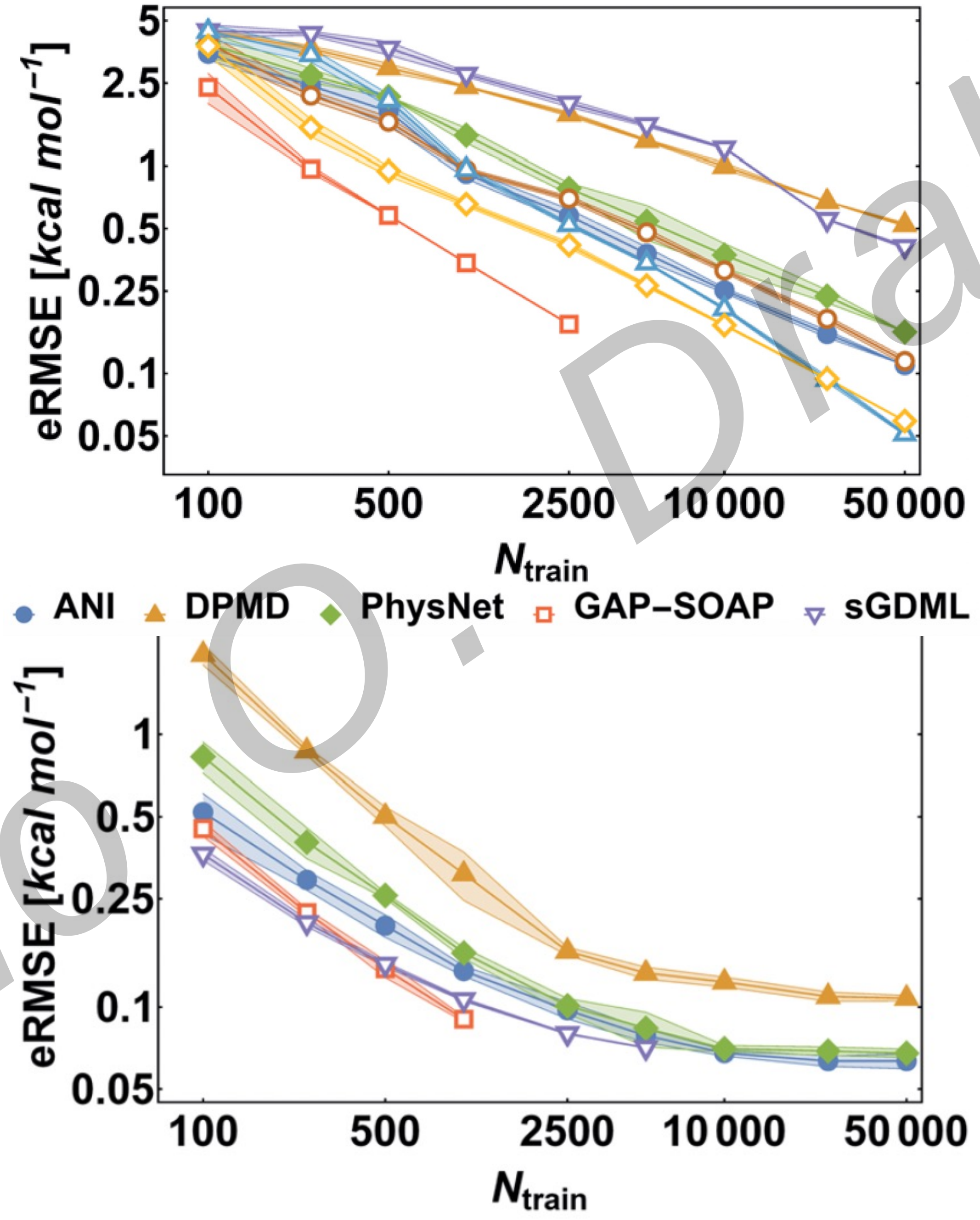
Timings!

For small training sets kernel methods (open markers) are often both more accurate and faster for training and prediction than neural networks (filled markers)

The same hardware!



M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413



Energies-only
Ethanol (MD17)

Energies+gradients
Ethanol (MD17)

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413

Machine learning potentials as force fields

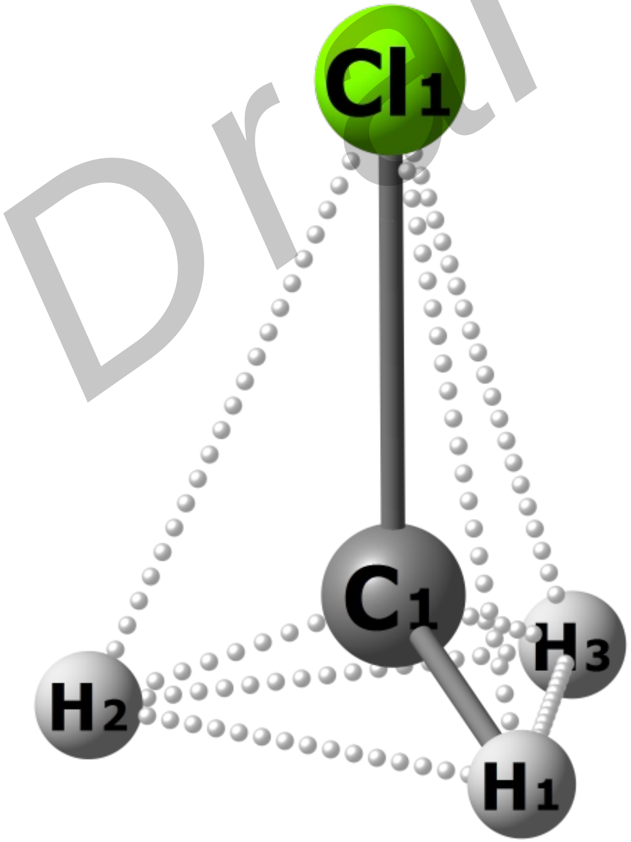
Where do we need gradients?

Gradients:

- Geometry optimization
- Molecular dynamics

$$\frac{\partial^2 \mathbf{r}_{A,d}}{\partial t^2} = -\frac{\partial E}{\partial \mathbf{r}_{A,d}} \frac{1}{M_A}$$

$$F = -\frac{\partial E}{\partial r}$$



How to calculate gradients with ML?

As first-order derivatives, e.g. in case of KRR:

$$\frac{\partial f(\mathbf{x}_i)}{\partial x_{i,d}} = \sum_{j=1}^{N_{\text{tr}}} \alpha_j \frac{\partial k(\mathbf{x}_i, \mathbf{x}_j)}{\partial x_{i,d}}$$

$$\frac{\partial k(x(\mathbf{M}_i), \mathbf{x}_j)}{\partial M_{i,at}} = \sum_{d=1}^{N_d} \frac{\partial k(\mathbf{x}_i, \mathbf{x}_j)}{\partial x_{i,d}} \frac{\partial x_{i,d}}{\partial M_{i,at}}$$

$$\frac{\partial^2 k(\mathbf{x}_i, \mathbf{x}_j)}{\partial M_{i,at} \partial M_{j,bu}} = \sum_{d=1}^{N_d} \sum_{e=1}^{N_d} \frac{\partial^2 k(\mathbf{x}_i, \mathbf{x}_j)}{\partial x_{i,d} \partial x_{j,e}} \frac{\partial x_{i,d}}{\partial M_{i,at}} \frac{\partial x_{j,e}}{\partial M_{j,bu}}$$

P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro Jr, J. Huang, M. Barbatti.
Top. Curr. Chem. **2021,** *379*, 27

How KREG calculates gradients?

In case of KREG:

$$\frac{\partial k\left(x(\mathbf{M}_i),\mathbf{x}_j\right)}{\partial M_{i,at}} = \sum_{d=1}^{N_d} \frac{\partial k\left(\mathbf{x}_i,\mathbf{x}_j\right)}{\partial x_{i,d}} \frac{\partial x_{i,d}}{\partial M_{i,at}}$$

$$\mathbf{x}^{\mathrm{T}} = \left[\cdots \quad \frac{r_{a,b\neq a}^{\text{ref}}}{r_{a,b\neq a}} \quad \cdots\right]$$

$$r_{a,b} = \sqrt{\left(\sum_{s=1}^{3}(M_{as}-M_{bs})^2\right)}$$

$$k\left(\mathbf{x}_i,\mathbf{x}_j\right) = \exp\left(-\frac{1}{2\sigma^2}\sum_{s}^{N_x}\left(x_{i,s}-x_{j,s}\right)^2\right)$$

$$\frac{\partial k\left(\mathbf{x}_i,\mathbf{x}_j\right)}{\partial x_{i,d}} = -\frac{1}{2\sigma^2}2\left(x_{i,d}-x_{j,d}\right)\exp\left(-\frac{1}{2\sigma^2}\sum_{s}^{N_x}\left(x_{i,s}-x_{j,s}\right)^2\right)$$

$$\frac{\partial x_d(M_{at})}{\partial M_{at}} = x_d\frac{1}{r_{a,b\neq a}^2}(M_{bt}-M_{at})$$

P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro Jr, J. Huang, M. Barbatti. *Top. Curr. Chem.* **2021,** *379*, 27

How to learn gradients with ML?

Gradients also provide valuable additional information for training ML models.

For N molecules, there are N energies, but $3N_{at}$ × N gradients!

Popular approach especially for NN MLPs:

$$f(\mathbf{x}_i;\boldsymbol{\beta}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots$$

$$\frac{\partial f(\mathbf{x}_i;\boldsymbol{\beta})}{\partial x_{i,d}} = \frac{\partial(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots)}{\partial x_{i,d}}$$

$$minimize\ Loss\ function = error\ for\ energies + error\ for\ gradients$$

$$L = L_{val} + w_{grxyz} L_{grxyz} \qquad L = \sqrt{L_{val} L_{grxyz}}$$

sGDML

sGDML – a symmetrized variant of the gradient-domain machine learning (GDML). It uses unnormalized permutationally invariant, Matérn kernel.

$$\mathbf{x}^{\mathrm{T}} = \left[\cdots \quad \frac{1}{r_{a,b}} \quad \cdots\right]$$

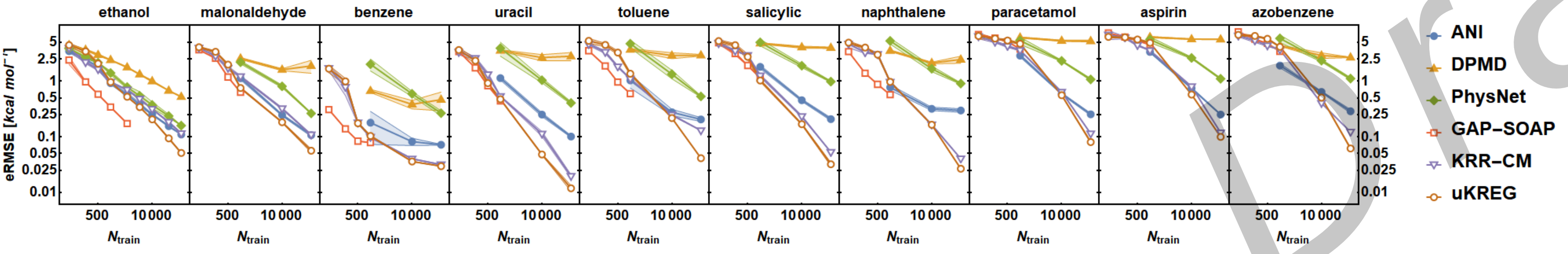
$$\frac{\partial f(\mathbf{x})}{\partial M_{at}} = \sum_{j=1}^{N_{\mathrm{tr}}} \sum_{b=1}^{N_{\mathrm{at}}} \sum_{u=1}^{3} \alpha_{j,bu} \frac{\partial^2 k(\mathbf{x}, \mathbf{x}_j)}{\partial M_{at} \partial M_{j,bu}}$$

$$\begin{pmatrix} \frac{\partial^2 k(\mathbf{x}_1, \mathbf{x}_1)}{\partial M_{1,11} \partial M_{1,11}} + \lambda & \cdots & \frac{\partial^2 k\left(\mathbf{x}_1, \mathbf{x}_{N_{\mathrm{tr}}}\right)}{\partial M_{1,11} \partial M_{N_{\mathrm{tr}}, N_{\mathrm{at}} 3}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 k\left(\mathbf{x}_{N_{\mathrm{tr}}}, \mathbf{x}_1\right)}{\partial M_{N_{\mathrm{tr}}, N_{\mathrm{at}} 3} \partial M_{1,11}} & \cdots & \frac{\partial^2 k\left(\mathbf{x}_{N_{\mathrm{tr}}}, \mathbf{x}_{N_{\mathrm{tr}}}\right)}{\partial M_{N_{\mathrm{tr}}, N_{\mathrm{at}} 3} \partial M_{N_{\mathrm{tr}}, N_{\mathrm{at}} 3}} + \lambda \end{pmatrix} \begin{pmatrix} \alpha_{1,11} \\ \vdots \\ \alpha_{N_{\mathrm{tr}}, N_{\mathrm{at}} 3} \end{pmatrix} = \begin{pmatrix} \frac{\partial y_1}{\partial M_{1,11}} \\ \vdots \\ \frac{\partial y_{N_{\mathrm{tr}}}}{\partial M_{N_{\mathrm{tr}}, N_{\mathrm{at}} 3}} \end{pmatrix}$$

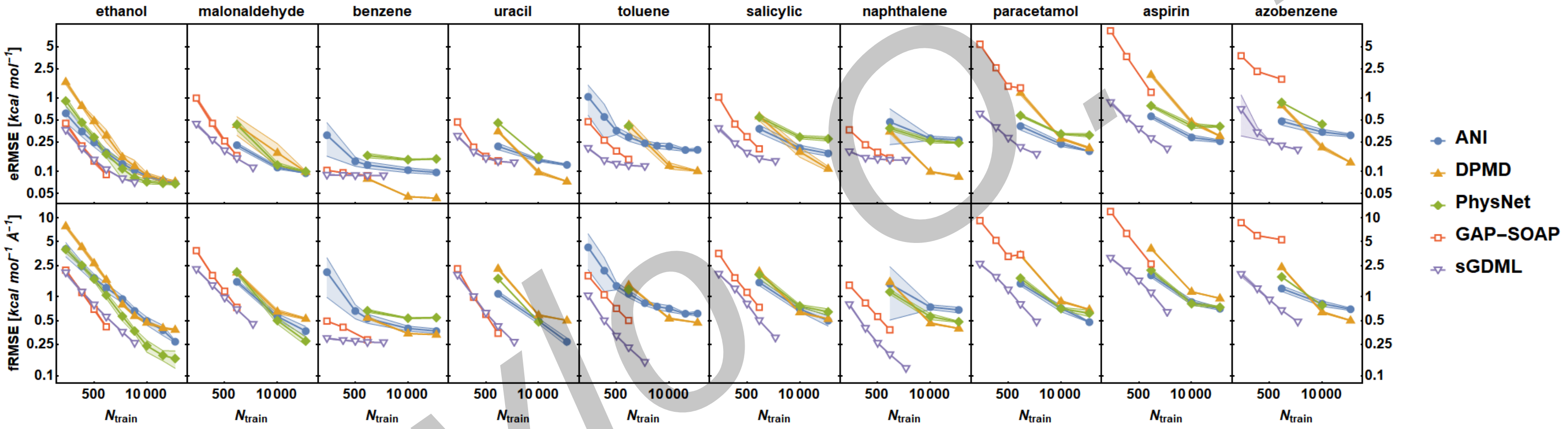
$$f(\mathbf{x}) = const + \sum_{j=1}^{N_{\mathrm{tr}}} \sum_{b=1}^{N_{\mathrm{at}}} \sum_{u=1}^{3} \alpha_{j,bu} \frac{\partial k(\mathbf{x}, \mathbf{x}_j)}{\partial M_{j,bu}}$$

S. Chmiela, H. E. Sauceda, K.-R. Müller, A. Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018,** *9*, 3887

Benchmark on MD17



Energies-only

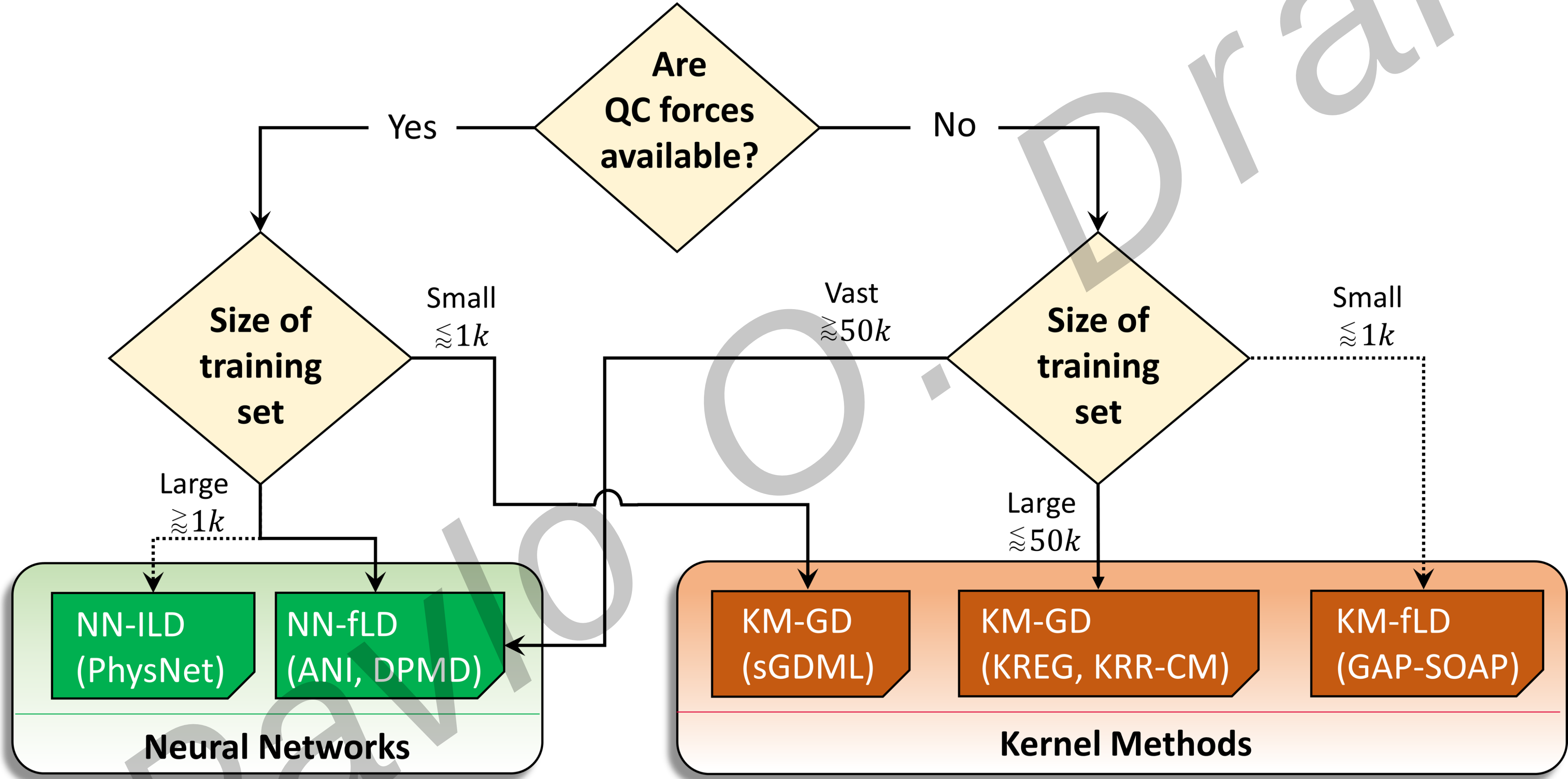


Energies + forces

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413

MD17: S. Chmiela, H. E. Sauceda, K.-R. Müller, A. Tkatchenko. *Nat. Commun.* **2018,** *9*, 3887; S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, K.-R. Müller. *Sci. Adv.* **2017,** *3*, e1603015.

Choosing the 'Right' ML Potential



Are QC forces available?

Yes

No

Size of training set

Small $\lesssim 1k$

Large $\gtrsim 1k$

Vast $\gtrsim 50k$

Size of training set

Small $\lesssim 1k$

Large $\lesssim 50k$

NN-lLD (PhysNet)

NN-fLD (ANI, DPMD)

Neural Networks

KM-GD (sGDML)

KM-GD (KREG, KRR-CM)

KM-fLD (GAP-SOAP)

Kernel Methods

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413

Platform for Benchmarking

It is an open project as the development of ML potentials is not a static field. You are welcome to use the same protocols to update our benchmark results http://mlatom.com/mlpbenchmark1/

The minimum input for MLatom – can be used as a platform for testing ML potentials:

```
learningCurve
XYZfile=ethanol.xyz
Yfile=ethanol_energies.dat
lcNtrains=100,200,300,400,500
lcNrepeats=30,30,30,30,10
MLmodelType=[your model]
```

MLatom.com

Fuchun Ge

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413

Good ML method for quantum dynamics?

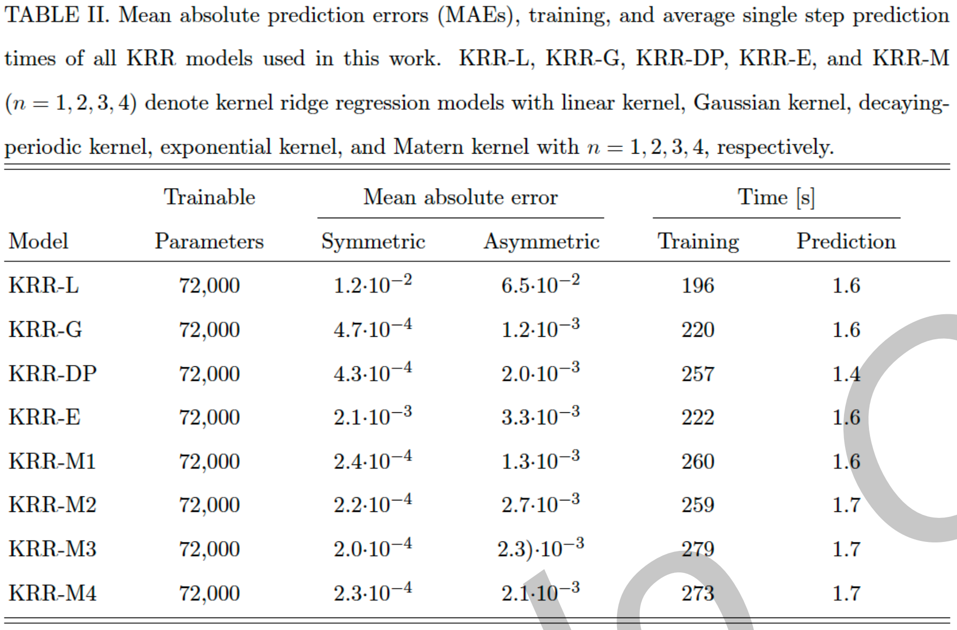


TABLE II. Mean absolute prediction errors (MAEs), training, and average single step prediction times of all KRR models used in this work. KRR-L, KRR-G, KRR-DP, KRR-E, and KRR-M ($n = 1, 2, 3, 4$) denote kernel ridge regression models with linear kernel, Gaussian kernel, decaying-periodic kernel, exponential kernel, and Matern kernel with $n = 1, 2, 3, 4$, respectively.

	Trainable	Mean absolute error		Time [s]	
Model	Parameters	Symmetric	Asymmetric	Training	Prediction
KRR-L	72,000	$1.2{\cdot}10^{-2}$	$6.5{\cdot}10^{-2}$	196	1.6
KRR-G	72,000	$4.7{\cdot}10^{-4}$	$1.2{\cdot}10^{-3}$	220	1.6
KRR-DP	72,000	$4.3{\cdot}10^{-4}$	$2.0{\cdot}10^{-3}$	257	1.4
KRR-E	72,000	$2.1{\cdot}10^{-3}$	$3.3{\cdot}10^{-3}$	222	1.6
KRR-M1	72,000	$2.4{\cdot}10^{-4}$	$1.3{\cdot}10^{-3}$	260	1.6
KRR-M2	72,000	$2.2{\cdot}10^{-4}$	$2.7{\cdot}10^{-3}$	259	1.7
KRR-M3	72,000	$2.0{\cdot}10^{-4}$	$2.3){\cdot}10^{-3}$	279	1.7
KRR-M4	72,000	$2.3{\cdot}10^{-4}$	$2.1{\cdot}10^{-3}$	273	1.7

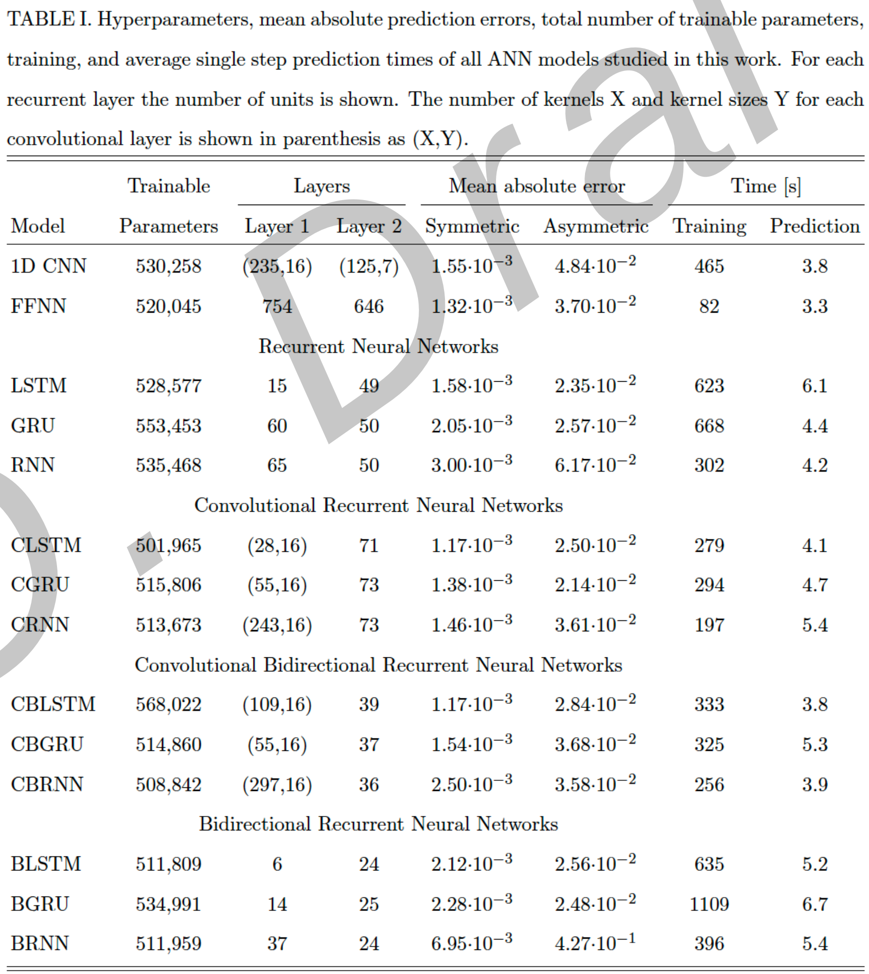


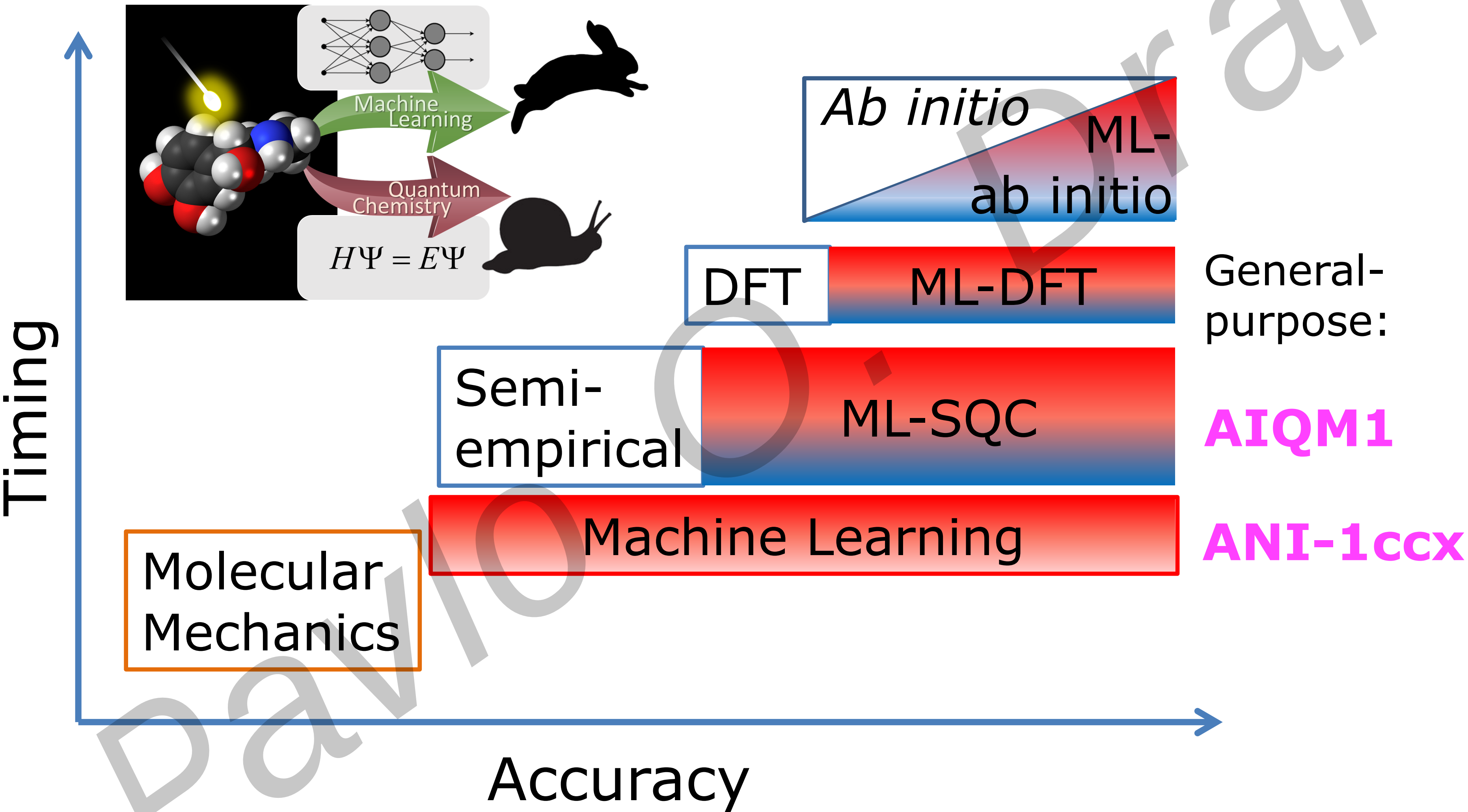
TABLE I. Hyperparameters, mean absolute prediction errors, total number of trainable parameters, training, and average single step prediction times of all ANN models studied in this work. For each recurrent layer the number of units is shown. The number of kernels X and kernel sizes Y for each convolutional layer is shown in parenthesis as (X,Y).

	Trainable	Layers		Mean absolute error		Time [s]	
Model	Parameters	Layer 1	Layer 2	Symmetric	Asymmetric	Training	Prediction
1D CNN	530,258	(235,16)	(125,7)	$1.55{\cdot}10^{-3}$	$4.84{\cdot}10^{-2}$	465	3.8
FFNN	520,045	754	646	$1.32{\cdot}10^{-3}$	$3.70{\cdot}10^{-2}$	82	3.3
Recurrent Neural Networks							
LSTM	528,577	15	49	$1.58{\cdot}10^{-3}$	$2.35{\cdot}10^{-2}$	623	6.1
GRU	553,453	60	50	$2.05{\cdot}10^{-3}$	$2.57{\cdot}10^{-2}$	668	4.4
RNN	535,468	65	50	$3.00{\cdot}10^{-3}$	$6.17{\cdot}10^{-2}$	302	4.2
Convolutional Recurrent Neural Networks							
CLSTM	501,965	(28,16)	71	$1.17{\cdot}10^{-3}$	$2.50{\cdot}10^{-2}$	279	4.1
CGRU	515,806	(55,16)	73	$1.38{\cdot}10^{-3}$	$2.14{\cdot}10^{-2}$	294	4.7
CRNN	513,673	(243,16)	73	$1.46{\cdot}10^{-3}$	$3.61{\cdot}10^{-2}$	197	5.4
Convolutional Bidirectional Recurrent Neural Networks							
CBLSTM	568,022	(109,16)	39	$1.17{\cdot}10^{-3}$	$2.84{\cdot}10^{-2}$	333	3.8
CBGRU	514,860	(55,16)	37	$1.54{\cdot}10^{-3}$	$3.68{\cdot}10^{-2}$	325	5.3
CBRNN	508,842	(297,16)	36	$2.50{\cdot}10^{-3}$	$3.58{\cdot}10^{-2}$	256	3.9
Bidirectional Recurrent Neural Networks							
BLSTM	511,809	6	24	$2.12{\cdot}10^{-3}$	$2.56{\cdot}10^{-2}$	635	5.2
BGRU	534,991	14	25	$2.28{\cdot}10^{-3}$	$2.48{\cdot}10^{-2}$	1109	6.7
BRNN	511,959	37	24	$6.95{\cdot}10^{-3}$	$4.27{\cdot}10^{-1}$	396	5.4

L. E. H. Rodrıguez, A. Ullah, K. J. R. Espinosa, P. O. Dral, A. A. Kananenka. A comparative study of different machine learning methods for dissipative quantum dynamics. **2022**, *submitted*. *arXiv*: https://arxiv.org/abs/2207.02417.

Machine learning

P. O. Dral, M. Barbatti, *Nat. Rev. Chem.* **2021**, *5*, 388



Timing

Accuracy

Ab initio

ML-ab initio

DFT

ML-DFT

Semi-empirical

ML-SQC

Machine Learning

Molecular Mechanics

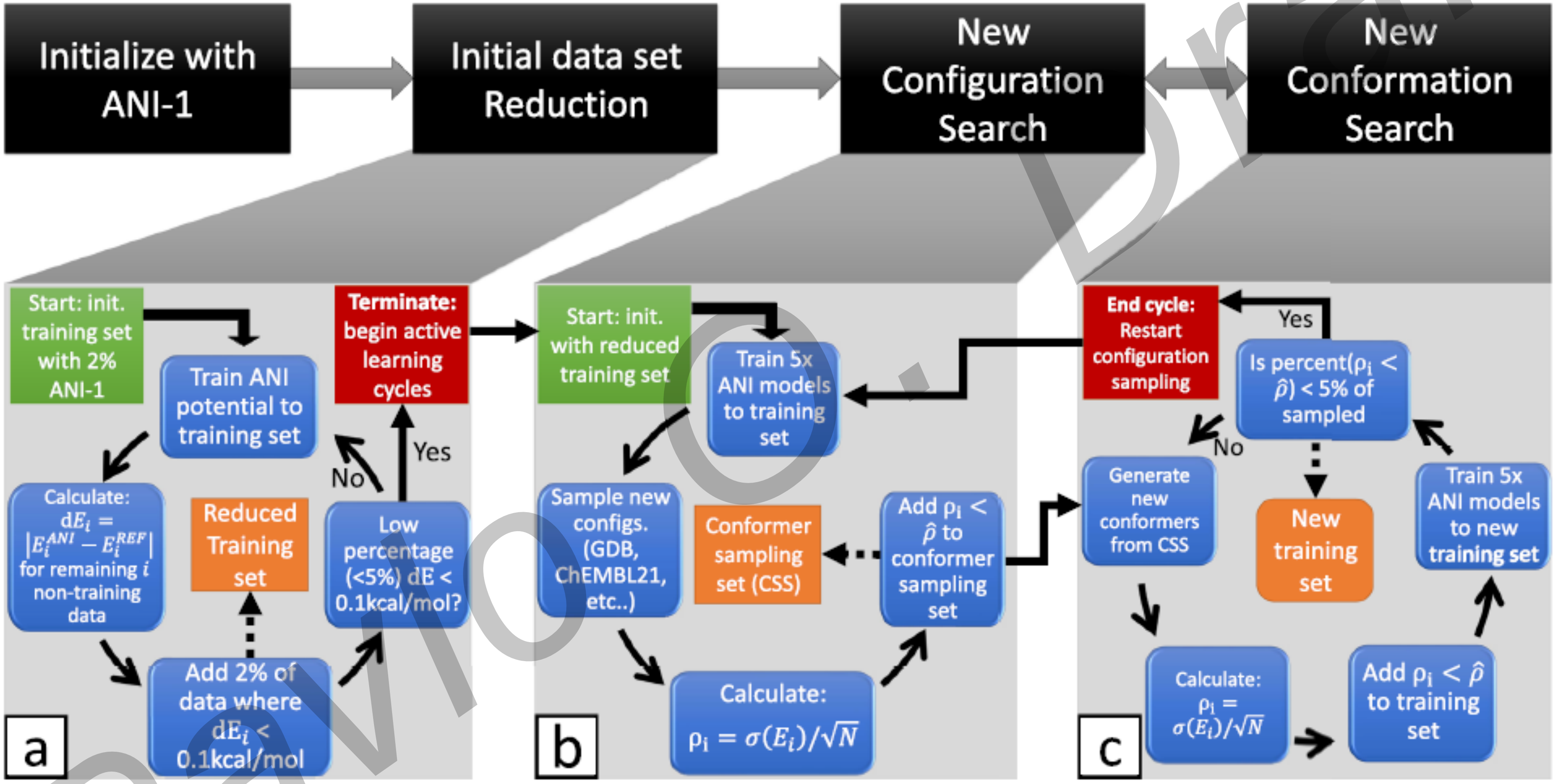
General-purpose:

AIQM1

ANI-1ccx

P. O. Dral, *J. Phys. Chem. Lett.* **2020**, *11*, 2336

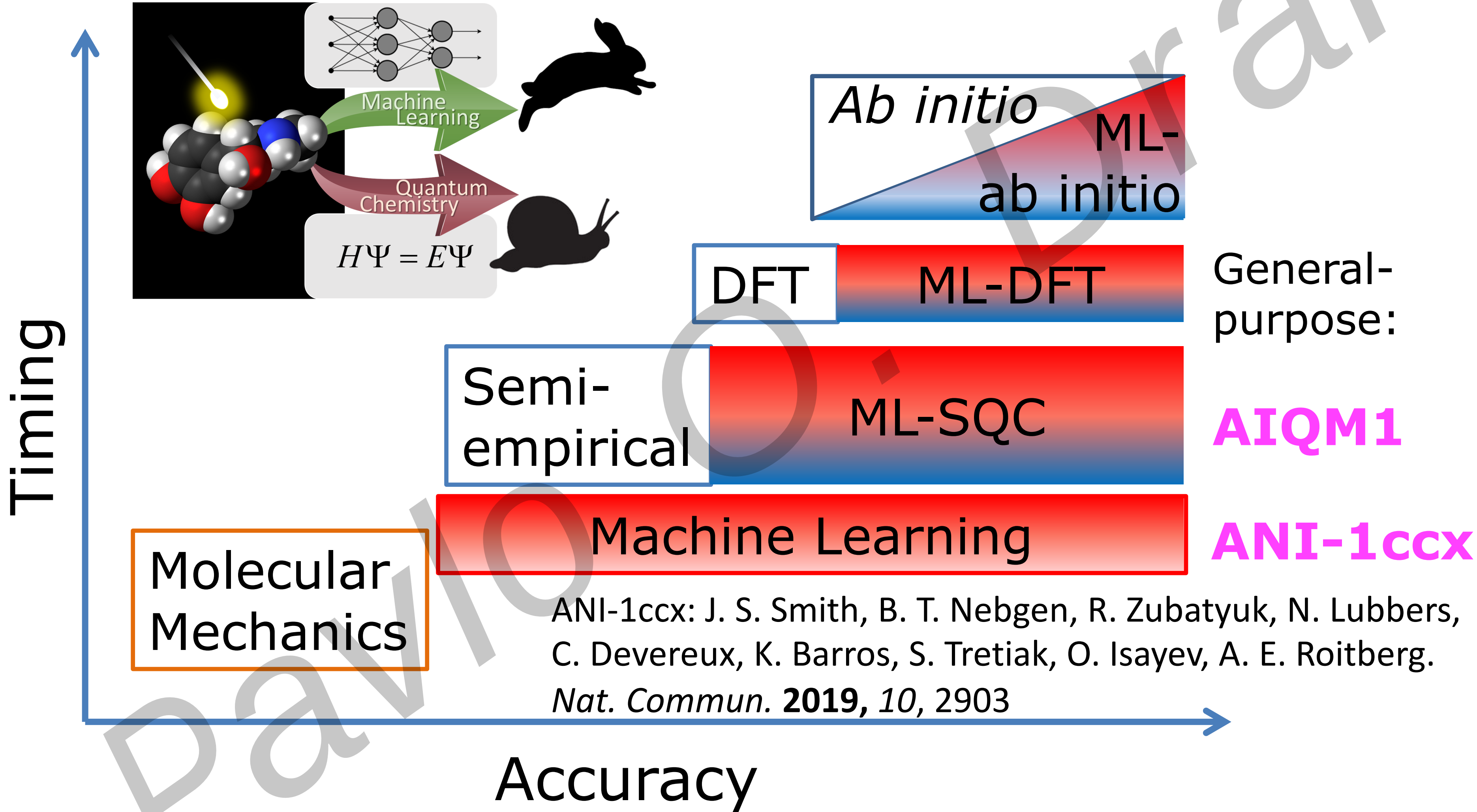
Active learning for chemical space



J. S. Smith, O. Isayev, A. E. Roitberg, *Chem. Sci.* **2017**, *8*, 3192

Machine learning

P. O. Dral, M. Barbatti, *Nat. Rev. Chem.* **2021,** *5*, 388



Ab initio

ML-ab initio

DFT

ML-DFT

General-purpose:

Semi-empirical

ML-SQC

AIQM1

Machine Learning

ANI-1ccx

Molecular Mechanics

ANI-1ccx: J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, A. E. Roitberg. *Nat. Commun.* **2019,** *10*, 2903

Timing

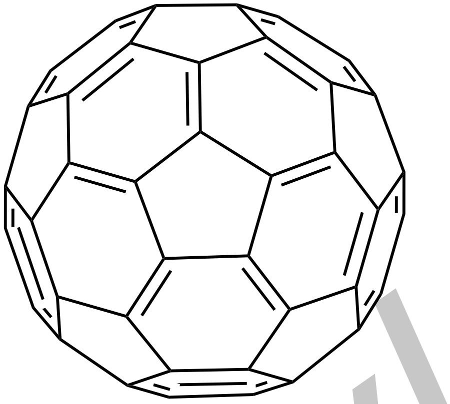
Accuracy

P. O. Dral, *J. Phys. Chem. Lett.* **2020**, *11*, 2336

Accuracy vs transferability vs cost

Artificial Intelligence-Enhanced Quantum Chemical Method 1

AIQM1

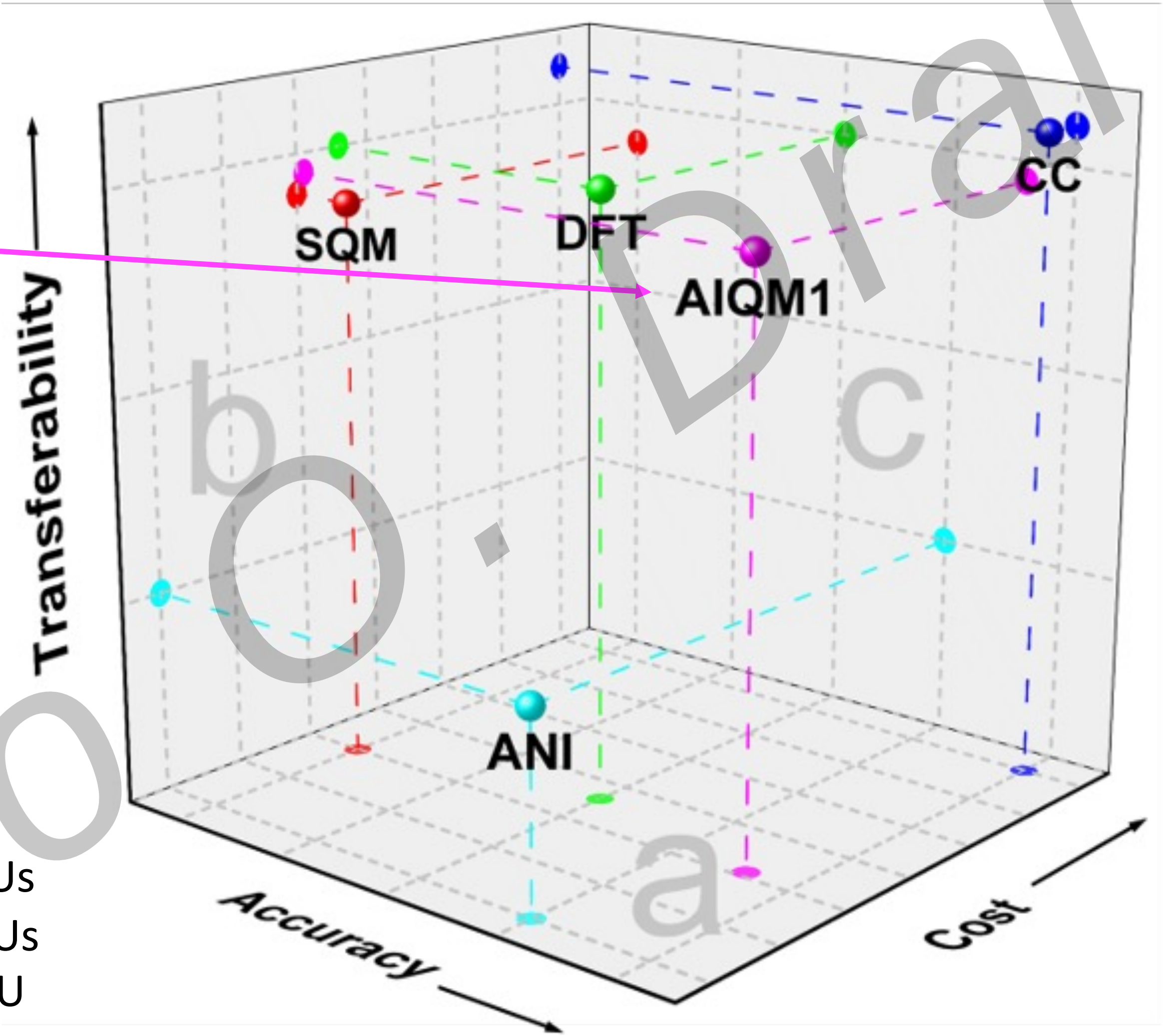


C_{60}

CCSD(T)*/CBS SP: 69 h /15 CPUs

DFT opt: 30 min /32 CPUs

AIQM1 opt: 14 sec / 1 CPU



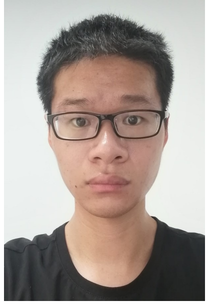
P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral. Artificial Intelligence-Enhanced Quantum Chemical Method with Broad Applicability. *Nat. Commun.* **2021**, *12*, 7022.

Wei Wu

Olexandr
Isayev

Artificial Intelligence-Enhanced Quantum Chemical Method 1

AIQM1



Peikun Zheng

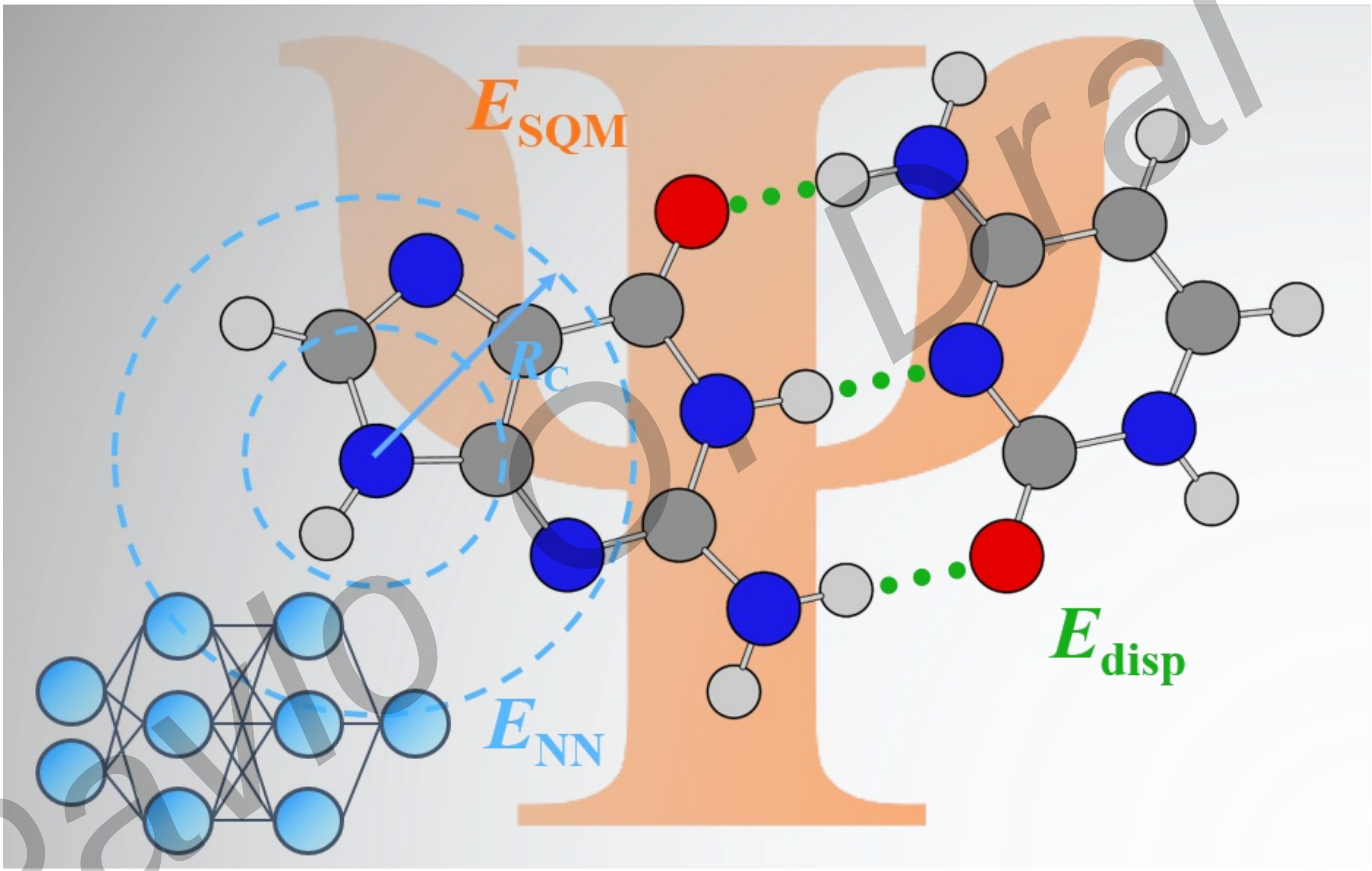
Roman Zubatyuk

accurate and highly transferable ML+QC approach

Code available at: http://mlatom.com/aiqm1/
(online calculations on a cloud)

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral. Artificial Intelligence-Enhanced Quantum Chemical Method with Broad Applicability. *Nat. Commun.* **2021**, *12*, 7022.

AIQM1: Semiempirical QM + NN + D4



P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

Orthogonalization- and Dispersion-Corrected Methods 2 & 3 (ODM2 & ODM3 ODMx)

P. O. Dral, X. Wu, W. Thiel, *J. Chem. Theory Comput.* **2019**, *15*, 1743

ODMx Methods

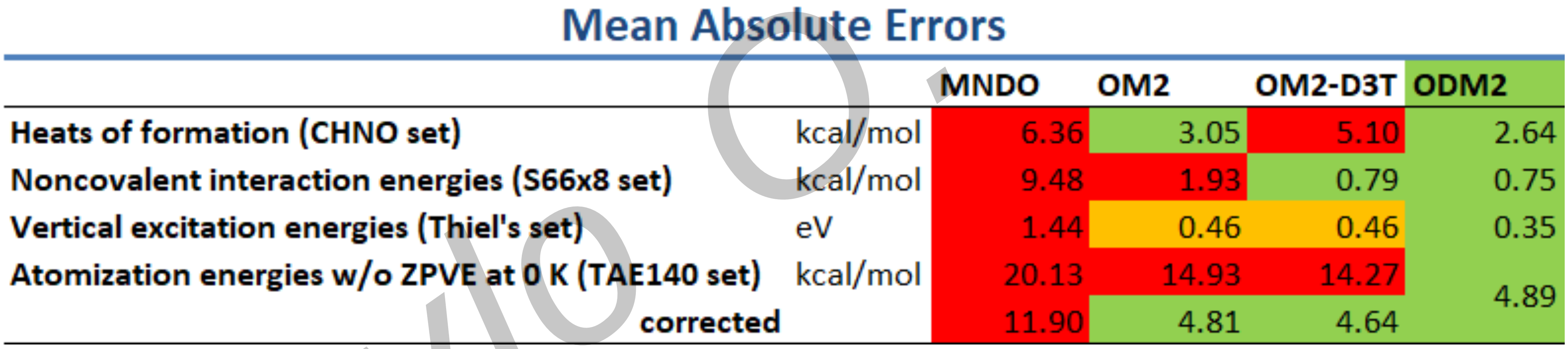
State-of-the-art ODMx methods solve many inconsistencies of NDDO-based semiempirical QM methods

- Include orthogonalization effects
- Include penetration integrals
- Include core-valence interactions
- Include dispersion corrections as their integral part
- Parametrized on a balanced set of ground-state, excited-state and noncovalent properties
- Calculate heats of formation in *ab initio* manner with explicit ZPVE and thermal enthalpic corrections

P. O. Dral, X. Wu, W. Thiel, *J. Chem. Theory Comput.* **2019**, *15*, 1743

ODMx Methods

ODMx methods are the most consistent and robust SQC methods to date



Mean Absolute Errors

		MNDO	OM2	OM2-D3T	ODM2
Heats of formation (CHNO set)	kcal/mol	6.36	3.05	5.10	2.64
Noncovalent interaction energies (S66x8 set)	kcal/mol	9.48	1.93	0.79	0.75
Vertical excitation energies (Thiel's set)	eV	1.44	0.46	0.46	0.35
Atomization energies w/o ZPVE at 0 K (TAE140 set)	kcal/mol	20.13	14.93	14.27	4.89
corrected		11.90	4.81	4.64	

P. O. Dral, X. Wu, W. Thiel, *J. Chem. Theory Comput.* **2019**, *15*, 1743

Excited States

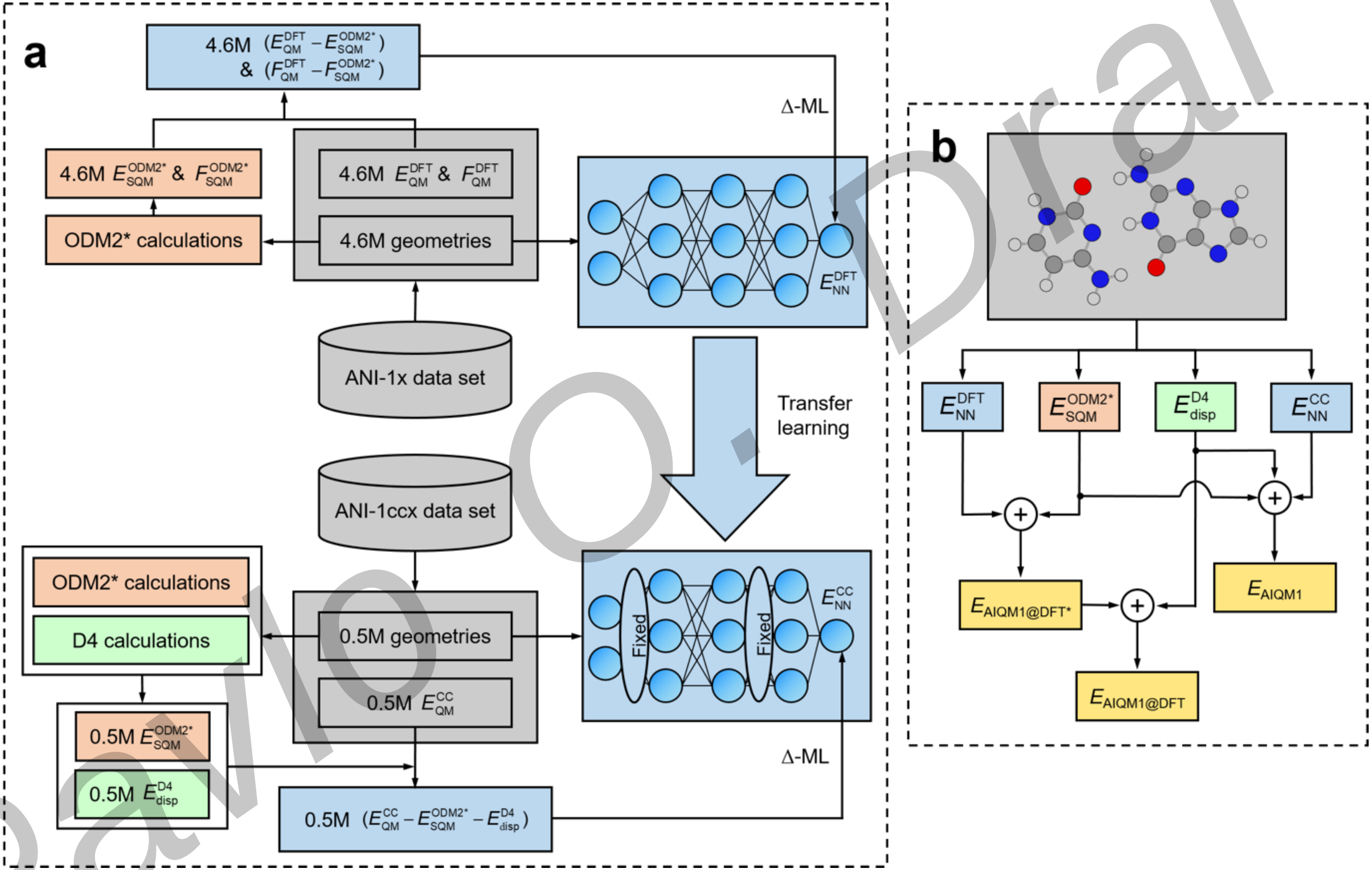
Errors in vertical excitation energies

Benchmarking on Thiel's 2008 set:

Method	Mean absolute error, eV
TD B3LYP/TZVP[1]	**0.33**
AM1/MRCI[2]	1.23
INDO/S[2]	0.56
OM2/MRCI[1]	0.47
OM3/MRCI[1]	0.42
ODM2/MRCI[3]	**0.35**
ODM3/MRCI[3]	**0.33**

[1] D. Tuna, Y. Lu, A. Koslowski, W. Thiel, *J. Chem. Theory Comput.* **2016**, *12*, 4400
[2] M. R. Silva-Junior, W. Thiel, *J. Chem. Theory Comput.* **2010**, *6*, 1546
[3] P. O. Dral, X. Wu, W. Thiel, *J. Chem. Theory Comput.* **2019**, *15*, 1743

Training



P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

Multi-fidelity data: Δ-learning

R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld,
J. Chem. Theory Comput. **2015**, *11*, 2087

H_2 dissociation curve

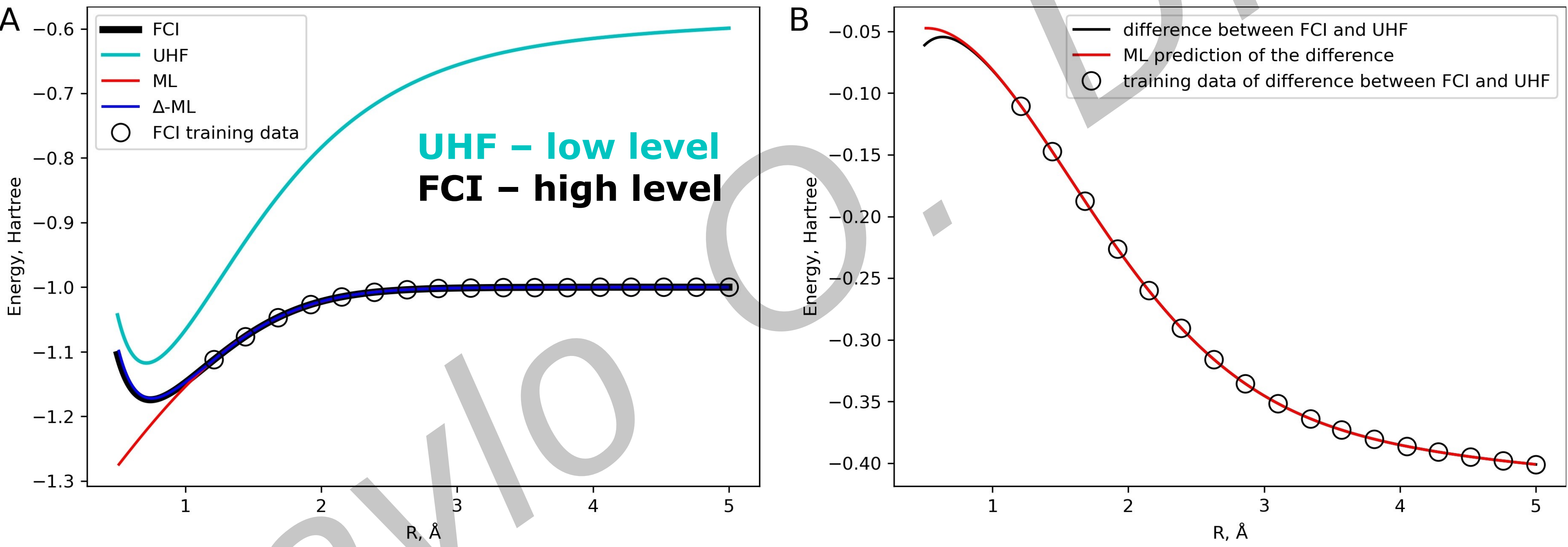
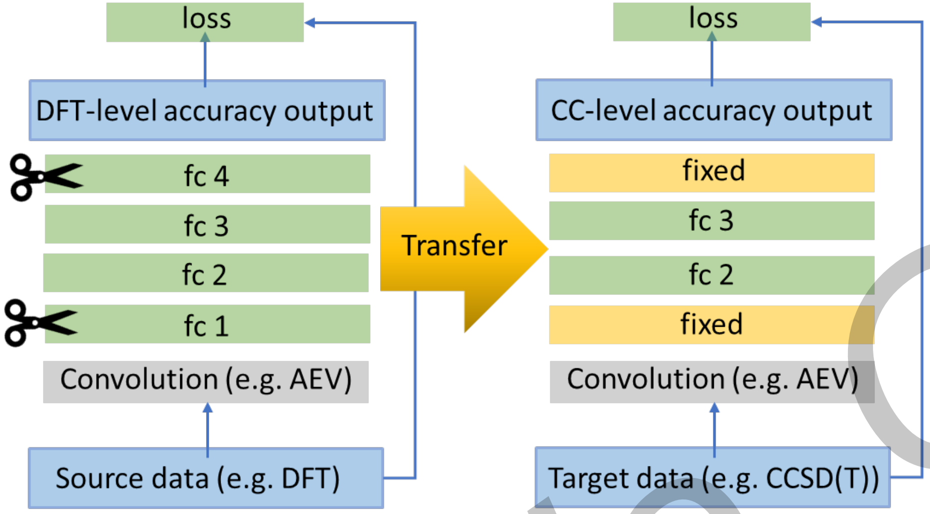


Figure: Pavlo O. Dral, Tetiana Zubatiuk, Bao-Xin Xue, Learning from multiple quantum chemical methods: Δ-learning, transfer learning, co-kriging, and beyond. In *Quantum Chemistry in the Age of Machine Learning*, Pavlo O. Dral, Ed. Elsevier: 2022, in press. Paperback ISBN: 9780323900492

Multi-fidelity data: Transfer learning



loss

DFT-level accuracy output

fc 4

fc 3

fc 2

fc 1

Convolution (e.g. AEV)

Source data (e.g. DFT)

Transfer

loss

CC-level accuracy output

fixed

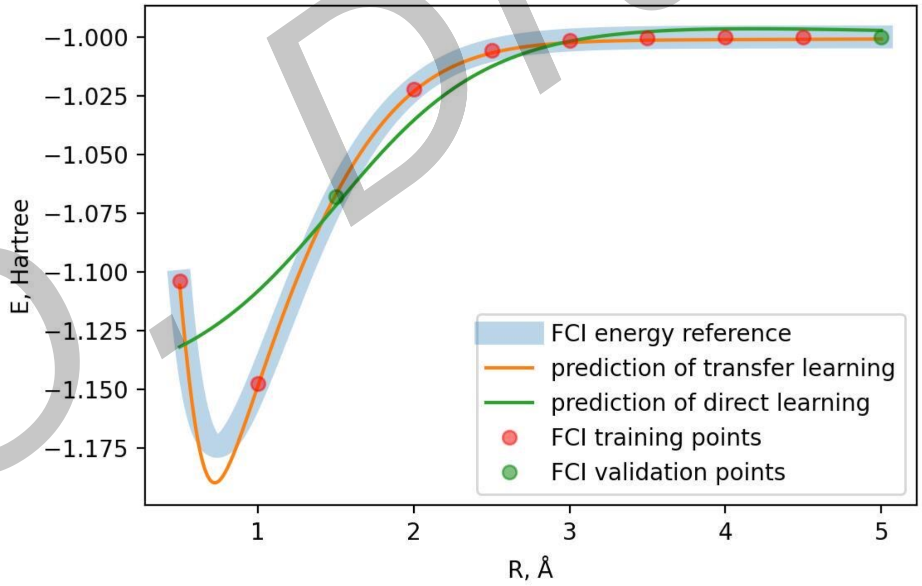
fc 3

fc 2

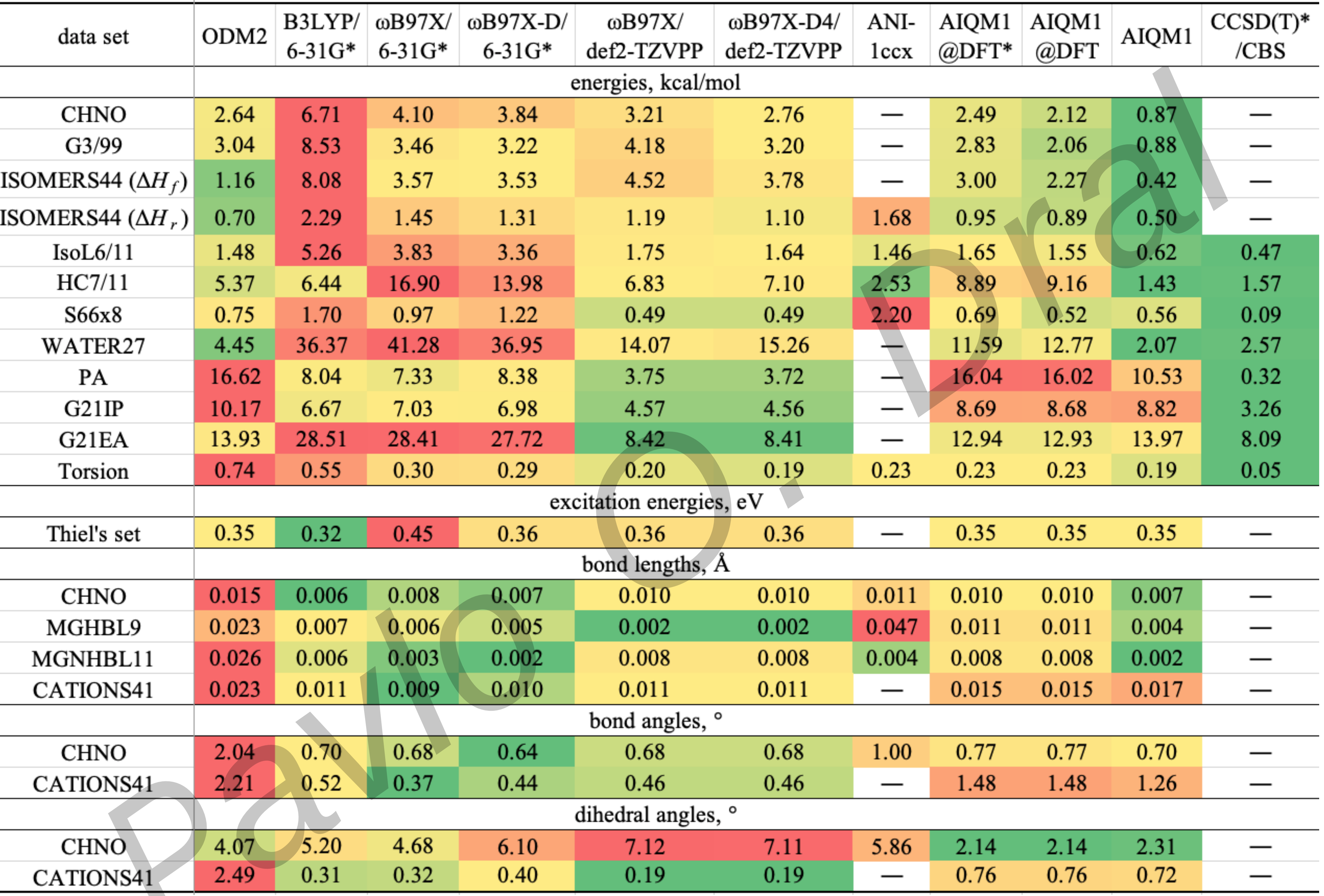
fixed

Convolution (e.g. AEV)

Target data (e.g. CCSD(T))

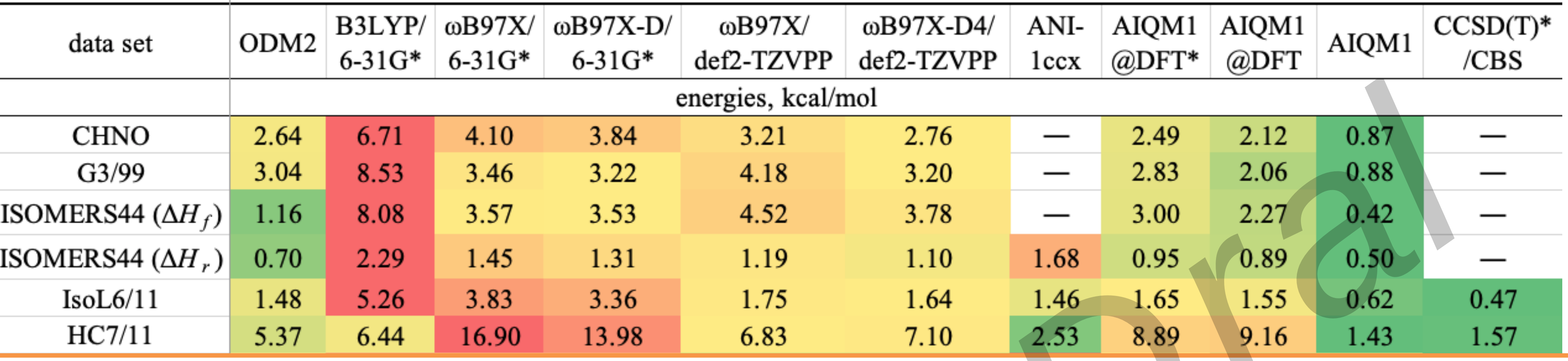


Figures: Pavlo O. Dral, Tetiana Zubatiuk, Bao-Xin Xue, Learning from multiple quantum chemical methods: Δ-learning, transfer learning, co-kriging, and beyond. In *Quantum Chemistry in the Age of Machine Learning*, Pavlo O. Dral, Ed. Elsevier: 2022, in press. Paperback ISBN: 9780323900492



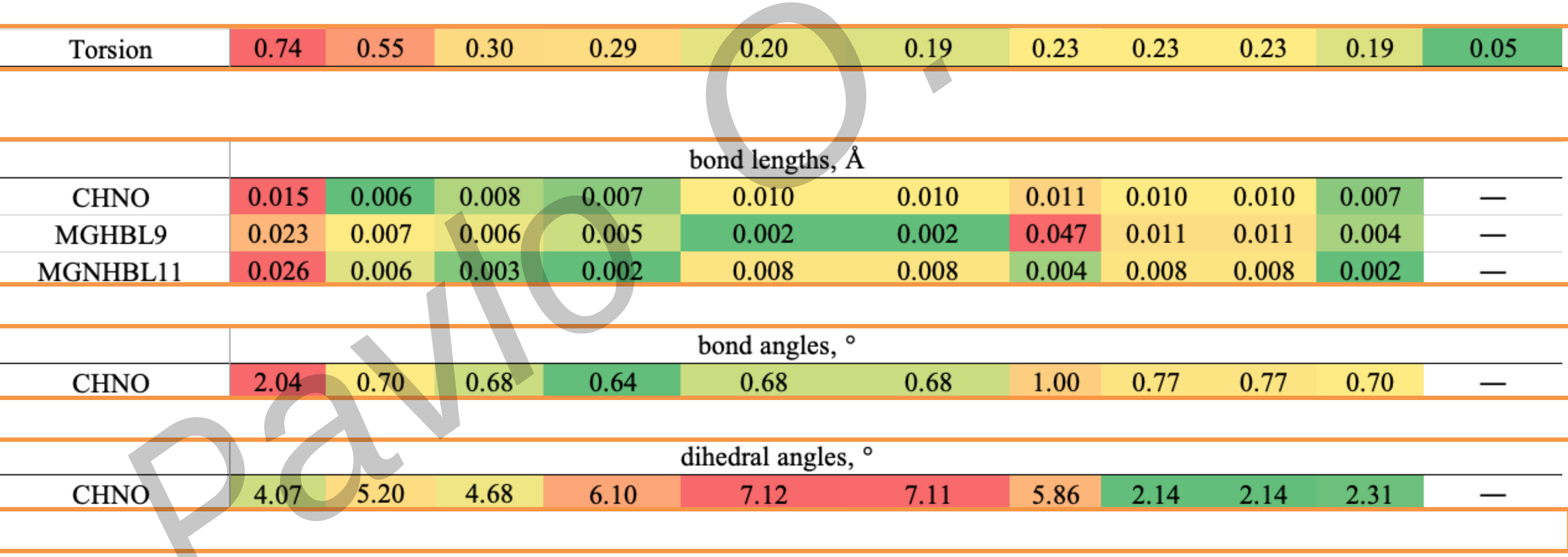
data set	ODM2	B3LYP/ 6-31G*	ωB97X/ 6-31G*	ωB97X-D/ 6-31G*	ωB97X/ def2-TZVPP	ωB97X-D4/ def2-TZVPP	ANI-1ccx	AIQM1 @DFT*	AIQM1 @DFT	AIQM1	CCSD(T)*/CBS
	energies, kcal/mol										
CHNO	2.64	6.71	4.10	3.84	3.21	2.76	—	2.49	2.12	0.87	—
G3/99	3.04	8.53	3.46	3.22	4.18	3.20	—	2.83	2.06	0.88	—
ISOMERS44 (ΔH_f)	1.16	8.08	3.57	3.53	4.52	3.78	—	3.00	2.27	0.42	—
ISOMERS44 (ΔH_r)	0.70	2.29	1.45	1.31	1.19	1.10	1.68	0.95	0.89	0.50	—
IsoL6/11	1.48	5.26	3.83	3.36	1.75	1.64	1.46	1.65	1.55	0.62	0.47
HC7/11	5.37	6.44	16.90	13.98	6.83	7.10	2.53	8.89	9.16	1.43	1.57
S66x8	0.75	1.70	0.97	1.22	0.49	0.49	2.20	0.69	0.52	0.56	0.09
WATER27	4.45	36.37	41.28	36.95	14.07	15.26	—	11.59	12.77	2.07	2.57
PA	16.62	8.04	7.33	8.38	3.75	3.72	—	16.04	16.02	10.53	0.32
G21IP	10.17	6.67	7.03	6.98	4.57	4.56	—	8.69	8.68	8.82	3.26
G21EA	13.93	28.51	28.41	27.72	8.42	8.41	—	12.94	12.93	13.97	8.09
Torsion	0.74	0.55	0.30	0.29	0.20	0.19	0.23	0.23	0.23	0.19	0.05
	excitation energies, eV										
Thiel's set	0.35	0.32	0.45	0.36	0.36	0.36	—	0.35	0.35	0.35	—
	bond lengths, Å										
CHNO	0.015	0.006	0.008	0.007	0.010	0.010	0.011	0.010	0.010	0.007	—
MGHBL9	0.023	0.007	0.006	0.005	0.002	0.002	0.047	0.011	0.011	0.004	—
MGNHBL11	0.026	0.006	0.003	0.002	0.008	0.008	0.004	0.008	0.008	0.002	—
CATIONS41	0.023	0.011	0.009	0.010	0.011	0.011	—	0.015	0.015	0.017	—
	bond angles, °										
CHNO	2.04	0.70	0.68	0.64	0.68	0.68	1.00	0.77	0.77	0.70	—
CATIONS41	2.21	0.52	0.37	0.44	0.46	0.46	—	1.48	1.48	1.26	—
	dihedral angles, °										
CHNO	4.07	5.20	4.68	6.10	7.12	7.11	5.86	2.14	2.14	2.31	—
CATIONS41	2.49	0.31	0.32	0.40	0.19	0.19	—	0.76	0.76	0.72	—

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022



data set	ODM2	B3LYP/ 6-31G*	ωB97X/ 6-31G*	ωB97X-D/ 6-31G*	ωB97X/ def2-TZVPP	ωB97X-D4/ def2-TZVPP	ANI- 1ccx	AIQM1 @DFT*	AIQM1 @DFT	AIQM1	CCSD(T)* /CBS
	energies, kcal/mol										
CHNO	2.64	6.71	4.10	3.84	3.21	2.76	—	2.49	2.12	0.87	—
G3/99	3.04	8.53	3.46	3.22	4.18	3.20	—	2.83	2.06	0.88	—
ISOMERS44 (ΔH_f)	1.16	8.08	3.57	3.53	4.52	3.78	—	3.00	2.27	0.42	—
ISOMERS44 (ΔH_r)	0.70	2.29	1.45	1.31	1.19	1.10	1.68	0.95	0.89	0.50	—
IsoL6/11	1.48	5.26	3.83	3.36	1.75	1.64	1.46	1.65	1.55	0.62	0.47
HC7/11	5.37	6.44	16.90	13.98	6.83	7.10	2.53	8.89	9.16	1.43	1.57

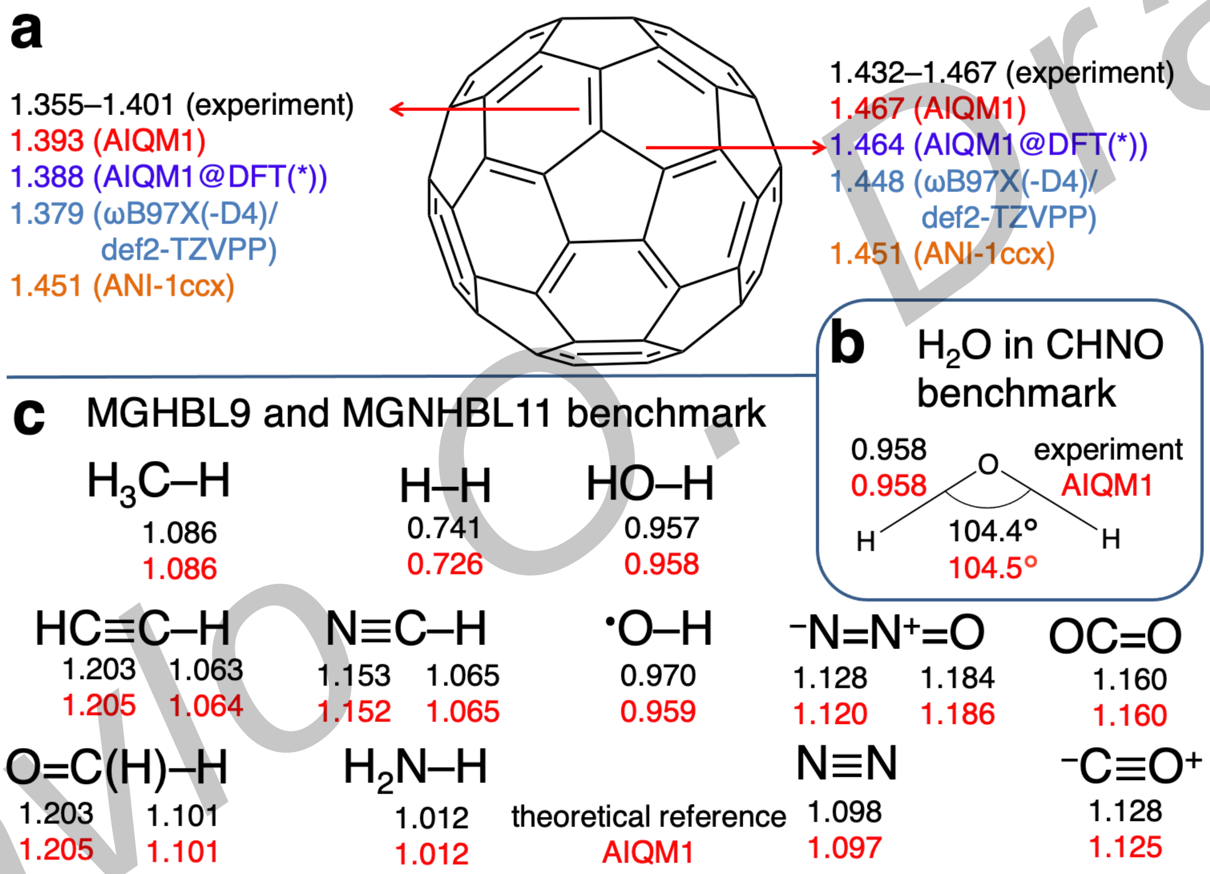
Ground state properties (heats of formation, reaction enthalpies, and ZPVE-exclusive reaction energies)



Torsion	0.74	0.55	0.30	0.29	0.20	0.19	0.23	0.23	0.23	0.19	0.05
	bond lengths, Å										
CHNO	0.015	0.006	0.008	0.007	0.010	0.010	0.011	0.010	0.010	0.007	—
MGHBL9	0.023	0.007	0.006	0.005	0.002	0.002	0.047	0.011	0.011	0.004	—
MGNHBL11	0.026	0.006	0.003	0.002	0.008	0.008	0.004	0.008	0.008	0.002	—
	bond angles, °										
CHNO	2.04	0.70	0.68	0.64	0.68	0.68	1.00	0.77	0.77	0.70	—
	dihedral angles, °										
CHNO	4.07	5.20	4.68	6.10	7.12	7.11	5.86	2.14	2.14	2.31	—

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

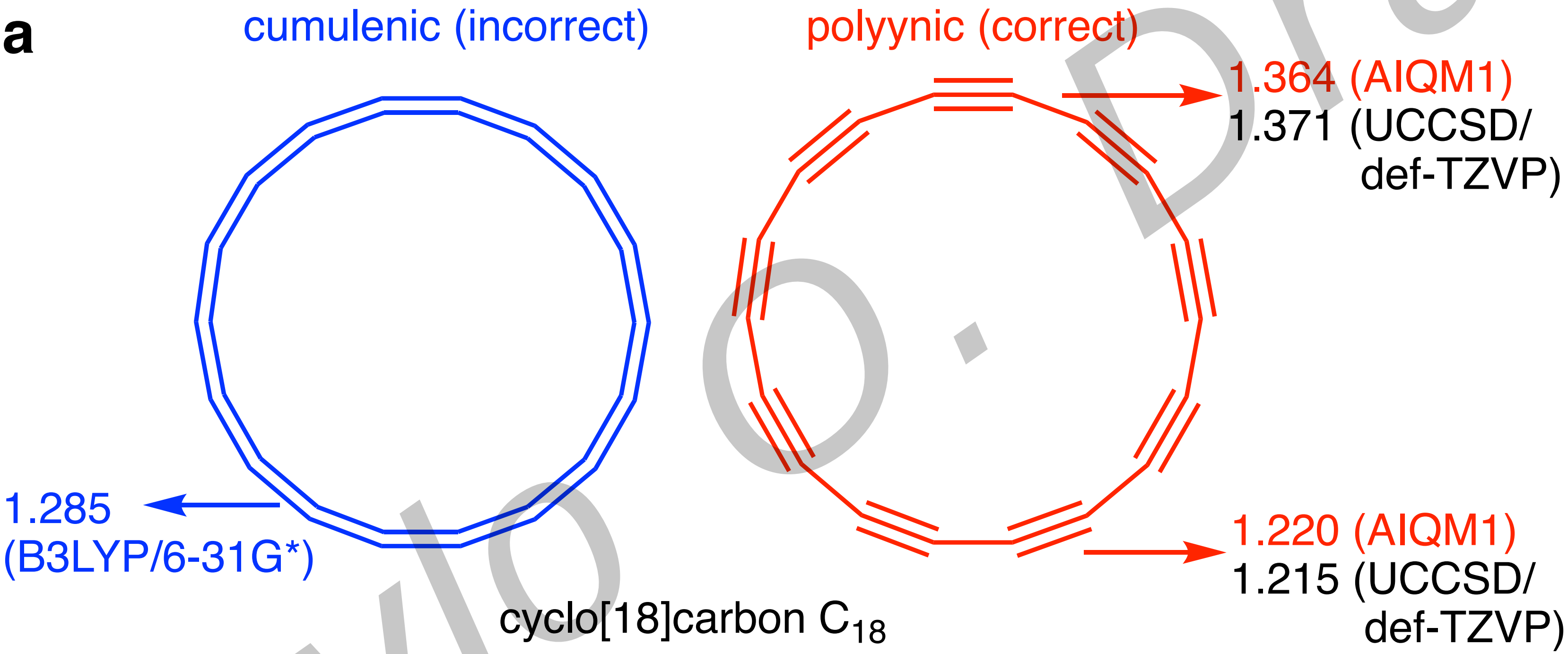
Ground-state geometries



P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

AIQM1 is better than DFT

Ground-state geometries



a

cumulenic (incorrect)

polyynic (correct)

1.364 (AIQM1)
1.371 (UCCSD/
def-TZVP)

1.285
(B3LYP/6-31G*)

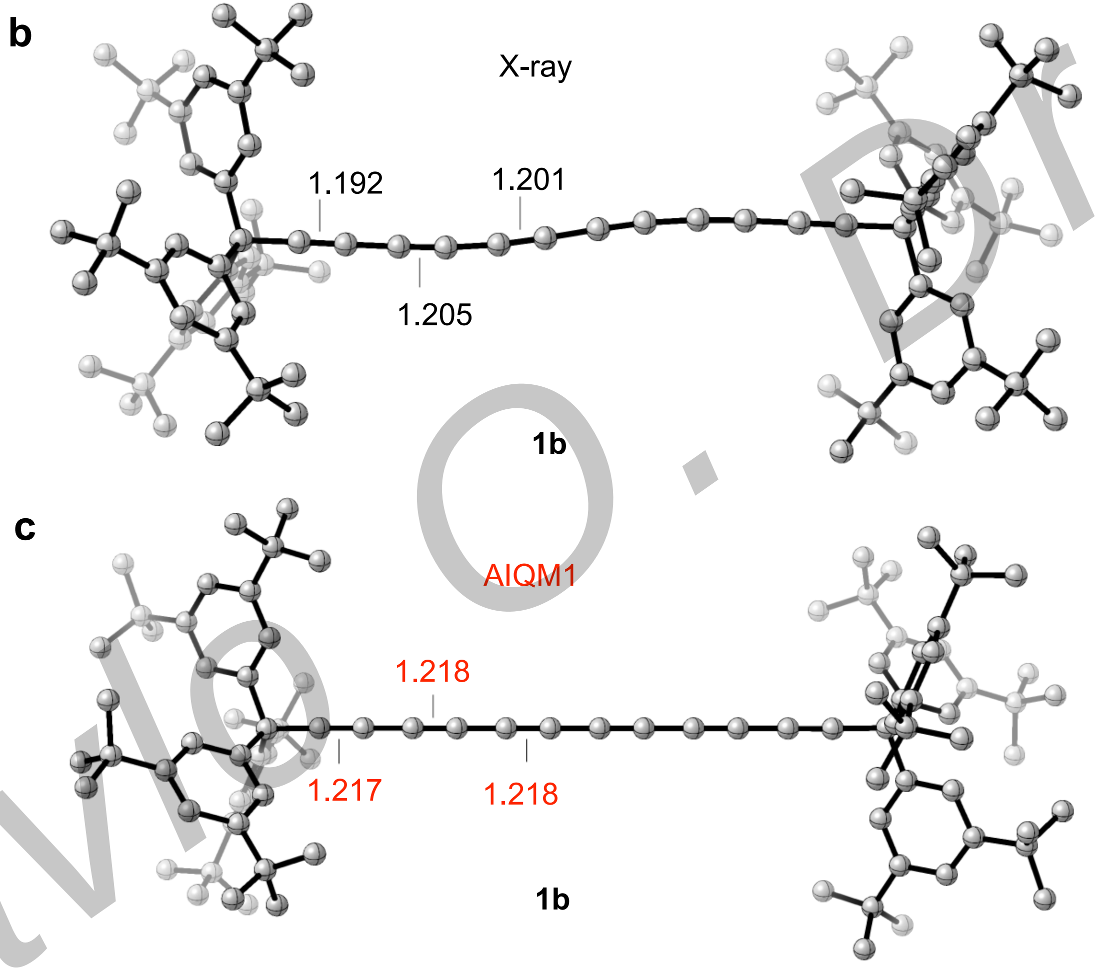
cyclo[18]carbon C_{18}

1.220 (AIQM1)
1.215 (UCCSD/
def-TZVP)

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

… and better than X-ray!

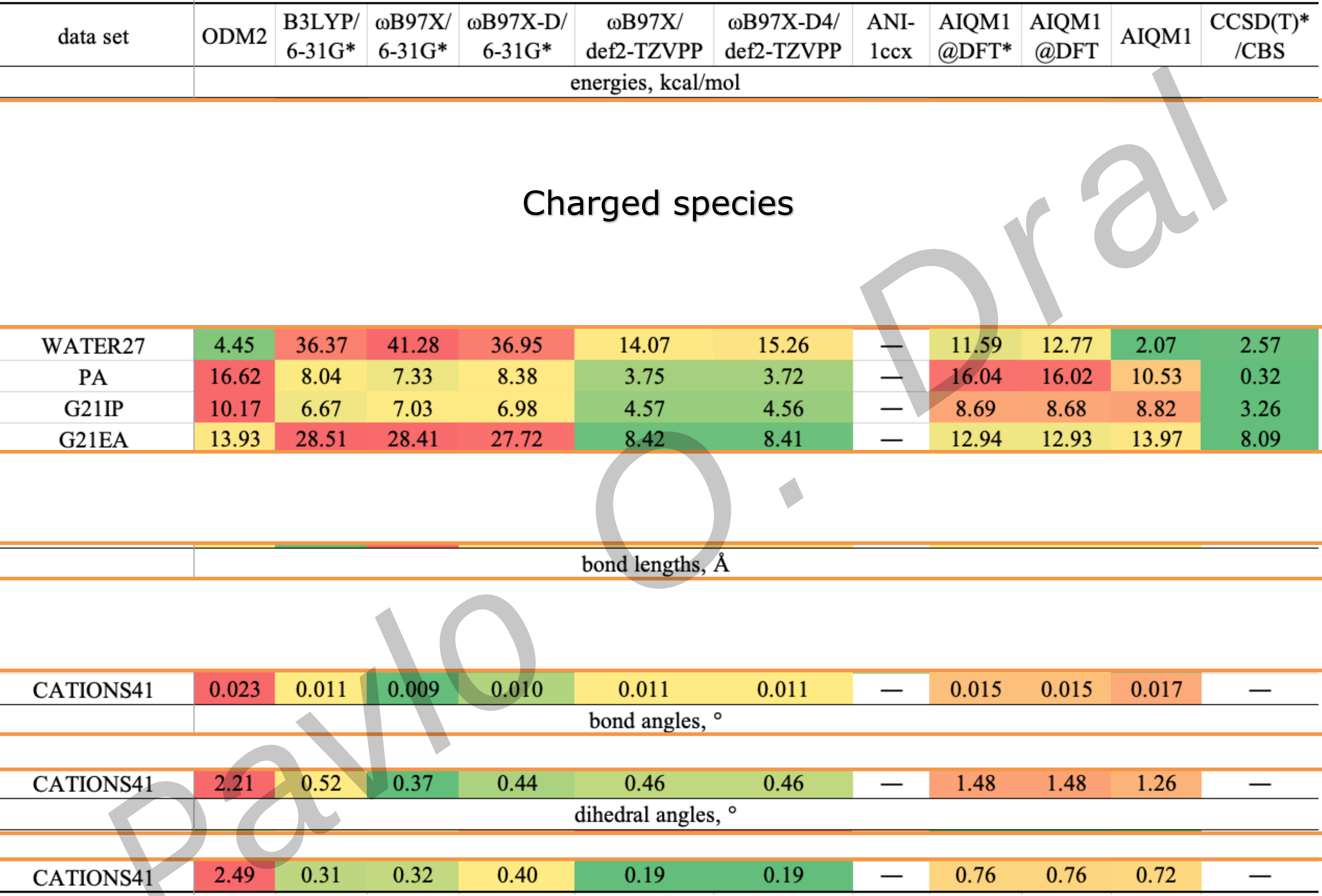
Ground-state geometries



P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

AIQM1 is transferable

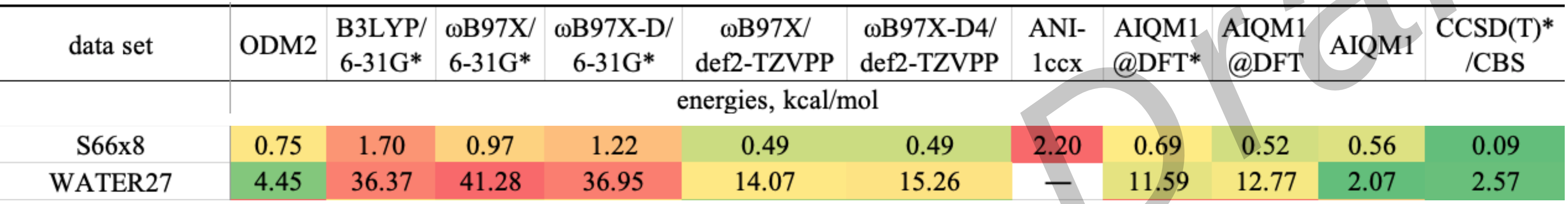
Charged species



data set	ODM2	B3LYP/ 6-31G*	ωB97X/ 6-31G*	ωB97X-D/ 6-31G*	ωB97X/ def2-TZVPP	ωB97X-D4/ def2-TZVPP	ANI-1ccx	AIQM1 @DFT*	AIQM1 @DFT	AIQM1	CCSD(T)* /CBS
	energies, kcal/mol										
WATER27	4.45	36.37	41.28	36.95	14.07	15.26	—	11.59	12.77	2.07	2.57
PA	16.62	8.04	7.33	8.38	3.75	3.72	—	16.04	16.02	10.53	0.32
G21IP	10.17	6.67	7.03	6.98	4.57	4.56	—	8.69	8.68	8.82	3.26
G21EA	13.93	28.51	28.41	27.72	8.42	8.41	—	12.94	12.93	13.97	8.09
	bond lengths, Å										
CATIONS41	0.023	0.011	0.009	0.010	0.011	0.011	—	0.015	0.015	0.017	—
	bond angles, °										
CATIONS41	2.21	0.52	0.37	0.44	0.46	0.46	—	1.48	1.48	1.26	—
	dihedral angles, °										
CATIONS41	2.49	0.31	0.32	0.40	0.19	0.19	—	0.76	0.76	0.72	—

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

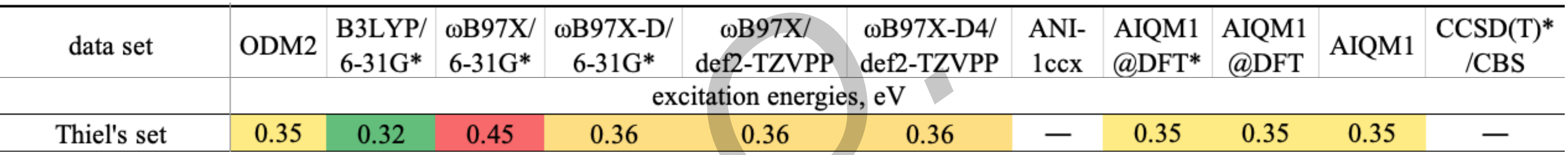
Noncovalent interactions



data set	ODM2	B3LYP/ 6-31G*	ωB97X/ 6-31G*	ωB97X-D/ 6-31G*	ωB97X/ def2-TZVPP	ωB97X-D4/ def2-TZVPP	ANI-1ccx	AIQM1 @DFT*	AIQM1 @DFT	AIQM1	CCSD(T)* /CBS
	energies, kcal/mol										
S66x8	0.75	1.70	0.97	1.22	0.49	0.49	2.20	0.69	0.52	0.56	0.09
WATER27	4.45	36.37	41.28	36.95	14.07	15.26	—	11.59	12.77	2.07	2.57

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

Excited states

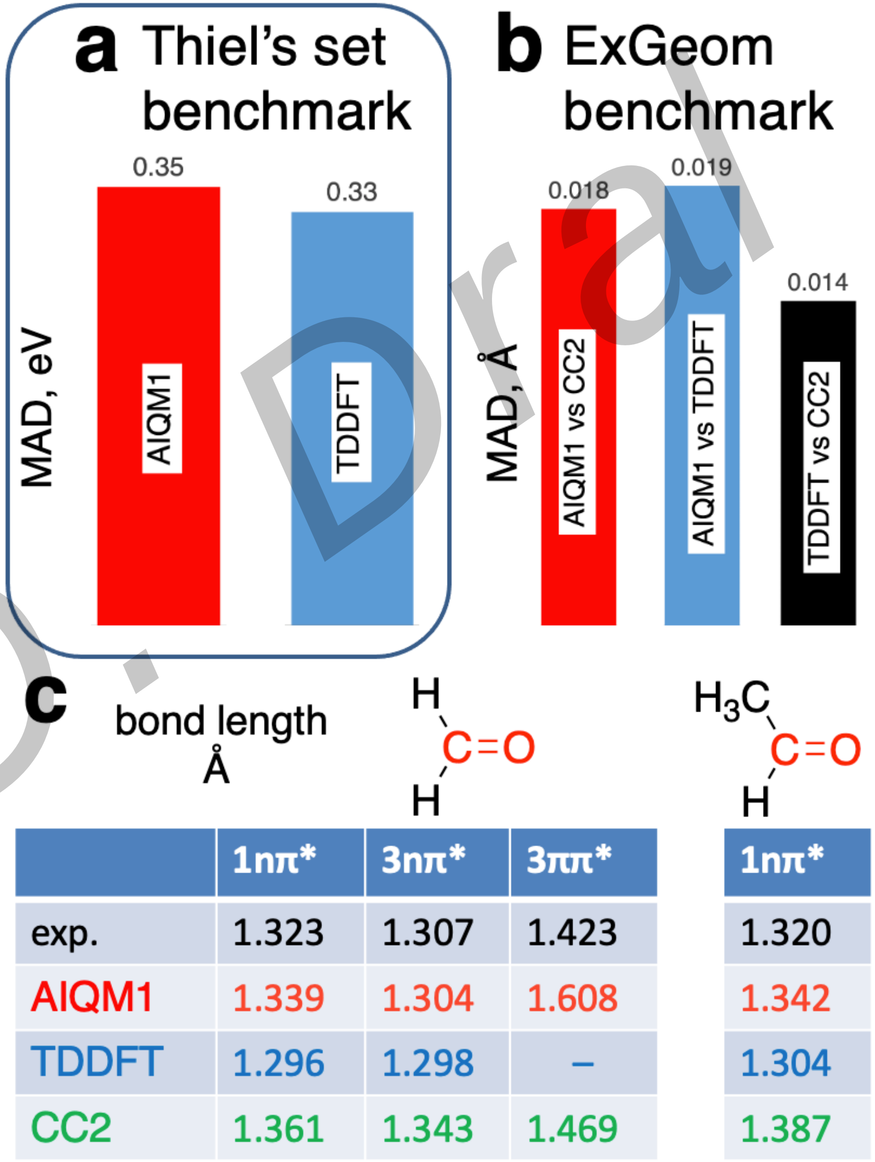


data set	ODM2	B3LYP/ 6-31G*	ωB97X/ 6-31G*	ωB97X-D/ 6-31G*	ωB97X/ def2-TZVPP	ωB97X-D4/ def2-TZVPP	ANI-1ccx	AIQM1 @DFT*	AIQM1 @DFT	AIQM1	CCSD(T)* /CBS
	excitation energies, eV										
Thiel's set	0.35	0.32	0.45	0.36	0.36	0.36	—	0.35	0.35	0.35	—

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

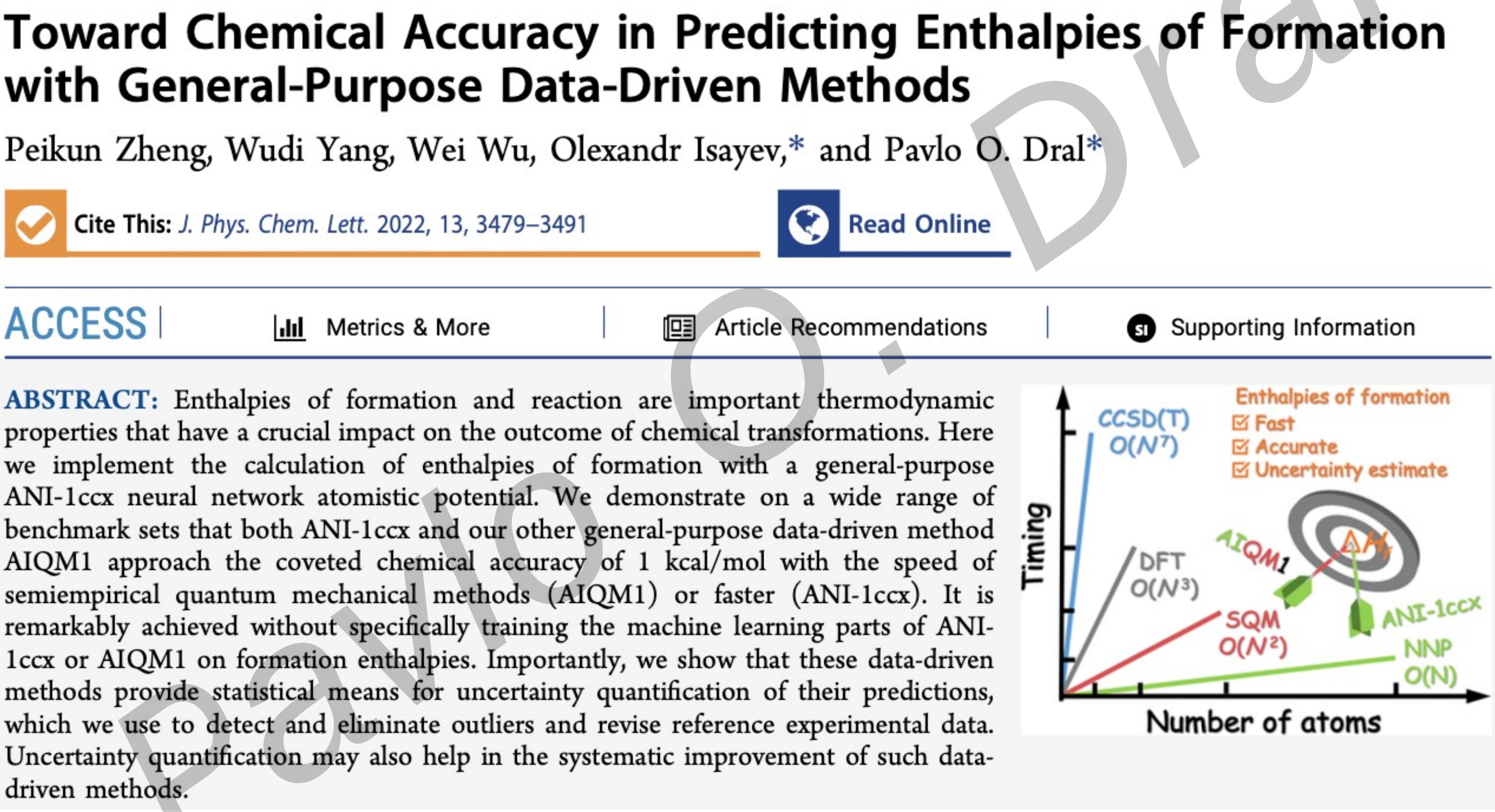
Excited states

TDDFT: B3LYP/TZVP



	1nπ*	3nπ*	3ππ*	1nπ*
exp.	1.323	1.307	1.423	1.320
AIQM1	1.339	1.304	1.608	1.342
TDDFT	1.296	1.298	–	1.304
CC2	1.361	1.343	1.469	1.387

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022



Toward Chemical Accuracy in Predicting Enthalpies of Formation with General-Purpose Data-Driven Methods

Peikun Zheng, Wudi Yang, Wei Wu, Olexandr Isayev,* and Pavlo O. Dral*

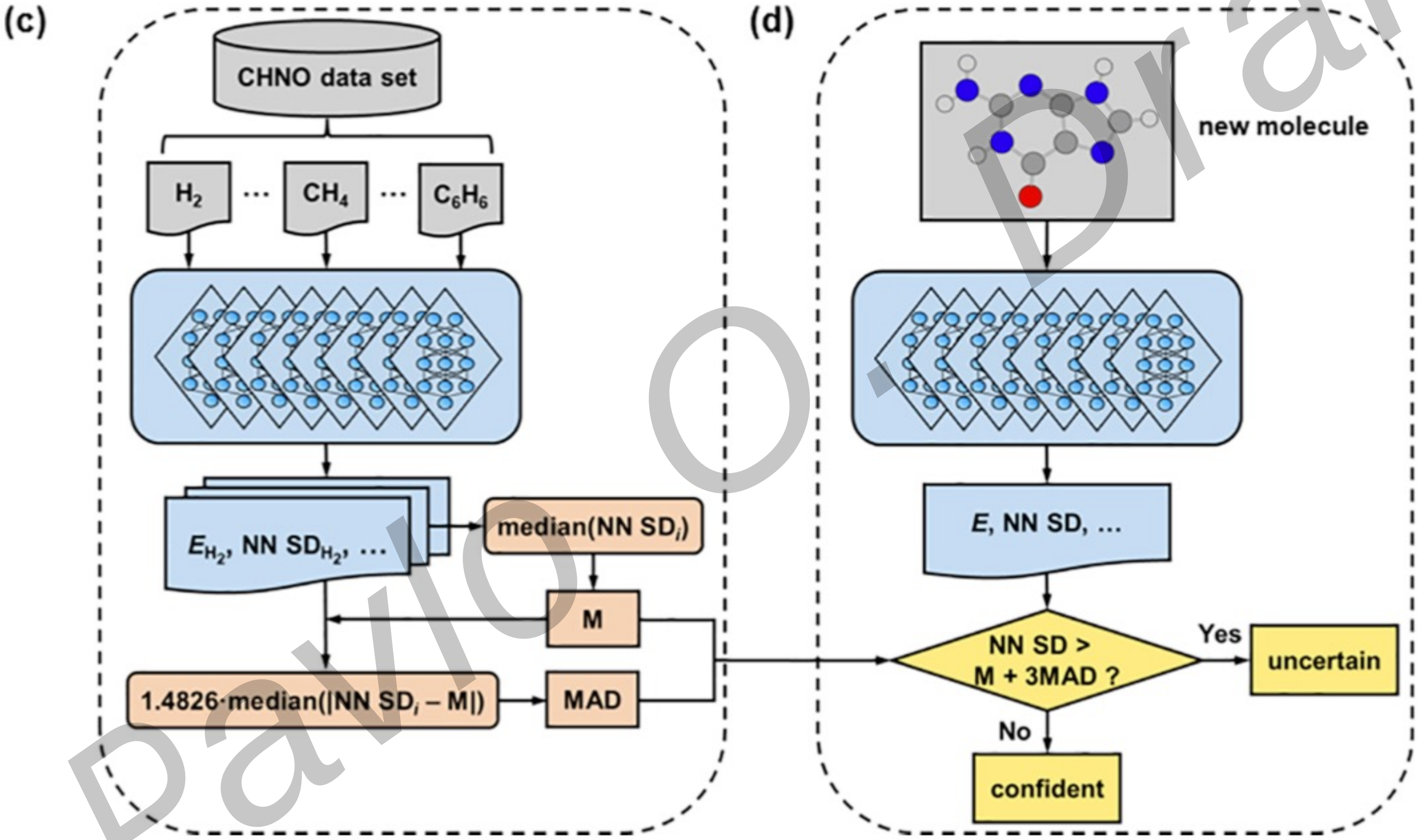
Cite This: *J. Phys. Chem. Lett.* 2022, 13, 3479−3491

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

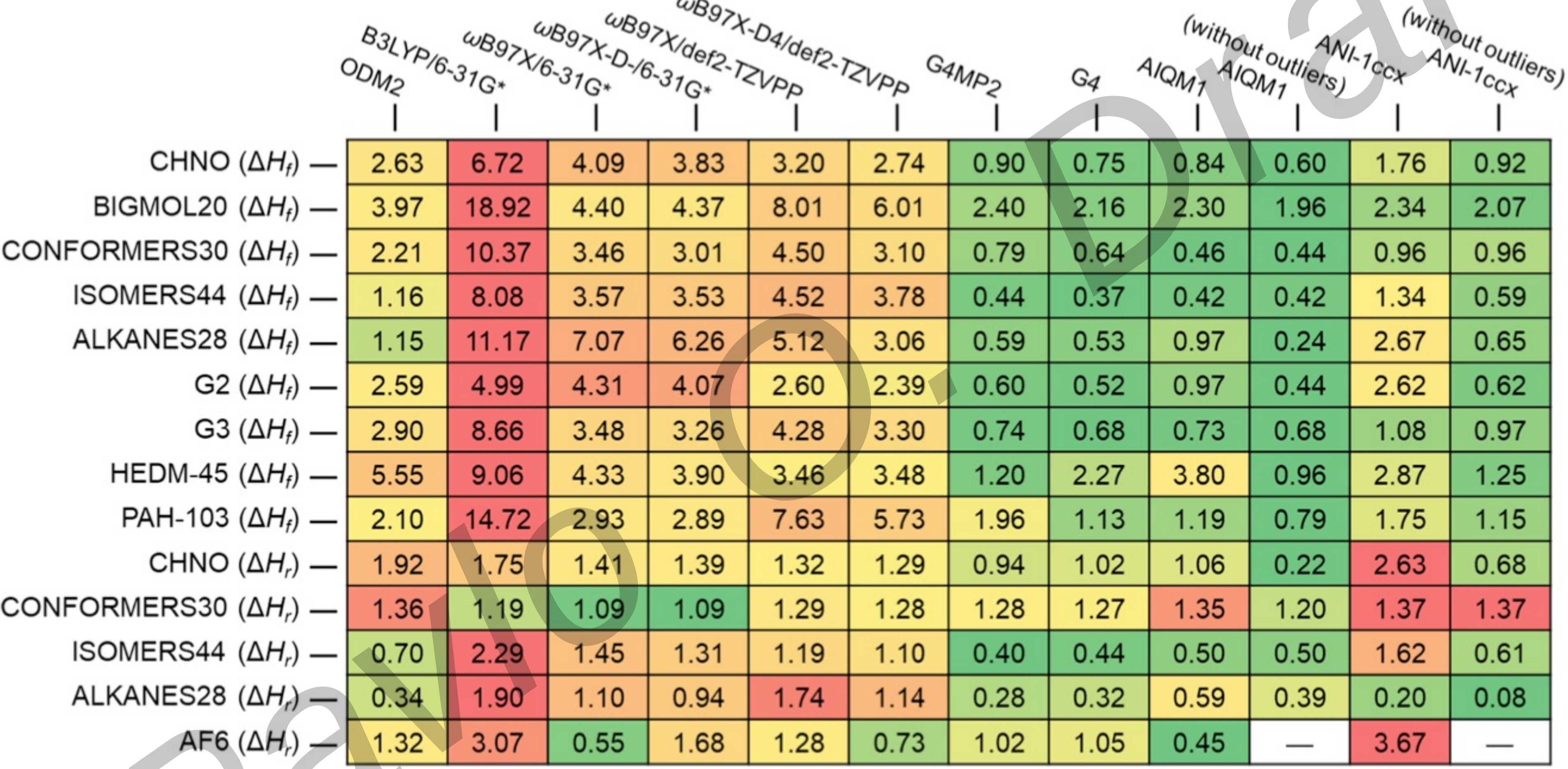
ABSTRACT: Enthalpies of formation and reaction are important thermodynamic properties that have a crucial impact on the outcome of chemical transformations. Here we implement the calculation of enthalpies of formation with a general-purpose ANI-1ccx neural network atomistic potential. We demonstrate on a wide range of benchmark sets that both ANI-1ccx and our other general-purpose data-driven method AIQM1 approach the coveted chemical accuracy of 1 kcal/mol with the speed of semiempirical quantum mechanical methods (AIQM1) or faster (ANI-1ccx). It is remarkably achieved without specifically training the machine learning parts of ANI-1ccx or AIQM1 on formation enthalpies. Importantly, we show that these data-driven methods provide statistical means for uncertainty quantification of their predictions, which we use to detect and eliminate outliers and revise reference experimental data. Uncertainty quantification may also help in the systematic improvement of such data-driven methods.

Uncertainty quantification



P. Zheng, W. Yang, W. Wu, O. Isayev, P. O. Dral, *J. Phys. Chem. Lett.* **2022**, *13*, 3479

Heats of Formation: Outliers from Uncertainty Quantification



	ODM2	B3LYP/6-31G*	ωB97X/6-31G*	ωB97X-D-/6-31G*	ωB97X/def2-TZVPP	ωB97X-D4/def2-TZVPP	G4MP2	G4	AIQM1	AIQM1 (without outliers)	ANI-1ccx	ANI-1ccx (without outliers)
CHNO (ΔH_f)	2.63	6.72	4.09	3.83	3.20	2.74	0.90	0.75	0.84	0.60	1.76	0.92
BIGMOL20 (ΔH_f)	3.97	18.92	4.40	4.37	8.01	6.01	2.40	2.16	2.30	1.96	2.34	2.07
CONFORMERS30 (ΔH_f)	2.21	10.37	3.46	3.01	4.50	3.10	0.79	0.64	0.46	0.44	0.96	0.96
ISOMERS44 (ΔH_f)	1.16	8.08	3.57	3.53	4.52	3.78	0.44	0.37	0.42	0.42	1.34	0.59
ALKANES28 (ΔH_f)	1.15	11.17	7.07	6.26	5.12	3.06	0.59	0.53	0.97	0.24	2.67	0.65
G2 (ΔH_f)	2.59	4.99	4.31	4.07	2.60	2.39	0.60	0.52	0.97	0.44	2.62	0.62
G3 (ΔH_f)	2.90	8.66	3.48	3.26	4.28	3.30	0.74	0.68	0.73	0.68	1.08	0.97
HEDM-45 (ΔH_f)	5.55	9.06	4.33	3.90	3.46	3.48	1.20	2.27	3.80	0.96	2.87	1.25
PAH-103 (ΔH_f)	2.10	14.72	2.93	2.89	7.63	5.73	1.96	1.13	1.19	0.79	1.75	1.15
CHNO (ΔH_r)	1.92	1.75	1.41	1.39	1.32	1.29	0.94	1.02	1.06	0.22	2.63	0.68
CONFORMERS30 (ΔH_r)	1.36	1.19	1.09	1.09	1.29	1.28	1.28	1.27	1.35	1.20	1.37	1.37
ISOMERS44 (ΔH_r)	0.70	2.29	1.45	1.31	1.19	1.10	0.40	0.44	0.50	0.50	1.62	0.61
ALKANES28 (ΔH_r)	0.34	1.90	1.10	0.94	1.74	1.14	0.28	0.32	0.59	0.39	0.20	0.08
AF6 (ΔH_r)	1.32	3.07	0.55	1.68	1.28	0.73	1.02	1.05	0.45	—	3.67	—

Mean absolute errors (MAEs) in kcal/mol of benchmarked methods for various data sets. MAEs of AIQM1 and ANI-1ccx

P. Zheng, W. Yang, W. Wu, O. Isayev, P. O. Dral, *J. Phys. Chem. Lett.* **2022**, *13*, 3479

Revising Experimental Heats of Formation

CHNO data set

vinylacetylene
original: 73
revised: 70.4
AIQM1: 69.1
G4: 69.1

acetylacetone
original: –91
revised: –89.9
AIQM1: –87.3
G4: –87.6

1,3-pyrimidine
exper.: 46.8

AIQM1: 44.5
G4: 44.3

acrolein
exper.: –18

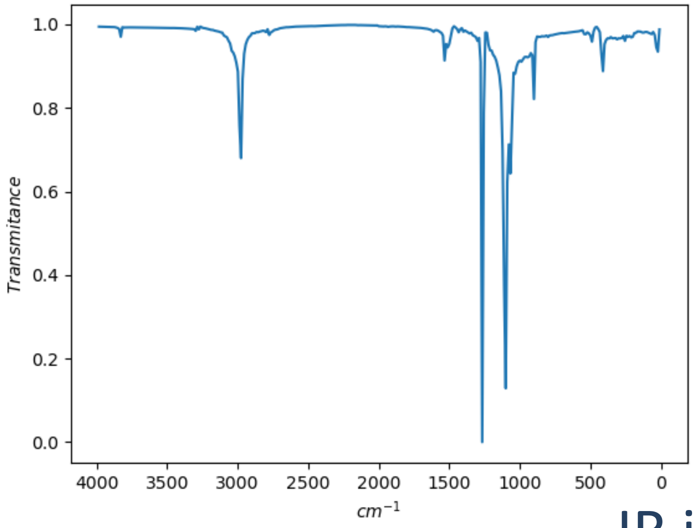
AIQM1: –16.0
G4: –15.9

P. Zheng, W. Yang, W. Wu, O. Isayev, P. O. Dral, *J. Phys. Chem. Lett.* **2022**, *13*, 3479

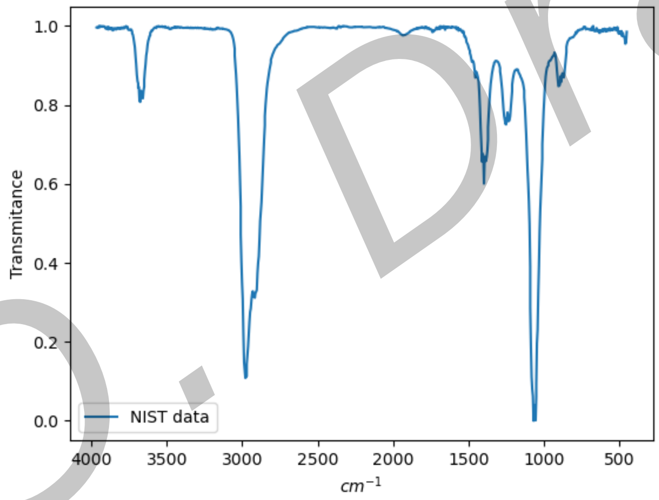
Infrared spectra with AIQM1

AIQM1

By Yifan Hou, unpublished



Experiment (NIST)



IR intensities:

$$P(\omega) \propto \int \langle \dot{\boldsymbol{\mu}}(\tau)\dot{\boldsymbol{\mu}}(t+\tau)\rangle_{\tau} e^{-i\omega t}\mathrm{d}t$$

$$P(\omega) \propto \omega \tanh\left(\frac{\hbar\omega}{2k_B T}\right)\int \langle \boldsymbol{\mu}(\tau)\boldsymbol{\mu}(t+\tau)\rangle_{\tau} e^{-i\omega t}\mathrm{d}t$$

Long molecular dynamic trajectory is needed!

Phys. Chem. Chem. Phys. **2013,** 15, 6608

Phys. Chem. Chem. Phys. **2016**, 28325

Pros and cons of AIQM1

✓ Very good for ground-state, neutral species

If you can use AIQM1, do not use B3LYP/6-31G*!

✓ Provides uncertainty estimates

× Still only limited to H, C, N, and O elements

- Can be improved for noncovalent, charged, and excited-state properties

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

Example of highly accurate, global, hole-free potential energy surface for rovibrational spectra with spectroscopic accuracy

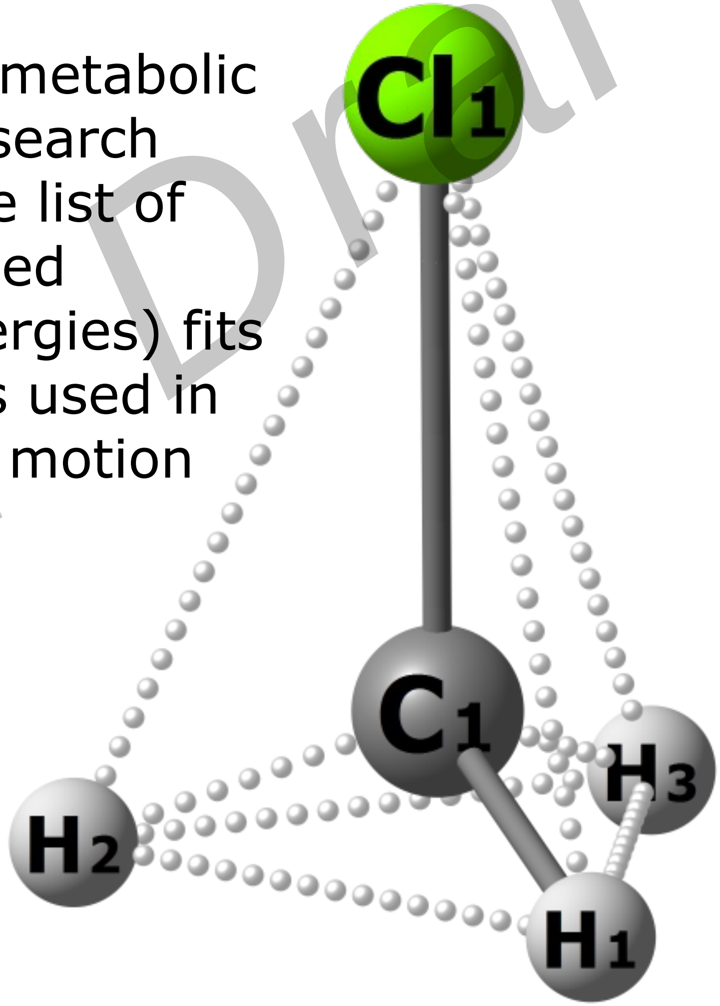
Spectroscopic accuracy: error less than
1 cm^{-1} ≈ 0.003 kcal/mol ≈ 0.01 kJ/mol ≈ 0.0001 eV ≈ 0.000005 Hartree

Potential Energy Surface of CH_3Cl

CH_3Cl

- Biomarker produced from a secondary metabolic process: useful for extraterrestrial life search
- Highly accurate and comprehensive line list of the rotation-vibration spectrum is needed
- **TROVE** (Theoretical ROVibrational Energies) fits analytic representation of PES, which is used in variational approach to solving nuclear motion problem

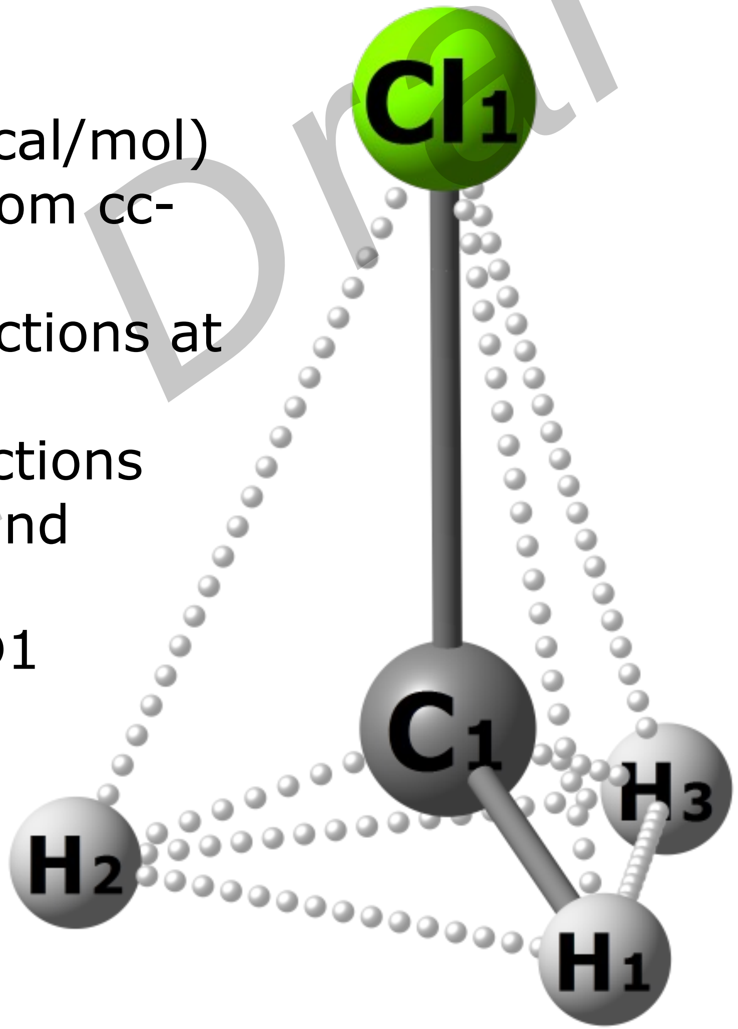
$$\left(\sum_{i=1}^{N}\frac{-\hbar^2}{2M_i}\nabla_i^2+V\right)\Psi=E\Psi$$



Owens, Yurchenko, Yachmenev, Tennyson, Thiel, *J. Chem. Phys.* **2015**, *142*, 244306
TROVE: Yurchenko, Thiel, Jensen, *J. Mol. Spectrosc.* **2007**, *245*, 126

Ab Initio PES of $CH_3{}^{35}Cl$

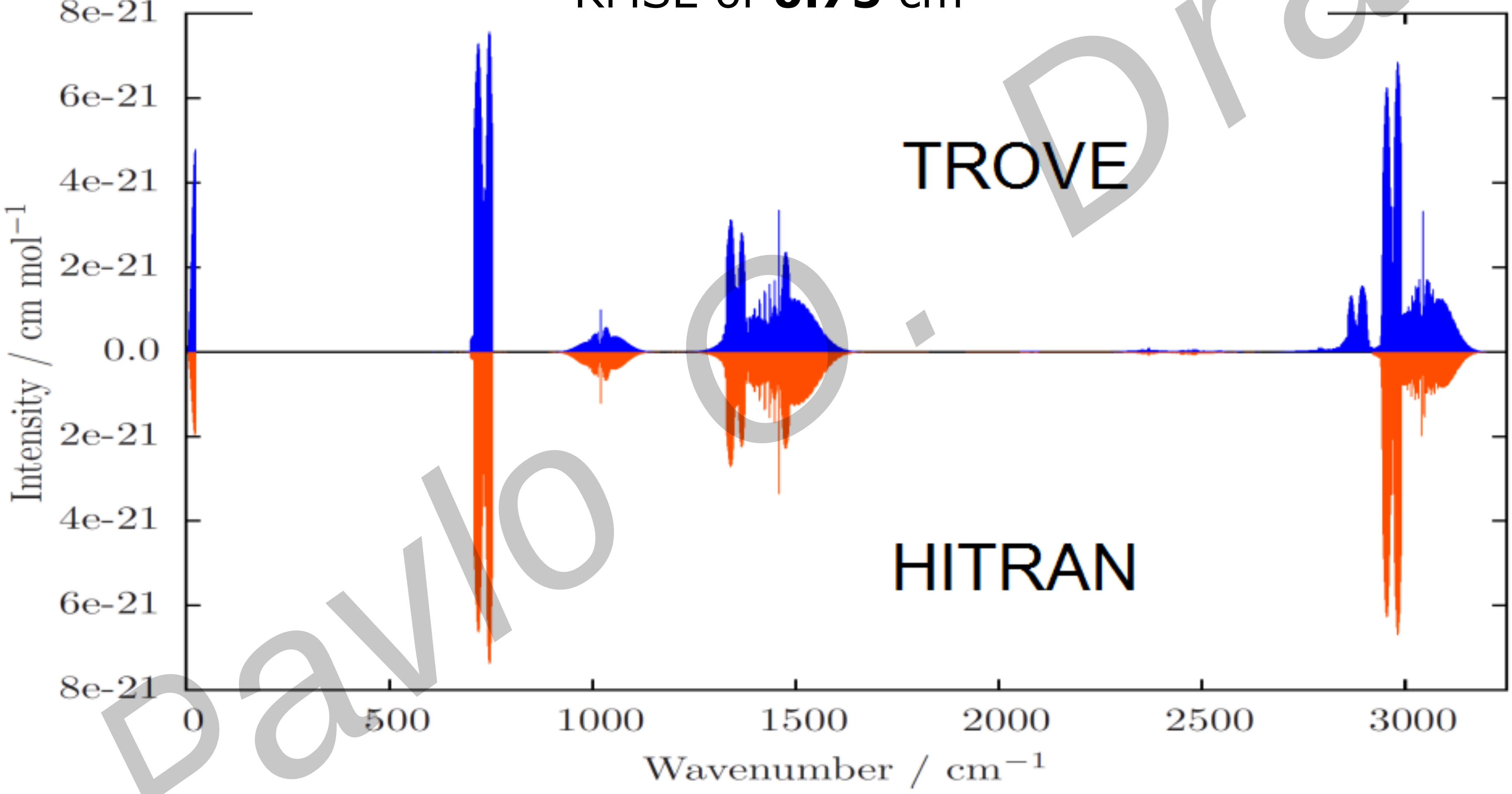
- 44820 grid points (**CBS-35^{HL}**)
- Energies up to *hc* · 50000 cm^{-1} (140 kcal/mol)
- CCSD(T,FC)-F12b/CBS (extrapolated from cc-pVTZ-F12 and cc-pVQZ-F12)
- Core-valence electron correlation corrections at CCSD(T)-F12b/cc-pCVQZ-F12
- Higher-order electron correlation corrections from CCSDT/aug-cc-pVTZ(+d for Cl), and CCSDT(Q)/aug-cc-pVDZ(+d for Cl)
- Scalar relativistic effects using the MVD1 approach at CCSD(T)/aug-cc-pCVTZ(+d for Cl)
- The diagonal Born-Oppenheimer correction at CCSD/aug-cc-pCVTZ(+d for Cl)



Owens, Yurchenko, Yachmenev, Tennyson, Thiel, *J. Chem. Phys.* **2015**, *142*, 244306

Simulated Absorption Line List

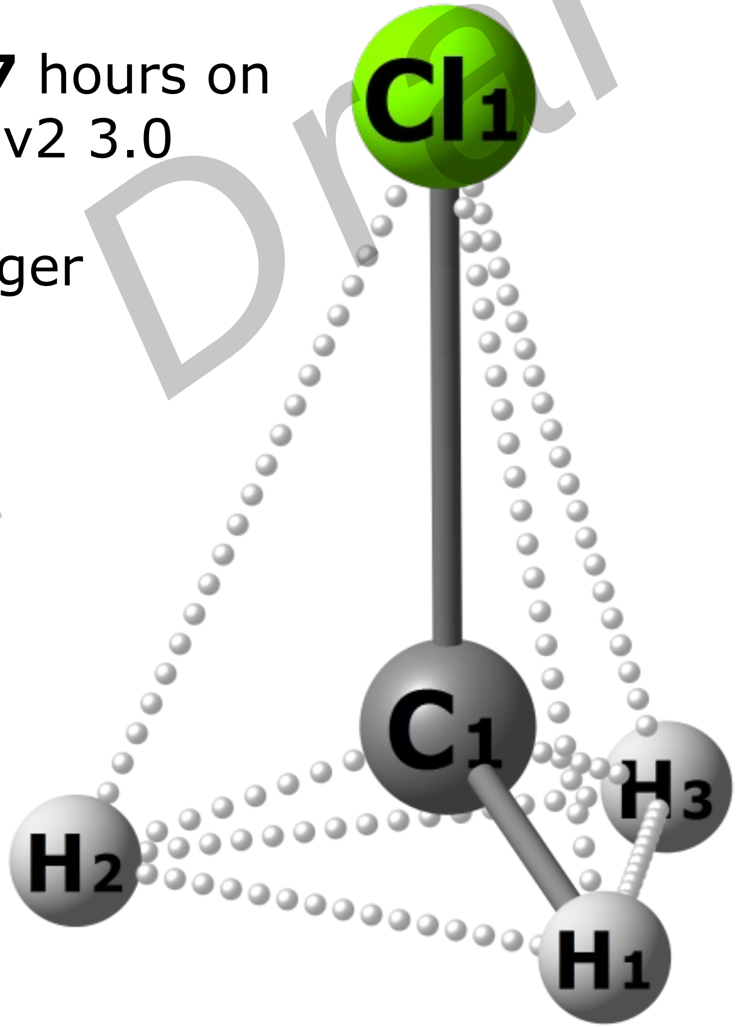
Fundamental term values are reproduced with RMSE of **0.75** cm^{-1}



Owens, Yurchenko, Yachmenev, Tennyson, Thiel, *J. Chem. Phys.* **2015**, *142*, 244306

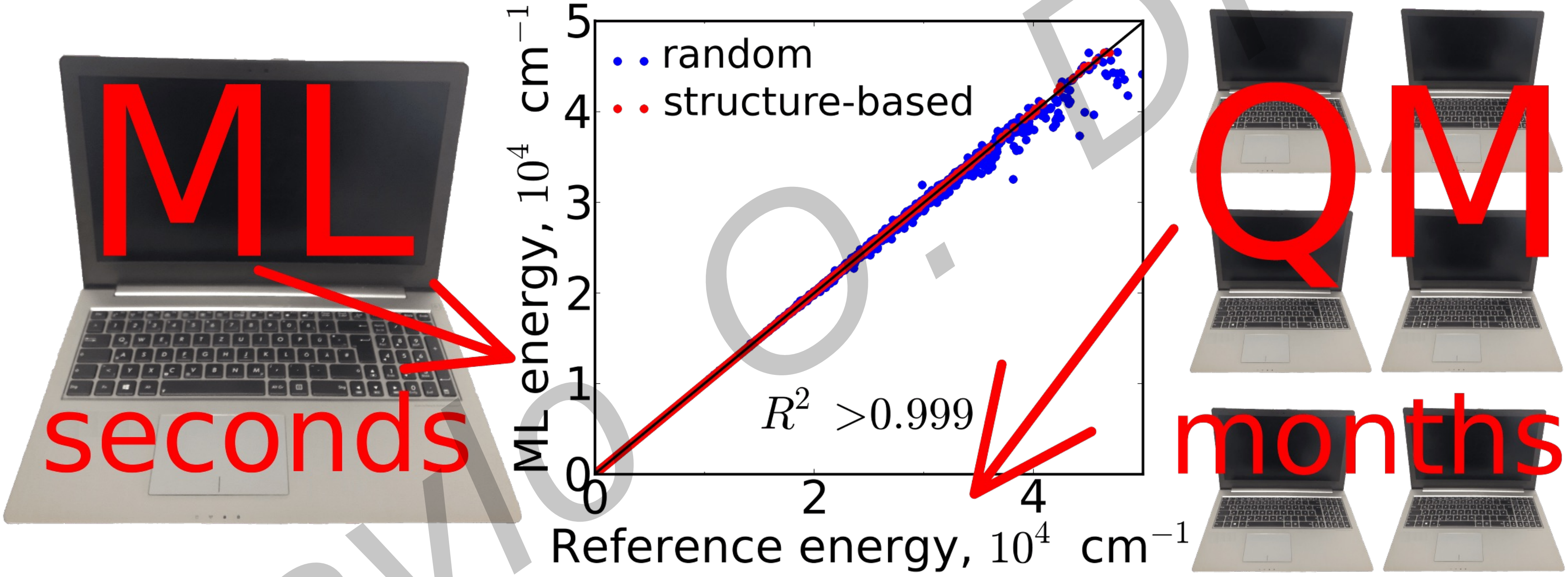
Ab Initio PES of $CH_3{}^{35}Cl$

- Equilibrium structure calculations: **26.7** hours on a single core of an Intel Xeon E5-2690 v2 3.0 GHz processor
- Distorted structures can take much longer



Owens, Yurchenko, Yachmenev, Tennyson, Thiel, *J. Chem. Phys.* **2015**, *142*, 244306

Motivation



Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Kernel Ridge Regression

$$Y^{\mathrm{ML}}(\mathbf{M}_a) = \sum_{i=1}^{N_{train}} \alpha_i K(\mathbf{M}_a, \mathbf{M}_i)$$

$$K(\mathbf{M}_a, \mathbf{M}_i) = \exp\left(-\frac{\left\| \mathbf{M}_i - \mathbf{M}_a \right\|_2^2}{2\sigma^2} \right)$$ Gaussian kernel

$$\min_{\alpha} \sum_{i}^{N_{train}} \left(Y_i^{est} - Y_i^{ref} \right)^2 \boxed{+ \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha}}$$

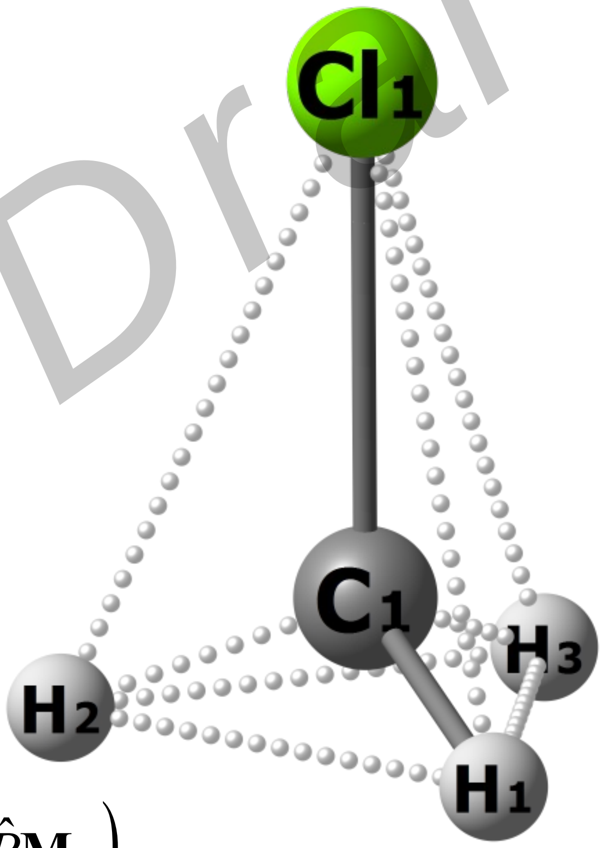
$$\boldsymbol{\alpha} = \left(\mathbf{K} \boxed{+ \lambda \mathbf{I}} \right)^{-1} \mathbf{Y}^{ref}$$

Hastie, Tibshirani, Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed.; Springer-Verlag, **2009**

RE descriptor for CH_3Cl

$$\mathbf{M} = \begin{pmatrix} \frac{r^{eq}_{C-Cl}}{r_{C-Cl}} \\ \frac{r^{eq}_{C-H1}}{r_{C-H1}} \\ \frac{r^{eq}_{Cl-H1}}{r_{Cl-H1}} \\ \frac{r^{eq}_{H1-H2}}{r_{H1-H2}} \\ \dots \end{pmatrix}$$

- All 10 atom pairs
- Hydrogens are sorted by their nuclear repulsion for structure-based sampling
- Normalized permutational invariant kernel [2] is used for ML calculations (decreases errors 2–3 times compared to sorting)



$$\overline{K}(\mathbf{M}_i,\mathbf{M}_j) = \frac{\sum_{\hat{P}}^{N_{perm}} K(\mathbf{M}_i,\hat{P}\mathbf{M}_j)}{\sqrt{\sum_{\hat{P}}^{N_{perm}} K(\mathbf{M}_i,\hat{P}\mathbf{M}_i)}\sqrt{\sum_{\hat{P}}^{N_{perm}} K(\mathbf{M}_j,\hat{P}\mathbf{M}_j)}}$$ [2]

[1] Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108
[2] Bartók, Csányi, *Int. J. Quantum Chem.* **2015**, *115*, 1051

Low-energy region

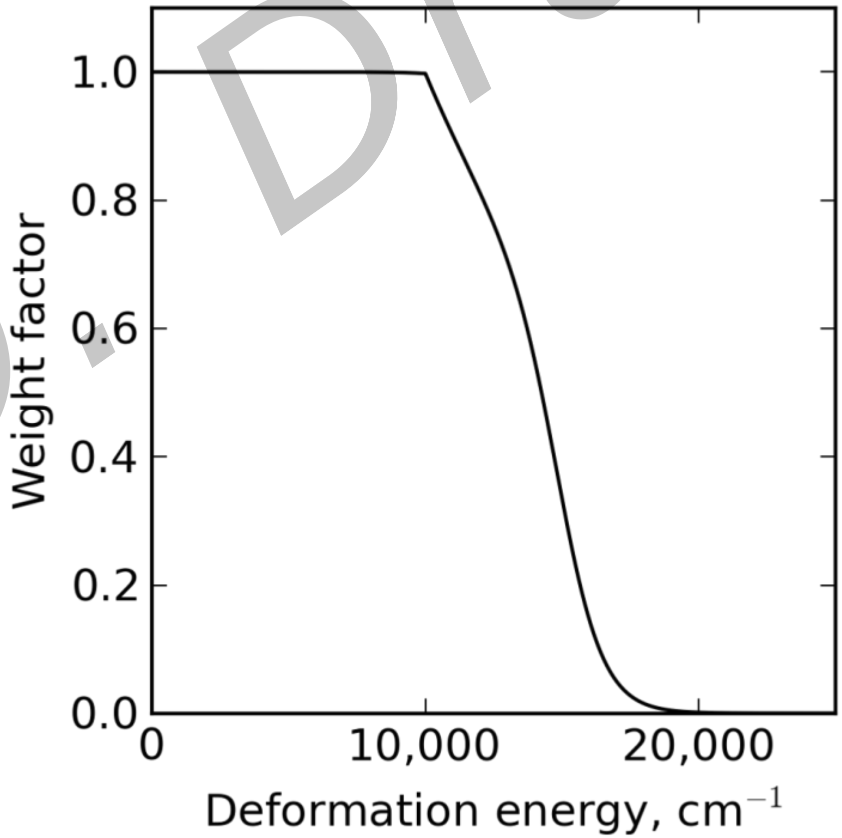
- Low-energy structures ($<10{,}000$ cm^{-1}) are most important:
  What is the error measure to take it into account?

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Weighted root-mean-square error

$$w\left(E_i\right) = \left(\frac{\tanh\left[-0.0006 \times \left(E_i - 15\,000\right)\right] + 1.002002002}{2.002002002}\right) \times \frac{1}{NE_i^{(w)}}$$

$$\mathrm{wRMSE} = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(E_i^{\mathrm{est}} - E_i^{\mathrm{ref}}\right)^2 \omega\left(E_i^{\mathrm{ref}}\right)}$$



[1] Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108
[2] Partridge, Schwenke, *J. Chem. Phys.* **1997**, *106*, 4618

Self-correcting ML

Self-correcting ML

$$Y^{\mathrm{ML}}(\mathrm{M}_a) = \sum_{i=1}^{N_{train}} \alpha_i K(M_a, M_i)$$

$$E^{\mathrm{layer\,1}}(\mathrm{M}_a) = \sum_{i=1}^{N_{train}} \alpha_i^{\mathrm{layer\,1}} K(M_a, M_i)$$ trained on $\mathbf{E}^{\mathrm{ref}}$

$$\Delta E^{\mathrm{ML,\,layer\,1}}(\mathrm{M}_a) = \sum_{i=1}^{N_{train}} \alpha_i^{\mathrm{layer\,2}} K(M_a, M_i)$$ trained on

$$\Delta\mathbf{E}^{\mathrm{ref,layer\,1}} = \mathbf{E}^{\mathrm{ref}} - \mathbf{E}^{\mathrm{layer\,1}}$$

$$E^{\mathrm{layer\,2}}(\mathrm{M}_a) = E^{\mathrm{layer\,1}}(\mathrm{M}_a) + \Delta E^{\mathrm{ML,\,layer\,1}}(\mathrm{M}_a)$$

and so on…

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Self-correcting ML

wRMSEs in cm^{-1} for the 'hold-out' prediction set for ML models trained on the training set with 10% points randomly sampled from the entire grid of 44819 points

Layer 1	Layer 2	Layer 3	Layer 4
14.73 ± 12.63	8.76 ± 1.67	8.73 ± 1.67	8.73 ± 1.67

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Structure-based sampling

- ML is good for interpolation
- ML can fail spectacularly for extrapolation
- **How can we make sure that ML interpolates rather than extrapolates?**

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Structure-based sampling

Kernel Ridge Regression

$$Y^{\mathrm{ML}}\left(\mathbf{M}_a\right)=\sum_{i=1}^{N_{train}}\alpha_i K\left(\mathbf{M}_a,\mathbf{M}_i\right)$$

$$K\left(\mathbf{M}_a,\mathbf{M}_i\right)=\exp\left(-\frac{\left\|\mathbf{M}_i-\mathbf{M}_a\right\|_2^2}{2\sigma^2}\right)$$

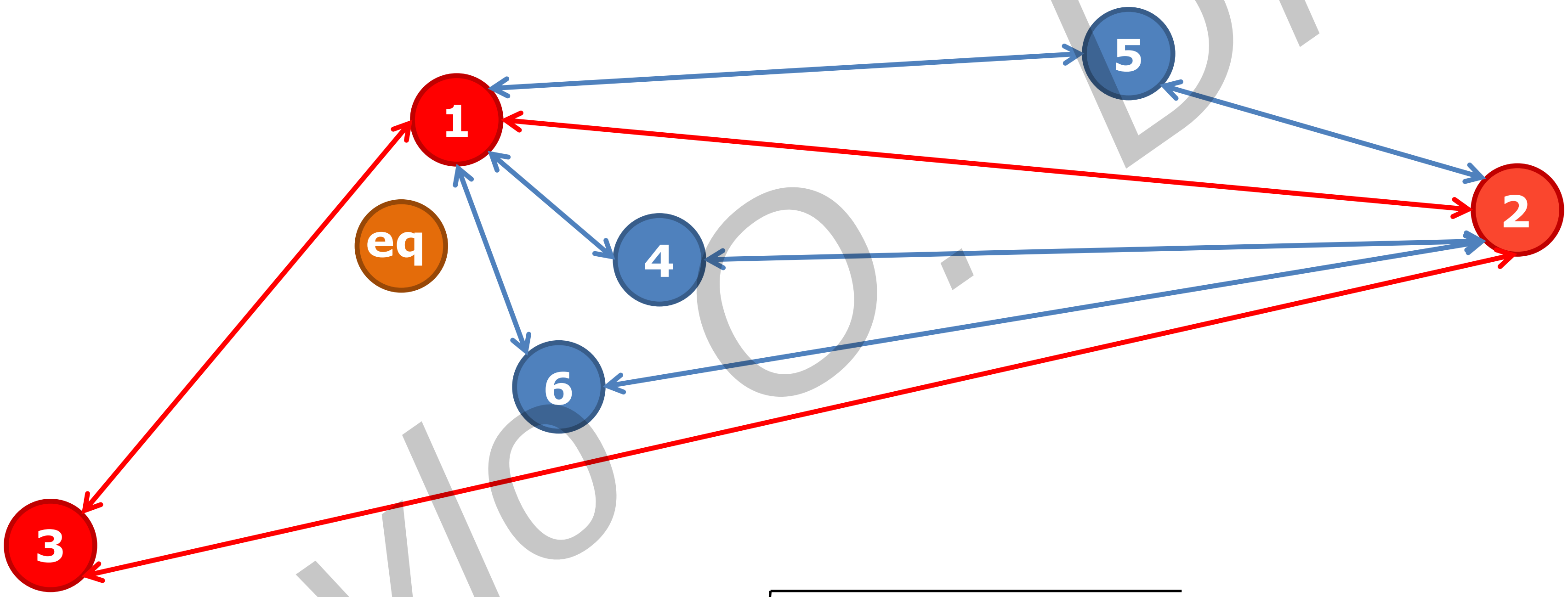
Gaussian kernel

$$\left\|\mathbf{M}_i-\mathbf{M}_a\right\|_2=D\left(\mathbf{M}_a,\mathbf{M}_i\right)=\sqrt{\sum_{\beta}^{N_{\mathbf{M}}}\left(\mathbf{M}_{i,\beta}-\mathbf{M}_{a,\beta}\right)^2}$$

2-norm is the Euclidean distance between molecular descriptors

Structure-based sampling

No preliminary energy calculations are needed!



$$D(\mathbf{M}_a, \mathbf{M}_i) = \sqrt{\sum_{\beta}^{N_{\mathbf{M}}} (\mathbf{M}_{i,\beta} - \mathbf{M}_{a,\beta})^2}$$

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Low-energy region

Low-energy region

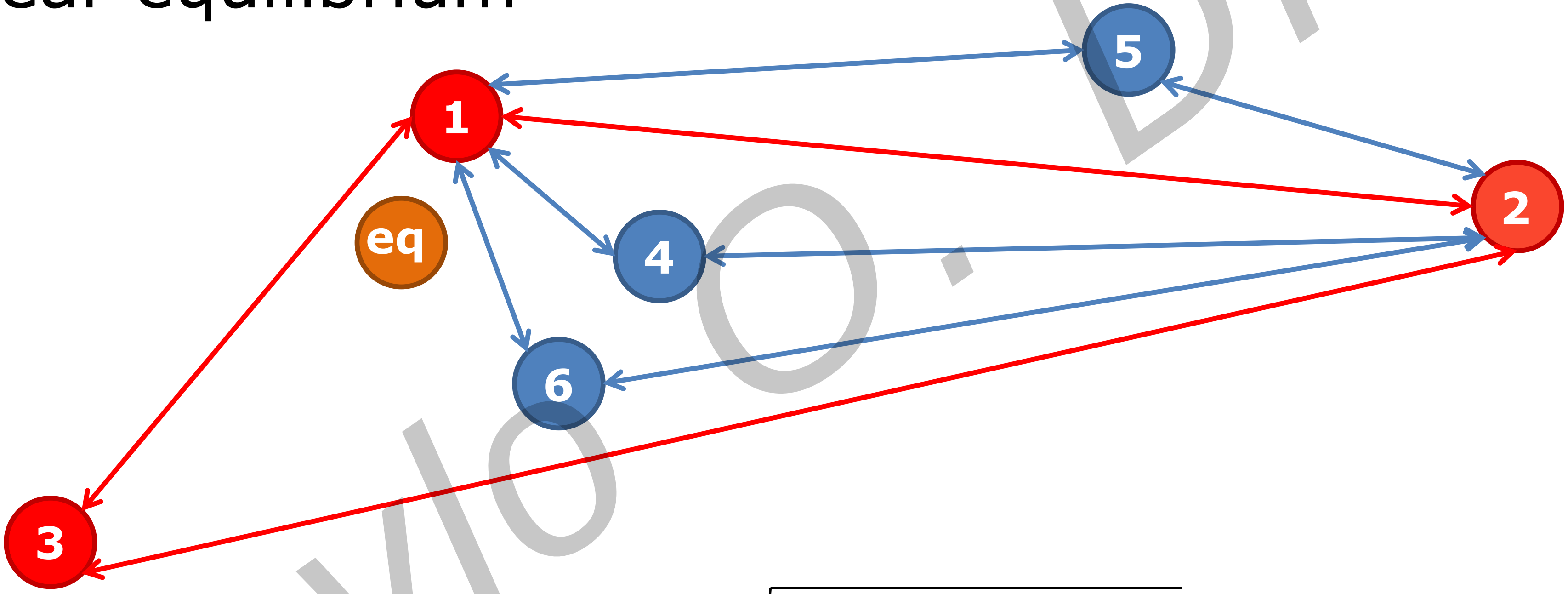
- Low-energy structures (<10,000 cm^{-1}) are most important:

  How to teach ML to favor low-energy region?

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Structure-based sampling

Take the first point to the training set, which is near equilibrium



$$D(\mathbf{M}_a, \mathbf{M}_i) = \sqrt{\sum_{\beta}^{N_{\mathbf{M}}} (\mathbf{M}_{i,\beta} - \mathbf{M}_{a,\beta})^2}$$

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Targeted Optimization of Hyperparameters

$$Y^{\mathrm{ML}}\left(\mathbf{M}_a\right) = \sum_{i=1}^{N_{train}} \alpha_i K\left(\mathbf{M}_a, \mathbf{M}_i\right)$$

$$K\left(\mathbf{M}_a, \mathbf{M}_i\right) = \exp\left(-\frac{\left\|\mathbf{M}_i - \mathbf{M}_a\right\|_2^2}{\enclose{circle}{2\sigma^2}}\right)$$

Gaussian kernel

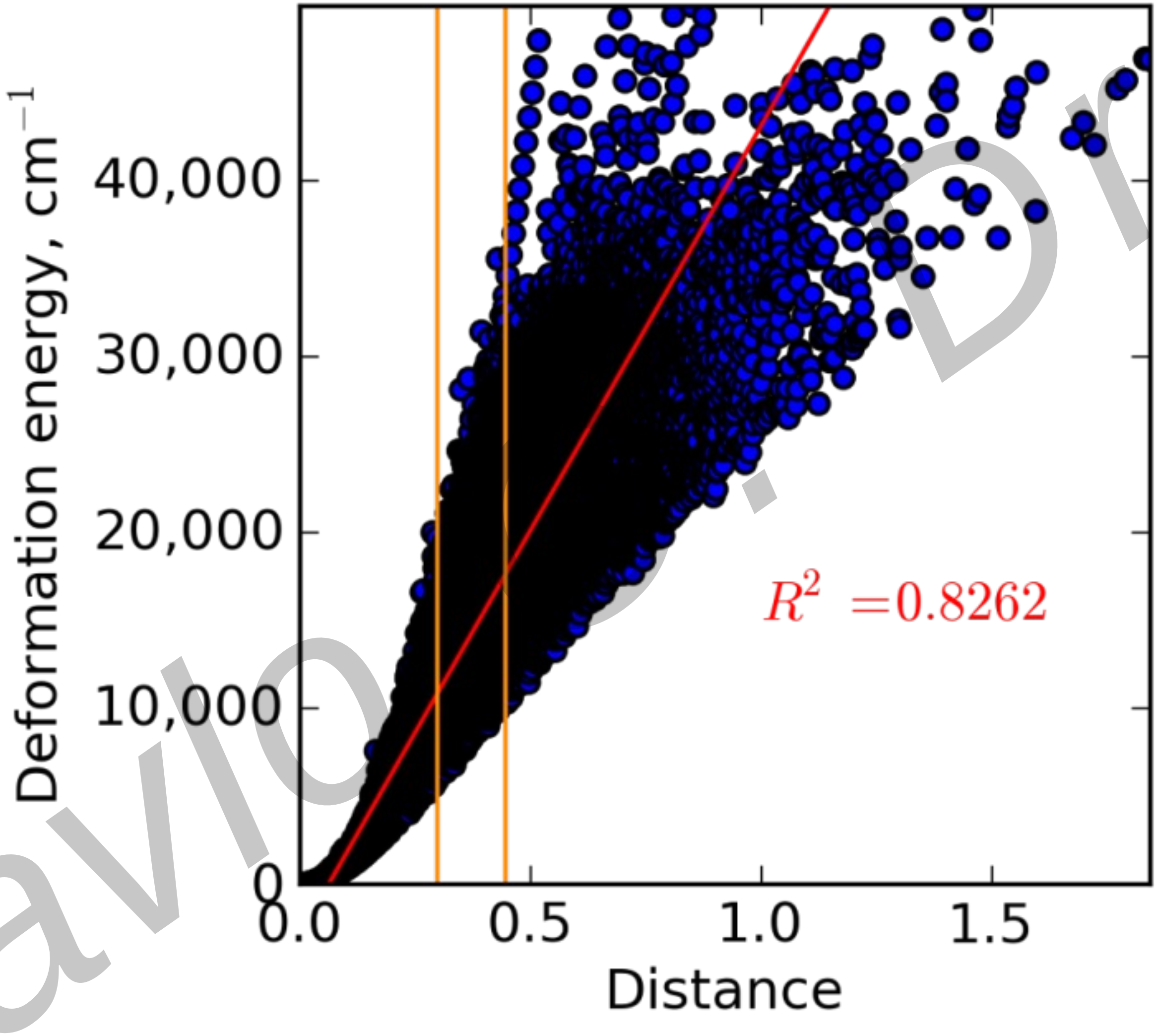
$$\min_{\alpha} \sum_{i}^{N_{train}} \left(Y_i^{est} - Y_i^{ref}\right)^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha}$$

$$\boldsymbol{\alpha} = \left(\mathbf{K} + \enclose{circle}{\lambda \mathbf{I}}\right)^{-1} \mathbf{Y}^{ref}$$

σ and λ are optimized by training ML model on 80% of the training set and minimizing error for the <10,000 cm^{-1} reference points in the remaining 20%

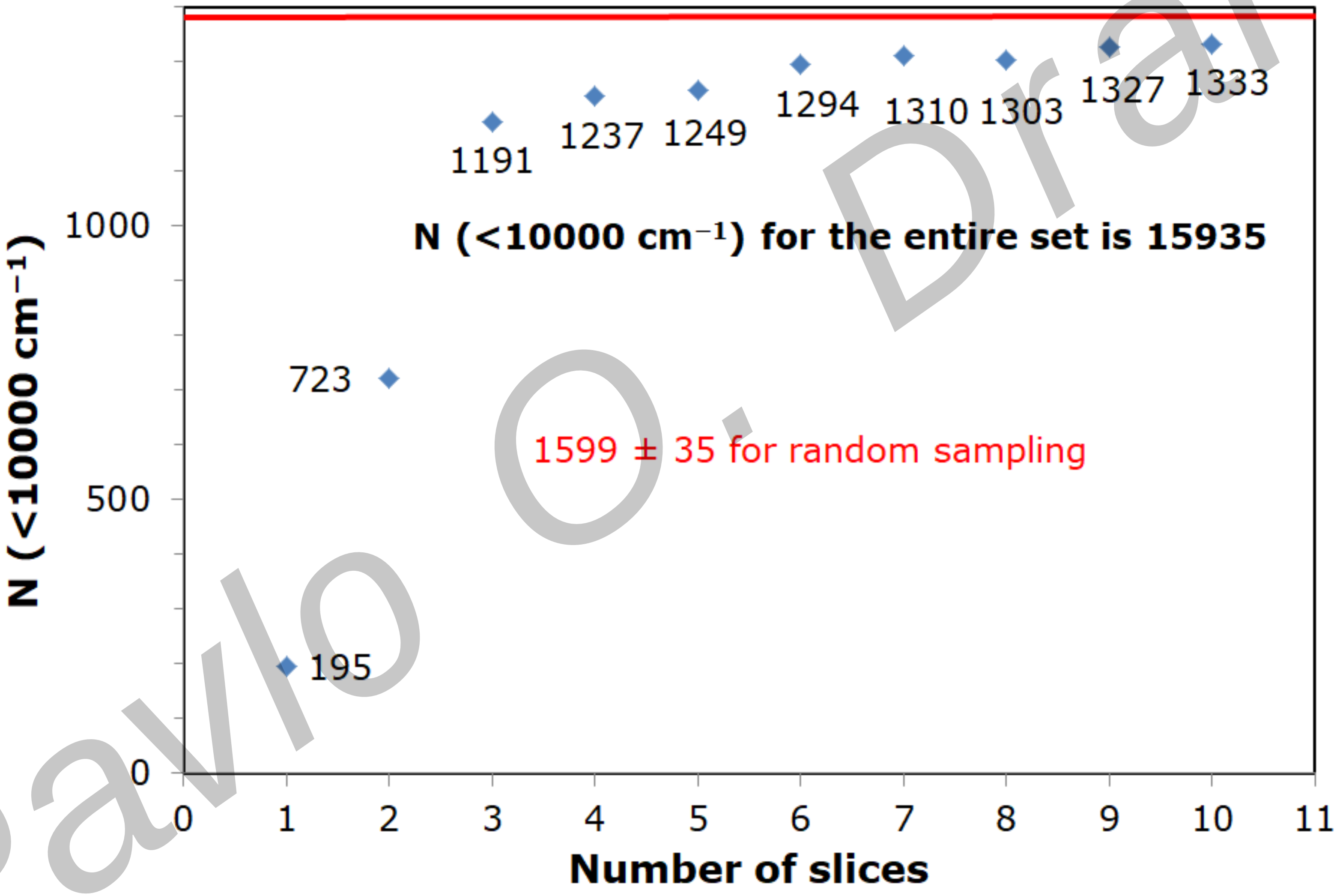
Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Slicing



Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Sampling 10% from 44819 Points



N (<10000 cm^{-1}) for the entire set is 15935

1599 ± 35 for random sampling

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

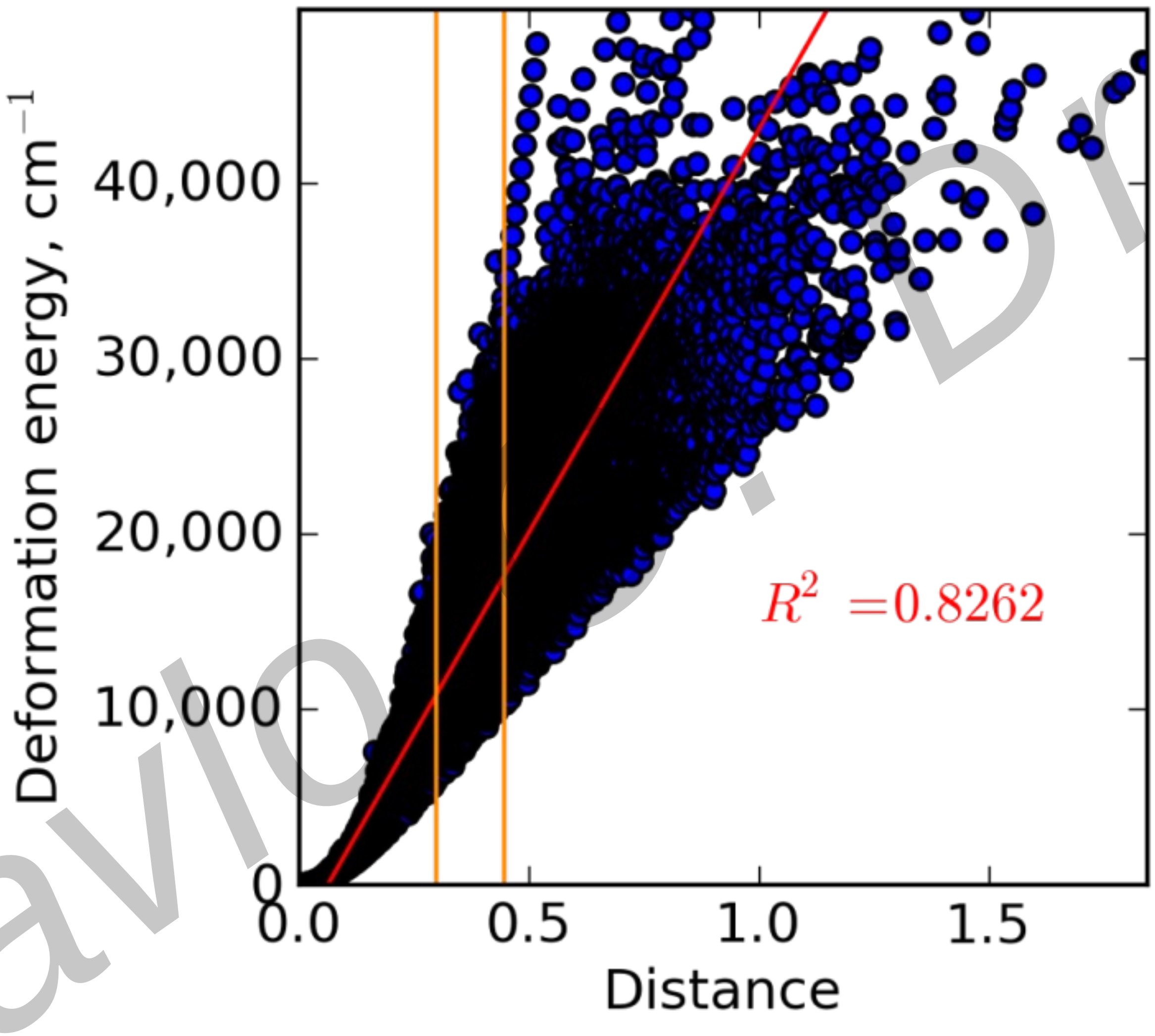
Effect of slices

wRMSEs in cm^{-1} for the 'out-of-sample' prediction sets for ML models trained on a fraction of 44819 grid points

Training set size	Unsliced	3 slices
50%	0.37	0.62
25%	1.60	2.58
10%	4.83	3.63

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

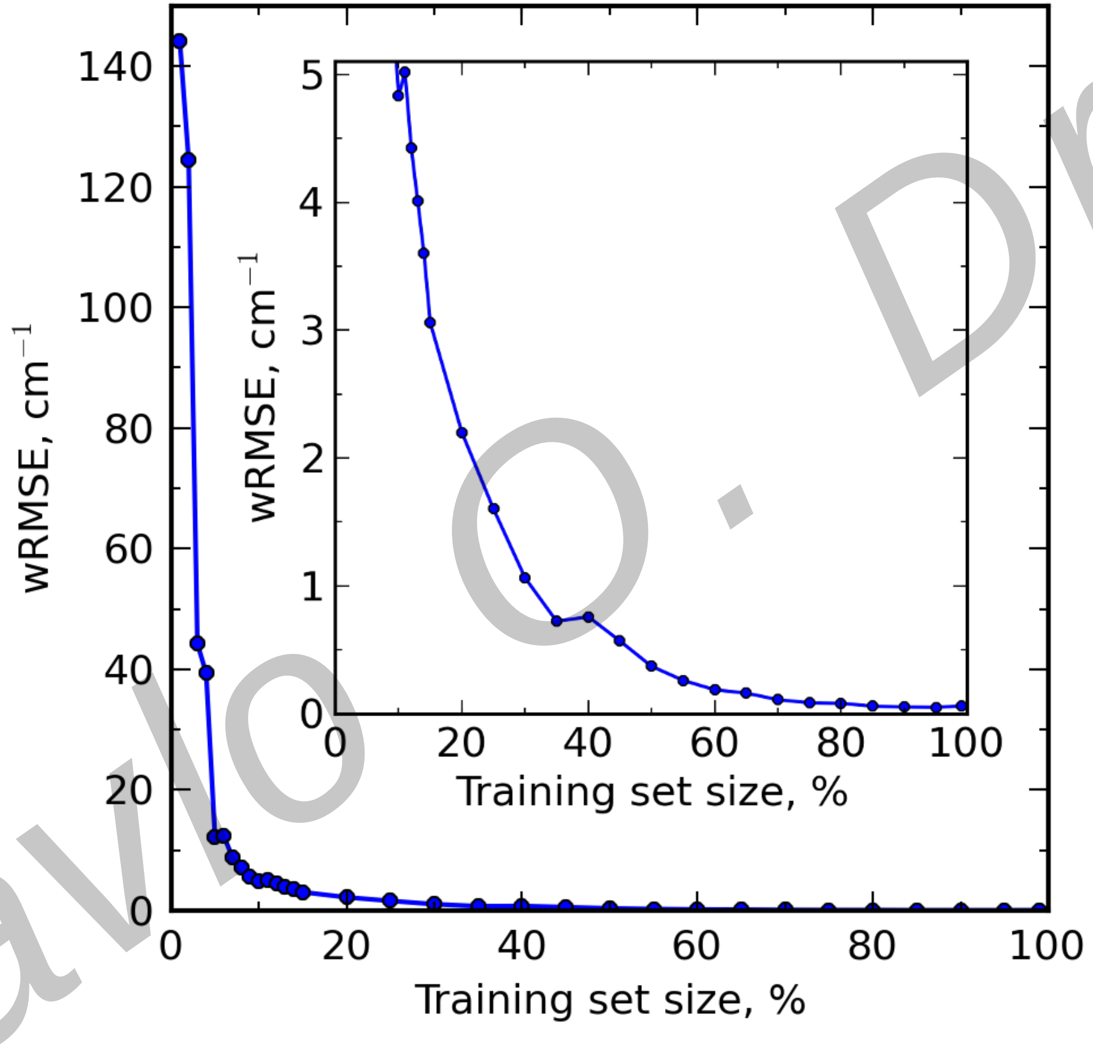
Slicing



Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

ML model performance

Learning curve



Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

ML models

Four-layer ML models, training sets obtained by:

1. 50%-ML – structure-based sampling of 50% of grid points
2. r50%-ML – random sampling of 50% of grid points
3. 25%-ML – structure-based sampling of 25% of grid points
4. 10%-ML – structure-based sampling of 10% of grid points
5. s10%-ML – structure-based sampling of 10% of grid points from **s**liced data

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

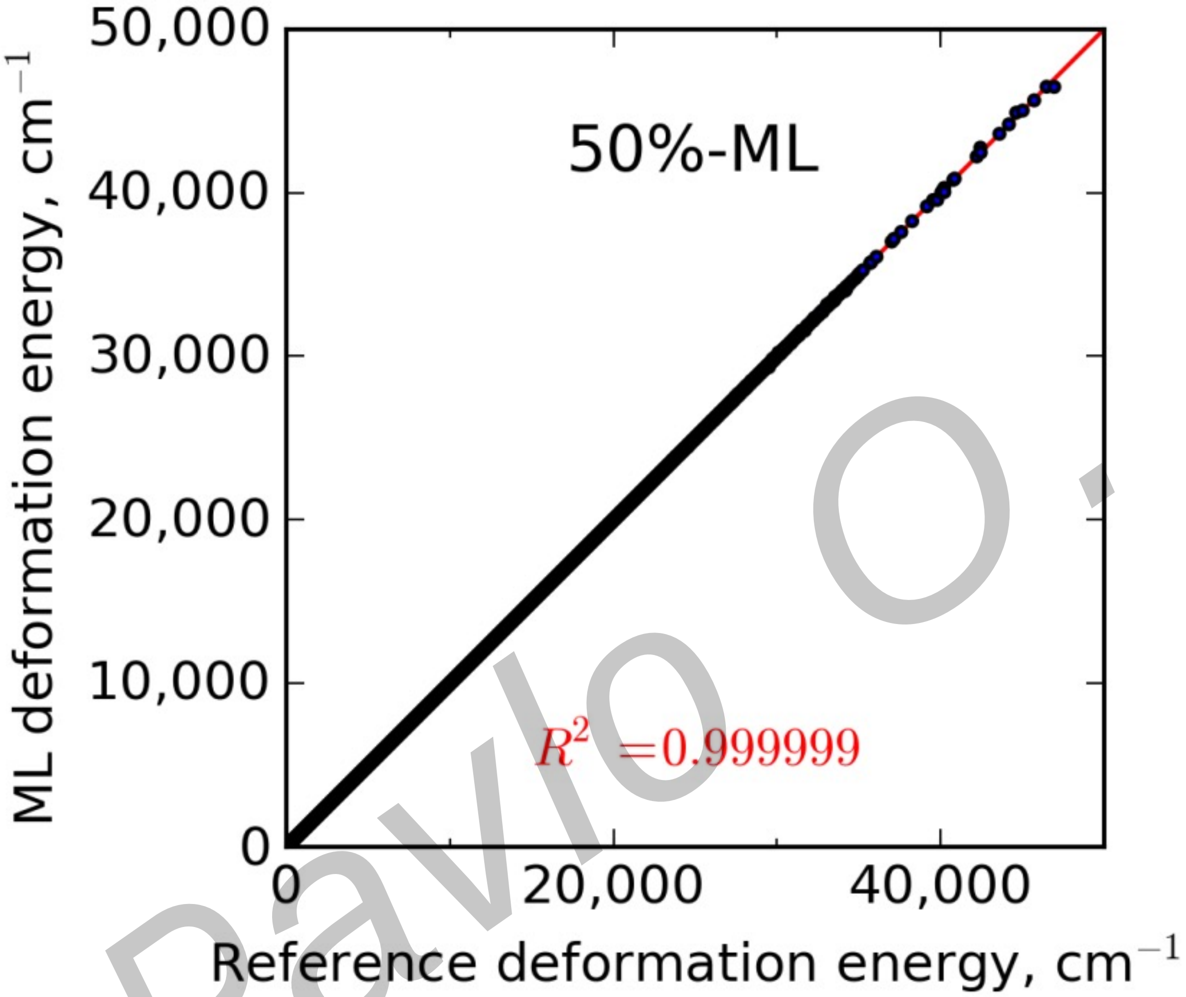
ML model performance

Errors in cm^{-1} for the ‘hold-out’ prediction sets

Model	wRMSE	RMSE	Largest outlier
50%-ML	0.37	6	476
r50%-ML	4.14	167	6919
25%-ML	1.60	16	1060
10%-ML	4.83	40	2859
s10%-ML	3.63	62	2190

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

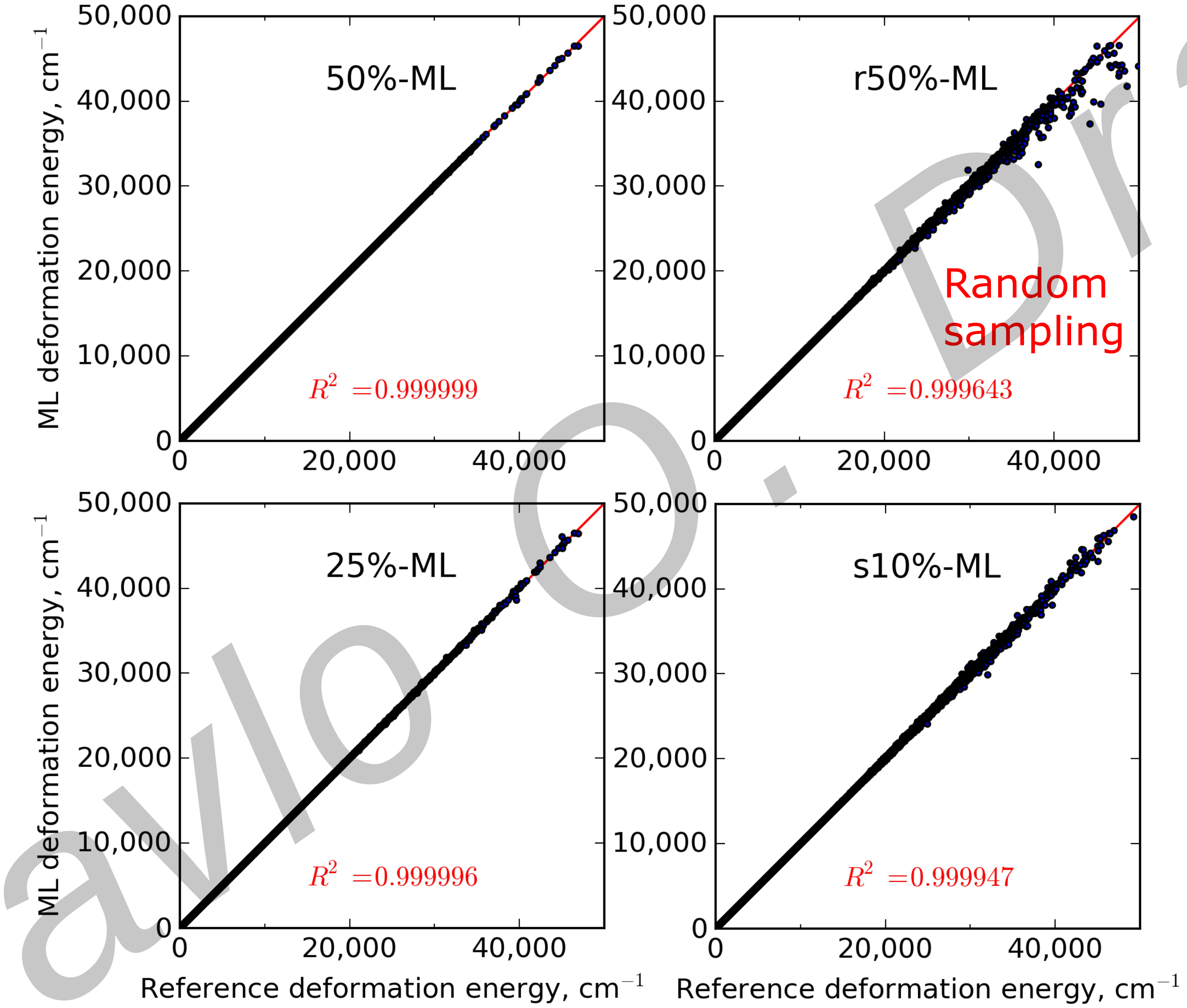
ML model performance



wRMSE, cm^{-1}	RMSE, cm^{-1}
0.37	6.23

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

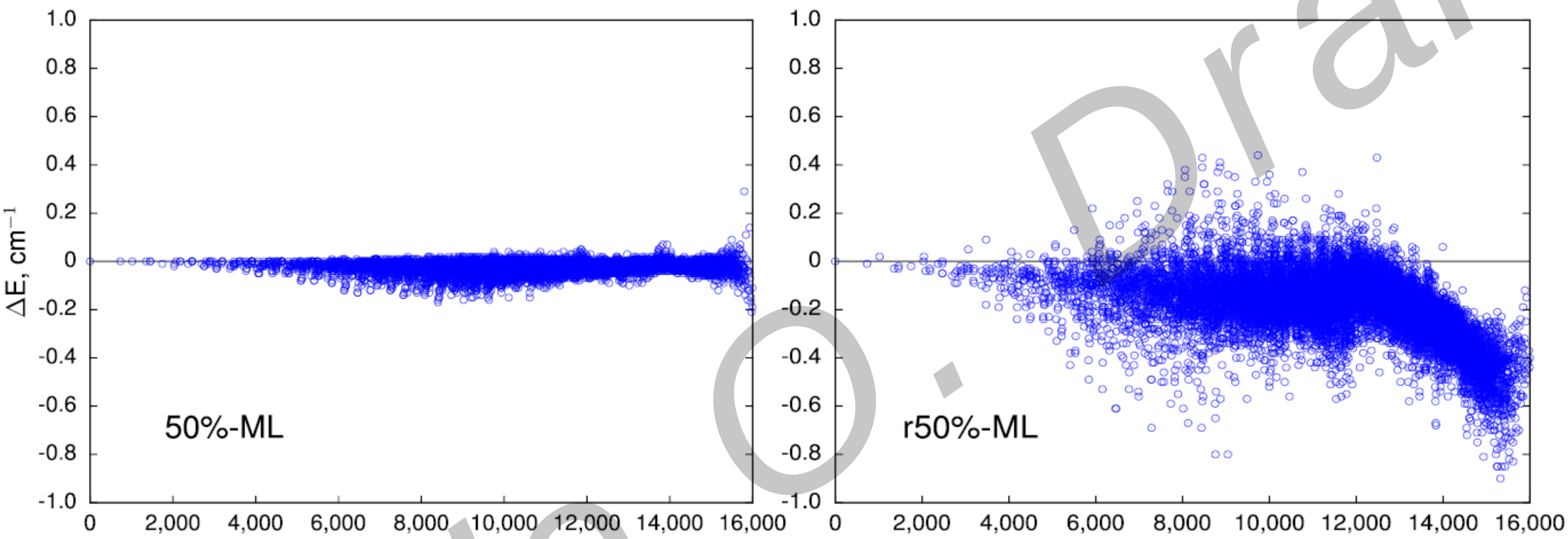
ML model performance



Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Vibrational levels

Accuracy of vibrational spectra

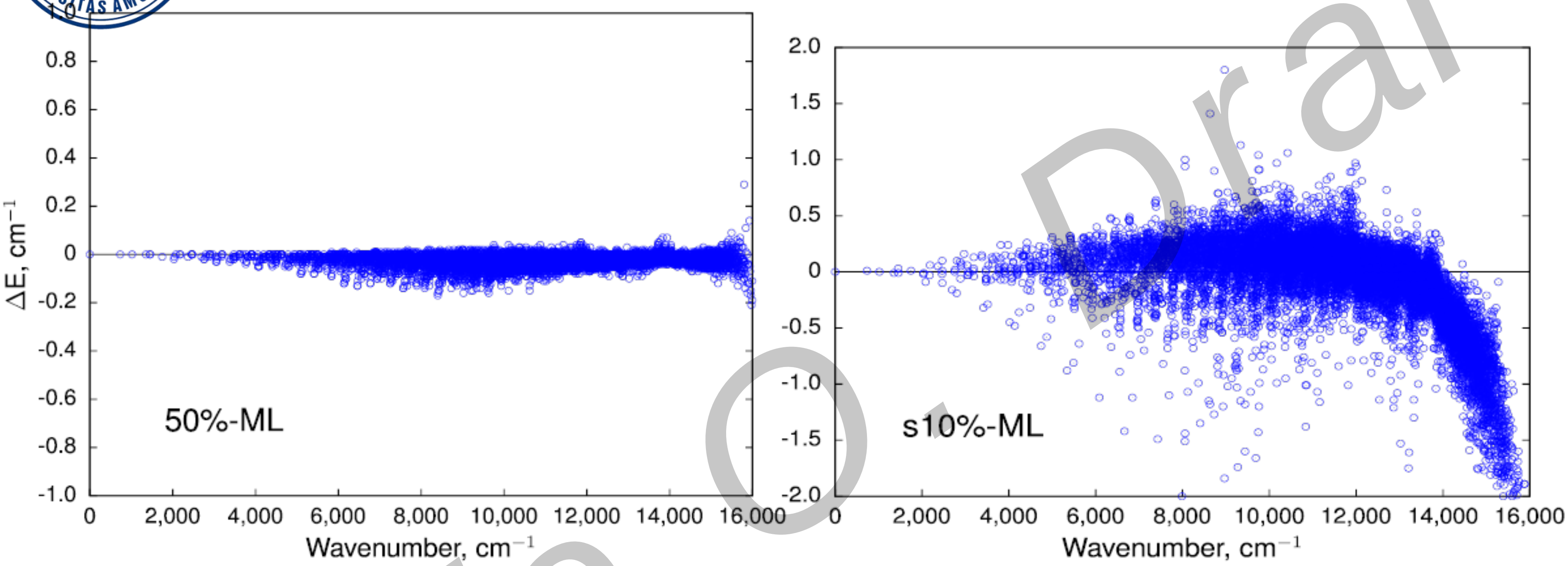


No of levels	Energy, cm^{-1}	RMSE, cm^{-1}
166	5,000	0.02
3,606	10,000	0.04

No of levels	Energy, cm^{-1}	RMSE, cm^{-1}
166	5,000	0.10
3,606	10,000	0.18

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Accuracy of vibrational spectra



No of levels	Energy, cm^{-1}	RMSE, cm^{-1}
166	5,000	0.02
3,606	10,000	0.04

No of levels	Energy, cm^{-1}	RMSE, cm^{-1}
166	5,000	0.14
3,606	10,000	0.28

Dral, Owens, Yurchenko, Thiel, *J. Chem. Phys.* **2017**, *146*, 244108

Hierarchical ML (hML) Combining Δ-ML with sampling

Automatic procedure to find optimal training points for each delta-model

$$\hat{y}_{\Sigma,i} = \sum_{M} \hat{y}_{M,i}(N_{\mathrm{tr},M})$$

$$\frac{e_{\Sigma}}{e_{\Sigma,\mathrm{th}}} + \frac{t_{\Sigma}/t_{\Sigma,\mathrm{max}}}{1 - s_{\mathrm{th}}} + \frac{p}{e_{\Sigma,\mathrm{th}}}$$

$$\frac{\hat{y}_{M,i}(N_{\mathrm{tr},M})}{\hat{y}_{1,i}(N_{\mathrm{tr},M})} \approx r_{M/1,i} = \frac{y_{M,i}}{y_{1,i}}$$

$$\mathrm{hML} = \Delta_{0}^{\mathrm{MP2}} + \Delta_{\mathrm{MP2}}^{\mathrm{VTZ}} + \Delta_{\mathrm{VTZ}}^{\mathrm{VQZ}} + \Delta_{\mathrm{VQZ}}^{\mathrm{CBS}} + \Delta_{0}^{\mathrm{SR}} + \Delta_{0}^{\mathrm{CV}} + \Delta_{0}^{\mathrm{HO}} + \Delta_{0}^{\mathrm{DBOC}}$$

P. O. Dral, A. Owens, A. Dral, G. Csányi, *J. Chem. Phys.* **2020**, *152*, 204110

Hierarchical ML (hML) Combining Δ-ML with sampling

$$\hat{y}_{\Sigma,i} = \sum_{M} \hat{y}_{M,i}(N_{\mathrm{tr},M})$$

Automatic procedure to find optimal training points for each delta-model

$$\frac{e_{\Sigma}}{e_{\Sigma,\mathrm{th}}} + \frac{t_{\Sigma}/t_{\Sigma,\max}}{1 - s_{\mathrm{th}}} + \frac{p}{e_{\Sigma,\mathrm{th}}}$$

$$\frac{\hat{y}_{M,i}(N_{\mathrm{tr},M})}{\hat{y}_{1,i}(N_{\mathrm{tr},M})} \approx r_{M/1,i} = \frac{y_{M,i}}{y_{1,i}}$$

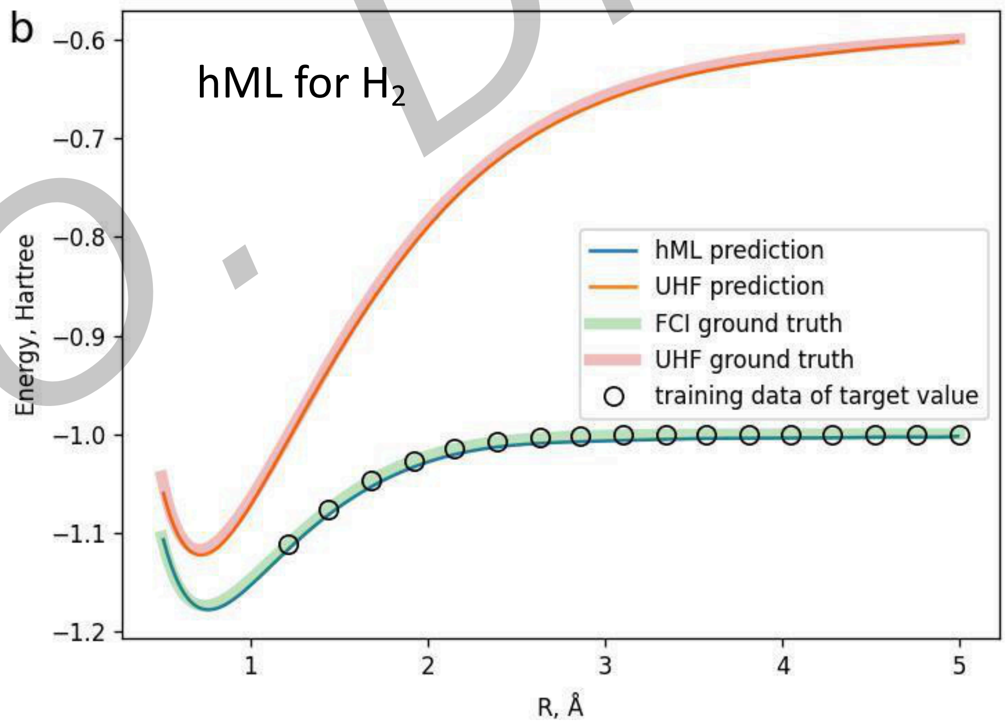


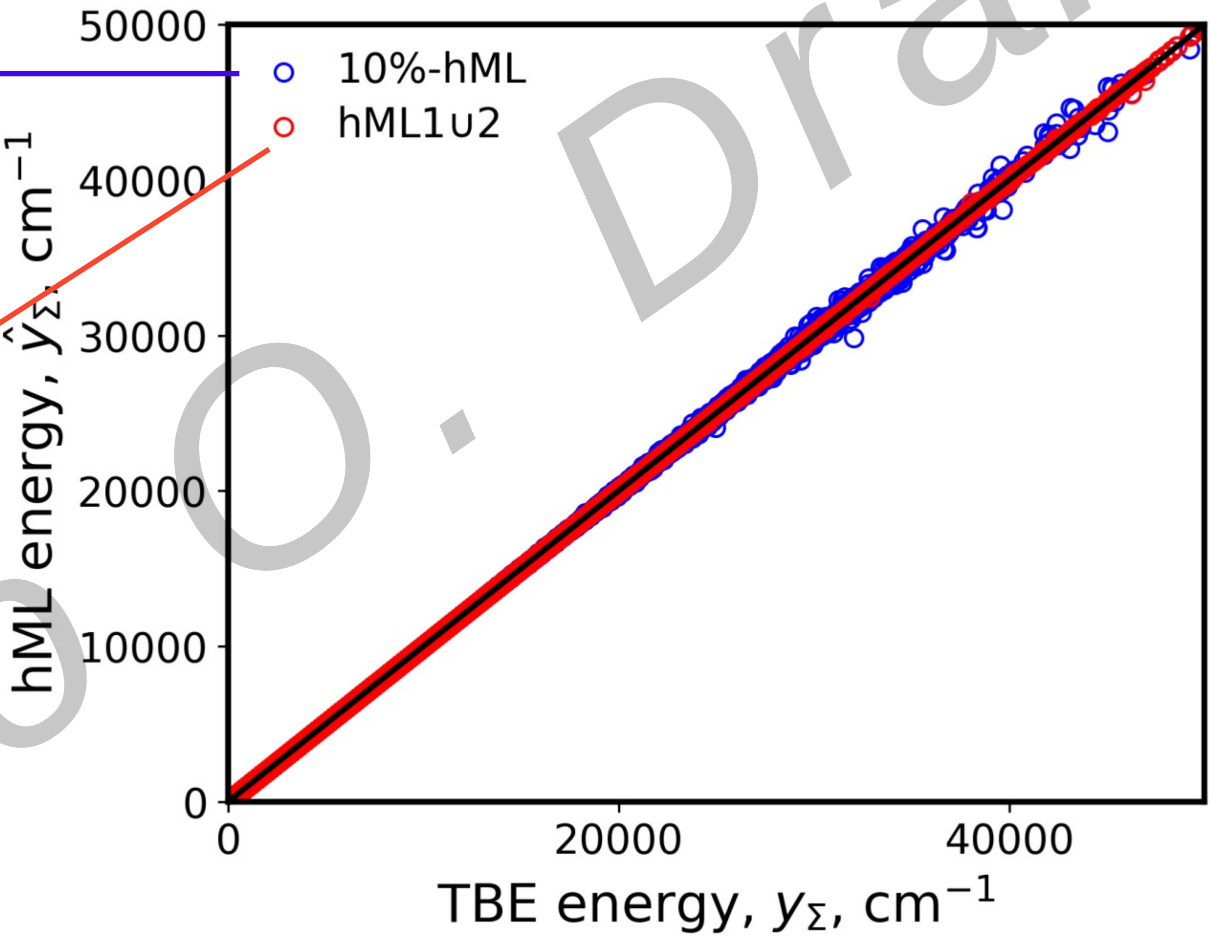
Figure: Pavlo O. Dral, Tetiana Zubatiuk, Bao-Xin Xue, Learning from multiple quantum chemical methods: Δ-learning, transfer learning, co-kriging, and beyond. In *Quantum Chemistry in the Age of Machine Learning*, Pavlo O. Dral, Ed. Elsevier: 2022, in press. Paperback ISBN: 9780323900492

Hierarchical ML (hML)
Combining Δ-ML with sampling

Pure ML-model trained on
saved 90% of CPU-time

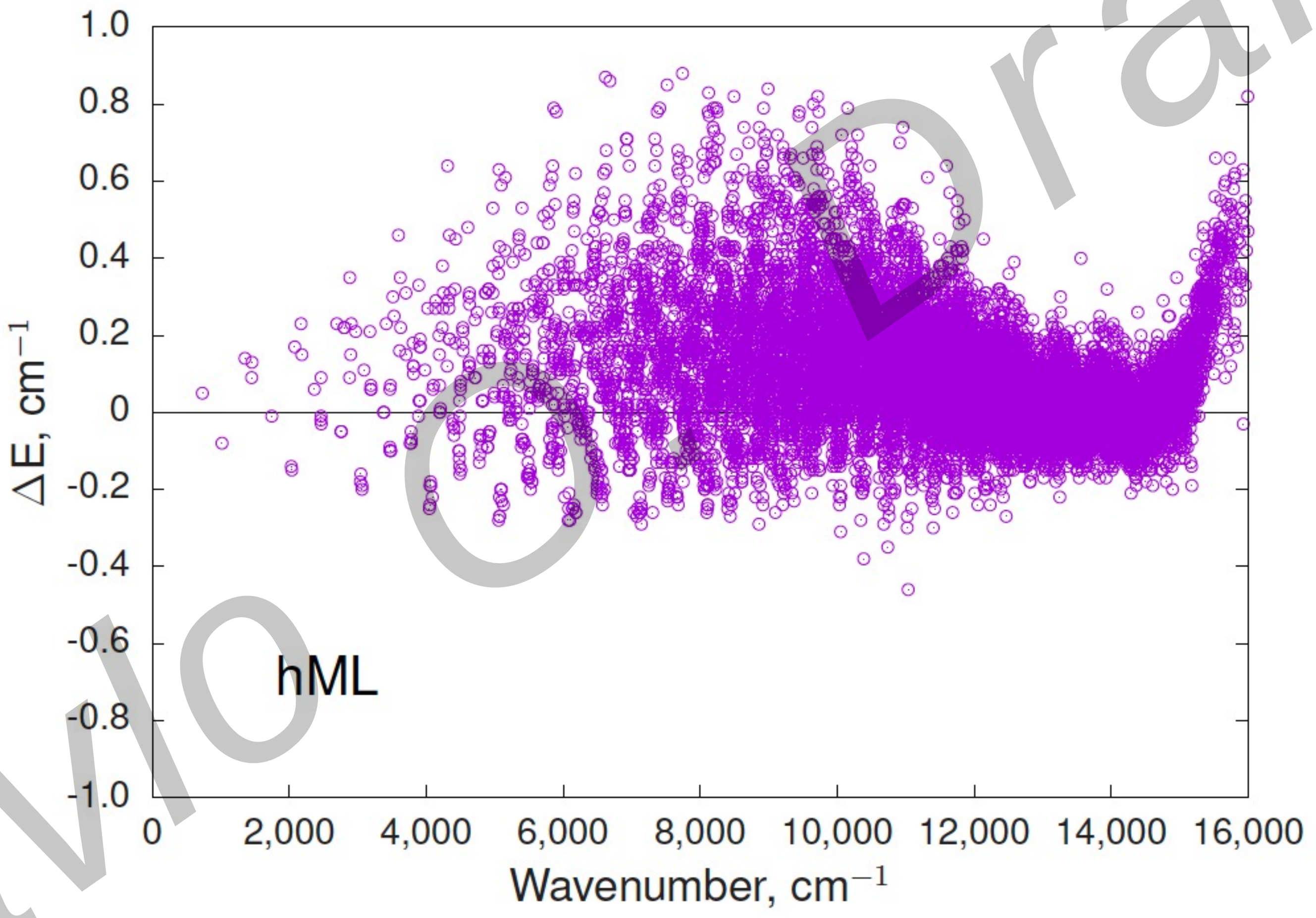
Weighted RMSE:
3.49 cm^{-1} = 0.01 kcal/mol

Δ-ML models:
saved 99% of CPU-time
Weighted RMSE:
1.12 cm^{-1} = 0.003 kcal/mol



P. O. Dral, A. Owens, A. Dral, G. Csányi, *J. Chem. Phys.* **2020**, *152*, 204110

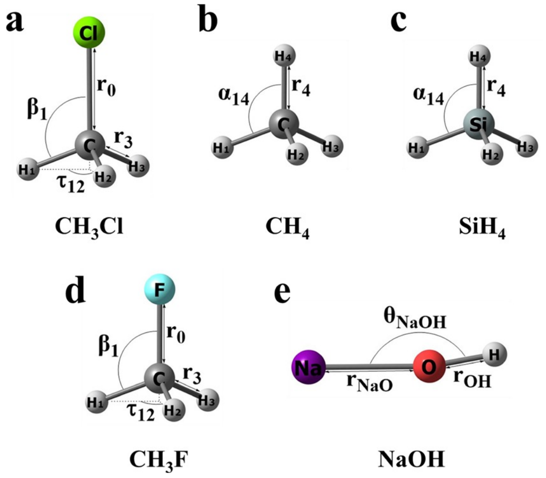
Accuracy of vibrational spectra



P. O. Dral, A. Owens, A. Dral, G. Csányi, *J. Chem. Phys.* **2020**, *152*, 204110

VIB5 database

Global, hole-free, highly accurate!

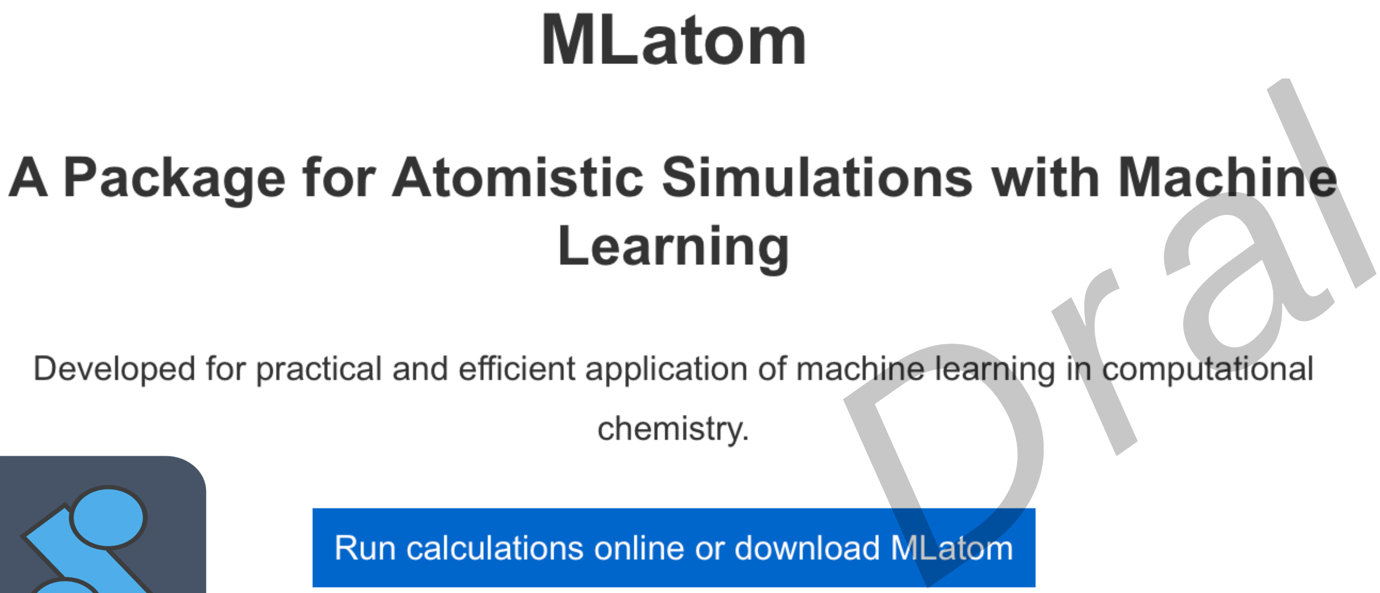


CH_3Cl CH_4 SiH_4 CH_3F NaOH

Molecule	Grid size
CH_3Cl	44819
CH_4	97217[a]
SiH_4	84002
CH_3F	82653
NaOH	15901
Total: 5 molecules	324592[a]

No.	Key	Description	Units
1	XYZ	Nuclear positions in Cartesian coordinates	Å
2	INT	Nuclear positions in internal coordinates	Å; degree
3	HF-TZ	Total energy at HF/cc-pVTZ	Hartree
4	HF-QZ	Total energy at HF/cc-pVQZ	Hartree
5	MP2	Total energy at MP2/cc-pVTZ	Hartree
6	CCSD-T	Total energy at CCSD(T)/cc-pVQZ	Hartree
7	TBE	Theoretical best estimate of ab initio deformation energies	cm^{-1}
8	MP2_grad_xyz	Energy gradient in Cartesian coordinates at MP2/cc-pVTZ	Hartree/Å
	MP2_grad_int	Energy gradient in internal coordinates at MP2/cc-pVTZ	Hartree/Å; Hartree/degree
	CCSD-T_grad_xyz	Energy gradient in Cartesian coordinates at CCSD(T)/cc-pVQZ	Hartree/Å
	CCSD-T_grad_int	Energy gradient in internal coordinates at CCSD(T)/cc-pVQZ	Hartree/Å; Hartree/degree
	CBS	Deformation energies at CCSD(T)-F12b/CBS	cm^{-1}
	VTZ	Deformation energies at CCSD(T)-F12b/cc-pVTZ-F12 (only for CH_3Cl molecule)	cm^{-1}
	VQZ	Deformation energies at CCSD(T)-F12b/cc-pVQZ-F12 (only for CH_3Cl molecule)	cm^{-1}
	CV	Deformation energy corrections to account for core-valence electron correlation	cm^{-1}
	HO	Deformation higher-order coupled cluster terms beyond perturbative triples	cm^{-1}
	SR	Deformation scalar relativistic (SR) effects	cm^{-1}
	DBOC	Deformation diagonal Born–Oppenheimer corrections (only for CH_3Cl, CH_4, CH_3F, and NaOH molecules)	cm^{-1}
19	MP2-aQZ	Deformation energies at MP2/aug-cc-pVQZ (only for CH_3Cl molecule)	cm^{-1}

L. Zhang, S. Zhang, A. Owens, S. N. Yurchenko, P. O. Dral. *Sci. Data* **2022,** *9*, 84



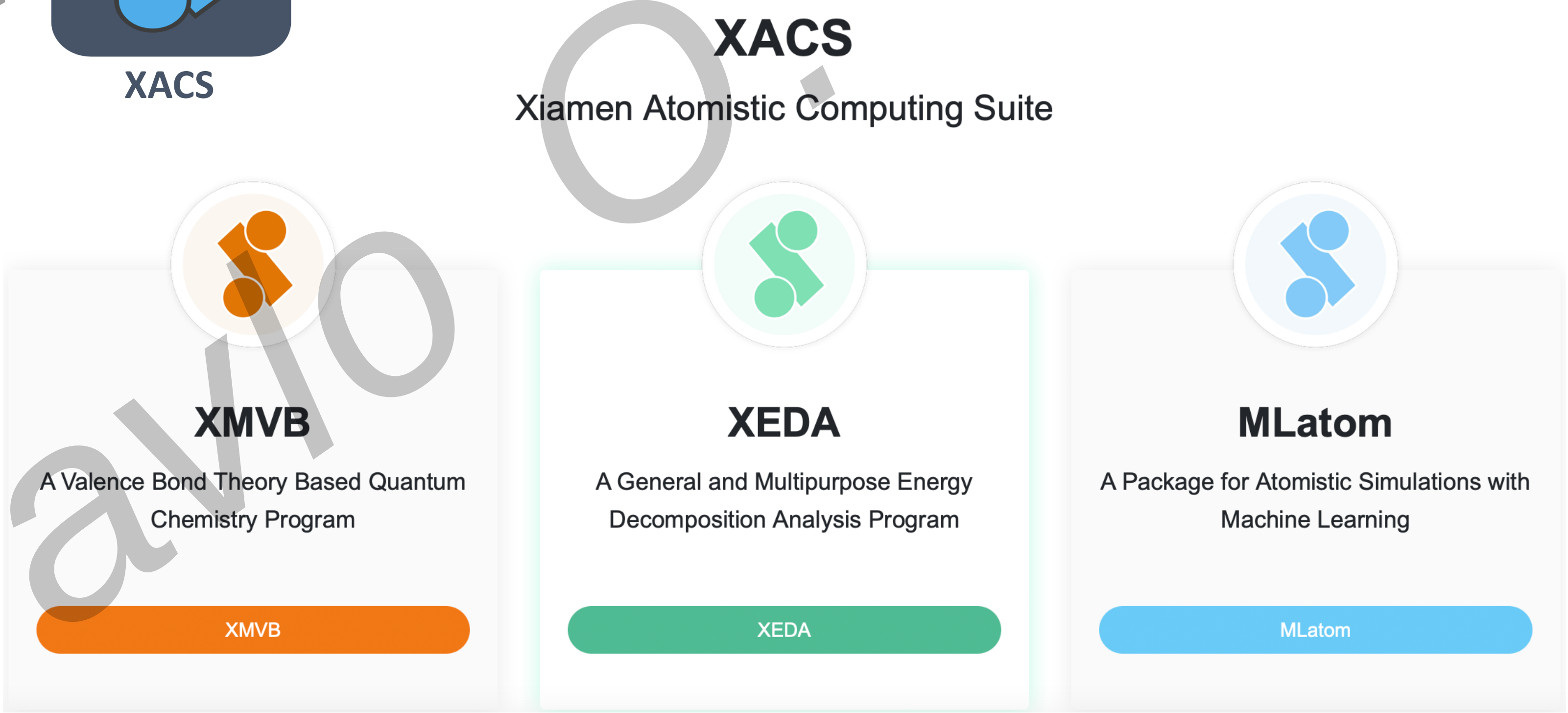
MLatom

A Package for Atomistic Simulations with Machine Learning

Developed for practical and efficient application of machine learning in computational chemistry.

MLatom.com

XACS



XACS

Xiamen Atomistic Computing Suite

XMVB

A Valence Bond Theory Based Quantum Chemistry Program

XEDA

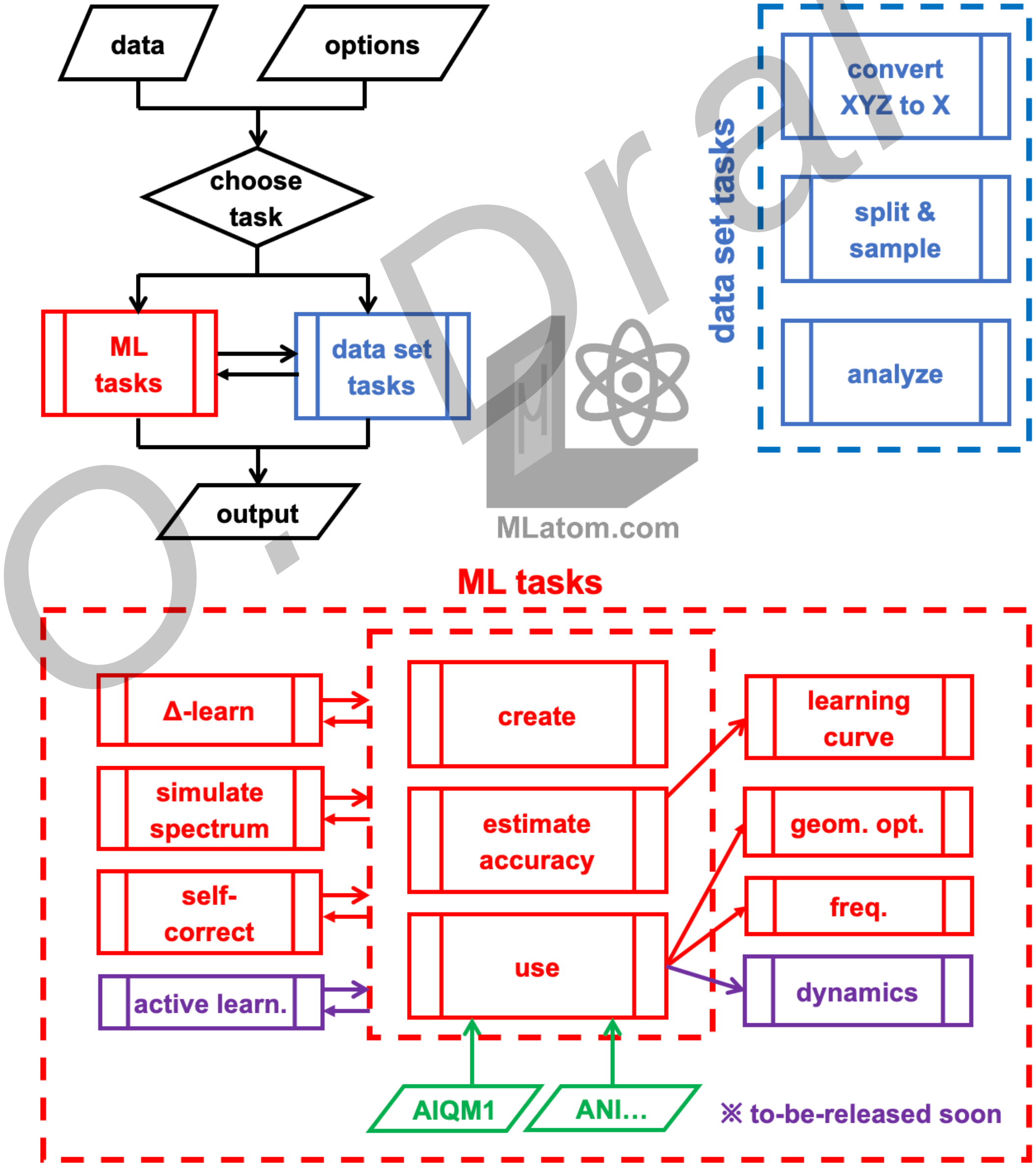
A General and Multipurpose Energy Decomposition Analysis Program

MLatom

A Package for Atomistic Simulations with Machine Learning

Flowchart

P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou,
M. Pinheiro Jr, J. Huang, M. Barbatti, **2021**, *379*, 27



3rd-party interfaces

DeePMD-kit	deepmodeling.com
GAP	libatoms.github.io/GAP
sGDML	sgdml.org
TorchANI	aiqm.github.io/torchani
PhysNet	github.com/MMunibas/PhysNet
ASE	wiki.fysik.dtu.dk/ase
Hyperopt	hyperopt.github.io/hyperopt
Sparrow	scine.ethz.ch/download/sparrow
NX	newtonx.org
Gaussian	gaussian.com
MNDO	mndo.kofo.mpg.de

Acknowledgements

Thank you for your attention!

Funding:

PCOSS

国家自然科学基金委员会
National Natural Science Foundation of China