Small and Large, Charged Molecules in Solution and at Interfaces

Tesei, Giulio

2018

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

Link to publication

Citation for published version (APA):
Tesei, G. (2018). *Small and Large, Charged Molecules in Solution and at Interfaces*. Lund University.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Small and Large, Charged Molecules in Solution and at Interfaces

Giulio Tesei

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Science, Lund University, Sweden.
To be defended on September 28 2018 at 13.15 in lecture hall B, Center for Chemistry and Chemical Engineering, Lund.

Faculty opponent

Professor Douglas J. Tobias
University of California, Irvine, USA

Organization
LUND UNIVERSITY
Center for Chemistry and Chemical Engineering
P. O. Box 124
SE–221 00 LUND
Sweden

Document name
DOCTORAL DISSERTATION

Date of issue
2018-09-04

Sponsoring organization

Author(s)
Giulio Tesei

Title and subtitle
Small and Large, Charged Molecules in Solution and at Interfaces

Abstract

The influence of small and large, charged molecules on the properties of aqueous interfaces has important implications in chemistry, biology, and medicine. For example, at the surface of marine aerosols, the presence of halide anions and tropospheric gases gives way to multiphase chemical reactions affecting climate and air quality. In this Thesis, we investigate three different systems and phenomena at the interface using molecular simulations, experiments, and theory.

A new accurate all-atom force field is developed to study thiocyanate ions in solution and at the interface. We show how the different cations affect the properties of NaSCN and KSCN bulk aqueous solutions and, supplementing vibrational sum frequency spectroscopy, our simulations indicate that the thiocyanate anion has a larger affinity for hydrophobic surfaces with exposed methyl groups than for the air–water interface.

Combining coarse-grained simulations and quartz crystal microbalance with dissipation monitoring experiments, we develop an analytical model to study the adsorption of amyloid fibrils onto oppositely charged lipid bilayers. The model shows that short, rigid fibrils adsorb more onto oppositely charged surfaces than long fibrils.

A computationally efficient coarse-grained model is developed to investigate the interaction of cationic peptides with lipid membranes. Additionally, we use small-angle X-ray scattering experiments and all-atom simulations to study the solution behavior of arginine-rich cell-penetrating peptides. Despite their large positive charge, we find that arginine decapeptides self-associate in aqueous solution. We elucidate the molecular origin of the attraction, and highlight its common occurrence in protein crystal structures.

Key words: molecular modeling, MD simulations, MC simulations, statistical mechanics, SAXS, QCM-D, Hofmeister series, amyloid fibril, lipid bilayer, cell-penetrating peptide

Classification system and/or index termes (if any):

Supplementary bibliographical information:

Language
English

ISSN and key title:

ISBN
978-91-7422-588-4 (print)
978-91-7422-589-1 (pdf)

Recipient's notes

Number of pages
182

Price

Security classification

Distribution by (name and address)

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Giulio Tesei

Date 2018-08-20

Small and Large, Charged Molecules in Solution and at Interfaces

Giulio Tesei

LUND
UNIVERSITY

Front cover: © Paolo Tesei and Paula Leckius 2018, photo by Das Sasha
Back cover: Coarse-grained representation of the adsorption of amyloid fibrils onto an oppositely charged lipid bilayer © Giulio Tesei 2018

Copyright © Giulio Tesei 2018
Doctoral Thesis

Division of Theoretical Chemistry
Department of Chemistry
Faculty of Science
Lund University

All rights reserved

ISBN 978-91-7422-588-4 (printed)
ISBN 978-91-7422-589-1 (digital)
Printed by Media-Tryck, Lund University

Abstract

The influence of small and large, charged molecules on the properties of aqueous interfaces has important implications in chemistry, biology, and medicine. For example, at the surface of marine aerosols, the presence of halide anions and tropospheric gases gives way to multiphase chemical reactions affecting climate and air quality. In this Thesis, we investigate three different systems and phenomena at the interface using molecular simulations, experiments, and theory.

A new accurate all-atom force field is developed to study thiocyanate ions in solution and at the interface. We show how different cations affect the properties of NaSCN and KSCN bulk aqueous solutions and, supplementing vibrational sum frequency spectroscopy, our simulations indicate that the thiocyanate anion has a higher affinity for hydrophobic surfaces with exposed methyl groups than for the air–water interface.

Combining coarse-grained simulations and quartz crystal microbalance with dissipation monitoring experiments, we develop an analytical model to study the adsorption of amyloid fibrils onto oppositely charged lipid bilayers. The model shows that short, rigid fibrils adsorb more onto oppositely charged surfaces than long fibrils.

A computationally efficient coarse-grained model is developed to investigate the interaction of cationic peptides with lipid membranes. Additionally, we use small-angle X-ray scattering experiments and all-atom simulations to study the solution behavior of arginine-rich cell-penetrating peptides. Despite their large positive charge, we find that arginine decapeptides self-associate in aqueous solution. We elucidate the molecular origin of the attraction, and highlight its common occurrence in protein crystal structures.

Populärvetenskaplig sammanfattning

Växelverkningar mellan små och stora molekyler har konsekvenser för reaktioner som kemister utför i sina provrör, lika mycket som för de livsnödvändiga reaktionerna som sker i vår kropp och våra celler. Förståelsen av dessa växelverkningar är nyckeln till att kunna utnyttja dem till vår fördel, till exempel i industriella processer eller i utveckling av mediciner. För att uppnå detta kan man utföra experiment med avancerade verktyg, till exempel när man försöker undersöka molekylära växelverkningar med röntgenstrålning. Tyvärr är dessa kraftfulla tekniker dyra och i allmänhet oförmögna att observera enskilda molekyler. I den här avhandlingen konstruerade vi nya datorbaserade modeller för att förutse och visa i detalj på hur vissa av dessa växelverkningar påverkar våra system, samtidigt som vi analyserar den information som kan erhållas från experiment. Specifikt undersökte vi de växelverkningar som sker vid gränsen mellan vatten—runt celler eller i ett provrör—och luft eller en mikroskopisk yta—såsom ett protein eller ett cellmembran.

I ett av dessa system, beskriver vi beteendet hos små molekyler som kallas *tiocyanatjoner*. Detta då de är viktiga inom såväl industri som biokemiska laboratorier. Den utvecklade datormodellen kan exakt förutse beteendet hos tiocyanatjoner vid olika gränsskikt.

Vi tittar också på märkliga biologiska processer som växelverkningen mellan en cells yttre membran och *amyloidfibriller*. Amyloidfibriller är stavformade klumpar av proteinliknande molekyler och är vanligtvis kopplade till flera sjukdomar såsom Alzheimers sjukdom. Vi har utvecklat datormodeller för att undersöka fibril–membran interaktionen, eftersom denna anses vara en viktig del mot att förstå fibrillernas giftiga verkan. Denna förvärvade kunskap kan få betydelse för att förklara de molekylära mekanismerna bakom sjukdomar och att utforma läkemedel mot dem.

Slutligen har vi utvecklat modeller och utfört experiment samt datorberäkningar i syfte att förstå hur och varför vissa molekyler som kallas *cellpenetrerande peptider* enkelt kan ta sig in i celler. Resultaten kan bidra till att designa nya cellpenetrerande peptider som kan styra läkemedel in i celler.

Popular Science Summary

Interactions between small and large molecules have implications not just for reactions that chemists conduct in their test tubes, but also for the life-supporting reactions that happen in our cells and body. Understanding these interactions is key to their exploitation for human benefit, including industrial processes and the design of drugs against diseases. This quest may involve performing experiments using advanced setups, for instance when trying to investigate molecular interactions using radiation such as X-rays. Unfortunately, these powerful techniques are expensive, and generally unable to shed light on single molecules. In this Thesis, we designed new accurate computer-based models to predict and explore in great detail how some of these interactions take place, while also analyzing the information that can be obtained from the experimental setups. Specifically, we investigated the interactions occurring at the junction between water—around the cells or in a test tube—and air or a microscopic surface—like that of a protein or a cell membrane.

In one such system, we describe the behavior of small molecules called *thiocyanate ions*, because of their remarkable interaction with proteins for industrial and technological applications. The developed computer model can accurately predict the behavior of thiocyanate ions at different interfaces.

We also look at critical biological processes, like the interaction between the outer membrane of a cell and *amyloid fibrils*. Amyloid fibrils are rod-shaped assemblies of protein-like molecules and are associated with several disorders such as Alzheimer's disease. We have developed computer models to probe the fibril–membrane interaction, as it is regarded as one of the first steps toward the toxic action of these assemblies. The acquired information can contribute to our understanding of these diseases and to design drugs against them.

Finally, we have developed models, performed computer simulations, and experiments, to understand how and why *cell-penetrating peptides* are able to enter cells very efficiently. Our findings may find application to designing new cell-penetrating peptides for transporting drugs into cells.

List of Papers

This Thesis is based on the following papers, which are referred to by Roman numerals, and are included in the Appendix.

I **Specific Cation Effects on SCN^- in Bulk Solution and at the Air–Water Interface**
Giulio Tesei, Vidar Aspelin, and Mikael Lund
The Journal of Physical Chemistry B, 122(19):5094–5105 DOI: 10/cphm, 2018

II **Do Monovalent Anions Preferentially Adsorb to Extended Hydrophobic Solid Interfaces Exposing Methyl Groups?**
Elizaveta Potapova, Giulio Tesei, Mikael Lund, and Eric Tyrode
Manuscript, 2018

III **Aggregate Size Dependence of Amyloid Adsorption onto Charged Interfaces**
Giulio Tesei, Erik Hellstrand, Kalyani Sanagavarapu, Sara Linse, Emma Sparr, Robert Vácha, and Mikael Lund
Langmuir, 34(4):1266–1273 DOI: 10/chz9, 2018

IV **Coarse-grained Model of Titrating Peptides Interacting with Lipid Bilayers**
Giulio Tesei, Mario Vazdar, and Mikael Lund
Manuscript, 2018

V **Self-association of a Highly Charged Arginine-rich Cell-penetrating Peptide**
Giulio Tesei, Mario Vazdar, Malene Ringkjøbing Jensen, Carolina Cragnell, Philip E. Mason, Jan Heyda, Marie Skepö, Pavel Jungwirth, and Mikael Lund
Proceedings of the National Academy of Sciences, 114(43):11428–11433
DOI: 10/cd5n, 2017

VI **Arginine "Magic": Guanidinium Like-Charge Ion Pairing from Aqueous Salts to Cell Penetrating Peptides**
Mario Vazdar, Jan Heyda, Philip E. Mason, Giulio Tesei, Christoph Allolio, Mikael Lund, and Pavel Jungwirth
Accounts of Chemical Research, 51(6):1455–1464 DOI: 10/gdsqqd, 2018

Papers Not Included in the Thesis

- **Micropatterned Carbon-on-Quartz Electrode Chips for Photocurrent Generation from Thylakoid Membranes**
  Ada-Ioana Bunea, Arto Heiskanen, Galina Pankratova, Giulio Tesei, Mikael Lund, Hans-Erik Åkerlund, Donal Leech, Niels Bent Larsen, Stephan Sylvest Keller, Lo Gorton, and Jenny Emnéus
  ACS Applied Energy Materials, 1(7):3313–3322 DOI: 10/cprv, 2018
- **Influence of Surface Concentration on Poly(vinyl alcohol) Behavior at the Water–Vacuum Interface: A Molecular Dynamics Simulation Study**
  Giulio Tesei, Gaio Paradossi, and Ester Chiessi
  The Journal of Physical Chemistry B, 118(24):6946–6955 DOI: 10/f57sfp, 2014

List of Contributions

I G.T., V.A., and M.L. designed research; G.T. and V.A. performed research; G.T. and V.A. analyzed data; G.T., V.A., and M.L. wrote the paper. G.T. was responsible for the submission and revision process.

II E.P., G.T., M.L., and E.T. designed research; E.P., G.T., and E.T. performed research; E.P., G.T., and E.T. analyzed data; G.T. and E.T. wrote the paper; G.T. performed MD simulations; and E.P. performed VSFS experiments.

III G.T., S.L., E.S., R.V., and M.L. designed research; G.T., E.H., K.S., S.L., R.V., and M.L. performed research; S.L. contributed new reagents/analytic tools; G.T., E.H., and K.S. analyzed data; G.T., S.L., E.S., R.V., and M.L. wrote the paper; G.T. performed QCM-D experiments and MC simulations; E.H. performed QCM-D experiments; and K.S. performed AFM experiments. G.T. was responsible for the submission and revision process.

IV G.T., M.V., and M.L. designed research; G.T. and M.V. performed research; G.T. and M.V. analyzed data; G.T. and M.L. wrote the paper; G.T. performed MC simulations and M.V. performed MD simulations.

V G.T., M.V., C.C., P.E.M., J.H., M.S., P.J., and M.L. designed research; G.T. and M.R.J. performed research; G.T. and M.R.J. analyzed data; G.T., P.J., and M.L. wrote the paper; G.T. performed SAXS experiments and MD simulations; and M.R.J. performed NMR experiments. G.T. and M.L. were responsible for the submission and revision process.

VI M.V., J.H., P.E.M., G.T., C.A., M.L., and P.J. designed research, performed research, and analyzed data; M.V. wrote the paper with contributions from the other authors.

List of Abbreviations

AMP	antimicrobial peptide
CG	coarse-grained
CPP	cell-penetrating peptide
DF	distribution function
GC	grand canonical
HSQC	heteronuclear single quantum coherence
KB	Kirkwood–Buff
KBI	Kirkwood–Buff integral
MC	Monte Carlo
MD	molecular dynamics
NMR	nuclear magnetic resonance
PMF	potential of mean force
QCM-D	quartz crystal microbalance with dissipation monitoring
QCR	quartz crystal resonator
RDF	radial distribution function
RMSD	root-mean-square deviation
RRP	arginine-rich peptide
SAXS	small-angle X-ray scattering
VSFS	vibrational sum frequency spectroscopy
WHAM	weighted histogram analysis method

Acknowledgements

I would like to acknowledge everyone who has supported me during the work that has lead to this Thesis. First of all, my supervisor Mikael Lund for his creative inspiration, encouragement, and appreciation. Working under his guidance has been enriching and fun. I wish to thank Ester Chiessi, my bachelor's and master's thesis supervisor, for inciting my interest in physical chemistry and for introducing me to the exciting field of molecular simulations. Many thanks to Marie Skepö, for introducing me to SAXS and for her guidance during our fruitful collaboration. I would also like to thank Mario Vazdar, Eric Tyrode, Robert Vácha, and Pavel Jungwirth for their valuable scientific input. For their kind help and advice with the experiments with lipids and fibrils, I am thankful to Emma Sparr and Sara Linse. I wish to thank my co-supervisor Jan Forsman for his contribution to this Thesis and toward my teaching of statistical thermodynamics. *Tusen tack till* Martin Trulsson as well for his advice in the statistical thermodynamics course. Bertil Halle's excellent notes on the theory of SAXS helped deepen my understanding of the technique, and are the main source for Chapter 5 in this Thesis. I wish to thank Gunilla Karlström, her Swedish classes will always be among the best memories of my PhD years. Steven Vancoillie, thank you for your help with quantum chemistry and for the runs in the woods. I wish to thank Paula Leckius for her assistance and precious suggestions. Björn Stenqvist, thank you for your generous and enthusiastic support to my mathematical questions. I wish to thank Carolina, Stephanie, Coralie, and Ellen for our trips to ESRF. Alexei, thank you for your help with the Thesis and for your great sense of humor. Vidar, I am grateful for our collaboration and for the discovery of Mässingshornet. I wish to thank all the members of my lab and everyone at the divisions of physical and theoretical chemistry for the nice environment, and in particular Helena Persson and Maria Södergren.

I am indebted to Francesco for his irreplaceable friendship; for his deep understanding, wise insight, and good humor, which I cherished. Jana, I feel lucky for the mutual affection we have built through the years, for all the fun we had sharing our office and in our free time, and of course, for our scientific discussions. Maria, thank you for standing by my side in good times and in bad, and for making my life at and outside the office much more fun. Giulia, thank you for all the moments we shared in Skåne, Göteborg, and Stockholm. Marco, thank you for being a great friend from day one and for your limitless patience. Veron-

ica, thank you for all the fun we have had and will keep having together. Regina, Vishal, and Damiano—I am so glad we met and thankful for the time we spent together. Serena, thank you for all our chats and for catching up as if time didn't pass. Viola, Irene, and Michi—thank you for coming to visit me in Sweden, and for always being close friends, despite the distance. Finally, I would like to thank Sandra, Kush, my aunt Patrizia and uncle Paolo, my brother Paolo, and my parents Gabriella and Ivo for their (un)conditional love and for everything I learned from them.

Contents

1 Introduction **1**

2 Kirkwood–Buff Solution Theory **5**
2.1 DFs in the Canonical Ensemble 5
2.2 DFs in the Grand Canonical Ensemble 7
2.3 DFs and Thermodynamic Quantities 8
2.4 Two-component Solutions 10

3 Biased Sampling Techniques **13**
3.1 Umbrella Sampling 15
3.2 Wang–Landau methods 18

4 Quartz Crystal Microbalance with Dissipation Monitoring **23**
4.1 Operating Principle 24
4.2 Modeling 26

5 Small-angle X-ray Scattering **29**
5.1 Form and Structure Factors 31
5.2 The Guinier Approximation 33
5.3 Structure Factor and Radial Distribution Function 34

6 Summary and Outlook **37**
6.1 Paper I 37
6.2 Paper II 38
6.3 Paper III 39
6.4 Paper IV 40
6.5 Paper V 41
6.6 Paper VI 42

References **43**

Appendix **49**

Chapter 1

Introduction

While chemical processes are commonly treated as bulk phenomena, complex molecular systems are often comprised of multiple phases. The natural consequence is the introduction of interfaces that serve as sites for chemical reactions and physical processes. The processes harbored at interfaces between an aqueous phase and another phase—liquid, solid, or vapor—often play crucial roles in chemistry, biology, and medicine. For instance, the lipid bilayer of the cell membrane meets the aqueous surroundings to generate a life-supporting interface, which acts as the medium for metabolite diffusion and intercellular signaling. Yet another common example is the air/water interface of marine aerosols, wherein ionic species are involved in chemical reactions with important implications in atmospheric chemistry.[1,2] Electrolytes are also present in physiological media, where the interplay of ions and the interface between biomolecules and aqueous phase affects protein stability and solubility[3] and the mechanical properties of lipid membranes.[4,5]

In 1888, Franz Hofmeister studied the effect of several anions and cations to yield precipitation of protein solutions.[6] While the seminal work used egg whites, the Hofmeister series naively rank ions based on their influence on the properties of the protein/water interface.[3] To understand these interactions in depth, in recent years, surface-sensitive experiments and molecular simulations have significantly advanced molecular-level understanding of aqueous interfaces. The two methods are highly complementary and interdependent. On the one hand, the analysis of experimental data requires an underlying theory, typically involving assumptions and approximations on the molecular-level structure of the system, which can be assessed by molecular simulations. On the other hand, molecular simulations of systems at the nanometer scale are commonly based on empirically derived models, *i.e.*, force fields, and hence extensive comparison with experimentally accessible data is required to substantiate simulation results.

This Thesis begins by developing such a force field. The development of force fields for molecular simulations of aqueous interfaces involves validation of properties of the bulk solution and of the surface region. In this Thesis, this challenging process is detailed in Paper I for force fields for sodium and potassium thiocyanate at the air/water interface, developed using Kirkwood–Buff theory (see Chapter 2). The interest in the thiocyanate anion is due to its prominent position at the extreme end of the Hofmeister series, *i.e.*, its ability to increase protein solubility. In Paper II, we use the developed force field for molecular simulations of sodium thiocyanate aqueous solutions, and complement it with nonlinear vibrational spectroscopy experiments, in a comparative study between the interface with air and with a hydrophobic solid surface. Both, the simulation predictions and experimental observations demonstrate that thiocyanate is attracted more to the hydrophobic surface than to the air/water interface.

The interface between aqueous solutions and biological lipid membranes is an area of interest, which has progressed in the last decade thanks to the synergy between molecular simulations and experiments. Now, we move toward biological and biochemical systems with interfaces and physical surface processes. The membrane/water interface is studied in Paper III and IV, focusing on the interactions of peptides and peptide aggregates with lipid bilayers. Phenomena pertaining to peptide–membrane interactions are typically characterized by high free energy barriers (see Chapter 3) as well as large time and size scales, *e.g.*, milliseconds and hundreds of nanometers.[7] Simulations of detailed models accounting for the interactions between all the atoms in the system are computationally expensive, if not infeasible. Therefore, a host of strategies, generically termed coarse-graining, have been devised to decrease the complexity of the numerical problem, by reducing the number of interacting particles.[8] Although lacking chemical details, the simplified representation conveyed by coarse-grained models has the advantage of highlighting the dominant microscopic features.

In Paper III, we develop and use a coarse-grained model, supplemented by quartz crystal microbalance with dissipation monitoring experiments (see Chapter 4), and theory to describe the adsorption of amyloid fibrils onto oppositely charged lipid membranes. Amyloid fibrils are self-assembled elongated aggregates of peptides associated with many neurodegenerative and other diseases.[9] A proposed mechanism of cellular toxicity of amyloid fibrils is based on their interaction with lipids, resulting in the disruption of the cell membrane.[10,11] In Paper III, we also study the dependence of adsorption on the ionic strength of the aqueous medium, and on the length of the particles.

In Paper IV, a coarse-grained model is developed to efficiently simulate the adsorption and translocation processes of cell-penetrating peptides across lipid bilayers. These molecules, rich in basic amino acids, spontaneously traverse cell membranes through the lipid bilayer/aqueous interface, and are promising candidates for drug delivery.[12] The emphasis of the coarse-grained model in this work is on the role of acid-base equilibria and the dependence of permeation energy on peptide length.

While the membrane translocation efficiency of cell-penetrating peptides is often attributed to peptide–membrane interactions, it has been suggested that the cooperative action of the peptides is also crucial for cellular uptake.[13] In Paper V, we compare intermolecular interactions of the highly positively charged cell penetrating peptide deca-arginine against deca-lysine, in aqueous solution. Via small-angle X-ray scattering (see Chapter 5), we show that deca-arginine self-associates, whereas deca-lysine does not. All-atom molecular dynamics simulations show that the attraction originates from interactions between the guanidinium moieties of arginine side chains, together with salt bridges involving carboxyl groups. We also analyzed publicly available protein crystal structures to find that this newly discovered mode of interaction in deca-arginine dimers is a common motif in biochemistry.

Finally, with the elevated interest in guanidinium–guanidinium interactions at interfaces, as well as in bulk, Paper VI is a review article concerning these interactions in aqueous solutions of guanidinium salts, the self-interaction of short and long arginine–rich peptides, and the aggregation and adsorption of long arginine-rich peptides onto lipid membranes.

In summary, this Thesis deals with disparate molecular systems related to aqueous interfaces, ranging from triatomic anions to peptide aggregates tens of nanometers in length. All-atom molecular dynamics force fields (Paper I), coarse-grained models (Paper III and IV), as well as analytical theory (Paper III) and a continuum model (Paper IV) have been developed to gain insight into phenomena covering a wide spectrum of length scales.
The following Chapters provide concise overviews of theories and methods that have played a key role in the work of this Thesis.

Chapter 2

Kirkwood–Buff Solution Theory

Kirkwood–Buff (KB) solution theory[14] is a general and exact theory relating thermodynamic properties of a multicomponent fluid to spatial correlations between pairs of molecules of the various species present in solution. The spatial correlations arise from interparticle interactions and are expressed by molecular distribution functions (DFs).[15,16] The molecular-level information conveyed by the DFs is related to local cross fluctuations between the densities of the components of the isotropic solution. Within the framework of the grand canonical ensemble, local density fluctuations provide the connection between DFs and macroscopic solution properties, *e.g.*, compressibility, partial molar volume, and the change in chemical potential in response to a change in composition.[17] In the derivation of KB theory, no assumptions are made regarding the shape of the particles or the pairwise additivity of the intermolecular interactions.[18] Therefore, the theory is valid for isotropic as well as anisotropic molecules, irrespective of the nature of the total potential energy. These aspects make KB theory suitable to the study of complex fluids, *e.g.*, protein solutions. On the one hand, KB theory is widely used in molecular simulations for force field parametrization and validation against experimental thermodynamic data[19] as well as to gain insight into phenomena such as ion pairing,[20] cosolute effects on association processes,[21] and folding equilibria.[22] On the other hand, the inverse KB theory[23] is used to estimate molecular-level properties, such as protein–solvent and protein–cosolute coordination numbers, from experimental measurements of macroscopic quantities.[24,25]

This chapter begins by introducing molecular DFs for closed and open systems; subsequently, the relationships between DFs and thermodynamic properties are described for one-component fluids; finally, the DFs are generalized for two-component solutions and the connection with activity derivatives is presented.

2.1 DFs in the Canonical Ensemble

For a system of N indistinguishable and rigid molecules in a closed system of volume V at constant temperature T, a configuration of the system is denoted by the vector $\mathbf{q} = (\mathbf{q_1}, .., \mathbf{q_N})$, where $\mathbf{q_i} = (\mathbf{r_i}, \mathbf{\Omega_i})$ defines the position, $\mathbf{r_i}$, and the orientation, $\mathbf{\Omega_i}$, of the ith molecule. Using the mathematical formalism of classical statistical mechanics, the system can be described by the canonical partition function[26]

$$Q(N, V, T) = \frac{1}{h^{3N} N!} \int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{q} \exp\left[-\beta H(\mathbf{p}, \mathbf{q})\right], \tag{2.1}$$

where h is Planck constant, $\mathbf{p}$ is the vector containing the momenta of the N molecules, $\beta = 1/(k_B T)$ is the inverse thermal energy, and $H(\mathbf{p}, \mathbf{q})$ is the Hamiltonian of the system.[26] In Equation 2.1, the first integral goes from $-\infty$ to ∞ over the three components of the N momenta, whereas the second integral is over V and over all possible orientations of the molecules. To simplify the notation in this Chapter, the probability density in the canonical ensemble is denoted by

$$P(\mathbf{q}) = \frac{\exp\left[-\beta U(\mathbf{q})\right]}{\int \mathrm{d}\mathbf{q} \exp\left[-\beta U(\mathbf{q})\right]}, \tag{2.2}$$

where $U(\mathbf{q})$ is the total potential energy of the system.
The average density of molecules at a given location $\mathbf{r}'$ is expressed by the singlet DF:[16]

$$\rho^{(1)}(\mathbf{r}') = \int \mathrm{d}\mathbf{q}\, P(\mathbf{q}) \sum_{i=1}^{N} \delta(\mathbf{r_i} - \mathbf{r}'), \tag{2.3}$$

where $\delta(\mathbf{r_i} - \mathbf{r}')$ is the 3D Dirac delta function. Due to the indistinguishability of molecules, the N integrals are identical, and for any molecule labeled 1, Equation 2.3 can be rewritten as

$$\rho^{(1)}(\mathbf{r}') = N \int \mathrm{d}\mathbf{q}\, P(\mathbf{q})\, \delta(\mathbf{r_1} - \mathbf{r}') = N P^{(1)}(\mathbf{r}'), \tag{2.4}$$

where $P^{(1)}(\mathbf{r}')$ is the probability density of Equation 2.2 evaluated at $\mathbf{r_1} = \mathbf{r}'$, and orientationally averaged. In a macroscopic, homogeneous system, the density is constant throughout V, *i.e.*, $\rho = N/V$, and the volume integral of $\rho^{(1)}(\mathbf{r}')$ is simply

$$\int_V \mathrm{d}\mathbf{r}' \rho^{(1)}(\mathbf{r}') = \rho \int_V \mathrm{d}\mathbf{r}' = N. \tag{2.5}$$

The probability density of simultaneously finding molecule 1 at $\mathbf{r}'$ and molecule 2 at $\mathbf{r}''$ is given by

$$P^{(2)}(\mathbf{r}', \mathbf{r}'') = \int \mathrm{d}\mathbf{q}\, P(\mathbf{q})\, \delta(\mathbf{r_1} - \mathbf{r}')\, \delta(\mathbf{r_2} - \mathbf{r}''), \tag{2.6}$$

whereas the joint probability of finding any molecule at $\mathbf{r}''$ and any other molecule at $\mathbf{r}'$, is expressed by the pair DF[16,27]

$$\begin{aligned} \rho^{(2)}(\mathbf{r}', \mathbf{r}'') &= \int \mathrm{d}\mathbf{q}\, P(\mathbf{q}) \sum_{i=1}^{N} \delta(\mathbf{r_i} - \mathbf{r}') \sum_{j=1}^{N} \delta(\mathbf{r_j} - \mathbf{r}'') \\ &= N(N-1)\, P^{(2)}(\mathbf{r}', \mathbf{r}''), \end{aligned} \tag{2.7}$$

where the factor $N(N-1)$ is the total number of pairs of molecules in the system. The normalization condition of $\rho^{(2)}(\mathbf{r}', \mathbf{r}'')$ is

$$\int_V \mathrm{d}\mathbf{r}' \int_V \mathrm{d}\mathbf{r}'' \rho^{(2)}(\mathbf{r}', \mathbf{r}'') = N(N-1). \tag{2.8}$$

The ratio $P^{(2)}(\mathbf{r}', \mathbf{r}'')/P^{(1)}(\mathbf{r}')$ is the conditional probability, $P^{(2)}(\mathbf{r}'|\mathbf{r}'')$, of finding molecule 2 at $\mathbf{r}''$ when molecule 1 is at $\mathbf{r}'$. This leads to the definition of the pair correlation function,[16] $g^{(2)}(\mathbf{r}', \mathbf{r}'')$, as the average number density of molecules in the volume element $\mathrm{d}\mathbf{r}''$, given a molecule in $\mathrm{d}\mathbf{r}'$:

$$\rho^{(1)}(\mathbf{r}'')\, g^{(2)}(\mathbf{r}', \mathbf{r}'')\mathrm{d}\mathbf{r}'' = \frac{\rho^{(2)}(\mathbf{r}', \mathbf{r}'')\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r}''}{\rho^{(1)}(\mathbf{r}')\, \mathrm{d}\mathbf{r}'}. \tag{2.9}$$

For an isotropic and homogeneous fluid, $\rho^{(2)}(\mathbf{r}', \mathbf{r}'') = \rho^2 g^{(2)}(\mathbf{r}', \mathbf{r}'')$; moreover, $g^{(2)}(\mathbf{r}', \mathbf{r}'')$ depends only on the scalar interparticle distance $r = |\mathbf{r}'' - \mathbf{r}'|$. Therefore, it is convenient to perform the change of variables $(\mathbf{r}', \mathbf{r}'') \rightarrow (\mathbf{r}', \mathbf{r} = \mathbf{r}'' - \mathbf{r}')$, followed by the transformation to polar coordinates $(\mathbf{r}', \mathbf{r}) \rightarrow (\mathbf{r}', r, \theta, \phi)$. The resulting $g(r)$ is the radial DF (RDF), defined so that $\rho g(r) 4\pi r^2 \mathrm{d}r$ equals the average number of molecules in a spherical shell of radius r and thickness $\mathrm{d}r$, centered around a fixed particle. The relationship between the RDF and the pair DF is given by

$$\rho g(r)\, 4\pi r^2\, \mathrm{d}r = \rho^{(1)}(\mathbf{r}'')\, g^{(2)}(\mathbf{r}', \mathbf{r}'')\, \mathrm{d}\mathbf{r}'' = \frac{\rho^{(2)}(\mathbf{r}', \mathbf{r}'')\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r}''}{\rho^{(1)}(\mathbf{r}')\, \mathrm{d}\mathbf{r}'}. \tag{2.10}$$

2.2 DFs in the Grand Canonical Ensemble

In an open system, *i.e.*, in a region of volume V of an infinitely large system, the number of molecules fluctuates. This scenario is described by the grand canonical (GC) ensemble, with partition function

$$\Xi(\mu, V, T) = \sum_{N=0}^{\infty} Q(N, V, T) \exp{(\beta \mu N)}, \tag{2.11}$$

where μ is the chemical potential of the molecules. The probability of finding exactly N particles in the open system of volume V is[16,17]

$$P(N) = \frac{Q(N, V, T)}{\Xi(\mu, V, T)} \exp{(\beta \mu N)}. \tag{2.12}$$

The ensemble average, $\langle \dots \rangle$, of N is readily obtained as

$$\langle N \rangle = \sum_{N=0}^{\infty} N\, P(N). \tag{2.13}$$

Analogously, the singlet and pair DFs for the GC ensemble are obtained as averages of the corresponding canonical DFs running over N and weighted by $P(N)$.

In the notation of this Chapter, DFs defined for the open system are labeled with the subscript O to distinguish them from the canonical DFs. The singlet DF is defined for systems of at least a single molecule, therefore,

$$\rho_O^{(1)}(\mathbf{r}') = \sum_{N=1}^{\infty} \rho^{(1)}(\mathbf{r}')\, P(N), \tag{2.14}$$

whereas, the pair DF requires a minimum of two particles for its definition:

$$\rho_O^{(2)}(\mathbf{r}', \mathbf{r}'') = \sum_{N=2}^{\infty} \rho^{(2)}(\mathbf{r}', \mathbf{r}'')\, P(N). \tag{2.15}$$

Considering the relationship between DFs and probability densities (Equations 2.4 and 2.7), the volume integrals of $\rho_O^{(1)}(\mathbf{r}')$ and $\rho_O^{(2)}(\mathbf{r}', \mathbf{r}'')$ can be derived as

$$\int_V \mathrm{d}\mathbf{r}' \rho_O^{(1)}(\mathbf{r}') = \sum_{N=1}^{\infty} N\, P(N) \int_V \mathrm{d}\mathbf{r}' P^{(1)}(\mathbf{r}') = \langle N \rangle \tag{2.16}$$

and

$$\begin{aligned} &\int_V \mathrm{d}\mathbf{r}' \int_V \mathrm{d}\mathbf{r}'' \rho_O^{(2)}(\mathbf{r}', \mathbf{r}'') \\ &= \sum_{N=2}^{\infty} N(N-1)\, P(N) \int_V \mathrm{d}\mathbf{r}' \int_V \mathrm{d}\mathbf{r}'' P^{(2)}(\mathbf{r}', \mathbf{r}'') \\ &= \langle N^2 \rangle - \langle N \rangle, \end{aligned} \tag{2.17}$$

respectively.

2.3 DFs and Thermodynamic Quantities

The volume integral of the RDF of an open system:

$$G_O = \int_0^{\infty} \mathrm{d}r 4\pi r^2 \left[g_O(r) - 1 \right] \tag{2.18}$$

takes the name of Kirkwood–Buff integral (KBI),[18,28] for its central role in KB solution theory. Using Equations 2.16, 2.17, and 2.10, KBI can be solved to obtain a relationship between DFs and the fluctuation in the number of molecules in the open system:

$$\begin{aligned} G_O &= \frac{1}{\rho} \int_0^{\infty} \mathrm{d}r 4\pi r^2 \left[\rho g_O(r) - \rho \right] \\ &= \frac{1}{\rho} \int_V \mathrm{d}\mathbf{r}'' \left[\rho_O^{(1)}(\mathbf{r}'') g_O^{(2)}(\mathbf{r}', \mathbf{r}'') - \rho_O^{(1)}(\mathbf{r}'') \right] \frac{\int_V \mathrm{d}\mathbf{r}' \rho_O^{(1)}(\mathbf{r}')}{\int_V \mathrm{d}\mathbf{r}' \rho_O^{(1)}(\mathbf{r}')} \\ &= \frac{1}{\rho} \left[\frac{\int_V \mathrm{d}\mathbf{r}' \int_V \mathrm{d}\mathbf{r}'' \rho_O^{(2)}(\mathbf{r}', \mathbf{r}'') - \int_V \mathrm{d}\mathbf{r}' \rho_O^{(1)}(\mathbf{r}') \int_V \mathrm{d}\mathbf{r}'' \rho_O^{(1)}(\mathbf{r}'')}{\int_V \mathrm{d}\mathbf{r}' \rho_O^{(1)}(\mathbf{r}')} \right] \\ &= \frac{1}{\rho} \left[\frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} - 1 \right], \end{aligned} \tag{2.19}$$

where $\rho = \langle N \rangle / V$. The fluctuation in N, $\sqrt{\langle N^2 \rangle - \langle N \rangle^2}$, a central quantity of the GC ensemble, is related to the chemical potential via[29]

$$\langle N^2 \rangle - \langle N \rangle^2 = \frac{1}{\beta} \left(\frac{\partial \langle N \rangle}{\partial \mu} \right)_{VT} = \frac{V}{\beta} \left(\frac{\partial \rho}{\partial \mu} \right)_T, \tag{2.20}$$

where the second equality follows from Equation 2.13. Using the chain rule and the Gibbs–Duhem equation, Equation 2.20 can be manipulated as follows:[29]

$$\begin{aligned} \langle N^2 \rangle - \langle N \rangle^2 &= \frac{V}{\beta} \left(\frac{\partial \rho}{\partial p} \right)_T \left(\frac{\partial p}{\partial \mu} \right)_T = \frac{V}{\beta} \left(\frac{\partial \rho}{\partial p} \right)_T \rho \\ &= \frac{\langle N \rangle}{\beta} \left(\frac{\partial \rho}{\partial V} \right)_T \left(\frac{\partial V}{\partial p} \right)_T = -\frac{\rho^2}{\beta} \left(\frac{\partial V}{\partial p} \right)_T, \end{aligned} \tag{2.21}$$

where p is the pressure. Equation 2.21 relates the fluctuation in N in an open system to the relative change in V, in response to a change in p, *i.e.*, to the isothermal compressibility:[29]

$$\chi_T = -\frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_T. \tag{2.22}$$

Combining Equation 2.19 and Equation 2.21, KBI is found to be related to the ratio between the χ_T of the fluid and the compressibility of an ideal gas,[30] $\chi_T^{id} = \beta / \rho$:

$$1 + \rho G_0 = \frac{\chi_T}{\chi_T^{id}}. \tag{2.23}$$

Equation 2.23 shows that the information on the interparticle correlations incorporated in KBI can be directly translated to a macroscopic quantity pertaining to the net intermolecular interactions in the system. However, this result is strictly valid in the GC ensemble, whereas for a closed system:

$$1 + \rho G = 1 + \rho \int_0^\infty \mathrm{d}r 4\pi r^2 \left[g(r) - 1 \right] = 0. \tag{2.24}$$

Molecular simulations are commonly performed for systems of finite size in the canonical or isothermal-isobaric ensembles. Several approaches have been suggested[31–35] to accurately estimate thermodynamic properties from KBIs truncated at an upper limit, R:

$$G(R) = \int_0^R \mathrm{d}r 4\pi r^2 \left[g(r) - 1 \right]. \tag{2.25}$$

In Paper I, these methods are discussed and applied to the calculation of activity derivatives of electrolyte solutions from all-atom molecular dynamics simulations in the isothermal-isobaric ensemble. Moreover, in Chapter 5, Figure 5.3 shows an example of the usage of Equation 2.25 for the calculation of the χ_T of a Lennard–Jones fluid from Metropolis Monte Carlo simulations in the canonical ensemble.

2.4 Two-component Solutions

For a solution of N_i molecules of species i and N_j molecules of species j, the vector of configurations can be written as[14,17,18]

$$\mathbf{q} = (\mathbf{q}_1^{\mathbf{i}}, ..., \mathbf{q}_{\mathbf{N_i}}^{\mathbf{i}}, \mathbf{q}_{\mathbf{N_i+1}}^{\mathbf{j}}, ..., \mathbf{q}_{\mathbf{N_i+N_j}}^{\mathbf{j}}). \tag{2.26}$$

The definition of the molecular DFs are analogous to the one-component case:

$$\rho_i^{(1)}(\mathbf{r}') = N \int \mathrm{d}\mathbf{q}\, P(\mathbf{q})\, \delta(\mathbf{r}_1^{\mathbf{i}} - \mathbf{r}') = N_i\, P_i^{(1)}(\mathbf{r}'), \tag{2.27}$$

and

$$\begin{aligned} \rho_{ij}^{(2)}(\mathbf{r}', \mathbf{r}'') &= N_i(N_j - \delta_{ij}) \int \mathrm{d}\mathbf{q}\, P(\mathbf{q})\, \delta(\mathbf{r}_1^{\mathbf{i}} - \mathbf{r}')\, \delta(\mathbf{r}_2^{\mathbf{j}} - \mathbf{r}'') \\ &= N_i(N_j - \delta_{ij})\, P_{ij}^{(2)}(\mathbf{r}', \mathbf{r}''), \end{aligned} \tag{2.28}$$

where $\mathbf{r}'$ and $\mathbf{r}''$ refer to species i and j, respectively, and δ_{ij} is the Kronecker delta, which is used to compactly discern between the case of like molecules, *i.e.*, $N_i(N_i - 1)$ or $N_j(N_j - 1)$ pairs, and the case of unlike molecules, *i.e.*, N_iN_j pairs. According to Equation 2.10, the RDFs for a binary mixture are defined as

$$\rho_j g_{ij}(r)\, 4\pi r^2\, \mathrm{d}r = \rho_j^{(1)}(\mathbf{r}'')\, g_{ij}^{(2)}(\mathbf{r}', \mathbf{r}'')\, \mathrm{d}\mathbf{r}'' = \frac{\rho_{ij}^{(2)}(\mathbf{r}', \mathbf{r}'')\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r}''}{\rho_i^{(1)}(\mathbf{r}')\, \mathrm{d}\mathbf{r}'}. \tag{2.29}$$

The partition function for a two-component system is

$$\Xi(\mu_i, \mu_j, V, T) = \sum_{N_i=0}^{\infty} \sum_{N_j=0}^{\infty} Q(N_i, N_j, V, T)\, \exp{(\beta\mu_i N_i)}\, \exp{(\beta\mu_j N_j)}. \tag{2.30}$$

In close analogy to Equation 2.13, an expression can be obtained for the ensemble average of one of the components:

$$\begin{aligned} \langle N_i \rangle &= \sum_{N_i=0}^{\infty} \sum_{N_j=0}^{\infty} N_i\, P(N_i, N_j) \\ &= \frac{1}{\Xi} \sum_{N_i=0}^{\infty} \sum_{N_j=0}^{\infty} N_i\;\, Q\, \exp{(\beta\mu_i N_i)}\, \exp{(\beta\mu_j N_j)} \\ &= \frac{1}{\beta} \left[\frac{\partial \ln \Xi}{\partial \mu_i} \right]_{\mu_j, V, T}. \end{aligned} \tag{2.31}$$

Therefore, the normalization conditions of the singlet and pair DFs for a two-component, open system are[14,18]

$$\int_V \mathrm{d}\mathbf{r}' \rho_{Oi}^{(1)}(\mathbf{r}') = \sum_{N_i=1}^{\infty} \sum_{N_j=0}^{\infty} N_i\, P(N_i, N_j) = \langle N_i \rangle \tag{2.32}$$

and

$$\begin{aligned}&\int_V \mathrm{d}\mathbf{r}' \int_V \mathrm{d}\mathbf{r}'' \rho^{(2)}_{Oij}(\mathbf{r}', \mathbf{r}'') \\ &= \sum_{N_i=1+\delta_{ij}}^{\infty} \sum_{N_j=1-\delta_{ij}}^{\infty} N_i(N_j - \delta_{ij})\, P(N_i, N_j) = \langle N_i N_j \rangle - \langle N_i \rangle \delta_{ij},\end{aligned} \tag{2.33}$$

respectively. The definition of KBIs follows as a generalization of Equation 2.19

$$\begin{aligned}G_{Oij} &= \int_0^\infty \mathrm{d}r 4\pi r^2 \left[g_{Oij}(r) - 1\right] = \frac{1}{\rho_j} \int_0^\infty \mathrm{d}r 4\pi r^2 \left[\rho_j g_{Oij}(r) - \rho_j\right] \\ &= \frac{1}{\rho_j} \left[\frac{\int_V \mathrm{d}\mathbf{r}' \int_V \mathrm{d}\mathbf{r}'' \rho^{(2)}_{Oij}(\mathbf{r}', \mathbf{r}'') - \int_V \mathrm{d}\mathbf{r}' \rho^{(1)}_{Oi}(\mathbf{r}') \int_V \mathrm{d}\mathbf{r}'' \rho^{(1)}_{Oj}(\mathbf{r}'')}{\int_V \mathrm{d}\mathbf{r}' \rho^{(1)}_{Oi}(\mathbf{r}')} \right] \\ &= \frac{V}{\langle N_j \rangle} \left[\frac{\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle}{\langle N_i \rangle} - \delta_{ij} \right] = G_{Oji}.\end{aligned} \tag{2.34}$$

This result shows that, for a two-component solution, KBI is related to the cross fluctuations of the number of particles of species i and j.[18] Moreover, $\rho_j G_{Oij}$ can be interpreted as the excess coordination number[20,35] quantifying the depletion or surplus of particles j around particle i with respect to $\langle N_j \rangle$.
The connection to the chemical potential is obtained by deriving $\langle N_i \rangle$ with respect to μ_j:

$$\left(\frac{\partial \langle N_i \rangle}{\partial \mu_j} \right)_{\mu_i, V, T} = \beta \left(\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle \right) = V \beta \rho_i \left(\delta_{ij} + \rho_j G_{Oij} \right), \tag{2.35}$$

where the second equality is based on Equation 2.34. Thermodynamic solution data is often measured at constant pressure and temperature as a function of composition. Equation 2.35 can be recast in the isothermal-isobaric ensemble to obtain an expression connecting the RDFs to the change in the chemical potential of species i in response to a change in the number density of species i:[14,17,18]

$$\beta \rho_i \left(\frac{\partial \mu_i}{\partial \rho_i} \right)_{p,T} = \frac{1}{1 + \rho_i \left(G_{Oii} - G_{Oij} \right)}, \tag{2.36}$$

or equivalently:

$$\left(\frac{\partial \ln a_i}{\partial \ln \rho_i} \right)_{p,T} = \frac{1}{1 + \rho_i \left(G_{Oii} - G_{Oij} \right)}, \tag{2.37}$$

where a_i is the activity of species i. Equation 2.37 provides a convenient route to obtain the activity derivatives of solutes from molecular simulation of solutions. In Paper I, this approach has been applied to three-component systems to parametrize force fields for aqueous electrolyte solutions of several salts against experimental activity data.

Chapter 3

Biased Sampling Techniques

Metropolis Monte Carlo (MC) and molecular dynamics (MD) simulations are computational techniques to explore the coordinate or phase space of a system described at a microscopic level. Since the number of possible states for a many-particle system is extremely large, molecular simulations aim at generating configurations that are *representative* of a particular ensemble.[36] For N particles in the canonical ensemble, a configuration is a set of $3N$ Cartesian coordinates, $\mathbf{r}$, with potential energy $U(\mathbf{r})$. Representative configurations are the thermally populated states, *i.e.*, those with negative, or small and positive $U(\mathbf{r})$ values, which contribute the most to the configurational partition function:[26,36]

$$Q_{NVT}^{ex} = \frac{1}{V^N} \int d\mathbf{r} \exp\left[-\beta U(\mathbf{r})\right] = \frac{Z_{NVT}}{V^N}, \tag{3.1}$$

where Z_{NVT} is the configurational integral, V is the volume of the system, and $\beta = 1/k_B T$. Denoting ensemble averages by $\langle \dots \rangle$, Equation 3.1 can be rewritten as

$$Q_{NVT}^{ex} = \frac{\int d\mathbf{r} \exp\left[-\beta U(\mathbf{r})\right]}{\int d\mathbf{r} \exp\left[\beta U(\mathbf{r})\right] \exp\left[-\beta U(\mathbf{r})\right]} = \frac{1}{\Big\langle \exp\left(\beta U\right) \Big\rangle}. \tag{3.2}$$

In principle, the excess Helmholtz free energy can be calculated as

$$A^{ex} = -k_B T \ln Q_{NVT}^{ex} = k_B T \ln \Big\langle \exp\left(\beta U\right) \Big\rangle, \tag{3.3}$$

however, conventional simulation methods are targeted to selectively sample regions of phase space with small values of $\exp\left(\beta U\right)$. This entails that the convergence of the ensemble average in Equation 3.3 is in general difficult to attain. In biophysics, it is often of interest to calculate the free energy difference between stable and metastable states as well as the height of the free energy barriers in-between. The free energy landscapes can be simplified as free energy profiles or surfaces by identifying one or two generalized coordinates that discriminate between the thermodynamic states of interest. These so-called reaction coordinates, $\Re(\mathbf{r})$, are functions of the coordinates of the particles such as the center-of-mass separation between two molecules (see Paper V), the end-to-end distance of a polymer chain, or the angle between the principal axis of a rod-like particle and the normal to an interface (see Paper IV). The probability density for the reaction coordinate at any $\Re$ is given by the ensemble average of the Dirac delta function $\delta[\Re(\mathbf{r}) - \Re]$:

$$\rho(\Re) = \Big\langle \delta[\Re(\mathbf{r}) - \Re] \Big\rangle = \frac{\int \mathrm{d}\mathbf{r}\, \delta[\Re(\mathbf{r}) - \Re] \exp\left[-\beta U(\mathbf{r})\right]}{Z_{NVT}}, \tag{3.4}$$

where at the numerator $\exp\left[-\beta U(\mathbf{r})\right]$ is integrated over all degrees of freedom but $\Re$ to obtain $Z_{NVT}(\Re)$.[36] The change in free energy along $\Re$ with respect to a reference $\Re$ value is expressed by[36]

$$\Delta\mathcal{F}(\Re) = -k_B T \ln \rho(\Re) = -k_B T \ln \frac{Z_{NVT}(\Re)}{Z_{NVT}}. \tag{3.5}$$

$\Delta\mathcal{F}(\Re)$ is commonly referred to as the potential of mean force (PMF) due to the following relationship with the mechanical force along $\Re$:

$$-\frac{\mathrm{d}\Delta\mathcal{F}(\Re)}{\mathrm{d}\Re} = \frac{k_B T}{\rho(\Re)} \frac{\mathrm{d}\rho(\Re)}{\mathrm{d}\Re} = \left\langle -\frac{\mathrm{d}U}{\mathrm{d}\Re} \right\rangle_{\Re(\mathbf{r})=\Re}. \tag{3.6}$$

If $\Re(\mathbf{r})$ is a nonlinear function of $\mathbf{r}$, a change of variables is required to perform the integral at the numerator of Equation 3.4.[37,38] Specifically, the transformation is $\mathbf{r} = (r_1, r_2, ..., r_{3N}) \to \mathbf{u} = (\Re, \mathbf{q})$, where $\mathbf{q} = (q_2, ..., q_{3N})$ is a set of $3N - 1$ coordinates. The resulting integral is

$$Z_{NVT}(\Re) = \int \mathrm{d}\mathbf{u}\, |\det(\mathbf{J})|\, \delta[\Re(\mathbf{q}) - \Re] \exp\left[-\beta U(\mathbf{r}(\mathbf{u}))\right], \tag{3.7}$$

where $|\det(\mathbf{J})|$ is the absolute value of the determinant of the Jacobian matrix

$$\mathbf{J}(\mathbf{q}) = \begin{bmatrix} \frac{\partial r_1}{\partial \Re} & \frac{\partial r_1}{\partial q_2} & \cdots & \frac{\partial r_1}{\partial q_{3N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_{3N}}{\partial \Re} & \frac{\partial r_{3N}}{\partial q_2} & \cdots & \frac{\partial r_{3N}}{\partial q_{3N}} \end{bmatrix}. \tag{3.8}$$

$|\det(\mathbf{J})|$ accounts for the change in the volume element accompanying the transformation. A simple illustration of the Jacobian correction is provided by the $\Re$ defined as the angle between the bond of a rigid diatomic molecule and the z-axis, *e.g.*, coinciding with the normal to an interface. The variable change to be performed in this case is $(x_1, y_1, z_1, x_2, y_2, z_2) \to (x_1, y_1, \Re, x_2, y_2, z_2)$, where x_1, y_1, z_1, x_2, y_2, z_2 are the atomic Cartesian coordinates. $\Re = \arccos(z_2 - z_1)$ whereas the Jacobian is

$$\begin{aligned} &\mathbf{J}(x_1, y_1, \Re, x_2, y_2, z_2) \\ &= \begin{bmatrix} \frac{\partial x_1}{\partial x1} & \cdots & \frac{\partial x_1}{\partial \Re} & \cdots & \frac{\partial x_1}{\partial z_2} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial z_1}{\partial x1} & \cdots & \frac{\partial z_1}{\partial \Re} & \cdots & \frac{\partial z_1}{\partial z_2} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial z_2}{\partial x1} & \cdots & \frac{\partial z_2}{\partial \Re} & \cdots & \frac{\partial z_2}{\partial z_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sin\Re & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -\sin\Re & 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{3.9}$$

$|\det(\mathbf{J})| = \sin\Re$ expresses the $\Re$-dependence of the configurational volume,[37] *e.g.*, the fact that a larger number of configurations are possible for $\Re = \pi/2$ than for $\Re = 0$.

If $|\det(\mathbf{J})|$ is a function of $\Re$ only, PMFs that are free from the trivial volume dependence expressed by the Jacobian can be obtained by introducing a correction term in Equation 3.5:

$$\Delta\mathcal{F}(\Re) = -k_B T \ln \rho(\Re) + k_B T \ln|\det(\mathbf{J})|. \tag{3.10}$$

The calculation of PMFs from molecular simulations can be effectively achieved by advanced techniques that enhance the sampling of high-energy states along $\Re$. A class of such methods is based on introducing a bias in the evolution of the system in the form of an additional $\Re$-dependent energy term to the potential energy. The present Chapter focuses on two implementations of this approach: the umbrella sampling technique and the Wang–Landau method.

3.1 Umbrella Sampling

In biomolecular simulations, the widely used variant of the umbrella sampling technique of Torrie and Valleau[39] involves the addition to the Hamiltonian of a biasing potential, $w_i(\Re)$, to restrain the system to a narrow range of $\Re$. An example of an umbrella potential is the harmonic function centered around $\Re_i$:

$$w_i(\Re) = \frac{1}{2}K(\Re - \Re_i)^2. \tag{3.11}$$

where K is the force constant. N_w simulations are restricted to partly overlapping windows along $\Re$, each thoroughly sampling $\rho_i(\Re)$ in the neighborhood of a different $\Re_i$.[40] The biased $\rho_i(\Re)^B$ can be expressed as

$$\rho_i(\Re)^B = \frac{\int d\mathbf{r}\, \delta[\Re(\mathbf{r}) - \Re] \exp[-\beta U(\mathbf{r})] \exp[-\beta w_i(\Re)]}{\int d\mathbf{r}\, \exp[-\beta U(\mathbf{r})] \exp[-\beta w_i(\Re)]}. \tag{3.12}$$

To obtain the unbiased PMF through Equation 3.5, the bias is removed from $\rho_i(\Re)^B$ and the resulting unbiased $\rho_i(\Re)^U$ for the N_w windows are combined together. Dividing Equations 3.4 by Equation 3.12 yields

$$\frac{\rho_i(\Re)^U}{\rho_i(\Re)^B} = \frac{\Big\langle \exp[-\beta w_i(\Re)] \Big\rangle \int d\mathbf{r}\, \delta[\Re(\mathbf{r}) - \Re] \exp[-\beta U(\mathbf{r})]}{\int d\mathbf{r}\, \delta[\Re(\mathbf{r}) - \Re] \exp[-\beta U(\mathbf{r})] \exp[-\beta w_i(\Re)]}. \tag{3.13}$$

As previously mentioned, the integral in the denominator of Equation 3.13 is carried out over all coordinates but $\Re$.[41] Therefore, the exponential factor $\exp(-\beta w_i)$ can be taken out of the integral to obtain

$$\frac{\rho_i(\Re)^U}{\rho_i(\Re)^B} = \Big\langle \exp[-\beta w_i(\Re)] \Big\rangle \exp[\beta w_i(\Re)]. \tag{3.14}$$

The unbiased PMF_i for a single simulation exploring the neighborhood of $\Re_i$ is given by

$$\beta\mathrm{PMF}_i = -\ln\rho_i(\Re)^U = -\ln\rho_i(\Re)^B + \beta\mathcal{F}_i - \beta w_i(\Re). \tag{3.15}$$

where $\mathcal{F}_i$ is a constant quantifying the free energy contribution originating from the bias in the ith window.[40] $\mathcal{F}_i$ is defined as

$$\exp\left[-\beta\mathcal{F}_i\right] = \Big\langle \exp[-\beta w_i(\Re)] \Big\rangle = \int \mathrm{d}\Re \exp[-\beta w_i(\Re)]\rho(\Re)^U, \tag{3.16}$$

where $\rho(\Re)^U$ is the equilibrium unbiased distribution of the probability density of $\Re$. An effective route to calculate $\rho(\Re)^U$ from the N_w biased distributions is given by the weighted histogram analysis method (WHAM).[42] $\rho(\Re)^U$ can be expressed as a linear combination of the various $\rho_i(\Re)^U$:

$$\rho(\Re)^U = \sum_{i=1}^{N_w} C_i(\Re)\rho_i(\Re)^U = \sum_{i=1}^{N_w} C_i(\Re)\rho_i(\Re)^B \exp[\beta w_i(\Re) - \beta\mathcal{F}_i], \tag{3.17}$$

where the second equality is derived from Equations 3.14 and 3.16; whereas the coefficients $C_i(\Re)$ satisfy the normalization requirement, $\sum_{i=1}^{N_w} C_i(\Re) = 1$. The method of Lagrange multipliers provides the following expression for the $C_i(\Re)$ yielding the optimal estimate of $\rho(\Re)^U$:[43]

$$C_i(\Re) = \frac{n_i \exp[\beta\mathcal{F}_i - \beta w_i(\Re)]}{\sum_{j=1}^{N_w} n_j \exp[\beta\mathcal{F}_j - \beta w_i(\Re)]}, \tag{3.18}$$

where n_i is the number of uncorrelated $\Re$ values that are used to generate $\rho_i(\Re)^B$. Equations 3.17 and 3.18 can be succinctly rewritten as

$$\rho(\Re)^U = \frac{\sum_{i=1}^{N_w} n_i \rho_i(\Re)^B}{\sum_{j=1}^{N_w} n_j \exp[\beta\mathcal{F}_j - \beta w_i(\Re)]}. \tag{3.19}$$

Since the set of $\mathcal{F}_i$ values for the N_w simulations, $\{\mathcal{F}_i\}$, is unknown, Equations 3.19 and 3.16 are solved iteratively.

As an example from Paper IV, Listing 3.1 shows an implementation of the WHAM algorithm in the Python coding language, which was used to calculate the PMF for the translocation of a neutral histidine monopeptide across a lipid bilayer. Umbrella sampling MD simulations were performed with GROMACS[44] using the all-atom model presented in Paper IV. $\Re$ is defined as the center-of-mass separation between amino acid and membrane along the axis perpendicular to the bilayer. 36 umbrella sampling simulations were performed using harmonic potentials with $K = 10^3$ kJ mol^{-1} nm^{-2}. $\Re$ values were saved at regular time intervals to different data files which are loaded for each simulation at line 13 in Listing 3.1. The $\Re$ value at time-zero is the equilibrium distance $\Re_i$ and is used to calculate $w_i(\Re)$ (line 14). $\rho_i(\Re)^B$ is calculated for the N_w simulations as a histogram of the $\Re$ values with bin width 0.02 nm (line 18). $\{\mathcal{F}_i\}$ and $\rho(\Re)^U$ are initially set to zero (lines 20–21) and successively estimated by solving

Equations 3.19 and 3.16 self-consistently (lines 26–35). The iterative process is interrupted when the maximal difference between the $\mathcal{F}_i$ values of two successive cycles is smaller than 10^{-6} kJ mol^{-1} (line 26). The N_w biased distributions and the PMF calculated using the WHAM implementation of Listing 3.1 are shown in Figure 3.1.

```
 1 import numpy as np
 2 
 3 RT = 8.3145*0.310 # thermal energy, (kJ/mol)
 4 K = 1e3. # harmonic force constant, (kJ/mol/nm^2)
 5 Nw = 36 # number of simulations
 6 edges = np.arange(-0.61,3.7,0.02) # edges of the histograms
 7 rc = edges[:-1]+(edges[1]-edges[0])*0.5 # reaction coordinate
 8 wi = np.zeros(shape=(Nw,rc.size)) # set of biasing potentials
 9 rhoBi = np.zeros(shape=(Nw,rc.size)) # set of biased distrib.
10 
11 for i in range(Nw):
12     # loading data file of rc vs time
13     t,x = np.loadtxt(str(i)+'.dat',unpack=True)
14     wi[i] = 0.5*K*np.power(rc-x[0],2) # biasing potential vs rc
15     # the first 10 ns are considered as equilibration
16     x = x[t>1e4]
17     # distributions are generated as histograms
18     rhoBi[i], edges = np.histogram(x,bins=edges,normed=False)
19 
20 Fi = np.zeros(shape=(Nw,)) # free energy constants {Fi}
21 rhoU = np.zeros(shape=(rc.size,)) # unbiased distribution
22 prevFi = np.zeros(shape=(Nw,)) # {Fi} in previous iteration
23 change = 1 # maximal difference between previous and current {Fi}
24 ni = rhoBi.sum(axis=1)[:,np.newaxis] # number of rc values {ni}
25 
26 while change > 1e-6: # iterative procedure
27     num = (ni*rhoBi).sum(axis=0)
28     den = (ni*np.exp((Fi[:,np.newaxis]-wi)/RT)).sum(axis=0)
29     rhoU = np.divide(num, den) # Eq. 5.14
30     for i in range(Nw):
31         Fi[i] = np.trapz(rhoU*np.exp(-wi[i]/RT),rc) # Eq. 5.11
32     Fi = -RT*np.log(Fi)
33     Fi -= Fi.min()
34     change = np.fabs(prevFi-Fi).max()
35     prevFi = np.copy(Fi)
36 
37 PMF = -RT*np.log(rhoU)
```

Listing 3.1: Example of implementation of the WHAM algorithm written in the Python coding language for the calculation of a PMF from biomolecular simulations.

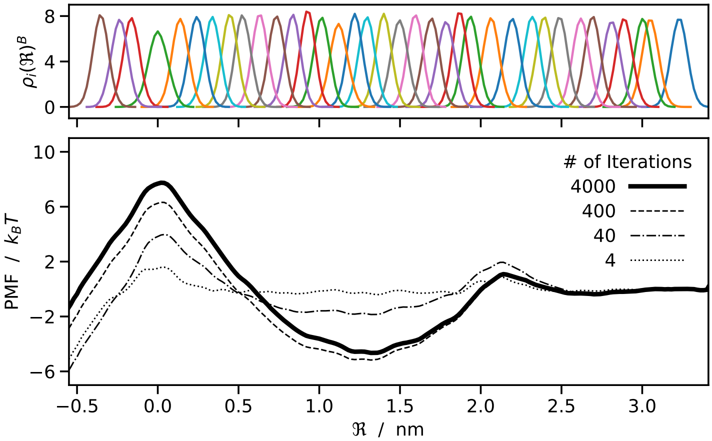

Figure 3.1: (*Top*) Probability density distributions, $\rho_i(\Re)^B$, of the reaction coordinate, $\Re$, from 36 umbrella sampling MD simulations. (*Bottom*) Potentials of mean force (PMFs) obtained by unbiasing and combining the various $\rho_i(\Re)^B$ spanning the whole range of $\Re$. The PMFs are calculated using the Python code of Listing 3.1 with 4000 (solid line), 400 (dashed line), 40 (dot-dashed line), and 4 (dotted line) iteration cycles.

3.2 Wang–Landau methods

The enhanced sampling techniques falling under the category of Wang–Landau methods are characterized by the optimization of a biasing potential aimed at obtaining equal probabilities for all points on $\Re$. The bias is an *a priori* unknown function of $\Re$, which is progressively modified to achieve uniform distributions of $\Re$. The implementation of Wang and Landau[45,46] represents a refinement of previous methods,[47–50] successfully applied to study, *e.g.*, phase transitions of classical spin systems[47,48] and of the Lennard–Jones fluid.[49] Here, the method is applied to the calculation of free energy profiles and surfaces.[50–52] As opposed to the multistage umbrella sampling technique described above, Wang–Landau methods require a single simulation with an undetermined biasing potential $w(\Re)$, which is initially set to zero. Throughout the simulation, $w(\Re)$ is continuously updated to penalize the sampling of the explored regions of $\Re$. Concurrently, the biased distribution, $\rho(\Re)^B$, is calculated and periodically monitored. $\rho(\Re)^B$ and $w(\Re)$ are generated as histograms, however, the former is built by unit increments whereas the bin counts of the latter are increased by f. If the degree of uniformity of $\rho(\Re)^B$ meets a given criterion, f is reduced, $\rho(\Re)^B$ is set to zero, and the criterion for uniformity is made more stringent. Convergence is reached when $\rho(\Re)^B$ shows an insignificant dependence on $\Re$. For $\rho(\Re)^B = C$, where C is an arbitrary constant, Equation 3.15 can be rewritten as

$$\beta\mathrm{PMF} = -\ln\left[C\left\langle\exp[-\beta w(\Re)]\right\rangle\right] - \beta w(\Re) = C - \beta w(\Re), \tag{3.20}$$

therefore, the converged $w(\Re)$ is directly related to the PMF. The criterion to assess the flatness of $\rho(\Re)^B$ can vary from implementation to implementation. A common requirement is that the ratio between the minimum and the mean values of $\rho(\Re)^B$ should be sufficiently close to one.[43] The initial value of f and of the scaling factor used for the successive downward adjustments are also implementation- and system-dependent. In the beginning of the simulation, a relatively large value of f is beneficial to overcome free energy barriers along $\Re$. Once the whole $\Re$-range has been visited, f needs to be downscaled to progressively reduce the fluctuations in $w(\Re)$ and capture the fine details of the PMF.

In MC implementations of Wang–Landau methods, trial moves are accepted according to the following criterion:[51]

$$\mathrm{acc}(\mathbf{r}_{old} \rightarrow \mathbf{r}_{new}) = \min\Big[1,\ \exp\left[-\beta U(\mathbf{r}) - \beta w(\Re)\right]\Big]. \tag{3.21}$$

Since the Hamiltonian is continuously modified throughout the simulation, detailed balance is not strictly observed.[43,51] Therefore, to sample properties as ensemble averages, it is necessary to perform long Wang–Landau simulations biased with the converged $w(\Re)$, and with $f = 0$.[43]

Wang–Landau methods offer the possibility for parallel computing implementations of the MC algorithm.[52,53] N_p MC simulations can be initialized with different random number seeds and carried out in parallel on separate processors. Periodically, $w_i(\Re)$ and $\rho_i(\Re)^B$ are collected from the N_p processes and the update of f is attempted. If at least one $w_i(\Re)$ meets the flatness criterion, each $w_i(\Re)$ is updated to the average $\sum_{i=1}^{N_p} w_i(\Re)/N_p$ and the single simulations proceed independently until the subsequent attempt. All the other aspects of the simulation are identical as in the serial implementation.

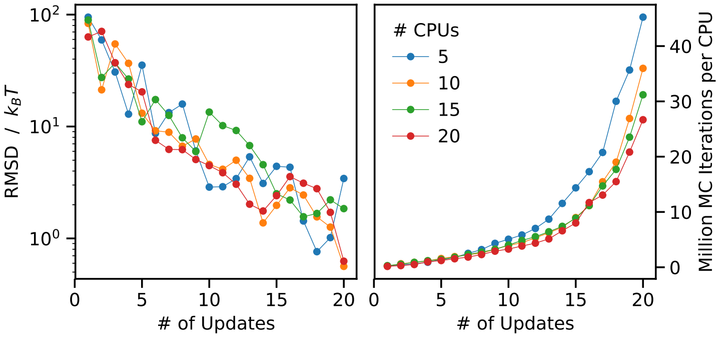

Figure 3.2: (*Left*) Root-mean-square deviation (RMSD) of intermediate PMFs with respect to the fully converged PMF as a function of the number of updates of `f` and `sampled` for simulations performed with 5 (blue), 10 (orange), 15 (green), and 20 (red) CPUs. (*Right*) Number of MC iterations as a function of the number of updates of `f` and `sampled` for simulations performed with 5 (blue), 10 (orange), 15 (green), and 20 (red) CPUs.

```
 1 void update(const vector<double> &coord) override {
 2     using namespace Faunus::MPI;
 3     double uold = penalty[coord]; // store current bias
 4     if (++cnt % nupdate == 0 && f > 0) {
 5         int min = histo.minCoeff(); // smallest count
 6         MPI_Barrier(mpi.comm);
 7         // gather counts of least sampled bin from all processes
 8         MPI_Allgather(&min, 1, MPI_INT, minCounts.data(), 1,
 9                 MPI_INT, mpi.comm);
10         // check if the largest of the smallest counts > sampled
11         if ( minCounts.maxCoeff() > sampled ) {
12             MPI_Gather(penalty.data(), penalty.size(),
13                 MPI_DOUBLE,buffer.data(), penalty.size(),
14                 MPI_DOUBLE, 0, mpi.comm);
15             if (mpi.isMaster()) {
16                 penalty.setZero(); // set penalty to zero
17                 // average penalty over all processes
18                 for (int i = 0; i < mpi.nproc(); i++)
19                     penalty += Eigen::Map<Eigen::MatrixXd>(
20                         buffer.data() + i*penalty.size(),
21                         penalty.rows(), penalty.cols() )
22                         / double(mpi.nproc());
23                 penalty = penalty.array() - penalty.minCoeff();
24             }
25             MPI_Bcast(penalty.data(), penalty.size(),
26                     MPI_DOUBLE, 0, mpi.comm);
27             histo.setZero(); // set histo to zero
28             f = f * scale; // decrease increment
29             // increase sampled
30             sampled = ceil( sampled / scale );
31         }
32     }
33     histo[coord]++; // update probability density histogram
34     penalty[coord] += f; // update bias histogram
35     udelta += penalty[coord] - uold; // energy change
36 }
```

Listing 3.2: Function to update the biasing potential in the implementation of the Wang–Landau method in the sofware package Faunus.

Listing 3.2 shows an excerpt from the code of the software package Faunus.[54] Faunus is written in the C++ programming language and uses the message passing library standard MPI. The function `update` takes as an argument the visited value of the reaction coordinate, `coord` (line 1), and increases the counts for the corresponding bins in the $\rho_i(\mathfrak{R})^B$ and $w_i(\mathfrak{R})$ histograms (lines 33–34), which are implemented as matrices named `histo` and `penalty`, respectively. If the number of MC steps is a multiple of `nupdate` (line 4), the processes exchange the counts of the least visited element of `histo` and collect all the values in the vector `minCounts` (lines 8–9). If the maximum element of `minCounts` is larger than a threshold `sampled` value, the manager process gathers the N_p `penalty` matrices (lines 12–14) and computes the average (lines 18–22). The average `penalty` is then redistributed by the manager to the remainder $N_p - 1$ processes (lines 25–26). Subsequently (lines 27–30), the `histo` of each process is set to zero, whereas the increment `f` and the `sampled` value are multiplied and divided by

$0 <$`scale`< 1, respectively.
An example of the Wang–Landau method is given by the calculation of the PMF of translocation of a single amino acid across a lipid bilayer. MC simulations are performed with Faunus using the coarse-grained model presented in Paper IV. The initial values of `f` and `sampled` are 0.05 k_BT and 3, respectively, whereas `scale` is set to 0.7. The left hand-side of Figure 3.2 shows the progress of the convergence of the PMF as a function of the number of updates of `f` and `sampled`, for simulations performed with 5, 10, 15, and 20 parallel processes (CPUs). The degree of convergence is quantified as the root-mean-square deviation (RMSD) of the intermediate PMF with respect to the fully converged PMF. The degree of convergence increases steadily as `f` is downscaled irrespective of the number of CPUs. However, the right hand-side of Figure 3.2 indicates that the increase in the number of processes decreases the number of MC iterations required to perform the same number of updates of `f`.

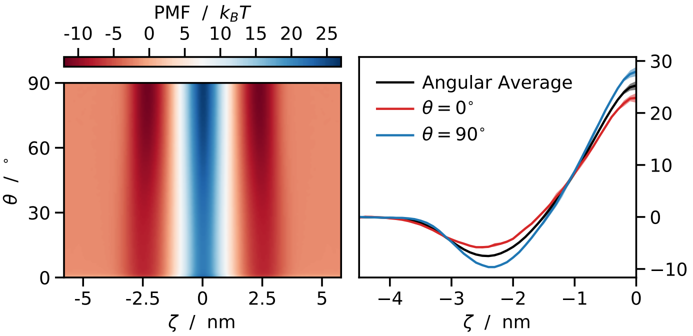

Figure 3.3: (*Left*) 2D PMF of translocation of a rigid linear three-bead molecule across a lipid bilayer calculated as a function of two reaction coordiantes, ζ and θ. (*Right*) 1D PMFs obtained from the 2D PMF by angular averaging (black), and by selecting constant-angle profiles for $\theta = 0°$ (red) and $\theta = 90°$ (blue). Shaded areas represent the uncertainties, estimated as the absolute difference between the two symmetrical halves of the 1D PMFs.

For the simulated system consisting of 487 particles, 30 million MC iterations per CPU are performed in approximately 30 minutes using 10-core 2.3 GHz processors. The Faunus implementation of the Wang–Landau method is suitable for the calculation of PMFs as a function of one or two reaction coordinates. The left hand-side of Figure 3.3 shows a 2D PMF obtained with Faunus using the coarse-grained model presented in Paper IV. The PMF pertains to the translocation of a linear rigid three-bead molecule across a lipid bilayer and is a function of a distance, ζ, and an angular coordinate, θ. ζ is the z-component of the center-of-mass separation between the molecule and the bilayer whereas θ is the angle between the principal axis of the molecule and the z-axis, coinciding with the normal to the membrane. The calculation of the 2D PMF required the Jacobian correction of Equation 3.10 applied along the angular coordinate, with

$|\det(\mathbf{J})| = \sin\theta$. The right hand-side of Figure 3.3 shows free energy profiles as a function of ζ obtained from the 2D PMF. Comparison of the angular average of the 2D PMFs with the profiles for $\theta = 0°$ and $\theta = 90°$ indicates that the molecule is preferentially oriented parallel to the z-axis at the midplane ($\zeta = 0$); whereas it favors the perpendicular orientation in the region of the lipid head groups.

Chapter 4

Quartz Crystal Microbalance with Dissipation Monitoring

The quartz crystal microbalance with dissipation monitoring (QCM-D) is a surface-sensitive technique based on the measurement of changes in oscillation of a quartz crystal in response to changes in the surrounding environment. It was in the late 1960s that the name microbalance was used for the first time to refer to this device.[55] Since then, the sensitivity of the sensor has greatly improved and the range of applications of the quartz crystal microbalance (QCM) have expanded. The sensing unit of QCM is the quartz crystal resonator (QCR), a piezoelectric disk of crystalline quartz of thickness d_c and radius R. Typical dimensions are $d_c = 330$ μm and $R = 7.5$ mm.[56] The resonance of a QCR occurs in a very narrow band of frequencies. For instance, the half-bandwidth, Γ_r, of a QCR with resonance frequency, f_r, of 5 MHz is typically smaller than 25 Hz.[56] The two sides of QCR crystals used for QCM-D measurements are shown in Figure 4.1. In the QCM-D instrument Q-sense E4 from Qsense (Göteborg, Sweden), the QCR is located in a measurement chamber where it is held horizontally on top of an O-ring. Each side is in contact with an electrode, generally made of gold. The upper electrode, which is exposed to the sample, has a larger surface than the lower one, and can be coated with various materials, *e.g.*, silica.[56,57]

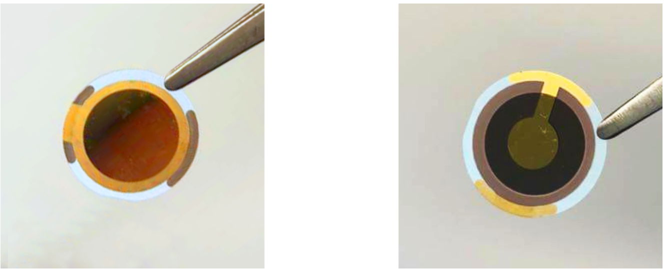

Figure 4.1: Pictures of QCM-D crystals from Qsense (Göteborg, Sweden) with uncoated gold electrodes taken from the front side (*Left*) and from the back side (*Right*).

4.1 Operating Principle

When a static voltage is applied, the QCR is deformed to a very small extent. However, by applying an AC voltage of frequency equal to f_r, more pronounced deformations are induced and a sizable oscillating electric current can be extracted by the electrodes. For an AT-cut QCR, the applied voltage mainly induces shear deformations, wherein the vibrations occur tangentially to the crystal surface and result in the generation of standing plane waves propagating in the direction normal to the surface.[56] Figure 4.2 illustrates the displacement patterns for the first six harmonics of the resonance frequency for a bare QCR and for a QCR covered with a rigid film. The wavelength of the shear waves is twice as large as the thickness of the QCR. The antinodes are located at the surface of the QCR, whereas, within the crystal, as many nodal planes are present as the overtone order, n. Since the two faces of the deformed QCR are oppositely charged, only odd n can be excited.[58] When a rigid film is deposited on the QCR, the increase in the thickness of the vibrating disk determines an increase in λ and a decrease in the frequency of the oscillations $f = c/\lambda$, where c is the speed of sound. The Sauerbrey equation[55] describes this phenomenon in terms of the change in frequency, Δf, upon deposition of a thin rigid film of mass m_f

$$m_f = -C\frac{\Delta f}{n}, \tag{4.1}$$

where C is a mass sensitivity constant.[57] Equation 4.1 is strictly valid for rigid samples in the gas phase and accounts for the use of QCM as a gravimetric technique. Due to the small Γ_r of the resonance frequency band, Δf can be measured with high precision and the gravimetric sensitivity is of the order of

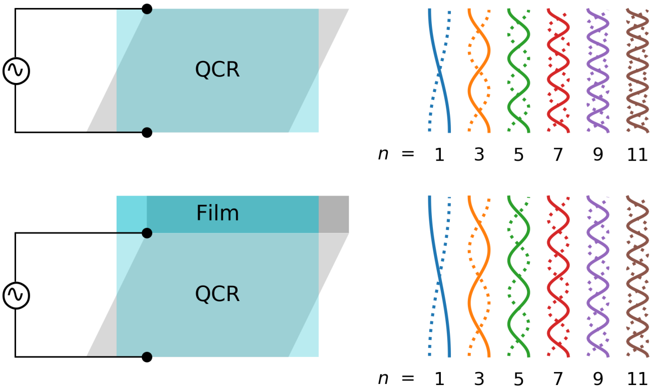

Figure 4.2: (*Left*) Schematic illustration of a QCR excited by an oscillating voltage. (*Right*) Corresponding displacement patterns for the first 6 odd overtone orders, n. (*Top*) The wavelength of the standing shear waves generated by the QCR operating at the fundamental f_r ($n = 1$) is twice as long as the thickness of the crystal. (*Bottom*) The wavelength increases upon the deposition of a rigid film on the surface of the QCR.

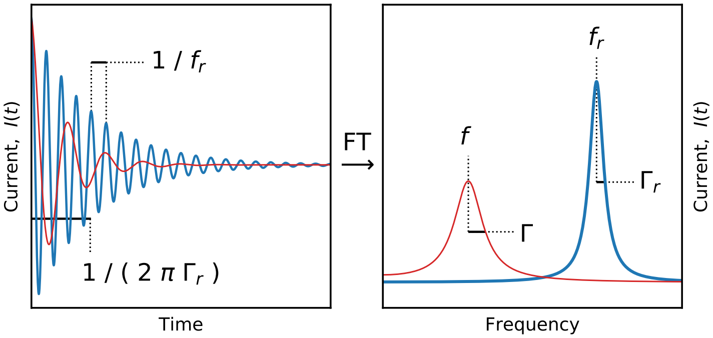

Figure 4.3: (*Left*) Time-domain signal for the current drawn by the electrodes of a bare QCR (blue) and of a QRC covered with a soft film (red) as obtained from Equation 4.2. (*Right*) Corresponding frequency-domain signals obtained by Fourier transform (FT). The deposited film increases the thickness of the resonator and decreases the frequency of the oscillations ($f - f_r < 0$). The viscoelastic forces acting on the soft film result in the rapid decay of the time-domain signal and in the broadening of the frequency-domain band.

1 ng cm^{-2}.[56] In water, a shear wave generated by a QCR oscillating at 5 MHz has a penetration depth of 250 μm.[56] QCM-D owes its surface specificity to this rapid decay of the shear waves in the liquid medium, which limits the material sensed by the QCR to a surface layer of thickness 1 μm.[56]

If the film is deformable or if the rigid adsorbate protrudes into the liquid phase and is acoustically coupled with the solvent molecules, frictional viscous forces considerably damp the amplitude of the oscillations of the QCR. In that case, Equation 4.1 does not apply, and the viscoelastic properties of the film have to be taken into account. To disentangle the effects of mass and viscosity in QCM-D measurements, the QCR is periodically excited by a radio-frequency (RF) pulse matching f_r. The current drawn by the electrodes, $I(t)$, has an initial amplitude $I(0)$ which freely decays in time, t. The dampened signal can be modeled as a cosine function of frequency f, decaying exponentially with decay constant $1/(2\pi\Gamma)$:[57,59]

$$I(t) = I(0)\,\exp\left(-2\pi\Gamma t\right)\,\cos\left(2\pi f t\right). \tag{4.2}$$

The current signal is digitally recorded and fitted to Equation 4.2 to simultaneously measure Γ and f. The RF pulses are applied approximately every second, typically alternating between the frequencies of the first 6 odd overtones.[57] In QCM-D experiments, the time evolutions of Γ and f are monitored as the sample flows through a thermostatted measurement chamber. Figure 4.3 illustrates the signal of Equation 4.2 in the time and frequency domains for a bare QCR (Γ_r and f_r), and for a QCR covered with a viscoelastic film (Γ and f). Γ is related to the energy dissipation D by[59]

$$D = \frac{2\Gamma}{f}, \tag{4.3}$$

which is proportional to the ratio between the energy dissipated during one oscillation, E_d, and the energy stored in the deposited film after the excitation, E_s,[57]

$$D = \frac{E_d}{2\pi E_s}. \tag{4.4}$$

4.2 Modeling

Generally, the length scale of inhomogeneities of the media surrounding the QCR is considerably smaller than the penetration depth of the shear waves. Therefore, continuum models are commonly used to analyze QCM-D signals in terms of energy storage and dissipation processes. Within the framework of continuum mechanics, the response of the viscoelastic film to the stress applied by the oscillating QCR is expressed by the complex frequency-dependent shear modulus

$$\tilde{G}(f) = G'(f) + iG''(f), \tag{4.5}$$

where $G'(f)$ is the storage modulus and $G''(f)$ is the loss modulus, which is related to the shear elastic viscosity, $\eta(f)$, by[60]

$$G''(f) = 2\pi f \eta(f). \tag{4.6}$$

In the literature, $G'(f)$ is often termed the shear elastic modulus, $\mu(f)$. A widely used model for viscoelastic homogeneous films in QCM-D data analysis is the one-layer Voigt model[60,61]. The QCR is treated as a harmonic oscillator covered by one viscoelastic layer. The bulk solution is considered as a semi-infinite Newtonian fluid of density ρ_s and frequency-independent viscosity η_s. Newtonian fluids are characterized by relaxation rates that are considerably faster than the f at which the QCR is operating; therefore, $G'(f) \ll G''(f)$ and $G'(f)$ can be neglected[61]. No-slip conditions are assumed to hold at the interfaces.[57,60,61] Beside the viscoelastic parameters $\mu(f)$ and $\eta(f)$, the layer is described by density, ρ_f, and thickness, d_f. In the viscoelastic regime, $\mu(f)$ and $\eta(f)$ vary monotonically with f and can be approximated by the following power laws[61]

$$\begin{aligned} \mu(f) &= \mu_0 \left(f/f_0\right)^{\alpha_\mu}, \\ \eta(f) &= \eta_0 \left(f/f_0\right)^{\alpha_\eta}. \end{aligned} \tag{4.7}$$

μ_0, η_0, and f_0 are reference values whereas α_μ and α_η are the power law exponents. The Voigt model provides the equations to relate Δf and ΔD to $G'(f)$ and $G''(f)$. Global fitting procedures are performed on the time-dependent data of Δf and ΔD of several overtones. Knowledge of the parameters ρ_s, η_s, and ρ_f is a prerequisite, whereas μ_0, η_0, α_μ, α_η, and d_f are obtained from the fit. Finally, the mass of the film can be obtained as

$$m_f = \pi R^2 \rho_f d_f. \tag{4.8}$$

For many samples, *e.g.*, protein solutions, the viscoelastic film is highly solvated and a substantial amount of solvent molecules are acoustically coupled to the

QCR. In that case, d_f has to be considered as an effective hydrodynamic thickness and m_f as wet mass[57].

QCM-D measurements are used in Paper III to investigate the adsorption of amyloid beta peptides and fragmented fibrils to supported lipid bilayers. The Δf and ΔD are modeled by treating the supported lipid bilayer as a rigid film and using the one-layer Voigt model described in this Chapter to obtain the viscoelastic properties and the wet mass of the fibril film.

Chapter 5

Small-angle X-ray Scattering

Small-angle X-ray scattering (SAXS) is a powerful experimental technique for the structural characterization of solutions of particles in the nanometer range. Exploiting the variation of the electron density throughout a sample volume, SAXS provides information on the size and shape of the particles, as well as on the spatial correlations between the particles.[62]

Figure 5.1 shows a schematic illustration of a SAXS experiment. A beam of collimated X-ray photons is incident on the sample. The scattered radiation is collected by a 2D area detector, while the transmitted radiation is absorbed by a beamstop to protect the detector.[63] The oscillating electric field of the incident electromagnetic waves is denoted by

$$\mathbf{E}(t) = \hat{e}E\exp[i(\mathbf{k}\cdot\mathbf{r} - \nu t)], \tag{5.1}$$

where ν is the frequency, $\hat{e}$ is the polarization versor and $\mathbf{k} = \hat{k}k$ is the wave vector defining the propagation direction, $\hat{k} \perp \hat{e}$. Synchrotron X-rays have wavelength $\lambda = 2\pi/k \approx 0.1$ nm which is considerably larger than the classical electron radius, r_e =2.818 fm. Therefore, the interaction between X-rays and

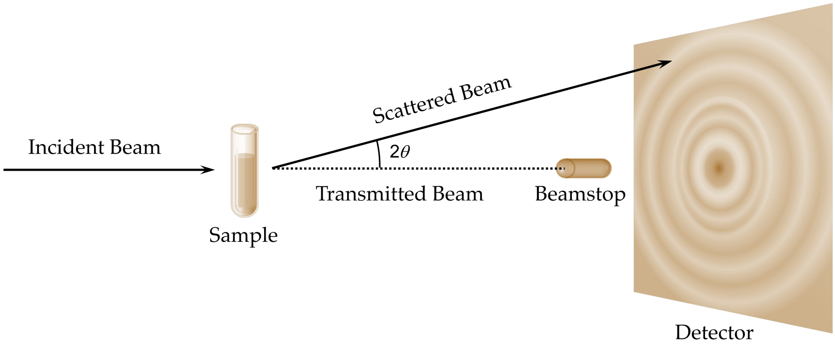

Figure 5.1: Schematic representation the experimental setup of a small-angle X-ray scattering measurement.

electrons can be treated as classical elastic Thomson scattering.[64] $\mathbf{E}(t)$ exerts a Lorentz force on a free electron, which makes the latter oscillate with acceleration

$$\mathbf{a} = \frac{\mathbf{F}(t)}{m_e} = -e\frac{\mathbf{E}(t)}{m_e}, \tag{5.2}$$

where m_e is the mass of the electron and e is the elementary charge. The oscillating free electron emits an electromagnetic wave of the same λ and a phase shift of π with respect to the incident wave. Since all incoming X-ray photons are in-phase, the scattering process is characterized by a well-defined phase relationship between incident and scattered waves.[64] The amplitudes of the scattered X-rays result in an interference pattern on the detector. The amplitude of a scattered wave is $\exp(i\phi)$, where ϕ is the phase angle. Figure 5.2 schematically illustrates ϕ for an electron located at position $\mathbf{r}$ with respect to the origin. The angle between the scattered and incident waves is known as the scattering angle. It is usually denoted by 2θ, and in SAXS experiments it is typically smaller than $5°$. The difference in optical path length of the electron at $\mathbf{r}$ with respect to the origin is $\delta_2 - \delta_1 = \mathbf{r}\cdot\hat{k_0} - \mathbf{r}\cdot\hat{k}$ where $\hat{k_0}$ and $\hat{k}$ are the propagation versors of the incident and the scattered wave, respectively. Therefore, the phase angle is

$$\phi = (\delta_2 - \delta_1)k = -\mathbf{q}\cdot\mathbf{r}, \tag{5.3}$$

where $\mathbf{q}$ is the scattering vector, $\mathbf{q} = \mathbf{k} - \mathbf{k_0}$, of magnitude $q = \frac{4\pi}{\lambda}\sin\theta$. The total amplitude, $F(\mathbf{q})$, of radiation at an angle 2θ is obtained by integrating the complex amplitude of the scattered wave emitted by a single electron, $\exp(-i\mathbf{q}\cdot\mathbf{r})$, over the sample volume:

$$F(\mathbf{q}) = \int \mathrm{d}\mathbf{r}\rho(\mathbf{r})\exp(-i\mathbf{q}\cdot\mathbf{r}), \tag{5.4}$$

where $\rho(\mathbf{r})$ is the continuous electron density. The scattering intensity, $I(\mathbf{q})$, is defined as the squared norm of the total amplitude:[64]

$$I(\mathbf{q}) = |F(\mathbf{q})|^2 = \int \mathrm{d}\mathbf{r}\int \mathrm{d}\mathbf{r}'\rho(\mathbf{r})\rho(\mathbf{r}')\exp(-i\mathbf{q}\cdot\mathbf{r})\exp(i\mathbf{q}\cdot\mathbf{r}'). \tag{5.5}$$

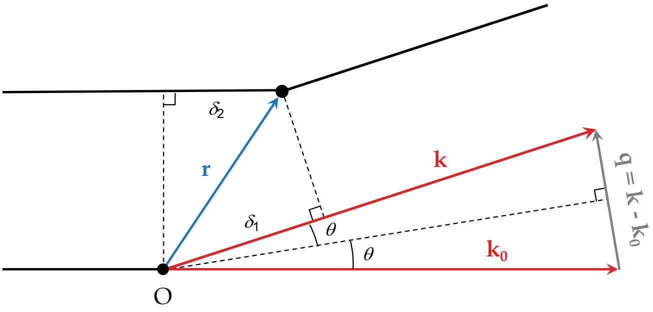

Figure 5.2: Schematic illustration of the difference in optical path length of the electron at $\mathbf{r}$ with respect to an electron at the origin.

By choosing $\mathbf{r}' = \mathbf{r} + \mathbf{R}$, the intensity is expressed in terms of interelectron vector distances $\mathbf{R}$,

$$I(\mathbf{q}) = V \int \mathrm{d}\mathbf{R} \langle \rho(\mathbf{r}) \rho(\mathbf{r} + \mathbf{R}) \rangle \exp(-i\mathbf{q} \cdot \mathbf{R}), \tag{5.6}$$

where V is the sample volume, whereas $\langle \rho(\mathbf{r}) \rho(\mathbf{r} + \mathbf{R}) \rangle$ is the spatial autocorrelation function of the electron density.[64] For isotropic and non-crystalline samples, $\langle \rho(\mathbf{r}) \rho(\mathbf{r} + \mathbf{R}) \rangle$ depends only on the magnitude R of the vector $\mathbf{R}$, whereas $I(\mathbf{q})$ can be integrated in spherical coordinates:[64]

$$\begin{aligned} I(\mathbf{q}) &= V \int_0^\infty \mathrm{d}R\, R^2 \langle \rho(\mathbf{r}) \rho(\mathbf{r} + \mathbf{R}) \rangle \int_0^{2\pi} \mathrm{d}\gamma \int_0^\pi \mathrm{d}\alpha \sin\alpha \exp(-iqR\cos\alpha) \\ &= V \int_0^\infty \mathrm{d}R\, 4\pi R^2 \langle \rho(\mathbf{r}) \rho(\mathbf{r} + \mathbf{R}) \rangle \frac{\sin(qR)}{qR} = I(q), \end{aligned} \tag{5.7}$$

where α is the angle between $\mathbf{q}$ and $\mathbf{R}$, and $\sin(qR)/qR$ is known as the Debye factor. Since the scattering intensity depends only on the magnitude of the scattering vector, the 2D scattering pattern recorded on the detector has cylindrical symmetry with respect to the incident X-ray beam along $\mathbf{k_0}$.

5.1 Form and Structure Factors

A particle solution can be treated as a binary system where the particles are immersed in a background of solvent and co-solutes.[64] The $I(q)$ of the sample is approximated to the sum of the scattering intensities originating from the particles, $I_P(q)$, and from the background, $I_B(q)$. $I_P(q)$ is thereby obtained by subtracting $I_B(q)$ from $I(q)$,

$$I_P(q) = I(q) - I_B(q). \tag{5.8}$$

where $I_B(q)$ is measured from a solution prepared to accurately match the composition of the medium in which the particles are dissolved.[63] In this kind of SAXS experiments, $F(\mathbf{q})$ is the Fourier transform of the excess electron density,

$$\Delta\rho(\mathbf{r}) = \rho_P(\mathbf{r}) - \overline{\rho_B}, \tag{5.9}$$

where $\rho_P(\mathbf{r})$ is the electron density of the particles and $\overline{\rho_B}$ is the average electron density of the background. The success of the measurement relies on the contrast, *i.e.*, $\overline{\Delta\rho} = \overline{\rho_P} - \overline{\rho_B}$, as well as on the accurate match between the compositions of the background solution and the medium surrounding the particles in the sample.[63]

For a system of N identical particles in a volume V, the scattering intensity arising from the particles can be expressed as a sum of contributions from the single particles

$$I_P(q) = \frac{N}{V} \langle |F(\mathbf{q})|^2 \rangle + \frac{1}{V} \left\langle \sum_{j=1}^{N} \sum_{k=1 \neq j}^{N} F_j(\mathbf{q}) F_k^*(\mathbf{q}) \exp[-i\mathbf{q} \cdot (\mathbf{r_j} - \mathbf{r_k})] \right\rangle, \tag{5.10}$$

where $F(\mathbf{q})$ is the Fourier transform of $\Delta\rho(\mathbf{r})$ extended over the particle volume, V_P,

$$F(\mathbf{q}) = \int_{V_P} \mathrm{d}\mathbf{r}\Delta\rho_j(\mathbf{r}) \exp(-i\mathbf{q}\cdot\mathbf{r}). \tag{5.11}$$

where $F_j(\mathbf{q})$ and $F_k(\mathbf{q})$ are the $F(\mathbf{q})$s of single particles with mass centers located at r_j and r_k, respectively. If the particle is non-spherical, $F(\mathbf{q})$ depends on the orientation, Ω, of the particle. In the decoupling approximation, the orientation of a particle is considered to be independent of its position and of the configuration of the other particles.[62] The configurational average in Equation 5.10 can be divided into orientational and positional averages:[64]

$$I_P(q) = \frac{N}{V}\langle|F(\mathbf{q})|^2\rangle_\Omega + \frac{|\langle F(\mathbf{q})\rangle_\Omega|^2}{V}\sum_{j=1}^{N}\sum_{k=1\neq j}^{N}\left\langle \exp[-i\mathbf{q}\cdot(\mathbf{r_j}-\mathbf{r_k})]\right\rangle_{\mathbf{r}}. \tag{5.12}$$

The spherically averaged scattering intensity of a single particle is readily obtained from Equation 5.7:

$$\langle|F(\mathbf{q})|^2\rangle_\Omega = (V_P\overline{\Delta\rho})^2\int_0^{D_{max}} \mathrm{d}R\ P(R)\frac{\sin(qR)}{qR} \tag{5.13}$$

where D_{max} is the largest separation between two points inside the particle. $P(R)$ is the pair distance distribution function,

$$P(R) = \langle\Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}+\mathbf{R})\rangle\frac{4\pi R^2}{V_P\overline{\Delta\rho}^2}, \tag{5.14}$$

where $\langle\Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}+\mathbf{R})\rangle$ is the spatial autocorrelation function of the excess electron density inside V_P. $P(R)$ is normalized as

$$P(R) = \int_0^{D_{max}} \mathrm{d}R\ P(R) = 1. \tag{5.15}$$

Equation 5.12 can be rewritten as

$$\begin{aligned} I_P(\mathbf{q}) &= n_P\langle|F(\mathbf{q})|^2\rangle_\Omega\left\{1+\frac{1}{N}\frac{|\langle F(\mathbf{q})\rangle_\Omega|^2}{\langle|F(\mathbf{q})|^2\rangle_\Omega}\sum_{j=1}^{N}\sum_{k=1\neq j}^{N}\langle\exp[-i\mathbf{q}\cdot(\mathbf{r_j}-\mathbf{r_k})]\rangle_{\mathbf{r}}\right\} \\ &= n_P(V_P\overline{\Delta\rho})^2P(q)S_{eff}(\mathbf{q}), \end{aligned} \tag{5.16}$$

where n_P is the number density of the solution and $S_{eff}(\mathbf{q})$ is the effective structure factor.[64] For an ideal solution, $S_{eff}(q) = 1$ and the scattering intensity is

$$I_P^{id.}(q) = \frac{N}{V}\langle|F(\mathbf{q})|^2\rangle_\Omega = n_P(V_P\overline{\Delta\rho})^2P(q) \tag{5.17}$$

where $P(q)$ is the form factor:[64]

$$P(q) = \int_0^{D_{max}} \mathrm{d}R\ P(R)\frac{\sin(qR)}{qR}. \tag{5.18}$$

The effective structure factor is defined as

$$S_{eff}(\mathbf{q}) = 1 + \frac{|\langle F(\mathbf{q})\rangle_\Omega|^2}{\langle |F(\mathbf{q})|^2\rangle_\Omega}(S(q) - 1), \tag{5.19}$$

where $S(\mathbf{q})$ is the structure factor:

$$S(\mathbf{q}) = \frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\langle \exp[-i\mathbf{q}\cdot(\mathbf{r_j} - \mathbf{r_k})]\rangle. \tag{5.20}$$

For a monodisperse solution of spherical particles $|\langle F(\mathbf{q})\rangle_\Omega|^2 = \langle |F(\mathbf{q})|^2\rangle_\Omega$ and

$$I_P(q) = n_P(V_P\overline{\Delta\rho})^2 P(q)S(q). \tag{5.21}$$

5.2 The Guinier Approximation

The form factor, $P(q)$, yields information on the overall shape and size of the particle. From $P(q)$ we can obtain the radius of gyration of the particle, R_g. The series expansion of $\sin(qR)/qR$ for small q gives

$$\frac{\sin(qR)}{qR} = 1 - \frac{(qR)^2}{6} + \mathcal{O}(qR)^4. \tag{5.22}$$

We can write $P(q)$ for small q values as

$$P(q) \approx \int_0^{D_{max}} \mathrm{d}R\ P(R) - \int_0^{D_{max}} \mathrm{d}R\ P(R)\frac{(qR)^2}{6} = 1 - \frac{(qR_g)^2}{3}, \tag{5.23}$$

where R_g is the radius of gyration for a particle of uniform density

$$R_g^2 = \frac{1}{2}\int_0^{D_{max}} \mathrm{d}R\ R^2 P(R). \tag{5.24}$$

Using the series expansion $e^{x^2} \approx 1 - x^2$, we derive the Guinier approximation[63]

$$\ln P(q) \approx -\frac{(qR_g)^2}{3}. \tag{5.25}$$

For a sample of low particle concentration, $\ln I(q)$ vs. q^2 can be fitted to

$$\ln I_P^{id.}(q) \approx \ln I_P^{id.}(0) - \frac{(qR_g)^2}{3}. \tag{5.26}$$

For globular proteins, Equation 5.26 is accurate for $q < 1.3/R_g$.[63] The slope gives an estimate of R_g, whereas the intercept is $I_P^{id.}(q)$ extrapolated at $q = 0$. Given the partial specific volume of the particle, $\overline{v}$, the mass concentration of the sample, c_P, and $\overline{\Delta\rho}$, the molecular weight of the particle, M_r, is readily obtained from $I_P^{id.}(0)$.[63] According to Equation 5.17,

$$I_P^{id.}(0) = n_P(V_P\overline{\Delta\rho})^2 = \frac{c_P(\overline{v}\overline{\Delta\rho})^2 M_r}{N_A}, \tag{5.27}$$

where N_A is the Avogadro constant.
The Guinier analysis was used to estimate the radii of gyration and the molecular weights of the decapeptides studied in Paper V, as shown in Figure S1 in the Supporting Information.

5.3 Structure Factor and Radial Distribution Function

The structure factor

$$S(\mathbf{q}) = \sum_{j=1}^{N} \langle \exp[-i\mathbf{q} \cdot (\mathbf{r_1} - \mathbf{r_j})] \rangle \tag{5.28}$$

can be factorized into a self part, $j = 1$, and $N - 1$ distinct parts

$$\begin{aligned} S(\mathbf{q}) &= \langle \exp[-i\mathbf{q} \cdot (\mathbf{r_1} - \mathbf{r_1})] \rangle + \sum_{j=2}^{N} \langle \exp[-i\mathbf{q} \cdot (\mathbf{r_1} - \mathbf{r_j})] \rangle \\ &= 1 + (N-1) \langle \exp[-i\mathbf{q} \cdot (\mathbf{r_1} - \mathbf{r_2})] \rangle. \end{aligned} \tag{5.29}$$

In classical statistical thermodynamics, the ensemble average of $\exp[-i\mathbf{q} \cdot (\mathbf{r_1} - \mathbf{r_2})]$, where $\mathbf{r_1}$ and $\mathbf{r_2}$ are the positions of the center of mass of two distinct particles, reads as

$$\langle \exp[-i\mathbf{q} \cdot (\mathbf{r_1} - \mathbf{r_2})] \rangle = \frac{\int d\mathbf{r_1} ... d\mathbf{r_N} \exp[-i\mathbf{q} \cdot (\mathbf{r_1} - \mathbf{r_2})] \exp[-\beta U(\mathbf{r_1}, ..., \mathbf{r_N})]}{\int d\mathbf{r_1} ... d\mathbf{r_N} \exp[-\beta U(\mathbf{r_1}, ..., \mathbf{r_N})]}, \tag{5.30}$$

where $\beta = (k_B T)^{-1}$ and $U(\mathbf{r_1}, ..., \mathbf{r_N})$ is the Hamiltonian of the system.
As mentioned in Chapter 2, the joint distribution function for finding a particle at position $\mathbf{r_1}$ and any other particle at $\mathbf{r_2}$ is[27,30]

$$n_P^{(2)}(\mathbf{r_1}, \mathbf{r_2}) = N(N-1) \frac{\int d\mathbf{r_3} ... d\mathbf{r_N} \exp[-\beta U(\mathbf{r_1}, ..., \mathbf{r_N})]}{\int d\mathbf{r_1} ... d\mathbf{r_N} \exp[-\beta U(\mathbf{r_1}, ..., \mathbf{r_N})]}. \tag{5.31}$$

Therefore,

$$\begin{aligned} S(\mathbf{q}) &= 1 + \frac{1}{N} \int d\mathbf{r_1} d\mathbf{r_2}\, n_P^{(2)}(\mathbf{r_1}, \mathbf{r_2}) \exp[-i\mathbf{q} \cdot (\mathbf{r_1} - \mathbf{r_2})] \\ &= 1 + \frac{1}{N} \int d\mathbf{r_1} \int d\mathbf{r_2}\, n_P(\mathbf{r_1}) n_P(\mathbf{r_2})\, g^{(2)}(\mathbf{r_1}, \mathbf{r_2}) \exp[-i\mathbf{q} \cdot (\mathbf{r_1} - \mathbf{r_2})]. \end{aligned} \tag{5.32}$$

For an isotropic solution, $g(\mathbf{r_1}, \mathbf{r_2})$ depends only on $|\mathbf{r_2} - \mathbf{r_1}| = |\mathbf{r_{12}}| = r_{12}$, hence $g(\mathbf{r_1}, \mathbf{r_2}) = g(r_{12})$. After the convenient change of variables $(\mathbf{r_1}, \mathbf{r_2}) \rightarrow (\mathbf{r_{12}}, \mathbf{r_1})$, we obtain

$$\begin{aligned} S(\mathbf{q}) &= 1 + \frac{1}{N} \int d\mathbf{r_1} \int d\mathbf{r_{12}}\, n_P^2\, g(r_{12}) \exp[-i\mathbf{q} \cdot \mathbf{r_{12}}] \\ &= 1 + \frac{V}{N} \int d\mathbf{r_{12}}\, n_P^2\, g(r_{12}) \exp[-i\mathbf{q} \cdot \mathbf{r_{12}}] \\ &= 1 + n_P \int d\mathbf{r_{12}}\, g(r_{12}) \exp[-i\mathbf{q} \cdot \mathbf{r_{12}}]. \end{aligned} \tag{5.33}$$

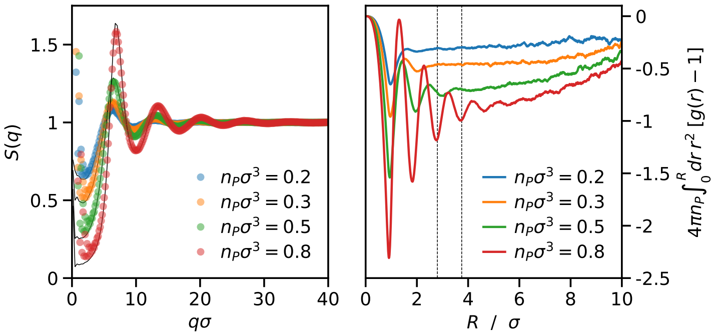

Figure 5.3: (*Left*) Structure factors, $S(q)$, of Lennard–Jones fluids of various densities, n_P, calculated from *NVT* Metropolis Monte Carlo simulations using Equation 5.20 (circles) and Equation 5.35 (lines). (*Right*) Kirkwood–Buff integrals vs the cutoff distance, R, calculated for the same systems using Equation 2.25. Dashed vertical lines indicate the R-interval over which the integrals are averaged to estimate the isothermal compressibilities of the fluids. Simulations are performed in the canonical ensemble using the software package Faunus.[54]

The forward scattering cannot be measured experimentally because forward scattered and transmitted photons are indistinguishable.[64] Furthermore a beamstop is used to absorb $I(0)$ to protect the detector. As a consequence, $S(q)$ is often defined as

$$\begin{aligned} S(\mathbf{q}) &= 1 + n_P \int \mathrm{d}\mathbf{r_{12}}\; g(r_{12}) \exp(-i\mathbf{q}\cdot\mathbf{r_{12}}) - n_P \int \mathrm{d}\mathbf{r_{12}}\; \exp(-i\mathbf{q}\cdot\mathbf{r_{12}}) \\ &= 1 + n_P \int \mathrm{d}\mathbf{r_{12}}\; [g(r_{12}) - 1] \exp(-i\mathbf{q}\cdot\mathbf{r_{12}}), \end{aligned} \tag{5.34}$$

where $n_P \int d\mathbf{r_{12}}\; \exp(-i\mathbf{q}\cdot\mathbf{r_{12}})$ corresponds to the radiation that passes through the sample unscattered.[15]

$S(q)$ can be integrated in spherical coordinates

$$\begin{aligned} S(q) &= 1 + n_P \int_0^\infty \mathrm{d}r\; r^2\; [g(r) - 1] \int_0^{2\pi} \mathrm{d}\gamma \int_0^\pi \mathrm{d}\alpha \sin\alpha \exp(-iqr\cos\alpha) \\ &= 1 + 4\pi n_P \int_0^\infty \mathrm{d}r\, r^2\; [g(r) - 1] \frac{\sin(qr)}{qr}. \end{aligned} \tag{5.35}$$

The structure factor extrapolated to $q = 0$ is a measure of the response in the number density to a change in osmotic pressure, Π. Using the series expansion of $\sin(qR)$ around $qr = 0$, we obtain

$$\begin{aligned} \lim_{q\to 0} S(q) &= \lim_{q\to 0} \left\{ 1 + \frac{4\pi n_P}{q} \int_0^\infty \mathrm{d}r\; r\; (qr - \frac{(qr)^3}{6} + ...)[g(r) - 1] \right\} \\ &= 1 + 4\pi n_P \int_0^\infty \mathrm{d}r\, r^2\; [g(r) - 1] = n_P \chi_{osm} k_B T, \end{aligned} \tag{5.36}$$

where χ_{osm} is the osmotic compressibility

$$\chi_{osm} = -\frac{1}{V}\left(\frac{\partial V}{\partial \Pi}\right)_T = \frac{1}{n_P}\left(\frac{\partial n_P}{\partial \Pi}\right)_T. \tag{5.37}$$

The derivation of Equation 5.36 is based on the relationship in the grand canonical ensemble between Kirkwood–Buff integral (KBI) and compressibility expressed by Equation 2.23 in Chapter 2.

For low particle concentrations, Π can be approximated to a truncated virial expansion

$$\begin{aligned} \frac{1}{n_P\chi_{osm}} &= \left(\frac{\partial \Pi}{\partial n_P}\right)_T = k_BT\left(\frac{\partial}{\partial n_P}[n_P + B_2n_P^2 + \mathcal{O}(n_P^3)]\right)_T \\ &= k_BT[1 + 2B_2n_P + \mathcal{O}(n_P^2)]. \end{aligned} \tag{5.38}$$

Therefore, at high dilution, the second virial coefficient, B_2, can be obtained from $S(0)$ via

$$S(0) \approx \frac{1}{1 + 2B_2n_P} \approx 1 - 2B_2n_P, \tag{5.39}$$

where, in the last step, we have used the series expansion of $(1 + 2B_2n_P)^{-1}$ around $n_P = 0$.

To conclude this Chapter, an example is given of the use of Metropolis Monte Carlo simulations[65] to calculate the RDF, the $S(q)$, as well as the isothermal compressibility of a model fluid. The left-hand side of Figure 5.3 shows structure factors obtained from Equation 5.20 and Equation 5.35 for Lennard–Jones fluids for closed systems of various densities in the canonical ensemble. The right-hand side of Figure 5.3 shows KBIs for the same systems as a function of the upper limit of integration, R. As discussed in Chapter 2, Equation 5.36 is valid only for open systems. A viable approach to accurately estimate $S(0)$ from a closed-system RDF is to average the KBI over one oscillation in the low-R region[32] (dashed vertical lines in Figure 5.3).

Chapter 6

Summary and Outlook

This Chapter summarizes key results from the papers included in the Thesis, as well as their future perspectives.

6.1 Paper I

Positioned at the end of the Hofmeister series, the thiocyanate ion binds to the NH and CH groups of the protein backbone,[3] and is one of the most surface-active small inorganic ions.[66] Previous SCN^- force fields failed to reproduce its orientational alignment at the interface. We developed a classical, all-atom force field for aqueous solutions of NaSCN and KSCN with unprecedented accuracy. Quantum-chemical calculations of the solvated anion were used to derive the charge distribution within SCN^-. The parameters describing the ions were adjusted to improve the agreement between experimental and calculated activity coefficients, obtained from molecular dynamics (MD) simulations using Kirkwood–Buff theory (see Chapter 2, Equation 2.37). The force field was shown to be suitable for simulations of the air/water interface by comparing experimental and simulated surface tension data, over a wide range of salt concentrations. Further, the force field was tested against available experimental data on reorientational dynamics and ion clustering in bulk solution.

The newly developed model suggests that K^+ preferentially interacts with the S atom of SCN^-, whereas Na^+ shows a stark affinity for the N atom. The difference in strength between K^+–S and Na^+–N interactions has notable consequences on the dynamic and structural properties of NaSCN and KSCN aqueous bulk solutions. Specifically, KSCN has a higher tendency to form ion pairs, whereas, large clusters are formed in concentrated NaSCN solutions resulting in slower SCN^- reorientational dynamics. On the contrary, the cation has a negligible influence on the properties of the air/water interface, wherein SCN^- molecules form an average angle of 44° with the normal to the surface, with the S atom pointing toward the vapor phase.

The current work tackles several limitations of empirical force fields employed in earlier studies. However, we observed that aqueous solutions of potassium salts underestimated experimental mass densities, and the surface tensions for KI and KCl deviated from experiments. In order to address this, reparametrization of the cation needs to be attempted.

6.2 Paper II

We further used the developed force field for SCN^- to study its surface-active behavior at air/water and hydrophobic interfaces. The simulations were complemented by vibrational sum frequency spectroscopy (VSFS) of monovalent salt solutions at the interface with a hydrophobic solid surface. Using VSFS, we probed non-hydrogen bonded OH groups of water in contact with the hydrophobic surface, to infer the relative amount of anions at the interface, as well as CN triple bonds with preferred orientation for direct estimation of SCN^-. SCN^- and I^- were found to be present in significant amounts in the outermost surface layer at the interface, whereas Cl^- and Br^- are expelled from the interface. Moreover, SCN^- was found to have a higher affinity for the alkane/water interface than I^-. MD simulations agreed with the accumulation of SCN^- in the outermost surface layer. However, we also found that the average concentration of SCN^- in the interfacial region is lower than in the bulk.

Both, simulations and experiments show that SCN^- forms an average angle of around 60° with the normal to the alkane/water interface, and the anion orientation is negligibly affected by bulk salt concentration. Using our simulations, we were also able to infer that the structure of the ion distribution at the interface is dictated by the strong Na^+–N interaction described in Paper I. Briefly, in the outermost surface layer, the S atom of SCN^- points toward the monolayer, while in the inner surface layers anions prefer the antiparallel orientation with respect to the normal to the interface. The two layers of roughly oppositely oriented SCN^- molecules are interposed by Na^+ ions.

The Langmuir model was used to analyze the VSFS data, and we found that SCN^- is attracted to the hydrophobic surface more than to the air/water interface. This trend was also found to be in agreement with MD simulation results. Further work for confirming the applicability of the Langmuir model would entail the calculation of 2D radial distribution functions for SCN molecules in the surface layers.
For a more quantitative discussion of the differences between air/water and monolayer/water interfaces, concentration profiles and orientational distributions should be calculated with respect to the instantaneous interface.

Further, simulations and experiments for KSCN at the alkane/water interface would be useful to confirm the weaker interaction of K^+ with the N atom, in comparison to Na^+. We expect a less pronounced antiparallel alignment for the SCN^- molecules in the deeper layer for KSCN solutions, which would result in more intense peaks for the CN triple bond.

6.3 Paper III

The focus of this study was the adsorption of amyloid aggregates onto oppositely charged lipid membranes. A combination of quartz crystal microbalance with dissipation monitoring (QCM-D, see Chapter 4), coarse-grained (CG) molecular simulations, and analytical theory was used to study the dependence of the adsorption of amyloid aggregates of different architecture on aggregate size and solution ionic strength. QCM-D experiments showed that $A\beta_{1-40}$ fibrils adsorb to positively charged lipid bilayers at low ionic strength, whereas no significant adsorption is observed at high ionic strength. Further, $A\beta_{1-40}$ peptides in monomeric form do not adsorb to positively charged lipid bilayers.

We explained the experimental results using Monte Carlo (MC) simulations for rigid aggregates of 2 to 100–200 monomers interacting with a surface of homogeneous charge density. The CG aggregates were constructed using a bottom-up approach, building on the available atomic-resolution structures and representing each amino acid as a bead with a certain charge and radius. The free energy of interaction between aggregate and surface was calculated for aggregates of various lengths and solutions of different ionic strengths. This data was used to develop an analytical theory accounting for the entropy loss associated with the aggregates approaching the surface, wherein fibrils are represented by line segments with homogenous charge distribution. Using generalized van der Waals theory,[67] the line-segment model was used to predict the amount of aggregates adsorbed onto the surface.

Our simulations showed that the aggregate–surface interaction is governed by the interplay between electrostatic attraction and entropic repulsion due to excluded volume. This determines a complex dependence of the adsorbed amount on the ionic strength of the bulk solution and on aggregate size. Notably, we observed preferential adsorption of smaller aggregates over larger ones at physiological ionic strength and pH. Further, the orientation of the aggregates on the surface was found to be dependent on aggregate size and ionic strength.

Besides providing new insight into the interaction between amyloid aggregates and membranes, our findings describe general physical principles pertaining to the interaction of elongated, rigid particles with oppositely charged interfaces. The CG model could be extended by introducing discrete representations of multivalent ions, and laterally diffusing charge sites on the surface. Additionally, it would be interesting to verify the predictions of our model, especially those regarding the orientation at the interface, on other systems consisting of rod-like particles and oppositely charged surfaces.

6.4 Paper IV

Cell-penetrating peptides (CPPs) and antimicrobial peptides (AMPs) are oligopeptides often rich in basic amino acids. To fully exploit their potential as candidates for intracellular drug delivery and treatment of infections, it is crucial to understand the interaction of short cationic peptides with lipid bilayers. Unfortunately, experimental techniques seldom provide highly resolved molecular-level information, and simulations of peptide–membrane interactions require large length scales and long equilibration times, especially when aimed at elucidating the concerted action of several peptides. Therefore, we developed a computationally efficient CG force field and outlined a convenient strategy to parametrize it against all-atom simulations as well as experimental data.

The CG force field is based on the Cooke lipid model,[68] which reproduces structural and mechanical properties of lipid bilayers. In our extended model, lipids and peptides are described as connected beads of given radius, charge, and excess polarizability embedded in a dielectric continuum. The excess polarizabilities quantify the difference in dielectric properties between the beads and the surrounding aqueous medium, and determine van der Waals and charge–induced dipole interactions between peptide and lipid beads. The parametrization of the model is targeted toward all-atom free energies for the monopeptide at the membrane interface and midplane, and a continuum model is developed to map the reference data to the parameters of the CG force field.

Biased sampling techniques (see Chapter 3) are used to calculate free energy profiles of translocation for protonated and neutral histidine, lysine, and arginine monopeptides, using the CG and the all-atom force fields. CG and all-atom models display similar deformations of the lipid membrane upon monopeptide permeation. Further, free energy maxima and minima of translocation profiles of CG di- and triarginine molecules are in agreement with reference all-atom data. Using constant-pH MC simulations, we demonstrate that the CG model can be used to predict the protonation states of mono-, di-, and tripeptides as a function of the depth of membrane penetration.

The model can be further applied to the investigation of the mechanism of action of CPPs and AMPs. Specifically, it can contribute to explain the chain-length dependence of the translocation efficiency of oligo-arginines, which is expected to be maximal for 6–15 residues.[69,70] Moreover, after parametrization of peptide–peptide interactions, the model could be employed to perform large-scale simulations involving the concurrent interaction of several peptides with a lipid membrane.

6.5 Paper V

Arginine-rich CPPs (RRPs) can traverse biological membranes via a direct, non-endocytotic mode.[71] It has been experimentally shown that the efficiency of cellular uptake of RRPs can be related to the formation of aggregates in solution. Simulation studies have suggested that RRPs can aggregate on the lipid bilayer surface, giving way to a cooperative internalization process. In this study, we present the non-trivial solution behavior of deca-arginine (R10), in comparison with deca-lysine (K10). Small angle X-ray scattering (SAXS, see Chapter 5) shows that counterintuitively, R10 self-associates at low-to-intermediate ionic strengths, despite its high positive charge. Conversely, K10 displays the expected solution behavior of polyelectrolytes. We used all-atom umbrella sampling MD simulations (see Chapter 3) and calculated free energy profiles of interaction between R10 and K10. The results are shown to be consistent with the trends observed in SAXS measurements. We found that the R10–R10 interaction is attributed to the stacking of the positively charged guanidinium (Gdm^+) side chains of the ninth residues, which favorably interact with the negatively charged C-terminal carboxyl groups. Inspection of the Protein Data Bank revealed that the double salt bridge between a like-charged Gdm^+ ion pair and two carboxyl groups is a common motif in protein crystal structures.

To confirm the importance of the ninth arginine residue, we also looked at R8KR and K8RK. SAXS and simulations showed that these peptides do not self-associate. However, R8KR–R8KR interactions are significantly less repulsive than K10–K10 interactions. These results suggest that, beside the double salt bridge involving the ninth residues, self-association requires additional Gdm^+–Gdm^+ attractive interactions between other residues. This non-specific pairing between arginine side chains has also been corroborated using HSQC NMR. As mentioned previously, the efficiency of uptake by cells is optimal for oligo-arginines of 6–15 residues.[69,70] Our findings suggest that the low bioavailability of shorter and longer peptide chains might be related to the suppression of self-association, due to either the lack of attractive Gdm^+–Gdm^+ interactions, or the build-up of net positive charge of the peptide.

It would be interesting to systematically test the dependence of oligo-arginine attraction on the chain length, *e.g.*, via SAXS and simulations of solutions of oligo-arginines of 8–20 residues. Moreover, to confirm the role of the double salt bridge, solutions of C-terminally amidated oligo-arginines could be investigated. The detailed information on oligo-arginine interactions acquired form SAXS and all-atom simulations could additionally be used to develop a CG peptide model to investigate, within the force field of Paper IV, the influence of self-association on membrane adsorption and translocation.

6.6 Paper VI

This paper presents an overview of the research on Gdm^+ ion pairing focusing on its implications for the mechanism of action of RRPs. Gdm^+ is the smallest cation that is able to form like-charge ion pairs in aqueous solution. Gdm^+ can be considered as an orientational amphiphile, owing to its planar structure with hydrophobic faces and the ability to form in-plane H bonds with water molecules. Like SCN^-, Gdm^+ is placed at the extreme end of the Hofmeister series, and GdmCl is a widely used protein denaturant. As previously mentioned, Gdm^+–Gdm^+ pairing plays crucial roles in protein structure and function, especially when further stabilized by a double salt bridge with two carboxyl groups.

MD simulations and quantum-chemical calculation of hydrated Gdm^+ dimers have shown that Gdm^+–Gdm^+ interactions are thermodynamically stable, owing to the compensation of the like-charge repulsion by cavitation, dispersion, and quadrupole–quadrupole attractions. Electrophoresis experiments and MD simulations indicate that Gdm^+ ions form like-charge ion pairs with arginine side chains, and the work of Paper V has further shown that ion paring also occurs between the arginine side chains of RRPs.

At dilute extracellular peptide concentrations, the internalization of RRPs has been shown to occur via direct translocation, whereas, at high peptide concentrations, experiments suggest a *transduction* mechanism involving the local aggregation of RRPs adsorbed onto the plasma membrane.[71] Comparative studies conducted using MD simulations, as well as experiments, have shown that polyarginines bind considerably more strongly than polylysines onto lipid bilayers, and that the interactions between adsorbed peptides are attractive for the former and repulsive for the latter. Further, MD simulations have suggested that cooperative actions of the self-associating RRPs can induce kinetic stabilization of membrane transient pores.

The findings reviewed in this paper may contribute to design new cell-penetrating peptides and to optimize their efficacy as intracellular delivery vectors.

References

[1] Jungwirth, P.; Tobias, D. J. Specific Ion Effects at the Air/Water Interface. *Chemical Reviews* **2006**, *106*, 1259–1281.

[2] Tobias, D. J.; Stern, A. C.; Baer, M. D.; Levin, Y.; Mundy, C. J. Simulation and Theory of Ions at Atmospherically Relevant Aqueous Liquid–Air Interfaces. *Annual Review of Physical Chemistry* **2013**, *64*, 339–359.

[3] Okur, H. I.; Hladílková, J.; Rembert, K. B.; Cho, Y.; Heyda, J.; Dzubiella, J.; Cremer, P. S.; Jungwirth, P. Beyond the Hofmeister Series: Ion-Specific Effects on Proteins and Their Biological Functions. *The Journal of Physical Chemistry B* **2017**, *121*, 1997–2014.

[4] Pabst, G.; Hodzic, A.; Štrancar, J.; Danner, S.; Rappolt, M.; Laggner, P. Rigidification of Neutral Lipid Bilayers in the Presence of Salts. *Biophysical Journal* **2007**, *93*, 2688–2696.

[5] Friedman, R. Membrane–Ion Interactions. *The Journal of Membrane Biology* **2018**, *251*, 453–460.

[6] Hofmeister, F. Zur Lehre von der Wirkung der Salze. *Archiv für Experimentelle Pathologie und Pharmakologie* **1888**, *24*, 247–260.

[7] Deserno, M. Mesoscopic Membrane Physics: Concepts, Simulations, and Selected Applications. *Macromolecular Rapid Communications* **2009**, *30*, 752–771.

[8] Menichetti, R.; Kremer, K.; Bereau, T. Efficient Potential of Mean Force Calculation from Multiscale Simulations: Solute Insertion in a Lipid Membrane. *Biochemical and Biophysical Research Communications* **2018**, *498*, 282–287.

[9] Fändrich, M. On the Structural Definition of Amyloid Fibrils and Other Polypeptide Aggregates. *Cellular and Molecular Life Sciences* **2007**, *64*, 2066–2078.

[10] Milanesi, L.; Sheynis, T.; Xue, W.-F.; Orlova, E. V.; Hellewell, A. L.; Jelinek, R.; Hewitt, E. W.; Radford, S. E.; Saibil, H. R. Direct Three-dimensional Visualization of Membrane Disruption by Amyloid Fibrils. *Proceedings of the National Academy of Sciences* **2012**, *109*, 20455–20460.

[11] Friedman, R.; Pellarin, R.; Caflisch, A. Amyloid Aggregation on Lipid Bilayers and Its Impact on Membrane Permeability. *Journal of Molecular Biology* **2009**, *387*, 407–415.

[12] Bechara, C.; Sagan, S. Cell-penetrating Peptides: 20 Years Later, Where Do We Stand? *FEBS Letters* **2013**, *587*, 1693–1702.

[13] Sun, D.; Forsman, J.; Lund, M.; Woodward, C. E. Effect of Arginine-rich Cell Penetrating Peptides on Membrane Pore Formation and Life-times: a Molecular Simulation Study. *Phys. Chem. Chem. Phys.* **2014**, *16*, 20785–20795.

[14] Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions. I. *The Journal of Chemical Physics* **1951**, *19*, 774–777.

[15] Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 4th ed.; Academic Press: Amsterdam, 2013; pp 105–110.

[16] Ben-Naim, A. *Water and Aqueous Solutions: Introduction to a Molecular Theory*, 1st ed.; Plenum Press: New York, 1974; pp 29–80.

[17] Newman, K. E. Kirkwood–Buff Solution Theory: Derivation and Applications. *Chem. Soc. Rev.* **1994**, *23*, 31–40.

[18] Ben-Naim, A. *Water and Aqueous Solutions: Introduction to a Molecular Theory*, 1st ed.; Plenum Press: New York, 1974; pp 123–144.

[19] Mamatkulov, S.; Fyta, M.; Netz, R. R. Force Fields for Divalent Cations Based on Single-ion and Ion-pair Properties. *The Journal of Chemical Physics* **2013**, *138*, 024505.

[20] van der Vegt, N. F. A.; Haldrup, K.; Roke, S.; Zheng, J.; Lund, M.; Bakker, H. J. Water-Mediated Ion Pairing: Occurrence and Relevance. *Chem. Rev.* **2016**, *116*, 7626–7641.

[21] Krishnamoorthy, A. N.; Holm, C.; Smiatek, J. Influence of Cosolutes on Chemical Equilibrium: a Kirkwood–Buff Theory for Ion Pair Association–Dissociation Processes in Ternary Electrolyte Solutions. *The Journal of Physical Chemistry C* **2018**, *122*, 10293–10302.

[22] van der Vegt, N. F. A.; Nayar, D. The Hydrophobic Effect and the Role of Cosolvents. *The Journal of Physical Chemistry B* **2017**, *121*, 9986–9998.

[23] Ben-Naim, A. Inversion of the Kirkwood–Buff Theory of Solutions: Application to the Water–Ethanol System. *The Journal of Chemical Physics* **1977**, *67*, 4884–4890.

[24] Shimizu, S. Estimating Hydration Changes Upon Biomolecular Reactions from Osmotic Stress, High pressure, and Preferential Hydration Experiments. *Proceedings of the National Academy of Sciences* **2004**, *101*, 1195–1199.

[25] Calero-Rubio, C.; Strab, C.; Barnett, G. V.; Roberts, C. J. Protein Partial Molar Volumes in Multicomponent Solutions from the Perspective of Inverse Kirkwood–Buff Theory. *The Journal of Physical Chemistry B* **2017**, *121*, 5897–5907.

[26] Hill, T. L. *An Introduction to Statistical Thermodynamics*, 1st ed.; Dover: New York, 1986; pp 112–123.

[27] Chandler, D. *Introduction to Modern Statistical Mechanics*, 1st ed.; Oxford University Press: New York, 1987; pp 195–209.

[28] Ben-Naim, A. *Molecular Theory of Water and Aqueous Solutions: Understanding Water*, 1st ed.; World Scientific: Singapore, 2009; pp 67–75.

[29] Hill, T. L. *An Introduction to Statistical Thermodynamics*, 1st ed.; Dover: New York, 1986; pp 36–37.

[30] Ben-Naim, A. *Water and Aqueous Solutions: Introduction to a Molecular Theory*, 1st ed.; Plenum Press: New York, 1974; pp 81–113.

[31] Krüger, P.; Vlugt, T. J. H. Size and Shape Dependence of Finite-volume Kirkwood–Buff Integrals. *Physical Review E* **2018**, *97*.

[32] Milzetti, J.; Nayar, D.; van der Vegt, N. F. A. Convergence of Kirkwood–Buff Integrals of Ideal and Nonideal Aqueous Solutions Using Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2018**, *122*, 5515–5526.

[33] Ganguly, P.; van der Vegt, N. F. A. Convergence of Sampling Kirkwood–Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 1347–1355.

[34] Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Vlugt, T. J. H.; Simon, J.-M. Kirkwood–Buff Integrals for Finite Volumes. *J. Phys. Chem. Lett.* **2013**, *4*, 235–238.

[35] Hess, B.; van der Vegt, N. F. A. Cation Specific Binding with Protein Surface Charges. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13296–13300.

[36] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, 2nd ed.; Oxford University Press: Oxford, 2017; pp 46–94.

[37] Trzesniak, D.; Kunz, A.-P. E.; van Gunsteren, W. F. A Comparison of Methods to Compute the Potential of Mean Force. *ChemPhysChem* **2007**, *8*, 162–169.

[38] Paci, E.; Ciccotti, G.; Ferrario, M.; Kapral, R. Activation Energies by Molecular Dynamics with Constraints. *Chemical Physics Letters* **1991**, *176*, 581–587.

[39] Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.

[40] Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Computer Physics Communications* **1995**, *91*, 275–282.

[41] Kästner, J. Umbrella Sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 932–942.

[42] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.

[43] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, 2nd ed.; Oxford University Press: Oxford, 2017; pp 297–341.

[44] Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

[45] Wang, F.; Landau, D. P. Determining the Density of States for Classical Statistical Models: A Random Walk Algorithm to Produce a Flat Histogram. *Physical Review E* **2001**, *64*.

[46] Wang, F.; Landau, D. P. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Physical Review Letters* **2001**, *86*, 2050–2053.

[47] Berg, B. A.; Celik, T. New Approach to Spin-glass Simulations. *Physical Review Letters* **1992**, *69*, 2292–2295.

[48] Lee, J. New Monte Carlo Algorithm: Entropic Sampling. *Physical Review Letters* **1993**, *71*, 211–214.

[49] Hunter, J. E.; Reinhardt, W. P. Finite-size Scaling Behavior of the Free Energy Barrier Between Coexisting Phases: Determination of the Critical Temperature and Interfacial Tension of the Lennard-Jones Fluid. *The Journal of Chemical Physics* **1995**, *103*, 8627–8637.

[50] Engkvist, O.; Karlström, G. A Method to Calculate the Probability Distribution for Systems with Large Energy Barriers. *Chemical Physics* **1996**, *213*, 63–76.

[51] Calvo, F. Sampling Along Reaction Coordinates with the Wang–Landau Method. *Molecular Physics* **2002**, *100*, 3421–3427.

[52] Khan, M. O.; Kennedy, G.; Chan, D. Y. C. A Scalable Parallel Monte Carlo Method for Free Energy Simulations of Molecular Systems. *Journal of Computational Chemistry* **2004**, *26*, 72–77.

[53] Vogel, T.; Li, Y. W.; Wüst, T.; Landau, D. P. Generic, Hierarchical Framework for Massively Parallel Wang-Landau Sampling. *Physical Review Letters* **2013**, *110*.

[54] Stenqvist, B.; Thuresson, A.; Kurut, A.; Vácha, R.; Lund, M. *Faunus*—a Flexible Framework for Monte Carlo Simulation. *Molecular Simulation* **2013**, *39*, 1233–1239.

[55] Sauerbrey, G. Verwendung von Schwingquarzen zur Wägung dünner Schichten und zur Mikrowägung. *Zeitschrift für Physik* **1959**, *155*, 206–222.

[56] Johannsmann, D. *The Quartz Crystal Microbalance in Soft Matter Research: Fundamentals and Modeling*; Springer International Publishing: Cham, 2015; pp 1–22.

[57] Höök, F.; Kasemo, B. In *Piezoelectric Sensors*; Janshoff, A., Steinem, C., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp 425–447.

[58] Blitz, J. *Fundamentals of Ultrasonics*, 1st ed.; Butterworths: London, 1963; pp 41–44.

[59] Johannsmann, D. *The Quartz Crystal Microbalance in Soft Matter Research: Fundamentals and Modeling*; Springer International Publishing: Cham, 2015; pp 23–31.

[60] Voinova, M. V.; Rodahl, M.; Jonson, M.; Kasemo, B. Viscoelastic Acoustic Response of Layered Polymer Films at Fluid-Solid Interfaces: Continuum Mechanics Approach. *Physica Scripta* **1999**, *59*, 391–396.

[61] Johannsmann, D. *The Quartz Crystal Microbalance in Soft Matter Research: Fundamentals and Modeling*; Springer International Publishing: Cham, 2015; pp 33–47.

[62] Li, T.; Senesi, A. J.; Lee, B. Small Angle X-ray Scattering for Nanoparticle Research. *Chemical Reviews* **2016**, *116*, 11128–11180.

[63] Jacques, D. A.; Trewhella, J. Small-angle Scattering for Structural Biology—Expanding the Frontier while Avoiding the Pitfalls. *Protein Science* **2010**, *19*, 642–657.

[64] Halle, B. Lecture Notes on the Theory of Small-angle X-ray Scattering for Nuclear Spin Relaxers and Other Novices. 2013.

[65] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **1953**, *21*, 1087–1092.

[66] Mizuno, H.; Rizzuto, A. M.; Saykally, R. J. Charge-Transfer-to-Solvent Spectrum of Thiocyanate at the Air/Water Interface Measured by Broadband Deep Ultraviolet Electronic Sum Frequency Generation Spectroscopy. *The Journal of Physical Chemistry Letters* **2018**, 4753–4757.

[67] Nordholm, S.; Woodward, C.; Forsman, J.; Freasier, B.; Penfold, R.; Abbas, Z. *Generalized van der Waals Theory of Molecular Fluids in Bulk and at Surfaces*; Elsevier, 2018.

[68] Cooke, I. R.; Deserno, M. Solvent-free Model for Self-assembling Fluid Bilayer Membranes: Stabilization of the Fluid Phase Based on Broad Attractive Tail Potentials. *The Journal of Chemical Physics* **2005**, *123*, 224710.

[69] Mitchell, D.; Steinman, L.; Kim, D.; Fathman, C.; Rothbard, J. Polyarginine Enters Cells More Efficiently than Other Polycationic Homopolymers. *Journal of Peptide Research* **2000**, *56*, 318–325.

[70] Futaki, S.; Suzuki, T.; Ohashi, W.; Yagami, T.; Tanaka, S.; Ueda, K.; Sugiura, Y. Arginine-rich Peptides. *Journal of Biological Chemistry* **2000**, *276*, 5836–5840.

[71] Brock, R. The Uptake of Arginine-Rich Cell-Penetrating Peptides: Putting the Puzzle Together. *Bioconjugate Chemistry* **2014**, *25*, 863–868.

Paper I

G. Tesei, V. Aspelin, and M. Lund
The Journal of Physical Chemistry B, **122**, 19, 5094–5105 (2018)
https://www.doi.org/10/cphm

Copyright © 2018, American Chemical Society.
Reprinted with permission.

Paper II

E. Potapova, G. Tesei, M. Lund, and E. Tyrode
Manuscript (2018)

Paper III

G. Tesei, E. Hellstrand, K. Sanagavarapu, S. Linse, E. Sparr,
R. Vácha, and M. Lund
Langmuir, **34**, 4, 1266–1273 (2018)
https://www.doi.org/10/chz9
Copyright © 2018, American Chemical Society.
Reprinted with permission.

Paper IV

G. Tesei, M. Vazdar, and M. Lund
Manuscript (2018)

Paper V

G. Tesei, M. Vazdar, M. Ringkjøbing Jensen, C. Cragnell, P. E. Mason, J. Heyda,
M. Skepö, P. Jungwirth, and M. Lund
Proceedings of the National Academy of Sciences, **114**, 43, 11428–11433 (2017)
DOI: https://www.doi.org/10/cd5n
Copyright © 2017, National Academy of Sciences.
Reprinted with permission.

Paper VI

M. Vazdar, J. Heyda, P. E. Mason, G. Tesei, C. Allolio,
M. Lund, and P. Jungwirth
Accounts of Chemical Research, **51**, 6, 1455–1464 (2018)
https://www.doi.org/10/gdsqqd
Copyright © 2018, American Chemical Society.
Reprinted with permission.