Available online at **www.joac.info**

ISSN: 2278-1862

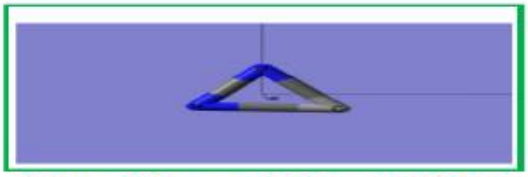
Journal of Applicable Chemistry

2019, 8 (1): 435-461

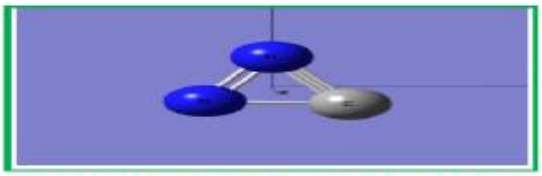
(International Peer Reviewed Journal)

New Chemistry News

$^{-}N=C=N^{-}$



New News of Chem (NNC)



ChemNewsNew (CNN)

Machine learning **in CQC**	**Information Source (is)** **ACS.org**

Screening metallic MOFs	Interatomic potentials		**MachLrn**

Task

- Electrical transport properties of
- MOFs

- DFT+PBE does not capture physics responsible for some part of potential affecting electrical transport properties

Methods

- Machine learning
- Statistical multi-voting
- ab initio calculations

MachLrnMethods

- Logistic regression (LR)
- Support vector classification (SVC)
- Neural network (NN)
- Random forest (RF)

Semi-local DFT→ to identify most promising candidates

Metallic Metal–Organic Frameworks Predicted by the Combination of Machine Learning Methods and Ab Initio Calculations

J. Phys. Chem. Lett., 2018, 9 (16), 4562–4569
DOI: 10.1021/acs.jpclett.8b01707

Yuping He, Ekin D. Cubuk, Mark D. Allendorf, Evan J. Reed

Material preparation	**Nano**	**MachLrn**

Task

Formation of Subnanometer Substructures in Nanoassembliesmaterials

Methods		
Expt ➔ Small-angle X-ray scattering (SAXS) ➔ X-ray absorption near-edge structure (XANES) spectroscopy	**CQC** o ab initio	**MachLrn** o NNs

Subnanometer Substructures in Nanoassemblies Formed from Clusters under a Reactive Atmosphere Revealed Using Machine Learning

J. Phys. Chem. C, 2018, 122 (37), 21686–21693
DOI: 10.1021/acs.jpcc.8b07952

Janis Timoshenko, Avik Halder, Bing Yang, Soenke Seifert, Michael J. Pellin, Stefan Vajda, and Anatoly I. Frenkel

SPR	**DFT**	**MachLrn**
Task	**Compounds**	
➔ Predictive universal SPR	o Over 12, 000 experimentally synthesized and characterized ones	

Machine learning-assisted discovery of solid Li-ion conducting materials

Chem. Mater., Just Accepted Manuscript
DOI: 10.1021/acs.chemmater.8b03272

Austin D. Sendek, Ekin D. Cubuk, Evan R Antoniuk, GowoonCheon, Yi Cui, and Evan J. Reed

Vibrational Properties			**MachLrn**
Methods	**Accuracy**		
➔ random-forest algorithm	- 121 different mechanically stable structures of KZnF3 reaches mean absolute error of 0.17 eV/Å2		

Vibrational Properties of Metastable Polymorph Structures by Machine Learning

J. Chem. Inf. Model., Article ASAP
DOI: 10.1021/acs.jcim.8b00279

Fleur Legrain, Ambroise van Roekeghem, Stefano Curtarolo,Jesús Carrete, Georg K. H. Madsen, Natalio Mingo

HOMO–LUMO			**MachLrn**
Task	**Methods**		Descriptors
HOMO level and HOMO–LUMO gapPrediction			
Method ➔ DFT	o MachLrn o molSimplify LASSO o Kernel ridge regression (KRR) o NNs		o Heuristic o Topological o Revised Autocorrelation

+ Performance of NNstrained with20–30 features is superior to RAC full (153 feature set)
  - full DFT evaluation of 15, 000 molecule design space requires n-days
+ MachLrn is fast (a few minutes)

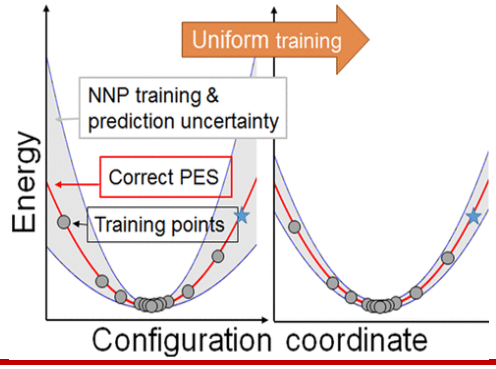
Strategies and Software or Machine Learning Accelerated Discovery in Transition Metal Chemistry

Ind. Eng. Chem. Res., **2018**, *57* (42), 13973–13986
DOI: 10.1021/acs.iecr.8b04015

Aditya Nandy, ChenruDuan, Jon Paul Janet, Stefan Gugler, and Heather J. Kulik

Interatomic potential	**MachLrn**
Task	

➔ NN interatomic potential -- a promising next-generation atomic potential
+ Self-learning capability and universal mathematical structure
- NNP suffers from highly inhomogeneous feature-space sampling in the training set



Toward Reliable andransferable Machine Learning Potentials: Uniform Training by Overcoming Sampling Bias

J. Phys. Chem. C, **2018**, *122* (39), 22790–22795
DOI: 10.1021/acs.jpcc.8b08063

WonseokJeong, Kyuhyun Lee, DongsunYoo, Dongheon Lee, and Seungwu Han

Correlation energies	**MachLrn**
Task	

➔ To predict structure correlation energies (which is a measure of interactions between electrons enabling chemists to model behavior of molecule(s)

Machine learning predicts electron energies

C&EN, 2018, *96* (33), 77
DOI: 10.1021/cen-09633-scicon3

Sam Lemonick

Thermodynamic properties			**MachLrn**
Task	**Prediction**		

Thermodynamic properties: [liquid densities, heats of vaporization, heat capacities, vapor–liquid equilibrium curves, critical temperatures, critical densities, surface tensions;]

+ Feasibility of expanding predictions beyond simulation using a machine learning model

Predicting Thermodynamic Properties of Alkanes by High-Throughput Force Field Simulation and Machine Learning

J. Chem. Inf. Model., Article ASAP
DOI: 10.1021/acs.jcim.8b00407

Zheng Gong, Yanze Wu, Liang Wu, and Huai Sun

Energy		MachLrn
Task	**Prediction**	
forces and energy of a molecule with only XYZ		

- ➔ Accurate NeurAlnetworKengINe for Molecular Energies (ANAKIN-ME)
- ➔ Third version called ANI-1ccx
  - Accuracy of CCSD(T) in the computational time of force fields

Machine learning offers fast, accurate calculations

C&EN, 2018, *96* (34), 7–7
DOI: 10.1021/cen-09634-scicon5

Sam Lemonick

CO_2 capture			MachLrn
Task		**Methods- multiscale**	
CO_2 capture enhancement metrics of MOF Effect of pore chemical/topological features		➔ DFT ➔ Grand canonical Monte Carlo ➔ machine learning	

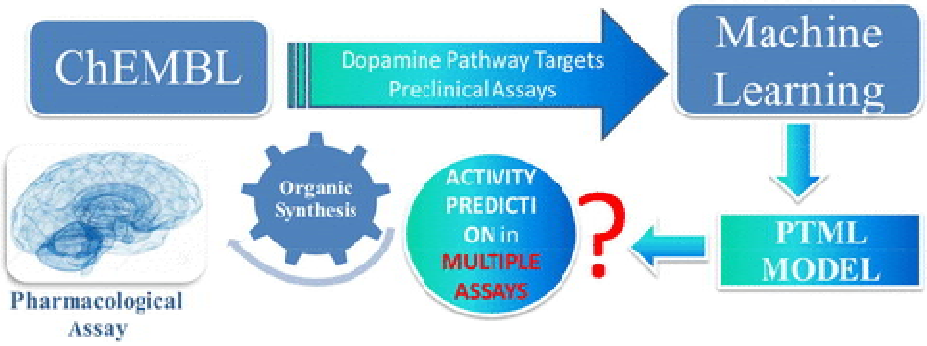
Role of Pore Chemistry and Topology in the CO_2 Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning

Chem. Mater., **2018**, *30* (18), 6325–6337
DOI: 10.1021/acs.chemmater.8b02257

Ryther Anderson, Jacob Rodgers, Edwin Argueta, AchayBiong, Diego A. Gómez-Gualdrón

IC_{50}, EC_{50}, K_i			MachLrn
Task Prediction of IC_{50}, EC_{50}, K_i Data Set large ChEMBL data set of preclinical assays [of compounds targeting dopamine pathway proteins]	**Methods** o NNs o Random Forest o Deep Learning	Perturbation Theory/Machine Learning (PTML) linear model of multiple pharmacological parameters + 50,000 cases with accuracy of 70–91% in training and external validation series	

- o Organic synthesis, chemical characterization, and pharmacological assay of a new series of l-prolyl-l-leucyl-glycinamide (PLG) peptidomimetic compoundsfor the first time
- o Molecular docking study for some of these compounds with software Vina AutoDock.
- o Perturbation Theory/Machine Learning (PTML) -- linear model of multiple pharmacological parameters

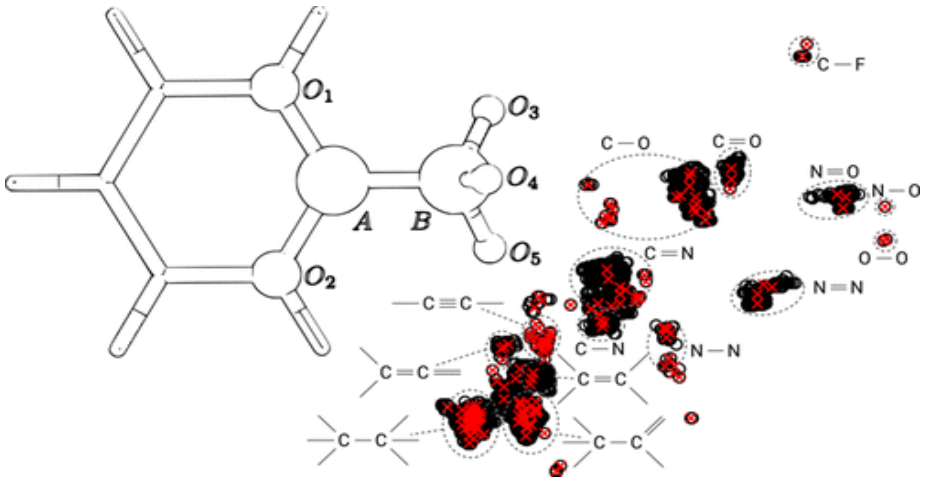


Perturbation Theory/Machine Learning Model of ChEMBL Data for Dopamine Targets: Docking, Synthesis, and Assay of New l-Prolyl-l-leucyl-glycinamide Peptidomimetics

ACS Chem. Neurosci., **2018**, *9* (11), 2572–2587
DOI: 10.1021/acschemneuro.8b00083

Joana Ferreira da Costa, David Silva, Olga Caamaño, José M. Brea, Maria Isabel Loza, Cristian R. Munteanu, Alejandro Pazos, Xerardo García-Mera, , Humbert González-Díaz

Atomization energies	**MachLrn**
Task	**Information**
▪ To automatize parametrization process of DFTB • To improve transferability and accuracy using large quantum chemical data	▪ Training with energies, forces of equilibrium and nonequilibrium structures of2100 molecules ▪ Testset:~130, 000 organic molecules containing O, N, C, H, and F atoms. ▪ Result: Atomization energies of the reference method can be reproduced within an error of ~2.6 kcal/molindicating drastic improvement over standard DFTB.
Methods ➔ Density-functional tight-binding (DFTB) method ➔ Unsupervised machine learning	✓ Generalized pair-potentials + Chemical environment is included during the learning process →specific effective two-body potentials

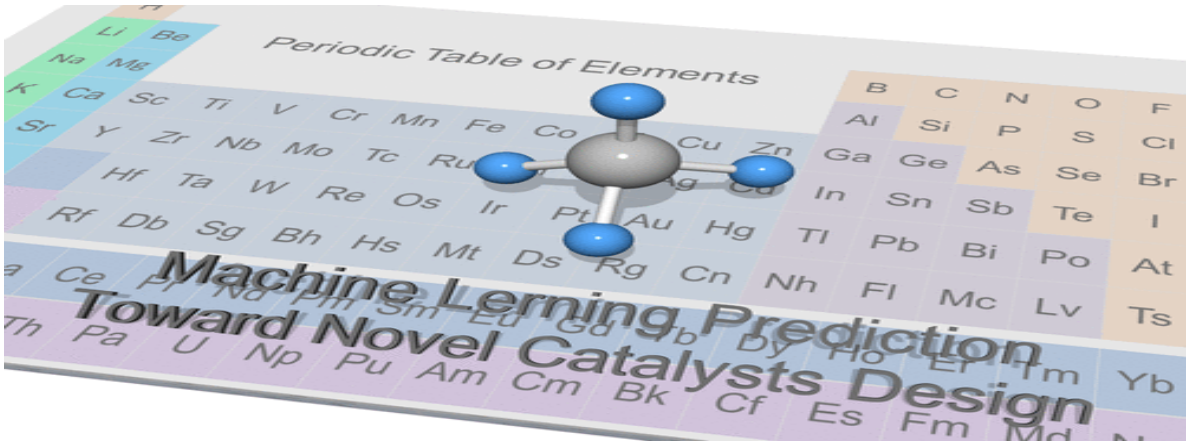


Generalized Density-Functional Tight-Binding Repulsive Potentials from Unsupervised MachineLearning

J. Chem. Theory Comput., **2018**, *14* (5), 2341–2352
DOI: 10.1021/acs.jctc.7b00933

Julian J. Kranz, Maximilian Kubillus, Raghunathan Ramakrishnan, O. Anatole von Lilienfeld, Marcus Elstner

Adsorption energies		MachLrn
Task	**Methods**	**X**
➔ To predict adsorption energies of CH_4 related species on the Cu-based alloys ➔ To predict the catalytic performances of the solid catalysts	▪ Ordinary linear regression ▪ Boosting regression (GBR) ▪ Extra tree regression (ETR)	12 descriptors ETR (RMSEs for energies below 0.3 eV)>> [OLS, GBR]

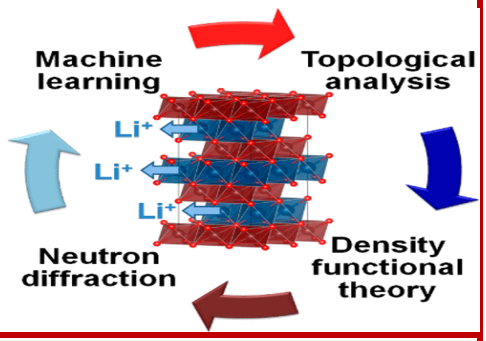


Toward Effective Utilization of Methane: Machine Learning Prediction of Adsorption Energies onMetal Alloys

J. Phys. Chem. C, **2018**, *122* (15), 8315–8326,
DOI: 10.1021/acs.jpcc.7b12670

Takashi Toyao, Keisuke Suzuki, Shoma Kikuchi, Satoru Takakusagi, Ken-ichi Shimizu, IchigakuTakigawa

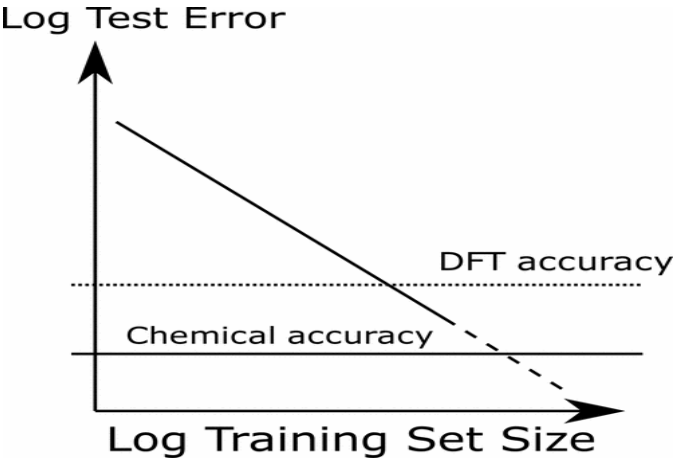
Materials		MachLrn
Task	**Methods**	
• Comparative analysis of cathode materials • $LiNiO_2$ (LNO) • $LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$ (NCA) ▪ NCA configurational space :20,760 configurations	➔ Combined topological analysis ➔ DFT ➔ *operando*neutron diffraction ➔ Machine learning algorithms	Strong dependence of the results of optimization on the initial structure guess



Li(Ni, Co, Al)O_2 Cathode Delithiation: A Combination of Topological Analysis, Density Functional Theory, Neutron Diffraction, and Machine Learning Techniques

J. Phys. Chem. C, **2017**, *121* (51), 28293–28305
DOI: 10.1021/acs.jpcc.7b09760

Roman A. Eremin, Pavel N. Zolotarev, Olga Yu. Ivanshina, Ivan A. Bobrikov

Ground-state properties	MachLrn
Task	**Database**
Machine learning (ML) models of 13 electronic ground-state properties of organic molecules	QM9 database [Ramakrishnan et al. *Sci. Data* 2014, *1*, 140022]
Methods.Regression	**Molecular representations**
➔ Bayesian ridge regression (BR) ➔ Linear regression (LR) with elastic net regularization (EN) ➔ Random forest (RF) ➔ Kernel ridge regression (KRR) ➔ Neural networks ▪ Graph convolutions (GC) ▪ Gated graph networks (GG)	o Coulomb matrix o Bag of bonds o Molecular graphs o (MG) BAML o ECFP4 o Distribution based variants Histograms of ▪ Distances (HD) ▪ Angles (HDA/MARAD) ▪ Dihedrals (HDAD)
	**Molecular Properties** ▪ Enthalpies ▪ Free energies of atomization ▪ HOMO/LUMO gap ▪ HOMO/LUMO energies ▪ Dipole moment ▪ Polarizability ▪ Zero point energy ▪ Vibrational energy ▪ Heat capacity ▪ Highest fundamental vibrational frequency

Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error

*J. Chem. TheoryComput. ,**2017**, 13 (11), 5255–5264*
***DOI:** 10.1021/acs.jctc.7b00577*

Felix A. Faber, Luke Hutchison, Bing Huang, Justin Gilmer, Samuel S. Schoenholz, George E. Dahl, Oriol Vinyals, Steven Kearnes, Patrick F. Riley, and O. Anatole von Lilienfeld

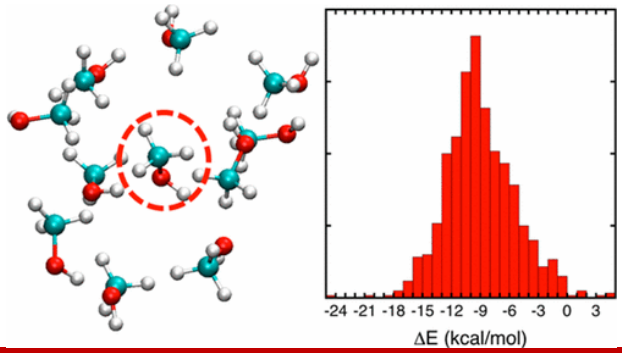
Catalytic activities based on local atomic configurations	MachLrn
Task	**Method**
➔ To direct NO decomposition on RhAu alloy nanoparticles	• Universal machine-learning scheme using a local similarity kernel.

Predicting Catalytic Activity of Nanoparticles by a DFT-Aided Machine-Learning Algorithm

*J. Phys. Chem. Lett., **2017**, 8 (17), 4279–4283*
***DOI:** 10.1021/acs.jpclett.7b02010*

Ryosuke Jinnouchi, Ryoji Asahi

Properties-- condensed phase	MachLrn
Task	**Methods**
➔ To determine a polarizable force field parameters using only ab initio data from quantum mechanics (QM) calculations of molecular clusters at the MP2/6-31G(d, p), DFMP2(fc)/jul-cc-pVDZ, and DFMP2(fc)/jul-cc-pVTZ levels ➔ Prediction of experimental condensed phase properties o Density o Heat of vaporization	▪ Machine learning (ML) techniques with the genetic algorithm (GA) **Systems** ▪ 4943 dimer electrostatic potentials ▪ 1250 cluster interaction energies for methanol

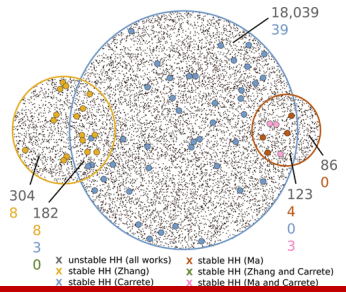


Machine Learning Force Field Parameters from Ab Initio Data

J. Chem. Theory Comput., ***2017***, *13 (9), 4492–4503*
DOI: *10.1021/acs.jctc.7b00521*

Ying Li, Hui Li, Frank C. Pickard, Badri Narayanan, Fatih G. Sen, Maria K. Y. Chan, Subramanian K. R. S. Sankaranarayanan, Bernard R. Brooks, Benoît Roux

Discovery of New Half-Heuslers		MachLrn
Task	Randomforest Method	
+ Predicting stability of half-Heusler (HH) compounds, using only experimentally reported compounds as a training set	✓ Screened 1:1:1 compositions with ML: 71 178 ✓ Likely stable candidates: 481	o Configurational entropies o Quasiharmonicontributions



Materials Screening for the Discovery of New Half-Heuslers: Machine Learning versus ab Initio Methods

J. Phys. Chem. B, ***2018***, *122 (2), 625–632*
DOI: *10.1021/acs.jpcb.7b05296*

Fleur Legrain, Jesús Carrete, Ambroise van Roekeghem, Georg K.H. Madsen, Natalio Mingo

Adaptive Basis Sets	MachLrn
Machine Learning Adaptive Basis Sets for Efficient Large Scale Density Functional Theory Simulation	*J. Chem. Theory Comput.*, **2018**, *14* (8), 4168–4175 **DOI:** 10.1021/acs.jctc.8b00378

Ole Schütt and Joost VandeVondele

Electronic structure	MachLrn
Task	**WorkFlow**
▪ To determine electronic structure of molecules without DFT	▪ Molecules' structure and properties = machLrn (maps of molecular electron density determined from molecules' PE)

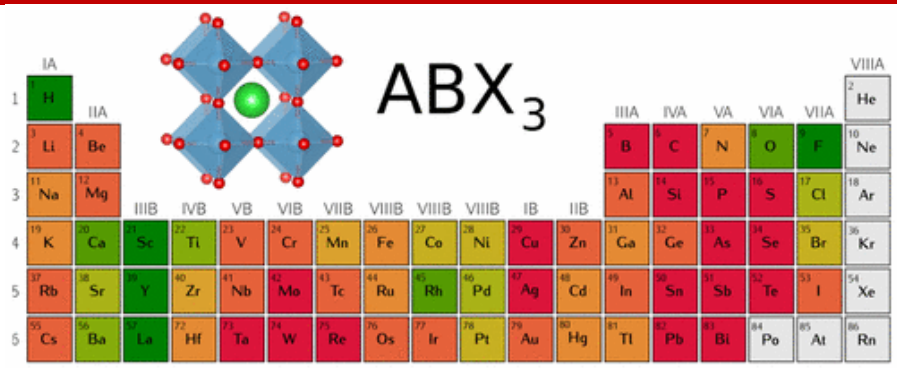
Machine learning streamlines electronic structure calculations for molecules — *C&EN*, 2017, *95* (42), 5–5
DOI: 10.1021/cen-09542-notw3

JyllianKemsley

Thermodynamic Stability		MachLrn
Task	**Data**	**Modeling**
➔ PredictingThermodynamic Stability of Solids (perovskites)	▪ 2, 50, 000 cubic perovskites (withelements from hydrogen to bismuth, excluding rare gases and lanthanides)	

	Perovskites
Trainingset	20, 000
Test set	2, 30, 000

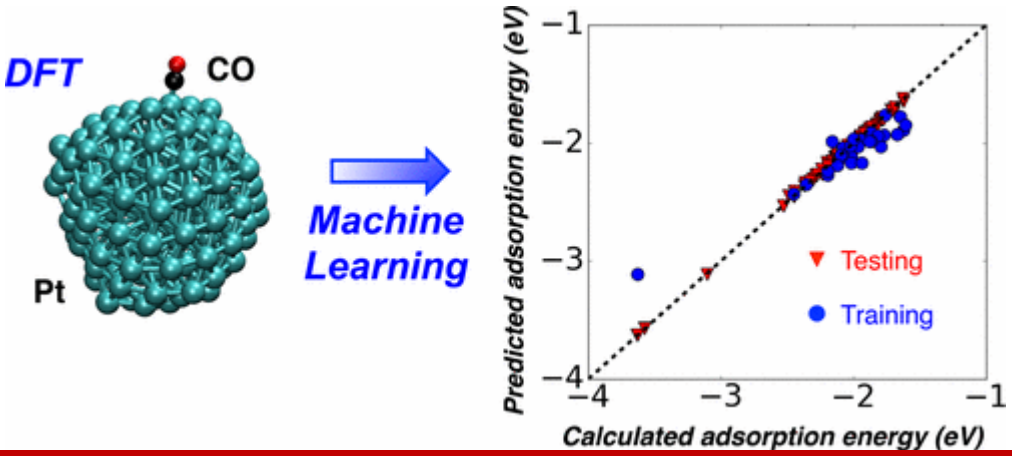
Methods	Information
▪ Ridge regression ▪ Random forests ▪ Extremely randomized trees (including adaptive boosting) ▪ Neural networks	o Extremely randomized trees: smallest mean absolute error of the distance to the convex hull (121 meV/atom) o Systems (around 500) that are thermodynamically stable but that are not present in crystal structure databases



Predicting the Thermodynamic Stability of Solids Combining Density Functional Theory and Machine Learning — *Chem. Mater.*, ***2017***, *29 (12), 5090–5103*
DOI: *10.1021/acs.chemmater.7b00156*

Jonathan Schmidt, Jingming Shi, Pedro Borlido, Liming Chen, Silvana Botti, and Miguel A. L. Marques

CO adsorption	MachLrn
Task	**Method**
➔ CO adsorption on Pt nanoclusters ➔ Predictive models for site-specific adsorption behavior	Empirical-potential-based GA-DFT
Information	**Descriptors**
o Absolute mean error in CO adsorption energy prediction of 0.12 eV o Similar to underlying error of DFT adsorption calculations	▪ Structural ▪ Electronic ▪ Fully frozen adsorption energy—computationally inexpensive probe of CO–Pt bond formation

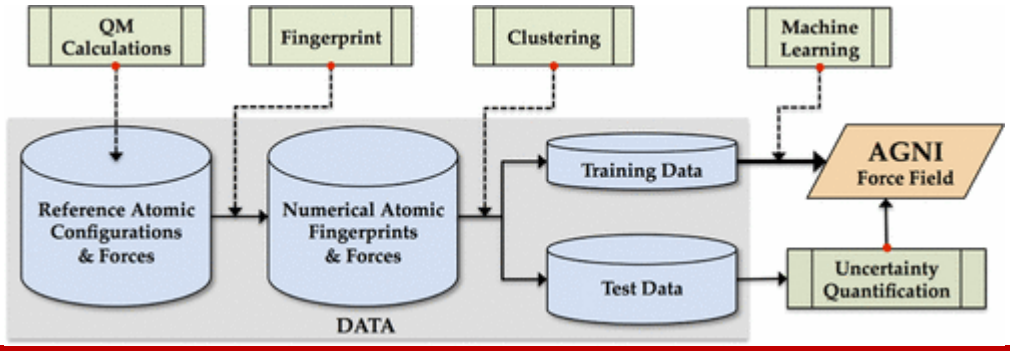


Adsorption of CO on Low-Energy, Low-Symmetry Pt Nanoparticles: Energy Decomposition Analysis and Prediction via Machine-Learning Models

J. Phys. Chem. C, ***2017***, *121 (10), 5612–5619*
DOI: *10.1021/acs.jpcc.6b12800*

Raymond Gasper, Hongbo Shi, Ashwin Ramasubramaniam

Forcefields:Construction,		MachLrn
Task	**multistep WorkFlow**	**Alg**
Force Fields	Machine learning methods in tandem with quantum mechanics → Force fields	▪ Generating diverse reference atomic environments and force data ▪ Choosing a numerical representation for the atomic environments ▪ Down selecting a representative training set ▪ Learning method • Simulating complex materials phenomena viz.Surface melting, stress–strain behavior [it truly goes beyond the realm of ab initio methods, both in length and time scales] • Validation of constructed force field



Machine Learning Forcefields:Construction, Validation, and Outlook

J. Phys. Chem. C, ***2017***, *121 (1), 511–522*
DOI: *10.1021/acs.jpcc.6b10908*

V. Botu, R. Batra, J. Chapman, and R. Ramprasad

Electrostatic multipole moments

MachLrn

Predict

- ✓ Electrostatic multipole moments for all topological atoms in any amino acid based on molecular geometry only
- ✓ Molecular electrostatic interaction energies

\+ Methodology can also handle amino acids with aromatic side chains, without the need for modification

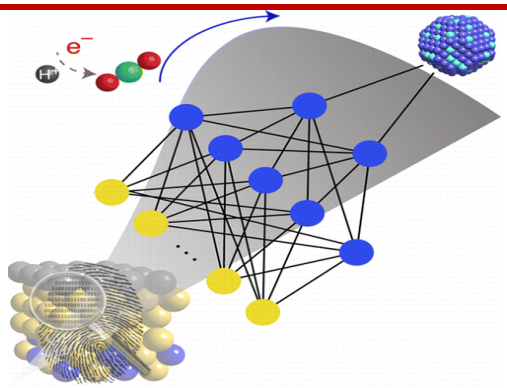
Fn(Error)	kJ mol^{-1}
Mean prediction error	< 5.3
Lowest error observed	2.8
Mean error across the entire set	4.2

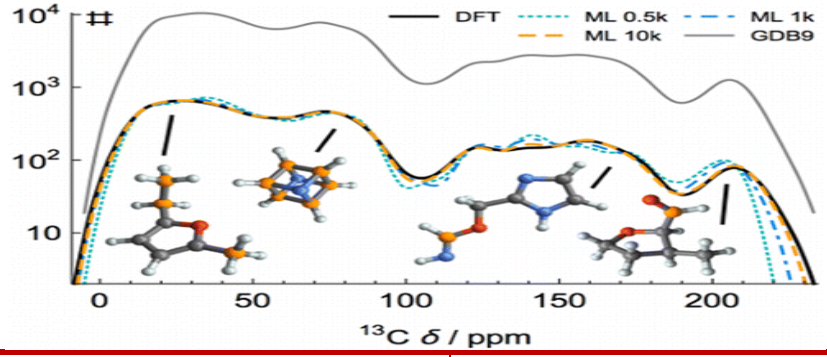
Method
➔ Kriging

Multipolar Electrostatic Energy Prediction for all 20 Natural Amino Acids Using Kriging Machine Learning

J. Chem. Theory Comput., **2016**, *12* (6), 2742–2751
DOI: 10.1021/acs.jctc.6b00457

Timothy L. Fletcher, Paul L. A. Popelier

Chemisorption	MachLrn	
Task	**Data**	**Information**
➔ Machine-learning-augmented chemisorption model →Prediction of the surface reactivity of metal alloys	 o *ab initio* adsorption energies and electronic fingerprints of idealized bimetallic surfaces→	• Captures complex, nonlinear interactions of adsorbates (e.g., CO) on multi-metallics • ~0.1 eV error
Machine-Learning-Augmented Chemisorption Model for CO2 Electroreduction Catalyst Screening	*J. Phys. Chem. Lett., 2015, 6 (18), pp 3528–3533* *DOI: 10.1021/acs.jpclett.5b01660*	
XianfengMa, Zheng Li, Luke E. K. Achenie, Hongliang Xin		

NMR		MachLrn
Task	**Methods**	
Prediction ▪ Proton and carbon nuclear chemical shifts, ▪ Atomic core level excitations, forces	▪ Machine learning models of quantum mechanical observables of atoms in molecules.	▪ Diverse set of 9 k small organic molecules ▪ accuracies on par with DFT
		
Machine Learning for Quantum Mechanical Properties of Atoms in Molecules	*J. Phys. Chem. Lett., **2015**, 6 (16), 3309–3313* ***DOI:** 10.1021/acs.jpclett.5b01456*	
Matthias Ru, Raghunathan Ramakrishnan, O. Anatole von Lilienfeld		

SEMO-parameters		MachLrn
Task	**WorkFlow**	**Information**

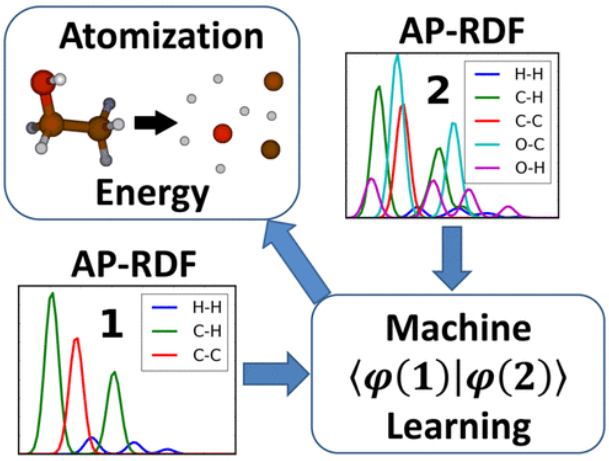
▪ Improvements in the accuracy of SEMO by o ML models for the parameters	o ML-SQC o Automatic tuning of SQC parameters for individual molecules + Improved accuracy + Without deteriorating transferability to molecules with molecular descriptors very different from those in the training set	▪ Semiempirical OM2 appliedto 6095 constitutional isomers $C_7H_{10}O_2$ o accurate *ab initio* atomization enthalpies are available ▪ Mean absolute errors in atomization enthalpies : [6.3 to 1.7 kcal/mol]

Machine Learning of Parameters for Accurate Semiempirical Quantum Chemical Calculations *J. Chem. Theory Comput.*, ***2015***, *11 (5), 2120–2125*
DOI: *10.1021/acs.jctc.5b00141*

Pavlo O. Dral, O. Anatole von Lilienfeld, and Walter Thiel

Atomization energies **MachLrn**

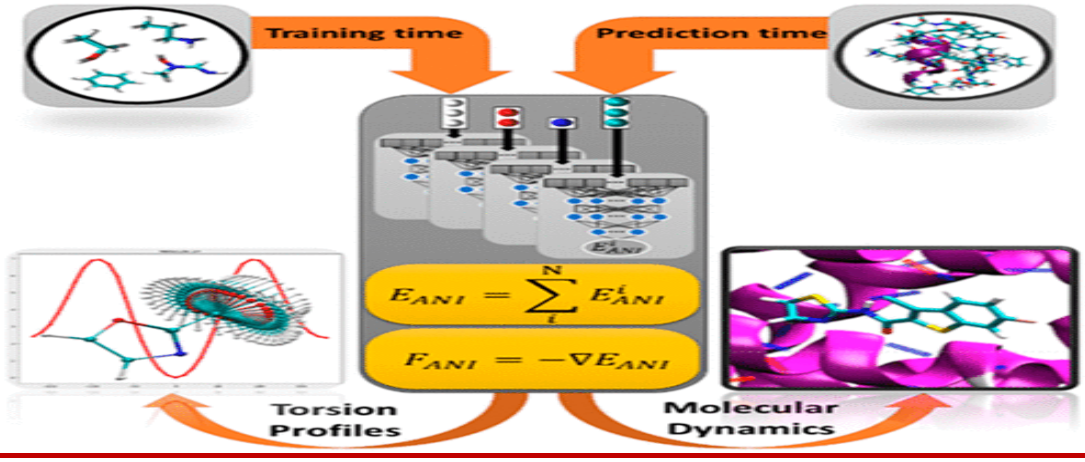
Task	Methods	Accuracy
▪ Molecular atomization energies	▪ Machine learning kernels in closed and analytic form o Atomic property weighted radial distribution function (AP-RDF) + descriptor with a Gaussian kernel. → improvement in performance of the Bag-of-Bonds descriptor when the bond type restriction is included in AP-RDF.	o MAE = 1.7 kcal/mol o QM7 data set



Bond Type Restricted Property Weighted Radial Distribution Functions for accurate MachineLearning Prediction of Atomization Energies *J. Chem. Theory Comput.*, ***2018***, *14 (10), 5229–5237*
DOI: *10.1021/acs.jctc.8b00788*

MykhayloKrykunov, Tom K. Woo

Drug Discovery--Computational **MachLrn**



Transforming Computational Drug Discovery with Machine Learning and AI

ACS Med. Chem. Lett., 2018, 9 (11), pp 1065–069
DOI: 10.1021/acsmedchemlett.8b00437 x

Justin S. Smith, Adrian E. Roitberg, OlexandrIsayev

To predict important new configurations | **MachLrn**

Task

- Multireference problem of the water molecule with elongated bonds.

Methods

- First-order perturbation
- Random selection
- Monte carlo configuration interaction

- NNs discriminate between important and unimportant configurations

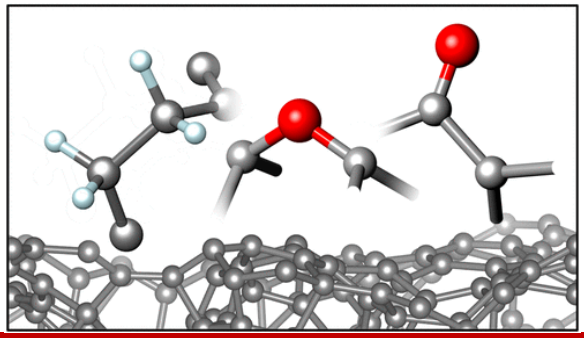
Machine Learning Configuration Interaction

J. Chem. Theory Comput., 2018, 14 (11), 5739–5749
DOI: 10.1021/acs.jctc.8b00849

J. P. Coe

Surface Chemistry | **MachLrn**

Task	**Methods**
Complex atomic-scale structures Chemical reactivity of ta-C (Tetrahedral amorphous carbon)surfaces	Machine learning Density functional tight binding DFT



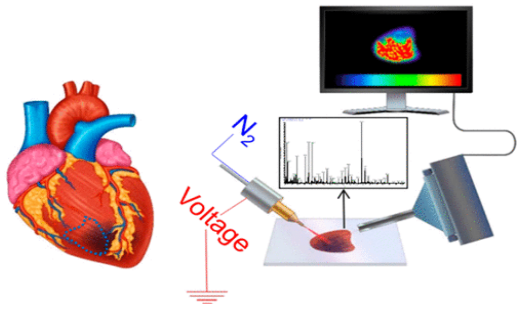
Computational Surface Chemistry of Tetrahedral Amorphous Carbon by Combining MachineLearning and Density Functional Theory

Chem. Mater., 2018, 30 (21), pp 7438–7445
DOI: 10.1021/acs.chemmater.8b02410

Volker L. Deringer, Miguel A. Caro, Richard Jana, Anja Aarva, Stephen R. Elliott, Tomi Laurila, GáborCsányi, and Lars Pastewka

---

Molecular Recognition of Myocardial Infarction | **MachLrn**

Task	
Identification ▪ Cardiac pathology molecules classification ▪ Chemical identity of small metabolites and lipids	
Method ▪ Gradient boosting tree ensemble	

Combining Desorption Electrospray Ionization Mass Spectrometry Imaging and Machine Learning for Molecular Recognition of Myocardial Infarction

Anal. Chem., 2018, 90 (20), 12198–12206
DOI: 10.1021/acs.analchem.8b03410

Katherine Margulis, Zhenpeng Zhou, Qizhi Fang, Richard E. Sievers, Randall J. Lee, and Richard N. Zare

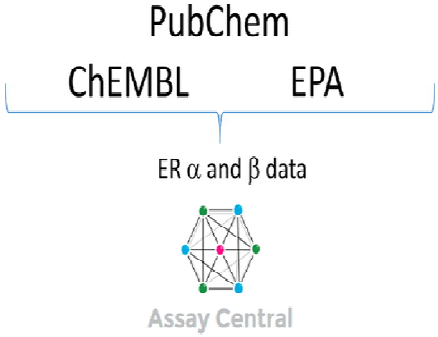
Prediction- Estrogen Receptor Binding | **MachLrn**

X

- Chemical features
- Binary fingerprints (ECFP6, FCFP6, ToxPrint, or MACCS keys)
- Continuous molecular descriptors from RDKit

classic MachLrnAlg
- Bernoulli Naive Bayes
- AdaBoost Decision Tree
- Random Forest
- SVM
- Deep NNs

Metrics
- Recall
- Precision
- F1-score
- Accuracy
- Area under receiver operating characteristic curve
- Cohen's Kaa
- Matthews correlation coefficient

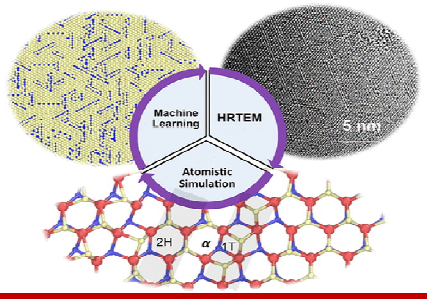


Comparing Multiple Machine Learning Algorithms and Metrics for Estrogen Receptor Binding Prediction

Mol. Pharmaceutics, 2018, 15(10),4361-4370
DOI:10.1021/acs.molpharmaceut.8b00546xx

Daniel P. Russo,Kimberley M. Zorn, Alex M. Clark, Hao Zhu, and Sean Ekins

Atomistic Simulations | **MachLrn**

Task	Methods	
▪ Defect Dynamics in 2-D MoS_2	o Supervised machine learning, o in situ high-resolution transmission electron microscopy (s) o MD o GA with MD→to identify the long-range structure of randomly distributed point defects (sulfur vacancies)	

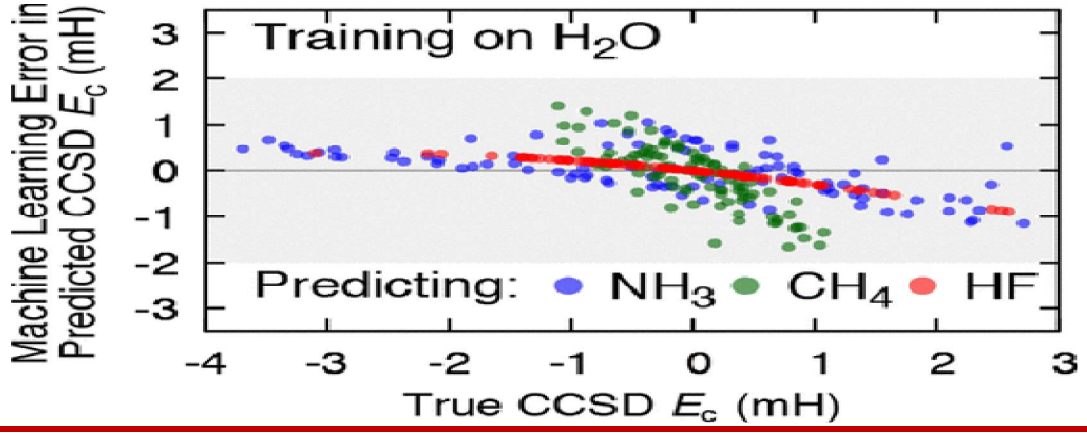
Defect Dynamics in 2-D MoS2 Probed by Using Machine Learning, Atomistic Simulations, and High-Resolution Microscopy

ACS Nano, 2018, 12 (8), pp 8006–8016
DOI: 10.1021/acsnano.8b02844

Tarak K. Patra, Fu Zhang, Daniel S. Schulman, Henry Chan, Mathew J. Cherukara, Mauricio Terrones , Saptarshi Das∇, Badri Narayanan, Subramanian K. R. S. Sankaranarayanan

Correlation energies	MachLrn

Task	WorkFlow
▪ Predicting electronic structure correlation energies using Hartree–Fock input. ▪ Maximizing transferability across chemical systems and compactness of the feature set,	▪ Total correlation energy =Fn(individual and pair contributions from occupied molecular orbitals) ▪ Gaussian process regression ▪ To predict the contributions from a feature set [molecular orbital properties Ex:Fock, Coulomb, and exchange matrix elements]

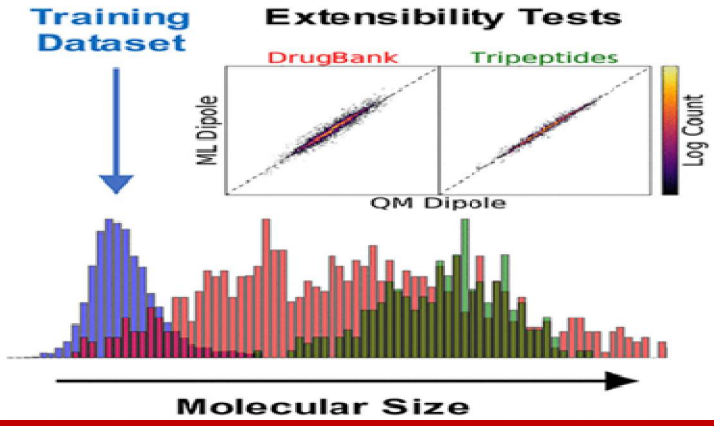


Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis *J. Chem. Theory Comput., 2018, 14 (9), 4772–4779*
DOI: 10.1021/acs.jctc.8b00636
Matthew Welborn, Lixue Cheng, and Thomas F. Miller, III

Drug Discovery	MachLrn

Machine Learning in Drug Discovery *J. Chem. Inf. Model., 2018, 58 (9), 1723–1724*
DOI: 10.1021/acs.jcim.8b00478
Se Hochreiter, Guenter Klambauer, and Matthias Rarey

Partial atomic charge assignment	MachLrn



Discovering a Transferable Charge Assignment Model Using Machine Learning

J. Phys. Chem. Lett., 2018, 9 (16), 4495–4501
DOI: 10.1021/acs.jpclett.8b01939

Andrew E. Sifain, Nicholas Lubbers, Benjamin T. Nebgen, Justin S. Smith, Andrey Y. Lokhov, OlexandrIsayev,Adrian E. Roitberg, Kipton Barros, and Sergei Tretiak

Crystal Structure — **MachLrn**

Task	Methods	Pearson Crystal Database
▪ To predict crystal structure	▪ Machine learning ▪ DFT	▪ 24 913 unique chemical formulas existing between 290 and 310 K They contain 10 711 unique crystal structures

Metrics	Range
Accuracy	97 ± 2 to 85 ± 2%;
Average precision	86 ± 2 to 79 ± 2%
Average recall	73 ± 2 to 54 ± 2%
Minimum-class representatives	150 to 10,

Machine Learning and Energy Minimization Approaches for Crystal Structure Predictions: A Review and New Horizons

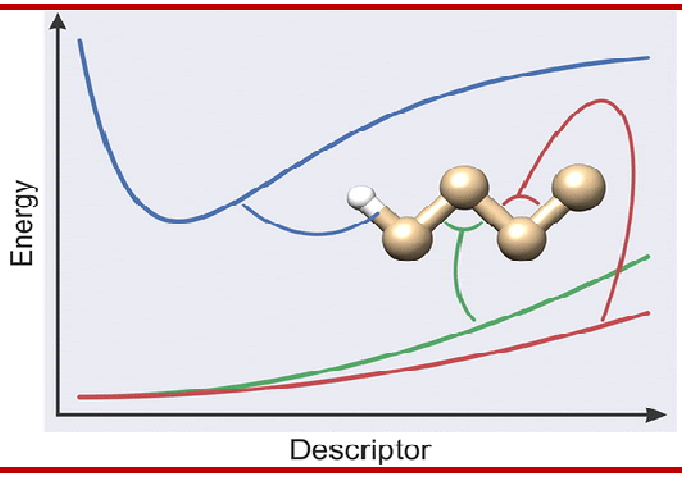
Chem. Mater., 2018, 30 (11), 3601–3612
DOI: 10.1021/acs.chemmater.7b05304

Jake Graser, Steven K. Kauwe, and Taylor D. Sparks

Molecular properties — **MachLrn**

Task	Descriptors	Methods
▪ Prediction of molecular properties across chemical compound space	▪ Two-body and three-body interaction descriptors ▪ Invariant to translation, rotation, and atomic indexing	▪ Kernel ridge regression ▪ DFT

Best model (Training set : [5000 random molecules])		
Accuracy (kcal/mol)	**On remaining molecules**	**DB**
0.8	1, 868	GDB-7
1.5	16, 722	GDB-9



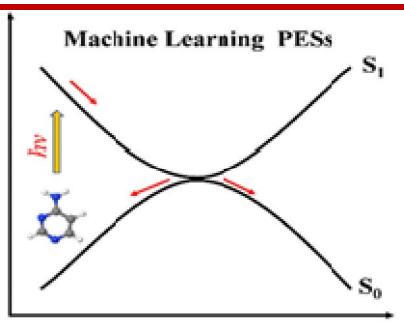
Many-Body Descriptors for Predicting Molecular Properties with Machine Learning: Analysis of Pairwise and Three-Body Interactions in Molecules

J. Chem. Theory Comput., 2018, 14 (6), 2991–3003
DOI: 10.1021/acs.jctc.8b00110xx

WiktorPronobis, Alexandre Tkatchenko, and Klaus-Robert Müller

Dynamics simulation of polyatomic systems	MachLrn

Task	Method
o Learning potential energy surfaces in nonadiabatic MD	▪ Kernel ridge regression ▪ ML-PESs are consistent with those based on CASSCF Potential Energy Surfaces



Inclusion of Machine Learning Kernel Ridge Regression Potential Energy Surfaces in On-the-Fly Nonadiabatic Molecular Dynamics Simulation

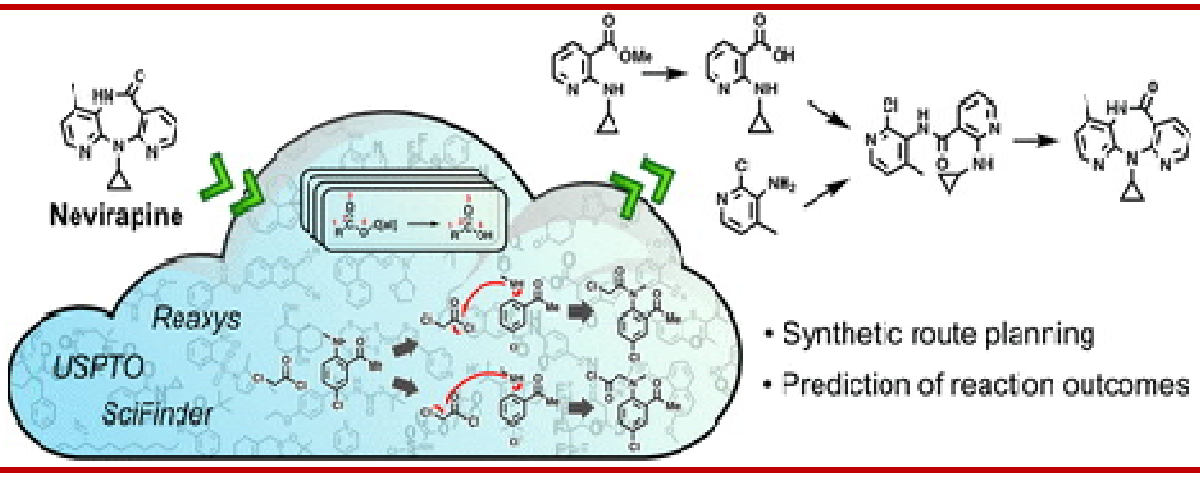
J. Phys. Chem. Lett., 2018, 9 (11), 2725–2732
DOI: 10.1021/acs.jpclett.8b00684xx

Deping Hu, Yu Xie, Xusong Li, Lingyue Li, and Zhenggang Lan

Computer-aided synthesis planning	MachLrn

Task	WorkFlow
▪ Data-driven approaches to synthesis planning.	▪ Input :Molecular structure ▪ Output:Sorted list of detailed reaction schemes ▪ Each connect the target to via a series of chemically feasible reaction through purchasable starting materials

Methods	Large reaction corpora
➔ Retrosynthetic planning ➔ Anticipating the products of chemical reactions ➔ Earlier Chemists decide how to synthesize small molecule compounds	➔ United States Patent and Trademark Office (USPTO) ➔ Reaxys, ➔ SciFinder databases o Millions of tabulated reaction examples



Machine Learning in Computer-Aided Synthesis Planning

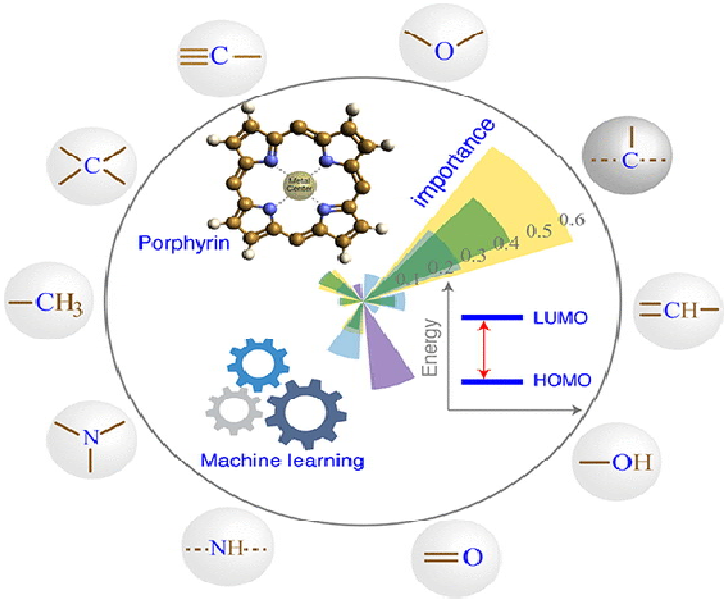
Acc. Chem. Res., 2018, 51 (5), 1281–1289
DOI: 10.1021/acs.accounts.8b00087

Connor W. Coley, William H. Green, and Klavs F. Jensen

Energy gap	MachLrn
Task	**Data & Descriptors**
➔ Predicting the energy gaps	▪ >12 000 porphyrins from the Computational Materials Repository ▪ Electrotopological-state index,

Methods

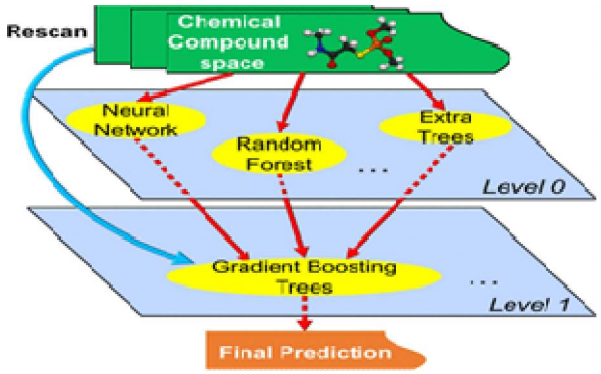
- Domain knowledge of chemical graph theory



Machine-Learning Energy Gaps of Porphyrins with Molecular Graph Representations

J. Phys. Chem. A, 2018, 122 (18), 4571–4578
DOI: 10.1021/acs.jpca.8b02842xx

Zheng Li, NoushinOmidvar, Wei Shan Chin, Esther Robb, Amanda Morris, Luke Achenie, Hongliang Xin

Molecular atomization energies	MachLrn
Task	
▪ Ground state molecular atomization energies	

Significantly Improving the Prediction of Molecular Atomization Energies by an Ensemble of Machine Learning Algorithms and Rescanning Input Space: A Stacked Generalization Aroach

J. Phys. Chem. C, 2018, 122 (16), 8868–8873
DOI: 10.1021/acs.jpcc.8b03405

Ruobing Wang

CQC	MachLrn

Task

- To reduce the dimensionality of a complex molecular system
- Essential/ internal coordinates
  - Specific interatomic distances
  - Dihedral angles

Machine Learning of Biomolecular Reaction Coordinates — *J. Phys. Chem. Lett., 2018, 9 (9), 2144–2150* *DOI: 10.1021/acs.jpclett.8b00759xx*

Simon Brandt, Florian Sittel, Matthias Ernst, and Gerhard Stock

Band gap	MachLrn

Task

- To predict band gap of inorganic solids

Methods

- Support vector classification
- DFT (PBE-level) calculated

Predicting the Band Gaps of Inorganic Solids by Machine Learning — *xx* J. Phys. Chem. Lett., 2018, 9 (7), 1668–1673 DOI: 10.1021/acs.jpclett.8b00124

YaZhuo, Aria Mansouri Tehrani, JakoahBrgoch

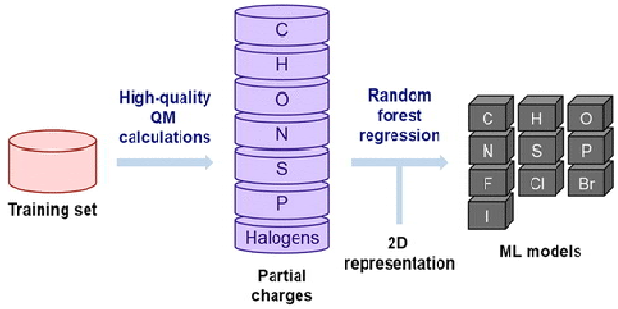
Partial charges	Prediction	MachLrn

Task

- Parametrization of small organic molecules for classical molecular dynamics simulations

Methods

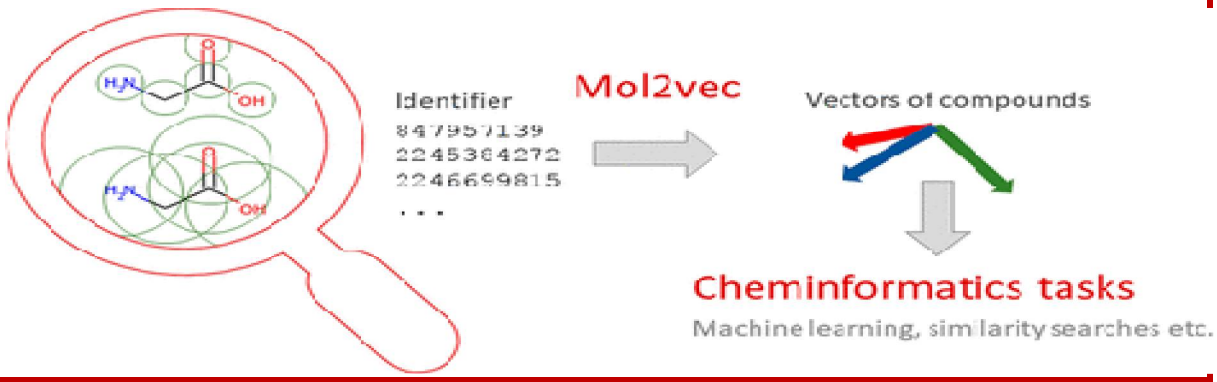
➔ DFT



Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations — *J. Chem. Inf. Model., 2018, 58 (3), 579–590* *DOI: 10.1021/acs.jcim.7b00663*

Patrick Bleiziffer, Kay Schaller, SereinaRiniker

Moleules→ vectors	**MachLrn**
Task	**Mol2vec**
▪ To learn vector representations of molecular substructures	▪ Learns vector representations of molecular substructures ▪ Compounds can finally be encoded as vectors by summing the vectors of the individual substructures
Methods ➔ Proteochemometric approach ➔ That is alignment-independent and used for proteins with low sequence similarities ➔ Unsupervised machine learning approach	o Yields dense vector representations o Overcomes drawbacks of common compound feature representations such as sparseness and bit collisions



Mol2vec: Unsupervised Machine Learning Aroach with Chemical Intuition

J. Chem. Inf. Model., 2018, 58 (1), 27–35
DOI: 10.1021/acs.jcim.7b00616

Sabrina Jaeger, Simone Fulle, Samo Turk

MaterialsDesirable properties	**MachLrn**
Task	**Methods**
▪ Acceleration of search for materials with desired properties ▪ For an arbitrary composition of a compound , what crystal structures are adopted	▪ Support vector machine ▪ Random forest algs.

o Discovery of RhCd, the first new binary AB compound to be found in over 15 years, with a CsCl-type structure

o New candidates for thermoelectric materials, including previously unknown compounds (e.g., TiRu2Ga with Heusler structure; Mn(Ru0.4Ge0.6) with CsCl-type structure) and previously reported compounds but counterintuitive candidates (e.g., Gd12Co5Bi).

- Machine-learning models are only as good as the experimental data used to develop them →

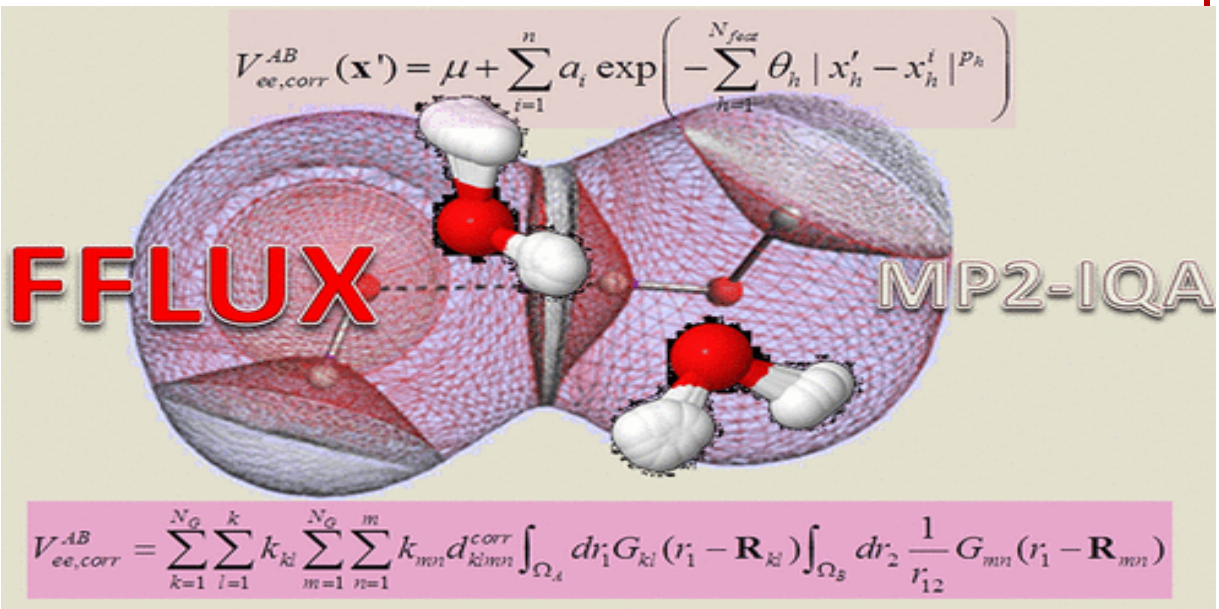
➔ Experimental work will continue to be necessary to improve the predictions made by machine learning

Discovery of Intermetallic Compounds from Traditional to Machine-Learning Aroaches

Acc. Chem. Res., 2018, 51 (1), 59–68
DOI: 10.1021/acs.accounts.7b00490

Anton O. Oliynyk and Arthur Mar

Electron correlation energy		MachLrn
Task	**Systems**	**Methods**
▪ To predict dynamic electron correlation energy of an atom or a bond in a molecule utilizing topological atoms	▪ Water monomer ▪ Water dimer ▪ van der Waals complex H2⋯He	➔ Kriging (Gaussian Process Regression with a non-zero mean function) ➔ New method by which dispersion potentials for molecular simulation can be generated

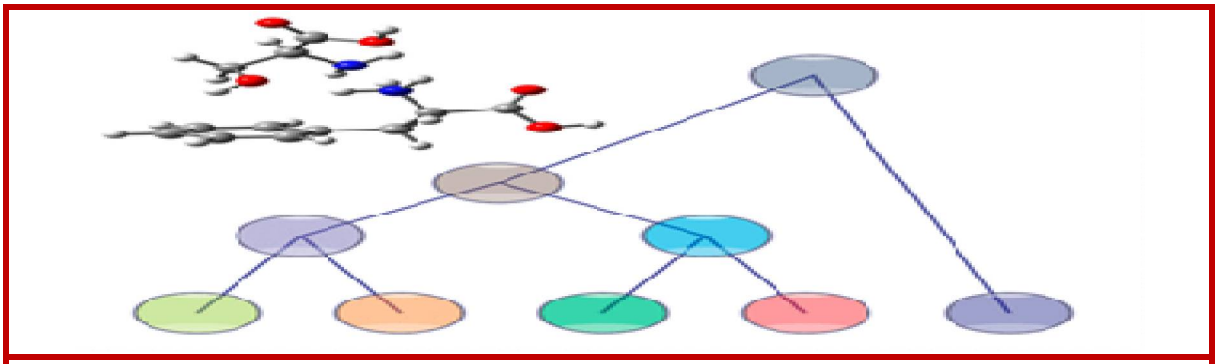


Machine Learning of Dynamic Electron Correlation Energies from Topological Atoms

J. Chem. Theory Comput., 2018, 14 (1), 216–224
DOI: 10.1021/acs.jctc.7b01157

James L. McDonagh, Arnaldo F. Silva, Mark A. Vincent, and Paul L. A. Popelier

PES	MachLrn
➔ 37 isomers are identified within 180 kJ·mol–1 of the global-minimum structure	- Cluster structures are grouped using hierarchical clustering ➔Partitions the PES in terms of nuclear configuration. o Calculated IR spectra for the various isomers are then compared with the isomer-specific IR spectra by means of the cosine distance metric o To facilitate spectral assignment o To identify which regions of the PES are populated in the electrospray ionization process.



Applying Machine Learning to Vibrational Spectroscopy

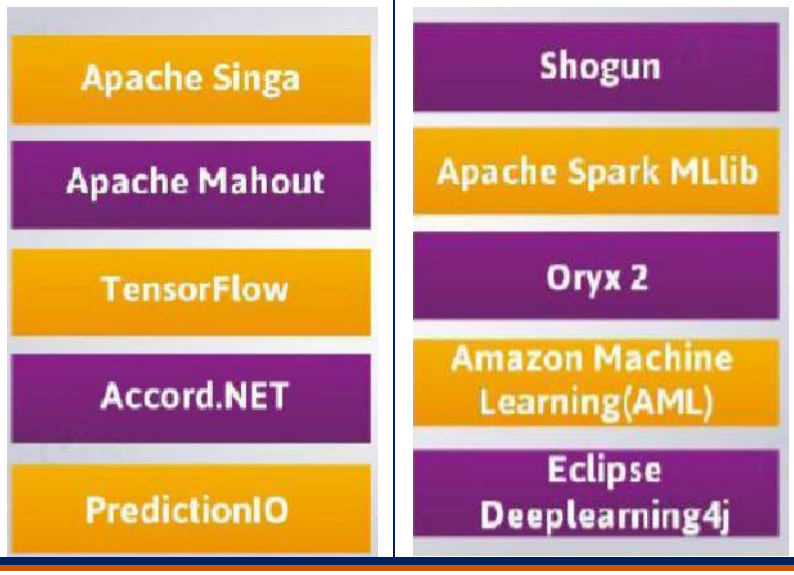
J. Phys. Chem. A, 2018, 122 (1), 167–171
DOI: 10.1021/acs.jpca.7b10303

Weiqiang Fu, W. Scott Hopkins

Typical Machine Learning Methods

Open-source libraries for MachLrn 	o Scikitlearn, o TensorFlow o PyTorch4
	o Google's TensorFlow. o Nervana Neon. o Amazon Web Services. o OpenNN. o Apache Spark MLlib. o Caffe. o Veles

Machine Learning Software



Machine Learning methods (Algs.)

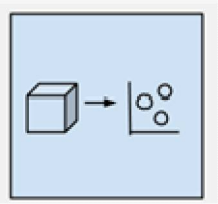
Regression

- Linear (LR)
  - Ordinary
  - With elastic net regularization (EN)
- Stepwise Regression
- Bayesian ridge (BR)
- Logistic Regression
- Kernel ridge (KRR)
- Boosting (GBR)
- Extra tree (ETR)
- Multivariate Adaptive Regression Splines (MARS)
- Locally Estimated Scatterplot Smoothing (LOESS)
- Ridge Regression
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Elastic Net
- Least-Angle Regression (LARS)

Bayesian algorithms

- Naive Bayes
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Averaged One-Dependence Estimators (AODE)
- Bayesian Belief Network (BBN)
- Bayesian Network (BN)

Dimensionality Reduction



- Principal Component Analysis (PCA)
- Principal Component Regression (PCR)
- Partial Least Squares Regression (PLSR)
- Sammon Mapping
- Multidimensional Scaling (MDS)
- Projection Pursuit

Discriminant Analysis

- Linear Discriminant Analysis (LDA)
- Mixture Discriminant Analysis (MDA)
- Quadratic Discriminant Analysis (QDA)
- Flexible Discriminant Analysis (FDA)

clustering algorithms

- k-Means
- k-Medians
- Expectation Maximisation (EM)
- Hierarchical Clustering

Decision Tree Algorithms

- Classification and Regression Tree (CART)
- Iterative Dichotomiser 3 (ID3)
- C4.5 and C5.0 (different versions)
- Chi-squared Automatic Interaction Detection (CHAID)
- Decision Stump
- M5
- Conditional Decision Trees

Instance-based Algorithms

- k-Nearest Neighbor (kNN)
- Learning Vector Quantization (LVQ)

Association Rule Learning Algorithms

- Apriori algorithm
- Eclat algorithm

- Self-Organizing Map (SOM)
- Locally Weighted Learning (LWL)

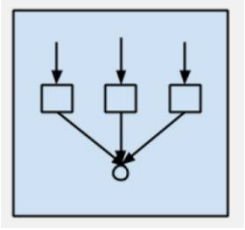
Deep Learning Algorithms

- Deep Boltzmann Machine (DBM)
- Deep Belief Networks (DBN)
- Convolutional Neural Network (CNN)
- Stacked Auto-Encoders

Artificial Neural Network

- Perceptron
- Hopfield Network
- Radial Basis Function Network (RBFN)
- Graph convolutions (GC)
- Gated graph networks (GG
- Deep Learning Deep NNs

Ensemble Algorithms



- Boosting
- Bootstrapped Aggregation (Bagging)
- AdaBoost
- Stacked Generalization (blending)
- Gradient Boosting Machines (GBM)
- Gradient Boosted Regression Trees (GBRT)
- Random Forest
- Extremely randomized trees (including adaptive boosting)

	Type	Size	Dimension
Data	Symbolic Boolean Numeric Pixel, ….	Small Large Big	[1, 2, 3] [4, 5, ….]
	Data Science		
Artificial	intelligence	Symbolic	
Computational	intelligence	Numeric	
Machine	learning		
Deep	learning		

Learning	[Life, Machine]
Life	[animal, Human]
Learning intensity	[Deep, normal]

Machine Learning	[Learning style; Similarity form or Function]
Learning Style	[Supervised; Unsupervised; Semi-Supervised]
Similarity form	[Regression;

Machine learning		or function	Instance-based Regularization; Decision Tree Clustering Deep Learning; Dimensionality Reduction Ensemble

- Unsupervised
- Semi-supervised
- supervised
- hybrid

Input	Multiple levels of representation Single source Fusion

Data	
Noise [Low, high] Bias	uncertainty
Redundant	Inconsistent

Models:
Inadequate; overambitious; Empirical ; Theoretical

Theories:
low level; complicated hierarchical, ;

Performance	Data uncertainity Human level or above

Real time, dynamic live molecular/atomic/electron level

Phenomena:	Completely known	but difficult to model
	Not completely known In a compartment or sub phenomenon	The rest can be modelled
	Vague	

High level scientific computing Software	Python/R/Matlab, Mathematica
Low-level libraries	C/C++/Fortran/Assembly

CQC-derived properties	CQC-derivedproperties
Electron energies	Thermodynamic properties
HOMO HOMO–LUMO gap Prediction	Chemisorption
Forces and energy	CO adsorption on Pt nanoclusters Force Fields
Interatomic potential	Discovery in Transition Metal Chemistry
Polarizable force field parameters	Vibrational
Electrostatic multipole moment	NMR (1H1;13C)

CQC-methods		Experimental
➔ DFT ➔ Grand canonical Monte Carlo ➔ Density-functional tight-binding (DFTB) method	➔ Combined topological analysis	o Small-angle X-ray scattering (SAXS) o X-ray absorption near-edge structure (XANES) spectroscopy o Neutron diffraction

Molecular representations	DataBases
o Coulomb matrix o Bag of bonds o Molecular graphs o (MG) BAML o ECFP4 o Distribution based variants o histograms of o distances (HD) o angles (HDA/MARAD) o dihedrals (HDAD	o QM9 database [Ramakrishnan et al. *Sci. Data* 2014, *1*, 140022]

ACS.org: Information Source (is)

R. Sambasiva Rao, School of Chemistry
Andhra University, Visakhapatnam
rsr.chem@gmail.com