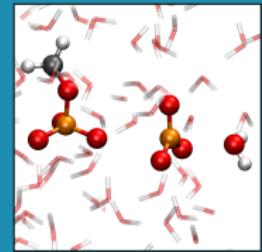


Ab Initio Molecular Dynamics Simulations of Phosphate Hydrolysis Using Neural Network Potentials



Master's thesis defence presentation

Albert Makhmudov

Supervisor: Prof. J. Harvey

June 30, 2025

Why is phosphate hydrolysis challenging to study?

Complex Reaction Mechanism

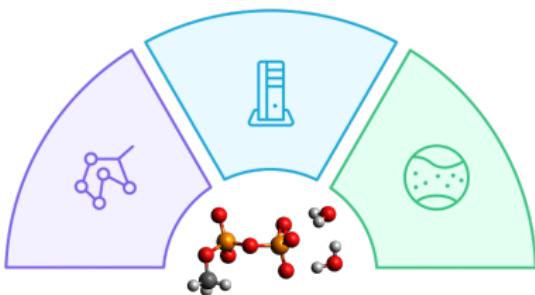
Involves multiple pathways and difficult-to-characterise transition states.

High Computational Cost

AIMD is expensive since it requires extensive sampling and large system sizes due to water's role.

Strong Solvent Effects

Complicated by hydrogen bonding and proton transfer in aqueous environments.



The ongoing debate about the reaction mechanism served as a motivation to revisit phosphate hydrolysis in this project.

Research goals

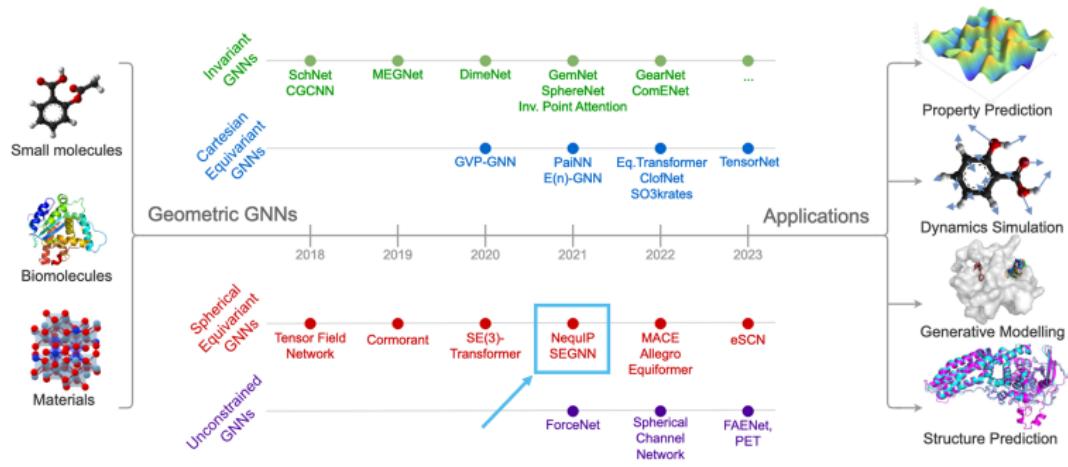
- ▶ Compose a comprehensive dataset covering all reaction steps.
- ▶ Train the NequIP neural network to fit a neural network potential (NNP).
- ▶ Assess the accuracy and performance of the NNP.
- ▶ Perform well-tempered metadynamics simulations to obtain the free energy surface.
- ▶ Gain insights into the kinetics and thermodynamics of the reaction.
- ▶ Gain insights into the proton transfer mechanism.



*It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems **without too much computation**.*

– Paul Dirac

Geometric graph neural networks



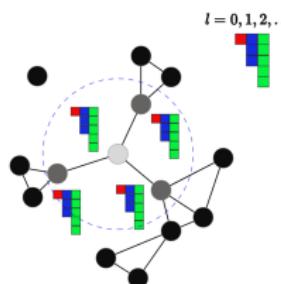
NequIP was named one of the most accurate and stable neural network potentials to run the extensive molecular dynamics simulations. NequIP is based on the equivariant graph neural network architecture.

Neural equivariant interatomic potentials (NequIP)

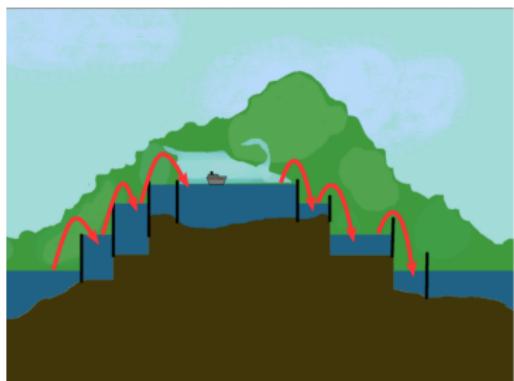
The neural network is trained using a loss function:

$$\mathcal{L} = \lambda_E \|\hat{E} - E\|^2 + \lambda_F \frac{1}{3N} \sum_{i=1}^N \sum_{\alpha=1}^3 \left\| -\frac{\partial \hat{E}}{\partial r_{i,\alpha}} - F_{i,\alpha} \right\|^2$$

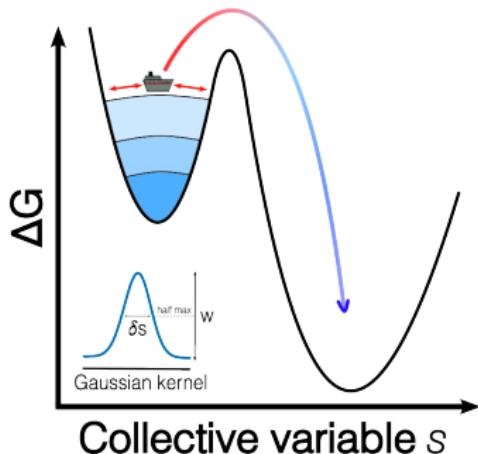
It is based on a weighted sum of energy and force loss terms. Where \hat{E} is the predicted energy, λ_E and λ_F are the energy and force weights, respectively, and N is the number of atoms. MSE loss is used.



Metadynamics



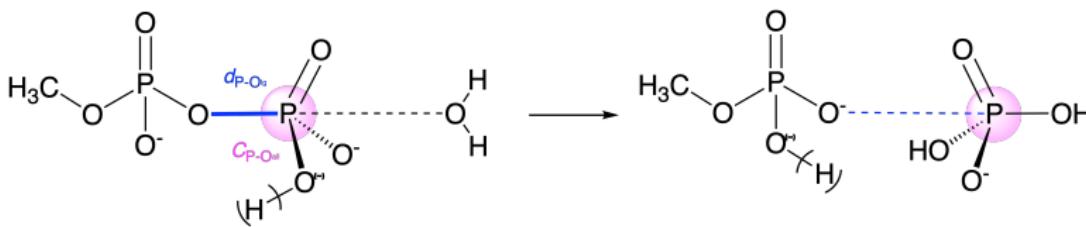
The Panama Canal



The beauty of metadynamics is that, after sufficiently long sampling, the biasing potential converges to the negative of the underlying free energy surface:

$$\lim_{t \rightarrow \infty} V(s, t) \sim -F(s)$$

System



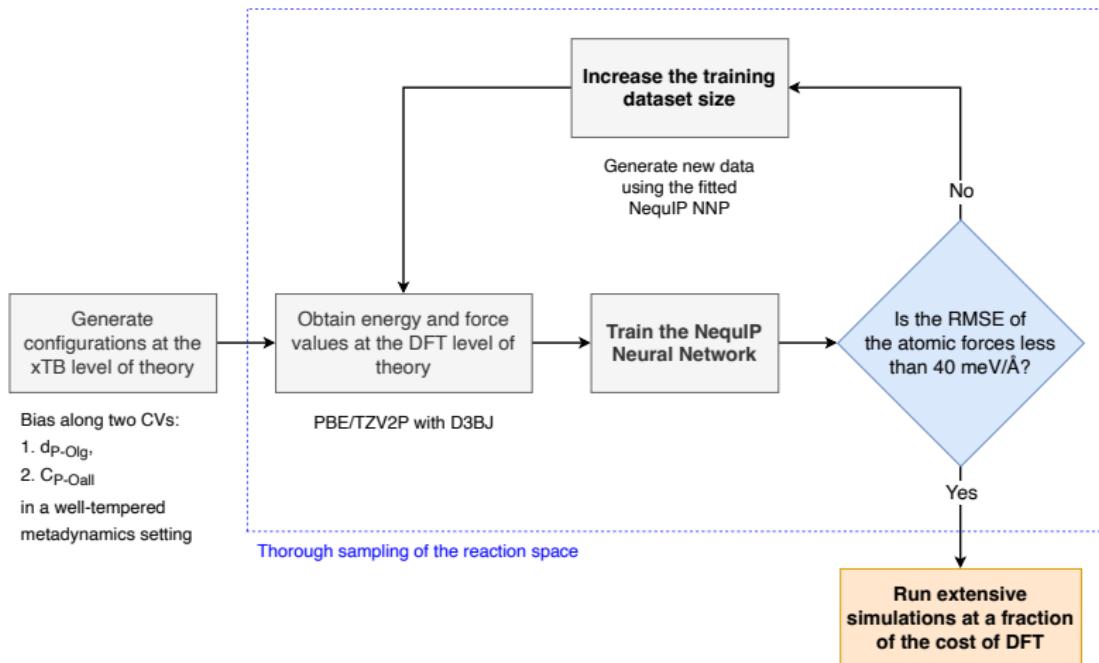
$$d_{P-O_{lg}} = d(P - O_{lg}), C_{P-O_{all}} = CN(P - O_{all})$$

The systems studied in this work are the following:

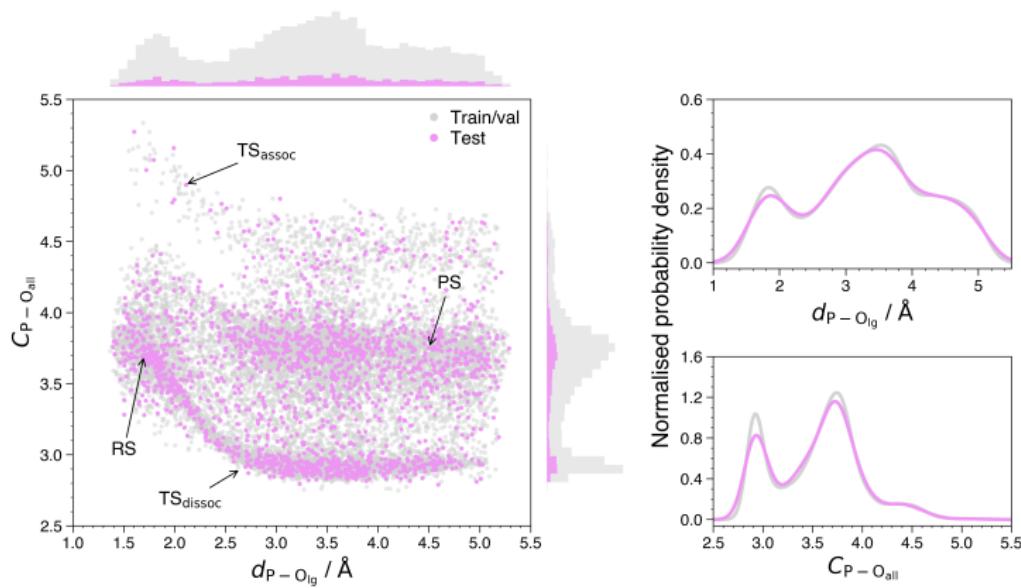
- ▶ MeDHP³⁻ with 3 Na⁺ counterions solvated by 119 H₂O.
- ▶ MeHDP²⁻ with 2 Na⁺ counterions solvated by 124 H₂O.

The box dimensions are 15.877 × 15.877 × 15.877 and 15.901 × 15.901 × 15.901 Å³, respectively.

Neural network training workflow

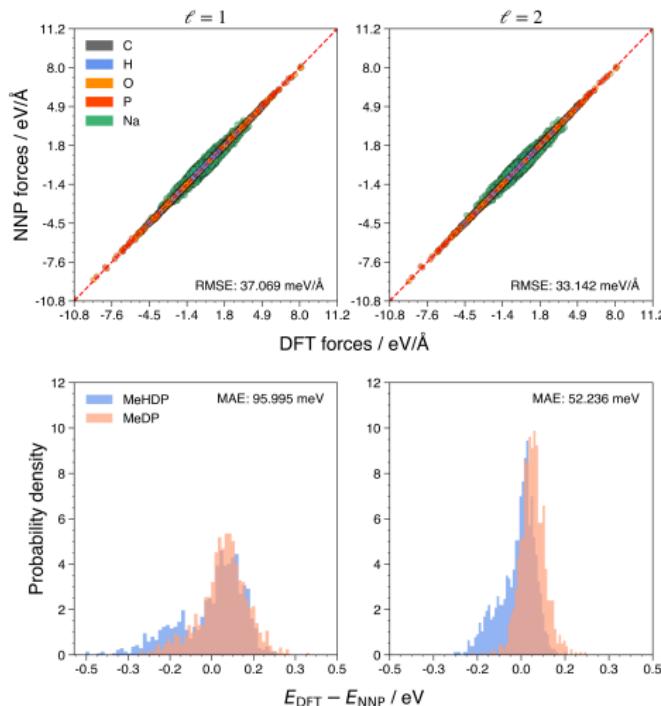


Final training and test datasets' composition



The final training and test datasets consist of 12,000 and 1,800 frames, respectively, and covers all reaction steps.

Accuracy of the fitted potential

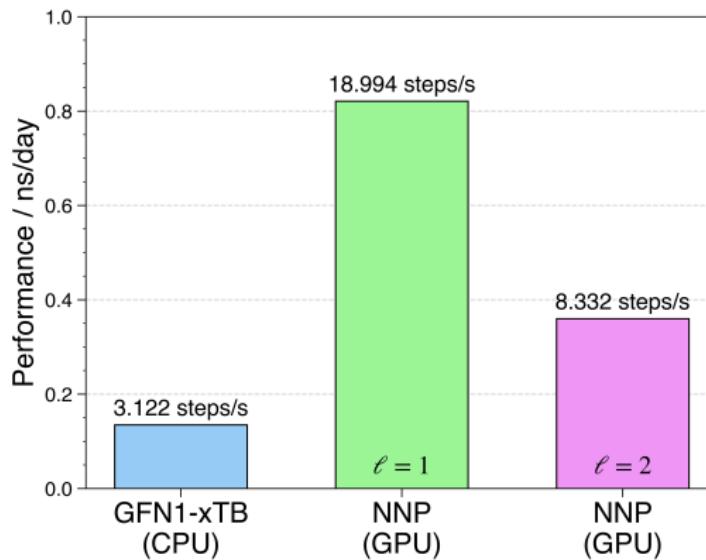


Community standards are as follows:

- ▶ 'very good fit'
 $\text{MAE}_E = 1\text{-}10 \text{ meV/atom}$
 $\text{RMSE}_F = 20\text{-}40 \text{ meV/}\text{\AA}$
- ▶ 'perfect fit'
 $\text{MAE}_E \sim 1 \text{ meV/atom}$
 $\text{RMSE}_F \sim 10 \text{ meV/}\text{\AA}$

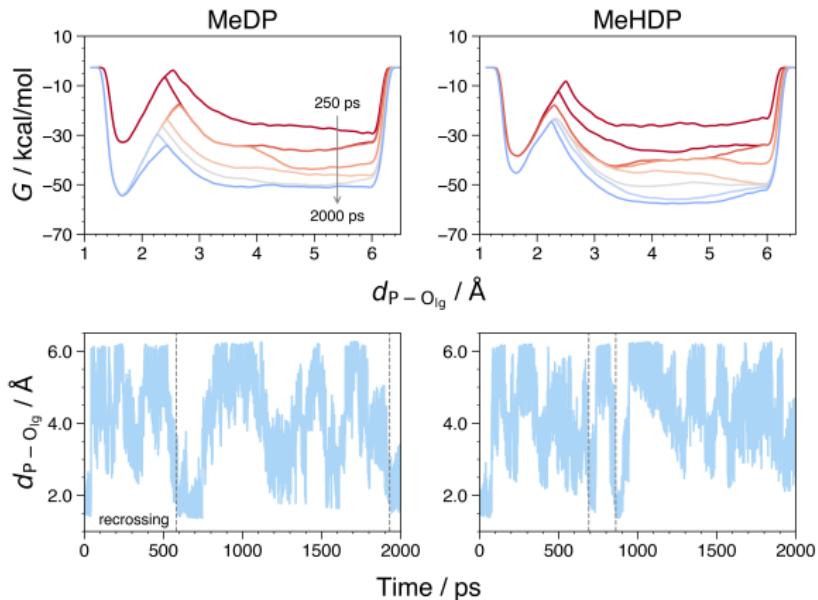
We fitted 2 potentials and both of them are very accurate exhibiting fairly small errors.

Performance of the potential



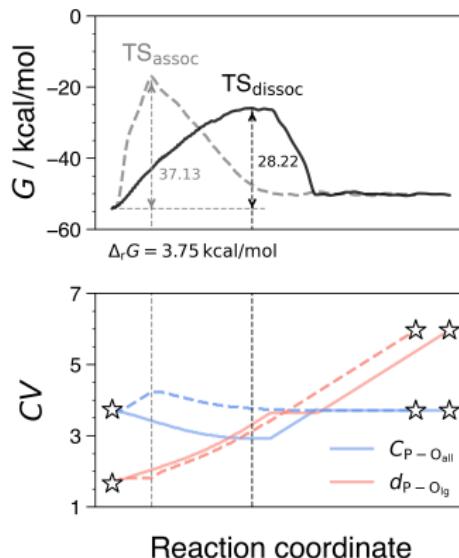
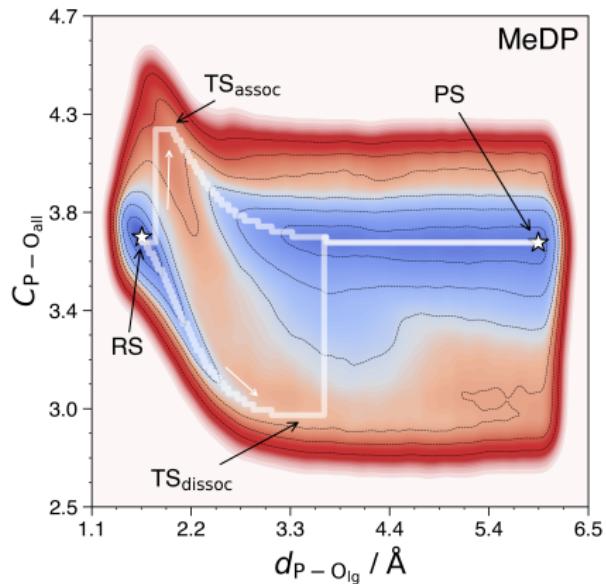
The $\ell = 1$ potential is about 2.5 times faster than the $\ell = 2$ potential, while the accuracy is comparable.

Convergence of the free energy profiles

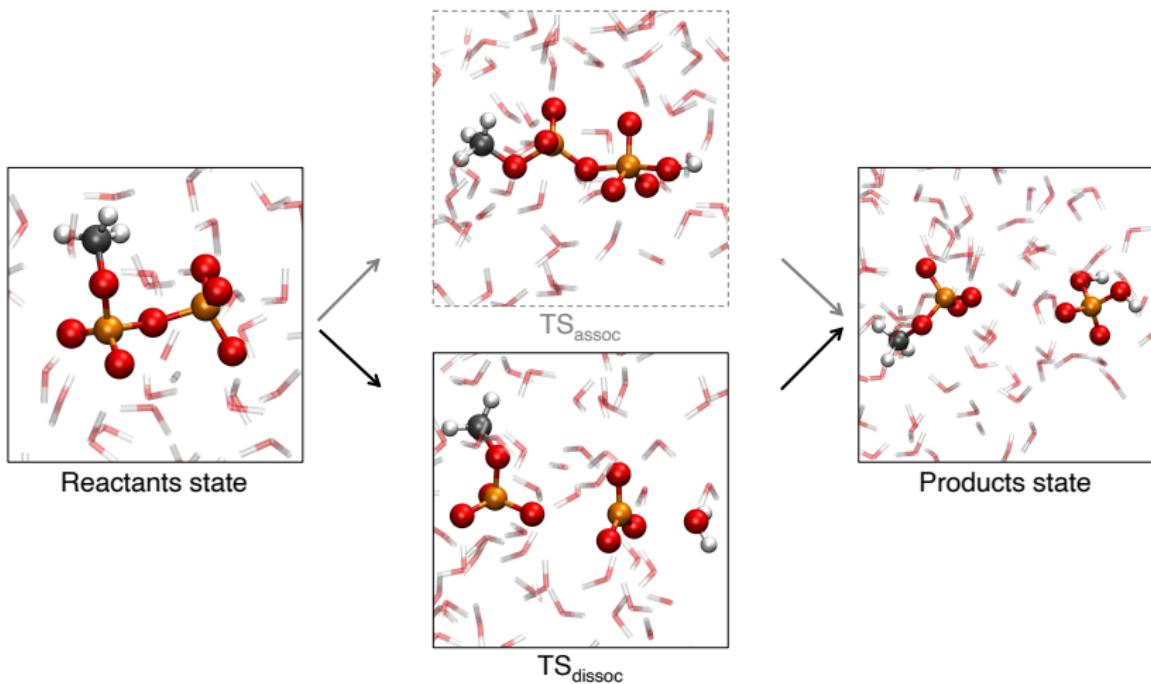


- ▶ The profiles are not fully converged.
- ▶ Hence, the results should be considered provisional.
- ▶ There are recrossing events between the reactants and products states.
- ▶ Which gives confidence in the soon approaching convergence.

Reaction mechanism for MeDP³⁻ at 300 K

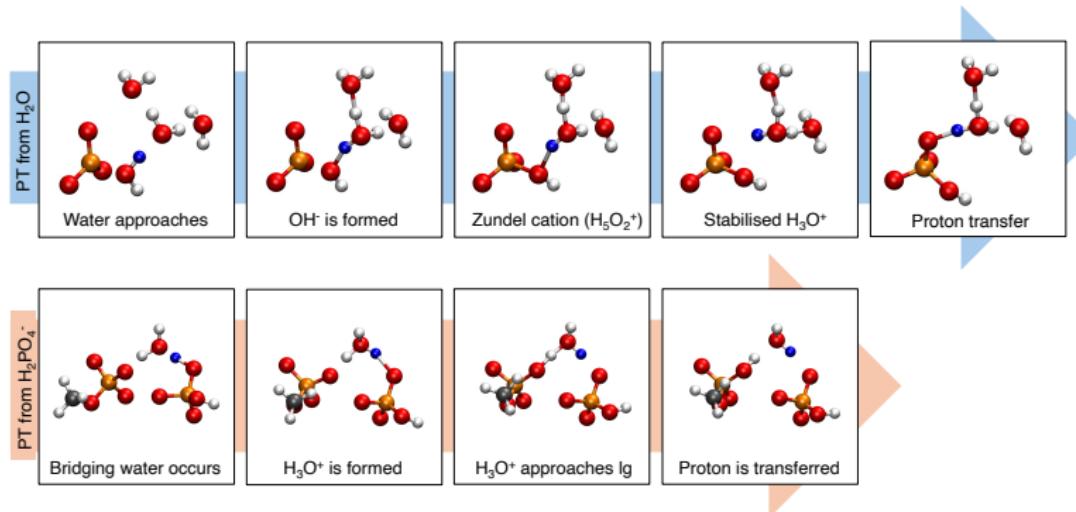


$\Delta G_{\text{exp}}^{\ddagger}$ for $\text{HP}_2\text{O}_7^{3-}$ hydrolysis at 25°C is 29.2 kcal/mol. $\Delta G_{\text{calc}}^{\ddagger}$ of 28.22 kcal/mol for the D_NA_N mechanism is within the chemical accuracy!



In the associative TS phosphorus is pentacoordinated. In the dissociative TS, there is clearly no bond between the phosphorus and the leaving group.

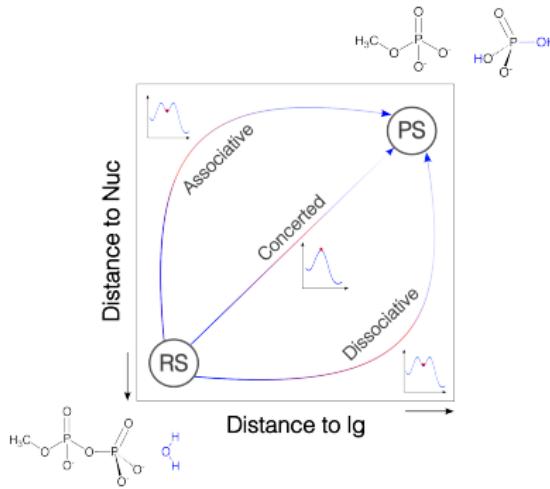
Proton transfer mechanism



From the limited number of inspected proton transfer (PT) events, it was found that the PT mechanism involves 1 or 3 water molecules and happens from either H_2O or H_2PO_4^- .

What we achieved so far?

- ▶ The NNP produced expected for the PBE-D3(BJ)/TZV2P results.
- ▶ For the first time, the 2 ns long sampling was performed.
- ▶ MeDP³⁻ hydrolysis goes through the D_NA_N mechanism.
- ▶ The agreement with the experiment is excellent.
- ▶ The proton transfer mechanism involving 1 and 3 water molecules was shown.

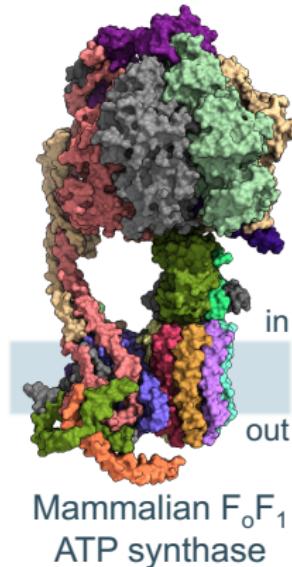


More O'Ferrall-Jencks diagram for MeDP³⁻ hydrolysis.

What can be done further?

- ▶ Extended sampling to reach the convergence.
- ▶ Transition state validation.
- ▶ Effect of enthalpy and entropy.
- ▶ Extension to more complex systems.

As neural network potentials continue to improve, they are likely to become an indispensable tool in the computational chemist's toolkit, enabling the exploration of chemical space with unprecedented accuracy and efficiency.



Thank you for your attention!

Acknowledgments:

- ▶ Jeremy Harvey and all members of the research group, as well as the QCPC division.
- ▶ Ehsan Moravveji (VSC), Hans Vansweevelt (KUL), and Anders Johansson (Harvard) for technical assistance.
- ▶ TCCM cohort for the coffee breaks, lunches, and a lot of fun times (in Aspet and beyond) we shared in 2 years.
- ▶ EMJMD TCCM for the financial support by means of Erasmus+ scholarship.



Group picture, Spring 2025



TCCM cohort 2023-2025



by Mauro Gascón Navas