Exploring Phosphine Ligation States through QM/MM Simulations with Equivariant Graph Neural Networks

Master's Thesis Computational Science and Engineering

Carried out at ETHZ (Zürich) in the Research Group Computational Chemistry
Supervised by Prof. Sereina Riniker, Dr. Felix Pultar and Univ.Prof. Georg Madsen

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Author

LinkedIn

paper preprint

Eva Doloszeski BSc MSc

- **Bachelors:** Technical Chemistry
- **Key Area Chemistry and Materials Science**
  - Understanding Advanced Molecular Simulation (EPFL)
  - Atomistic and Quantum Simulation of Materials (EPFL)
  - Physical and Theoretical Solid State Chemistry (P. Blaha)
  - Concepts in Condensed Matter Physics (J. Kunes)
  - Introduction to Atomistic Calculations (J. C. Montaña)
- **Key Area Electronics**
  - Introduction to Finite Element Methods in Solid Mechanics (D. Pahr)
  - Introduction to Semiconductor Physics and Devices (M. Waltl)
  - Properties of Semiconductors and related Nanostructures (EPFL)
  - Nanotechnology (EPFL)
- **Current Occupation:** PhD Candidate in the Theoretical Chemistry Group of Georg Madsen at TU Vienna
- **Favorite Lectures:** Atomistic and Quantum Simulation of Materials, Introduction to Atomistic Calculations, Computational Science on Many-Core Architectures, NSSC II

Aim

The aim of this work was to validate a previously developed equivariant graph neural network, featuring anisotropic message passing (AMP), investigating ligation states of phosphine ligands within transition metal complexes. AMP was used to substitute quantum mechnical calculations in a QM/MM scheme using umbrella sampling with the final goal of predicting reaction free energies.

Introduction

QM/MM

Hybrid QM/MM methods combine quantum mechanics (QM) with molecular mechanics (MM) for describing the potential energy surface of a system necessary for molecular dynamics (MD) simulations. QM calculations allow for greater accuracy and description of electronic processes necessary for the ligation state reaction. MM descriptions are more efficient and good for bulk properties ideal for dacription of solvents.

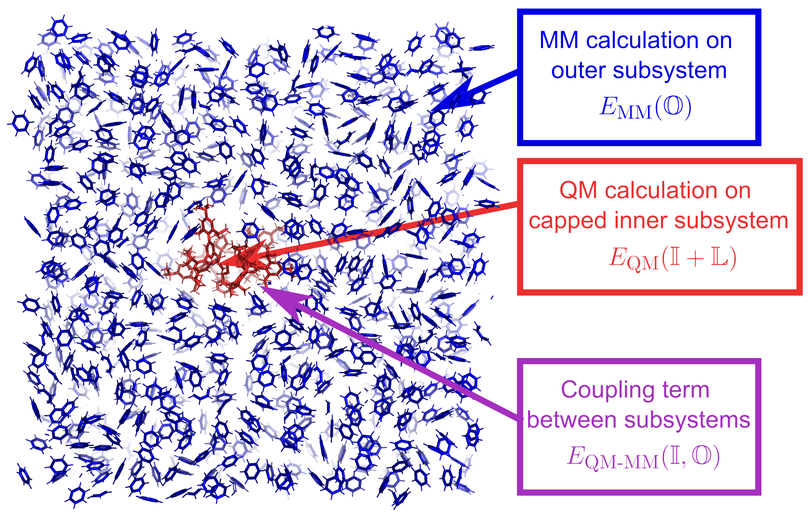


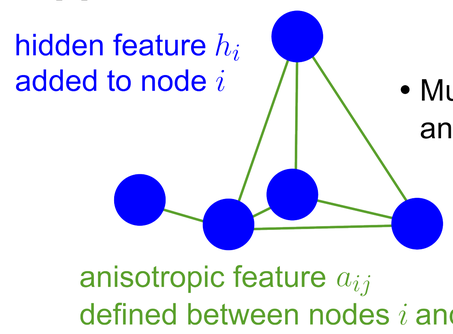
Figure: PteroPhos complex in benzene solvent

Additive QM/MM scheme

$$E^{\mathrm{add}}_{\mathrm{QM/MM}}(\mathbb{S}) = E_{\mathrm{MM}}(\mathbb{O}) + \underbrace{E_{\mathrm{QM}}(\mathbb{I}) + E^{\mathrm{el}}_{\text{QM-MM}}}_{\text{Substitute with ML potential}} + E^{\mathrm{vdW}}_{\text{QM-MM}}$$

AMP

The **A**nisotropic **M**essage **P**assing model was used in the QM/MM MD simulations to substitute the QM part. [2]



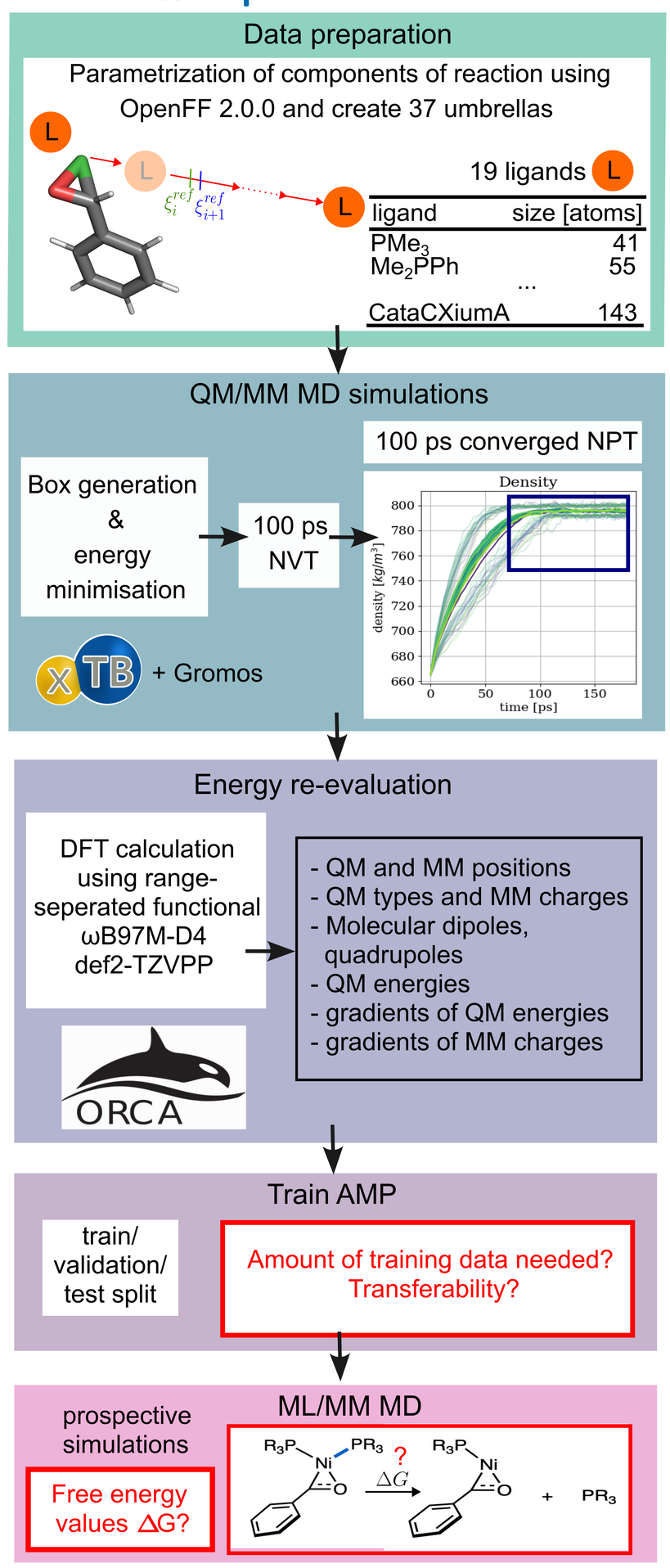
- Multipoles added as anisotropic state to h_i
- Message passing altered by anisotropic feature a_{ij}

$$\mathbf{h}_i^{l+1} = \phi_h(\mathbf{h}_i^l, \textstyle\sum_{j\in N_{QM}(i)} \phi_e(\mathbf{h}_i^l, \mathbf{h}_j^l, \mathbf{a}_{ij}))$$

- Interaction with MM particles added

$$E_{\mathrm{pot}} = \textstyle\sum_i^{QM} \phi_V(\mathbf{h}_i^n) + \sum_i^{QM}\sum_j^{MM} \frac{\phi_\rho(\mathbf{h}_i^n)\cdot q_j}{||r_{ij}||}$$

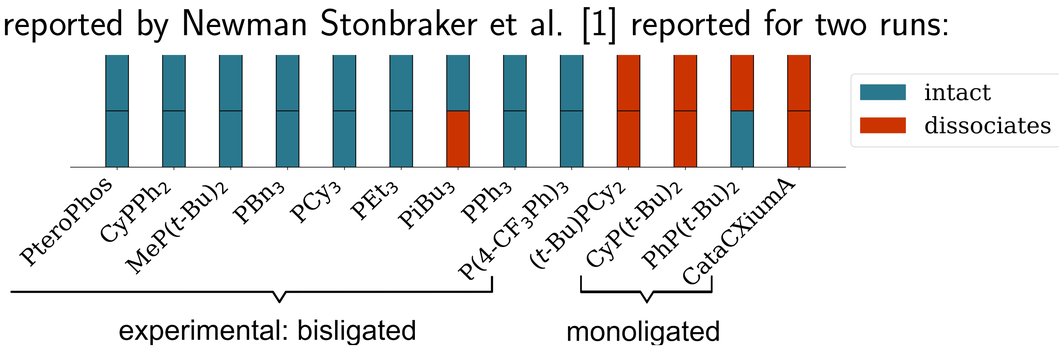
Method & Implementation



Results and Discussion

Transferability: A model with 606 568 parameter trained with 42 000 frames yielded in a mean absolute error (MAE) for the energies of the QM atoms of 2.2 kJ/mol. Transferability was evaluated for 4 ligands (PEt_3, $PhP(t\text{-}Bu)_2$, PCy_3, CataCXiumA), by not including the data in the training. The MAE for the energies stayed within chemical accuracy only for CataCXiumA it raised to 6.2 kJ/mol due to the unique adamantane group. However, the error could be reduced by 1.8 kJ/mol if only eight frames of CataCXiumA were included in the training set.

Dissociation of monoligated and bisligated complexes: Prospective simulations of all monoligated complexes showed that none of them dissociated. Whereas the dissociation of the bisligated complexes coincided well with the experimental results reported by Newman Stonbraker et al. [1] reported for two runs:



Free energy: The free energy was calculated for complexes with PCy_3 and PteroPhos using umbrella sampling and prospective simulations. The correct ligation state was determined in both cases even if the training data did not contain data of the two compelexes.

Conclusions

The AMP model was effectively applied in the context of (QM)ML/MM simulations utilizing electrostatic embedding. The pipeline allowed for substituting the traditional QM approach with a machine learning model, aiming to achieve DFT like accuracy. For different bisligated and monoligated transition metal complexes stable simulations of 1 ns were produced, yielding a combined simulation time of 195 ns. A significant acceleration of approximately a factor of 10^4 to 10^5 was estimated when comparing the DFT method to the machine learning model. Therefore results could be yielded for the PteroPhos ligand, comprising 385 atoms, otherwise inaccessible at this level of theory. Exploration to new ligands was shown by the transferability experiments and the free energy values obtained.

[1] Samuel H Newman-Stonebraker et al. "Univariate classification of phosphine ligation state and reactivity in cross-coupling catalysis". In: *Science* 374.6565 (2021), pp. 301–308.

[2] Moritz Thürlemann and Sereina Riniker. "Anisotropic message passing: Graph neural networks with directional and long-range interactions". In: *The Eleventh International Conference on Learning Representations*. 2023.