KU LEUVEN

FACULTY OF SCIENCE

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

Albert MAKHMUDOV

Supervisor: Prof. J. Harvey
KU Leuven

Thesis presented in
fulfillment of the requirements
for the degree of Master of Science
in Theoretical Chemistry and Computational Modelling

Academic year 2024-2025

Foreword

Contribution statement

Summary

List of abbreviations

Contents

Chapter 1

Introduction

1.1 Role of phosphates in biological systems

1.2 Enzymes involved in phosphate hydrolysis

1.3 Reaction mechanism

1.4 Research goals

Chapter 2

Theory

2.1 A brief introduction to statistical mechanics

2.1.1 Classical forcefields and molecular dynamics

2.1.2 The canonical ensemble and free energy calculations

2.1.3 Enhanced sampling techniques

Metadynamics and its well-tempered flavour

Kinetics from metadynamics

2.2 Density functional theory

2.2.1 The Kohn-Sham approach

2.2.2 Generalised gradient approximation and PBE functional

2.2.3 *Ab initio* molecular dynamics and GPW method

2.3 Extended tight binding

2.4 Neural network potentials

2.4.1 Deep neural networks

Multilayer perceptron

Graph neural networks

Message passing neural networks

2.4.2 Invariance and equivariance

Chapter 3

Computational Details

This chapter provides the details of the computational methods used in this work. The first section describes the generation of the training dataset, including the preparation of the system, initial equilibration using the molecular mechanics, exploration of the configuration space at the xTB level, further data labeling, and iterative training of the neural network potential. The second section discusses the production runs at different temperatures using the fitted neural network potential. The third section describes the workflow of validating the transition states obtained from the simulations following the partial Hessian formalism. Finally, the fourth section presents the data analysis and visualisation techniques employed to interpret the results.

3.1 Training dataset generation

3.1.1 System preparation

The systems were prepared using the `CHARMM-GUI` webserver's functionality [1]. In particular, the `Multicomponent Assembler` interface [2] was utilised.

As a first step, the singly protonated and deprotonated forms of the methyl diphosphate were parametrised in `CGenFF` [3], i.e. CHARMM General Forcefield. These states of the methyl diphosphate were chosen based on the fact that pyrophosphoric (diphosphoric) acid has the following dissociation constants [4]:

$$\begin{aligned} H_4P_2O_7 &\rightleftharpoons [H_3P_2O_7]^- + H^+, \quad pK_a = 0.91 \\ [H_3P_2O_7]^- &\rightleftharpoons [H_2P_2O_7]^{2-} + H^+, \quad pK_a = 2.10 \\ [H_2P_2O_7]^{2-} &\rightleftharpoons [HP_2O_7]^{3-} + H^+, \quad pK_a = 6.70 \\ [HP_2O_7]^{3-} &\rightleftharpoons [P_2O_7]^{4-} + H^+, \quad pK_a = 9.32 \end{aligned}$$

Thus, at the physiological pH of 7.4 this acid exists as an equillibrium between the doubly and singly protonated forms. As an assumption, the methyl group can be considered as a proton, therefore we condsidered the methyl diphosphate molecule to exist as a mixture of the singly (MeHDP) and deprotonated (MeDP) forms at the physiological pH.

After succesfully parametrising the molecules, the system was solvated in a cubic box of water molecules together with the sodium counterions Na^+ to neutralise the charge. The final system composition can be seen in Table 3.1.

3.1.2 Initial equillibration using the classical forcefields

Table 3.1: System composition and simulation box details. [1]The final dimensions were obtained after the NPT run using the CHARMM36m forcefield.

System	Equillibrated box dimensions[1] (Å)	No. of water molecules	No. of Na^+
MeDP	$15.877 \times 15.877 \times 15.877$	119	3
MeHDP	$15.901 \times 15.901 \times 15.901$	124	2

3.1.3 xTB based exploration of the configuration space

3.1.4 Data labeling

3.1.5 Iterative training of the neural network potential

First round

Second round

Third round

3.2 Production runs at different temperatures

3.3 Validation of the transition states

3.4 Data analysis and visualisation

Chapter 4

Results and Discussion

Chapter 5

Conclusions

Bibliography

[1] Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859–1865 (2008). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20945. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20945.

[2] Kern, N. R., Lee, J., Choi, Y. K. & Im, W. CHARMM-GUI Multicomponent Assembler for modeling and simulation of complex multicomponent systems. *Nature Communications* **15**, 5459 (2024). URL https://www.nature.com/articles/s41467-024-49700-4. Publisher: Nature Publishing Group.

[3] Kim, S. *et al.* CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules: CHARMM-GUI Ligand Reader and Modeler for CHARMM Force Field Generation of Small Molecules. *Journal of Computational Chemistry* **38**, 1879–1886 (2017). URL https://onlinelibrary.wiley.com/doi/10.1002/jcc.24829.

[4] Haynes, W. M. *CRC Handbook of Chemistry and Physics* (CRC Press, 2016). Google-Books-ID: VVezDAAAQBAJ.

Appendix A

Supplementary information

Quantum Chemistry and Physical Chemistry
Celestijnenlaan 200F bus 2404
3001 LEUVEN, BELGIË
tel. + 32 16 37 21 98
jeremy.harvey@kuleuven.be
www.kuleuven.be