Environmental Phosphorus Recovery Based on Molecular Bioscavengers

From Quantum Mechanics to Continuum Physics

Gruber, Mathias Felix; Hélix-Nielsen, Claus

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Technical University of Denmark

Environmental Phosphorus Recovery Based on Molecular Bioscavengers

– From Quantum Mechanics to Continuum Physics

Mathias F. Gruber

DTU Environment
Department of Environmental Engineering

PhD Thesis
June 2016

Environmental Phosphorus Recovery Based on Molecular Bioscavengers

– From Quantum Mechanics to Continuum Physics

Mathias F. Gruber

PhD Thesis

June, 2016

DTU Environment
Department of Environmental Engineering
Technical University of Denmark

Mathias F. Gruber

Environmental Phosphorus Recovery Based on Molecular Bioscavengers
– From Quantum Mechanics to Continuum Physics

PhD Thesis, June 2016

Address:	DTU Environment Department of Environmental Engineering Technical University of Denmark Miljoevej, building 113 2800 Kgs. Lyngby Denmark
Phone reception:	+45 4525 1600
Fax:	+45 4593 2850
Homepage:	http://www.env.dtu.dk
E-mail:	info@env.dtu.dk

Printed by:	GraphicCo June 2016
Cover:	Torben Dolin

Preface

This thesis is submitted in partial fulfillment of the requirements for obtaining the Ph.D. degree from the Technical University of Denmark (DTU). The project was funded by *IBISS: Industrial Biomimetic Sensing and Separation, supported by Innovation Fund Denmark* and by *Copenhagen Cleantech Cluster*. The main academic supervisor of the project was Claus Hélix-Nielsen (DTU Environment). The work took place between October 2012 and March 2016, and was carried out at Aquaporin A/S, DTU Physics (October 2012 – April 2014) and DTU Environment (May 2014 – March 2016). During this time, a total of six months were spent with unrelated work during a leave of absence from the project. Another six months were spent on an external stay in New Zealand (March 2013 to August 2013) at the University of Auckland, under professor Vijayalekshmi Sarojini.

Publications

There are four papers integrated into this thesis, which are referred to in the text by Roman numerals **I - IV**:

Paper I Mathias F. Gruber, Per Jr. Greisen, Caroline M. Junker, and Claus Hélix-Nielsen, Phosphorus binding sites in proteins: structural preorganization and coordination. *The Journal of Physical Chemistry. B*, **2014**, 118, 1207-1215.

Paper II Mathias F. Gruber, Andrea Bordoni, Claus Hélix-Nielsen, Describing Phosphate, Sulphate and Arsenate with Quantum Methods: Does one-size-fits-all apply?, Submitted to *Journal of Computational Chemistry*.

Paper III Mathias F. Gruber, Elizabeth Wood, Sigurd Truelsen, Thomas Østergaard, and Claus Hélix-Nielsen, Computational Design of Biomimetic Phosphate Scavengers, *Environmental Science & Technology*, **2015**, 49, 9469–9478.

Paper IV Mathias F. Gruber, Ulf Aslak, Claus Hélix-Nielsen, Open-Source CFD Model for Optimization of Forward Osmosis and Reverse Osmosis Membrane Modules, *Separation and Purification Technology*, **2016**, 158, 183-192

Contributions to the Papers

Paper I Designed and performed all the numerical research and data analysis. Wrote the initial manuscript and managed input from co-authors Handled the journal submission and revision process.

Paper II Designed and performed all the numerical research and data analysis. Wrote the initial manuscript and managed input from co-authors Handled the journal submission process.

Paper III Designed and performed all the numerical research and data analysis. Wrote the initial manuscript and managed input from co-authors Handled the journal submission and revision process.

Paper IV Designed and performed all the numerical research and data analysis. Wrote the initial manuscript and managed input from co-authors Handled the journal submission and revision process.

In this online version of the thesis, **Papers I - IV** are not included but can be obtained from electronic article databases e.g. via www.orbit.dtu.dk or on request from:

DTU Environment,
Technical University of Denmark,
Miljoevej, Building 113,
2800 Kgs. Lyngby,
Denmark,

info@env.dtu.dk.

Dansk sammenfatning

Fosfor er et essentielt grundstof, der indgår i alt kendt liv. Det findes således i en lang række centrale molekyler, der indgår i forskellige cellulære funktioner. På globalt plan anvendes fosfor i produktionen af gødningsprodukter, og uden fosfor ville vi ikke være i stand til at brødføde verdens befolkning. Mængden af de fossile reserver af fosfor er desværre begrænset, og enkelte estimater forudser, at vi indenfor de næste 15-25 år forbruger mere fosfat, end vi kan producere. Der er derfor et internationalt pres for at implementere bæredygtige retningslinjer for, hvordan fosfat forbruges, samt for at udvikle nye teknologier til genanvendelse af fosfat fra vores spildevand.

Naturen har brugt milliarder af år på at raffinere proteiner, der interagerer med en række forskellige fosfat-forbindelser. Dette er den primære inspiration for det indeværende arbejde, og de overordnede ambitioner for rapporten er således: at bidrage til den fremtidige udvikling af en genanvendelses-teknologi baseret på biologiske molekyler, at belyse fundamentale molekylære aspekter der er relevante ved en sådan teknologi, samt vise hvordan computermodeller kan bruges i udviklingen af en sådan teknologi. Det foreliggende arbejde kombinerer således et bredt spektrum af computer-modeller, helt fra atomare kvanteberegninger til makroskala fluid dynamik, til at anskueliggøre en genanvendelsesteknologi baseret på biomolekyler.

Rapporten indledes med en statistisk analyse af eksperimentelle data fra kendte proteiner. Dette giver et indblik i, hvordan proteiner i naturen interagerer med forskellige fosfat-forbindelser i form af, hvilke aminosyrer der forefindes i proteinernes bindingssteder for fosfat. Derefter bruges metoder fra kvantemekanikken til at undersøge fosfat-molekyler isoleret, og de kvantemekaniske metoder kombineres derpå med metoder fra molekylær dynamik for at vise, hvordan det dynamiske samspil mellem fosfater og proteiner kan beskrives og kvantificeres – Det vises således, at nogle almindeligt brugte metoder, herunder B3LYP, ikke egner sig til at beskrive interaktioner med fosfat, men såfremt andre computer-modeller bruges, eksempelvis wB97XD eller PM6, så kan det godt lade sig gøre at simulere interaktioner med fosfat nøjagtigt. Slutteligt vises det, hvordan en open source fluid-dynamik model, som vi har gjort frit tilgængeligt online, kan bruges til at optimere industrielt design, og det diskuteres, hvilke fordele og ulemper der er ved en række potentielle udviklingsmuligheder for den foreslåede teknologi.

Summary

Phosphorus is a ubiquitous element of all known life and as such it is found throughout numerous key molecules related to various cellular functions. The supply of phosphorus is tightly linked to global food security, since phosphorus is used to produce agricultural fertilizers, without which it would not be possible to feed the world population. Sadly, the current supply of phosphorus is based on the gradual depletion of limited fossil reserves, and some estimates predict that within 15-25 years we will consume more phosphorus than we can produce. There is therefore a strong international pressure to develop sustainable phosphorus practices as well as new technologies for phosphorus recovery.

Nature has spent billions of years refining proteins that interact with phosphates. This has inspired the present work where the overall ambitions are: to facilitate the development of a recovery technology based on biological phosphorus scavengers, to examine fundamental molecular system aspects relevant for such a technology, and to motivate the use of computational techniques throughout an iterative design process of such a technology. A wide spectrum of computational methods, from atomic-scale quantum calculations to macro-scale fluid simulations, are employed to hint at the potential of a recovery technology based on molecular bioscavengers.

As a first approach, data mining is used to obtain statistical information about how proteins in nature interact with phosphate groups, thereby revealing characteristic amino acid distributions of the binding sites. Quantum mechanical methods are used to investigate how phosphate moieties are described using electronic structure methods, and molecular dynamics in combination with quantum mechanics are used to show how the dynamical interaction between phosphates and proteins can be described – it is found that certain commonly used computational methods, including B3LYP, are ill-suited for characterizing interactions with phosphate groups, but nevertheless that phosphate-protein interactions can efficiently be quantified using other methods, e.g. wB97XD or PM6. Finally, it is shown how computational fluid dynamics can be used to optimize large-scale industrial processes using an open-source model, which we have made freely available online to the membrane community, and the advantages/disadvantages of different potential physical implementations of the proposed scavenger technology are discussed.

Acknowledgements

I extend my sincerest thanks to my supervisor Claus Hélix-Nielsen for his great enthusiasm and support during the project. Throughout the project he has encouraged me and granted me the freedom to pursue my interests, as well as allowed me to work very flexible hours, which I am immensely grateful for.

I thank Per Jr. Greisen for spiking my initial interest for the field of computational biochemistry, both before and during the first stages of the project. Andrea Bordoni, for aiding me with DFT and QM simulations, and later (along with Sebastian Borchert) in always providing excellent support for the high-performance computer at DTU. Without their help, much of my work would have been hindered. I also thank all the exemplary bachelor and master students I have had the pleasure of supervising, especially: Ulf Aslak, Sigurd Truelsen, Thomas Østergaard and Morten Gruber, all of whom have helped to extend my own knowledge within the scientific studies they undertook during their projects.

I would like to thank all the people at Aquaporin A/S, which is where I have spent most of my time, for providing a friendly, educational and inspiring working environment. Especially thanks to the people of the biomimetic membrane group at DTU, with whom I've shared many laughs as well as serious discussions.

Lastly, I thank Emilie Baron for putting up with me throughout the entire project. Without your support I would not have been able to pull through and I look forward to our future life together.

Contents

List of Symbols

Physical constants

Constant	Description	Value
e	Electron charge	$1.602\,176 \times 10^{-19}\,\mathrm{C}$
ϵ_0	Vacuum permittivity	$8.854\,187 \times 10^{-12}\,\mathrm{F\,m^{-1}}$
m_e	Electron mass	$9.109\,38 \times 10^{-31}\,\mathrm{kg}$
$\hbar$	Planck's constant, $\hbar = h/2\pi$	$1.054\,571 \times 10^{-34}\,\mathrm{J\,s}$

Symbols used in Chapter 3

Latin	Description
E_n	Total (nuclear-electron) energy
E_e	Electronic energy
E_H	Coulomb (Hartree) energy
E_{XC}	Exchange-Correlation Functional
h	One-body Hamiltonian
H	Hamiltonian
J_{ij}	Coulomb integral
K_{ij}	Exchange integral
N	Number of electrons
$\mathbf{r}$	Coordinates of all electrons
$\mathbf{r}_i$	Coordinates of electron i
$\mathbf{R}$	Coordinates of all nuclear particles
T_R	Kinetic energy of electrons
V_{ext}	External Potential

Greek	Description
ε_i	KS orbital energy
ρ	Electron density
ϕ	Electronic spatial orbital
ψ	Electronic spin-orbital
Ψ_n	Total (electron-nuclear) wave function

Symbols used in Chapter 4

Latin	Description
A_{ij}	Repulsive Lennard-Jones parameter
B_{ij}	Attractive Lennard-Jones parameter
f_i	Force experienced by particle i
K_r	Force constant for bonds
K_θ	Force constant for angles
m_i	Mass of particle i
n	Periodicity
r	Bond lengths
r_{eq}	Equilibrium Bond length
r_i	Position of particle i
q	Charge of particle
R_{ij}	Distance between particles i and j
t	time
v_i	Velocity of particle i
$V_n/2$	Amplitude
V_{pot}	Potential energy function

Greek	Description
γ	Torsional displacement
θ	Bond angles
θ_{eq}	Equilibrium Bond angle
ϕ	Torsional angle

Symbols used in Chapter 5

Latin	Description	Unit
A	pure water permeability coefficient	$\mathrm{m\,(s\,Pa)^{-1}}$
B	solute permeation coefficient	$\mathrm{m\,s^{-1}}$
D_{AB}	binary solute diffusion coefficient	$\mathrm{m^2\,s^{-1}}$
$\mathbf{f}$	body force per unit mass	$\mathrm{N\,kg^{-1}}$
$\mathbf{J}_s$	solute flux	$\mathrm{kg\,(m^2\,s)^{-1}}$
$\mathbf{J}_w$	water permeation flux	$\mathrm{m\,s^{-1}}$
K	membrane mass transfer coefficient	$\mathrm{s\,m^{-1}}$
m_A	solute mass fraction	$\mathrm{kg\,kg^{-1}}$
$\mathbf{n}$	surface normal unit vector	-
p	pressure	Pa
t	time	s
$\mathbf{U}$	fluid velocity vector	$\mathrm{m\,s^{-1}}$

Greek	Description	Unit
μ	viscosity of fluid	Pa s
π	osmotic pressure	Pa
ρ	fluid density	$\mathrm{kg\,m^{-3}}$

Subscript	Description
bulk	bulk value
D	draw side of membrane
eff	effective value
F	feed side of membrane
m	membrane surface value

List of Abbreviations

Abbreviation	Interpretation
2D	Two-Dimensional
3D	Three-Dimensional
ADP	Adenosine Diphosphate
AL-FS	Active Layer – Feed Side
aMD	Accelerated Molecular Dynamics
ATP	Adenosine Triphosphate
BC	Boundary Condition
CFD	Computational Fluid Dynamics
CI	Configuration Interaction
cMD	Conventional Molecular Dynamics
CP	Concentration Polarization
CPU	Central Processing Unit
DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
dGMP	Deoxyguanosine Monophosphate
EBPR	Enhanced Biological Phosphorus Removal
ECP	External Concentration Polarization
FO	Forward Osmosis
FVM	Finite Volume Method
GGA	Generalized Gradient Approximation
GppNp	5'-Guanylyl Imidodiphosphate
GTP	Guanosine-5'-triphosphate
HF	Hartree-Fock
HSP70	Heat Shock Protein 70
ICP	Internal Concentration Polarization

KS	Kohn-Sham
LDA	Local Density Approximation
LSDA	Local Spin Density Approximation
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
MP2	2nd Order Möller-Plesset
MP3	3rd Order Möller-Plesset
MP4	4th Order Möller-Plesset
NPT	Isothermal–isobaric Statistical Ensemble
NVE	Microcanonical Statistical Ensemble
NVT	Canonical Statistical Ensemble
P	Phosphorus
PBP	Periplasmic Phosphate Binding Protein
PES	Potential Energy Surface
PLP	Pyridoxal-5'-Phosphate
PRO	Pressure Retarded Osmosis
QM	Quantum Mechanics
QM/MM	Quantum Mechanics / Molecular Mechanics
REMD	Replica Exchange Molecular Dynamics
RNA	Ribonucleic Acid
RO	Reverse Osmosis
SCF	Self-Consistent Field
SGLD	Self-Guided Langevin Dynamics
vdW	van der Waals
WFT	Wave Function Theory
XC	Exchange Correlation

CHAPTER 1

Introduction

Phosphorus is an essential element for all known life and a key component of a wide range of cellular systems. On the global scale, phosphorus is used by humans to produce fertilizers, without which farmers could not achieve the high crop yields that are needed to feed the world population. If the supply of such fertilizers was ever to fail, humanity would face widespread famine on a scale unlike anything ever seen before in known history.

By far the largest source of phosphorus is mined phosphate rock, a finite resource that is becoming increasingly expensive and subject to geopolitical tensions as around 85% of the world's remaining reserves are controlled by just five countries [1, 2]. It is predicted that the growing demand for phosphorus could potentially surpass the production rate from mined rocks in the coming decades, leading to a terminal decline in the availability of phosphorus, yet there are currently no explicit international guidelines, policies or organizations responsible for ensuring a long-term supply of phosphorus [3, 4]. Estimates of when such *peak phosphorus* production may be reached vary, ranging between 15-25 years [4, 5], leading to estimates for when the reserves are fully depleted of 100-400 years [4, 6, 7].

A majority of the mined phosphate rock (approximately 80%) is used to produce mineral fertilizers for agricultural applications [8]. Unfortunately, only a fraction of the phosphorus (approximately 40%) makes it into the intended crops while the rest is lost along the way, causing serious environmental problems in the form of eutrophication of lakes and other aquatic ecosystems [8, 9]. In a sense, we are facing a unique problem of both having too little and too much phosphorus, and the current environmental challenge along with the looming threat of a future phosphorus crisis calls for a need to stimulate sustainable phosphorus practices and technology development.

From the technology point-of-view, phosphorus has been a subject of interest for several decades [10]. The focus has however mainly been directed on how to most efficiently remove phosphorus and only recently is resource recovery starting to get increasing attention [4, 6, 8]. Common approaches for phosphorus removal involve chemical precipitation of the inorganic phosphate anions with cations of iron, aluminium, calcium, magnesium, and/or ammonium, thereby producing a chemi-

cal sludge that is usually disposed off in a landfill [4]. Another popular removal method is enhanced biological phosphorus removal (EBPR), where bacteria capable of capturing and storing high amounts of phosphorus are used [11]. The phosphorus contained in the resulting biomass can subsequently be used as a soil amendment, or be disposed off in a landfill. These methods have been developed for removing phosphorus from municipal wastewaters; however, the job at hand is to capture phosphorus from the largest phosphorus flows in a manner where it can be reused on a global scale, and the current methods are not necessarily going to be effective to address that issue [8].

In 2012, Satorius *et al.* conducted a survey of experts in the field of phosphorus recovery, showing that phosphorus recovery is expected to become a well-established process over the next 15-20 years for economic reasons and that the market for recovery technologies is projected to increase in the immediate future [12]. Many full-scale installations for phosphorus recovery are already operational in Europe, North America and Asia, demonstrating that recovery is technically feasible using existing technologies in the form of chemical precipitation and EBPR [4]. However, beyond technical feasibility also economical aspects, legislation, and national policies have an influence on the large-scale implementation of a technology. Adding in considerations of socio-cultural feasibility [3], environmental benefits [13], and multi-stakeholder supply chain risks [2], it is clear that the overall viability of any recovery technology is difficult to quantify. Looking only at the price of phosphorus in phosphate rock (0.2–0.8 €/kg P) there is currently no economic incentive for phosphorus recovery, simply because the operating costs of current technologies are too high (1.6–8.8 €/kg P) [4]. Going a step further and taking into account the environmental benefit of avoiding phosphorus discharge and/or the cost of already-in-place phosphorus removal technologies, the recovery process can be shown to be economically profitable at certain wastewater treatment plants [4, 13]. Even so, it is difficult to improve on the current technologies [4], and there is room for the development of new technologies that are (i) capable of recovering phosphorus with high efficiency, (ii) capable of handling phosphorus flows that are difficult to intercept (e.g. erosion and runoff), and (iii) can operate at even lower costs such that the recovered resource can compete with naturally mined phosphate rock, thereby greatly facilitating implementation of the technology [4, 8].

1.1 A New Biomimetic Technology

In the quest of pursuing new technology humans have always looked at nature for inspiration and solutions to our problems [14]. For the problem at hand, it is remarkable to note that nature has spent billions of years refining the quintessential methods with which cells utilize phosphorus [15]. In fact, nearly 20% of all known proteins are known to interact with phosphate either in its inorganic form or bound to an organic moiety such as a nucleotide [16] – the plural term *phosphates* is therefore often used in this thesis to refer to any compound containing the inorganic phosphate anion PO_4^{3-}.

An example of phosphorus utilization in bacteria and archaea can be found in the *Phosphate specific transport system* (Pst), where high-affinity periplasmic phosphate binding proteins (PBPs) play a central role in enabling phosphate transport to operate with high affinity and high selectivity at very low phosphate concentrations [17, 18]. Even when looking at arsenate, which has striking similarities to phosphate (nearly identical pK_a-values, similarly charged oxygens, only ∼4% larger thermochemical radii), some PBPs are capable of up to 4,500-fold discrimination between the very similar molecules [19]. The PBPs are not consumed in the process, nor does the process involve chemical bonding to the phosphate (as in the case of chemical precipitation or EBPR); the PBPs bind in a reversible manner, capturing and releasing phosphate as required to facilitate phosphate transport in the cell [18], see Fig. 1.1. Such highly specific and regulated interactions with phosphate are an integral part of biological systems, and it motivates the idea that these systems could be imitated in new, *biomimetic* phosphorus recovery technologies.

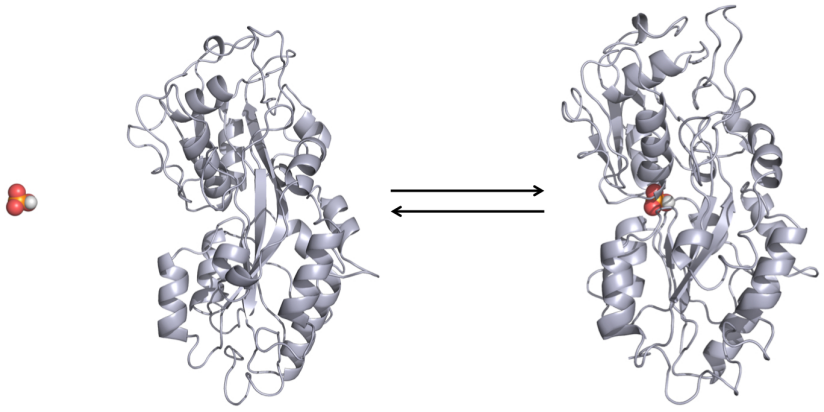

Figure 1.1: Illustration demonstrating the binding of phosphate (HPO_4^{2-}) by a periplasmic phosphate-binding protein (PBP). PDB #4F1V.

The idea of using proteins in biomimetic technologies is not new, one recent and prominent example being the incorporation of aquaporin proteins into commercial water filtration membranes [20]. Aquaporins are proteins used by cells to selectively transport water across lipid bilayer membranes: by purifying and inserting the aquaporins into membrane support structures, it is possible to achieve highly efficient biomimetic membranes capable of operating for extended periods of time with high water permeability and solute rejection coefficients [20–22]. Another example is the application of enzymes in industrial processes, enzymes being biological catalysts capable of enhancing the rate of chemical reactions in living organisms. The use of extracted enzymes in industry represents a biomimetic technology, where the number of applications has exploded in recent years, mainly owing to the advances in protein engineering technologies and environmental/economical necessity [23–25].

Turning to consider the creation of a biomimetic technology for phosphorus recovery, one intuitive option is to base the design on *protein bioscavengers* capable of capturing and releasing phosphorus compounds in a controlled manner and incorporating these proteins into a setup suitable for large-scale industry. From a practical point of view, such a technology raises a series of fundamental questions at different temporal and spatial scales, among these being (i) how proteins chemically interact with phosphates and how to design an optimal binding site, (ii) how can proteins be engineered to capture and release phosphates in a controlled and reversible manner and what are the dynamics of such molecular mechanisms, and (iii) how can the designed bioscavengers be implemented in robust and scalable industrial setups. Addressing these questions are challenging, in part because the length scales of importance range from sub-nanometer for the interactions responsible for selective binding of phosphates, to potentially several meters for large-scale industrial applications.

Especially questions about the smallest length scales can be difficult to address since those scales are not easily observed experimentally. Therefore, theoretical models capable of describing biomolecules are very attractive, and with advances in modern computing power in recent years, the use of computational techniques capable of describing molecular systems in atomistic detail have therefore become increasingly popular. The scientific fields surrounding such numerical models are vast, with focus being divided into many different basic methodologies and focus areas, including *Quantum Mechanical* (QM) calculations, which provide information about the electronic structure of atoms and molecules [26, 27], *Molecular Dynamics* (MD) simulations, which provide transient information about molecular movements [28], *Monte Carlo* (MC) simulations, which can be used to sample molecular con-

formations [29], and data mining techniques, which can be used to analyse large amounts of data such as known protein structures and sequences [30, 31].

Turning from the design of the biomolecules to the design of application modules, a bioscavenger technology can be imagined in a variety of different setups. These range from having the scavengers free in solution, to attaching the scavengers to porous substrates, and/or combining the scavengers with existing technologies, e.g. membrane technologies such as reverse osmosis (RO) and forward osmosis (FO) [32]. Common for many potential applications is that they involve fluid flows, and also at these larger length scales computational methods are firmly established as an integral part of the industrial development process, namely, a broad range of techniques grouped together under the common name of *Computational Fluid Dynamics* (CFD) can be used to simulate and analyse problems involving fluid flows [33].

As a first step towards developing a biomimetic resource recovery technology, it is compelling to gain as much theoretical insight about the systems of interest as possible beforehand. Computational techniques provide an inexpensive way of obtaining such information, and can therefore greatly assist subsequent efforts towards realizing the technology in experimental settings.

1.2 Aims of the Thesis

The intent of the work performed in this thesis is to apply computational techniques to acquire information relevant for the creation of a biomimetic technology for phosphorus recovery. More specifically, the aims can be listed as follows:

Aim 1 Use data mining to analyse experimentally determined protein structures known to interact with phosphate moieties, thereby obtaining information about how binding sites for phosphates are designed in nature.

Aim 2 Apply electronic structure methods to investigate the fundamental molecular properties of phosphates and the differences in properties between phosphates and analogues molecules such as arsenates and sulphates.

Aim 3 Gain an understanding of the dynamics involved in phosphate binding in proteins by using molecular dynamics to sample the conformational ensemble of an intrinsically disordered peptide known to selectively bind phosphate.

Aim 4 Develop and demonstrate how a computational fluid dynamics model can be used to optimize module design in a technology based on forward osmosis.

1.3 Thesis Outline

The thesis is written around the four studies presented in **Papers I - IV** and the four aims listed in Sec. 1.2, with each aim having a full chapter devoted to motivating the study in question, providing relevant background information and theory, as well as discussing the results and perspectives of the study.

Chapter 2 Describes how data mining can be used to obtain information about how proteins in nature bind phosphates. The chapter introduces and reviews how proteins in general interact with phosphates.

Chapter 3 Gives a rudimentary introduction to electronic structure methods, reviews how such methods are used in the biochemical literature surrounding phosphate moieties, and investigates the use of such methods for describing phosphate, sulphate and arsenate molecules.

Chapter 4 Reports on how a combination of quantum mechanics and molecular dynamics can be used to examine the dynamical interaction between phosphate and a short phosphate-binding peptide in solution.

Chapter 5 Taking forward osmosis as a basic membrane technology upon which a phosphorus recovery application could be based, it is shown how computational fluid dynamics can be used to optimize module design.

Chapter 6 Summarizes the obtained results and discusses their implications for the future development of a biomimetic phosphorus recovery technology.

The thesis is structured in this manner because there is a high diversity in the theoretical background underlying each study, and therefore readability is increased by taking each study at a time rather than clumping together all the theoretical knowledge in one initial chapter.

CHAPTER 2

Data Mining

During the last few decades, biologists have continuously accelerated their efforts in understanding biological processes, an undertaking which has resulted in a surge of biological data. The resulting information is disseminated in various heterogeneous databases and more than 500 of such databases can now be found in the scientific literature [34]. Examples include the GenBank database, which at the time of writing contains more than 100.000.000 DNA sequences [35], and the RCSB Protein Data Bank which contains more than 110.000 experimentally determined protein structures [36].

Exploiting biological databases for the discovery of new knowledge poses a series of fundamental data analysis difficulties, e.g. issues with handling noisy and incomplete data, the processing of computationally intensive tasks, and integrating data from different sources [37]. Bioinformatics, or computational biology, is the interdisciplinary science of interpreting biological data using information technology and computer science. A particularly active area of research is the application and development of data mining techniques, data mining being a subfield of computer science where computational algorithms are used for discovering patterns in large data sets (”big data”) using methods derived from artificial intelligence, machine learning, statistics, and database systems [38, 39].

Of all the known protein structures, nearly 20% interact with phosphates, either in the inorganic form or bound in an organic moiety [16], and as a first step towards understanding how proteins in nature bind phosphates, it is therefore compelling to extract as much information from these structures as possible. In this chapter a literature review for what is known about how proteins in nature bind phosphates is presented and the data mining study in **Paper I** is motivated and discussed.

2.1 Phosphate Binding Proteins in Nature

The ubiquitous nature of phosphates in biology means that their interactions with proteins have been extensively studied in the scientific literature, both in the case of individual proteins binding to inorganic phosphates and for proteins binding to phosphate moieties such as nucleotides or co-factors [40]. Going back to some of the earliest discoveries, in 1974 Rossmann *et al.* identified a protein fold now known as the "Rossmann fold", which is commonly found in mono and dinucleotide binding proteins [41]. From the Rossmann fold, two glycine-rich consensus sequences were identified, GXXGXGK(S,T) and GXGXXG, which were found to be involved in mono- and dinucleotide binding, respectively [42, 43]. Later in 1982, Walker *et al.* investigated distantly related sequences of a series of ATP-requiring enzymes and identified highly conserved consensus sequences, which were denoted *Walker A* and *Walker B* motifs [44]. The Walker A motif is now also commonly known as the "P-loop", for the role it plays in interacting with phosphate compounds, see Fig. 2.1.

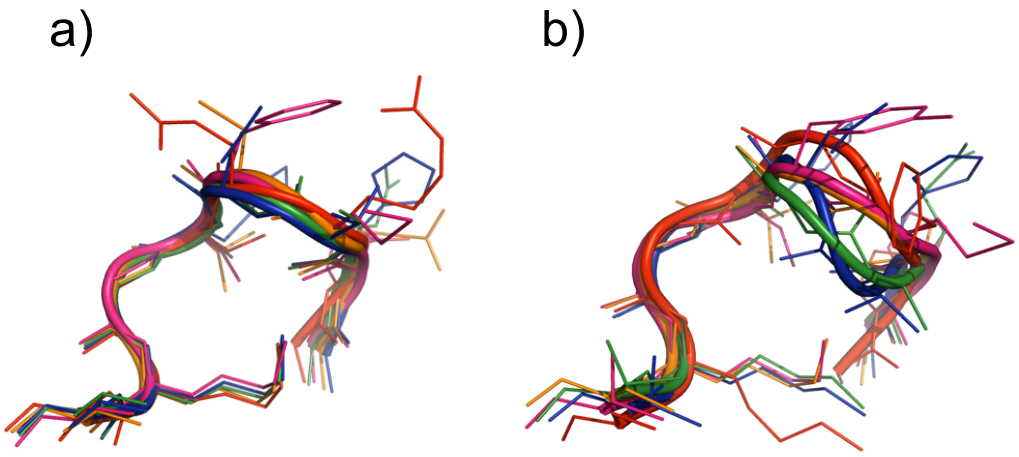

Figure 2.1: Alignment of randomly selected P-loops from different protein families. **a)** loops from structures with bound ligands (ligands not shown), **b)** loops from structures without bound ligands.

.

The consensus sequence for the P-loop is GXXXXGK(S,T), although within some protein families it is possible to refine the sequence, such that for example in adenylate kinases the consensus sequence is GXPGXGKGT [45]. In general, the glycine residues (G) in the P-loop and Rossmann fold sequences are believed to adopt conformations that would not be tolerated by any other amino acid with a

side chain, and the conserved lysine residues (K) are postulated to be important both for the conformation of the P-loop, as well as for stabilization of the interaction with the β and γ phosphates in mono-nucleotides [40]. Whereas in di-nucleotide binding motifs, a glycine-rich Rossmann fold is often located at a tight turn between a β strand and an α helix [46], the mono-nucleotide binding proteins often contain rather long P-loop sequences that connect a β strand and an α helix, and the P-loop is therefore sometimes referred to as "a giant anion hole" [47]. The classical P-loop generally binds the common *anti* conformation of ADP/ATP; however, a "novel P-loop" with the highly conserved sequence PXXXGLGSSAA has also been identified, which binds the unusual *syn* conformation instead [48].

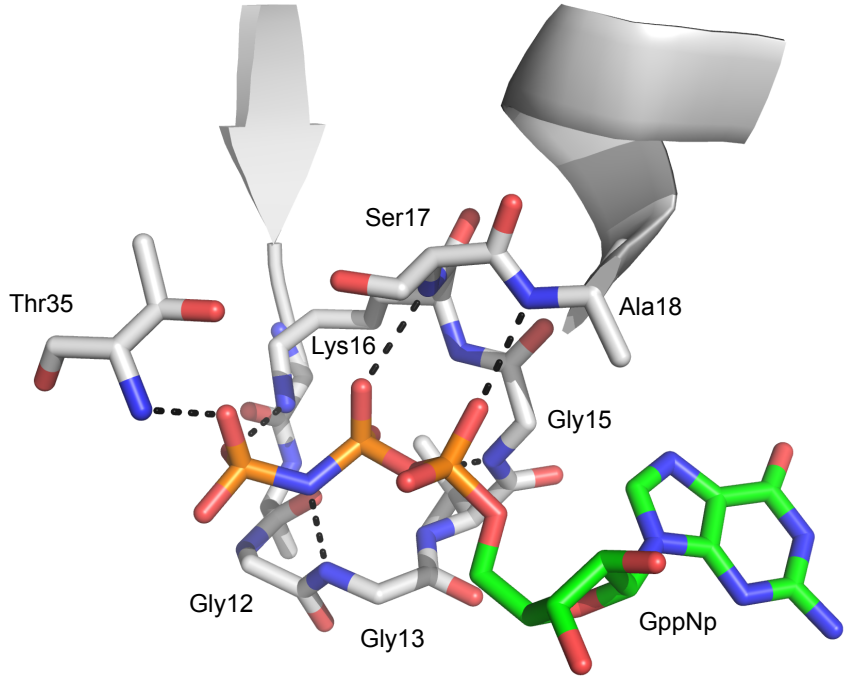

Figure 2.2: Region of the crystal structure for the protein p21, showing how a GTP analogue (GppNp) is bound in a site containing a P-loop (GAGGVGKS) connecting an α helix and a β strand. Atoms are colored as follows: carbon atoms in protein (gray), carbon atoms in ligand (green), oxygen (red), nitrogen (blue), phosphorus (orange). PDB #5P21.

Beyond the Rossmann folds and P-loops, Denesyuk *et al.* identified another novel anion binding motif known as the $C^{\alpha}NN$ structural motif, which includes one C^{α} atom and two backbone N atoms interacting with pyridoxal-5'-phosphate (PLP) [49]. Another commonly observed structural motif which is often used by Actin,

HSP70 (heat shock protein 70) and some sugar kinases, consists of residues from β hairpins which can tightly interact with phosphates [48].

Taken together, there are a variety of conserved structural motifs in phosphate-binding proteins, but the consensus for many of these motifs are that they often feature binding sites containing several glycine residues as part of a loop structure, an adjacent lysine residue participating in the phosphoryl interaction, and also often the positively polarized N terminus of an α helix [40]. In many binding sites, it is astounding how the P-loops "wrap" tightly around the phosphate moieties with a highly specific and near-macrocyclic organization of NH-residues towards the bound anions, see Fig. 2.2. These finely tuned spatial arrangements of residues in the binding sites are suggested to be what allows the proteins to differentiate between molecules such as phosphate and sulphate, even though the differences in size and charge for these are very subtle [50].

For a long time investigations of phosphate binding proteins had primarily been focused on the information that could be obtained from single X-ray crystal structures. In 2007, however, Hirsch *et al.* made one of the first statistical evaluations of a large number of crystal structures obtained from phosphate-binding proteins (3003 structures in total) [40], thereby providing more comprehensive insight into the nature of how proteins interact with phosphates. These analyses revealed several interesting observations; until then, it was believed that most phosphates were bound by proteins with the assistance of a co-bound metal cation, but the statistical analysis showed that this was only the case for one third of the sampled structures, and 1070 of the structures even bound phosphates in the absence of both metal cations as well as any positively charged amino acid residues such as lysine or arginine. Their study furthermore revealed characteristic distributions of amino acids in the binding sites which could be used as "fingerprints" for the different classes of phosphate binding proteins, as well as concluded that phosphate binding sites show a strong dependence on location, such that towards the protein surface cationic residues and metals are commonly found, while binding sites deep inside the proteins have a higher tendency of neutral amino acids [40].

2.2 Study Motivation & Paper I

The studies by Hirsch *et al.*, as well as previous investigations into individual phosphate binding structures, focus primarily on the direct interactions between the phosphate moiety and the protein amino acids. The coordination of phosphorus compounds in a protein binding site is, however, likely also influenced by interactions that go beyond these first shell interactions. The ”shell” terminology is borrowed from the field of metalloproteins, where proteins are known to direct the interactions with metal ions not only by first shell protein-metal interactions but also indirectly (via protein-protein interactions) in a ”second shell”, i.e. protein residues interacting with protein residues in the first shell of the binding site [51]. For metalloproteins, this second shell is known to be of great importance, e.g. in protecting and shielding the binding site [52], stabilizing the binding site [51, 53], for enhancing the binding site affinity [54, 55] and, in general, fine-tune the environment of the binding site [56, 57].

Because second shell interactions play an essential role for metalloproteins, we hypothesized that also for phosphate binding sites the second shell interactions might play a significant role in tuning the physiochemical properties of the binding sites. Our motivation for the study at hand, as presented in **Paper I**, was therefore to perform a structural survey of all known phosphate binding proteins, and address the importance of both first and second shell interactions in the phosphate binding sites.

2.3 Conclusions & Outlook for Paper I

A total of 8307 protein structures were selected from the RCSB Protein Data Bank by their ability to bind different classes of phosphorus compounds and similarly to what Hirsch *et al.* observed, we also found that ∼74% of the structures bound the phosphorus compounds without the assistance of co-bound metal cations. Furthermore, it was shown that the amino acid distributions of the first shell are not simply determined by the type of phosphorus compounds, but also by the tendency of the different compounds to co-bind with metal cations: e.g. Mg^{2+} is rarely found in binding sites for inorganic phosphate, whereas it is more common in binding sites for pyrophosphate. The second shell of the phosphate binding sites was observed to be remarkably conserved across the investigated structures, suggesting that it is indeed important for stabilization and potentially also for fine-tuning the internal environment such as to reinforce specificity and affinity of the binding site.

The study as a whole contributes to the overall understanding of how phosphorus compounds are bound by proteins in nature. In addition to such fundamental information it also provides a tool for sampling information about binding sites for specific phosphorus compounds, which can be of use when designing bio-scavengers for phosphorus recovery. Designing a universal bio-scavenger may be challenging, but for a specific molecule such as inorganic phosphate, the study quantifies the relative importance of individual residues in both the first and second shell of the binding site. The information obtained by data mining as used in this study inherently contains an unquantified degree of noise because of the large number of structures investigated – in actively developing a bioscavenger for a specific compound, the grouping of structures can however be further refined based on e.g. protein family or structural features of interest, thereby reducing the amount of noise. Data mining may thus be an invaluable tool that can aid an iterative design process by providing information about the importance of the individual structural components of a binding site, or even provide suggestions for improving the binding site.

CHAPTER 3

Electronic Structure Methods

A central goal of this project was to investigate phosphate molecules and their interactions with proteins using various computational methods. Such computational methods are derived from different scientific theories, a theory being a well-substantiated explanation of some aspect of the natural world as acquired through scientific methods and repeated experimentation and observation. Theories generally aim to be inductive in nature and have predictive/explanatory capability, and perhaps one of the most basic theories is that of quantum mechanics (QM). Said in another way, QM is a fundamental branch of science that is concerned with, among other things, the behaviour of atoms, molecules and electrons. Electronic structure methods, which are the computational methods investigated in this chapter, derive from the laws of QM rather than classical physics, and as suggested by the name specifically relate to predicting the electronic structures of atomic and molecular systems.

In this chapter, a short introduction to QM and computational chemistry is given. A comprehensive overview of the scientific fields surrounding electronic structure calculations would require several weighty volumes and the following textbooks can be recommended for the interested reader: Jensen [26], Atkins and Friedman [58] and Kohanoff [59]. Computational QM methods are frequently used by molecular biologists for investigating systems of interest, and therefore it is reviewed in this chapter how these methods are employed in the scientific literature to describe phosphate-binding proteins. Finally, potential pitfalls of using QM methods for describing protein-phosphate interactions are highlighted, which motivates the study presented in **Paper II**.

3.1 Molecular Quantum Mechanics

Neglecting relativistic effects, QM theory stipulates that the energy and all other related properties of a collection of electrons and nuclei are governed by a Hamiltonian H, and that all these properties can in principle be obtained by solving the time-independent Schrödinger equation [59]:

$$H\Psi_n(\mathbf{R}, \mathbf{r}) = E_n\Psi_n(\mathbf{R}, \mathbf{r}) \tag{3.1}$$

where E_n are the energy eigenvalues and $\Psi_n(\mathbf{R}, \mathbf{r})$ the corresponding many-body wave functions, $\mathbf{R}$ is a set of nuclear coordinates and $\mathbf{r}$ is a set of electron coordinates. In more practical terms, the Hamiltonian is an operator that includes terms for the kinetic energies of the nuclei and electrons, the electron-nucleus attractive potential, and the nucleus-nucleus and electron-electron repulsive potential energies [59]. The meaning of the wave functions is provided in Born's statistical interpretation, which states that $|\psi(\mathbf{r}_i, t)|^2$ is the probability of finding a particle with wave function ψ at a given location $\mathbf{r}_i$ at time t [60]. Solving Eq. (3.1) typically poses a daunting challenge given the fact that often it is a multi-component many-body system interacting through two-body Coulomb interactions, and analytical solutions are thus limited to only a few cases such as the hydrogen atom, and even numerical solutions are limited to small molecules [58].

Instead of attempting to solve the Schrödinger equation for all particles simultaneously, the Born-Oppenheimer approximation can be employed (under appropriate conditions [59]), which essentially assumes that electrons can respond almost instantaneously to a displacement of the nuclei, and therefore that the nuclei can be regarded as fixed [58]. The electronic configurations can then be calculated in the "static" potential of the nuclei, and from a set of such calculations (at different nuclear conformations), a Potential Energy Surface (PES) can be constructed, see Fig. 3.1. The nuclei are also quantum objects with associated wave functions; however, if the nuclei are considered as classical particles, the force on them could be calculated from the gradient of the electronic energy $\nabla E_e(\mathbf{R})$, and thereby their equations of motion could be derived from electronic calculations. Luckily this approximation turns out to be justified by looking at the thermal wavelengths of the nuclei at room temperature which in the worst-case scenario of hydrogen is only 0.2Å [59].

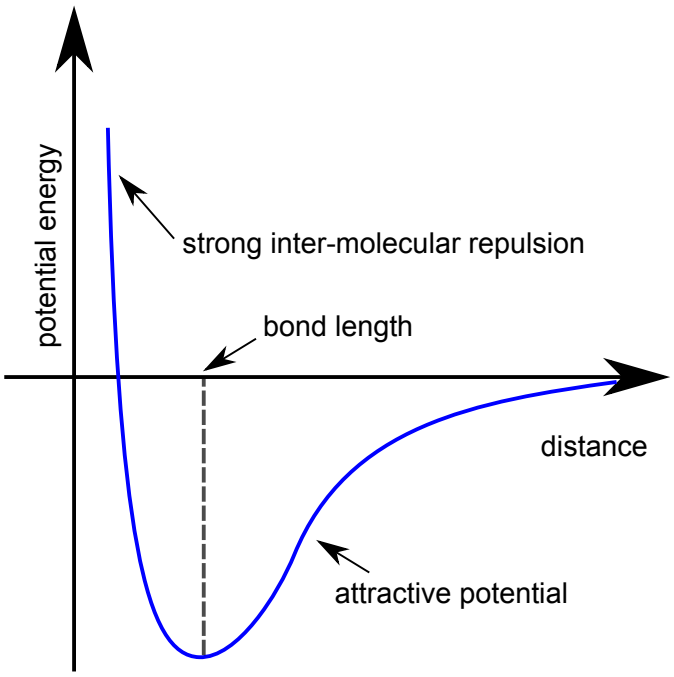

Figure 3.1: Sketch of potential energy curve for a di-atomic molecule. The distance represents the distance between the two atoms..

A typical goal is to solve $\nabla E_e(\mathbf{R}) = 0$ and thereby find a local minimum on the PES, a procedure which is known as a *geometry optimization*. However, to obtain $\nabla E_e(\mathbf{R})$ it is still necessary to solve the time-independent Schrödinger equation, a procedure which is known as an *electronic structure calculation*. In the following Secs. 3.1.1 - 3.1.2, two fundamental approaches to such electronic calculations are introduced, namely the wave function theory (WFT) and density functional theory (DFT) methods.

3.1.1 Wave Function Methods

In *ab initio* (first principles) WFT methods, a model is chosen for describing the electronic wave function, and based on this model the wave function of the system is determined using only fundamental physical constants and the atomic numbers. The accuracy of the method is thus determined solely by the form of model wave function and as the accuracy of the model wave function is increased, its solution accordingly approach that of the real solution of the Schrödinger equation [27].

One of the most fundamental WFT methods still in use is that of Hartree-Fock (HF), see [61] for a recent review. In HF theory, a complete set $\{\psi_i\}$ of orthonormal spin-orbitals are considered, where each spin orbital is given as the product of a

spatial orbital $\phi(\mathbf{r}_i)$ and a spin function (either spin up, α, or spin down, β). Given that electrons are fermions, the total wave function must be anti-symmetric under exchange of particles: this is satisfied in HF theory by assuming that the ground-state wave function Ψ_0 can be constructed as a product of orbitals as arranged in a single Slater determinant, noting that determinants possess the general property that they change sign whenever two rows or columns are interchanged [62]:

$$\Psi_0 = \frac{1}{\sqrt{N!}}\det|\psi_1\psi_2...\psi_N| \tag{3.2}$$

where N is the number of electrons. The ground-state energy E_0^{HF} for the HF method can be shown to be [61]:

$$E_0^{\mathrm{HF}} = \sum_i^N \underbrace{\langle\psi_i|\hat{h}|\psi_i\rangle}_{h_i} + \frac{1}{2}\sum_{i,j}\left[\underbrace{\langle\psi_i\psi_j|\frac{1}{r}|\psi_i\psi_j\rangle}_{J_{ij}} - \underbrace{\langle\psi_i\psi_j|\frac{1}{r}|\psi_j\psi_i\rangle}_{K_{ij}}\right] \tag{3.3}$$

where $\hat{h}$ is a one-body *core Hamiltonian* with corresponding one-electron integral h_i, which describes the motion of a single electron interacting with all the nuclei in the system [61]. Eq. (3.3) introduces the two-electron integrals J_{ij} and K_{ij}, which are known as the Coulomb and Exchange integrals, and it is noted that when $i = j$ they lead to exact cancellation [58]. Turning back to the electronic calculation, i.e. the solving of Eq. (3.3), this can be performed by using the Euler-Lagrange variational principle, where the N spin-orbitals $\{\psi_i\}$ are varied until the energy E_0^{HF} reaches a minimum value [58]. Practically this is done by forming an initial guess of $\{\psi_i\}$, which are used to form a so-called *Fock operator*, that in turn can be used to estimate a new guess of the spin-orbitals, a process which can be repeated until a convergence criterion related to the energy change has been met [58]. The resulting spin orbitals are said to be *self consistent* and calculations of this type are known as self-consistent field (SCF) calculations [26].

Although HF theory has been a major achievement in computational physics and chemistry, it is a somewhat crude approximation to the true many-body ground state, given that it is based on a single determinantal wave function as specified in Eq. (3.2). A single-determinant approximation does not correctly take correlation effects into account and the energy of HF is always higher than the exact energy [59]. The difference, $E_0^{\mathrm{HF}} - E_0^{\mathrm{Exact}}$, was coined by Löwdin as the *correlation energy* [63] and in practise is a term used to describe the inadequacy of the HF model. The word *correlation*, however, has to be used with caution since HF does include a certain amount of electron correlation arising from the antisymmetry of the wave function in Eq. (3.2), which prevents two electrons of parallel spin from being found at the

same point in space (accordingly also known as *Fermi correlation*) [64]. Another contribution is that of *Coulomb correlation*, which describes the correlation between spatial positions of electrons due to their Coulomb repulsion – electrons in HF can be said to only experience the average electron density of the other electrons, while in reality electrons are particles that experience instantaneous Coulomb repulsion upon encountering other electrons [26, 58, 59]. For a recent review of electron correlation in computational chemistry see ref. [64].

To account for electron correlation, several post-HF methods have been developed, including but not limited to Möller-Plesset methods (MP2, MP3, MP4 etc.) [65], where electron correlation effects are added by means of perturbation theory, and Configuration Interaction (CI) methods, where instead of a single determinant wave function, a linear combination of the ground state and excited state determinants is used to construct the wave function [66]. For more detailed information on the post-HF methods see ref. [26].

3.1.2 Density Functional Theory

In HF and post-HF methods, spin-orbital wave functions are sought that can be used to solve the Schrödinger equation. HF by itself however lacks a proper description of the electron correlation energy, and post-HF methods are often many times more computationally expensive than HF and can thus only be applied to very small molecules of interest. DFT methods are therefore an attractive alternative and are often used for larger systems since they include electron correlation effects and only require about as many computational resources as HF [27].

It is somewhat of a controversial question whether DFT methods can be denoted as *ab initio* given that many methods are parametrized [27]; however, DFT in its purest form as a theory is principally exact and like any *ab initio* theory it is based only on fundamental physical constants and atomic numbers. DFT is based on the Hohenberg-Kohn theorem, published in 1964, which establishes that the ground state energy (and all other ground-state electronic properties) of a system can be uniquely determined from the electron density, $\rho(\mathbf{r})$ [67]. Unfortunately, the theorem does not provide the functional form for how the energy depends on the electron density; it only proves that such a functional exists. In 1965 W. Kohn and L. J. Sham showed that the energy of an n-electron system can be written as a function of the electron density as follows [59, 68]:

$$E[\rho] \;=\; T_R[\rho] + \int \rho(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) \mathrm{d}\mathbf{r} + E_H[\rho] + E_{XC}[\rho] \tag{3.4}$$

with:

$$T_R[\rho] = -\frac{\hbar^2}{2m_e}\sum_{i=1}^{N}\langle\psi_i(\mathbf{r})|\nabla^2|\psi_i(\mathbf{r})\rangle \tag{3.5}$$

$$E_H[\rho] = \frac{1}{2}\int\int\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{3.6}$$

Looking at Eq. (3.4), the first term $T_R[\rho]$ represents the kinetic energy of the electrons, the second term is the external potential acting on the system (at minimum the electron-nucleus interaction), and the third term, E_H, is the Coulomb (or Hartree) energy. The last term is denoted the exchange-correlation energy, and it is the only term for which an exact expression is not known [26]; the Hohenberg-Kohn theorem demonstrates that E_{XC} is a functional of the electron density ρ, but does not provide an analytical form for it. The spin-orbitals of Eq. (3.4) are known as Kohn-Sham (KS) orbitals, and once these have been computed, the electron density can be calculated from the sum of the orbitals, namely:

$$\rho(\mathbf{r}) = \sum_{i=1}^{N}|\psi_i(\mathbf{r})|^2 \tag{3.7}$$

Different approximations of the E_{XC} functional are discussed in Sec. 3.1.3. The variational principle can be applied to the electronic energy in Eq. (3.4) to obtain the *KS equations* for the one-electron orbitals $\psi_i(\mathbf{r}_1)$ [59]:

$$\left\{-\frac{\hbar^2}{2m_e}\nabla^2 + v_{\text{eff}}(\mathbf{r})\right\}\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}) \tag{3.8}$$

with:

$$v_{\text{eff}}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + \int\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}' + \frac{\delta E_{XC}[\rho]}{\delta\rho(\mathbf{r})} \tag{3.9}$$

where ε_i are the KS orbital energies. Given an approximation for E_{XC}, Eq. (3.8) can be used in a SCF calculation, where initially a guess is provided for the electron density that is then used to calculate the approximate E_{XC} functional, which is used in Eq. (3.8) to obtain the KS orbitals, and from this new set of orbitals a new electron density can be computed using Eq. (3.7). In this manner the self-consistent procedure can be repeated until convergence, at which point the ground state energy and electron density has been determined.

3.1.3 Exchange-Correlation Functionals

The primary source of error in DFT typically stems from the approximate nature of the exchange-correlation functional, and the search for ever more accurate and computationally efficient functionals is therefore still an active area of research. By now numerous schemes for approximating the E_{XC} energy can be found in the scientific literature, and in this section only a bird's-eye view of the different functionals is presented. See ref. [26, 59] for more details.

The different exchange-correlation functionals often belong to a specific group, e.g. functionals may be based on the local density approximation (LDA) which assumes that the electron density varies slowly and that E_{XC} can be calculated using formulas derived for a uniform electron density, or alternatively be based on the generalized gradient approximation (GGA) where also the gradient of the electron density is considered [59]. The applicability of the different types of functionals depends largely on the system of interest, e.g. GGA can lead to large improvements over LDA especially in systems where the electron density undergoes substantial changes, such as in some molecules.

The exchange-correlation functionals are often separated into an exchange functional and a correlation functional for describing exchange and correlation energies, respectively. In this way many functionals are named by their combination of exchange and correlation functional, e.g. the popular functional BLYP is a combination of the GGA exchange functional by A.D. Becke [69] and the GGA correlation functional by C. Lee, W. Yang and R.G. Parr [70]. Beyond combining different exchange and correlation functionals, in 1993 A. Becke introduced the concept of hybrid functionals, where the DFT exchange term is mixed with the exact exchange from HF [71]. Within the field of organic computational chemistry, especially the hybrid functional B3LYP gained enormous popularity in the 90s, having the following functional form [72]:

$$\begin{aligned} E_{XC}^{\mathrm{B3LYP}} &= E_X^{\mathrm{LSDA}} + \alpha_0(E_X^{\mathrm{HF}} - E_X^{\mathrm{LSDA}}) + \alpha_x(\Delta E_X^{\mathrm{B88}}) \\ &\quad + \alpha_C E_C^{\mathrm{LYP}} + (1-\alpha_C)E_C^{\mathrm{VWN}} \end{aligned} \tag{3.10}$$

where E_X^{LSDA} is the local spin density approximation (LSDA) exchange functional [73], E_X^{B88} is Becke's gradient correction to the LSDA exchange functional [69], E_C^{LYP} is the Lee-Yang-Par correlation functional [70], and E_C^{VWN} is local correlation functional by Vosko, Wilk and Nusair [74]. The parameters $\alpha_0 = 0.20$, $\alpha_X = 0.72$ and $\alpha_C = 0.81$ of Eq. (3.10) have been chosen to best fit a database of molecular properties [71]. Although B3LYP and similar functionals have been an

immense contribution to the field of computational chemistry and are successful in many applications, they still lead to qualitative failures, noticeably in asymptotic regions of molecular systems where they can predict an exponential electronic density decay rather than the correct $1/r$ decay [75]. The self-interaction errors causing these incorrect trends can be qualitatively resolved by using long-range corrected functionals, where the amount of HF exchange is increased with the distance r, such that long-range electron-electron interactions are described almost 100% by HF exchange [76]. Another drawback of many DFT functionals is that they fail in properly describing dispersion interactions at inter-atomic distances, dispersion being defined as an attractive part of the van der Waals (vdW) interaction between non-bonded atoms, and therefore the development of dispersion-corrected functionals is also actively pursued [77].

Overall, all current DFT functionals are approximations and no functional (so far) is accurate for all properties of interest in all cases. Functionals are continuously improved, e.g. the "modern" long-range corrected functional ωB97X [75] was shortly after release updated to also include dispersion corrections leading to the functional ωB97X-D [76]. In a review paper by A. D. Becke from 2014, he starts off by stating *"Density-functional theory (DFT) is a subtle, seductive, provocative business. Its basic premise, that all the intricate motions and pair correlations in a many-electron system are somehow contained in the total electron density alone, is so compelling it can drive one mad."* [78] – highlighting both the advantages and challenges that computational chemists are faced with in working with DFT. In the end, it is left up to the user to use experience and intuition to decide which functional to use for a given problem, giving also consideration to the available amount of computational resources.

3.1.4 Basis Sets

A final aspect of electronic structure methods that deserves mention is their practical implementation, specifically in terms of the use of *basis sets*. Basis sets are mathematical descriptions of the orbitals within a system (which in turn are used to model the electronic wave function) and *larger* basis sets generally refer to mathematical descriptions that impose fewer restrictions on the locations of the electrons in space [27].

Common basis sets for QM methods are based on linear combinations of gaussian functions, mainly because these are computationally efficient when it comes to integral evaluation. A minimum basis set includes the minimum number of basis

functions required for representing all the electrons of each atom, e.g. the basis set STO-3G is a minimal basis set where three gaussian functions (accounting for the "3G") are used for each basis function. The basis sets can be made larger by increasing the number of basis functions per atom, e.g. in the *X-YZg* basis sets introduced by John Pople, X is the number of gaussian functions used for the core atomic orbitals, and Y and Z indicate that the valence orbitals are composed of two basis functions, one with Y gaussians and one with Z gaussians [79]. Examples include *split valence basis sets* such as 3-21G and 6-31G, or *triple-split valence basis sets* such as 6-311G [27]. To allow orbitals to change shape, it is common to add polarized basis functions to basis sets, such as in 6-31(d,p) where d functions are added to heavy atoms and p functions to hydrogens. Finally, for atoms and molecules where electrons are relatively far from the nucleus, diffusive basis functions can be added, e.g. in 6-31++G diffusive functions are added to both heavy atoms and hydrogen atoms, which allow the orbitals to occupy a larger region of space. More information on basis sets can be found in standard textbooks on computational chemistry, see refs. [26, 58, 59].

For the remainder of this thesis a *model chemistry* is used to refer to the particular combination of QM method and basis set, e.g. B3LYP/6-311++G(d,p) refers to calculations performed with the DFT functional B3LYP and the basis set 6-311++G(d,p).

3.2 QM and Phosphate Binding Proteins

Computational QM methods have had a tremendous impact on our understanding of small molecular systems during the last few decades and it is compelling to believe that the same level of knowledge can also be brought to biological systems [80]. Indeed, in 2011 a popular paper was published in Nature with the title "The Dawn of Quantum Biology", arguing that coherent quantum processes are ubiquitous in the natural world, noting that "*biology has a knack for using what works. And if that means quantum hanky-panky, then quantum hanky-panky it is*" [81].

One of the main obstacles in applying QM methods to biological systems, which typically involves on the order of thousands of atoms, are the computational bottlenecks of most electronic structure methods. Nevertheless, with the advance of faster computing resources and increased accessibility of computational chemistry software, QM methods are today almost routinely being used by biochemists around the world to investigate biological systems of interest [80]. As discussed in Chapter

2, interactions with phosphates are quintessential in biological systems and indeed the scientific literature is rich with examples where QM methods are used for describing phosphates and interactions with phosphates. A short overview of some of these studies is given in the following.

Small Phosphate Molecules

It is compelling first to look at the phosphate molecules themselves, all the way from the simple oxyacid H_3PO_4 and its anions ($H_2PO_4^-$, HPO_4^{2-} and PO_4^{3-}), to more complex organophosphates such as adenosine triphosphate (ATP). The earliest calculations were mainly performed in the gas phase, e.g. in 2000, Rustad *et al.* performed gas-phase calculations of a series of oxyacids (including phosphates) using B3LYP with different basis sets, and from the results they obtained geometric structures, gas-phase acidities, and vibrational frequencies for the molecules that are in good agreement with experimental values [82]. In another study performed by Kish *et al.* in 1999, HF/3-21(d) was used to identify substrates that could potentially substitute for phosphate in a dehydrogenase enzyme, which was done by looking at bond lengths and charge distributions in phosphate along with a series of phosphate analogous [83].

In 2012, M. Rudbeck published a study that showed how several biologically relevant phosphate molecules are described with different model chemistries, demonstrating 1) that inclusion of diffusive and polarization basis functions influence the obtained vibrational frequencies in the calculations, and 2) that also the inclusion of an implicit continuum solvent model had a substantial impact on the frequencies of these small organic phosphate molecules [84]. This study highlights the care that must be taken in ensuring that appropriate basis sets are used for phosphate calculations and the importance of distinguishing between gas-phase and solvent calculations. Investigations of the more detailed electronic structure of phosphate molecules can also be found, e.g. in 2008 Burrow *et al.* used B3LYP/6-31G(d) to investigate several molecules containing the P=O group, to see if these have the characteristics of the π^* resonances typically associated with multiple bonds; their results suggest that such resonances do not occur in trimethyl phosphate, and therefore that they are also unlikely occur in molecules such as DNA [85].

Beyond looking at the intrinsic properties of phosphate molecules, numerous studies can also be found that investigate how phosphates interact with other molecules, for example Bianciotto *et al.* have investigated the anionic zwitterion $CH_3O^+(H)PO_3^{2-}$ using B3LYP in both gas phase, with an implicit solvent model,

and even with the inclusion of explicit water molecules, the results supporting previous theories that these zwitterions exist as intermediates in the dissociative hydrolysis of the methyl phosphate anion $CH_3OPO_3H^-$ [86]. Other recent studies have used B3LYP/6-31G(d) to study the ion-radical mechanisms for enzymatic ATP synthesis [87], B3LYP/6-31++G(d,p) to study the affinity of Al(III) with phosphates [88], and B3LYP/6-311+G(df,p) to study proton transfers in the hydrolysis reactions of phosphate dianions and sulfate monoanions [89].

The take-home message of all these studies is that numerous reports can be found in the literature that investigate phosphate molecules using QM methods and that most of these studies are performed using B3LYP, demonstrating the impact this functional has had within the field of biochemistry.

Interaction between Phosphates and Proteins

An excellent example of the use of QM methods to investigate a phosphate binding protein can be found in a paper from 2004, where Dittrich *et al.* studied the enzyme F_1-adenosine triphosphatase (ATPase), the function of which is to catalyse the decomposition of ATP into ADP and a free phosphate anion, releasing energy in the process which can be harnessed to drive other chemical reactions in the cell [90], see Fig. 3.2. By applying B3LYP/6-31G in combination with molecular mechanics to the protein-ATP complex, they identified critical protein residues for the hydrolysis reaction as well as reported on how the mechanical motions of the protein are coupled to its catalytic activity, thereby aiding in the fundamental understanding of how the protein operates [90].

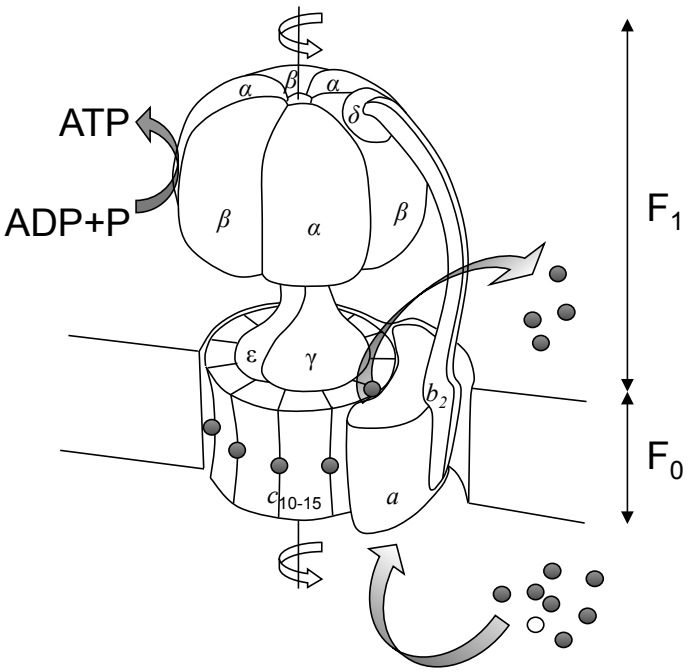

Figure 3.2: Schematic illustration of ATP synthase which is composed of two motor units, F_0 and F_1. The protein couples proton translocation across the inner membrane into ATP synthesis/hydrolysis potential. The water soluble F_1 motor unit (F_1-ATPase) contains both hydrolysis and synthesis catalysis sites, and isolated F_1-ATPase hydrolyses ATP to rotate the shaft against the stator ring. For recent reviews of ATP synthases, see refs. [91, 92]. Figure by supervisor Claus Hélix-Nielsen.

Many proteins in nature are phosphorylated, and interactions between such phosphorylated sites and other proteins are crucial to many cellular functions. To investigate such interactions, Laskin *et al.* published a paper in 2011 where they used B3LYP/6-31++G(d,p) to investigate the interaction between short phosphorylated peptides and model cationic ligands [93]. From the results they reported dissociation parameters between the two that corresponded well with experimental values, and also showed that phosphate abstraction from the peptides is mainly affected by the character of the phosphorylated side chain while hydrogen bonding in the peptide and properties of the ligand only play minor roles in determining the energetics and dynamics of the abstraction process [93].

Numerous other studies can be listed, recent examples being investigations of the molecular interaction between dGMP and the amino acid glycine (which is predominant in phosphate binding sites, see Chapter 2) using B3LYP/6+311+G(d,p) [94], or B3LYP/6-31(d) investigations into how the co-factor pyridoxal-5'-phosphate (PLP) behaves in the PLP-dependent enzyme dopa decarboxylase [95]. Typical for many

studies is that they use a combination of molecular mechanics and QM (described in more detail in Chapter 4) to overcome the computational restraints imposed by pure DFT-calculations, and typically studies are carried out using the B3LYP functional, with only a few exceptions.

Can Arsenates Replace Phosphates?

In 2011 Wolfe-Simon *et al.* published a study which attracted much media attention as it suggested that a bacterial strain, GFAJ-1, could sustain its growth by replacing phosphorus with arsenic [96]. This study gave rise to much controversy, and it was later experimentally disproved in that it was shown that GFAJ-1 is actually an arsenic-resistant and phosphorus-dependant organism, although high-resolution mass spectrometry did show the presence of a few abiotically formed arsenylated compounds, including C6 sugar arsenates, in the GJAF-1 extracts [97, 98]. In the end, arsenate esters are highly unstable in water compared to phosphate analogous, which is why phosphates predominate over arsenates in nature [15, 99].

Although disproved, the 2011 study by Wolfe-Simon spiked renewed interest in using QM methods to investigate differences between phosphate and arsenate compounds, e.g. Mládek *et al.* used high-level dispersion-corrected DFT calculations to analyze the electronic structure of the arsenate analogue of the DNA backbone, finding that arsenates may serve as a potential substitute for phosphate [100]. In another study, Jissi *et al.* showed using DFT calculations (with B3LYP and a dispersion-corrected functional M06-2X) that from a structural, electronic and kinetic point of view, substitutions of phosphates with arsenates are possible [101].

3.3 Study Motivation & Paper II

The B3LYP functional is beloved among many biochemists, and despite the fact that this functional is more than a decade old, it is still widely used today. As discussed in Sec. 3.1.3, improved functionals are continuously being developed, with much focus being dedicated to promoting, among other things, more accurate long-range and dispersion behaviour of the functionals. Specifically, it is a known fact that classic functionals such as B3LYP perform poorly for describing highly anionic systems; e.g. in 2010 F. Jensen published a study where he showed that many of such "standard" XC functionals describe anions as only having a fraction of the extra electron bound, and that this error extends to molecular anions as well as intermolecular systems where it leads to unrealistic electron transfer [102]. It is compelling to see if such errors are also present in B3LYP calculations of anionic phosphates, which is a primary motivation for the study presented in **Paper II**.

A variety of different basis sets are used in the scientific literature for modelling phosphate compounds, and especially in the case of large systems and older studies, limitations in computational resources mean that small basis sets such as 6-31G have been used to study the phosphates. These studies have shown remarkable results [90], indicating that in some systems this level of description is adequate, yet it is tempting to investigate how differences in basis sets actually influence the molecular properties of phosphate molecules, especially considering their anionic nature which would generally suggest that less restrictive (larger) basis sets might be required.

A final inquiry which motivated the study presented in **Paper II** is based on the fact that many studies are carried out in the gas phase. Phosphates in nature can however be found in a wide range of chemical environments, ranging from aqueous solution to deep within a protein core. Investigating all such possible conditions is intractable, but by assigning an average dielectric constant to the environment through an implicit solvent model, the polarizability of the local surroundings can be systematically varied, thereby making it possible to get a rough idea of how the phosphate molecules behave in different environments.

3.4 Conclusions & Outlook for Paper II

The oxyacids H_3PO_4, H_3AsO_4 and H_2SO_4 and their corresponding eight anions were investigated using a range of model chemistries, including WFT methods HF and MP2, standard DFT functionals B3LYP and BLYP, long-range corrected functionals LC-BLYP [103], CAM-B3LYP [104] and wB97XD [76], as well as the semi-emperical method PM6 [105] and the composite method CBS-QB3 [106]. Calculations were carried out both in gas-phase and with an implicit solvent model where the dielectric constant of the environment was systematically varied, and novel values for molecular properties such as orbital energies, electron affinity, bond lengths, charge distribution and molecular volumes are reported. These properties are derived in a more rigorous manner than has previously been described in the scientific literature, thus giving insight into the differences between e.g. arsenates and phosphates.

To consistently describe the molecules across all the tested theories, it was found that a minimum basis set should include both diffusive functions as well as polarized basis functions, such as e.g. in 6-311++G(d,p), even though some individual properties such as bond lengths could be decently described by only including extra polarized basis functions. Calculations that were ill-defined in vacuum were found to be stabilized already at low dielectric constants of around $\sim$5, corresponding to very hydrophobic protein pockets [107]; however, the local environment in general was found to have only little effect on the overall geometric structure or electron density distribution in the molecules.

Finally, from calculations of the investigated molecules and an ammonium cation NH_4^+, an erroneous description of electron transfer was observed at intermediary and long-range distances for B3LYP. This is especially pronounced in gas-phase calculations and solvent calculations with low dielectric constants. It is unknown whether such effects would be observed e.g. for calculations deep within protein binding pockets, but it highlights that extreme care must be taken when interpreting results of phosphate molecules obtained with B3LYP, and it urges a shift towards using long-range corrected functionals for describing phoshate-containing systems.

CHAPTER 4

Molecular Dynamics

In principle, all dynamic properties of a given molecular system could be derived from solving the Schrödinger equation. Unfortunately, many problems are too large to tackle using QM methods alone, as such calculations are often restricted to below 100 atoms depending on the procedure in question [27]. Force field methods, otherwise known under the general term *molecular mechanics* (MM), ignore the electronic structure calculations and instead use *force fields* to calculate the energy of the systems based only nuclear coordinates [108]. In a sense, they bypass the need for electronic structure calculations by parameterising the electronic energy as a function of nuclear coordinates only – the parameters in turn are found by fitting to experimental data and to results obtained in higher level electronic structure calculations. A typical application of MM can be found in *molecular dynamics* (MD), where force fields are used to calculate the forces on individual nuclei in a system, and with a suitable integrator the dynamics of the particles can then be predicted as a function of time.

In this chapter, the theory underlying MD as used during this project is outlined, and it is discussed how MD methods can be combined with QM in hybrid calculations (QM/MM) suitable for simulating interactions between phosphates and proteins. Ideally, trajectories obtained from MD simulations are "converged", such that they sufficiently sample all conformations in the canonical ensemble of the molecule being investigated. Advanced sampling techniques are also discussed in this chapter, which leads to the motivation of the study in **Paper III**, where the interaction between a highly disordered soluble P-loop peptide and inorganic phosphate is investigated using QM/MM.

4.1 The AMBER Force Field

"Amber", in the context of this thesis, can refer to two things; a set of MM force fields specifically designed for simulating biomolecules (proteins, DNA and RNA), and a package of molecular simulation programs, the latter of which is henceforth referred to as Amber14 since version 14 of the software package was used throughout this project. The amber force fields are some of the most widely applied parameter sets used for simulating biomolecules and as such they have been extensively tested, evaluated and revised during the last two decades [109]. The basis for the force fields is the following potential energy function [110]:

$$\begin{aligned} V_{\text{pot}} &= \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \\ &\quad + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right] + \sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_0 R_{ij}} \end{aligned} \tag{4.1}$$

where for bond and angle terms K_r and K_θ are the force constants, r is the bond length and θ is the bond angle, with the subscript eq denoting equilibrium values. For the dihedral term, $V_n/2$ is an amplitude, n is periodicity, and ϕ is the torsional angle at displacement γ. The final terms describe non-bonded interactions between particles i and j at a distance R_{ij}, namely in the form of a Lennard-Jones potential based on parameters A_{ij} and B_{ij} for repulsive and attractive interactions, respectively, and finally the last term which is the Coulomb electrostatic interactions with q denoting particle charge and ϵ_0 being the vacuum permittivity.

The latest instalments of the amber force fields include the sets ff14SB and ff14ipq, both of which contain all required parameters for simulating proteins composed of natural amino acids, as well as DNA and RNA [111].

4.2 Equations of Motion

To gain a basic understanding of MD it is intuitive first to consider a micro-canonical ensemble where the number of particles N, system volume V, and energy E is kept constant (NVE ensemble). A simulation in such an ensemble corresponds to an adiabatic process without heat exchange, i.e. during an MD simulation the potential and kinetic energies are directly exchanged while the total energy is conserved. Assuming that the particles in the system can be treated classically and thus that they can be addressed using Newtonian mechanics, the force f_i experienced by particle i in the micro-canonical ensemble can be written [29]:

$$f_i = m_i \frac{d^2 r_i}{dt^2} = -\nabla V_{\text{pot}} \tag{4.2}$$

where m_i and r_i are the mass and position of the particle i, respectively, and V_{pot} is the energy as defined by the force field in Eq. (4.1). Given Eq. (4.2), the next step is to integrate the equations of motion in order to obtain the time evolution of the particles. The criteria for good integration algorithms are that they are computationally efficient, permit the use of relatively large time steps, show good conservation of energy and that they are reversible in time [29]. A classical integration scheme is the Verlet algorithm, where a Taylor expansion is used for the position r_i of a particle i after (and before) a time step Δt:

$$r_i(t + \Delta t) = r_i(t) + v_i(t)\Delta t + \frac{f_i(t)}{2m_i}\Delta t^2 + \frac{\Delta t^3}{3!}\Delta t^2 \dddot{r}_i + \mathcal{O}(\Delta t^4) \tag{4.3}$$

$$r_i(t - \Delta t) = r_i(t) - v_i(t)\Delta t + \frac{f_i(t)}{2m_i}\Delta t^2 - \frac{\Delta t^3}{3!}\Delta t^2 \dddot{r}_i + \mathcal{O}(\Delta t^4) \tag{4.4}$$

where $v_i(t)$ is the velocity of particle i and $\mathcal{O}(\Delta t^4)$ describes an error on the order of Δt^4. Summing Eqs. (4.3 - 4.4) gives the Verlet algorithm [112]:

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + \frac{f_i(t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{4.5}$$

which can be used to predict new positions for the particles as a function of time with an error on the order of Δt^4. Several adaptations and improvements of the Verlet algorithm as well as a multitude of higher-order schemes can be found in the scientific literature, see ref. [29]. It is often more convenient to perform simulations in other ensembles, e.g. NVT or NPT, which can be achieved by using an approach based on the reformulation of the Lagrangian equations of motion of the system [29].

4.3 Combining MD with QM in QM/MM

In the hybrid *QM/MM* approach, a small region of a large system is described with a QM method, while the rest of the system is described using MM. The approach is often applied to systems where parts of the system are not well described using classical mechanics, e.g. in the case of highly polarizable molecular moieties or for modelling of chemical reactions. The effective Hamiltonian for a QM/MM system consists of a term for the QM region, a term for the MM region, and a term describing the interaction between the two regions. The effective energy of the system can thus be written [111]:

$$E_{\text{eff}} = \langle \Psi | H_{\text{QM}} + H_{\text{QM/MM}} | \Psi \rangle + E_{\text{MM}} \tag{4.6}$$

where the Hamiltonian H_{QM} is evaluated according to the chosen QM method and the energy E_{MM} is calculated as in a regular MM calculation. The interaction term $H_{\text{QM/MM}}$ is slightly more complicated, especially if covalent bonds between the MM and QM region are involved. Typically an electrostatic embedding scheme is employed where interactions between MM nuclei charges and the electrons of the QM system, as well as interactions between the MM nuclei and the QM nuclei, are explicitly considered, effectively resulting in a scheme where the MM region can polarize the electron density of the QM region. In cases where such polarization of the QM region is not important, it is sufficient to use a mechanical embedding scheme where interactions between the QM and MM regions are simply treated in the same classical approximation that is used for MM; see ref. [111] for more details on the QM/MM approaches in Amber14.

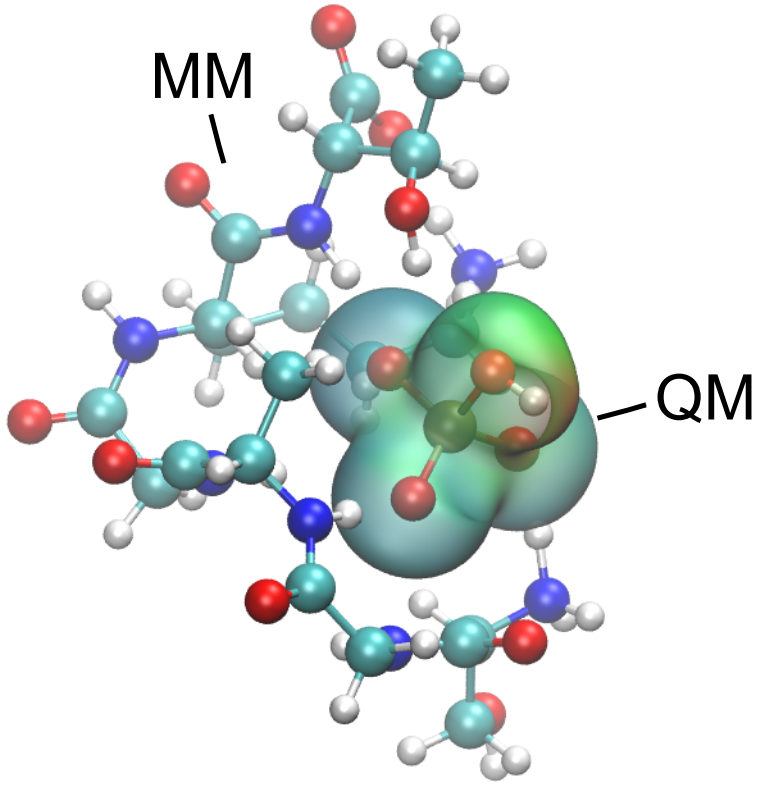

Figure 4.1: Illustration of a case where QM is used to describe an inorganic phosphate molecule, interacting with a SGAGKT peptide described by MM (see study motivation in Sec. 4.5 for details about peptide).

It was found in the study presented in Chapter 3 that both polarized and diffusive basis functions are essential for adequately describing phosphate molecules using electronic structure methods. These findings suggest that phosphates are presumably best described using a QM/MM approach with an electronic embedding scheme rather than a mechanical embedding scheme, since in such a scheme the electron density of the phosphate moieties can be polarized by the MM region.

4.4 Enhanced Sampling Algorithms

The application of MD simulations is often limited because it is difficult to obtain trajectories that are ”converged”, that is, trajectories which accurately sample the conformational space of the system of interest and not only a small region of the conformational space. The difficulty arises because energy landscapes are often

rough, especially for biomolecules, meaning that the simulation can be ”trapped” in local minima from which it does not escape during the simulation [113].

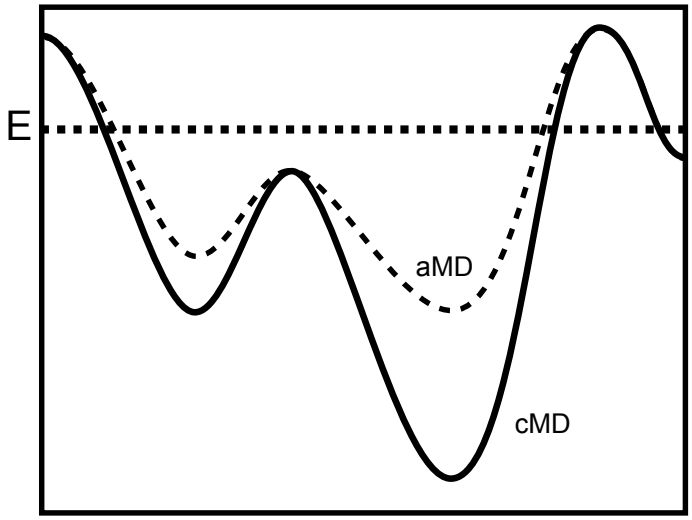

Figure 4.2: Schematic showing how accelerated MD (aMD) enhances sampling by applying a potential bias to the energy landscape of conventional MD (cMD).

During the last few decades several methods have been proposed for addressing the sampling limitation of MD, examples including different variations of replica exchange MD (REMD) [114, 115], self-guided Langevin dynamics (SGLD) [116], and accelerated MD (aMD) [117]; see ref. [113] for a recent review of different enhanced sampling methods. These methods all work to improve conformational sampling; however, they also all have disadvantages, e.g. in the form of being computationally prohibitive, unable to guarantee convergence, being dependent on prior knowledge about the system, or requiring complicated post-processing analysis [111].

Taking aMD as an example, the method is based on applying a bias potential to the simulation which effectively lowers the energy barriers, see Fig. 4.2, thereby allowing sampling to continue faster because the simulation is less likely to be caught in local minima. It represents a method where only a single copy of the system needs to be simulated (unlike REMD which requires multiple replicas to be simulated at once), and therefore it is a comparably efficient approach. The applied bias in aMD is typically chosen to be proportional to the height of the energy barriers such that the overall energy landscape is conserved – the canonical ensemble can thereby, in theory, subsequently be recovered by appropriate reweighting of the simulated conformational ensemble [118].

4.5 Study Motivation & Paper III

In the literature review presented in Sec. 3.2 multiple studies are listed where QM/MM is used to investigate interactions between biomolecules and phosphorus compounds. Such studies are of particular interest because they provide information not only about a single static binding conformation between e.g. a protein and the phosphate ligand, but also about the dynamics of these interactions. The difficulty of ensuring that the full conformational ensemble has been sampled during an MD simulation however increases with the size of the system being investigated. Therefore, from a computational point of view, it is desirable to initiate investigations of the interaction between phosphate and biomolecules by looking at smaller prototype systems for which it is easier to achieve converged ensembles. A system that lends itself well to such an investigation is the interaction between the P-loop sequence SGAGKT and inorganic phosphate, noting that it was shown in 2012 by Bianchi *et al.* that this particular hexapeptide is capable of binding HPO_4^{2-} selectively in solution [119].

The SGAGKT-HPO_4^{2-} system represents an interesting study case because the binding between the peptide and HPO_4^{2-} is likely to be highly reversible (because of the short size of the peptide), while still being selective (the peptide was found experimentally not to bind e.g. SO_4^{2-} or $H_2PO_4^-$ [119]), and both of these properties are desirable for a bioscavenger. Due to the small size of the system, it might additionally be possible to sample close to the canonical ensemble using MD and moderate computing resources, at least when using an enhanced sampling method such as aMD. As discussed in Sec. 4.3, it is arguably necessary to use QM to describe the anionic phosphate molecule HPO_4^{2-}. To keep the simulations tractable, the semi-empirical QM method PM6 was chosen to describe the phosphate molecules, noting that PM6 was shown in the work performed in **Paper II** to represent inorganic phosphates satisfyingly when compared to high-level methods such as MP2/6-311++G(d,p) and wB97XD/6-311++G(d,p).

4.6 Conclusions & Outlook for Paper III

The interaction between the peptide SGAGKT and inorganic phosphate was investigated using a combination of QM/MM and aMD. By using tools from information theory along with statistical methods, it was confirmed that the conformational space of the hexapeptide by itself can be fully sampled using such an approach and close-to-converged ensembles can be achieved for the peptide-phosphate system. The conformational ensemble of the peptide was found to be significantly stabilized by phosphate binding, and contrary to popular belief, the simulations suggest that binding of phosphate does not take place inside a single tightly knit and stable P-loop nest conformation, but rather that binding is supported by multiple binding modes.

Taking a critical standpoint, peptide-based bioscavengers inherently pose many difficulties: specifically, it may prove to be difficult to control the binding mechanism, which is influenced by both enthalpic interactions as well as entropy, and for certain applications the selectivity and binding affinity of the peptides may not be sufficient. In Chapter 2 it was shown that the protein secondary structure, specifically the second shell of the binding sites, is highly conserved in natural phosphate binding proteins. Using full-sized proteins may thus be required for attaining sufficient phosphate selectivity and affinity, and the larger structural domain, in addition, opens up for controlling the binding mechanism by targeting protein regions further away from the binding site. See further discussions on this subject in Chapter 6.

Despite the potential disadvantages, small peptides capable of binding phosphate selectively represent an attractive starting point for the development of novel bioscavengers. The advantages of using such small biomolecules are that they might be easier to produce and use in industrial applications, and the complex binding mechanisms may allow for delicate control of the binding process. The studies performed here suggest that these systems can adequately be described using QM/MM and from the obtained information about the peptide-phosphate binding modes, it might be possible to deduce how to optimize design of such peptide bioscavengers.

CHAPTER 5

Computational Fluid Dynamics

Using specialized software and hardware platforms, MD can now be used to simulate millions of particles on time scales of multiple microseconds of physical time per day [120]. Even though that is an impressive feat, it is nowhere near the requirements of most full-scale technology applications, which often involve length scales on the order of meters, and time-scales of minutes, hours or even days. At large scales there is however also much to be gained from computational investigations, and during the last two decades especially Computational Fluid Dynamics (CFD) has been increasingly applied to a wide array of complicated systems, ranging from describing blood flow in arteries [121] to optimization of ship and automobile designs [122] and weather forecasting [123].

In this chapter the basic theory underlying CFD is presented and it is discussed how osmotic membrane technologies might tie together with the implementation of a phosphorus recovery technology, i.e. a technology based on combining osmotic membrane technology with molecular bioscavengers. Specifically, the osmotically-driven process of forward osmosis (FO) is considered, which in recent years has emerged as a popular alternative to conventional pressure-driven processes such as reverse osmosis (RO). Finally, the study of **Paper IV** is motivated, where CFD is used to investigate and optimize flow in FO module prototypes, and the implications of the study are discussed.

5.1 Computational Fluid Dynamics

CFD is a branch of fluid mechanics where numerical methods and algorithms are used to solve and analyse fluid flows as described by a set of governing partial differential equations. These governing equations are typically based on one or more of three fundamental principles, namely, conservation of mass, momentum and energy – by applying these principles to a suitable description of the flow, the mathematical form of the governing equations can be derived [124]. Looking at the equations for conservation of mass and momentum, as would be relevant for an incompressible and isothermal fluid, these equations can be written as follows when using a fixed Finite Volume Method (FVM) representation of the flow of a Newtonian fluid [124]:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{U}) = 0 \tag{5.1}$$

$$\frac{\partial \rho \mathbf{U}}{\partial t} + \nabla \cdot (\rho \mathbf{U}\mathbf{U}) = -\nabla p + \nabla \cdot \left[\mu \left(\nabla \mathbf{U} + (\nabla \mathbf{U})^{\mathrm{T}} \right) \right] + \rho \mathbf{f} \tag{5.2}$$

with ρ being the fluid density, $\mathbf{U}$ the fluid velocity vector, p the pressure, μ the fluid viscosity, and $\mathbf{f}$ external body force per unit mass. Eq. (5.1) and Eq. (5.2) describe mass and momentum conservation, respectively, and by solving these partial differential equations, information can be obtained about the fluid flow in the form of pressure and velocity fields as a function of time.

When it comes to the practical part of solving the governing equations, the general approach is first to discretize the equations, thereby transforming the partial differential equations into a set of algebraic equations, which can then subsequently be solved using a variety of numerical methods [125]. The discretization procedure inherently involves not only a discretization of the equations but also of the solution domain, such that the latter is decomposed into a finite number of smaller regions, or *cells*, at which the algebraic equations can be evaluated [124].

5.2 Osmotic Membrane Filtration

In conventional membrane technologies a hydraulic pressure is applied to drive separation of solute from solvent across a semi-permeable membrane. Such technologies are roughly separated into different categories depending on the size of the solute, namely micro-filtration (0.1-5 μm), ultrafiltration (1-100 nm), nanofiltration (0.5-10 nm) and reverse osmosis (<0.5 nm) [126].

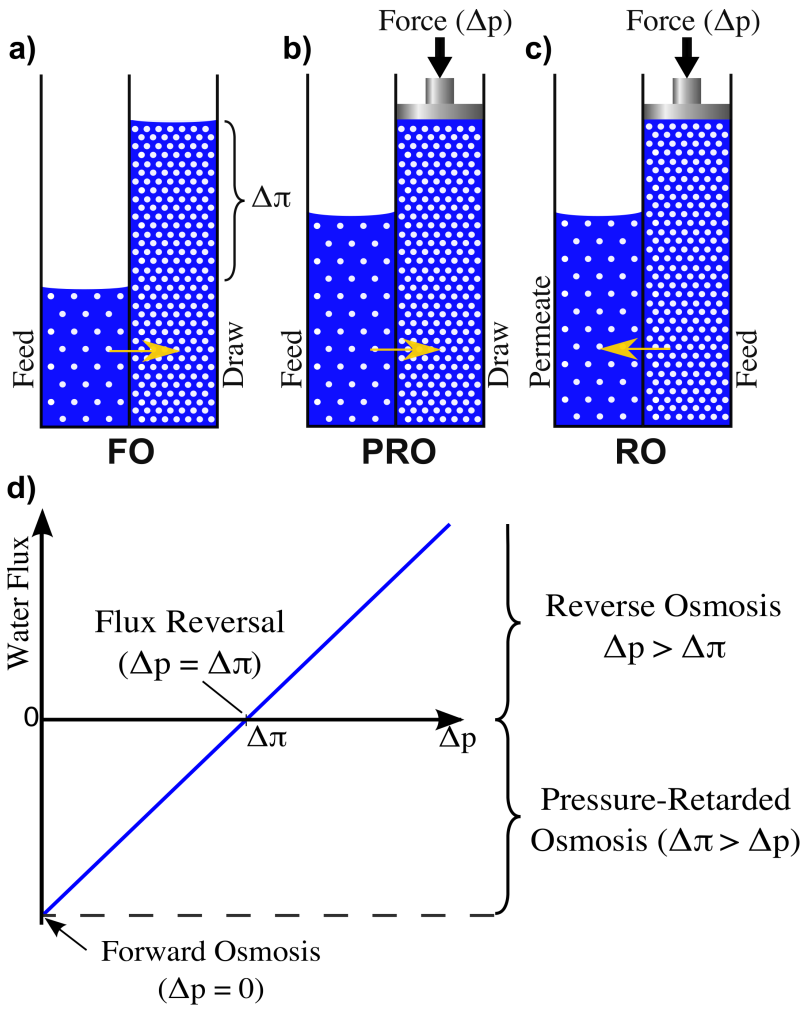

Figure 5.1: Sketches of FO, PRO and RO. **a)** an osmotic pressure difference drives flow from a feed solution into a concentrated *draw* solution. **b)** an osmotic gradient drives flow from a feed solution into a pressurized draw solution, thus enabling the conversion of osmotic pressure into hydraulic pressure. **c)** a hydraulic pressure is used to force water from a feed solution through a membrane. **d)** illustration of the FO, PRO and RO regimes. In FO the applied hydraulic pressure Δp is zero, in PRO Δp is lower than the osmotic pressure difference $\Delta\pi$, and in RO $\Delta p > \Delta\pi$.

Focus in this work is restricted to osmotic membrane separation processes. Conventionally osmosis can be defined as *the net movement of water across a selective and permeable membrane driven by a difference in osmotic pressure across the membrane*, noting that osmotic pressure is a quantity that increases with solute concentration [126]. Three categories of osmotic separation processes are typically defined, namely, RO where a hydraulic pressure is applied to overcome the osmotic pressure, FO where an osmotic gradient drives flow across the membrane, and Pressure Retarded Osmosis (PRO) where the osmotic gradient drives flow into a pressurized chamber, see Fig. 5.1. RO is by far the most widely applied osmotic membrane technology in the industry; however, FO has received increasing attention in recent years because it theoretically does not require the application of a hydraulic pressure and therefore potentially is more cost-effective [32].

One of the major complications in all the osmotic membrane separation processes is the phenomenon known as *concentration polarization* (CP), which is a general term used to refer to accumulation or dilution of solute concentration near the membrane that leads to decreased membrane performance. The situation is illustrated in Fig. 5.2 for FO, where both external concentration polarization (ECP) and internal concentration polarization (ICP) lead to a reduced effective osmotic pressure across an asymmetric membrane. A central goal of designing industrial modules for osmotic membrane technologies is thus to reduce these detrimental effects, e.g. by modulating fluid flow so as minimize CP. This is a job well-suited for CFD simulations as they allow the user to rapidly prototype different module designs and investigate complex and coupled mass-transfer effects near the membrane in high detail.

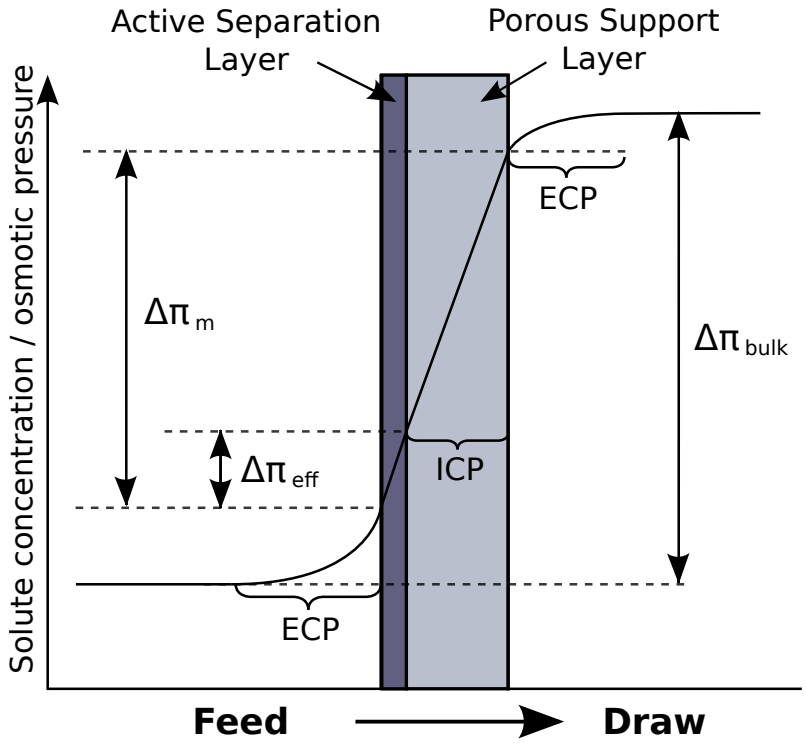

Figure 5.2: Sketch of external and internal concentration polarization (ECP/ICP) in FO with an asymmetric membrane. $\Delta\pi_{\text{bulk}}$ is the osmotic pressure difference between the bulk of the feed and draw solutions, $\Delta\pi_m$ is the difference across the membrane and $\Delta\pi_{\text{eff}}$ is the effective osmotic pressure difference which drives the separation process over the active layer of the membrane.

5.2.1 CFD & Osmotic Membrane Separation

In 2011 our group formulated and demonstrated the use of a CFD model capable of simulating both RO and FO membrane processes [127]. The model was developed in the open source framework OpenFOAM® and is based on a weakly compressible formulation of the governing equations, meaning that an isothermal Newtonian fluid is assumed where the density is dependent on the solute concentration only and not the pressure. The model implements the membrane as a 2D plane without any thickness, and mass transfer across the membrane is specified at each point of the membrane using boundary conditions (BCs) derived from analytical expressions, e.g. the velocity BC in FO is based on the original analytical work performed by Lee *et al.* [128] and Loeb *et al.* in [129]:

$$\mathbf{J}_w = \frac{1}{K} \ln \frac{B + A\pi_{\text{D},m}}{B + |\mathbf{J}_w| + A\pi_{\text{F},m}} \, \mathbf{n}_{\text{D}} \qquad \text{AL - FS orientation} \qquad (5.3)$$

where $\mathbf{J}_w$ is the water permeation flux, K is a membrane mass transfer coefficient, A is the pure water permeability, B is the solute permeability, π is osmotic pressure, $\mathbf{n}$ is the membrane surface normal unit vector, AL-FS refers to the configuration where the active layer of an asymmetric membrane faces the feed solution, and the subscripts D, F refer to the draw and feed side of the membrane, respectively, while the subscript m specifies that the value is at the membrane surface.

Beyond the governing equations for mass and momentum conservation, the CFD model uses the diffusion-convection equation to describe the solute mass fraction m_A of a single solute [130]. The convective and diffusive solute fluxes in and out of the membrane plane must be balanced with the solute flux through the membrane, which leads to another key component of the CFD model, namely the BC for the solute mass fraction on the membrane:

$$\underbrace{-\rho_m D_{\mathrm{AB}} \frac{\partial m_A}{\partial n_{\mathrm{D}}} \mathbf{n}_{\mathrm{D}}}_{\text{diffusive}} + \underbrace{\rho_m m_{A,m} \mathbf{J}_w}_{\text{convective}} = \mathbf{J}_s \tag{5.4}$$

where D_{AB} is the solute diffusion coefficient and $\mathbf{J}_s$ is the solute flux through the membrane. Having defined the membrane BCs and the weakly compressible formulation of the governing equations, the OpenFOAM model as presented in ref. [127] was implemented using a PISO algorithm for treating the inter-equation pressure-velocity coupling [131], resulting in a transient CFD model capable of simulating pressure, velocity and solute mass fraction as a function of time in a given system geometry.

5.3 Study Motivation & Paper IV

Taking a few steps back and considering the overall topic of this thesis, the use of bioscavengers as a basis for a resource recovery technology is not straightforward, and especially the challenge of how to optimally integrate scavengers into large-scale industrial applications remain an open question. The specifics of the integration depends on the bioscavenger, e.g. if we consider the situation where a bioscavenger has been designed to bind phosphate selectively and the release mechanism is designed such that the bound phosphate is released upon being targeted with a laser impulse of a specific frequency – in that case it makes no sense to integrate the scavengers deep within a thick porous structure, since then it might not be possible to engage the release mechanism of the buried scavengers.

Besides the specifics of the bioscavengers, also the details of the phosphorus input stream must be considered; i.e. if it is in the form of solid waste, sludge, or different kind of waste water. Given that we are considering a technology based on protein bioscavengers, the ideal input stream is a pre-treated aqueous solution with a high concentration of the phosphorus compounds we are attempting to recover. One way to obtain such a concentrated solution is to use low-cost FO in a pre-treatment process, that is, to use a concentrated draw solution to up-concentrate a feed solution containing phosphorus, which can then subsequently be passed to the operational unit containing the biomimetic phosphorus scavengers. To assist in the development of such FO modules, in the study presented in **Paper IV** we sought to improve our previous CFD model and demonstrate its use by optimizing common lab-scale FO membrane module characteristics.

Our original OpenFOAM model was initially verified by comparison to analytical results for 2D chambers [127] and later also validated against FO experiments performed in complex 3D lab-scale modules [132]. In summary, the model was shown to accurately be able to predict the performance of different membrane module geometries; however, the benchmarking of a single 3D module on a personal computer took on the order of 1-2 weeks of simulation time. Such long calculation times are prohibitive in cases where hundreds or even thousands of different module geometries need to be tested in order to find the optimal configuration. This prompted us to attempt to optimize the numerical implementation of the model, with the goal of being able to benchmark a single geometry on a single CPU in no more than 24h, meaning that numerous geometries could be benchmarked in parallel every day.

5.4 Conclusions & Outlook for Paper IV

In this work we implemented a new CFD model based on our previous work [127], but with several numerical improvements: a complete re-implementation for use with the latest OpenFOAM v.3.0 (the previous model was implemented using v.1.7), a steady-state solver based on the SIMPLE algorithm [133], as well as empirical optimization of the choices of discretization and solution schemes for the governing equations. The result is a model capable of benchmarking FO/RO membrane modules/chambers containing solution domains with millions of cells within 24h using a single CPU.

The model was used to benchmark several hundred membrane modules by varying the properties of several parameters, e.g. the dimensions of the chambers, the number of inlets, the inlet angles, the cross-flow velocity, and even several different flow-promoting spacer geometries. The model effectively lets the user estimate the spatial distribution of water and solute flux across the membrane and thus help to identify ”dead areas” in the modules where severe CP effects dominate and mass-transfer across the membrane is inefficient.

The numerical implementation of CFD models capable of describing osmotic membrane processes is not straightforward, and therefore an open source platform may facilitate an increased use of CFD within the field of membrane module engineering in the future. The models developed in the study were therefore made available free-of-charge online along with **Paper IV**, thereby providing such an open source platform for other researchers to extend upon.

The paper demonstrates that it is possible to rapidly optimize FO modules using CFD. Besides being capable of describing osmotic membrane processes, the open source model with a solver based on a weakly compressible formulation of the governing equations may, however, also provide an excellent starting point for future CFD models capable of optimizing the design of a biomimetic phosphorus recovery technology. As an example, if scavengers were incorporated into a porous matrix, the CFD model could be modified to investigate flow within that porous medium in order to benchmark how different porosities, tortuosities, etc. perform in terms of maximizing the available surface area and optimizing the distribution of feed solution in the porous matrix simultaneously.

CHAPTER 6

Discussion & Future Perspectives

A wide range of length scales have been explored using computational techniques in this thesis, ranging from atomic quantum chemical descriptions of phosphate anions to macroscopic scale CFD simulations of osmotic membrane separation modules. From the performed studies, as presented in **Papers I - IV**, an understanding has been acquired about how proteins in nature bind phosphates, how phosphate molecules and interactions with them can be theoretically described, and how potential large-scale applications can be effectively optimized. All the aims listed in Sec. 1.2 have thus been accomplished successfully.

The very fundamental information obtained in this work does not directly call for the implementation of a specific technology; however, the attained knowledge is of immense interest for any future development and experimental implementation of biomimetic phosphorus recovery technologies. To support this claim of importance, this final chapter is dedicated to discussing future perspectives of biomimetic recovery technologies, both in terms of developing biomimetic phosphate scavengers and industrial-scale application modules.

6.1 Biomimetic Phosphorus Scavengers

The design of bioscavengers can take at least two overall paths, namely, the design of short peptide-based scavengers and the design of larger protein-based scavengers. The peptide analysed in **Paper III** was originally proposed by Bianchi *et al.*, who chose to synthesize and analyse that particular peptide based on the P-loop consensus sequence [119]. The peptide proposed by Bianchi *et al.* is however but one possible case and countless other peptide sequences might well be preferable both in terms of selectivity, affinity, and/or stability. As briefly touched upon in Sec. 4.6, using small peptides as opposed to larger proteins offer both advantages and disadvantages, as discussed a bit more at length in the following:

Synthesis Short peptides can readily be synthesized, making the iterative design process of the optimal peptide much faster. Several commercial companies now offer the purchase of 10mg peptide samples from $\sim$ \$3/amino acid, a price that can be significantly reduced when up-scaled in the industry [134, 135]. In comparison, expression and purification of new recombinant proteins can be a laborious and potentially error-prone task which may take weeks and cost hundreds or even thousands of dollars if purchased from a supplier.

Unusual Amino Acids An advantage of synthesizing peptides, in contrast to recombinant expression of peptides or proteins, is that special amino acids can readily be incorporated into the sequence; this opens up for creating cyclic peptides or peptides conjugated to other compounds, which might significantly improve stability and functionality of the scavengers in ways not possible with natural amino acids.

Molecule Size Because peptides are typically smaller than proteins they may be used at higher densities and it might be possible to implement them in more restricted environments, e.g. in porous support structures. On the other hand, proteins may provide more opportunities for modifications: i.e. whereas it is unlikely that a peptide will retain its affinity for phosphate upon chemical modification, since that may directly affect the binding interactions, similar modification far from the active site in a protein may facilitate new functionality without disturbing the native binding site.

Secondary & Tertiary Structure The lack of global structure in peptides means that they have a lower potential for selectivity and affinity towards phosphate than proteins. Simply, a peptide in solution is unlikely to be able to distinguish

HPO_4^{2-} from $HAsO_4^{2-}$, nor bind phosphate as strongly as the PBPs introduced in Sec. 1.1. Part of the reason for the higher affinity and selectivity of the proteins is that the protein structures are more rigid, a fact which may also be an advantage when it comes to tweaking the design of existing proteins: i.e. compared to a single point mutation in a rigid protein structure, it is harder to predict the effect of such a mutation on the conformational space of a short peptide.

Binding Mechanism & Applicability The mechanism suggested by MD for the interaction between SGAGKT and HPO_4^{2-} may prove to be both a blessing and a curse in terms of using peptides as industrial bioscavengers: it may be an efficient "handle" for controlling the binding/release mechanism, e.g. by raising/lowering the temperature, or it may effectively mean that the applicability of the peptides is restricted to a limited number of environments, with small perturbations disturbing the binding mechanism to such a degree that it becomes ineffective.

The above list includes a few of the circumstances which have to be considered during the design of a phosphorus bioscavenger. Although the binding mechanisms might be slightly different, techniques such as MD and QM/MM can, to a certain extent, be used to facilitate the design process of both peptides and proteins, e.g. by screening libraries of different scavengers in various environments for binding affinity and selectivity, thereby potentially reducing the need for extensive laborious experimental work.

Although there are some advantages of using peptides as phosphate bioscavengers, I believe these are outweighed by the potential difficulties and limitations of such an approach. Specifically, the complex binding mechanism between short peptides and phosphate suggests that the design process becomes an exercise in screening a large number different peptide sequences, chemical modifications, and chemical environments, constantly aiming to keep selectivity and affinity high, as well as for each design trying to achieve some kind of binding/release mechanism, e.g. in the form of changing the conformational space of the peptide by lowering/raising the temperature, or by attaching the peptides to a more complex molecule that influences its conformational space and can be controlled with outside stimuli; e.g. a light-sensitive protein.

Instead of peptide-based phosphate bioscavengers, it is more conducive to pursue the design protein-based bioscavengers through the re-design of natural phosphate binding proteins. Perhaps the most compelling starting point for such research are the PBPs introduced in Sec. 1.1, given that these are highly conserved in nature, have exceptionally high affinity and selectivity, and last but not least have

an intuitive ”pac-man”-like binding mechanism. Especially the binding mechanism of the PBPs may be contributive towards the design of a release mechanism, e.g. through the incorporation of magnetic nanoparticles or light-sensitive molecules on the ”backside” of the protein, which can then used for inducing the opening/closing of the pac-man structures, so to speak.

6.1.1 Stability Considerations for Bioscavengers

A quality of the biomolecules that deserves special attention during the design process is stability. Indeed, the advantages of applying biomolecules in industrial applications, namely high specificity, high activity under mild conditions, high turnover number and biodegradability, are notoriously offset by the disadvantage of intrinsic instability [23, 136]. Unlike proteins, which often depend on an intact tertiary structure to function, smaller peptides cannot be denatured; i.e. peptides can only be irreversibly damaged by covalent modification or breakage of peptide bonds. On the other hand, the stability of a protein can be many-fold enhanced by its rigid structure, which reduces the propensity for irreversible chemical changes [137, 138]. In the end, considerations of protein/peptide stability are entirely dependent on the given application in which they are used, e.g. temperatures, pH, if proteolytic activity has to be considered etc., and therefore only some general considerations are made in this subsection.

Looking at the degradation pathways for peptides, they typically include reactions such as hydrolysis, deamination, oxidation and diketopiperazine and pyroglutamic acid formation [135, 139, 140]. The propensity for these different pathways depends largely on the amino acid composition of the peptide, such that half-life's for peptides can range from hours to weeks at room temperature, depending on their composition [135, 139]. Peptides can be stabilized using a multitude of different chemical modifications, e.g. glycosylation, deamination, N-acetylation, N-formylation, amidation of C-terminus, by incorporation of unusual amino acids, or by using D-amino acids instead of the natural L-residues [141]. Also, cross-linking or cyclization of peptides can work to improve the stability of peptides [142], an example being Bogdanowich-Knipp *et al.* who showed a 30-fold increase in stability of a cyclic peptide as compared to its linear counterpart [140].

Larger proteins are inherently susceptible to the same degradation pathways as peptides, with the added complication of secondary and tertiary protein structure. According to the widely accepted Lumry-Eyring description, inactivation of proteins is a two-stage phenomenon, including first a reversible unfolding of the original

protein followed by kinetically irreversible steps [136, 138]:

$$\mathrm{N} \overset{\mathrm{K}}{\rightleftharpoons} \mathrm{D} \xrightarrow{\mathrm{k}} I \tag{6.1}$$

where N is the active protein, D is the reversibly denatured and inactive protein, and I is the irreversibly inactivated form, with K and k being the equilibrium and rate constants, respectively. Unlike peptides where stability is mainly determined by amino acid composition, the secondary and tertiary structure of proteins also play an immense role in determining stability, which can be observed when comparing mesophilic (prefers moderate temperatures) and thermophilic (prefers high temperature) organisms, where it is generally observed that proteins of the latter have more internal interactions (hydrogen bonds, electrostatic interactions, hydrophobic interactions, disulfide bonds, metal binding), and a superior conformational structure in terms of being more rigid, having higher packing efficiency, reduced entropy of unfolding, etc. [143–145].

Similar to the peptides, protein stability can be increased in many different ways, e.g. by various chemical modifications [24, 136], or by rational protein engineering where mutations are introduced into existing proteins, a process that can be facilitated by both experimental methods such as directed evolution [146], or using computational tools optimized for protein design and redesign of proteins [147]. A stabilization strategy that deserves special mention is *immobilization* of proteins, which not only increases stability but also offers advantages of easier operational control and product recovery in industrial applications. Among the various methods for protein immobilization [136], especially multipoint covalent attachment is effective when it comes to thermal stabilization and examples can be found where up to 30.000-fold increases in stability are achieved by immobilization on glyoxyl agarose gels [148].

Stability is an imperative quality to be considered in the design of any biological molecules intended for use in industry. It is a quality that must be thoroughly analysed not simply as an intrinsic static property of the biomolecule, but in the particular environment in which it is to be applied. It is therefore compelling to note that even in some of the harshest environments certain microorganisms thrive and feed on phosphorus [149], and therefore there is good reason to believe that a highly stable protein capable of binding phosphate can either be found in nature, or designed based on information obtained from such organisms, e.g. through data mining approaches such as the one presented in Chapter 2.

6.2 Possible Application Modules

Having designed a phosphorus bioscavenger, the next step is implementation of the scavenger into a scalable industrial application. Such an application can be theorized to take numerous forms, but in the following a few overall considerations are made for the characteristic cases where biomolecules are either used in solution or immobilized.

6.2.1 Solution-Based Bioscavengers

There are at least two particular advantages of a solution-based application over an immobilization-based application: 1) assuming that aggregation of the molecules is not an issue, it potentially allows for using higher densities of the bioscavengers, noting that in an immobilization-based application the density is limited by surface area rather than volume, and 2) the complication of the immobilization step is avoided, and subsequent disposal of inactive bioscavengers may be more straightforward. A simple illustrative example of a potential setup for a solution-based application is shown in Fig. 6.1, highlighting how any resource recovery application needs both a capture-phase and a release-phase.

Fig. 6.1 is only a rough illustration that serves the purpose of demonstrating some of the potential steps involved in a solution-based recovery application; specifically, it highlights that even though the step of immobilizing the bioscavengers covalently is avoided by having the bioscavengers in the solution, it introduces analogous complications as one needs to be able to retain the phosphate-scavenger complex from the rest of the feed solution during the capture phase, and at the end of the release phase the scavengers need to be separated from the phosphorus compound. Both of these complications impose additional design requirements for the bioscavenger molecule; e.g. in the form of a HIS-tag or similar peptide tag on the protein, which can be used for separation purposes.

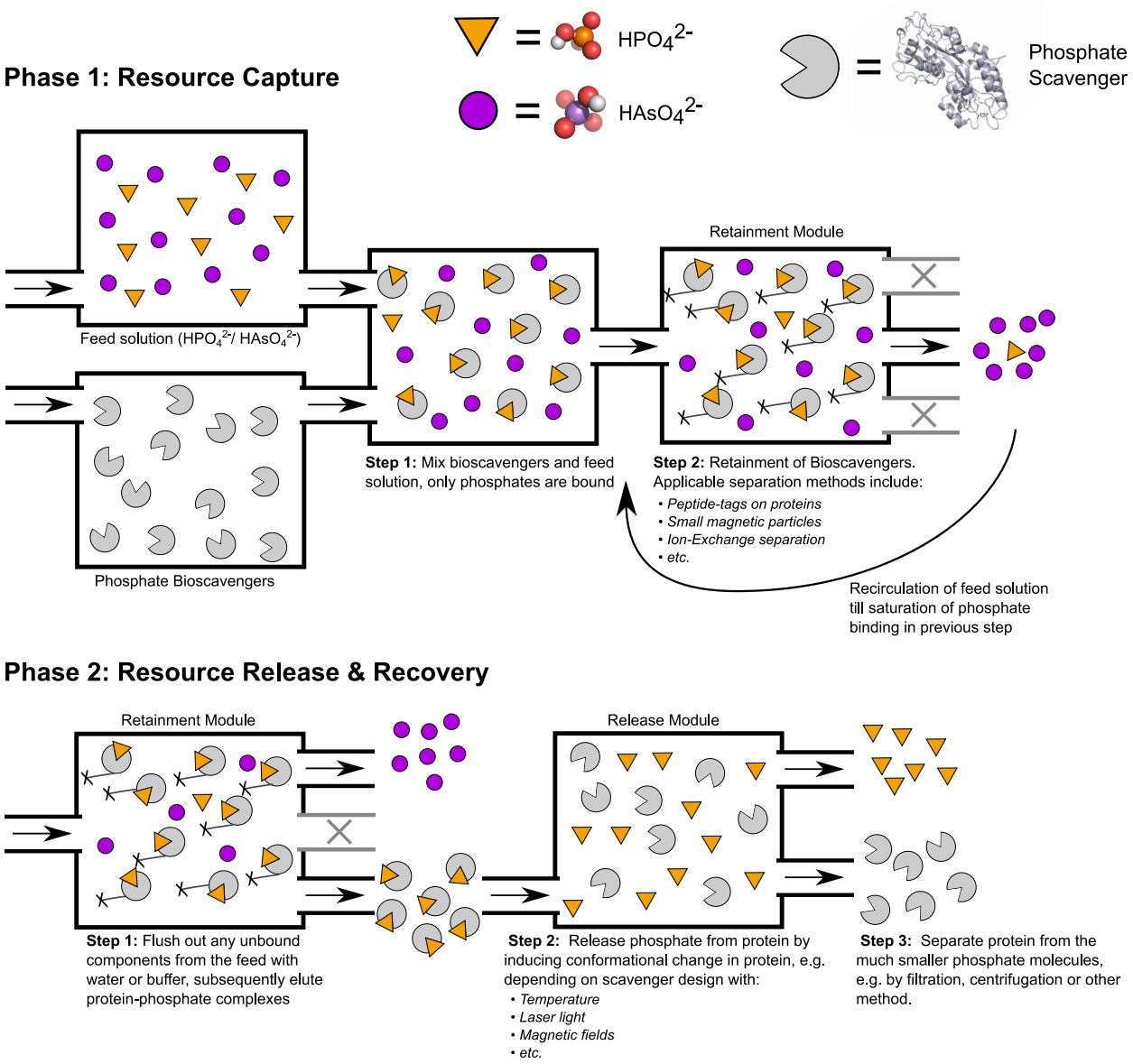

Figure 6.1: A schematic diagram demonstrating basic considerations for implementing a biomimetic phosphorus recovery application using solution-based phosphorus bioscavengers. In phase 1, a feed solution (exemplified as a mixture of HPO_4^{2-} and $HAsO_4^{2-}$) is mixed with a solution of bioscavengers, which selectively bind only the phosphate molecules. Next, the mixture is directed through a ”retainment module”, which selectively retains only the bioscavengers, allowing the rest of the feed solution to pass through and potentially recirculate back through the retainment module in order to saturate phosphate binding. In phase 2, the remnants of the feed solution is flushed out of the retainment module from step 2 in phase 1, whereafter the bioscavengers are allowed to elute, being transferred to the next module where the release mechanism for the bioscavengers is activated. Finally, bioscavengers must be separated from the phosphate such that they can be re-used in phase 1.

6.2.2 Immobilized Bioscavengers

As briefly mentioned in Sec. 6.1.1, immobilization of proteins offer advantages not only in terms of improved operational control of the proteins, but also the protein stability can be significantly increased by several orders of magnitude [136]. An illustration showing potential steps involved in an immobilization-based application is shown in Fig. 6.2. Similarly to the solution-based approach two phases are involved, capture and release, but looking at the schematic diagrams it is qualitatively suggested that operation of the immobilization-based technology is simpler and more efficient than the solution-based technology, mainly because the proposed "retainment" and separation steps are avoided as the scavengers are already irreversibly attached to a support structure.

The schematic in Fig. 6.2 provides little to no suggestions or assumptions for the specific details of the immobilization procedure or the nature of the "reaction module". The design of the reaction module depends on many factors, e.g. how the scavengers are immobilized (adsorption, covalent binding, entrapment, or membrane confinement [136]), the scavenger resource release mechanism, the feed solution, etc. As a result, numerous different setups can be theorized, but perhaps one of the most apparent ideas for the immobilization would be to incorporate the bioscavengers into a porous structure, so as to optimize surface area and scavenger density. Kim *et al.* recently published a review paper in which they provide an overview of immobilization methods and chemistries that might be relevant for such an approach, and in the review they also list advantages and disadvantages of different surface materials [150].

The first step in designing a biomimetic phosphorus recovery technology based on bioscavengers is the molecular design of the scavengers – subsequently, an application that accommodates the properties of the bioscavenger can be designed. Regardless of the specific application, it likely is going to involve flow of a presumably aqueous feed solution, and it might therefore be possible to use computational techniques such as CFD throughout the design process; i.e. to evaluate the feasibility of different application designs, to prototype different porous materials, and in the end optimize fluid flow in the application modules.

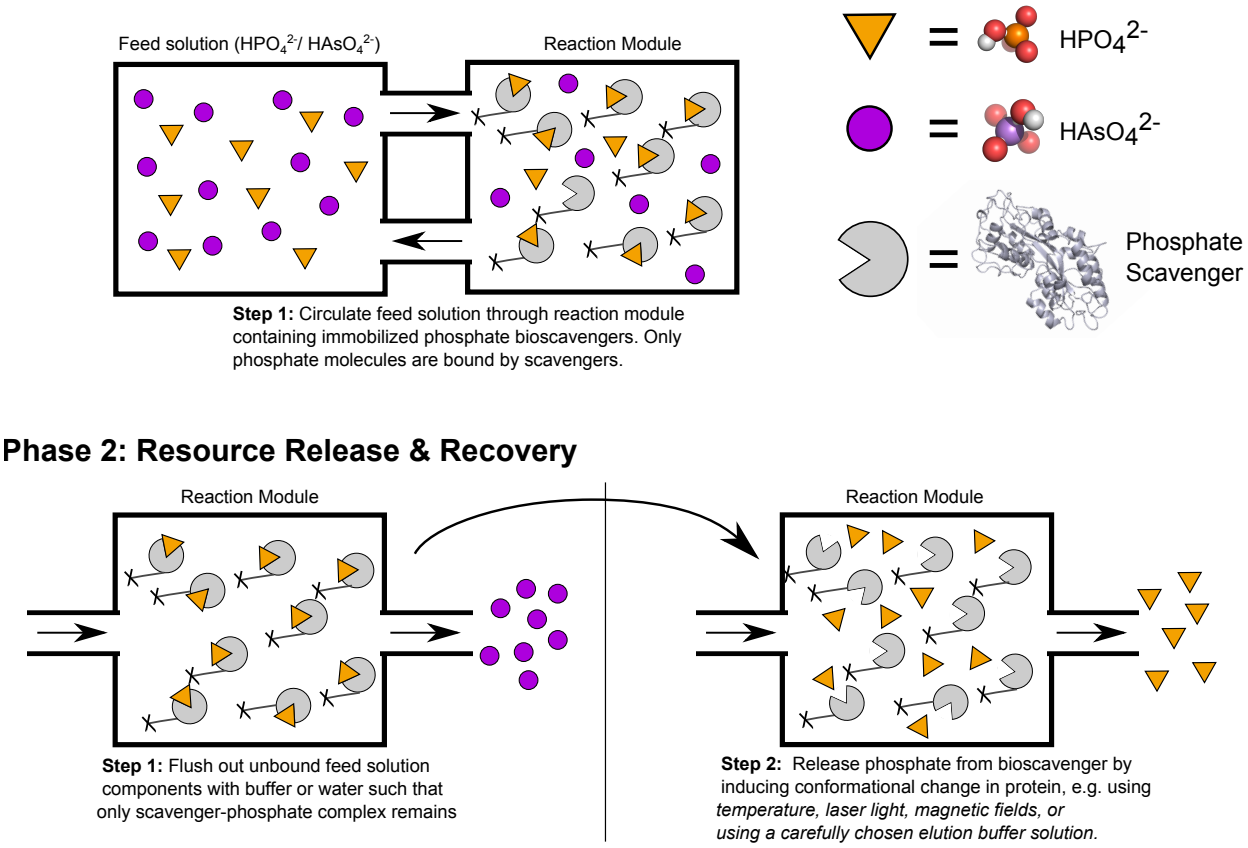

Figure 6.2: A schematic diagram demonstrating basic considerations for implementing a biomimetic phosphorus recovery application using immobilized phosphorus bioscavengers. In phase 1, a feed solution (exemplified as a mixture of HPO_4^{2-} and $HAsO_4^{2-}$) is circulated through a reaction chamber containing immobilized bioscavengers until binding is saturated. In phase 2, any unbound feed components are flushed out of the reaction chamber from phase 1, and subsequently the release mechanism of the bioscavengers is invoked, and the phosphorus compound can be recovered.

6.3 Final Conclusions & Remarks

Phosphorus is an essential and finite resource, and there is a call for the implementation of sustainable usage practises as well as technology development that can ensure that the resource is used and recycled efficiently in the future. Interactions between phosphates and proteins have been refined for billions of years in nature, which makes it attractive to pursue the design of a recovery application based on biological scavengers. Such a technology is however in its infancy and as the first step towards its implementation, in this work we have addressed some of the fundamental questions about both the design of bioscavengers and application modules.

Computational resources and software frameworks are becoming increasingly accessible, in effect giving us as researchers the ability to gain insight into systems of interest that can not readily be obtained experimentally and which can aid us in our pursuit of knowledge and technology advancement. In this work, aforesaid computational techniques have been applied to investigate different fundamental aspects of the development of a biomimetic phosphorus recovery technology, addressing how the techniques can be used to interpret large amounts of known data (data mining), gain new information about the systems of interest (QM and MD), as well as optimize a particular technology (CFD). We have thereby 1) obtained information about how proteins in nature interact with phosphates, which can be used in future rational design of phosphorus bioscavengers, 2) quantified how consistently phosphates are described using a variety of electronic structure methods, which is a crucial basis for running future simulations with phosphates, 3) shown how the interaction between a flexible hexapeptide and inorganic phosphate can be accurately described, which motivates the idea that computational methods can be used to prototype different molecular bioscavengers, and 4) demonstrated how complex macro-scale designs can be optimized and prototyped *in silico*.

The studies presented in this thesis not only elucidate central aspects for the systems in question but also effectively demonstrate how computational tools can and should be an integral part of an iterative experimental and theoretical design process during the development of advanced biomimetic technologies. Finally, it is crucial to note that the development of a biomimetic recovery technology is not limited to phosphorus – rather, phosphorus recovery can be seen as a model system for the development of broad range of biomimetic water treatment technologies, capable of not only recovering different valuable resources (phosphorus, nitrogen, lithium, biopolymers, etc.) but also removing challenging species such as pesticides and heavy metals from water streams.

Bibliography

1. U.S. Geological Survey. *Mineral Commodity Summaries 2015* tech. rep. (2015), 1–188.
2. Cordell, D., Turner, A. & Chong, J. The hidden cost of phosphate fertilizers: mapping multi-stakeholder supply chain risks and impacts from mine to fork. *Global Change, Peace & Security* **27,** 323–343 (2015).
3. Cordell, D. & Neset, T. Phosphorus vulnerability: A qualitative framework for assessing the vulnerability of national and regional food systems to the multi-dimensional stressors of phosphorus scarcity. *Global Environmental Change* **24,** 108–122 (2014).
4. Desmidt, E. *et al.* Global Phosphorus Scarcity and Full-Scale P-Recovery Techniques: A Review. *Critical Reviews in Environmental Science and Technology* **45,** 336–384 (2015).
5. Cordell, D., Drangert, J. & White, S. The story of phosphorus: Global food security and food for thought. *Global Environmental Change* **19,** 292–305 (2009).
6. Schröder, J. J., Cordell, D., Smit, A. L. & Rosemarin, A. *Sustainable Use of Phosphorus* tech. rep. (Plant Research International, 2009), 1–109.
7. Kauwenbergh, S. J. V. *World Phosphate Rock Reserves and Resources* tech. rep. (International Fertilizer Development Center, 2010), 1–48.
8. Rittmann, B. E., Mayer, B., Westerhoff, P. & Edwards, M. Capturing the lost phosphorus. *Chemosphere* **84,** 846–853 (2011).
9. Selman, M. & Greenhalgh, S. *Policy Note: Eutrophication: Policies, Actions, and Strategies to Address Nutrient Pollution* tech. rep. 3 (World Resources Institute, 2009), 1–16.
10. Rittmann, B. E. & McCarty, P. L. *Environmental biotechnology: principles and applications* (McGraw-Hill, 2001).

11. Yuan, Z., Pratt, S. & Batstone, D. J. Phosphorus recovery from wastewater through microbial processes. *Current Opinion in Biotechnology* **23,** 878–883 (2012).

12. Sartorius, C., Horn, J. V. & Tettenborn, F. Phosphorus Recovery from Wastewater—Expert Survey on Present Use and Future Potential. *Water Environment Research* **84,** 313–322 (2012).

13. Molinos-Senante, M., Hernández-Sancho, F., Sala-Garrido, R. & Garrido-Baserba, M. Economic feasibility study for phosphorus recovery processes. *AMBIO* **40,** 408–416 (2011).

14. Vincent, J. F. V., Bogatyreva, O. A., Bogatyrev, N. R., Bowyer, A. & Pahl, A. Biomimetics: its practice and theory. *Journal of The Royal Society Interface* **3,** 471–482 (2006).

15. Westheimer, F. H. Why Nature Chose Phosphates. *Science* **235,** 1173–1178 (1987).

16. Gruber, M. F., Greisen, P., Junker, C. M. & Hélix-Nielsen, C. Phosphorus binding sites in proteins: structural preorganization and coordination. *The Journal of Physical Chemistry. B* **118,** 1207–1215 (2014).

17. Willsky, G. R. & Malamy, M. H. Effect of arsenate on inorganic phosphate transport in Escherichia coli. *Journal of Bacteriology* **144,** 366–374 (1980).

18. Rao, N. N. & Torriani, A. Molecular aspects of phosphate transport in Escherichia coli. *Molecular Microbiology* **4,** 1083–1090 (1990).

19. Elias, M. *et al.* The molecular basis of phosphate discrimination in arsenate-rich environments. *Nature* **491,** 134–137 (2012).

20. Tang, C., Wang, Z., Petrinić, I., Fane, A. G. & Hélix-Nielsen, C. Biomimetic aquaporin membranes coming of age. *Desalination* **368,** 89–105 (2015).

21. Zhao, Y. *et al.* Synthesis of robust and high-performance aquaporin-based biomimetic membranes by interfacial polymerization-membrane preparation and RO performance characterization. *Journal of Membrane Science* **423-424,** 422–428 (2012).

22. Tang, C. Y., Zhao, Y., Wang, R., Hélix-Nielsen, C. & Fane, A. G. Desalination by biomimetic aquaporin membranes: Review of status and prospects. *Desalination* **308,** 34–40 (2013).

23. Jemli, S., Ayadi-Zouari, D., Hlima, H. B. & Bejar, S. Biocatalysts: application and engineering for industrial purposes. *Critical Reviews in Biotechnology* **36,** 246–258 (2016).

24. Robinson, P. K. Enzymes: principles and biotechnological applications. *Essays In Biochemistry* **59,** 1–41 (2015).

25. Choi, J., Han, S. & Kim, H. Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnology Advances* **33,** 1443–1454 (2015).

26. Jensen, F. *Introduction to computational chemistry* 2nd ed. (John Wiley & Sons Ltd, 2007).

27. Foresman, J. B. & Frish, Æ. *Exploring Chemistry with Electronic Structure Methods* 2nd ed. (Gaussian, Inc. Pittsburgh PA, 1996).

28. Rapaport, D. C. *The art of molecular dynamics simulation* 2nd ed. (Cambridge university press, 2004).

29. Frenkel, D. & Smit, B. *Understanding molecular simulation: from algorithms to applications* 2nd ed. (Academic Press, 2001).

30. Elo, L. L. & Schwikowski, B. Mining proteomic data for biomedical research. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2,** 1–13 (2012).

31. Inhester, T. & Rarey, M. Protein-ligand interaction databases: Advanced tools to mine activity data and interactions on a structural level. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4,** 562–575 (2014).

32. Cath, T. Y., Childress, A. E. & Elimelech, M. Forward osmosis: Principles, applications, and recent developments. *Journal of Membrane Science* **281,** 70–87 (2006).

33. Wesseling, P. *Principles of Computational Fluid Dynamics* (Springer, 2001).

34. Discala, C., Benigni, X., Barillot, E. & Vaysseix, G. DBcat: A catalog of biological databases. *Nucleic Acids Research* **28,** 8–9 (2000).

35. Benson, D. A. *et al.* GenBank. *Nucleic Acids Research* **41,** 36–42 (2013).

36. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28,** 235–242 (2000).

37. Li, X., Ng, S. & Wang, J. T. L. *Biological data mining and its applications in healthcare* (World Scientific, 2013).

38. Lee, P. M. *Bayesian statistics: an introduction* 4th ed. (John Wiley & Sons, 2012).

39. Chen, M., Han, J. & Yu, P. S. Data Mining: An Overview from a Database Perspective. *IEEE Transactions on Knowledge and Data Engineering* **8,** 866–883 (1996).

40. Hirsch, A. K. H., Fischer, F. R. & Diederich, F. Phosphate recognition in structural biology. *Angewandte Chemie International Edition* **46,** 338–352 (2007).

41. Rossmann, M. G., Moras, D. & Olsen, K. W. Chemical and biological evolution of nucleotide-binding protein. *Nature* **250,** 194–199 (1974).

42. Möller, W. & Amons, R. Phosphate-binding sequences in nucleotide-binding proteins. *FEBS Letters* **186,** 1–7 (1985).

43. Schulz, G. E. Binding of nucleotides by proteins. *Current Opinion in Structural Biology* **2,** 61–67 (1992).

44. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal* **1,** 945–951 (1982).

45. Saraste, M., Sibbald, P. R. & Wittinghofer, A. The P-loop—a common motif in ATP and GTP-binding proteins. *Trends in Biochemical Sciences* **15,** 430–434 (1990).

46. Hol, W. G., van Duijnen, P. T. & Berendsen, H. J. The alpha-helix dipole and the properties of proteins. *Nature* **273,** 443–446 (1978).

47. Dreusicke, D. & Schulz, G. E. The glycine-rich loop of adenylate kinase forms a giant anion hole. *FEBS Letters* **208,** 301–304 (1986).

48. Zhou, T., Daugherty, M., Grishin, N. V., Osterman, A. L. & Zhang, H. Structure and mechanism of homoserine kinase: Prototype for the GHMP kinase superfamily. *Structure* **8,** 1247–1257 (2000).

49. Denesyuk, A. I., Denessiouk, K. A., Korpela, T. & Johnson, M. S. Functional attributes of the phosphate group binding cup of pyridoxal phosphate-dependent enzymes. *Journal of Molecular Biology* **316,** 155–172 (2002).

50. Luecke, H. & Quiocho, F. A. High specificity of a phosphate transport protein determined by hydrogen bonds. *Nature* **347,** 402–406 (1990).

51. Dudev, T., Lin, Y., Dudev, M. & Lim, C. First-second shell interactions in metal binding sites in proteins: a PDB survey and DFT/CDM calculations. *Journal of the American Chemical Society* **125,** 3168–3180 (2003).

52. Maynard, A. T. & Covell, D. G. Reactivity of zinc finger cores: analysis of protein packing and electrostatic screening. *Journal of the American Chemical Society* **123,** 1047–1058 (2001).

53. Dudev, T. & Lim, C. Factors governing the protonation state of cysteines in proteins: an Ab initio/CDM study. *Journal of the American Chemical Society* **124,** 6759–6766 (2002).

54. He, Q. *et al.* Mutations at Nonliganding Residues Tyr-85 and Glu-83 in the N-Lobe of Human Serum Transferrin. *Journal of Biological Chemistry* **273,** 17018–17024 (1998).

55. DiTusa, C. A. *et al.* Thermodynamics of Metal Ion Binding. 2. Metal Ion Binding by Carbonic. *Biochemistry* **40,** 5345–5351 (2001).

56. Christianson, D. W. & Cox, J. D. Catalysis by Metal-Activated Hydroxide in Zinc and Manganese Metalloenzymes. *Annual Review of Biochemistry* **68,** 33–57 (1999).

57. Lipscomb, W. N. & Sträter, N. Recent Advances in Zinc Enzymology. *Chemical reviews* **96,** 2375–2434 (1996).

58. Atkins, P. W. & Friedman, R. S. *Molecular Quantum Mechanics* 4th ed. (Oxford University Press, USA, 2005).

59. Kohanoff, J. *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods* 1st ed. (Cambridge University Press, 2006).

60. Griffiths, D. J. *Introduction to Quantum Mechanics* 2nd ed. (Pearson Prentice Hall, 2004).

61. Echenique, P. & Alonso, J. L. A mathematical and computational review of Hartree–Fock SCF methods in quantum chemistry. *Molecular Physics* **105,** 3057–3098 (2007).

62. Slater, J. C. The theory of complex spectra. *Physical Review* **34,** 1293–1322 (1929).

63. Löwdin, P. Quantum theory of many-particle systems. III. Extension of the Hartree-Fock scheme to include degenerate systems and correlation effects. *Physical Review* **97,** 1509–1520 (1955).

64. Tew, D. P., Klopper, W. & Helgaker, T. Electron correlation: the many-body problem at the heart of chemistry. *Journal of Computational Chemistry* **28,** 1307–1320 (2007).

65. Møller, C. & Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **46,** 618–622 (1934).

66. Pople, J. A., Seeger, R. & Krishnan, R. Variational configuration interaction methods and comparison with perturbation theory. *International Journal of Quantum Chemistry* **12,** 149–163 (1977).

67. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Physical Review* **136,** 864–871 (1964).

68. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review* **140,** 1133–1138 (1965).

69. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **38,** 3098–3100 (1988).

70. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **37,** 785–789 (1988).

71. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **98,** 5648–5652 (1993).

72. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry* **98,** 11623–11627 (1994).

73. Becke, A. D. in *The challenge of d and f electrons* 165–179 (American Chemical Society: Washington, DC, 1989).

74. Vosko, S. H., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics* **58,** 1200–1211 (1980).

75. Chai, J. & Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *The Journal of Chemical Physics* **128,** 084106 (2008).

76. Chai, J. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Physical Chemistry Chemical Physics* **10,** 6615–6620 (2008).

77. Grimme, S. Density functional theory with London dispersion corrections. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1,** 211–228 (2011).

78. Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. *Journal of Chemical Physics* **140,** 18A301 (2014).

79. Hehre, W. J., Ditchfield, R. & Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **56,** 2257–2261 (1972).

80. Merz, K. M. Using quantum mechanical approaches to study biological systems. *Accounts of Chemical Research* **47,** 2804–2811 (2014).

81. Ball, P. Physics of life: The dawn of quantum biology. *Nature* **474,** 272–274 (2011).

82. Rustad, J. R., Dixon, D. A., Kubicki, J. D. & Felmy, A. R. Gas-phase acidities of tetrahedral oxyacids from ab initio electronic structure theory. *The Journal of Physical Chemistry A* **104,** 4051–4057 (2000).

83. Kish, M. M. & Viola, R. E. Oxyanion specificity of aspartate-beta-semialdehyde dehydrogenase. *Inorganic Chemistry* **38,** 818–820 (1999).

84. Rudbeck, M. Basis set dependence of phosphate frequencies in density functional theory calculations. *International Journal of Quantum Chemistry* **112,** 2435–2439 (2012).

85. Burrow, P. D., Gallup, G. A. & Modelli, A. Are there pi* shape resonances in electron scattering from phosphate groups? *The Journal of Physical Chemistry. A* **112,** 4106–4113 (2008).

86. Bianciotto, M., Barthelat, J. & Vigroux, A. Reactivity of phosphate monoester monoanions in aqueous solution. 2. A theoretical study of the elusive zwitterion intermediates RO+(H)PO3 2-. *The Journal of Physical Chemistry A* **106,** 6521–6526 (2002).

87. Buchachenko, A. L., Kuznetsov, D. A. & Breslavskaya, N. N. Ion-radical mechanism of enzymatic ATP synthesis: DFT calculations and experimental control. *The Journal of Physical Chemistry. B* **114,** 2287–2292 (2010).

88. Luque, N. B., Mujika, J. I., Rezabal, E., Ugalde, J. M. & Lopez, X. Mapping the affinity of aluminum(III) for biophosphates: interaction mode and binding affinity in 1:1 complexes. *Physical Chemistry Chemical Physics* **16,** 20107–20119 (2014).

89. Yamabe, S., Zeng, G., Guan, W. & Sakaki, S. A DFT study on proton transfers in hydrolysis reactions of phosphate dianion and sulfate monoanion. *Journal of Computational Chemistry* **35,** 2195–2204 (2014).

90. Dittrich, M., Hayashi, S. & Schulten, K. ATP hydrolysis in the betaTP and betaDP catalytic sites of F1-ATPase. *Biophysical Journal* **87,** 2954–2967 (2004).

91. Iino, R. & Noji, H. Operation mechanism of F(0) F(1)-adenosine triphosphate synthase revealed by its structure and dynamics. *IUBMB life* **65,** 238–246 (2013).

92. Pedersen, P. L., Ko, Y. H. & Hong, S. ATP synthases in the year 2000: evolving views about the structures of these remarkable enzyme complexes. *Journal of Bioenergetics and Biomembranes* **32,** 325–332 (2000).

93. Laskin, J., Yang, Z. & Woods, A. S. Competition between covalent and non-covalent bond cleavages in dissociation of phosphopeptide-amine complexes. *Physical Chemistry Chemical Physics* **13,** 6936–6946 (2011).

94. Singh, S. *et al.* Molecular interactions of 2'-deoxyguanosine 5'-monophosphate with glycine in aqueous media probed via concentration and pH dependent Raman spectroscopic investigations and DFT study. *Physical Chemistry Chemical Physics* **14,** 14315–14324 (2012).

95. Chakrabarty, K., Gupta, S. N., Das, G. K. & Roy, S. Theoretical studies on the pyridoxal-5'-phosphate dependent enzyme dopa decarboxylase: Effect of thr 246 residue on the co-factor-enzyme binding and reaction mechanism. *Indian Journal of Biochemistry and Biophysics* **49,** 155–164 (2012).

96. Wolfe-Simon, F. *et al.* A bacterium that can grow by using arsenic instead of phosphorus. *Science* **332,** 1163–1166 (2011).

97. Erb, T. J., Kiefer, P., Hattendorf, B., Günther, D. & Vorholt, J. A. GFAJ-1 is an arsenate-resistant, phosphate-dependent organism. *Science* **337,** 467–470 (2012).

98. Reaves, M. L., Sinha, S., Rabinowitz, J. D., Kruglyak, L. & Redfield, R. J. Absence of detectable arsenate in DNA from arsenate-grown GFAJ-1 cells. *Science* **337,** 470–473 (2012).

99. Baer, C. D., Edwards, J. O. & Rieger, P. H. Kinetics of the hydrolysis of arsenate(V) triesters. *Inorganic Chemistry* **20,** 905–907 (1981).

100. Mládek, A., Šponer, J., Sumpter, B. G., Fuentes-Cabrera, M. & Šponer, J. E. On the geometry and electronic structure of the As-DNA backbone. *The Journal of Physical Chemistry Letters* **2,** 389–392 (2011).

101. Jissy, A. K. & Datta, A. Can arsenates replace phosphates in natural biochemical processes? A computational study. *The Journal of Physical Chemistry B* **117,** 8340–8346 (2013).

102. Jensen, F. Describing anions by density functional theory: Fractional electron affinity. *Journal of Chemical Theory and Computation* **6,** 2726–2735 (2010).

103. Iikura, H., Tsuneda, T., Yanai, T. & Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *The Journal of Chemical Physics* **115,** 3540–3544 (2001).

104. Yanai, T., Tew, D. P. & Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters* **393,** 51–57 (2004).

105. Stewart, J. J. P. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling* **13,** 1173–1213 (2007).

106. Montgomery, J. A., Frisch, M. J., Ochterski, J. W. & Petersson, G. A. A complete basis set model chemistry. VII. Use of the minimum population localization method. *The Journal of Chemical Physics* **112,** 6532–6542 (2000).

107. Li, L., Li, C., Zhang, Z. & Alexov, E. On the dielectric "constant" of proteins: Smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *Journal of Chemical Theory and Computation* **9,** 2126–2136 (2013).

108. Leach, A. R. *Molecular modelling: principles and applications* 2nd ed. (Pearson Education Ltd, 2001).

109. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **65,** 712–725 (2006).

110. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117,** 5179–5197 (1995).

111. Case, D. A. *et al. AMBER 2015* (University of California, San Fransisco).

112. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159,** 98–103 (1967).

113. Bernardi, R. C., Melo, M. C. R. & Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochemica et Biophysica Acta* **1850,** 872–877 (2015).

114. Nymeyer, H., Gnanakaran, S. & García, A. E. Atomic Simulations of Protein Folding, Using the Replica Exchange Algorithm. *Methods in Enzymology* **383,** 119–149 (2004).

115. Cheng, X., Cui, G., Hornak, V. & Simmerling, C. Modified replica exchange simulation methods for local structure refinement. *The Journal of Physical Chemistry. B* **109,** 8220–8230 (2005).

116. Wu, X. & Brooks, B. R. Self-guided Langevin dynamics simulation method. *Chemical Physics Letters* **381,** 512–518 (2003).

117. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics* **120,** 11919–11929 (2004).

118. Miao, Y. *et al.* Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *Journal of Chemical Theory and Computation* **10,** 2677–2689 (2014).

119. Bianchi, A., Giorgi, C., Ruzza, P., Toniolo, C. & Milner-White, E. J. A synthetic hexapeptide designed to resemble a proteinaceous P-loop nest is shown to bind inorganic phosphate. *Proteins* **80,** 1418–1424 (2012).

120. Shaw, D. E. *et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer* in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (2014), 41–53.

121. Taylor, C. A., Fonte, T. A. & Min, J. K. Computational Fluid Dynamics Applied to Cardiac Computed Tomography for Noninvasive Quantification of Fractional Flow Reserve. *Journal of the American College of Cardiology* **61,** 2233–2241 (2013).

122. Thévenin, D. & Janiga, G. *Optimization and Computational Fluid Dynamics* 1st ed. (Springer, 2008).

123. Aubry, R., Vázquez, M., Houzeaux, G., Cela, J. M. & Marras, S. An unstructured CFD approach to numerical weather prediction. *48th AIAA Aerospace Sciences Meeting,* 1–26 (2010).

124. Anderson, J. *Computational Fluid Dynamics* 1st ed. (McGraw-Hill Education, 1995).

125. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing* 3rd ed. (Cambridge University Press, New York, USA, 2007).

126. Mulder, J. *Basic Principles of Membrane Technology* 2nd ed. (Springer, 1996).

127. Gruber, M. F. *et al.* Computational fluid dynamics simulations of flow and concentration polarization in forward osmosis membrane systems. *Journal of Membrane Science* **379,** 488–495 (2011).

128. Lee, K. L., Baker, R. W. & Lonsdale, H. K. Membranes for power generation by pressure-retarded osmosis. *Journal of Membrane Science* **8,** 141–171 (1981).

129. Loeb, S., Titelman, L., Korngold, E. & Freiman, J. Effect of porous support fabric on osmosis through a Loeb-Sourirajan type asymmertric membrane. *Journal of Membrane Science* **129,** 243–249 (1997).

130. Bird, R. B., Stewart, W. E. & Lightfoot, E. N. *Transport Phenomena* 2nd ed. (John Wiley & Sons, Inc, 2007).

131. Issa, R. I. Solution of the Implicitly Discretised Fluid Flow Equations by Operator-Splitting. *Journal of Computational Physics* **62,** 40–65 (1985).

132. Gruber, M. F. *et al.* Validation and Analysis of Forward Osmosis CFD Model in Complex 3D Geometries. *Membranes* **2,** 764–82 (2012).

133. Patankar, S. *Numerical heat transfer and fluid flow* (CRC Press, 1980).

134. Bray, B. L. Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nature Reviews Drug Discovery* **2,** 587–593 (2003).

135. Sewald, N. & Jakubke, H. *Peptides: Chemistry and Biology* 2nd ed. (Wiley, 2009).

136. Iyer, P. V. & Ananthanarayan, L. Enzyme stability and stabilization-Aqueous and non-aqueous environment. *Process Biochemistry* **43,** 1019–1032 (2008).

137. Zale, S. E. & Klibanov, A. M. On the role of reversible denaturation (unfolding) in the irreversible thermal inactivation of enzymes. *Biotechnology and Bioengineering* **25,** 2221–2230 (1983).

138. Lumry, R. & Eyring, H. Conformation changes of proteins. *The Journal of Physical Chemistry* **58,** 110–120 (1954).

139. Bell, L. N. Peptide stability in solids and solutions. *Biotechnology Progress* **13,** 342–346 (1997).

140. Bogdanowich-Knipp, S. J., Chakrabarti, S., Williams, T. D., Dillman, R. K. & Siahaan, T. J. Solution stability of linear vs. cyclic RGD peptides. *The Journal of Peptide Research* **53,** 530–541 (1999).

141. Gentilucci, L., De Marco, R. & Cerisoli, L. Chemical modifications designed to improve peptide stability: incorporation of non-natural amino acids, pseudo-peptide bonds, and cyclization. *Current Pharmaceutical Design* **16,** 3185–3203 (2010).

142. Henchey, L. K., Jochim, A. L. & Arora, P. S. Contemporary strategies for the stabilization of peptides in the α-helical conformation. *Current Opinion in Chemical Biology* **12,** 692–697 (2008).

143. Mozhaev, V. V. & Martinek, K. Structure-stability relationships in proteins: a guide to approaches to stabilizing enzymes. *Advanced Drug Delivery Reviews* **4,** 387–419 (1990).

144. Fágáin, C. Ó. Understanding and increasing protein stability. *Biochimica et Biophysica Acta* **1252,** 1–14 (1995).

145. Ladero, M. *et al.* Studies on the activity and the stability of beta-galactosidases from Thermus sp strain T2 and from Kluyveromyces fragilis. *Enzyme and Microbial Technology* **30,** 392–405 (2002).

146. Eijsink, V. G. H., Gåseidnes, S., Borchert, T. V. & van den Burg, B. Directed evolution of enzyme stability. *Biomolecular Engineering* **22,** 21–30 (2005).

147. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. Practically useful: What the Rosetta protein modeling suite can do for you. *Biochemistry* **49,** 2987–2998 (2010).

148. Fernandez-Lafuente, R., Cowan, D. A. & Wood, A. N. P. Hyperstabilization of a thermophilic esterase by multipoint covalent attachment. *Enzyme and Microbial Technology* **17,** 366–372 (1995).

149. Pikuta, E. V., Hoover, R. B. & Tang, J. Microbial Extremophiles at the Limits of Life. *Critical Reviews in Microbiology* **33,** 183–209 (2007).

150. Kim, D. & Herr, A. E. Protein immobilization techniques for microfluidic assays. *Biomicrofluidics* **7,** 041501 (2013).

Papers

Paper I Mathias F. Gruber, Per Jr. Greisen, Caroline M. Junker, and Claus Hélix-Nielsen, Phosphorus binding sites in proteins: structural preorganization and coordination. *The Journal of Physical Chemistry. B*, **2014**, 118, 1207-1215.

Paper II Mathias F. Gruber, Andrea Bordoni, Claus Hélix-Nielsen, Describing Phosphate, Sulphate and Arsenate with Quantum Methods: Does one-size-fits-all apply?, Submitted to *Journal of Computational Chemistry*.

Paper III Mathias F. Gruber, Elizabeth Wood, Sigurd Truelsen, Thomas Østergaard, and Claus Hélix-Nielsen, Computational Design of Biomimetic Phosphate Scavengers, *Environmental Science & Technology*, **2015**, 49, 9469–9478.

Paper IV Mathias F. Gruber, Ulf Aslak, Claus Hélix-Nielsen, Open-Source CFD Model for Optimization of Forward Osmosis and Reverse Osmosis Membrane Modules, *Separation and Purification Technology*, **2016**, 158, 183-192

In this online version of the thesis, **Papers I - IV** are not included but can be obtained from electronic article databases e.g. via www.orbit.dtu.dk or on request from:

DTU Environment,
Technical University of Denmark,
Miljoevej, Building 113,
2800 Kgs. Lyngby,
Denmark,

info@env.dtu.dk.

The Department of Environmental Engineering (DTU Environment) conducts science-based engineering research within four sections:
Water Resources Engineering, Urban Water Engineering,
Residual Resource Engineering and Environmental Chemistry & Microbiology.

The department dates back to 1865, when Ludvig August Colding, the founder of the department, gave the first lecture on sanitary engineering as response to the cholera epidemics in Copenhagen in the late 1800s.

DTU Environment
Department of Environmental Engineering
Technical University of Denmark

Miljoevej, building 113
2800 Kgs. Lyngby
Denmark

Phone: +45 4525 1600
Fax: +45 4593 2850
e-mail: info@env.dtu.dk
www.env.dtu.dk