



## Abstract

To validate a previously developed equivariant graph neural network (GNN) featuring anisotropic message passing, investigation of ligation states of phosphine ligands within transition metal complexes was chosen. These ligands play a key role in cross-coupling reactions. By integrating anisotropic states inspired by the Cartesian multipole formalism into the neural network, the network captures directional information. The neural network was successfully used to replace the quantum mechanical calculations in a quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) simulation of 20 distinct phosphine ligands attached to a nickel-benzaldehyde complex.

A speed-up factor of  $10^5$  was obtained for the largest complex in the training data with the ligand CataCXiumA using the ML model substituting the DFT method. The model with approximately 600 000 parameters achieved a mean absolute error (MAE) for energies of 2.2 kJ/mol when it was trained on 42 000 data points including coordinates and charges to reproduce forces, energies and molecular multipole information. The model showed good transferability for four ligands, which were not present in the training data set, maintaining a mean absolute error for energies between 2.3 and 6.2 kJ/mol. Chemical accuracy was achieved for three of these ligands. For all complexes ML/MM MD trajectories could be produced for extensive periods of time, spanning several hundred picoseconds. Prospective simulations indicate comparable trends across all 13 ligands, comparable to experimental data from Newman-Stonebraker et al.<sup>1</sup> For two complexes, (PteroPhos)<sub>2</sub>Ni(benzaldehyde) and (PCy<sub>3</sub>)<sub>2</sub>Ni(benzaldehyde), free energy profiles obtained from prospective simulations using umbrella sampling correctly predicted the experimentally observed ligation state, when transferred to the monoligated equivalent. Remarkably, the largest system, PteroPhos, comprising 385 QM atoms and 4 000 MM atoms, was accurately predicted despite its size and not being present in the training or validation set demonstrating the scalability of the method.

This new method allows the description of systems that were previously difficult or impossible to describe due to their size and complexity, which prevented the use of more established methods such as pure QM calculations or the force field formalism.

## Acknowledgement

I would like to express my gratitude to Dr. Felix Pultar, who supervised my master's thesis. He provided me with invaluable guidance, answered all my questions, and was always available to provide constructive feedback and engage in productive discussions. It was a pleasure to work with someone who is as dedicated and enthusiastic about his work as he is. I would also like to thank Prof. Dr. Sereina Riniker for giving me the opportunity to pursue my master's thesis in her research group at ETH Zurich and for her invaluable input during my computational work whenever it was needed. My sincere appreciation also goes to Prof. Dr. Philippe Hünenberger for his enlightening discussions and insights. I would also like to thank Univ. Prof. Georg Kent Hellerup Madsen, who made my project at ETH Zurich possible by supervising my thesis at my home university, the Vienna University of Technology. Furthermore, I am grateful for the MARVEL INSPIRE Potentials Master's Fellowships, which provided me with financial support and the opportunity to participate in all the events and seminars they offered. I am also grateful to all the other members of the Computational Chemistry group for the many enjoyable lunches and coffee breaks, as well as the football games and ski-touring events, which provided me with a warm welcome within the group. Finally, I would like to thank my parents for their guidance and support throughout my Master's and Bachelor's degrees. I would also like to thank all my friends for their emotional support whenever needed.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgment</b>	<b>iii</b>
<b>List of Abbreviations</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Phosphine Ligands in Cross Coupling Reactions . . . . .	1
1.2 Molecular Model . . . . .	3
1.2.1 Molecular Dynamics (MD) and Stochastic Dynamics (SD) . .	3
1.2.2 Quantum Mechanics/ Molecular Mechanics (QM/MM) . . . .	5
1.3 Umbrella Sampling . . . . .	9
1.4 Neural Network Potentials in Computational Chemistry . . . . .	12
1.4.1 Anisotropic Message Passing - AMPv2 Model . . . . .	13
1.5 Aim of This Work . . . . .	15
<b>2 Computational Details</b>	<b>17</b>
2.1 Data Preparation . . . . .	19
2.2 QM/MM MD Runs . . . . .	20
2.2.1 Training Data Generation . . . . .	20
2.2.2 Lennard-Jones Parameters and Force Constants . . . . .	22
2.2.3 Energy Re-Evaluation . . . . .	22
2.3 AMP Training and Testing . . . . .	23
2.4 Prospective Simulations . . . . .	25
2.5 QM/MM MD Simulation for $L_xNi(4\text{-fluorobenzaldehyde})$ Complexes .	28
<b>3 Results and Discussion</b>	<b>29</b>
3.1 Effect of LJ Parameters . . . . .	29
3.2 Effect of Force Constants . . . . .	33
3.3 Training Data Generation . . . . .	35
3.4 QM/MM MD for $L_xNi(4\text{-fluorobenzaldehyde})$ . . . . .	42
3.5 xTB versus DFT . . . . .	44
3.6 Training AMP Models . . . . .	48
3.6.1 Effect of Amount of Training Data Used . . . . .	48

---

3.6.2	Transferability Experiment . . . . .	50
3.7	Prospective ML/MM MD Simulations . . . . .	56
3.7.1	Stability Test for Bisligated Complexes Using ML/MM MD . . . . .	58
3.8	Free Energy Calculation Using ML/MM MD . . . . .	61
<b>4</b>	<b>Conclusion and Outlook</b>	<b>66</b>
	<b>References</b>	<b>76</b>
	<b>Supplementary Information</b>	<b>77</b>

## List of Abbreviations

AMP	anisotropic message passing
ANN	artificial neural networks
DFT	density functional theory
DFTB	density functional tight binding
GNN	graph neural network
HDNNP	high-dimensional neural network potential
KS	Kohn–Sham
LJ	Lennard-Jones
MAE	mean absolute error
MD	molecular dynamics
NN	neural network
PES	potential energy surface
RDF	radial distribution function
RMSD	root-mean-square deviation
RT	room temperature
SCC	self-consistent cycle
SD	stochastic dynamics
WHAM	weighted histogram analysis method
HF	Hartree-Fock

## List of Figures

1	Catalytic cycle of cross coupling reactions. . . . .	1
2	Hierarchy of methods in atomic scale modeling. . . . .	6
3	Additive QM/MM scheme for (PteroPhos) <sub>2</sub> Ni(benzaldehyde) in solvent simulation box. . . . .	8
4	Umbrella sampling with harmonic potential. . . . .	10
5	Reaction for ligation state determination. . . . .	15
6	Schematic representation of the principal stages in the generation of training data. . . . .	17
7	Ligands L used in L <sub>x</sub> Ni(benzaldehyde) complexes with x=1,2 (monoligated vs bisligated complex). . . . .	18
8	Additional ligands L used in L <sub>x</sub> Ni(4-fluorobenzaldehyde) complexes with x=1,2 (monoligated vs bisligated). . . . .	28
9	LJ potential for three LJ parameter sets tested for nickel. . . . .	31
10	Effect of different LJ parameters investigated using RDF (hydrogen of solvent taken as reference). . . . .	32
11	Explanatory images for force constant determination. . . . .	34
12	Effect of different force constants on phosphine ligand as it is moved away from the complex. . . . .	35
13	Overall runtime of xTB to run 100 ps MD for differently sized systems. . . . .	36
14	Monitored density and temperature values in the training data. . . . .	37
15	Distribution of distances between nickel and phosphine ligands in the training data for the L <sub>2</sub> Ni(benzaldehyde) complex. . . . .	39
16	Configurational composition of training data for the first ten umbrellas for highest force constant. . . . .	41
17	Distribution of distances between nickel and phosphine ligands in the training data for the L <sub>2</sub> Ni(4-fluorobenzaldehyde) complex. . . . .	43
18	Energy values of xTB versus DFT in training data set. . . . .	45
19	Forces for QM atoms evaluated by xTB versus DFT in training data set. . . . .	47
20	Effect of different amounts of training data. . . . .	49
21	Ligands in data sets 2 and 3. . . . .	51
22	Model versions 5 - 9: Transferability experiment. . . . .	53
23	Model versions 5 - 9: Transferability experiment analyzed for different ligands. . . . .	54
24	Model versions 11 - 15: Transferability experiment analyzed for different ligands. . . . .	55

25	MAEs for models used for prospective simulations. . . . .	56
26	Overall runtime of PyTorch to run 100 ps ML/MM MD for differently sized systems. . . . .	57
27	Stability test with ML/MM MD simulations for all bisligated complexes. . . . .	58
28	Stability test with ML/MM MD simulations for all monoligated complexes. . . . .	60
29	Example of probability densities for individual umbrellas. . . . .	62
30	Free energy profile for $(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$ . . . . .	63
31	Free energy profile for $(\text{PteroPhos})_x\text{Ni}(\text{benzaldehyde})$ . . . . .	64
S1	Effect of different LJ parameters investigated using RDF (carbon of solvent taken as reference). . . . .	80
S2	Effect of different force constants on the distances between nickel and the neighboring atoms. . . . .	81
S3	RDF for all monoligated and bisligated complexes in the training data. . . . .	82
S4	Distribution of distances between nickel and benzaldehyde in training data for $\text{L}_1\text{Ni}(\text{benzaldehyde})$ complex. . . . .	83
S5	Distribution of distances between nickel and phosphine ligand in training data for $\text{L}_1\text{Ni}(\text{benzaldehyde})$ complex. . . . .	84
S6	Distribution of distances between nickel and benzaldehyde in training data for $\text{L}_2\text{Ni}(\text{benzaldehyde})$ complex. . . . .	85
S7	Configurational composition of training data for the first ten umbrellas for lowest force constant. . . . .	86
S8	Forces for MM atoms evaluated by xTB versus DFT in training data set. . . . .	87
S9	Effect of different amounts of training data, part 2. . . . .	88
S10	Model versions 5 - 9: Transferability experiment, part 2. . . . .	88
S11	Model versions 5 - 9: Transferability experiment analyzed for different ligands, part 2. . . . .	89
S12	Model versions 11 - 15: Transferability experiment analyzed for different ligands, part 2. . . . .	89
S13	Models 11 - 15: Transferability experiment. . . . .	90

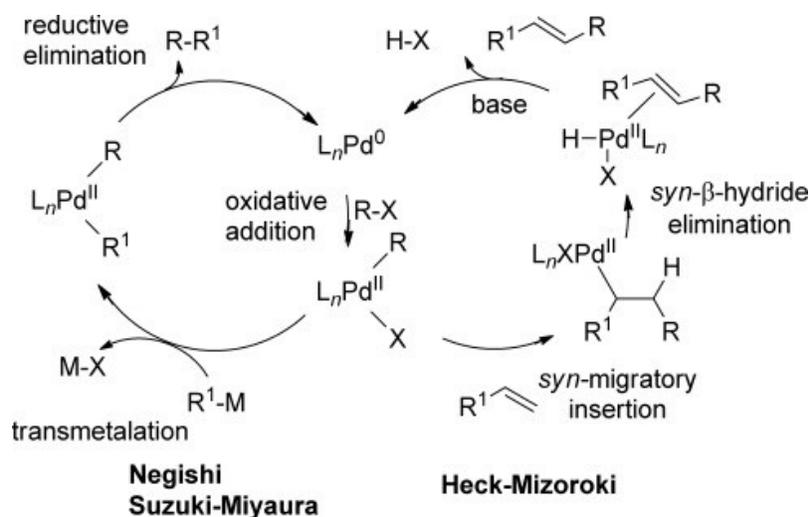
## List of Tables

1	Main differences between minimal and precision model. . . . .	24
2	Composition of train/validation/test set of all model versions. . . . .	26
3	Composition of data sets 1, 2, and 3. . . . .	27
4	Details for hyperparameter tuning. . . . .	27
5	Lennard-Jones parameters and van der Waals radii for atoms H, C, and Ni. . . . .	30
6	Performance of xTB versus DFT. . . . .	46
7	Train/validation/test split for testing effect of amount of training data. . . . .	49
8	Train/validation/test split for testing transferability. . . . .	52
9	Stability results for bisligated complexes from ML/MM MD runs. . . . .	59
10	Free energy differences from umbrella sampling. . . . .	65
S1	Distances between nickel and ligands for MD runs for determining LJ parameters. . . . .	77
S2	List of abbreviations and long-form names of ligands. . . . .	78
S3	Additional information about ligands and their complexes (Kraken ID, ligation state and complex size). . . . .	79

# 1 Introduction

## 1.1 Phosphine Ligands in Cross Coupling Reactions

Transition metal-catalyzed cross-coupling reactions were first employed to form new carbon-carbon bonds using palladium as a catalyst. The contributions to this reaction were the subject of a Nobel Prize awarded in 2010 to Richard Heck, Ei-ichi Negishi, and Akira Suzuki. The foundation that they established has enabled the development of numerous other reactions, including carbon-heteroatom coupling,  $\alpha$ -arylation, and direct arylation by C-H activation.<sup>2-5</sup> The catalytic cycle for both Mizoroki-Heck and Negishi, and Suzuki-Miyaura reactions are shown in Figure 1.



**Figure 1: Catalytic cycle of cross coupling reactions.** The principal stages of the catalytic cycle for Mizoroki-Heck coupling and Negishi and Suzuki-Miyaura reaction. With  $L_nPd^0$  being the catalyst with ligand L,  $R-X$  an aryl halide or pseudohalide. For the Heck-Mizoroki reaction an alkene is added. For the Negishi Suzuki-Miyaura cycle, an organometallic species  $R^1-M$  is introduced. Image from C. Johansson et al.<sup>3</sup>

Cross-coupling reactions, together with amide bond formation or  $S_NAr$  reactions, are among the most important reactions used in medicinal chemistry for drug synthesis. The largest category of cross-coupling reactions is the Suzuki-Miyaura reaction, which employs palladium as a catalyst.<sup>6</sup> Some examples where these reactions play a key role in synthesis include the synthesis of losartan, which regulates blood pressure; discodermolide, a cell growth inhibitory agent; and Merck's Singulair process, which

is used to treat asthma.<sup>3</sup>

The development of coupling reactions can be divided into several waves, with the first wave being used primarily for the investigation of potential catalysts, the second wave for the expansion to different coupling partners, and the third wave including fine-tuning, which was achieved, for example, by varying the ligand or substrate. A combination of steric and electronic effects of the ligands was found to influence the outcome of the individual steps of the catalytic cycle and thus the overall outcome of the reaction.<sup>3</sup> Examples of ligand design for palladium-catalyzed cross-coupling reactions include diadamantyl ligands such as CataCXium, dialkylbiaryl phosphines or Buchwald ligands.<sup>7,8</sup>

In addition to palladium-catalyzed reactions, nickel can also be employed as a transition metal catalyst. The cost of nickel is lower than that of palladium, and due to its different properties, it can access new reactions or alter their efficiency. It is important to note that ligands that are useful in palladium-catalyzed reactions are not necessarily effective in nickel-catalyzed reactions. Therefore, new ligands must be developed.<sup>9,10</sup> One approach for identifying new ligands is to utilize descriptors, which encapsulate the steric and electronic properties of ligands, thereby enabling the prediction of the outcome of different reactions.<sup>11</sup>

S. Newman-Stonebraker et al. identified a descriptor, the minimum percentage buried volume  $\%V_{\text{bur}}(\text{min})$ , which could divide different data sets of monodentate phosphines for nickel and palladium with respect to how much yield was achieved in cross-coupling reactions.<sup>1</sup>  $\%V_{\text{bur}}(\text{min})$  represents the smallest  $\%V_{\text{bur}}$  among all conformers. This value indicates the amount of the steric bulk of the given ligand which fits into a sphere with a radius of 3.5 Å. The values for the steric descriptors were obtained a priori from the *kraken* virtual library and then compared to the yield of Ni-catalyzed Suzuki-Miyaura coupling (SMC) reactions and several Pd-catalyzed reactions.<sup>12</sup> Their hypothesis that the outcome of the reaction depends on the ligation state of the metal with respect to the ligand, was supported by an additional experiment. Ni(COD)<sub>2</sub> (COD, 1,5-cyclooctadiene), 4-fluorobenzaldehyde, and two equivalents of phosphine were brought to a reaction, and <sup>1</sup>H, <sup>19</sup>F as well as <sup>31</sup>P NMR spectroscopy were performed to determine the ligation state. The only descriptor that could differentiate between the ligation state results for the complexes was the  $\%V_{\text{bur}}(\text{min})$  with a value of 32%. In order to support their conclusions, the researchers employed a series of bulky ligands, namely DinoPhos ligands, with high cone angle values but low  $\%V_{\text{bur}}(\text{min})$ . These were created by K. Wu and A.

Doyle, who demonstrated their efficacy in the Ni-catalyzed C(sp<sup>3</sup>) Suzuki coupling of acetals with boronic acids.<sup>9</sup> Furthermore, an additional DinoPhos ligand, namely PteroPhos, was designed with an even larger size and a cone angle of 235°, but %V<sub>bur</sub>(min) of 27.2% which is below their cutoff found. In fact, the ligand was also found to be reactive in the cross-coupling reaction and to be most stable as the bisligated L<sub>2</sub>Ni(4-fluorobenzaldehyde) complex.

## 1.2 Molecular Model

Molecular models employ physical laws to elucidate the behavior of molecules. A number of fundamental choices must be made at the outset, including which degrees of freedom to consider, how to describe the interactions in terms of the degrees of freedom, the method of generating the configurations, and what boundary conditions to use. The first two choices strongly influence each other. The degrees of freedom are the entities that are modeled in the system. These can include nuclei, electrons, atoms or united atoms. The choice of which entities to include depends on whether the interaction is described quantum mechanically or classically. This aspect will be discussed in greater detail in Section 1.2.2. The third choice determines how the values of the degrees of freedom are sampled. This can be based on random search techniques, such as Monte Carlo methods, or on molecular dynamics, where integration of time-dependent Schrödinger, Newtonian or Langevin equations of motion is performed. The latter two options are outlined in greater detail in Section 1.2.1.<sup>13</sup>

### 1.2.1 Molecular Dynamics (MD) and Stochastic Dynamics (SD)

The trajectories obtained from dynamics simulations can be used to obtain averages, distributions, and time series of various physical quantities, as well as to predict the macroscopic behavior of the system.<sup>13,14</sup> One of the major advantages is that a significant number of properties are accessible that cannot be measured experimentally, cannot be physically observed, or are too expensive to study. In addition, simulations provide temporal resolution in the femtosecond time domain and spatial information at the atomic level. For example, dynamic properties such as transport phenomena, conformational changes, protein folding or ligand binding can be observed, as well as the flexibility and stability of conformations. It is also possible to precisely control and manipulate the environment. Examples of such perturbations include changes in temperature or pressure, observation of the response of the system to such changes, addition or removal of ligands, mutation of atoms, application of forces or external fields, and modification of the environment in general. As a result, these simulations serve as a valuable complement to experimental work, facilitating the elucidation of

the underlying mechanisms.<sup>15</sup>

In molecular dynamics (MD) simulations, Newton's equations of motion are the underlying laws of nature that govern the movement of particles. These can be expressed in the Hamiltonian description by the following equation.<sup>16</sup>

$$\dot{q} = M^{-1}p, \quad \dot{p} = F = -\nabla E_{\text{pot}}(q) \quad (1)$$

$(q, p)$  denote the configuration of the system, where  $q$  are the coordinates and  $p$  is the momentum vector.  $M$  is a constant diagonal mass matrix of the system. Furthermore,  $E_{\text{pot}}$  denotes the potential energy surface and  $F$  is the force vector. These equations can be expressed in terms of a system with  $N$  particles, where  $q^N$  are the Cartesian coordinates for all particles and  $p$  is expressed in terms of  $v^N$ , the velocity vector. Using these equations, the positions and velocities of all particles can be updated to propagate the system.

$$\dot{q}^N(t) = v^N(t), \quad M\dot{v}^N(t) = -\nabla E_{\text{pot}}(q^N(t)) \quad (2)$$

Stochastic dynamics is seen as an extension of molecular dynamics, where the stochastic Langevin equation of motion is integrated to update positions and momenta in the system. The second equation of Equation 2 changes to:<sup>14,16,17</sup>

$$M\dot{v}^N(t) = -\nabla \overline{E_{\text{pot}}}(q^N(t)) - M\gamma v^N(t) + f^{st,N}(t) \quad (3)$$

If we compare this to Newton's equation of motion, we can see that the main form of the equation remains the same, but additional terms are added to the forces.  $-\nabla \overline{E_{\text{pot}}}(q^N(t))$  is the mean force including all the forces between the explicit particles, plus accounts for the average of the implicit particles. The coefficient  $\gamma$  is the friction coefficient or collision rate and the whole term  $M\gamma v^N(t)$  is the frictional force. The stochastic force  $f^{st}$  for a single atom  $i$  can be expressed by:

$$\langle (f_i^{st})^2 \rangle = 6M_i\gamma_i k_B T_{ref} \quad (4)$$

Here  $T_{ref}$  is the reference temperature that can be set to determine the temperature to use stochastic dynamics like an external heat bath.  $k_B$  is the Boltzmann's constant and the values  $f_i^{st}$  are distributed with a Gaussian probability distribution with zero mean. Instead of using it as a thermostat, one can also use it to replace solvent molecules in a simulation.<sup>14</sup>

### 1.2.2 Quantum Mechanics/ Molecular Mechanics (QM/MM)

Combined quantum mechanics (QM)/ molecular mechanics (MM) methods are hybrid methods where parts of the system are treated quantum mechanically and other parts classically to obtain a potential energy surface. The main goal is to combine the advantages of both approaches. QM calculations generally allow for greater accuracy and generality. For example, a QM description is required for all cases where electronic processes occur, including electronic excitations, charge transfer, or the breaking/formation of new bonds. Most importantly, MM calculations are more efficient and allow access to larger systems, involving several thousand atoms, and to longer timescales. They also allow modeling of bulk properties such as density, dielectricity, isothermal compressibility or heat capacity, which is important for the description of solvents.<sup>18-23</sup>

Possible applications are therefore all systems where a quantum mechanical approach is required to describe an electronic event, and where the interaction with the environment is also crucial, but can be adequately described classically. It is usually used in energy minimization, molecular dynamics (MD), or Monte Carlo (MC) methods. Biomolecular applications are often suitable, especially because of the large systems to be described. Most commonly, enzymatic reactions are modeled where all atoms directly involved in the enzyme-substrate binding are treated quantum-mechanically, allowing for a change in charge distribution, and all other atoms and the solvent are treated classically, still allowing for steric and electrostatic interactions.<sup>24</sup> Furthermore, it can be used to incorporate solvent effects by treating the solvent at the classical level and the solute at the QM level.<sup>21</sup>

#### Molecular Mechanics (MM) Using Fixed-Charge Force Fields

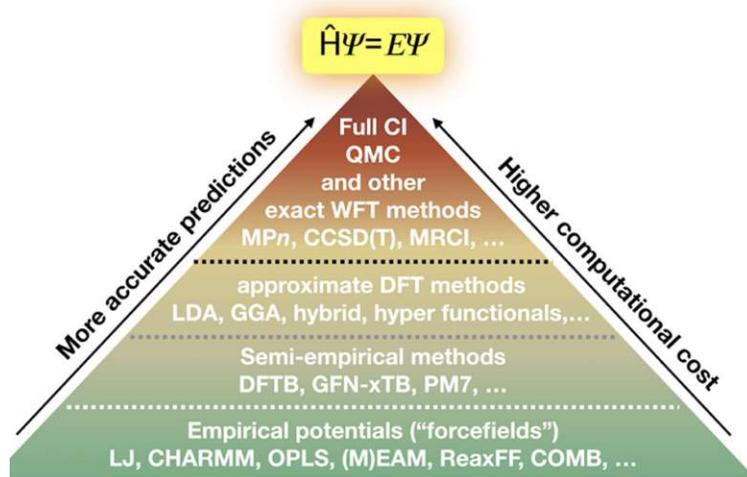
The MM region can be described by fixed-charge force fields, which are a classical functional form for the potential energy consisting of a sum of atomic interaction terms of the following or a similar form.<sup>13,14,18,25</sup>

$$E_{\text{pot}}(\vec{r}) = \sum_{\text{bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{\text{torsion}} K_\phi [1 + \cos(n\phi - \delta)] + \sum_{\text{improper}} K_\xi (\xi - \xi_0)^2 + \sum_{\text{pairs } (i,j)} \left\{ \left[ \frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^6} \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \right\} \quad (5)$$

The first four terms are covalent terms, accounting for bond stretching, bond-angle

bending, dihedral-angle torsion, and improper dihedral-angle bending. The last two terms are non-bonded contributions, including van der Waals interactions and electrostatic Coulomb terms. Depending on the force field used, the different terms are parameterized using a mixture of *ab initio* QM methods and experimental data, including structural, spectroscopic, thermodynamic, dielectric, and transport data.<sup>13,26</sup> The constants parameterized within a force field are  $K_b$ ,  $K_\theta$ ,  $K_\phi$ ,  $K_\xi$ ,  $C_{12}$ ,  $C_6$ . Furthermore,  $b$  is the bond length;  $\theta$  is the bond angle;  $\phi$ ,  $\delta$  and  $n$  are the torsion angle, the phase angle, and the multiplicity; and  $\xi$  is the improper dihedral angle.  $r_{ij}$  is the distance between atoms  $i$  and  $j$ ;  $q$  the partial charge;  $\epsilon_0$  the electric constant; and  $\epsilon_1$  the background electric permittivity. If the subscript is 0, the reference property is denoted. Van der Waals and electrostatic interactions are usually approximated or neglected after a certain cutoff.

## Quantum Mechanics (QM)



**Figure 2: Hierarchy of methods in atomic scale modeling.** Ranking of atomistic modeling methods used for creating potential energy surfaces (PES) with respect to computational cost and accuracy. Image from J. Keith et al.<sup>27</sup>

For the QM regime, one has to choose any quantum mechanical approach. Depending on the level of accuracy one wants to achieve, one can choose between semi-empirical approaches, other density-functional theory (DFT) methods, or higher-level post-Hartree-Fock methods, including, for example, coupled-cluster theory. These three levels are shown in the Figure 2, which shows how the computational cost increases with the accuracy of the latter methods.<sup>27</sup>

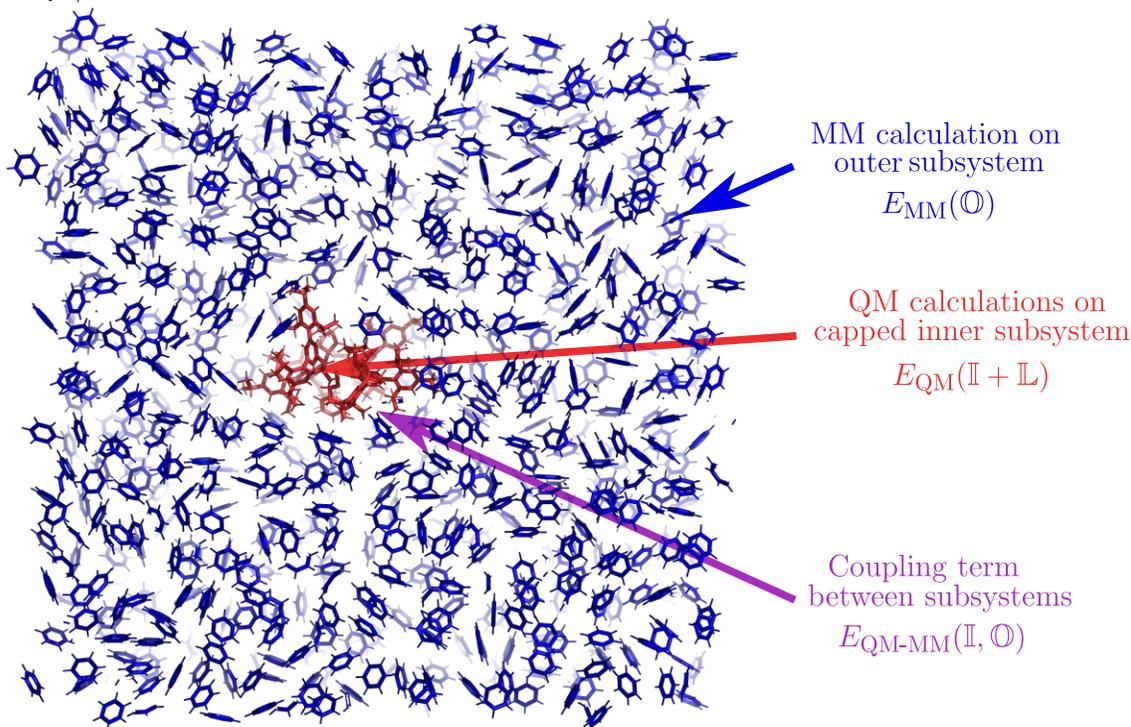
The top level of the pyramid (see Figure 2) introduces the smallest approximations for solving the Schrödinger equation of a system, commonly the Born-Oppenheimer approximation together with the Slater determinant approximation. These methods are called wavefunction theory methods. The simplest of these is the Hartree-Fock (HF) method, which does not allow for electron-electron interactions beyond the mean-field approximation of these interactions. To reintroduce these interactions, correlated wavefunction methods must be used, including Møller-Plesset perturbation theory (e.g. MPn) or coupled cluster theory (e.g. CCSD(T)). One level below are density functional theory (DFT) methods, which are based on the electron density rather than the wavefunction. The most commonly used is the Kohn-Sham(KS)-DFT, which assumes non-interacting electrons with a defined ground state density. Here, some terms in the energy functional, such as the kinetic energy of noninteracting electrons, the nuclear-electron potential and the classical Coulombic interaction, can be expressed exactly. Others resulting from electron interactions, commonly called the exchange-correlation term, must be approximated. This term can be expressed in a variety of ways by choosing a functional, such as LDA, GGA, range-separated, or hybrid functionals. The last level of QM methods in Figure 2 are semi-empirical methods that still consider electrons, but introduce more approximations using parameterized expressions.<sup>27,28</sup> These methods are derived by introducing approximations into either HF methods or KS DFT. Self-consistent field calculations are still employed, but are restricted to a valence-only minimal basis set. An example is density functional tight binding (DFTB) methods, which are derived from KS-DFT. In this approach, the electron density is decomposed into a density of free and neutral atoms and density fluctuations. The semi-empirical method employed in this study is GFN2-xTB, an extended DFTB method that utilizes a maximum of third-order density fluctuations. Its name is derived from its main applications, which include the computation of **g**eometries, **f**requencies, and **n**on-covalent interaction energies.<sup>27,29</sup>

### Coupling QM and MM

In QM/MM simulations, the QM and MM parts are combined and a coupling term is added to account for the interaction between the two zones. There are various schemes for combining the individual energy terms, including both subtractive and additive schemes. There are also different ways of describing the interaction between the two zones, namely mechanical embeddings, electrostatic embeddings and polarizable embeddings.<sup>18-20</sup> In this work, the additive scheme with electrostatic embedding is used. Therefore, only this scheme will be explained. The review by H. Senn and

Additive QM/MM scheme

$$E_{\text{QM/MM}}^{\text{add}}(\mathcal{S}) = E_{\text{MM}}(\mathcal{O}) + E_{\text{QM}}(\mathbb{I} + \mathbb{L}) + E_{\text{QM-MM}}(\mathbb{I}, \mathcal{O})$$



**Figure 3: Additive QM/MM scheme for  $(\text{PteroPhos})_2\text{Ni}(\text{benzaldehyde})$  in solvent simulation box.** Visual presentation of single contributions in additive scheme for QM/MM calculations: QM description for  $(\text{PteroPhos})_2\text{Ni}(\text{benzaldehyde})$  complex and MM description for benzene solvent.

Q. Thiel provides the notations and equations in the following section.<sup>18</sup>

In combined QM/MM methods, the system investigated is divided into a QM region and a MM region as shown in Figure 3. Here one of the complexes studied in this work is shown, where the chemically active and QM region is the complex and the MM region is the solvent contained in the system. As indicated in the figure the MM calculation is performed for the outer subsystem  $\mathcal{O}$ . The QM calculation is performed for the capped inner subsystem  $\mathbb{I} + \mathbb{L}$ , where  $\mathbb{L}$  accounts for linking atoms that are necessary when covalent bonds go from QM to MM regions. Linking atoms introduce a lot of approximations and should be avoided if possible. In addition to

these two terms another expression, the coupling term  $E_{\text{QM-MM}}(\mathbb{I}, \mathbb{O})$ , is needed for the additive scheme, as seen in Equation 6.

$$E_{\text{QM/MM}}^{\text{add}}(\mathbb{S}) = E_{\text{MM}}(\mathbb{O}) + E_{\text{QM}}(\mathbb{I} + \mathbb{L}) + E_{\text{QM-MM}}(\mathbb{I}, \mathbb{O}) \quad (6)$$

The coupling term has a non-negligible contribution and accounts for interactions between the two zones. These include bonded interactions  $E_{\text{QM-MM}}^b$ , van der Waals interactions  $E_{\text{QM-MM}}^{\text{vdW}}$ , and electrostatic interactions  $E_{\text{QM-MM}}^{\text{el}}$ .

$$E_{\text{QM-MM}}(\mathbb{I}, \mathbb{O}) = E_{\text{QM-MM}}^b + E_{\text{QM-MM}}^{\text{vdW}} + E_{\text{QM-MM}}^{\text{el}} \quad (7)$$

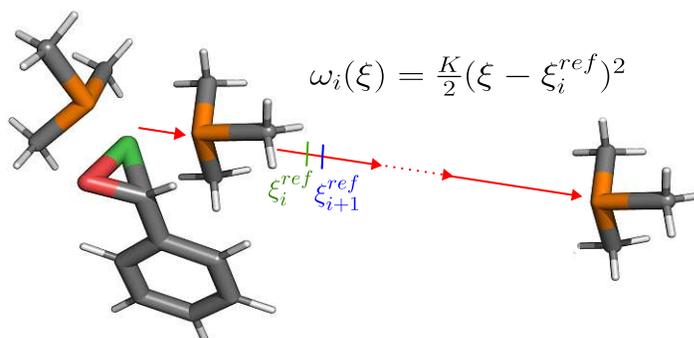
The expression for the electrostatic term  $E_{\text{QM-MM}}^{\text{el}}$  depends on the embedding used. For the simplest embedding, the mechanical embedding, the electrostatic interactions are treated classically by assigning point charges to the QM region similar to the MM region, or by calculating them anew in each step. For the electrostatic embedding, the chosen QM method must be able to incorporate the MM point charges as an electronic field for the following description.

$$\hat{H}_{\text{QM-MM}}^{\text{el}} = - \sum_i^N \sum_{J \in \mathbb{O}}^L \frac{q_J}{|r_i - R_J|} + \sum_{\alpha \in \mathbb{I} + \mathbb{L}}^M \sum_{J \in \mathbb{O}}^L \frac{q_J Q_\alpha}{|R_\alpha - R_J|} \quad (8)$$

The first term accounts for the interaction between the total number of electrons  $N$ , and a total of  $L$  partial charges  $q_J$  of the MM zone. The second term accounts for the Coulomb interaction between all partial charges of the MM zone and a total of  $M$  QM nuclear charges  $Q_\alpha$ .  $R$  and  $r$  are the coordinates of the nuclear charges and electrons. The most important disadvantages are that the atomic partial charges of the MM zone theoretically cannot be mixed with the charge distribution of the QM zone and further, the MM charges are fixed. To additionally introduce changes in the MM charges, one has to add another level of complexity using polarizable embedding. The other two terms of the additive scheme are treated purely classically, independently of the embedding. For the van der Waals interactions  $E_{\text{QM-MM}}^{\text{vdW}}$  a Lennard-Jones (LJ) potential is applied. A difficulty arises when one has to assign suitable LJ parameters, as these are sometimes not available or may change depending on the chemical environment, especially during a reaction.

### 1.3 Umbrella Sampling

Many biomolecular problems, such as folding, molecular complexation or membrane formation, have free energy differences in the order of 1-10  $k_B T$  (tens of kJ/mol).<sup>13</sup>



**Figure 4: Umbrella sampling with harmonic potential.** Explanatory image for umbrella sampling for reaction of  $(\text{PMe}_3)_2\text{Ni}(\text{benzaldehyde})$  to monoligated version  $\text{PMe}_3\text{Ni}(\text{benzaldehyde})$ . Biased potential  $\omega_i$  calculated for reaction coordinate  $\xi$  applied for each window  $i$ , depending on reference position  $\xi_i^{\text{ref}}$  and force constant  $K$ . Atoms displayed using different colors (P: orange, Ni: green, O: red, C: grey, H: white).

Crossing these high energy barriers in MD simulations is difficult or nearly impossible due to the time needed for the transition scaling exponentially with the height of the energy barrier.<sup>30</sup> Various methods have been developed to address this issue, including umbrella sampling, replica exchange methods, solute tempering, and local elevation.<sup>31–35</sup>

The umbrella sampling method was developed by M. Torrie and J. P. Valleau in 1976 as a means of circumventing the inefficiencies inherent in Boltzmann-weighted sampling in Monte Carlo (MC) methods for calculating free energy differences. In their work, they modified the underlying Boltzmann distribution for sampling by introducing a nonphysical distribution with an added weighting function. This new distribution was designed to (a) ensure that regions previously sampled correctly under the Boltzmann distribution are still accurately represented and (b) include regions not previously adequately sampled but which are relevant to the free energy calculation, thereby improving the efficiency and accuracy of these estimations.<sup>36</sup>

In practice, one must select a reaction coordinate,  $\xi$ , along which one desires to derive a free energy profile. This is typically done by connecting the two thermodynamic states of interest. The objective is to sample this space using a biased distribution that is as close to truly uniform as possible. Various approaches are employed, including the use of an adaptive bias that is iteratively adjusted, or harmonic bias potentials applied to multiple windows. In the latter approach, the range of the

reaction coordinate is divided into individual intervals. In each of these windows, a simple harmonic potential with a force constant  $K$  is defined, which penalizes any deviation of the system from the reference position  $\xi_i^{ref}$  (see Equation 9).<sup>31</sup> An example of defining a reaction coordinate  $\xi$  and how to divide it into windows is shown in Figure 4.

$$\omega_i(\xi) = \frac{K}{2}(\xi - \xi_i^{ref})^2 \quad (9)$$

This term is a simple additive term in the biased (superscript  $b$ ) potential energy surface  $E^b$  derived from the unbiased potential  $E^u$ :

$$E^b(r) = E^u(r) + \omega_i(\xi) \quad (10)$$

The Helmholtz free energy can easily be expressed in terms of the unbiased distribution  $Q$  as  $A = -\frac{1}{\beta} \ln Q$ , where  $\beta = \frac{1}{k_B T}$ ,  $k_B$  is the Boltzmann constant, and  $T$  is the temperature. The free energy profile can be obtained by establishing the relationship between the biased and unbiased distributions along the reaction coordinate  $Q(\xi)$ . This can be achieved by inserting the expressions for  $E^b(r)$  and  $E^u(r)$  into the following ensemble average expression:

$$Q(\xi) = \frac{\int \delta[\xi(r) - \xi] \exp[(-\beta E)] d^N r}{\int \exp[(-\beta E)] d^N r} \quad (11)$$

As a result, one obtains an expression for  $A$  for every window  $i$ , where  $P_i^b(\xi)$  is the biased distribution obtained from each MD run, and  $F_i$  is a constant independent of  $\xi$ :

$$A_i(\xi) = -\frac{1}{\beta} \ln P_i^b(\xi) - \omega_i(\xi) + F_i \quad (12)$$

with

$$F_i = -\frac{1}{\beta} \ln \int P^u(\xi) \exp[-\beta \omega_i(\xi)] d\xi \quad (13)$$

Following the completion of the simulation, a post-processing method, such as the weighted histogram analysis method (WHAM), can be employed to integrate the results of the individual windows into a final, unbiased, free energy profile (see Equation 14). The underlying theory of the method was derived and validated by Kumar et al. between 1992 and 1994.<sup>37,38</sup>

$$P^u(\xi) = \sum_i^{windows} N_i \exp[-\beta \omega_i(\xi) + \beta F_i] P_i^u(\xi) \quad (14)$$

In Equation 14, the unbiased distribution is obtained from the summation over all windows and the total number of steps  $N_i$  in each window. Equations 14 and 13 can be combined in an iterative manner to minimize the statistical error of the unbiased distribution to obtain final values for  $F_i$ .

## 1.4 Neural Network Potentials in Computational Chemistry

Neural networks are universal approximators of any smooth nonlinear function.<sup>39</sup> They can therefore find patterns in large data sets and are not only relevant in computational chemistry, but have advanced dramatically speech recognition, visual object recognition, and other forms of object recognition, which are part of our everyday lives.<sup>40</sup>

Artificial neural networks (ANNs) are composed of interconnected nodes organized in layers with parameters that must be learned. In the feed-forward pass, the input is propagated through all the layers, undergoing linear transformations on the output of the previous layer and finally applying a non-linear activation function. In a second step, the backpropagation updates the parameters according to the derivative of a loss function that calculates the discrepancy between the output and the reference data.<sup>27</sup> Given that NNs are universal approximators, given enough data, they can predict functional relationships regardless of the underlying physical principles. Furthermore, the use of high-quality data during the fitting process, combined with regularization, introduces generality that allows extrapolation to new data. These characteristics make them optimal candidates for use as force fields to predict potential energy surfaces, approximating the accuracy of *ab initio* methods and exhibiting the efficiency of classical force fields. To achieve this, chemical descriptors that encode the structure of a chemical system are used as input, while properties obtained from high-level *ab initio* methods are used as output.<sup>41</sup>

However, several challenges arise when using machine learning approaches to develop force fields. High-quality *ab initio* data are computationally demanding to generate, meaning that the ML models used must be both accurate and data efficient. Moreover, scalability to large system sizes remains problematic, as does transferability from one chemical system to another. In addition, the laws of physics introduce numerous symmetries that must be considered, such as energy conservation, rototranslational invariance of the potential energy, and the indistinguishability of identical atoms.<sup>41</sup>

A practical solution to address these challenges involves the use of high-dimensional neural network potentials (HDNNPs). These models calculate the total energy of a system as a sum of atomic contributions. Additionally, they employ symmetry functions to effectively capture and represent the local environment of individual atoms.<sup>42</sup> Several descriptor-based NNs have been developed.<sup>43,44</sup>

A major drawback of descriptor-based NNs is that the descriptors need to be chosen a priori, and they can be computationally expensive when the number of input dimensions increases. An alternative is to use graph neural network (GNN) models that extend the idea of neural networks and enable the direct processing of graph structured data. This is highly advantageous when the initial data is originally already structured like a graph, as in the case of molecules, where edges in the graph can be used to represent bonds and nodes can be used for atoms.<sup>45–48</sup>

#### 1.4.1 Anisotropic Message Passing - AMPv2 Model

In this study, the AMPv2 machine learning model is employed in the context of (QM)ML/MM simulations. The initial concept is described by M. Thürlemann and S. Riniker.<sup>49</sup> The code was developed by F. Pultar and M. Thürlemann, but it has not yet been published.

AMP, which stands for **A**nisotropic **M**essage **P**assing, adapts traditional message passing in a GNN to account for anisotropic features. The anisotropic features that have been incorporated are Cartesian multipoles, which are expressed as linear combinations of local frames. The combination of these two features includes directional information, thereby making the GNN equivariant under rotations and invariant for translations.

In the AMPv2 program, typical message passing is extended by two additional steps. The QM zone is represented as a graph  $G = (V, E)$ , where  $\mathbf{v}_i \in V$  denotes a set of nodes and  $\mathbf{e}_{ij} \in E$  the set of edges. The hidden feature vector  $\mathbf{h}_i^l \in \mathbb{R}^n$  of the node  $\mathbf{v}_i$  can be updated to  $\mathbf{h}_i^{l+1}$  for the next iteration step  $l + 1$  by means of the following equation:

$$\mathbf{h}_i^{l+1} = \phi_h(\mathbf{h}_i^l, \sum_{j \in N_{QM}(i)} \phi_e(\mathbf{h}_i^l, \mathbf{h}_j^l, \mathbf{a}_{ij})) \quad (15)$$

The functions  $\phi_e$  and  $\phi_h$  are learnable functions for edges and nodes. The index  $j$  runs over all neighbors of  $i$  in the QM zone,  $N_{QM}(i)$ , and  $\mathbf{a}_{ij}$  is an anisotropic

feature defined between two nodes. This equation has the same form as message passing in traditional GNNs except that  $\mathbf{a}_{ij}$  is used instead of an unmodified edge feature typically employed in this context.<sup>45,46</sup> For the anisotropic feature  $\mathbf{a}_{ij}$  and its treatment in the message passing step, two additional steps must be added, as expressed by the following equations.

$$\mathbf{M}_i^k = \sum_{j \in N_{QM}(i)} \phi_{MQM(k)}(\mathbf{h}_i^l, \mathbf{h}_j^l, \mathbf{a}_{ij}) \mathbf{R}_{ij}^k + \sum_{j \in N_{MM}(i)} \phi_{MMM(k)}(\mathbf{h}_i^l, \mathbf{a}_{ij}) \mathbf{R}_{ij}^k \quad (16)$$

$$\mathbf{a}_{ij} = \phi_b(\|\mathbf{r}_{ij}\|) \circ G_{ij}(\hat{r}_{ij}, \mathbf{M}_i, \mathbf{M}_j) \quad (17)$$

The Equation 16 is employed to derive  $\mathbf{M}_i^k$ , the Cartesian multipoles up to order  $k$ . The first term of the equation represents the expansion of the multipole for each node  $i$ . This term uses a linear combination of a local basis, which is the traceless tensor product of order  $k$  of the unit vector  $\hat{r}_{ij}$  connecting the nodes  $i$  and  $j$ .<sup>50</sup>

$$\mathbf{R}_{ij}^k = \hat{r}_{ij} \otimes \hat{r}_{ij} \otimes \dots \quad (18)$$

The learnable function, denoted by  $\phi_{MQM(k)}$ , is responsible for predicting the coefficients of the expansion. The second term accounts for the electrostatic interaction of the MM particles with the QM particles, which is derived from Equation 8. It should be noted that the MM particles are not part of the graph, but possess a partial charge,  $q_i$ . The updating function, denoted by  $\phi_{MMM(k)}$ , takes as input the hidden feature of the QM atom  $\mathbf{h}_i^l$  and the anisotropic feature  $\mathbf{a}_{ij}$  defined in terms of the multipole of the QM atom and the partial charge of the MM atom.

Equation 17 defines the contraction of the multipoles between nodes. The interaction between multipoles is defined by the product of radial components, represented by  $\phi_b(\|\mathbf{r}_{ij}\|)$  and  $G_{ij}(\hat{r}_{ij}, \mathbf{M}_i, \mathbf{M}_j)$ , concatenated multipole interaction coefficients for all orders  $k$ . The multipole interaction formalism is used from C. Burnham and N. English.<sup>51</sup>

The forces are then the negative gradient of the potential energy  $E_{\text{pot}}$ , which is given by:

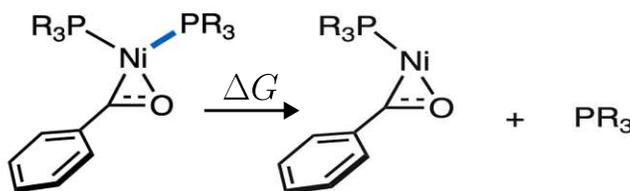
$$E_{\text{pot}} = \sum_i^{QM} \phi_V(\mathbf{h}_i^n) + \sum_i^{QM} \sum_j^{MM} \frac{\phi_\rho(\mathbf{h}_i^n) \cdot q_j}{\|\mathbf{r}_{ij}\|} \quad (19)$$

The first term represents the contribution of each atom in the quantum mechanical zone to the potential energy. The second term represents the Coulomb interaction,

where  $\phi_\rho(\mathbf{h}_i^n)$  is the charge density for the QM atoms.  $n$  is the total number of graph layers.

Consequently, molecular multipoles are obtained by the multipole formalism. When used embedded in MD simulations, forces are required which are automatically obtained by the backpropagation step.

## 1.5 Aim of This Work



**Figure 5: Reaction for ligation state determination.** Primary aim of this work is finding the free energy value  $\Delta G$  between bisligated  $L_2Ni(\text{benzaldehyde})$  complex and  $L_1Ni(\text{benzaldehyde})$  monoligated complex. The ligand  $L$  are different phosphine ligands. Image modified from S. Newman-Stonebraker et al.<sup>1</sup>

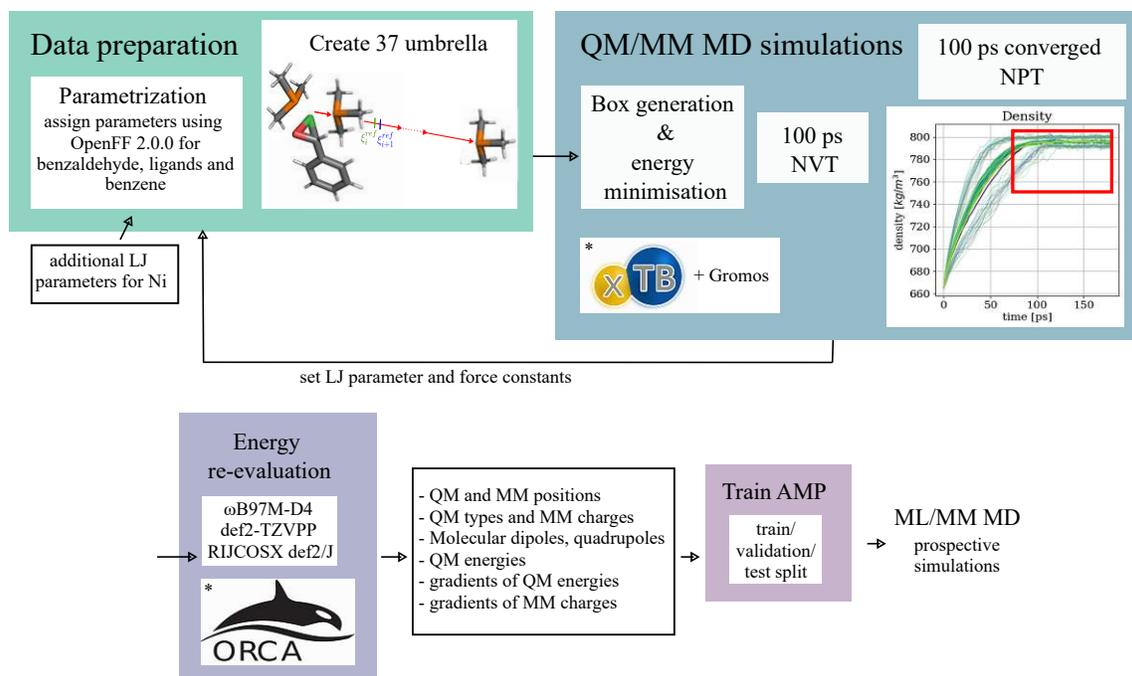
The aim of this work was to investigate ligation states of phosphine ligands  $L$  in  $L_xNi(\text{benzaldehyde})$  complexes by the use of the AMPv2 model in a (QM)ML/MM simulations with electrostatic embedding. The reaction is presented in Figure 5.

The principal scientific questions that must be addressed along the way are as follows:

- Are there any difficulties encountered during the process of generating the training data? Could the non-standard Lennard-Jones parameter for nickel be considered a problem in the generation of the training data?
- How accurate is the electronic energy predicted by the AMPv2 model in comparison to DFT and semi-empirical methods?
- How do the errors for different properties behave if the training/validation/test split is changed?
- How is the transferability for unseen data? How do the errors behave if data pertaining to specific complexes is entirely excluded from the training process?

- Can prospective simulations using the AMPv2 model in ML/MM simulations be performed? Are those stable and ensure that no extreme forces are predicted?
- Is it possible to model a real-world problem and reproduce experimental data? Are the values of free energy calculations obtained by umbrella sampling comparable to experimental results?

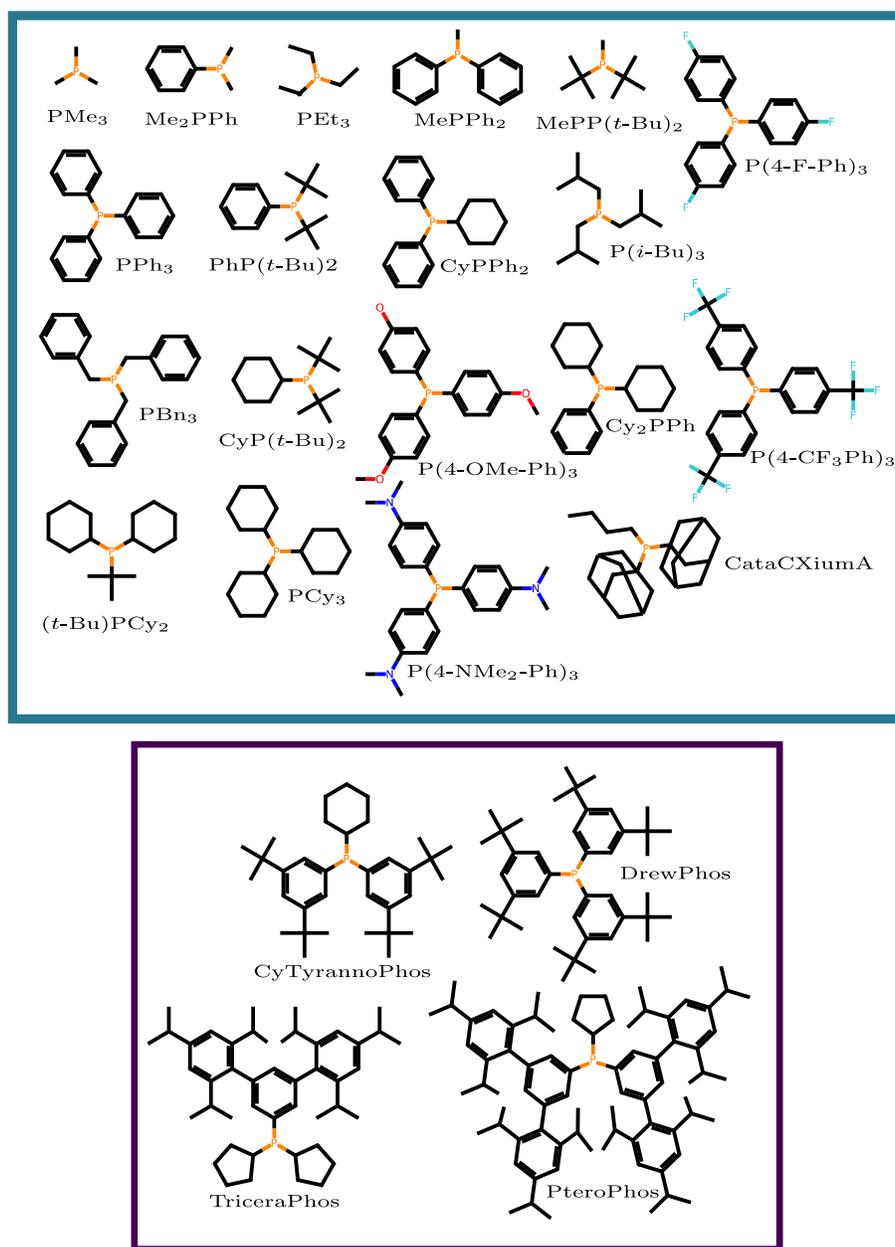
## 2 Computational Details



**Figure 6: Schematic representation of the principal stages in the generation of training data.** For performing ML/MM MD simulations, where the QM computation is successfully replaced by an ML model, four main steps must be completed beforehand: The data preparation step (parametrization and coordinate file generation for all 37 umbrellas), the QM/MM MD simulations (generation of simulation box, NVT and NPT ensemble simulations), the energy re-evaluation (using the  $\omega$ B97M-D4 DFT range-separated functional and def2-TZVPP basis set) and the training of the AMP model (performing a train/validation/test split beforehand). Some images were included from external sources and subsequently modified, marked with \*.<sup>29,52</sup>

The final objective was to investigate the behavior of  $L_x$ Ni(benzaldehyde) complexes, where  $x=1$  represents the monoligated complex and  $x=2$  the bisligated complex. For this purpose, the training data was collected from 37 MD simulations from the umbrella sampling in which the ligand L was gradually displaced.

Figure 6 depicts the computational setup. First, the parametrization of the compounds and the generation of the initial coordinate files was performed. The subsequent stage was a restrained QM/MM MD simulation including the generation of



**Figure 7:** Ligands  $L$  used in  $L_xNi(\text{benzaldehyde})$  complexes with  $x=1,2$  (mono-ligated vs bisligated complex). All ligands used in training data are in upper box (petrol-colored box), large ligands, which were computationally too expensive, are in lower box; Bn, benzyl; Ph, phenyl; *t*-Bu, *tert*-butyl; *i*-Bu, isobutyl.

the simulation box, NVT and NPT ensemble simulations. In this step, the values of the Lennard-Jones parameters as well as force constants for the restraints were determined. The energy re-evaluation provides the input for the training data for the AMPv2 model, which can then be employed for prospective simulations once the training process is completed.

## 2.1 Data Preparation

The objective of the data preparation was to generate xyz coordinate files for each complex version, with variations applied to the ligand L, and the topology files, which encode the complex parametrization. For benzene, the solvent, the parametrization was stored for the creation of the simulation box in the subsequent stages.

For the complexes, DFT-preoptimized structures from Newman-Stonebraker et al. were used.<sup>1</sup> All ligands for which input structures were available are shown in Figure 7. The parametrization was performed using a package developed in the Riniker group, qmpeg. The coordinate files were generated using RDKit 2023.09.1.<sup>53</sup> The parametrization was performed using the OpenFF 2.0.0 force field, with the help of certain functionalities provided by OpenMM 8.0.0 and openmmforcefields 0.11.2.<sup>54,55</sup> The topology files were saved in AMBER format.<sup>56</sup> The files were then converted to Gromos topology files using `amber2gromos` from Gromos++, and some functionality from ParmEd 4.2.2. and PyGromosTools 3.0.<sup>57-59</sup> For nickel, an alternative route was required, by manually setting the Lennard-Jones parameters to literature values.

Subsequently, topology and coordinate files were combined to create the final complexes. For the addition of the solvent, either 3 000 or 4 000 benzene molecules were added.<sup>1</sup> A total of 37 initial xyz files were generated for the umbrella sampling. One ligand was fixed in position and the other was gradually shifted along the axis established between the nickel atom and the shifted ligand. The ligand was initially positioned at a distance of 2 Å from the nickel atom and then moved incrementally until it was 20 Å away. Each movement measured 0.5 Å.

---

<sup>1</sup>4 000 benzene molecules were only necessary for the complex with PteroPhos, as a larger simulation box was required due to its size.

## 2.2 QM/MM MD Runs

### 2.2.1 Training Data Generation

All molecular dynamics (MD) simulations were conducted using the MD++ program from the GROMOS software package (custom Riniker group branch) interfaced with GFN2-xTB/6.5.1.<sup>29,60,61</sup> K. Meier et al. were the first to interface the GROMOS software to quantum-chemical program packages MNDO and TURBOMOLE.<sup>60</sup> This was extended to other external QM programs such as xTB or ORCA in the Riniker group. Additional features were implemented as well by F. Pultar, including printing of the QM zone, necessary for the energy re-evaluation.

For the QM/MM scheme, electrostatic embedding was chosen, with a cutoff for the inclusion of MM atoms in QM calculations set to 1.4 nm. The classical potential energy surface is evaluated using the GROMOS software, similar to Equation 5. Harmonic bond-stretching and angle-bending terms are used as recommended for topologies derived from AMBER.<sup>62</sup> Non-bonded van der Waals interactions were evaluated by means of the 12-6 Lennard-Jones potential, while electrostatic interactions are approximated by Coulomb's law within a cutoff radius of 1.4 nm. Beyond this cutoff, long-range electrostatic contributions are represented by a Poisson-Boltzmann generalized reaction field.<sup>63</sup> The dielectric constant of benzene was set to 2.27.<sup>64</sup> The electrostatic 1-4 interactions were scaled by a factor of  $\frac{1}{1.2}$ , accounting for the Amber topology, which was converted to GROMOS files.<sup>62</sup>

According to the principle of electrostatic embedding, expressed by equations 6, 7 and 8, the classical potential energy expression is extended by an additional term  $E_{QM-MM}^{vdW}$ . Because of this term, Lennard-Jones parameters are needed for all QM atoms. Also, the Hamiltonian for the QM zone (accounting for kinetic energy of electrons and interaction between electrons and nuclei of QM atoms), is extended by  $\hat{H}_{QM-MM}^{el}$ , according to the electrostatic embedding. The term  $E_{QM-MM}^b$ , which is normally treated by link atom schemes, is not used as no bonds between QM and MM atoms exist.

The xTB program used was GFN2-xTB/6.5.1.<sup>29</sup> The default accuracy settings were employed, with the maximum number of iterations in the self-consistent cycle (SCC) set to 250 and an accuracy multiplier set to 1. This determined the convergence criteria and the accuracy in the integral calculation. Fermi-smearing was used, with an electronic temperature set to 1000 K, for improved convergence. All complexes were assigned a charge of 0 and a spin multiplicity of 1.

A total of 37 umbrella restrained simulations were carried out for the training data generation. These were repeated for three different force constants used for the restraints. For all simulations two distance restraints were applied to the distances between each of the phosphorus atoms of the ligand and the nickel atom. In the MD++ program, a fully harmonic attractive potential of form of Equation 9 was selected. The reference position in Å is set to values of  $\xi_i^{ref} = \{2, 2.5, 3, 3.5, \dots, 20\}$  for the individual umbrellas  $i$ . The force constant  $K$  for the biasing potential was set to values of 2 000, 20 000, and 100 000 kJ/(mol nm<sup>2</sup>) for the different MD runs. These values were determined in Section 3.2.

For each umbrella and force constant, a MD simulations was performed. The simulation time step was set to 0.5 fs and cubic periodic boundary conditions were applied. Each MD simulation consisted of three steps: energy minimization, equilibration and production run. First, the GROMOS++ function `ran_box` is employed to randomly arrange the solute molecules within the simulation box.<sup>57</sup> The box size is determined by the density, since the number of solvent molecules is fixed. The generated simulation box was subjected to 200 steps of energy minimization via a gradient descent method. Subsequently, an equilibration run being a NVT (constant number of particles, volume and temperature) ensemble simulation was conducted for a total of 100 ps. Stochastic dynamics (SD) was used, acting as a Langevin thermostat, maintaining a constant temperature of  $T_{ref} = 350$  K. The friction coefficient in Equations 3 and 4 is set to  $\gamma = 1$ . Initial velocities were generated from a Maxwell distribution at 350 K. This temperature was chosen because it is just below the boiling point of benzene (353.2 K).<sup>65</sup> The high temperature results in out-of-equilibrium structures and ensures wider sampling by overcoming more local energy barriers. Finally, another 180 ps NPT ensemble simulation was performed, maintaining a constant pressure rather than volume. In addition to the SD temperature control, a Berendsen barostat was incorporated for constant pressure.<sup>66</sup> The weak coupling to the barostat is performed at a frequency of 0.5 ps, a reference pressure of 1 atm and an isothermal compressibility of  $4.575 \times 10^{-4}$  kJ/(mol nm<sup>3</sup>). Isotropic pressure scaling was employed, adjusting the box lengths to the same extent. The final 100 ps were utilised for subsequent post-processing with the final objective to be used as training data. The GROMOS program was used with enabled MPI parallelization using two MPI cores.

The analysis package GROMOS++ was employed for some preprocessing steps and, in particular, for the analysis of the trajectories.<sup>57</sup>

### 2.2.2 Lennard-Jones Parameters and Force Constants

Before producing the training data, Lennard-Jones (LJ) parameters and values for the force constants for the distance restraints needed to be found. MD simulations performed for both experiments were similar to those described in the previous section, Section 2.2.1.

For the evaluation of the LJ parameters, literature values were chosen to be used in the MD simulation. The results can be found in Section 3.1, where the literature values are shown in Table 5. These MD simulations were performed for the bisligated  $L_2Ni(\text{benzaldehyde})$  and monoligated  $L_1Ni(\text{benzaldehyde})$  complex for two ligands  $L$ , which were  $PMe_3$  and  $PEt_3$ . The simulation temperature was set to 298 K and the equilibration run time was reduced to 80 ps.

For the second experiment, the determination of the force constant, 100 ps of an equilibration run was performed. All 37 umbrellas were simulated for complexes with the ligands  $PEt_3$  and  $PhP(t-Bu)_2$ . The force constants tested were 2 000, 20 000, 40 000, 80 000, and 100 000  $\text{kJ}/(\text{mol nm}^2)$ . The results can be found in section 3.2.

### 2.2.3 Energy Re-Evaluation

To train the AMPv2 model, frames including total energies, forces for QM and MM atoms, as well as dipoles and quadrupoles calculated at DFT level of theory are required. This procedure will be referred to as energy re-evaluation. 50 DFT energy re-evaluations were performed for each MD simulation using snapshots of the last 100 ps of the NPT ensemble, as explained in Section 2.2.1. For each complex a total number of 5 550 energy re-evaluations were performed, resulting from multiplying the 50 calculations per MD run by 37 umbrellas and 3 force constants.

Single point energy and gradient calculations were performed using ORCA/5.0.4.<sup>52</sup> The range-separated hybrid functional  $\omega B97M-D4$ , a modified version of the  $\omega B97X-V$  functional with DFT-D4 correction by Najibi and Goerigk, was used.<sup>67,68</sup> This functional was chosen due to its demonstrated efficacy in several benchmark studies for transition metals.<sup>69-71</sup> The functional also includes exact exchange for long-range interactions, which is considered important. Atom-pairwise dispersion correction, D4, is used.<sup>72,73</sup> The orbital basis set employed was a triple zeta valence basis augmented by polarization functions (def2-TZVPP), while an auxiliary basis set def2/J was also employed, both of which were designed by F. Weigend and A. Reinhard.<sup>74,75</sup> Further RIJCOSX was added for speed-up, which introduces approximations for

Coulomb integrals and for the numerical integration for HF exchange.<sup>76</sup> The convergence criterion was set to TightSCF and the integration grid was defgrid2. All point charges were included in an electrostatic embedding scheme in the QM calculations.

## 2.3 AMP Training and Testing

The models were trained using the AMPv2 program (code not published yet), developed by M. Thürleman and F. Pultar. Its theory and architecture of which were described in the introduction, in Section 1.4.1. AMPv2 models were trained with PyTorch 2.0.1.<sup>2</sup>

The training data for the AMPv2 models consisted of the re-evaluated results of the DFT single point calculations, as well as the coordinates of QM and MM particles, the charges of the MM zone, and the atomic numbers of the QM atoms. The latter were extracted from the MD trajectories. For each complex, results were shuffled and then saved collectively in batches of 8 frames. This was done because reduced accuracy was previously reported for unshuffled data, showing the model’s limited generalizability.<sup>49</sup>

The Adam optimizer was used for the training process to update the weights of the neural network (NN).<sup>77</sup> For all models, an initial learning rate of  $3 \times 10^{-4}$  was passed to the Adam optimizer. Additionally, an exponential learning rate scheduler was employed, with a decay rate  $r_{\text{decay}}$  that depended on the total number of epochs  $N_{\text{epochs}}$ .

$$r_{\text{decay}} = \exp\left(\frac{\log(f_{\text{decay}})}{N_{\text{epochs}}}\right) \quad (20)$$

$f_{\text{decay}}$  was fixed to 0.02 for all models. A lower number of epochs thus resulted in a faster decay of the global learning rate.

The loss function  $L$  comprises the mean square error (MSE) of the energy  $E$ , the gradients  $F$  of QM and MM particles, and contributions from dipole  $M^1$  and quadrupole  $M^2$ .  $N_{QM}$  is the total number of QM atoms, and  $N_{MM}$  is the total number of MM atoms.

<sup>2</sup>Other packages used: torchvision 0.15.2, torchaudio 2.0.2, pytorch-cuda 11.8, torch-scatter 2.1.2, torchmetrics 1.3.2, psutil 5.9.8 and torchlayers 0.1.1

$$L = (1-\alpha)(E-\tilde{E})^2 + \alpha \frac{1}{3N_{QM}} \sum_i^{N_{QM}} \sum_\alpha^3 (F_{i\alpha} - \tilde{F}_{i\alpha})^2 + \alpha\beta \frac{1}{3N_{MM}} \sum_i^{N_{MM}} \sum_\alpha^3 (F_{i\alpha} - \tilde{F}_{i\alpha})^2 + \gamma \frac{1}{3} \sum_\alpha^3 (M_\alpha^1 - \tilde{M}_\alpha^1)^2 + \gamma \frac{1}{6} \sum_\alpha^3 \sum_{\beta=\alpha}^3 (M_{\alpha\beta}^2 - \tilde{M}_{\alpha\beta}^2)^2 \quad (21)$$

The scaling factors are set to the same values for each training process being  $\alpha = 0.99$ ,  $\beta = 100$ , and  $\gamma = 100$ . Note that for the energies, the relative energy is learned for each batch, resulting in a value for the implicit energy offset. The default type for these tensors is set to float32. The loss is backpropagated and used to update all weights. For regularization, early stopping and auxiliary learning tasks are used. The training process is stopped as soon as the error, similar to expression 21 but without factor  $\beta$ , is no longer decreasing on the validation data set.

**Table 1: Main differences between minimal and precision model.** Total number of message passing steps, kernels for encoding the distance between QM atoms, channels for encoding the multipoles for each atom and number of trainable parameters of the model. The cutoff for the polarization of QM and MM atoms is given in Å.

properties/model	minimal	precision
# message passing steps	2	3
# kernels (distance between QM atoms)	8	20
# channels (multipoles/atom)	32	64
cutoff MM-QM polarization [Å]	9	10
# parameters	606 568	2 683 392

Two different models were trained, which will be referred to as the minimal and precision models in all subsequent sections. The main differences between these models are shown in Table 1. The minimal model encodes various properties with a lower degree of precision, resulting in a reduction of trainable parameters by a factor of four. In the AMPv2 model, three different cutoffs need to be set. The first is the graph cutoff for the atoms in the QM zone, which is set to 5 Å for all models. The second is the long-range cutoff, which determines the extent of the polarization introduced for QM-MM interactions. This cutoff has different values for minimal and precision models (see Table 1). The last cutoff defines the QM-MM pairs for the Coulomb interaction in Equation 19, where all possible pairs are evaluated. The

number of Bessel functions used for encoding distances must be chosen. For all distances between QM and MM atoms, this value is always set to 8, while for the distances between QM atoms, a value of 20 is chosen for the precision model (see Table 1). The total number of message passing steps and channels used to encode the multipoles is also variable according to Table 1. The feature vector has a size of 128 for the nodes in the graph.

For the minimal and precision model, a total amount of 15 different model versions were trained. Each of these used a different composition of the training, validation, and test data. The details of how many frames are used for training, validation, and testing are shown in Table 2. In this table, the terms data set 1, 2, and 3 are used denoting subsets of the entire training data. These subsets are explained in Table 3.

A minimal hyperparameter tuning was performed. For each model version, both minimal and precision models, the training was performed three times, only changing the total number of epochs used. The number of epochs used for each training are shown in Table 4. The number of epochs multiplied by the total amount of training data remained approximately constant. According to Equation 20, a lower number of epochs results in a faster decay of the learning rate. For these three runs, the model with the lowest MAE on the QM gradients for the validation data set was selected as the final model. Consequently, in the following chapters, results will be shown for all model versions listed in Table 2 as minimal and precision models.

## 2.4 Prospective Simulations

For the prospective simulations, the previously trained PyTorch models were exported and used in the GROMOS software package (custom Riniker group branch). The interface utilises the coordinates of QM and MM particles, as well as the charges of the MM zone and atomic numbers of QM atoms to predict the energy by a forward passing step. The autograd functionality in PyTorch provides the gradients that are then used to propagate the system in the MD run. This setup allows the PyTorch model to perform all the tasks previously performed by the xTB method. PyTorch was parallelized using CUDA to run on GPUs, while the GROMOS computations were performed on CPUs using MPI.

A preliminary experiment was conducted to determine the stability, in terms of dissociation, of all monoligated  $L_1Ni(\text{benzaldehyde})$  and bisligated  $L_2Ni(\text{benzaldehyde})$  complexes. For this purpose, all ligands employed in the training data generation

**Table 2: Composition of train/validation/test set of all model versions.** Total number (#) and percentages (%) of frames used in training, validation and test sets listed for each model version. Data used from three different data sets (data sets 1, 2 and 3) is explained in more detail in Table 3.

version	data set	train	validation	test	train	validation	test
		[# frames]	[# frames]	[# frames]	[%]	[%]	[%]
1	1	84 360	1 976	1 976	80.0	1.9	1.9
2	1	42 104	1 976	1 976	39.9	1.9	1.9
3	1	2 0976	1 976	1 976	19.9	1.9	1.9
4	1	10 488	1 976	1 976	9.9	1.9	1.9
5	2	33 240	2 400	0	39.9	2.9	0.0
	3	0	0	17 760	0.0	0.0	80.0
6	2	33 240	2 400	0	39.9	2.9	0.0
	3	32	32	17 760	0.1	0.1	80.0
7	2	33 240	2 400	0	39.9	2.9	0.0
	3	64	64	17 760	0.3	0.3	80.0
8	2	33 240	2 400	0	39.9	2.9	0.0
	3	416	416	17 760	1.9	1.9	80.0
9	2	33 240	2 400	0	39.9	2.9	0.0
	3	896	896	17 760	4.0	4.0	80.0
10	1	102 144	3 040	0	96.9	2.9	0.0
11	2	80 640	2 400	0	96.9	2.9	0.0
	3	0	0	17 760	0.0	0.0	80.0
12	2	80 640	2 400	0	96.9	2.9	0.0
	3	96	96	17 760	0.4	0.4	80.0
13	2	80 640	2 400	0	96.9	2.9	0.0
	3	192	192	17 760	0.9	0.9	80.0
14	2	80 640	2 400	0	96.9	2.9	0.0
	3	1 088	1 088	17 760	4.9	4.9	80.0
15	2	80 640	2 400	0	96.9	2.9	0.0
	3	2 208	2 208	17 760	9.9	9.9	80.0

process, including PteroPhos were simulated using an unrestrained ML/MM MD simulation. The experiment was conducted twice, using model version 2, trained with 40% of all data, and model version 5, trained with the same data but excluding data from four ligands (see Table 2). Both model versions were minimal models.

**Table 3: Composition of data sets 1, 2, and 3.** Description of composition of dataset 1, dataset 2, and dataset 3 with respect to complexes included and total number of frames. Dataset 1 denotes all data available, while datasets 2 and 3 are subsets of dataset 1.

data set 1	data from complexes using all 19 ligands (105 450 frames)
data set 2	data from complexes using 14 ligands (excluded were $\text{PEt}_3$ , $\text{PhP}(t\text{-Bu})_2$ , $\text{PCy}_3$ and $\text{CataCXiumA}$ ) (83 250 frames)
data set 3	data from complexes using 4 ligands, $\text{PEt}_3$ , $\text{PhP}(t\text{-Bu})_2$ , $\text{PCy}_3$ and $\text{CataCXiumA}$ (22 200 frames)

**Table 4: Details for hyperparameter tuning.** For each model version the total number of epoch is displayed used for training.

model versions	# epochs
1, 11 - 15	32, 64 and 128
2, 5 - 9	64, 128 and 256
3	128, 256 and 512
4	256, 512 and 1024

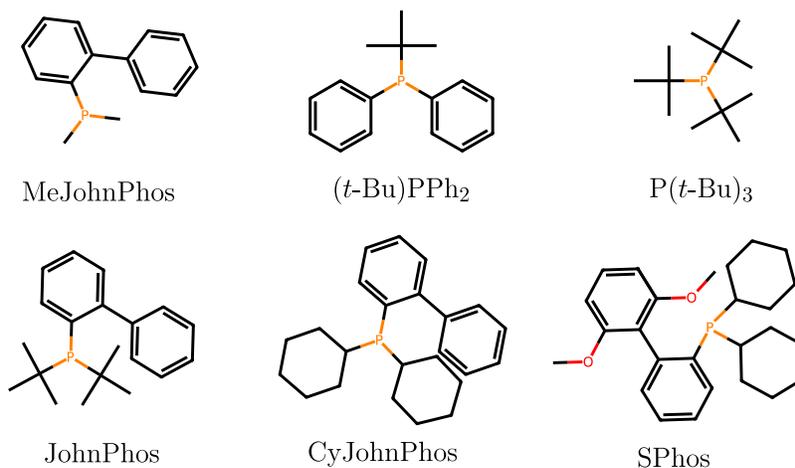
For complexes that did not dissociate as monoligated and bisligated complexes, the energetic minima are unclear. Umbrella sampling can be performed to generate free energy profiles to resolve this issue. This was performed for two complexes, those with ligands PteroPhos and  $\text{PCy}_3$ . 37 individual ML/MM MD runs were performed in which one ligand was moved stepwise 0.5 Å away from the nickel atom, commencing at 2.0 Å. Only the ligand that was moved was restrained with a fully harmonic potential, using a force constant of 2 000 kJ/(mol nm<sup>2</sup>). The runs were re-weighted using the WHAM (weighted histogram analysis method) program, version 2.0.10 by A. Grossfield, to generate free energy curves.<sup>78</sup> The method requires sufficient sampling in each window. This was not achieved using a force constant of 2 000 kJ/(mol nm<sup>2</sup>) in the range of 2.5 Å to 5.5 Å because the restraint was too weak for yielding the target geometries. Further sampling was added by performing another 10 to 11 MD runs with a force constant of 10 000 kJ/(mol nm<sup>2</sup>). The ML/MM MD runs for the umbrella sampling were performed using minimal and precision model, model version 10, for the PteroPhos complex. For the complex with the  $\text{PCy}_3$  ligand, minimal models were employed, utilizing two distinct versions, namely 5 and 8.

For all ligands, the final frame from the NPT run of the training data generation process was used as the initial starting configuration. This approach eliminated the

need for a box minimization and equilibration run. For PteroPhos, however, a newly generated box was required, and the equilibration run was omitted. All simulations were performed at 298 K.

## 2.5 QM/MM MD Simulation for $L_xNi(4\text{-fluorobenzaldehyde})$ Complexes

QM/MM MD simulations were also performed for  $L_xNi(4\text{-fluorobenzaldehyde})$ , where the benzaldehyde was substituted by 4-fluorobenzaldehyde. The same 19 ligands from the previous sections (see Figure 7) were used, as well as another six additional ligands that were handcrafted (contributed by Igor Gordiy). These ligands were constructed modifying the xyz file of the other ligands and then performing an energy minimization using xTB, while fixing the hapto-2 configuration of the 4-fluorobenzaldehyde with respect to the nickel. The additional ligands are shown in Figure 8. The only differences in the MD simulations are that now 4 distance restraints are applied: between the Ni and P atoms for both phosphine ligands, as well as the O and the C-atoms of 4-fluorobenzaldehyde, which are both coordinated to the nickel atom.



**Figure 8:** Additional ligands  $L$  used in  $L_xNi(4\text{-fluorobenzaldehyde})$  complexes with  $x=1,2$  (monoligated vs bisligated complex). Additional structures were handcrafted;  $L$ , ligand.

## 3 Results and Discussion

### 3.1 Effect of LJ Parameters

The effect of three different sets of Lennard-Jones (LJ) parameters from the literature was analyzed by examining the trajectories obtained when using these parameters in a QM/MM MD simulation. The values obtained from three different literature sources are shown in Table 5 (denoted by Ni: LJ case  $x$ ,  $x = 1, 2, 3$ ). The values were provided for a 12-6 LJ functional form using either the parameter  $\sigma$ , which is the distance where the energy is zero, or as  $r_{\min}$ , which is the distance where the minimum of the potential energy is located, and  $\epsilon$ , the well depth of the potential.  $\sigma$  and  $r_{\min}$  are connected by the following relationship:

$$\sigma = \frac{r_{\min}}{2^{1/6}} \quad (22)$$

The LJ potential for the different literature values used is plotted in Figure 9. The LJ parameters are converted internally to the parameters  $A$  and  $B$ , which are also displayed in Table 5 and can be used to express the potential energy  $E$  in the 12-6 LJ functional form as follows:

$$E(r) = \frac{A}{r^{12}} - \frac{B}{r^6} \quad (23)$$

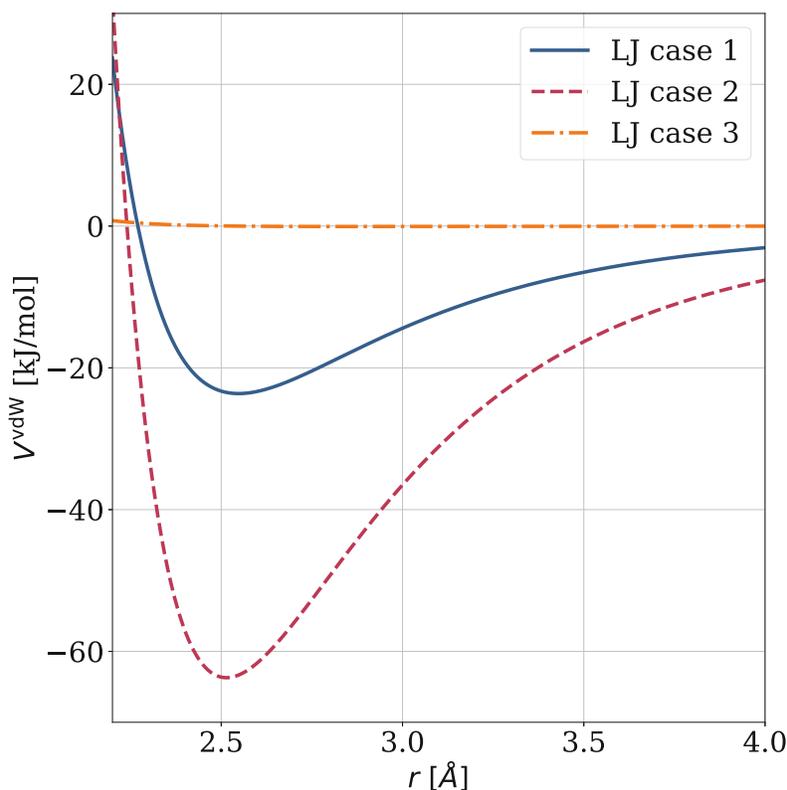
The repulsion coefficient  $A$  and the dispersion coefficient  $B$  can be converted using the equations  $\sigma = \left(\frac{A}{B}\right)^{\frac{1}{6}}$  and  $\epsilon = \frac{B^2}{4A}$ . All values in Table 5 were used in the MD runs. While the values for nickel were collected from literature, the values for hydrogen and carbon originate from the OpenFF 2.0.0 force field. Finally, van der Waals radii for all elements were stated using literature values as well. Van der Waals radii for several main group elements and some metals were published by A. Bondi in 1964, based on experimental data from X-ray diffraction, corrected to be consistent with gas kinetic collision cross sections, critical densities, and liquid state properties.<sup>79</sup> The values for the main group elements were confirmed by the use of additional experimental data or alternative techniques, including computational methods.<sup>80,81</sup> For hydrogen, experiments demonstrated that the value proposed by A. Bondi was overestimated by 0.1 Å.<sup>80</sup> The sum of van der Waals radii is frequently employed to ascertain interatomic distances. However, it has been observed in literature that these often underestimate the value of the potential energy curve.<sup>82</sup>

**Table 5: Lennard-Jones parameters and van der Waals radii for atoms H, C, and Ni.** Van der Waals radii  $r_{vdw}$ , potential well depth  $\epsilon$ , the distance where the potential is zero  $\sigma$ , the repulsion coefficient  $A$ , and the dispersion coefficient  $B$ . For nickel, values from three different literature sources are given.

element	$r_{vdw}$ [Å]	$\epsilon$ [kcal/mol]	$\sigma$ [Å]	$r_{min}$ [Å]	$A$ [(kcal Å <sup>12</sup> )/mol]	$B$ [(kcal Å <sup>6</sup> )/mol]
H	1.10 <sup>80</sup>	0.016	2.57	2.89	$5.247 \times 10^3$	$1.810 \times 10^1$
C	1.70 <sup>79</sup>	0.087	3.48	3.91	$1.099 \times 10^6$	$6.179 \times 10^2$
Ni: LJ case 1	1.63 <sup>79</sup>	5.65 <sup>83</sup>	2.27	2.55 <sup>83</sup>	$4.229 \times 10^5$	$3.091 \times 10^3$
Ni: LJ case 2	1.63 <sup>79</sup>	15.23 <sup>84</sup>	2.24 <sup>84</sup>	2.51	$9.720 \times 10^5$	$7.694 \times 10^3$
Ni: LJ case 3	1.63 <sup>79</sup>	0.015 <sup>85</sup>	2.53	2.84 <sup>85</sup>	$4.123 \times 10^3$	$1.572 \times 10^1$

To determine which LJ parameters should be employed in the training data generation process, three factors were evaluated: (a) the stability of the structures in the MD simulation, (b) the presence of a peak resulting from the nickel-solvent interaction in the radial distribution function (RDF), and (c) the positioning of this peak with respect to the sum of van der Waals radii or the mean of the  $r_{min}$  values. The MD simulations were conducted for the monoligated and bisligated complexes with the ligands PMe<sub>3</sub> and PEt<sub>3</sub>, which are among the smallest ligands. The LJ parameters of nickel influence the term  $E_{QM-MM}^{vdW}$  in Equation 7, which determines the van der Waals interactions between QM and MM zone. Therefore, it can be expected that the tested LJ parameters will have the most pronounced effect for complexes where the solvent can get closer to the nickel atom, which is true for monoligated complexes and typically for smaller ligands. Furthermore, the RDF of benzene with respect to nickel was chosen as a decision criterion because it describes the density of the benzene atoms with respect to nickel, which is influenced by long-range interactions.

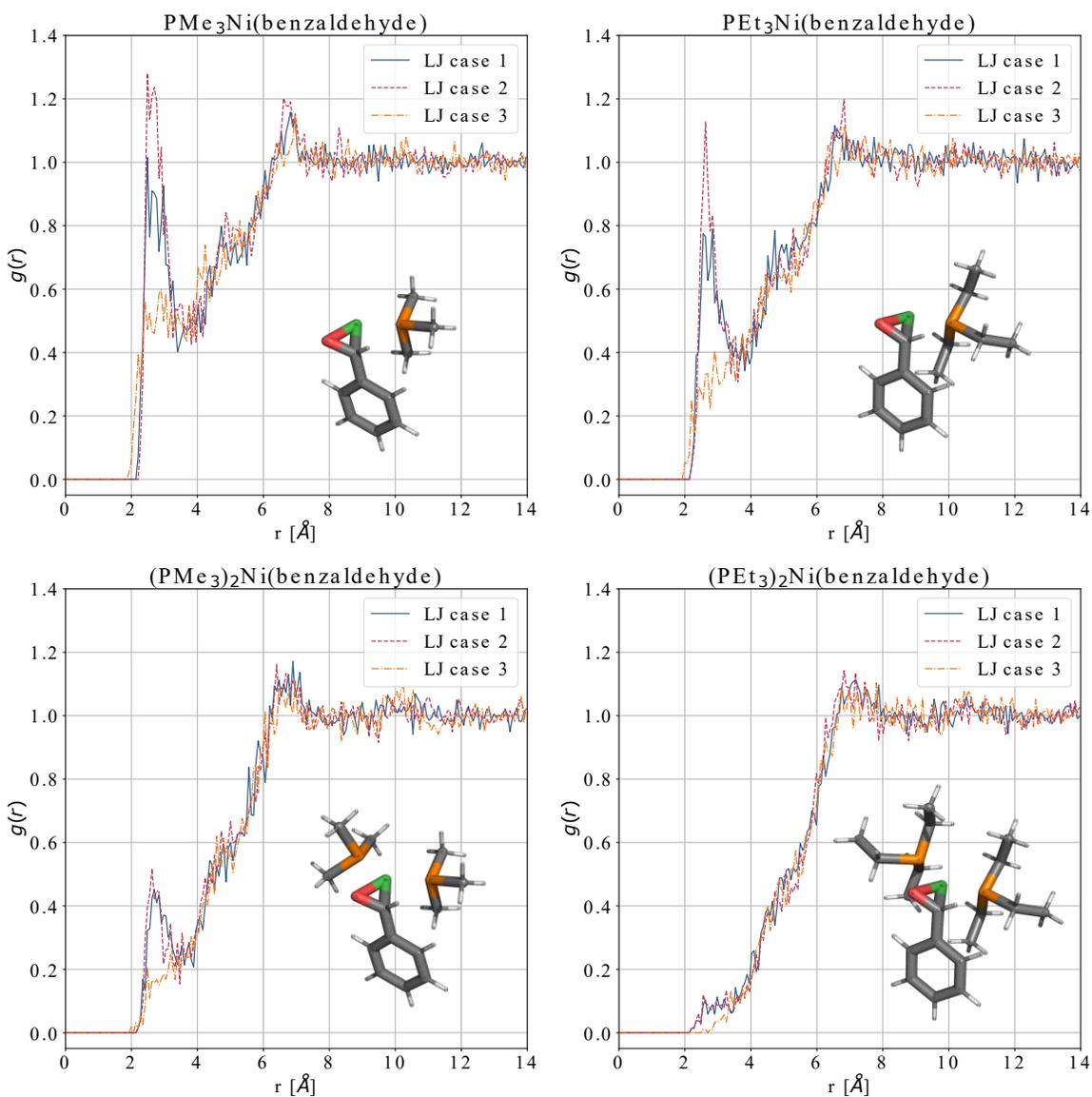
For (a) the first 100 ps of the production run were discarded and the subsequent 100 ps were used for the analysis. In the second half of the trajectory, the energy was converged, as was the temperature ( $293.3 \pm 1.7$  K), the density ( $863.3 \pm 1.8$  kg/m<sup>3</sup>, exp. 876 kg/m<sup>3</sup><sup>65</sup>), and the box volume ( $451.5 \pm 0.9$  nm<sup>3</sup>). The complexes were stable throughout the simulation, meaning that the ligands remained intact and coordinated to the nickel atom. The latter can be seen by evaluating the distances between nickel and the atoms coordinated to nickel (see supplementary information, Table S1). Furthermore, it is evident that both phosphorus atoms undergo a maxi-



**Figure 9: LJ potential for three LJ parameter sets tested for nickel.** Functional form of LJ potentials used in MD runs applying parameters from literature: LJ case 1<sup>83</sup>, LJ case 2<sup>84</sup>, and LJ case 3<sup>85</sup>; LJ, Lennard Jones.

mum change in distance to the nickel atom of 0.5 Å. In contrast, the carbon atoms exhibit a relative positional change of up to 1.38 Å. For larger distances between carbon and nickel, the benzaldehyde transitions to a hapto-1 state where the carbon atom is no longer coordinated to the nickel atom.

For (b) and (c), the RDF was plotted in Figure 10 choosing the hydrogen atom as reference for the benzene molecules. In the supplementary information, the same plot is visible using the carbon atom as reference for the benzene molecule (see supplementary information, Figure S1). The sum of the van der Waals radii of nickel and hydrogen would be expected to yield a peak close to 2.73 Å (for nickel and carbon this value corresponds to 3.33 Å). To get a rough estimate of where the first peak should be placed, it is also possible to use the arithmetic mean of



**Figure 10: Effect of different LJ parameters investigated using RDF (hydrogen of solvent taken as reference).** RDF  $g(r)$  for monoligated,  $x=1$ , (upper plots) and bisligated,  $x=2$ , (lower plots)  $L_x\text{Ni}(\text{benzaldehyde})$  complexes with ligands  $L=\text{PMe}_3$  (left plots) and  $L=\text{PEt}_3$  (right plots). Complex shown in lower right corner; LJ, Lennard Jones; MD, Molecular Dynamics; RDF, radial distribution function.

the corresponding  $r_{\text{min}}$  values. Using hydrogen as a reference and considering all three LJ parameter sets, one would expect the peak somewhere between 2.7 Å and

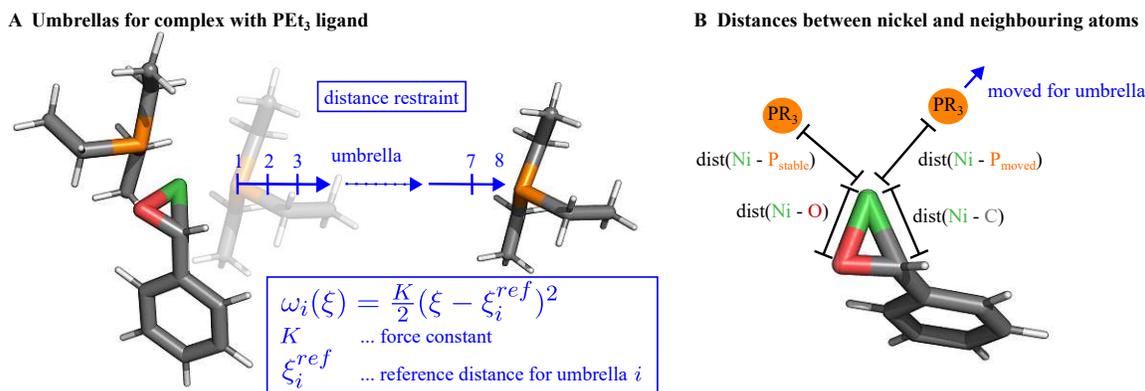
2.9 Å, for carbon between 3.2 Å and 3.4 Å. This is fulfilled as the peak for the RDF choosing hydrogen as a reference is placed between 2.6 Å and 2.8 Å (see Figure 10) and for carbon between 3.4 Å and 3.5 Å (see supplementary information, Figure S1). A notable shift in the position of the peak for the different parameter sets is not discernible. A clear distinction between the LJ parameters tested can be seen in the height of the peak, with the second parameter set resulting in the highest peak, the first set in second, and the third set in last. The LJ parameter set case 3 was excluded from further discussion as the peak was barely visible. Finally, the LJ parameter case 1 with  $\epsilon=5.65$  kcal/mol and  $\sigma=2.27$  Å was chosen for all further simulations. These values were selected as the authors of the paper introduce them specifically for the usage in MD simulations, focusing on different metals.<sup>84</sup> Their objective was to reproduce physical properties such as surface tension, interface properties with solvent as well as mechanical properties.

### 3.2 Effect of Force Constants

MD simulations were performed for two ligands, namely PhP(*t*-Bu)<sub>2</sub> and PEt<sub>3</sub>, using five different force constants with different random seeds for the initial velocity generation. Umbrella sampling was applied for both systems as shown in Figure 11A, applying the force constant  $K$  in the harmonic potential  $\omega_i$ . The data analyzed were obtained from 100 ps of an equilibration run with two distance restraints for both ligands with respect to nickel.

Since an NVT ensemble was performed, the density was lower than that of the equilibrated NPT ensembles, which remained at a constant value of 665 kg/m<sup>3</sup>. However, the set temperature value of 350 K was almost reached, with a value of  $343.1 \pm 1.8$  K. It should be noted that in this experiment, reaching equilibrium was not the primary objective; rather, the distribution of distances between atoms, especially for the two restrained ligands, was the main focus of interest.

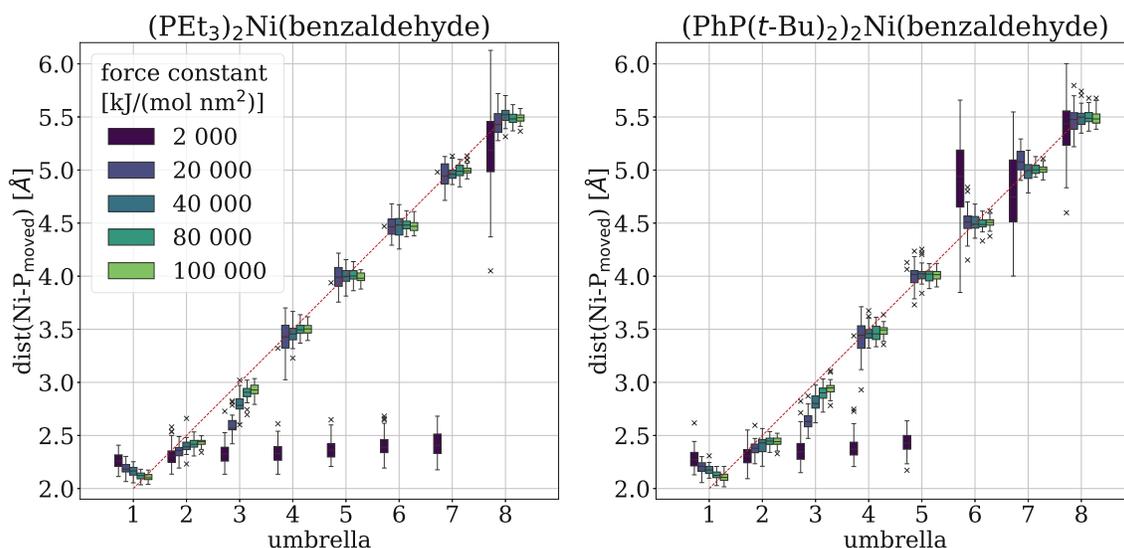
The distances between the nickel atom and all coordinated neighboring atoms were collected by 100 equidistant points from each trajectory. These distances are graphically shown in Figure 11B. The force constants used were 2 000, 20 000, 40 000, 80 000, and 100 000 kJ/(mol nm<sup>2</sup>). Monitoring of all umbrellas showed that the complexes remained intact, in the sense that all three ligands were coordinated to the nickel atom. In the first nine umbrellas, it was observed that the distributions of distances underwent further changes. These are illustrated in Figure 12 and Figure S2 in the supplementary information. The first plot, Figure 12, shows the distances



**Figure 11: Explanatory images for force constant determination.** (A) Schematic image of umbrella sampling for  $(\text{PEt}_3)_x\text{Ni}(\text{benzaldehyde})$  moving from bisligated to monoligated state for higher umbrellas. In box the harmonic potential applied for each umbrella  $i$  is presented. (B) Schematic image of distances analysed in this section: Distance between nickel atom (green) and phosphine ligands  $\text{PR}_3$  (orange), including ligand which is moved away for higher umbrella denoted by  $\text{P}_{\text{moved}}$  and ligand which is kept coordinated to nickel, named  $\text{P}_{\text{stable}}$ . As well as distances from nickel atom to carbon (grey) and oxygen (red) of benzaldehyde both coordinated to nickel. Hydrogen atoms are denoted by white coloring.

between the ligand as it is moved away step by step in each umbrella. This plot served as the basis for selecting different force constants used in the subsequent training data generation process. In order to introduce a significant degree of variety into the distances available in the training data, the force constants 2 000, 20 000 and 100 000  $\text{kJ}/(\text{mol nm}^2)$  were chosen.

Figures 12 and S2 also show that for the  $\text{PhP}(t\text{-Bu})_2$  ligand, the environment around the nickel atom for the bisligated complex is much more crowded than for the  $\text{PEt}_3$  complex. In Figure 12, one can see that the ligand being moved away remains in close proximity to the complex for the first five to seven umbrellas, despite the presence of a restraint to leave the complex. It remains coordinated longer for the  $\text{PEt}_3$  ligand. The second plot (see supplementary information, Figure S2) provides information about the behavior of all the other neighboring atoms of the nickel center. It is interesting to note that for the bisligated complex with  $\text{PEt}_3$ , benzaldehyde acts as haptο-2 ligand in the majority of configurations. In the bisligated  $\text{PhP}(t\text{-Bu})_2$  the benzaldehyde molecule is mostly coordinated only by the oxygen atom as a  $\eta^1$ -ligand. Also, the equilibrium distances of the oxygen of the benzaldehyde and the two phosphorus atoms to the nickel are higher for low umbrella and the  $\text{PhP}(t\text{-Bu})_2$  ligand. The significantly more crowded environment for the  $(\text{PhP}(t\text{-Bu})_2)_2\text{Ni}(\text{benzaldehyde})$



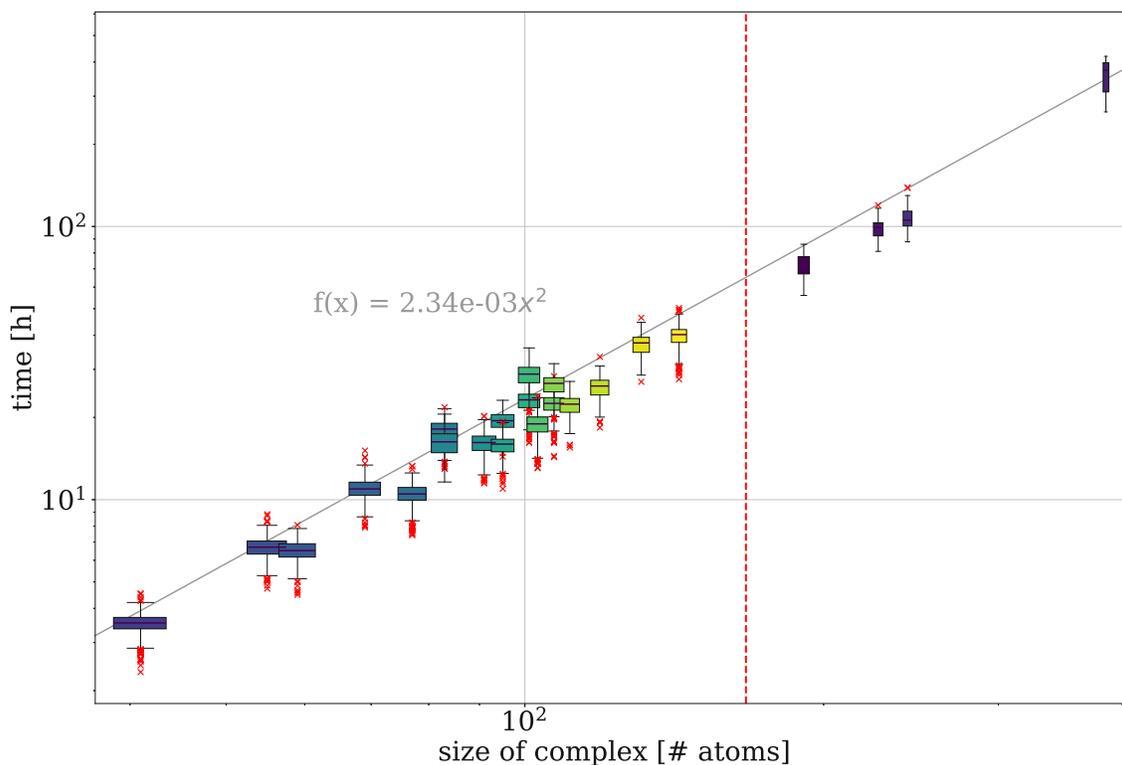
**Figure 12: Effect of different force constants on phosphine ligand as it is moved away from the complex.** Boxplot for distances from restrained MD run, between nickel atom and phosphorus of ligand moved away for the first eight umbrellas. Shown for ligands PEt<sub>3</sub> and PhP(*t*-Bu)<sub>2</sub> in complex. Red-dashed line indicates restrained distance.

complex compared to the (PEt<sub>3</sub>)<sub>2</sub>Ni(benzaldehyde) complex can be compared to the findings of the NMR experiments of Newman-Stonebraker et al.<sup>1</sup> They showed that the ligand PhP(*t*-Bu)<sub>2</sub> was stable in the monoligated complex with nickel and 4-fluorobenzaldehyde, and PEt<sub>3</sub> in the bisligated version. These results were compared with the reasoning outlined in section 3.4.

### 3.3 Training Data Generation

The objective of this section is to analyze various properties of the training data. Convergence across different properties is investigated, along with an examination of the properties of the solvent using the radial distribution function (RDF). Additionally, the configurations of the complexes are explored by analyzing the distances between nickel and its ligands.

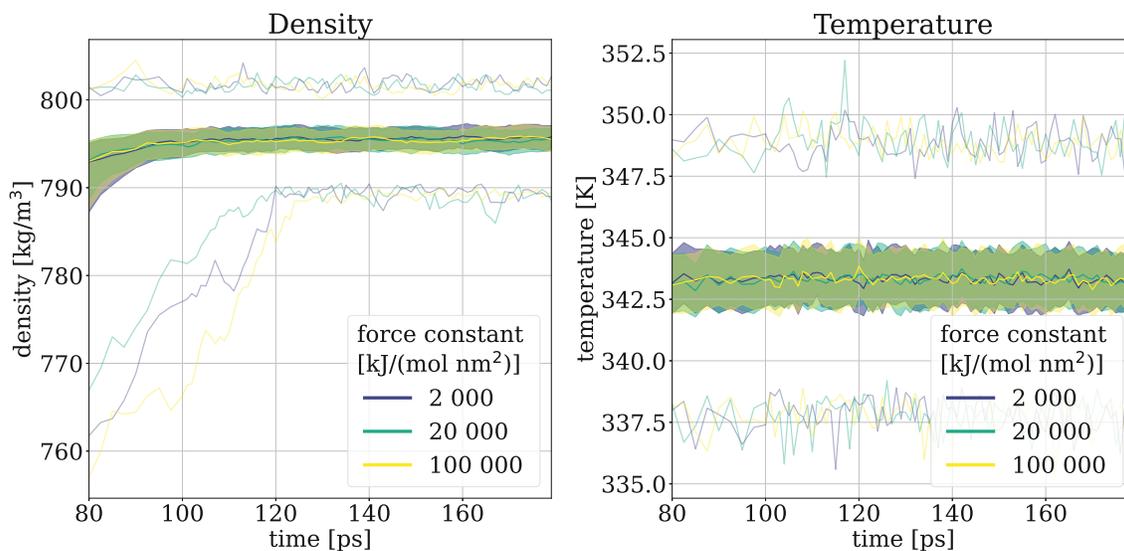
The training data was generated using the LJ parameters and the three force constants determined in the subsections 3.1 and 3.2. The L<sub>2</sub>Ni(benzaldehyde) complex was transformed into the monoligated version by moving one phosphine ligand 37 umbrellas stepwise away for 19 different ligands L. The ligands CyTyrannoPhos,



**Figure 13: Overall runtime of xTB to run 100 ps MD for differently sized systems.** Total time measured in hours that it takes xTB to run 100 ps MD, corresponding to 40 000 energy calculations. The data plotted in the log-log plot is a collection of the equilibration and production runs of the training data generation process (boxplots left of red dotted line). Some data points (minimum of 36) were collected for the ligands CyTyrannoPhos, DrewPhos, TriceraPhos, and PteroPhos (boxplots right of red dotted line).

DrewPhos, TriceraPhos, and PteroPhos were not included in the training data process due to the quadratic scaling observed in the log-log plot in Figure 13 (see power law  $f(x) = ax^2$ ). (CataCXiumA)<sub>2</sub>Ni(benzaldehyde) was the largest system included in the training data with a total of 143 atoms. The time xTB spent on calculations was already over 40 h for 100 ps of MD simulation. The next larger system, comprising the ligand CyTyrannoPhos with 191 atoms, already required more than 72 h; the PteroPhos system (385 atoms) more than 15 d. All the ligands in plot 13 on the right side of the red dashed line were not included in the training data; however, a minimum of 36 timings were collected in test runs. Note that these timings are not

accounting for calculations for the MM zone. However, it can be reasonably assumed that these calculations will not increase the time significantly, given that the total number of solvent molecules remains unchanged. Although the main motivation for the development of semi-empirical methods was to reduce computational resources, they still scale as  $O(N^2)$ .<sup>27</sup> This is problematic for larger systems, which was shown here for all systems larger than CataCXiumA.



**Figure 14: Monitored density and temperature values in the training data.** For the density (left) and temperature (right), the median (solid line), minimum and maximum (transparent line) as well as the range between the Q1 (25% quantile) and Q3 (75% quantile) (colored area) are shown. The data comprises all MD runs for all 19 ligands of the training data combined. The time is displayed from ps 80 until 180, as the first 80 ps were discarded.

The training data frames were extracted from the NPT ensemble in the picosecond range from 80 to 180, as the trajectories were found to be sufficiently converged for the purpose of training data generation (see Figure 14). The density value of the data utilized as training data was  $795.5 \pm 2.0 \text{ kg/m}^3$  (exp. value  $876 \text{ kg/m}^3$  at room temperature<sup>65</sup>). This resulted in a total volume of the simulation box of  $490.6 \pm 1.2 \text{ nm}^3$ . With regard to the temperature, the values exhibited convergence towards  $343.3 \pm 1.8 \text{ K}$  (set to  $350 \text{ K}$ ). These values, along with the density, are presented in Figure 14.

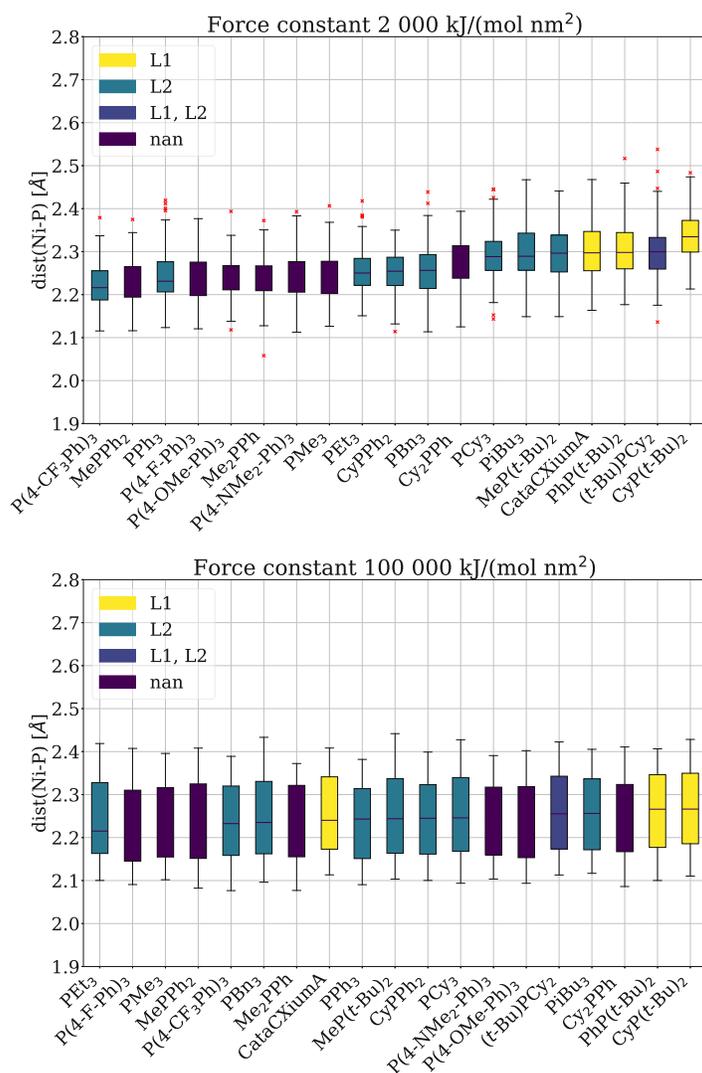
Furthermore, the RDF of the bisligated and monoligated complexes in Figure S3

in the supplementary material was found to be much more diffuse than observed in section 3.1. This can be attributed to the altered temperature, which is close to the boiling point of benzene. The characteristic nickel-solvent interaction peak is observed only for some of the monoligated complexes. For all complexes, however, the RDF demonstrates that the solvent can approach the nickel atom to a greater extent for the monoligated versions than the bisligated ones. Figure S3 shows the RDF for all ligands for the smallest and largest umbrella, corresponding to the monoligated and bisligated complexes, for the force constant  $100\,000\text{ kJ}/(\text{mol nm}^2)$ . The RDF for all different force constants gave similar profiles and showed the same trends. Furthermore, the plot shown was constructed using hydrogen as the reference for the benzene solvent molecules since the peak resulting from the van der Waals interaction of nickel with the solvent was more pronounced there.

The distribution of distances is of great importance to obtain a comprehensive sampling of the region of interest. First, it was observed that none of the complexes dissociated. Second, the distances were examined in more detail for nickel with respect to all of its neighboring atoms. These are the phosphorus atom of the phosphine ligands and the carbon and oxygen atoms of the benzaldehyde to which nickel is coordinated. A different behavior was observed for the distances in umbrellas one to seven compared to all other umbrellas. Depending on the simulated complex, both phosphine ligands remained coordinated to the nickel atom for the first five to seven umbrellas. However, for higher umbrellas, the restraint is too strong, and the phosphine ligand separates from the complex. The distances are therefore analyzed separately for the first umbrellas, representing the  $\text{L}_2\text{Ni}(\text{benzaldehyde})$  state, and the rest, representing the  $\text{L}_1\text{Ni}(\text{benzaldehyde})$  state.

The distances for the  $\text{L}_1\text{Ni}(\text{benzaldehyde})$  complex for all ligands are shown in Figure S4 and S5 in the supplementary information. The data presented in these figures are for a single force constant, as no significant variation was observed across different force constants. The distributions of nickel-oxygen and nickel-phosphorus distances are narrow and similar for all complexes. The distribution for nickel-carbon distances is also comparable across different ligands, although there are some outliers that indicate that the hapticity of benzaldehyde varies with respect to nickel.

Regarding the  $\text{L}_2\text{Ni}(\text{benzaldehyde})$  complexes, two aspects can be highlighted. Firstly, the force constant exerts a significant influence on the distances. Secondly, the experimental data are in agreement with the trends observed in the data. The distances of nickel to both phosphine ligands are shown in Figure 15 for the first three umbrellas



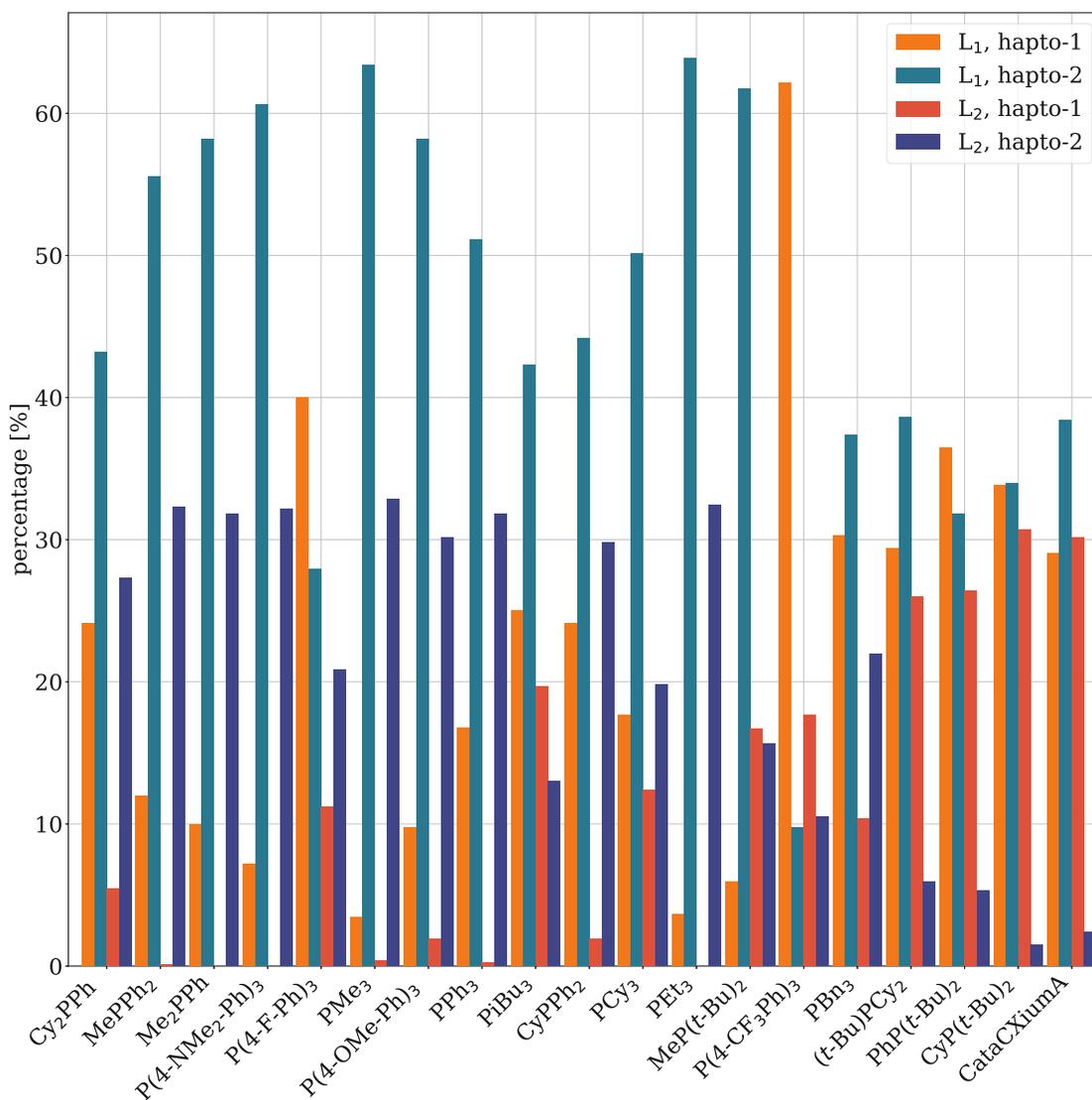
**Figure 15: Distribution of distances between nickel and phosphine ligands in the training data for the L<sub>2</sub>Ni(benzaldehyde) complex.** The distances between nickel and phosphorus are presented for all ligands for the lowest force constant (2 000 kJ/(mol nm<sup>2</sup>), top) and the highest force constant (100 000 kJ/(mol nm<sup>2</sup>), bottom). Data from the first three umbrellas combined are shown. Experimental ligation state results from Newman-Stonebraker et al. are indicated with colored boxes, being the monoligated "L1", the bisligated "L2" complex, or the presence of both indicated by "L1, L2". If no experiment was conducted for the ligand, this is indicated by "nan".<sup>1</sup>

combined. For the higher force constant, it can be observed that the distribution becomes wider as the phosphine ligands are restrained to a distance of either 2, 2.5, or 3 Å. In the case of the lower force constants, the measured distance deviates more from the restrained distance, with the majority of ligands situated between 2.2 and 2.35 Å with respect to the nickel atom. In particular, when comparing the ligation state determined experimentally by Newman-Stonebraker et al., trends can be observed.<sup>1</sup> For the lowest force constant, it can be seen that those complexes which exhibit a preference for the monoligated state have a greater distance between the nickel and the phosphine ligands.

The distribution of distances between the nickel atom and the initially coordinated carbon and oxygen atoms of the benzaldehyde molecule for the  $L_2Ni$ (benzaldehyde) complexes is also investigated (compare Figure S6 in supplementary material). The distances of the oxygen atoms remain relatively constant across the different ligands. In contrast, the distances of the carbon atoms exhibit a pronounced change, which accounts for a change in hapticity. Once more, for those complexes that have been demonstrated in experiments to favour monoligation, benzaldehyde occurs predominantly as a  $\eta^1$  ligand.

Figures S7 and 16 show the composition of the training data set for force constants 2 000 kJ/(mol nm<sup>2</sup>) and 100 000 kJ/(mol nm<sup>2</sup>). In these plots, configurations are partitioned using the ligation state and hapticity of the complexes. The occurrences are counted for the first ten umbrellas, where the environment is more crowded. A complex was designated as  $L_1Ni$ (benzaldehyde) if the phosphorus atom of the ligand was more than 3 Å away from the nickel atom. The hapticity of the benzaldehyde is categorized as hapto-2 if the distance to the carbon atom is less than 2.5 Å, and as hapto-1 if the distance is above the same value.

Figure S7 illustrates that using the lowest force constant results in some complexes never occurring as bisligated complexes with a hapto-2 benzaldehyde ligand. This is problematic because this is the configuration occurring in the reaction which is investigated (see Figure 5). Nevertheless, using the force constant of 100 000 kJ/(mol nm<sup>2</sup>) solves this problem partially. In Figure 16 one can see that for the higher force constant the required configuration combining benzaldehyde as  $\eta^2$  ligand and the bisligated  $L_2Ni$ (benzaldehyde) complex, occurs for all ligands. For some ligands, especially CataCXiumA and CyP(*t*-Bu)<sub>2</sub>, this configuration is still strongly under-represented.



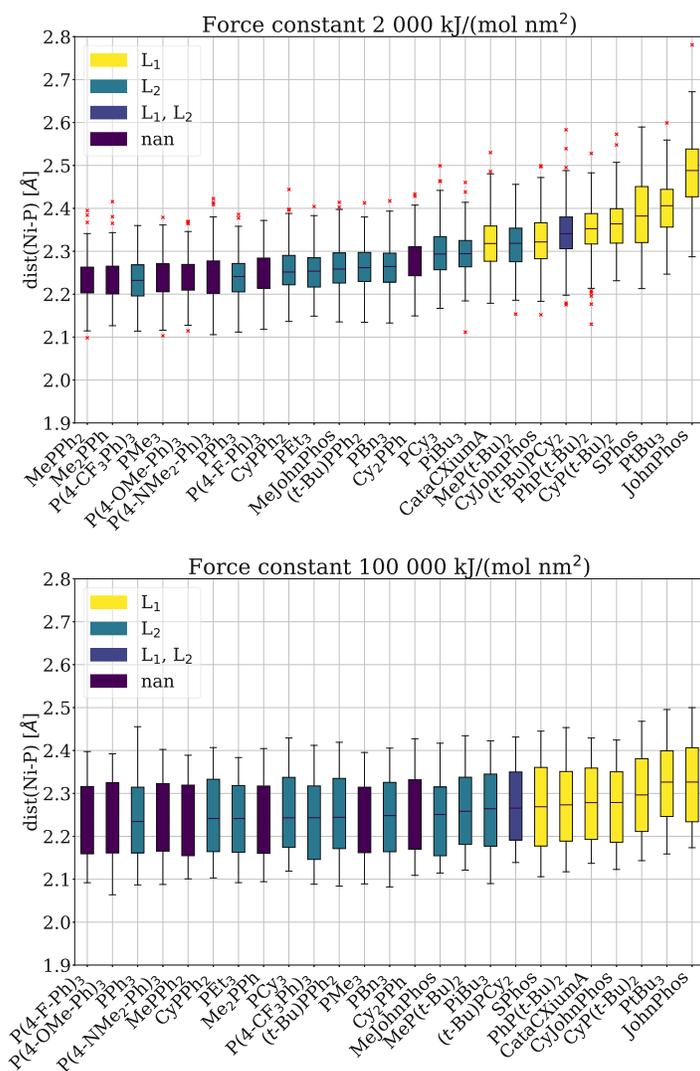
**Figure 16: Configurational composition of training data for the first ten umbrellas for highest force constant.** All configurations are divided into four different cases, distinguishing between L<sub>1</sub>Ni(benzaldehyde) complexes and L<sub>2</sub>Ni(benzaldehyde) complexes as well as hapto-1 and hapto-2 benzaldehyde ligand. Occurrence is given in percentages adding up to 100% for each ligand. Force constant is 100 000 kJ/(mol nm<sup>2</sup>). Red shades indicate hapto-1, blue shades hapto-2.

### 3.4 QM/MM MD for $L_xNi(4\text{-fluorobenzaldehyde})$

The ligation state experiments conducted by Newman-Stonebraker et al. used 4-fluorobenzaldehyde instead of benzaldehyde as the  $\eta^2$  ligand present in the complexes.<sup>1</sup> Other than that, the reaction was the same as shown in Figure 5. In their study, benzaldehyde was used for all DFT calculations, while 4-fluorobenzaldehyde was used for the NMR study, allowing the measurement of  $^{19}\text{F}$  NMR spectra. In order to facilitate a meaningful comparison between our results and the experimental NMR experiments, it is necessary to ascertain whether the fluorine atom exerts a significant influence on the ligation state of the complex.

As outlined in the introductory section 1.1, the primary effects observed to influence the ligation state are attributed to the properties of the ligand. These include the ligand size, flexibility, and steric hindrance. In addition, solvent effects, thermodynamic or kinetic factors may also influence the outcome. It is likely that the electronic effect is most influenced by the addition of fluorine to the substrate, since fluorine has a higher electronegativity than hydrogen. To investigate this question, QM/MM MD runs were performed for the same complexes (and some additional ones, see Figure 8) in which benzaldehyde was substituted with 4-fluorobenzaldehyde. The distribution of distances for the phosphine ligands with respect to the nickel atom for the  $L_2Ni(4\text{-fluorobenzaldehyde})$  complex, Figure 17, exhibits similar trends to those observed for the  $L_2Ni(\text{benzaldehyde})$  complex, Figure 15. Both figures show that the distances correlate with the experimental binary outcome of the ligation state experiment performed by Newman-Stonebraker et al.<sup>1</sup> This trend is evident regardless of the substituent in the para position. The trend is also evident for the four newly added complexes that have been shown to be experimentally stable as  $L_1Ni(4\text{-fluorobenzaldehyde})$ .

Furthermore, the work of Newman-Stonebraker et al. indicates that the ligation state depends solely on the steric bulk within the first coordination sphere, which is a consequence of the ligands and not the substrate. The steric bulk within the first coordination sphere is quantified by  $\%V_{\text{bur}}$ , which is a value for each ligand obtained from the *kraken* library.<sup>12</sup> A ligand with a  $\%V_{\text{bur}}$  value greater than 32% was found to be unsuitable for Ni-catalyzed Suzuki-Miyaura reactions. For a subset of 28 of these ligands, the ligation state was determined (using 4-fluorobenzaldehyde as a substrate) and it was shown that all complexes were divided into monoligated and bisligated complexes by the same cutoff value of  $\%V_{\text{bur}}$ . This result provides compelling evidence for a correlation between the ligation states observed in both experiments, despite the fact that the substrate used in the reactions was not 4-



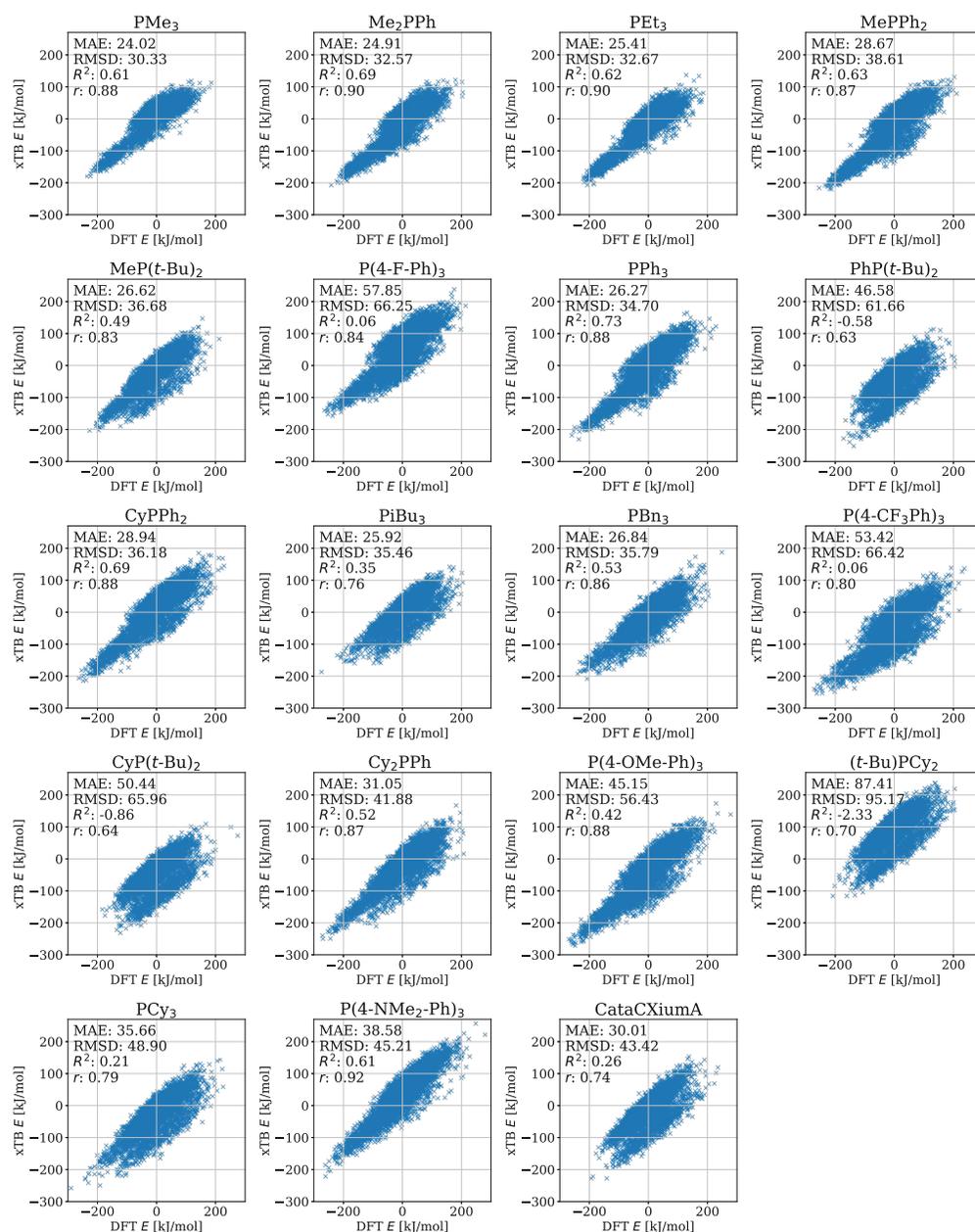
**Figure 17: Distribution of distances between nickel and phosphine ligands in the training data for the  $L_2Ni(4\text{-fluorobenzaldehyde})$  complex.** The distances between nickel and phosphorus are presented for all ligands for the lowest force constant ( $2\,000\text{ kJ}/(\text{mol nm}^2)$ , top) and the highest force constant ( $100\,000\text{ kJ}/(\text{mol nm}^2)$ , bottom). Data from the first three umbrellas combined are shown. Experimental ligation state results from Newman-Stonebraker et al. are indicated with colored boxes, being the monoligated "L1", the bisligated "L2" complex, or the presence of both indicated by "L1, L2". If no experiment was conducted for the ligand, this is indicated by "nan".<sup>1</sup>

fluorobenzaldehyde.<sup>1</sup>

Regarding the other properties of the QM/MM MD runs for the complex with 4-fluorobenzaldehyde, it should be noted that they were comparable to those obtained from the benzaldehyde runs. Once more, only the data that had converged were analyzed, where the temperature, density, and box volume had reached comparable values. The radial distribution function (RDF) exhibited comparable curves. The additional restraint applied to the carbon and oxygen atoms of the hapto-2 4-fluorobenzaldehyde had a significant impact on the distance distribution. In particular, for high force constants, the 4-fluorobenzaldehyde always acted as an  $\eta^2$  ligand.

### 3.5 xTB versus DFT

A total of 5 550 frames were extracted from the trajectories for each ligand obtained from the QM/MM MD simulations, where xTB was the QM method used for generating the trajectories within the QM/MM MD simulations. The trajectories were then re-evaluated using a higher-level DFT method, by using the ORCA program. The energy values of both methods can be directly compared and are presented in Figure 18. For the comparison of the forces, the derivatives were calculated for QM and MM atoms, respectively. Those values are shown in Figure 19 and in the supplementary material in Figure S8. These values are significant as they serve as the basis for the propagation of the system in each time step of the MD run. The values are also presented in a combined manner in Table 6 for all complexes. In the figures and the table, the mean absolute error (MAE), the root-mean-square deviation (RMSD), the  $R^2$  score, and the Pearson correlation coefficient  $r$  are presented. The MAE is a measure of the average magnitude of the difference between both methods, while the RMSD uses the squares of the errors to give an estimate. Both methods exhibit comparable trends in the properties, with the only distinction being that the RMSD assigns greater significance to larger errors. The  $R^2$  score, the coefficient of determination, is typically employed to assess the quality of a linear regression model, taking into account both predicted and actual values. The optimal value for the  $R^2$  score is 1. A value of 0 indicates that the mean is predicted, while a value below 0 indicates that the model performs worse than expected. The Pearson correlation coefficient is a measure of the linear relationship between two values, with a range of -1 to 1. A value of -1 indicates a negative linear relationship, while a value of 1 indicates a positive linear relationship.



**Figure 18: Energy values of xTB versus DFT in training data set.** For all frames included in the training data (5 550 per complex) the energies  $E$  obtained from xTB are compared to the re-evaluated values of DFT for each complex individually.

**Table 6: Performance of xTB versus DFT.** Mean absolute error (MAE), root-mean-square deviation (RMSD),  $R^2$  score, Pearson coefficient  $r$ , minimum and maximum values for the difference between the values obtained by xTB and DFT. The minimum and maximum values, spanning the range, are obtained from the measures of all 19 different ligands (upper part of the table). Further, the measures are given for the data of all ligands combined (lower part of the table).

	energies [kJ/mol]	QM gradients [kJ/mol · Å]	MM gradients [kJ/mol · Å]
range for all ligands			
MAE	[24.0 - 87.4]	[12.45 - 16.26]	[0.04 - 0.07]
RMSD	[30.3 - 95.2]	[21.24 - 27.21]	[0.10 - 0.18]
$R^2$	[-2.33 - 0.73]	[0.89 - 0.94]	[-0.37 - 0.71]
$r$	[0.63 - 0.92]	[0.96 - 0.98]	[0.73 - 0.88]
min	$[1.1 \times 10^{-4} - 8.1 \times 10^{-2}]$	$[1.4 \times 10^{-7} - 8.0 \times 10^{-5}]$	$[3.0 \times 10^{-8} - 2.0 \times 10^{-6}]$
max	[119.8 - 219.4]	[297.62 - 494.02]	[3.03 - 10.45]
result for all ligands combined			
MAE	37.6	14.02	0.05
RMSD	50.4	23.12	0.15
$R^2$	0.44	0.93	0.44
$r$	0.71	0.97	0.78
min	$1.1 \times 10^{-4}$	$1.4 \times 10^{-7}$	$3.0 \times 10^{-8}$
max	219.4	494.02	10.45

The discrepancies between the energy predictions of xTB and DFT are approximately one order of magnitude higher than the chemical accuracy if the DFT calculation is assumed to be the ground truth ( $1 \text{ kcal/mol} = 4.184 \text{ kJ/mol}$ ).<sup>27</sup> The MAE for all frames and across all systems is stated in Table 6 with a value of 37.6 kJ/mol. The range of possible values for MAE varies considerably depending on the complex, with a minimum of 24.0 kJ/mol and a maximum of 87.4 kJ/mol (see Table 6 and Figure 18). Also the minimum and maximum values indicate an even wider distribution. The observed trend in these values across different complexes appears to be random and is heavily influenced by the data selected. Figure 18, however, provides a clear illustration of the range of values that can be observed. The value of the  $r$  coefficient above 0.5 indicates the presence of a linear relationship between the two methods. However, the value of  $R^2$  indicates that the proportion of the variance explained by

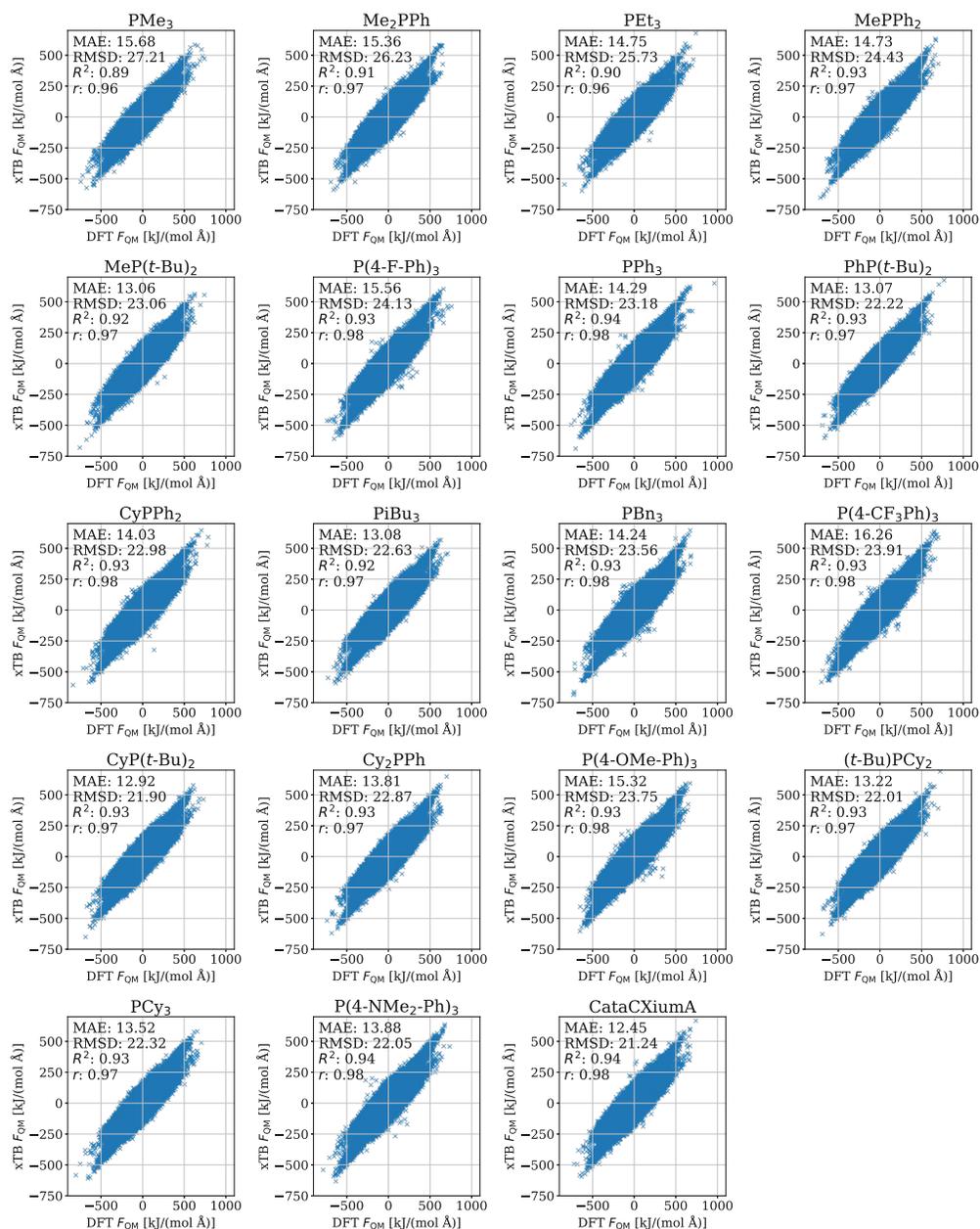


Figure 19: Forces for QM atoms evaluated by xTB versus DFT in training data set. For all frames used in the training data (5 550 per complex) the forces in the QM region  $F_{QM}$  obtained from xTB are compared to the re-evaluated values of DFT.

the relationship is relatively small.

The gradients of the energies in the QM zone are shown in Figure 19. The MAE values and their ranges are again summarized in Table 6. It is also noteworthy that the  $R^2$  score now falls within the range of 0.89 to 0.94, indicating that the variance of the xTB values can be predicted from the variance of the DFT gradient values. Moreover, the Pearson correlation coefficient is considerably higher. The aforementioned values demonstrate that the QM gradients of both methods are considerably more interchangeable than those of the absolute energy values. This behavior is logical when one considers that the absolute total energy values are not particularly significant, whereas differences between them are. The forces reflect these differences, as they are the change in the total energy with respect to a direction, either being x, y, or the z-coordinate.

In the supplementary material, the same plots are presented for the gradients on the MM atoms, and the values are once again summarized in Table 6. Here, the linear relationship deteriorates, and the coefficient of determination indicates that the variance of one method cannot be predicted by taking the other method into account. This is to be expected, given that the MM charges are only included in the QM calculation as external partial charges and are not explicitly taken into account.

### 3.6 Training AMP Models

The mean absolute error was selected as the metric to facilitate comparisons between the results of the various models in the subsequent sections. While the  $R^2$  score and the Pearson correlation coefficient  $r$  could also be compared, but are not presented here. The values of these coefficients were, in general, extremely high, indicating a near-perfect linear relationship between the true and predicted values. For the energies and QM gradients, the values of  $r$  and  $R^2$  were found to be between 0.99 and 1.00.

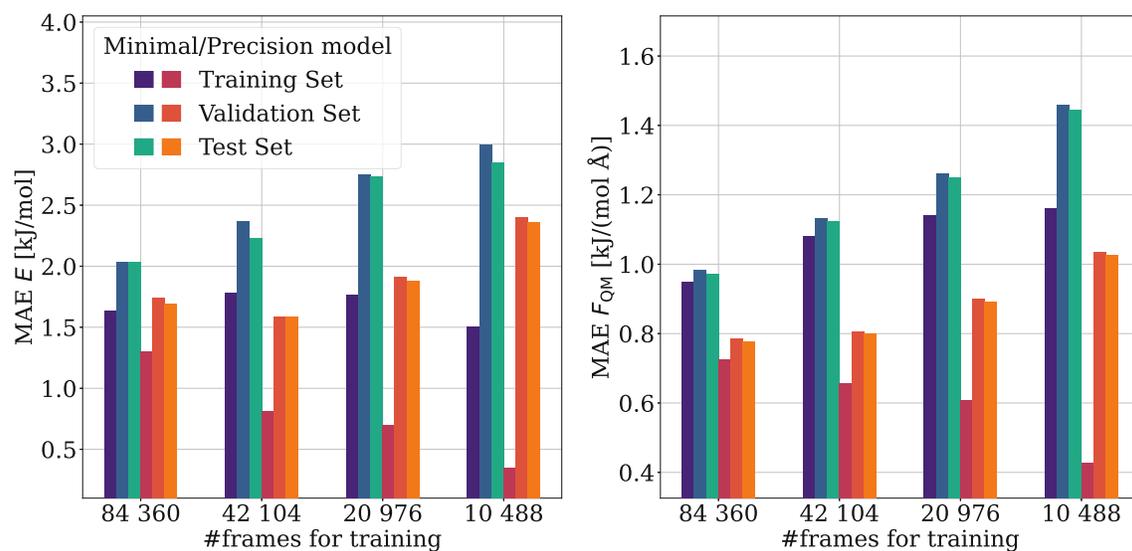
#### 3.6.1 Effect of Amount of Training Data Used

In this experiment, the mean absolute error (MAE) was compared for model versions 1 to 4. The train/validation/test split is described in Table 7, which is a subset of Table 2 from the Computational Details section. The MAEs for energies  $E$  and QM energy gradients  $F_{\text{QM}}$  are shown in Figure 20.

**Table 7: Train/validation/test split for testing effect of amount of training data.**

For model versions 1 - 4 the total amount of frames for each split (train/validation/test) as well as the percentages are listed. This table is a subset of Table 2.

version	data set	train [# frames]	validation [# frames]	test [# frames]	train [%]	validation [%]	test [%]
1	1	84 360	1 976	1 976	80.0	1.9	1.9
2	1	42 104	1 976	1 976	39.9	1.9	1.9
3	1	20 976	1 976	1 976	19.9	1.9	1.9
4	1	10 488	1 976	1 976	9.9	1.9	1.9



**Figure 20: Effect of different amounts of training data.** The MAE is shown for the QM energy  $E$  (left) as well as the forces on the QM atoms  $F_{QM}$  (right) for the model versions 1 - 4. MAEs are shown for training, validation, and test sets. Minimal models (blue color scheme) and precision models (red color scheme) are shown. The total number of frames in the training data is 84 360, 42 104, 20 976 and 10 488.

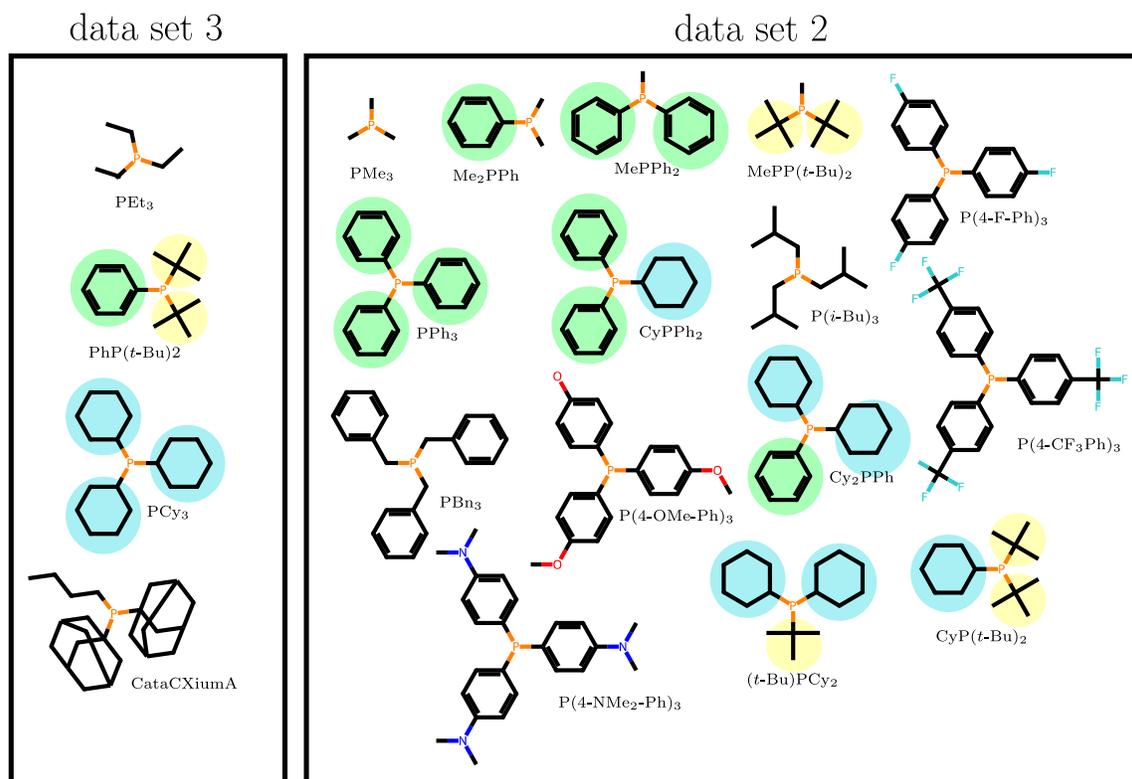
As anticipated, the errors for the minimal model (indicated by the blue bars) are higher than those for the precision model (red bars). This is a direct consequence of the precision model having a greater number of parameters and a more precise representation of physical properties, such as the inclusion of additional parameters for encoding distances or multipoles per atom (see Table 1).

It is desirable to perform as well on the training data set as on unseen data, and therefore to minimize both bias and variance, called the bias-variance tradeoff.<sup>86</sup> The objective is to construct a model that is both generalizable to new data and capable of capturing regularities present in the training data. This concept is less effectively realized when the training data set is comprised of a smaller number of data frames. In Figure 20 it can be seen how the MAE difference between training and test data sets increases if fewer and fewer frames are used in the training. As the number of training data points is reduced, the degree of overfitting increases. This phenomenon is particularly evident in the case of the precision model. This might be attributed to the higher complexity of this model, which results in a lack of generalization ability when few training data points are available.

It is notable that all errors are below 4.184 kJ/mol (1 kcal/mol), which is generally considered to be chemical accuracy.<sup>27</sup> The MAE for the MM gradients, dipoles and quadrupoles also exhibits similar trends (see supplementary information, Figure S9). The data indicates that approximately 40 000 frames should be sufficient for training purposes, as the discrepancy between the errors on the test set and the training set is relatively minor. Consequently, for the next experiment, approximately 40% of the training data will be utilized, in addition to conducting the experiment with all data.

### 3.6.2 Transferability Experiment

The second experiment was conducted with the objective of evaluating the transferability of a model when exposed to new data that was not included in the training data set. As previously described in section 2.3, the data was divided into two subsets, referred to as data sets 2 and 3. The composition of both data sets with respect to the ligands present in them is shown in Figure 21. Data set 3 comprised the data for four ligands, namely  $\text{PEt}_3$ ,  $\text{PhP}(t\text{-Bu})_2$ ,  $\text{PCy}_3$  and  $\text{CataCXiumA}$ . In contrast, data set 2 encompassed all the other ligands. The transferability experiment was conducted twice, using 39.9% and 96.9% of data set 2. For the first experiment with less data, model version 5 - 9 were trained. Their composition with respect to train/validation/test split is shown in Table 8. Model version 5 is the only model with no data from data set 3. Model versions 6 to 9 were augmented with small amounts from data set 3. These small amounts were 0.1%, 0.3%, 1.9%, and 4.0% of all the data available in data set 3. However, a more meaningful comparison can be made between the amount of data set 3 data used and the total amount of data used in the training. Subsequently, the percentages are reduced to values of 0.1%, 0.2%,



**Figure 21: Ligands in data sets 2 and 3.** Illustrative representation of ligands in data sets 2 and 3. Subunits that occur in data set 2 and 3 are indicated with color coding: green for phenyl group, yellow for *tert*-butyl group, and blue for hexyl group.

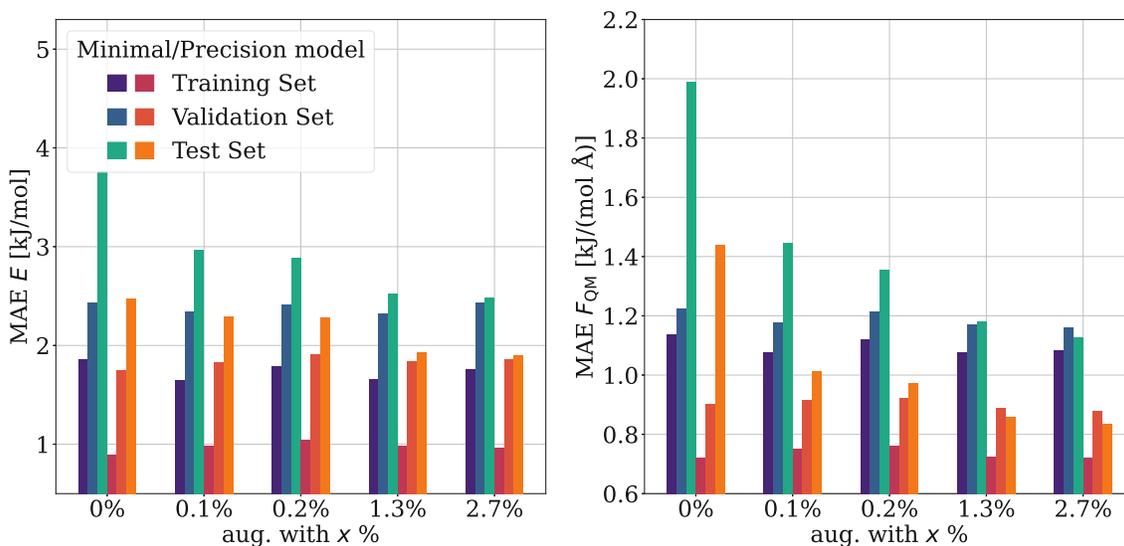
1.3%, and 2.7%. For the second transferability experiment, using 96.9% of data set 2, the model versions 11 - 15 were used. Again, their composition is shown in Table 8. For model versions 12 until 15, data from data set 3 was added. The percentages were increased compared to model versions 6 to 9, yielding the same percentages when compared to the overall data, being 0.1%, 0.2%, 1.3%, and 2.7%. The main analysis will be done for the models 5 - 9, as Section 3.6.1 showed that less data in the training is sufficient.

Figure 22 depicts the MAE on energies  $E$  and QM gradients  $F_{\text{QM}}$  for the training, validation, and test data. As anticipated, the errors on the training and validation data remain largely unchanged when transitioning from model version 5 to 9. This phenomenon can be readily explained by the observation that the primary composition of the data sets remains largely unchanged, with a maximum of 2.7% of distinct

**Table 8: Train/validation/test split for testing transferability.** The total amount of frames for each split (train/validation/test) as well as the percentages are listed for the transferability experiment 1 and 2. Transferability experiment 1 includes model versions 5 - 9 and experiment 2 model version 11 - 15. The data shown in this table is a subset of Table 2.

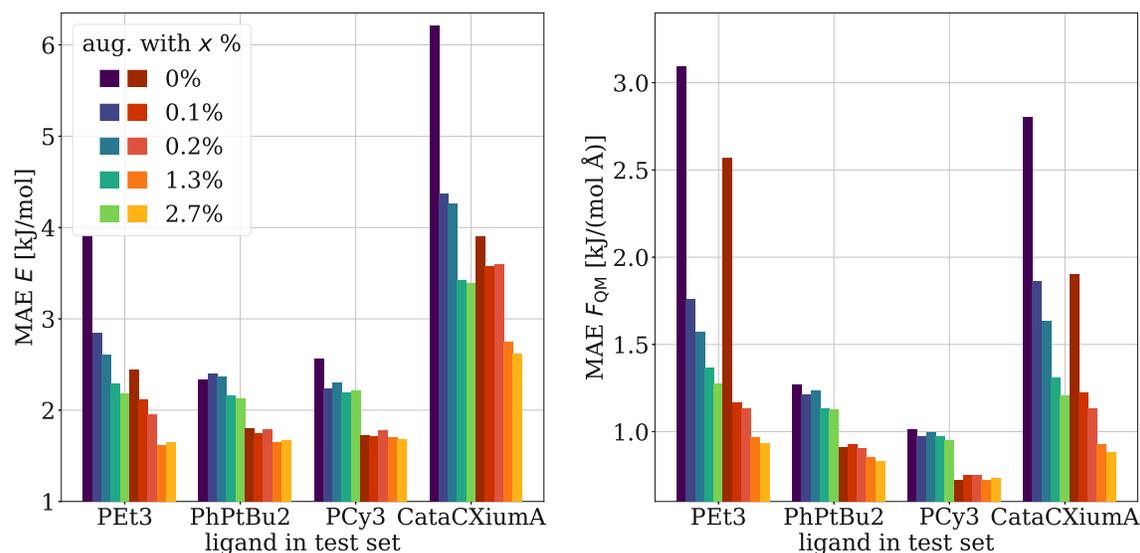
version	data set	train [# frames]	validation [# frames]	test [# frames]	train [%]	validation [%]	test [%]
transferability experiment 1							
5	2	33 240	2 400	0	39.9	2.9	0.0
	3	0	0	17 760	0.0	0.0	80.0
6	2	33 240	2 400	0	39.9	2.9	0.0
	3	32	32	17 760	0.1	0.1	80.0
7	2	33 240	2 400	0	39.9	2.9	0.0
	3	64	64	17 760	0.3	0.3	80.0
8	2	33 240	2 400	0	39.9	2.9	0.0
	3	416	416	17 760	1.9	1.9	80.0
9	2	33 240	2 400	0	39.9	2.9	0.0
	3	896	896	17 760	4.0	4.0	80.0
transferability experiment 2							
11	2	80 640	2 400	0	96.9	2.9	0.0
	3	0	0	17 760	0.0	0.0	80.0
12	2	80 640	2 400	0	96.9	2.9	0.0
	3	96	96	17 760	0.4	0.4	80.0
13	2	80 640	2 400	0	96.9	2.9	0.0
	3	192	192	17 760	0.9	0.9	80.0
14	2	80 640	2 400	0	96.9	2.9	0.0
	3	1 088	1 088	17 760	4.9	4.9	80.0
15	2	80 640	2 400	0	96.9	2.9	0.0
	3	2 208	2 208	17 760	9.9	9.9	80.0

data being introduced. Conversely, the addition of a small amount of data set 3 to the training and validation process has a much greater impact on the error of the test set. It is noteworthy that even when only 0.1% of the data is added to the training set, the MAE on the test set can be reduced by 0.8 kJ/mol for the minimal model. The improvement in terms of MAE for adding 0% to 0.1% is the most pronounced, indicating that even a minimal quantity of data can markedly reduce the MAE of the system.



**Figure 22: Model versions 5 - 9: Transferability experiment.** The MAE is shown for the QM energy  $E$  (left) as well as the forces on the QM atoms  $F_{QM}$  (right) for the model versions 5 - 9. MAEs are shown for training, validation and test sets. Minimal models (blue color scheme) and precision models (red color scheme) are shown. Training and validation sets consisted of data set 2 but were gradually augmented with frames from data set 3 (0%, 0.1%, 0.2%, 1.3%, and 2.7%). For testing, data set 3 was used.

In order to ascertain whether any of the aforementioned four systems may have contributed to the occurrence of the larger error within the test set, the model versions 5 to 9 were subjected to testing on data pertaining to each ligand in isolation. The MAE on the test data split for the four ligands is shown in Figure 23. This plot illustrates the MAEs on energies and QM gradients. The MAEs on the other properties are presented in the supplementary material in Figure S11. A comparison of the errors for all properties reveals that the CataCXiumA system performs the worst. Additionally, the performance of the system with the ligand  $\text{PEt}_3$  is not optimal for most properties, including QM energy gradients and energies. For the systems with the ligands  $\text{PhP}(t\text{-Bu})_2$  and  $\text{PCy}_3$ , the MAE is as low as if the model was actually trained on these ligands. This can be explained by comparing the subunits present in the other 15 ligands that were used to train the models, see Figure 21. Four of the ligands contain cyclohexyl groups, either twice or only once. The system with the tricyclohexylphosphine group,  $\text{PCy}_3$ , also has the lowest test set error. A further three ligands in the dataset contain *tert*-butyl groups, while five

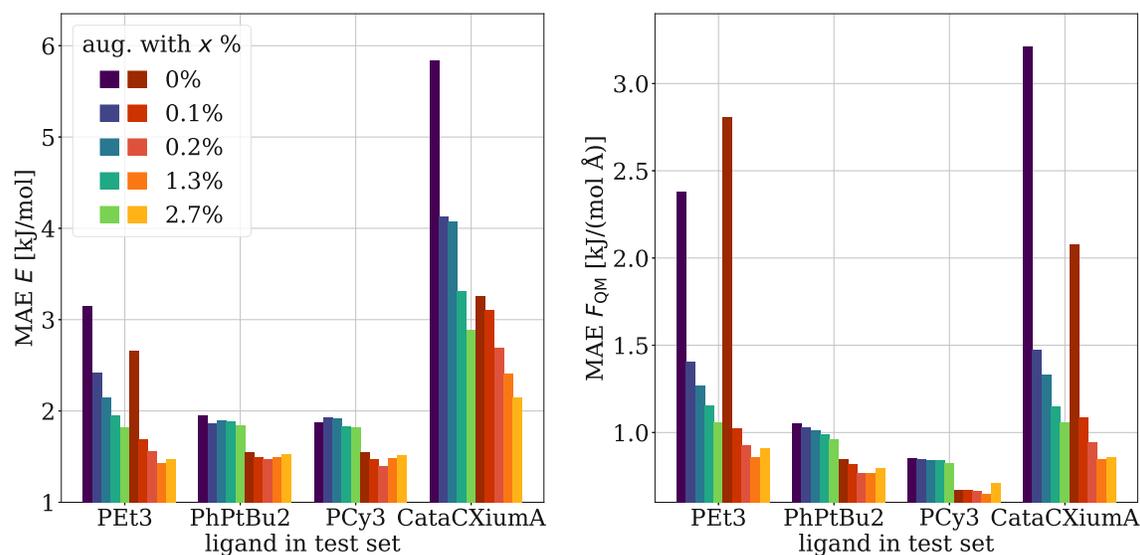


**Figure 23: Model versions 5 - 9: Transferability experiment analyzed for different ligands.** The MAE is shown for the QM energy  $E$  (left) as well as the forces on the QM atoms  $F_{QM}$  (right) for the model versions 5 - 9. MAEs for the test data set are shown only. MAEs are reported for complexes  $(PEt_3)_xNi(\text{benzaldehyde})$ ,  $(PhP(t-Bu)_2)_xNi(\text{benzaldehyde})$ ,  $(PCy_3)_xNi(\text{benzaldehyde})$ , and  $(CataCXiumA)_xNi(\text{benzaldehyde})$ . Minimal models (blue color scheme) and precision models (red color scheme) are shown. Training and validation sets consisted of data set 2 but were gradually augmented with frames from data set 3 (0%, 0.1%, 0.2%, 1.3%, and 2.7%). For testing, data set 3 was used.

ligands contain phenyl groups, if those not being substituted are counted. These two groups are present in  $PhP(t-Bu)_2$ , which also performs well. Conversely, neither ethyl nor adamantane groups are present in other ligands. This can explain the poor performance for the complex with the  $PEt_3$  and  $CataCXiumA$  ligand. As previously observed, augmenting the training data with frames of the four ligands has the most pronounced effect when adding the first 32 frames. For the systems  $PhP(t-Bu)_2$  and  $(PCy_3)_xNi(\text{benzaldehyde})$ , which already performed well, only a slight improvement is observed. In the case of the other two ligands, particularly  $CataCXiumA$  the MAE is reduced by approximately 2 kJ/mol.

Those four ligands were selected first and foremost, for their distinct subunits. The  $PhP(t-Bu)_2$  and  $PCy_3$  ligands have both subunits occurring in the training data, whereas the other two ligands,  $PEt_3$  and  $CataCXiumA$ , have subunits being new to

the tested model. As expected, the error on the test data increases for the latter two systems. The second reason for choosing those four systems, was that the results by Newman Stonebraker et al. suggest that  $\text{PhP}(t\text{-Bu})_2$  and  $\text{CataCXiumA}$  prefer the monoligated complex, while  $\text{PEt}_3$  and  $\text{PCy}_3$  prefer the bisligated complex. This fact seems to be irrelevant to the predictive power of the model; what matters is the representation of the distinct subunits in the training and validation data.

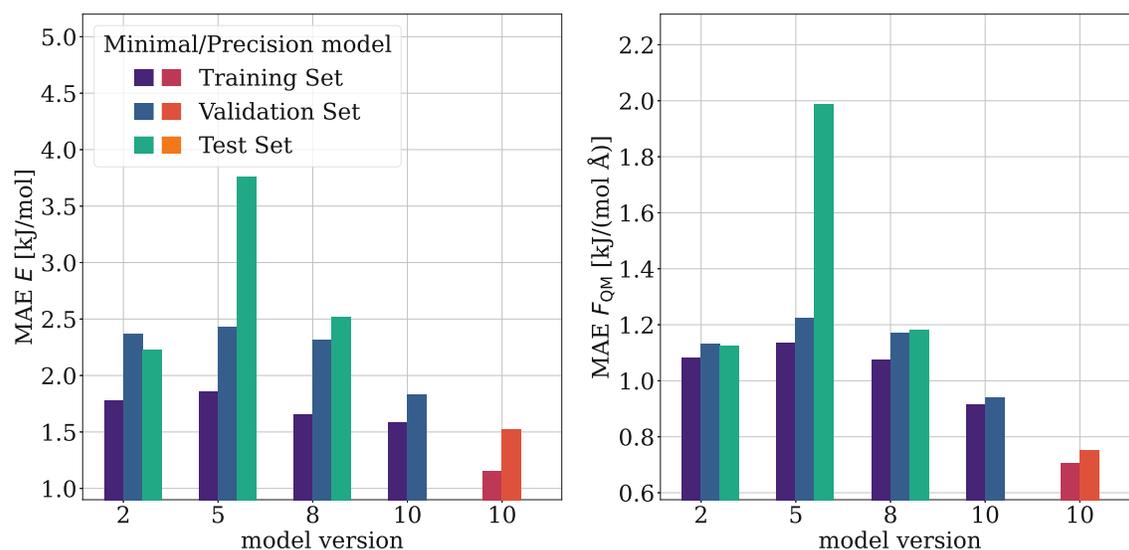


**Figure 24: Model versions 11 - 15: Transferability experiment analyzed for different ligands.** The MAE is shown for the QM energy  $E$  (left) as well as the forces on the QM atoms  $F_{\text{QM}}$  (right) for the models 11 - 15. MAEs for the test data set are shown only. MAEs are reported for complexes  $(\text{PEt}_3)_x\text{Ni}(\text{benzaldehyde})$ ,  $(\text{PhP}(t\text{-Bu})_2)_x\text{Ni}(\text{benzaldehyde})$ ,  $(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$ , and  $(\text{CataCXiumA})_x\text{Ni}(\text{benzaldehyde})$ . Minimal models (blue color scheme) and precision models (red color scheme) are shown. Training and validation sets consisted of data set 2 but were gradually augmented with frames from data set 3 (0%, 0.1%, 0.2%, 1.3%, and 2.7%). For testing data set 3 was used.

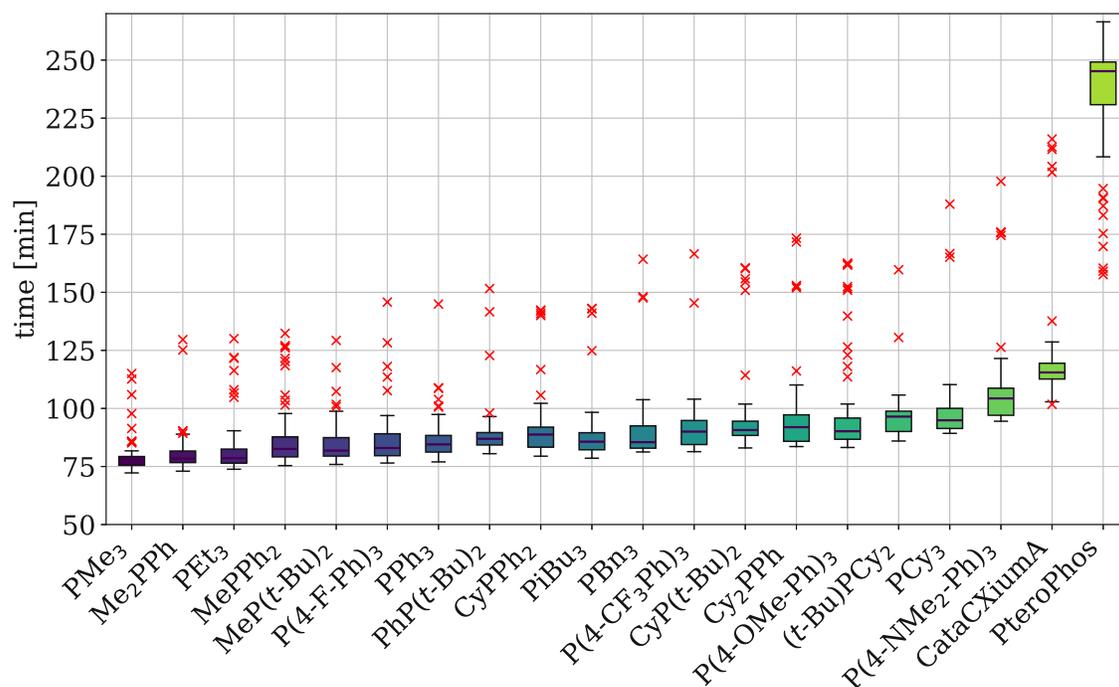
The second transferability experiment used the maximum amount of data set 2, comprising 96.9% instead of 39.9% of the data set. This experiment used model versions 11 to 15. 3% of the data were again reserved for the validation process (see Table 8). The MAEs for all properties on training, validation and test data are provided in the supplementary information, Figure S13. The result is comparable to

the equivalent experiment that was trained on less than half of the data, with the only difference being slightly lower MAE values for models trained on more data. Once more, the test errors were collected and plotted for the four ligands from data set 3 individually. The result for MAEs of energies and gradients of energies for QM particles is displayed in Figure 24. The same trends can be seen as for the model versions 5 to 9 (compare Figure 23). The presented plots demonstrate that, even when repeating the experiment with a different model version, the same two systems can be accurately represented. Whereas, difficulties appear for the same two ligands in both cases. The supplementary material contains the MM energy gradients, dipoles and quadrupole errors, as shown in Figure S12.

### 3.7 Prospective ML/MM MD Simulations



**Figure 25: MAEs for models used for prospective simulations.** The MAE is shown for the energy values  $E$  (left), as well the forces on the QM atoms  $F_{\text{QM}}$  (right) for the models used for the prospective simulations. MAEs are shown for training, validation, and test sets. Minimal models (blue color scheme) and precision models (red color scheme) are shown. Model versions 2 and 5 (both minimal models) are used for the unrestrained ML/MM MD runs. For the umbrella sampling of  $(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$ , model version 5 is used again as well as 8 (minimal model) and for  $(\text{PteroPhos})_x\text{Ni}(\text{benzaldehyde})$  model version 10 (minimal and precision model).



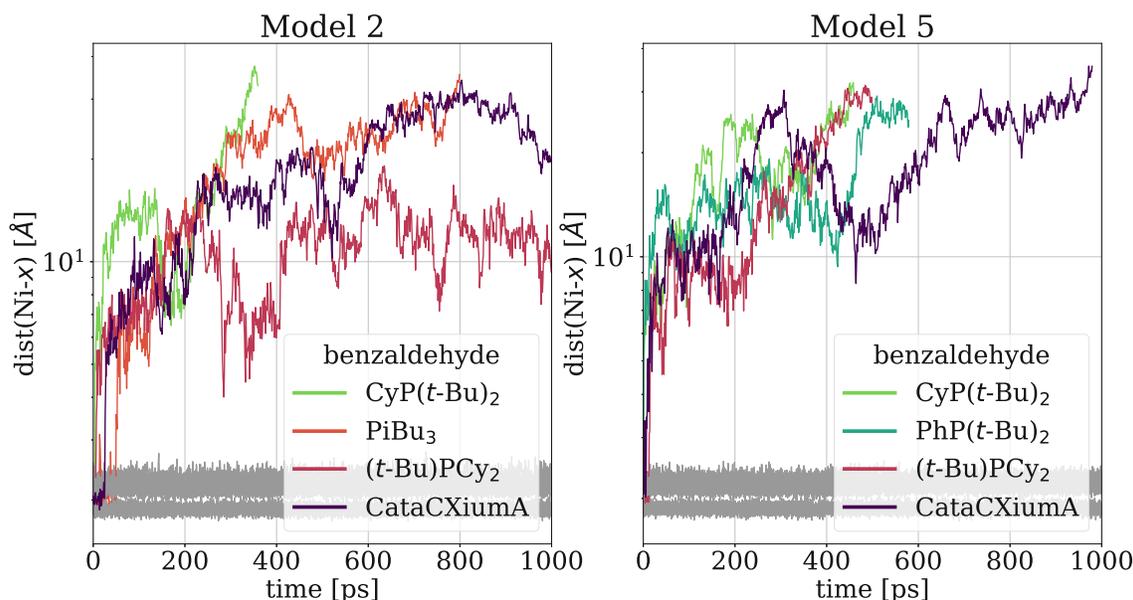
**Figure 26: Overall runtime of PyTorch to run 100 ps ML/MM MD for differently sized systems.** Total time measured in minutes, PyTorch needs to run 100 ps MD, corresponding to 40 000 energy and force calculations. The times are collected from calculations performed with model version 5.

The prospective simulations using ML/MM MD involved two distinct tasks: firstly, a stability test for all bisligated complexes, and secondly, a free energy calculation using umbrella sampling.

In order to investigate the stability of the bisligated  $(L)_2Ni(\text{benzaldehyde})$  complexes, unrestrained simulations were performed. Two minimal models, designated as versions 2 and 5, were employed. The total number of frames used to train, validate, and test these models are summarized in Table 2. The errors of model versions 2 and 5 were discussed in sections 3.6.1 and 3.6.2. For clarity, the MAEs for energies and gradients are shown together with all other models used for prospective simulations in Figure 25. It should be noted that model version 2 was trained on approximately 40% of the entire data available and model version 5 on approximately 40% of the data set 2, excluding data from the four ligands,  $PEt_3$ ,  $PhP(t-Bu)_2$ ,  $PCy_3$ , and  $CataCXiumA$  (see Table 3 and Figure 21).

For the umbrella sampling, model versions 5 and 8 were used as minimal models for  $(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$ , and model version 10 as minimal and precision models for  $(\text{PteroPhos})_x\text{Ni}(\text{benzaldehyde})$ . Model 5 was explained in the previous paragraph, and model 8 was first mentioned in section 3.6.2. The latter model is based on the same training data as model 5, but was augmented with 1.3% of the data from the four ligands,  $\text{PEt}_3$ ,  $\text{PhP}(t\text{-Bu})_2$ ,  $\text{PCy}_3$ , and  $\text{CataCXiumA}$  (see Table 8). For the complex with PteroPhos, two models were used, model 10 in minimal and precision versions. They were trained on the maximum amount of data possible, using 97% of all data for training and 3% for validation. The MAEs of these models are also shown in Figure 25.

### 3.7.1 Stability Test for Bisligated Complexes Using ML/MM MD

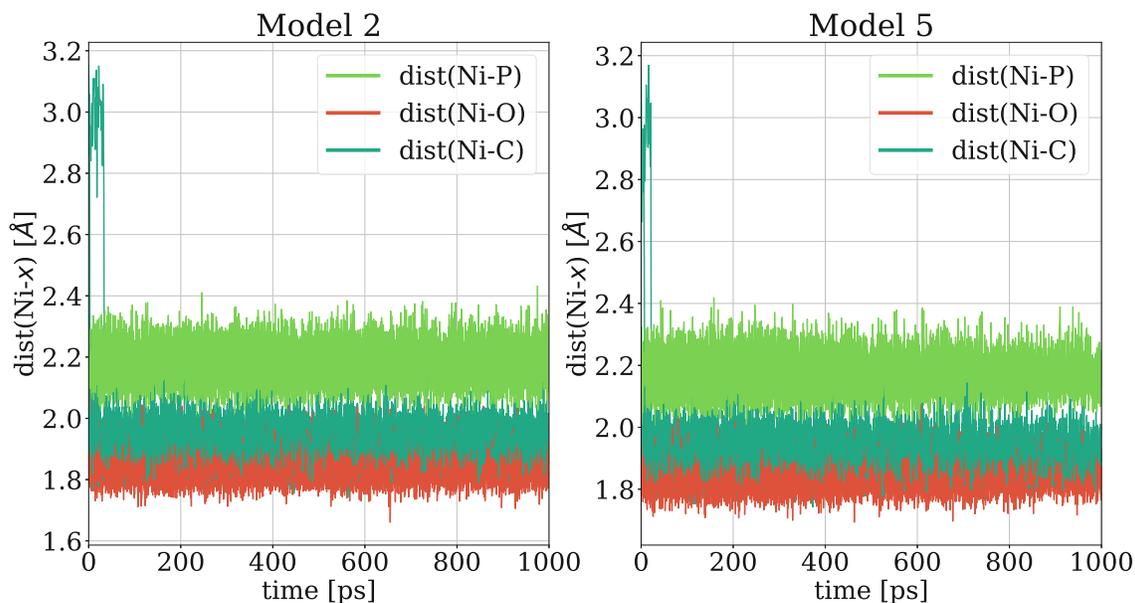


**Figure 27: Stability test with ML/MM MD simulations for all bisligated complexes.** Distances between nickel and both phosphine ligands as well as benzaldehyde ( $x = \text{P}_1, \text{P}_2$  for phosphine ligands or O for benzaldehyde) for all unrestrained ML/MM simulations for bisligated complexes. Simulations were performed using model versions 2 (left) and 5 (right). All distances are displayed in grey; only if distances exceeded a threshold of 4 Å, is their component colored. All leaving molecules were, without any exception, benzaldehyde. The ligand of the complex which dissociated is stated in the legend.

**Table 9: Stability results for bisligated complexes from ML/MM MD runs.** For complexes with different ligands, the result of the NMR experiment by Newman-Stonebraker et al.<sup>1</sup> is given (denoted exp. ligation state). **X** indicates that benzaldehyde dissociates in the ML/MM MD runs using model versions 2 and 5. If **X** is not indicated, the bisligated complex remained intact for 1 ns of ML/MM MD runs.

ligand	exp. ligation state <sup>1</sup>	dissociates (model 2)	dissociates (model 5)
CataCXiumA	1	<b>X</b>	<b>X</b>
CyP( <i>t</i> -Bu) <sub>2</sub>	1	<b>X</b>	<b>X</b>
PhP( <i>t</i> -Bu) <sub>2</sub>	1		<b>X</b>
<i>t</i> -BuPCy <sub>2</sub>	1 or 2	<b>X</b>	<b>X</b>
PCy <sub>3</sub>	2		
MeP( <i>t</i> -Bu) <sub>2</sub>	2		
PCy <sub>3</sub>	2		
CyPPh <sub>2</sub>	2		
PPh <sub>3</sub>	2		
PEt <sub>3</sub>	2		
PBn <sub>3</sub>	2		
P( <i>i</i> -Bu) <sub>3</sub>	2	<b>X</b>	
PteroPhos	2		

Before discussing the results of all unrestrained ML/MM MD simulations, it is important to highlight the significant time savings achieved by using the ML model instead of a QM method. The time required by PyTorch to perform 100 ps ML/MM MD runs is shown in Figure 26 for the minimal model, model version 5. This plot can be compared to the time taken by xTB, shown in Figure 13. The speedup ranges from 2.7 (complex with PMe<sub>3</sub>) to 91 (complex with PteroPhos). For CataCXiumA, the last complex included in the training data, the speedup factor is 20. Since the ML model approximates the values obtained from ORCA DFT calculations, one can also compare the time taken by this program. The time required for a single DFT calculation varies between 21 min (complex with PMe<sub>3</sub>) and 330 min (complex with CataCXiumA). To obtain the speedup over the ML model, one needs to multiply the time values of a single DFT energy calculation by a factor of 200 000, the total number of steps performed in 100 ps MD. The speedup then ranges from  $5.5 \times 10^4$  to  $5.7 \times 10^5$  for (PMe<sub>3</sub>)<sub>2</sub>Ni(benzaldehyde) to (CataCXiumA)<sub>2</sub>Ni(benzaldehyde). ORCA calculations were not performed for the PteroPhos complex.



**Figure 28: Stability test with ML/MM MD simulations for all monoligated complexes.** Distances between nickel and one phosphine ligands as well as benzaldehyde ( $x = \text{P}$  for phosphine ligand or O and C for atoms of benzaldehyde coordinated to nickel) for all unrestrained ML/MM simulations for monoligated complexes. Simulations were performed using model versions 2 (left) and 5 (right).

The results for the stability of all bisligated complexes are presented in Figure 27. In Figure 27, the distances between the nickel atom and the ligands, including benzaldehyde, are shown for both runs using different models, model version 2 and model version 5. It can be observed that a total of five bisligated complexes are unstable either in both runs or in only one of the runs. In all cases, the benzaldehyde leaves the complex. These results can again be compared to the ligation state experiments of Newman-Stonebraker et al., as can be seen in Table 9.<sup>1</sup> The table shows that those complexes that have been experimentally determined to be monoligated tend to dissociate more frequently. In order to ascertain a free energy minimum for the stable bisligated complexes, it is necessary to determine the free energy profile when transitioning to the monoligated version. The energetic minimum was determined for two complexes in section 3.8.

Figure 28 illustrates the distances between nickel and the phosphine ligand and benzaldehyde for the monoligated complex. It can be observed that all complexes

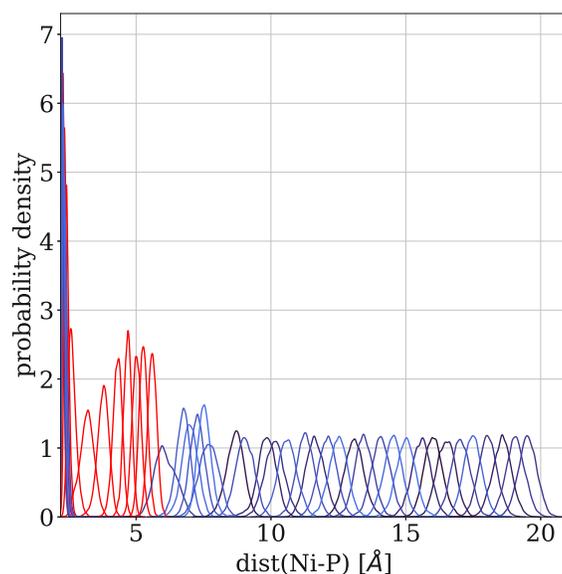
remain intact. Consequently, in cases where the bisligated version dissociates, it can be concluded that the monoligated complex is the favored configuration. Therefore, calculating a free energy profile is unnecessary.

With regard to the bisligated complexes, it is noteworthy that benzaldehyde is consistently the leaving group. This suggests that the  $L_2Ni$  configuration is also stable. Other potential reasons for the dissociation of benzaldehyde may be related to its hapticity. The original reaction studied was that of a monoligated versus bisligated complex, with phosphine serving as the ligand. It was assumed that benzaldehyde would always act as a  $\eta^2$  ligand. As demonstrated in section 3.3, the configurations including benzaldehyde as a hapto-1 ligand were particularly frequent in the training data for the five complexes that dissociate. Furthermore, the initial configuration utilized in this section was the final frame of the training data generation process, which included the  $\eta^1$  benzaldehyde ligand for the same five complexes. During the simulation, a transition of benzaldehyde from hapto-1 to hapto-2 ligand was observed, although it did not occur with great frequency. It would be of interest to repeat the experiment with a hand-picked configuration from the training data, where all complexes have a  $\eta^2$  benzaldehyde ligand.

Finally, Figure 27 shows comparable results for both model versions 2 and 5. Although model version 5 was not trained on data including the ligands  $PEt_3$ ,  $PhP(t-Bu)_2$ ,  $PCy_3$ , and  $CataCXiumA$ , it yielded the same results as model 2, which was trained on the entire data set. Notably, the complex with  $PteroPhos$  was also simulated for 1 ns as both a monoligated and bisligated complex, which would have taken about 150 days using xTB (see Figure 13).

### 3.8 Free Energy Calculation Using ML/MM MD

For  $(PCy_3)_xNi(\text{benzaldehyde})$  and  $(PteroPhos)_xNi(\text{benzaldehyde})$  complexes, the free energy profile was calculated when moving from the bisligated to the monoligated version. This was achieved using the weighted histogram analysis method (WHAM) and the distribution of nickel-phosphorus distances for all umbrellas of the restrained ML/MM MD simulations. Figure 29 displays the probability densities for the individual umbrellas for the complex  $(PteroPhos)_xNi(\text{benzaldehyde})$ . It is evident that the initial sampling was not evenly distributed (Figure 29, shown in blue). However, by incorporating additional MD runs into the gap, the discrepancy could be rectified (Figure 29, shown in red). For this purpose, additional umbrellas were added in the region between 2.5 Å and 5.2 Å. For the simulation using the precision model

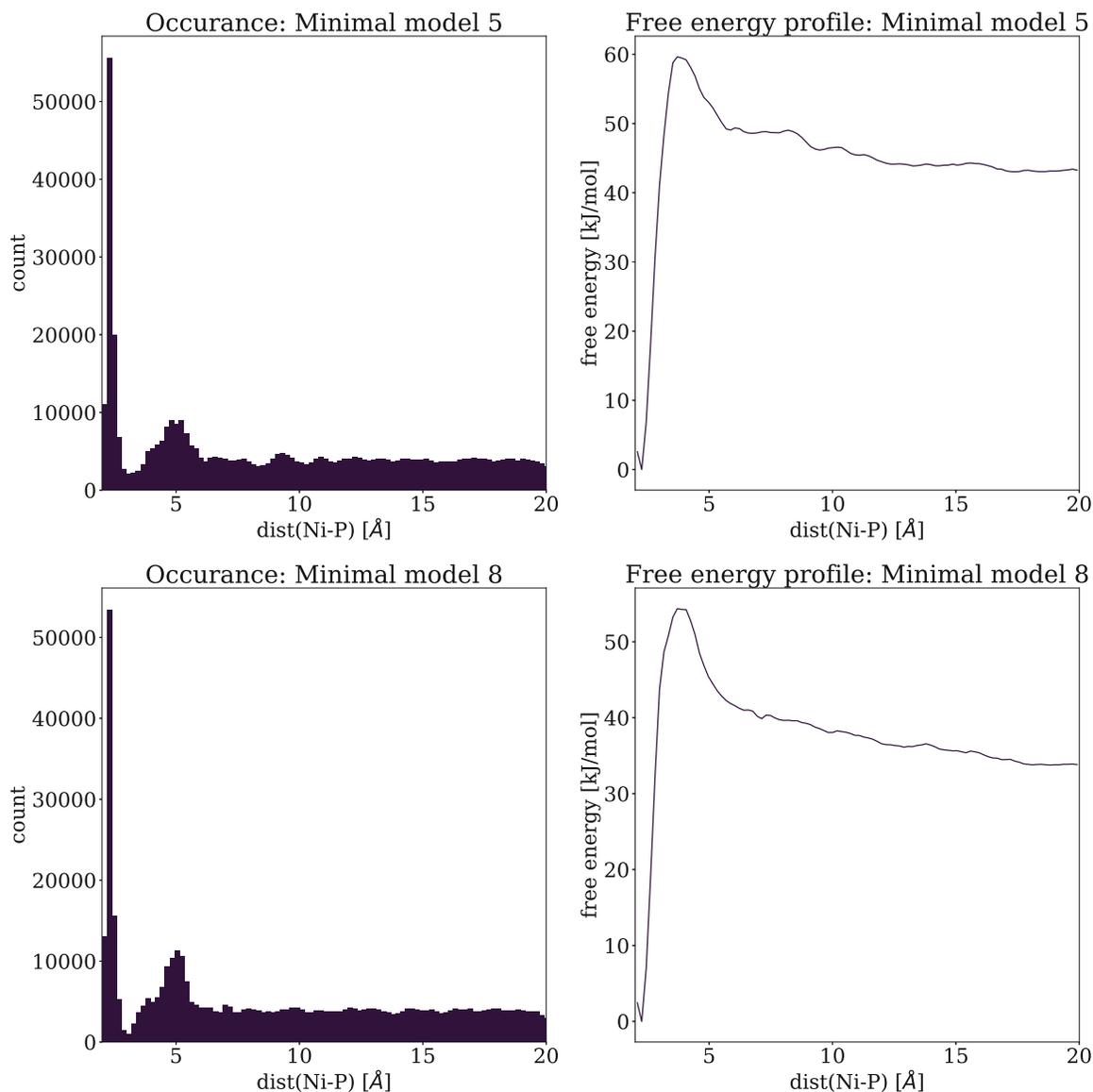


**Figure 29: Example of probability densities for individual umbrellas.** Probability densities for all umbrellas for the system (PteroPhos)<sub>x</sub>Ni(benzaldehyde) simulated using precision model version 10. Initial 37 umbrellas with increment 0.5 Å and force constant 2 000 kJ/(mol nm<sup>2</sup>) (blue lines) as well as additional 11 umbrellas using force constant 10 000 kJ/(mol nm<sup>2</sup>) (red lines) are shown.

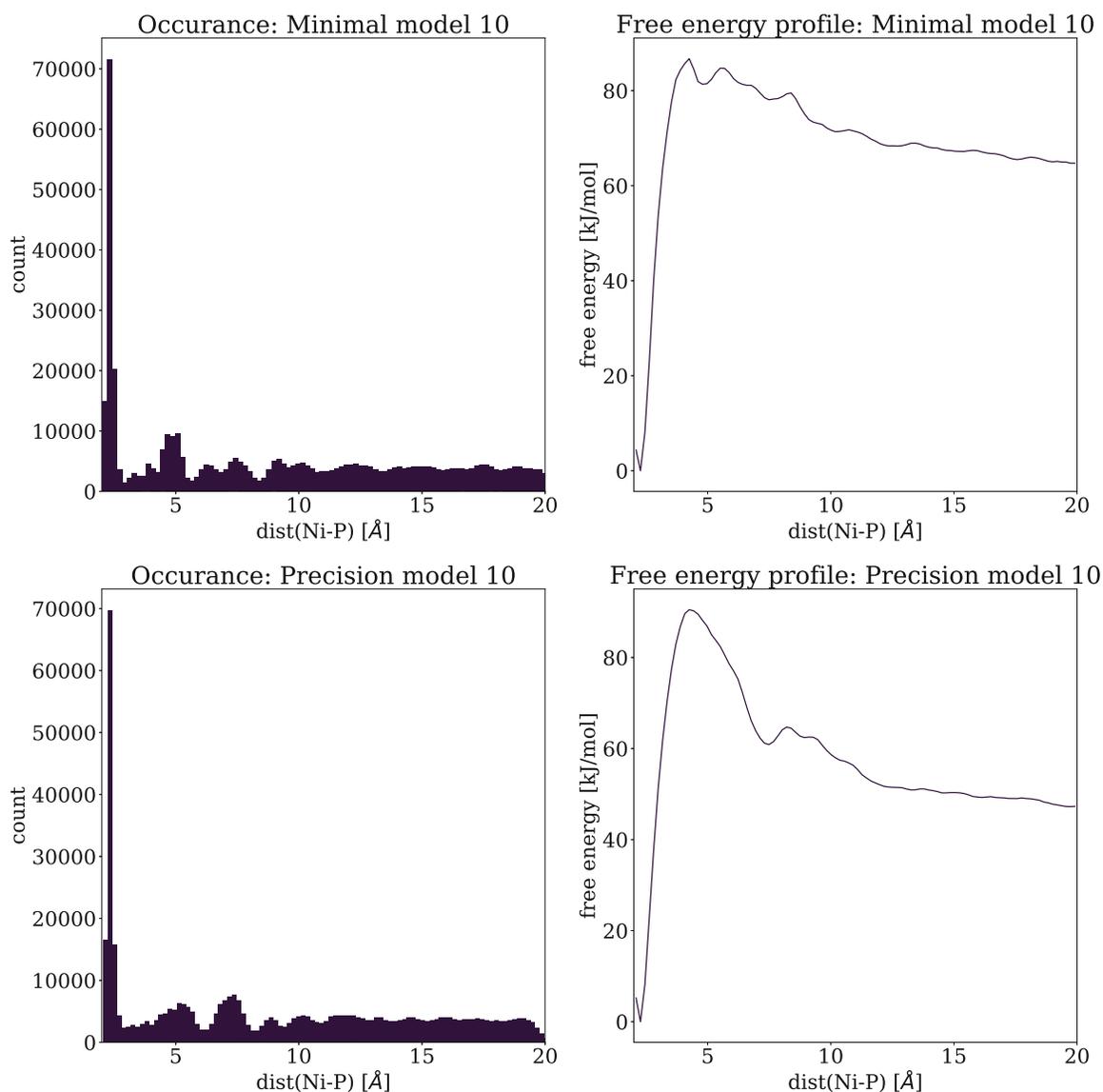
for (PteroPhos)<sub>x</sub>Ni(benzaldehyde), one extra umbrella was added at 5.5 Å. In order to fill this gap, additional sampling was performed by using a force constant of 10 000 kJ/(mol nm<sup>2</sup>) and adding additional restrained MD runs every 0.3 Å.

The final total occurrence of all distances and the final free energy profile are displayed in Figures 30 and 31 for the systems (PCy<sub>3</sub>)<sub>x</sub>Ni(benzaldehyde), as well as (PteroPhos)<sub>x</sub>Ni(benzaldehyde). It is important to note that the distribution of distances should be as uniform as possible. This could be improved by additional sampling. More importantly, there are no gaps visible. For all model versions and both systems, it is clearly visible that the L<sub>2</sub>Ni(benzaldehyde) is more stable than the L<sub>1</sub>Ni(benzaldehyde). This is consistent with the NMR experiments conducted by Newman-Stonebraker et al.<sup>1</sup>, which demonstrated that the (PCy<sub>3</sub>)<sub>2</sub>Ni(4-fluorobenzaldehyde) and (PteroPhos)<sub>2</sub>Ni(4-fluorobenzaldehyde) complexes were the predominant species in the ligation state experiment.

The differences in free energy between the monoligated and bisligated complexes are



**Figure 30: Free energy profile for  $(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$ .** The free energy profile for  $(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$  when moving one of the  $\text{PCy}_3$  ligands stepwise away until a total of 20 Å distance with respect to the nickel atom. Two different minimal models, model version 5 (upper plots) and model version 8 (lower plots) are shown. On the left the occurrence of the distances is showed and on the right the final free energy profile.



**Figure 31: Free energy profile for  $(\text{PteroPhos})_x\text{Ni}(\text{benzaldehyde})$ .** The free energy profile for  $(\text{PteroPhos})_x\text{Ni}(\text{benzaldehyde})$  when moving one of the PteroPhos ligands stepwise away until a total of 20 Å distance with respect to the nickel atom. The output of minimal model version 10 (upper plots) and precision version 10 (lower plots) are shown. On the left the occurrence of the distances is showed and on the right the final free energy profile.

presented in Table 10. While it is evident that the bisligated complex is energetically favored, it remains unclear which free energy difference is more significant. To resolve this, further experiments would be necessary, along with a verification of the convergence of the sampling.

However, model version 5, which was not trained on data of  $(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$  predicted the same qualitative output as model version 8, which was augmented with data of the complex. Furthermore, both models result in the same outcome for the  $(\text{PteroPhos})_x\text{Ni}(\text{benzaldehyde})$  complex, which is in agreement with literature.<sup>1</sup>

**Table 10: Free energy differences from umbrella sampling.** Free energy difference between monoligated and bisligated state for two complexes, with ligand  $\text{PCy}_3$  and PteroPhos. Umbrella sampling was conducted twice using two different model versions 5 and 10 as minimal model.

		$\Delta G$ [kJ/mol]
$(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$	Minimal model 5	-43.2
	Minimal model 8	-33.8
$(\text{PteroPhos})_x\text{Ni}(\text{benzaldehyde})$	Minimal model 10	-65.2
	Precision model 10	-47.9

## 4 Conclusion and Outlook

The AMPv2 model, a GNN with physically motivated features, uses anisotropic message passing and anisotropic states derived from the Cartesian multipole formalism. This 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 accuracy similar to DFT. This was accomplished by creating MD trajectories using a semi-empirical approach and re-evaluating the energies and multipoles with the DFT method for the training data. The ligation states of 20 distinct phosphine ligands L, which are relevant for cross-coupling reactions, were investigated in the context of  $L_2Ni(\text{benzaldehyde})$  complexes.

A series of prospective simulations were conducted utilizing a range of AMPv2 models. All bisligated and monoligated transition metal complexes were subjected to simulations of 1 ns, yielding a combined simulation time of 195 ns. The aforementioned simulations were found to be stable, in that no extreme forces were observed, and only realistic behavior occurred. Stable MD simulations are often not given by state-of-the-art ML FFs even though predictive errors on forces and energies are sufficient.<sup>87</sup> Remarkably, stable simulations were even achieved when benzaldehyde dissociated from the complex, which was not sampled in the training data. These findings suggest that the AMPv2 model exhibits behaviour comparable to a QM method.

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. Compared to the semi-empirical method, the speedup ranged from 2.7 to 20. It should be noted that these results only include all complexes used in the training data, with the ligand CataCXiumA being the largest complex, encompassing 143 atoms in total. For the complex with PteroPhos, comprising 385 atoms, the speed-up factor for the comparison to the semi-empirical method rose to 91. These figures demonstrate how ML/MM can facilitate the investigation of large systems otherwise inaccessible at this level of theory.

Furthermore, the ML model demonstrates an optimal linear correlation and predictive capacity, rendering it a viable substitute for the DFT method. When a model with 606 568 parameter was trained with 42 000 frames, a good bias-variance trade-off was achieved. For this model, the mean absolute error (MAE) for the energies of the QM atoms is 2.2 kJ/mol and the forces acting on the QM atoms are 1.1 kJ/(mol Å).

Generally, the MAE for the energies, except for one special case, is within chemical accuracy, defined as 1 kcal/mol (4.184 kJ/mol).

The transferability of an AMPv2 model was demonstrated to be relatively robust when tested with new data not included in the training. This robustness was observed across four distinct ligands. For ligands containing subunits that were prevalent in the training data set, the model achieved remarkable accuracy. However, for systems with unique chemical subunits, like the adamantane group in CataCXiumA, the mean absolute error (MAE) tripled, reaching a maximum MAE of 6.2 kJ/mol. However, the error could be reduced by 1.8 kJ/mol if only eight frames of CataCXiumA were included in the training set. This significant improvement prompts the question of whether the same effect can be achieved with frames collected from a shorter trajectory. This approach could be used to learn on larger complexes where training data is limited.

Finally, it is important to compare the results to the experimental ligation state NMR results reported by Newman Stonebraker et al.<sup>1</sup> The authors presented the results of the binary ligation state for 13 comparable complexes: three were monoligated, nine were bisligated, and one exhibited both states with equal prevalence. For the three monoligated complexes, the benzaldehyde dissociated for most runs during the prospective simulations (2 repeats; 5/6 times dissociated). Consequently, these complexes are correctly categorized as monoligated complexes, as the monoligated version did not dissociate. In contrast, nearly all bisligated complexes remained intact during the simulation (2 repeats; 17/18 times did not dissociate). For these complexes it is unclear which ligation state is energetically preferred. Umbrella sampling was used for the complexes with the ligands PCy<sub>3</sub> and PteroPhos for deriving a free energy profile. The free energy difference  $\Delta G$  was  $-33.8$  and  $-43.2$  kJ/mol for the PCy<sub>3</sub> complex and  $-47.9$  and  $-65.2$  kJ/mol for the PteroPhos complex. The bisligated state is energetically preferred, in accordance with the experimental findings. For the PCy<sub>3</sub> ligand comparable results for the free energy difference were yielded even if no data of this complex was included in the training data. Additionally, the correct ligation state for the PteroPhos ligand was successfully predicted, despite the challenging example presented by the ligand's large cone angle and extreme size.<sup>1</sup>

For all prospective simulations, the total number of repeats was limited to a maximum of two due to computational cost, which makes statistical analysis challenging. Additionally, in the prospective simulations, benzaldehyde was identified as the leaving molecule in the event of dissociation. This finding cannot be contradicted

by experimental data.<sup>1</sup> However, it is unclear whether the under-representation of hapto-2 benzaldehyde is the cause of the observed behavior or if the  $L_2Ni$  is an experimentally stable species.

In conclusion, the AMPv2 model can be used in ML/MM simulations to investigate the ligation states of various phosphine ligands in transition-metal complexes, which is crucial for understanding of cross-coupling reactions. The scheme enables the inclusion of explicit solvent effects and an approximate QM description at the DFT level for the complex itself, thus facilitating the study of reactions. Once the model is trained, the behavior of phosphine ligands can be investigated with significantly reduced computational resources in molecular dynamics simulations, providing dynamic information that static descriptors are unable to provide.<sup>11,15</sup> Its transferability also enables study of new ligands, especially those with similar chemical subunits. Ligations states are also crucial in palladium catalyzed reactions, which could be studied with the same approach.<sup>1,88,89</sup> Further, one could also consider expanding the (QM)ML/MM scheme to other use cases like conformational search or the investigation of other steps of the catalytic cycle of cross coupling reactions, like oxidative addition, transmetalation, or reductive elimination.<sup>90–92</sup>

To conclude in molecular models a tradeoff between sampling, system size, and method accuracy must be always taken into account.<sup>93</sup> Using a machine learning model, like AMPv2, which accurately approximates a high level *ab initio* method takes the accuracy out of the equation, allowing for more sampling and larger systems. This opens up many new opportunities, especially for large biomolecular systems.

## References

- [1] S. 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] S. J. Firsan, V. Sivakumar, and T. J. Colacot. “Emerging trends in cross-coupling: twelve-electron-based L1Pd (0) catalysts, their mechanism of action, and selected applications”. In: *Chemical Reviews* 122.23 (2022), pp. 16983–17027.
- [3] C. C. Johansson Seechurn et al. “Palladium-catalyzed cross-coupling: a historical contextual perspective to the 2010 Nobel Prize”. In: *Angewandte Chemie International Edition* 51.21 (2012), pp. 5062–5085.
- [4] N. Miyaura and A. Suzuki. “Palladium-catalyzed cross-coupling reactions of organoboron compounds”. In: *Chemical reviews* 95.7 (1995), pp. 2457–2483.
- [5] S. Z. Tasker, E. A. Standley, and T. F. Jamison. “Recent advances in homogeneous nickel catalysis”. In: *Nature* 509.7500 (2014), pp. 299–309.
- [6] D. G. Brown and J. Bostrom. “Analysis of past and present synthetic methodologies on medicinal chemistry: where have all the new reactions gone? Miniperspective”. In: *Journal of medicinal chemistry* 59.10 (2016), pp. 4443–4458.
- [7] A. Zapf, A. Ehrentraut, and M. Beller. “A new highly efficient catalyst system for the coupling of nonactivated and deactivated aryl chlorides with arylboronic acids”. In: *Angewandte Chemie International Edition* 39.22 (2000), pp. 4153–4155.
- [8] J. P. Wolfe et al. “Highly active palladium catalysts for Suzuki coupling reactions”. In: *Journal of the American Chemical Society* 121.41 (1999), pp. 9550–9561.
- [9] K. Wu and A. G. Doyle. “Parameterization of phosphine ligands demonstrates enhancement of nickel catalysis via remote steric effects”. In: *Nature chemistry* 9.8 (2017), pp. 779–784.
- [10] E. C. Hansen et al. “New ligands for nickel catalysis from diverse pharmaceutical heterocycle libraries”. In: *Nature chemistry* 8.12 (2016), pp. 1126–1130.
- [11] D. J. Durand and N. Fey. “Computational ligand descriptors for catalyst design”. In: *Chemical reviews* 119.11 (2019), pp. 6561–6594.

- [12] T. Gensch et al. “A comprehensive discovery platform for organophosphorus ligands for catalysis”. In: *Journal of the American Chemical Society* 144.3 (2022), pp. 1205–1217.
- [13] W. F. Van Gunsteren et al. “Biomolecular modeling: goals, problems, perspectives”. In: *Angewandte Chemie International Edition* 45.25 (2006), pp. 4064–4092.
- [14] W. F. Van Gunsteren and H. J. Berendsen. “Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry”. In: *Angewandte Chemie International Edition in English* 29.9 (1990), pp. 992–1023.
- [15] S. A. Hollingsworth and R. O. Dror. “Molecular dynamics simulation for all”. In: *Neuron* 99.6 (2018), pp. 1129–1143.
- [16] B. Leimkuhler and C. Matthews. “Molecular dynamics”. In: *Interdisciplinary applied mathematics* 39 (2015), p. 443.
- [17] W. F. Van Gunsteren and H. J. Berendsen. “A leap-frog algorithm for stochastic dynamics”. In: *Molecular Simulation* 1.3 (1988), pp. 173–185.
- [18] H. M. Senn and W. Thiel. “QM/MM methods for biomolecular systems”. In: *Angewandte Chemie International Edition* 48.7 (2009), pp. 1198–1229.
- [19] H. Lin and D. G. Truhlar. “QM/MM: what have we learned, where are we, and where do we go from here?” In: *Theoretical Chemistry Accounts* 117 (2007), pp. 185–199.
- [20] G. Groenhof. “Introduction to QM/MM simulations”. In: *Biomolecular simulations: methods and protocols* (2013), pp. 43–66.
- [21] J. Gao. “Hybrid quantum and molecular mechanical simulations: an alternative avenue to solvent effects in organic chemistry”. In: *Accounts of chemical research* 29.6 (1996), pp. 298–305.
- [22] M. J. Field, P. A. Bash, and M. Karplus. “A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations”. In: *Journal of computational chemistry* 11.6 (1990), pp. 700–733.
- [23] C. Caleman et al. “Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant”. In: *Journal of chemical theory and computation* 8.1 (2012), pp. 61–74.

- [24] A. Warshel and M. Levitt. “Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme”. In: *Journal of molecular biology* 103.2 (1976), pp. 227–249.
- [25] S. Riniker. “Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: An overview”. In: *Journal of chemical information and modeling* 58.3 (2018), pp. 565–578.
- [26] P. S. Nerenberg and T. Head-Gordon. “New developments in force fields for biomolecular simulations”. In: *Current opinion in structural biology* 49 (2018), pp. 129–138.
- [27] J. A. Keith et al. “Combining machine learning and computational chemistry for predictive insights into chemical systems”. In: *Chemical reviews* 121.16 (2021), pp. 9816–9872.
- [28] M. P. Johansson, V. R. Kaila, and D. Sundholm. “Ab initio, density functional theory, and semi-empirical calculations”. In: *Biomolecular Simulations: Methods and Protocols* (2013), pp. 3–27.
- [29] C. Bannwarth, S. Ehlert, and S. Grimme. “GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions”. In: *Journal of chemical theory and computation* 15.3 (2019), pp. 1652–1671.
- [30] K. Huang. *Statistical mechanics*. John Wiley & Sons, 2008.
- [31] J. Kästner. “Umbrella sampling”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.6 (2011), pp. 932–942.
- [32] Y. Sugita and Y. Okamoto. “Replica-exchange molecular dynamics method for protein folding”. In: *Chemical physics letters* 314.1-2 (1999), pp. 141–151.
- [33] P. Liu et al. “Replica exchange with solute tempering: A method for sampling biological systems in explicit water”. In: *Proceedings of the National Academy of Sciences* 102.39 (2005), pp. 13749–13754.
- [34] T. Huber, A. E. Torda, and W. F. V. Gunsteren. “Local elevation: a method for improving the searching properties of molecular dynamics simulation”. In: *Journal of computer-aided molecular design* 8 (1994), pp. 695–708.
- [35] L. Zheng et al. “Molecular dynamics and simulation”. In: *Encyclopedia of Bioinformatics and Computational Biology* (2019).
- [36] G. M. Torrie and J. P. Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *Journal of computational physics* 23.2 (1977), pp. 187–199.

- [37] S. Kumar et al. “The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”. In: *Journal of computational chemistry* 13.8 (1992), pp. 1011–1021.
- [38] S. Kumar et al. “Multidimensional free-energy calculations using the weighted histogram analysis method”. In: *Journal of Computational Chemistry* 16.11 (1995), pp. 1339–1350.
- [39] K. Hornik, M. Stinchcombe, and H. White. “Multilayer feedforward networks are universal approximators”. In: *Neural networks* 2.5 (1989), pp. 359–366.
- [40] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444.
- [41] O. T. Unke et al. “Machine learning force fields”. In: *Chemical Reviews* 121.16 (2021), pp. 10142–10186.
- [42] J. Behler and M. Parrinello. “Generalized neural-network representation of high-dimensional potential-energy surfaces”. In: *Physical review letters* 98.14 (2007), p. 146401.
- [43] J. S. Smith, O. Isayev, and A. E. Roitberg. “ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost”. In: *Chemical science* 8.4 (2017), pp. 3192–3203.
- [44] K. Yao et al. “The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics”. In: *Chemical science* 9.8 (2018), pp. 2261–2269.
- [45] F. Scarselli et al. “The graph neural network model”. In: *IEEE transactions on neural networks* 20.1 (2008), pp. 61–80.
- [46] J. Gilmer et al. “Neural message passing for quantum chemistry”. In: *International conference on machine learning*. PMLR. 2017, pp. 1263–1272.
- [47] K. Schütt et al. “SchNet: A continuous-filter convolutional neural network for modeling quantum interactions”. In: *Advances in neural information processing systems* 30 (2017).
- [48] O. T. Unke and M. Meuwly. “PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges”. In: *Journal of chemical theory and computation* 15.6 (2019), pp. 3678–3693.
- [49] M. Thürlmann and S. Riniker. “Anisotropic Message Passing: Graph Neural Networks with Directional and Long-Range Interactions”. In: *The Eleventh International Conference on Learning Representations*. 2022.

- [50] M. Thürlmann, L. Bösel, and S. Riniker. “Learning atomic multipoles: prediction of the electrostatic potential with equivariant graph neural networks”. In: *Journal of Chemical Theory and Computation* 18.3 (2022), pp. 1701–1710.
- [51] C. J. Burnham and N. J. English. “A new relatively simple approach to multipole interactions in either spherical harmonics or cartesian, suitable for implementation into Ewald Sums”. In: *International Journal of Molecular Sciences* 21.1 (2019), p. 277.
- [52] F. Neese. “The ORCA program system”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2.1 (2012), pp. 73–78.
- [53] G. Landrum. “RDKit: Open-source cheminformatics 2023.09.1”. In: (2020).
- [54] P. Eastman et al. “OpenMM 7: Rapid development of high performance algorithms for molecular dynamics”. In: *PLoS computational biology* 13.7 (2017), e1005659.
- [55] S. Boothroyd et al. *Development and benchmarking of open force field 2.0.0*. 2022.
- [56] R. Salomon-Ferrer, D. A. Case, and R. C. Walker. “An overview of the Amber biomolecular simulation package”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3.2 (2013), pp. 198–210.
- [57] A. P. Eichenberger et al. “GROMOS++ software for the analysis of biomolecular simulation trajectories”. In: *Journal of chemical theory and computation* 7.10 (2011), pp. 3379–3390.
- [58] M. R. Shirts et al. “Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset”. In: *Journal of computer-aided molecular design* 31 (2017), pp. 147–161.
- [59] M. T. L. et al. “PyGromosTools V3.0”. In: (2022).
- [60] K. Meier, N. Schmid, and W. F. Van Gunsteren. “Interfacing the GROMOS (bio) molecular simulation software to quantum-chemical program packages”. In: *Journal of computational chemistry* 33.26 (2012), pp. 2108–2117.
- [61] N. Schmid et al. “Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation”. In: *Computer Physics Communications* 183.4 (2012), pp. 890–903.
- [62] GROMOS. *Volume 4: Data Structures and Formats*. 2021.
- [63] I. G. Tironi et al. “A generalized reaction field method for molecular dynamics simulations”. In: *The Journal of chemical physics* 102.13 (1995), pp. 5451–5459.

- [64] S. Lata, K. Singh, and S. Ahlawat. “A study of dielectric properties and refractive indices of aniline+ benzene,+ toluene,+ o-xylene, and+ p-xylene at 298.15 K”. In: *Journal of Molecular Liquids* 147.3 (2009), pp. 191–197.
- [65] S. Kim et al. “PubChem substance and compound databases”. In: *Nucleic acids research* 44.D1 (2016), pp. D1202–D1213.
- [66] H. J. C. Berendsen et al. “Molecular dynamics with coupling to an external bath”. In: *The Journal of chemical physics* 81.8 (1984), pp. 3684–3690.
- [67] N. Mardirossian and M. Head-Gordon. “ $\omega$ B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 non-local correlation”. In: *The Journal of chemical physics* 144.21 (2016).
- [68] A. Najibi and L. Goerigk. “DFT-D4 counterparts of leading meta-generalized-gradient approximation and hybrid density functionals for energetics and geometries”. In: *Journal of Computational Chemistry* 41.30 (2020), pp. 2562–2572.
- [69] M. A. Iron and T. Janes. “Evaluating transition metal barrier heights with the latest density functional theory exchange–correlation functionals: The MOBH35 benchmark database”. In: *The Journal of Physical Chemistry A* 123.17 (2019), pp. 3761–3781.
- [70] R. Grotjahn and M. Kaupp. “A Look at Real-World Transition-Metal Thermochemistry and Kinetics with Local Hybrid Functionals”. In: *Israel Journal of Chemistry* 63.7-8 (2023), e202200021.
- [71] S. Dohm et al. “Comprehensive thermochemical benchmark set of realistic closed-shell metal organic reactions”. In: *Journal of chemical theory and computation* 14.5 (2018), pp. 2596–2608.
- [72] E. Caldeweyher, C. Bannwarth, and S. Grimme. “Extension of the D3 dispersion coefficient model”. In: *The Journal of chemical physics* 147.3 (2017).
- [73] E. Caldeweyher et al. “A generally applicable atomic-charge dependent London dispersion correction scheme”. In: (2019).
- [74] F. Weigend and R. Ahlrichs. “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy”. In: *Physical Chemistry Chemical Physics* 7.18 (2005), pp. 3297–3305.
- [75] F. Weigend. “Accurate Coulomb-fitting basis sets for H to Rn”. In: *Physical chemistry chemical physics* 8.9 (2006), pp. 1057–1065.

- [76] F. Neese et al. “Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A ‘chain-of-spheres’ algorithm for the Hartree–Fock exchange”. In: *Chemical Physics* 356.1-3 (2009), pp. 98–109.
- [77] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [78] A. Grossfield. “An implementation of WHAM: the Weighted Histogram Analysis Method, Version 2.0.10”. In: (2021).
- [79] A. v. Bondi. “van der Waals volumes and radii”. In: *The Journal of physical chemistry* 68.3 (1964), pp. 441–451.
- [80] R. S. Rowland and R. Taylor. “Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der Waals radii”. In: *The Journal of Physical Chemistry* 100.18 (1996), pp. 7384–7391.
- [81] M. Mantina et al. “Consistent van der Waals radii for the whole main group”. In: *The Journal of Physical Chemistry A* 113.19 (2009), pp. 5806–5812.
- [82] I. Y. Chernyshov, I. V. Ananyev, and E. A. Pidko. “Revisiting van der Waals radii: From comprehensive structural analysis to knowledge-based classification of interatomic contacts”. In: *ChemPhysChem* 21.5 (2020), pp. 370–376.
- [83] H. H. et al. “Accurate simulation of surfaces and interfaces of face-centered cubic metals using 12-6 and 9-6 Lennard-Jones potentials”. In: *The Journal of Physical Chemistry C* 112.44 (2008), pp. 17281–17290.
- [84] K. Daun, J. Titantah, and M. Karttunen. “Molecular dynamics simulation of thermal accommodation coefficients for laser-induced incandescence sizing of nickel particles”. In: *Applied Physics B* 107 (2012), pp. 221–228.
- [85] A. K. Rappé et al. “UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations”. In: *Journal of the American chemical society* 114.25 (1992), pp. 10024–10035.
- [86] S. Geman, E. Bienenstock, and R. Doursat. “Neural networks and the bias/variance dilemma”. In: *Neural computation* 4.1 (1992), pp. 1–58.
- [87] X. Fu et al. “Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations”. In: *arXiv preprint arXiv:2210.07237* (2022).
- [88] F. Schoenebeck and K. Houk. “Ligand-controlled regioselectivity in palladium-catalyzed cross coupling reactions”. In: *Journal of the American Chemical Society* 132.8 (2010), pp. 2496–2497.

- [89] M. Garcia-Melchor et al. “Computational perspective on Pd-catalyzed C–C cross-coupling reaction mechanisms”. In: *Accounts of chemical research* 46.11 (2013), pp. 2626–2634.
- [90] M. Besora et al. “The importance of conformational search: a test case on the catalytic cycle of the Suzuki–Miyaura cross-coupling”. In: *Theoretical Chemistry Accounts* 128 (2011), pp. 639–646.
- [91] A. A. Braga et al. “Computational characterization of the role of the base in the Suzuki–Miyaura cross-coupling reaction”. In: *Journal of the American Chemical Society* 127.25 (2005), pp. 9298–9307.
- [92] A. A. Braga, G. Ujaque, and F. Maseras. “A DFT study of the full catalytic cycle of the Suzuki–Miyaura cross-coupling on a model system”. In: *Organometallics* 25.15 (2006), pp. 3647–3658.
- [93] V. Vennelakanti et al. “Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins”. In: *Current opinion in structural biology* 72 (2022), pp. 9–17.

## Supplementary Information

**Table S1: Distances between nickel and ligands for MD runs for determining LJ parameters.** Statistical data for distances between Ni and neighboring atoms - O, C of benzaldehyde and P of phosphine ligands. The values for phosphorus are separated into values for the monoligated,  $x=1$ , and bisligated,  $x=2$ ,  $L_x\text{Ni}(\text{benzaldehyde})$  complex. Displayed are the minimum value, the 25% quantile (Q1), the median (Q2), the 75% quantile (Q3) and the maximum value.

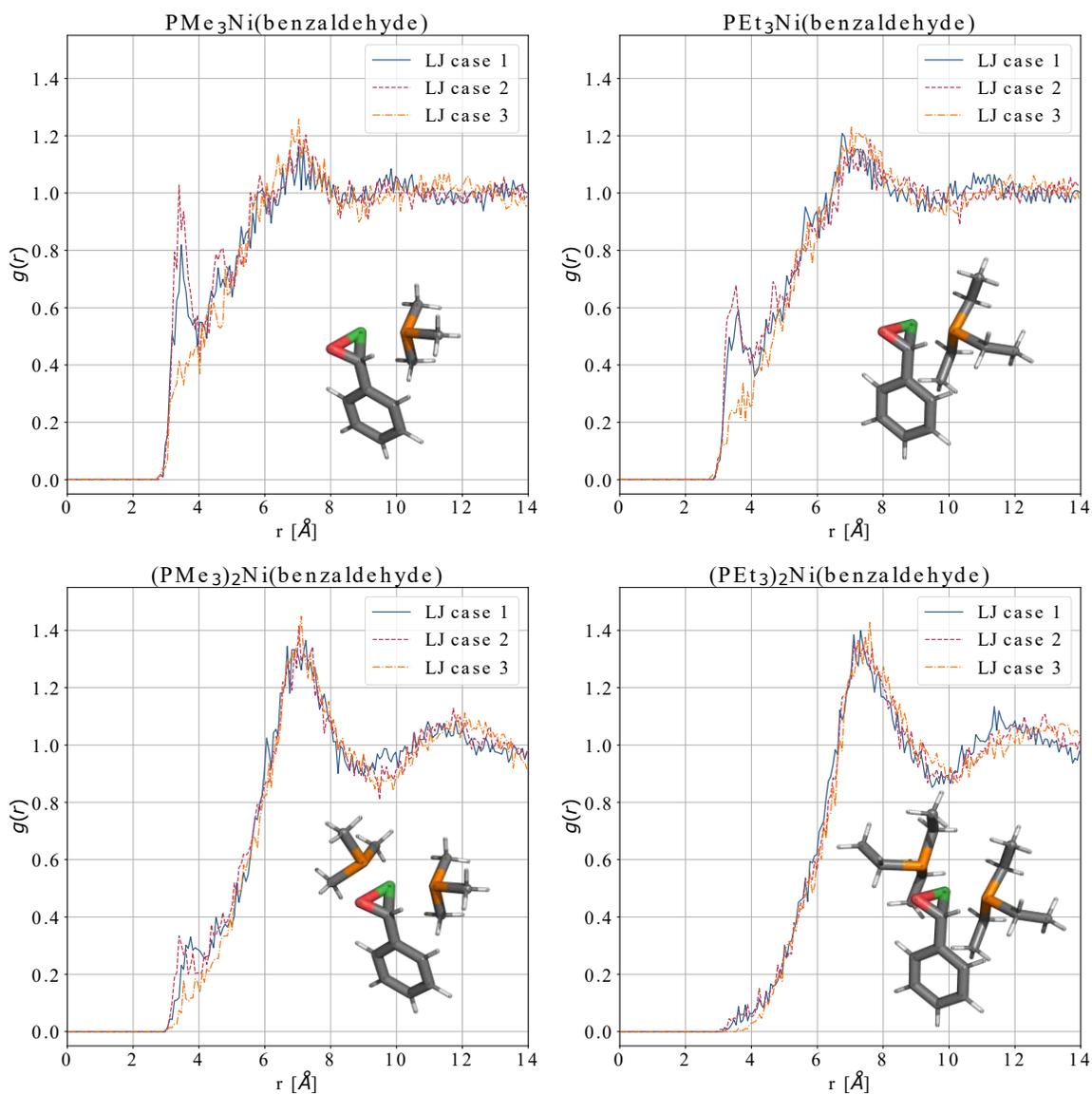
	O	C	P ( $L_1$ complex)	P ( $L_2$ complex)
min	1.67	1.75	2.05	2.01
Q1	1.82	1.93	2.21	2.17
Q2	1.87	1.99	2.25	2.21
Q3	1.92	2.06	2.30	2.26
max	2.35	3.13	2.54	2.52

**Table S2: List of abbreviations and long-form names of ligands.** All ligands with abbreviation (left) and long-form name (right). Ligands in the upper part were used for experiments with both benzaldehyde and 4-fluorobenzaldehyde, the lower part only for the latter.

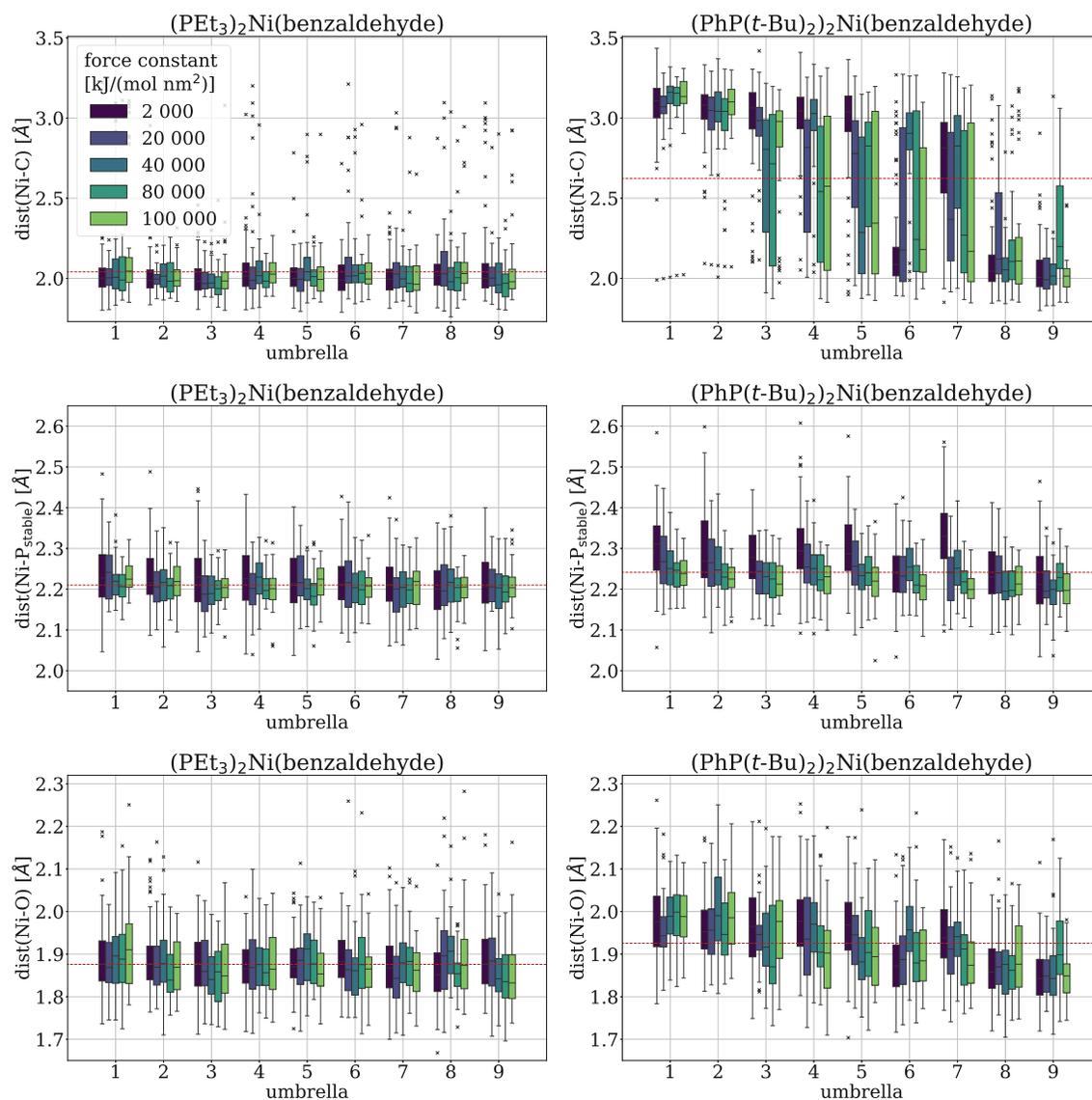
ligand	ligand (chemical name)
PMe <sub>3</sub>	Trimethylphosphine
Me <sub>2</sub> PPh	Dimethylphenylphosphine
PEt <sub>3</sub>	Triethylphosphine
MePPh <sub>2</sub>	Methyldiphenylphosphine
MeP( <i>t</i> -Bu) <sub>2</sub>	Di- <i>tert</i> -butylmethylphosphine
PPh <sub>3</sub>	Triphenylphosphine
P(4-F-Ph) <sub>3</sub>	Tris(4-fluorophenyl)phosphine
PhP( <i>t</i> -Bu) <sub>2</sub>	Di- <i>tert</i> -butylphenylphosphine
CyPPh <sub>2</sub>	Cyclohexyldiphenylphosphine
P( <i>i</i> -Bu) <sub>3</sub>	Triisobutylphosphine
P(4-CF <sub>3</sub> Ph) <sub>3</sub>	Tris(4-trifluoromethylphenyl)phosphine
PBn <sub>3</sub>	Tribenzylphosphine
CyP( <i>t</i> -Bu) <sub>2</sub>	Cyclohexyldi- <i>tert</i> -butylphosphine
P(4-OMe-Ph) <sub>3</sub>	Tris(4-methoxyphenyl)phosphine
Cy <sub>2</sub> PPh	Dicyclohexylphenylphosphine
( <i>t</i> -Bu)PCy <sub>2</sub>	<i>tert</i> -Butyldicyclohexylphosphine
PCy <sub>3</sub>	Tricyclohexylphosphine
P(4-NMe <sub>2</sub> -Ph) <sub>3</sub>	Tris(4-dimethylaminophenyl)phosphine
CataCXiumA	Di-(1-adamantyl)- <i>n</i> -butylphosphine
CyTyrannoPhos	Cyclohexylbis(3,5-di- <i>tert</i> -butylphenyl)phosphine
DrewPhos	Tris(3,5-di- <i>tert</i> -butylphenyl)phosphine
TriceraPhos	Dicyclopentyl(3,5-di-(2,4,6-triisopropylphenyl)phenyl)phosphine
PteroPhos	Cyclopentyl(3,5-di-(2,4,6-triisopropylphenyl)phenyl)phosphine
MeJohnPhos	(2-Biphenyl)di-methylphosphine
( <i>t</i> -Bu)PPh <sub>2</sub>	<i>tert</i> -Butyldiphenylphosphine
P( <i>t</i> -Bu) <sub>3</sub>	Tri( <i>tert</i> -butyl)phosphine
JohnPhos	(2-Biphenyl)di- <i>tert</i> -butylphosphine
CyJohnPhos	(2-Biphenyl)di-cyclohexylphosphine
SPhos	2-Dicyclohexylphosphino-2',6'-dimethoxybiphenyl
BrettPhos	Dicyclohexyl(2',4',6'-triisopropyl-3,6-dimethoxy-[1,1'-biphenyl]-2-yl)phosphine
CyTriceraPhos	Dicyclohexyl(3,5-di-(2,4,6-triisopropylphenyl)phenyl)phosphine

**Table S3: Additional information about ligands and their complexes (Kraken ID, ligation state and complex size).** All ligands used for  $L_2Ni(\text{benzaldehyde})$  complexes (upper part) and for  $L_2Ni(4\text{-fluorobenzaldehyde})$  complexes (lower part), including the Kraken ID, the ligation state result (NMR experiment by Newman-Stonebraker et al.<sup>1</sup>), as well as the total size of the complex.

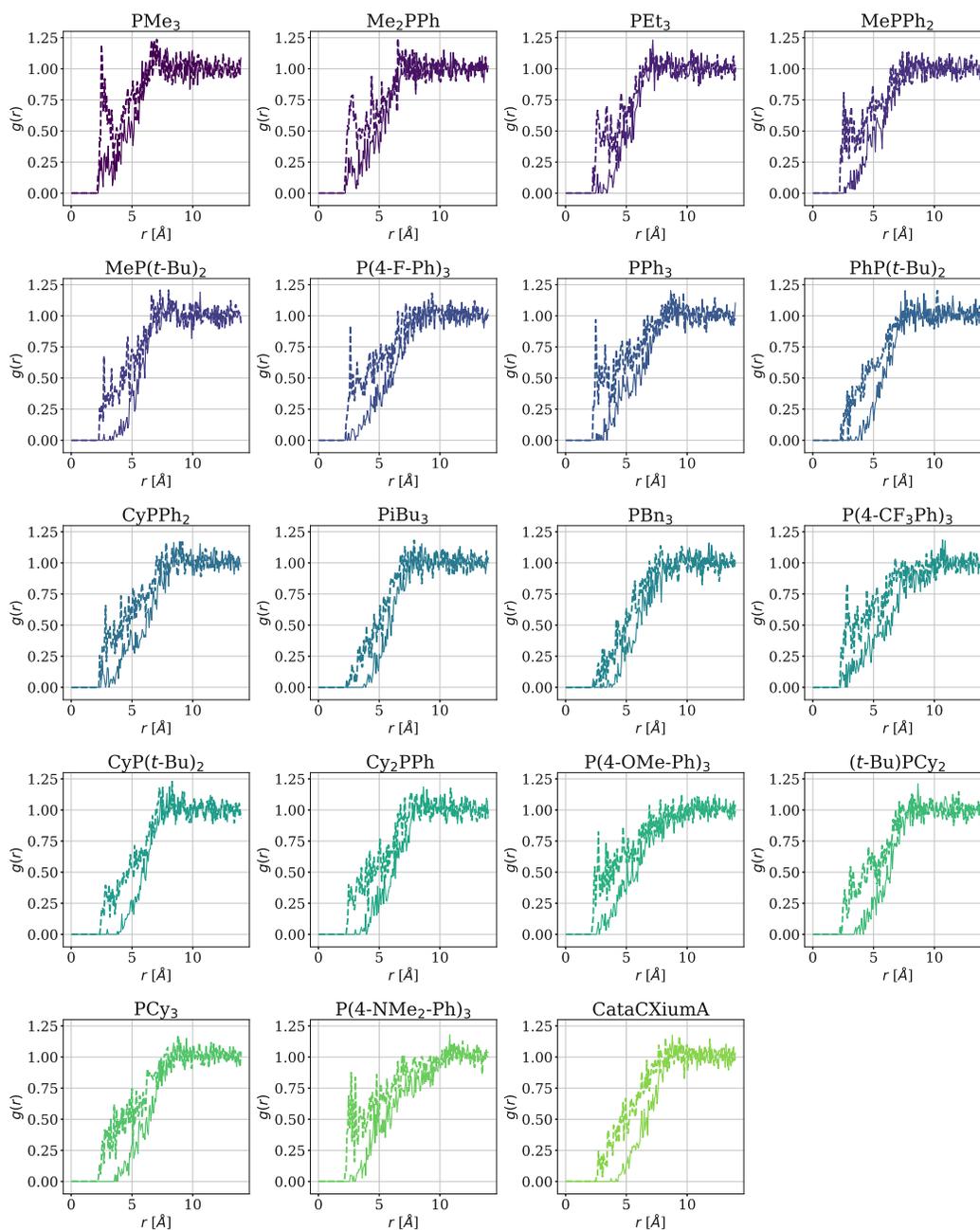
ligand	Kraken ID	ligation state	complex size [atoms]
PMe <sub>3</sub>	22	NaN	41
Me <sub>2</sub> PPh	80	NaN	55
PEt <sub>3</sub>	21	2	59
MePPh <sub>2</sub>	81	NaN	69
MeP( <i>t</i> -Bu) <sub>2</sub>	14	2	77
PPh <sub>3</sub>	17	2	83
P(4-F-Ph) <sub>3</sub>	61	NaN	83
PhP( <i>t</i> -Bu) <sub>2</sub>	31	1	91
CyPPh <sub>2</sub>	162	2	95
P( <i>i</i> -Bu) <sub>3</sub>	252	2	95
P(4-CF <sub>3</sub> Ph) <sub>3</sub>	16	2	101
PBn <sub>3</sub>	65	2	101
CyP( <i>t</i> -Bu) <sub>2</sub>	30	1	103
P(4-OMe-Ph) <sub>3</sub>	62	NaN	107
Cy <sub>2</sub> PPh	68	NaN	107
( <i>t</i> -Bu)PCy <sub>2</sub>	32	2	111
PCy <sub>3</sub>	11	2	119
P(4-NMe <sub>2</sub> -Ph) <sub>3</sub>	455	NaN	131
CataCXiumA	10	1	143
CyTyrannoPhos	158	2	191
DrewPhos	566	2	227
TriceraPhos	159	2	243
PteroPhos	183	2	385
MeJohnPhos	862	2	75
( <i>t</i> -Bu)PPh <sub>2</sub>	13	2	87
P( <i>t</i> -Bu) <sub>3</sub>	8	1	95
JohnPhos	42	1	111
CyJohnPhos	5	1	127
SPhos	3	1	143
BrettPhos	102	1	197
CyTriceraPhos	171	2	255



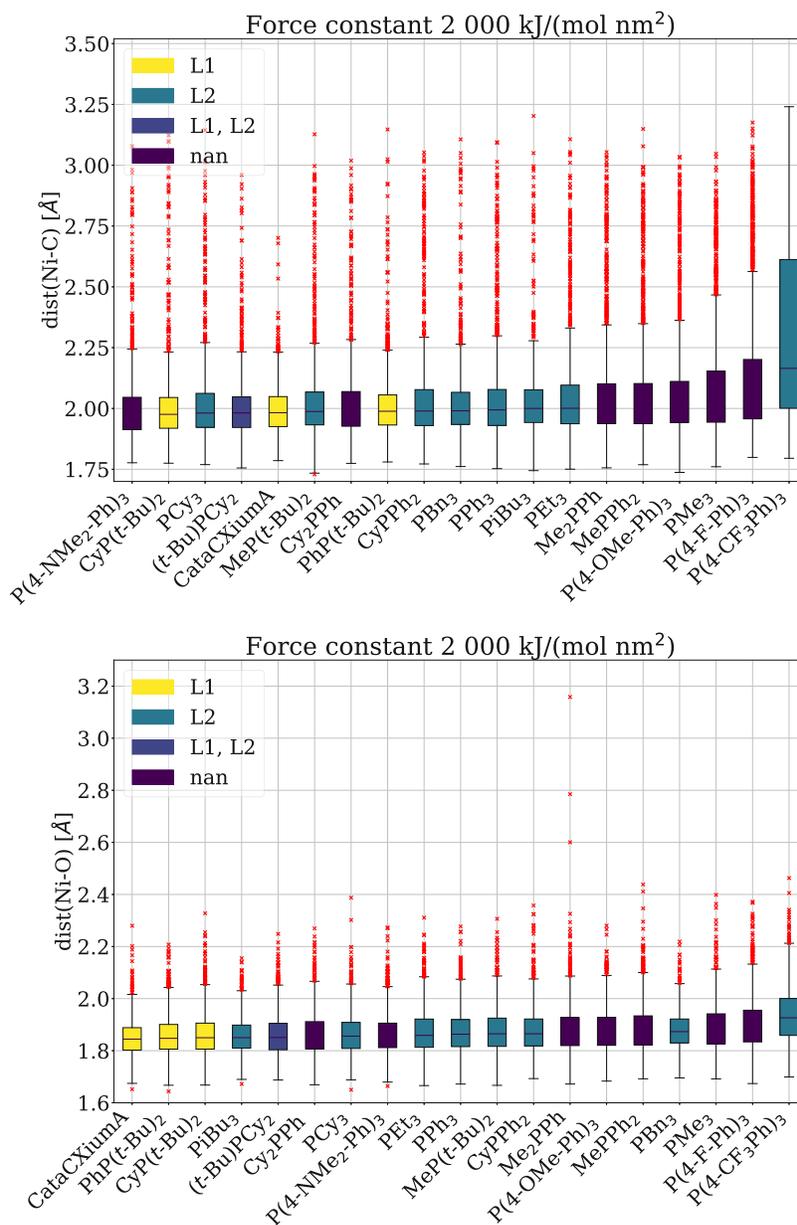
**Figure S1: Effect of different LJ parameters investigated using RDF (carbon of solvent taken as reference).** RDF  $g(r)$  for monoligated,  $x=1$ , (upper plots) and bisligated,  $x=2$ , (lower plots)  $L_x$ Ni(benzaldehyde) complexes with ligands  $L$ =PMe<sub>3</sub> (left plots) and  $L$ =PEt<sub>3</sub> (right plots). Complex structure shown in lower right corner; LJ, Lennard Jones; MD, Molecular Dynamics; RDF, radial distribution function.



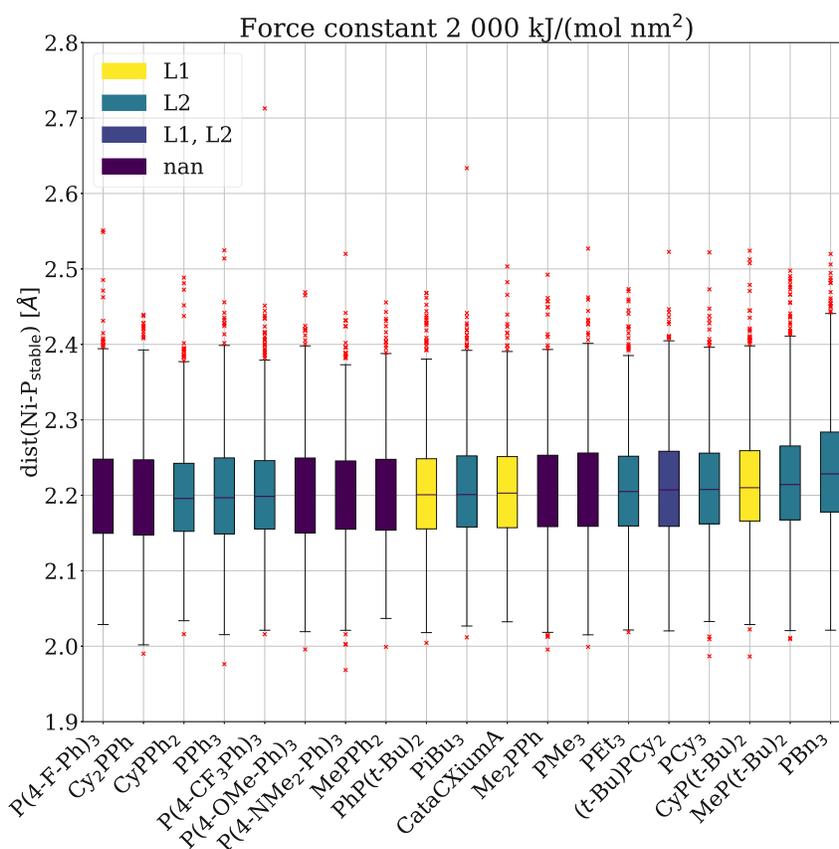
**Figure S2: Effect of different force constants on the distances between nickel and the neighboring atoms.** Boxplot for distances from restrained MD run between nickel atom and phosphorus of ligand which is kept close to nickel, as well as carbon and oxygen of benzaldehyde for the first 9 umbrellas. Shown for ligands  $\text{PEt}_3$  and  $\text{PhP}(t\text{-Bu})_2$  in complex. Red-dashed line indicates mean of distances.



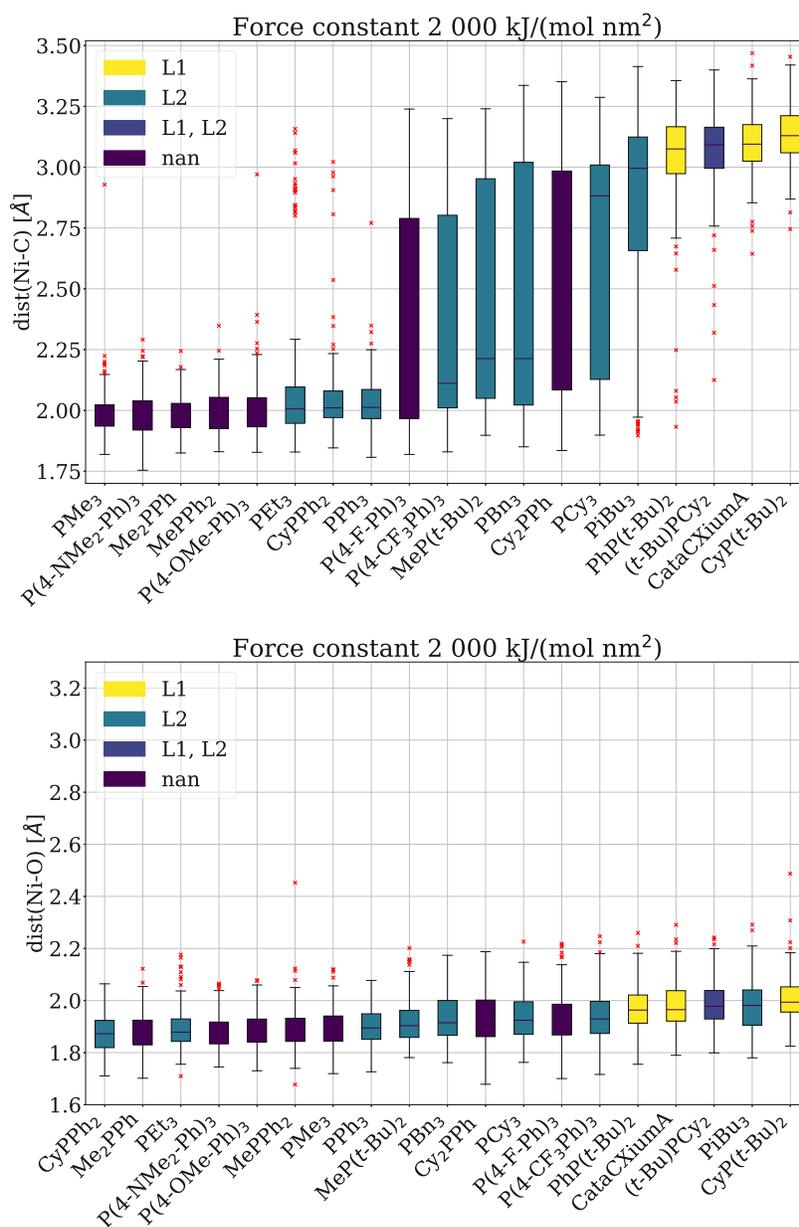
**Figure S3: RDF for all monoligated and bisligated complexes in the training data.** The RDF  $g(r)$  for all ligands L (title of subplot) used in the training data is shown for their monoligated,  $x = 1$ , (dashed curve) and bisligated,  $x = 2$ , (solid curve)  $L_xNi(\text{benzaldehyde})$  complex.  $r$  is the distance between the nickel atom and hydrogen of the benzene solvent. Force constant 100 000 kJ/(mol nm<sup>2</sup>) was used; RDF, radial distribution function.



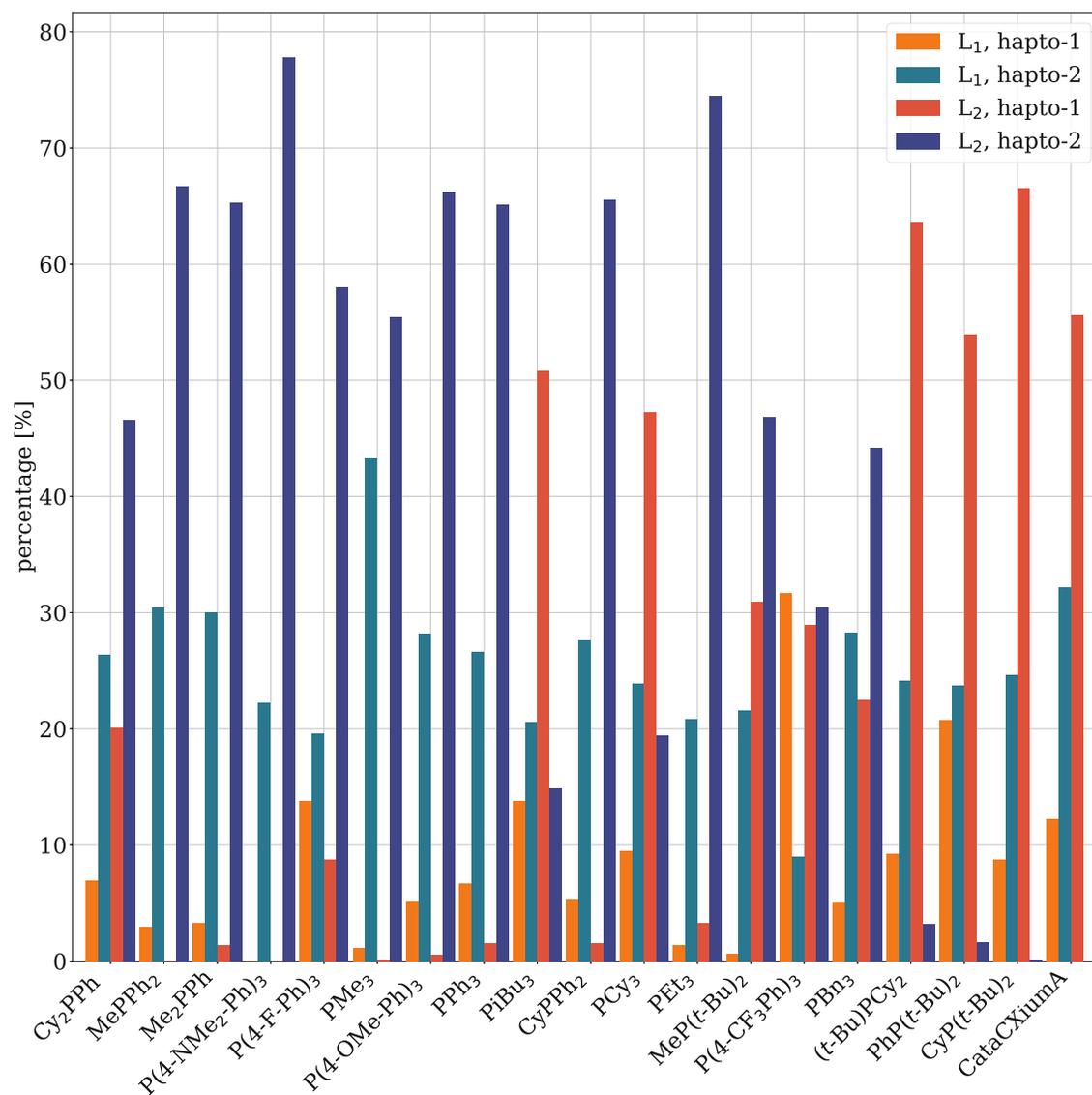
**Figure S4: Distribution of distances between nickel and benzaldehyde in training data for  $L_1Ni(\text{benzaldehyde})$  complex.** For all ligands displayed on the x-axis, for the lowest force constant  $2000 \text{ kJ}/(\text{mol nm}^2)$ , the distances for nickel-carbon (top), and nickel-oxygen (bottom) for the last 17 umbrellas combined are shown. Experimental ligation state results from Newman-Stonebraker et al. are indicated with colored boxes, being the monoligated "L1", the bisligated "L2" complex, or the presence of both indicated by "L1, L2". If no experiment was conducted for the ligand, this is indicated by "nan".<sup>1</sup>



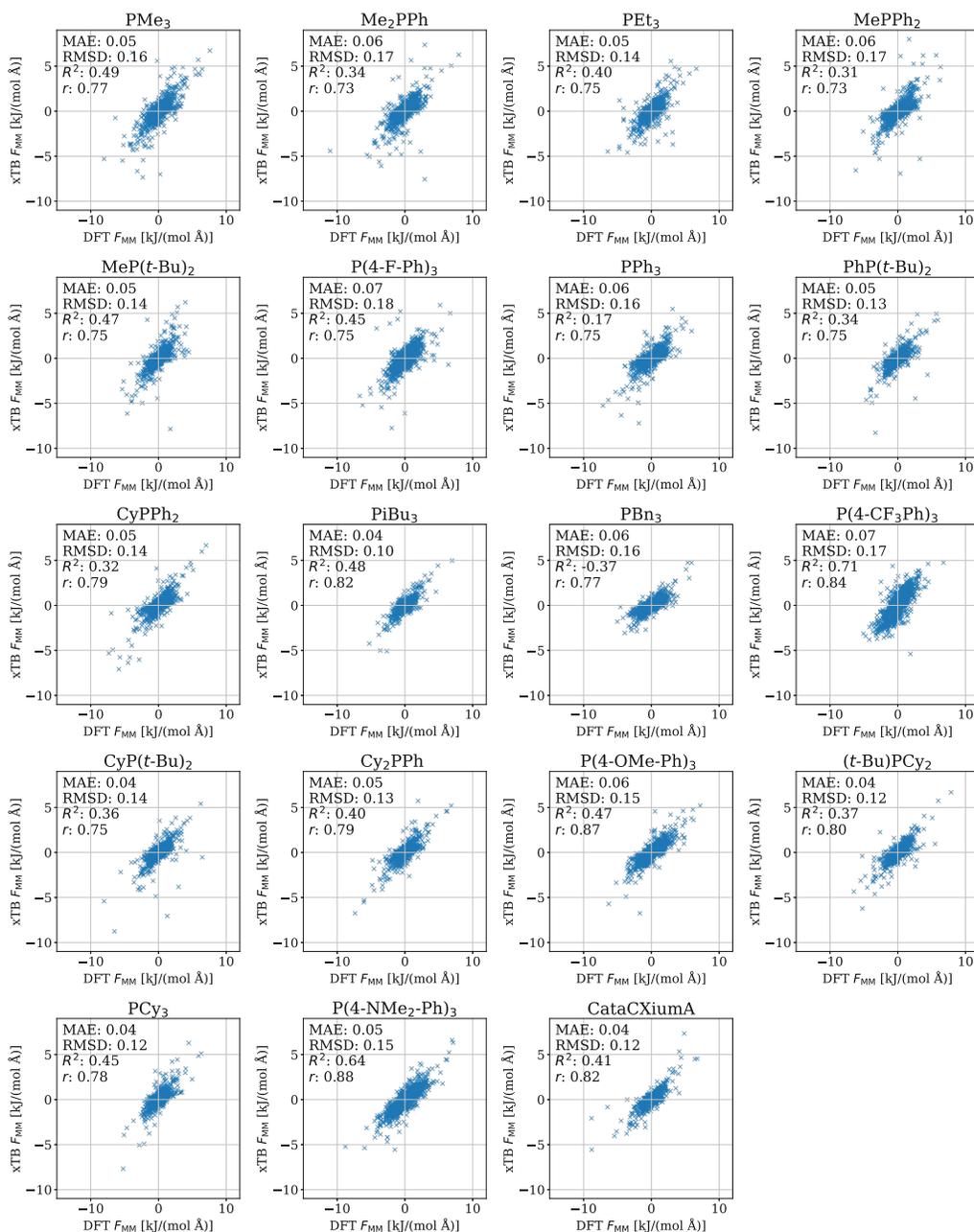
**Figure S5: Distribution of distances between nickel and phosphine ligand in training data for L<sub>1</sub>Ni(benzaldehyde) complex.** For all ligands, for the lowest force constant 2 000 kJ/(mol nm<sup>2</sup>), the distances for nickel-phosphorus for the last 17 umbrellas are shown. Experimental ligation state results from Newman-Stonebraker et al. are indicated with colored boxes, being the monoligated "L1", the bisligated "L2" complex, or the presence of both indicated by "L1, L2". If no experiment was conducted for the ligand, this is indicated by "nan".<sup>1</sup>



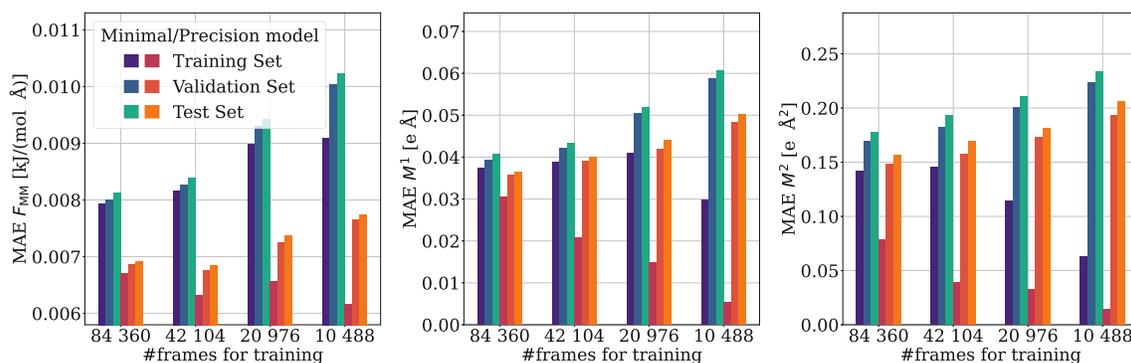
**Figure S6: Distribution of distances between nickel and benzaldehyde in training data for  $L_2Ni(\text{benzaldehyde})$  complex.** For all ligands, for the lowest force constant  $2000 \text{ kJ}/(\text{mol nm}^2)$ , the distances for nickel-carbon (top) and nickel-oxygen (bottom) for the first three umbrellas are shown. Experimental ligation state results from Newman-Stonebraker et al. are indicated with colored boxes, being the monoligated "L1", the bisligated "L2" complex, or the presence of both indicated by "L1, L2". If no experiment was conducted for the ligand, this is indicated by "nan".<sup>1</sup>



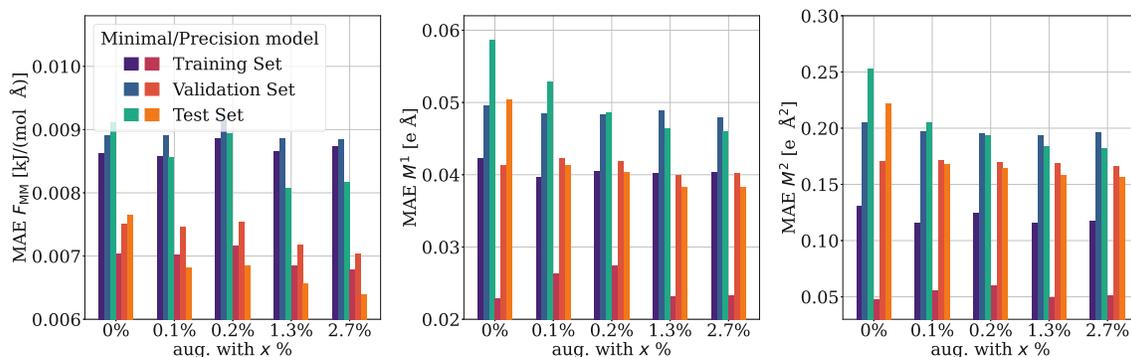
**Figure S7: Configurational composition of training data for the first ten umbrellas for lowest force constant.** All configurations are divided into four different cases, distinguishing between L<sub>1</sub>Ni(benzaldehyde) complexes and L<sub>2</sub>Ni(benzaldehyde) complexes as well as hapto-1 and hapto-2 benzaldehyde ligand. Occurrence is given in percentages adding up to 100% for each ligand. Force constant is 2 000 kJ/(molnm<sup>2</sup>). Red shades indicate hapto-1, blue shades hapto-2 benzaldehyde.



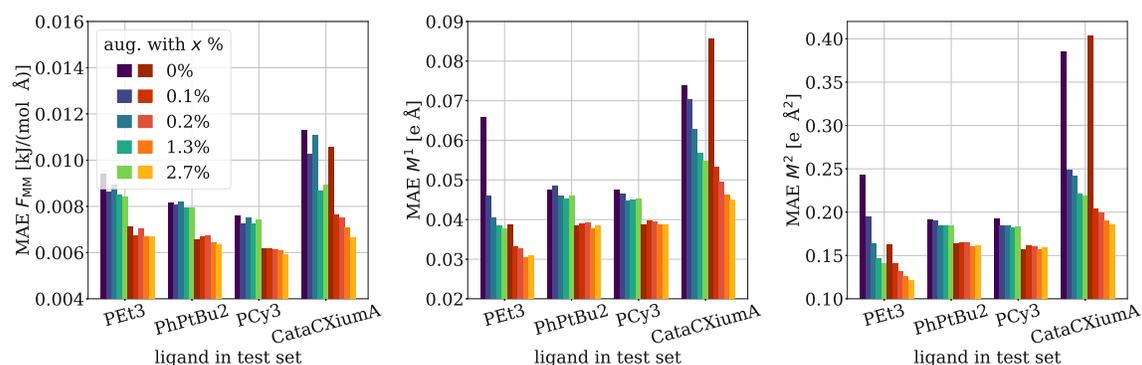
**Figure S8: Forces for MM atoms evaluated by xTB versus DFT in training data set.** For all frames used in the training data (5 550 per complex) the gradients of the energies of the MM atoms  $F_{MM}$  obtained from xTB are compared to the re-evaluated values of DFT for each complex individually.



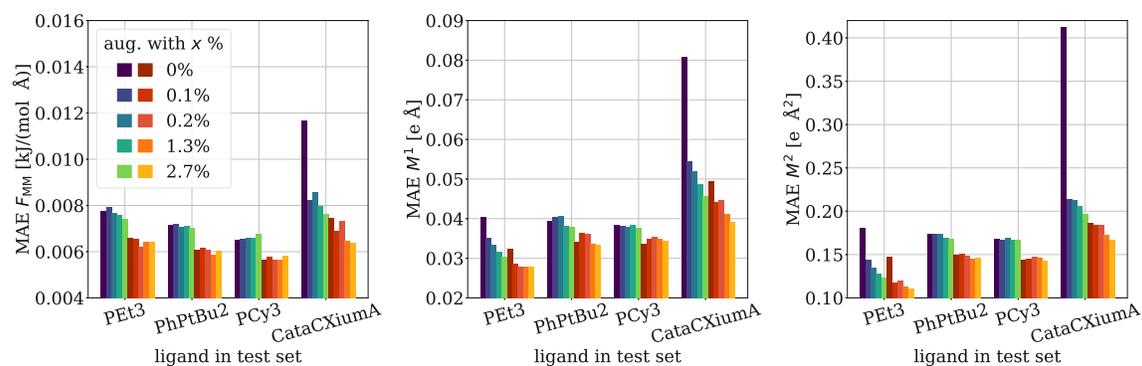
**Figure S9: Effect of different amounts of training data, part 2.** The MAE is shown for the forces on the MM atoms  $F_{MM}$  (left), the dipoles  $M^1$  (middle) and quadrupoles  $M^2$  (right) for the model versions 1 - 4. Minimal models (blue color scheme) and precision models (red color scheme) are shown. The total number of frames in the training data is 84 360, 42 104, 20 976 and 10 488.



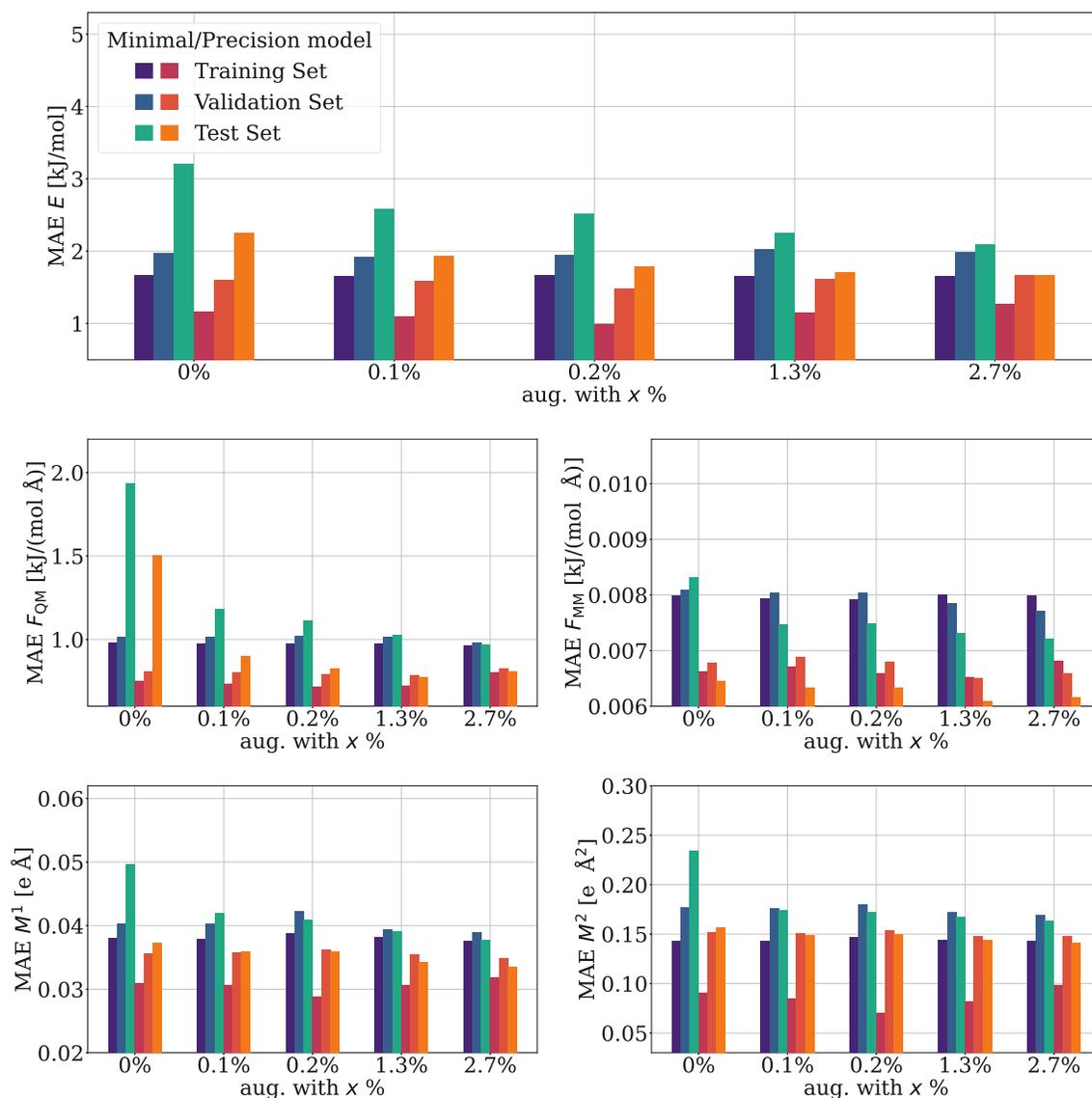
**Figure S10: Model versions 5 - 9: Transferability experiment, part 2.** The MAE is shown for the forces on the MM atoms  $F_{MM}$  (left), the dipoles  $M^1$  (middle) and quadrupoles  $M^2$  (right) for the model versions 5 - 9. MAEs are shown for training, validation and test sets. Minimal models (blue color scheme) and precision models (red color scheme) are shown. Training and validation sets consisted of data set 2, but were gradually augmented with frames from data set 3 (0%, 0.1%, 0.2%, 1.3%, and 2.7%). For testing data set 3 was used.



**Figure S11: Model versions 5 - 9: Transferability experiment analyzed for different ligands, part 2.** The MAE is shown for the forces on the MM atoms  $F_{MM}$  (left), the dipoles  $M^1$  (middle) and quadrupoles  $M^2$  (right) for the model versions 5 - 9. MAEs for the test data set are shown only. MAEs are reported for complexes  $(\text{PET}_3)_x\text{Ni}(\text{benzaldehyde})$ ,  $(\text{PhP}(t\text{-Bu})_2)_x\text{Ni}(\text{benzaldehyde})$ ,  $(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$ , and  $(\text{CataCXiumA})_x\text{Ni}(\text{benzaldehyde})$ . Minimal models (blue color scheme) and precision models (red color scheme) are shown. Training and validation sets consisted of data set 2 but were gradually augmented with frames from data set 3 (0%, 0.1%, 0.2%, 1.3%, and 2.7%). For testing data set 3 was used.



**Figure S12: Model versions 11 - 15: Transferability experiment analyzed for different ligands, part 2.** The MAE is shown for the forces on the MM atoms  $F_{MM}$  (left), the dipoles  $M^1$  (middle) and quadrupoles  $M^2$  (right) for the model versions 11 - 15. MAEs for the test data set are shown only. MAEs are reported for complexes  $(\text{PET}_3)_x\text{Ni}(\text{benzaldehyde})$ ,  $(\text{PhP}(t\text{-Bu})_2)_x\text{Ni}(\text{benzaldehyde})$ ,  $(\text{PCy}_3)_x\text{Ni}(\text{benzaldehyde})$ , and  $(\text{CataCXiumA})_x\text{Ni}(\text{benzaldehyde})$ . Minimal models (blue color scheme) and precision models (red color scheme) are shown. Training and validation sets consisted of data set 2 but were gradually augmented with frames from data set 3 (0%, 0.1%, 0.2%, 1.3%, and 2.7%). For testing data set 3 was used.



**Figure S13: Models 11 - 15: Transferability experiment.** The MAE is shown for energy values  $E$  (top), forces on QM atoms  $F_{QM}$  (middle left), forces on the MM atoms  $F_{MM}$  (middle right), the dipoles  $M^1$  (bottom left) and quadrupoles  $M^2$  (bottom right) for the model versions 11 - 15. MAEs are shown for training, validation, and test sets. Minimal models (blue color scheme) and precision models (red color scheme) are shown. Training and validation sets consisted of data set 2, but were gradually augmented with frames from data set 3 (0%, 0.1%, 0.2%, 1.3%, and 2.7%). For testing data set 3 was used.