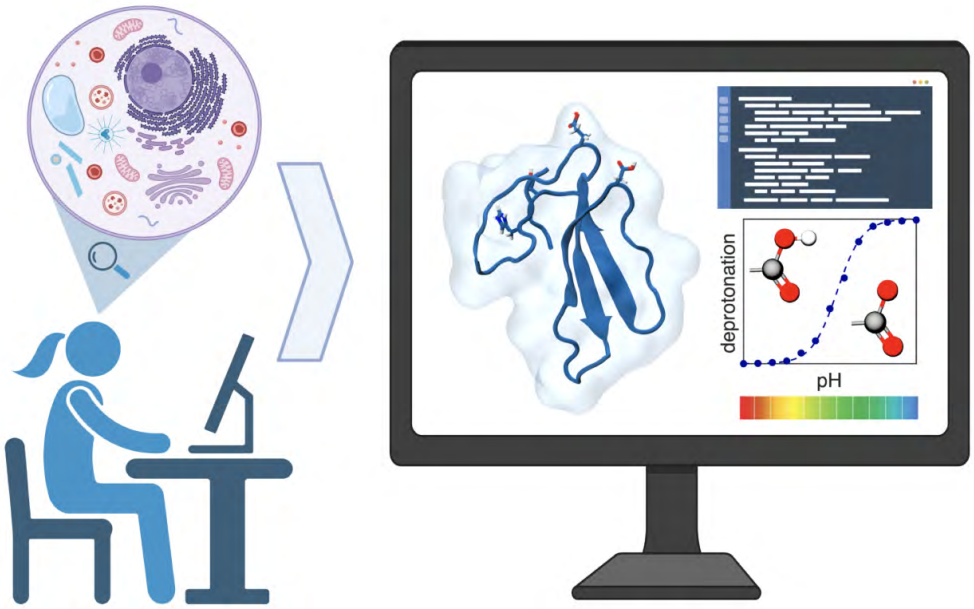
JYU DISSERTATIONS 614

Noora Aho

Molecular Dynamics Simulations of Acids and Bases in Biomolecular Environments



UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF MATHEMATICS
AND SCIENCE

JYU DISSERTATIONS 614

Noora Aho

Molecular Dynamics Simulations of Acids and Bases in Biomolecular Environments

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Ylistönrinteen auditoriossa Kem4
maaliskuun 31. päivänä 2023 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in building Ylistönrinne, auditorium Kem4, on March 31, 2023, at 12 o'clock.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2023

Editors
Gerrit Groenhof
Department of Chemistry
Päivi Vuorio
Open Science Centre, University of Jyväskylä

Cover picture by Noora Aho, edited using BioRender.

ISBN 978-951-39-9504-1 (PDF)
URN:ISBN:978-951-39-9504-1
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-9504-1

ABSTRACT

Aho, Noora
Molecular dynamics simulations of acids and bases in biomolecular environments
Jyväskylä: University of Jyväskylä, 2023, 78 p. (+ included articles)
(JYU Dissertations
ISSN 2489-9003; 614)
ISBN 978-951-39-9504-1 (PDF)

Proton transfer and pH are key features in determining the function of many biological entities, such as lipid membranes and enzymes. Local pH defines the protonation states of titratable groups, which by altering the electrostatic interactions can affect the chemical processes in these biomolecules. In addition to experimental research, continuously advancing computational methods are providing atomistic details of diverse biological processes. One of the common simulation methods for biological systems is molecular dynamics (MD). However, in conventional MD the protonation states of titratable groups such as amino acids in proteins are fixed, preventing the investigation of protonation events. To overcome this limitation, so called constant pH molecular dynamics has been developed during the last decades. In constant pH MD the pH is fixed and the protonation of selected groups can alter during the simulation, enabling the study of pH-dependent phenomena. In this thesis, MD simulations are performed to complement experimental findings of proton diffusion on the surface of phospholipid membranes. An accurate and efficient constant pH MD routine is implemented into the GROMACS simulation software, together with corrections to the CHARMM force field in order to more reliably sample the rotamers of titratable amino acids. Together, these results demonstrate the benefits of a *fast, flexible, and free* constant pH MD method in describing pH-dependent processes in biology.

Keywords: pH, proton transfer, protonation, molecular dynamics, force field constant pH molecular dynamics

TIIVISTELMÄ (FINNISH ABSTRACT)

Aho, Noora
Molekyylidynamiikkasimulaatioita hapoista ja emäksistä biomolekulaarisessa ympäristössä
Jyväskylän yliopisto, 2023

Useiden biologisten kokonaisuuksien, kuten entsyymien tai solukalvojen, toimintaan vaikuttaa niitä ympäröivän liuoksen happamuus, jota kuvataan yleisesti pH-arvolla. Paikallinen pH vaikuttaa atomitasolla biomolekyylien kemiallisten ryhmien protonaatioasteeseen eli siihen, onko kyseisiin ryhmiin sitoutunut protoni. Protonien sitoutuminen muuttaa molekyylien varauksia, jolloin niiden ja ympäristön välisten sähköisten vuorovaikutusten suuruus voi muuttua. Tällä on vaikutus biomolekyylien rakenteeseen ja niiden ominaisuuksiin, vaikuttaen lopulta koko biologisen kokonaisuuden toimintaan. Kokeellisen tutkimuksen lisäksi biomolekyylien toimintaa atomitasolla voidaan tutkia laskennallisesti modernien tietokonesimulaatiomenetelmien avulla. Yksi yleisimmistä biologisten systeemien tutkimiseen käytettävistä simulaatiomenetelmistä on molekyylidynamiikka (MD). Perinteisten Newtonin mekaniikkaan perustuvien MD-simulaatioiden rajoituksena on kuitenkin se, että kemiallisten ryhmien, kuten aminohappojen, protonaatioaste ja siten varaus ei voi muuttua simulaatioiden aikana. Tällöin pH:sta riippuvaisia muutoksia biologisissa kokonaisuuksissa ei voida suoraan tutkia simulaatioissa. Tämän väitöskirjatutkimuksen tarkoituksena on ensin täydentää kokeellisia havaintoja protoninsiirrosta fosfolipidikalvon pinnalla hyödyntämällä klassisia MD-simulaatioita. Väitöskirjan keskiössä on niin kutsutun vakio-pH MD-simulaatiomenetelmän jatkokehittäminen GROMACS-ohjelmistolle. Tämä laskennallisesti tehokas ja tarkka vakio-pH MD mahdollistaa protonaatioasteen muuttumisen epäsuorasti MD-simulaation aikana, jolloin pH-riippuvuus saadaan huomioitua simulaatioissa. Lisäksi vuorovaikutuksia kuvaavaan CHARMM-voimakenttään lisätään korjauksia, jotta aminohappojen sivuketjujen rotameerien mallintaminen olisi todenmukaisempaa vakio-pH simulaation aikana. Kokonaisuutena väitöskirjassa esitetyt tulokset havainnollistavat pH:n tärkeyttä simulaatioparametrina, ja vakio-pH MD:n merkitystä biologisten kokonaisuuksien laskennallisessa tutkimuksessa.

Avainsanat: pH, protoninsiirto, protonaatio, molekyylidynamiikka
voimakenttä, vakio-pH molekyylidynamiikka

Author

Noora Aho
Department of Chemistry
Nanoscience Center
University of Jyväskylä
Jyväskylä, Finland

Supervisor

Professor Gerrit Groenhof
Department of Chemistry
Nanoscience Center
University of Jyväskylä
Jyväskylä, Finland

Reviewers

Professor Matthias Ullmann
Theoretical and Computational Biochemistry
University of Bayreuth
Bayreuth, Germany

Professor Andre Juffer
Biocenter Oulu
University of Oulu
Oulu, Finland

Opponent

Professor David van der Spoel
Department of Cell and Molecular Biology
Uppsala University
Uppsala, Sweden

PREFACE

The research presented in this thesis was carried out at the Nanoscience Center, Department of Chemistry at the University of Jyväskylä during 2018-2022. I would like to acknowledge Academy of Finland for funding (grant 311031), and CSC Finnish IT Center for Science and Finnish Grid Infrastructure for providing computational resources.

First of all, I would like to express my gratitude to my supervisor Professor Gerrit Groenhof, for giving me an opportunity to explore the simulation world of biomolecules and to grow as a young scientist. I would like to thank Dr. Pavel Buslaev for support and supervision during the research, especially for numerous bright ideas and co-implementation of the constant pH code.

I am beyond grateful for the active collaboration with the GROMACS developers Professor Berk Hess and Dr. Paul Bauer, who's efforts made the constant pH code from a rough sketch to an actual working product. I also thank Anton Jansen for peer support along the way. For giving me an opportunity to consolidate his experimental results with simulations, I would like to thank Associate Professor Nadav Amdursky from the Israel Institute of Technology.

I am thankful for the discussions, guidance and everyday company of all the current or previous members of our research group: Dr. Dmitry Morozov, Dr. Ruth Tichauer, Dr. Emmi Pohjolainen, Vaibhav Modi, Ilya Sokolowski and Kushagri Arora. In addition, a warm thank you to the whole community of the Nanoscience Center, for creating an inspiring atmosphere for carrying out research.

I also thank the pre-reviewers of this thesis, Professor Matthias Ullmann and Professor Andre Juffer for their time to read my work and offer their valuable comments.

Haluan kiittää jokaista läheistäni ja ystävääni, joka on uskonut tämän pienen tutkijanalun löytävän vielä joskus perille. Vaikka harva teistä on todella ymmärtänyt laskennallisen nanotieteen tutkimukseni syvintä sisintä, olette kannatelleet minua elämäni poluilla, ja se jos mikä on ollut arvokasta. Kiitos erityisesti Jyväskylän naisfyysikoille lounasseurasta, vertaistuesta, sekä kaikista hiihto- ja suunnistusretkistä, jotka pitivät mielen kirkkaana tieteen keskellä. Kiitos kaikille etelän ja pohjoisen ystäville, jotka ottivat tämän tutkijanalun usein viikonlopun viettoon tai etätyömökkiviikolle tuulettamaan ajatuksia. Isoin kiitos siskolleni Niinalle, joka on arkena tsempannut tutkimaan ja muulloin ollut aina valmis pieneen tai isompaan seikkailuun pakoon arjen aherrusta.

Viimeisenä suurkiitos ja tuhannet rapsutukset koiralleni Pojulle, joka piti uskollisesti huolta hyvinvoinnistani väitöstutkimuksen aikana ilahduttamalla jokaista etätyöpäivääni ja pakottamalla karkaamaan kävelyille keskellä päivää.

Jyväskylä, March 2023

Noora Aho

ABBREVIATIONS

AA	all-atom
AWH	accelerated weight histogram
CG	coarse-grained
CI	charge interpolation
CpHMD	constant pH molecular dynamics
CPU	central processing unit
cryo-EM	cryogenic electron microscopy
ESPT	excited-state proton transfer
FF	force field
GAFF	generalized Amber force field
GB	generalized Born
GPU	graphics processing unit
L-J	Lennard-Jones
MD	molecular dynamics
MM	molecular mechanics
MS	multisite
NMR	nuclear magnetic resonance
NPT	isothermal-isobaric ensemble
NVE	microcanonical ensemble
NVT	canonical ensemble
PB	Poisson-Boltzmann
pHREX	pH replica-exchange
PBC	periodic boundary conditions
PD	proton diffusion
PDB	Protein Data Bank
PME	particle mesh Ewald
POPA	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-1-rac-glycerol
PT	proton transfer
QM	quantum mechanics
REX	replica-exchange
STD	standard deviation
TI	thermodynamic integration

LIST OF INCLUDED ARTICLES

I Nadav Amdursky, Yiyang Lin, Noora Aho, Gerrit Groenhof. Exploring fast proton transfer events associated with lateral proton diffusion on the surface of membranes. *Proceedings of the National Academy of Sciences*, 6*(7)*, 2443-2451, **2019**.

II Noora Aho, Pavel Buslaev, Anton Jansen, Paul Bauer, Gerrit Groenhof, Berk Hess. Scalable constant pH molecular dynamics in GROMACS. *Journal of Chemical Theory and Computation*, *18(10)*, 6148-6160, **2022**.

III Pavel Buslaev, Noora Aho, Anton Jansen, Paul Bauer, Gerrit Groenhof, Berk Hess. Best practices in constant pH MD Simulations: accuracy and sampling. *Journal of Chemical Theory and Computation*, *18(10)*, 6134-6147, **2022**.

For publication **I**, the author of this thesis took part in designing of the computational part of the research and executed the plan. She performed the force field parameterization of the photoacid, assembled the simulation systems, ran all molecular dynamics simulations and carried out the analysis. She participated in writing the computational part of the manuscript.

For publications **II** and **III**, the author implemented the revised constant pH molecular dynamics (CpHMD) into the GROMACS software. She contributed to the main parts of the code and extensively tested the performance of the implementation, working in close collaboration with the other authors of the publications. She designed, performed and analyzed CpHMD simulations and participated in writing the manuscripts for both publications.

CONTENTS

ABSTRACT
TIIVISTELMÄ (FINNISH ABSTRACT)
PREFACE
ABBREVIATIONS
LIST OF INCLUDED ARTICLES
CONTENTS

1 INTRODUCTION 13

2 PH AND PROTON TRANSFER IN BIOLOGY 15

3 MOLECULAR DYNAMICS 24
3.1 Dynamics 25
3.2 Thermodynamics 26
3.3 Potentials 28
3.4 Simulation systems 30
3.5 Limitations 31

4 CONSTANT PH MOLECULAR DYNAMICS 33
4.1 A brief history of constant pH MD 34
4.2 Constant pH MD for GROMACS 39
4.2.1 λ-dynamics 39
4.2.2 Multisite titratable groups 44
4.2.3 Charge neutrality 46
4.2.4 Applicability and limitations 46

5 RESULTS 48
5.1 Proton diffusion on the surface of phospholipid membranes (publication **I**) 48
5.2 Fast, flexible, and free constant pH MD for GROMACS (publication **II**) 51
5.3 Best practices in constant pH MD simulations (publication **III**) 58

6 SUMMARY AND OUTLOOK 66

REFERENCES 69

INCLUDED ARTICLES

1 INTRODUCTION

As scientists, our aim is to understand nature. Traditionally, this understanding has grown from carrying out experiments, measuring relevant quantities and drawing conclusions from the data. With the constantly growing power of modern computers, however, computational simulation techniques have become a comparable tool to study the laws, processes and phenomena of nature. From a simulation point of view, understanding means developing models and simulation techniques that are able to reproduce, supplement and predict experimental findings.

If one is interested in studying the dynamics of biomolecules at atomistic scale, as the author of this thesis, one of the most common computational methods is classical molecular dynamics. Based on the fundamental laws of classical mechanics, it can be applied to various sizes of biological systems ranging from single molecules to entities such as viruses, spanning timescales from pico- to milliseconds.

With the goal of mimicking experimental conditions, parameters such as temperature and pressure are usually kept constant in classical MD. In turn, the possibility to fix the pH of the surrounding solution constant during MD has not conventionally been possible. Being unable to simulate at constant pH, and dynamically change the protonation states of titratable groups prevents investigating pH-dependent phenomena of biological systems with MD.

In the recent decades, however, so-called constant pH MD approaches have been developed, which include pH as an input parameter and allow dynamic changes of the protonation states and thus atomic charges. In this thesis, the author presents an accurate and efficient constant pH MD approach implemented in

GROMACS, which is one of the most popular open-access MD software used to study biological systems computationally. The implementation, together with best practices, provides the GROMACS user community a possibility to explore the world of pH-dependent processes in biology, while openly providing details of the actual constant pH program.

To summarize, the objective of this thesis is to connect the concepts of pH and proton transfer from the biological point of view into the area of computational techniques, especially molecular dynamics simulations. This is achieved by:

(1) presenting an example of computational work to support experimental findings of proton diffusion on the surface of membranes (publication **I**)
(2) emphasizing the importance of including pH dependence in MD simulations, by providing revised constant pH MD method for the popular GROMACS simulation software (publication **II**), and
(3) providing general best practices for constant pH MD, together with corrections to a common force field to better sample rotameric configurations of amino acids (publication **III**)

The structure of the thesis is as follows: in the beginning, the basic theory of pH, proton transfer, MD and constant pH MD are introduced, after which the main results of the three publications listed above are presented and further discussed. Together, the thesis provides a comprehensive look into the perspectives of pH and protonation in MD simulations.

2 PH AND PROTON TRANSFER IN BIOLOGY

pH is a scale for describing the acidity or basicity of an aqueous solution. The higher the concentration of protons (H^+), or correspondingly the concentration of hydronium ions (H_3O^+) in a solution, the more acidic it is measured, and vice versa. By definition, pH is the decimal logarithm of the proton activity a_{H^+} in aqueous solution:

$$\mathrm{pH} = -\log_{10} a_{H^+} \approx -\log_{10}[H^+] \tag{1}$$

The latter approximate definition states that when considering common applications with low concentration of protons, the proton activity, meaning the effective concentration of a non-ideal solution, can be replaced with the proton concentration of the aqueous solution denoted with square brackets.

In chemistry and biology, solution pH is a key factor for determining the structure, function and dynamics of various macromolecules. As examples of pH dependent processes in nature, most enzymes have evolved to require a narrow optimal pH range in order to reach the highest rate of activity, [1] and the ATP synthase protein utilizes a transmembrane pH gradient to produce ATP as an energy source for the cells. [2] As another example, changing the intracellular pH affects multiple processes, such as cell growth, metabolism, and membrane potential and transport. [3]

A simple acid-base dissociation reaction for a monoacid reactant can be written as $HA \rightleftharpoons A^- + H^+$, where compound HA is the acid, A^- the base and H^+ the hydrogen ion, or proton. For an acid-base reaction the equilibrium constant is called the dissociation coefficient K_a, which describes the relative amounts of the product and the reactant (namely the base and the acid) after the system has reached equilibrium. The dissociation constant, however, does not specify the

rate of the chemical reaction, but only its directionality. K_a is commonly defined as $\mathrm{p}K_\mathrm{a}$, using the negative logarithm expression

$$\mathrm{p}K_\mathrm{a} = -\log_{10} K_\mathrm{a} = -\log_{10} \frac{a_{\mathrm{A}^-} a_{\mathrm{H}^+}}{a_{\mathrm{HA}}} \approx -\log_{10} \frac{[\mathrm{A}^-][\mathrm{H}^+]}{[\mathrm{HA}]} \tag{2}$$

As for the pH, the above expression for the dissociation constant is formally defined for the activities of the compounds in solution a_i. However, in usual applications where the aqueous solution has a relatively low concentration, activities can again be replaced by concentrations. Similarly to any chemical reaction, the acid dissociation constant is also related to the standard Gibbs energy of the reaction. A change in the $\mathrm{p}K_\mathrm{a}$ or a certain compound corresponds to a change in the Gibbs energy of the reaction as follows:

$$\Delta G^\ominus = -RT \ln K_\mathrm{a} = (RT \ln 10)\mathrm{p}K_\mathrm{a} \tag{3}$$

Acid-base reactions are among the key reactions in biological systems, and being able to quantitatively analyse the behaviour of these compounds in the contexts of biological processes is essential. An acid donating a proton, and thereby changing its protonation state alters the electrostatic interactions of the compound relative to the surroundings. As the physical interactions and forces change at the atomistic scale, the structure and even function of the entire macromolecule may change.

For the scope of this thesis, a particularly interesting group of chemical compounds is the organic amino acids. The most common alpha-amino acids consist of an amino group, a carboxylic group and a side-chain group bonded to a central α-carbon. A chain of amino acids connected from the amino and carboxylic acid groups forms a peptide, and the peptides fold into proteins which take part in many of the essential processes of cells. More precisely, in this thesis we are mostly interested in *titratable* amino acids, which refer to those able to donate or accept one or more protons in the side chain: aspartic and glutamic acids, histidine, lysine, arginine, cysteine and tyrosine. Example structures, together with the $\mathrm{p}K_\mathrm{a}$ values of the amino acid side chains in water, are shown in Figure 1. It is important to notice that moving from aqueous solution to a different electrostatic environment, such as a protein, the $\mathrm{p}K_\mathrm{a}$ value of the side chain group may change.

Depending on the pH of the surrounding solution and the local electrostatic environment in a protein, a titratable amino acid can be in different protonation states, at simplest protonated (acid) or deprotonated (base). The protonation states define the charges and the electrostatic interactions, and changes in

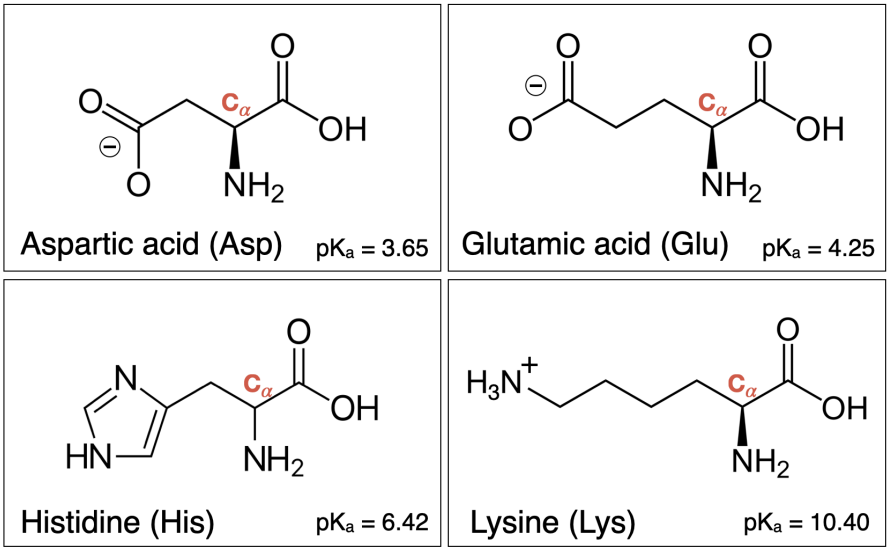


FIGURE 1 Structures of four titratable amino acids being particularly considered in this thesis: glutamic acid (Glu), aspartic acid (Asp), histidine (His) and lysine (Lys). The pK_a values for the sidechains of these groups are given [4,5]. For clarity, only the macroscopic pK_a for His is given. For all amino acids, the structure of the most probable protonation state at physiological pH = 7 is shown: deprotonated Glu and Asp (- COO^-), deprotonated His-E with the proton at Nϵ, and protonated Lys (- N_3H^+).

protonation influence the structure and dynamics of proteins. To demonstrate, Figure 2 shows an illustration of a protein with the key titratable amino acids and dependence between protonation states and the pH. As mentioned above, for example many enzymes have a relatively narrow optimal pH range. If the pH of the surrounding solution is altered to a value beyond the optimal range, the protonation states of titratable amino acids in this enzyme will change. This can influence the structure of the active site, prohibit the optimal substrate binding and the catalysis process, leading to even completely losing the catalytic function of the enzyme. [1,6]

Knowing the pK_a values of certain groups is critical for describing a biological process, which is why being able to measure these values is of high importance. pK_a values of titratable groups are commonly measured using titration methods, both experimental and computational, where the common underlying idea is to compute the relation between the concentration of the acid and the pH in solution. The famous relation between pH, pK_a and measured concentrations is denoted as Henderson-Hasselbalch equation:

$$\mathrm{pH} = \mathrm{p}K_a + \log_{10} \frac{[\mathrm{A}^-]}{[\mathrm{HA}]} \tag{4}$$

which can be straightforwardly derived from the definitions of pH and pK_a

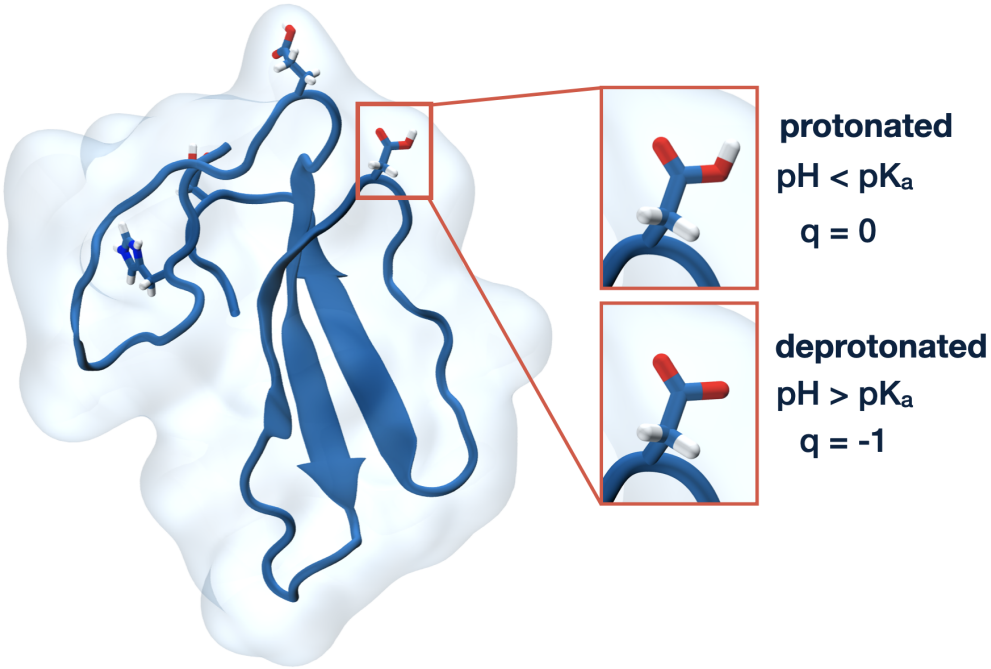


FIGURE 2 Effect of pH on the protonation states of amino acid sidechains. As an example, the water-soluble cardiotoxin V protein (PDB ID: 1CVO [7]), with important titratable groups shown as stick representation. The zoomed-in view shown one of the titratable aspartic acids. If the pH of the surrounding solution is higher than the pK_a of the Asp sidechain, the carboxyl group prefers the deprotonated state with negative charge. Vice versa, if the pH is lower than pK_a, the protonated neutral state is preferred.

defined in in Equations 1 and 2, respectively. With the scope of this thesis, the validity of the Henderson-Hasselbalch equation lies on the following assumptions: the concentration of the protons in solution is low enough, the acid of interest is monobasic following the simple dissociation reaction $HA \rightleftharpoons A^- + H^+$, and the acid is completely dissociating in water and not forming for example ion-pairs.

The first experimental method from early 1900's to determine the pK_a and to obtain a titration curve was potentiometric titration, where a known amount of titrant is added step-wise to the unknown analyte solution. [8, 9] The change in potential, corresponding to pH, is measured upon the reaction. An example of a typical titration curve from a potentiometric titration experiment is presented in the left panel of Figure 3. The equivalence point is reached as the amount of added titrant completely neutralizes the analyte solution, and at the half-equivalence point exactly half of the acid in the analyte solution has reacted with the titrant, resulting in $[HA] = [A^-]$ or $pH = pK_a$. However for extreme pH values, higher temperatures, mixtures or complex molecules, other more precise methods have been developed over the years. For example, pK_a values for smaller compounds such as drug molecules and photoactive dyes can be measured using spectrophotometry. [10]. For proteins with multiple titratable

amino acids determining pK_a values becomes more challenging, and practically only nuclear magnetic resonance (NMR) titration is capable to fully characterise the pK_a for each titratable site. [8,11] As the chemical shifts of NMR-active nuclei depend on protonation states, measuring those as a function of solution pH enables the determination of pK_a values by NMR.

The lack of simple experimental techniques to determine pK_a values for proteins and other larger biomolecules has encouraged the development of computational methods, especially during the last decades. The basic idea behind most of the computational approaches is to estimate the free energy $\Delta G^{\ominus}$ shift, and thus the pK_a shift, related to transferring the titratable group from solvent to the protein environment presented as

$$\mathrm{p}K_a(\text{protein}) = \mathrm{p}K_a(\text{solvent}) + \Delta\mathrm{p}K_a(\text{solvent} \rightarrow \text{protein}) \tag{5}$$

as the free energy $\Delta G^{\ominus}$ is directly related to the pK_a as shown in equation 3.

The largest effect on the pK_a shift from the solvent to the protein environment arises from changes in the electrostatic interactions, whereas the van der Waals interactions contribute much less. [12, 13] Therefore, the key of computational methods to determine pK_a values for titratable groups in proteins, relies on calculating the electrostatic interactions accurately enough. There are different levels of theory to describe the electrostatic interactions and to estimate the pK_a values, often divided to (i) empirical methods, (ii) macroscopic methods based on continuum electrostatics and (iii) microscopic methods calculating the electrostatic interactions at the atomic level.

Firstly, so-called empirical methods, [14–17] such as PROPKA [16] use the structure of the protein to identify the location, interactions with the local environment and desolvation effect of titratable groups, and based on these estimate the pK_a values by comparing to a large data set of measured values. Despite being faster than macroscopic or microscopic approaches, the mentioned methods rely entirely on static protein structures, and may result in remarkable errors for example flexible proteins, deeply buried titratable sites or transmembrane proteins. [18, 19] Recently, machine learning based methods to predict pK_a values of titratable groups in proteins have also been developed, [20, 21] which use experimental pK_a databases for training the machine learning model.

The second category consists of macroscopic methods which utilize implicit continuum electrostatics to compute the electrostatic part of the free energy related to changing the local environment of the titratable group from solvent to the protein. They are based on techniques numerically solving the Poisson-

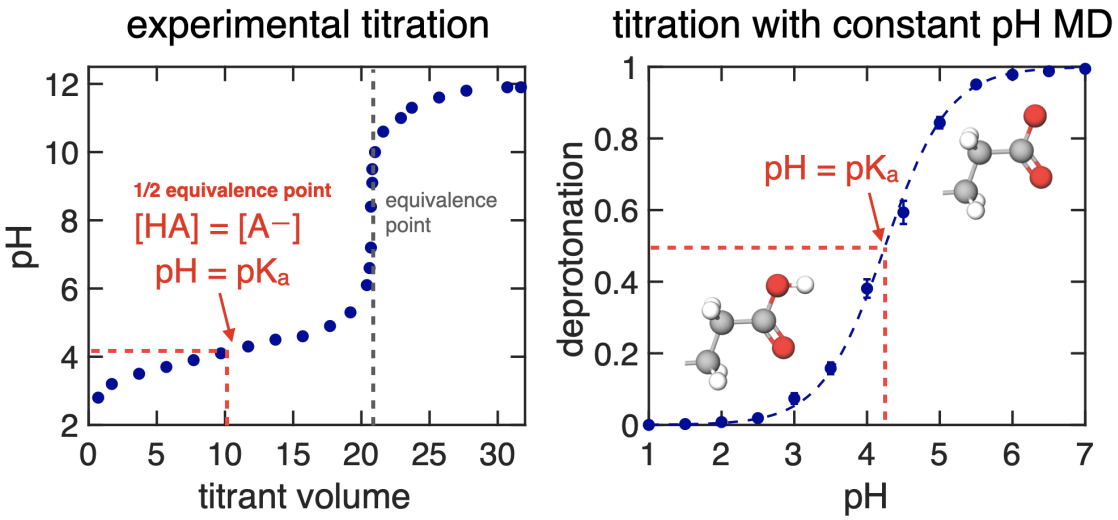


FIGURE 3 An example titration curve of glutamic acid side chain from experimental potentiometric titration (left panel) and computational constant pH MD simulation (right panel). Experimentally, the curve illustrates the relationship between the titrant volume and pH, and the pK_a is read from the half-equivalence point (1/2 equivalence point), where exactly half of the acid in the analyte solution has reacted with the titrant. In the computational titration, the amount of deprotonated acid during the simulation is plotted as a function of pH, and the pK_a is read from the pH at which half of the compound is deprotonated. The dashed line represents the fit to the Henderson-Hasselbalch equation.

Boltzmann (PB) equation [22–30] or the Generalized Born approach. [31, 32] In these methods, the protein and solvent are modelled as continuous media with low and high dielectric constants, respectively, and the titratable groups inside protein as point charges. To improve the pK_a estimation, the macroscopic methods have also been refined and extended to include for example conformational flexibility of the proteins. [33, 34]

Lastly, the atomic level microscopic methods are more accurate but also more computationally expensive than the continuum based macroscopic ones. In addition to the atomic level description of electrostatic interactions, their advantage lies in the ability to include dynamic motion of the protein, which is important for the local interactions of titratable groups and thus pK_a values. Despite the increasing power of modern computers, the most fundamental microscopic quantum mechanical (QM) methods [35, 36] are still excessively expensive for size scales of a protein system. Instead, hybrid quantum mechanics/molecular mechanics techniques (QM/MM) or even entirely classical molecular mechanics (MM) techniques with implicit or explicit solvent description are applicable to determine the free energies. [13, 37, 38]

In addition to estimating the pK_a values of titratable groups, it is essential to gain knowledge of the *dynamics* of the protein or biomolecule at a specific pH,

and connect changes in dynamics to changes in protonation states. Molecular dynamics (MD) is one of the widely used computational methods to study atomic level motions and interactions at the size scales of proteins. However, classical MD does not enable changing the protonation states during the simulations, but instead those must be set prior the simulations. This is especially problematic for protein systems where the pK_a values are close to the pH and the choice for the protonation states, based on the aforementioned methods, becomes difficult.

In the last two decades so-called constant pH molecular dynamics (CpHMD) methods have also been developed, and as they are the focus area in this thesis, the CpHMD methods are reviewed in more detail in Chapter 4. In addition to pK_a estimation CpHMD can reveal details of the dynamics behind the protonation events, and couple the sampling of atomic coordinates to sampling of protonation states. Thus, CpHMD enables the investigation of pH-dependent processes at the atomic level using computers. To demonstrate the possibility of computational pK_a estimation, a typical titration curve obtained using CpHMD is presented in Figure 3, together with the experimental one. In this thesis, we have utilized both classical MD simulations with fixed protonation states and CpHMD simulations, both of which have their strengths depending on the system of interest.

As the acid-base dissociation takes place, the released proton H^+ has to *transfer* somewhere. The process of proton transfer from one molecule to another is a central phenomena in various biological processes, such as proton diffusion on the surface of membranes. If we consider an acid AH undergoing the dissociation reaction and a negative base B^- accepting the proton, the proton transfer reaction can be presented as

$$B^- + HA \rightleftharpoons [B^- \cdots HA] \rightleftharpoons [BH \cdots A^-] \rightleftharpoons BH + A^-$$

where the square brackets denote intermediate hydrogen bonded complexes. [39] Proton transfer can also occur from a protonated molecule to a neutral molecule. As a chemical bond is broken while a proton transfer event takes place, accompanied with motion along multiple degrees of freedom, the required free energy for proton transfer can be relatively high. The acidity of the donor HA and the basicity of acceptor B^- in the local environment, namely their pK_a values, define the extent of proton transfer events and the free energy required in order for the transfer to occur.

One of the fundamental proton transfer processes in biology is proton diffusion (PD) across membranes. In order to follow PD and these fast proton transfer events at nanosecond time scales, one possible technique is to embed a photoacid (aryl-OH molecule) into a lipid vesicle mimicking the membrane. Upon excitation of the light-induced photoacid a proton is released, and its diffusion

along the membrane can be followed. [40] For the purpose of this thesis, an example of such photoacid is shown in Figure 4, together with the excited-state proton transfer (ESPT) scheme.

In ESPT, the photoacid is excited, and upon light excitation due to a large shift in the pK_a between the ground (ROH) and excited (ROH*) states of the acid, it releases a proton during the excited-state lifetime. As the proton is released with rate k_{PT}, ROH* deprotonates to RO^{-*}. The released proton can then further diffuse to the surroundings or recombine to reform ROH* with rate k_a.

The rate coefficients k_{PT} and k_a for ROH*, together with the diffusion constant and dimensionality, can be obtained experimentally using time-resolved fluorescence techniques. The measured intensity of the ROH* decay profile is fitted to a theoretical model, from which the parameters are resolved. [40] To further investigate the behaviour of the system in addition to experiments, classical MD simulations can be performed to reveal details of the photoacid orientation in the membrane and the interactions between the proton-releasing -OH group and the surroundings.

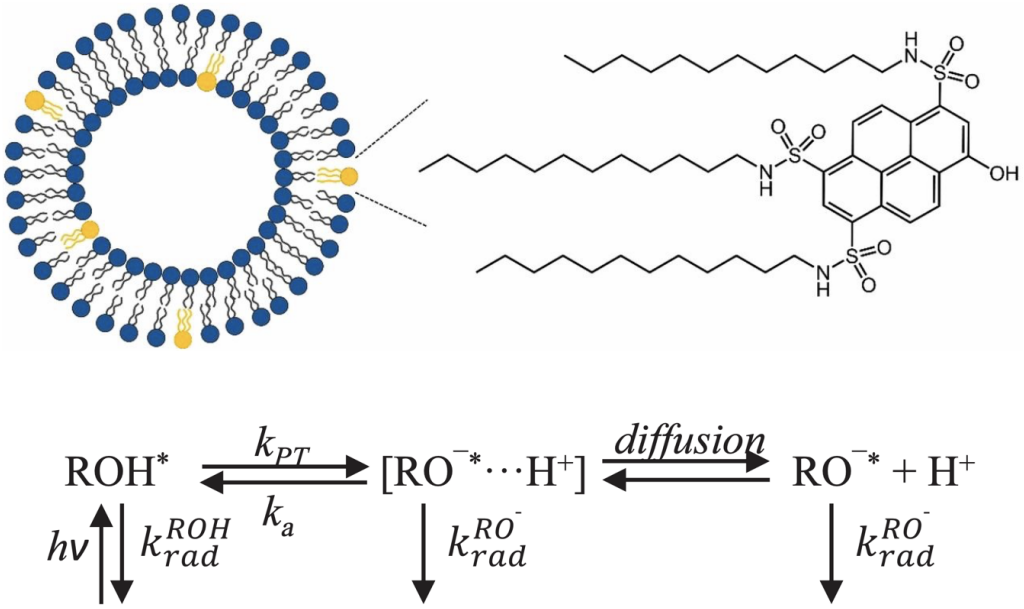


FIGURE 4 Above: Molecular scheme of a photoacid (C12-HPTS) embedded into a lipid vesicle, and the atomic structure of this photoacid. Below: The excited-state proton transfer (ESPT) cycle of a photoacid. The excited state photoadic is denoted as ROH*. Reprinted with permission from Nadav Amdursky, Lin Yiyang, Noora Aho and Gerrit Groenhof. Exploring fast proton transfer events associated with lateral proton diffusion on the surface of membranes. *Proceedings of the National Academy of Sciences* 116, no. 7 (2019): 2443-2451. Copyright 2019 National Academy of Sciences.

To conclude, pH and proton transfer play an important role in biological processes by controlling the function of various biological entities. Changing the protonation states of titratable groups, by accepting or donating a proton leads to changes in the electrostatic interactions. By modelling these interactions reliably, we are able to simulate the dynamics of both atomic positions and protonation states of biomolecules computationally, for which the methods will be introduced in the next chapters.

3 MOLECULAR DYNAMICS

The origin of biological processes lies in the interactions between atoms and molecules. As the basic physical laws of nature regarding interactions and forces between molecules are known, theoretical predictions of the corresponding processes are possible. Computer simulations have become a standard tool to investigate atomistic scale phenomena in biochemistry. [41–45] In the recent decades, the power and efficiency of computers has increased remarkably, and new algorithms, software and hardware are constantly developed to better fulfill the needs of researchers. Already since decades, molecular dynamics (MD) has been one of the key computational methods to study various biological systems, from picosecond to millisecond timescales, and from small biomolecules to cell-scale complexes.

Within the scope of this thesis, the focus lies on classical MD simulations of proteins and membranes at atomistic scale, with timescales up to a microsecond. For the research presented in this thesis, MD provided the possibility to use classical level description to explore the phenomena related to proton transfer and protonation events in phospholipid membranes, complementing experimental findings (publication **I**). In addition to performing classical simulations, a constant pH MD method was implemented for the popular MD software (GROMACS), enabling the sampling of protonation states along with conventional dynamics (publications **II** and **III**).

The underlying idea of a molecular dynamics simulation is to take an atomic structure of a *molecular* system, and evolve the time-dependent *dynamics* of the system by applying classically described principles of physical interactions. In practice, the system is built to mimic an experimental setup, often utilizing experimentally resolved structures of biological macromolecules. Then, the

simulation propagates in time by the laws of Newtonian mechanics for which the force is integrated from the chosen potential function describing the physical interactions between the atoms in the system. From the simulation trajectory, which combines the coordinates of all atoms resolved for each time step, various kinetic and thermodynamic properties can be analyzed in order to compare and validate the used model, and to discover novel knowledge of atomistic scale properties of the studied system.

In this chapter, the basic principles of MD are introduced, together with discussion on the choices made within the scope of this thesis and limitations of the method. In the next chapter 4, the theory of MD is extended in order to take protonation dynamics into consideration.

3.1 Dynamics

The Hamiltonian function $H(\mathbf{R})$ specifying the total energy of a classical simulation system of N atoms can be written as the sum of kinetic $K(\mathbf{R})$ and potential $V(\mathbf{R})$ energies:

$$H(\mathbf{R}) = K(\mathbf{R}) + V(\mathbf{R}) = \sum_{i}^{N_{\text{atoms}}} \frac{m_i}{2} \dot{\mathbf{r}}_i^2 + V(\mathbf{R}) \tag{6}$$

where $\mathbf{R}$ contains the Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ and velocities $\dot{\mathbf{r}} = (\dot{\mathbf{r}}_1, \dot{\mathbf{r}}_2, \ldots, \dot{\mathbf{r}}_N)$ of all N atoms in the 3D cartesian space. The potential energy term $V(\mathbf{R})$ is revised in detail in the next section.

In the course of the MD simulation, both the positions $\mathbf{r}_i$ and velocities $\mathbf{v}_i$ of each atom in the system are evolved by solving the differential equations of Newton's second law:

$$m_i \frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} = \mathbf{F}_i(\mathbf{R}), \qquad i = 1 \ldots N \tag{7}$$

where $\mathbf{F}_i$ is the force acting on each individual atom i. The force is computed by solving the gradient of the interaction potential $V(\mathbf{R})$ as

$$\mathbf{F}_i(\mathbf{R}) = -\frac{\partial V(\mathbf{R})}{\partial \mathbf{r}_i} \tag{8}$$

Practically, the equations of motions are impossible to be solved analytically for the many-body system of N atoms, as the motions of these atoms are coupled. Therefore, finite difference methods have to be used, where the integration is performed in small timesteps Δt. [46] The assumption is that the force $\mathbf{F}_i$ is constant during a timestep and calculated for each atom i. The positions and velocities are calculated using algorithms based on formulating the positions and velocities as truncated Taylor series expansions, and most popular algorithms being the Verlet [47] and leap-frog [48] algorithms. In addition to aforementioned, also stochastic (Langevin) dynamics can be used in molecular simulations. In Langevin dynamics, an additional term for friction and noise are added to the Newton's equations of motion. [49,50] The idea behind these terms is to describe damping due to surrounding solution or air and perturbation due to occasional collisions, and to control the temperature of the system. For a schematic diagram of the MD routine, please see the corresponding illustration in Figure 9 of the Results chapter.

3.2 Thermodynamics

In an experiment, macroscopic average properties over a large number of particles and long enough time are measured. From a simulation trajectory, which is a collection of frames along a much shorter time than real-world experiments, only microscopic properties can be calculated. The essence of being able to reliable compare MD simulations to experiments lies in the ergodic hypothesis: it is assumed, that during the long enough stochastic simulation the system would sample sufficiently the typical configurations, after which the macroscopic properties can be obtained as time averages of the microscopical quantities by means of statistical mechanics. [51]

Statistical mechanics connects the microscopic ensembles of a large number of particles to the macroscopic, measurable world. In terms of statistical mechanics, for a classical system in thermal equilibrium the ensemble average for quantity A can obtained by integrating over the phase space, and be written as [52]

$$\bar{A} = \frac{\int A e^{-H(\mathbf{r},\mathbf{p})/kT} \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{p}}{\int e^{-H(\mathbf{r},\mathbf{p})/kT} \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{p}} \tag{9}$$

where $\bar{A}$ is the ensemble average, $H(\mathbf{r},\mathbf{p})$ the Hamiltonian of the system in terms of $3N$ position coordinates $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ and $3N$ momenta coordinates $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N)$ of the N atoms in the system, k the Boltzmann constant

and T the temperature of the system.

As the ergodicity assumption holds, and the MD simulation is assumed to sample the relevant phase space reliably over the simulation time, the ensemble average may be replaced by a time average over the discrete simulation steps τ_i [52, 53]:

$$\bar{A} = \frac{1}{\tau} \sum_{\tau_i=1}^{\tau} A(\mathbf{r}(\tau_i), \mathbf{p}(\tau_i)) \tag{10}$$

where τ denotes the total simulation time, and $A(\mathbf{r}(\tau_i), \mathbf{p}(\tau_i)$ is the quantity A that is measurable at each time step τ_i in the simulation. Crucial is that the quantity A can be expressed as a function of positions $\mathbf{r}$ and momenta $\mathbf{p}$ of the atoms recorded in the simulation. The above also assumes that the observed quantity does not depend on the initial coordinates and momenta.

If all forces in the system are due to the interaction potential energy of the system, the total energy (E) of the simulation system is conserved. If in addition the number of atoms (N) and the volume (V) of the simulation unit cell are constant, the MD simulation is being performed in the microcanonical (NVE) ensemble. Furthermore, if the temperature (T) instead of total energy is kept constant, the simulation is said to be performed in the canonical (NVT) ensemble. The NVT ensemble is often used in the shorter equilibration simulations prior the actual MD. Most often the actual MD simulations are then performed in the isobaric-isothermal (NPT) ensemble, mimicking the experimental conditions, where both temperature and pressure (P) are kept constant.

Maintaining constant temperature and pressure during MD is done by utilizing thermo- and barostats, for which common methods include the Berendsen, [54] stochastic velocity rescale (v-rescale) [55] or Nose-Hoover [56, 57] thermostats, and Berendsen, [54] Parrinello-Rahman [58] and the recent cell rescaling (c-rescale) [59] barostats. Currently, in the GROMACS software [60] which was used for MD simulations in the work related to this thesis, the v-rescale and c-rescale thermo- and barostats are recommended, as they both sample the canonical ensemble correctly and decay fast to the desired temperature or pressure. Before these rescaling algorithms were introduced, a common approach was to use the Berendsen thermo- and barostats during equilibration simulations to converge fast to the desired values, after which change to Nose-Hoover or Parrinello-Rahman algorithms in order to preserve the correct ensemble.

3.3 Potentials

In classical MD, quantum mechanical effects and particles such as electrons are ignored in the simulations, and the total potential energy of the system is calculated by only considering the nuclear positions of the atoms. [46]. Validation for this simplification originates in the Born-Oppenheimer approximation, which states that nuclei and electrons can be considered separately due to their significant mass difference. [61] The interaction potential $V(\mathbf{R})$ in Equation 6 can have various forms, depending of the level of accuracy in describing the physical interactions between the individual atoms in the system.

The functional forms of these potentials, with a set of parameters to describe the atoms, is called a *force field*. Commonly, a force field consists of relatively simple intra- and intermolecular interaction terms, that are often divided to bonded and non-bonded interactions. Bonded interactions include the stretching of covalent bonds, opening and closing of angles defined with three atoms, and dihedral rotations around bonds defined with four atoms. All bonded interactions originate from "energetic penalties" due to deviations of bond lengths or angles from their equilibrium values. Non-bonded terms include the pair-additive van der Waals and electrostatic interactions between the atoms in the system. One example of a common force field description (of the CHARMM36 force field, [62], used for many simulations of this thesis) is:

$$
\begin{aligned}
V(\mathbf{R}) = & \sum_{\text{bonds}} \frac{1}{2} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi (1 + \cos(n\phi - \delta)) \\
& + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{U-B}} k_u (u - u_0)^2 \\
& + \sum_{i,j<i} 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) + \sum_{i,j<i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}
\end{aligned}
\tag{11}
$$

In the above expression, the first five potential terms describe the bonded interactions. Bonds and angles are modelled as harmonic potentials, with bond length b between two atoms, angle θ between three atoms, equilibrium values b_0 and θ_0, and force constants k_b, and k_θ. The bond lengths and angles are calculated from the atomic positions $\mathbf{r}_i$. The equilibrium values and force constants are defined in the force fields. When simulating biological systems with MD, the highest frequency vibrations of hydrogen bonds are usually prevented by using

special constraints to fix their bond lengths (and angles for water molecules). This is done typically by applying the LINCS [63] and SETTLE [64–66] constraint algorithms for non-water and water molecules, respectively. By neglecting these fast degrees of freedom a larger timestep can be used, which improves the efficiency of MD simulations. A typical timestep for MD simulations is thus 2 fs, as using an even larger one would introduce high errors during the integration of the equations of motion.

The dihedral potential is divided into two: (1) the periodic proper dihedral potential for torsion angle ϕ, with force constant k_ϕ, multiplicity n (*i.e.*, the number of minima) and the reference value δ, and (2) the harmonic improper dihedral potential describing planar structures, with torsional angle ω, force constant k_ω and reference value ω_0. The potential of CHARMM force fields also contains a special Urey-Bradley (U-B) term, purpose of which is to restrain the bond stretching due to angle vibrations. In addition to the introduced terms, the recent CHARMM force fields also contain a CMAP correction potential. CMAP is a correction map for the backbone dihedral energies for pairs of phi/psi angles, and its purpose is to improve sampling of the backbone configurations. [67]

The two last potential terms describe the van der Waals and electrostatic interactions, respectively. The van der Waals interaction is often described using the Lennard-Jones (L-J) potential, [46] which includes (1) the short-range repulsive forces between any two atoms due to Pauli's exclusion principle ($\propto 1/r^{-12}$) and (2) the long-range attractive forces due to dispersion ($\propto 1/r^{-6}$). While the latter $1/r^{-6}$ term describing the dispersion has a quantum mechanical justification, the form $1/r^{-12}$ for the repulsive term is mainly a convenient computational choice: easy to calculate as the square of $1/r^{-6}$, yet approximating the Pauli repulsion well enough. In the equation, σ is the separation between atoms at which the potential is zero, and ϵ the well depth of the potential. The long-range electrostatic interaction between any two atoms is described using the Coulomb's law, where q_i and q_j are the charges of the atoms, ϵ_0 the vacuum permittivity and r_{ij} the distance between the atoms.

Computing the long-range forces between all atoms in the system is an extremely demanding task. Practically in an MD simulation, either cut-offs, reaction field methods [68] or Fourier-based algorithms are applied to simplify the calculations. In the context of the simulations part of this thesis, L-J potential was simply truncated and shifted to zero after a certain cut-off distance, since the attractive long-range part decays relatively fast. For the electrostatic interactions, where including the long-range part is more crucial, a common Fourier-based Particle mesh Ewald (PME) method was used. [69,70]

The force field includes both the functional form of the potentials describing the physical interactions between classical atoms, and the associated parameter

sets. A usual procedure in force fields of biomolecules is to define certain *atom types*, for which a set of parameters is derived either from experimental data or quantum chemistry calculations. [71] Each atom in the simulations system is assigned an atom type, and thus a set of parameters, that most reliably represent the properties of this particular atom with respect to the local environment. As an example, the carbon atom can be assigned a different atom type depending on the bond multiplicity. In addition to all-atom (AA) force fields, the representation can be reduced in order to increase the efficiency and simulate larger systems for longer timescales. So-called coarse-grained (CG) force fields use beads instead of atoms (*i.e* mapping four heavy atoms to one bead) and different potential forms.

It is crucial to remember that all force fields are empirical, and there is no unique method on how to translate the QM effects into parameters describing a classical system. Therefore, different force fields are often non-compatible, and should bot be mixed to describe separate molecules inside a simulation system. [71] Further limitations regarding the usage of force fields is discussed in section 3.5.

In the simulations of this thesis, the following force fields were used: all-atom CHARMM36 [72] and coarse-grained Martini [73, 74] for MD and CpHMD simulations of soluble proteins, and GAFF (Generalized Amber Force Field) [75] and Amber-compatible SLipids [76, 77] for describing a photoacid dye embedded in phospholipid membrane. Most widely used force fields for simulating biological systems commonly contain a different set of parameters for amino acids, lipids, nucleic acids, solution molecules and ions, additionally also carbohydrates, polymers, small ligands or drug molecules. If the simulation system of interest would contain molecules not defined in the force field, the molecules need to be *parameterized* for a specific force field prior to the MD simulations. In the Results section regarding publication **I**, the parameterization process of a photoacid dye molecule is shared.

3.4 Simulation systems

Setting up a simulation system for MD begins with selecting a relevant structure for the system of interest, either from experimental or available simulation data. In this thesis, the simulation systems were water-soluble proteins or phospholipid membranes. In the case of proteins, the initial atomistic structures are most commonly resolved experimentally by NMR [78], x-ray crystallography [79] or cryo-EM [80] methods. For membranes, different tools and libraries such as CHARMM-GUI [81, 82] can be utilized to setup a bilayer consisting of various types of phospholipids. A highly important feature when selecting the

components for a simulations system is that there must exist comparable force field parameters for all parts.

After selecting the initial structure for the macromolecules of interest, another important step is the choice of description for the surround solution. The most common solute in biological system is water, for which various models of different accuracy have been developed. Water can be modelled either with an implicit or explicit method. The implicit water model represents the water as a continuous medium with approximate properties, and it is computationally less demanding. In the explicit water model, which is commonly used nowadays on efficient computers, each water molecule is represented explicitly as a collection of atoms. For the all-atom simulations of this thesis, the explicit TIP-3P [62, 83] water model was selected, due to its applicability and validated performance in protein simulations. For the coarse-grained simulations, polarizable water particles were used. [74] In addition to solution, a physiologically relevant concentration of ions, specific for each simulation system, are added. Ions, such as Na^+ and Cl^-, are important for both biological processes (such as protein or membrane kinetics) and neutralization of the total charge of the system. [84–86]

In practice, the simulation system is not infinite, but has boundaries. The effect of these boundaries on the system of interest should be kept as modest as possible, commonly achieved by using periodic boundary conditions (PBC). The PBC approximate an infinitely large system by repeating the simulation box as a space-filling array. With PBC, the so-called minimum-image convention is used, meaning that for each particle only the nearest image is considered when computing the short-range non-bonded interaction terms. [53, 87] The commonly used PME method for long range electrostatic interactions is designed for periodic systems. [69,70]

3.5 Limitations

Computer simulations of many-body systems are never exact. The methods are always based on approximations and verified against experimental data or other simulation results. In order to perform trustworthy, quality simulations that provide useful results for the community, the simulations must be designed carefully for the specific scientific question.

There are numerous limitations affecting the accuracy of all computational methods, including MD. First of all, even though the force fields for MD simulations are constantly improving, they are essentially approximate. The choices of functional forms and parameters that describe the interactions in

molecules are ultimately based on choices, and in most cases highly specific for a certain type of biomolecule. [88] Secondly, in addition to force field accuracy, also the accuracy of the initial structure limits the reliability of the simulation results. If no high-resolution initial structure corresponding to the desired conditions exists, the results of MD can be highly biased. [45] The validity of the simulation model should be confirmed either by experimental results or corresponding simulation data, by comparing for example hydrogen bonding, order parameters, protein folding or membrane properties. [72, 89, 90] Thirdly, the classical level of theory prevents bond breaking or formation in standard MD simulations. Protonation states of titratable groups are chosen for the initial structure using experimental of empirical knowledge of the system, and they stay fixed throughout the simulation time. With the scope of this thesis, the limitation of no bond breaking is bypassed using a special constant pH MD method, where titratable groups can change the protonation state. Still, even in the constant pH MD approach, physical bonds are not broken, just the partial charges of the proton are interpolated to zero.

Lastly, a major limitation for simulations of larger biomolecular systems, such as proteins or membranes, is the ability to sample long enough timescales. Sampling the relevant conformational space of larger systems might take up to and over micro- or milliseconds, which might be beyond the limits even for modern supercomputers. Using statistical analysis methods to draw conclusions from a too small set of simulation trajectories is dangerous, and can considerably bias the results. Whenever possible, the results should be validated either by experimental data or comparing to simulations of similar systems, before drawing significant conclusions.

4 CONSTANT PH MOLECULAR DYNAMICS

Molecular dynamics (MD) can simulate systems in different thermodynamics ensembles in order to model experimental conditions, most often constant temperature, volume or pressure. An important, yet often ignored variable in MD simulations is the aqueous proton concentration, pH. As discussed in Chapter 2, pH together with the local environment defines the most probable protonation states of titratable groups, which in turn can affect even the dynamics and function of entire macromolecules.

Even though pH is a key variable for various biological processes, classical MD simulations were originally not designed to be performed at constant pH, in contrast to experiments. The ordinary approach to indirectly include pH in MD simulations is to set up the simulation system by selecting the most probable protonation states for each titratable group. The selection is carried out for the initial structure of the simulated molecule, and is based on experimental or computational pK_a prediction for which different methods are discussed in Chapter 2, or simply on an enlightened guess based on the pH. As the initial protonation states are kept fixed throughout the simulation, the protonation of titratable groups, such as amino acids or lipids, is not allowed to change dynamically. Therefore, pH-dependent processes have been challenging to study with classical MD simulations.

In order to include the effect of pH in MD simulations, various methods have been introduced, referred to as *constant pH molecular dynamics* (CpHMD). In this chapter, the various CpHMD methods and their main properties and differences are summarized. Then, the theory of a recent CpHMD implementation by the author of this thesis and her colleagues for GROMACS is presented in detail.

4.1 A brief history of constant pH MD

As the name of the method reveals, constant pH molecular dynamics enables simulating biomolecular systems at constant pH (in addition to constant temperature, volume or pressure). CpHMD makes it feasible to carry out simulations that enable proton transfer between titratable groups and a imaginary proton bath at a desired pH. CpHMD is designed not only to predict the pK_a values of certain titratable groups by titration, but also to study the dynamic changes of protonation states of titratable groups during a MD simulation. As the charge, and thus electrostatic interactions, depend on the protonation state of a group, a direct correlation exists between pH and sampling of the configurational space. The possibility of titratable groups to alter the protonation states during a simulation depending on the pH and local environment provides a possibility to follow pH-dependent processes with MD simulations. [18, 91]

Since the beginning of this millennium, CpHMD routines with different underlying methods and focus areas have emerged. To clarify the history and available CpHMD methods for different softwares, Table 1 lists the main CpHMD methods released for the most popular MD simulation softwares: GROMOS96/GROMACS [60], AMBER [92], CHARMM [93] and NAMD [94]. In the table, the CpHMD routines from various research groups for different MD software are categorized based on the two main features: representation of the protonation states of a titratable group (discrete or continuous) and representation of the solvent (implicit or explicit).

In the discrete CpHMD, [12, 96, 97, 100, 101, 109] the underlying idea is to use a Monte Carlo (MC) scheme between a certain number of MD steps to estimate the free energies for changing the protonation state, and make instantaneous discrete changes between the states. In contrast to the continuous methods, discrete CpHMD only samples the physically relevant protonation states. The different discrete CpHMD approaches developed over the years vary mostly in the solvent description, discussed in detail in a little while, used during the MD and MC steps, and the choice of method to estimate of the free energy of changing the protonation state for the MC step. The early implementations in GROMOS96 software from Bürgi *et al.* [12] and Baptista *et al.* [96] describe the solvent explicitly during MD, while the free energy for the MC step is estimated using more expensive thermodynamic integration (TI) for the first and computationally cheaper PB continuum electrostatics for the latter. The discrete CpHMD methods implemented in AMBER and CHARMM rely on the implicit GB description for both the MD and MC steps, [97,100,101] making them computationally efficient. The discrete methods have been shown to accurately

				protonation		solvent	
year	**corr.author**	**cite**	**software**	**disc**	**cont**	**impl**	**expl**
2001	Hünenberger	[95]	GROMOS96	empirical ($\star$)			
2002	Bürgi	[12]	GROMOS96	×		×	
2002	Baptista	[96]	GROMOS96	×		×	
2004	McCammon	[97]	AMBER	×		×	
2004	Brooks	[98]	CHARMM		×	×	
2005	Brooks	[99]	CHARMM		×	×	
2010	Roitberg	[100]	AMBER	×		×	
2011	B. Brooks	[101]	CHARMM	×		×	
2011	Shen	[102]	CHARMM		×	×	
2011	Grubmüller	[103]	GROMACS		×		×
2012	Brooks	[104]	CHARMM		×		×
2012	Shen	[105]	CHARMM		×		×
2014	Brooks	[106]	CHARMM		×		×
2016	Grubmüller	[107]	GROMACS		×		×
2016	Shen	[108]	CHARMM		×		×
2017	Roux	[109]	NAMD	×			×
2018	Shen	[110]	AMBER		×	×	
2022	Hess	[111]	GROMACS		×		×
2022	Shen	[112]	AMBER		×		×

TABLE 1 A brief summary of the main constant pH MD methods. For each method the following information is listed: year of publication, name of corresponding author, citation, software, representation of protonation states (discrete or continuous) and representation of solvent (implicit, explicit or hybrid if × is in between). Note: there have been arguments to describe the first proposed method as an empirical one [113] and therefore it will be left uncategorized.

predict pK_a values of proteins [114] and lipids [115], and to study the pH-induced conformational changes in proteins and other biomolecules. [116–118]

Continuous CpHMD methods [98, 99, 102–108, 110–112] are based on the λ-dynamics approach [119] developed by Brooks *et al.* for calculating free energies from simulations. Originating from the early work of Mertz and Pettitt, [120] the basic idea of continuous CpHMD is to assign a λ-coordinate for each titratable state, and to update its value continuously during the MD simulation along with the atomic coordinates or the system. The underlying theory behind the λ-dynamics based continuous CpHMD is described in more detail in section 4.2, where our implementation for GROMACS is introduced. As well as the discrete methods, continuous CpHMD can be combined with both implicit or explicit solvent, but especially the recent implementations of continuous CpHMD utilize the fully explicit solvent, being more accurate but also more computationally expensive. [103–106, 108, 112] Continuous CpHMD approaches

have been successfully applied to for example predict pK_a values [121–123] and to study the pH-dependent conformational changes in proteins. [124–127] In contrast to instantaneous jumps between physical protonation states in discrete CpHMD, the continuous CpHMD contains intermediate states that leave time for the local environment to relax upon the charge chance. At the same time, sampling of these unphysical intermediate states should also be kept minimal, achieved by using additional biasing potentials.

In both discrete and continuous CpHMD, the proton transfer is not explicitly modelled and covalent bonds not broken, but instead charges are interpolated to zero and the proton-solvation free energy is taken into account when updating the protonation states. In addition to CpHMD, also methods combining classical MD with explicit proton transfer have been developed. These methods utilize various approaches, such as reactive FFs or proton hopping, and they include the breakage of covalent bonds and transfer of the excess proton explicitly. [128–134] Challenges of these approaches to incorporate pH into large-scale simulations include a heavy parameterization process, computational inefficiency and sampling the incorrect thermodynamic ensemble.

As mentioned, the solvent in MD can be modelled either implicitly or explicitly. Implicit solvent models represent the solvent as a continuous low-dielectric medium instead of explicit solvent molecules, making them computationally efficient while still being able to reproduce the general behaviour of solvents in a biomolecular environment. Most of the implicit solvent models used in CpHMD simulations are based on the Generalized Born (GB) solvent model, which by approximating the Poisson-Boltzmann equation describe the electrostatics of a solvent surrounding a solute. [135–137] However, implicit solvent models are not able to describe more complex heterogeneous systems such as membrane proteins or systems in which modelling interactions with explicit water molecules are crucial.

In contrast, the more realistic explicit solvent models, where each solvent molecule is treated as separate, are both computationally more expensive but also provide the direct solvent interactions with the solute. Examples of widely used explicit water models include the SPC/E [138] and TIP3P [83, 139] (or TIP4P/TIP5P [83] with more sites) models. In addition to purely implicit or explicit, a discrete CpHMD approach can also utilize a hybrid scheme in order to improve the efficiency. In the hybrid scheme, an explicit solvent model is used during MD simulations for sampling of atomic coordinates, and an implicit solvent model for calculating the free energy of (de)protonation events during the MC steps.

In addition to the two main features of CpHMD for the representation of the protonation states and the solvent, there are also more details distinguishing

the various implementations. Firstly, in the CpHMD methods utilizing explicit solvent, it is highly important to maintain neutrality of the simulation box during the simulations. [86] As protonation states, and thus partial charges, are allowed to dynamically change during CpHMD simulations, the total charge of the simulation box can fluctuate. A non-neutral box could lead to artefacts when the electrostatic interactions are calculated using Ewald-based methods, [86] which is why the total charge should be kept neutral in CpHMD simulations that utilize PME. Charge neutrality in has been achieved either by directly coupling the titratable groups to a remote water molecule or ion, [105, 108] or by introducing titratable buffer particles that collectively take up the excess charge of the titratable groups, [107, 111] which has a few advantages. Firstly, there is no need to add position restraints to the many buffer particles as they only carry a small charge, and strong interactions or binding to the system of interest are avoided. Secondly, introducing many buffers with small charges also results in smaller disruption of the hydrogen bond network of the water molecules. For the CpHMD methods utilizing only implicit solvent description, the change of charges during (de)protonation events has no effect by nature.

Secondly, the continuous CpHMD methods can be divided based on the possibility to include chemically coupled titratable sites within one titratable group. In these groups, the force field parameters of one site are dependent on the protonation of the other sites in the same group. A common example of a multisite group is histidine, an amino acid with two titratable sites defining three relevant protonation states. Such groups have been represented in few continuous CpHMD methods by directly coupling the potential energy functions of these titratable sites. [99, 103] Another, computationally more efficient, approach is to use a so-called *multisite* approach, [140] where separate λ-coordinates are introduced to describe each protonation state, and the sampling of these coordinates is restricted to a plane connecting the physical states, using a constraint function. [101, 104, 106, 111]

Lastly, also relevant for the continuous CpHMD methods, the choice of interpolation when calculating the electrostatic interactions of the λ-coordinates affects the computational efficiency. In general, the continuous CpHMD methods utilize linear interpolation of either potential energy functions [96,97,101,103,110, 141] or partial charges [98,99,108,111,112] to compute these interactions of the λ-coordinates. Interpolation of potential energy functions results in additional force calculations during the simulation, introducing a linear scaling of simulation speed as more titratable groups are added. [111] Described in detail in the next section, charge interpolation linearly interpolates between the partial charges of both states, in principle resulting in no dependence on the number of titratable groups but a comparable performance to standard MD. In the latest CpHMD implementations with CI and efficient algorithms to calculate the electrostatic

interactions, the superior computational efficiency enables the study of larger and more complex pH-dependent systems and phenomena.

In CpHMD, the conformational dynamics and sampling of protonation states are coupled, which can lead to slow convergence, especially related to dihedral degrees of freedom in proteins [142, 143] and when obtaining accurate pK_a values. Slow convergence can prevent studying of larger systems, such as pH-dependent protein complexes. In order to overcome this sampling challenge, the proposed solution has been to modify the torsional barriers of amino acids side chains. [98, 143, 144] Another attempt to enhance sampling of protonation states and conformatios has been the introduction of so-called pH replica-exchange (pHREX), [100, 102, 105, 106, 108, 110] which combines CpHMD with pH replica-exchange. In pHREX, independent replicas at different pH are run simultaneously, and with a given frequency the configurations at different pH runs are exchanged based on a Metropolis criteria. In order to access longer simulation times and to increase sampling of both protonation and configurational spaces, various CpHMD methods have also been recently implemented to run on GPUs. [111, 112, 145, 146]

The methodological choices for the various CpHMD implementations together with the selected software affect accuracy, generality and efficiency of method. It can of course be debated which choices lead to most accurate results and should be taken into account when implementing a new CpHMD method. As mentioned, both the aforementioned discrete and continuous CpHMD have their pros and cons, depending on the simulation system of interest. However, the continuous CpHMD with explicit water description has become more popular in the recent years, due to its generality, efficiency and availability in various softwares. When using the continuous CpHMD, explicit water and multisite description should be default as the modern computers offer enough resources to utilize these accurate features. Choosing a method that utilizes charge interpolation has a better performance, which is highly important for systems with a high number of titratable groups. A charge levelling scheme should be default if PME is used for calculating electrostatic interactions, simply because it has been shown that PME does not perform accurately with non-neutral simulation systems. The scheme introduced above with multiple buffer particles is preferred, as the excess charge smeared to many small-charge particles causes less artifacts to the simulation system. In addition, the author of this thesis argues that all of the chosen properties should be well described in the corresponding publication, along with access to the actual code (which has not been the obvious case for all published methods).

After the brief review of the recent CpHMD implementations, the next section will focus on the theory behind the latest CpHMD developed for GROMACS [111] by the author of this thesis and her colleagues. This implementation

includes continuous representation of protonation states with fully explicit solvent, is capable to represent multisite titratable groups with no extra cost, has a charge levelling scheme, and utilizes charge interpolation when calculating the underlying λ-dependent potential energies. These features combined, and implemented in the free and popular MD software GROMACS, make the CpHMD accurate, *fast, flexible and free*.

4.2 Constant pH MD for GROMACS

GROMACS [60] is the most popular open-source software for performing atomistic MD simulations of biological systems. The large user community of the software appreciates and also motivates the implementation of special simulation methods, such as CpHMD, that provide insights into systems beyond conventional MD. There has been a request for years to update and reshape the previous GROMACS CpHMD method [103] to be in line with the other software packages for biomolecular simulations, such as CHARMM [93] and AMBER [92]. The author of this thesis, together with colleagues, accepted the request and implemented an efficient CpHMD method for the most recent version of GROMACS. The implementation is discussed in publication **II** [111] and, in addition, the accuracy and sampling of our CpHMD is discussed in the accompanying publication **III**. [143]

In this section, the underlying theory of the improved CpHMD method for GROMACS will be presented, including discussion of the following: description of the protonation with λ-dynamics and pH-dependent potentials, superior computational efficiency of the method, description of multisite titratable groups, and conservation of neutrality in the total charge of the system upon protonation state changes during CpHMD simulations.

4.2.1 λ-dynamics

Introduced by Brooks *et al.* by over two decades ago, λ-dynamics [119] has become the basis of most continuous CpHMD methods. The key idea of CpHMD methods based on λ-dynamics is that the protonation state of a titratable group is described by a continuous λ-coordinate, which can dynamically change during the simulations.

A titratable site i can be protonated or deprotonated, depending on the interactions with the local environment. In λ-dynamics based CpHMD, a fictious

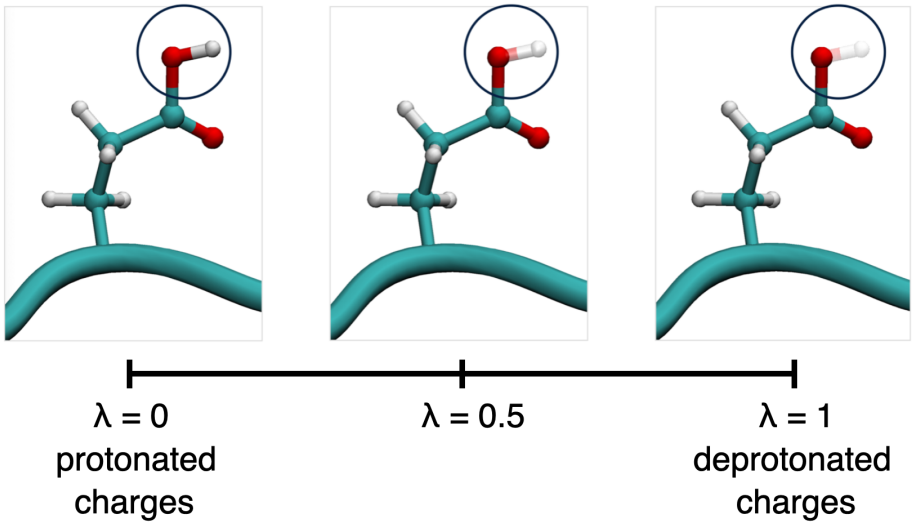


FIGURE 5 In the implemented CpHMD method for GROMACS, protonation states of titratable groups are described using the λ-dynamics. Here, the protonated and deprotonated states of aspartic acid with respect to the λ-coordinate are shown. $\lambda = 0$ and $\lambda = 1$ correspond to the protonated and deprotonated states of Asp, respectively. During the CpHMD simulation, the values of λ-coordinates of all titratable groups are changing dynamically, and updated along with the Cartesian coordinates during the MD algorithm.

coordinate λ_i with mass m_i is introduced for each titratable site i in the system. These additional λ-coordinates continuously interpolate between the protonated and deprotonated states of this site, and they are updated along the spacial coordinates of the system using Newton's equations of motion. Figure 5 clarifies the λ-dynamics: for $\lambda = 0$ the titratable site, here the carboxyl group of aspartic acid sidechain, is protonated, and towards $\lambda = 1$ the group deprotonates. To emphasize the description of (de)protonation events in CpHMD simulations, the hydrogen atom is never actually released by breaking the bond but only its partial charge is interpolated to zero.

In many previous continuous CpHMD implementations in different software packages, the interpolation was carried out between the potential energy functions of the protonation states. [102, 103, 106] While being relative easy to implement into the MD code, potential energy interpolation introduces a critical reduction in the simulation efficiency with respect to increasing the number of titratable groups. [111] Since the choice of interpolation in the Hamiltonian does not affect the outcome as long as the potential energy of the protonated and deprotonated states remains the same, interpolating the partial charges of the atoms part of titratable groups is another possibility, [98,108,111,112] also utilized in our implementation.

The total Hamiltonian for a system with titratable groups i is

$$H(\mathbf{R},\boldsymbol{\lambda}) = \sum_{j}^{N_{\text{atoms}}} \frac{m_j}{2}\dot{\mathbf{r}}_j^2 + \sum_{i}^{N_{\text{sites}}} \frac{m_\lambda}{2}\dot{\lambda}_i^2 + V(\mathbf{R},\boldsymbol{\lambda}) \tag{12}$$

where $\mathbf{R}$ includes the Cartesian coordinates $\mathbf{r}_j$ of all N_{atoms} atoms with mass m_j, and $\boldsymbol{\lambda}$ includes the λ_i coordinates of all N_{sites} titratable sites. Kinetic energies of atoms j are accompanied by kinetic energies of the additional fictitious λ-particles λ_i with mass m_λ. In addition, the potential energy of the system now also depends on $\boldsymbol{\lambda}$, as the partial charges of atoms affect the electrostatic interactions, and those are λ-dependent.

The total potential energy of the system $V(\mathbf{R},\boldsymbol{\lambda})$ can be written as

$$V(\mathbf{R},\boldsymbol{\lambda}) = V^{\text{R}}(\mathbf{R}) + V^{\text{coul}}(\mathbf{R},\boldsymbol{\lambda}) + V^{\text{MM}}(\boldsymbol{\lambda}) + V^{\text{bias}}(\boldsymbol{\lambda}) + V^{\text{pH}}(\boldsymbol{\lambda}) \tag{13}$$

where $V^{\text{R}}(\mathbf{R})$ contains the potential energy terms *not* dependent on λ (bonded interactions, L-J interactions), and $V^{\text{coul}}(\mathbf{R},\boldsymbol{\lambda})$ the potential energy terms that are λ-dependent (Coulomb interactions). Three potential terms that only depend on λ are added: $V^{\text{MM}}(\boldsymbol{\lambda})$, $V^{\text{bias}}(\boldsymbol{\lambda})$ and $V^{\text{pH}}(\boldsymbol{\lambda})$, which are the correction, biasing, and pH potentials, respectively. To clarify, the bonded and L-J interactions are assumed not to depend on λ-coordinates but to be constant upon the protonation state changes, as their contribution is small. [111] Thus, atom types of titratable groups' atoms are not altered during the CpHMD simulations.

The term $V^{\text{coul}}(\mathbf{R},\boldsymbol{\lambda})$ contains the electrostatic Coulomb interactions, which by depending on the protonation state of the titratable groups naturally depend on the λ-coordinates. Using the interpolation of partial charges for the λ-dynamics to describe titratable groups, the pair-wise Coulomb potential can be written as

$$\begin{aligned}
V^{\text{coul}}(\mathbf{R},\boldsymbol{\lambda}) &= V^{\text{rest-rest}}(\mathbf{R}) + V^{\boldsymbol{\lambda}\text{-rest}}(\mathbf{R},\boldsymbol{\lambda}) + V^{\boldsymbol{\lambda}\text{-}\boldsymbol{\lambda}}(\mathbf{R},\boldsymbol{\lambda}) + V^{\boldsymbol{\lambda}}(\mathbf{R},\boldsymbol{\lambda}) \\
&= \tfrac{1}{2}\textstyle\sum_i^{n_{\text{rest}}}\sum_j^{n_{\text{rest}}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_k^{N_{\text{sites}}}\sum_i^{n_k}\sum_j^{n_{\text{rest}}} \frac{\left((1-\lambda_k)q_i^A + \lambda_k q_i^B\right)q_j}{4\pi\epsilon_0 r_{ij}} \\
&\quad + \sum_k^{N_{\text{sites}}}\sum_{m;m\neq k}^{N_{\text{sites}}}\sum_i^{n_k}\sum_j^{n_m} \frac{\left[(1-\lambda_k)q_i^A + \lambda_k q_i^B\right]\left[(1-\lambda_m)q_j^A + \lambda_m q_j^B\right]}{4\pi\epsilon_0 r_{ij}} \\
&\quad + \tfrac{1}{2}\sum_k^{N_{\text{sites}}}\sum_i^{n_k}\sum_j^{n_k} \frac{\left[(1-\lambda_k)q_i^A + \lambda_k q_i^B\right]\left[(1-\lambda_k)q_j^A + \lambda_k q_j^B\right]}{4\pi\epsilon_0 r_{ij}}
\end{aligned} \tag{14}$$

where the four potentials are divided as follows:

1. $V^{\text{rest-rest}}(\mathbf{R})$ contains interactions between atoms that are not part of any titratable λ-group, and do not depend on λ-coordinates. The sum runs over all n_{rest} atoms not part of any titratable group.
2. $V^{\lambda\text{-rest}}(\mathbf{R}, \boldsymbol{\lambda})$ contains interactions between atoms $i = 1...n_k$ of each titratable λ-group $k = 1...N_{\text{sites}}$ with atoms $j = 1...n_{\text{rest}}$ that are not part of any λ-group. For atoms part of titratable groups, the partial charges are interpolated with λ_k.
3. $V^{\lambda\text{-}\lambda}(\mathbf{R}, \boldsymbol{\lambda})$ contains interactions between atoms belonging to two different titratable λ-groups, with interpolated partial charges.
4. $V^{\lambda}(\mathbf{R}, \boldsymbol{\lambda})$ contains interactions between atoms within each titratable group, with interpolated partial charges.

In the above, q^A and q^B are the partial charges of atoms part of a titratable group, where A denotes the protonated and B deprotonated states. In order to update the values of the λ-coordinates along with spacial coordinates, the gradient of the potential $V^{\text{coul}}(\mathbf{R}, \boldsymbol{\lambda})$ with respect to each λ-coordinate must be computed. The gradient with respect to λ_k can be written as [111]

$$\frac{\partial V_{\text{coul}}(\mathbf{R},\boldsymbol{\lambda})}{\partial \lambda_k} = \sum_i^{n_k} \Phi(\mathbf{R_i}, \boldsymbol{\lambda}) \Delta q_i \tag{15}$$

where the electrostatic potential $\Phi(\mathbf{R_i}, \boldsymbol{\lambda})$ at the position of atom i is used to calculate the gradient. The electrostatic potential is caused by all other atoms in the system, including the atoms of all titratable sites, whose the partial charges are interpolated:

$$\Phi(\mathbf{R_i}, \boldsymbol{\lambda}) = \sum_j^{n_{\text{rest}}} \frac{q_j}{4\pi\epsilon_0 r_{ij}} + \sum_m^{N_{\text{sites}}} \sum_j^{n_m} \frac{(1-\lambda_m) q_j^A + \lambda_m q_j^B}{4\pi\epsilon_0 r_{ij}}$$

and $\Delta q_i = q_i^B - q_i^A$ is the difference between the partial charges of atoms part of titratable group i in the protonated (A) and deprotonated (B) states.

In GROMACS, electrostatic Coulomb interactions are preferably calculated using the smooth Particle Mesh Ewald method. [69, 70] Extracting the electrostatic potential, needed to compute the gradient of the λ-dependent potential, along with the standard PME calculation adds only a small computational overhead. Therefore, using the interpolation of charges for the propagation of λ-coordinates provides a CpHMD method comparable with the computational effiency of a standard simulation. This is a major improvement compared to the previous CpHMD method developed for GROMACS using the potential energy interpolation, [103] where 2N additional PME calculations were needed for N titratable groups.

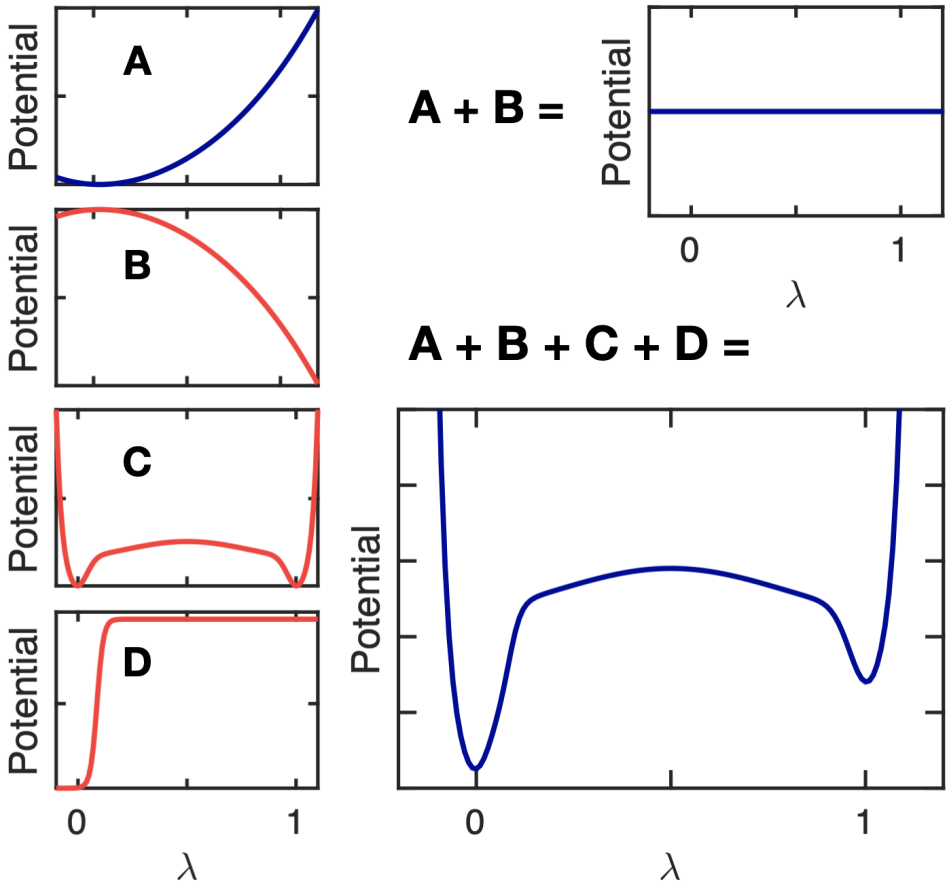


FIGURE 6 λ-dependent potential energy terms in CpHMD. (A) Free energy of deprotonation of a titratable group at force field level, $\Delta G_i^{\mathrm{MM}}(\lambda)$. (B) To compensate for $\Delta G_i^{\mathrm{MM}}(\lambda)$, a correction potential $V_i^{\mathrm{MM}}(\lambda)$ is added, such that the free energy difference between the two states is zero (A+B). (C) A biasing potential, $V^{\mathrm{bias}}(\lambda)$ is added to prevent the sampling of non-physical states. (D) The proton chemical potential (pH) is modelled by $V^{\mathrm{pH}}(\lambda_i)$. The total contribution of λ-dependent potential terms is shown on the A+B+C+D panel. Adapted with permission from Noora Aho, Pavel Buslaev, Anton Jansen, Paul Bauer, Gerrit Groenhof and Berk Hess. Scalable Constant pH Molecular Dynamics in GROMACS. *Journal of Chemical Theory and Computation*, 18 (10), pp. 6148-6160, **2022**. Copyright © 2022 American Chemical Society.

Equation 13 contains three more λ-dependent potential terms: a correction potential $V^{\mathrm{MM}}(\boldsymbol{\lambda})$, a biasing potential $V^{\mathrm{bias}}(\boldsymbol{\lambda})$ and a pH-dependent potential $V^{\mathrm{pH}}(\boldsymbol{\lambda})$. All of these potentials are evaluated separately for each titratable group λ_i. Forms of these potentials are presented in Figure 6, and discussed next.

For a titratable group in water, there is a non-zero free energy associated with changing the protonation state, making one of the states more favourable. In order to compensate for this free energy, originating from the missing quantum mechanical contributions of breaking the chemical bond and solvating the proton, a correction potential $V^{\mathrm{MM}}(\mathbf{R}, \boldsymbol{\lambda})$ is added (Figure 6B) for each λ-coordinate.

Addition of the correction potential makes the interpolated potential flat at pH = pK_a^{ref} (Figure 6A+B), ensuring equal sampling of the λ-coordinate across its interval $V^{MM}(\mathbf{R}, \boldsymbol{\lambda})$ is calculated prior to the actual CpHMD simulations, and it corresponds to the free energy required to deprotonate the reference group in water:

$$V_i^{MM}(\lambda_i) = -\Delta G_i^{MM}(\lambda_i) \tag{16}$$

Secondly, in order to reduce the sampling of nonphysical states in between the physical end states $\lambda = 0$ and $\lambda = 1$, the biasing potential $V^{bias}(\boldsymbol{\lambda})$ is added. The biasing potential has a double well shape with an adjustable height (Figure 6C), that creates two local minima around the physical end states, while still allowing transitions between them and preventing λ-values far outside the range of $0 < \lambda < 1$. The detailed form of $V^{bias}(\boldsymbol{\lambda})$ can be found in the original publication **II**. [111]

Lastly, the pH-dependent potential $V^{pH}(\boldsymbol{\lambda})$ is added for each λ-coordinate, in order to include the effect of the solution pH. The pH-dependent potential adds a free energy difference $\Delta V^{pH} = RT \ln(10) [pK_{a,i} - pH]$ between the protonation states. The difference depends on the difference between the desired pH and the reference pK_a of the reference group in water. The pH potential is plotted in Figure 6D, and it has a smooth step-wise potential form, also found in detail in the original publication **II**. [111] To summarize, the total λ-dependent potential, containing the three described potentials added to the interpolated potential for a titratable group, is illustrated in Figure 6 lower-right panel A+B+C+D.

4.2.2 Multisite titratable groups

Multisite titratable groups contain two or more chemically coupled titratable sites. This means, that changing the force field parameters of one site, which in our case are only the partial charges as discussed above, directly affects the parameters of another site, and *vice versa*. Multisite groups can be described by an extended version of the method developed by Brooks *et al.* [140]: independent λ-coordinates are added to represent each protonation state of the titratable group, and these coordinates are constrained onto a plane connecting the physical states.

In our implementation, each protonation state of a titratable group is represented using a one-dimensional λ_i^k, where i denotes the index of the multisite titratable group, and k the protonation state within this group. Value $\lambda_i^k = 1$ corresponds to the physical protonation state, and $\lambda_i^k = 0$ is a non-physical state common to all λ_i^k describing the same multisite group. In order to sample the

relevant protonation states of a multisite titratable group, the λ_i^k-coordinates are propagated as independent two-state λ-coordinates, but in addition constrained onto the (hyper)-plane connecting the protonation states, namely constraining the sum $\sum_k^i \lambda_k^i = 1$. This linear constraint is solved analytically (together with the constraint for charge neutrality, introduced below), by using a constraint algorithm based on minimizing the Lagrangian multipliers. [107,111] In general, this constraint can be applied to any number of titratable states. Details on the constraint can be found in the original publication.

Common example of a chemically coupled multisite group is histidine (His), which contains two titratable sites, namely the N_δ and N_ϵ nitrogens, and can exist in three protonation states. Figure 7 shows the histidine side chain along with definitions of λ-coordinates for the multisite representation. The three protonation states of histidine are doubly protonated HSP (proton bound to both nitrogens, described with λ_1), HSD (proton bound to N_δ, described with λ_2) and HSE (proton bound to N_ϵ, described with λ_3). The $\lambda = 0$ state is common for all

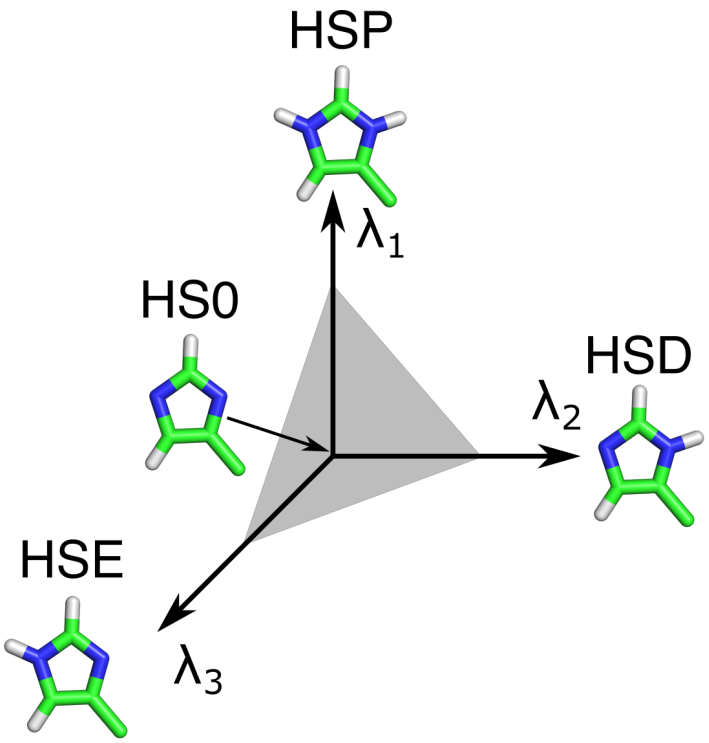


FIGURE 7 Multisite representation applied for histidine. The three relevant protonation states are described by λ_1 (doubly protonated HSP), λ_2 (HSD) and λ_3 (HSE). The $\lambda = 0$ state HS0 is common for all λ-coordinates. The sampling is restricted to the plane $\lambda_1 + \lambda_2 + \lambda_3 = 1$, and a biasing potential is added for all λ-coordinates in order to avoid non-physical states. Adapted with permission from Noora Aho, Pavel Buslaev, Anton Jansen, Paul Bauer, Gerrit Groenhof and Berk Hess. Scalable Constant pH Molecular Dynamics in GROMACS. *Journal of Chemical Theory and Computation*, 18 (10), pp. 6148-6160, **2022**. Copyright © 2022 American Chemical Society.

the λ-coordinates describing a multisite group, and does correspond to an actual physical state (but for example a set of zero partial charges).

4.2.3 Charge neutrality

As mentioned in the previous section, the titratable sites of the system are allowed to (de)protonate during a CpHMD simulation, which results in fluctuations of the total charge of the simulation box. This is different from conventional MD simulations, where the total charge stays constant with fixed protonation states. Nonzero total charge should be compensated to avoid artefacts for the electrostatic interaction calculations using PME. [86]

In our GROMACS CpHMD implementation, the total charge is kept constant during CpHMD by a scheme in which a number of buffer particles are added to the system to uptake the excess charge, and to keep the simulation box neutral. [107, 111] One λ-coordinate describes the charges of all buffers simultaneously, and the charges of each buffer are the same at each step. The number of buffer particles is adjusted based on the number of titratable groups in the system of interest, and a constraint keeps the total charge constant: $N^{\mathrm{BUF}}q^{\mathrm{BUF}} + q^{\mathrm{SYS}} = \mathrm{const}$, where N^{BUF} is the number of buffers, q^{BUF} the charge on an individual buffer and q^{SYS} the charge of the rest of the system. In practice, the constraint for charge neutrality is combined with the constraint for multisite titratable groups. Both constraints, if applied, are solved simultaneously at each simulation step after updating the λ-coordinates, using an algorithm based on minimizing the Lagrangian multipliers. [107, 111] In practice, the buffers are particles with small maximal charges $|0.5|e$ and small Lennard-Jones radii. With these parameters strong interactions with the rest of the system and clustering with other buffer particles are prevented, as discussed in more detail in the Results chapter. [143]

4.2.4 Applicability and limitations

Currently, the developments of constant pH MD provide a wide range of methods for various simulation softwares. As the properties and performances vary a lot between methods (Table 1), it is of high importance for a novice user to select a method and software suitable for their problem. The introduced CpHMD implementation for GROMACS in publication **II** is open access, accurate and fast in performance, which makes it an attractive tool to include pH induced protonation dynamics into MD simulations, especially for users with previous experience of GROMACS.

For a researcher familiar with MD but novice in CpHMD simulations, it should

be highly emphasized that despite its applicability to various systems of interest, CpHMD is not a magic bullet to include protonation state dynamics for each possible system. First of all, the initial structure, force field parameters and partial atomic charges must be known for the titratable molecule one plans to treat with CpHMD. Especially for small ligands, lipid headgroups or other more special compounds, these might be previously unknown, meaning that they must be derived from first principles before CpHMD can be applied.

As with standard MD, the often most crucial limitation in a simulation is the sampling. For CpHMD, in addition to sampling the configurational space, also the protonation states described with λ-coordinates must be sufficiently sampled within the simulation time. As an example, the barriers of torsions of aliphatic side chains have been found to be too high to converge on microsecond timescales using CpHMD, preventing the λ-degrees of freedom to reach equilibrium. Therefore, as discussed in publication **III**, modifications to the CHARMM36 force field are suggested when using CpHMD. In order to reliably calculate correct time averages from the simulations, there must be enough transitions between the relevant protonation states of all titratable groups in the system. [97] For example, if many titratable groups interact strongly with the local electrostatic environment, the convergence for protonation states can be extremely slow. If even the sampling of a single titratable amino acid does not reach convergence within tens of nanoseconds, [18, 143], reaching convergence for pK_a values for tens of residues in a protein requires multiple times longer timescales. Therefore, estimating the errors and investigating the sampling of both configurations and protonation is of high importance when applying the method for a new system of interest.

5 RESULTS

For this thesis, molecular dynamics simulations were both performed and further developed, in order to study proton transfer and the effect of pH on various biological systems. First, conventional MD simulations were carried out to complement experimental findings of proton diffusion on the surface of phospholipid membranes (Chapter 5.1). As the main work, accurate and efficient constant pH MD routine was implemented into the GROMACS simulation software program (Chapter 5.2). Lastly, best practices for CpHMD simulations are provided, including corrections to the CHARMM36 force field in order to more reliably sample the rotamers of titratable amino acids (Chapter 5.3). Together, these results demonstrate the benefits of an accurate and efficient constant pH MD method when studying pH-dependent processes in biology.

5.1 Proton diffusion on the surface of phospholipid membranes (publication I)

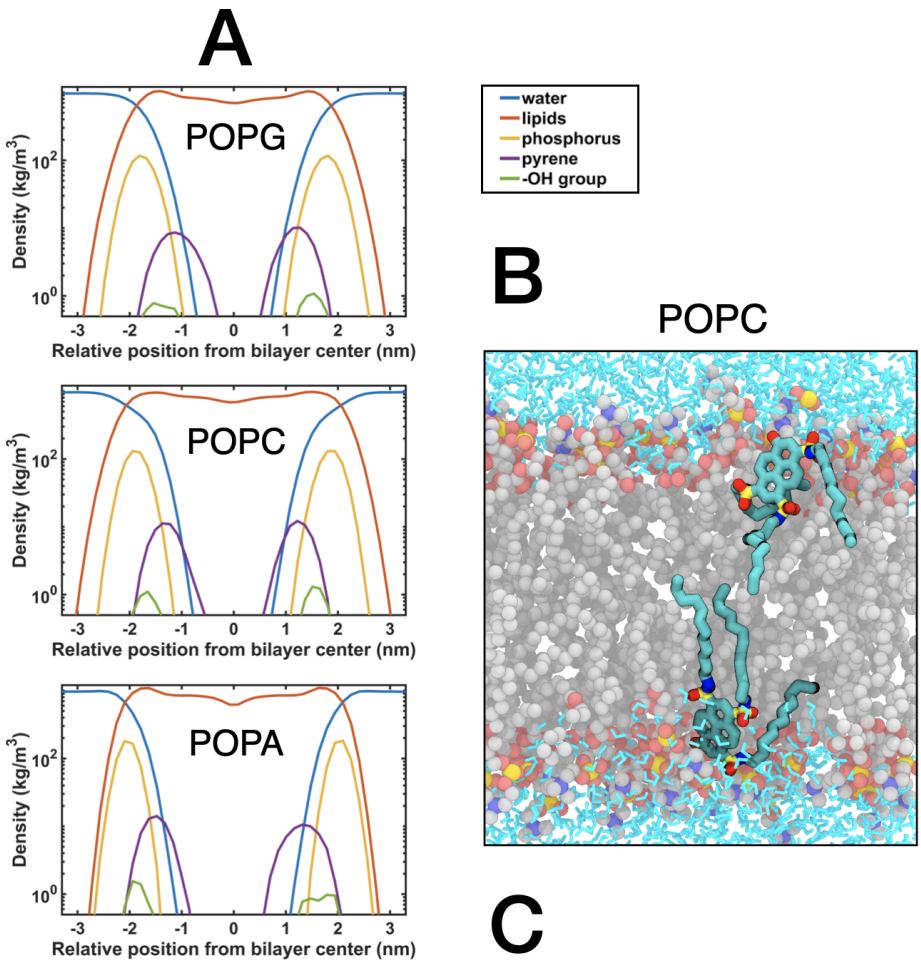
As a collaboration with an experimental research group, we carried out MD simulations of the C12-HPTS photoacid (structure shown in Figure 4) embedded in phospholipid bilayers with different headgroups: a negative headgroup (POPG), a zwitterionic headgroup (POPC) and a headgroup composed only of the negative phosphate group (POPA). [40] The goal of these simulations was to supplement the time-resolved fluorescence experiments of proton transfer on the surface of membranes, by providing atomistic level details on the location of the photoacid in the bilayer, and the interactions between the photoacid with its surroundings.

First, before performing the MD simulations, we parametrized the C12-HPTS photoacid used in the experiments for MD simulations with the GROMACS software. We decided on using the GAFF (Generalized Amber Force Field) [75] to describe the interactions of the photoacid, due to the straightforward parameterization process. For the parameterization, atomic structure for the pyrene headgroup of the photoacid was created using CHARMM-GUI [81, 141]. The electrostatic potential of the headgroup was calculated using the Gaussian16 package [147] with the B3LYP exchange correlation functional and 6-31G* basis set. [148–151] From the electrostatic potential, partial charges for the probe head group atoms were derived from a restricted electrostatic potential (RESP) fit with the Merz-Singh- Kollman scheme, [152] using AnteChamber [153] and acpype. [154] Atomtypes and parameters for L-J and bonded interactions for the photoacid head were also derived using AnteChamber. For the hydrocarbon tails of the probe, atom types, parameters and charges were assigned based on the the saturated POPC lipid tail of the SLipids force field. [77, 155]. In the MD simulations, POPC and POPG were modelled with the SLipids force field, whereas parameters for POPA had to be parametrized by us with the SLipids guidelines [76].

For each of the phospholipid types a bilayer with 122 lipids was created. Two photoacids were embedded to each bilayer, one to the upper and one to the lower leaflet. We performed microsecond MD simulations of each photoacid-membrane-water system. An example snapshot of the photoacid in the POPC membrane is shown in Figure 8B.

From the simulation trajectories, we first investigated the location and orientation of the photoacid in the membranes. The goal was to verify the experimental hypothesis that the hydroxyl group of the photoacid is exposed towards the surface of the membrane. Density distributions for all three systems are shown in Figure 8A where the density of water, lipids, the phosphorus atoms of the lipids, pyrene head of the photoacid, and the hydroxyl group of the photoacid are separately plotted. From the densities of the hydroxyl groups relative to the densities of the phosphorus atoms of the lipids, we can conclude that the proton-releasing -OH group is located at the membrane-water interface.

The second objective of the MD simulations was to analyze the interactions between the hydroxyl group of the photoacid with its surroundings, in order to estimate the possibility of the photoacid to release a proton to the nearby water or lipid molecules. From the time-resolved fluorescence measurements performed by our experimental colleagues, the kinetic rate constants k_{PT} and k_a could be obtained, and those are listed in Figure 8C.



Lipid system	hydrogen bonds with water molecules, %	hydrogen bonds with lipid phosphates, %	total hydrogen bonds with everything, %	k_{PT}, ns^{-1}	k_a, ns^{-1}
POPG	31.9	65.9	97.8	2.6	18.5
POPC	27.1	71.3	98.4	2.8	17.0
POPA	37.4	56.2	93.6	5.5	11.2

FIGURE 8 Main simulation results from publication **I**. (A) Density distributions along the bilayer normal for water (blue), phospholipids (red), phosphorus atoms of the lipid headgroups (yellow), pyrene moieties of the photoacid (violet), and hydroxyl groups of the photoacid (green). (B) An example snapshot from an MD trajectory of the POPC system, showing the position of the photoacid in the upper and lower leaflets. (C) Percentage of time (averaged between the upper and lower leaflets) that a hydrogen bond is formed between the -OH group of the photoacid and the nearby waters or lipid phosphates in the MD simulations, and proton transfer rate constants k_{PT} and k_a obtained from the experiments, to compare simulation results to. Adapted with permission from Nadav Amdursky, Yiyang Lin, Noora Aho and Gerrit Groenhof. Exploring fast proton transfer events associated with lateral proton diffusion on the surface of membranes. *Proceedings of the National Academy of Sciences*, 116 (7), pp. 2443-2451, **2019**. Copyright © 2019 National Academy of Sciences.

Introduced in the ESPT scheme in Figure 4, k_{PT} and k_a are the rate coefficients for proton release as the excited state ROH^* deprotonates to RO^{-*} and the recombination to reform ROH^* from RO^{-*}, respectively. We focus on these parameters, as their relative order with respect to the three lipid types gives us insights into the diffusion of protons on the surface of the membranes, and can be correlated with the simulation results. The higher k_{PT} and lower k_a obtained for POPA suggest that protons released on the surface of the membrane are more probable to escape from the membrane to the bulk water compared to POPC and POPG.

The simulation trajectories were exploited to inspect the interactions of the photoacid with the surroundings. As the proton is released from the hydroxyl group of the photoacid located at the interface between the bilayer and water, we calculated the number of hydrogen bonds from the -OH to the nearby water molecules and lipid headgroups, correlating these results with the experimental rate constants. The percentages of time a hydrogen bond is formed are given in the table in Figure 8D. From these values, the ratio is obtained between the time a hydrogen bond is formed between the surrounding water molecules and the lipid headgroups. The simulations show the same behaviour as seen in the experiments: POPA, probably due to its small negative headgroup, forms more hydrogen bonds with the bulk water molecules corresponding to the higher k_a of the experiments, therefore suggesting that proton transfer from the membrane surface to the bulk is more likely to happen.

Overall, the simulations part of the publication was an example of how proton transfer properties of biomolecules can be investigated with classical MD, without a possibility to directly detach the proton from the molecules. Providing necessary details of the experimentally studied system by MD complemented the fluorescence measurements, and both approaches together provided new insights to the picture of the proton transfer mechanisms on the surface of phospholipid membranes.

5.2 Fast, flexible, and free constant pH MD for GROMACS (publication II)

Within the scope of this thesis, publication **II** can be considered as the main research work performed by the author (naturally together with colleagues and the GROMACS development team). In brief, the heart and soul of the publication was to implement an efficient and accurate constant pH MD method in the popular GROMACS software, with all the key properties listed in Section 4.1.

Even though CpHMD has been implemented in GROMACS about a decade ago, the previous version [103] was outdated and inefficient, and not applicable for simulating large biomolecular systems at long timescales.

As GROMACS is among the most popular open-source MD software designed for simulating biological systems, the inclusion of an efficient CpHMD method has been desired by the user community since the publication of the previous implementation. Many pH-dependent processes in biology contain larger entities, such as proteins or membranes, which require relatively large amounts of computational power. Therefore, the ideal CpHMD routine (proudly presented by us) should have a performance comparable to the conventional MD simulations and, in addition, include the treatment of titratable groups by enabling the changing of protonation states during the simulation.

When implementing CpHMD, our aim was to combine and extend existing features of the previous CpHMD methods available for different MD software: GROMACS, AMBER and CHARMM. Our version of CpHMD includes the usage of charge interpolation, working with both commonly used PME and reaction field methods to calculate the electrostatic interactions. Combined with the continuous representation of protonation states by utilizing λ-dynamics, the features together result in a comparable computational efficiency to standard MD. The solvent is represented explicitly, there exists a possibility to include multisite titratable groups with more than two protonation states, and the total charge of the simulation system can remain neutral upon protonation events by a charge levelling scheme. Together, these properties make the efficiency of our CpHMD implementation comparable to standard MD, which enables longer simulations of larger systems, and in general applicable to various pH-dependent systems regardless the structure or force field.

The schematic representation of our CpHMD implementation is presented in Figure 9. The dark blue boxes contain the core loop of MD simulations, which consists of force calculation from the potential, solving Newton's equations of motion according to Equation 7, and maintaining the NPT ensemble with thermo- and barostats, after which the simulation is continued until the desired amount of sampling is reached. The red boxes illustrate the preparation and analysis steps carried out before and after the MD loop. When running MD at a constant pH, the topology, that contains all information about the interaction potential of the system, must be known for all protonation states included in the CpHMD simulation. In addition, the reference pK_a-values and the parameters of the correction potential $V^{\mathrm{MM}}(\lambda)$ (both discussed in 4.2.1) must be available for all titratable groups.

In CpHMD, along with atomic positions, the λ-coordinates describing the protonation states of titratable groups are updated. The forces acting on λ-

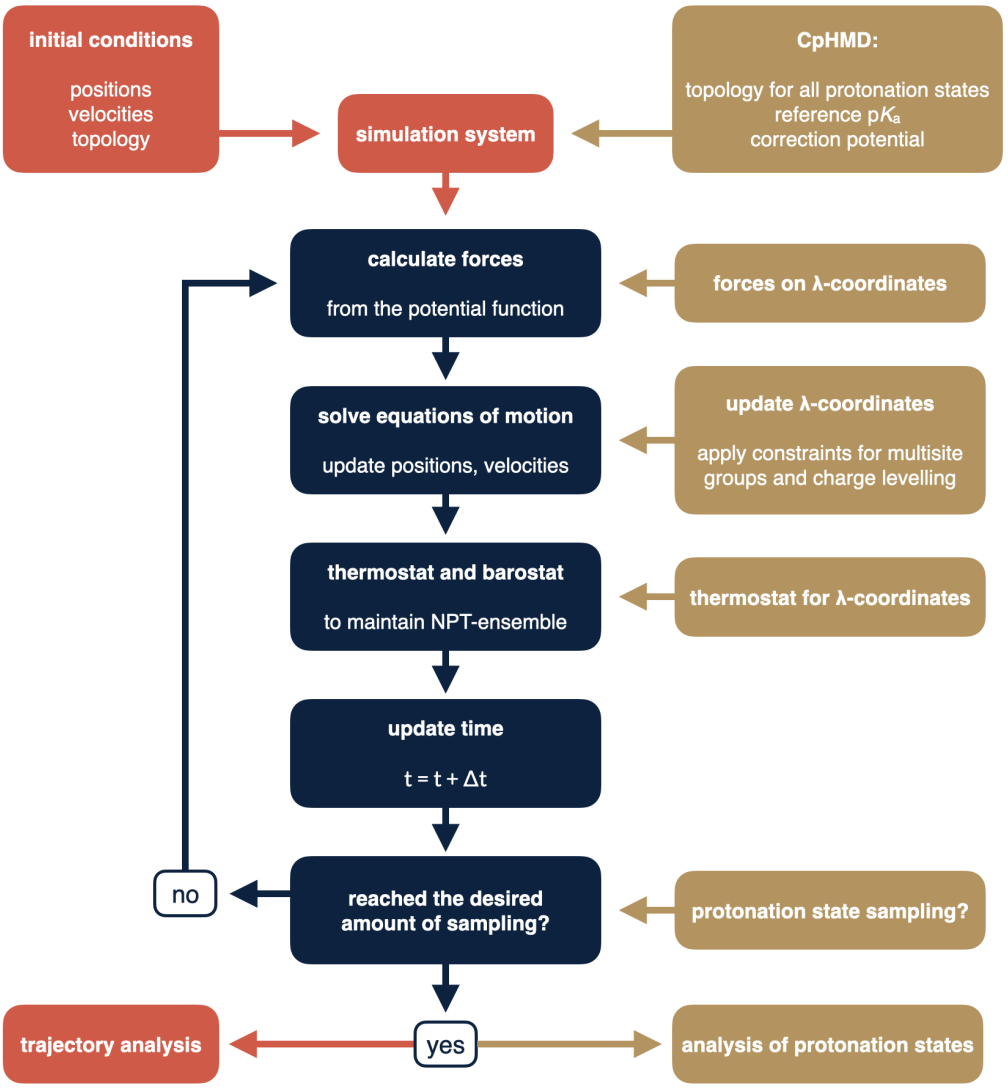


FIGURE 9 Schematic representation of MD simulations, and the additional steps performed in a CpHMD simulation (yellow boxes). The MD core loop is presented in the blue boxes, and the red boxes annotate steps before and after the simulation.

coordinates are calculated, as presented in 4.2.1, after which the values are propagated using Newton's equations of motion. To represent the multisite titratable groups, and to maintain the charge neutrality of the simulation box, a linear constraint is applied for the λ-coordinates (as described in 4.2.2 and 4.2.3). The temperature of these fictitious λ-particles is also kept constant by applying a separate thermostat.

After the CpHMD simulation has reached the desired amount of sampling, for both spatial and λ-coordinates, the conventional analysis of the MD trajectory can be extended to analyzing the trajectory of the λ-coordinates, representing the protonation states. For the specified pH, information about the coupled dynamics of positions and protonation can be extracted. These dynamics together can reveal pH-dependent conformations of the studied system, as the titratable

groups are allowed to change protonation state as the simulation proceeds.

For a system with titratable groups, for example a protein, a set of CpHMD simulations at different pH can be performed to investigate the titration behaviour and to calculate the pK_a values for groups of interest. Examples of such titration is shown in Figure 10. Using our implementation of CpHMD in GROMACS, we first obtained the titration curve for tripeptides Ala-X-Ala, X being Asp, Glu and His, using both all-atom (AA) CHARMM36* [72] and coarse-grained (CG) Martini 2.0 [73,74] force fields. CHARMM36* refers to the modified CHARMM36 FF with decreased torsion barriers to better sample rotational degrees of freedom in CpHMD simulations, presented in Section 5.3. [143] Our CpHMD code is able to produce the correct fractions of simulation time when each group is deprotonated at a given pH (dots in the titration plots), to which the Henderson-Hasselbalch equation can be fitted (dashed lines). The obtained pK_a values are given in Table 2, and they match well with the used reference pK_a values as they should for titrations in water environment. Importantly, to validate the treatment of the multisite titratable groups, also the microscopic pK_a values of His are well reproduced for CHARMM36 FF.

As an example of a slightly more complex system, we then performed a set of CpHMD simulations to titrate the water-soluble cardiotoxin V protein (PDB ID: 1CVO [7]), for which four residues were held as titratable: His-4, Glu-17, Asp-42 and Asp-59. The structure of the protein together with the titration curves for both CHARMM36* and Martini FFs are shown in Figure 10. The obtained pK_a values are given in Table 2. For this protein, we were able to capture the large downward shift of pK_a compared to the reference pK_a for Asp-59, both with AA and CG levels of theory. The pK_a values obtained with the AA CHARMM36* FF correspond well to the experimentally obtained ones. For the CG Martini simulations, even though the pK_a values are further away from the experimental results, the order of pK_a-s is conserved. Importantly, in all titration curves plotted from the CpHMD simulation data, the shape of the calculated fraction of deprotonated group (dots) is matching well with the fitted Henderson-Hasselbalch equation (dashed lines). In the original publication, we also present the titration of another protein (hen egg white lyzozyme, HEWL [156]).

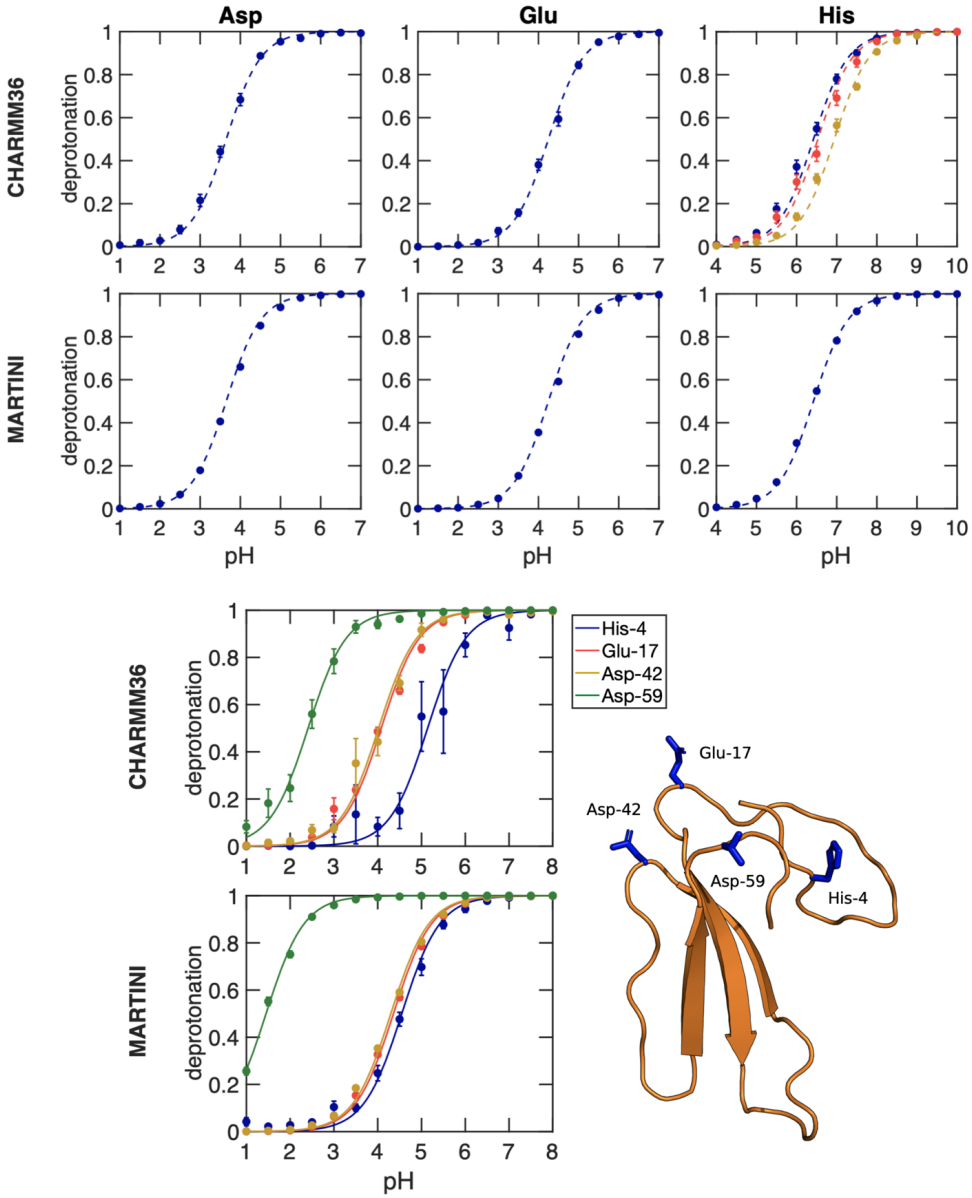


FIGURE 10 Above: Titrations of tripeptides (Asp, Glu, His) in water, with both AA CHARMM36 and CG Martini 2.0 force fields. Below: Titration of the cardiotoxin V protein (PDB ID: 1CVO [7]) with both AA CHARMM36 and CG Martini 2.0 force fields. In all plots, dots show the fraction of frames in which the group is deprotonated, and the dashed lines are the fits to the Henderson-Hasselbalch equation. Obtained pK_a values, with comparison to model or experimental values, are given in Table 2. Adapted with permission from Noora Aho, Pavel Buslaev, Anton Jansen, Paul Bauer, Gerrit Groenhof and Berk Hess. Scalable Constant pH Molecular Dynamics in GROMACS. *Journal of Chemical Theory and Computation*, 18 (10), pp. 6148-6160, **2022**. Copyright © 2022 American Chemical Society.

	pK_a value		
Tripeptides [4,5]	CHARMM36	MARTINI	Ref.
Asp	3.61 ± 0.03	3.69 ± 0.02	3.65
Glu	4.26 ± 0.04	4.30 ± 0.03	4.25
His macroscopic	6.34 ± 0.08	6.40 ± 0.03	6.42
His HSD	6.56 ± 0.06		6.53
His HSE	6.90 ± 0.05		6.94
Cardiotoxin V [157,158]	CHARMM36	MARTINI	Exp.
His-4	5.14 ± 0.16	4.54 ± 0.09	5.5
Glu-17	4.08 ± 0.08	4.36 ± 0.04	4
Asp-42	4.02 ± 0.10	4.30 ± 0.05	3.2
Asp-59	2.41 ± 0.07	1.45 ± 0.03	< 2

TABLE 2 pK_a values obtained from titration simulations for tripeptides and the cardiotoxin V protein (PDB ID: 1CVO [7]). The reference pK_a values for tripeptides, and the experimental pK_a values for the protein are given in the last column. The values for Asp and Glu are from [4], while the microscopic and macroscopic pK_a values for His are taken from [5]. Experimentally obtained pK_a values for cardiotoxin V are from [157, 158]. Adapted with permission from Noora Aho, Pavel Buslaev, Anton Jansen, Paul Bauer, Gerrit Groenhof and Berk Hess. Scalable Constant pH Molecular Dynamics in GROMACS. *Journal of Chemical Theory and Computation*, 18 (10), pp. 6148-6160, **2022**. Copyright © 2022 American Chemical Society.

In general, the CpHMD routine is not bound to any specific FF. Any chemical group of the system can be included as a titratable one, as long as the partial charges, FF parameters are reference pK_a values are known for all protonation states. This is demonstrated in the above results by providing the titration results with both AA and CG descriptions. In addition, the original publications **II** and **III** provide guidelines on how to parametrize new titratable compounds, for which the correction potential V^{MM} has to be obtained. With our improved representation of multisite titratable groups, the parametrization of such groups is general, despite the number of protonation states.

As our CpHMD implementation is accurately able to include the protonation into conventional MD simulations, the second important validation is the performance of the code. As mentioned in Chapter 4, the previous implementation of CpHMD in a fork of GROMACS 3.3.3 [103, 107] was extremely inefficient when multiple titratable groups were added. When the calculation of the derivative of λ-dependent potentials (ie. force acting on the λ-coordinates) is carried out using the charge interpolation (CI) scheme, it was shown that only one electrostatic calculation suffices (Equation 15). To prove the efficiency of our implementation, we performed benchmark simulations

of both the new implementation with CI, and the previous implementation utilizing interpolation of the potential energies (Figure 11A). For the benchmark of the small turkey ovomucoid inhibitor protein, CI provides a comparable performance as traditional MD with only an about 20% computational overhead, whereas the performance of the previous implementation decreases dramatically as the number of titratable groups increases.

Additionally, Figure 11B presents the performance of our CpHMD implementation with respect to traditional MD for a larger membrane-protein simulation system with in total 185 residues held as titratable. Again, the performance of our CpHMD is comparable to standard MD, with about 20% computational overhead when CPU-only is used, and about 35% overhead with the CPU+GPU implementation. These results demonstrate the applicability of our CpHMD code for larger biological systems, for which the previous attempts of implementing CpHMD in GROMACS have not succeeded and hence became outdated as the software has continued to develop.

To summarize, we implemented an accurate and efficient CpHMD in the popular,

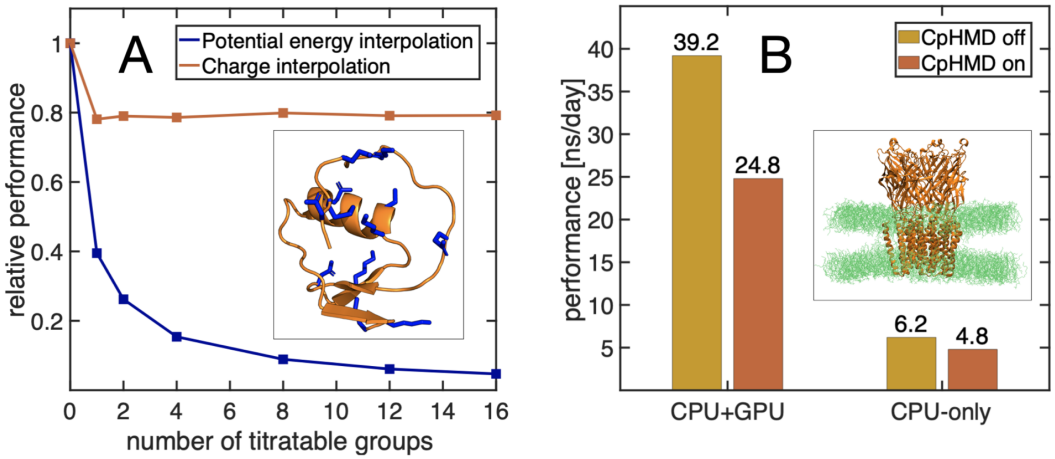


FIGURE 11 Performance of the CpHMD implementation in GROMACS. (A) The relative performance of the previous implementation of CpHMD into a fork of GROMACS 3.3 release (blue) and of our new implementation (red) as the function of the number of titratable sites, obtained from simulations of turkey ovomucoid inhibitor protein. The *relative* performance is the ratio between the average number of integration steps per time unit for a simulation with CpHMD, and the average number of integration steps per time unit for a normal simulation without CpHMD (B) Comparison of the performance between CPU-only and CPU+GPU implementations for the ligand-gated ion channel GLIC [159] with 185 titratable sites. Adapted with permission from Noora Aho, Pavel Buslaev, Anton Jansen, Paul Bauer, Gerrit Groenhof and Berk Hess. Scalable Constant pH Molecular Dynamics in GROMACS. *Journal of Chemical Theory and Computation*, 18 (10), pp. 6148-6160, **2022**. Copyright © 2022 American Chemical Society.

open-source MD software GROMACS. The strength of our implementation lies in the clever combination of existing methods for treating protonation states continuously, utilizing the charge interpolation, representing multi-site titratable groups and keeping the total charge of the simulation system constant while the protonation states change. As the development of the CpHMD code has been done in close collaboration with the GROMACS development team, the pitfall of inability in maintaining the code for future versions of the software can be avoided. As a consequence of the implementation efforts of the thesis author and her colleagues, CpHMD has become an attractive addition to supplement the conventional MD among the large GROMACS user community.

5.3 Best practices in constant pH MD simulations (publication III)

In brief, as the title suggests, publication **III** includes details and best practices regarding CpHMD method, and the main results can be divided in three points:

1. force field modifications to improve sampling of side chain rotamers, and to reach convergence for protonation states and torsion angles
2. importance of the fitting order of $\langle \partial V^{\mathrm{MM}} / \partial \lambda \rangle_{\lambda}$ for the correction potential
3. buffer particle parameters for the charge levelling scheme

The presented results are relevant not only for our implementation of CpHMD in GROMACS but also for other CpHMD packages and softwares, which makes the publication of high importance when aiming for accurate long-timescale CpHMD simulations. In addition to CpHMD, the sampling of rotamers could also be significant in λ-dynamics based free energy calculation methods for protonatable residues.

With the recent efficient CpHMD approaches that utilize the charge interpolation scheme for updating the λ-coordinates describing protonation states, simulations comparable to traditional MD timescales are possible. During the test phase of our CpHMD implementation when running multiple replicas at the same pH, we discovered that distributions of λ-coordinates were not consistent between replicas, even after running a hundred nanoseconds of a tripeptide with only one λ-coordinate. We then tracked down a correlation between inconsistent λ-distributions and insufficient sampling of the dihedral degrees of freedom in the amino acid side chains. The similar conclusion had been drawn also before us, for both standard MD [142] and CpHMD [98, 144]. In the previous

CpHMD applications, the proposed solution for the unconverged syn- and anti-conformations of the carboxyl group proton was to add corrections to the force field in order to (i) lower the torsional barrier and enhance the rotational sampling [98] or (ii) increase the barrier and prevent any rotations from the syn- to anti-configurations. [144]

Our analysis revealed that not only the carboxylic acid group conformations, but also other amino acid sidechain torsions, were not sampled equally between

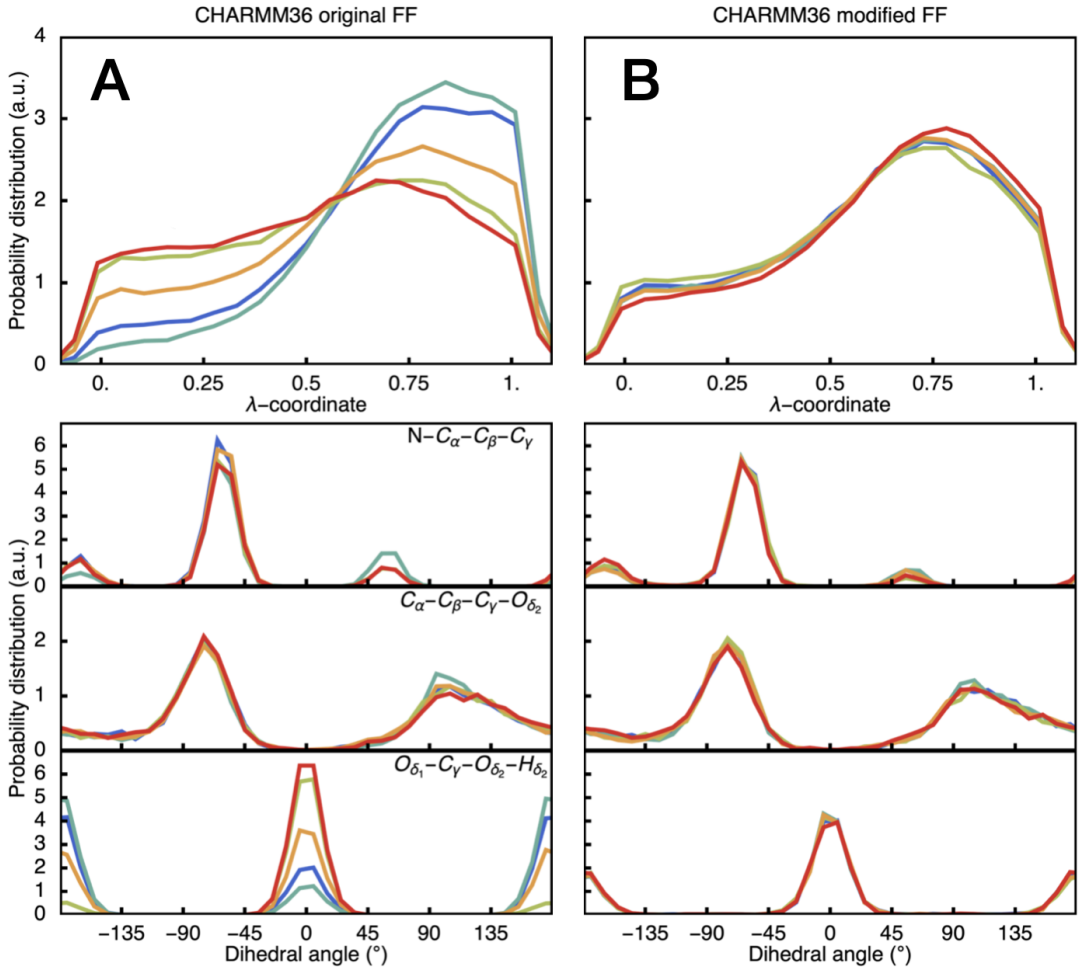


FIGURE 12 Probability distribution of the λ-coordinate and dihedral angles of tripeptide AlaAspAla, simulated at pH=pK_a, height of biasing potential set to zero, and 3rd order fits for $\langle \partial V^{MM}/\partial\lambda \rangle_\lambda$, using (A) original and (B) modified CHARMM36 FF. Different colors correspond to independent replicas. Using the modified FF results in identical distributions for λ-coordinates and dihedral angles N-C_α-C_β-C_γ and O_{δ_1}-C_γ-O_{δ_2}-H_{δ_2}. Adapted with permission from Pavel Buslaev, Noora Aho, Anton Jansen, Paul Bauer, Berk Hess and Gerrit Groenhof. Best practices in constant pH MD simulations: accuracy and sampling *Journal of Chemical Theory and Computation*, 18 (10), pp. 6134-6147, **2022**. Copyright © 2022 American Chemical Society.

different replicas of CpHMD simulations when the original CHARMM36 FF was used. This is demonstrated in Figure 12A, where the distributions of the λ-coordinate and three torsional angles, namely N-C_α-C_β-C_γ and O_{δ_1}-C_γ-O_{δ_2}-H_{δ_2} are presented. We performed 5 replicas of 100 ns of CpHMD for a tripeptide aspartic acid (Asp) in water (Glu, His, Lys, N-ter and C-ter can be found in the supplementary of the original manuscript). CpHMD was run at pH=pK_a, height of biasing potential set to zero, and with 3rd order fits for $\langle \partial V^{MM}/\partial\lambda \rangle_\lambda$.

When the maximas of the torsion potentials were reduced, by adding a small dihedral angle (ϕ) dependent correction, the distributions of the λ-coordinate and the torsional angles are consistent (Figure 12B). As a note, the non-flat λ-distributions originate from using a 3rd order fit for $\langle \partial V^{MM}/\partial\lambda \rangle_\lambda$ instead of a higher order, which will be demonstrated in the following results. In general, a linear, [108] or 3rd order [103] fit to $\langle \partial V^{MM}/\partial\lambda \rangle_\lambda$ has been commonly used.

The dihedral angle (ϕ) dependent correction to the torsion potentials was designed in a way that it will both lower the torsional barriers and still preserve the shape of the minimas. The correction had the following form:

$$\Delta V_i(\phi) = \epsilon_i \left[\cos(n_i\phi) - (-1)^{n_i} \frac{1}{4} \cos(2n_i\phi) \right], \tag{17}$$

where ϵ_i is an empirical coefficient optimized to decrease the barriers of the torsion potential and converge the ϕ_i-distributions, and n_i is the multiplicity of torsion angle i ($n_i = 2$ for conjugated bonds and $n_i = 3$ for aliphatic bonds). For each torsion in the amino acid side chain, for which we observed inconsistency between CpHMD replicas, ϵ_i was obtained using an iterative scheme: (1) with an initial guess of ϵ_i we calculated the free energy profiles for each torsion, (2) using the ϵ_i we recalculated the correction potential $V^{MM}(\lambda)$ by fitting a 3rd order polynomial to the $\langle \partial V/\partial\lambda \rangle_\lambda$ values, and (3) performed five CpHMD replicas at pH = pK_a and with biasing potential barrier set to zero. After these steps, we compared the distributions of the λ-coordinates, and iteratively increased the ϵ_i by the listed three steps, until the distributions were converged.

The introduced force field modifications for Asp sidechain are shown in Figure 13, together with the atom nomenclature. Panel A presents the forms of the correction potentials $\Delta V_i(\phi)$ from Equation 17 for both $n = 2$ and $n = 3$. Panel B contains the torsional potentials obtained using the original and modified CHARMM36 FF, for both deprotonated and protonated states of the group. From the presented torsional potentials we can also confirm that the locations of minimas are not altered upon the correction, but only the height of the barriers. FF modifications and replicas of sampling runs for the other titratable amino acid side chains (Glu, His, Lys, N-ter and C-ter) can be found in the supplementary

of the original manuscript. In the original publication, we also compare the potential energy profile for the Asp C_α-C_β dihedral potential profile obtained using original and modified CHARMM36 FFs, to a profile obtained by MP2 level of theory QM calculation (publication **II** Figure S9). By comparing the deviation between MM and QM profiles for the dihedral, we can conclude that the modified FF is an equally good fit to the QM potential.

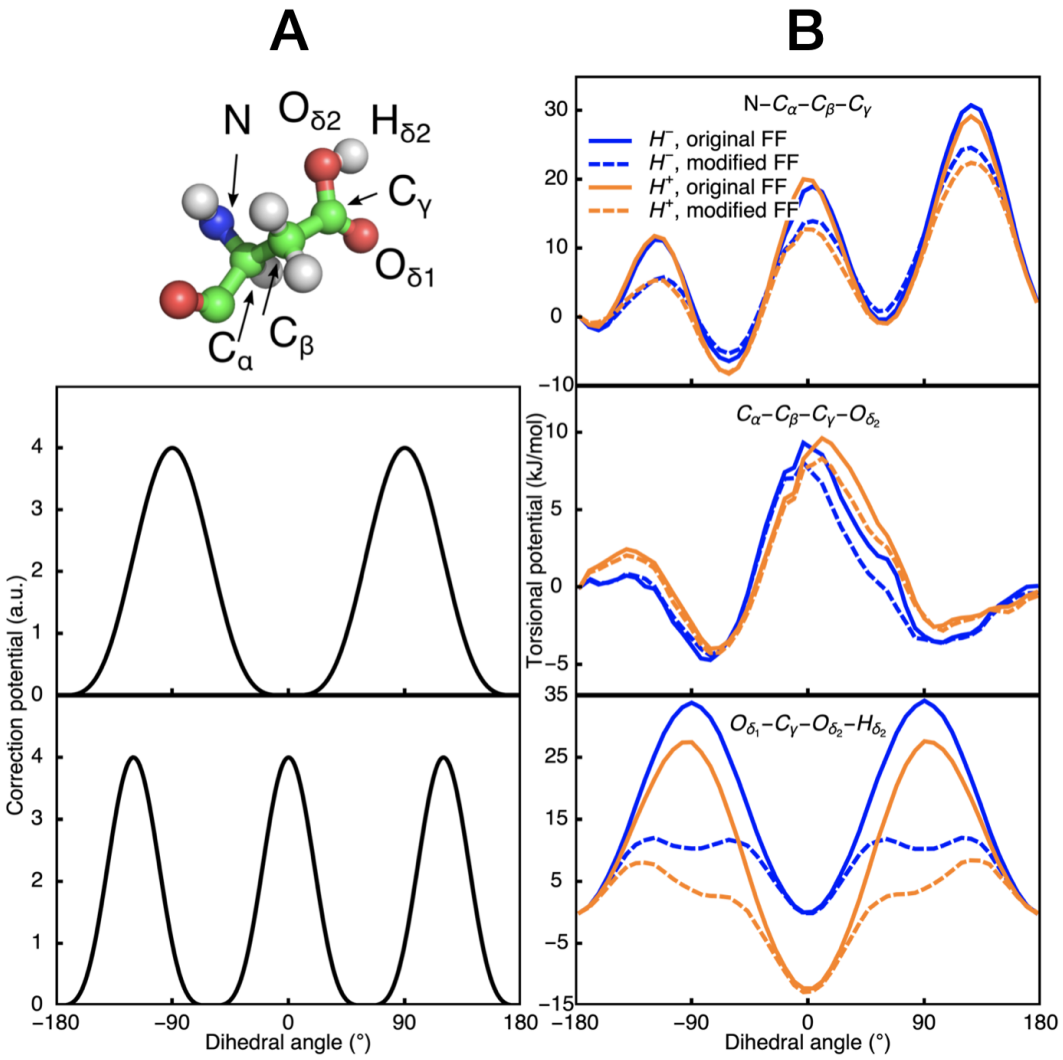


FIGURE 13 Introduced force field modifications of CHARMM36 force field for the aspartic acid sidechain torsions. (A) Structure and nomenclature of the atoms of Asp, below which are the correction potentials added to dihedral torsions of the original force field, for the cases of two and three local minima. (B) Original and modified torsional barriers of Asp, for protonated (H+) and deprotonated (H-) states. Modifications were introduced for three torsions: N-C_α-C_β-C_γ, C_α-C_β-C_γ-O_{δ_2} and O_{δ_1}-C_γ-O_{δ_2}-H_{δ_2}. Adapted with permission from Pavel Buslaev, Noora Aho, Anton Jansen, Paul Bauer, Berk Hess and Gerrit Groenhof. Best practices in constant pH MD simulations: accuracy and sampling *Journal of Chemical Theory and Computation*, 18 (10), pp. 6134-6147, **2022**. Copyright © 2022 American Chemical Society.

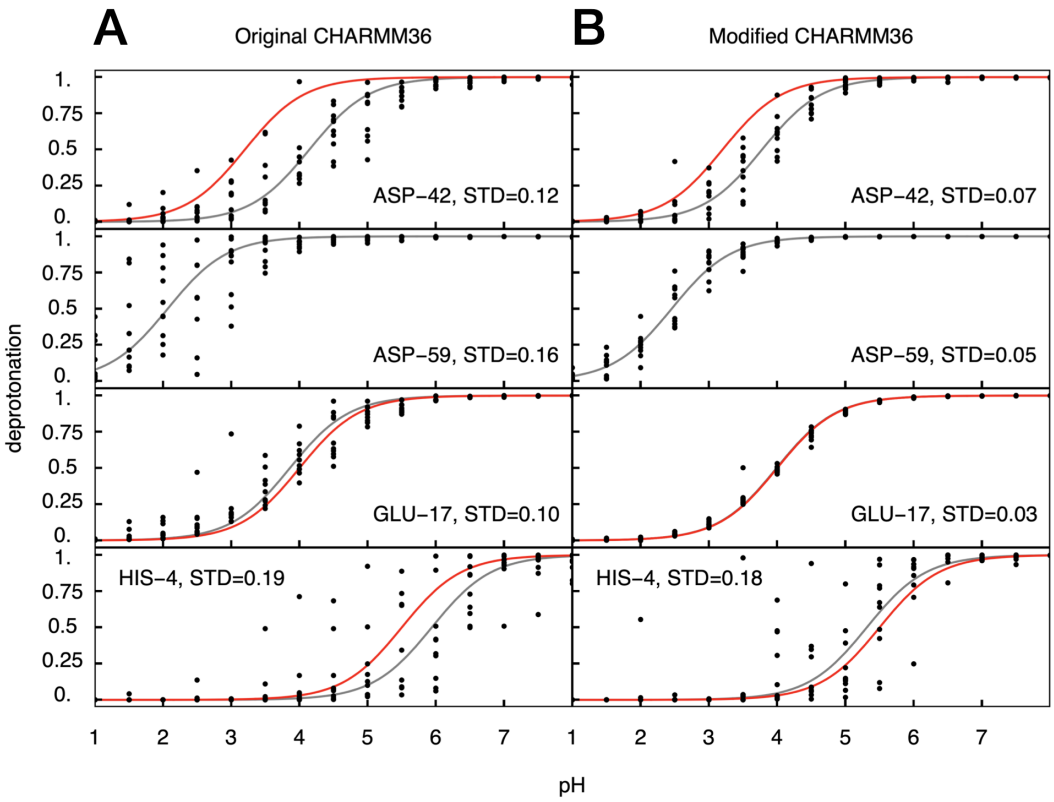


FIGURE 14 Titration of cardiotoxin V using the (A) original and (B) modified CHARMM36 force fields. Points show the fractions of deprotonated acid for each individual replica. Gray and red lines show the fitted and experimental Henderson-Hasselbalch curves, respectively. For each titratable group, the standard deviation between the fitted average deprotonation and the individual replicas is calculated. Adapted with permission from Pavel Buslaev, Noora Aho, Anton Jansen, Paul Bauer, Berk Hess and Gerrit Groenhof. Best practices in constant pH MD simulations: accuracy and sampling *Journal of Chemical Theory and Computation*, 18 (10), pp. 6134-6147, **2022**. Copyright © 2022 American Chemical Society.

As an example of the force field modifications in practice, Figure 14 presents a comparison of titrations of the cardiotoxin V protein, using CpHMD with original and modified CHARMM36 FFs. In order to demonstrate the effect of lowering the torsional barriers resulting in more consistent rotamer sampling and thus better convergence of protonation states, we performed 10 replicas of 100 ns CpHMD simulations of the protein at a pH range from 1 to 8, and calculated the fraction of deprotonated acid for the individual replicas. When the modified FF is used, the standard deviation (STD) between the fitted average deprotonation fraction and the individual replicas decreases, corresponding to improved sampling (STD values shown in the legends of Figure 14). In addition, the estimated pK_a values are closer to the experimentally obtained ones when the modified CHARMM36 FF is used, which also verifies the usage of the FF modifications.

To conclude the part of force field modifications, our extensive simulations of tripeptides revealed inconsistencies in both the sampling of protonation states (λ-coordinates), and the torsion angles of amino acid sidechains. We made a detailed check of the free energy profiles of the torsional angles, and added a systematic correction, which made the λ- and dihedral distributions consistent between different replicas, thus improving the sampling of dihedrals. Using the introduced FF corrections also improves the consistency of titration replicas and pK_a estimation, as demonstrated for the cardiotoxin V protein.

During the investigation of torsion angles and convergence of protonation states we also noticed that even for the modified FF at pH=pK_a and height of biasing potential set to zero, the distributions of λ-coordinates were not flat, as can be seen in Figure 12. At these conditions the underlying potential should be flat for a titratable residue in water, if the correction potential accurately represents the free energy of deprotonating the titratable group (Figure 6 A+B). As using 3rd order fits for $\langle \partial V^{MM}/\partial\lambda \rangle_\lambda$ resulted in non-flat distributions of the long sampling replicas, presented in Figure 12, this suggested that the order of the polynomial fit was not sufficient. Figure 15 presents the results regarding the importance of higher order polynomial fits for the $\langle \partial V^{MM}/\partial\lambda \rangle_\lambda$. In panel A, an example parameterization is shown: 10 ns CpHMD simulations of tripeptide AlaAspAla at fixed λ-values, from which the average $\langle \partial V^{MM}/\partial\lambda \rangle_\lambda$ as a function of λ is plotted. Even though the polynomial fits of different orders (1st, 3rd and 7th) seem to be consistent, the zoomed in view shown on the right reveals deviations. Additionally, since the $\langle \partial V^{MM}/\partial\lambda \rangle_\lambda$ values are in the order of hundreds of kJ/mol, relatively small differences in the plots are actually remarkable.

In panel B of Figure 15 the mean error is shown as a function of fitting order (left) and λ-distributions for 100ns sampling runs for tripeptide AlaAspAla (right). The distributions clearly show the need of higher 7th order polynomial fit in order to obtain a flat distribution during CpHMD. Similar analysis was performed for the other titratable amino acid side chains (Glu, His, Lys, N-ter and C-ter), and the suggested fitting orders can be found in the original manuscript. Based on these analysis, the author of this thesis recommends all CpHMD users to perform a similar analysis of the fitting order for their titratable groups.

In addition to the force field modifications, publication **III** justifies the parameter selections for the buffer particles introduced for the charge levelling scheme. As the protonation states and thus charges of the titratable groups change dynamically during the simulation, charge levelling is used to keep the total charge of the simulation box neutral during the CpHMD simulations. The scheme is introduced in Chapter 4, but in a nutshell, the idea is to add buffer particles to the system, and constrain the total charge of these buffers and the titratable groups such that the total charge remains neutral.

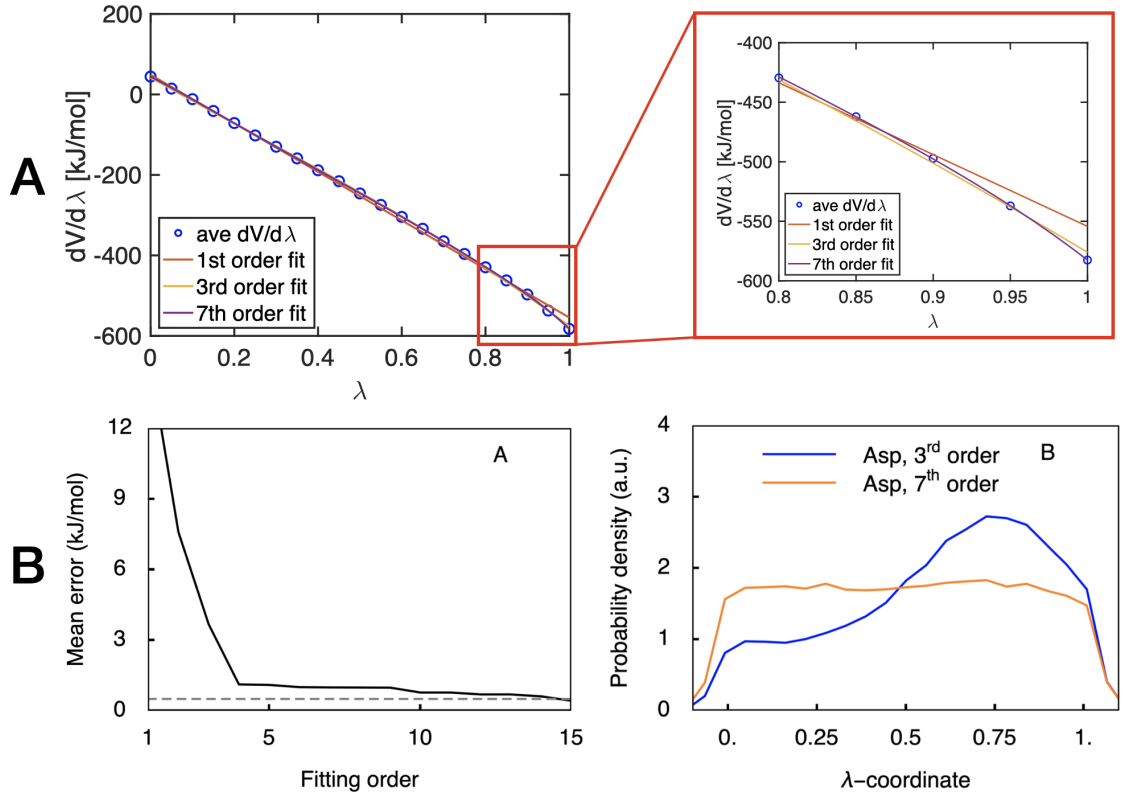


FIGURE 15 Importance of the fitting order of $\langle \partial V^{\mathrm{MM}}/\partial\lambda \rangle_\lambda$. (A) Example of a polynomial fit to $\langle \partial V^{\mathrm{MM}}/\partial\lambda \rangle_\lambda$ for tripeptide AlaAspAla. Points show the average $\langle \partial V^{\mathrm{MM}}/\partial\lambda \rangle_\lambda$ obtained from 10 ns simulations at fixed λ-values, and red, yellow and violet lines show the fitted polynomial of 1$^{\mathrm{st}}$, 3$^{\mathrm{rd}}$ and 7$^{\mathrm{th}}$ order, respectively. The right plot is zoomed in to the range $\lambda = [0.8, 1.0]$, to highlight the poor quality of the fit for lower orders. (B) On the left, mean error of the polynomial fit with respect to the fitting order, calculated for tripeptide AlaAspAla. On the right, distribution of the λ-coordinate from CpHMD simulations at pH=pK_{a}, height of biasing potential set to zero for 3$^{\mathrm{rd}}$ and 7$^{\mathrm{th}}$ order fits. Panel B adapted with permission from Pavel Buslaev, Noora Aho, Anton Jansen, Paul Bauer, Berk Hess and Gerrit Groenhof. Best practices in constant pH MD simulations: accuracy and sampling *Journal of Chemical Theory and Computation*, 18 (10), pp. 6134-6147, **2022**. Copyright © 2022 American Chemical Society.

There are few points which one should consider when adding charged buffer particles into the solvent:

1. minimizing the impact of charging the buffer, in order to avoid local rearrangement of the hydrogen bonding network
2. preventing the clustering of multiple buffer particles in the solvent
3. preventing the permeation or interaction of buffer particles into or with a hydrophobic environments, such as a membranes or proteins

The first point was addressed by finding out an optimal range for the charge of a single buffer particle, based on accelerated weight histogram (AWH) simulations [160]. During this adaptive free energy simulation, a so-called friction metric is computed, which corresponds to the efficiency of sampling. In order to keep the friction less than 50% larger as compared to zero charges, and to avoid slowing down the dynamics of the local environment, we settled upon a charge range from -0.5 e to +0.5 e for the buffer particle. The points 2 and 3 regarding buffer clustering and permeation were covered by adjusting the L-J parameters of the buffers. By performing test simulations with different values, where radial distribution functions between buffers and a protein, and densities of buffers with respect to a membrane were compared, we found suitable parameters of $\sigma = 0.25$ nm and $\epsilon = 4$ kJ/mol.

When combined, the results of publication **III** not only complement our implementation of CpHMD in GROMACS, but also suggest guidelines on parameterizational aspects of CpHMD simulations in general, regardless of force field or software used. In our publication, the need and the effect of a well-defined protocol to enhance rotamer sampling was clearly seen for a single titratable group in water and further tested for a protein system. In addition, the importance of the fitting order for the correction potential was justified. As the efficiency of the CpHMD methods is approaching standard MD and longer timescales can be reached, any detail leading to better convergence is worth paying attention to, especially when parameterizing novel titratable groups for CpHMD.

As a final remark, the author of this thesis wants to point out that all of our work regarding CpHMD is open access and available for the community. The source code for the CpHMD in GROMACS is public, supplied with a manual and tutorials, and can be downloaded from here: https://gitlab.com/gromacs-constantph/constantph/. The modified force field, all correction potential fitting coefficients, and buffer parameters are published along with the publications, together with the standard information about protocols, force field and input parameters. Our aim is to provide the MD community a possibility to test and apply our implementation of CpHMD for exploring pH-dependent biomolecular processes, along with suggesting improvements or providing critical comments to the method itself. The open data policy is of "greatest importance for the scientific quality of simulations" [161]

6 SUMMARY AND OUTLOOK

To expand the fundamental understanding of nature, the work of this thesis focuses on pH-dependent processes in biology at the atomistic scale. As the current experimental methods are still unable to capture the nanoscale dynamics of these processes in full detail, our goal was to further develop the computational methods aiming to model pH and proton transfer in classical molecular dynamics simulations.

The main result and product of this thesis is the efficient constant pH molecular dynamics implementation to the GROMACS software (publication **II**). Constant pH MD enables dynamically changing the protonation states during classical MD simulations, for which simulating at a fixed pH is traditionally not possible. With constant pH MD, the sampling of protonation states together with atomic coordinates is possible, which enables the investigation of pH-dependent phenomena in proteins, membranes and other biomolecular systems usually studied by means of MD simulations.

Constant ph MD is not an entirely novel method. There exists a previous CpHMD developed for GROMACS from a decade ago, [103] but due to the speed decrease with the increasing number of titratable groups it was never imported to the main branch of GROMACS. With the improved CpHMD method, we succeed to achieve a comparable efficiency to traditional MD, allowing the investigation of larger biomolecules and longer timescales. There also exist efficient CpHMD routines based on the same principles to compute the electrostatic potential for other softwares, namely CHARMM and AMBER. [108, 112] In contrast to the others, GROMACS is a free and open-access software with a large userbase, which is why an efficient CpHMD has been a decade long request from the simulation community. In addition to being efficient, our implementation also

combines the important features of CpHMD: description of groups with multiple titratable sites, a scheme to keep the total charge of the box neutral upon (de)protonation events, and the generality to any available force field.

The close collaboration with the core development team ensures the further support of our CpHMD code in the future releases of GROMACS. A highly efficient, open-access CpHMD in the popular GROMACS software will provide the MD community the possibility to dive into the world of pH-dependent processes of biomolecules. In addition to the open-access code, we also share in detail the parameterization process of titratable groups and relevant input files, and provide a tutorial regarding CpHMD simulations with GROMACS. Being transparent about the code and input increases the impact of the work, as other people in the field can openly learn, use, and also criticize the CpHMD implementation.

In addition to providing a new CpHMD implementation, we also suggest general best practices for CpHMD simulations, that hold regardless of software or force field used (publication **III**). In order to enhance the sampling of dihedral angles of the amino acid sidechains, thus speeding up the convergence of the protonation states, we introduce a small correction term to the torsion potentials of common amino acids for the CHARMM36 force field. We also share details about parameterizing new titratable groups (*ie.* calculating V^{MM} accurately) and keeping the total charge of the simulation box neutral upon protonation state changes.

Alongside with CpHMD, this thesis presents an example of supporting experimental results of proton transfer on phospholipid membranes using conventional MD with fixed protonation states (publication **I**). We parameterized the C12-HPTS photoacid used in the experimental research and simulated it inside three lipid bilayers with differently charged headgroups. We were able to provide atomistic details about the location of the photoacid embedded in the membranes by analyzing the density of different molecules in the system. Additionally, the proton transfer from the photoacid to either the headgroups of the lipids or the surrounding bulk water was investigated, by analyzing the hydrogen bonding of the hydroxyl group of the photoacid.

As illustrated by the latter, the choice of the computational method depends on the studied phenomena. CpHMD provides a possibility to track protonation state changes dynamically in biomolecules but the method might not be a necessity, or even a possibility, for all situations and systems. It is important to remember, that CpHMD is not a magic bullet to treat every pH-dependent phenomena involving proton transfer, but the method has to be relevant for the specific research question in mind. In many situations, traditional MD with fixed protonation states still has its advantages.

To conclude, being able to connect dynamics of atomic positions with dynamics of protonation states using CpHMD with barely any extra computational cost will broaden the scope of traditional MD. As the methods to include pH and protonation state changes are being further developed, the possibilities to study even more complex processes by simulations increase. This will both support and motivate the experimental research of pH-dependent process in biology, and ultimately enhance our understanding of the fundamentals of nature.

REFERENCES

[1] A. Warshel, "Calculations of enzymic reactions: calculations of pK_a, proton transfer reactions, and general acid catalysis reactions in enzymes," *Biochemistry*, vol. 20, no. 11, pp. 3167–3177, 1981.

[2] C. Von Ballmoos, A. Wiedenmann, and P. Dimroth, "Essentials for ATP synthesis by F_1F_0 ATP synthases," *Annual review of biochemistry*, vol. 78, pp. 649–672, 2009.

[3] R. W. Putnam, "Intracellular pH regulation," in *Cell physiology source book*, pp. 357–372, Elsevier, 2001.

[4] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, and C. N. Pace, "pK_a values of the ionizable groups of proteins," *Protein science*, vol. 15, no. 5, pp. 1214–1218, 2006.

[5] M. Tanokura, "1H-NMR study on the tautomerism of the imidazole ring of histidine residues: I. microscopic pK_a values and molar ratios of tautomers in histidine-containing peptides," *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, vol. 742, no. 3, pp. 576–585, 1983.

[6] T. K. Harris and G. J. Turner, "Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites," *IUBMB life*, vol. 53, no. 2, pp. 85–98, 2002.

[7] A. K. Singhal, K. Y. Chien, W. G. Wu, and G. S. Rule, "Solution structure of cardiotoxin V from naja naja atra," *Biochemistry*, vol. 32, no. 31, pp. 8036–8044, 1993.

[8] D. L. Rabenstein and T. L. Sayer, "Determination of microscopic acid dissociation constants by nuclear magnetic resonance spectrometry," *Analytical Chemistry*, vol. 48, no. 8, pp. 1141–1146, 1976.

[9] J. Reijenga, A. Van Hoof, A. Van Loon, and B. Teunissen, "Development of methods for the determination of pk_a values," *Analytical chemistry insights*, vol. 8, pp. ACI–S12304, 2013.

[10] L. A. Flexser, L. P. Hammett, and A. Dingwall, "The determination of ionization by ultraviolet spectrophotometry: Its validity and its application to the measurement of the strength of very weak bases1," *Journal of the American Chemical Society*, vol. 57, no. 11, pp. 2103–2115, 1935.

[11] J. Song, M. Laskowski, M. Qasim, and J. L. Markley, "NMR determination of pK_a values for asp, glu, his, and lys mutants at each variable contiguous enzyme- inhibitor contact position of the turkey ovomucoid third domain," *Biochemistry*, vol. 42, no. 10, pp. 2847–2856, 2003.

[12] R. Bürgi, P. A. Kollman, and W. F. Van Gunsteren, "Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation," *Proteins: Structure, Function, and Bioinformatics*, vol. 47, no. 4, pp. 469–480, 2002.

[13] T. Simonson, J. Carlsson, and D. A. Case, "Proton binding to proteins: pKa calculations with explicit and implicit solvent models," *Journal of the American Chemical Society*, vol. 126, no. 13, pp. 4167–4180, 2004.

[14] H. Li, A. D. Robertson, and J. H. Jensen, "Very fast empirical prediction and rationalization of protein pKa values," *Proteins: Structure, Function, and Bioinformatics*, vol. 61, no. 4, pp. 704–721, 2005.

[15] F. Milletti, L. Storchi, and G. Cruciani, "Predicting protein pKa by environment similarity," *Proteins: Structure, Function, and Bioinformatics*, vol. 76, no. 2, pp. 484–495, 2009.

[16] M. H. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen, "PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions," *Journal of chemical theory and computation*, vol. 7, no. 2, pp. 525–537, 2011.

[17] K. P. Kilambi and J. J. Gray, "Rapid calculation of protein pKa values using rosetta," *Biophysical journal*, vol. 103, no. 3, pp. 587–595, 2012.

[18] W. Chen, B. H. Morrow, C. Shi, and J. K. Shen, "Recent development and application of constant pH molecular dynamics," *Molecular simulation*, vol. 40, no. 10-11, pp. 830–838, 2014.

[19] E. Alexov, E. L. Mehler, N. Baker, A. M. Baptista, Y. Huang, F. Milletti, J. Erik Nielsen, D. Farrell, T. Carstensen, M. H. Olsson, *et al.*, "Progress in the prediction of pK_a values in proteins," *Proteins: structure, function, and bioinformatics*, vol. 79, no. 12, pp. 3260–3275, 2011.

[20] H. Gokcan and O. Isayev, "Prediction of protein pK_a with representation learning," *Chemical science*, vol. 13, no. 8, pp. 2462–2474, 2022.

[21] A. Y. Chen, J. Lee, A. Damjanovic, and B. R. Brooks, "Protein p k a prediction by tree-based machine learning," *Journal of Chemical Theory and Computation*, vol. 18, no. 4, pp. 2673–2686, 2022.

[22] C. Tanford and J. G. Kirkwood, "Theory of protein titration curves. i. general equations for impenetrable spheres," *Journal of the American Chemical Society*, vol. 79, no. 20, pp. 5333–5339, 1957.

[23] M. K. Gilson, A. Rashin, R. Fine, and B. Honig, "On the calculation of electrostatic interactions in proteins," *Journal of molecular biology*, vol. 184, no. 3, pp. 503–516, 1985.

[24] D. Bashford and M. Karplus, "pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model," *Biochemistry*, vol. 29, no. 44, pp. 10219–10225, 1990.

[25] A.-S. Yang, M. Gunner, R. Sampogna, K. Sharp, and B. Honig, "On the calculation of pK_as in proteins," *Proteins: Structure, Function, and Bioinformatics*, vol. 15, no. 3, pp. 252–265, 1993.

[26] J. Antosiewicz, J. A. McCammon, and M. K. Gilson, "The determinants of pK_as in proteins," *Biochemistry*, vol. 35, no. 24, pp. 7819–7833, 1996.

[27] J. J. Havranek and P. B. Harbury, "Tanford–kirkwood electrostatics for protein modeling," *Proceedings of the National Academy of Sciences*, vol. 96, no. 20, pp. 11145–11150, 1999.

[28] R. E. Georgescu, E. G. Alexov, and M. R. Gunner, "Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins," *Biophysical journal*, vol. 83, no. 4, pp. 1731–1748, 2002.

[29] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev, "H++ 3.0: automating pK_a prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations," *Nucleic acids research*, vol. 40, no. W1, pp. W537–W541, 2012.

[30] C. Li, Z. Jia, A. Chakravorty, S. Pahari, Y. Peng, S. Basu, M. Koirala, S. K. Panday, M. Petukh, L. Li, *et al.*, "DelPhi suite: new developments and review of functionalities," *Journal of computational chemistry*, vol. 40, no. 28, pp. 2502–2508, 2019.

[31] D. Bashford and D. A. Case, "Generalized born models of macromolecular solvation effects," *Annual review of physical chemistry*, vol. 51, no. 1, pp. 129–152, 2000.

[32] M. Feig and C. L. Brooks III, "Recent advances in the development and application of implicit solvent models in biomolecule simulations," *Current opinion in structural biology*, vol. 14, no. 2, pp. 217–224, 2004.

[33] T. J. You and D. Bashford, "Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility," *Biophysical journal*, vol. 69, no. 5, pp. 1721–1733, 1995.

[34] E. Alexov and M. Gunner, "Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties," *Biophysical journal*, vol. 72, no. 5, pp. 2075–2093, 1997.

[35] M. D. Liptak and G. C. Shields, "Accurate pK_a calculations for carboxylic acids using complete basis set and gaussian-n models combined with cpcm continuum solvation methods," *Journal of the American Chemical Society*, vol. 123, no. 30, pp. 7314–7319, 2001.

[36] Y. E. Zevatskii and D. Samoilov, "Modern methods for estimation of ionization constants of organic compounds in solution," *Russian Journal of Organic Chemistry*, vol. 47, no. 10, pp. 1445–1467, 2011.

[37] H. Li, A. D. Robertson, and J. H. Jensen, "The determinants of carboxyl pKa values in turkey ovomucoid third domain," *Proteins: Structure, Function, and Bioinformatics*, vol. 55, no. 3, pp. 689–704, 2004.

[38] B. Kuhn, P. A. Kollman, and M. Stahl, "Prediction of pK_a shifts in proteins using a combination of molecular mechanical and continuum solvent calculations," *Journal of computational chemistry*, vol. 25, no. 15, pp. 1865–1872, 2004.

[39] K. Szalewicz, "Hydrogen bond," in *Encyclopedia of Physical Science and Technology (Third Edition)* (R. A. Meyers, ed.), pp. 505–538, New York: Academic Press, third edition ed., 2003.

[40] N. Amdursky, Y. Lin, N. Aho, and G. Groenhof, "Exploring fast proton transfer events associated with lateral proton diffusion on the surface of membranes," *Proceedings of the National Academy of Sciences*, vol. 116, no. 7, pp. 2443–2451, 2019.

[41] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature structural biology*, vol. 9, no. 9, pp. 646–652, 2002.

[42] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, "Long-timescale molecular dynamics simulations of protein structure and function," *Current opinion in structural biology*, vol. 19, no. 2, pp. 120–127, 2009.

[43] J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, "Molecular dynamics simulations of large macromolecular complexes," *Current opinion in structural biology*, vol. 31, pp. 64–74, 2015.

[44] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, "Molecular dynamics simulations: advances and applications," *Advances and applications in bioinformatics and chemistry: AABC*, vol. 8, p. 37, 2015.

[45] S. A. Hollingsworth and R. O. Dror, "Molecular dynamics simulation for all," *Neuron*, vol. 99, no. 6, pp. 1129–1143, 2018.

[46] A. R. Leach and A. R. Leach, *Molecular modelling: principles and applications*. Pearson education, 2001.

[47] L. Verlet, "Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules," *Physical review*, vol. 159, no. 1, p. 98, 1967.

[48] R. W. Hockney, "The potential calculation and some applications," *Methods Comput. Phys.*, vol. 9, p. 136, 1970.

[49] W. Van Gunsteren and H. Berendsen, "Algorithms for brownian dynamics," *Molecular Physics*, vol. 45, no. 3, pp. 637–647, 1982.

[50] N. Goga, A. Rzepiela, A. De Vries, S. Marrink, and H. Berendsen, "Efficient algorithms for langevin and DPD dynamics," *Journal of chemical theory and computation*, vol. 8, no. 10, pp. 3637–3649, 2012.

[51] D. Frenkel and B. Smit, "Chapter 2 - statistical mechanics," in *Understanding Molecular Simulation (Second Edition)* (D. Frenkel and B. Smit, eds.), pp. 9–22, San Diego: Academic Press, second edition ed., 2002.

[52] R. Pathria and P. D. Beale, "2 - elements of ensemble theory," in *Statistical Mechanics (Third Edition)* (R. Pathria and P. D. Beale, eds.), pp. 25–38, Boston: Academic Press, third edition ed., 2011.

[53] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2 ed., 2004.

[54] H. J. Berendsen, J. v. Postma, W. F. Van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of chemical physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[55] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *The Journal of chemical physics*, vol. 126, no. 1, p. 014101, 2007.

[56] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular physics*, vol. 52, no. 2, pp. 255–268, 1984.

[57] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Physical review A*, vol. 31, no. 3, p. 1695, 1985.

[58] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied physics*, vol. 52, no. 12, pp. 7182–7190, 1981.

[59] M. Bernetti and G. Bussi, "Pressure control using stochastic cell rescaling," *The Journal of Chemical Physics*, vol. 153, no. 11, p. 114107, 2020.

[60] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1, pp. 19–25, 2015.

[61] M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln," *Annalen der physik*, vol. 389, no. 20, pp. 457–484, 1927.

[62] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *The journal of physical chemistry B*, vol. 102, no. 18, pp. 3586–3616, 1998.

[63] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, "LINCS: a linear constraint solver for molecular simulations," *Journal of computational chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997.

[64] S. Miyamoto and P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," *Journal of computational chemistry*, vol. 13, no. 8, pp. 952–962, 1992.

[65] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of computational physics*, vol. 23, no. 3, pp. 327–341, 1977.

[66] G. Ciccotti and J.-P. Ryckaert, "Molecular dynamics simulation of rigid molecules," *Computer Physics Reports*, vol. 4, no. 6, pp. 346–392, 1986.

[67] A. D. MacKerell Jr, "Empirical force fields for biological macromolecules: overview and issues," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1584–1604, 2004.

[68] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, "A generalized reaction field method for molecular dynamics simulations," *The Journal of chemical physics*, vol. 102, no. 13, pp. 5451–5459, 1995.

[69] T. Darden, D. York, and L. Pedersen, "Particle mesh ewald: An N·log (N) method for ewald sums in large systems," *The Journal of chemical physics*, vol. 98, no. 12, pp. 10089–10092, 1993.

[70] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald method," *The Journal of chemical physics*, vol. 103, no. 19, pp. 8577–8593, 1995.

[71] L. Monticelli and D. P. Tieleman, *Force Fields for Classical Molecular Dynamics*, pp. 197–213. Totowa, NJ: Humana Press, 2013.

[72] J. Huang and A. D. MacKerell Jr, "CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data," *Journal of computational chemistry*, vol. 34, no. 25, pp. 2135–2145, 2013.

[73] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries, "The MARTINI force field: coarse grained model for biomolecular simulations," *The journal of physical chemistry B*, vol. 111, no. 27, pp. 7812–7824, 2007.

[74] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, and S. J. Marrink, "Polarizable water model for the coarse-grained MARTINI force field," *PLoS computational biology*, vol. 6, no. 6, p. e1000810, 2010.

[75] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *Journal of computational chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.

[76] J. P. Jämbeck and A. P. Lyubartsev, "Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids," *The journal of physical chemistry B*, vol. 116, no. 10, pp. 3164–3179, 2012.

[77] J. P. Jambeck and A. P. Lyubartsev, "An extension and further validation of an all-atomistic force field for biological membranes," *Journal of chemical theory and computation*, vol. 8, no. 8, pp. 2938–2948, 2012.

[78] K. Wüthrich, "Protein structure determination in solution by NMR spectroscopy.," *Journal of Biological Chemistry*, vol. 265, no. 36, pp. 22059–22062, 1990.

[79] A. Ilari and C. Savino, "Protein structure determination by x-ray crystallography," *Bioinformatics*, pp. 63–87, 2008.

[80] K. M. Yip, N. Fischer, E. Paknia, A. Chari, and H. Stark, "Atomic-resolution protein structure determination by cryo-EM," *Nature*, vol. 587, no. 7832, pp. 157–161, 2020.

[81] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: a web-based graphical user interface for CHARMM," *Journal of computational chemistry*, vol. 29, no. 11, pp. 1859–1865, 2008.

[82] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, *et al.*, "CHARMM-GUI input generator for NAMD, GROMACS, AMBER, openMM, and CHARMM/openMM simulations using the CHARMM36 additive force field," *Journal of chemical theory and computation*, vol. 12, no. 1, pp. 405–413, 2016.

[83] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.

[84] G. T. Ibragimova and R. C. Wade, "Importance of explicit salt ions for protein stability in molecular dynamics simulation," *Biophysical journal*, vol. 74, no. 6, pp. 2906–2911, 1998.

[85] A. Cordomí, O. Edholm, and J. J. Perez, "Effect of ions on a dipalmitoyl phosphatidylcholine bilayer. a molecular dynamics simulation study," *The Journal of Physical Chemistry B*, vol. 112, no. 5, pp. 1397–1408, 2008.

[86] J. S. Hub, B. L. de Groot, H. Grubmüller, and G. Groenhof, "Quantifying artifacts in ewald simulations of inhomogeneous systems with a net charge," *Journal of chemical theory and computation*, vol. 10, no. 1, pp. 381–390, 2014.

[87] D. Frenkel and B. Smit, "Chapter 4 - molecular dynamics simulations," in *Understanding Molecular Simulation (Second Edition)* (D. Frenkel and B. Smit, eds.), pp. 63–107, San Diego: Academic Press, second edition ed., 2002.

[88] M. González, "Force fields and molecular dynamics simulations," *École thématique de la Société Française de la Neutronique*, vol. 12, pp. 169–200, 2011.

[89] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic validation of protein force fields against experimental data," *PloS one*, vol. 7, no. 2, p. e32131, 2012.

[90] D. Poger, B. Caron, and A. E. Mark, "Validating lipid force fields against experimental data: Progress, challenges and perspectives," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1858, no. 7, pp. 1556–1565, 2016.

[91] J. Mongan and D. A. Case, "Biomolecular simulations at constant pH," *Current opinion in structural biology*, vol. 15, no. 2, pp. 157–163, 2005.

[92] D. A. Case, H. M. Aktulga, K. Belfon, I. Ben-Shalom, S. R. Brozell, D. Cerutti, T. Cheatham III, G. Cisneros, V. Cruzeiro, T. Darden, *et al.*, *Amber 2021*. University of California Press, 2021.

[93] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, *et al.*, "CHARMM: the biomolecular simulation program," *Journal of computational chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.

[94] J. C. Phillips, D. J. Hardy, J. D. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, *et al.*, "Scalable molecular dynamics on CPU and GPU architectures with NAMD," *The Journal of chemical physics*, vol. 153, no. 4, p. 044130, 2020.

[95] U. Börjesson and P. H. Hünenberger, "Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines," *The Journal of chemical physics*, vol. 114, no. 22, pp. 9706–9719, 2001.

[96] A. M. Baptista, V. H. Teixeira, and C. M. Soares, "Constant-pH molecular dynamics using stochastic titration," *The Journal of chemical physics*, vol. 117, no. 9, pp. 4184–4200, 2002.

[97] J. Mongan, D. A. Case, and J. A. McCammon, "Constant pH molecular dynamics in generalized born implicit solvent," *Journal of computational chemistry*, vol. 25, no. 16, pp. 2038–2048, 2004.

[98] M. S. Lee, F. R. Salsbury Jr, and C. L. Brooks III, "Constant-pH molecular dynamics using continuous titration coordinates," *Proteins: Structure, Function, and Bioinformatics*, vol. 56, no. 4, pp. 738–752, 2004.

[99] J. Khandogin and C. L. Brooks, "Constant pH molecular dynamics with proton tautomerism," *Biophysical journal*, vol. 89, no. 1, pp. 141–157, 2005.

[100] Y. Meng and A. E. Roitberg, "Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model," *Journal of chemical theory and computation*, vol. 6, no. 4, pp. 1401–1412, 2010.

[101] S. G. Itoh, A. Damjanović, and B. R. Brooks, "pH replica-exchange method based on discrete protonation states," *Proteins: Structure, Function, and Bioinformatics*, vol. 79, no. 12, pp. 3420–3436, 2011.

[102] J. A. Wallace and J. K. Shen, "Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange," *Journal of chemical theory and computation*, vol. 7, no. 8, pp. 2617–2629, 2011.

[103] S. Donnini, F. Tegeler, G. Groenhof, and H. Grubmüller, "Constant pH molecular dynamics in explicit solvent with λ-dynamics," *Journal of chemical theory and computation*, vol. 7, no. 6, pp. 1962–1978, 2011.

[104] G. B. Goh, J. L. Knight, and C. L. Brooks, "Constant pH molecular dynamics simulations of nucleic acids in explicit solvent," *Journal of chemical theory and computation*, vol. 8, no. 1, pp. 36–46, 2012.

[105] J. A. Wallace and J. K. Shen, "Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH," *The Journal of chemical physics*, vol. 137, no. 18, p. 184105, 2012.

[106] G. B. Goh, B. S. Hulbert, H. Zhou, and C. L. Brooks III, "Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism," *Proteins: structure, function, and bioinformatics*, vol. 82, no. 7, pp. 1319–1331, 2014.

[107] S. Donnini, R. T. Ullmann, G. Groenhof, and H. Grubmüller, "Charge-neutral constant pH molecular dynamics simulations using a parsimonious proton buffer," *Journal of chemical theory and computation*, vol. 12, no. 3, pp. 1040–1051, 2016.

[108] Y. Huang, W. Chen, J. A. Wallace, and J. Shen, "All-atom continuous constant pH molecular dynamics with particle mesh ewald and titratable water," *Journal of chemical theory and computation*, vol. 12, no. 11, pp. 5411–5421, 2016.

[109] B. K. Radak, C. Chipot, D. Suh, S. Jo, W. Jiang, J. C. Phillips, K. Schulten, and B. Roux, "Constant-pH molecular dynamics simulations for large biomolecular systems," *Journal of chemical Theory and Computation*, vol. 13, no. 12, pp. 5933–5944, 2017.

[110] Y. Huang, R. C. Harris, and J. Shen, "Generalized born based continuous constant pH molecular dynamics in amber: Implementation, benchmarking and analysis," *Journal of chemical information and modeling*, vol. 58, no. 7, pp. 1372–1383, 2018.

[111] N. Aho, P. Buslaev, A. Jansen, P. Bauer, G. Groenhof, and B. Hess, "Scalable constant ph molecular dynamics in GROMACS," *Journal of Chemical Theory and Computation*, vol. 18, no. 10, pp. 6148–6160, 2022.

[112] J. A. Harris, R. Liu, V. M. de Oliveira, E. V. Montelongo, J. A. Henderson, and J. Shen, "GPU-accelerated all-atom particle-mesh ewald continuous constant ph molecular dynamics in amber," *bioRxiv*, 2022.

[113] A. M. Baptista, "Comment on "explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines" [J. Chem. Phys. 114, 9706 (2001)]," *The Journal of chemical physics*, vol. 116, no. 17, pp. 7766–7768, 2002.

[114] J. A. Wallace, Y. Wang, C. Shi, K. J. Pastoor, B.-L. Nguyen, K. Xia, and J. K. Shen, "Toward accurate prediction of pKa values for internal protein residues: the importance of conformational relaxation and desolvation energy," *Proteins: Structure, Function, and Bioinformatics*, vol. 79, no. 12, pp. 3364–3373, 2011.

[115] D. Vila-Viçosa, V. H. Teixeira, H. A. Santos, A. M. Baptista, and M. Machuqueiro, "Treatment of ionic strength in biomolecular simulations of charged lipid bilayers," *Journal of Chemical Theory and Computation*, vol. 10, no. 12, pp. 5483–5492, 2014.

[116] S. R. Campos, M. Machuqueiro, and A. M. Baptista, "Constant-ph molecular dynamics simulations reveal a β-rich form of the human prion protein," *The Journal of Physical Chemistry B*, vol. 114, no. 39, pp. 12692–12700, 2010.

[117] N. V. Di Russo, D. A. Estrin, M. A. Martí, and A. E. Roitberg, "ph-dependent conformational changes in proteins and their effect on experimental pkas: the case of nitrophorin 4," *PLoS computational biology*, vol. 8, no. 11, p. e1002761, 2012.

[118] A. Sarkar, P. L. Gupta, and A. E. Roitberg, "pH-dependent conformational changes due to ionizable residues in a hydrophobic protein interior: The study of L25K and L125K variants of snase," *The Journal of Physical Chemistry B*, vol. 123, no. 27, pp. 5742–5754, 2019.

[119] X. Kong and C. L. Brooks III, "λ-dynamics: A new approach to free energy calculations," *The Journal of chemical physics*, vol. 105, no. 6, pp. 2414–2423, 1996.

[120] J. E. Mertz and B. M. Pettitt, "Molecular dynamics at a constant ph," *The international journal of supercomputer applications and high performance computing*, vol. 8, no. 1, pp. 47–53, 1994.

[121] J. A. Wallace and J. K. Shen, "Predicting pKa values with continuous constant ph molecular dynamics," in *Methods in enzymology*, vol. 466, pp. 455–475, Elsevier, 2009.

[122] W. D. Bennett, A. W. Chen, S. Donnini, G. Groenhof, and D. P. Tieleman, "Constant pH simulations with the coarse-grained MARTINI model—application to oleic acid aggregates," *Canadian Journal of Chemistry*, vol. 91, no. 9, pp. 839–846, 2013.

[123] J. Koivisto, X. Chen, S. Donnini, T. Lahtinen, H. Häkkinen, G. Groenhof, and M. Pettersson, "Acid–base properties and surface charge distribution of the water-soluble au102 (p mba) 44 nanocluster," *The Journal of Physical Chemistry C*, vol. 120, no. 18, pp. 10041–10050, 2016.

[124] J. Khandogin, J. Chen, and C. L. Brooks III, "Exploring atomistic details of pH-dependent peptide folding," *Proceedings of the National Academy of Sciences*, vol. 103, no. 49, pp. 18546–18550, 2006.

[125] G. B. Goh, J. L. Knight, and C. L. Brooks III, "ph-dependent dynamics of complex RNA macromolecules," *Journal of chemical theory and computation*, vol. 9, no. 2, pp. 935–943, 2013.

[126] N. Verma, J. A. Henderson, and J. Shen, "Proton-coupled conformational activation of sars coronavirus main proteases and opportunity for designing small-molecule broad-spectrum targeted covalent inhibitors," *Journal of the American Chemical Society*, vol. 142, no. 52, pp. 21883–21890, 2020.

[127] B. W. Zhang, L. Brunetti, and C. L. Brooks III, "Probing ph-dependent dissociation of hdea dimers," *Journal of the American Chemical Society*, vol. 133, no. 48, pp. 19393–19398, 2011.

[128] S. Lammers, S. Lutz, and M. Meuwly, "Reactive force fields for proton transfer dynamics," *Journal of computational chemistry*, vol. 29, no. 7, pp. 1048–1063, 2008.

[129] S. Kale and J. Herzfeld, "Proton defect solvation and dynamics in aqueous acid and base," *Angewandte Chemie International Edition*, vol. 51, no. 44, pp. 11029–11032, 2012.

[130] A. Warshel and R. M. Weiss, "An empirical valence bond approach for comparing reactions in solutions and in enzymes," *Journal of the American Chemical Society*, vol. 102, no. 20, pp. 6218–6226, 1980.

[131] U. W. Schmitt and G. A. Voth, "The computer simulation of proton transport in water," *The Journal of chemical physics*, vol. 111, no. 20, pp. 9361–9381, 1999.

[132] M. A. Lill and V. Helms, "Molecular dynamics simulation of proton transport with quantum mechanically derived proton hopping rates (q-hop md)," *The Journal of Chemical Physics*, vol. 115, no. 17, pp. 7993–8005, 2001.

[133] M. G. Wolf and G. Groenhof, "Explicit proton transfer in classical molecular dynamics simulations," *Journal of computational chemistry*, vol. 35, no. 8, pp. 657–671, 2014.

[134] F. Grünewald, P. C. Souza, H. Abdizadeh, J. Barnoud, A. H. de Vries, and S. J. Marrink, "Titratable martini model for constant ph simulations," *The Journal of chemical physics*, vol. 153, no. 2, p. 024118, 2020.

[135] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics," *Journal of the American Chemical Society*, vol. 112, no. 16, pp. 6127–6129, 1990.

[136] M. E. Davis and J. A. McCammon, "Electrostatics in biomolecular structure and dynamics," *Chemical Reviews*, vol. 90, no. 3, pp. 509–521, 1990.

[137] A. V. Onufriev and D. A. Case, "Generalized born implicit solvent models for biomolecules," *Annual review of biophysics*, vol. 48, pp. 275–296, 2019.

[138] H. Berendsen, J. Grigera, and T. Straatsma, "The missing term in effective pair potentials," *Journal of Physical Chemistry*, vol. 91, no. 24, pp. 6269–6271, 1987.

[139] D. J. Price and C. L. Brooks III, "A modified TIP3P water potential for simulation with ewald summation," *The Journal of chemical physics*, vol. 121, no. 20, pp. 10096–10103, 2004.

[140] J. L. Knight and C. L. Brooks III, "Multisite λ dynamics for simulated structure–activity relationship studies," *Journal of chemical theory and computation*, vol. 7, no. 9, pp. 2728–2739, 2011.

[141] S. Kim, J. Lee, S. Jo, C. L. Brooks III, H. S. Lee, and W. Im, "CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules," 2017.

[142] P. V. Klimovich and D. L. Mobley, "Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations," *Journal of Computer-Aided Molecular Design*, vol. 24, no. 4, pp. 307–316, 2010.

[143] P. Buslaev, N. Aho, A. Jansen, P. Bauer, B. Hess, and G. Groenhof, "Best practices in constant pH MD simulations: Accuracy and sampling," *Journal of Chemical Theory and Computation*, vol. 18, no. 10, pp. 6134–6147, 2022.

[144] P. Dobrev, S. Donnini, G. Groenhof, and H. Grubmüller, "Accurate three states model for amino acids with two chemically coupled titrating sites in explicit solvent atomistic constant pH simulations and pK a calculations," *Journal of chemical theory and computation*, vol. 13, no. 1, pp. 147–160, 2017.

[145] E. J. Arthur and C. L. Brooks III, "Efficient implementation of constant ph molecular dynamics on modern graphics processors," *Journal of computational chemistry*, vol. 37, no. 24, pp. 2171–2180, 2016.

[146] R. C. Harris and J. Shen, "GPU-accelerated implementation of continuous constant ph molecular dynamics in amber: pKa predictions with single-pH simulations," *Journal of chemical information and modeling*, vol. 59, no. 11, pp. 4821–4832, 2019.

[147] M. Frisch *et al.*, "GAUSSIAN16. revision A. 03. gaussian inc., wallingford, CT, USA," 2016.

[148] D. B. Axel, "Density-functional thermochemistry. III. the role of exact exchange," *The Journal of chemical physics*, vol. 98, no. 7, pp. 5648–5652, 1993.

[149] C. Lee, W. Yang, and R. G. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density," *Physical review B*, vol. 37, no. 2, p. 785, 1988.

[150] S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," *Canadian Journal of physics*, vol. 58, no. 8, pp. 1200–1211, 1980.

[151] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields," *The Journal of physical chemistry*, vol. 98, no. 45, pp. 11623–11627, 1994.

[152] U. C. Singh and P. A. Kollman, "An approach to computing electrostatic charges for molecules," *Journal of computational chemistry*, vol. 5, no. 2, pp. 129–145, 1984.

[153] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," *Journal of molecular graphics and modelling*, vol. 25, no. 2, pp. 247–260, 2006.

[154] A. W. Sousa da Silva and W. F. Vranken, "ACPYPE-antechamber python parser interface," *BMC research notes*, vol. 5, no. 1, pp. 1–8, 2012.

[155] J. P. Jambeck and A. P. Lyubartsev, "Another piece of the membrane puzzle: extending slipids further," *Journal of chemical theory and computation*, vol. 9, no. 1, pp. 774–784, 2013.

[156] M. Ramanadham, L. Sieker, and L. Jensen, "Refinement of triclinic lysozyme: Ii. the method of stereochemically restrained least squares," *Acta Crystallographica Section B: Structural Science*, vol. 46, no. 1, pp. 63–69, 1990.

[157] C.-M. Chiang, K.-Y. Chien, H.-j. Lin, J.-F. Lin, H.-C. Yeh, P.-l. Ho, and W.-g. Wu, "Conformational change and inactivation of membrane phospholipid-related activity of cardiotoxin V from taiwan cobra venom at acidic pH," *Biochemistry*, vol. 35, no. 28, pp. 9167–9176, 1996.

[158] C.-M. Chiang, S.-L. Chang, H.-j. Lin, and W.-g. Wu, "The role of acidic amino acid residues in the structural stability of snake cardiotoxins," *Biochemistry*, vol. 35, no. 28, pp. 9177–9186, 1996.

[159] L. Sauguet, F. Poitevin, S. Murail, C. Van Renterghem, G. Moraga-Cid, L. Malherbe, A. W. Thompson, P. Koehl, P.-J. Corringer, M. Baaden, *et al.*, "Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels," *The EMBO journal*, vol. 32, no. 5, pp. 728–741, 2013.

[160] V. Lindahl, J. Lidmar, and B. Hess, "Accelerated weight histogram method for exploring free energy landscapes," *The Journal of chemical physics*, vol. 141, no. 4, p. 044110, 2014.

[161] A. Elofsson, B. Hess, E. Lindahl, A. Onufriev, D. Van der Spoel, and A. Wallqvist, "Ten simple rules on how to create open access and reproducible molecular simulations of biological systems," 2019.

ORIGINAL PAPERS

I

EXPLORING FAST PROTON TRANSFER EVENTS ASSOCIATED WITH LATERAL PROTON DIFFUSION ON THE SURFACE OF MEMBRANES

by

Nadav Amdursky, Yiyang Lin, Noora Aho, Gerrit Groenhof. *Proceedings of the National Academy of Sciences, 6,(7),* 2443-2451, **2019.**

DOI: 10.1073/pnas.18123511169

Exploring fast proton transfer events associated with lateral proton diffusion on the surface of membranes

Nadav Amdursky[a,1], Yiyang Lin[b], Noora Aho[c], and Gerrit Groenhof[c]

[a]Schulich Faculty of Chemistry, Technion – Israel Institute of Technology, Haifa 3200003, Israel; [b]Department of Materials, Imperial College London, SW7 2AZ London, United Kingdom; and [c]Department of Chemistry and NanoScience Center, University of Jyväskylä, FIN-40014, Jyväskylä, Finland

Edited by Harry B. Gray, California Institute of Technology, Pasadena, CA, and approved December 19, 2018 (received for review July 18, 2018)

Proton diffusion (PD) across biological membranes is a fundamental process in many biological systems, and much experimental and theoretical effort has been employed for deciphering it. Here, we report on a spectroscopic probe, which can be tightly tethered to the membrane, for following fast (nanosecond) proton transfer events on the surface of membranes. Our probe is composed of a photoacid that serves as our light-induced proton source for the initiation of the PD process. We use our probe to follow PD, and its pH dependence, on the surface of lipid vesicles composed of a zwitterionic headgroup, a negative headgroup, a headgroup that is composed only from the negative phosphate group, or a positive headgroup without the phosphate group. We reveal that the PD kinetic parameters are highly sensitive to the nature of the lipid headgroup, ranging from a fast lateral diffusion at some membranes to the escape of protons from surface to bulk (and vice versa) at others. By referring to existing theoretical models for membrane PD, we found that while some of our results confirm the quasi-equilibrium model, other results are in line with the nonequilibrium model.

proton diffusion | excited-state proton transfer | lipid vesicles | photoacid | molecular dynamics

The translocation of protons across the two sides of biological membranes by transmembrane proteins is a fundamental biochemical process, such as the transport of protons in the aerobic respiration system in the mitochondria of eukaryotes or in the chloroplast of plants and prokaryotes during photosynthesis (1, 2). In this proton translocation process, protons arrive at the entrance of the protein-based proton pump either from the membrane or from the bulk solution. The diffusion of protons in bulk aqueous solution is highly efficient, with a diffusion coefficient (D_{H^+}) of $D_{H^+} = 9.3 \times 10^{-5}$ cm^2·s^{-1}, as commonly explained by the Grotthuss mechanism, which describes a hydrogen bond-assisted proton hopping between water molecules (3, 4). However, the diffusion of protons along the surface of the membrane toward the entrance of the pump is still under investigation. The protein structure of proton pumps suggests that a protein structural motif in the form of ionizable side chains of amino acids is responsible for shuttling the protons from the surface of the membrane to the protein pump (1, 2, 5). It was established more than 20 years ago that proton transfer (PT) from the surface of membranes to bulk water (and vice versa) is different from PT events in bulk water (6–10). Proton diffusion (PD) along a membrane surface is more complicated than bulk PD since the structure of the lipid molecule, the distance between adjacent lipid molecules, and the water molecules in and around the lipid membrane all affect PD. Due to advances in fast spectroscopy experiments and molecular dynamics (MD) simulations, significant progress in our understanding of this type of PD has been made in the past decade (11–23).

In a pioneering type of experiment, the groups of Pohl and coworkers (11–14) and Widengren and coworkers (15, 16) have explored the role of the lipid and bulk proton concentration on PD along membranes. Widengren and coworkers (15, 16) have used a fluorescent chromophore that was tethered to the membrane surface; by fluorescence correlation spectroscopy, they followed the protonation of the chromophore. Their main conclusion was that the lipids act as proton-collecting antennas from the bulk solution. They further found that (*i*) incorporating phosphatidic acid (PA) in vesicles of phosphatidylcholine (PC) or phosphatidylglycerol (PG) caused an increase in PD (16) and (*ii*) both the ionic strength of the solution and the mode of protonation of the proton-collecting antennas influence the PD parameters (15). The lateral PD coefficient on the membrane was indirectly estimated by the authors while assuming a proton hopping between adjacent lipid molecules (with a distance of 1 nm), and resulted in a diffusion coefficient of $D_{H^+} = 2 \times 10^{-7}$ cm^2·s^{-1}. Pohl and coworkers used a similar chromophore as in the latter experiments, but the release of protons was either photoinitiated from a proton release chromophore that was situated dozens of micrometers from the probed chromophore (11, 13) or by proton injection a few micrometers from the detection site (12, 23). This experimental approach can only be used with planar membranes (instead of vesicles), but it enables the direct calculation of the diffusion coefficient from the kinetic curves. The authors' main conclusions were (*i*) PD along a PC membrane was similar to the one along a phosphatidylethanolamine (PE) membrane sharing a diffusion coefficient of ~4×10^{-5} cm^2·s^{-1}; (*ii*) surprisingly, the diffusion along a glycerol monooleate (GMO) membrane was similar to the PC and PE membranes; and (*iii*) PD is most probably due to protons in the second layer of water around the membrane.

Here, we explore PD along the surface of lipid vesicles by following the light-initiated proton release from a membrane-anchored photoacid, which dissociates upon light excitation due to the large difference in pK_a value between its ground and excited states. Using time-resolved fluorescence enables us to extract the PT efficiency to and from the photoacid, the PD coefficient, and

Significance

Proton diffusion (PD) across biological membranes is a fundamental process in many biological systems, such as in aerobic respiration or in the photosynthesis process. Even though this diffusion process has been studied for many decades, it is still highly enigmatic. Here, we suggest an experimental technique to study the very fast PD events taking place upon the photoinduced release of a proton on the surface of the membrane. We show that the diffusion, as well as the interaction between protons on the surface and protons in bulk media can be influenced by the nature of the membrane's lipids. Our results shed some light on this unique PD process and help to overcome previous experimental contradictions in the literature.

Author contributions: N. Amdursky, Y.L., and G.G. designed research; N. Amdursky, Y.L., and N. Aho performed research; N. Amdursky, Y.L., N. Aho, and G.G. contributed new reagents/analytic tools; N. Amdursky, N. Aho, and G.G. analyzed data; and N. Amdursky wrote the paper.

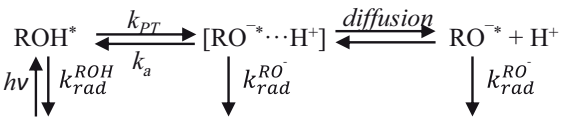
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: amdursky@technion.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1812351116/-/DCSupplemental.

Published online January 24, 2019.



Scheme 1. ESPT photoprotolytic cycle of photoacids. *hν*, photon absorption; *rad*, radiative decay.

the dimensionality of the diffusion. We explore four types of membranes with different phospholipid headgroups: a zwitterionic headgroup (PC), a negative headgroup (PG), a headgroup that is composed only of the negative phosphate group (PA), and a positive headgroup without the phosphate group [trimethylammonium-propane (TAP)] (24). The use of photoacids to follow proton dynamics on the surface of lipid membranes started in the pioneering works of Nachliel and Gutman (25) nearly 30 years ago. In contrast to that study, in which the photoacid was solvated in solution, we have tightly tethered the photoacid to the surface of the lipid, and thus follow only protons that have been released on the surface.

Theoretical Model

Photoacids (Brønsted type) are aryl-OH molecules, whose pK_a values differ between ground and excited states. For instance, the common photoacid of 8-hydroxy-1,3,6-pyrenetrisulfonate (HPTS) has a pK_a value of 7.4 in the ground state and a pK_a* value of 1.3 in the excited state (26). Accordingly, the molecule dissociates during the excited-state lifetime, and the excited-state proton transfer (ESPT) photoprotolytic cycle can be described as follows.

Following light excitation of the protonated species (ROH), the excited molecule (ROH*) deprotonates with a PT rate of k_{PT}. The proton can then either recombine with the anion (RO^{-*}), with a recombination rate of k_a, or diffuse along the media, which can be described by the Debye–Smoluchowski equation (27). The arrows pointing down in Scheme 1 represent the radiative decay of the ROH* and RO^{-*}, which share a similar radiative lifetime, τ, of ~5.4 ns, which enables following the PD for at least 30 ns using time-resolved experimental methods. The fluorescence intensity and decay of the ROH* and RO^{-*} forms can be easily followed individually since each form emits at a different wavelength. Photoacids have been used to follow the hydration of biological surfaces of proteins (26, 28–30) and polysaccharides (26, 31, 32). The photoprotolytic cycle of photoacids (Scheme 1) has been theoretically modeled by Agmon and coworkers (27, 33–35), who suggested the following expression for the ROH* fluorescence decay:

$$I_f^{ROH^*} * \exp\left(-t/\tau\right) \cong \frac{\pi a^2 k_a \exp[-R_D/a]}{2k_{PT}(\pi D_{H^+} t)^{d/2}}. \quad [1]$$

This model describes an ESPT process occurring on a surface of a reaction sphere with a radius of a, where d is the dimensionality of the PD and R_D is the Debye radius. In their model, R_D represents the distance at which the coulombic attraction between the negative excited-state anion (RO^{-*}) and the positive proton equals the thermal energy (k_BT):

$$R_D = \frac{|Z_1 Z_2| e^2}{4\pi\varepsilon_0 \varepsilon_r k_B T}, \quad [2]$$

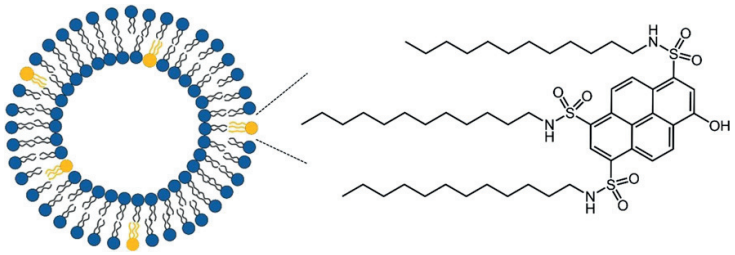
where Z_1 and Z_2 are the electron charge (e) units of RO^{-*} and the dissociated proton, ε_0 is the permittivity of free space, and ε_r is the relative dielectric permittivity of the medium. The relative dielectric permittivity of the water layer surrounding the lipid vesicle is hard to estimate, and we have used an averaged value of 20–27 for the first nanometers around the lipid vesicles as was found experimentally and theoretically (36, 37). Accordingly, the calculated Debye radius for our membrane-bound modified photoacid (Scheme 2) is R_D = 20–28 Å. HPTS in solution has a fourfold higher Z_1 value (its RO^{-*} form has four electron charges), but due to the high relative dielectric permittivity of water (ε_r = 80), HPTS has a value of R_D = 28 Å.

Results

For tethering the photoacid to the surface of a lipid vesicle, we synthesized a hydrophobic photoacid where a dodecyl group has been attached to each of the sulfonate groups of HPTS (C_{12}-HPTS; Scheme 2). Due to the hydrophobic nature of our newly synthesized photoacid, it could not be dissolved in water. Comparing the steady-state emission of HPTS and C_{12}-HPTS in ethanol and methanol has allowed us to conclude that C_{12}-HPTS can be considered as a stronger photoacid than HPTS (*SI Appendix*, Fig. S1 and further discussion). A stronger photoacid is more likely to undergo ESPT. Our finding is in line with the works of Jung and coworkers (38, 39), who observed an increase in the photoacid strength upon modifications of the HPTS sulfonic groups.

C_{12}-HPTS was tethered to the membrane (at a ratio of 1:100 probe/lipid; *Materials and Methods*), which was composed of the following lipids: 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-1′-*rac*-glycerol (POPG), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphate (POPA), or 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP). We have used dynamic light scattering to quantify the size distribution of the different formed vesicles, showing similar distribution of our different vesicles (*SI Appendix*, Fig. S2).

MD Study. Our working hypothesis here is that our newly designed probe can integrate into the membrane while exposing its hydroxy group toward the membrane surface (i.e., in the water membrane interface) and release a proton upon light irradiation. To test this hypothesis, we performed atomistic MD simulations of the C_{12}-HPTS incorporated in lipid bilayers of POPC, POPG, and POPA. The incorporation in DOTAP vesicles was not simulated since, as will be discussed below, ESPT from C_{12}-HPTS in



Scheme 2. Molecular scheme of C_{12}-HPTS incorporated into a lipid vesicle.

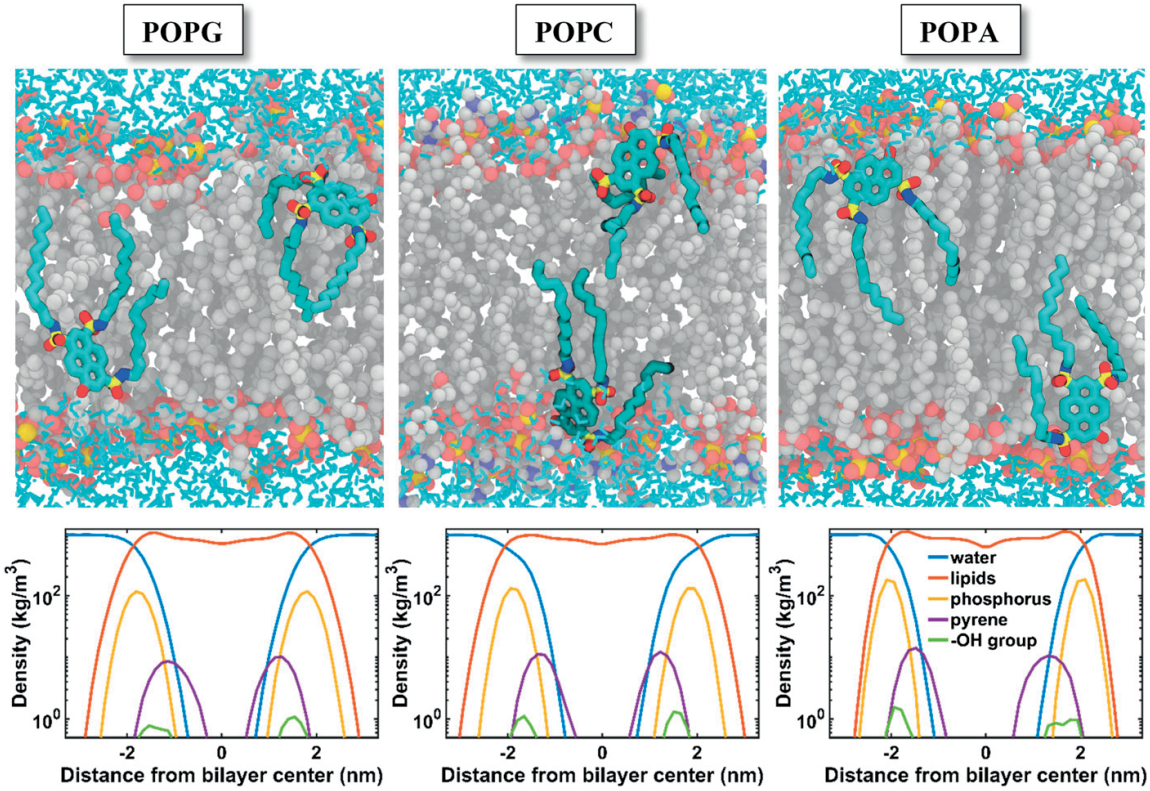


Fig. 1. (*Top*) Snapshots from atomistic MD trajectories showing the position of the C_{12}-HPTS dye in the upper and lower leaflets of the three phospholipid bilayers (water molecules are shown as cyan sticks). (*Bottom*) Density distributions along the bilayer normal of the water (blue), lipids (red), phosphorus atoms of the lipid headgroups (yellow), C_{12}-HPTS pyrene moieties (violet), and C_{12}-HPTS hydroxy groups (green).

DOTAP could not be followed in physiological aqueous conditions. The molecular snapshots of the MD simulations (Fig. 1, *Top*) show that C_{12}-HPTS is well integrated within the membrane, exposing its −OH toward the water–membrane interface, in close proximity to the location of the phosphate groups, as can be observed in the density plots (Fig. 1, *Bottom*). The MD simulations also allowed us to estimate whether the −OH of the photoacid can release a proton by following the ability of the photoacid's hydroxy group to form hydrogen bonds with nearby water molecules or phosphate groups (Table 1). The table shows that the −OH can form hydrogen bonds with nearby moieties for the vast majority of the simulation time. Furthermore, the ratio between the number of hydrogen bonds to water molecules and the number of hydrogen bonds to nearby phosphate groups depends on lipid, with the maximum number of hydrogen bonds to water for the POPA membrane. As will be discussed later (see *Discussion*), these results support our spectroscopic measurements.

Steady-State Spectroscopy. Steady-state absorption (Fig. 2) studies of the photoacid-containing lipid vesicles allowed us to verify that C_{12}-HPTS has been tethered to the membrane and to determine the pK_a of the photoacid by ratiometric titration with a strong acid/base. Fig. 2 and Table 2 show that the pK_a of the photoacid is highly sensitive to the charge of the lipid headgroup (molecular schemes in Fig. 2). We found that the pK_a in the negatively charged headgroup vesicles (POPG and POPA) is higher than in the zwitterionic headgroup vesicle (POPC), while the pK_a in the positively charged headgroup vesicles (DOTAP) is extremely low. It has been shown that the pK_a of HPTS in aqueous solutions can be changed as a function of salt concentration (40) at different water/organic solvent ratios (41) or next to a protein surface (30),

Table 1. Percentage of time that a hydrogen bond is formed between the −OH group of C_{12}-HPTS and the surrounding water molecules or lipid phosphate groups

Lipid system	Hydrogen bonds with water molecules, %	Hydrogen bonds with lipid phosphates, %	Total hydrogen bonds with everything, %
POPG	31.9	65.9	97.8
POPC	27.1	71.3	98.4
POPA	37.4	56.2	93.6

The values are averaged between the upper and lower membrane leaflets in the MD simulations.

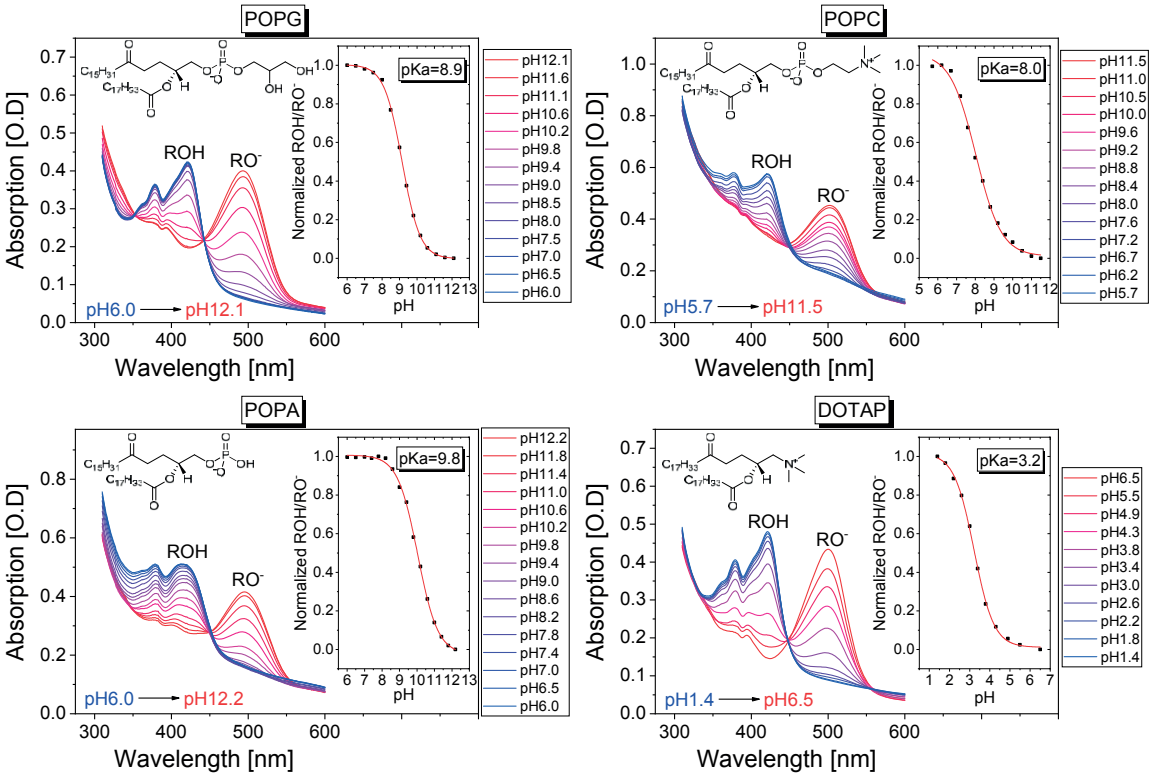


Fig. 2. Absorption spectrum of C_{12}-HPTS incorporated in the different lipid vesicles at different pH values. (*Insets*) pK_a calculations and the molecular schemes of the lipid. O.D, optical density.

but, as far as we are aware, such drastic change in the pK_a value of a given photoacid has never been observed before. Our latter finding further suggests that the photoacid is in tight proximity to the headgroup of the lipid, in line with the results from our MD simulation. In a similar manner, Widengren and coworkers (15) also observed an increase in the pK_a values of fluorescein and Oregon green when these photoacids were tethered to PG and PC vesicles. We have used the absorption and emission maxima (Table 2) to calculate the ΔpK_a values ($\Delta pK_a = pK_a - pK_a^*$) of the photoacid at different lipid vesicles using the Förster cycle (42):

$$\Delta pK_a = \frac{\Delta E}{\ln(10)RT}, \qquad [3]$$

where ΔE is the energy difference between the protonated (ROH) and deprotonated (RO^-) forms of the photoacid, R is the gas constant, and T is temperature. To calculate ΔE, we have used the averaged values for the absorption (E_{Ab}) and emission (E_{Em}) maxima to compensate the solvent contribution to the Stokes' shifts (43):

$$\Delta E = \left(E_{Ab}^{ROH} + E_{Em}^{ROH}\right)/2 - \left(E_{Ab}^{RO^-} + E_{Em}^{RO^-}\right)/2. \qquad [4]$$

For following the light-induced proton release from photoacids, they have to be excited in their ROH form. Accordingly, the above-calculated pK_a values limit the pH range of the measurement, in which only low pH values could be used for DOTAP vesicles. Fig. 3 shows the normalized emission spectrum with different lipid vesicles and at different pH values, where several important observations can be made. The first one is that for all of our measurements, the main form of the molecule is RO^-, meaning that the photoacid can dissociate and release a proton. This is a further validation that the photoacid is on the surface of the lipid vesicle, since if it were buried within the hydrophobic membrane, we would expect to see the ROH peak as the main one, as was previously shown for HPTS within hydrophobic binding sites (26).

The second important observation is that the emission profile is different for different lipid vesicles, which indicates that the lipid headgroup has a major role in determining the ESPT parameters of the photoprotolytic cycle of C_{12}-HPTS (Scheme 1). We can use the RO^-/ROH ratio of the emission spectrum at different pH values (*SI Appendix*, Table S1) as a first indication for the role of the lipid headgroup in ESPT to/from the photoacid and for PD. In general, high RO^-/ROH ratios (low-intensity ROH peak; Fig. 3) can be a result of a fast PT rate, a slow recombination rate, or a fast PD

Table 2. Absorption and emission maxima of ROH and RO^-, together with the calculated pK_a, ΔpK_a, and pK_a^* values

System	POPG	POPC	POPA	DOTAP
λ_{abs}^{ROH}, nm	419	420	411	422
$\lambda_{abs}^{RO^-}$, nm	501	502	500	501
λ_{em}^{ROH}, nm	475	476	473	459
$\lambda_{em}^{RO^-}$, nm	543	543	541	544
pK_a	8.9	8.0	9.8	3.2
ΔpK_a	7.0	6.9	7.5	7.6
p$K_a^{*\dagger}$	1.9	1.1	2.3	−4.4

†The values were estimated by $pK_a^* = pK_a - \Delta pK_a$.

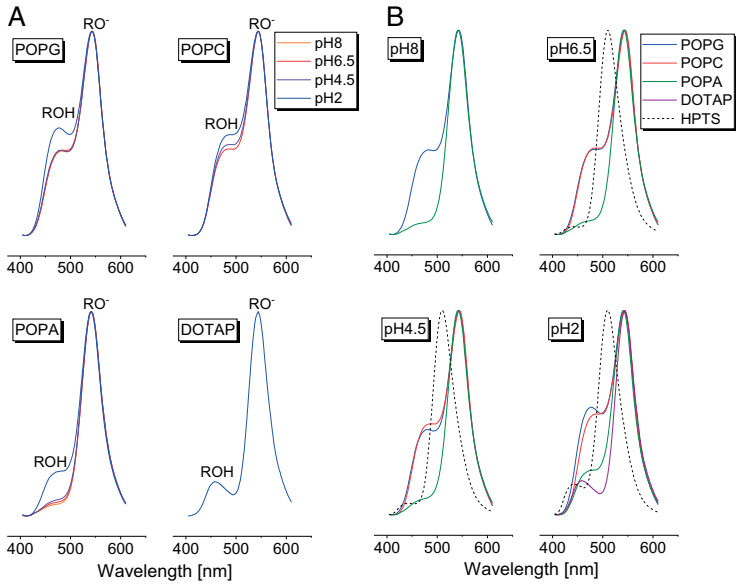


Fig. 3. Emission spectrum of C_{12}-HPTS incorporated in the different lipid vesicles at different pH values, below the pK_a of the photoacid. (*A*) Spectrum for each of the lipid vesicles as a function of pH. (*B*) Spectrum at each of the pH values as a function of the lipid vesicle. The dashed lines in *B* represent the emission spectrum of HPTS.

coefficient, and vice versa, while only with transient fluorescence can we distinguish between them (*vide infra*). The RO^-/ROH ratio and the emission peak positions of C_{12}-HPTS in POPC and POPG are very similar. This means that although the surface charge is different (POPG is negatively charged, while POPC is zwitterionic), the ESPT parameters on their surface are similar. The higher RO^-/ROH ratios of the less bulky POPA (negatively charged) and DOTAP (positively charged) lipid vesicles imply some role of the lipid headgroup bulkiness in PD; however, as will be discussed below (see *Discussion*), the cause for it is very different.

Time-Resolved Fluorescence. Following our steady-state measurements that confirmed the ability of C_{12}-HPTS to release a proton, and already suggested that PD is different on the surface of lipid vesicles composed of different headgroups, we have turned to time-resolved fluorescence measurements. Figs. 4 and 5 show the fluorescence time-resolved decay of the ROH* form at different lipid vesicles and at different pH values, respectively, in comparison to the decay of HPTS (the decay of HPTS at different pH values is shown in *SI Appendix*, Fig. S3). Qualitatively, we can divide the decay profile of ROH* into three components (Table 3), where the first fast component is directly related to the PT from the photoacid (k_{PT}), which takes place in the first nanosecond, and the other slower components are indicative of the PD and the recombination rate (k_a). Following PT, the proton can either diffuse or recombine with the RO^{-*} (Scheme 1), where the PD efficiency and dimensionality influence mainly the 1- to 6-ns regime after excitation, and the recombination can be associated with a long-lived (high-intensity) exponential tail that can span to tens of nanoseconds (27, 34). This qualitative understanding of the ROH* decay profile can be nicely explained by referring to free HPTS in solution at different pH values (Table 3 and *SI Appendix*, Fig. S3 and text within). Accordingly, we can divide the decay profile of C_{12}-HPTS (Figs. 4 and 5 and Table 3) into three different profiles:

i) The decays of POPG and POPC have a similar shape and around the same intensities, with a much slower k_{PT} (τ_1) compared with HPTS in bulk solution (Fig. 5), and their change in decay as a function of pH (both τ_{1-3} and a_{1-3}) is relatively moderate (Fig. 4).

ii) The decay of POPA is very different from that of POPG and POPC, but it has some resemblance to HPTS, both in its fast k_{PT} (τ_1) and especially in the presence and magnitude of the exponential long-lived fluorescent tail (a_3 and τ_3) at low pH.

iii) DOTAP, with its single measurement at low pH, has a unique decay profile, different from all other measurements, with a τ_1 between the ones of POPA and POPC/POPG and a relatively "small" long-lived fluorescent tail (τ_3).

For trying to quantify the different PT rate constants, diffusion coefficient, and dimensionality of the PD, we have used the above theoretical approach, utilizing a software package developed by Krissinel and Agmon (44) for solving diffusion problems, and specifically ESPT from photoacids [spherical symmetric diffusion problem (SSDP) program, version 2.66]. Since the model was not designed for highly concentrated proton solutions (low pH), manifested as a high-intensity, long-lived exponential fluorescent tail (as discussed above), we have used it only for the results obtained at pH 6.5, and thus did not fit the decay obtained with the DOTAP vesicles (fits are displayed in *SI Appendix*, Fig. S4). Below, we summarize our results (Table 4):

i) The PT rate constant, k_{PT}, on the surface of all lipid vesicles is lower than the one for HPTS in bulk water, whereas the value for POPA is the highest among all other vesicles.

ii) The proton recombination rate constant, k_a, for POPG and POPC is much higher than the one for HPTS in water, while the one for POPA is closer to the one of HPTS.

iii) The PD coefficient (D_{H^+}) from the photoacid on the lipid vesicles is reduced from the high proton mobility in bulk water and has different values for different lipid vesicles. POPA exhibits a very close value to HPTS, and POPG and POPC have a similar lower coefficient.

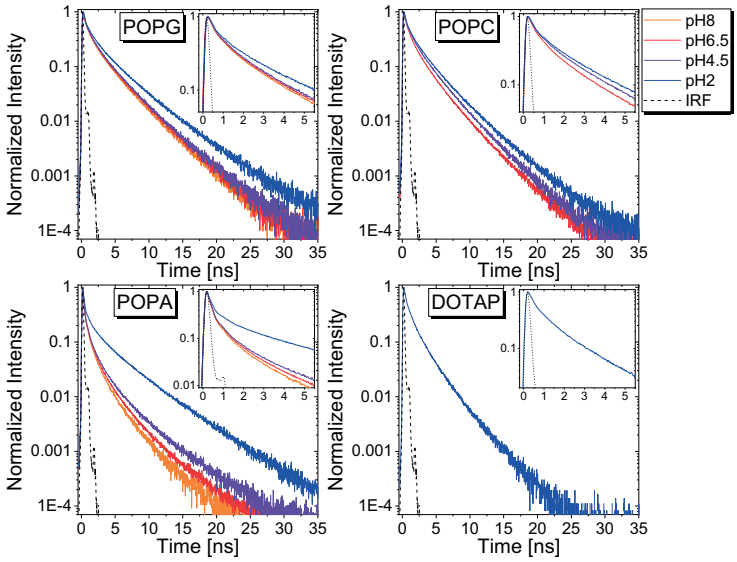


Fig. 4. Time-resolved fluorescence decay of the ROH* form in the four lipid vesicles as a function of pH value. (*Insets*) Zoomed-in views of the first nanoseconds. IRF, instrument response function.

iv) The dimensionality (d) of the PD from the photoacid on the lipid vesicles is reduced from the dimensionality value of 3 of HPTS in bulk water. POPC and POPG have a dimensionality value of 2, and POPA has a higher dimensionality value of 2.5.

v) The R_D values differ between different lipid vesicles, indicating that the dielectric constant next to the lipid vesicle changes for different lipid headgroups.

Discussion

According to the time-resolved fluorescence measurements and their fitting, we can divide the PD behavior into three different types of diffusion patterns.

The first one relates to the surfaces of POPC and POPG, which, despite a different surface charge, share the same PT and PD properties. They both exhibit lateral (2D) PD on the surface of the membrane, with a coefficient of 3.5×10^{-5} $cm^2 \cdot s^{-1}$, nearly

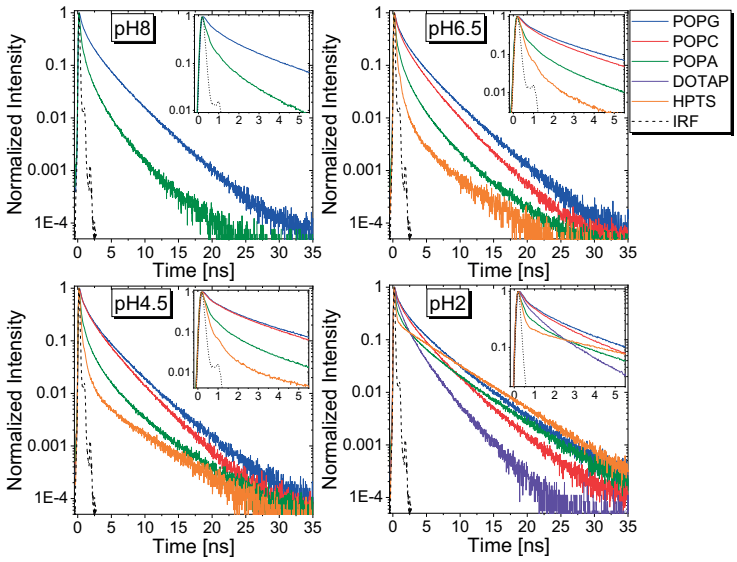


Fig. 5. Time-resolved fluorescence decay of the ROH* form at each of the pH values as a function of the lipid vesicle, in comparison to the decay of HPTS. (*Insets*) Zoomed-in views of the first nanoseconds. IRF, instrument response function.

Table 3. Fitting parameters of a three-exponent decay function for C_{12}-HPTS fluorescence in different lipid vesicles and at different pH values and for HPTS

System	pH	a_1	τ_1, ns	a_2	τ_2, ns	a_3	τ_3, ns
POPG	8.0	0.45	0.42	0.37	1.48	0.18	3.14
	6.5	0.47	0.46	0.35	1.55	0.18	3.25
	4.5	0.45	0.45	0.36	1.55	0.19	3.25
	2.0	0.43	0.51	0.36	1.77	0.21	3.78
POPC	6.5	0.46	0.41	0.38	1.34	0.16	2.88
	4.5	0.43	0.40	0.38	1.51	0.19	2.93
	2.0	0.44	0.42	0.36	1.67	0.20	3.21
POPA	8.0	0.70	0.14	0.26	0.86	0.04	2.58
	6.5	0.69	0.14	0.26	0.88	0.05	2.76
	4.5	0.69	0.15	0.26	0.93	0.05	3.16
	2.0	0.59	0.16	0.27	1.48	0.14	4.22
DOTAP	2.0	0.52	0.36	0.35	1.24	0.13	2.52
HPTS	6.5	0.82	0.13	0.17	0.55	0.006	3.55
	4.5	0.82	0.14	0.17	0.57	0.006	4.14
	2.0	0.72	0.13	0.15	2.53	0.13	5.21

The preexponential factors (a) have been normalized to $a_1 + a_2 + a_3 = 1$.

threefold slower than the one for protons in bulk water. The PT from the photoacid on the surface of these vesicles is also slower, with a k_{PT} that is more than threefold lower than the one of HPTS in bulk water (also τ_1 in Table 3). The k_a value for these lipid surfaces is around twofold larger than the one of HPTS (also τ_3 in Table 3). The high proton recombination rate on the surface of POPC and POPG can be explained by the constrained lateral PD that increases the possibility for a recombination event between a proton and the excited anion species (RO^{-*}). Interestingly, while τ_3 and its magnitude (a_3) are only weakly influenced by the pH of the solution, they change dramatically for HPTS in bulk water. This indicates a limited influence of bulk protons on the PD properties of these membranes.

The second pattern is observed for the negatively charged surface of POPA. The PD from the photoacid exhibits a mixed dimensionality of ~2.5, meaning that unlike the surfaces of POPG and POPC, the protons on the surface of POPA do not diffuse laterally, but also escape to bulk water. Indeed, all of the parameters for POPA have intermediate values between the ones for HPTS in bulk water and the ones obtained for POPC and POPG. POPA has the largest PD coefficient, the fastest PT rate, and the slowest recombination rate among the three lipids (at pH 6.5). Also, the change in the long-lived fluorescent tail magnitude (τ_3 and especially a_3) is strongly modulated by the pH, similar to HPTS, which suggests that the tethered photoacid in POPA vesicles is much more exposed to bulk solution than in the other lipid vesicles. This is in line with the results of the MD simulations, which also suggest that the photoacid's hydroxy group is more solvent-exposed in POPA than in POPG or POPC (Table 1).

The third pattern relates to the positively charged surface of DOTAP. Due to the very low pK_a of C_{12}-HPTS in DOTAP, we could measure its ESPT only at a very low pH value. The main conclusion is the relative absence of the long-lived fluorescent tail (τ_3) and the relatively slow PT (τ_1). This type of pattern is common in photoacids within binding sites of proteins (28, 45), which might imply that the hydration next to DOTAP vesicles is restricted (more hydrophobic) compared to the other vesicles.

Finally, we will try to correlate our study to previous results and available PD mechanisms. Membrane PD can be followed for different time regimes depending on what experimental approach is used. Pohl and coworkers (11–14) probed PD on the millisecond-second regime, while Widengren and coworkers (15, 16) focused on the microsecond-millisecond regime. Accordingly, our results complement the fast time regime for PD, where we focus on fast nanosecond diffusion processes. Nevertheless, our results share some similarities with the slower type of measurements and can help in resolving the discrepancies in the literature. Widengren and coworkers (15, 16) followed the protonation of a lipid-tethered dye, which bears some resemblance to our approach. They found that PT on the surface of DOPG and DOPC vesicles is faster than proton exchange between surface and bulk water. This finding is in line with the high recombination rates on the surfaces of POPG and POPC observed in our work. They further found that the addition of DOPA (to DOPG vesicles) increases the protonation of their chromophore, particularly at low pH. Our observation that POPA is highly accessible to protons in solution partially supports this finding. Nevertheless, one of their main conclusions is that the surface of the membrane acts as a proton-collecting antenna, which is not supported by our measurements for POPG, POPC, and DOTAP, but can be considered compatible with our POPA results. Moreover, led by that conclusion, they estimated a low PD coefficient (~10^{-7} $cm^2 \cdot s^{-1}$) based on the distance between titratable groups in the lipid structure. According to our measurements, which enable us to estimate the diffusion coefficient directly from the data, this low value should be revised. Pohl and coworkers (11, 13) have used a similar chromophore as the latter studies, but their conclusion was different. Their experimental setup allowed them to calculate the PD coefficient directly from their kinetic curves, where they showed that all of the different membranes they used, namely, PC, PE, and GMO, shared the same diffusion coefficient on the order of 5×10^{-5} $cm^2 \cdot s^{-1}$. This latter value is of the same magnitude as our derived values, even though we followed PD on the nanosecond scale, while Pohl and coworkers (11, 13) followed diffusion on the millisecond-second regime. Because the GMO membrane shares the same glycerol termination as PG membranes, the similarity in the diffusion parameters in their work can be nicely correlated to our finding of similar diffusion parameters between POPC and POPG. They concluded that PT on the membrane surface cannot involve proton hopping between titratable groups. More recently, they suggested that proton wires on the surface of the membrane can facilitate PT (13). Our results confirm that PD across POPG and POPC cannot be explained by proton hopping across titratable groups, but we do show that POPA vesicles can act as proton-collecting antennas, while DOTAP vesicles cannot support high PD.

Two main PD mechanisms have been proposed to explain the membrane PD process (13, 18, 21, 46, 47). The first mechanism suggested by Medvedev and Stuchebrukhov (18, 21, 47) discusses the coupling between bulk diffusion and surface diffusion. They claim that protons can go from bulk to surface (the "antenna effect," where protons populate titratable groups) with a rate constant of k_{on}, and from surface to bulk with a rate constant of k_{off}. They further distinguish between two regimes, the slow- and

Table 4. Summary of the PT kinetic parameters (k_{PT} and k_a), proton diffusion coefficient (D_{H^+}), dimensionality (d), and R_D for C_{12}-HPTS in different lipid vesicles (at pH 6.5) as obtained using the SSDP program

System	k_{PT}, ns^{-1}	k_a, Å·ns^{-1}	D_{H^+}, $cm^2 \cdot s^{-1}$	d	R_D, Å
POPG	2.6	18.5	3.5×10^{-5}	2.0	20
POPC	2.8	17.0	3.5×10^{-5}	2.0	21
POPA	5.5	11.2	8.5×10^{-5}	2.5	26
HPTS	8.9	8.5	9.0×10^{-5}	3.0	28

To limit the degrees of flexibility of this model, we have fixed the value for the radius of the reaction sphere, a, to 6 nm for all fittings, which is a common value for the HPTS photoacid. τ (Eq. **1**) was set to 5.4 ns for HPTS and to 5.7 ns for C_{12}-HPTS, as determined from our lifetime measurements. SSDP, spherical symmetric diffusion problem.

fast-exchange regimes, which refer to the rate of proton exchange between surface and bulk. They suggest that the fast regime is the most probable one, and that long-distance PD processes that take place on the millisecond-second time scale can be explained by numerous k_{on}/k_{off} events that lead to a quasi-equilibrium state between surface and bulk protons. The second mechanism proposed by Agmon and coworkers (13) and Agmon and Gutman (46) suggests that protons on the surface are not in equilibrium with bulk protons. They further propose that the membrane is not acting as a proton-collecting antenna and that protons do not escape from surface to bulk (i.e., protons have very low k_{off} values). They suggest that protons on the surface diffuse via proton wires in what they envision as a propagation of an advancing front of protons, where protons at the front protonate all possible titratable sites, and that the protons following that front diffuse in an unhindered fashion. Both theories have been developed to explain results observed in the (sub)second regime. Our measurements follow only diffusion events taking place in the excited state and, accordingly, are limited to the nanosecond regime. Nevertheless, we believe that our results can shed insights on the possible fast kinetics of this diffusion process.

As discussed, our results with POPA vesicles are very different from the ones of POPC and POPG. The strong influence of the solution pH on the kinetic parameters observed for the POPA vesicles suggests that protons released on the surface of the membrane can escape to bulk, and that during the excited-state lifetime, protons from the bulk can go to the surface, which supports the quasi-equilibrium model. It is important, however, to mention that PA lipids are frequently referred to as "buffered lipids" (i.e., due to their bare phosphate group, they act as buffer molecules in solution, and can therefore be easily titratable). On the other hand, our POPC and POPG results suggest PD with limited surface-to-bulk and bulk-to-surface PT events, which, in combination with a relatively high PD coefficient, support the nonequilibrium model, as well as disproving the possibility of diffusion by proton hopping between titratable groups (11). Nevertheless, since we observed some pH dependency of the kinetic parameters, the protons cannot be completely nonequilibrated with the solution, and there is probably some bulk-to-surface PT. Thus, while our results seem to support the nonequilibrium model, we cannot ascribe all of our results to a single suggested PD mechanism, but hope that our results will ignite new theoretical efforts toward a unifying approach for PD on the surface of membranes.

Conclusions

In summary, we have developed a photoacid-based probe that can be tethered to membranes for studying light-induced PD processes on the surface of membranes. While the photoacid enables the exploration of fast PT events within the PD process, the time interval in which PD can be studied is limited to the excited-state lifetime of the probe, which is on the order of a few dozen nanoseconds. As a consequence, the accessible length scales are also limited to few dozen nanometers (48). We used our photoacid probe to explore PD across phospholipid membranes with different lipid headgroups, namely, PA, PC, PG, and TAP, as well as the effect of pH on the PD process. We found that protons diffuse laterally on the surface of PC and PG vesicles with a diffusion coefficient of 3.5×10^{-5} $cm^2 \cdot s^{-1}$, which cannot be explained by proton hopping between titratable groups, with only a moderate pH dependence on the kinetic parameters. For PA vesicles, we found that protons could escape to bulk, and vice versa (the antenna effect), which also leads to a more prominent effect of the solution pH on the kinetic parameters. For TAP vesicles, we found a slower PT from the photoacid with little influence of bulk protons, which resembles the behavior of hydrophobic pockets.

Materials and Methods

Synthesis of C_{12}-HPTS Derivative. To a mixture of dodecanamine (0.5 mmol) and triethanolamine (0.5 mmol) in 10 mL of dichloromethane (0 °C), 0.1 mmol of 8-acetoxypyrene-1,3,6-trisulfonyl chloride (38) in 10 mL of dichloromethane was added dropwise. After warming to room temperature, the reaction mixture was stirred for another 48 h. The organic phase was washed with sodium bicarbonate and saturated sodium chloride solution before being dried over sodium sulfate. After evaporation, the crude product was purified via column chromatography (3:7 ethyl acetate/petrol ether). The product was confirmed by MALDI-TOF.

Preparation of Lipid Vesicles. All lipids (POPA, POPC, POPG, and DOTAP) were purchased from Avanti Polar Lipids, Inc. The lipid was mixed with the C_{12}-HPTS probe in chloroform at a 100:1 (lipid/probe) molar ratio. The mixture was dried under a stream of N_2 and placed under vacuum for at least 2 h to remove any residual solvent. The dried lipid film was then rehydrated in 5.0 mM phosphate buffer and vortexed for 30 s. The resulting solution was extruded 31 times through a polycarbonate membrane with 200-nm pores (0.2 μm; Nucleopore). The liposome solution was diluted to a lipid concentration of 1.0 mg/mL and stored at 4 °C. Before later use, the pH of the solution was adjusted using either HCl or NaOH.

Spectroscopic Measurements. The above-mentioned lipid vesicle solutions or a 35-μM HPTS solution was used for the spectroscopic measurements. Dynamic light scattering measurements were recorded with a Malvern Zetasizer using a 1-cm-pathlength cuvette. The UV-visible absorption measurements were recorded with a PerkinElmer spectrometer using a 1-cm-pathlength quartz cuvette. The steady-state fluorescence measurements were recorded with a Horiba Fluorolog system using an excitation wavelength of 395 nm with 1-nm bandpass slits for both entrance (excitation) and exit (emission). The time-resolved fluorescence measurements were acquired with a Horiba Deltaflex system, using a 405-nm diode laser with a pulse duration of <100 ps as the light excitation source. The ROH fluorescent decay was collected at 445 nm and 455 nm for HPTS and C_{12}-HPTS, respectively, where at least 30,000 counts (at peak) were collected. The instrument response function was collected at the excitation wavelength using a diluted Ludox solution in water.

MD Simulations. Details of MD experiments are given in *SI Appendix*.

ACKNOWLEDGMENTS. We thank D. Huppert and N. Agmon for fruitful discussions. G.G. and N. Aho thank the CSC – IT Center for Science (Espoo, Finland) for computational resources. N. Amdursky received financial support from the Chaya Career Advancement Chair, the Russel Berrie Nanotechnology Institute, and the Grand Technion Energy Program, and G.G. and N. Aho were supported by the Academy of Finland (Grant 311031).

1. Mulkidjanian AY, Heberle J, Cherepanov DA (2006) Protons @ interfaces: Implications for biological energy conversion. *Biochim Biophys Acta* 1757:913–930.
2. Williams RJP (1988) Proton circuits in biological energy interconversions. *Annu Rev Biophys Biophys Chem* 17:71–97.
3. Agmon N (1995) The Grotthuss mechanism. *Chem Phys Lett* 244:456–462.
4. Knight C, Voth GA (2012) The curious case of the hydrated proton. *Acc Chem Res* 45:101–109.
5. Õjemyr LN, Lee HJ, Gennis RB, Brzezinski P (2010) Functional interactions between membrane-bound transporters and membranes. *Proc Natl Acad Sci USA* 107:15763–15767.
6. Alexiev U, Mollaaghababa R, Scherrer P, Khorana HG, Heyn MP (1995) Rapid long-range proton diffusion along the surface of the purple membrane and delayed proton transfer into the bulk. *Proc Natl Acad Sci USA* 92:372–376.
7. Gopta OA, Cherepanov DA, Junge W, Mulkidjanian AY (1999) Proton transfer from the bulk to the bound ubiquinone Q(B) of the reaction center in chromatophores of Rhodobacter sphaeroides: Retarded conveyance by neutral water. *Proc Natl Acad Sci USA* 96:13159–13164.
8. Heberle J, Riesle J, Thiedemann G, Oesterhelt D, Dencher NA (1994) Proton migration along the membrane surface and retarded surface to bulk transfer. *Nature* 370:379–382.
9. Lechner RE, Dencher NA, Fitter J, Dippel T (1994) Two-dimensional proton diffusion on purple membrane. *Solid State Ion* 70:296–304.
10. Teissié J, Prats M, Soucaille P, Tocanne JF (1985) Evidence for conduction of protons along the interface between water and a polar lipid monolayer. *Proc Natl Acad Sci USA* 82:3217–3221.
11. Springer A, Hagen V, Cherepanov DA, Antonenko YN, Pohl P (2011) Protons migrate along interfacial water without significant contributions from jumps between ionizable groups on the membrane surface. *Proc Natl Acad Sci USA* 108:14461–14466.
12. Zhang C, et al. (2012) Water at hydrophobic interfaces delays proton surface-to-bulk transfer and provides a pathway for lateral proton diffusion. *Proc Natl Acad Sci USA* 109:9744–9749.

13. Weichselbaum E, et al. (2017) Origin of proton affinity to membrane/water interfaces. *Sci Rep* 7:4553.
14. Serowy S, et al. (2003) Structural proton diffusion along lipid bilayers. *Biophys J* 84: 1031–1037.
15. Sandén T, Salomonsson L, Brzezinski P, Widengren J (2010) Surface-coupled proton exchange of a membrane-bound proton acceptor. *Proc Natl Acad Sci USA* 107: 4129–4134.
16. Brändén M, Sandén T, Brzezinski P, Widengren J (2006) Localized proton microcircuits at the biological membrane-water interface. *Proc Natl Acad Sci USA* 103:19766–19770.
17. Wolf MG, Grubmüller H, Groenhof G (2014) Anomalous surface diffusion of protons on lipid membranes. *Biophys J* 107:76–87.
18. Medvedev ES, Stuchebrukhov AA (2013) Mechanism of long-range proton translocation along biological membranes. *FEBS Lett* 587:345–349.
19. Yamashita T, Voth GA (2010) Properties of hydrated excess protons near phospholipid bilayers. *J Phys Chem B* 114:592–603.
20. Re S, Nishima W, Tahara T, Sugita Y (2014) Mosaic of water orientation structures at a neutral zwitterionic lipid/water interface revealed by molecular dynamics simulations. *J Phys Chem Lett* 5:4343–4348.
21. Medvedev ES, Stuchebrukhov AA (2011) Proton diffusion along biological membranes. *J Phys Condens Matter* 23:234103.
22. Toppozini L, et al. (2015) Anomalous and anisotropic nanoscale diffusion of hydration water molecules in fluid lipid membranes. *Soft Matter* 11:8354–8371.
23. Antonenko YN, Pohl P (2008) Microinjection in combination with microfluorimetry to study proton diffusion along phospholipid membranes. *Eur Biophys J* 37:865–870.
24. Åkerman J (2005) Applied physics. Toward a universal memory. *Science* 308:508–510.
25. Nachliel E, Gutman M (1988) Time-resolved proton-phospholipid interaction. Methodology and kinetic analysis. *J Am Chem Soc* 110:2629–2635.
26. Amdursky N, Simkovitch R, Huppert D (2014) Excited-state proton transfer of photoacids adsorbed on biomaterials. *J Phys Chem B* 118:13859–13869.
27. Pines E, Huppert D, Agmon N (1988) Geminate recombination in excited-state proton-transfer reactions: Numerical solution of the Debye–Smoluchowski equation with backreaction and comparison with experimental results. *J Chem Phys* 88:5620–5630.
28. Amdursky N (2015) Photoacids as a new fluorescence tool for tracking structural transitions of proteins: Following the concentration-induced transition of bovine serum albumin. *Phys Chem Chem Phys* 17:32023–32032.
29. Cohen B, Álvarez CM, Carmona NA, Organero JA, Douhal A (2011) Proton-transfer reaction dynamics within the human serum albumin protein. *J Phys Chem B* 115: 7637–7647.
30. Awasthi AA, Singh PK (2017) Excited-state proton transfer on the surface of a therapeutic protein, protamine. *J Phys Chem B* 121:10306–10317.
31. Simkovitch R, Huppert D (2015) Excited-state proton transfer of weak photoacids adsorbed on biomaterials: Proton transfer to glucosamine of chitosan. *J Phys Chem A* 119:641–651.
32. Simkovitch R, Huppert D (2015) Excited-state proton transfer of weak photoacids adsorbed on biomaterials: 8-hydroxy-1,3,6-pyrenetrisulfonate on chitin and cellulose. *J Phys Chem A* 119:1973–1982.
33. Agmon N (2005) Elementary steps in excited-state proton transfer. *J Phys Chem A* 109: 13–35.
34. Agmon N, Pines E, Huppert D (1988) Geminate recombination in proton-transfer reactions. II. Comparison of diffusional and kinetic schemes. *J Chem Phys* 88: 5631–5638.
35. Huppert D, Goldberg SY, Masad A, Agmon N (1992) Experimental determination of the long-time behavior in reversible binary chemical reactions. *Phys Rev Lett* 68: 3932–3935.
36. Cherepanov DA, Feniouk BA, Junge W, Mulkidjanian AY (2003) Low dielectric permittivity of water at the membrane interface: Effect on the energy coupling mechanism in biological membranes. *Biophys J* 85:1307–1316.
37. Teschke O, Ceotto G, de Souza EF (2001) Interfacial water dielectric-permittivity-profile measurements using atomic force microscopy. *Phys Rev E Stat Nonlin Soft Matter Phys* 64:011605.
38. Finkler B, et al. (2014) Highly photostable "super"-photoacids for ultrasensitive fluorescence spectroscopy. *Photochem Photobiol Sci* 13:548–562.
39. Spies C, et al. (2014) Solvent dependence of excited-state proton transfer from pyranine-derived photoacids. *Phys Chem Chem Phys* 16:9104–9114.
40. Barnadas-Rodríguez R, Estelrich J (2008) Effect of salts on the excited state of pyranine as determined by steady-state fluorescence. *J Photochem Photobiol Chem* 198: 262–267.
41. Pines E, Fleming GR (1991) Proton transfer in mixed water-organic solvent solutions: Correlation between rate, equilibrium constant, and the proton free energy of transfer. *J Phys Chem* 95:10448–10457.
42. Jacquemin D, Perpète EA, Ciofini I, Adamo C (2008) Fast and reliable theoretical determination of pKa* for photoacids. *J Phys Chem A* 112:794–796.
43. Pines D, Pines E (2007) Solvent assisted photoacidity. *Hydrogen-Transfer Reactions* (Wiley-VCH, Weinheim, Germany), pp 377–415.
44. Krissinel EB, Agmon N (1996) Spherical symmetric diffusion problem. *J Comput Chem* 17:1085–1098.
45. Amdursky N, Rashid MH, Stevens MM, Yarovsky I (2017) Exploring the binding sites and proton diffusion on insulin amyloid fibril surfaces by naphthol-based photoacid fluorescence and molecular simulations. *Sci Rep* 7:6245.
46. Agmon N, Gutman M (2011) Bioenergetics: Proton fronts on membranes. *Nat Chem* 3: 840–842.
47. Medvedev ES, Stuchebrukhov AA (2006) Kinetics of proton diffusion in the regimes of fast and slow exchange between the membrane surface and the bulk solution. *J Math Biol* 52:209–234.
48. Amdursky N, Orbach R, Gazit E, Huppert D (2009) Probing the inner cavities of hydrogels by proton diffusion. *J Phys Chem C* 113:19500–19505.

II

SCALABLE CONSTANT PH MOLECULAR DYNAMICS IN GROMACS

by

Noora Aho, Pavel Buslaev, Anton Jansen, Paul Bauer, Gerrit Groenhof, Berk Hess. *Journal of Chemical Theory and Computation, 18(10),* 6148-6160, **2022.**

DOI: 10.1021/acs.jctc.2c00516

JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

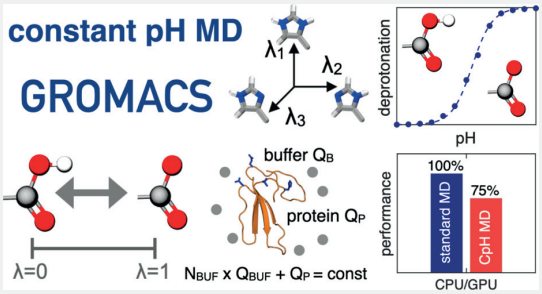
Scalable Constant pH Molecular Dynamics in GROMACS

Noora Aho,* Pavel Buslaev,* Anton Jansen, Paul Bauer, Gerrit Groenhof,* and Berk Hess*

Cite This: https://doi.org/10.1021/acs.jctc.2c00516 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Molecular dynamics (MD) computer simulations are used routinely to compute atomistic trajectories of complex systems. Systems are simulated in various ensembles, depending on the experimental conditions one aims to mimic. While constant energy, temperature, volume, and pressure are rather straightforward to model, pH, which is an equally important parameter in experiments, is more difficult to account for in simulations. Although a constant pH algorithm based on the λ-dynamics approach by Brooks and co-workers [Kong, X.; Brooks III, C. L. *J. Chem. Phys.***1996**, *105*, 2414−2423] was implemented in a fork of the GROMACS molecular dynamics program, uptake has been rather limited, presumably due to the poor scaling of that code with respect to the number of titratable sites. To overcome this limitation, we implemented an alternative scheme for interpolating the Hamiltonians of the protonation states that makes the constant pH molecular dynamics simulations almost as fast as a normal MD simulation with GROMACS. In addition, we implemented a simpler scheme, called multisite representation, for modeling side chains with multiple titratable sites, such as imidazole rings. This scheme, which is based on constraining the sum of the λ-coordinates, not only reduces the complexity associated with parametrizing the intramolecular interactions between the sites but also is easily extendable to other molecules with multiple titratable sites. With the combination of a more efficient interpolation scheme and multisite representation of titratable groups, we anticipate a rapid uptake of constant pH molecular dynamics simulations within the GROMACS user community.



INTRODUCTION

Since their introduction more than four decades ago, molecular dynamics (MD) computer simulations have come of age.[1] Thanks to improvements in computer hardware, algorithmic developments, as well as increased accuracy of force fields, MD simulation has evolved into a predictive technique that can complement experiments by providing atomistic insights into the dynamics of complex systems.[1,2] While many experimental conditions can be modeled with good accuracy, the aqueous proton concentration, or pH, is typically accounted for indirectly by constraining the protonation states of titratable residues to their, presumed, most probable form at the start of the simulation. Because the electrostatic interactions depend critically on the protonation state of the residues, the pH affects the conformational ensemble. Conversely, because the conformation can influence the proton affinity of the residues, or pK_a, a direct correlation exists between pH and conformational dynamics, which cannot be captured if the protonation state is kept fixed in the simulation.[3]

To overcome this limitation in classical MD simulations and include the effect of pH on the conformational sampling directly, several solutions have been proposed in the last decades[4,5] and used to investigate pH-dependent protein−protein[6] and protein-RNA interactions,[7] drug binding,[8,9] and structural changes.[10,11] These solutions can be roughly divided into a category that relies on discrete changes in protonation states[12−19] and a second category in which a protonation state can change continuously.[20−33] More recently, a third category that relies on the transfer of proton-like particles between titratable sites, including protein residues and solvent molecules, was proposed for the Martini force field.[34]

In the discrete constant pH approaches, the protonation state of a residue can change at regular intervals of the simulation according to a Metropolis Monte Carlo criterion.[16,17,28,35,36] To avoid a low acceptance rate due to unfavorable solvent configurations, the Monte Carlo step is performed based on free energies calculated using the approximation of either an implicit solvent representation,[13,15] or a short all-atom thermodynamic integration.[14,37]

Most continuous approaches for MD at constant pH are based on the λ-dynamics technique developed by Brooks and co-workers.[38] A one-dimensional λ-coordinate with fictitious mass m_λ is introduced for each titratable site, and the equations of motion for these additional degrees of freedom are integrated along with the Cartesian positions of the atoms.[21] The

Received: May 18, 2022
Published: September 21, 2022

λ-coordinate defines the protonation state of the residue: at $\lambda = 0$ the residue is protonated and interacts with the rest of the system as such, while at $\lambda = 1$ it is deprotonated. The energy function that acts on the λ-coordinates depends on (i) the intrinsic proton affinity (reference pK_a) of the titratable site in water, (ii) the interactions with the environment, which are mostly electrostatic,[39] and (iii) the pH of the solvent, which is set by the user. In addition, potentials are introduced to bias sampling toward the physical states at $\lambda = 0$ and $\lambda = 1$. Protons are not transferred directly between the titratable residues and the solvent molecules but rather exchanged with an external proton bath. Because the chemical potential of this bath is determined by the proton concentration (pH) of the aqueous solution, constant pH MD (CpHMD) simulations based on λ-dynamics are performed in a grand canonical ensemble for the proton degrees of freedom.

While λ-dynamics-based constant pH approaches were originally developed for implicit solvent simulations,[21] they have since then been adapted for explicit solvent simulations as well.[18,23−26,28,30] The key computational challenge for explicit solvent implementations is the long-range electrostatic interaction, for which multiple solutions have been suggested, including a shifted cutoff scheme,[26] a hybrid scheme combining the particle mesh Ewald (PME) treatment for the Cartesian coordinates with the generalized Born model for the λ-particles,[24,40] and a fully consistent PME treatment for both λ and Cartesian degrees of freedom.[23,30]

In addition to accurate modeling of the long-range electrostatic interactions, also sampling can pose a serious challenge to simulations at constant pH. While the choice for the PME method in the original implementation of λ-dynamics in the fork of GROMACS 3.3 release[23] was motivated by its accurate description of long-range electrostatics, the linear increase of the computational effort with the number of titratable sites limited the sampling efficiency, which meant that systems with many titratable sites could not be studied in practice.

To remove this bottleneck and enable constant pH MD with GROMACS at a modest additional cost compared to a standard simulation, we switch to an alternative scheme for computing the long-range electrostatic interactions of the λ-particles under periodic boundary conditions. The alternative scheme is based on a linear interpolation of partial charges[21] rather than the potential energy functions as in the original implementation of constant pH MD in a GROMACS fork.[23]

Although the previous implementation of constant pH in a GROMACS fork was documented and shared with the community as an open-source program, there has been some misunderstanding about how electrostatic interactions were computed for λ-particles.[24,25,30] To resolve this controversy, we first explain in detail how the electrostatic interactions were calculated in the previous GROMACS implementation of constant pH MD. We next contrast this linear interpolation between the potential energy functions of the protonated and deprotonated states of a residue on the one hand to the interpolation between the partial charges of both states on the other hand[21] and show why the latter is computationally much more efficient. We then demonstrate the superior performance of the charge-interpolation scheme by running a series of constant pH MD simulations of amino acids and proteins. To emphasize that the new constant pH implementation in GROMACS is not restricted to a specific force field (or the resolution of a force field model) nor to a specific algorithm for evaluating electrostatic interactions, we also show the results of constant pH MD simulations with the Martini coarse-grained force field,[41] in combination with a shifted cutoff electrostatics model. Because of GROMACS' large user community, we expect our work to increase the popularity of constant pH MD simulations.

THEORY

Before discussing the differences between linear interpolating of the potential energy functions on the one hand,[23] and of partial charges on the other hand,[21] for computing the potential energy landscape of the titration coordinates, we briefly review the λ-dynamics approach[38] that forms the basis for the constant pH molecular dynamics algorithm in GROMACS.[23]

λ-Dynamics-Based Constant pH MD Simulations. A titratable site i can exist in a protonated or deprotonated state. The protonation state affects the interactions between the site and the rest of the system. In constant pH MD simulations based on λ-dynamics,[38] an additional coordinate λ_i is introduced for each site i, and the potential energy function of the total system is continuously interpolated between the two protonation states along this coordinate, i.e., $V(\lambda_i)$.[21] A fictitious mass m_λ is assigned to each λ_i-coordinate, and the coordinates evolve along with the Cartesian degrees of freedom of all atoms in the system, based on Newton's equations of motion. Thus, the total Hamiltonian of the system is

$$H(\mathbf{R}, \boldsymbol{\lambda}) = \sum_{i}^{N_{\text{sites}}} \frac{m_\lambda}{2}\dot{\lambda}_i^2 + \sum_{j}^{N_{\text{atoms}}} \frac{m_j}{2}\dot{\mathbf{r}}_j^2 + V(\mathbf{R}, \boldsymbol{\lambda}) \tag{1}$$

where $\mathbf{R}$ is the vector of the Cartesian coordinates $\mathbf{r}_j$ of all N_{atoms} atoms with mass m_j and $\boldsymbol{\lambda}$ is the vector of the λ_i coordinates of all N_{sites} titratable sites.

λ-Dependent Potential Energy Function. In addition to the interpolation between the potentials of the protonated $V_A(\mathbf{R})$ and deprotonated states $V_B(\mathbf{R})$, three more λ-dependent terms are included in the potential energy function of the total system $V(\mathbf{R}, \boldsymbol{\lambda})$, as illustrated in Figure 1: (i) a correction potential $V_i^{\text{MM}}(\lambda_i)$ to compensate for missing quantum mechanical contributions to proton affinities (Figure 1B); (ii) a biasing potential $V_i^{\text{bias}}(\lambda_i)$ that enhances sampling of the physical end states at $\lambda_i = 0$ and $\lambda_i = 1$ (Figure 1C); and (iii) a pH-dependent term $V^{\text{pH}}(\lambda_i)$ to model the chemical potential of protons in water (Figure 1D).

The purpose of adding the correction term $V_i^{\text{MM}}(\lambda_i)$ (Figure 1B) is to make the interpolated potential function flat if the titratable site i is in its reference state, for which the proton affinity is known experimentally, at pH = p$K_{a,i}$. This potential is determined by evaluating the deprotonation free energy of the single residue in water (reference state) at the force field level by thermodynamic integration along the λ-coordinate (Figure 1A):

$$V_i^{\text{MM}}(\lambda_i) = -\Delta G_i^{\text{MM}}(\lambda_i) \tag{2}$$

To prevent sampling of the nonphysical states between $\lambda_i = 0$ and $\lambda_i = 1$ on this flat potential energy surface while still enabling sufficient transitions between the physical end-states to sample both protonation states with the correct thermodynamic weight, we introduce the biasing potential $V_i^{\text{bias}}(\lambda_i)$ suggested by Donnini et al.[42] (Figure 1C).

The pH-dependent term $V^{\text{pH}}(\lambda_i)$ (Figure 1D) is a correction that includes the effect of the solution pH on the free energy difference between the protonated and deprotonated states, such that this difference is

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

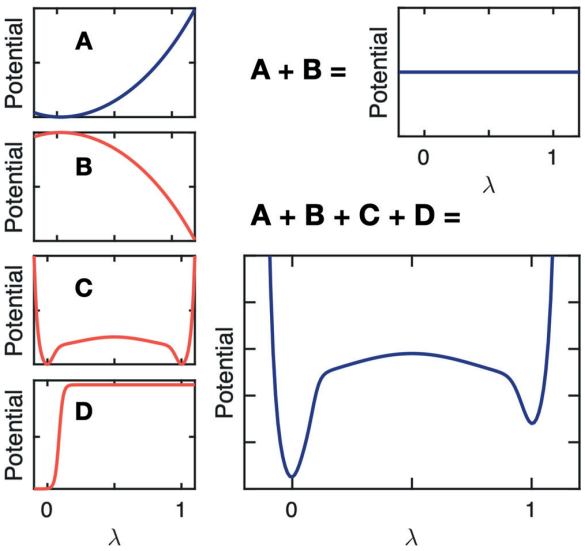


Figure 1. Illustration of the additional λ-dependent potential energy terms (B−D). Panel A shows the protonation free energy of a titratable residue in its reference state obtained at the force field level, $\Delta G_i^{\mathrm{MM}}(\lambda_i)$. To compensate for the shortcomings of the force field and obtain a zero free energy difference between the two protonation states (A + B), we add the correction potential, $V_i^{\mathrm{MM}}(\lambda)$, shown in panel B. A biasing potential, $V_i^{\mathrm{bias}}(\lambda)$,[42] is introduced to avoid sampling of nonphysical states (panel C). To model the proton chemical potential (pH), we add a pH-dependent term, $V^{\mathrm{pH}}(\lambda_i)$ (panel D). For a titratable residue at pH ≠ pK_a, the total λ-dependent potential, including the interpolated force field functions and the three additional terms, is illustrated in panel (A + B + C + D).

$$V^{\mathrm{pH}}(\lambda_i = 1) - V^{\mathrm{pH}}(\lambda_i = 0) = RT\ln(10)[\mathrm{p}K_{\mathrm{a},i} - \mathrm{pH}] \tag{3}$$

where we use the experimentally determined p$K_{\mathrm{a},i}$ value of residue i in its reference state. Although various forms for this potential have been suggested,[29,42,43] we propose a smooth step-function-based potential:

$$\begin{aligned} V^{\mathrm{pH}}(\lambda_i) &= RT\ln(10)[\mathrm{p}K_{\mathrm{a},i} - \mathrm{pH}] \\ &\quad \times \frac{1}{1 + \exp(-2k_1(\lambda_i - 1 + x_0))}, \text{ if pH} > \mathrm{p}K_{\mathrm{a}} \\ V^{\mathrm{pH}}(\lambda_i) &= RT\ln(10)[\mathrm{p}K_{\mathrm{a},i} - \mathrm{pH}] \times \frac{1}{1 + \exp(-2k_1(\lambda_i - x_0))}, \\ &\quad \text{if pH} \le \mathrm{p}K_{\mathrm{a}} \end{aligned} \tag{4}$$

where k_1 and x_0 define the steepness and kink position of the step function. In this form, illustrated in Figure 1D, the pH-dependent potential also aids in preventing the sampling of nonphysical states, i.e., $0.1 < \lambda < 0.9$.

Linear Interpolation of Potential Energy Functions. In the previous implementation of constant pH MD in a GROMACS fork, the smooth interpolation of the force field potential energy function between the protonated and deprotonated states was achieved by linearly interpolating the force field potentials of these states.[23] Thus, for a single titratable site, the λ-dependent potential is given by

$$\begin{aligned} V(\mathbf{R}, \lambda) &= (1 - \lambda)V_A(\mathbf{R}) + \lambda V_B(\mathbf{R}) + V^{\mathrm{MM}}(\lambda) + V^{\mathrm{bias}}(\lambda) \\ &\quad + V^{\mathrm{pH}}(\lambda) \end{aligned} \tag{5}$$

with $V^{\mathrm{MM}}(\mathbf{R}, \lambda)$, $V^{\mathrm{bias}}(\lambda)$, and $V^{\mathrm{pH}}(\lambda)$ the correction, biasing, and pH-dependent potentials, respectively, that were briefly discussed above, and with short-hand notations for

$$V_A(\mathbf{R}) = V(\mathbf{R}, \lambda = 0)$$
$$V_B(\mathbf{R}) = V(\mathbf{R}, \lambda = 1)$$

The gradient required for updating λ according to Newton's equations of motion is

$$\begin{aligned} \frac{\partial V(\mathbf{R}, \lambda)}{\partial \lambda} &= V_B(\mathbf{R}) - V_A(\mathbf{R}) + \frac{\partial}{\partial \lambda}V^{\mathrm{MM}}(\lambda) + \frac{\partial}{\partial \lambda}V^{\mathrm{bias}}(\lambda) \\ &\quad + \frac{\partial}{\partial \lambda}V^{\mathrm{pH}}(\lambda) \end{aligned} \tag{6}$$

Thus, the evaluation of the force on the λ-particle requires that the potential energy, including the long-range electrostatic interactions, is computed twice: once for $\lambda = 0$ (i.e., V_A) and once more for $\lambda = 1$ (*i.e.*, V_B). If the Particle-Mesh-Ewald (PME) method is used to compute those long-range electrostatic interactions,[44,45] separate PME grid builds are needed because the charge distributions are not identical in states A and B.

For systems with many titratable sites, multiple λ-groups are introduced. Because the analytical expressions for the correction, biasing, and pH-dependent terms in eq 5 are additive, we no longer consider them explicitly in what follows and focus exclusively on the interpolation of the force field potential energies between the multiple protonation states of the system. For N λ-coordinates, there are 2^N such states and the interpolation generalizes to[20]

$$V(\mathbf{R}, \boldsymbol{\lambda}) = \sum_{\mathbf{l}}^{2^N} V(\mathbf{R}, \mathbf{l})\Pi_i^N \lambda_i^{(l_i)}(1 - \lambda_i)^{1 - l_i} \tag{7}$$

Here, we represent the N λ_i-coordinates as an N-dimensional vector $\boldsymbol{\lambda}$. The 2^N possible protonation states of the system are represented by the N-dimensional vector $\mathbf{l}$ with elements l_i equal to 0 or 1 that indicate whether a site i is protonated ($\lambda_i = 0$) or deprotonated ($\lambda_i = 1$). The sum runs over all 2^N possible combinations of $l_i = 0$ and $l_i = 1$. The gradient required for updating λ_i is obtained by deriving the interpolated potential, $V(\mathbf{R}, \boldsymbol{\lambda})$, with respect to λ_i:

$$\begin{aligned} \frac{\partial}{\partial \lambda_i}V(\mathbf{R}, \boldsymbol{\lambda}) &= \frac{\partial}{\partial \lambda_i}\left[\sum_{\mathbf{l}}^{2^N} V(\mathbf{R}, \mathbf{l})\Pi_k^N \lambda_k^{(l_k)}(1 - \lambda_k)^{1 - l_k}\right] \\ &= \frac{\partial}{\partial \lambda_i}\left[\sum_{\mathbf{l}, l_i = 0}^{2^{N-1}} V(\mathbf{R}, \mathbf{l})\Pi_k^N \lambda_k^{(l_k)}(1 - \lambda_k)^{1 - l_k}(1 - \lambda_i)\right. \\ &\quad \left. + \sum_{\mathbf{l}, l_i = 1}^{2^{N-1}} V(\mathbf{R}, \mathbf{l})\Pi_k^N \lambda_k^{(l_k)}(1 - \lambda_k)^{1 - l_k}\lambda_i\right] \\ &= \sum_{\mathbf{l}, l_i = 1}^{2^{N-1}} V(\mathbf{R}, \mathbf{l})\Pi_k^N \lambda_k^{(l_k)}(1 - \lambda_k)^{1 - l_k} \\ &\quad - \sum_{\mathbf{l}, l_i = 0}^{2^{N-1}} V(\mathbf{R}, \mathbf{l})\Pi_k^N \lambda_k^{(l_k)}(1 - \lambda_k)^{1 - l_k} \\ &= V(\mathbf{R}, \boldsymbol{\lambda}', \lambda_i = 1) - V(\mathbf{R}, \boldsymbol{\lambda}', \lambda_i = 0) \end{aligned} \tag{8}$$

where the $'$ indicates that λ_i is omitted from vector $\boldsymbol{\lambda}$. Note that, as we focus only on the interpolated potentials, the biasing, correction, and pH-dependent terms are left out.

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

In general, the number of terms in the potential (eq 7) increases exponentially with the number of titratable sites. However, for pairwise interactions involving titratable sites whose nonbonded force field parameters do not depend on the protonation state of the other sites (chemically uncoupled sites), the number of terms required to evaluate the interpolated potential scales linearly. For systems with such "chemically", or "topologically" uncoupled sites, the interpolated potential contains four types of interactions

$$V(\mathbf{R}, \lambda) = V^{\text{rest}-\text{rest}}(\mathbf{R}) + V^{\lambda-\text{rest}}(\mathbf{R}, \lambda) + V^{\lambda-\lambda}(\mathbf{R}, \lambda) + V^{\lambda}(\mathbf{R}, \lambda) \tag{9}$$

For pairwise electrostatic interactions, the terms on the right-hand side are defined as

1. Interactions between atoms that are not part of any λ-group, and hence independent of the λ_i's:

$$V^{\text{rest}-\text{rest}}(\mathbf{R}) = \frac{1}{2}\sum_{i}^{n_{\text{rest}}}\sum_{j}^{n_{\text{rest}}}\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{10}$$

where the sums run over all n_{rest} atoms that are not part of a λ-group.

2. Interpolated interactions between atoms of each λ-group with atoms that are not part of any λ-group:

$$V^{\lambda-\text{rest}}(\mathbf{R}, \lambda) = \sum_{k}^{N_{\text{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_{\text{rest}}}\left\{(1-\lambda_k)\frac{q_i^A q_j}{4\pi\epsilon_0 r_{ij}} + \lambda_k\frac{q_i^B q_j}{4\pi\epsilon_0 r_{ij}}\right\} \tag{11}$$

where the first sum runs over all titratable sites, the second one runs over all n_k atoms of the k th λ-group, and the final sum runs over all atoms that are not part of any λ group.

3. Interpolated interactions between atoms belonging to two different λ-groups:

$$\begin{aligned} V^{\lambda-\lambda}(\mathbf{R}, \lambda) &= \sum_{k}^{N_{\text{sites}}}\left\{(1-\lambda_k)\sum_{m,m\neq k}^{N_{\text{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_m}\left[(1-\lambda_m)\frac{q_i^A q_j^A}{4\pi\epsilon_0 r_{ij}} + \lambda_m\frac{q_i^A q_j^B}{4\pi\epsilon_0 r_{ij}}\right] + \lambda_k\sum_{m,m\neq k}^{N_{\text{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_m}\left[(1-\lambda_m)\frac{q_i^B q_j^A}{4\pi\epsilon_0 r_{ij}} + \lambda_m\frac{q_i^B q_j^B}{4\pi\epsilon_0 r_{ij}}\right]\right\} \\ &= \sum_{k}^{N_{\text{sites}}}\sum_{m,m\neq k}^{N_{\text{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_m}\frac{[(1-\lambda_k)q_i^A + \lambda_k q_i^B][(1-\lambda_m)q_j^A + \lambda_m q_j^B]}{4\pi\epsilon_0 r_{ij}} \end{aligned} \tag{12}$$

4. Interpolated interactions within each of the λ-groups:

$$V^{\lambda}(\mathbf{R}, \lambda) = \frac{1}{2}\sum_{k}^{N_{\text{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_k}\left\{(1-\lambda_k)\frac{q_i^A q_j^A}{4\pi\epsilon_0 r_{ij}} + \lambda_k\frac{q_i^B q_j^B}{4\pi\epsilon_0 r_{ij}}\right\} \tag{13}$$

From eq 8, the gradient with respect to λ_k is

$$\begin{aligned} \frac{\partial}{\partial\lambda_k}V_{\text{Coul}}(\mathbf{R}, \lambda) &= V_{\text{Coul}}(\mathbf{R}, \lambda', \lambda_k = 1) - V_{\text{Coul}}(\mathbf{R}, \lambda', \lambda_k = 0) \\ &= \sum_{i}^{n_k}\sum_{j}^{n_{\text{rest}}}\left[\frac{q_i^B q_j}{4\pi\epsilon_0 r_{ij}} - \frac{q_i^A q_j}{4\pi\epsilon_0 r_{ij}}\right] + \sum_{m,m\neq k}^{N_{\text{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_m}\left[\frac{q_i^B[(1-\lambda_m)q_j^A + \lambda_m q_j^B]}{4\pi\epsilon_0 r_{ij}} - \frac{q_i^A[(1-\lambda_m)q_j^A + \lambda_m q_j^B]}{4\pi\epsilon_0 r_{ij}}\right] + \frac{1}{2}\sum_{i}^{n_k}\sum_{j}^{n_k}\left[\frac{q_i^B q_j^B}{4\pi\epsilon_0 r_{ij}} - \frac{q_i^A q_j^A}{4\pi\epsilon_0 r_{ij}}\right] \end{aligned} \tag{14}$$

Thus, the evaluation of the Coulomb contribution to the gradient for each λ_k-group requires two electrostatic computations, with the interpolated partial charges of the other λ_m sites (i.e., $q_j(\lambda_m) = (1-\lambda_m)q_j^A + \lambda_m q_j^B$):

$$\begin{aligned} \frac{\partial}{\partial\lambda_k}V_{\text{Coul}}(\mathbf{R}, \lambda) &= \sum_{i}^{n_k} q_i^B \times \left[\sum_{a}^{n_{\text{rest}}}\frac{q_a}{4\pi\epsilon_0 r_{ia}} + \sum_{m,m\neq k}^{N_{\text{sites}}}\sum_{j}^{n_m}\frac{(1-\lambda_m)q_j^A + \lambda_m q_j^B}{4\pi\epsilon_0 r_{ij}} + \frac{1}{2}\sum_{j}^{n_k}\frac{q_j^B}{4\pi\epsilon_0 r_{ij}}\right] - \sum_{i}^{n_k} q_i^A \times \left[\sum_{a}^{n_{\text{rest}}}\frac{q_a}{4\pi\epsilon_0 r_{ia}} + \sum_{m,m\neq k}^{N_{\text{sites}}}\sum_{j}^{n_m}\frac{(1-\lambda_m)q_j^A + \lambda_m q_j^B}{4\pi\epsilon_0 r_{ij}} + \frac{1}{2}\sum_{j}^{n_k}\frac{q_j^A}{4\pi\epsilon_0 r_{ij}}\right] \\ &= \sum_{i}^{n_k} q_i^B\Phi_k^B(\mathbf{R}_i, \lambda') - \sum_{i}^{n_k} q_i^A\Phi_k^A(\mathbf{R}_i, \lambda') \end{aligned} \tag{15}$$

Here, we introduced the electrostatic potential $\Phi_k^A(\mathbf{R}_i, \lambda')$ of the system with partial charges of λ-group k in the protonated state (q_i^A) and interpolated charges for all other λ-groups. As before, λ' is the vector with all λ_m's except λ_k. Likewise, electrostatic potential $\Phi_k^B(\mathbf{R}_i, \lambda')$ is evaluated with the partial charges of λ-group k in the deprotonated state (q_i^B) and the same interpolated charges for all other λ-groups. Thus, $2N_{\text{sites}}$ computations are needed to evaluate the gradients for all titratable sites. The same arguments apply to pairwise Lennard-Jones interactions, but because the contribution of Lennard-Jones interaction to pK_a shift is minor, we neglected them in this work (see Supporting Information).

Linear Interpolation of Partial Charges. While the linear scaling of the gradients for the pairwise potentials in eq 15 in principle is a great improvement over the formal exponential scaling in eq 8, the requirement of performing $2N_{\text{sites}}$ calculations per MD step still poses a computational bottleneck, in particular for larger systems. To overcome this bottleneck for electrostatic interactions, we follow the suggestion by Brooks and co-workers to interpolate charges rather than interaction functions.[21] When interpolating the partial charges between the protonation states of N_{sites} chemically uncoupled titratable sites, the λ-dependent Coulomb energy becomes

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$V_{\mathrm{coul}}(\mathbf{R}, \lambda) = V^{\mathrm{rest-rest}}(\mathbf{R}) + V^{\lambda-\mathrm{rest}}(\mathbf{R}, \lambda) + V^{\lambda-\lambda}(\mathbf{R}, \lambda) + V^{\lambda}(\mathbf{R}, \lambda)$$
$$= \frac{1}{2}\sum_{i}^{n_{\mathrm{rest}}}\sum_{j}^{n_{\mathrm{rest}}}\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{k}^{N_{\mathrm{sites}}}\sum_{i}^{n_k}\sum_{i}^{n_{\mathrm{rest}}}\frac{((1-\lambda_k)q_i^A + \lambda_k q_i^B)q_j}{4\pi\epsilon_0 r_{ij}}$$
$$+ \sum_{k}^{N_{\mathrm{sites}}}\sum_{m;m\neq k}^{N_{\mathrm{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_m}\frac{[(1-\lambda_k)q_i^A + \lambda_k q_i^B][(1-\lambda_m)q_j^A + \lambda_m q_j^B]}{4\pi\epsilon_0 r_{ij}}$$
$$+ \frac{1}{2}\sum_{k}^{N_{\mathrm{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_k}\frac{[(1-\lambda_k)q_i^A + \lambda_k q_i^B][(1-\lambda_k)q_j^A + \lambda_k q_j^B]}{4\pi\epsilon_0 r_{ij}} \quad (16)$$

The gradient of the potential energy with respect to λ_k is

$$\frac{\partial V_{\mathrm{coul}}(\mathbf{R}, \lambda)}{\partial\lambda_k} = \sum_{i}^{n_k}\sum_{j}^{n_{\mathrm{rest}}}\frac{(q_i^B - q_i^A)q_j}{4\pi\epsilon_0 r_{ij}}$$
$$+ \sum_{m=1;m\neq k}^{N_{\mathrm{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_m}\frac{(q_i^B - q_i^A)[(1-\lambda_m)q_j^A + \lambda_m q_j^B]}{4\pi\epsilon_0 r_{ij}}$$
$$+ \sum_{i}^{n_k}\sum_{j}^{n_k}\frac{(q_i^B - q_i^A)[(1-\lambda_k)q_j^A + \lambda_k q_j^B]}{4\pi\epsilon_0 r_{ij}} = \sum_{i}^{n_k}\sum_{j}^{n_{\mathrm{rest}}}\frac{(q_i^B - q_i^A)q_j}{4\pi\epsilon_0 r_{ij}}$$
$$+ \sum_{m=1}^{N_{\mathrm{sites}}}\sum_{i}^{n_k}\sum_{j}^{n_m}\frac{(q_i^B - q_i^A)[(1-\lambda_m)q_j^A + \lambda_m q_j^B]}{4\pi\epsilon_0 r_{ij}}$$
$$= \sum_{i}^{n_k}\Phi(\mathbf{R}_i, \lambda)\Delta q_i \quad (17)$$

where $\Phi(\mathbf{R}_i, \lambda)$ is the electrostatic potential at the position of atom i due to the charge distribution of all other atoms in the system, including the atoms of all titratable sites, for which the partial charges are interpolated:

$$\Phi(\mathbf{R}_i, \lambda) = \sum_{j}^{n_{\mathrm{rest}}}\frac{q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{m}^{N_{\mathrm{sites}}}\sum_{j}^{n_m}\frac{(1-\lambda_m)q_j^A + \lambda_m q_j^B}{4\pi\epsilon_0 r_{ij}}$$

and Δq_i is the difference between the atomic charges of titratable residue i in the protonated (A) and deprotonated (B) states:

$$\Delta q_i = q_i^B - q_i^A$$

In contrast to when potential energy functions are interpolated, the same electrostatic potential is used to evaluate the electrostatic forces on both the atoms and the λ-particles. Therefore, a single electrostatic calculation per time step suffices. If the electrostatic interactions are modeled with the smooth Particle Mesh Ewald method,[44,45] the short-range real-space interactions and long-range reciprocal-space interactions are computed separately. For the pairwise short-range interactions in real space, an additional calculation for each interacting pair and a subsequent accumulation of the potential at each atom is needed. Whereas this calculation comes at no extra computational cost if the standard pair interaction kernels are used, the accumulation leads to a measurable computational overhead, as we will show later. For the mesh part of the PME calculation, a gathering of potentials from the grid is required for charges in λ-groups only, but this also comes at a negligible computational overhead. Because the extra effort required to compute the gradients on the λ-particles is rather small, a constant pH MD implementation based on charge interpolation is computationally not much more expensive than a normal MD simulation, which is a major improvement with respect to the previous CpHMD implementation in GROMACS.[23]

Multisite Representation of Chemically Coupled Titratable Sites. If titratable sites are "chemically" or "topologically" coupled, the force field parameters of one site depend on the value of the λ-coordinate of the other site, and *vice versa*. For example, histidine can exist in three protonation states, as shown in Figure 2. In most force fields, the partial charges of all atoms in the His side chain, including the two sites, depend on the protonation state. Hence, if the N_δ site changes protonation, the electrostatic interactions of the N_ϵ site are also affected.

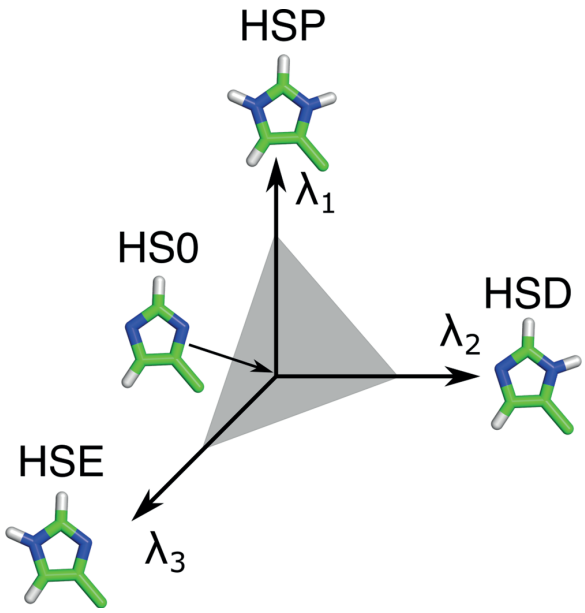


Figure 2. Multisite representation illustrated for a histidine side chain. Each possible protonation state is represented by its own λ-coordinate. HSP refers to doubly protonated histidine, HSD, and HSE to histidine singly protonated at the N_δ or N_ϵ, respectively. HS0 is a common, nonphysical state of the residue. To restrict the sampling to the plane connecting the physical states, a constraint $\lambda_1 + \lambda_2 + \lambda_3 = 1$ is applied (gray plane). A biasing potential is also applied to enhance sampling at the end states, where one of the λ-coordinates is close to one, while the other coordinates are close to zero.

To model the chemically coupled sites in the histidine side chain, Khandogin and Brooks introduced two λ-coordinates:[22] one that interpolates between the double and single protonated forms and a second coordinate switching between protonation at the N_δ and the N_ϵ atoms. Donnini et al. introduced separate λ-coordinates for N_δ and N_ϵ.[23] In both solutions, the coupling between the coordinates is achieved with a two-dimensional correction potential.

Because extending the dimensionality beyond two coordinates is difficult from both the implementation and parametrization perspective, Brooks and co-workers introduced a multisite representation,[46,47] where a separate $\lambda_{i,k}$-coordinate is assigned to each physical state k of a titratable group i. For a residue with multiple "chemically"-coupled titratable sites, each $\lambda_{i,k}$-coordinate has the same state at $\lambda_{i,k} = 0$, while at $\lambda_{i,k} = 1$, the group is in one of the n_i possible protonation states (i.e., state k) of residue i. The state at $\lambda_{i,k} = 0$ is the same for all $\lambda_{i,k}$-coordinates in residue i but does not correspond to a physical protonation state of the residue and neither do states for which the sum of the $\lambda_{i,k}$-s is not equal to 1 (Figure 2). To restrict sampling to the (hyper-)plane connecting the physical states, the sum of the $\lambda_{i,k}$ values is constrained ($\sum_k \lambda_{i,k} = 1$). Since the λ-dynamics

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

implementation in a fork of GROMACS relies on linear λ-coordinates, rather than on auxiliary circular coordinates that would fulfill the constraints by construction,[46,48] we apply a constraint on the sum of $\lambda_{i,k}$-coordinates. To efficiently apply this constraint, we use an analytical expression to solve a generalized version of the charge constraint introduced by Donnini et al.[42] (see the Appendix). While an atom can be part of multiple $\lambda_{i,k}$-coordinates in residue i, each affecting its charge, we show in the Supporting Information that the expression for the contribution of this atom to the total Coulomb energy is identical to that of an uncoupled site (eq 17).

In the multisite representation, each λ-coordinate is independent of the others and thus evolves on a one-dimensional potential (eq 4), similar to that of "chemically" uncoupled sites. However, in contrast to the uncoupled sites, the correction potential V^{MM} is multidimensional as its value depends on all $\lambda_{i,k}$-coordinates representing each of the possible protonation states of residue i. These potentials are obtained through a least-squares fit of a multidimensional polynomial to the ensemble-averaged gradients of the potentials with respect to $\lambda_{i,k}$ evaluated on the $(n_i - 1)$-dimensional grid of the n_i coupled λ-coordinates, *i.e.*, $\langle \partial V/\partial\lambda_{i,k}\rangle_{\lambda_1 \ldots \lambda_{ni}}$. The fitting procedure is explained in detail in the Supporting Information.

The multisite representation can be applied to residues with any number of titratable sites, including residues with only a single titratable site. In the latter case, two λ-coordinates, corresponding to the protonated state ($\lambda_{i,1} = 1$, $\lambda_{i,2} = 0$) and deprotonated state ($\lambda_{i,1} = 0$, $\lambda_{i,2} = 1$), are introduced with a constraint on their sum ($\lambda_{i,1} + \lambda_{i,2} = 1$).

METHODS

We have implemented the algorithms for CpHMD with charge interpolation in a fork of GROMACS software package (2021 release).[49] The code and manuals are available for free at https://gitlab.com/gromacs-constantph/constantph. Here we verify the validity of our implementation for reproducing pK_a values of peptides and proteins. To demonstrate that the linear interpolation of charges (eq 17) scales better with the number of titratable sites in the system than the linear interpolation of interaction functions (eq 15), we compared the scaling between our new implementation, which is based on linear charge interpolation on the one hand, and a previous implementation in a fork of the GROMACS 3.3 release, which is based on linear interpolation of the force field potentials on the other hand.[23] To estimate the additional computational effort required for performing CpHMD with the new implementation, we also compared the performance of a CpHMD simulation to that of a normal MD simulations on both CPUs and GPUs.

Simulated Systems. To test the implementation, we performed constant pH MD simulations of the following systems: (1) glutamic acid (Glu), (2) aspartic acid (Asp), (3) histidine (His), (4) Cardiotoxin V (PDB ID: 1CVO[50]), (5) hen egg white lyzozyme (HEWL, PDB ID: 2LZT[51]), (6) the GLIC pentameric ligand-gated ion channel (PDB ID: 4HFI[52]), and (7) turkey ovomucoid inhibitor (PDB ID: 2GKR[53]).

In systems 1−6, the interactions were modeled with the CHARMM36[54] all-atom (AA) force field, with some modifications in the torsion parameters to accelerate the convergence. These modifications are presented and validated in an accompanying paper, in which we report the application of our constant pH implementation to lysine, C-, and N-termini.[55] Systems 1−5 were also simulated with the Martini 2.0 coarse grained (CG) force field.[41] The martinize.py script was used to automatically generate the CG representation of these systems.[56] System 7 was simulated to compare the efficiency of interpolating charges and potentials. The interactions in this system were modeled with the OPLS force field[57] because the GROMACS 3.3 release, on which the linear interpolation of potentials implementation was based, does not support the CMAP correction that is needed for the CHARMM36 force field.[58]

The amino acids Glu, Asp, and His were modeled as tripeptides Ala-X-Ala with acetylated N-terminus (ACE) and *N*-methylamidated C-terminus (CT_3). The proteins were simulated with charged termini. The tripeptides were placed in a periodic rectangular box of dimensions $5 \times 5 \times 5$ nm^3 with approximately 4000 CHARMM TIP3P[59,60] water molecules in the AA simulations and 950 polarizable water particles in the CG simulations.[61] The water-soluble protein cardiotoxin V was placed in a periodic rectangular box of $7.9 \times 7.9 \times 7.9$ nm^3 and filled with 16500 CHARMM TIP3P water molecules in the AA simulations. In the CG simulations, the protein was placed inside a periodic rectangular box of $5.7 \times 5.7 \times 5.7$ nm^3 and filled with 1800 polarizable water particles. The larger water-soluble protein HEWL was placed in a periodic rectangular box of $8.9 \times 8.9 \times 8.9$ nm^3 and filled with 23000 CHARMM TIP3P water molecules in the AA simulations and 5400 polarizable water particles in the CG simulations. Na^+ and Cl^- ions were added to all systems at 0.15 M concentration to neutralize the protein systems. The turkey ovomucoid inhibitor protein was placed in a box of $4.9 \times 4.9 \times 4.9$ nm^3 with 3086 SPC[62] water molecules. The GLIC protein was embedded into a bilayer membrane containing 498 phosphatidylcholine (POPC) lipids, placed in a box of $14.0 \times 14.0 \times 15.9$ nm^3, and filled with 66494 CHARMM TIP3P waters, 58 Na^+, and 123 Cl^- ions. The system contained 292135 atoms in total. The simulation of this system was performed with the GROMACS 2021.4 release as reference. The GLIC benchmarks were run with default settings on an Intel i9−7920X 12-core CPU and an Nvidia RTX 2080 Ti GPU. All input configurations are provided as the Supporting Information.

In the AA simulations, Coulomb interactions were modeled with the smooth PME method with a real-space cutoff of 1.2 nm and a grid spacing of 0.14 nm,[44,45] while Lennard-Jones interactions were smoothly switched to zero in a range from 1.0 to 1.2 nm. In the CG simulations, Coulomb interactions were modeled by a Reaction Field potential with a 1.1 nm cutoff, $\epsilon_r = 2.5$, and $\epsilon_{RF} = \infty$, while Lennard-Jones interactions were truncated at 1.1 nm.[63] To keep the temperature constant at 300 K, we used the v-rescale thermostat[64] with time constants of 0.5 and 1.0 ps^{-1} for AA and CG simulations, respectively. The pressure was kept constant at 1 bar with the Parrinello−Rahman barostat[65] with relaxation times of 2.0 and 12.0 ps for AA and CG simulations, respectively. A leapfrog integrator was used with an integration step of 2 and 20 fs for AA and CG simulations, respectively. In the AA simulations, the LINCS algorithm[66] was used to constrain h-bond lengths of the solutes, while the SETTLE[67] algorithm was used to constrain internal degrees of freedom of the water molecules. Prior to the constant pH MD simulations, the potential energy of each system was minimized using the steepest descent method, followed by 1 ns of equilibration.

Constant pH MD Simulation Setups. In the atomistic simulations, the multisite representation was used to model the protonation states of titratable residues. Two λ-coordinates were

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

introduced to model the two forms of the carboxylic acid side chain in Asp and Glu, while three coordinates were used to describe the three protonation states of the imidazole side chain in His. In the CG simulations, the single-site representation was used, in which the A and B states represent the protonated and deprotonated states of the titratable beads. Because, in contrast to AA force fields, there is no distinction between the two neutral forms of the His side chain in the Martini force field, the single-site description for HIS suffices in the CG simulations. In both atomistic and coarse-grained simulations, the transformations between the different protonation states were achieved by changing the charges of the ionizable groups. The Lennard-Jones and bonded terms (bonds, angles, and torsions) were kept in the protonated and deprotonated states in AA and CG simulations, respectively. We show in Figure S3 that the contribution of these terms is sufficiently small to be neglected without significant error. We note, however, that these terms can be made λ-dependent as well, but this is beyond the scope of the current work since the efforts to implement this are high.

The mass of the λ-particles was set to 5 atomic units, and their temperature was maintained at 300 K by using a separate v-rescale thermostat for the λ-coordinates with a time constant of 2.0 ps^{-1}. For all λ-coordinates the biasing potential $V_i^{bias}(\lambda_i)$ was defined by equation S1 in the Supporting Information. The barrier height of the double-well potential was set to 5.0 and 7.5 kJ/mol for AA and CG simulations, respectively. The parameters for the double-well potential and the pH-dependent potential (eq 4) are provided in Table S1.

For the tripeptides, we calculated five independent CpHMD trajectories of 20 ns each at 13 pH values, ranging from 1.0 to 7.0 for the peptides with Glu and Asp, and from 4.0 to 10.0 for the peptides with His. For the cardiotoxin V protein (three Asp and one His titratable residues), we performed five independent CpHMD simulations of 50 ns at 15 pH values between 1.0 to 8.0. For the HEWL protein (seven Asp, two Glu, and one His titratable residues), we performed five independent CpHMD simulations of 75 ns at 21 pH values between −1.0 to 9.0. The values of the λ-coordinates were written to the output file with a frequency of 1 ps^{-1}.

Reference States and Force Field Correction Potentials. The constant pH simulations of the aforementioned systems require reference states for Asp, Glu, and His, in which the proton affinity (pK_a) is known from the experiment. The measured and calculated (force field) deprotonation free energies of these reference states were used to include the effect of the pH bath, $V^{pH}(\lambda)$, as well as the effects of the breaking and forming of chemical bonds in the simulation, i.e., V^{MM} in eq 2. The measured reference pK_a values used in this work are included in Table 1. Note that the experimental values were obtained for pentapeptides, while tripeptides were used for computing V^{MM}. This however did not affect the results, as shown in Figure S2.

Thermodynamic integration was used to compute the reference free energies as follows: the partial charges in tripeptide systems representing the reference states of Glu, Asp, and His were linearly interpolated between $\lambda = -0.1$ and $\lambda = 1.1$ with a step of 0.05 under the constraint $\lambda_1 + \lambda_2 = 1$ for Glu and Asp, while for His, the constraint was $\lambda_1 + \lambda_2 + \lambda_3 = 1$. For each set of λ values, called a grid point, a 10 ns MD simulation was performed. The $\partial V/\partial\lambda_i$ values were saved every ps, which is approximately equal to the autocorrelation times for the λ-coordinates. The total charge of the system was kept neutral by simultaneously changing the charge of a single buffer particle, as discussed below. The $\partial V/\partial\lambda_i$ values were averaged over the last 9 ns of the trajectories. To obtain an analytical expression for V^{MM}, a fifth-order polynomial was fitted to these averages for Asp and Glu, while an eighth-order polynomial was fitted for His, taking into account possible linear dependencies of the coefficients (see the Supporting Information). Fitting errors were below 0.5 kJ/mol for Asp and Glu and below 1 kJ/mol for His, which are of similar magnitude as the statistical accuracy of the derivatives.

Table 1. pK_a Values Obtained from Titration Simulations[a]

	tripeptide simulations[69,70]		
	pK_a values		
amino acid	CHARMM36	MARTINI	exp.
Asp	3.61 ± 0.03	3.69 ± 0.02	3.65
Glu	4.26 ± 0.04	4.30 ± 0.03	4.25
His macroscopic	6.34 ± 0.08	6.40 ± 0.03	6.42
His HSD	6.56 ± 0.06		6.53
His HSE	6.90 ± 0.05		6.94
	simulation of cardiotoxin V[71,72]		
	pK_a values		
amino acid	CHARMM36	MARTINI	exp.
His-4	5.14 ± 0.16	4.54 ± 0.09	5.5
Glu-17	4.08 ± 0.08	4.36 ± 0.04	4
Asp-42	4.02 ± 0.10	4.30 ± 0.05	3.2
Asp-59	2.41 ± 0.07	1.45 ± 0.03	<2
	r = 0.96	r = 0.80	
	MSE = 0.24	MSE = 0.64	
	RMSE = 0.49	RMSE = 0.80	
	simulation of HEWL[73]		
	pK_a values		
amino acid	CHARMM36	MARTINI	exp.
Glu-7	2.82 ± 0.07	4.86 ± 0.05	2.6
His-15	4.84 ± 0.05	5.42 ± 0.05	5.5
Asp-18	3.35 ± 0.05	3.31 ± 0.03	2.8
Glu-35	7.64 ± 0.13	6.36 ± 0.05	6.1
Asp-48	0.99 ± 0.07	3.36 ± 0.05	1.4
Asp-52	5.69 ± 0.12	7.18 ± 0.10	3.6
Asp-66	1.70 ± 0.10	5.22 ± 0.05	1.2
Asp-87	1.73 ± 0.03	3.47 ± 0.05	2.2
Asp-101	5.43 ± 0.11	4.20 ± 0.06	4.5
Asp-119	2.77 ± 0.05	3.80 ± 0.05	3.5
	r = 0.90	r = 0.49	
	MSE = 0.96	MSE = 4.01	
	RMSE = 0.98	RMSE = 2.00	

[a]The reference pK_a values for tripeptides are given in the last column that contain the experimental pK_a values. The values for Asp and Glu are taken from ref 69, while the microscopic and macroscopic pK_a values for His are taken from ref 70. Experimental pK_a values for cardiotoxin V are from refs 71 and 72 and for HEWL from ref 73. For both proteins Pearson correlation (r), MSE and RMSE errors are provided.

Buffer Particles. Dynamically changing partial charges can affect the total charge of the simulation unit cell, which can lead to artifacts, as documented for instance in Hub et al. for Ewald-based methods.[68] To avoid such artifacts, it is essential to keep the total charge of the unit cell constant. Two approaches have been proposed: (i) direct coupling between each titratable residue and a water,[27] or ion,[25] and (ii) titratable buffers that collectively compensate for changes in charge of all titratable residues.[42]

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

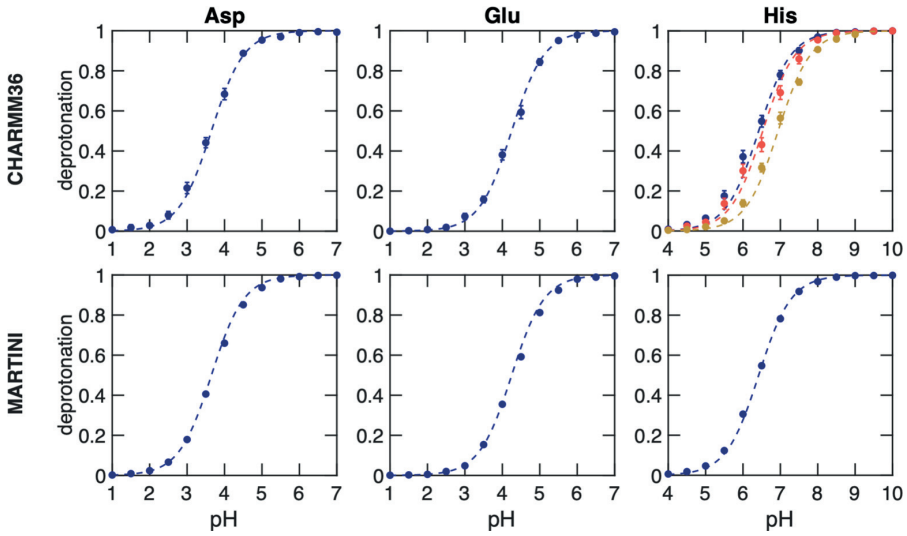


Figure 3. Titrations of tripeptide amino acids (Glu, Asp, and His) in water. The top and bottom rows show titrations performed with the modified AA CHARMM36[55] and CG Martini[41] force fields, respectively. In all simulations, neutrality was maintained by including ten buffer particles in combination with the charge constraint. Dots show the fraction of frames in which the residue is deprotonated, and the dashed lines represent the fits to the Henderson−Hasselbalch equation. For His, the blue color represents the macroscopic pK_a, while yellow and red represent the microscopic pK_a values for HSD (proton on N_δ) and HSE (proton on N_ϵ), respectively. In the Martini 2.0 model, HSD and HSE are indistinguishable and hence only the macroscopic titration curve is shown. Errors of S^{deprot} were estimated from the standard error of the mean for the different replicas. From the fits, the pK_a values were estimated and listed in Table 1.

Here, we follow the latter approach, but with several improvements for all-atom simulations. First, to avoid restraints, which were needed to minimize interactions between the buffers and the titrable sites in previous work,[42] we introduced buffer particles with both small LJ radius and small partial charges of maximal |0.5|e, such that they do not disturb the hydrogen bond network, nor interact too strongly with the titratable sites or other buffers. Second, to also prevent strong interactions with hydrophobic regions in the system, the $C^{(6)}$ dispersion parameter with anything other than water was set to zero, including the other buffers. The latter also avoids the clustering of buffers during the simulation. Thus, the buffer particles have an σ of 0.25 nm and an ϵ of 4 kJ/mol. Further details on buffer parametrization are provided in the accompanying paper.[55] In coarse-grained simulations, standard Na^+ ions were used as buffer particles.

As in Donnini et al.,[42] the buffers were collectively coupled to the titratable sites in the system via a charge constraint. The charges of all buffers were thus simultaneously interpolated between −0.5e and 0.5e, keeping the simulation box neutral. For all peptide simulations, 10 such buffers were introduced into the system, while 20 and 50 buffers were added to the simulation boxes with cardiotoxin V and HEWL proteins (systems 4−5), respectively, in both AA and CG models. 185 buffer particles were used in GLIC simulations.

Analysis of the Constant pH Trajectories. To estimate the pK_a values of titratable groups from multiple simulations at various pH values, we computed the average fraction of deprotonated frames (S^{deprot}) over all replicas. For a group with a single titratable site, this average was obtained as

$$S^{\text{deprot}}(\text{pH}) = \frac{N^{\text{deprot}}}{N^{\text{prot}} + N^{\text{deprot}}} \tag{18}$$

where N^{prot} and N^{deprot} are the total number of frames in which the site is protonated and deprotonated, respectively. For titratable sites modeled in the single-site representation, we considered it protonated if λ is below 0.2 and deprotonated if λ is above 0.8. For sites that are described with the multisite description, we considered a state protonated if the λ associated with the protonated form of the residue is above 0.8 and deprotonated if the λ associated with the deprotonated form of the residue is above 0.8.

To estimate the macroscopic pK_a values of histidine, which contains two titratable sites N_ϵ and N_δ, we calculated for each pH value the average fraction of frames in which the residue is deprotonated at either of the two sites:

$$S^{\text{deprot}}_{\text{macro}}(\text{pH}) = 1 - \frac{N^{\lambda_p}}{N^{\lambda_p} + N^{\lambda_\epsilon} + N^{\lambda_\delta}} \tag{19}$$

where N^{λ_p}, N^{λ_ϵ}, and N^{λ_δ} are the numbers of frames in which $\lambda_p > 0.8$, $\lambda_\epsilon > 0.8$, and $\lambda_\delta > 0.8$ (Figure 2). To estimate the microscopic pK_a values for the two sites of His, we calculated for each site the average fraction of frames in which that site was deprotonated:

$$S^{\text{deprot}}_{\text{micro}(\delta/\epsilon)}(\text{pH}) = \frac{N^{\lambda_\delta/\lambda_\epsilon}}{N^{\lambda_p} + N^{\lambda_\epsilon/\lambda_\delta}} \tag{20}$$

Errors were estimated from the standard error of the mean for the different replicas.

The averaged fractions at each pH value were fitted to the Henderson−Hasselbalch equation:

$$S^{\text{deprot}} = \frac{1}{10^{(\text{p}K_a - \text{pH})} + 1} \tag{21}$$

which yielded the pK_a values as fitting parameters. The error in the pK_a was estimated from the 95% confidence interval for the nonlinear least-squares fit to the average (S^{deprot}) values.

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

RESULTS AND DISCUSSION

Here we discuss the results obtained with our new implementation of constant pH into the fork of the GROMACS 2021 release.[49] While here our focus is on the validity and performance of the constant pH MD implementation, the convergence of the conformational and λ degrees of freedom are investigated systematically in the accompanying paper.[55]

Titration of Single Amino Acids. In Figure 3, we show the titration curves for AlaAspAla, AlaGluAla, and AlaHisAla tripeptides, obtained from simulations with the modified all-atom CHARMM36[55] and coarse-grained Martini 2.0 force fields.[41] Fitting the deprotonated fractions as a function of pH value to the Henderson−Hasselbalch equation (dashed lines in Figure 3) yields pK_a values for the tripeptides that are within 0.1 pK_a units from the reference values. Comparing the titration curves obtained with the Martini 2.0 force field in our implementation to those computed with the constant pH approach developed explicitly for this coarse-grained model,[34] our results suggest a much better agreement with the experiment than the latter. We attribute this difference to the more sophisticated explicit treatment of proton-like particles in the Martini constant pH approach. The rather good agreement between the titration curves obtained for both force fields on the one hand and the experiment on the other hand suggests that our implementation has little to no dependency on the force field, in line with the GROMACS philosophy of supporting a wide range of popular force fields.

Titration of Proteins. The titration curves of cardiotoxin V and HEWL proteins are shown in Figures 4 and 5, respectively. The pK_a values obtained from fitting the Henderson−Hasselbalch equation to the degree of deprotonation in the all-atom simulations of both proteins with the CHARMM36 force field, listed in Table 1, are in good agreement with previous constant pH MD simulations[30,74] and in reasonable agreement with experimental estimates from NMR spectroscopy [Pearson correlation coefficient (r) 0.96 and 0.9, RMSE 0.49 and 0.98 for cardiotoxin V and HEWL, respectively].[71−73] The pK_a values and S^{deprot} are converged in 50 ns, as discussed in the Supporting Information section 5 (Figures S6−S9). Analysis of the RMS deviation of the backbone and of the RMS fluctuation of the residues, plotted in Figures S10−S13, suggest no major influence of the pH on the structural stability of these proteins. The titration correlates with solvent exposure, which contributes to the stabilization of the charged protonation state (see Supporting Information section 7) (Figures S14−S16). The trends in the pK_a shifts are well reproduced, including the downshift of Asp-59 in cardiotoxin V, and, with the exception of Glu-35 and Asp-52 in HEWL, the deviations are below 1 pK_a unit. We note that also in previous constant pH simulations with the CHARMM force field,[22,30] similar deviations were found for these two residues (see Figure S5). This suggests that the origin of the discrepancy might lie beyond the implementation, and could be due to either a lack of sampling or systematic shortcomings in the force field, as was discussed in Huang et al.[30] For detailed insights into the structural origins of these pK_a shifts, we refer the reader to the paper of Swails and Roitberg.[19]

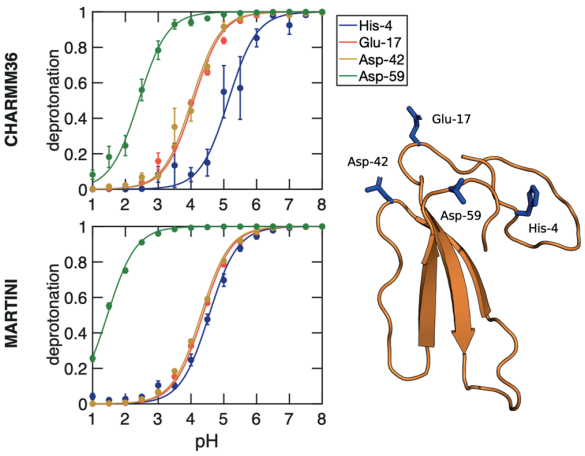


Figure 4. Titration curves of the cardiotoxin V protein obtained from constant pH MD simulations with the CHARMM36 (top) and Martini 2.0 force fields (bottom). For each of the four titratable residues in this protein, the dots show the fraction of frames in which the residue is deprotonated. Errors of S^{deprot} were estimated from the standard error of the mean for the different replicas. The lines show the best fits to the Henderson−Hasselbalch equation. The pK_a values for each titratable residue were estimated from these fits and listed in Table 1. The right panel shows the protein structure with the four titratable residues highlighted in stick representation.

The pK_a values estimated from the Martini 2.0 force field simulations of these proteins do not agree as well with the experiment as those derived from the all-atom simulations (Pearson correlation coefficient (r) 0.8 and 0.49; RMSE 0.8 and 2.0 for cardiotoxin V and HEWL, respectively). We speculate that the larger deviation of the pK_a's in the coarse-grained constant pH simulations could be due to the lower accuracy of the electrostatic interactions. Although we still consider the results obtained with the Martini simulations reasonable, in particular for the peptides, the discrepancies for the titratable residues in proteins suggest that additional parametrization efforts may be required to systematically improve the force field for constant pH MD simulations based on λ-dynamics. Such improvements would be particularly worthwhile considering coarse grained simulations pave the way to perform MD simulations of complete organelles,[75] in which many processes have a strong pH dependence.

Efficiency of the Implementation. To demonstrate that linear interpolation of charges is computationally more efficient than the linear interpolation of the potential energy functions for systems with many titratable sites, we investigated how the computational cost of a simulation scales with the number of titratable sites in the system for both approaches. Because we have implemented the interpolation of the charges rather than potential energy functions into the fork of GROMACS 2021 release, whereas the potential energy function interpolation was implemented in a fork of GROMACS 3.3 release, we compare the relative performances of both codes for an increasing number of titratable sites in the system. We define the relative performance as the ratio between the average number of integration steps per time unit for a simulation with constant pH on the one hand and the average number of integration steps per time unit for a normal simulation without constant pH on the other hand.

Figure 6A shows that the relative performance of constant pH simulations with charge interpolation does not decrease when the number of titratable sites included in the simulation increases. Most of the 30−40% drop in performance compared to a normal MD simulation with the same version of GROMACS is caused by the additional calculations and reductions in the nonbonded pair-interaction kernels that are

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

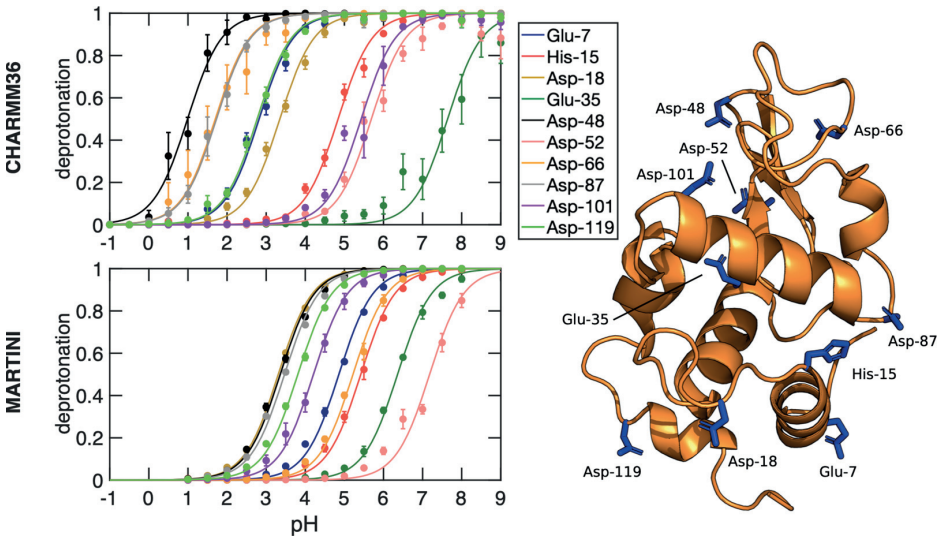


Figure 5. Titration curves of the HEWL protein obtained from constant pH MD simulations with the CHARMM36 (top) and Martini 2.0 force fields (bottom). For each of the ten titratable residues, the dots show the fraction of frames in which that residue is deprotonated. Errors of S^{deprot} were estimated from the standard error of the mean for the different replicas. The lines show the best fit to the Henderson-Hasselbach equation. The pK_a values for each titratable residue were estimated from these fits and listed in Table 1. The right panel shows the protein structure with the ten titratable residues highlighted in stick representation.

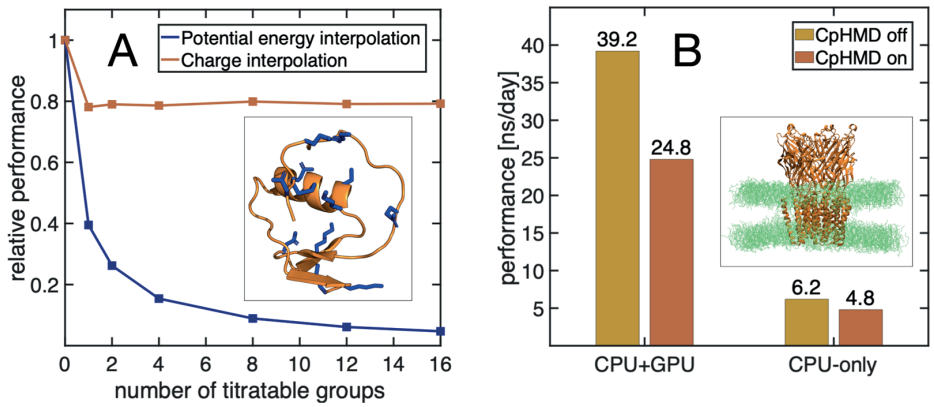


Figure 6. (A) Relative performance of interpolating potentials in a previous implementation of CpHMD into a fork of GROMACS 3.3 release (blue) and of charge interpolation in our new implementation (red) as a function of the number of titratable sites. The simulations were performed for the turkey ovomucoid inhibitor protein (PDB ID: 2GKR[53]), shown in the inset, where the titratable sites are highlighted in stick representation. (B) Comparison of the performance between CPU-only and CPU+GPU implementations for the ligand-gated ion channel GLIC (PDB ID: 4HFI[52]) with 185 titratable sites. In total, the GLIC system contained 292135 atoms.

required to obtain the real-space part of the electrostatic potential ($\Phi(\mathbf{R}_i, \boldsymbol{\lambda})$ in eq 17).

In contrast, the relative performance of constant pH simulations based on the linear interpolation of potential energy functions decreases with the number of titratable sites in the system. This comparison thus demonstrates that by replacing linear interpolation of potentials with linear interpolation of partial charges, we have overcome the major bottleneck in the earlier constant pH implementation in the fork of GROMACS 3.3 release and paved the way toward simulations of large biomolecular systems at constant pH.

An example of such a large system is the proton-gated ion channel GLIC, a membrane protein with 185 titratable residues. Figure 6B shows the performance of the new implementation for this large system when running the simulation on CPU and on a combination of CPU and GPU. While the computational overhead is somewhat larger when using a GPU in addition to a CPU, the overall performance still improves significantly when adding a GPU. We have also implemented a parallel version using MPI. For the GLIC system, the code scales up to 256 cores, on the Mahti supercomputer at CSC, with a performance of 42 ns/day, compared with 61 ns/day without constant pH.

CONCLUSIONS

We have presented and validated a new implementation of λ-dynamics based constant pH molecular dynamics in the GROMACS software. Our implementation combines several developments in this field into a single MD program, including the multisite representation of titratable groups,[46] charge interpolation,[21] Particle Mesh Ewald electrostatics,[30] and charge

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

constraints.[42] Test calculations on amino acids and proteins suggest that the new implementation is efficient, accurate, and agnostic to force fields. Combined with user-friendly parametrization protocols, presented in the accompanying paper,[55] we expect that this implementation will pave the way toward routinely including the effect of pH in biomolecular MD simulations.[76]

APPENDIX: CONSTRAINT ALGORITHM

We use constraints to restrict sampling to the correct protonation states in the multisite representation as well as to maintain a neutral charge of the simulation box. Both multisite and charge constraints keep the linear combination of a subset of λ-coordinates constant and are applied simultaneously. Thus, there are N_c constraint equations

$$\sigma^k(\lambda) = \sum_{i=1}^{N_{\text{sites}}} \alpha_i^k \lambda_i = C^k \tag{22}$$

for $k \leq N_c$. Here, $\boldsymbol{\lambda}$ is the vector of all λ_i-coordinates, and C^k is the value of constraint k, which can be zero. If $\sigma^k(\boldsymbol{\lambda})$ is a multisite constraint, $\alpha_i^k = 1$ for λ_i-coordinates that represent one of the protonation states of a residue, while $\alpha_i^k = 0$ for all other λ_i-coordinates. If $\sigma^k(\boldsymbol{\lambda})$ is a constraint for keeping the overall charge constant, $\alpha_i^k = \sum_j^{N_{\text{atoms}}^i} q_{j,i}^B - q_{j,i}^A$, with N_{atoms}^i the number of atoms whose charges change as a function of λ_i.

During a leapfrog integration step, all λ_i-coordinates are first propagated without constraints to their unconstrained new values λ_i^u, which do not fulfill the constraints in eq 22. To obtain the constrained λ_i^c values, we first connect the constrained and unconstrained λ_i values using the definition of $\sigma^k(\boldsymbol{\lambda})$:

$$\sigma^k(\lambda^c) = \sigma^k(\lambda^u) + \sum_{i=1}^{N_{\text{sites}}} \alpha_i^k [\lambda_i^c - \lambda_i^u] \tag{23}$$

Because the unconstrained and constrained λ_i-coordinates are also connected by the constraint forces $G_i^k = -\zeta^k \frac{\partial \sigma^k}{\partial \lambda_i}$, we have in addition that

$$\begin{aligned} \lambda_i^c &= \lambda_i^u + \sum_{k=1}^{N_c} G_i^k \frac{\Delta t^2}{m_i} = \lambda_i^u - \sum_{k=1}^{N_c} \zeta^k \frac{\partial \sigma^k}{\partial \lambda_i} \frac{\Delta t^2}{m_i} \\ &= \lambda_i^u - \sum_{k=1}^{N_c} \zeta^k \alpha_i^k \frac{\Delta t^2}{m_i} \end{aligned} \tag{24}$$

where ζ^k is the Lagrange multiplier for constraint k, m_i the fictitious mass of λ_i, and Δt the integration time step. Substituting eq 24 in 23 yields

$$\sigma^k(\lambda^c) = \sigma^k(\lambda^u) - \sum_{i=1}^{N_{\text{sites}}} \alpha_i^k \sum_{l=1}^{N_c} \zeta^l \alpha_i^l \frac{\Delta t^2}{m_i} \tag{25}$$

which after rearranging can be expressed as

$$\Delta\sigma^k = \sigma^k(\lambda^u) - \sigma^k(\lambda^c) = \sum_{l=1}^{N_c} \zeta^l \sum_{i=1}^{N_{\text{sites}}} \alpha_i^k \alpha_i^l \frac{\Delta t^2}{m_i} \tag{26}$$

The last expression can be rewritten in matrix form

$$\boldsymbol{\Delta\sigma} = (\Delta\sigma^1, ..., \Delta\sigma^{N_c})^T = \mathbf{A}\boldsymbol{\zeta} \tag{27}$$

where $\boldsymbol{\zeta} = (\zeta^1, \cdots, \zeta^{N_c})^T$ and

$$\mathbf{A} = \left(\sum_{i=1}^{N_{\text{sites}}} \alpha_i^k \alpha_i^l \frac{\Delta t^2}{m_i} \right) \tag{28}$$

Because the α_i^k coefficients remain the same, matrix $\mathbf{A}$ is computed once at the start of the simulation. At each step, the elements $\Delta\sigma^k$ are evaluated as the difference between the $\sigma^k(\lambda^u)$ and $\sigma^k(\lambda^c) = C^k$:

$$\Delta\sigma^k = \sigma^k(\lambda^u) - C^k \tag{29}$$

The Lagrange multipliers ζ^k are then obtained from eq 27 and used to correct the unconstrained λ_i values (eq 24).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00516.

The input files and parameters used for MD simulations in the presented work (ZIP)

A Mathematica notebook with instructions and routines to fit V^{MM} (ZIP)

Description of λ-potentials, the effect of neglecting the interpolation of Lennard-Jones interactions, titration results for Asp and Glu within the single-site representation, a comparison of pK_a values for HEWL obtained with various λ-dynamics-based constant pH methods, demonstration that charge interpolation requires a single evaluation of the electrostatic potential for both single- and multisite representations (PDF)

AUTHOR INFORMATION

Corresponding Authors

Noora Aho − *Nanoscience Center and Department of Chemistry, University of Jyväskylä, 40014 Jyväskylä, Finland*; Email: noora.s.aho@jyu.fi

Pavel Buslaev − *Nanoscience Center and Department of Chemistry, University of Jyväskylä, 40014 Jyväskylä, Finland*; orcid.org/0000-0003-2031-4691; Email: pavel.i.buslaev@jyu.fi

Gerrit Groenhof − *Nanoscience Center and Department of Chemistry, University of Jyväskylä, 40014 Jyväskylä, Finland*; orcid.org/0000-0001-8148-5334; Email: gerrit.x.groenhof@jyu.fi

Berk Hess − *Department of Applied Physics and Swedish e-Science Research Center, Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden*; Email: hess@kth.se

Authors

Anton Jansen − *Department of Applied Physics and Swedish e-Science Research Center, Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden*

Paul Bauer − *Department of Applied Physics and Swedish e-Science Research Center, Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden*; orcid.org/0000-0002-2268-0065

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00516

Author Contributions

N.A. and P.B. contributed equally.

Notes

The authors declare no competing financial interest.

The fork of GROMACS 2021 with constant pH implemented as described here, is available for download free of charge from https://gitlab.com/gromacs-constantph/constantph. In addi-

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

tion to the source code, also instructions on how to set up and perform MD simulations are available

■ ACKNOWLEDGMENTS

This research was supported by the Swedish Research Council (Grant 2019-04477), Academy of Finland (Grants 311031 and 332743), and the BioExcel CoE (Grant H2020-INFRAEDI-02-2018-823830). The simulations were performed on resources provided by the CSC — IT Center for Science, Finland, and the Swedish National Infrastructure for Computing (SNIC 2021/1-38). We also thank Erik Lindahl, Helmut Grubmüller, Dmitry Morozov, Serena Donnini, and Plamen Dobrev for their support during the project.

■ REFERENCES

(1) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129−1143.
(2) Groenhof, G.; Modi, V.; Morozov, D. Observe while it happens: catching photoactive proteins in the act with non-adiabatic molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 106−112.
(3) Warshel, A.; Sussman, F.; King, G. Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry* **1986**, *25*, 8368−8372.
(4) Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Erik Nielsen, J.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. Progress in the prediction of pKa values in proteins. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3260−3275.
(5) Chen, W.; Morrow, B. H.; Shi, C.; Shen, J. K. Recent development and application of constant pH molecular dynamics. *Mol. Simul.* **2014**, *40*, 830−838.
(6) Zeng, X.; Mukhopadhyay, S.; Brooks, C. L., III Residue-level resolution of alphavirus envelope protein interactions in pH-dependent fusion. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2034−2039.
(7) Law, S. M.; Zhang, B. W.; Brooks, C. L., III pH-sensitive residues in the p19 RNA silencing suppressor protein from carnation Italian ringspot virus affect siRNA binding stability. *Protein Sci.* **2013**, *22*, 595−604.
(8) Ellis, C. R.; Shen, J. pH-dependent population shift regulates BACE1 activity and inhibition. *J. Am. Chem. Soc.* **2015**, *137*, 9543−9546.
(9) Kim, M. O.; Blachly, P. G.; McCammon, J. A. Conformational dynamics and binding free energies of inhibitors of BACE-1: from the perspective of protonation equilibria. *PLoS computational biology* **2015**, *11*, e1004341.
(10) Sarkar, A.; Gupta, P. L.; Roitberg, A. E. pH-dependent conformational changes due to ionizable residues in a hydrophobic protein interior: The study of L25K and L125K variants of SNase. *J. Phys. Chem. B* **2019**, *123*, 5742−5754.
(11) Sarkar, A.; Roitberg, A. E. PH-Dependent Conformational Changes Lead to a Highly Shifted p K a for a Buried Glutamic Acid Mutant of SNase. *J. Phys. Chem. B* **2020**, *124*, 11072−11080.
(12) Baptista, A. M.; Martel, P. J.; Petersen, S. B. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. *Proteins: Struct., Funct., Bioinf.* **1997**, *27*, 523−544.
(13) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **2002**, *117*, 4184−4200.
(14) Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 469−480.
(15) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **2004**, *25*, 2038−2048.
(16) Meng, Y.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Biomolecules Using a Discrete Protonation Model. *J. Chem. Theory Comput.* **2010**, *6*, 1401−1412.
(17) Itoh, S. G.; Damjanovic, A.; Brooks, B. R. pH replica-exchange method based on discrete protonation states. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3420−3436.
(18) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341−1352.
(19) Swails, J. M.; Roitberg, A. E. Enhancing conformation and protonation state sampling of hen egg white lysozyme using pH replica exchange molecular dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 4393−4404.
(20) Börjesson, U.; Hünenberger, P. H. Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. *J. Chem. Phys.* **2001**, *114*, 9706−9719.
(21) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III Constant-pH molecular dynamics using continuous titration coordinates. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 738−752.
(22) Khandogin, J.; Brooks, C. L. Constant pH Molecular Dynamics with proton tautomerism. *Biophys. J.* **2005**, *89*, 141−157.
(23) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmüller, H. Constant pH molecular dynamics in explicit solvent with λ-dynamics. *J. Chem. Theory Comput.* **2011**, *7*, 1962−1978.
(24) Wallace, J. A.; Shen, J. K. Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange. *J. Chem. Theory Comput.* **2011**, *7*, 2617−2629.
(25) Wallace, J. A.; Shen, J. K. Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. *J. Chem. Phys.* **2012**, *137*, 184105.
(26) Goh, G. B.; Knight, J. L.; Brooks, C. L. Constant pH Molecular Dynamics Simulations of Nucleic Acids in Explicit Solvent. *J. Chem. Theory Comput.* **2012**, *8*, 36−46.
(27) Chen, W.; Wallace, J. A.; Yue, Z.; Shen, J. K. Introducing Titratable Water to All-Atom Molecular Dynamics at Constant pH. *Biophys. J.* **2013**, *105*, L15−L17.
(28) Lee, J.; Miller, B. T.; Damjanovic, A.; Brooks, B. R. Constant pH Molecular Dynamics in Explicit Solvent with Enveloping Distribution Sampling and Hamiltonian Exchange. *J. Chem. Theory Comput.* **2014**, *10*, 2738−2750.
(29) Dobrev, P.; Donnini, S.; Groenhof, G.; Grubmüller, H. Accurate Three States Model for Amino Acids with Two Chemically Coupled Titrating Sites in Explicit Solvent Atomistic Constant pH Simulations and p K a Calculations. *J. Chem. Theory Comput.* **2017**, *13*, 147−160.
(30) Huang, Y.; Chen, W.; Wallace, J. A.; Shen, J. All-atom continuous constant pH molecular dynamics with particle mesh Ewald and titratable water. *J. Chem. Theory Comput.* **2016**, *12*, 5411−5421.
(31) Huang, Y.; Harris, R. C.; Shen, J. Generalized Born based continuous constant pH molecular dynamics in Amber: Implementation, benchmarking and analysis. *J. Chem. Inf. Model.* **2018**, *58*, 1372−1383.
(32) Harris, R. C.; Shen, J. GPU-accelerated implementation of continuous constant pH molecular dynamics in amber: pKa predictions with single-pH simulations. *J. Chem. Inf. Model.* **2019**, *59*, 4821−4832.
(33) Harris, R. C.; Liu, R.; Shen, J. Predicting reactive cysteines with implicit-solvent-based continuous constant pH molecular dynamics in amber. *J. Chem. Theory Comput.* **2020**, *16*, 3689−3698.
(34) Grünewald, F.; Souza, P. C.; Abdizadeh, H.; Barnoud, J.; de Vries, A. H.; Marrink, S. J. Titratable Martini model for constant pH simulations. *J. Chem. Phys.* **2020**, *153*, 024118.
(35) Mongan, J.; Case, D. A. Biomolecular simulations at constant pH. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157−163.
(36) Damjanovic, A.; Miller, B. T.; Okur, A.; Brooks, B. R. Reservoir pH replica exchange. *J. Chem. Phys.* **2018**, *149*, 072321.
(37) Chen, Y.; Roux, B. Constant-pH Hybrid Nonequilibrium Molecular Dynamics - Monte Carlo Simulation Method. *J. Chem. Theory Comput.* **2015**, *11*, 3919−3931.

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(38) Kong, X.; Brooks, C. L., III λ-dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414−2423.
(39) Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: p K a calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167−4180.
(40) Chen, W.; Shen, J. K. Effects of System Net Charge and Electrostic Truncation on All-Atom Constant pH Molecular Dynamics. *J. Comput. Chem.* **2014**, *35*, 1986−1996.
(41) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(42) Donnini, S.; Ullmann, R. T.; Groenhof, G.; Grubmüller, H. Charge-neutral constant pH molecular dynamics simulations using a parsimonious proton buffer. *J. Chem. Theory Comput.* **2016**, *12*, 1040−1051.
(43) Dobrev, P.; Vemulapalli, S. P. B.; Nath, N.; Griesinger, C.; Grubmüller, H. Probing the accuracy of explicit solvent constant pH molecular dynamics simulations for peptides. *J. Chem. Theory Comput.* **2020**, *16*, 2561−2569.
(44) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(45) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(46) Knight, J. L.; Brooks, C. L., III Multisite λ dynamics for simulated structure-activity relationship studies. *J. Chem. Theory Comput.* **2011**, *7*, 2728−2739.
(47) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks, C. L., III Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 1319−1331.
(48) Knight, J. L.; Brooks, C. L., III Applying efficient implicit nongeometric constraints in alchemical free energy simulations. *Journal of computational chemistry* **2011**, *32*, 3423−3432.
(49) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19−25.
(50) Singhal, A. K.; Chien, K. Y.; Wu, W. G.; Rule, G. S. Solution structure of cardiotoxin V from Naja naja atra. *Biochemistry* **1993**, *32*, 8036−8044.
(51) Ramanadham, M.; Sieker, L.; Jensen, L. Refinement of triclinic lysozyme: II. The method of stereochemically restrained least squares. *Acta Crystallographica Section B: Structural Science* **1990**, *46*, 63−69.
(52) Sauguet, L.; Poitevin, F.; Murail, S.; Van Renterghem, C.; Moraga-Cid, G.; Malherbe, L.; Thompson, A. W.; Koehl, P.; Corringer, P.-J.; Baaden, M.; Delarue, M. Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. *EMBO journal* **2013**, *32*, 728−741.
(53) Lee, T.-W.; Qasim, M., Jr; Laskowski, M., Jr; James, M. N. Structural insights into the non-additivity effects in the sequence-to-reactivity algorithm for serine peptidases and their inhibitors. *Journal of molecular biology* **2007**, *367*, 527−546.
(54) Huang, J.; MacKerell, A. D., Jr CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry* **2013**, *34*, 2135−2145.
(55) Buslaev, P.; Aho, N.; Jansen, A.; Bauer, P.; Hess, B.; Groenhof, G. Best practices in constant pH MD simulations: accuracy and precision. *J. Chem. Theory Comput.* **2022**, DOI: 10.1021/acs.jctc.2c00517.
(56) Martini-website, General purpose coarse-grained force field. http://cgmartini.nl/index.php/tools2/proteins-and-bilayers/204-martinize (accessed October 01, 2021).
(57) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(58) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D., Jr Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophysical journal* **2006**, *90*, L36−L38.
(59) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-induced forces between two hydrophilic groups. *J. Phys. Chem.* **1994**, *98*, 2198−2202.
(60) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(61) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable water model for the coarse-grained MARTINI force field. *PLoS computational biology* **2010**, *6*, e1000810.
(62) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J.*Intermolecular forces*; Springer, 1981; pp 331−342.
(63) De Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **2016**, *199*, 1−7.
(64) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(65) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(66) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997**, *18*, 1463−1472.
(67) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **1992**, *13*, 952−962.
(68) Hub, J. S.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.* **2014**, *10*, 381−390.
(69) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. pK values of the ionizable groups of proteins. *Protein science* **2006**, *15*, 1214−1218.
(70) Tanokura, M. 1H-NMR study on the tautomerism of the imidazole ring of histidine residues: I. Microscopic pK values and molar ratios of tautomers in histidine-containing peptides. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1983**, *742*, 576−585.
(71) Chiang, C.-M.; Chien, K.-Y.; Lin, H.-j.; Lin, J.-F.; Yeh, H.-C.; Ho, P.-l.; Wu, W.-g. Conformational change and inactivation of membrane phospholipid-related activity of cardiotoxin V from Taiwan cobra venom at acidic pH. *Biochemistry* **1996**, *35*, 9167−9176.
(72) Chiang, C.-M.; Chang, S.-L.; Lin, H.-j.; Wu, W.-g. The role of acidic amino acid residues in the structural stability of snake cardiotoxins. *Biochemistry* **1996**, *35*, 9177−9186.
(73) Webb, H.; Tynan-Connolly, B. M.; Lee, G. M.; Farrell, D.; O'Meara, F.; Søndergaard, C. R.; Teilum, K.; Hewage, C.; McIntosh, L. P.; Nielsen, J. E. Remeasuring HEWL pKa values by NMR spectroscopy: Methods, analysis, accuracy, and implications for theoretical pKa calculations. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 685−702.
(74) Lee, J.; Miller, B. T.; Damjanovic, A.; Brooks, B. R. Enhancing constant-pH simulation in explicit solvent with a two-dimensional replica exchange method. *J. Chem. Theory Comput.* **2015**, *11*, 2560−2574.
(75) Pezeshkian, W.; König, M.; Wassenaar, T. A.; Marrink, S. J. Backmapping triangulated surfaces to coarse-grained membrane models. *Nat. Commun.* **2020**, *11*, 2296.
(76) Wolfram Research, Inc.*Mathematica, Version 11.3*. Champaign: IL, 2018.

https://doi.org/10.1021/acs.jctc.2c00516
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

III

BEST PRACTICES IN CONSTANT PH MD SIMULATIONS: ACCURACY AND SAMPLING

by

Pavel Buslaev, Noora Aho, Anton Jansen, Paul Bauer, Gerrit Groenhof, Berk Hess. *Journal of Chemical Theory and Computation, 18(10),* 6134-6147, **2022.**

DOI: 10.1021/acs.jctc.2c00517

JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

Best Practices in Constant pH MD Simulations: Accuracy and Sampling

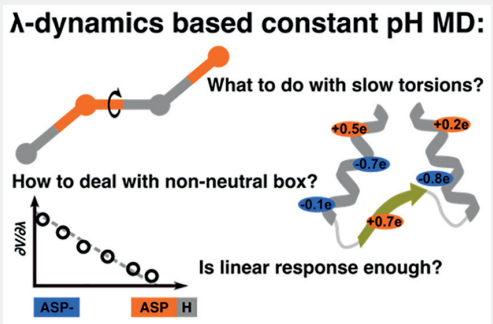
Pavel Buslaev,*,# Noora Aho,# Anton Jansen, Paul Bauer, Berk Hess,* and Gerrit Groenhof*

J. Chem. Theory Comput. **2022**, 10.1021/acs.jctc.2c00516

Cite This: https://doi.org/10.1021/acs.jctc.2c00517 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Various approaches have been proposed to include the effect of pH in molecular dynamics (MD) simulations. Among these, the λ-dynamics approach proposed by Brooks and co-workers [Kong, X.; Brooks III, C. L. *J. Chem. Phys.* **1996**, **105**, 2414−2423] can be performed with little computational overhead and hfor each typeence be used to routinely perform MD simulations at microsecond time scales, as shown in the accompanying paper [Aho, N. et al. *J. Chem. Theory Comput.* **2022**, DOI: 10.1021/acs.jctc.2c00516]. At such time scales, however, the accuracy of the molecular mechanics force field and the parametrization becomes critical. Here, we address these issues and provide the community with guidelines on how to set up and perform long time scale constant pH MD simulations. We found that barriers associated with the torsions of side chains in the CHARMM36m force field are too high for reaching convergence in constant pH MD simulations on microsecond time scales. To avoid the high computational cost of extending the sampling, we propose small modifications to the force field to selectively reduce the torsional barriers. We demonstrate that with such modifications we obtain converged distributions of both protonation and torsional degrees of freedom and hence consistent pK_a estimates, while the sampling of the overall configurational space accessible to proteins is unaffected as compared to normal MD simulations. We also show that the results of constant pH MD depend on the accuracy of the correction potentials. While these potentials are typically obtained by fitting a low-order polynomial to calculated free energy profiles, we find that higher order fits are essential to provide accurate and consistent results. By resolving problems in accuracy and sampling, the work described in this and the accompanying paper paves the way to the widespread application of constant pH MD beyond pK_a prediction.



INTRODUCTION

Thanks to improvements in algorithms, force fields, and computer hardware, molecular dynamics (MD) simulations have become a versatile tool for investigating the conformational landscape of complex biomolecular systems at the atomic level.[1−5] An important algorithmic improvement has been the explicit inclusion of pH in MD simulations,[6−12] as pH is an important experimental parameter that affects the structure and dynamics of biomolecules. To provide the users of the GROMACS MD package[13] with access to simulations at constant pH, we have implemented the λ-dynamics-based constant pH approach by Brooks and co-workers.[9,14] In contrast to the previous implementation in a fork of GROMACS 3.3,[10] the new implementation, which is described in an accompanying paper,[15] is efficient and constant pH MD simulations can be performed with about 25% for CPU, 30−40% for CPU + GPU computational overhead compared to normal MD simulations irrespective of the number of titratable sites in the system. In the accompanying methodological paper,[15] we also demonstrate that the method can be successfully applied to calculate pK_a values of titratable sites in a protein. The purpose of this paper is to provide users with guidelines and recommendations on how to set up and perform constant pH MD simulations, including the necessary parametrization steps.

In constant pH MD simulations, titratable groups can dynamically change their protonation state. These changes are driven by interactions between the group and the chemical environment (modeled with a force field) and the aqueous proton concentration (modeled with a pH potential). Because at the force field level a number of contributions to the free energy of (de)protonation are not included explicitly, i.e., quantum mechanical interactions associated with bond breakage and formation as well as the actual proton particle, corrections to the force field are needed in λ-dynamics-based

Received: May 18, 2022

Table 1. Table of Simulated Systems[a]

system	box size (nm^3)	no. of waters	no. of ions	force field
BUF_1	5 × 5 × 5	~4000	11 Na, 11 Cl, 2 Buf	CHARMM36m
BUF_2	3 × 3 × 3	~3000	1 or 2 Buf	CHARMM36m
ADA	5 × 5 × 5	~4000	11 Na, 11 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
ADA_3	3 × 3 × 3	~850	2 Na, 2 Cl, 10 Buf	CHARMM36m-cph
ADA_7	7 × 7 × 7	~11100	31 Na, 31 Cl, 10 Buf	CHARMM36m-cph
$ADA_{low\ salt}$	5 × 5 × 5	~4000	4 Na, 4 Cl, 10 Buf	CHARMM36m-cph
$ADA_{high\ salt}$	5 × 5 × 5	~4000	38 Na, 38 Cl, 10 Buf	CHARMM36m-cph
AEA	5 × 5 × 5	~4000	11 Na, 11 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
AKA	5 × 5 × 5	~4000	11 Na, 12 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
AHA	5 × 5 × 5	~4000	11 Na, 12 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
AAA_1	5 × 5 × 5	~4000	11 Na, 12 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
AAA_2	5 × 5 × 5	~4000	11 Na, 11 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
$1CVO_1$	6.8 × 7.8 × 7.1	~12400	35 Na, 48 Cl, 20 Buf	CHARMM36m, CHARMM36m-cph
$1CVO_2$	7.1 × 7.1 × 7.1	~11400	13 Cl, 150 Buf	CHARMM36m
$MEMB_1$	5.9 × 5.9 × 8.1	~4200	8 Na, 8 Cl, 50 Buf	CHARMM36m
$MEMB_2$	5.9 × 5.9 × 8.1	~4200	8 Na, 8 Cl, 1 Buf	CHARMM36m
SOL	4.3 × 4.3 × 4.3	~2600	50 Buf	CHARMM36m

[a]We denote the modified CHARMM36m force field as CHARMM36m-cph.

constant pH MD. In GROMACS, these corrections are implemented as analytical functions, $V^{MM}(\lambda_j)$, fitted to the free energy profile associated with the deprotonation of a titratable residue j at the force field level.[15]

The accuracy of such free energy profiles depends not only on how closely the force field model represents the true potential energy surface but also on the convergence of sampling of all other degrees of freedom in the system. Therefore, whereas in normal MD the accuracy of the dynamics depends solely on the quality of the force field, the accuracy of λ-dynamics-based constant pH MD depends additionally on whether all relevant degrees of freedom are sampled sufficiently in the simulations required for parametrizing the correction potentials.

We found that insufficient sampling of the dihedral degrees of freedom in the amino acid side chains can lead to poor convergence in the deprotonation free energy profiles, as also observed by Klimovich and Mobley in simulations without constant pH.[16] We traced the lack of the dihedral sampling to the barriers that separate the minima in the torsion potentials. These barriers are too high to reach a converged sampling of the dihedral free energy landscape on the time scales of typical constant pH MD simulations. Because the interaction between the titratable group and the environment depends critically on the dihedral angles of the side chain, a lack of convergence in these dihedral angles also affects the sampling of the protonation states.

Rather than increasing the time scale of the MD simulations to obtain converged dihedral and protonation state distributions or introducing enhanced sampling techniques,[12,17−21] we propose reducing the barriers for dihedral rotations in a systematic way. We will demonstrate that such optimized dihedral force field parameters improve pK_a estimates of amino acids without compromising the overall conformational sampling of the protein.

With a higher accuracy of the underlying deprotonation free energy profiles, we found that the correct sampling of protonation states also depends critically on the order of the polynomial fit used to obtain an analytical form for the correction potential. We show that the commonly accepted first-order fit,[12] although firmly based on linear response theory,[22] is not sufficiently accurate and can lead to erroneous protonation dynamics in constant pH MD simulations.

Because the dominant energetic contribution to proton affinity comes from electrostatic interactions,[22] it is of utmost importance to use an accurate description of such interactions. Constant pH MD simulations have been performed with various electrostatic models, including generalized Born,[9] shifted cutoff,[23] and particle mesh Ewald (PME).[10,12] Of these methods, the Ewald summation-based PME method[24,25] is generally considered to provide the most accurate description of the electrostatic interactions in periodic biomolecular systems.[26] Because Ewald summation can only provide accurate results if the simulation box remains neutral,[27] the charge fluctuations associated with the dynamic protonation and deprotonation in constant pH MD simulations need to be compensated to prevent artifacts. Titratable sites can be directly coupled to special particles, modeled as ions or water molecules,[21,28] such that charge is transferred directly between the titratable site and that particle. Alternatively, all sites can be coupled collectively to a sufficiently large number of buffer particles.[29] The latter approach has the advantage that spontaneous fluctuations in the interaction of the buffer particles with their environment affect all titratable sites to the same extent. The disadvantage is that the setup and parametrization of the buffer approach are more involved, as these require selecting the number of buffers and parametrizing their interaction with the rest of the system. To facilitate the use of buffers in constant pH MD, we provide a parametrization strategy aimed at preventing buffer clustering, buffer binding to titratable sites, and buffer permeation into hydrophobic regions. We demonstrate that buffers parametrized with this strategy also avoid finite-size effects associated with the periodicity of small simulation boxes.[30,31]

METHODS

Here, we go through all of the important aspects of the simulation setups.

Simulated Systems. We performed standard and constant pH MD simulations of the systems listed in Table 1. The original and modified (described in detail below) CHARMM36m[32,33] force fields were used in all simulations.

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

The table also presents the box size, the number of CHARMM36 TIP3P water molecules,[34−36] the number of ions, and buffer particles included in each system. The fitting coefficients of the V^{MM} correction potential for the buffer particles were obtained with system BUF_1. To find the optimal charge range and Lennard−Jones parameters for the buffer particles, enhanced sampling simulations with the accelerated weight histogram (AWH) method were performed on system BUF_2. Systems ADA, AEA, AKA, and AHA are alanine tripeptides with capped termini and as the central residue aspartic, glutamic, lysine, and histidine amino acids, respectively. AAA_1 and AAA_2 systems are alanine tripeptides with protonated termini. C- and N-termini were made titratable in AAA_1 and AAA_2 systems, respectively. Two sets of simulations of the cardiotoxin V protein were performed. System $1CVO_1$ was used to calculate the pK_a values of titratable residues, while the larger system $1CVO_2$ was used to compute the radial distribution function of the buffer particles around the protein. The membrane systems $MEMB_1$ and $MEMB_2$ contained 106 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) lipids. The starting coordinates and topologies of all systems are provided as Supporting Information.

Simulation Setup. Periodic boundary conditions were applied in all systems. Electrostatic interactions were modeled with the particle mesh Ewald method,[24,25] while van der Waals interactions were modeled with Lennard−Jones potentials which were smoothly switched to zero in the range from 1.0 to 1.2 nm. Simulations were performed at a constant temperature of 300 K, maintained by the v-rescale thermostat,[37] with a time constant of 0.5 ps^{-1} and at a constant pressure of 1 bar, maintained by the Parrinello−Rahman barostat,[38] with a period of 2.0 ps. The leapfrog integrator with an integration step of 2 fs was used. Bond lengths to hydrogens in the solute were constrained with the LINCS algorithm,[39] while the internal degrees of the CHARMM TIP3P water molecules[35,36] were constrained with the SETTLE algorithm.[40] Prior to the constant pH MD simulations, the energy of all systems was minimized using the steepest descent method, followed by a 1 ns equilibration.

Constant pH MD Simulation Setup. In the constant pH MD simulations, the mass of λ-particles was set to 5 atomic units and the temperature was kept constant at 300 K with a separate v-rescale thermostat for the λ degrees of freedom[37] with a time constant of 2.0 ps^{-1}. The single-site representation, defined and described in the accompanying paper,[15] was used for Asp, Glu, Lys, C-ter, and N-ter, whereas the multisite representation, also described in that paper, was used for His.[15,41,42] The multisite representation models each physical state of groups with chemically coupled titratable sites with an independent λ-coordinate and takes chemical coupling into account by applying the linear constraint on these λ-coordinates requiring $\sum_{i\in\text{group}} \lambda_i^{\text{group}} = 1$.[15] The same pH and biasing potentials were used as in Aho et al.[15] In the sampling simulations of single titratable residues, the pH was set equal to the pK_a and the barrier height of the biasing potential was set to zero. The titration of the cardiotoxin V (PDB ID 1CVO[43]) protein was performed by running 10 independent replicas of 100 ns each for 15 equidistantly spaced pH values in the range from 1.0 to 8.0 using both the original and a modified CHARMM36m force field. In the titration simulations, the barrier height of the biasing potential was set to 7.5 kJ mol^{-1} for groups modeled with a single-site representation and to 5 kJ mol^{-1} for groups modeled with a multisite representation.

Reference Simulations. The constant pH simulations require a correction potential $V^{MM}(\lambda_j)$ for each titratable residue j. These correction potentials are the integrals of polynomial fits to the expectation value of $\langle \partial V/\partial\lambda \rangle_\lambda$ in reference state simulations at fixed λ-values.[15] Thus, after integration, an nth-order polynomial fit to $\langle \partial V/\partial\lambda \rangle_\lambda$ yields an $(n + 1)$th-order polynomial function that represents $V^{MM}(\lambda)$. However, in our implementation, the fit to $\langle \partial V/\partial\lambda \rangle_\lambda$ was used, rather than $V^{MM}(\lambda)$. We thus refer to the fitting order as the order of the polynomial fit to $\langle \partial V/\partial\lambda \rangle_\lambda$.

We performed the reference simulations as follows: The partial charges in the tripeptide systems were linearly interpolated between $\lambda = -0.1$ and $\lambda = 1.1$ with a step of 0.05. For His, all three λ coordinates were changed under the constraint $\lambda_1 + \lambda_2 + \lambda_3 = 1$. For each set of λ-values, called a grid point, we performed an 11 ns MD simulation in which the $\partial V/\partial\lambda_j$ were saved every picosecond and accumulated. The total charge of the system was kept neutral by simultaneously changing the charge of a single buffer particle. The fitting procedure is described in full detail in the accompanying paper.[15]

Dihedral Free Energy Profiles. Because the sampling of protonation states is tightly coupled to the sampling of side chain dihedral degrees of freedom, we computed the free energy profiles associated with the rotation of the dihedrals in the side chain of the central amino acid in the capped tripeptide systems (Table 1) by means of umbrella sampling.[44,45] As the first step, we performed 20 ns MD simulations with a time-dependent potential on the dihedral angle with a force constant of 418.4 kJ mol^{-1} rad^{-2}. The center of this potential was moved from 0° to 360° with a rate of 18° ns^{-1}. From these simulations, frames with dihedral angles closest to 0°, 10°, 20°, etc., were selected as references for the umbrella replicas. The difference between the dihedral angle in the selected frames and the target angle was always below 0.1°. Then, we performed 36 umbrella sampling simulations of 11 ns with a harmonic restraining potential centered at the reference dihedral angle and a force constant of 418.4 kJ mol^{-1} rad^{-2}. We used the WHAM procedure,[46] implemented in GROMACS,[47] to unbias these umbrellas and obtain free energy profiles associated with the full rotation of the dihedral angle.

Dihedral Potential Energy Profiles at the QM and MM Levels. To check the validity of the proposed force field modifications, we computed the potential energy profiles for the N−C_α−C_β−C_γ dihedral of aspartic acid with capped residues. The profiles were computed at both quantum mechanical (QM) and molecular mechanical (MM) levels. The QM profiles were computed at the MP2/6-31+G* level of theory using the Firefly QC package,[48] which is partially based on the GAMESS (US)[49] source code. The MM profiles were computed for both the original and the modified CHARMM36m force fields. The potential energy was computed for the N−C_α−C_β−C_γ dihedral angle with 10° increments. For each dihedral value, the structures were energy minimized prior to potential energy calculation.

Accelerated Weight Histogram Alchemical Simulations. Buffer particles are used in constant pH MD to maintain the neutrality of the simulated system. Ideally, buffers should not introduce any artifacts due to binding to titratable groups, binding to each other, or penetrating into hydrophobic regions.

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

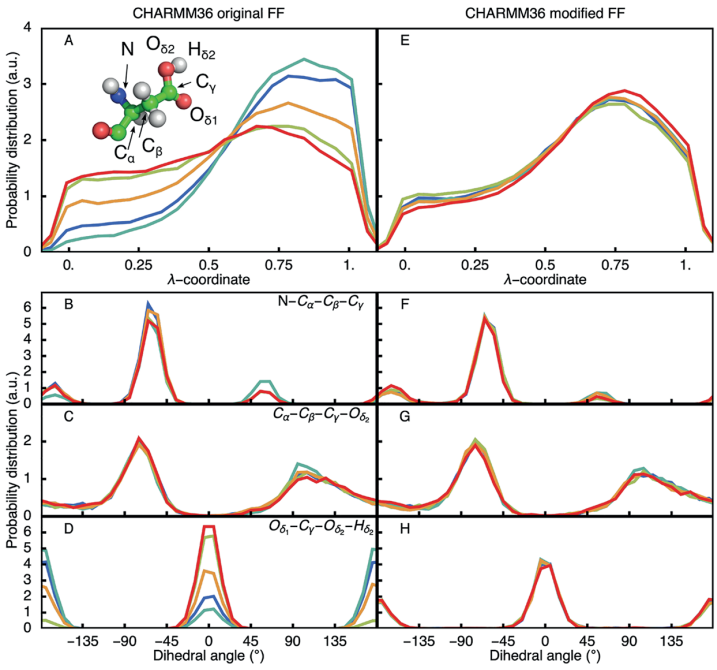


Figure 1. Distributions of the λ-coordinate (A and E) and dihedrals (B–D and F–H) in constant pH MD simulation of Asp with the original (A–D) and modified CHARMM36m (E–H) force field. While the simulations were performed for an ADA tripeptide, only the central aspartic acid is shown for clarity in the inset of A. (A and D) Distributions for third-order fits for $\langle \partial V/\partial\lambda \rangle_\lambda$ obtained with the original force field. Different colors correspond to independent replicas. Distributions for the λ-coordinate (A) as well as distributions for N–C_α–C_β–C_γ (B) and O_{δ_1}-C_γ-O_{δ_2}-H_{δ_2} (D) dihedrals are not identical. Right column (E–H) shows the distributions for the modified CHARMM36m force field with third-order fit for $\langle \partial V/\partial\lambda \rangle_\lambda$. Distributions are identical.

To prevent such behavior, we optimized the charge range and Lennard–Jones parameters of the buffers. To this end, we performed a series of enhanced sampling simulations with the accelerated weight histogram method (AWH).[50,51] In one set of simulations with two buffers in the simulation box (BUF_2, Table 1), we quantified the sampling efficiency from the friction metric[50,51] as a function of the absolute charge per buffer particle. In these simulations, the charge of one buffer was changed from 0 to +0.8 while simultaneously the charge of the other buffer was changed from 0 to −0.8 in order to maintain neutrality. The CHARMM36m Lennard–Jones parameters for sodium were used for the buffers in these simulations. In the other set of simulations, we computed the free energy difference between introducing a neutral buffer in water (BUF_2) and inside the hydrophobic region of a POPC bilayer system ($MEMB_2$, Table 1) for various values of the Lennard–Jones parameters of the buffer. In these simulations, the Lennard–Jones interactions between the buffer and the rest of the system were increased from noninteracting at $\lambda = 0$ to fully interacting ($\lambda = 1$) in 10 discrete steps. For all systems, we simulated 10 replicas of 10 ns, from which the free energy differences and friction coefficients were obtained as averages over the replicas.

Dihedral Analysis. Distributions of side chain dihedral angles in proteins were derived from publicly shared MD trajectories of proteins with PDB IDs 1U19,[52] 2RH1,[53] 2Y02,[54] and 5UEN[55] obtained from the GPCRMD[56] and SARS-CoV-2 databases (https://covid.molssi.org/).[57] For each trajectory, the distributions of the following dihedral angles were calculated:

(1) N–C_α–C_β–C_γ in aspartic acid
(2) N–C_α–C_β–C_γ in glutamic acid
(3) C_α–C_β–C_γ–C_δ in glutamic acid
(4) N–C_α–C_β–C_γ in histidine
(5) H–$O_{\epsilon 2}$–C–$O_{\epsilon 1}$ in aspartic and glutamic acids.

In this work, we also computed the distributions of these dihedrals from standard MD trajectories of cardiotoxin V (PDB ID 1CVO[43]).

Comparisons of Distributions. To compare two distributions $P_i(x)$ and $P_j(x)$, with the corresponding cumulative distribution functions $F_i(x)$ and $F_j(x)$, we used Kolmogorov–Smirnov statistics (KSS)[58]

$$\mathrm{KSS}(F_i, F_j) = \sup_x [F_i(x) - F_j(x)] \tag{1}$$

The larger the KSS, the less similar the distributions are. The distributions were considered consistent when $\mathrm{KSS}(F_i, F_j) < 0.03$. The KSS was computed using the script from the PCAlipids package.[59,60]

Titration. To estimate the pK_a values of titratable groups from multiple simulations at various pH values, we computed the average fraction of deprotonated frames (S^{deprot}) over all replicas at each pH value. For a group with a single titratable site, this average was obtained as

$$S^{\mathrm{deprot}}(\mathrm{pH}) = \frac{N^{\mathrm{deprot}}}{N^{\mathrm{prot}} + N^{\mathrm{deprot}}} \tag{2}$$

where N^{prot} and N^{deprot} are the total number of frames in which the site is protonated and deprotonated, respectively. For titratable sites modeled with the single-site representation, we

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

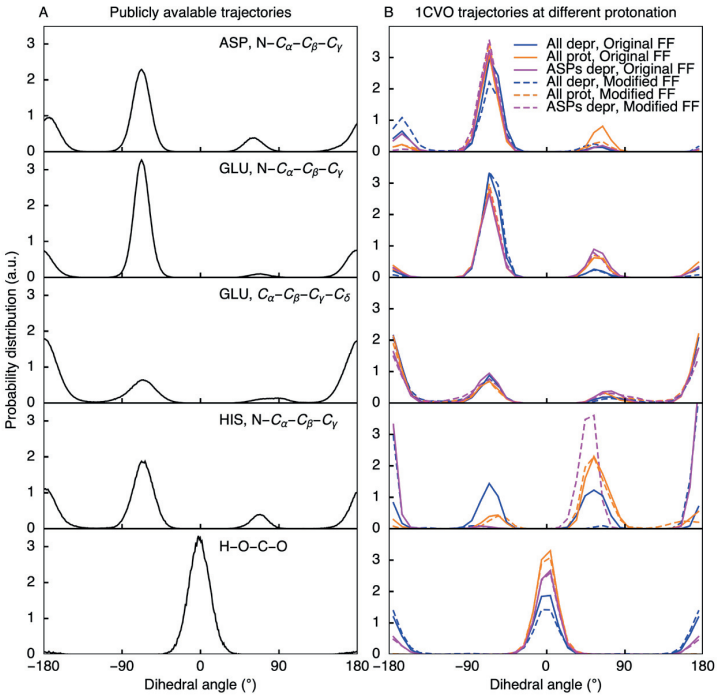


Figure 2. Distributions of dihedral angles for which the torsion potentials were corrected from standard MD simulations. (A) Dihedral distributions from publicly available trajectories[56,57] (total simulation time ≈ 10 μs). Probability density is significant only around local minima. (B) Dihedral distributions obtained from simulations of cardiotoxin V with both the original and the modified barriers for different protonation states of the titratable residues. Local minima are preserved, and no additional configurations are observed.

considered the site protonated if λ is below 0.2 and deprotonated if λ is above 0.8. For sites that are described with the multisite description, we considered a state protonated if the λ associated with the protonated form of the residue is above 0.8 and deprotonated if the λ associated with the deprotonated form of the residue is above 0.8.

To estimate the macroscopic pK_a value of histidine, which contains two titratable sites N_ϵ and N_δ, we calculated for each pH value the average fraction of frames in which the residue is deprotonated at either of the two sites

$$S_{\text{macro}}^{\text{deprot}}(\text{pH}) = 1 - \frac{N^{\lambda_p}}{N^{\lambda_p} + N^{\lambda_\epsilon} + N^{\lambda_\delta}} \tag{3}$$

where N^{λ_p}, N^{λ_ϵ}, and N^{λ_δ} are the number of frames in which $\lambda_p > 0.8$, $\lambda_\epsilon > 0.8$, and $\lambda_\delta > 0.8$.

The averaged fractions at each pH value were fitted to the Henderson–Hasselbalch equation

$$S^{\text{deprot}} = \frac{1}{10^{(\text{p}K_a - \text{pH})} + 1} \tag{4}$$

which yielded the pK_a values as fitting parameters.

RESULTS AND DISCUSSION

First, we demonstrate that a lack of sampling of the relevant dihedral degrees of freedom in amino acid side chains with the CHARMM36m force field reduces the accuracy of the correction potentials for λ-dynamics. To overcome these convergence problems, we modify the force field by reducing the barriers in the torsion potential and show that this significantly improves the accuracy of the correction potentials and hence the results of constant pH simulations, including pK_a estimates, without affecting the protein conformational dynamics. After the validation of the modified force field parameters, we show how the buffer particles that maintain the neutrality of the simulation box have to be parametrized to prevent finite-size effects on proton affinities due to periodicity.[30,31]

Sampling. Klimovic and Mobley have shown that calculated hydration free energies of single amino acids depend on the starting conformation.[16] Because a few picoseconds typically suffice to sample bond and angle degrees of freedom in the amino acid as well as the rotational degrees of freedom of the water molecules, we speculate that their observation implies a lack of sampling in the dihedral degrees of freedom in the amino acid side chain. Therefore, we systematically analyzed the convergence of both λ-coordinates and dihedral angles during constant pH simulations of single amino acids in water.

We performed 100 ns constant pH MD simulations at pH = pK_a of systems ADA, AEA, AKA, AHA, AAA_1, and AAA_2 (Table 1). To enhance the sampling of the λ-coordinate in these systems, we ran the simulations without a barrier in the biasing potential ($V^{\text{bias}}(\lambda)$, eq 5 in Aho et al.[15]). The correction potential ($V^{\text{MM}}(\lambda)$, eq 5 in Aho et al.[15]) was obtained by fitting a third-order polynomial function to the $\langle \partial V/\partial\lambda \rangle_\lambda$ values of the reference trajectories. We will show later that for accurate and reproducible constant pH MD results, a higher order fit is required. Nevertheless, in spite of its limited accuracy, using the same third-order fit for all system suffices to systematically compare the distributions of the relevant degrees of freedom and assess their convergence.

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

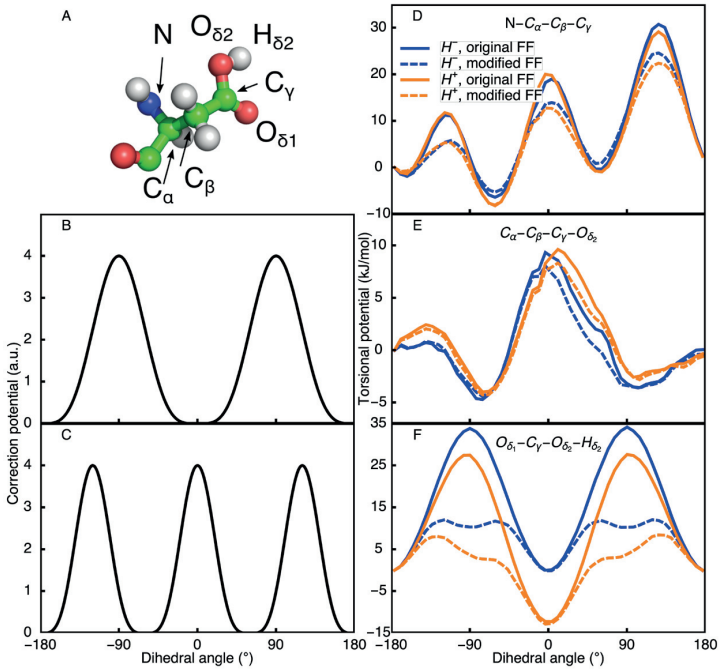


Figure 3. Modification of the torsional barriers in the Asp side chain. (A) Aspartic acid and its atomic nomenclature. (B and C) Corrections added to dihedral torsions of the original force field. Corrections with two (B) and three (C) local minima were used for torsions with two and three local minima, respectively. Heights of the corrections were selected through the iterative process, aimed at achieving consistent λ-distributions without introducing additional local minima in the free energy profiles. (D–F) Original and modified torsional barriers of Asp for both protonated (H^+) and deprotonated (H^-) states.

In Figure 1A, we show the distributions of the λ-coordinate in five constant pH MD replicas of the ADA system with the original CHARMM36m force field parameters. Distributions of the λ-coordinates in the other systems (AEA, AKA AHA, and AAA_1 and AAA_2) are shown in the Supporting Information (SI, Figures S1–S5). The dissimilarity between the λ-distributions in the replicas (maximum KSS between replicas of 0.29, 0.11, 0.04, and 0.095 for ADA, AEA, AHA, and AAA_1, respectively) indicates a lack of convergence. In addition, the distributions of the dihedral angles, shown in Figure 1B–D, are also not identical for all replicas. Because there is no barrier from the biasing potential for the λ-coordinate, we conclude that the lack of convergence in λ is due to insufficient sampling of the dihedral degrees of freedom.

Force Field Modifications. To overcome the lack of sampling, one can increase the simulation time or enhance sampling by means of special algorithms such as replica exchange MD.[61,62] Replica exchange has been applied frequently in the context of constant pH MD with exchanges between replicas at different temperature (T-REMD), pH (pH-REMD), or both.[20,63−66] However, because the protonation state has little effect on the torsion barrier height (Figure 3), changing the pH across the pH ladder would do little to enhance the sampling of the dihedral degrees of freedom. Furthermore, REMD methods are computationally more demanding than performing a single MD simulation and also prevent access to the dynamical properties of the system due to the jumps between replicas. As for some applications the dynamical properties may be highly needed, we consider it important that all relevant protonation states can be sampled correctly in a single constant pH MD trajectory. Because the sampling of λ-coordinates is tightly coupled to the sampling of the dihedral angles, convergence within a single trajectory can in principle also be reached by lowering the barriers of the torsion potentials.

Previously, modifications to the force field have been introduced for carboxyl groups. To improve the sampling of the syn and anti conformations of the carboxyl proton in Glu, Asp, and the C-terminus, Brooks and co-workers reduced the barrier for this rotation by a factor of 8 and also scaled the carboxyl oxygen radii by 0.95.[9] In contrast, Grubmüller and co-workers modified this torsion potential to prevent sampling the anti-conformation altogether.[67] However, according to our analysis (Figure 1), there is a lack of convergence not only in the carboxyl dihedral angle but also in the other side chain dihedral angles.

Reducing the torsional barriers without affecting the overall sampling of the conformational space is possible only if the regions near such barriers are sparsely sampled. We, therefore, analyzed the distributions of the dihedral angles in the side chains of titratable amino acids in the publicly available trajectories of G-protein-coupled receptors and of SARS-CoV-2 proteins.[56,57] The distributions of these dihedral angles, plotted in Figure 2A, reflect the shape of the underlying torsion potentials with maxima coinciding with local minima of the potential profiles. The low density near barriers suggests that these barriers are rather high and might be reduced without affecting the dihedral distributions.

To achieve convergence of both dihedral angles and λ-coordinates on a 100 ns time scale, we thus altered the maxima of the torsion potentials by adding a small dihedral

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

angle (ϕ)-dependent correction to the CHARMM36m force field

$$\Delta V_i(\phi) = \epsilon_i\left[\cos(n_i\phi) - (-1)^{n_i}\frac{1}{4}\cos(2n_i\phi)\right] \quad (5)$$

where n_i is the multiplicity of the torsion angle i (i.e., the number of minima) with $n_i = 2$ for conjugated bonds and $n_i = 3$ for aliphatic bonds. The parameter ϵ_i is an empirical coefficient that is optimized such that the barriers are low enough to converge the distribution of ϕ_i without introducing additional minima on the potential energy surface.

For each side chain dihedral angle of the titratable amino acids, the coefficient ϵ_i was optimized in an iterative fashion: After an initial guess, we computed the free energy profiles associated with rotation of the dihedral as well as five unbiased 100 ns trajectories at pH = pK_a with different starting conditions and a biasing potential ($V^{bias}(\lambda)$, eq 5 in Aho et al.[15]) without barrier. Prior to these constant pH MD simulations, we recomputed the correction potential, $V^{MM}(\lambda)$, by fitting a third-order polynomial to the $\langle\partial V/\partial\lambda\rangle_\lambda$ values obtained from thermodynamic integration simulations performed with the current value of ϵ_i. Free energy profiles were inspected visually for artificial minima, while distributions of both dihedral angles and λ-coordinates were compared between the five unbiased replica runs based on their similarity. The coefficient ϵ_i was gradually increased until the distributions in the different replicas were sufficiently similar (KSS < 0.03), while at the same time no additional minima appeared in the free energy profiles.

In Figure 3 we show the optimized torsion potentials as well as their effect on the free energy profiles of the dihedral angles in Asp. The corrections and free energy profiles of Glu, His, and the C-terminus are shown in Figures S6−S8 of the SI. With the exception of the C_β−C_γ−O_δ−H dihedral, the corrections introduce no additional minima on the free energy profile of these torsions. Furthermore, as shown on the right panels of Figure 1, the distributions of the dihedrals and λ-coordinates are nearly indistinguishable for all replicas after correction. Because with the corrected potentials the Kolmogorov−Smirnov statistics for Asp, Glu, His, and the C-terminus are 0.028, 0.015, 0.027, and 0.022, we conclude that the corrections improve the convergence of both the λ-coordinates and the dihedral degrees of freedom in constant pH simulations. Note that although the distributions of the λ-coordinate are sufficiently similar, the sampling of these coordinates is not yet uniform (Figure 1E). We will show below that this discrepancy is due to the low order of the polynomial fit used for obtaining the correction potential $V^{MM}(\lambda)$.

With the exception of the O_{δ_1}−C_γ−O_{δ_2}−H dihedral, the corrections we propose here lead to changes in the torsion barrier of at most 16 kJ mol^{-1} (i.e., for the Glu N−C_α−C_β−C_γ torsion). For many biomolecular force fields, the parameters of the torsion potentials are obtained by fitting suitable periodic functions to energies evaluated at the MP2 level of theory.[68−71] The parameters for each type of torsional potential are simultaneously fitted for multiple amino acids. Therefore, the average root mean squared (RMS) difference between the torsional energy at the CHARMM36m level and that at the MP2 level of theory is on the order of 10 kJ mol^{-1}. The RMS deviation between the modified and the original torsion potentials is at most 8 kJ mol^{-1}, and the RMS deviation between the ab initio potential at the MP2/6-31+G* level and the N−C_α−C_β−C_γ torsion potential in ASP is reduced from 4 kJ mol^{-1} for the original CHARMM36m force field to 3.5 kJ mol^{-1} for the modified CHARMM36m force field (Figure S9). Therefore, we conclude that with the corrections of the torsion potentials, the modified force field provides an equally good fit to QM potential profiles as the original CHARMM36m force field.[33,71]

We also performed standard MD simulations and simulated five replicas for 100 ns for the two protonation states of the Asp tripeptide in water using both the original and the modified CHARMM36m force field parameters. Without the modifications, the local minima are not consistently sampled in all replicas (Figure S10). In contrast, with the corrections, identical distributions of the dihedral angles are obtained also in standard MD simulations. In addition, the modifications are essential to sample both syn and anti conformations of the carboxyl proton, in line experiment.[72] We note, however, that the correction required for sampling both of these conformations significantly alters the shape of the barrier (Figure 3F). Nevertheless, because of their low mass, proton can tunnel through such barriers, and therefore, we consider the shape and height of the torsional barrier less relevant for this specific dihedral than for the other dihedrals.

Finally, we demonstrate that the modifications do not alter the distributions of the dihedral angles in protein simulations. We performed MD simulations of the $1CVO_1$ system both with and without the modifications to the torsion potentials of titratable amino acids with either (i) all of these residues protonated, (ii) all deprotonated, or (iii) all Asp residues deprotonated and all other residues protonated. In Figure 2B, we plot the distributions of the dihedral angles for which corrections were introduced. The high similarity between the distributions suggests that the corrections do not lead to the sampling of different dihedral distributions, even if the relative weights of the minima are slightly altered, in particular, for the H−O−C−O dihedral. We conclude, therefore, that the corrections introduced to facilitate sampling of the dihedral and λ-coordinates do not significantly alter the protein conformational landscape and can hence be used to perform both normal and constant pH MD simulations.

Quality of the Correction Potential V^{MM}. *Reference Potential.* To verify that the modified torsion potentials overcome the convergence problems, we performed constant pH simulations at pH = pK_a and without a barrier in the biasing potential. Because with such a setup the potential energy profile for the λ-particle should be flat, we expected a uniform λ-distribution, provided that the dihedral degrees of freedom are sufficiently sampled. However, as shown in Figure 1E, the distributions are identical between replicas but not uniform despite the corrections to the torsion potentials.

Because both the pH-dependent potential $V^{pH}(\lambda)$ and the biasing potential are flat by construction at pH = pK_a, the deviations must originate from discrepancies between the correction potential $V^{MM}(\lambda)$ and the underlying free energy profile associated with deprotonation. The correction potential is obtained as a polynomial fit to the $\langle\partial V/\partial\lambda\rangle_\lambda$ values from thermodynamic integration simulations. Because linear response (LR) theory predicts a linear dependence between the hydration free energy and the magnitude of a (point) charge, a first-order fit has often been used to obtain the correction potential for constant pH MD.[9,12] However, even if the change in the charge dominates the free energy of changing the

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

protonation state, hydrogen-bond rearrangements can contribute as well. Because the effects due to such structural rearrangements are neglected in LR theories, we hypothesized that higher order fits may be necessary for obtaining sufficiently accurate correction potentials.

To test our hypothesis, we investigated the accuracy of the polynomial fit to the correction potential. In Figure 4A, we

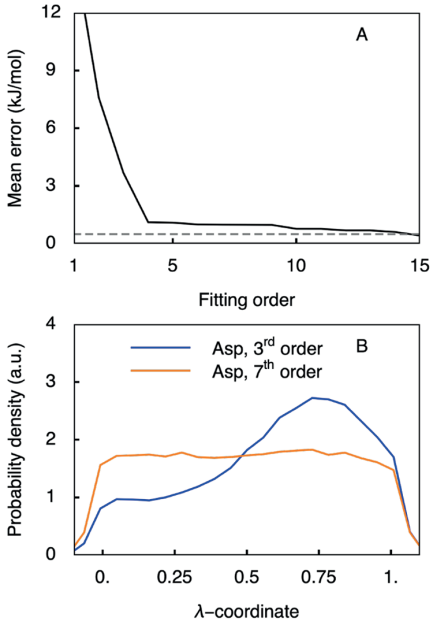


Figure 4. Quality of the $V^{MM}(\lambda)$ correction potential as a function of the order of the polynomial fit to $\langle \partial V/\partial \lambda \rangle_\lambda$ for Asp. (A) Fitting error as a function of fitting order (black line). Gray dashed line shows the average error in the calculated $\langle \partial V^{MM}/\partial \lambda \rangle$. (B) Distributions of λ-coordinates for the third- and seventh-order polynomial fits to $\langle \partial V/\partial \lambda \rangle_\lambda$. Whereas with the lower order fit the distribution is significantly rugged, the distribution becomes nearly flat and uniform on the [0, 1] interval of the λ-coordinate if a seventh-order fit is used.

show the mean error of the correction potential with respect to the computed free energy difference associated with deprotonation as a function of the fitting order. For the LR approximation the fitting errors are higher than 10 kJ/mol. In the worst-case scenario, such errors could lead to deviations in predicted pK_a values of more than one pK_a unit. With an error of 4 kJ mol^{-1}, the third-order fit, used above to address the convergence issues, does not yield a sufficiently accurate representation of the underlying free energy profile. Increasing the order of the polynomial fit reduces this error, and as shown in Figure 4B, at least a seventh-order fit is required to provide a uniform distribution of the λ coordinate for the Asp tripeptide in constant pH MD simulations at pH = pK_a.

Also, for carboxyl groups in the side chains of Glu and in the C-terminus, a polynomial fit to $\langle \partial V/\partial \lambda \rangle_\lambda$ of at least seventh-order is needed to provide a sufficiently accurate correction potential (Figure 4 and Figures S11 and S14 in SI). For the imidazole ring of His with three coupled titratable sites, a seventh-order fit suffices as well (Figure S13), while for the amino bases in the side chain of Lys and the N-terminus, at least an eighth-order fit is required (Figures S12 and S15 in SI). We speculate that the higher order fit is needed for the latter sites due to the larger change in the charge on the central nitrogen atom from −0.3 e to −0.96 e upon deprotonation. The change in the charge of the carboxylic oxygen from 0.55 e to −0.76 e is smaller as are the changes on the nitrogen atoms of the imidazole ring of His (from −0.36 e to −0.7 e).

Parameterization of Buffer Particles. A change in the protonation state affects the total charge of the simulated system, which can lead to artifacts when Ewald summation is used to treat the electrostatic interactions.[27,31] In our implementation of constant pH MD, we avoid this problem by introducing titratable buffers into the simulation box that compensate for the charge fluctuations of the titratable residues.[29]

In the original implementation of constant pH MD in GROMACS,[10] the buffers were hydronium molecules that compensated for the overall charge fluctuations by changing their charge between 0 and +1 e. To prevent sampling charges beyond this interval, a biasing potential with steep edges at λ = 0 and 1 was introduced to restrict the range of λ-values. However, this potential still introduces additional forces at the edges of the λ-interval. To avoid the effects of such forces, we use a completely flat biasing potential for the buffers, also outside of the charge interval.

Because changing the charge of a buffer particle in solution induces local rearrangements of the hydrogen-bonding network that in turn could affect the proton affinity of a nearby titratable group, we want to minimize the impact of charging the buffer particles. To determine the charge range in which the buffers do not cause significant hydrogen-bond network rearrangement, we ran AWH simulations with two ions, the charges of which are changed simultaneously in opposite directions (BUF_2 system). From the friction metric available in the AWH method,[50] we estimated the local diffusion coefficient, which is related to the efficiency of sampling: The higher the friction, the slower the dynamics and the more sampling is required to reach convergence. We calculated the friction coefficient (Figure 5A) for the coordinate associated with changing the charge on the buffer. For charges higher than 0.5 e, the friction was more than 50% higher than that for zero charges, reflecting longer correlation times and hence slower dynamics. We, therefore, conclude that the optimal range for the buffer charge is between −0.5 e and 0.5 e.

The collective λ-coordinate of the buffer particles is not restricted to a fixed interval by a wall-like potential. To avoid the buffer charge exceeding the optimal range, multiple buffer particles are needed in the simulation box. The optimal number of buffers can be calculated based on the analysis of charge fluctuations performed by Donnini et al.[29]

With a small charge, a buffer particle is apolar. To prevent clustering of such apolar particles in water, permeation into hydrophobic areas, such as membrane interiors, or interactions with the protein, the Lennard−Jones parameters (σ and ϵ) of the buffers were chosen such that the buffers have only repulsive interactions with all other atoms, except water. After experimenting with the parameters for the buffer particles, we settled on a σ of 0.25 nm and an ϵ of 4 kJ mol^{-1}. This choice leads to decreased clustering of buffers, low buffer concentrations in the proximity of titratable sites, and reduced penetration into hydrophobic regions (Figure 5). The resulting free energies of neutral buffer insertion into water and the hydrophobic region of the membrane are −2.09 ± 0.07 and 1.2 ± 0.6 kJ mol^{-1} compared to 8.45 ± 0.05 and 7.8 ± 0.3

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

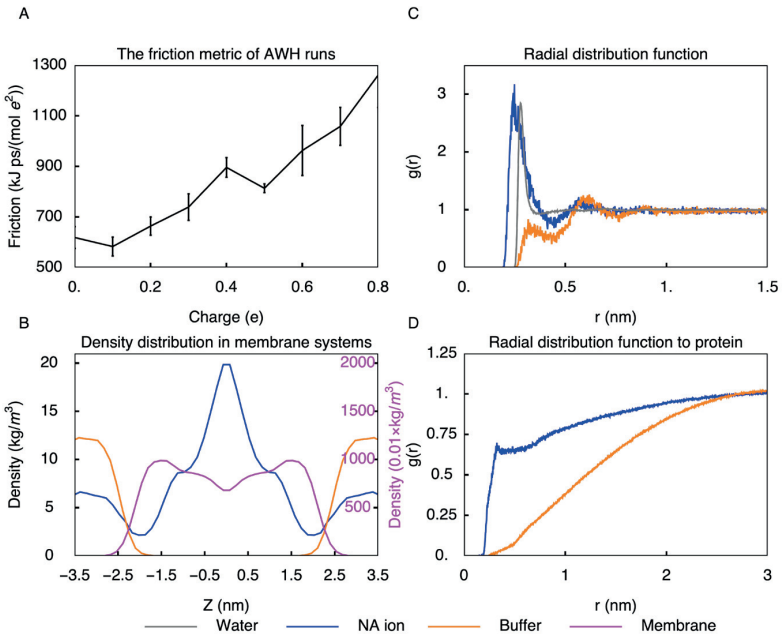


Figure 5. Parameterization of the buffer. (A) AWH friction metric as a function of buffer charge. Since the higher the friction metric is the slower the sampling, the buffers should ideally have low charges. (B) Density distribution of buffers in a membrane system. Optimized buffers do not penetrate into the lipid bilayer, while uncharged sodium ions do. (C) Radial distribution function between buffers. When standard sodium-ion parameters are used as buffers, the tendency to form clusters is high. Optimization of buffer parameters prevents clustering. (D) Radial distribution functions of buffers with respect to the protein. With the original sodium parameters the buffers have a higher tendency to bind to the protein than with the optimized parameters.

kJ mol^{-1} for insertion of an uncharged sodium ion into these regions.

While the primary goal of introducing buffers is to avoid the artifacts associated with a non-neutral periodic simulation box,[27] there can be other artifacts as well.[30,31] In particular, the undersolvation caused by solvent orientational restraints due to periodic boundary conditions[30] could lead to finite-size effects especially for small boxes. To understand if such finite-size effects affect the results of constant pH simulations, we investigated how the distribution of the λ-coordinates depends on the system size. We, therefore, performed constant pH MD simulations for three different box sizes and at various ionic strengths. All simulations were performed at pH = pK_a and without a barrier in the biasing potential. The uniformity of the distributions in these simulations, shown in Figure 6, suggests that for the box sizes tested, the finite-size effects are negligible.

Use Case: Consistent Protein Titrations. To demonstrate that with the modifications of the torsional barriers, a correction potential obtained by fitting at least a seventh-order polynomial to the $\langle \partial V/\partial \lambda \rangle_\lambda$ values of reference simulations, and buffer particles with optimized parameters, it is possible to perform accurate constant pH MD simulations, we calculated the pK_a values of all four cardiotoxin V titratable residues. We performed the pH titration simulations with both the original and the modified CHARMM36m force fields. In Figure 7, we show the titration curves obtained in the simulations and compare them to the experiment. Because there is no exact experimental estimate for the pK_a of ASP59, we only compare the pK_a values obtained for the other residues. The comparison suggests that the force field corrections improve the pK_a estimates, but more importantly, the lower deviation between

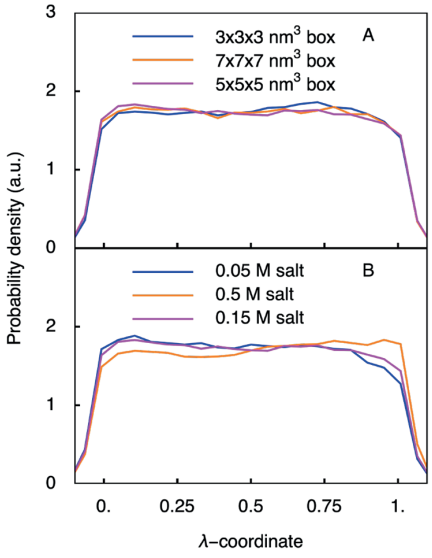


Figure 6. Effect of the box size (A) and ionic strength (B) on the distribution of the λ-coordinate of the ADA systems (ADA, ADA_3, ADA_7, $ADA_{low\ salt}$, $ADA_{high\ salt}$).

the individual replicas (from 0.12 to 0.07 for ASP42, from 0.16 to 0.05 for ASP59, from 0.1 to 0.03 for GLU17, and from 0.19 to 0.18 for HIS4, Figure 7) suggests that the sampling improves when the modified force field is used.

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

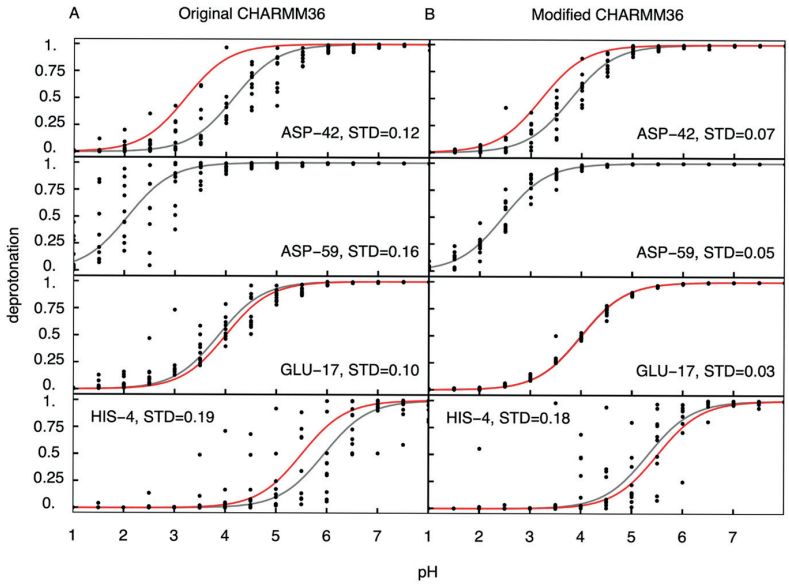


Figure 7. Titration of cardiotoxin V. (A and B) Titration curves of the titratable residues in $1CVO_1$ obtained from constant pH simulations performed with the original and modified CHARMM36 force fields, respectively. Black dots show the deprotonation ratio for the individual replica. Gray lines show the fits to the Henderson−Hasselbalch equation. Red lines are Henderson−Hasselbalch curves computed for the experimental pK_a value of the corresponding residue.[73,74] For ASP59, the exact pK_a is not known. For each curve, the standard deviation between the calculated and the fitted deprotonation ratio, averaged over all pH values and replicas, is shown.

Whereas for Asp and Glu the individual titration replicas are consistent when the modified force field is used, the replicas for HIS4 are not converged. As shown in Figure 8A, HIS4 interacts with TYR12 and PHE10. At pH = 5.5, close to the experimental pK_a of HIS4, we find three dominant conformations of HIS4-TYR12 pair in both normal and constant pH MD trajectories: A close contact (around 5 Å), a medium-range contact (around 6 Å), and a long-range contact (larger than 7 Å, Figure 8B and 8D). The distributions of these distances are not the same in all replicas (Figure 8E), and neither are the distributions of the λ-coordinates associated with the doubly protonated state of HIS4 (Figure 8F). These differences suggest a lack of convergence of the conformational dynamics of the protein in constant pH MD. To test if the protonation state of HIS4 correlates with the distance, we performed standard MD of cardiotoxin V with HIS4 in the three different protonation states (Figure. S16). Because there is no clear correlation between the protonation state and the HIS4-TYR12 distance, we cannot conclude that the protonation states are coupled to the conformation of the pair, at least not directly. Instead, the differences between the replicas suggest a lack of sampling of these conformations. Because the local environment differs between the states, we speculate that this lack of conformational sampling also affects the λ-distributions. We therefore can only conclude that the sampling of these distances would require more than 100 ns to converge. Thus, even if the corrections to the torsion potentials overcome the convergence issues associated with sampling the intrinsic dihedral degrees of freedom in single amino acids, reaching converged sampling of the protonation states in proteins may still require longer time scales if the inherent conformational dynamics is too slow.

Nevertheless, we observe that compared to normal MD, constant pH MD can increase the sampling of the local conformational space of the protein. In Figures 8C and S17 we show that the HIS4 samples configurations more efficiently in constant pH simulations. Specifically, the hydrogen bond between PHE10 and the HIS4 δ-hydrogen is much more stable in normal MD with a fixed protonation of δ-nitrogen (Figure S17), whereas in constant pH MD, the HIS4 also samples configurations in which the H-bond is broken (Figure S17), in particular around pH = pK_a, as evidenced by the distribution of the $N-C_\alpha-C_\beta-C_\gamma$ dihedral angle in Figure 8C. Thus, by keeping the protonation states flexible, constant pH MD facilitates the sampling of local conformations, which in turn may lead to faster convergence of the global conformational sampling.

CONCLUSIONS

It is now possible to run accurate constant pH molecular dynamics simulations on time scales of normal simulations, for example, with the new implementation in the GROMACS package presented in the accompanying paper.[15] Here, we have addressed the accuracy of constant pH simulation at longer time scales. We could demonstrate, on the basis of the CHARMM36m force field, that molecular force fields are not optimal for constant pH simulations because torsion barriers of titratable side chains are too high to reach convergence of the λ-coordinates associated with protonation. To overcome this sampling bottleneck, we proposed a systematic procedure to selectively reduce the barriers of the torsion potentials. In standard MD simulations, these modifications do not introduce noticeable artifacts but facilitate the convergence of side chain conformational sampling. Combined with the optimal fitting of V^{MM}, these force field modifications constitute an essential preparation step for constant pH simulations.

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

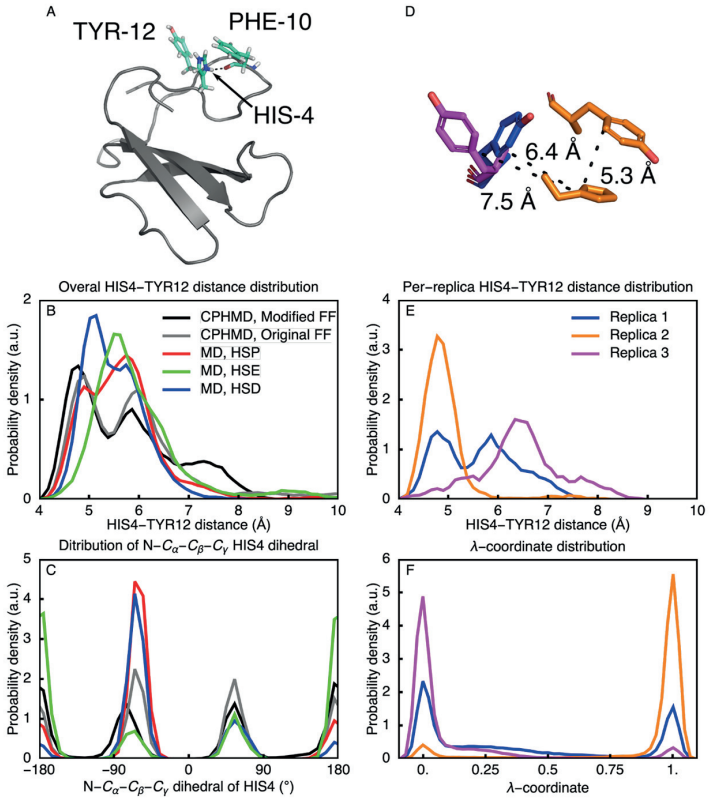


Figure 8. Sampling of cardiotoxin V N-terminal loop. (A) Structure of cardiotoxin V. Protein is shown in cartoon representation with HIS4, PHE10, and TYR12 shown as sticks. (B) Distributions of the distance between the centers of mass of HIS4 and TYR12 for all replicas combined. Combined distributions were calculated for constant pH MD with both the modified and the original force fields as well as for standard MD with HIS4 in all possible protonation states. (C) Distribution of HIS4 N–C_α–C_β–C_γ torsion. (D) Three observed conformations of TYR12 around HIS4. (E) Distributions of the distance between the centers of mass of HIS4 and TYR12 for individual replicas, computed for constant pH MD with the modified force field. (F) λ distributions of HIS4 protonated state for individual replicas of cardiotoxin V from the constant pH simulations with the modified force field.

The modifications of the CHARMM36m force field and the optimized parameters for the $\partial V^{MM}/\partial\lambda$ of correction potentials for Asp, Glu, His, Lys, and the C- and N-termini are available at https://gitlab.com/gromacs-constantph. Parameters for other force fields and residues will be made available there as well, once these are validated. We kindly ask the community to share with us also any parameters they may derive in their research, so that also these parameters can be made available.

The fork of GROMACS 2021 with constant pH MD implemented, as described here, is available for download free of charge from https://gitlab.com/gromacs-constantph/constantph. In addition to the source code, instructions on how to set up and perform MD simulations are available.

In addition to accurate parameters, it is also essential to keep the simulation system neutral during constant pH MD simulations. To achieve this, we introduced buffer particles with variable charges that dynamically compensate for the charge fluctuations of the titratable residues. To avoid that these buffers cluster, bind to the solute, disrupt hydrogen-bond networks, or penetrate into hydrophobic regions, we proposed a systematic parametrization procedure that can be used for any combination of force field and water model.

We expect the parametrization protocols proposed in this work to facilitate the application of constant pH MD not only within the user community of GROMACS but also of other MD programs as well. We also want to appeal to force field developers to take constant pH MD into consideration when developing their force fields.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00517.

Input files for MD simulations presented in this work (ZIP)

λ and angle distributions for Glu, His, Lys, and C- and N-termini; torsion corrections for Glu, His, and C-terminus; effects of force field modifications on dihedral dynamics and energy profiles; effects of fitting order on λ-distributions; conformations of HIS4 of cardiotoxin V (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Pavel Buslaev − *Nanoscience Center and Department of Chemistry, University of Jyväskylä, 40014 Jyväskylä,*

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Finland; orcid.org/0000-0003-2031-4691; Email: pavel.i.buslaev@jyu.fi

Berk Hess – *Department of Applied Physics and Swedish e-Science Research Center, Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden*; Email: hess@kth.se

Gerrit Groenhof – *Nanoscience Center and Department of Chemistry, University of Jyväskylä, 40014 Jyväskylä, Finland*; orcid.org/0000-0001-8148-5334; Email: gerrit.x.groenhof@jyu.fi

Authors

Noora Aho – *Nanoscience Center and Department of Chemistry, University of Jyväskylä, 40014 Jyväskylä, Finland*

Anton Jansen – *Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden*

Paul Bauer – *Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden*; orcid.org/0000-0002-2268-0065

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00517

Author Contributions

#P.B. and N.A. are co-first authors and contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by the Swedish Research Council (grant no. 2019-04477), Academy of Finland (grants 311031, 332743, and 342908), and BioExcel CoE (Grant No. H2020-INFRAEDI-02-2018-823830). The simulations were performed on resources provided by the CSC-IT Center for Science, Finland, and the Swedish National Infrastructure for Computing (SNIC 2021/1-38). We thank Dmitry Morozov for help with obtaining the torsion potentials at the MP2 level of theory.

■ REFERENCES

(1) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143.

(2) Groenhof, G.; Modi, V.; Morozov, D. Observe while it happens: catching photoactive proteins in the act with non-adiabatic molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 106–112.

(3) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **2020**, *153*, 134110.

(4) Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355–360.

(5) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T.; Dirks, R. M.; Dror, R. O.; Eastwood, M. P.; Edwards, B.; Even, A.; Feldmann, P.; Fenn, M.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Gorlatova, M.; Greskamp, B.; Grossman, J.; Gullingsrud, J.; Harper, A.; Hasenplaugh, W.; Heily, M.; Heshmat, B. C.; Hunt, J.; Ierardi, D. J.; Iserovich, L.; Jackson, B. L.; Johnson, N. P.; Kirk, M. M.; Klepeis, J. L.; Kuskin, J. S.; Mackenzie, K. M.; Mader, R. J.; McGowen, R.; McLaughlin, A.; Moraes, M. A.; Nasr, M. H.; Nociolo, L. J.; O'Donnell, L.; Parker, A.; Peticolas, J. L.; Pocina, G.; Predescu, C.; Quan, T.; Salmon, J. K.; Schwink, C.; Shim, K. S.; Siddique, N.; Spengler, J.; Szalay, T.; Tabladillo, R.; Tartler, R.; Taube, A. G.; Theobald, M.; Towles, B.; Vick, W.; Wang, S. C.; Wazlowski, M.; Weingarten, M. J.; Williams, J. M.; Yuh, K. A. Anton 3: twenty microseconds of molecular dynamics simulation before lunch. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. In *SC '21: The International Conference for High Performance Computing, Networking, Storage and Analysis*, St. Louis, MO, USA, Nov 14–19, 2021; de Supinski, B. R., Hall, M. W., Gamblin, T., Eds.; ACM, 2021; pp 1–11.

(6) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **2002**, *117*, 4184–4200.

(7) Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 469–480.

(8) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **2004**, *25*, 2038–2048.

(9) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III Constant-pH molecular dynamics using continuous titration coordinates. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 738–752.

(10) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmüller, H. Constant pH molecular dynamics in explicit solvent with λ-dynamics. *J. Chem. Theory Comput.* **2011**, *7*, 1962–1978.

(11) Chen, Y.; Roux, B. Constant-pH Hybrid Nonequilibrium Molecular Dynamics–Monte Carlo Simulation Method. *J. Chem. Theory Comput.* **2015**, *11*, 3919–3931.

(12) Huang, Y.; Chen, W.; Wallace, J. A.; Shen, J. All-atom continuous constant pH molecular dynamics with particle mesh Ewald and titratable water. *J. Chem. Theory Comput.* **2016**, *12*, 5411–5421.

(13) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(14) Kong, X.; Brooks, C. L., III λ-dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414–2423.

(15) Aho, N.; Buslaev, P.; Jansen, A.; Bauer, P.; Groenhof, G.; Hess, B. Scalable Constant pH Molecular Dynamics in GROMACS. *J. Chem. Theory Comput.* **2022**, DOI: 10.1021/acs.jctc.2c00516.

(16) Klimovich, P. V.; Mobley, D. L. Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 307–316.

(17) Vila-Viçosa, D.; Reis, P. B.; Baptista, A. M.; Oostenbrink, C.; Machuqueiro, M. A pH replica exchange scheme in the stochastic titration constant-pH MD method. *J. Chem. Theory Comput.* **2019**, *15*, 3108–3116.

(18) Meng, Y.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Biomolecules Using a Discrete Protonation Model. *J. Chem. Theory Comput.* **2010**, *6*, 1401–1412.

(19) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341–1352.

(20) Damjanovic, A.; Miller, B. T.; Okur, A.; Brooks, B. R. Reservoir pH replica exchange. *J. Chem. Phys.* **2018**, *149*, 072321.

(21) Wallace, J. A.; Shen, J. K. Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. *J. Chem. Phys.* **2012**, *137*, 184105.

(22) Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: p K a calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167–4180.

(23) Goh, G. B.; Knight, J. L.; Brooks, C. L. Constant pH Molecular Dynamics Simulations of Nucleic Acids in Explicit Solvent. *J. Chem. Theory Comput.* **2012**, *8*, 36–46.

(24) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(25) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(26) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical electrostatics for biomolecular simulations. *Chem. Rev.* **2014**, *114*, 779−814.
(27) Hub, J. S.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.* **2014**, *10*, 381−390.
(28) Chen, W.; Wallace, J. A.; Yue, Z.; Shen, J. K. Introducing Titratable Water to All-Atom Molecular Dynamics at Constant pH. *Biophys. J.* **2013**, *105*, L15−L17.
(29) Donnini, S.; Ullmann, R. T.; Groenhof, G.; Grubmüller, H. Charge-neutral constant pH molecular dynamics simulations using a parsimonious proton buffer. *J. Chem. Theory Comput.* **2016**, *12*, 1040−1051.
(30) Hünenberger, P. H.; McCammon, J. A. Ewald artifacts in computer simulations of ionic solvation and ion−ion interaction: A continuum electrostatics study. *J. Chem. Phys.* **1999**, *110*, 1856−1872.
(31) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.* **2013**, *139*, 184103.
(32) Huang, J.; MacKerell, A. D., Jr CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry* **2013**, *34*, 2135−2145.
(33) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.
(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulatin liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(35) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-induced forces between two hydrophilic groups. *J. Phys. Chem.* **1994**, *98*, 2198−2202.
(36) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(37) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(38) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(39) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997**, *18*, 1463−1472.
(40) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **1992**, *13*, 952−962.
(41) Knight, J. L.; Brooks, C. L., III λ-Dynamics free energy simulation methods. *Journal of computational chemistry* **2009**, *30*, 1692−1700.
(42) Knight, J. L.; Brooks, C. L., III Multisite λ dynamics for simulated structure−activity relationship studies. *J. Chem. Theory Comput.* **2011**, *7*, 2728−2739.
(43) Singhal, A. K.; Chien, K. Y.; Wu, W. G.; Rule, G. S. Solution structure of cardiotoxin V from Naja naja atra. *Biochemistry* **1993**, *32*, 8036−8044.
(44) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(45) Torrie, G. M.; Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **1974**, *28*, 578−581.
(46) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **1992**, *13*, 1011−1021.
(47) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713−3720.
(48) Granovsky, A. A. *Firefly version 7.1.G*; 2012; http://classic.chem.msu.su/gran/gamess/index.html (accessed 01−07−2022).
(49) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr General atomic and molecular electronic structure system. *Journal of computational chemistry* **1993**, *14*, 1347−1363.
(50) Lindahl, V.; Lidmar, J.; Hess, B. Accelerated weight histogram method for exploring free energy landscapes. *J. Chem. Phys.* **2014**, *141*, 044110.
(51) Lundborg, M.; Lidmar, J.; Hess, B. The accelerated weight histogram method for alchemical free energy calculations. *J. Chem. Phys.* **2021**, *154*, 204103.
(52) Okada, T.; Sugihara, M.; Bondar, A.-N.; Elstner, M.; Entel, P.; Buss, V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *Journal of molecular biology* **2004**, *342*, 571−583.
(53) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-resolution crystal structure of an engineered human β2-adrenergic G protein−coupled receptor. *science* **2007**, *318*, 1258−1265.
(54) Warne, T.; Moukhametzianov, R.; Baker, J. G.; Nehmé, R.; Edwards, P. C.; Leslie, A. G.; Schertler, G. F.; Tate, C. G. The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. *Nature* **2011**, *469*, 241−244.
(55) Glukhova, A.; Thal, D. M.; Nguyen, A. T.; Vecchio, E. A.; Jörg, M.; Scammells, P. J.; May, L. T.; Sexton, P. M.; Christopoulos, A. Structure of the adenosine A1 receptor reveals the basis for subtype selectivity. *Cell* **2017**, *168*, 867−877.
(56) Rodriguez-Espigares, I.; Torrens-Fontanals, M.; Tiemann, J. K. S.; Aranda-Garcia, D.; Ramirez-Anguita, J. M.; Stepniewski, T. M.; Worp, N.; Varela-Rial, A.; Morales-Pastor, A.; Medel-Lacruz, B.; Pandy-Szekeres, G.; Mayol, E.; Giorgino, T.; Carlsson, J.; Deupi, X.; Filipek, S.; Filizola, M.; Gomez-Tamayo, J. C.; Gonzalez, A.; Gutierrez-de-Teran, H.; Jimenez-Roses, M.; Jespers, W.; Kapla, J.; Khelashvili, G.; Kolb, P.; Latek, D.; Marti-Solano, M.; Matricon, P.; Matsoukas, M.-T.; Miszta, P.; Olivella, M.; Perez-Benito, L.; Provasi, D.; Rios, S.; Torrecillas, I. R.; Sallander, J.; Sztyler, A.; Vasile, S.; Weinstein, H.; Zachariae, U.; Hildebrand, P. W.; De Fabritiis, G.; Sanz, F.; Gloriam, D. E.; Cordomi, A.; Guixa-Gonzalez, R.; Selent, J.; et al. GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nat. Methods* **2020**, *17*, 777−787.
(57) Pavlova, A.; Zhang, Z.; Acharya, A.; Lynch, D. L.; Pang, Y. T.; Mou, Z.; Parks, J. M.; Chipot, C.; Gumbart, J. C. Machine Learning Reveals the Critical Interactions for SARS-CoV-2 Spike Protein Binding to ACE2. *J. Phys. Chem. Lett.* **2021**, *12* (23), 5494−5502.
(58) Massey, F. J., Jr The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association* **1951**, *46*, 68−78.
(59) Buslaev, P.; Gordeliy, V.; Grudinin, S.; Gushchin, I. Principal component analysis of lipid molecule conformational changes in molecular dynamics simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1019−1028.
(60) Buslaev, P.; Mustafin, K.; Gushchin, I. Principal component analysis highlights the influence of temperature, curvature and cholesterol on conformational dynamics of lipids. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2020**, *1862*, 183253.
(61) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **1999**, *314*, 141−151.
(62) Rosta, E.; Hummer, G. Error and efficiency of replica exchange molecular dynamics simulations. *J. Chem. Phys.* **2009**, *131*, 165102.
(63) Wallace, J. A.; Shen, J. K. Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange. *J. Chem. Theory Comput.* **2011**, *7*, 2617−2629.

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(64) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks, C. L., III Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 1319−1331.
(65) Khandogin, J.; Brooks, C. L. Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry* **2006**, *45*, 9363−9373.
(66) Itoh, S. G.; Damjanović, A.; Brooks, B. R. pH replica-exchange method based on discrete protonation states. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3420−3436.
(67) Dobrev, P.; Donnini, S.; Groenhof, G.; Grubmüller, H. Accurate Three States Model for Amino Acids with Two Chemically Coupled Titrating Sites in Explicit Solvent Atomistic Constant pH Simulations and p K a Calculations. *J. Chem. Theory Comput.* **2017**, *13*, 147−160.
(68) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(69) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(70) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.
(71) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(72) Sofronov, O. O.; Giubertoni, G.; Perez de Alba Ortiz, A.; Ensing, B.; Bakker, H. J. Peptide Side-COOH Groups Have Two Distinct Conformations under Biorelevant Conditions. *journal of physical chemistry letters* **2020**, *11*, 3466−3472.
(73) Chiang, C.-M.; Chien, K.-Y.; Lin, H.-j.; Lin, J.-F.; Yeh, H.-C.; Ho, P.-l.; Wu, W.-g. Conformational change and inactivation of membrane phospholipid-related activity of cardiotoxin V from Taiwan cobra venom at acidic pH. *Biochemistry* **1996**, *35*, 9167−9176.
(74) Chiang, C.-M.; Chang, S.-L.; Lin, H.-j.; Wu, W.-g. The role of acidic amino acid residues in the structural stability of snake cardiotoxins. *Biochemistry* **1996**, *35*, 9177−9186.

https://doi.org/10.1021/acs.jctc.2c00517
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

DEPARTMENT OF CHEMISTRY, UNIVERSITY OF JYVÄSKYLÄ
RESEARCH REPORT SERIES

1. Vuolle, Mikko: Electron paramagnetic resonance and molecular orbital study of radical ions generated from (2.2)metacyclophane, pyrene and its hydrogenated compounds by alkali metal reduction and by thallium(III)trifluoroacetate oxidation. (99 pp.) 1976
2. Pasanen, Kaija: Electron paramagnetic resonance study of cation radical generated from various chlorinated biphenyls. (66 pp.) 1977
3. Carbon-13 Workshop, September 6-8, 1977. (91 pp.) 1977
4. Laihia, Katri: On the structure determination of norbornane polyols by NMR spectroscopy. (111 pp.) 1979
5. Nyrönen, Timo: On the EPR, ENDOR and visible absorption spectra of some nitrogen containing heterocyclic compounds in liquid ammonia. (76 pp.) 1978
6. Talvitie, Antti: Structure determination of some sesquiterpenoids by shift reagent NMR. (54 pp.) 1979
7. Häkli, Harri: Structure analysis and molecular dynamics of cyclic compounds by shift reagent NMR. (48 pp.) 1979
8. Pitkänen, Ilkka: Thermodynamics of complexation of 1,2,4-triazole with divalent manganese, cobalt, nickel, copper, zinc, cadmium and lead ions in aqueous sodium perchlorate solutions. (89 pp.) 1980
9. Asunta, Tuula: Preparation and characterization of new organometallic compounds synthesized by using metal vapours. (91 pp.) 1980
10. Sattar, Mohammad Abdus: Analyses of MCPA and its metabolites in soil. (57 pp.) 1980
11. Bibliography 1980. (31 pp.) 1981
12. Knuuttila, Pekka: X-Ray structural studies on some divalent 3d metal compounds of picolinic and isonicotinic acid N-oxides. (77 pp.) 1981
13. Bibliography 1981. (33 pp.) 1982
14. 6^{th} National NMR Symposium, September 9-10, 1982, Abstracts. (49 pp.) 1982
15. Bibliography 1982. (38 pp.) 1983
16. Knuuttila, Hilkka: X-Ray structural studies on some Cu(II), Co(II) and Ni(II) complexes with nicotinic and isonicotinic acid N-oxides. (54 pp.) 1983
17. Symposium on inorganic and analytical chemistry May 18, 1984, Program and Abstracts. (100 pp.) 1984
18. Knuutinen, Juha: On the synthesis, structure verification and gas chromatographic determination of chlorinated catechols and guaiacols occuring in spent bleach liquors of kraft pulp mill. (30 pp.) 1984
19. Bibliography 1983. (47 pp.) 1984
20. Pitkänen, Maija: Addition of BrCl, B_2 and Cl_2 to methyl esters of propenoic and 2-butenoic acid derivatives and ^{13}C NMR studies on methyl esters of saturated aliphatic mono- and dichlorocarboxylic acids. (56 pp.) 1985
21. Bibliography 1984. (39 pp.) 1985
22. Salo, Esa: EPR, ENDOR and TRIPLE spectroscopy of some nitrogen heteroaromatics in liquid ammonia. (111 pp.) 1985

23. Humppi, Tarmo: Synthesis, identification and analysis of dimeric impurities of chlorophenols. (39 pp.) 1985
24. Aho, Martti: The ion exchange and adsorption properties of sphagnum peat under acid conditions. (90 pp.) 1985
25. Bibliography 1985 (61 pp.) 1986
26. Bibliography 1986. (23 pp.) 1987
27. Bibliography 1987. (26 pp.) 1988
28. Paasivirta, Jaakko (Ed.): Structures of organic environmental chemicals. (67 pp.) 1988
29. Paasivirta, Jaakko (Ed.): Chemistry and ecology of organo-element compounds. (93 pp.) 1989
30. Sinkkonen, Seija: Determination of crude oil alkylated dibenzothiophenes in environment. (35 pp.) 1989
31. Kolehmainen, Erkki (Ed.): XII National NMR Symposium Program and Abstracts. (75 pp.) 1989
32. Kuokkanen, Tauno: Chlorocymenes and Chlorocymenenes: Persistent chlorocompounds in spent bleach liquors of kraft pulp mills. (40 pp.) 1989
33. Mäkelä, Reijo: ESR, ENDOR and TRIPLE resonance study on substituted 9,10-anthraquinone radicals in solution. (35 pp.) 1990
34. Veijanen, Anja: An integrated sensory and analytical method for identification of off-flavour compounds. (70 pp.) 1990
35. Kasa, Seppo: EPR, ENDOR and TRIPLE resonance and molecular orbital studies on a substitution reaction of anthracene induced by thallium(III) in two fluorinated carboxylic acids. (114 pp.) 1990
36. Herve, Sirpa: Mussel incubation method for monitoring organochlorine compounds in freshwater recipients of pulp and paper industry. (145 pp.) 1991
37. Pohjola, Pekka: The electron paramagnetic resonance method for characterization of Finnish peat types and iron (III) complexes in the process of peat decomposition. (77 pp.) 1991
38. Paasivirta, Jaakko (Ed.): Organochlorines from pulp mills and other sources. Research methodology studies 1988-91. (120 pp.) 1992
39. Veijanen, Anja (Ed.): VI National Symposium on Mass Spectrometry, May 13-15, 1992, Abstracts. (55 pp.) 1992
40. Rissanen, Kari (Ed.): The 7. National Symposium on Inorganic and Analytical Chemistry, May 22, 1992, Abstracts and Program. (153 pp.) 1992
41. Paasivirta, Jaakko (Ed.): CEOEC'92, Second Finnish-Russian Seminar: Chemistry and Ecology of Organo-Element Compounds. (93 pp.) 1992
42. Koistinen, Jaana: Persistent polychloroaromatic compounds in the environment: structure-specific analyses. (50 pp.) 1993
43. Virkki, Liisa: Structural characterization of chlorolignins by spectroscopic and liquid chromatographic methods and a comparison with humic substances. (62 pp.) 1993
44. Helenius, Vesa: Electronic and vibrational excitations in some

biologically relevant molecules. (30 pp.) 1993

45. Leppä-aho, Jaakko: Thermal behaviour, infrared spectra and x-ray structures of some new rare earth chromates(VI). (64 pp.) 1994
46. Kotila, Sirpa: Synthesis, structure and thermal behavior of solid copper(II) complexes of 2-amino-2- hydroxymethyl-1,3-propanediol. (111 pp.) 1994
47. Mikkonen, Anneli: Retention of molybdenum(VI), vanadium(V) and tungsten(VI) by kaolin and three Finnish mineral soils. (90 pp.) 1995
48. Suontamo, Reijo: Molecular orbital studies of small molecules containing sulfur and selenium. (42 pp.) 1995
49. Hämäläinen, Jouni: Effect of fuel composition on the conversion of fuel-N to nitrogen oxides in the combustion of small single particles. (50 pp.) 1995
50. Nevalainen, Tapio: Polychlorinated diphenyl ethers: synthesis, NMR spectroscopy, structural properties, and estimated toxicity. (76 pp.) 1995
51. Aittola, Jussi-Pekka: Organochloro compounds in the stack emission. (35 pp.) 1995
52. Harju, Timo: Ultrafast polar molecular photophysics of (dibenzylmethine)borondifluoride and 4-aminophthalimide in solution. (61 pp.) 1995
53. Maatela, Paula: Determination of organically bound chlorine in industrial and environmental samples. (83 pp.) 1995
54. Paasivirta, Jaakko (Ed.): CEOEC'95, Third Finnish-Russian Seminar: Chemistry and Ecology of Organo-Element Compounds. (109 pp.) 1995
55. Huuskonen, Juhani: Synthesis and structural studies of some supramolecular compounds. (54 pp.) 1995
56. Palm, Helena: Fate of chlorophenols and their derivatives in sawmill soil and pulp mill recipient environments. (52 pp.) 1995
57. Rantio, Tiina: Chlorohydrocarbons in pulp mill effluents and their fate in the environment. (89 pp.) 1997
58. Ratilainen, Jari: Covalent and non-covalent interactions in molecular recognition. (37 pp.) 1997
59. Kolehmainen, Erkki (Ed.): XIX National NMR Symposium, June 4-6, 1997, Abstracts. (89 pp.) 1997
60. Matilainen, Rose: Development of methods for fertilizer analysis by inductively coupled plasma atomic emission spectrometry. (41 pp.) 1997
61. Koistinen, Jari (Ed.): Spring Meeting on the Division of Synthetic Chemistry, May 15-16, 1997, Program and Abstracts. (36 pp.) 1997
62. Lappalainen, Kari: Monomeric and cyclic bile acid derivatives: syntheses, NMR spectroscopy and molecular recognition properties. (50 pp.) 1997
63. Laitinen, Eira: Molecular dynamics of cyanine dyes and phthalimides in solution: picosecond laser studies. (62 pp.) 1997
64. Eloranta, Jussi: Experimental and theoretical studies on some

quinone and quinol radicals. (40 pp.) 1997

65. Oksanen, Jari: Spectroscopic characterization of some monomeric and aggregated chlorophylls. (43 pp.) 1998
66. Häkkänen, Heikki: Development of a method based on laser-induced plasma spectrometry for rapid spatial analysis of material distributions in paper coatings. (60 pp.) 1998
67. Virtapohja, Janne: Fate of chelating agents used in the pulp and paper industries. (58 pp.) 1998
68. Airola, Karri: X-ray structural studies of supramolecular and organic compounds. (39 pp.) 1998
69. Hyötyläinen, Juha: Transport of lignin–type compounds in the receiving waters of pulp mills. (40 pp.) 1999
70. Ristolainen, Matti: Analysis of the organic material dissolved during totally chlorine-free bleaching. (40 pp.) 1999
71. Eklin, Tero: Development of analytical procedures with industrial samples for atomic emission and atomic absorption spectrometry. (43 pp.) 1999
72. Välisaari, Jouni: Hygiene properties of resol-type phenolic resin laminates. (129 pp.) 1999
73. Hu, Jiwei: Persistent polyhalogenated diphenyl ethers: model compounds syntheses, characterization and molecular orbital studies. (59 pp.) 1999
74. Malkavaara, Petteri: Chemometric adaptations in wood processing chemistry. (56 pp.) 2000
75. Kujala Elena, Laihia Katri, Nieminen Kari (Eds.): NBC 2000, Symposium on Nuclear, Biological and Chemical Threats in the 21st Century. (299 pp.) 2000
76. Rantalainen, Anna-Lea: Semipermeable membrane devices in monitoring persistent organic pollutants in the environment. (58 pp.) 2000
77. Lahtinen, Manu: *In situ* X-ray powder diffraction studies of Pt/C, CuCl/C and Cu_2O/C catalysts at elevated temperatures in various reaction conditions. (92 pp.) 2000
78. Tamminen, Jari: Syntheses, empirical and theoretical characterization, and metal cation complexation of bile acid-based monomers and open/closed dimers. (54 pp.) 2000
79. Vatanen, Virpi: Experimental studies by EPR and theoretical studies by DFT calculations of α-amino-9,10-anthraquinone radical anions and cations in solution. (37 pp.) 2000
80. Kotilainen, Risto: Chemical changes in wood during heating at 150-260 °C. (57 pp.) 2000
81. Nissinen, Maija: X-ray structural studies on weak, non-covalent interactions in supramolecular compounds. (69 pp.) 2001
82. Wegelius, Elina: X-ray structural studies on self-assembled hydrogen-bonded networks and metallosupramolecular complexes. (84 pp.) 2001
83. Paasivirta, Jaakko (Ed.): CEOEC'2001, Fifth Finnish-Russian Seminar: Chemistry and Ecology of Organo-Element Compounds. (163 pp.) 2001
84. Kiljunen, Toni: Theoretical studies on spectroscopy and

atomic dynamics in rare gas solids. (56 pp.) 2001

85. Du, Jin: Derivatives of dextran: synthesis and applications in oncology. (48 pp.) 2001
86. Koivisto, Jari: Structural analysis of selected polychlorinated persistent organic pollutants (POPs) and related compounds. (88 pp.) 2001
87. Feng, Zhinan: Alkaline pulping of non-wood feedstocks and characterization of black liquors. (54 pp.) 2001
88. Halonen, Markku: Lahon havupuun käyttö sulfaattiprosessin raaka-aineena sekä havupuun lahontorjunta. (90 pp.) 2002
89. Falábu, Dezsö: Synthesis, conformational analysis and complexation studies of resorcarene derivatives. (212 pp.) 2001
90. Lehtovuori, Pekka: EMR spectroscopic studies on radicals of ubiquinones Q-*n*, vitamin K_3 and vitamine E in liquid solution. (40 pp.) 2002
91. Perkkalainen, Paula: Polymorphism of sugar alcohols and effect of grinding on thermal behavior on binary sugar alcohol mixtures. (53 pp.) 2002
92. Ihalainen, Janne: Spectroscopic studies on light-harvesting complexes of green plants and purple bacteria. (42 pp.) 2002
93. Kunttu, Henrik, Kiljunen, Toni (Eds.): 4th International Conference on Low Temperature Chemistry. (159 pp.) 2002
94. Väisänen, Ari: Development of methods for toxic element analysis in samples with environmental concern by ICP-AES and ETAAS. (54 pp.) 2002
95. Luostarinen, Minna: Synthesis and characterisation of novel resorcarene derivatives. (200 pp.) 2002
96. Louhelainen, Jarmo: Changes in the chemical composition and physical properties of wood and nonwood black liquors during heating. (68 pp.) 2003
97. Lahtinen, Tanja: Concave hydrocarbon cyclophane π-prismands. (65 pp.) 2003
98. Laihia, Katri (Ed.): NBC 2003, Symposium on Nuclear, Biological and Chemical Threats – A Crisis Management Challenge. (245 pp.) 2003
99. Oasmaa, Anja: Fuel oil quality properties of wood-based pyrolysis liquids. (32 pp.) 2003
100. Virtanen, Elina: Syntheses, structural characterisation, and cation/anion recognition properties of nano-sized bile acid-based host molecules and their precursors. (123 pp.) 2003
101. Nättinen, Kalle: Synthesis and X-ray structural studies of organic and metallo-organic supramolecular systems. (79 pp.) 2003
102. Lampiselkä, Jarkko: Demonstraatio lukion kemian opetuksessa. (285 pp.) 2003
103. Kallioinen, Jani: Photoinduced dynamics of $Ru(dcbpy)_2(NCS)_2$ – in solution and on nanocrystalline titanium dioxide thin films. (47 pp.) 2004
104. Valkonen, Arto (Ed.): VII Synthetic Chemistry Meeting and XXVI Finnish NMR Symposium. (103 pp.) 2004

105. Vaskonen, Kari: Spectroscopic studies on atoms and small molecules isolated in low temperature rare gas matrices. (65 pp.) 2004
106. Lehtovuori, Viivi: Ultrafast light induced dissociation of $Ru(dcbpy)(CO)_2I_2$ in solution. (49 pp.) 2004
107. Saarenketo, Pauli: Structural studies of metal complexing Schiff bases, Schiff base derived *N*-glycosides and cyclophane π-prismands. (95 pp.) 2004
108. Paasivirta, Jaakko (Ed.): CEOEC'2004, Sixth Finnish-Russian Seminar: Chemistry and Ecology of Organo-Element Compounds. (147 pp.) 2004
109. Suontamo, Tuula: Development of a test method for evaluating the cleaning efficiency of hard-surface cleaning agents. (96 pp.) 2004
110. Güneş, Minna: Studies of thiocyanates of silver for nonlinear optics. (48 pp.) 2004
111. Ropponen, Jarmo: Aliphatic polyester dendrimers and dendrons. (81 pp.) 2004
112. Vu, Mân Thi Hong: Alkaline pulping and the subsequent elemental chlorine-free bleaching of bamboo (*Bambusa procera*). (69 pp.) 2004
113. Mansikkamäki, Heidi: Self-assembly of resorcinarenes. (77 pp.) 2006
114. Tuononen, Heikki M.: EPR spectroscopic and quantum chemical studies of some inorganic main group radicals. (79 pp.) 2005
115. Kaski, Saara: Development of methods and applications of laser-induced plasma spectroscopy in vacuum ultraviolet. (44 pp.) 2005
116. Mäkinen, Riika-Mari: Synthesis, crystal structure and thermal decomposition of certain metal thiocyanates and organic thiocyanates. (119 pp.) 2006
117. Ahokas, Jussi: Spectroscopic studies of atoms and small molecules isolated in rare gas solids: photodissociation and thermal reactions. (53 pp.) 2006
118. Busi, Sara: Synthesis, characterization and thermal properties of new quaternary ammonium compounds: new materials for electrolytes, ionic liquids and complexation studies. (102 pp.) 2006
119. Mäntykoski, Keijo: PCBs in processes, products and environment of paper mills using wastepaper as their raw material. (73 pp.) 2006
120. Laamanen, Pirkko-Leena: Simultaneous determination of industrially and environmentally relevant aminopolycarboxylic and hydroxycarboxylic acids by capillary zone electrophoresis. (54 pp.) 2007
121. Salmela, Maria: Description of oxygen-alkali delignification of kraft pulp using analysis of dissolved material. (71 pp.) 2007
122. Lehtovaara, Lauri: Theoretical studies of atomic scale impurities in superfluid ^{4}He. (87 pp.) 2007
123. Rautiainen, J. Mikko: Quantum chemical calculations of structures, bonding, and spectroscopic properties of some sulphur and selenium iodine cations. (71 pp.) 2007
124. Nummelin, Sami: Synthesis, characterization, structural and

retrostructural analysis of self-assembling pore forming dendrimers. (286 pp.) 2008

125. Sopo, Harri: Uranyl(VI) ion complexes of some organic aminobisphenolate ligands: syntheses, structures and extraction studies. (57 pp.) 2008
126. Valkonen, Arto: Structural characteristics and properties of substituted cholanoates and *N*-substituted cholanamides. (80 pp.) 2008
127. Lähde, Anna: Production and surface modification of pharmaceutical nano- and microparticles with the aerosol flow reactor. (43 pp.) 2008
128. Beyeh, Ngong Kodiah: Resorcinarenes and their derivatives: synthesis, characterization and complexation in gas phase and in solution. (75 pp.) 2008
129. Välisaari, Jouni, Lundell, Jan (Eds.): Kemian opetuksen päivät 2008: uusia oppimisympäristöjä ja ongelmalähtöistä opetusta. (118 pp.) 2008
130. Myllyperkiö, Pasi: Ultrafast electron transfer from potential organic and metal containing solar cell sensitizers. (69 pp.) 2009
131. Käkölä, Jaana: Fast chromatographic methods for determining aliphatic carboxylic acids in black liquors. (82 pp.) 2009
132. Koivukorpi, Juha: Bile acid-arene conjugates: from photoswitchability to cancer cell detection. (67 pp.) 2009
133. Tuuttila, Tero: Functional dendritic polyester compounds: synthesis and characterization of small bifunctional dendrimers and dyes. (74 pp.) 2009
134. Salorinne, Kirsi: Tetramethoxy resorcinarene based cation and anion receptors: synthesis, characterization and binding properties. (79 pp.) 2009
135. Rautiainen, Riikka: The use of first-thinning Scots pine (*Pinus sylvestris*) as fiber raw material for the kraft pulp and paper industry. (73 pp.) 2010
136. Ilander, Laura: Uranyl salophens: synthesis and use as ditopic receptors. (199 pp.) 2010
137. Kiviniemi, Tiina: Vibrational dynamics of iodine molecule and its complexes in solid krypton - Towards coherent control of bimolecular reactions? (73 pp.) 2010
138. Ikonen, Satu: Synthesis, characterization and structural properties of various covalent and non-covalent bile acid derivatives of N/O-heterocycles and their precursors. (105 pp.) 2010
139. Siitonen, Anni: Spectroscopic studies of semiconducting single-walled carbon nanotubes. (56 pp.) 2010
140. Raatikainen, Kari: Synthesis and structural studies of piperazine cyclophanes – Supramolecular systems through Halogen and Hydrogen bonding and metal ion coordination. (69 pp.) 2010
141. Leivo, Kimmo: Gelation and gel properties of two- and three-component Pyrene based low molecular weight organogelators. (116 pp.) 2011
142. Martiskainen, Jari: Electronic energy transfer in light-harvesting complexes isolated from *Spinacia oleracea* and from three

photosynthetic green bacteria *Chloroflexus aurantiacus*, *Chlorobium tepidum*, and *Prosthecochloris aestuarii*. (55 pp.) 2011

143. Wichmann, Oula: Syntheses, characterization and structural properties of [O,N,O,X´] aminobisphenolate metal complexes. (101 pp.) 2011
144. Ilander, Aki: Development of ultrasound-assisted digestion methods for the determination of toxic element concentrations in ash samples by ICP-OES. (58 pp.) 2011
145. The Combined XII Spring Meeting of the Division of Synthetic Chemistry and XXXIII Finnish NMR Symposium. Book of Abstracts. (90 pp.) 2011
146. Valto, Piia: Development of fast analysis methods for extractives in papermaking process waters. (73 pp.) 2011
147. Andersin, Jenni: Catalytic activity of palladium-based nanostructures in the conversion of simple olefinic hydro- and chlorohydrocarbons from first principles. (78 pp.) 2011
148. Aumanen, Jukka: Photophysical properties of dansylated poly(propylene amine) dendrimers. (55 pp.) 2011
149. Kärnä, Minna: Ether-functionalized quaternary ammonium ionic liquids – synthesis, characterization and physicochemical properties. (76 pp.) 2011
150. Jurček, Ondřej: Steroid conjugates for applications in pharmacology and biology. (57 pp.) 2011
151. Nauha, Elisa: Crystalline forms of selected Agrochemical actives: design and synthesis of cocrystals. (77 pp.) 2012
152. Ahkola, Heidi: Passive sampling in monitoring of nonylphenol ethoxylates and nonylphenol in aquatic environments. (92 pp.) 2012
153. Helttunen, Kaisa: Exploring the self-assembly of resorcinarenes: from molecular level interactions to mesoscopic structures. (78 pp.) 2012
154. Linnanto, Juha: Light excitation transfer in photosynthesis revealed by quantum chemical calculations and exiton theory. (179 pp.) 2012
155. Roiko-Jokela, Veikko: Digital imaging and infrared measurements of soil adhesion and cleanability of semihard and hard surfaces. (122 pp.) 2012
156. Noponen, Virpi: Amides of bile acids and biologically important small molecules: properties and applications. (85 pp.) 2012
157. Hulkko, Eero: Spectroscopic signatures as a probe of structure and dynamics in condensed-phase systems – studies of iodine and gold ranging from isolated molecules to nanoclusters. (69 pp.) 2012
158. Lappi, Hanna: Production of Hydrocarbon-rich biofuels from extractives-derived materials. (95 pp.) 2012
159. Nykänen, Lauri: Computational studies of Carbon chemistry on transition metal surfaces. (76 pp.) 2012
160. Ahonen, Kari: Solid state studies of pharmaceutically important molecules and their derivatives. (65 pp.) 2012

161. Pakkanen, Hannu: Characterization of organic material dissolved during alkaline pulping of wood and non-wood feedstocks. (76 pp.) 2012
162. Moilanen, Jani: Theoretical and experimental studies of some main group compounds: from closed shell interactions to singlet diradicals and stable radicals. (80 pp.) 2012
163. Himanen, Jatta: Stereoselective synthesis of Oligosaccharides by *De Novo* Saccharide welding. (133 pp.) 2012
164. Bunzen, Hana: Steroidal derivatives of nitrogen containing compounds as potential gelators. (76 pp.) 2013
165. Seppälä, Petri: Structural diversity of copper(II) amino alcohol complexes. Syntheses, structural and magnetic properties of bidentate amino alcohol copper(II) complexes. (67 pp.) 2013
166. Lindgren, Johan: Computational investigations on rotational and vibrational spectroscopies of some diatomics in solid environment. (77 pp.) 2013
167. Giri, Chandan: Sub-component self-assembly of linear and non-linear diamines and diacylhydrazines, formulpyridine and transition metal cations. (145 pp.) 2013
168. Riisiö, Antti: Synthesis, Characterization and Properties of Cu(II)-, Mo(VI)- and U(VI) Complexes With Diaminotetraphenolate Ligands. (51 pp.) 2013
169. Kiljunen, Toni (Ed.): Chemistry and Physics at Low Temperatures. Book of Abstracts. (103 pp.) 2013
170. Hänninen, Mikko: Experimental and Computational Studies of Transition Metal Complexes with Polydentate Amino- and Aminophenolate Ligands: Synthesis, Structure, Reactivity and Magnetic Properties. (66 pp.) 2013
171. Antila, Liisa: Spectroscopic studies of electron transfer reactions at the photoactive electrode of dye-sensitized solar cells. (53 pp.) 2013
172. Kemppainen, Eeva: Mukaiyama-Michael reactions with α-substituted acroleins – a useful tool for the synthesis of the pectenotoxins and other natural product targets. (190 pp.) 2013
173. Virtanen, Suvi: Structural Studies of Dielectric Polymer Nanocomposites. (49 pp.) 2013
174. Yliniemelä-Sipari, Sanna: Understanding The Structural Requirements for Optimal Hydrogen Bond Catalyzed Enolization – A Biomimetic Approach.(160 pp.) 2013
175. Leskinen, Mikko V: Remote β-functionalization of β'-keto esters. (105 pp.) 2014
176. 12th European Conference on Research in Chemistry Education (ECRICE2014). Book of Abstracts. (166 pp.) 2014
177. Peuronen, Anssi: N-Monoalkylated DABCO-Based N-Donors as Versatile Building Blocks in Crystal Engineering and Supramolecular Chemistry. (54 pp.) 2014
178. Perämäki, Siiri: Method development for determination and recovery of rare earth elements from industrial fly ash. (88 pp.) 2014

179. Chernyshev, Alexander, N.: Nitrogen-containing ligands and their platinum(IV) and gold(III) complexes: investigation and basicity and nucleophilicity, luminescence, and aurophilic interactions. (64 pp.) 2014
180. Lehto, Joni: Advanced Biorefinery Concepts Integrated to Chemical Pulping. (142 pp.) 2015
181. Tero, Tiia-Riikka: Tetramethoxy resorcinarenes as platforms for fluorescent and halogen bonding systems. (61 pp.) 2015
182. Löfman, Miika: Bile acid amides as components of microcrystalline organogels. (62 pp.) 2015
183. Selin, Jukka: Adsorption of softwood-derived organic material onto various fillers during papermaking. (169 pp.) 2015
184. Piisola, Antti: Challenges in the stereoselective synthesis of allylic alcohols. (210 pp.) 2015
185. Bonakdarzadeh, Pia: Supramolecular coordination polyhedra based on achiral and chiral pyridyl ligands: design, preparation, and characterization. (65 pp.) 2015
186. Vasko, Petra: Synthesis, characterization, and reactivity of heavier group 13 and 14 metallylenes and metalloid clusters: small molecule activation and more. (66 pp.) 2015
187. Topić, Filip: Structural Studies of Nano-sized Supramolecular Assemblies. (79 pp.) 2015
188. Mustalahti, Satu: Photodynamics Studies of Ligand-Protected Gold Nanoclusters by using Ultrafast Transient Infrared Spectroscopy. (58 pp.) 2015
189. Koivisto, Jaakko: Electronic and vibrational spectroscopic studies of gold-nanoclusters. (63 pp.) 2015
190. Suhonen, Aku: Solid state conformational behavior and interactions of series of aromatis oligoamide foldamers. (68 pp.) 2016
191. Soikkeli, Ville: Hydrometallurgical recovery and leaching studies for selected valuable metals from fly ash samples by ultrasound-assisted extraction followed by ICP-OES determination. (107 pp.) 2016
192. XXXVIII Finnish NMR Symposium. Book of Abstracts. (51 pp.) 2016
193. Mäkelä, Toni: Ion Pair Recognition by Ditopic Crown Ether Based bis-Urea and Uranyl Salophen Receptors. (75 pp.) 2016
194. Lindholm-Lehto, Petra: Occurrence of pharmaceuticals in municipal wastewater treatment plants and receiving surface waters in Central and Southern Finland. (98 pp.) 2016
195. Härkönen, Ville: Computational and Theoretical studies on Lattice Thermal conductivity and Thermal properties of Silicon Clathrates. (89 pp.) 2016
196. Tuokko, Sakari: Understanding selective reduction reactions with heterogeneous Pd and Pt: climbing out of the black box. (85 pp.) 2016
197. Nuora, Piia: Monitapaustutkimus LUMA-Toimintaan liittyvissä oppimisympäristöissä tapahtuvista kemian oppimiskokemuksista. (171 pp.) 2016

198. Kumar, Hemanathan: Novel Concepts on The Recovery of By-Products from Alkaline Pulping. (61 pp.) 2016
199. Arnedo-Sánchez, Leticia: Lanthanide and Transition Metal Complexes as Building Blocks for Supramolecular Functional Materials. (227 pp.) 2016
200. Gell, Lars: Theoretical Investigations of Ligand Protected Silver Nanoclusters. (134 pp.) 2016
201. Vaskuri, Juhani: Oppiennätyksistä opetussuunnitelman perusteisiin - lukion kemian kansallisen opetussuunnitelman kehittyminen Suomessa vuosina 1918-2016. (314 pp.) 2017
202. Lundell Jan, Kiljunen Toni (Eds.): 22nd Horizons in Hydrogen Bond Research. Book of Abstracts. 2017
203. Turunen, Lotta: Design and construction of halogen-bonded capsules and cages. (61 pp.) 2017
204. Hurmalainen, Juha: Experimental and computational studies of unconventional main group compounds: stable radicals and reactive intermediates. (88 pp.) 2017
205. Koivistoinen Juha: Non-linear interactions of femtosecond laser pulses with graphene: photo-oxidation, imaging and photodynamics. (68 pp.) 2017
206. Chen, Chengcong: Combustion behavior of black liquors: droplet swelling and influence of liquor composition. (39 pp.) 2017
207. Mansikkamäki, Akseli: Theoretical and Computational Studies of Magnetic Anisotropy and Exchange Coupling in Molecular Systems. (190 p. + included articles) 2018.
208. Tatikonda, Rajendhraprasad: Multivalent N-donor ligands for the construction of coordination polymers and coordination polymer gels. (62 pp.) 2018
209. Budhathoki, Roshan: Beneficiation, desilication and selective precipitation techniques for phosphorus refining from biomass derived fly ash. (64 pp.) 2018
210. Siitonen, Juha: Synthetic Studies on 1-azabicyclo[5.3.0]decane Alkaloids. (140 pp.) 2018
211. Ullah, Saleem: Advanced Biorefinery Concepts Related to Non-wood Feedstocks. (57 pp.) 2018
212. Ghalibaf, Maryam: Analytical Pyrolysis of Wood and Non-Wood Materials from Integrated Biorefinery Concepts. (106 pp.) 2018

DEPARTMENT OF CHEMISTRY, UNIVERSITY OF JYVÄSKYLÄ
DISSERTATIONS PUBLISHED IN THE JYU DISSERTATIONS RESEARCH SERIES

1. Bulatov, Evgeny: Synthetic and structural studies of covalent and non-covalent interactions of ligands and metal center in platinum(II) complexes containing 2,2′-dipyridylamine or oxime ligands. (58 pp.) 2019. JYU Dissertations 70.
2. Annala, Riia: Conformational Properties and Anion Complexes of Aromatic Oligoamide Foldamers. (80 pp.) 2019. JYU Dissertations 84.
3. Isoaho, Jukka Pekka: Dithionite Bleaching of Thermomechanical Pulp - Chemistry and Optimal Conditions. (73 pp.) 2019. JYU Dissertations 85.
4. Nygrén, Enni: Recovery of rubidium from power plant fly ash. (98 pp.) 2019. JYU Dissertations 136.
5. Kiesilä, Anniina: Supramolecular chemistry of anion-binding receptors based on concave macromolecules. (68 pp.) 2019. JYU Dissertations 137.
6. Sokolowska, Karolina: Study of water-soluble p-MBA-protected gold nanoclusters and their superstructures. (60 pp.) 2019. JYU Dissertations 167.
7. Lahtinen, Elmeri: Chemically Functional 3D Printing: Selective Laser Sintering of Customizable Metal Scavengers. (71 pp.) 2019. JYU Dissertations 175.
8. Larijani, Amir: Oxidative reactions of cellulose under alkaline conditions. (102 pp.) 2020. JYU Dissertations 217.
9. Kolari, Kalle: Metal-metal contacts in late transition metal polymers. (60 pp.) 2020. JYU Dissertations 220.
10. Kauppinen, Minttu: Multiscale computational investigation of catalytic properties of zirconia supported noble metals. (87 pp.) 2020. JYU Dissertations 231.
11. Ding, Xin: Halogen Bond in Crystal Engineering: Structural Studies on Crystals with Ruthenium Centered Complexes and 1-(4-Pyridyl)-4-thiopyridine Zwitterion as Halogen Bond Acceptors. (59 pp.) 2020. JYU Dissertations 323.
12. Neuvonen, Antti: Toward an Understanding of Hydrogen-Bonding Bifunctional Organocatalyst Conformations and Their Activity in Asymmetric Mannich Reactions. (77 pp.) 2020. JYU Dissertations 336.
13. Kortet, Sami: 2,5-Diarylpyrrolidines and Pyroglutamic-Acid-Derived 2-Diarylmethyl-5-Aryl-Pyrrolidines: Their Synthesis and Use in Asymmetric Synthesis. (221 pp.) 2020. JYU Dissertations 337.
14. Saarnio, Ville: Fluorescent probes, noble metal nanoparticles and their nanocomposites: detection of nucleic acids and other biological targets. (80 pp.) 2021. JYU Dissertations 361.
15. Chernysheva, Maria: σ-hole interactions: the effect of the donors and acceptors nature in selenoureas, thioureas, halogenated species, substituted benzenes, and their adducts. (72 pp.) 2021. JYU Dissertations 370.
16. Bulatova, Margarita: Noncovalent interactions as a tool for supramolecular self-assembly of metallopolymers. (62 pp.) 2021. JYU Dissertations 377.

17. Romppanen, Sari: Laser-spectroscopic studies of rare earth element- and lithium-bearing minerals and rocks. (66 pp.) 2021. JYU Dissertations 393.
18. Kukkonen, Esa: Nonlinear optical materials through weak interactions and their application in 3D printing. (58 pp.) 2021. JYU Dissertations 441.
19. Kuosmanen, Riikka: The Effect of Structure on the Gel Formation Ability and the Properties of Bile Acid Based Supramolecular Organogels. (68 pp.) 2021. JYU Dissertations 465.
20. Reuna, Sini: Development of a Method for Phosphorus Recovery from Wastewaters. (67 pp.) 2022. JYU Dissertations 486.
21. Taipale, Essi: Synthetic and Structural Studies on the Effect of Non-Covalent Interactions on N(*sp2*)-Heterocyclic Molecules. (67 pp.) 2022. JYU Dissertations 496.
22. Järvinen, Teemu: Molecular Dynamics View on Matrix Isolation. (143 pp.) 2022. JYU Dissertations 544.
23. Kumar, Parveen: Synthesis and Structural Studies on Halogen(I) Complexes. (160 pp.) 2022. JYU Dissertations 549.
24. Forsblom, Samu: Design and Construction of Metal-Organic Polyhedra. (212 pp.) 2022. JYU Dissertations 569.
25. Korpelin, Ville: Computational Studies of Catalytic Active Site Properties and Reactions at the Metal–Oxide Interface. (78 pp.) 2022. JYU Dissertations 578.
26. Vuori, Hannu: Extending Benson Group Increment Theory to Compounds of Phosphorus, Silicon, and Boron with Computational Chemistry. (59 pp.) 2022. JYU Dissertations 581.
27. Pallerla, Rajanish: Studies Towards Synthesis of Favipiravir & Humilisin E. (139 pp.) 2023. JYU Dissertations 611.