

Theoretical Anharmonic Vibrational Spectroscopy on Machine Learning Optimized Potential Energy Surfaces

Jonathan Smith

Visiting Professor, Hylleraas Centre for Quantum Molecular Science
Associate Professor, Temple University

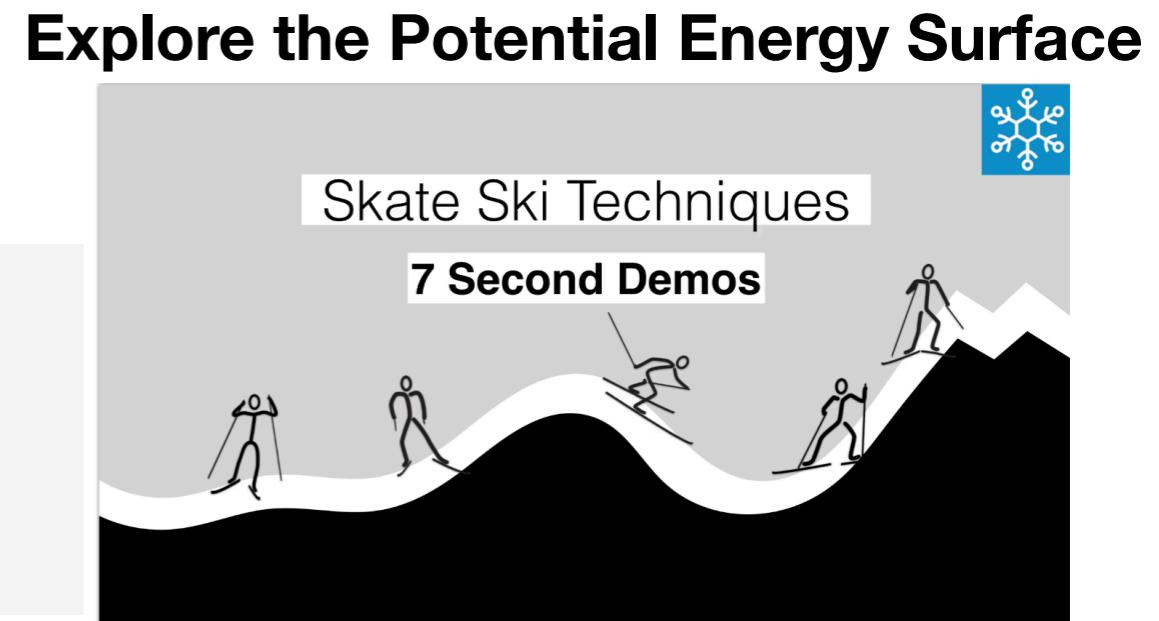
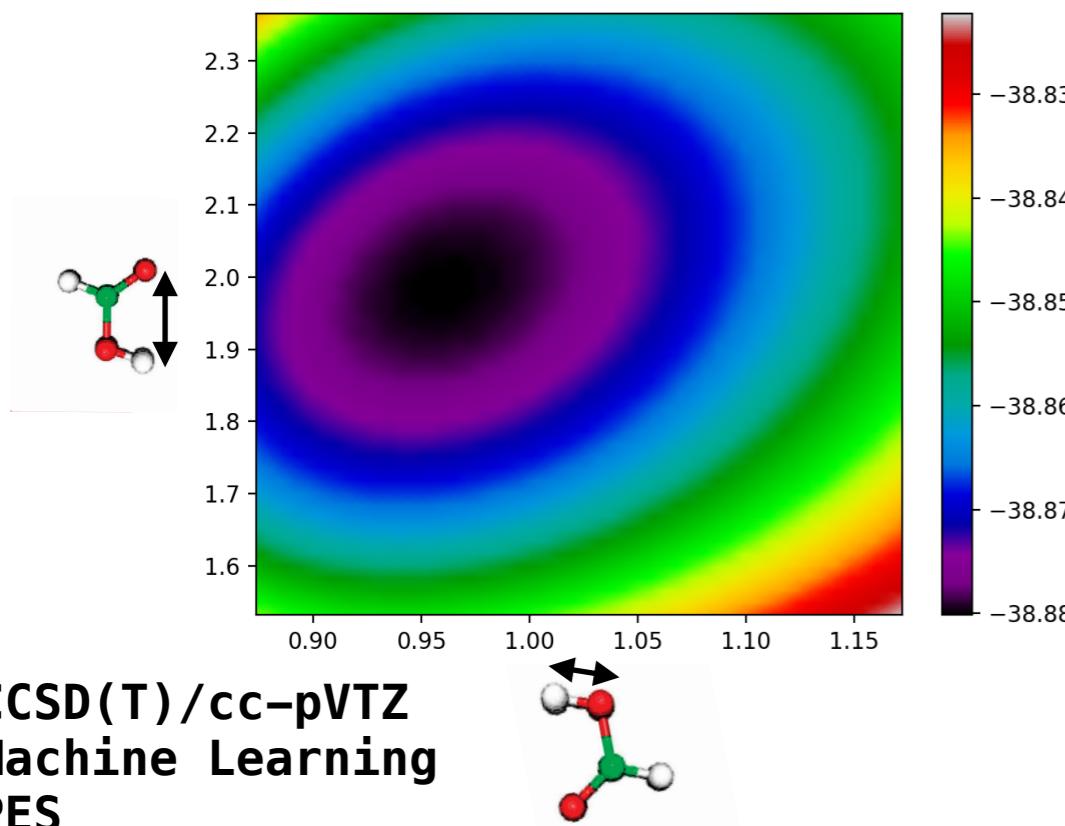


Abstract

Molecular vibrational spectra encode structure, conformation, energetics, and, in condensed phases, local solvent environment. In many cases sampling a wide range of an anharmonic potential energy surface is necessary to reproduce spectra and to better elucidate the underlying encoded chemistry. Computing spectra from ab initio molecular dynamics (MD) affords some advantages in the above cases due to sampling the potential energy surface without resorting to solely the potential's gradient near minima in static calculations. MD also implicitly includes mechanical and electrical anharmonicity. Ab initio molecular dynamics on coupled cluster level surfaces or directly afford a best case in accurately capturing these effects but direct dynamics at this level of theory are presently only viable for small molecular systems. Here we present an approach which leverages machine learning to optimize accurate potential energy surfaces with up to the coupled cluster level accuracy and fast evaluation of energetics and gradients at approximately 10^5 faster per gradient. This permits sufficiently long and statistically averaged trajectories and even permits the treatment of nuclear dynamics through the path integral molecular dynamics approximation which can be particularly important for hydroxyl groups and solvation dynamics. Early results on acetylene, formic acid, and a formic acid-water complex highlight the promise of this approach.

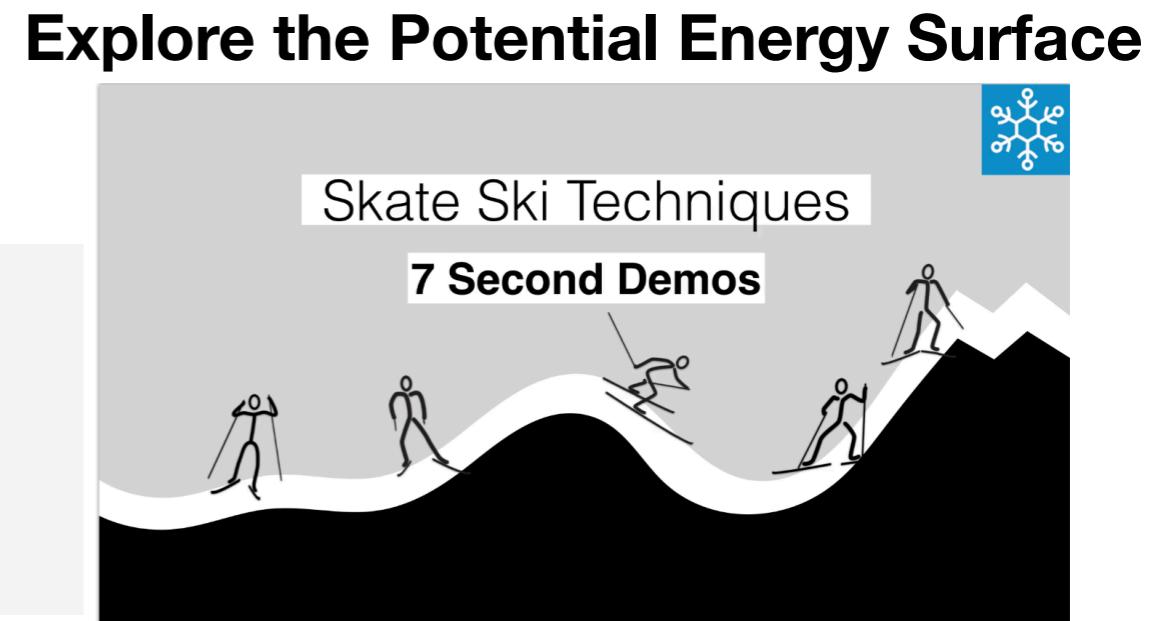
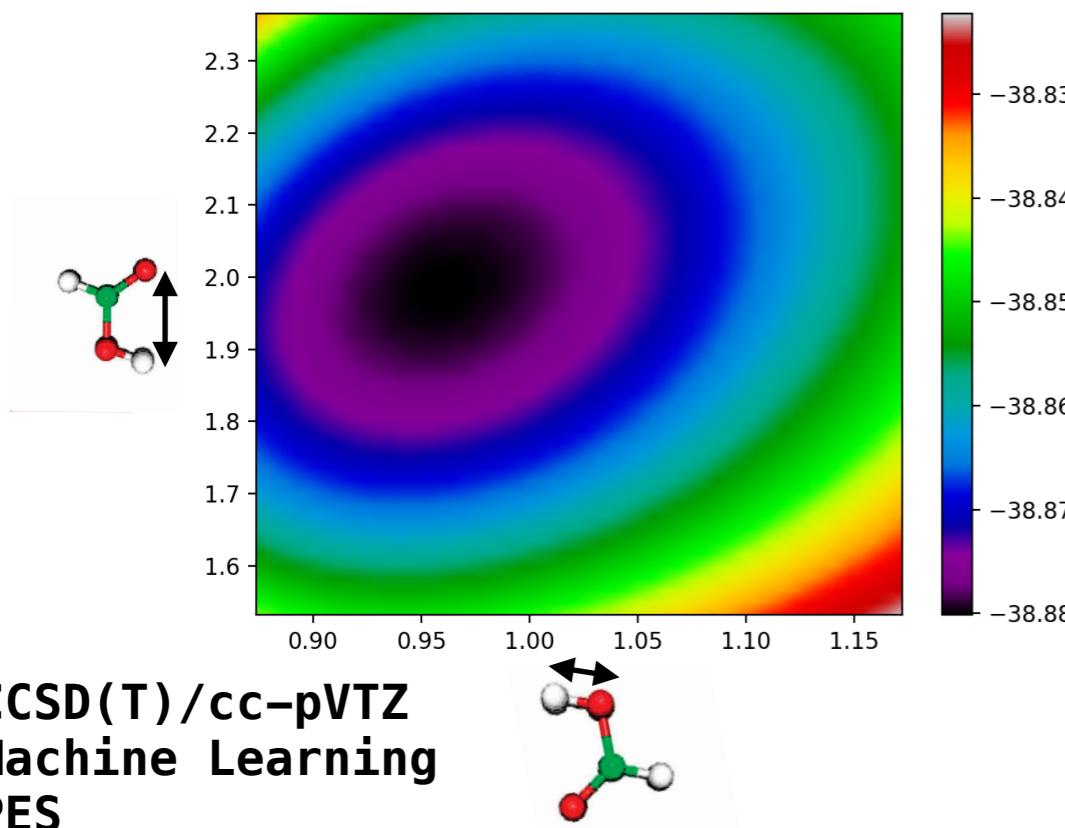
Motivation

- Vibrational signature: encode molecular energetics and reaction dynamics = **Molecular Insights**
- Energized, floppy, solvated molecules sample anharmonic parts of potential energy surface
- Nuclear Dynamics: Hydrogen bonding, conformational sampling



Motivation

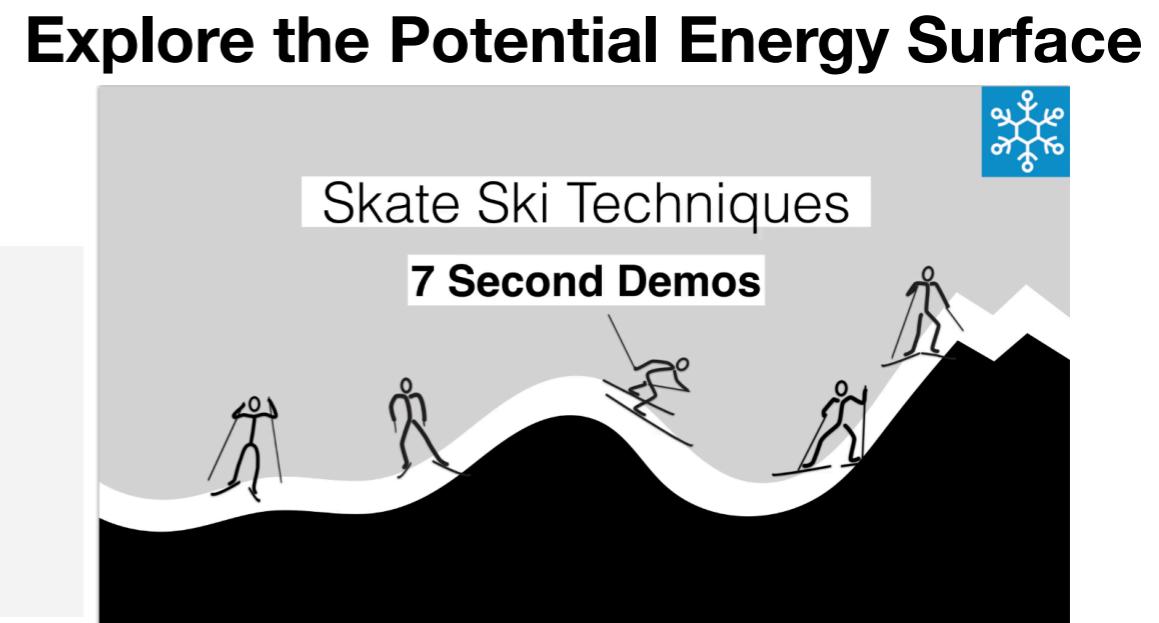
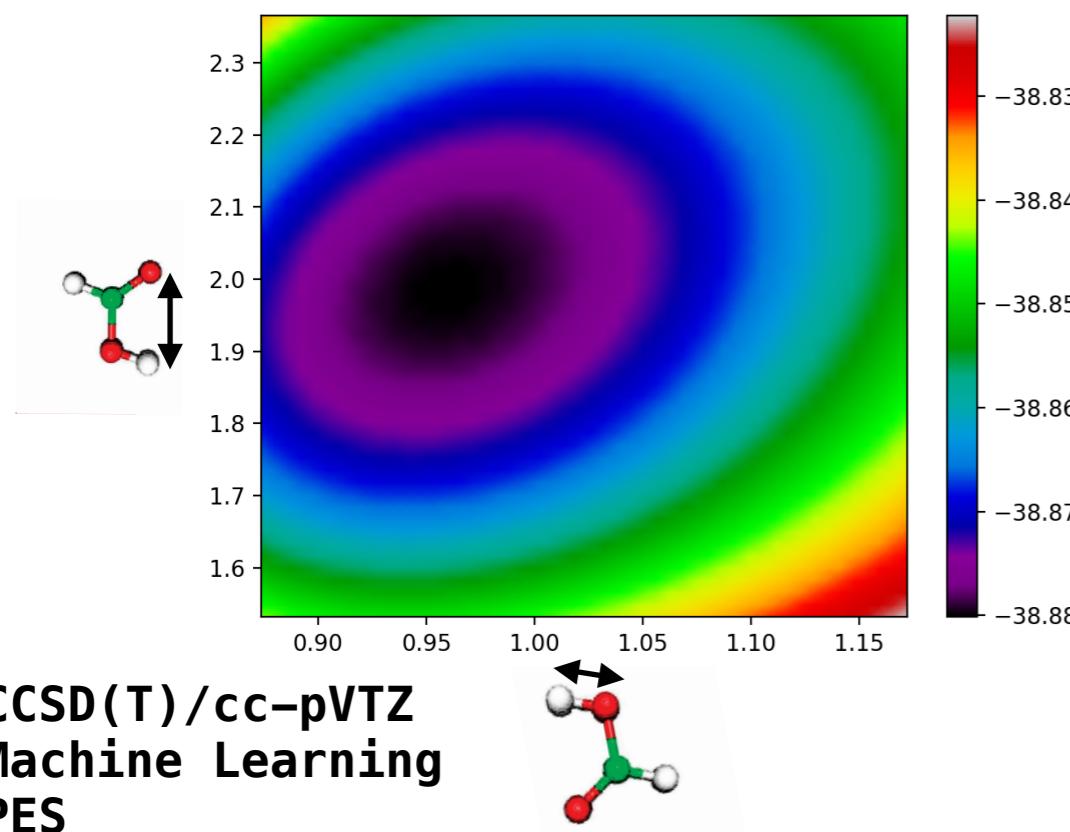
- Vibrational signature: encode molecular energetics and reaction dynamics = **Molecular Insights**
- Energized, floppy, solvated molecules sample anharmonic parts of potential energy surface
- Nuclear Dynamics: Hydrogen bonding, conformational sampling



Motivation

- Vibrational signature: encode molecular energetics and reaction dynamics = **Molecular Insights**
- Energized, floppy, solvated molecules sample anharmonic parts of potential energy surface
- Nuclear Dynamics: Hydrogen bonding, conformational sampling

Solution: Sample an accurate surface through *ab initio* molecular dynamics

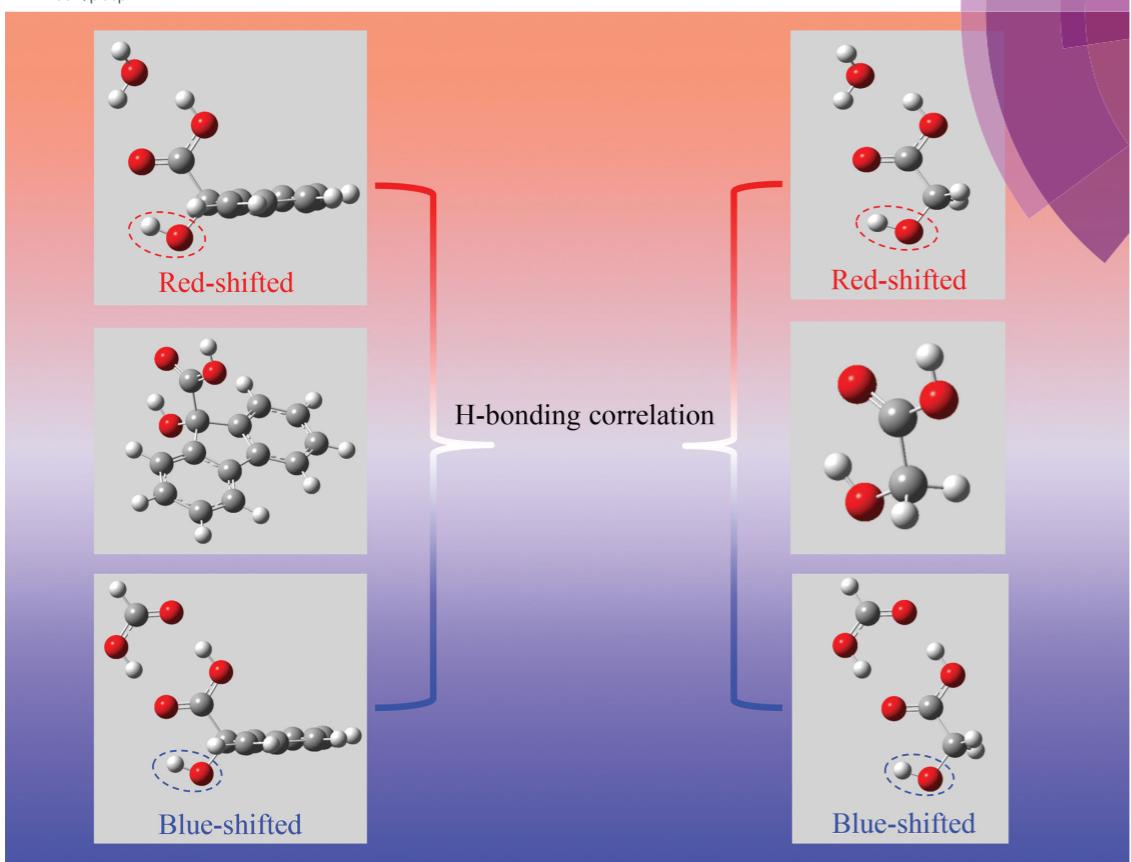


Motivation I

Volume 19 | Number 36 | 28 September 2017 | Pages 24373–25226

PCCP

Physical Chemistry Chemical Physics
rsc.li/pccp



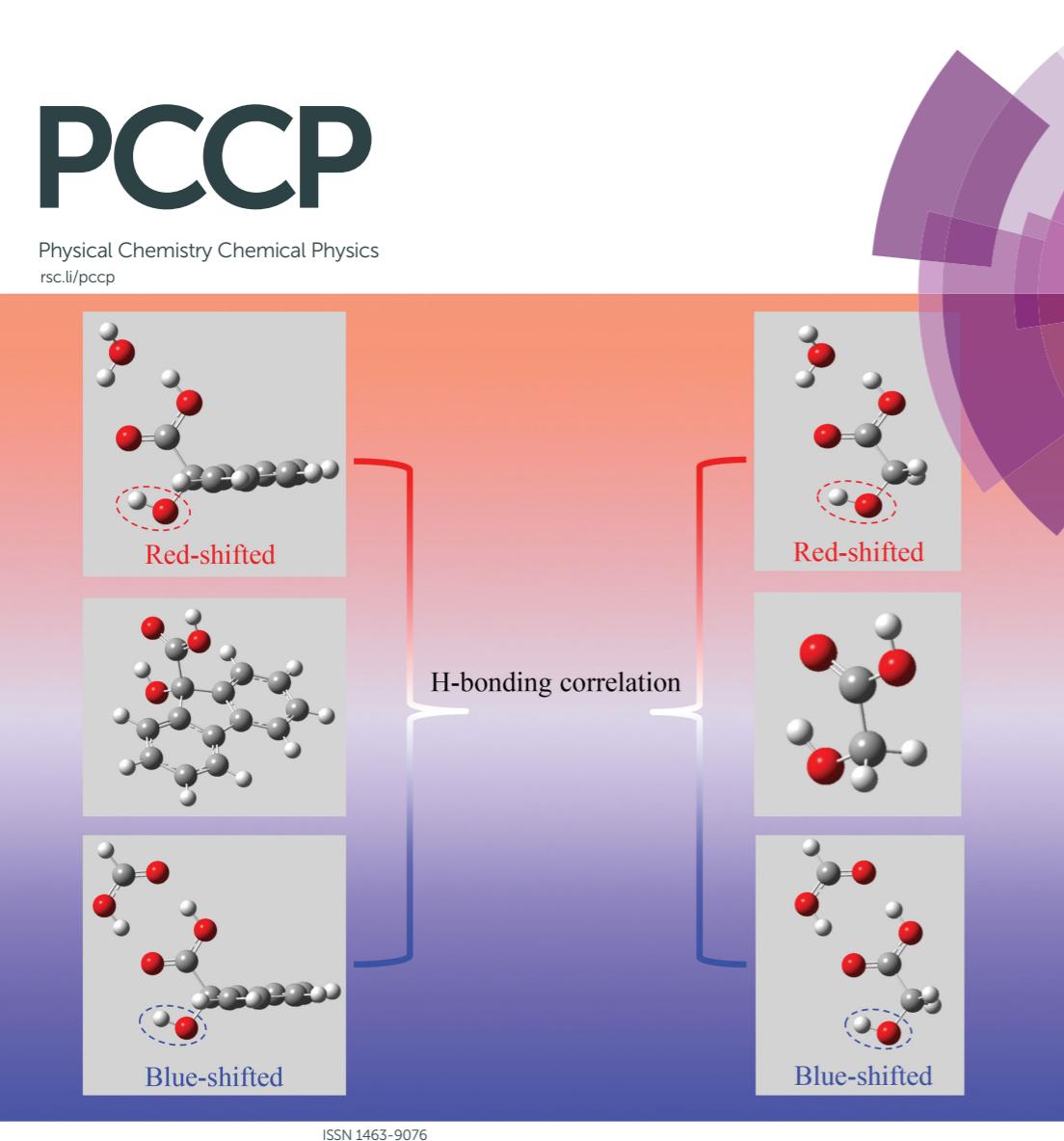
ISSN 1463-9076



PERSPECTIVE

Quanli Gu, Peifeng Su, Yong Xia, Zhijun Yang, Carl O. Trindle and Joseph L. Kneé
Quantitative probing of subtle interactions among H-bonds in alpha hydroxy carboxylic acid complexes

Motivation I



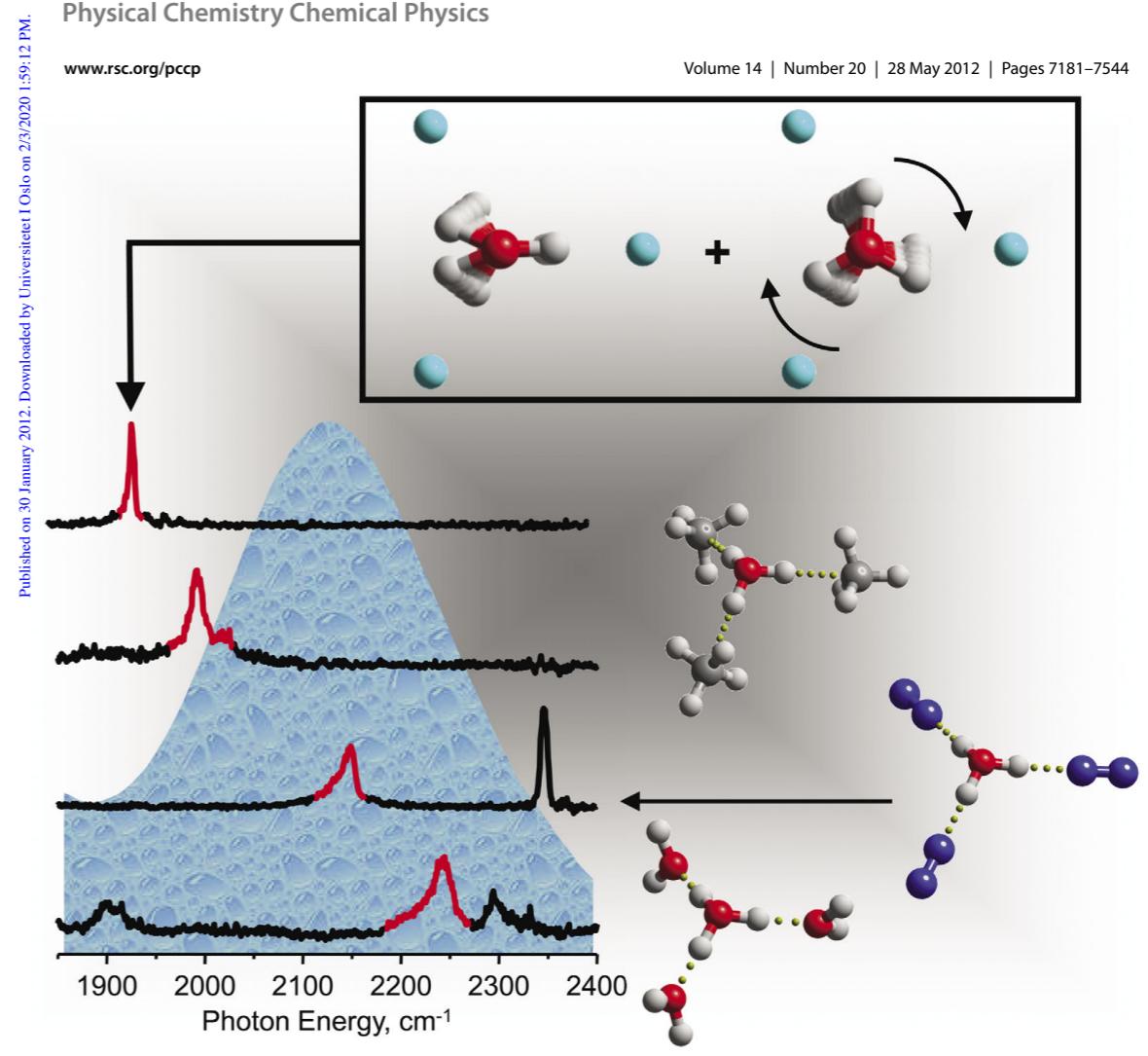
PERSPECTIVE
Quanli Gu, Peifeng Su, Yong Xia, Zhijun Yang, Carl O. Trindle and Joseph L. Kneé
Quantitative probing of subtle interactions among H-bonds in alpha hydroxy carboxylic acid complexes

PCCP

Physical Chemistry Chemical Physics

www.rsc.org/pccp

Volume 14 | Number 20 | 28 May 2012 | Pages 7181–7544



Carboxylic Acids

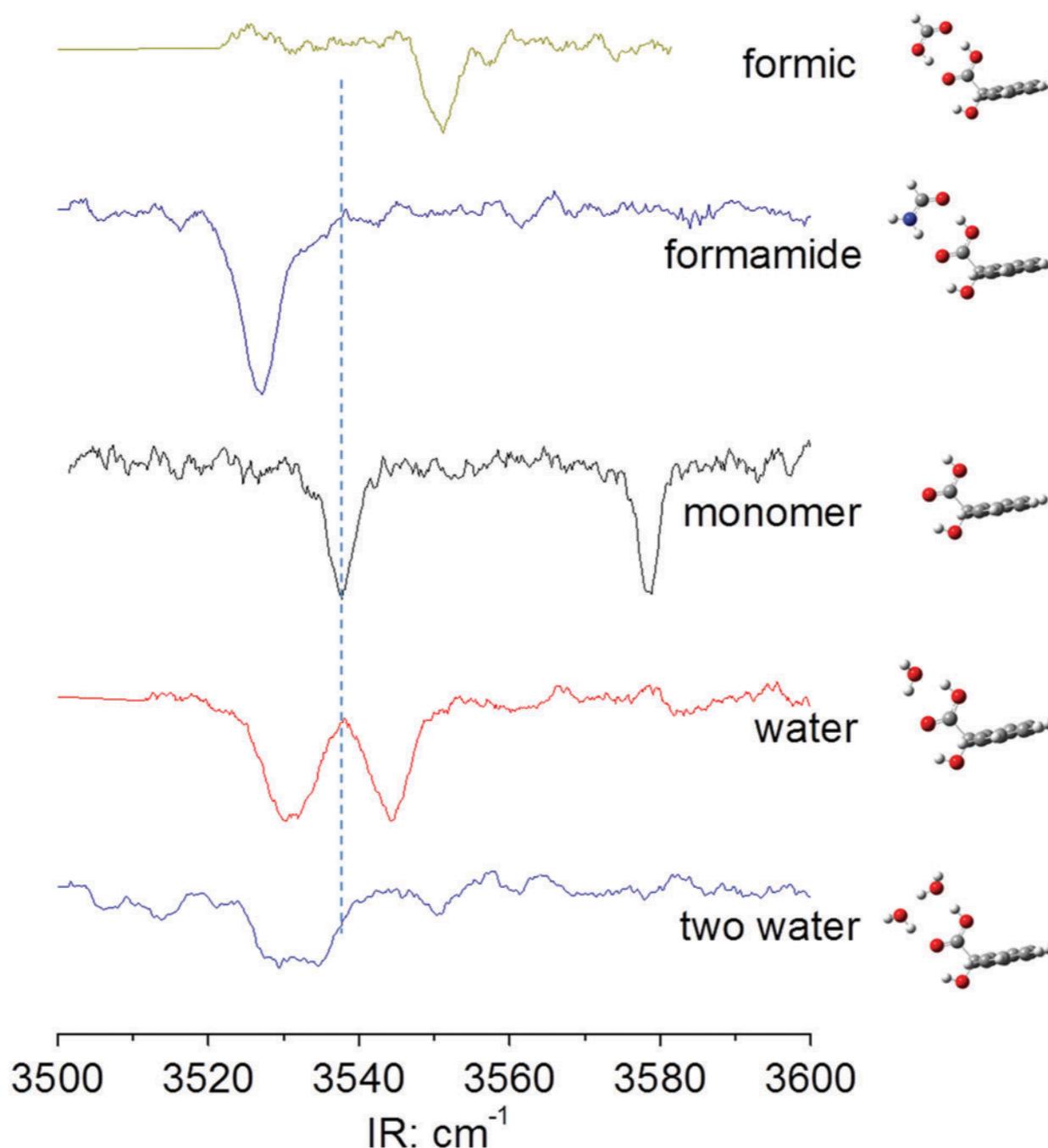


Fig. 6 Experimental IR spectra of the alpha OH stretch frequency in 9HFCA and its complexes. The vertical dash line indicates the alpha OH frequency position in monomer.

Motivation II

 nature astronomy

Letter | Published: 13 January 2020

Infrared detection of aliphatic organics on a cometary nucleus

A. Raponi , M. Ciarniello, [...] C. Leyrat

Nature Astronomy (2020) | Cite this article

192 Accesses | 24 Altmetric | Metrics

Abstract

The European Space Agency's Rosetta mission¹ has acquired unprecedented measurements of the surface of the nucleus of comet 67P/Churyumov-Gerasimenko (hereafter, 67P), the composition of which, as determined by in situ and remote-sensing instruments, including the VIRTIS instrument², seems to be an assemblage of ices, minerals and organic material³. We performed a refined analysis of infrared observations of the nucleus of 67P carried out by the VIRTIS-M hyperspectral imager. We find that the overall shape of the infrared spectrum of 67P is similar to that of other carbon-rich outer Solar System objects, suggesting a possible genetic link with them. More importantly, we also confirm the complex spectral structure of the wide 2.8–3.6 μm absorption feature populated by fainter bands. Among these, we unambiguously identify the presence of aliphatic organics by their ubiquitous 3.38 μm, 3.42 μm and 3.47 μm bands. This infrared detection of aliphatic species on a cometary surface has strong implications for the evolutionary history of the primordial Solar System and is evidence that comets provide an evolutionary link between interstellar material and Solar System bodies⁴.



pubs.acs.org/JPL

Letter

Infrared Spectroscopy of Neutral Water Dimer Based on a Tunable Vacuum Ultraviolet Free Electron Laser

Bingbing Zhang,[#] Yong Yu,[#] Zhaojun Zhang,[#] Yang-Yang Zhang,[#] Shukang Jiang, Qinming Li, Shuo Yang, Han-Shi Hu, Weiqing Zhang, Dongxu Dai, Guorong Wu, Jun Li,* Dong H. Zhang,* Xueming Yang,* and Ling Jiang*

 Cite This: *J. Phys. Chem. Lett.* 2020, 11, 851–855

 Read Online

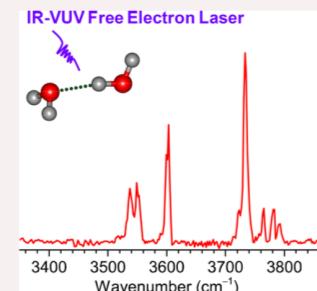
ACCESS

 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: Infrared (IR) spectroscopy provides detailed structural and dynamical information on clusters at the fingerprint level. Herein, we demonstrate the capability of a tunable vacuum ultraviolet free electron laser (VUV-FEL) for selective detection of a wide variety of neutral water clusters and for recording the size-dependent IR spectra. The present technique does not require the presence of an ultraviolet chromophore or a dipole moment and is generally applicable for IR spectroscopy of neutral clusters free from confinement. To show the features of our technique, we report here the IR spectra of neutral water dimer in the OH stretch region, providing benchmarks for theoretical study of the accurate description of hydrogen bonding structures involved in liquid water and ice. Quantum mechanical calculations on a 12-dimensional ab initio potential energy surface are utilized to simulate the anharmonic vibrational spectra of water dimer. These results help to resolve the controversy of the exact vibrational assignment of each band feature of the water dimer.



pubs.acs.org/JPCA

Article

Gas-Phase Reaction Kinetics of Pyruvic Acid with OH Radicals: The Role of Tunneling, Complex Formation, and Conformational Structure

Jonathan R. Church, Veronica Vaida, and Rex T. Skodje*

 Cite This: *J. Phys. Chem. A* 2020, 124, 790–800

 Read Online

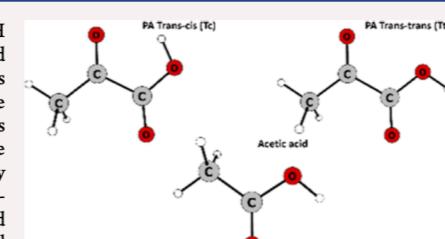
ACCESS

 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: The gas-phase reaction of pyruvic acid (PA) with the OH radical is studied theoretically using accurate quantum chemistry and transition state theory. Two chemically distinct H-atom abstraction reactions and two distinct OH addition reactions have been identified. The rate coefficients for these four processes were calculated. Quantum tunneling was included in each rate using the small curvature tunneling method. The influence of the conformational structure of PA was found to be particularly intriguing. While the trans-cis structure was found to dominantly react by H-atom abstraction from the methyl site, the trans-trans conformer was found to react mostly through H-atom abstraction from the acid site. A general formalism was developed to model the kinetics of the reactions that involve multiple conformers, interconverting prereactive complexes, and multiple transition states. Comparison of the results obtained with available experimental rate observations reveals agreement with the trans-trans conformer of PA but disagreement with the results obtained for a full statistical mixture of reagents. The role of these reactions in the atmospheric processing of PA is discussed.



Machine learning Potential Energy Surface

JCTC
Journal of Chemical Theory and Computation

Article
Cite This: *J. Chem. Theory Comput.* 2019, 15, 4386–4398

pubs.acs.org/JCTC

Gaussian Process and Neural Net PyTorch

Energy Surface

PES-Learn: An Open-Source Software Package for the Automated Generation of Machine Learning Models of Molecular Potential Energy Surfaces

Adam S. Abbott,[†] Justin M. Turney,[†] Boyi Zhang,[†] Daniel G. A. Smith,[‡] Doaa Altarawy,^{‡,§} and Henry F. Schaefer, III^{§,†}

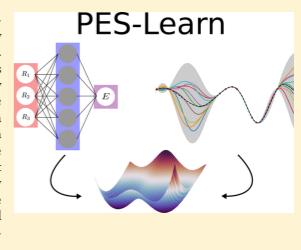
[†]Center for Computational Quantum Chemistry, The University of Georgia, Athens, Georgia 30602, United States

[‡]Molecular Sciences Software Institute, Virginia Tech, Blacksburg, Virginia 24061, United States

[§]Computer and Systems Engineering Department, Alexandria University, Alexandria, Egypt

 Supporting Information

ABSTRACT: We introduce a free and open-source software package (PES-Learn) which largely automates the process of producing high-quality machine learning models of molecular potential energy surfaces (PESs). PES-Learn incorporates a generalized framework for producing grid points across a PES that is compatible with most electronic structure theory software. The newly generated or externally supplied PES data can then be used to train and optimize neural network or Gaussian process models in a completely automated fashion. Robust hyperparameter optimization schemes designed specifically for molecular PES applications are implemented to ensure that the best possible model for the data set is fit with high quality. The performance of PES-Learn toward fitting a few semiglobal PESs from the literature is evaluated. We also demonstrate the use of PES-Learn machine learning models in carrying out high-level vibrational configuration interaction computations on water and formaldehyde.



Abbott, A. S.; Turney, J. M.; Zhang, B.; Smith, D. G. A.; Altarawy, D.; Schaefer, H. F., 3rd, "PES-Learn: An Open-Source Software Package for the Automated Generation of Machine Learning Models of Molecular Potential Energy Surfaces". *J Chem Theory Comput* **2019**, 15 (8), 4386-4398.

Kernel Ridge Regression PyTorch

Energy Conserving Surface

- Gradient based
- Symmetry

Computer Physics Communications 240 (2019) 38–45

Contents lists available at ScienceDirect
Computer Physics Communications
journal homepage: www.elsevier.com/locate/cpc


sGDML: Constructing accurate and data efficient molecular force fields using machine learning



Stefan Chmiela^a, Huziel E. Sauceda^b, Igor Poltavsky^c, Klaus-Robert Müller^{a,d,e,*}, Alexandre Tkatchenko^{c,**}

^aMachine Learning Group, Technische Universität Berlin, 10587 Berlin, Germany

^bFritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany

^cPhysics and Materials Science Research Unit, University of Luxembourg L-1511 Luxembourg, Luxembourg

^dDepartment of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 02841, Republic of Korea

^eMax Planck Institute for Informatics, Stuhlsatzweg 66123 Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 22 October 2018
Received in revised form 6 February 2019
Accepted 10 February 2019
Available online 28 February 2019

Keywords:
Machine learning potential
Machine learning force field
Ab initio molecular dynamics
Path integral molecular dynamics
Coupled cluster calculations
Molecular property prediction
Quantum chemistry
Gradient domain machine learning

ABSTRACT

We present an optimized implementation of the recently proposed symmetric gradient domain machine learning (sGDML) model. The sGDML model is able to faithfully reproduce global potential energy surfaces (PES) for molecules with a few dozen atoms from a limited number of user-provided reference molecular conformations and the associated atomic forces. Here, we introduce a Python software package to reconstruct and evaluate custom sGDML force fields (FFs), without requiring in-depth knowledge about the details of the model. A user-friendly command-line interface offers assistance through the complete process of model creation, in an effort to make this novel machine learning approach accessible to broad practitioners. Our paper serves as a documentation, but also includes a practical application example of how to reconstruct and use a PBE0+MBD FF for paracetamol. Finally, we show how to interface sGDML with the FF simulation engine ASE (Larsen et al., 2017) and i-PI (Kapil et al., 2019) to run numerical experiments including structure optimization, classical and path integral molecular dynamics and nudged elastic band calculations.

© 2019 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Chmiela, Stefan; Sauceda, Huziel E.; Poltavsky, Igor; Müller, Klaus-Robert; Tkatchenko, Alexandre, "sGDML: Constructing accurate and data efficient molecular force fields using machine learning". *Computer Physics Communications* **2019**, 240, 38-45.

Machine learning generated PES: Approach

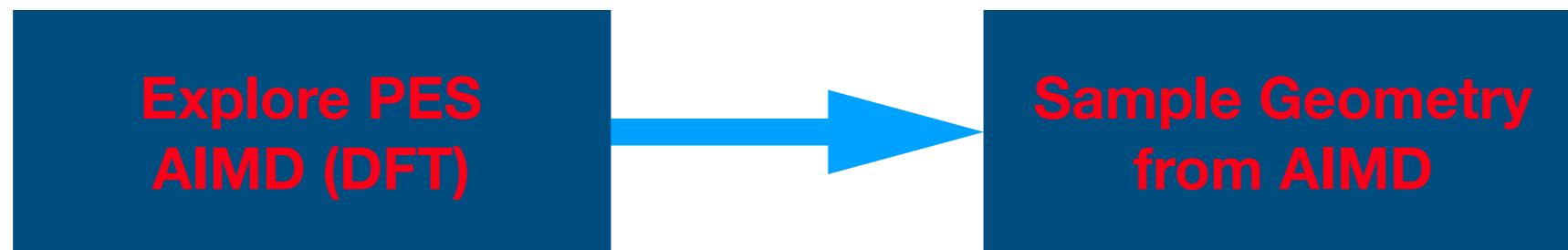
Machine learning generated PES: Approach

**Explore PES
AIMD (DFT)**

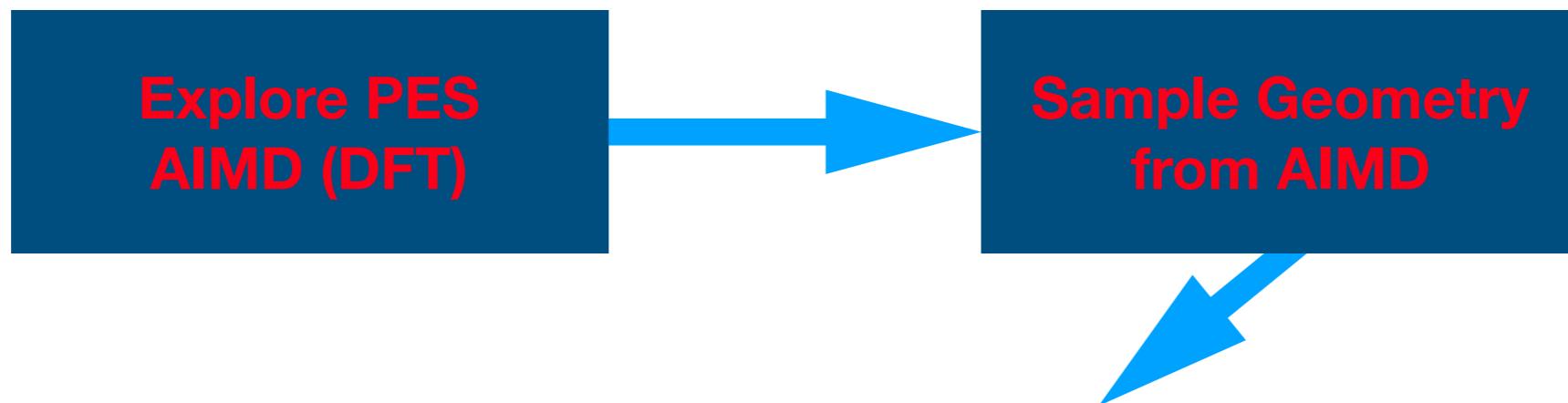
Machine learning generated PES: Approach



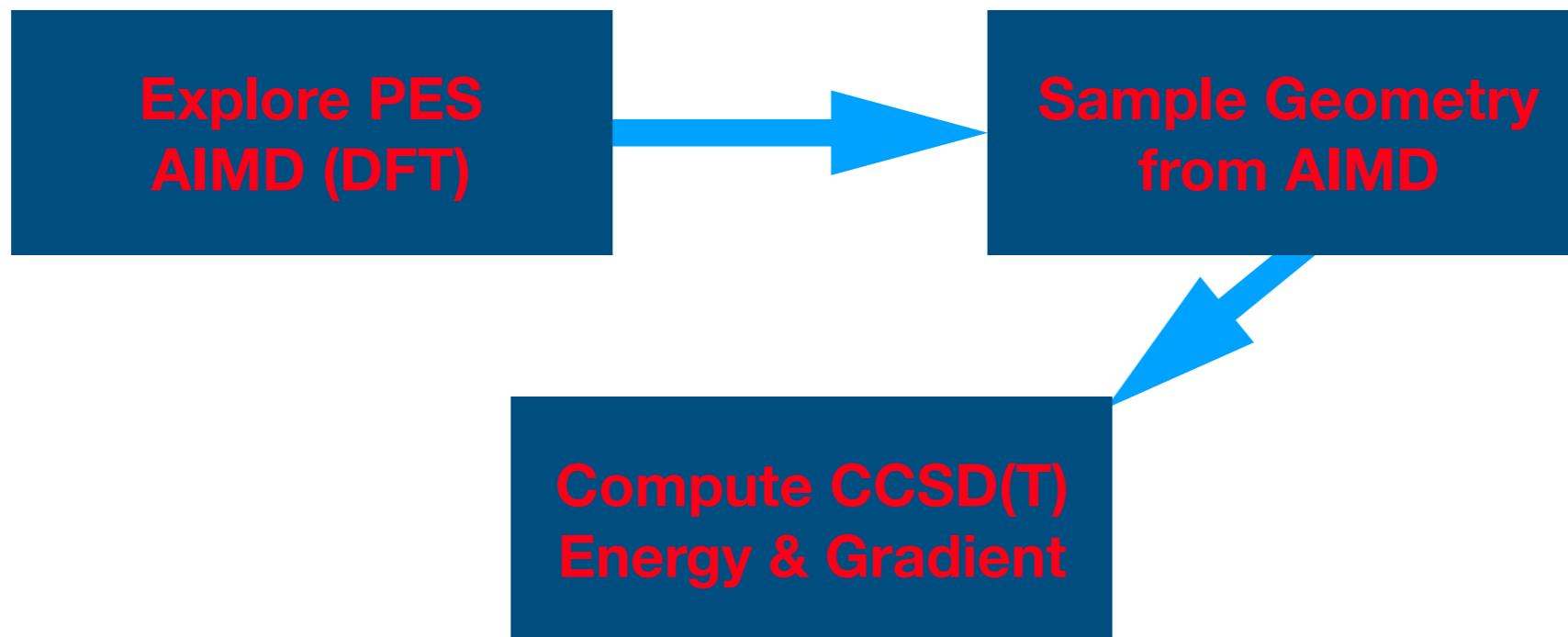
Machine learning generated PES: Approach



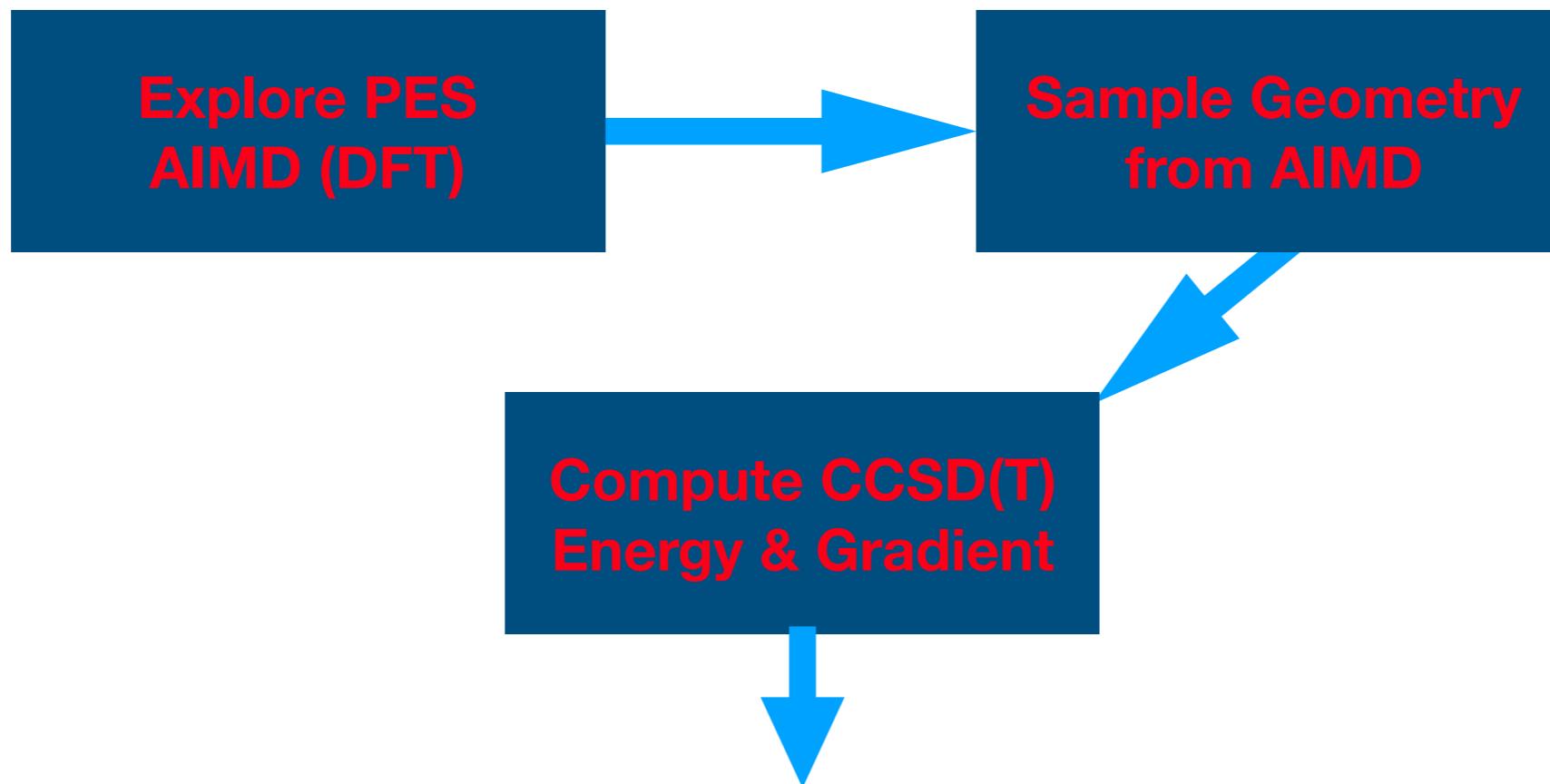
Machine learning generated PES: Approach



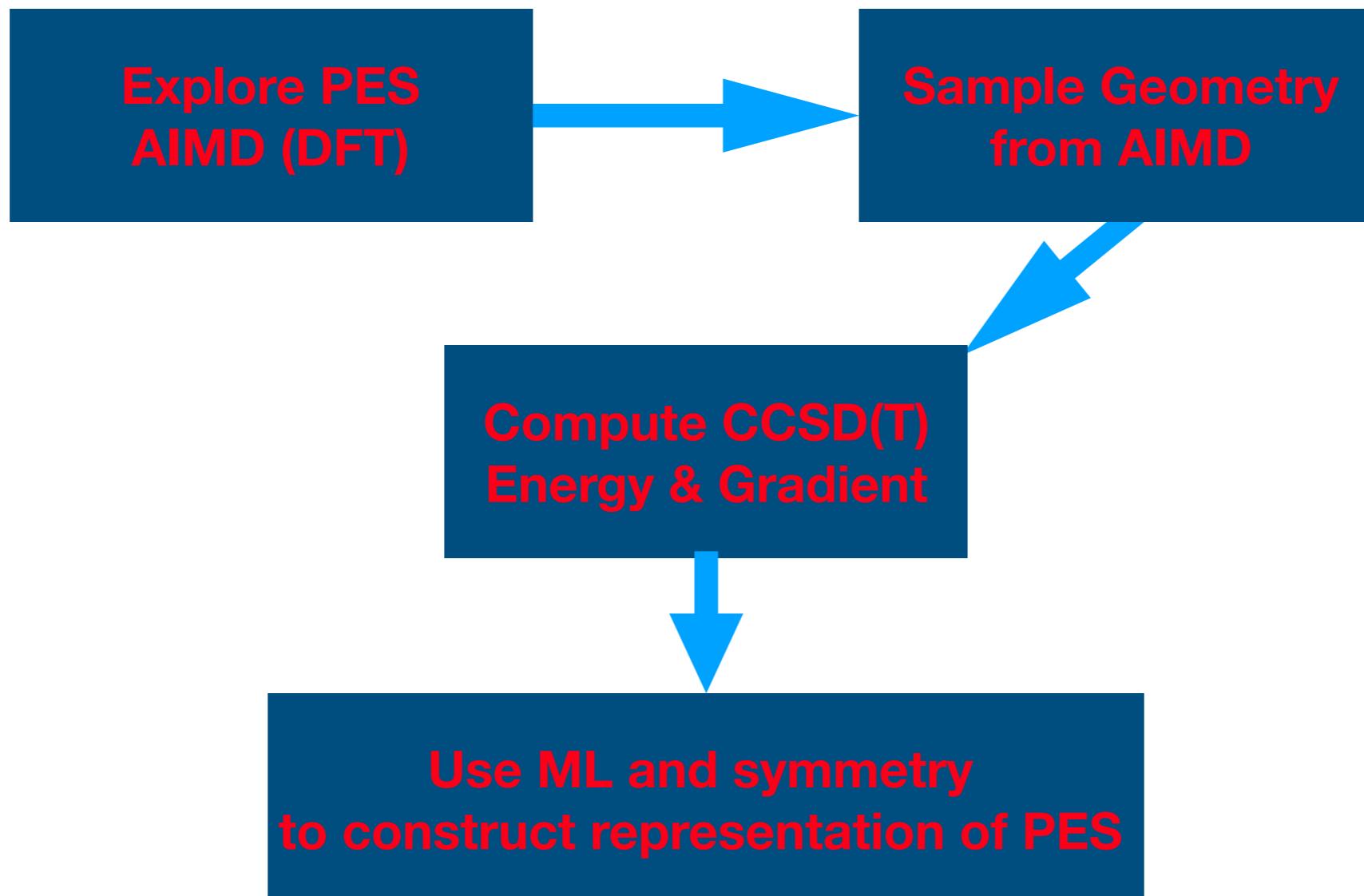
Machine learning generated PES: Approach



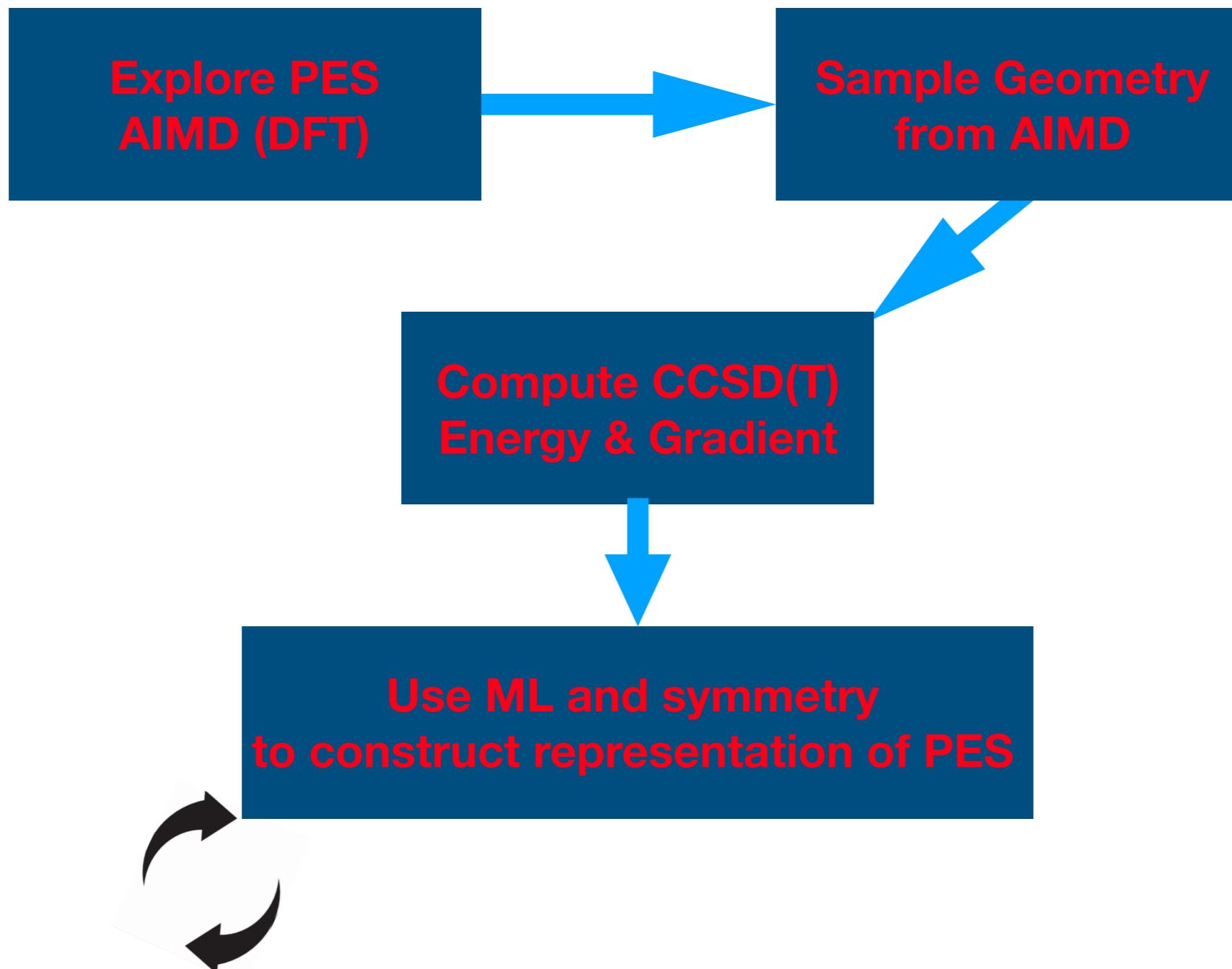
Machine learning generated PES: Approach



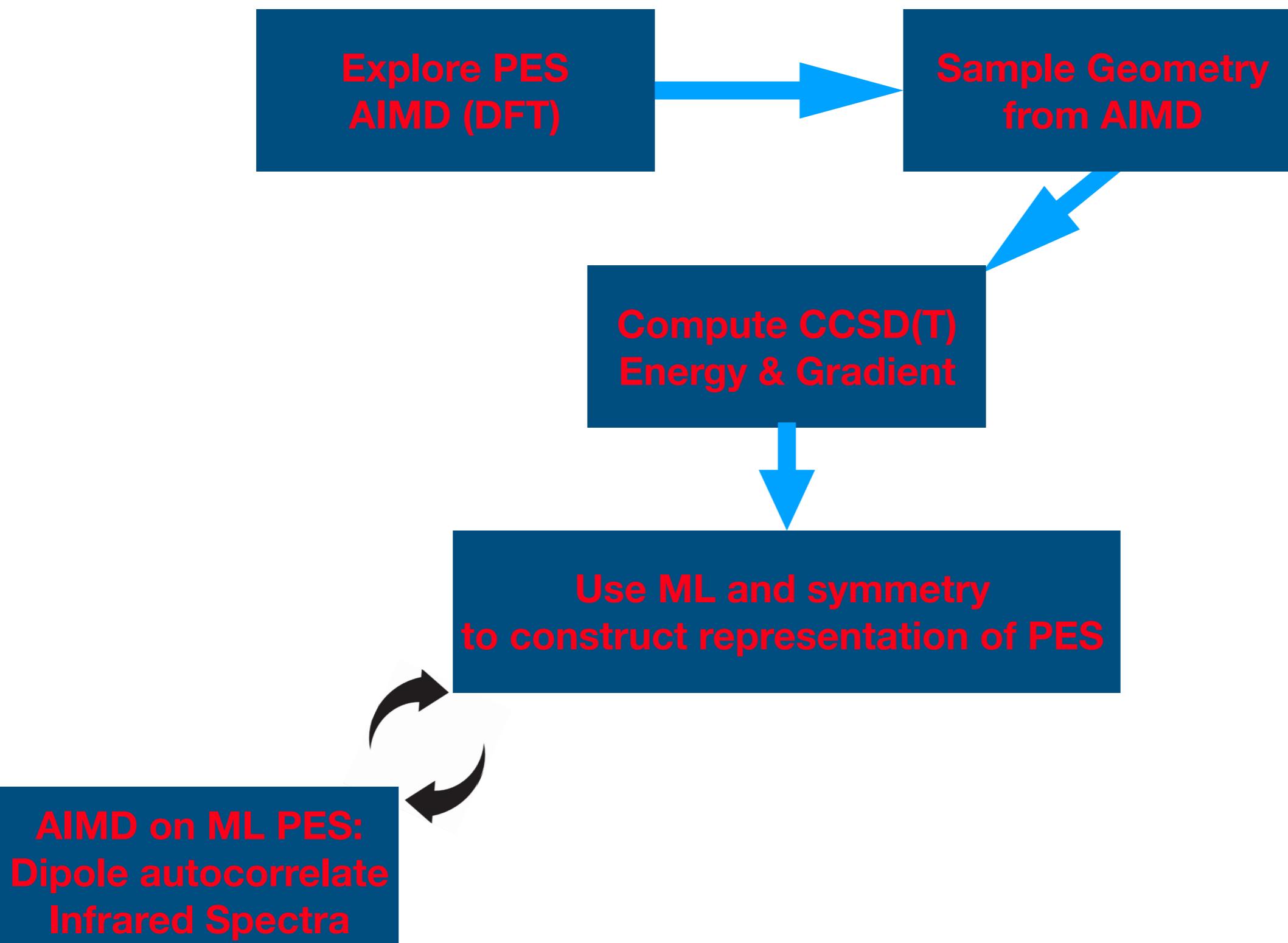
Machine learning generated PES: Approach



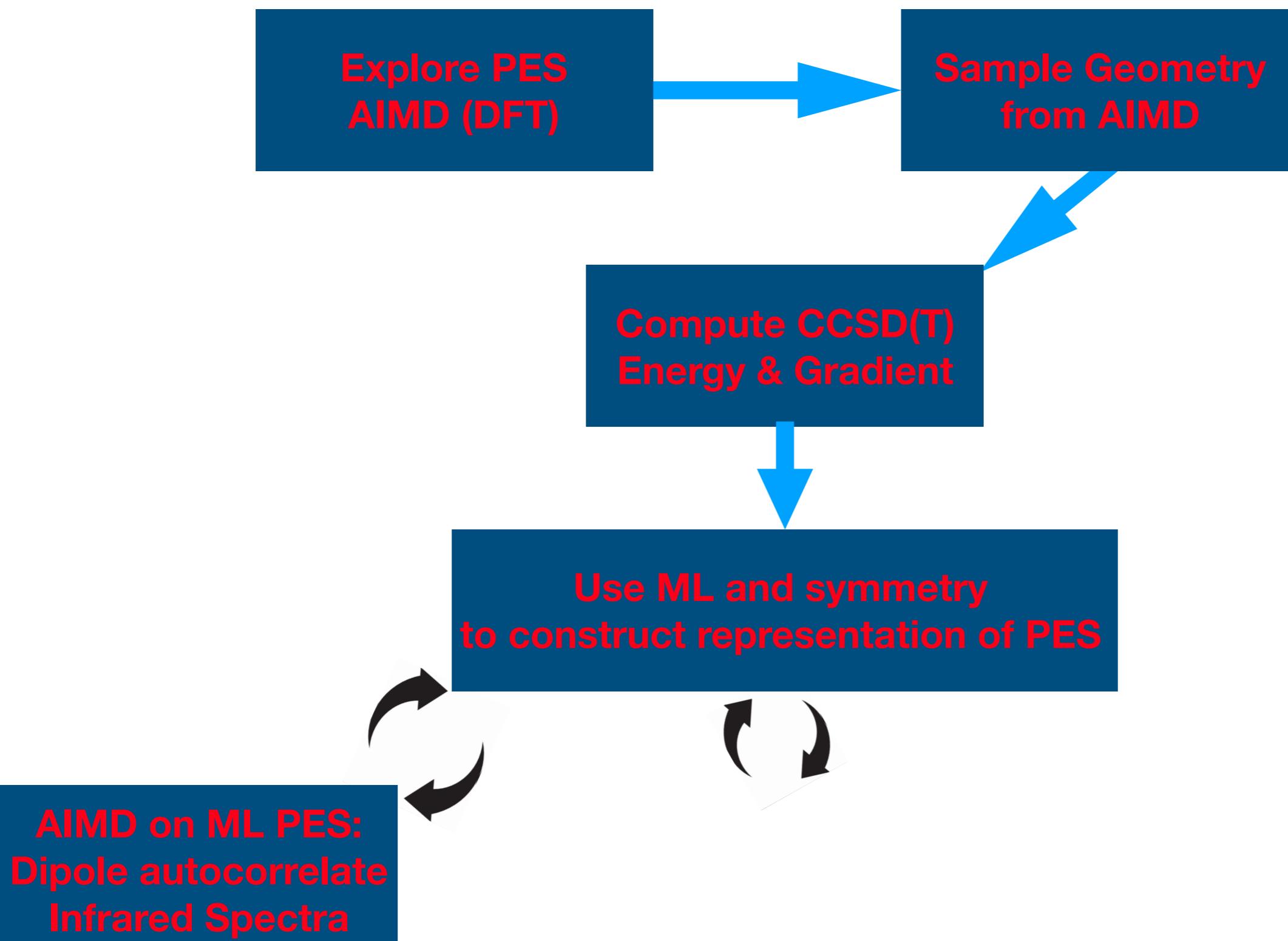
Machine learning generated PES: Approach



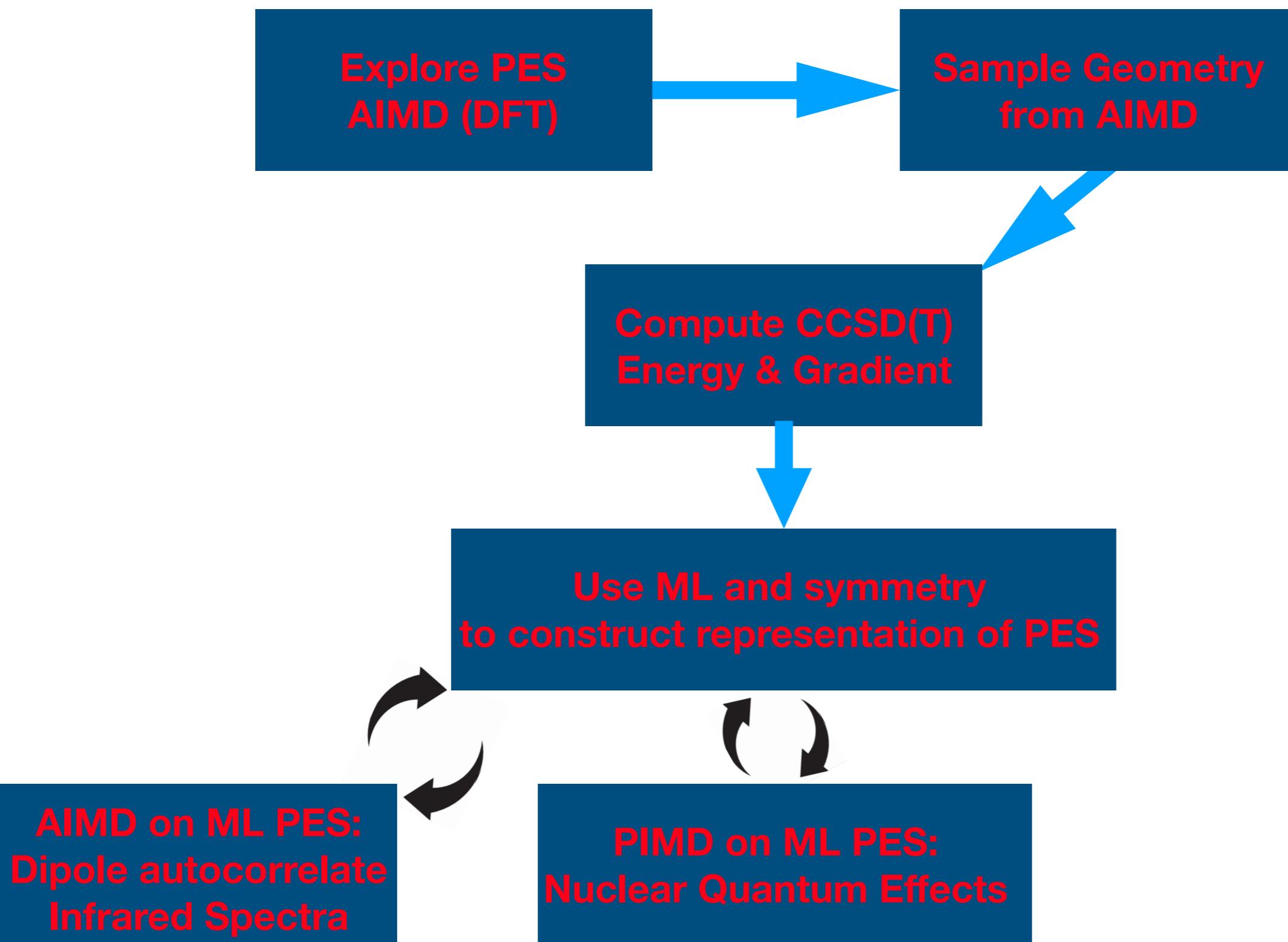
Machine learning generated PES: Approach



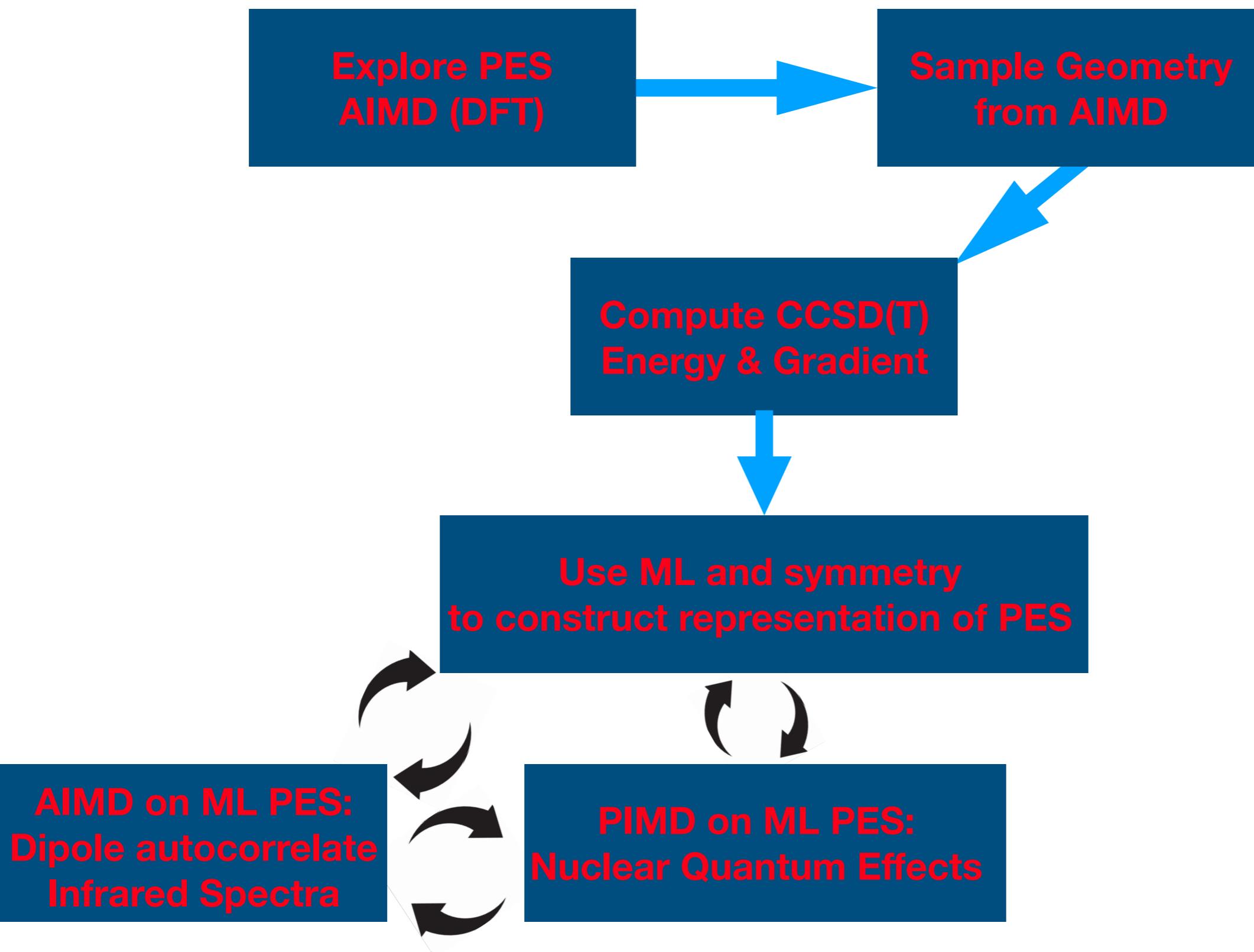
Machine learning generated PES: Approach



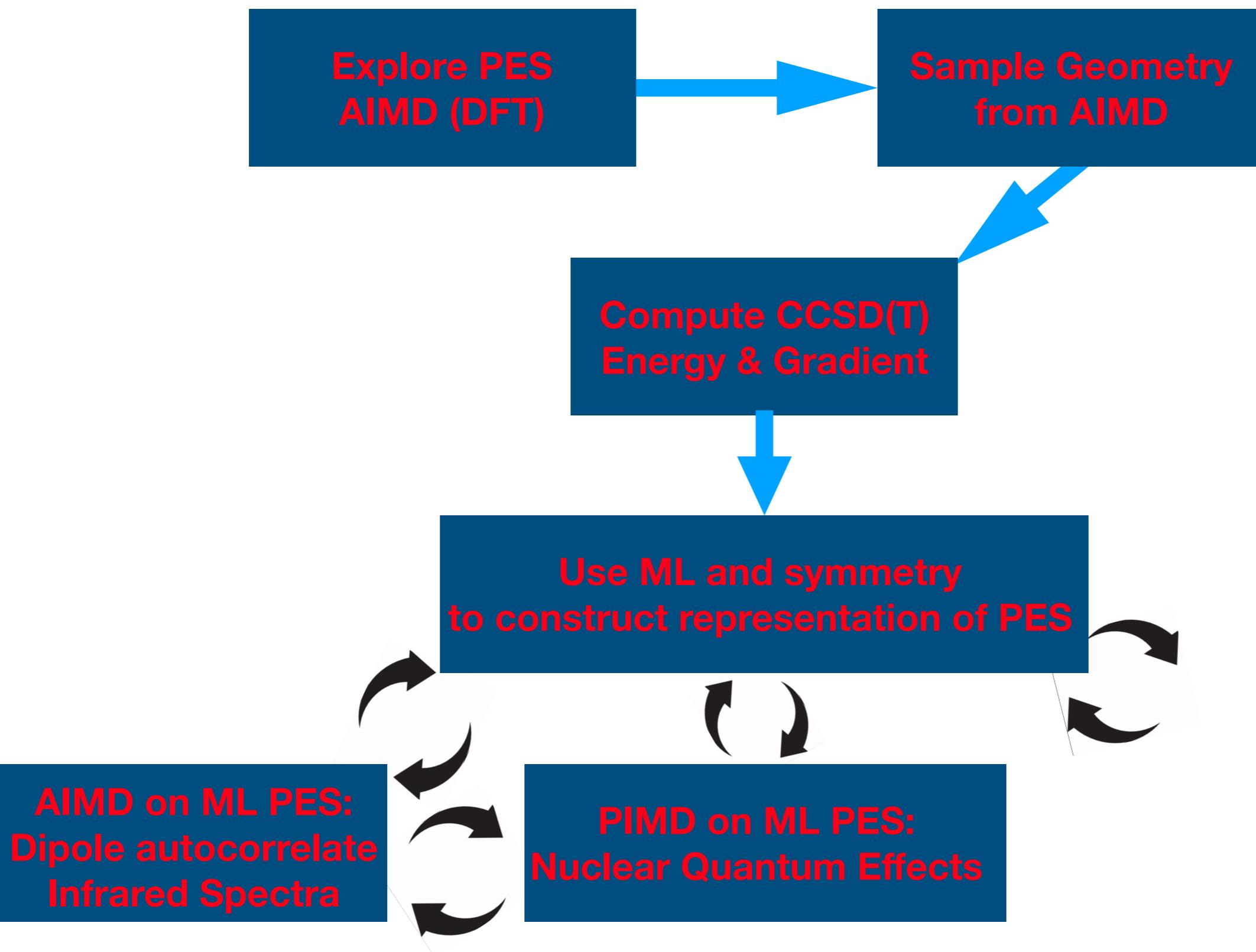
Machine learning generated PES: Approach



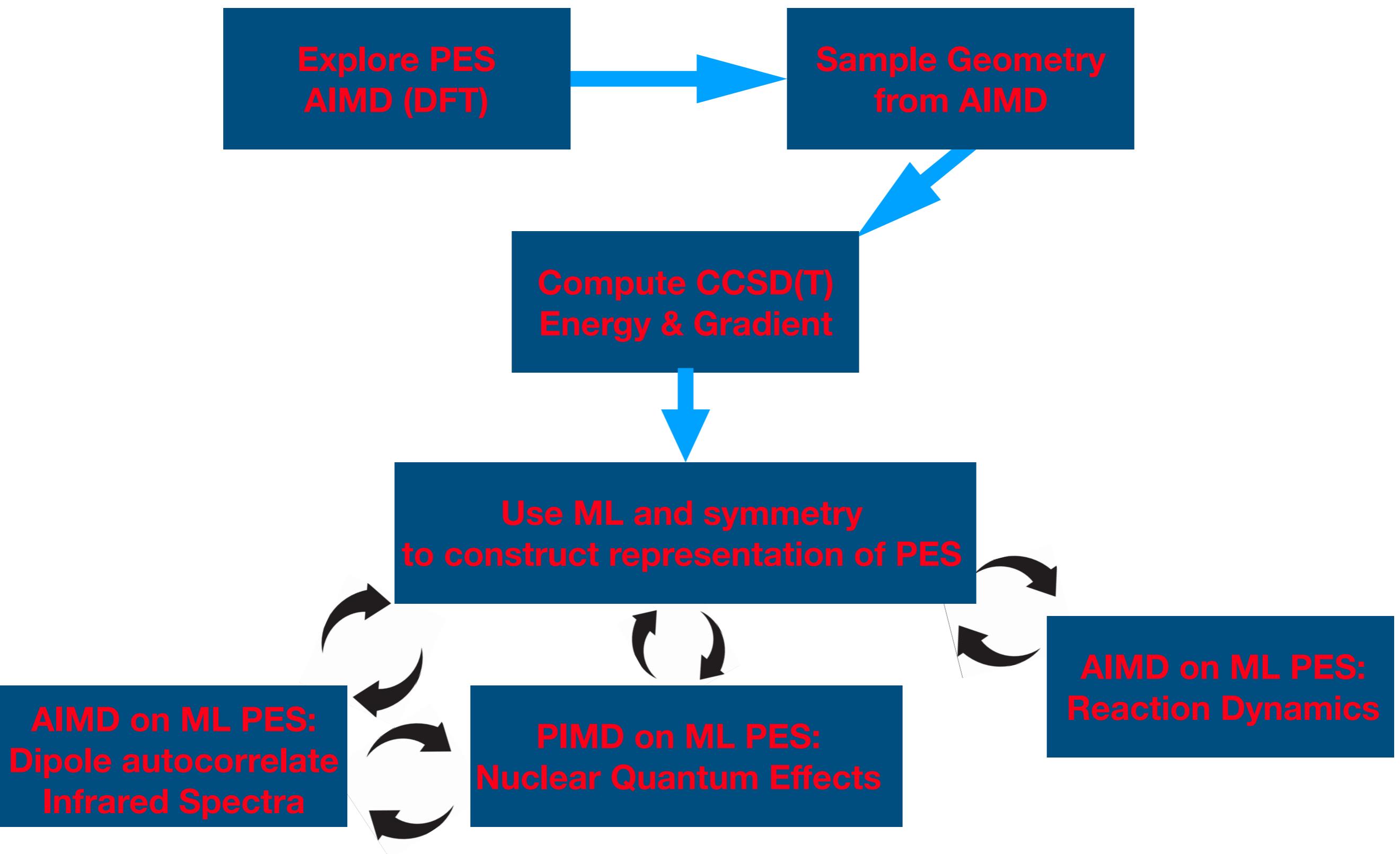
Machine learning generated PES: Approach



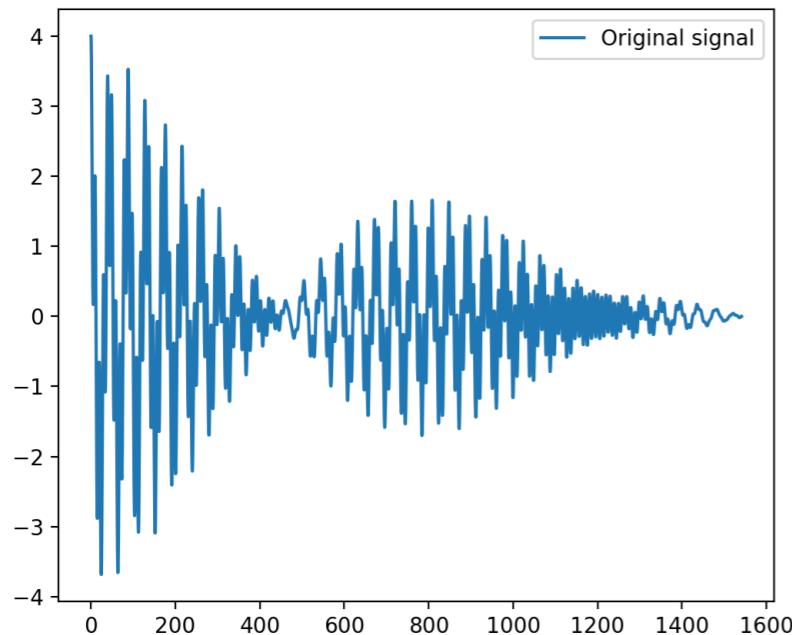
Machine learning generated PES: Approach



Machine learning generated PES: Approach



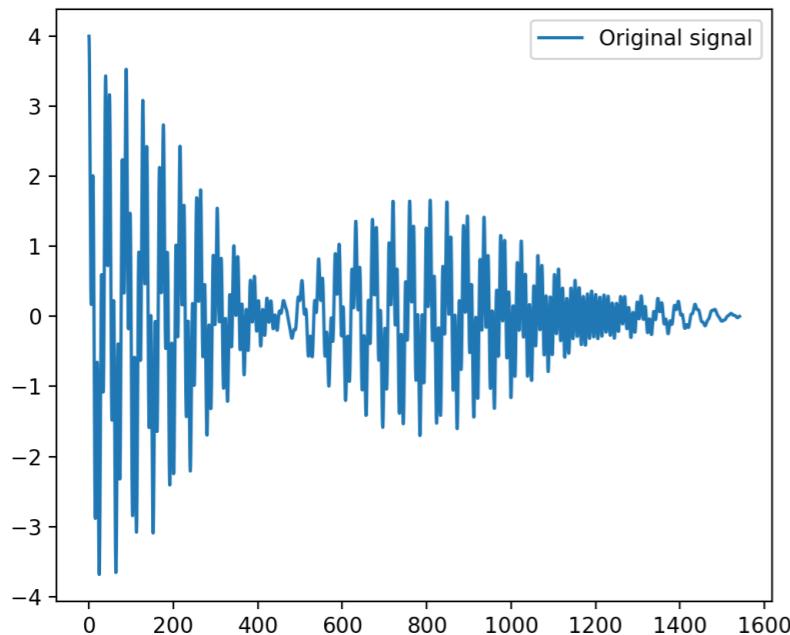
Computational Methodology



- 20+ psec AIMD direct dynamics
- NVE, Velocity Verlet
- CFOUR
 - CCSD(T)/cc-pVTZ
- GAMESS 2018 R3
 - B3LYP/cc-pVTZ (Dunning)
- i-PI
 - RPMD
- CP2K
 - DFTB
 - PIMD
- Python code:
 - Localization
 - Dipole autocorrelation function
 - FFT
- Benchmark
 - Infrared Spectra
 - Static Vibrational Computations
 - APFD/6-311++G(3df,3pd), VPT2

IApproach: Chmiela, S.; Sauceda, H. E.; Muller, K. R.; Tkatchenko, A., "Towards exact molecular dynamics simulations with machine-learned force fields". Nat Commun **2018**, 9 (1), 3887.

Computational Methodology

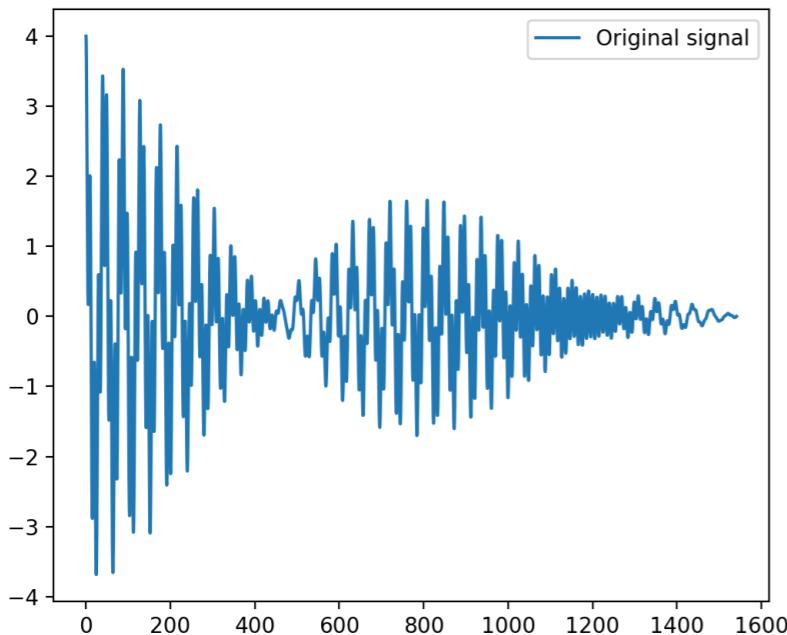


$$I_{cl} = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp[-i\omega t] \langle \vec{\mu}(0) \cdot \vec{\mu}(t) \rangle$$

- 20+ psec AIMD direct dynamics
- NVE, Velocity Verlet
- CFOUR
 - CCSD(T)/cc-pVTZ
- GAMESS 2018 R3
 - B3LYP/cc-pVTZ (Dunning)
- i-PI
 - RPMD
- CP2K
 - DFTB
 - PIMD
- Python code:
 - Localization
 - Dipole autocorrelation function
 - FFT
- Benchmark
 - Infrared Spectra
 - Static Vibrational Computations
 - APFD/6-311++G(3df,3pd), VPT2

IApproach: Chmiela, S.; Sauceda, H. E.; Muller, K. R.; Tkatchenko, A., "Towards exact molecular dynamics simulations with machine-learned force fields". Nat Commun **2018**, 9 (1), 3887.

Computational Methodology



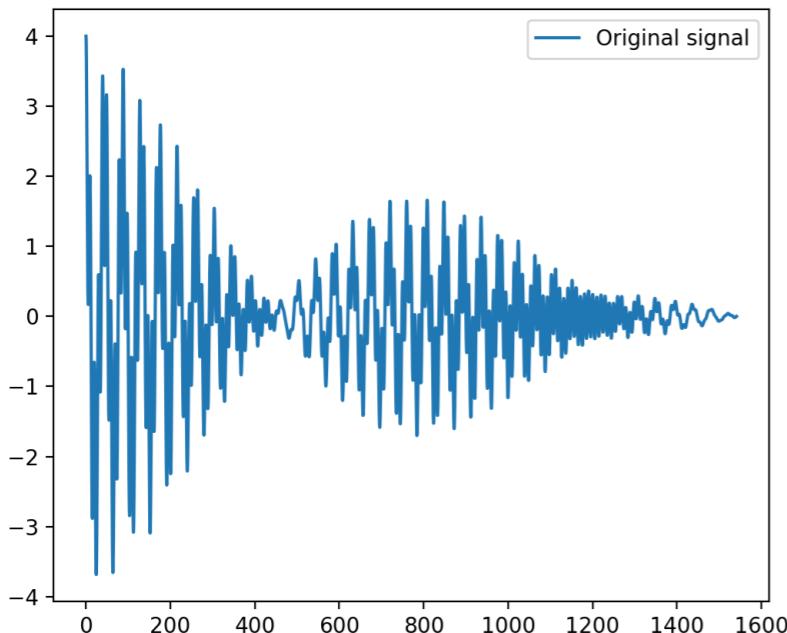
$$I_{cl} = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp[-i\omega t] \langle \vec{\mu}(0) \cdot \vec{\mu}(t) \rangle$$

$$\alpha_{QC}(\omega) = \left[\frac{4\pi^2 \omega}{3V\hbar c n(\omega)} \right] \beta \hbar \omega I_{cl}(\omega).$$

- 20+ psec AIMD direct dynamics
 - NVE, Velocity Verlet
 - CFOUR
 - CCSD(T)/cc-pVTZ
 - GAMESS 2018 R3
 - B3LYP/cc-pVTZ (Dunning)
 - i-PI
 - RPMD
 - CP2K
 - DFTB
 - PIMD
 - Python code:
 - Localization
 - Dipole autocorrelation function
 - FFT
- Benchmark
- Infrared Spectra
 - Static Vibrational Computations
 - APFD/6-311++G(3df,3pd), VPT2

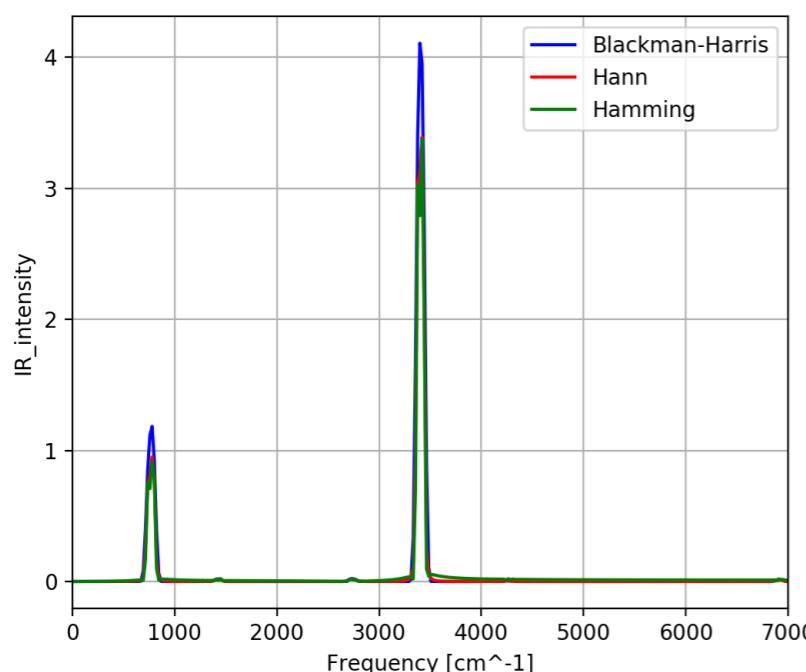
IApproach: Chmiela, S.; Sauceda, H. E.; Muller, K. R.; Tkatchenko, A., "Towards exact molecular dynamics simulations with machine-learned force fields". Nat Commun **2018**, 9 (1), 3887.

Computational Methodology



$$I_{cl} = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp[-i\omega t] \langle \vec{\mu}(0) \cdot \vec{\mu}(t) \rangle$$

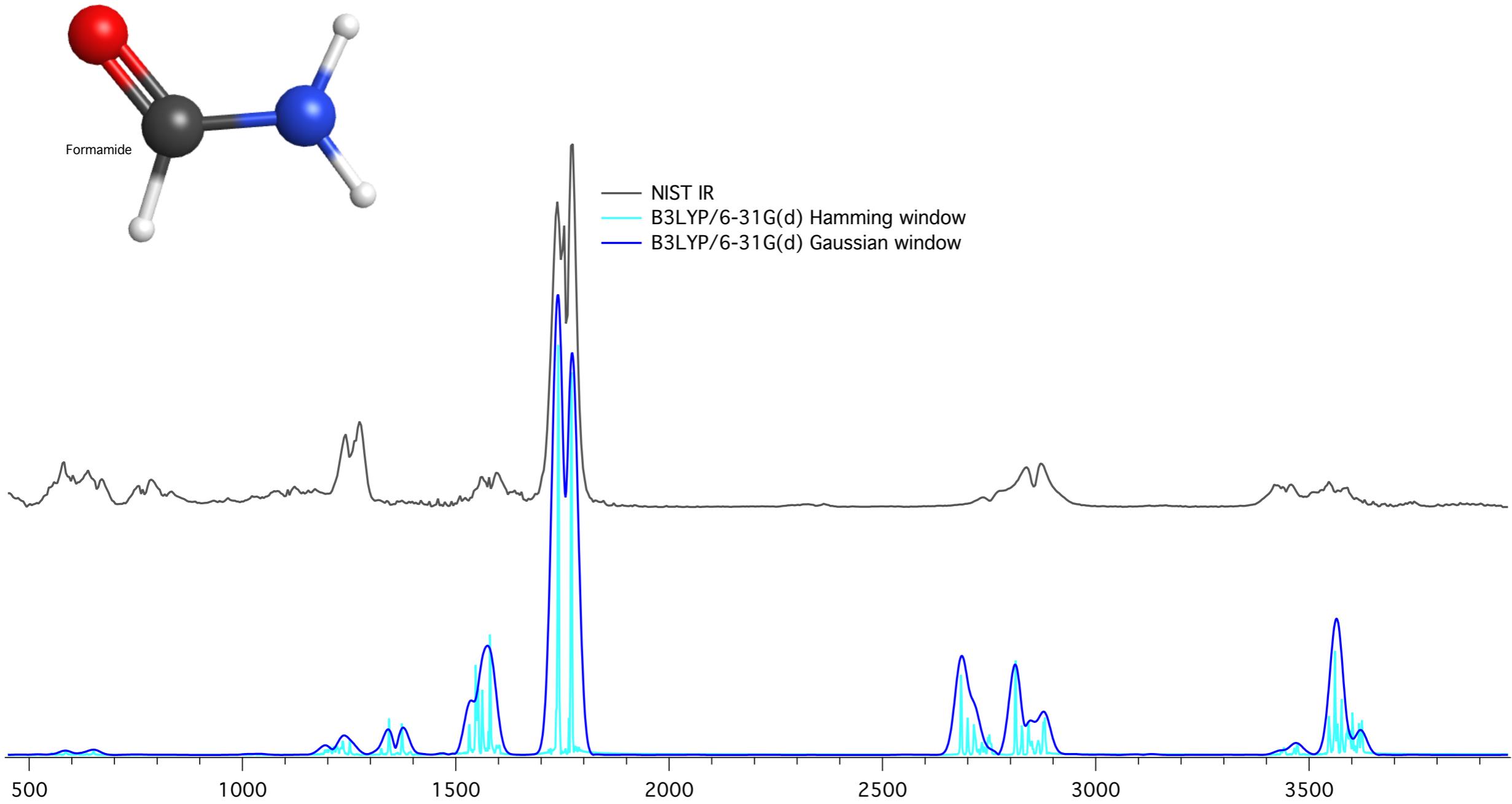
$$\alpha_{QC}(\omega) = \left[\frac{4\pi^2\omega}{3V\hbar cn(\omega)} \right] \beta \hbar \omega I_{cl}(\omega).$$



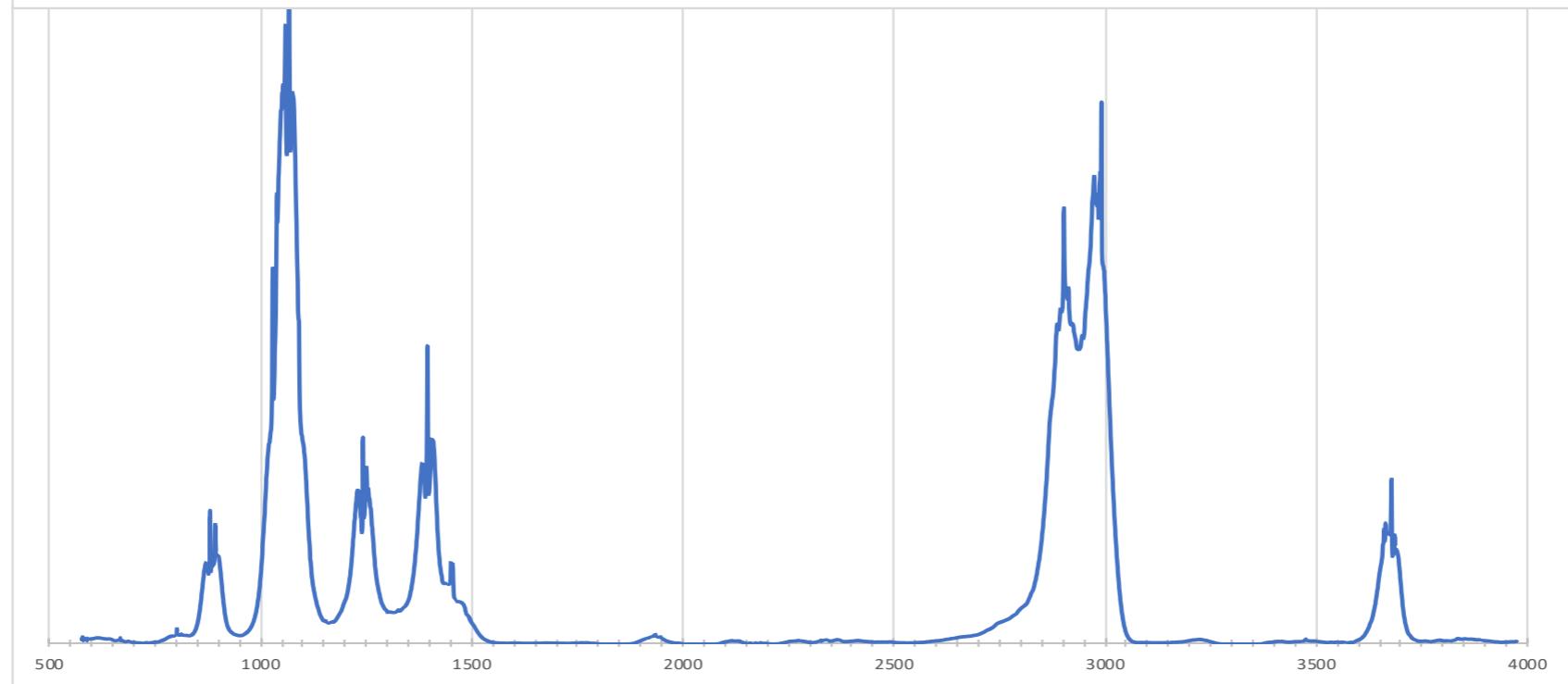
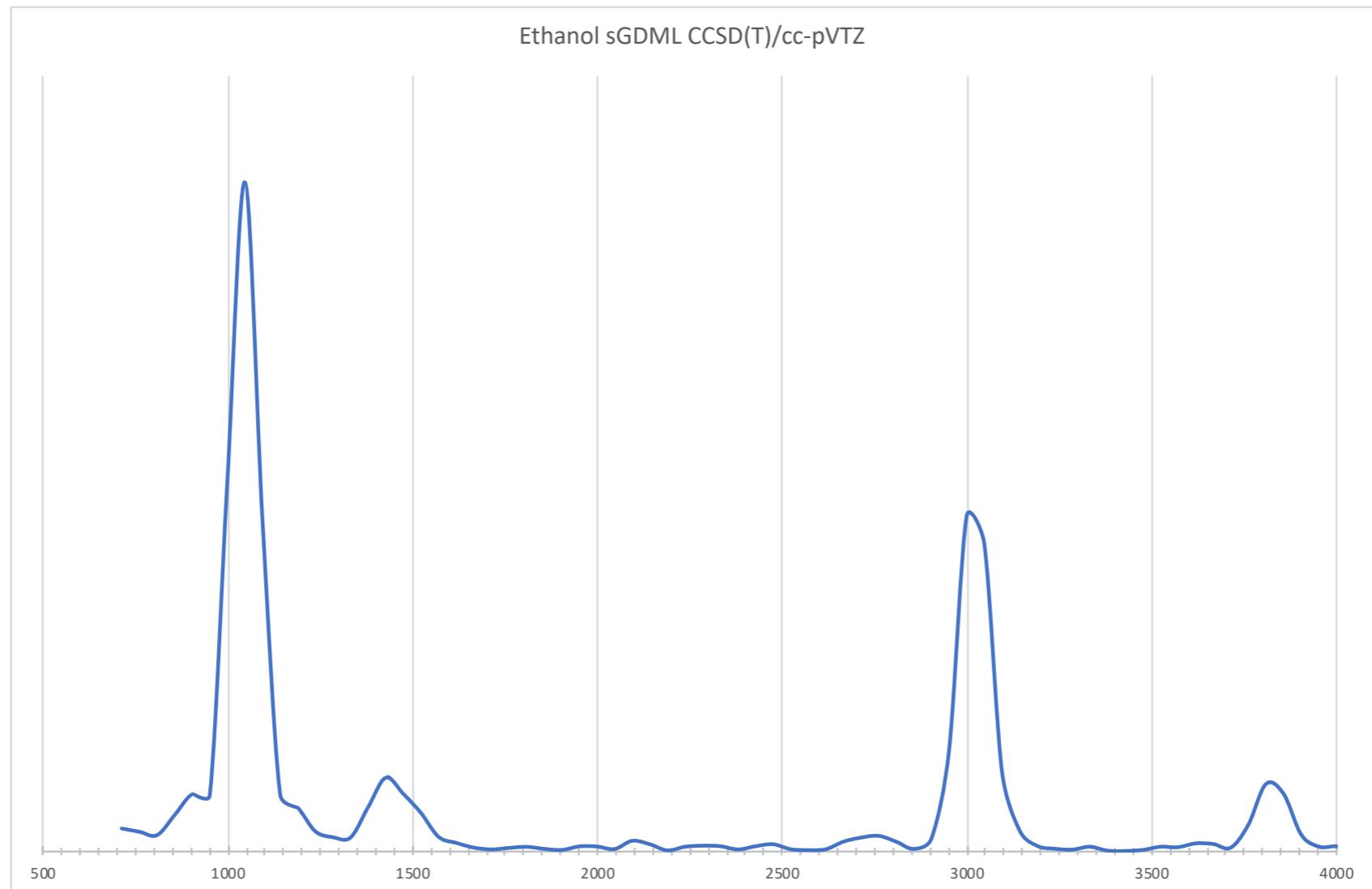
- 20+ psec AIMD direct dynamics
 - NVE, Velocity Verlet
 - CFOUR
 - CCSD(T)/cc-pVTZ
 - GAMESS 2018 R3
 - B3LYP/cc-pVTZ (Dunning)
 - i-PI
 - RPMD
 - CP2K
 - DFTB
 - PIMD
 - Python code:
 - Localization
 - Dipole autocorrelation function
 - FFT
- Benchmark
- Infrared Spectra
 - Static Vibrational Computations
 - APFD/6-311++G(3df,3pd), VPT2

IApproach: Chmiela, S.; Sauceda, H. E.; Muller, K. R.; Tkatchenko, A., "Towards exact molecular dynamics simulations with machine-learned force fields". Nat Commun **2018**, 9 (1), 3887.

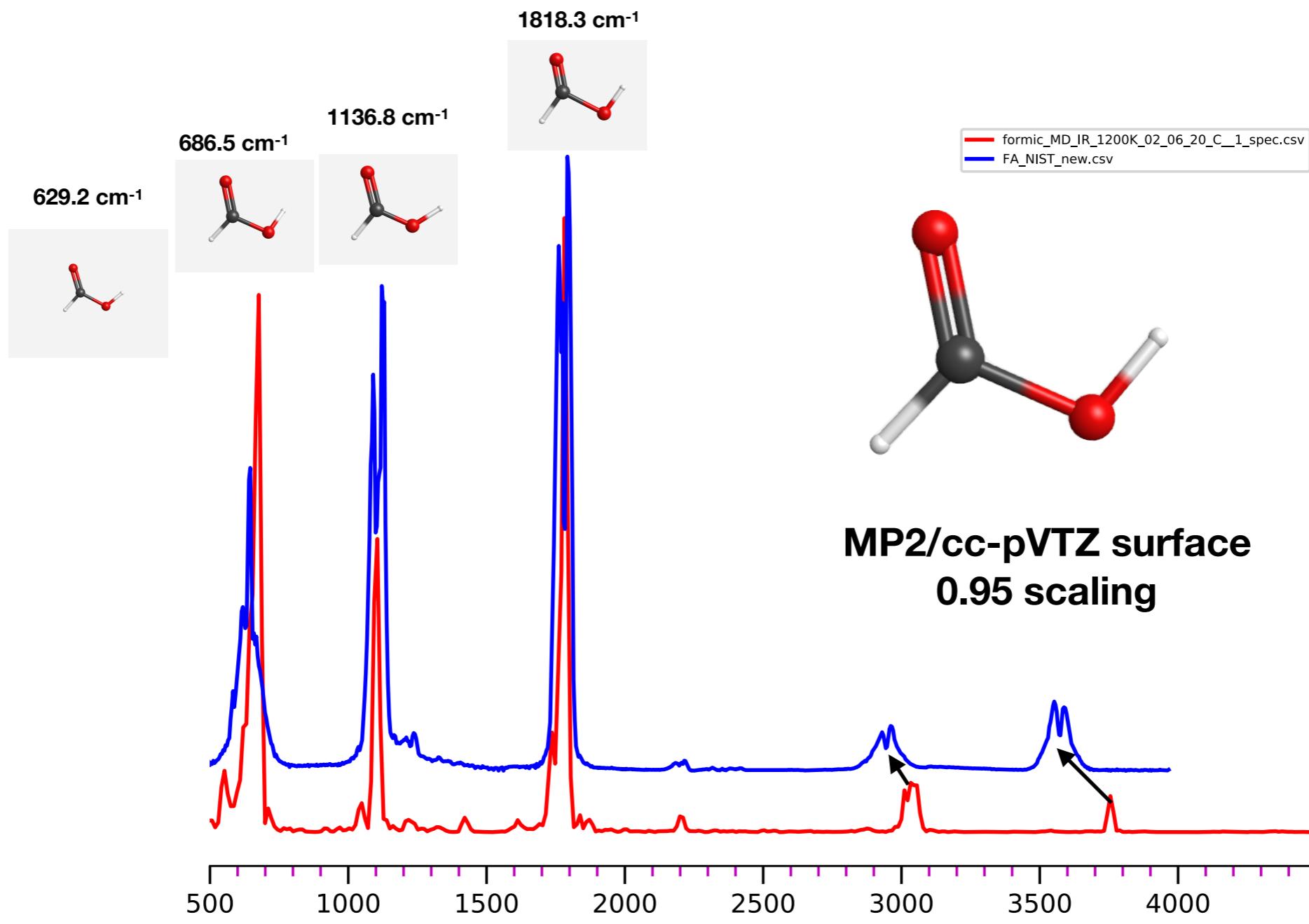
Benchmark: Formamide



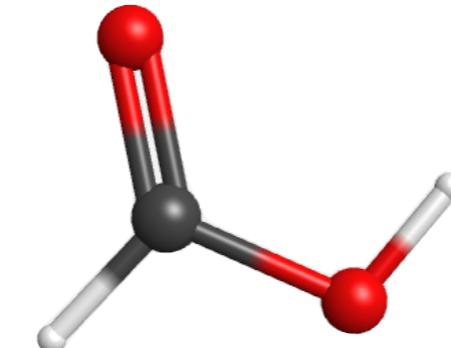
Ethanol



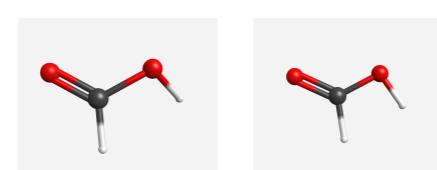
Formic acid



MP2/cc-pVTZ frequencies

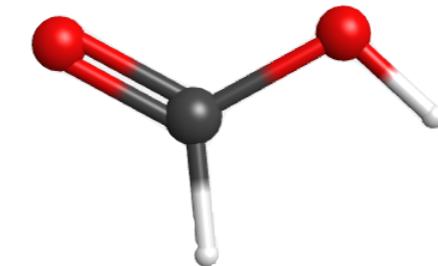


**MP2/cc-pVTZ surface
0.95 scaling**

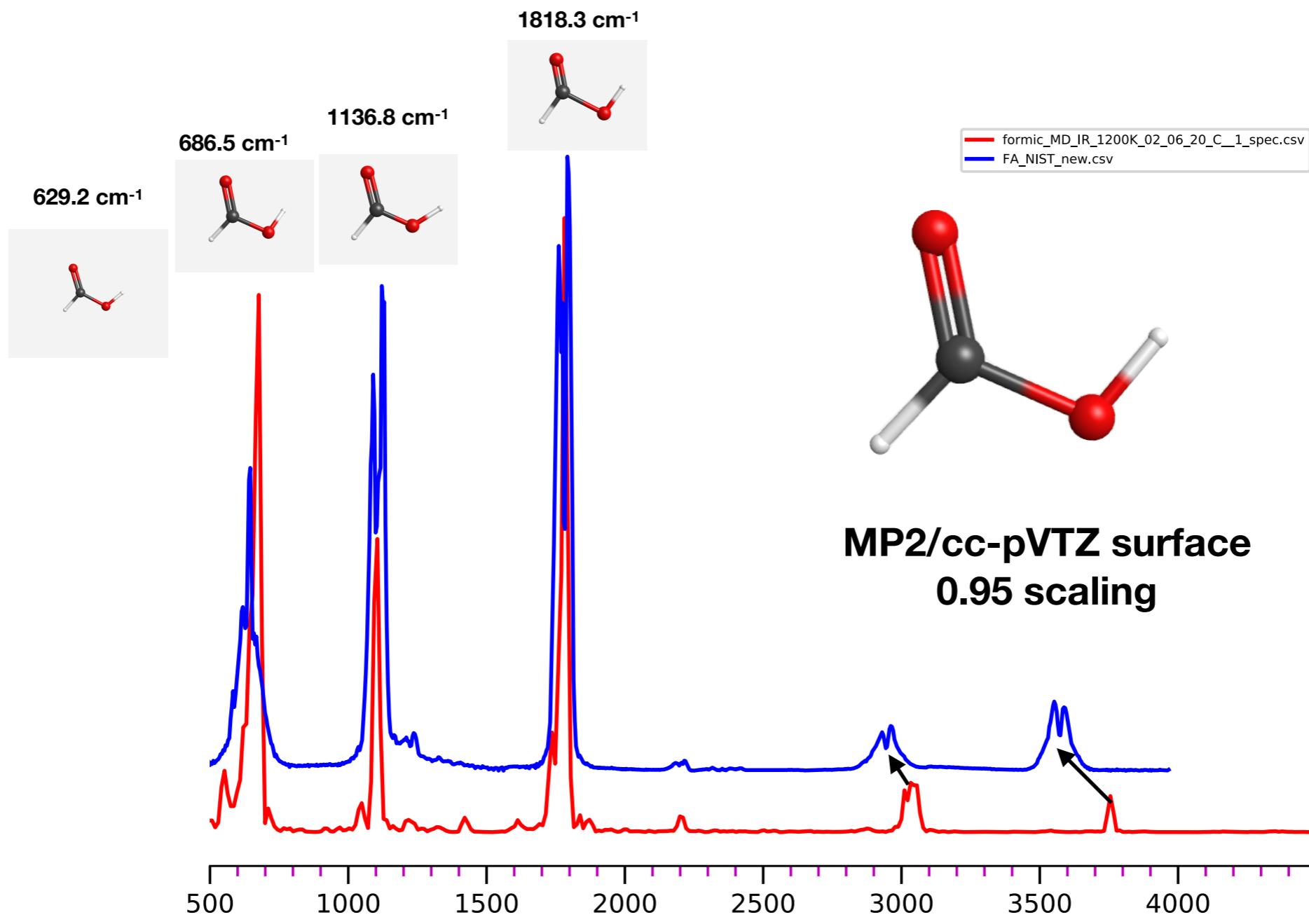


529.1 cm⁻¹

1130.4 cm⁻¹

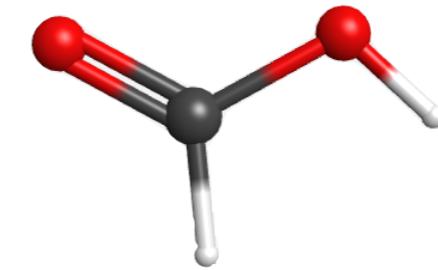


Formic acid

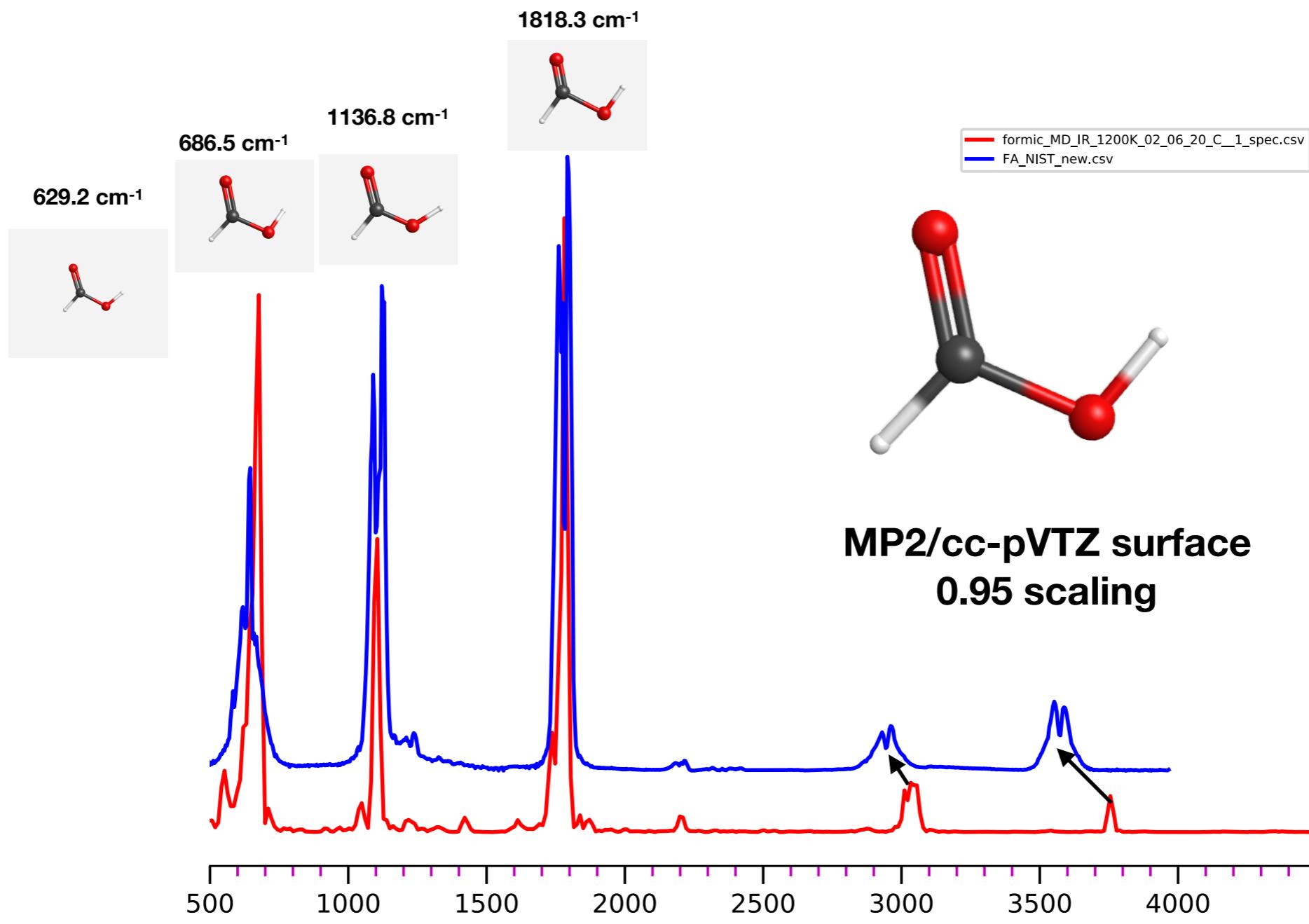


MP2/cc-pVTZ frequencies

529.1 cm⁻¹ 1130.4 cm⁻¹

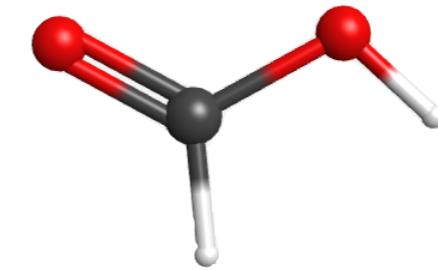


Formic acid

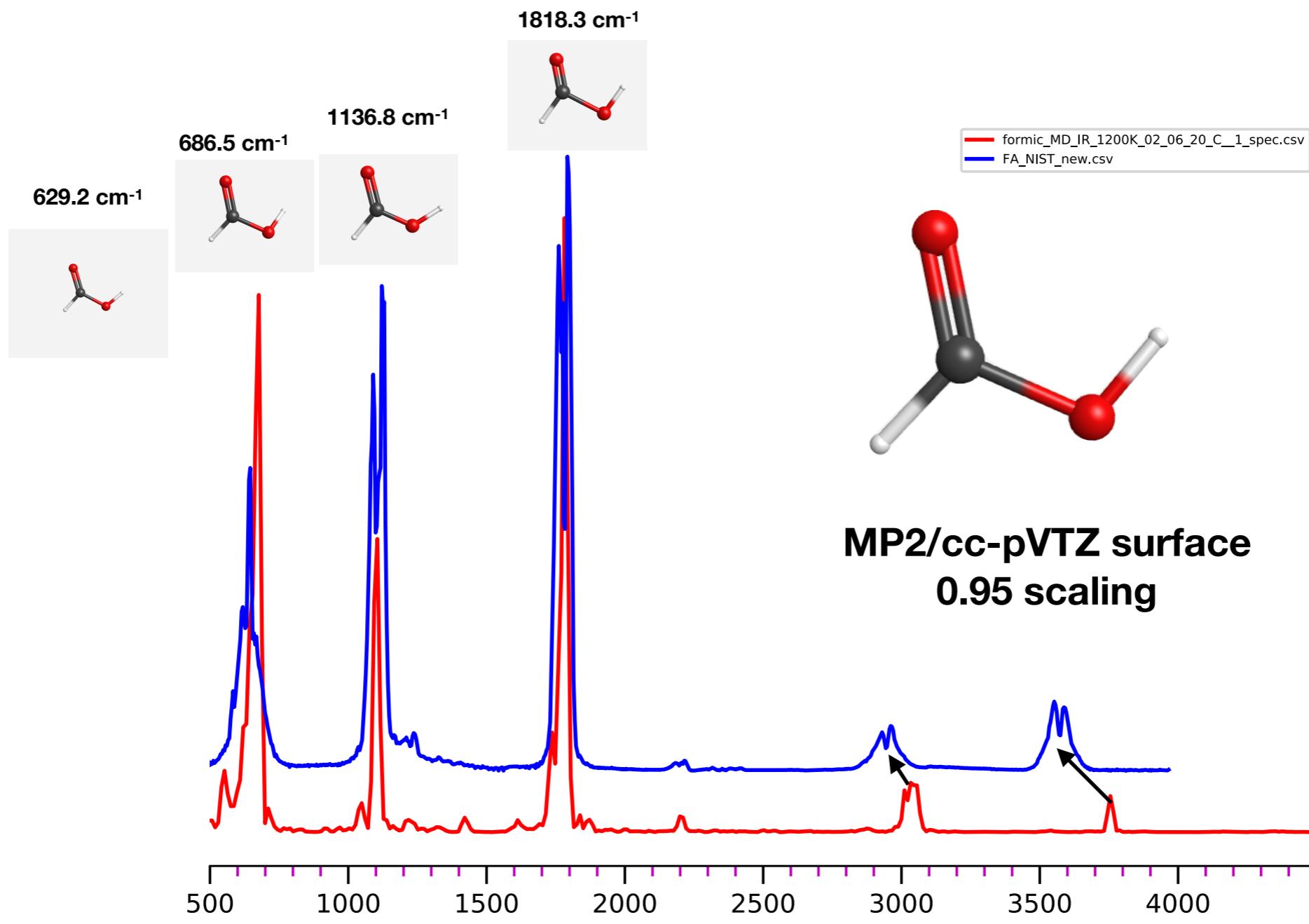


MP2/cc-pVTZ frequencies

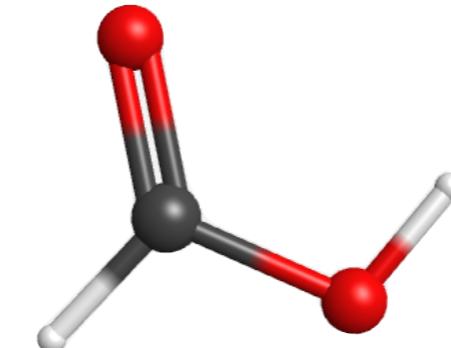
529.1 cm⁻¹ 1130.4 cm⁻¹



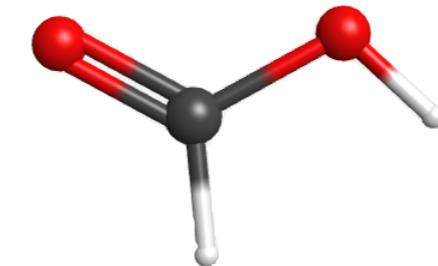
Formic acid



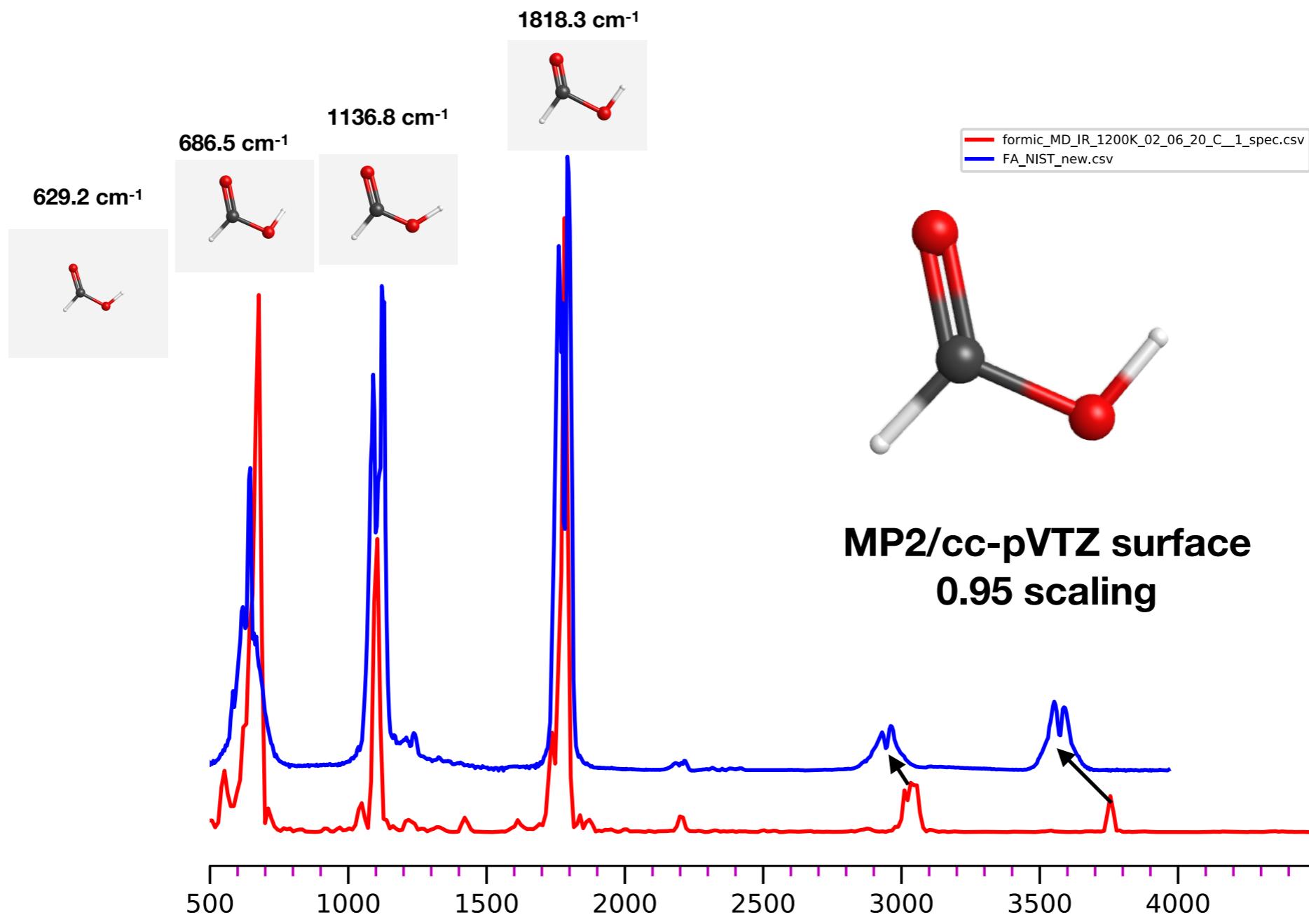
MP2/cc-pVTZ frequencies



**MP2/cc-pVTZ surface
0.95 scaling**

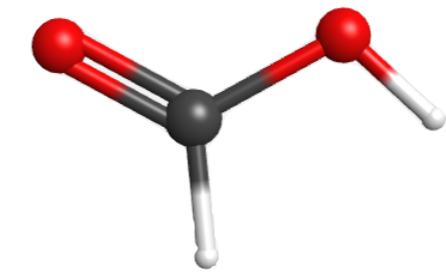


Formic acid

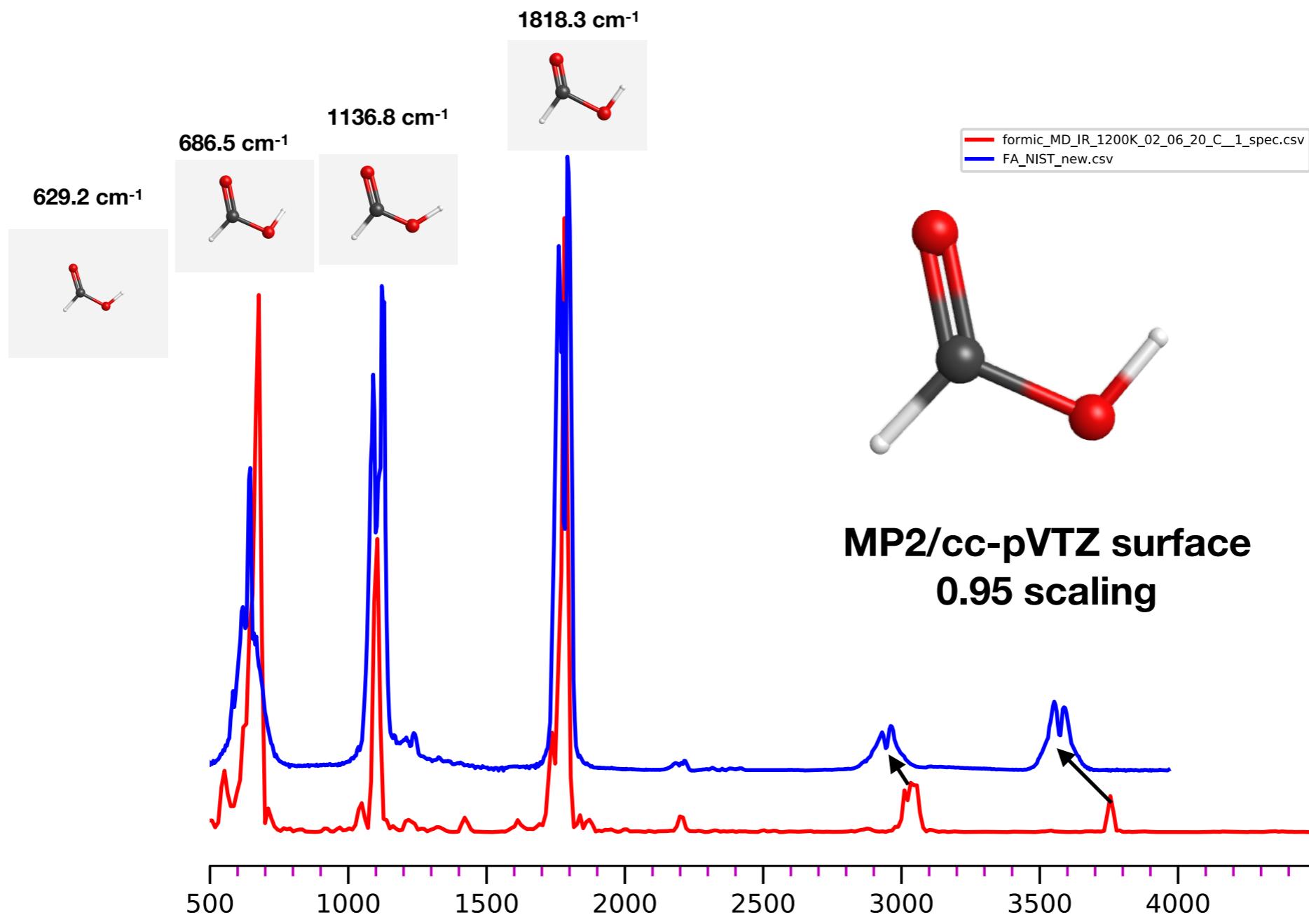


MP2/cc-pVTZ frequencies

529.1 cm⁻¹ 1130.4 cm⁻¹

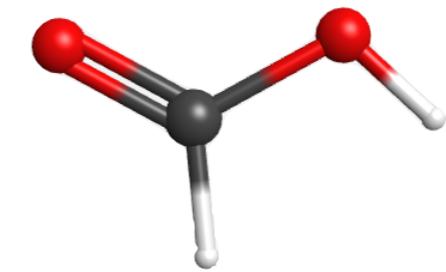


Formic acid

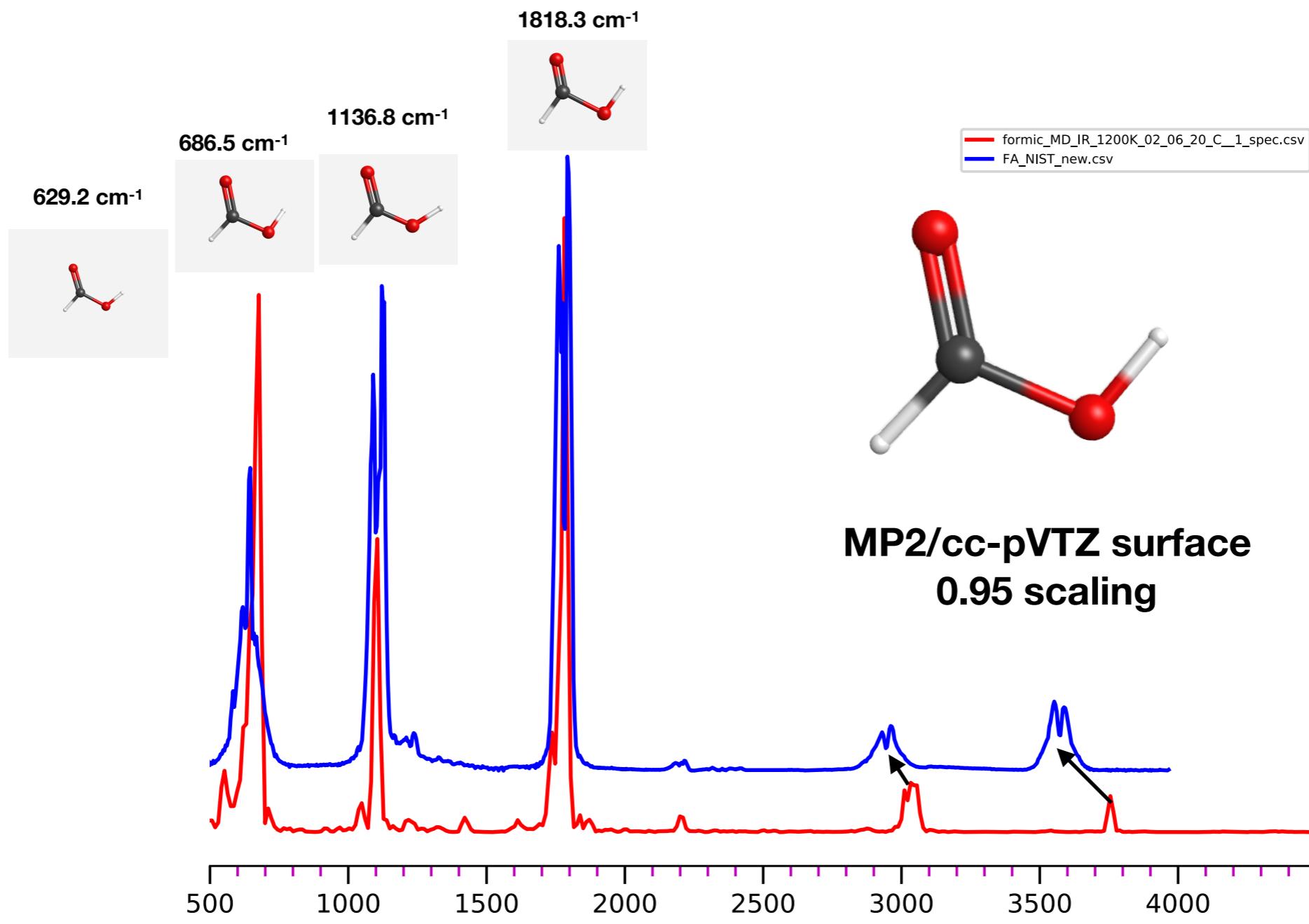


MP2/cc-pVTZ frequencies

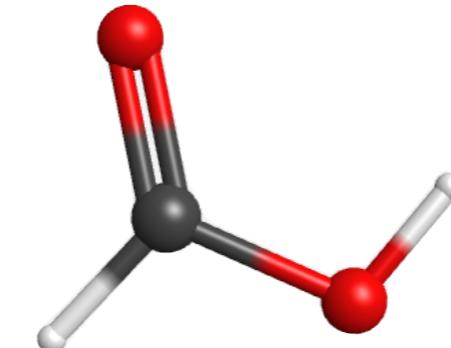
529.1 cm⁻¹ 1130.4 cm⁻¹



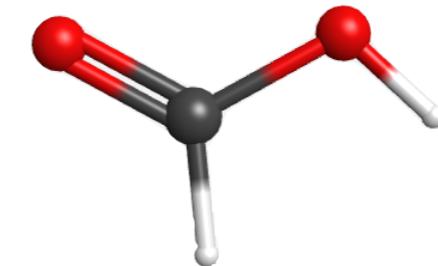
Formic acid



MP2/cc-pVTZ frequencies

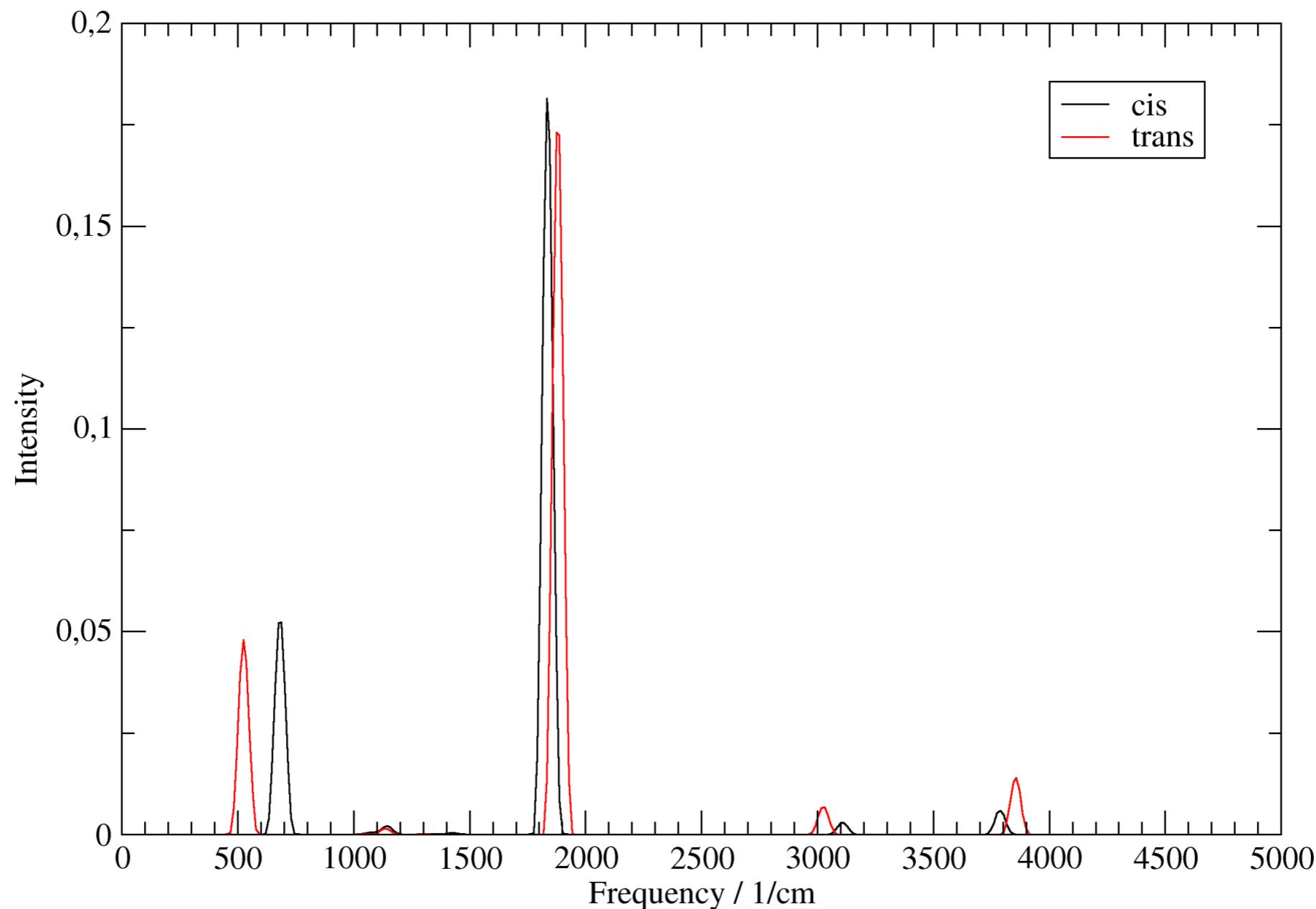


**MP2/cc-pVTZ surface
0.95 scaling**



Coupled Cluster MD

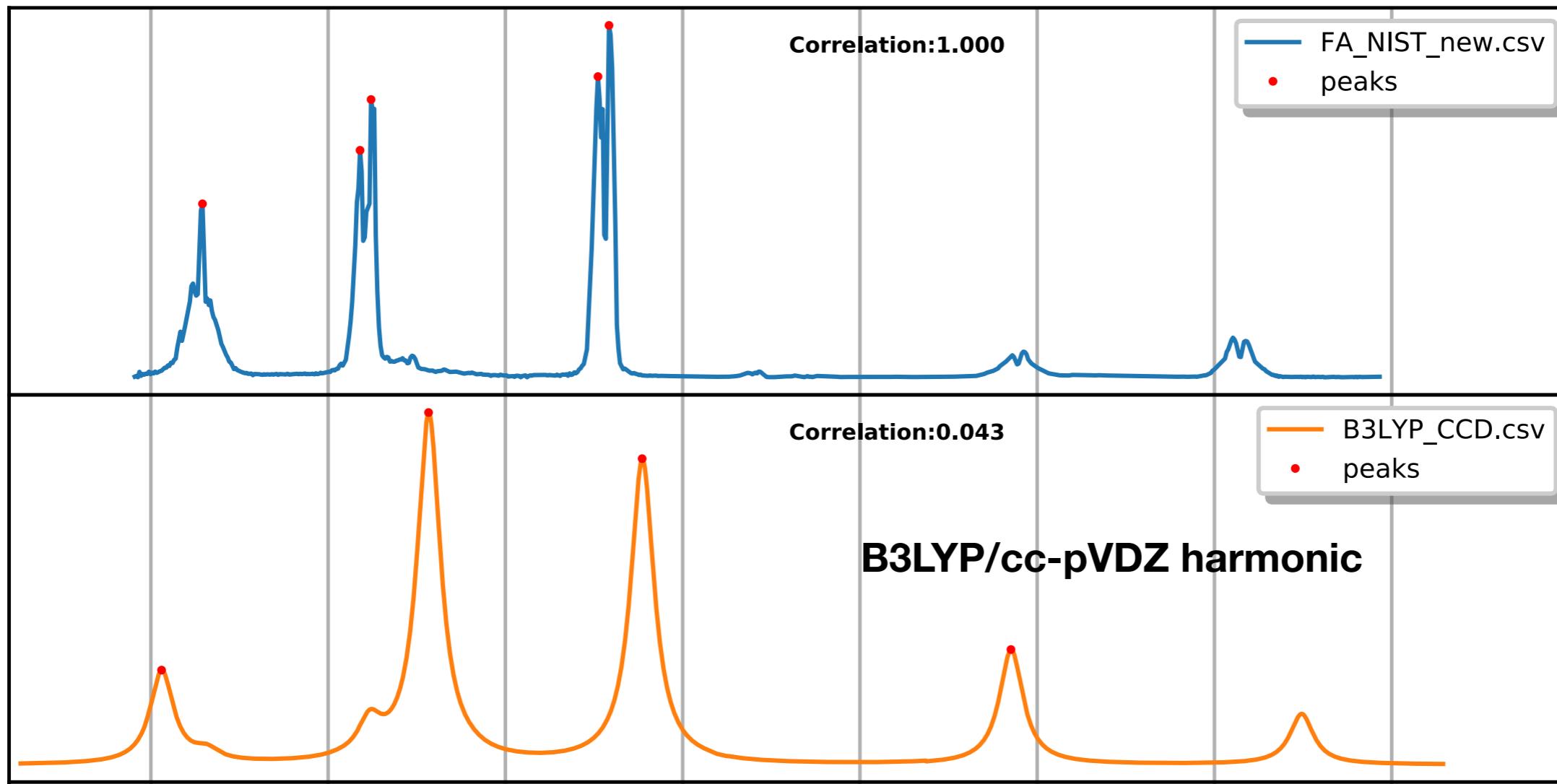
Formic Acid



with Jürgen Gauss and Till Kirsch

Formic Acid

Harmonic Frequency Comparison



Formic Acid

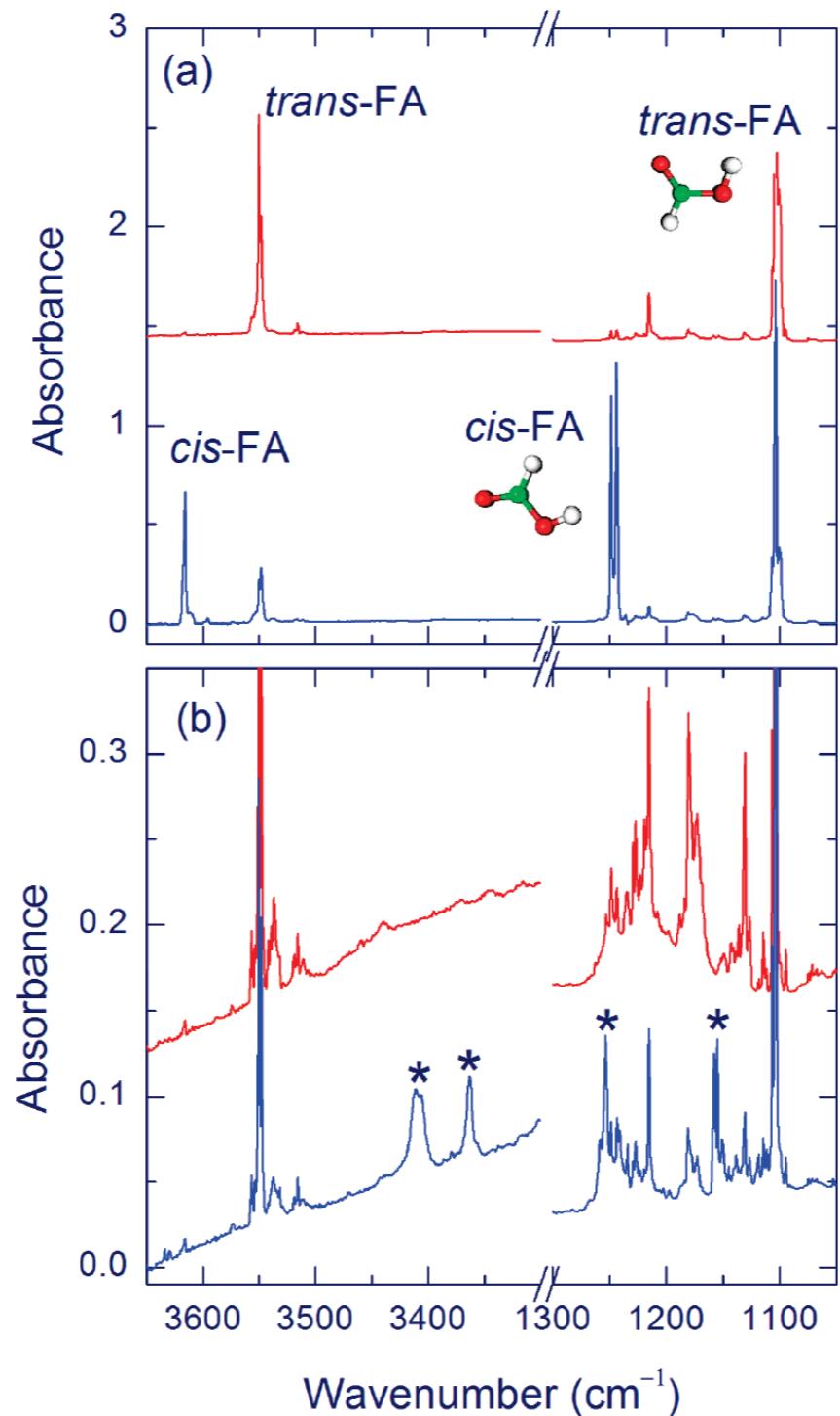
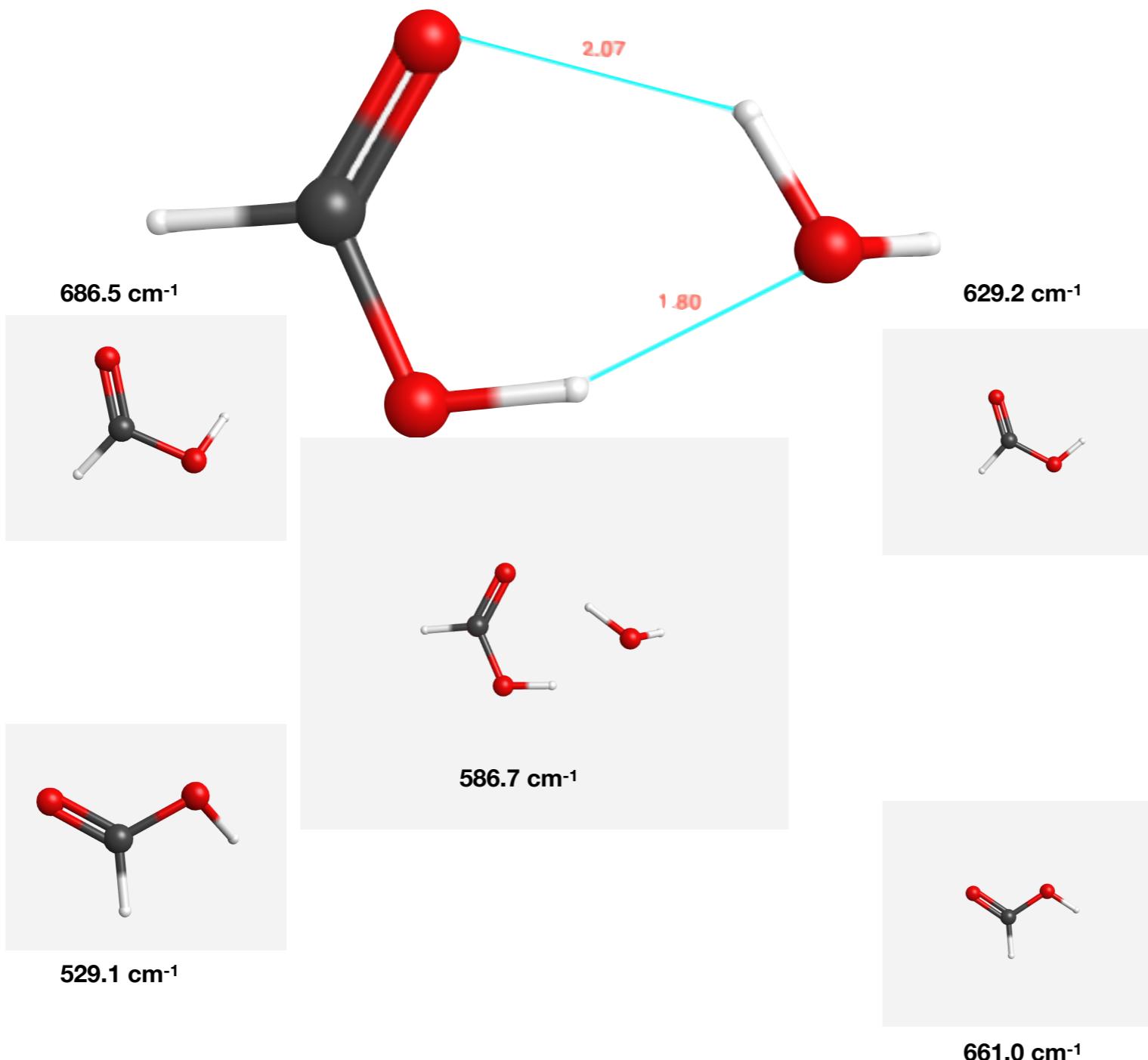


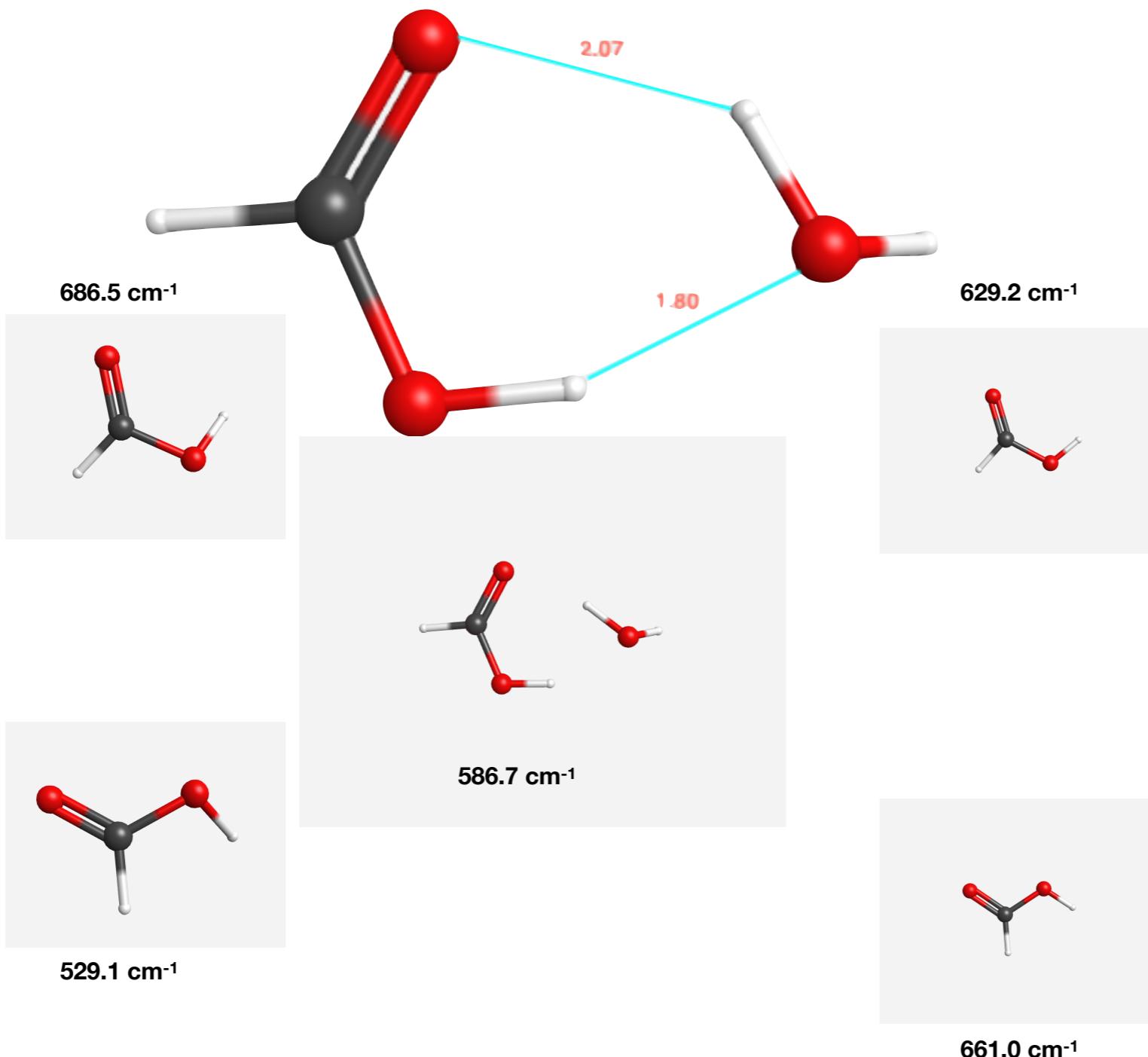
Figure 1. FTIR spectra of FA structures in an argon matrix (from top to bottom). (a) Spectra after deposition and shortly after excitation of *trans*-FA at 3550.3 cm^{-1} with the CW OPO. The structures of the FA conformers are shown following ref 1. (b) Spectra after annealing at 34 K of matrices containing mainly *trans*-FA and after enrichment with *cis*-FA. The new dimeric bands associated with the presence of *cis*-FA are marked with asterisks. The spectra were measured at 8.5 K.

Formic Acid + Water



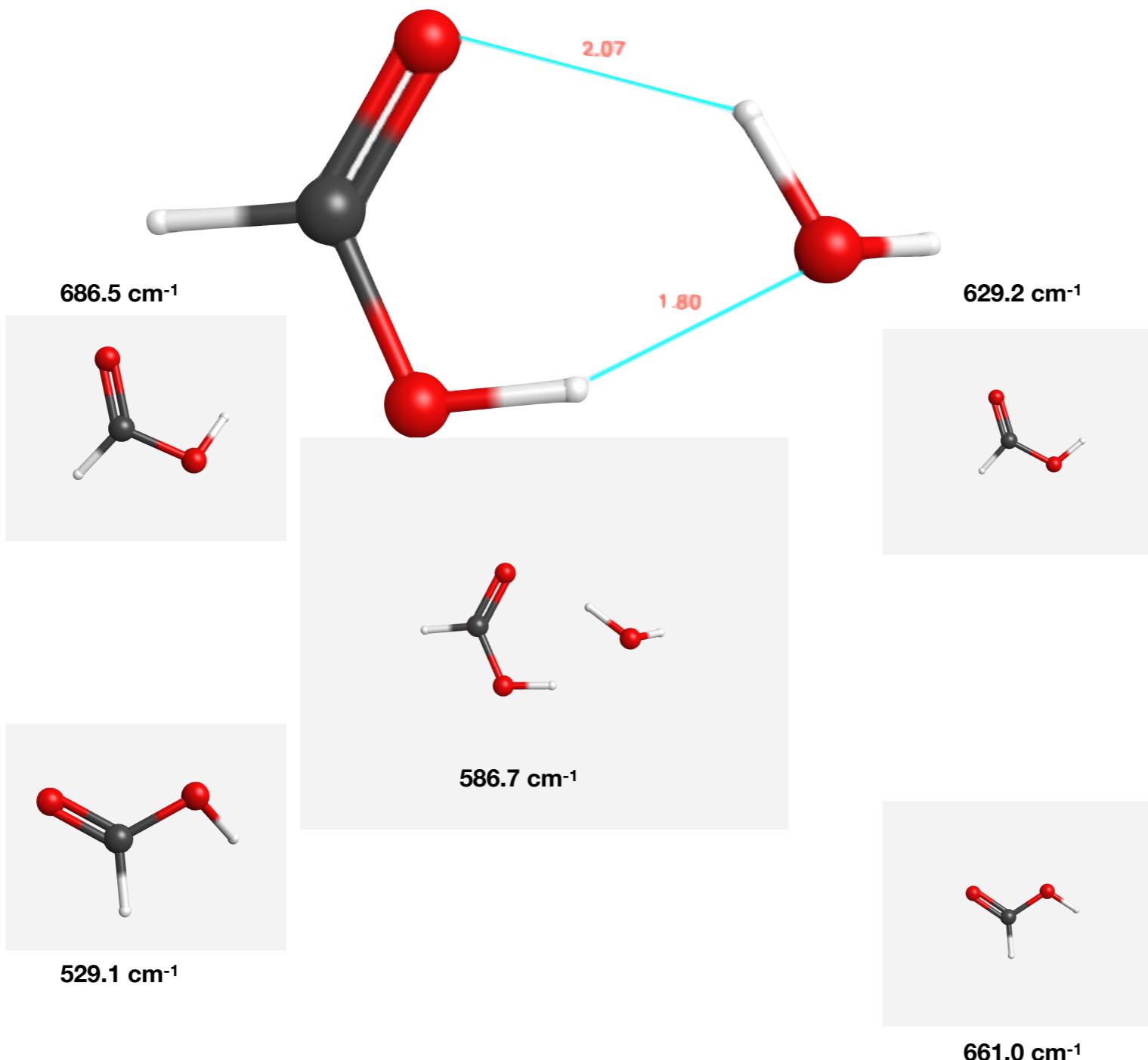
MP2/cc-pVTZ frequencies

Formic Acid + Water



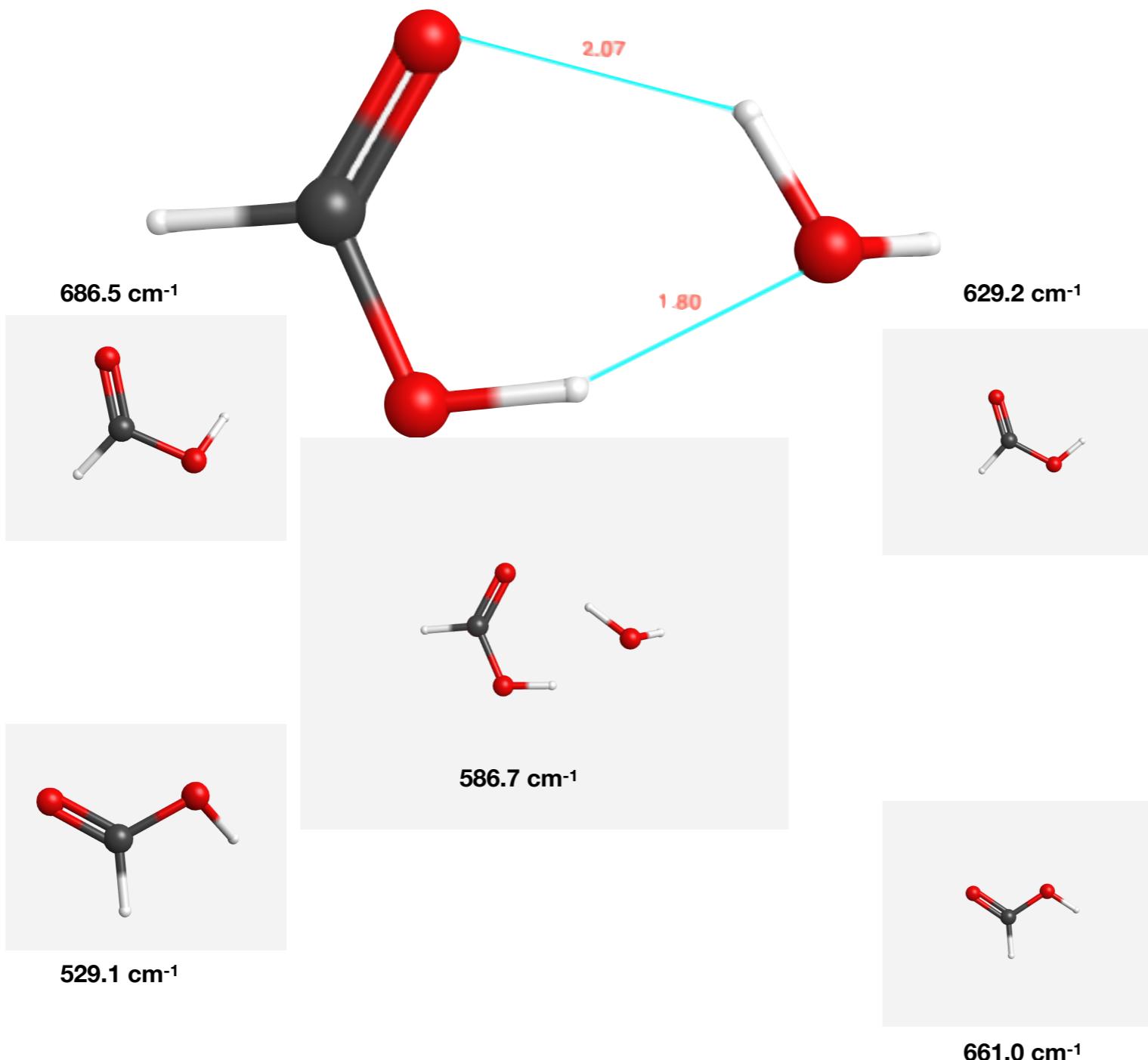
MP2/cc-pVTZ frequencies

Formic Acid + Water



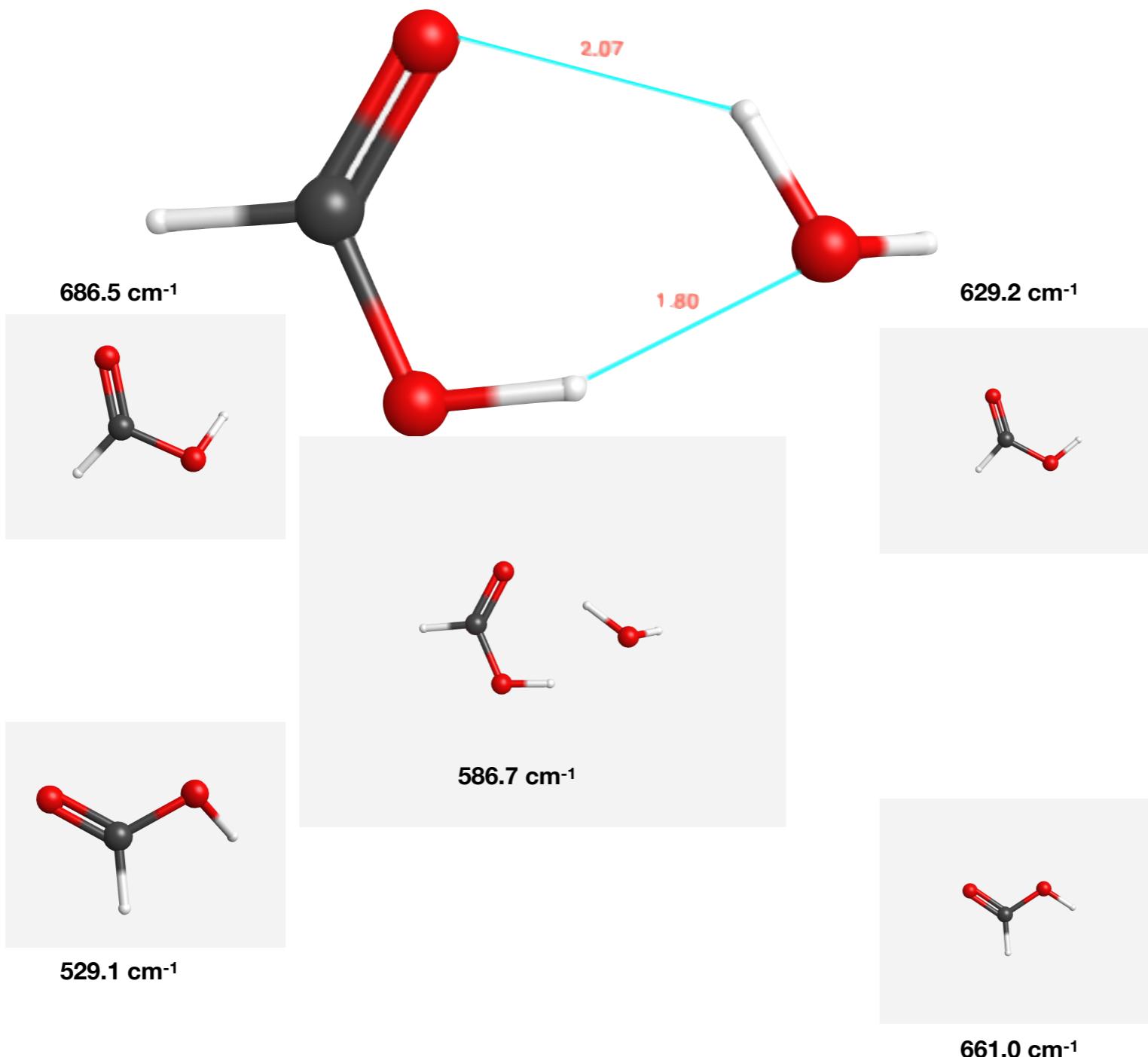
MP2/cc-pVTZ frequencies

Formic Acid + Water



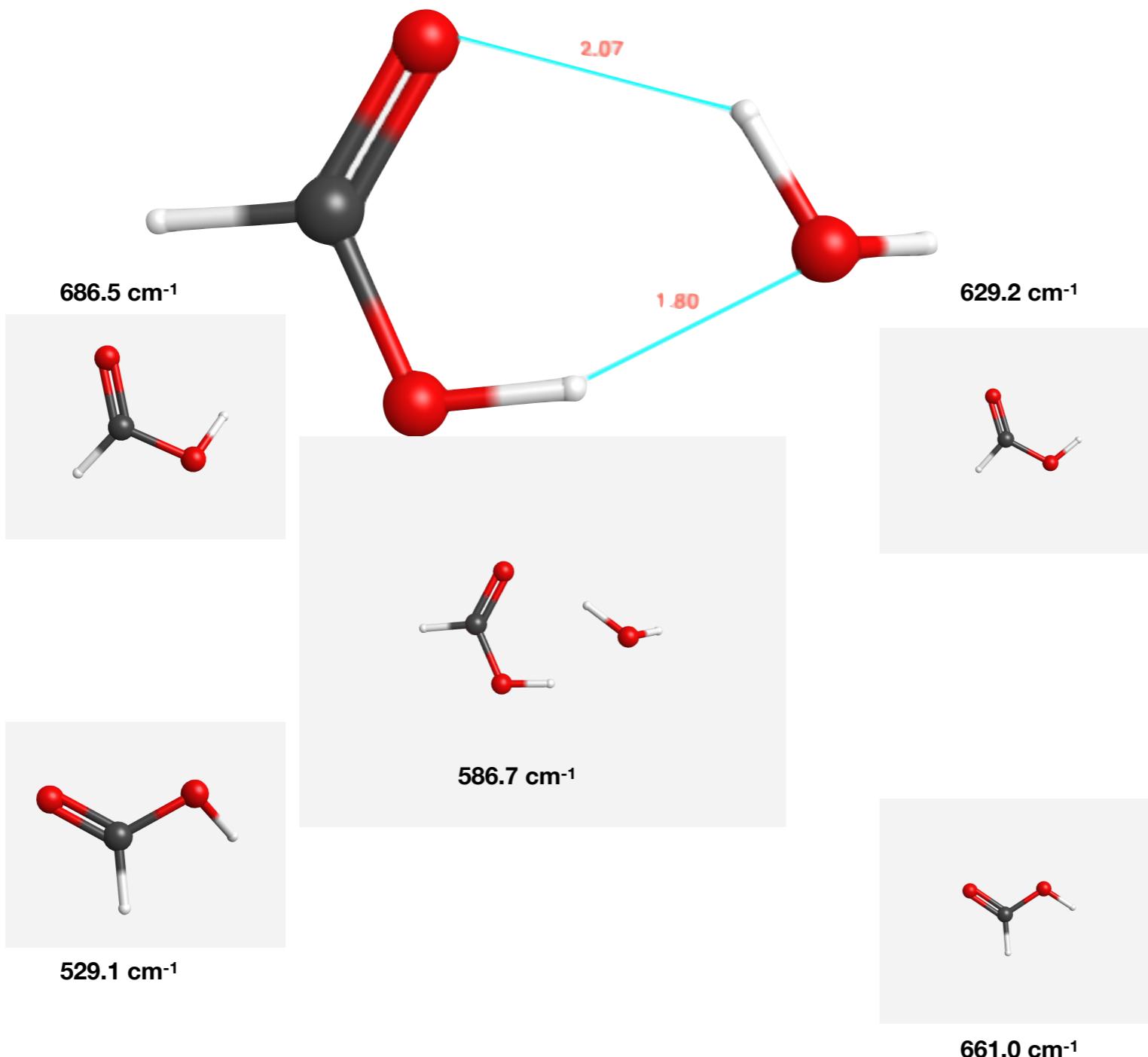
MP2/cc-pVTZ frequencies

Formic Acid + Water



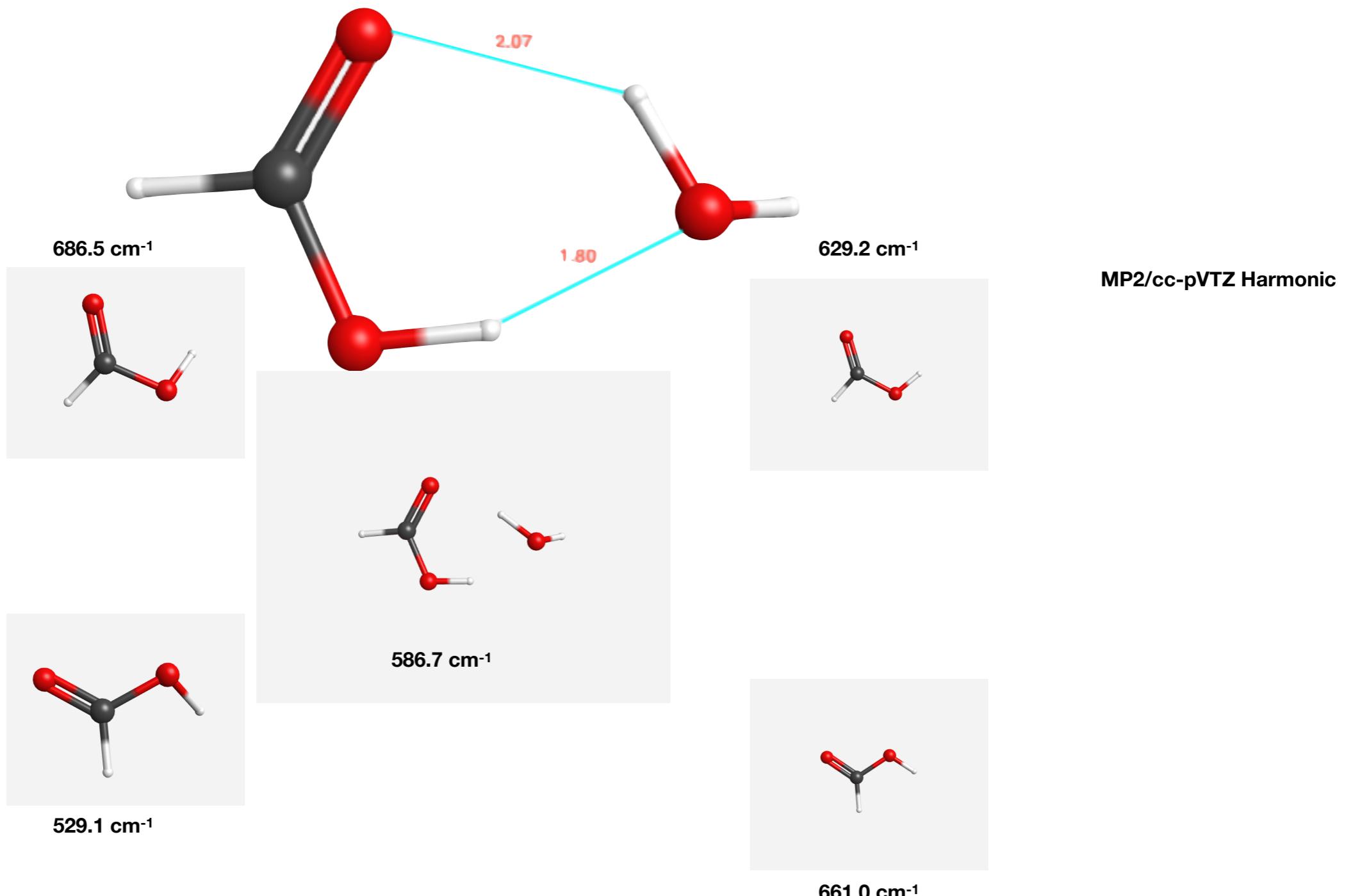
MP2/cc-pVTZ frequencies

Formic Acid + Water

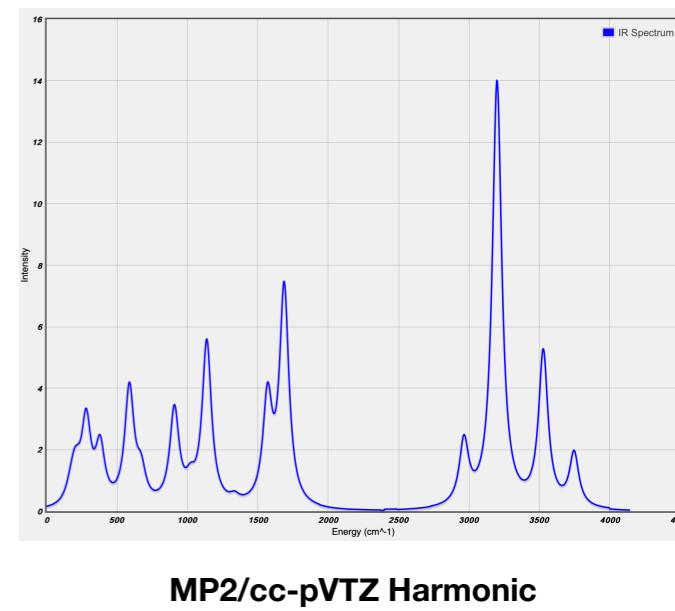
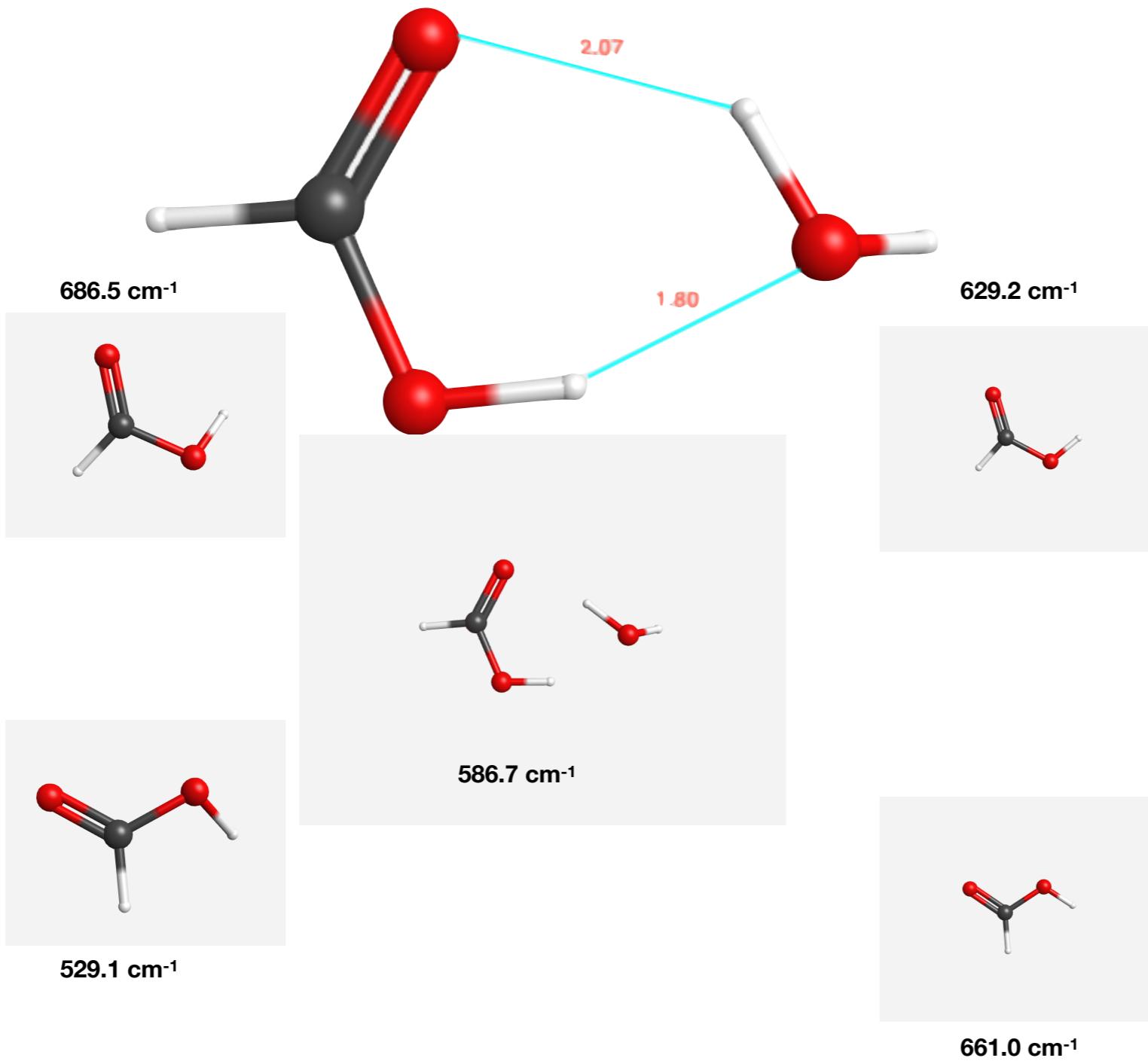


MP2/cc-pVTZ frequencies

Formic Acid + Water



Formic Acid + Water



MP2/cc-pVTZ frequencies

Timing

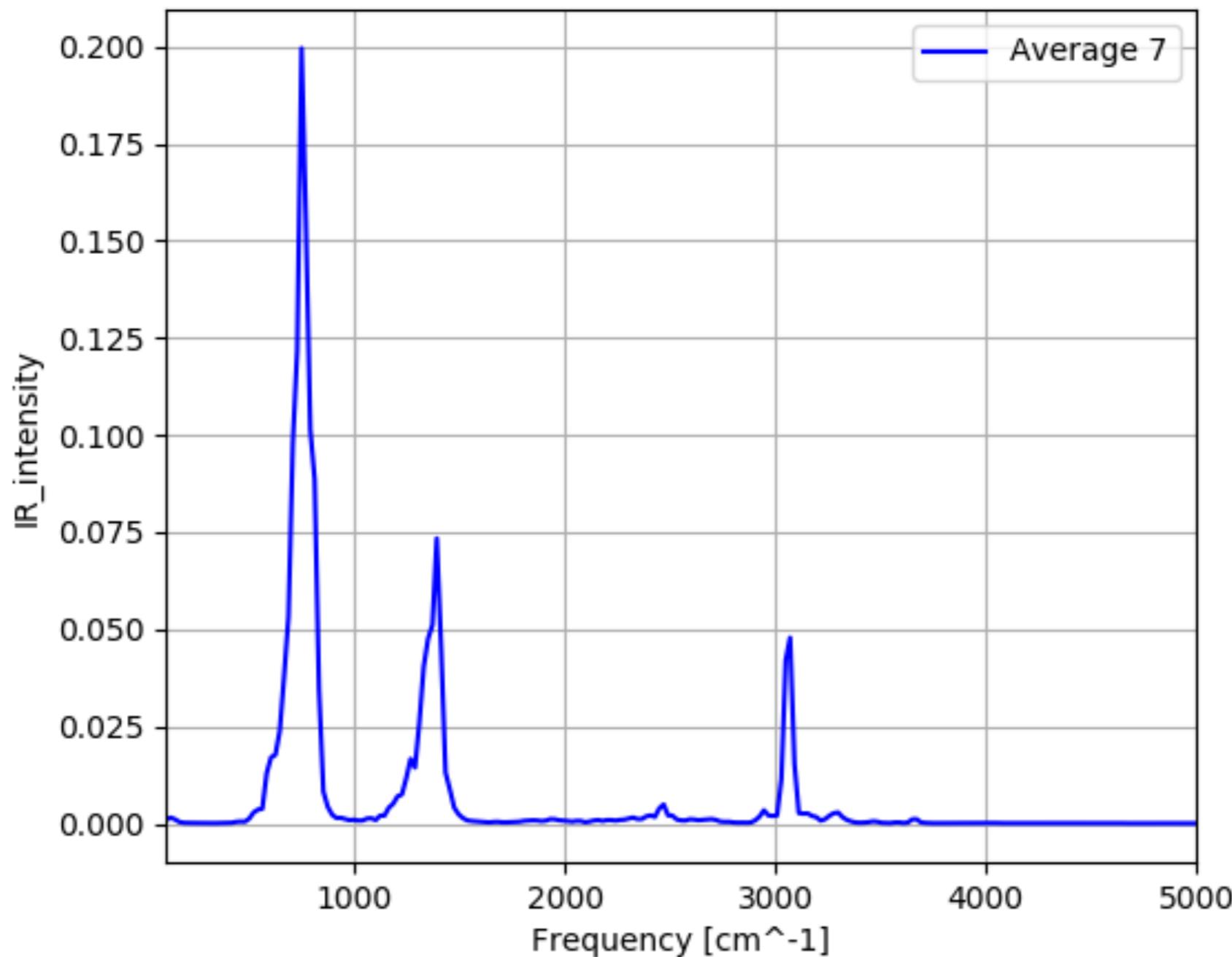
Process	Target	timing	
ML Train 2000 points	Formic Acid	51 minutes	28 core HPC node
PIMD i-PI 16 beads	Formic Acid	0.040 sec/step	4 core i7
MD on ML PES	Formic Acid	0.0012 sec/step	4 core i7
CCSD(T)/cc-pVTZ SP+gradient	Formic Acid	220 sec	28 core HPC node

Accuracy

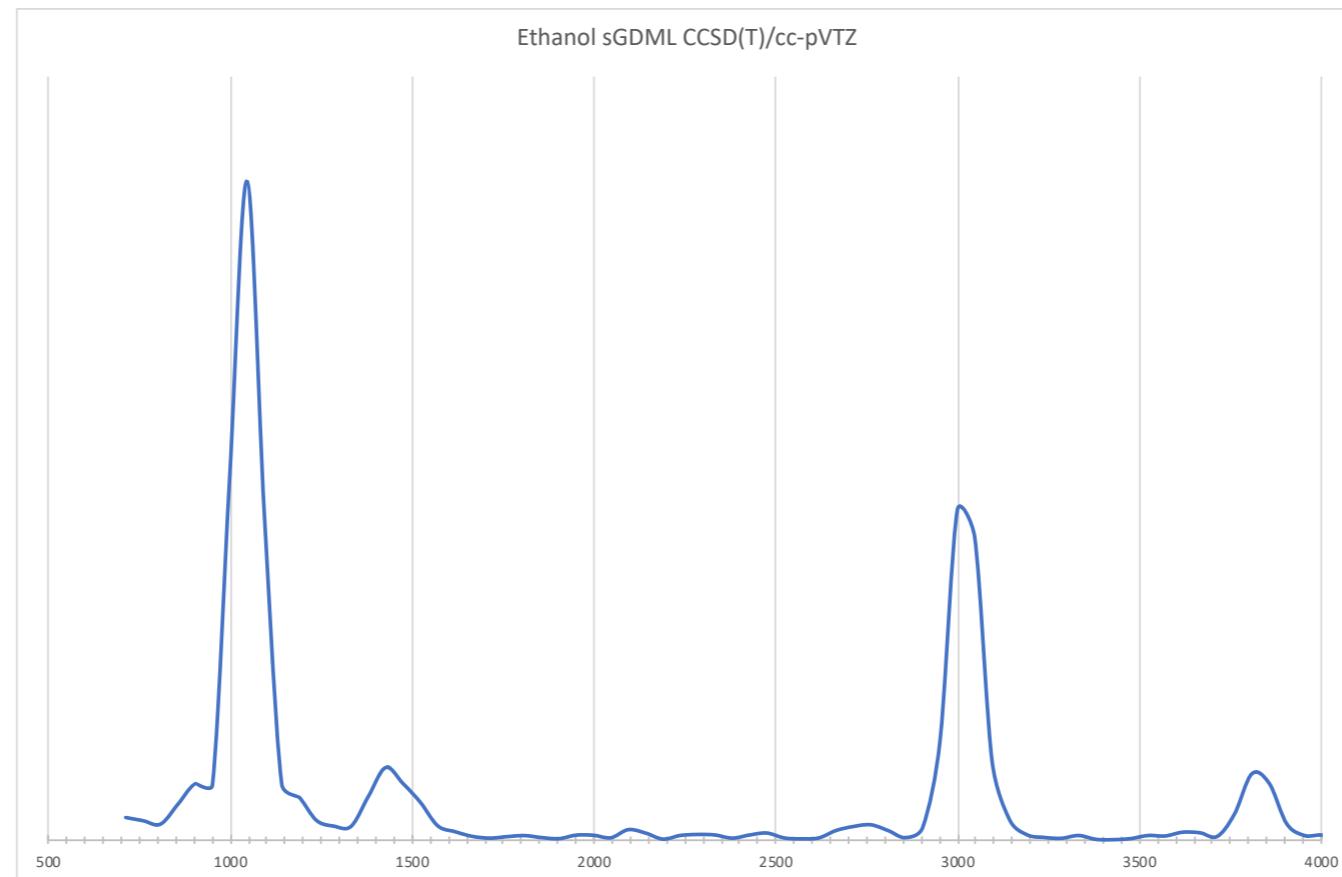
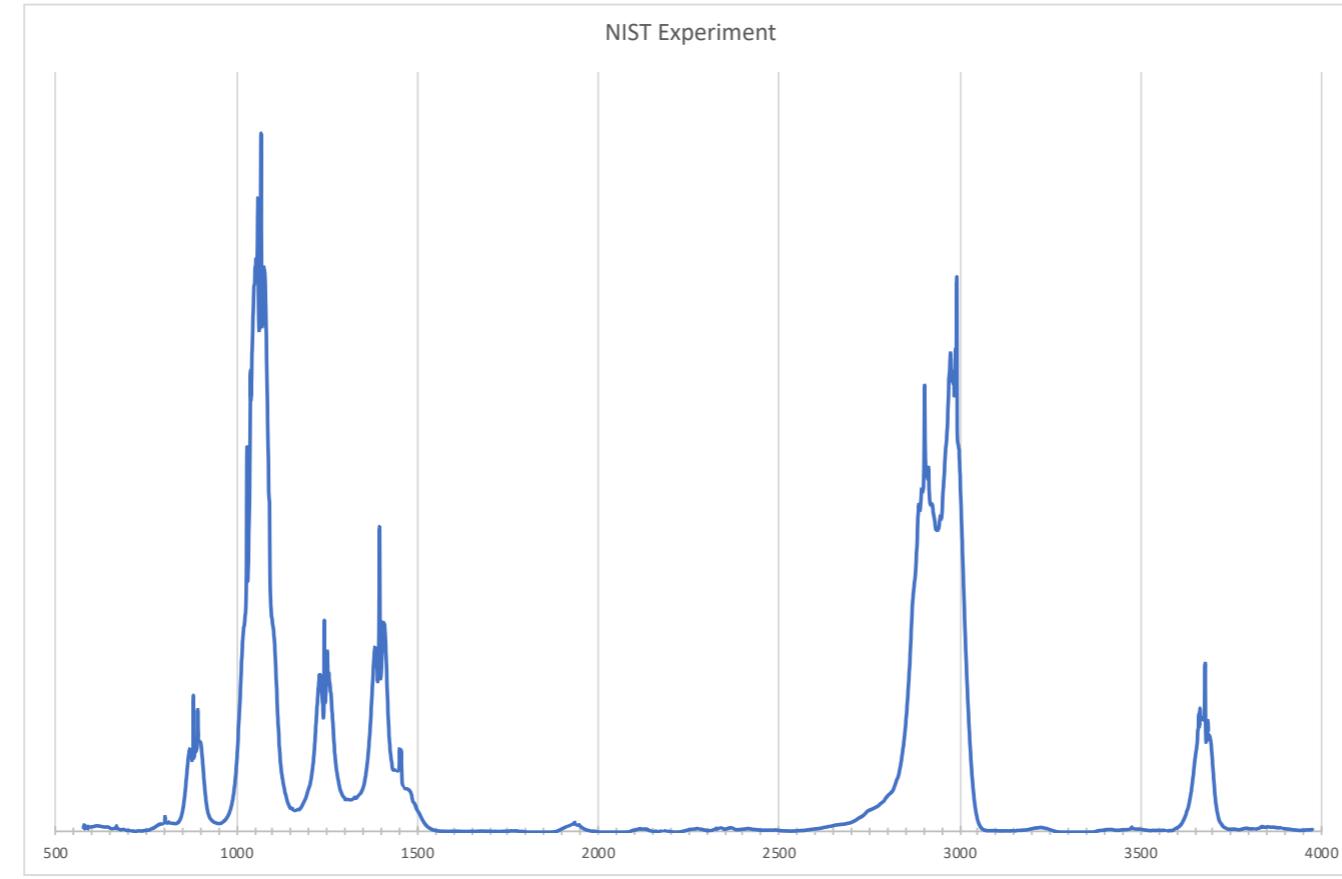
Process	Target	Energy [kcal/mole]	Forces	
sGML MLPES	Formic Acid + water	0.090/0.137 (MAE, RMSE)	0.375/0.583 (MAE, RMSE)	
sGML MLPES	Formic Acid	0.021/0.054 (MAE, RMSE)	0.225/0.571 (MAE, RMSE)	742.3 geo/s

Acetylene

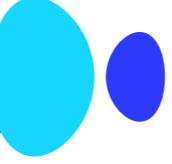
ML-PES
2000 cm⁻¹ energy



Machine learning PES: Ethanol IR



Acknowledgements

- Drew Behrendt (Temple '20)
- Michael Wilhelm (Temple University)
- Hai-Lung Dai (Temple University)
- Robert Field (MIT)
- Hylleraas Centre for Quantum Molecular Science  **Hylleraas**
 - NRC NOTUR NN4654K
 - DOE 02-86ER 134584
 - Temple University's HPC resources and thus was supported in part by NSF MRI-1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189.
 - Midwest Undergraduate Computational Chemistry Consortium: NSF-MRI 1039925

Extra Slides

Molecular Dynamics on Machine Learning Optimized Potential Energy Surfaces

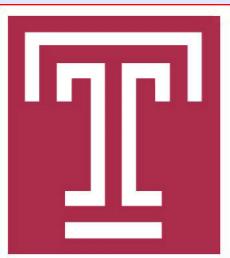
Jonathan Smith^{1,2}

¹Temple University Department of Chemistry

²Hylleraas Centre for Quantum Molecular Sciences, University of Oslo

Overview

Machine learning provides a new analytical tool to interpret and extrapolate computed quantum mechanical properties. In one application (I) vibrational frequencies of a library of molecules are estimated using density functional theory with a modest model chemistry and these estimates are improved by using a neural network with molecular properties as input and training on experimental frequencies. In a second application (II), molecular potential energy surfaces are interpolated from high level *ab initio* calculations using machine learning to provide a surface for molecular dynamics in which forces and gradients can be determined 4 orders of magnitude faster than high level *ab initio*. These surfaces are robust to using molecular dynamics to simulate anharmonic vibrational spectra.



Acknowledgements

Hylleraas HPC (Notur NRC NOTUR NN4654K)

Temple HPC (NSF 1625061)

Midwest Undergraduate Computational Chemistry Consortium

I. Vibrational frequency prediction neural network

A straightforward application of machine learning is presented which shows improved prediction of vibrational frequencies following training a neural network. The steps in constructing this neural network demonstrate the common tasks associated with applying machine learning to prediction.

1. Construction of a dataset to train the model. In this case experimental vibrational frequencies of hundreds of molecules are collated along with corresponding vibrational frequencies computed using the B3LYP/6-31G(d) model chemistry.
2. A neural network is designed and coded using Python with the PyTorch (<https://pytorch.org>) module. In this design an input layer consisting of 11 inputs is connected to three hidden layers feeding into a single output reflecting the vibrational frequency. This design is reflected in Figure 1. Each input is connected to each node of the first hidden layer using a weight between 0-1. Each node between hidden layers is activated based on what is known as a RELU activation function which is the max(0,x).
3. The network is trained on experimental vibrational frequencies and the connections between the nodes reflected in weights are optimized through a technique known as back propagation using a loss function. In this case a smoothL1loss is used. This process is repeated through several epochs until a tolerance in the loss function is achieved.
4. Following training, the model is validated on the quality of prediction of experimental data typically as a function of parameters (hyper parameters).
5. The model is tested on a second set of different data to estimate the accuracy and error bars. This is shown in Figure 2.

This neural network achieves a mean squared error of about an order of magnitude better than standard scaling of computed vibrational frequencies (0.96 in this case, see Radom [1]).

References

- 1) Merrick, Jeffrey P.; Moran, Damian; Radom, Leo, "An Evaluation of Harmonic Vibrational Frequency Scale Factors". *J. Phys. Chem. A* 2007, 111 (Copyright (C) 2010 American Chemical Society (ACS). All Rights Reserved.), 11683-11700.
- 2) Smith, Jonathan M.; Nikow, Matthew; Ma, Jianqiang; Wilhelm, Michael J.; Han, Yong-Chang; Sharma, Amit R.; Bowman, Joel M.; Dai, Hai-Lung, "Chemical Activation through Super Energy Transfer Collisions". *Journal of the American Chemical Society* 2014, 136 (5), 1682-1685.
- 3) Kapil, Venkat; Rossi, Mariana, *et al.* "i-PI 2.0: A universal force engine for advanced molecular simulations". *Computer Physics Communications* 2019, 236, 214-223.
- 4) Chmiela, Stefan; Sauceda, Huziel E., *et al.*, "sGMDL: Constructing accurate and data efficient molecular force fields using machine learning". *Computer Physics Communications* 2019, 240, 38-45.

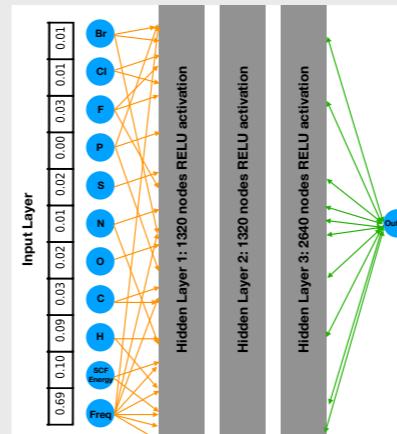


Figure 1. Schematic diagram of neural network for vibrational frequency prediction. Orange and green arrows are representative of strongly weighted connections. Table to left of inputs reflect the square of the sum of coefficients connected to that node representing the importance of each descriptor in predicting frequency.

II. Machine learning potential energy surfaces

Ab initio molecular dynamics is used to study chemical reaction dynamics [2] and predict other properties including vibrational spectra of energized or floppy molecules. Accurate model chemistries can require evaluation of energy and gradient for each step of a trajectory amounting to ~10 cpu sec/step with >10⁶ steps for a single trajectory at DFT accuracy for modest sized systems of 10-20 atoms. To achieve coupled cluster (CCSD(T)/cc-pVTZ) accuracy trajectories quickly become costly.

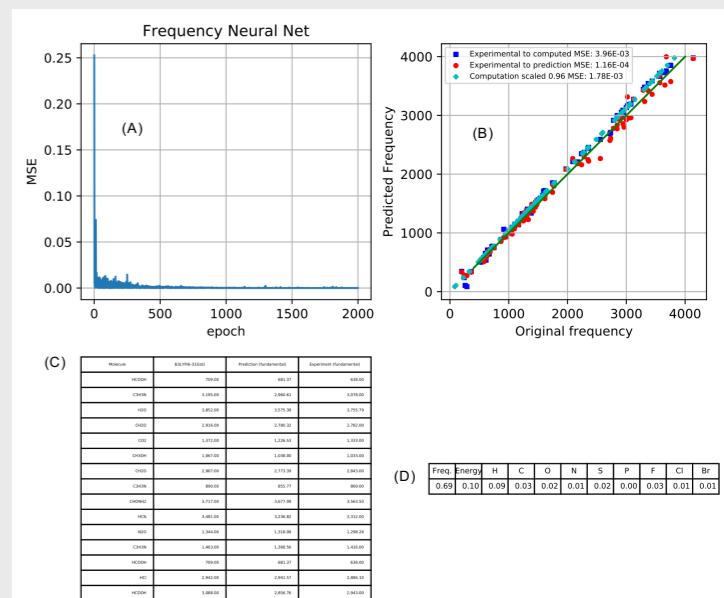


Figure 2. Selected results following training of neural network diagrammed in Figure 1. A) Evolution of mean squared error as a function of iterative refinement of coefficients. B) Plot of predicted frequencies versus experimental frequencies for B3LYP/6-31G(d) (blue), neural network prediction (red), and scaled, 0.96, B3LYP/6-31G(d) (cyan). C) Select frequency comparisons. D) Normalized coefficients on inputs including B3LYP/6-31G(d) harmonic frequency, SCF energy, # of H, C, O, N, S, P, F, Cl, Br in molecule.

Here we demonstrate the use of accurate potential energy and gradient points to inform construction of a machine learning potential energy surface upon which we run molecular dynamics at millisecond cpu sec/step at high accuracy. The scheme is outlined in Figure 3. The calculations are carried out using i-PI [3] for path integral molecular dynamics. Infrared spectra are simulated by using the Fourier transform of the dipole autocorrelation function and averaged over several trajectories. Machine learning potential energy surface is refined with the sGMDL package [4]. Figure 4 shows results for formic acid.

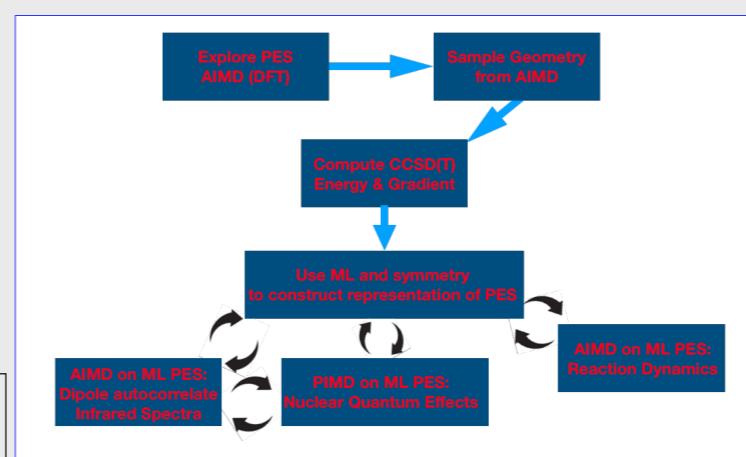


Figure 3. Schematic diagram showing steps of machine learning optimized potential energy surface creation and application

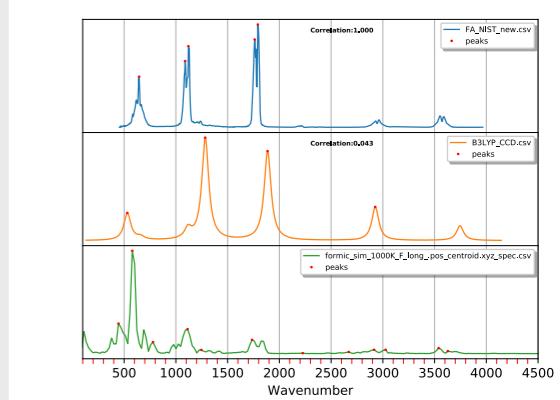
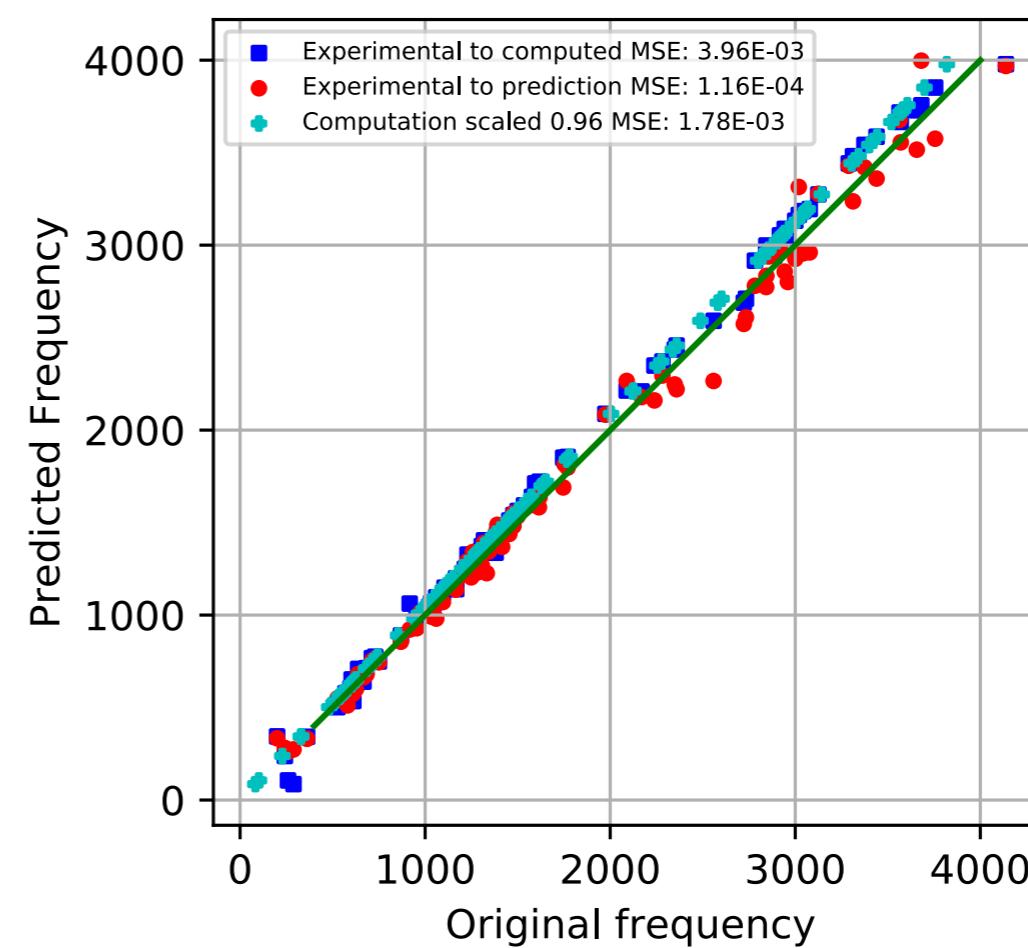
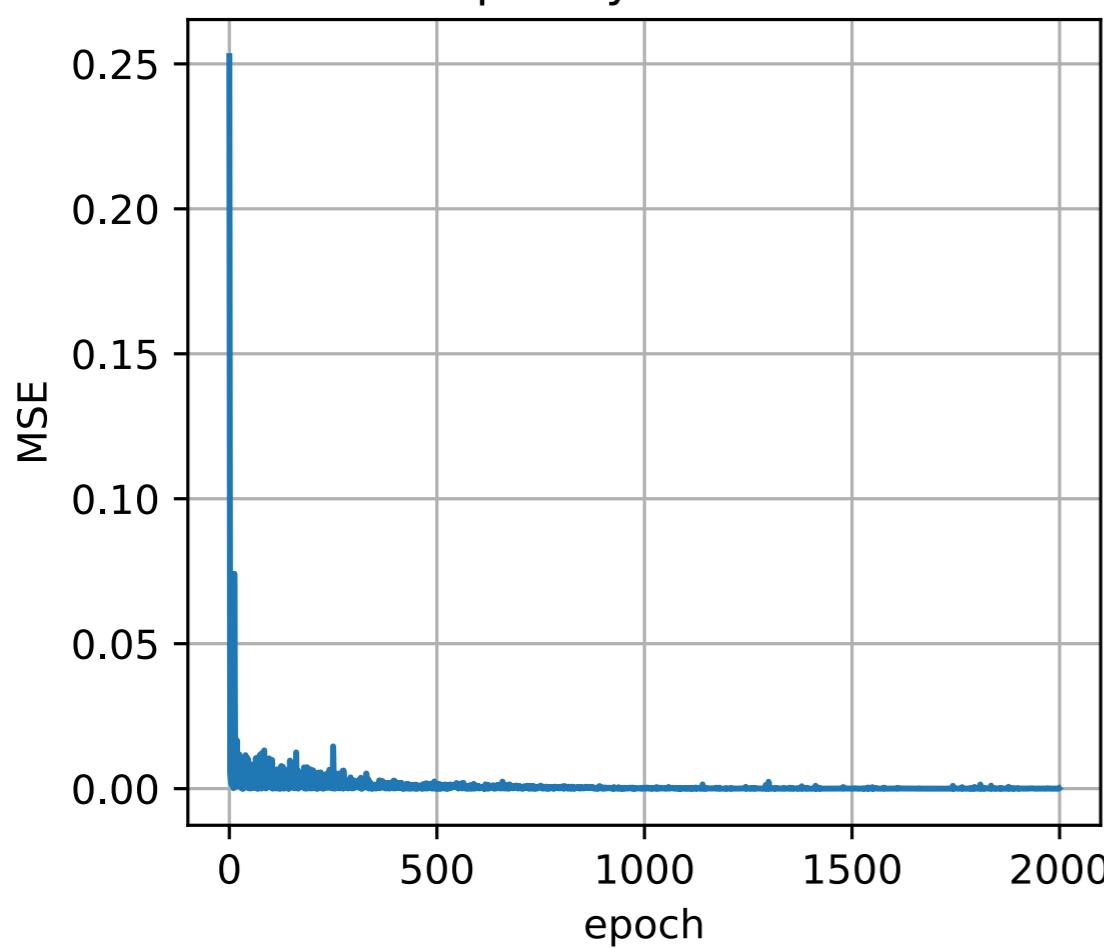


Figure 4. Formic acid gas phase experimental spectra, B3LYP/cc-pVDZ harmonic and PIMD (24 beads, 9.0 psec NVE, 0.1 fs/sec/step) on machine learning PES simulated infrared spectrum.

Machine Learning

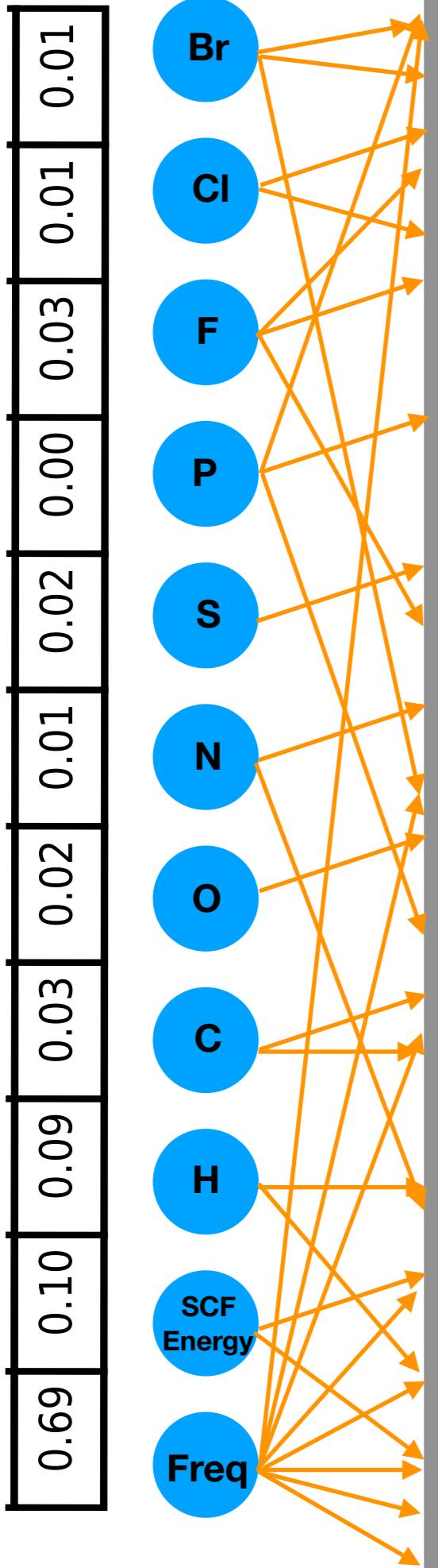
Frequency Neural Net



Molecule	B3LYP/6-31G(d)	Prediction (fundamental)	Experiment (fundamental)
HCOOH	709.00	681.37	638.00
C3H3N	3,195.00	2,960.61	3,078.00
H2O	3,852.00	3,575.38	3,755.79
CH2O	2,916.00	2,780.32	2,782.00
CO2	1,372.00	1,226.53	1,333.00
CH3OH	1,067.00	1,038.00	1,033.00
CH2O	2,967.00	2,773.39	2,843.00
C3H3N	890.00	855.77	869.00
CHONH2	3,717.00	3,677.08	3,563.50
HCN	3,481.00	3,236.82	3,312.00
N2O	1,344.00	1,318.08	1,298.28
C3H3N	1,463.00	1,368.56	1,416.00
HCOOH	709.00	681.37	638.00
HCl	2,942.00	2,941.57	2,886.10
HCOOH	3,088.00	2,856.76	2,943.00

Freq.	Energy	H	C	O	N	S	P	F	Cl	Br
0.69	0.10	0.09	0.03	0.02	0.01	0.02	0.00	0.03	0.01	0.01

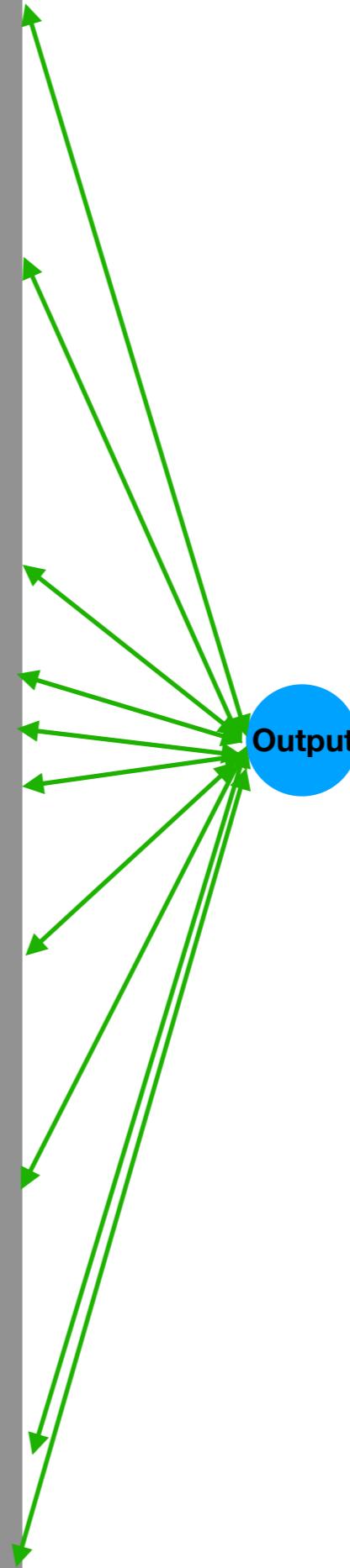
Input Layer



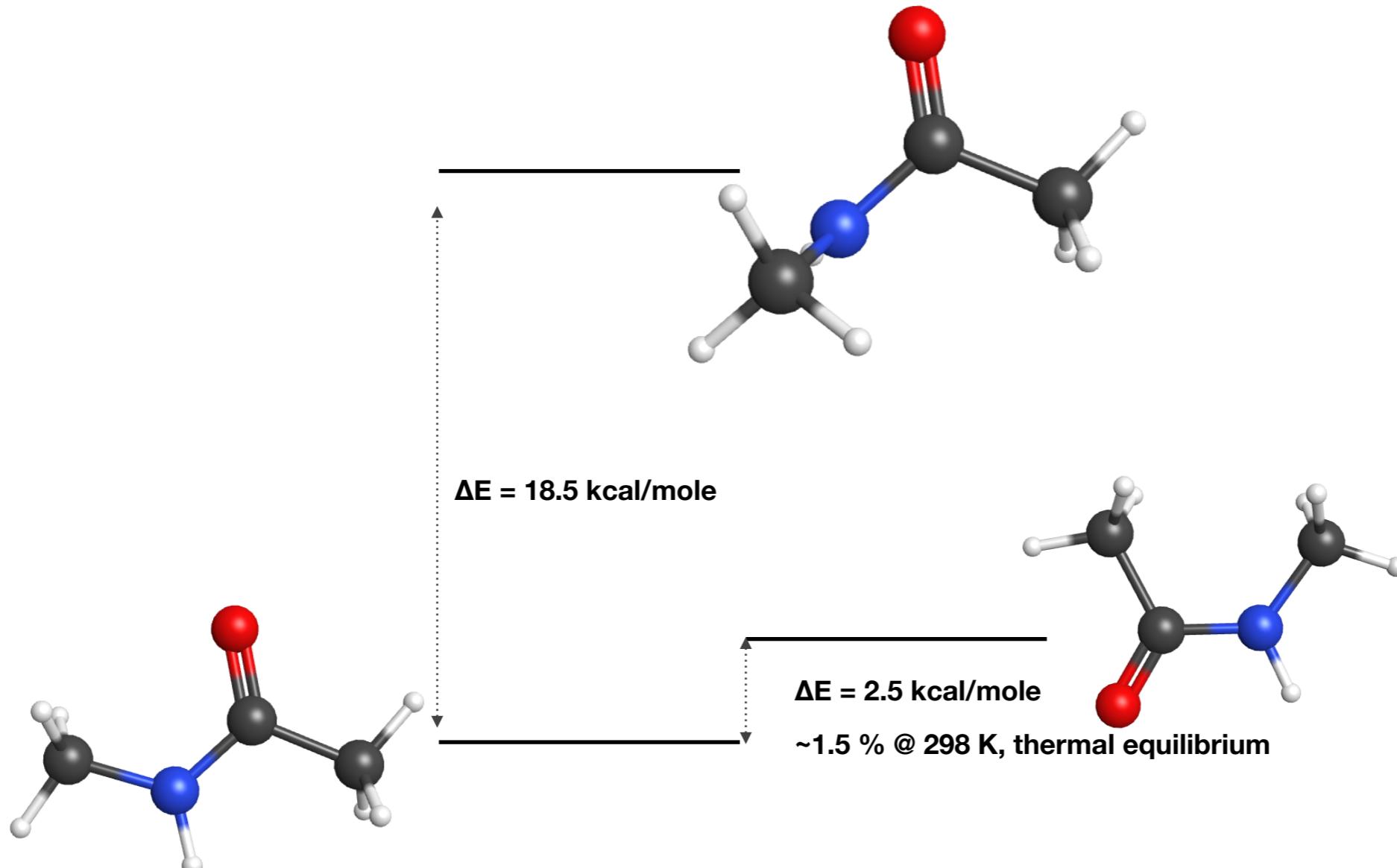
Hidden Layer 1: 1320 nodes RELU activation

Hidden Layer 2: 1320 nodes RELU activation

Hidden Layer 3: 2640 nodes RELU activation



N-methylacetimide



APFD/6-311++G(3df,3pd) + Zero Point
CCSD(T)/cc-pVTZ

Machine Learning: H₂O PES

H₂O CCSD(T)/cc-pVTZ Surface

from Hessian	scaled	Stated frequency	comparison
797.52222926	1615.6	1572.76	97.3
1836.08421487	3719.5	3712.22	99.8
1886.01406182	3820.6	3820.62	100.0
	2.03		

outhess_symm

0.0016	0.0000	0.0000	-0.0036	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.1817	-0.0600	0.0000	-0.0172	0.1869	0.0000	-0.1623	-0.1267
0.0000	-0.0600	1.9406	0.0000	-0.0346	-0.0548	0.0000	0.1002	-1.8829
-0.0036	0.0000	0.0000	0.0037	0.0000	0.0000	0.0018	0.0000	0.0000
0.0000	-0.0172	-0.0346	0.0000	1.7964	-0.4790	0.0000	-1.7788	0.5165
0.0000	0.1869	-0.0548	0.0000	-0.4790	0.3216	0.0000	0.2908	-0.2646
0.0000	0.0000	0.0000	0.0018	0.0000	0.0000	-0.0012	0.0000	0.0000
0.0000	-0.1623	0.1002	0.0000	-1.7788	0.2908	0.0000	1.9416	-0.3924
0.0000	-0.1267	-1.8829	0.0000	0.5165	-0.2646	0.0000	-0.3924	2.1482

Conclusions

Conclusions

- Mapping Spectra onto energetics and dynamics requires computations

Conclusions

- Mapping Spectra onto energetics and dynamics requires computations
 - Static with high order gradients effective for cold gas phase

Conclusions

- Mapping Spectra onto energetics and dynamics requires computations
 - Static with high order gradients effective for cold gas phase
 - Spectra computed from dynamics trajectory promising for energized and floppy molecules

Conclusions

- Mapping Spectra onto energetics and dynamics requires computations
 - Static with high order gradients effective for cold gas phase
 - Spectra computed from dynamics trajectory promising for energized and floppy molecules
- DFTB within CP2K fast, benchmark accuracy

Conclusions

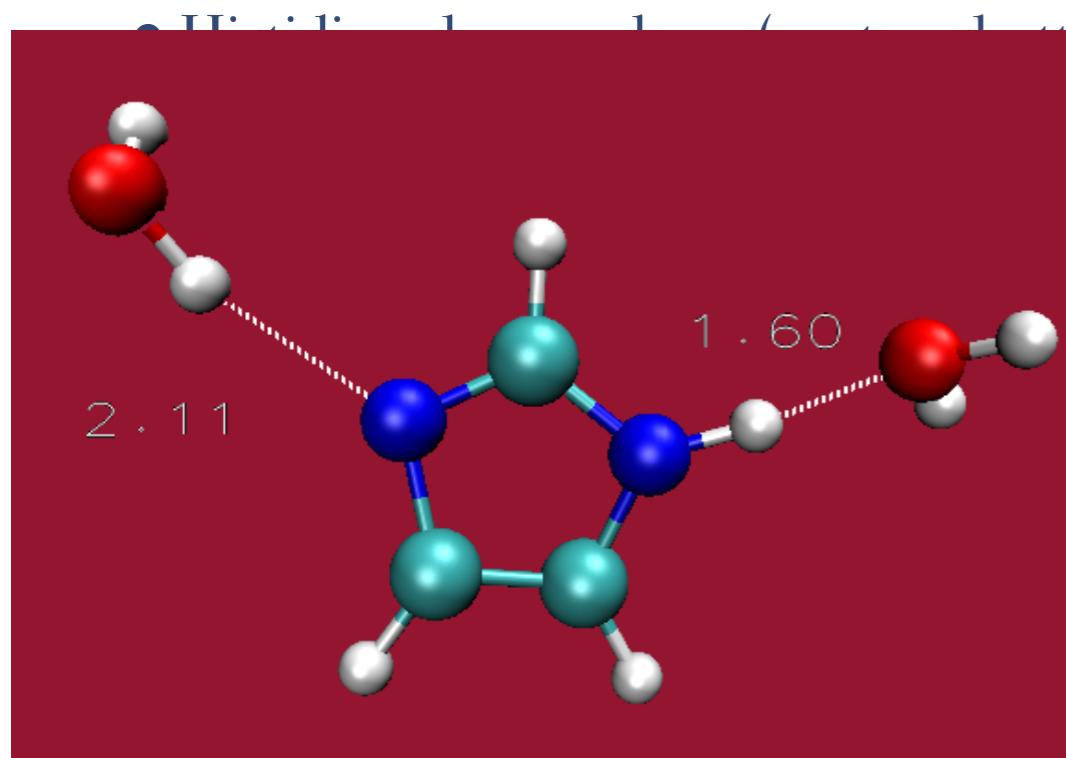
- Mapping Spectra onto energetics and dynamics requires computations
 - Static with high order gradients effective for cold gas phase
 - Spectra computed from dynamics trajectory promising for energized and floppy molecules
- DFTB within CP2K fast, benchmark accuracy
 - Can include nuclear quantum effect: PIMD

Conclusions

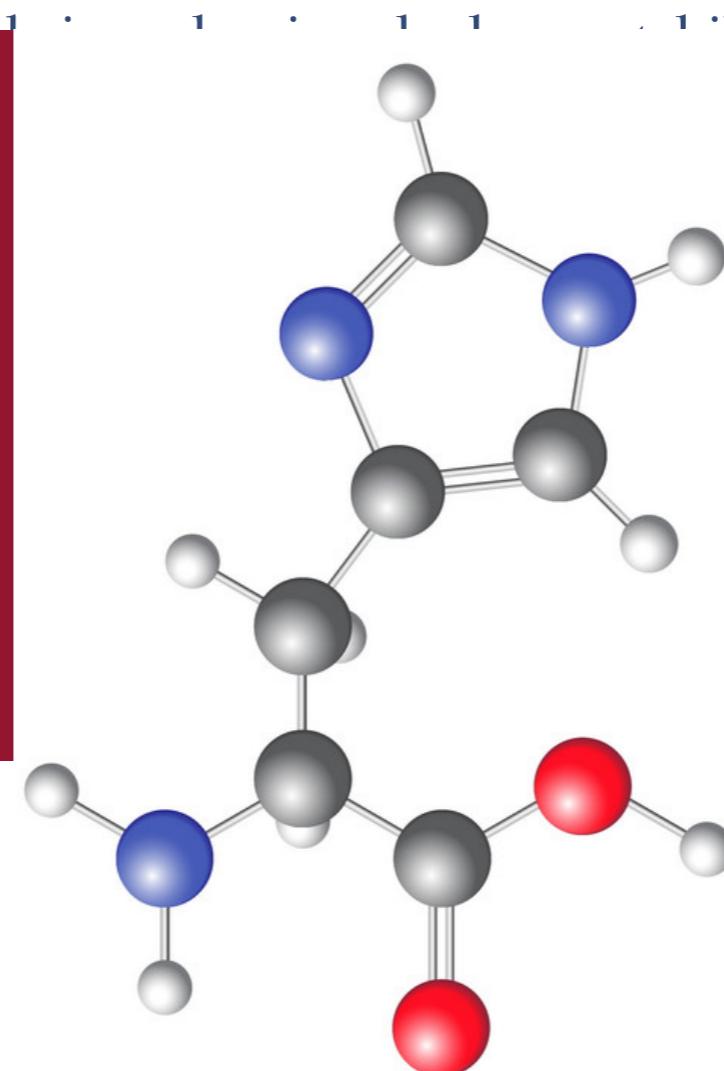
- Mapping Spectra onto energetics and dynamics requires computations
 - Static with high order gradients effective for cold gas phase
 - Spectra computed from dynamics trajectory promising for energized and floppy molecules
- DFTB within CP2K fast, benchmark accuracy
 - Can include nuclear quantum effect: PIMD
- AIMD with DFT moderate speed, further benchmarking

Imidazole

Imidazole



Imidazole



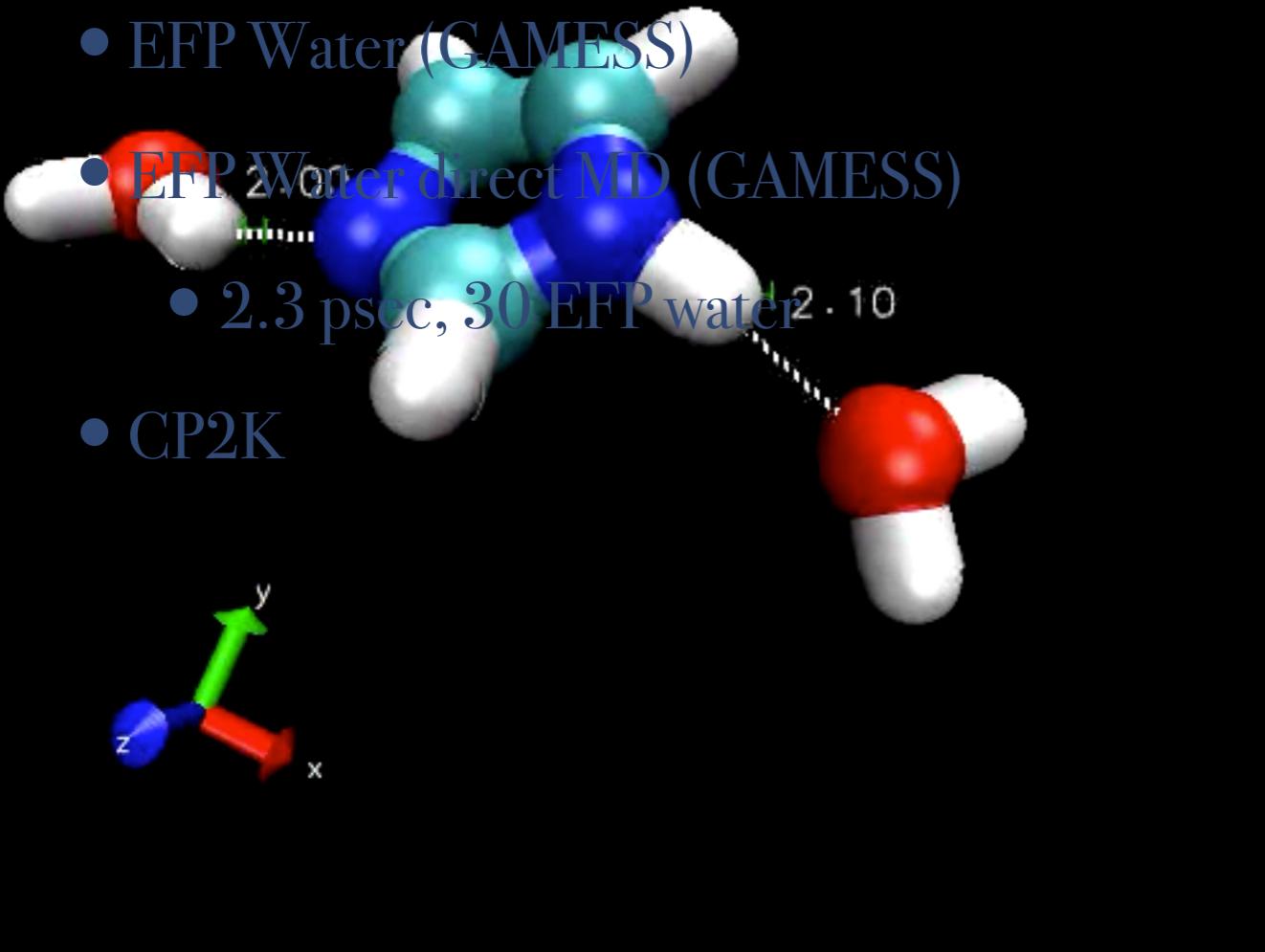
Histidine

uses oxyhemoglobin)

Solvated

Computational Approach

- Anharmonic gas phase frequencies
- Continuum solvated
- 1-4 explicit waters (Monte Carlo Sampled Geometry)+ continuum
- EFP Water (GAMESS)
- EFP Water direct MD (GAMESS)
 - 2.3 psec, 30 EFP water^{2 · 10}
- CP2K



Solvated

Computational Approach

- Anharmonic gas phase frequencies
- Continuum solvated
- 1-4 explicit waters (Monte Carlo Sampled Geometry)+ continuum
- EFP Water (GAMESS)
- EFP Water direct MD (GAMESS)
 - 2.3 psec, 30 EFP water^{2 · 10}
- CP2K

