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

© Copyright by KU Leuven

Without written permission of the promotors and the authors it is forbidden to reproduce or adapt in any form or by any means any part of this publication. Requests for obtaining the right to reproduce or utilize parts of this publication should be addressed to KU Leuven, Faculteit Wetenschappen, Celestijnenlaan 200H bus 2100, 3001 Leuven (Heverlee), telephone +32 16 32 14 01.

A written permission of the promotor is also required to use the methods, products, schematics and programs described in this work for industrial or commercial use, and for submitting this publication in scientific contests.

This thesis is an exam document that obtained no further correction of possible errors after the defense. Referring to this thesis in papers and analogous documents is only allowed after written consent of the supervisor(s), mentioned on the title page.

Foreword

Contribution statement

Summary

List of abbreviations

Contents

1 Introduction **1**

1.1 Role of phosphates in biological systems 1

1.2 Enzymes involved in phosphate hydrolysis 1

1.3 Reaction mechanism 1

1.4 Research goals 1

2 Theory **2**

2.1 A brief introduction to statistical mechanics 3

2.1.1 Classical forcefields and molecular dynamics 3

2.1.2 The canonical ensemble and free energy calculations 3

2.1.3 Enhanced sampling techniques 3

2.2 Density functional theory 3

2.2.1 The Kohn-Sham approach 3

2.2.2 Generalised gradient approximation and PBE functional 3

2.2.3 *Ab initio* molecular dynamics and GPW method 3

2.3 Extended tight binding 3

2.4 Neural network potentials 3

2.4.1 Deep neural networks 3

2.4.2 Invariance and equivariance 3

2.4.3 Behler-Parrinello neural network potentials 3

2.4.4 Equivariant neural network potentials 3

3 Computational Details **4**

3.1 Training dataset generation 4

3.1.1 System preparation 4

3.1.2 Initial equilibration using the classical forcefields 5

3.1.3 GFN1-xTB based exploration of the configuration space 6

3.1.4 Data labeling 6

3.1.5 Iterative training of the neural network potential 6

3.2 Production runs at different temperatures 6

3.3 Validation of the transition states 6
3.4 Data analysis and visualisation 6

4 Results and Discussion 7

5 Conclusions 8

Bibliography 9

A Supplementary information 10

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

2.4.3 Behler-Parrinello neural network potentials

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]:

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

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 TIP3 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 equilibration using the classical forcefields

The equilibration of the system was performed following the standard protocol generated by the `CHARMM-GUI` webserver [1]. The system was first energy minimised using the steepest descent algorithm for 5000 steps.

Subsequently the system was equillibrated in the NVT (constant number of particles, volume, and temperature) ensemble for 5 ns. During the minimisation and NVT equilibration, the heavy atoms of the solute were restrained using a harmonic potential with a force constant of 400 kJ mol^{-1} nm^{-2}.

As a last step, the system was equilibrated in the NPT (constant number of particles, pressure, and temperature) ensemble for 45 ns. Throughout the whole protocol, the temperature was set to 300 K and the pressure was set to 1 bar. To ensure the constant temperature and pressure, the system was coupled to a ν-rescale thermostat with a coupling constant of 1 ps and an isotropic *c*-rescale barostat with a coupling constant of 5 ps. During the NPT run, the cut-off for the non-bonded interactions was set to 0.6 nm and the long-range electrostatics were treated using the Particle Mesh Ewald (PME) method.

All simulations were conducted in `GROMACS 2021.4` [5] using the CHARMM36m forcefield [6] and the Leap-Frog integration method with a time step of 1 fs. All hydrogen involving bonds were constrained using the LINCS algorithm. The final dimensions of the box for all further calculations were obtained after the NPT run and are shown in Table 3.1.

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 GFN1-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.

[5] Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015). URL `https://www.sciencedirect.com/science/article/pii/S2352711015000059`.

[6] Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**, 71–73 (2017). URL `https://www.nature.com/articles/nmeth.4067`. Publisher: Nature Publishing Group.

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